

# **Search for Higgs Decays to Dark Matter and Trigger Studies**

João Carlos Arnauth Pela

Imperial College London  
Department of Physics

A dissertation submitted to Imperial College London  
for the degree of Doctor of Philosophy

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

## Abstract

The Compact Muon Solenoid (CMS) is a general-purpose particle detector at the CERN Large Hadron Collider (LHC). The goal of this experiment is to search for the Higgs boson and evidence of new physics and to test the prediction of the Standard Model (SM) at the TeV scale. This thesis describes the analysis of proton-proton collision data recorded by CMS during 2012 and support work for data taking during the same period.

A search for the Vector Boson Fusion (VBF) produced Higgs boson invisible decays, using  $19.5 \text{ fb}^{-1}$  of data recorded with prompt reconstruction triggers at a center of mass energy of 8 TeV, is presented. Events are selected with two forward jets and large Missing Transverse Momentum. Assuming the SM VBF production cross section and acceptance, the observed (expected) upper limit at the 95% confidence level on the  $\mathcal{B}(H \rightarrow \text{inv})$  is determined to be of 65% (49%) for  $m_H = 125 \text{ GeV}$ .

A second search for the VBF Higgs boson invisible decays, using  $19.2 \text{ fb}^{-1}$  of data recorded with delayed reconstruction (parked) triggers at a center of mass energy of 8 TeV, is also presented. A new event selection was developed taking advantage of the lower trigger requirements. Assuming the SM VBF Higgs production cross section and acceptance, the observed (expected) upper limit at the 95% confidence level on the  $\mathcal{B}(H \rightarrow \text{inv})$  is determined to be 57% (40%) for  $m_H = 125 \text{ GeV}$ .

Monitoring for the CMS Level 1 Trigger system has been developed and used during the 2012 and subsequent LHC data acquisition periods. Contributing to the high reliability of this system during data taking and providing crucial information for validation of the data quality.

## Declaration

I declare that the work contained in this thesis is my own, and all results and figures taken from other sources are indicated in the text and referenced appropriately. The analyses presented in this thesis were developed in close collaboration with other members of the CMS experiment.

I joined the Vector Boson Fusion (VBF) Higgs to invisible prompt analysis at a late stage in its preparation, so my contributions were modest. I contributed with a review of the literature of final background cross section input values for the background normalization. I also preformed QCD multi-jet background studies with the target of improving the final selection and as preparation for the parked analysis. The prompt analysis is presented to give context for the more important work developed later.

For the VBF Higgs to invisible parked analysis I developed the parked trigger, which was used to record the majority of the analysed data. I continued the QCD multi-jet background studies which have led to the production of dedicated Monte Carlo simulations and novel approaches to reject this type of events. I was solely responsible for the cross check analysis of the main result which validated the implementation of the main analysis [1]. It is a standard requirement for many CMS publications to have a cross check analysis implemented independently from the main result in order to be able to ensure the accuracy of the final results.

I have also participated in the preparation of the Run II analysis where I have led the development of both the triggers used for data recording during 2015. Additionally, I have developed a novel method to create the first QCD multi-jet Monte Carlo sample with no MET requirements with signal-like properties.

As part of the CMS Level 1 Trigger (L1T) Detector Performance Group (DPG), I developed monitoring tools, which were used both for real-time

monitoring and posterior data certification for physics analysis usage. My work in this group has led to my appointment for two years to the position of coordinator of the Compact Muon Solenoid (CMS) Level 1 Trigger (L1T) Data quality Monitoring software development team. Field work was also performed by doing shifts as Trigger and Shift Leader in the experiment control room and on-call shifts as the Trigger Detector on Call (DOC) expert.

João Pela

## Acknowledgements

I would like to thank the Imperial College HEP Group and the FCT for supporting and giving me the opportunity to pursue research including my time spent at CERN. Thanks to my supervisor, David Colling, for his guidance during the years of my PhD, the many useful discussions and his enthusiasm for particle physics research. Thanks must also go to all my colleagues of the CMS VBF Higgs to invisible group with whom I have worked directly, Patrick Dunne, Anne-Marie Magnan, Alexandre Nikitenko, Jim Brooke and Chayanit Asawatangtrakuldee. Thanks to everyone in CMS L1T DPG group, specially to Carlo Battilana and Arno Heister for their support. Thanks to all my friends that supported me on my way to finish this degree, it is difficult to mention all of you, but here is a small list, Antonio Pacheco, André David, Alice Mitchell, Chris Holdsworth-Swan, Francisco Sousa, Glenn Spiers, Guillermo Marrero, João Filipe, Miguel Cunhal, Miguel Machado, Miguel Romão, Nadia Silva, Raquel Monteiro, and many, many more. Finally, and most importantly, I thank my parents and grandparents for their constant support, love and encouragement.

The work presented in this thesis was supported by the Portuguese Government through Fundação para a Ciência e a Tecnologia (FCT) in the form of my PhD grant with the reference SFRH/BD/77979/2011. I am thankful for their support which allowed me to attain higher education.



*“To my grandmother.”*

# Contents

<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xvi</b>
<b>1 Theory and motivations</b>	<b>2</b>
1.1 Standard Model of Particle Physics . . . . .	2
1.1.1 Fundamental matter particles . . . . .	3
1.1.2 Fundamental forces . . . . .	4
1.1.3 Electroweak Gauge Symmetry . . . . .	4
1.1.4 The Higgs Mechanism in the Standard Model . . . . .	6
1.1.5 Searching for the SM Higgs boson . . . . .	9
1.1.6 Invisible Higgs decay . . . . .	13
<b>2 Experimental Apparatus</b>	<b>14</b>
2.1 The Large Hadron Collider . . . . .	14
2.1.1 Running and performance . . . . .	17
2.2 The Compact Muon Solenoid Experiment . . . . .	20
2.2.1 Geometry and conventions . . . . .	21
2.2.2 Inner tracking system . . . . .	22
2.2.3 Electromagnetic Calorimeter . . . . .	23
2.2.4 Hadronic Calorimeter . . . . .	25
2.2.5 Solenoid Magnet . . . . .	27
2.2.6 Muon System . . . . .	27
2.2.7 Data Acquisition System . . . . .	30
2.2.8 Trigger System . . . . .	30
2.2.9 Computing . . . . .	32
2.2.10 Level 1 Trigger: Stage I Upgrade . . . . .	33

<b>3 Event Reconstruction and simulation</b>	<b>34</b>
3.1 Tracks . . . . .	34
3.2 Vertex Reconstruction . . . . .	35
3.3 Particle Flow . . . . .	37
3.3.1 Isolation . . . . .	38
3.4 Electrons . . . . .	38
3.4.1 Isolation . . . . .	41
3.4.2 Veto electrons . . . . .	41
3.4.3 Tight electrons . . . . .	42
3.5 Muons . . . . .	42
3.5.1 Isolation . . . . .	43
3.5.2 Loose Muons . . . . .	44
3.5.3 Tight Muons . . . . .	44
3.6 Jets . . . . .	44
3.6.1 Jet Clustering . . . . .	45
3.6.2 Particle Flow Jet Identification . . . . .	46
3.6.3 Pileup Jet Identification . . . . .	46
3.6.4 Lepton cleaning . . . . .	47
3.6.5 Jet Energy Corrections . . . . .	47
3.7 Hadronic Taus . . . . .	48
3.7.1 Isolation and Discriminants . . . . .	49
3.8 Missing Transverse Energy . . . . .	49
3.9 Monte Carlo Simulation . . . . .	51
<b>4 L1 Trigger Data Quality Monitoring</b>	<b>54</b>
4.1 Data Quality Monitoring . . . . .	54
4.1.1 Online Monitoring . . . . .	54
4.1.2 Offline Monitoring . . . . .	55
4.2 Level 1 Trigger Data Quality Monitoring . . . . .	55
4.2.1 Rates Monitoring . . . . .	56
4.2.2 Synchronization Monitoring . . . . .	57
4.2.3 BPTX Monitoring . . . . .	59
4.2.4 Occupancy Monitoring . . . . .	61
4.3 Tests Summary . . . . .	66
<b>5 Search for H(<math>\text{Inv}</math>) decays in the VBF channel with CMS prompt data</b>	<b>69</b>
5.1 Event Selection . . . . .	70

5.2	Background Estimation . . . . .	71
5.3	Sources of uncertainty . . . . .	76
5.4	Results and conclusions . . . . .	77
<b>6</b>	<b>Run I parked data analysis preparation</b>	<b>79</b>
6.1	L1T parked trigger development . . . . .	80
6.1.1	VBF Higgs to Invisible Higgs Level 1 trigger development . . . . .	80
6.1.2	VBF Higgs inclusive Level 1 trigger development . . . . .	82
6.1.3	Final proposal . . . . .	85
6.2	Monte Carlo simulation of QCD multi-jet events with VBF jets and MET	86
6.2.1	Pre-selection for data comparison . . . . .	88
6.3	QCD control studies . . . . .	92
6.3.1	Dijet-MET system topological variables . . . . .	92
6.3.2	Track distribution variables . . . . .	94
6.3.3	Summary . . . . .	96
<b>7</b>	<b>Search for H(Inv) decays in the VBF channel with CMS parked data</b>	<b>98</b>
7.1	The Cross Check Analysis . . . . .	98
7.2	Data and MC samples . . . . .	99
7.2.1	Data . . . . .	99
7.2.2	Monte Carlo Samples . . . . .	100
7.3	Monte Carlo simulation to Data correction factors . . . . .	102
7.3.1	Pile-up . . . . .	102
7.3.2	Trigger efficiency . . . . .	102
7.3.3	Lepton Identification . . . . .	103
7.4	Signal event selection . . . . .	103
7.5	Control Regions . . . . .	104
7.5.1	Z background estimation . . . . .	105
7.5.2	W background estimation . . . . .	106
7.5.3	Top background estimation . . . . .	111
7.5.4	QCD background estimation . . . . .	111
7.6	Systematics . . . . .	113
7.7	Results . . . . .	114
7.7.1	Comparison with the cross-check analysis . . . . .	117
7.8	Limits on the cross section of invisibly decaying Higgs bosons . . . . .	117

<b>8 Run II preparation</b>	<b>119</b>
8.1 Run II trigger studies . . . . .	119
8.1.1 Methodology . . . . .	120
8.1.2 L1T algorithm development . . . . .	121
8.1.3 HLT algorithm development . . . . .	123
8.1.4 Summary . . . . .	133
8.2 Run II QCD Monte Carlo samples . . . . .	134
8.2.1 Goals and first attempt . . . . .	134
8.2.2 MADGRAPH parton level simulation . . . . .	135
8.2.3 Hadronization with PYTHIA 8 . . . . .	135
8.2.4 Generator level cuts . . . . .	137
8.2.5 Migration study . . . . .	141
8.2.6 Summary . . . . .	144
<b>9 Conclusions</b>	<b>146</b>
<b>Bibliography</b>	<b>149</b>
<b>Acronyms</b>	<b>162</b>

# List of Tables

1.1	List of leptons and their fundamental properties . . . . .	3
1.2	List of bosons and their fundamental properties . . . . .	4
2.1	LHC parameters relevant for detectors . . . . .	17
2.2	Parameters of the CMS superconducting solenoid . . . . .	28
3.1	Details of the CMS Electron-Gamma Particle Object Group (POG) recommendations for a <i>veto electron</i> . Here barrel is defined as $ \eta_{supercluster}  \leq 1.479$ and endcap is $1.479 <  \eta_{supercluster}  < 2.5$ . . . . .	41
3.2	Details of the CMS Electron-Gamma POG recommendations for a <i>tight electron</i> . Here barrel is defined as $ \eta_{supercluster}  \leq 1.479$ and endcap is $1.479 <  \eta_{supercluster}  < 2.5$ . . . . .	42
3.3	Table of the minimum values of full BDT method score for a PF jet to be accepted as coming from the PV using a loose working point. . . . .	46
3.4	Summary of the hadronic tau decay modes. . . . .	48
5.1	Summary of the estimated number of background and signal events, together with the observed yield, in the VBF search signal region. . . . .	76
5.2	Summary of the uncertainties in the total background and signal yields in the VBF channel. . . . .	77
6.1	Tables showing the L1T rate for different selection criteria for $5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ and an average PU of 28 interactions (scenario A). . . . .	81
6.2	Tables showing the L1T rate for different selection criteria for $7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$ and an average PU of 32 interactions (scenario B). . . . .	81
6.3	Table of the key parameters of each simulated MC event sample ordered by $p_T$ hat. . . . .	87
7.1	Relevant parked datasets from Run I and their total analysed integrated luminosity. Total analysed and certified also showed. . . . .	100

7.2	Table of the MC processes, corresponding cross sections (at NLO or NNLO when available) and equivalent integrated luminosity analysed. . . . .	101
7.3	Table of the step-by-step event yields for the signal region obtained by the cross check analysis. Yields are discriminated by Run I period. . . . .	104
7.4	Summary of the W background estimates as calculated by the cross check analysis. . . . .	107
7.5	Summary of the uncertainties on the total background and signal yields. . . . .	114
7.6	Summary of the estimated number of background and signal events, together with the observed yield, in the Vector Boson Fusion (VBF) search signal region. . . . .	115
7.7	Comparison of the data event yields in all relevant regions, between the main and cross check analyses. . . . .	117
8.1	Results of the search for L1T algorithms with a maximum rate of 5 kHz for the <i>TSG high luminosity scenario</i> . . . . .	122
8.2	Results of the automatic optimization of possible HLT paths for a maximum rate of 5 Hz for the <i>TSG high luminosity scenario</i> . All HLT algorithms are seeded by L1.ETM70. . . . .	125
8.3	Results of the automatic optimization of possible HLT paths for a maximum rate of 5 Hz for the <i>TSG high luminosity scenario</i> . All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF, $p_T^{jets} > 30 \text{ GeV}$ , and $\Delta\eta > 3.5$ and $ETM \geq 50 \text{ GeV}$ . . . . .	128
8.4	Results of the automatic optimization of possible HLT paths for a maximum rate of 5 Hz for the <i>TSG high luminosity scenario</i> . All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF, $p_T^{jets} > 30 \text{ GeV}$ , and $\Delta\eta > 3.5$ and a single jet $p_T^{jets} > 96 \text{ GeV}$ . . . . .	130
8.5	Summary of the results of the Hadronization with Pythia 8 of 1.4M MadGraph events passing the parton level filter. . . . .	137
8.6	Table showing the percentage of generator anti- $k_T$ jets with $R = 0.4$ passing cuts $p_T > 40 \text{ GeV}$ and $ \eta  < 4.8$ for events with at least one dijet with $\Delta\eta < 3.0$ and $m_{jj} < 1000 \text{ GeV}$ and according to an additional dijet $\Delta\phi$ cut. . . . .	140
8.7	Table showing the percentage of generator anti- $k_T$ dijets with $R = 0.4$ passing cuts $p_T^{jet} > 40 \text{ GeV}$ , $ \eta ^{jet} < 4.8$ , $\Delta\eta < 3.0$ and $m_{jj} < 1000 \text{ GeV}$ and according to an additional dijet $\Delta\phi$ cut. . . . .	140

- 8.8 Table showing the percentage of partons successfully matched to a generator AK4 jets. Numbers obtained for a total of 88282 events over all 3 possible hard scattering processes and for events with at least one di-parton with  $m_{jj} > 600 \text{ GeV}$  where each parton has  $p_T < 10 \text{ GeV}$  and  $|\eta| < 5.0$  . 142

# List of Figures

1.1	Feynman diagrams for the main production processes of the SM Higgs boson at the LHC. . . . .	9
1.2	Cross sections for Higgs production processes at $\sqrt{s} = 8$ TeV for a range of Higgs boson masses. . . . .	11
1.3	Higgs boson branching ratios in the SM for a range of Higgs boson masses. . . . .	12
2.1	Underground diagram of the Geneva area showing the LHC and its experiments location. . . . .	14
2.2	Diagram of the CERN accelerator complex. . . . .	16
2.3	Cross sections for several processes for collisions of antiproton-proton and proton-proton as a function of the center of mass energy. . . . .	18
2.4	Cumulative luminosity versus day delivered to CMS during stable beams and for p-p collisions for 2010, 2011 and 2012 data-taking. . . . .	19
2.5	Mean number of interactions per bunch crossing at the CMS experiment during 2012. . . . .	19
2.6	Diagram of CMS experiment showing the experiment in an open configuration and highlighting the position of its sub-detectors. . . . .	20
2.7	Schematic cross section of the CMS tracker. The inner tracker components are highlighted: TEC, TOB, TID, TIB and Pixels. . . . .	23
2.8	Diagram of the ECAL layout illustrating the positions of its components. . . . .	24
2.9	Longitudinal view of the CMS detector highlighting the location of the HCAL components: HB, HE, HO and HF. . . . .	26
2.10	Diagram of the CMS muon systems. The location of each muon chamber for each subsystem is shown: DT, CSC and RPC. . . . .	29
2.11	Diagram of the CMS DAQ system and its data flow. . . . .	30
2.12	Diagram of the L1T system and its data flow. . . . .	31
3.1	Primary vertex resolution in the $z$ coordinate and vertex reconstruction efficiency as a function of the number of constituent tracks. . . . .	36

3.2	Distributions for the variables $\Delta\eta_{in}$ , $\Delta\phi_{in}$ , $\sigma_{in\eta}$ and $H/E$ for simulated electrons and misidentified jets. . . . .	40
3.3	Distributions of the particle flow $E_T^{\text{miss}}$ in $Z \rightarrow \mu\mu$ and $\gamma+\text{jets}$ events in $\sqrt{s} = 8 \text{ TeV}$ data and simulation. . . . .	50
3.4	Illustration a proton-proton collision as implemented in MC event generators. . . . .	52
4.1	Monitoring plots produced by the L1T rates online monitoring tool for run 207269 and the electron/gamma object category. . . . .	57
4.2	Monitoring plots produced by the L1T synchronization online monitoring tool for run 207269 and the electron/gamma object category. . . . .	59
4.3	Monitoring plots produced by the L1T BPTX online monitoring tool for CMS run 207269 . . . . .	60
4.4	Detail of L1T BPTX online monitoring tool histogram for veto efficiency during its field test at run 207269. . . . .	61
4.5	Graphic showing the results of the determination of $T_{\text{crit}}$ as a function of $\mu_0$ and the corresponding fit for each bad cell hypothesis. . . . .	64
4.6	Graphic showing the results of $\mu_0^{\min}$ as a function of the number of bins and the corresponding fit for each bad cell hypothesis. . . . .	65
4.7	Monitoring plots produced by the L1T occupancy online monitoring tool for run 207099 while testing the GCT histogram for isolated EM region occupancy in $\eta - \phi$ . . . . .	66
4.8	Example of the plots produced by the L1T test summary online monitoring	67
5.1	Distributions of $M_{jj}$ , $\Delta\eta_{jj}$ , $\Delta\phi_{jj}$ and central jet $p_T$ in background and signal MC simulation. The distributions are shown after requiring two jets with $p_T^{\text{jet}1}, p_T^{\text{jet}2} > 50 \text{ GeV}$ , $ \eta  < 4.7$ , $\eta_{\text{jet}1} \cdot \eta_{\text{jet}2} < 0$ , $M_{jj} > 150 \text{ GeV}$ , and $\text{MET} > 130 \text{ GeV}$ . . . . .	72
5.2	Distribution for MET and $M_{jj}$ for a relaxed $Z$ control region, with no requirements on $\Delta\eta_{jj}$ , $\Delta\phi_{jj}$ or CJV and with a $M_{jj} > 1000 \text{ GeV}$ requirement. . . . .	73
5.3	Expected and observed 95% CL upper limits on the VBF production cross section times invisible branching fraction and normalized to the SM Higgs boson VBF production cross section. . . . .	78
6.1	Level 1 rate as a function of dijet $p_T^{\text{jets}}$ while selecting events with at least one dijet with $\Delta\eta > 3$ and $M_{jj} > 700 \text{ GeV}$ where both jets are in opposite sides of the detector for scenario A. . . . .	83

6.2	Level 1 rate as a function of dijet $p_T^{jets}$ while selecting events with at least one dijet with $\Delta\eta > 3$ and $M_T > 50$ GeV where both jets are in opposite sides of the detector for scenario A. . . . .	84
6.3	Level 1 rate as a function of dijet $p_T^{jets}$ while selecting events with at least one dijet with $\Delta\eta > 3$ where are in opposite sides of the detector and $HT > 100$ GeV for scenario A. . . . .	85
6.4	Reconstructed PF MET as a function of generator-level MET in the inclusive QCD multi-jet samples $80 < p_T < 600$ GeV before any selection. . . . .	87
6.5	Plots of the distributions for some of the key analysis variables after the proposed pre-selection using MC simulation including the new QCD multi-jet samples and $19.5 \text{ fb}^{-1}$ of data. . . . .	90
6.6	Distributions of jets $\Delta\phi$ and $\min(\Delta\phi(\text{MET}), jets)$ , for a signal selection criteria requiring the leading dijet to be on opposite sides of the detector and pass $p_T^{jets} > 50, 40$ GeV, $ \eta_{jets}  < 4.7$ , $M_{jj} > 800$ GeV, $\Delta\eta_{jj} > 3.6$ , $\min(\Delta\phi(\text{MET}), jets) > 1.5$ , with $\text{MET} > 60$ GeV and $\text{MET}_{sig} > 3$ . . . . .	91
6.7	Shape distributions for $\alpha_{\text{scale}}$ and $\alpha_{\text{vec}}$ . . . . .	93
6.8	Plots comparing the distribution of $\alpha_{\text{balance}}$ for VBF Higgs to invisible decays with $m_H = 120$ GeV and Quantum Chromodynamics (QCD) multi-jets samples. . . . .	94
6.9	Distributions of $\beta_{\text{PV Tracks}}^{\text{dijet } n}$ and $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$ for track $p_T > 1$ GeV . . . . .	95
6.10	Distributions of $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$ for track $p_T > 1$ GeV for events passing a central jet veto. . . . .	96
6.11	Distributions of $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$ for track $p_T > 1$ GeV for events where both and only one selected jets inside tracker acceptance. . . . .	97
7.1	Distributions of Pseudorapidity difference between the two selected VBF jets, $\Delta\eta_{jj}$ , Dijet mass $M_{jj}$ , Missing Transverse Energy (MET) significance $\mathcal{S}$ and MET, in the $Z \rightarrow \mu\mu$ control region as seen by the cross check analysis. . . . .	106
7.2	Dijet mass $M_{jj}$ for the $W \rightarrow e\nu$ , $W \rightarrow \mu\nu$ and $W \rightarrow \tau\nu$ control regions obtained with the cross check analysis. . . . .	108
7.3	MET for the $W \rightarrow e\nu$ , $W \rightarrow \mu\nu$ and $W \rightarrow \tau\nu$ control regions obtained with the cross check analysis. . . . .	109
7.4	Minimum azimuthal angle separation between any jet with $p_T > 30$ GeV and the MET $\Delta\phi(\text{MET}, jets)$ for the $W \rightarrow e\nu$ , $W \rightarrow \mu\nu$ and $W \rightarrow \tau\nu$ control region obtained with the cross check analysis. . . . .	110

7.5	$E_T^{\text{miss}}$ significance $\text{MET}_{\text{sig}}$ for events with $\Delta\phi(\text{MET}, \text{jets}) < 1.0$ and $\Delta\phi(\text{MET}, \text{jet}_{1,2}) > 2.3$ . MET significance, $\text{MET}_{\text{sig}}$ , for events with $\Delta\phi(\text{MET}, \text{jets}) > 1.0$ and at least 3 jets with $p_T > 30 \text{ GeV}$ .	112
7.6	Pseudorapidity difference between the two selected VBF jets $\Delta\eta_{jj}$ , dijet mass $M_{jj}$ , MET significance $\text{MET}_{\text{sig}}$ and MET, in the signal region as produced by the cross check analysis.	116
7.7	The 95% C.L. limit on $\mathcal{B}(H \rightarrow \text{inv})$ of a SM Higgs boson and the 95% C.L. limit on the cross section times $\mathcal{B}(H \rightarrow \text{inv})$ as a function of the Higgs boson mass, assuming SM Higgs boson acceptances.	118
8.1	Plots produced by the optimization process of possible L1T algorithms with a maximum rate of 5 kHz for the <i>TSG high luminosity scenario</i> .	123
8.2	Plots showing the scan over HLT PF MET for different algorithm base selection in the <i>TSG high luminosity scenario</i> . All HLT algorithms are seeded by L1_ETM70.	125
8.3	Trigger turn ons for the VBF Higgs to invisible dedicated trigger for $\text{MET}_{\text{no-}\mu}$ , dijet sub-lead jet $p_T$ , dijet $\Delta\eta_{jj}$ and dijet $M_{jj}$ .	127
8.4	Plots showing the scan over HLT PF MET for different algorithm base selection in the <i>TSG high luminosity scenario</i> . All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF, $p_T^{\text{jets}} > 30 \text{ GeV}$ , and $\Delta\eta > 3.5$ and $ETM \geq 50 \text{ GeV}$ .	129
8.5	Plots showing the scan over HLT PF MET for different algorithm base selection in the <i>TSG high luminosity scenario</i> . All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF, $p_T^{\text{jets}} > 30 \text{ GeV}$ , and a single jet $p_T^{\text{jets}} > 96 \text{ GeV}$ .	131
8.6	Results of the automatic optimization of possible HLT path for a maximum rate of 500 Hz for the <i>TSG high luminosity scenario</i> . The HLT algorithm is seeded by L1_ETM50.	132
8.7	Key variables of events passing the parton level filter.	136
8.8	Relevant distributions for the two jets comprising the leading dijet passing a generator filter requiring at least one dijet with $\Delta\eta > 3.0$ and $M_{jj} > 1000 \text{ GeV}$ where the jets have $p_T > 40 \text{ GeV}$ and $ \eta  < 4.8$	138
8.9	Relevant distributions for the leading dijet passing a generator filter requiring at least one dijet with $\Delta\eta > 3.0$ and $M_{jj} > 1000 \text{ GeV}$ where the jets have $p_T > 40 \text{ GeV}$ and $ \eta  < 4.8$	139

8.10 Plots for relevant variables of the selected di-parton against its matched dijet: Lead parton $p_T$ , sub-leading parton $p_T$ , lead parton $\eta$ , sub-leading parton $\eta$ , di-parton $\Delta\eta$ and di-parton $M_{jj}$	143
--	-----

# Chapter 1

## Theory and motivations

The goal of particle physics is to study the most fundamental constituents of matter and understand how they interact with each other. The Standard Model (SM) of particle physics will be briefly introduced, the Higgs mechanism explained and the search for Higgs boson decaying invisibly motivated. Throughout this chapter Einstein summation convention, Feynman slash notation and natural units are used, where  $\hbar = c = 1$ . Additionally, greek letters are used to label the four vectors, and gauge group generators use roman letters.

### 1.1 Standard Model of Particle Physics

The SM of particle physics is a Quantum Field Theory (QFT) including both relativistic and quantum mechanical effects. It describes the electromagnetic, weak and strong nuclear forces and their interaction with matter. This theory is one of the most successful theories ever, able to describe data from a wide range of experimental measurements. Before its discovery in 2012 [2, 3] the Higgs boson was the only missing particle that was predicted by this theory and not yet found.<sup>1</sup>

Despite its success, the SM does not explain some phenomena observed in nature, like the neutrino mass, the presence of large quantity of *dark matter* in the universe, or the even more mysterious *dark energy*. The discovery of the Higgs boson could allow us to probe the production of dark matter directly, through its decay into these elusive particles.

---

<sup>1</sup>A detailed description of the SM can be found in [4] and [5].

### 1.1.1 Fundamental matter particles

Within the SM all fundamental matter particles are spin- $\frac{1}{2}$  fermions. The equation of motion for a spin- $\frac{1}{2}$  fermion with a mass  $m$  is the Dirac equation (Equation 1.1).

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \quad (1.1)$$

In this equation the matrices  $\gamma^\mu$ ,  $\mu \in 0, 1, 2, 3$ , are defined by the anti-commutator relation  $\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 2\eta^{\mu\nu} I_4$  where  $\eta^{\mu\nu}$  is the flat space-time metric  $(+, -, -, -)$  and  $I_4$  is the  $4 \times 4$  identity matrix. The solutions for the Dirac fermion equation of motion,  $\psi$ , are the massive particle and anti-particle states, with momentum  $\mathbf{p}$  and energy  $E$ , which satisfy the relativistic expression,  $E^2 = \mathbf{p} \cdot \mathbf{p} + m^2$ .

Fundamental fermions can be split in two categories depending on whether they interact (quarks) or not (leptons) with the strong nuclear force. Both categories of particles can be grouped into three generations, with similar properties between them but increasing mass. While leptons can be examined isolated, free quarks are not observed in nature, they are confined in composite structures of three (baryons) or two (mesons) quarks. Table 1.1 shows a summary of the known fundamental matter particles.

	Leptons (J=1/2)			Quarks (J=1/2)		
Generation	Symbol	Mass	Q/e	Symbol	Mass	Q/e
$1^{st}$	e	511 keV	1	u	2.3 MeV	2/3
	$\nu_e$	< 2 eV	0	d	4.8 MeV	-1/3
$2^{nd}$	$\mu$	106 MeV	1	c	1.275 GeV	2/3
	$\nu_\mu$	< 0.19 MeV	0	s	95 MeV	-1/3
$3^{rd}$	$\tau$	1777 MeV	1	t	173.21 GeV	2/3
	$\nu_\tau$	< 18.2 MeV	0	b	4.66 GeV	-1/3

**Table 1.1:** List of fermions grouped in generations and split in fermions and quarks and their fundamental properties [6].

### 1.1.2 Fundamental forces

Gauge bosons mediate the fundamental forces of nature. All the currently observed force mediators are spin-1 particles, which is consequence of symmetries that the relevant theory possesses. The QFT that describes the electromagnetism is Quantum Electrodynamics (QED), and the strong nuclear force is Quantum Chromodynamics (QCD). Both these theories describe massless mediator bosons, the photon and the gluons, which appear as a direct consequence of the local gauge invariance of those theories. A fundamental difference between these interactions is their range, while the electromagnetism is infinite, the scale of the strong force is of around  $10^{-15}$  m.

The  $W^\pm$  and  $Z$  bosons are responsible for mediating the weak force and have a non-zero mass which has been measured experimentally [6–8]. The weak and electromagnetic forces mediator appear from the unification of the weak and electromagnetic interactions theories and the mixing of the associated gauge fields. Table 1.2 contains a summary of the fundamental gauge bosons of the Standard Model. The description of gravity is currently not included in the Standard Model, but as its interaction strength is much smaller than the other three forces it should not have any impact in its predictions.

Bosons			
Particle Name	Mass (GeV)	Q/e	Spin
Photon ( $\gamma$ )	0	0	1
$W^\pm$	$80.385 \pm 0.015$	$\mp 1$	1
$Z^0$	$91.1876 \pm 0.0021$	0	1
Gloun (g)	0	0	1

**Table 1.2:** List of force carrying bosons and their fundamental properties [6].

### 1.1.3 Electroweak Gauge Symmetry

Symmetries in nature normally appear as a direct consequence of the fundamental law. It can be shown that if a physical system can be described within the Lagrangian formalism, all symmetries that can be found on the system Lagrangian have an associated conserved quantity [9]. These properties can be applied when using dynamical quantum theories to constrain the Lagrangian of particle interactions, where the characteristics of the

interaction itself allow the identification of transformation under which the Lagrangian should be symmetric.

The development of the SM had one of its greatest successes in the unification of the electromagnetic and weak interactions [10–12]. The unification of these theories appears by combining their individual symmetry groups. The characteristics of the electroweak interaction are described by a Lagrangian which is invariant under transformations of the group  $SU(2)_L \times U(1)_Y$ , where the subscript  $L$  indicates the coupling to left-handed fermions only and  $Y$  indicates that the group generator is the weak hypercharge. The quantum numbers in this electroweak theory are the weak isospin  $t_{1,2,3}$  and hypercharge  $y$ , which are related to the electric charge as expressed in equation 1.2.

$$Q = t_3 + \frac{y}{2} \quad (1.2)$$

These quantum numbers are associated with gauge fields. The weak isospin fields  $W_\mu^i$ ,  $i = 1, 2, 3$  and the hypercharge field  $B_\mu$ . The weak isospin fields act on doublets like the ones in equation 1.3.

$$\psi_L = \begin{pmatrix} u_i \\ d_i \end{pmatrix}_L, \begin{pmatrix} \nu_i \\ l_i \end{pmatrix}_L \quad (1.3)$$

In this equation the  $u_i$  and  $d_i$  are up-type and down-type quarks respectively the  $l_i$  are charged leptons and the  $\nu_i$  are the corresponding neutrinos. The index  $i$  identifies the generation of fermions. The weak force only interacts with left-handed fermions, which is indicated by the subscript  $L$ , which makes it maximally parity violating. The fermion right-handed projections  $\psi_R$  are invariant under  $SU(2)_L$  and transform as singlet states.

The physical electroweak boson fields,  $W_\mu^\pm$ ,  $A_\mu$  the photon field and  $Z_\mu^0$  result from the mixing between the electroweak gauge fields, as it can be seen in equation 1.4 and 1.5.

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp W_\mu^2), \quad (1.4)$$

$$\begin{pmatrix} A_\mu \\ Z_\mu^0 \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \cdot \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix}, \quad (1.5)$$

Where  $\theta_W$  is the weak mixing angle [11], which is connected to the couplings of the weak ( $g$ ) and electromagnetic interactions ( $g'$ ) through the relation  $\theta_W = \tan^{-1} \frac{g'}{g}$ . The Gargamelle bubble chamber experiment at CERN discovered the weak neutral currents in 1973 [13], while the Z and  $W^\pm$  were discovered by the UA1 and measured by the UA2 collaborations at CERN in 1983 [14–17].

This model construction leads to a Lagrangian that does not have any mass terms. Directly adding mass terms of the form  $-M^2 W_\mu W^\mu$  cannot be done as it would break gauge invariance. Adding fermion mass terms of the form  $-m \bar{\psi} \psi = -m(\bar{\psi}_R \psi_L + \bar{\psi}_L \psi_R)$ , where  $\bar{\psi}$  is the adjoint of the field  $\psi$ , has field pairs of left- and right-handed components which will transform differently under the  $SU(2)_L$  and  $U(1)_Y$  groups, and as a consequence will also break gauge invariance.

The photon mass has been experimentally measured to be compatible with zero within errors, but W and Z have masses of the order of  $\approx 80$  GeV and  $\approx 91$  GeV respectively [6]. Therefore the electroweak symmetry must be spontaneously broken to reconcile theory and observation. The Higgs mechanism is the electroweak spontaneous symmetry breaking in the SM.

#### 1.1.4 The Higgs Mechanism in the Standard Model

In quantum field theory, a symmetry is spontaneously broken when the Lagrangian itself remains invariant but the vacuum state, where the Hamiltonian of the theory is at its minimum, does not [18]. For the electroweak theory, this symmetry breaking is obtained through the introduction of a complex scalar field which has a non-zero vacuum expectation value (VEV) [19–23]. This field is an  $SU(2)$  doublet as represented in equation 1.6.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (1.6)$$

The electroweak Lagrangian can now be expressed in the simple form present in equation 1.7.

$$\mathcal{L}_{EW} = -\frac{1}{4}(\mathbf{F}_{\mu\nu} \cdot \mathbf{F}^{\mu\nu} + G_{\mu\nu}G^{\mu\nu}), \quad (1.7)$$

In this Lagrangian,  $\mathbf{F}_{\mu\nu}$  is the weak isospin and  $G_{\mu\nu}$  is the field strength tensor, which are related to the fields through equations 1.8 and 1.9

$$\mathbf{F}_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu - g \mathbf{W}_\mu \times \mathbf{W}_\nu \quad (1.8)$$

$$G_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (1.9)$$

where  $\mathbf{W}_\mu = (W_\mu^1, W_\mu^2, W_\mu^3)$ . An additional term appears as a consequence of the introduction of the complex scalar field as expressed in equation 1.10.

$$\mathcal{L}_\phi = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad (1.10)$$

$$\text{with } D_\mu = \partial_\mu - \frac{1}{2}(igT_i W_\mu^i - ig' B_\mu) \quad (1.11)$$

where  $T_i$  are the  $SU(2)$  group generators, and  $V(\phi)$  is the potential term which can be found in equation 1.12

$$V(\phi) = \lambda(\phi^\dagger \phi)^2 - \mu_{SM} \phi^\dagger \phi \quad (1.12)$$

where  $\lambda$  and  $\mu_{SM}$  are constants which take into account the self-interactions and the masses of the scalar fields. The vacuum states correspond to the minima of  $V(\phi)$  and its expectation values of  $\langle 0|\phi|0\rangle$ , which are expressed in equation 1.13.

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \text{ with } v = \sqrt{\frac{\mu_{SM}^2}{\lambda}} \quad (1.13)$$

In order to obtain physical particles, the perturbations around the vacuum state are taken into account. If  $\theta_i$  and  $H_{SM}$  are small variations in four degrees of freedom of the field,  $\phi$  can be expressed as equation 1.14.

$$\phi = \exp(-i\theta_i T^i / 2v) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H_{SM} \end{pmatrix} \quad (1.14)$$

The phase fields  $\theta_i$  can be set to zero by an appropriate gauge transformation which only leaves  $H_{SM}$ . This result can now be inserted into the Lagrangian, where  $H_{SM}$  is a scalar field with mass  $\sqrt{2}\mu_{SM}$  which in turn means the  $W_\mu^\pm$  and  $Z_\mu^0$  fields acquire mass terms  $m_W$  and  $m_Z$  as expressed in equation 1.15.

$$m_W = m_Z \cos \theta_W = \frac{gv}{2}. \quad (1.15)$$

Finally, equation 1.16 shows the form of mass terms for the fermions which are introduced via Yukawa interactions between the fermion and Higgs fields

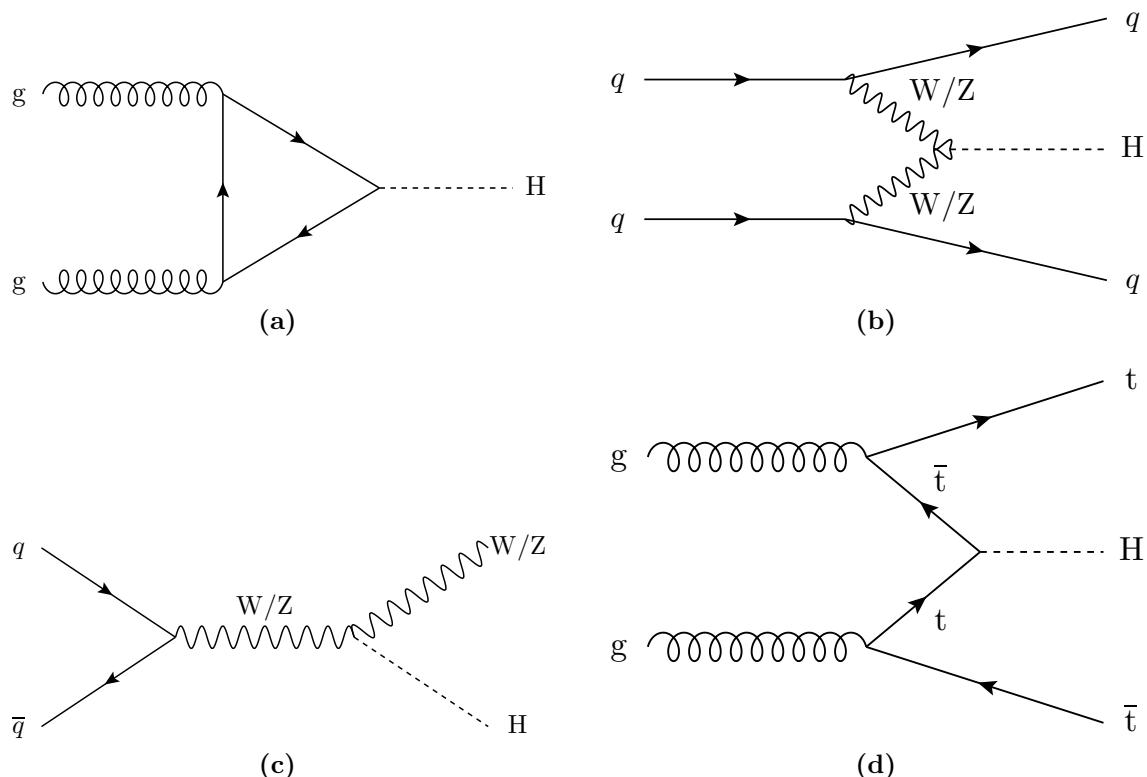
$$-\lambda_f (\overline{\psi}_L \phi \psi_R + \overline{\psi}_R \phi \psi_L), \quad (1.16)$$

where  $\lambda_f$  is the coupling for each fermion. Heavier fermions will have a stronger coupling to the Higgs boson, the value of  $\lambda_f$  will vary proportionally to the mass of the fermion,  $m_f$ . The value of each coupling is not predicted in the SM and have to be determined experimentally. The values of  $\sin \theta_W$ ,  $v$  and  $g$  can be determined with experimental values

of the  $W$  and  $Z$  masses and the fine structure constant, but the value of  $\mu_{SM}$  cannot be predicted. The mass of a Higgs boson associated with the Higgs field is  $m_H = \sqrt{2}\mu_{SM}$ , so it cannot be predicted directly by the SM, but indirect constraints can be imposed from theoretical considerations with the help of precision electroweak data, like the  $W$  and top quark mass measurements [24].

### 1.1.5 Searching for the SM Higgs boson

As described in the previous section, the Higgs boson mass is not directly predicted by the SM, which implied that searches for this particle needed to be performed covering the widest possible mass range. In particle accelerators, these searches are performed by looking for specific Higgs boson decays to either bosons or fermions. The coupling of the Higgs to a specific final state depends on both the mass of the Higgs and the masses of the particles of the final state.



**Figure 1.1:** Feynman diagrams for the main production processes of the SM Higgs boson at the Large Hadron Collider (LHC). Shown is (a) gluon fusion, (b) vector boson fusion and associated production with (c) vector bosons and (d) top quarks.

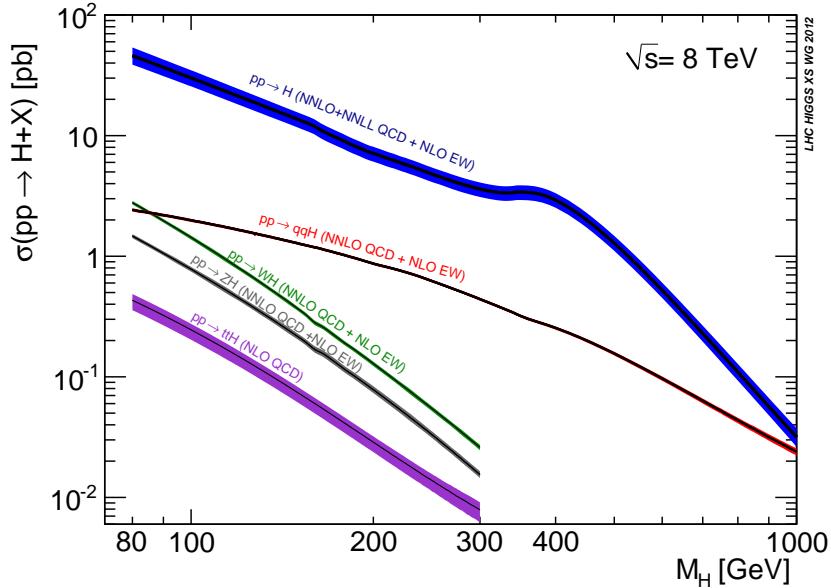
The production of the SM Higgs boson in proton-proton collisions occurs primarily through gluon-gluon fusion, Vector Boson Fusion (VBF) and associated production with either a vector boson or a pair of top quarks, the processes are illustrated in figure 1.1. The Higgs boson only couples to particles that have mass, hence it does not couple to gluons which are massless. The gluon fusion production process goes mainly through a top quark loop. The top quark is the heaviest quark and thus also has the highest coupling with the higgs. All these four processes are accessible at the LHC.

Searches for the Higgs boson have been already carried out at Large Electron Positron collider (LEP) and the Tevatron. The LEP was a particle accelerator colliding electrons and positrons at a center of mass energy ( $\sqrt{s}$ ) between 90 and 209 GeV. For this type of colliding particles the dominant production channel is associated production with a Z boson, this process is sometimes referred to as “Higgsstrahlung”. In the experiments of this accelerator the searches performed were predominantly looking for decays to  $b\bar{b}$  and  $\tau^+\tau^-$  pairs. The Higgs boson was not observed at LEP, leading to the exclusion of a SM Higgs with  $m_H < 114.4$  GeV at the 95% Confidence Level (CL) [25].

Searches were also performed by the CDF and D0 Collaborations at the Tevatron accelerator, which collided protons and antiprotons with  $\sqrt{s} = 1.96$  TeV. The experiments performed searches in the mass range of 90 – 200 GeV over the Higgs boson decays  $b\bar{b}$ ,  $W^+W^-$ ,  $\gamma\gamma$  and  $\tau^+\tau^-$  pairs, with the most sensitive being  $b\bar{b}$  and  $W^+W^-$ . The combined results of the Tevatron experiments resulted in an exclusion of a SM Higgs boson with  $m_H$  in the ranges 90 – 109 GeV and 149 – 182 GeV [26]. The results from the LEP and Tevatron direct searches can be combined with precision measurements of electroweak observables at LEP and by the SLAC Large Detector (SLD) to constrain the Higgs mass to  $94^{+29}_{-24}$  GeV [24], the quoted uncertainty only accounts for the experimental effects. Although, this value has been experimentally excluded, the mass point where the Higgs boson would eventually be found is just about the upper one sigma error limit.

The LHC is capable of colliding particles with a significantly higher centre of mass energy then the Tevatron. As a consequence it gives access to processes with smaller cross sections and allows searches in a wider mass range. Figure 1.2 shows the production cross section of different processes for proton-proton collisions with  $\sqrt{s} = 8$  TeV as operated during the 2012 period of LHC running. The dominant production process is gluon fusion by over one order of magnitude in cross section, for most of the mass range. Other processes, even with such low relative production rate are also useful. Their topological characteristics can be exploited to isolate signal-like events from the large quantities of background events. The second most likely production process is the VBF which includes

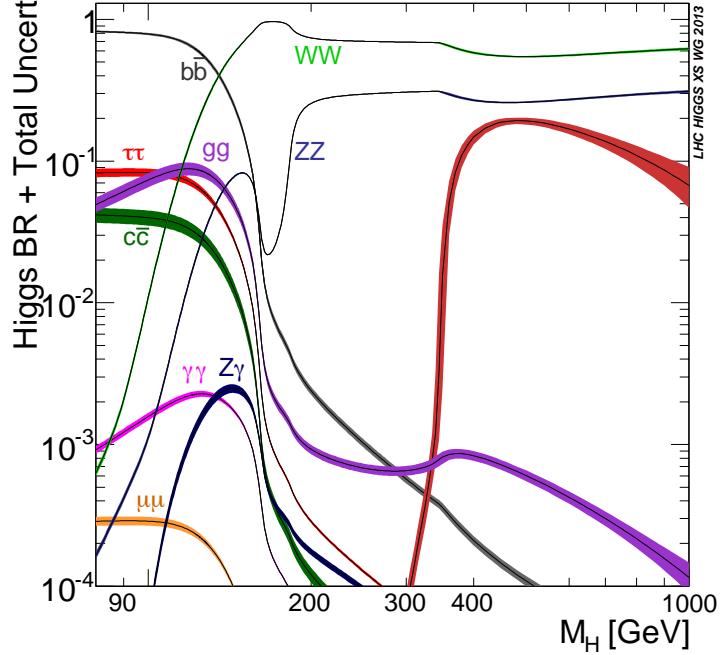
well separated high momentum quark jets in addition to the Higgs boson decay products. The associated production processes include a vector boson or a pair of top quarks in the final state. Since these particles can decay to leptons and b-quarks, those modes allow for good background rejection.



