

**Universidade do Minho**

Escola de Ciências

Juan Pedro Araque Espinosa

## **Search for heavy fermions with the ATLAS experiment at the LHC collider**

Tese de Doutoramento em Física

Trabalho efetuado sob orientação do

**Professor Doutor Nuno Filipe da Silva Fernandes de Castro**

e coorientação do

**Professor Doutor José Santiago Pérez**

e do

**Professor Doutor António Joaquim Onofre de Abreu**

**Ribeiro de Gonçalves**

outubro de 2015

## DECLARAÇÃO

Nome: Juan Pedro Araque Espinosa

E-mail: jp.araque@cern.ch

Telefone: +351964640378

Número de Bilhete de Identidade: 77356769-H

Título da tese:

Search for heavy fermions with the ATLAS experiment at the LHC collider.

Orientador:

Professor Doutor Nuno Filipe da Silva Fernandes de Castro

E co-orientação de:

Professor Doutor José Santiago Pérez

Professor Doutor António Joaquim Onofre de Abreu Ribeiro de Gonçalves

Ano de Conclusão: 2015

Designação do Doutoramento:

Programa Doutoral em Física

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, 28/10/2015

Assinatura: Juan Pedro Araque Espinosa

---

# Declaração de integridade

Declaro ter atuado com integridade na elaboração da presente tese. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou qualquer forma de falsificação de resultados, referenciando devidamente todo o trabalho utilizado para a elaboração da tese aqui apresentada.

Universidade do Minho, 28 de Outubro de 2015

Juan Pedro Araque Espinosa





---

# Acknowledgments

If there is a word that needs to be mentioned once this thesis has been finished, that word is thank you. The work presented here would have not been possible without the help of a lot of people.

I would like to thank, of course, my advisor Nuno Castro for so many reasons that I am not able to enumerate, but most importantly for being more than an advisor and helping me to achieve the goals we proposed when this thesis was conceived. I would also like to thank José Santiago for his help in different aspects of this thesis and for all the useful conversations we had. I would like to thank as well António Onofre for everything I learned from him at the beginning of my Ph.D. and for all his support.

As part of the Portuguese ATLAS group I have to thank Amelia Maio and Patricia Conde Muiño for their support and the hard work they have done as group leaders during the past four years.

During these years there has been people I have worked with that I need to thank because they have made possible a big part of the work presented here. I have to thank Mark Cook, without his help and support the  $Z$ -tag group, and the analysis I present in this thesis, would not have been the same. A special thank goes to Clare Bernard and Matt Leone for all the hours we have spent checking those event yields and trying to match everything in all the channels.

I would really like to thank Ana Henriques for her support in the TileCal group and for the help she offered me whenever I needed it. And of course, my colleagues that worked with me in the TileCal noise description Pavel Storovoitov, Claudio Santoni and Aliaksei Hrynevich.

Needless to say, I need to thank all the people in the  $HQT$  subgroup in ATLAS for all the discussions and work we have done since my first presentation (when it was still called  $4gt$ ). I have to especially mention Tobias Golling, Johannes Erdman, James Ferrando, Xiaowen Lei and Clement Helsen. Thank you all for being there

and making all this possible.

Coming back to Portugal I have to thank all the people of the ATLAS group at LIP, especially Bruno Galhardo, Susana Santos, Miguel Fiolhais and Filipe Veloso for being there when I needed either someone to check a code or discuss a technical aspect of an analysis, or talk about whatever nonsense that day was in my mind. I want to thank Vítor Oliveira for all the hard work he has done to maintain the computational infrastructure up and running, making all this possible. Without you everything would have been much harder. And also Henrique Carvalho, who knows that my geek side is always looking for new tools to build.

All that I am today as a scientist was defined in Granada and of course I cannot forget the people who inspired me. I want to thank Paco del Águila, Fernando Cornet, Jose Ignacio Illana and Manuel Masip for being so accessible and inspiring.

Talking about Granada I have to mention all my good friends I have shared lots of moments with. I would like to thank Miki, Jordi, Migue, Rafa, Fernando, Maria del Mar, Esperanza, Javi, Laura and I am sure I have forgotten someone, thank you all. Special thanks to Miki for all the endless discussion about the theoretical side of almost everything. And I cannot forget my favourite *guiri* Aaron Key, who has helped me in very different areas of the work presented here.

And last but, of course, not least, my family. They are those who define me, who make me who I am and have supported me in everything. I want to thank my parents Marien and Kiko, and my two brothers Francis and Enrique. I want to thank Vane, who has been by my side all these years, as well as her family, who I also want to thank.

This thesis has been supported by different projects and grants. It was funded by the project *Colaboração na experiência ATLAS*, Ref. CERN/FP/116346/2010 from October 1<sup>st</sup>, 2011 until April 1<sup>st</sup>, 2012, the project *Colaboração na experiência ATLAS*, Ref. CERN/FP/116346/2011 from April 2<sup>nd</sup>, 2011 until December 31<sup>st</sup> and by the FCT grant SFRH / BD / 52002 / 2012, within the IDPASC doctoral network, since January 1<sup>st</sup>, 2013.



---

*A toda mi familia,  
por siempre apoyarme y por esa frase que nunca olvidaré:  
“Haz lo que quieras, porque sabemos que lo harás bien.”*

*A Vane y su familia,  
por haber sido la mía  
durante todos estos años.*

*A María, Lela,  
gracias.*



---

# Abstract

## Search for heavy fermions with the ATLAS experiment at the LHC collider

In the present thesis a search for new heavy fermions using LHC data collected in 2012 by the ATLAS experiment is presented. In particular, a search for pair and single production of vector-like quarks with electric charge  $2/3$  ( $T$ ) and  $-1/3$  ( $B$ ) decaying to a  $Z$  boson is discussed. For this search the analysis was divided in two channels, depending on the lepton multiplicity, and both channels were combined at the end for the final results. Since no evidence for signal was found, limits on the mass of the vector-like quarks were evaluated. The observed (expected) limit on the mass of an  $SU(2)$  singlet  $T$  quark is 655 GeV (625 GeV), while the observed (expected) limit on the mass of a  $T$  quark in a  $(T, B)$  doublet is 735 GeV (720 GeV). The observed (expected) limit on the mass of an  $SU(2)$  singlet  $B$  quark is 685 GeV (670 GeV), while the observed (expected) limit on the mass of a  $B$  quark in a  $(B, Y)$  doublet is 755 GeV (755 GeV).

The impact of a heavy gluon in the searches for pair production of vector-like quarks has also been studied. Even if kinematic differences were found at the parton level, these seem to be negligible after taking the detector simulation and event reconstruction effects into account. The ATLAS results obtained within the scope of the present thesis were recasted in a realistic scenario and it was found that the selection efficiencies for signal, as well as the final results, are not sensitive to the presence of a heavy gluon. Such conclusion allowed to reinterpret all the relevant ATLAS and CMS results by simply scaling the cross-section, obtaining limits on the mass of the heavy gluon.



---

# Resumo

## Pesquisa de novos fermiões na experiência ATLAS do LHC

Na presente tese é apresentada uma pesquisa de novos fermiões usando os dados de LHC coletados em 2012 pela experiência ATLAS. Em particular, a produção simples e de pares de quarks vetoriais com carga elétrica de  $2/3$  ( $T$ ) e  $-1/3$  ( $B$ ) que decaem para bosões  $Z$  é discutida. Neste estudo a análise de dados foi dividida em dois canais, em função da multiplicidade de léptons, sendo estes combinados para obter um resultado final. Não tendo sido observada nenhuma evidência de sinal nos dados analisados, limites na massa dos quarks vetoriais foram obtidos. O limite observado (esperado) na massa de um quark  $T$  é 655 GeV (625 GeV) para o caso de um singlete de  $SU(2)$ , sendo 735 GeV (720 GeV) no caso de um dubleto ( $T, B$ ). Os correspondentes limites para um quark vectorial  $B$  são 685 GeV (670 GeV) para um singlete e 755 GeV (755 GeV) para um dubleto ( $B, Y$ ).

O impacto de um hipotético glúão pesado na pesquisa da produção de pares de quarks vetoriais foi também estudado. Concluiu-se que apesar de algumas diferenças cinemáticas existirem a nível partónico, estas diluem-se consideravelmente após considerar a simulação do detetor e os efeitos de reconstrução de acontecimentos. Foram então reinterpretados os resultados de ATLAS obtidos no contexto desta tese, observando-se que as diferenças de eficiência de seleção de sinal, bem como os resultados finais, não mudam significativamente na presença de glúões pesados. Esta observação permitiu justificar a reinterpretação dos restantes resultados de ATLAS e CMS considerando apenas a diferença de secções eficazes de forma a obter limites na massa do glúão pesado.





---

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theoretical introduction</b>	<b>3</b>
2.1	The Standard Model of particle physics . . . . .	3
2.2	Beyond the Standard Model . . . . .	7
2.2.1	Vector-like quarks . . . . .	8
2.2.2	Composite Higgs models . . . . .	14
<b>3</b>	<b>Experimental setup</b>	<b>21</b>
3.1	CERN . . . . .	21
3.2	The Large Hadron Collider . . . . .	23
3.3	The ATLAS detector . . . . .	25
3.3.1	Inner detector . . . . .	26
3.3.2	Calorimetry . . . . .	29
3.3.3	Muon Spectrometer . . . . .	30
3.3.4	Magnet system . . . . .	31
3.3.5	Luminosity subdetectors . . . . .	32
3.3.6	Triggering system . . . . .	33
3.4	Worldwide LHC Computer Grid . . . . .	34
<b>4</b>	<b>Pile-up noise characterisation in the TileCal Calorimeter</b>	<b>37</b>
4.1	Pile-up inside the ATLAS detector . . . . .	37
4.2	Pile-up noise in TileCal . . . . .	38
4.2.1	Analysis procedure . . . . .	41
4.2.2	Root mean square as the noise estimator . . . . .	43
4.2.3	Energy distribution quantiles as noise estimator . . . . .	46
<b>5</b>	<b>Search for new vector-like quarks</b>	<b>53</b>
5.1	Data sample and object reconstruction . . . . .	53
5.2	Analysis strategy . . . . .	58
5.3	Signal modelling . . . . .	64
5.3.1	Pair production . . . . .	64
5.3.2	Single production . . . . .	67

5.4	Background modelling . . . . .	69
5.5	Data and Monte Carlo prediction comparison . . . . .	72
5.5.1	Dilepton channel for the pair production hypothesis . . . . .	72
5.5.2	Trilepton channel for the pair production hypothesis . . . . .	77
5.5.3	Event selection modification for single production . . . . .	80
5.6	Systematic uncertainties . . . . .	83
5.7	Results . . . . .	88
5.7.1	Limits on the pair production hypothesis . . . . .	89
5.7.2	Limits on the single production hypothesis . . . . .	91
5.8	Other searches for VLQ . . . . .	95
<b>6</b>	<b>Heavy gluon impact in vector-like quark searches</b>	<b>101</b>
6.1	Heavy gluon in CHM . . . . .	102
6.2	Signal and background MC generation . . . . .	103
6.3	Comparison of kinematical distributions . . . . .	104
6.3.1	Kinematical differences at parton level . . . . .	105
6.3.2	Kinematical differences after detector simulation . . . . .	109
6.3.3	Recasting of the VLQ searches at the LHC . . . . .	112
6.4	Limits on heavy gluons from VLQ searches . . . . .	119
6.4.1	Current limits . . . . .	120
6.4.2	Early Run-2 expectations . . . . .	123
<b>7</b>	<b>Conclusions</b>	<b>127</b>
<b>A</b>	<b>Noise vs <math>\langle\mu\rangle</math> using the quantiles estimator</b>	<b>131</b>
<b>B</b>	<b>Data vs. MC comparison using the ALPGEN generator</b>	<b>135</b>
<b>C</b>	<b>Software tools</b>	<b>139</b>
<b>D</b>	<b>The <math>CL_s</math> method</b>	<b>149</b>
<b>E</b>	<b>ATLAS summary results for all channels</b>	<b>153</b>
	<b>Bibliography</b>	<b>156</b>

---

# List of Figures

2.1	Number of neutrinos. . . . .	9
2.2	Best fit results for the production and decay signal strength of the Higgs boson for the ATLAS and CMS combination. . . . .	10
2.3	Branching ratios of vector-like $T$ and $B$ quarks for all decay modes. .	12
2.4	Pair and single production processes for vector-like $T$ and $B$ . . . . .	13
2.5	Pair and single production processes for vector-like $T$ and $B$ . . . . .	14
3.1	CERN accelerators and detector facilities . . . . .	22
3.2	LHC dipole magnet . . . . .	24
3.3	The ATLAS detector . . . . .	25
3.4	Pile-up and luminosity representation for 2012 . . . . .	26
3.5	Event display with representation of 23 vertices in the same collision .	27
3.6	ATLAS Inner Detector . . . . .	28
3.7	ATLAS calorimeter system . . . . .	29
3.8	ATLAS Muon Spectrometer . . . . .	31
3.9	ATLAS magnet system and magnetic field in the toroidal magnets . .	32
3.10	ATLAS trigger and data acquisition systems [88]. . . . .	34
3.11	WLCG tasks share in the past three years . . . . .	35
4.1	Different steps of the topological clustering algorithm. . . . .	39
4.2	Energy distributions for different TileCal cells. . . . .	40
4.3	TileCal schematic: lateral view and cells . . . . .	42
4.4	Noise dependence with $\eta$ : 50 ns bunch-spacing . . . . .	44
4.5	Noise dependence with $\eta$ : 25 ns bunch-spacing . . . . .	45
4.6	Noise dependence with $\mu$ for different TileCal layers and 50 ns bunch-spacing. . . . .	47
4.7	Noise dependence with $\mu$ for different TileCal layers and 25 ns bunch-spacing. . . . .	48
4.8	Quantiles estimator: noise dependence with $\mu$ . . . . .	50

4.9	Comparison between the quantile estimator and the RMS estimator. .	51
5.1	Single electron trigger efficiencies. . . . .	55
5.2	Single muon trigger efficiencies. . . . .	56
5.3	Electron and muons reconstruction efficiencies. . . . .	57
5.4	MV1 $b$ -tag efficiency as a function of the jet $p_T$ . . . . .	58
5.5	Discriminating variables used for the event selection. . . . .	62
5.6	Discriminant variables for the dilepton and trilepton channel. . . . .	63
5.7	Cross-section of pair and single production vector-like $T$ and $B$ . . . .	65
5.8	Comparison of the $m(Zb)$ distribution for singlet and doublet hypotheses. . . . .	66
5.9	Parton level distributions for VLQ pair production signal samples. . .	68
5.10	Comparison between fast and full detector simulation for signal samples.	69
5.11	<b>JetFitter</b> sum distribution for each $N_{\text{tag}}$ region. . . . .	73
5.12	Background correction to the $N_{\text{tag}}$ distribution . . . . .	74
5.13	Invariant mass of the $Z$ boson candidate in the dilepton channel. . . .	75
5.14	Transverse momentum distribution for the $Z$ boson candidate before the $p_T(Z)$ correction. . . . .	76
5.15	Corrections applied to the $p_T(Z)$ distribution for the $Z$ +jets background. . . . .	77
5.16	Transverse momentum distribution for the $Z$ boson candidate after the $p_T(Z)$ correction. . . . .	78
5.17	Distributions for $H_T(\text{jets})$ and $m(Zb)$ for the dilepton channel for $N_{\text{tag}} = 1$ and $N_{\text{tag}} \geq 2$ . . . . .	80
5.18	Mass of the $Z$ boson candidate and central jets multiplicity for the trilepton channel. . . . .	82
5.19	Transverse momentum the $Z$ boson candidate and $b$ -tagged jets multiplicity for the trilepton channel. . . . .	82
5.20	Transverse momentum the $Z$ boson candidate and $b$ -tagged jets multiplicity for the trilepton channel. . . . .	83
5.21	Forward jet multiplicity and invariant mass of the $Zb$ system for the dilepton channel and single production mechanism. . . . .	84
5.22	Forward jet multiplicity and $H_T(\text{jets} + \text{leptons})$ distributions for the dilepton channel and single production mechanism. . . . .	85
5.23	Cross section upper limit at a 95% CL in the pair production hypothesis for the $B$ and $T$ quarks in the singlet and doublets hypothesis. .	92

5.24	Observed and expected lower mass limit at a 95% CL for different branching ratios to all decay modes for vector-like $T$ . . . . .	93
5.25	Observed and expected lower mass limit at a 95% CL for different branching ratios to all decay modes for vector-like $B$ . . . . .	94
5.26	Observed and expected upper limit in $\sigma \times \text{BR}$ for vector like $T$ and $B$ in the single production hypothesis . . . . .	95
5.27	Observed and expected ATLAS summary distributions for lower mass limit 95% CL for each for different branching ratios to all decay modes for vector-like $T$ . . . . .	97
5.28	Observed and expected ATLAS summary distributions for lower mass limit 95% CL for each for different branching ratios to all decay modes for vector-like $B$ . . . . .	98
5.29	Observed and expected ATLAS summary distributions for lower mass limit 95% CL for each for different branching ratios to all decay modes for vector-like $T$ . . . . .	99
5.30	Observed and expected ATLAS summary distributions for lower mass limit 95% CL for each for different branching ratios to all decay modes for vector-like $B$ . . . . .	100
6.1	Distribution of the $p_T(Z)$ at parton level for the QCD, HG and QCD+HG productions channels. . . . .	106
6.2	Distribution of $p_T(X'_{2/3})$ , at parton level for the QCD, HG and QCD+HG productions channels. . . . .	107
6.3	Distribution of the invariant mass of the $X'_{2/3}, \bar{X}'_{2/3}$ system, $M_{X'_{2/3}\bar{X}'_{2/3}}$ , at parton level for the QCD, HG and QCD+HG productions channels. . . . .	108
6.4	Distribution of $p_T(X'_{2/3})$ , at parton level for the QCD, HG and QCD+HG productions channels ( $M_G = 2.5$ TeV and $M_{X'_{2/3}} = 1.3$ TeV). . . . .	109
6.5	Comparison between QCD and QCD+HG after hadronisation, detector simulation and event reconstruction are included. . . . .	111
6.6	Different reconstruction steps to identify the origin of the disappearance of the different found at parton level . . . . .	113
6.7	The $b$ -jet multiplicity, $p_T(Z)$ , $H_T(\text{jets})$ and $M(Zb)$ distribution at each stage of the selection for the recast of the $Zt/b$ analysis. . . . .	115
6.8	Limits found at $\sqrt{s} = 8$ TeV represented in the $M_G - M_{X'_{2/3}}$ plane . . . . .	116
6.9	Input variables for the NN used at $\sqrt{s} = 8$ TeV . . . . .	118

6.10	Output of the NN and limits computed using the output of the NN as the discriminant variable . . . . .	119
6.11	Dependence of $M_{T'}$ with $M_{X'_{2/3}}$ and comparison between expected limits including only $X'_{2/3}$ and including $T'$ as well . . . . .	122
6.12	Dependence of $M_{T'}$ with $M_{X'_{2/3}}$ and comparison between expected limits including only $X'_{2/3}$ and including $T'$ as well . . . . .	124
6.13	Run-2 expectations for the mass limits on $M_G$ and $M_{X'_{2/3}}$ . . . . .	125
6.14	Integrated luminosity currently collected in Run-2. . . . .	126
A.1	Noise with quantiles estimator for layer A. . . . .	132
A.2	Noise with quantiles estimator for layer BC. . . . .	133
A.3	Noise with quantiles estimator for layer D. . . . .	134
A.4	Noise with quantiles estimator for the gap/crack scintillating cells. . .	134
B.1	Invariant mass of the $Z$ boson candidate in the dilepton channel using the <b>ALPGEN</b> generator for the $Z$ +jets background. . . . .	136
B.2	Transverse momentum distribution for the $Z$ boson candidate before the $p_T(Z)$ correction using the <b>ALPGEN</b> generator for the $Z$ +jets back- ground. . . . .	136
B.3	Transverse momentum distribution for the $Z$ boson candidate after the $p_T(Z)$ correction using the <b>ALPGEN</b> generator for the $Z$ +jets back- ground. . . . .	137
B.4	Distributions for $H_T(\text{jets})$ and $m(Zb)$ for the dilepton channel for $N_{\text{tag}} = 1$ and $N_{\text{tag}} \geq 2$ using the <b>ALPGEN</b> generator for the $Z$ +jets background. . . . .	138
D.1	Example of test statistics for signal and background. . . . .	150
D.2	Example of test statistics for signal and background in an analysis with no sensitivity. . . . .	151
E.1	Summary of the ATLAS results on lower bounds in the vector-like $B$ mass. . . . .	154
E.2	Summary of the ATLAS results on lower bounds in the vector-like $B$ mass. . . . .	155

---

# List of Tables

2.1	Constituents of the SM . . . . .	4
2.2	Possible multiplets of vector-like . . . . .	11
3.1	Designed ATLAS subsystem performance. . . . .	28
4.1	Correspondence between CI and percentage of the population in a normal distribution . . . . .	49
5.1	Event selection for both dilepton and trilepton channels and for single and pair production mechanism . . . . .	59
5.2	Predicted and observed event yields in the dilepton channel for pair production in different stages of the analysis in the $N_{\text{tag}} = 1$ category. . . . .	79
5.3	Predicted and observed event yields in the dilepton channel for pair production in different stages of the analysis in the $N_{\text{tag}} \geq 2$ category. . . . .	79
5.4	Event yields in the trilepton channel for pair production . . . . .	81
5.5	Number of observed and expected events for the single production mechanism . . . . .	86
5.6	Systematic uncertainties in the final selection yields for the dilepton channel . . . . .	88
5.7	Systematic uncertainties in the final selection yields for the trilepton channel . . . . .	89
5.8	Summary of mass limits from pair production . . . . .	90
6.1	Efficiencies comparison for the recast of the $Zt/b$ analysis for the QCD and QCD+HG channels . . . . .	114





---

# Acronyms

ALICE	A Large Ion Collider Experiment	EWSB	Electroweak symmetry breaking
ATLAS	A Toroidal LHC Apparatus	FCal	Forward Calorimeter
BCM	Beam Conditions Monitor	FCNC	Flavor changing neutral currents
BR	Branching ratio	Fermilab	Fermi National Accelerator Laboratory
C.L.	Confidence level	HEC	Hadronic End-cap Calorimeter
CDF	Collider Detector at Fermilab	IBL	Inserted $b$ -layer
CERN	European Organization for Nuclear Research	ID	Inner Detector
CHM	Composite Higgs model	ISOLDE	Isotope Separator On Line Device
CKM	Cabbibo-Kobayashi-Maskawa	LAr	Liquid argon electromagnetic calorimeter
CMS	Compact Muon Solenoid	LEP	Large Electron-Positron Collider
CNGS	CERN Neutrinos to Gran Sasso	LH	Left handed
DESY	Deutsches Elektronen-Synchrotron	LHC	Large Hadron Collider
DONUT	Direct Observation of the NU Tau	LO	Leading order
EW	Electroweak		

---

LUCID	Luminosity measurement using a Cherenkov Integrating Detector	QCD	Quantum Chromodynamics
MC	Monte Carlo	RH	Right handed
MCHM	Minimal Composite Higgs model	SCT	Semiconductor Tracker
NGB	Nambu-Goldstone boson	SCT	Semiconductor tracker
NLO	Next-to-leading order	SF-OS	Same-Flavour Opposite-Sign
NNLL	Next-to-next-to-leading log	SLAC	Stanford Linear Accelerator Center
OF	Optimal Filter	SM	Standard Model
PD	Pixel Detector	SPS	Super Proton Synchrotron
PD	Pixel detector	TileCal	Hadronic Tiles Calorimeter
PDF	Parton-distribution function	TRT	Transition Radiation Tracker
PMT	Photomultiplier	VEV	Vacuum expectation value
pNGB	pseudo Nambu-Goldstone boson	VLQ	Vector-like quark
		WCLG	Worldwide LHC Computer Grid

---

# Chapter 1

## Introduction

The work presented in this thesis was defined in the summer of 2011, a few years after the Large Hadron Collider started to accelerate and collide proton bunches. This allowed scientists at CERN to analyse the products of highly energetic proton-proton collisions without precedents, and, therefore, provided the perfect environment to start working in experimental high-energy physics. In fact, in the summer of 2012, the announcement of a new particle found by both the ATLAS and CMS Collaborations compatible with the Standard Model Higgs boson set one of the key dates in the history of physics. Even though the discovery of the Higgs boson provided the missing piece of the puzzle, needed to complete the Standard Model, some questions remain open. In this context, this thesis was conceived to analyse the data collected by the ATLAS detector and to search for new particles predicted by models that aim to answer these open questions, in particular, new heavy quarks.

One might ask now, *why are these new heavy quarks necessary?* In order to answer that question, Chapter 2 presents the Standard Model of particle physics initially, and what we learn from it. Vector-like quarks are also presented as new particles naturally appearing in different models, which try to answer some of the open questions in the SM. For instance, composite Higgs models, which introduce the idea of a Higgs boson as a bound state, instead of an elementary particle, are also discussed, as a special interest of this thesis.

In order to be able to search for vector-like quarks, an entire experimental setup is needed. The ATLAS detector is used to measure proton-proton collisions produced in the LHC. In Chapter 3, CERN, the LHC, the ATLAS detector and the largest computing grid in the world (the Worldwide LHC Computing Grid) are described.

The LHC produces proton-proton collisions at a rate of 40 MHz. In order to

achieve that rate several proton-proton collisions take place in the same bunch-crossing inside the ATLAS detector. On top of this, the time spacing between bunches is smaller than the integration time of some subsystem of the ATLAS detector. In this environment the ATLAS detector has to be able to distinguish between different interactions and identify interesting events. All the adjacent and simultaneous collisions that take place have an impact on the energy measured for an interesting event by the hadronic calorimeter of the ATLAS detector. This is called pile-up noise. In Chapter 4 the study and characterisation of the pile-up noise under different circumstances will be presented.

The search for vector-like quarks will be described in Chapter 5. A detailed description of the analysis performed using data measured with the ATLAS detector at  $\sqrt{s} = 8$  TeV will be presented. The analysis strategy, the definition of each channel considered, the signal and background Monte Carlo simulation used in the analysis, the data and object reconstruction used, the comparison between the data measurement and background expectation and the final results will be shown in detail. The analysis presented has an important role within the vector-like quark searches currently published which are also discussed in this chapter.

The vector-like quark analysis considers both pair and single production. While single production is model dependent, since the vector-like coupling with the Standard Model quarks directly mediates the production, the pair production is considered model independent because it is mediated via QCD interactions. Nonetheless, in composite-Higgs models a new heavy color octet, heavy gluon, can also appear in the spectrum. In this scenario the pair production mechanism is not model independent. In Chapter 6 a study is presented to evaluate the sensitivity of current vector-like quark searches to this new heavy-gluon as well as the expected sensitivity for early Run-2 data.

---

# Chapter 2

## Theoretical introduction

In this chapter an overview of the theoretical framework used to develop the work presented in this thesis is shown. First, the Standard Model of particle physics will be presented as the best description of particles and their interactions found so far. However, there are still some open questions for which the SM does not provide an answer. Several models beyond the SM have been proposed to answer these questions, often predicting the existence of new particles. In this chapter, the concept of vector-like quarks will be introduced as well as heavy gluons in the context of composite Higgs models.

### 2.1 The Standard Model of particle physics

The Standard Model (SM) reflects our best knowledge of the elementary particles and their interactions. It is a quantum gauge field theory based on the symmetry group  $SU(3)_c \times SU(2)_L \times U(1)_Y$ . Each component of the symmetry group is related to a given interaction mediated by its associated gauge boson. The component  $SU(2)_L \times U(1)_Y$  is known as the electroweak (EW) group and its associated interaction is the EW interaction. The  $SU(3)_c$  component is known as the color group and its associated force is the strong force. Table 2.1 shows a list of the SM particles (fermions, with spin 1/2, and gauge bosons, with spin 1). The fermions with color charge are called quarks. Their left-handed (LH) component is represented by  $q_L = (u_L, d_L)$ , including the heavier replicas  $(c_L, s_L)$  and  $(t_L, b_L)$  and their right-handed (RH) component are represented by  $u_R$  and  $d_R$ , including as well their heavier replicas  $c_R, s_R, t_R$  and  $b_R$ . Fermions without color charge are called leptons and are represented in the same way as quarks with  $L_L = (\nu_L, l_L)$  where  $l = e, \mu, \tau$ , and  $l_R$ . In the SM there are no RH neutrinos  $\nu_R$  since they have not been observed

<i>Field</i>	<i>Spin</i>	$SU(3)_c$	$SU(2)_L$	$U(1)_Y$
$q_L = (u_L, d_L)$	1/2	3	2	1/6
$u_R$	1/2	3	1	2/3
$d_R$	1/2	3	1	-1/3
$L_L = (\nu_L, l_L)$	1/2	1	2	-1/2
$l_R$	1/2	1	1	-1
$g_\mu^a$	1	8	1	0
$W_\mu^I$	1	1	3	0
$B_\mu$	1	1	1	0
$\phi$	0	1	2	1/2

Table 2.1: Constituents of the SM. Particles with spin 1/2 are the matter particles while particles with spin 1 are the gauge bosons of the theory. The  $\nu_R$  is not shown because it is not considered in the SM.

in nature [1]. This makes up the entire set of matter constituents within the SM: 6 quarks, and 6 leptons. There are a total of 12 gauge bosons in the theory: 8 of them are called *gluons*,  $g_\mu^a$ , with  $a = 1, 2, \dots, 8$ , and they are the force carriers of the strong interaction and correspond to the  $SU(3)_c$  group. The remaining four vector bosons,  $W^I$  and  $B_\mu$ , with  $I = 1, 2, 3$ , are related to the EW interaction and correspond to the  $SU(2)_L$  and  $U(1)_Y$  groups respectively.

The interactions between boson and fermions in the SM are defined by the Lagrangian

$$\begin{aligned} \mathcal{L}_{gauge} = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + i\bar{q}_{Li}\not{D}q_{Li} + i\bar{L}_{Li}\not{D}L_{Li} \\ & + i\bar{d}_{Ri}\not{D}d_{Ri} + i\bar{u}_{Ri}\not{D}u_{Ri} + i\bar{l}_{Ri}\not{D}l_{Ri} \end{aligned} \quad (2.1)$$

where  $F_{\mu\nu}F^{\mu\nu} = G_{\mu\nu}^a G_a^{\mu\nu} + W_{\mu\nu}^I W_I^{\mu\nu} + B_{\mu\nu}B^{\mu\nu}$ ,  $\not{D} = \gamma^\mu D_\mu$  and the index  $i$  runs for the three families of leptons and quarks. The tensors  $G_{\mu\nu}^a$ ,  $W_{\mu\nu}^I$  and  $B_{\mu\nu}$  are defined as

$$\begin{aligned} G_{\mu\nu}^a &= \partial_\mu g_\nu^a - \partial_\nu g_\mu^a - g_s f_{bc}^a g_\mu^b g_\nu^c, \\ W_{\mu\nu}^I &= \partial_\mu W_\nu^I - \partial_\nu W_\mu^I - g\epsilon_{jk}^I W_\mu^j W_\nu^k, \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu, \end{aligned}$$

where  $f_{bc}^a$  and  $\epsilon_{kj}^I$  are the group structure constants of  $SU(3)_c$  and  $SU(2)_L$  respectively. The covariant derivative acts on a given field  $\varphi$  as

$$D_\mu \varphi = \left( \partial + ig_s \frac{\lambda_a}{2} g_\mu^a - ig \frac{\sigma_I}{2} W_\mu^I + ig' Y B_\mu \right) \varphi$$

where  $\lambda_a$  are the Gell-Mann matrices and  $\sigma_I$  are the Pauli matrices.

Defining  $\sigma^\pm \equiv \sigma^1 \pm i\sigma^2$ ,  $W^\pm \equiv 1/\sqrt{2} (W_\mu^1 \mp iW_\mu^2)$  and introducing a rotation in the fields  $B_\mu$  and  $W_\mu^3$

$$\begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} = \begin{pmatrix} c_W & s_W \\ -s_W & c_W \end{pmatrix} \begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix},$$

where  $Z_\mu$  is the  $Z$  boson,  $A_\mu$  is the photon and  $s_W$  ( $c_W$ ) is the sine (cosine) of the Weinberg angle ( $\theta_W$ ), the action of the covariant derivative can be redefined as

$$D_\mu \varphi = \left[ \partial_\mu + ig_s \frac{\lambda_a}{2} g_\mu^a - \frac{ig}{2\sqrt{2}} (\sigma^+ W_\mu^+ + \sigma^- W_\mu^-) + \frac{ig}{c_W} (T_3 - s_W^2 Q) Z_\mu - ieQ A_\mu \right] \varphi, \quad (2.2)$$

where  $Q = T_3 + Y$  is the electric charge and  $T_3 = \frac{\sigma_3}{2}$  the third component of the weak isospin.

It is clear from (2.2) that two fermions in the same doublet ( $\psi, \psi'$ ) interact through the  $W$  boson (charged currents) and, analogously, a given fermion  $\psi$  presents neutral interactions via the photon and the  $Z$  boson (neutral currents):

$$\begin{aligned} \mathcal{L}_{CC} &= -\frac{g}{2\sqrt{2}} \left[ \bar{\psi} \gamma^\mu (1 - \gamma^5) \psi' W_\mu^+ + \bar{\psi}' \gamma^\mu (1 - \gamma^5) \psi W_\mu^- \right], \\ \mathcal{L}_{NC} &= -e \bar{\psi} \gamma^\mu \psi A_\mu - \frac{g}{2c_W} \bar{\psi} \gamma^\mu \left( T_3 (1 - \gamma^5) - 2Q s_W^2 \right) \psi Z_\mu. \end{aligned}$$

where  $\gamma^\mu = \{\gamma^0, \gamma^1, \gamma^2, \gamma^3\}$  are the four gamma matrices, or Dirac matrices. They have specific anticommutation relations and generate a matrix representation of the Clifford algebra. The  $\gamma^5$  is defined as  $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ .  $\mathcal{L}_{NC}$  is also defined for  $\psi'$ , although only the  $\psi$  part of the Lagrangian is written for simplicity.

So far, based on the symmetries of the SM, the different particles have been presented and their interactions derived. However, there is a missing term: a mass term for fermions and the massive gauge bosons (the  $W$  and  $Z$  bosons). In order

to accommodate a mass term for the massive particles the  $SU(2)_L \times U(1)_Y$  symmetry of the SM needs to be broken. This is achieved by spontaneously breaking the EW symmetry via the mechanism known as the *Higgs mechanism*. It was independently introduced by Robert Brout and François Englert [2], Peter Higgs [3] and Gerald Guralnik, C. R. Hagen and Tom Kibble [4]. It was incorporated to the EW interactions theory by Weinberg [5] and Salam which earned them the Nobel Prize in 1979, together with Sheldon Glashow, who proposed the EW interactions formalism in 1961 [6]. A new colorless scalar field  $\phi = (\phi^+, \phi^0)$ , which is a doublet under  $SU(2)_L$ , is introduced, the Higgs field. The development of a vacuum expectation value (VEV) by the Higgs field is responsible for the EW symmetry breaking (EWSB). The Higgs sector in the SM is represented by the Lagrangian

$$\begin{aligned} \mathcal{L}_{Higgs} = & (D_\mu \phi)^\dagger (D^\mu \phi) - \mu_h^2 \phi^\dagger \phi - \lambda_h (\phi^\dagger \phi)^2 \\ & - \left[ y_{ij}^u \overline{q_{Li}} \tilde{\phi} u_{Rj} + y_{ij}^d \overline{q_{Li}} \phi d_{Rj} + y_{ij}^l \overline{L_{Li}} \phi l_{Rj} + \text{h.c.} \right], \end{aligned} \quad (2.3)$$

where  $\tilde{\phi} = i\sigma_2 \phi^*$ , being  $\sigma_2$  the second Pauli matrix, has hypercharge  $Y = -1/2$ . The Higgs field can be written in the unitary gauge as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ h + v \end{pmatrix} \quad (2.4)$$

where  $v = \sqrt{-\mu^2/\lambda}$ . The EW symmetry  $SU(2)_L \times U(1)_Y$  is spontaneously broken<sup>1</sup> to  $U(1)_Q$  once  $\phi$  takes a given VEV,  $v$ . After EWSB the first term of (2.3) can be written as

$$(D_\mu \phi)^\dagger (D^\mu \phi) = \frac{1}{2} \partial_\mu h \partial^\mu h + \frac{g^2}{4} \left( v^2 + 2vh + h^2 \right) \left( W_\mu^+ W^{\mu-} + \frac{1}{2c_W^2} Z_\mu Z^\mu \right). \quad (2.5)$$

Now the  $W$  and the  $Z$  bosons have acquired mass  $m_W = gv/2$  and  $m_Z = m_W/c_W$ . In the same way, the terms describing the fermionic interactions with the Higgs field

---

<sup>1</sup>The remaining  $U(1)_Q$  symmetry is the electromagnetic group generated by  $T_3 + Y$  which was previously defined as the electric charge.



(the Yukawa sector) can be rewritten as

$$-\frac{v}{\sqrt{2}}\left(1 + \frac{h}{v}\right)\left(y_{ij}^u \overline{u_L^i} u_R^j + y_{ij}^d \overline{d_L^i} d_R^j + y_{ij}^l \overline{l_L^i} l_R^j + \text{h.c.}\right). \quad (2.6)$$

Until now everything has been shown in the interaction eigenstates. The matrices  $y_{ij}$  can be diagonalised using two unitary matrices  $\mathcal{U}_L^{u,d,l}$  and  $\mathcal{U}_R^{u,d,l}$  satisfying  $(\mathcal{U}_L^u)^\dagger y^u \mathcal{U}_R^u = \text{diag}(y_u, y_c, y_t)$ ,  $(\mathcal{U}_L^d)^\dagger y^d \mathcal{U}_R^d = \text{diag}(y_d, y_s, y_b)$  and  $(\mathcal{U}_L^l)^\dagger y^l \mathcal{U}_R^l = \text{diag}(y_e, y_\mu, y_\tau)$ . This transformation defines the mass eigenstates, which are linear combinations of the interaction eigenstates, in which a fermion  $\psi$  has acquired a mass  $m_\psi = v y_\psi / \sqrt{2}$ . In the mass eigenstates basis, flavour-changing charged currents are introduced via the Cabbibo-Kobayashi-Maskawa (CKM) matrix, defined as  $V = (\mathcal{U}_L^u)^\dagger \mathcal{U}_L^d$ . However, when considering the neutral currents in the mass eigenstates, the diagonalisation of the fermion fields leads to  $(\mathcal{U}_L^u)^\dagger \mathcal{U}_L^u = (\mathcal{U}_L^d)^\dagger \mathcal{U}_L^d = 1$ , which implies that no flavour-changing neutral currents (FCNCs) are present in the SM at tree-level.

The Higgs mechanism has been known since 1968 but the Higgs boson, the only elementary scalar particle in the SM, has been missing since then. The last piece to the SM puzzle was found in 2012, when both ATLAS and CMS experiments reported the observation of a scalar boson with properties consistent with those predicted for the SM Higgs Boson [7, 8]. After this observation François Englert and Peter Higgs were awarded with the Nobel Prize in 2013.

## 2.2 Beyond the Standard Model

The SM has been very well tested in the past four decades. The interactions and observables predicted by the SM have been measured in agreement with the theory with an unprecedented level of accuracy, most notably by LEP and SLAC [9, 10]. A global fit to different observables of the EW sector has been performed in [11–13], providing a high precision measurements of different EW observables of the SM. Based on the symmetries introduced in the SM, several particles were predicted before they were discovered. That is the case for the light quarks [14, 15], discovered in 1968 at SLAC, the tau [16], discovered by the SLAC-LBL group in 1975, the bottom quark [17], discovered by the Fermilab E288 experiment (at Tevatron) in 1977, the gluon [18, 19], discovered in 1978 at PETRA (DESY), the EW gauge bosons [20, 21], discovered in 1983 by the UA1 and UA2 collaborations (at CERN), the top quark [22, 23], discovered in 1995 by the CDF and D0 collaborations (at Tevatron, Fermilab), the tau neutrino [24], discovered in 2000 at the DONUT Collab-

oration and finally the Higgs boson,  $\sim 50$  years after being proposed, was discovered in 2012 by the ATLAS and CMS collaborations [7, 8] (at CERN).

As has been stated before, the SM has proven to be a reliable theory during the course of the past 40 years. However, the nature of the symmetry-breaking mechanism is unknown and there are also some open questions that the SM cannot answer, for example, the number of fermion generations or the mass hierarchy are not predicted by the SM. The SM does not explain some cosmological questions such as the matter anti-matter asymmetry nor provides a viable dark matter candidate. Since no RH neutrinos are present in the SM they do not get mass via the Higgs mechanism but they are known to have mass from neutrino oscillations observation in solar neutrino experiments [25]. Another issue with the SM is the naturalness problem, which arises when trying to access energies higher than the electroweak energy scale due to quadratic divergences in the scalar fields interactions. Thus, even though the Higgs Boson has been observed and the puzzle of the Higgs mechanism has been completed, the SM cannot be considered a complete theory.

In this chapter vector-like quarks will be presented as new particle candidates which arise in models aiming to solve the naturalness problem. Composite Higgs models (CHMs) feature vector-like quarks, which naturally appear due to the partial compositeness mechanism.

### 2.2.1 Vector-like quarks

The SM does not predict the number of lepton families, therefore the only reason why only three quark families are considered is that there is no experimental signature of new ones. The ALEPH, DELPHI, L3 and OPAL collaborations found data to be consistent with three families of light neutrinos [10] (and therefore leptons), which is shown in Figure 2.1. However, this reasoning is only valid if the mass of the new quarks is expected to be in the mass range observed so far. New quark families can be heavier than the top quark, in particular with masses over the TeV scale and therefore not have been observed due to the limit energy reach of the experiments. New quarks with the same symmetries as the ones already existent in the SM, i.e. chiral quarks with different charges under the  $SU(2)_L \times U(1)_Y$  group for the LH and RH components, obtain mass through the Higgs mechanism which implies that the mass of these heavy quark is proportional to the EWSB scale, and therefore they are expected to be relatively light,  $m_{chiral} = \lambda v \lesssim v$ . The only way they can be made heavy is by increasing the coupling  $\lambda$ , introducing, at the same time,

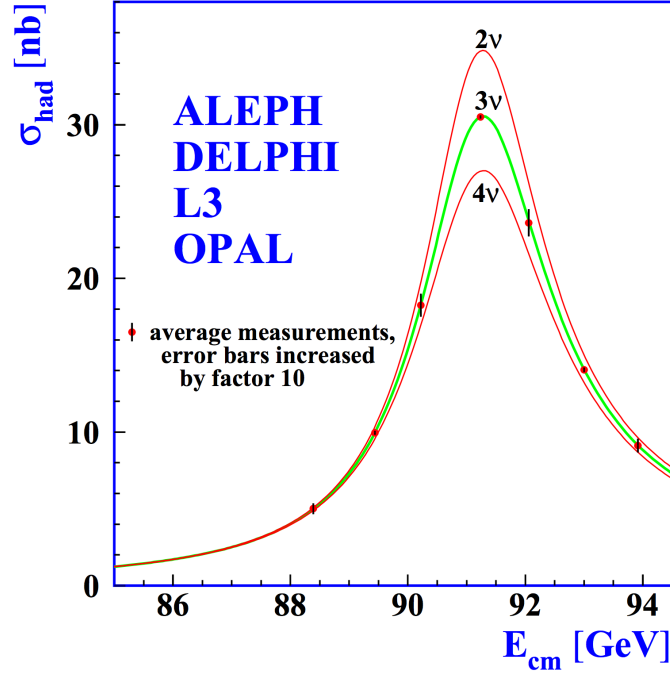


Figure 2.1: Measurements of the hadron production cross-section around the Z resonance. The curves indicate the predicted cross-section for two, three and four neutrino species with SM couplings and negligible mass [10].

large one-loop corrections to the Higgs cross-section predictions, which show a good agreement with the SM predictions [26] (see Figure 2.2). A new family of chiral quarks are very constrained [27] after the recent discovery of a light Higgs boson and the measurement of its properties [26] as well as from direct searches at the LHC [28–30]. Furthermore, the effects introduced in electroweak observables by the corrections induced by the presence of such new quarks imply strong constraints from electroweak precision tests as well [31].

On the other hand, vector-like quarks have the same transformation rules under the  $SU(2)_L \times U(1)_Y$  group for both LH and RH components. In this case a bare mass term, which does not violate gauge invariance, can be introduced avoiding the need of introducing large dimensionless couplings

$$\mathcal{L}_{bare} = \overline{Q}_L M^0 Q_R + h.c. \quad (2.7)$$

New vector-like quarks at the TeV scale are well motivated by different BSM models. They appear in theories for which the Higgs boson is a pseudo Nambu-Goldstone boson (pNGB), with a mass naturally in the EW scale [32–34]. They are also found in composite theories in which the mass term for the fermions appear

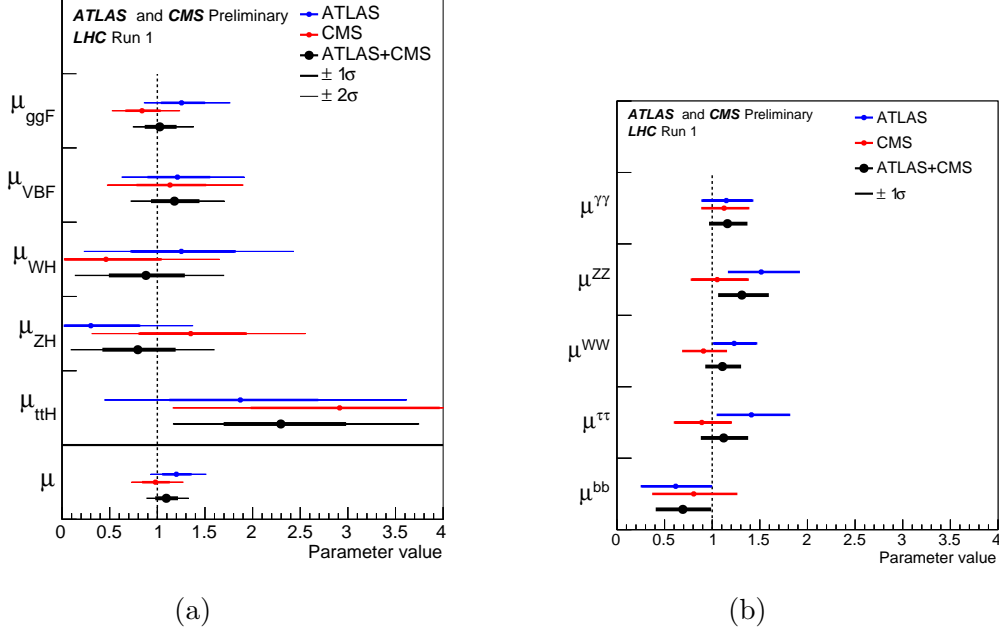


Figure 2.2: The best fit results for the production (a) and decay (b) signal strength for the Higgs boson are shown for the combination of the ATLAS and CMS in Run-1 [26].

through the mechanism of *partial compositeness* [35, 36]. In this mechanism the elementary sector mix linearly with the composite sector, producing a tower of massive vector-like fermionic resonances which mix with the elementary fermions, producing states which are a mixture of elementary and composite states. Thus, vector-like quarks are predicted in models implementing these ideas such as composite Higgs or little Higgs models. They appear as well in grand unification and string theories based on the  $E_6$  symmetry group [37].

In order for vector-like quarks to mix with the SM quarks [38, 39] and gauge bosons, some constraints appear on their possible quantum numbers, in order to preserve the gauge invariance. Therefore, if one wants to produce Yukawa terms, without changing the scalar sector, only seven possibilities [40] are allowed, which are summarised in Table 2.2. The vector-like  $T$  and  $B$  have electric charge  $2/3$  and  $-1/3$ . There are also vector-like quarks with exotic electric charge like the  $X$  and  $Y$  quarks with  $5/3$  and  $-4/3$  respectively.

The mixing with the SM quarks is obtained after EWSB via the mass mixing. In the weak eigenstate basis (denoted by a 0 superscript), the third generation and

<b>Q</b>	$T$	$B$	$\begin{pmatrix} T \\ B \end{pmatrix}$	$\begin{pmatrix} X \\ T \end{pmatrix}$	$\begin{pmatrix} B \\ Y \end{pmatrix}$	$\begin{pmatrix} X \\ T \\ B \end{pmatrix}$	$\begin{pmatrix} T \\ B \\ Y \end{pmatrix}$
<b>Isospin</b>	0	0	1/2	1/2	1/2	1	1
<b>Hypercharge</b>	2/3	-1/3	1/6	7/6	-5/6	2/3	-1/3

Table 2.2: Possible vector-like multiplets allowed to mix with the SM quarks through Yukawa couplings.

heavy quark mass terms are

$$\begin{aligned} \mathcal{L}_{mass} = & - \begin{pmatrix} \bar{t}_L^0 & \bar{T}_L^0 \end{pmatrix} \begin{pmatrix} m_{33}^u & m_{34}^u \\ m_{43}^u & M^0 \end{pmatrix} \begin{pmatrix} t_R^0 \\ T_R^0 \end{pmatrix} \\ & - \begin{pmatrix} \bar{b}_L^0 & \bar{B}_L^0 \end{pmatrix} \begin{pmatrix} m_{33}^d & m_{34}^d \\ m_{43}^d & M^0 \end{pmatrix} \begin{pmatrix} b_R^0 \\ B_R^0 \end{pmatrix} + h.c. \end{aligned} \quad (2.8)$$

where  $m_{ij}^{u,d}$ , are the mass mixings,  $v = 246$  GeV is the Higgs VEV and  $M^0$  is the bare mass term. The mass matrix can be the diagonalised by the unitary transformation

$$\mathcal{U}_L^q M^q (\mathcal{U}_R^q)^\dagger = M_{diag}^q \quad (2.9)$$

where  $\mathcal{U}_{L,R}^q$  are the mixing matrices and  $M_{diag}^q$  is the mass diagonal matrix. The relation between the vector-like quarks and the SM quarks using the mixing matrices can be written as

$$\begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix} = \mathcal{U}_{L,R}^u \begin{pmatrix} t_{L,R}^0 \\ T_{L,R}^0 \end{pmatrix} = \begin{pmatrix} c_{L,R}^u & -s_{L,R}^u e^{i\phi_u} \\ s_{L,R}^u e^{i\phi_u} & c_{L,R}^u \end{pmatrix} \begin{pmatrix} t_{L,R}^0 \\ T_{L,R}^0 \end{pmatrix}, \quad (2.10)$$

$$\begin{pmatrix} b_{L,R} \\ B_{L,R} \end{pmatrix} = \mathcal{U}_{L,R}^d \begin{pmatrix} b_{L,R}^0 \\ B_{L,R}^0 \end{pmatrix} = \begin{pmatrix} c_{L,R}^d & -s_{L,R}^d e^{i\phi_d} \\ s_{L,R}^d e^{i\phi_d} & c_{L,R}^d \end{pmatrix} \begin{pmatrix} b_{L,R}^0 \\ B_{L,R}^0 \end{pmatrix}, \quad (2.11)$$

where  $s_{L,R}^{u,d} \equiv \sin \theta_{L,R}^{u,d}$  and  $c_{L,R}^{u,d} \equiv \cos \theta_{L,R}^{u,d}$ , being  $\theta_{L,R}$  the mixing angle for the LH and RH components respectively and  $\phi_{u,d}$  are complex phases.

After diagonalising the mass matrix the following relation can be found (see [41]):

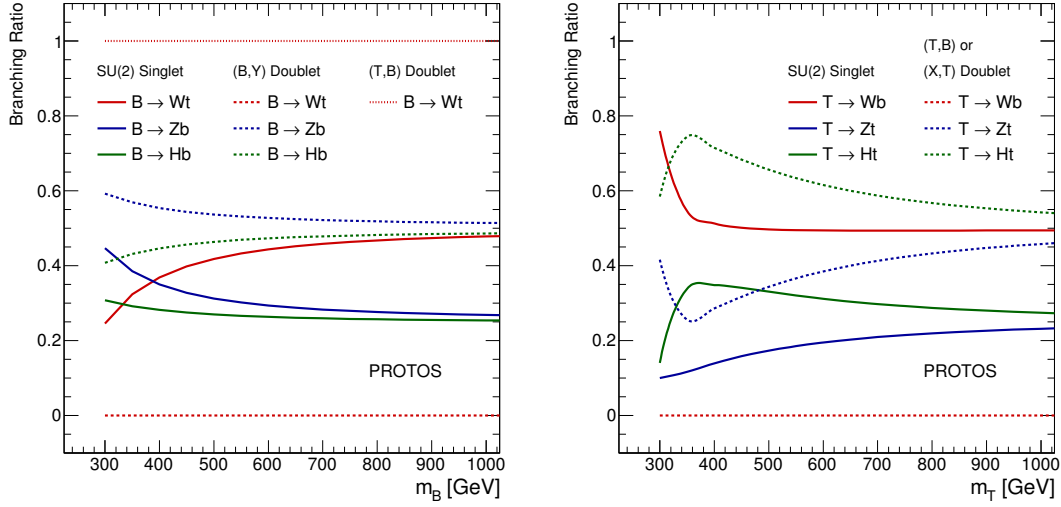


Figure 2.3: Branching ratios for all decaymodes of the vector-like  $T$  and  $B$  quarks as a function of the vector-like quark mass.

$$\tan \theta_R^q = \frac{m_q}{m_Q} \tan \theta_L^q, \quad \text{for singlets and triplets,} \quad (2.12)$$

$$\tan \theta_L^q = \frac{m_q}{m_Q} \tan \theta_R^q, \quad \text{for doublets,} \quad (2.13)$$

with  $(q, m_q, m_Q) = (u, m_t, m_T), (d, m_b, m_B)$ . Note that the R and L mixings are not independent and one of the chiralities is enhanced. For simplicity, and because it is a common occurrence in many natural models of physics beyond the SM, it is assumed that the mixing occurs mainly with the third generation.

The mixing with the SM quarks also modifies the coupling of the SM quarks with the  $Z$ ,  $W$  and  $H$  bosons. In fact, in the presence of vector-like quarks, the Glashow-Iliopoulos-Maiani mechanism [42] is broken and FCNC at tree-level are allowed [43]. Vector-like quarks decay to the SM quarks and one of the gauge bosons and the coupling is defined by the matrix  $V_{qQ}$ <sup>2</sup>, which is the extension of the Cabibo-Kowayashi-Maskawa matrix including the new vector-like quarks and defines the charge current mixings, the matrix  $X_{Qq}$  which defines the neutral current mixings and the Yukawa couplings  $Y_{Qq}$  (see [41] for a complete list of values of these couplings for different models). As shown before the mixing depends on the mass of the vector-like quarks and thus the branching-ratio for each decay channel. The

<sup>2</sup>The subindex  $Q$  and  $q$  represent a given vector-like quark and the SM quark it is mixing with.

multiplet each vector-like quark belongs to also has an impact on the branching ratio to each decay channel, for instance, considering the vector-like quark  $B$  in a singlet the possible decays are

$$B \rightarrow tZ, \quad B \rightarrow tH, \quad B \rightarrow tW^-, \quad (2.14)$$

but if the vector-like  $B$  is considered in the  $(TB)$  doublet, in a scenario where  $V_{Tb} \ll V_{Bt}$ , i.e. the top quark couples stronger to its partner than the bottom quark, the only allowed decay is through charged currents  $B \rightarrow tW^-$  [44]. Figure 2.3 shows the branching ratios for each decay channel as a function of the vector-like quark mass for some of the multiplets presented.

There are two main ways in which vector-like quarks can be produced.

- **Pair production:** vector-like quarks are produced in pairs via QCD interaction, as depicted in Figure 2.4(a). Since the coupling of vector-like quarks with gluons is the same as SM quarks this production mechanism is largely model independent.
- **Single production:** the single production mechanism of vector-like quarks is sensitive to the EW couplings between the vector-like quarks and the SM quarks. It is not mediated via QCD as it is the case in pair production but they are produced via charged or neutral currents, as can be seen in Figures 2.4(b) and (c) where the two processes considered in Chapter 5 are shown.

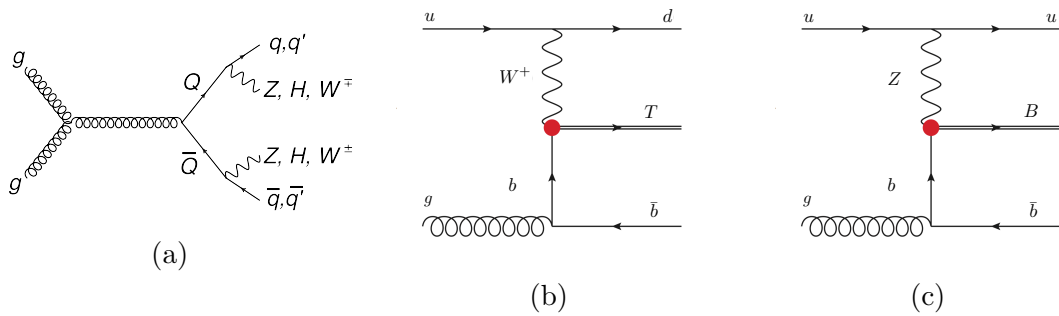


Figure 2.4: Pair production (a) and single production processes of vector-like  $T$  (b) and  $B$  (c).

Both production mechanisms are of interest depending on the vector-like quark masses which are still allowed. Direct searches performed by the ATLAS [28, 45–48]

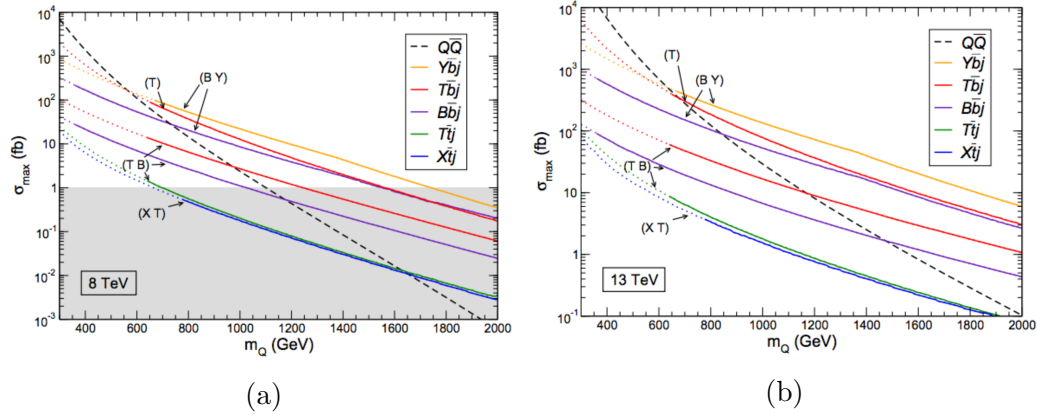


Figure 2.5: Pair production (a) and single production processes of vector-like  $T$  (c) and  $B$  (c) [41].

and CMS [49–51] collaborations<sup>3</sup>, have set lower bounds for the vector-like quark masses at  $m_Q \sim 750$  GeV for all possible branching-ratios, reaching masses near  $m_Q = 1$  TeV in certain phase space regions. The pair production mechanism is the dominant one when considering relatively light vector-like quarks, as can be seen in Figure 2.5 for 8 TeV (a) and 13 TeV (b). This is why pair production has been the main channel considered so far in direct searches. However, as higher masses are excluded, the single production mechanism might become dominant and therefore dedicated searches are going to be of interest. The second operation phase of the LHC at  $\sqrt{s} = 13$  TeV will probably allow the proven region to increase up to 1 TeV, which implies that single production will likely begin to be dominant. Nonetheless, the gain in pair production cross-section from the increase of the center-of-mass energy will probably drive the searches at the beginning of Run-2.

