



Université Paris Cité
Laboratoire de Physique Nucléaire et de Hautes Energies (LPNHE) (UMR 7585)
Ecole Doctorale de Sciences de la Terre et de l'Environnement et Physique de l'Univers de
Paris (560)

Identification of boosted Higgs bosons for new physics searches and improvement of the ATLAS tracker for the high luminosity phase of LHC

Par YAJUN HE

Thèse de doctorat en PHYSIQUES DE L'UNIVERS
Dirigée par Giovanni CALDERINI et par Reina CAMACHO TORO

Présentée et soutenue publiquement le 27 Octobre 2022

Devant un jury composé de:

Mme	Daniela BORTOLETTO	Professeur	Oxford University	Rapportrice
Mme	Anne Catherine LE BIHAN	CR-HDR	Université de Strasbourg	Rapportrice
M.	Valerio DAO	Chercheur	CERN	Examineur
Mme	Fabienne LEDROIT	DR	Université Grenoble Alpes	Examinatrice
M.	Giovanni CALDERINI	DR	Université Paris Cité	Directeur
Mme	Reina CAMACHO TORO	CR (ADT)	CNRS	Co-directrice



未一。

《后汉书·陈元传》

ACKNOWLEDGEMENT

First and foremost, my deepest gratitude goes to my supervisors, Giovanni Calderini and Reina Camacho Toro. Throughout my PhD journey, they have provided continuous guidance and endless support, and I would not have arrived at the end without their patience. Reina has always been there for all my questions and doubts about physics and life. Giovanni has guided my detector work and lightning me when I was lost.

In addition to my supervisors, this thesis is a culmination of the ideas and work of a brilliant team of physicists who I had the pleasure to work with. During the development of $X \rightarrow b\bar{b}$ calibration, part of the work has been shared with Shubham Bansal, Valerio Dao and Tatjana Lenz. The improvements have been made together with Daariimaa Bat-tulga, Ligang Xia and Jinchao Zheng. A lot of discussions were made with Arely Cortes Gonzalez and Eleanor Jones. The measurement of the $V(\rightarrow qq')H(\rightarrow b\bar{b})$ production has been shared with Viviana Cavaliere, Caterina Vernieri and Zhi Zheng. Much gratitude also goes to Haider Abidi, Andrea Sciandra and Stephane Willocq, who have regularly advised and supervised the progress of this analysis. I am indebted to Marco Bomben, Francesco Crescioli, Stefano Terzo and the testbeam team for their support on the detector work.

My past three years at LPNHE have been a pleasure, and much of this is owed to the ATLAS group at this lab. My time in Paris would not have been complete without the friends I have made at this lab and in the past three years. Thank you for the beers (non-alcohol;)), karaoke, hotpot and bouldering sessions throughout my time at Paris.

I would like to extend gratitude to my loving parents and sisters who have tolerated my stubbornness yet supported every decision I made in my life, including going abroad and pursuing a PhD. Last but not least, words cannot describe how grateful I am to my boyfriend, Da Yu Tou, who helped me maintained my sanity throughout the craziness that is 2022. Thank you for being the light during my PhD.

Now, onto formalities. This thesis is funded by the IDEX program of Universite Paris-Cite from 2019 to 2022.

CONTENTS

Acknowledgement	iii
List of figures	xvii
List of tables	xix
Abstract	1
Résumé	3
Résumé	5
Introduction	9
1 Theoretical framework	11
1.1 The Standard Model (SM)	11
1.1.1 Quantum electrodynamics (QED)	12
1.1.2 The electroweak unification	14
1.1.3 Quantum chromodynamics (QCD)	16
1.1.4 Spontaneous symmetry breaking (SSB) and the Higgs mechanism .	19
1.1.5 Limitations of the Standard Model	22
1.2 The Higgs boson at the Large Hadron Collider	22
1.2.1 Production modes	23
1.2.2 Decay modes	23
1.2.3 Boosted Higgs bosons	26
1.3 Summary	28
2 The ATLAS detector	29
2.1 The Large Hadron Collider (LHC)	29
2.2 The ATLAS detector	32
2.2.1 The inner detector	33
2.2.2 The calorimetry system	37
2.2.3 The muon system	40
2.2.4 The magnet system	42
2.2.5 The forward detectors	42
2.3 Data acquisition	44
2.3.1 The hardware trigger system: Level 1	45
2.3.2 The software trigger system: High Level Trigger	45
2.4 Summary	47
3 Event reconstruction	49
3.1 Track reconstruction	49
3.1.1 Primary tracks	51
3.1.2 Secondary tracks	52
3.2 Vertex reconstruction	53
3.3 Jets	53

3.3.1	Jet inputs	53
3.3.2	Jet algorithms	55
3.3.3	Jet grooming	56
3.3.4	Jet calibration	57
3.3.5	Jet truth labelling	58
3.4	b -tagging	59
3.4.1	Low-level b -tagging algorithms	60
3.4.2	High-level b -tagging algorithms	61
3.4.3	b -tagging performance	62
3.4.4	b -tagging calibration	62
3.5	Electrons	63
3.5.1	Electron reconstruction	63
3.5.2	Electron identification	63
3.5.3	Electron isolation	64
3.5.4	Electron-charge identification	64
3.5.5	Electron efficiency	65
3.6	Muons	65
3.6.1	Muon reconstruction	65
3.6.2	Muon identification	66
3.6.3	Muon isolation	67
3.6.4	Muon efficiency	67
3.7	Overlap removal	67
3.8	Summary	67
4	In-situ calibration of the $X \rightarrow b\bar{b}$ tagger using Z + jets events	69
4.1	Boosted $H \rightarrow b\bar{b}$ tagging	69
4.1.1	Double b -tagging method	70
4.1.2	Boosted $X \rightarrow b\bar{b}$ tagger	70
4.2	Methodology	71
4.3	Measurement of $\mu^{post-tag}$	74
4.3.1	Data and Monte Carlo simulations	74
4.3.2	Event selection	75
4.3.3	Inclusive $Z \rightarrow q\bar{q}$ template	77
4.3.4	Background modelling	78
4.3.5	Systematic uncertainties	81
4.3.6	Statistical model	82
4.3.7	Results	83
4.4	Measurement of $\mu^{pre-tag}$	83
4.4.1	Data and Monte Carlo simulations	83
4.4.2	Event selection	85
4.4.3	Systematic uncertainties	86
4.4.4	Results	87
4.5	Calibration results: scale factors	87
4.6	Moving to a binned template fit	88
4.7	Summary and outlook	90

5	Measurement of the $V(\rightarrow qq')H(\rightarrow b\bar{b})$ production at high transverse momentum	93
5.1	Overview	93
5.2	Analysis strategy	96
5.3	Data and MC simulations	97
5.4	Event selection	98
5.4.1	VH assignment	98
5.4.2	Signal, control and validation region definition	100
5.4.3	Significance studies	100
5.5	$X \rightarrow b\bar{b}$ tagger calibration	104
5.6	Background modelling	107
5.6.1	BDT reweighting strategy	107
5.6.2	BDT systematic uncertainties	109
5.6.3	SR template tests	112
5.7	Systematic uncertainties	122
5.7.1	Experimental systematic uncertainties	122
5.7.2	Modelling systematic uncertainties	122
5.8	Summary	123
6	Silicon sensors for High Energy Physics	125
6.1	Semiconductor basics	125
6.1.1	Charge carrier concentration	127
6.1.2	The pn-junction	128
6.1.3	Signal formation	130
6.2	Pixel detectors	131
6.2.1	Planar pixel sensors	132
6.2.2	3D pixel sensors	134
6.2.3	Monolithic pixel sensors	134
6.3	Radiation damage	135
6.4	Summary	136
7	Planar pixel modules for the ATLAS Inner Tracker (ITk)	139
7.1	ATLAS Inner Tracker (ITk)	139
7.2	LPNHE within the ITk project (and personal contributions)	141
7.3	Planar pixel module	142
7.3.1	RD53A readout chip	142
7.3.2	Sensor layouts	142
7.3.3	Hybridisation	143
7.3.4	Module assembly	143
7.4	Characterisation of planar pixel modules	143
7.4.1	I-V curve measurement	144
7.4.2	Pixel module calibration	147
7.4.3	Test-beam campaigns for characterisation	148
7.5	Test-beam results for planar pixel modules	153
7.5.1	LPHNE R&D modules	154
7.5.2	ITk market survey modules	155
7.5.3	AIDA active edge modules	158

7.6 Summary	162
Conclusion	163
Appendix	165
A Results of $\mu^{post-tag}$ and $\mu^{pre-tag}$	167
B SHERPA 2.2.11 V + jets samples	169
C Fully hadronic VH analysis	171
C.1 Data and MC comparisons	171
C.2 $X \rightarrow b\bar{b}$ tagger calibration	180
C.3 Transfer factor method	181
References	196

LIST OF FIGURES

1.1	The Standard Model of elementary particles [14].	12
1.2	Measurements of α_s as a function of the energy scale Q [21].	18
1.3	Main leading order Feynman diagrams contributing to the Higgs boson production in (a) the gluon fusion, (b) the vector-boson fusion, (c) the associated production with a gauge boson at tree level from a quark-quark interaction, (d) the associated production with a gauge boson (at loop level from a gluon-gluon interaction), (e) the associated production with a pair of top quarks (there is a similar diagram for the associated production with a pair of bottom quarks), (f-g) the associated production with a single top quark.	24
1.4	(a) Production cross sections for a SM Higgs boson with a mass of GeV from the proton-proton collisions at the LHC [38]. The theoretical uncertainties are indicated as bands. (b) The branching ratios for the main decays of the SM Higgs boson near $m_H = 125$ GeV. The theoretical uncertainties are indicated as bands.	24
1.5	Transverse momentum distribution of the Higgs boson at the LHC with $\sqrt{s} = 13$ TeV. The upper panel shows absolute predictions at LO and NLO in the full SM and in the infinite top-mass approximation (HEFT). The lower panel shows respective NLO/LO correction factors. The bands indicate theoretical errors of the full SM result due to scale variation [8].	27
1.6	Impact of the chromomagnetic operator on the p_T spectrum of the Higgs boson in the region allowed by the current experimental constraints [9].	27
2.1	Overview of the CERN accelerator complex [42].	30
2.2	LHC/HL-LHC upgrade plan [43].	31
2.3	Integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018 [44].	32
2.4	Luminosity-weighted distribution of the mean number of interactions per bunch crossing for the different years of operations during Run 2. The mean number of interactions per crossing is calculated from the instantaneous per bunch luminosity as $\mu = \mathcal{L}_{bunch} \times \sigma_{inel} / f_r$ where $\mathcal{L}_{bunch}$ is the per bunch instantaneous luminosity, σ_{inel} is the inelastic cross section which we take to be 80 mb for 13 TeV collisions, and f_r is the LHC revolution frequency [44].	33
2.5	(a) Schematic of the ATLAS detector including the relevant sub-detectors. (b) ATLAS system of coordinates. [46]	33
2.6	Schematic view of the interaction of different particles with the ATLAS detector [47].	34
2.7	Schematic overview of the Inner Detector [48].	35
2.8	A 3D visualisation of the structure of the ID barrel: the beam pipe, the Insertable B Layer (IBL), the Pixel layers of the Pixel Detector, the four cylindrical layers of the SCT and the three layers of the TRT barrel modules [49].	35
2.9	Schematic view of the Pixel Detector [50].	36
2.10	Schematic view of the Semiconductor Tracker (SCT) [51].	37
2.11	Schematic view of the ATLAS calorimetry system [53].	38

2.12	The ATLAS muon system [62].	41
2.13	The ATLAS magnet system [66].	42
2.14	Layout of the ATLAS Forward Detectors [72].	44
2.15	Layout of the ATLAS Trigger and Data Acquisition system in Run 2 data-taking [73].	44
2.16	An overview of the L1 Muon trigger chambers [76].	46
3.1	The ATLAS tracking flow chart [90].	50
3.2	The perigee representation [90].	50
3.3	An overview of the calibration chain for both the small- R and the large- R jets is shown. [108]	57
3.4	Graphical representation of b -jet (left), c -jet (middle) and l -jet (right). PV: Primary Vertex. IP: Impact parameter. SV: Secondary Vertex.	60
3.5	Distribution of the output discriminant of MV2 (left) and DL1 (right) for b -jets, c -jets and l -jets in the baseline $t\bar{t}$ simulated events. [112]	61
3.6	The c -jets (left) and l -jets (right) rejection as function of the b -tagging efficiency [116].	62
3.7	A schematic illustration of the path of an electron through the detector. The red trajectory shows the hypothetical path of an electron. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system. [121]	64
4.1	Efficiency of double subjet b -labelling at the truth level of a Higgs jet as a function of the Higgs jet p_T . The efficiency plots are for VR track jets with $\rho = 30$ GeV and $R_{min} = 0.02$ for varying values of R_{max} . The Higgs bosons are obtained from the BSM physics simulation of a Randall-Sundrum graviton. Details can be found in [106].	70
4.2	Pictorial representation of ordinary quark and gluon jets (top left), b -jets (top center), and boosted-jet topologies, emerging from high- p_T W and Z bosons (top right), Higgs bosons (bottom left), and top quarks (bottom right) decaying to all-quark final states. [127]	71
4.3	Dijet (left) and top (right) jet rejection as a function of large- R jet p_T with a constant $H \rightarrow b\bar{b}$ tagging efficiency of 60% [126] using the $X \rightarrow b\bar{b}$ tagger (blue), the double b -tagging method with DL1r (green), the double b -tagging method with MV2 (dark gray) and the double b -tagging method with MV2 but using fixed radius ($R = 0.2$) jets (light gray).	72
4.4	Tagging efficiency of the $X \rightarrow b\bar{b}$ tagger for $Z \rightarrow b\bar{b}$ (red) and $H \rightarrow b\bar{b}$ (blue) events as a function of p_T at working point 60%. The large- R jets used are the Higgs candidate jets after the selection described in Section 5.4. The $Z \rightarrow b\bar{b}$ events are from SHERPA 2.2.8 $Z \rightarrow b\bar{b}$ +jets sample. The $H \rightarrow b\bar{b}$ events are from ggF sample described in Section 5.3. The bottom panel shows the ratio of tagging efficiency for $Z \rightarrow b\bar{b}$ and $H \rightarrow b\bar{b}$	72

4.5	Distribution of the p_T asymmetry, $(p_{T,1} - p_{T,2})/(p_{T,1} + p_{T,2})$ (left) and the pseudorapidity difference between the first two p_T leading large- R jet, $ \Delta y_{1,2} $ (right) for $Z \rightarrow b\bar{b}$ (blue), $Z \rightarrow q\bar{q}$ ($q = c, l$) (yellow), $W \rightarrow qq$ (purple), dijet (red) and $t\bar{t}$ (green) process and collision data (black). The leading large- R jets are required to have p_T larger than 450 GeV and mass higher than 50 GeV. The subleading large- R jets are required to have p_T larger than 200 GeV.	76
4.6	Signal efficiency of the $X \rightarrow b\bar{b}$ tagger as a function of p_T at working point 60% with different definitions of $Z \rightarrow b\bar{b}$ using SHERPA 2.2.8 $Z \rightarrow b\bar{b}$ +jets sample. The Z boson candidate jets are required to have a mass between 50 and 150 GeV. <i>Matched Z: Contained Z + Other from V</i> . The label definition can be found in Section 3.3.5.	77
4.7	Data and MC comparison for Z + jets for differential p_T before the $X \rightarrow b\bar{b}$ tagger applied. Figure 4.7a shows the Z boson candidate jet p_T distribution. Figure 4.7b, 4.7c and 4.7d show the Z boson candidate jet mass distribution for $450 < p_T < 500$ GeV, $500 < p_T < 600$ GeV and $600 < p_T < 1000$ GeV. In each plot, the stacked distribution is from MC simulations and the black dots are from data. The red shadow band shows the MC statistical uncertainty. The bottom panel shows the ratio of event yields between the data and the MC predictions.	78
4.8	Data and MC comparison for Z + jets for differential p_T after $X \rightarrow b\bar{b}$ the tagger 60% WP. Figure 4.8a shows the Z boson candidate jet p_T distribution. Figure 4.8b, 4.8c and 5.7c show the Z boson candidate jet mass distribution for $450 < p_T < 500$ GeV, $500 < p_T < 600$ GeV and $600 < p_T < 1000$ GeV. In each plot, the stacked distribution is from MC simulations and the black dots are from data. The red shadow band shows the MC statistical uncertainty. The bottom panel shows the ratio of event yields between the data and the MC predictions.	79
4.9	Mass distributions of the large- R jets for $Z \rightarrow b\bar{b}$ (blue), $Z \rightarrow q\bar{q}$ ($q = c, l$) (yellow), $W \rightarrow qq$ (purple) and fully contained top (<i>contained $tq\bar{q}b$</i> , red) and other remnants from top (cyan) after $X \rightarrow b\bar{b}$ tagger 60% WP.	80
4.10	The Z boson candidate jet mass distribution after 60% WP of the $X \rightarrow b\bar{b}$ tagger for events with the large- R jet in the $450 < p_T < 500$ GeV (a); $500 < p_T < 600$ GeV (b) and $600 < p_T < 1000$ GeV (c) range. The dashed yellow line represents the Z + jets signal process. The dashed blue line represents the sum of all background processes. The solid red line is the sum of all signal and background processes.	84
4.11	Pull and ranking plots of the nuisance parameters for systematic uncertainties.	85
4.12	The p_T distribution of $Z \rightarrow \ell\ell$ (red) and $Z \rightarrow b\bar{b}$ (green) after the selections. The bottom panel shows the ratio of the event yields of $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$	87
4.13	The $Z \rightarrow \ell\ell$ p_T distribution (a) and mass distribution in three Z -boson candidate p_T bins: $450 < p_T < 500$ GeV (b), $500 < p_T < 600$ GeV (c) and $600 < p_T < 1000$ GeV (d). All distributions are shown after applying the event selection for the measurement of $\mu^{pre-tag}$	88
4.14	Signal efficiency scale factors for $X \rightarrow b\bar{b}$ tagger at 60% efficiency working point.	90

4.15	Truth label of the leading large- R jets (left) and the sub-leading large- R jets (right) in $Z \rightarrow b\bar{b}$, $Z \rightarrow q\bar{q}$ ($q = c, l$), $W \rightarrow qq$, dijet and $t\bar{t}$ samples. The truth label definition is detailed in Table 3.1.	91
5.1	The measured $V(\rightarrow leptons)H(\rightarrow b\bar{b})$ cross section with the resolved (left) and boosted (right) topologies in differential p_T bins.	94
5.2	The signal strengths measured using the differential signal regions in the inclusive Higgs production with $H \rightarrow b\bar{b}$ in ATLAS [153].	95
5.3	Chart flow for the region definition used in the fully hadronic $V(\rightarrow qq')H(\rightarrow b\bar{b})$ analysis.	101
5.4	The muon corrected large- R jet mass distribution of the Higgs candidate jet from Data and MC simulation for differential p_T in SR (left) and $\overline{\text{SR}}$ (right). The stacked plots are from the MC simulation. The distributions in black are from collision data. The event yields and its statistical uncertainties for each MC process and data are shown in the legend. The light grey box shows the MC statistical uncertainties. The bottom panels are the ratio of the event yields in the data and the MC prediction.	102
5.5	The large- R jet η distribution of the Higgs candidate jet from Data and MC simulation for differential p_T in SR (left) and $\overline{\text{SR}}$ (right). The stacked plots are from the MC simulation. The distributions in black are from collision data. The event yields and its statistical uncertainties for each MC process and data are shown in the legend. The light grey box shows the MC statistical uncertainties. The bottom panels are the ratio of the event yields in the data and the MC prediction.	103
5.6	The black curve is the leading large- R jet mass distribution after the selections for the measurement of $\mu^{post-tag}$ in Chapter 4. The red curve is the the leading large- R jet mass distribution after the selections for the measurement of $\mu^{post-tag}$ in the calibration of $X \rightarrow b\bar{b}$ tagger for the $V(\rightarrow qq')H(\rightarrow b\bar{b})$ analysis. Compared to the black curve, an additional selection is applied to get the red curve: i.e. <i>the events with the subleading jet mass in [60, 105] GeV are vetoed</i> . The p_T of the leading large- R jet is between 450 and 1000 GeV. The WP of the $X \rightarrow b\bar{b}$ tagger is 60%.	105
5.7	The leading large- R jet mass distribution after the selections for the measurement of $\mu^{post-tag}$ in the calibration of $X \rightarrow b\bar{b}$ tagger for the $V(\rightarrow qq')H(\rightarrow b\bar{b})$ analysis for $450 < p_T < 500$ GeV (a), $500 < p_T < 600$ GeV (b), $600 < p_T < 1000$ GeV (c). The 60% WP of the $X \rightarrow b\bar{b}$ tagger is used. Green: $Z \rightarrow q\bar{q}$; Pink: $W \rightarrow qq$; Purple: $t\bar{t}$; Blue: dijets. Black: data. Red: the statistical uncertainty band. The bottom panels are the ratio of the event yields from data and MC prediction.	105
5.8	Post-fit leading large- R jet mass distribution after 60% WP of the $X \rightarrow b\bar{b}$ tagger by vetoing the events with the subleading jet mass in [60, 105] GeV. Black dots: data. Red line: total post-fit distribution. Dashed blue line: post-fit dijet distribution. Dashed yellow line: post-fit signal distribution. Dashed green line: post-fit other backgrounds distribution.	106

5.9	Data and MC comparison after $Z \rightarrow \ell\ell$ selection for differential p_T with additionally vetoing the events with the leading jet mass in $[60, 105]$ GeV. Black dots: data. Stacked plots: MC simulation. Dark grey band: statistical uncertainty. Light grey band: statistical and systematic uncertainty.	106
5.10	Strategy of the BDT reweighting method. The BDT reweighting algorithm is trained with data passing the CR requirements and learns to reweight CR_failXbb large- R jet mass distribution to CR_passXbb. The training result is then applied to VR1 and VR2 for validations. Finally, a SR template is obtained by applying the training result to $\overline{\text{SR}}$	108
5.11	The shape of muon corrected Higgs candidate jet mass in CR_failXbb before (Red) and after (Blue) BDT reweighting in CR_passXbb (Black) for differential p_T bins: $250 < p_T^H < 450$ GeV (a), $450 < p_T^H < 650$ GeV (b) and $p_T^H > 650$ GeV (c). Black: target data distribution.	110
5.12	The shape of muon corrected Higgs candidate jet mass in $\overline{\text{SR}}$ before (Red) and after (Blue) BDT reweighting in SR (Black) for differential p_T bins: $250 < p_T^H < 450$ GeV (a), $450 < p_T^H < 650$ GeV (b) and $p_T^H > 650$ GeV (c). The black line is the target data distribution. Only the side-band data is shown in the plots.	110
5.13	ROC of VR1 for $250 < p_T^H < 450$ GeV. The X-axis is the True Positive Rate (TPR). The Y-axis is False Positive Rate (FPR). The red curve is before reweighting. The blue curve is after reweighting. The gray dashed curve is the random guess curve if there is no difference between the compared dataset and the target dataset.	111
5.14	ROC of VR1 for $450 < p_T^H < 650$ GeV. The X-axis is the True Positive Rate (TPR). The Y-axis is False Positive Rate (FPR). The red curve is before reweighting. The blue curve is after reweighting. The grey dashed curve is the random guess curve if there is no difference between the compared dataset and the target dataset.	111
5.15	ROC of VR1 for $p_T^H > 650$ GeV. The X-axis is the True Positive Rate (TPR). The Y-axis is False Positive Rate (FPR). The red curve is before reweighting. The blue curve is after reweighting. The grey dashed curve is the random guess curve if there is no difference between the compared dataset and the target dataset.	112
5.16	The ratio among each of N bootstraps SR templates and the nominal one for differential p_T . The blue line is the median obtained from the bootstrap distributions. The dashed red lines are the nominal $\pm$ half of the interquartile (IQR) range, i.e. the up and down variations of the statistical uncertainty estimated by the bootstrap method.	112
5.17	Non-closure systematic uncertainties of BDT templates for differential p_T bins. The dashed red lines are the up and down variations of the non-closure systematic uncertainties.	113

5.18	The muon corrected large- R jet mass distribution in SR for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. In each plot, the distribution of data is shown in black dots. Only side-band data is shown in the plots. The total statistical and systematic uncertainty is shown with the dashed band. Figure (a)-(c) are pre-fit muon corrected large- R jet mass distribution. The predicted dijet distribution using the BDT reweighting method is shown in blue. Other processes are predicted by MC simulations. Figure (d)-(f) are post-fit muon corrected large- R jet mass distribution from the inclusive fit setup and Figure (g)-(i) are from the differential fit setup. The post-fit dijet distribution using the BDT reweighting method is shown in blue.	114
5.19	Nuisance parameters of side-band fits. pT1stat, pT2stat and pT3stat are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; pT1syst, pT2syst and pT3syst are the non-closure systematic uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1, mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.	115
5.20	Correlation map for side-band fits. pT1stat, pT2stat and pT3stat are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; pT1syst, pT2syst and pT3syst are the non-closure systematic uncertainty for for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1, mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.	116
5.21	Comparisons of the fitted Higgs candidate muon corrected jet mass distributions with the BDT reweighting method (black) and the TF method (red) from the inclusive side-band fit. The error bars show the total fitted uncertainties in each method. The bottom panels show the ratio of the event yields fitted from the TF method and the BDT reweighting method.	117
5.22	The muon corrected large- R jet mass distribution in SR for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. In each plot, the distribution of Asimov data is shown in black dots. The total statistical and systematic uncertainty is shown with the dashed band. Figure (a)-(c) are pre-fit muon corrected large- R jet mass distribution. The predicted dijet distribution using the BDT reweighting method is shown in blue. Other processes are predicted by MC simulations. Figure (d)-(f) are post-fit muon corrected large- R jet mass distribution from the inclusive Asimov fit setup and Figure (g)-(i) are from the differential Asimov fit setup. The post-fit dijet distribution using the BDT reweighting method is shown in blue.	119

5.23	Nuisance parameters of Asimov fits. pT1stat, pT2stat and pT3stat are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; pT1syst, pT2syst and pT3syst are the non-closure systematic uncertainty for for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1, mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.	120
5.24	Correlation map for Asimov fits. pT1stat, pT2stat and pT3stat are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; pT1syst, pT2syst and pT3syst are the non-closure systematic uncertainty for for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1, mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.	121
6.1	Energy band structure of several materials [178].	126
6.2	Illustration of the Fermi function for a semiconductor [179].	126
6.3	Doping silicon with the donor atoms. (Left) Atom bonds with the donor dopant; (right) Energy bands diagram after the donor doping [180].	128
6.4	Doping silicon with the acceptor atoms. (Left) Atom bonds with the acceptor dopant; (right) Energy bands diagram after the acceptor doping [180].	128
6.5	The pn-junction formation. (Left) Two oppositely doped semiconductors are compared. Doping silicon with the donor atoms. (Right) The pn-junction is formed [180].	129
6.6	Sketch of a hybrid pixel [183].	132
6.7	Cross-sectional layout of segmented planar n-in-p pixel sensor with p-spray for inter-pixel insulation [184].	132
6.8	Scheme of the edge of a sensor with traditional cut technology (left) and with active edge technology (right).	133
6.9	Cross-sectional layout of 3D pixel sensor [184].	134
6.10	Types of point defects in a simple lattice [187].	135
6.11	The normalised rate of increase of leakage current α as function of the cumulated annealing time [188].	136
6.12	Type inversion of a n-type bulk into a p-type bulk [183].	136
7.1	Schematic layout of the ITk for the HL-LHC phase of ATLAS. Here only one quadrant and only active detector elements are shown. The horizontal axis is the axis along the beam line with zero being the interaction point. The vertical axis is the radius measured from the interaction point [184].	140
7.2	Design for $50 \times 50 \mu m^2$ (top) and $25 \times 100 \mu m^2$ (bottom) pixel cells, compatible with the RD53A chip. The structures colored in yellow represent the n^+ implants, in green the Aluminum layer, while the brown circles show the position of the UBM pads. [194]	142

7.3	Cross sectional schematic view of a module consisting of a single sensor and a readout chip bump bonded together, glued and wire bonded into a PCB. [184]	144
7.4	A RD53A single chip card.	144
7.5	The I-V curve of module W398-1 before and after irradiation. Pixel geometry: $50 \times 50 \mu m^2$, $130 \mu m$ thick.	145
7.6	Different layouts of pixel sensors for thin $50 \mu m$ thick sensors.	146
7.7	The IV-curves for $50 \mu m$ thick sensors before irradiation and after irradiation. Pixel geometry: $50 \times 50 \mu m^2$, $50 \mu m$ thick. Irradiation fluence: $\Phi = 5.3 \times 10^{15} n_{eq}/cm^2$	146
7.8	Illustration of the setup for module calibration at LPNHE.	147
7.9	Picture of test-beam setup with EUDET telescope at DESY.	148
7.10	Workflow of Eutelescope. [201]	149
7.11	Residuals for the EUDET telescope planes measured at DESY with a particle beam energy of 5 GeV. Residuals on the x and y-axis are superimposed for the same telescope plane.	152
7.12	An example of the residuals of measured DUT modules at DESY for different pitch size: $50 \times 50 \mu m^2$ (left) and $25 \times 100 \mu m^2$ (right).	153
7.13	Schematic overview of the working principle of TBmon2 [206].	153
7.14	Hit efficiency as a function of the bias voltage for the sensors of $50 \mu m$ [184].	155
7.15	Hit efficiency as a function of the bias voltage applied to the sensor after irradiation at the fluence of $2 \times 10^{15} n_{eq}/cm^2$ (left) and $5 \times 10^{15} n_{eq}/cm^2$ (right).	156
7.16	Schematic presentation of AIDA-2020 $150 \mu m$ thick wafer, AIDA-2020-G1 and AIDA-2020-G2. The orientation is illustrated.	158
7.17	Global efficiency of AIDA-2020-G1 and AIDA-2020-G2 as function of different bias voltage.	159
7.18	In-pixel efficiency map for AIDA-2020-G1.	159
7.19	In-pixel efficiency map for AIDA-2020-G2.	159
7.20	Edge efficiency for AIDA-2020-G1.	160
7.21	Edge efficiency for AIDA-2020-G2.	161
B.1	The comparison of $V + jets$ events generated from SHERPA 2.2.8 and SHERPA 2.2.11.	169
C.1	The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in CR_passXbb and CR_failXbb.	172
C.2	The Higgs jet η distribution from Data and MC simulations for differential p_T in CR_passXbb and CR_failXbb.	173
C.3	The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in VR1_failD2_passXbb and VR1_failD2_failXbb.	174
C.4	The Higgs jet η distribution from Data and MC simulations for differential p_T in VR1_failD2_passXbb and VR1_failD2_failXbb.	175
C.5	The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in VR1_failNtrk_passXbb and VR1_failNtrk_failXbb.	176
C.6	The Higgs jet η distribution from Data and MC simulations for differential p_T in VR1_failNtrk_passXbb and VR1_failNtrk_failXbb.	177

C.7	The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in VR1_failMass_passXbb and VR1_failMass_failXbb.	178
C.8	The Higgs jet η distribution from Data and MC simulations for differential p_T in VR1_failMass_passXbb and VR1_failMass_failXbb.	179

LIST OF TABLES

1.1	Production cross sections of the $m_H = 125$ GeV Higgs boson at the LHC [38].	25
1.2	Branching ratios and the relative uncertainty calculated for a SM Higgs boson with $m_H = 125$ GeV [38].	26
3.1	Truth label of the large- R jets.	59
4.1	The $X \rightarrow b\bar{b}$ tagger threshold values for all $X \rightarrow b\bar{b}$ WPs with $f_{top} = 0.25$.	71
4.2	Summary of the triggers used in the analysis. They are applied as an OR and all are required to be active. The offline threshold is the offline jet cut above which the triggers are 99% efficient.	75
4.3	The fraction of the $Z \rightarrow b\bar{b}$ events in the inclusive $Z \rightarrow q\bar{q}$ template after the $X \rightarrow b\bar{b}$ tagger at 60% WP.	80
4.4	The optimal function for differential p_T for 60% WP and its spurious signal uncertainty.	81
4.5	$\mu^{pre-tag}$ for differential p_T using inclusive $Z \rightarrow \ell\ell$ channel.	89
4.6	Pre-tag ($\mu^{pre-tag}$) and post-tag ($\mu^{post-tag}$) signal strength and the resulting signal efficiency scale factors (SF) for the $X \rightarrow b\bar{b}$ tagger at 60% efficiency working point measured using the $Z \rightarrow b\bar{b}$ calibration methods. Systematic uncertainties on the scale factor measurement are listed as well.	89
4.7	Scale factors derived for the signal efficiency of the $X \rightarrow b\bar{b}$ tagger at 60% WP using the binned template fit framework.	90
5.1	Cross section predictions used to normalise the simulated SM events for the ggF [163, 164], VBF [165], VH [161, 162], $t\bar{t}H$ [166], tH [167] and the SM fraction of each production mode.	98
5.2	Accuracy of assignment with different VH assignment schemes.	99
5.3	Signal significance in SR in the fully hadronic VH analysis and the inclusive $H \rightarrow b\bar{b}$ analysis.	101
5.4	Scale factors derived for the $X \rightarrow b\bar{b}$ tagger in fully hadronic VH analysis using $Z + \text{jets}$ events.	106
5.5	The optimised training variables for BDT reweighting method.	109
5.6	Post-fitted μ_{VH} in signal+background fit to the Asimov data with $\mu_{inj} = 1$ with the BDT reweighting method and the TF method.	118
7.1	Binary resolution for Mimosa planes and DUTs with each pixel geometry.	150
7.2	Summary of hit efficiency results for LPNHE R&D modules	154
7.3	ITk pixel market survey requirements. V_{bias} is the typical operational voltage before and after irradiation with different fluences.	156
7.4	Summary of hit efficiency results from market survey modules. Modules were tuned to mean threshold of 1500e.	157
A.1	Results of $\mu^{post-tag}$ and its uncertainties.	167
C.1	Scale factors derived for the $X \rightarrow b\bar{b}$ tagger in fully hadronic VH analysis using $Z + \text{jets}$ events.	180

ABSTRACT

Title: Identification of boosted Higgs bosons for new physics searches and improvement of the ATLAS tracker for the high luminosity phase of LHC

This thesis covers the main contributions of the author during the PhD program: the in-situ calibration of the $X \rightarrow b\bar{b}$ tagger using $Z + \text{jets}$ events, the measurement of the $V(\rightarrow qq)H(\rightarrow b\bar{b})$ production at high transverse momentum and the characterisation of the planar pixel modules for the ATLAS Inner Tracker (ITk) upgrade of the inner detector and for future colliders.

The in-situ calibration of the $X \rightarrow b\bar{b}$ tagger uses $Z + \text{jets}$ events to derive the signal scale factors for $p_T > 450$ GeV with data collected by the ATLAS experiment from 2015 to 2018. The dominant background, dijet, is modelled by fitting directly the data using an exponentiated polynomial or polynomial function. The scale factors are measured for the first time. The measured scale factors at 60%WP are $1.30^{+0.55}_{-0.56}$ for $450 < p_T < 500$ GeV, $1.17^{+0.35}_{-0.34}$ for $500 < p_T < 600$ GeV and $0.55^{+0.29}_{-0.28}$ for $600 < p_T < 1000$ GeV. Several improvements and applications are also discussed.

The measurement of VH production at high transverse momentum uses the final state of $V \rightarrow qq$ and $H \rightarrow b\bar{b}$ with data collected by the ATLAS experiment from 2015 to 2018. It is the first time this channel is explored in the ATLAS collaboration. The $X \rightarrow b\bar{b}$ tagger is used to tag the Higgs boson. The expected signal significance estimated by MC is around 1.28. The boosted decision tree reweighting method is one of the methods used for dijet modelling. The signal strength will be measured first in inclusive and differential p_T^H bins: $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. Then, the cross-section will be measured within the simplified template cross section (STXS) framework and will be used as well to constrain anomalous couplings in the Standard Model Effective Field Theory (SMEFT).

The LHC will enter the high-luminosity phase in 2026. The characterisation of the planar pixel modules is presented in the context of the ATLAS Inner Tracker (ITk) for the high luminosity LHC. The modules presented in this thesis include LPNHE R&D modules for ITk and future colliders, ITk market survey modules and the active edge modules. Their electrical properties and detection efficiencies have been measured.

Keywords: ATLAS, boosted Higgs boson, in-situ calibration, $X \rightarrow b\bar{b}$ tagger, $H \rightarrow b\bar{b}$, VH associated production, ITk, planar pixel module

RÉSUMÉ

Titre: Identification de bosons de Higgs boostés pour des recherches des nouvelles physiques et amélioration du tracker ATLAS pour la phase de haute luminosité du LHC

Cette thèse couvre trois sujets principaux : la calibration du marqueur $X \rightarrow b\bar{b}$ en utilisant des événements Z+jets, la mesure de la production $V(\rightarrow qq)H(\rightarrow b\bar{b})$ à haut moment transversal et la caractérisation des modules de capteurs planaires pour la mise à niveau du détecteur ATLAS Inner Tracker (ITk) et pour les futurs collisionneurs.

La calibration du marqueur $X \rightarrow b\bar{b}$ utilise les événements Z+jets pour dériver les facteurs d'échelle du signal pour $p_T > 450$ GeV avec les données collectées par l'expérience ATLAS de 2015 à 2018. Le fond dominant, dijet, est modélisé en ajustant directement les données à l'aide d'une fonction polynomiale ou polynomiale exponentielle. Les facteurs d'échelle sont mesurés pour la première fois. Les facteurs d'échelle mesurés à 60%WP sont de $1.30^{+0.55}_{-0.56}$ pour $450 < p_T < 500$ GeV, $1.17^{+0.35}_{-0.34}$ pour $500 < p_T < 600$ GeV et $0.55^{+0.29}_{-0.28}$ pour $600 < p_T < 1000$ GeV. Plusieurs améliorations et applications sont également discutées.

La mesure de la production de VH à haut moment transversal utilise l'état final de $V \rightarrow qq$ et $H \rightarrow b\bar{b}$ avec les données collectées par l'expérience ATLAS de 2015 à 2018. C'est la première fois que ce canal est exploré dans le cadre de la collaboration ATLAS. Le tagger $X \rightarrow b\bar{b}$ est utilisé pour marquer le boson de Higgs. La signification attendue du signal estimée par MC est d'environ 1,28. La méthode de boosted decision tree reweighting est l'une des méthodes utilisées pour la modélisation des dijets. L'intensité du signal sera mesurée dans des intervalles inclusifs et différentiels de p_T^H : $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV et $p_T^H > 650$ GeV.

Le LHC entrera dans la phase de haute luminosité en 2026. La caractérisation des capteurs de pixels planaires est présentée dans le contexte de l'ATLAS Inner Tracker (ITk) pour le LHC à haute luminosité. Les modules présentés dans cette thèse comprennent les modules LPNHE R&D pour ITk et les futurs collisionneurs, les modules ITk pour l'étude de marché et les modules de bord actifs. Leurs propriétés électriques et leurs efficacités de détection ont été mesurées.

Mots-clés: ATLAS, boson de Higgs boosté, calibration, tagger $X \rightarrow b\bar{b}$, $H \rightarrow b\bar{b}$, production associée VH , ITk, capteur de pixels planaires

RÉSUMÉ

Titre: Identification de bosons de Higgs boostés pour des recherches des nouvelles physiques et amélioration du tracker ATLAS pour la phase de haute luminosité du LHC

Le modèle standard de la physique des particules est une théorie relativiste des champs quantiques qui tente de décrire les particules fondamentales connues et leurs interactions à travers les forces fortes, faibles et électromagnétiques. Le SM a été construit sur le concept de symétrie de jauge. Le mécanisme de Higgs joue un rôle essentiel dans le modèle standard car il explique la masse des particules élémentaires, notamment les masses des bosons W et Z qui résultent de la brisure spontanée de symétrie de la force électrofaible.

Le modèle standard a passé avec succès des décennies de tests expérimentaux. Le grand collisionneur de hadrons (LHC) [1] de l'organisation européenne pour la recherche nucléaire (CERN), sous la frontière franco-suisse près de Genève, est aujourd'hui le plus grand et le plus puissant accélérateur de particules pour sonder la physique des particules élémentaires. Dans l'anneau du LHC, il y a quatre détecteurs principaux pour enregistrer les données de collision à différentes objectives de physique: ALICE[2], ATLAS [3], CMS [4] et LHCb [5].

En 2012, les collaborations ATLAS et CMS [6, 7] ont annoncé la découverte du boson de Higgs. Et toutes les particules prédites par le modèle standard ont été détectées expérimentalement. Cependant, le modèle standard a plusieurs limites théoriques et expérimentales. Cela motive la communauté des physiciens à chercher la physique au-delà du modèle standard, également connue sous le nom de la nouvelle physique, en particulier à la frontière des hautes énergies où la sensibilité à la nouvelle physique est plus élevée [8, 9].

L'étude des processus du boson Higgs à haut moment transversal est très intéressante du point de vue théorique et expérimental. D'un point de vue théorique, la forme p_T^H prédite par le modèle standard est remarquablement stable par rapport aux ordres supérieurs et peut fournir un moyen d'identifier une nouvelle physique qui entre dans le régime p_T élevé. Du point de vue expérimental, de plus en plus de bosons de Higgs boostés sont produits à mesure que le LHC fonctionne à plus haute énergie. Pendant des années d'exploitation et de mises à niveau, le LHC a été exploité à partir de $\sqrt{s} = 7$ to 13 TeV et la production du boson de Higgs à $m_H = 125$ GeV a augmenté de 19,1 fb à 55,1 fb. L'énergie de collision pp du LHC est passée à $\sqrt{s} = 13,6$ TeV depuis l'été 2022. À l'avenir, le LHC fonctionnera à une énergie encore plus élevée. Cela nous permet d'augmenter la précision des mesures du boson de Higgs à haut moment transversal. D'un autre côté, l'étude du boson de Higgs boosté est un défi et nécessite des techniques de reconstruction particulières, comme nous le verrons dans cette thèse.

En conséquence, les expériences du LHC ont continuellement développé de nouvelles technologies de détection, des techniques d'analyse physique et des logiciels pour améliorer la précision des mesures du modèle standard et de recherches en nouvelle physique. Ma thèse est donc définie à couvrir à la fois l'analyse physique et le développement de détecteurs au sein de la collaboration ATLAS.

Le détecteur ATLAS a un vaste programme de physique comprenant plusieurs mesures

de précision au modèle standard et des recherches de nouvelle physique. Physiquement, le détecteur se compose de six sous-détecteurs différents enroulés concentriquement autour du point d'interaction. Lors de collisions, les particules sont produites et puis traversent le détecteur. Elles interagissent avec les sous-systèmes qui enregistrent leur trajectoire, leur moment et leur énergie. Les données utilisées dans cette thèse ont été enregistrées par le détecteur ATLAS de 2015 à 2018.

Le mode de désintégration dominant d'un boson de Higgs est $H \rightarrow b\bar{b}$, et est donc le canal utilisé dans cette thèse. Au cours de cette thèse, trois sujets principaux ont été étudiés au sein de la collaboration ATLAS: la calibration du marqueur $X \rightarrow b\bar{b}$ à l'aide d'événements Z + jets, la mesure de la production de $V(\rightarrow qq)H(\rightarrow b\bar{b})$ à haut moment transversal et la caractérisation des capteurs de pixels planaires pour la mise à niveau ATLAS Inner Tracker (ITk) pour la phase de haute luminosité du LHC.

La calibration du marqueur $X \rightarrow b\bar{b}$ à l'aide d'événements Z + jets

Le marquage fait référence à l'identification d'un processus d'intérêt. Il est essentiel pour les analyses de physique où des processus et des topologies spécifiques sont utilisés. Dans les analyses physiques actuelles, les événements $H \rightarrow b\bar{b}$ sont identifiés comme les large- R jets avec deux jets à rayon variable (VR) associés à étiquette b , ce que l'on appelle la méthode de double étiquetage b . Les performances, les calibrations et les applications ont été étudiés sur la base de jet b isolé. Beaucoup de résultats intéressants ont été produits. Cependant, à très haute énergie, les deux jets b du boson de Higgs sont fortement collimatés, de sorte que la séparation des deux jets b devient moins efficace. Par conséquent, nous sommes motivés pour développer de nouvelles techniques de marquage plus efficaces. Le marqueur $X \rightarrow b\bar{b}$, récemment développé dans ATLAS, se concentre sur le marquage d'un jet de grand rayon contenant deux hadrons b . Le marqueur $X \rightarrow b\bar{b}$ est un algorithme basé sur un réseau de neurones. Le marqueur $X \rightarrow b\bar{b}$ montre un rejet de fond plus puissant que la méthode de double étiquetage b à la même efficacité de signal.

Étant donné que le marqueur est formé à l'aide de la simulation Monte Carlo qui ne peut pas décrire complètement les données réelles, ses performances dans la simulation MC ne sont pas les mêmes que celles des données réelles. En conséquence, la calibration du marqueur $X \rightarrow b\bar{b}$ est une tâche importante pour permettre son utilisation par les analyses physiques intéressées par les topologies $b\bar{b}$ boostées. Les événements Z + jets sont utilisés pour dériver les facteurs d'échelle du signal pour $p_T > 450$ GeV en utilisant les données collectées par l'expérience ATLAS de 2015 à 2018. Les facteurs d'échelle du signal sont mesurés par $\mu^{\text{post-tag}} / \mu^{\text{pre-tag}}$, où $\mu^{\text{post-tag}}$ et $\mu^{\text{pre-tag}}$ sont la force du signal après et avant de passer le tagger $X \rightarrow b\bar{b}$.

Dans la mesure de $\mu^{\text{post-tag}}$, les principaux bruits de fond sont dijet, $t\bar{t}$ et W + jets. Le dijet est modélisé en ajustant directement les données à l'aide d'un polynôme exponentiel ou d'une fonction polynomiale, selon le p_T . Par contre, avant l'application du marqueur $X \rightarrow b\bar{b}$, le rapport signal sur bruit de fond est trop faible pour mesurer $\mu^{\text{pre-tag}}$. Par conséquent, $\mu^{\text{pre-tag}}$ est obtenu par la mesure auxiliaire à l'aide d'événements $Z \rightarrow \ell\ell$.

Les facteurs d'échelle mesurés à l'aide d'événements Z + jets pour le marqueur $X \rightarrow b\bar{b}$ à 60% WP sont de $1,30^{+0,55}_{-0,56}$ pour $450 < p_T < 500$ GeV, $1,17^{+0,35}_{-0,34}$ pour $500 < p_T < 600$ GeV et $0,55^{+0,29}_{-0,28}$ pour $600 < p_T < 1000$ GeV. C'est la première fois que le marqueur $X \rightarrow b\bar{b}$ est calibré. La méthodologie a été établie, ainsi que le cadre de calibrage. Je suis un contributeur

principal pour ce travail. La méthodologie et les résultats sont documentés dans la Réf. [10].

Le marqueur $X \rightarrow b\bar{b}$ et sa calibration sont les recommandations officielles pour les analyses au sein de la collaboration ATLAS et ont déjà été utilisés par les récentes recherches de résonance VH [11]. De plus, la méthodologie et les logiciels de la calibration sont réutilisés pour d'autres études, telles que la calibration du tagger boosté $W/Z(\rightarrow qq)$ à l'aide d'événements V + jets, l'effet de proximité de la méthode de double étiquetage b et la analyse entièrement hadronique VH présentée dans cette thèse.

La mesure de la production de $V(\rightarrow qq)H(\rightarrow b\bar{b})$ à haut moment transversal

Les propriétés du boson de Higgs sont mesurées par les collaborations ATLAS et CMS après sa découverte, telles que les sections efficaces de production, les taux de désintégration, le couplage aux bosons et aux fermions. Toutes les mesures étaient en accord avec les prédictions pour un boson de Higgs du modèle standard dans les incertitudes. En conséquence, les physiciens des particules ont recherché de nouvelles idées et techniques pour améliorer encore les connaissances dans le secteur du boson de Higgs.

La mesure de la production de VH à haut moment transversal utilise l'état final de $V \rightarrow qq$ et $H \rightarrow b\bar{b}$. La force du signal sera d'abord mesurée dans les tranches p_T^H inclusives et différentielles : $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV et $p_T^H > 650$ GeV. Ensuite, la section efficace sera mesurée dans le cadre de Simplified Template Cross Sections (STXS) et sera également utilisée pour contraindre les couplages anormaux dans une théorie efficace du modèle standard. C'est la première fois que ce canal est exploré dans la collaboration ATLAS. Les études de simulation montrent une bonne complémentarité avec les canaux VH semileptoniques mesurés précédemment.

Les données Run 2 complètes sont utilisées. Les deux large- R jets avec p_T les plus hauts sont désignés comme un candidat du boson de Higgs et un candidat du boson vecteur. Le marqueur W/Z est utilisé pour marquer le boson vecteur. Le marqueur $X \rightarrow b\bar{b}$ est utilisé pour marquer le boson de Higgs. Le candidat du boson de Higgs doit réussir le marqueur $X \rightarrow b\bar{b}$ à 60% WP et le candidat du boson vecteur doit réussir le marqueur W/Z à 50% WP.

Les facteurs d'échelle du marqueur $X \rightarrow b\bar{b}$ sont remesurés dans l'analyse VH pour $p_T > 450$ GeV en utilisant la méthode j'ai développée pour la calibration du marqueur $X \rightarrow b\bar{b}$ à l'aide d'événements Z + jets et en utilisant une sélection orthogonale pour obtenir une donnée indépendante. Les facteurs d'échelle mesurés sont $1, 44^{+0,50}_{-0,57}$ pour $450 < p_T < 500$ GeV, $1,30^{+0,39}_{-0,44}$ pour $500 < p_T < 600$ GeV et $0,86^{+0,49}_{-0,55}$ pour $600 < p_T < 1000$ GeV.

Deux méthodes basées sur les données ont été développées pour estimer le bruit de fond principal dans cette analyse, le bruit de fond dijet : la méthode de repondération avec Boosted Decision Trees (BDT) et la méthode Transfer Factor (TF). Des comparaisons des deux méthodes d'estimation de fond ont été faites. Les distributions des ajustements de bande latérale de la région du signal entre les deux méthodes sont cohérentes l'une avec l'autre, ce qui tient compte de leurs incertitudes systématiques. Pour les ajustements d'Asimov, les erreurs ajustées avec la méthode TF sont plus petites que celles de la méthode de repondération BDT. Les deux méthodes ont une flexibilité très différente dans les ajustements statistiques.

Cette analyse est encore en cours et, par conséquent, aucun résultat final n'est présenté dans ce document. La méthode de modélisation dijet sera choisie parmi les méthodes de

repondération BDT et TF. L'ensemble complet des incertitudes systématiques, y compris les incertitudes systématiques expérimentales et les incertitudes systématiques de modélisation, est préparé et sera ajouté à la mesure finale. D'autres résultats arriveront dans les deux mois à venir, qui seront présentés lors de la soutenance de thèse. Ce document sera mis à jour en conséquence.

À long terme, la combinaison de cette analyse avec les résultats obtenus à partir de la production de la $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ et de la fusion des gluons-gluons avec $H \rightarrow b\bar{b}$ permettra de gagner en précision et d'améliorer la compréhension de la $H \rightarrow b\bar{b}$.

La caractérisation des capteurs de pixels planaires pour la phase de haute luminosité du LHC

Cette thèse présente également mes travaux sur la caractérisation des capteurs de pixels planaires dans le contexte de l'ATLAS Inner Tracker (ITk) pour le LHC à haute luminosité. Les modules d'intérêt sont les modules R&D du LPNHE pour l'ITk ou les futurs collisionneurs, les modules pendant la campagne d'étude de marché de l'ITk, et les modules avec bord actif dans le cadre du programme AIDA. La caractérisation de ces modules a été réalisée en plusieurs étapes. Tout d'abord, la courbe I-V des modules a été mesurée au laboratoire local pour déterminer les limites opérationnelles des appareils testés. Après irradiation, le courant de fuite des capteurs augmente du fait des défauts créés dans la zone appauvrie. Ensuite, les modules ont été calibrés et l'efficacité des modules a été mesurée à l'aide d'installations des faisceaux d'essai. Les données obtenues à partir des faisceaux d'essai ont été reconstruites et analysées à l'aide d'EUTelescope et de TBmon2.

Les modules R&D du LPNHE pour ITk et les futurs collisionneurs ont été testés. L'expérience des tests de différents modules a été utilisée dans la conception du capteur ITk. Les capteurs avec épaisseur de $50\ \mu\text{m}$ sont un candidat prometteur pour les futurs collisionneurs. Deux modules avec épaisseur de $50\ \mu\text{m}$ ont été testés avant irradiation. Les modules ont été irradiés récemment et seront à nouveau testés sous des faisceaux d'essai. Au-delà de la R&D pour l'ITk, des modules plus fins sont assemblés au LPNHE pour mener une étude approfondie. Les modules pour l'étude de marché ITk ont également été testés pour s'assurer que les différentes sociétés produisant des modules répondaient aux exigences de qualité. La campagne d'étude de marché ITk est terminée et les modules officiels pour ITk vont être produits avec les fournisseurs validés. Les modules avec bord actif sont développés dans le cadre de la collaboration AIDA-2020. L'efficacité de bord des capteurs est étendue grâce à la conception d'anneaux de protection. Les capteurs ont maintenant été irradiés pour de futures études de performance.

En résumé, le travail de cette thèse comprend des études sur la calibration du marqueur $X \rightarrow b\bar{b}$, des analyses physiques avec le boson de Higgs et le développement de détecteurs. Les études de la calibration $H \rightarrow b\bar{b}$ fournissent des données essentielles pour les analyses physiques utilisant $H \rightarrow b\bar{b}$ boosté. L'analyse physique avec le boson de Higgs utilisant $V(\rightarrow qq)H(\rightarrow b\bar{b})$ fournit une nouvelle perspective et des informations supplémentaires à la mesure combinée de Higgs pour sonder le modèle standard et la nouvelle physique. Le travail sur le développement des détecteurs permet de mieux comprendre nos futurs détecteurs et, par conséquent, le potentiel physique du détecteur.

Mots-clés: ATLAS, boson de Higgs boosté, calibration, remarqueur $X \rightarrow b\bar{b}$, $H \rightarrow b\bar{b}$, production associée VH , ITk, capteur de pixels planaires

INTRODUCTION

The Standard Model (SM) of particle physics is a relativistic quantum field theory that describes the known fundamental particles and their interactions through the strong, weak and electromagnetic forces. It has successfully passed decades of experimental tests. In the SM, the inertial mass of all elementary particles is given by their interactions with the Higgs field. In 2012, the Higgs boson with mass of around 125 GeV was discovered by the ATLAS and CMS collaborations [6, 7] at the Large Hadron Collider (LHC), the world's largest and most powerful particle accelerator to probe elementary particle physics. And all particles predicted by the SM have been experimentally detected. However, the SM has several theoretical and experimental limitations. This motivates the physics community to search for beyond-SM physics, also known as new physics, especially at the high energy frontier where the sensitivity to new physics is higher [8, 9].

At the same time, the LHC has been extending its capability of probing new physics at high energies by increasing its collision energy and luminosity. Accordingly, the ATLAS experiment, one of the main experiments at the LHC, has continuously developed novel detector technologies, physics analysis technique, and software to improve the precision of SM measurements and new physics searches. My contributions to the ATLAS collaboration cover both physics analysis and detector development. The outline of this thesis and my contributions are presented as follow.

Chapter 1 introduces the SM, the theoretical framework of particle physics. The SM was built on the concept of gauge symmetry. The Higgs mechanism plays an essential part in the SM because it explains the mass of the elementary particles, especially the masses of W and Z boson which arise from the spontaneous symmetry breaking of the electroweak force. The main Higgs production mechanisms in proton-proton (pp) collisions are gluon fusion, vector boson fusion, associated production with a gauge boson, and associated production with top quarks. The dominant decay mode of a Higgs boson with mass of 125 GeV is $H \rightarrow b\bar{b}$, and is thus the channel used in this thesis. The main physics result in this thesis is the measurement of the associated production of a Higgs boson with a vector boson, both decaying hadronically, at high transverse momentum.

Chapter 2 describes the LHC and the ATLAS detector. The ATLAS detector, located with the LHC ring, has a vast physics program including precision SM measurements and searches for new physics. The detector consists of six different subdetectors wrapped concentrically around the interaction point. When the particles produced in the collisions pass through, they interact with the sub-systems that record their trajectory, momentum, and energy. The pp collision data used in this thesis was recorded by the ATLAS detector from 2015 to 2018.

Chapter 3 focuses on the reconstruction, identification and calibration of the physics objects starting from the ATLAS detector signals, with a focus on the objects used in this thesis.

Chapter 4 presents the identification of high transverse momentum Higgs bosons decaying to b -quarks and their corresponding calibration using Z + jets events. The identification tools are required to efficiently discriminate the $H \rightarrow b\bar{b}$ events from QCD and top backgrounds. Two methods are reviewed in this chapter: double b -tagging method and the

$X \rightarrow b\bar{b}$ tagger. This chapter focuses on the signal efficiency calibration of the $X \rightarrow b\bar{b}$ tagger with data using $Z + \text{jets}$ events. I am one of the main contributors for this calibration work. This work was documented in Ref. [10]. These results are being used in analyses using the $X \rightarrow b\bar{b}$ tagger and the full dataset collected between 2015 and 2018.

Chapter 5 presents the measurement of the Higgs boson production associated with a vector boson. The final state used for the Higgs boson is $H \rightarrow b\bar{b}$; whereas the vector boson is detected via $V \rightarrow qq$. The tool used to tag the $H \rightarrow b\bar{b}$ events is the $X \rightarrow b\bar{b}$ tagger presented in Chapter 4. The analysis is ongoing and the corresponding results will be updated by the time of the thesis defense. I am one of the main analysts and my contributions so far include: event selection optimisation, the development and validation of a background estimation and the calibration of the tagger used to identify high transverse momentum $H \rightarrow b\bar{b}$ objects.

Chapter 6 introduces the basics of silicon sensors. The silicon detectors play a crucial role in tracking and b -tagging. The building block of the silicon particle detectors is the pn-junction. Electron-hole pairs are generated and collected when a charged particle passes through the depleted region of the sensor. Pixel detectors are made up of the pn-junctions organized in a matrix of two-dimensional arrays to have a good spatial resolution. This chapter concludes with a brief review of radiation damage in silicon sensors.

Chapter 7 presents the characterisation of the planar pixel modules for ATLAS Inner Tracker (ITk). ITk is designed to cope with the increasingly high data readout requirement and the harsh radiation environment during the high luminosity phase of the LHC (HL-LHC). The ATLAS group at LPNHE is involved in the design, development and characterisation of pixel sensor for ITk, as well as in the module assembly process for the central part of the ITk detector. This chapter presents results of the planar pixel modules characterisation studies in which I had a primary role. All modules have been tested with particle beams, and the test-beam results are presented. I have participated in the test-beam activities, including setup building, module testing/tuning, and data-taking. I have also contributed to reconstructing and analysing test-beam data. The results can be found in this thesis and in Refs. [12, 13].

THEORETICAL FRAMEWORK

What are the fundamental blocks of matter? How do they interact? What are the fundamental rules of the Universe? These are puzzling questions that humans have been trying to answer throughout history. The Standard Model (SM) of particle physics is the best theoretical framework that we have today to answer some of these questions. However, it has limitations. For example, it can not explain dark matter and the gravitational interaction is not included. Therefore, we are motivated to search for new physics beyond SM. A great place to look for new physics is in the Higgs sector as it is tightly related to the electroweak symmetry breaking (EWSB). In this chapter, the SM is presented in Section 1.1. Then, Section 1.2 focuses on the Higgs sector.

1.1 The Standard Model (SM)

In the SM, the Universe is made by fermions of half-integer spin. The fermions interact with each other through interactions mediated by elementary bosons of integer spin. The illustration of the elementary particles in the SM is shown in Figure 1.1. There are three types of interactions in the SM: electromagnetic, weak, and strong forces. The gauge bosons that mediate these interactions are the photon (γ), the vector bosons ($W^\pm/Z$) and the gluons (g), respectively. An additional scalar particle, the Higgs boson, is the force carrier of the Higgs field; the Higgs field is responsible for generating the masses of the elementary particles and the differences between the electromagnetic and the weak forces. The fermions are divided into two categories depending on their interactions: leptons and quarks. Leptons do not carry a “colour” charge and thus are not affected by the strong force, while quarks are sensitive to all the fundamental interactions. For both leptons and quarks there are three generations, consisting of two particles: an up-type (charge $q = 2/3$) and a down-type ($q = -1/3$) quark, or a charged lepton ($q = -1$) and a neutrino ($q = 0$). The unit of charges is the absolute value of electron charge ($|e|$). The following sections show how the SM was built.

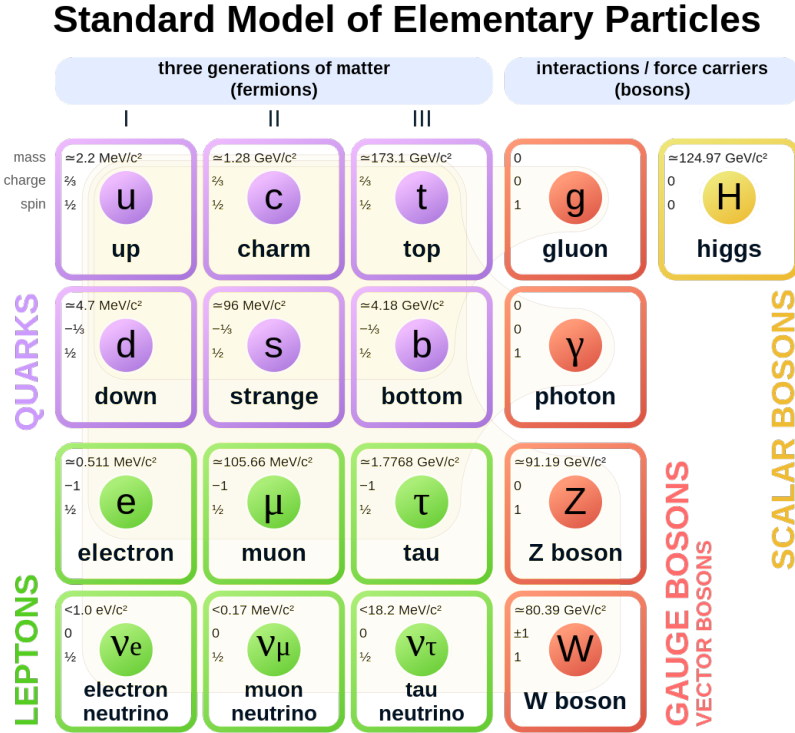


Figure 1.1: The Standard Model of elementary particles [14].

1.1.1 Quantum electrodynamics (QED)

The Standard Model is built on the idea that the interactions between elementary particles correspond to symmetries of the Lagrangian under a group of local transformations. To do so, we start from a fermion field ψ in the free theory; the Dirac Lagrangian can be written as:

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

where m is the mass of the fermion, γ^μ are the 4×4 Dirac matrices:

$$\gamma^0 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \gamma^i = \begin{pmatrix} 0 & \tau^i \\ \tau^i & 0 \end{pmatrix} \quad (1.2)$$

where τ^i are the Pauli matrices:

$$\tau^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tau^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \tau^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (1.3)$$

In Equation (1.1), the first part is denoted as the kinetic term, while the second as the mass term. The Lagrangian is invariant under the global U(1) phase transformation. Let us consider the local gauge U(1) phase transformation of the field (interpreted as a rotation),

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

where x is the point in the 4-dimensional phase space considered, q is the real parameter relating to the fermion and $\theta(x)$ the local phase transform.

Under this transformation, the Lagrangian, written as Equation (1.1), is not invariant anymore:

$$\begin{aligned}\mathcal{L}_{free} \rightarrow \mathcal{L}' &= (e^{-iq\theta(x)}\bar{\psi})(i\gamma^\mu\partial_\mu - m)e^{iq\theta(x)}\psi \\ &= \mathcal{L}_{free} - q\partial_\mu\theta(x)\bar{\psi}\gamma^\mu\psi\end{aligned}\quad (1.5)$$

In order to restore the invariance, a corrective term is introduced and one choice could be:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + i q A_\mu \quad (1.6)$$

where A_μ is an additional vector field (i.e. described as a boson) which is required to transform under the U(1) transformation as:

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu\theta(x) \quad (1.7)$$

The new Lagrangian is therefore written as:

$$\begin{aligned}\mathcal{L}_{U(1)} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \\ &= \mathcal{L}_{free} - qJ^\mu A_\mu\end{aligned}\quad (1.8)$$

where $J^\mu = \bar{\psi}\gamma^\mu\psi$ is a term associated to the electromagnetic current, and thus this term can be interpreted as an interaction term between the fermion and the field A_μ . Experimentally, the A_μ field is interpreted as the mediator of the interaction (i.e. the photon in this case), so-called the electromagnetic field. q is interpreted as the coupling constant of the fermion with the electromagnetic field (i.e. charge of the fermion). Therefore, imposing the invariance of the free fermion Lagrangian under the local gauge U(1) transformation leads to the introduction of a boson (γ) and its interaction with the fermion. Moreover, by introducing the electromagnetic field tensor $F_{\mu\nu}$ defined as:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (1.9)$$

which can be obtained by computing the transformation of the commutation operator $[D_\mu, D_\nu]$ with respect to the U(1) local gauge symmetry [15] and which is directly gauge invariant, the full QED Lagrangian can thus be rewritten as :

$$\begin{aligned}\mathcal{L}_{QED} &= \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ &= \mathcal{L}_{free} - \underbrace{qJ^\mu A_\mu}_{\text{interaction term}} - \underbrace{\frac{1}{4}F_{\mu\nu}F^{\mu\nu}}_{\text{mediator kinetic term}}\end{aligned}\quad (1.10)$$

The QED Lagrangian can be expressed in terms of the left-handed (ψ_L) and right-handed (ψ_R) components of the fermion field ψ :

$$\mathcal{L}_{QED} = \underbrace{\bar{\psi}_L i\gamma^\mu \partial_\mu \psi_L + \bar{\psi}_R i\gamma^\mu \partial_\mu \psi_R}_{\text{kinetic term}} - \underbrace{qA_\mu(\bar{\psi}_L \gamma^\mu \psi_L + \bar{\psi}_R \gamma^\mu \psi_R)}_{\text{interaction term}} - \underbrace{m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R)}_{\text{mass term}} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.11)$$

The mass term mixes the left-handed and right-handed fields, while the kinetic and the interaction terms do not. For the massless neutrinos, its Lagrangian only includes the kinetic terms of left-handed component, since $m_\nu = 0$, $q_\nu = 0$.

1.1.2 The electroweak unification

In 1957, the weak interaction was observed only to couple left-handed electrons and neutrinos together by Madame Wu's experiment [16]. This observation could not be interpreted using the U(1) transformation described in the previous section. To describe it, Bludman proposed in 1958 to use SU(2) as symmetry group [17]. The following representation is used:

$$L_L = \begin{pmatrix} \nu_L^e \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_L^\mu \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_L^\tau \\ \tau_L \end{pmatrix}, \quad L_R = e_R, \mu_R, \tau_R \quad (1.12)$$

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix}, \quad U_R = u_R, c_R, t_R \quad (1.13)$$

$$D_R = d_R, s_R, b_R \quad (1.14)$$

Neglecting the masses of the fermions, the resulting Lagrangian for the leptons l can be expressed only with the kinematic part:

$$\mathcal{L}_{\text{kin.}} = i\bar{L}_L \gamma^\mu \partial_\mu L_L + i\bar{L}_R \gamma^\mu \partial_\mu L_R \quad (1.15)$$

Under the SU(2) transformations, the fields transform as :

$$\begin{aligned} \psi_L &\rightarrow \psi'_L = e^{i\theta(\vec{x}) \cdot \vec{\tau}} \psi_L \\ \psi_R &\rightarrow \psi'_R = \psi_R \end{aligned} \quad (1.16)$$

where $\vec{\tau} = (\tau_1, \tau_2, \tau_3)$ are the Pauli matrices. Similarly to the U(1) case, the invariance of the Lagrangian in Equation (1.15) under the SU(2) local transformation is obtained by introducing:

$$\begin{aligned} \partial_\mu \psi_L &\rightarrow D_\mu \psi'_L = (\partial_\mu - i g \frac{\tau_i}{2} W_\mu^i) e^{i\theta(\vec{x}) \cdot \vec{\tau}} \psi_L \\ \partial_\mu \psi_R &\rightarrow \partial_\mu \psi'_R = \partial_\mu \psi_R \end{aligned} \quad (1.17)$$

where $\vec{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ are massless isovector triplet. The new Lagrangian becomes:

$$\mathcal{L}_{\text{kin.}} = \underbrace{\frac{g}{\sqrt{2}} (J^{+, \mu} W_\mu^+ + J^{-, \mu} W_\mu^-)}_{\text{charged current}} + \underbrace{g (\bar{\nu}_L \gamma^\mu \nu_L - \bar{l}_L \gamma^\mu l_L) W_\mu^3}_{\text{neutral current}} \quad (1.18)$$

where the charged currents $J^{\pm, \mu}$ are defined as:

$$J^{+, \mu} = (J^{+, \mu})^\dagger = \bar{\nu}_L \gamma^\mu l_L \quad (1.19)$$

and the positive and negative vector bosons:

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \pm W_\mu^2) \quad (1.20)$$

The requirement of the SU(2) gauge invariance incorporates 3 bosons: W^+ , W^- and W^3 . The $W^\pm$ mediate the charged weak interactions. W^3 is a neutral boson but cannot be identified with the electromagnetic one. The similarity in the U(1) and the SU(2) gauge

symmetry motivates a new theory to unify the electromagnetic and weak interaction. In 1961, Sheldon Glashow proposed to use the $SU(2) \times U(1)$ group [18].

In order to keep the invariance of the Lagrangian (1.15) under the local transformation of $SU(2) \times U(1)$, the following transformation laws are introduced :

$$\begin{aligned}\psi_L &\rightarrow \psi'_L = e^{i(\vec{\theta}(\vec{x}) \cdot \vec{\tau} + \alpha(x) \cdot Y)} \psi_L \\ \psi_R &\rightarrow \psi'_R = e^{i\alpha(x) \cdot Y} \psi_R \\ \partial_\mu &\rightarrow D_\mu = (\partial_\mu - i g \frac{\tau_i}{2} W_\mu^i - i g' \frac{Y}{2} B_\mu)\end{aligned}\tag{1.21}$$

where Y is the weak hypercharge, $Y = -1$ for the left-handed doublet and $Y = -2$ for the right-handed singlet. $\vec{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ and B_μ are fields introduced. The original kinetic term of the Lagrangian is thus written as:

$$\mathcal{L}_{\text{kin.}} = \frac{g}{\sqrt{2}}(J^{+, \mu} W_\mu^+ + J^{-, \mu} W_\mu^-) + g(\bar{\psi}_L \gamma^\mu \frac{\tau_3}{2} \psi_L) W_\mu^3 + g'(-\bar{\psi}_L \gamma^\mu \frac{1}{2} \psi_L - \bar{\psi}_R \gamma^\mu \psi_R) B_\mu\tag{1.22}$$

Defining the lepton weak isospin ($J^{3, \mu}$) and weak hypercharge ($J^{Y, \mu}$) currents as:

$$J^{3, \mu} = \bar{\psi}_L \gamma^\mu \frac{\tau_3}{2} \psi_L\tag{1.23}$$

and

$$J^{Y, \mu} = -\bar{\psi}_L \gamma^\mu \frac{1}{2} \psi_L - \bar{\psi}_R \gamma^\mu \psi_R\tag{1.24}$$

The neutral part of the Lagrangian in (1.22) can be written as:

$$\mathcal{L}_{\text{neutral}} = g J^{3, \mu} W_\mu^3 + g' J^{Y, \mu} B_\mu\tag{1.25}$$

In order to recover the electromagnetic interaction, the rotation of the neutral vector boson fields is introduced:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \cdot \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}\tag{1.26}$$

where θ_W is the mixing weak angle, so-called the Weinberg angle.

The neutral part of the Lagrangian becomes:

$$\mathcal{L}_{\text{neutral}} = g J^{3, \mu} (\cos \theta_W Z_\mu + \sin \theta_W A_\mu) + g' J^{Y, \mu} (-\sin \theta_W Z_\mu + \cos \theta_W A_\mu)\tag{1.27}$$

In particular, the neutral current interaction mediated by the field (A_μ) can be extracted as:

$$\mathcal{L}_{\text{em.}} = (g \sin \theta_W J^{3, \mu} + g' \cos \theta_W J^{Y, \mu}) A_\mu\tag{1.28}$$

Since :

$$J^{3, \mu} + J^{Y, \mu} = \bar{\psi}_R \gamma^\mu \psi_R + \bar{\psi}_L \gamma^\mu \psi_L = J^\mu\tag{1.29}$$

by fixing $g \sin \theta_W = g' \cos \theta_W$, the electromagnetic part of the interaction can be restored and the full Lagrangian in Equation (1.22) can be rewritten as:

$$\begin{aligned}\mathcal{L}_{\text{kin.}} &= \frac{g}{\sqrt{2}}(J^{+, \mu} W_\mu^+ + J^{-, \mu} W_\mu^-) + g J^{3, \mu} W_\mu^3 + g' J^{Y, \mu} B_\mu \\ &= \frac{g}{\sqrt{2}}(J^{+, \mu} W_\mu^+ + J^{-, \mu} W_\mu^-) + \frac{g}{\cos \theta_W} (J^{3, \mu} - \sin^2 \theta_W J^\mu) Z_\mu + g \cdot \sin \theta_W J^\mu A_\mu\end{aligned}\tag{1.30}$$

where the $W^\pm$ bosons are responsible for mediating the weak charge currents, Z boson for mediating a neutral current and the photon γ for the electromagnetic current. However, the mass term can not be included because it breaks the gauge symmetry. The theory predicting the massless particles presents this limitation until 1964 when the Brout-Englert-Higgs mechanism was proposed. The details will be shown in Section 1.1.4.