**Figure 1.2:** Cross sections for Higgs production processes at  $\sqrt{s} = 8$  TeV for a range of Higgs boson masses  $m_H$  [27]. Across the mass range, the gluon-fusion mode dominates, followed by the vector boson fusion and associated production modes. The widths of the lines represent the theoretical uncertainties on the cross section calculation.

Figure 1.3 shows the branching fractions to the different decay channels depending on the Higgs boson mass. At low Higgs boson mass, many possible decay channels are accessible. The decay into two photons, which cannot happen directly as the photon is massless, happens via fermion or W loops. Other decays are also possible to  $W^+W^-$ ,  $ZZ$ ,  $b\bar{b}$  and  $\tau^+\tau^-$ . For Higgs masses above 130 GeV, decays to  $W^+W^-$ ,  $ZZ$  dominate as they become kinematically favourable. The observation of a Higgs boson in any of these decay channels is in itself important and gives a handle to probe the couplings of the particles involved in the Higgs production and decay. The most sensitive decays at the low mass region are  $\gamma\gamma$  and  $ZZ$  due to their clean signatures.

The ATLAS and CMS Collaborations announced the discovery of a new boson with a mass around 125 GeV [2, 3]. To achieve this result both experiments analysed approximately  $5 \text{ fb}^{-1}$  of data collected at  $\sqrt{s} = 7 \text{ TeV}$  and  $5 - 6 \text{ fb}^{-1}$  at  $\sqrt{s} = 8 \text{ TeV}$ . This discovery was made by combining searches using  $ZZ$  and  $\gamma\gamma$  decay modes in both experiments, the



**Figure 1.3:** Higgs boson branching ratios in the SM for a range of Higgs boson masses  $m_H$  [27]. At high masses, above their kinematic thresholds, the WW, ZZ and  $t\bar{t}$  (shown in red) decay modes dominate. At lower masses a wide range of different final states is possible. The widths of the lines represent the theoretical uncertainties on the branching ratio calculation.

observed combined excess of events yielded a  $5\sigma$  deviation from the background-only expectation.

The LHC Run I was completed in early 2013, with ATLAS and Compact Muon Solenoid (CMS) recording  $\approx 20 \text{ fb}^{-1}$  at 8 TeV. This increase in luminosity allowed access to less sensitive decay modes. The combination of both experiments ZZ and  $\gamma\gamma$  measurements results in best fit mass of  $125.09 \pm 0.21(\text{stat}) \pm 0.11(\text{syst}) \text{ GeV}$  and a signal strength relative to the SM of  $1.24^{+0.18}_{-0.16}$  [28]. The CMS collaboration has also performed a signal strength measurement combining the information from  $\gamma\gamma$ ,  $W^+W^-$ , ZZ,  $b\bar{b}$ ,  $\tau^+\tau^-$  and  $\mu^+\mu^-$  final states, obtaining a signal strength of  $1.00 \pm 0.09(\text{stat})^{+0.08}_{-0.07}(\text{theo}) \pm 0.01(\text{syst})$  [29]. Individual channels were studied in their compatibility with the SM, and all of them show consistency with the SM predictions for a 125 GeV Higgs boson. The ATLAS collaboration has also performed similar studies over production rates and couplings for various channels [30–32], and both collaborations have performed studies on the spin-parity quantum numbers [33–35] and limits have also been set over the invisible branching fraction which will be analysed in depth in this document. No significant deviations from the predictions of the SM have been observed to date.

### 1.1.6 Invisible Higgs decay

The discovery of the Higgs boson described in the previous chapter, and the absence of any experimental hints of new physics beyond the SM at the LHC, have strongly limited proposed models for new physics. Currently, the uncertainties associated with the Higgs boson are still large enough to allow for the possibility of non-SM properties. Although additional SM-like Higgs bosons have been excluded over a wide range of masses, there is still the possibility of additional Higgs bosons with exotic decay modes.

Invisible Higgs boson decays are predicted in a wide range of models, for example in supersymmetric models to neutralinos [36] or graviscalars in models with extra dimensions [37, 38]. If the Higgs boson can interact with the currently unknown dark matter sector, invisible decay modes could be possible, and bounds on these decays can constrain dark matter models.

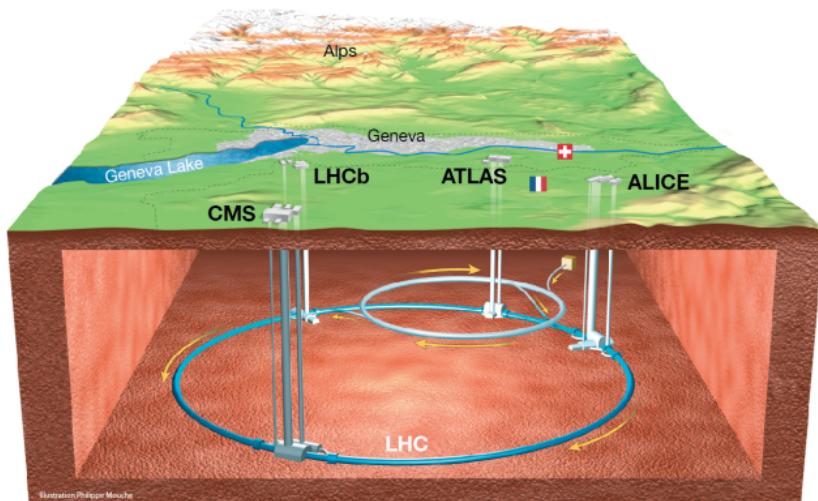
The SM Higgs boson can decay to neutrinos through ZZ intermediate decay, but this decay has a very small branch ratio of only 0.106%. The observation of a significant branching fraction to invisible would be clear evidence of new physics and would point to direct production of dark matter.

# Chapter 2

## Experimental Apparatus

### 2.1 The Large Hadron Collider

The LHC [39, 40] is currently the world's largest particle accelerator and is capable of producing the highest energy particle beams ever made by mankind. This machine has total perimeter of 26.7 km and was built at the European Organization for Nuclear Research (CERN) in a circular tunnel, which previously housed the LEP collider [41], at an average depth of 100 m below ground under the Franco-Swiss border near Geneva, Switzerland. A diagram of the LHC tunnel and its experiments can be found at figure 2.1.



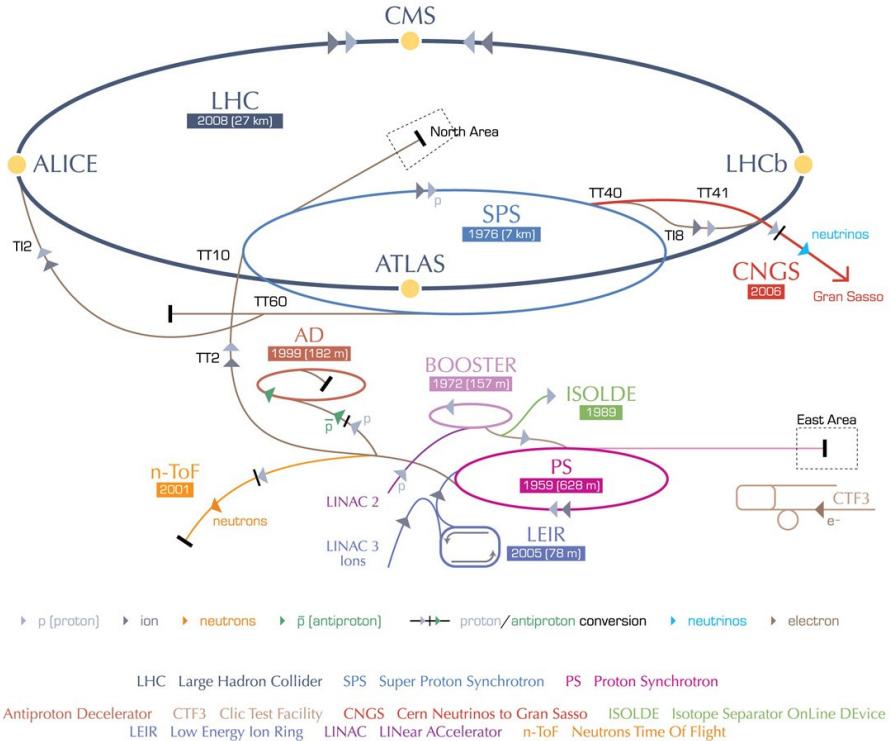
**Figure 2.1:** Underground diagram of the Geneva area showing the LHC and its experiments location [42].

The LHC is a synchrotron machine with the capability of accelerating two particle beams in opposite directions in two separate beam pipes. These beams only cross and are forced to collide in four points of the accelerator where particle detectors are installed to observe the products of such collisions. This experiments are: A Toroidal LHC ApparatuS (ATLAS) [43], CMS [44], Large Hadron Collider beauty (LHCb) [45] and A Large Ion Collider Experiment (ALICE) [46].

The objective of the LHC program is to investigate physics at the TeV scale, more specifically to understand the electroweak symmetry breaking and if this phenomenon could be explained by the Higgs mechanism. There are many Beyond the Standard Model (BSM) theories that predict new physics at this energy regime, making the LHC the perfect machine to investigate such phenomena. ATLAS and CMS are general-purpose detectors which aim to investigate a broad spectrum of physics. The LHCb detector is used to study processes that involve the decay of b-flavoured hadrons. The ALICE detector is optimised to look at heavy-ion collisions and to investigate the properties of the extreme density medium that is formed.

The LHC is only the last element of a complex accelerator chain which step-by-step increases the energy of the particles that will eventually be collided [40]. Protons are initially obtained by stripping the electrons of hydrogen gas. This process happens at the beggining of the Linear Particle Accelerator 2 (LINAC2) which then accelerates them up to the energy of 50 MeV. After this initial step, protons are injected into the Proton Synchrotron Booster (PSB) and the energy ramps ups to 1.4 GeV. Particles are then passed to the Proton Synchrotron (PS) where the energy further increases to 25 GeV. Subsequently they are injected into the Super Proton Synchrotron (SPS) where the particle energy reaches 450 GeV. Finally, protons pass to the LHC where they can be accelerated to a maximum energy of 7 TeV. A simplified diagram of the CERN accelerator chain can be found in figure 2.2. Normal operation of the LHC therefore depends on all the upstream accelerators availability. The turn around time, the time necessary to stop the accelerator from running and restart collisions, can be as low as 2 hours. When stable beams are achieved, a single proton fill can be used to collide protons for up to 24 hours, but it is common to restart more frequently to profit from the higher collision rates possible right at the beginning of a new fill.

Each beam pipe can be filled with proton or heavy ions. Three modes of operation have been tried: proton-proton, proton-lead ion and lead ion-lead ion. By changing the incoming particles we are changing the quantity of nucleons present at each interaction. The maximum design energy per proton is 7 TeV and is 2.76 TeV for each lead nucleon. The



**Figure 2.2:** Diagram of the CERN accelerator complex [47].

maximum design luminosity for proton-proton collisions is  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  and  $10^{27} \text{ cm}^{-2}\text{s}^{-1}$  for lead ion-lead ion collisions.

Particle beams trajectories are curved by 1232 niobium-titanium superconducting dipole magnets each with a length of 14.3 m. They are cooled with superfluid helium to 1.9 K and produce the necessary magnetic field of 8.4 T. Eight Radio Frequency (RF) cavities located at the LHC point 4 are used to accelerate the beams. At nominal operation the LHC will steer 2808 bunches composed up to  $10^{11}$  protons separated by 25 ns in each direction. Some of the key parameters of the LHC proton-proton and lead-lead operation can be found in table 2.1.

At the LHC we are looking for extremely rare processes. As it can be seen in figure 2.3, the production cross section of a SM Higgs boson is more than 9 orders of magnitude smaller than the total proton-proton cross section.

To be able to record and study such rare processes we need to produce a significant number of collisions. For this purpose the LHC was designed to operate at high instantaneous luminosity,  $L$ . This quantity is defined as,

		<i>pp</i>	<b>HI</b>	
Energy per nucleon	E	7	2.76	TeV
Dipole field at 7 TeV	B	8.33	8.33	T
Design Luminosity*	$\mathcal{L}$	$10^{34}$	$10^{27}$	$\text{cm}^{-2}\text{s}^{-1}$
Bunch separation		25	100	ns
No. of bunches	$k_B$	2808	592	
No. particles per bunch	$N_p$	$1.15 \times 10^{11}$	$7.0 \times 10^7$	
<b>Collisions</b>				
$\beta$ -value at IP	$\beta^*$	0.55	0.5	m
RMS beam radius at IP	$\sigma^*$	16.7	15.9	$\mu\text{m}$
Luminosity lifetime	$\tau_L$	15	6	h
Number of collisions/crossing	$n_c$	$\approx 20$	-	

\* For heavy-ion (HI) operation the design luminosity for Pb-Pb collisions is given.

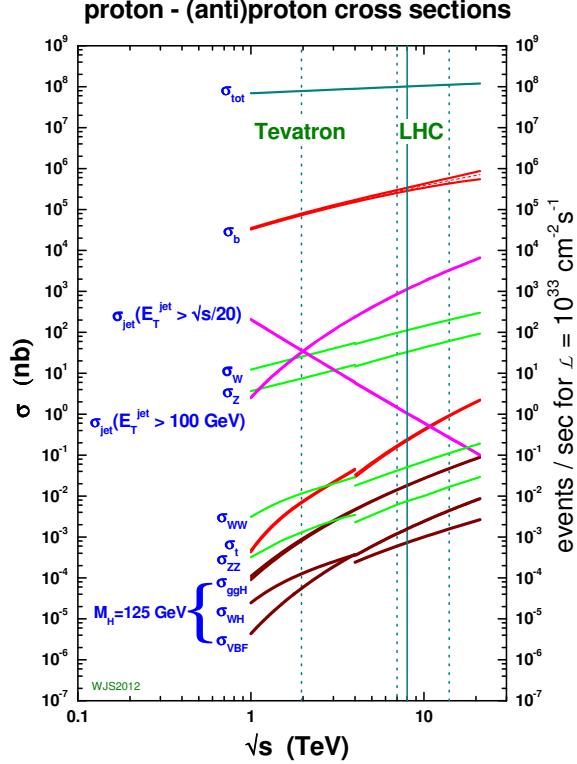
**Table 2.1:** The machine parameters relevant for the LHC detectors.[48]

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma}{4\pi \epsilon_n \beta^*} F, \quad (2.1)$$

where  $N_b$  is the number of protons per bunch,  $n_b$  is the number of bunches,  $f_{\text{rev}}$  is the frequency of revolution,  $\gamma$  is the Lorentz factor,  $\epsilon_n$  is the normalized emittance,  $\beta^*$  is the beta function at the collision point and  $F$  is the reduction factor due to the crossing angle.

### 2.1.1 Running and performance

Operation of the LHC started when the first beams circulated in the machine in September 2008. Unfortunately, only a few days later, a faulty weld between two dipole magnets caused a significant magnet quench which in turn damaged several dipoles and a simultaneous leak of a significant amount of helium happened. The event showed that beyond the repair of the affected systems the accelerator needed a significant consolidation program to allow it to return to activity [49]. This consolidation program took over one year to finalise and to prevent further possible problems and allow better understanding of the machine while maximizing physics reach, it was decided to initially run the LHC at

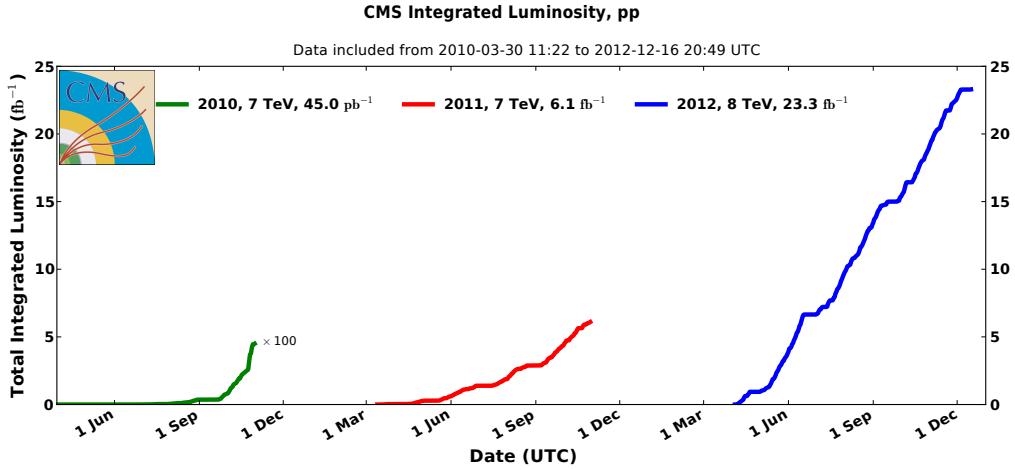


**Figure 2.3:** Cross sections for several processes for collisions of antiproton-proton and proton-proton as a function of the center of mass energy [44].

7 TeV center-of-mass energy. First collisions happened at November 2009 just at the SPS injection energy of 450 GeV giving start to the LHC run I.

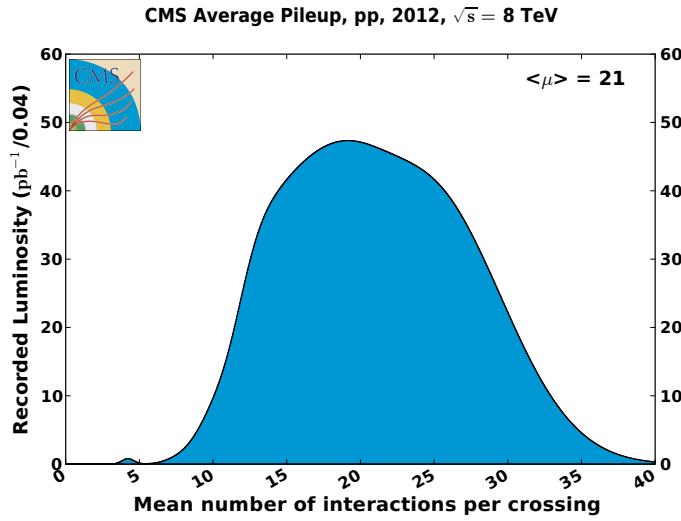
The collision energy was finally ramped up to 7 TeV with first collisions being observed during March 2010. Operation at this energy continued until the end of 2011, with a peak luminosity being achieved of  $3.7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ . The total integrated luminosity delivered to CMS was  $6.1 \text{ fb}^{-1}$  with the total actually recorded being  $5.6 \text{ fb}^{-1}$ . During 2012 with the increased knowledge of the accelerator it was possible to increase the centre-of-mass energy further to 8 TeV, eventually reaching peak luminosity of  $7.7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-2}$  and delivering integrated luminosity of  $23.3 \text{ fb}^{-1}$  to CMS of which  $21.79 \text{ fb}^{-1}$  were recorded. Figure 2.4 shows the delivered luminosity in the period 2010-2013 over time.

For physics usage, data needs to undergo a certification process. In this process specialists from each CMS subsystem check that no problem has happened during data taking that would bias or invalidate the recorded events. For 2011 a total of  $5.1 \text{ fb}^{-1}$  and for 2012 a total  $19.7 \text{ fb}^{-1}$  were considered of good quality for physics.



**Figure 2.4:** Cumulative luminosity versus day delivered to CMS during stable beams and for p-p collisions. This is shown for 2010 (green), 2011 (red) and 2012 (blue) data-taking [50].

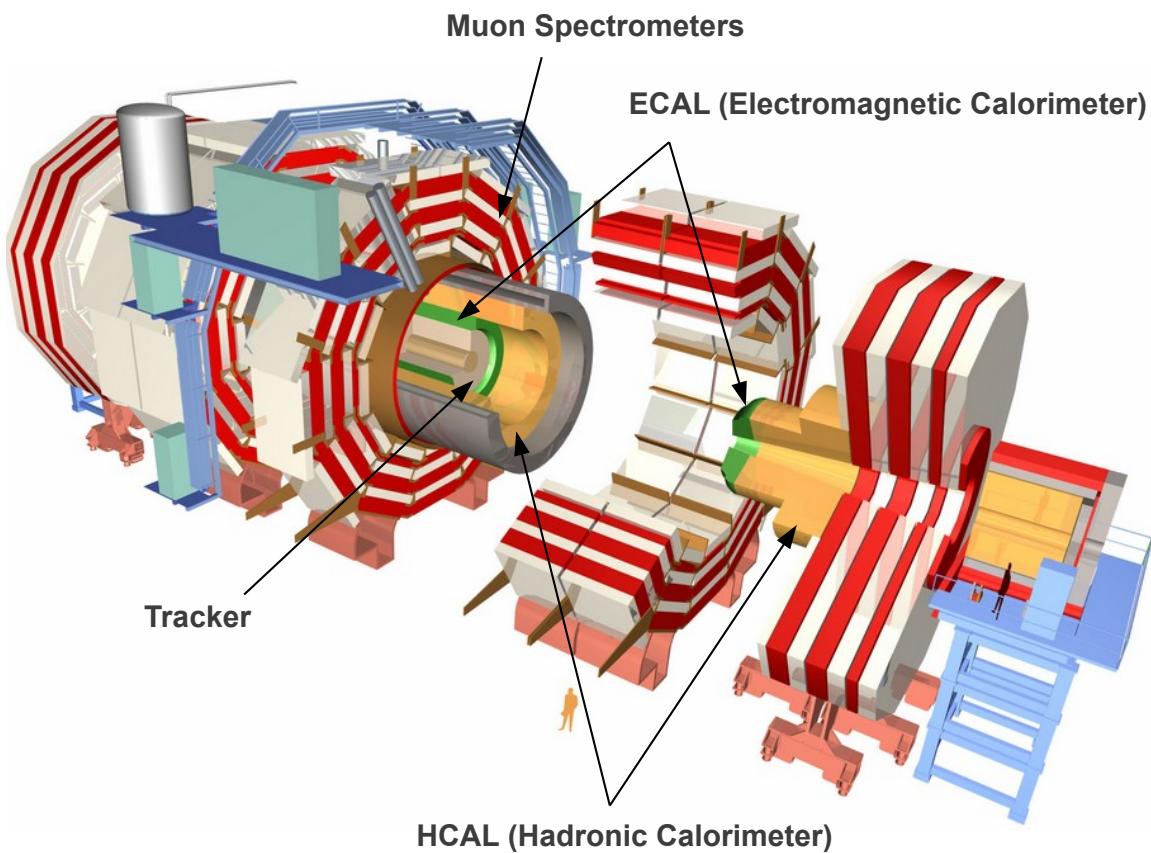
In order to achieve high integrated luminosity LHC collides particle bunches up to 40 million times a second, and many interactions may happen simultaneously, this effect is called Pile-Up (PU). Figure 2.5 shows the distribution of the mean number of interaction per bunch crossing during 2012 at the CMS experiment.



**Figure 2.5:** Mean number of interactions per bunch crossing at the CMS experiment during 2012. The year's total mean number of interactions per crossing,  $\langle \mu \rangle$ , was of 21 collisions [51].

## 2.2 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) experiment [44] is a general purpose experiment located at the LHC point 5, near the village of Cessy, France. It was designed to be a high performance detector studying collisions at its centre. It is composed of several subsystems in a classic onion shaped structure. A diagram of the experiment can be found in figure 2.6.



**Figure 2.6:** Diagram of CMS experiment showing the experiment in an open configuration and highlighting the position of its sub-detectors [52].

The main driving motivation for its design was to investigate the electroweak symmetry breaking and the Higgs mechanism which at the design time was presumed to be the most likely explanation. Many alternative theories to the standard model predict new particles which could be observed at the TeV scale. CMS as a multi-purpose experiment is well suited to search for these new scenarios. If found, such new physics may allow us to understand some of the currently open questions in particle physics, like providing particle candidates for dark matter. Furthermore, some of these possible new physics

signals could point the way towards a grand unified theory. CMS is also capable of operating while the LHC is colliding heavy ions and has a rich program covering the study of matter at extreme temperatures, densities and parton momentum fraction (low-x).

The requirements imposed on the CMS design to meet its physics goals can be summarized in the following table [44, 53]:

- Good muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution ( $\approx 1\%$  at 100 GeV), and the ability to determine unambiguously the charge of muons with  $p_T < 1$  TeV.
- Good charged-particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of  $\tau$ 's and b-jets, requiring pixel detectors close to the interaction region.
- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution ( $\approx 1\%$  at 100 GeV), wide geometric coverage,  $\pi^0$  rejection, and efficient photon and lepton isolation at high luminosities.
- Good missing-transverse-energy and dijet-mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

The final detector design fulfils all these requirements. The experiment is compact compared to the other LHC experiments being 22 m long and 15 m in diameter. Although small, it is the heaviest of the four big detectors at 12500 tonnes. Its high density is a direct consequence of it producing the highest magnetic field at 4 T and therefore needing more material for it to be contained in its return yoke.

### 2.2.1 Geometry and conventions

The adopted coordinate system has its origin in the center of CMS, where the nominal collision point is located, the  $y$ -axis points vertically upwards, and the  $x$ -axis points radially inward in the direction of the centre of the LHC. The  $z$ -axis points along the beam line towards the Jura mountains from the LHC point 5. The azimuthal angle  $\phi$  is measured from the  $x$ -axis in the  $x$ - $y$  plane. The polar angle  $\theta$  is measured from the  $z$ -axis.

We define pseudorapidity as  $\eta = -\ln(\tan(\theta/2))$ . All transverse quantities, like the transverse momentum ( $p_\perp$ ), are measured in the transverse plane to the beam axis. The imbalance of energy is also measured in the  $x$ - $y$  plane and is denoted as  $E_\perp^{miss}$ .

## 2.2.2 Inner tracking system

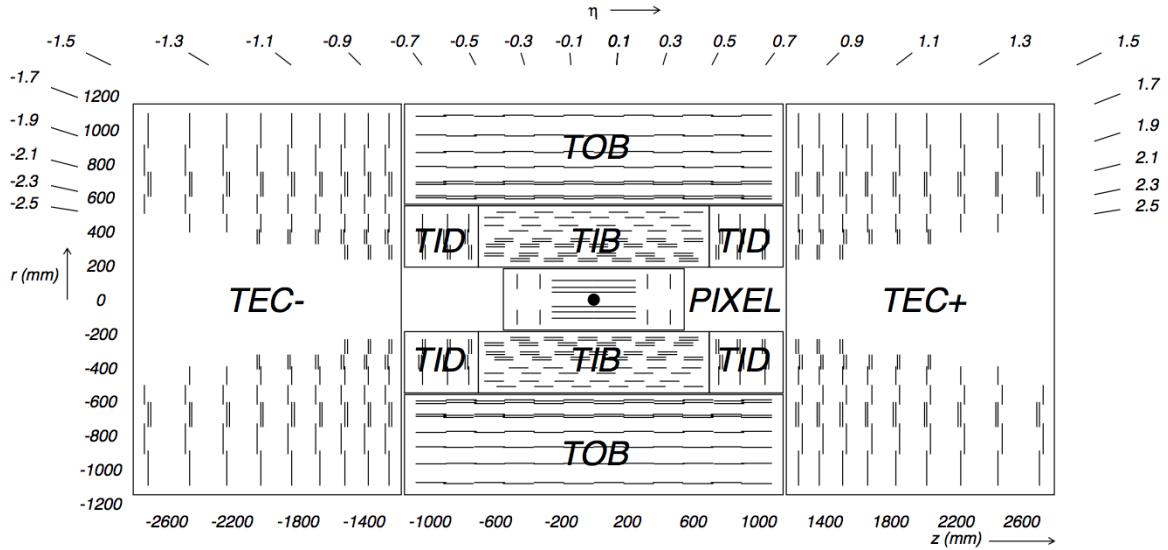
The inner tracking system is the closest detector to the beam axis and the interaction region [54, 55]. Its function is to measure the trajectory of all charged particles with momentum above 1 GeV being produced at each LHC collision. With the help of the strong magnetic field produced by the CMS magnet, particle trajectories are bent allowing for charge and momentum determination. With the resulting tracks it is then possible to determine the primary vertex as well as secondary vertices like other lower energy proton-proton collisions or displaced vertices from the decay of long-lived particles like B-hadrons.

Building a tracking system for an experiment at the LHC is very challenging. Such system at design luminosity will be hit by an average of 1000 particles per beam crossing at a rate approaching 40 MHz. It needs to be a fast, efficient, high granularity detector, radiation hard and as thin as possible to not deflect the incoming particles trajectory. At each layer the occupancy should be of the order of 1% or lower. These design requirements have led to a tracker design entirely based on silicon detector technology.

The volume near the interaction point can be split according to the charged particle flux into three regions:

- $r < 10$  cm: highest particle flux, up to  $\approx 10^8 \text{ cm}^{-2}\text{s}^{-1}$  at  $r \approx 4\text{cm}$ , pixel detectors are used. The pixel size is  $\approx 100 \times 150 \mu\text{m}^2$ , which translates into an occupancy of  $10^{-4}$  per LHC bunch crossing.
- $20 < r < 55$  cm: particle flux decreases enough to use silicon micro-strips with a minimum cell size of  $10\text{ cm} \times 80 \mu\text{m}$ , leading to an occupancy of  $\approx 2 - 3\%$  per LHC bunch crossing.
- $50 < r < 110$  cm: most outer region of the tracker, particle flux is low enough to use larger pitch silicon micro-strips. The maximum cell size is of  $25\text{ cm} \times 180 \mu\text{m}$ , and occupancy is of the order of  $\approx 1\%$ .

The CMS tracker final configuration is composed of a pixel detector with three barrel layers at radii between 4.4 cm and 10.2 cm and 2 disks on each side of the barrel. A silicon strip tracker with 10 barrel detection layers extending up to 1.1 m with 3 inner tracker plus 9 endcap disks on each side of the barrel. A schematic of this detector module distribution can be found at figure 2.7. This detector has an acceptance covering up to pseudorapidity of  $|\eta| < 2.5$  and has a total active area of about  $200 \text{ m}^2$  making it the largest silicon tracker ever built.



**Figure 2.7:** Schematic cross section of the CMS tracker [44]. Each line represents a detector module. Double lines represent dual surface back-to-back detector modules. The inner tracker components are highlighted: Tracker Endcaps (TEC), Tracker Outer Barrel (TOB), Tracker Inner Disks (TID), Tracker Inner Barrel (TIB) and Pixels.

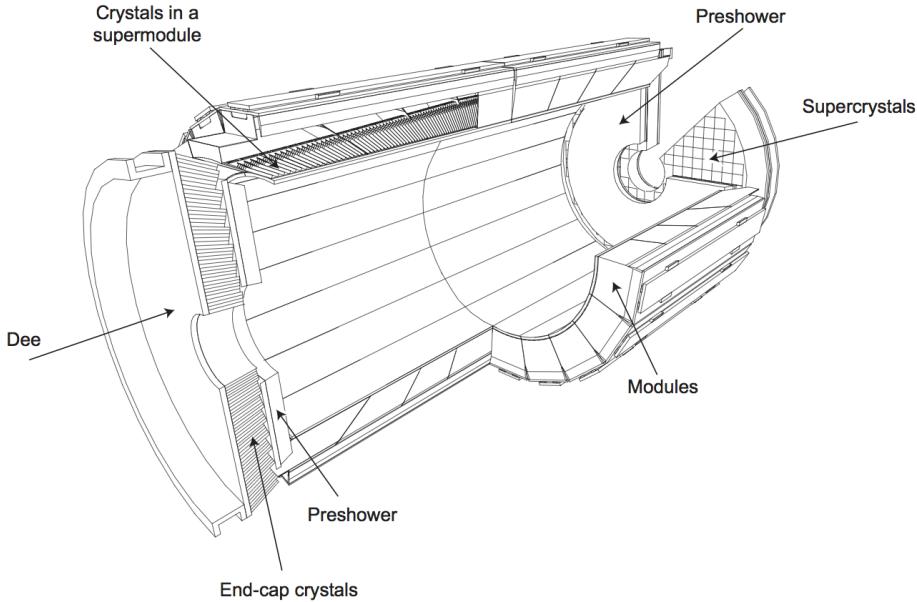
### 2.2.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) is the detector responsible for measuring the energy of electrons and photons [56, 57]. It is an almost hermetic energy measurement system comprising 61200 lead-tungstate ( $PbWO_4$ ) crystals mounted in the barrel and 7324 crystals in each of the 2 endcaps and it has an acceptance up to  $|\eta| < 3.0$ .

Lead tungstate has a fairly high density ( $8.28 \text{ g/cm}^3$ ), a short radiation length (0.89 cm) and a small Moliere redius (2.2 cm). The crystals also have a fast scintillation decay time, emitting 80% of the light yield in 25 ns (the minimal bunch crossing time at the LHC). These characteristics make it a good choice for an electromagnetic calorimeter allowing a compact design with fine granularity. However, this crystals emit a fairly low light yield ( $30 \gamma/\text{MeV}$ ) which requires the use of photo-detectors with intrinsic gain which will preform well inside a magnetic field. In the barrel region silicon Avalanche photo-diodes (APD) are used and Vacuum Photo-Triodes (VPT) are used in the endcaps. To guarantee good response from both crystals and APD it is necessary to have system thermal stability, with the goal being temperature variation of less than  $0.1^\circ\text{C}$ , which is achieved with a water cooling system running at  $18^\circ\text{C}$ .

The barrel sections, the ECAL Barrel (EB), has an inner radius of 129 cm and is composed of 36 identical “supermodules”, each covers the barrel length and corresponding to a pseudo-rapidity interval of  $0 < |\eta| < 1.479$ . The crystals are quasi-projective (the axes are tilted at  $3^\circ$  with respect to the line from the nominal vertex position) and cover 0.0174 (i.e.  $1^\circ$ ) in  $\Delta\phi$  and  $\Delta\eta$ . The crystals have a front face cross-section of  $\approx 22 \times 22 \text{ mm}^2$  and a length of 230 mm, corresponding to  $25.8 X_0$ .

The endcap section, the ECAL Endcaps (EE), are at a distance of 314 cm from the vertex and covering a pseudorapidity range of  $1.479 < |\eta| < 3.0$ , each structured as 2 “Dees”, consisting of semi-circular aluminium plates from which are cantilevered structural units of  $5 \times 5$  crystals, known as “supercrystals”. A diagram of the ECAL can be found on figure 2.8.



**Figure 2.8:** Diagram of the ECAL layout illustrating the positions of its components. The barrel has an inner radius of 129 cm and the encaps are at a distance of 315.4 cm from the vertex [44].

The energy resolution of the ECAL can be expressed as:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \quad (2.2)$$

Here  $E$  is the energy of the incoming particle in GeV,  $S$  is the stochastic term, which quantifies the fluctuations in scintillation and lateral containment of the shower,  $N$  the

noise term, which relates to the electronics and digitisation process, and finally  $C$  is a constant term that quantifies the non-uniform longitudinal response and inter-calibration errors. These parameters have been measured to be  $S = 0.028 \text{ GeV}^{1/2}$ ,  $N = 0.12 \text{ GeV}$  and  $C = 0.003$  with the help of an electron beam [58] and in the absence of magnetic field.

### Preshower detector

The CMS Preshower is a detector located in front of each endcap covering the fiducial region of  $1.653 < |\eta| < 2.6$ . Its mission is to identify neutral pion decays, help to identify electrons against minimum ionizing particles and improve electron and photon position determination.

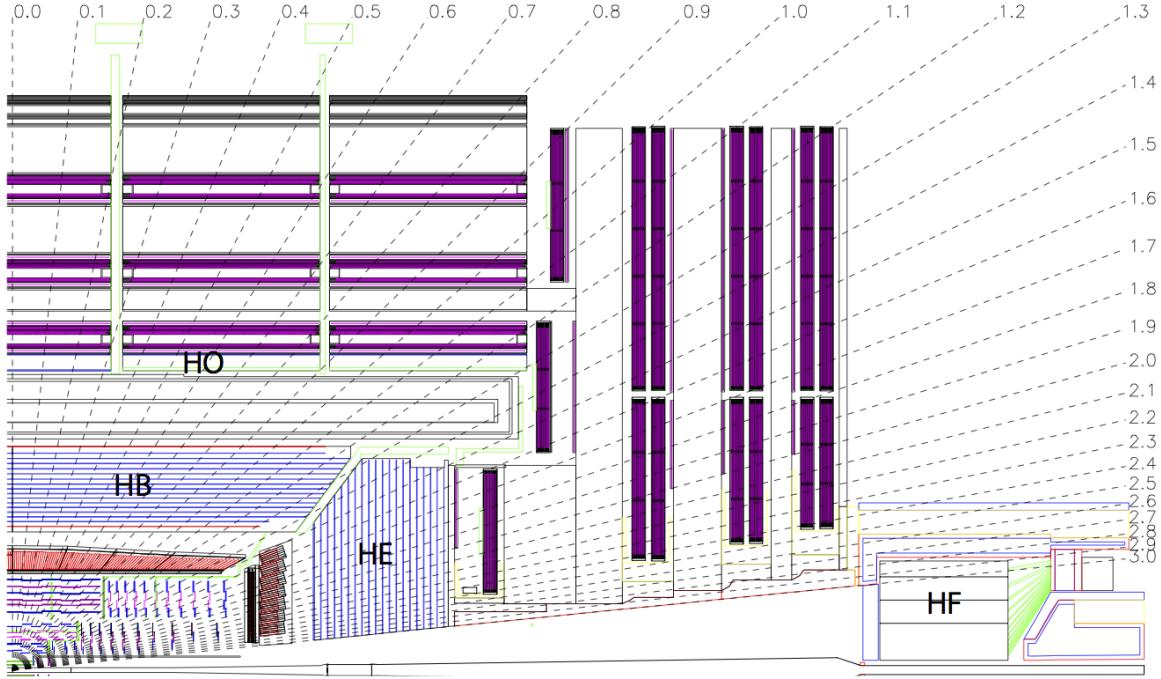
This detector is a sampling calorimeter composed of two layers of lead radiators each followed by silicon strip sensors. The lead layers have the function of forcing the incoming particles to initiate an electromagnetic shower. The first lead layer has  $2X_0$  while the second has  $1X_0$ , which results in 95% of the single incident photons starting their shower before hitting the first sensor [44]. The shape of the lead layers edge matches the ECAL crystal behind them to facilitate calculations at the Level 1 Trigger (L1T).

Each silicon sensor has an active area of  $61 \times 61 \text{ mm}$  and is  $320 \mu\text{m}$  thick. The sensors are divided into 32 strips, each  $1.9 \text{ mm}$  long. The preshower system has a total thickness of  $20 \text{ cm}$  and 137000 individual read-out channels.

#### 2.2.4 Hadronic Calorimeter

The Hadronic Calorimeter (HCAL) is a sampling calorimeter which is designed to measure the properties of hadron jets and indirectly neutrinos or other undiscovered particles that would result in apparent missing energy [59]. The design of the Hadronic Calorimeter (HCAL) was strongly influenced by the choice of the magnet parameters as most of the calorimetry is inside of the magnet. A diagram of the HCAL subsystems and their location inside CMS can be found in figure 2.9.

The HCAL Barrel (HB) covers the region up to  $|\eta| < 1.3$  and is limited from the beam side by the ECAL at radius  $r = 1.77 \text{ m}$  and outwards by the magnet at radius  $r = 2.95 \text{ m}$ . This is a strict limitation on the amount of absorber material to be used. This detector is composed of 36 identical azimuthal wedges split in two half-barrels. They are constructed



**Figure 2.9:** Longitudinal view of the CMS detector highlighting the location of the HCAL components: HCAL Barrel (HB), HCAL Endcap (HE), HCAL Outer (HO) and HCAL Forward (HF). The dashed lines show different pseudo-rapidity values. The experiment has an approximate length of 21.6 m long and 15 m [44].

of brass absorber plates alternated with plastic scintillator. Brass has a short interaction length ( $X_0 = 16.42$  cm) and is non-magnetic. The detector is composed of 2304 towers with a segmentation of  $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$  which correspond to the same area of the  $5 \times 5$  arrays of ECAL crystals.

To improve the measurement capability, an outer calorimeter, the HCAL Outer (HO), is placed outside of the magnet as a *tail catcher*. It increases the effective thickness of the hadronic calorimeter by over 10 interaction lengths. This detector covers the range  $|\eta| < 1.26$ , it is composed by an iron absorber and scintillator and is subdivided into sectors that cover  $30^\circ$  azimuthal angle in each of the barrel wheels.

The HCAL Endcap (HE) covers the range of  $1.3 < |\eta| < 3.0$ . It is composed by 2034 towers segmented in  $\eta$  (14 strips) and  $\phi$  ( $5^\circ$  sectors). In the 8 innermost towers the segmentation is  $10^\circ$  in  $\phi$ , whilst the  $\eta$  segmentation increases in  $\eta$  from 0.09 to 0.35.

Additionally, to extend acceptance to  $|\eta| < 5.2$  the HF is installed at 11.2 m from the interaction point providing excellent hermeticity for  $E_{\perp}^{\text{miss}}$  measurement. Its steel absorber is 1.65 m deep and has quartz fibres running through it, parallel to the beam

line. The energy measurement is made via Cerenkov light produced by the incoming particles inside the fibres. There are 13 towers in  $\eta$  with segmentation of  $\approx \Delta\eta = 0.175$  except the lowest  $\eta$  tower with  $\approx \Delta\eta = 0.1$  and highest  $\eta$  tower with  $\approx \Delta\eta = 0.3$ . The segmentation in  $\phi$  is  $\Delta\phi = 10^\circ$  except in the highest  $\eta$  towers where  $\Delta\phi = 20^\circ$ . There are a total of 900 towers per HF module.

Similarly to the ECAL, the energy resolution of the HCAL was tested using a test beam of single charged pions [58], and it was obtained that:

$$\frac{\sigma}{E} = \frac{94.3\%}{\sqrt{E}} \oplus 8.4\%. \quad (2.3)$$

### 2.2.5 Solenoid Magnet

The design requirements for correct charge assignment and  $p_T$  determination for charged particles, and especially muons, drive the magnet parameters choice. For muons, unambiguous charge determination requires momentum resolution of  $\Delta p/p \approx 10\%$  at  $p = 1\text{TeV}$ . These requirements are especially difficult to obtain in the forward regions but with the correct length/radius ratio can be obtained with a modestly sized solenoid magnet but with large field [48, 60].

The choice of the CMS collaboration was to build a Niobium-Titanium (NbTi) superconducting solenoid magnet which has been designed to operate at fields up to 4 T. It has a diameter of 6 m and a length of 12.5 m and at maximum field the stored energy reaches 2.7 GJ. Typically, the magnet is only run at 3.8 T in order to maximize its lifetime. To contain such an enormous magnetic flux, a 10 kt return yoke envelopes the magnet with 5 wheels in the barrel region and 2 endcaps composed of 3 disks closing the sides [44]. A summary of the most important magnet parameters can be found at table 2.2.

### 2.2.6 Muon System

The muon detection is an important part of the mission of CMS [61]. Muons are fairly easy to detect when compared with other elementary particles and are only rarely produced in proton-proton collisions. Taking the example of a SM Higgs boson with  $m_H = 125\text{ GeV}$ , while the decay mode involving a pair of Z bosons is fairly unlikely, only happening  $\approx 2.64\%$  of the times [6], the Z bosons can decay into 4 muons. This decay, while rare,

Parameter	Value
Field	4 T
Inner Bore	5.9 m
Length	12.9 m
Number of turns	2168
Current	19.5 kA
Stored Energy	2.7 GJ
Hoop Stress	64 atm

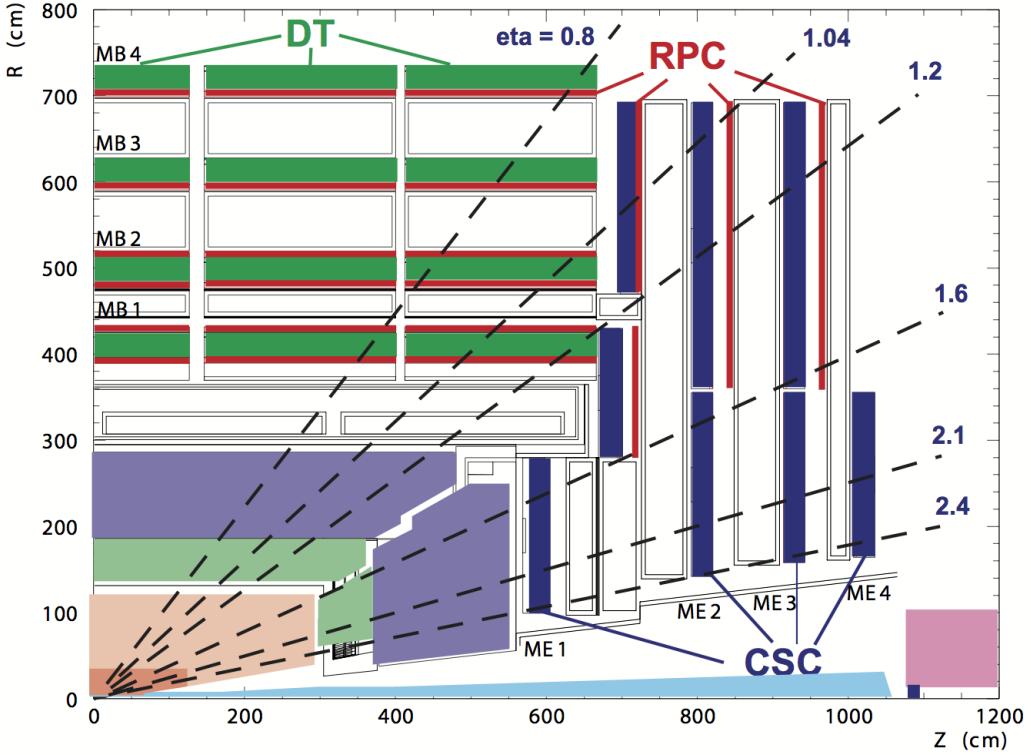
**Table 2.2:** Parameters of the CMS superconducting solenoid [48]

does not have significant backgrounds making it a “golden channel” for discovery, which indeed has proven the case [62]. Many other models, like Supersymmetry (SUSY), use muon final states in their searches for exactly the same reason. The CMS muon system is composed of 3 types of gaseous detectors depending on the location and momentum reconstruction needs. A diagram of the disposition of this system inside CMS can be found on figure 2.10.

In the barrel and up to  $|\eta| < 1.2$ , Drift Tube (DT) are used, since the neutron background is small and the field is constant. This system is composed of 250 chambers and is arranged in 4 concentric cylindrical layers which are installed inside of the return yoke. These chambers have a total of 172000 wires with a length of 2.4 m which are housed inside of tubes filled with a mixture of argon and carbon-dioxide. Each of the barrel wheels is split into 12 sectors covering  $30^\circ$  azimuthal angle. The maximum gas ionization drift is 2.0 cm and results in a single point resolution of  $\approx 200 \mu\text{m}$  per wire. For each station and each measured muon, the  $\phi$  resolution is better than  $200 \mu\text{m}$  and the direction resolution is  $\approx 1 \text{ mrad}$ .