### 2.2.2 Composite Higgs models

Composite Higgs models (CHM) were first introduced by Georgi and Kaplan [52, 53] as a solution for the hierarchy problem. In these models the Higgs boson is assumed to be a bound state of a new strong sector instead of a elementary scalar. The fact that the Higgs is a bound state protects its mass, being unaffected by radiative corrections above a certain compositeness scale. However no new physics has been observed so far up to the TeV scale and the Higgs mass has been measured to be  $m_H \sim 125$  GeV. In order to be in the EW scale, the Higgs is considered a pNGB of a new strongly interacting sector which is generated via the spontaneous breaking

<sup>3</sup>The analysis presented in this thesis in Chapter 5 presents the most stringent mass limit up to date on vector-like quarks for the  $Z$  decay mode.



of a global symmetry  $\mathcal{G} \rightarrow \mathcal{H}$ , making the Higgs boson naturally light while keeping the new physics effects in the TeV scale. In order to be able to accommodate the current description of the SM, the symmetry group  $SU(3)_c \times SU(2)_L \times SU(2)_R$ , where  $SU(2)_R$  is included to protect the custodial symmetry<sup>4</sup>, must be contained in  $\mathcal{G}$  and  $\mathcal{H}$  and the NGB of  $\mathcal{G}/\mathcal{H}$  must contain a colorless bidoublet of the SM group, which corresponds to the Higgs degrees of freedom.

Several groups satisfy these requirements. In the minimal CHM (MCHM) [54] in which  $SO(5)$  is spontaneously broken to  $SO(4)$  is of interest in the context of this thesis as will be used in Chapter 6. In this case  $SO(5)$  has 10 generators which is broken to  $SO(4)$  with 6, thus 4 generators are broken. Four NGB appear which are identified with the four components of the Higgs boson. In the context of MCHM, the coupling between the Higgs and the gauge bosons is not linear. The scalar sector Lagrangian reads

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu h \partial^\mu h + \frac{g^2}{4} \sin^2 \left( \frac{h}{f} \right) W_\mu^+ W^{\mu-} + \frac{g^2}{8c_W^2} f^2 \sin^2 \left( \frac{h}{f} \right) Z_\mu Z^\mu \quad (2.15)$$

The previous equation can be expanded to second order in  $h$  which, after symmetry breaking, reads

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu h \partial^\mu h + \frac{g^2}{4} \left[ v^2 + 2v\sqrt{1-\xi}h + (1-2\xi)h^2 \right] \left( W_\mu^+ W^{\mu-} + \frac{1}{2c_W^2} Z_\mu Z^\mu \right) \quad (2.16)$$

where

$$v = f \sin \left( \frac{\langle h \rangle}{f} \right), \quad \text{and } \xi = \frac{v^2}{f^2}. \quad (2.17)$$

Here  $v \sim 246$  GeV fixes the EW scale,  $f$  is the composite scale and thus  $\xi$  is called the degree of compositeness. Note that  $\langle h \rangle$ , which is the Higgs VEV, in MCHM is not the same as  $v$  but when the composite scale is large enough  $\xi \rightarrow 0$  recovering the SM couplings between the Higgs and the gauge boson, i.e. the composite sector decouples.

The global symmetry  $G$  is explicitly broken by the coupling of elementary fermions to composite operators  $\mathcal{O}(x)$ . When the coupling between the composite and ele-

---

<sup>4</sup>The custodial symmetry ensures that the  $\rho$  parameter, defined as the ratio involving the masses of the weak bosons and the Weinberg angle, is  $\sim 1$ .

mentary sector is linear, a tower of fermionic resonances is excited by the composite operator. The linear coupling introduces a mass mixing between the composite states and the elementary fermions which implies that the physical states are a admixture of elementary and composite states. This mechanism to obtain mass via linear couplings is called *partial compositeness*.

Lets focus on a particular version of MCHM [55, 56], the MCHM<sub>45</sub>, in which the RH top quark is a fully composite singlet, the LH SM quarks are embedded in a 5 representation of SO(5) and the vector-like quarks span a 4 representation of SO(4)<sup>5</sup>:

$$Q_L^5 = \frac{1}{\sqrt{2}} \begin{pmatrix} ib_L \\ b_L \\ it_L \\ -t_L \\ 0 \end{pmatrix}, \quad \Psi = \frac{1}{\sqrt{2}} \begin{pmatrix} i(B - X_{5/3}) \\ B + X_{5/3} \\ i(T + X_{2/3}) \\ -T + X_{2/3} \end{pmatrix}. \quad (2.18)$$

The top partners, in terms of  $SU(2)_L \times U(1)_Y$  representations, give rise to two doublets  $(T, B)$  and  $(X_{5/3}, X_{2/3})$  with hypercharges 1/6 and 7/6 respectively. The vector-like  $T$  and  $X_{2/3}$  have electric charge 2/3, the  $B$  quark has electric charge -1/3 and the quark  $X_{5/3}$  has an exotic electric charge of 5/3. After EWSB, a linear combination of  $T$  and  $X_{2/3}$ , that we denote  $X'_{2/3}$  remains degenerate with  $X_{5/3}$ . The orthogonal combination, that we call  $T'$ , and  $B$  are somewhat heavier with a small mass splitting. In most of the parameter space their decay BR are:

$$BR(X_{5/3} \rightarrow tW^+) = BR(B \rightarrow tW^-) = 1, \quad (2.19)$$

$$BR(X'_{2/3} \rightarrow tZ) \approx BR(X'_{2/3} \rightarrow tH) \approx \frac{1}{2}, \quad (2.20)$$

$$BR(T' \rightarrow tZ) \approx BR(T' \rightarrow tH) \approx \frac{1}{2}. \quad (2.21)$$

In general, vector resonances also appear in composite models. Due to partial compositeness and that the elementary quarks have charge under  $SU(3)_c$ , the vector-like quarks will also have color charge in order to preserve the color symmetry. In this scenario, in order to preserve the color symmetry in the composite sector, it is natural that heavy vector resonances appear as a spin-1 color octet [57], called heavy gluon, which can mix with the SM gluon.

---

<sup>5</sup>This is the model used for the analysis presented later in Chapter 6 with the addition of a heavy color octet.

The Lagrangian which describes the MCHM4<sub>5</sub> including the heavy gluon, in the elementary-composite basis, reads

$$\begin{aligned}\mathcal{L} = & \bar{q}_L i \not{D} q_L + \bar{t}_R i \not{D} t_R + \bar{\Psi} i (\not{D} + i \not{e}) \Psi - M_\Psi \bar{\Psi} \Psi \\ & + \left[ i c_1 (\bar{\Psi}_R)_i \gamma^\mu d_\mu^i t_R + y f (\bar{Q}_L^5)^I U_{Ii} \Psi_R^i + y c_2 f (\bar{Q}_L^5)^I U_{I5} t_R + \text{h.c.} \right] \\ & - \frac{1}{2} \text{Tr}[G_{\mu\nu}^e]^2 - \frac{1}{2} \text{Tr}[G_{\mu\nu}^c]^2 + \frac{1}{2} M_c^2 \left( G_\mu^c - \frac{g_e}{g_c} G_\mu^e \right)^2, \quad (2.22)\end{aligned}$$

where  $y$ ,  $c_1$  and  $c_2$  are three dimensionless couplings,  $f$  is the composite scale,  $M_c$  is the composite gluon mass,  $G_\mu^{e,c}$  are the elementary and composite gluon, respectively, and  $g_{e,c}$  are the elementary and composite gluon couplings.

The first line include the gauge interaction for the SM and vector-like quarks and the bare mass term for vector-like quarks. The covariant derivative acts over each field as follows

$$\begin{aligned}i D_\mu q_L &= \left( i \partial_\mu + g \frac{\sigma^i}{2} W_\mu^i + \frac{g'}{6} B_\mu + g_e G_\mu^e \right) q_L, \\ i D_\mu t_R &= \left( i \partial_\mu + \frac{2g'}{3} B_\mu + g_c G_\mu^c \right) t_R, \\ i D_\mu \Psi &= \left( i \partial_\mu + \frac{2g'}{3} B_\mu + g_c G_\mu^c \right) \Psi. \quad (2.23)\end{aligned}$$

The remaining terms read

$$\begin{aligned}i \bar{\Psi}_R^i \not{D}_i t_R &= \frac{g}{\sqrt{2}} s_h [(\bar{X}_{5/3})_R \mathcal{W}^+ - \bar{B}_R \mathcal{W}^-] t_R \\ &\quad - \frac{g}{2c_W} s_h [\bar{T}_R + (\bar{X}_{2/3})_R] Z t_R + i [(\bar{X}_{2/3})_R - \bar{T}_R] \frac{\not{\partial} \rho}{f} t_R, \quad (2.24)\end{aligned}$$

$$\begin{aligned}\bar{\Psi} \left( \frac{2g'}{3} \not{B} - \not{e} \right) \Psi &= \frac{g}{c_W} \left( -\frac{1}{2} + \frac{s_W^2}{3} \right) \bar{B} Z B + \frac{g}{c_W} \left( \frac{1}{2} - \frac{5s_W^2}{3} \right) \bar{X}_{5/3} Z X_{5/3} \\ &\quad + \frac{g}{c_W} \left( \frac{1}{2} c_h - \frac{2s_W^2}{3} \right) \bar{T} Z T + \frac{g}{c_W} \left( -\frac{1}{2} c_h - \frac{2s_W^2}{3} \right) \bar{X}_{2/3} Z X_{2/3} \\ &\quad + \frac{g}{\sqrt{2}} \left\{ \bar{B} \mathcal{W}^- [c_{h/2}^2 T + s_{h/2}^2 X_{2/3}] + \bar{X}_{5/3} \mathcal{W}^+ [s_{h/2}^2 T + c_{h/2}^2 X_{2/3}] + \text{h.c.} \right\} \\ &\quad + \text{photon couplings}, \quad (2.25)\end{aligned}$$

$$(\bar{Q}_L^5)^I U_{Ii} \Psi_R^i = \bar{b}_L B_R + \bar{t}_L [c_{h/2}^2 T_R + s_{h/2}^2 (X_{2/3})_R], \quad (2.26)$$

$$(\bar{Q}_L^5)^I U_{I5} t_R = -\frac{1}{\sqrt{2}} s_h \bar{t}_L t_R, \quad (2.27)$$

where

$$s_x \equiv \sin \frac{x}{f}, \quad c_x \equiv \cos \frac{x}{f}, \quad (2.28)$$

except for  $s_W$  and  $c_W$ , which are the sine and cosine of the Weinberg angle.  $\rho$  is the physical Higgs boson and  $h$  reads, in the unitary gauge

$$h \equiv \langle h \rangle + \rho, \quad (2.29)$$

with

$$f s_{\langle h \rangle} = v \approx 246 \text{ GeV}. \quad (2.30)$$

The  $d_\mu$  and  $e_\mu$  symbols are required by the Callan-Coleman-Wess-Zumino construction [58, 59].

As can be seen the heavy gluon, i.e. the composite color octet, only couples to composite quarks ( $t_R$  and  $\Psi$ ) while the elementary gluon only couples with elementary quarks. The physical states of the elementary and composite gluons are defined by the rotation

$$\begin{pmatrix} G_\mu^e \\ G_\mu^c \end{pmatrix} = \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 \\ \sin \theta_3 & \cos \theta_3 \end{pmatrix} \begin{pmatrix} g_\mu \\ G_\mu \end{pmatrix}, \quad (2.31)$$

where the mixing is fixed by the ratio of elementary to composite couplings  $\tan \theta_3 = g_e/g_c$ . After the rotation a massless color octet, the SM gluon  $g_\mu$ , and the heavy gluon  $G_\mu$  are produced with mass

$$M_G = \frac{M_c}{\cos \theta_3}. \quad (2.32)$$

Now, the SM gluon couples universally with coupling strength  $g_s = g_e \cos \theta_3 = g_c \sin \theta_3$ . The heavy gluon now couples to elementary and composite fields with coupling strength given by

$$G \bar{\psi}_{\text{elem}} \psi_{\text{elem}} : -\frac{g_s^2}{\sqrt{g_c^2 - g_s^2}}, \quad G \bar{\psi}_{\text{comp}} \psi_{\text{comp}} : \sqrt{g_c^2 - g_s^2}. \quad (2.33)$$

The setup presented in this section based on the MCHM4<sub>5</sub> with the addition of the heavy gluon via de partial compositeness mechanism will be used in the analysis presented in Chapter 6 in which pair production of vector-like quarks will be studied when, in addition to QCD, a heavy gluon is considered. The pair production cross-

---

section is increased when the heavy gluon channel is open. There is a balance between the mass of the heavy gluon and the mass of the top partners when assessing the impact of the heavy gluon channel in the production rate of vector-like quarks. As the mass of the heavy gluon increases the QCD process becomes dominant due to the rapidly decay of the PDFs. As the mass of the vector-like quark increases the heavy gluon channel becomes dominant since QCD only interactions are not able to produce a pair of heavy vector-like quarks. The impact of a heavy gluon in the kinematics of the vector-like quarks and their decays will be discussed in detail in Chapter 6.



---

## Chapter 3

# Experimental setup

In this chapter a review of the experimental setup used to develop the work presented in this thesis is presented. The European Organization for Nuclear Research (CERN), the Large Hadron Collider, the ATLAS detector and the Worldwide LHC Grid are described in the following sections.

### 3.1 CERN

The CERN laboratory, which is based at the border between Switzerland and France, was founded in September 29<sup>th</sup> of 1954 after the ratification by the 12 founding states: Belgium, Denmark, France, the Federal Republic of Germany, Greece, Italy, the Netherlands, Norway, Sweden, Switzerland, the United Kingdom and Yugoslavia. The idea was to build an European laboratory for research on the atomic nuclei and bring scientists together from all over the world. It was soon applied to a more wide range of high-energy physics experiments mainly focused in the study of the interaction of subatomic particles. In order to achieve this goal CERN has built and run several accelerators and detectors which have allowed to improve our knowledge about particle physics. Figure 3.1 shows the current status of the intricate web of accelerators and detectors working at CERN. These accelerators work together as a source of particles for different detectors such as n-Tof, which study the neutron-nucleus interactions with energy ranging from a few MeV to several GeV and measures this energy from the neutrons time of flight, and ISOLDE, a facility placed at the Proton Synchrotron booster dedicated to the production of a large variety of radioactive ion beams for many different experiments in the fields of nuclear and atomic physics, solid-state physics, materials science and life sciences. The high-energy particles produced by CERN accelerators are used even hundreds of

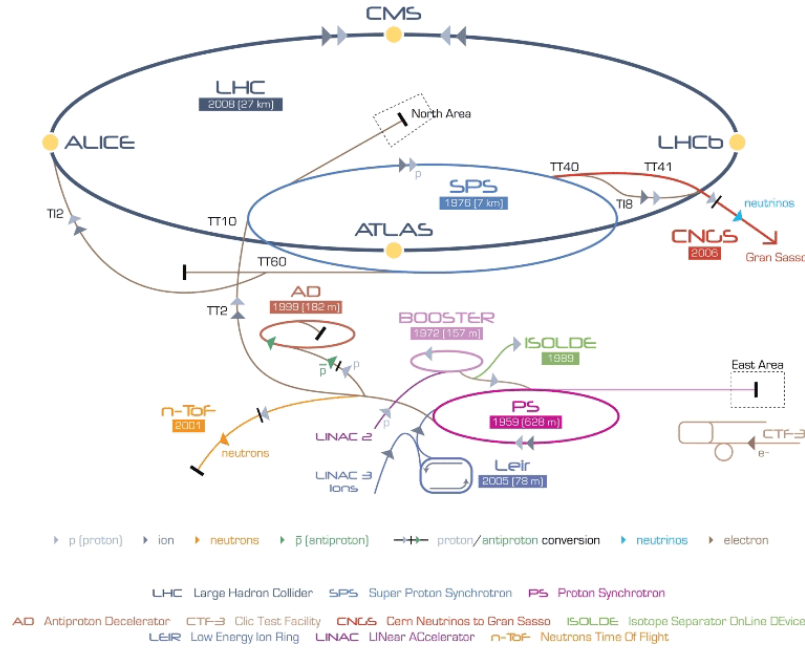


Figure 3.1: Representation of the accelerators, detectors and facilities currently working at CERN [60].

kilometres away when beams of neutrinos produced from protons extracted from the Super Proton Synchrotron (SPS) are sent to Gran Sasso, in Italy, to study neutrinos properties as part of the CERN Neutrinos to Gran Sasso (CNGS) project.

Since its foundation, CERN has witnessed great achievements in particles physics and major discoveries: the discovery of the neutral currents with the Gargamelle bubble chamber in 1973, the discovery of the  $W^\pm$  and  $Z$  bosons with the UA1 and UA2 experiments in 1983, the determination of the number of light neutrino families at the Large ElectronPositron Collider (LEP) in 1989 [10], the first creation of anti-hydrogen in 1995 [61], the discovery of direct CP violation in 1999 [62], the isolation of anti-hydrogen in 2010 [63], the stabilisation of anti-hydrogen for over 15 minutes in 2011 [64] and the latest discovery of the Higgs boson with a mass of 125 GeV which was observed both by the CMS and ATLAS Collaboration in 2012 [7, 8] which was the last missing piece of the SM.

Not only particle physics finds at CERN an important play to evolve but engineering and computer science are also important pieces of the science ecosystem at CERN. The main example is ENQUIRE, a project by Tim Berners-Lee at CERN aiming to provide a new information management system which allows to reference content in a network by a link that could be embedded in readable text. It was proposed as an answer of a need by the scientific community to easily share information. It quickly evolved and became what we know today as the World Wide Web, be-



coming a pillar of modern communications and opening business opportunities for some of the more important companies in the world.

CERN has come a long journey since it was born more than 60 years ago. Today it is not only an European research facility but it has become an international laboratory with 22 member states, observer states for which their membership is not possible or not yet feasible like UNESCO, USA, the European Union or Japan among others and non-member states with co-operation agreements with CERN like Argentina, Korea or South Africa among others. Around 10.000 scientists from over 600 institutions and 113 nationalities are visiting CERN regularly for their research.

## 3.2 The Large Hadron Collider

The Large Hadron Collider [65] (LHC), located at CERN, is the largest and highest-energy accelerator ever built. It lies in a 27 kilometres circular tunnel, built for LEP, at a depth ranging from 50 up to 175 meters underground. It accelerates two proton beams in opposite directions and uses older accelerators like the PS and SPS as preaccelerators to boost the protons to the desired energy. The first proton beams circulated in the main ring for the first time in September of 2008 and the first collisions at 3.5 TeV took place on March 30<sup>th</sup>, 2010. Since then, the LHC have been performing really well running at  $\sqrt{s} = 7$  TeV for the rest of 2010 and 2011 and upgrading the energy to  $\sqrt{s} = 8$  TeV until the beginning of 2013. The first long shutdown took place in February 14<sup>th</sup> of 2013. This long shutdown had the goal of upgrading the LHC to be able to provide higher energy than before and start producing proton-proton collisions at  $\sqrt{s} = 13$  TeV. It ended in April 5<sup>th</sup> of 2015, when the first proton beams started circulating the LHC ring again. In June 3<sup>rd</sup> stable beams started to run in the accelerator producing collisions recorded by the detectors and ready for physics studies.

The LHC has two parallel beam pipes where two beams of protons circulate in opposite directions and intersect in four interaction points in which the four main experiments of the LHC are placed: ATLAS (A Toroidal LHC Aparatus), LHCb, CMS (Compact Muon Solenoid) and ALICE (A Large Ion Collider Experiment). While CMS and ATLAS are both general purpose detectors, LHCb and ALICE are focused on the study of CP violation and  $b$ -physics, and heavy-ion collisions, respectively. The beams of protons are accelerated by 16 radiofrequency (RF) cavities housed in four cylindrical refrigerators called cryomodules. The RF cavities are metallic chambers with the shape and size ideal to make the oscillating electro-

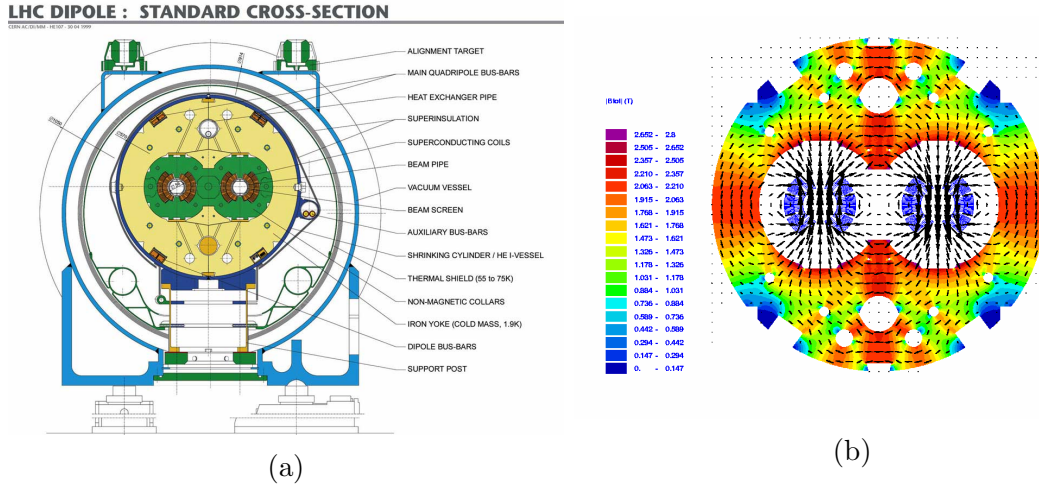


Figure 3.2: Front view of an LHC dipole magnet (a) and a representation of the magnetic field inside it (b) [66, 67].

magnetic field inside resonate. The oscillating field is tuned in a way that a proton with the ideal energy will see zero acceleration while protons with slightly different energies arriving later or earlier will be accelerated or decelerated so that they stay close to the energy of the ideal particle. In order to curve the proton-beams into the circular orbit, the LHC has 1232 dipole magnets, each 14.3 meters long, which are shown in Figure 3.2. In addition, in order to focus the proton beams, 392 quadrupoles are placed along the LHC. These superconducting magnets made of copper-clad niobium-titanium (NbTi) operate at an average temperature of 1.9K, kept by approximately 96 tones of superfluid liquid helium He-II insulated from the exterior of the pipe by a vacuum vessel. The dipole magnets are designed to be able to produce two different magnetic fields in opposite directions in order to properly bend the two beams of protons traveling in different directions, producing a magnetic field of 8.3T at its designed beam energy of 7 TeV.

Over 2800 proton bunches with a time spacing of 25 nanoseconds are supported by the design of the LHC. The proton bunches provide a collision rate of 40 MHz at a design luminosity of  $10^{24}\text{cm}^{-2}\text{s}^{-1}$  and a center-of-mass energy of 14 TeV.

The LHC was developed to investigate the predictions of different theories trying to solve some of the remaining open questions the SM cannot answer. A common denominator of these theories is the possible presence of new particles or new interactions at the TeV scale, which is only possible if high energy collisions are produced. Another common feature of theories beyond the SM is that the processes in which they are involved are rare, which means that a large amount of data is needed in order to be able to obtain statistically significant measurements. In any case, the LHC goal is not only try to find answers to the open questions in particle physics but also

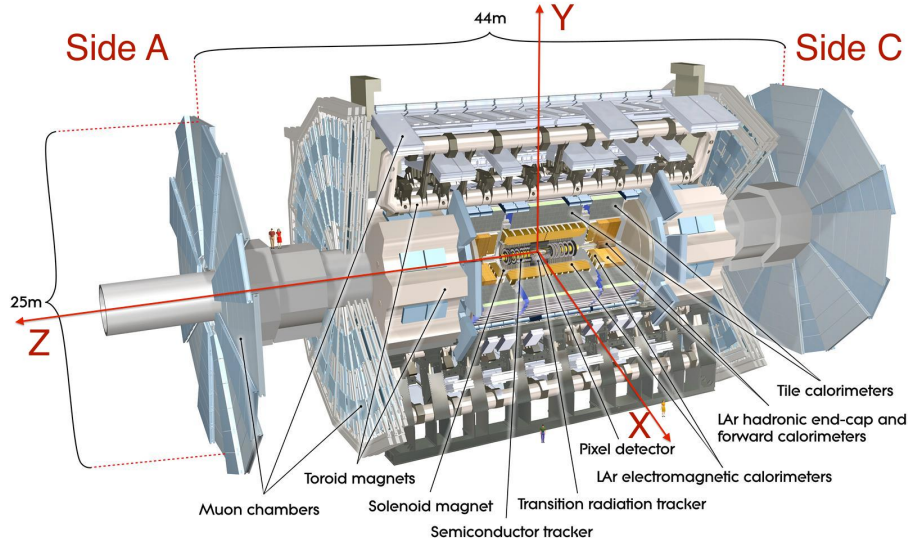


Figure 3.3: Representation of the ATLAS detector. All the subsystems are represented and the coordinate system is marked with red axis [69].

improve the knowledge we already have. By being able to open phase-space regions which were almost closed in previous accelerators and providing enough collisions we could improve the precision of different measurement already done, in particular those concerning the top quark sector given that the LHC is considered as a top quark factory. The LHC also made possible to observe the long awaited Higgs boson with a mass of 125 GeV. But now that the second operation phase has started it would be interesting to see how the Higgs boson couples with different particles and further study its properties using the high amount of data that the LHC is expected to deliver ( $O(300) \text{ fb}^{-1}$  at the end of Run-2 or  $O(3) \text{ ab}^{-1}$  considering the high-luminosity LHC).

### 3.3 The ATLAS detector

The ATLAS detector [68], shown in Figure 3.3, is one of the four main detectors of the LHC. It is a general-purpose detector which targets a wide variety of physics processes arising from the LHC proton-proton collisions. It is 44 meters long and 25 meters tall, composed by cylindrical layers around the beam-pipe, each one with a specific purpose, and two end-caps in order to cover the maximum angle around the interaction point.

Given the high energy and collision rate that the LHC is able to deliver, the ATLAS detector operates under extreme conditions. A high collision rate means that the ATLAS detector needs to be able to handle all the information from the collisions fast enough to be able to differentiate between multiple collisions and have

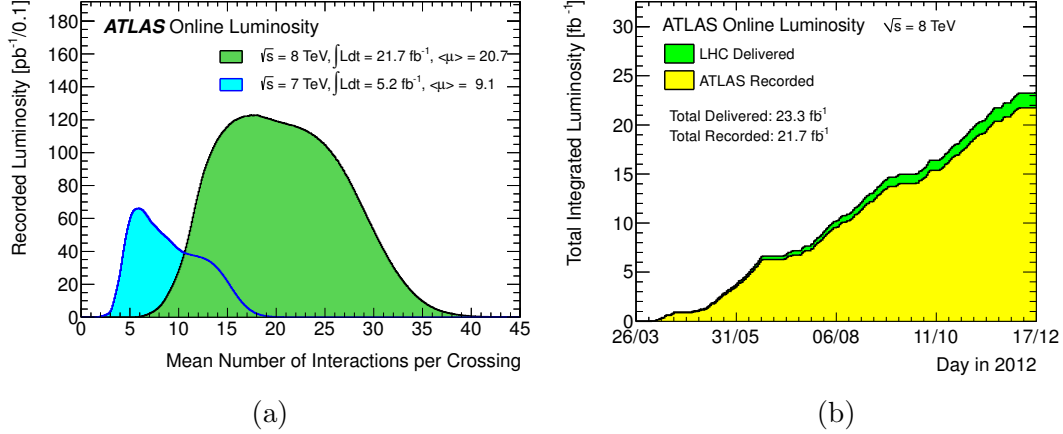


Figure 3.4: Recorded luminosity as a function of the mean number of interactions per bunch-crossing (a) at 7 TeV (blue) and 8 TeV (green). The full datasets at both energies are represented. The total integrated luminosity delivered by the LHC and recorded by ATLAS in the 8 TeV full data-set (b) [70].

enough segmentation to differentiate between different interactions within the same bunch crossing. In a proton-proton collision, tenths of different interactions can take place<sup>1</sup> and the ATLAS detector needs to identify possible interesting candidates and decide if they are worth to be stored for later use. As can be seen in Figure 3.4, the luminosity recorded as a function of the mean number of interactions per bunch-crossing ( $\mu$ ) in the 7 TeV run peaks around  $\mu = 5$  while for the 8 TeV dataset it peaks around  $\mu = 15$ . This is due to the fact that in order for the LHC to deliver higher energy and luminosity the price to pay is harder conditions in terms of pile-up. Figure 3.5 shows an event display taken on April 15<sup>th</sup> 2012 of a  $Z$  boson candidate decaying into two muons. The bottom panel of the figure shows that the  $Z$  boson interaction occurs among other 23 pile-up events which eventually will need to be cleaned. In such a complicated environment, different systems helps the ATLAS detector to properly identify each collision. The general performance goals for each subsystem of the ATLAS detector are summarised in Table 3.1 and each of them is explained in the following sections.

### 3.3.1 Inner detector

The most inner layer of the ATLAS detector is the Inner Detector (ID) [72, 73], shown in figure 3.6. The purpose of the ID is to track charged particles close to the interaction point, a few centimetres away from the beam-pipe. The ID present an

<sup>1</sup>These are called pile-up events as will be described in Chapter 4.

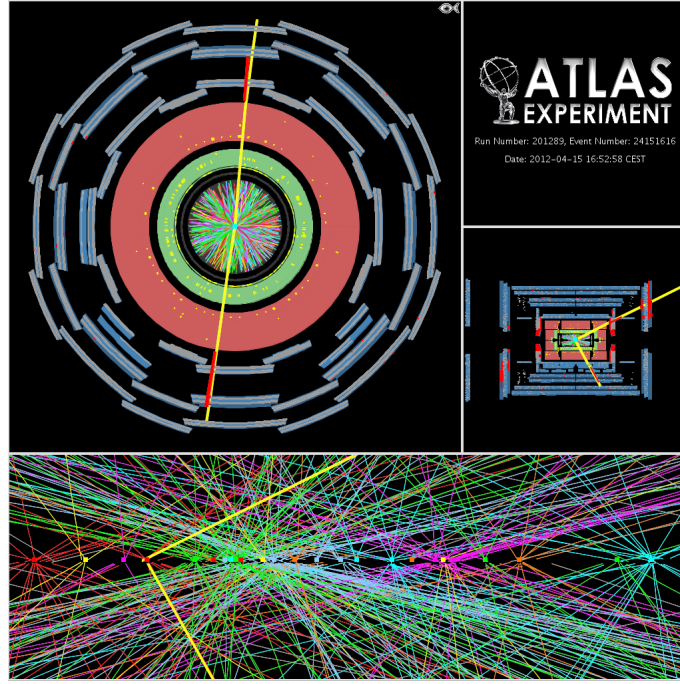


Figure 3.5: Event display of an interaction inside the ATLAS detector which took place in April 15<sup>th</sup> 2012. A  $Z$  boson candidate can be seen decaying into two muons (highlighted yellow lines) among 23 additional pile-up collisions which are shown in the bottom panel of the figure [71].

acceptance in pseudorapidity<sup>2</sup> of  $\eta < 2.5$  and full coverage in the azimuthal angle ( $\phi$ ). The detector provides a transverse momentum ( $p_T$ ) resolution, in the plane perpendicular to the beam axis, of  $\sigma_{p_T}/p_T = 0.05\%p_T/GeV \oplus 1\%$ .

The inner detector is composed by three sub-detectors: the Pixel Detector (PD), the Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). The PD is the most inner layer of the ID. It consists of 3 cylindrical barrel and 2 end-caps of 3 discs, each composed by pixel sensors and microstrips, for a total of 1744 silicon pixel modules. The 3 silicon layers provide a three point tracking of charged particles that travel through them. Due to the short distance between the PD and the interaction point it must be very resistant to radiation. The SCT is very similar to the PD but covers a much larger area. It consists of 4 concentric barrels and 2 end-caps of 9 discs each for a total of 4088 silicon-strip detectors and provides eight strips measurements (four space-points) for the particles traveling from the interaction point. The TRT is based on the use of straw detectors with 4 mm in diameter which contain a gold-plated wire in the center surrounded by a mixture of

<sup>2</sup>The pseudorapidity is defined as  $\eta = -\ln[\tan(\theta/2)]$ , where  $\theta$  is the polar angle, and the distance  $\Delta R$  in the pseudorapidity-azimuthal angle space is defined as  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ . The ATLAS coordinate system is represented with the red axis in Figure 3.3.



<i>Component</i>	<i>Resolution</i>	$\eta$	<i>Trigger (<math>\eta</math>)</i>
ID	$\frac{\sigma_{p_T}}{p_T} = 0.05\% p_T \oplus 1\%$	$\pm 2.5$	
EM. Cal.	$\frac{\sigma_{p_T}}{p_T} = 10\%/\sqrt{E} \oplus 0.7\%$	$\pm 3.2$	$\pm 2.5$
Hadronic Cal.			
Barrel and end-cap	$\frac{\sigma_{p_T}}{p_T} = 50\%/\sqrt{E} \oplus 3\%$	$\pm 3.2$	$\pm 3.2$
Forward	$\frac{\sigma_{p_T}}{p_T} = 100\%/\sqrt{E} \oplus 10\%$	$3.1 <  \eta  < 4.9$	$3.1 <  \eta  < 4.9$
Muon Chambers	$\frac{\sigma_{p_T}}{p_T} = 10\%$ at $p_T = 1$ TeV	$\pm 2.7$	$\pm 2.4$

Table 3.1: General performance goals of the ATLAS subsystems. The unit for  $E$  and  $p_T$  is GeV.

Xenon gas. The Xenon mixture allows to also identify electrons by the detection of transition-radiation photons inside the Xenon gas. The barrel contains around 50000 straws, each one divided in two at the center with readout at both ends of the straw while the end-caps contain 320000 radial straws with the readout in the outer part of the straws. In total, the TRT is read out by 420000 channels. Using two different thresholds for the drift time the TRT is capable of differentiate between tracking hits or transition-radiations hits. The combination of both provides a robust pattern recognition and high precision in azimuthal angle  $\phi$  and  $z$  coordinate.

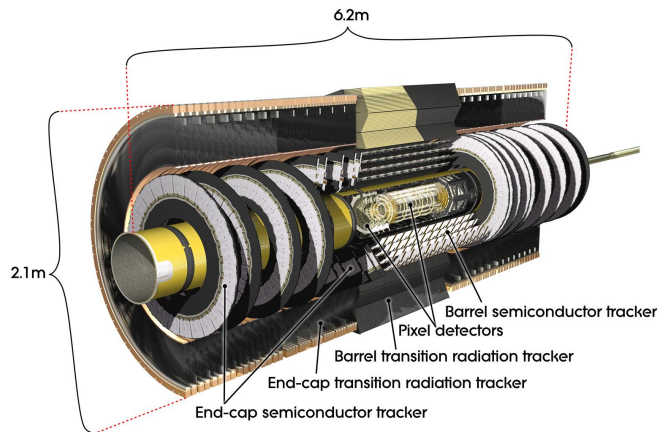


Figure 3.6: Representation of the ATLAS Inner Detector [74].

The ATLAS ID is crucial for the identification of  $\tau$  leptons and  $b$ -jets. A high performance tracking algorithm is needed to identify not only the track of each charged particle, which also allows for a good charged-particle momentum resolution, but also for the identification of secondary vertices near the interaction point. These secondary vertices are used to differentiate between jets arising from the hadronization of light quarks or from  $b$ -jets due to the small time-of-flight of the  $B$  mesons.

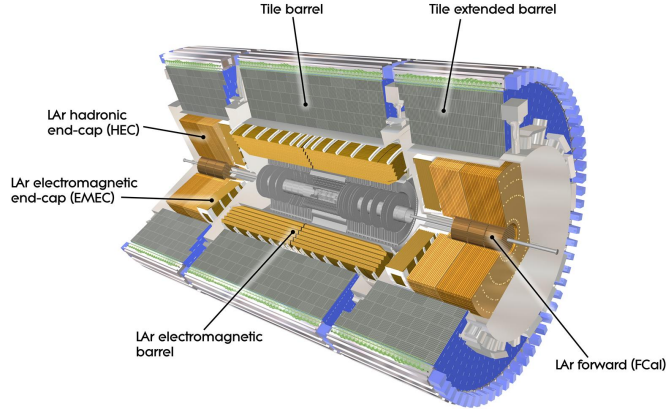


Figure 3.7: Open view of the ATLAS calorimeter system [77].

This small time-of-flight produces a displaced vertex that makes possible to develop a  $b$ -tagging algorithm used in many analyses. A new sub-detector has been introduced in the ID during the long shutdown leading to the Run-II of the LHC which started last April. This new sub-detector, the Insertable  $b$ -layer (IBL) is a new layer of pixel detectors inserted between the PD and the beam-pipe, contributing to an improved  $b$ -tagging algorithm.

### 3.3.2 Calorimetry

Following the Inner Detector the next two layers are the ATLAS calorimeter system, shown in Figure 3.7 and composed by the electromagnetic [75] and the hadronic [76] calorimeters, covering a region of  $|\eta| < 3.2$  and  $|\eta| < 4.9$  respectively. The electromagnetic calorimeter is designed to force the decay and then measure the energy of electromagnetic-interacting particles like charged leptons and photons. The hadronic calorimeter is designed to measure the energy deposition of strongly-interacting particles and hadrons. Both calorimeters usually have to deal with showering of particles produced after they interact with the calorimeter, so a good segmentation and spacial coverage is crucial to properly identify the origin of the shower and position.

The electromagnetic calorimeter (LAr), is a lead (passive material) and liquid argon (active material) sampling detector composed by one barrel and two end-caps. It has an accordion shape which allows for a full azimuthal angle coverage and symmetry without dead paths between different components of the calorimeter. The barrel part and the end-cap part cover a region of  $|\eta| < 1.475$  and  $1.375 < |\eta| < 3.2$ . A pre-sampler is placed in front of the calorimeter for  $|\eta| < 1.8$  to evaluate the energy loss along the way to the calorimeter. The copper/liquid argon hadronic endcap calorimeter (HEC) covers a region of  $1.5 < |\eta| < 3.2$  and a copper/tungsten-

liquid argon forward calorimeter (FCal) covers the region closest to the beam pipe ( $3.1 < |\eta| < 4.9$ ). In total, all the transversally segmented LAr detectors comprise 182468 readout cells.

The hadronic calorimeter (TileCal) is composed by steel as absorber material and scintillating plastic tiles as active material and was designed to measure the energy deposition of strongly-interacting particles. It consists of one central barrel (divided in two parts LBA and LBC and covering the  $|\eta| < 1.0$  region) and two extended barrels (EBA and EBC covering the  $3.1 < |\eta| < 4.9$  region) composed of 64 wedges, also known as modules. The light produced by the particles interacting with the scintillating tiles is collected by photomultipliers (PMTs) via wavelength shifting fibres which are bundled together to form cells with three sampling depths in the radial direction (A, BC and D), and with a granularity of  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  ( $0.1 \times 0.2$  for the last layer). Every cell is read from both sides in order to provide redundancy and the readout of each module is grouped in assemblies of 48 readouts units for a total of 9856 channels, hosted at the outer radius of the calorimeter. A more detailed description of the TileCal can be found in Chapter 4.

### 3.3.3 Muon Spectrometer

The outer part of the ATLAS detector is the Muon Spectrometer [78], shown in Figure 3.8. It is designed to provide both momentum measurement, with an uncertainty in the transverse momentum varying from 3% at 100 GeV and 10% at 1 TeV, and tracking of muons, providing a muon trigger with different  $p_T$  thresholds. It extends from a radius of 4.25 m around the calorimeters to the outer radius of the detector. The momentum measurement is done by measuring the curvature of the muon trajectory which are deflected by the magnetic field provided by three toroidal magnets: one in the central part covering  $|\eta| < 1.1$  and one in each end-cap covering  $1.1 < |\eta| < 2.7$ . In order to measure the curvature of the muons four different tracking technologies are used. Cathode Strip Chambers are used in most of detector, composed of two MultiLayer made of three or four layer of tubes. In the end-cap inner region ( $|\eta| < 2.0$ ) Cathode Strip Chambers are used because of their capability to cope with higher backgrounds rates. In order to provide fast muon triggering, Resistive Plate Chambers and Thin Gap Chambers are used in the barrel and end-cap regions.



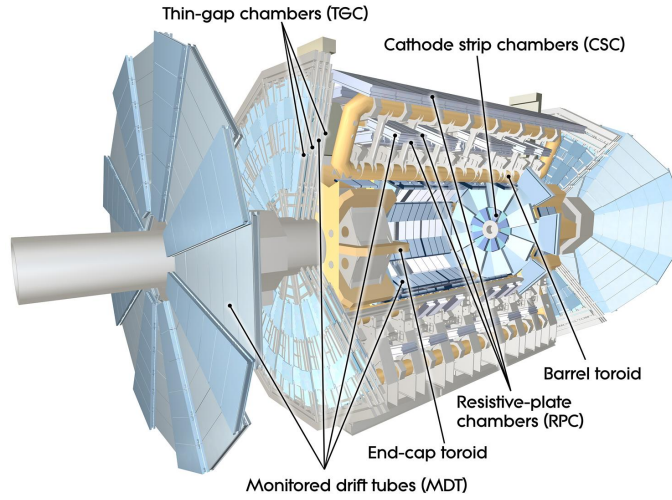


Figure 3.8: Representation of the ATLAS Muon Spectrometer [79].

### 3.3.4 Magnet system

The ATLAS magnetic system [80], shown in Figure 3.9 (a) is composed of four superconducting magnets: the inner solenoid [81], which is placed between the calorimeters and the ID, is used to produce a magnetic field of 2 T that curves charged particles inside the ID and allows the measurement of their momentum by the tracking system; the barrel and en-caps toroidal magnets [82] which provides a toroidal magnetic field between 0.5 T and 1 T used to deflect muons which will interact with the muon spectrometer.

The inner solenoid inner radius 2.46 m, the outer radius is 2.63 m and it is 5.29 m long. It produces a strong and uniform magnetic field which allows a measurements to be made very precisely but due to the field strength, low energy particles (hundreds of MeV) momentum can not be measured. The barrel toroid is formed by eight air-core superconducting coil placed around the calorimeters with a inner radius of 9.4 m, an outer radius of 20.1 m and 25.3 m long. The two end-cap toroids are needed to increase the magnetic field in the end-cap region. They weight 240 tones and are subject to a Lorentz force of 280 tones pushing them to the barrel toroids. In contrast with the inner solenoid, the magnetic field produced by the toroids is not uniform as can be seen in Figure 3.9 (b).

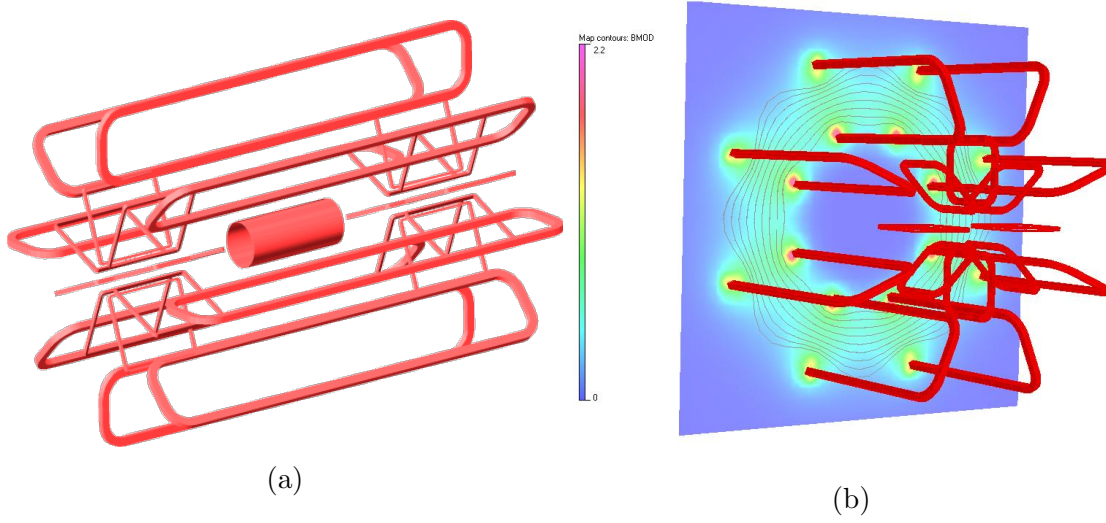


Figure 3.9: Representation of the ATLAS magnet system (a) in which the inner solenoid and the barrel and end-cap toroids can be seen. Representation of the toroid magnetic field produce by the barrel toroids (b) [83, 84].

### 3.3.5 Luminosity subdetectors

The instantaneous luminosity in the ATLAS detector is inferred from direct measurements of beam parameters. It can be written as

$$\mathcal{L} = \frac{n_b f_r n_1 n_2}{2\pi \Sigma_x \Sigma_y}, \quad (3.1)$$

where  $n_b$  is the number of bunches colliding in each revolution around the LHC,  $f_r$  is the revolution frequency,  $n_1$  and  $n_2$  are the bunch population (protons per bunch) in the beam 1 and 2, respectively, and  $\Sigma_x$  and  $\Sigma_y$  characterise the longitudinal and vertical beam widths. In order to measure the beam widths dedicated beam-separation scans, also known as van der Meer (*vdM*) scans. In a *vdM* scan, the beams are separated by steps of a known distance, which allows a direct measurements of  $\Sigma_x$  and  $\Sigma_y$  [85].

A fundamental ingredient of the ATLAS strategy to assess and control the systematic uncertainties affecting the absolute luminosity determination is to compare the measurements of several detectors, most of which use more than one algorithm to assess the luminosity, using either event or hits counting algorithms.

The previously described subsystems, the ID and the hadronic calorimeter, are used to determine the primary vertex in a  $pp$  collision and to measure average particle rates over longer time scales respectively.

In addition to the ID and the hadronic calorimeter, the Beam Conditions Monitor (BCM) and LUCID are also used to evaluate the luminosity. The BCM consists

of four small diamond sensors, approximately  $1 \text{ cm}^2$  in cross section each, arranged around the beam-pipe in a cross pattern on each side of the IP, at a distance of  $z = \pm 184 \text{ cm}$ . It was originally designed to issue beam-abort requests when beam losses start to risk damaging the ID. The fast readout of the BCM also provides a bunch-by-bunch luminosity signal at  $|\eta| = 4.2$  with a time resolution of  $\sim 0.7 \text{ ns}$ . LUCID is a Cherenkov detector specifically designed for measuring the luminosity. Two sets of 16 of mechanically polished aluminium tubes filled with  $\text{C}_4\text{F}_{10}$  gas surround the beam-pipe on each side of the interaction point at a distance of 17 m, covering the pseudorapidity range  $5.6 < |\eta| < 6.0$ . The Cherenkov light produced by charged particles in the gas is collected by the PMTs and are used to produce luminosity measurements for each LCH bunch crossing using *hit counting* algorithms. These algorithms are programmed in Field Programmable Gate Arrays (FPGAs), providing a fast hardware processing which can be reprogrammed.

### 3.3.6 Triggering system

The high collision rate inside the ATLAS provide a huge amount of data which needs to be analysed and store ( $\sim 40$  millions events per second for  $\sim 1.5 \text{ MB}$  per event for raw data). In order to be able to deal with it the ATLAS triggering system [86, 87] is composed of three trigger levels (L1, L2 and event filter) designed to identify, in real time, interesting events to be stored. Each trigger level refines the criteria applied to each event reducing the data taking rate.

The L1 trigger is designed to make fast decisions based on a limited amount of all the detector information. It searches for high-transverse momentum particles or high-energy deposition in the calorimeters as well as high missing and total transverse energy. The muons are identified using the trigger chambers in the barrel and end-cap regions of the muon spectrometer and calorimeter selections are based on reduced-granularity information from all calorimeters. The L1 trigger lower the event rate from  $\sim 1 \text{ GHz}$  to  $\sim 75 \text{ kHz}$  with an event processing time of  $2.5 \text{ ms}$  and identifies regions of interest (ROIs) which passes to the next trigger stages including the region in  $\eta$  and  $\phi$  in the detector identified with interesting regions and the criteria satisfied.

The L2 trigger uses the full detector information, including inner tracker information which was not used in the L1, collected in the ROIs passed by the L1 trigger through a dedicated channel. It identifies further features reduces the event rate to  $\sim 3.5 \text{ kHz}$  with an average event processing time of  $40 \text{ ms}$ . The events analysed

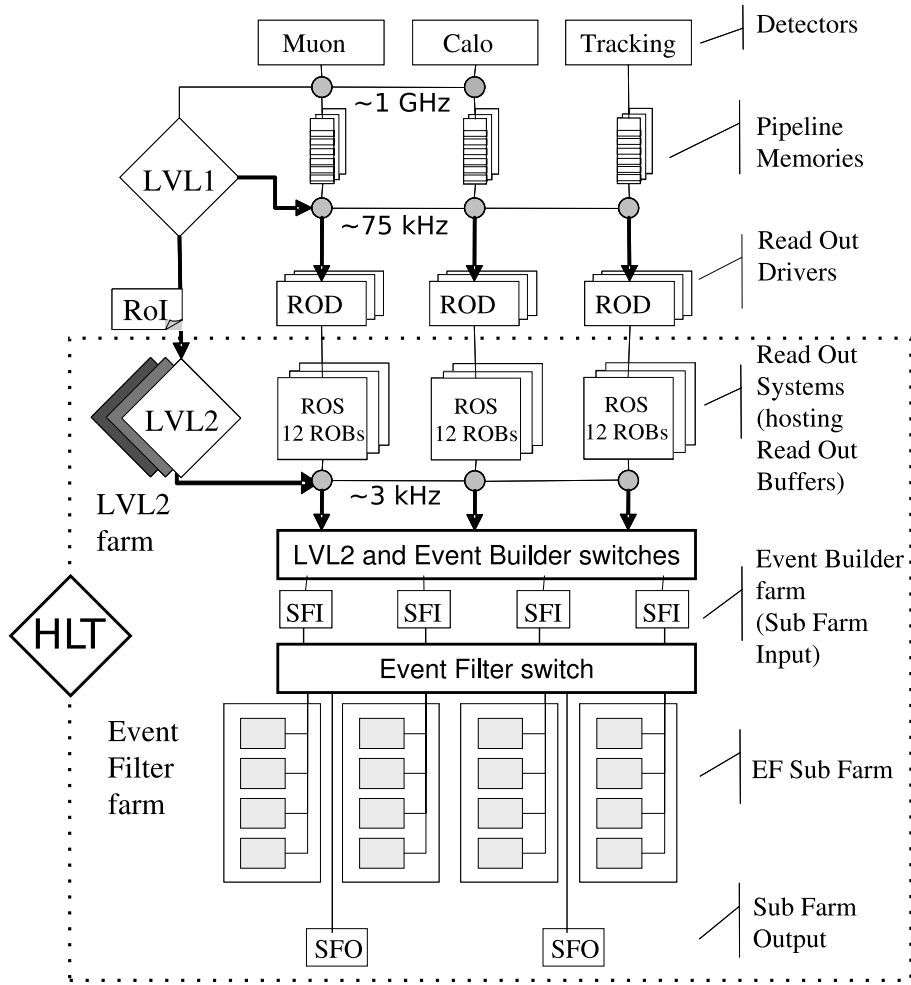


Figure 3.10: ATLAS trigger and data acquisition systems [88].

and selected by the L2 trigger are forwarded to the event filter which applies the last event selection and reduce the event rate to 200 Hz with an average processing time per event of 4 s. The events satisfying all the trigger criteria are stored permanently for later use. The triggering system and data acquisition system is shown in Figure 3.10.

### 3.4 Worldwide LHC Computer Grid

The Worldwide LHC Computer Grid (WLCG) is a worldwide network connecting more than 170 computing centres in 42 countries, linking up national and international grid infrastructures. The goal of the WLCG is to provide the computational infrastructure to store, organize and analyse the  $\sim 30$  Pb of data produced annually by the LHC Experiment [89]. The WLCG is structured in different layers, called *tiers*, each one with a specific set of services.

The first layer, or Tier 0 is the CERN Data Center which his located at CERN

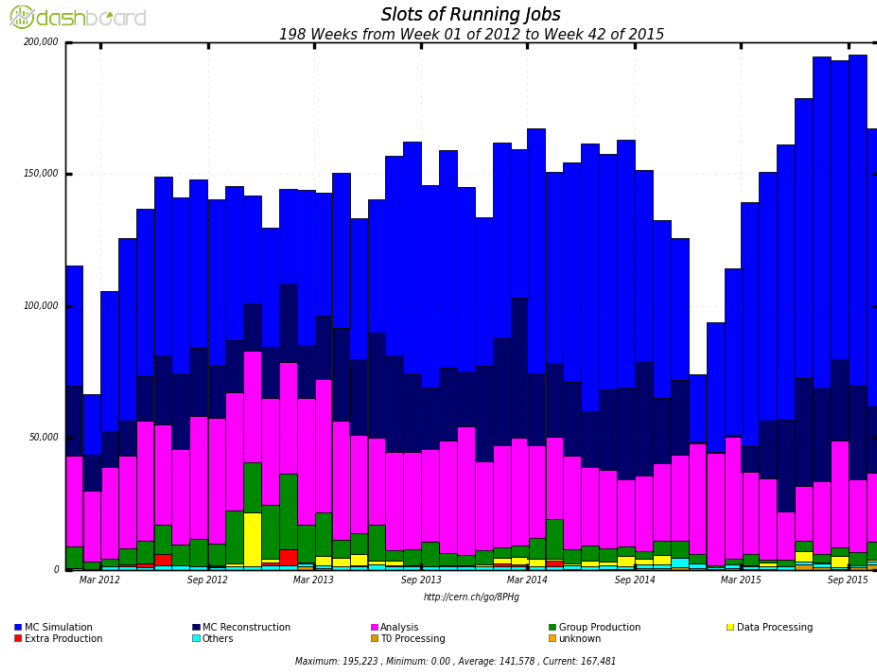


Figure 3.11: Tasks carried out in the WLCG grouped by category from January of 2012 until October of 2015 [90].

and, since 2013, also in the Wigner Research Centre for Physics in Budapest, Hungary over 1200km away. The two sites are connected by two dedicated 100 Gbit/s data links. The Tier 0 is responsible for the first copy of raw data, first reconstruction and replication of raw data and reconstruction output to the Tier 1s. It also reprocess data during LHC down-times.

The Tier 1s are 13 large computer centres with enough computational power to store a proportional share of raw data and distribute it accordingly among the Tier 2s. They also provide large-scale reprocessing of data and safekeeping of the output.

The Tier 2s are computational centres of scientific institutes and universities with enough power for specific analysis tasks and proportional share of simulated event production and reconstruction. There are around 120 Tier 2s around the world.

The Tier 3s are the local machines in which scientist access the WLCG facilities. They can be local clusters in an university or scientific institute or even an individual PC. There is no formal engagement between WLCG and Tier 3 resources.

The WLCG provides computational power for different tasks needed to analyse the  $pp$  collision data. Figure 3.11 shows the number of jobs run by the WLCG from January of 2012 until October of 2015. As can be seen data processing is a marginal portion of the computational tasks carried out by the WLCG.



---

## Chapter 4

# Pile-up noise characterisation in the TileCal Calorimeter

As mentioned in the previous chapter, the LHC is designed to provide a high rate of  $pp$  collisions and under these circumstances the ATLAS detector subsystems need to deal with several simultaneous interactions which need to be properly identified.

For the case of the hadronic calorimeter it means that the energy deposition measured will have overlapping contributions from different interactions which needs to be understood. In this chapter the studies done under the scope of this thesis to characterise the noise introduced in the energy measurements by these simultaneous interactions are presented.

### 4.1 Pile-up inside the ATLAS detector

One of the key features of the LHC is the high amount of statistics it is designed to produce. In order to be able to increase the luminosity, proton bunches are collided at a very high rate, producing several simultaneous interactions in the same bunch-crossing. Nonetheless, increasing the luminosity also means that the operation conditions will be harder since the probability of having simultaneous interactions also increases. These simultaneous interactions are called pile-up events and they need to be carefully studied (characterised in data and modelled in Monte Carlo simulation) in order to properly reconstruct physics objects used in data analysis.

There are two different types of pile-up:

- **In-time pile-up:** pile-up events that occurs in the same bunch crossing. When two bunches of protons collide inside the ATLAS detector several interactions take place simultaneously.

- **Out-of-time pile-up:** when the time-spacing between proton bunches is shorter than the integration time of a sub-detector, different bunch-crossings might be recognised as only one.

Part of the pile-up effects can be reduced by reconstructing each collision independently and identifying each primary vertex. The sophisticated tracking algorithms, designed to recognise patterns using the detailed track information provided by the ID, are able to identify the primary vertex of several simultaneous interactions, which helps to recognise pile-up vertices in order to reduce the impact on the reconstruction of the physics objects.

For the energy measurements in the calorimeters, the information coming from different interactions might not be so easily differentiated. Two tracks which are close together might be well recognised by the tracking algorithms but their energy depositions in the calorimeter might overlap, being hard to know how much energy a given track deposited. The degradation in the energy measurement introduced by pile-up events is known as pile-up noise.

## 4.2 Pile-up noise in TileCal

Particles produced in a  $pp$  interaction inside the ATLAS detector travel through the cells of the TileCal producing light when interacting with the active material. The light collected by the PMTs and read by the readout system is transformed into an energy measurement. The optimal filter (OF) algorithm [91] is used to transform the electrical signal produced by the PMTs into an energy measurement. The main contribution to this energy deposition comes from the interaction which triggered the event but the energy measurement has contribution of pile-up events as well. The component of the energy measurement coming from these pile-up events is known as pile-up noise, which increases as the number of simultaneous interactions increase. In addition to the noise produced by pile-up, the electronics also introduces noise in the readout system of the cells which is known as electronic noise.

The noise in TileCal has a direct impact on the measurements of quantities like jets energy or transverse missing energy  $E_T^{\text{miss}}$ . The ATLAS reconstruction algorithm for jets is based on the topological clustering (topo-cluster) algorithm [92] which relies on the identification of energy depositions which are not likely to be noise fluctuations using an iterative process (Figure 4.1) where adjacent cells are grouped together to form the energy cluster. The noise in a given cell of the TileCal needs to be characterised in order to evaluate if a given energy deposition is likely to be a noise



fluctuation. The algorithm starts from a cell with a measured energy measured over a threshold defined as  $4\sigma$  of the energy distribution randomly triggered events. Then the neighbour cells with an energy up to  $2\sigma$  are included and finally the immediate neighbouring cells are included.