1.1.3 Quantum chromodynamics (QCD)

Experimentally, no isolated quark has ever been observed as they form either mesons or baryons. The model to describe the experimental observations is quantum chromodynamics (QCD), the theory of strong interaction of quarks. QCD is based on the SU(3) color symmetry and assumes each flavour of quarks comes in three different colors. The quark fields are represented using 3-dimensional vectors. The color states of quarks are SU(3) triplets:

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} \quad (1.31)$$

The local gauge transformation is defined as:

$$\psi \rightarrow \psi' = e^{\frac{i}{2}\theta^a(x).t^a} \psi \quad (1.32)$$

where a represents the Gell-Mann matrix indices (from 1 to 8). The Gell Mann matrices are defined as:

$$\begin{aligned} t^1 &= \frac{1}{2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & t^2 &= \frac{1}{2} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ t^3 &= \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} & t^4 &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ t^5 &= \frac{1}{2} \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} & t^6 &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ t^7 &= \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} & t^8 &= \frac{1}{2\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \end{aligned} \quad (1.33)$$

and couple different colours to each other by $[t^a, t^b] = if^{abc}t^c$. The quarks are therefore coloured objects. Similarly to the U(1) and the SU(2) case, the following transformation laws are introduced:

$$\begin{aligned} \partial_\mu \rightarrow (D_\mu) &= \partial_\mu - ig_s A_\mu^a t^a \\ A_\mu^a t^a &\rightarrow A_\mu^a t^a + \frac{1}{g_s} \partial_\mu \theta^a t^a + i[t^a, t^b] \theta^a A_\mu^b \end{aligned} \quad (1.34)$$

where g_s is the strong coupling constant and A_μ^a are the eight massless gluons created from the gauge symmetries. The QCD Lagrangian is thus derived including the kinetic term for

the gluons:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi} (i\gamma^\mu D_\mu - m) \psi - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (1.35)$$

where $G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c$. When expanding the Lagrangian of Equation (1.35), one can find the quark-gluon interaction term and the self-interactions of the gluon fields with a trilinear and quadrilinear coupling. The self-interaction term explains that neither quarks nor gluons are observed as free particles, which was not the case in QED. Furthermore, in order to properly quantize the QCD Lagrangian, one needs to add the so-called Gauge-fixing and Fadeev-Popov ghosts terms [19], also known as the ghost-gluon interaction. In the QCD Lagrangian, all interaction terms are given in terms of a single coupling g_s . Unfortunately, only using a single constant g_s does not help to understand the strong-interacting phenomena across different energy regimes because we have learnt from experiments that the strength of the strong force changes with the energy scale: quarks are confined at low energies but behave as free particles at high energies. Moreover, if we neglect the quark masses, there is no energy scale in the Lagrangian. Thus, QCD should provide some dynamical scale if it is the right theory.

Let's consider the scattering amplitude which is a function of the four momentum of the n final-states particles in an event. The leading order (LO) calculation of the scattering amplitude is carried out for $2 \rightarrow n$ processes at tree level. The next-to-leading order (NLO) calculation involves supplementing the LO result with the squared $2 \rightarrow n+1$ processes at tree level and the interference of the $2 \rightarrow n$ processes at tree level and 1-loop level. The next-to-next-to-leading order (NNLO) is considerably more complicated than the NLO as it consists of: the squared $(n+2)$ -parton tree-level amplitude, the interference of the $(n+1)$ -parton tree-level and 1-loop amplitudes, the interference of the n -parton tree-level and 2-loop amplitudes, and the squared n -parton 1-loop amplitude. Although it is straightforward to compute all kind of scattering amplitudes at LO, the computation of corrections at NLO/NNLO diverge, known as ultra-violet (UV) divergences. The UV divergences exist in both QCD and QED. In order to solve these issues, the dimensional regularisation and renormalisation of the perturbative QCD (pQCD) is discussed here.

The **dimensional regularization** is a method proposed by Hooft and Veltman [20]. Let's consider the self-energy gluon loop. The divergent term comes from the momentum integration. We could introduce a cut-off λ in the integrals such that only momentum scales smaller than λ are considered in the integration. The integral is thus well-defined via dimensional regularization and λ becomes a regularizing parameter. Nevertheless, λ could be related to some unknown additional physics at very short distances. To avoid this, we introduce an arbitrary energy scale μ to form a dimensionless quantity q^2/μ^2 where q^2 is the momentum transfer in the loop. In this way we can determine the effect of the energy scale which was not shown in the QCD Lagrangian.

The **renormalisation** is a collection of techniques to reabsorb all UV divergences by redefining the original fields and couplings. After redefinition, the scattering amplitudes are usually finite and give rise to definite predictions for the cross section, which can be compared with the experiment. The renormalised coupling is thus measurable. The run-

ning couplings depend on the energy scale, μ , and its reference, μ_R :

$$\begin{aligned}\alpha_{QED}(\mu) &= \frac{\alpha_{QED}(\mu_R)}{1 - \frac{\alpha_{QED}(\mu_R)}{6\pi} \ln\left(\frac{\mu}{\mu_R}\right)} \\ \alpha_s(\mu) &= \frac{\alpha_s(\mu_R)}{1 + \frac{\alpha_s(\mu_R)}{6\pi} (11N_C - 2N_f) \ln\left(\frac{\mu}{\mu_R}\right)}\end{aligned}\tag{1.36}$$

where $g_{QED/s} = \sqrt{4\pi\alpha_{QED/s}}$, N_C is the number of colours of quarks and N_f is the number of generation of fermions. One important difference is that the correction in the denominator is negative for the QED structure while for QCD is positive. Therefore, for QCD, the lowest the energy is, the strongest the coupling is, as shown in Figure 1.2.

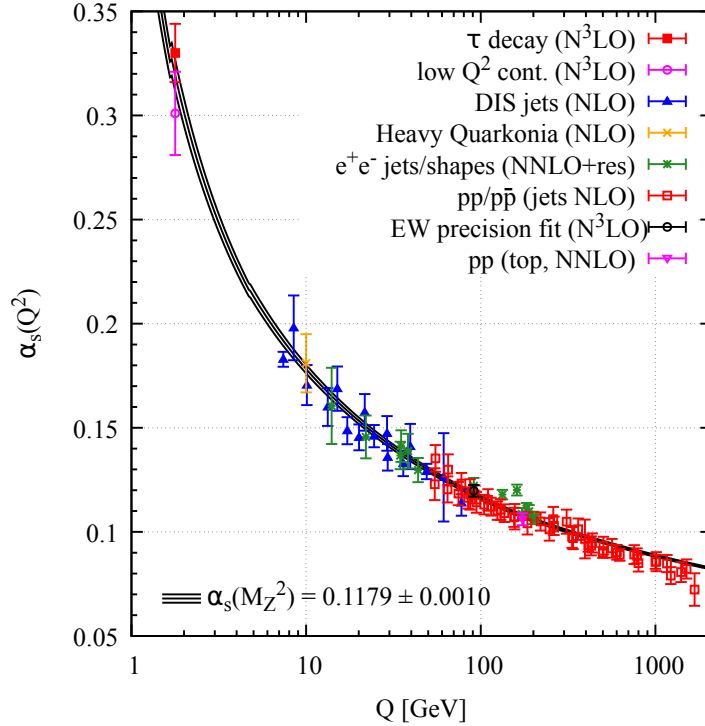


Figure 1.2: Measurements of α_s as a function of the energy scale Q [21].

In the experimental studies of QCD, since we are not able to directly measure partons (quarks and gluons), but only hadrons and their decay products, we have to establish a correspondence between observables obtained at the parton and the hadron level. The correspondence is achieved by means of quantities which are infrared and collinear safe. In other words, the quantities should be finite at any order of QCD. Besides, a correspondence between the hadron level and the actual measured quantities in the detector has to be established. Jets are an important objects in the experimental studies of QCD. The performance and reconstruction of jets in the ATLAS experiment is presented in Section 3.3.

1.1.4 Spontaneous symmetry breaking (SSB) and the Higgs mechanism

The theoretical structure presented in Section 1.1.1, 1.1.2 and 1.1.3 only predicts massless particles which contradicts the experimental measurements of non-zero particle mass. For example, the electrons was observed having mass by Arthur Schuster in 1890, eighty years before the theoretical structure proposed. The Brout-Englert-Higgs mechanism (the Higgs mechanism in short) resolved this tension through the application of spontaneous symmetry breaking (SSB). The concept of SSB was first introduced by Peter Higgs, Robert Brout and Francois Englert, and Gerald Guralnik, C. R. Hagen, and Tom Kibble [22–24]. Then, it was proposed as an application to the electroweak unification for the generation of masses for gauge bosons by Salam and Weinberg [25, 26]. Let's focus directly on the electroweak symmetry breaking (EWSB) through which the Higgs field breaks the $SU(2) \times U(1)$ symmetry in the SM. A complex $SU(2)$ doublet field is introduced:

$$\phi = \begin{pmatrix} \phi^\dagger \\ \phi_0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.37)$$

The kinetic term and potential has the following forms:

$$\begin{aligned} \mathcal{L}_\phi &= (D^\mu \phi)^\dagger D_\mu \phi - V(\phi) \\ &= (D^\mu \phi)^\dagger D_\mu \phi - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 \end{aligned} \quad (1.38)$$

When $\mu^2 < 0$, ϕ obtains as non-zero vacuum expectation value (vev) of $\nu^2 = \frac{-\mu^2}{\lambda}$, around which one may expand:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.39)$$

after applying a gauge transformation to the unitary gauge. $H(x)$ is the field which describes variations from the vacuum, i.e. the field associated with the SM Higgs boson. The kinetic term of ϕ provides masses to the weak gauge bosons:

$$\begin{aligned} (D^\mu \phi)^\dagger D_\mu \phi &= |[\partial_\mu + i\frac{g}{2}\tau^i W_\mu^i + i\frac{g'}{2}B_\mu]\phi|^2 \\ &= \frac{1}{2}\partial^\mu H \partial_\mu H + \frac{1}{8}(g^2(v+H)^2|W_\mu^1 + iW_\mu^2|^2 + (v+H)^2|gW_\mu^3 - g'B_\mu|^2) \end{aligned} \quad (1.40)$$

We define the $W^\pm$, Z , and A gauge bosons as:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \\ Z_\mu &= \frac{1}{\sqrt{g^2 + g'^2}}(gW_\mu^3 - g'B_\mu) \\ A_\mu &= \frac{1}{\sqrt{g^2 + g'^2}}(g'W_\mu^3 + gB_\mu) \end{aligned} \quad (1.41)$$

The Lagrangian in Equation 1.38 is therefore rewritten as:

$$\begin{aligned}
\mathcal{L}_\phi = & \frac{1}{2}(\partial_\mu H)^2 - \frac{1}{2}2 \cdot \mu^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4 \\
& + \frac{(gv)^2}{4}W^{+\mu}W_\mu^- + \frac{g^2 + g'^2}{8}v^2 Z^\mu Z_\mu \\
& + \frac{g^2 v}{2}HW^{+\mu}W_\mu^- + \frac{g^2 + g'^2}{4}v H Z^\mu Z_\mu \\
& + \frac{g^2}{4}H^2 W^{+\mu}W_\mu^- + \frac{g^2 + g'^2}{8}H^2 Z^\mu Z_\mu
\end{aligned} \tag{1.42}$$

The first comment is that the photon field is not appearing in Equation 1.42 leading to the creation of a massless field. The first line describes the scalar kinetic and self coupling as in the U(1) case, the second gives rise to the mass of the weak vector boson:

- the mass-degenerate W bosons:

$$m_{W^+} = m_{W^-} = m_W = \frac{gv}{2} \tag{1.43}$$

- the massive Z boson:

$$m_{Z^0} = \frac{\sqrt{g^2 + g'^2}}{2}v \tag{1.44}$$

- the massless photon:

$$m_A = 0 \tag{1.45}$$

The relationship between the Z and W masses are:

$$\frac{m_W}{m_Z} = \frac{g}{\sqrt{g^2 + g'^2}} = \cos \theta_W \tag{1.46}$$

g and g' are linked to the Fermi constant G_F , m_W and the electrical charge e . In the latest report published in 2021 by Particle Data Group [21], m_W is measured to be 80.379 ± 0.012 GeV and m_Z is 91.1876 ± 0.0021 GeV. And their ratio is in agreement with the Equation 1.46. The third and fourth lines are showing the linear and quadratic couplings of the Higgs boson to the gauge bosons. The direct coupling of the Higgs boson to the gauge fields from the third line is shown to be proportional to the square of their gauge boson masses. The mass of the Higgs boson can be identified as following by looking at the SSB in the U(1) only case:

$$m_H = \sqrt{2\lambda}v^2 \tag{1.47}$$

The Higgs v.e.v. (v) in itself is also determined by the Fermi constant derived from the muon lifetime:

$$v = \sqrt{\frac{1}{\sqrt{2}G_F}} \simeq 246 \text{ GeV} \tag{1.48}$$

Since λ is free parameter, the mass of the Higgs boson is not predicted by the SM.

The fermion mass term can also be generated through spontaneous symmetry breaking.

For leptons:

$$\mathcal{L}_{H,\text{leptons}} = -Y \bar{L}_L \phi L_R + \text{h.c.} \quad (1.49)$$

The couplings Y also called the Yukawa couplings, parameters that are not fixed a priori and depending on the lepton types. The h.c. stands for the hermitian conjugate part of the expression. Expanding ϕ around its ground state gives the following result:

$$\begin{aligned} \mathcal{L}_{H,\text{leptons}} &= -Y (\bar{\nu}_L, \bar{l}_L) \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H) \end{pmatrix} l_R + \text{h.c.} \\ &= -\frac{1}{2} Y \sqrt{2} v \bar{l}_L l_R - \frac{Y}{\sqrt{2}} H \bar{l}_L l_R + \text{h.c.} \end{aligned} \quad (1.50)$$

The first term is directly a mass term for the leptons with $m_l = Y \sqrt{2} v$, showing the link between the mass of a lepton and its Yukawa parameter. The couplings with the Higgs boson are also proportional to the mass of the leptons.

For quarks, the coupling of the Higgs doublet is given by:

$$\mathcal{L}_{H,\text{quarks}} = -Y_d \bar{Q}_L \phi D_R - Y_u \bar{Q}_L \tilde{\phi} U_R + \text{h.c.} \quad (1.51)$$

Expanding about the vacuum, $\tilde{\phi} = i\tau^2 \phi = \begin{pmatrix} \frac{1}{\sqrt{2}}(v + H) \\ 0 \end{pmatrix}$, gives:

$$\begin{aligned} \mathcal{L}_{H,\text{quarks}} &= -Y_d \bar{Q}_L \phi D_R - Y_u \bar{Q}_L \tilde{\phi} U_R + \text{h.c.} \\ &= -Y_d (\bar{u}_L, \bar{d}_L) \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}(v + H) \end{pmatrix} D_R - Y_u (\bar{u}_L, \bar{d}_L) \begin{pmatrix} \frac{1}{\sqrt{2}}(v + H) \\ 0 \end{pmatrix} U_R + \text{h.c.} \\ &= -Y_d \bar{d}_L \frac{1}{\sqrt{2}}(v + H) D_R - Y_u \bar{u}_L \frac{1}{\sqrt{2}}(v + H) U_R + \text{h.c.} \end{aligned} \quad (1.52)$$

The 3×3 matrices Y_d and Y_u are not necessarily diagonal but must be diagonalized in order to identify the quark masses and mass eigenstates according to the observations. We write the Y_d and Y_u in terms of diagonal matrices as:

$$Y_d = U_d \Lambda_d S_d^\dagger \quad \text{and} \quad Y_u = U_u \Lambda_u S_u^\dagger \quad (1.53)$$

Thus, the quark fields transform as:

$$\begin{aligned} u_L &\rightarrow U_u u_L \quad \text{and} \quad U_R \rightarrow S_u U_R \\ d_L &\rightarrow U_d d_L \quad \text{and} \quad D_R \rightarrow S_d D_R \end{aligned} \quad (1.54)$$

where U and S are unitary 3×3 matrices. Then the Lagrangian in Equation 1.52 becomes:

$$\begin{aligned} \mathcal{L}_{H,\text{quarks}} &= -\frac{1}{\sqrt{2}}(v + H) (\bar{d}_L U_d^\dagger U_d \Lambda_d S_d^\dagger S_d D_R + \bar{u}_L U_u^\dagger U_u \Lambda_u S_u^\dagger S_u U_R) + \text{h.c.} \\ &= -\frac{1}{\sqrt{2}}(v + H) (\bar{d}_L \Lambda_d D_R + \bar{u}_L \Lambda_u U_R) + \text{h.c.} \end{aligned} \quad (1.55)$$

since the matrices $\Lambda_{u,d}$ are diagonal, we may get the mass term of quarks by defining $m_{u,d}^i = \Lambda_{u,d}^{ii}$ where i indicates the generation index. Furthermore, it is particularly important when looking at the flavour changing currents such as in the electroweak theory:

$$\bar{u}_L \gamma^\mu d_L W_\mu^+ \rightarrow \bar{u}_L \gamma^\mu (U_u^\dagger U_d) d_L W_\mu^+ \quad (1.56)$$

The $(U_u^\dagger U_d)$ is known as the Cabibbo–Kobayashi–Maskawa (CKM) matrix [27, 28] which can be written in terms of the mixing angles $(\theta_1, \theta_2, \theta_3)$ and a CP-violating phase. The CKM matrix are free parameters of the SM and need to be determined experimentally.

1.1.5 Limitations of the Standard Model

All the particles predicted by the SM have been discovered. Experimentally, the SM is the most successful theory of particle physics to date. However, the SM is not perfect:

Unexplained phenomena The SM is inherently incomplete. Here are some examples. The SM does not include the gravitational interaction, one of the four fundamental interactions in the Universe. The SM is also incompatible with the general relativity, the most successful theory of gravity [29]. The SM does not well explain dark matter and dark energy, which present 95% of the mass-energy in the Universe according to the cosmological observations [30]. The SM predicts massless neutrinos which contradicts the neutrino oscillation experiments where the neutrinos do have mass [31, 32]. There is no mechanism in the SM to explain the disproportionate noted matter relative to antimatter [33].

Theoretical problems The SM was built in an ad hoc way. These ad hoc features imply a lack of understandings. For example, in the SM, the mass of the Higgs boson gets some very large quantum corrections due to the presence of virtual particles (mostly virtual top quarks). These corrections are much larger than the actual mass of the Higgs boson. It is the famous hierarchy problem and deemed unnatural by many theorists.

Hints from experimental results There are some experimental evidence contradicting the SM at the 5σ level. The choice of the 5σ level is explained in Section 5.1. The measured value of the muon’s anomalous magnetic dipole moment (known as $g - 2$) is significantly different from the SM prediction [34]. The published results in 2021 from Fermilab exhibit a deviation of 4.2 standard deviation with respect to the SM predictions [35]. The latest measurement of $B -$ meson decay reported a 5σ deviation from the SM [36]. Last, but not the end, the CDF collaboration reported a measurement for the mass of the W boson which exceeds the mass predicted by the Standard Model with a significance of 7σ [37].

Therefore, both experimentalists and theorists in particle physics are motivated to search for new physics. For experimentalists, the new physics searches employ one of two strategies: direct searches where decays of BSM particles are observed directly and indirect searches which look for discrepancy between experimentally measured and theoretically predicted values of the properties of SM particles. In this thesis, I focus on the indirect searches using the Higgs boson.

1.2 The Higgs boson at the Large Hadron Collider

The discovery of the Higgs boson was the necessary step to prove that the Higgs mechanism is the correct theory to describe the mass of the elementary particles. Motivated by this, scientists designed high energy colliders like LEP, Tevatron and LHC for probing a variety of physics processes and aiming at discovering the Higgs boson. Finally, with the announcement of the discovery of Higgs boson in July 2012, the ATLAS and CMS collaborations at the LHC confirmed experimentally the Higgs mechanism with a Higgs boson

mass around 125 GeV [6, 7]. It was an important step for the SM but the journey is far from over. An age of precision measurements and verification of the predicted Higgs properties started.

1.2.1 Production modes

The main Higgs production mechanisms at the LHC are gluon fusion, vector boson fusion, associated production with a gauge boson, associated production with heavy quarks. The detailed Feynman diagrams of these processes are shown in Figure 1.3. The cross sections for the production of a SM Higgs boson as a function of the center of mass energy $\sqrt{s}$, for pp collisions, are summarized in Figure 1.4a. Table 1.1 summarizes the Higgs boson production cross sections and relative uncertainties for a Higgs boson mass of 125 GeV for $\sqrt{s}$ at 7, 8, 13 and 14 TeV.

The gluon fusion (ggF) production is the production mode with the largest cross section at high-energy hadron colliders. As shown in Figure 1.3(a), the Higgs boson is mediated by the exchange of a virtual, heavy top (bottom) quark. Since the Higgs coupling to fermions is proportional to their masses, the process is enhanced and can be found only at the LHC thanks to loops. This production mode is in the order of ten times larger than the second largest one.

The vector boson fusion (VBF) production is the production mode with the second-largest cross section at the LHC. As shown in Figure 1.3(b), the production proceeds by the scattering of two quarks, mediated by exchange of a W or Z boson, with the Higgs boson radiated off the weak-boson propagator. The scattered quarks give rise to two back-to-back hard jets in the forward and backward regions of the detector. Because of the color-singlet nature of the weak-gauge boson exchange, gluon radiation from the central-rapidity regions is strongly suppressed.

The associated production to a vector boson (VH) is a process where a Higgs boson is radiated (Higgs-strahlung) from an initial particle which is present in the final state. Amongst all the possible initial particles, the vector bosons are favoured due to the high coupling with the Higgs boson, as shown in the LO diagram in Figure 1.3(c).

The production in association with heavy quarks gives a small contribution to the total Higgs production cross section. As shown in 1.3(e-g), the Higgs boson couples directly to fermions in the initial state. Due to its important mass, couplings with top quarks are dominant. This mode is an unique way to understand the direct coupling to the top-quark as it only enters in loops for the other production modes. However, due to the high $t\bar{t}$ QCD production at the LHC, this channel is very challenging to tackle experimentally.

1.2.2 Decay modes

The lifetime of the Higgs boson is too short to be measured directly. Only its decay products are accessible to the experimentalists. The width of each decay channel of the Higgs boson

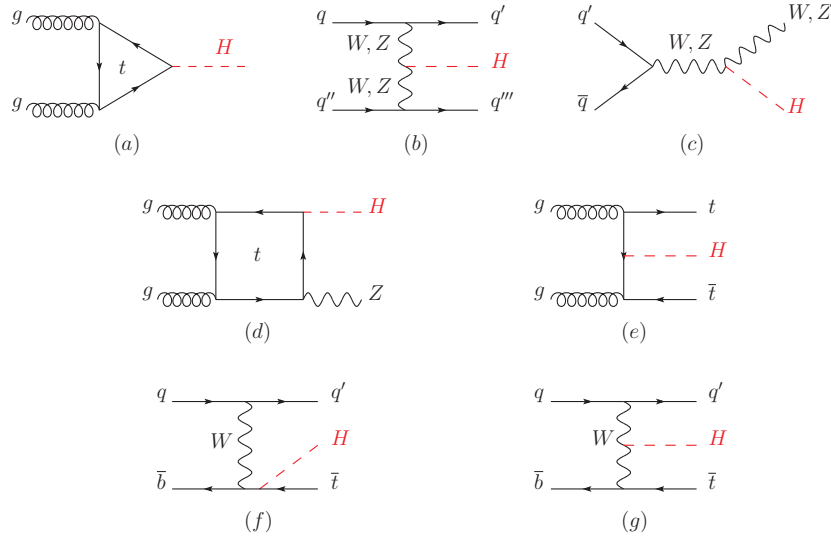


Figure 1.3: Main leading order Feynman diagrams contributing to the Higgs boson production in (a) the gluon fusion, (b) the vector-boson fusion, (c) the associated production with a gauge boson at tree level from a quark-quark interaction, (d) the associated production with a gauge boson (at loop level from a gluon-gluon interaction), (e) the associated production with a pair of top quarks (there is a similar diagram for the associated production with a pair of bottom quarks), (f-g) the associated production with a single top quark.

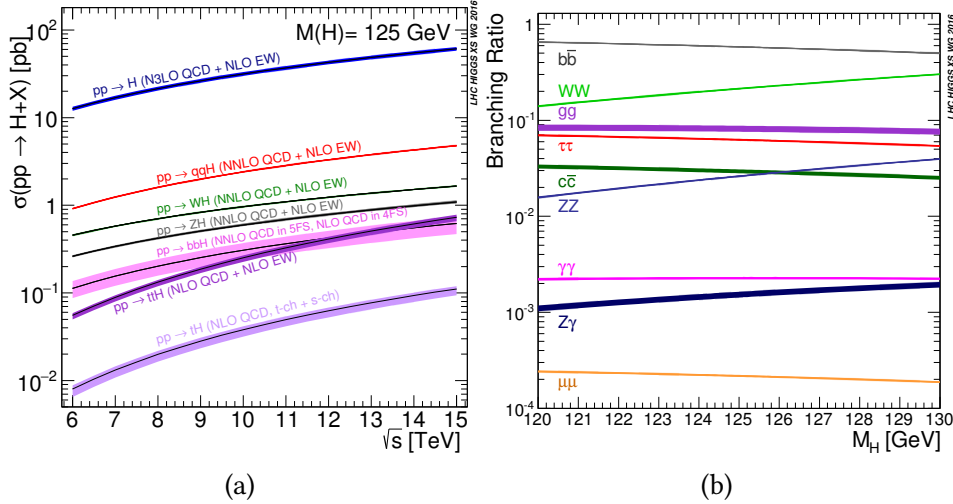


Figure 1.4: (a) Production cross sections for a SM Higgs boson with a mass of GeV from the proton-proton collisions at the LHC [38]. The theoretical uncertainties are indicated as bands. (b) The branching ratios for the main decays of the SM Higgs boson near $m_H = 125$ GeV. The theoretical uncertainties are indicated as bands.

can be derived using the rules in Section 1.1.4:

$$\begin{aligned} \Gamma(H \rightarrow f\bar{f}) &= \frac{m_f^2}{v^2} N_C \frac{m_H}{8\pi} \left(1 - 4 \frac{m_f^2}{m_H^2}\right)^{3/2} \\ \Gamma(H \rightarrow VV^*) &= \delta_V \frac{m_H^2}{v^2} N_C \frac{m_H}{32\pi} \left(1 - 4 \frac{m_V^2}{m_H^2}\right)^{1/2} \left(1 - 4 \frac{m_V^2}{m_H^2} - 12 \left(\frac{m_V^2}{m_H^2}\right)^2\right) \end{aligned} \quad (1.57)$$

Table 1.1: Production cross sections of the $m_H = 125$ GeV Higgs boson at the LHC [38].

$\sqrt{s}$ (TeV)	Production cross section (in pb) for $m_H = 125$ GeV					
	ggF	VBF	WH	ZH	$t\bar{t}H$	total
7	$16.9^{+5.5\%}_{-7.6\%}$	$1.24^{+2.2\%}_{-2.2\%}$	$0.58^{+2.2\%}_{-2.3\%}$	$0.34^{+3.1\%}_{-3.0\%}$	$0.09^{+5.6\%}_{-10.2\%}$	19.1
8	$21.4^{+5.4\%}_{-7.6\%}$	$1.60^{+2.1\%}_{-2.1\%}$	$0.70^{+2.1\%}_{-2.2\%}$	$0.42^{+3.4\%}_{-2.9\%}$	$0.13^{+5.9\%}_{-10.1\%}$	24.2
13	$48.6^{+5.6\%}_{-7.4\%}$	$3.78^{+2.1\%}_{-2.1\%}$	$1.37^{+2.0\%}_{-2.0\%}$	$0.88^{+4.1\%}_{-3.5\%}$	$0.50^{+6.8\%}_{-9.9\%}$	55.1
14	$54.7^{+5.6\%}_{-7.4\%}$	$4.28^{+2.1\%}_{-2.1\%}$	$1.51^{+1.8\%}_{-1.9\%}$	$0.99^{+4.1\%}_{-3.7\%}$	$0.61^{+6.9\%}_{-9.8\%}$	62.1

where N_c is the colour factor ($N_c = 3$ when considering quarks, $N_c = 1$ for leptons), and δ_V a factor that normalises the number of diagrams giving the same final state ($\delta_W = 2$ while $\delta_Z = 1$). In order to interpret further on the probability of decay, a Branching Ratio is defined as:

$$BR(H \rightarrow XX) = \frac{\Gamma(H \rightarrow XX)}{\sum_i \Gamma(H \rightarrow ii)} \quad (1.58)$$

From Equation 1.57, the decay width of each channel are highly correlated to the mass of the decay particles. The value for the branching ratio for the different modes of decay as a function of the Higgs boson mass is given in Figure 1.4b. A Higgs boson mass of about 125 GeV allows to explore the Higgs boson couplings to many SM particles. In particular the dominant decay modes are $H \rightarrow b\bar{b}$ and $H \rightarrow W^+W^-$, followed by $H \rightarrow gg$, $H \rightarrow \tau^+\tau^-$, $H \rightarrow c\bar{c}$ and $H \rightarrow ZZ$. With much smaller rates follow the Higgs boson decays into $H \rightarrow \gamma\gamma$, $H \rightarrow Z\gamma$ and $H \rightarrow \mu^+\mu^-$. Since the decays into gluons, diphotons and $Z\gamma$ are loop induced, they provide indirect information on the Higgs boson couplings to WW , ZZ and $t\bar{t}$ in different combinations. The uncertainties in the branching ratios include the missing higher-order corrections in the theoretical calculations as well as the errors in the SM input parameters, in particular fermion masses and the QCD gauge coupling, involved in the decay. The theoretical calculations of the BR for each decay channel for a SM Higgs boson with $m_H = 125$ GeV are shown in Table 1.2.

Having the largest BR among all decay modes, $H \rightarrow b\bar{b}$ is the most promising channel for probing the coupling of the Higgs field to fermions at hadron colliders. However, after the announcement of the Higgs discovery, experimentalists have spent six years before observing $H \rightarrow b\bar{b}$ because the presence of very large QCD backgrounds makes the identification of a Higgs boson signal in this channel quite challenging. The identification of $H \rightarrow b\bar{b}$ is thus a crucial topic. Part of my thesis work focused in studying the $H \rightarrow b\bar{b}$ decay produced in association to a vector boson, for the first time, in the fully hadronic channel, $V(\rightarrow qq)H(\rightarrow b\bar{b})$. The Higgs bosons in my studies have a large transverse momentum and have a large Lorentz boost. The motivations of using boosted Higgs bosons are presented in the following section.

Table 1.2: Branching ratios and the relative uncertainty calculated for a SM Higgs boson with $m_H = 125$ GeV [38].

Decay channel	Branching ratio	Rel. uncertainty
$H \rightarrow b\bar{b}$	5.82×10^{-1}	$+1.2\%$ -1.3%
$H \rightarrow W^+W^-$	2.14×10^{-1}	$\pm 1.5\%$
$H \rightarrow \tau^+\tau^-$	6.27×10^{-2}	$\pm 1.6\%$
$H \rightarrow c\bar{c}$	2.89×10^{-2}	$+5.5\%$ -2.0%
$H \rightarrow ZZ$	2.62×10^{-2}	$\pm 1.5\%$
$H \rightarrow \gamma\gamma$	2.27×10^{-3}	$\pm 2.1\%$
$H \rightarrow Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \rightarrow \mu^+\mu^-$	2.18×10^{-4}	$\pm 1.7\%$

1.2.3 Boosted Higgs bosons

The high collision energies at the LHC can result in the production of Higgs bosons with transverse momenta, p_T , much larger than their mass. Such Higgs bosons are boosted: their decay products are highly collimated, and for fully hadronic decays they can be reconstructed as a single hadronic jet (a useful rule of thumb is $2M/p_T \sim R$: twice the jet mass divided by the p_T is roughly equal to the maximum opening angle of the two decay products). The study of boosted boson Higgs processes is very interesting from the theoretical and experimental point of view.

From the theoretical point of view, the SM predicted p_T^H shape is remarkably stable with respect to higher order corrections and it may provide a way to identify new physics that enters in the high p_T regime. As shown by Figure 1.5, the QCD corrections to the Higgs boson transverse momentum distribution increase the leading order result by almost a factor of two. However, their magnitude appears to be quite similar to the QCD corrections computed in the infinite top-mass approximation (HEFT) [8]. The difference of the two results is close to five percent. If new physics enter in this region, the p_T distribution will be distorted and deviations from SM predictions will be observed. The Standard Model Effective Field Theory (SMEFT) is one of the current theoretical frameworks available to parameterise the deviations from the SM. Any deviation is parameterised with higher-dimension operators, in the first approximation dimension six [39–41]. For example, the chromomagnetic dipole moment operator is the parameter to modify the interactions between gluons and the top quark. The studies of the impact of the chromomagnetic operator found that, as shown by Figure 1.6, it can significantly distort the transverse-momentum spectrum of the Higgs boson at large p_T and the impact depends on the actual size of the corresponding Wilson coefficient describing the chromomagnetic dipole moment operator. Therefore, it is very interesting to explore the Higgs boson production in the high p_T regime.

From the experimental point of view, more and more boosted Higgs bosons are produced as the LHC operates at higher energy. During years of runs and upgrades, the

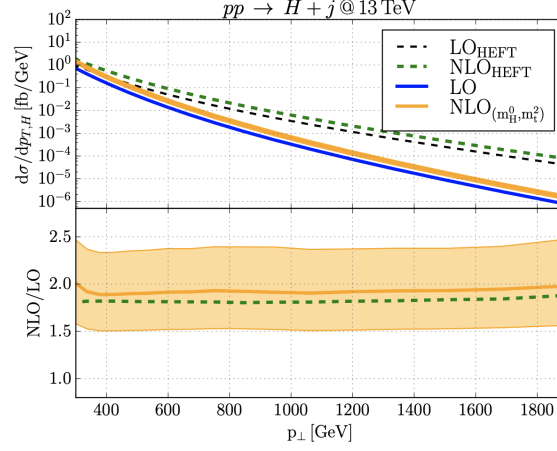


Figure 1.5: Transverse momentum distribution of the Higgs boson at the LHC with $\sqrt{s} = 13$ TeV. The upper panel shows absolute predictions at LO and NLO in the full SM and in the infinite top-mass approximation (HEFT). The lower panel shows respective NLO/LO correction factors. The bands indicate theoretical errors of the full SM result due to scale variation [8].

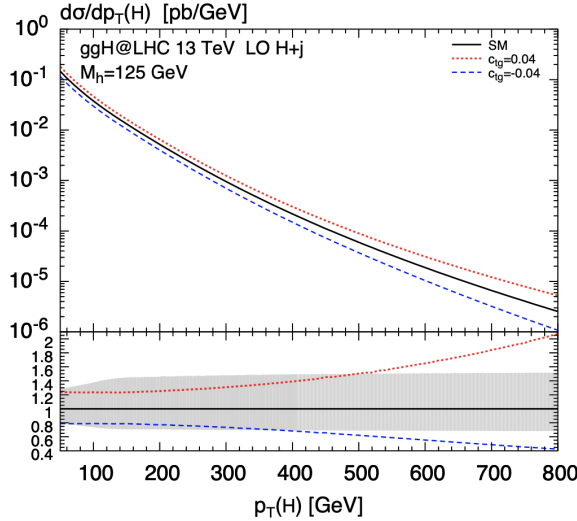


Figure 1.6: Impact of the chromomagnetic operator on the p_T spectrum of the Higgs boson in the region allowed by the current experimental constraints [9].

LHC has been operated from $\sqrt{s} = 7$ to 13 TeV and the production of the Higgs boson at $m_H = 125$ GeV from pp collisions increased from 19.1 fb to 55.1 fb as shown in Table 1.1. The pp collision energy of the LHC increased to $\sqrt{s} = 13.6$ TeV since the summer of 2022. In the future, the LHC will operate at an even higher energy. More data will be collected. This allows us to increase the precision of the Higgs boson measurements at high transverse momentum. On the other hand, studying the boosted Higgs boson is challenging and requires special reconstruction techniques as will be discussed in Chapter 4.

1.3 Summary

The SM has been presented in this chapter, including its limitations, our motivations to search new physics. The Higgs boson sector was presented in particular as it is essential for the mass generation of the particles in the SM. Furthermore, the boosted Higgs boson presents advantages from both a theoretical and an experimental point of view. Therefore, the studies in Chapter 4 and 5 will focus on the boosted Higgs bosons. The LHC is also moving to higher collision energies in order to produce more boosted Higgs bosons, thus, the work in Chapter 7 will focus on the improvement of the ATLAS tracker for high luminosity phase of LHC.

THE ATLAS DETECTOR

In order to probe the SM and to search for new physics, high energy colliders and particle detectors have been developed in the last decades. The work in this thesis uses proton-proton collision data recorded by the **A Toroidal LHC ApparatuS** (ATLAS) experiment at the Large Hadron Collider (LHC). In this chapter, the LHC is presented in Section 2.1, followed by the structure and the design of the ATLAS detector in Section 2.2 and the data acquisition system in Section 2.3.

2.1 The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) [1] in the European organization for nuclear research (CERN), beneath the France–Switzerland border near Geneva, is today’s largest and most powerful particle accelerator to probe the elementary particle physics. The LHC uses protons and ions (Lead and Xenon) as primary particles which have been boosted by CERN accelerator complex. The current CERN accelerator complex includes several accelerators as shown by Figure 2.1. Since the studies presented in the manuscript uses proton collisions, the ion-related acceleration and collisions are not described here. The initial protons originate from bottled dihydrogen. The hydrogen atoms are ionized and separated, then injected to the first accelerator, the LINAC 2, where they are to be accelerated to 50 MeV. Then the protons reach the BOOSTER synchrotron to be accelerated to 1.4 GeV. Then in the Proton Synchrotron (PS), they become ultra-relativistic proton bunches with an energy of 25 GeV and a bunch structure of 81 bunch packets with 25 ns spacing. Afterwards, the protons are directed to the Super Proton Synchrotron (SPS) and accelerated to 450 GeV. Eventually, the LHC is filled with 2808 bunches per beam and are accelerated up to 6.5 TeV. Each bunch contains about 10^{11} protons. There are two streams of protons circulating in the LHC in counter way. While the two beams circulate in different beam pipes, when they approach the Interaction Points (IP) they are focalised to cross each other. The collisions happen every 25 ns, i.e. at the bunch crossing frequency of 40 MHz. There are four IPs in the LHC ring, corresponding to four main detectors to record the collision data for different physics purposes: **A Large Ion Collider Experiment** (ALICE)[2], **A Toroidal LHC ApparatuS** (ATLAS) [3], **Compact Muon Solenoid** (CMS) [4] and **Large Hadron Collider beauty** (LHCb) [5].

CERN's Accelerator Complex

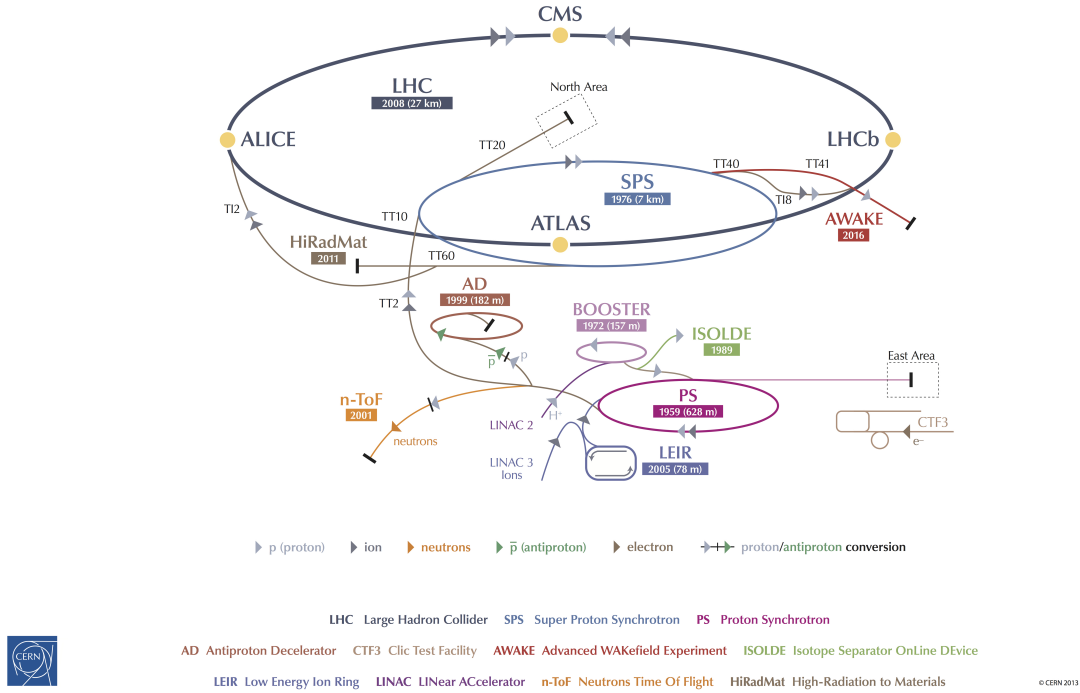


Figure 2.1: Overview of the CERN accelerator complex [42].

The major figure of merit for collisions is the **instantaneous luminosity** $\mathcal{L}$. It is defined as the number of events N produced per second for a process which has a cross-section of σ at that given energy:

$$\sigma \mathcal{L} = \frac{dN}{dt} \quad (2.1)$$

This factor is linked to the collision scheme provided by the accelerator machine.

$$\mathcal{L} = \frac{N_p^2 n_b f \gamma}{4\pi \sigma_1 \sigma_2} F \quad (2.2)$$

where N_p is the number of protons per bunch, n_b the number of bunches circulating in the collider, f the revolution frequency of revolving N_p particles in each beam, γ the relativistic factor associated to the particles in the bunch, $\sigma_{1,2}$ is the transverse width of the bunch 1 and 2, F the geometrical factor of the collision scheme that takes into account the fact that collisions are not happening head-on but with a small crossing angle.

To describe the total data delivered by the LHC and recorded by the experiments, the **integrated luminosity** is used:

$$\mathcal{L}_{int} = \int L dt \quad (2.3)$$

The LHC has been continuously upgraded to increase the collision's energy and luminosity. Figure 2.2 shows the operation and upgrade plan of the LHC. The LHC started its



Figure 2.2: LHC/HL-LHC upgrade plan [43].

first run in 2011 and ended at the beginning of 2013 running at centre-of-mass energies of 7 TeV and 8 TeV. During that period, the LHC operated at 75% of its original designed luminosity and the total integrated luminosity delivered was 30 fb⁻¹. In 2013-2014, the LHC went the first Long Shutdown (LS1) period for upgrades and tests. After the LS1, the LHC reached its designed instantaneous luminosity of $1 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. Between 2015 and 2018, corresponding to the LHC Run 2, the LHC operated a centre-of-mass energy with luminosity higher than the designed value. The peak values attained in 2018 were at 13 TeV and reached approximately $2.4 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ under normal running conditions thanks to the excellent beam quality. The total integrated luminosity delivered to the LHC was 190 fb⁻¹. After the second Long Shutdown (LS2) in 2019-2022, the LHC is going to operate at an instantaneous luminosity around twice the designed value from 2022 to 2025, so-called Run 3. The centre-of-mass energy will be at 13.6 TeV. The total integrated luminosity delivered from LHC is expected to be 450 fb⁻¹. After Run 3, the LHC will enter into three years of LS3 for installation of new upgrades for High Luminosity phase of LHC (HL-LHC). The HL-LHC runs are planned to start from 2029 with luminosity 5 to 7.5 times the design one.

The work in this thesis is based on proton-proton collisions data recorded by ATLAS during LHC Run 2. Figure 2.3 shows that the total integrated luminosity delivered to ATLAS at end of Run 2 in 2018 is 156 fb⁻¹. However, the integrated luminosity recorded by ATLAS and good quality data are not equal to the delivered one. The recorded luminosity reflects the data acquisition inefficiency, as well as the inefficiency of the so-called "warm start": when the stable beam flag is raised, the tracking detectors undergo a ramp of the high-voltage and, for the pixel system, the preamplifiers are turned on. The good quality data is data passing the data quality assessment in ATLAS [45]. Only the good quality data is used for physics analyses.

The number of pp interactions per bunch crossing can be increased in order to increase the luminosity, according to Equation 2.1. At each bunch crossing, several collisions can happen which is helpful to increase the probability to get rare and interesting events. However, the surrounding soft collisions are a challenge for the reconstruction of the physics

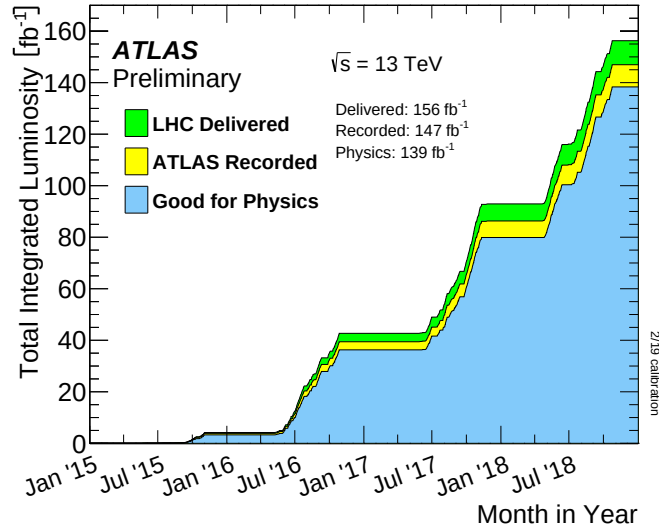


Figure 2.3: Integrated luminosity versus time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015-2018 [44].

objects, degrading the resolution of the quantities measured by the detectors. The effect is known as pile-up and can be caused by particles from the same bunch-crossing (in-time), or due to remnants of the interactions of the previous bunch crossing (out-of-time). The major contribution of pile-up is out-of-time. The mean number of interactions, $\langle \mu \rangle$, is used to describe the out-of-time pile-up. As shown in Figure 2.4, the mean value for the Run 2 is approximately 34. The HL-LHC will have to face a mean number of interactions of roughly 200 with consequent experimental challenges to detector operations and event reconstructions, more details will be given in Section 7.1.

2.2 The ATLAS detector

The ATLAS [3] is a multi-purpose detector to study a large variety of physics from SM measurements to BSM searches. In order to increase the geometrical acceptance and the hermiticity, the detector has the shape of a cylinder 46 m long and 25 m in diameter, as shown in Figure 2.5a. From its shape, a right-handed coordinate system is defined as the convention for all studies in ATLAS, as shown in 2.5b. The z -axis is the axis of the beam direction. The x -axis points towards the centre of the LHC ring. The y -axis points upwards. The x - and y -axes form the transverse plane at the collision point. Cylindrical coordinates (r, ϕ) are used in the transverse plane and the pseudorapidity η is defined in terms of the polar angle θ as

$$\eta = -\ln \tan(\theta/2) \quad (2.4)$$

The detector is a multi-layered instrument. It consists of six different detecting sub-systems wrapped concentrically in layers around the interaction point. Figure 2.5a shows the structure of the detector. It includes the Inner Detector (Section 2.2.1), the calorimetry system (Section 2.2.2), the muon system (Section 2.2.3) and a huge magnet system (Section

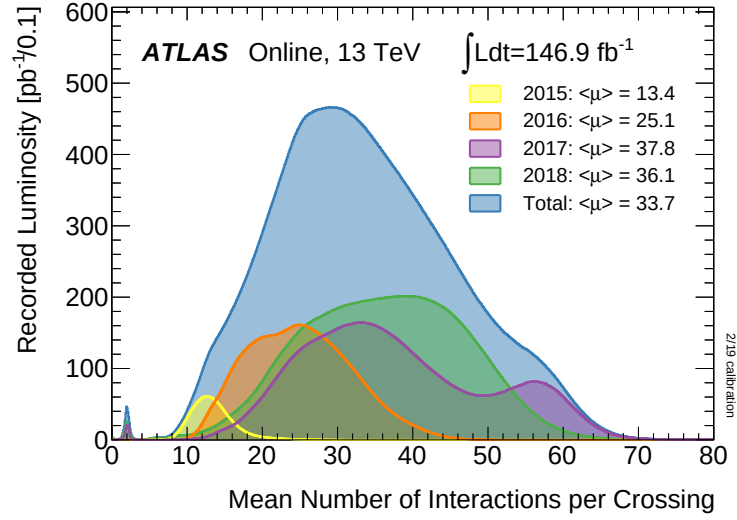


Figure 2.4: Luminosity-weighted distribution of the mean number of interactions per bunch crossing for the different years of operations during Run 2. The mean number of interactions per crossing is calculated from the instantaneous per bunch luminosity as $\mu = \mathcal{L}_{\text{bunch}} \times \sigma_{\text{inel}} / f_r$ where $\mathcal{L}_{\text{bunch}}$ is the per bunch instantaneous luminosity, σ_{inel} is the inelastic cross section which we take to be 80 mb for 13 TeV collisions, and f_r is the LHC revolution frequency [44].

2.2.4). When particles are passing through, they interact with the sub-systems which record their trajectory, momentum, and energy. The interaction of particles with the sub-systems varies with each particle, as shown in Figure 2.6 and determines the materials used by the sub-detectors, the particle reconstruction and identification methods used in ATLAS. The data acquisition system of ATLAS is described in Section 2.3.

2.2.1 The inner detector

The Inner Detector (ID) [48] is the innermost sub-detector, closest to the collision point. It consists of three different systems using three technologies: silicon pixel detectors, silicon

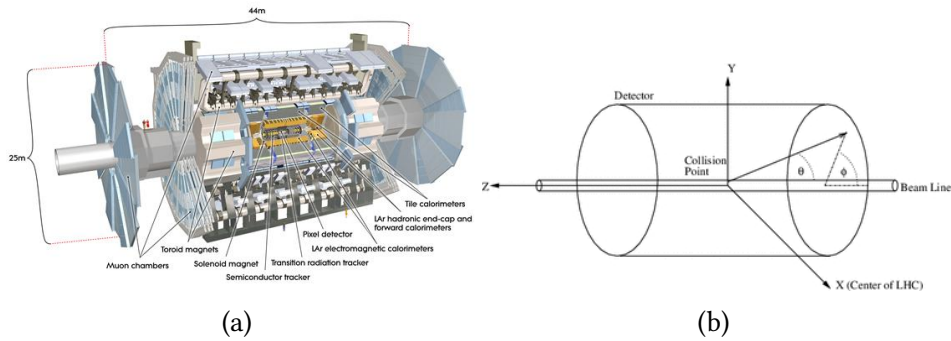


Figure 2.5: (a) Schematic of the ATLAS detector including the relevant sub-detectors. (b) ATLAS system of coordinates. [46]

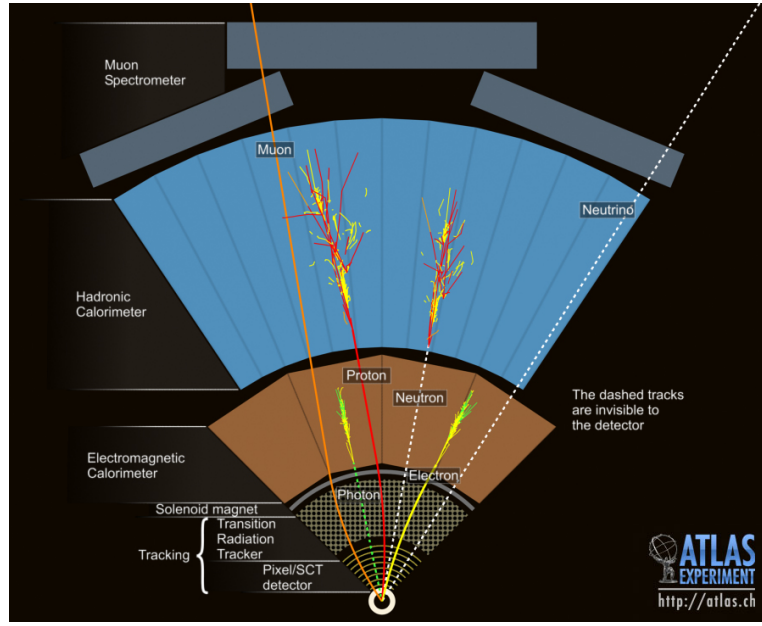


Figure 2.6: Schematic view of the interaction of different particles with the ATLAS detector [47].

strip detectors and straw drift tubes, all surrounded by a thin superconducting solenoid providing a 2 T axial magnetic field. The ID is designed to perform charged particle tracking within $|\eta| < 2.5$. The main components of the ID are the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT). Figure 2.7 shows the overview of the Inner Detector, and the Figure 2.8 shows a 3D visualisation of the structure of the barrel of the ID.

2.2.1.1 The Pixel Detector

The flux of particles is particularly large near the collision point, thus a radiation-hard and fine segmented detector is needed. The Pixel Detector [50] is based on the semi-conductor technology which is able to cope with the required radiation hardness. The fundamental of the semi-conductor technology and the particle detection principle are presented in Chapter 6. The Pixel Detector consists of four layers of silicon pixels (Figure 2.9), with over 92 million pixels. The inner-most layer is called the Insertable B Layer (IBL), a new tracking layer added for Run 2 which is the closest to the beam line and designed to improve the precision and robustness of track reconstruction. The IBL consists of 280 silicon pixel modules arranged on 14 azimuthal carbon fibre staves surrounding the beam pipe at a radius of 33.25 mm. Each staff is instrumented with 12 two-chip planar modules, covering the region of $\eta < 2.7$, and 8 single-chip modules with 3D sensors, four at each end of the staff ($2.7 < |\eta| < 3$). The thickness of the IBL pixels is between $150 - 250 \mu\text{m}$. The pixel size for IBL pixels is $50 \mu\text{m} \times 250 \mu\text{m}$, corresponding to an intrinsic spatial resolution of $10 \mu\text{m} \times 60 \mu\text{m}$. The remaining part of the Pixel Detector consists of 1744 silicon pixel modules arranged in three barrel layers and two endcaps with three disks each. The three barrel layers are known as B-Layer, Layer-1 and Layer-2, respectively, and their radii are shown in Figure 2.9. They provide an η coverage up to 2.5 and a complete ϕ coverage. The

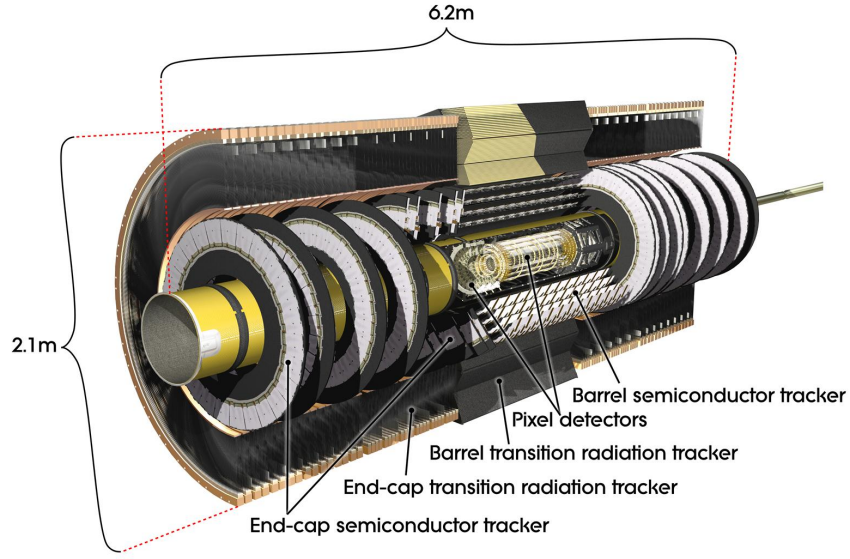


Figure 2.7: Schematic overview of the Inner Detector [48].

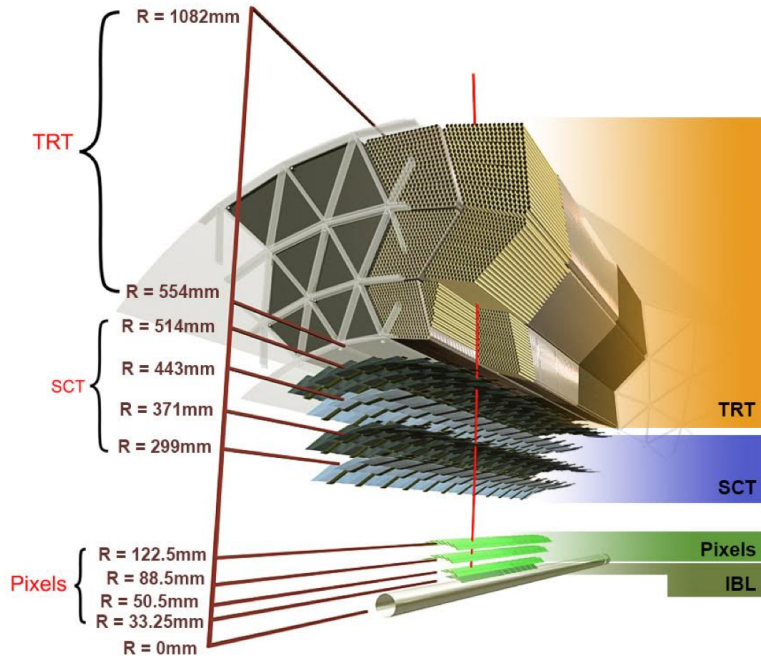


Figure 2.8: A 3D visualisation of the structure of the ID barrel: the beam pipe, the Insertable B Layer (IBL), the Pixel layers of the Pixel Detector, the four cylindrical layers of the SCT and the three layers of the TRT barrel modules [49].

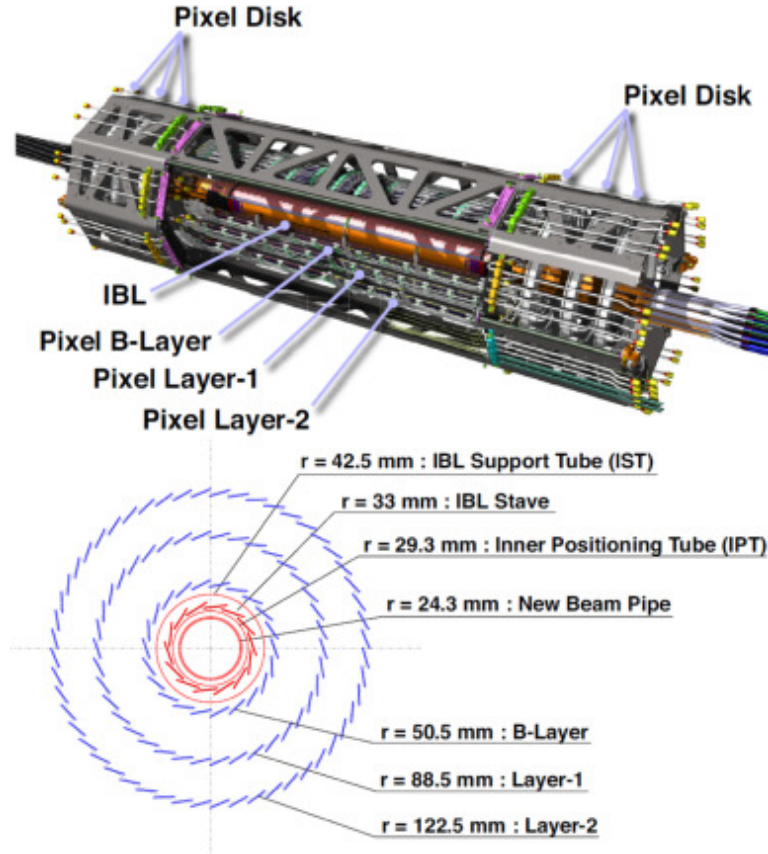


Figure 2.9: Schematic view of the Pixel Detector [50].

barrel modules were assembled on staves of 13 modules each, whilst the endcap modules were assembled directly on the disks. Each module is built with a $250\text{ }\mu\text{m}$ thick layer of silicon with pixel size of $50\text{ }\mu\text{m} \times 400\text{ }\mu\text{m}$, corresponding to an intrinsic spatial resolution of $10\text{ }\mu\text{m}$ in the $r - \phi$ direction and $115\text{ }\mu\text{m}$ in the z -direction.

2.2.1.2 The Semiconductor Tracker

The Semiconductor Tracker (SCT) [51] is the middle part of the ID. It provides an η coverage up to 2.5 and a complete ϕ coverage. It consists of 4088 silicon strip modules. They are arranged in four barrel layers and two endcaps with nine disks each. Each module consists of a pair of single-sided strip sensors glued back-to-back with a 40 mrad angle between them and 1 mm distance. Each module comprises twelve 128-channel chips. The barrel modules are mounted directly on the cylindrical support structures of each layer in 12 rings, whilst the endcap modules are assembled in 3 rings on the disks. The barrel SCT sensors have a uniform pitch strip of $80\text{ }\mu\text{m}$, corresponding the resolution of $17\text{ }\mu\text{m}$ along $r - \phi$, while the endcap sensor strips run radially with a $161.5\text{ }\mu\text{rad}$ angular pitch. Its detection principle is similar to that of the Pixel Detector, although the lower particle density allows using silicon strips rather than small rectangular pixels. The strips are configured in two layers under a small angle with respect to each other, such that a position measurement along the strip length can be obtained from hits in overlapping strips.

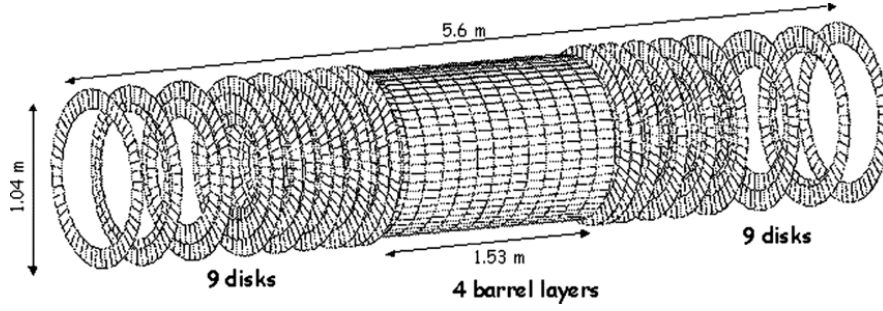


Figure 2.10: Schematic view of the Semiconductor Tracker (SCT) [51].

2.2.1.3 Transition Radiation Tracker (TRT)

The Transition Radiation Tracker (TRT) [52] is the outermost subdetector of the tracking system and extends track reconstruction radially outwards to a radius of 1082 mm. It is made of 298 304 gas-filled straw tubes of 4 mm diameter. The barrel contains 96 modules in 3 layers (32 modules per layer). The straws are parallel to the beam axis and are 144 cm long. The two endcaps are built with 160 layers of 37 cm long straws. Each endcap contains 40 disks. It has η coverage up to 2.0 and only provide $r - \phi$ information. The resolution per straw in the $r - \phi$ direction is $130 \mu\text{m}$. Particles passing through the tubes ionize the gas inside the tube, and the electrons drift under the electric field between the walls of the tube and the thin wire going through the center of the tube. The electrons are collected by the thin wire and used to build a signal that allows determining if there was a hit, and how far from the center of the straw it occurred (since the drift time is converted into the distance of the track to the wire).

2.2.2 The calorimetry system

The calorimetry system is designed to measure the energy a particle loses as it passes through the detector and to stop all interacting particles apart from muons and neutrinos. Muons are highly penetrating and are not absorbed in the calorimeters. Neutrinos interact with very low probability and remain mostly undetected. The energy loss of the incoming particles is measured through the shower development created in the calorimeters. Ionisation, nuclear reactions and atom excitations are triggered in showers. The secondary particles are then generated and the initial energy of the incoming particles is degraded. The shower development is directly linked to the nature of the particle, as well as to the materials used. There are two basic types of showers according to their interactions. **Electromagnetic showers** are produced by a particle that interacts primarily via the electromagnetic force (where bremsstrahlung and photon conversion effects dominate, usually a photon or electron). **Hadronic showers** are produced by hadrons and proceed mostly via the strong nuclear force. Therefore separate detectors are constructed in ATLAS to enhance those effects, allowing a discrimination of the particles depending on whether they interact more with the former or the latter. Globally, the electromagnetic calorimeter measures the energy of electrons and photons and the hadronic calorimeter samples the energy of hadrons. By computing the expected energy balance of the event in the transverse plane, the transverse component of the undetected energy by neutrinos, known as the **missing transverse en-**

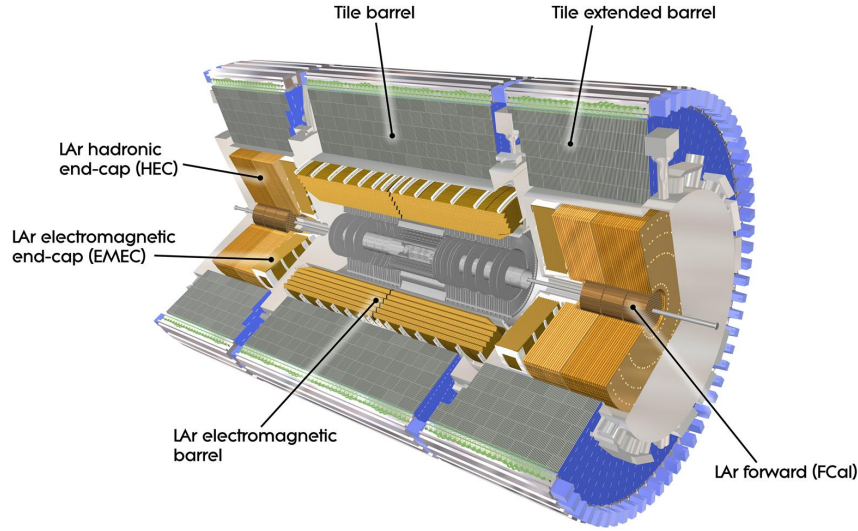


Figure 2.11: Schematic view of the ATLAS calorimetry system [53].

ergy (MET), can be also estimated. The geometry of the calorimeters is designed to have the largest possible coverage. The full system is presented in Figure 2.11. ATLAS includes an electromagnetic and a hadronic calorimeter, with different characteristics in order to account for the different properties of electromagnetic and hadronic showers.

The ATLAS calorimeter system uses sampling technology, i.e. the layers of passive dense material with high stopping power (the absorber) alternated with layers of active material. Particle showers are produced through interactions of the incident particle in both the absorber and the active material. Particles deposit their energy in both the absorber and the active material, even though only the energy deposited in the active material is measured. If the number of successive layers is large, the primary particle will lose almost all of its initial energy. But this energy is not completely measured by the calorimeter system even with an optimal design, due mainly to energy that remains in the absorber, in non-instrumented regions of the detector and to the non compensation of the calorimeter. For this reason, calibrations need to be applied to properly measure the energy of the incident particle.

One important characteristic of calorimeters is their resolution. The resolution of the calorimeter improves with increasing energy since, for sampling, the calorimetry is based on statistical processes. The energy resolution can be expanded as:

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E[\text{GeV}]}} \oplus C \quad (2.5)$$

The first term represents the stochastic term linked to the statistical fluctuation of the shower, and the last term is due to the detector non-uniformities, alignment and electronics calibrations.

2.2.2.1 The electromagnetic calorimeter

The electromagnetic (EM) calorimeter is a lead-liquid argon (LAr) sampling calorimeter with accordion geometry. More details can be found in [54, 55]. It is divided into a barrel region, covering $|\eta| < 1.475$, and two endcaps covering $1.375 < |\eta| < 3.2$. The width of the active LAr gaps is 2 mm in the barrel and ranges from 1.2 mm to 2.7 mm in the endcap. The calorimeter is segmented into three longitudinal layers for $|\eta| < 2.5$. The first layer is finely segmented in the η direction to provide $\gamma - \pi^0$ separation and photon direction measurement. The cell size in η is for instance 0.003125 in the barrel region. The second layer collects most of the shower energy and has a granularity of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ for $|\eta| < 2.5$. The last layer is used to correct for leakage behind the electromagnetic calorimeter. The EM calorimeter is located in the cryostat since liquid argon needs to be kept at a temperature about 88 K. The inner radius of the cryostat is 1385 mm, and its outer radius is 2132 mm. The barrel component shares its cryostat vessel with the solenoid magnet (see Section 2.2.4) in order to minimize the amount of inactive material. Between the barrel and each endcap, some space (known as crack region $1.375 < |\eta| < 1.52$) is available for cables and services for the inner detector. In addition, a thin presampler layer is located between the cryostat and the calorimeter, covering the region $|\eta| < 1.8$. It is used to correct for fluctuations in the energy loss upstream the calorimeter.

The characteristic interaction distance of an electromagnetic shower is the radiation length of the material, defined as the mean distance over which an electron loses all but $1/e$ of its energy. Due to the position of the central solenoid coil inside the LAr barrel calorimeter and the aluminum cryostat walls, the amount of material traversed by particles before reaching the electromagnetic calorimeter typically ranges between 3 and 6 radiation lengths (apart from a small transition region with higher values) depending on the pseudorapidity. All layers of the EM calorimeter together correspond to more than 22 radiation lengths, which is sufficient to stop a particle like electrons and photons traversing the EM calorimeter and enough to fully contain the shower generated by the particle.

The energy resolution of the EM calorimeter is also measured. In Equation 2.5, the first term is obtained from test-beam measurements [56] with $S = 10\%$, the second one can be measured from $Z \rightarrow e^+e^-$ [57]. Its value varies in the detector between 0.1 and 0.3 %.

2.2.2.2 The hadronic calorimeter

Hadrons deposit some fraction of their energy in the EM calorimeter but they are not fully absorbed. The hadronic calorimeter is thus designed to contain the showers of high-energy hadrons. The depth of material required to contain them is expressed in terms of the nuclear interaction length λ of the passive material. Different technologies are adopted in the barrel and in the endcaps. The barrel is equipped with Tile Calorimeter [58], an iron-scintillator (Tile) sampling calorimeter. It is divided into four partitions, two long barrels (LB) (covering $0 < |\eta| < 0.8$) and two extended barrels (EB) (covering $0.8 < |\eta| < 1.7$) and is segmented into 64 modules along the azimuth ϕ . Wavelength shifting fibers collect the light generated in the plastic scintillators and carry it to photomultipliers (PMT). Two fibers, attached to every tile from different sides in ϕ , go to different PMTs, providing redundant double read-out of a signal. Each PMT receives signals from multiple tiles which are grouped into cells

of different size depending on their pseudorapidity and depth. Three longitudinal layers A, BC, D are defined inside the modules and the dimensions of the cells are optimised to obtain a structure of projective towers with granularity of $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the first two layers and $\Delta\eta \times \Delta\phi = 0.2 \times 0.1$ in the last layer. Cells of an additional special layer E, attached to the extended barrel modules, are read out by a single PMT each. In total, the tile has 5182 cells and 9852 channels. The thickness of the hadronic calorimeter is approximately 10λ .

The endcaps also use the LAr as the active material but use copper instead of lead as the absorber. The detectors are arranged in two independent wheels with the same diameter covering different η regions of LBs and EBs. Their modules are grouped with tower granularity ranging from $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ in the front wheels to $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ in the rear ones.

The energy resolution of the hadronic calorimeter is measured in test-beams [59]. In Equation 2.5, S is measured to be around 55% and C is measured to be around 3.4%.

2.2.2.3 The forward calorimeter

Located 4.7 m away from the collision point, the Forward Calorimeter (FCal) [60] is composed of an electromagnetic and hadronic part and allows measurements in a large pseudorapidity regions, $3.1 < |\eta| < 4.9$. The system must have a radiation tolerance higher than the other system due to its position. The absorber must be dense. The electromagnetic part is made of copper for the absorber, while the hadronic part uses tungsten to increase the number of total radiation lengths of the system. The FCal is approximately 10 interaction lengths deep. A compact design was implemented to avoid any energy leakage and consists of concentric rods for the electronics aligned with the beam pipe. The FCal is necessary because it allows an improvement in the estimation of p_T of neutrinos or other particles that could escape the detector, through the improvement in the measure of forward jets.

2.2.3 The muon system

The Muon Spectrometer [61] is designed to have a high quality stand-alone muon measurement with large acceptance both for muon triggering and reconstruction. The conceptual design of the Muon Spectrometer is illustrated in Figure 2.12. The spectrometer is divided into three regions: a barrel extending in the rapidity region $|\eta| \leq 1.2$ and two endcaps covering the rapidity regions ($1 < |\eta| < 2.7$). They are placed in the proximity of the toroidal magnets (see Section 2.2.4) in order to give a high momentum precision measurement. The barrel system is composed of three cylindrical layers at a radius of 5, 7.5, 10 m from the beam pipe, whilst a system of wheels is set for the end-caps at 7.4, 10.8, 14, 21.5 m from the IP along the z -axis. The detecting parts of the system are Monitored Drift Tubes (MDTs) and Cathode Strip Chambers (CSCs). In order to give a much faster signal for the trigger than what the detecting devices can produce, two sets of detectors complete the muon system: Resistive Plate Chambers (RPCs) and Thin Gap Chambers (TGCs). The total momentum resolution obtained with such system ranges from $\sigma_{p_T}/p_T = 10\%$ for 1 TeV muons to 2-3% in the low- p_T range, where the Inner Detector measurement provides

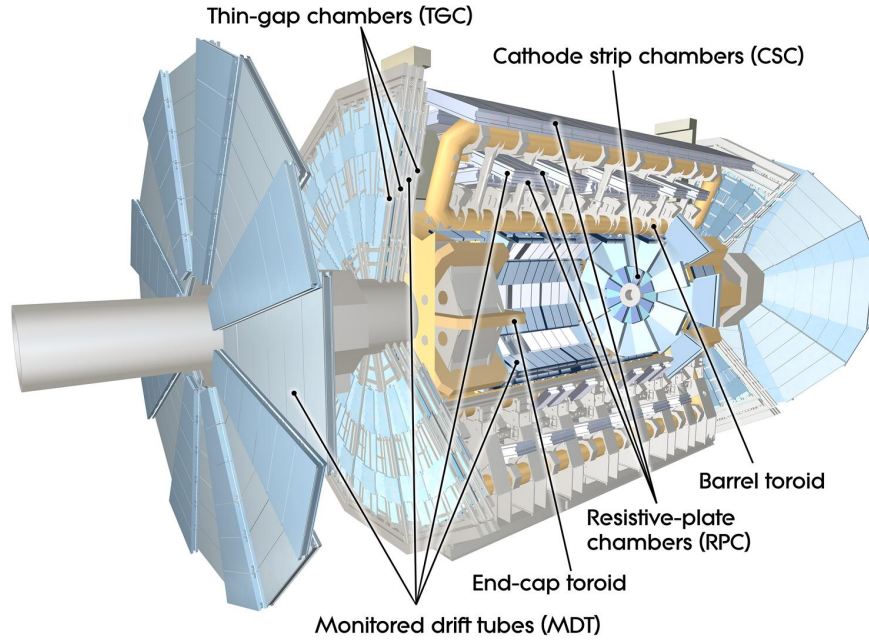


Figure 2.12: The ATLAS muon system [62].

additional information.