In the endcaps, CSC are used in the region between  $2.4 > |\eta| > 0.9$ . Here, muon and background rates are high and the magnetic field is not uniform. This system has fast response and is radiation resistant. It is composed of 468 chambers arranged in 4 stations per side. Each chamber is trapezoidal in shape and made of 6 gas gaps and covers either  $10^\circ$  or  $20^\circ$  in  $\phi$ . Each gap contains a plane of cathode strips and a plane of anode wires. For each chamber the spatial resolution is of the order of  $200 \mu\text{m}$  and the angular resolution is  $\approx 10 \text{ mrad}$  in  $\phi$ .

Finally the RPC covers the  $|\eta| < 1.6$  range. This system overlaps with the 2 other muon systems. It is very fast with an ionization event being much faster than the bunch



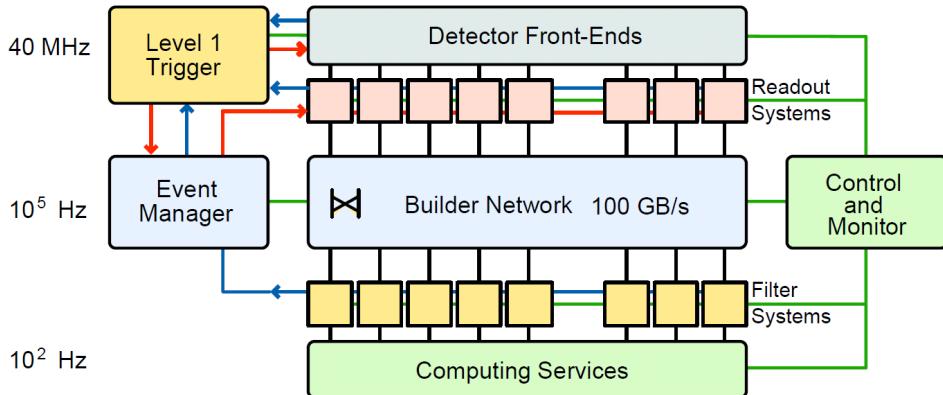
**Figure 2.10:** Diagram of the CMS muon systems. The location of each muon chamber for each subsystem is shown: Drift Tube (DT), Cathode Strip Chamber (CSC) and Resistive Plate Chamber (RPC) [48].

crossing time. This fast response allows, in conjunction with a dedicated trigger system, the correct bunch crossing associated with the detection of a muon to be selected. In the barrel, there are 480 rectangular chambers arranged in 4 stations with 6 RPC layers (2 layers are present in the 2 stations closest to the beam pipe). In the endcaps there are 3 RPC disk shaped stations on each side, which are composed by trapezoidally shaped detectors.

The combined muon system offline momentum resolution is of the order of 9% for small values of  $\eta$  and  $p$  and for transverse momenta of up to 200 GeV. At higher energies of around 1 TeV the standalone momentum resolution is in the range of 15-40% depending on  $|\eta|$ . These values are limited by the muon multiple-scattering before arriving to the muon system. If we combine the tracker information into a global fit the resolution for lower  $p_T$  tracks improves an order of magnitude while at higher momenta (around 1 TeV) it is about 5%, which is well inside the CMS design requirements.

## 2.2.7 Data Acquisition System

The CMS Data Acquisition (DAQ) system is designed to process, analyse and ultimately store the information collected by the detector [63]. The LHC produces bunch crossings at a rate of up to 40 MHz but we are only capable of storing between  $10^2 - 10^3$  events per second. At design luminosity, each bunch crossing will have an average of over 20 simultaneous collisions and produce a zero-suppressed data payload of around 1 MByte. A first level of trigger was developed in order to reduce the event rate to a maximum of 100 kHz and only the selected events are fully retrieved from the detector event buffers. Even with this event suppression the DAQ has to retrieve and move  $\approx 100$  GByte/s from the detector to the surface. This data comes from approximately 650 data sources and has to be merged into a single event package. The information is then passed to a computer farm where software filters serve as a second level of trigger. In this system the event rate is further reduced by a factor of up to 1000 making the output rate compatible with what can be saved into permanent storage. A diagram of this system can be found on figure 2.11.



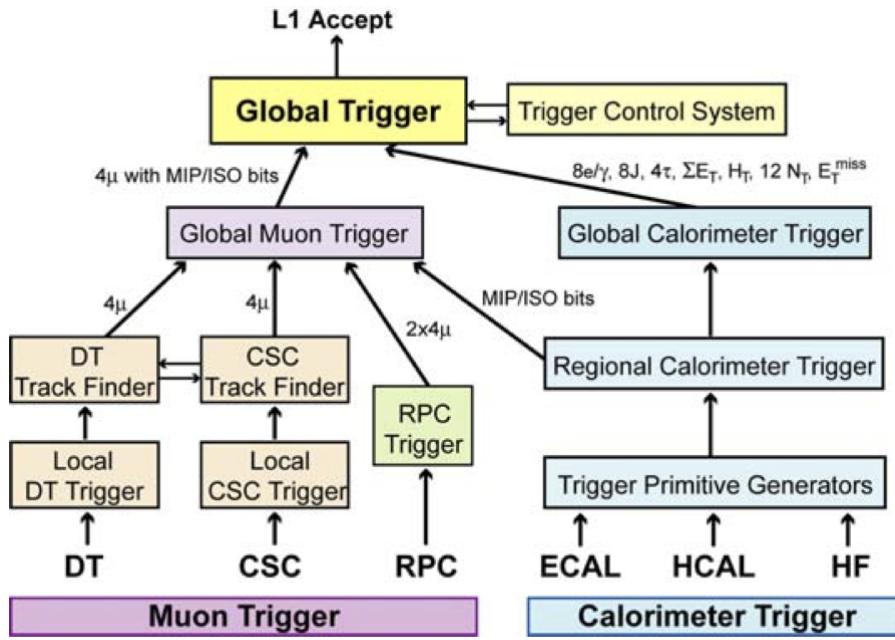
**Figure 2.11:** Diagram of the CMS DAQ system. Data flow is shown as the lines connecting each electronics or computing units [44].

## 2.2.8 Trigger System

The CMS trigger system is responsible for selecting which collisions are recorded in real-time. We can only save  $\mathcal{O}(10^2 - 10^3)$  events per second with the current systems. This implies that the trigger system needs to obtain a data reduction of a factor of  $\mathcal{O}(10^6 - 10^7)$ . This is achieved with a two level trigger system, the first is a dedicated

hardware system named Level 1 Trigger (L1T) [64] and the second is a commercial computer farm running dedicated software called the High Level Trigger (HLT) [63].

Initially, all data is stored for 128 bunch crossings which corresponds to  $3.2\ \mu\text{s}$ . This is the time we have to make a first decision to keep or discard an event. This is the task of the L1T which has the target to reduce the data to a maximum rate of 100 kHz. There isn't enough time to get all the information from the detector, so only a coarse version of the calorimetry and muon systems data, and some correlation between them is accessed. With this information the L1T produces a set of particle candidates and energy sums over which custom, user-defined algorithms can be used to filter events. A diagram of the L1T trigger components and the data flow across the system is present on figure 2.12.



**Figure 2.12:** Diagram of the L1T system. The arrows indicate data flow and the number of particle candidates at each step is indicated [44].

The HLT receives events accepted by the L1T and needs to perform further event reduction of  $\mathcal{O}(10^3 - 10^2)$  to a final output rate of  $\mathcal{O}(10^2 - 10^3)\text{Hz}$ . This system is composed of standard computing hardware in the form of a computing farm with  $\approx 15\text{k}$  Central Processing Units (CPUs). This system, using the additional latency created by the L1T event selection, is able to make use of the complete detector information including the tracker data. More sophisticated and precise algorithms are therefore possible which can be tailored to select any desired physical final state. An HLT path is

defined by the sequence of requirements for an event to be accepted, starting with the selection of the seeding L1T algorithms followed by HLT object requirements.

Event selection algorithms at both the L1T and HLT are frequently updated during data taking. The selection thresholds may be tuned in order to control the rate with the changes of LHC luminosity. Novel methods or strategies to identify particles more efficiently can be implemented, like PU subtraction or new calibrations. Analysis groups may also show interest in recording new event final states for which new selection criteria may be developed. The set of algorithms used for data taking is normally referred to as the *trigger menu*.

After events pass both levels of the trigger they are recorded into permanent storage. During 2012-13 operation, two output streams were saved. The *prompt data stream*, with a rate of approximately 300 Hz, was composed of high priority trigger paths which were immediately reconstructed. And the *parked data stream*, with an average rate of 600 Hz, was stored without reconstruction. These data waited until computing resources were free to go through reconstruction [65]. This process was finalised a few months after the LHC Run I was finished.

Even with such measures to reduce the data to be stored, each LHC experiment records several petabytes of data every year in addition to similarly sized amounts of simulated events.

### 2.2.9 Computing

The quantity of data produced by the LHC and the necessary processing capability is so big that it would be difficult to have all computing resources in a single place. For this reason a tiered system was developed, where all participating computing sites are connected and have specific roles and responsibilities in the data taking, processing and storing. This global computing system is known as the Grid [66].

The CERN Data Centre is the Tier 0 of this network, all data produced by the LHC experiments is handled by this facility. Only about 20% of the total capacity of the Grid is hosted here, but the CERN Tier 0 has the very important mission of safe keeping all the raw data produced by the experiments. It also has the task of doing the first pass of event reconstruction, which is the process identifying meaningful physics objects in data.

There are 7 CMS Tier 1 computer centres around the world. They are responsible to store a proportional amount of raw and reconstructed data for safe keeping. If any

reprocessing of the data is needed, these centres are responsible for this task and for storing the resulting output as well.

Local research centres like universities or scientific laboratories are normally at the Tier 2 level. These centres have the responsibility of handling a proportional share of simulated data production and reconstruction. Currently there are over 150 Tier 2 centres around the world. All physics analysis happen at this level.

Individual computers or local clusters without any formal engagement with the Grid structure, are considered to be the Tier 3 level of the Grid.

### 2.2.10 Level 1 Trigger: Stage I Upgrade

An extensive upgrade program for the L1T electronics was planned and is being executed in order to cope with the increase of luminosity and pile-up predicted for Run II [67, 68]. The center-of-mass energy has almost doubled from 8 TeV to 13 TeV, instantaneous luminosity will also increase, as will the average pile-up. Also, the bunch separation has changed from 50 ns to 25 ns making out-of-time pile-up a significant problem.

To ensure physics performance during 2015 and beyond, only a partial upgrade was executed for the 2015 run which is known as the *Stage-1* upgrade. The main feature of this upgrade program is the replacement of the existing Global Calorimeter Trigger (GCT). Two key enhancements were possible:

- Event-by-event pile-up energy subtraction for jet reconstruction,  $e/\gamma$  isolation,  $\tau$  isolation.
- Smaller feature size  $\tau$  candidates, which have significantly better energy estimation and background rejection.

The intermediate system will have significantly better performance than the now legacy system. The full 2016 calorimeter trigger system will additionally provide finer granularity which will lead to increased position and energy resolution.

# Chapter 3

## Event Reconstruction and simulation

This chapter describes how the CMS detector produces physics objects from the information collected at each event. The VBF Higgs to invisible analysis uses almost all the physics objects reconstructed by the detector, making use of information from all the experiment sub-detectors. The following sections describe in detail how each of these objects are reconstructed and what filters are placed upon them. The last section describes how Monte Carlo (MC) methods are used to simulate physics processes and emulate the detector response.

### 3.1 Tracks

Reconstructing the trajectories of charged particles allows us to measure their momentum and determining their charge. This is possible by analysing the hit patterns in the inner tracking system. In CMS this reconstruction is made with the Combinatorial Track Finder (CTF) algorithm [69]. The relevant steps for track generation are described below:

- Seed generation is made with hits at the pixel detector. A track seed can be made with two or three hits. In the first case a known vertex or the beam spot is used to constrain the seed momentum. The parameters of each seed are estimated using the assumption that the trajectory is a helix, but it takes into account hit errors and multiple scattering [70].
- The track seed is extrapolated through the tracker layers with a combinatorial Kalman filter [71]. For each additional layer, the best matching hit, if any, is added

and track parameters are recomputed. This procedure continues until the last layer is reached [69].

- Ambiguity resolution may be necessary since it is possible to have the same track being reconstructed from different seeds, or a seed may result in more than a single trajectory candidate. To resolve this possible double counting, when considering a pair of tracks with more than 50% of shared hits, we discard the one with the fewer hits. In case of equal number of hits, the track with the lowest  $\chi^2$  is kept.
- After the track building and cleaning stages are done, final refitting is performed. This procedure is aimed at removing possible bias by constraints at the seed forming stage. A standard Kalman filter and smoother are used.

The process of track finding is repeated up to six times, with the hits for each successfully reconstructed track removed for the next iteration. Using early LHC data and a dataset of pions and muons, it was possible to estimate that the tracking efficiency is  $> 98\%$  for all track  $p_T > 500 \text{ MeV}$  and  $> 99\%$  for tracks with  $p_T > 2 \text{ GeV}$  [72].

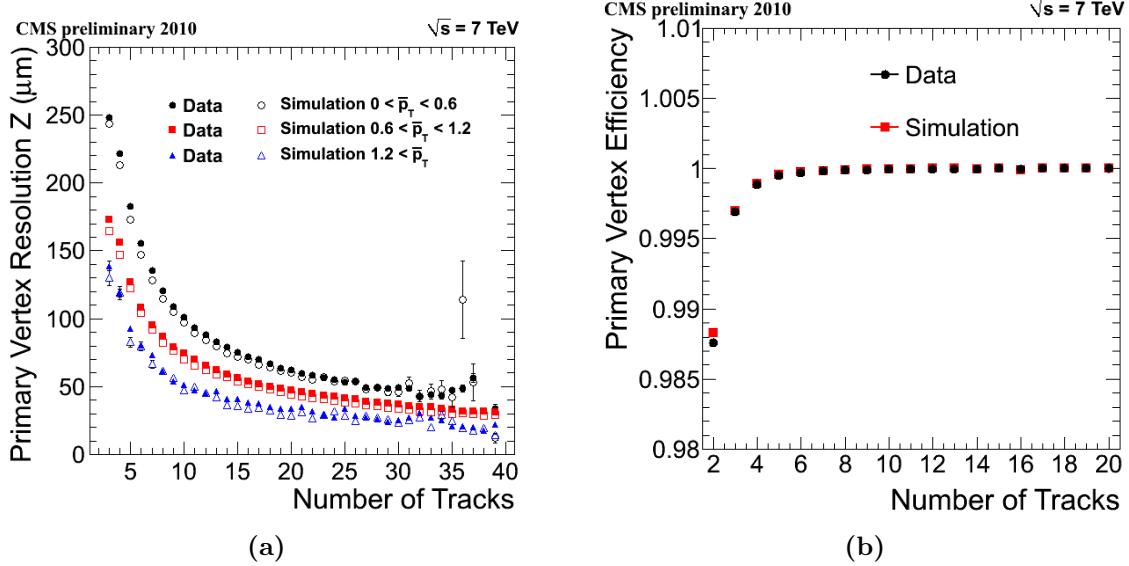
## 3.2 Vertex Reconstruction

The LHC can produce extreme collision intensities which are obtained partially by having multiple collisions happening at each bunch crossing. As discussed in section 2.1.1 an average of 21 simultaneous collisions happened per bunch crossing during 2012. In this environment, it is crucial to identify the Primary Vertex (PV) and the particles that come from it. This information can then be used to reject particles coming from other additional collisions and to identify displaced vertices which can be the signature of long lived particles like b-hadrons.

The individual tracks are reconstructed making use of the inner tracker. Each vertex is initially seeded by two tracks with separation in  $z$  less than 1 cm. Then remaining tracks are clustered to the vertex seeds with the Deterministic Annealing (DA) algorithm [73]. After the clustering process is done, the position of each vertex is recomputed using the Adaptive Vertex Fitter (AVF) algorithm [74]. In this algorithm, weights  $w_i$ , are assigned to tracks according to how compatible they are with the fitted vertex position. Weights vary from 1 to 0, tracks assigned weights of close 1 are highly compatible with the vertex and close 0 are given to low compatibility tracks. Then we can define the number of degrees of freedom of the new fit as:

$$n_{dof}(vertex) = 2 \sum_i^{tracks} w_i - 3 \quad (3.1)$$

This variable can be used to distinguish real proton-proton interactions from misclustered vertices, since it is correlated with the number of tracks compatible with that specific vertex [75]. The vertex position and resolution have been measured with LHC data and compared with simulation. The resulting plots can be found in figure 3.1 as a function of the number of tracks.



**Figure 3.1:** (a) Primary vertex resolution in the  $z$  coordinate as function of the number of associated tracks. Results are give for three ranges of average track  $p_T$ . (b) Primary vertex efficiency as a function of the number of associated track [75]

The PV is defined as the vertex with the highest sum of associated tracks  $p_T$  squared. In situations where no vertex can be reconstructed, for example if there is a tracking failure, the beam spot position is assumed. Knowing precisely the interaction point allows us to determine particle candidate quantities relative to it, which allow for better object identification and pile-up control.

Most CMS analyses, including the ones presented in this thesis, require explicitly that a good vertex is reconstructed with the following characteristics:

- Real reconstructed vertex from tracks, not the beam spot.
- A minimum number of degrees of freedom:  $n_{dof} > 4$ .
- Collision must be near the interaction region. We require longitudinal distance to be  $|z| \leq 24$  cm (longitudinal impact parameter).
- Collision must be close to the beam line. Radial distance to beam line:  $d_{xy} < 2$  cm (transverse impact parameter).

### 3.3 Particle Flow

The Particle Flow (PF) algorithm [76–78] is used in the CMS experiment with the objective of reconstructing every stable particle produced in the event. This is achieved by combining information from all CMS sub-detectors in order to identify electrons, photons, muons, charged hadrons and neutral hadrons and measure their direction, energy and type. The identified particles can in turn be used in jet clustering, determining the missing transverse energy, reconstructing and identifying taus, calculating particle isolation, identifying b-quark jets, etc.

The CMS experiment is very well suited to this approach as it is equipped with a high precision silicon tracker, which is immersed in a uniform axial magnetic field and has a dual calorimeter design with high hermeticity and resolution. The tracker system allows very precise direction/momenta reconstruction for charged particles, down to transverse momentum as low as 150 MeV. The high granularity of the ECAL allows for photons to be identified through deposit separation even inside high energy jets. In turn electrons can be reconstructed by combining their track and the energy deposits of the electron itself and its emissions, this algorithm will be explained further in section 3.4. The tracker information also allows the separation of charged and neutral hadrons in close proximity, a task which is not possible with just the HCAL due to its coarser granularity. We can determine the charged hadron momentum from the track information, and then, by removing its deposit from the calorimeter system we can determine the neutral hadron deposits. In areas outside the tracker and/or ECAL coverage, measurements are more coarse since we have less information available.

The clustering is performed separately in the ECAL and HCAL. We start by identifying *seed clusters* which are local maxima of calorimeter cell energy deposits. We add neighbouring cells into *topological clusters* if their energy deposit is bigger than two standard deviations of the electronics noise. This value was determined to be 80 MeV for the ECAL barrel, up to 300 MeV for the ECAL endcap and 800 MeV for the HCAL. The energy of each cell may be shared between multiple clusters.

Tracks and clusters are PF elements that need to be linked together to reconstruct the particle they came from and also to avoid double counting. We pair elements based on a metric of distance between elements and if they are compatible we merge them into *blocks* which can be interpreted as particle candidates. As an example, a pair of a track and energy cluster on the calorimeter system would be linked if you could extrapolate the track to the cluster volume.

### 3.3.1 Isolation

To reduce the probability of misidentification of a lepton coming from QCD jets as opposed to a lepton originating directly from the hard scatter process we can require isolation [79, 80]. We compute the isolation by summing the transverse momenta of all particles inside a cone around the selected lepton. In this sum, we include all charged particles, neutral hadrons and photons. But we do not want to include the PU contribution to this sum so we only include the charged candidates with a PV impact parameter smaller than 0.1 cm. Different methods are used for each particle to estimate and subtract the neutral component of the PU depending on Particle Object Group (POG) recommendations.

Normally, for physics analysis we defined the more meaningful *relative isolation* as  $I_{rel} = I/p_T^{lepton}$ . By using, a quantity that is relative to the candidate  $p_T$  and not an absolute cut we avoid wrongly accepting low energy candidates or rejecting high energy candidates. In the next sections, the steps taken to calculate this quantity for each particle candidate are explained.

## 3.4 Electrons

In the CMS experiment, electrons are reconstructed by matching energy clusters in the ECAL with tracks coming from the inner tracking system. Unfortunately, electrons can

loose and disperse significant amounts of energy before they reach the ECAL. While they go through the inner tracker they may emit photons via bremsstrahlung and in turn these photon can convert to  $e^+e^-$  pairs. About 35% of the electrons radiate at least 70% of their energy in this way [81]. This spread of energy is mostly in  $\phi$  due to the applied magnetic field [82]. Dedicated algorithms were developed to combine the the ECAL energy deposits, into a so called *supercluster*, of the initial electron and its emissions.

Different algorithms are used in the barrel and endcap regions. In the barrel region, we explore the simple  $\eta - \phi$  geometry with the *hybrid clustering algorithm* [79]. The procedure starts by identifying *seed crystals* with  $E_T > 1\text{ GeV}$ . A domino-shaped cluster is formed around this seed in the  $\eta$  direction of  $3 \times 1$  or  $5 \times 1$  crystals centred at the seed. Additional dominoes are added in both  $\phi$  directions in an attempt to collect the bremsstrahlung emissions up to  $\Delta\phi \approx 0.3\text{ rad}$ . Any domino with energy below 100 MeV is disregarded. The resulting additional sub-clusters must have their own seed with  $E_T > 350\text{ MeV}$  and they are all combined to form the final *supercluster*.

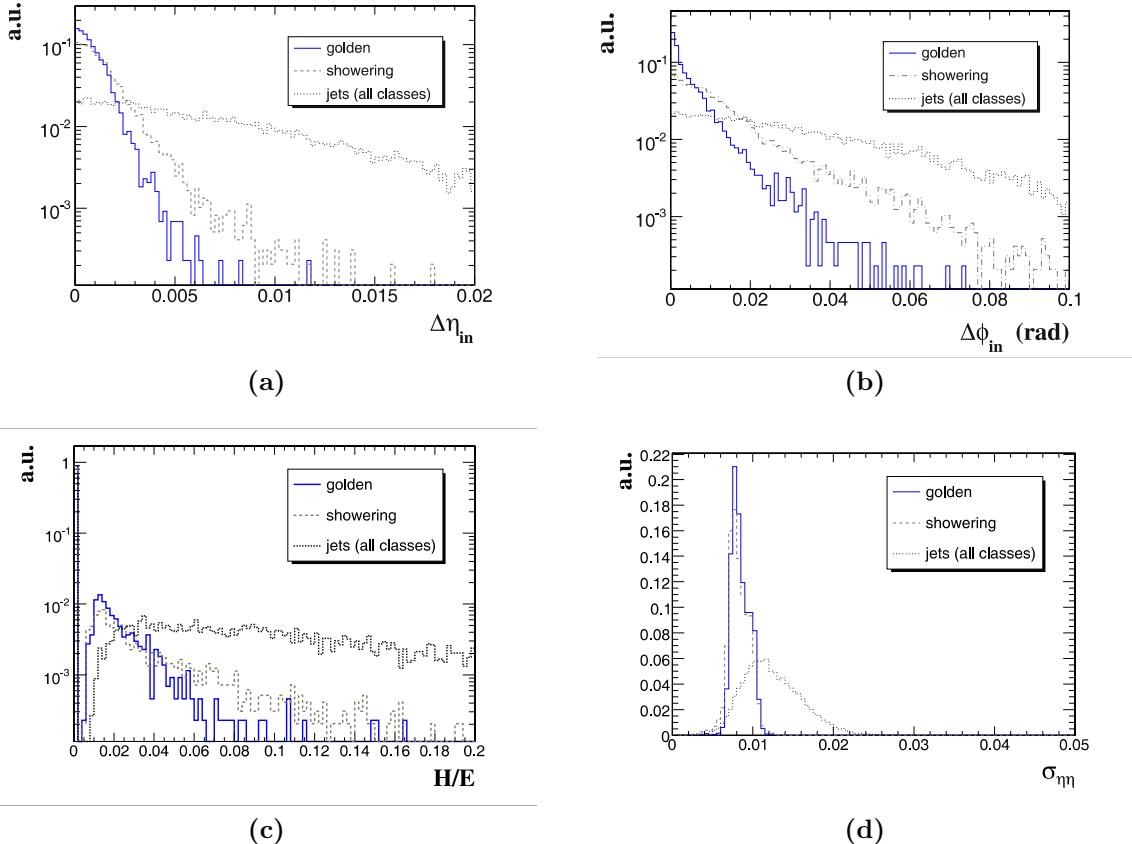
In the encaps, the *Multi- $5 \times 5$  algorithm* [79] is used. In this region of the detector the geometry is more complex and does not follow a simple  $\eta - \phi$  symmetry. The seeds for this clustering procedure are the crystals which are local maxima over their four direct neighbours and have a deposit of  $E_T > 0.18\text{ GeV}$ . Then, and starting with the seeds with highest  $E_T$ , we collect the energy around them into clusters of  $5 \times 5$  crystals. We then search for similar seeds and form clusters that can overlap within  $\Delta\eta < 0.07$  and  $\Delta\phi < 0.3\text{ rad}$  of the initial seed. Those clusters are then combined into a single *supercluster* which needs to have at least  $E_T > 1\text{ GeV}$ . The *supercluster* is then extrapolated to the ECAL preshower by clustering the energy within  $\Delta\eta < 0.15$  and  $\Delta\phi < 0.45$  around the most energetic cluster and adding it to the *supercluster* itself [79].

In order to reconstruct the electron track, we need to take into account the bremsstrahlung emissions. The CTF algorithm, which was described in section 3.1, is not appropriate for this purpose so a different track-finding algorithm had to be developed. For high  $p_T$  electrons, we use the ECAL supercluster energy deposit weighted mean impact point as a seed. If we combine this information with the determined  $E_T$  we can define two  $\eta - \phi$  search regions in the pixel detector depending on the charge hypothesis. If we find two compatible hits, the electron trajectory is updated. From this point normal track building is performed but instead of a Kalman filter algorithm we use a Gaussian Sum Filter (GSF) algorithm [83]. This method performs better in the presence of non-Gaussian losses like the one coming from the bremsstrahlung emissions.

The typical background to real electrons are collimated hadronic jets, with  $\pi^0$  and  $\pi^\pm$  overlap or from  $\pi^\pm$  showers [81]. There are many useful variables that may be used to reduce such background and which are often used in *electron identification* criteria:

- $\Delta\eta_{in}$  and  $\Delta\phi_{in}$ , are the distance between the track direction at the vertex and extrapolated to the ECAL and supercluster.
- $\sigma_{\eta\eta}$  is the energy-weighted  $\eta$  width of the cluster. For real prompt electrons this is normally small, since this quantity is not significantly affected by the magnetic field.
- $H/E$  is the ration of hadronic to electromagnetic energy in the region of the seed cluster.

Distributions of these variables for simulated electrons and jets can be found in figure 3.2.



**Figure 3.2:** Distributions for (a)  $\Delta\eta_{in}$ , (b)  $\Delta\phi_{in}$ , (c)  $H/E$  and (d)  $\sigma_{\eta\eta}$ . Here *golden electrons* are those who emit minimal bremmstrahlung photons, *showering* are electrons that lose a large fraction of their energy in emissions and *jets* are the typical distributions for hadronic jets [81].

### 3.4.1 Isolation

For electrons, we calculate isolation with the *effective area-corrected isolation* method over a cone of  $\Delta R < 0.3$  around the electron. For the neutral PU subtraction we use a look-up table of effective areas according to electron  $|\eta|$  which is multiplied by the estimated neutral PU energy density by unit of effective area. The definition for this isolation can be found in equation 3.2.

$$I = \sum_{\text{charged non-pileup}} p_T + \max \left( 0, \sum_{\text{neutral}} p_T + \sum_{\text{photon}} p_T - \rho(\text{lepton}) \times \text{Eff. Area}(\text{lepton}) \right) \quad (3.2)$$

### 3.4.2 Veto electrons

The minimum electron candidate requirements to veto an event if an electron is present is defined as the *veto electron* criteria. For this purpose we required an electron candidate with  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.4$  which passes the CMS Electron/Gamma POG [84] requirements of the cut-based electron Identification (ID) *veto electron* working point. A summary of these conditions can be found in table 3.1.

Variable	Barel	Endcap
$ \Delta\eta(\text{track}, \text{supercluster}) $	< 0.007	< 0.1
$ \Delta\phi(\text{track}, \text{supercluster}) $	< 0.8	< 0.7
$\sigma(i\eta, i\eta)$	< 0.01	< 0.03
$H/E$	< 0.15	-
$ d_0(\text{vertex}) $	< 0.04	< 0.04
$ d_Z(\text{vertex}) $	< 0.2	< 0.2
$\frac{PF_{\text{isolation}}}{p_\perp}$ for $\Delta R_{\text{cone}} = 0.3$	< 0.15	< 0.15

**Table 3.1:** Details of the CMS Electron-Gamma POG recommendations for a *veto electron*. Here barrel is defined as  $|\eta_{\text{supercluster}}| \leq 1.479$  and endcap is  $1.479 < |\eta_{\text{supercluster}}| < 2.5$ .

### 3.4.3 Tight electrons

We define a *tight electron* as an electron candidate with  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.4$  which passes the CMS Electron/Gamma POG requirements of the cut-based electron ID *tight electron* working point. This working point is similar to the 2011 very tight WP70 working point. A summary of these conditions can be found in table 3.2.

Variable $p_T > 20 (p_T \leq 20)$	Barel	Endcap
$ \Delta\eta(\text{track}, \text{supercluster}) $	$< 0.004$	$< 0.005$
$ \Delta\phi(\text{track}, \text{supercluster}) $	$< 0.3$	$< 0.2$
$\sigma(i\eta, i\eta)$	$< 0.01$	$< 0.03$
$H/E$	$< 0.12$	$< 0.10$
$ d_0(\text{vertex}) $	$< 0.02 \text{ cm}$	
$ d_Z(\text{vertex}) $	$< 0.1 \text{ cm}$	
$ \frac{1}{E} - \frac{1}{p} $	$< 0.05$	
$\frac{PF_{\text{isolation}}}{p_\perp}$ for $\Delta R_{\text{cone}} = 0.3$	$< 0.10$	$< 0.10(0.07)$
Conversion rejection: vertex fit probability	$< 1 \times 10^6$	
Conversion rejection: missing hits	$= 0$	

**Table 3.2:** Details of the CMS Electron-Gamma POG recommendations for a *tight electron*. Here barrel is defined as  $|\eta_{\text{supercluster}}| \leq 1.479$  and endcap is  $1.479 < |\eta_{\text{supercluster}}| < 2.5$ .

## 3.5 Muons

Muon track reconstruction starts independently at the inner-tracker (*tracker track*) and in the muon systems (*standalone muon track*) [80]. Then this information can be combined into a single muon track in two possible ways.

*Global muon reconstruction* is an *outside-in algorithm*. We start by finding a tracker track match for each standalone muon track. This is done by propagating the match candidate pair to a common surface and comparing track parameters. For each matched pair, a *global-muon fit* is performed using all hits from the two tracks in a Kalman-filter algorithm [71]. For muons of  $p_T \gtrsim 200 \text{ GeV}/c$ , it has been shown that a *global-muon fit* improves the momentum resolution compared to a *tracker-only fit* [48, 85].

*Tracker muon reconstruction* is an *inside-out algorithm*. In this method we start by selecting all tracker tracks with  $p_T > 0.5 \text{ GeV}$  and  $p > 2.5 \text{ GeV}$ . We extrapolate those tracks to the muon system, taking into account the magnetic field, energy loss and scattering. If we find a match with at least one muon segment in the muon system (track stub in the DT or CSC) this tracker track now becomes a Tracker Muon.

Tracker muon reconstruction is more efficient than the global muon reconstruction at low momenta ( $p \lesssim 5 \text{ GeV}$ ). This difference is due to tracker muon reconstruction only requiring one segment on the muon system. Global muon reconstruction is more efficient for higher energies, where the muons are more likely to pass several muon stations.

Muons can also be classified as prompt or non-prompt. Prompt muons are the ones produced directly in the hard process, like the decays of vector bosons or quarkonia particle decays. On the other hand, non-prompt muons typically come from in-flight decays of light hadrons, from taus or heavy quark decays.

The Global muon reconstruction has high purity reconstructing prompt muons, but sometimes hadronic activity can “punch-through” the calorimeter system and appear in the muon system, generating fakes. To reduce this type of background we can use different muon identification criteria.

Studies have been performed with the CMS detector to measure muon reconstruction efficiency [80]. Muon candidates were obtained using a global fit of both tracker and muon chamber hits with a  $\chi^2$  per degree of freedom of less than 10. This fit must include at least one segment in the muon chamber, track segments in at least 2 muon stations, use more than 10 hits in the inner tracker of which at least one in the a pixel layer and finally a small transverse impact parameter  $|d_{xy}| < 2 \text{ mm}$ . The efficiency for such a criteria for muon candidates with  $p_T > 10 \text{ GeV}$  has been measured both in data and Monte Carlo using  $J/\psi \rightarrow \mu^+ \mu^-$  and  $Z \rightarrow \mu^+ \mu^-$  and it plateaus at 96-99%.

### 3.5.1 Isolation

For muons we use the *combined isolation* over a cone of  $\Delta R < 0.4$  around the muon. For neutral PU subtraction we use the charged PU component inside the cone and multiply it by a factor of 0.5 which is determined from simulation. The definition for this isolation can be found in equation 3.3.

$$I = \sum_{\text{charged non-pileup}} p_T + \max \left( 0, \sum_{\text{neutral}} p_T + \sum_{\text{photon}} p_T - \frac{1}{2} \sum_{\text{charged pileup}} p_T \right) \quad (3.3)$$

### 3.5.2 Loose Muons

We can define a *loose muon* using the cut-based definitions recommend by the CMS Muon POG [86] with the same name, where we require the muon candidate to be a PF muon that is also a tracker or global muon. We exclude only standalone muons, which are only  $\approx 0.01\%$  of the PF muons. Additionally we require the muon candidate to have  $p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.1$  and relative combined isolation  $< 0.2$ .

### 3.5.3 Tight Muons

We can also define *tight muon* as a muon candidate with  $p_T > 20 \text{ GeV}$ ,  $|\eta| < 2.1$  passing relative combined isolation  $< 0.12$ . Additionally, we require compatibility of being produced at the primary vertex by requiring  $d_{xy} < 0.045 \text{ cm}$  and  $d_z < 0.2 \text{ cm}$ . We also require the muon to pass the CMS Muon POG recommended cut-based *tight muon* identification criteria that require the candidate to be a PF muon which is also a global muon, where the global track fit has at least one muon chamber hit and  $\chi^2/\text{ndof} < 10$ , the presence of muon segments in at least two chambers, at least five tracker layers with hits and at least one pixel hit.

## 3.6 Jets

When we collide hadrons the most probable hard processes will be the scattering quarks and gluons. However, these do not reach our detectors. They quickly hadronize and fragment generating a collimated spray of particles which is commonly referred to as a jet. To determine the properties of these outgoing quarks and gluons, we need to look at the characteristics of their associated jets. To achieve this goal, we need to combine the measured jet remnants in a way that preserves the physical properties of the original parton.

### 3.6.1 Jet Clustering

Jet clustering algorithms are sets of rules that allow us to combine particle candidates into jets [87]. These algorithms normally are controlled by parameters that define how close particles need to be in order to be associated into a jet and a way to combine their momentum. However, a jet definition should be robust and provide consistent measurements about the parton underlying the jet. There are two major families of problems that may affect jet algorithms. These problems appear when the number of jets in an event changes by adding a soft collinear gluon emission (collinear safety) or by parton splitting (infrared safety).

In CMS, we use a sequential recombination algorithm known as anti- $k_T$  [88], which is both infrared and collinear safe. This algorithm starts by determining a measurement of distance between every pair of objects  $d_{ij}$  and the distance of each object to the beamline  $d_{iB}$ . The definition of these distances can be found in equations 3.4 and 3.5 respectively

$$d_{ij} = \min(p_{\mathrm{T}}_i^{2p}, p_{\mathrm{T}}_j^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (3.4)$$

$$d_{iB} = p_{\mathrm{T}}_i^{2p} \quad (3.5)$$

where  $\Delta R$  is the separation in the  $\eta - \phi$  plane and  $R$  is the maximum radius for the jet. The parameter  $p$  determines the type of algorithm. When  $p$  is equal to 1 it is a  $k_T$  algorithm, 0 for the Cambridge/Aachen algorithm and -1 for the anti- $k_T$ .

After determining all the  $d_{ij}$  and  $d_{iB}$ , we determine which is the minimum distance. If it is a  $d_{ij}$  we combine those two objects and recalculate all the distances. If the minimum is a  $d_{iB}$  we declare  $i$  to be a *final state jet*, remove it from the list of particles and recalculate all distances again. The procedure continues until there are no more objects remaining.

The anti- $k_T$  algorithm tends to cluster particles around the hardest particle in a region, which normally leads to a cone-like jet area in the  $\eta - \phi$  plane. In the VBF Higgs to invisible analysis, the clustering is made over PF particle candidates using the implementation in the FASTJET software package [89]. The CMS recommended cone radius size for 2012-13 analysis is 0.5 while for 2015 it is 0.4.

### 3.6.2 Particle Flow Jet Identification

The CMS Jet-MET POG has defined criteria to reject fake, badly reconstructed, and noisy PF jets while keeping 98-99% real jets [90]. All the presented analyses in this thesis have used the recommended PF jet ID in the loose working point. In this working point all jets are required to have at least two constituents, and both the neutral hadron fraction and the neutral Electromagnetic (EM) fraction to be below 99%. Additionally, for jets inside the tracker acceptance with  $|\eta| < 2.4$ , we require the charged particle multiplicity and charged hadron fraction to be greater than zero, and the charged EM fraction to be less than 99%.

### 3.6.3 Pileup Jet Identification

To identify if a PF jet has come from PU or from the primary vertex we make use of a Boosted Decision Tree (BDT). This machine learning algorithm was trained with information about the trajectory of the tracks associated with the jet, the jet shape, and object multiplicity. In the presented analyses we have used the recommended loose working point of the *full BDT method* [91]. This method was applied to each jet, which would only be accepted if the BDT output score would pass the cuts defined in table 3.3 depending on jet  $p_T$  and  $\eta$ .

Jet $p_T$	Jet $ \eta $	$BDT_{score}$
$20 < p_T \leq 30$	$ \eta  < 2.5$	$> -0.80$
$20 < p_T \leq 30$	$2.50 \leq  \eta  < 2.75$	$> -0.85$
$20 < p_T \leq 30$	$2.75 \leq  \eta  < 3.00$	$> -0.84$
$20 < p_T \leq 30$	$3.00 \leq  \eta  < 5.00$	$> -0.85$
$30 < p_T$	$ \eta  < 2.5$	$> -0.80$
$30 < p_T$	$2.50 \leq  \eta  < 2.75$	$> -0.74$
$30 < p_T$	$2.75 \leq  \eta  < 3.00$	$> -0.68$
$30 < p_T$	$3.00 \leq  \eta  < 5.00$	$> -0.77$

**Table 3.3:** Table of the minimum values of *full BDT method* score for a PF jet to be accepted as coming from the PV using a loose working point. Required minimum values have been binned in jet  $p_T$  and  $\eta$ .

### 3.6.4 Lepton cleaning

To avoid having leptons being misidentified as jets we filter out all jets which are located at  $\Delta R < 0.5$  to any veto electron or loose muons.

### 3.6.5 Jet Energy Corrections

When reconstructing a jet, the clustered energy often does not match the original parton energy. There are many reasons for this effect, like non-linearity of the calorimeters response, detector noise, overlap with problematic detector areas, additional energy from PU, miscalibration, etc. To fix this problem corrections are determined and applied to each jet in order to have, on average, an energy measurement that is equal to the original parton. These corrections can be factorized into components as it is represented in equation 3.6 [92].

$$P_{\text{corr}}^{\mu} = C_{\text{offset}}(p_{\text{T}}^{\text{raw}}, \eta) \cdot C_{\text{rel}}(p_{\text{T}}^{\text{off}}, \eta) \cdot C_{\text{abs}}(p_{\text{T}}^{\text{rel}}, \eta) \cdot P_{\text{raw}}^{\mu} \quad (3.6)$$

The  $C_{\text{offset}}$  term accounts for and subtracts the contribution of PU and noise in the detector measurements. Its value is determined by taking into account the specific event  $p_{\text{T}}$ -density,  $\rho$ , and the individual jet area  $A$  [93]. The event  $\rho$  is calculated as the median  $p_{\text{T}}$ -density of all jets present in the event. Since the median is taken, it will not be affected by the presence of hard jets. Unfortunately, the Underlying Event (UE) activity has similar characteristics to the PU and should not be subtracted. To avoid this effect, the correction takes the form of  $\rho - \langle \rho \rangle_{\text{UE}} \cdot A$ , where  $\langle \rho \rangle_{\text{UE}}$  is the average expected UE contribution.

The  $C_{\text{rel}}$  term is applied to make the energy response flat as a function of  $\eta$ . It is applied to the offset-corrected transverse momentum  $p_{\text{T}}^{\text{off}}$ . To determine its value the  $p_{\text{T}}$ -balancing method is used [92]. In this method, we select a reference jet located in the central region where energy measurement is expected to be flat and a probe jet at any value of  $\eta$ . We can calculate the average of balance quantity as  $(p_{\text{T}}^{\text{probe}} - p_{\text{T}}^{\text{reference}})/p_{\text{T}}^{\text{average}}$  which is used to determine the correction to response in bins of jet  $\eta$  and dijet average  $p_{\text{T}}$ .

The  $C_{\text{abs}}$  term is intended to make the response uniform in  $p_{\text{T}}$ . It is applied to the  $\eta$ -corrected transverse momentum  $p_{\text{T}}^{\text{rel}}$  and is calculated using the Missing Transverse

Energy Projection Fraction (MPF) method [94]. In this method, we use the good experimental resolution for leptons and photons in processes like  $\gamma + \text{jets}$  and  $Z + \text{jets}$  to infer on the properties of the recoil jets. Since these processes should not have Missing Transverse Energy (MET), if MET is observed, it can be used to calibrate the jet response for the jets present in the event.

The total uncertainty on the jet energy scale is obtained by summing in quadrature the estimated uncertainties of each one of the correction terms. The total uncertainty is in the range of  $\approx 3 - 5\%$  depending on  $p_T$  and  $\eta$  [92].

### 3.7 Hadronic Taus

Taus can decay leptonically and hadronically. In leptonic decays the tau decays directly to an electron or a muon and two additional neutrinos. Therefore it is very difficult to identify such decays experimentally. On the other hand, a hadronic tau decay produces a characteristic signature of a narrow jet containing an odd number of charged particles and additional neutral hadrons, as well as a tau neutrino. In all the analyses presented in this thesis, when referring to a tau, we refer an hadronically decaying tau. The most probable decay modes have one or three charged  $\pi$  mesons and are summarized in table 3.4.

Decay Channel	Resonance	Mass [MeV]	Branching Fraction [%]
$\tau^\pm \rightarrow \pi^\pm \nu_\tau$			11.6
$\tau^\pm \rightarrow \pi^\pm \pi^0 \nu_\tau$	$\rho$	770	26.0
$\tau^\pm \rightarrow \pi^\pm \pi^0 \pi^0 \nu_\tau$	$a_1$	1260	10.8
$\tau^\pm \rightarrow \pi^\pm \pi^\mp \pi^\pm \nu_\tau$	$a_1$	1260	9.8
$\tau^\pm \rightarrow \pi^\pm \pi^\mp \pi^\pm \pi^0 \nu_\tau$			4.8
Other hadronic modes			1.7
Total			64.7

**Table 3.4:** Summary of the hadronic tau decay modes, with the branching fractions and intermediate resonances listed where relevant [95].

Reconstruction of hadronic tau neutrinos with PF is done by identifying the specific decay mode visible products. The approach is at the core of the Hadron Plus Strips (HPS) algorithm [96, 97]. It combines reconstructed charged hadrons with strips of clustered

photons which are interpreted as  $\pi_0$ . The reconstructed system is constrained by the tau mass and intermediate resonances and is a highly collimated jet when compared with a typical QCD jet.

### 3.7.1 Isolation and Discriminants

Isolation for taus is calculated in a similar way to electrons and muons. The isolation variable is defined by summing the  $p_T$  of all PF hadron and photon candidates in a cone of  $\Delta R < 0.5$  around the tau axis. Here the charged hadron tracks are required to have  $d_z < 2\text{ cm}$  to the tau production vertex. We can subtract the contribution to isolation coming from PU estimating its density in a cone of  $\Delta R < 0.8$  around the tau and considering tracks with  $d_z > 2\text{ cm}$ . All tau constituents are ignored in this sum. Working points have been defined for loose, medium and tight isolation [97].

Electrons can be reconstructed as taus when they make isolated deposits in the calorimeter or emit enough energy via bremsstrahlung to form a strip. A BDT has been trained with a set of variables similar to the ones used in electron identification to exclude such misreconstructions. Similarly to the electron and muon isolation three working points have been defined [96, 97].

Muons are less likely to be reconstructed as a tau. We can exclude such tau candidates by requiring that the track of the leading charged hadron is not also a tracker muon. This discriminator also has three possible working points [96, 97].

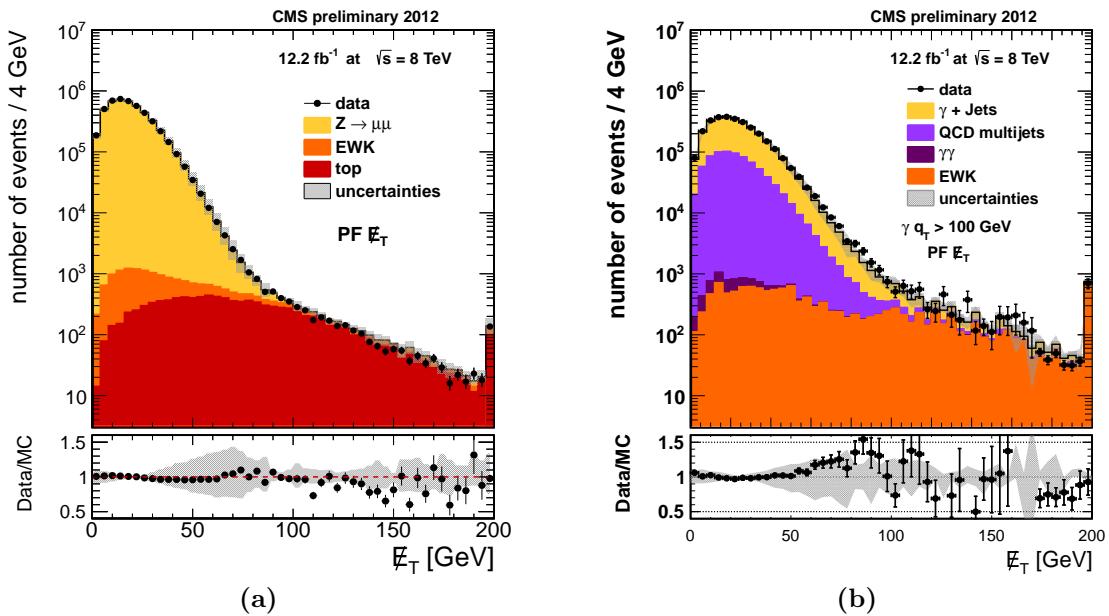
We can now define a hadronic tau candidate as a PF tau candidate with  $p_T > 20\text{ GeV}$ ,  $|\eta| < 2.3$  and  $d_z < 0.2\text{ cm}$  to the primary vertex. We require that the candidate passes decay-mode identification, tight isolation and finally tight discriminators against electrons and muons.