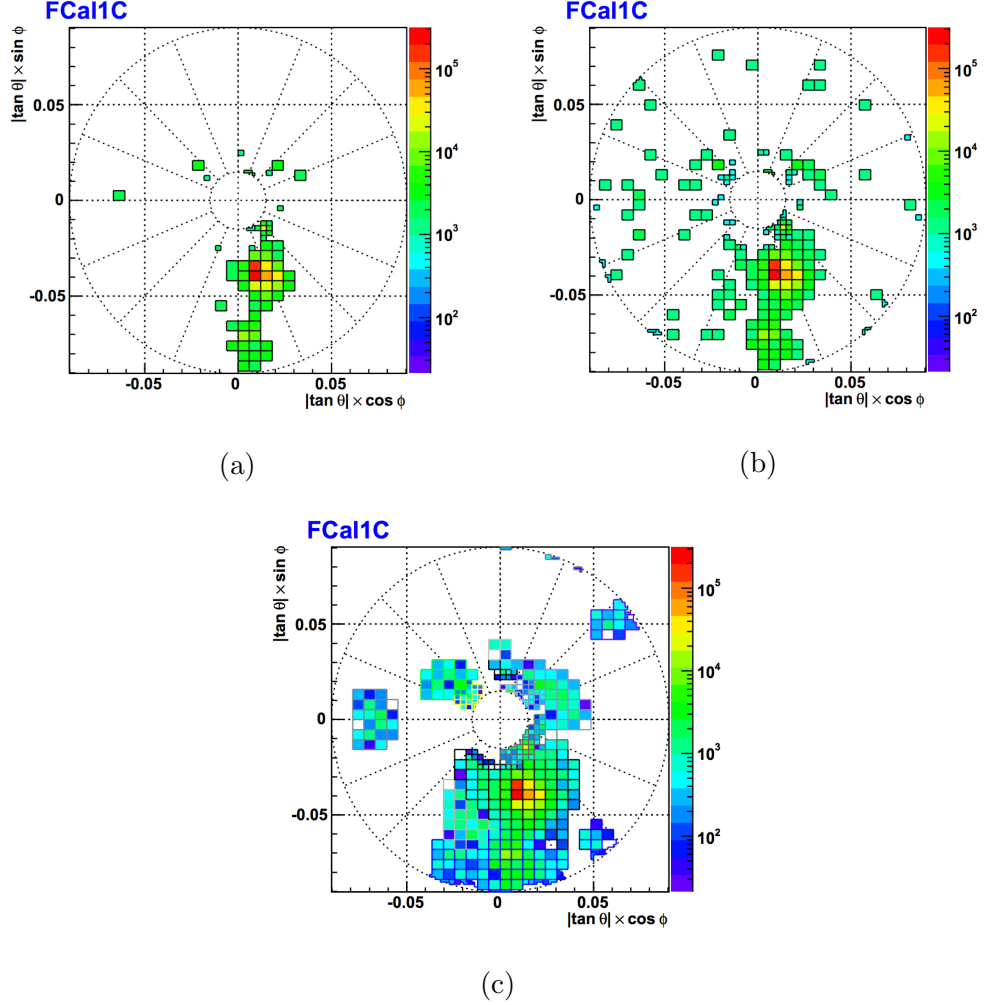


Figure 4.1: The different steps of the topological clustering algorithm are shown for the FCal calorimeter as an illustrative example. First, cells with energy over the  $4\sigma$  threshold are selected (a). Afterward, cells with energy up to  $2\sigma$  (b) and later the adjacent cells (c) are included to form the energy cluster [93].

The expected energy distribution of randomly triggered events needs to be characterised in order to properly define what is a deviation of  $4\sigma$ . Figure 4.2 shows the energy distribution of the TileCal cell A0 (layer A and  $|\eta| < 0.1$ ) and E14 (a cell of located in the gap between the long barrel and the end cap with  $1.4 < |\eta| < 1.6$ )

in figures (a) and (b) respectively. A high pile-up run (in red) is compared to a low pile-up run (in blue) for randomly triggered events and a bunch spacing of 50 ns. There are two main observations that can be made. First, for higher pile-up the energy distribution tends to be wider, due to the degradation of the energy measurements referred earlier due to pile-up. This degradation affects in a different way different parts of the detector. For instance, Figure 4.2 (c) shows the same distribution but for the D0 cell, which is located at the same  $|\eta|$  than A0 but further from the beam-pipe. It can be seen how the effect is smaller for the layer D than for the layer A. The second feature is that the energy distribution are not gaussian. In fact, when comparing the energy distribution in a cell near the beam-pipe in the central part of the detector (A0) with a gap/crack scintillating cell (E14) one can see that the shape may vary drastically.

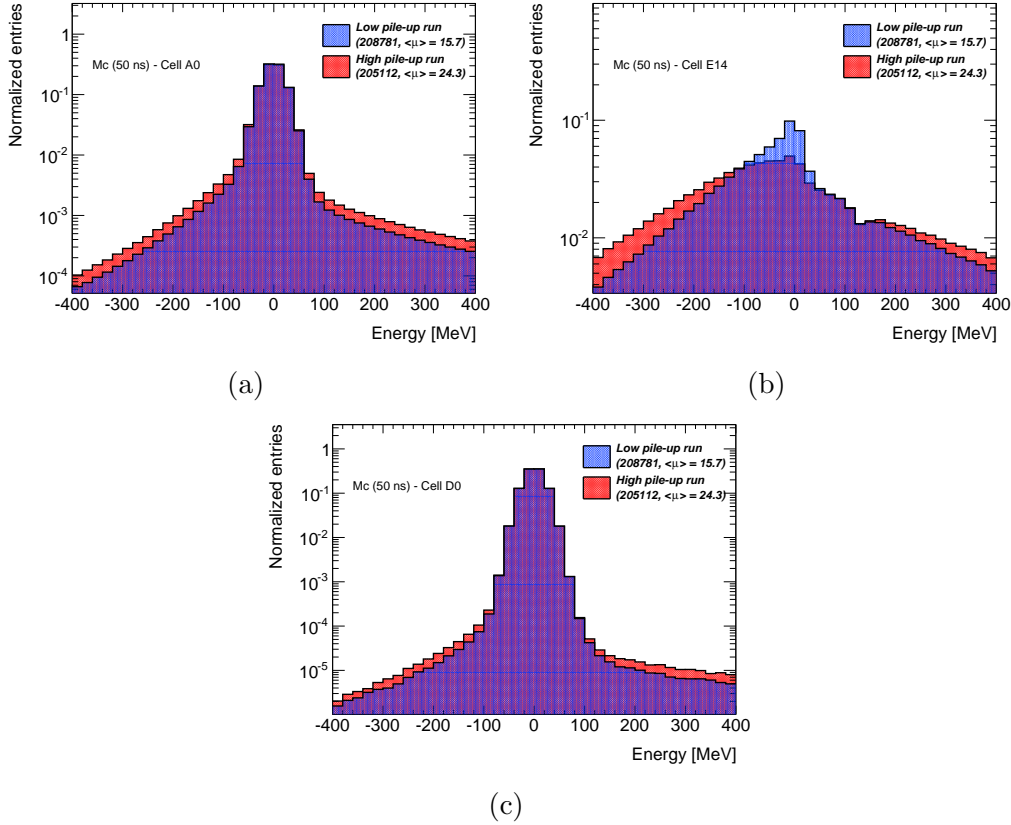


Figure 4.2: The energy distribution of TileCal cells, the A0 (a), E14 (b) and D0 (c) are shown. The red distribution correspond to the energy distribution measured in a high pile-up run while the blue distribution correspond to a low pile-up run. Data collected at  $\sqrt{s} = 8$  TeV and a bunch spacing of 50 ns has been analysed.

In order to properly define the energy threshold for the seed of the topo-cluster algorithm, a two gaussian template was designed [94] and has been used to describe the energy distribution, improving the reconstruction of jets and  $E_T^{\text{miss}}$ . In the following sections the different noise estimators studied under the scope of this thesis will be presented.

### 4.2.1 Analysis procedure

In order to evaluate the noise in each TileCal cell zero bias data (i.e. data randomly collected by the ATLAS detector without any trigger criteria applied which is dominated by pile-up) collected during 2012 at  $\sqrt{s} = 8$  TeV were used. Both 50 ns and 25 ns bunch spacing datasets have been considered with only a few test runs with 25 ns being available. Monte Carlo generated events were also analysed to evaluate the agreement between data and the detector simulation. In order to generate Monte Carlo events mainly composed by pile-up a process which will not produce any signal in the detector, usually  $Z \rightarrow \nu\nu$ , was generated and multijets events were injected as extra hits in the detector simulation.

Figure 4.3 (a) shows a representation of the lateral view of TileCal with only the positive values of  $\eta$ , i.e. the LBA and EBA partitions, being represented since TileCal is symmetric in  $\eta$ . In the representation the different layers (A,BC, D and the gap/crack scintillating tiles, also known as E cells) and towers (each 0.1 step in  $\eta$  represented with dotted lines) can be seen. The E cells are placed in the gap between the long barrel and the extended barrel acting as active material in a region filled with cables, services and power supplies for the other ATLAS sub-detectors. The E cells are used to calibrate the energy lost when passing through the gap in order to recover this energy after the reconstruction.

Each cell of the Tile Calorimeter is connected via wavelength shifting fibres to two PMTs which collect the light produced by the particle interacting with the scintillating tiles and transform it into electric signal. One of the 64  $\phi$  wedges can be seen in Figure 4.3(b) where the scintillating tiles, the fibres and the PMTs are shown.

In order to analyse the data collected by TileCal, an analysis framework has been designed. For each event collected in the different runs all the cells in the detector where analysed. The steps to analyse the data are the following:

- Both PMTs of the cell being analysed are checked to ensure that the cell is active. Since for MC simulated events all cells are active, the list of inactive

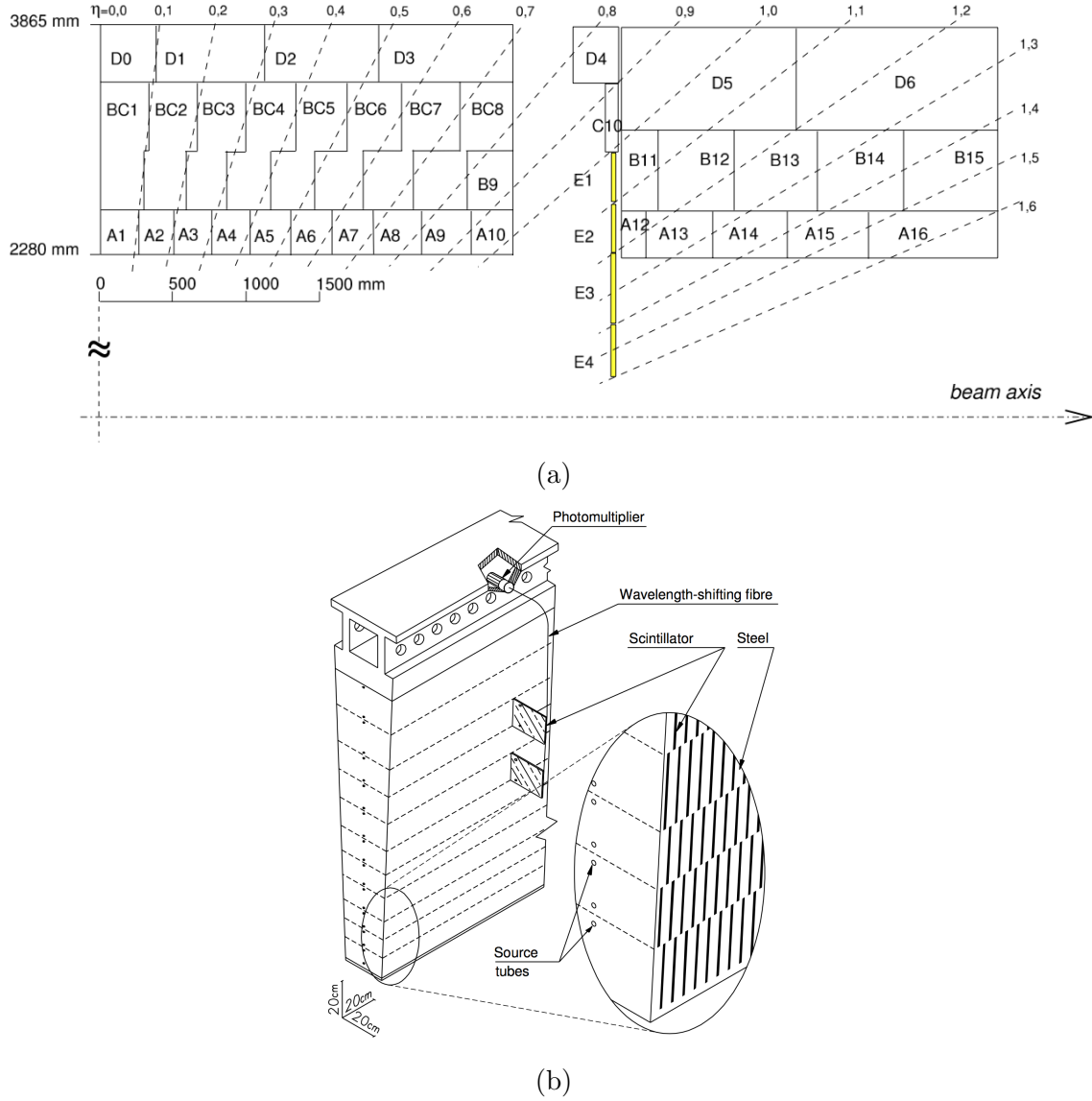


Figure 4.3: Lateral view of TileCal(a) and a  $\phi$  wedge in which the assembly of scintillating tiles, the fibres and the photomultipliers are shown (b) [94].

cells needs to be saved in order to filter them when analysing MC.

- For each active cell, the energy measurement and other properties (pile-up conditions of the event, cell partition, layer, tower and module) are stored to be analysed later.
- The previous steps are repeated for each event and cell. After all events have been analysed the energy distribution of each cell corresponding to different conditions is obtained and can be used to estimate the noise.

Given the TileCal azimuthal symmetry, the information around the 64 divisions in  $\phi$  was integrated.

#### 4.2.2 Root mean square as the noise estimator

Once the energy deposition in each TileCal cell has been read and stored the noise can be evaluated. As shown before, the effect of higher pile-up noise is the increase in the width of the energy distribution. A good estimator to evaluate the width is defined as the root mean square (RMS) of the energy distribution,

$$RMS = \sqrt{\langle E^2 \rangle - \langle E \rangle^2}, \quad (4.1)$$

where  $\langle E^2 \rangle$  is the mean value of the square of the energy and  $\langle E \rangle^2$  is the square of the mean energy deposited in a cell. The uncertainty on the noise is derived using the error of the RMS for a normal distribution. This estimator is meant to describe the total noise, not attempting to separate pile-up and electronic noise.

Figure 4.4 shows the noise as a function of  $|\eta|$  for the layers A, BC, D and E for a given run of 8 TeV with a pile-up  $\langle \mu \rangle = 15.7$  and a bunch-spacing of 50 ns. The distributions are shown as a function  $|\eta|$  given that it is symmetrical for both sides of the detector. As mentioned before, the information around the 64 wedges in the azimuthal angle has been integrated. Different key behaviours can be observed:

- As we move further from the interaction point (moving from layer A to D) the noise tends to decrease. This is due to the fact that the decay products of pile-up events tends to be soft and are likely to be stopped in the lower layers of the calorimeter.
- It can also be seen how the noise is rather constant for the central part of the detector ( $|\eta| < 1$  for layers A and BC and  $|\eta| < 0.8$  for the layer D) but it suddenly increases after reaching  $|\eta| \sim 1.1$ . This is due to the fact that the separation between the TileCal long barrel and extended barrel is located in that region. The gap between both barrels is used to pass upstream material like service cables which implies that the interaction between the particles and this *non-active* material introduces degradation in the energy measurements.
- After this sudden increase the noise starts to drop in the extended barrel. This is due to the fact that the particles reaching the long barrel have a long distance of flight which implies that soft particles as the ones coming from pile-up events will not be able to reach that far.

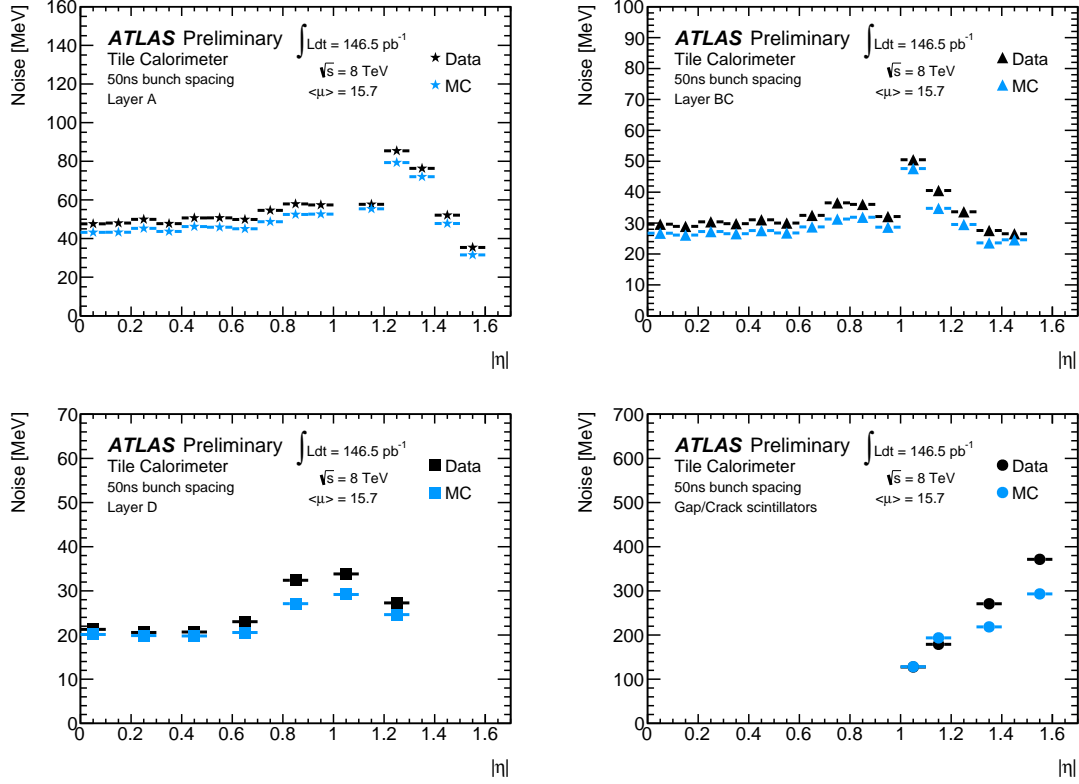


Figure 4.4: Noise dependence with  $|\eta|$  for the different TileCal layers and  $\sqrt{s} = 8 \text{ TeV}$  data with a bunch-spacing of 50 ns and  $\langle \mu \rangle = 15.7$ .

The case for the gap/crack scintillating tiles, the so called E cells, is different. All the passive material in that region introduces a degradation in the energy measured by this cells resulting in higher values of the noise. Nonetheless they are not used for energy measurement purposes but only for calibration.

The bunch spacing is another important factor to include when characterising the noise. During Run-1 the LHC has been operating with a time between bunches of 50 ns. This has been decreased in Run-2 to 25 ns and this is why a few test runs were made with 25 ns before the long shutdown of the LHC. Figure 4.5 shows the same distributions as Figure 4.4 but in this case the data analysed correspond to 8 TeV data with  $\langle \mu \rangle = 10$  and a bunch-spacing of 25 ns. Even if the noise is expected to increase with pile-up, Figures 4.4 and 4.5 show noise values which are similar. When the time-spacing between bunches is decreased the same collision rate can be achieved with fewer simultaneous interactions, decreasing the in-time pile-up, since proton bunches will be colliding at a higher rate. On the other hand, the reduced time-spacing increases the out-of-time pile-up since it is easier to have

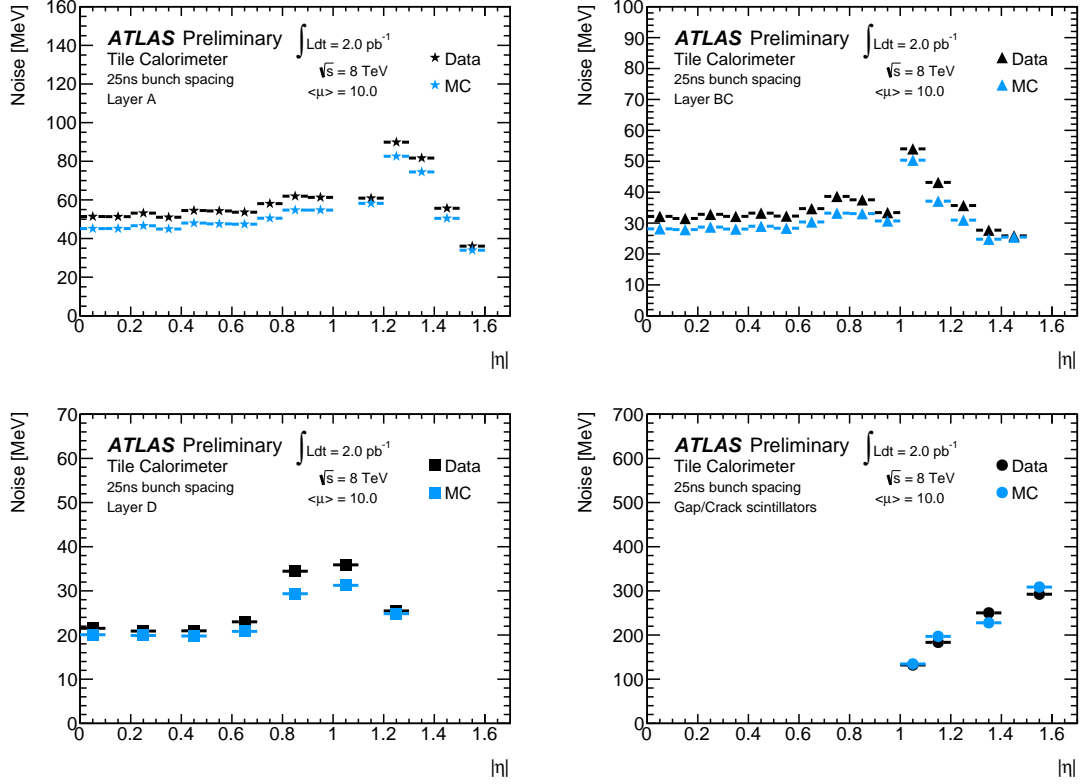


Figure 4.5: Noise dependence with  $|\eta|$  for the different TileCal layers and  $\sqrt{s} = 8$  TeV data with a bunch-spacing of 25 ns and  $\langle\mu\rangle = 10$ .

contamination from different bunch crossing in the energy measurement of a given event. These two factors compensate thus making the noise distribution shown in Figure 4.5 comparable to the noise distribution in Figure 4.4.

Concerning the data and Monte Carlo simulation comparison a good agreement was found in the noise description for different values of  $\eta$  and different layers. When comparing both measurements, discrepancies of a few MeVs can be seen. This is several order of magnitude below the usual energies measured for different physics objects (over  $10^4$  MeV). It is important to notice the big effort done by the ATLAS Collaboration to simulate the detector in order to have good data/simulation agreement.

After describing the noise behaviour for different values of  $|\eta|$  the noise dependence with pile-up has been studied. Several runs were analysed for 50 and 25 ns bunch-spacing in order to get a sizeable sample of different pile-up conditions. Monte Carlo simulations, including high pile-up samples, were also used to compare with data measurements. For this case, since several data runs are analysed together no

pile-up reweighting has been made. All Monte Carlo simulated events have been analysed with its corresponding pile-up conditions. Figure 4.6 shows the noise as a function of  $\langle\mu\rangle$  for different TileCal cells using data collected at 8 TeV with a bunch-spacing of 50 ns while Figure 4.7 shows the same study done with 20 ns bunch-spacing. In this case, since the dependence of pile-up is the same across different towers the information about different values of  $\eta$  has been integrated. As expected, the noise increases with the mean number of interaction per bunch crossing. As it was seen before, pile-up events have more influence in layers closer to the beam pipe as can be seen comparing different layers in Figure 4.6 (a) and (c). For Layer A the noise increases from  $\sim 50$  MeV up to  $\sim 200$  MeV while for Layer D the noise only increase a total of  $\sim 40$  MeV. Although with a limited amount of data, the noise as a function of  $\langle\mu\rangle$  for a bunch spacing of 25 ns seems to agree with the MC simulation, which seems to indicate that the noise increases faster with  $\langle\mu\rangle$  than it does for 50 ns.

### 4.2.3 Energy distribution quantiles as noise estimator

The RMS of the energy distribution is a fair noise estimator, although somehow limited. The RMS gives information on how wide the energy distribution is but it does not give any information about the tails give the non-gaussian regime of the energy distribution. For this reason the RMS alone cannot be used to identify the seed threshold for the clustering algorithm presented at the beginning of this section. The seed of the algorithm needs to have an energy deposition which is equivalent to  $4\sigma$  of a normal distribution. The double gaussian template is able to fit the non-gaussian behaviour of the energy distribution but it is shape dependent. In order to overcome this, a new estimator is proposed for the noise based on the quantiles of the energy distribution itself. In this way, the noise estimation is completely independent of the shape of the energy distribution and might be a candidate to derive the energy threshold for the topo-cluster algorithm.

The quantiles of the energy distribution can be used to not only characterise the width of the energy distribution but also to be able to obtain information about its tails. The main idea is to identify different regions in the energy distribution containing a given percentage of events.

A given distribution can be divided in  $Q$  equal-sized intervals, known as  $q$ -quantiles, and the quantiles are the boundaries of these intervals. The energy  $\epsilon$



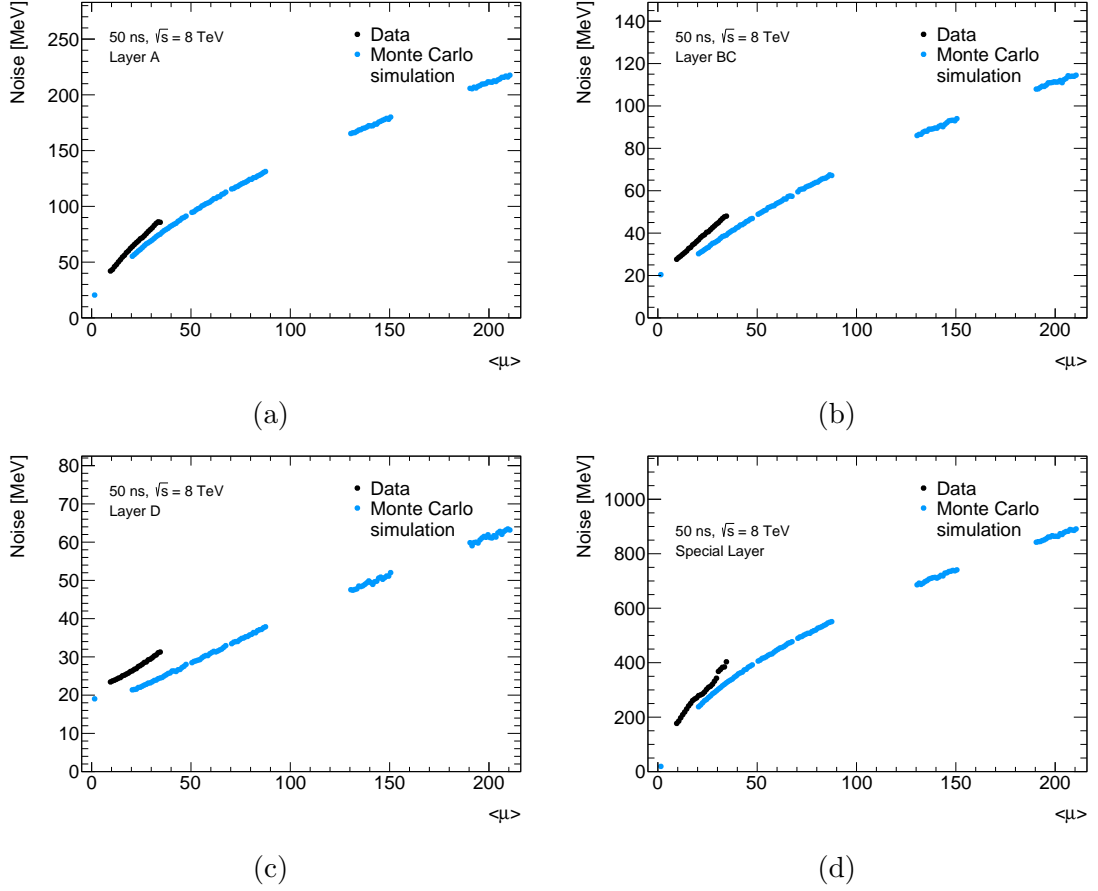


Figure 4.6: Noise dependence with  $\langle\mu\rangle$  for different TileCal layers. The information for different towers has been integrated. Both data and Monte Carlo simulation (including high pile-up samples) are shown.

is said to be the  $k_{th}$  quantile of the energy distribution using a total of  $Q$  intervals if

$$P(E \leq \epsilon) = k/Q. \quad (4.2)$$

For example, a distribution divided in two intervals, the only quantile which is the value dividing the distribution in two intervals is the median of the distribution. In order to evaluate the uncertainty of each quantile a bootstrap method is used. With this method 500 replicas of each energy distribution are created using Poisson fluctuations in each bin of the original energy histogram. A given quantile is calculated in each one of the 500 replicas and a *quantile distribution* is defined using the 500 quantiles obtained before. The RMS of the *quantile distribution* will be the uncertainty on the quantile.

With this estimator, an equivalence can be made between the energy distribution

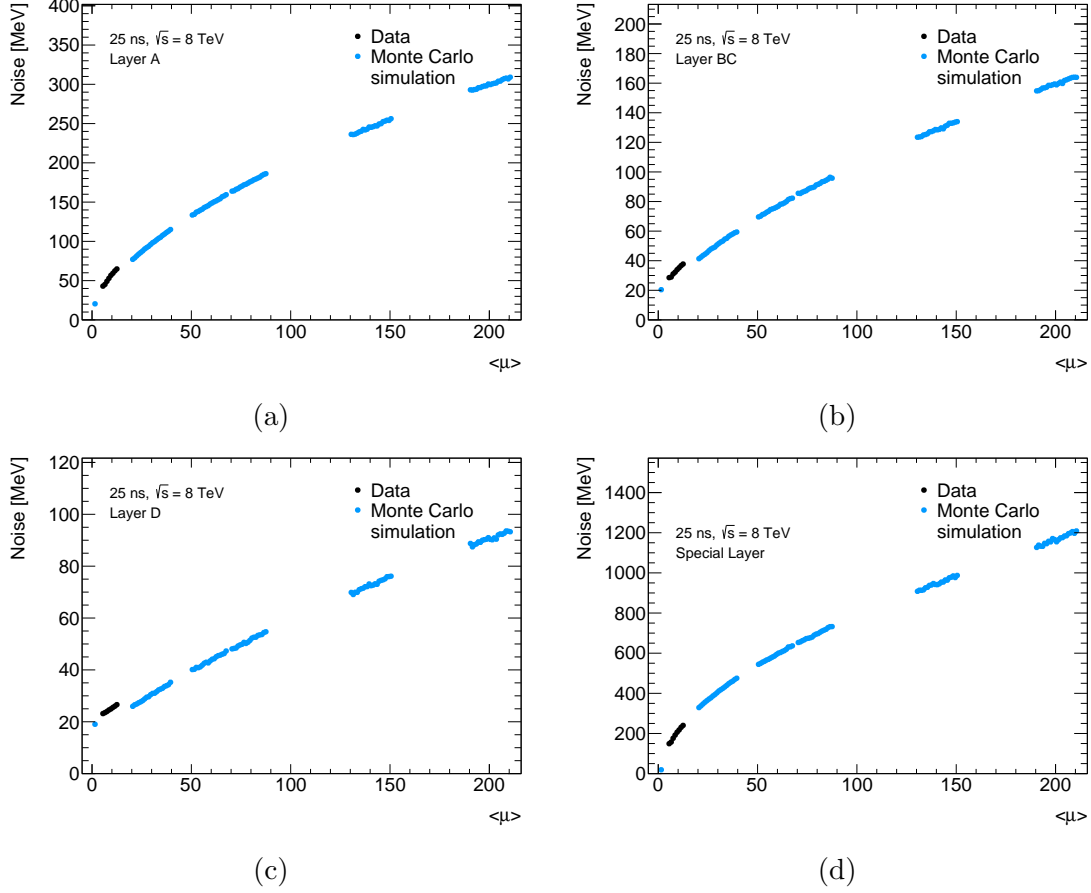


Figure 4.7: Noise dependence with  $\langle\mu\rangle$  for different TileCal layers. The information for different towers has been integrated. Both data and Monte Carlo simulation (including high pile-up samples) are shown.

and a gaussian distribution by finding the corresponding quantiles that define the corresponding confidence interval  $CI = (-n\sigma, n\sigma)$ . Table 4.1 shows the percentage of events which are enclosed within a confidence interval  $CI$  in a normal distribution.

The noise dependence with  $\langle\mu\rangle$  for different cells has been studied using the quantiles estimator. One cell of each layer is shown as an example in Figure 4.8 for data collected with a bunch-spacing of 50 ns. In Appendix A the noise dependence with  $\langle\mu\rangle$  for all the cells is shown. The non-gaussian regime of the energy distribution can be observed with much larger positive tails. This asymmetry increases with pile-up since the negative tails of the distribution remain almost constant while the positive tails get more important.

In order to better estimate the non-gaussian regime of the energy distribution a comparison is made between the two estimators. If the energy distribution were

$1\sigma$	68.27%
$2\sigma$	95.45%
$3\sigma$	99.73%
$4\sigma$	99.99%

Table 4.1: Percentage of the population within a confidence level  $CI=(-n\sigma, n\sigma)$  and the quantiles which define the interval enclosing the same percentage of events.

gaussian, the 99.99% confidence interval could be defined just by knowing the RMS and scaling it by 4 since that confidence interval correspond to  $4\sigma$ . What is shown in Figure 4.9 is a comparison between the estimated using the quantiles estimator for a confidence interval containing the 99.99% and 95.45%, full blue and yellow circles respectively, and compare it to the noise estimated using the RMS scales by a factor of 4 and 2, hollow blue and yellow circles respectively. As can be seen, the noise derived using the quantiles estimator disagrees with the noise derived using the RMS scaled by a factor 2 and 4. This disagreement increases when going further to the tails with the RMS being a good approximation when looking for the central part of the distribution. However it should be noticed that in some cases, like in Figure 4.9(c), this approximation cannot be done.

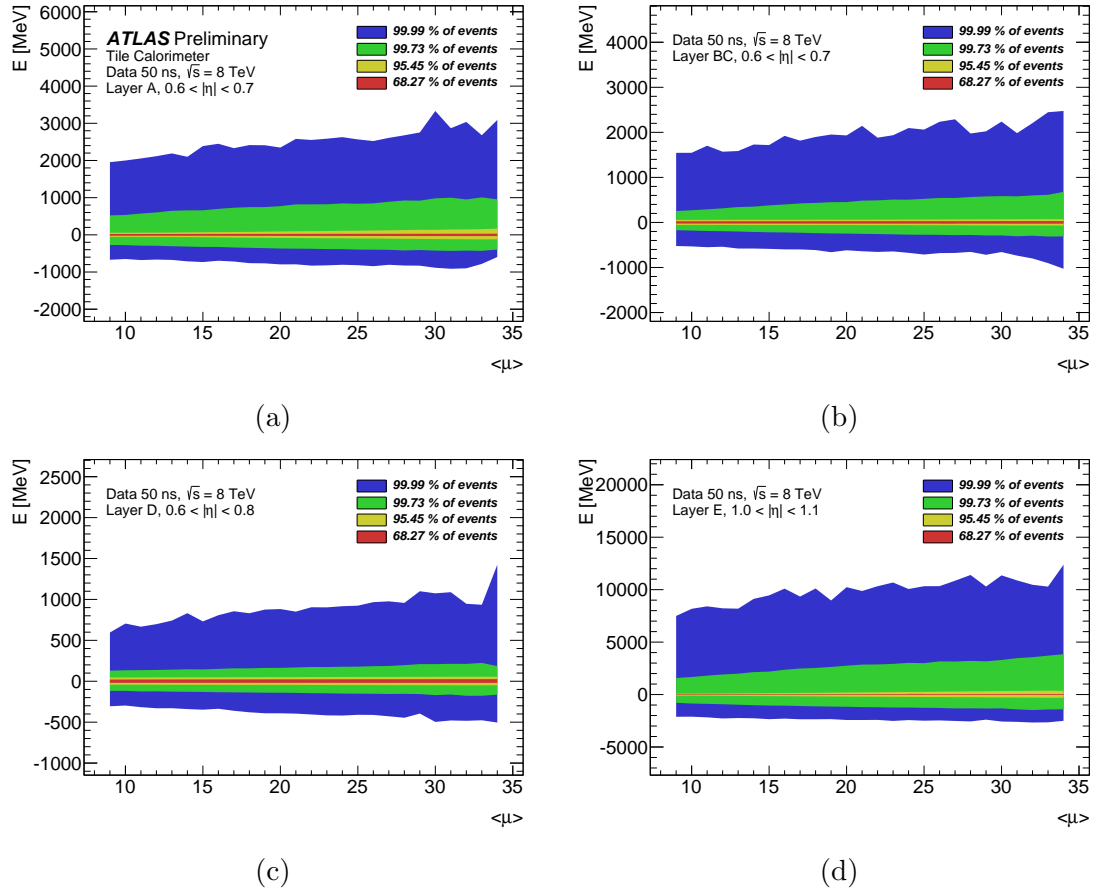


Figure 4.8: Noise dependence with  $\langle\mu\rangle$  for different TileCal cells using the quantiles estimator. Data collected a 8 TeV with a bunch-spacing of 50 ns has been analysed. The information for different partitions and modules has been integrated.

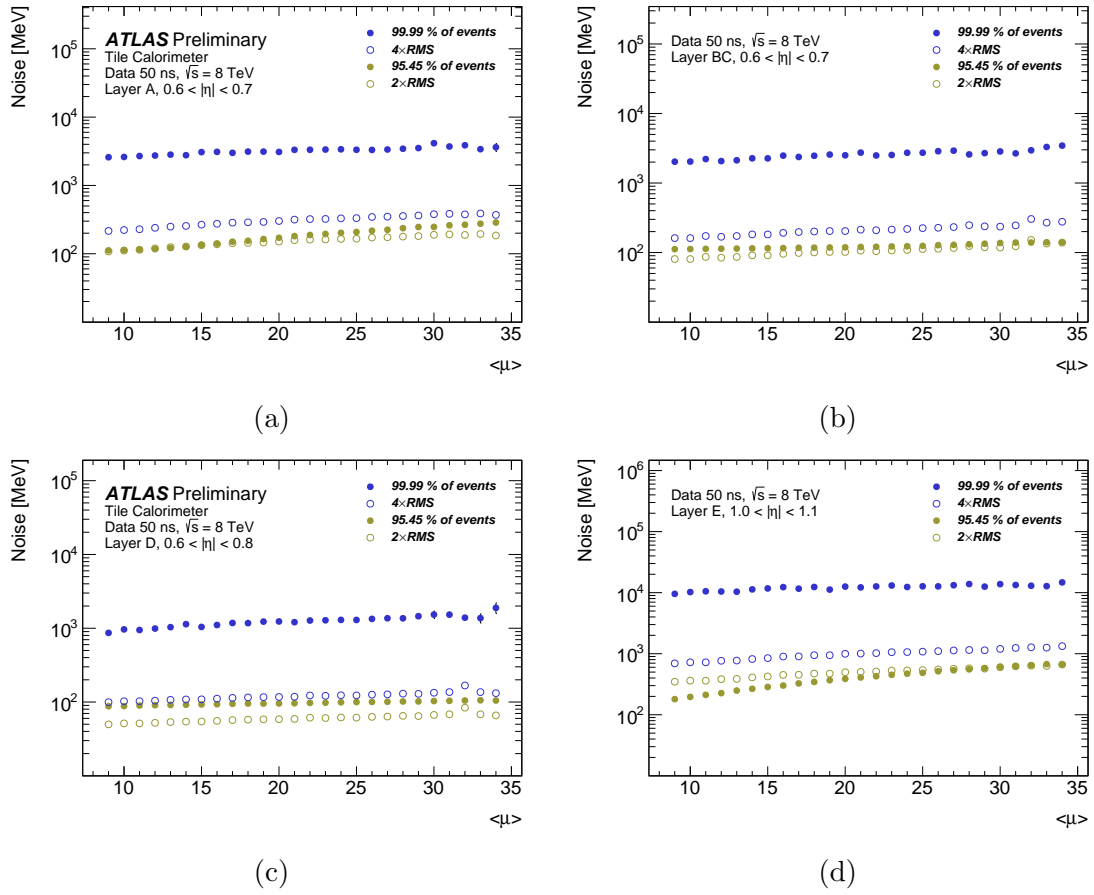


Figure 4.9: Comparison of the noise derived using the quantiles estimator for the regions containing the 95.45% and the 99.99% of the events and the RMS estimator scaled by 2 and 4. One cell of each layer is shown as an example. The information for different partitions and modules has been integrated.



---

## Chapter 5

# Search for new vector-like quarks

In Chapter 2 the SM was presented as the best known description of the nature and interactions of particle physics. Nonetheless some questions are still opened and new models try to answer them in different ways. A common signature of many new physics models is the prediction of new particles arising from new symmetries introduced, for instance, vector-like quarks are a common signature in several non-supersymmetric models. Searching for these new particles is a direct way of testing these models.

Different searches have been performed in both ATLAS and CMS experiments by building a complete and comprehensive set of analyses systematically covering a wide range of possible scenarios for vector-like quark production. In the scope of this thesis, the search of pair- and singly-produced vector-like quark decaying into a  $Z$  boson in a multi-leptonic topology has been performed studying the di-leptonic (2 lepton exclusive selection) and tri-leptonic ( $\geq 3$  leptons) channels. In this chapter an overview of this analysis will be presented discussing the analysis strategy, object reconstruction, data and Monte Carlo simulation samples which have been used, an overview of the systematics uncertainties and finally the results of the analysis.

### 5.1 Data sample and object reconstruction

Data collected by the ATLAS detector during between April and December of 2012 at  $\sqrt{s} = 8$  TeV has been analysed. This analysis was first released with an integrated luminosity of  $14.3 \pm 0.5 \text{ fb}^{-1}$  [85] as a preliminary result in June, 2013 [95]. For this preliminary result only the di-lepton channel and pair-production was considered. The analysis was updated to analyse the full data-set with an integrated luminosity of  $20.3 \pm 0.6 \text{ fb}^{-1}$  and it was improved introducing the single-production mechanism

and dividing the analysis in tri-leptonic and di-leptonic channel [47, 96]. The final version of the analysis is the one which is discussed in this thesis.

Single-lepton triggers with different  $p_T$  thresholds are combined to increase the overall efficiency. The  $p_T$  thresholds are 24 GeV and 60 GeV for the electron triggers and 24 GeV and 36 GeV for the muon triggers. The lower-threshold triggers include isolation requirements on the candidate leptons, resulting in inefficiencies at higher  $p_T$  that are recovered by the higher- $p_T$  threshold triggers. Figure 5.1 shows the efficiency of the combined trigger with lower and higher  $p_T$  thresholds as a function of the  $p_T$  of the electron. The combined efficiency is shown after the hardware-based Level-1 trigger (L1), the software-based Level-2 trigger (L2) and event filter (EF). The efficiency without the higher  $p_T$  threshold trigger included is also shown after the event filter (green line) where the inefficiencies for higher  $p_T$  can be observed. Figure 5.2 shows the combined trigger efficiency for the lower and higher  $p_T$  threshold single muon trigger for the barrel (a) and the endcap (b) region after the Level-1 trigger, Level-2 trigger and event filter. The trigger efficiency is estimated using a data-driven tag-and-probe method using a sample of enriched  $Z \rightarrow \ell\ell$  events. A tag lepton is defined as an offline lepton, reconstructed using the full detector information, matching<sup>1</sup> a lepton passing the unpre-scaled single lepton trigger with  $p_T > 25(20)$  GeV for electrons (muons). The probe lepton is defined as an oppositely charged lepton yielding an invariant mass  $|m_{(tag,probe)} - m_Z| < 10$  GeV (requiring also  $\Delta\phi_{(tag,probe)} > 2$  for muons). The efficiency is defined as the fraction of probe leptons matching an associated online lepton [97, 98].

Events satisfying the trigger requirements must also have a reconstructed vertex with at least five associated tracks with  $p_T > 400$  MeV, consistent with the beam collision region in the  $(x, y)$  plane. If more than one such vertex is found, the primary vertex selected is the one with the largest sum of the squared transverse momenta of its associated tracks.

The physics objects used in this analysis are electrons, muons and hadronic jets, including  $b$ -jets.

Electron candidates are reconstructed from clusters of energy depositions in the EM calorimeter (requiring small leakage in the hadronic calorimeter) which are matched with tracks in the inner detector. The candidates are required to have  $p_T > 25$  GeV and  $|\eta_{cluster}| < 2.47$  (where  $\eta_{cluster}$  is the pseudorapidity of the cluster associated with the electron candidate). Electrons in the region between the barrel

---

<sup>1</sup>Both leptons are said to match if the distance between them is  $\Delta R < 0.15$ .



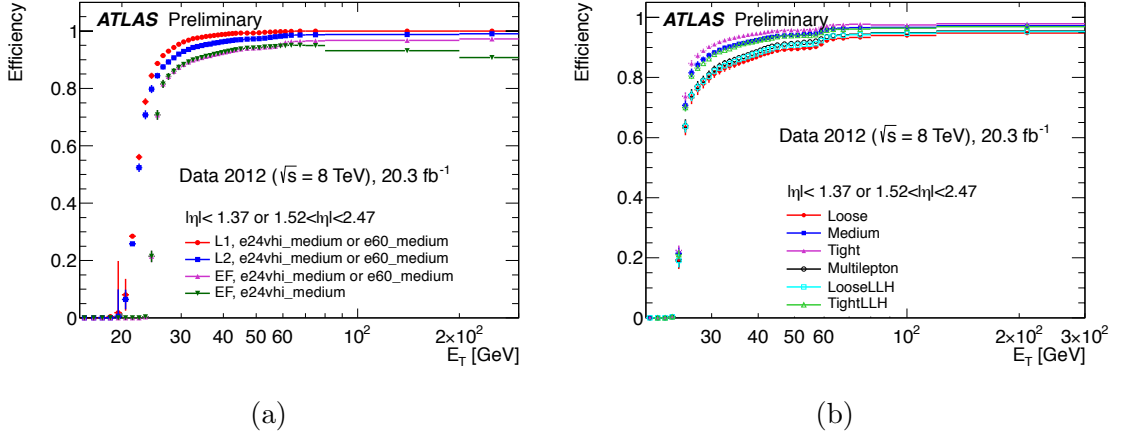


Figure 5.1: Combined single electron trigger efficiency for the lower threshold trigger and higher threshold trigger as a function of the electron  $p_T$  after the Level-1 trigger (L1), Level-2 trigger (L2) and the event filter (EF) stage [99].

and the end-cap of the EM calorimeter ( $1.37 < |\eta_{cluster}| < 1.52$ ) are not considered since the separation between both parts of the calorimeter is used to support the detector and for the passing of service material. Electrons used to reconstruct the  $Z$  boson candidate must satisfy medium quality criteria, requiring at least 2 hits in the pixel detector, at least 1 hit in the  $b$ -layer for  $|\eta| < 2.01$ , a transverse impact parameter  $d_0 < 5$  mm and a separation between the energy cluster and the associated track  $\Delta\eta_1 < 0.0005$  among other shower shape requirements [101]. Electrons not associated with the  $Z$  boson candidate (the extra leptons in the tri-leptonic channel) are required to satisfy tighter identification criteria including requirements in the ratio between the cluster energy and the track momentum and rejection of electrons originated a previously reconstructed photon conversion [101], in order to reduce the contribution from jets misidentified as electrons (“fakes”). Figure 5.3(a) shows the electron reconstruction efficiency for different reconstruction criteria comparing data and MC prediction using enriched data samples of  $Z \rightarrow ee$ . In order to reduce contribution from semi-leptonic decays from  $B$ - and  $C$ -hadron decays inside jets these electrons are required to be isolated. A calorimeter isolation requirement is applied based on the scalar sum of transverse energy deposited inside a cone of radius  $\Delta R < 0.2$  around the electron candidate, as well as a track isolation requirement based on the scalar sum of the track transverse momenta inside a cone of radius of  $\Delta R < 0.3$  around the associated electron track. Both isolation criteria are chosen to have an efficiency of 90% for electrons coming from  $W$  and  $Z$  bosons.

Muons candidates [102, 103] are reconstructed using the tracks found in the

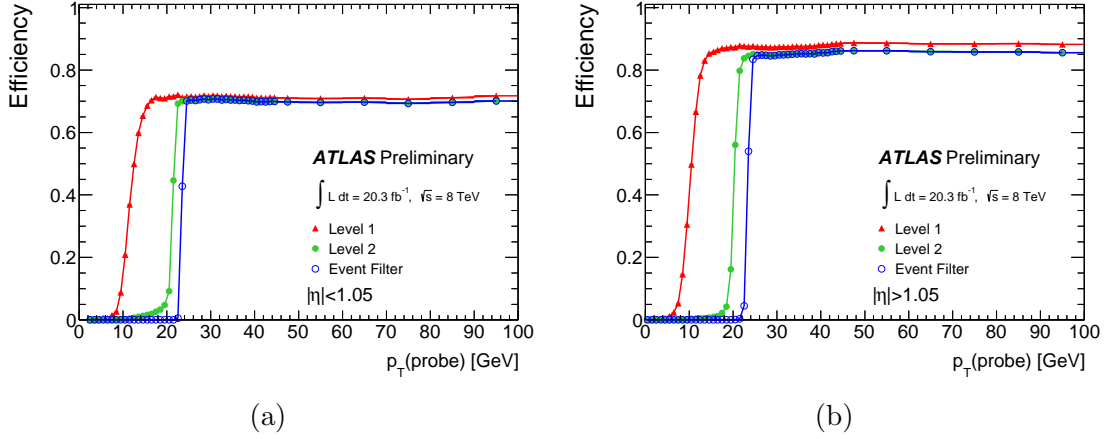


Figure 5.2: Combined single muon trigger efficiency for the lower threshold trigger and higher threshold trigger as a function of the muon  $p_T$  after the Level-1 trigger, Level-2 trigger and the event filter stages in the barrel (a) and endcap region (b) [100].

muon spectrometer and matched with the corresponding tracks in the inner detectors (requiring at least 1 hit in the pixel detector, at least 5 hits in the SCT and at least 5 hits in the TRT for  $0.1 < |\eta| < 1.9$ ), fitting both tracks to find the final muon candidate. Figure 5.3 (b) shows the reconstruction efficiency of muons comparing data and MC simulation as a function of the  $p_T$  of the muon. Selected muons are required to have  $p_T > 25$  GeV and  $|\eta| < 2.5$  with a longitudinal impact parameter with respect to the interaction point smaller than 2 mm. This requirement ensures that the muons candidates are muons produced in the collisions and not cosmic muons, which are muons coming from the atmosphere that penetrates the Earth and pass through the detector. A  $p_T$  dependent isolation criteria is imposed for the muons candidates in which the scalar sum of the track  $p_T$  must be smaller than 5% of the muon  $p_T$  inside a variable cone of  $\Delta R < 10 \text{ GeV}/p_T^\mu$ . This isolation criteria is designed to better define the exclusion cone around the muon depending taking into account that more boosted muons are expected to be more collimated than less boosted muons, defining a narrower cone around the muon candidate.

Jets are reconstructed using the anti- $k_t$  algorithm [105–107] with a radius  $\Delta R = 0.4$  using the topological clusters of energy deposition in the hadronic calorimeter. The topological cluster is calibrated [108, 109] to correct the cluster energy for the effects of non-compensation<sup>2</sup>, dead material, and out-of-cluster leakage effects. The

<sup>2</sup>The response of the calorimeter to electromagnetic and hadronic energy depositions is not the same due to the different scales of the energy measured.

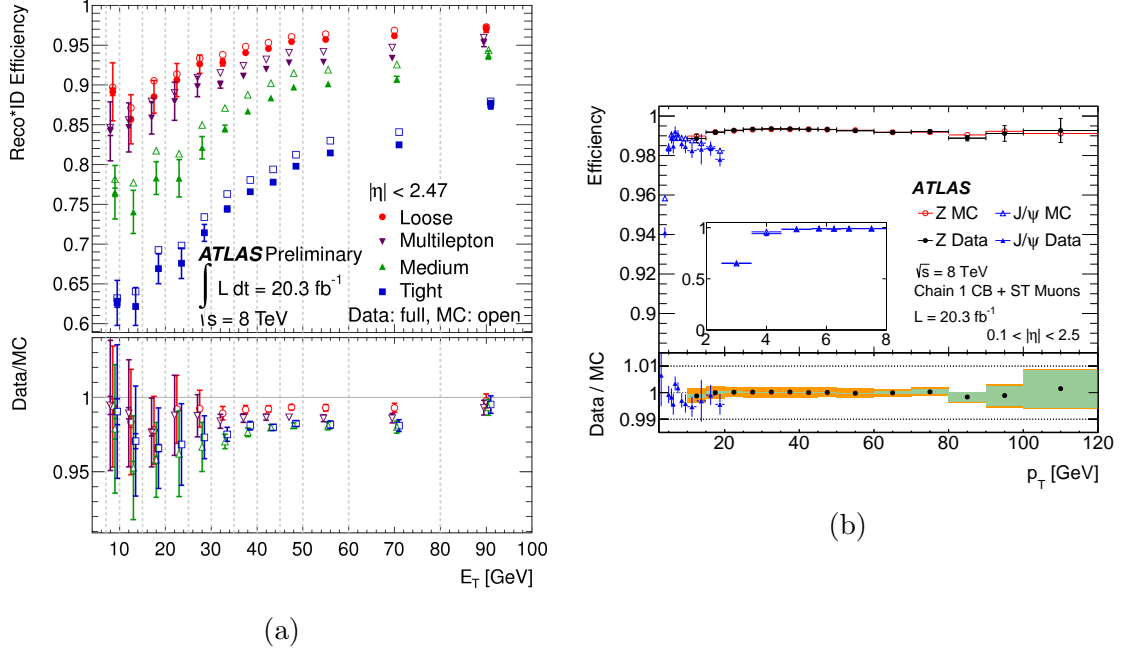


Figure 5.3: Efficiencies for different electron reconstruction criteria for data and MC simulation as a function of the electron  $p_T$  in a  $Z \rightarrow ee$  enriched sample (a) and for muon reconstruction as a function of the muon  $p_T$  comparing data and MC simulation in  $Z \rightarrow \mu\mu$  and  $J/\psi$  samples for CB+ST muons (b). [102, 104].

corrections are obtained from simulation of charged and neutral particles. After energy calibration [110, 111], central jets are selected with  $p_T > 25$  GeV and  $|\eta| < 2.5$ . In order to reduce the contributions from jets originated in pile-up events a requirement is made in the scalar sum of the transverse track momenta in jets with  $p_T < 50$  GeV and  $|\eta| < 2.4$  to ensure that at least the 50% of the scalar sum of track transverse momenta associated with the jet comes from tracks also compatible with originating from the primary vertex. Forward jets used for the single-production analysis selection are defined as those with  $|\eta| > 2.5$  and  $p_T > 35$  GeV. During jets reconstruction no distinction is made between the object identified as electrons and hadronic energy deposition. In order to remove the overlap between electron and jets, any jet identified within a cone of radius  $\Delta R < 0.2$  is discarded. Once these jets are discarded, any remaining electrons or muons within a radius  $\Delta R < 0.4$  of a selected jet are discarded.

Central jets identified as the hadronisation of a  $b$ -quark,  $b$ -jets, are defined using the MV1 tagger which is the output of a multivariate algorithm which uses information from the impact parameters of displaced tracks as well as topological properties of secondary and tertiary decay vertices reconstructed within the jet [112, 113]. The

MV1 operation point used corresponds to a 70% efficiency on  $b$ -jets selection and rejection factors for light and  $c$ -jets of 130 and 5 respectively, calibrated using dileptonic  $t\bar{t}$  events for  $b$ -jets with  $p_T > 20$  GeV and  $|\eta| < 2.5$ . The efficiency of tagging a jet as a  $b$ -jet as a function of the jet  $p_T$  is shown in Figure 5.4.

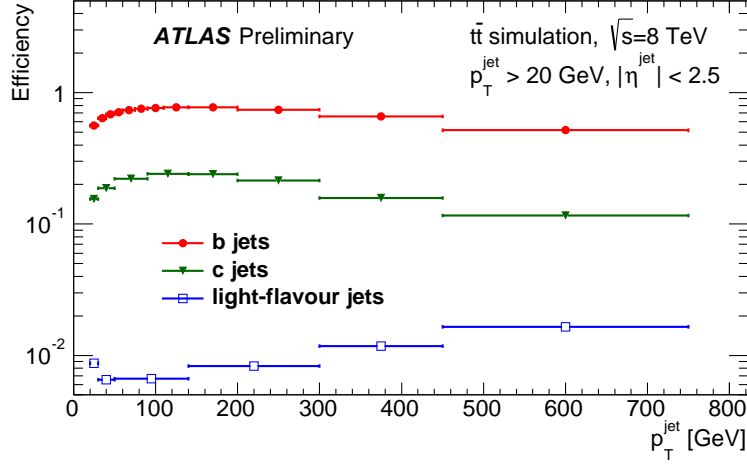


Figure 5.4: Efficiency of the MV1 tagger to select  $b$ -,  $c$ -, and light-jets, as a function of jet  $p_T$  [112].

## 5.2 Analysis strategy

The strategy followed when developing this analysis focuses on a first set of requirements which are valid for both pair and single production mechanisms and small variations in the selection to accommodate different features which are characteristic of each type of mechanism. The event selection detailed in this section is summarised in Table 5.1.

Since the analysis is focused on the search of vector-like quarks decaying to a  $Z$  boson, the main requirement which is common to pair and single production is the presence of a  $Z$  boson candidate. The  $Z$  boson candidate is reconstructed with a pair of same-flavour opposite-sign leptons<sup>3</sup> with an invariant mass that satisfies  $|m_{\ell\ell} - m_Z| < 10$  GeV, where  $m_Z$  is set to be 91.19 GeV. The analysis is then divided in two different channels based on the lepton multiplicity ( $N_{\text{lep}}$ ) requiring exactly two leptons for the dilepton channel and at least three leptons for the trilepton channel. Figure 5.5(a) shows the lepton multiplicity distribution after requiring the presence of a  $Z$  boson candidate for the  $SU(2)$  singlet model with a mass of 650 GeV.

<sup>3</sup>Only electrons and muons are considered as leptons in the following.