The Monitored Drift Tubes (MDTs) consist of two multilayers of aluminium pressurized drift tubes of about 3 cm in diameter, using an argon/ CO_2 mixture as the drift gas. A tungsten-rhenium anode wire in the tube collects the charge produced through the ionization caused by the muons traversing the gas. The measured drift time is used to determine the coordinate of the muon with the required precision. The tubes are placed into multilayers in order to improve the tracking performance. MDTs do not operate properly in very high particle density environments. For this reason, they are not placed close to the beam.

The Cathode Strip Chambers (CSCs) are multiwire proportional chambers. The wires are at a voltage of 1900 V. They lay sandwiched between strip cathodes in a gas mixture (argon/ CO_2), creating an electric field that can be used to measure the charge created by an ionizing particle. The coordinates of a traversing charged particle are obtained from the relative measurement of induced charge on adjacent cathode strips. The strips on each of the two cathode planes are positioned orthogonally, thus allowing the determination of two coordinates: in the bending direction and in ϕ .

The Resistive Plate Chambers (RPCs) consist of two rectangular detectors. Each one being a gaseous detector where the gas mixture (primarily $C_2H_2F_4$ is used) is enclosed between two resistive rectangular plates separated by 2 mm. A voltage of 9.8 kV is applied between the rectangular plates (leading to a drift time of about 5 ns), such that the charge created by the ionizing particle is lead towards the anode plane, where the signal is read out. The RPCs are built up of two detectors to provide redundancy that reduces the noise and other backgrounds, such as photons and low-energy neutrons in the cavern. RPCs are placed on the same support structure than some of the barrel MDTs.

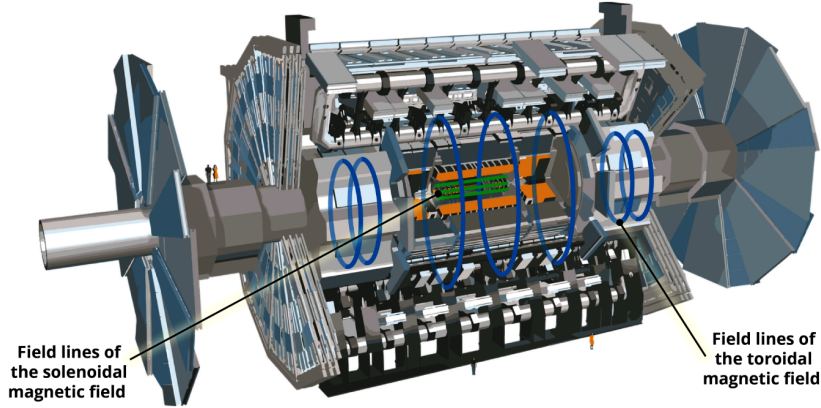


Figure 2.13: The ATLAS magnet system [66].

The Thin Gap Chambers (TGCs) uses the similar technology to the one used by the RPCs. The gas is a mixture of CO_2 and $n - C_5H_{12}$. The distance between the wires is 1.8 mm, while the distance from the wire to the cathode is 1.4 mm. The drift time combined with the signal propagation time in the electrodes guarantee that the signals arrive to the read-out system within 25 ns (the LHC bunch separation). The TGCs are built of two or three gaseous detector planes. These trigger chambers measure both coordinates of the track, one in the bending (η) plane and one in the non-bending (ϕ) plane. TGCs, unlike the RPCs, have their own support structure.

2.2.4 The magnet system

The Magnet System is designed to bend the trajectories of charged particles so that ATLAS can measure their momentum and charge. Two systems of magnets are used: a solenoid [63] provides the magnetic field for the Inner Detector, while a toroidal system, composed by a barrel [64] and two endcaps [65] magnets, provide the magnetic field for the muon spectrometer as shown in Figure 2.13. The solenoid is located in front of the calorimeters. Its inner and outer radius are respectively 1.23 and 1.28 m for a length of 5.8 m. It creates a 2 T axial field inside the inner-detector. The toroidal system covers a much larger volume, with an inner radius of 4.7 m and an outer one of 10.05 m for the barrel part, 25.3 m long, whilst the end-caps are 5 m long and extend from 0.9 to 5.3 m in radius. It provides a field ranging from 0.2 to 2.5 T in the barrel and 0.2 to 3.5 T in the end-caps.

2.2.5 The forward detectors

The Forward Detectors allow additional measurements for physics and monitoring purposes. In the forward region five extra detectors are located outside the ATLAS detector volume. Figure 2.14 shows their layouts and locations (except the Beam Condition Monitor (BCM)).

The Absolute Luminosity For ATLAS (ALFA) [67] Roman Pots (RP) are located at 240 m from the interaction point. There are two RPs on each side, the distance between these two stations is 4 m. Each station consists of 2 scintillating fiber detectors. The system is designed to approach as close as 1 mm to the beam and provides a pseudorapidity coverage $10.6 < |\eta| < 13.5$. Its goal is to measure the total pp cross section and absolute luminosity in special LHC runs at low luminosity ($10^{27} \text{ cm}^{-2} \text{ s}^{-1}$). The absolute luminosity can be determined to 2-3 % accuracy.

The Zero Degree Calorimeter (ZDC) [68] is placed at 140 m on both sides of interaction point in the TAN region (target absorber for neutrals) between the tubes at the point where the single beam pipe splits into two. It is a sampling calorimeter composed of four modules: one electromagnetic and three hadronic tungsten/quartz calorimeters. The ZDC is able to measure neutral particles at $|\eta| > 8.3$. The main purpose of the ZDC is to measure the centrality of the collisions in heavy ion runs. During proton-proton collision, ZDC is used for beam halo, beam gas suppression, luminosity monitor and also as an additional minimum bias trigger. It is also used to tag diffractive processes. When the luminosity reaches around $10^{33} \text{ cm}^{-2} \text{ s}^{-1}$, the ZDC modules are removed in order to minimize the radiation damage and will be reinstalled for heavy ion runs.

The Luminosity measurement using Cherenkov Integrating Detector (LUCID) [69] is composed of two modules located at 17 m from the interaction point on both side of ATLAS. Each module is composed of 20 Cherenkov tubes at the end of which photomultiplier tubes (PMT) are placed to collect the light. Each Cherenkov tube is made of aluminium, 15 mm in diameter, filled with C_4F_{10} . It covers a pseudorapidity range $5.6 < |\eta| < 5.9$ for charged particles with a Cherenkov threshold of 10 MeV for electrons and 2.8 GeV for charged pions. It is one of the main ATLAS online monitors for instantaneous and integrated luminosity measurement. The measurement is based on the fact that the average number of interactions in a bunch crossing is proportional to the number of particles detected in the detector. LUCID is calibrated using information from ALFA measurements that allows to reach an accuracy of 2-3 % on the absolute luminosity.

The Minimum Bias Trigger Scintillators (MBTS) [70] are segmented scintillator paddles quite close to the beam-pipe and located at 3.6 m from the interaction point. The system consists of 32 scintillator paddles, 2 cm thick, organised into 2 disks, one on each side of the interaction point of ATLAS. The system is placed between inner detector and end-cap cryostat and provides a pseudorapidity coverage $2.1 < |\eta| < 3.8$. The main purpose is to provide a trigger on minimum bias collision activity during the proton-proton collisions at low luminosities.

The Beam Condition Monitor (BCM) [71] consists of two sets of diamond sensors located 184 cm away from the interaction point in the direction of the beam and 5.5 cm away in the radial distance. The sensors are $1 \text{ cm} \times 1 \text{ cm}$ in size and $500 \text{ } \mu\text{m}$ thick. It has the role of monitoring the beam against beam losses that could cause detector damages, but it is also used as luminosity monitor. If large beam losses are detected, BCM sends a signal to the LHC, which causes an abort and a controlled dump of the accelerator ring. Its signal is also sent to the ATLAS detector in order to take the necessary actions to minimize the damage.

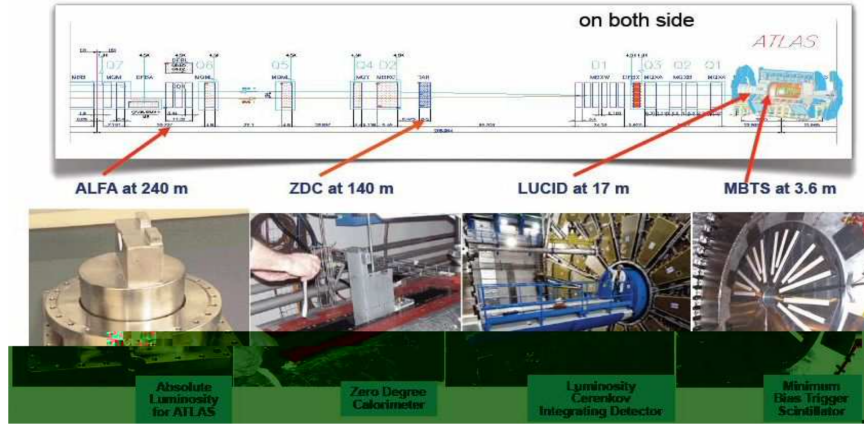


Figure 2.14: Layout of the ATLAS Forward Detectors [72].

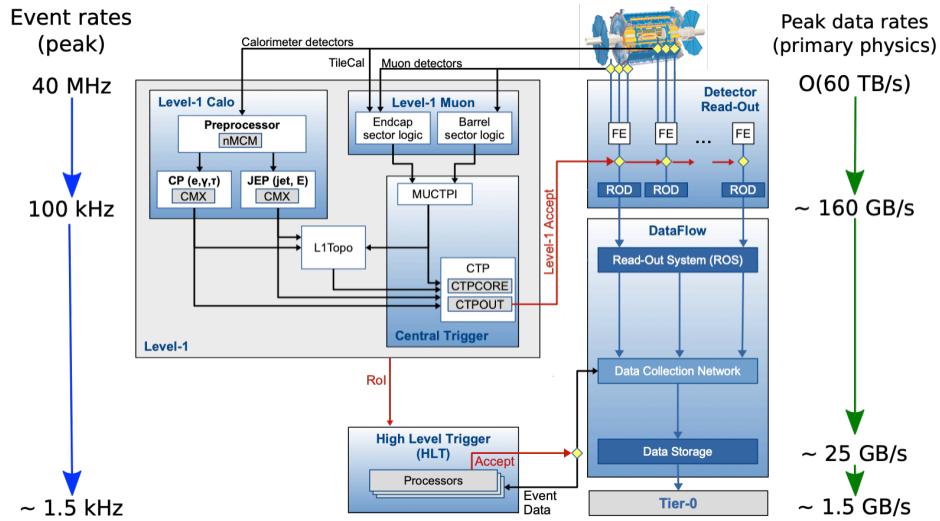


Figure 2.15: Layout of the ATLAS Trigger and Data Acquisition system in Run 2 data-taking [73].

2.3 Data acquisition

As shown in Figure 2.4, in Run 2, 34 collisions were happening for each bunch crossing on average with the individual event size of around 1.5 MB, yielding a theoretical 60 TB/s data rate. Currently, there is no hardware device that can write down or transport data with such a rate on disk or tape with current electronic technologies. The trigger system [73] is designed to reduce the rate of data recording by selecting the hard scattered events which are more interesting for further studies. In order to take into account the different latency and dead time of detectors as well as the complexity of the algorithms used for the trigger, two stages are set for the trigger as shown in Figure 2.15: The Level 1 and High Level Trigger (HLT).

2.3.1 The hardware trigger system: Level 1

The first reduction step called Level 1 [74] is operated by custom-made electronics with a guaranteed maximum latency of $2.5 \mu\text{s}$. It receives information from both the calorimeter and muon chamber Front-End (FE) electronics and uses the sum of signals from calorimeters or muon chambers to select events. It triggers the detector readout at a maximum rate of 100 kHz, imposing the necessary dead-times to protect the detector front-end buffers. It identifies a so-called region-of-interest (RoI): a part of the detector that recorded interesting signals that is used as a seed for the High-Level Trigger. This step only focuses on the high transverse-energy electrons, photons, muons, taus and jets as well as MET. To gain from the various specifics of the sub-detectors, the L1 is composed of several sub-modules.

Level 1 Calorimeter Trigger The goal of the L1 Calo [75] is to process both electromagnetic and hadronic objects coming from the two calorimeters. The analog detector signals from coarse calorimeter cells are digitised and calibrated by the preprocessor, and sent in parallel to the Cluster Processor (CP) and Jet/Energy-sum Processor (JEP). The CP system identifies electron, photon, and τ -lepton candidates above a programmable threshold, and the JEP system identifies jet candidates and produces global sums of total and missing transverse energy. The output of the L1 Calo is sent to the Level 1 Topological Trigger and the Level 1 Central Trigger Processor.

Level 1 Muon Trigger The L1 Muon trigger uses hits from the RPCs (in the barrel) and TGCs (in the endcaps) [76, 77] to find a coincidence of hits across multiple layers that are consistent with a muon originating from the interaction point. First, a hit has to be found on a pivot plane (middle station for the RPC, outer station for the TPC) as shown in Figure 2.16. Then, to reduce the rate in the endcap regions of particles not originating from the interaction point, tracks are reconstructed by applying coincidence between the outer and inner TGC stations, as well as between the TGCs and the tile calorimeter. The coincidence windows are p_T dependent. The output of L1 Muon is sent to the Level 1 Central Trigger Processor.

The Level 1 Topological Trigger The L1 Topo Trigger was designed to cope with the increase of the trigger rates in Run 2 as well as to preserve the final trigger and analysis efficiency. It uses outputs from the L1 Calo and the L1 Muon to compute angular and kinematic quantities. It is also possible to transfer and analyse the angular and energy information of each trigger object in the event in one system and compute topological angular and kinematic selections thus providing a largely improved background rejection with minimal or no signal loss.

The Level 1 Central Trigger Processor The L1 trigger decision is formed by the Central Trigger Processor (CTP) [73, 78], which receives inputs from the L1 Calo trigger, the L1 Muon trigger through the L1 Muon Central Trigger Processor Interface (MUCTPI) and the L1 topological trigger as well as trigger signals from several detector subsystems.

2.3.2 The software trigger system: High Level Trigger

After an event has been accepted by the L1 trigger, it is moved from the detector specific front-end buffers via the ReadOut Drivers (RODs) into a common readout system (ROS), as

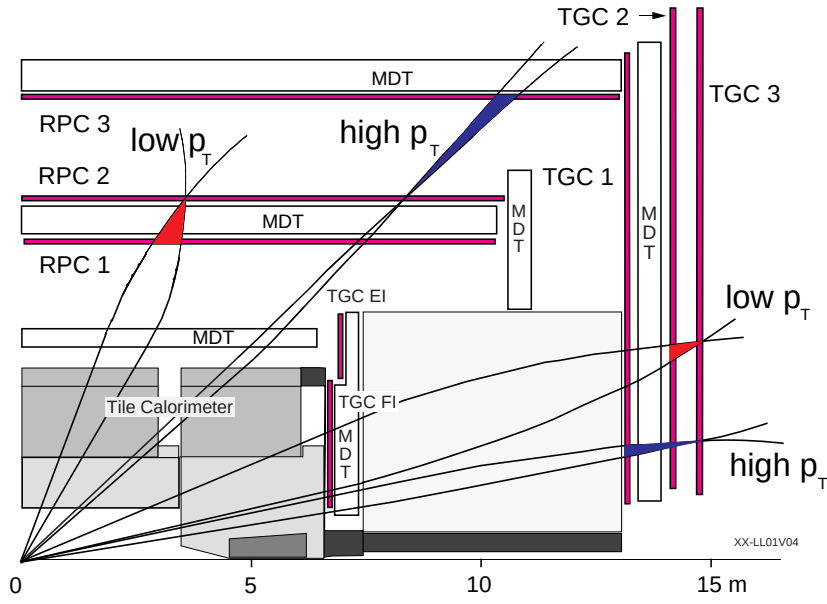


Figure 2.16: An overview of the L1 Muon trigger chambers [76].

shown by Figure 2.15. From here the High Level Trigger [79] reduces the triggered event rate from the L1 rate (100 kHz) to a final output rate of roughly 1 kHz. After HLT, the data is massively stored in the CERN computing centre and the Event Builder has access to the data via an ethernet-based network. The software-based HLT itself is composed of several subsystems. The direct interface of HLT to L1 triggers is called the L2 trigger. After the L2 trigger has generated a decision, the event is either discarded or built at the L2 accept rate. The full event data is passed to the Event Builder where it is assembled and formatted. Then the complete event becomes available to the Event Filter (EF). The EF reduces the data volume so that it can be handled offline by the mass storage operations and by the subsequent offline data reconstruction and analysis steps.

Since the number of objects passing the thresholds can be quite high, some additional random selection is set for some trigger items. This process is called **prescaling**. The prescale determines how often an event that passed a given trigger is accepted. A prescale of 1 means that all events selected by the trigger are accepted, while a prescale of 1000 means that events passing the trigger will only be accepted one out of a thousand times.

The sequence of algorithms that defines a certain trigger object at each trigger level will be referred to as a trigger chain. The final energy threshold and quality requirements are naturally determined by the last trigger level used. The naming convention for a trigger chain is:

$$[\text{LEVEL}][\text{N}][\text{TYPE(S)}][\text{THRESHOLD}][\text{ISOLATION}][\text{QUALITY}] \quad (2.6)$$

where LEVEL refers to the trigger level used, TYPE(S) specifies the object(s) candidate(s) (i.e. electron, muon, photon, jet, etc. or a combination of them), N indicates its multiplicity, THRESHOLD is a number corresponding to a transverse momentum (or energy) threshold applied, ISOLATION indicates the object isolation and QUALITY refers to the severity of requirements in the algorithm. For example, *L1J100* triggers the jets (see Section 3.3) at

the L1 level with a transverse momentum larger than 100 GeV. More complicated trigger chains can be defined as well, for instance, *HLT_j420_a10_lcw_L1J100* triggers the large-radius jets (see Section 3.3) at the HLT level with a transverse momentum larger than 420 GeV after *L1J100*. The trigger chain used in the different analyses presented in this thesis are discussed at the beginning of each analysis chapter.

2.4 Summary

The LHC and the ATLAS detector have been presented in this chapter. Each component of the ATLAS detector is designed and optimised for different purposes. The reconstruction of the physics objects makes optimal use of the different parts of the detector as will be described in Chapter 3. The ATLAS description in this chapter correspond to the detector available during the LHC Run 2 data taking period. The data recorded during Run 2 between 2015 and 2018 is used in the studies shown in Chapter 4 and 5. In 2018 the LHC went into a long shutdown to prepare for higher energy collisions, $\sqrt{s} = 13.6$ TeV, starting this year. Several upgrades were made to the ATLAS detector during this long shutdown period, e.g. improvements in the L1 trigger architectures and the installation of the New Small Wheels [80] detector. Further updates are planned to prepare for the HL-LHC phase. One of the main updates planned is the new inner detector, the Inner Tracker project (ITk), a new all-silicon tracker, which will be discussed in Chapter 7.

EVENT RECONSTRUCTION

The detector signals generated by the particles produced in the LHC collisions are processed by a set of algorithms to reconstruct, identify and calibrate the physics objects used in physics analyses. This chapter describes the reconstruction and calibration of the physics objects in ATLAS. The chapter focuses primarily on the physics objects used in this thesis, i.e. Inner Detector tracks, vertices, jets, electrons and muons. Other physics objects like photons [81], taus [82] and missing transverse energy [83], are not used in this thesis and are hence not presented in detail here.

3.1 Track reconstruction

Tracks are reconstructed in the Inner Detector using a sequence of algorithms [84, 85]. The studies of tracking performance have demonstrated good agreement between data and simulation at low pile-up environments [86, 87]. Further studies show the robustness of the algorithms with pile-up up to 31 at $\sqrt{s} = 7$ TeV [88]. However, as LHC entered into the new energy regime of $\sqrt{s} = 13$ TeV, events with TeV-scale jets showering in the detectors now occur at rates high enough to be studied in detail. The reconstructed tracks using the Inner Detector information are the crucial inputs for reconstruction of jets for boosted bosons, top quarks, b -hadrons or τ -leptons, see Section 3.3. In the cores of highly energetic hadronic jets and τ -leptons, the average separation between highly collimated charged particles is comparable to the granularity of individual sensors of the Inner Detector. This can create confusion within the algorithms used to reconstruct charged-particle trajectories. The current track reconstruction algorithms in ATLAS have been optimised to cope with dense environments from Run 2 [89]. The tracks which are particularly interesting for physics analyses are the primary tracks and the secondary tracks following different strategies. As shown in Figure 3.1, the primary tracking is done inside-out (Section 3.1.1) and the secondary tracking is done outside-in (Section 3.1.2).

Tracks are described using a perigee representation, with five parameters and a reference point, as shown by Figure 3.2. The reference point used is the average position of the pp interactions (beamspot position). The five parameters are:

- The transverse impact parameter d_0 which is the projection of the point of closest approach in the transverse direction.

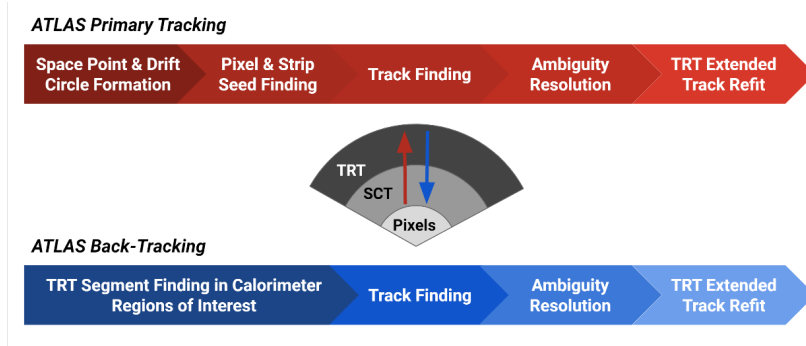


Figure 3.1: The ATLAS tracking flow chart [90].

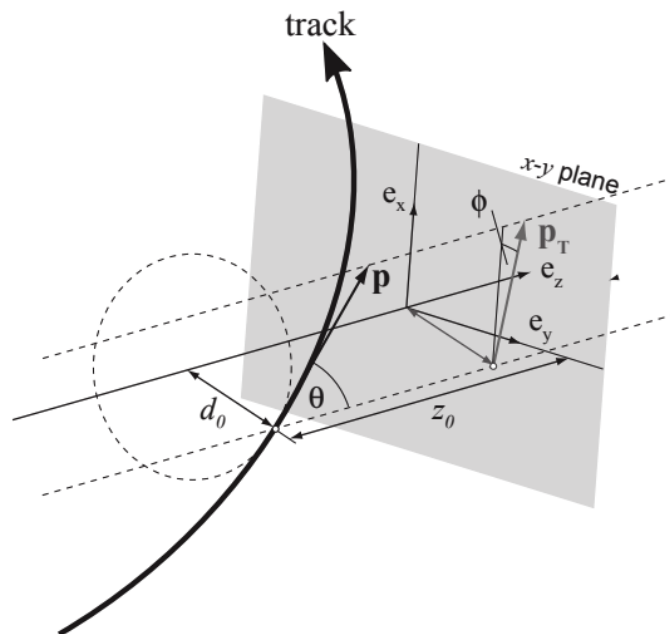


Figure 3.2: The perigee representation [90].

- The longitudinal impact parameter $z_0 \sin \theta$ which is the projection of the point of closest approach in the z-axis direction.
- The inverse transverse momentum p/p_T which gives details about the curvature of the particle track.
- The track azimuthal angle ϕ which represents the direction of the track in the $r - \phi$ plane at the point of closest approach, ϕ varies between $[-\pi, \pi]$.
- The track polar angle η , which varies between $[0, \pi]$.

d_0 , $z_0 \sin \theta$ and p/p_T are main parameters to evaluate the tracking performance.

3.1.1 Primary tracks

The primary tracks are from primary particles, which are defined as particles with a mean lifetime of greater than 3×10^{-11} s directly produced in a pp interaction or from the subsequent decays or interactions of particles with a lifetime shorter than 3×10^{-11} s. The tracks reconstructed by the inside-out algorithm are required to have transverse momentum $p_T > 400$ MeV. As shown by Figure 3.1, the main steps for primary track reconstruction are:

Step 1: Space-points determination The space-points are the final objects obtained from the raw data of the Inner Detector. In the Pixel and SCT detector, the space-points are the three-dimensional measurements of the points where a charged particle traverses the active material of the detectors. When a charged particle passes through the detectors, the deposited charge is often collected by multiple adjacent channels, referred as a "cluster". In the Pixel, each cluster determines one space-point. While in the SCT, clusters from both sides of a strip layer must be combined to obtain a space-point. The space-points are determined from the channels contributing to the clusters using a linear approximation refined with a charge interpolation technique [91]. In the TRT, the space-points are two-dimensional since the measurement is the drift radius. Once the direction of the particle is known, the drift radius translates into a closest distance to the wire with an ambiguity (which side of the wire the particle passed-by).

Step 2: Track seeding and finding Track seeds are formed from triplets of space-points in either the Pixel or the SCT. Then a combinatorial Kalman filter [92] is used to build track candidates from the chosen seeds. An alignment happens at the beginning of each run in order to solve the misalignment among planes due to mechanical movements and thermal variations. A global fitting method is used to minimize the track-hit residuals in each detector plane [93, 94]. Track-hit residuals are the differences between the hit in the detector and the intersection of the detector with the track extrapolation. The fit is done for all the tracks at the same time.

Step 3: Ambiguity solving This step is performed to resolve overlaps between track candidates and reject fake tracks which combine unrelated clusters. First the track scores are calculated by an ambiguity solver using a robust approach [89]. The tracks with a bad score are rejected. In addition, track candidates are rejected by the ambiguity solver if they fail to meet any of the following basic quality criteria (*Loose selection*):

- $p_T > 400$ MeV

- $|\eta| < 2.5$
- Minimum of 7 pixel and SCT clusters (12 are expected.)
- Maximum of either one shared pixel cluster or two shared SCT clusters on the same layer
- No more than two holes¹ in the combined pixel and SCT detectors
- No more than one hole in the pixel detector
- $|d_0| < 2.0$ mm with respect to the beam line
- $|z \sin \theta| < 3.0$ mm with respect to the beam line

In dense environments where boosted objects are highly collimated, and some particles can be detected simultaneously by adjacent pixels, more stringent selections in addition to the *loose* selection requirements are usually required (*Tight Primary* selection) [95]:

- Minimum of 9 pixel and SCT clusters (if $|\eta| \leq 1.65$)
- Minimum of 11 pixel and SCT clusters (if $|\eta| \geq 1.65$)
- At least one hit on one of the two innermost pixel layers
- No pixel holes

We should note that the selections have evolved with time. The defined *loose* and *tight primary* selections are used for analysis with full Run 2 data.

The refined and purified track candidates resulting from the ambiguity resolution are then refitted using a global χ^2 method to obtain the final, high-precision track parameter estimate.

Step 4: TRT extension TRT extension is applied to profit of additional measurements on the track, in particular for momentum resolution and particle identification. It is similar to what is done for track candidates built in previous steps but starts from the position estimation of the track candidate in the TRT volume. A Kalman filter is used to build candidate extensions. Then the TRT hits are added to the tracks and the whole tracks are refitted again with the global χ^2 fitter. Si-only tracks are kept if the fit worsens with the addition of the TRT hits.

3.1.2 Secondary tracks

The secondary tracks are mainly generated from particle decays, photon conversions or hadronic interactions, referred as secondary particles. The secondary tracks are reconstructed using a back-tracking pass. Only the detector hits not already assigned to tracks from the primary reconstruction are used. The track reconstruction is only attempted in RoI determined by deposits in the EM calorimeter. It starts with segments of hits in the TRT which are compatible with the RoI. In these TRT segments, the seeds consists of two space-points which are constructed in the close part of the Pixel and the SCT to the TRT.

¹Holes are defined as intersections of the reconstructed track trajectory with a sensitive detector element that does not contain a matching cluster.

These seeds are then extended into track candidates using the same procedure as for the primary track reconstruction.

The b -tagging described in Section 3.4 is based on the reconstructed primary and secondary tracks. For this reason it is thus important to understand the track reconstruction performance. The evaluated *tight primary* track reconstruction efficiency is overall lower (up to $\sim 10\%$ for $1.5 \leq |\eta| \leq 2.5$, $\sim 5\%$ for $|\eta| < 1.5$ and $\sim 5\%$ in the full p_T range) than the *loose* track reconstruction efficiency because of the more stringent requirements, but the *tight primary* selection improves fake rejection in dense environment [89, 95, 96].

3.2 Vertex reconstruction

The input of the vertex reconstruction is a collection of tracks reconstructed as described in Section 3.1. Tracks can be originated from a primary vertex (PV), the interaction point between two partons, or from a secondary vertex (SV) from particles decays, photon conversions or hadronic interactions. In one event there is one hard-scatter PV, while the others are associated, by convention, to the pile-up interactions. This section focuses on the PVs. More details can be found in [97, 98]. Secondary vertices (SV) correspond to the decays of short-lived particles, which decays at a measurable distance from the PVs. They are the signature of heavy flavour particles, especially b -hadrons which decays at a hundred of micrometers ($c\tau_b \simeq 0.5\text{mm}$), taus, or other long-living particles.

In general, the PV reconstruction is done in three steps. A seed position is first selected based on the beam spot in the transverse plane. Then, the selected tracks and the seed are fitted to have a best estimated vertex position. After the vertex position is determined, tracks that are incompatible with the vertex are discarded. The discarded tracks are used to seed a new PV. This procedure is repeated until no unassociated tracks are left or no additional vertex can be found. PVs are required to have at least two associated tracks. The same track can be associated to multiple vertices. The PV with the largest sum of squared tracks' transverse momenta is chosen as the hard-scatter PV. In high dense environment, the increasing number of fake tracks increases the probability to reconstruct a fake vertex. In order to maximise the number of available tracks used for the vertex reconstruction, the *loose* selection requirements (see Section 3.1.1) are applied to the tracks [99].

3.3 Jets

Jets are an useful way to represent collimated streams of particles from the hadronisation of quarks and gluons. They are reconstructed using tracks or deposited energy of the particles in calorimeters or a combination of both. The inputs and algorithms used for jet reconstruction are presented in Section 3.3.1 and Section 3.3.2, respectively, followed by the jet calibration procedure in Section 3.3.4.

3.3.1 Jet inputs

ID tracks The ID tracks are presented in detail in Section 3.1. The *loose* selection requirement is used. In addition, the tracks are required to have $p_T > 500\text{ MeV}$, $|\eta| < 2.5$ and

$z_0 \sin \theta < 2.0 \text{ mm}$ with respect to the PV which has the highest scalar p_T^2 sum of tracks-associated.

Topological clusters (Topo-clusters) The topo-clusters are three-dimensional groupings of noise-suppressed clusters of calorimeter cells [100]. First the cells with a signal significance over $4\sigma_{noise}$ in an event are selected as seeds. σ_{noise} is the standard deviation of the cell-signal fluctuations introduced by electronics and pile-up events. Then, their neighbouring cells with significance over $2\sigma_{noise}$ are subsequently added until all adjacent cells have a significance below $2\sigma_{noise}$. Afterwards, all neighbouring cells are added to the cluster as well. If in a cluster, there are two or more local energy maxima, a splitting algorithm is used to separate the showers. All clusters are taken to be massless. The topo-clusters can be calibrated either at the raw (electromagnetic, EM) scale or at the local cell weighting (LCW) scale. In EM scale, the energy of an isolated topo-cluster is the sum of its constituent cells energy. In LCW scale, the difference between electromagnetic and hadronic interactions is taken into account. After calibration, the angular coordinates (η and ϕ) of topo-clusters are recalculated relative to the PV of the event.

Particle-flow objects (PFOs) The particle-flow (PFlow) reconstruction [101] uses the *tight primary* tracks (Section 3.1.1) and the EM topo-clusters as inputs. First, the PFlow algorithm attempts to match each selected track to a single topo-cluster in the calorimeter. Then, expected energy deposited in the calorimeter by the particle that produced the track is subtracted, cell-by-cell, from the associated topo-clusters. If the associated calorimeter energy following this subtraction is consistent with the expected shower fluctuations of a single particle, the remaining calorimeter energy is removed. The topo-clusters which are not matched to any tracks are assumed to contain energy deposited by neutral particles and are left unmodified. However, the subtraction is gradually disabled when the p_T of the tracks increases to account for the degradation of the inner tracker performance at high p_T . The subtraction is completely disabled for the tracks with $p_T > 100 \text{ GeV}$. Therefore, the jets reconstructed with PFOs have better performance at low p_T compared with the baseline topo-cluster-based jet definition. After the application of the PFlow algorithm, the collection of PFOs consists of the charged PFOs and the neutral PFOs. The charged PFOs are classified into different categories: the charged PFOs which were used to subtract energy from a topo-cluster, and the ones for which no subtraction was performed due to their high momentum or being located in a dense environment. The neutral PFOs are the unmodified topo-clusters after the PFlow algorithm.

Track-Calo Clusters (TCCs) The inputs of the TCCs [102] are the tracks and the topo-clusters as well but are used in a different way from the PFlow algorithm. The TCC algorithm uses the *loose* tracks (Section 3.1.1) and the LCW topo-clusters. First, it attempts to match each selected track to a single topo-cluster in the calorimeter. If one track matches one topo-cluster, the p_T of the TCC object is taken from the topo-cluster, while its η and ϕ coordinates are taken from the track because the calorimeter has better energy resolution but worse spatial resolution than the ID. If multiple tracks are matched to multiple topo-clusters, several TCC objects are created (where the TCC multiplicity is equal to the track multiplicity): each TCC object is given some fraction of the momentum of the topo-cluster, where that fraction is determined from the ratios of momenta of the matched tracks. TCC angular properties (η , ϕ) are taken directly from the unmodified ID tracks, and their mass is set to zero. Contrary to the PFOs, the jets reconstructed with the TCCs have improved

performance at high p_T , however their performance is worse than the topo-clusters-based jets at low p_T . After the TCC algorithm, the unmatched topo-clusters are included as unmodified neutral TCCs and the matched topo-clusters are included as the modified TCCs.

Unified flow objects (UFOs) The unified flow objects (UFOs) [103] combine desirable aspects of PFOs and TCCs in order to achieve optimal overall performance across the full kinematic range. The process begins by applying the standard ATLAS PFlow algorithm. The charged PFOs which are matched to pile-up vertices are removed. Then, a modified TCC algorithm is applied to the remaining PFOs and the *tight primary* tracks. Any tracks which have been used for the PFlow subtraction are not considered to avoid overlap. The UFOs are outputs from the modified TCC algorithm. This approach provides the maximum benefit of the PFlow subtraction at lower p_T , and cluster splitting where the benefit is maximal at high p_T .

Generator-level particles The generator-level particles are used to reconstruct particle-level jets, or ‘truth jets’. All detector-stable particles from the hard-scattering process with a lifetime, τ , in the laboratory frame such that $c\tau > 10 \text{ mm}$ are used. Particles that are expected to leave only negligible energy depositions in the calorimeters, i.e. muons and neutrinos, are excluded. Particle-level jets are used as the reference objects for simulation-based ATLAS jet calibrations, and for studies of the jet energy and mass resolution. They are also used as reference in physics studies.

3.3.2 Jet algorithms

The software package used for jet reconstruction in ATLAS is FASTJET [104]. The most used algorithm in ATLAS is the anti- k_t algorithm [105]. In this algorithm, two distances d_{ij} are introduced: d_{ij} the distance between entities (jet input objects) i and j , and d_{iB} the distance between entity i and the beam (B). If $d_{ij} < d_{iB}$, the entity j is combined to the entity i . Otherwise, the entity i is called a jet and removed from the list of entities. The distances are recalculated within the remaining entities and the procedure is repeated until no entities are left. The definition of the two distances are:

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad (3.1)$$

$$d_{iB} = p_{Ti}^{2p} \quad (3.2)$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ and p_{Ti} , y_i and ϕ_i are the transverse momentum, rapidity and azimuth of the entity i , respectively. The parameter (p) is introduced in addition to the usual radius parameter (R) to govern the relative power of the energy versus geometrical (Δ_{ij}) scales. If p is set to -1, it is called **anti- k_t algorithm**; if p is set to 1, it is called **k_t algorithm**. Compared to other jet algorithms as discussed in [105], the anti- k_t algorithm is soft-resilient and gives more circular jets of radius (R). Different distance parameters R are typically set in ATLAS for different purposes.

Small- R jets The small- R jets are jets reconstructed with $R = 0.4$. Jets reconstructed using only the origin-corrected EM scale topo-clusters are referred as **the EMTopo jets**.

Jets using PFOs are referred as **the PFlow jets**. These jets are not used in the analyses in this thesis but are mentioned for the sake of comparison.

Large- R jets The large- R jets are jets reconstructed with $R = 1.0$. Similar to the small- R jets, there are different types of large- R jets. Jets reconstructed using the LCW scale topoclusters are **the LCTopo jets**. Accordingly, there are **the TCC jets** and **the UFO jets**. Furthermore, the large- R jets are used to reconstruct 2-body and 3-body decays at high p_T , like the hadronically decaying Higgs boson, W/Z , the top quark. In this thesis, the LCTopo jets are used for $H \rightarrow b\bar{b}$ tagging and are adopted in physics analyses. The TCC jets are developed for the vector boson tagging and largely used by diboson related analyses. The UFO jets are stepping into the stage in ATLAS and we are looking forward to the results using the UFO jets.

Variable- R (VR) track jets The VR track jets are reconstructed using Inner Detector tracks as inputs but with R which is as function of the jet p_T as:

$$R \rightarrow R_{eff}(p_T) = \frac{\rho}{p_T} \quad (3.3)$$

The parameter ρ determines how fast the effective jet size decreases with the transverse momentum of the jet. In addition to ρ , the VR algorithm requires two additional parameters, R_{min} and R_{max} , to impose lower and upper cut-offs on the jet size, respectively. The additional parameters prevent the jets from becoming too large at low p_T and from shrinking below the detector resolution at high p_T . The effective jet size varies smoothly between R_{min} and R_{max} . The VR track jets are mostly useful for $H \rightarrow b\bar{b}$ tagging at high energies where (see Chapter 4) the fixed radius jets overlap and the tagging efficiency is highly degraded [106]. The ρ , R_{min} and R_{max} of the VR track jets are hence scanned to have the best performance in $H \rightarrow b\bar{b}$ tagging. The optimal values for VR track jets are: $\rho = 30$ GeV and $R_{max} = 0.4$, $R_{min} = 0.02$ [106].

3.3.3 Jet grooming

The large- R jets are expected to fully contain the hadronic massive particle decays depending of the p_T . However, since the large fraction of the calorimeter volume is used during the large- R jets reconstruction, it is unavoidable to have more pile-up constituents enclosed within the jet volume. Since the pile-up constituents are low-energy and randomly distributed, they can obscure the angular structure within the jet that is the key to identifying massive particle decays. To get over these limitations, the large- R jets are typically groomed. Jet grooming is a class of algorithms to remove the pile-up constituents following a defined strategy, rebuilding the final jet from the remaining constituents. For the LCTopo jets used in this thesis, the trimming procedure [107] is adopted: trimming a jet consists of taking the original anti- k_t jet, reclustering the constituents in it using the k_t algorithm with a distance parameter of R_{sub} , and throwing away any constituents in a given k_t subjet if the subjet p_T is less than f_{cut} of the large- R jet p_T . The optimal parameters for the LCTopo jets are $f_{cut} = 5\%$ and $R_{sub} = 0.2$. This approach yields a stable jet mass distribution from the 2012 pile-up conditions up to an average of 200 collisions per bunch crossing. For the UFO jets, a different technique called Soft-drop (SD) is found to be optimal [103].

3.3.4 Jet calibration

Jets have to be calibrated to correct the reconstructed jet energy (and mass for the large- R jets) to the particle level energy (mass). The calibration procedure is shown in Figure 3.3. It can be split into three primary stages: first the pile-up contributions are suppressed at the jet level, then the jet is calibrated to the Monte Carlo (MC) truth scale, and finally the differences between MC and data are accounted for. Only the calibration of the LCTopo jets is presented in this section since it is used in the physics analyses in this thesis.

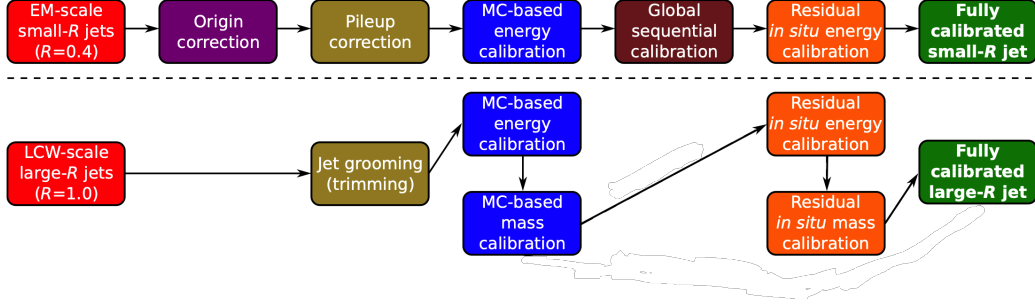


Figure 3.3: An overview of the calibration chain for both the small- R and the large- R jets is shown. [108]

MC-based jet energy calibration The MC-based jet energy calibration is the first step of calibration after jet grooming. It is essential because a large fraction of the energy in a hadronic shower is invisible to the detector, for example the hadronic shower energy ending up in dead volumes of the detector. This invisible energy must be recovered to accurately measure the Jet Energy Scale (JES). To do it, the reconstructed jets are compared with the MC truth jets described in Section 3.3.1. The jet energy response $\mathcal{R}$ is defined as:

$$\mathcal{R} = \langle p_T^{\text{reco}} / p_T^{\text{truth}} \rangle \quad (3.4)$$

$p_T^{\text{reco}} / p_T^{\text{truth}}$ distribution is calculated in fine bins of p_T^{truth} as well as η . In each bin, the response distribution is fitted with a gaussian function. The jet energy scale is defined as the mean of the gaussian fit. The standard deviation of the gaussian fit ($\sigma_{\mathcal{R}}$) is used to define jet energy resolution (JER), $\sigma_{\mathcal{R}} / \mathcal{R}$. The calibration corrects the reconstructed jet energy scale by applying $1/\mathcal{R}$ in each bin.

MC-based jet mass calibration The jet mass is also an important variable for the large- R jets, as they are used to represent the hadronic decay of a massive particle, and it is desirable for the large- R jet to have a mass corresponding to the parent massive particle to aid in the identification and interpretation of hadronic events. After the jet energy calibration, the same procedure of matching truth to reconstructed jets is followed as in the jet energy calibration. It is referred as the Jet Mass Scale (JMS) calibration. The jet mass response is defined as:

$$\mathcal{R}_m = \langle m^{\text{reco}} / m^{\text{truth}} \rangle \quad (3.5)$$

$m^{\text{reco}} / m^{\text{truth}}$ is parameterised in the p_T^{truth} , m^{truth} , and η . As for the jet energy calibration, the response distribution is fitted by a gaussian function. The jet mass response $\mathcal{R}_m$

is the mean of the gaussian fit. The standard deviation of the gaussian fit ($\sigma_{\mathcal{R}_m}$) is used to define jet mass resolution (JMR), $\sigma_{\mathcal{R}_m}/\mathcal{R}_m$. The calibration is correcting the reconstructed jet mass scale by applying $1/\mathcal{R}_m$ in each bin.

In-situ calibration In-situ calibration is used to understand how well the JES and JMS calibrations are modelled in data. Any differences between data and MC simulation can lead to bias in the jet energy (or p_T) and mass and thus must be corrected. The in-situ calibration is determined in two separate steps. Detailed techniques and results for the large- R jets are shown in [109]. A variety of final states are used to cover the entire detector. For the jet energy response, the correction factor is a product of two terms. The absolute calibration is derived from a statistical combination of three measurements from Z + jets, γ +jet and multijet events in the central region of the detector. A relative intercalibration, derived using dijet events, propagates the well-measured central JES into the forward region of the detector. The correction is applied as a four-momentum scale factor to jets in data. The jet mass response is measured using two methods: the R_{trk} method and the forward folding method. The mass response is measured in lepton+jets top quark pair production ($t\bar{t}$ production) with a fit to the peaks in the jet mass distribution of high- p_T W . A full set of in-situ uncertainties is thus derived for the JES, JER, JMS and JMR. The performance of these uncertainties depends on the truth information of the large- R jets, see Section 3.3.5.

3.3.5 Jet truth labelling

The small- R jet flavour labelling in simulation is done by matching geometrically the jet with truth hadrons. If a weakly decaying b -hadron with p_T above 5 GeV is found within $\Delta R = 0.2$ of the track-jet axis, the track-jet is labelled as a b -jet. In the case that the b -hadron could match more than one track-jets, only the closest track-jet is labelled. If no b -hadron is found, the procedure is repeated for weakly decaying c -hadrons to label c -jets. If no c -hadron is found, the procedure is repeated for τ -leptons to label τ -jets. A jet for which no such matching can be made is labelled as a light-flavour jet.

The large- R jet truth labelling defines the acceptance of the analysis. The truth matching of the large- R jets in ATLAS is shown in [110]. For the LCTopo jets, the truth matching is done in two steps:

- First match the truth particle to truth jets such that the angular distance between the truth particle and the truth jets is smaller than 0.75, and tag the jet according to the truth particle. If multiple truth jets match the truth particle, the closest truth jet to the particle gets the label of the truth particle.
- Then match the reconstructed jet to the truth jet such that the angular distance between the reconstructed jet and the truth jet is smaller than 0.75, and assign correspondingly the tag to the reconstructed jet. If multiple truth jets match the reconstructed jet, the closest truth jet to the reconstructed jet gives the label.

Table 3.1 summarises the truth labels used in this thesis and its requirements.

Table 3.1: Truth label of the large- R jets.

Num.	Label	Requirements
1	<i>contained tqqb</i>	There are a matched top quark and W boson; there is at least one B -hadron; the truth jet mass is larger than 140 GeV.
2	<i>contained Wqq</i>	There is at least one c or u -quark in the first 2 leading partons; and their invariant mass is in 60-140 GeV range; the truth jet mass is in 50-100 GeV range.
3	<i>contained Zqq</i>	The first 2 leading partons are with opposite charge; and their invariant mass is in 60-140 GeV range; the truth jet mass is in 60-110 GeV range.
4	<i>Wqq from top</i>	There are a matched top quark and W -boson; there is no b -hadron; the truth jet mass is in 50-100 GeV range.
5	<i>Other from top</i>	There is only a matched top quark.
6	<i>Other from V</i>	There is at least one c or u -quark in the first 2 leading partons OR the first 2 leading partons have opposite charge; their invariant mass is required in 60-140 GeV range.
7	<i>no truth</i>	There is no matched truth jets.
8	<i>QCD</i>	The truth jets are not matched to any heavy particles.
0	<i>default</i>	Without any label.

3.4 b -tagging

The b -jets are jets originating from b -hadron fragmentation. They are essential objects for many physics analyses, such as the property measurements of the top quark, searches via $H \rightarrow b\bar{b}$ and searches for new physics phenomena using final states where the b -quarks are present. Figure 3.4 shows different types of jets: b -jet (left), c -jet (middle) and l -jet (right). c -jet is a jet from a charmed hadron, typically a D -meson and l -jet is defined as a jet corresponding to neither b -, c -, nor τ -jets. The b -hadrons have longer lifetime and larger mass than charmed mesons and light quarks. The lifetime of b -hadrons is in the order of 1.5 ps ($c\tau \approx 450\mu m$). A b -hadron with $p_T = 50$ GeV has a path of 3 mm in the transverse direction before decaying. Due to the large mass of b -hadrons, a b -jet has measurable Secondary Vertex (SV) and high decay product multiplicity. These features make a b -jet distinguishable from c -jets and l -jets.

b -tagging is the identification of jets containing b -hadrons. The key objects for b -tagging are the hadronic jets, the tracks associated with the jets and the primary vertices. The hadronic jets are small- R jets or variable- R (VR) track jets. Tracks are reconstructed as described in Section 3.1.1. They are associated to the jets based on their angular distribution $\Delta R(track, jet)^2$. The ΔR cut varies as a function of the jet p_T , resulting in a narrower cone for jets at high p_T , which are more collimated. At 20 GeV, it is 0.45, while for more energetic jets with a p_T of 150 GeV, the ΔR cut is 0.26. Any given track is associated with at most one jet; if it satisfies the association criteria to more than one jet, the jet with the smallest ΔR is chosen.

² $\Delta R(track, jet) = \sqrt{(\Delta\eta)^2 + (\Delta\Phi)^2}$

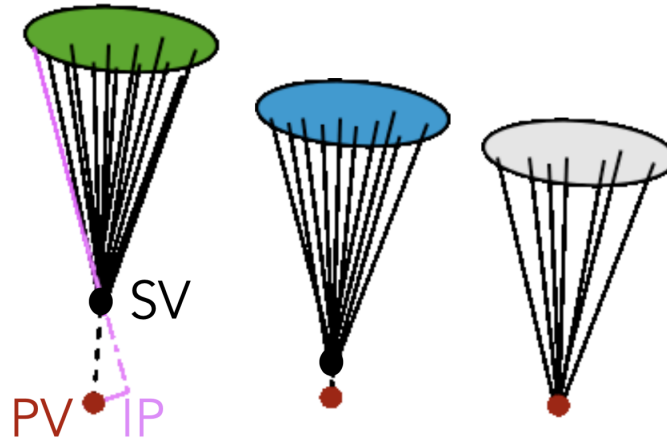


Figure 3.4: Graphical representation of b -jet (left), c -jet (middle) and l -jet (right). **PV**: Primary Vertex. **IP**: Impact parameter. **SV**: Secondary Vertex.

3.4.1 Low-level b -tagging algorithms

The jets and the tracks introduced in previous section are first processed in the low-level algorithms: IP2D, IP3D, RNNIP, SV1 and JetFitter.

IP2D and IP3D [111, 112] are impact parameter-based algorithms. IP2D uses the signed transverse impact parameter significance of tracks to construct a discriminating variable, whereas IP3D uses both the signed transverse and the longitudinal impact parameter significance in a two-dimensional template to account for their correlation. In both methods, the probability density functions are obtained from reference histograms by MC simulation. Then, they are used to calculate the ratio of the b - and l -jet probabilities for each track. Afterwards, a log-likelihood ratio (LLR) discriminant is computed as the sum of the per-track contributions, $\sum_{i=1}^N \log(\frac{p_b}{p_u})$, where N is the number of tracks of a given jet and p_b, p_u are the probability for the b - and l -jet flavour hypotheses, respectively. We assume no correlation among the various tracks contributing to the sum in the calculation. In addition to the LLR separating the b - and l -jets, LLR functions are also computed to separate the b -jets from the c -jets, or the c -jets from the l -jets.

RNNIP [113] is an impact parameter-based recurrent neural network algorithm. It uses the same inputs as IP3D, but it takes advantage of the correlations among the various tracks. The network provides the b -jet, c -jet, l -jet and τ -jet probabilities as output.

SV1 [114] is a secondary vertex finding algorithm which reconstructs a single displaced secondary vertex in a jet. The reconstruction starts from the possible two-track vertices and runs iteratively on all tracks contributing to the cleaned two-track vertices. In each iteration, the track-to-vertex association is evaluated using a χ^2 test. The track with the largest χ^2 is removed. The vertex fit is repeated until an acceptable vertex χ^2 and a vertex invariant mass less than 6 GeV are obtained. A single secondary vertex is reconstructed as output.

JetFitter [115] is a topological multi-vertex algorithm which exploits the topological structure of weak b - and c -hadron decays inside the jet. It tries to reconstruct the complete

b -hadron decay chain. It assumes that the B - and D -meson share the same flight directions. It starts from the possible two-track vertices as in SV1. Then a dedicated Kalman Filter is used to assign the tracks to vertices. Finally, the vertex fitting procedure yields the optimal values for the b -hadron flight axis direction and the position of the fitted vertices.

3.4.2 High-level b -tagging algorithms

High-level algorithms produce more performant discriminants using the outputs from the low-level algorithms. The standard high-level algorithms in ATLAS are based on machine learning techniques: MV2 and DL1(r).

The MV2 algorithm [113] is a boosted decision tree (BDT) algorithm. Its inputs are the kinematic properties of the jets and the outputs of IP2D/IP3D, SV1 and JetFitter. The list of inputs is described in Table 1 of Ref. [112]. The algorithm is trained on the SM $t\bar{t}$ and BSM $Z' \rightarrow t\bar{t}$ sample. To avoid differences in the kinematic distributions of signal (b -jets) and background (c -jets and l -jets), the b -jets and c -jets are reweighted in p_T and $|\eta|$ to match the spectrum of l -jets. MV2c10 is the most used algorithm in ATLAS. In MV2c10, the c -jet fraction in the background sample is set to 7% to enhance the charm rejection whilst preserving a high light-flavour jet rejection.

The DL1 algorithm [113] is a deep feed-forward neural network algorithm. Its inputs are the same as those used in the MV2 algorithm. The output of DL1 NN is the probabilities for a jet to be a b -jet, a c -jet or a l -jet. The DL1r algorithm has the same architecture as the DL1 algorithm but it uses additional inputs from RNNIP.

The output of MV2c10 and DL1, based on the simulated events, are shown in Figure 3.5.

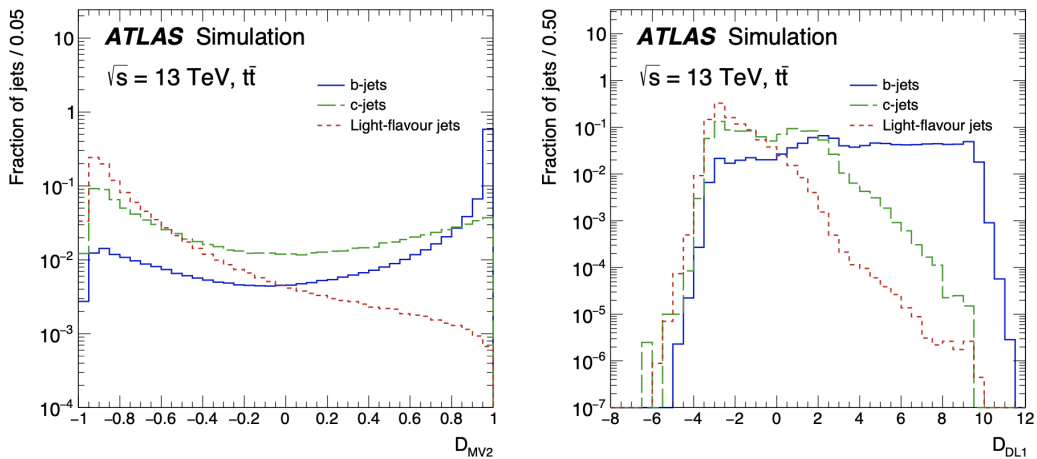


Figure 3.5: Distribution of the output discriminant of MV2 (left) and DL1 (right) for b -jets, c -jets and l -jets in the baseline $t\bar{t}$ simulated events. [112]

3.4.3 b -tagging performance

The output of the MV2 algorithm is a single probability of a jet being classified as a b -jet. DL1 and DL1r provide three outputs, p_b , p_c and p_u , corresponding to the probability of a jet to be classified as a b -jet, c -jet or a l -jet. The DL1 and DL1r final discriminants are obtained by calculating the b -jet log-likelihood:

$$D_{DL1(r)} = \ln \frac{p_b}{f_c p_c + (1 - f_c) p_u} \quad (3.6)$$

Where f_c is the weight of the c -jet rejection in the variable and is optimised separately for the different DL1(r) variants and optimization campaigns. In the final recommendation for analyses using full Run 2 data, f_c is optimised to be 0.018.

The fraction of b -jets can be improved by applying cuts on the discriminant from high-level algorithms. The b -tagging efficiency (ϵ_b) is defined as the fraction of the b -jets passing the chosen cuts over the total number of b -jets. The misidentification efficiency of a c -jet and l -jet is referred to as mistag rate (ϵ_c and ϵ_u). It is defined as the fraction of the c -jets or l -jets passing the chosen cuts for the b -tagging discriminant over the total number of c -jets or l -jets. Figure 3.6 shows the c -jets and l -jets rejection ($1/\epsilon_c$, $1/\epsilon_u$) as function of the b -tagging efficiency. The performance of MV2 is similar to DL1 but worse than DL1r. The added RNNIP outputs significantly improve c -jets and l -jets rejection.

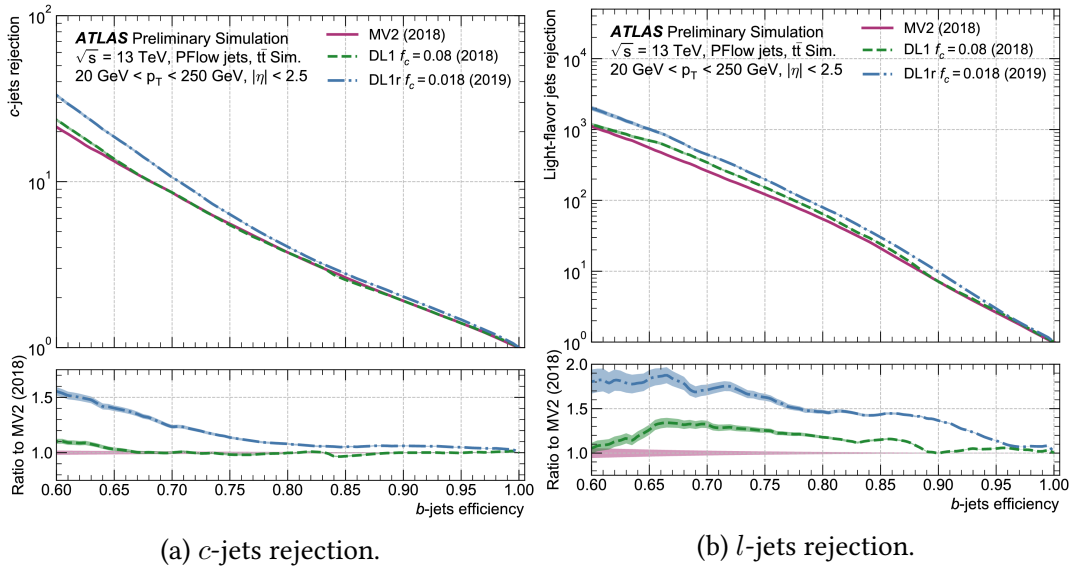


Figure 3.6: The c -jets (left) and l -jets (right) rejection as function of the b -tagging efficiency [116].

3.4.4 b -tagging calibration

The b -jets tagging working points (WP) used in physics analysis are defined by a fixed selection requirement on the b -tagging algorithm discriminant distribution, ensuring a specific b -tagging efficiency (ϵ_b). The tagging algorithms are developed using specific (nominal)

MC simulations. The nominal MC samples may have a different performance from MC samples used in physics analyses, and they can not fully describe data. Therefore, the MC-to-MC scale factors ($\epsilon_{alt. MC}/\epsilon_{nom. MC}$) and the data-to-MC scale factors ($\epsilon_{data}/\epsilon_{nom. MC}$) are needed.

The MC-to-MC scale factors are measured with several MC event generators and parton shower models with respect to the nominal generator used for the b -tagging efficiency measurement. Details and results are reported in [117]. The data-to-MC scale factors are measured using well defined data sets. The b -jet tagging efficiency is calibrated using the fully leptonic decaying $t\bar{t}$ events (2L $t\bar{t}$), where the two W -bosons decay leptonically. The tag-and-probe method [118] is used. The c -jet mistag rate is calibrated using single leptonic decaying $t\bar{t}$ events (1L $t\bar{t}$), where one of the W -bosons decays leptonically and the other decays to a c - and s -quark or other quark pair combinations. A likelihood fit is used to extract the c -jet mistag rate from the pair of jets associated with W -boson decays [119]. The l -jet mistag rate is calibrated using Z + jets events with enriched l -jets [120].

3.5 Electrons

From detector signals to the physical object used in the analysis, several steps are necessary in the selection of electrons in ATLAS: reconstruction, identification, isolation and electron-charge identification.

3.5.1 Electron reconstruction

When an electron passes through the detectors, its trajectory is illustrated as in Figure 3.7. The electron first traverses the tracking system (Pixel detector, then the SCT and finally the TRT) and then enters the electromagnetic calorimeter. Hence, the electron candidates are reconstructed using information from electromagnetic calorimeter and the Inner Detector [121]. The reconstruction starts with a sliding window algorithm which aims to find electromagnetic clusters with a threshold $E_T > 2.5$ GeV, referred as seed-cluster. When a seed-cluster is found, the algorithm looks for a reconstructed ID track to match. The track reconstruction is detailed in Section 3.1. The close matching in $\eta \times \phi$ space of the tracks to the clusters is considered as an electron candidate and calibrated as such. Otherwise it is calibrated as a photon.

3.5.2 Electron identification

After reconstruction, prompt electrons entering the central region of the detector ($|\eta| < 2.47$) are selected using a likelihood-based (LH) identification. The inputs to the LH consist of measurements from the tracking system, the calorimeter system, and quantities that combine both the tracking and the calorimeter information. The probability density functions used in the LH model are based on the simulated events. The calculated LH discriminant is used to define working points (WPs). To cover the various required prompt-electron signal efficiencies and corresponding background rejection factors needed by the physics analyses, four different WPs are defined with four fixed values (from low to high) of the LH

discriminant: VeryLoose, Loose, Medium, and Tight. More details can be found in [121]. The detailed criteria used in this thesis will be presented in the corresponding sections.

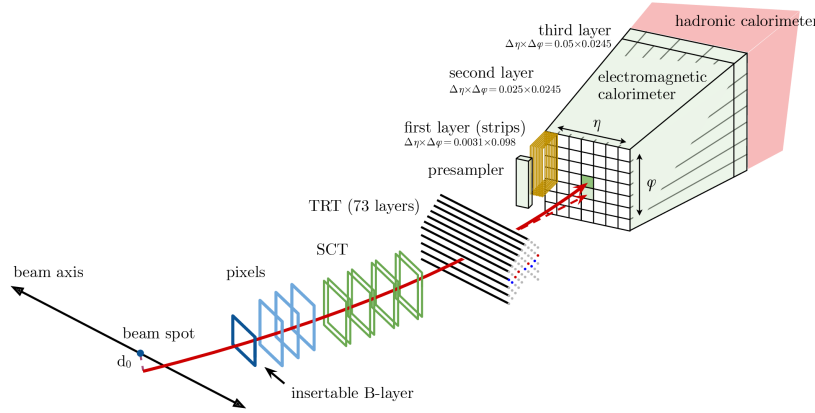


Figure 3.7: A schematic illustration of the path of an electron through the detector. The red trajectory shows the hypothetical path of an electron. The dashed red trajectory indicates the path of a photon produced by the interaction of the electron with the material in the tracking system. [121]

3.5.3 Electron isolation

The isolation requirement is used to differentiate the prompt production of electrons, muons, and photons in signal processes (from the hard-scattering vertex, or from the decay of heavy resonances such as the Higgs, W , and Z bosons) from background processes such as semileptonic decays of heavy quarks, hadrons misidentified as leptons and photons. The variables used for electron isolation are based on the calorimeter and tracking measurements [121]. The calorimeter-based isolation variable, $E_{T,cone}^{isol}$, is calculated by summing all topo-clusters within the cone of radius ΔR around the electron and after subtracting the core energy deposited by the candidate electron (with leakage correction if needed) and the contribution from pile-up events. The track-based isolation variable, $p_{T,var}^{isol}$, is calculated by summing the transverse momenta of the tracks found within a variable radius cone aligned with the electron track, excluding the candidate's own contribution.

A large variety of isolation working points are defined with $E_{T,cone}^{isol}$ and $p_{T,var}^{isol}$ to fit requirements in different physics analyses. The detailed criteria used in this thesis will be presented in the corresponding sections.

3.5.4 Electron-charge identification

The electric charge of an electron is determined from the curvature of the associated track reconstructed in the Inner Detector. The misidentification of the electron charge can result from the matching of an incorrect track to the electron candidate or from a mismeasurement of the curvature of the primary electron track. The charge-misidentification rate for reconstructed electron candidates can be reduced with an additional selection criterion

based on the output discriminant of a boosted decision tree (BDT). The measurements of the electron-charge misidentification are associated to the electron identification WPs.

3.5.5 Electron efficiency

The total electron efficiency (ϵ_{total}) is factorised as a product of different efficiency terms:

$$\epsilon_{\text{total}} = \epsilon_{\text{EMclus}} \times \epsilon_{\text{reco}} \times \epsilon_{\text{id}} \times \epsilon_{\text{iso}} \times \epsilon_{\text{trig}} = \left(\frac{N_{\text{cluster}}}{N_{\text{all}}} \right) \times \left(\frac{N_{\text{reco}}}{N_{\text{cluster}}} \right) \times \left(\frac{N_{\text{id}}}{N_{\text{reco}}} \right) \times \left(\frac{N_{\text{iso}}}{N_{\text{id}}} \right) \times \left(\frac{N_{\text{trig}}}{N_{\text{iso}}} \right) \quad (3.7)$$

The efficiency to reconstruct the EM topo-clusters (localised energy deposits) associated with all produced electrons, ϵ_{EMclus} , is given by the number of reconstructed EM topo-clusters N_{cluster} divided by the number of produced electrons N_{all} . This efficiency is evaluated entirely from simulation, where the reconstructed cluster is associated to a genuine electron produced at generator level. The reconstruction efficiency, ϵ_{reco} , is given by the number of reconstructed electron candidates N_{reco} divided by the number of EM-cluster candidates N_{cluster} . The identification efficiency, ϵ_{id} , is given by the number of identified and reconstructed electron candidates N_{id} divided by N_{reco} . The isolation efficiency is calculated as the number of identified electron candidates satisfying the isolation, identification, and reconstruction requirements N_{iso} divided by N_{id} . Finally, the trigger efficiency is calculated as the number of triggered (and isolated, identified, reconstructed) electron candidates N_{trig} divided by N_{iso} . All the efficiencies as well as their scale factors are measured using $Z \rightarrow e^+e^-$ and $J/\psi \rightarrow e^+e^-$ events. Whereas the $Z \rightarrow e^+e^-$ sample is used to extract all terms, the $J/\psi \rightarrow e^+e^-$ sample is only used to extract the identification efficiency since the significant background as well as the difficulties in designing a trigger for this process prevent its use in determining the reconstruction efficiency [121].

3.6 Muons

3.6.1 Muon reconstruction

Since muons are feebly interacting with calorimeters, they usually have long tracks. The muon track reconstruction is primarily based on the ID and MS tracking detectors. Information from the calorimeters is also used to account for cases of large energy loss in the calorimeters in the determination of track parameters. The reconstruction of tracks in ID are described in detail in Section 3.1. The reconstruction of tracks in the MS starts with the identification of short straight-line local track segments reconstructed from hits in an individual MS station. Segments in the different stations are combined into preliminary track candidates using a loose pointing constraint based on the IP and a parabolic trajectory that constitutes a first-order approximation to the muon bending in the magnetic field. Finally a global χ^2 fit of the muon trajectory through the magnetic field is performed, taking into account the effects of possible interactions in the detector material as well as the effects of possible misalignments between the different detector chambers. By using the complete detector information, five main reconstruction strategies are adopted and result in different

muon types [122]: combined (CB), inside-out combined (IO), muon-spectrometer extrapolated (ME), segment-tagged (ST), and calorimeter-tagged (CT).

Combined (CB) muons The CB muons are reconstructed by matching the MS tracks to the ID tracks and performing a combined track fit based on the ID and MS hits, taking into account the energy loss in the calorimeters. Most of muons in the region of $|\eta| < 2.5$ are CB muons. For $|\eta| < 2.5$, the MS tracks may be combined with short track segments reconstructed from hits in the pixel and SCT detectors leading to a subset of CB muons referred as silicon-associated forward (SiF) muons. The SiF muons are introduced for Run 2 data for the first time, and this allows to make better use of the ID near the boundary of its acceptance.

Inside-out (IO) combined muons The IO muons are reconstructed using a complementary inside-out algorithm, which extrapolates the ID tracks to the MS and searches for at least three loosely-aligned MS hits. The ID tracks, the energy loss in the calorimeters and the MS hits are then used in a combined track fit. It recovers some muons in regions of limited MS coverage and for low- p_T muons which may not reach the middle MS station.

Muon-spectrometer extrapolated (ME) muons If a MS track cannot be matched to an ID track, its parameters are extrapolated to the beamline and used to define an ME muon. Such muons are used to extend the acceptance outside that of the ID, thus fully exploiting the full MS coverage up to $|\eta| = 2.7$.

Segment-tagged (ST) muons The ST muons are reconstructed by requiring that an ID track extrapolated to the MS satisfies tight angular matching requirements to at least one reconstructed MS segment. This selection recovers low- p_T muons.

Calorimeter-tagged (CT) muons The CT muons are identified by extrapolating ID tracks through the calorimeters to search for energy deposits consistent with a minimum-ionising particle. Such deposits are used to tag the ID track as a muon, and the muon parameters are again taken directly from the ID track fit. This selection recovers muons in the forward region ($2.5 < |\eta| < 2.7$).

3.6.2 Muon identification

After reconstruction, high-quality muon candidates used for physics analyses are selected by a set of requirements on the number of hits in the different ID subdetectors and different MS stations, on the track fit properties, and on variables that test the compatibility of the individual measurements in the two detector systems. Several WPs are defined to suit the needs of the wide variety of physics analyses involving final states containing muons. The standard selections WPs are Loose, Medium and Tight. Two additional selection WPs are also designed for analysis targeting extreme phase space regions: High- p_T and Low- p_T . More details can be found in [122]. The criteria used in this thesis will be presented in the corresponding sections.

3.6.3 Muon isolation

Muons from the prompt decays of the SM bosons or the hypothetical BSM particles can be discriminated from muons from hadronic sources by measuring the amount of hadronic activity in their proximity. The transverse energy (or momentum if considering only tracks) reconstructed in a cone around a muon and divided by the muon p_T defines the muon isolation. Depending on the topology, most non-prompt muons that can be rejected using isolation criteria originate from heavy-flavour hadron decays. Isolation can be measured independently using either the ID tracks (track-based) or the topological cell clusters from calorimeters (calorimeter-based), or through a combination of the two (particle-flow-based isolation). Several isolation WPs are defined, balancing prompt-muon acceptance, rejection of non-prompt muons, and performance in close proximity to other objects. More details can be found in [122]. The detailed criteria used in this thesis will be presented in the corresponding sections.

3.6.4 Muon efficiency

The measurement of muon efficiency is more complicated than the one of electron efficiency because it depends on the muon types [122]. In the $|\eta| < 2.5$ region, both ID and MS are available and the tag-and-probe method is adopted using $Z \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ events. With $Z \rightarrow \mu^+\mu^-$ events, the measurement of the reconstruction and identification efficiencies for muons with $p_T > 15$ GeV and of the isolation efficiency down to $p_T = 3$ GeV are extracted. $J/\psi \rightarrow \mu^+\mu^-$ events are deployed to further extend the reconstruction and identification efficiencies down to $p_T = 3$ GeV. In the $2.5 < |\eta| < 2.7$ region, muons are reconstructed only as ME or SiF muons. In this region, instead of measuring the efficiency, the scale factors are calculated directly via a double-ratio method.

3.7 Overlap removal

The following overlap removal procedure [123] is applied to resolve ambiguities in which multiple electrons, muons or jets are reconstructed from the same detector signature, in particular for studies at high energy regime. First, any electrons which share one track are removed. Second, if an electron and muon share a track, the electron is rejected if the muon was associated with a signature in the MS, otherwise the muon is rejected. Third, any jet within $\Delta R < 0.2$ of an electron is rejected. Fourth, any jet within $\Delta R < 0.2$ of a muon is rejected if it has less than three associated tracks. Fifth, any electrons or muons within $\Delta R < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^{\text{lep}})$ of a jet passing the previous requirements are rejected. Finally, any large- R jet found within $\Delta R < 1.0$ of an electron is rejected.

3.8 Summary

This chapter described the algorithms and methodologies used in ATLAS for the reconstruction and calibration of the physics objects used in Chapter 4 and 5, including tracks, vertices, jets, electrons and muons. The reconstruction has been presented in the context of LHC Run 2 conditions as only Run 2 data is used in this thesis. In Run 3 and HL-LHC, the al-

gorithms for object reconstruction need to be improved and re-optimised. New techniques will be needed due to the planned detector upgrades and changes in the run conditions.