## 3.8 Missing Transverse Energy

The Standard Model describes neutrinos as particles which only interact via the weak force. They can pass through our detectors without interacting and therefore not allowing any direct measurement. Many new models describe additional particles that would also be able to escape detection by leaving very small or no energy deposits in our experiments. The appearance of such particles can only be inferred through the measurement of an

imbalance of transverse momentum of all detected particles. This effect can be quantified as the negative sum off all visible particle candidates transverse momentum in an event.

The magnitude of that vector is referred to as Missing Transverse Energy (MET). Particle flow methodology provides a complete list of object candidates in the event with excellent resolution achieved by combining all available information, making it well suited to be the input for MET calculation. Although CMS has an excellent individual particle resolution, the calculation of MET is affected by the combined resolution of the measurement of all particles in the event. Figure 3.3 shows the distributions of PF MET for both data and simulation for event selections of  $Z \rightarrow \mu\mu$  and  $\gamma + jets$  processes at  $\sqrt{s} = 8$  TeV.



**Figure 3.3:** Distributions of the particle flow  $E_T^{\text{miss}}$  in (a)  $Z \rightarrow \mu\mu$  and (b)  $\gamma + \text{jets}$  events in  $\sqrt{s} = 8$  TeV data and simulation. The uncertainty in the muon, photon, jet and neutral hadron energy responses is showed by the shaded band [98].

Both photon and muon energy measurements have good resolution in the CMS experiment and these processes do not involve real MET. The observed distributions in both plots are predominantly shaped by the resolution of jet energy measurement.

During data taking, issues with the detector or data acquisition can happen creating anomalously high MET and rendering such events unusable. The groups responsible for each part of the detector and individual physics objects, check the data after collection to detect such problems. Event filters are produced to reject such problematic events when performing physics analyses. The CMS JET-MET POG compiled a list of the recommended filters for analysis using 2012-13 data to remove events affected by energy

deposits from beam halo, noise in HCAL readout electronics, particles directly hitting the ECAL photodiodes, track reconstruction problems and finally ECAL and HCAL mistimed laser calibration sequence. These filters have been used in both prompt and parked VBF Higgs to invisible analyses.

There are many factors that affect MET response and resolution. These include zero suppression thresholds, which dictate the minimum energy a calorimeter cell will report, dead or non-instrumented regions of the detector and reconstruction inefficiencies. Techniques have been developed to correct both response and resolution when using PF MET [99]. These corrections include accounting for the bias in response due to using incorrect energy scale of the jets, and reducing the impact of pileup on the resolution [98].

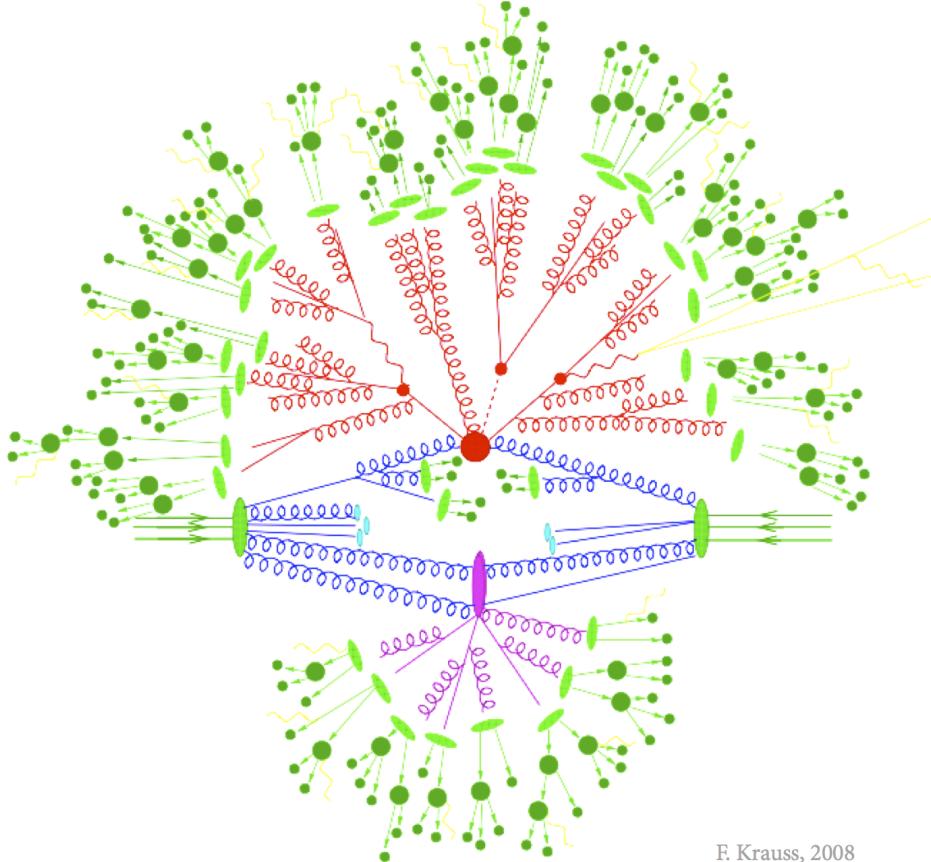
In the VBF Higgs to invisible analysis, the irreducible background of  $Z \rightarrow \nu\nu$  is investigated using  $Z \rightarrow \mu\mu$  as a proxy. To record this process with the same HLT trigger that is used for signal, MET is calculated without including muons,  $MET_{no-\mu}$ . This quantity is also used in the offline analysis, where muons are vetoed in the signal region to recover the usual MET value, and are required in some control regions to estimate backgrounds.

### 3.9 Monte Carlo Simulation

To simulate one event in the CMS experiment, we first start by the physics process itself. It can be split into two sub-processes: hard scattering and hadronization. There are many purpose-built software programs that will perform each of these steps. An illustration of how the simulation of proton-proton collision is done with MC programs can be found in figure 3.4. A review of the available generators for LHC physics can be found in reference [100].

General purpose particle physics event generators like PYTHIA 8 [102, 103], HERWIG++ [104] and SHERPA [105] are able to do both hard scattering and hadronization steps for a wide variety of physics processes. Typically these programs are restricted to  $2 \rightarrow 2$  and  $2 \rightarrow 1$  hard processes calculated at Leading Order (LO).

There are many other dedicated matrix-element generators, like MADGRAPH 5 [106], ALPGEN [107] and also SHERPA that focus on the hard process simulation. These programs provide  $2 \rightarrow X$  hard scattering, where a higher number of partons in the final state is possible. Some generators have also implemented Next to Leading Order



**Figure 3.4:** Illustration of a proton-proton collision as implemented in some MC event generators [101]. Sub-processes are represented, the hard-scattering in the centre of the diagram, the parton showering in red, hadronization in green. We can also observe the UE interaction and its showering in purple.

(NLO) calculations, which provide better kinematics description and lower uncertainties. Two examples of such generators are aMC@NLO [108] and POWHEG [109–116]. Both these generators are extensively used by the CMS collaboration, and the latter was used to simulate the signal sample in the analysis presented in this thesis. The simulated parton level events then need to be passed to a general purpose event generator for hadronization.

Overlapping of the phase-space description of matrix-element and showering programs needs to be avoided when simulating multi-jet events. This problem rises from software like PYTHIA or HERWIG describing parton radiation as a Markov Chain process based on Sudakov form factors. This approach is only formally correct in the limit of soft and collinear emissions. On the other hand Matrix Element (ME) programs like

MADGRAPH work well in the hard scattering high energy limit but diverge when the partons become soft or collinear.

There are a few jet-parton matching schemes developed to account for this overlap [117]. Showering can be vetoed and the event reweighed accordingly, like in the Catani-Krauss-Kuhn-Webber (CKKW) scheme [118–120] or events can be rejected altogether like in the MLM scheme [121]. Depending on the generator used for the showering, different schemes are implemented and care must be taken in the definition of the matching parameters.

Most event generators can be finely tuned so important aspects of the simulation can be adjusted to experimental conditions. As an example, in the CMS experiment, PYTHIA is used with the Z2 tune, which was produced using measurements made using minimum bias data at the Tevatron and at the LHC [122].

After the physics event is simulated, the interaction with the detector and the corresponding electronics response is estimated using a precise model of the experiment. In the CMS experiment GEANT 4 [123, 124] software is used for this task which also relies heavily on MC methods.

# Chapter 4

## L1 Trigger Data Quality Monitoring

The CMS L1T makes the first event selection of any physics analysis. Therefore, its correct operation is crucially for the physics output of the experiment. The Data Quality Monitoring (DQM) is one of the CMS monitoring systems which is able to provide both real-time monitoring and offline analysis of the detector operation. This chapter describes the L1T DQM applications developed and used for online monitoring and data certification for physics analyses of the during 2012-13 data taking.

### 4.1 Data Quality Monitoring

The DQM is a critical monitoring system that has an important role in detector and operations efficiency. It is also important in the certification of recorded data for physics analysis [64, 125]. The DQM system is an end-to-end solution that provides tools to create, fill, display and archive histograms and scalar monitors. It provides the ability to monitor the detector and DAQ in real-time, analyse the reconstruction process, validate the experiment’s software releases and its simulated data. The purpose of this system is to identify problems or errors in both hardware and software as early and accurately as possible.

#### 4.1.1 Online Monitoring

The online DQM system is composed of several applications that are part of the CMS data processing work flow. The software is executed on the CMS computing cluster at point 5. Applications fall into two categories: *high level trigger modules* and *data quality*

*monitoring modules.* The *high level trigger modules* are run directly on the HLT filter farm and can only produce a limited number of histograms which monitor that system or specific HLT path. The *data quality monitoring modules* run over events coming from a dedicated DQM event stream with a rate of 5 – 10 Hz. These events contain only the raw detector and trigger information. Each subsystem has its own application which can analyse all events from the stream or filter a subset with a predefined trigger selection. At the end of every luminosity section, which corresponds to 23.31 s, histograms are gathered from the nodes where the applications are run and are merged together. The results are shown in real time in a web based application which is accessible by the shift crew and on-call experts.

#### 4.1.2 Offline Monitoring

The offline DQM is used in numerous workflows including monitoring of the event reconstruction process, alignment and calibration validation, CMS software release validation, etc. For all these tasks, a standardized two-step process is run.

In the first step, histograms are produced in the same computing jobs as the task to be monitored and are stored along with the rest of the event data. This happens in multiple simultaneous jobs which, depending on the task, can be at the Tier 0 or Tier 1.

In the second, *harvesting step*, the histograms are extracted from the event data and summed together. The resulting histograms contain the full event yields from each run for each processed dataset. Applications running at this step have access to the detector conditions from the Detector Control System (DCS) and the DAQ, and can produce new histograms like summaries of the relevant quantities for each run.

### 4.2 Level 1 Trigger Data Quality Monitoring

The L1T DQM is composed of four applications. The first two applications run as part of the online DQM system and monitor the trigger hardware and emulation in real-time. The second pair of applications runs in the offline DQM system as part of the (re-)reconstruction workflow and provide information for physics data certification.

The first of the two online applications directly monitors the operation of the trigger. Each trigger subsystem produces plots of its own relevant quantities including information

which allows the origin of problems to be pin pointed. Additionally, a set of monitoring tools observe the final objects and global behaviour of the system. Key aspects are analysed such as the value of reference algorithm rates, synchronization of trigger, finding regions of the detector that show unexpected high/low rate.

The second online application compares the results of the trigger against a real-time software emulation of the system which should allow the quick detection of trigger misconfiguration or degradation of quality of operation.

Both offline monitoring applications replicate the analysis preformed by their online counterparts but over the complete recorded dataset for each run.

In the next sections we will focus on the trigger monitoring tools that the author developed or significantly improved.

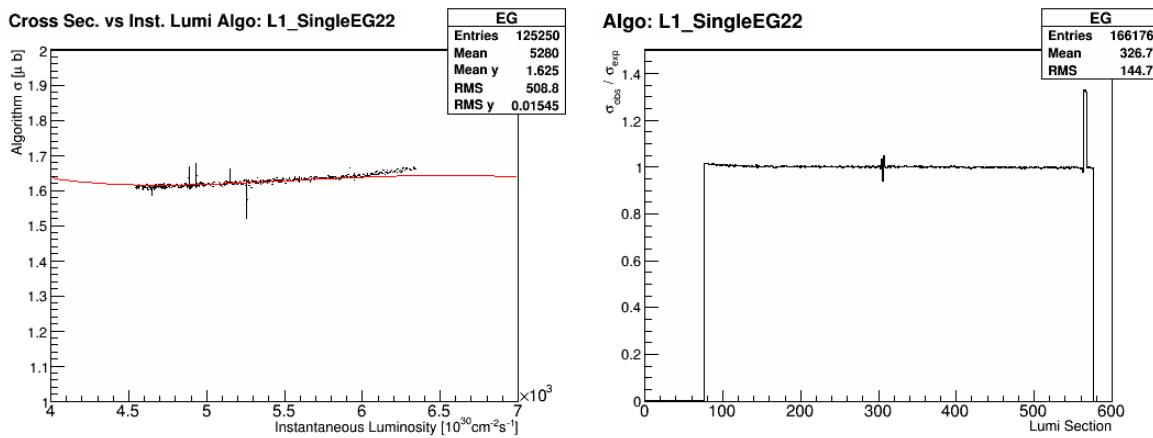
### 4.2.1 Rates Monitoring

The rates monitoring tool inspects the firing rate of each L1T object category. At the beginning of each run, the L1T menu is analysed and, for each object category, the lowest threshold unprescaled algorithm is selected. If no unprescaled algorithm is available the lowest prescale and threshold trigger is selected. If the selected trigger algorithms are  $\eta$  restricted a warning is shown in the produced histograms to identify that the tests performed do not cover the full acceptance of the monitored object. The following categories of objects can be monitored: Electron-Gamma, Isolated Electron-Gamma, Central Jets ( $|\eta| < 3$ ), Forward Jets ( $3 < |\eta| < 5$ ), All Jets ( $|\eta| < 5$ ), Taus, Muon, total energy (ETT), total energy in jets (HTT), missing transverse energy (ETM) and jets missing transverse energy (HTM).

When the algorithms to be monitored are determined, the tool retrieves the expected algorithm cross sections as a function of the instantaneous luminosity from an external database. These functions are updated daily by fitting runs from the previous days with similar conditions. This task is executed by the Web base Monitoring (WbM), which is a CMS monitoring system that runs in parallel to the DQM. The algorithm cross section for each luminosity section is calculated with equation 4.1

$$\sigma_{\text{Algo}} = \frac{\text{Prescale}_{\text{Algo}} * \overline{\text{Rate}_{\text{Algo}}}}{\overline{\text{Inst. Lumi.}} * (1 - \text{CMS dead time fraction})} \quad (4.1)$$

where  $\sigma_{\text{Algo}}$  is the observed algorithm cross section,  $\text{Prescale}_{\text{Algo}}$  is the prescale applied to the algorithm,  $\overline{\text{Rate}_{\text{Algo}}}$  is the algorithm average rate,  $\overline{\text{Inst. Lumi.}}$  is the average instantaneous luminosity delivered to CMS and the *CMS dead time fraction* is the observed fraction of time CMS was not recording data. The measured value is compared with predictions from previous runs for each luminosity section. The monitor presents these results in histograms with the measured value and the relative value to prediction. Example of these histograms can be found in figure 4.1



**Figure 4.1:** Monitoring plots produced by the L1T rates online monitoring tool for run 207269 and the electron/gamma object category. The automatically selected algorithm was L1\_SingleEG22 for this run. On the left histogram the algorithm cross section as a function of instantaneous luminosity is plotted. The red line is the prediction obtained from fitting data from previous runs while the black points are the measurements for this run. On the right histogram the fraction of the measured value over the prediction is showed as a function of the luminosity section.

Automatic tests are configured to monitor the produced histograms and flag as bad the luminosity sections that show deviation from prediction above 20%. Marking a specific luminosity section as bad does not invalidate its use for physics analysis, but references it for further investigation by the CMS shift crew or certification experts.

### 4.2.2 Synchronization Monitoring

The synchronization monitoring tool tests if each L1T object category is being produced and associated with the correct bunch crossing. Similarly to the L1T rates monitoring tool described in the previous section, at the beginning of each run we select the lowest threshold unprescaled algorithm for each object category. If none are available the algorithm with lowest prescale and lowest threshold is selected.

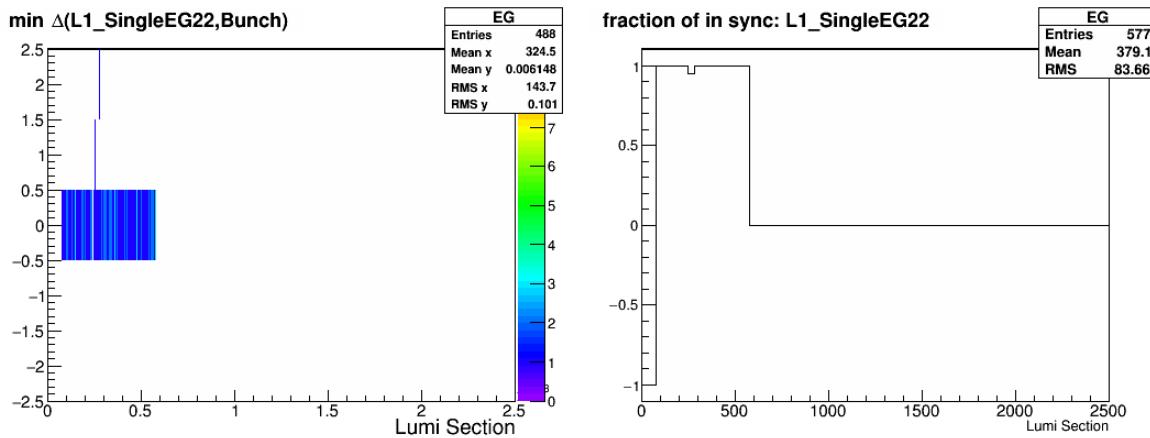
The information of which bunch crossings are filled is retrieved at the beginning of each run by CMS. That information is stored in a database at point 5 and later replicated to the CMS offline conditions database. At the same time, the synchronization monitoring tool determines the LHC fill number from the L1T Global Trigger (GT) system. Data from the LHC is obtained via the Data Interchange Protocol (DIP) which allows exchange of information between detector and accelerator. With this information the bunch crossing information is retrieved from the Online Master Database System (OMDS) when running online and from Offline Reconstruction Condition DB ONline (ORCON) when running offline.

When selected events are desynchronized from the correct bunch crossing at the L1T level, these events will appear empty to the HLT and offline. Therefore it is unlikely that they will pass any HLT triggers, making it very difficult to spot this type of problems. For this reason, the synchronization monitoring looks only at events that come from a special HLT trigger, the HLT pass-through paths. These triggers are highly prescaled and only required that a specific L1T condition is fired. All available HLT pass-through paths of single object L1T trigger are monitored by this tool.

All events triggered passing HLT pass-through paths are analysed and all selected algorithms firing are compared to the actual LHC bunch crossing filling and the results are recorded. Additionally, for each event we query the GT about the LHC beam mode and if for any event the status is not Stable Beams, the luminosity section is immediately marked as bad.

As we ran this monitoring only over the events that pass HLT pass-through paths, a single luminosity section will typically not have enough statistics to take conclusions on the behaviour of the system. To provide reliable results at the end of each luminosity section a decision is made as to whether the current luminosity section has enough statistics by itself or needs to be grouped with the previous ones. Blocks of luminosity section are made until a minimum (configurable) number of events is reached for each individual monitored trigger. At this point the histogram of the fraction of events in time with bunch crossings is updated. If the LHC beam mode changes or the run ends, the current open luminosity sections block is closed with the current statistics. The histograms produced by this tool for run 207269 can be found in figure 4.2.

Similarly to the L1T rates monitoring tool, automatic tests are configured to flag as bad luminosity sections that show deviation from prediction above 20%.



**Figure 4.2:** Monitoring plots produced by the L1T synchronization online monitoring tool for L1 single electron/gamma object category, which is automatically monitoring algorithm L1\_SingleEG22 for the run 207269. The left plot shows the minimum difference between the triggered bunch crossing and a LHC filled bunch crossing. The right plot shows the fraction of events that triggered in a LHC filled bunch crossing for each luminosity section block.

### 4.2.3 BPTX Monitoring

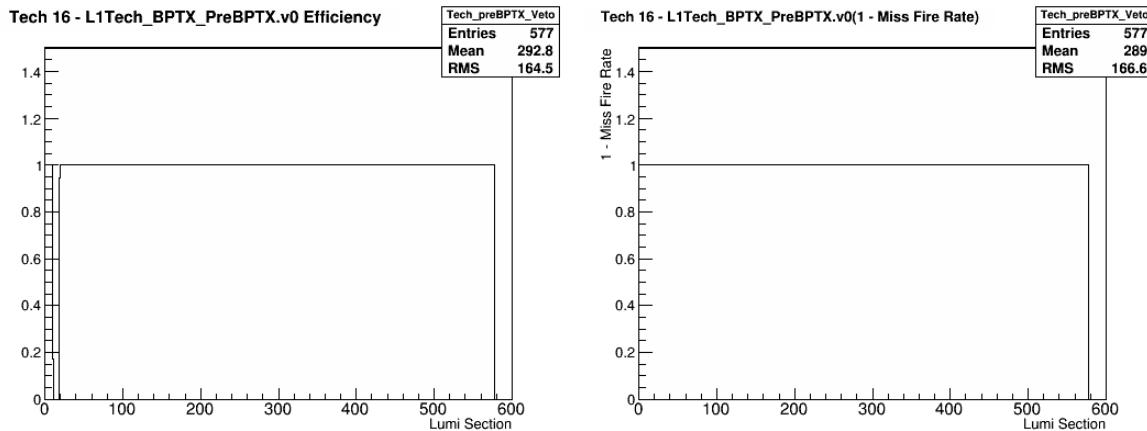
The Beam Pickup for Timing for the eXperiments (BPTX) system is composed of two beam detectors located in each beam pipe 175 m upstream of the CMS experiment [44]. These detectors were designed to provide precise information about the bunch structure, timing of each beam and to have sensitivity to time structures under 25 ns.

In early 2012, a problem was identified in the L1T, where some events would fire on the bunch crossing before the actual event. It was discovered that this effect was most likely connected to sensors in the calorimeter system being directly hit by particles causing a large out-of-time signal. Unfortunately, the trigger has a set of rules intended to limit the event rate. These are necessary in order to allow for sufficient latency to extract the information from the detector when a collision is accepted. One of these rules states that if a collision is accepted by the L1T the next 2 collisions are ignored by the system [64]. As a consequence, if the L1T pre-fires due to out-of-time signals subsequent real beam crossings will be vetoed. The solution found was to veto L1T triggering on the previous bunch crossing to any filled LHC bunch. This was achieved by defining a new veto technical trigger bit corresponding to the signal of both BPTX detectors logical AND advanced by one bunch crossing.

Although this was a successful solution to this problem it caused concern in the Trigger Studies Group (TSG) and L1T Detector Performance Group (DPG). If the BPTX bunch

detection threshold was set too high this veto would be ineffective, leading to no bunches being detected and no veto being applied. If the BPTX bunch detection threshold was set too low, residual amounts of protons or noise in the unfilled bunch spaces could lead to vetoing filled bunch spaces. The development and commissioning of a monitoring tool was requested as priority task.

A new tool was developed to compare the LHC filling scheme with the firing of the technical trigger associated with the L1T veto. Following the ideas of the L1T synchronization monitoring tool, the same procedure was used to retrieve the LHC filling scheme and algorithm firing results. For each selected event, the GT records the results of each L1T trigger algorithm for the two previous and the two subsequent bunch crossings. For this tool, all five recorded bunch crossings are compared with the LHC bunch structure. In this case we are interested in both efficiency, as low efficiency would mean that the BPTX bunch detection threshold was too low, and misfire rate, which would mean that the BPTX bunch detection threshold was being too high. Examples of the histograms produced by this tool can be found in figure 4.3.

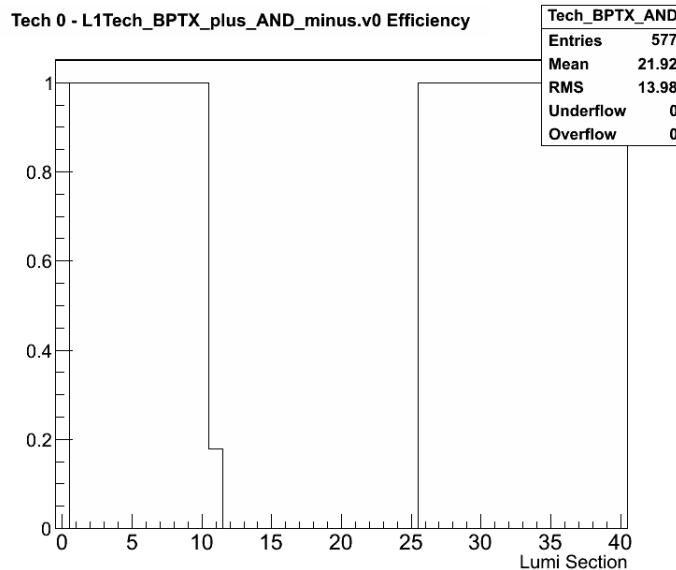


**Figure 4.3:** Monitoring plots produced by the L1T BPTX online monitoring tool for CMS run 207269. On the left the BPTX veto efficiency in relation to the LHC fill bunch structure is showed. On the right for the same algorithm  $1 - \text{Miss Fire}$  fraction is showed.

## Implementation Tests

To test that the BPTX monitoring tool would be successful in detecting the possible failure of the BPTX system, a field test was necessary. During run 207269 in which real data was recorded the author with the permission of the L1T DPG disabled the veto technical L1T bit associated with the BPTX logical and advanced by one luminosity

section. The bit was kept disabled for a few luminosity sections which was promptly identified by the monitoring tool as it can be seen in figure 4.4.



**Figure 4.4:** Detail of L1T BPTX online monitoring tool histogram for veto efficiency during its field test at run 207269. The monitored bit was disabled manually leading to the monitored drop off efficiency.

After this successful test, the trigger shifter instructions were updated to include this histogram in the periodic checks to be done.

#### 4.2.4 Occupancy Monitoring

The occupancy monitoring tool identifies regions of the detector where the trigger system response has been degraded. A region is considered *dead* if the number measurements is null or its rate is consistently smaller than what would be expected for that area. Alternatively, a region can become *hot* if the measurement rate is consistently bigger than expected for that area. This tool identifies both of these categories of problem by analysing histograms produced by the trigger subsystems.

This tool uses the  $\eta$  and  $\phi$  symmetries of both physics processes and the experimental design. The collisions in CMS happen in the centre of the experiment with beams of the same energy colliding head-on. Additionally, the detector is symmetric around the beam lines transverse plane passing on the collision point and also to the beam line itself. Both these factors imply that the response in a strip of cells across  $\phi$  at constant  $\eta$  should

be the same, on average, in every cell and that the response should be equivalent in a similar strip and constant  $-\eta$ .

The test consists of initially selecting a histogram of a quantity that is expressed in absolute event counts per region and that exhibits the described  $\eta$  and  $\phi$  symmetries. The histogram is integrated for as many luminosity sections as necessary to have enough statistics for conclusive results. When enough statistics are gathered and starting from the centre, a strip of cells is defined along  $\phi$  to one side of that symmetry line. The value of the median of the selected cells is determined. Each cell of the opposing strip is compared to this median with statistical tests tuned to detect significant deviations. If any tests are failed, the cell is marked as bad for the period of the histogram integration. The same procedure is repeated reversing the role of both strips. After all cells in the first strip pair is tested we move to the next two strips of cells in increasing  $\eta$  and repeat the procedure until all cells in the histogram have been tested. For histograms where the symmetry line falls in the middle of a strip of cell, that strip is tested against itself. The median is used to avoid bias from outliers like the *hot* or *dead* cells we are aiming to identify.

Cells which are already known to be problematic can be masked from this tool to avoid being always marked as bad and contributing to the calculation of the fraction of problematic cells.

### Statistical hypotheses test

Since we are analysing histograms of absolute number of entries, such as the location on L1T Electron-Gamma candidates, each cell will follow Poisson statistics [126]. The probability of obtaining a histogram cell with value  $x$  when the expected value is  $\mu$  is described by equation 4.2.

$$P(x; \mu) = \frac{\exp(-\mu) \cdot \mu^x}{x!} \quad (4.2)$$

The implemented statistical tests will evaluate each cell over two hypotheses. The null hypothesis  $H_0$ , considers that the cell is behaving as expected and that the average number of events is  $\mu_0$ . The alternative hypothesis  $H_1$ , proposes that this is a problematic cell with an average number of events of  $\mu_1$ . We can now define a test statistic  $T$  as the log-likelihood ratio of the two hypotheses as defined in equation 4.3.

$$T = \ln \frac{P(x, \mu_1)}{P(x, \mu_0)} \quad (4.3)$$

The test statistic  $D = 2 \cdot T$  will be  $\chi^2$ -distributed in the limit of infinite sample size. Two tests need to be performed for the *dead* and *hot* hypotheses. The relationship between  $\mu_0$  and  $\mu_1$  for both tests can be defined as  $\mu_1 = f \cdot \mu_0$  where  $f$  is the factional deviation from  $\mu_0$  to flag a cell as bad. *Dead* cell is defined as a cell that triggers with a rate of 10% or less than the expected rate ( $f_{\text{dead}} = 0.1$ ) and *hot* cell is defined as a cell that triggers with rate of at least two times the expected rate ( $f_{\text{hot}} = 2.0$ ).

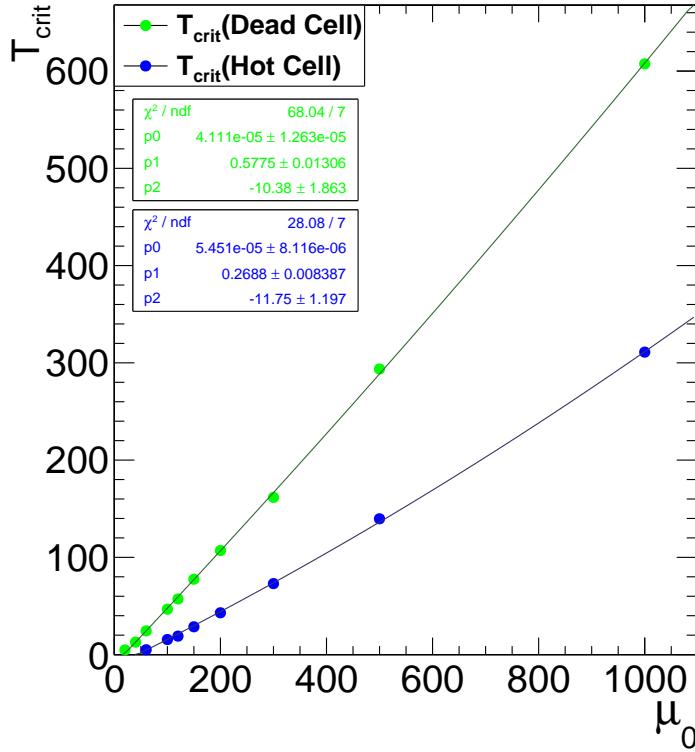
A test efficiency of 99% with a fake rate of 1% were chosen as key parameters to constrain the test behaviour, where efficiency is the probability of correctly identifying a problematic cell and fake rate is the probability of marking one or more cells as bad in a single histogram. The choice of these parameters defines the test threshold  $T_{\text{crit}}$  and implies a requirement on the minimum average number of events per cell depending on the number of bins per histogram.

If the test statistic  $T$  is above the critical value  $T_{\text{crit}}$  we reject  $H_0$  and consider the cell as bad, if it is below  $T_{\text{crit}}$  we do not reject  $H_0$  and consider the cell as good. The critical value is set by a choice of confidence level of finding a problematic cell and depends on  $\mu_0$ . To determine  $T_{\text{crit}}(\text{DeadCell})$  and  $T_{\text{crit}}(\text{HotCell})$  as a function of  $\mu_0$ , two sets of MC toy experiments were made. For each experiment, the variable  $\mu_0$  was set between 0 and 1000, which is its typical range in the histograms to be monitored, and  $\mu_1$  was set according to which bad cell hypothesis. For each toy experiment, 500 values were determined around  $\mu_1$  with poison distribution for which the test statistic  $T$  was determined. The critical value for 99% efficiency is the value of 0.01-*percentile* of  $T$  distribution. The results obtained for  $T_{\text{crit}}$  were fit with second order polynomial as a function of  $\mu_0$  (eq. 4.4).

$$T_{\text{crit}} = a \cdot \mu_0^2 + b \cdot \mu_0 + c \quad (4.4)$$

The results of the determination of each  $T_{\text{crit}}$  for each set of toys for both bad cell hypothesis and the corresponding fits can be found on figure 4.5.

To determine the minimum  $\mu_0$  as a function of the number of bins, we need to fulfil both efficiency and fake rate conditions, this can also be determined with the help of

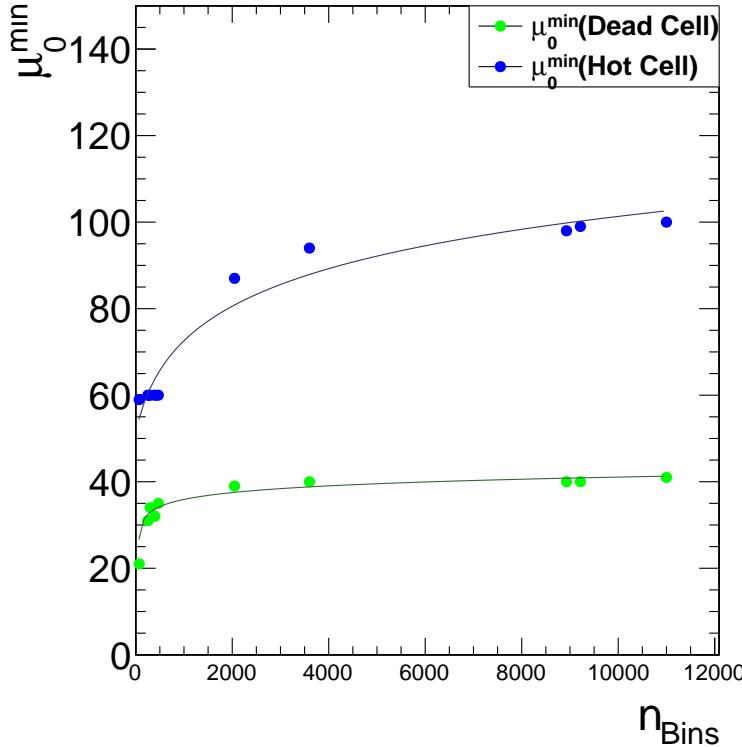


**Figure 4.5:** Graphic showing the results of the determination of  $T_{\text{crit}}$  as a function of  $\mu_0$  and the corresponding fit for each bad cell hypothesis. Results are for an efficiency of 99% and a fake rate of 1%.

MC. Searches for  $\mu_{min}$  were preformed using histograms with a predefined number of cells. For each tested  $\mu_0$ , five hundred experiments where made by filling all cells with a Poisson random numbers around  $\mu_0$ . Resulting cells where then tested with  $T$  against the critical value determined for that specific  $\mu_0$ . The fake rate is the fraction of experiments where one or more cells are marked as bad. The procedure is repeated for different  $\mu_0$  until the minimum value for this variable is found that exhibits a fake rate of 0.01 or lower. The procedure was repeated for the number of cells of all histograms to be initially monitored by this tool. The obtained values of  $\mu_0^{min}$  were fitted with a logarithm function as showed in equation 4.5.

$$\mu_0^{min} = a \cdot \ln(b \cdot n_{\text{Bins}} + c) + d \quad (4.5)$$

On figure 4.6, all the calculated  $\mu_0^{min}$  values and the corresponding fits for both bad cell hypotheses tests are showed.

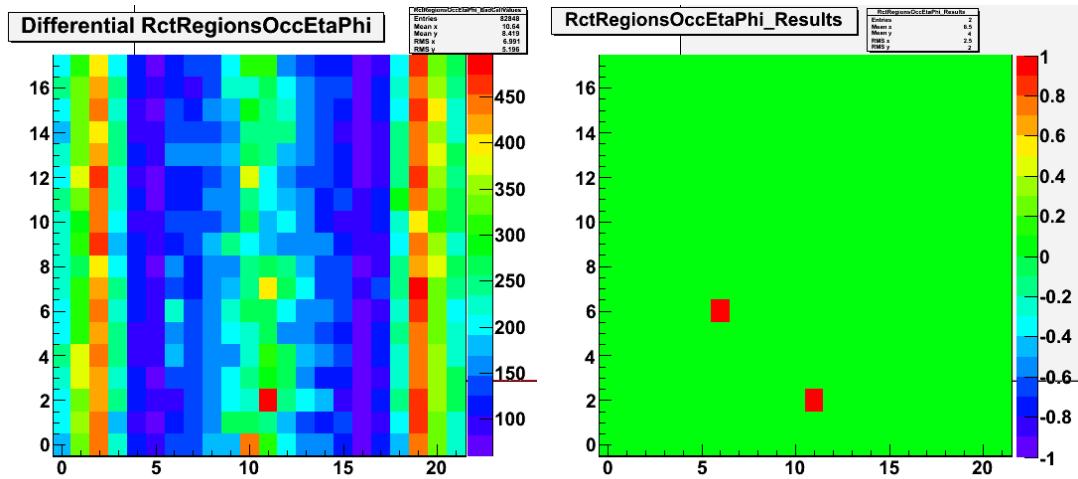


**Figure 4.6:** Graphic showing the results of  $\mu_0^{\min}$  as a function of the number of bins and the corresponding fit for each bad cell hypothesis. Results are for an efficiency of 99% with a fake rate of 1%.

### Implemented monitoring tool

The L1T occupancy monitor integrates the histograms in blocks of luminosity sections to ensure they have enough statistics. At the end of each luminosity section, we use the previously obtained fits to test each cell strip median against the histograms  $\mu_0^{\min}$  for both bad cell hypothesis tests. Cell and strips that are masked as bad are ignored. If all strips have enough statistics the bad cell tests are performed and all cells that fail are marked as bad for the period of integration of the histogram. An example of a histogram integrated for a few luminosity sections and the results of the bad cell search are shown in figure 4.7.

An additional plot is produced for each one of the monitored histograms showing the fraction of unmasked cells that pass both bad cell tests. Automatic tests are attached to these histograms and are configured to flag as bad the luminosity sections that show more than 30% bad cells. This value is too high but was set in order to allow testing of



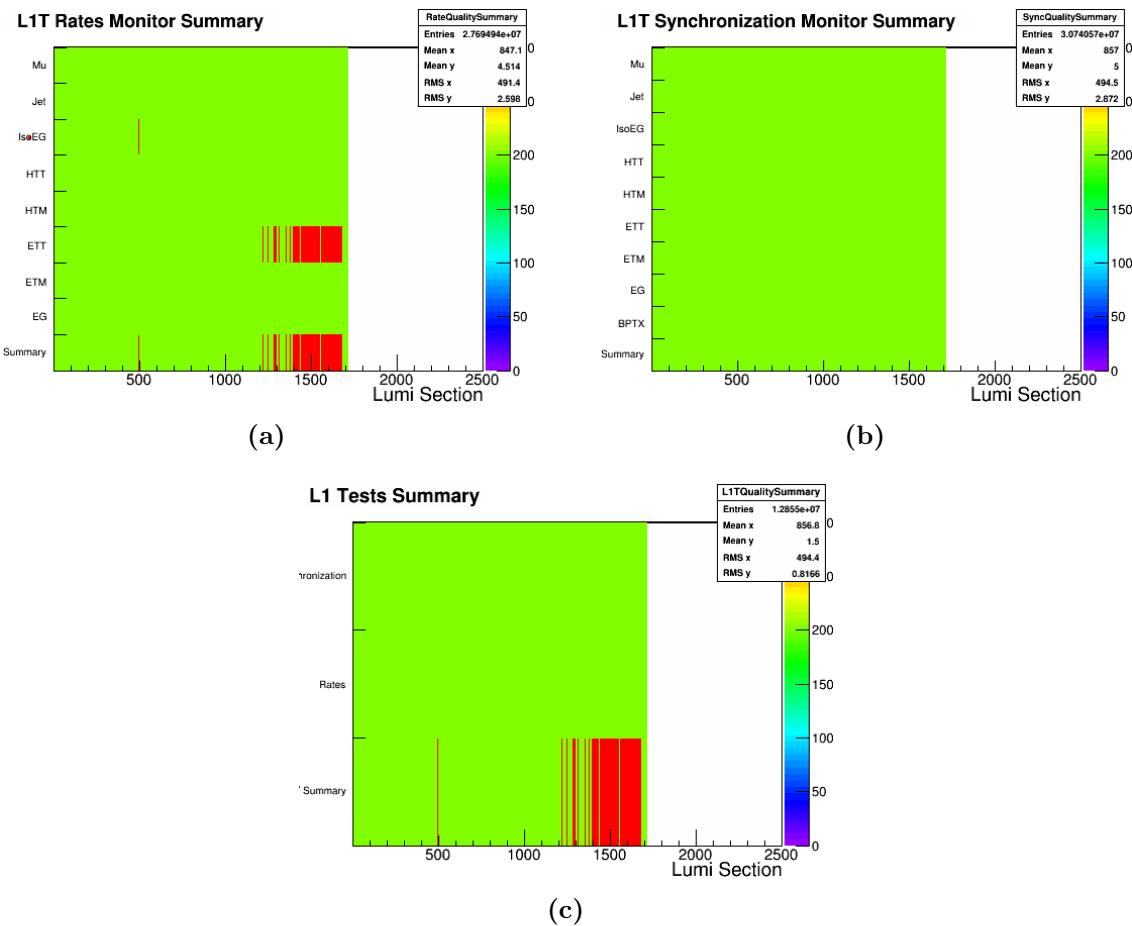
**Figure 4.7:** Monitoring plots produced by the L1T occupancy online monitoring tool for run 207099 while testing the GCT histogram for isolated EM region occupancy in  $\eta - \phi$ . On the left is the histogram under test which have been integrated for enough luminosity sections for meaningful results. On the right is an histogram where the cells that have passed the test are marked in green and red for the cells that failed. Two cells were found that fail the preformed tests.

the full implementation of the tool and not to flag luminosity sections as bad while some of the original plots need intervention by the subsystem experts.

### 4.3 Tests Summary

To simplify the task of the shift crew and the certification for physics analysis, a test summary application was developed. This tool collects the results of other tests and presents them in a single set of plots as a function of the luminosity section. Three plots are produced, summaries of the L1T rates and L1T synchronization monitoring tools, and a global tests summary. In each histogram, the bottom horizontal line is the summary of the lines above, which is marked as bad (red) if any of the tests above fail. This scheme allows the user to quickly identify a problem by back tracing information from what tests where marked as bad starting from the summary line on the *L1T Tests Summary* histogram. An example of plots produced by this application can be found in figure 4.8.

The L1T occupancy monitoring was executed over histograms produced by other developers. Some of these histograms suffered from pathological problems that needed intervention from their authors. This caused the summary from that monitoring tool



**Figure 4.8:** Example of the plots produced by the L1T test summary online monitoring. Figure (a) summary of all tests made by L1T rates monitor. Figure (b) summary of all tests made by L1T synchronization monitor, Figure (c) global summary of all tests performed.

to always be flagged as bad. Although implemented, in order to avoid confusion it was decided to not enable this summary plot or add its results to the global summary until necessary changes to the original histograms are made.

# Chapter 5

## Search for H(Inv) decays in the VBF channel with CMS prompt data

The search for Higgs boson invisible decays has already been attempted in past experiments and is also the topic of many LHC analyses. Searches were performed by the LEP experiments [127–129] and at the LHC with the full 7 and 8 TeV datasets, by the ATLAS collaboration [130–133] and by the CMS collaboration [134]. With the assumption of the SM production cross section and acceptance for the Higgs boson and using the VBF production mode, the ATLAS collaboration has placed a preliminary observed (expected) upper limit on the Higgs boson branching fraction to Invisible,  $\mathcal{B}(H \rightarrow \text{inv})$ , of 0.29 (0.35) at 95% confidence level for  $m_H = 125.5$  GeV [133]. The CMS collaboration had combined both the VBF, as presented in this chapter, and ZH production modes to set an observed (expected) upper limit on the  $\mathcal{B}(H \rightarrow \text{inv})$  at  $m_H = 125$  GeV of 0.58 (0.44) at 95% confidence level [134].

In this analysis, we focus on Higgs boson decays into invisible particles produced in association with two final state well separated quark jets. These jets will have large rapidity separation and high invariant mass. An event selection has been developed to take advantage of this distinct topology, by selecting two jets with VBF characteristics and large MET in order to separate this signal from other background processes. We have drawn inspiration from the selection criteria proposed in [135].

The main backgrounds for this analysis are the decays of  $Z$  to neutrinos with additional jets ( $Z(\nu\nu)+\text{jets}$ ), and  $W$  to a charged lepton and a neutrino with additional jets ( $W(\ell\nu)+\text{jets}$ ), where the charged lepton was not reconstructed or properly identified. These backgrounds are estimated from yields in control regions, where we select each boson decay into charged leptons together with a dijet with VBF characteristics. These

yields are extrapolated to the signal region, using conversion factors determined with the help of MC simulation. The background from QCD processes is completely estimated from control regions in data, as MC simulation is not reliable due to insufficient statistics for the extrapolation to the signal region. All other minor backgrounds like from  $t\bar{t}$ , single-top, diboson, and Drell-Yan( $\ell\ell$ )+jets processes are estimated directly from MC. A total integrated luminosity of 19.7 fb was analysed.

The observed data yield together with the estimations of the yields for the signal and backgrounds, allow us to perform a single counting experiment and draw limits on the Higgs branching fraction to invisible decay products.