Event selection			
$Z$ candidate preselection $\geq 2$ central jets $p(Z) \geq 150$ GeV			
Dilepton channel		Trilepton channel	
$= 2$ leptons		$\geq 3$ leptons	
$\geq 2$ b-tagged jets		$\geq 1$ b-tagged jet	
Pair production	Single production	Pair production	Single production
$H_T(jets) \geq 600$ GeV	$\geq 1$ fwd. jet	–	$\geq 1$ fwd. jet
Final discriminant			
$m(Zb)$		$H_T(jets+leptons)$	

Table 5.1: Event selection for both dilepton and trilepton channels and for single and pair production mechanism

The dashed region is the complete background prediction and the red and blue lines are the signal hypotheses for the vector-like  $T$  (red) and  $B$  (blue) for pair (solid) and single (dotted) production. It can be noticed that only  $\sim 20\%$  of the events satisfy  $N_{lep} \geq 3$  but it is a region where signal dominates over background since the decay products of the vector-like quarks are more likely to produce higher lepton multiplicity than background processes. In the case of single production of the  $B$  quark, the neutral current decay won't produce as much leptons as for the  $T$  quark. This is why trilepton channel is not considered in single production of  $B$  quarks.

Another common feature is the high number of jets ( $N_{jet}$ ) and  $b$ -tagged jets ( $N_{tag}$ ) in the decay products of vector-like quarks. Figure 5.5b shows the  $b$ -tagged jet multiplicity in the dilepton channel. It can be noticed that for the pair production scenario  $\sim 50\%$  of signal events will be produced with  $N_{tag} > 2$ . It is easily understood when taking into account that for a pair of  $B$  quarks there will always be at least one  $b$  jet in the decay products (i.e.  $B \rightarrow H/Zb$  or  $B \rightarrow Wt \rightarrow WWb$ ) and for a pair of  $T$  quarks the same reasoning applies. In the case of single production the  $N_{tag}$  distribution is not so wide, instead it peaks in  $N_{tag} = 1$  which is understood since only one vector-like quark is present in the event. In both dilepton and trilepton channel at least two central jets are required. In the dilepton channel at least two  $b$ -tagged are also required while at least one is required in the trilepton channel. This difference in the selection is chosen to retain statistic in the trilepton channel. In the dilepton channel two control regions are defined for  $N_{tag} = 0$  and  $N_{tag} = 1$  which are used for background corrections as will be discussed later.

Given that the  $Z$  boson is a decay product of the new heavy vector-like quarks,

it is expected to be considerably harder than the  $Z$  bosons produced in background processes as can be seen in Figure 5.5 (c), which shows the  $p_T(Z)$  distribution for the dilepton channel. The  $p_T(Z)$  is required to be greater than 150 GeV in order to suppress the background contamination.

A kinematical signature of signal events produced via pair production is the production of several high energetic jets. Under these circumstances the scalar sum of the  $p_T$  of the selected central jets ( $H_T(\text{jets})$ ) is considered in order to further suppress background contamination. Figure 5.5 (d) shows the  $H_T(\text{jets})$  distribution for the dilepton channel after the  $Z + \geq 2$  central jets selection. It is not the case for single production where less jet activity is present. Events in the dilepton channel for the pair production selection are required to satisfy  $H_T(\text{jets}) > 600$  GeV. The  $p_T$  of the leptons is not considered for the dilepton channel since this information is already taken into account for the  $p_T(Z)$  requirement. Nonetheless, the  $p_T$  of the leptons is included in the trilepton channel, in particular to be able to consider the  $p_T$  of the third lepton, which is not taken into account in the  $p_T(Z)$  requirement. The  $H_T(\text{jets} + \text{leptons})$  distribution is used as final discriminant in the trilepton channel for pair production rather than imposing a minimum value requirement to further suppress background contamination.

The light-flavour quark produced in association with the singly-produced vector-like quark gives rise to an energetic forward jet. Figure 5.5 (e) shows the forward jet multiplicity in the trilepton channel. For the single-production hypothesis at least one forward jet is required for both dilepton and trilepton channel.

The invariant mass of the  $Z$  boson candidate and the highest- $p_T$   $b$ -jet system,  $m(Zb)$ , is used as final discriminant used for hypotheses testing for the dilepton channel. As mentioned before, the  $H_T(\text{jets} + \text{leptons})$  distribution is used for the trilepton channel. Figure 5.6 (a) shows the  $m(Zb)$  distribution Figure 5.6 (b) shows the  $H_T(\text{jets} + \text{leptons})$  distribution.

In order not to bias the analysis design a common strategy is to define a *blinding policy*. This implies that in a signal-dominated region data measurements are not represented (neither in the distributions nor in the yields count) until the background prediction is well modelled and understood. The blinded region was established for the di-lepton (tri-lepton) channel after selecting a  $Z$  boson candidate with  $p_T \geq 150$  GeV and  $N_{\text{tag}} \geq 2$  ( $N_{\text{tag}} \geq 1$ ). Once the background prediction was understood in the unblinded region and after a formal approval of the Collaboration, the data measurements were introduced in the blinded region to test the signal

hypothesis. From this point on the event selection remained frozen.

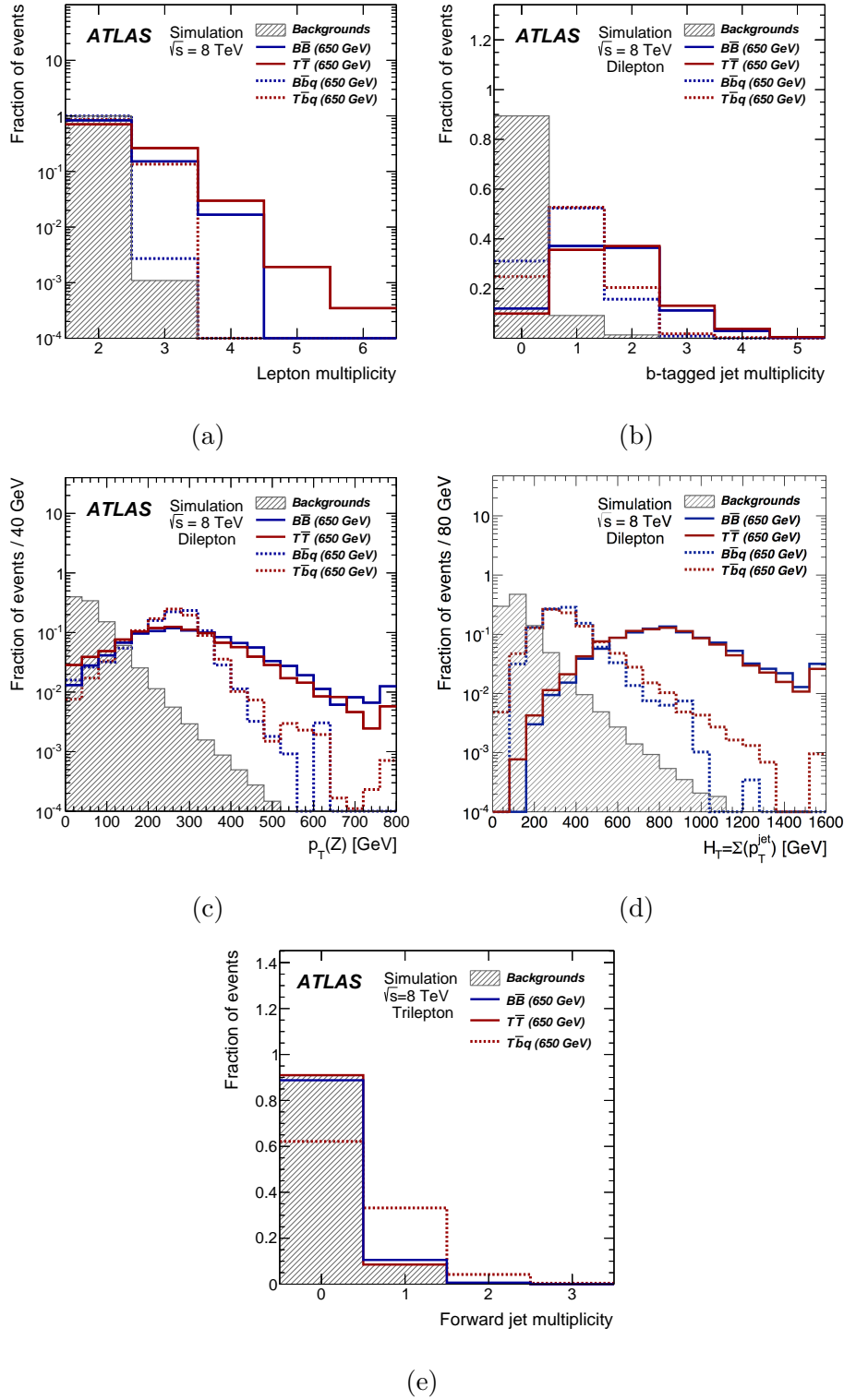


Figure 5.5: Unit normalised distributions of the variables used to discriminate between signal and background. The shaded region represent the background prediction while the lines represent the different signal hypotheses for pair (solid) and single (dotted) production and  $T$  (red) and  $B$  (blue) vector-like quarks.



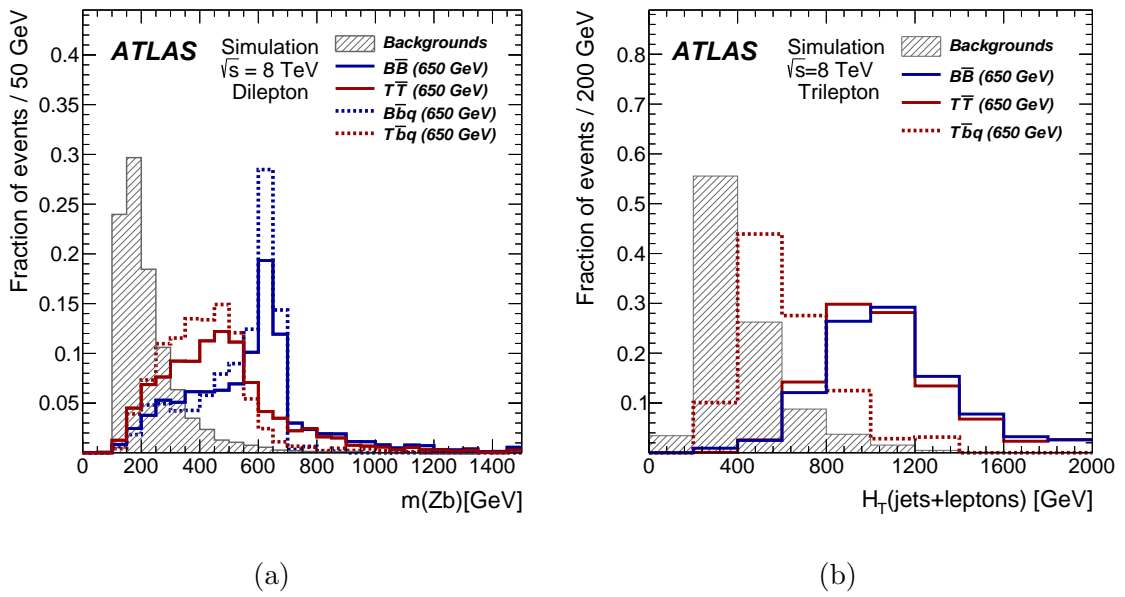


Figure 5.6: Unit normalised distributions of the final discriminant used for hypothesis testing for the dilepton (a) and trilepton (b) channel. The discriminant variable for the dilepton channel is the invariant mass of the  $Zb$  system reconstructed as explained in the text. For the trilepton channel the  $H_T$  (jets + leptons) distribution is used as final discriminant.

## 5.3 Signal modelling

For this analysis both single as pair production have been used. As mentioned in Chapter 2, the pair production is mediated via QCD interaction, remaining model independent while the single production mechanism depends on the model which is being considered since the production rate is mediated by the EW coupling between the VLQ and the gauge bosons. In this section the different approaches taken for pair and single production mechanisms are detailed.

### 5.3.1 Pair production

Monte Carlo generated samples of leading-order (LO) pair-production of  $T\bar{T}$  and  $B\bar{B}$  pairs have been generated with **PROTOS v2.2** [44, 114] interfaced with **PYTHIA v6.421** [115] for parton shower and fragmentation, and using the MSTW 2008 LO [116] set of parton distribution functions (PDFs). The cross-sections used to normalise these samples has been computed using **Top++ v2.0** [117], a next-to-next-to-leading-order (NNLO) calculation in QCD including resummation of next-to-next-to-leading logarithm (NNLL) soft gluon terms, using the MSTW 2008 NLO PDFs [116, 118]. Figure 5.7 shows the cross-section as a function of the vector-like quark mass for pair and single production. The pair production cross-section (solid line) is computed with **Top++**, as mentioned before, while the single production cross section (dashed and dotted line) is provided by **PROTOS** and **MadGraph** [119] using two different electroweak coupling parameters which will be discussed later.

Within the scope of this thesis the generation and validation of pair production signal samples was performed. Figure 5.9 shows some distributions produced to validate the pair-production signal samples. Figure 5.9(a), (b) and (c) show the mass,  $p_T$  and pdg-id of the decays of the vector-like  $T$  while (d), (e) and (f) show the same distributions for the vector-like  $B$  for a mass of the vector-like quark of 600 GeV. Figure 5.9(g), (h) and (i) show the mass of the  $Z$ ,  $H$  and  $W$  boson respectively while (j), (k) and (l) show their  $p_T$ . The mass of the  $H$  boson is fixed at 125 GeV. Since these signal samples were produced to be used by all the vector-like quarks analyses in the ATLAS Collaboration the idea is to follow a general approach that suits all possible scenarios when considering what regions of the phase space to generate. In this sense, a given model or a given branching ratio alone are not desired since analyses with different topologies might be more sensitive to different phase-space regions and may suffer from a lack of statistic in the Monte Carlo samples. Masses ranging between 350 GeV and 1050 GeV were produced in steps of 50 GeV using

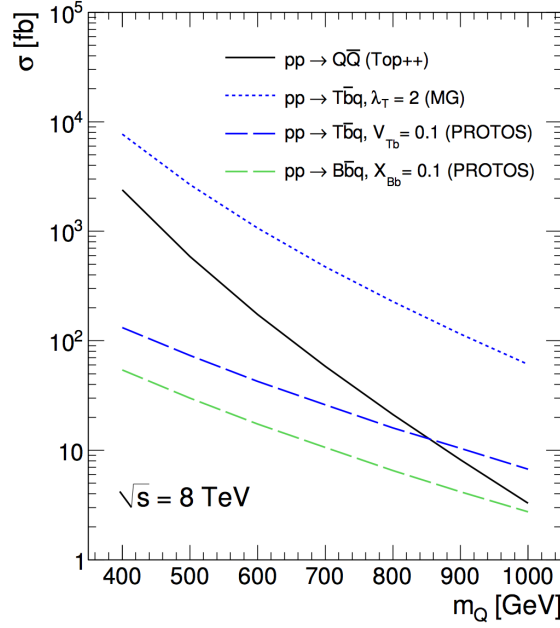


Figure 5.7: Cross section for pair production (solid line) and single production using PROTOS (dashed lines) and MadGraph (dotted line) with two different coupling parameterisation.

the singlet model presented in [44]. In the referenced text it is discussed that the kinematic differences between doublet and singlet<sup>4</sup> smear after detector simulation is applied. In order to evaluate the impact of these differences in the vector-like quark analyses, masses of 350 GeV and 600 GeV were also generated for a doublet model. Figure 5.8 shows the comparison between the singlet and doublet prediction for the  $m(Zb)$  distribution, the final discriminant of the di-leptonic channel for pair production, for the  $ee$  and  $\mu\mu$  channel separately using a vector-like  $T$  with a  $m(T) = 600$  GeV. A fairly good agreement is found within statistical uncertainties. Since no differences were found in the overall performance of the analyses the singlet model was used in all cases and a reweight in the relative composition of the decay modes was done to obtain any desired branching ratio.

In order to produce signal samples which are usable for the entire collaboration the production strategy needs to be designed with the goal of producing enough statistics for all possible vector-like analyses. In order to achieve this goal, the pair production generation was done imposing a branching ratio of 1/3 for the three

<sup>4</sup>It should be noticed that the mixing between vector-like quarks and the SM quarks is larger for the LH component in singlets, while it is larger for the RH component in doublets. This introduces kinematical differences which can be seen, for instance, in the polarisation of the top quark in VLQ decays.

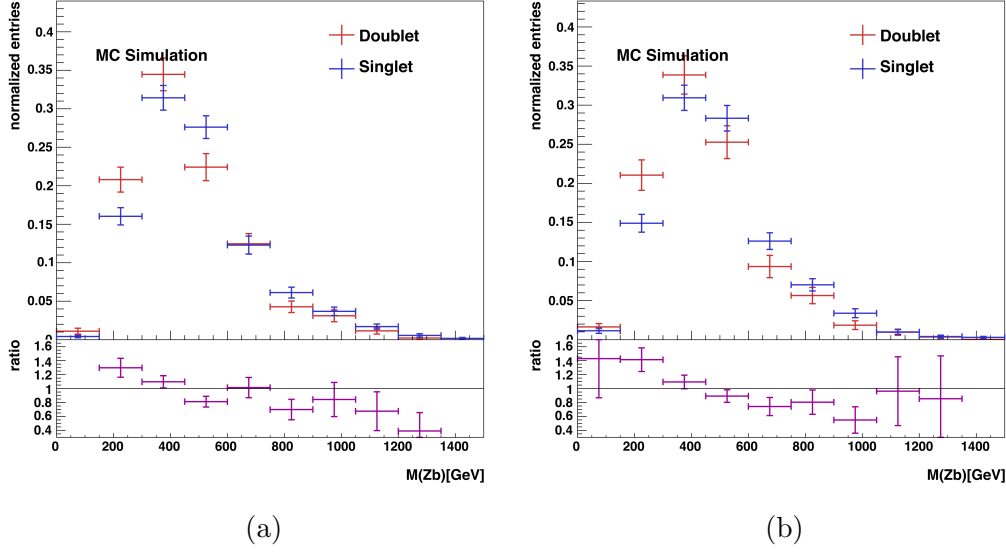


Figure 5.8: Comparisson between the singlet and doublet model prediction for the final discriminant of the di-leptonic channel for the  $ee$  channel (a) and  $\mu\mu$  channel (b). Vector-like  $T$  with a mass of 600 GeV is shown.

decay modes ( $Z$ ,  $W$  and  $H$ ). With this approach, regardless of what decay mode an analysis is more sensitive to, all possibilities are fairly covered. In order to get a given set of branching-ratios the analysis distributions just need to be corrected properly. The singlet and doublet models are the two benchmark models for the pair production analysis. Since it has been shown that no significant kinematical difference is expected between singlet and doublet, the only difference left is the different branching-ratios expected in both cases. The singlet model is used and the contribution of each decay channel to the final discriminant is reweighted to get the proper BR when reporting results for the doublet.

Particle level information is used to recognise which decay mode has been generated in that particular and a reweight factor is defined to obtain the desired BR. For a given set of branching ratios for  $Z$ ,  $W$  and  $H$  decay modes, the weight of the event is corrected by a factor  $\omega$  defined as

$$\omega = \frac{B_1 B_2}{1/9} \quad (5.1)$$

where  $B_1$ ,  $B_2$  are the desired branching ratios for the two decay modes present in the event. Once the event is reweighted, the relative proportion of each decay mode will be the one expected for the branching ratios desired.

All samples were passed through a simulation of the ATLAS detector that em-

employs a fast simulation of the response of the calorimeters [120]. Additional samples with quark masses of 400 GeV, 600 GeV, and 800 GeV were also produced using the standard **GEANT** v4 [121] based simulation of all the detector components, to test the agreement. Figure 5.10 shows the comparison between both detector simulations for the  $ee$  (top) and  $\mu\mu$  (bottom) channels using a vector-like  $T$  with  $m(T) = 800$  GeV. The main differences that might appear between fast and full detector simulation are found in distributions related with jets structure precisely because the calorimeter simulation is simplified in fast detector simulation. The jet multiplicity and the  $H_T$  (jets) distributions are shown in Figure 5.10(a), (d),(b) and (e). These are the distributions most related to jets used in this analysis and as can be seen no significant difference is observed. The comparison in the final discriminant is shown in Figure 5.10(c) and (f) showing that, as expected, both distributions are compatible.

### 5.3.2 Single production

As it was discussed in Chapter 2, as the vector-like quark mass increases the single production mechanism becomes dominant. This implies that, as lower masses are excluded, the single production gains relevance thus making this production mechanism of interest in the search for vector-like quarks.

The dashed lines in Figures 5.7 are the cross sections for the processes  $B\bar{b}q$  and  $T\bar{b}q$  for coupling values of  $V_{Tb} = 0.1$  and  $X_{bb} = 0.1$  [41, 44] which are chosen to reflect the magnitude of indirect upper bounds on mixing [41, 56] from precision electroweak data when assuming a single vector-like multiplet. This indirect constraints may be relaxed in realistic CHM where several vector-like multiplet are present [56] which is why several authors [55, 56, 122] have emphasised single production as a preferred discovery channel. The dotted line in Figure 5.7 represents the cross-section for the process  $T\bar{b}q$  process using the CHM presented in [122] in which the coupling is parameterised with  $\lambda_T$  which is related to the Yukawa coupling and the degree of compositeness of the third-generation SM quarks. The value  $\lambda_T = 2$  is used as a benchmark.

Using the model provided by the authors, fast-simulation samples for the process  $T\bar{t}q$  have been generated using **MadGraph** 5. Masses of the vector-like quark ranging from 400 GeV up to 1050 GeV have been considered setting the coupling to  $\lambda_T = 2$ . In order to test the dependence of the experimental acceptance to different  $\lambda_T$  values, samples with a vector-like mass of 700 GeV were generated for values of  $\lambda_T$  from 1 to 5. Different values of  $\lambda_T$  produce different VLQ width but no differences in the

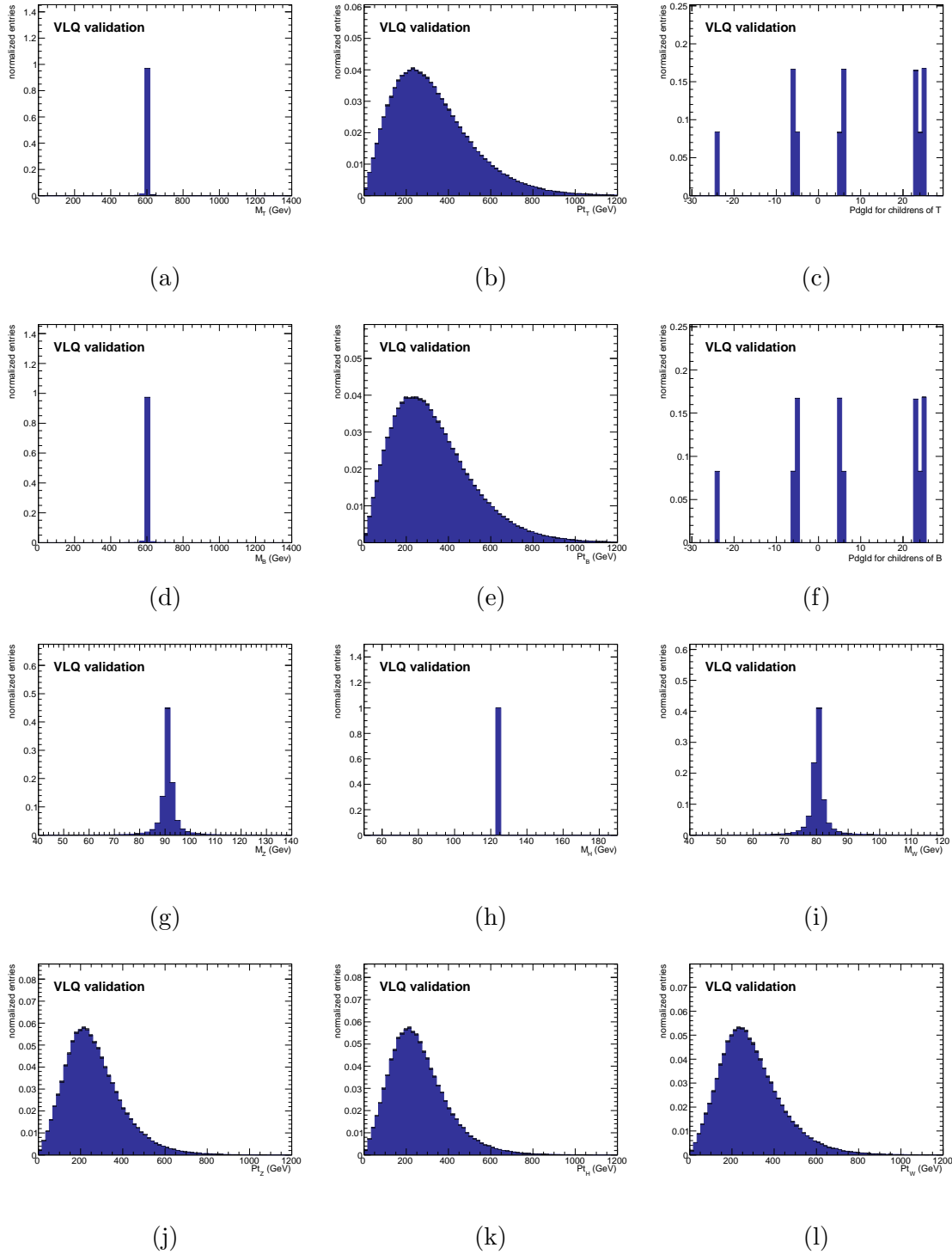


Figure 5.9: Parton level distributions for the VLQ pair-production samples. The first and second row shows the mass,  $p_T$  and pdg-id of the decays of the  $T$  and  $B$  quark respectively. The mass ( $p_T$ ) of the  $Z$ ,  $H$  and  $W$  boson are shown in the third (fourth) row. Vector-like quarks with a mass of 600 GeV have been generated.

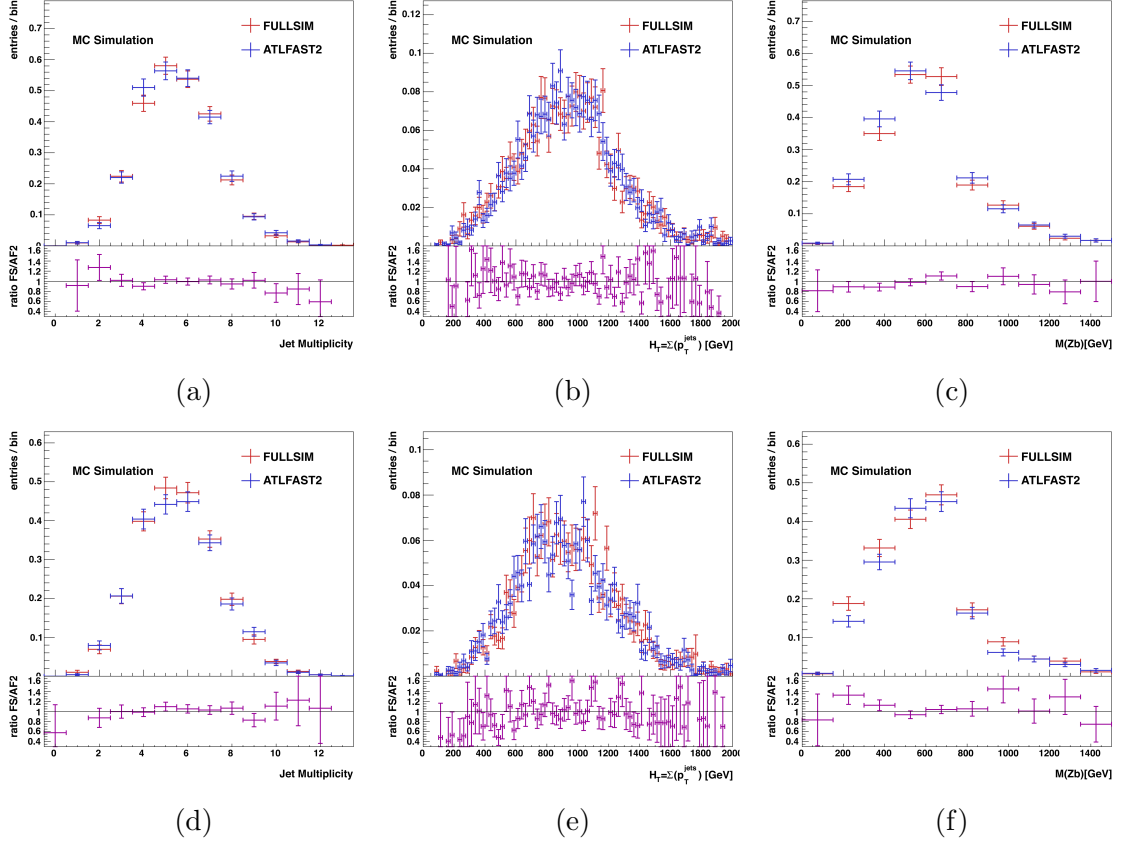


Figure 5.10: Comparison between fast and full detector simulation using a vector-like  $T$  with  $m(T) = 800$  GeV for the  $ee$  (top) and  $\mu\mu$  (bottom) channels. The jet multiplicity distribution is shown after requiring at least one  $Z$  boson candidate (a and d), the  $H_T(\text{jets})$  distribution is shown after requiring  $\geq 2$  jets,  $\geq 2$   $b$ -jets and  $p_T(Z) \geq 150$  GeV (b and e). The  $m(Zb)$  distribution is shown after the final selection (c and f).

analysis acceptance was found when the  $\lambda_T$  parameter was varied, being a variation in the total cross-section the only effect introduced by this variation. The  $B\bar{b}q$  process is absent on some CHM [55, 122]. For this analysis, fully simulated samples for the process  $B\bar{b}q$  have been generated with PROTOS for the  $SU(2)$  singlet for masses ranging from 400 GeV up to 1200 GeV in steps of 50 GeV and a coupling value of  $X_{bB} = 0.1$ . Both generators were compared and no significant differences were found within uncertainties.

## 5.4 Background modelling

Sizeable contribution from SM processes is expected from processes with high jet and  $b$ -jet multiplicity, multileptonic topologies and processes containing a  $Z$  boson.

The background processes are predicted using MC samples normalised to the NLO cross-section. Unless stated otherwise, all samples have been passed through a full detector simulation.

The main background in the dilepton channel analysis are processes including a  $Z$  boson and additional jets ( $Z$ +jets). Two leading-order MC generators have been used to generate this background, **ALPGEN** [123] and **SHERPA** [124]. **SHERPA** is the default generator which has been used and **SHERPA** has been used as a cross-check to ensure a good description of the main background in the dilepton channel. Both samples have been normalised to the NNLO cross-section prediction calculated with the **DYNNLO** program [125].

The  $Z$ +jets **ALPGEN** samples have been generated using the v2.13 version with the CTEQ6L1 [126] PDFs set and interfaced to **PYTHIA** v6.426 for parton-shower and hadronization. Separate inclusive  $Z$ +jets and dedicated  $Z + bb$  and  $Z + cc$  have been generated. In the inclusive samples, heavy-flavour quarks arise from the parton showering while in the dedicated samples they can be produced in the matrix element. In the case where heavy-flavour quarks are produced in the matrix element a matching algorithm [127] is applied to avoid the double counting due to the same parton configurations being produced in the matrix element and in the parton showering. Another double counting might occur when combining both the inclusive and the dedicated sample, in this case, another algorithm based on the angular separation between the heavy-flavour quarks is used. The matrix element generation is preferred if  $\Delta R(h_h, \bar{q}_h) < 0.4$ , where  $q_h = c, b$ , and the parton-showering generation is preferred otherwise.

The **SHERPA**  $Z$ +jets background has been produced with the v1.4.1 version using the CT10 [128] PDFs set setting the charm and bottom quarks to be massive. Different filters have been applied to divide the generation in samples with a bottom quark, a charm quark and no bottom quark and neither a charm nor a bottom quark present in the event. In the following, when referring to  $Z$ +bottom the filter used requires the presence of a bottom quark and when referring to  $Z$ +light the two last filters will be used (i.e. those without a bottom quark). Each heavy-flavour filtered sample of the **SHERPA** generation has also been divided in several sliced samples based on the  $p_T$  of the  $Z$  boson ( $p_T(Z)$ ): inclusive, 70-140 GeV, 140-280 GeV, 280-500 GeV, and greater than 500 GeV. Since the analysis will be focused on high- $p_T$  objects and particularly a high- $p_T$   $Z$  boson will be required, the different slices in which the sample is divided provides higher statistic for high values of  $p_T(Z)$  which



is the main reason this background has been chosen as the default one.

In the trilepton channel, the main source of background processes are diboson processes ( $ZZ$ ,  $ZW$ ,  $WW$ ) and processes with a  $Z$  boson produced in association with a top quark pair ( $t\bar{t} + V$ , where  $V = Z, W$ ). The diboson processes are the dominant background in the early stages of the analysis while the  $t\bar{t} + V$  processes dominate the later stages. The diboson samples have been generated with **SHERPA** using the CT10 PDFs set and normalized to NLO cross-section predictions obtained with **MCFM** [129]. The  $t\bar{t} + V$  samples are generated with **MadGraph** using the CTEQ6L1 PDFs set and **PYTHIA** for parton-shower and hadronization. These samples are also normalized to NLO cross-section predictions [130].

Background sources which doesn't contain a  $Z$  boson candidate constitute sub-leading background processes in the analysis. Simulated  $t\bar{t}$  events are produced with **POWHEG** [131–134] for the matrix method generation using the CT10 PDF set. The parton showering and hadronisation has been done with **PYTHIA** v6.421. The  $t\bar{t}$  cross section has been evaluated with **TOP++** as for the pair production signal samples setting the top quark mass to 172.5 GeV. Single-top production has also been considered being produced via  $s$ - and  $t$ -channel as well as in association with a  $W$  boson ( $Wt$  processes). The **mc@nlo** [135, 136] samples interfaced to **HERWIG** 6.520.2 [137–139] for the parton showering and hadronisation has been used for the  $Wt$  and  $s$ -channel processes while **ACERMC** [140] interfaced with **PYTHIA** is used for the  $t$ -channel process. The single-top processes are normalized to NLO cross-section predictions [141].

Processes in which events with no  $Z$  boson candidate enter the event selection due to a non-prompt or fake lepton passing the lepton selection criteria are estimated with data relaxing or reverting certain lepton identification requirement. These events are less than 5% of the total background estimation at early stages of the analysis and negligible after the final selection is applied.

In the following, the dilepton channel sources of background processes are shown for the  $Z$ +jets processes as explained before, for the  $t\bar{t}$  process and a category defined as *Other bkg.* contains contribution from single-top, diboson processes and  $t\bar{t} + V$  processes. For the trilepton channel distributions and yield tables the  $WZ$  and  $t\bar{t} + V$  processes will be shown while the subleading backgrounds ( $Z$ +jets, diboson  $ZZ$  and  $t\bar{t}$ ) are grouped into the *Other bkg.* category.

## 5.5 Data and Monte Carlo prediction comparison

In the previous session the selection criteria for the different channels proposed in this analysis was presented. This section shows the agreement between the  $\sqrt{s} = 8$  TeV data analysed and the MC prediction for the different processes considered in this analysis. Unless stated otherwise only the **SHERPA**  $Z$ +jets sample will be shown. Appendix B shows **ALPGEN** distributions used to compare both background samples for the dilepton channel in pair production. Under the scope of this thesis the analysis for the dilepton channel for pair production has been developed as well as the New Physics Analysis package (NPA) and software tools that have been used in the other channels of the analysis as well as by others analyses in the collaboration (See Appendix C). The di-lepton and tri-lepton channel for the pair production hypothesis is presented in Section 5.5.1 and Section 5.5.2 respectively. Section 5.5.3 describes the event selection modifications introduced for the single production mechanism.

### 5.5.1 Dilepton channel for the pair production hypothesis

In the dilepton channel exactly two same-flavour opposite-sign leptons are selected. These leptons are used to define the  $Z$  boson candidate which is required to satisfy  $|m_{\ell\ell} - m_Z| < 10$  GeV. At this stage of the analysis  $\sim 12.5 \times 10^6$  data events are selected which is in agreement with the MC prediction ( $12.1 \times 10^6$ ) within uncertainties dominated at this stage by luminosity uncertainties and lepton reconstruction systematic uncertainties. Events with at least two central jets are selected and divided into three channels based on  $N_{\text{tag}}$ . Events with  $N_{\text{tag}} \geq 2$  define the signal region while events with  $N_{\text{tag}} = 0$  and  $N_{\text{tag}} = 1$  define two control regions used for background corrections.

The main background for the dileptonic channel is the  $Z$ +jets background. For this background the  $N_{\text{tag}} = 0$  region  $Z + \text{light}$  will dominate and  $Z + \text{bottom}$  will become dominant as the number of  $b$ -tagged jets required increases. This makes this selection sensitive to the heavy-flavour composition of the  $Z$ +jets sample. In order to assess the correct heavy-flavour composition of the  $Z$ +jets background the **JetFitter** variable was studied. The **JetFitter** variable is a tagger, similar to the MV1 tagger used in this analysis to define  $b$ -jets, which produces higher values for jets more likely to be tagged as  $b$ -jets. The distribution of the sum of the **JetFitter** value of the two jets with higher MV1 will group events with more likelihood to contain  $b$ -jets in the higher values of the distributions while events with less  $b$ -jets will populate the lower tails of the distribution. Using this procedure to

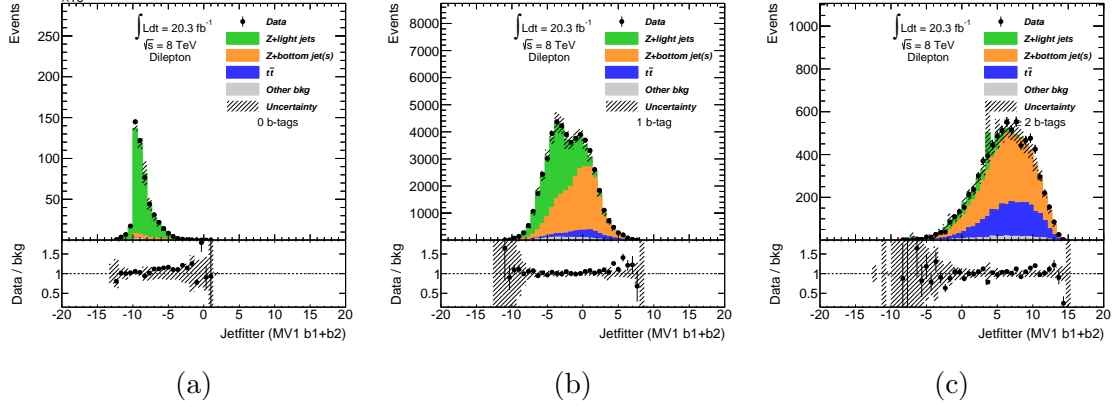


Figure 5.11: The sum of the **JetFitter** value of the two highest MV1 jets is shown for the  $N_{\text{tag}} = 0$  (a),  $N_{\text{tag}} = 1$  (b) and  $N_{\text{tag}} \geq 2$  (c).

separate heavy-flavour composition (shown in Figure 5.11), it can be seen that the default heavy-flavour composition for the **SHERPA** generator is well modelled within uncertainties.

The MC prediction when using the **SHERPA**  $Z$ +jets sample agrees with data within uncertainties while differences are observed when using the **ALPGEN** sample for which the MC prediction is low by a  $\sim 20\%$  and  $\sim 15\%$  for the  $N_{\text{tag}} = 1$  and  $N_{\text{tag}} \geq 2$  categories respectively. Checking a  $Z$ +jets depleted region, i.e. outside the 10 GeV window around the  $Z$  mass peak and requiring  $N_{\text{tag}} \geq 2$ , where the  $t\bar{t}$  background is dominant, a good agreement is found between data and MC which implies that the **ALPGEN** sample underestimate the  $Z$ +jets predictions. This differences justify a scaling applied to the  $Z$ +jets background in order to match the data expectation in a signal depleted region for each  $N_{\text{tag}}$  category. The correction is done to the  $Z$ +jets overall prediction since, as shown before, the default heavy-flavour composition is well modelled. Events which satisfy  $p_{\text{T}}(Z) < 100$  GeV are selected and a weight value is defined as

$$\omega_{N_{\text{tag}}} = \frac{N_{\text{data}} - N_{\text{other}}}{N_{Z+\text{jets}}} \quad (5.2)$$

where  $N_{\text{data}}$  is the number of data events,  $N_{\text{other}}$  is the number of events from non- $Z$ +jets background processes and  $N_{Z+\text{jets}}$  is the number of events from  $Z$ +jets background. Figure 5.12 shows the correction value applied to the  $ee$  and  $\mu\mu$  channel separately. It can be noticed how the **SHERPA** sample is almost uncorrected nonetheless the correction is also applied to correct in the same way both generators.

Two uncertainties are associated to this correction and are included in the sys-

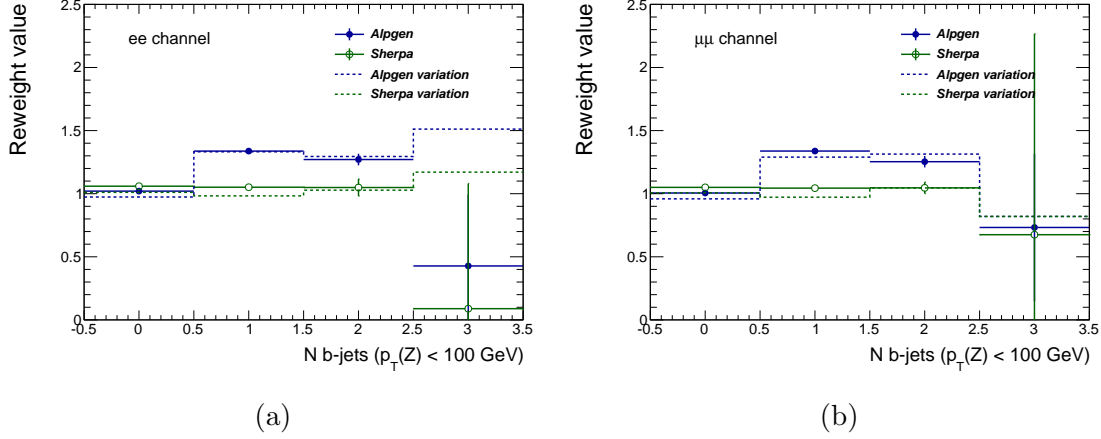


Figure 5.12: Correction values applied to the  $Z$ +jets background samples for each  $N_{\text{tag}}$  category. Correction values for SHERPA and ALPGEN are shown (dots) as well as the reweight value derived in a different control region used as a systematic uncertainty (dashed line). Correction values for electrons (a) and muons (b) are shown.

tematic uncertainties evaluation:

- The statistical uncertainty of each sample is taken into account when deriving the reweight value. It is represented as error bars in the reweight value shown in Figure 5.12. The difference between the nominal reweight value and the reweight increased and decreased by its error is taken as a systematic uncertainty.
- The choice of the control region also has an impact on the value of the correction applied. A different control region, also depleted in signal, is defined with events satisfying  $50 \text{ GeV} \geq p_T(Z) \geq 150 \text{ GeV}$ . The difference between the nominal correction and the correction derived in the alternative control region is used as a systematic uncertainty. The reweight value obtained is shown with dashed lines in Figure 5.12.

Figure 5.13 shows the  $m(Z)$  distribution for the  $N_{\text{tag}} = 1$  region (a) and the  $N_{\text{tag}} \geq 2$  region (b).

After the analysis have been divided in different  $N_{\text{tag}}$  categories the  $p_T$  of the  $Z$  candidate (shown in Figure 5.14) is required to satisfy  $p_T(Z) > 150 \text{ GeV}$ . A trend can be noticed in the background expectation which seems to overestimate high- $p_T$   $Z$  bosons in the  $N_{\text{tag}} = 1$  region. When applying the  $p_T(Z) > 150$  requirement this trend will result in a global overestimation of  $\sim 14\%$  compared with an uncertainty of  $8\%$ . To correct this trend a new correction is applied to the  $Z$ +jets background.

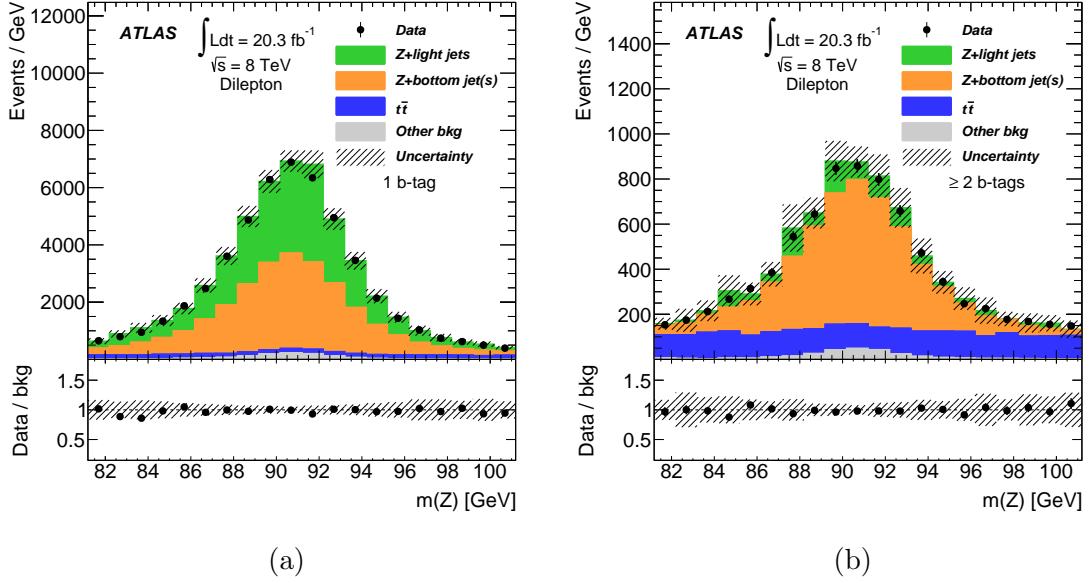


Figure 5.13: Invariant mass distribution for the dilepton channel in the  $N_{\text{tag}} = 1$  region (a) and  $N_{\text{tag}} \geq 2$  region (b).

Using the  $N_{\text{tag}} = 1$  region as the signal depleted region (with  $\sim 10$  signal events and  $\sim 5 \times 10^4$  background events) a weight  $\omega_i$  is defined as

$$\omega_i = \left[ \frac{N_{\text{data}} - N_{\text{other}}}{N_{Z+\text{jets}}} \right]_i, \quad (5.3)$$

for each bin  $i$  of the  $p_T(Z)$  distribution, in the same way as is done for the previous correction. In order to minimise statistical fluctuations, but still be able to apply the correction derived for each bin, a third degree polynomial is used to fit the correction values. Figure 5.15 shows the correction values derived in the  $N_{\text{tag}} = 1$  region for the  $ee$  and  $\mu\mu$  channel separately. Each event is corrected by the value of the fit for the corresponding  $p_T(Z)$  value in the event. For events with a  $p_T(Z)$  value above the fitted range the correction is supposed to be constant above the last fitted value. Table 5.2 shows the expected and observed event yields at different stages of the analysis for the  $N_{\text{tag}} = 1$  control region. Background event yields are shown before and after the  $p_T(Z)$  correction is applied to the  $Z$ +jets background. Table 5.3 shows the predicted and observed event yields for the  $N_{\text{tag}} \geq 2$  signal region at different stages of the analysis with both  $Z$ +jets background corrections applied.

Similarly to the first correction, two uncertainties are defined:

- The statistical uncertainty on the correction values related to the finite size of

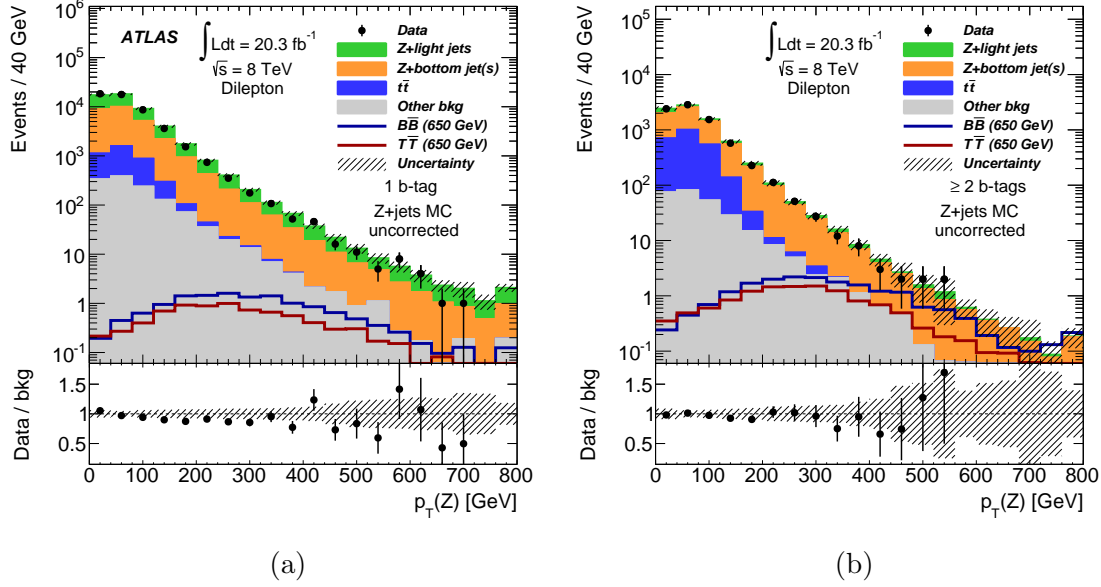


Figure 5.14: The  $p_T(Z)$  distribution is shown for the  $N_{\text{tag}} = 1$  region (a) and the  $N_{\text{tag}} \geq 2$  region (b).

the MC samples. The green shaded region in Figure 5.15 (a) and (b) shows the envelope of the fit and variations up and down for each event are done and assigned as systematic uncertainties to the analysis. The uncertainty of the fit represent the actual error for each point.

- In order to evaluate the effect of the control region chosen to derive the correction, an alternative control region populated by events satisfying that  $N_{\text{tag}} = 0$  is used to re-derive the correction. The difference between both corrections is used as a systematic uncertainty. Figure 5.15 (c) and (d) show the correction derived in this alternative region.

Figure 5.16 shows the  $p_T(Z)$  distribution after the correction is applied.

All the events passing the  $p_T(Z) \geq 150$  GeV requirement are then required to satisfy  $H_T(\text{jets}) \geq 600$  GeV. Figure 5.17 shows the  $H_T(\text{jets})$  distribution for the  $N_{\text{tag}} = 1$  (a) and for  $N_{\text{tag}} \geq 2$  (b). The final discriminant for the di-lepton channel, the invariant mass of the  $Zb$  system, composed by the  $Z$  boson candidate and the highest- $p_T$   $b$ -jet, is shown in Figure 5.17 (c) and (d), for the  $N_{\text{tag}} = 1$  control region and the  $N_{\text{tag}} \geq 2$  final signal region respectively.

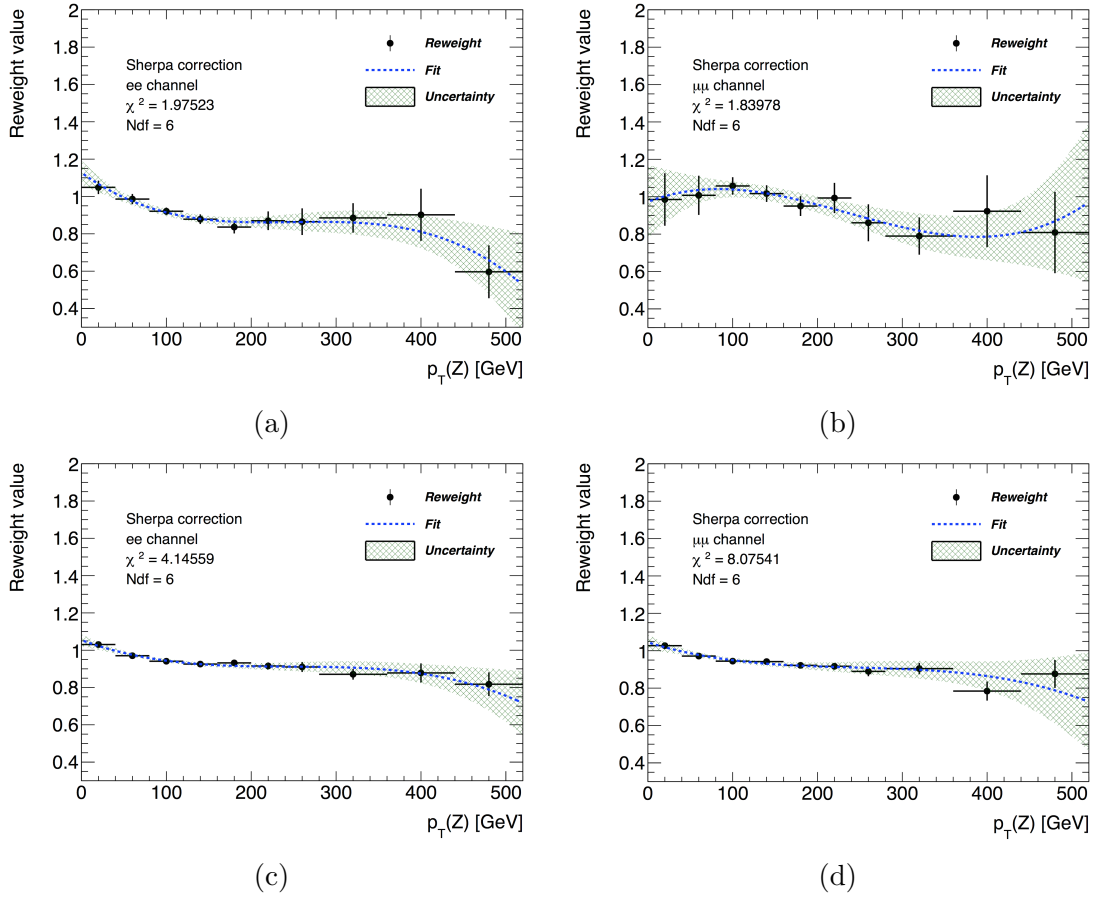


Figure 5.15: Correction values for the  $p_T(Z)$  distribution derived in the  $N_{\text{tag}} = 1$  control region (top) and the  $N_{\text{tag}} = 0$  region (bottom) for the  $ee$  (left) and  $\mu\mu$  (right) channel. A third degree polinomial is used to minimize the impact of statistical fluctuations.

### 5.5.2 Trilepton channel for the pair production hypothesis

The trilepton channel of the analysis targets the vector-like production with three leptons in the final state. This channel has fewer events than the dilepton channel but the relative population of signal events is higher given that trilepton events are less likely to occur in the SM background processes. In particular after selecting a pair of same-flavour opposite-sign leptons with the presence of a third isolated lepton only  $1760$  data events are selected compared with the  $12.5 \times 10^5$  selected events in the dilepton channel. Table 5.4 shows the event yields in the trilepton channel for different stages of the analysis.

Figure 5.18 (a) shows the mass of the  $Z$  boson candidate after requiring  $|m_{\ell\ell} - m_Z| < 10$  GeV and  $N_{\text{lep}} \geq 3$ . In the trilepton channel the dominant background is no longer  $Z$ +jets as it was the case for the dilepton channel since this background

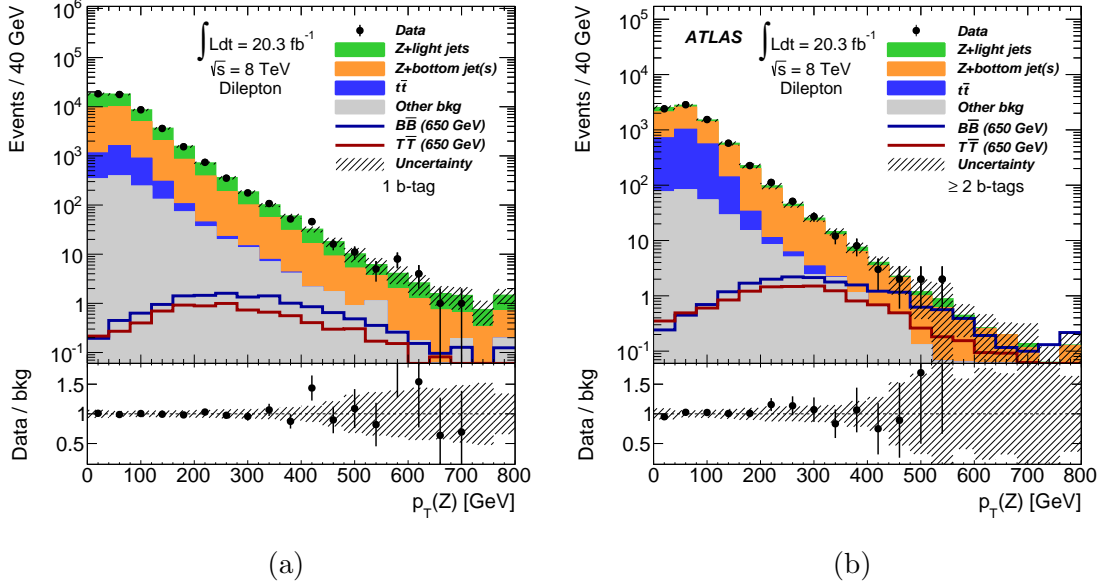


Figure 5.16: The  $p_T(Z)$  distribution is shown for the  $N_{\text{tag}} = 1$  region (a) and the  $N_{\text{tag}} \geq 2$  region (b) after the  $p_T(Z)$  correction is applied.

process will hardly produce more than 2 leptons. For the trilepton channel diboson processes (mostly  $WZ$  and small contribution from  $ZZ$  as well) are the dominant SM background along with  $t\bar{t} + V$  which becomes more important after requiring  $N_{\text{tag}} \geq 1$ . Smaller contribution from  $ZZ$ ,  $Z$ +jets and  $t\bar{t}$  processes are included in the *Other bkg.* category shown in the distribution. Figure 5.18 (b) shows the central jet multiplicity for the trilepton channel before selecting  $\geq 2$  central jets.

Figure 5.19 (a) shows the  $p_T(Z)$  distribution after requiring  $\geq 2$  central jets. Good agreement between data and MC prediction is observed within uncertainties at this stage of the analysis. Events in the trilepton channel are required to satisfy  $p_T(Z) \geq 150$  GeV and then two regions are defined based again in the  $b$ -tagged jet multiplicity. The  $N_{\text{tag}} = 0$  region is used to verify background modelling and the  $N_{\text{tag}} \geq 1$  is used as signal region. Figure 5.19 (b) shows the  $N_{\text{tag}}$  distribution after requiring  $p_T(Z) \geq 150$  GeV.