IN-SITU CALIBRATION OF THE $X \rightarrow b\bar{b}$ TAGGER USING $Z + \text{JETS}$ EVENTS

Tagging refers to the identification of a process of interest. It is essential for physics analyses where specific processes and topologies are used. In the previous chapter, the single b -tagging techniques and corresponding calibrations were discussed. This chapter focuses on boosted $H \rightarrow b\bar{b}$ tagging techniques, in particular on the boosted $X \rightarrow b\bar{b}$ tagger and its corresponding calibration. The work is documented in Ref. [10].

Section 4.1 introduces two tagging methods for boosted $H \rightarrow b\bar{b}$ events: the double b -tagging method and the boosted $X \rightarrow b\bar{b}$ tagger. Section 4.2 to 4.4 explain the methodology of the signal efficiency calibration for the $X \rightarrow b\bar{b}$ tagger. The final signal scale factors are reported in Section 4.5. Further improvements to the fit framework are presented in Section 4.6 followed by a summary and outlook in Section 4.7.

4.1 Boosted $H \rightarrow b\bar{b}$ tagging

The angular distance between the two b -jets (ΔR) from a Higgs boson is proportional to $2m_H/p_T^H$, where m_H is the mass of the Higgs boson and p_T^H is its transverse momentum. In the resolved region where $p_T^H < 250$ GeV, the two b -jets from the Higgs boson are well separated and the $H \rightarrow b\bar{b}$ is reconstructed using two small- R jets which are tagged as b -jets. In the boosted region where $p_T^H > 250$ GeV, the two b -quarks from the Higgs boson are collimated. The $H \rightarrow b\bar{b}$ is reconstructed as a large- R jet and the b -quarks are reconstructed as subjets. The choice of subjets are VR track jets (Section 3.3.2). Figure 4.1 shows the double subjet b -labelling (Section 3.3.5) efficiency at the truth level of a Higgs jet as a function of the Higgs jet p_T . The double subjet b -labelling efficiency decreases drastically when using fixed $R = 0.2$ track jets for $p_T^H > 800$ GeV. The double subjet b -labelling efficiency for VR track jets with $R_{max} = 0.2$ is worse for $p_T^H < 800$ GeV compared to VR track jets with $R_{max} = 0.3$ or $R_{max} = 0.4$. The double subjet b -labelling for VR track jets with $R_{max} = 0.4$ achieves the best performance. Similar studies is also done to optimise R_{min} and ρ leading to values of 0.02 and 30 GeV, respectively. Hence, in this thesis, the VR track jets used are built with the following parameters: $\rho = 30$ GeV, $R_{max} = 0.4$, $R_{min} = 0.02$ if there is no specific note.

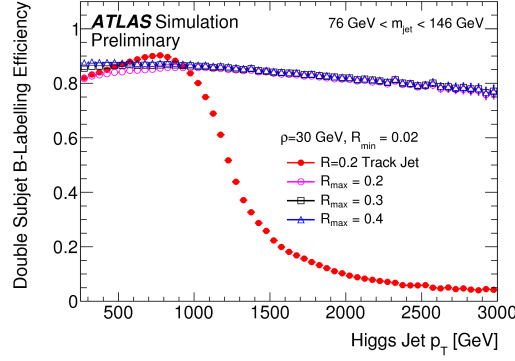


Figure 4.1: Efficiency of double subject b -labelling at the truth level of a Higgs jet as a function of the Higgs jet p_T . The efficiency plots are for VR track jets with $\rho = 30$ GeV and $R_{min} = 0.02$ for varying values of R_{max} . The Higgs bosons are obtained from the BSM physics simulation of a Randall-Sundrum graviton. Details can be found in [106].

4.1.1 Double b -tagging method

The double b -tagging method for identifying $H \rightarrow b\bar{b}$ tags the VR track jets ghost-associated to the large- R jets using the b -tagging algorithms, discussed in Section 3.4. $H \rightarrow b\bar{b}$ is identified as the large- R jets with at least two b -tagged ghost-associated VR track jets. The **ghost-association** [124] method is used because it provides a more robust matching procedure than a simple geometric matching. In the ghost-association procedure, the "ghosts" are the 4-vector of the track jets with the track jet p_T set to an infinitesimal amount. Then, the large- R jet is reconstructed using the topo-clusters and the ghost with the anti- k_t algorithm. The reclustered large- R jet is identical to the ungroomed parent of the trimmed large- R jet which is used for the identification, with the addition of the associated track jets ghosts. By only using the direction of the track jets, the kinematic of the trimmed large- R jet is still the same as without the ghost-association performed. The performance and calibrations are based on the isolated VR track jets and jet flavours as presented in Section 3.4. This method is adopted by most of analyses on the boosted Higgs measurements in Section 5.1. However, the double b -tagging method is degraded when probing extremely high p_T . For example, in the searches for heavy resonance decaying to HH via $b\bar{b}b\bar{b}$, the efficiency of the double b -tagging method is degraded if the resonance mass is higher than 1.5 TeV, thus, we have to use looser tagging requirements [125]. Therefore, new methods exploiting the two-body kinematics have been developed in the last 2-3 years (Section 4.1.2).

4.1.2 Boosted $X \rightarrow b\bar{b}$ tagger

Instead of tagging b -jets, the new $X \rightarrow b\bar{b}$ tagger [126] tags the $H \rightarrow b\bar{b}$ object. The main backgrounds are thus dijet from QCD (also called multijet) and top jets, which eventually contain a b -quark inside the large- R jets as shown by Figure 4.2. The new $X \rightarrow b\bar{b}$ tagger is a neural network based algorithm. The inputs of the $X \rightarrow b\bar{b}$ tagger are the p_T , η of the large- R jets and the outputs of the DL1r algorithm (see Section 3.4) for the first three VR track jets. Its output is the probability of the large- R jets coming from QCD (p_{QCD}), top

Table 4.1: The $X \rightarrow b\bar{b}$ tagger threshold values for all $X \rightarrow b\bar{b}$ WPs with $f_{top} = 0.25$.

Working point	50%	60%	70%
Threshold value	3.09	2.44	1.76

(p_{top}) and Higgs (p_{Higgs}). And the discriminant used to define working points is:

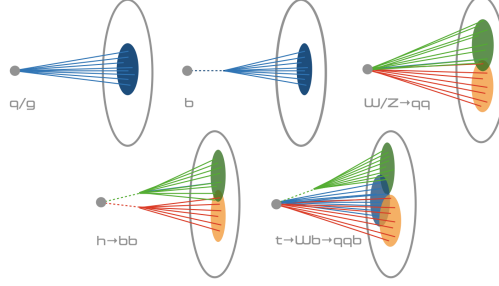


Figure 4.2: Pictorial representation of ordinary quark and gluon jets (top left), b-jets (top center), and boosted-jet topologies, emerging from high- p_T W and Z bosons (top right), Higgs bosons (bottom left), and top quarks (bottom right) decaying to all-quark final states. [127]

$$D_{Xbb} = \ln \frac{p_{Higgs}}{f_{top} \cdot p_{top} + (1 - f_{top}) \cdot p_{QCD}} \quad (4.1)$$

Where f_{top} determines the fraction of top background and is set to 0.25 [10]. The optimised working points are defined in Table 4.1.

The new $X \rightarrow b\bar{b}$ tagger shows improved background rejection than the double b -tagging method at the same signal efficiency. Figure 4.3 shows the dijet and top jet rejection as a function of large- R jet p_T for a constant $H \rightarrow b\bar{b}$ tagging efficiency of 60%. The double b -tagging method with fixed-radius track-jets of $R = 0.2$ begins to lose discriminating power for $p_T > 800$ GeV. It is consistent with what we see from Figure 4.1. The double b -tagging method with VR track-jets can resolve multiple subjets in boosted objects with p_T higher than 1 TeV. The $X \rightarrow b\bar{b}$ tagger has the best performance in terms of multijet and top jets rejection. Its performance depends mainly on flavour compositions of the large- R jets. The tagger has similar tagging efficiency for $H \rightarrow b\bar{b}$ and $Z \rightarrow b\bar{b}$ (See Figure 4.4). Since the tagger is trained using the MC simulation which can not fully describe the real data, its performance in the MC simulation is not the same as the one in real data. To use the tagger in physics analysis, we have to calibrate it. The calibration of the signal efficiency of the tagger, which is part of my thesis work, are detailed in this chapter.

4.2 Methodology

It is the first time that the calibration of signal efficiency for the $X \rightarrow b\bar{b}$ tagger is performed. The calibration of the signal efficiency is performed using two topologies. $Z\gamma$ events are used to cover the p_T regime from 250 to 450 GeV because the statistics are limited beyond

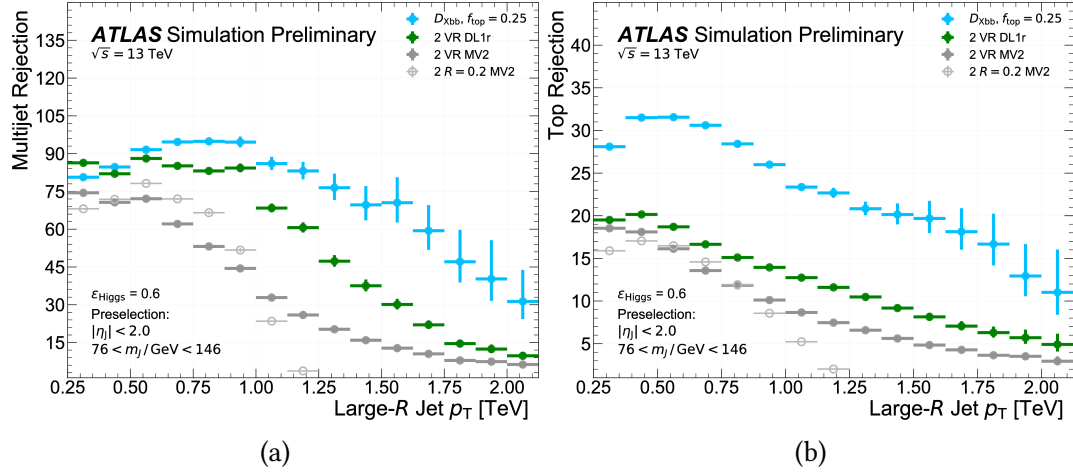


Figure 4.3: Dijet (left) and top (right) jet rejection as a function of large- R jet p_T with a constant $H \rightarrow b\bar{b}$ tagging efficiency of 60% [126] using the $X \rightarrow b\bar{b}$ tagger (blue), the double b -tagging method with DL1r (green), the double b -tagging method with MV2 (dark gray) and the double b -tagging method with MV2 but using fixed radius ($R = 0.2$) jets (light gray).

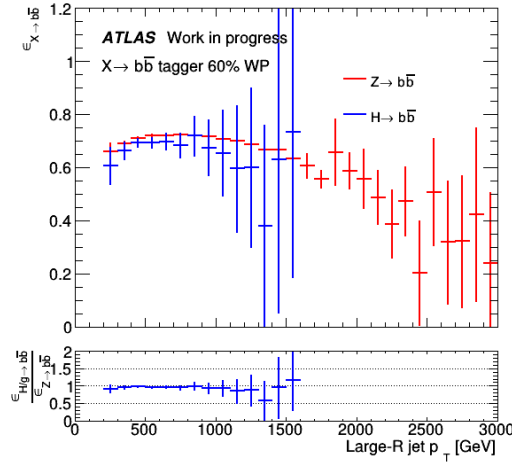


Figure 4.4: Tagging efficiency of the $X \rightarrow b\bar{b}$ tagger for $Z \rightarrow b\bar{b}$ (red) and $H \rightarrow b\bar{b}$ (blue) events as a function of p_T at working point 60%. The large- R jets used are the Higgs candidate jets after the selection described in Section 5.4. The $Z \rightarrow b\bar{b}$ events are from SHERPA 2.2.8 $Z \rightarrow b\bar{b}$ + jets sample. The $H \rightarrow b\bar{b}$ events are from ggF sample described in Section 5.3. The bottom panel shows the ratio of tagging efficiency for $Z \rightarrow b\bar{b}$ and $H \rightarrow b\bar{b}$.

450 GeV. Z + jets events are used for $p_T > 450$ GeV. The calibration of the mistag rate of top jets uses $t\bar{t}$ events. The calibration using $g \rightarrow b\bar{b}$ events is also studied. All studies are documented in Ref. [126]. My work focuses on the calibration using Z + jets events.

The main motivations to use Z + jets events are: 1) There are more statistics at the high p_T regime; 2) The calibration results could be applied to the $H \rightarrow b\bar{b}$ topology since the $X \rightarrow b\bar{b}$ tagger is trained mass-independently. However, the dijet events are overwhelming at the high p_T regime, which is one of the difficulties in this calibration.

The data-to-simulation scale factor is defined as:

$$SF = \frac{\epsilon^{data}}{\epsilon^{MC}} \quad (4.2)$$

Where ϵ^{data} and ϵ^{MC} are the efficiency of $X \rightarrow b\bar{b}$ tagger in data and MC, respectively, for a given cut on the $D_{Xb\bar{b}}$ (see Table 4.1). Using the definition of signal efficiency, we can reformulate Equation 4.2 as:

$$SF = \frac{\frac{N_{passed}^{sig,data}}{N_{passed}^{sig,MC}}}{\frac{N_{tot}^{sig,data}}{N_{tot}^{sig,MC}}} = \frac{\mu^{post-tag}}{\mu^{pre-tag}} \quad (4.3)$$

$N_{tot}^{sig,MC}$ and $N_{tot}^{sig,data}$ are the total number of signal events in MC simulation and data, respectively. $N_{passed}^{sig,MC}$ and $N_{passed}^{sig,data}$ are the number of signal events passing the $X \rightarrow b\bar{b}$ tagger in MC simulation and data, respectively. $\mu^{post-tag}$ and $\mu^{pre-tag}$ are the signal strength after and before passing the $X \rightarrow b\bar{b}$ tagger. In this calibration, the signal events are $Z(\rightarrow b\bar{b}) + \text{jets}$. The dominant backgrounds are dijet, $t\bar{t}$, $Z(\rightarrow qq) + \text{jets}(q = c, l)$ and $W(\rightarrow qq) + \text{jets}$. $\mu^{post-tag}$ and $\mu^{pre-tag}$ are measured using different methods due to the rejection of a large fraction of backgrounds by the $X \rightarrow b\bar{b}$ tagger.

Since it is impossible to separate $Z \rightarrow b\bar{b}$ from $Z \rightarrow q\bar{q}$ ($q = c, l$) events without flavour information, we are measuring inclusive $Z \rightarrow q\bar{q}$ after the $X \rightarrow b\bar{b}$ tagger. The notation $Z \rightarrow q\bar{q}$ is used to include $Z \rightarrow b\bar{b}$ and $Z \rightarrow q\bar{q}$ ($q = c, l$) events. The μ , defined as $N_{passed}^{Z \rightarrow q\bar{q},data} / N_{passed}^{Z \rightarrow q\bar{q},MC}$, is measured by fitting the large- R jet mass distribution. The inclusive $Z \rightarrow q\bar{q}$, $W \rightarrow qq$ and $t\bar{t}$ are described by MC templates. The MC simulations are detailed in Section 4.3.1.2. The dijet background is estimated directly from the data. Assuming that the $Z \rightarrow b\bar{b}$ and $Z \rightarrow q\bar{q}$ ($q = c, l$) have the same behaviour and get the same normalisation factor from the fit, the fitted μ equals to $\mu^{post-tag}$ in Equation 4.3. The benchmark working point of the $X \rightarrow b\bar{b}$ tagger used to establish the method is 60%. The measurement of $\mu^{post-tag}$ is presented in Section 4.3.

Before the $X \rightarrow b\bar{b}$ tagger is applied, the signal-to-background ratio is too low to measure $\mu^{pre-tag}$. Therefore, $\mu^{pre-tag}$ is obtained through the auxiliary measurement using $Z \rightarrow \ell\ell$ events. Leptons are used to trigger the events, the backgrounds of $Z \rightarrow \ell\ell$ are efficiently rejected. The signal strength of $Z \rightarrow \ell\ell$ can be measured as:

$$\mu^{lep} = \frac{N_{ll}^{data} - N_{bkg,ll}^{MC}}{N_{Z \rightarrow \ell\ell}^{MC}} \quad (4.4)$$

where $N_{Z \rightarrow q\bar{q}}^{MC}$ is the predicted number of $Z \rightarrow q\bar{q}$ events before tagging, N_{ll}^{data} is the total observed number of events after the $Z \rightarrow \ell\ell$ selection. $N_{bkg,ll}^{MC}$ is the predicted number of events of the background of $Z \rightarrow \ell\ell$. $N_{Z \rightarrow \ell\ell}^{MC}$ is the predicted number of $Z \rightarrow \ell\ell$ events.

The observed number of $Z \rightarrow q\bar{q}$ events before tagging ($N_{Z \rightarrow q\bar{q}}^{data}$) can be estimated at the first order by using μ^{lep} :

$$N_{Z \rightarrow q\bar{q}}^{data} = \mu^{lep} \times N_{Z \rightarrow q\bar{q}}^{MC} \quad (4.5)$$

μ^{lep} is equal to $\mu^{pre-tag}$ in this calibration. $\mu^{pre-tag}$ depends only on the leptonic channel. There is no need to apply corrections from the leptonic channel to the hadronic channel for two reasons. First the MC setup used to generate $Z \rightarrow \ell\ell$ samples is the same as for $Z \rightarrow q\bar{q}$ samples. Any discrepancy due to different MC setups is eliminated. Second, from the p_T distribution of $Z \rightarrow \ell\ell$ and $Z \rightarrow q\bar{q}$, the ratio between $Z \rightarrow \ell\ell$ and $Z \rightarrow q\bar{q}$ is nearly constant for $p_T < 1000$ GeV (See Figure 4.12). No bias is introduced if we look at differential p_T bins. The measurement of $\mu^{pre-tag}$ is presented in Section 4.4.

The scale factors are reported in Section 4.5 followed by a discussion on the improvement of the fit framework in Section 4.6.

4.3 Measurement of $\mu^{post-tag}$

4.3.1 Data and Monte Carlo simulations

4.3.1.1 Data

The full Run 2 dataset collected by the ATLAS detector during 2015-2018 are used in the study. The total luminosity of the full Run 2 data is 139 fb^{-1} .

4.3.1.2 Monte Carlo simulations

The simulations are crucial to understand the detector, to develop analysis strategies, to estimate the sensitivity to different physics processes, to develop and validate object reconstruction algorithms, and so on. In ATLAS, the simulation chain consists of four steps. First, the pp collision events are calculated perturbatively and simulated by the MC event generators. The event generators take care of the production and decay of particles in a given process. The specific generators will be introduced for each process used in this thesis. Then, the interactions between the generated particles and the ATLAS detector are simulated. The simulation is based on GEANT 4 [128] and requires an accurate description of the detector material and geometry. After, digitization is performed to convert the deposited energy in detector. The output format of the simulation is identical to the real detector output format. Finally, the object reconstruction are processed as described in Chapter 3. The algorithms used for simulated data are same as the ones used for real data.

The hadronically decaying Z and W events, with one additional parton at NLO order accuracy and up to 4 additional partons at LO, are generated using SHERPA 2.2.8 [129] with NNPDF3.0 [130] NNLO parton distribution function (PDF). The samples used in this calibration are generated with $p_T^V > 200$ GeV. The SHERPA 2.2.8 sample allows for comparing the SHERPA cluster model and fragmentation model. The $Z + \text{jets}$ samples are separated into $Z \rightarrow b\bar{b}$ and $Z \rightarrow q\bar{q}$ ($q = c, l$) datasets, which helps to increase the statistics of $Z + \text{jets}$ events after applying the $X \rightarrow b\bar{b}$ tagger. The HERWIG ++ [131] samples are used as the alternative samples for hadronically decaying $Z + \text{jets}$. The events are generated with one additional parton at LO.

The simulated $t\bar{t}$ events are generated at tree-level using POWHEG-Box V2 [132] with $hdamp = 1.5m_t$ and the NNPDF3.0 [130] NLO PDF. The hadronization is performed us-

Table 4.2: Summary of the triggers used in the analysis. They are applied as an **OR** and all are required to be active. The offline threshold is the offline jet cut above which the triggers are 99% efficient.

Year	Trigger	Offline Treshhold [GeV]
2015	HLT_j360_a10_lcw_sub_L1J100	$p_{T,J} > 410 \text{ GeV}$
2016	HLT_j420_a10_lcw_L1J100	$p_{T,J} > 450 \text{ GeV}$
2017	HLT_j440_a10t_lcw_jes_L1J100	$p_{T,J} > 470 \text{ GeV}$
	HLT_j390_a10t_lcw_jes_30smcINF_L1J100	$p_{T,J} > 420 \text{ GeV}, m_J > 50 \text{ GeV}$
2018	HLT_j460_a10t_lcw_jes_L1J100	$p_{T,J} > 490 \text{ GeV}$
	HLT_j420_a10t_lcw_jes_35smcINF_L1J100	$p_{T,J} > 450 \text{ GeV}, m_J > 60 \text{ GeV}$
	HLT_j420_a10t_lcw_jes_35smcINF_L1SC111	$p_{T,J} > 450 \text{ GeV}, m_J > 60 \text{ GeV}$

ing PYTHIA 8.230 [133] with A14 tune and the NNPDF23 LO PDF [134]. The decay of b -hadrons is performed using EvtGen [135]. Since the lepton veto is applied in this calibration, only fully hadronically decaying $t\bar{t}$ are used. Regarding the alternative samples, the parton shower uncertainty is estimated by replacing PYTHIA 8.230 with HERWIG, and the matrix element uncertainty is estimated by replacing AMC@NLO [136] with POWHEG.

Simulated QCD dijet events are generated using PYTHIA 8.235 [133] generator with the A14 tune and the NNPDF23 LO PDF [134]. Moreover, to increase the statistics for developing the background estimation method for the QCD background, we select the events with at least one b -hadron in the leading large- R truth jet [137].

4.3.2 Event selection

The data and MC simulations are preprocessed with pre-selection requirements to have high-quality events. The events are required to have a primary vertex, see Section 3.2. The non-collision backgrounds originating from the calorimeter noise, the beam-halo interactions, or the cosmic rays which can lead to the spurious calorimeter signals, are suppressed. Events with isolated e or μ with $p_T > 25 \text{ GeV}$ are vetoed. The large- R jets used in this calibration are the trimmed LCTopo jets. The events are required to have at least two large- R jets with $p_T > 200 \text{ GeV}$. The events are required to pass the HLT for single large- R jets. The summary of the triggers is shown in Table 4.2. The name convention is explained in Section 2.3. The offline threshold is the offline jet cut above which the triggers are 99% efficient, corresponding the unified offline trigger cuts for the leading large- R jets with $p_T > 450 \text{ GeV}$ and $m_J > 50 \text{ GeV}$. The overlap removal procedure is also applied as described in Section 3.7.

In addition to pre-selection, extra event selections are applied to eliminate the mis-modelled phase space and increase the signal significance.

p_T balance and rapidity cut Two additional cuts are applied:

- $\frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}} < 0.15$
- $|\Delta y_{1,2}| < 1.2$

Where $p_{T,1}$ is the transverse momentum of the leading large- R jets and $p_{T,2}$ is the trans-

verse momentum of the sub-leading large- R jets. $\Delta y_{1,2}$ is the difference of rapidity of the leading large- R jet and the sub-leading large- R jet. Figure 4.5 shows the distribution of $(p_{T,1} - p_{T,2})/(p_{T,1} + p_{T,2})$ and $|\Delta y_{1,2}|$ for $Z \rightarrow b\bar{b}$, $Z \rightarrow q\bar{q}$ ($q = c, l$), $W \rightarrow qq$, dijet and $t\bar{t}$ process and in data. From the comparisons between data and dijet, these two selections help reducing the mis-modelling of dijet. For the p_T asymmetry distribution, the ratio starts to decrease from 0.15. For the rapidity difference, it is tightened to 1.2 to reject more dijet events.

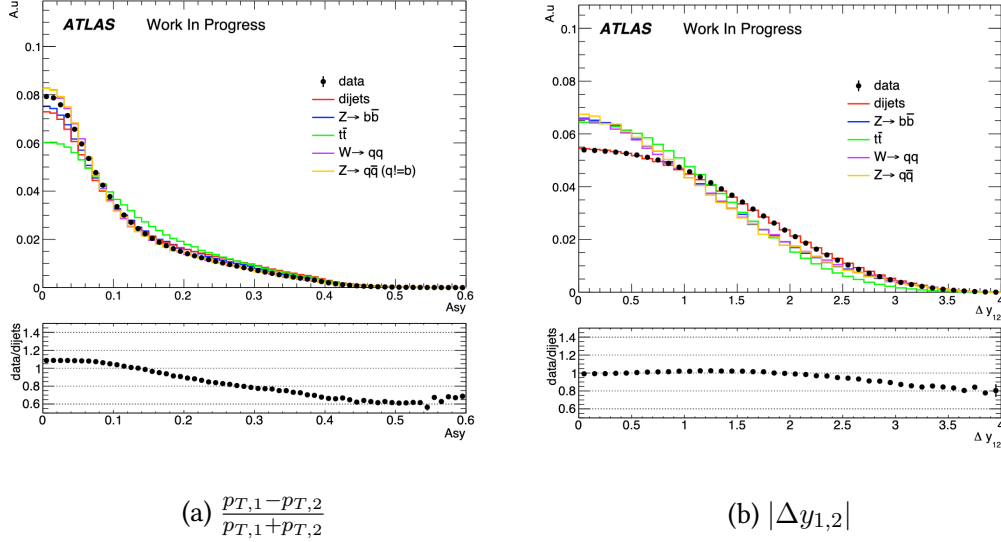


Figure 4.5: Distribution of the p_T asymmetry, $(p_{T,1} - p_{T,2})/(p_{T,1} + p_{T,2})$ (left) and the pseudorapidity difference between the first two p_T leading large- R jet, $|\Delta y_{1,2}|$ (right) for $Z \rightarrow b\bar{b}$ (blue), $Z \rightarrow q\bar{q}$ ($q = c, l$) (yellow), $W \rightarrow qq$ (purple), dijet (red) and $t\bar{t}$ (green) process and collision data (black). The leading large- R jets are required to have p_T larger than 450 GeV and mass higher than 50 GeV. The subleading large- R jets are required to have p_T larger than 200 GeV.

Z boson candidate jet In an event, the large- R jet with the highest transverse momentum is chosen as the Z boson candidate jet. The Z boson candidate jets are required to have at least two ghost-associated VR track-jets with $p_T > 7$ GeV.

The signal efficiency of the $X \rightarrow b\bar{b}$ tagger at 60% working point with different definitions of $Z \rightarrow b\bar{b}$ is shown in Figure 4.6. The studies of the signal efficiency are based on the truth label of the large- R jets and the VR track jets. The label definition can be found in Section 3.3.5. The efficiency varies as a function of the transverse momentum of the large- R jet. The signal efficiency is closer to 60% if the $Z \rightarrow b\bar{b}$ candidate is labelled as *matched* Z boson in $Z(\rightarrow b\bar{b}) + \text{jets}$ samples. *Matched* Z means that the large- R jet is labelled as *contained* Z or as *Other from V*. If the additional $2b$ flavour requirements are added, the signal efficiency increases.

The signal significance S , defined as $S/\sqrt{S+B}$ is calculated after each selection for the large- R jet mass window: $70 < m < 110$ GeV and for $p_T > 450$ GeV. After selections, the signal significance is 2.82 before the $X \rightarrow b\bar{b}$ tagger applied and 17.63 after the $X \rightarrow b\bar{b}$ tagger applied at 60% WP.

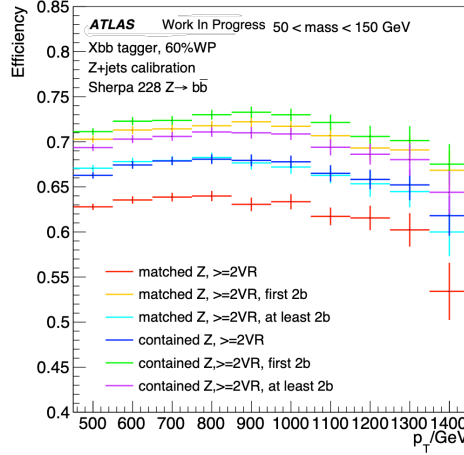


Figure 4.6: Signal efficiency of the $X \rightarrow b\bar{b}$ tagger as a function of p_T at working point 60% with different definitions of $Z \rightarrow b\bar{b}$ using SHERPA 2.2.8 $Z \rightarrow b\bar{b} + \text{jets}$ sample. The Z boson candidate jets are required to have a mass between 50 and 150 GeV. *Matched Z*: *Contained Z + Other from V*. The label definition can be found in Section 3.3.5.

Figure 4.7 and 4.8 show data and MC comparisons before and after the $X \rightarrow b\bar{b}$ tagger 60% WP. Before the $X \rightarrow b\bar{b}$ tagger is applied, for all the p_T bins, the data/MC ratio is seen to be smaller than 1 because the cross section of dijet is overestimated. The data and MC simulation ratio is around 0.8, roughly constant as a function of p_T . After the $X \rightarrow b\bar{b}$ tagger is applied at 60% WP, the ratio of data and MC is not constant. The ratio increases as function of the large- R jet mass for all p_T bins. The flavour composition modelling might be one of the important issues causing the disagreement between data and MC after tagging (We should note that the $X \rightarrow b\bar{b}$ tagger selects flavours). Since there is a non-negligible mis-modelling of QCD for data, a data-driven method is adopted for the background estimation based on a direct fit to data.

4.3.3 Inclusive $Z \rightarrow q\bar{q}$ template

The inclusive $Z \rightarrow q\bar{q}$ template is created from the combined samples of $Z(\rightarrow b\bar{b}) + \text{jets}$ and $Z(\rightarrow q\bar{q}) + \text{jets}(q = c, l)$. The large- R jets are required to be labelled as *contained Zqq* or *Other from V*, i.e. *matched Z*. The label definition can be found in Section 3.3.5. Different models were tested to describe the $Z \rightarrow q\bar{q}$: the sum of three Gaussian functions, the double-sided crystal ball function (DSCB), and the Bukin function. The DSCB is chosen because it gives a better description of the template shape and a better understanding of the systematic uncertainties with its mean and width parameters.

The fraction of $Z \rightarrow b\bar{b}$ events in the template is computed after the $X \rightarrow b\bar{b}$ tagger at 60% WP. The results are shown in Table 4.3. There are around 90% $Z \rightarrow b\bar{b}$ events in the template.

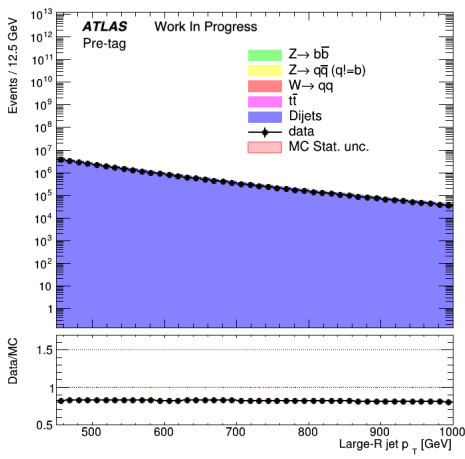
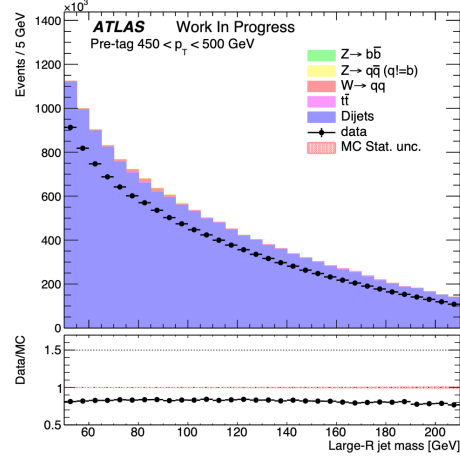
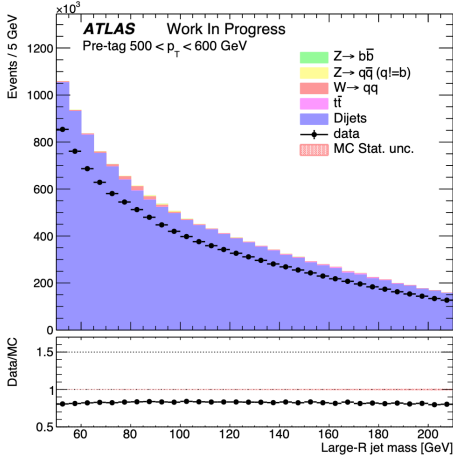
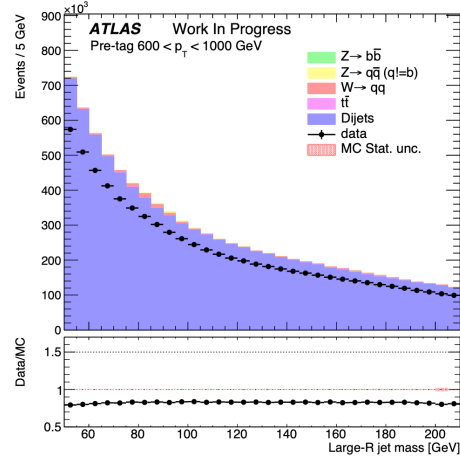
(a) p_T distribution(b) $450 < p_T < 500$ GeV(c) $500 < p_T < 600$ GeV(d) $600 < p_T < 1000$ GeV

Figure 4.7: Data and MC comparison for Z + jets for differential p_T before the $X \rightarrow b\bar{b}$ tagger applied. Figure 4.7a shows the Z boson candidate jet p_T distribution. Figure 4.7b, 4.7c and 4.7d show the Z boson candidate jet mass distribution for $450 < p_T < 500$ GeV, $500 < p_T < 600$ GeV and $600 < p_T < 1000$ GeV. In each plot, the stacked distribution is from MC simulations and the black dots are from data. The red shadow band shows the MC statistical uncertainty. The bottom panel shows the ratio of event yields between the data and the MC predictions.

4.3.4 Background modelling

4.3.4.1 Resonant background

The mass distribution of $Z \rightarrow b\bar{b}$, $Z \rightarrow q\bar{q}$ ($q = c, l$), $W \rightarrow qq$ and $t\bar{t}$ after tagging are shown in Figure 4.9. As mentioned in Section 4.3.3, an inclusive $Z \rightarrow q\bar{q}$ template is used for $Z \rightarrow b\bar{b}$ and $Z \rightarrow q\bar{q}$ ($q = c, l$).

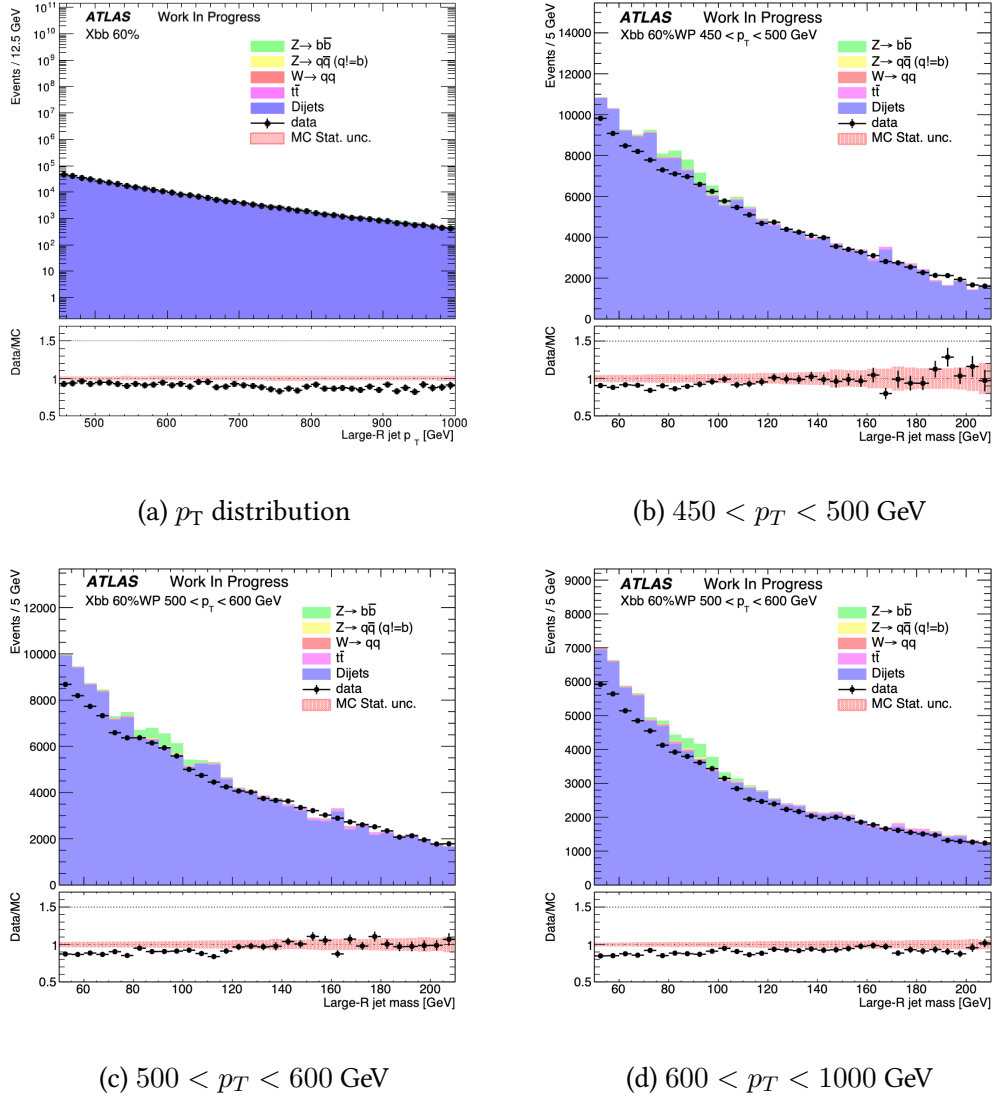


Figure 4.8: Data and MC comparison for $Z + \text{jets}$ for differential p_T after $X \rightarrow b\bar{b}$ the tagger 60% WP. Figure 4.8a shows the Z boson candidate jet p_T distribution. Figure 4.8b, 4.8c and 5.7c show the Z boson candidate jet mass distribution for $450 < p_T < 500$ GeV, $500 < p_T < 600$ GeV and $600 < p_T < 1000$ GeV. In each plot, the stacked distribution is from MC simulations and the black dots are from data. The red shadow band shows the MC statistical uncertainty. The bottom panel shows the ratio of event yields between the data and the MC predictions.

$W \rightarrow q\bar{q}$ The $W \rightarrow q\bar{q}$ template is created from the $W(\rightarrow q\bar{q}) + \text{jets}$ sample. The large- R jets are required to be labelled as *contained Wqq* or *Other from V*. The label *matched W* will be used in the next sections. The contribution of $W \rightarrow q\bar{q}$ events is small and modelled by a DSCB function.

$t\bar{t}$ The $t\bar{t}$ template is created from the $t\bar{t}$ sample. The contribution from $Wq\bar{q}$ from *top* is negligible compared to others. In order to facilitate the modelling, two templates are created: $t\bar{t}$ (tqqb) which consists of the large- R jets labelled as *contained tqqb*, $t\bar{t}$ (no tqqb)

Table 4.3: The fraction of the $Z \rightarrow b\bar{b}$ events in the inclusive $Z \rightarrow q\bar{q}$ template after the $X \rightarrow b\bar{b}$ tagger at 60% WP.

p_T	Ratio
$450 \leq p_T < 500$ GeV	0.92
$500 \leq p_T < 600$ GeV	0.92
$600 \leq p_T < 1000$ GeV	0.90

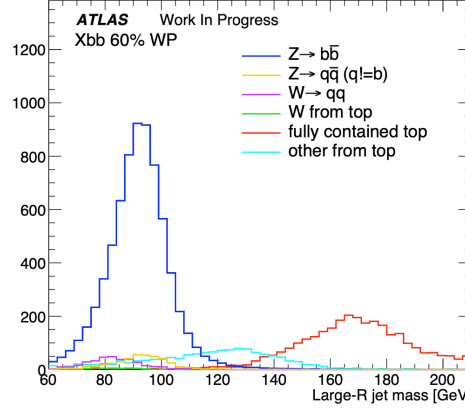


Figure 4.9: Mass distributions of the large- R jets for $Z \rightarrow b\bar{b}$ (blue), $Z \rightarrow q\bar{q}$ ($q = c, l$) (yellow), $W \rightarrow qq$ (purple) and fully contained top (*contained tq q b*, red) and other remnants from top (cyan) after $X \rightarrow b\bar{b}$ tagger 60% WP.

which consists of large- R jets labelled as *Wqq from top* or *Other from top*. A DSCB function models each category. This method needs to be checked with looser WPs where the contribution from *Wqq from top* can not be neglected.

4.3.4.2 Non-resonant background model

The modelling of the QCD background is of prime importance as it is the leading background in this calibration, and the combined signal plus background fit to data is highly sensitive to the functional form used to describe this background. The strategy of the QCD background modelling is described in the following:

First, two families of functions are considered to describe the QCD background form: exponentiated polynomials and standard polynomials. The background-only fit is performed to the side-band in data with functions of different order from each family. The data side-band includes the lower side-band range of the Z boson candidate jet mass from 50 GeV to 70 GeV, and the upper side-band range of the Z boson candidate jet mass from 110 GeV to 150 GeV. An unbinned maximum likelihood fit is performed (See Section 4.3.6).

Then, a F -test [138] is performed to decide the optimal order of function for each family. The F -test provides a quantitative justification of the suitable number of the free param-

ters from the two sets of functions. The test statistic $F_{p,q}$ is calculated from each fit as:

$$F_{p,q} = \frac{\frac{\chi_p^2 - \chi_q^2}{n_q - n_p}}{\frac{\chi_q^2}{n_{bins} - n_q}} \quad (4.6)$$

Where χ_p^2 and χ_q^2 are the χ^2 values computed in n bins of two fits, having n_p and n_q as the number of free parameters for the p^{th} and q^{th} function respectively. In the asymptotic limit, the test statistic $F_{p,q}$ follows a Fischer distribution, $\mathcal{F}(F|n_q - n_p, n_{bins} - n_q)$.

Afterwards, a spurious signal test is performed to choose an optimal function out of the two suitable functions resulting from F -test. The spurious signal is the bias introduced by the parameterised function in the modelling of the background. In order to quantify the spurious signal, the signal+background model is applied to fit the Z boson candidate jet mass distribution of the background-only MC sample, i.e. dijet simulated samples. The fitted signal strength μ^{SS} is treated as the spurious signal strength. Dijet simulated samples are reweighted w.r.t data side-band to have a good description of dijet distribution. The spurious signal test is performed for all functions passing the F -test. The function with the smallest μ^{SS} is chosen as the optimal function for the final fit. The maximum between μ^{SS} and its statistical uncertainty is taken into account as the spurious signal uncertainty. The implementation of the spurious signal uncertainty in the statistical model is described in Section 4.3.6 (Equation 4.7). The optimal functions to describe the multijet background and the corresponding spurious signal uncertainties are summarised in Table 4.4.

Table 4.4: The optimal function for differential p_T for 60% WP and its spurious signal uncertainty.

p_T	Optimal function	Spurious signal uncertainty
$450 \leq p_T < 500 \text{ GeV}$	$\sum_{i=0}^3 a_i \left(\frac{m}{100[\text{GeV}]} \right)^i$	0.113
$500 \leq p_T < 600 \text{ GeV}$	$\sum_{i=0}^3 a_i \left(\frac{m}{100[\text{GeV}]} \right)^i$	0.065
$600 \leq p_T < 1000 \text{ GeV}$	$a_0 \exp(\sum_{i=1}^3 a_i \left(\frac{m}{100[\text{GeV}]} \right)^i)$	0.069

4.3.5 Systematic uncertainties

Theoretical systematic uncertainties as well as experimental systematic uncertainties are considered in this calibration.

Theoretical systematic uncertainties The uncertainties on the modelling of the Z and $W + \text{jets}$ events are evaluated using the alternative generator samples and alternative fragmentation models. The renormalisation and factorisation scale and PDF+ α_s uncertainties are also considered. A 10% normalisation uncertainty is assigned to the $W + \text{jets}$ background [139], while the normalisation of the $Z + \text{jets}$ signal is a free parameter in the fit.

The $t\bar{t}$ background is small in the analysis. The $t\bar{t}$ normalisation is fixed to MC prediction in the final fit. A 24% normalisation uncertainty on the $t\bar{t}$ production cross section

is applied [140]. The alternative parton shower (PS) model and the alternative matrix element (ME) generator uncertainties are considered (See Section 4.3.1.2). The initial state radiation (ISR) and the final state radiation (FSR) variation shown in [141] are smaller than the systematic uncertainties mentioned above and hence neglected.

Spurious signal systematic uncertainties The spurious signal systematic uncertainties are estimated in Section 4.3.4.2 and considered in the fit model.

Large- R jet systematic uncertainties The large- R jet systematic uncertainties are from the in-situ jet calibration mentioned in Section 3.3.4, including systematic uncertainties for JES, JER, JMS and JMR. For each systematic uncertainty, the correlation is set in the fit model for $Z \rightarrow q\bar{q}$, $W \rightarrow qq$ and $t\bar{t}$.

Luminosity and pileup reweighting uncertainties The total Run 2 pp collision data sample corresponds to an integrated luminosity of 139 fb^{-1} with an uncertainty of 1.7% [142]. The pileup reweighting uncertainties are considered as well.

4.3.6 Statistical model

The main target of this measurement is to extract $\mu^{\text{post-tag}}$ for different p_T bins: $450 \leq p_T < 500 \text{ GeV}$, $500 \leq p_T < 600 \text{ GeV}$, $600 \leq p_T < 1000 \text{ GeV}$. The binning is chosen to have decent statistics for the studies in each bin. An unbinned likelihood fit is applied to the jet mass distribution of Z candidates after applying the $X \rightarrow b\bar{b}$ tagger. The likelihood is defined as a product of global Poisson distributions with products of probability for each event according to signal+background model and the constraints for systematic uncertainties:

$$\mathcal{L}(\vec{m}|\mu, \vec{\theta}, N_B, \vec{p}) \quad (4.7)$$

$$= \text{Pois}(n|N) \quad (4.8)$$

$$\times \left[\frac{N_Z(\mu, \vec{\theta}) + N_{SS}(\theta_{SS})}{N} \times \mathcal{S}_v(\vec{m}|\vec{\theta}) + \sum_{t \in \text{Cats}T} \frac{N_t(\vec{\theta})}{N} \times \mathcal{T}_t(\vec{m}|\vec{\theta}) + \frac{N_B}{N} \times \mathcal{B}(\vec{m}|\vec{p}) \right] \quad (4.9)$$

$$\times \prod_{s \in S} \mathcal{G}(0|\theta_s, 1) \quad (4.10)$$

$\vec{m}$ is the set of observables for each event, i.e. the large- R jet mass. The observed number of events is given as n . The total expected number of events $N = N_Z + N_{SS} + N_V + \sum_{t \in \text{Cats}T} N_t + N_B$, where $\text{Cats}T = \{t\bar{t}^{\text{Top}}, t\bar{t}^{\text{Other}}\}$, is the sum of the expected number of events for the Z boson signal (N_Z), the contribution from spurious signal (N_{SS}), the $t\bar{t}$ process ($N_{t\bar{t}}^{\text{Top}}$ and $N_{t\bar{t}}^{\text{Other}}$), and the non-resonant background (N_B). The signal strength of the Z boson μ , defined as the ratio of the measured cross section to the SM prediction, $\mu = \sigma^{\text{Measured}}/\sigma^{\text{Predicted}}$, is the parameter of interest (POI) of the fit. The probability of the Z boson, W boson, $t\bar{t}^{\text{Top}}$ and $t\bar{t}^{\text{Other}}$ process is modelled as $\mathcal{S}_Z(\vec{m}|\vec{\theta})$, $\mathcal{S}_W(\vec{m}|\vec{\theta})$,

$\mathcal{T}_{t\bar{t}}^{Top}(\vec{m}|\vec{\theta})$, $\mathcal{T}_{t\bar{t}}^{Other}(\vec{m}|\vec{\theta})$, respectively. The non-resonant background is described as $\mathcal{B}(\vec{m}|\vec{p})$, where $\vec{p} = \{p_1, \dots\}$ are coefficients of the background function and are kept floating in the fit. $S = \{s_1, \dots\}$ is a set of systematic uncertainties, and $\mathcal{G}(0|\theta_s, 1)$ is the constraint of each systematic uncertainty.

4.3.6.1 Treatment of systematic uncertainties

Each systematic uncertainty affects the normalisation, peak and/or resolution of the jet mass distribution described by the DSCB function. The effect is evaluated by: $\frac{(Val_{syst} - Val_{nominal})}{Val_{nominal}}$. In the fit model, all effects are considered simultaneously with the following exceptions:

- Only effect on the peak and the resolution of the jet mass distribution are considered for the modelling systematic uncertainty of the $Z + \text{jets}$ process from the alternative generator.
- Only effect on the normalisation is considered for the scale uncertainties and PDF+ α_s uncertainty of the $Z/W + \text{jets}$ process.
- Only effect on the normalisation is considered for the PS and ME uncertainties of $t\bar{t}$.

4.3.7 Results

Figure 4.10 shows the post-fit plots for the 3 different p_T bins defined for the calibration. Globally, the pulls between the fitted model and the data are smaller than 3, which means that the fit model describes the data distribution well. The impact of each systematic uncertainty in the fitted μ is estimated by fixing the corresponding nuisance parameter to its best-fit value $+/-1\sigma$ and repeating the fit. The difference between the new fitted μ with the nominal fitted μ is taken as the systematic uncertainty. The total uncertainty of the fitted μ is the sum in quadrature of the impact of all the systematic uncertainties. The fit results are summarised in Table A.1. Figure 4.11 shows the pulls and ranking of the impact from each systematic uncertainty. The leading systematic uncertainty is JMR, whose impact can be up to 20% according to the p_T bins. It is due to the fact that the JMR uncertainty is overestimated in the jet in-situ calibration.

4.4 Measurement of $\mu^{pre-tag}$

The $Z \rightarrow \ell\ell + \text{jets}$ channel is used to measure $\mu^{pre-tag}$ since it is less contaminated by backgrounds. The signal events are $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$, the main background contributions are from WZ and ZZ .

4.4.1 Data and Monte Carlo simulations

The dataset used in the measurement of $\mu^{pre-tag}$ is the same as the one used for $\mu^{post-tag}$.

The $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events used to measure $\mu^{pre-tag}$ are generated using SHERPA 2.2.8. The $Z \rightarrow \ell\ell$ samples are generated with the same QCD accuracy as for $Z \rightarrow$

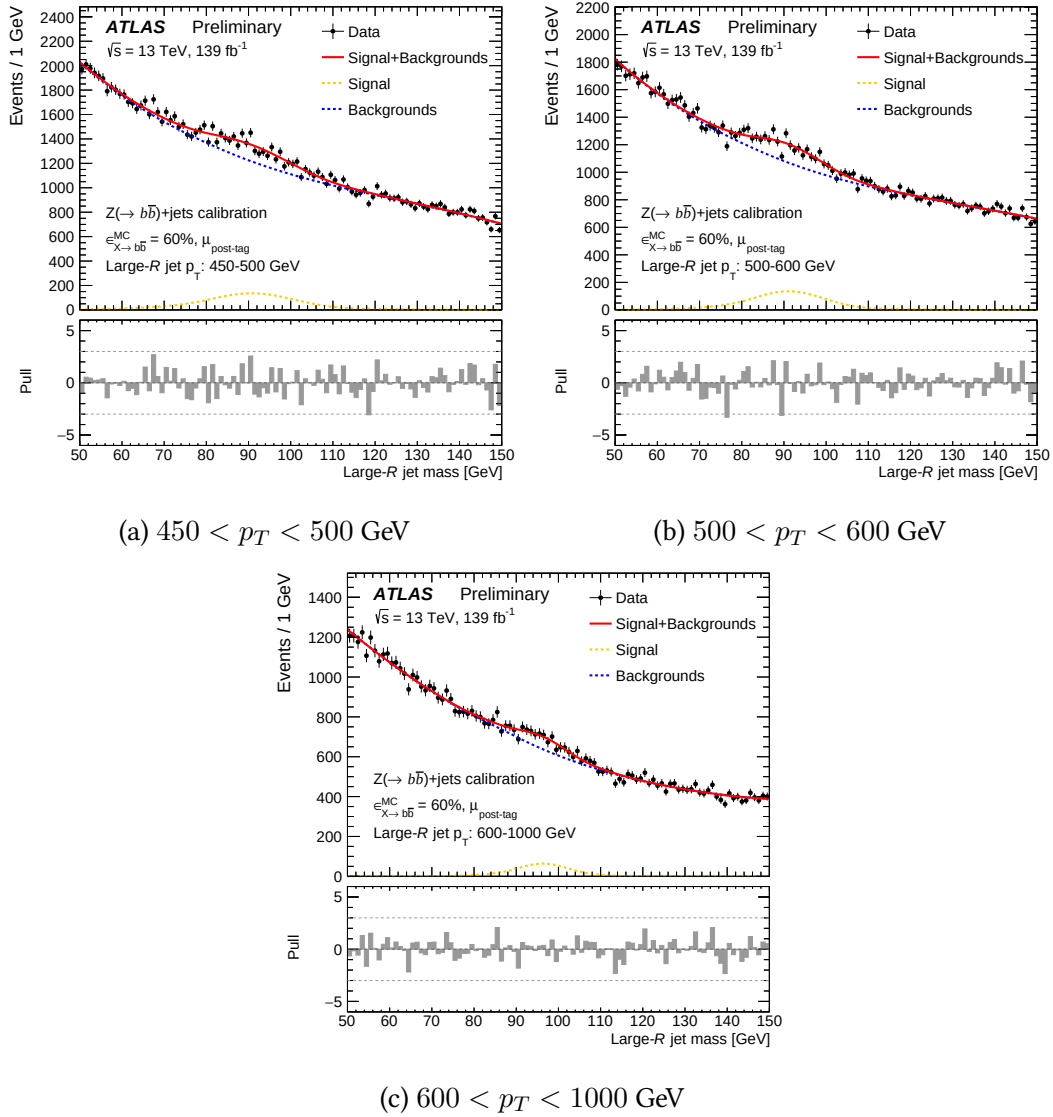


Figure 4.10: The Z boson candidate jet mass distribution after 60% WP of the $X \rightarrow b\bar{b}$ tagger for events with the large- R jet in the $450 < p_T < 500$ GeV (a); $500 < p_T < 600$ GeV (b) and $600 < p_T < 1000$ GeV (c) range. The dashed yellow line represents the Z + jets signal process. The dashed blue line represents the sum of all background processes. The solid red line is the sum of all signal and background processes.

$q\bar{q}$ samples. An on-shell Z -decay model is used. The difference created by the different Z -decay models is negligible. The alternative $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ events are generated with MADGRAPH using a multileg LO approach up to 4 additional partons. PYTHIA 8.230 is used in the parton shower.

The diboson background consists of the final states arising from WZ and ZZ events. The nominal MC sample for $q\bar{q}$ -initiated diboson processes are generated by SHERPA 2.2.1.

muons are required to pass impacter parameter selections for e (μ): $|d_0| < 5(3) \text{ mm}$ and $|\Delta z_0 \sin \theta| < 0.5 \text{ mm}$. The *medium* identification working point and the *tight* isolation working point are used for electrons and muons. At least two leptons with the same flavor are required in the events. The overlap removal procedure is applied as described in Section 3.7. The events are required to have at least one trimmed LCTopo large- R jets with $p_T > 200 \text{ GeV}$.

4.4.2.1 Reconstruction of $Z \rightarrow \ell\ell$

$Z \rightarrow \mu^+\mu^-$ is reconstructed using the two leading muons with $p_T > 27 \text{ GeV}$. The charge of the two leading muons is opposite. $Z \rightarrow e^+e^-$ is reconstructed by the two leading electrons with $p_T > 25 \text{ GeV}$. No requirement is applied on the charge of electrons due to a high rate of charge misidentification. An additional p_T balance requirement on leptons is applied: $\frac{(p_T^{l_1} - p_T^{l_2})}{p_T^{ll}} < 0.8$. $p_T^{l_1}$ and $p_T^{l_2}$ are the transverse momentum of the leading and subleading leptons, respectively. p_T^{ll} is the transverse momentum of the Z boson reconstructed using the two leading leptons.

The 4-vector of the reconstructed Z boson is calculated by the sum of the 4-vector of the two leading leptons. The reconstructed Z is required to have: $66 < m_{ll} < 116 \text{ GeV}$ and $p_T^{ll} > 450 \text{ GeV}$. In analogy with the $\mu^{post-tag}$ selection, the following selections are also applied (p_T^{leadj} is the transverse momentum of the recoil large- R jet):

- $p_T^{ll} > p_T^{leadj}$
- p_T asymmetry cut: $0 \leq \frac{p_T^{ll} - p_T^{leadj}}{p_T^{ll} + p_T^{leadj}} < 0.15$
- $|\Delta y_{ll-leadj}| < 1.2$

The validation of this method is based on the consistency of $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$. After all the selections in our analysis, the p_T distribution of $Z \rightarrow \ell\ell$ is consistent with that of $Z \rightarrow b\bar{b}$, as shown in Figure 4.12. There is a small slope in the ratio of $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$ in a function of p_T of the Z boson. For $p_T < 1000 \text{ GeV}$, the slope is considered negligible.

The main backgrounds are WZ and ZZ process. Figure 4.13 shows data/MC distribution using $Z \rightarrow \ell\ell$ channel for $450 < p_T < 500 \text{ GeV}$, $500 < p_T < 600 \text{ GeV}$ and $600 < p_T < 1000 \text{ GeV}$ respectively. There is a good agreement between data and MC. In the measurement of $\mu^{pre-tag}$, signals and backgrounds are modelled using the MC simulation and only the yields are needed according to Equation 4.4.

4.4.3 Systematic uncertainties

The following systematic uncertainties are considered in the measurement of $\mu^{pre-tag}$.

- Theoretical uncertainties: Three main systematic uncertainties are considered for $Z \rightarrow \ell\ell$: generator uncertainties (SHERPA 2.2.8 VS MADGRAPH +PYTHIA 8), scale uncertainties and PDF+ α_s uncertainties as described in Section 4.3.5 since the same MC setups are used as for hadronically decaying Z . The contribution of diboson is so small that its theoretical uncertainties are not considered.

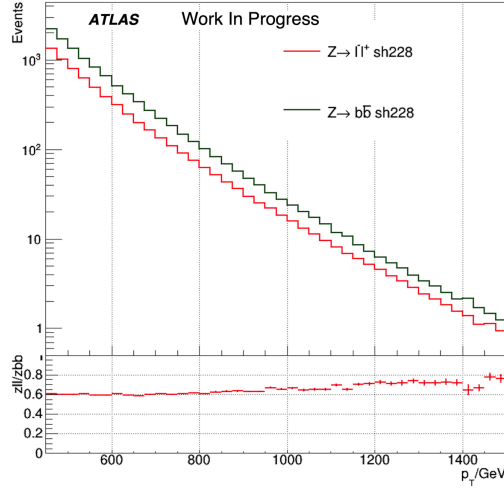


Figure 4.12: The p_T distribution of $Z \rightarrow \ell\ell$ (red) and $Z \rightarrow b\bar{b}$ (green) after the selections. The bottom panel shows the ratio of the event yields of $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$.

- Integrated luminosity and pileup reweighting uncertainty.
- Lepton-related systematic uncertainties: uncertainties in the trigger, reconstruction, identification and isolation efficiencies, and in the momentum of electrons [121] and muons [122] are considered. More details can be found in Section 3.5 for electrons and Section 3.6 for muons.
- Large- R jet related systematic uncertainties: For $\mu^{\text{pre-tag}}$, the large- R jet uncertainties are not dominant for the final $\mu^{\text{pre-tag}}$. However, it is important to study the impact of JES on the consistency of $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$. The ratio of $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$ varies within 5% in for $p_T < 1000$ GeV indicating a good consistency between $Z \rightarrow \ell\ell$ and $Z \rightarrow b\bar{b}$.

4.4.4 Results

Only the yields are needed according to Equation 4.4. The results of $\mu^{\text{pre-tag}}$ using inclusive $Z \rightarrow \ell\ell$ channel are shown in Table 4.5. The $\mu^{\text{pre-tag}}$ values are consistent with 1 considering their systematic uncertainties. The leading systematic uncertainties for $\mu^{\text{pre-tag}}$ is from the theoretical uncertainties (Z -modelling) where an alternative generator samples and scale uncertainties and PDF+ α_s uncertainties are contributed.

4.5 Calibration results: scale factors

The scale factor is calculated by Equation 4.3. The propagation of each systematic uncertainty is calculated as:

$$\delta_{SF}^{\text{syst}} = \frac{\delta_{\mu^{\text{post-tag}}}^{\text{syst}} \mu^{\text{pre-tag}} - \delta_{\mu^{\text{pre-tag}}}^{\text{syst}} \mu^{\text{post-tag}}}{\mu^{\text{pre-tag}} \mu^{\text{pre-tag}}} \quad (4.11)$$

The values for $\mu^{\text{pre-tag}}$, $\mu^{\text{post-tag}}$ and the scale factors are summarised in Table 4.6. This

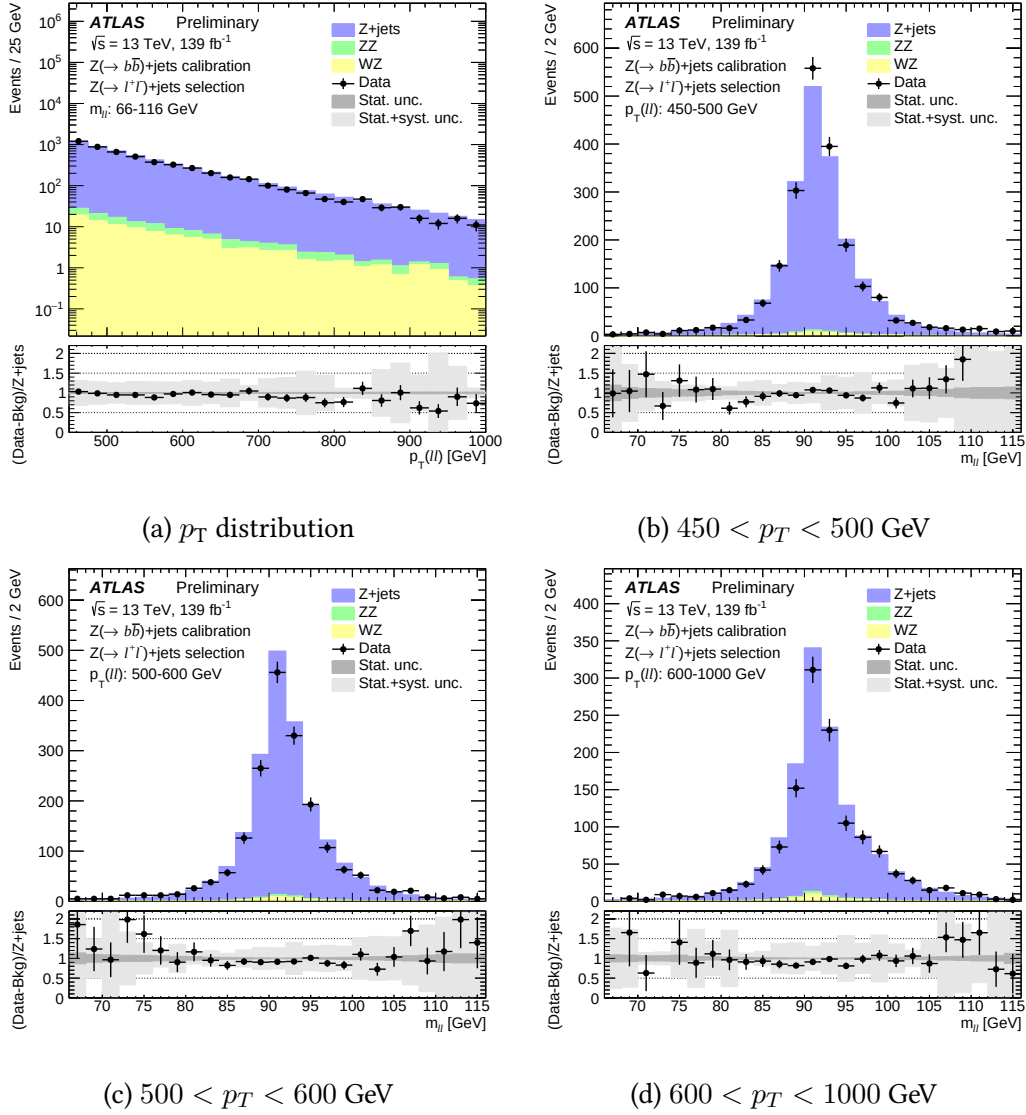


Figure 4.13: The $Z \rightarrow \ell\ell$ p_T distribution (a) and mass distribution in three Z -boson candidate p_T bins: $450 < p_T < 500$ GeV (b), $500 < p_T < 600$ GeV (c) and $600 < p_T < 1000$ GeV (d). All distributions are shown after applying the event selection for the measurement of $\mu^{\text{pre-tag}}$

table includes also the systematic uncertainties of the scale factors. The leading systematic uncertainties for $Z + \text{jets}$ are from the fit model, the $Z + \text{jets}$ modelling and the statistical uncertainty. The fit model uncertainty is a special uncertainty obtained by varying the fitted mass region in the measurement of $\mu^{\text{post-tag}}$. In addition, the scale factors depend on the p_T and plotted in Figure 4.14. The SF for the first p_T bin is from the measurement using $Z\gamma$ events to cover the low p_T regime.

4.6 Moving to a binned template fit

A binned maximum likelihood fit is an attractive option since we have large datasets. However, the widely used statistical modelling package RooFit [143] was strongly biased when

Table 4.5: $\mu^{\text{pre-tag}}$ for differential p_T using inclusive $Z \rightarrow \ell\ell$ channel.

p_T	$450 < p_T < 500 \text{ GeV}$	$500 < p_T < 600 \text{ GeV}$	$600 < p_T < 1000 \text{ GeV}$
$\mu^{\text{pre-tag}}$	1.014	0.939	0.931
Uncertainties ($+\sigma/-\sigma$)			
Statistical	0.018	0.019	0.025
Z modelling	+0.179/-0.176	+0.151/-0.147	+0.179/-0.176
e-related	+0.038/-0.038	+0.037/-0.037	+0.043/-0.043
μ-related	+0.031/-0.039	+0.029/-0.035	+0.032/-0.036
JMS	+0.002/-0.001	0.000/0.000	0.000/0.000
JMR	0.000/0.000	0.000/0.000	0.000/0.000
JES	+0.039/-0.039	+0.037/-0.036	+0.040/-0.038
Others	+0.017/-0.017	+0.017/-0.017	+0.017/-0.017
Total uncertainty	+0.191/-0.190	+0.164/-0.162	+0.192/-0.191

 Table 4.6: Pre-tag ($\mu^{\text{pre-tag}}$) and post-tag ($\mu^{\text{post-tag}}$) signal strength and the resulting signal efficiency scale factors (SF) for the $X \rightarrow b\bar{b}$ tagger at 60% efficiency working point measured using the $Z \rightarrow b\bar{b}$ calibration methods. Systematic uncertainties on the scale factor measurement are listed as well.

Calibration p_T [GeV]	$Z(\rightarrow b\bar{b})\gamma$ 200 – 450	$Z(\rightarrow b\bar{b}) + \text{jets}$		
	200 – 450	450 – 500	500 – 600	600 – 1000
$\mu^{\text{post-tag}}$	1.33	1.32	1.10	0.51
$\mu^{\text{pre-tag}}$	0.92	1.01	0.94	0.93
SF	1.45	1.30	1.17	0.55
Uncertainties ($\pm\sigma$)				
Statistical	± 0.37	± 0.18	± 0.13	± 0.09
Z -boson modelling	$^{+0.24}_{-0.19}$	–	–	–
$Z + \text{jets}$ modelling	–	$^{+0.21}_{-0.28}$	± 0.15	± 0.18
Fit model	0.14	0.39	0.22	0.16
Spurious signal	± 0.26	± 0.11	± 0.07	± 0.07
Other background modelling	± 0.05	± 0.03	± 0.02	± 0.01
Lepton & Photon related	± 0.02	$^{+0.06}_{-0.07}$	$^{+0.06}_{-0.07}$	± 0.03
Jet mass scale	± 0.05	$^{+0.02}_{-0.01}$	± 0.01	$^{+0.02}_{-0.01}$
Jet mass resolution	$^{+0.03}_{-0.02}$	$^{+0.22}_{-0.15}$	$^{+0.11}_{-0.09}$	$^{+0.09}_{-0.07}$
Jet energy scale	$^{+0.06}_{-0.07}$	± 0.09	± 0.09	± 0.05
Others	$^{+0.14}_{-0.16}$	± 0.01	$< \pm 0.01$	$< \pm 0.01$
Total uncertainty	$^{+0.53}_{-0.56}$	$^{+0.55}_{-0.56}$	$^{+0.35}_{-0.34}$	$^{+0.29}_{-0.28}$

computing the likelihood of continuous PDFs in bins. That is why the unbinned maximum likelihood fit has been used to measure $\mu^{\text{post-tag}}$. The bias has been fixed recently [144] by introducing a new PDF class to RooFit. When using a binned maximum likelihood fit, the $Z \rightarrow q\bar{q}$, $W \rightarrow qq$ and $t\bar{t}$ are modelled using histograms instead of functional form. Only one template is enough to describe the $t\bar{t}$ process, and it is not necessary to validate the

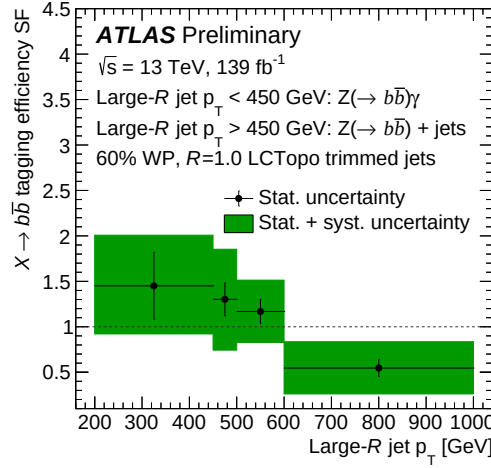


Figure 4.14: Signal efficiency scale factors for $X \rightarrow b\bar{b}$ tagger at 60% efficiency working point.

functional form the WP is changed. I have repurposed the fit framework in order to use a binned maximum likelihood fit. The binned fit framework has been validated, and the scale factors have been measured at 50%, 60% and 70% WP. They have been released as official recommendation to the ATLAS physics analyses. Table 4.7 shows the measured SFs at 60% WP with the binned template fit framework. The fits for the $\mu^{post-tag}$ measurements have been performed with 2 GeV/bin. The 4th order polynomial function was chosen via a F-test for all p_T bins at 60% WP. The major systematic uncertainties are the Z + jets modelling and the spurious signal uncertainties.

Table 4.7: Scale factors derived for the signal efficiency of the $X \rightarrow b\bar{b}$ tagger at 60% WP using the binned template fit framework.

Calibration	$Z(\rightarrow b\bar{b}) + \text{jets}$		
p_T [GeV]	450 – 500	500 – 600	600 – 1000
$\mu^{post-tag}$	1.24	1.33	0.97
$\mu^{pre-tag}$	1.00	0.94	0.90
SF	1.24	1.42	1.07
Uncertainties ($\pm\sigma$)			
Statistical	± 0.20	± 0.18	± 0.19
Total uncertainty	$+0.43$ -0.49	$+0.30$ -0.37	$+0.31$ -0.44

4.7 Summary and outlook

After establishing the methodology, the $X \rightarrow b\bar{b}$ tagger has been calibrated for the first time. The SFs have been reported for the benchmark 60% WP. The signal efficiency scale factors vary between $1.30^{+0.55}_{-0.56}$ and $0.55^{+0.29}_{-0.28}$ for $450 < p_T < 1000$ GeV when using the unbinned fit framework. The signal efficiency scale factor vary between $1.42^{+0.43}_{-0.49}$ and $1.07^{+0.31}_{-0.44}$ when using the binned fit framework. A few improvements can be made in future studies:

- **MC samples** The SHERPA 2.2.11 [145] MC generator is now available. It is shown to improve the description of data for various measurements and it provides additional higher-order electroweak corrections. I have contributed to the production and validation of the SHERPA 2.2.11 hadronic $V + \text{jets}$ samples. Comparison studies of SHERPA 2.2.8 and SHERPA 2.2.11 has been briefly reported in Appendix B. The new samples are ready for the calibration studies.
- **Selection** Some truth-level studies on the large- R jet candidates have been performed to understand the topologies, see Figure 4.15. The truth labelling strategy and definition are detailed in Section 3.3.5. In the $Z(\rightarrow b\bar{b}) + \text{jets}$ process, the $Z \rightarrow b\bar{b}$ could be either the leading large- R jet or the sub-leading large- R jet. Since we only considered the leading large- R jet, we were losing 40% of our signal events. A similar observation was made for other processes. Thus, the contribution of the sub-leading large- R jet can be added through additional requirements on the candidate assignment in future studies.