## 5.1 Event Selection

In this analysis, we use the recorded data by a purpose designed trigger that selects events with at least one dijet with VBF characteristics and MET. The dijet is required to have its jets in opposite sides of the detector and pass  $p_T^{jet_1}, p_T^{jet_2} > 40 \text{ GeV}$ , jets  $\eta$  separation ( $\Delta\eta$ ) of at least 3.5 and dijet invariant mass ( $M_{jj}$ ) of a least 800 GeV. By requiring any dijet instead of the leading dijet we avoid rejecting events where a PU jet a leading jet or the effects of the lower energy resolution of the trigger versus offline. We also require  $MET_{no-\mu} > 65 \text{ GeV}$ , the use of MET without muons allows us to record with the same trigger, a control sample of processes  $W(\mu\nu)+\text{jets}$  and  $Z(\mu\mu)+\text{jets}$ . The MC simulated events are re-weighted according to the probability of passing the trigger. The trigger weights are determined in a dataset of events recorded with trigger condition requiring a single muon. They are a function of the offline measurements of sub-leading jet  $p_T$ ,  $M_{jj}$  and  $MET_{no-\mu}$ .

The signal region is defined by selecting events with a tighter version of the trigger conditions with additional cuts and vetoes. Building on the trigger requirements, we select events where the leading pair of particle flow anti- $k_T$  jets with radius of 0.5 have  $p_T^{jet_1}, p_T^{jet_2} > 50 \text{ GeV}$ ,  $|\eta_{jets}| < 4.7$ ,  $\eta_{jet_1} \cdot \eta_{jet_2} < 0$ ,  $\Delta\eta_{jj} > 4.2$ ,  $M_{jj} > 1100 \text{ GeV}$  and missing transverse energy of at least 130 GeV. Where  $jet_1$  and  $jet_2$  are respectively the leading and sub-leading jets in decreasing  $p_T$  order of the event. We veto events with identified veto electrons or loose muons, as defined in chapter 3, to suppress processes with  $Z$  or  $W$  boson decays. To reduce QCD multi-jet backgrounds we additionally request the selected dijet to pass  $\Delta\phi < 1.0 \text{ rad}$ . Typically QCD jets will be back to back and therefore will be rejected by this requirement. Finally, a Central Jet Veto (CJV)

is applied where no additional jet can be present between the two leading jets with  $p_T > 30 \text{ GeV}$ .

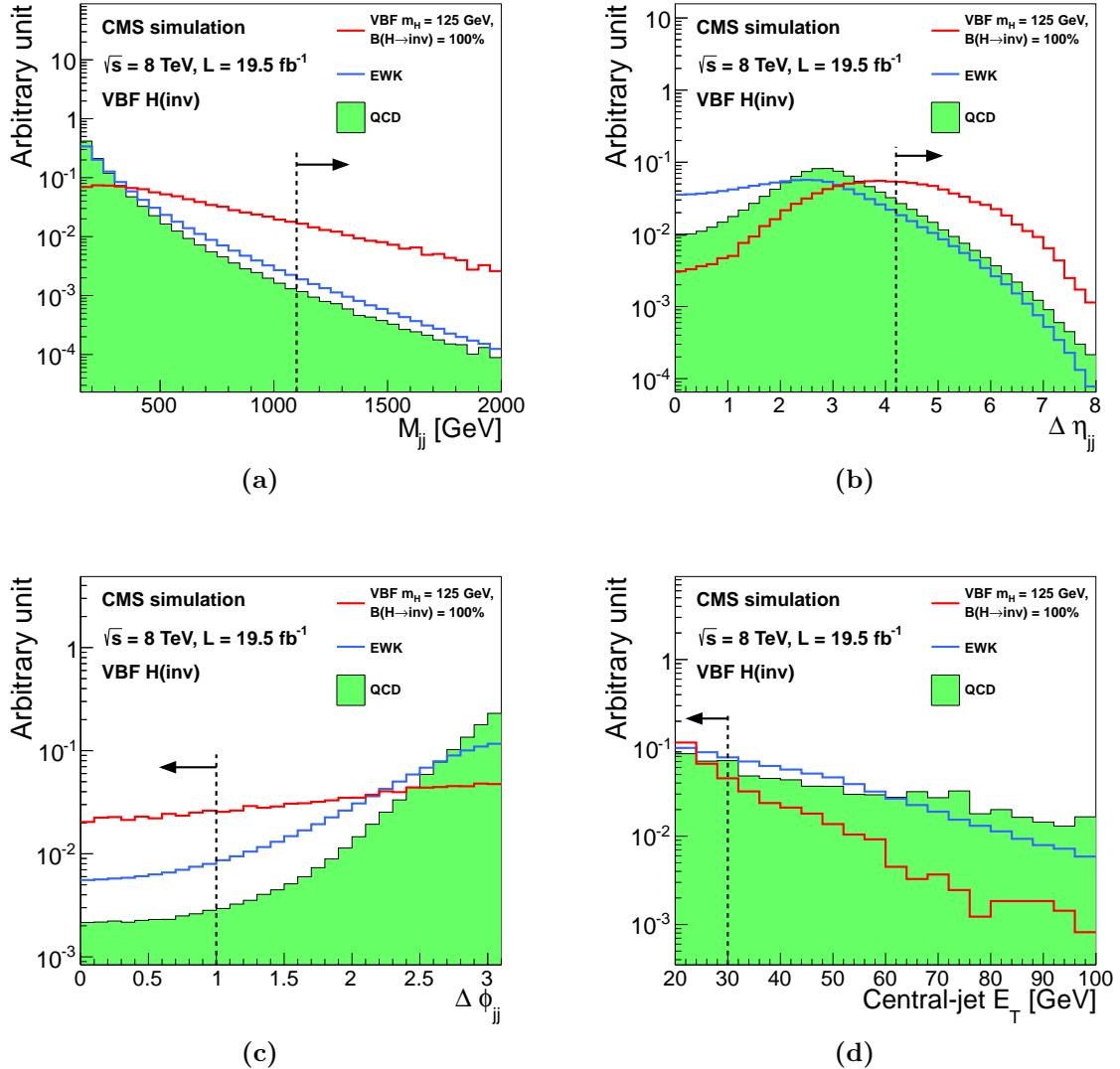
The event selection was optimized by setting the lepton vetoes to the recommended values by the relevant POG and the Central Jet Veto (CJV) to a value where its behaviour is well understood. All other thresholds were optimised to obtain the best possible signal significance which was calculated with a profile likelihood method that takes into account all relevant systematics. In this calculation, the Higgs mass was assumed to be 125 GeV and a branching fraction to invisible of 100%. The variables involved in the trigger (jet  $p_T$ ,  $M_{jj}$  and MET) are constrained to be above the 95% efficiency working point of the trigger. Distributions of the selected dijet  $M_{jj}$ ,  $\Delta\eta$ ,  $\Delta\phi$  and of the CJV obtained using MC simulation are shown on figure 5.1 together with the optimized cut thresholds.

To estimate the signal yields the POWHEG MC generator [109–116] was used to create events with a Higgs boson produced via the VBF channel with SM couplings and with mass of 125 GeV. The obtained signal efficiency was  $(6.8 \pm 0.3) \times 10^{-3}$ , which corresponds to an event yield of  $210 \pm 29$  (syst). The signal efficiency dependency on jet  $p_T$ , dijet  $M_{jj}$ , and MET are correlated and of comparable amounts. Additionally, a small amount of gluon-fusion signal, where the Initial State Radiation (ISR) emissions take the role of the VBF jets, is also expected to pass the signal event selection. Using the same MC event generator this contribution has been estimated to be of  $14 \pm 10$  (syst) events.

## 5.2 Background Estimation

The irreducible background  $Z(\nu\nu)+\text{jets}$  is estimated from data using as proxy  $Z(\mu\mu)$  decays. A control region for the  $Z$  background is defined with the same event selection as the signal region, with the following changes: instead of the muons veto, a pair of opposite charge tight muons is required with an invariant mass compatible with a  $Z$  decay of  $60 < M_{\mu\mu} < 120 \text{ GeV}$ . We veto the event if any more additional veto electrons or loose muons are present. We use  $\text{MET}_{no-\mu}$  to emulate the signature from a  $Z$  decay into neutrinos. We can extrapolate the number of events in signal region using equation 5.1.

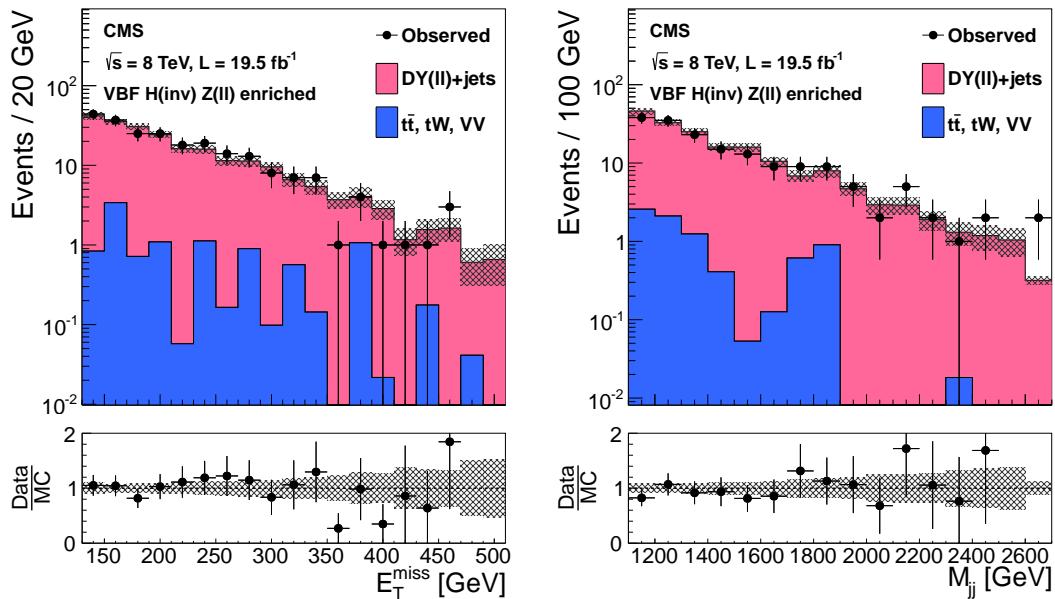
$$N_{\nu\nu}^s = (N_{\mu\mu\text{obs}}^c - N_{\text{bkg}}^c) \cdot \frac{\sigma(Z \rightarrow \nu\nu)}{\sigma(Z/\gamma^* \rightarrow \mu\mu)} \cdot \frac{\varepsilon_{Z\text{MC}}^s}{\varepsilon_{Z\text{MC}}^c}. \quad (5.1)$$



**Figure 5.1:** Distributions of (a)  $M_{jj}$ , (b)  $\Delta\eta_{jj}$ , (c)  $\Delta\phi_{jj}$ , (d) and central jet  $p_T$  in background and signal MC simulation. The distributions are shown after requiring two jets with  $p_T^{jet1}, p_T^{jet2} > 50 \text{ GeV}$ ,  $|\eta| < 4.7$ ,  $\eta_{jet1} \cdot \eta_{jet2} < 0$ ,  $M_{jj} > 150 \text{ GeV}$ , and  $\text{MET} > 130 \text{ GeV}$ . The arrows correspond to the thresholds applied for the final selection, after optimization [134].

where  $N_{\nu\nu}^s$  is the estimated number of  $Z(\nu\nu)$  events in the signal region,  $N_{\mu\mu\text{obs}}^c$  is the number of observed events in the control region in data,  $N_{\text{bkg}}^c$  is the number of other backgrounds events in the control region estimated from simulation,  $\sigma(Z \rightarrow \nu\nu)/\sigma(Z/\gamma^* \rightarrow \mu\mu)$  is the ratio of cross sections for both  $Z(\nu\nu)$  and  $Z(\mu\mu)$  processes,  $\varepsilon_{Z\text{MC}}^s$  and  $\varepsilon_{Z\text{MC}}^c$  are the  $Z$  background selection efficiency for the signal region and control region estimated both from MC. The MCFM MC generator [136] was used to estimate the ratio of

cross sections in equation 5.1 as  $\sigma(Z \rightarrow \nu\nu)/\sigma(Z/\gamma^* \rightarrow \mu\mu) = 5.651 \pm 0.023$  (syst) for  $m_{Z/\gamma^*} > 50$  GeV. The selection efficiency terms are calculated using a DY( $\ell\ell$ )+jets MC simulation, for the signal region the muons are ignored and the obtained efficiency is  $\varepsilon_{ZMC}^s = (1.65 \pm 0.27$  (syst))  $\times 10^{-6}$  and  $\varepsilon_{ZMC}^c = (1.11 \pm 0.17$  (syst))  $\times 10^{-6}$  for the control region. The event yield observed in this control region is of  $N_{\mu\mu obs}^c = 12$  events. The other backgrounds in the control region are estimated using MC simulation of the  $t\bar{t}$ , diboson and single-top processes being  $N_{bkg}^c = 0.23 \pm 0.15$  (syst) event. Using these results the contribution of the  $Z(\nu\nu)$  background in the signal region is estimated as  $99 \pm 29$  (stat)  $\pm 25$  (syst) events. Figure 5.2 shows the MET and dijet invariant mass distributions with a less strict  $Z$  control region event selection, where  $\Delta\eta_{jj}$ ,  $\Delta\phi_{jj}$  and CJV requirements are not enforced and requiring dijet  $M_{jj} > 1000$  GeV.



**Figure 5.2:** Distribution for MET on the left and  $M_{jj}$  on the right for a relaxed  $Z$  control region, with no requirements on  $\Delta\eta_{jj}$ ,  $\Delta\phi_{jj}$ , or CJV and with a  $M_{jj} > 1000$  GeV requirement. Backgrounds are shown cumulatively, normalized to data, and with systematic uncertainty shown as a hatched region. The lower panels show the ratio of data to the simulated background [134].

The  $W$  boson backgrounds,  $W(e\nu)+\text{jets}$  and  $W(\mu\nu)+\text{jets}$ , are estimated in control regions that select a single lepton. Two regions are defined following the approach used for the  $Z$  boson background. The  $W(\mu\nu)$  control region is defined by replacing the *loose muon* veto by a requirement of one *tight muon* and vetoing any event with additional *loose muons*. For this control region, the  $MET_{no-\mu}$  is used to replicate what would be expected if the muon was misreconstructed or misidentified. The  $W(e\nu)$  control region is defined

by replacing the electron veto by a requirement of one *tight electron* and vetoing any event with additional *veto electrons*. For this control region, we do not recompute MET as it is included in the trigger requirements. Equation 5.2 can be used to extrapolate the events in both these control regions to the signal region

$$N_\ell^s = (N_{\ell\text{obs}}^c - N_{\text{bkg}}^c) \cdot \frac{N_{\text{WMC}}^s}{N_{\text{WMC}}^c}, \quad (5.2)$$

where  $N_\ell^s$  is the estimated number of background events in the signal region,  $N_{\ell\text{obs}}^c$  is the number of observed events in the control region in data,  $N_{\text{bkg}}^c$  is the number of other backgrounds in the control region estimated from simulation,  $N_{\text{WMC}}^s$  and  $N_{\text{WMC}}^c$  are the number of  $W(\ell\nu)$ +jets background events in signal and control regions estimated from MC simulation. These ratios are estimated and  $N_{\text{WMC}}^s/N_{\text{WMC}}^c$  is equal to  $0.347 \pm 0.045$  (syst) for  $W(\mu\nu)$  and  $1.08 \pm 0.21$  (syst) for  $W(e\nu)$ . In data, the observed yields in the  $W(\mu\nu)$  control region is 223 events, with estimated backgrounds from other processes of  $30.4 \pm 7.0$  (syst) events. For the  $W(e\nu)$  control region the observed yield is 65 events with estimated backgrounds from other processes of  $7.1 \pm 4.7$  (syst) events. The extrapolated background in the signal region is  $66.8 \pm 5.2$  (stat)  $\pm 15.7$  (syst) events for the  $W(\mu\nu)$  background and  $62.7 \pm 8.7$  (stat)  $\pm 18.1$  (syst) for the  $W(e\nu)$  background.

The  $W(\tau\nu)$ +jets process where the tau decays hadronically  $\tau_{had}$  is estimated in a similar way to  $W(e\nu)$ +jets and  $W(\mu\nu)$ +jets. The  $W(\tau_{had}\nu)$  control region is defined like the signal region with the additional requirement of one tau following the description of chapter 3, no other additional leptons are allowed, and the CJV is not applied to increase the yield. We estimated the yield in the signal region  $N_{\tau_{had}}^s$  using equation 5.2. The conversion factor is derived again from the prediction of the number of events in the signal and control regions for this process from MC simulation. The extrapolated signal region contribution for the  $W(\tau_{had}\nu)$  background of  $53 \pm 18$  (stat)  $\pm 18$  (syst) events.

To increase the confidence in the MC background model and the extrapolations to the signal regions, we compute the expected data yields from one control region to another using conversion factors determined from MC simulation. For example, the  $W(\mu\nu)$  control region data yield is used to compute the yield of the  $Z(\mu\mu)$  region as given by equation 5.3. In all cases, the estimations agreed, within uncertainties, with the observed yields in data.

$$N_{\mu\mu}^c = (N_{\mu\text{obs}}^c - N_{\text{bkg}}^c) \cdot \frac{N_{Z\text{MC}}^c}{N_{W\text{MC}}^c}, \quad (5.3)$$

The QCD multijet background is estimated in the signal region by defining four regions depending on passing or failing the MET and CJV requirements. We define these regions, after the full remaining selection, as follows:

- A: fail MET criteria, fail CJV criteria;
- B: pass MET criteria, fail CJV criteria;
- C: fail MET criteria, pass CJV criteria;
- D: pass MET criteria, pass CJV criteria.

We use regions A, B and C to estimate the QCD multijet contribution in D. These three regions are first subtracted of the electroweak backgrounds, which are already estimated using other control regions, with event yield estimations from MC simulation. The QCD multijet yield in region D is then estimated using  $N_D = N_B N_C / N_A$ , where  $N_i$  is the number of events in region  $i$ . This method is based on the assumption that the four regions are uncorrelated, which is tested by comparing the MET distributions below  $\text{MET} < 130 \text{ GeV}$  of both pass and fail CJV. The maximum observed difference was of 40%, which is assigned as a method systematic uncertainty. Using this method, the contribution to the signal region of QCD multijet processes is  $30.9 \pm 4.8 \text{ (stat)} \pm 23.0 \text{ (syst)}$  events. To increase the confidence on this method, it was applied to a QCD multijet dominated area by changing the  $\Delta\phi_{jj}$  requirement to  $\Delta\phi_{jj} > 2.6 \text{ rad}$ . In this region  $2551 \pm 57 \text{ (stat)}$  events were observed, after subtraction of other backgrounds which were estimated from MC simulation. The prediction of this method is of  $2959 \pm 58 \text{ (stat)}$  events, which is compatible with the observation within systematic uncertainties. Furthermore, a cross-check was performed using as variables MET and  $\Delta\phi_{jj}$ , and the obtained prediction is consistent with the main method.

The remaining SM background, for the processes  $t\bar{t}$ , single-top, VV and DY( $\ell\ell$ )+jets, in the signal region are estimated directly from MC simulation to be  $20.0^{+6.0}_{-8.2} \text{ (syst)}$  events. Table 5.1 summarizes all background estimations along with the prediction of the yield for a signal with  $m_H = 125 \text{ GeV}$  and  $\mathcal{B}(H \rightarrow \text{inv}) = 100\%$ . The total expected background yield in the signal region is  $332 \pm 36 \text{ (stat)} \pm 45 \text{ (syst)}$ .

Process	Event yields
$Z(\nu\nu)$ +jets	$99 \pm 29$ (stat) $\pm 25$ (syst)
$W(\mu\nu)$ +jets	$67 \pm 5$ (stat) $\pm 16$ (syst)
$W(e\nu)$ +jets	$63 \pm 9$ (stat) $\pm 18$ (syst)
$W(\tau_{had}\nu)$ +jets	$53 \pm 18$ (stat) $\pm 18$ (syst)
QCD multijet	$31 \pm 5$ (stat) $\pm 23$ (syst)
Sum ( $t\bar{t}$ , single top quark, $VV$ , DY)	$20.0 \pm 8.2$ (syst)
Total background	$332 \pm 36$ (stat) $\pm 45$ (syst)
VBF H(inv.)	$210 \pm 29$ (syst)
ggF H(inv.)	$14 \pm 10$ (syst)
Observed data	390
S/B	70%

**Table 5.1:** Summary of the estimated number of background and signal events, together with the observed yield, in the VBF search signal region. The signal yield is given for  $m_H = 125$  GeV and  $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ . [134]

### 5.3 Sources of uncertainty

The small size of the data control event samples for the V+jets backgrounds translates into a large statistical uncertainty on the estimates in the signal region ranging from 5 – 30%. The systematic uncertainty also associated with this channels is dominated by the MC samples statistical uncertainty when calculating the conversion factors from control to signal regions. Important sources of systematic uncertainty also arise from the effects of the jet and MET energy scale and resolution. These effects are estimated by varying the scales and resolutions within their uncertainty, applying them to the jets and unclustered energy and recalculating the MET. The resulting systematic uncertainty is 13% in the signal acceptance, 7-15% in the V+jets background estimates, and 60% in the QCD multijet background estimate. As described in the previous section, an additional uncertainty of 40% is associated with the QCD multijet estimation, but this background yield is small compared with the total. Muon and electron efficiency uncertainties appear due to the scale factors used to correct MC simulation to data and are small.

For the minor backgrounds which were estimated from MC simulation, the dominating uncertainties come from the used physics process cross sections, which are set according to measurements made by other CMS collaboration analyses, and the jet/MET scale uncertainties. Theoretical uncertainties on the VBF signal yields result from Parton

Distribution Function (PDF) uncertainties, factorization and renormalization scale uncertainties. For the gluon fusion signal the dominating uncertainties arise from MC modelling of ISR and other effects. It is estimated by comparing the predictions from different MC generators and is estimated to be 60%. Gluon fusion represents a small amount of the total signal so this uncertainty only has a modest effect. Table 5.2 summarizes the uncertainties taken into account in relation to signal or total background yields. The combined effect of all uncertainties associated with the backgrounds results in an increase of about 65% in the expected upper limit on the  $\mathcal{B}(H \rightarrow \text{inv})$ .

Source	Total background	Signal
Control region statistics	11%	-
MC statistics	11%	4%
Jet/ $E_T^{\text{miss}}$ energy scale/resolution	7%	13%
QCD background estimation	4%	-
Lepton efficiency	2%	-
Tau ID efficiency	1%	-
Luminosity	0.2%	2.6%
Cross sections	0.5–1%	-
PDFs	-	5%
Factorization/renormalization scale	-	4%
Gluon fusion signal modelling	-	4%
Total	18%	14%

**Table 5.2:** Summary of the uncertainties in the total background and signal yields in the VBF channel. All uncertainties affect the normalization of the yield, and are quoted as the change in the total background or signal estimate, when each systematic effect is varied according to its uncertainties. The signal uncertainties are given for  $m_H = 125\text{ GeV}$  and  $\mathcal{B}(H \rightarrow \text{inv}) = 100\%$ . [134]

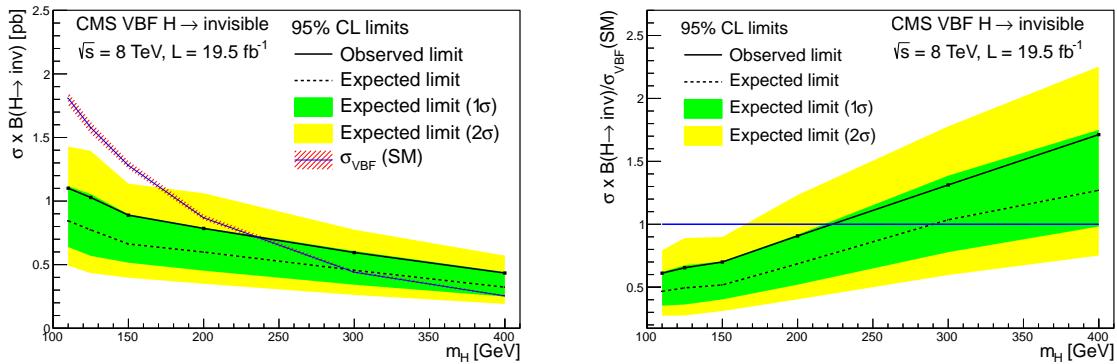
## 5.4 Results and conclusions

As shown in table 5.1, 390 data events were observed in the signal region. This yield is compatible with the background-only prediction. Since no evidence of signal is observed, 95% CL upper limits on the Higgs boson production cross section times branching fraction are computed. The limits are calculated using the  $\text{CL}_s$  method [137–139] based on asymptotic formulae [140], following the standard CMS Higgs boson searches

combination technique [3, 141]. Systematic uncertainties are incorporated as nuisance parameters and treated according to the frequentist paradigm [141]. The 95% CL limits on the Higgs boson production cross section times invisible branching fraction are also presented normalised to the SM production cross section [137, 142], which is denoted as  $\xi = \sigma \cdot \mathcal{B}(H \rightarrow \text{inv})/\sigma_{\text{SM}}$ . The choice of the SM production cross section is arbitrary, since the existence of a sizeable invisible cross section width would indicate physics beyond the SM, which could mean also modification of the production cross-sections. An alternative choice of model for Higgs boson production would not provide additional information since it essentially would scale the limits.

If SM production cross sections and acceptances are assumed,  $\xi$  can be interpreted as a limit on the invisible branching of the 125 GeV Higgs boson.

Figure 5.3 shows on the left plot the observed and median expected 95% CL limits on the Higgs boson production cross section times invisible branching fraction, as a function of the Higgs boson mass, for the VBF production mode. The right plot shows the corresponding limit on  $\xi$ . Assuming the SM VBF production cross section and acceptance, this corresponds to an observed (expected) upper limit on  $\mathcal{B}(H \rightarrow \text{inv})$  of 0.65 (0.49) for  $m_H = 125$  GeV.



**Figure 5.3:** Expected and observed 95% CL upper limits on the VBF production cross section times invisible branching fraction (left figure), and normalized to the SM Higgs boson VBF production cross section (right figure) [134].

# Chapter 6

## Run I parked data analysis preparation

The Run I CMS VBF Higgs to invisible analysis was performed over two overlapping datasets. The promptly reconstructed data became available almost immediately after recording and its analysis was already presented in chapter 5, and became known as the Run I *prompt analysis*. Simultaneously, a second data stream was recorded with lower trigger thresholds which only became available for analysis after the full LHC Run I was finished. This chapter describes the studies made to prepare the Run I *parked analysis* over this additional dataset.

A description of the studies performed to develop the trigger conditions used to record data for the VBF Higgs to invisible analysis can be found in section 6.1. This study was extended to create a condition to select VBF Higgs events independently of the final state.

The lower trigger thresholds required a re-optimization of the analysis event selection. Due to the lack of sufficient statistics of simulated events, the *prompt analysis* was tuned to suppress the selection of QCD multi-jet processes to a negligible level. The ability to lower the selection thresholds to take advantage of the lower parked data trigger requirements was therefore limited by our understanding of the QCD multi-jet background. Having a MC description of this type of events would allow easier threshold optimization and would create the opportunity for the analysis to evolve to a shape-based or multivariate-based analysis.

The production and characterization of QCD multi-jet samples with VBF characteristics and real MET is described in section 6.2. Further studies of possible approaches to suppress the QCD multi-jet background are presented in sections 6.3.1 and 6.3.2.

## 6.1 L1T parked trigger development

The first step of any analysis is defining or selecting a trigger to collect data. This trigger should have a high signal efficiency while recording at an acceptable rate.

At the beginning of 2012 the possibility of recording data without promptly reconstructing it was introduced. This additional data is now known as *parked data*. The CMS VBF Higgs to invisible analysis saw this as an opportunity to develop a secondary set of triggers with lower thresholds to allow more signal to be collected in comparison with the already developed prompt trigger. As this effort developed, it became clear that an inclusive trigger that would record VBF events regardless of final state could be implemented.

### 6.1.1 VBF Higgs to Invisible Higgs Level 1 trigger development

Data recorded during the special high PU run in late 2011, was used to study L1T trigger algorithms to be used during the 2012 proton run. During this LHC fill, the average PU was  $\approx 30$  simultaneous interactions.

The investigated algorithms select the typical topology of our signal. They look for events with MET and two jets located in opposite sides of the detector by requiring  $\eta_{jet1} \times \eta_{jet2} < 0$ , and pseudo-rapidity separation of at least  $\Delta\eta_{jj} > 3$ . The possibility of using  $\Delta\phi_{jj}$  was also studied but was disfavoured since it could cause bias in BSM searches.

The conditions expected for early 2012 were of instantaneous luminosity of  $5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$  and an average PU of 28 interactions (scenario A). For late 2012 conditions were expected to increase to instantaneous luminosity of  $7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$  and an average PU of 32 interactions (scenario B).

Algorithm parameters were optimized for both these scenarios of LHC running and several benchmark L1T rates were considered. The proposed target rate for the algorithm

suggested by the TSG was 2 kHz. Additional working points were calculated with the intention of adjusting the selection cuts according to higher or lower bandwidth available on the trigger menu. The two key variables were the selected jets transverse momentum,  $p_T^{\text{jets}}$ , and the MET and they were optimized separately. Each of these variables was set in turn to the lowest reasonable value while the other was scanned until the necessary rate value was achieved. Results for scenario A can be found in table 6.1 and for scenario B in table 6.2.

MET [GeV] ( $p_T^{\text{jets}} > 20 \text{ GeV}$ )				
$\Delta\phi$	no cut	2.5	2.1	1.8
10 kHz	32	32	32	32
5 kHz	35	35	35	35
2 kHz	41	41	41	41
1 kHz	47	47	47	46
0.5 kHz	54	54	54	53

$p_T^{\text{jets}} \text{ [GeV]} (\text{MET} > 30 \text{ GeV})$				
$\Delta\phi$	no cut	2.5	2.1	1.8
10 kHz	28	28	24	24
5 kHz	32	32	32	32
2 kHz	52	48	44	44
1 kHz	68	68	64	64
0.5 kHz	92	92	88	88

**Table 6.1:** Tables showing the L1T rate for different selection criteria for  $5 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$  and an average PU of 28 interactions (scenario A). In selected events the leading two jets are in opposite sides of the detector. On the left table the MET cut is calculated while requiring the two leading jets to have  $p_T^{\text{jets}} > 20 \text{ GeV}$ . Similarly, on the right table  $p_T^{\text{jets}}$  cut is calculated while requiring MET  $> 30 \text{ GeV}$ .

MET [GeV] ( $p_T^{\text{jets}} > 20 \text{ GeV}$ )				
$\Delta\phi$	no cut	2.5	2.1	1.8
10 kHz	36	36	36	36
5 kHz	40	40	40	40
2 kHz	47	47	47	46
1 kHz	54	54	54	54
0.5 kHz	67	66	66	64

$p_T^{\text{jets}} \text{ [GeV]} (\text{MET} > 30 \text{ GeV})$				
$\Delta\phi$	no cut	2.5	2.1	1.8
10 kHz	32	32	32	32
5 kHz	40	40	40	40
2 kHz	64	60	60	56
1 kHz	76	76	76	76
0.5 kHz	100	100	96	92

**Table 6.2:** Tables showing the L1T rate for different selection criteria for  $7 \times 10^{33} \text{ cm}^{-2}\text{s}^{-1}$  and an average PU of 32 interactions (scenario B). In selected events the leading two jets are in opposite sides of the detector. On the left table the MET cut is calculated while requiring the two leading jets to have  $p_T^{\text{jets}} > 20 \text{ GeV}$ . Similarly, on the right table  $p_T^{\text{jets}}$  cut is calculated while requiring MET  $> 30 \text{ GeV}$ .

These results were used to define working points for this trigger, which were proposed to the TSG to be included on L1T menu. The proposed trigger options were:

- Algorithm A: Lead dijet (opp. sides +  $p_T^{\text{jets}} > 20 \text{ GeV}$  +  $\Delta\eta_{jj} > 3$ ) + MET  $> 40 \text{ GeV}$
- Algorithm B: Lead dijet (opp. sides +  $p_T^{\text{jets}} > 50 \text{ GeV}$  +  $\Delta\eta_{jj} > 3$ ) + MET  $> 30 \text{ GeV}$

It can be observed that with the predicted increase of instantaneous luminosity and PU from scenario A to scenario B, the necessary rate for such proposed algorithms would escalate up to  $\approx 5 \text{ kHz}$ . For the rate to be maintained at  $2 \text{ kHz}$  the value of the MET cut in algorithm A would have to be raised to  $47 \text{ GeV}$  and the value of  $p_T^{\text{jets}}$  cut in algorithm B would have to increase to  $64 \text{ GeV}$ .

The L1T algorithm selecting only MET bigger than  $40 \text{ GeV}$  remained unprescaled in the trigger menus used during Run I. A decision was made to use this algorithm as the seed for the VBF Higgs to invisible parked HLT trigger path, thus avoiding the additional complication of having to implement a dijet plus MET L1T algorithm for minimal threshold gain.

### 6.1.2 VBF Higgs inclusive Level 1 trigger development

It would be desirable to have a dedicated VBF Higgs inclusive L1T trigger that would be decay independent. Such an algorithm would allow analysts to have a single trigger for all VBF produced Higgs decay signatures, which would imply less systematics in their comparison. Additionally, if an algorithm is used by more analyses, it will become better understood.

When selecting events based only on the presence of a dijet with VBF characteristics, we can remove the dependency on the Higgs decay. This approach would be suitable since it ignores the Higgs decays themselves. Since we are not making any assumptions of the Higgs model, we could study all possible decays, even those predicted by yet to be defined models, with a single trigger. Thus, it would be a model-independent trigger.

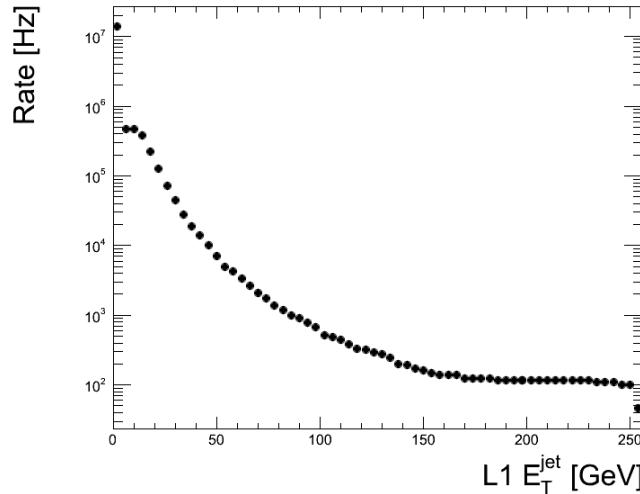
This trigger can also be used for analysis of WW scattering, which in the case of the absence of the now discovered Higgs boson would allow to eventually exclude the standard model itself, since without a Higgs boson the prediction of the SM would violate unitarity at the TeV scale [143].

Such a trigger would have to select two forward jets with VBF characteristics, and would be limited to what is possible to implement on the current L1T hardware. The following variables were considered to suppress the trigger rate of this algorithm by constraining the dijet system: dijet invariant mass, dijet transverse invariant mass ( $M_T$ ), and event

Hadronic Total (HT) energy. For this study, we always require a L1T dijet with both jets in opposite sides of the detector and  $\Delta\eta > 3$ . The selected jets  $p_T$  is scanned and L1T algorithm rate is calculated for a grid of points of each considered dijet variable.

### Dijet invariant mass

For VBF processes the outgoing dijet system is expected to have high invariant mass making this quantity a possible handle to select this type of events with a L1T trigger algorithm. Calculation of dijet invariant mass was not implemented in the L1T hardware but according to trigger experts it was in principle possible. To obtain the target rate of 5 kHz the required thresholds for scenario A when selecting a dijet  $M_{jj} > 700$  GeV a dijet  $p_T^{jets} > 55$  GeV is obtained. Figure 6.1 shows the rate scan used to obtain this result.



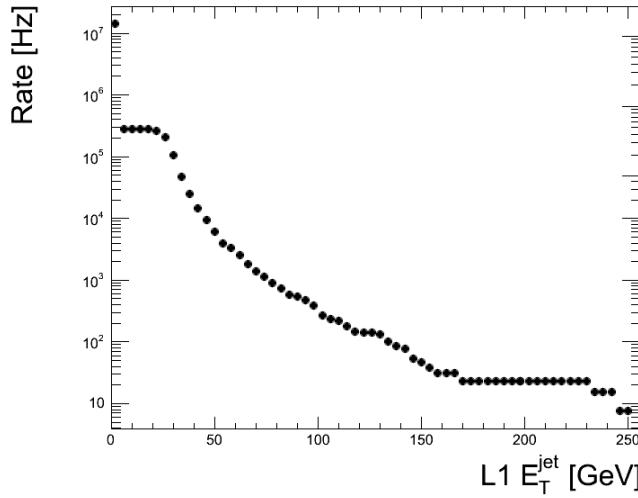
**Figure 6.1:** Level 1 rate as a function of dijet  $p_T^{jets}$  while selecting events with at least one dijet with  $\Delta\eta > 3$  and  $M_{jj} > 700$  GeV where both jets are in opposite sides of the detector for scenario A. Results based on data from the high pileup special run taken late 2011. Selecting a L1T rate of 5 kHz would result in a  $p_T^{jets} \gtrsim 55$  GeV threshold.

Unfortunately, the required dijet jets  $p_T$  is already above the prompt VBF Higgs to invisible HLT path by 15 GeV even with the high L1T  $M_{jj}$  requirement.

### Dijet transverse invariant mass

The dijet transverse mass, which is the transverse component of the dijet invariant mass, was also considered and proved to be more effective at suppressing the selection

of QCD multi-jet events. This quantity is also not implemented in the L1T hardware but, similarly to  $M_{jj}$ , could be implemented. A possible working point for scenario A was obtained with a predicted L1T rate of 5kHz. Events would be selected with at least one dijet with  $\Delta\eta > 3$  and  $M_T > 50$  GeV where both jets have  $p_T > 45$  GeV and are in opposite sides of the detector. The expected signal efficiency was calculated to be  $\approx 70\%$  for a VBF produced SM Higgs with  $m_H = 125$  GeV decaying to  $\tau\tau$ . Figure 6.2 shows the rate scan used to obtain this result.



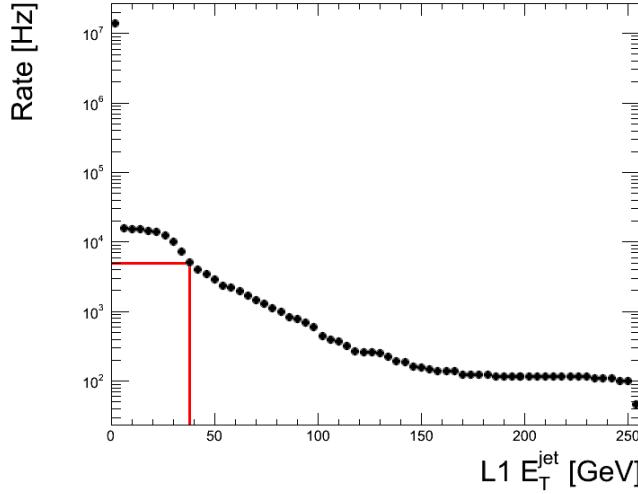
**Figure 6.2:** Level 1 rate as a function of dijet  $p_T^{jets}$  while selecting events with at least one dijet with  $\Delta\eta > 3$  and  $M_T > 50$  GeV where both jets are in opposite sides of the detector for scenario A. Results based on data from the high pileup special run taken late 2011. Selecting a L1T rate of 5 kHz would result in a  $p_T^{jets} \gtrsim 55$  GeV threshold.

The required dijet jets  $p_T$  is above the prompt VBF Higgs to invisible HLT path by 5 GeV but this should be compensated by the low  $M_T$  requirement and absence of MET requirement.

### Event scalar sum of the transverse hadronic energy

The event scalar sum of the transverse hadronic energy is the sum of the transverse energy of all L1T jets in the event. This variable has the advantage of being already implemented in the L1T hardware. A possible working point for scenario A was obtained with a predicted L1T rate of 5kHz. Events would be selected with  $HT > 100$  GeV and with at least one dijet with  $\Delta\eta > 3$ , where both jets have  $p_T > 40$  GeV and are in opposite sides of the detector. The expected signal efficiency was calculated to be  $\approx 98\%$ .

for a SM VBF Higgs with  $m_H = 125$  GeV decaying to  $\tau\tau$ . A plot of the rate scan over L1T  $p_T^{jets}$  for  $HT > 100$  GeV can be found in figure 6.3.



**Figure 6.3:** Level 1 rate as a function of dijet  $p_T^{jets}$  while selecting events with at least one dijet with  $\Delta\eta > 3$  where are in opposite sides of the detector and  $HT > 100$  GeV for scenario A. Results based on data from the high pileup special run taken late 2011. The red lines indicate the selected working point of L1T rate of 5 kHz resulting in a  $p_T^{jets} \gtrsim 40$  GeV threshold.

The two selected jets will contribute at least 80 GeV to the required HT. The additional necessary 20 GeV can come from the Higgs decay products or other unrelated energy deposits in the events. This characteristic adds some dependency on the decay channel through the need of visible decay products, but this was deemed acceptable.

### 6.1.3 Final proposal

The L1T algorithm selecting MET bigger than 40 GeV was defined as the seed for the invisible component of VBF Higgs decays. For the VBF channels with visible decays, taking into consideration the findings of this study, a similar simplified approach was taken. Depending on the delivered instantaneous luminosity, events with L1T HT bigger than 150 or 170 GeV would be selected. The developed inclusive VBF Higgs HLT path was seeded by a logical OR of these three L1T seeds.

The VBF Higgs signal efficiency for the logical or L1T MET  $> 40$  GeV and L1T HT  $> 150$  GeV is higher than 38% for all studied visible VBF produced Higgs decays including Higgs to  $\tau\tau$  with 39% efficiency and higgs to  $\gamma\gamma$  with 63% efficiency.

## 6.2 Monte Carlo simulation of QCD multi-jet events with VBF jets and MET

Simulating and reconstructing quantities of QCD multi-jet events comparable to the ones produced at the LHC experiments is impractical. At every second of LHC physics operation several millions of bunch crossings happen, each one able to create several simultaneous collisions. With the currently available hardware, it takes in excess of one minute to fully simulate one of such bunch crossing. Furthermore, most of these events have only low transverse momentum collisions and are unlikely to be picked up by any physics analysis selections.

These constraints lead to QCD multi-jet events being simulated in  $p_T$  hats, where the first simulated collision outgoing particles summed  $p_T$  is generated within a predefined range. Then several other collisions are added to the event as PU. These additional collisions are generated without any constraints in  $p_T$ .

This binned method allows the user to have access to QCD hard scattering samples with increasing energies. Such event samples allow studying the contribution of each QCD multi-jet energy range to a hypothetical analysis selection. As a practical example, we do not need to look over millions of QCD events to find high energy jets. We can just start from the highest QCD  $p_T$  hat and add lower bins until the contributing to the event selection is negligible. On the other hand, analyses like the CMS VBF Higgs to invisible analysis, search for event topologies with low energy jets and/or MET. In those cases, available inclusive QCD multi-jet event samples will not have enough statistics to provide insight into this backgrounds behaviour.

The signal for this analysis has well separated jets in  $\eta$  and large MET. Generating QCD multi-jet processes with such characteristics could allow us to simulate enough event statistics to compare with the Run I recorded data and would hopefully provide an accurate description for this background. To create such a simulated event sample, a generator level filter had to be implemented. The PYTHIA 6 MC event generator was used. Events would only be kept if at least one generator anti- $k_T^{R=0.5}$  dijet would be found with  $p_T^{jets} > 20 \text{ GeV}$ ,  $|\eta_{jets}| < 5.0$ ,  $\Delta\eta > 3.2$  and  $M_{jj} > 700 \text{ GeV}$ . Additionally, the vectorial sum of all the neutrinos in the event was required to be bigger than 40 GeV.

Similarly to the official inclusive QCD multi-jet samples, these new simulated datasets were produced in the same  $p_T$  hat bins, in the range from 80 to 600 GeV. Higher  $p_T$  bins were not simulated, since the official samples already had enough equivalent luminosity

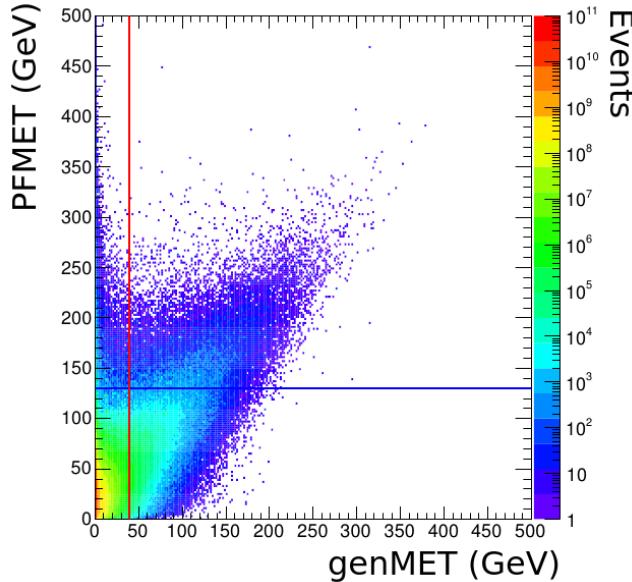
to be directly compared to data. Table 6.3 summarizes the key parameters for each produced  $p_T$  hat.

$p_T$ hat [GeV]	Filter Efficiency	Produced Events	Cross Section [pb]	Equivalent Integrated Luminosity [ $fb^{-1}$ ]
80-120	0.000049	1614416	1033680	38.09
120-170	0.000283	2051000	156293.3	44.79
170-300	0.000987	1391500	34138.15	40.28
300-470	0.002659	207840	1759.549	45.47
470-600	0.004127	104675	113.8791	219.53

**Table 6.3:** Table of the key parameters of each simulated MC event sample ordered by  $p_T$  hat.

Each sub-sample has approximately twice the integrated luminosity recorded during Run I except the bin 470 – 600 GeV which had approximately ten times the equivalent luminosity.

To illustrate the applied generator-level MET requirement, figure 6.4 shows the generator MET versus the reconstructed PF MET for the **inclusive** QCD multi-jet samples after being reweighed by their respective cross section.



**Figure 6.4:** Reconstructed PF MET as a function of generator-level MET in the inclusive QCD multi-jet samples  $80 < p_T < 600$  GeV before any selection. The red is the generator cut applied to the privately produced QCD multi-jet samples and the blue line is the *prompt analysis* signal region PF MET requirement.

Two event populations can be observed in this plot. The first is along the diagonal of the two variables, where the generator MET is correctly reconstructed into PF MET. The second population is along the vertical line at zero generator level MET, which corresponds to events with mis-reconstructed MET. When selecting events with PF MET  $> 130$  GeV, the *prompt analysis* offline requirement, the number of events with generator MET bigger than 40 GeV is only 17-31% of the total, showing that most QCD multi-jet events with high PF MET are mis-measured. The produced QCD multi-jet samples do not simulate miss measurement, but it was hoped at this time that a suitable event selection would suppress this type of events leaving only the real MET topologies.