In the trilepton channel the  $H_T(\text{jets}) \geq 600$  GeV requirement is dropped and the  $H_T(\text{jets} + \text{leptons})$  is used as final discriminant. In this case the  $p_T$  of the leptons is included in the definition of  $H_T$  in order to introduce the information of the third lepton which is not taken into account when selecting events with  $p_T(Z) \geq 150$ . Figure 5.20 shows the  $H_T(\text{jets} + \text{leptons})$  distribution for the  $N_{\text{tag}} = 0$  (a) control region and  $N_{\text{tag}} \geq 1$  (b) signal regions. In the  $N_{\text{tag}} = 0$  control region a good agreement is observed between observed data events and MC prediction.



	$Z + \geq 2 \text{ jets } (N_{\text{tag}} = 1)$	$p_T(Z) > 150 \text{ GeV}$	$H_T(\text{jets}) > 600 \text{ GeV}$
$Z + \text{light}$	$24000 \pm 1500$	$1940 \pm 190$	$104.6 \pm 8.6$
<i>After <math>p_T(Z)</math> corr.</i>	$23600 \pm 1500$	$1700 \pm 150$	$89 \pm 12$
$Z + \text{bottom}$	$24100 \pm 1700$	$1970 \pm 240$	$82.5 \pm 8.0$
<i>After <math>p_T(Z)</math> corr.</i>	$23600 \pm 1700$	$1730 \pm 160$	$71 \pm 11$
$t\bar{t}$	$2850 \pm 230$	$68 \pm 11$	$8.0 \pm 2.9$
Other SM	$1250 \pm 370$	$180 \pm 60$	$17.9 \pm 5.7$
Total SM	$52200 \pm 2300$	$4150 \pm 310$	$213 \pm 13$
<i>After <math>p_T(Z)</math> corr.</i>	$51300 \pm 2300$	$3690 \pm 230$	$186 \pm 16$
Data	51291	3652	171
$B\bar{B}$	$13.6 \pm 1.0$	$11.7 \pm 0.9$	$9.6 \pm 0.8$
$T\bar{T}$	$7.9 \pm 0.5$	$6.5 \pm 0.5$	$5.2 \pm 0.5$

Table 5.2: Predicted and observed number of events in the dilepton channel at different stages of the analysis in the  $N_{\text{tag}} = 1$  category. The  $Z$ +jets predictions, as well as the total background prediction, are shown before and after the  $p_T(Z)$  correction described in the text. Reference  $B\bar{B}$  and  $T\bar{T}$  signal yields are provided for  $m_{B/T} = 650 \text{ GeV}$  and  $SU(2)$  singlet branching ratios. The uncertainties on the predicted yields include statistical and systematic sources.

	$Z + \geq 2 \text{ jets } (N_{\text{tag}} \geq 2)$	$p_T(Z) \geq 150 \text{ GeV}$	$H_T(\text{jets}) \geq 600 \text{ GeV}$
$Z + \text{light}$	$900 \pm 210$	$63 \pm 14$	$4.0 \pm 1.3$
$Z + \text{bottom}$	$4420 \pm 300$	$382 \pm 49$	$19.3 \pm 3.6$
$t\bar{t}$	$2190 \pm 230$	$33.0 \pm 8.0$	$4.6 \pm 1.5$
Other SM	$270 \pm 70$	$42 \pm 11$	$4.0 \pm 1.1$
Total SM	$7780 \pm 440$	$519 \pm 53$	$32.0 \pm 4.2$
Data	7790	542	31
$B\bar{B}$	$18.7 \pm 1.5$	$16.5 \pm 1.4$	$14.2 \pm 1.3$
$T\bar{T}$	$12.1 \pm 0.8$	$10.0 \pm 0.7$	$8.6 \pm 0.7$

Table 5.3: Predicted and observed number of events in the dilepton channel at different stages of the analysis in the  $N_{\text{tag}} \geq 2$  category. The  $Z$ +jets predictions, as well as the total background prediction, are shown before and after the  $p_T(Z)$  correction described in the text. Reference  $B\bar{B}$  and  $T\bar{T}$  signal yields are provided for  $m_{B/T} = 650 \text{ GeV}$  and  $SU(2)$  singlet branching ratios. The uncertainties on the predicted yields include statistical and systematic sources.

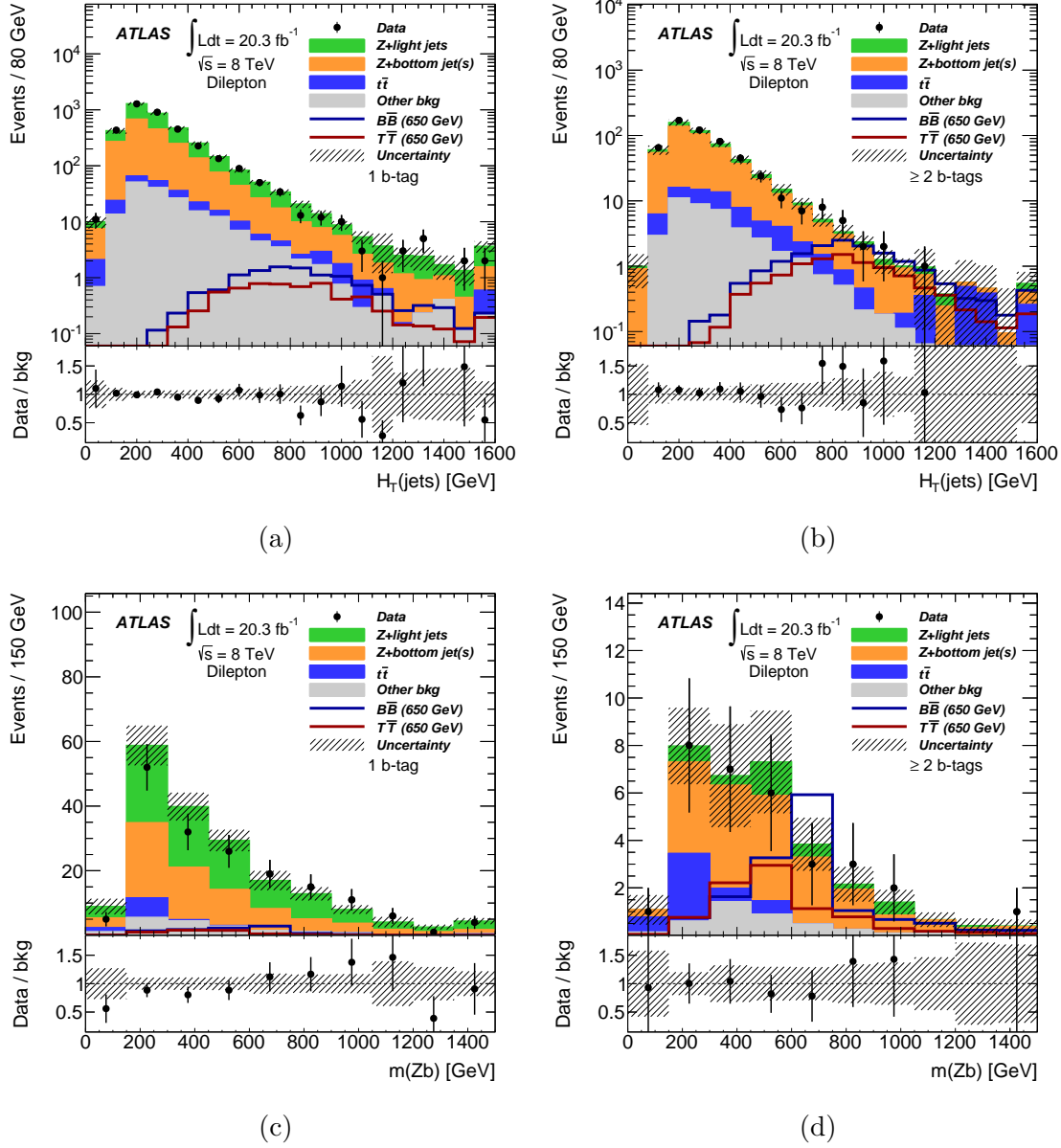


Figure 5.17: The  $H_T(\text{jets})$  (top) and  $m(Zb)$  (bottom) distributions are shown for  $N_{\text{tag}} = 1$  (left) and  $N_{\text{tag}} \geq 2$  (right). The  $m(Zb)$  distribution in the  $N_{\text{tag}} \geq 2$  region is the final discriminant used for hypothesis testing.

### 5.5.3 Event selection modification for single production

The selection criteria for the single production mechanism are the same as for the pair production mechanism with a couple of modifications. A characteristic feature of singly-produced vector-like quarks is the production of an associated energetic forward light-flavour jet. With this consideration, for both the dilepton and tripleton

	Trilepton ch.	$\geq 2$ central jets	$p_T(Z) > 150$ GeV	$N_{\text{tag}} \geq 1$
$WZ$	$1170 \pm 130$	$219 \pm 32$	$51.5 \pm 8.9$	$5.8 \pm 1.4$
$t\bar{t} + V$	$23.5 \pm 6.7$	$22.0 \pm 6.3$	$7.0 \pm 2.1$	$5.8 \pm 1.8$
Other SM	$435 \pm 50$	$67 \pm 13$	$10.4 \pm 9.2$	$2.6 \pm 1.3$
Total SM	$1630 \pm 170$	$309 \pm 39$	$69 \pm 14$	$14.3 \pm 2.6$
Data	1760	334	78	16
$B\bar{B}$	$5.8 \pm 0.4$	$5.7 \pm 0.4$	$4.99 \pm 0.33$	$4.17 \pm 0.30$
$T\bar{T}$	$7.4 \pm 0.5$	$7.4 \pm 0.5$	$6.7 \pm 0.5$	$5.5 \pm 0.4$

Table 5.4: Predicted and observed number of events in the trilepton channel at different stages of the analysis from left to right. The final column represents the signal region for testing the pair production hypotheses. Reference  $B\bar{B}$  and  $T\bar{T}$  signal yields are provided for  $m_{B/T} = 650$  GeV and  $SU(2)$  singlet branching ratios. The uncertainties on the predicted yields include statistical and systematic sources.

channel, at least one forward jet is required when adapting the selection criteria from the pair production to the single production mechanism. In addition, for single production there is less jet activity in the event which implies that the  $H_T$  (jets) distribution is not as efficient as for the pair production discriminating signal from background. Due to this, for the dilepton channel the  $H_T$  (jets)  $\geq 600$  GeV requirement is dropped and instead  $\geq 1$  forward jet is required. Figure 5.21 (a) and (b) show the forward jet multiplicity distribution in the  $N_{\text{tag}} = 1$  control region and  $N_{\text{tag}} \geq 2$  signal region respectively for the dilepton channel. These distributions are shown in the same stage of the analysis as the  $H_T$  (jets) distribution in Figure 5.17. At least one forward jet is required, reducing the background contamination over an order of magnitude while maintaining a large fraction of signal events for the single production hypothesis. Figure 5.21 (c) and (d) shows the invariant mass of the  $Zb$  system after requiring at least one forward jet in both  $N_{\text{tag}} = 1$  control region and  $N_{\text{tag}} \geq 2$  signal region as well.

For the trilepton channel Figure 5.22 shows the forward jet multiplicity in the  $N_{\text{tag}} \geq 1$  signal region (a) and the final discriminant  $H_T$  (jets + leptons) after requiring at least one forward jet (b). The trilepton channel has no sensitivity to the  $B\bar{b}q$  single production process since trilepton topologies are less favorable. Only the  $T\bar{t}q$  process is considered. Table 5.5 shows the observed and expected event yields for the single production hypothesis at different stages of the analysis.

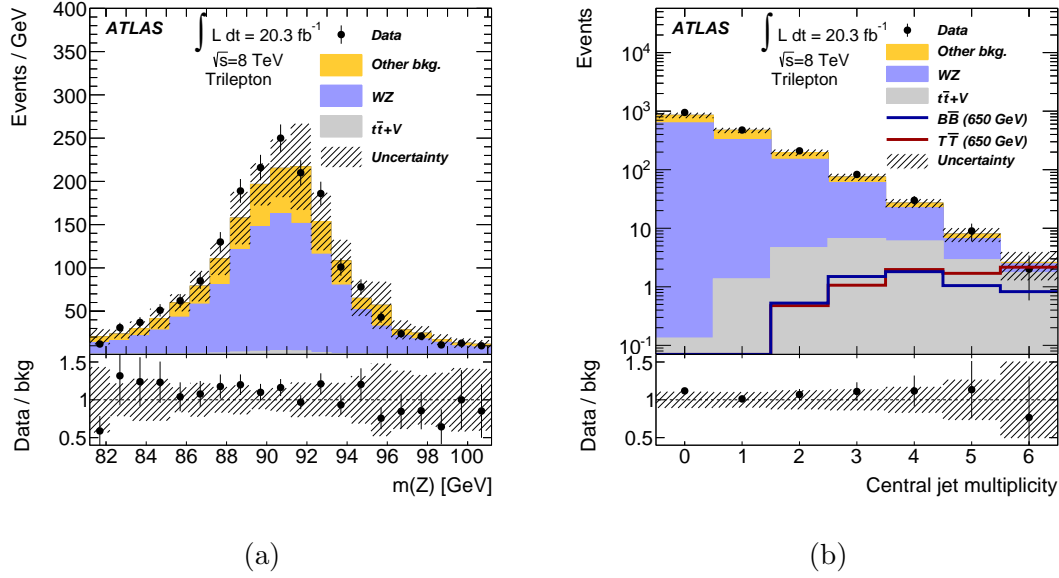


Figure 5.18: The  $m(Z)$  (a) and central jet multiplicity distributions (b) are shown for the tripleton channel selection.

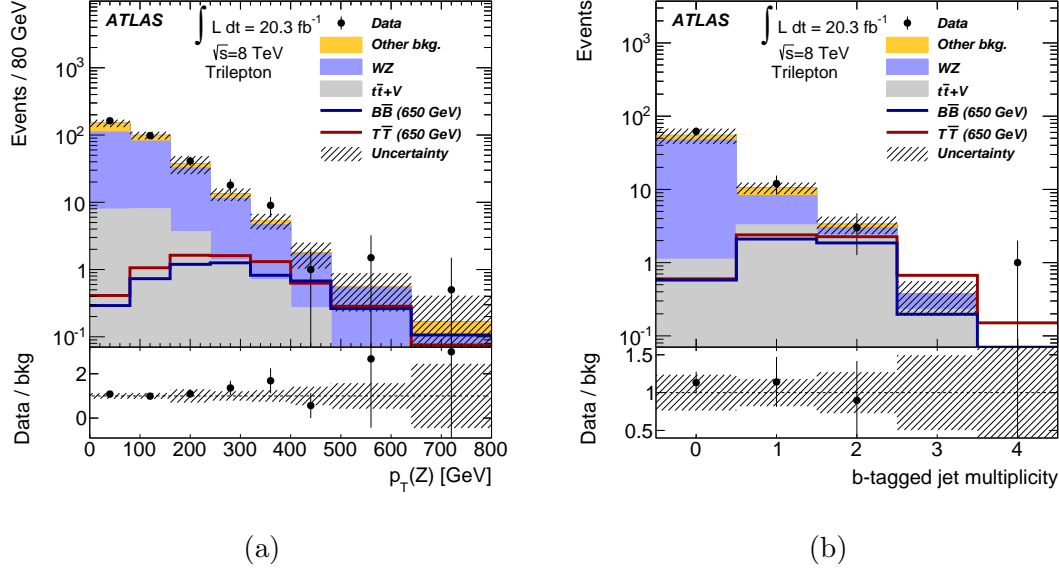


Figure 5.19: The  $p_T(Z)$  (a) and the  $b$ -tagged jet multiplicity distributions (b) are shown for the tripleton channel selection. Events in both distributions are required to have  $\geq 2$  central jets while events in the  $b$ -tagged jet multiplicity distribution also satisfy  $p_T(Z) \geq 150$  GeV.

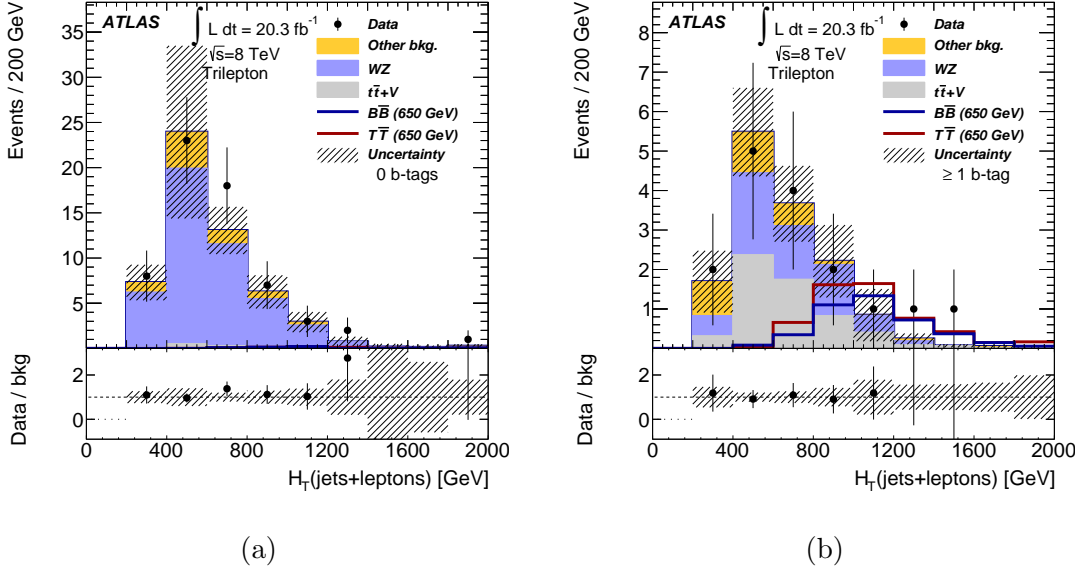


Figure 5.20: The  $p_T(Z)$  (a) and the  $b$ -tagged jet multiplicity distributions (b) are shown for the trilepton channel selection. Events in both distributions are required to have  $\geq 2$  central jets while events in the  $b$ -tagged jet multiplicity distribution also satisfy  $p_T(Z) \geq 150$  GeV.

## 5.6 Systematic uncertainties

Different sources of systematic uncertainty affect the background and signal yields and also the shape of the final discriminants  $m(Zb)$  and  $H_T(\text{jets} + \text{leptons})$ . In order to evaluate this uncertainties, each variation has been applied to the input samples and the full analysis has been performed to each varied sample. The  $Z$ +jets background corrections presented for the di-lepton channel (the scaling derived for each  $N_{\text{tag}}$  region and the correction applied to the  $p_T(Z)$  distribution) have also been re-derived when taking into account each systematic variation. In this section the different sources of systematic uncertainties are described and Table 5.6 and Table 5.7 show the variation in percentage of each source of systematic uncertainty for the dilepton and trilepton channel respectively.

### Luminosity

All background sources are normalised to the integrated luminosity of the observed data,  $\mathcal{L} = 20.3 \text{ fb}^{-1}$ . This luminosity and its uncertainty are estimated using the same procedure as ref. [85] and the uncertainty is measured to be 2.8% which translate into a global normalisation uncertainty for all background and signal samples.

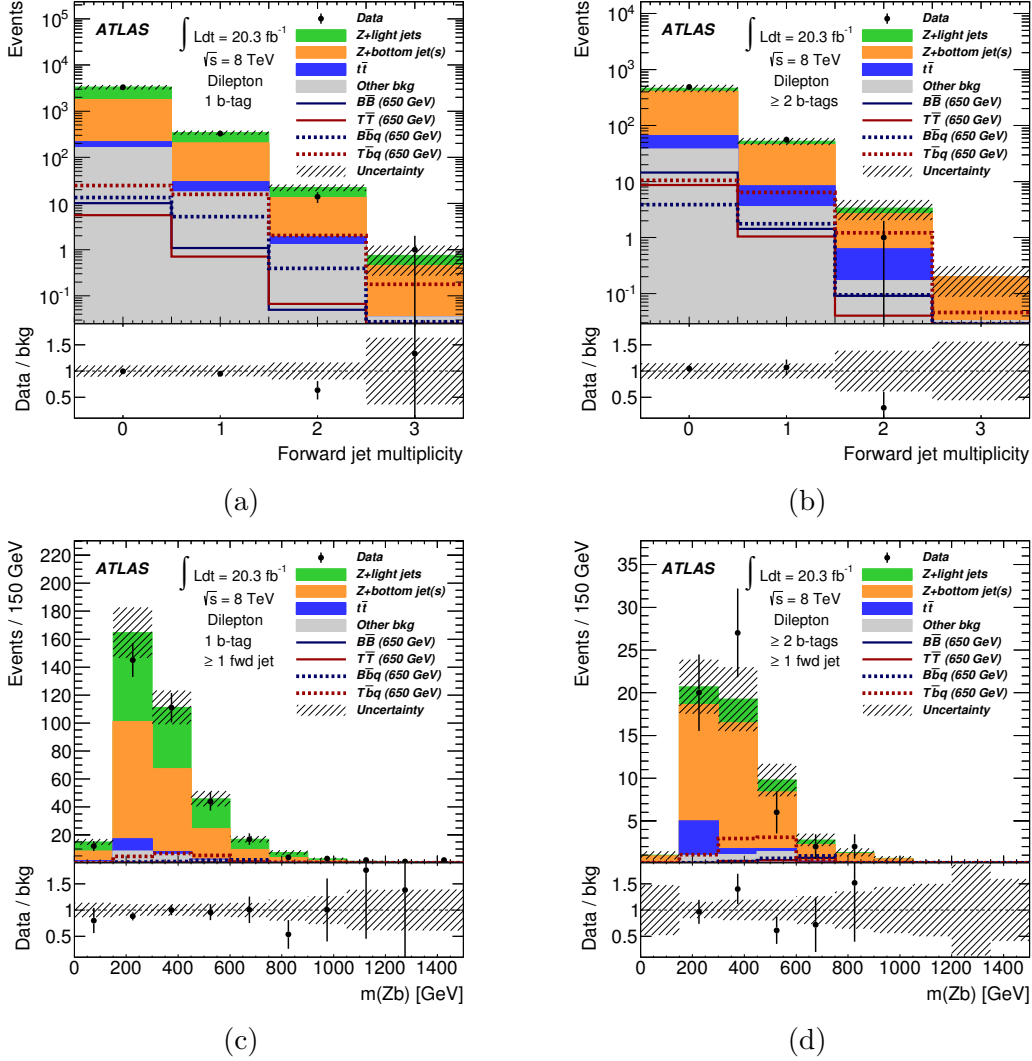


Figure 5.21: The forward jet multiplicity (top) and the invariant mass of the  $Zb$  system (bottom) distributions after requiring  $Z + \geq 2$  central jets,  $p_T(Z) \geq 150$  GeV and  $N_{\text{tag}} = 1$  (left) and  $N_{\text{tag}} \geq 2$  (right) are shown. The predicted  $Tbq$  signal assumes a coupling parameter  $\lambda_T = 2$ , while the predicted  $B\bar{b}q$  signal assumes a coupling parameter  $X_{bB} = 0.5$ .

## Signal and background cross sections

The cross section of each background and signal sample also has a given uncertainty which is taken into account [117, 129, 130, 142]. The cross section uncertainty for the  $Z$ +jets background is not considered since it is been normalised by the background corrections applied. Nonetheless cross-section uncertainties from other backgrounds impact the yields of the  $Z$ +jets background indirectly through these corrections. While for the dilepton channel the impact is small (0.7% uncertainty

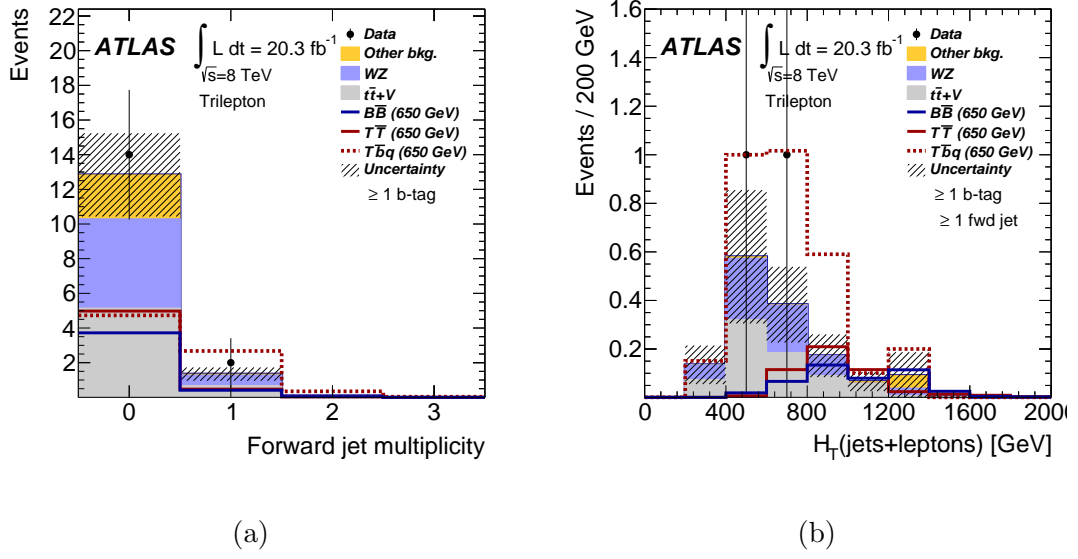


Figure 5.22: The forward jet multiplicity (a) and the  $H_T(\text{jets} + \text{leptons})$  (b) distributions after requiring  $Z + \geq 2$  central jets,  $p_T(Z) \geq 150$  GeV and  $N_{\text{tag}} \geq 1$  are shown. At least one forward jet is required for events in the  $H_T(\text{jets} + \text{leptons})$  distribution as well. The predicted  $T\bar{b}q$  signal assumes a coupling parameter  $\lambda_T = 2$ .

in the total background yields) this uncertainty is the dominant one in the trilepton channel. The uncertainties for the two main background sources in the trilepton channel are the following: for the  $t\bar{t} + V$  processes is conservatively assessed to be 30% [130] and the uncertainty for the  $WZ$  background is estimated as  $50\% \times H_T(\text{jets} + \text{leptons}) / 1 \text{ TeV}$  following ref. [143].

## Jet reconstruction

Different factors affect the jet reconstruction and they impact the final observable of the analysis. Jets are reconstructed from topological energy deposition in the calorimeter and when the energy of the jet is measured a correction scale is applied to the energy (jet energy scale) in order to account for difference between data and MC [110]. This uncertainty varies between 0.8% and 6% for different values of the  $p_T$  and  $\eta$  of the jet under consideration. Additional uncertainties associated with pile-up events can be as large as 5%. An addition uncertainty of up to 2.5% depending on the  $p_T$  of the jet is applied to  $b$ -tagged jets. The jet energy resolution has been measured in di-jets events and agrees with prediction from MC simulation within 10%. To evaluate this uncertainty the jet energy is smeared accordingly. The uncertainty associated with the jet reconstruction efficiency, evaluated in minimum-bias and dijet events, is taken into account by randomly removing simulated jets. In

Dilepton channel		Trilepton channel	
$Z+\text{light}$	$7.3 \pm 2.0$	$WZ$	$0.62 \pm 0.27$
$Z+\text{bottom}$	$40 \pm 10$	$t\bar{t} + V$	$0.74 \pm 0.24$
$t\bar{t}$	$5.2 \pm 2.1$		
Other SM	$3.8 \pm 1.3$	Other SM	$0.07 \pm 0.10$
Total SM	$56 \pm 12$	Total SM	$1.4 \pm 0.4$
Data	57	Data	2
$B\bar{b}q$ ( $X_{bB} = 0.5$ )	$1.88 \pm 0.27$	$T\bar{b}q$ ( $\lambda_T = 2$ )	$3.1 \pm 0.5$
$T\bar{b}q$ ( $\lambda_T = 2$ )	$7.7 \pm 1.0$		
$B\bar{B}$	$1.53 \pm 0.24$	$B\bar{B}$	$0.45 \pm 0.10$
$T\bar{T}$	$1.08 \pm 0.15$	$T\bar{T}$	$0.50 \pm 0.10$

Table 5.5: Number of predicted and observed dilepton and trilepton channel events after the final selection for testing the single-production hypotheses, which includes a forward-jet requirement. The expected yield of  $T\bar{b}q$  and  $B\bar{b}q$  events is listed for  $SU(2)$  singlet  $T$  and  $B$  quarks with a mass of 650 GeV and for reference mixing parameters. The predicted contribution of pair-production events in the single-production signal regions is also provided. The uncertainties on the predicted yields include statistical and systematic sources.

order to minimize the contribution from jets originated in pile-up events an isolation requirement is made on the track energy of the reconstructed central jets. The performance of this requirement was evaluated and compared in data and simulated events with a dileptonic  $Z$  boson and an associated jets, selecting separately events enriched in hard-scatter jets and events enriched in pile-up jets. Correction factors for simulations were derived for both types of events. For hard-scatter jets the correction decreased from  $\sim 1.03$  at  $p_T = 25$  GeV to  $\sim 1.01$  at  $p_T \geq 50$  GeV, while for pile-up jets they are consistent with unity. The uncertainty of this correction is also taken into account. Since in the dilepton channel the requirement  $H_T(\text{jets}) \geq 600$  GeV is applied, the final discriminant is sensitive to these uncertainties. The jet reconstruction uncertainties are the dominant ones in the dilepton channel with an impact of 11% in the final yields for the total background prediction.

## Corrections for b-tagging efficiencies

The  $b$ -tagging algorithms are complex algorithms that provides an easy way to identify if a given jet is tagged as coming from a  $B$  meson. Detailed studies have been performed to derive calibration factors for different operation points providing different  $b$ -tagging efficiencies [112, 113]. Efficiencies for tagging  $b$ -jets ( $c$ -jets) are



corrected by  $p_T$ -dependent factors in the range 0.9–1.0 (0.9–1.1), whereas the light-jet efficiency is corrected by  $p_T$ - and  $\eta$ -dependent factors in the range 1.2–1.5. The uncertainties in these corrections are between 2–6% for  $b$ -jets, 10–15% for  $c$ -jets, and 20–40% for light jets.

## Lepton reconstruction and trigger

In order to correct differences observed between data and MC simulation on the electron and muons reconstruction efficiencies as well as the single-lepton trigger efficiency used in this analysis, different scale-factors are applied to simulated events. The uncertainties on these correction scale-factor are taken into account and, when combined, they contribute to an uncertainty on the final signal and background estimates at the level of 5%. The energy scale and resolution of leptons have been determined using data events containing leptonic  $Z$  bosons and corrections factors have been applied to simulated events in order to correct for differences between data and MC predictions [101–103]. As it was done for jet energy scale and resolution the uncertainties on the lepton energy scale are taken into account repeating the event selection varying the momentum of the lepton accordingly and the uncertainties on the lepton momentum resolution are taken into account smearing the lepton momentum according to the corresponding uncertainty. The impact of the momentum resolution uncertainty was evaluated by smearing the lepton momentum in simulation accordingly. The lepton momentum uncertainties contribute to an uncertainty on the final signal and background estimates at the level of 1%.

## Systematic uncertainties related to $Z$ +jets modelling

Two systematic uncertainties are related with the  $Z$ +jets background corrections and are explained in detail in Section 5.5.1. The first one is related with the statistical uncertainty of the correction values themselves. In order to take this uncertainty into account the complete event selection is passed over the corrected samples varying the correction factor up and down by the statistical uncertainty. The second uncertainty is related to the choice of control region where the correction is derived. For the first correction where the  $N_{\text{tag}}$  multiplicity is rescaled to match data in a control region defined by events with  $p_T(Z) \leq 100$  GeV a different control region is defined as events with  $50 \leq p_T(Z) \leq 150$ . The  $Z$ +jets samples are rescaled using the scale values derived in this region and all the event selection is applied again. The difference between the nominal correction and this new correction is treated as the uncertainty related to the selection of a given control region. The same procedure

Fractional uncertainties (%): dilepton channel						
	$Z$ +jets	$t\bar{t}$	Other bkg.	Total bkg.	$B\bar{B}$	$T\bar{T}$
Luminosity	1.4	2.8	2.8	0.3	2.8	2.8
Cross section	5.5	6.4	29	0.7	-	-
Jet reconstruction	13	10	14	11	2.0	2.1
$b$ -tagging	9.1	13	9.9	5.7	7.2	5.9
$e$ reconstruction	2.9	16	5.9	4.6	2.5	1.5
$\mu$ reconstruction	3.8	7.8	7.2	4.2	3.2	1.3
$Z$ +jets $p_T(Z)$ correction	9.0	-	-	6.5	-	-
$Z$ +jets rate correction	6.9	-	-	5.0	-	-
MC statistics	5.0	25	12	5.4	2.4	2.9

Table 5.6: The fractional uncertainties (%) in the yields of signal and background events after the final dilepton channel selection for testing the pair production hypotheses. The signals correspond to  $SU(2)$  singlet  $T$  and  $B$  quarks with a mass of 650 GeV.

is used to assign an uncertainty to the  $p_T(Z)$  correction. The nominal correction is derived in the  $N_{\text{tag}} = 1$  control region and for the alternative control region  $N_{\text{tag}} = 0$  is used. The difference again is treated as the systematic uncertainty.

Uncertainties related to the factorisation, renormalisation and matching scale as well as the uncertainty related to the PDF set used were also studied for the  $Z$ +jets Monte Carlo generation using **SHERPA**. In order to evaluate these uncertainties dedicated samples were produced varying the factorisation and renormalisation scale by a factor 2 up and down as well as changing the matching scale from the baseline value ( $Q = 20$  GeV) to  $Q = 15$  GeV. These uncertainties were found to be, at most 10% for the di-lepton channel. The PDF uncertainties were found to have an impact of 1% in the most extreme case. Since the uncertainties defined in the  $Z$ +jets corrections already cover the uncertainties related with the background modelling they were not included in the final analysis.

## 5.7 Results

The final discriminant distribution for the dilepton and trilepton channel for pair production are shown in Figures 5.17 (d), 5.20 (b) while the final discriminant for both channel in the case of the single production hypothesis are shown in Figures 5.21 (d) and 5.22 (b). A good agreement is seen between observed data events and expected background simulated events. This agreement is also seen in the final

Fractional uncertainties (%): trilepton channel						
	$WZ$	$t\bar{t} + V$	Other bkg.	Total bkg.	$B\bar{B}$	$T\bar{T}$
Luminosity	2.8	2.8	2.8	2.8	2.8	2.8
Cross section	17	30	8.9	21	-	-
Jet reconstruction	5.4	1.2	8.1	3.1	4.0	1.8
$b$ -tagging	13	3.6	13	6.7	5.6	5.5
$e$ reconstruction	9.3	3.9	37	11	5.9	12
$\mu$ reconstruction	14	3.9	18	4.2	6.2	5.7
MC statistics	11	3.1	27	6.6	4.8	8.3

Table 5.7: The fractional uncertainties (%) in the yields of signal and background events after the final trilepton channel selection for testing the pair production hypotheses. The signals correspond to  $SU(2)$  singlet  $T$  and  $B$  quarks with a mass of 650 GeV.

selection event yields shown in Tables 5.3, 5.4 and 5.5. Since no statistically significant excess is seen in data upper limits are set in the production cross section for pair and single production of vector-like quarks. A 95% confidence level (CL) is used.

The  $CL_s$  method [144, 145] (briefly described in Appendix D) is used to evaluate the compatibility of the observed data with the background-only and signal-plus-background hypotheses. The shape of the final discriminant is used and a binned Poisson likelihood is performed where the statistical test is defined as a log-likelihood ratio function

$$-2 \ln \left( \frac{L_{s+b}}{L_b} \right) \quad (5.4)$$

where  $L_{s+b}$  is the Poisson probability to observe data under the signal+background hypothesis while  $L_b$  is the probability to observe data under the background only hypothesis. Poisson pseudo-experiments are generated for the two hypothesis and the impact of the systematic uncertainties is evaluated in both the normalisation and shape of the final discriminant, which is done when producing the pseudo-experiment varying each bin content accordingly.

### 5.7.1 Limits on the pair production hypothesis

Figure 5.23 (a) and (b) show the cross section upper limit at a 95% CL for the pair production hypothesis of vector-like  $T$  for the singlet and the doublet ( $T, B$ )

Hypothesis	Singlet mass limit [GeV]			Doublet mass limit [GeV]		
	Dilepton	Trilepton	Comb.	Dilepton	Trilepton	Comb.
$B\bar{B}$	690 (665)	610 (610)	685 (670)	765 (750)	540 (530)	755 (755)
$T\bar{T}$	620 (585)	620 (620)	655 (625)	705 (665)	700 (700)	735 (720)

Table 5.8: Observed (expected) 95% CL limits on the  $T$  and  $B$  quark mass (GeV) assuming pair production of  $SU(2)$  singlet and doublet quarks, and using the dilepton and trilepton channels separately, as well as combined.

benchmark models respectively. The upper limit cross section is shown as a function of the mass of the vector-like quark with the uncertainty bands for  $\pm 1\sigma$  and  $\pm 2\sigma$  in green and yellow. The theoretical cross section derived with **Top++** is shown in red with an uncertainty corresponding to  $\text{PDF} + \alpha_s$  and scale uncertainties. The trilepton and dilepton channel are combined for the cross section limit evaluation. The observed (expected) limit on the mass of an  $SU(2)$  singlet  $T$  quark is 655 GeV (625 GeV), while the observed (expected) limit on the mass of a  $T$  quark in a  $(T, B)$  doublet is 735 GeV (720 GeV). Table 5.8 shows the excluded mass for the vector-like  $T$  and  $B$  for the dilepton and trilepton channel separately and assuming singlet and doublet benchmark models.

Figure 5.23 (c) and (d) show the cross section upper limit at a 95% CL for the pair production of vector like  $B$  quark for the singlet and the  $(B, Y)$  doublet model respectively. In this case the  $(B, Y)$  doublet has been used because the  $(T, B)$  doublet only present charged current decays and thus this analysis has no sensitivity to it. The observed (expected) limit on the mass of an  $SU(2)$  singlet  $B$  quark is 685 GeV (670 GeV), while the observed (expected) limit on the mass of a  $B$  quark in a  $(B, Y)$  doublet is 755 GeV (755 GeV). The dilepton and trilepton channel are again combined to evaluate the limits and Table 5.8 shows the lower mass bounds for the dilepton and trilepton channel separately as well as the combination.

It can be noticed (comparing expected limits) how the dilepton channel drives the sensitivity of the analysis for the  $B\bar{B}$  production while it is the trilepton channel which is more sensitive for the  $T\bar{T}$  production. The final discriminant of the dilepton channel, the  $m(Zb)$  distribution, represents the invariant mass of the  $B$  quark while for the case of a  $T$  quark a  $W$  boson is missing from the decay ( $T \rightarrow tZ \rightarrow WbZ$ ). This implies, as can be noticed in Figure 5.17 (d), that the  $m(Zb)$  distribution has more discriminating power when considering  $B\bar{B}$  than when considering  $T\bar{T}$ . The addition of the trilepton channel with a different final discriminant improves the

final result for the  $T\bar{T}$  production.

In addition to lower bounds on the vector-like quarks mass for the singlet and doublet benchmark models a scan in all possible branching ratios for the decay modes ( $Z$ ,  $W$  and  $H$ ) has been performed and lower bounds for the vector-like quark mass are given using the combination of both dilepton and trilepton channel. Figure 5.24 shows the expected (a) and observed (b) lower limit on  $m(T)$  for all possible branching ratios. The branching ratio for  $T \rightarrow Wb$  increases along the  $x$  axis while the branching ratio of  $T \rightarrow Ht$  increases along the  $y$  axis. Since all the branching ratios must sum to unity, the diagonal represent the branching ratio of  $T \rightarrow Zt$ . It can be noticed how the sensitivity increases in the lower left corner where high branching ratios to the  $Z$  decay mode are found reaching a maximum lower bound of 809 GeV (811 GeV) for the observed (expected) limit. Figure 5.25 shows the same results considering vector-like  $B$  instead. The higher sensitivity can be found again for high branching ratios to the  $Z$  decay mode with a maximum observed (expected) limit of 787 GeV (820 GeV).

### 5.7.2 Limits on the single production hypothesis

In the case of the single production hypothesis the cross section upper limits for the  $pp \rightarrow T\bar{b}q$  and  $pp \rightarrow B\bar{b}q$  at a 95% CL are shown in Figure 5.26. The upper limit is multiplied by the branching ratio to the  $Z$  boson decay mode. In the case of the vector-like  $T$  limits are given for the combination of the di-lepton and tri-lepton channel while for the vector-like  $B$  only the di-lepton channel is used since the tri-lepton channel has no sensitivity to this process.

The sensitivity of the analysis to constraints on the coupling parameter  $\lambda_T$  of the chosen composite Higgs model was assessed. In this model, the SM  $t_R$  is a partially-composite state after EWSB. The composite part of  $t_R$  is parametrised by an angle  $\theta$  which is assumed to be small in the definition of the  $\lambda_T$  parameter, which is not the case for small masses and high values of  $\lambda_T$  (discussed in [146]). This implies that for the mass range considered in the analysis values of  $\lambda_T > 1.5$  are not physical. Values of  $\lambda_T < 1.5$  are neither observed nor expected to be excluded by the current sensitivity of the analysis for the single production hypothesis. The sensitivity of the coupling parameters  $V_{Tb}$  and  $X_{Bb}$  has been also assessed for the  $T\bar{b}q$  and  $B\bar{b}q$  processes in the singlet model and no expected sensitivity is found for the  $T$  or  $B$  quark for values of  $V_{Tb} < 1$  and  $X_{Bb} < 0.5$ . Only values of  $V_{Tb}$  as low as 0.7 are observed to be excluded due to a small downward fluctuation in data relative

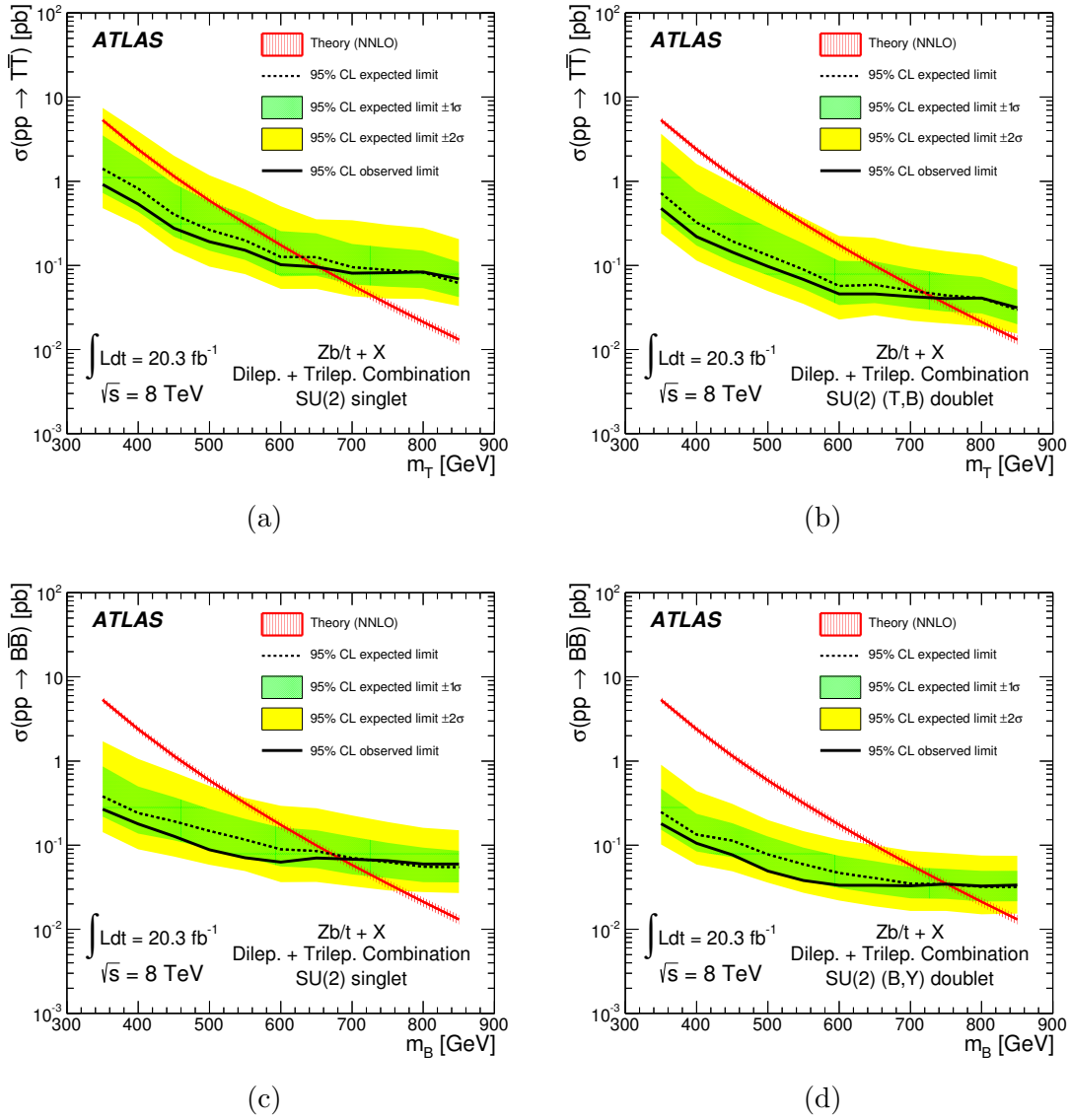
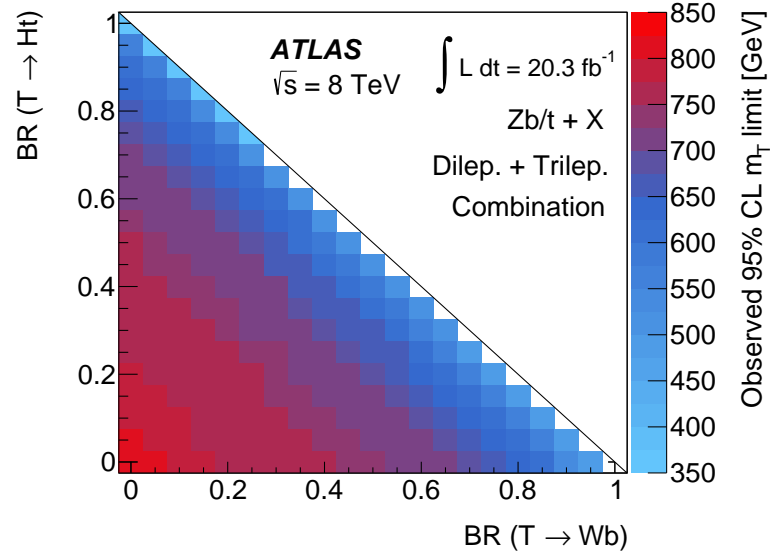
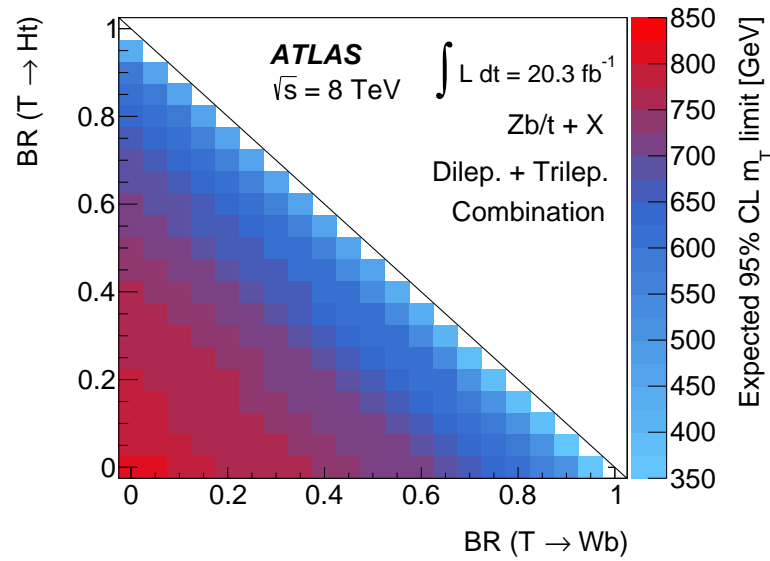


Figure 5.23: Observed and expected upper limit at 95% CL on the cross section for pair production of vector-like  $T$  for the singlet (a) and the  $(T, B)$  doublet (b) and vector-like  $B$  for the singlet (c) and  $(B, Y)$  (d) doublet.

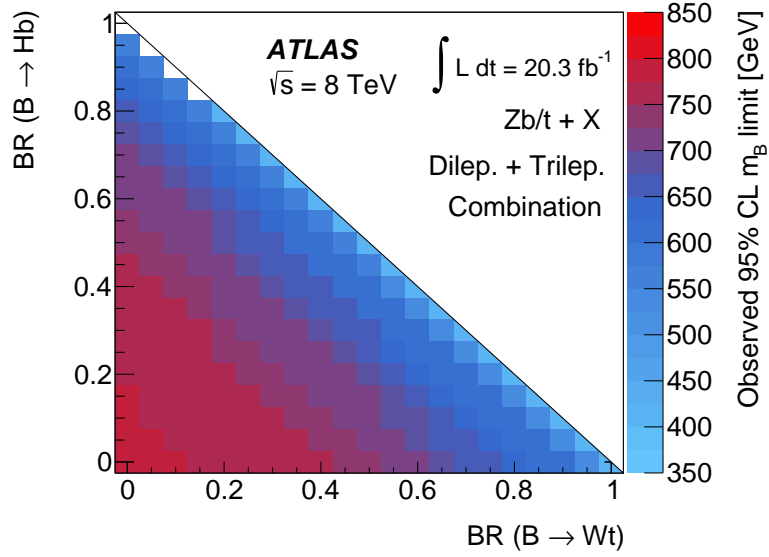


(a)

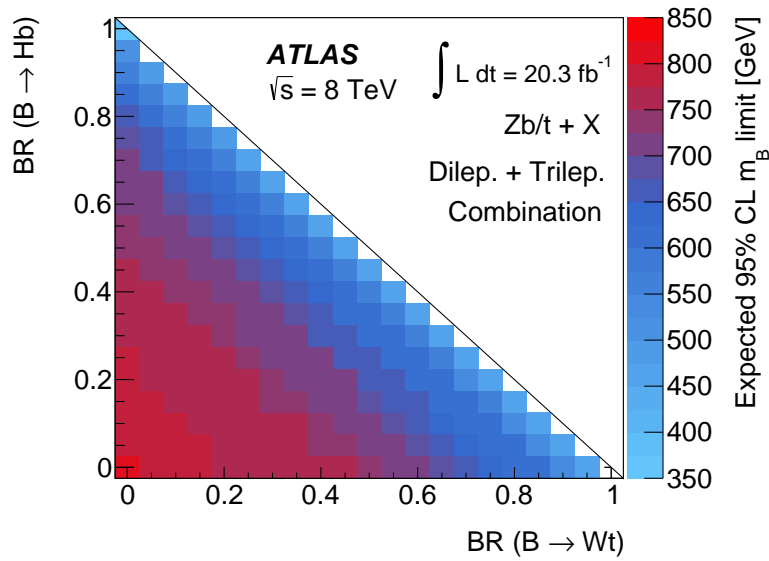


(b)

Figure 5.24: Observed and expected lower mass limit for the mass of the vector-like  $T$  quark at a 95% CL presented as a function of the branching ratios for the  $W$  and  $H$  decay modes.



(a)



(b)

Figure 5.25: Observed and expected lower mass limit for the mass of the vector-like  $B$  quark at a 95% CL presented as a function of the branching ratios for the  $W$  and  $H$  decay modes.



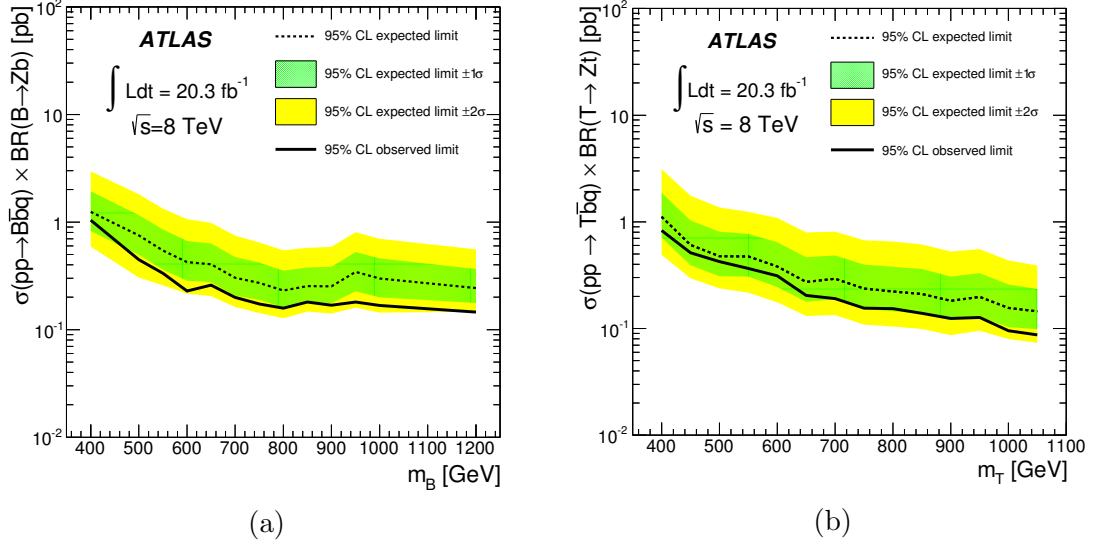


Figure 5.26: Observed and expected upper limit at 95% CL in the cross section for single production of  $B\bar{b}q$  (a) and  $T\bar{b}q$  (b) processes is shown. The upper limit cross section is multiplied by the branching ratio to the  $Z$  decay mode. The benchmark coupling parameters  $\lambda_T = 2$  and  $X_{Bb} = 0.1$  are used for the  $T\bar{b}q$  and  $B\bar{b}q$  processes respectively.

to background in the mass range of 450–650 GeV.

## 5.8 Other searches for VLQ

Several models predict the presence of vector-like quarks which is why both ATLAS and CMS collaborations are committed with several analysis trying to cover as much phase space as possible searching for them. A great job has been done in Run-1 in vector-like quark searches and it is still being done preparing for searches in Run-2 in order not only to increase the reach of the searches but also in providing the results in the most useful format for the community to make use of them.

The analysis presented in this thesis is part of the set of analysis performed by the ATLAS Collaboration searching for both vector-like  $B$  and  $T$  [28, 45–47]. The different analyses are designed to target several final states topologies in order to have greater sensitivity to each corner of the branching-ratio plane (as shown in Figures 5.25 and 5.24). This analysis is the only one sensitive to high branching ratios to the  $Z$  decay channel covering the lower left corner. As part of the work done in the scope of this thesis the results from all analyses in the ATLAS collaboration were taken into account to produce the summary results which represent the most up-to-date results in the search for pair production of vector like quarks scanning all

possible branching-ratios which are shown in Figures 5.27 and 5.28 for vector-like  $T$  and  $B$  respectively. The figures represent the most stringent lower mass limit for each branching ratio taking into account all the published results of the ATLAS Collaboration. As can be seen vector-like  $T$  quarks with masses below  $\sim 740$  GeV are excluded for any branching ratio while masses below  $\sim 940$  GeV are excluded for high branching ratios for the Higgs decay mode. In the case of vector-like  $B$  masses below  $\sim 800$  GeV are excluded for high branching-ratios to the  $Z$  and  $W$  as well as in the limit of  $BR(B \rightarrow Zb) \rightarrow 0$  which are the regions covered by the different analyses. Lower sensitivity is found for high branching-ratios to the Higgs boson decay mode since no dedicated search has been performed. Appendix E shows the results with the different analyses shown in different colors.

The CMS collaboration also has a complete set of vector-like quark searches covering several final states topologies in order to cover all possible branching-ratios. The combined final results [49, 50] on the observed and expected lower bound in the mass of the vector-like  $T$  and  $B$  is shown in Figures 5.29 and 5.30 respectively. CMS excludes vector-like  $T$  with masses below 720 GeV for all possible branching ratios with higher sensitivity for high branching-ratios to the  $W$  decay channel where the lower bound limit goes up to 920 GeV. For the case of vector-like  $B$ , masses below 740 GeV are excluded for all possible branching ratios with high sensitivity to both the  $H$  and  $W$  decay modes where the lower bound increases up to 900 GeV and 850 GeV respectively.

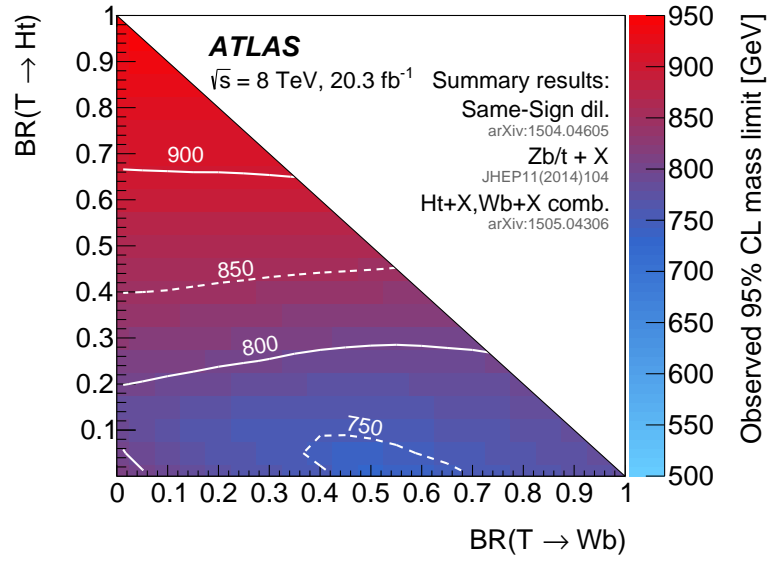
The analysis presented in this thesis was also the first approach to search for the single production mechanism. As higher masses of the vector-like quarks are being excluded single production becomes more important<sup>5</sup>. The ATLAS collaboration recently published a new result focused on singly-produced vector-like  $B$  [48] which shows that the Collaborations are starting to search for the single-production mechanism.

It is important to mention that not only vector-like  $B$  and  $T$  with electric charge  $-1/3$  and  $2/3$  respectively are being searched for. Analyses from ATLAS [46] and CMS [51] have also set limits on the mass of the  $T_{5/3}$  vector-like quark, with an exotic electric charge of  $5/3$ , excluding masses below 840 GeV and 800 GeV respectively.