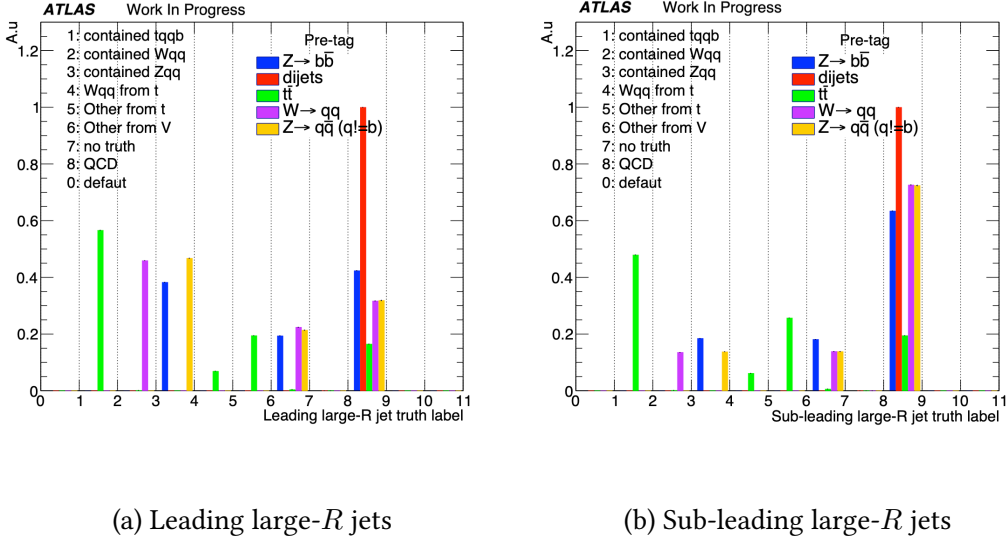


Figure 4.15: Truth label of the leading large- R jets (left) and the sub-leading large- R jets (right) in $Z \rightarrow b\bar{b}$, $Z \rightarrow q\bar{q}$ ($q = c, l$), $W \rightarrow q\bar{q}$, dijet and $t\bar{t}$ samples. The truth label definition is detailed in Table 3.1.

Many analyses are interested in using this tagger. For example, in the recently published research work for a heavy resonance particle decaying into a SM Higgs boson and a new particle in a fully hadronic final state, the $X \rightarrow b\bar{b}$ tagger is used for the Higgs boson identification and the scale factors derived using $Z + \text{jets}$ method are applied [11]. In the analysis presented in this thesis (see Chapter 5), the $X \rightarrow b\bar{b}$ tagger is also used and its scale factors are derived according to the needs of the analysis. Furthermore, the methodology and the calibration framework can be easily repurposed for other studies:

- **$V + \text{jets}$ calibration for boosted W/Z taggers** The current signal efficiency calibration for W/Z taggers uses W bosons from $t\bar{t}$ events. However, the statistical error for the highest p_T bin [350, 600] GeV is quite large, around 50% [146]. The use of $V + \text{jets}$ events could improve the SFs precision at high p_T . The $Z + \text{jets}$ calibration

framework has been modified by changing the taggers and the selection for $V + \text{jets}$. The studies are ongoing for the W/Z taggers of the UFO jets (See 3.3.1).

- **The double b -tagging method: close-by effect** As we introduced in Section 4.1.1, the calibrations used in the double b -tagging method are based on the isolated VR track jets. Since the b -jets from the Higgs bosons or Z bosons are collimated at the high p_T regimes, the validity of the single/isolated b -tagging calibrations is not ensured and the close-by effect needs to be considered. The $Z + \text{jets}$ calibration framework is used to evaluate the close-by effect in some analyses.

MEASUREMENT OF THE $V(\rightarrow qq')H(\rightarrow b\bar{b})$ PRODUCTION AT HIGH TRANSVERSE MOMENTUM

This chapter presents the measurement of the Higgs boson production associated with a vector boson using the fully hadronic $V(\rightarrow qq')H(\rightarrow b\bar{b})$ final state at high transverse momentum. This is the first time ever this analysis is being performed in ATLAS. The vector boson and the Higgs boson are reconstructed using large- R jets. The analysis is ongoing. The signal region is still blinded. Although no final results can be presented, the strategy and analysis optimisation are documented in this chapter. Section 5.1 gives an overview of the Higgs boson measurements using $H \rightarrow b\bar{b}$ and motivations for studying the fully hadronic $V(\rightarrow qq')H(\rightarrow b\bar{b})$ final state. Section 5.2 describes briefly the analysis strategy. Section 5.3 presents the data and MC samples used in this analysis. Section 5.4 presents the event selection, including the MC significance studies. Section 5.5 shows the calibration of the $X \rightarrow b\bar{b}$ tagger for this analysis. An overview of the background estimation is provided in Section 5.6, including the side-band and Asimov fit results. The systematic uncertainties are described in Section 5.7. Finally, a summary and outlook are included in Section 5.8.

5.1 Overview

As introduced in Section 1.1.4, the Higgs boson results from the Brout-Englert-Higgs mechanism in the SM where it is needed to give mass to the W and Z gauge bosons and fermions. During LHC Run 1, the Higgs particle was discovered by the ATLAS and CMS collaborations [6, 7] with a mass of 125.09 ± 0.24 GeV [147].

The channel $H \rightarrow b\bar{b}$, is the strongest proof of the coupling of the Higgs field to fermions at hadron colliders. Nevertheless, large backgrounds make isolating a Higgs boson signal in this channel quite challenging. The particle physicists did not observe it during LHC Run 1 due to insufficient statistics. In LHC Run 2, the centre-of-mass of pp collisions was increased to 13 TeV. In 2018, the $H \rightarrow b\bar{b}$ was finally observed in the associated production of the Higgs boson with a vector boson, W or Z [148]. In this analysis, $W \rightarrow \ell\nu$ and $Z \rightarrow \ell\ell$ were used for triggering. The backgrounds were efficiently rejected, and the signal

was purified. Higgs boson candidates were reconstructed from two b -tagged small- R jets. With this observation, a new era of detailed measurements in the Higgs sector opened up, through which the SM will be further challenged.

At the same time, the experimental techniques used for studies have been improved to increase the signal significance as much as possible. A large variety of techniques were adopted according to different production modes, kinematic phase space etc. The reconstruction of the $H \rightarrow b\bar{b}$ sector is always crucial, and it is mentioned briefly in the following. The details about jets can be found in Section 3.3. Details and discussions on the b -tagging and $H \rightarrow b\bar{b}$ tagging are presented in Section 3.4 and Section 4.1, respectively.

After the observation of $H \rightarrow b\bar{b}$ in 2018, the associated production of a Higgs boson with a W or Z boson decaying into leptons and the Higgs boson decaying to a $b\bar{b}$ pair, $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$, was measured in the resolved [149] and boosted [150] topologies. The measured signal strengths (see Equation 5.2) with the resolved and boosted topologies were:

$$\begin{aligned}\mu_{V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})}^{\text{resolved}} &= 1.17 \pm 0.16(\text{stat.})_{-0.16}^{+0.19}(\text{syst.}) \\ \mu_{V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})}^{\text{boosted}} &= 0.72_{-0.28}^{+0.29}(\text{stat.})_{-0.22}^{+0.26}(\text{syst.})\end{aligned}\quad (5.1)$$

$\mu_{V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})}^{\text{resolved}}$ and $\mu_{V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})}^{\text{boosted}}$ were consistent within systematic uncertainties. With the resolved topology, the Higgs boson candidate was reconstructed from two b -tagged small- R jets in the event. The $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ cross section could be measured for p_T^V as low as 75 GeV. However, due to the highly collimated b -jets at large transverse momentum, with the resolved topology, the Higgs boson candidate reconstruction efficiency was highly degraded, and the $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ cross section could only be measured inclusively at $p_T^V > 250$ GeV. On the contrary, in the boosted topology, the Higgs candidate was reconstructed by a large- R jet with two ghost-associated b -tagged VR track jets. The $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ cross section could be measured inclusively and differentially for $p_T^V > 250$ GeV. Figure 5.1 shows the measured $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ cross section with the resolved topology (Figure 5.1a) and the boosted topology (Figure 5.1b).

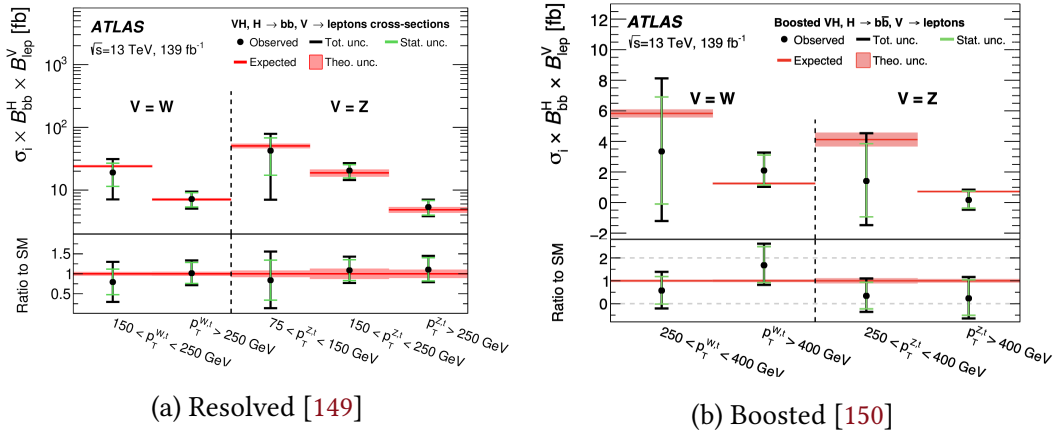


Figure 5.1: The measured $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ cross section with the resolved (left) and boosted (right) topologies in differential p_T bins.

The vector boson fusion production mode with $H \rightarrow b\bar{b}$ was also measured using

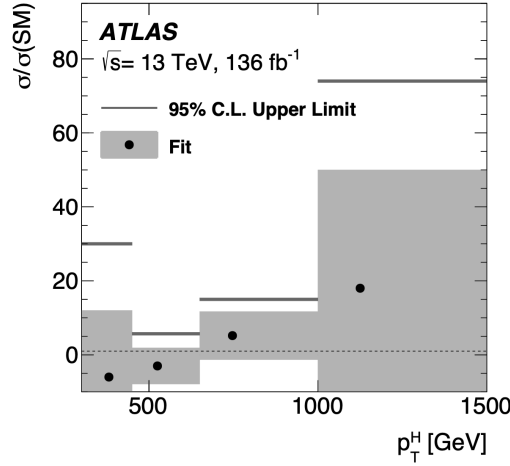


Figure 5.2: The signal strengths measured using the differential signal regions in the inclusive Higgs production with $H \rightarrow b\bar{b}$ in ATLAS [153].

full Run 2 dataset, the paper published in 2021 [151]. Higgs boson candidates were reconstructed from two b -tagged jets in the event as in the resolved $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$. The observed excess for the background-only hypothesis had a significance of 2.9 standard deviations (2.9 standard deviations expected). The measured signal strength was $\mu_{VBF} = 0.99^{+0.30}_{-0.30}(\text{stat.})^{+0.18}_{-0.16}(\text{syst.})$. This corresponded to an observed $\sigma_{VBF} \times B_{H \rightarrow b\bar{b}} = 2.16^{+0.67}_{-0.66}(\text{stat.})^{+0.40}_{-0.35}(\text{syst.})$ pb.

The production mode associated with top quarks with $H \rightarrow b\bar{b}$ was measured using full Run 2 dataset, a paper published in 2021 [152]. In the resolved category, Higgs boson candidates were reconstructed from two b -tagged jets in the event. In the boosted category, a Deep Neural Network (DNN) algorithm was used to identify $H \rightarrow b\bar{b}$ from the complex and highly collimated final products in this mode. The measured signal strength was $\mu_{t\bar{t}H} = 0.35^{+0.20}_{-0.20}(\text{stat.})^{+0.30}_{-0.28}(\text{syst.})$, corresponding to an observed (expected) significance of 1.0 (2.7) standard deviations with respect to the background-only hypothesis.

The inclusive Higgs production mode with $H \rightarrow b\bar{b}$ was also measured for $p_T^H > 250$ GeV using full Run 2 dataset, a paper published in 2021 [153]. The Higgs candidate was reconstructed by a large- R jet with two ghosted-associated b -tagged VR track jets. Differential measurements for $p_T^H > 450$ GeV were achieved. Figure 5.2 shows the measured signal strength. The 95% confidence-level upper limit on the cross section for the Higgs boson production with transverse momentum above 450 GeV was 115 fb, and above 1 TeV was 9.6 fb. The SM predictions in the same kinematic regions were 18.4 fb and 0.13 fb, respectively. A similar measurement for $p_T^H > 450$ GeV was also performed in the CMS experiment [154]; the inclusive Higgs boson signal strength was measured as $\mu_H = 3.7^{+1.2}_{-1.2}(\text{stat.})^{+0.6}_{-0.7}(\text{syst.})^{+0.8}_{-0.5}(\text{theo.})$, corresponding to an observed significance of 2.5 standard deviations with respect to the background-only hypothesis, while the expected significance of the SM signal was 0.7 standard deviations. The differential cross section, assuming that the other production modes occur at the SM rates, was measured. An excess was seen for $p_T^H > 650$ GeV with a local significance of 2.6 standard deviations with respect to the SM expectation.

The ATLAS and CMS collaborations also measured other properties of the Higgs boson, such as the decay rates, the coupling to bosons and fermions. So far, all measurements agreed with the predictions for a SM Higgs boson within uncertainties. As a result, particle physicists have been looking for new ideas and techniques to further improve the knowledge in the Higgs sector.

The study documented in this chapter is based on the SM Higgs boson produced in association with a vector boson with such a final state that the vector boson decays to two quarks and that the Higgs boson decays to a pair of b -quarks. This measurement is carried out for the first time in ATLAS. The measurement provides complementarity to the $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ analysis by extending to the higher Higgs momentum. Furthermore, as introduced in Section 1.2.3, the sensitivity of probing new physics at a high p_T^H regime is enhanced from both the theoretical and experimental points of view. As shown in the studies presented in the previous paragraphs, when the boosted techniques are used for high p_T^H , we can gain more sensitivity. The cross section measurement of $V(\rightarrow qq)H(\rightarrow b\bar{b})$ will be performed within the Simplified Template Cross Sections (STXS) framework [155]. The measurement will then be used to constrain anomalous couplings in a Standard Model Effective Field Theory (SMEFT) [156].

The expected signal significance estimated by MC is around 1.28 (see Section 5.4.3). The dominant background in this analysis is represented by the QCD dijet process, which is difficult to simulate. Thus, the dijet is estimated in an entirely data-driven way. The selected events are subdivided into p_T^H categories to perform both the inclusive measurement and the measurement as a function of p_T . The plan is to use a binned likelihood fit to the Higgs-candidate jet mass distribution to extract the signal.

5.2 Analysis strategy

The measurement will be performed in different stages.

The first stage is to maximise the overall sensitivity to the Higgs boson signal. The event selection is optimised to maximise the signal significance using MC simulation. For a given final state, for example, the fully hadronic $V(\rightarrow qq)H(\rightarrow b\bar{b})$ final state, the signal strength is defined as:

$$\mu(V(\rightarrow qq)H(\rightarrow b\bar{b})) = \frac{[\sigma^{VH} \times BR(H \rightarrow b\bar{b}) \times BR(V \rightarrow qq)]_{measured}}{[\sigma^{VH} \times BR(H \rightarrow b\bar{b}) \times BR(V \rightarrow qq)]_{predicted}} \quad (5.2)$$

The numerator is extracted from the measurement. The value μ shows us how large the deviation is from the SM prediction. The measurement of μ relies on the selection of the analysis region. The uncertainty of μ is highly dependent on the uncertainty associated with the signal process. The measurement will be performed in inclusive and differential binning of the reconstructed jet p_T : $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV.

Then, a STXS measurement will be performed. The STXS is a common framework for the Higgs cross section measurements for the LHC experiments [155]. It is a compromise between experimental sensitivity and independence from theoretical assumptions. It is

used to combine the measurements of ATLAS and CMS and to interpret the LHC Higgs data. The STXS measurements are physical cross section measurements with the phase spaces that are agreed upon commonly across experiments.

In the STXS framework, the Higgs boson signal is split according to production modes. The different production modes are further split into differential regions, i.e. STXS 'bins'. The binning is chosen to have a flat experimental acceptance within each bin. The binning is done in a variety of variables, such as the p_T of the Higgs boson, the p_T of the vector boson for $V(\rightarrow leptons)H(\rightarrow b\bar{b})$, or the invariant dijet mass for VBF topologies. For the fully hadronic VH measurement, the STXS measurement is binning on truth p_T of the Higgs boson: $300 < p_T^{\text{truth}}(H) < 450$ GeV, $450 < p_T^{\text{truth}}(H) < 650$ GeV and $p_T^{\text{truth}}(H) > 650$ GeV. All events with $p_T^{\text{truth}}(H) < 300$ GeV will be fixed to the SM prediction within uncertainties. The migration of the reconstructed variables like p_T^H to the truth variables like $p_T^{\text{truth}}(H)$ will be needed. If there is no specific mention in the thesis, the variables used in this chapter are reconstructed.

Afterwards, the STXS measurement can be used for several interpretations. In this analysis, the STXS measurement will be used to constrain anomalous couplings in a Standard Model Effective Field Theory (SMEFT) [157]. As introduced in Section 1.2.3, the sensitivity of new physics is larger with SMEFT with respect to the SM. In the SMEFT, the SM Lagrangian is extended with higher-dimensional operators, in the first approximation dimension six. The operators affecting the signal process will be used. The Higgs boson production cross section in the STXS can be corrected by adding their contributions to the SM one. The STXS and the Higgs boson decay rates will be parameterised using the Wilson coefficients [158].

5.3 Data and MC simulations

The full Run 2 dataset collected by the ATLAS detector during 2015-2018 is used in this work.

The signal is the Higgs boson production associated with a vector boson (VH). It is simulated using POWHEG-Box V2 [132] at NLO in QCD and the NNPDF3.0 [130] NLO PDF. The loop-induced $gg \rightarrow ZH$ process is generated separately at LO with POWHEG-Box. In all cases, the events are showered using PYTHIA 8.240 with the AZNLO tune [159] and the CTEQ6L1 PDF [160]. The $q\bar{q}/qg \rightarrow VH$ cross section is calculated at NNLO in QCD with NLO electroweak corrections. The $gg \rightarrow ZH$ cross section is calculated at NLO and next-to-leading-logarithm (NLL) accuracy [161, 162]. The decays of bottom and charm hadrons are performed by EvtGen [135]. The Higgs boson events from gluon-gluon fusion production (ggF), vector boson fusion (VBF) and in association with two top quarks ($t\bar{t}H$) are generated using POWHEG-Box V2 [132], and as with signal are interfaced to PYTHIA 8.240 with the AZNLO tune and the NNPDF3.0 NLO PDF.

The predicted cross sections used to normalise simulated SM events for different Higgs production modes are shown in Table 5.1.

The dominant background in this analysis is dijet, $t\bar{t}$ and V + jets. The same simulated samples are used as in Section 4.3.1.2. The diboson processes (VV) also have minor contributions; the same simulation setup is used as for the V + jets processes.

Table 5.1: Cross section predictions used to normalise the simulated SM events for the ggF [163, 164], VBF [165], VH [161, 162], $t\bar{t}H$ [166], tH [167] and the SM fraction of each production mode.

Process	Cross section [fb]	Fraction [%]
ggF	48.58	87.2
VBF	3.782	6.8
VH	2.257	4.0
$t\bar{t}H + tH$	0.599	1.1

5.4 Event selection

The events are required to have no isolated muons and electrons. The large- R jets used in this work are **the trimmed LCTopo jets** as introduced in Section 3.3. At least two large- R jets are required to be present in the events. The leading large- R jet is required to have p_T larger than 450 GeV and mass larger than 50 GeV so that the efficiency of the triggers is higher than 99%. The subleading large- R jet is required to have p_T larger than 200 GeV and mass larger than 40 GeV to be in the validity region of the large- R jet calibration.

5.4.1 VH assignment

The identification of W/Z boson relies on several properties of the large- R jets. For the W/Z tagger [146] used in this study, the discriminants are:

- Combined mass of the large- R jets.
- $D_2^{\beta=1}$. It exploits the two-body structure of the $W/Z \rightarrow qq'$ decays, absent from typical QCD jets.
- N_{trk} . It is the number of tracks matched to the large- R jets by ghost association. This quantity is higher for gluon-induced jets than quark-induced jets due to the distinct energy scales involved and the different colour factors for gluons and quarks.

The mass window and the upper cut on $D_2^{\beta=1}$ and n_{trk} are optimised depending on p_T to achieve 50% signal efficiency and to maximize the background rejection in each p_T bin. The background rejection factor is 70-200 depending on p_T .

The boosted $X \rightarrow b\bar{b}$ tagger is used to identify the Higgs bosons. More details about the $X \rightarrow b\bar{b}$ tagger are presented in Section 4.1.2. The working point used in this analysis is 60% as defined in Table 4.1.

Three ambiguity procedures were studied to determine the optimal choice based on the accuracy of the assignment and background modelling abilities.

- Scheme A - Vtag method:
We check first which one of the large- R jets is V -tagged; if the large- R jet is tagged and assigned as V candidate ($VCand$); the other large- R is considered to be the Higgs

boson candidate (*HCand*). If both large- R jets are V -tagged, the large- R jet mass is compared; the one with the larger mass is chosen as *HCand*. This method is NOT favoured because the background shape in the control region and validation regions is far from the one in the signal region.

- Scheme B - $X \rightarrow b\bar{b}$ tagger method:

This method relies on the $X \rightarrow b\bar{b}$ tagger and the D_{Xbb} defined by Equation 4.1 for each large- R jet. It selects the *HCand* as the large- R jet (of the leading two in the event) with the highest D_{Xbb} . This method is favoured for background modelling studies. However, only the $X \rightarrow b\bar{b}$ tagger WPs are calibrated, not the D_{Xbb} distribution. Using the uncalibrated D_{Xbb} distribution might create different assignment efficiency of VH between data and MC simulations.

- Scheme C - mixed method:

This method is a modified version of Scheme B to solve the uncalibrated D_{Xbb} issue. If no large- R jet passes the $X \rightarrow b\bar{b}$ tagger at 60% WP, the VH candidates are assigned as Scheme B. If only one large- R jet passes the $X \rightarrow b\bar{b}$ tagger at 60% WP, the passing one is assigned as *HCand*. If both large- R jets pass the $X \rightarrow b\bar{b}$ tagger, the one with higher mass is selected as *HCand*. *This method is adopted in this analysis.*

After the VH assignment, the Higgs boson candidate is required to have at least two ghost-associated VR track jets. The accuracy of the assignment, which is estimated as the fraction of the truth VH events with respect to the total signal events after the selection, is studied for all schemes. Truth VH events require that the *HCand* matches the truth Higgs boson geometrically and the *VCand* matches geometrically the truth vector boson. Table 5.2 shows the accuracy of assignment with different VH assignment schemes. Scheme C gives the best accuracy values.

Table 5.2: Accuracy of assignment with different VH assignment schemes.

	$250 < p_T^H < 450 \text{ GeV}$	$450 < p_T^H < 650 \text{ GeV}$	$p_T^H > 650 \text{ GeV}$
Scheme A	74.64%	67.12%	70.76%
Scheme B	75.68%	76.18%	75.20%
Scheme C	76.43%	76.24%	75.41%

Additionally, since the neutrinos and muons are not included in the large- R jet reconstruction and jet calibrations, the large- R jet energy is corrected to account for the semileptonic b -hadron decay, i.e. muon-in-jet correction. The used muons are required to pass a minimum set of quality criteria and to be found within $\Delta R = \min(0.4, 0.04 + 10/p_T^\mu)$ of a VR track jet associated with the large- R jet. The four-vector of the accepted muon with the largest p_T is added to the large- R jet. The correction significantly impacts the mass resolution [153]. We should note that the jet mass and p_T used for the event selection are without the muon-in-jet correction; however, this analysis is binned with muon-corrected p_T , and the final mass distribution is muon-corrected jet mass. Three p_T bins are used: $250 < p_T^H < 450 \text{ GeV}$, $450 < p_T^H < 650 \text{ GeV}$ and $p_T^H > 650 \text{ GeV}$.

5.4.2 Signal, control and validation region definition

The regions used in this analysis are defined according to the discriminants of the $X \rightarrow b\bar{b}$ tagger and the W/Z tagger. As shown in Figure 5.3, the 3-variables cuts are applied to the $VCand$ and the $X \rightarrow b\bar{b}$ tagger is applied to the $HCand$.

From the $VCand$ side, the events can be categorised as:

- **Pass V** events: All cuts on the W/Z 3-variables tagger are passed.
- **VR1** events: Only one cut on the the W/Z 3-variables tagger is failed. If the $D_2^{\beta=1}$ cut is failed, the event is categorised into VR1_failD2. A similar definition for VR1_failMass and VR1_failNtrk.
- **VR2** events: Only two cuts on the W/Z 3-variables tagger are failed. If the $D_2^{\beta=1}$ and mass cuts are failed, the event is categorised into VR2_failD2failMass. A similar definition for VR2_failNtrkfailMass and VR2_failD2failNtrk.
- **CR** events: All cuts on the the W/Z 3-variables tagger are failed.

From the $HCand$ side, the events can be classified as:

- **Pass Xbb** events: if the D_{Xbb} of the $HCand$ is larger than 2.44, i.e. 60% WP.
- **Fail Xbb** events: if the D_{Xbb} of the $HCand$ is smaller than 2.44, i.e. 60% WP.

The **signal region (SR)** events are the intersection of the Pass V and Pass Xbb events. The **$\overline{\text{SR}}$** events are the intersection of the Pass V and Fail Xbb events. The SR is used to extract the signal strength. The **control region (CR)** is used to study dijet modelling. The **validation regions (VR1 and VR2)** are used to validate the dijet modelling strategy. Figure 5.4 and 5.5 shows the mass and η distribution of the data and MC simulations of SR events and $\overline{\text{SR}}$ events. The MC simulations disagree with the data because it is difficult to simulate the dijet process. In this analysis, the dijet is modelled using a data-driven method. The mass and η distributions for other regions are shown in Appendix C.1.

5.4.3 Significance studies

The signal significance is calculated by using all bins in the mass distribution from 60 to 200 GeV with 5 GeV/bin, i.e. $\sqrt{\sum_{i=1}^{28} 2((s_i + b_i) \times \ln(1 + s_i/b_i) - s_i)}$ where s_i and b_i are the signal event yields and background events yields, respectively, in the i^{th} bin. Table 5.3 shows the signal significance in SR for differential and inclusive p_T bins for the fully hadronic VH analysis and the inclusive $H \rightarrow b\bar{b}$ analysis. The signal significance for $250 < p_T^H < 450$ GeV is the smallest. The signal significance for $450 < p_T^H < 650$ GeV is the largest among the differential p_T bins. The signal significance in the inclusive $p_T^H > 250$ GeV is 1.28. The significance reported in the inclusive $H \rightarrow b\bar{b}$ analysis is 2.40 for inclusive p_T^H [141]. The expected significance in the boosted $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ analysis is 2.73 for inclusive p_T^V , 2.16 for $450 < p_T^V < 400$ GeV and 2.00 for $p_T^V > 400$ GeV [150]. This underlines the importance of pursuing this analysis.

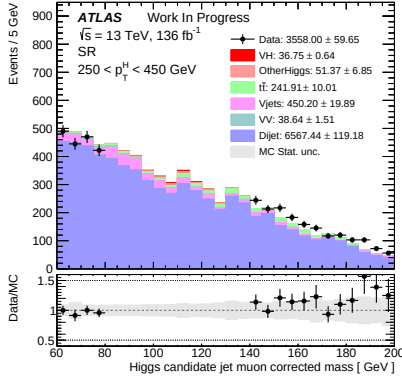
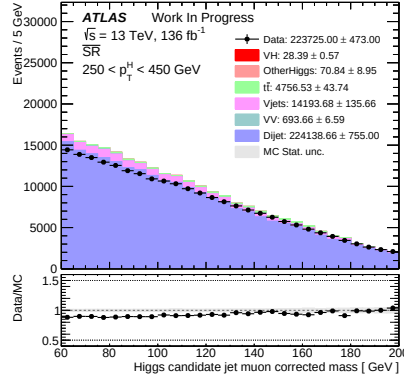
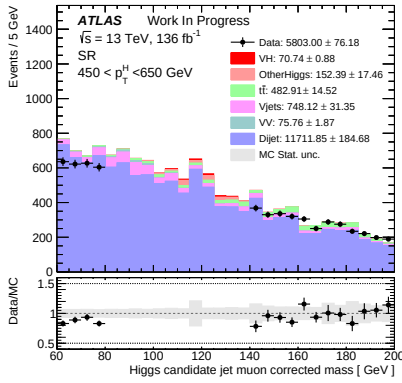
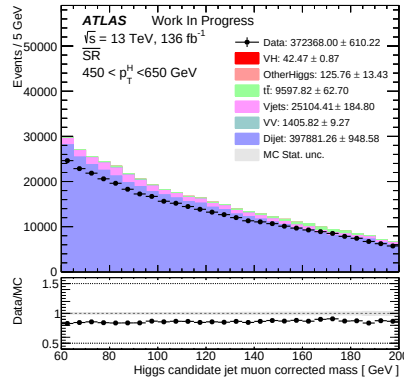
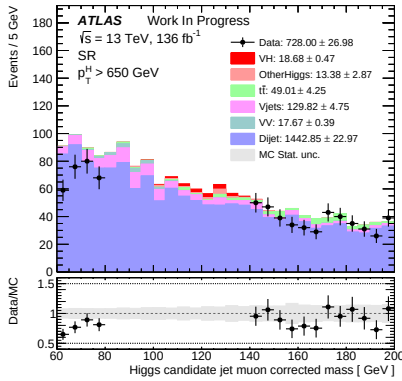
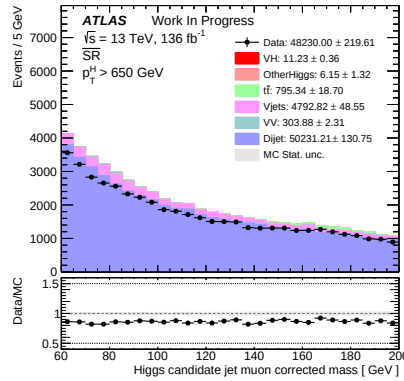
(a) $250 < p_T < 450 \text{ GeV}, \text{SR}$ (b) $250 < p_T < 450 \text{ GeV}, \overline{\text{SR}}$ (c) $450 < p_T < 650 \text{ GeV}, \text{SR}$ (d) $450 < p_T < 650 \text{ GeV}, \overline{\text{SR}}$ (e) $p_T > 650 \text{ GeV}, \text{SR}$ (f) $p_T > 650 \text{ GeV}, \overline{\text{SR}}$

Figure 5.4: The muon corrected large- R jet mass distribution of the Higgs candidate jet from Data and MC simulation for differential p_T in SR (left) and $\overline{\text{SR}}$ (right). The stacked plots are from the MC simulation. The distributions in black are from collision data. The event yields and its statistical uncertainties for each MC process and data are shown in the legend. The light grey box shows the MC statistical uncertainties. The bottom panels are the ratio of the event yields in the data and the MC prediction.

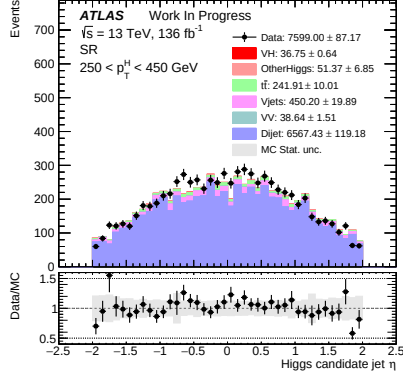
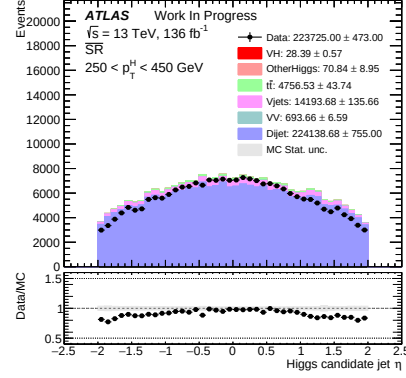
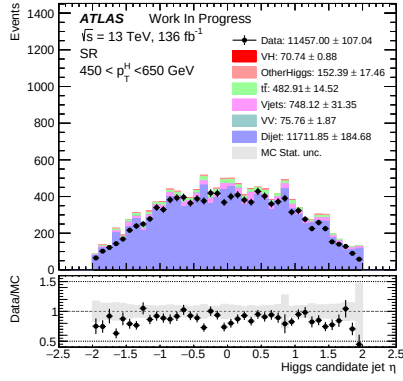
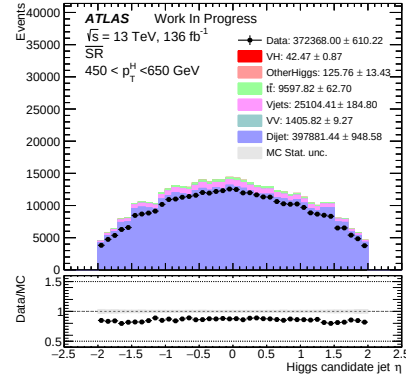
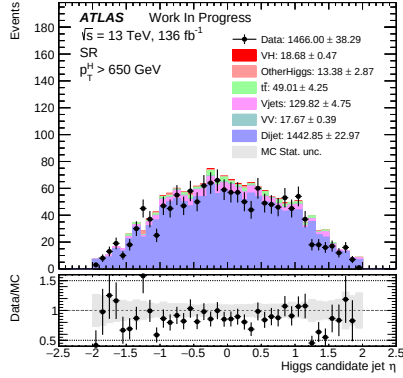
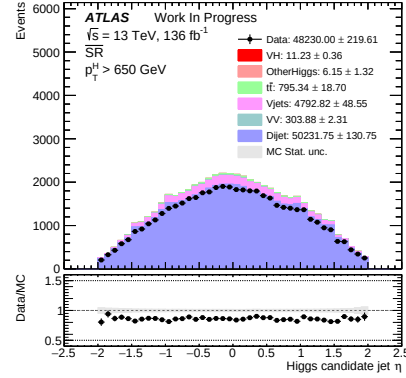

 (a) $250 < p_T < 450$ GeV, SR

 (b) $250 < p_T < 450$ GeV, $\overline{\text{SR}}$

 (c) $450 < p_T < 650$ GeV, SR

 (d) $450 < p_T < 650$ GeV, $\overline{\text{SR}}$

 (e) $p_T > 650$ GeV, SR

 (f) $p_T > 650$ GeV, $\overline{\text{SR}}$

Figure 5.5: The large- R jet η distribution of the Higgs candidate jet from Data and MC simulation for differential p_T in SR (left) and $\overline{\text{SR}}$ (right). The stacked plots are from the MC simulation. The distributions in black are from collision data. The event yields and its statistical uncertainties for each MC process and data are shown in the legend. The light grey box shows the MC statistical uncertainties. The bottom panels are the ratio of the event yields in the data and the MC prediction.

5.5 $X \rightarrow b\bar{b}$ tagger calibration

As the W/Z tagger and the $X \rightarrow b\bar{b}$ tagger are used in this analysis, calibrating those taggers is crucial. Semileptonic $t\bar{t}$ events are used for calibrating the W tagger [146] and then the scale factors are extrapolated to the Z tagger. For the $X \rightarrow b\bar{b}$ tagger, the signal efficiency is calibrated using two topologies. $Z\gamma$ events cover the p_T regime from 250 to 450 GeV, while $Z + \text{jets}$ events are used for $p_T > 450$ GeV. Since the tagging efficiency for $Z \rightarrow b\bar{b}$ and $H \rightarrow b\bar{b}$ is similar, the scale factors which are obtained using $Z + \text{jets}$ events can be propagated to the $H \rightarrow b\bar{b}$ events without corrections, see Figure 4.4. The mistag rate of top jets is calibrated using the semileptonic $t\bar{t}$ events. The $W + \text{jets}$ has a minor contribution in this analysis; the calibration of the mistag rate of the W jets is not needed. Since the dijet background is data-driven in this analysis, the $g \rightarrow b\bar{b}$ calibration is not needed either. More details are presented in [126]. However, as presented in Chapter 4, the triggers used in $Z + \text{jets}$ calibration are the same as in this analysis. Moreover, both studies require at least two large- R jets and very similar kinematic selection. Thus, the events used for the $X \rightarrow b\bar{b}$ calibration in Chapter 4 overlap highly with the events used in this analysis in SR. To avoid the overlap, an additional cut to the event selection presented in Chapter 4 is applied: *the events with the subleading jet mass in the mass region used for the W/Z tagger to tag the W/Z events, i.e. $[60, 105]$ GeV, are vetoed*; the scale factors are thus remeasured. The $\mu^{\text{post-tag}}$ and $\mu^{\text{pre-tag}}$ are measured using the same methodology presented in Chapter 4. The same data and MC simulated samples are used as in Chapter 4.

Measurement of $\mu^{\text{post-tag}}$ With this additional cut, 20% of events are removed. Figure 5.6 shows the comparison of jet mass distribution in data with and without the additional cut. The ratio is flat across the fitted mass region for the measurement of $\mu^{\text{post-tag}}$. Figure 5.7 shows the data and MC comparison for the large- R jet mass distribution in differential p_T bins. The binned template fit (see Section 4.7) is adopted with 5 GeV/bin. The $Z \rightarrow q\bar{q}$, $W \rightarrow qq$ and $t\bar{t}$ processes are described by the MC histograms. The functional form for the dijet background is optimised using the F-test. The 4th order polynomial function is chosen for all p_T bins. The spurious signals are also measured but not included in the fit model since this does not change the central value of $\mu^{\text{post-tag}}$. The same theoretical and experimental systematic uncertainties described in Chapter 4 are included. Figure 5.8 shows the post-fit plots. The fitted model describes well the data for all p_T bins. The jet response topology uncertainty, corresponding to the difference of the jet response modelling of hadronic decay from QCD, has the major contribution among all the systematic uncertainties included in the fit.

Measurement of $\mu^{\text{pre-tag}}$ In analogy with the $\mu^{\text{post-tag}}$ measurement, the additional cut is applied: *the events are vetoed if the leading large- R jet mass is in $[60, 105]$ GeV*. The other selections are the same as the one described in Section 4.4. The same systematic uncertainties are considered as in Section 4.4. Figure 5.9 shows the data and MC comparison after the $Z \rightarrow \ell\ell$ selection. The main systematic uncertainties for $\mu^{\text{pre-tag}}$ are from Z modelling, which shows the comparison between Sherpa simulation and Madgraph+Pythia simulation.

Scale factors The scale factors are finally computed, as shown in Table 5.4. They vary from 0.86 to 1.44 and follow the same trend as in the previous chapter. The dominant systematic uncertainties are from the spurious signal and the Z -boson modelling. The contribution of

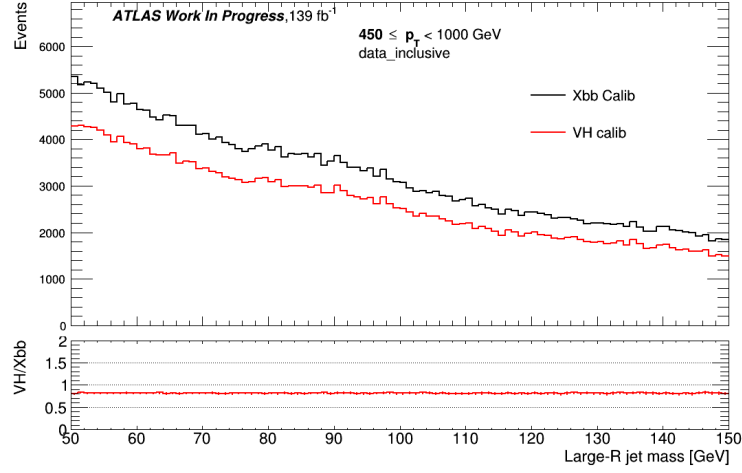


Figure 5.6: The black curve is the leading large- R jet mass distribution after the selections for the measurement of $\mu^{\text{post-tag}}$ in Chapter 4. The red curve is the the leading large- R jet mass distribution after the selections for the measurement of $\mu^{\text{post-tag}}$ in the calibration of $X \rightarrow b\bar{b}$ tagger for the $V(\rightarrow qq')H(\rightarrow b\bar{b})$ analysis. Compared to the black curve, an additional selection is applied to get the red curve: i.e. *the events with the subleading jet mass in $[60, 105]$ GeV are vetoed*. The p_T of the leading large- R jet is between 450 and 1000 GeV. The WP of the $X \rightarrow b\bar{b}$ tagger is 60%.

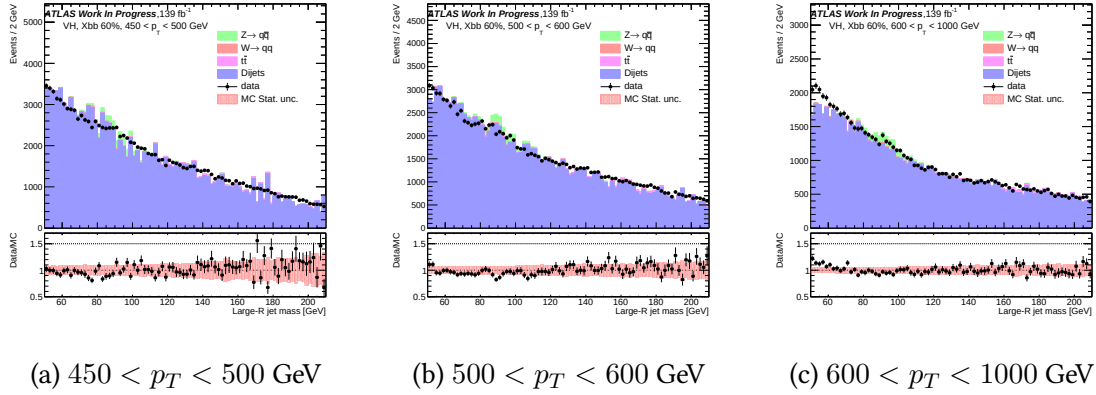


Figure 5.7: The leading large- R jet mass distribution after the selections for the measurement of $\mu^{\text{post-tag}}$ in the calibration of $X \rightarrow b\bar{b}$ tagger for the $V(\rightarrow qq')H(\rightarrow b\bar{b})$ analysis for $450 < p_T < 500$ GeV (a), $500 < p_T < 600$ GeV (b), $600 < p_T < 1000$ GeV (c). The 60% WP of the $X \rightarrow b\bar{b}$ tagger is used. Green: $Z \rightarrow q\bar{q}$; Pink: $W \rightarrow qq$; Purple: $t\bar{t}$; Blue: dijets. Black: data. Red: the statistical uncertainty band. The bottom panels are the ratio of the event yields from data and MC prediction.

the spurious signal uncertainty to the total uncertainty is larger than 50%. The breakdown of the systematic uncertainties can be found in Appendix C.2. The SFs are consistent with the ones measured using the binned template fit framework within uncertainties, see Table 4.7.

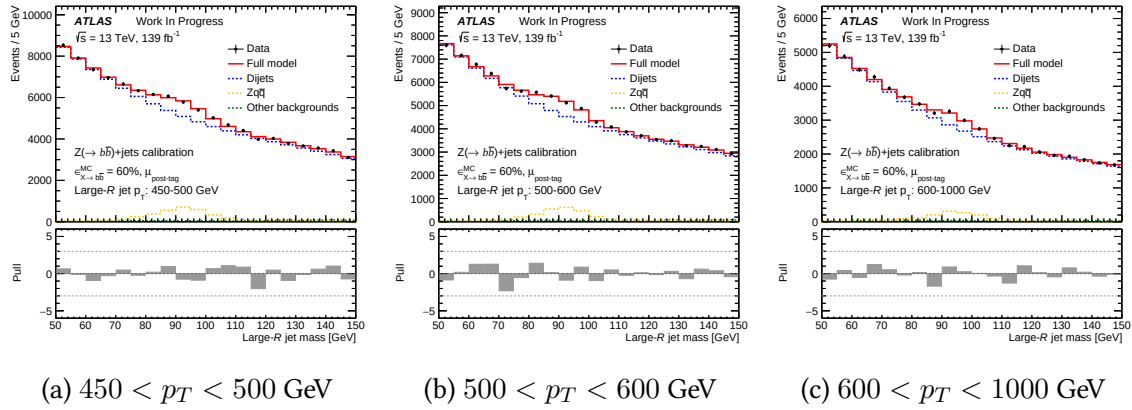


Figure 5.8: Post-fit leading large- R jet mass distribution after 60% WP of the $X \rightarrow b\bar{b}$ tagger by vetoing the events with the subleading jet mass in $[60, 105]$ GeV. Black dots: data. Red line: total post-fit distribution. Dashed blue line: post-fit dijet distribution. Dashed yellow line: post-fit signal distribution. Dashed green line: post-fit other backgrounds distribution.

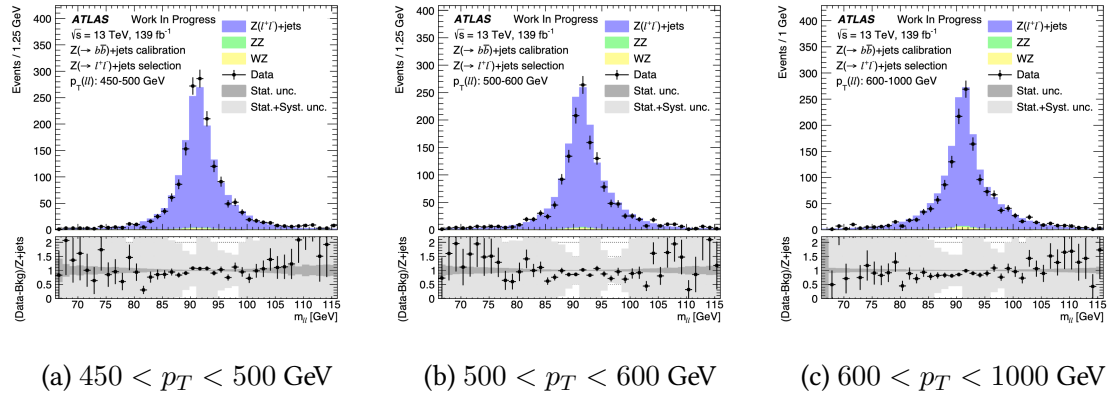


Figure 5.9: Data and MC comparison after $Z \rightarrow \ell\ell$ selection for differential p_T with additionally vetoing the events with the leading jet mass in $[60, 105]$ GeV. Black dots: data. Stacked plots: MC simulation. Dark grey band: statistical uncertainty. Light grey band: statistical and systematic uncertainty.

Table 5.4: Scale factors derived for the $X \rightarrow b\bar{b}$ tagger in fully hadronic VH analysis using Z + jets events.

Calibration p_T [GeV]	$Z(\rightarrow b\bar{b}) + \text{jets}$		
	450 – 500	500 – 600	600 – 1000
$\mu^{\text{post-tag}}$	1.43	1.19	0.78
$\mu^{\text{pre-tag}}$	0.99	0.92	0.91
SF	1.44	1.30	0.86
Uncertainties ($\pm\sigma$)			
Statistical	± 0.14	± 0.13	± 0.13
Total uncertainty	$+0.50$ -0.57	$+0.39$ -0.44	$+0.49$ -0.55

5.6 Background modelling

The dijet modelling is based on data because of the insufficient statistics and difficulties to model QCD in simulation. Two dijet modelling methods have been developed in this analysis. One is the boosted decision tree (BDT) reweighting method [168]. The idea is to get a SR dijet template shape from failXbb region to passXbb region using a multivariate method. The other method is the transfer factor (TF) method. The TF is the ratio between the number of the dijet events in SR and in $\overline{\text{SR}}$. This ratio is modelled by a function of large- R jet mass and p_T . The simultaneous fit to SR and $\overline{\text{SR}}$ is performed to extract the signal strength and the dijet distribution. The two methods are cross-checks, and several tests are defined to compare and understand their performance in dijet modelling. This section focuses on the BDT reweighting method, where I have made a major contribution. The TF method is briefly described in Appendix C.3.

5.6.1 BDT reweighting strategy

5.6.1.1 BDT reweighting basics

The BDT reweighting method is a Machine Learning (ML) technique for reweighting distributions. In the method, the space of variables used for reweighting is split into a few large regions by a decision tree instead of simply splitting each variable into several bins. The decision trees split the space of variables into regions by checking simple conditions. Each region is associated with some leaves of the tree [169]. The regions are obtained by maximizing the χ^2 defined as following:

$$\chi^2 = \sum_{\text{leaf}} \frac{(w_{\text{leaf},o} - w_{\text{leaf},t})^2}{w_{\text{leaf},o} + w_{\text{leaf},t}} \quad (5.3)$$

where $w_{\text{leaf},o}$ is the number of events in the original dataset in a leaf and $w_{\text{leaf},t}$ is the number of events in the target dataset in the same leaf. The prediction in the leaves is thus $w_{\text{leaf},o}/w_{\text{leaf},t}$. For events from the original dataset in the leaf, the dataset is reweighted with this ratio. The BDT reweighting uses many such trees. The prediction of the different trees is summed up.

In this analysis, the strategy is illustrated by Figure 5.10. The original dataset is from the failXbb data, and the target dataset is from the passXbb data. The BDT reweighting algorithm is trained with data passing the CR requirements and learns to reweight CR_failXbb large- R jet mass distribution to CR_passXbb. The training result is then applied to VR1 and VR2 for validations. Finally, a SR template is obtained by applying the training result to $\overline{\text{SR}}$. The training is performed separately in differential p_T bins: $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. This reweighting provides a shape for the dijet background while the normalisation of the template is obtained by:

$$\mu_{CR} = \frac{N_{CR}^{\text{passXbb}}}{N_{CR}^{\text{failXbb}}}, \quad N_{SR}^{\text{dijet}} = \mu_{CR} \times N_{\overline{\text{SR}}} \quad (5.4)$$

Where $N_{CR}^{passXbb}$ and $N_{CR}^{failXbb}$ are the number of events in CR_passXbb and CR_failXbb, respectively. $N_{\overline{SR}}$ is the number of events in $\overline{SR}$. N_{SR}^{dijet} is the number of predicted dijet events in SR. The normalisation does not impact the fits since it is floating but is used to generate all the following prefit plots.

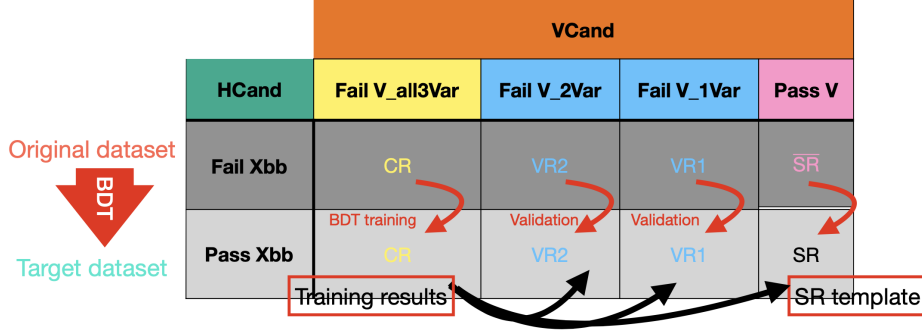


Figure 5.10: Strategy of the BDT reweighting method. The BDT reweighting algorithm is trained with data passing the CR requirements and learns to reweight CR_failXbb large- R jet mass distribution to CR_passXbb. The training result is then applied to VR1 and VR2 for validations. Finally, a SR template is obtained by applying the training result to $\overline{SR}$.

We are allowed to use this strategy to get SR dijet templates for the following reasons:

- Enough statistics for training and for validations. There are 10 times more events in failXbb than in passXbb region.
- Largely dominant dijet events in CR, VR1 and $\overline{SR}$. The Higgs contribution, $V + \text{jets}$, VV and $t\bar{t}$ in CR_passXbb, CR_failXbb contribute less than 5%, which means the BDT reweighting algorithm can be trained with pure dijet contribution in CR. In $\overline{SR}$, the fraction of dijet is larger than 90%. Furthermore, to investigate the impact of ggF contribution in the training dataset, ggF events are injected to enrich their fraction by a factor of ten. The impact is negligible.
- Similar flavour compositions among regions. The $X \rightarrow b\bar{b}$ tagger is flavour-based. The flavour compositions are similar for each passXbb region, and the flavour compositions are similar for each failXbb region.

5.6.1.2 Algorithm and training variables

The Gradient Boosted Reweigher (GBR) [170] is used in this analysis. The GBR is capable of reweighting events in a high-dimensional space by splitting the space of variables to capture the differences between the original dataset and the target dataset. Furthermore, the GBR always has good statistical precision on weights. This section shows only the optimal training setup and results. The number of trees is optimised as 100. The depth of trees is set to 3. The minimal number of events in the leaf is required to be 500. The training variables used in this analysis are shown in Table 5.5.

Table 5.5: The optimised training variables for BDT reweighting method.

Variable	Meaning
p_T^H	Muon-corrected $HCand$ jet p_T
m^H	Muon-corrected $HCand$ jet mass
η^H	$HCand$ jet η
ϕ^H	$HCand$ jet ϕ
N_{trk}^H	Number of ungroomed tracks associated to the $HCand$ jet
$N_{trk\ jet}^H$	Number of ghost-associated VR track jets to the $HCand$ jet

5.6.1.3 BDT reweighting results

The training is done in each differential p_T bin separately.

The two-sample Kolmogorov-Smirnov (KS), non-parametric, is used to compare the reweighted shape with the target shape. The KS statistic is the KS distance [171]. Figure 5.11 shows the comparison of the shape of m_H in CR before and after BDT reweighting. After BDT reweighting, the reweighted CR_failXbb shape is similar to CR_passXbb. The KS distance of reweighted CR_failXbb to CR_passXbb data is reduced to 0.004. The GBR is well trained to learn the shape in CR_passXbb for all p_T bins. Figure 5.12 shows the shape of the dijet template obtained for the SR. Only the side-band areas are shown in the plots. For $250 < p_T^H < 450$ GeV, the SR and $\overline{SR}$ have a close shape even before reweighting with a KS distance of 0.006. There is less margin to get a better description after reweighting. The KS distance after reweighting is 0.014, slightly larger than the value before reweighting, but the BDT reweighting shape is still a good description of the SR data side-band. For $450 < p_T^H < 650$ GeV and $p_T^H > 650$, the KS distance decreases after reweighting in SR. The BDT reweighting shape can describe well the SR data side-band.

The area under the ROC (AUC) provides an aggregate measure of performance [172]. Reweighting aims to bring the original distribution as similar as possible to the target distribution. Thus, the AUC after reweighting should be closer to the random guess value of 0.5. Figure 5.13, 5.14 and 5.15 show the ROC curves for different p_T bins for VR1. After reweighting, the ROC is closer to the random guess curve and smaller AUCs are obtained, which means the failXbb data is less distinguishable from the passXbb data after reweighting. It is what we expect from the BDT reweighting algorithm.

5.6.2 BDT systematic uncertainties

5.6.2.1 Statistical uncertainty via bootstrap method

In order to estimate the systematic uncertainties introduced by the statistics of the training datasets (CR_failXbb, CR_passXbb), the bootstrap method is adopted [173]. A set of N GBRs (bootstrap reweighters) is trained, each time varying the training datasets by re-sampling it with replacement. The replacement is obtained by simply weighting each event in the dataset with a Poisson variable having a mean of 1. Finally, N histogram templates are extracted from the bootstrap reweighters for the SR. The nominal background template

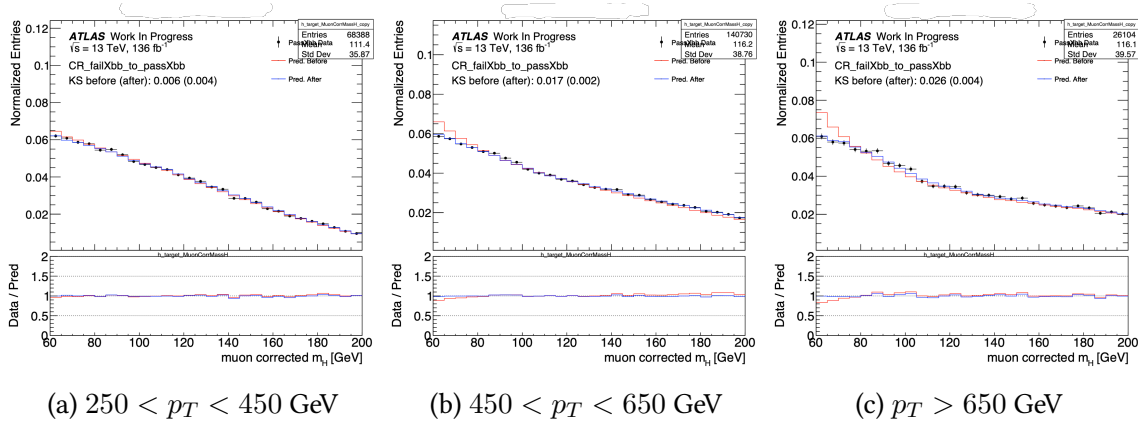


Figure 5.11: The shape of muon corrected Higgs candidate jet mass in CR_failXbb before (Red) and after (Blue) BDT reweighting in CR_passXbb (Black) for differential p_T bins: $250 < p_T^H < 450$ GeV (a), $450 < p_T^H < 650$ GeV (b) and $p_T^H > 650$ GeV (c). Black: target data distribution.

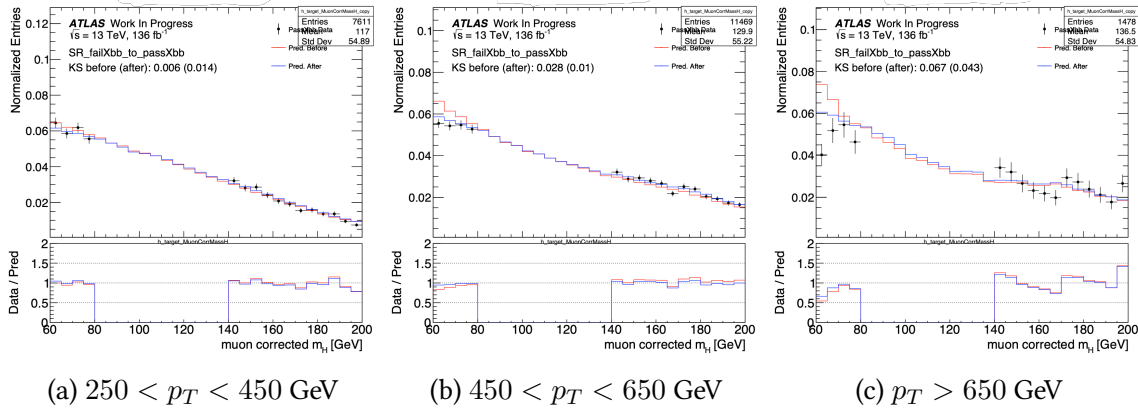


Figure 5.12: The shape of muon corrected Higgs candidate jet mass in $\overline{\text{SR}}$ before (Red) and after (Blue) BDT reweighting in SR (Black) for differential p_T bins: $250 < p_T^H < 450$ GeV (a), $450 < p_T^H < 650$ GeV (b) and $p_T^H > 650$ GeV (c). The black line is the target data distribution. Only the side-band data is shown in the plots.

is obtained from the GBR with the training datasets without replacement. Two additional templates are defined by using the nominal $\pm$ half of the interquartile (IQR) range as the upper and lower symmetric error bands.

Since the multijet normalisation is floating during the fit, the up and down templates share the same normalisation factor with the nominal template.

In Figure 5.16, an estimate of the magnitude of error calculated with $N=250$ bootstrap GBRs is shown for differential p_T bins for the m_H in the SR. The bootstrap variations are smaller than 2% for all p_T bins. A discrepancy is observed between the nominal and the bootstrap median distribution. In order to check the impact of the discrepancy, the same tests have been done with the bootstrap median distribution as the dijet template; there is no difference found in the final results.

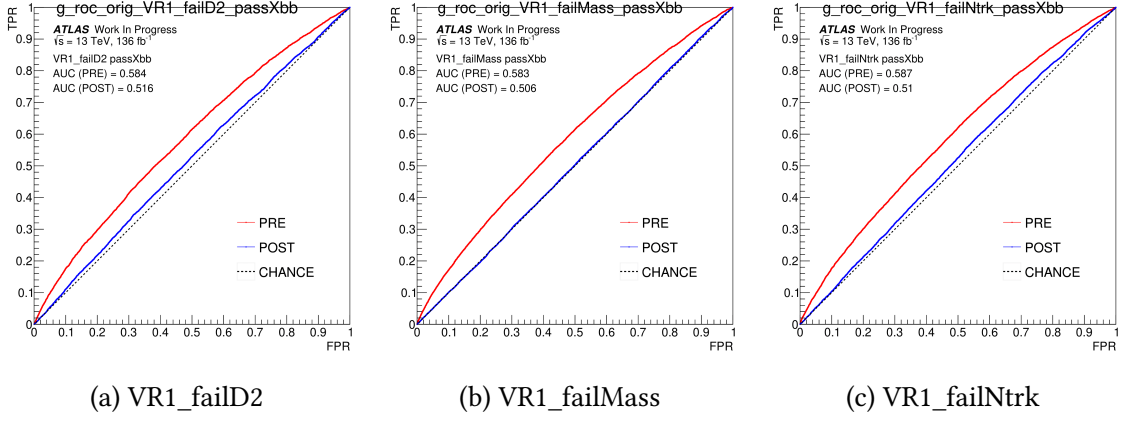


Figure 5.13: ROC of VR1 for $250 < p_T^H < 450$ GeV. The X-axis is the True Positive Rate (TPR). The Y-axis is False Positive Rate (FPR). The red curve is before reweighting. The blue curve is after reweighting. The gray dashed curve is the random guess curve if there is no difference between the compared dataset and the target dataset.

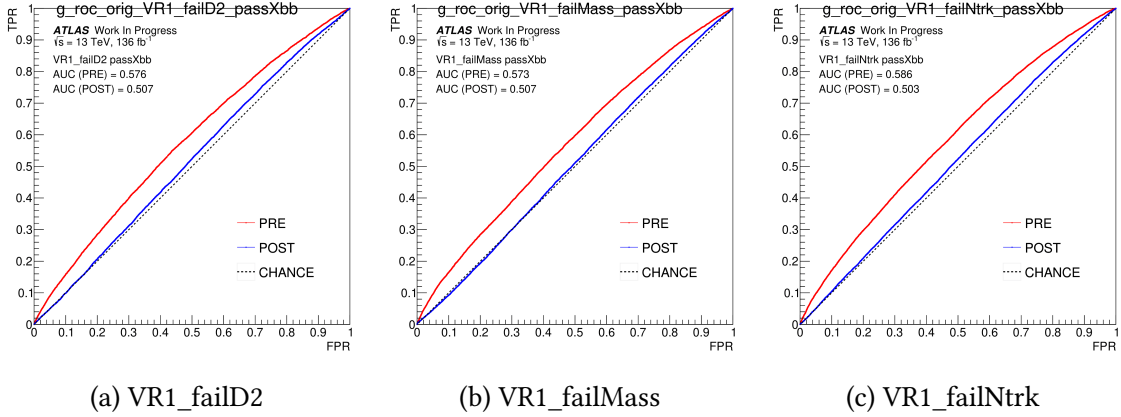


Figure 5.14: ROC of VR1 for $450 < p_T^H < 650$ GeV. The X-axis is the True Positive Rate (TPR). The Y-axis is False Positive Rate (FPR). The red curve is before reweighting. The blue curve is after reweighting. The grey dashed curve is the random guess curve if there is no difference between the compared dataset and the target dataset.

5.6.2.2 Non-closure systematic uncertainties

A non-closure systematic uncertainty is applied to consider the modelling imperfections in the dijet estimation method. For each VR1, the difference between the reweighted VR1 and VR1 data is computed. Only VR1 is used since their compositions are similar to SR. The non-closure systematic uncertainty is the envelope of all the histograms showing the differences. All the histograms are rebinned to have a smoother template, as shown in Figure 5.17.

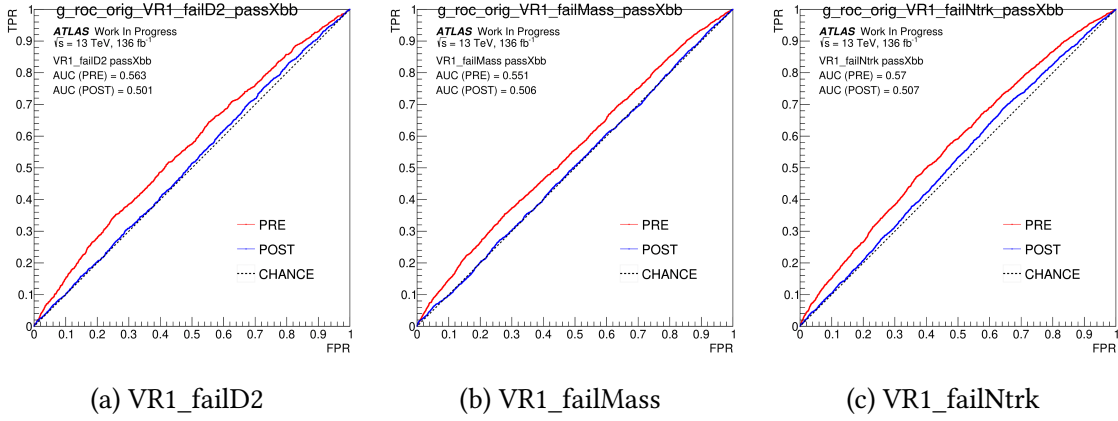


Figure 5.15: ROC of VR1 for $p_T^H > 650$ GeV. The X-axis is the True Positive Rate (TPR). The Y-axis is False Positive Rate (FPR). The red curve is before reweighting. The blue curve is after reweighting. The grey dashed curve is the random guess curve if there is no difference between the compared dataset and the target dataset.

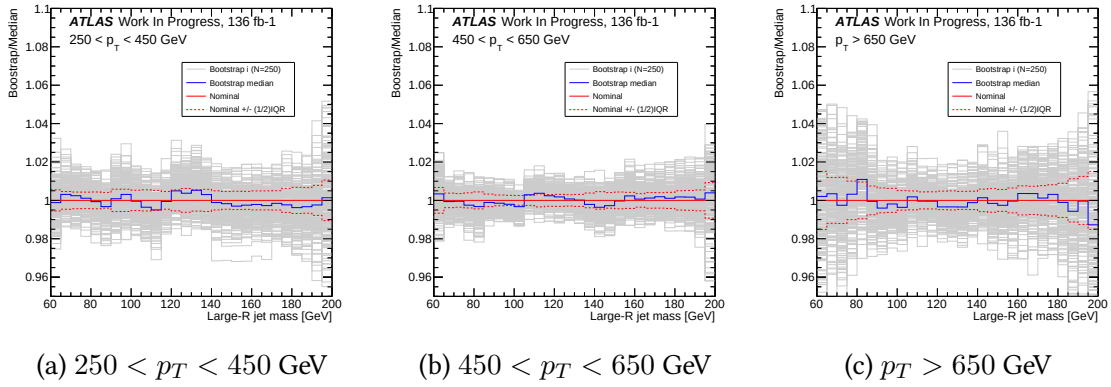


Figure 5.16: The ratio among each of N bootstraps SR templates and the nominal one for differential p_T . The blue line is the median obtained from the bootstrap distributions. The dashed red lines are the nominal $\pm$ half of the interquartile (IQR) range, i.e. the up and down variations of the statistical uncertainty estimated by the bootstrap method.

5.6.3 SR template tests

In order to test dijet templates extracted with the BDT reweighting method and to compare with the TF method, we perform two tests:

- Fit to the side-band of the SR data, see 5.6.3.1.
- Fit to the Asimov data with $\mu_{inj} = 1$, see 5.6.3.2.

In both tests, the common fit settings are the following. The VH is described by the MC histogram with a yield strength μ_{VH} . The other Higgs contributions, i.e. ggF , $t\bar{t}H$, VBF, are described by their MC histogram and fixed to their prediction. The $V + \text{jets}$ is described by its MC histogram and fixed to its prediction. A 10% normalisation uncertainty is assigned according to Ref. [139]. The VV is described by its MC histogram and fixed

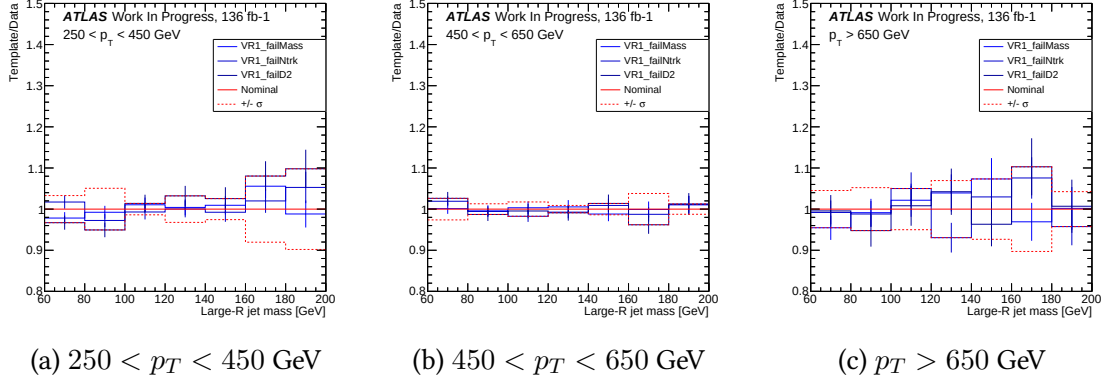


Figure 5.17: Non-closure systematic uncertainties of BDT templates for differential p_T bins. The dashed red lines are the up and down variations of the non-closure systematic uncertainties.

to its prediction. A 20% normalisation uncertainty is assigned. The $t\bar{t}$ is described by its MC histogram. The MC normalisation is scaled by a factor of 0.83, and a 12% normalisation uncertainty is assigned according to Ref. [174]. The MC statistical uncertainty is considered bin-by-bin. The statistical uncertainties smaller than 1% are neglected. A 1.7% of luminosity uncertainty [142] is considered for all MC histograms. Simulated JMR uncertainties via a smearing process are also applied for all MC histograms. The dijet background is modelled by the BDT templates. The bootstrap statistical uncertainty and non-closure systematic uncertainties are included as shape variations of the BDT templates. The normalisation of the templates is set to free (μ_{QCD}).

5.6.3.1 Side-band fits

The test is checks the behaviour of systematic uncertainties in the fit to SR data side-band by excluding the Higgs mass range 80-140 GeV. The VH process is fixed to its MC prediction in the fits. The inclusive fit is performed simultaneously for all p_T bins. The differential fits are performed separately for each p_T bin. Figure 5.18 shows the pre-fit, inclusive post-fit and differential post-fit plots for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. The post-fit plots show a good agreement between fit results and the data side-band. The pulls from the fits are shown in Figure 5.19. The non-closure systematic uncertainty for the first p_T bin is pulled in the inclusive and differential fits. Figure 5.20 shows the correlation map of the fitted variables in the inclusive and differential side-band fits. μ_{QCD} has a correlation to the $t\bar{t}$ and $V + \text{jets}$ cross section uncertainties in both inclusive and differential fits and μ_{QCD} also correlates to the non-closure systematic uncertainty of $250 < p_T^H < 450$ GeV and $450 < p_T^H < 650$ GeV. The fitted distributions with BDT reweighting method and the TF method are compared in Figure 5.21. The error bars in the plots include all the systematic uncertainties considered in the side-band fits. Globally, the two methods have consistent post-fit distributions within uncertainties.

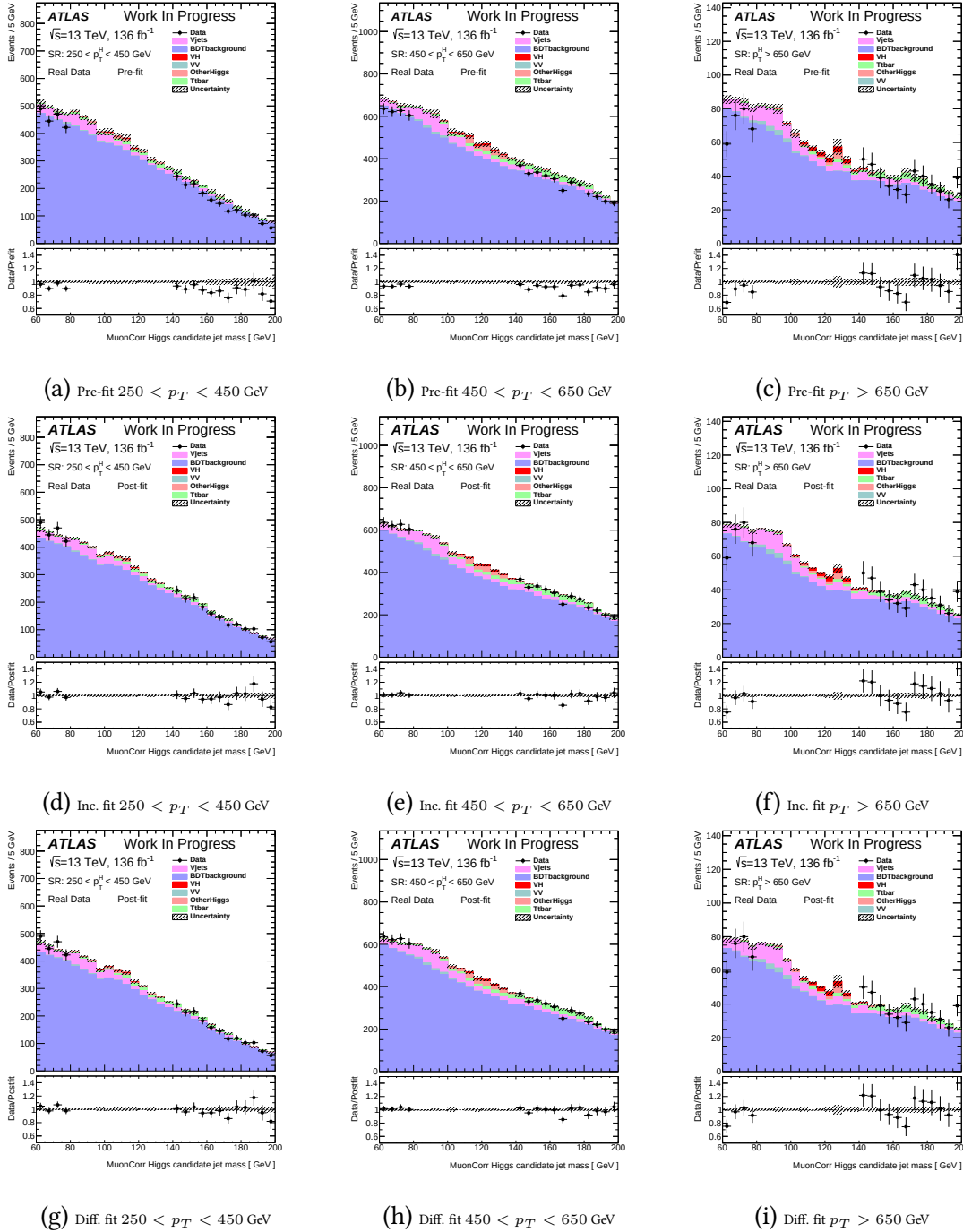
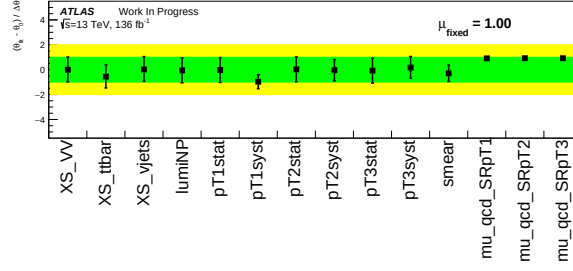


Figure 5.18: The muon corrected large- R jet mass distribution in SR for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. In each plot, the distribution of data is shown in black dots. Only side-band data is shown in the plots. The total statistical and systematic uncertainty is shown with the dashed band. Figure (a)-(c) are pre-fit muon corrected large- R jet mass distribution. The predicted dijet distribution using the BDT reweighting method is shown in blue. Other processes are predicted by MC simulations. Figure (d)-(f) are post-fit muon corrected large- R jet mass distribution from the inclusive fit setup and Figure (g)-(i) are from the differential fit setup. The post-fit dijet distribution using the BDT reweighting method is shown in blue.