### 6.2.1 Pre-selection for data comparison

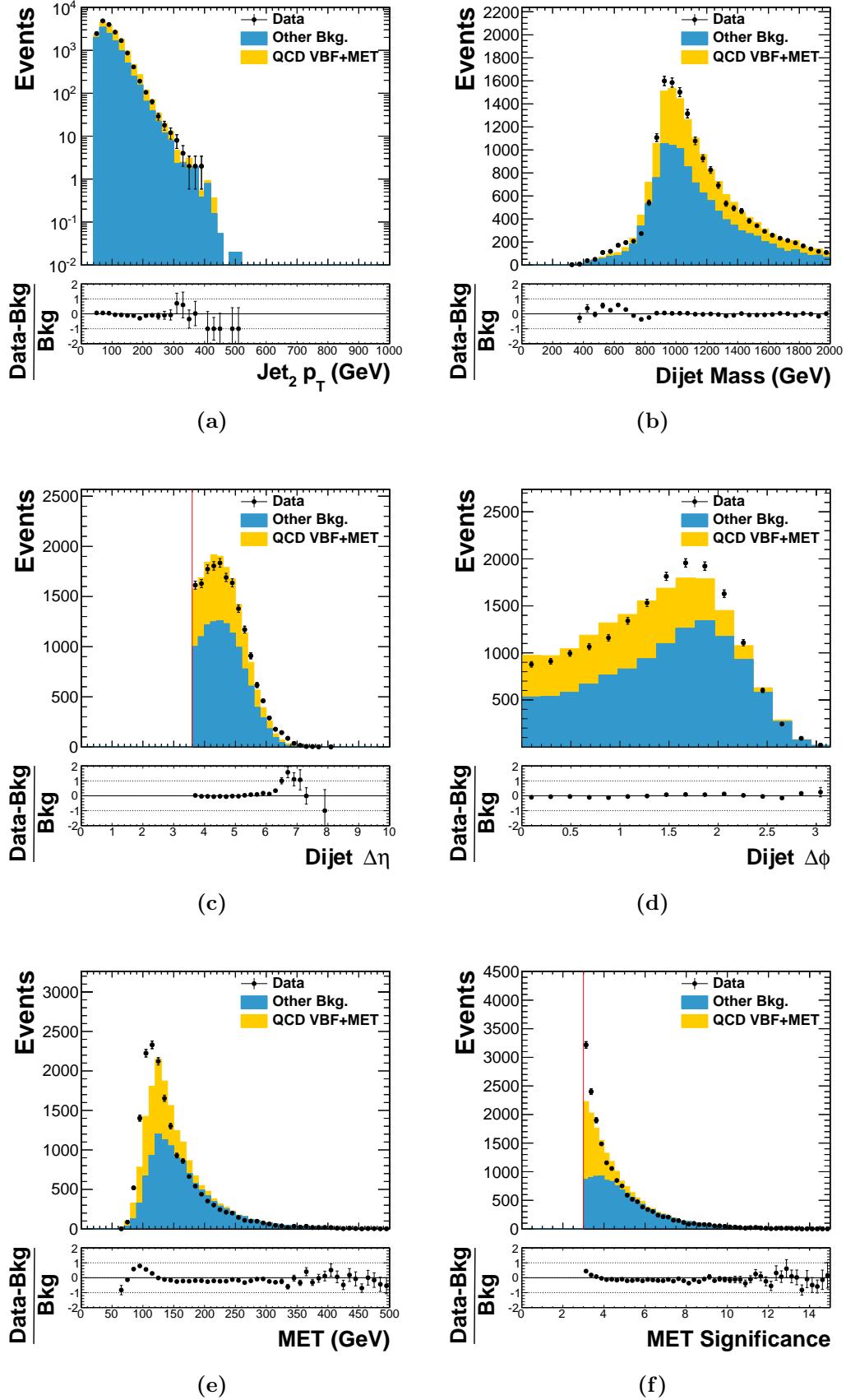
In an attempt to use these samples effectively, an event pre-selection was defined where the key variables of the VBF Higgs to invisible analysis could be described by the MC simulated datasets including the new QCD multi-jet samples. If proven reliable, such a pre-selection would be the ideal starting point for the optimization of a cut-based analysis or to train a machine learning discriminator like a BDT.

The approach taken was to methodically removed variable ranges on reference distributions where fake MET events are likely to dominate. Starting by requiring that the trigger bits were passed for data and for MC simulated events, the trigger reweighing was applied. The lepton vetos were applied to both data and simulation. Furthermore, the event leading dijet had to pass  $\Delta\eta > 3.6$ , with its composing jets passing  $p_T > 50$  GeV,  $|\eta| < 4.7$ , and  $\Delta\phi$  to the MET vector at least 1.5. The significance of the MET,  $\text{MET}/\sigma_{\text{MET}}$ , was required to be at least 3.

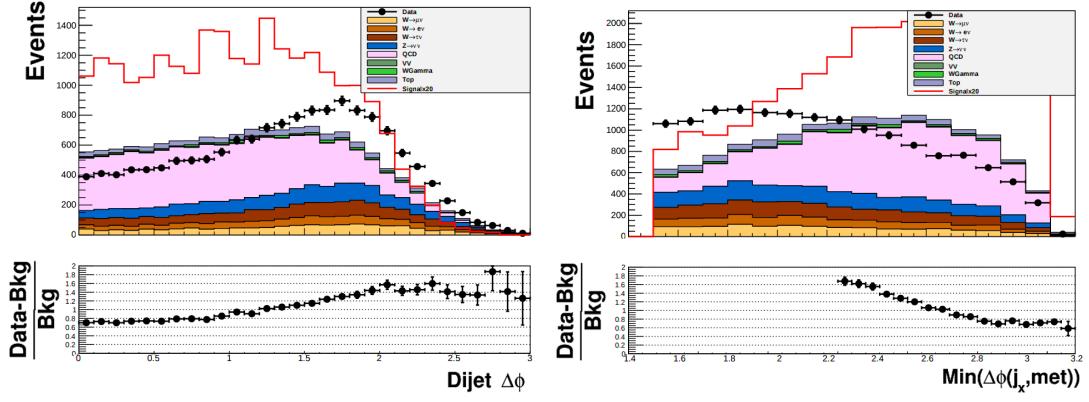
The QCD multi-jet events passing this pre-selection were weighted so that their total number would be the difference between data and all other backgrounds estimated from MC simulation. The selection signal efficiency for a VBF produced SM Higgs decay to invisible with  $m_H = 125$  GeV was estimated to be 76%. Figure 6.5 contains plots comparing data and MC simulation for some of the key variables used in the *prompt analysis* signal event selection.

Data description by the MC event samples was seen as acceptable in all variables to proceed with further studies. The biggest observed discrepancies were at low values of MET and MET significance and high values of dijet  $\Delta\eta$ .

Unfortunately, when analysing new signal and control regions inspired in the *prompt analysis* with the additional pre-selection requirements, the agreement became unacceptable for many variables, making it not possible to use these samples for shape comparison of BDT training. Figure 6.6 shows plots of  $\Delta\phi$  and  $\min(\Delta\phi(\text{MET}), \text{jets})$  for a possible loose signal selection, requiring the leading dijet to be on opposite sides of the detector and pass  $p_T^{jets} > 50, 40 \text{ GeV}$ ,  $|\eta_{jets}| < 4.7$ ,  $M_{jj} > 800 \text{ GeV}$ ,  $\Delta\eta_{jj} > 3.6$ ,  $\min(\Delta\phi(\text{MET}), \text{jets}) > 1.5$ , with  $\text{MET} > 60 \text{ GeV}$  and  $\text{MET}_{sig} > 3$ . When imposing these criteria, the new QCD multi-jet samples have significant difficulty modelling the shape difference between the other MC samples and data. The new QCD multi-jet samples had special difficulty describing events where the MET was aligned with a jet axis which points to MET mis-measurement. Although, these QCD multi-jet event samples could not fully describe this component of data, they still model a significant part of the QCD multi-jet background and have been used to study new variables to attempt to suppress at least those types of events.



**Figure 6.5:** Plots of the distributions for some of the key analysis variables after the proposed pre-selection using MC simulation including the new QCD multi-jet samples and  $19.5 \text{ fb}^{-1}$  of data. Plots are, (a) sub-leading dijet  $p_T$ , (b) dijet invariant mass, (c) dijet  $\Delta\eta$ , (d) dijet  $\Delta\phi$ , (e) MET and (f) MET significance. The red vertical lines represent the applied cuts by the pre-selection.



**Figure 6.6:** Distributions of (a) jets  $\Delta\phi$  (b) and  $\min(\Delta\phi(\text{MET}), \text{jets})$ , for a signal selection criteria requiring the leading dijet to be on opposite sides of the detector and pass  $p_T^{\text{jets}} > 50, 40 \text{ GeV}$ ,  $|\eta_{\text{jets}}| < 4.7$ ,  $M_{jj} > 800 \text{ GeV}$ ,  $\Delta\eta_{jj} > 3.6$ ,  $\min(\Delta\phi(\text{MET}), \text{jets}) > 1.5$ , with  $\text{MET} > 60 \text{ GeV}$  and  $\text{MET}_{\text{sig}} > 3$ . Monte Carlo samples are compared with  $19.5 \text{ fb}^{-1}$  of data. [144]

## 6.3 QCD control studies

Some of the requirements imposed in the *prompt analysis* signal selection will have to be lowered to take advantage of the parked data relaxed trigger thresholds. This tuning of variables may allow backgrounds to also be accepted. The QCD multi-jet background is of particular concern due to its high cross section at low energies and lack of complete and statistically sufficient MC description. To control and suppress this background, new variables were studied for possible inclusion in the new event selection, taking advantage of the physical properties expected of the signal process. These new variables were being considered as inputs for both a cut-based event selection and also for the training of a BDT discriminator against all backgrounds or simply against the QCD multi-jet background.

### 6.3.1 Dijet-MET system topological variables

When analysing events from VBF produced Higgs decays, the products of the Higgs boson should be recoiling from the outgoing VBF jets. When the Higgs decays invisibly, the measured MET should have the same magnitude as the dijet system transverse momentum and they should be pointing in opposite directions. These topological constraints do not apply to QCD multi-jet events and therefore can be used to discriminate between these processes. Two variables are defined to explore these properties:  $\alpha_{\text{vec}}$  (eq. 6.1) and  $\alpha_{\text{balance}}$  (eq. 6.2). Additionally, a “jets plus MET” energy scale variable similar to HT can be defined as  $\alpha_{\text{scale}}$  (eq. 6.3).

$$\alpha_{\text{vec}} = |\vec{p}_{\text{T}}^{\text{jet}1} + \vec{p}_{\text{T}}^{\text{jet}2} + \vec{MET}| \quad (6.1)$$

$$\alpha_{\text{balance}} = \frac{\text{Dijet } p_{\text{T}}}{\text{Dijet } p_{\text{T}} + \text{MET}} \quad (6.2)$$

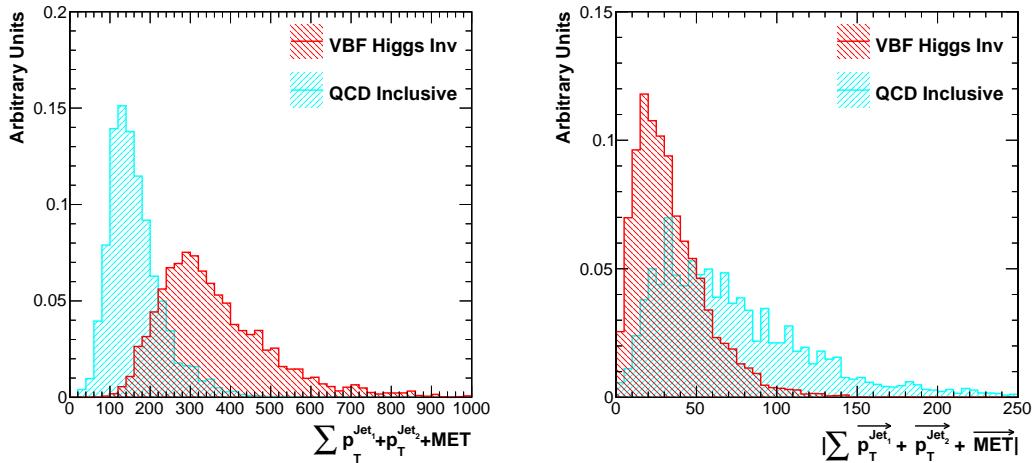
$$\alpha_{\text{scale}} = p_{\text{T}}^{\text{jet}1} + p_{\text{T}}^{\text{jet}2} + \text{MET} \quad (6.3)$$

$$(6.4)$$

Variable  $\alpha_{\text{vec}}$  should be zero for a balanced dijet plus MET system, like in the signal process. The measurement will only be smeared by the detector energy/momentum and direction resolutions. While this is a good variable, resolution changes with jets and MET magnitudes, affecting the spread of this variable. The system energy balance variable

$\alpha_{\text{balance}}$  should peak at 0.5 for signal and should only depend on the energy/momentum resolution. Signal events are only possible when large amounts of momentum are exchanged by the incoming partons, contrarily to the vast majority of QCD multi-jet events. The  $\alpha_{\text{scale}}$  allows us to set a minimum on the total amount of momentum present in the objects of interest and can be combined with  $\alpha_{\text{vec}}$  or  $\alpha_{\text{balance}}$  to further isolate the signal.

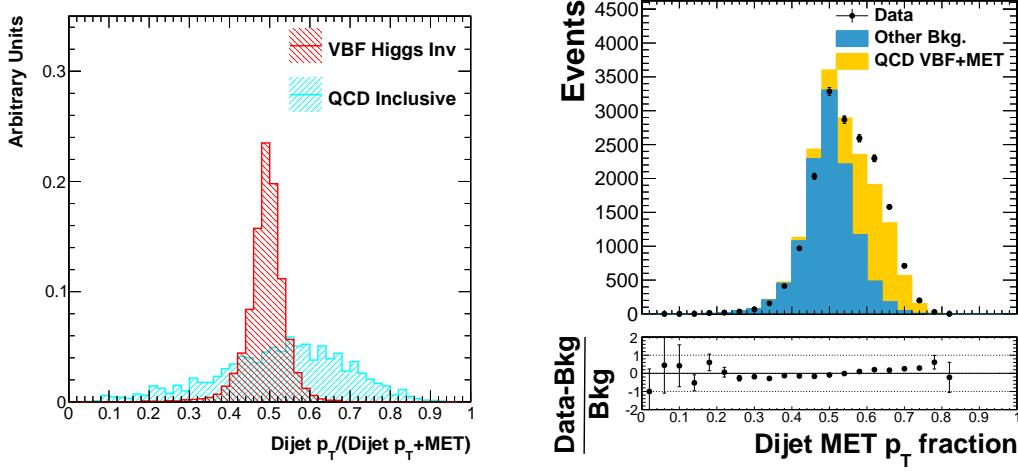
To study the shapes of these variables, MC simulation was used. Simulated VBF Higgs to invisible decays with  $m_H = 120 \text{ GeV}$  events were used for signal and inclusive QCD multi-jets events with  $\hat{p}_T > 80 \text{ GeV}$  for background. Events were selected if the leading two jets had  $p_T > 50 \text{ GeV}$  and  $|\eta| < 4.7$ ; these jets would compose the selected dijet. Figure 6.7 and 6.8 show the shape of these variables for both signal and QCD multi-jet events in this minimal selection.



**Figure 6.7:** Shape distributions for  $\alpha_{\text{scale}}$  (left) and  $\alpha_{\text{vec}}$  (right). Signal is VBF Higgs to invisible decays with  $m_H = 120 \text{ GeV}$  and QCD multi-jets ( $\hat{p}_T > 80 \text{ GeV}$ ) inclusive decays are the background. Shapes integral is normalized to 1.

It can be observed that in  $\alpha_{\text{scale}}$  the signal, as expected, is harder than most background. In  $\alpha_{\text{vec}}$  the signal is closer to zero, as expected, but there is a significant overlap in shape with the background. The most promising variable is  $\alpha_{\text{balance}}$ , where the signal peaks sharply around 0.5. Figure 6.8 also shows a plot for the pre-selection in section 6.2.1 where the new QCD multi-jet samples are used.

Good description of the data is observed and it is shown that the new QCD multi-jet samples are able to describe effects from this background that were not modelled before,



**Figure 6.8:** Plots of  $\alpha_{\text{balance}}$  distribution. The left plot shows the comparison of VBF Higgs to invisible decays with  $m_H = 120 \text{ GeV}$  and QCD multi-jets ( $p_T^{\text{jet}} > 80 \text{ GeV}$ ) inclusive decays, each shape integral is normalized to 1. The right plot shows events selected using the pre-selection defined in section 6.2.1. Simulated backgrounds including the new QCD multi-jet samples are compared with  $19.5 \text{ fb}^{-1}$  of data.

but also that  $\alpha_{\text{balance}}$  could be used to suppress a significant part of the QCD multi jet background.

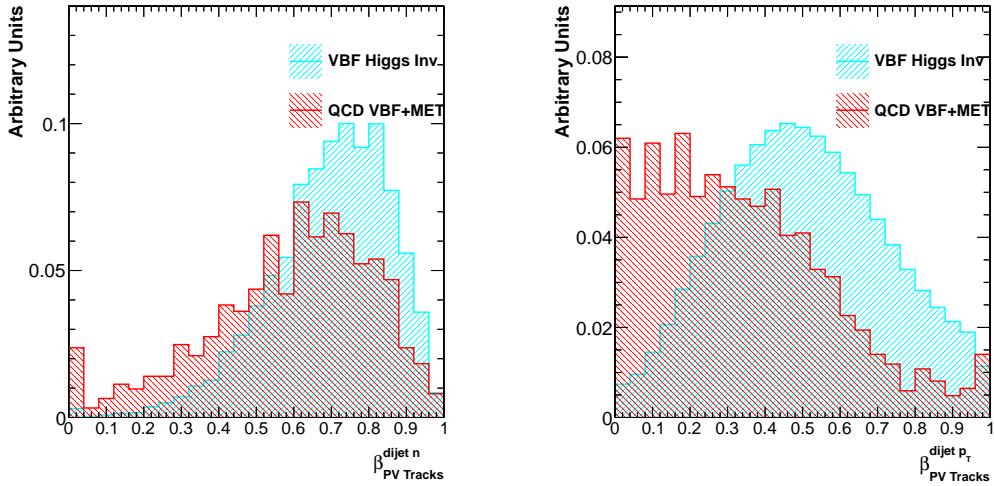
### 6.3.2 Track distribution variables

One of the important features of VBF processes is that there is no colour connection between the outgoing quark jets. As QCD multi-jet processes do exhibit this property, it is expected that these background events will have significant hadronic activity between the event jets, motivating the use of a central jet veto in the Run I *prompt analysis*. Events were vetoed if a jet was identified as coming from PV with enough  $p_T$  was found between the two leading jets. But colour connection can result simply in a spread of energy in the event which may be clustered into PU jets or remain unclustered. Also by having a minimum  $p_T$  requirement on a single veto jet may result in accepting events where multiple low  $p_T$  jets coming from the PV are present, which combined would pass the veto threshold.

For colour connected jets, it is expected that a significant number of tracks coming from the primary vertex would be spread in the event, especially between the selected jets. Forward physics analyses which target similar unconnected dijet signal topologies, use variables like the fraction of tracks inside the selected jets, to discriminate the signal

from background [145]. Using tracks has the advantage of not depending on jet clustering or veto jet definition and uses directly the excellent resolution of the CMS inner tracking system.

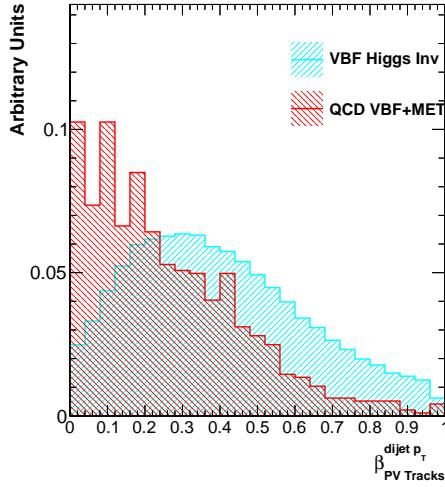
Two variables were considered for inclusion in a cut- or Multivariate Analysis (MVA)-based signal selection criteria using tracks from the PV. The fraction of tracks contained inside the selected jets ( $\beta_{\text{PV Tracks}}^{\text{dijet } n}$ ), and the fraction of summed momentum of the tracks contained inside the selected jets ( $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$ ). In both cases, only tracks above 1 GeV were considered, but this threshold could be further optimised. Figure 6.9 shows plots for both these variables when selecting events where the leading dijet has  $M_{jj} > 700$  GeV and  $\Delta\eta > 3.5$ , where both jets are in opposite sides of the detector, pass PU identification criteria,  $p_T > 50$  GeV and  $|\eta| < 4.7$ . Simulated MC VBF Higgs to invisible decays with  $m_H = 125$  GeV are used for signal, and the new QCD multi-jet samples are used for background.



**Figure 6.9:** Distributions of  $\beta_{\text{PV Tracks}}^{\text{dijet } n}$  (left) and  $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$  (right) for track  $p_T > 1$  GeV. Events are selected where the leading dijet has  $M_{jj} > 700$  GeV and  $\Delta\eta > 3.5$ , where both jets are in opposite sides of the detector, pass PU identification criteria, have  $p_T > 50$  GeV and  $|\eta| < 4.7$ .

From both analysed variables in this event selection, the  $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$  is the most promising variable to separate signal and QCD multi-jet. To prove that this variable allows additional discrimination gains to the usage of a CJV, this variable was plotted in figure 6.10 after vetoing events where a jet with  $p_T > 30$  GeV passing PU identification criteria is found between the two leading jets.

It can be observed that this variable still has good discrimination power to isolate the signal from background. There is a caveat in its usage which is the tracker coverage. The



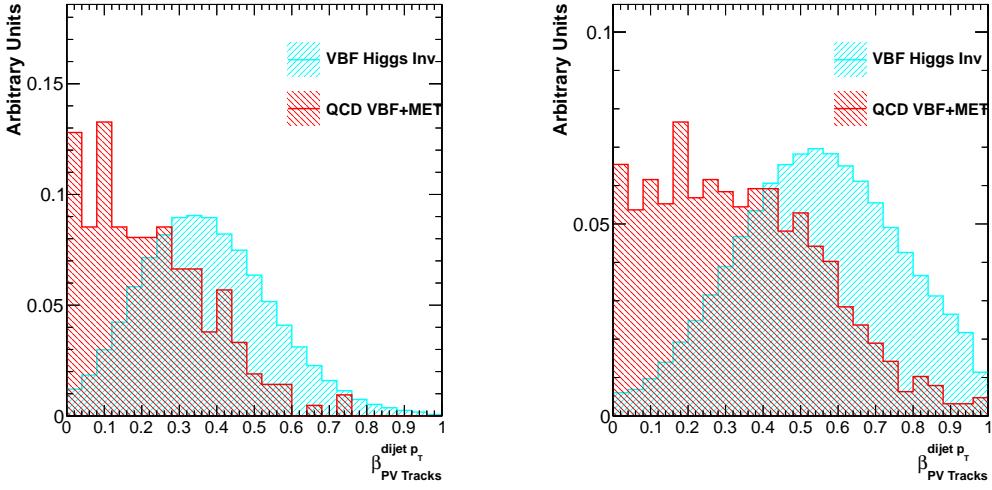
**Figure 6.10:** Distributions of  $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$  for track  $p_T > 1 \text{ GeV}$ . Events are selected where the leading dijet has  $M_{jj} > 700 \text{ GeV}$  and  $\Delta\eta > 3.5$  where both jets are in opposite sides of the detector, pass PU identification criteria, have  $p_T > 50 \text{ GeV}$  and  $|\eta| < 4.7$ . Events are vetoed when a jet with  $p_T > 30 \text{ GeV}$  passing PU identification criteria is found between the two leading jets.

majority of the signal events have both or at least one of the selected jets inside of the tracker acceptable, but there is a category of events that will have both jets in regions where tracking is not possible, which in turn will also make it very difficult to determine which is the primary vertex. Figure 6.11 shows the  $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$  distributions for events with one or both selected jets inside the tracker acceptance without the application of a CJV. In both cases, this variable remains able to discriminate signal from QCD backgrounds.

There are three categories for this quantity depending on how many selected jets are inside the tracker acceptance. To use this type of variables in a possible event selection, it may be necessary to split the selection into categories depending on the jets location. Such problem would not arise from a MVA based analysis, since these categories would be automatically learned by this methods and even possibly extended.

### 6.3.3 Summary

New quantities were found that are able to discriminate between signal and background which could be used to improve the criteria used for signal selection in the *prompt analysis*. Because of the lack of a full description of the QCD multi-jet background it was difficult to optimise the signal event selection in a relaxed threshold region with a significant component of this background. To avoid this problem, it was decided to



**Figure 6.11:** Distributions of  $\beta_{\text{PV Tracks}}^{\text{dijet } p_T}$  for track  $p_T > 1 \text{ GeV}$  for events where both (left) and only one (right) selected jets inside tracker acceptance. Events are selected where the leading dijet has  $M_{jj} > 700 \text{ GeV}$  and  $\Delta\eta > 3.5$  where both jets are in opposite sides of the detector, pass PU identification criteria, have  $p_T > 50 \text{ GeV}$  and  $|\eta| < 4.7$ .

take again the approach to suppress the QCD multi-jet background to negligible levels introducing a hard cut in the minimum  $\Delta\phi$  between jets and MET. At this point the presented variables could not contribute significantly to the improvement of the event selection and therefore were not used.

However, for the Run II re-optimisation of the CMS Higgs to invisible analysis these quantities are again being considered for inclusion in the signal event selection.

# Chapter 7

## Search for H(Inv) decays in the VBF channel with CMS parked data

This chapter describes the analysis performed over the CMS Run I proton-proton parked data collected over 2012. A total integrated luminosity of  $19.2 \pm 0.5 \text{ fb}$  was analysed. This data was recorded and stored without reconstruction and only became fully available a few months after data taking during the Long Shutdown 1 (LS1). The advantage of this approach was the possibility of using lower threshold triggers which can collect more signal but unfortunately also more backgrounds. To take full advantage of this data, the analysis had to be redesigned and extended with new control regions. To validate the obtained results a new cross check analysis was also preformed.

### 7.1 The Cross Check Analysis

It is a requirement for nearly all CMS publications to have a cross check analysis implemented independently from the main result in order to be able to ensure the accuracy of the final results due to possible errors with the software implementation. For this purpose, the CMS VBF Higgs to invisible result using prompt data presented in chapter 5, was produced by two different and independent code frameworks. Before publication a good level of synchronization was obtained validating the obtained measurement. Due to lack of man power it was initially decided to perform the 2012 parked data analysis with only a single framework. At a later stage of the analysis it was thought that at least some level of cross check would be required to take this analysis to a public result.

This cross check analysis starts from the physics object files, *ntuples*, produced by the main analysis for all the relevant datasets. The software used for this object extraction process and its data formats are also used by other analyses at Imperial College London, including both the SM and Minimal Supersymmetric Standard Model (MSSM) Higgs to  $\tau\bar{\tau}$ , the Higgs to  $\tau\bar{\tau}b\bar{b}$ , and prompt Higgs to invisible analyses. These past analysis have been cross checked independently and therefore that part of the software is considered to be sufficiently validated. No event requirements are applied at this physics object production level except the official CMS list of certified good luminosity sections for physics usage.

The analysis of these *ntuples* was performed by an independent code framework which was developed in order to replicate all relevant event yields produced by the main analysis for data and MC simulation. The data yields were produced simultaneously with the main analysis and before the results publication while the MC simulation yields and the extrapolation to the signal region were obtained at a later date for completeness.

## 7.2 Data and MC samples

### 7.2.1 Data

This analysis used the full  $\sqrt{s} = 8 \text{ TeV}$  Run I proton-proton certified collision data. The total integrated luminosity analysed is  $19.2 \pm 0.5 \text{ fb}^{-1}$  [146]. The LHC Run I was composed of four periods A, B, C and D which identify major changes to either the LHC or CMS operation, such as the deployment of new reconstruction software.

The triggers used in this analysis selected events with two jets with the distinct VBF topology and MET. Three triggers were used during 2012 depending on the data taking period. All selected trigger paths are seeded by the same L1T condition which required the event to have L1T  $\text{MET} > 40 \text{ GeV}$ . This quantity was calculated using calorimeter trigger towers up to  $|\eta| < 3.0$ . The parked triggers are additionally seeded by *HT* L1T seeds, but in this analysis events are explicitly required have been seed by L1T MET. The trigger used during Run A is the same as the one used in the *prompt analysis* already presented in chapter 5, and selects events with one pair of PF jets in opposite side of the detector with  $p_T > 40 \text{ GeV}$ ,  $M_{jj} > 800 \text{ GeV}$ , and  $\Delta\eta_{jj} > 3.5$  and PF  $\text{MET}_{no-\mu} > 65 \text{ GeV}$ . The triggers used in Run B and C (D) were the new parked data triggers, which select events with at least one pair of calorimeter jets in opposite sides of the detector with

$p_T > 35(30)$  GeV,  $M_{jj} > 700$  GeV and  $\Delta\eta > 3.5$ . A summary of the integrated luminosity collected according to the each data taking period and provenance can be found in table 7.1.

Era	Type	$\int Luminosity [pb^{-1}]$
Run A	Prompt Data	889
Run B	Parked Data	3871
Run C	Parked Data	7152
Run D	Parked Data	7317
Total analysed		19229
Total certified luminosity		19789

**Table 7.1:** Relevant parked datasets from Run I and their total analysed integrated luminosity. Total analysed and certified also showed.

The VBF Higgs inclusive parked trigger only became available in the beginning of the 2012 Run B. The difference between the certified and analysed numbers is due to the new VBF Higgs inclusive parked trigger being present but not active for the first few runs of the 2012 Run B.

### 7.2.2 Monte Carlo Samples

A variety of event generators was used to simulate the backgrounds to this analysis. The VBF Higgs to invisible signal was simulated using the POWHEG 2 event generator [109–116] and its hadronization was performed with PYTHIA 6.4.26 [102]. The main backgrounds arising from W and Z decays associated with jets (W/Z+jets) and  $t\bar{t}$  also with associated additional jets are simulated using MADGRAPH 5.1.1 [106, 108] and hadronization is also done using PYTHIA. Additional samples are used for Electroweak (EWK) Z and W processes. Table 7.2 shows the cross sections for the used samples and their equivalent integrated luminosity.

The equivalent integrated luminosities for the inclusive QCD multi-jet samples are small compared to the amount of analysed data up to the  $p_T \hat{<} 470 < p_T < 600$  GeV. Motivating the production and usage of the dedicated QCD multi-jet samples with VBF like jets and real MET described in section 6.2.

Dataset	$\sigma$ [pb]	Equivalent $\int L$ [ $\text{fb}^{-1}$ ]
$(Z \rightarrow \nu\nu) + \text{jets } (50 < HT < 100 \text{ GeV})$	381.2	10.6
$(Z \rightarrow \nu\nu) + \text{jets } (100 < HT < 200 \text{ GeV})$	160.3	27.6
$(Z \rightarrow \nu\nu) + \text{jets } (200 < HT < 400 \text{ GeV})$	41.49	122
$(Z \rightarrow \nu\nu) + \text{jets } (400 < HT < \infty \text{ GeV})$	5.274	191
$(W \rightarrow l\nu) + \text{jets } (\text{inclusive})$	37509(NNLO)	2.03
$(W \rightarrow l\nu) + 1 \text{ jet}$	5400	42.9
$(W \rightarrow l\nu) + 2 \text{ jet}$	1750	19.5
$(W \rightarrow l\nu) + 3 \text{ jet}$	519	29.9
$(W \rightarrow l\nu) + 4 \text{ jet}$	214	62.5
$(Z/\gamma \rightarrow ll) + \text{jets } (M_{ll} > 50 \text{ GeV})$	3503.71(NNLO)	8.7
$(Z/\gamma \rightarrow ll) + 1 \text{ jets } (M_{ll} > 50 \text{ GeV})$	561	42.9
$(Z/\gamma \rightarrow ll) + 2 \text{ jets } (M_{ll} > 50 \text{ GeV})$	181	121
$(Z/\gamma \rightarrow ll) + 3 \text{ jets } (M_{ll} > 50 \text{ GeV})$	51.1	216
$(Z/\gamma \rightarrow ll) + 4 \text{ jets } (M_{ll} > 50 \text{ GeV})$	23.04	278
EWK $(Z/\gamma \rightarrow ll) + 2 \text{ jets}$	0.888	3354
EWK $(W^+ \rightarrow l\nu) + 2 \text{ jets}$	6.48	1388
EWK $(W^- \rightarrow l\nu) + 2 \text{ jets}$	4.09	1466
WW	54.838(NLO)	182
WZ	33.21(NLO)	301
ZZ	17.654(NLO)	555
$W\gamma$	461.6	10.4
$t\bar{t} + \text{jets}$	245.8(NNLO)	28.2
$t$ (t-channel)	56.4(NLO)	66.6
$t$ (tW-channel)	11.1(NLO)	44.8
$t$ (s-channel)	3.79(NLO)	68.6
$\bar{t}$ (t-channel)	30.7(NLO)	63.0
$\bar{t}$ (tW-channel)	11.1(NLO)	44.5
$\bar{t}$ (s-channel)	1.76(NLO)	79.5
QCD ( $30 < p_T < 50 \text{ GeV}$ )	66285328.0	0.00009
QCD ( $50 < p_T < 80 \text{ GeV}$ )	8148778.0	0.00074
QCD ( $80 < p_T < 120 \text{ GeV}$ )	1033680.0	0.0058
QCD ( $120 < p_T < 170 \text{ GeV}$ )	156293.3	0.038
QCD ( $170 < p_T < 300 \text{ GeV}$ )	34138.15	0.170
QCD ( $300 < p_T < 470 \text{ GeV}$ )	1759.549	3.40
QCD ( $470 < p_T < 600 \text{ GeV}$ )	113.8791	34.8
QCD ( $600 < p_T < 800 \text{ GeV}$ )	26.9921	148
QCD ( $800 < p_T < 1000 \text{ GeV}$ )	3.550036	1130
QCD ( $1000 < p_T < 1400 \text{ GeV}$ )	0.737844	1310
QCD ( $1400 < p_T < 1800 \text{ GeV}$ )	0.03352235	60000
QCD ( $1800 < p_T < \infty \text{ GeV}$ )	0.001829005	534000

**Table 7.2:** Table of the MC processes, corresponding cross sections (at NLO or Next to Next to Leading Order (NNLO) when available) and equivalent integrated luminosity analysed.

## 7.3 Monte Carlo simulation to Data correction factors

To compare MC simulation with data, events must be reweighed to match observed key distributions. Weights for each event are calculated to match the data PU distribution, the probability of passing the trigger and to get the correct lepton identification probability.

### 7.3.1 Pile-up

The distribution of PU in data and MC simulated events samples is not the same. Each MC dataset is reweighed event-by-event in order to match the observed distribution in the analysed data. The average PU is estimated to be 21 interactions per bunch crossing.

### 7.3.2 Trigger efficiency

The initial event selection for data in this analysis starts with a dedicated set of triggers. During Run A the same trigger as in the *prompt analysis* is taken, while for runs B, C and D new parked data triggers are used. To maximize the usage of the event statistics of the selected MC samples, events are not vetoed when failing the trigger conditions. Instead, an event-by-event weight is calculated which takes into account how much luminosity was recorded by each one of the individual triggers. The applied weights depend on offline quantities corresponding to the ones used in the trigger conditions: PF  $MET_{no-\mu}$ , leading dijet  $M_{jj}$  and sub-leading jet  $p_T$ .

To determine the weights, turn-on curves were obtained according to these offline variables as a function of PF  $MET_{no-\mu}$  in bins of dijet  $M_{jj}$  and sub-leading jet  $p_T$  using independently recorded events by a single-muon trigger. This approach allows the determination of weights which include the correlations between these variables. The turn-on curves are obtained by fitting equation 7.1 to each bin.

$$\frac{\varepsilon_{max}}{2} \text{Erf} \left( \frac{x - x_0}{\sqrt{\Gamma}} \right) + 1, \quad (7.1)$$

where  $\varepsilon_{max}$  is the maximum efficiency of the trigger in the bin,  $x_0$  is the mid-value of the turn-on and  $\Gamma$  is the width of the turn-on [134].

### 7.3.3 Lepton Identification

The used lepton identification criteria follows the CMS Electron-Gamma and Muon POG recommendations. The same POG have measured the efficiency of identifying each lepton criteria as a function of  $p_T$  and  $\eta$ . When selecting leptons the events are reweighed using scale factors per lepton. When vetoing events in the presence of leptons, veto efficiencies are applied per lepton identified at generator level passing the acceptance requirements of electrons (muons) of  $p_T > 10 \text{ GeV}$  and  $|\eta| < 2.4(2.1)$ .

## 7.4 Signal event selection

Most events recorded by the triggers originate from QCD multi-jet processes. In this type of events, the MET requirement is fulfilled in two different ways, events with mis-measured jets creating *fake* MET and events with genuine MET involving neutrinos coming typically from heavy-flavour decays. The introduction of a MET significance,  $\text{MET}_{sig}$ , hard requirement reduces the contribution from *fake* MET events significantly. Both types of QCD multi-jet events can be suppressed by requiring that all jets in the event with  $p_T > 30 \text{ GeV}$  are separated by a minimum azimuthal angle from MET vector,  $\Delta\phi(\text{MET}, \text{jets})$ .

The trigger requirements and the need to reduce the QCD multi-jet contribution to a negligible level drive the choice of the following base criteria. Events are selected where the leading pair of PF anti- $k_T^{R=0.5}$  jets pass the requirements  $\eta_{j1} \cdot \eta_{j2} < 0$ ,  $p_T^{jet_1}, p_T^{jet_2} > 50, 40 \text{ GeV}$ ,  $|\eta_{jets}| < 4.7$ ,  $\Delta\eta_{jj} > 3.6$ ,  $M_{jj} > 1000 \text{ GeV}$ , where  $jet_1$  and  $jet_2$  are respectively the leading and sub-leading jets in terms of  $p_T$  in the event. Missing transverse energy is required to be at least  $90 \text{ GeV}$ ,  $\text{MET}_{sig} > 4.0$  and  $\Delta\phi(\text{MET}, \text{jets}) > 2.0$ . Additionally, events are rejected if a veto electron or a loose muon are found with identification criteria defined in chapter 3.

The events selected by these criteria are in order of decreasing predicted yield:  $W(\ell\nu)$  and  $Z(\nu\nu) + \text{jets}$ , QCD multi-jet,  $t\bar{t}$  and single top, dibosons, and Drell–Yan( $\ell\ell$ )+jets. The selection is further optimised by tightening the requirements in order to obtain the best 95% C.L. expected limit on  $\mathcal{B}(H \rightarrow \text{inv})$  for a  $m_H = 125 \text{ GeV}$  Higgs boson. The optimal selection, which is defined at the signal region, is found to be the presented base criteria with the additional tighter requirements of  $\Delta\phi(\text{MET}, \text{jets}) > 2.3$ ,  $p_T^{jet_2} > 45 \text{ GeV}$

and  $M_{jj} > 1200 \text{ GeV}$ . Table 7.3 shows the obtained yield for each step of the event selection obtained by the cross check analysis.

	Prompt Run A	Parked Run B	Parked Run C	Parked Run D	Total Data
Vertex Filter	3606391	132346320	228049748	308041846	672044305
Event Quality Filters	2658960	131554431	226680352	305918529	666812272
ECAL Laser Filter	2634271	131543040	226680352	305918529	666776192
HCAL Laser Filter	2634080	131543040	226679741	305918529	666775390
L1T MET $\geq 40 \text{ GeV}$	2461217	88174347	160560859	227801622	478998045
HLT Trigger	97522	75100422	137527238	152041761	364766943
$N(Electron_{veto}) = 0$	96600	74947192	137241812	151725585	364011189
$N(Muon_{loose}) = 0$	94864	74913002	137179173	151652654	363839693
Dijet requirement	18338	13678405	25090291	24082304	62869338
MET $> 90 \text{ GeV}$	4167	38178	68047	79723	190115
MET <sub>sig</sub> $> 4$	786	3396	5988	5567	15737
$\Delta\phi(\text{MET}, jets) > 2.3$	34	91	205	178	508

**Table 7.3:** Table of the step-by-step event yields for the signal region obtained by the cross check analysis. Yields are discriminated by Run I period. Exact matching was achieved with the main analysis in each run period.

The observed yield for this region is 508 events in both main and cross check analyses. Synchronization was also achieved for each Run I data taking period.

## 7.5 Control Regions

The dominant backgrounds in this analysis come from W and Z decays, and they are estimated using independent data control regions which are extrapolated to the signal region with the help of MC simulation. A new control region is introduced to estimate the minor background from  $t\bar{t}$ ; the procedure used is the same as for the W and Z backgrounds. The QCD multi-jet background is directly estimated from data. The

remaining minor backgrounds coming from dibosons and Drell Yan, are taken directly from MC simulation.

### 7.5.1 Z background estimation

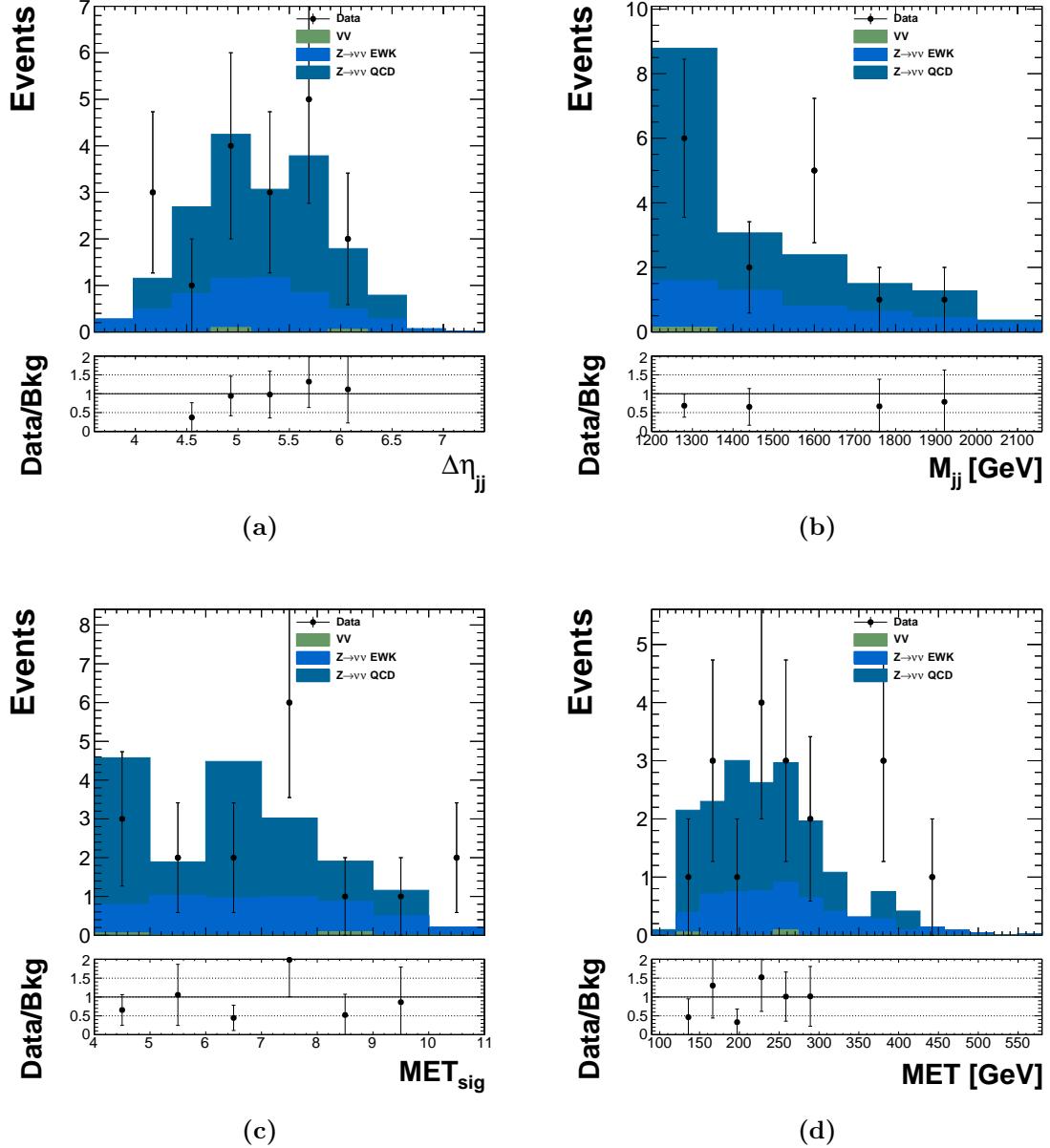
The contribution of the  $Z(\nu\nu)$ +jets process in the signal region is estimated by selecting the visible decay  $Z \rightarrow \mu\mu$ . The extrapolation to the signal region takes into consideration the difference of cross sections of these processes. The control region criteria are the same as the ones used for the signal region with the exception that the lepton veto is replaced by requiring that the only leptons in the event are a pair of *tight muons* with an invariant mass,  $M_{\mu\mu}$ , compatible with a  $Z$  decay,  $60 < M_{\mu\mu} < 120$  GeV. The events on the signal region are estimated using equation 7.2.

$$N_S^{Z \rightarrow \nu\nu} = \left( N_C^{Data} - N_C^{bkg} \right) \cdot \frac{\sigma(Z \rightarrow \nu\nu)}{\sigma(Z/\gamma^* \rightarrow \mu\mu)} \cdot \frac{\epsilon_S^{ZMC}}{\epsilon_C^{ZMC}}, \quad (7.2)$$

The ratio of cross sections  $\sigma(Z \rightarrow \nu\nu)/\sigma(Z/\gamma^* \rightarrow \mu\mu)$  was determined to be  $5.651 \pm 0.023$  (syst) and the selection efficiencies for the signal and control regions are respectively  $\epsilon_S^{ZMC} = (1.8 \pm 0.1) \times 10^{-6}$  and  $\epsilon_C^{ZMC} = (1.2 \pm 0.1) \times 10^{-6}$ . The number of observed events, by both the main and cross check analysis, in this data control region is  $N_C^{Data} = 18 \pm 4.2$ , and the number of events estimated from MC simulation in this control region originating from other backgrounds is  $N_C^{bkg} = 0.2 \pm 0.1$  (stat). The estimated contribution from EWK produced  $Z$ +jets is 21%. Figure 7.1 shows distributions of  $\Delta\eta_{jj}$ ,  $M_{jj}$ ,  $\text{MET}_{sig}$  and  $\text{MET}$ .

Within the event statistics available, a good agreement between data and MC is observed. The final estimate of the  $Z \rightarrow \nu\nu$  background is  $N_S^{Z \rightarrow \nu\nu} = 158.1 \pm 37.8$  (stat)  $\pm 21.2$  (syst).

Good agreement between data and MC simulation is observed considering the available statistics. The final estimation of the contribution of the  $Z \rightarrow \nu\nu$  background to the signal region is  $N_S^{Z \rightarrow \nu\nu} = 158.1 \pm 37.8$  (stat)  $\pm 21.2$  (syst) for the main analysis and  $151.4 \pm 39.0$  (stat) for the cross check analysis.



**Figure 7.1:** Distributions of (a) Pseudorapidity difference between the two selected VBF jets,  $\Delta\eta_{jj}$ , (b) Dijet mass  $M_{jj}$ , (c) MET significance  $S$ , (d) and MET, in the  $Z \rightarrow \mu\mu$  control region as seen by the cross check analysis.