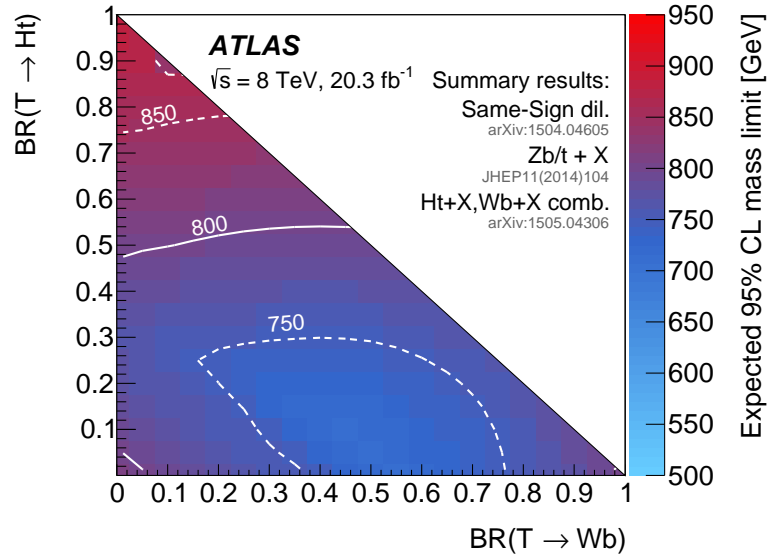
In summary, the analysis presented in this thesis is a key part of the set of

---

<sup>5</sup>At the beginning of Run-2 pair production will be the main production mechanism since the gain in center-of-mass energy makes pair production dominant again at the mass range under consideration.



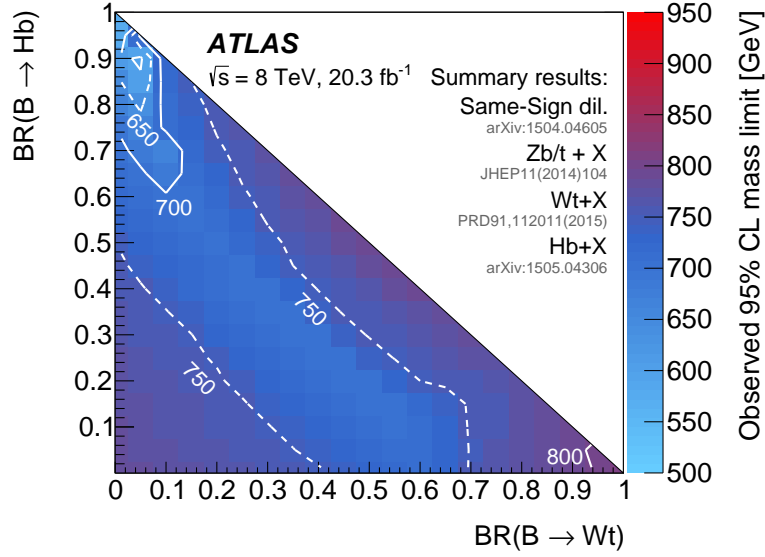
(a)



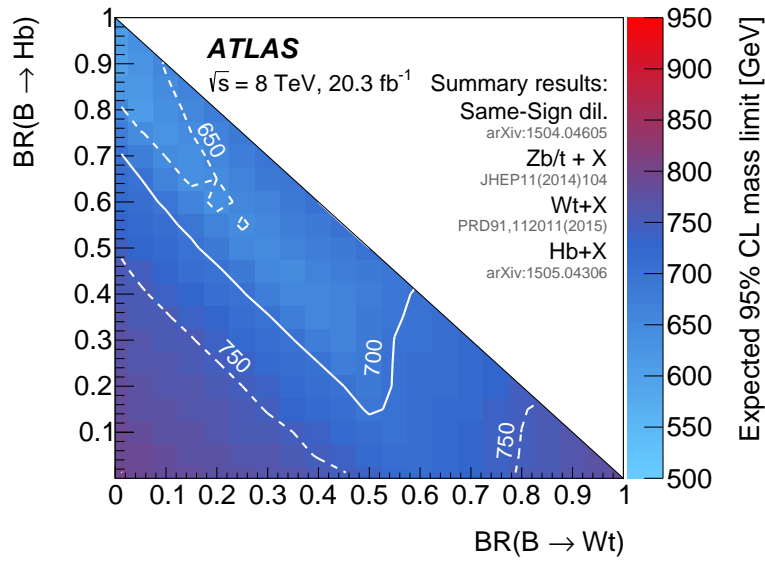
(b)

Figure 5.27: The most stringent observed (a) and expected (b) lower mass limit at 95% CL for pair production vector-like  $T$  using all the published ATLAS results from vector-like quarks searches.

analyses published by the ATLAS Collaboration being the only one which covers the region for high branching ratios to the  $Z$  decay mode and presents the most stringent limits for this channel.

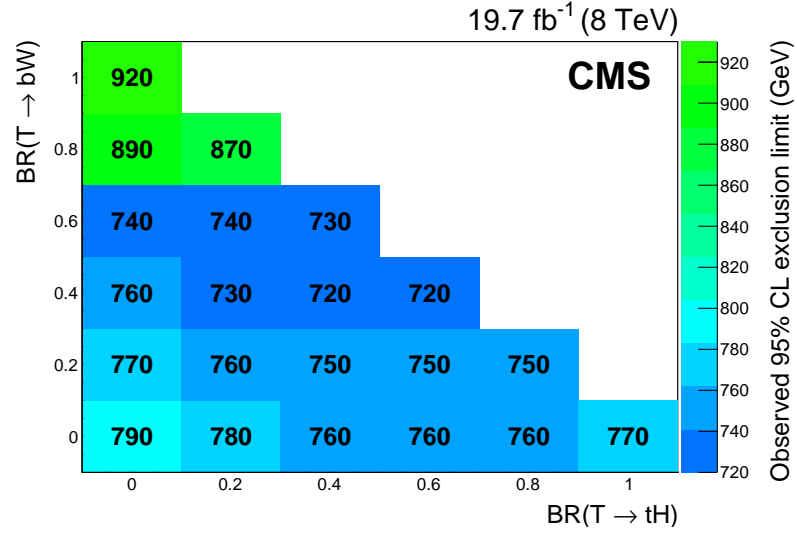


(a)

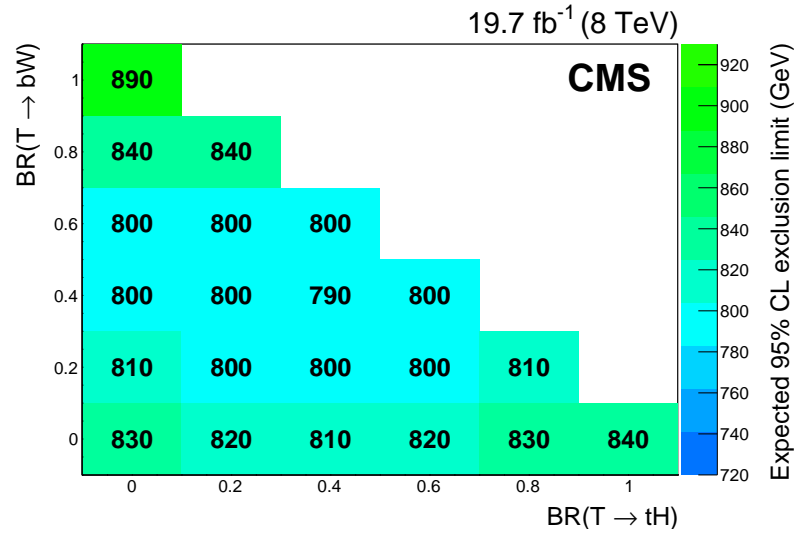


(b)

Figure 5.28: The most stringent observed (a) and expected (b) lower mass limit at 95% CL for pair production vector-like  $B$  using all the published ATLAS results from vector-like quarks searches.



(a)



(b)

Figure 5.29: The most stringent observed (a) and expected (b) lower mass limit at 95% CL for pair production vector-like  $T$  using all the published ATLAS results from vector-like quarks searches [50].

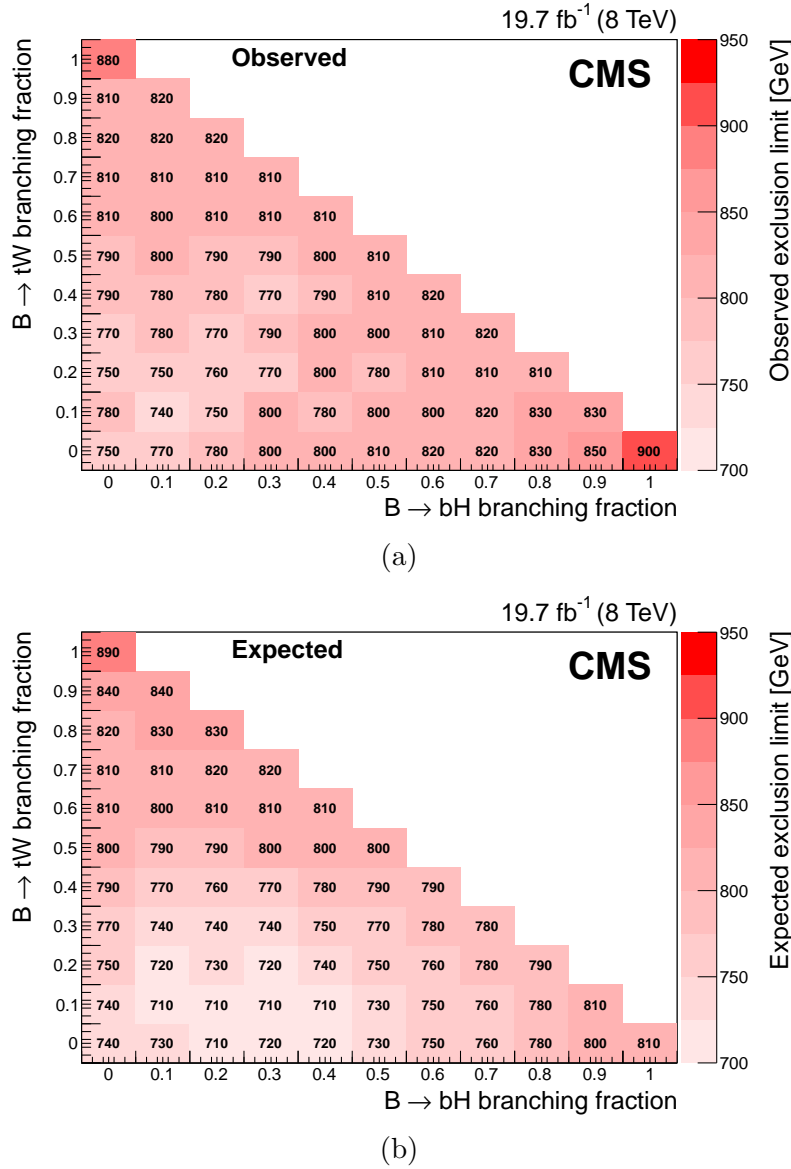


Figure 5.30: The most stringent observed (a) and expected (b) lower mass limit at 95% CL for pair production vector-like  $B$  using all the published ATLAS results from vector-like quarks searches [49].

---

## Chapter 6

# Heavy gluon impact in vector-like quark searches

In Chapter 5 the search for pair and single production of vector-like quarks, with at least one of them decaying to a  $Z$  boson, was presented. Both single and pair production mechanisms were taken into account being the first analysis approaching the single production mechanism. The main concern when presenting the results of a given search is to provide them in a way that the theory community can use, regardless of the particular model being considered. For single production it is particularly challenging because one should take the most model independent approach possible but, at the same time, the production cross section is model dependent since it is the VLQ EW couplings with the gauge bosons what controls the cross section. Different approaches have been discussed, and still are under study, in order to be able to provide results in the most useful way. The best approach seems to report results on the mixing between SM quarks and VLQs ( $\theta_{L,R}$  as shown in Section 2.2.1) instead of the coupling itself.

In the case of pair production it is easier to be rather model independent since the process is mediated via QCD interactions. Nonetheless this might not be entirely the case since a new vector resonance can be present in the model, changing the pair production rate. This is the case of the heavy gluons presented in Section 2.2.2. In this chapter, the impact of a heavy gluon, on the context of CHM, in VLQ searches [147] will be presented. After discussing the details of the CHM model used for this study, the signal and background MC samples which have been generated will be presented. The main goal of this work is to find the possible kinematical differences between the pair production with and without heavy gluon. The

first approach will be to see the behaviour of both production mechanisms at the parton level. After that, a full analysis chain will be introduced taken into account hadronisation and detector simulation. Finally the results will be presented and the limits published by the ATLAS and CMS collaborations will be reinterpreted for the case in which a heavy gluon is included.

## 6.1 Heavy gluon in CHM

For this study the MCHM, in particular the MCHM4<sub>5</sub> presented in Section 2.2.2, with the addition of a heavy gluon, is used. As a reminder, in this model the RH SM top quark is fully composite. The LH components of the SM quarks transform in a 5 representation of  $SO(5)$  while the top partners transform in a 4 representation of  $SO(4)$

$$Q_L^5 = \frac{1}{\sqrt{2}} \begin{pmatrix} ib_L \\ b_L \\ it_L \\ -t_L \\ 0 \end{pmatrix}, \quad \Psi = \frac{1}{\sqrt{2}} \begin{pmatrix} i(B - X_{5/3}) \\ B + X_{5/3} \\ i(T + X_{2/3}) \\ -T + X_{2/3} \end{pmatrix}. \quad (6.1)$$

The relevant part of the lagrangian, in the elementary-composite basis, reads

$$\begin{aligned} \mathcal{L} &= \bar{q}_L i \not{D} q_L + \bar{t}_R i \not{D} t_R + \bar{\Psi} i (\not{D} + i \not{e}) \Psi - M_\Psi \bar{\Psi} \Psi \\ &+ \left[ i c_1 (\bar{\Psi}_R)_i \gamma^\mu d_\mu^i t_R + y f (\bar{Q}_L^5)^I U_{Ii} \Psi_R^i + y c_2 f (\bar{Q}_L^5)^I U_{I5} t_R + \text{h.c.} \right] \\ &- \frac{1}{2} \text{Tr}[G_{\mu\nu}^e]^2 - \frac{1}{2} \text{Tr}[G_{\mu\nu}^c]^2 + \frac{1}{2} M_c^2 \left( G_\mu^c - \frac{g_e}{g_c} G_\mu^e \right)^2. \end{aligned} \quad (6.2)$$

where the elementary and composite gluons,  $G^e$  and  $G^c$  respectively, can be expressed in the physical basis via the rotation

$$\begin{pmatrix} G_\mu^e \\ G_\mu^c \end{pmatrix} = \begin{pmatrix} \cos \theta_3 & -\sin \theta_3 \\ \sin \theta_3 & \cos \theta_3 \end{pmatrix} \begin{pmatrix} g_\mu \\ G_\mu \end{pmatrix}, \quad (6.3)$$

where the rotation angle is related to the ratio between the elementary and composite gluon coupling  $\tan \theta_3 = g_e/g_c$ . After this rotation the massless SM gluon is introduced,  $g_\mu$ , as well as the heavy gluon  $G_\mu$ , whose mass is defined as

$$M_G = \frac{M_c}{\cos \theta_3}. \quad (6.4)$$



For this study the parameters of the model have been set to

$$g_c = 3, \quad f = 800 \text{ GeV}, \quad c_1 = 0.7, \quad c_2 = 1.7, \quad (6.5)$$

where the value of  $g_c$ , which mainly affects the cross-section of the pair-produced VLQs. The value for  $g_c$  has been fixed to 3 to study the effect of the heavy gluon in different kinematics observables and it has been varied in  $2 \leq g_c \leq 5$  to get the bounds on  $M_G$  from current vector-like searches and the early Run-2 expectations (as will be discussed in detail in Section 6.4.2. The parameter  $y$  measures the degree of compositeness of the LH top quark and it is fixed by the top mass. It is sensitive to the value of  $c_2$  with a small dependence on the values of the top partners mass. The top partners masses considered in this study have been chosen to be  $600 \text{ GeV} \leq M_\Psi \leq 1.6 \text{ TeV}$  in steps of 100 GeV. For these mass values the  $y$  parameter ranges from  $y \approx 0.95$  to  $y \approx 0.61$ . The degree of compositeness of the LH top quark, defined as

$$s_L = \frac{yf}{\sqrt{y^2 f^2 + M_\Psi^2}}, \quad (6.6)$$

ranges from  $s_L = 0.78$  and  $s_L = 0.29$ . The masses of the heavy gluons used for this analysis are chosen to be  $1.5 \text{ TeV} \leq M_G \leq 4.5 \text{ TeV}$  in steps of 500 GeV.

A final remark must be made about the model used in this analysis. The large couplings of the heavy gluon with the composite sector and the high number of possible decays translate into a naturally wide heavy gluon [148–150]. Due to the rapid decay of the PDFs, wide heavy gluons will produce top partners decays more similar to the QCD continuum than narrower ones. In order to make the differences, if any, between QCD only and pair production including a heavy gluon more visible, the couplings of the heavy gluon with the composite sector (including the top quark and the LH component of the bottom quark) are scaled down, when necessary, to maintain the heavy gluon width being a 20% of its mass. These scales are done only to emphasise possible kinematic differences; if no kinematical difference is found in this setup it will not be visible when the full width is taken into account.

## 6.2 Signal and background MC generation

In order to generate MC samples of the model presented in the previous section, an Universal Feynrules Output [151] (UFO) mode has been implemented using Feynrules [152]. The MadGraph 5 [119] generator has been used for the parton level generation while Pythia 6 [115] and Delphes 3 [153] have been used for the

hadronisation and fast detector simulation of the ATLAS detector respectively. For the detector simulation the default ATLAS configuration file included in the **Delphes** distribution has been used, defining the  $R$  parameter in the **FastJet** [107] algorithm for the jets and  $b$ -jets to 0.4; the  $b$ -tagging efficiency has been set to 70% and the mistag probabilities for  $c$ -jets and light-quark/gluon jets have been set to 20% and 0.7%, respectively. A total of 300k events have been produced for each pair of masses  $(M_Q, M_G)$ , where  $Q$  runs over all top partners in the model, including all possible decays channels. The MC production has been replicated in three different scenarios: pair production of vector-like quarks via QCD only (labelled as QCD), via heavy gluon only (labelled HG) and considering both QCD and HG production mechanisms, with the proper interference included (labelled QCD+HG).

In order to recast the search presented in Chapter 5 (which will be explained in detail in Section 6.3.3) including the heavy gluon, the two main backgrounds of the analysis have been generated using the same tools as for the signal production.

For the  $Z$ +jets background two matched LO samples of  $Z + bb$  and  $Z + cc$  have been generated with up to three extra jets in the final state. The generation was divided into two subsamples: a sample with  $H_T(\text{jets}) < 300$  GeV and a complementary sample with  $H_T(\text{jets}) > 300$  GeV. This was done to ensure enough statistic in the high  $H_T(\text{jets})$  region. A small statistics sample was generated of  $Z$ +light to test the sensitivity of the analysis. It was found to be negligible (almost no event passed the  $\geq 2b$ -jets selection) and thus it has not been included in the analysis. The full statistics samples for the  $Z$ +jets background contain a total of  $\sim 14$  million events for the  $Z + bb$  sample and  $\sim 3$  million events for the  $Z + cc$  sample.

The  $t\bar{t}$  sample was generated at NLO with **MadGraph 5**, using **MadSpin** [154] to generate the dileptonic decay of the  $t\bar{t}$  events. The hadronisation and detector simulation have been performed with the same tools as before. A total of  $\sim 4$  millions of events have been generated.

Each event generation has been done for two center-of-mass energy configurations (8 and 13 TeV) in order to assess both the actual sensitivity of the analysis and to estimate the sensitivity of early Run-2 analyses to the presence of heavy gluons. Each sample has been normalised to the proper cross section provided by **MadGraph**.

### 6.3 Comparison of kinematical distributions

The purpose of this study is to find if current VLQ analyses would be sensitive to the presence of a heavy gluon in the pair production mechanism. If that would

be the case, dedicated searches would need to be designed in order to fully exploit the differences. Furthermore, if an evidence of signal is found in any top partner search, these deviations from the pair production mediated only by QCD, would be an evidence of heavy gluon as well. On the other hand, if no differences are found the efficiencies of the currently designed and published analyses are likely to be very similar when considering the heavy gluon and therefore, setting bounds on the heavy gluon mass would be straightforward since the only effect would be a rescale in the production cross-section.

First, several kinematical distributions are shown at parton level to evaluate the possible differences without hadronisation or detector simulation effects. After this comparison has been made, hadronisation and detector simulation are introduced to check the impact on each kinematical distribution. The recast of the analysis presented in Chapter 5 is done afterwards to see the impact of the heavy gluon in a realistic scenario.

### 6.3.1 Kinematical differences at parton level

In this section the kinematical differences at the parton level, observed in the pair production of  $X'_{2/3}$  via QCD only and including the heavy gluon in the  $s$ -channel are presented. In order to disentangle the origin of any possible kinematical difference the three production mechanisms (QCD, HG and QCD+HG) will be studied separately. Also, each kinematical distribution is shown for  $\sqrt{s} = 8$  and 13 TeV.

In experimental searches it is common to start looking at objects which are closest to the detector since they are easier to reconstruct. Focusing on the  $Z$  decay of the  $X'_{2/3}$  ( $X'_{2/3} \rightarrow Zt \rightarrow ZW^+b$ ), the  $p_T(Z)$  would be an example since it is rather easy to reconstruct with a pair of same-flavour opposite-sign leptons. Figure 6.1 shows the  $p_T(Z)$  distribution<sup>1</sup> for the masses  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV. These masses have been chosen as benchmark because they are outside the current excluded region and present a sizeable contribution from the HG channel.

Small differences can be observed between QCD and HG but they are almost negligible (even now at the parton level) when comparing QCD and QCD+HG (both at 8 and 13 TeV), which represent the realistic scenario since, if the heavy gluon is present, it will show together with QCD production. The same behaviour has been

---

<sup>1</sup>For all the figures in this section the panels on the left correspond to the distribution properly normalised to its cross section while the panels on the right correspond to unit-normalise distributions to better evaluate the shape differences; the panels on the top show the distribution at  $\sqrt{s} = 8$  TeV while in the bottom  $\sqrt{s} = 13$  TeV is shown.

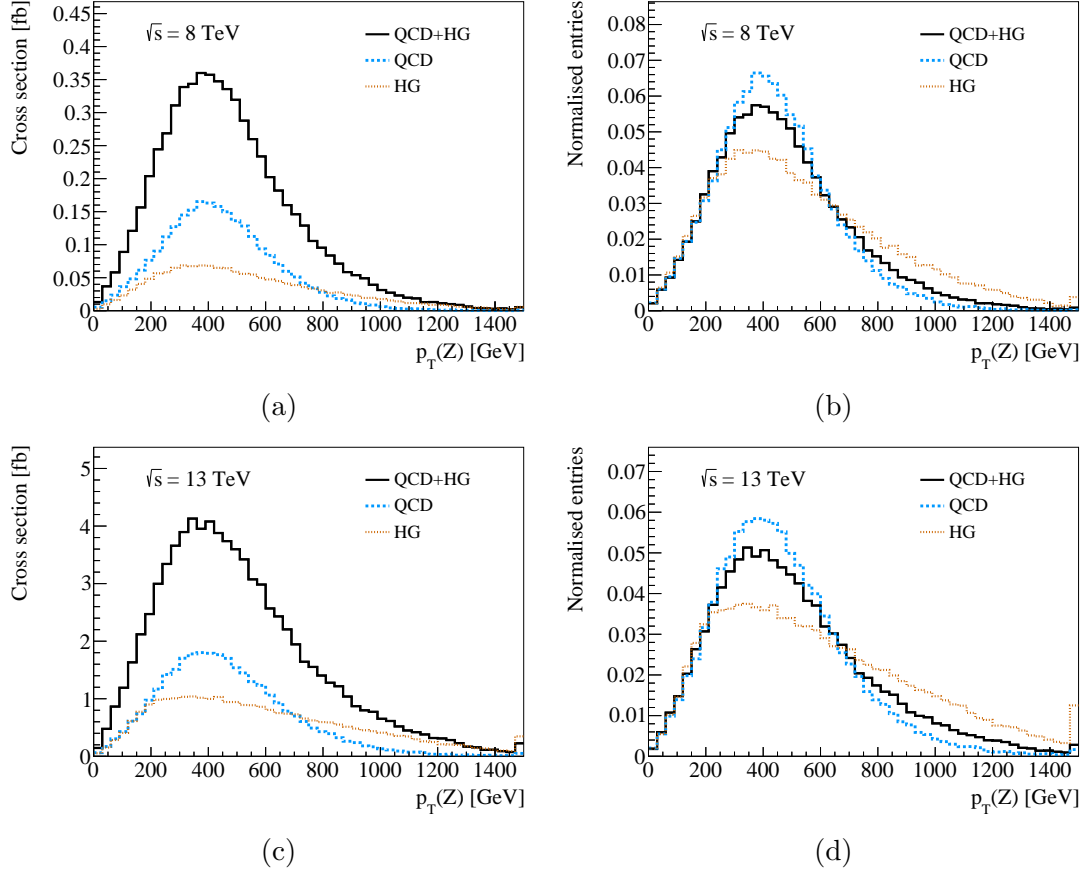


Figure 6.1: Distribution of the  $Z$  boson transverse momentum for  $\sqrt{s} = 8$  TeV (top) and 13 TeV (bottom) at parton level with  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV.

found in other distributions of objects close to the detector such as  $p_T$  of the leptons coming from the decay of the  $Z$  boson or in the  $p_T$  of the boson present in other decay channels. Only the  $p_T(Z)$  distribution is shown as an example since it is an important variable in the analysis presented in the previous Chapter.

This behaviour is expected since objects further up the decay chain would be more sensitive to the presence of the heavy gluon. Figure 6.2 and 6.3 show  $p_T(X'_{2/3})$  and  $M_{X'_{2/3}\bar{X}'_{2/3}}$  respectively for the same masses as before. As expected, when moving from the  $Z$  boson up in the decay chain to the  $X'_{2/3}$  quark, bigger differences start to appear. These differences are also larger for 13 TeV than for 8 TeV. This can be understood when looking at Figure 6.3 and comparing the distribution for 8 TeV (top) and 13 TeV (bottom). At 8 TeV there is not enough energy to produce on-shell heavy gluons and most of the time it is produced off-shell, as can be seen looking at the HG curve. The off-shell heavy gluon produces a top partner spectrum

closer to the QCD continuum. On the other hand, at 13 TeV the increase in energy makes it possible for the heavy gluon to be produced on-shell, generating harder top partners. Of course, there is a balance between the best observable to look for differences and what can be achieved in a real analysis. Reconstructing the full decay chain for both top partners to be able to find a heavy gluon candidate turns out to be a rather ambitious goal due to the high number of topologies in play. Furthermore, the more complex the reconstruction the more degradation is introduced by hadronisation and detector effects in the different observables involved in such reconstruction.

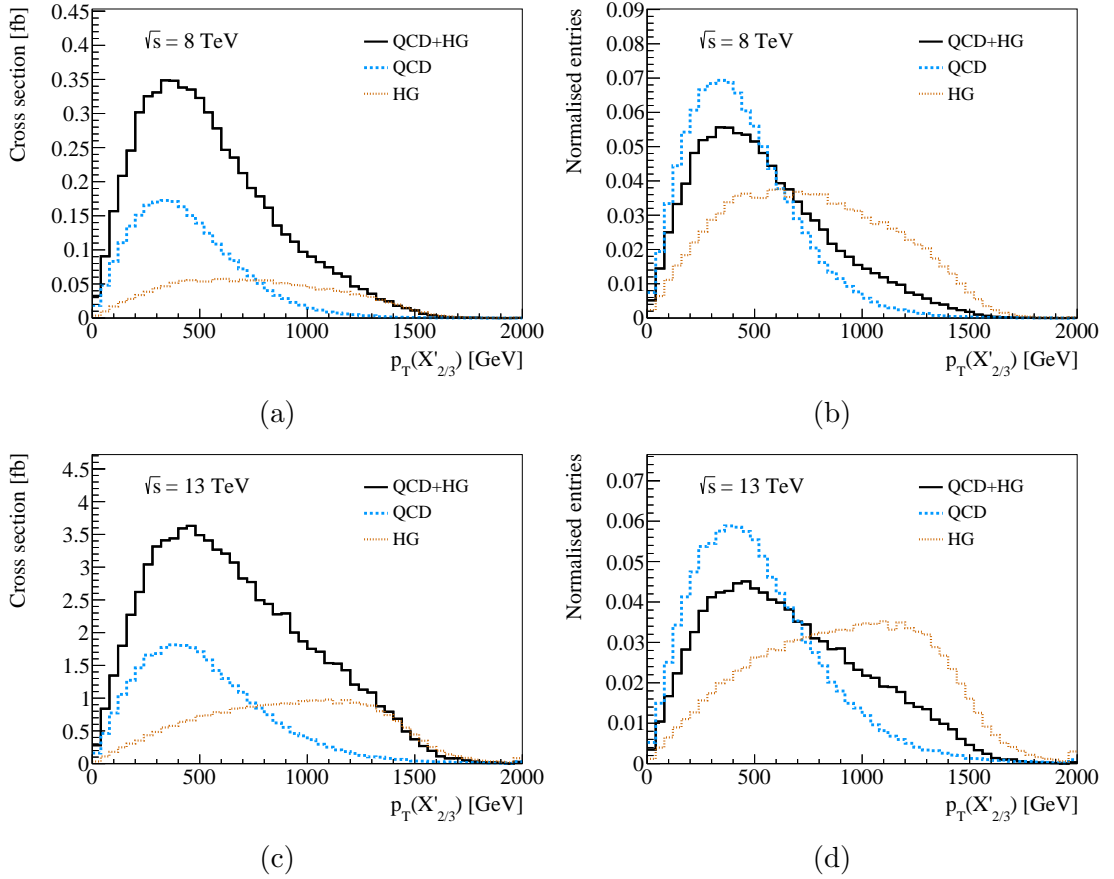


Figure 6.2:  $X'_{2/3}$   $p_T$  distribution for  $\sqrt{s} = 8$  (top) and 13 TeV (bottom) at parton level with  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV.

Until now, the usual behaviour found is that the presence of a heavy gluon produces harder top partners. Nonetheless it is worth noting that it might not be the case. With a heavy gluon of  $M_G = 3.5$  TeV, as chosen for the benchmark in this analysis, a VLQ with mass  $M_{X'_{2/3}} = 1$  TeV is always allowed to be produced

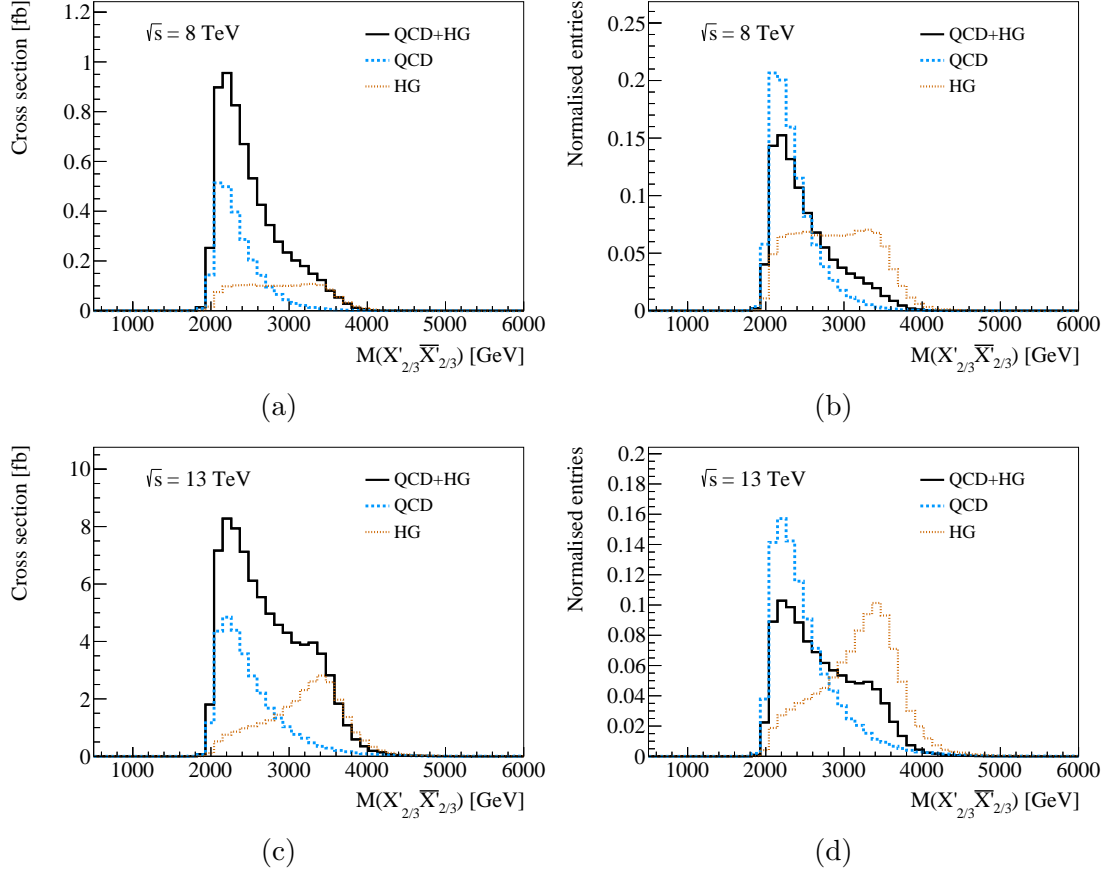


Figure 6.3:  $X'_{2/3}\bar{X}'_{2/3}$  invariant mass distribution for  $\sqrt{s} = 8$  (top) and 13 TeV (bottom) at parton level with  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV.

on-shell. An interesting situation occurs when the mass of the heavy gluon is below the kinematical threshold to produce a pair of VLQ (i.e.  $M_G < 2M_{X'_{2/3}}$ ) and the HG production dominates over QCD<sup>2</sup>. In this scenario the presence of a heavy gluon will translate into softer VLQs than expected in the QCD case as can be seen in Figure 6.4 which show both the  $p_T(X'_{2/3})$  and  $M_{X'_{2/3}\bar{X}'_{2/3}}$  distributions at 8 TeV (for this case the 13 TeV distributions don't show any difference) for a pair of masses  $M_G = 2.5$  TeV and  $M_{X'_{2/3}} = 1.3$  TeV.

<sup>2</sup>As the mass of the heavy gluon increases the production for QCD dominates due to the rapid decay of the PDFs. On the other hand, as the mass of the top partners increases the coupling with the composite sector increases making the HG production dominate over QCD.

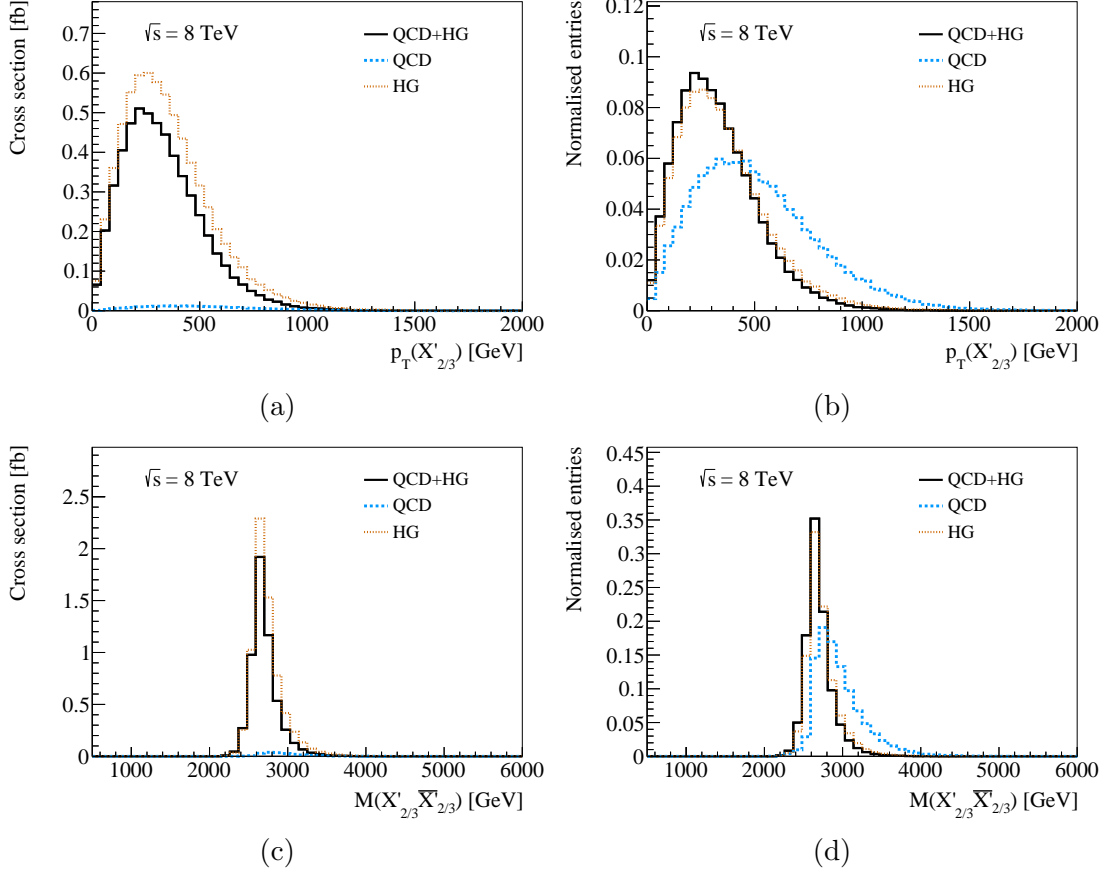


Figure 6.4:  $X'_{2/3}$   $p_T$  distribution for  $\sqrt{s} = 8$  (top) and 13 TeV (bottom) at parton level with  $M_G = 2.5$  TeV and  $M_{X'_{2/3}} = 1.3$  TeV.

### 6.3.2 Kinematical differences after detector simulation

After showing the effect of the heavy gluon at the parton level in the previous section, the effects of hadronisation and detector simulation are now taken into account. After hadronisation and detector simulation are introduced the different particles in each event needs to be properly reconstructed. As mentioned before, trying to reconstruct the full decay chain of both top partners is too ambitious for Run-1 or early Run-2 analyses. Instead, the decay  $X'_{2/3} \rightarrow Zt \rightarrow ZbW$  will be reconstructed assuming hadronic decays of the  $W$  boson. First, a selection is applied in which each event must contain at least a pair of same-flavour opposite-sign leptons satisfying  $|M_{\ell\ell} - M_Z| < 10$  GeV which will define the  $Z$  candidate. Then, events are required

to have at least 4 jets and at least 2  $b$ -jets.<sup>3</sup> For each event passing the selection, a modified  $\chi^2$  is used to identify the best object combination to reconstruct each particle. For a particle  $X$  which decays to two particles  $i, j$  the modified  $\bar{\chi}_X^2$  is defined as

$$\bar{\chi}_X^2 = \frac{(M_{ij} - M_X)^2}{\sigma_X^2} \Delta R(i, j), \quad (6.7)$$

where  $M_{ij}$  is the invariant mass of the  $i, j$  system,  $M_X$  is the mass of the particle  $X$ ,  $\sigma_X$  is its width and  $\Delta R(i, j)$  is the angular separation between the two decay products. For each event all the possible combinations of jets and  $b$ -jets are done and the one that minimises

$$(\bar{\chi}_W^2 + \bar{\chi}_t^2) \Delta R(Z, t). \quad (6.8)$$

is chosen as the correct one. The  $\Delta R$  term in the  $\bar{\chi}_X^2$  definition is used to include the angular separation between the decay products in the object reconstruction.

After detector simulation is applied and the event reconstruction is done, the previous differences seem to almost disappear. Figure 6.5 shows different distribution for the benchmark pair of masses ( $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV) and  $\sqrt{s} = 13$  TeV, which showed greater differences at the parton level. Figure 6.5(a) shows the invariant mass of the  $X'_{2/3}$  candidate to show that the top partner mass is correctly recovered after the reconstruction. In Figures 6.5(b), (c) and (d) the  $p_T(Z)$ ,  $p_T(X'_{2/3})$  and  $H_T(\text{jets})$  are shown. As it can be seen no difference is observed when comparing the QCD and QCD+HG expectation. This seems to point out that, in a realistic scenario, the impact of the heavy gluon in the final discriminant is negligible and thus it would be straightforward to reinterpret the results only with a change in the cross-section.

When comparing the different production mechanisms in this section, different effects have been included: hadronisation, detector simulation and the event reconstruction itself; this makes unclear what is the main process responsible for the disappearance of the differences observed at the parton level. In order to find which effect is responsible, the  $p_T(X'_{2/3})$  distribution has been obtained step by step, which are defined as follows:

- **Parton level:** this corresponds to the distribution showed in the previous

---

<sup>3</sup>This selection will be better explained in the next section when this reconstruction is used to evaluate the impact on a realistic scenario.



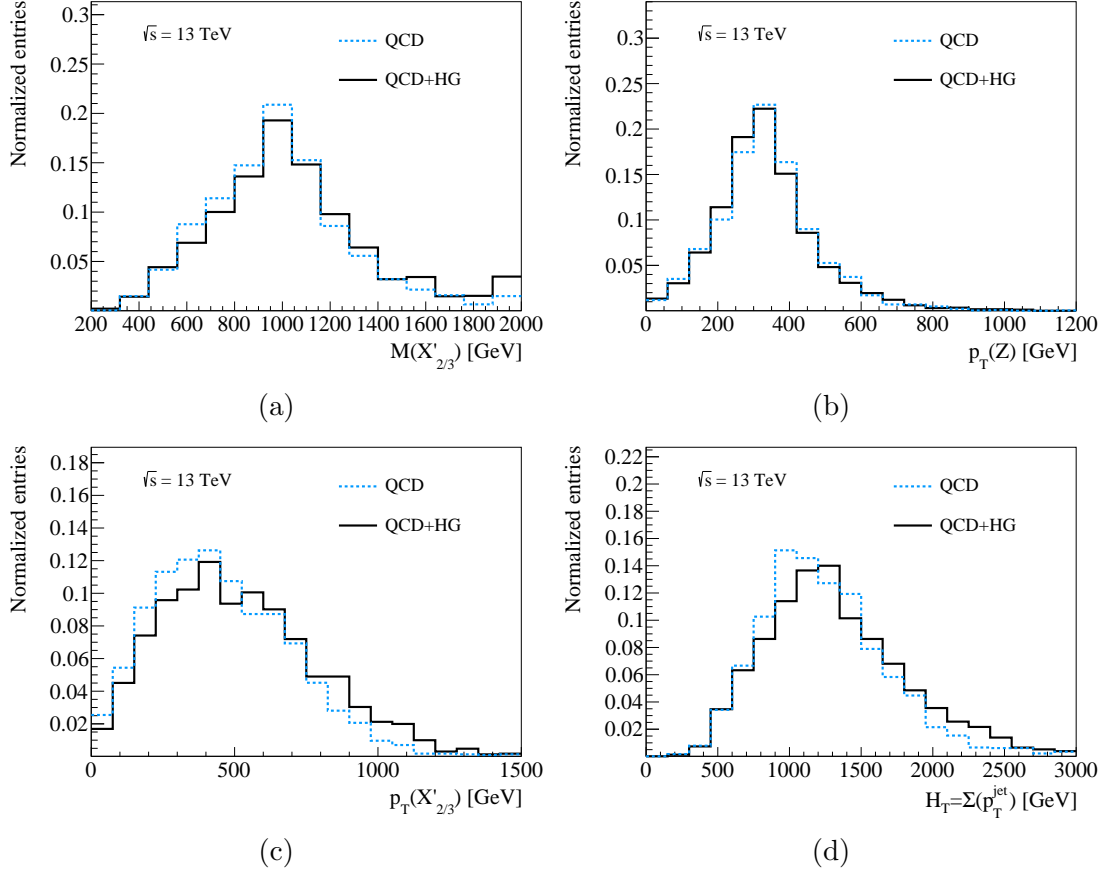


Figure 6.5: Kinematical distributions after detector simulation for  $\sqrt{s} = 13$  TeV with  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV. The distributions shown are the invariant mass of the objects reconstructing  $X'_{2/3}$  (top left), the  $Z$  boson transverse momentum (top right), the reconstructed  $X'_{2/3}$  transverse momentum (bottom left) and  $H_T$  (bottom right). All distributions are normalised to unit area.

section where the  $X'_{2/3}$  can be recognised by the particle information, i.e. its pdgid.

- **Truth matching:** in this case the **FastJet** algorithm has been used to do the clustering of the particles after hadronisation, without including the effects of the detector simulation. The two jets closest to the  $W$  boson and the jet closest to the  $b$ -jet decaying from the top are matched, rejecting those which are outside a cone of  $\Delta r > 1$  from each particle. This provides a method to evaluate the impact of hadronisation alone.
- **Reco (no Delphes):** here the reconstruction algorithm presented in this

section is applied using the jets created for the previous scenario. The  $Z$  candidate is chosen based on parton level information. This procedure allows to evaluate the effect of the reconstruction itself without including detector simulation.

- **Reco (with Delphes):** this correspond the final distribution as shown in Figure 6.5(c) including the entire simulation chain.

These steps are all shown in Figure 6.6 for QCD (a), HG (b) and QCD+HG (c) showing that while the QCD channel shows no difference at each step and QCD+HG shows a mild difference, the biggest one is found in the HG channel. A sequential degradation can be observed with each step contributing to the softening of  $p_T(X'_{2/3})$ . Nonetheless it is the reconstruction itself which seems to be drifting the distribution towards the QCD region. This can be better understood in Figure 6.6(d) where the mass of the reconstructed  $W$  boson is shown for the *Truth matching* and *Reco (no Delphes)* cases. For the truth matching scenario, it can be seen that there is a peak at the  $W$  mass but also a secondary peak at  $m_t$  (even a small bump at the  $X'_{2/3}$  mass). This implies that in a considerable amount of events<sup>4</sup> the two jets which have been reconstructed closest to the  $W$  boson decay products do not reconstruct a  $W$  boson but a top quark instead, i.e. boosted topologies are present. This secondary peak disappears after the usual reconstruction is done, which is expected since this reconstruction algorithm is designed to group objects to produce a given mass. In this case, the differences might be recovered using a reconstruction method which uses larger jets and jets substructure variables. Nonetheless this is beyond the scope of this analysis which aims to study the impact on currently published analysis or early Run-2 analyses.

### 6.3.3 Recasting of the VLQ searches at the LHC

The goal of this section is to evaluate the impact of the heavy gluon in a realistic scenario. As seen in the previous section, after hadronisation and detector simulation is introduced, the differences observed at the parton level are washed out. In this case, the efficiencies of a given analysis, and therefore the experimental sensitivity, are likely to be very similar in models with and without heavy gluons in the spectrum.

---

<sup>4</sup>The 52% of the events with  $M(W) < 220$  GeV will give a  $W$  mass which peaks at the SM top mass.

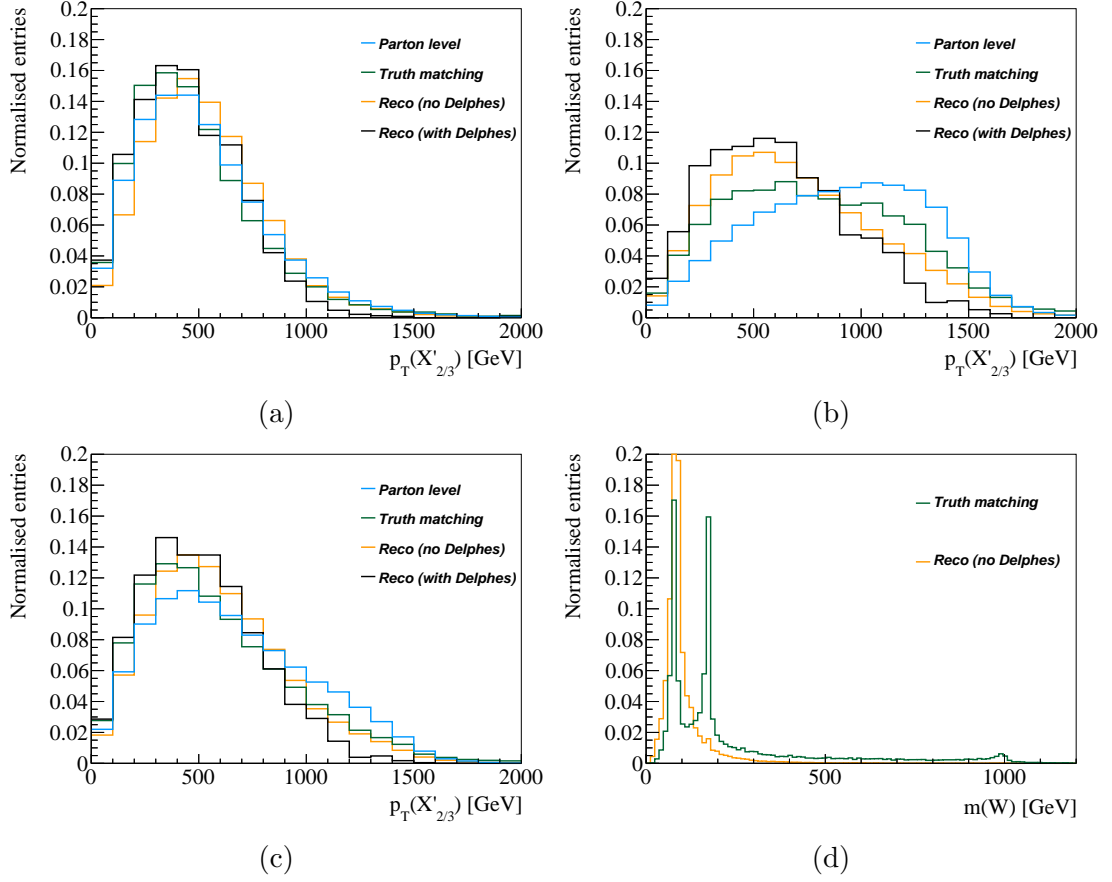


Figure 6.6: The  $p_T(X'_{2/3})$  distribution at parton level (blue line), reconstructed using truth matching and FastJet clustering (green line), using the reconstruction algorithm presented in the paper without detector simulation (yellow line) and using the full analysis chain as presented in the paper (black line). The distribution is shown in the three possible scenarios, i.e. only QCD (a), only HG (b) and QCD+HG (c). The mass of the reconstructed  $W$  boson (d) is shown, only for the HG channel, comparing the truth matching step and the reconstruction algorithm without Delphes.

In order to assess the impact of the small differences remaining, a recast of the analysis presented in Chapter 5 is done. The same selection criteria presented in the previous chapter will be applied to the signal and background samples and the expected limits will be derived in two different ways to identify the effect of the heavy gluon. The first approach is to compute the limits performing the analysis for the models in which both  $X'_{2/3}$  and the heavy gluon are included. For the second approach the limits are derived assuming only QCD production and that the only effect of the heavy gluon is a rescaling of the cross-section to the corresponding

value when including the heavy gluon. These limits will be computed for each pair of masses and represented in the  $M_{X'_{2/3}} - M_G$  plane. If the effect of the heavy gluon is not visible both limits are expected to be similar while divergences between both limits will be seen otherwise.

As before, the benchmark masses used to show the background and signal comparison are  $M_{X'_{2/3}} = 1$  TeV and  $M_G = 3.5$  TeV, although all masses have been analysed to be able to derive the limits at the end. The event selection is the same as shown in Table 5.1: the  $Z$  boson candidate is reconstructed with a pair of same-flavour opposite-sign leptons with an invariant mass satisfying  $|M_{\ell\ell} - M_Z| < 10$  GeV. After the  $Z$  boson has been defined, events are selected with at least 2 jets, at least 2  $b$ -jets, with a  $Z$  boson candidate with  $p_T(Z) > 150$  GeV and  $H_T(\text{jets}) > 600$  GeV. For this selection criteria the Table 6.1 show the efficiency after each cut applied comparing the QCD and QCD+HG channels. As expected, the efficiencies are similar for both channels which implies that it is likely that limits will not show any difference when including the heavy gluon.

Selection	QCD cut efficiency (%)	QCD+HG cut efficiency (%)
Leptonic $Z$	1.28	1.36
$\geq 2$ jets	99.82	99.90
$\geq 2$ $b$ -jets	64.25	64.00
$p_T(Z) > 150$ GeV	93.09	92.50
$H_T > 600$ GeV	94.19	93.42

Table 6.1: Event selection and cut efficiencies for both QCD and QCD+HG productions at  $\sqrt{s} = 8$  TeV for  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV. Each efficiency is derived based on the number of events which passed the cut before. The efficiency of the first cut includes the leptonic BR of the  $Z$ .

Figure 6.7 show the background and signal comparison (both for QCD and QCD+HG) at different stages of the analysis with all the distributions normalised to unit area. The  $b$ -jet multiplicity is shown in (a) after requiring a  $Z$  candidate and at least 2 jets, the  $p_T(Z)$  distribution is shown in (b) after requiring at least 2  $b$ -jets, the  $H_T(\text{jets})$  distribution is shown in (c) after selecting events with  $p_T(Z) > 150$  GeV and the final discriminant,  $M(Zb)$  is shown in (d) after the full selection is applied. As can be seen, no significant difference is found in the signal shapes when including the heavy gluon.

After the full selection has been applied the  $CL_s$  method is used to compute the VLQ mass limit as a function of  $M_G$ , shown in Figure 6.8(a), in the two ways

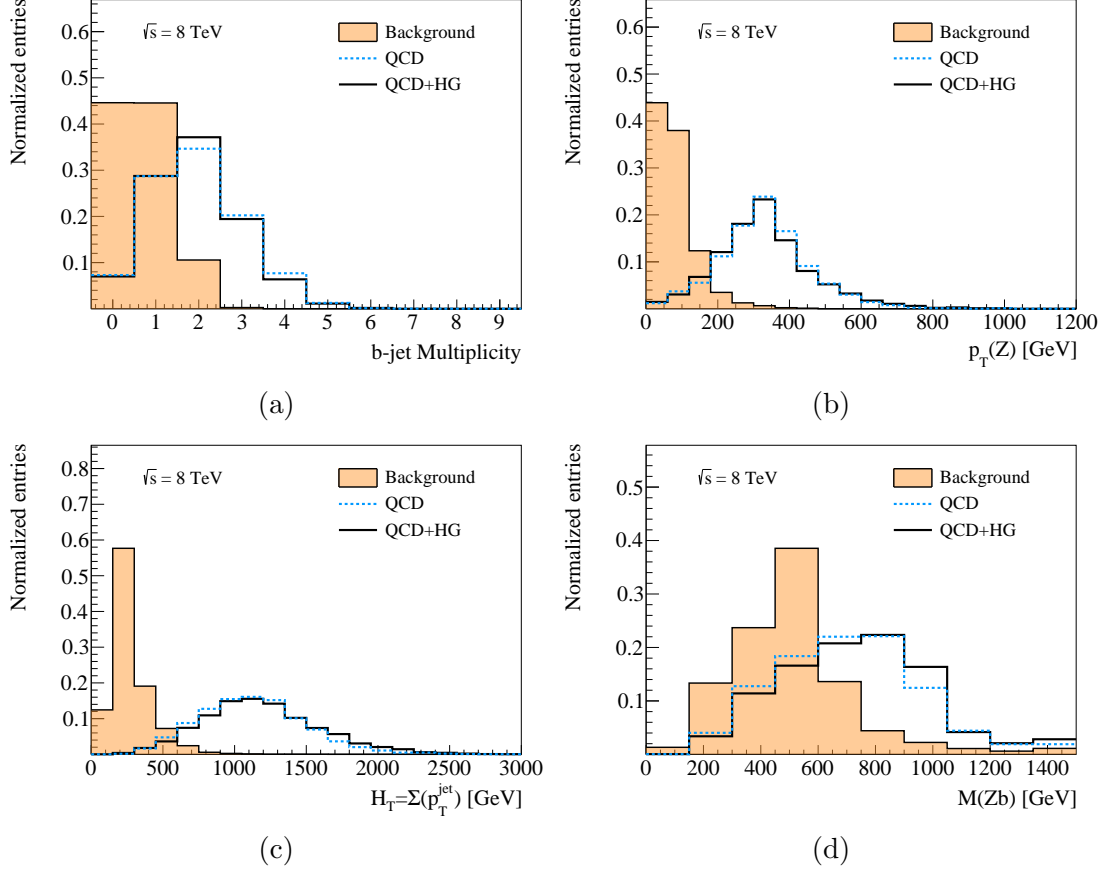


Figure 6.7: Comparison between signal (QCD and QCD+HG) and background for the  $b$ -jet multiplicity,  $p_T(Z)$ ,  $H_T$  (jets) and the final discriminant  $M(Zb)$ . Distributions are shown for  $\sqrt{s} = 8$  TeV and  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV.

mentioned before. The solid purple line corresponds to the bounds computed when properly including the heavy gluon in the signal hypothesis. The green line, labeled as *Scaled QCD limit*, corresponds to the bounds computed assuming that the experimental sensitivity is the same in the QCD and QCD+HG production mechanisms. Figure 6.8(b) illustrates the second method in which the intersection between the experimental upper limit in the production cross section at a 95% C.L., when assuming only QCD, and the theory curves for different values of  $M_G$  defines the lower bound in  $M_{X'_{2/3}}$  for each mass of the heavy gluon. As can be seen, both methods provide nearly identical results. This indicates that with the current analyses<sup>5</sup> the already published experimental limits can be used to set bounds on the mass of

<sup>5</sup>Assuming that the VLQ is not reconstructed and that no boosted techniques are used.

the heavy gluon. In Figure 6.8(a) are also shown the limits for the QCD only case (dotted blue line) and the limit reported for the dilepton channel in the ATLAS analysis (dashed red line). The agreement of both values, including also the limit for high  $M_G$ , show the consistency of the results.

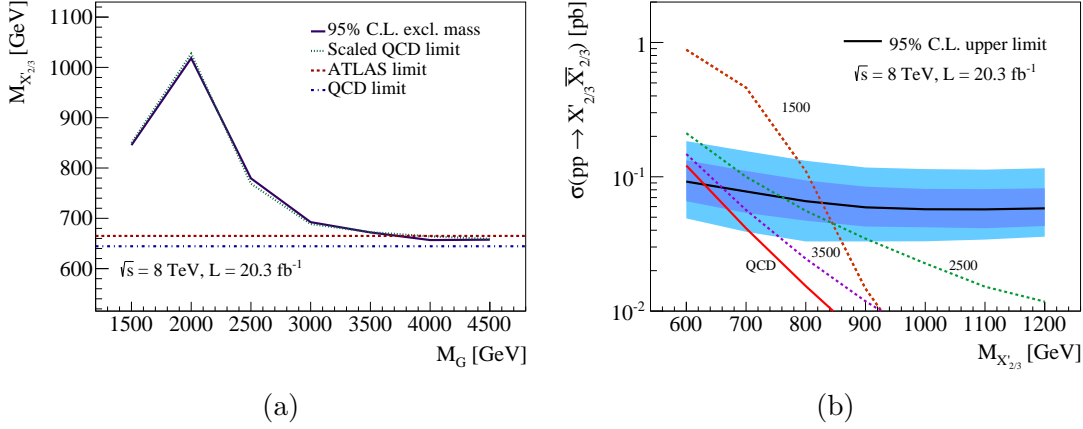


Figure 6.8: The 95% confidence level (C.L.) lower limits in the  $M_G - M_{X'_{2/3}}$  plane (a) computed with the two methods discussed in the text. The limit reported by ATLAS ( $\sqrt{s} = 8 \text{ TeV}$ ) in dashed red and the corresponding number obtained with the recast of the analysis in dashed blue are also shown. The 95% C.L. upper limits on the pair production cross section as a function of  $M_{X'_{2/3}}$  (b) assuming QCD production (solid black) with 1- and 2- $\sigma$  bands in blue, together with the theoretical production cross sections in the model for different values of  $M_G$  are represented.