(a) Inclusive fit.

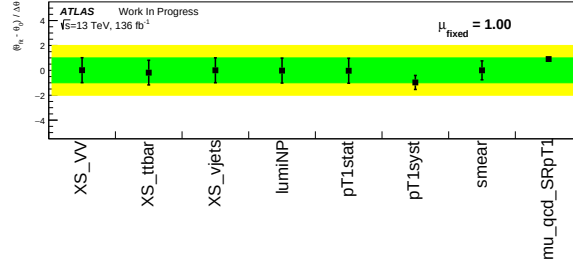
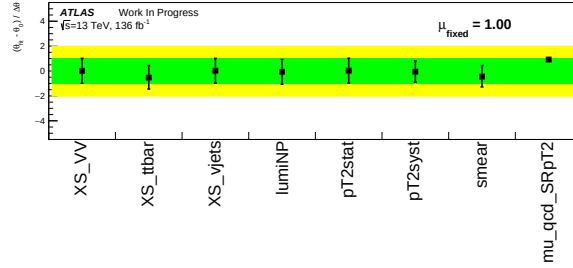
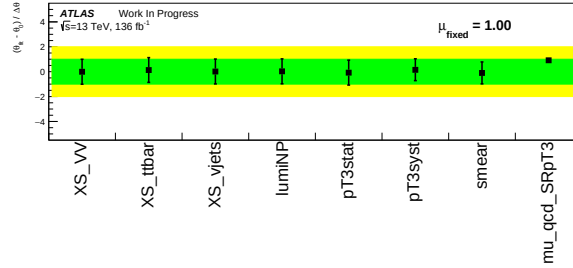

 (b) Differential fit $250 < p_T^H < 450$ GeV.

 (c) Differential fit $450 < p_T^H < 650$ GeV.

 (d) Differential fit $p_T^H > 650$ GeV.

Figure 5.19: Nuisance parameters of side-band fits. pT1stat, pT2stat and pT3stat are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; pT1syst, pT2syst and pT3syst are the non-closure systematic uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1, mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.

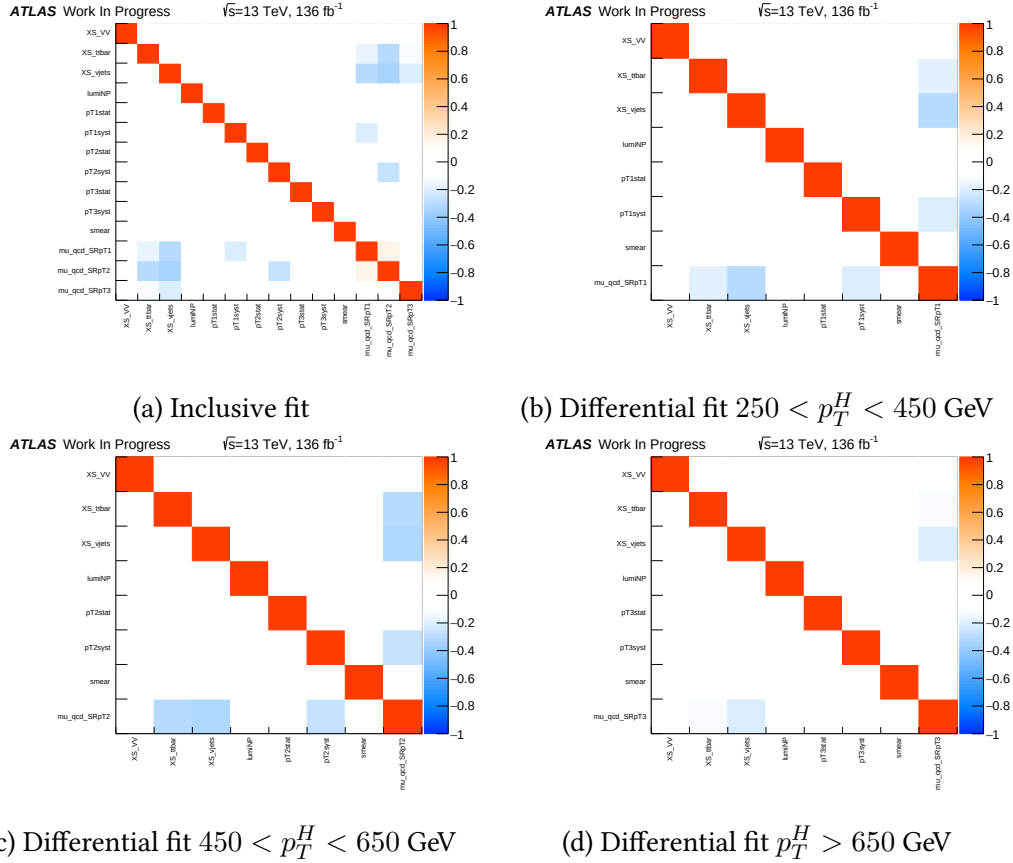


Figure 5.20: Correlation map for side-band fits. $pT1stat$, $pT2stat$ and $pT3stat$ are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; $pT1syst$, $pT2syst$ and $pT3syst$ are the non-closure systematic uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1 , mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.

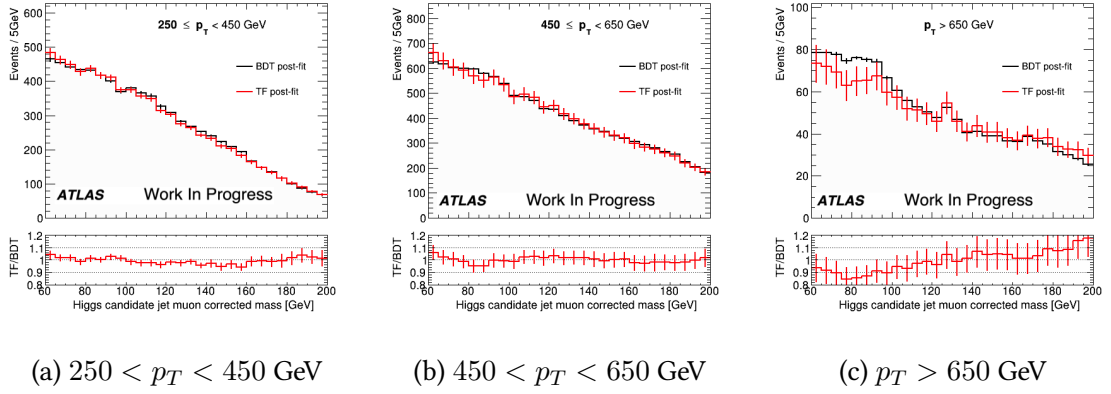


Figure 5.21: Comparisons of the fitted Higgs candidate muon corrected jet mass distributions with the BDT reweighting method (black) and the TF method (red) from the inclusive side-band fit. The error bars show the total fitted uncertainties in each method. The bottom panels show the ratio of the event yields fitted from the TF method and the BDT reweighting method.

5.6.3.2 Asimov fits

The data used in the Asimov fits is the Asimov data [175]. The Asimov data is generated with $\mu_{inj} = 1$ for each p_T bin. The VH , other Higgs contributions, $V + \text{jets}$, VV and $t\bar{t}$ are from their MC templates with normalisation fixed to their prediction. The dijet events are from the BDT templates rescaled to the μ_{QCD} obtained from the inclusive side-band fit, i.e. $\mu_{QCD}^{pT1} = 0.91$, $\mu_{QCD}^{pT2} = 0.92$ and $\mu_{QCD}^{pT3} = 0.92$. The normalisation of the VH process is set free in the Asimov fits. The common fit settings are used. Figure 5.22 shows the pre-fit, inclusive post-fit and differential post-fit plots for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. The pulls of the nuisance parameters from the fits are shown in Figure 5.23. The non-closure systematic uncertainty for $250 < p_T^H < 450$ GeV and JMR uncertainty via smearing are constrained in the inclusive fit. The non-closure systematic uncertainty for $250 < p_T^H < 450$ GeV is constrained in the differential fit. The JMR uncertainty via smearing is constrained for $450 < p_T^H < 650$ GeV in the differential fit. Figure 5.24 shows the correlation map of the fitted variables in the inclusive and differential Asimov fits. The μ_{QCD} behaves as in the side-band fits. Moreover, the μ_{VH} has a correlation to the μ_{QCD} in both inclusive and differential fits and the μ_{VH} also correlates to the non-closure systematic uncertainty of $250 < p_T^H < 450$ GeV and $450 < p_T^H < 650$. The fitted μ_{VH} with the BDT reweighting method and the TF method is summarised in Table 5.6. The 2% disagreement with 1 in the BDT method for the inclusive fit and differential fit $450 < p_T^H < 650$ GeV needs to be understood. The μ_{VH} is consistent with the injected one in the TF method. The fitted errors with the TF method are smaller than the ones with the BDT reweighting method.

Both the BDT reweighting method and the TF method provide a decent model for the measurement. However, they have very different flexibilities in the statistical fits. In the BDT reweighting method, the dijet template is used and the estimated systematic uncertainties are smaller than 10%. The flexibility of the fits is from the normalisation of the dijet background and the systematic uncertainties added in fits. The flexibility will increase when more systematic uncertainties are added to the fits, see Section 5.7. In the TF method, the parameters of the TF function are fully data-driven. The floating parameters provide more freedom to fit the data distribution than the BDT reweighting method. When more systematic uncertainties are added, the stability of the fits needs to be studied.

Table 5.6: Post-fitted μ_{VH} in signal+background fit to the Asimov data with $\mu_{inj} = 1$ with the BDT reweighting method and the TF method.

p_T	μ_{VH} (BDT)	μ_{VH} (TF)
$250 < p_T^H < 450$ GeV	$1.00^{+2.91}_{-2.87}$	$1.00^{+1.88}_{-1.85}$
$450 < p_T^H < 650$ GeV	$0.98^{+1.33}_{-1.25}$	$1.00^{+0.99}_{-0.99}$
$p_T^H > 650$ GeV	$1.00^{+1.37}_{-1.31}$	$1.00^{+1.01}_{-0.97}$
Inclusive	$0.99^{+0.90}_{-0.86}$	$1.00^{+0.67}_{-0.66}$

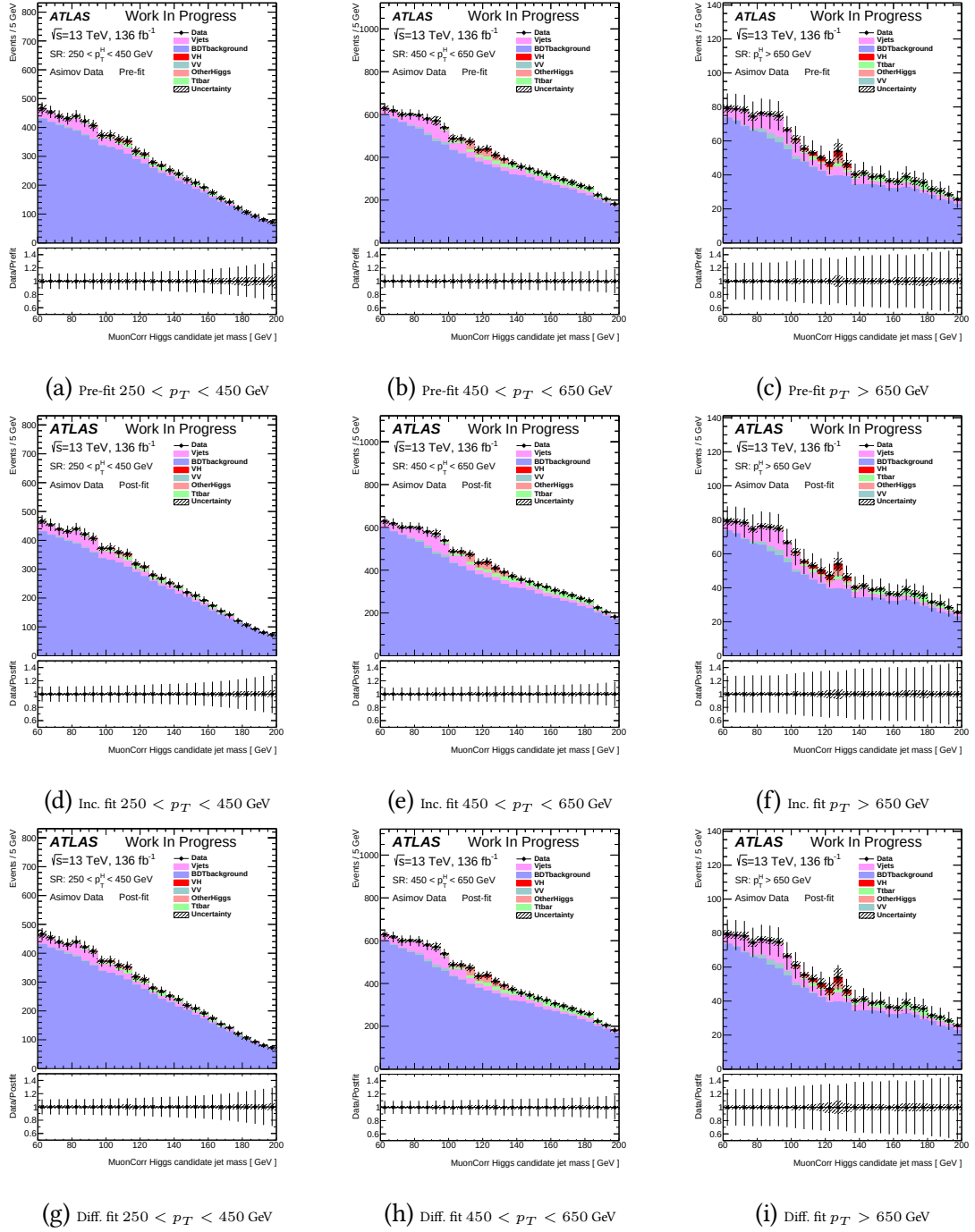
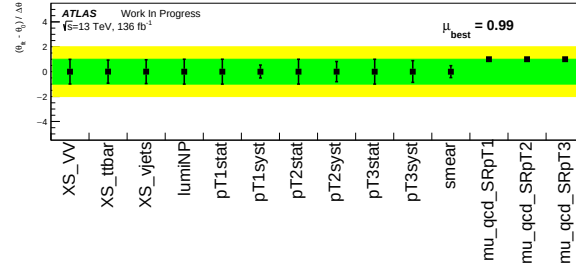


Figure 5.22: The muon corrected large- R jet mass distribution in SR for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. In each plot, the distribution of Asimov data is shown in black dots. The total statistical and systematic uncertainty is shown with the dashed band. Figure (a)-(c) are pre-fit muon corrected large- R jet mass distribution. The predicted dijet distribution using the BDT reweighting method is shown in blue. Other processes are predicted by MC simulations. Figure (d)-(f) are post-fit muon corrected large- R jet mass distribution from the inclusive Asimov fit setup and Figure (g)-(i) are from the differential Asimov fit setup. The post-fit dijet distribution using the BDT reweighting method is shown in blue.



(a) Inclusive fit.

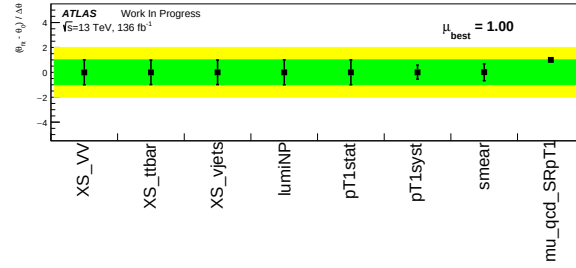
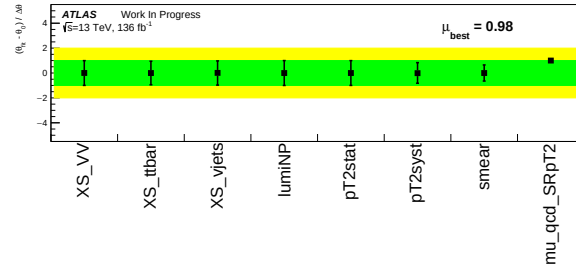
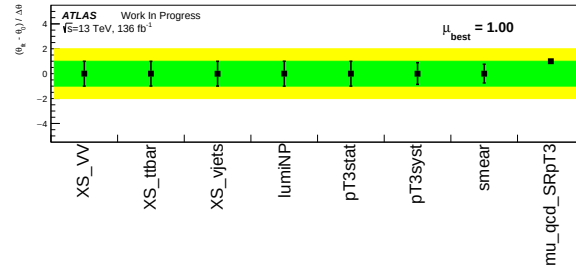
(b) Differential fit $250 < p_T^H < 450$ GeV.(c) Differential fit $450 < p_T^H < 650$ GeV.(d) Differential fit $p_T^H > 650$ GeV.

Figure 5.23: Nuisance parameters of Asimov fits. $pT1stat$, $pT2stat$ and $pT3stat$ are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; $pT1syst$, $pT2syst$ and $pT3syst$ are the non-closure systematic uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1 , mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.

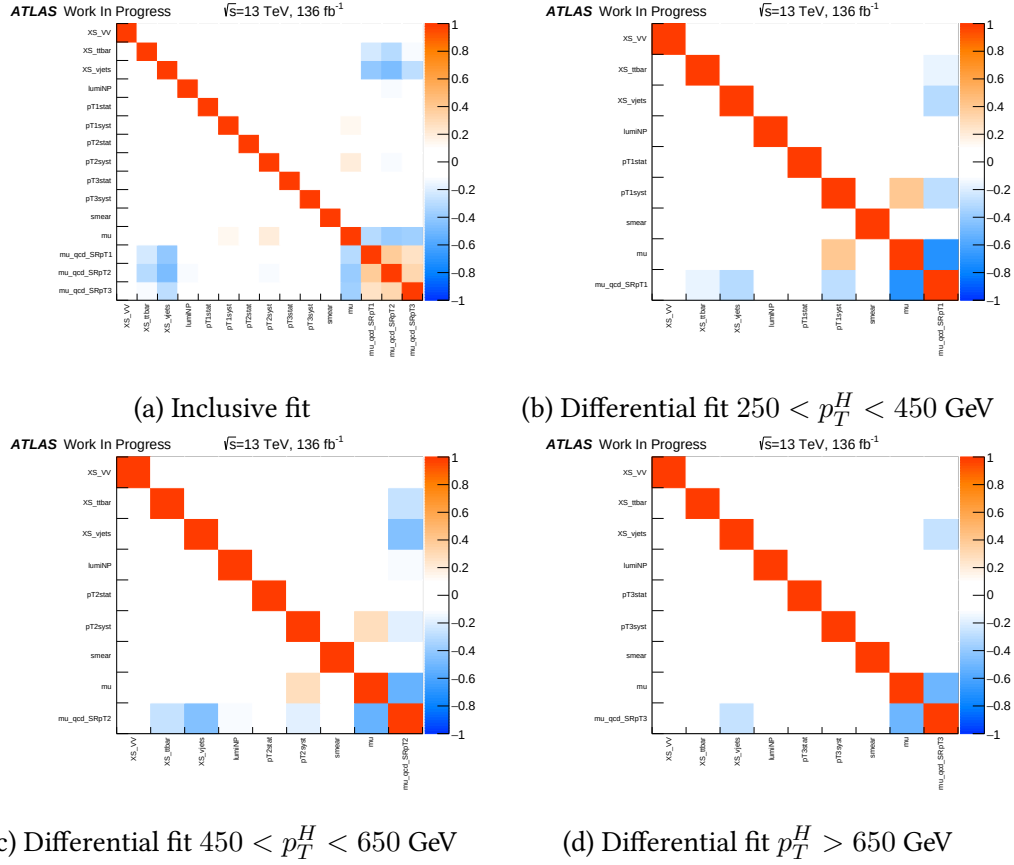


Figure 5.24: Correlation map for Asimov fits. pT1stat, pT2stat and pT3stat are the bootstrap statistical uncertainty for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; pT1syst, pT2syst and pT3syst are the non-closure systematic uncertainty for for $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV respectively; smear means the simulated JMR uncertainty via the smearing process. mu_qcd_SRpT1, mu_qcd_SRpT2 and mu_qcd_SRpT3 are the normalisations of the dijet, which is free parameters in the fits.

5.7 Systematic uncertainties

Systematic uncertainties arise from the reconstruction of the various physics objects and from theoretical and/or modelling uncertainties affecting the predictions for both the backgrounds and signals. These uncertainties manifest themselves as uncertainties both in the overall yield and shape of the final observable.

5.7.1 Experimental systematic uncertainties

The following experimental systematic uncertainties are considered in this analysis:

- Large- R jet systematic uncertainties. This is a set of systematic uncertainties derived from jet calibration described in Section 3.3.4, including uncertainties for JES, JER, JMS and JMR. They will be applied to all MC processes except dijet.
- V -tagging related systematic uncertainties. This is a set of systematic uncertainties derived from the W/Z tagger calibration [146]. They will be applied to all MC processes except dijet.
- H -tagging related systematic uncertainties. It includes all systematic uncertainties obtained from the $X \rightarrow b\bar{b}$ tagger calibration. The uncertainties from $Z + \text{jets}$ calibration are summarised in Table 5.4. The uncertainties from $Z\gamma$ and $t\bar{t}$ calibration are summarised in Ref. [126]. The uncertainties will be applied to MC processes according to the large- R jet topology. The uncertainties from $Z + \text{jets}$ and $Z\gamma$ calibrations will be applied to the Z and H jets. The uncertainties from the $t\bar{t}$ calibration will be applied to the top jets.
- μ -related systematic uncertainties. Since the muon-in-jet correction is applied, the uncertainties in the trigger, reconstruction, identification and isolation efficiencies for muons [122] will be considered.
- Others. A 1.7% of luminosity uncertainty [142] is considered. The pileup reweighting uncertainty will be added for all MC histograms.

5.7.2 Modelling systematic uncertainties

For the signal process, the uncertainty on the cross section for VH is taken to be 5% [176]. It includes the scale variations. The NNLO corrections are minor, and no additional systematic uncertainty is assigned. The electroweak (EW) corrections are also considered. The uncertainties for the STXS bins will be evaluated as Ref. [177].

Different modelling systematic uncertainties are considered for different background processes:

- ggF: The uncertainty on the cross section for ggF will be taken to be 20% [176]. It includes variations of the factorisation and renormalisation scales, PDF, and parton

shower model to account for their respective uncertainties. The NLO correction is small, and no additional systematic uncertainty will be assigned.

- VBF: The uncertainty on the cross section for VBF will be taken to be 0.5% [176]. The electroweak (EW) corrections will also be considered.
- $t\bar{t}H$: The uncertainty on the cross section for $t\bar{t}H$ will be taken to be 13% [176]. The EW corrections will also be considered.
- V + jets: A 10% normalisation uncertainty is considered [139]. The renormalisation and factorisation scale uncertainties, the EW corrections, and the uncertainties from the alternative generator samples and alternative fragmentation models will also be considered.
- VV : A 20% normalisation uncertainty is considered. The renormalisation and factorisation scale uncertainties and the EW corrections will be considered.
- $t\bar{t}$: A 12% normalisation uncertainty is considered [174]. The alternative parton shower (PS) model and the alternative matrix element (ME) generator uncertainties will be considered.
- dijet: In the BDT reweighting method, the shape variations from the bootstrap statistical uncertainty and the non-closure systematic uncertainties are considered as described in Section 5.6.2. In the TF method, the difference between the post-fit distribution and the data in the VR are taken as a systematic uncertainty.

5.8 Summary

The analysis focuses on the associated production of a vector boson decaying to a pair of quarks and a Higgs boson decaying to a pair of b -quarks at high transverse momentum. The signal strength is being measured using the full Run 2 data in inclusive and differential p_T^H bins: $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. Then, the cross section within the STXS framework will be measured. These will be used to constrain anomalous couplings in a SMEFT. The W/Z tagger at 50% WP is used to tag the vector boson. The $X \rightarrow b\bar{b}$ tagger at 60% WP is used to tag the Higgs boson. The leading two large- R jets are assigned using the discriminant of the $X \rightarrow b\bar{b}$ tagger as the Higgs boson candidate and the vector boson candidate. The MC signal significance is around 1.28 for $p_T^H > 250$ GeV.

The scale factors of the $X \rightarrow b\bar{b}$ tagger have been remeasured in the VH analysis for $p_T > 450$ GeV using the method developed in Chapter 4 and using an orthogonal selection to achieve statistical independence. The measured scale factors are $1.44_{-0.57}^{+0.50}$ for $450 < p_T < 500$ GeV, $1.30_{-0.44}^{+0.39}$ for $500 < p_T < 600$ GeV and $0.86_{-0.55}^{+0.49}$ for $600 < p_T < 1000$ GeV.

Two background modelling methods have been developed: the BDT reweighting method and the TF method. I have contributed the most to the development and validation of the BDT reweighting method. For the BDT reweighting method, a SR template has been obtained. The statistical uncertainty and the non-closure systematic uncertainties of the template shapes have been estimated. The templates have been tested by the SR side-band fits

and the Asimov fits. The TF method is presented in Appendix C.3. The post-fit distribution in the SR side-band fits of the two methods is consistent within systematic uncertainties. The fitted errors with the TF method are smaller than the ones with the BDT reweighting method in the Asimov fits.

The analysis is still ongoing. The dijet modelling method will be chosen among the BDT reweighting and TF methods. The full systematic uncertainties described in Section 5.7 have been included in the frameworks and will be added in the final measurement. The STXS measurement will be performed and the Wilson coefficients to be used will be defined in the setup for the SMEFT interpretation.

SILICON SENSORS FOR HIGH ENERGY PHYSICS

Pixel detectors realised on high resistivity silicon substrates are one of the most common technologies for high-energy physics experiments for studying short-lived particles and coping with the increasing interaction rates and energy. In this chapter, we introduce silicon detectors, focusing on aspects relevant for tracking and vertexing. First, the basics of semiconductor detectors are introduced in Section 6.1. Then the pixel sensor concept and its properties are presented in Section 6.2 followed by a discussion on the effects of radiation damage on the silicon pixel sensors in Section 6.3.

6.1 Semiconductor basics

When atoms are organised in a lattice, due to the Pauli exclusion principle, electrons in the atoms influence each other and organise in energy bands: the energy levels can be categorised generically in valence band and conduction band. In conductors the two bands are overlapping; in semiconductors they are close in energy to each other while in insulators they are far apart as shown in Figure 6.1. For semiconductors and insulators the gap between the two bands is called forbidden energy gap (E_g). The valence band is populated by the valence electrons, and the conduction band can be or not partially populated depending on the class of material and the temperature. Conduction happens when enough free electrons have the energy to be in the conduction band. In semiconductors this happens through transition of electrons from the valence band to the conduction band. When an electron moves to the conduction band, it leaves a vacancy in the valence band, called a "hole". The hole behaves like a particle with positive charge. Both electrons and holes are charge carriers. The typical value of the gap for semiconductors is about 1 eV which is higher than the thermal energy at room temperature (about 25 meV). However, the energy gap of semiconductors usually decreases when their temperature increases.

In order to understand the electrical and thermal properties of crystalline materials, we introduce the concept of the Fermi energy or 'Fermi level' (E_F). All electrons pack into the lowest available energy states at absolute zero temperature and build up a 'Fermi sea'. The Fermi level is the surface of that sea at absolute zero temperature, where no electrons will have enough energy to rise above the surface. The Fermi function gives the probability that a given available electron energy (E) state will be occupied at a given temperature (T in

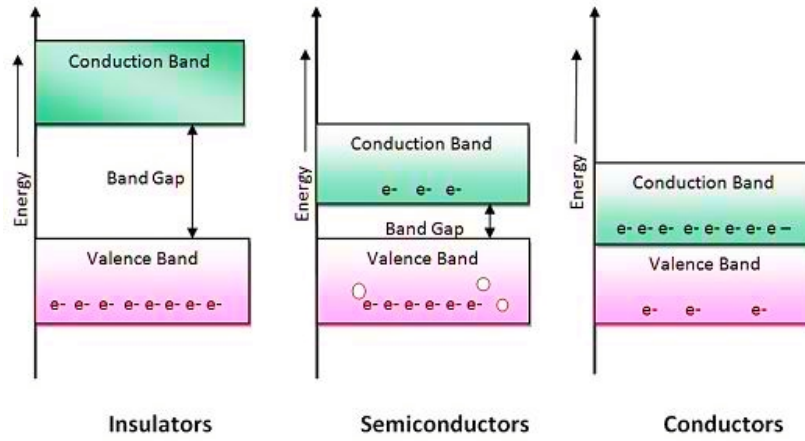


Figure 6.1: Energy band structure of several materials [178].

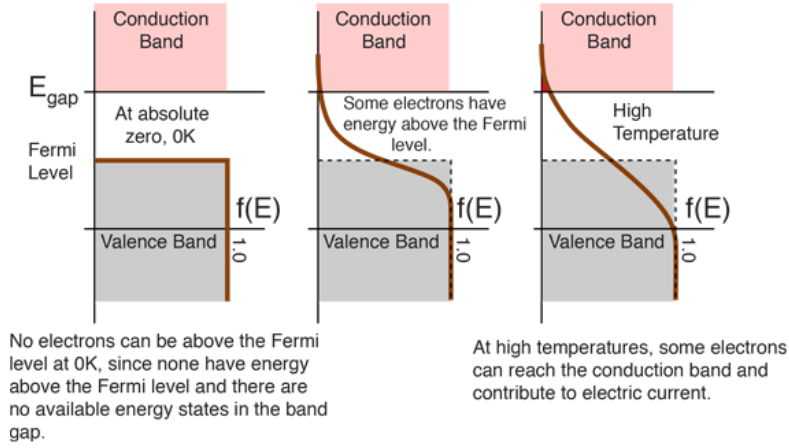


Figure 6.2: Illustration of the Fermi function for a semiconductor [179].

Kelvin):

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \approx e^{-(E-E_F)/kT} \text{ (at room temperature)} \quad (6.1)$$

where k is the Boltzmann constant. Figure 6.2 illustrates the Fermi function for a semiconductor. At 0 K, there is a gap between the Fermi level and the conduction band. At higher temperatures, a larger fraction of the electrons can bridge this gap and participate in electrical conduction. Thus it is important to use semiconductors whose energy gap will not significantly change with the temperature. The most commonly used semiconductors are thus silicon and germanium. At 0 K, the energy gap is 1.21 eV for silicon and 0.79 eV for germanium. At room temperature around 300 K, the energy gap is 1.1 eV for silicon and 0.72 eV for germanium. For HEP detectors, silicon is preferred over germanium because it can work at higher temperatures due to a higher energy gap. Furthermore, other important characteristics can explain silicon's success, such as its large abundance and the possibility of changing its properties by doping.

6.1.1 Charge carrier concentration

6.1.1.1 Intrinsic semiconductor

A piece of semiconductor is called intrinsic if the concentration of impurities is negligible compared to the thermally generated free electrons and holes. The concentration of free charge carriers in the conduction band and the valence band are the same; thus, the Fermi level is roughly in the middle. According to Equation 6.1, the concentration of intrinsic charge carriers can be calculated as:

$$\begin{aligned} n &= 2 \underbrace{\left(\frac{2\pi m_n kT}{h^2} \right)^{3/2}}_{N_C} e^{-(E_C - E_F)/kT} \\ p &= 2 \underbrace{\left(\frac{2\pi m_p kT}{h^2} \right)^{3/2}}_{N_V} e^{-(E_F - E_V)/kT} \end{aligned} \quad (6.2)$$

Where $n(p)$ are the concentration of free electrons (holes), $m_n(m_p)$ the effective mass of electrons (holes), $E_C(E_V)$ the energy of the conduction (valence) band, and E_F the Fermi energy. The quantity $N_C(N_V)$ is the effective density of states in the conduction (valence) band and only depends on the temperature. The Fermi level is defined by the requirement that the semiconductor in total is electrically neutral. n and p should be equal with a result of an intrinsic Fermi level, E_i , very close to the middle of the band gap, as shown by the following equation:

$$E_i = \frac{E_C - E_V}{2} + \frac{3kT}{4} \ln\left(\frac{m_p}{m_n}\right) \quad (6.3)$$

One should notice that the product of electron and hole concentrations depends only on the temperature under thermal equilibrium:

$$np = n_i^2 = N_C N_V e^{-E_g/kT} \quad (6.4)$$

Where n_i is the concentration of the charge carriers in intrinsic semiconductors.

6.1.1.2 Extrinsic semiconductor

In most cases, the material used for semiconductor devices is not intrinsic but doped with a tiny fraction of other materials to alter its conductivity. There are two categories of dopants to create extrinsic silicon:

- **Donors:** the dopants like Phosphorus have one more electron in the outer shell and release their 'extra' electron to the conduction band, as illustrated by Figure 6.3. At room temperature, all donor states are usually ionised, and the concentration of electrons, n , equals the concentration of donor atoms, N_D . As electrons are the majority

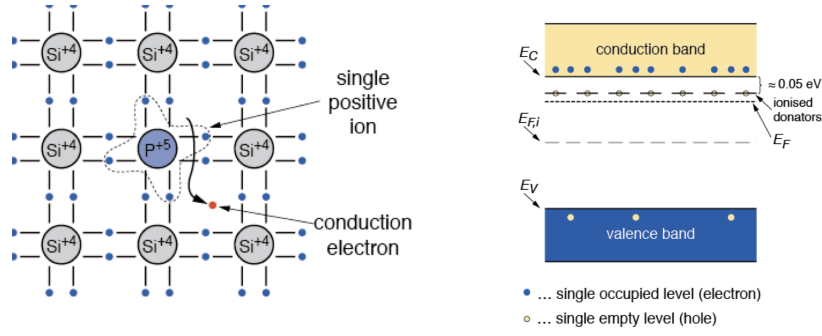


Figure 6.3: Doping silicon with the donor atoms. (Left) Atom bonds with the donor dopant; (right) Energy bands diagram after the donor doping [180].

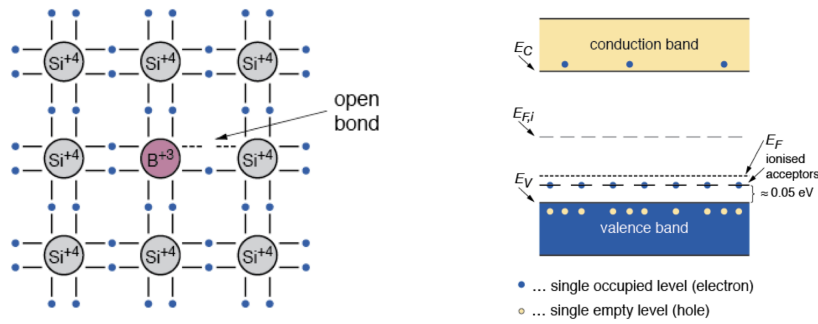


Figure 6.4: Doping silicon with the acceptor atoms. (Left) Atom bonds with the acceptor dopant; (right) Energy bands diagram after the acceptor doping [180].

charge carriers, donor doped silicon is called n-type silicon. By Equation 6.4, the Fermi level shifts towards the conduction band:

$$E_F = E_i + kT \ln\left(\frac{N_D}{n_i}\right) \quad (6.5)$$

- **Acceptors:** the dopants like Boron have one less electrons in their outer shell and trap an electron from the valence band to establish a covalent bond to the neighboring silicon atom, as illustrated by Figure 6.4. At room temperature, all acceptor states are ionised and the concentration of the acceptor, N_A and hole concentrations p are almost equal. As holes are the majority charge carriers, acceptor doped silicon is called p-type silicon. By Equation 6.4, the Fermi level shifts towards the valence band:

$$E_F = E_i - kT \ln\left(\frac{N_A}{n_i}\right) \quad (6.6)$$

6.1.2 The pn-junction

The building block of the silicon particle detector is the pn-junction where the p- and n-type silicon regions are placed into contact as shown by Figure 6.5. At the contact interface, some of the majority charge carriers of one side diffuse into the differently doped side due to the concentration difference. This mechanism creates a zone depleted of free charge

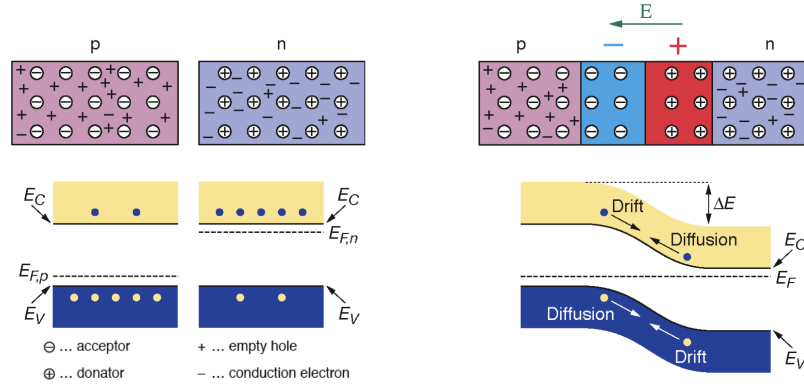


Figure 6.5: The pn-junction formation. (Left) Two oppositely doped semiconductors are compared. Doping silicon with the donor atoms. (Right) The pn-junction is formed [180].

carriers called the depletion region. This region is electrically charged and therefore creates an electrical field counteracting the diffusion, leading to the so-called built-in voltage. In the thermal equilibrium state, the value of built-in voltage is in the order of 0.7 V. When a charged particle crossed the depletion region, the electron-hole pairs created by ionisation along the trajectory will drift under the influence of the electrical field, generating a signal which can be read out by electrodes. When an external voltage in the same direction as the built-in voltage is applied, the depletion zone will be extended to ease particle detection. By looking at Figure 6.5, it is clear that to deplete the junction volume, a higher potential on the n-side should be applied than on the p-side. This polarisation will be referred to the reverse bias voltage. In most cases, the reverse bias voltage is much higher than the built-in voltage. With this assumption, the total width of the depletion zone can be estimated as:

$$W = \sqrt{\frac{2\epsilon_0\epsilon_{Si}}{e|N_{eff}|}} V \quad (6.7)$$

where V is the reverse bias voltage applied, ϵ_0 and ϵ_{Si} are the permittivity of vacuum and silicon respectively, e is the charge of the electron, $|N_{eff}| = |N_D - N_A|$ is the effective doping concentration. The depletion voltage (V_{depl}) is the voltage needed to achieve the complete depletion of the junction volume corresponding to the thickness of the pn-junction (d):

$$V_{depl} = \frac{e|N_{eff}|d^2}{2\epsilon_0\epsilon_{Si}} \quad (6.8)$$

The net generation of carriers in the depletion region creates a current flow called leakage current (I_{leak}). The current can be obtained:

$$I_{leak} = \frac{en_i A d}{2\tau_g} \quad (6.9)$$

where A is the cross-sectional area of the junction and τ_g is the average generation lifetime. When the sensor is fully depleted, I_{leak} is roughly independent of the reverse bias

voltage but depends on the temperature:

$$\frac{I(T)}{I(T_0)} = \frac{T^2}{T_0^2} e^{-\frac{E_a}{2k}(\frac{1}{T} - \frac{1}{T_0})} \quad (6.10)$$

where E_a is the equivalent of an activation energy (the experimental value of E_a for silicon is about 1.21 eV).

However, if the reverse bias exceeds the full depletion voltage, the pn-junction is over-depleted, and a constant field $(V - V_{depl})/d$ is present. If it is increased to very high values, an avalanche breakdown occurs. The leakage current increases exponentially, and the device may be damaged. The voltage at which the breakdown occurs is called the breakdown voltage V_{bd} , which defines the maximum operational bias voltage of the sensor. Therefore, it is crucial to measure the current-voltage curve of each sensor before setting the operational conditions.

6.1.3 Signal formation

When a charged particle passes through the depleted region of the sensor, electron-hole pairs are generated by ionisation along its path. The number of electron-hole pairs created is proportional to the energy left by the particle in the material. The mean energy loss of a charged particle in the material can be described by the Bethe-Bloch formula [181]. In HEP, the particles are typically relativistic; most of the particles crossing the sensor are Minimum Ionizing Particles (MIP) which have the ionisation compatible with the minimum of the Bethe-Bloch formula. A Landau-Vavilov distribution can describe the energy loss spectrum of a MIP in the silicon. The long tail is due to the possibility of energy transfers from the rare inelastic scattering of δ ray or a knock-on electrons. For a silicon sensor, the Most Probable Value (MPV) of a charged particle with such distribution is equivalent to 80 electron-hole pairs per μm .

Once the charge carriers are created, they will move toward the collecting electrode according to several effects:

- **Electrical drift:** The charge carriers drift towards the collecting electrode, following the electric field $\vec{E}$. The charge carrier velocity for the electrons v_e (holes, v_h) depends on the mobility of electrons μ_e (holes, μ_h) in the silicon: $v_e = -\mu_e \vec{E}$ ($v_h = \mu_h \vec{E}$).
- **Thermal diffusion:** The charge carriers are also driven under the thermal diffusion if there is a gradient in the carrier concentration. Since the detectors are usually operated cold, the thermal diffusion is limited and has a minor contribution compared to the electrical drift.
- **Generation and recombination:** Since the energy gap of silicon is relatively low, the generation and recombination of electron-hole pairs are prevalent. An electron in the conduction band can transit directly into a vacancy in the valence band. Alternatively, the electron can recombine with a hole via the intermediary states in the band gap. The latter has a much more significant contribution, especially when the material has been irradiated, see Section 6.3.

- Lorentz angle deflection (if a magnetic field is present): the charge carriers move with an angle compared to the direction of $\vec{E}$.

Once the charge carriers are collected when they arrive at the electrode, the total induced charge Q in the electrode is:

$$Q = q(V_{Ramo}(x_i) - V_{Ramo}(x_f)) \quad (6.11)$$

Where $V_{Ramo}(x_i)$ is the Ramo potential at the initial point of the carrier, i.e. where it is created; $V_{Ramo}(x_f)$ is the Ramo potential at the endpoint of the carrier, i.e. the position of the electrode which collects the carrier. The Ramo potential is calculated using the Shockley-Ramo theorem [182]. Then the induced signal will be processed in an electronic front-end read-out chip; a read-out chip example is presented in Section 7.3.1.

6.2 Pixel detectors

Pixel detectors are made by the pn-junctions which are organised in a matrix of two-dimensional arrays. With such organisation, the position of the tracks across the detector can be measured unambiguously. In general, the start point of a silicon sensor in HEP is a piece of the high purity silicon wafer of the goal thickness which is already doped on one face. The wafer is first thermally oxidised at high temperature to form a protection layer of SiO_2 on the surface. Then, the lithographic steps are performed to structure the wafer surface. The local dopant is then implanted by etching the SiO_2 layers to create contacts and metal is deposited to create the electrodes. Finally, the wafer is lapped down to its required thickness (usually ten to hundreds of micrometers) if it uses a mechanical wafer to reinforce it during the process. There are several possible combinations for the doped bulk and the implant:

- p-type bulk with heavily doped n-type implant (n-in-p)
- p-type bulk with heavily doped p-type implant (p-in-p)
- n-type bulk with heavily doped n-type implant (n-in-n)
- n-type bulk with heavily doped p-type implant (p-in-n)

Since the p-type implants collect holes and the mobility of holes is lower than the electrons, the carrier collection of the p-type implants will be strongly affected by charge trapping of the holes after irradiation. The n-type implants collect the electrons that are less affected. Furthermore, for n-in-n and p-in-p, since they are using the same type of doped material, a doubled-sided process is required to implant guard rings in the backside of the structure, which are more expensive to produce. Therefore, the n-in-p technology is preferred. The current ATLAS planar pixel detectors are built with n-in-n technology. The preferred planar pixel design for the future ATLAS inner tracker (Section 7.1) for HL-LHC is n-in-p.

In this section, we introduce different sensors with a special emphasis on the planar pixel sensors, which have been exploited in the current ATLAS pixel sensors or foreseen for future pixel productions. The planar pixel sensors are also the ones studied in Chapter 7.

6.2.1 Planar pixel sensors

The hybrid planar pixel sensors implants are segmented on the top surface of the bulks. The hybrid refers to their property of being coupled to their electronic read out via a small metallic bond as shown by Figure 6.6. Figure 6.7 illustrates the hybrid planar pixel sensor layout. Necessary features for the sensor operation are discussed in the following.

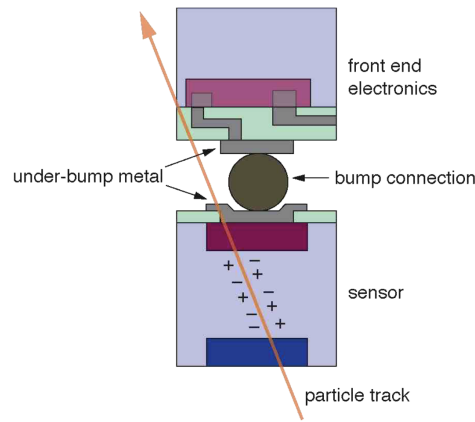


Figure 6.6: Sketch of a hybrid pixel [183].

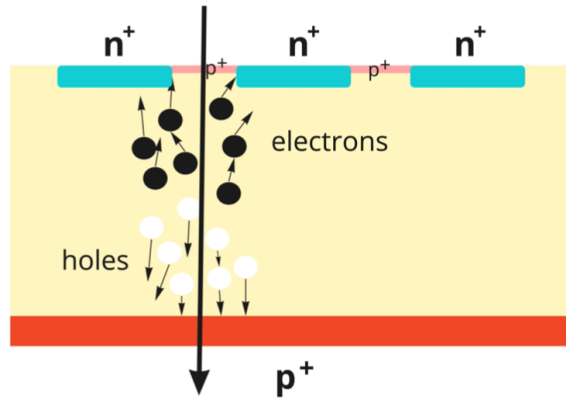


Figure 6.7: Cross-sectional layout of segmented planar n-in-p pixel sensor with p-spray for inter-pixel insulation [184].

Inter-pixel insulation The pixel isolation would be compromised in case of formation of an electron layer just underneath the oxide. The electron layer could short together two pixels, implying a critical loss in terms of resolution. For n-in-p and n-in-n sensors, the pixel isolation can be achieved either using p-spray (low dose p-implantation on all the sensors)

or p-stop (high dose p-implantation between pixels) which would mitigate the effect of the charge layer.

Edges and guard rings The traditional technique to separate sensors from a wafer is performed by dicing with a diamond saw. The sensors are diced from the wafers, with the unavoidable consequence of creating micro-cracks at the edges of the sensors and hence defects in the lattice as shown by Figure 6.8. Therefore, the guard rings are adopted to decrease the voltage gradually at the edge avoiding the generation of leakage current at the border, but these represent a region of low hit efficiency due to the lower electrical field in this part of the sensor. For example, the edge width between the cutting edge and the edge pixels in the current ATLAS pixel detectors is about $1100\ \mu\text{m}$ [185] corresponding to 15 guard rings. However, thanks to a technique called "active edge", it is possible to reduce the insensitive area at the border of the sensor thanks to an ion-etched trench which prevents the effects of the crystal damages produced by the standard mechanical dicing process. The etched trench is filled with doped polysilicon to assure that it will be equipotential with the backside, thus avoiding current generation at the edge, see Figure 6.8. The number of guard rings can also be considerably reduced, or an active edge sensor can even be produced without guard rings. The use of active edges allows for reducing the un-instrumented area at the border of the sensor up to $50\ \mu\text{m}$, as there is no need to add multiple guard rings in this zone.

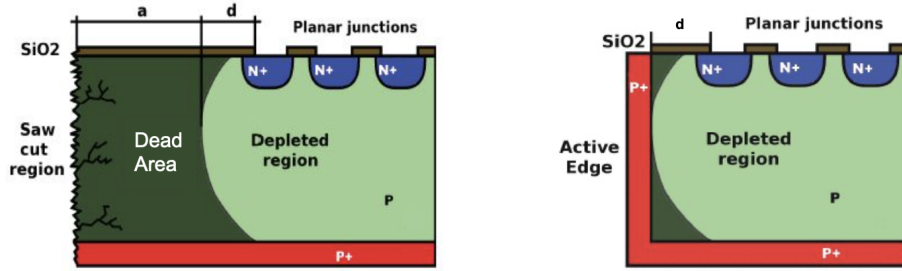


Figure 6.8: Scheme of the edge of a sensor with traditional cut technology (left) and with active edge technology (right).

Protection layers In n-in-p sensors, there is a very small drop in potential from the backside, at high voltage, across the edge. The front face still presents a significant potential close to the cut, while the pixels are kept at ground through the readout electronics, whose surface is also at ground potential. When the n-in-p sensor is irradiated, its operation voltage can be up to a few hundred volts. Hence a protection layer is mandatory to prevent discharges between the sensor area with high voltage (edges and guard rings) and the readout chip, which is in close proximity (of the order of $O(10\ \mu\text{m})$). The typical isolation material can be either parylene or BCB (Benzocyclobutene).

Bias structure While after bump-bonding the polarisation of the pixels is assured by the readout chip, bias structures are necessary if we want to bias the sensor before flip-chip, for testing purpose. Two kinds of biasing strategy are adopted: bias dots/grids or temporary metal. The bias dots and grids approach exploits the punch-through (PT) mechanism or small polysilicon resistors. In the case of PT or resistors, the bias rail network running along the inter-pixel area, is connected to a structure at the corner of pixels, called PT

dot, or directly to the pixel implant, in the case of polysilicon. The PT dot determines the pixel potential via the punch-through mechanism. The bias network is kept at ground voltage during the measurement, thus controlling the pixel potential. The PT dots usually degrade the sensor's performance in terms of hit efficiency in the PT region or via the resistor. Instead, in the case of biasing with temporary metal (TM), a network of metal lines is deposited during the wafer processing. The metal shortens all pixels of each column. This temporary metal layer is afterwards removed via chemical etching. The hit efficiency is more homogeneous as no permanent structures are present after bump-bonding.

6.2.2 3D pixel sensors

3D pixel sensors are also hybrid pixel sensors. A schematic view can be found in Figure 6.9. In contrast to planar pixel sensors, they are composed of heavily doped pillars passing through a low-doped silicon bulk and the electric field is parallel to the surface of the sensor. Thus the electrode distance of the 3D sensors is typically shorter than in the planar sensors, and this allows to obtain the same electrical field at a much smaller voltage. The possibility to operate the sensors with a relatively low bias voltage is very important especially after irradiation. For example, for 3D at IBL lifetime dose, less than 200 V are needed to fully deplete the sensor, compared with a planar sensor which requires 1000V or more [186]. On the other hand, the capacitance of 3D sensors is in general higher than the one of planar sensors of the same thickness.

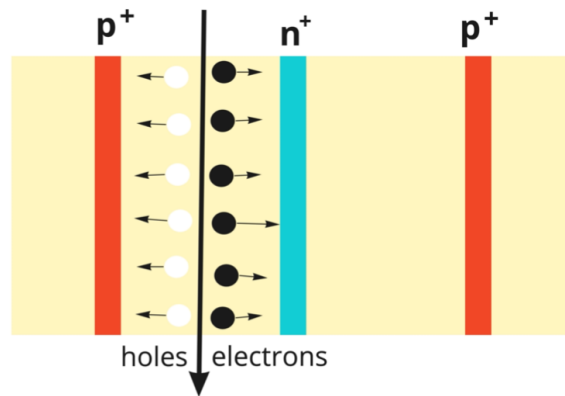


Figure 6.9: Cross-sectional layout of 3D pixel sensor [184].

6.2.3 Monolithic pixel sensors

In contrast to hybrid pixel sensors, monolithic pixel sensors are intended to incorporate the electronics and the sensors on the same wafers, for example, complementary metal-oxide-semiconductor (CMOS) pixel sensors. Since the bump bonding process is not needed anymore, they have promising potential cost reductions for large detector surfaces and manufacturing simplifications. However, only very recently, they have been developed to provide a radiation hardness close to the one needed in some of the HEP experiments.

6.3 Radiation damage

The current installed inner detector of ATLAS accumulated about $10^{15} n_{eq}/cm^2$ of non ionizing dose, and about 300 *kGy* of total ionizing dose. At the HL-LHC phase, it is anticipated that pixels will integrate radiation levels 10 times higher than the current level. Therefore, it is crucial to understand the radiation damage mechanism and detector behaviour under such high flux of particles to develop HEP detectors.

Radiation damages any part of the pixel sensors, such as the bulk, the surface, the interface between the silicon and the oxide. The damage to the surfaces can create charge in the oxide, it can be mitigated by using surface isolation technologies (p-spray or p-stop). The majority of damages are in the bulk, which represents the most of the volume of the sensor. This section focuses on the damage in bulk and discusses its effects on particle detection.

From the microscopic view, an atom in the silicon lattice may be displaced from its original position under the inelastic collisions with the incident particles, as illustrated by Figure 6.10. A displaced atom represents microscopic radiation damage and is called an interstitial defect. The hole left by a displaced atom is called a vacancy. A vacancy combined with an adjacent interstitial site forms a Frenkel defect pair. The defects can be point defects or clusters of defects. In the band gap, the point defects act as intermediate states. They can be charged, consequently influencing the effective doping concentration. This leads to several impacts of radiation: an increase of the leakage current, a change in the operational bias voltage, and a decrease of the charge collection efficiency due to the charge trapping phenomenon.

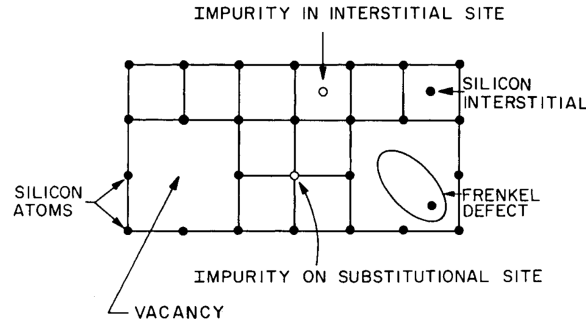


Figure 6.10: Types of point defects in a simple lattice [187].

Leakage current One of the main characteristics of radiation damage is the increase of leakage current. The defects in the band gap act as generation-recombination centers. They lead to a decrease of the generation lifetime τ_g , hence an increase of the leakage current I_{leak} according to Equation 6.9. The rate of increase of leakage current ΔI per unit of fluence Φ and per unit of volume V is called α : $\alpha = \frac{\Delta I}{\Phi V}$. α is in the order of 10^{-17} A/cm and is reducing with the time, as shown in Figure 6.11, an effect called annealing. As the leakage current increases, it heats more the sensor and can give raise to breakdown; one way to mitigate the leakage current increase is to keep the sensor cold.

Operational bias voltage The effective doping concentration depends highly on the fluence; thus, the depletion voltage of the sensors after irradiation shifts according to Equation 6.8. Figure 6.12 shows the change of the effective doping concentration and the depletion

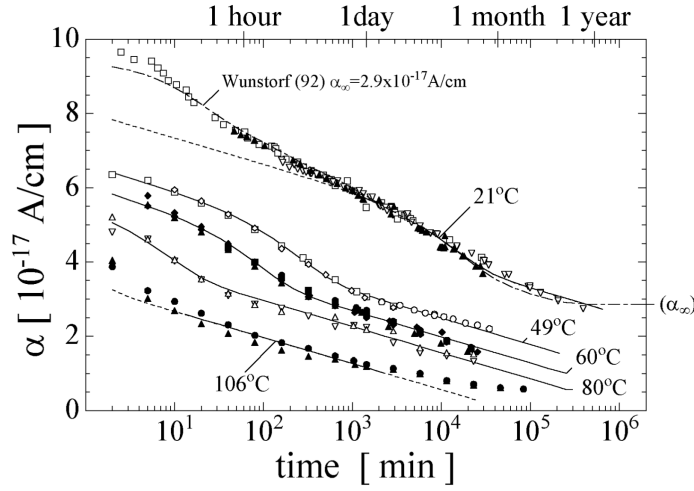


Figure 6.11: The normalised rate of increase of leakage current α as function of the cumulated annealing time [188].

voltage as a function of the fluence. In the case of n-type bulk, the exposure to the fluence of the order of $10^{13} - 10^{14} \text{ n}_{eq}/\text{cm}^2$ leads to a type inversion of the n-type bulk into a p-type bulk and to a steady increase of the full depletion voltage after type inversion. There is no inversion for the n-in-p sensors while the depletion bias voltage increases as well.

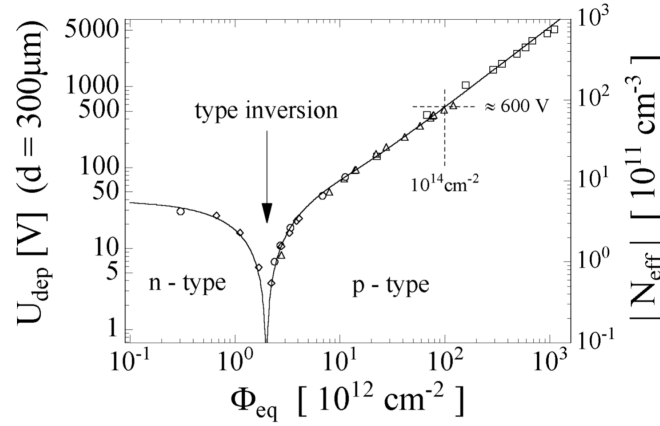


Figure 6.12: Type inversion of a n-type bulk into a p-type bulk [183].

Charge trapping In the band gap, charge carriers can be trapped by the new states created due to defects in the bulk. The typical trapping time τ_{trap} is inversely proportional to the fluence. As a result, a loss of charge collection efficiency with the fluence is observed. This effect can be very important in some HEP application where the sensors are exposed to a very high radiation fluence.

6.4 Summary

The basics of silicon sensors have been presented in this chapter. They are crucial to understand the performance of the silicon detectors in general. The pixel sensor and its properties

have been focused, including the radiation damage. The pixel sensor will be focused on in Chapter 7 in the context of the ATLAS Inner Tracker.

PLANAR PIXEL MODULES FOR THE ATLAS INNER TRACKER (ITk)

The Inner Tracker (ITk) of the ATLAS detector is a complete new silicon tracker designed for HL-LHC to cope with expected event rate and track multiplicity and with the radiation hardness requirements, in order to maintain the same tracking performance as ATLAS has now. The ITk installation is planned during the next LHC long shutdown to start in 2026 and to last for 3 years. Currently the ITk project focuses on the detector design, testing and protocols definitions aspects for the production phase. The design and the requirements of ITk are reviewed in Section 7.1. Then, Section 7.2 gives an overview of the activities of LPNHE in the context of ITk and my contributions. Next, Section 7.3 focuses on the detailed descriptions of the planar pixel modules, an item to which I have given my major contributions. Afterwards, the facilities, setup and software tools used to characterize the modules are presented in Section 7.4 followed by the test-beam results in Section 7.5.

7.1 ATLAS Inner Tracker (ITk)

As described in Section 2.1, high luminosity means a higher event rate, more pile-up events, and higher radiation dose and fluence, which implies a much harsher environment than the present running conditions for the ATLAS sub-detectors. The innermost pixel layers will receive a radiation fluence of $10^{16} - 10^{17} \text{ n}_{eq}/\text{cm}^2$ (4-6 times the IBL fluence in LHC Run 3) during data collection of the integrated luminosity of 4000 fb^{-1} [189]. During HL-LHC, ATLAS will have to face a number of superimposed interactions of roughly 200. More information about HL-LHC conditions can be found in Section 2.1. Hence, to cope with the severe data-taking environment expected at the HL-LHC, the ATLAS detector will be fully upgraded, including a new silicon tracker: the Inner Tracker (ITk).

The proposed ITk layout is shown in Figure 7.1. ITk consists of inner pixel layers and outer strip layers. The beam pipe with a maximum external radius of 28.2 mm will be reused from Run 2 of the LHC [190]. The inner pixel layers will cover a pseudorapidity range down to $|\eta| = 4$. A quadrant of the inner pixel layers is depicted at the bottom of Figure 7.1. It consists of an inner replaceable system and an outer permanent system. The inner barrel and endcap (Yellow) consist of pixel layers 0 and 1. The outer barrel (Green)

consists of pixel layers 2-4. The outer endcaps (Blue) consist of rings for the outer system in the forward region. The outer strip region comprises two short strip regions, two long strip layers, and one endcap on each side with six disks. The strip system covers up to $|\eta| = 2.7$. Including the strip detector, the total length of the ITk in the z-direction is $z = 661.0$ cm.

To maintain the current performance in the HL-LHC regime, a thinner and more granular detector is required. The ITk pixel sensor design [191] includes both planar pixels (for Layers 1-4) and 3D pixels (for the innermost Layer 0, due to their higher cost). For planar pixels a sensor thickness of $100\ \mu\text{m}$ is chosen for Layer-1 and $150\ \mu\text{m}$ for the others. The pixel geometry is $50 \times 50\ \mu\text{m}^2$ for the planar pixel modules. For the 3D sensors, the pixel size is $25 \times 100\ \mu\text{m}^2$ in the flat barrel and $50 \times 50\ \mu\text{m}^2$ in the endcaps.

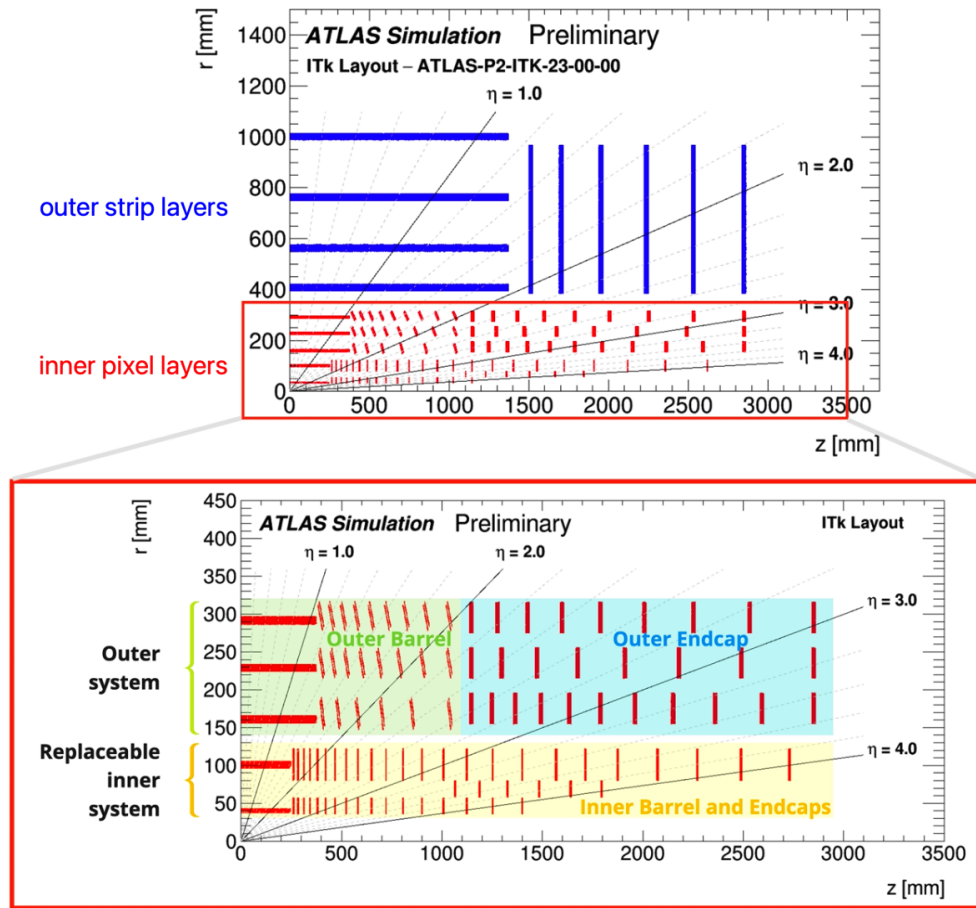


Figure 7.1: Schematic layout of the ITk for the HL-LHC phase of ATLAS. Here only one quadrant and only active detector elements are shown. The horizontal axis is the axis along the beam line with zero being the interaction point. The vertical axis is the radius measured from the interaction point [184].

7.2 LPNHE within the ITk project (and personal contributions)

The ATLAS group of LPNHE has traditionally been involved in the design and development of pixel sensors for high energy physics. Along the years we have studied several technical solutions for pixels of future detectors, with special attention to the R&D for the ATLAS ITk. Many different prototypes were produced. They can be roughly summarised as follows:

- Sensors with different thickness. The thickness of sensors in LPNHE R&D production is 50, 100, 130 and 150 μm . This experience has been used in the ITk sensor design, where the 150 μm thick sensors of the outer layers, and the 100 μm thick sensors of the Layer-1, designed in collaboration with our group, represent a large improvement with respect to the sensors used previously in the ATLAS pixel system and in the IBL. The 50 μm thick sensors are promising candidates for future colliders as they have low material and better expected radiation hardness.
- Sensors with different layouts. The pitch sizes of the LPNHE R&D productions are mainly $50 \times 50 \mu m^2$ and $25 \times 100 \mu m^2$. Different bias structures combining with different pitch sizes are also considered.
- Active edge sensors. It is one of the innovative techniques we have studied with an eye to the ITk but also to other applications like imaging or medical physics, where sensor tiles cannot be staggered to reduce the dead zones. Two active edge productions have been done by the LPNHE group in collaboration with FBK Trento, and I had a primary role in the characterisation of the second one, designed in the framework of the AIDA-2020 European project [192, 193].

The test and the characterisation of these prototypes has been performed and contributed by the researchers and the PhD students in the group. I am one of the contributors. The modules I have tested are presented in this Chapter, including three modules with 130 μm thickness, four modules with 50 μm thickness, 2 modules with active edge technology. The modules cover different sensor layouts at different fluences of radiation. The details of the modules are presented in Section 7.5.3 for active edge modules and Table 7.2 in Section 7.5.1 for other R&D modules.

After the design of the sensor pixels for the ATLAS ITk, the collaboration entered the phase of the evaluation of possible vendors through a process managed by CERN on behalf of the collaboration and called market survey. Prototyping productions from several vendors have been evaluated through electrical measurements and beam tests before and after irradiation campaigns. Our group, as one of the groups participating to the design, has been selected as one of the natural candidate groups to perform this characterisation. As I will discuss later in Section 7.5.2, I contributed to the test-beam operations and to the subsequent data reconstruction, which allowed us to determine the properties of the sensors under evaluation and to validate the participating vendors.

7.3 Planar pixel module

7.3.1 RD53A readout chip

The RD53A integrated circuit [194] is the first prototype chip developed by the RD53 collaboration for the high hit rate and high radiation tolerance for both ATLAS and CMS at HL-LHC. All pixel modules studied in this thesis are RD53A modules. The 65 nm CMOS technology is chosen for RD53A. Design variations in the analog part of the chip are also contained for testing purposes. The chip pixels are arranged in the matrix of 400×192 (column $\times$ row). The dimension of the active area is $20.0 \times 9.6 \text{ mm}^2$. This is half the size of the final version of the ITk chip. The pixel cell size of the chip is reduced to $50 \times 50 \mu\text{m}^2$ to maintain the occupancy at the desired level in the HL-LHC regime. Three different versions of the analog front-end (FE) were implemented in the RD53A chip to study and compare them in one single readout chip: synchronous FE, linear FE and differential FE. The details about the FEs and signal processing can be found in [194]. The differential FE is chosen for ATLAS ITk in the updated versions of the chip. For ATLAS the RD53A has been followed by the ITkPixV1 [195] successor, while the final version used in the ITk will be a further evolution called ITkPixV2, with basically the same architecture and small improvements.

7.3.2 Sensor layouts

The n-in-p process is chosen for ITk planar pixel sensors. As already mentioned, the thickness of the sensors is chosen to be 100 or 150 μm . The sensors used for the design development matched the array of RD53A. The sensor's design is shown in Figure 7.2. The pixel cells are $50 \times 50 \mu\text{m}^2$ (top) and $25 \times 100 \mu\text{m}^2$ but the position of the bump pads is such that both of them can be flip-chip to the readout chip. The designs on the left are relative to devices where a biasing structure has been implemented, with a punch-through dot common to four-pixel cells. On the right side, pixel cells without a biasing structure are shown.

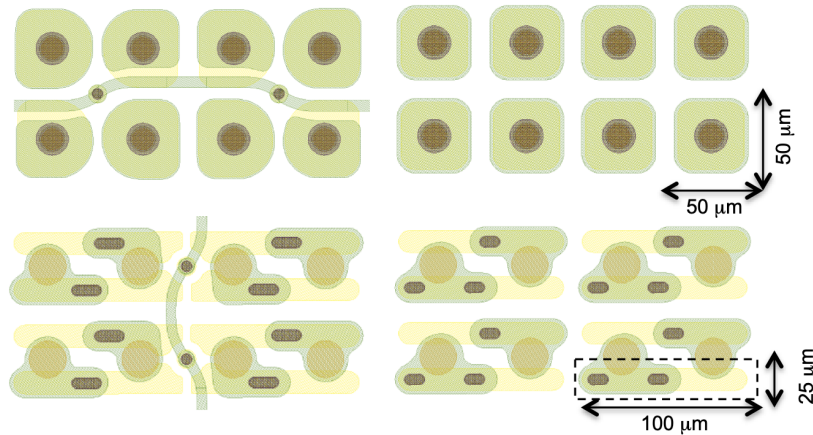


Figure 7.2: Design for $50 \times 50 \mu\text{m}^2$ (top) and $25 \times 100 \mu\text{m}^2$ (bottom) pixel cells, compatible with the RD53A chip. The structures colored in yellow represent the n^+ implants, in green the Aluminum layer, while the brown circles show the position of the UBM pads. [194]

7.3.3 Hybridisation

The unit of ITk pixel detector is a hybrid pixel module. As described by Figure 6.6 in Section 6.2.1, the hybrid refers to the property of sensors being coupled to their electronic readout via an array of small metal bonds, one per channel. Hybridisation is the high-density inter-connection technique to connect the pixel sensor to the front-end chip. The hybridisation process consists of two steps. First, the bumps are deposited on the pads of the FE chip to connect. This operation is performed at the wafer level. It is usually coupled with other wafer processes, particularly the deposition of a metal layer in the bump area (UBM, under bump metallisation) on both the FE wafer and the sensor wafer. Then the wafers are cut, and the individual dies are connected using a “flip-chip” technique. The hybridisation is challenging for the ATLAS ITk with respect to the present ATLAS pixel system for two reasons: 1) The requirement is more stringent due to reduced wafer thickness for assemblies targeting a front-end chip thickness of $150\ \mu\text{m}$ and a sensor thickness of 100 to $150\ \mu\text{m}$ compared to the ones for IBL module where the front-end chip and sensor are $190\ \mu\text{m}$ and $250\ \mu\text{m}$ respectively. 2) The required bump density is five times more than the IBL FE-I4 chip due to the pixel pitch of $50\ \mu\text{m}$ in both directions. The ATLAS ITk pixel chip has 400 columns and 384 rows, giving a total of 153 600 bump bonds per chip, compared to 26 880 bonds for an FE-I4 chip in IBL.

7.3.4 Module assembly

A hybrid module in the configuration we used to develop and characterise the sensors using Single Chip Cards (SCC), has three main components: sensor, chip and Printed Circuits Board (PCB). Their simplified schematics are shown in Figure 7.3. The sensor and the front-end chip are connected via hybridisation process, see Section 7.3.3. The PCB is a carrier to connect power for front-end chip and sensor biasing and data transmission. The front-end chip is glued to the PCB, while the back of the sensor remains exposed. The utilized glue must be robust against radiation, environmental conditions and thermal stress. Wire bonds provide the electrical connection between the front-end chips and the PCB. They are used for command, data communication, powering, and grounding, while a wirebond connects the back of the sensor to a PCB pad to provide the bias voltage. Figure 7.4 shows a real, well-assembled SCC used in the thesis. The main components of the card are labeled in the picture. All the modules tested in the thesis are SCCs. The RD53A module should be powered at 1.2V for analog and digital front-ends together via direct powering mode [194]. To bias the sensor, the module should be connected to a high voltage power supply, e.g., Keithley 2410, which provides the sensor bias voltage and measures the sensor’s leakage current. Please note that only the assembly of the SCC which is used for testing is mentioned above. In the real modules of ITk, a flex is used instead of a PCB. The flex is glued on the back of the sensor and wirebonded to the FE chip.