### 7.5.2 W background estimation

A similar approach is taken for the W background. The control region is defined by the signal region criteria except that we explicitly require the presence of exactly one single lepton (tight electron or tight muon) or hadronic tau. The data event yield obtained in this region is extrapolated to the signal region with a conversion factor determined from

MC simulation. Equation 7.3 is used to obtain the predicted number of events in the signal region.

$$N_S^W = N_S^{W\,MC} \frac{N_C^{Data} - N_C^{bkg}}{N_C^{W\,MC}} = N_S^{W\,MC} \cdot SF \quad (7.3)$$

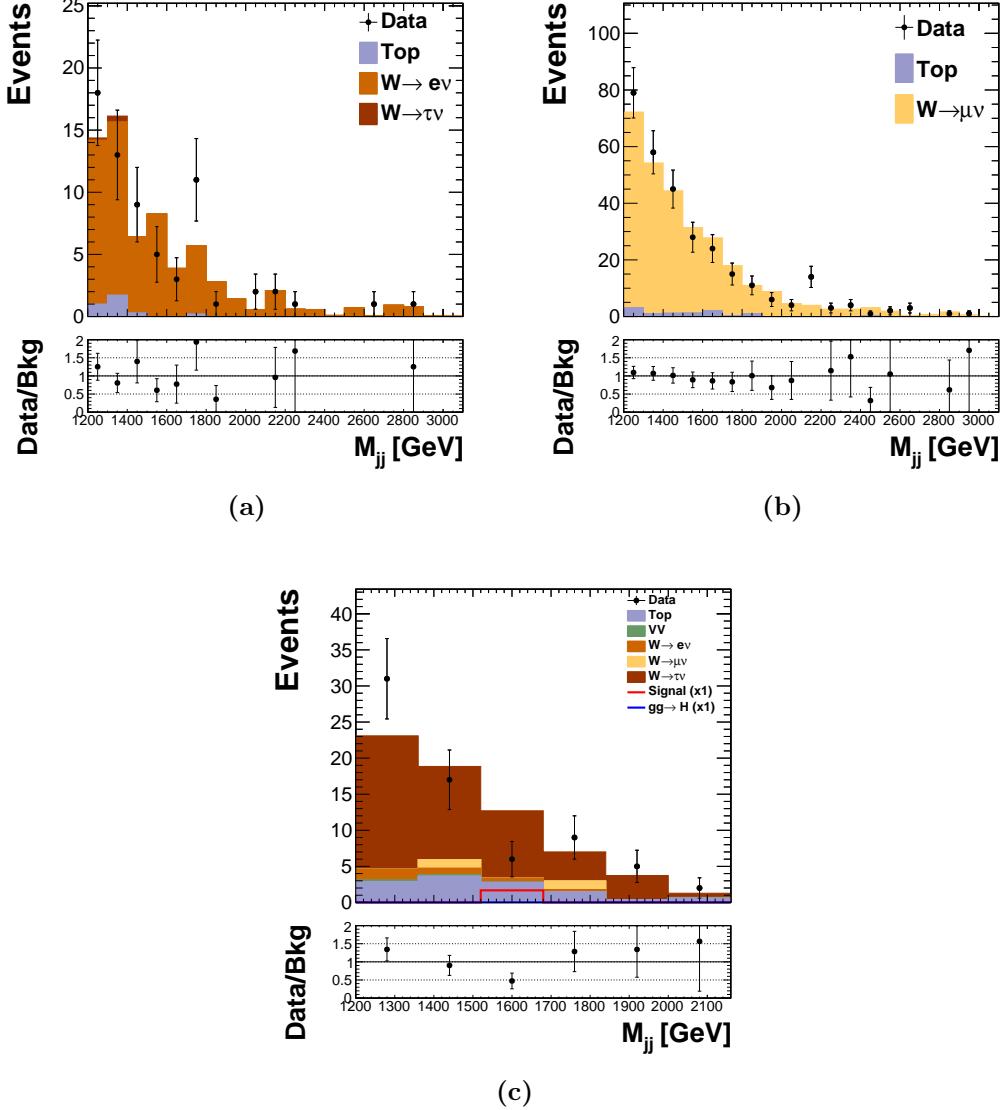
The prediction of each W decay channel is calculated separately for e,  $\mu$  (which include  $W \rightarrow \tau\nu \rightarrow \nu_\tau l\nu_l$  with  $l$  equal to e and  $\mu$  respectively) and hadronic  $\tau$ . In these control regions the number of events from other backgrounds  $N_C^{bkg}$  is mainly composed of events from top processes which are estimated from MC.

In the  $W \rightarrow \tau_h \nu$  control region, a very small number of events pass the  $\Delta\phi(\text{MET}, jets)$  requirement. In order to increase the event statistics, this is replaced by a requirement that the minimal azimuthal angle separation between the MET and one of the leading two jets  $\Delta\phi(\text{MET}, jet_{1,2})$  is greater than 1. To reject events from QCD multi-jet processes, an additional requirement on the transverse mass of the W to be greater than 20 GeV is used. The  $W \rightarrow \mu\nu$  region has enough statistics to study the full range of  $\Delta\phi(\text{MET}, jets)$ . A 20% systematic uncertainty is added to account for the observed difference in shape of the  $\Delta\phi(\text{MET}, jets)$  variable observed between MC simulation and data. Distributions of  $M_{jj}$ , MET and  $\Delta\phi(\text{MET}, jets)$  are shown in figures 7.2, 7.3 and 7.4

All presented distributions show good agreement between data and MC simulation. The event yields in each control region for both data and MC simulation and the final estimations of the  $W \rightarrow l\nu$  backgrounds in the signal region can be found in table 7.4.

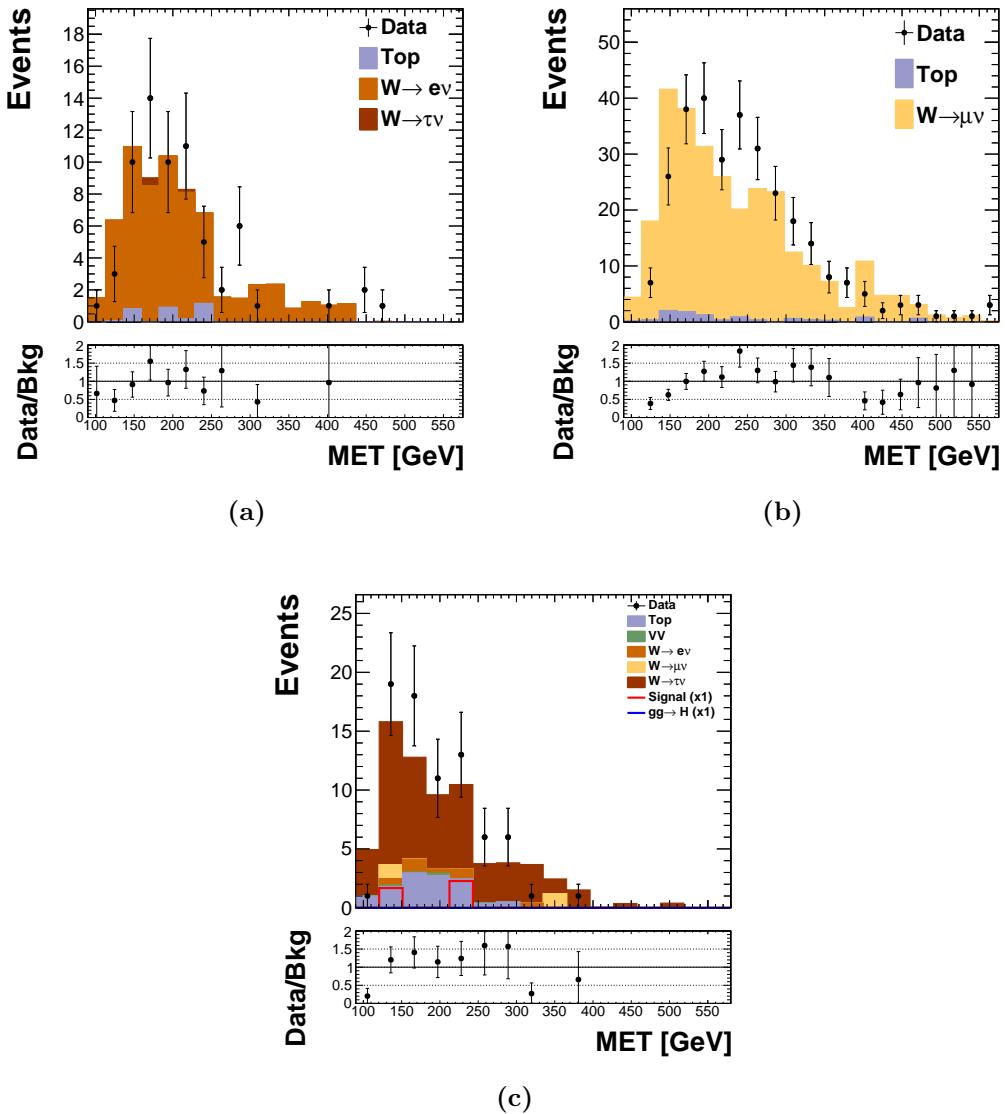
	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$
$N_C^{data}$	$68 \pm 8.2$	$300 \pm 17.3$	$76 \pm 8.7$
$N_C^{bkg}$	$6.4 \pm 1.5$	$19.3 \pm 2.4$	$23.9 \pm 3.6$
$N_C^{MC}$	$87.3 \pm 7.7$	$227.6 \pm 12.0$	$60.3 \pm 6.0$
$N_S^{MC}$	$90.3 \pm 8.7$	$119.9 \pm 9.9$	$94.0 \pm 8.5$
SF	$0.71 \pm 0.11$	$1.23 \pm 0.10$	$0.86 \pm 0.18$
$N_S$	$63.8 \pm 12.0$	$147.9 \pm 17.2$	$81.2 \pm 18.3$

**Table 7.4:** Summary of the W background estimates as calculated by the cross check analysis. The quoted uncertainties are of statistical origin.

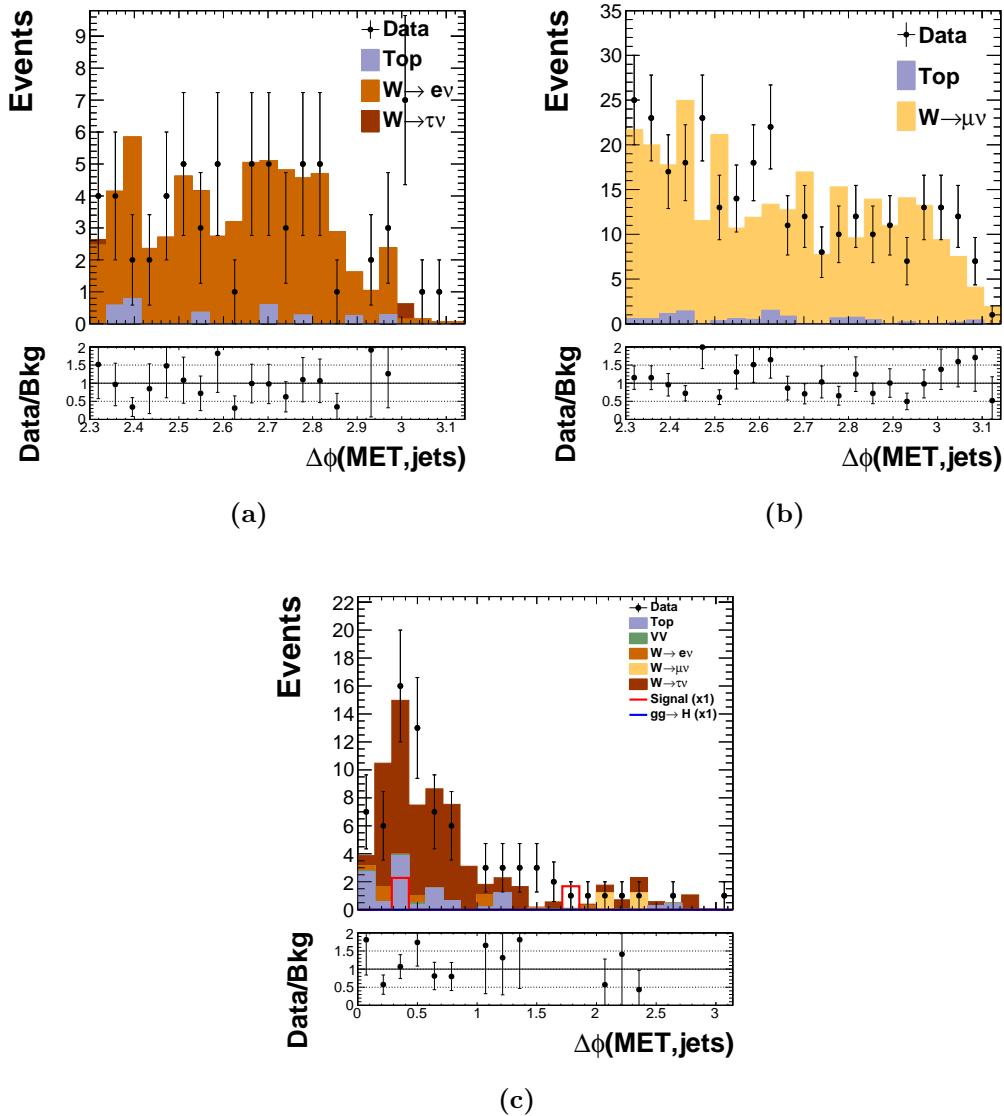


**Figure 7.2:** Dijet mass  $M_{jj}$  for the (a)  $W \rightarrow e\nu$ , (b)  $W \rightarrow \mu\nu$  and (c)  $W \rightarrow \tau\nu$  control regions obtained with the cross check analysis.

The cross check analysis has determined exactly the same data event yields for all W control selection regions. However, some discrepancies were observed in the MC simulation yields possibly pointing to a problem with the weighting of events.



**Figure 7.3:** MET for the (a)  $W \rightarrow e\nu$ , (b)  $W \rightarrow \mu\nu$  and (c)  $W \rightarrow \tau\nu$  control regions obtained with the cross check analysis.



**Figure 7.4:** Minimum azimuthal angle separation between any jet with  $p_T > 30 \text{ GeV}$  and the MET  $\Delta\phi(\text{MET}, \text{jets})$  for the (a)  $W \rightarrow e\nu$ , (b)  $W \rightarrow \mu\nu$  and (c)  $W \rightarrow \tau\nu$  control region obtained with the cross check analysis.

### 7.5.3 Top background estimation

To estimate the contribution of processes involving the top quark in the signal region, two control regions were defined. The first region is the same as the one used for the signal region with the exception that the lepton vetoes are replaced by selecting two tight leptons, exactly one tight electron and one tight muon and no other additional leptons. The  $\Delta\phi(\text{MET}, \text{jets})$  is not performed to increase statistics. This region is found to be composed almost entirely of  $t\bar{t}$  events. The data event yield in this region was determined to be 21 events, by both main and cross check analysis. The data-to-simulation scale factor obtained is  $1.21 \pm 0.19$  (data stat.)  $\pm 0.16$  (syst.).

A second region selects events with same criteria as the signal region with the exception that the lepton vetoes are replaced by selecting a single tight lepton ( $e$  or  $\mu$ ) and no other additional leptons. Additionally, one of the leading jets is required to be identified as a jet from a b quark (using the Combined Secondary Vertex algorithm [147]). The composition of this region was determined to be 10% single top, 50%  $t\bar{t}$  and 40% W+jets using MC simulation. For this region the main analysis observed 429 events which lead to a data-to-simulation scale factor of  $0.88 \pm 0.07$  (data stat.)  $\pm 0.08$  (syst.). The systematic uncertainties associated with the determined scale factors are dominated by the small statistics available in MC simulation. Taking into account these results, a 20% systematic uncertainty is assigned to the top quark contribution to the signal region.

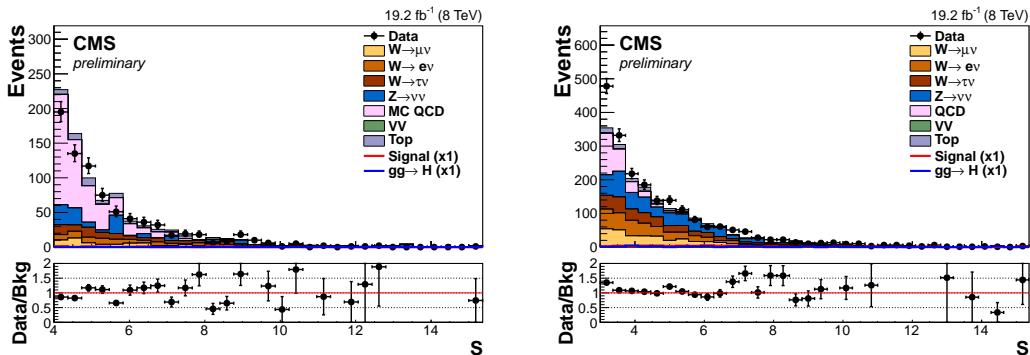
### 7.5.4 QCD background estimation

The contribution of QCD multi-jet processes is determined with a data-driven method using non-isolated MET. Three regions are defined: Region I is denoted as “inverted” and gives the description of the QCD multi-jet shape; Region II is denoted as “3-jet” where a cross check is preformed to see how well the QCD multi-jet shapes are described; and Region III is denoted as “sideband” in this region the normalization of the QCD multi-jet shapes is extracted to apply to the signal region.

The QCD multi-jet *inverted region*, is selected by changing the  $\Delta\phi(\text{MET}, \text{jets})$  requirement to  $\min(\Delta\phi(\text{MET}), \text{jets}) < 1.0$  while requiring  $\Delta\phi(\text{MET}, \text{jet}_{1,2}) > 2.3$ . The change in requirements provides two leading jets which are signal-like, but at the same time ensures MET will not be isolated. The distribution of  $\text{MET}_{\text{sig}}$  in the inverted region is shown in figure 7.5. The selected events in this region are expected to originate about 20% from W, Z and top processes. The QCD shape is defined as the shape observed

in data after the subtraction of non-QCD backgrounds, which are normalised by scale factors determined in their respective control regions, but with the same inverted selection. Good agreement between data and the VBF enriched QCD MC simulation in this region is observed as shown in figure 7.5.

The *3-jet region* is defined by requiring  $\Delta\phi(\text{MET}, \text{jets}) > 1.0$ ,  $\text{MET}_{\text{sig}} > 3$  and at least three jets with  $p_T > 30 \text{ GeV}$ . Using MC simulation the contribution from signal to this region was determined to be negligible. This region is used to ensure that the QCD shape is adequate. The expected number of QCD events in the *3-jet region*,  $n_{QCD}^{3j}$ , is the data yield in this region subtracted from backgrounds with the W and Z predictions being normalised to their control leptonic regions. The QCD shape can now be compared between the data in this region and the *inverted region* with  $\Delta\phi(\text{MET}, \text{jet}_{1,2}) > 1.0$  and normalizing it to  $n_{QCD}^{3j}$ . The distribution of  $\text{MET}_{\text{sig}}$  in the *3-jet region* is shown in figure 7.5. The discrepancy between data and the prediction is found to be less than 20%. Unfortunately, since most of the events in the signal region will only have two jets, this control region cannot be used for the final QCD multi-jet estimation.



**Figure 7.5:** (left)  $E_T^{\text{miss}}$  significance  $\text{MET}_{\text{sig}}$  for events with  $\Delta\phi(\text{MET}, \text{jets}) < 1.0$  and  $\Delta\phi(\text{MET}, \text{jet}_{1,2}) > 2.3$ . MC QCD is the QCD MC normalised to the background-subtracted data yield. (right) MET significance,  $\text{MET}_{\text{sig}}$ , for events with  $\Delta\phi(\text{MET}, \text{jets}) > 1.0$  and at least 3 jets with  $p_T > 30 \text{ GeV}$ . The QCD is modelled by data using the inverted  $\Delta\phi(\text{MET}, \text{jets}) < 1.0$  and  $\Delta\phi(\text{MET}, \text{jet}_{1,2}) > 1$  selection, after background subtraction, and normalised to the background-subtracted data yield. In both figures, the W and Z backgrounds have been normalised to their respective control regions in the same conditions. The last bin represents all those events falling above the range of the histogram [1].

The *sideband region* is used to determine the normalization of the QCD shape to be used in the signal region. This region is defined by selecting events with  $3 < \text{MET}_{\text{sig}} < 4$  and  $1.0 < \Delta\phi(\text{MET}, \text{jet}_{1,2}) < 2.0$ . In this region it is observed that the normalisation factor decreases rapidly with increasing requirements as a function of  $\text{MET}_{\text{sig}}$  and as a function of  $\Delta\phi(\text{MET}, \text{jet}_{1,2})$ . This normalization factor variation is fitted and extrapolated to the

signal region requirements. The average of the two extrapolation factors is used as the central prediction and the envelope is used to assign the systematic uncertainty on the QCD multi-jets normalisation.

## 7.6 Systematics

The dominant source of uncertainty are the statistical uncertainties associated with the yields of the control regions both in data and MC simulation, which are used for the estimation of the different backgrounds in the signal region.

The errors associated with jet energy scale, unclustered energy scale and jet energy resolution are estimated for both signal and background processes by varying each quantity independently by its uncertainties [92]. For each variation, the MET is recomputed and the signal and background predictions are recalculated. A similar procedure is applied for the lepton efficiency and PU scale factors which are applied to the MC simulation, which are also varied by their uncertainties and propagated through the analysis [79, 80].

The uncertainty associated with the integrated luminosity measurement is of 2.6% and is only applied to the MC simulated signal and backgrounds [146]. The main backgrounds are normalised using a data-driven method which takes into account the trigger efficiency, while the impact of the luminosity measurement in the signal and minor backgrounds was found to be negligible.

Uncertainties associated with diboson cross sections are taken from CMS measurements [148], while the theoretical uncertainties due to PDF and QCD scales associated to the signal cross section are taken from the LHC Higgs Cross Section Working Groups Yellow Report 3 [137, 142].

The uncertainty on the extrapolation of the  $Z \rightarrow \nu\nu$  background was obtained by comparing the QCD produced  $Z/\gamma^* \rightarrow \mu\mu$  prediction from MADGRAPH and aMC@NLO\_MG5 [108].

The results from both generators were compatible within statistical uncertainty, leading to no additional uncertainties being added.

Table 7.5 shows a summary of the overall size of each uncertainty as a percentage of the total signal and background predictions.

Source	Total background	Signal
Control region data statistics	9.3	-
MC statistics	5.4	3.8
Jet energy scale	4.6	11
$W \rightarrow \tau\nu$ control region extrapolation	4.3	-
QCD normalisation	3.2	-
Jet energy resolution	3.0	1.8
Lepton ID efficiency	2.4	-
Unclustered energy scale	1.9	1.6
Pileup weight	1.1	1.5
Top MC scale factor uncertainties	0.25	-
Luminosity	0.02	2.6
QCD scale, PDF and cross section uncertainties	0.01	5.2

**Table 7.5:** Summary of the uncertainties on the total background and signal yields. All uncertainties affect the normalization of the yield, and are quoted as the change in % in the total background or signal estimate, when each systematic effect is varied according to its uncertainties. The signal uncertainties are given for  $m_H = 125$  GeV and  $\mathcal{B}(H \rightarrow \text{inv}) = 100\%$  [1].

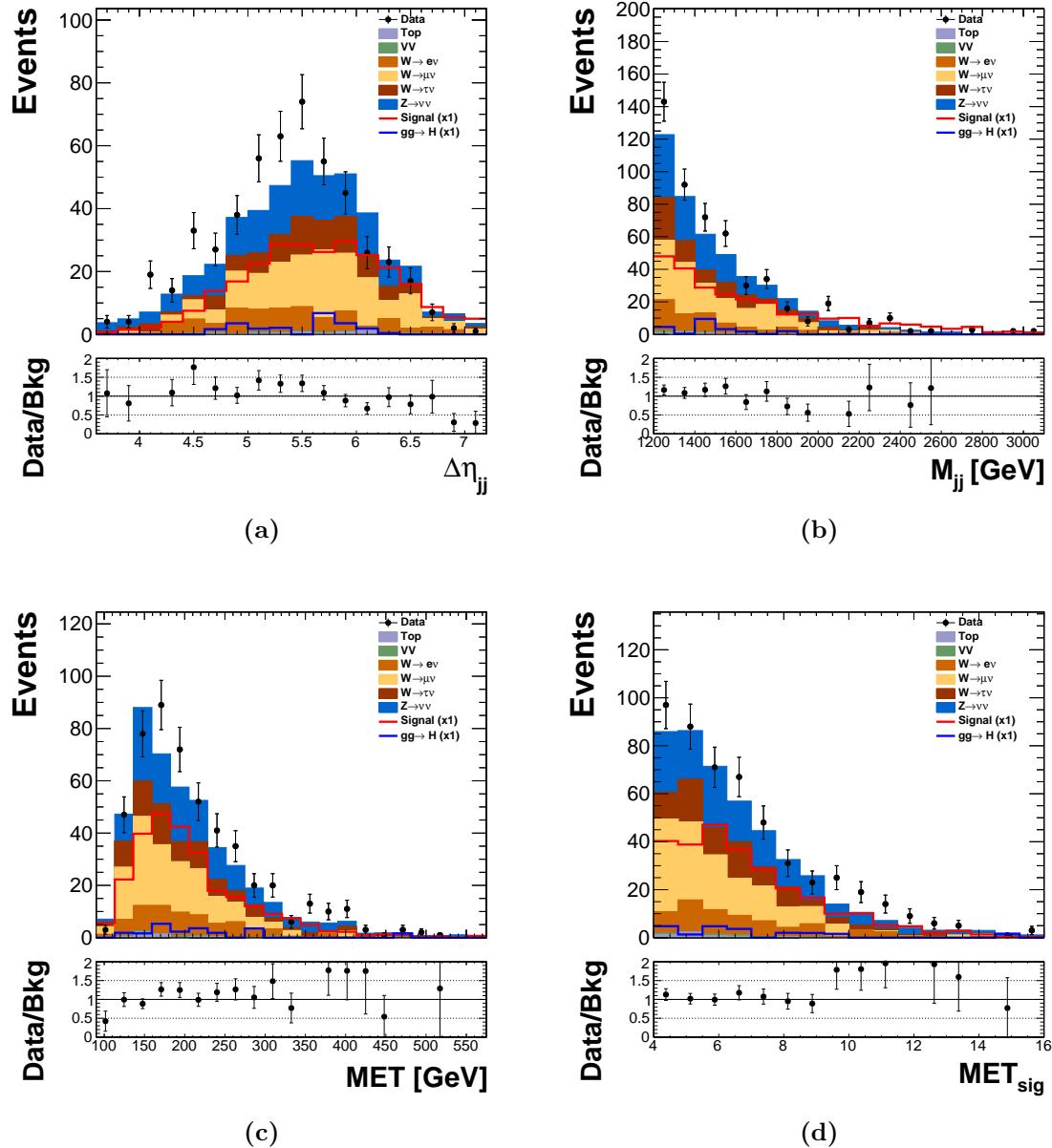
## 7.7 Results

The final estimations of the number of events predicted in the signal region are derived using scale factors determined with MC simulation for the W, Z where data control regions are used to normalise the predictions. In the case of the QCD multi-jet background the contribution to the signal region is determined using non-isolated MET events. The remaining backgrounds are estimated directly from MC simulation. These results are summarized in table 7.6 for both the main and the cross check analysis.

Distributions of the  $\Delta\eta_{jj}$ ,  $M_{jj}$ ,  $\text{MET}_{sig}$  and MET variables, in the signal region, are shown in figure 7.6.

Process	Event yields		
	Main	Cross Check	Difference [%]
$Z \rightarrow \nu\nu$	$158.1 \pm 37.3 \pm 21.2$	$151.4 \pm 39.0$	-4.25
$W \rightarrow e\nu$	$57.9 \pm 7.4 \pm 7.7$	$63.8 \pm 12.0$	+10.2
$W \rightarrow \mu\nu$	$102.5 \pm 6.2 \pm 11.7$	$147.9 \pm 17.2$	+43.4
$W \rightarrow \tau\nu$	$94.6 \pm 13.1 \pm 23.8$	$81.3 \pm 18.3$	-14.1
top	$5.5 \pm 1.8$	$5.0 \pm 1.2$	-9.1
VV	$3.9 \pm 0.7$	$3.5 \pm 0.7$	-10.3
QCD multijet	$17 \pm 14$	-	-
Total Background	$439.4 \pm 40.7 \pm 43.5$	$452.9 \pm 47.93$	+3.0
Signal(VBF)	$273.1 \pm 31.2$	$267.7 \pm 9.1$	-2.0
Signal(ggH)	$23.1 \pm 15.9$	$21.7 \pm 6.0$	-6.1
Observed data	508	508	0.0

**Table 7.6:** Summary of the estimated number of background and signal events, together with the observed yield, in the VBF search signal region. The signal yield is given for  $m_H = 125$  GeV and  $\mathcal{B}(H \rightarrow \text{inv}) = 100\%$ . For the main analysis, where two errors quoted, they are the statistical and systematic uncertainties respectively, where only one is quoted it is the systematic uncertainty [1]. For the cross check analysis only statistical errors are quoted.



**Figure 7.6:** (a) Pseudorapidity difference between the two selected VBF jets  $\Delta\eta_{jj}$ , (b) dijet mass  $M_{jj}$ , (c) MET significance  $\text{MET}_{\text{sig}}$  and (d) MET, in the signal region as produced by the cross check analysis.

### 7.7.1 Comparison with the cross-check analysis

The cross check analysis has successfully validated the main analysis by reproducing the data event yields in all the relevant regions. All event yields were measured to be exactly the same in both analysis except the yield in the QCD sideband region where a discrepancy of +1.1% was observed. Since the QCD multi-jet background is a minor background, representing less than 4% of the total background, this level of synchronization was deemed acceptable. Table 7.7 shows a comparison of the event yield obtained by both analysis and the fractional difference for each region.

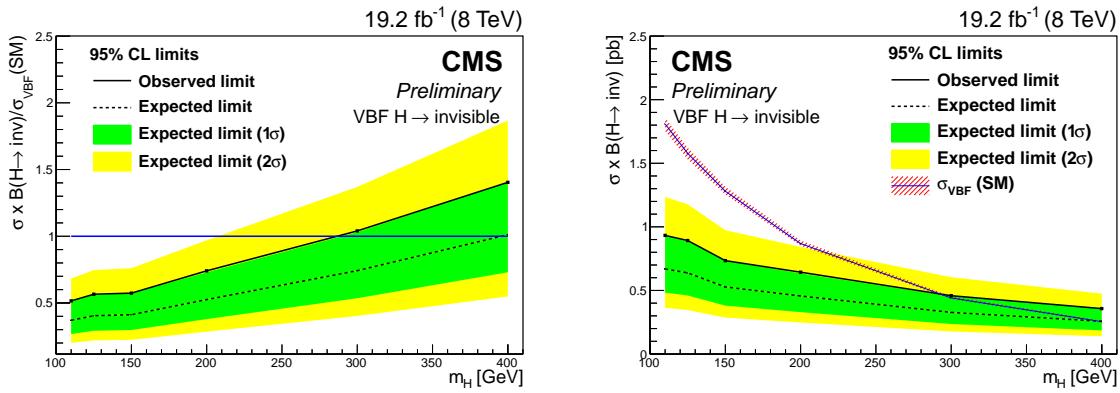
Region	Main	Cross Check	Difference [%]
$Z \rightarrow \nu\nu$	18	18	0.0
$W \rightarrow \mu\nu$	300	300	0.0
$W \rightarrow e\nu$	68	68	0.0
$W \rightarrow \tau\nu$	76	76	0.0
top (Region 1)	21	21	0.0
QCD Sideband region	1586	1603	+1.1
Signal	508	508	0.0

**Table 7.7:** Comparison of the data event yields in all relevant regions, between the main and cross check analyses. The column "Difference" is defined a  $(N_{\text{Cross Check}} - N_{\text{main}})/N_{\text{main}}$ .

## 7.8 Limits on the cross section of invisibly decaying Higgs bosons

As shown in table 7.6, 508 data events were observed in the signal region, this yield is compatible with the background-only prediction. Since no evidence of signal is observed, 95% CL upper limits on the Higgs boson production cross section times branching fraction are computed. The limits are calculated using the asymptotic  $\text{CL}_s$  method [137–139] based on asymptotic formulae [140], following the standard CMS Higgs boson searches combination technique [3, 141]. Systematic uncertainties are incorporated as nuisance parameters and treated according to the frequentist paradigm [141] and all correlations between processes are taken into account.

If SM production cross sections and acceptances are assumed, the observed (expected) 95% C.L. limit on  $\mathcal{B}(H \rightarrow \text{inv})$  of a SM 125 GeV Higgs boson is 57% (40%). Figure 7.7 shows the the 95% C.L. limit on  $\mathcal{B}(H \rightarrow \text{inv})$ and 95% C.L. limit on the cross section times  $\mathcal{B}(H \rightarrow \text{inv})$ , both under the assumption of SM Higgs boson acceptances, as a function of Higgs mass.



**Figure 7.7:** The 95% C.L. limit on  $\mathcal{B}(H \rightarrow \text{inv})$  of a SM Higgs boson (left) and the 95% C.L. limit on the cross section times  $\mathcal{B}(H \rightarrow \text{inv})$  (right) as a function of the Higgs boson mass, assuming SM Higgs boson acceptances [1].

As it can be observed in table 7.5, similarly to the *prompt analysis*, the dominant source of systematic uncertainty is the limited number of events present in the control regions both in data and MC simulation. This effect is particularly noticeable in the  $Z$  control region, if this region statistical uncertainty was of the order of the one measured in the  $W \rightarrow \mu\nu$  control region the expected 95% C.L. limit on the cross section times  $\mathcal{B}(H \rightarrow \text{inv})$ for a SM 125 GeV Higgs boson would be reduced to 33%.

# Chapter 8

## Run II preparation

After the successful completion of the first data taking period, the Run I of the LHC, the accelerator and detectors went through a two year long technical shut-down which was designated the LS1. During this period the accelerator completed a consolidation and improvement program to allow a ramp up of the beams energy up to the design value of 7 TeV per beam in proton-proton mode. At the same time the experiments also performed maintenance, repair and improvement programs.

Data analysis continued during this period of no data taking using the datasets already available or the newly reconstructed parked data. Gradually CMS physics analysis groups started finishing their Run I analyses and shifted their focus to the preparation for the Run II of the LHC, where higher collision energies, higher values of PU and more recorded integrated luminosity are expected. As part of this global effort, the CMS VBF Higgs to invisible analysis also started its own preparation work.

The first step is always the definition of a trigger condition for data taking. The effort made to define and study an adequate set of triggers is documented in section 8.1. Additionally, a study was conducted that led to the proposal of the creation of a dedicated QCD MC sample with signal-like characteristics expanding on similar samples made for Run I. This study can be found in section 8.2.

### 8.1 Run II trigger studies

The first step of any CMS physics analysis is to define which trigger to use for data taking. The TSG develops generic usage trigger conditions, known as trigger paths, which can be used by any analysis. Typically these conditions cover all possible single objects

(single electron, single jets, etc), multiple objects (double electron, triple muon, etc), cross triggers (single electron + single muon, etc) and sums (MET, HT, etc). In some cases, as for this analysis, it is better to define a custom condition to obtain maximum physics content at trigger level. The following reasons drove the decision to create a set of dedicated trigger paths.

- Maximize signal collection efficiency by selecting our signal topology with reduced trigger thresholds compared to generic triggers;
- Use a trigger condition with  $\text{MET}_{no-\mu}$  instead of MET to study the irreducible EWK Z background;
- Create a new dedicated prescaled trigger path with reduced thresholds with objective of reducing systematics;

In order to propose new triggers, it was decided to consider low rate and high rate scenarios in terms of the available L1T and HLT bandwidth. For the L1T, the low rate scenario was the usage of only the lowest threshold unprescaled MET algorithm on the menu, while for the high rate scenario a dedicated additional L1T seed algorithm with a pure rate (without accounting for overlaps with other trigger) of up to 5 kHz would be proposed. For the HLT signal trigger path, rates of 1.5 Hz and 5.0 Hz were considered and of 0.1 Hz and 0.5 Hz for the background path.

### 8.1.1 Methodology

To study new L1T algorithms for a never before attempted collision energy, MC simulation must be used. At this level the system has to analyse all collisions which are produced by the LHC, which implies that the test simulated event sample cannot have significant generation cuts. For this purpose, the so called neutrino gun event samples are used. In these event simulations, the hard process is replaced by the production of a single neutrino which will escape the detector without leaving any deposit. Pile-Up (PU) events are added to the event following a Poissonian distribution with its centre chosen according to predicted LHC performance scenarios. This PU events are selected randomly from a large QCD multi-jet sample, generated with minimum restrictions. This type of sample normally is denominated by *Minimum Bias QCD*. The final event content will be the overlap of many minimum bias events without any hard process as expected from the great majority of collisions.

At the HLT, the events are already pre-selected by the L1T and the dominating processes at this point are dependent on the characteristics of both the L1T seeding algorithm and the HLT conditions. For the CMS VBF Higgs to invisible analysis, the trigger conditions take advantage of the topology of the VBF jets and MET. These characteristics make high energy QCD multi-jet events the dominating source of rate for any HLT paths that will collect specifically our signal process.

The trigger system that was present in the beginning of Run II was an upgraded version of the one previously used. As such, its response had to be emulated using available MC samples. The latest version available of the L1T stage 2 and HLT systems description was used to perform these studies.

The target conditions of LHC running used in this study and required for the development of new algorithms, the *TSG high luminosity scenario*, are an instantaneous luminosity of  $1.4 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ , an average PU of 40 collisions and a bunch separation of 25 ns.

### 8.1.2 L1T algorithm development

In the L1T trigger menu, the reference algorithm was selected as the lowest unprescaled MET trigger available, which was L1\_ETM70. This algorithm selects events with L1T MET of 70 GeV which is only calculated in  $|\eta| < 3.0$  range. It has an expected pure rate of  $\approx 7 \text{ kHz}$  and a signal efficiency of 27-28%.

When designing an offline analysis, normally it is desirable to select events in a parameter space where the trigger efficiency is close to 100%, avoiding the need to re-weight the MC simulated events to match the trigger behaviour. Unfortunately, even if the default L1T algorithm has a reasonable signal collection efficiency, it is likely to only be fully efficient when selecting events with offline PF MET two to three times higher than the trigger threshold, implying a significant increase of this offline variable requirement when compared with that used during Run I. For these reasons, the priority was to find a solution that would allow a lower threshold to be applied to L1T MET by requiring additional conditions.

To determine the best possible algorithm, an automatic optimization method was implemented. Several base dijet configurations were defined with one key variable being allowed to float to achieve a target rate. A maximum rate of 5 kHz was set as an optimistic acceptable pure rate for such an algorithm. All base configurations start by requiring at least one L1T dijet being on opposite sides of the detector (VBF condition). Many

possible configurations were tested requiring the selected jets to have  $p_T > 30, 40$  and dijet  $\Delta\eta > 3.0, 3.5$ . Scanned variables included lead jet  $p_T$ , sub-lead jet  $p_T$ , dijet  $M_{jj}$ , L1T MET and L1T Missing Hadronic Total (MHT) (minus the momentum sum of all L1T jets).

As the reference trigger already collects a significant fraction of the signal, for each possible L1T selection criteria the additional signal efficiency to L1\_ETM70 is calculated. Table 8.1 shows the best results obtained by the automatic procedure, ordered in descending value of additional efficiency.

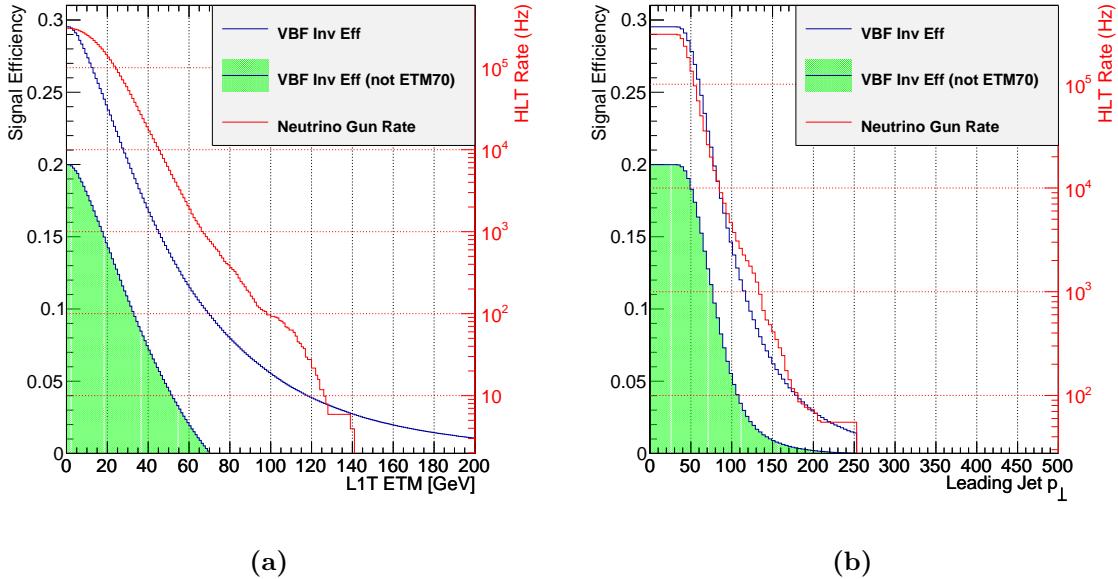
Event Selection Criteria			L1T	Signal Efficiency [%]	
Base	Additional	Value [GeV]	Rate [Hz]	Individual	Additional
Dijet VBF + $p_T^{jets} > 30 + \Delta\eta > 3.5$	Lead Jet $p_T$	97	4632	14.6	5.5
Dijet VBF + $p_T^{jets} > 40 + \Delta\eta > 3.5$	Lead Jet $p_T$	97	4356	13.5	5.2
Dijet VBF + $p_T^{jets} > 30 + \Delta\eta > 3.5$	ETM	51	4961	13.6	4.0
Dijet VBF + $p_T^{jets} > 30 + \Delta\eta > 3.0$	ETM	56	4890	17.6	3.9
Dijet VBF + $p_T^{jets} > 40 + \Delta\eta > 3.0$	Dijet $M_{jj}$	1760	4991	6.5	3.7
Dijet VBF + $p_T^{jets} > 40 + \Delta\eta > 3.5$	ETM	51	4482	12.4	3.7
Dijet VBF + $p_T^{jets} > 40 + \Delta\eta > 3.5$	MHT	47	4963	12.5	3.7
Dijet VBF + $p_T^{jets} > 40 + \Delta\eta > 3.5$	Dijet $M_{jj}$	1760	4991	6.5	3.7
Dijet VBF + $p_T^{jets} > 40 + \Delta\eta > 3.0$	ETM	56	4589	16.4	3.6
Dijet VBF + $p_T^{jets} > 30 + \Delta\eta > 3.5$	MHT	49	4518	13.1	3.6

**Table 8.1:** Results of the search for L1T algorithms with a maximum rate of 5 kHz for the *TSG high luminosity scenario*. Base criteria is fixed while an additional variable is scanned and its value is set by the allowed maximum rate. Results are presented in descending order of additional signal collection efficiency relative to L1\_ETM70, while individual efficiency is the sole algorithm signal efficiency.

The obtained results are surprising as the highest additional efficiency algorithm found does not include any L1T MET requirement. Instead, it requires that the leading L1T jet in the event is at least 97 GeV. As expected, by adding a dijet requirement the MET requirement could be reduced to about 50 GeV. Both a MHT and dijet  $M_{jj}$  requirements perform significantly worse than the single jet and MET options. The two following criteria were selected after rounding to the closest possible L1T thresholds for further studies:

- Dijet VBF +  $p_T^{jets} > 30 \text{ GeV} + \Delta\eta > 3.5 + ETM \geq 50 \text{ GeV}$
- Dijet VBF +  $p_T^{jets} > 30 \text{ GeV} + \Delta\eta > 3.5 + \text{Single Jet } p_T > 96 \text{ GeV}$

For both of these algorithms the plots resulting from the scan over the additional variable can be found in figure 8.1.



**Figure 8.1:** Plots produced by the optimization process of possible L1T algorithms with a maximum rate of 5 kHz for the *TSG high luminosity scenario*. Both figures require events with at least one dijet in opposite sides of the detector passing  $p_T^{jets} > 30$  and  $\Delta\eta > 3.5$ . Figure (a) shows the scan over L1T MET while figure (b) shows the scan over leading jet  $p_T$ . The red line is the estimated rate in Hz, the blue line the fraction of accepted signal and the green shaded area the additional efficiency relative to L1.ETM70.

### 8.1.3 HLT algorithm development

In the HLT trigger menu, the reference trigger was selected as the lowest unprescaled PF MET trigger available which was `HLT_PFMET170_NoiseCleaned`. This algorithm is seeded by `L1_ETM70` and selects events with HLT PF MET over 170 GeV, which is calculated in the full  $\eta$  coverage of the detector. This trigger was calculated to have an expected HLT pure rate of  $\approx 4.5$  Hz and a signal collection efficiency of 9.4%, while the Run I prompt trigger efficiency was  $\approx 7.6\%$  and the parked trigger between  $\approx 10.0\%$  and  $\approx 12.9\%$  depending on the acquisition era.

Similarly to the L1T search algorithm, an automatic approach was developed to obtain the best possible algorithm thresholds. In this case, the only variable scanned to fulfil the desired rate algorithm was HLT PF MET. A *grid search* method was implemented, where

all possible base configurations of thresholds were tested. For each set of thresholds, the signal efficiency, selection rates, and additional efficiency to reference HLT path were calculated.

Events were selected when at least one HLT dijet was found, where both jets were on opposite sides of the detector (VBF condition) passing all the requirements of the base selection and the additional PF MET minimum. The following base configuration variables and values were tested:

- Symmetric dijet  $p_T^{jets} > 40, 50, 60$
- Asymmetric dijet  $p_T^{jet_1}, p_T^{jet_2} > (50, 40), (60, 40), (70, 40), (80, 40), (90, 40), (100, 40)$
- Dijet  $M_{jj} > 500, 600, 700, 800, 900, 1000, 1100$
- Dijet  $\Delta\eta > 3.5, 3.7, 3.9, 4.1, 4.3, 4.5$

The PF MET minimum threshold was optimised for the signal HLT path considering a conservative low target rate of  $\approx 1.5$  Hz and an aggressive high target rate of 5.0 Hz. The assumption was that the high target rate trigger would be the base of the proposal to the TSG group, but if there was limited available bandwidth in the menu we would have the fallback conservative trigger proposal. For the sake of brevity, and as the high rate scenario was accepted by the TSG group only these results are shown in the following section.