The recast of the analysis presented in Chapter 5 shows that there is no sensitivity to the presence of a heavy gluon. Nonetheless, the  $X'_{2/3}$  is not reconstructed and the expected differences after detector simulation are not expected to be large. In order to try to exploit the small differences found after detector to better evaluate the impact of a heavy gluon a more sophisticated analysis has been developed. The event selection is slightly modified and the  $X'_{2/3}$  is reconstructed using the same procedure shown in Section 6.3.2. The  $Z$  boson candidate is reconstructed in the same way as before. Once the  $Z$  boson is identified, events are required to have at least 4 jets, with at least 2 of them being  $b$ -tagged jets. The jet multiplicity selection is increased from the  $\geq 2$  jets requirement used in the ATLAS analysis given that for the signal considered in this analysis a higher jet multiplicity is expected. At least one pair of the selected jets is required to have an invariant mass inside a 10 GeV window of the  $W$  mass and the two leptons decaying from the  $Z$  boson are required to satisfy  $\Delta R(\ell_1, \ell_2) < 1.5$ . Finally the cut on the  $H_T(\text{jets})$  variable is dropped. The two

last cuts are introduced in order to still reduce the background contribution while leaving the  $p_T(Z)$  and  $H_T(\text{jets})$  variables with discriminant power. The reason to do this is that a multivariate analysis (MVA) is used after the final selection has been applied in order to exploit the small differences that can appear in different variables instead of using only one discriminant variable.

The MVA used is a neural network (NN) implemented using the TMVA[155] using 7 inputs variables (shown in Figure 6.9): Jet multiplicity,  $p_T(t)$ ,  $p_T(X'_{2/3})$ ,  $p_T(Z)$ ,  $H_T$ ,  $M_{X'_{2/3}}$  and  $p_T^{\text{align}}$ .  $p_T^{\text{align}}$  is a variable that is sensitive to both the energy of the decay products of the vector-like quark and the angular separation between them. It increases as we move to a more boosted regime and is defined as

$$p_T^{\text{align}} = \frac{\sqrt{p_T(Z)^2 + p_T(W)^2 + p_T(b)^2}}{\max\left(\Delta R(X'_{2/3}, Z), \Delta R(X'_{2/3}, W), \Delta R(X'_{2/3}, b)\right)}. \quad (6.9)$$

These variables are selected to provide good discrimination between signal and background and for their potential to provide discrimination between both production mechanisms (QCD and QCD+HG).

The NN first needs to be trained with a subset of signal and background events to *learn* to recognise signal and background events. This training is done for each pair of masses  $M_G - M_{X'_{2/3}}$  separately. After the NN has been trained a test run is done with the remaining events to ensure that no overtraining (or undertraining) is present<sup>6</sup>. The training events for signal and background are taken to be randomly selected from all the available events, splitting the sample in half. The output of the NN for the training and tests set of events, both for the signal and background samples, presented the same behaviour without any bias observed.

---

<sup>6</sup>Overtraining or undertraining might appear when the algorithm fails to properly recognise signal and background events producing a bias in the output to the signal or background region respectively.

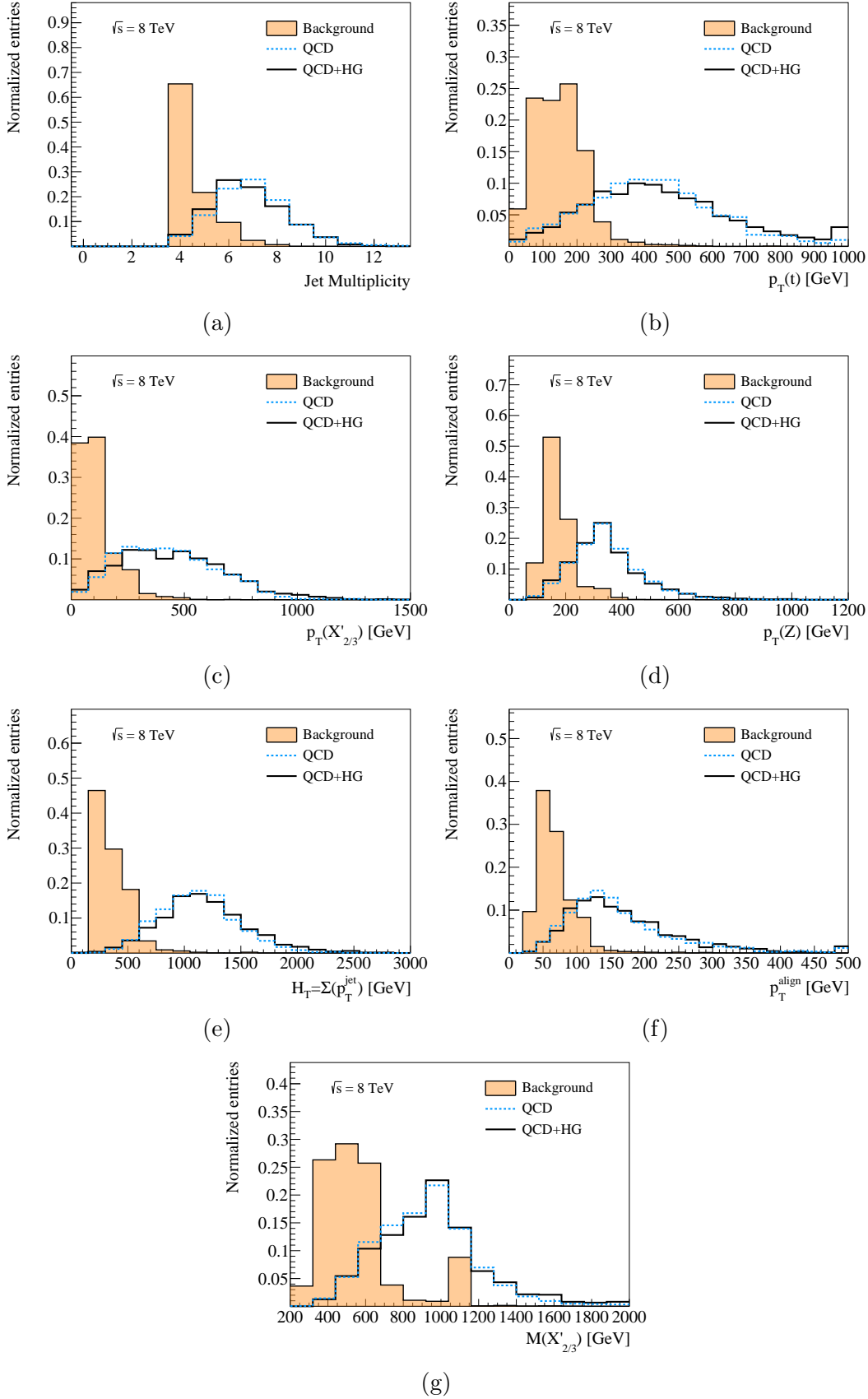


Figure 6.9: The 7 inputs for the NN are shown for the benchmark masses  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$ : Jet multiplicity (a),  $p_T(t)$  (b),  $p_T(X'_{2/3})$  (c),  $p_T(Z)$  (d),  $H_T$  (e),  $M_{X'_{2/3}}$  (f) and  $p_T^{\text{align}}$  (g).



Finally, the NN output (shown in Figure 6.10 (a) for the benchmark masses) is used to derive the limits in the  $M_G - M_{X'_{2/3}}$  plane. These limits are shown in Figure 6.10 (b) and, even though the MVA analysis seems to be more sensitive to small variations, the difference between the properly computed limit, taking into account the presence of HG, and the one in which the QCD experimental limit is used is only marginal. Thus, the conclusion drawn from the previous analysis that one can simply take the experimentally published limits that assume QCD production and use them to compute the bounds on models with a heavy gluon seems to be robust and apply also to more sophisticated analyses with a richer object reconstruction.

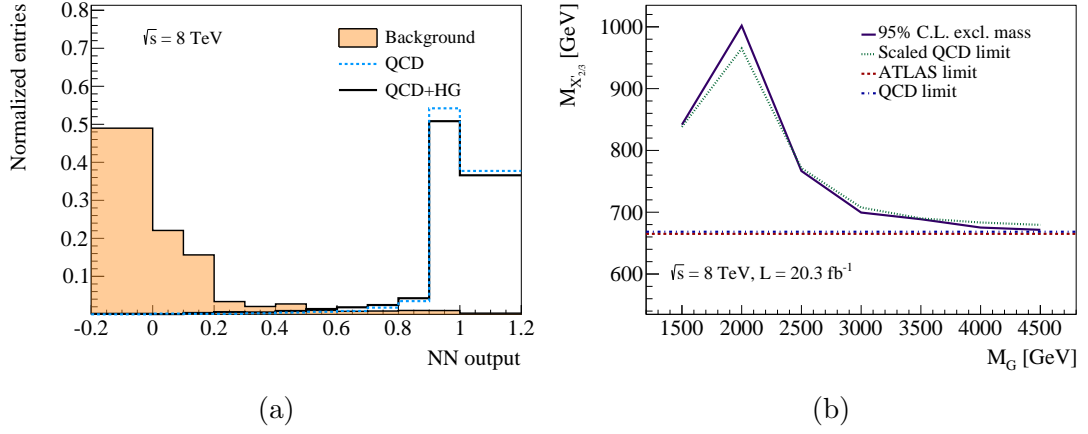


Figure 6.10: The output of the NN (a) for the masses  $M_G = 3.5$  TeV and  $M_{X'_{2/3}} = 1$  TeV is shown. The background estimation corresponds to the training done in the QCD+HG case. The 95% C.L. lower limit on  $M_{X'_{2/3}}$  as a function of  $M_G$  (b) for  $\sqrt{s} = 8$  TeV with the full run-1 integrated luminosity is also shown. The green line shows the lower mass limit assuming that the only effect is an increase in the cross section and the purple line shows the lower mass limit when considering the correct cross section upper limit for each pair of masses.

## 6.4 Limits on heavy gluons from VLQ searches

After showing that the presence of a heavy gluon does not modify the efficiencies, and hence the final results on the limits obtained by different analyses, this section aim to reinterpret the results published by the ATLAS and CMS collaborations for direct searches of VLQs and extract the current bounds to the mass of the heavy gluon and the VLQs which appear in the model at hands. The results will be presented in the  $M_G - M_Q$  plane, where  $Q$  runs for all the VLQ present in the

spectrum. Afterwards the same analysis presented in the previous section, using the NN is applied to 13 TeV signal and background samples to evaluate the reach of early Run-2 VLQ searches with an integrated luminosity of  $10 \text{ fb}^{-1}$ , which correspond to a realistic expectation for early Run-2 data.

### 6.4.1 Current limits

From the study shown in this chapter one can extract the conclusion that the presence of a heavy gluon does not impact the performance of a given analysis<sup>7</sup>. In this case a reinterpretation of the results can be done by only rescaling the cross section accordingly for each VLQ and heavy gluon mass.

There are two main obstacles which were found when attempting to do this reinterpretation. The first one is that even though the heavy gluon does not change the efficiencies of a given analysis, the decay BR for the VLQ does. This is due to the contamination from other decay channels into the one for which an analysis is designed to be most sensitive to. In this aspect, if the decay BRs for the VLQ implemented in a given analysis are not compatible with the ones presented by the model described in this section the efficiencies of the analysis will change, leading to different results when including this signal hypothesis. For this reinterpretation, a given published analysis has been used if the BR of the channel the analysis is most sensitive to is equal or smaller than the BR in the CHM5<sub>4</sub> model. When the BR is the same than the one predicted by the CHM5<sub>4</sub> the reinterpretation is straightforward. If the BR is smaller, the results are scaled up to match the BR of the CHM5<sub>4</sub>. This approach provides a conservative bound since a lower or equal global efficiency than the one obtained with the model at hands is assumed, due to the contamination from channels with lower efficiency.

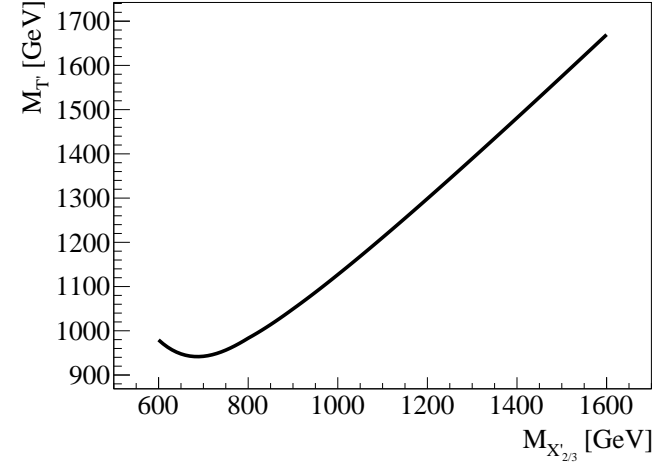
The second obstacle is that, usually, results are reported for only one given VLQ, nonetheless several VLQ may appear in the same model and sometimes a given analysis won't be able to distinguish between two different VLQs. In the model used in this section the VLQs can be organised in two groups: the  $X_{5/3}$  and  $B$ , both decaying exclusively to  $Wt$ , which unless specific reconstruction is done for each VLQ to identify the charge of the lepton decaying from the  $W$ , they are difficult to distinguish (a detailed study on how to identify different top partners at the LHC

---

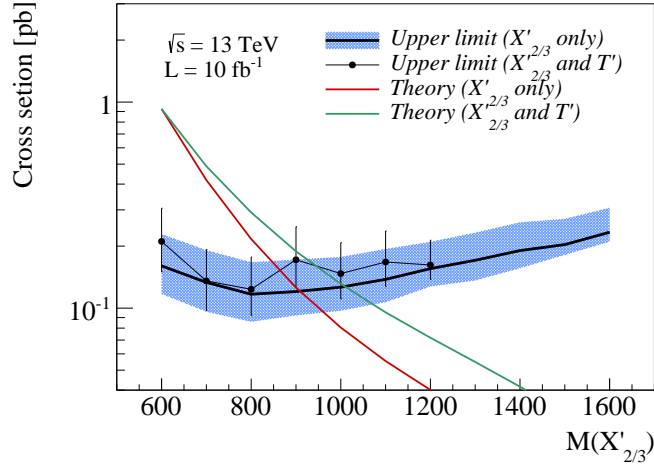
<sup>7</sup>It is important to remember that this statement is valid for analysis in which no boosted variables, such as jets substructures, are used as was shown in Section 6.3.2. This is the case for the Run-1 analyses and a realistic scenario for early Run-2 analyses, so this conclusion remains valid for the study shown in this section.

can be found in [44]); the second group is formed by  $X'_{2/3}$  and  $T'$  which have the exact same decay patterns and therefore are undistinguishable. In the CHM5<sub>4</sub> used in this section, the  $X_{5/3}$  and the  $X'_{2/3}$  are degenerate and lighter than  $T'$  and  $B$ . In the region of parameter space that is being considered the mass splitting is of the order of 10%, which is large enough to make interference effects small but not so large that two distinct peaks would show up in the relevant distributions. Furthermore, the heavier ones,  $T'$  and  $B$ , are heavy enough to make their contribution to the total cross section much smaller than the one from the lighter ones,  $X_{5/3}$  and  $X'_{2/3}$ , in the searches to which both contribute. The  $T'$  mass as function of the  $X'_{2/3}$  mass is shown in Figure 6.11 (a) for the masses considered in this analysis. In these circumstances, the global efficiencies are likely to be insensitive to the presence of both quarks simultaneously. In order to quantify this difference, the MVA has been repeated including the two 2/3 quarks in the same signal modelling and evaluating the limits in this case for a selected set of masses. The results, in Figure 6.11 (b), show that the efficiencies in both cases are very similar. The 95% C.L. cross-section upper limit as a function of  $M_{X'_{2/3}}$  is shown considering only the  $X'_{2/3}$  (black line) with  $\pm 1\sigma$  band in blue and including both the  $X'_{2/3}$  and the  $T'$  (black dots). The theory curve in red is the cross section considering only  $X'_{2/3}$  while the green line shows the cross-section when including both VLQs. It can be seen how, for the 13 TeV case for which the differences are expected to be larger, the expected limits are compatible within uncertainties and the intersection between both curves and each of the expected limits are compatible as well. Therefore, it is a good approximation to derive the mass limit overlying the total cross-section curve over published limits where only one VLQ has been taken into account.

The published analyses which have been used to derived the current bounds on the VLQ and heavy gluon masses have been the following. Starting from the  $B$  quark, the analyses presented in [46] (ATLAS) and [156] (CMS) have been used. The analysis published by ATLAS has a lower BR to the  $W$  channel than the one predicted by the model at hands. However, the analysis is most sensitive to the  $W$  channel and, as discussed before, a conservative bound can be establish rescaling the published results to the needed 100% BR. For the case of the CMS analysis the results are already published with a 100% BR to the  $W$  channel. Unfortunately, the  $B$  masses considered in this analysis are heavier than the ones to which these analysis are sensitive to. However, in some cases the limits can be established. For instance, using the ATLAS results some values of  $M_B$  can be excluded ranging from



(a)



(b)

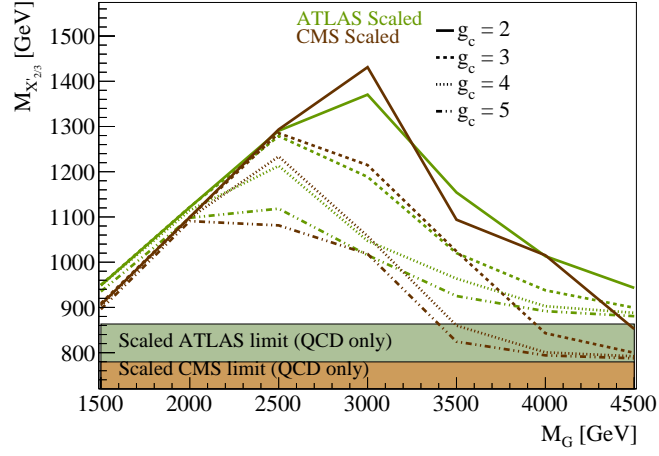
Figure 6.11: The  $M_{T'}$  mass as a function of  $M_{X'_{2/3}}$  mass (a) for the chosen parameters and the comparison between the 95% C.L. experimental limit obtained assuming only  $X'_{2/3}$  production (black solid line) and assuming production of both  $X'_{2/3}$  and  $T'$  (black dots) are shown. The solid red line represents the cross section of pair production of  $X'_{2/3}$  only while the dashed green line represent the cross section of pair production of both  $X'_{2/3}$  and  $T'$ . In this figure  $\sqrt{s} = 13$  TeV have been chosen since larger differences are observed. The benchmark mass for the heavy gluon  $M_G = 3.5$  TeV is shown.

$M_B \simeq 750$  GeV for  $M_G = 4$  TeV to  $M_B \simeq 1$  TeV for  $M_G = 2$  TeV. Using the CMS results we can excluded values of  $M_B$  ranging from  $M_B \simeq 800$  GeV for  $M_G = 3$  TeV to  $M_B \simeq 950$  GeV for  $M_G = 2$  TeV. These limits have been found for  $g_c = 3$ .

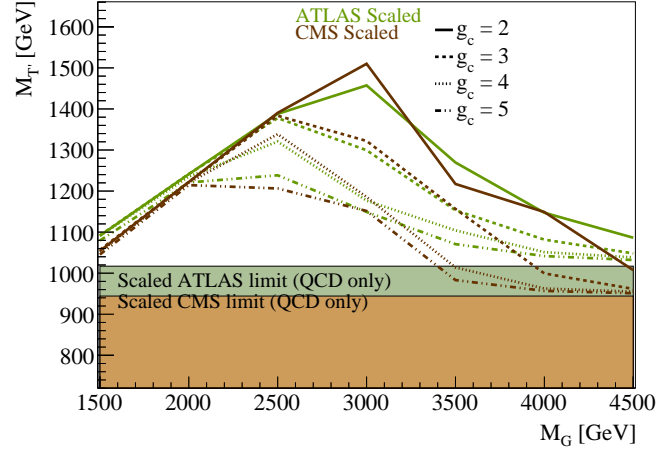
For the  $X'_{2/3}$  and  $T'$  quarks the results in [45] (ATLAS) and [157] (CMS) have been used. The ATLAS results contain a combination of two analyses that target decays of the VLQ into  $H$  and  $W$ , respectively. Since charged current decays are not present for the  $X'_{2/3}$  and  $T'$  quarks in the present model the results quoted for the doublet (with a 50% BR to the  $Z$  and  $H$  channels) only in the  $H$  channel have been used. The CMS analysis reports limits for a 2:1:1 BR pattern into  $W$ ,  $H$  and  $Z$ , respectively, and the analysis used is very sensitive to the  $Z$  channels. Thus, their results have been scaled to our BR as discussed above. The results can be found in Figure 6.12 (a) for different values of  $g_c$ . These results can be transformed in limits on the mass of the  $T'$ , shown in Figure 6.12 (b) using the dependence shown in Figure 6.11. In the case of the  $X_{5/3}$  the same ATLAS analysis used for the  $B$  quark was used. The analysis also published results for the search of a vector-like quark with a electric charge of 5/3 assuming a BR of 100% to  $Wt$ . The CMS analysis [51] searching for VLQ with an electric charge of 5/3 in the same-sign dilepton final state has been also used. They assume a 100% BR into  $Wt$  which make their results straightforward to reinterpret. The mass limits can be found in Figure 6.12 (c).

### 6.4.2 Early Run-2 expectations

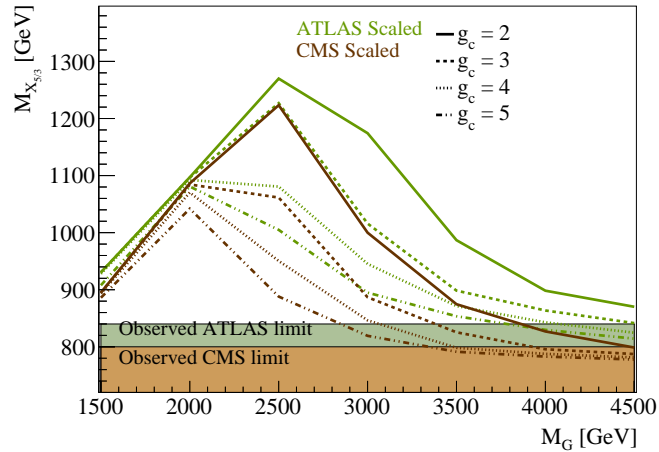
Once the current bounds on VLQ and heavy gluon masses have been derived for the CHM5<sub>4</sub> model the next step is to evaluate what would be the reach of the early analyses performed with Run-2 data. The idea of this study is not only to see how the current bounds are expected to improve with early Run-2 data ( $L = 10 \text{ fb}^{-1}$  and  $\sqrt{s} = 13 \text{ TeV}$ ) but to see if the higher center-of-mass energy will introduce any kinematical differences that can be observed after the full analysis chain.



(a)



(b)



(c)

Figure 6.12: The 95% C.L. lower limits on the  $M_G - M_{X'_{2/3}}$  (a),  $M_G - M_{T'}$  (b) and  $M_G - M_{X_{5/3}}$  (c), derived from full Run-1 published data ( $\sqrt{s} = 8$  TeV), for different values of  $g_c$  are shown. The filled orange and green region represent the limits for CMS and ATLAS, respectively, assuming QCD production of  $X'_{2/3}$  and  $T'$  for figures (a) and (b), and  $X_{5/3}$  in figure (c).

The analysis performed in this case is the same as described in Section 6.3.3 using the same MVA. The limits are then computed in the two different ways as previously described to assess the impact of the heavy gluon in the sensitivity of the analysis. The results are shown in Figure 6.13 in which the same color coding has been used: the solid purple line represents the mass limits at a 95% C.L. when the heavy-gluon is properly considered for each mass and the dotted green line represents the mass limit assuming that the only change introduced by the heavy gluon is a rescale of the cross-section. The dashed blue line shows the QCD only mass limit for  $M_{X'_{2/3}}$ .

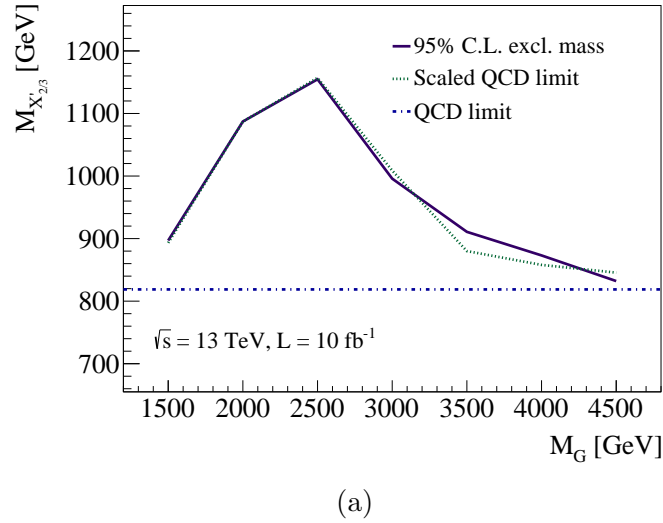
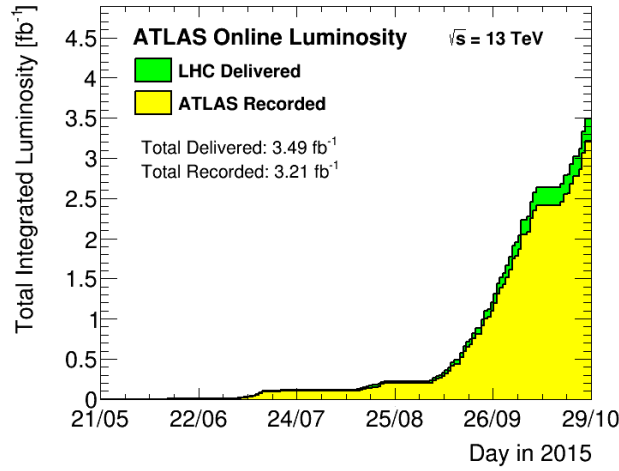


Figure 6.13: The expected lower limits at 95% C.L. on the  $X'_{2/3}$  mass as a function of  $M_G$  at  $\sqrt{s} = 13$  TeV and an integrated luminosity of  $10 \text{ fb}^{-1}$ . The color scheme is the same as in Figure 6.10 and explained in the text.

There are two points worth emphasising. The first one is about the impact of the heavy gluon in the pair production mechanism at 13 TeV. The expected limit assuming only QCD production and rescaling the production cross-section accordingly for each pair of masses is perfectly compatible with what is found when properly introducing the heavy gluon in the signal hypothesis. This points out that the design of the current analysis are likely insensitive to the presence of a heavy gluon at least in the early days of Run-2. The second important point is that, even with only an integrated luminosity of  $10 \text{ fb}^{-1}$ , which is likely to be achieved in the next year (Figure 6.14 shows the current integrated luminosity when this thesis was written), the limits on the mass of both the VLQ and the heavy gluon show a significant improvement. The expected limits on the mass of the VLQ with

early Run-2 data are in the range of  $\sim 820 - 1160$  GeV, to be compared with the  $\sim 600 - 1000$  GeV seen at  $\sqrt{s} = 8$  TeV. A similar improvement can be expected with respect to the complete current limits presented in Fig. 6.12. Given these two points, it is clear that as soon as the experimental collaborations publish their results on VLQ production during Run-2, they can be easily translated to bounds on the parameter space of models with heavy vectors in the spectrum.



(a)

Figure 6.14: Integrated luminosity collected for stable beams of  $pp$  collision during 2015 at  $\sqrt{s} = 13$  TeV.



---

## Chapter 7

# Conclusions

In the present thesis, three different aspects of high energy physics were covered: a more technical work in the ATLAS Tile Calorimeter, a search for new quarks in multilepton final states using LHC data collected by ATLAS and a phenomenology study on the interpretation of the current searches for vector-like quarks at the LHC.

On the technical side, the work directly related with the detector has been presented in Chapter 4. A proper characterisation of the noise in the Tile Calorimeter is crucial in order to improve the object reconstruction. All the studies were performed at 50 ns bunch spacing with a few runs at 25 ns, getting a glimpse of the impact it could have on the energy measurements for each cell of the calorimeter. Now that Run-2 has started and 25 ns has become the default bunch spacing in the LHC, the conditions under which the ATLAS detector will operate will be even harder. A new noise estimator was presented, the quantiles estimator. The work is now in progress in order to include the noise estimation provided by this method into MC simulated events to evaluate the impact and jet reconstruction.

The second part covered in this thesis is the analysis of data collected by the ATLAS detector. As previously mentioned, the SM represents a triumph for our understanding of elementary particles and their interactions. Nonetheless, one could not do anything but ask for an answer to the unsolved problems which still remain. As such, the search for vector-like quarks has been developed, in order to explore a new phase space region opened with the 8 TeV data. Different parts of the analysis were developed, from the signal modelling and validation for the pair production mechanism to the final results published in November of 2014. The analysis was divided in the dilepton and trilpeton channel and it was decided to approach, for the first time, the single production mechanism. The analysis is part of the ATLAS

strategy for VLQ searches and it is the only one covering the region with high branching ratios to the  $Z$  decay. The final results show that no significant deviation from the SM expectation is found for the data analysed. The combination of the dilepton and trilepton channel was used to report the final results. For the pair production mechanism observed (expected) limits were set at a 95% C.L. for the vector-like  $T$ , excluding masses below 655 GeV (625 GeV) for singlet and 735 GeV (720 GeV) for doublet, and excluding masses of the vector-like  $B$  below 685 GeV (670 GeV) for singlet and 755 GeV (755 GeV) for doublet. These results are summarised along with the rest of the analyses in Figures 5.28 and 5.27.

Various tools have been developed under the scope of this thesis. Particularly of note are the analysis interface and common tools used by the group involved in the multilepton  $Z$ -tag analysis as well as other vector-like quark searches within the ATLAS Collaboration. As mentioned before, the signal production and validation were done at the beginning of the analysis for the pair production mechanism. Afterwards, the data pre-processing tools were developed to produce the analysis inputs used for the rest of the groups working on the analysis. In order to correct the  $Z$ +jets background as explained in the text, a correction tool was developed and maintained. The task of producing the summary results (Figures 5.28 and 5.27) was also completed and the code to analyse the results from the different groups and produce the summary results was developed becoming the official tool for the 2D branching ratio figures.

The third part of the work of this thesis is the phenomenological study presented in Chapter 6. Along with vector-like quarks, which are predicted by CHM as fermion resonances, heavy vector resonances, heavy gluons, may appear as well. The question is if the eventual kinematic effects of a heavy gluon in the pair production of vector-like quarks are significant enough to have an impact on current searches or if it might be the case for early Run-2 data. Signal hypotheses including heavy gluon and only QCD pair production, as well as the proper combination of both, including the appropriate interference, were generated. After an intensive study of different kinematical distribution before and after detector simulation it was found that the kinematical difference that can be observed at the parton level almost disappear after detector simulation is introduced. Even with a more sophisticated analysis in which a neural network was used to exploit the small differences between several inputs distributions, the effect of a heavy gluon was not observed in the final results. This implies that current VLQ results can be used to set bounds on the masses of

VLQs and heavy gluons expected by the model, which was also done.

It was shown that for the masses and energies considered, a significant amount of the events presented boosted topologies. It might be the case that using jet substructure variables or larger jets some of the differences might be recovered, although it was beyond the scope of the work. Nonetheless it is important to keep an eye on boosted topologies, will surely become relevant in Run-2 given the increased center-of-mass energy and the range of masses that will be considered ( $\sim 1$  TeV).

Run-2 has already started but it is still in the beginning. For now, masses of vector-like quarks have been excluded between  $\sim 750$  GeV and  $\sim 1$  TeV. Vector-like quarks appear in models trying to solve the naturalness problem and fix the quadratic divergences of the Higgs sector. Nonetheless, this remains interesting if VLQ masses are not in the multi-TeV scale. With the higher energies achieved in Run-2 the gap between the current excluded masses and the  $\sim 2$  TeV edge might be covered. If that is the case, then they will be either discovered or stronger limits will be set on their masses.

In summary, the LHC is a powerful machine which is able to answer questions such as *is the Higgs boson there?*, but it is also capable of inspiring new ones as new horizons are being explored.



---

## Appendix A

### Noise vs $\langle \mu \rangle$ using the quantiles estimator

In this appendix a complete overview of the noise as a function of  $\langle \mu \rangle$  with the quantiles estimator is shown. Figures [A.1](#), [A.2](#), [A.3](#) and [A.4](#) show the noise for the cells in the layer A, BC, D and E (gap/crack scintillating cells) respectively.

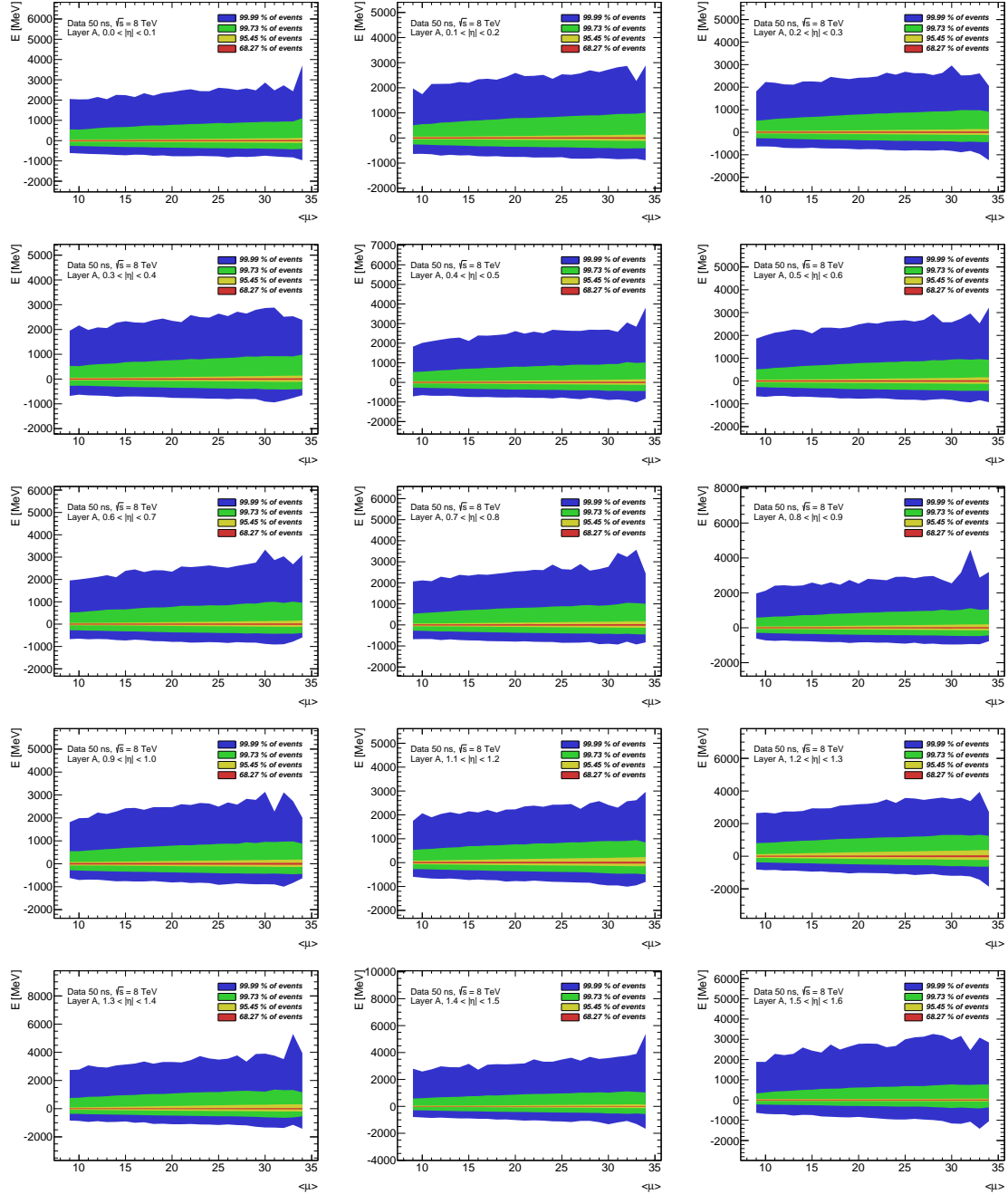


Figure A.1: Noise dependence with  $\langle\mu\rangle$  for the different TileCal cells in the layer A using the quantiles estimator. Data collected a 8 TeV with a bunch-spacing of 50 ns has been analysed. The information for different partitions and modules has been integrated.

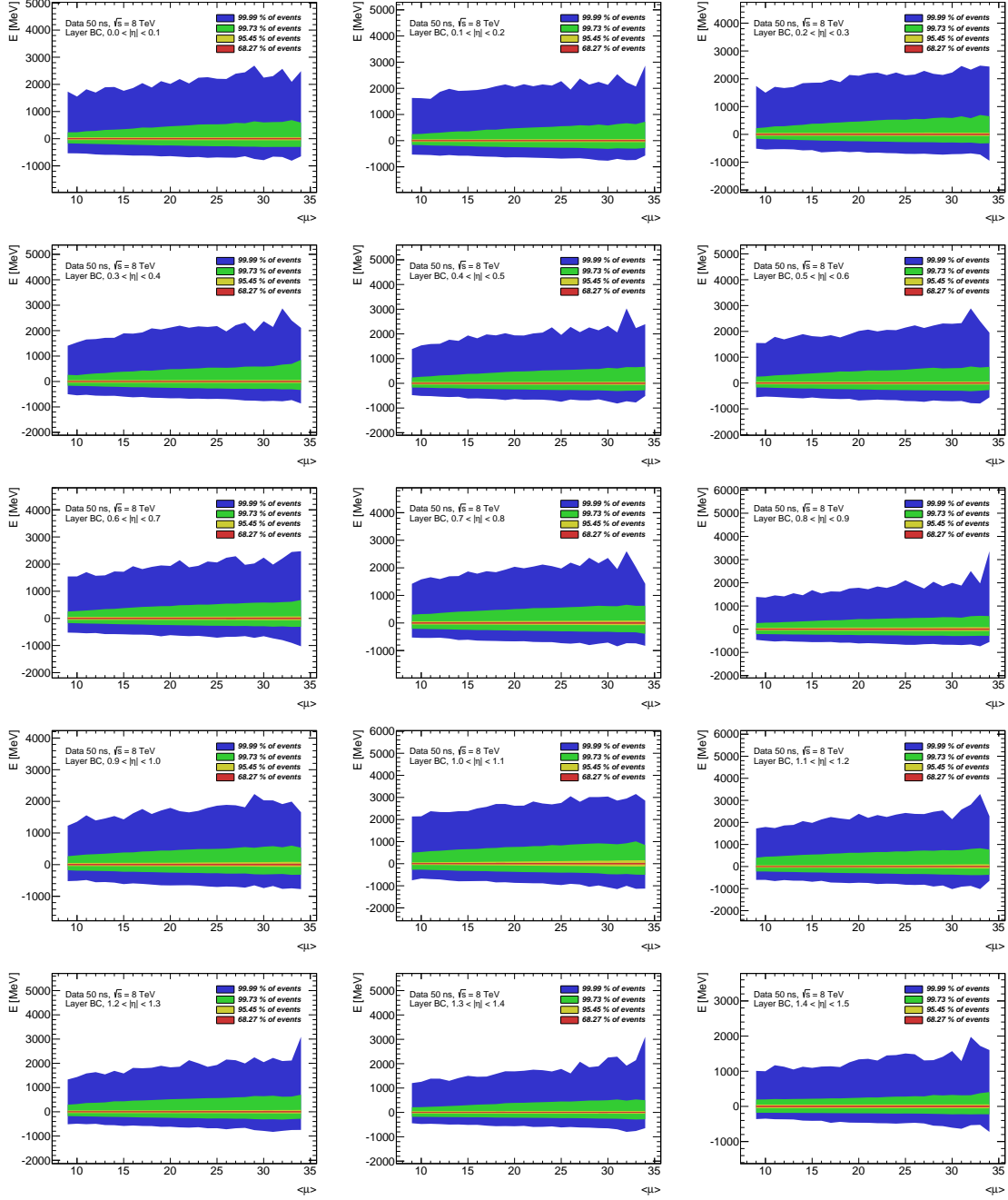


Figure A.2: Noise dependence with  $\langle \mu \rangle$  for the different TileCal cells in the layer BC using the quantiles estimator. Data collected at 8 TeV with a bunch-spacing of 50 ns has been analysed. The information for different partitions and modules has been integrated.

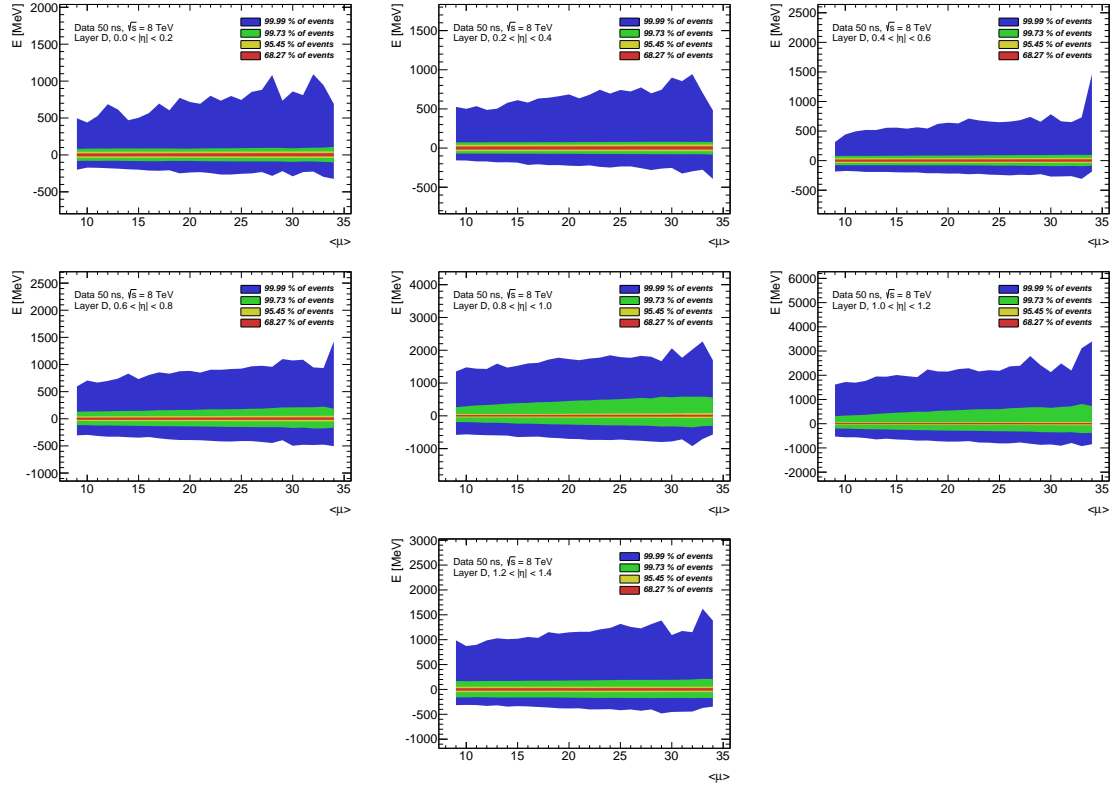


Figure A.3: Noise dependence with  $\langle\mu\rangle$  for the different TileCal cells in the layer D using the quantiles estimator. Data collected a 8 TeV with a bunch-spacing of 50 ns has been analysed. The information for different partitions and modules has been integrated.

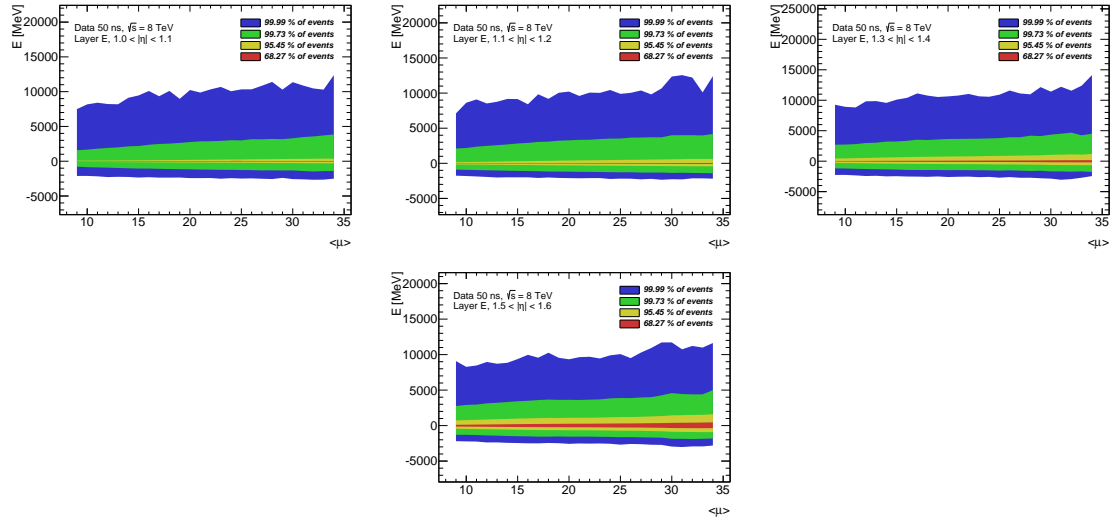


Figure A.4: Noise dependence with  $\langle\mu\rangle$  for the different TileCal cells in the special layer using the quantiles estimator. Data collected a 8 TeV with a bunch-spacing of 50 ns has been analysed. The information for different partitions and modules has been integrated.



---

## Appendix B

# Data vs. MC comparison using the ALPGEN generator

As mentioned in Chapter 5 the default generator used for the  $Z$ +jets background modelling was **Sherpa**. This decision was made based on a better description of the observed data (as can be seen in the corrections applied to the  $N_{\text{tag}}$  distribution in Figure 5.12) and the higher statistic available for that generator. Nonetheless,  $Z$ +jets samples generated with **ALPGEN** were compared in parallel to check that everything was consistent. In this appendix the comparison between the observed data and the expected background contribution using the **ALPGEN** samples is shown at different stages of the analysis. The corrections shown in Chapter 5 were applied as well to the **ALPGEN** samples. The invariant mass of the  $Z$  boson candidate is shown in Figure B.1 for the  $N_{\text{tag}} = 1$  (a) and  $N_{\text{tag}} \geq 2$  (b) regions after the first correction was applied. The  $p_T(Z)$  distribution is shown in Figure B.2 for both control regions  $N_{\text{tag}} = 1$  (a) and  $N_{\text{tag}} \geq 2$  (b) before the  $p_T(Z)$  correction is applied. The same distribution is shown in Figure B.3 after the  $p_T(Z)$  correction is applied. Finally, the  $H_T(\text{jets})$  ( $m(Zb)$ ) distribution is shown in the top (bottom) panels in Figure B.4 for the  $N_{\text{tag}} = 1$  and  $N_{\text{tag}} \geq 2$  categories in the left and right panels respectively. In these final stages of the analysis the lack of statistic in the **ALPGEN** sample is clearly seen.

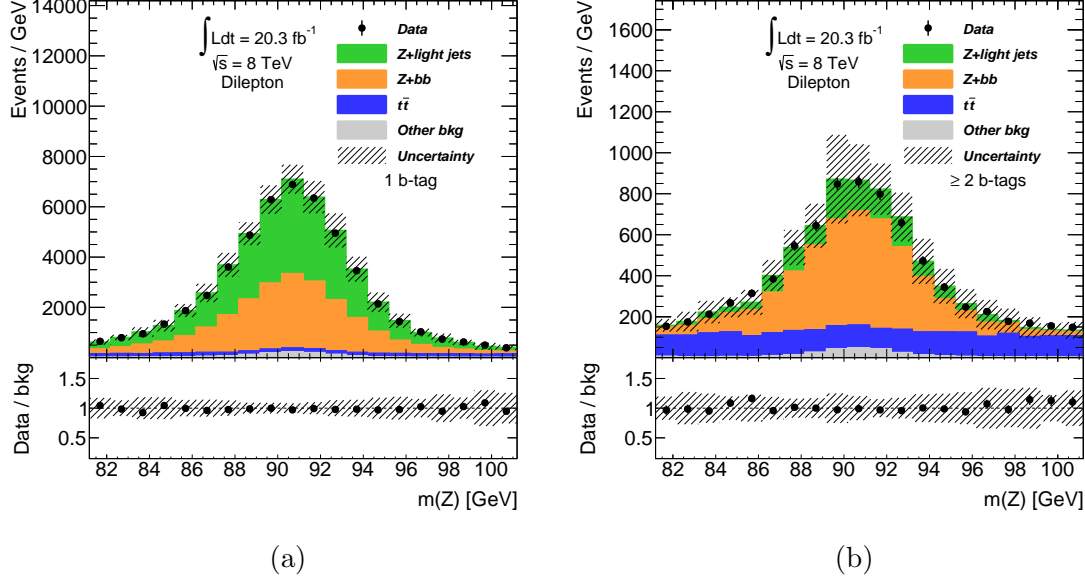


Figure B.1: Invariant mass distribution for the dilepton channel in the  $N_{\text{tag}} = 1$  region (a) and  $N_{\text{tag}} \geq 2$  region (b). The ALPGEN generator has been used to generate the  $Z$ +jets background samples.

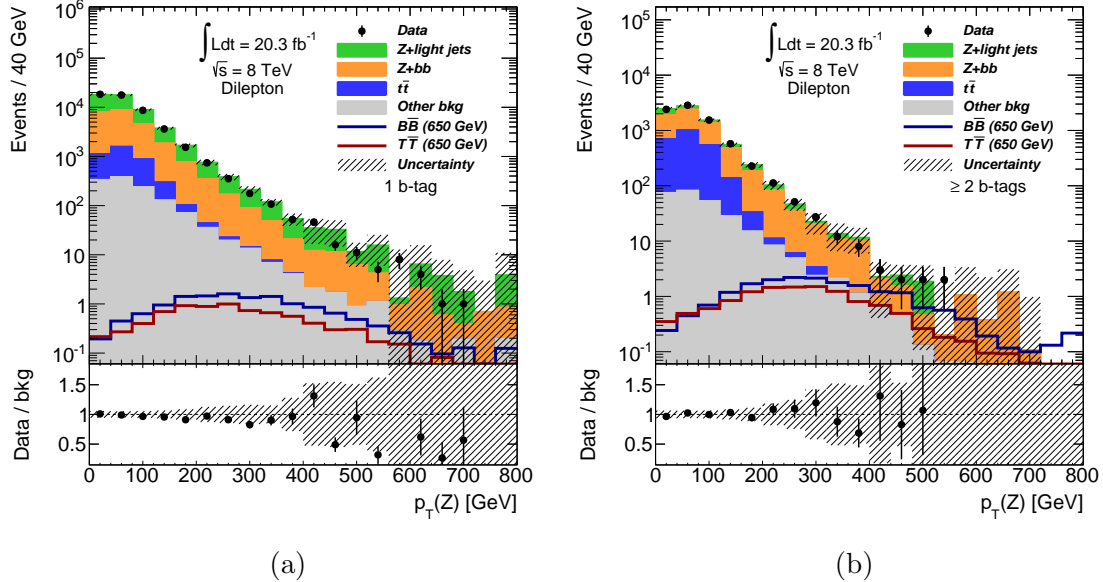


Figure B.2: The  $p_T(Z)$  distribution is shown for the  $N_{\text{tag}} = 1$  region (a) and the  $N_{\text{tag}} \geq 2$  region (b) before the  $p_T(Z)$  correction is applied. The ALPGEN generator has been used to generate the  $Z$ +jets background samples.

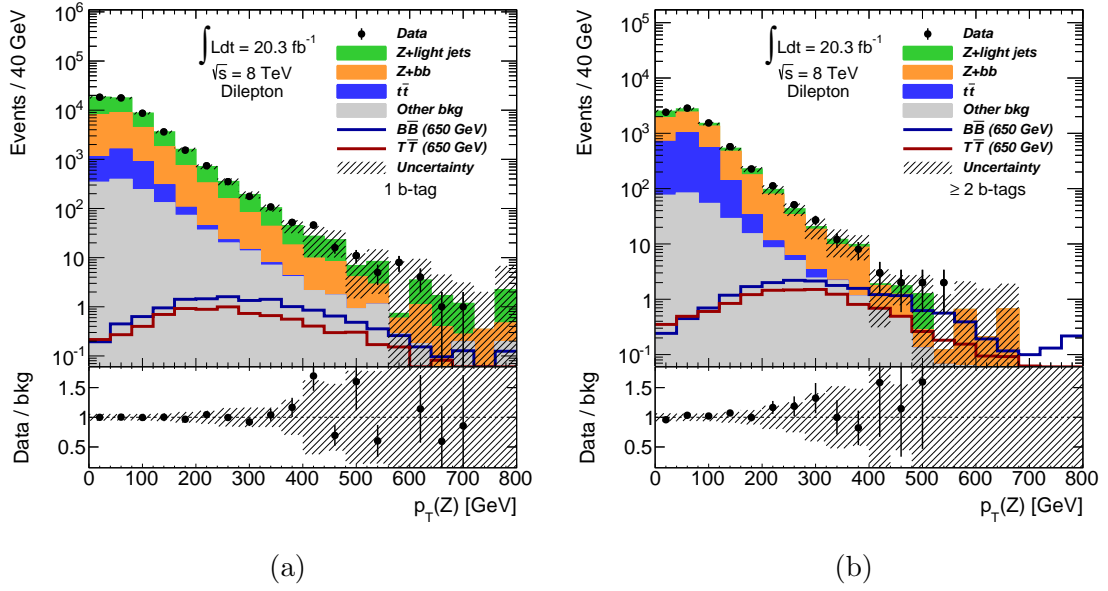


Figure B.3: The  $p_T(Z)$  distribution is shown for the  $N_{\text{tag}} = 1$  region (a) and the  $N_{\text{tag}} \geq 2$  region (b) after the  $p_T(Z)$  correction is applied. The ALPGEN generator has been used to generate the  $Z$ +jets background samples.

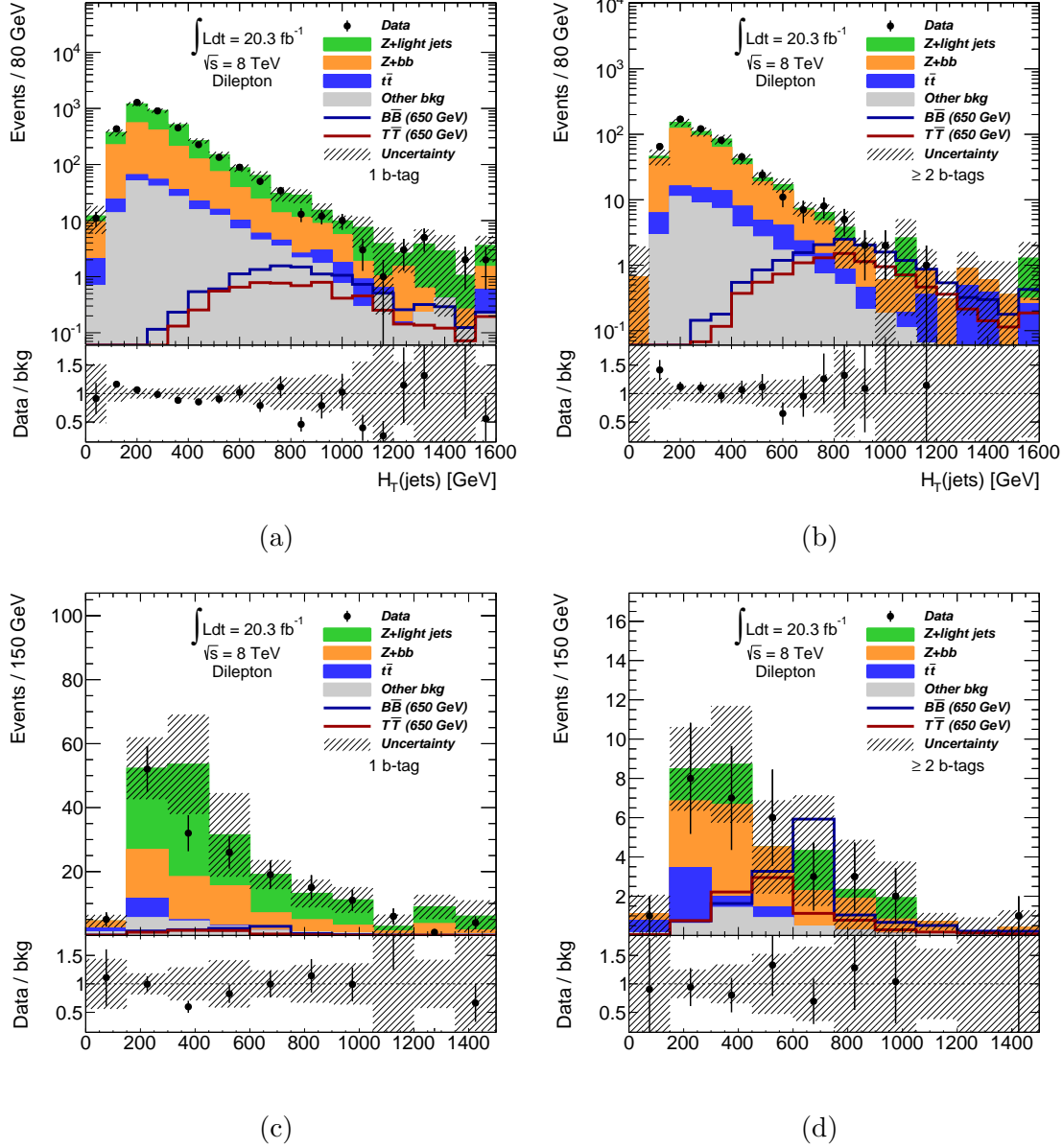


Figure B.4: The  $H_T(\text{jets})$  (top) and  $m(Zb)$  (bottom) distributions are shown for  $N_{\text{tag}} = 1$  (left) and  $N_{\text{tag}} \geq 2$  (right). The  $m(Zb)$  distribution in the  $N_{\text{tag}} \geq 2$  region is the final discriminant used for hypothesis testing. The ALPGEN generator has been used to generate the Z+jets background samples.

---

# Appendix C

## Software tools

Different software tools have been developed to produce the results obtained in this thesis. In this appendix a quick overview of the most important ones is presented

### New Physics Analysis package

The main tool which has been developed is the analysis package *New Physics Analysis* (NPA). NPA is an analysis framework developed as a C++ interface which has been built on top of LipCbrAnalysis [158], the previous analysis framework used in the group. The LipCbrAnalysis package was conceived and designed several years ago and several parts of it could be improved and made more flexible. This is the idea behind NPA.