7.4 Characterisation of planar pixel modules

One of the most important characterisation tools for our modules is a test-beam. During test-beams, a particle beam from a particle accelerator illuminates a device under test, and

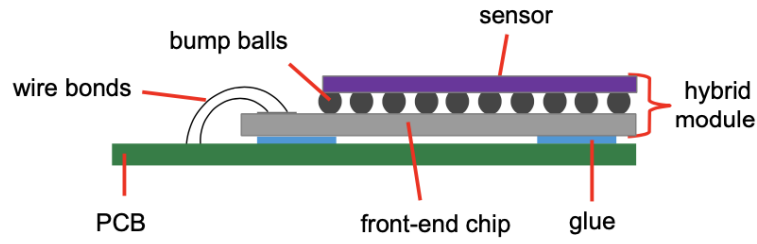


Figure 7.3: Cross sectional schematic view of a module consisting of a single sensor and a readout chip bump bonded together, glued and wire bonded into a PCB. [184]

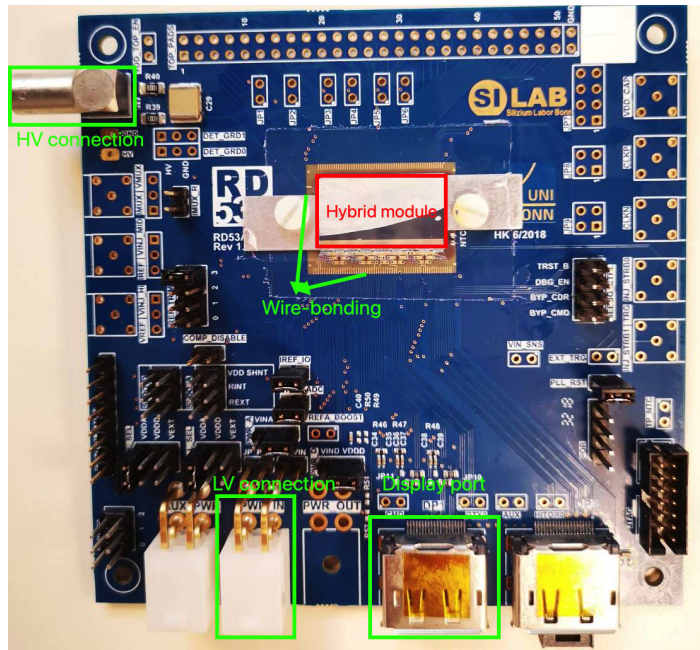


Figure 7.4: A RD53A single chip card.

data is collected for further analysis. This section explains essential operations to do during a test-beam campaign.

7.4.1 I-V curve measurement

The I-V curve measures the leakage current of the sensor as a function of the bias voltage. As introduced in Section 6.1.2, it is fundamental to identify the range of operational limit of the sensor as well as the breakdown voltage, which helps selecting good sensors that could be used for production. The I-V curve differs from one sensor to another. For the same sensor, the I-V curve also depends on the received fluence, thickness, temperature, etc. The I-V curve can be measured at the wafer level and at module level. At test-beams, the I-V curves are measured at module level. The setup is trivial: the module is connected to a voltage supply e.g., Keithley 2410 electrometer through the HV connection as shown in Figure 7.4. The electrometer reads the leakage current at each voltage step. Usually, to gain more data-taking times during the test-beams, the I-V curves are measured at local

laboratories with conditions as similar as possible to the ones during the test-beams. For unirradiated modules, the I-V curves are measured at room temperature; For irradiated modules, they are measured at $-50\text{ }^{\circ}\text{C}$.

During the ITk sensors R&D, we have tried different sensor thicknesses and module W398_1 is an example of an intermediate thickness. It was produced from $130\text{ }\mu\text{m}$ thick wafer with $50\times 50\text{ }\mu\text{m}^2$ pixel geometry without punch through (PT). Figure 7.5 shows the I-V curve of module W398_1 before and after irradiation. Unfortunately, while we have data at wafer level, only a few bias voltage steps were taken before the module got irradiated, and there is no record of the breakdown voltage value after hybridisation. Nevertheless, the sensor is already over-depleted at 140 V before irradiation. The module was irradiated to the fluence $\Phi = 1.7 \times 10^{15}\text{ n}_{eq}/\text{cm}^2$. After irradiation, the leakage current is doubled than before irradiation and reaches $16\text{ }\mu\text{A}$ at 600 V . The operation voltage range where the sensor is fully depleted is chosen when the leakage current is linear and stable, up to the breakdown voltage. Besides, for irradiated modules, we will not go beyond 600 V for safety reasons. The breakdown does not happen even with an extended operation limit at 700 V .

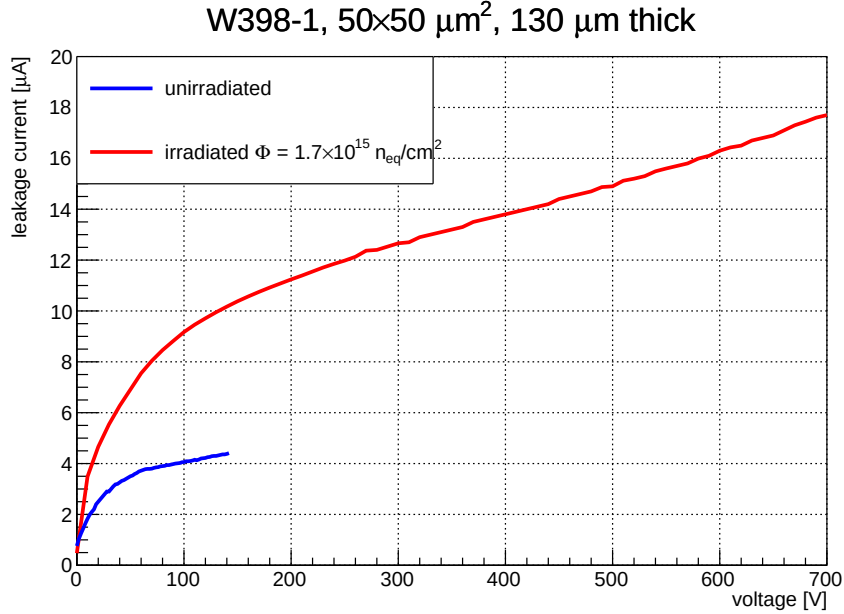


Figure 7.5: The I-V curve of module W398-1 before and after irradiation. Pixel geometry: $50\times 50\text{ }\mu\text{m}^2$, $130\text{ }\mu\text{m}$ thick.

As we introduced in Section 7.2, LPNHE has also developed $50\text{ }\mu\text{m}$ thick sensors, and some results are presented in this Section. Reduced sensor thickness leads to reduced leakage current, and the operational bias voltage is lower. The thinner sensor is also more tolerant to an increased radiation fluence. A pixel geometry of $50\times 50\text{ }\mu\text{m}^2$ is used. Different biasing structures are implemented: no PT, straight PT and wavy PT. The structures are illustrated in Figure 7.6. The different bias structures create different patterns of inefficiency in the sensor. Another interesting feature of these sensors is the width of the guard ring (GR) region around the pixel matrix. As we introduced in the previous chapter, the guard rings are used to control gradient of the high potential from the edge to the last pixel column/row. A wider band of guard rings would be able to control the decrease of high po-

tential present at the last pixel column/row. Thus, the sensors with a larger GR width can be safely operated at higher bias voltage values before the breakdown occurs at the edges of the sensors. GRs are important after irradiation, where a higher voltage is required to fully deplete the sensor. Figure 7.7 shows the IV-curves for 50 μm thick modules. The module without PT has the highest breakdown voltage because the potential gradient in the sensor is flatter without PT than with PT. It is followed by modules with wavy PT. The modules with straight PT have the lowest breakdown voltage. Comparing W50_17 and W50_105, the breakdown voltage increases from 250 V to 380 V when the GRs width increases from 250 μm to 450 μm . The modules with PT have a higher leakage current than those without PT. The 50 μm thick modules have a lower leakage current compared with W398_1 with the thickness of 130 μm as expected. At 50 V, The leakage current of W398_1 is four times higher than W50_11 before irradiation. W50_11 and W50_105 are irradiated to fluence of $\Phi = 5.3 \times 10^{15} n_{eq}/cm^2$. The leakage current for these two modules is not as stable as before irradiation.

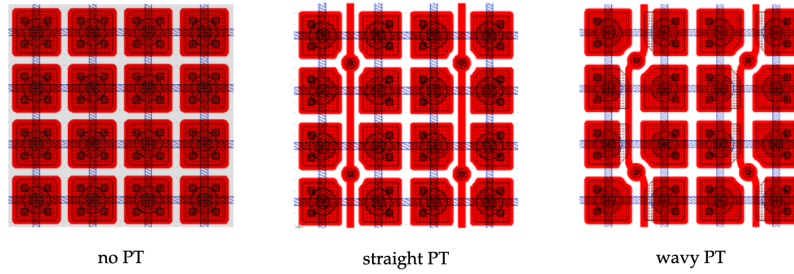


Figure 7.6: Different layouts of pixel sensors for thin 50 μm thick sensors.

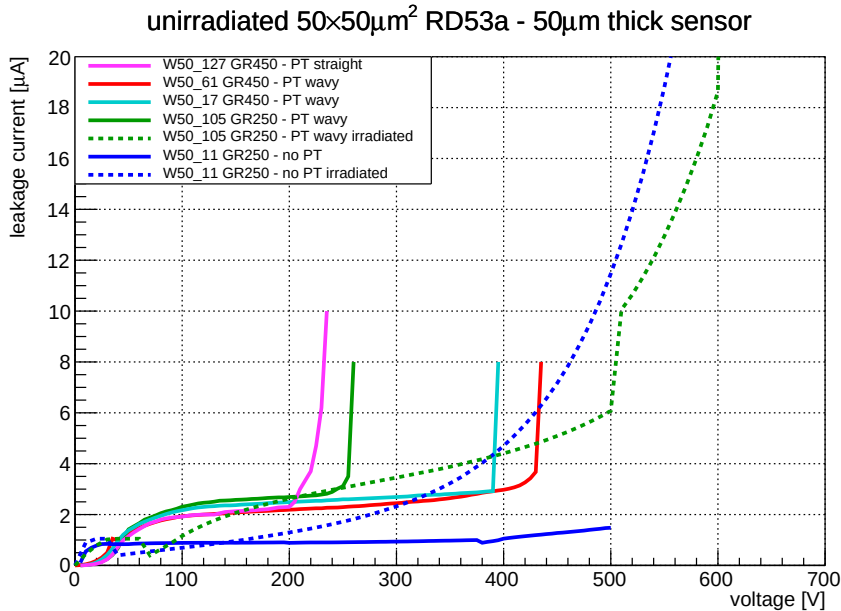


Figure 7.7: The IV-curves for 50 μm thick sensors before irradiation and after irradiation. Pixel geometry: 50×50 μm^2 , 50 μm thick. Irradiation fluence: $\Phi = 5.3 \times 10^{15} n_{eq}/cm^2$

7.4.2 Pixel module calibration

In order to have a homogeneous response over the whole pixel matrix, the front-end chip of the modules must be calibrated. The calibration can account for small differences in the analog behaviour of the FE chip, channel by channel, and for differences in the sensor from one pixel to the other. At local laboratories, tuning is helpful to check if the modules are operational or not. Modules should be calibrated again before data-taking since the tuning results will be impacted by the environmental conditions, like temperature and humidity. The setup used at LPNHE for module tuning is illustrated by Figure 7.8. The module is measured inside a climate chamber for shielding from light and to provide constant temperature and humidity. Unirradiated modules are tuned at 20°C, and irradiated modules are kept at low temperature and tuned at -50°C . The HV and LV power supplies are detailed in Section 7.3.4 for RD53A modules. The module is read out using the YARR system [196]. The chip tuning is done step-by-step: digital scan, analog scan, threshold tuning, Time over Threshold (ToT) tuning, and noise occupancy scan.

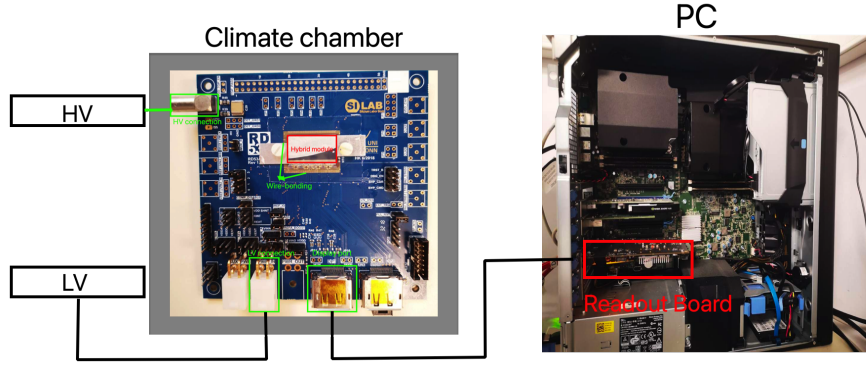


Figure 7.8: Illustration of the setup for module calibration at LPNHE.

Digital scan In a digital scan, a number of digital hits is injected to simulate the discriminator's output. Any disfunction along the readout chain after the pixel discriminator is revealed if the number of hits for each pixel differs from the number of injections. The malfunctioning pixels are disabled in the module.

Analog scan In an analog scan, a certain number of charges is injected N times for each pixel. If the pixel is dead, it will show 0 occupancy. If the pixel is noisy, it will go over threshold a number of times which is larger than expected. From the analog scan, the response of the readout chip can be tested. The threshold also plays a role, the analog scan is usually performed at least twice, i.e. before and after the threshold tuning, in a calibration chain.

Threshold tuning The signal can be registered by a pixel only if the collected charge is above the set threshold. The threshold of each pixel must be tuned to avoid unnecessary readout due to noise and have a stable hit detection and reliable charge/ToT measurements everywhere in the whole chip. The threshold tuning is performed as a multi-scanning process. A well-defined charge is injected in each scan. The charge value above the threshold results in a hit. The discriminator has a 0 and 1 response that correspondingly results in a digital square wave function. Due to the electronic noise and other factors, the resultant re-

sponse of the discriminator can be modelled as a step function convoluted with a Gaussian distribution. The threshold is defined to be the charge for which 50% of injected hits are recorded. A successful threshold tuning is achieved by injecting the charge for each pixel and producing a response with a narrow and symmetrical Gaussian distribution around the target value. For example, a typical threshold target for 100-150 μm thick sensor is 1000e. If the sensor is irradiated, a higher threshold might be needed due to the high noise level.

Time over Threshold (ToT) tuning The tuning of the ToT is similar to the tuning of the threshold. The mean ToT is tuned to the expected signal amplitude generated in the sensor. A typical ToT tuning for 100-150 μm thick sensor targets 8 ToT for a 10 ke injected charge. ToT time unit corresponds to 25 ns, bunch crossing time of collisions in LHC.

Noise occupancy scan In the noise scan, random triggers are sent to the chip and the number of hits from each pixel is counted. Pixels that are fired at a more than expected rate of hits due to the global noise level are disabled during data-taking. The noise occupancy is required to be 10^{-5} or below for the ATLAS pixel detector not to influence the tracking efficiency [191]. This scan is performed after tuning and before data taking for all modules.

7.4.3 Test-beam campaigns for characterisation

7.4.3.1 Test-beam setup

The test-beam setup is shown in Figure 7.9. The beamlines, the tracking telescope and data acquisition systems are described in the following.

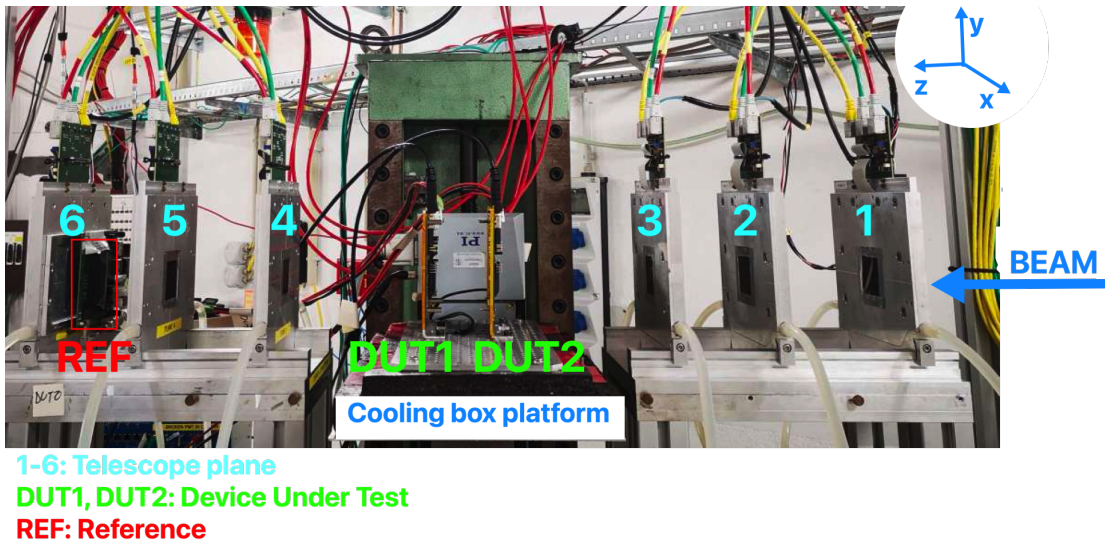


Figure 7.9: Picture of test-beam setup with EUDET telescope at DESY.

CERN SPS beam The beam at CERN ¹ [197] used to test our modules is the 120 GeV pion beam, which is originated from the SPS.

¹Unfortunately, we have not managed to have useful data form CERN SPS beam for our modules under test due to the long SPS shutdown.

DESY beam The beam provided by DESY ² [198] is 3-6 GeV electrons with a quasi continuous flow.

EUDET telescope At CERN and DESY, the telescope we used for tracking studies is the EUDET telescope [199]. It consists of six detection planes (3 upstream, 3 downstream) equipped with the Mimosa26 monolithic pixel sensors. A Mimosa26 sensor [200] has thickness of $50 \mu\text{m}$. Each sensor has pixel size of $18.4 \times 18.4 \mu\text{m}^2$ arranged in $1152 \text{ columns} \times 576 \text{ rows}$ making an active area of $21.2 \times 10.6 \text{ mm}^2$. The data read out was triggered by the coincidence of plastic scintillators, whose areas were about 1 cm^2 . The Data Acquisition software for the EUDET telescope is EUDAQ.

After data-taking, the track reconstruction and analysis were performed with Eutelescope [201] and TBmon2 [202], respectively.

7.4.3.2 EUTelescope

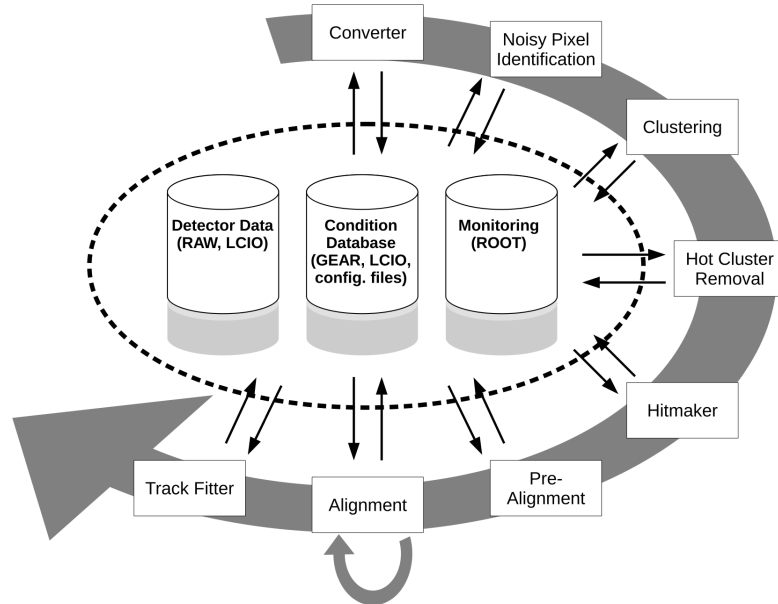


Figure 7.10: Workflow of Eutelescope. [201]

The EUTelescope is a modular, comprehensive software framework for reconstructing particle trajectories recorded with beam telescopes [201]. Figure 7.10 shows different steps to process raw data into tracks. Initially, data is converted from the native format of the DAQ system to the format used in the framework: LCIO format. Then, noisy pixels are tagged if they fire at a frequency higher than a certain threshold. The step creates a noisy pixel database which is the foundation of any subsequent noise treatment. Next, the reconstruction groups adjacent pixel hits into clusters. Different algorithms can deal with binary and non-binary readout as well as equidistant and non-equidistant geometries. Using the noisy pixel database, clusters containing at least one noisy pixel can be masked and removed, resulting in a noise-free cluster collection. Hit positions in the local reference frame (the sensor reference frame) are thus derived based on the cluster shape. These

²All test-beam results in this thesis are from DESY beam.

hit positions are used to determine a preliminary alignment by investigating the correlations between the upstream sensors and the downstream sensors in the beam telescope as well as DUT sensors. Starting from the pre-aligned setup, the next step derives correction constants to the alignment using General Broken Lines (GBL) track fitting method [203], which takes into account the multiple scattering³. This step might be repeated to converge towards a final set of alignment constants iteratively. The last reconstruction step produces the final track fits and possibly exports the tracks for use in an external analysis framework.

7.4.3.3 Hit residuals and spatial resolution

The study of the modules' performance requires a good understanding of the track reconstruction, which provides the inputs. The hit residuals and the spatial resolution are two of the most important variables we use during the track reconstruction.

The **binary resolution** [204] can primarily describe the spatial resolution of the sensor and is usually used as input for the track reconstruction. It is calculated with the simplest choice of the readout mode (i.e., single threshold binary readout) for the pixel sensor, i.e.:

$$\sigma_{binary} = \frac{pitch\ size}{\sqrt{12}} \quad (7.1)$$

Table 7.1: Binary resolution for Mimosa planes and DUTs with each pixel geometry.

Name	Pixel geometry [μm^2]	σ_x [μm]	σ_y [μm]
Mimosa	18.4 × 18.4	5.3	5.3
DUT	50 × 50	14.4	14.4
	25 × 100	28.9	7.2

Table 7.1 summarises the binary resolution of Mimosa planes of the telescope and DUTs with each pixel geometry. The **intrinsic resolution** (σ_{int}) of the sensor can be better than the binary resolution of the pitch when the average cluster size contains more than one pixel indicating some charge sharing between pixels. Besides, the pointing resolution of the telescope planes at the location of the DUTs are also taken into account in the σ_{int} . The pointing resolution is primarily affected by the multiple scattering. More discussion about the pointing resolution can be found in Ref. [205]. In a short summary, the higher the beam energy is, the smaller the effect of the multiple scattering is; the thinner the sensor is, the smaller the effect of the multiple scattering is.

The hit residual is defined as the difference between a hit position in a plane under study and the position of the intersection between the reconstructed track associated to the hit and the plane. The standard deviation of the residual distribution, i.e., residual width, indicates the spatial resolution of the sensor, folded with multiple effects such as the choice of the readout mode, the reconstruction algorithm, the amount of charge sharing between pixels, multiple scattering.

³When passing through matter, a charged particle is subject to a large number of scattering events mostly due to Coulomb interaction with the nuclei.

If the hits from the sensor are included in the track fitting procedure, the residuals are **biased**. The multiple scattering effect is thus taken into account in that sensor. If the hits from the sensor are NOT included in the track fitting procedure, the residuals are **unbiased**. The multiple scattering effect is NOT taken into account in that sensor.

Figure 7.11 shows the biased residuals for EUDET telescope planes measured at DESY with a particle beam energy of 5 GeV. The residual width is smaller than σ_{binary} due to the charge sharing among pixels and varies from 3.0 to 4.0 μm according to the position of planes. Figure 7.12 shows an example of unbiased residuals of measured DUT modules at DESY for different pitch sizes. The residual widths are twice larger than σ_{binary} for sensors with pixel geometry of $50 \times 50 \mu m^2$. For sensors with pixel geometry of $25 \times 100 \mu m^2$, the residual width in the x-direction is 30% wider than the binary one, and the residual width in the y-direction is 4 times larger than the binary one. This is due to the large impact of the multiple scattering at this beam energy.

7.4.3.4 TBmon2

TBmon2 was the ATLAS Pixel test-beam analysis framework [202]⁴. It operates on output files produced by the EU Telescope. The architecture of TBmon2 is shown in Figure 7.13. A test-beam analysis usually consists of multiple TBmon2 analyses, which can be configured through a dedicated analysis configuration. The DUT geometry for every device is included in the corresponding DUT configuration file. The user needs to set the central configuration parameters, like the selection of fiducial regions to accept or reject tracks, track quality criteria, or reference plane assignment. Once TBmon2 is executed, it will select a set of good tracks and hits defined by the criteria set by the user. All tracks and hits that pass the preprocessor checks are described as ‘good’ and used for the efficiency analyses. Some important feature of this analysis are:

Global efficiency A track is considered good if it has a matched hit in the reference plane. A hit in a DUT plane is considered good if it is within the matching radius of a good track and if it belongs to a cluster that does not contain noisy or hot pixels or if it is not masked. The efficiency ϵ is defined as:

$$\epsilon = \frac{n_{hits}}{n_{tracks}} \quad (7.2)$$

and its uncertainty σ_ϵ is defined as:

$$\sigma_\epsilon = \sqrt{\frac{\epsilon(1 - \epsilon)}{n_{tracks}}} \quad (7.3)$$

The global efficiency is calculated for the entire sensor area.

In-pixel efficiency Instead of being calculated over the entire sensor area, the in-pixel efficiency is calculated for individual pixels. The hits and tracks from all pixels of the same geometry are mapped into one so-called in-pixel maps. It allows us to have a detailed view inside the pixels with a high statistics.

⁴A new framework called Corryvreckan [207] is currently adopted.

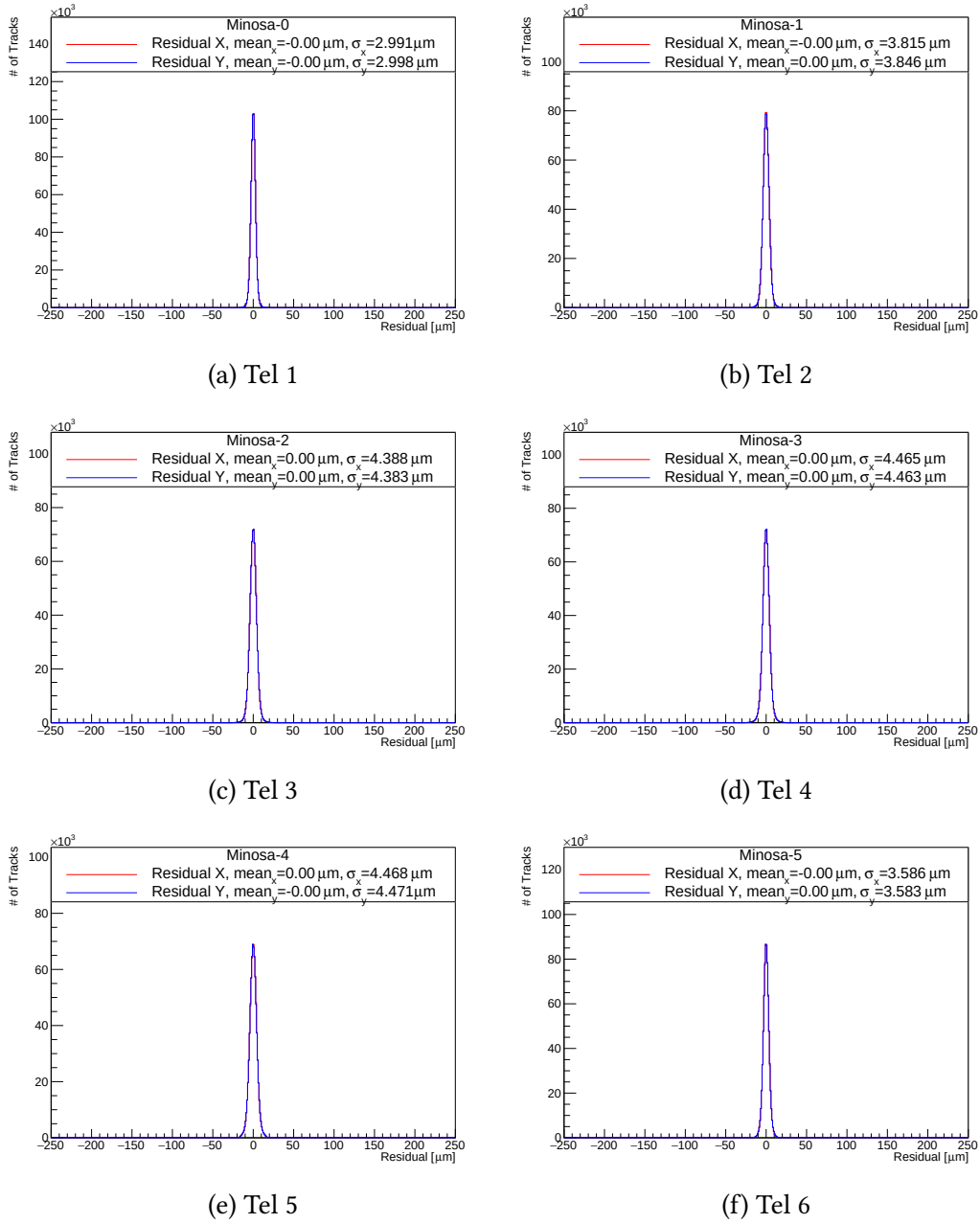


Figure 7.11: Residuals for the EUDET telescope planes measured at DESY with a particle beam energy of 5 GeV. Residuals on the x and y-axis are superimposed for the same telescope plane.

Edge efficiency The edge efficiency focuses on the area between the last-pixel rows/columns and the doped trench for active-edge sensors. An efficiency measurement as a function of the track position in the edge area is performed, using data collected with the beam focused on the edge area.

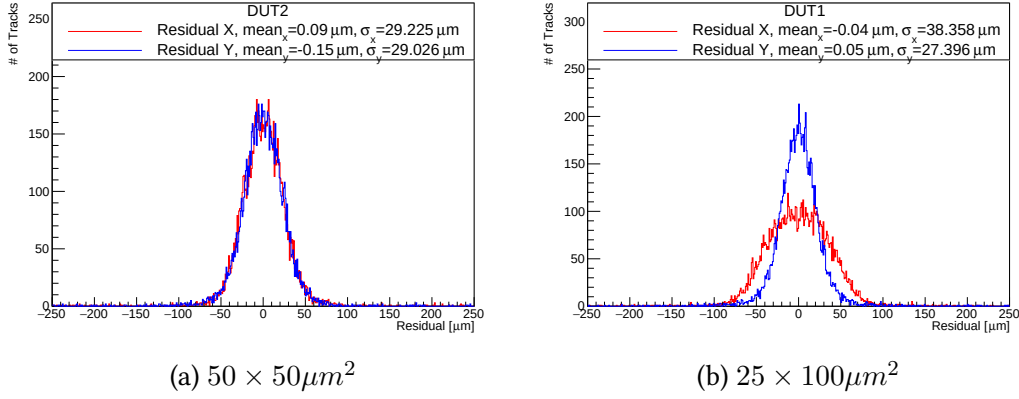


Figure 7.12: An example of the residuals of measured DUT modules at DESY for different pitch size: $50 \times 50 \mu\text{m}^2$ (left) and $25 \times 100 \mu\text{m}^2$ (right).

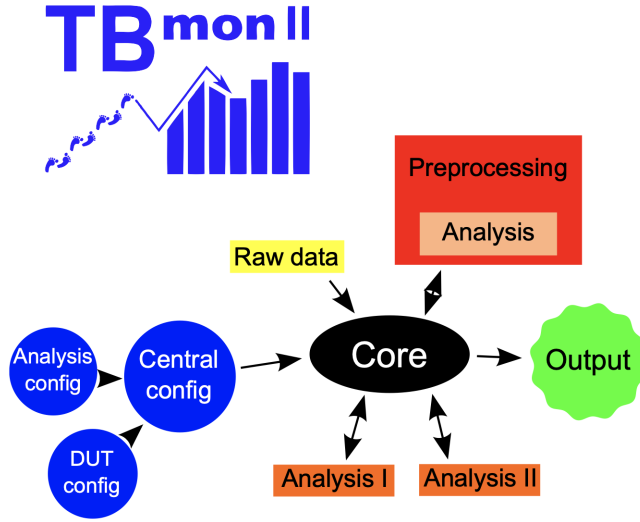


Figure 7.13: Schematic overview of the working principle of TBmon2 [206].

7.5 Test-beam results for planar pixel modules

Several devices have been tested during my PhD for different purposes as introduced in Section 7.2. They are listed here and the test-beam results are presented in different sections:

- LPNHE R&D modules for ITk or future colliders, including 130 and $50 \mu\text{m}$ thick modules. These modules are built to converge on the design of ITk pixel sensors. The tested modules and results are presented in Section 7.5.1.
- ITk market survey modules. The market survey campaign is to validate the vendors for ITk modules. The tested modules and results are presented in Section 7.5.2.
- Active edge modules in the context of the AIDA program. The active edge technology is not used in the ITk but it is very interesting to study the performance for imaging and medical applications. The tested modules and results are presented in Section

7.5.3.

7.5.1 LPHNE R&D modules

In this section, we focus on the R&D ITk planar pixel modules of LPNHE, which were assembled from sensors that were produced at FBK. The modules includes:

- Modules with 130 μm thick sensors. They are modules to converge on the sensor design of ITk pixel modules. The measurements are made on modules both before and after irradiation.
- Modules with 50 μm thick sensors. They are radiation-hard modules for future colliders. The measurements are only done with unirradiated modules. The devices have been irradiated later but the test-beam data was not in time to be included into this manuscript.

All results mentioned in this section have been achieved together with Dr. Reem Taibah, a former PhD student at LPNHE. The results are summarised in Table 7.2. The tested 130 μm thick modules have $25 \times 100 \mu m^2$ and $50 \times 50 \mu m^2$ pixels with different bias structures. One of the modules is irradiated at the fluence of $1.7 \times 10^{15} n_{eq}/cm^2$. The modules have efficiency higher than 97%. It is pitiful that the data collected (with my contribution) are not enough for thorough discussions. However, more discussions are presented in Dr. Reem Taibah's thesis [184]. The 50 μm thick sensors are developed for future colliders. The flip-chip technique is more challenging with such thin sensors as we mentioned in Section 7.3.3. All tested 50 μm thick sensors have $50 \times 50 \mu m^2$. Different bias structures are tested while only the results before irradiation are available due to the lack of beam availability during the COVID pandemic. The hit efficiency as function of the bias voltage for the s 50 μm is shown in Figure 7.14. W50_11, without PT structure, has higher hit efficiency compared to W50_105 as expected. When comparing the sensors with the same PT structure, W50_105 with narrower GRs region has higher hit efficiency.

Table 7.2: Summary of hit efficiency results for LPNHE R&D modules

Fluence [$\times 10^{15} n_{eq}/cm^2$]	Name	Pixel ge- ometry [μm^2]	Thickness [μm]	Biasing structure	Voltage (V)	Fiducial effi- ciency [%]
unirradiated	W360_4	25×100	130	TM	25	97.0 ± 0.6
	W398_4	50×50	130	PT	60	98.8 ± 0.1
1.7	W398_1	50×50	130	TM	600	97.7 ± 0.1
unirradiated	W50_61	50×50	50	GR450 PT wavy	100	95.7 ± 0.1
	W50_17		50	GR450 PT wavy	100	97.3 ± 0.2
	W50_105		50	GR250 PT wavy	100	98.9 ± 0.1
	W50_11		50	GR250 TM	100	99.0 ± 0.1

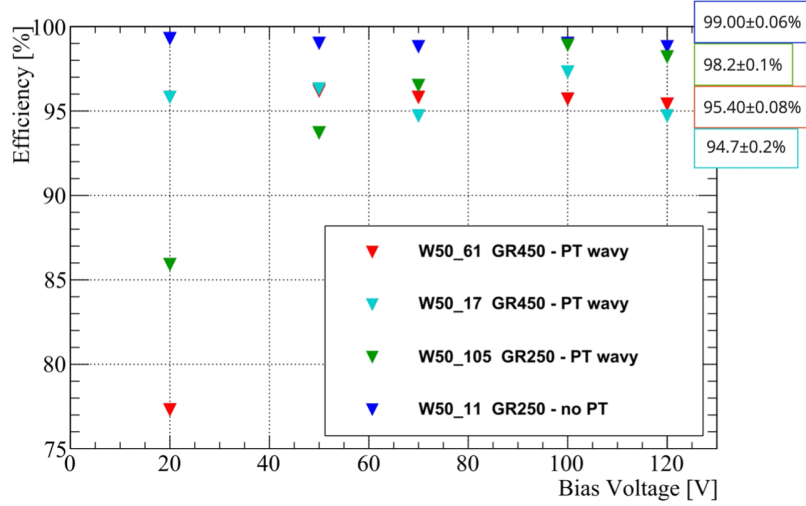


Figure 7.14: Hit efficiency as a function of the bias voltage for the sensors of $50 \mu m$ [184].

7.5.2 ITk market survey modules

I have also participated to many market survey test-beam campaigns for ITk pixel detectors. The market survey is aimed to qualify the potential suppliers during the prototyping phase. Modules from six different foundries have been tested. In order to avoid any bias on the qualification, all modules are labelled with numbers first instead of the foundry names and then assigned to different institutes for qualification. Nine institutes⁵ have been contributing to the test of the modules. More detailed information about the qualification process and requirements is documented in ITk Technical Design Report [191], the requirements for test-beam are summarised in Table 7.3. In this section, only the modules assigned to LPNHE and the hit efficiency measurements from test-beam are reported.

The results are summarised in Table 7.4. The last column of the table indicates if the modules meet the requirements. Four unirradiated modules are presented in the table, one fails the requirements. Seven modules irradiated at the fluence of $2 \times 10^{15} n_{eq}/cm^2$ are listed in the table, three of them fail the requirements (Table 7.3). Eight modules irradiated at the fluence of $5 \times 10^{15} n_{eq}/cm^2$ are presented, they all meet the requirements. For some modules, the fiducial efficiency as function of bias voltage is measured as we published in [12]. The results are shown in Figure 7.15.

⁵KEK, LAL, Lancaster University, LPNHE, TU Dortmund, University of Glasgow, University of Göttingen, University of Liverpool, University of Oxford

Table 7.3: ITk pixel market survey requirements. V_{bias} is the typical operational voltage before and after irradiation with different fluences.

Thickness [μm]	V_{bias} [V]	Fluence [$\times 10^{15} n_{eq}/cm^2$]	Efficiency [%]
100/150	$V_{dep} + 50$	Before irradiation	98.5
100	300	2	97.0
	400	5	
150	400	2	97.0
	600	5	

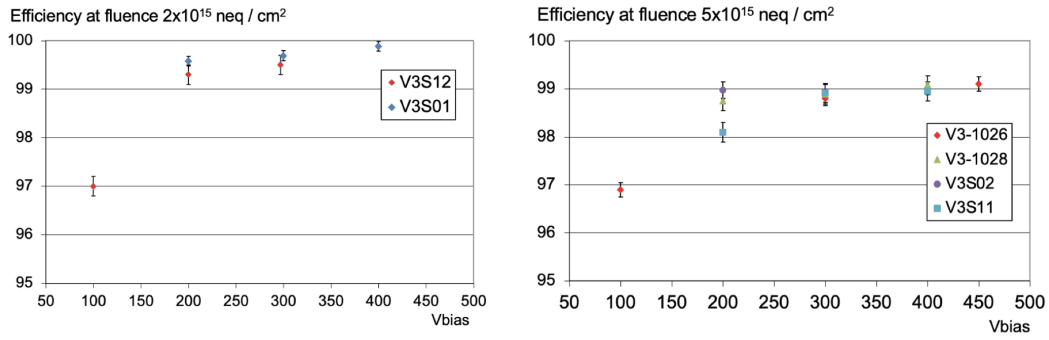


Figure 7.15: Hit efficiency as a function of the bias voltage applied to the sensor after irradiation at the fluence of $2 \times 10^{15} \text{ neq / cm}^2$ (left) and $5 \times 10^{15} \text{ neq / cm}^2$ (right).

Table 7.4: Summary of hit efficiency results from market survey modules. Modules were tuned to mean threshold of 1500e.

Fluence [$\times 10^{15} n_{eq}/cm^2$]	Name	Pixel geometry [μm^2]	Thickness [μm]	Biasing structure	Voltage (V)	Fiducial efficiency [%]	meet require- ment?
unirradiated	V3S13	50×50	100	TM	80	99.8 ± 0.1	yes
	V6S04		150	PT	50	97.57 ± 0.03	no
unirradiated	V3S05	25×100	150	PT	130	98.87 ± 0.1	yes
	V3S06		150	PT	80	99.94 ± 0.03	yes
2	V6S11	50×50	100	TM	300	99.73 ± 0.1	yes
	V1S01		150	-	400	94.93 ± 0.29	no
	V3S01		150	TM	400	99.88 ± 0.04	yes
	V4S01		150	-	400	98.93 ± 0.03	yes
2	V2S11	25×100	100	-	300	92.03 ± 0.4	no
	V2S05		150	-	400	99.43 ± 0.1	yes
	V6S02		150	-	400	91.74 ± 0.1	no
5	V1S02	50×50	100	PT	400	98.9 ± 0.0	yes
	V3S1x1028		100	PT	400	99.08 ± 0.11	yes
	V1S03		150	-	400	98.5 ± 0.01	yes
	V3S02		150	TM	400	99.91 ± 0.03	yes
	V4S03		150	-	600	99.86 ± 0.03	yes
5	V3S11	25×100	100	TM	300	98.89 ± 0.04	yes
	V3S1x1026		100	PT	300	98.53 ± 0.05	yes
	V6S03		150	-	400	97.20 ± 0.04	yes

7.5.3 AIDA active edge modules

Developing sensors with a reduced dead area is necessary to ensure an optimal geometrical acceptance of tracking detectors. Active edge sensors have a reduced dead area at their borders as presented in Section 6.2.1. The sensors I have tested are active edge sensors built at FBK in the framework of the AIDA-2020 collaboration. They are labelled G1 and G2 and are $150\text{ }\mu\text{m}$ thick. The geometry of the two sensors is detailed in Figure 7.16:

- AIDA-2020-G1: The pitch size is $25 \times 100\text{ }\mu\text{m}^2$. The thickness is $150\text{ }\mu\text{m}$. It is designed with PT wavy biasing structure and one grounded bias ring. There are 3 GRs of $150\text{ }\mu\text{m}$ wide between the last pixel and the trench.
- AIDA-2020-G2: The pitch size is $50 \times 50\text{ }\mu\text{m}^2$. The thickness is $150\text{ }\mu\text{m}$. It is designed with temporary metal technology. There are 4 GRs of $170\text{ }\mu\text{m}$ wide between the last pixel and the trench.

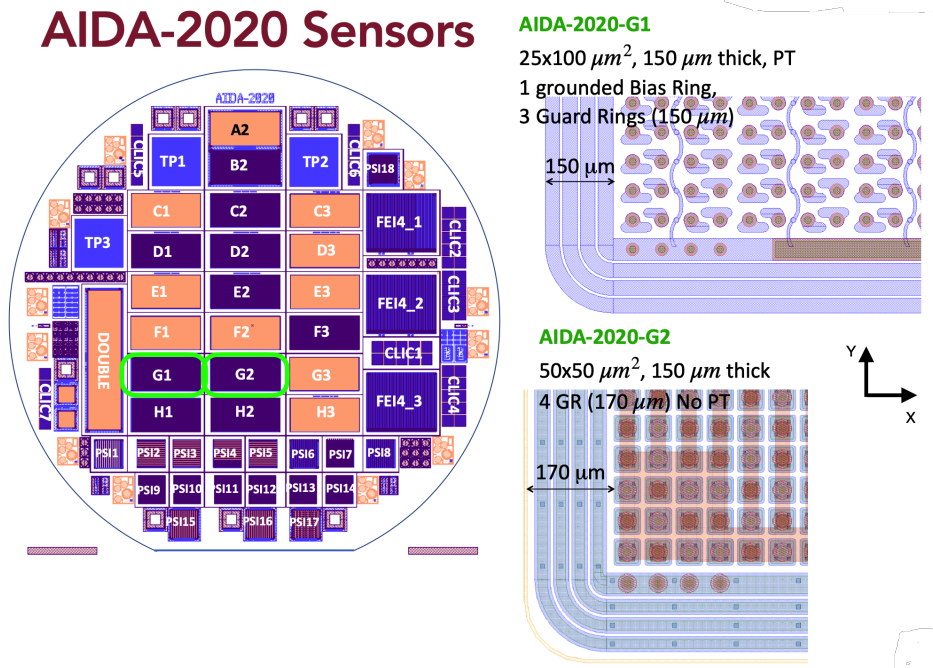


Figure 7.16: Schematic presentation of AIDA-2020 $150\text{ }\mu\text{m}$ thick wafer, AIDA-2020-G1 and AIDA-2020-G2. The orientation is illustrated.

AIDA-2020-G1 and AIDA-2020-G2 have been tested before irradiation in a test-beam at DESY. The beam energy was set to 5.2 GeV and the beam was focused on the edge of the sensors. Both of the sensors are tuned with a target threshold at $1500e$. Only the linear FE is used to ease the beam focusing operations. The test-beam results are presented as following:

Global efficiency Figure 7.17 shows the global efficiency of AIDA-2020-G1 and AIDA-2020-G2 as function of different bias voltage values. At 0 V , both of the modules have $\sim 40\%$ of global efficiency. At $>5\text{ V}$ bias voltage, they have $\sim 99\%$ of global efficiency.

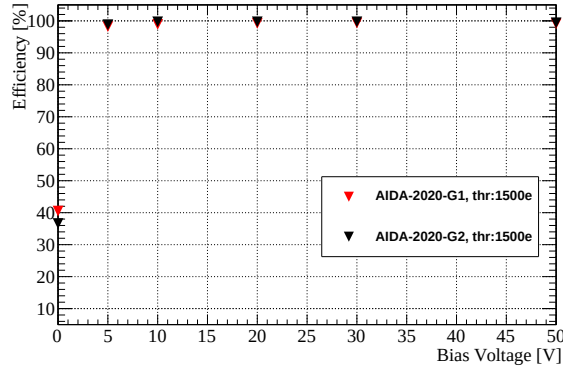


Figure 7.17: Global efficiency of AIDA-2020-G1 and AIDA-2020-G2 as function of different bias voltage.

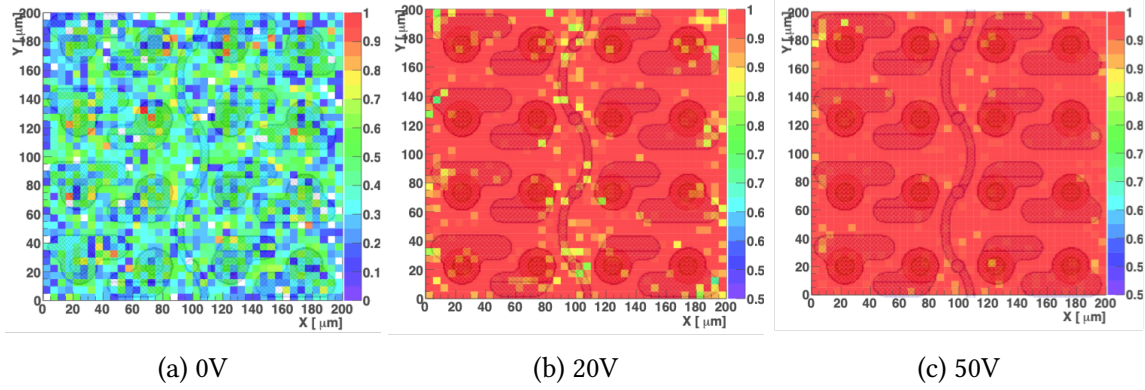


Figure 7.18: In-pixel efficiency map for AIDA-2020-G1.

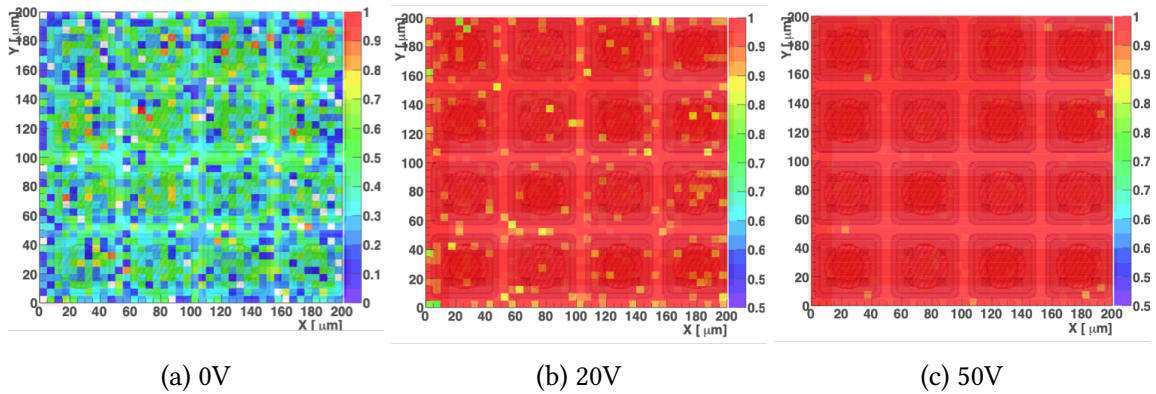


Figure 7.19: In-pixel efficiency map for AIDA-2020-G2.

In-pixel efficiency Figure 7.18 shows the in-pixel efficiency map for AIDA-2020-G1 at different bias voltages. The efficiency maps are overlapped with sensor pixel geometry map in order to have a better understanding of the measurements. At 0 V, the efficiency is about 20%-59% in the whole map. At 20 V, the fiducial region (by excluding the PT bias structure region) shows efficiency higher than 99%. Due to the presence of the PT dots and the electric field effect induced by the potential in the PT network, the PT bias structure

region appears less efficient than the fiducial region. At 50 V, due to high bias voltage, the PT bias structure region can reach a similar efficiency to the fiducial region. Figure 7.19 shows the in-pixel efficiency map for AIDA-2020-G2 at different bias voltage. Same as for AIDA-2020-G1, the efficiency maps are overlapped with sensor pixel geometry map in order to have a better understanding between them. At 0 V, the inefficiency shows in the whole map. At 20 V, the efficiency is higher and the map is homogeneous because there is no PT in this sensor. At 50 V, a very nice and homogeneous efficiency map is obtained.

Edge efficiency Figure 7.20 shows the edge efficiency for AIDA-2020-G1 and Figure 7.21 for AIDA-2020-G2. For both AIDA-2020-G1 and AIDA-2020-G2, the hit efficiency is extended in the region between the last pixels and the trench. The efficiency in the extended region is more significant when the bias voltage increases. For AIDA-2020-G1, the 50% efficiency point reaches to 40 μm and 80 μm from the last pixel at 20V and 50V respectively. For AIDA-2020-G2, the 50% efficiency point reaches to 25 μm and 65 μm from the last pixel at 20V and 50V respectively. This is particularly important after irradiation where usually very high bias voltage needs to be applied. There are some hits recorded beyond the trench region which is an artifact of multiple scattering effect. Now, the sensors have been irradiated at the IRRAD facility at CERN and are going to be tested again.

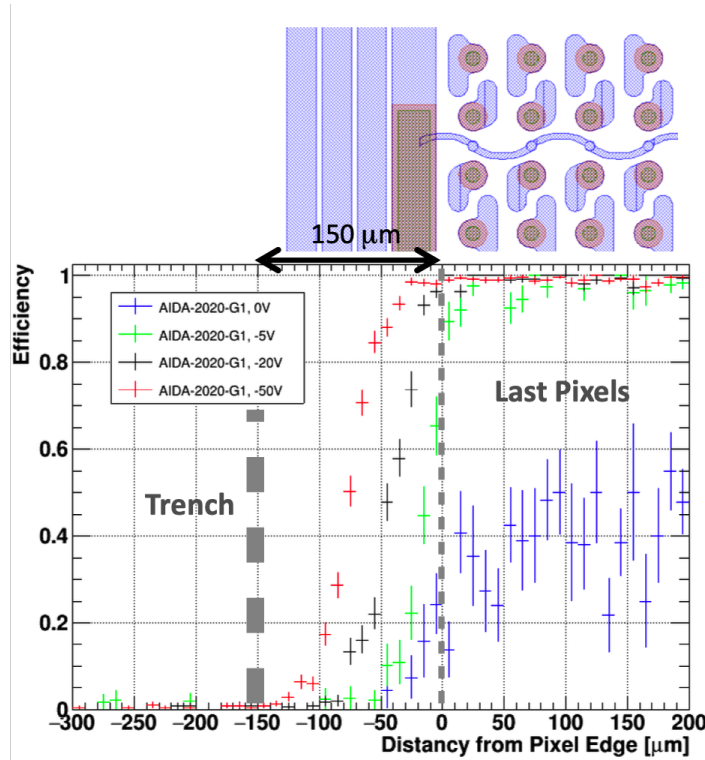


Figure 7.20: Edge efficiency for AIDA-2020-G1.

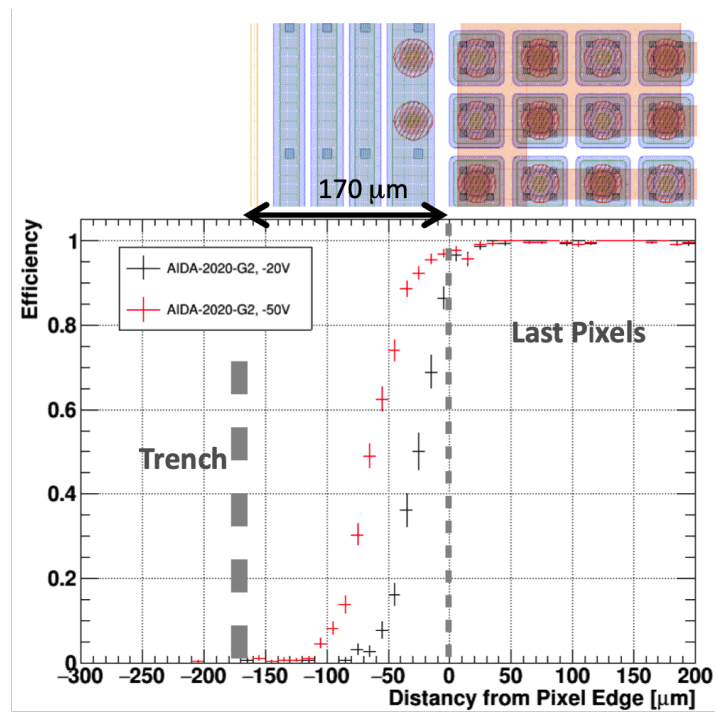


Figure 7.21: Edge efficiency for AIDA-2020-G2.

7.6 Summary

This chapter presents the characterisation of planar pixel modules developed in the context of the ITk project. The ITk and the basics of planar pixel modules have been reviewed. The I-V curve of the modules was measured at LPNHE to determine the operational limits of the devices under test. After irradiation, the leakage current of the sensors increases due to the different types of defects created in the depleted region. The efficiency of the modules was measured using test-beam facilities. The data obtained from test-beams was reconstructed and analysed using EUTelescope and TBmon2.

LPNHE R&D modules for ITk and future colliders have been tested. The experience from testing different modules has been used in the ITk sensor design. The 50 μm thick sensors are a promising candidate for future colliders. Two thin modules have been tested before irradiation. The modules have been irradiated and will be tested at beam lines in the future. Beyond R&D of ITk, more thin modules are assembled in LPNHE to allow a thorough study.

The modules for ITk market survey have also been tested and compared with the requirements. The ITk market survey campaign is finished, and the modules for ITk will be produced with the validated vendors.

The active edge modules have been developed in the framework of the AIDA-2020 collaboration. The edge efficiency of the sensors is extended with the design of guard rings. Now, the sensors have been irradiated and going to be tested again.

CONCLUSION

The Higgs boson produced at high transverse momentum is sensitive to the new physics. To study the Higgs boson at high transverse momentum, we need more powerful colliders and detectors, advanced techniques for physical objects reconstruction and identification as well as new ideas on the physic analysis strategies. Over the course of this thesis, three main topics covering all these aspects for boosted Higgs boson physics have been studied: the in-situ calibration of the $X \rightarrow b\bar{b}$ tagger using Z + jets events, the measurement of the $V(\rightarrow qq)H(\rightarrow b\bar{b})$ production at high transverse momentum and the characterisation of the planar pixel modules for the ATLAS Inner Tracker (ITk) upgrade of the inner detector and future colliders.

The calibration of the $X \rightarrow b\bar{b}$ tagger is an important task to allow its use by physics analyses interested in boosted $b\bar{b}$ topologies. The Z + jets events are used to derive the signal scale factors for $p_T > 450$ GeV using data collected by the ATLAS experiment from 2015 to 2018. The main backgrounds are dijet, $t\bar{t}$ and W + jets. The dijet is modelled by fitting the data directly using an exponentiated polynomial or polynomial function, depending on the p_T . The scale factors measured using Z + jets events for the $X \rightarrow b\bar{b}$ tagger at 60% WP are $1.30^{+0.55}_{-0.56}$ for $450 < p_T < 500$ GeV, $1.17^{+0.35}_{-0.34}$ for $500 < p_T < 600$ GeV and $0.55^{+0.29}_{-0.28}$ for $600 < p_T < 1000$ GeV. It is the first time that the $X \rightarrow b\bar{b}$ tagger has been calibrated. The methodology has been established, as well as the calibration framework. The methodology and results are documented in Ref. [10].

The $X \rightarrow b\bar{b}$ tagger and the corresponding calibration are the official recommendations for ATLAS analyses using the full Run 2 data and have already been used by recent VH resonance searches [11]. Furthermore, the methodology and the calibration framework are repurposed for other studies, such as the calibration of boosted $W/Z(\rightarrow qq)$ tagger using V + jets events, the close-by effect of the double b -tagging method, and the fully hadronic VH analysis presented in this thesis.

The measurement of VH production at high transverse momentum uses the final state of $V \rightarrow qq$ and $H \rightarrow b\bar{b}$. The signal strength will be measured first in inclusive and differential p_T^H bins: $250 < p_T^H < 450$ GeV, $450 < p_T^H < 650$ GeV and $p_T^H > 650$ GeV. Then, the cross-section will be measured within the STXS framework and will be used as well to constrain anomalous couplings in a SMEFT. It is the first time this channel is explored in the ATLAS collaboration. Simulation studies show a good complementarity with the semileptonic VH channels measured previously.

The full Run 2 data is used. The first two leading- p_T large- R jets are assigned as the Higgs boson candidate and the vector boson candidate. The W/Z tagger is used to tag the vector boson. The $X \rightarrow b\bar{b}$ tagger is used to tag the Higgs boson. The Higgs boson candidate is required to pass the $X \rightarrow b\bar{b}$ tagger at 60% WP and the vector boson candidate is required to pass the W/Z tagger at 50% WP. The scale factors of the $X \rightarrow b\bar{b}$ tagger have been remeasured in this analysis.

Two data-driven methods have been developed to estimate the main background in this analysis, the dijet background: the BDT reweighting method and the TF method. The BDT reweighting method, to which I have significantly contributed, is presented in detail. Comparisons of both background estimation methods have been made. The distributions

from the signal region side-band fits between the two methods are consistent with one another, accounting for their systematic uncertainties. For the Asimov fits, the fitted errors with the TF method is smaller than that of the BDT reweighting method. The two methods have very different flexibility in the statistical fits which are discussed in detail in Section 5.6.3.2.

This analysis is still ongoing, and therefore no final results are presented in this document. All results can be found in the internal note, Ref. [208]. The dijet modelling method will be chosen from the BDT reweighting and TF methods. The full set of systematic uncertainties, including experimental systematic uncertainties and modelling systematic uncertainties, are prepared and will be added to the final measurement.

In the long term, combining this analysis with the results obtained from the $V(\rightarrow \text{leptons})H(\rightarrow b\bar{b})$ and the ggF $H \rightarrow b\bar{b}$ production will allow to gain more precision and improve the $H \rightarrow b\bar{b}$ understanding.

The tracking detectors play an important role in track reconstruction and are thus crucial for jet reconstruction and flavour tagging. The present Inner detector will not be able to cope with the HL-LHC conditions, an all-silicon ITk is proposed to achieve HL-LHC physics goals. This thesis presents my work on the planar silicon pixel module characterisation for ITk. The characterisation of these modules was performed in several steps. First, the I-V curve of the modules was measured at the local laboratory to determine the operational limits of the devices under test. Then, the modules were calibrated, and the efficiency of the modules was measured using test-beam facilities. The data obtained from test beams was reconstructed and analysed using EU Telescope and TBmon2.

The modules of interest consist of the LPNHE R&D modules for the ITk, modules during the market survey campaign of the ITk and R&D modules beyond ITk. The experience of LPNHE from testing different modules has been used in the ITk sensor design. Sensors with different bias structures and geometries have been tested. The bias structures depend on the suppliers. The geometry of $50 \times 50 \mu\text{m}^2$ is chosen for ITk planar pixel layers to have a standard production. The 100 and 150 μm thick sensors are chosen by ITk. During R&D stage, we have also found that the 50 μm ultra-thin sensors show good efficiency, and they are radiation-hard candidates for future colliders. After R&D, market survey is an important stage to validate the potential suppliers. The sensors from 6 foundries have been tested to ensure that the different companies producing modules met the quality requirements. The ITk market survey campaign has concluded, and the modules for ITk are being produced with the validated vendors. Beyond ITk R&D, the active edge modules are developed in the framework of the AIDA-2020 collaboration. The edge efficiency of the sensors is extended with the design of guard rings.

Milestones for the development of ITk pixel detector have been achieved regardless of the COVID pandemic. The system is moving from the R&D phase to a reconstruction mode and is expected to be in service from 2029. Beyond ITk, more understanding about the silicon tracking detector has been achieved for future detectors.

The studies of boosted Higgs boson for new physics searches will benefit from the work presented in this thesis: the $X \rightarrow b\bar{b}$ tagger calibration, physics analysis and detector development. All studies are very important for LHC Run 3 and HL-LHC studies, and we are looking forward to more results coming in the future for boosted Higgs boson studies and

new physic searches.

RESULTS OF $\mu^{post-tag}$ AND $\mu^{pre-tag}$ Table A.1: Results of $\mu^{post-tag}$ and its uncertainties.

	$450 < p_T < 500 \text{ GeV}$	$500 < p_T < 600 \text{ GeV}$	$600 < p_T < 1000 \text{ GeV}$
μ	1.320	1.097	0.508
Uncertainties (+σ/ - σ)			
Statistical	0.081	0.071	0.062
Z modelling	+0.120/-0.214	+0.028/-0.026	0.026/-0.030
W modelling	+0.003/-0.007	+0.005/-0.005	+0.0042/-0.006
ttbar modelling	+0.028/-0.032	+0.017/-0.017	+0.010/-0.012
Background modelling	-0.112/0.114	-0.065/0.066	-0.067/0.068
JMS	+0.019/-0.011	+0.011/-0.003	+0.021/-0.007
JMR	-0.220/0.148	-0.105/0.088	-0.082/0.065
JES	+0.072/-0.071	+0.063/-0.061	+0.031/-0.029
Others	+0.023/-0.025	+0.020/-0.020	0.009/-0.010
Total uncertainties	+0.298/-0.307	+0.161/-0.149	+0.132/-0.122

SHERPA 2.2.11 $V + \text{JETS}$ SAMPLES

The hadronically decaying Z and W events, with two additional parton at NLO order accuracy and up to 5 additional partons at LO, are generated using SHERPA 2.2.11 [145] with the NNPDF3.0 [130] NNLO PDF. The events are required to have at least two large- R jets and no isolated electrons and muons present. The leading large- R jets are required to have p_T larger than 450 GeV. The subleading large- R jets are required to have p_T larger than 200 GeV. The truth label of the leading large- R jets is required to *matched* V . Figure B.1 shows the leading large- R jet mass and p_T distribution for events generated from SHERPA 2.2.8 and SHERPA 2.2.11. The electroweak virtual corrections can be seen from the ratio panel of these distribution. The improvement on the inclusive cross-section is around 20%. The improvement is p_T dependent and varies from 10% to 30% when the p_T of the large- R jets goes from 450 GeV to 1.5 TeV. These observations are consistent with the results published in Ref. [145]. The calibration team is moving to use the SHERPA 2.2.11 $V + \text{jets}$ samples.

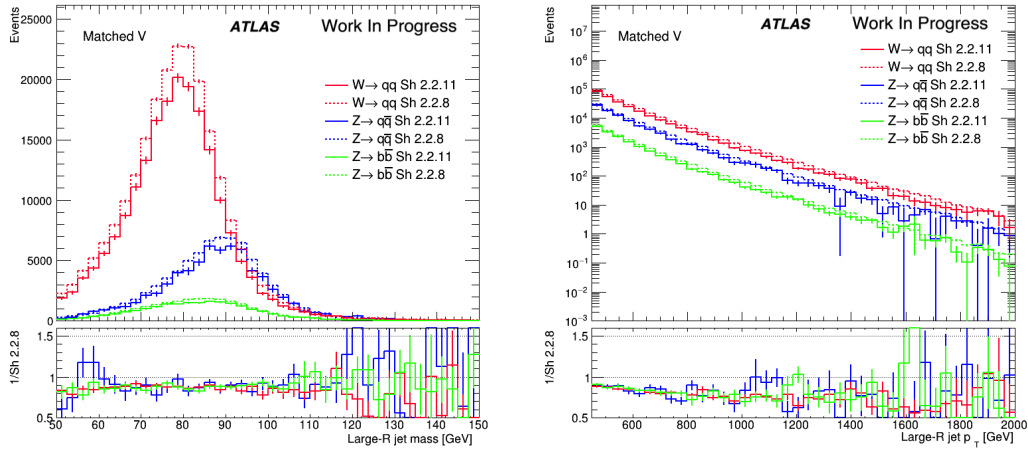


Figure B.1: The comparison of $V + \text{jets}$ events generated from SHERPA 2.2.8 and SHERPA 2.2.11.

FULLY HADRONIC VH ANALYSIS

C.1 Data and MC comparisons

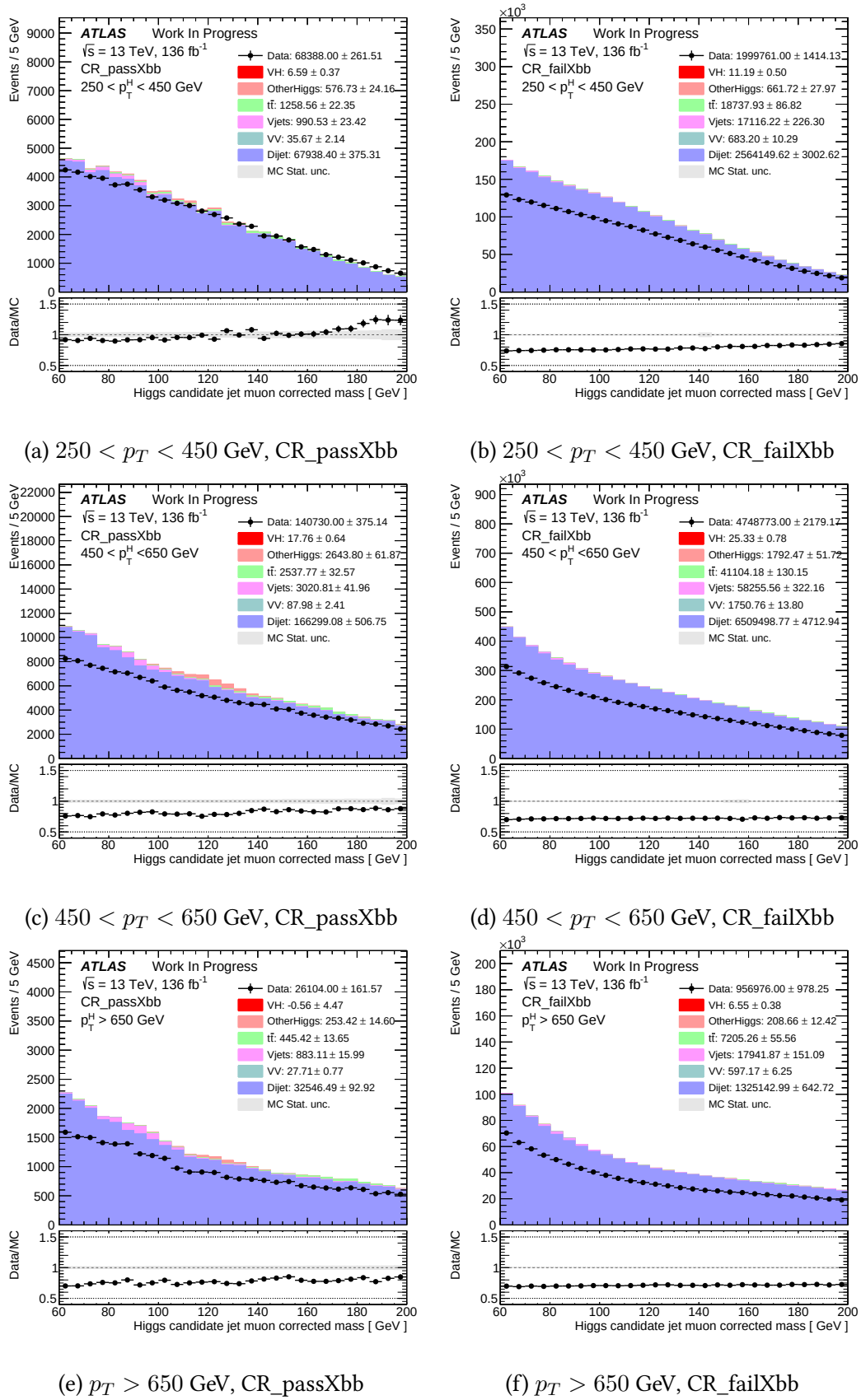


Figure C.1: The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in CR_passXbb and CR_failXbb.