Every analysis framework is based on three main steps:

- Initialisation: the first phase of an analysis framework is to initialise every tool needed and prepare the input files which are going to be read.
- Event loop: the main part of any analysis is the loop over all the events which are going to be analysed. It is usually the one that takes most of the time and where event reconstruction or kinematical corrections are applied.
- Finalisation: this is the wrap-up phase of the analysis. Outputs are prepared and organised and everything is cleaned before closing the application.

The initialisation part of NPA is still very similar to the one done in the previous package. First, all the inputs for the analysis (in this thesis all the inputs are ROOT [159] TTree objects inside a TFile structure) are organised in samples and the information for each sample is listed in a configuration file (i.e. name, number

of events, cross-section etc.). Each set of input files is grouped in samples for data and MC simulation and each MC sample is normalised to the cross section supplied in the configuration file. A set of parameters are used to configure the code such as luminosity, name of the output file and any other parameter that the user wants to introduce. Each time the analysis run, one of the samples defined will be used. Several functions allow the user in the initialisation phase to configure the code as desired before the event loop.

The event loop is mainly a set of decisions with a boolean outcome which decides if a given event passes a given selection, or cut. A per-event variable will keep track of which cuts a given event has passed. In *NPA* all information observables defined by the user are stored in histograms which are also filled in the event loop phase. Each event will fill an entry on the desired histogram corresponding to each of the cuts passed by the event. In this way, after the analysis is finished, the desired observable can be accessed at each level of the selection.

The finalisation part recovers all information collected in the event loop such as event yields for each cut and histograms and produces three output files. Two text files containing the errors (if any) and a text output file with the log, including event yields. Another TFile structure is produced with all the histograms declared in the code properly filled.

*NPA* is designed to be independent of the type of inputs the code will analyse. It should be irrelevant, from the analysis design point of view, where a jet or an electron comes from. It shouldn't matter if the input come from official ATLAS production or from a Delphes simulation (the same *NPA* code framework was used for the analyses presented in Chapter 5 and 6). In order to achieve this independence, the framework does not read the information directly from the input but from a *reader* which can be designed to translate the inner structure of the input data structure into the objects used by the framework. This structure can still be improved and default readers for commonly used input types are underway.

There are two main features included in *NPA*. The first one is a set of tools written in *Ruby*, on a object oriented approach, that have been developed in order to automatise the process of running a given analysis based on *NPA*. The *NPAJobManager* class receives an array of *Jobs* objects (which can be defined in different formats, even in a configuration json file) which define all the different runs for a given analysis. An example of this array might be the next piece of code.

```

1  config = [{:ID => "ttbar_NLO_ee",
2             :sample => 1,
3             :output => "outputpath/NPAOutput_ttbar_NLO_ee",
4             :options => {"ZChannel"=>"ee", "TMVA"=>"MLPBNN"}
5  },
6  {:ID => "Zbb_ee",
7       :sample => 3,
8       :output => "outputpath/NPAOutput_Zbb_ee",
9       :options => {"ZChannel"=>"ee", "TMVA"=>"MLPBNN"}
10 }]

```

Here two different runs have been defined. One of them with the ID `ttbar_NLO_ee` will run over the sample number 1, which should have been configured before to be  $t\bar{t}$  at NLO, and another run with ID `Zbb_ee` will run over the sample 3, configured to be  $Z+bb$ . Both of them receive the output name and a set of options. The options are the remaining parameters read by the code, in this case `ZChannel = ee` is an option defined by the user which is used to accept only dielectronic decays of the  $Z$  boson candidate if present.

Once everything has been configured, an `NPAJobManager` will take care of running every NPA job.

```

1  manager = NPAJobManager.new(:command => "RunVLQAnalysis")
2  manager.load(config)
3  # Only ttbar
4  manager.filter(["ttbar"])
5  manager.submit(:cluster => "PBS", :flags => {"-V"=>""},
6               :max => 10)
7  manager.generate_xml

```

Notice that the `NPAJobManager` is created with a *command* option, this defines the executable of the analysis. Then the jobs are loaded and, optionally, a filtering can be applied, selecting only the jobs whose ID matches the regexp, or list of regexps, introduced. In this case only the  $t\bar{t}$  sample will be run. This makes sense because in an usual analysis hundreds of jobs might be defined and this will rarely be modified, nonetheless it is common to run over only a subset of samples, which can be done just by defining a filter. The `submit` command receives some parameters such as what cluster type is going to be used and any flags for the

submission. Different options can be defined like the possibility of having a dry run using the `test` flag or how many jobs will run at the same time with the `max` option. NPAJobManager will track the jobs and submit new ones as needed. It is important to point out that the NPAJobManager class is capable of support any cluster type since a custom Cluster object can be defined by the user following simple rules.

Besides this class, several other classes helps to manage the analysis as well as to produce analysis skeletons which are useful to start developing an analysis based on NPA

The other new feature added to NPA is a set of tools designed to produce plots based on the histograms defined in the analysis. Hundreds of histograms have been produced for this thesis in order to check that everything is under control or to try to exploit some kinematical signature. It is obvious that the production of so many histograms needs to be automatised. A set of C++ classes take care of this task in NPA. A function *generate.xml* is defined in NPAJobManager. This function will produce an XML configuration file which defines all the output files generated by the analysis for all the samples and group them with their corresponding systematic variations if defined.

```

1 <?xml version='1.0'?>
2 <Inputs>
3   <Configuration InputDir='output_8TeV'/>
4     <Input ID='ttbar_NLO_ee' File='NPAOutput_ttbar_NLO_ee.root'>
5       <ShapeSyst ID="JES"
6         FileDown="VLQAnalysis_BBS_650_AF_elemu_jes_down.root"
7         FileUp="VLQAnalysis_BBS_650_AF_elemu_jes_up.root" />
8       <ShapeSyst ID="JER"
9         File="VLQAnalysis_BBS_650_AF_elemu_jer.root"/>
10    </Input>
11    <Input ID='Zbb_ee' File='NPAOutput_Zbb_ee.root'>
12      <ShapeSyst ID="JES"
13        FileDown="VLQAnalysis_BBS_650_AF_elemu_jes_down.root"
14        FileUp="VLQAnalysis_BBS_650_AF_elemu_jes_up.root" />
15      <ShapeSyst ID="JER"
16        File="VLQAnalysis_BBS_650_AF_elemu_jer.root"/>
17    </Input>
18 </Inputs>

```



Here, all the job outputs are organised to be used as inputs to produce the different plots. As can be seen, each output of the analysis is defined as an input for the plotting tools using the **Input** node. Each input can have a set of systematic variations associated to it. In this case, the up and down variations are defined for the case of the JES systematic. Nonetheless a systematic can be defined as one-sided as is the case of the JER uncertainty or even just a global variation. The amount of output files generated in a complex analysis can be very large, this is the reason why this XML file is automatically generated and the ID of each input and systematic is the one defined in the run configuration.

Each set of inputs can be organised as desired in the last configuration file, which is created by the user. This file will contain the sample structure used to produce the plots.

```

1 <?xml version="1.0" encoding="ISO-8859-1"?>
2 <Tables>
3   <Configuration OutputDir="newrew/2tag/Plots/" />
4   <Samples Name="$ee+\mu\mu$" InputsFile="inputs.xml">
5     <!-- Data -->
6     <Sample ID="Data" Name="Data" Type="data">
7       <Input ID="Data_elemu" />
8     </Sample>
9
10    <!-- Signal -->
11    <Sample ID="BBS" Name="BBS (650)" Type="signal">
12      <Input ID="BBS_650_AF_elemu" />
13    </Sample>
14    <Sample ID="TTS" Name="TTS (650)" Type="signal">
15      <Input ID="TTS_650_AF_elemu" />
16    </Sample>
17
18    <!-- Background -->
19    <Sample ID="Zjets_light" Name="$Z$+light jets" Type="bck">
20      <Input ID="Sherpa_Zlight_elemu" />
21      <Input ID="Sherpa_Zlight_sliced_elemu" />
22      <Input ID="Sherpa_Zcc_elemu" />
23      <Input ID="Sherpa_Zcc_sliced_elemu" />

```

```

24     </Sample>
25     <Sample ID="Zjets_bb" Name="$Z+bb$" Type="bck">
26         <Input ID="Sherpa_Zbb_elemu"/>
27         <Input ID="Sherpa_Zbb_sliced_elemu"/>
28     </Sample>
29     <Sample ID="ttbar" Name="$t\bar{t}$" Type="bck">
30         <Input ID="ttbar_elemu"/>
31     </Sample>
32     <Sample ID="Other" Name="Other Bck." Type="bck">
33         <Input ID="ttbarX_elemu"/>
34         <Input ID="singletop_elemu"/>
35         <Input ID="Diboson_elemu"/>
36         <Input ID="Triboson_elemu"/>
37         <Input ID="Wjets_elemu"/>
38     </Sample>
39 </Samples>

```

Each node `Sample` will be a sample in the final plot and is composed by one or more inputs. Different NPA C++ classes will read the configuration files and provide easy to use objects which identify each sample as well as derive the error bars including systematic variations. An example code to obtain histograms is as follows

```

1     NPASampCol *samples = new NPASampCol("xmlfile.xml",
2                                         "Plots",
3                                         "containerName");
4     TGraphAsymmErrors systerr;
5     TH1D *he = samples->getHistogram("histoname",systerr);

```

In the previous code the histogram `histoname` has been obtained from a sample which has been defined before with ID `sampleID`. All the inputs included in the defined sample will be taken into account in the final histogram. The object `systerr` is a `TGraph` (a set of  $x$  and  $y$  coordinates) with asymmetric errors corresponding to the error band that can be drawn on top of the histogram. This error band will already include all systematic variations defined previously. The same tools can be used to produced latex tables with event yields and systematic variations for each sample.

During this thesis several approaches have been followed to produce the different results (plots and tables) and the use of this new feature of **NPA** has proven to be highly time efficient. There are several XML files that need to be created but it rarely takes more than half an hour and any change that want to be done later on is just a matter of modifying some configurations lines which saves a lot of time.

## Common tools for vector-like quark searches

A couple of tools have been developed for the different teams working in the different vector-like quark searches, all of them related to the final results shown.

The first contribution was the setup of a common framework to produce the final results of the analyses and produce the limit plots both in the 2D branching ratio plane and the upper limit on the cross section. The tools to run the jobs were already done but they were focused only in one type of analysis.

During this thesis the tools were changed to make it usable for other teams. A common framework was set up with the idea of just downloading configuration files and automatically compile everything needed and be ready to produce results. The biggest contribution was to the plotting codes, specially the 2D branching ratio scan. Since different analyses were sensitive to different regions it was not trivial to produce a tool that can make the final results look good for every analysis using ROOT. After the final code was developed all the analyses could just plugin their results and obtain the code.

Eventually a new format was designed for the analysis presented in Chapter 5, specifically in Figures 5.25 and 5.24. When the plotting code for this style was developed it was done thinking that other people might need to use it and a general approach was taken when designing the input reading and configuration. At the end this became the common plotting code to produce the published ATLAS results in the search for pair production of VLQ at  $\sqrt{s} = 8$  TeV, as shown in Figure 5.28 and 5.27.

## MadGraph submission tools

In Chapter 6, the **NPA** framework presented before was used to analysed the samples after detector simulation. All the samples were produced using **MadGraph** and several background processes and mass points for signal (including masses of heavy gluon and vector-like quarks) needed to be generated both for 8 and 13 TeV. In this sense it became clear that some tool was needed to produce and organise all the sample.

The idea is similar to the one followed by the NPAJobManager. Written in Ruby, the JobManager class store a set of jobs that will be run by MadGraph and then they are submitted and organised.

```

1 vlq = JobManager.new(:ID => "job_ID", :submax=>30,
2                       :output_path => "output/path")
3 VERSION = "v2_8TeV"
4 tags.each do |key,value|
5   vlq.add_job do |j|
6     j.model = "MCH45_UFO"
7     j.name = value
8     j.procs<<"p p > ht1 ht1~, (ht1 > Z t),(ht1~ > Z t~)"
9     j.procs<<"p p > ht1 ht1~, (ht1 > H t),(ht1~ > H t~)"
10    j.procs<<"p p > ht1 ht1~, (ht1 > Z t),(ht1~ > H t~)"
11    j.procs<<"p p > ht1 ht1~, (ht1 > H t),(ht1~ > Z t~)"
12    # Set options
13    j.opts << "multi_run 10"
14    j.opts << "pythia=ON"
15    j.opts << "delphes=ON"
16    j.set "nevents 30000"
17    if VERSION.include? "8TeV"
18      j.set "ebeam1 4000"
19      j.set "ebeam2 4000"
20    else
21      j.set "ebeam1 6500"
22      j.set "ebeam2 6500"
23    end
24    j.set "drll 0"
25    j.set "drjl 0.2"
26    j.set "ptj 10"
27    j.set "etal 3"
28    #Set files to read
29    j.files << "./delphes_myATLAS2.dat" #DELPHES card
30    j.files << key # Params
31  end
32 end

```

In the previous code a `JobManager` class was defined with a given ID, a maximum number of job submitted of 30 and the output path for the job. After that any job can be added to the manager. The advantage is that it can be done iteratively in order to produce similar **MadGraph** production with slightly modified parameters. For instance, in the code above, the variable `tags` contains different parameter cards (different heavy quark and heavy gluon masses) which are used in each job. For each card, a job,  $j$ , is added and the processes to be generated are indicated in the `procs` member of the Job class. Using the `opts` member the options provided to **MadGraph** can be introduced and using the `set` function different parameters of the `run_card.dat` file can be updated. Finally, any file needed such as a **Delphes** card or the param card (which in this example is the `key` variable in the iterator) can be supplied by the `files` member of the Job class. These tools for **MadGraph** production helped to quickly define the samples to be generated at the same time as they were automatically submitted.



---

# Appendix D

## The $CL_s$ method

In this thesis the  $CL_s$  method has been used for deriving upper limits on the cross section for vector-like quark production.

In a given analysis a test statistic  $q$  has been derived to distinguish between signal and background. Under a given hypothesis  $\mathcal{H}$ , the  $p$ -value is defined as the probability of finding a test statistics  $q$  equally or less compatible with  $\mathcal{H}$  than the measured  $q_{obs}$ .

In a search like the one presented in this thesis there are two hypotheses: background only ( $b$ ) and the hypothesis in which signal is also present ( $s+b$ ). Figure D.1 shows two distribution corresponding to the test statistics in the case of background only and  $s+b$ .

In this case, the  $s+b$  hypothesis distribution,  $f(q|s+b)$ , is shifted to the left, which implies that the  $p$ -value of the  $s+b$  hypothesis is the probability of finding a test statistics  $q$  which is greater than  $q_{obs}$

$$p_{s+b} = \int_{q_{obs}}^{\infty} f(q|s+b) dq. \quad (D.1)$$

One could use this  $p$ -value to evaluate if a given measurement agrees with the  $s+b$  hypothesis. Defining a confidence level CL, a given signal hypothesis is rejected if

$$p_{s+b} < \alpha \quad (D.2)$$

where  $\alpha = CL - 1$ . The value of CL is defined by convention to 95% which implies that the  $s+b$  is rejected for  $p$ -values under this hypothesis smaller than 5%. In other words, if the probability of finding something which is less compatible with the  $s+b$  hypothesis than what has been measured is less than 5% the the signal is

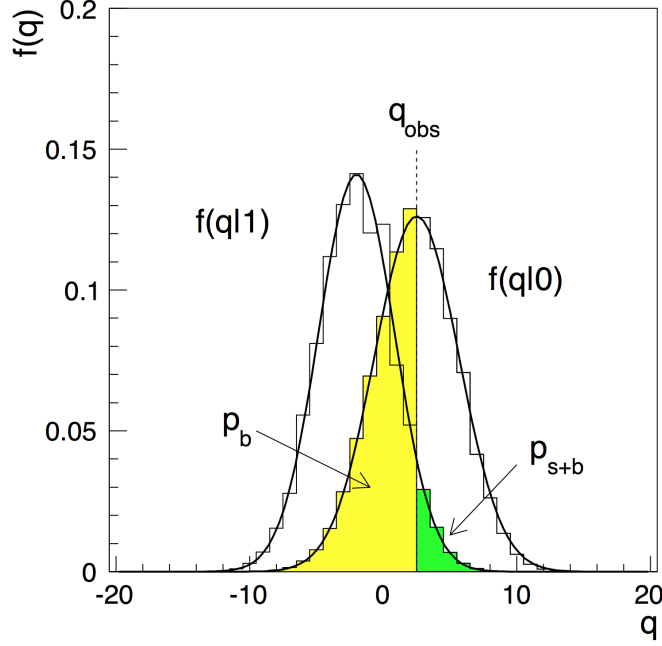


Figure D.1: Distributions of a test statistic  $q$  under the  $s + b$  and  $b$  hypotheses [160].

excluded.

This is what is called the  $CL_{s+b}$  method and seems to make sense. Nonetheless when looking for signal that hasn't been observed before, the natural scenario is that the signal contribution will be small. In this scenario the distributions of the tests statistics look like Figure D.2.

In this case the analysis has no sensitivity (or very small) to the signal which is being considering. If one derive  $p_{s+b}$  in this case, it might lead to a exclusion of the signal model for a small downward fluctuation of the observed numbers of events. In fact, imagine that we now know that signal is there and we are no longer trying to exclude it, we are trying to discover it. In a discovery, the null hypothesis is  $s + b$  while the hypothesis which is being tested is  $b$ . In this case, when the analysis has no sensitivity to the signal, it might even happen that  $p_b < \alpha$ , or in other words, the background hypothesis will be excluded, i.e. the SM hypothesis will be excluded.

It is obvious that in this cases one should not claim an exclusion and therefore the method should penalise the little sensitivity of the analysis. This is when the  $CL_s$  method was introduced. Now, the  $p$ -value of the background only hypothesis is introduced as

$$p_b = \int_{q_{obs}}^{\infty} f(q|b) dq. \quad (D.3)$$

For analysis with low sensitivity  $s \ll b$  and therefore  $s + b \sim b$ . In this case,  $p_b$



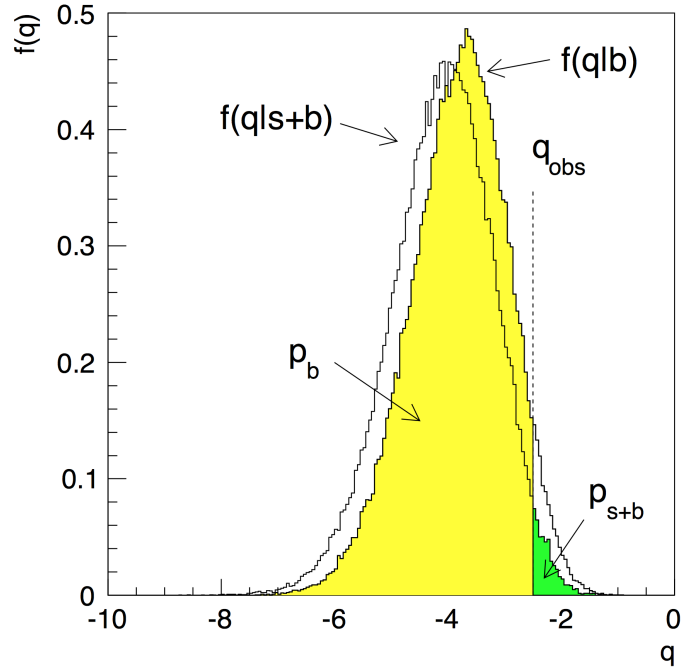


Figure D.2: Distributions of a test statistic  $q$  under the  $s + b$  and  $b$  hypotheses in an example with little sensitivity to the signal model [160].

will become larger for small values of  $p_{s+b}$ . An effective  $p$ -value is introduced,  $CL_s$ , defined by

$$CL_s \equiv \frac{p_{s+b}}{1 - p_b}. \quad (\text{D.4})$$

The  $CL_s$  value is by construction always greater than  $p_{s+b}$  which implies that all models excluded by  $CL_s$  would have been excluded by  $CL_{s+b}$ . Nonetheless, the term  $1 - p_b$  penalises the cases for which both distributions  $s + b$  and  $b$  are very similar. On the other hand, if the two distributions are fairly separated the term  $1 - p_b$  will be smaller and the effective  $p$ -value is similar to what is obtained by  $CL_{s+b}$ .

In summary, for two hypothesis  $s + b$  and  $b$  the distribution of the tests statistics  $q$  is obtained. The signal model is excluded at a 95% CL if  $CL_s < 0.05$ .



---

## Appendix E

# ATLAS summary results for all channels

Figure E.1 and E.2 show the lower bounds to the mass of vector-like  $B$  and  $T$  respectively in a 2D plane for all possible branching ratios. Results from all analyses searching for pair production of vector-like quarks, each of them covering a region of the branching ratio plane, published by the ATLAS Collaboration are included.

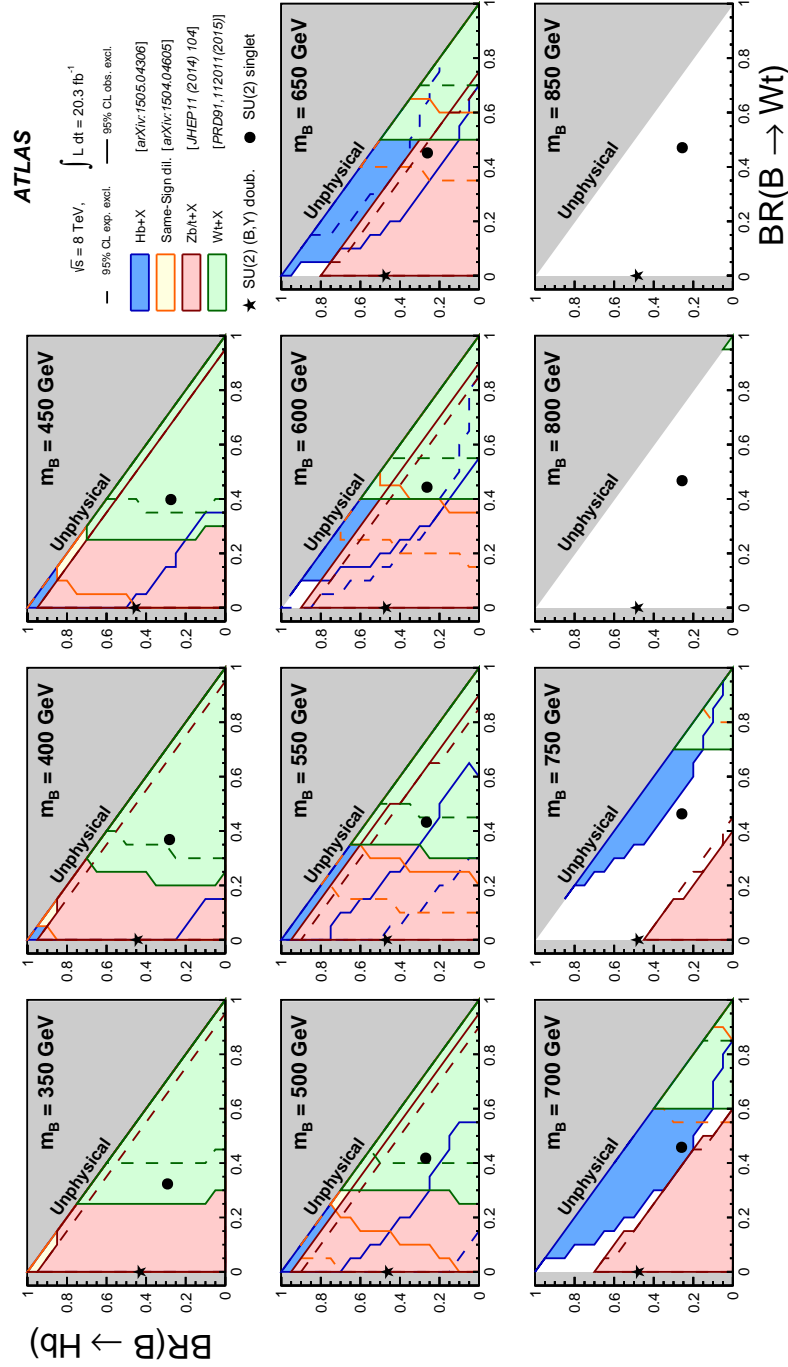


Figure E.1: The lower bound on the mass of the vector-like  $B$  for all searches for pair production of vector-like  $B$  published by the ATLAS Collaboration. Different colours represent the different analyses which are sensitive to different decay modes. The filled area represents the observed limits while the dashed line represent the expected limits.

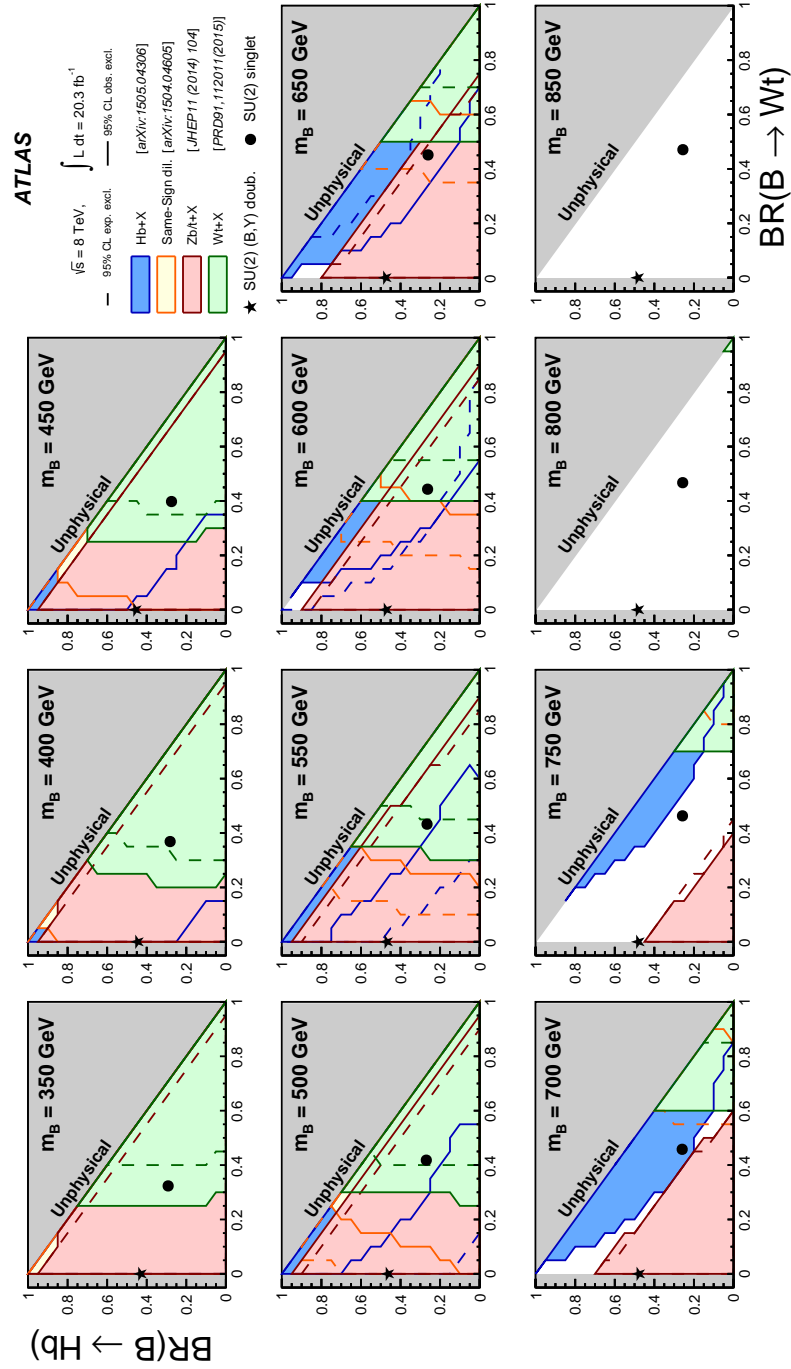


Figure E.2: The lower bound on the mass of the vector-like  $B$  for all searches for pair production of vector-like  $B$  published by the ATLAS Collaboration. Different colours represent the different analyses which are sensitive to different decay modes. The filled area represents the observed limits while the dashed line represent the expected limits.



---

# Bibliography

- [1] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes and R. P. Hudson, *Experimental test of parity conservation in beta decay*, *Phys. Rev.* **105** (Feb, 1957) 1413–1415.
- [2] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, *Phys. Rev. Lett.* **13** (1964) 321–323.
- [3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, *Phys. Rev. Lett.* **13** (1964) 508–509.
- [4] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, *Global Conservation Laws and Massless Particles*, *Phys. Rev. Lett.* **13** (1964) 585–587.
- [5] S. Weinberg, *A Model of Leptons*, *Phys. Rev. Lett.* **19** (1967) 1264–1266.
- [6] S. L. Glashow, *Partial Symmetries of Weak Interactions*, *Nucl. Phys.* **22** (1961) 579–588.
- [7] **ATLAS** Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012), no. CERN-PH-EP-2012-218 1–29 [[1207.7214](#)].
- [8] **CMS** Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012), no. CMS-HIG-12-028, CERN-PH-EP-2012-220 30–61 [[1207.7235](#)].
- [9] **DELPHI, OPAL, LEP Electroweak, ALEPH, L3** Collaboration, *Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP*, *Phys. Rept.* **532** (2013), no. CERN-PH-EP-2013-022 119–244 [[1302.3415](#)].

- [10] **SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3 Collaboration**, *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006), no. SLAC-R-774 257–454 [[hep-ex/0509008](#)].
- [11] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, R. Kogler, K. Moenig, M. Schott and J. Stelzer, *The Electroweak Fit of the Standard Model after the Discovery of a New Boson at the LHC*, *Eur. Phys. J.* **C72** (2012), no. DESY-12-154 2205 [[1209.2716](#)].
- [12] F. del Aguila and J. de Blas, *Electroweak constraints on new physics*, *Fortsch. Phys.* **59** (2011) 1036–1040 [[1105.6103](#)].
- [13] J. de Blas, *Electroweak limits on physics beyond the Standard Model*, *EPJ Web Conf.* **60** (2013) 19008 [[1307.6173](#)].
- [14] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. C. DeStaebler, J. Drees, L. W. Mo and R. E. Taylor, *Observed Behavior of Highly Inelastic electron-Proton Scattering*, *Phys. Rev. Lett.* **23** (1969), no. SLAC-PUB-0650 935–939.
- [15] E. D. Bloom *et. al.*, *High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees*, *Phys. Rev. Lett.* **23** (1969), no. SLAC-PUB-0642 930–934.
- [16] M. L. Perl *et. al.*, *Evidence for Anomalous Lepton Production in e+ - e- Annihilation*, *Phys. Rev. Lett.* **35** (1975), no. SLAC-PUB-1626, LBL-4228 1489–1492.
- [17] S. W. Herb *et. al.*, *Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions*, *Phys. Rev. Lett.* **39** (1977), no. FERMILAB-PUB-77-058-EXP, PRINT-77-0640 (COLUMBIA), FERMILAB-PUB-77-058-E 252–255.
- [18] **TASSO Collaboration**, *Evidence for Planar Events in e+ e- Annihilation at High-Energies*, *Phys. Lett.* **B86** (1979), no. DESY-79-53 243.
- [19] **PLUTO Collaboration**, *Jet Analysis of the  $\Upsilon$  (9.46) Decay Into Charged Hadrons*, *Phys. Lett.* **B82** (1979), no. DESY-78-71 449.



- [20] **UA2** Collaboration, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti- $p$   $p$  Collider*, *Phys. Lett.* **B122** (1983), no. CERN-EP-83-25 476–485.
- [21] **UA1** Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $s^{*}(1/2) = 540$ -GeV*, *Phys. Lett.* **B122** (1983), no. CERN-EP-83-13 103–116.
- [22] **D0** Collaboration, *Search for high mass top quark production in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV*, *Phys. Rev. Lett.* **74** (1995), no. FERMILAB-PUB-94-354-E, FSU-SCRI-95-48 2422–2426 [[hep-ex/9411001](#)].
- [23] **CDF** Collaboration, *Observation of top quark production in  $p\bar{p}$  collisions*, *Phys. Rev. Lett.* **74** (1995), no. FERMILAB-PUB-95-022-E, CDF-PUB-TOP-PUBLIC-3040, ANL-HEP-PR-95-44 2626–2631 [[hep-ex/9503002](#)].
- [24] **DONUT** Collaboration, *Observation of tau neutrino interactions*, *Phys. Lett.* **B504** (2001), no. FERMILAB-PUB-00-335-E 218–224 [[hep-ex/0012035](#)].
- [25] R. Davis, D. S. Harmer and K. C. Hoffman, *Search for neutrinos from the sun*, *Phys. Rev. Lett.* **20** (May, 1968) 1205–1209.
- [26] **ATLAS, CMS** Collaboration, *Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC  $pp$  collision data at  $\sqrt{s} = 7$  and 8 TeV*, .
- [27] A. Djouadi and A. Lenz, *Sealing the fate of a fourth generation of fermions*, *Phys. Lett.* **B715** (2012), no. LPT-ORSAY-12-31, CERN-PH-TH-2012-087 310–314 [[1204.1252](#)].
- [28] **ATLAS** Collaboration, *Analysis of events with  $b$ -jets and a pair of leptons of the same charge in  $pp$  collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, [1504.04605](#).
- [29] **ATLAS** Collaboration, *Search for pair production of heavy top-like quarks decaying to a high- $p_T$   $W$  boson and a  $b$  quark in the lepton plus jets final*

- state at  $\sqrt{s}=7$  TeV with the ATLAS detector, *Phys. Lett.* **B718** (2013), no. CERN-PH-EP-2012-258 1284–1302 [[1210.5468](#)].
- [30] CMS Collaboration, *Search for pair produced fourth-generation up-type quarks in pp collisions at  $\sqrt{s} = 7$  TeV with a lepton in the final state*, *Phys. Lett.* **B718** (2012), no. CMS-EXO-11-099, CERN-PH-EP-2012-244 307–328 [[1209.0471](#)].
- [31] O. Eberhardt, G. Herbert, H. Lacker, A. Lenz, A. Menzel, U. Nierste and M. Wiebusch, *Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations*, *Phys. Rev. Lett.* **109** (2012), no. TTP12-034 241802 [[1209.1101](#)].
- [32] M. Perelstein, M. E. Peskin and A. Pierce, *Top quarks and electroweak symmetry breaking in little Higgs models*, *Phys. Rev.* **D69** (2004), no. SLAC-PUB-10185 075002 [[hep-ph/0310039](#)].
- [33] O. Matsedonskyi, G. Panico and A. Wulzer, *Light Top Partners for a Light Composite Higgs*, *JHEP* **01** (2013) 164 [[1204.6333](#)].
- [34] R. Contino, L. Da Rold and A. Pomarol, *Light custodians in natural composite Higgs models*, *Phys. Rev.* **D75** (2007), no. UAB-FT-619, ROMA1-1445-2006 055014 [[hep-ph/0612048](#)].
- [35] D. B. Kaplan, *Flavor at SSC energies: A New mechanism for dynamically generated fermion masses*, *Nucl. Phys.* **B365** (1991), no. UCSD-PTH-91-04 259–278.
- [36] R. Contino, T. Kramer, M. Son and R. Sundrum, *Warped/composite phenomenology simplified*, *JHEP* **05** (2007), no. ROMA1-1441-2006 074 [[hep-ph/0612180](#)].
- [37] J. L. Hewett and T. G. Rizzo, *Low-Energy Phenomenology of Superstring Inspired  $E(6)$  Models*, *Phys. Rept.* **183** (1989), no. MAD-PH-446, IS-J-3005 193.
- [38] F. del Aguila and M. J. Bowick, *The Possibility of New Fermions With  $\Delta I = 0$  Mass*, *Nucl. Phys.* **B224** (1983), no. UFTP-82-19 107.

- [39] F. del Aguila, M. Perez-Victoria and J. Santiago, *Effective description of quark mixing*, *Phys. Lett.* **B492** (2000), no. UG-FT-116-00, MIT-CTP-2996 98–106 [[hep-ph/0007160](#)].
- [40] F. del Aguila, M. Perez-Victoria and J. Santiago, *Observable contributions of new exotic quarks to quark mixing*, *JHEP* **09** (2000), no. UG-FT-118-00, MIT-CTP-2997 011 [[hep-ph/0007316](#)].
- [41] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer and M. Prez-Victoria, *Handbook of vectorlike quarks: Mixing and single production*, *Phys. Rev.* **D88** (2013), no. 9 094010 [[1306.0572](#)].
- [42] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, *Phys. Rev.* **D2** (1970) 1285–1292.
- [43] F. del Aguila, L. Ametller, G. L. Kane and J. Vidal, *Vector Like Fermion and Standard Higgs Production at Hadron Colliders*, *Nucl. Phys.* **B334** (1990), no. FTUV/89-27, UAB-FT-221/89, UM-TH-89-17A 1.
- [44] J. A. Aguilar-Saavedra, *Identifying top partners at LHC*, *JHEP* **11** (2009) 030 [[0907.3155](#)].
- [45] **ATLAS** Collaboration, *Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, [1505.04306](#).
- [46] **ATLAS** Collaboration, *Search for vectorlike B quarks in events with one isolated lepton, missing transverse momentum and jets at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *Phys. Rev.* **D91** (2015), no. 11 112011 [[1503.05425](#)].
- [47] **ATLAS** Collaboration, *Search for pair and single production of new heavy quarks that decay to a Z boson and a third-generation quark in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, *JHEP* **11** (2014), no. CERN-PH-EP-2014-188 104 [[1409.5500](#)].
- [48] **ATLAS** Collaboration, *Search for the production of single vector-like and excited quarks in the Wt final state in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, [1510.02664](#).

- [49] **CMS** Collaboration, *Search for pair-produced vector-like  $B$  quarks in proton-proton collisions at  $\sqrt{s} = 8$  TeV*, [1507.07129](#).
- [50] **CMS** Collaboration, *Search for vector-like charge  $2/3$   $T$  quarks in proton-proton collisions at  $\sqrt{s} = 8$  TeV*, [1509.04177](#).
- [51] **CMS** Collaboration, *Search for top-quark partners with charge  $5/3$  in the same-sign dilepton final state*, *Phys. Rev. Lett.* **112** (2014), no. 17 171801 [[1312.2391](#)].
- [52] D. B. Kaplan and H. Georgi,  *$SU(2) \times U(1)$  Breaking by Vacuum Misalignment*, *Phys. Lett.* **B136** (1984), no. HUTP-83/A069 183.
- [53] D. B. Kaplan, H. Georgi and S. Dimopoulos, *Composite Higgs Scalars*, *Phys. Lett.* **B136** (1984), no. HUTP-83/A079 187.
- [54] K. Agashe, R. Contino and A. Pomarol, *The Minimal composite Higgs model*, *Nucl. Phys.* **B719** (2005), no. UAB-FT-567 165–187 [[hep-ph/0412089](#)].
- [55] A. De Simone, O. Matsedonskyi, R. Rattazzi and A. Wulzer, *A First Top Partner Hunter’s Guide*, *JHEP* **04** (2013), no. CERN-PH-TH-2012-323, SISSA-31-2012-EP 004 [[1211.5663](#)].
- [56] C. Anastasiou, E. Furlan and J. Santiago, *Realistic Composite Higgs Models*, *Phys. Rev.* **D79** (2009) 075003 [[0901.2117](#)].
- [57] M. Chala, J. Juknevich, G. Perez and J. Santiago, *The Elusive Gluon*, *JHEP* **01** (2015), no. CAFPE-185-14, CERN-PH-TH-2014-216, DESY-14-213, SISSA-60-2014-FISI, UG-FT-315-14, –UG-FT-315-14 092 [[1411.1771](#)].
- [58] S. R. Coleman, J. Wess and B. Zumino, *Structure of phenomenological Lagrangians. 1.*, *Phys. Rev.* **177** (1969) 2239–2247.
- [59] C. G. Callan, Jr., S. R. Coleman, J. Wess and B. Zumino, *Structure of phenomenological Lagrangians. 2.*, *Phys. Rev.* **177** (1969) 2247–2250.
- [60] “Accelerators and experiments at CERN.” [http://meroli.web.cern.ch/meroli/lecture\\_machine\\_experiment\\_cern.html](http://meroli.web.cern.ch/meroli/lecture_machine_experiment_cern.html). Accessed: 2015-08-02.
- [61] G. Baur *et. al.*, *Production of anti-hydrogen*, *Phys. Lett.* **B368** (1996), no. PRINT-96-011 (JULICH) 251–258.

- [62] **NA48** Collaboration, *A New measurement of direct CP violation in two pion decays of the neutral kaon*, *Phys. Lett.* **B465** (1999), no. CERN-EP-99-114 335–348 [[hep-ex/9909022](#)].
- [63] **ALPHA** Collaboration, *Trapped antihydrogen*, *Nature* **468** (2010) 673–676.
- [64] **ALPHA** Collaboration, *Confinement of antihydrogen for 1000 seconds*, *Nature Phys.* **7** (2011), no. TRI-PP-11-02 558–564 [[1104.4982](#)].
- [65] O. S. Brning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole and P. Proudlock, *LHC Design Report*. CERN, Geneva, 2004.
- [66] “Diagram of an LHC dipole magnet.” <https://cds.cern.ch/record/40524>. CERN-DI-9906025, Accessed: 2015-08-18.
- [67] “LHC dipole magnetic flux.” <http://www.quantumdiaries.org/2011/04/24/the-cern-accelerator-complex/>. Accessed: 2015-08-02.
- [68] **ATLAS** Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003.
- [69] **ATLAS** Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J.* **C70** (2010) 823–874 [[1005.4568](#)].
- [70] “ATLAS Luminosity public results.” <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults>. Accessed: 2015-07-31.
- [71] “ATLAS event display of  $Z \rightarrow \mu\mu$  taken in 2012.” [https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone#2012\\_Z\\_mu\\_mu\\_event\\_with\\_high\\_pil](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EventDisplayStandAlone#2012_Z_mu_mu_event_with_high_pil). Accessed: 2015-07-31.
- [72] **ATLAS** Collaboration, *ATLAS inner detector: Technical design report. Vol. 2*, CERN-LHCC-97-17, 1997.
- [73] **ATLAS** Collaboration, *ATLAS inner detector: Technical design report. Vol. 1*, CERN-LHCC-97-16, ATLAS-TDR-4, 1997.
- [74] “ATLAS Inner Detector pictures.” <http://www.atlas.ch/photos/inner-detector-combined.html>. Accessed: 2015-07-31.

- 
- [75] **ATLAS** Collaboration, *ATLAS liquid argon calorimeter: Technical design report*, CERN-LHCC-96-41, 1996.
- [76] **ATLAS** Collaboration, *ATLAS tile calorimeter: Technical design report*, CERN-LHCC-96-42, 1996.
- [77] “ATLAS calorimeter system.”  
<http://www.atlas.ch/photos/calorimeters-combined-barrel.html>.  
Accessed: 2015-07-31.
- [78] **ATLAS** Collaboration, *ATLAS muon spectrometer: Technical design report*, CERN-LHCC-97-22, ATLAS-TDR-10, 1997.
- [79] “ATLAS Muon Spectrometer.”  
<http://www.atlas.ch/photos/muons-combined.html>. Accessed:  
2015-07-31.
- [80] **ATLAS** Collaboration, *ATLAS magnet system: Technical design report*, CERN-LHCC-97-18, 1997.
- [81] **ATLAS** Collaboration, *ATLAS central solenoid: Technical design report*, CERN-LHCC-97-21, 1997.
- [82] **ATLAS** Collaboration, *ATLAS endcap toroids: Technical design report*, CERN-LHCC-97-20, 1997.
- [83] “ATLAS magnets system.”  
<http://atlas-magnet.web.cern.ch/atlas-magnet/>. Accessed:  
2015-07-31.
- [84] “ATLAS toroidal magnetic field.”  
<http://atlas-magnet.web.cern.ch/atlas-magnet/field/>. Accessed:  
2015-07-31.
- [85] **ATLAS** Collaboration, *Improved luminosity determination in pp collisions at  $\sqrt{s} = 7$  TeV using the ATLAS detector at the LHC*, *Eur. Phys. J. C* **73** (2013), no. 8 2518 [[1302.4393](#)].
- [86] **ATLAS** Collaboration, *ATLAS first level trigger: Technical design report*, CERN-LHCC-98-14, ATLAS-TDR-12, 1998.

- [87] **ATLAS** Collaboration, *ATLAS high-level trigger, data acquisition and controls: Technical design report*, CERN-LHCC-2003-022, ATLAS-TRD-016, 2003.
- [88] F. M. A. Veloso, *Study of ATLAS sensitivity to FCNC top quark decays*. PhD thesis, Coimbra U., Coimbra, 2008. Presented on 12<sup>nd</sup> Nov 2008.
- [89] **ATLAS** Collaboration, *ATLAS computing: Technical design report*, CERN-LHCC-2005-022, ATLAS-TRD-017, 2005.
- [90] “WLCG monitoring tools.” <http://wlcg.web.cern.ch/tools>. Accessed: 2015-10-20.
- [91] E. Fullana *et. al.*, *Optimal Filtering in the ATLAS Hadronic Tile Calorimeter*, ATL-TILECAL-2005-001, ATL-COM-TILECAL-2005-001, CERN-ATL-TILECAL-2005-001, 2005.
- [92] W. Lampl, S. Laplace, D. Lelas, P. Loch, H. Ma, S. Menke, S. Rajagopalan, D. Rousseau, S. Snyder and G. Unal, *Calorimeter clustering algorithms: Description and performance*, .
- [93] S. Menke, “Talk on Topological clustering and local hadron calibration.” <https://agenda.linearcollider.org/event/5037/session/1/contribution/4/material/slides/0.pdf>. Accessed: 2015-07-31.
- [94] **ATLAS** Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, *Eur. Phys. J.* **C70** (2010), no. CERN-PH-EP-2010-024 1193–1236 [[1007.5423](#)].
- [95] **ATLAS** Collaboration, *Search for pair production of new heavy quarks that decay to a Z boson and a third generation quark in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, ATLAS-CONF-2013-056, CERN, Geneva, Jun, 2013.
- [96] **ATLAS** Collaboration, *Search for pair and single production of new heavy quarks that decay to a Z boson and a third generation quark in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector*, ATLAS-CONF-2014-036, CERN, Geneva, Jul, 2014.

- [97] **ATLAS** Collaboration, *Performance of the ATLAS Muon Trigger in 2011*, ATLAS-CONF-2012-099, 2012.
- [98] **ATLAS** Collaboration, *Performance of the ATLAS Electron and Photon Trigger in  $p$ - $p$  Collisions at  $\sqrt{s} = 7$  TeV in 2011*, ATLAS-CONF-2012-048, 2012.
- [99] “Single electron trigger efficiencies public plots at  $\sqrt{s} = 8$  TeV.”  
[https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EgammaTriggerPublicResults#Electron\\_trigger\\_performance\\_AN2](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/EgammaTriggerPublicResults#Electron_trigger_performance_AN2).  
Accessed: 2015-08-18.
- [100] “Single muon trigger efficiencies public plots at  $\sqrt{s} = 8$  TeV.”  
[https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonTriggerPublicResults#Muon\\_trigger\\_performances\\_in\\_AN1](https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonTriggerPublicResults#Muon_trigger_performances_in_AN1).  
Accessed: 2015-08-18.
- [101] **ATLAS** Collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, *Eur. Phys. J.* **C74** (2014), no. 7 2941 [[1404.2240](#)].
- [102] **ATLAS** Collaboration, *Measurement of the muon reconstruction performance of the ATLAS detector using 2011 and 2012 LHC protonproton collision data*, *Eur. Phys. J.* **C74** (2014), no. 11 3130 [[1407.3935](#)].
- [103] **ATLAS** Collaboration, *Muon reconstruction efficiency and momentum resolution of the ATLAS experiment in proton-proton collisions at  $\sqrt{s} = 7$  TeV in 2010*, *Eur. Phys. J.* **C74** (2014), no. 9 3034 [[1404.4562](#)].
- [104] **ATLAS** Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton collision data*, ATLAS-CONF-2014-032, 2014.
- [105] M. Cacciari, G. P. Salam and G. Soyez, *The Anti- $k(t)$  jet clustering algorithm*, *JHEP* **04** (2008), no. LPTHE-07-03 063 [[0802.1189](#)].
- [106] M. Cacciari and G. P. Salam, *Dispelling the  $N^3$  myth for the  $k_t$  jet-finder*, *Phys. Lett.* **B641** (2006), no. LPTHE-05-32 57–61 [[hep-ph/0512210](#)].



- [107] M. Cacciari, G. P. Salam and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012), no. CERN-PH-TH-2011-297 1896 [[1111.6097](#)].
- [108] T. Barillari, E. Bergeaas Kuutmann, T. Carli, J. Erdmann, P. Giovannini, K. J. Grahm, C. Issever, A. Jantsch, A. Kiryunin, K. Lohwasser, A. Maslennikov, S. Menke, H. Oberlack, G. Pospelov, E. Rauter, P. Schacht, F. Span, P. Speckmayer, P. Stavina and P. Strzenec, *Local Hadronic Calibration*, ATL-LARG-PUB-2009-001-2, CERN, Geneva, Jun, 2008.
- [109] **ATLAS** Collaboration, *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*, ATLAS-CONF-2015-037, 2015.
- [110] **ATLAS** Collaboration, G. Aad *et. al.*, *Jet energy measurement and its systematic uncertainty in proton-proton collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector*, *Eur. Phys. J.* **C75** (2015), no. CERN-PH-EP-2013-222 17 [[1406.0076](#)].
- [111] *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, CERN, Geneva, Aug, 2013.
- [112] *Calibration of the performance of b-tagging for c and light-flavour jets in the 2012 ATLAS data*, ATLAS-CONF-2014-046, CERN, Geneva, Jul, 2014.
- [113] *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, CERN, Geneva, Feb, 2014.
- [114] “Protos - PROgram for TOp Simulations.”  
<http://jaguar.web.cern.ch/jaguar/protos/>. Accessed: 2015-08-16.
- [115] T. Sjostrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006), no. FERMILAB-PUB-06-052-CD-T, LU-TP-06-13 026 [[hep-ph/0603175](#)].
- [116] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Parton distributions for the LHC*, *Eur. Phys. J.* **C63** (2009), no. IPPP-08-95, DCPT-08-190, CAVENDISH-HEP-08-16 189–285 [[0901.0002](#)].

- [117] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, *Comput. Phys. Commun.* **185** (2014), no. CERN-PH-TH-2011-303, TTK-11-58 2930 [[1112.5675](#)].
- [118] A. D. Martin, W. J. Stirling, R. S. Thorne and G. Watt, *Uncertainties on  $\alpha(S)$  in global PDF analyses and implications for predicted hadronic cross sections*, *Eur. Phys. J.* **C64** (2009), no. IPPP-09-33, DCPT-09-66, CAVENDISH-HEP-09-06 653–680 [[0905.3531](#)].
- [119] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer and T. Stelzer, *MadGraph 5: Going Beyond*, *JHEP* **06** (2011), no. FERMILAB-PUB-11-448-T 128 [[1106.0522](#)].
- [120] **ATLAS** Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, CERN, Geneva, Oct, 2010.
- [121] **GEANT4** Collaboration, *Geant4: a simulation toolkit*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3 250 – 303.
- [122] N. Vignaroli, *Early discovery of top partners and test of the Higgs nature*, *Phys. Rev.* **D86** (2012) 075017 [[1207.0830](#)].
- [123] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, *JHEP* **07** (2003), no. CERN-TH-2002-129, FTN-T-2002-06 001 [[hep-ph/0206293](#)].
- [124] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, *Event generation with SHERPA 1.1*, *JHEP* **02** (2009), no. FERMILAB-PUB-08-477-T, SLAC-PUB-13420, ZU-TH-17-08, DCPT-08-138, IPPP-08-69, EDINBURGH-2008-30, MCNET-08-14 007 [[0811.4622](#)].
- [125] S. Catani, L. Cieri, G. Ferrera, D. de Florian and M. Grazzini, *Vector boson production at hadron colliders: a fully exclusive QCD calculation at NNLO*, *Phys. Rev. Lett.* **103** (2009) 082001 [[0903.2120](#)].

- [126] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky and W. K. Tung, *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002), no. MSU-HEP-011101 012 [[hep-ph/0201195](#)].
- [127] M. L. Mangano, M. Moretti and R. Pittau, *Multijet matrix elements and shower evolution in hadronic collisions:  $Wb\bar{b} + n$  jets as a case study*, *Nucl. Phys.* **B632** (2002), no. CERN-TH-2001-215 343–362 [[hep-ph/0108069](#)].
- [128] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin and C. P. Yuan, *New parton distributions for collider physics*, *Phys. Rev.* **D82** (2010), no. MSUHEP-100707, SMU-HEP-10-10 074024 [[1007.2241](#)].
- [129] J. M. Campbell and R. K. Ellis, *An Update on vector boson pair production at hadron colliders*, *Phys. Rev.* **D60** (1999), no. FERMILAB-PUB-99-146-T 113006 [[hep-ph/9905386](#)].
- [130] M. V. Garzelli, A. Kardos, C. G. Papadopoulos and Z. Trocsanyi,  *$t\bar{t}W^{+-}$  and  $t\bar{t}Z$  Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, *JHEP* **11** (2012) 056 [[1208.2665](#)].
- [131] S. Frixione, P. Nason and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007), no. BICOCCA-FT-07-12, GEF-TH-19-2007 126 [[0707.3088](#)].
- [132] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007), no. BICOCCA-FT-07-9, GEF-TH-21-2007 070 [[0709.2092](#)].
- [133] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010), no. DESY-10-018, SFB-CPP-10-22, IPPP-10-11, DCPT-10-22 043 [[1002.2581](#)].
- [134] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, *JHEP* **11** (2004), no. BICOCCA-FT-04-11 040 [[hep-ph/0409146](#)].

- [135] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002), no. CAVENDISH-HEP-02-01, LAPTH-905-02, GEF-TH-2-2002 029 [[hep-ph/0204244](#)].
- [136] S. Frixione, E. Laenen, P. Motylinski, B. R. Webber and C. D. White, *Single-top hadroproduction in association with a W boson*, *JHEP* **07** (2008), no. CERN-TH-2008-104, ITP-UU-08-27, NIKHEF-2008-005, ITFA-2008-16, CAVENDISH-HEP-08-06 029 [[0805.3067](#)].
- [137] G. Corcella *et. al.*, *HERWIG 6.5 release note*, [hep-ph/0210213](#).
- [138] G. Marchesini *et. al.*, *HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons*, *Comput. Phys. Commun.* **67** (1992), no. CAVENDISH-HEP-91-26, DESY-91-048 465–508.
- [139] G. Corcella *et. al.*, *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, *J. High Energy Phys.* **0101** (2001), no. CAVENDISH-HEP-99-03, CERN-TH-2000-284, RAL-TR-2000-048 10 [[hep-ph/0011363](#)].
- [140] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, *Comput. Phys. Commun.* **184** (2013), no. TPJU-6-2004 919 [[hep-ph/0405247](#)].
- [141] N. Kidonakis, *Differential and total cross sections for top pair and single top production*, [1205.3453](#).
- [142] N. Kidonakis, *Differential and total cross sections for top pair and single top production*, in *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012)*, pp. 831–834, 2012. [1205.3453](#). [[831\(2012\)](#)].
- [143] F. Campanario and S. Sapeta, *WZ production beyond NLO for high- $p_T$  observables*, *Phys. Lett.* **B718** (2012), no. FTUV-12-0920, KA-TP-36-2012, LPN12-102, SFB-CPP-12-70, IPPP-12-70, DCPT-12-140, –DCPT-12-140, –DCPT-12-140 100–104 [[1209.4595](#)].
- [144] T. Junk, *Confidence level computation for combining searches with small statistics*, *Nucl. Instrum. Meth.* **A434** (1999),

- no. CARLETON-OPAL-PHYS-99-01, CERN-EP-99-041 435–443  
[[hep-ex/9902006](#)].
- [145] A. L. Read, *Presentation of search results: the  $CL(s)$  technique*, *J. Phys.* **G28** (2002) 2693.
- [146] R. Contino and G. Servant, *Discovering the top partners at the LHC using same-sign dilepton final states*, *JHEP* **06** (2008),  
no. CERN-PH-TH-2007-233, SACLAY-T07-149 026 [[0801.1679](#)].
- [147] J. P. Araque, N. F. Castro and J. Santiago, *Interpretation of Vector-like Quark Searches: Heavy Gluons in Composite Higgs Models*, [[1507.05628](#)].
- [148] A. Azatov, D. Chowdhury, D. Ghosh and T. S. Ray, *Same sign di-lepton candles of the composite gluons*, *JHEP* **08** (2015), no. CERN-PH-TH-2015-10 140 [[1505.01506](#)].
- [149] R. Barcelo, A. Carmona, M. Masip and J. Santiago, *Stealth gluons at hadron colliders*, *Phys. Lett.* **B707** (2012), no. UG-FT-287-11, CAFPE-157-11 88–91 [[1106.4054](#)].
- [150] R. Barcelo, A. Carmona, M. Chala, M. Masip and J. Santiago, *Single Vectorlike Quark Production at the LHC*, *Nucl. Phys.* **B857** (2012) 172–184 [[1110.5914](#)].
- [151] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, *UFO - The Universal FeynRules Output*, *Comput. Phys. Commun.* **183** (2012), no. CP3-11-25, IPHC-PHENO-11-04, IPPP-11-39, DCPT-11-78, MPP-2011-68 1201–1214 [[1108.2040](#)].
- [152] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014), no. CERN-PH-TH-2013-239, MCNET-13-14, IPPP-13-71, DCPT-13-142, –PITT-PACC-1308 2250–2300 [[1310.1921](#)].
- [153] **DELPHES 3** Collaboration, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057 [[1307.6346](#)].

- [154] P. Artoisenet, R. Frederix, O. Mattelaer and R. Rietkerk, *Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations*, *JHEP* **03** (2013), no. NIKHEF-2012-021, CERN-PH-TH-2012-329 015 [[1212.3460](#)].
- [155] A. Hocker *et. al.*, *TMVA - Toolkit for Multivariate Data Analysis*, *PoS ACAT* (2007), no. CERN-OPEN-2007-007 040 [[physics/0703039](#)].
- [156] **CMS** Collaboration, *Search for a vector-like bottom quark partner in same sign di-lepton final states*, .
- [157] **CMS** Collaboration, *Inclusive search for a vector-like  $T$  quark with charge  $\frac{2}{3}$  in  $pp$  collisions at  $\sqrt{s} = 8$  TeV*, *Phys. Lett.* **B729** (2014), no. CMS-B2G-12-015, CERN-PH-EP-2013-215 149–171 [[1311.7667](#)].
- [158] N. Castro, *Study of the  $Wtb$  vertex structure at the ATLAS experiment*. PhD thesis, U. Coimbra, Coimbra, 2008. Presented on 17<sup>th</sup> Oct 2008.
- [159] R. Brun and F. Rademakers, *ROOT - An Object Oriented Data Analysis Framework*, in *Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996*, vol. A 389, pp. 81–86, 1997. See also <http://root.cern.ch/>.
- [160] “The  $CL_s$  method.” <http://www.pp.rhul.ac.uk/~cowan/stat/cls/CLsInfo.ps>. Accessed: 2015-10-28.