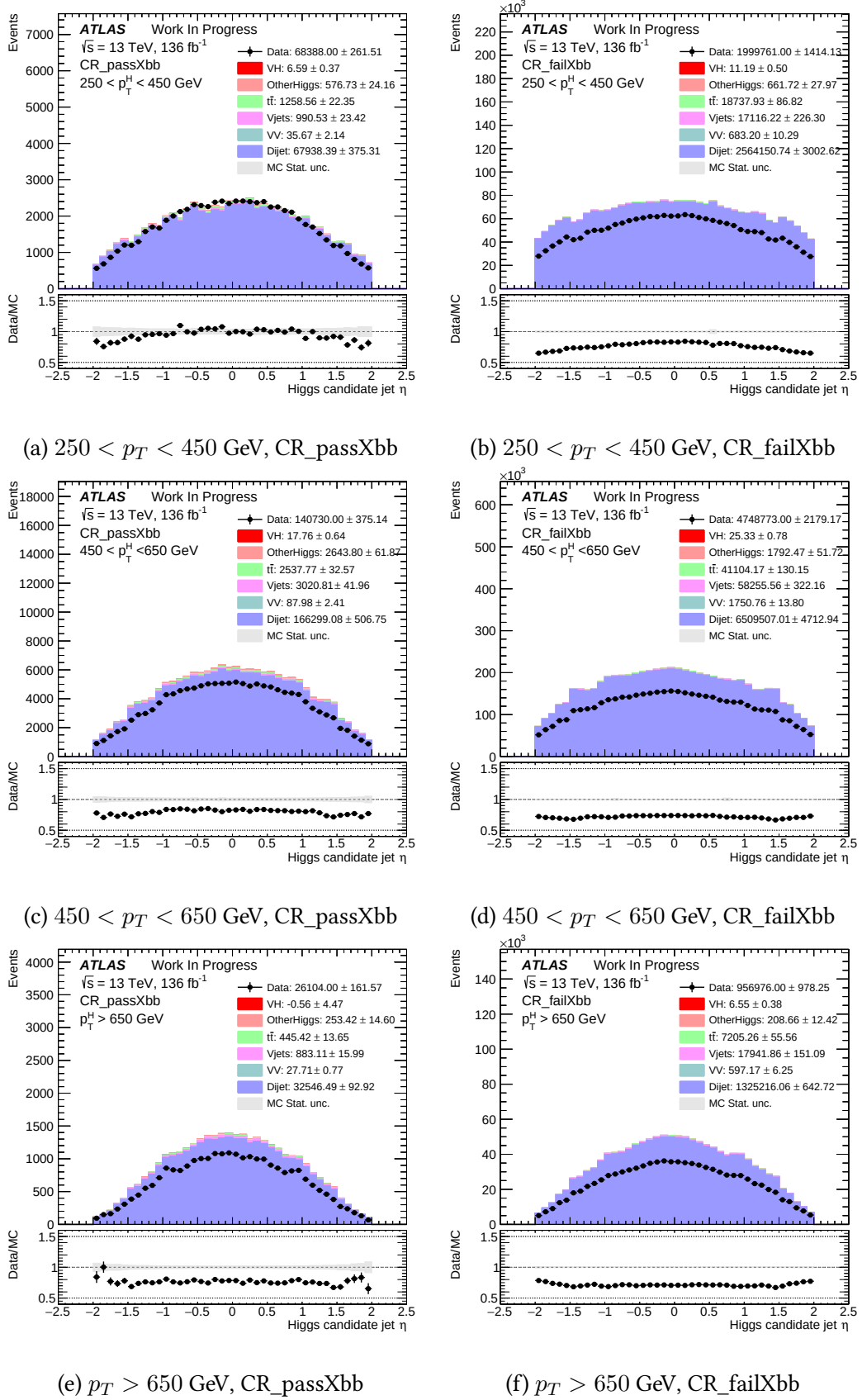
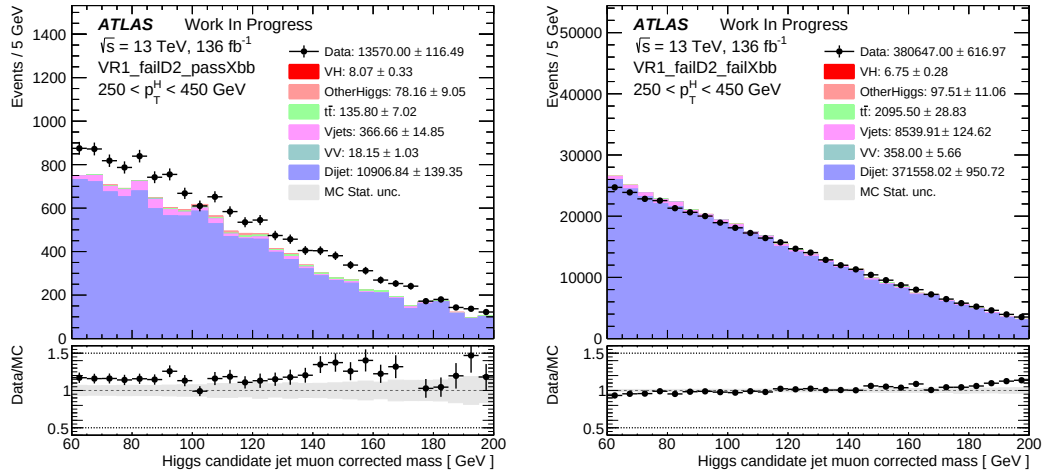
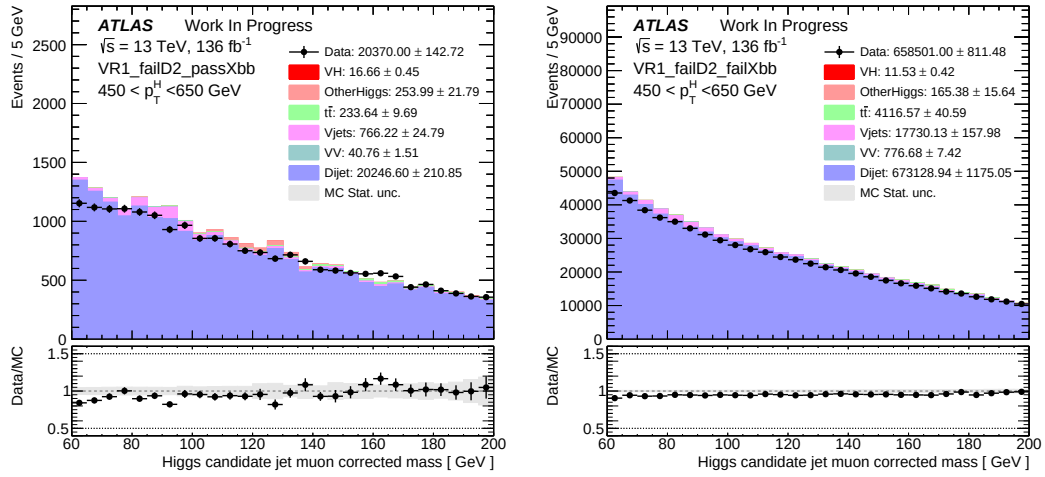


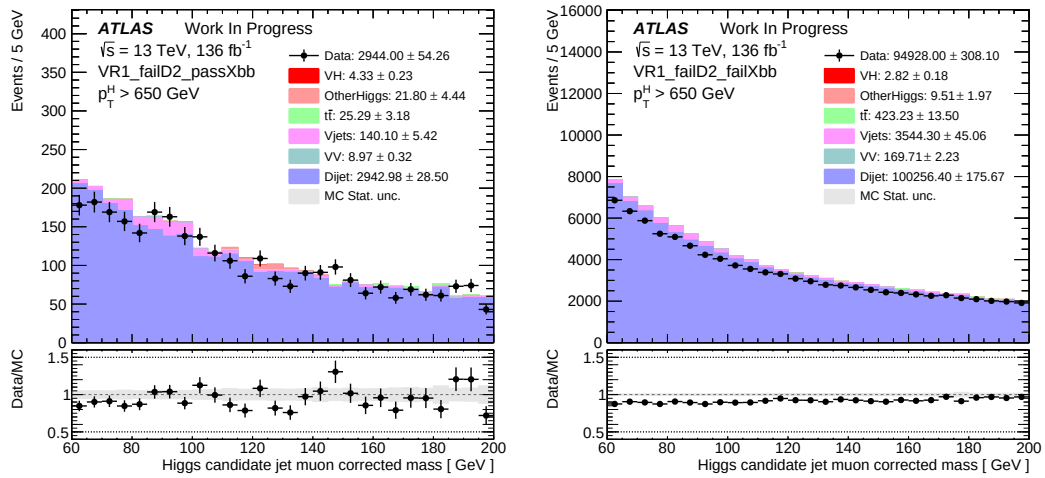
Figure C.2: The Higgs jet η distribution from Data and MC simulations for differential p_T in CR_passXbb and CR_failXbb.



(a) $250 < p_T < 450 \text{ GeV}$, VR1_failD2_passXbb (b) $250 < p_T < 450 \text{ GeV}$, VR1_failD2_failXbb



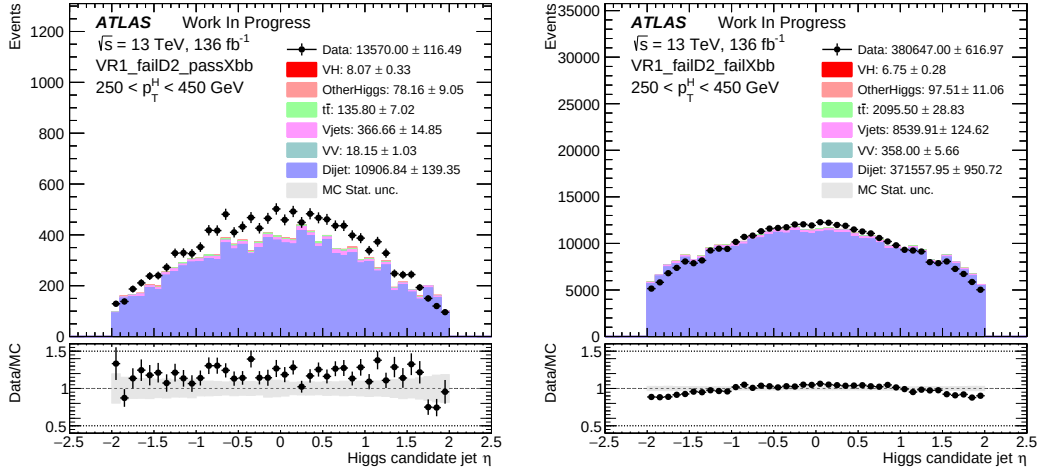
(c) $450 < p_T < 650 \text{ GeV}$, VR1_failD2_passXbb (d) $450 < p_T < 650 \text{ GeV}$, VR1_failD2_failXbb



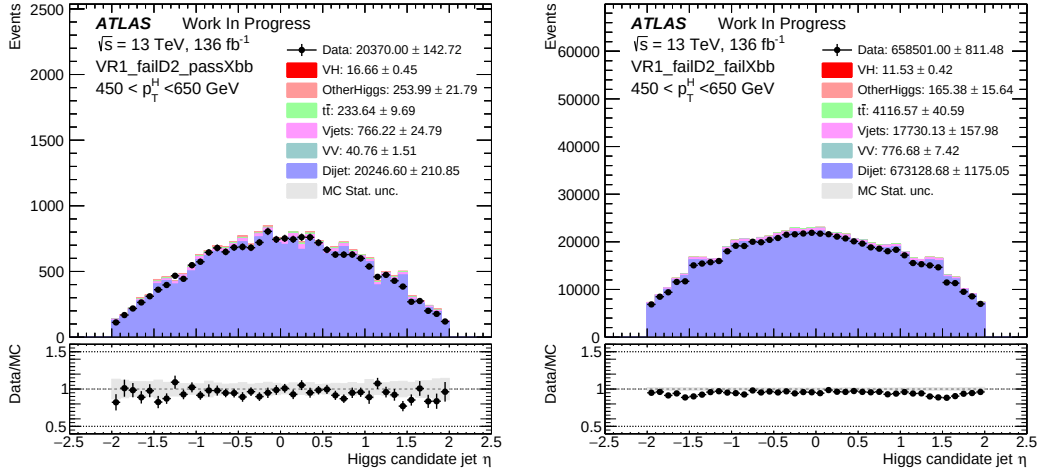
(e) $p_T > 650 \text{ GeV}$, VR1_failD2_passXbb

(f) $p_T > 650 \text{ GeV}$, VR1_failD2_failXbb

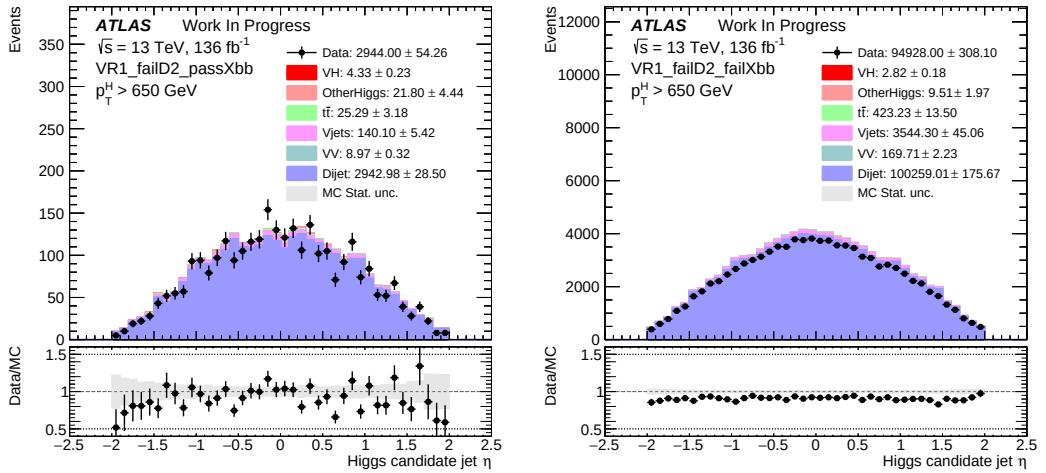
Figure C.3: The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in VR1_failD2_passXbb and VR1_failD2_failXbb.



(a) $250 < p_T < 450 \text{ GeV}$, VR1_failD2_passXbb (b) $250 < p_T < 450 \text{ GeV}$, VR1_failD2_failXbb

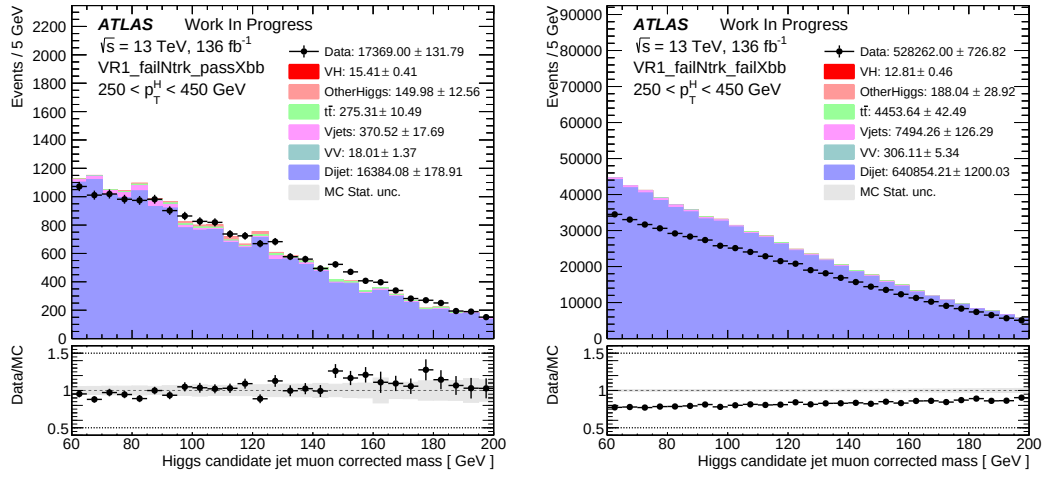


(c) $450 < p_T < 650 \text{ GeV}$, VR1_failD2_passXbb (d) $450 < p_T < 650 \text{ GeV}$, VR1_failD2_failXbb

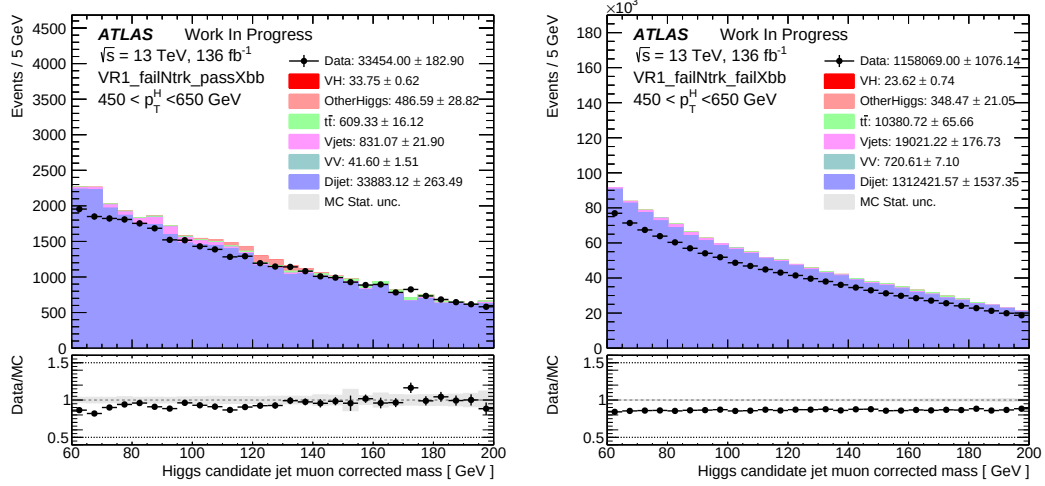


(e) $p_T > 650 \text{ GeV}$, VR1_failD2_passXbb (f) $p_T > 650 \text{ GeV}$, VR1_failD2_failXbb

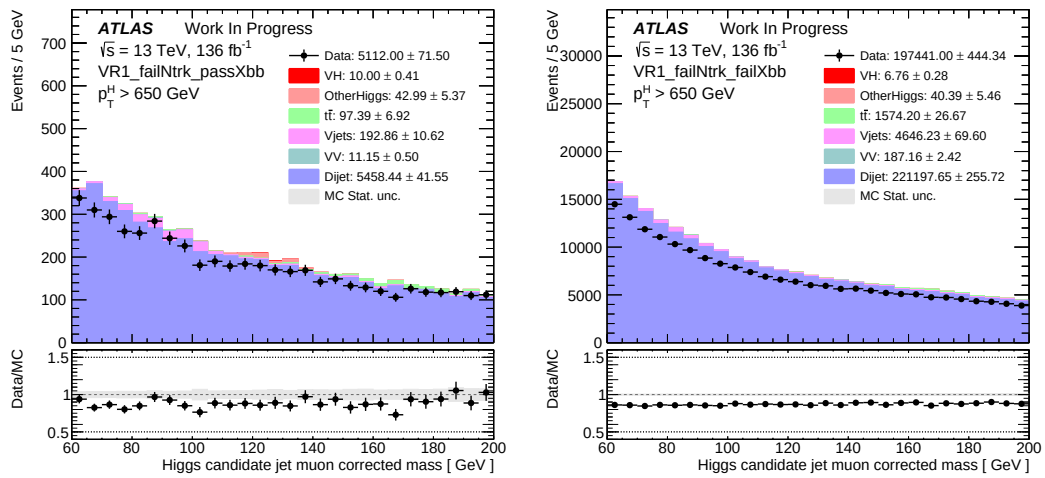
Figure C.4: The Higgs jet η distribution from Data and MC simulations for differential p_T in VR1_failD2_passXbb and VR1_failD2_failXbb.



(a) $250 < p_T < 450 \text{ GeV}$, VR1_failNtrk_passXbb (b) $250 < p_T < 450 \text{ GeV}$, VR1_failNtrk_failXbb

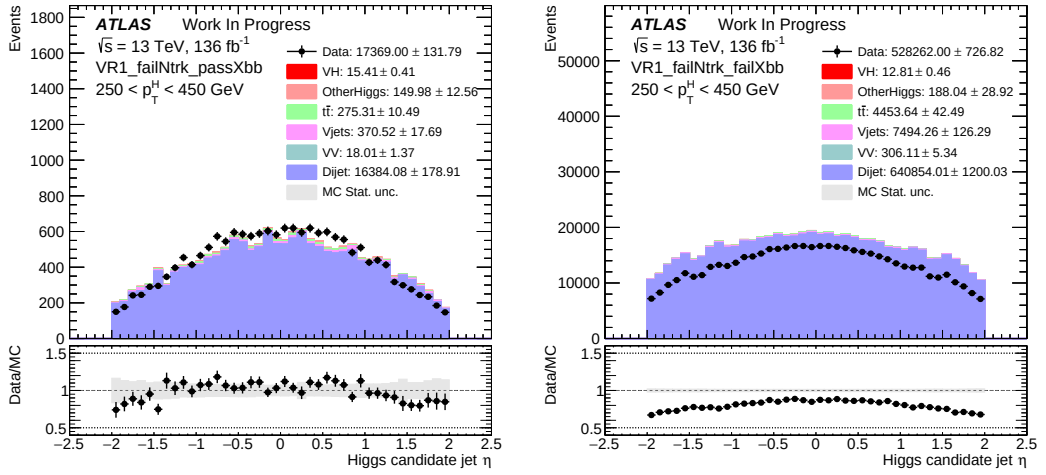


(c) $450 < p_T < 650 \text{ GeV}$, VR1_failNtrk_passXbb (d) $450 < p_T < 650 \text{ GeV}$, VR1_failNtrk_failXbb

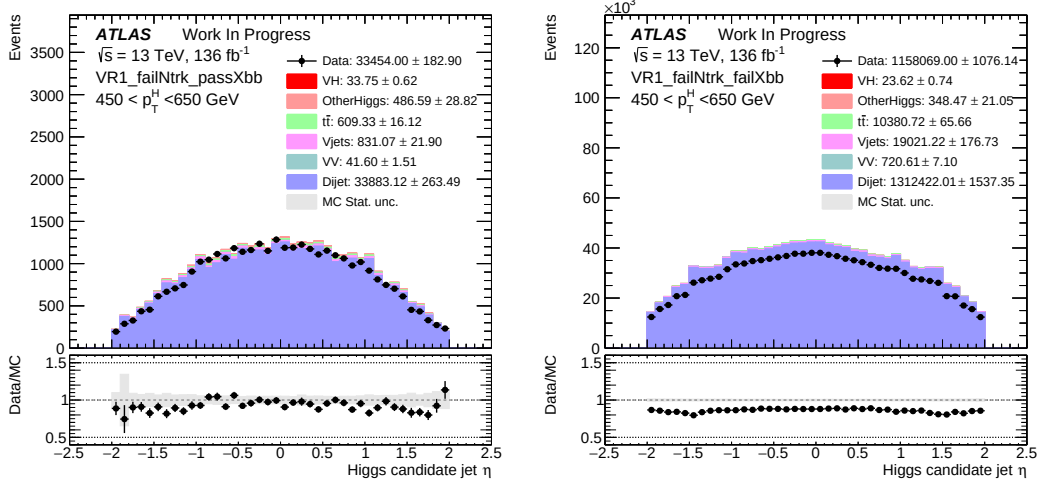


(e) $p_T > 650 \text{ GeV}$, VR1_failNtrk_passXbb (f) $p_T > 650 \text{ GeV}$, VR1_failNtrk_failXbb

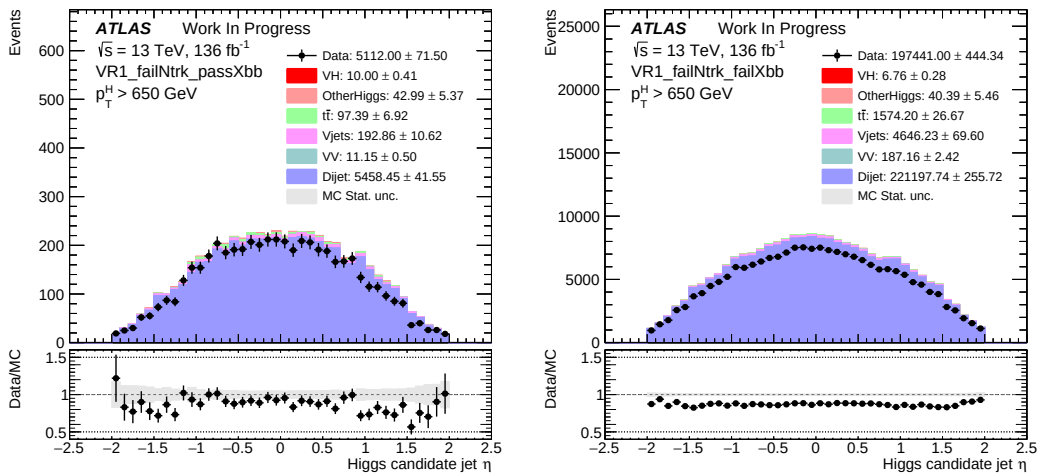
Figure C.5: The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in VR1_failNtrk_passXbb and VR1_failNtrk_failXbb.



(a) $250 < p_T < 450 \text{ GeV}$, VR1_failNtrk_passXbb (b) $250 < p_T < 450 \text{ GeV}$, VR1_failNtrk_failXbb



(c) $450 < p_T < 650 \text{ GeV}$, VR1_failNtrk_passXbb (d) $450 < p_T < 650 \text{ GeV}$, VR1_failNtrk_failXbb



(e) $p_T > 650 \text{ GeV}$, VR1_failNtrk_passXbb (f) $p_T > 650 \text{ GeV}$, VR1_failNtrk_failXbb

Figure C.6: The Higgs jet η distribution from Data and MC simulations for differential p_T in VR1_failNtrk_passXbb and VR1_failNtrk_failXbb.

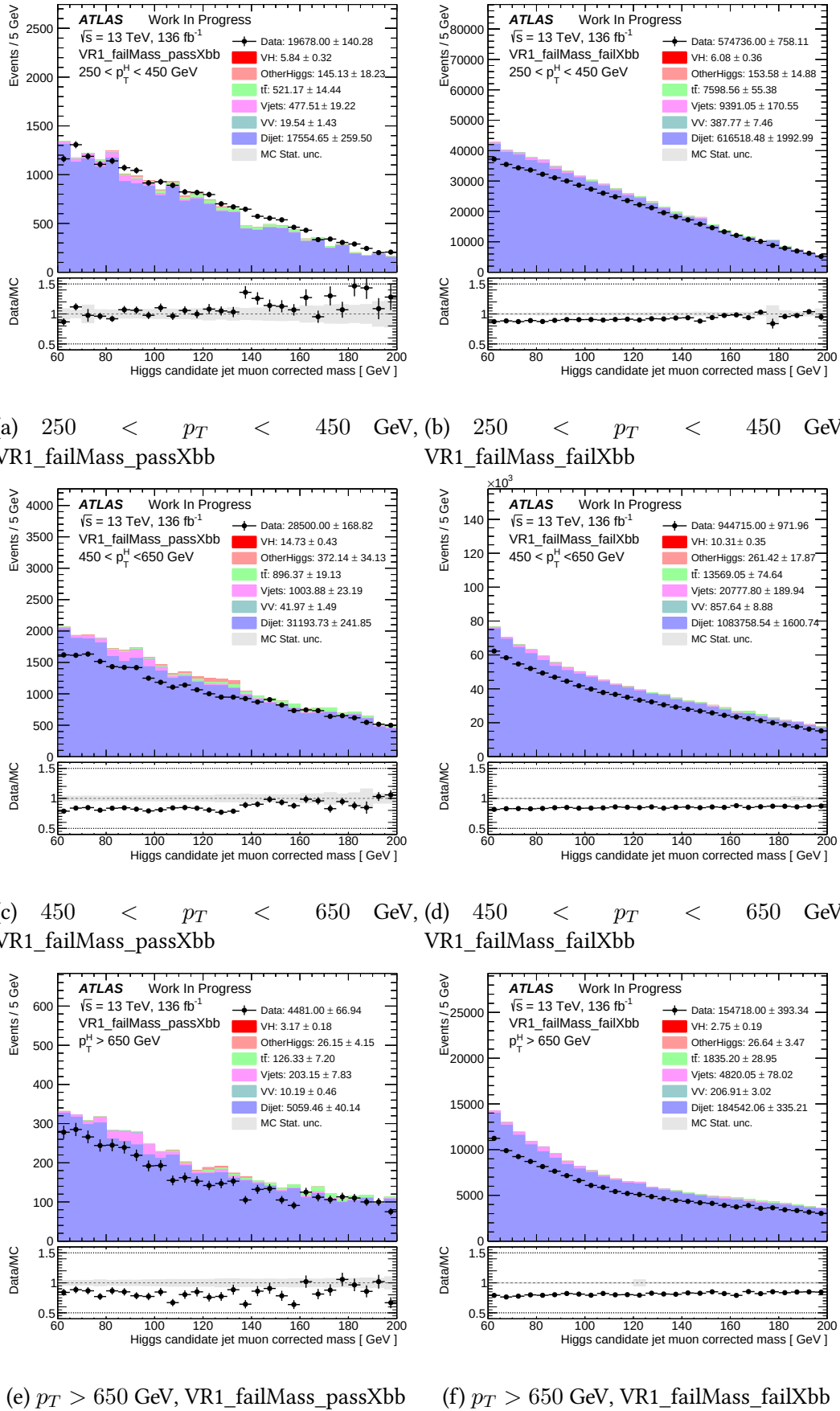
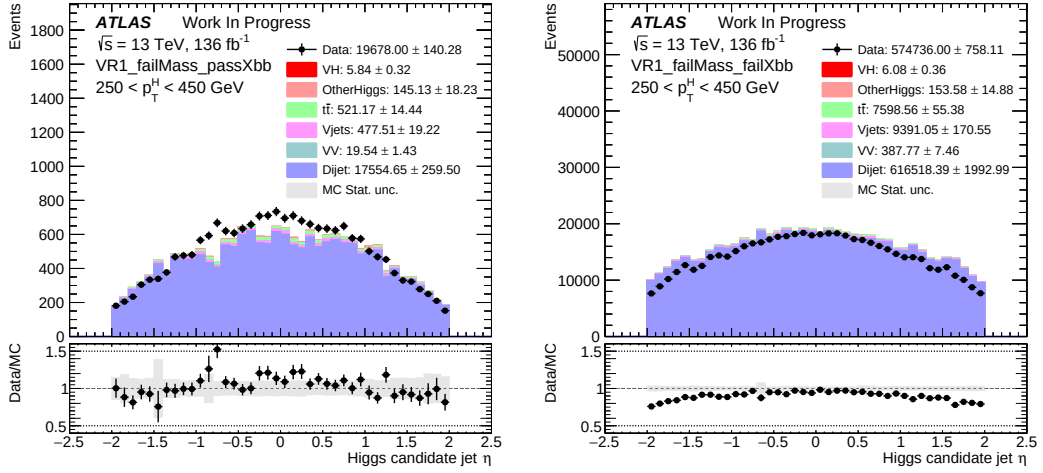
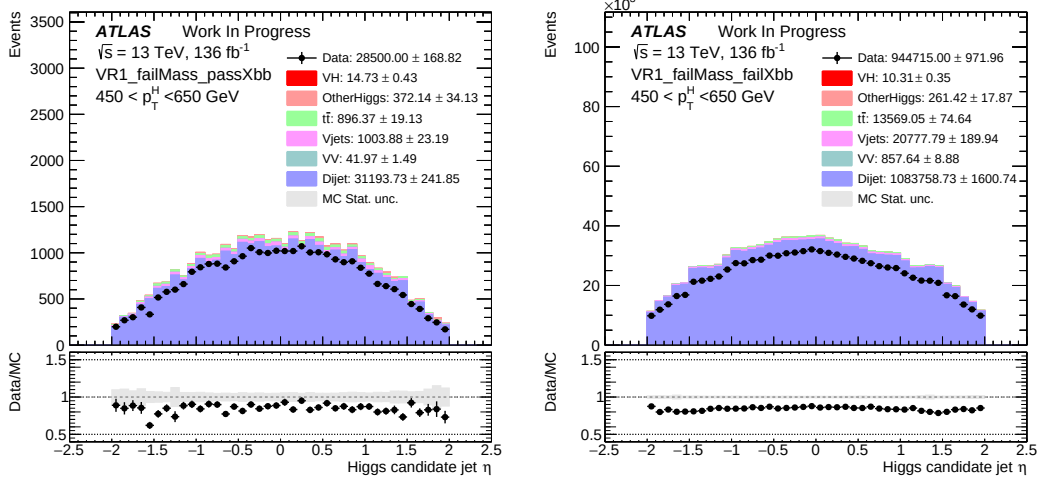


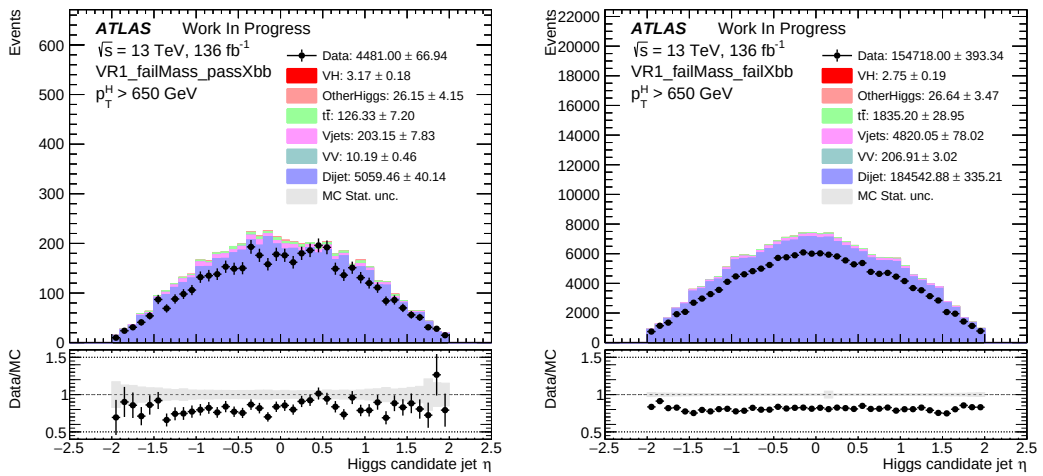
Figure C.7: The muon-corrected Higgs jet mass distribution from Data and MC simulations for differential p_T in VR1_failMass_passXbb and VR1_failMass_failXbb.



(a) $250 < p_T < 450 \text{ GeV}$, VR1_failMass_passXbb (b) $250 < p_T < 450 \text{ GeV}$, VR1_failMass_failXbb



(c) $450 < p_T < 650 \text{ GeV}$, VR1_failMass_passXbb (d) $450 < p_T < 650 \text{ GeV}$, VR1_failMass_failXbb



(e) $p_T > 650 \text{ GeV}$, VR1_failMass_passXbb (f) $p_T > 650 \text{ GeV}$, VR1_failMass_failXbb

Figure C.8: The Higgs jet η distribution from Data and MC simulations for differential p_T in VR1_failMass_passXbb and VR1_failMass_failXbb.

C.2 $X \rightarrow b\bar{b}$ tagger calibration

Table C.1: Scale factors derived for the $X \rightarrow b\bar{b}$ tagger in fully hadronic VH analysis using Z + jets events.

Calibration p_T [GeV]	$Z(\rightarrow b\bar{b}) + \text{jets}$		
	450 – 500	500 – 600	600 – 1000
$\mu^{post-tag}$	1.43	1.19	0.78
$\mu^{pre-tag}$	0.99	0.92	0.91
SF	1.44	1.30	0.86
Uncertainties ($\pm\sigma$)			
Statistical	± 0.14	± 0.13	± 0.13
Z -boson modelling	$^{+0.22}_{-0.36}$	$^{+0.15}_{-0.24}$	$^{+0.15}_{-0.28}$
Spurious signal	± 0.40	± 0.32	± 0.44
Other background modelling	± 0.05	± 0.04	± 0.03
Lepton related	± 0.04	± 0.04	± 0.02
Jet mass scale	± 0.01	$^{+0.01}_{-0.02}$	$< \pm 0.01$
Jet mass resolution	$^{+0.01}_{-0.02}$	± 0.02	± 0.01
Jet energy scale	$^{+0.10}_{-0.09}$	± 0.09	± 0.05
Jet energy resolution	± 0.01	$< \pm 0.01$	$< \pm 0.01$
PRW	± 0.01	$< \pm 0.01$	$< \pm 0.01$
Total uncertainty	$^{+0.50}_{-0.57}$	$^{+0.39}_{-0.44}$	$^{+0.49}_{-0.55}$

C.3 Transfer factor method

The transfer factor (TF) method is a data-driven approach to estimate the dijet background. It was first used in the search for highly boosted Higgs to bottom quark-antiquark pairs with the CMS detector [154]. The rate of dijet events in the SR, which are selected with the passXbb requirement, is calculated from the events in the failXbb region ($\overline{\text{SR}}$), as a function of the jet mass and p_T with a transfer factor (TF) as defined:

$$N_{dijet}^{passXbb}(p_T, m) = TF \times N_{dijet}^{failXbb}(p_T, m), \quad (\text{C.1})$$

where $N_{dijet}^{failXbb}(p_T, m)$ is the number of events in the failXbb region as a function of jet p_T and mass.

If the $X \rightarrow b\bar{b}$ tagger would be completely uncorrelated from the jet p_T and jet mass, the transfer factor would be a constant. However, the $X \rightarrow b\bar{b}$ tagger has not negligible dependencies on both the jet p_T and jet mass.

To account for these correlations, the TF is Taylor-expanded as a function of p_T and ρ , where $\rho = \log(m^2/p_T^2)$. The QCD scaling variable, ρ is a more natural variable to consider as it should have a roughly invariant shape as a function of p_T .

$$N_{dijet}^{passXbb}(p_T, m) = TF(p_T, \rho) \times N_{dijet}^{failXbb}(p_T, m). \quad (\text{C.2})$$

Then the number of events in the passXbb region in the jet p_T bin p_{T_i} and mass bin m_j , $N_{dijet}^{passXbb}(p_{T_i}, m_j)$, can be written as

$$N_{dijet}^{passXbb}(p_{T_i}, m_j) = \sum_{k,l} \alpha_{kl} \rho_{ij}^k p_{T_i}^l \times N_{dijet}^{failXbb}(p_{T_i}, m_j), \quad (\text{C.3})$$

where α_{kl} is the polynomial coefficient for the k th order on ρ and l th order on p_T , and $N_{dijet}^{failXbb}(p_{T_i}, m_j)$ is the number of events in the region failXbb in the jet p_T bin p_{T_i} and mass bin m_j .

The coefficients α_{kl} are determined from the likelihood fit. Based on Equation C.3, dijet background estimate and signal extraction are performed simultaneously:

$$\mathcal{L}(\text{data}|\mu, \theta) = \prod_{i,j} P(N_{passXbb,ij}^{data} | N_{passXbb,ij}^{dijet} + N_{passXbb,ij}^{other\ bkg} + N_{passXbb,ij}^{VH}) \quad (\text{C.4})$$

$$\times \prod_{i,j} P(N_{failXbb,ij}^{data} | N_{failXbb,ij}^{dijet} + N_{failXbb,ij}^{other\ bkg} + N_{failXbb,ij}^{VH}) \quad (\text{C.5})$$

$$(\text{C.6})$$

BIBLIOGRAPHY

1. *The Large Hadron Collider* <https://home.cern/science/accelerators/large-hadron-collider>.
2. *A Large Ion Collider Experiment* <https://home.cern/science/experiments/alice>.
3. *ATLAS Luminosity Public Results Run2* <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
4. *Compact Muon Solenoid (CMS)* <https://home.cern/science/experiments/cms>.
5. *Large Hadron Collider beauty (LHCb)* <https://home.cern/science/experiments/lhcb>.
6. Aad, G. *et al.* Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B* **716**, 1–29. arXiv: [1207.7214 \[hep-ex\]](#) (2012).
7. Chatrchyan, S. *et al.* Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B* **716**, 30–61. arXiv: [1207.7235 \[hep-ex\]](#) (2012).
8. Lindert, J. M., Kudashkin, K., Melnikov, K. & Weber, C. Higgs bosons with large transverse momentum at the LHC. *Phys. Lett. B* **782**, 210–214. arXiv: [1801.08226 \[hep-ph\]](#) (2018).
9. Grazzini, M., Ilnicka, A. & Spira, M. Higgs boson production at large transverse momentum within the SMEFT: analytical results. *Eur. Phys. J. C* **78**, 808. arXiv: [1806.08832 \[hep-ph\]](#) (2018).
10. *Efficiency corrections for a tagger for boosted $H \rightarrow b\bar{b}$ decays in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector* tech. rep. (CERN, Geneva, Aug. 2021). <http://cds.cern.ch/record/2777811>.
11. *Anomaly detection search for new resonances decaying into a Higgs boson and a generic new particle X in hadronic final states using $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector* tech. rep. (CERN, Geneva, July 2022). <https://cds.cern.ch/record/2816323>.
12. Calderini, G. *et al.* LPNHE-FBK Thin n-on-p Pixel Sensors for HL-LHC Upgrade and Beyond. *JPS Conf. Proc.* **34**, 010021. 7 p. <https://cds.cern.ch/record/2801064> (2021).
13. *Test-beam results from AIDA-2020 active edge modules* <https://indico.cern.ch/event/1104064/contributions/4790585/>.
14. *Standard Model of Elementary Particles* https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
15. Peskin, M. E. & Schroeder, D. V. *An Introduction to quantum field theory* ISBN: 9780201503975, 0201503972. <http://www.slac.stanford.edu/~mpeskin/QFT.html> (Addison-Wesley, Reading, USA, 1995).

16. Wu, C. S., Ambler, E., Hayward, R. W., Hoppes, D. D. & Hudson, R. P. Experimental Test of Parity Conservation in Beta Decay. *Phys. Rev.* **105**, 1413–1415. <https://link.aps.org/doi/10.1103/PhysRev.105.1413> (4 Feb. 1957).
17. Bludman, S. A. On the universal Fermi interaction. *Nuovo Cim.* **9**, 433–445 (1958).
18. Glashow, S. L. Partial Symmetries of Weak Interactions. *Nucl. Phys.* **22**, 579–588 (1961).
19. Faddeev, L. D. & Popov, V. N. Feynman Diagrams for the Yang-Mills Field. *Phys. Lett.* **B25**, 29–30 (1967).
20. 't Hooft, G. & Veltman, M. Regularization and renormalization of gauge fields. *Nuclear Physics B* **44**, 189–213. ISSN: 0550-3213. <https://www.sciencedirect.com/science/article/pii/0550321372902799> (1972).
21. Group, P. D. Review of Particle Physics. *Progress of Theoretical and Experimental Physics* **2020**. 083C01. ISSN: 2050-3911. eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>. <https://doi.org/10.1093/ptep/ptaa104> (Aug. 2020).
22. Higgs, P. W. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.* **13**, 508–509. <https://link.aps.org/doi/10.1103/PhysRevLett.13.508> (16 Oct. 1964).
23. Englert, F. & Brout, R. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.* **13**, 321–323. <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (9 Aug. 1964).
24. Guralnik, G. S., Hagen, C. R. & Kibble, T. W. B. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.* **13**, 585–587 (1964).
25. Weinberg, S. A Model of Leptons. *Phys. Rev. Lett.* **19**, 1264–1266 (1967).
26. Salam, A. Weak and Electromagnetic Interactions. *Conf. Proc.* **C680519**, 367–377 (1968).
27. Cabibbo, N. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.* **10**, 531–533 (1963).
28. Kobayashi, M. & Maskawa, T. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.* **49**, 652–657 (1973).
29. Sushkov, A. O., Kim, W. J., Dalvit, D. A. R. & Lamoreaux, S. K. New Experimental Limits on Non-Newtonian Forces in the Micrometer Range. *Phys. Rev. Lett.* **107**, 171101. arXiv: [1108.2547](https://arxiv.org/abs/1108.2547) [quant-ph] (2011).
30. *Dark Energy, Dark Matter* <https://science.nasa.gov/astrophysics/focus-areas/what-is-dark-energy>.
31. Fukuda, Y. *et al.* Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.* **81**, 1562–1567. arXiv: [hep-ex/9807003](https://arxiv.org/abs/hep-ex/9807003) (1998).
32. Mikaelyan, L. A. & Sinev, V. V. Neutrino oscillations at reactors: What next? *Phys. Atom. Nucl.* **63** (eds Bednyakov, V. A., Brudanin, V. B. & Kovalenko, S. G.) 1002–1006. arXiv: [hep-ex/9908047](https://arxiv.org/abs/hep-ex/9908047) (2000).

33. Canetti, L., Drewes, M. & Shaposhnikov, M. Matter and Antimatter in the Universe. *New J. Phys.* **14**, 095012. arXiv: [1204.4186 \[hep-ph\]](#) (2012).
34. Blum, T. *et al.* The Muon (g-2) Theory Value: Present and Future. arXiv: [1311.2198 \[hep-ph\]](#) (Nov. 2013).
35. Abi, B. *et al.* Measurement of the Positive Muon Anomalous Magnetic Moment to 0.46 ppm. *Phys. Rev. Lett.* **126**, 141801. arXiv: [2104.03281 \[hep-ex\]](#) (2021).
36. Capdevila, B., Crivellin, A., Descotes-Genon, S., Matias, J. & Virto, J. Patterns of New Physics in $b \rightarrow s\ell^+\ell^-$ transitions in the light of recent data. *JHEP* **01**, 093. arXiv: [1704.05340 \[hep-ph\]](#) (2018).
37. CDF Collaboration *et al.* High-precision measurement of the W boson mass with the CDF II detector. *Science* **376**, 170–176 (Apr. 2022).
38. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector* (eds Heinemeyer, S., Mariotti, C., Passarino, G. & Tanaka, R.) <https://cds.cern.ch/record/2227475> (Oct. 2016).
39. Buchmüller, W. & Wyler, D. Effective lagrangian analysis of new interactions and flavour conservation. *Nuclear Physics B* **268**, 621–653. ISSN: 0550-3213. <https://www.sciencedirect.com/science/article/pii/0550321386902622> (1986).
40. Grzadkowski, B., Iskrzynski, M., Misiak, M. & Rosiek, J. Dimension-Six Terms in the Standard Model Lagrangian. *JHEP* **10**, 085. arXiv: [1008.4884 \[hep-ph\]](#) (2010).
41. Ellis, J. *SMEFT Constraints on New Physics beyond the Standard Model in Beyond Standard Model: From Theory to Experiment* (May 2021). arXiv: [2105.14942 \[hep-ph\]](#).
42. Mobs, E. The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019. General Photo. <https://cds.cern.ch/record/2684277> (July 2019).
43. *The HL-LHC project* <https://hilumilhc.web.cern.ch/content/hl-lhc-project>.
44. *ATLAS Luminosity Public Results Run2* <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
45. Aad, G. *et al.* ATLAS data quality operations and performance for 2015–2018 data-taking. *JINST* **15**, P04003. arXiv: [1911.04632 \[physics.ins-det\]](#) (2020).
46. D’Eramo, L. *Study of the decay properties of the Higgs boson into two b quarks and upgrade of the ATLAS inner tracker. Étude de la désintégration du boson de Higgs en deux quarks b et amélioration du trajectographe de l’expérience ATLAS* Presented 23 Sep 2019 (2019). <https://cds.cern.ch/record/2698495>.
47. Pequenaio, J. *Event Cross Section in a computer generated image of the ATLAS detector*. Mar. 2008. <https://cds.cern.ch/record/1096081>.
48. *The Inner Detector* <https://atlas.cern/Discover/Detector/Inner-Detector>.
49. Et al., A. Alignment of the ATLAS Inner Detector in Run-2. *Eur. Phys. J. C* **80**, 1194. 41 p. arXiv: [2007.07624](#). <https://cds.cern.ch/record/2724037> (July 2020).

50. Giordani, M. The upgraded Pixel Detector of the ATLAS experiment for Run-2 at the Large Hadron Collider. <https://cds.cern.ch/record/2229166> (Nov. 2016).
51. Sfyrla, A. *et al.* The Detector Control System for the ATLAS Semiconductor Tracker Assembly Phase. *Nuclear Science, IEEE Transactions on* **52**, 938–943 (Sept. 2005).
52. Vogel, A. *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates* tech. rep. (CERN, Geneva, Apr. 2013). <https://cds.cern.ch/record/1537991>.
53. *The ATLAS calorimeter* <https://atlas.cern/Discover/Detector/Calorimeter>.
54. *ATLAS liquid-argon calorimeter: Technical Design Report* <https://cds.cern.ch/record/331061> (CERN, Geneva, 1996).
55. Collaboration, A. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* **3**. Also published by CERN Geneva in 2010, S08003. 437 p. <https://cds.cern.ch/record/1129811> (2008).
56. Aharrouché, M. *et al.* Energy Linearity and Resolution of the ATLAS Electromagnetic Barrel Calorimeter in an Electron Test-Beam. *Nucl. Instrum. Methods Phys. Res., A* **568**, 601–623. 48 p. <http://cds.cern.ch/record/976098> (Aug. 2006).
57. Collaboration, A. Electron and photon energy calibration with the ATLAS detector using 2015-2016 LHC proton-proton collision data. *Submitted to: JINST*. arXiv: [1812.03848 \[hep-ex\]](https://arxiv.org/abs/1812.03848) (2018).
58. Henriques Correia, A. M. *The ATLAS Tile Calorimeter* tech. rep. (CERN, Geneva, Mar. 2015). <https://cds.cern.ch/record/2004868>.
59. Khramov, E. *et al.* *Study of the Response of the Hadronic Barrel Calorimeter in the ATLAS Combined Test-beam to Pions of Energies from 20 to 350 GeV for Beam Impact Points from 0.2 to 0.65* tech. rep. ATL-TILECAL-PUB-2009-007. ATL-COM-TILECAL-2009-006 (CERN, Geneva, Apr. 2009). <http://cds.cern.ch/record/1172156>.
60. and, D. G. Performance of the ATLAS Forward Calorimeters in First LHC Data. *Journal of Physics: Conference Series* **293**, 012041. <https://doi.org/10.1088/1742-6596/293/1/012041> (Apr. 2011).
61. *ATLAS muon spectrometer: Technical Design Report* <https://cds.cern.ch/record/331068> (CERN, Geneva, 1997).
62. *The ATLAS muon spectrometer* <https://atlas.cern/Discover/Detector/Muon-Spectrometer>.
63. Yamamoto, A. *et al.* The ATLAS central solenoid. *Nucl. Instrum. Meth.* **A584**, 53–74 (2008).
64. ATLAS barrel toroid: Technical design report (1997).
65. ATLAS endcap toroids: Technical design report (1997).
66. *The ATLAS magnet system* <https://atlas.cern/Discover/Detector/Magnet-System>.

-
67. Jenni, P., Nordberg, M., Nessi, M. & Jon-And, K. ATLAS Forward Detectors for Measurement of Elastic Scattering and Luminosity (2008).
68. White, S. The ATLAS zero degree calorimeter. *Nucl. Instrum. Meth. A* **617** (eds Chiarelli, G., Cervelli, F., Forti, F. & Scribano, A.) 126–128 (2010).
69. Caforio, D. *Luminosity measurement using Cherenkov Integrating Detector (LUCID) in ATLAS in 10th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications* (2008), 413–417.
70. Sidoti, A. Minimum Bias Trigger Scintillators in ATLAS Run II. *JINST* **9**, C10020 (2014).
71. Cindro, V. *et al.* The ATLAS Beam Conditions Monitor. *Journal of Instrumentation* **3**, P02004–P02004. <https://doi.org/10.1088/1748-0221/3/02/p02004> (Feb. 2008).
72. (eds Ferro, F. & Lami, S.) *Proceedings, Workshop on Forward Physics at the LHC: La Biodola, Isola d’Elba, Italy, May 27-29, 2010* (Dec. 2010). arXiv: [1012.5169](https://arxiv.org/abs/1012.5169) [hep-ex].
73. collaboration, A. Performance of the ATLAS trigger system in 2015. *The European Physical Journal C* **77**, 317. ISSN: 1434-6052. <https://doi.org/10.1140/epjc/s10052-017-4852-3> (May 2017).
74. *ATLAS level-1 trigger: Technical Design Report* tech. rep. (Geneva, 1998). <https://cds.cern.ch/record/381429>.
75. Collaboration, A. The ATLAS Level-1 Calorimeter Trigger. *Journal of Instrumentation* **3**, P03001. <http://stacks.iop.org/1748-0221/3/i=03/a=P03001> (2008).
76. Collaboration, A. The Level-1 Trigger Muon Barrel System of the ATLAS experiment at CERN. *Journal of Instrumentation* **4**, P04010. <http://stacks.iop.org/1748-0221/4/i=04/a=P04010> (2009).
77. Pontecorvo, L. The ATLAS Muon Spectrometer. revised version number 1 submitted on 2003-07-27 16:31:16. <https://cds.cern.ch/record/676896> (June 2003).
78. Collaboration, A. The ATLAS central level-1 trigger logic and TTC system. *Journal of Instrumentation* **3**, P08002 (Aug. 2008).
79. Collaboration, T. A. T. The ATLAS Data Acquisition and High Level Trigger system. *Journal of Instrumentation* **11**, P06008. <http://stacks.iop.org/1748-0221/11/i=06/a=P06008> (2016).
80. Kawamoto, T. *et al.* New Small Wheel Technical Design Report (June 2013).
81. Aad, G. *et al.* Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data. *JINST* **14**, P12006. arXiv: [1908.00005](https://arxiv.org/abs/1908.00005) [hep-ex] (2019).
82. Aad, G. *et al.* Reconstruction of hadronic decay products of tau leptons with the ATLAS experiment. *Eur. Phys. J. C* **76**, 295. arXiv: [1512.05955](https://arxiv.org/abs/1512.05955) [hep-ex] (2016).
83. Aaboud, M. *et al.* Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **78**, 903. arXiv: [1802.08168](https://arxiv.org/abs/1802.08168) [hep-ex] (2018).

84. Cornelissen, T. *et al.* *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)* tech. rep. (CERN, Geneva, Mar. 2007). <https://cds.cern.ch/record/1020106>.
85. Cornelissen, T. *et al.* The new ATLAS track reconstruction (NEWT). *Journal of Physics: Conference Series* **119**, 032014. <https://doi.org/10.1088/1742-6596/119/3/032014> (July 2008).
86. *Tracking Results and Comparison to Monte Carlo simulation at $\sqrt{s} = 900$ GeV* tech. rep. (CERN, Geneva, July 2010). <https://cds.cern.ch/record/1276323>.
87. *Performance of the ATLAS Silicon Pattern Recognition Algorithm in Data and Simulation at $\sqrt{s} = 7$ TeV* tech. rep. (CERN, Geneva, July 2010). <https://cds.cern.ch/record/1281363>.
88. *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment* tech. rep. (CERN, Geneva, Mar. 2012). <https://cds.cern.ch/record/1435196>.
89. Collaboration, A. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC run 2. *Eur. Phys. J. C* **77**, 673. 44 p. arXiv: [1704.07983](https://cds.cern.ch/record/2261156). <https://cds.cern.ch/record/2261156> (Apr. 2017).
90. Perigee representation and ATLAS tracking flow chart. <https://atlassoftwaredocs.web.cern.ch/trackingTutorial/idooverview/>.
91. Belau, E. *et al.* Charge collection in silicon strip detectors. *Nuclear Instruments and Methods in Physics Research* **214**, 253–260. ISSN: 0167-5087. <https://www.sciencedirect.com/science/article/pii/0167508783905914> (1983).
92. Gavrilenko, I. *Description of Global Pattern Recognition Program (XKalman)* tech. rep. (CERN, Geneva, Apr. 1997). <https://cds.cern.ch/record/686017>.
93. *Alignment of the ATLAS Inner Detector and its Performance in 2012* tech. rep. (CERN, Geneva, July 2014). <https://cds.cern.ch/record/1741021>.
94. Collaboration, A. Alignment of the ATLAS Inner Detector in Run-2. *Eur. Phys. J. C* **80**, 1194. 41 p. arXiv: [2007.07624](https://cds.cern.ch/record/2724037). <https://cds.cern.ch/record/2724037> (July 2020).
95. *Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV* tech. rep. (CERN, Geneva, Dec. 2015). <http://cds.cern.ch/record/2110140>.
96. *Charged-particle distributions in $\sqrt{s} = 13$ TeV pp interactions measured with the ATLAS detector at the LHC* tech. rep. (CERN, Geneva, July 2015). <https://cds.cern.ch/record/2037701>.
97. Piacquadio, G., Prokofiev, K. & Wildauer, A. Primary vertex reconstruction in the ATLAS experiment at LHC. *Journal of Physics: Conference Series* **119**, 032033. <https://doi.org/10.1088/1742-6596/119/3/032033> (July 2008).

-
98. Collaboration, A. 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* **77**, 332. 52 p. arXiv: [1611.10235](#). <https://cds.cern.ch/record/2235651> (Nov. 2016).
99. *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13\text{TeV}$* tech. rep. (CERN, Geneva, July 2015). <http://cds.cern.ch/record/2037717>.
100. Aad, G. *et al.* Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C* **77**, 490. arXiv: [1603.02934](#) [[hep-ex](#)] (2017).
101. Aaboud, M. *et al.* Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C* **77**, 466. arXiv: [1703.10485](#) [[hep-ex](#)] (2017).
102. *Improving jet substructure performance in ATLAS using Track-CaloClusters* tech. rep. (CERN, Geneva, July 2017). <https://cds.cern.ch/record/2275636>.
103. Aad, G. *et al.* Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions. *Eur. Phys. J. C* **81**, 334. arXiv: [2009.04986](#) [[hep-ex](#)] (2021).
104. Cacciari, M., Salam, G. P. & Soyez, G. FastJet user manual. *The European Physical Journal C* **72**. <https://doi.org/10.1140%2Fepj%2Fs10052-012-1896-2> (Mar. 2012).
105. Cacciari, M., Salam, G. P. & Soyez, G. The anti-kt jet clustering algorithm. *Journal of High Energy Physics* **2008**, 063–063. <https://doi.org/10.1088%2F1126-6708%2F2008%2F04%2F063> (Apr. 2008).
106. *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS* tech. rep. (CERN, Geneva, June 2017). <https://cds.cern.ch/record/2268678>.
107. Aad, G. *et al.* Performance of jet substructure techniques for large- R jets in proton–proton collisions at $\sqrt{s} = 7\text{ TeV}$ using the ATLAS detector. *JHEP* **09**, 076. arXiv: [1306.4945](#) [[hep-ex](#)] (2013).
108. Schramm, S. *ATLAS Jet Reconstruction, Calibration, and Tagging of Lorentz-boosted Objects* tech. rep. (CERN, Geneva, Nov. 2017). <https://cds.cern.ch/record/2291608>.
109. Aaboud, M. *et al.* In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector. *Eur. Phys. J. C* **79**, 135. arXiv: [1807.09477](#) [[hep-ex](#)] (2019).
110. *JetTruthLabelingTool.cxx* <https://gitlab.cern.ch/atlas/athena/-/blob/21.2/PhysicsAnalysis/AnalysisCommon/ParticleJetTools/Root/JetTruthLabelingTool.cxx>.
111. *Optimisation of the ATLAS b -tagging performance for the 2016 LHC Run* tech. rep. (CERN, Geneva, June 2016). <https://cds.cern.ch/record/2160731>.

112. Aad, G. *et al.* ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **79**, 970. arXiv: [1907.05120 \[hep-ex\]](#) (2019).
113. *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run* tech. rep. (CERN, Geneva, July 2017). <https://cds.cern.ch/record/2273281>.
114. *Secondary vertex finding for jet flavour identification with the ATLAS detector* tech. rep. (CERN, Geneva, June 2017). <https://cds.cern.ch/record/2270366>.
115. *Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC* tech. rep. (CERN, Geneva, Oct. 2018). <https://cds.cern.ch/record/2645405>.
116. *Expected performance of the 2019 ATLAS b -taggers* <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/>.
117. *Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration* tech. rep. (CERN, Geneva, May 2020). <https://cds.cern.ch/record/2718610>.
118. Aaboud, M. *et al.* Measurements of b -jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP* **08**, 089. arXiv: [1805.01845 \[hep-ex\]](#) (2018).
119. *Measurement of b -tagging Efficiency of c -jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector* tech. rep. (CERN, Geneva, Mar. 2018). <https://cds.cern.ch/record/2306649>.
120. *Calibration of light-flavour b -jet mistagging rates using ATLAS proton-proton collision data at $\sqrt{s} = 13$ TeV* tech. rep. (CERN, Geneva, Apr. 2018). <https://cds.cern.ch/record/2314418>.
121. Aaboud, M. *et al.* Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **79**, 639. arXiv: [1902.04655 \[physics.ins-det\]](#) (2019).
122. Aad, G. *et al.* Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **81**, 578. arXiv: [2012.00578 \[hep-ex\]](#) (2021).
123. *Overlap removal strategy* https://indico.cern.ch/event/631313/contributions/2683959/attachments/1518878/2373377/Farrell_ORTools_ftaghbb.pdf.
124. Cacciari, M., Salam, G. P. & Soyez, G. The Catchment Area of Jets. *JHEP* **04**, 005. arXiv: [0802.1188 \[hep-ph\]](#) (2008).
125. Aad, G. *et al.* Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D* **105**, 092002. arXiv: [2202.07288 \[hep-ex\]](#) (2022).
126. *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS* tech. rep. (CERN, Geneva, July 2020). <https://cds.cern.ch/record/2724739>.

127. Moreno, E. A. *et al.* Interaction networks for the identification of boosted $H \rightarrow b\bar{b}$ decays. *Phys. Rev. D* **102**, 012010. arXiv: [1909.12285 \[hep-ex\]](#) (2020).
128. Agostinelli, S. *et al.* Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506**, 250–303. ISSN: 0168-9002. <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (2003).
129. Gleisberg, T. *et al.* Event generation with SHERPA 1.1. eprint: [arXiv:0811.4622](#) (2008).
130. Ball, R. D. *et al.* Parton distributions for the LHC Run II. *JHEP* **04**, 040. arXiv: [1410.8849 \[hep-ph\]](#) (2015).
131. Bähr, M. *et al.* Herwig++ physics and manual. *The European Physical Journal C* **58**, 639–707. ISSN: 1434-6052. <http://dx.doi.org/10.1140/epjc/s10052-008-0798-9> (Nov. 2008).
132. Alioli, S., Nason, P., Oleari, C. & Re, E. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP* **06**, 043. arXiv: [1002.2581 \[hep-ph\]](#) (2010).
133. Sjöstrand, T. e. a. An Introduction to PYTHIA 8.2. eprint: [arXiv:1410.3012](#) (2014).
134. Carrazza, S., Forte, S. & Rojo, J. *Parton Distributions and Event Generators in Proceedings, 43rd International Symposium on Multiparticle Dynamics (ISMD 13)* (2013), 89–96. arXiv: [1311.5887 \[hep-ph\]](#).
135. Lange, D. J. The EvtGen particle decay simulation package. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **462**. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, 152–155. ISSN: 0168-9002. <http://www.sciencedirect.com/science/article/pii/S0168900201000894> (2001).
136. Frixione, S., Stoeckli, F., Torrielli, P., Webber, B. R. & White, C. D. The MCanLO 4.0 Event Generator. arXiv: [1010.0819 \[hep-ph\]](#) (Oct. 2010).
137. *Request of multi-jet JZXWithSW samples with a truth Large-R jet + b-hadron filter* <https://its.cern.ch/jira/browse/ATLMCPROD-8878>.
138. ATLAS Collaboration. *Search for Higgs Boson Production via Weak Boson Fusion in Association with a High-Energy Photon at $\sqrt{s} = 13$ TeV with the ATLAS detector* tech. rep. ATL-COM-PHYS-2019-496 (CERN, Geneva, 2019). <https://cds.cern.ch/record/2673667>.
139. Lindert, J. M. *et al.* Precise predictions for $V + \text{jets}$ dark matter backgrounds. *Eur. Phys. J. C* **77**, 829. arXiv: [1705.04664 \[hep-ph\]](#) (2017).
140. Aaboud, M. *et al.* Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Phys. Rev. D* **98**, 012003. arXiv: [1801.02052 \[hep-ex\]](#) (2018).
141. Battaglia, M. *et al.* *Measurement of Inclusive Higgs Boson Production at High p_T^H in the $H \rightarrow b\bar{b}$ Decay Mode* tech. rep. ATL-COM-PHYS-2019-1425 (CERN, Geneva, Nov. 2019). <https://cds.cern.ch/record/2703097>.

142. *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, June 2019). <https://cds.cern.ch/record/2677054>.
143. Verkerke, W. & Kirkby, D. *The RooFit toolkit for data modeling* 2003. <https://arxiv.org/abs/physics/0306116>.
144. Gligorov, V. V., Hageboeck, S., Nanut, T., Sciandra, A. & Tou, D. Y. Avoiding biases in binned fits. *JINST* **16**, T08004. arXiv: [2104.13879 \[physics.data-an\]](#) (2021).
145. Aad, G. *et al.* Modelling and computational improvements to the simulation of single vector-boson plus jet processes for the ATLAS experiment. arXiv: [2112.09588 \[hep-ex\]](#) (Dec. 2021).
146. Aaboud, M. *et al.* Performance of top-quark and W -boson tagging with ATLAS in Run 2 of the LHC. *Eur. Phys. J. C* **79**, 375. arXiv: [1808.07858 \[hep-ex\]](#) (2019).
147. Aad, G. *et al.* Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.* **114**, 191803. arXiv: [1503.07589 \[hep-ex\]](#) (2015).
148. Aaboud, M. *et al.* Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector. *Phys. Lett. B* **786**, 59–86. arXiv: [1808.08238 \[hep-ex\]](#) (2018).
149. Aad, G. *et al.* Measurements of WH and ZH production in the $H \rightarrow b\bar{b}$ decay channel in pp collisions at 13 TeV with the ATLAS detector. *Eur. Phys. J. C* **81**, 178. arXiv: [2007.02873 \[hep-ex\]](#) (2021).
150. Aad, G. *et al.* 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* **816**, 136204. arXiv: [2008.02508 \[hep-ex\]](#) (2021).
151. Aad, G. *et al.* Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C* **81**, 537. arXiv: [2011.08280 \[hep-ex\]](#) (2021).
152. Aad, G. *et al.* Measurement of Higgs boson decay into b -quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. arXiv: [2111.06712 \[hep-ex\]](#) (Nov. 2021).
153. Aad, G. *et al.* Constraints on Higgs boson production with large transverse momentum using $H \rightarrow b\bar{b}$ decays in the ATLAS detector. arXiv: [2111.08340 \[hep-ex\]](#) (Nov. 2021).
154. Sirunyan, A. M. *et al.* Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP* **12**, 085. arXiv: [2006.13251 \[hep-ex\]](#) (2020).
155. Berger, N. *et al.* Simplified Template Cross Sections - Stage 1.1. arXiv: [1906.02754 \[hep-ph\]](#) (June 2019).
156. Contino, R., Ghezzi, M., Grojean, C., Muhlleitner, M. & Spira, M. Effective Lagrangian for a light Higgs-like scalar. *JHEP* **07**, 035. arXiv: [1303.3876 \[hep-ph\]](#) (2013).

157. De Florian, D. *et al.* Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. **2/2017**. arXiv: [1610.07922 \[hep-ph\]](#) (Oct. 2016).
158. Methodology for EFT interpretation of Higgs boson Simplified Template Cross-section results in ATLAS (Oct. 2019).
159. Example ATLAS tunes of Pythia8, Pythia6 and Powheg to an observable sensitive to Z boson transverse momentum (2013).
160. Pumplin, J. *et al.* New generation of parton distributions with uncertainties from global QCD analysis. *JHEP* **07**, 012. arXiv: [hep-ph/0201195](#) (2002).
161. Ferrera, G., Somogyi, G. & Tramontano, F. Associated production of a Higgs boson decaying into bottom quarks at the LHC in full NNLO QCD. *Phys. Lett. B* **780**, 346–351. arXiv: [1705.10304 \[hep-ph\]](#) (2018).
162. Caola, F., Luisoni, G., Melnikov, K. & Röntsch, R. NNLO QCD corrections to associated WH production and $H \rightarrow b\bar{b}$ decay. *Phys. Rev. D* **97**, 074022. arXiv: [1712.06954 \[hep-ph\]](#) (2018).
163. Anastasiou, C. *et al.* High precision determination of the gluon fusion Higgs boson cross-section at the LHC. *JHEP* **05**, 058. arXiv: [1602.00695 \[hep-ph\]](#) (2016).
164. Bonetti, M., Melnikov, K. & Tancredi, L. Higher order corrections to mixed QCD-EW contributions to Higgs boson production in gluon fusion. *Phys. Rev. D* **97**. [Erratum: *Phys.Rev.D* 97, 099906 (2018)], 056017. arXiv: [1801.10403 \[hep-ph\]](#) (2018).
165. Bolzoni, P., Maltoni, F., Moch, S.-O. & Zaro, M. Higgs production via vector-boson fusion at NNLO in QCD. *Phys. Rev. Lett.* **105**, 011801. arXiv: [1003.4451 \[hep-ph\]](#) (2010).
166. Frixione, S., Hirschi, V., Pagani, D., Shao, H. -S. & Zaro, M. Electroweak and QCD corrections to top-pair hadroproduction in association with heavy bosons. *JHEP* **06**, 184. arXiv: [1504.03446 \[hep-ph\]](#) (2015).
167. Demartin, F., Maltoni, F., Mawatari, K. & Zaro, M. Higgs production in association with a single top quark at the LHC. *Eur. Phys. J. C* **75**, 267. arXiv: [1504.00611 \[hep-ph\]](#) (2015).
168. Rogozhnikov, A. Reweighting with Boosted Decision Trees. *J. Phys. Conf. Ser.* **762** (eds Salinas, L. & Torres, C.) 012036. arXiv: [1608.05806 \[physics.data-an\]](#) (2016).
169. Roe, B. P., Yang, H.-J. & Zhu, J. *Boosted decision trees, a powerful event classifier in PHYSTAT (2005): Statistical Problems in Particle Physics, Astrophysics and Cosmology* (Sept. 2005), 139–142.
170. *Gradient Boosted Reweighter* https://arogozhnikov.github.io/hep_ml/reweight.html.
171. Hodges, J. L. The significance probability of the smirnov two-sample test. *Arkiv för Matematik* **3**, 469–486. <https://doi.org/10.1007/BF02589501> (1958).
172. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**. ROC Analysis in Pattern Recognition, 861–874. ISSN: 0167-8655. <https://www.sciencedirect.com/science/article/pii/S016786550500303X> (2006).

173. Michelucci, U. & Venturini, F. Estimating Neural Network's Performance with Bootstrap: A Tutorial. *Machine Learning and Knowledge Extraction* **3**, 357–373. ISSN: 2504-4990. <https://www.mdpi.com/2504-4990/3/2/18> (2021).
174. Differential $t\bar{t}$ cross-section measurements using boosted top quarks in the all-hadronic final state with 139 fb^{-1} of ATLAS data. arXiv: [2205.02817 \[hep-ex\]](#) (May 2022).
175. Cowan, G., Cranmer, K., Gross, E. & Vitells, O. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C* **71**. [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)], 1554. arXiv: [1007.1727 \[physics.data-an\]](#) (2011).
176. Becker, K. *et al.* Precise predictions for boosted Higgs production. arXiv: [2005.07762 \[hep-ph\]](#) (May 2020).
177. Evaluation of theoretical uncertainties for simplified template cross section measurements of V -associated production of the Higgs boson (2018).
178. *Energy gap of different materials* <https://physics.stackexchange.com/questions/467901/what-is-energy-band-gap>.
179. *Fermi function* <http://hyperphysics.phy-astr.gsu.edu/hbase/Solids/Fermi.html>.
180. *Silicon Detectors* <https://indico.cern.ch/event/124392/contributions/1339904/attachments/74582/106976/IntroSilicon.pdf>.
181. Sigmund, P. Particle Penetration and Radiation Effects. *Particle Penetration and Radiation Effects*, by Peter Sigmund. Berlin: Springer, 2008. ISBN: 978-3-540-72622-7 (Jan. 2008).
182. Ramo, S. Currents Induced by Electron Motion. *Proceedings of the IRE* **27**, 584–585 (1939).
183. Bomben, M. *Silicon Trackers for High Luminosity Colliders* PhD thesis (Paris U., VI-VII, 2018).
184. Taibah, R. H. M. *Upgrade of the ATLAS tracking detector in preparation for the High-Luminosity phase of the LHC : Planar pixel module characterisation and calibration for the current and future ATLAS inner tracker*. *Mise jour du trajectographe d'ATLAS en préparation de la phase haute luminosité du CERN LHC : étalonnage et caractérisation de modules à pixels du détecteur de traces actuel et futur d'ATLAS* Presented 17 Sep 2021 (Oct. 2021). <https://cds.cern.ch/record/2791088>.
185. Alam, M. S. *et al.* The ATLAS silicon pixel sensors. *Nucl. Instrum. Meth. A* **456**, 217–232 (2001).
186. Darbo, G. Experience on 3D Silicon Sensors for ATLAS IBL. *JINST* **10**, C05001. arXiv: [1411.6937 \[physics.ins-det\]](#) (2015).
187. Lutz, G. *Semiconductor Radiation Detectors: Device Physics* ISBN: 978-3-540-64859-8 (Springer, New York, 1999).
188. Moll, M. *Radiation damage in silicon particle detectors: Microscopic defects and macroscopic properties* PhD thesis (Hamburg U., 1999).

189. Radiation Simulation Public Results in ATLAS experiment. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/RadiationSimulationPublicRes>
190. Aaboud, M. *et al.* Study of the material of the ATLAS inner detector for Run 2 of the LHC. *JINST* **12**, P12009. arXiv: [1707.02826](https://arxiv.org/abs/1707.02826) [[hep-ex](#)] (2017).
191. Collaboration, A. *Technical Design Report for the ATLAS ITk Pixel Detector* tech. rep. (CERN, Geneva, Mar. 2018). <https://cds.cern.ch/record/2310230>.
192. AIDA-2020 <https://aida2020.web.cern.ch/aida2020/>.
193. AIDainnova <https://aidainnova.web.cern.ch>.
194. Garcia-Sciveres, M. *The RD53A Integrated Circuit* tech. rep. (CERN, Geneva, Oct. 2017). <https://cds.cern.ch/record/2287593>.
195. Garcia-Sciveres, M., Loddo, F. & Christiansen, J. *RD53B Manual* tech. rep. (CERN, Geneva, Mar. 2019). <https://cds.cern.ch/record/2665301>.
196. Whallon, N. L. *et al.* Upgrade of the YARR DAQ system for the ATLAS Phase-II pixel detector readout chip. *PoS TWEPP-17*, 076 (2018).
197. Gerbershagen, A. Overview over CERN SPS test beams. *6th beam telescope and test beam workshop* (2018).
198. Diener, R. *et al.* The DESY II Test Beam Facility. *Nucl. Instrum. Meth. A* **922**, 265–286. arXiv: [1807.09328](https://arxiv.org/abs/1807.09328) [[physics.ins-det](#)] (2019).
199. Rubinskiy, I. An EUDET/AIDA Pixel Beam Telescope for Detector Development. *Phys. Procedia* **37** (ed Liu, T.) 923–931 (2012).
200. Jansen, H. *et al.* Performance of the EUDET-type beam telescopes. *EPJ Tech. Instrum.* **3**, 7. arXiv: [1603.09669](https://arxiv.org/abs/1603.09669) [[physics.ins-det](#)] (2016).
201. Bisanz, T. *et al.* EU Telescope: A modular reconstruction framework for beam telescope data. *JINST* **15**, P09020. arXiv: [2011.10356](https://arxiv.org/abs/2011.10356) [[physics.ins-det](#)] (2020).
202. *TBmon2 framework* <https://gitlab.cern.ch/tbmon2>.
203. Kleinwort, C. General Broken Lines as advanced track fitting method. *Nucl. Instrum. Meth. A* **673**, 107–110. arXiv: [1201.4320](https://arxiv.org/abs/1201.4320) [[physics.ins-det](#)] (2012).
204. Rossi, L., Fischer, P., Rohe, T. & Wermes, N. *Pixel detectors: from fundamentals to applications* <https://cds.cern.ch/record/976471> (Springer, Berlin, 2006).
205. Terzo, S. *Development of radiation hard pixel modules employing planar n-in-p silicon-sensors with active edges for the ATLAS detector at HL-LHC* (2015). <https://cds.cern.ch/record/2676574>.
206. Bisanz, T. Test-beam activities and results for the ATLAS ITk pixel detector. *JINST* **12**, C12053 (2017).
207. Kröger, J., Spannagel, S. & Williams, M. User Manual for the Corryvreckan Test Beam Data Reconstruction Framework, Version 1.0. arXiv: [1912.00856](https://arxiv.org/abs/1912.00856) [[physics.ins-det](#)] (Dec. 2019).
208. Abidi, H. *et al.* Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a pair of *b*-quarks in the fully hadronic channel tech. rep. (CERN, Geneva, Apr. 2022). <https://cds.cern.ch/record/2806588>.