### Signal path with L1T seed L1\_ETM70

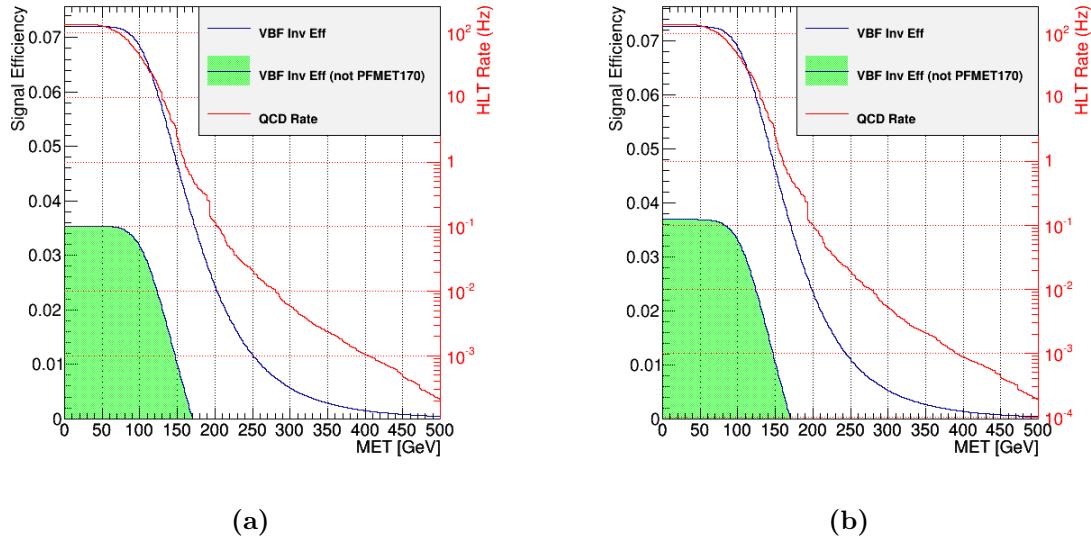
The baseline HLT signal path was optimized for events passing the already available L1T reference algorithm which was L1\_ETM70. All the obtained paths for a 5 Hz maximum HLT rate have significantly lower signal collection efficiency than the reference HLT path. Table 8.2 shows the best algorithm thresholds for the maximum total signal efficiency, maximum additional signal efficiency and lowest lowest PF MET threshold. For each category of results, the best dijet symmetric and asymmetric  $p_T$  thresholds results are presented.

The best algorithms for both maximum total and additional efficiencies are asymmetric. However, the difference to the best symmetric algorithms is small and comes mostly at a cost of an increased PF MET requirement or by lowering  $M_{jj}$ , while increasing the lead jet  $p_T$ . The algorithms that minimize PF MET, have only about  $\approx 1\%$  signal

Algorithm	Event Requirements					Rate	Signal Efficiency	
	Type	$p_T^{jet_1}, p_T^{jet_2}$ [GeV]	VBF	$\Delta\eta$	$M_{jj}$ [GeV]	MET [GeV]	HLT [Hz]	Total [%]
Maximum Total Signal Efficiency								
Asymmetric	70,40	Yes	3.5	500	144	4.99	5.18	1.37
Symmetric	40,40	Yes	3.5	600	140	4.68	5.16	1.49
Maximum Additional Signal Efficiency								
Asymmetric	60,40	Yes	3.7	500	140	4.84	5.13	1.55
Symmetric	40,40	Yes	3.5	600	140	4.68	5.16	1.49
Lowest PF MET Threshold								
Symmetric	60,60	Yes	4.1	800	119	4.99	< 3%	1.04
Asymmetric	100,40	Yes	4.3	1000	122	4.94	< 3%	1.01

**Table 8.2:** Results of the automatic optimization of possible HLT paths for a maximum rate of 5 Hz for the *TSG high luminosity scenario*. All HLT algorithms are seeded by L1\_ETM70. Results are presented for the best dijet symmetric and asymmetric  $p_T$  thresholds, for maximum total signal efficiency, maximum additional signal efficiency to HLT\_PFMET170\_NoiseCleaned, and lowest PF MET.

efficiency but reduce that threshold by  $\approx 50$  GeV when compared with the reference HLT path. Plots obtained during the scan of PF MET for the two maximum additional signal efficiency algorithms can be found in figure 8.2.



**Figure 8.2:** Plots showing the scan over HLT PF MET for different algorithm base selection in the *TSG high luminosity scenario*. All HLT algorithms are seeded by L1\_ETM70. Figure (a) base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet_1}, p_T^{jet_2} > 40, 40$  GeV,  $\Delta\eta > 3.5$  and  $M_{jj} > 600$  GeV while figure (b) base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet_1}, p_T^{jet_2} > 60, 40$  GeV,  $\Delta\eta > 3.7$  and  $M_{jj} > 500$  GeV.

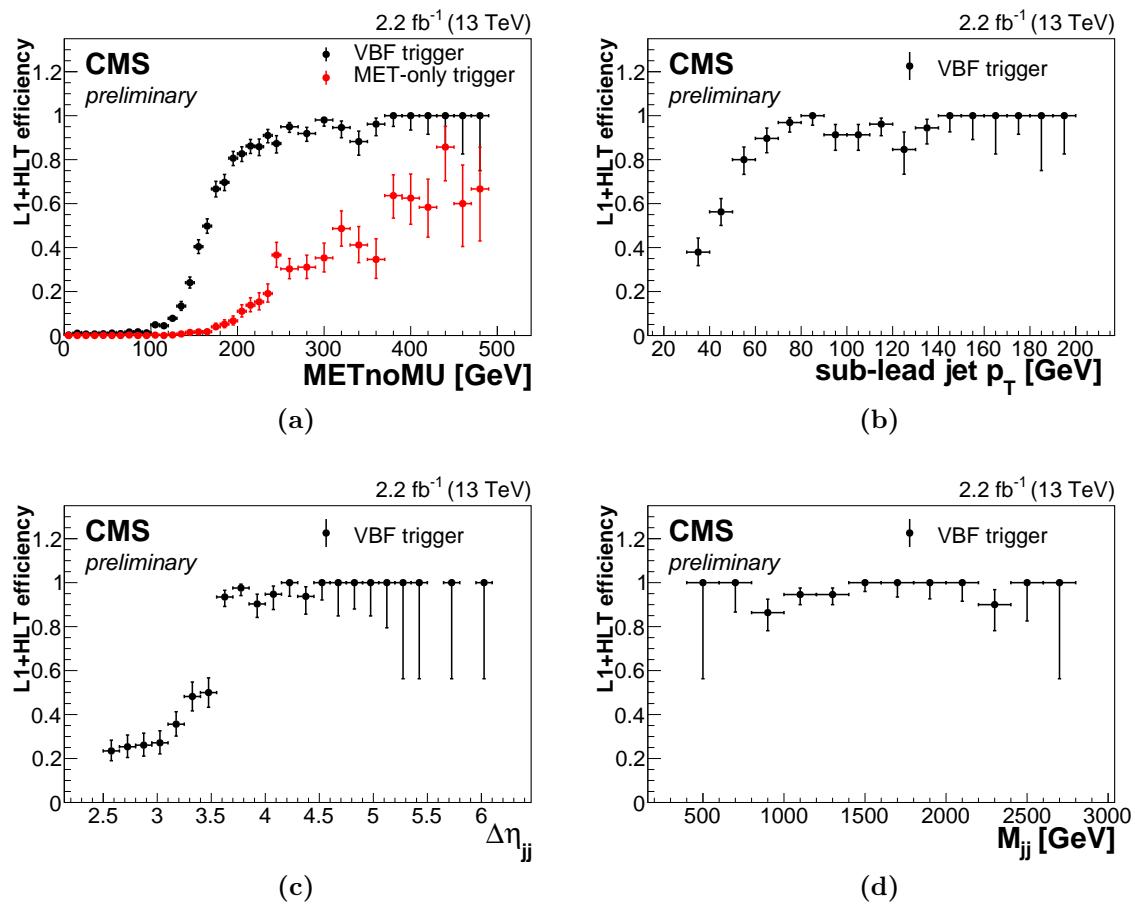
Since the difference was small between these two trigger paths, for simplicity the symmetric path was chosen. This dedicated trigger, combined with the HLT reference trigger, records 10.9% of the simulated signal process, which corresponds to an increase of signal collection efficiency of 15.8% when compared with just the reference trigger. Due to the lack of time and manpower to study and implement a new L1T algorithm, this was the proposed solution for data taking during 2015. This proposal was accepted by the TSG and the resulting HLT trigger path used PF MET<sub>no- $\mu$</sub> , which does not increase the rate significantly. It was integrated into the standard CMS trigger menu and was used to record data during the full 2015 Run II campaign.

### VBF Higgs to invisible trigger turn ons

In the end of the 2015 data taking run, trigger analysis was performed with the full luminosity recorded with 8 TeV. To study the trigger efficiency in an unbiased way events were used that were recorded through a single muon trigger. The VBF Higgs to invisible dedicated HLT trigger has requirements in four variables, MET<sub>no- $\mu$</sub> , dijet  $p_T^{jets}$ , dijet  $\Delta\eta$  and dijet  $M_{jj}$ . To study the trigger efficiency in selecting events depending on the offline values of these variables (trigger turn ons) they must be analysed individually.

To decouple the dependency between them, we require all except the one being analysed to be above a threshold where in that specific variable the trigger is 100% efficient. The offline selection used requires events with lead dijet  $p_T^{jets} > 80$  GeV, dijet  $M_{jj} > 600$  GeV, dijet  $\Delta\eta_{jj} > 3.6$  and MET<sub>no- $\mu$</sub>   $> 300$  GeV. When analysing a variable turn-on, the corresponding event selection requirement is removed. Figure 8.3 shows the trigger turn ons for all variables used on the trigger as a function of their offline counterparts.

It can be seen that the dedicated trigger performs significantly better than the MET only trigger which selects events with PF MET  $> 170$  GeV, when considering the MET<sub>no- $\mu$</sub>  turn-on. In this variable, 90% efficiency is reached at around 240 GeV, for the dijet sub-leading jet  $p_T$  around 70 GeV, for dijet  $\Delta\eta_{jj}$  around 3.7. For dijet  $M_{jj}$  a precise turn-on cannot be described with the current statistics, but it appears to sharply rise to 100% just after the trigger requirement of 600 GeV. With these results we can conclude that the trigger is performing as expected and is significantly better than the MET-only trigger, considering the analysis needs.



**Figure 8.3:** Trigger turn ons for the VBF Higgs to invisible dedicated trigger for (a)  $\text{MET}_{no-\mu}$  (includes comparison with the trigger menu MET only trigger), (b) dijet sub-lead jet  $p_T$ , (c) dijet  $\Delta\eta_{jj}$  and (d) dijet  $M_{jj}$  [149].

### Signal path with L1T seed Dijet + MET

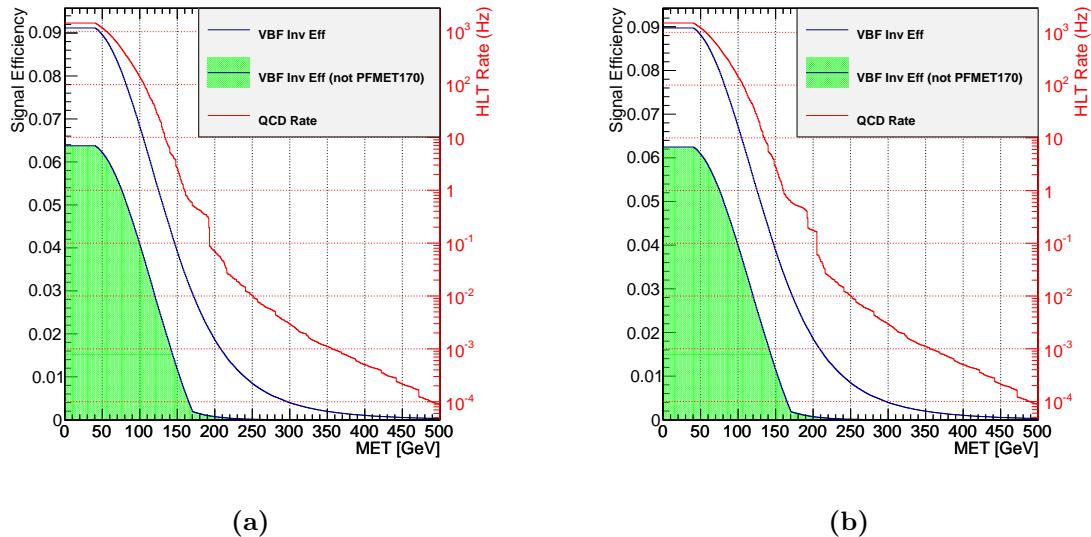
Optimization can also be done for the two new proposed L1T seeds. Events are selected with a L1T dijet with its jets in opposite sides of the detector where  $p_T^{jets} > 30 \text{ GeV}$ ,  $\Delta\eta > 3.5$  and L1T  $ETM \geq 50 \text{ GeV}$ . The same procedure was applied as described in previous section. Results for the best algorithm threshold combinations for maximum total signal efficiency and lowest PF MET threshold can be found in table 8.3. Again, for each category of results, the best dijet symmetric and asymmetric  $p_T$  thresholds results are presented.

Algorithm	Event Requirements					Rate	Signal Efficiency	
	Type	$p_T^{jet_1}, p_T^{jet_2} [\text{GeV}]$	VBF	$\Delta\eta$	$M_{jj} [\text{GeV}]$		HLT [Hz]	Total [%]
Maximum efficiency for jets								
Asymmetric	60,40	Yes	3.7	500	144	4.76	4.59	1.57
Symmetric	40,40	Yes	3.5	600	145	4.69	4.48	1.52
Maximum Additional Efficiency								
Asymmetric	60,40	Yes	4.1	500	140	4.77	4.39	1.71
Symmetric	40,40	Yes	4.1	600	141	4.95	4.23	1.62
Lowest MET Threshold								
Symmetric	60,60	Yes	4.5	1000	122	4.93	2.28	1.02
Asymmetric	100,40	Yes	4.5	1100	125	4.87	2.64	1.00

**Table 8.3:** Results of the automatic optimization of possible HLT paths for a maximum rate of 5 Hz for the *TSG high luminosity scenario*. All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF,  $p_T^{jets} > 30 \text{ GeV}$ , and  $\Delta\eta > 3.5$  and  $ETM \geq 50 \text{ GeV}$ . Results are presented for the best dijet symmetric and asymmetric  $p_T$  thresholds, for maximum total signal efficiency, maximum additional signal efficiency to HLT\_PFMET170\_NoiseCleaned, and lowest PF MET.

As expected lowering the L1T MET threshold allows a bigger additional signal efficiency to be achieved. The best algorithms in this benchmark quantity, similarly to the previous study, also requires HLT MET  $> 140 \text{ GeV}$  and similar dijet thresholds, implying the added efficiency comes from recovering events that fail L1\_ETM70. It is also interesting that the total efficiency is lower than seen in table 8.2, which could be caused by the additional L1T jet requirements. Plots of the two best additional signal efficiency algorithms scans of PF MET can be found in figure 8.4.

The tail of additional efficiency can be seen continuing above PF MET  $> 170 \text{ GeV}$ , further supporting the hypothesis that efficiency is being recovered from events lost by the L1\_ETM70 requirement. The best additional signal efficiency algorithm optimized for



**Figure 8.4:** Plots showing the scan over HLT PF MET for different algorithm base selection in the *TSG high luminosity scenario*. All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF,  $p_T^{jets} > 30$  GeV, and  $\Delta\eta > 3.5$  and  $ETM \geq 50$  GeV. Figure (a) base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet1}, p_T^{jet2} > 60, 40$  GeV,  $\Delta\eta > 4.1$  and  $M_{jj} > 500$  GeV while figure (b) base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet1}, p_T^{jet2} > 40, 40$  GeV,  $\Delta\eta > 4.1$  and  $M_{jj} > 600$  GeV.

this L1T dijet plus MET seed combined with the HLT reference trigger records 11.11% of the simulated signal process. This dedicated HLT trigger, when compared with the proposed dedicated trigger for Run II, gives 14.8% additional efficiency with an increase of 2.0% in total signal efficiency.

### Signal path with L1T seed Dijet + Single Jet

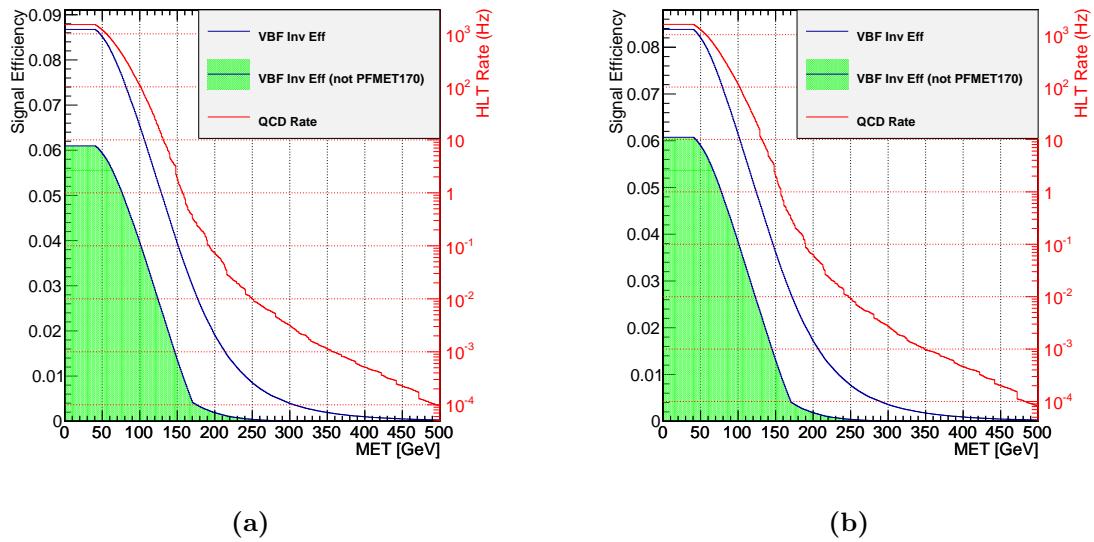
Finally, optimization of HLT algorithms was also preformed over a possible L1T seed without any MET requirement. This seed selected events with a L1T dijet with its jets in opposite sides of the detector, where  $p_T^{jets} > 30 \text{ GeV}$ ,  $\Delta\eta > 3.5$  and the highest  $p_T$  L1T jet in the event has at least 96 GeV. This high  $p_T$  jet can and should be in the selected dijet but for simplicity of algorithm design both in hardware and software coupled with development time constraints, forced these conditions to be kept separated. The same procedure from previous section is applied and the results can be found in table 8.4.

Algorithm Type	Event Requirements					Rate HLT [Hz]	Signal Efficiency	
	$p_T^{jet_1}, p_T^{jet_2} [\text{GeV}]$	VBF	$\Delta\eta$	$M_{jj} [\text{GeV}]$	MET [GeV]		Total [%]	Additional [%]
Maximum efficiency								
Symmetric	40,40	Yes	3.5	500	148	4.93	4.79	1.75
Asymmetric	50,40	Yes	3.5	500	148	4.92	4.78	1.74
Maximum Additional Efficiency								
Asymmetric	90,40	Yes	4.1	500	140	4.51	4.44	1.86
Symmetric	40,40	Yes	4.3	800	140	4.81	4.09	1.78
Lowest MET Threshold								
Symmetric	60,60	Yes	4.3	1100	123	4.89	2.53	1.20
Asymmetric	100,40	Yes	4.3	1100	128	4.98	3.18	1.42

**Table 8.4:** Results of the automatic optimization of possible HLT paths for a maximum rate of 5 Hz for the *TSG high luminosity scenario*. All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF,  $p_T^{jets} > 30 \text{ GeV}$ , and  $\Delta\eta > 3.5$  and a single jet  $p_T^{jets} > 96 \text{ GeV}$ . Results are presented for the best dijet symmetric and asymmetric  $p_T$  thresholds, for maximum total signal efficiency, maximum additional signal efficiency to HLT\_PFMET170\_NoiseCleaned, and lowest PF MET.

The best additional signal efficiency HLT trigger algorithm found was highly asymmetric, as expected. Surprisingly, this is the best additional efficiency algorithm obtained in all studies. Once again, the determined PF MET was 140 GeV, suggesting that the recovered efficiency comes from the absence of an L1T MET restriction, combined with the asymmetric topological requirements. The total efficiency is also below the algorithms

based on the L1\_ETM70 seed. Since the best additional efficiency configurations are selected, this implies that phase space lost by reference algorithm is being recovered at the cost of total efficiency. Plots of the two best additional signal efficiency algorithms PF MET scans can be found in figure 8.5.



**Figure 8.5:** Plots showing the scan over HLT PF MET for different algorithm base selection in the *TSG high luminosity scenario*. All HLT algorithms are seeded by proposed L1T algorithm selecting a dijet passing requirements VBF,  $p_T^{jets} > 30$  GeV, and a single jet  $p_T^{jets} > 96$  GeV. Figure (a) base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet1}, p_T^{jet2} > 90, 40$  GeV,  $\Delta\eta > 4.1$  and  $M_{jj} > 500$  GeV while figure (b) base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet1}, p_T^{jet2} > 40, 40$  GeV,  $\Delta\eta > 4.3$  and  $M_{jj} > 800$  GeV.

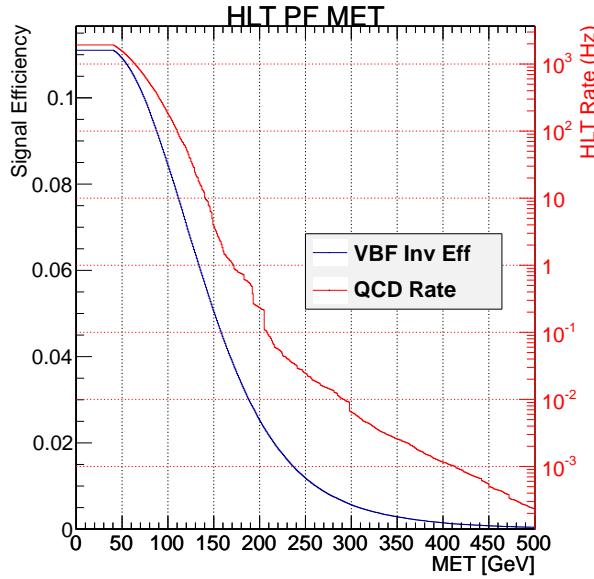
The best additional signal efficiency algorithm optimized for this L1T dijet plus single jet seed combined with the HLT reference trigger records 11.26% of the simulated signal process. This dedicated HLT trigger, when compared with the proposed dedicated trigger for Run II, selects 19.8% additional efficiency with an increase of 3.4% in total signal efficiency. This trigger configuration most interesting feature is that it avoids completely the slow turn on of L1T MET which may allow lower offline MET selection threshold.

### Systematics HLT trigger development

One of the main systematics for the Run II analysis was the lack of statistics on the background control regions. In an attempt to solve this problem in Run II, it was seen as desirable to design a trigger algorithm with lower thresholds to study these regions.

To allow the lowering of thresholds the trigger has to be prescaled. The target rate is for the low bandwidth scenario 0.1 Hz and 0.5 Hz for the high bandwidth scenario.

Considering the chosen dedicated HLT trigger path presented in section 8.1.3, it was decided to make a background path with the lowest possible MET requirement both at L1T and HLT. The L1T algorithm chosen to seed this trigger path was L1\_ETM50 which in the TSG proposed menu is prescaled by 1000. The same automatic procedure used for the signal path studies was used here again. The maximum rate for this seed before any prescale applied was 500 Hz for the high bandwidth scenario of an HLT output rate of 0.5 Hz. The same base dijet configuration as the dedicated signal path was used, selecting events with a dijet on opposite sides of the detector passing requirements:  $p_T^{jet_1}, p_T^{jet_2} > 40, 40 \text{ GeV}$ ,  $\Delta\eta > 3.5$  and  $M_{jj} > 600 \text{ GeV}$ . The obtained HLT PF MET threshold was 80 GeV for an unprescaled rate of 505.75 Hz. The final predicted rate for an HLT prescale of 1 was 0.5 Hz and 0.1 Hz for an HLT prescale of 5, as required by the study targets for each scenario. A plot of the scan over HLT PF MET results from the optimization procedure can be found in figure 8.6.



**Figure 8.6:** Results of the automatic optimization of possible HLT path for a maximum rate of 500 Hz for the *TSG high luminosity scenario*. The HLT algorithm is seeded by L1\_ETM50. Figure shows the scan over HLT PF MET for base configuration requiring a dijet on opposite sides of the detector,  $p_T^{jet_1}, p_T^{jet_2} > 40, 40$ ,  $\Delta\eta > 3.5$  and  $M_{jj} > 600 \text{ GeV}$

It was decided to propose to the TSG the lowest rate background path with prescale 1000 at L1T and 5 at HLT. This proposal was accepted and a version of this trigger using

PF MET<sub>no- $\mu$</sub>  was implemented and integrated into the standard CMS trigger menu and was used to record data during the full 2015 Run II campaign.

### 8.1.4 Summary

A complete study was preformed to define trigger conditions to record data during the LHC Run II to be used by the VBF Higgs to invisible analysis. Two HLT triggers were proposed and accepted by the TSG and included in the 2015 trigger menus.

Considering a SM VBF Higgs to Invisible process with  $m_H = 125 \text{ GeV}$ , the signal is gathered by a combination of a general purpose trigger path, HLT\_PFMET170\_NoiseCleaned and a dedicated trigger. The dedicated trigger is seeded by L1\_ETM70 and selects events with at least one dijet on opposite sides of the detector passing requirements  $p_T^{jets} > 40 \text{ GeV}$ ,  $\Delta\eta > 3.5$ ,  $M_{jj} > 600 \text{ GeV}$  and PF MET<sub>no- $\mu$</sub>   $> 140 \text{ GeV}$ . Together these triggers collect 10.9% of the simulated signal process where the dedicated path corresponds to an increase of signal collection efficiency of 15.8% over the MET only trigger. For the *TSG high luminosity scenario* the dedicated trigger is expected to have a pure rate under 5.0 Hz. Turn on curves has been produced for this trigger with the full luminosity recorded during 2015, showing it is operating as expected.

A new background trigger path was developed, it is seeded by L1\_ETM50 and selects events with at least one dijet on opposite sides of the detector passing requirements  $p_T^{jets} > 40 \text{ GeV}$ ,  $\Delta\eta > 3.5$ , and  $M_{jj} > 600 \text{ GeV}$  and PF MET<sub>no- $\mu$</sub>   $> 80 \text{ GeV}$ . For the *TSG high luminosity scenario* an HLT pure rate of  $\approx 0.1 \text{ Hz}$  is expected when applying a L1T prescale of 1000 and 5 at HLT.

Additionally, a study of new dedicated L1T algorithms was preformed with the HLT algorithms optimised for the highest additional rate when compared with the baseline triggers. It was demonstrated that the dedicated trigger additionally efficiency can be increased by 14.8% by using a dedicated L1T algorithm selecting events with a dijet plus MET and by as much as 19.8% when selecting events with a dijet plus a hard single jet. The latter would have the advantage of avoiding the slow turn-on the L1T MET. Both these options are being considered for the 2016 Run II campaign.

## 8.2 Run II QCD Monte Carlo samples

During the preparation of the Run I VBF Higgs to invisible analysis, a set of QCD samples with VBF-like jets and real MET were produced, described in section 6.2. These samples enabled the understanding of the mechanisms that create real MET in QCD and how those could be mitigated. However, events with fake MET were found to be the dominating type of QCD multi-jet events passing the analysis selection criteria.

In the preparation for Run II, it was considered once again to be useful to have similar samples remade and possibly extended. It was identified that not only real MET is significant but also fake MET coming from detector mis-measurement. The QCD background is currently the only background without any adequate representative MC event sample. If such a sample could be produced, it could allow the analysis to evolve into a shape-based analysis or the use of machine learning techniques, since full description of the signal and backgrounds would be possible with simulations.

### 8.2.1 Goals and first attempt

Building on the knowledge gained from the samples produced during Run I, we can define the goals for these new samples. Cuts at generator level involving MET should be avoided in order to not filter out events where the MET comes from mis-measurement. Variables that may bias the  $\Delta\phi(jet - jet)$  distribution should also be avoided, as the Run I analysis uses inverted cuts in this variable to perform data-driven QCD estimation. All cuts should be below the event selections used during Run I and, if possible, around or even below the Run II trigger conditions. The sample to be simulated should be equivalent to at least  $1\text{ fb}^{-1}$  of data but of a size comparable with the current official QCD Inclusive sample. This last requirement is to ensure that the computing resources necessary for making such sample do not go above those currently used to produce similar purpose samples.

The first attempt to produce a proposal for the production of this QCD VBF-like sample was based on filtering events produced by PYTHIA 8. This filtering was done by first clustering the generator particles in anti- $k_T$  jets with  $R = 0.4$ , where muons were ignored. Only events with at least one dijet with VBF characteristics would be kept. Unfortunately, this approach leads to a very large number of event being generated (hard scattering and hadronization) and clustered, only to be discarded. The computing time was considered too large to be feasible considering the physics case by the CMS

team responsible for official sample production. However, it was recommended to take a different approach by using a ME generator, like MADGRAPH, and to cut at the parton level, before any hadronization or clustering. After this initial event selection, a second layer of cuts could be applied after hadronization to ensure the actual outgoing jets would pass offline selection criteria. Furthermore, using a ME generator should provide a more accurate description of multi-jet events, while the two-step approach should allow a significant reduction of the necessary computing time.

### 8.2.2 MADGRAPH parton level simulation

The MADGRAPH event generator was selected to produce the parton level simulation. With this generator it is possible to simulate events from the interaction of two proton partons and to obtain a final state with any number of partons. Each additional parton on the final state comes at the cost of an exponential increase of the possible diagrams, which in turn means more time is necessary to create events. It was chosen to only produce final states with 2, 3 and 4 outgoing partons. This generator has been used to create similar QCD samples used by some CMS SUSY analyses.

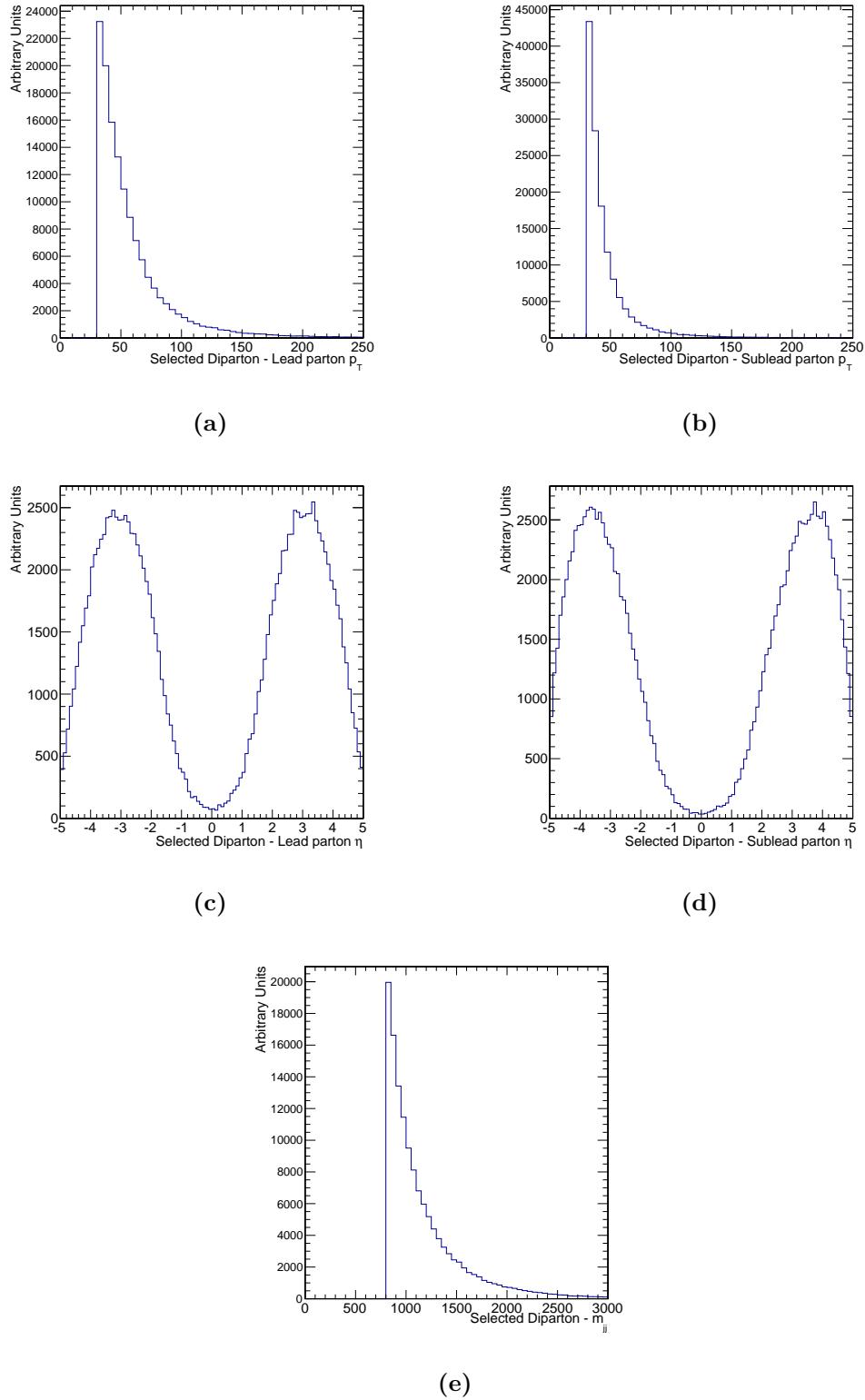
The outgoing partons are defined to be a gluon or a quark (u, d, c, s or b). We do not allow diagrams with top quarks as they do not hadronize and lead to event topologies which are already accounted by other MC samples. The outgoing partons are the seeds of our final state jets.

A custom parton level filter was implemented inside the MADGRAPH code to select events with VBF characteristics. To pass the filter, the event must have at least one outgoing di-parton with invariant mass of 800 GeV, where both parton are inside the detector volume with  $|\eta| < 5.0$  and have  $p_T > 30$  GeV. The distributions of these variables for events passing this cuts can be found in figure 8.7.

The estimated cross section for these processes and selection is  $(1.029 \pm 0.002) \times 10^7$  pb and we request the production of  $1.2 \times 10^{10}$  events. That corresponds to just over  $1.1 \text{ fb}^{-1}$  of equivalent integrated luminosity.

### 8.2.3 Hadronization with PYTHIA 8

The parton level events that have passed the initial filter are then hadronized. Similarly to other samples produced in CMS, PYTHIA 8 was chosen for this. As described in



**Figure 8.7:** Parton  $p_T$ ,  $\eta$  and di-parton  $m_{jj}$  distributions for the leading di-parton passing cuts: parton  $p_T > 30$  GeV and  $|\eta| < 5.0$ , di-parton  $M_{jj} > 800$  GeV.

section 3.9, when using a ME generator with a shower generator we need to filter the overlapping phase-space. As recommended by the CMS generator group we used the MLM scheme [121] with the same parameters used for the production of previous official samples. The results of the hadronization process are summarized in table 8.5.

Process	Events			Cross Section [pb]	
	Tried	Passed	accepted [%]	Before	After
$pp \rightarrow jj$	231789	54291	$23.4 \pm 1.01$	$1.675 \times 10^6 \pm 4.536 \times 10^3$	$3.924 \times 10^5 \pm 1.817 \times 10^3$
$pp \rightarrow jjj$	502287	36250	$7.2 \pm 0.03$	$3.622 \times 10^6 \pm 9.809 \times 10^3$	$2.614 \times 10^5 \pm 1.500 \times 10^3$
$pp \rightarrow jjjj$	692600	44299	$6.4 \pm 0.03$	$4.972 \times 10^6 \pm 1.346 \times 10^4$	$3.180 \times 10^5 \pm 1.697 \times 10^3$
Total	1426676	134840	$9.45 \pm 0.03$	$1.027 \times 10^7 \pm 1.727 \times 10^4$	$9.718 \times 10^5 \pm 2.903 \times 10^3$

**Table 8.5:** Summary of the results of the Hadronization with Pythia 8 of 1.4M MadGraph events passing the parton level filter.

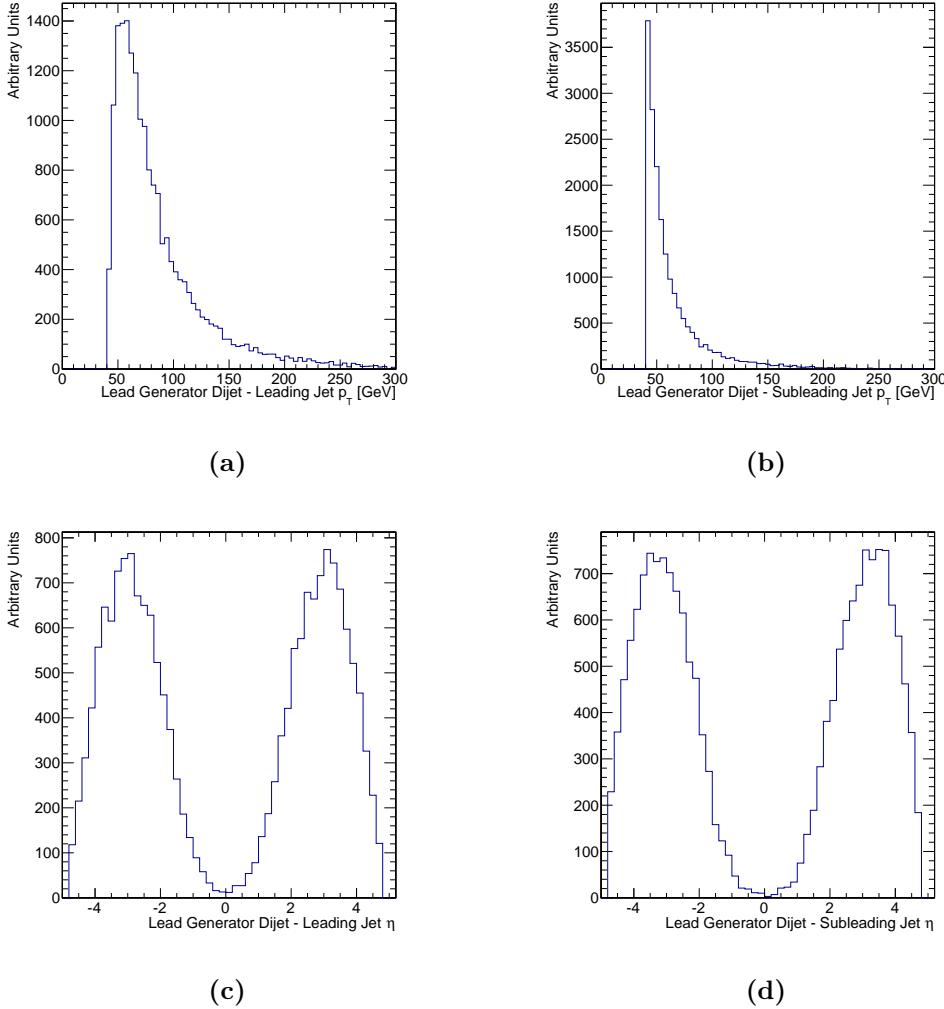
The efficiency of the post-hadronization event matching has been estimated to be  $9.45\% \pm 0.03\%$ , leading to a sample cross section of  $(9.718 \pm 0.03) \times 10^5$  pb. The lower matching efficiency in the 3 and 4 jets final states is due to the absence of a restriction on minimum jet  $p_T$  on any additional jets to the dijet passing the parton level cuts. These additional generator jets, if low enough in energy, will hardly be picked up by the clustering algorithm and therefore cannot be matched to their seed parton.

### 8.2.4 Generator level cuts

After hadronization we cluster the outgoing stable particles with the anti- $k_T$  algorithm with  $R = 0.4$  while ignoring muons. The reason for ignoring muons is that CMS muon detector coverage only goes up to  $|\eta| < 2.4$ , so all muons outside this region will not be seen by the experiment and therefore will not be clustered into jets. Most of our signal like events will have at least one jet in the region  $|\eta| > 2.4$ . For the jets inside the muon systems acceptance, the jet  $p_T$  will be underestimated when a muon is present, but this effect was assumed to be small.

We start by making an initial selection of the events with at least one generator level dijet passing  $\Delta\eta > 3.0$ ,  $M_{jj} > 1000$  GeV where both jets pass  $p_T > 40$  GeV and  $|\eta| < 4.8$ , this selection is thresholds are below all offline requirements of the parked data analysis. The events passing this selection are split into two sub-samples. Sub-sample A has events where the selected dijet passes  $\Delta\phi \leq 2.15$  and sub-sample B where at least one dijet passes all initial conditions and an inverted  $\Delta\phi$  requirement. Plots over all the relevant

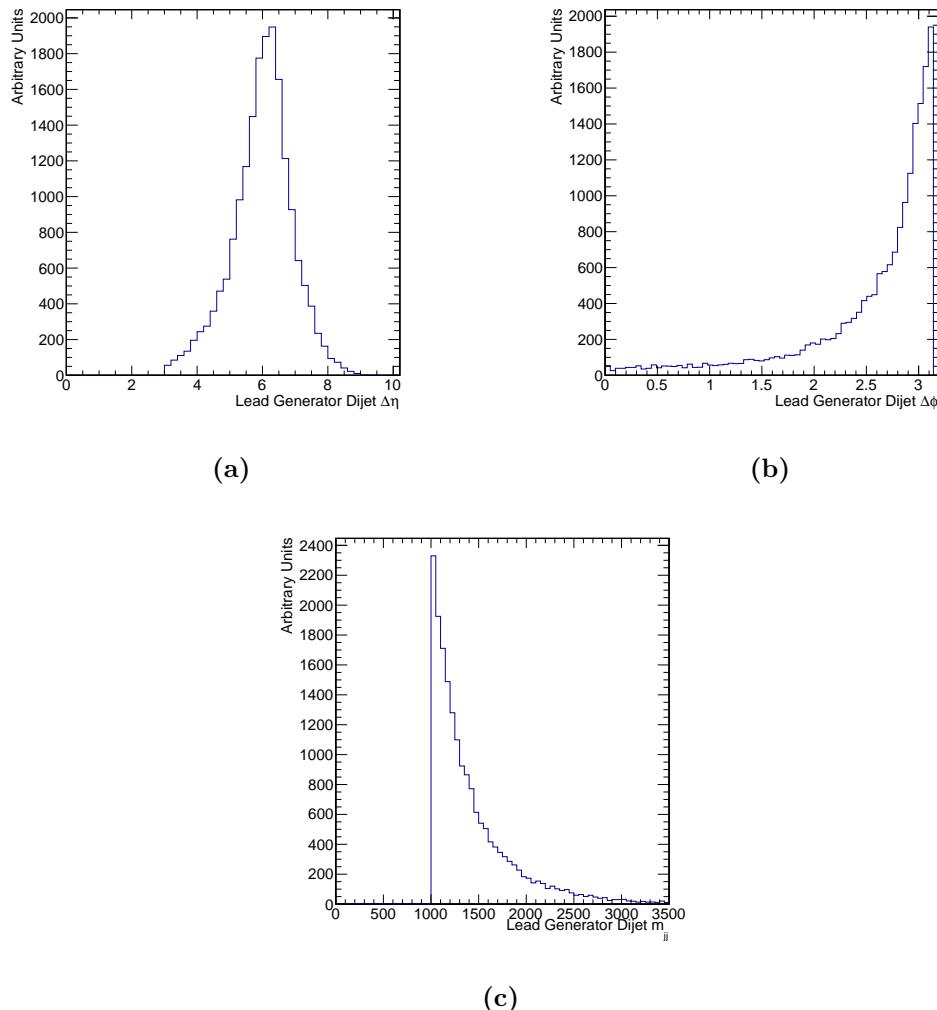
variable before the  $\Delta\phi$  requirement and for the leading dijet passing the cuts can be found in figure 8.8 and 8.9.



**Figure 8.8:** Relevant distributions for the two jets comprising the the leading dijet passing a generator filter requiring at least one dijet with  $\Delta\eta > 3.0$  and  $M_{jj} > 1000$  GeV where the jets have  $p_T > 40$  GeV and  $|\eta| < 4.8$

All the distributions show the expected features of the generator level filter cuts. As expected the peak of the  $\Delta\phi$  distribution is at  $\pi$  when the 2 jets are back to back, but a tail of events is visible down to zero.

Sub-sample A will be produced by running over all the events produced up to the hadronization step. Its estimated filter efficiency is  $(2.938 \pm 0.005) \times 10^{-1}$  which would lead to a sample size of approximately 29 million events and corresponding to an equivalent luminosity of over  $1.1 \text{ fb}^{-1}$ .



**Figure 8.9:** Relevant distributions for the leading dijet passing a generator filter requiring at least one dijet with  $\Delta\eta > 3.0$  and  $M_{jj} > 1000$  GeV where the jets have  $p_T > 40$  GeV and  $|\eta| < 4.8$

Sub-sample B will result from running over only 10% of the events available at the hadronization step. This filter has an efficiency  $(1.125 \pm 0.009) \times 10^{-1}$  and would lead to sample of about 14 million events, corresponding to an equivalent luminosity of over 110 pb. If additional computing resources would become available this sample could be expanded up to 100% of the base sample to a total of 141 million events and equivalent luminosity over  $1.1 \text{ fb}^{-1}$ .

It is necessary to determine the overlap between these two sub-samples. From table 8.6, we can see that a significant number of events of each sub-sample have additional jets, passing all required jet conditions.

$N_{Jets}$	$\Delta\phi$ cut		
	no cut	$< 2.15$	$\gtrsim 2.15$
2	$63.83 \pm 0.59$	$15.80 \pm 0.63$	$73.38 \pm 0.70$
3	$23.53 \pm 0.36$	$50.21 \pm 1.13$	$17.59 \pm 0.34$
4	$9.43 \pm 0.23$	$24.34 \pm 0.78$	$6.70 \pm 0.21$
5	$2.42 \pm 0.11$	$7.14 \pm 0.42$	$1.70 \pm 0.11$
+6	$0.79 \pm 0.07$	$2.50 \pm 0.25$	$0.63 \pm 0.06$

**Table 8.6:** Table showing the percentage of generator anti- $k_T$  jets with  $R = 0.4$  passing cuts  $p_T > 40 \text{ GeV}$  and  $|\eta| < 4.8$  for events with at least one dijet with  $\Delta\eta < 3.0$  and  $m_{jj} < 1000 \text{ GeV}$  and according to an additional dijet  $\Delta\phi$  cut.

These additional jets lead to additional combinations that may pass the criteria of the opposite sub-sample. As it can be seen in table 8.7 in as much as 5% of the events in the  $\Delta\phi \leq 2.15$  sub-sample there is a second combination of two jets that would pass the criteria to be in that sub-sample.

$N_{Dijets}$	$\Delta\phi$ cut		
	no cut	$< 2.15$	$\gtrsim 2.15$
1	$93.53 \pm 0.71$	$94.29 \pm 1.54$	$97.51 \pm 0.80$
2	$5.84 \pm 0.18$	$5.35 \pm 0.37$	$2.39 \pm 0.13$
3	$0.44 \pm 0.05$	$0.30 \pm 0.09$	$0.07 \pm 0.02$
+4	$0.19 \pm 0.03$	$0.05 \pm 0.04$	$0.03 \pm 0.01$

**Table 8.7:** Table showing the percentage of generator anti- $k_T$  dijets with  $R = 0.4$  passing cuts  $p_T^{jet} > 40 \text{ GeV}$ ,  $|\eta|^{jet} < 4.8$ ,  $\Delta\eta < 3.0$  and  $m_{jj} < 1000 \text{ GeV}$  and according to an additional dijet  $\Delta\phi$  cut.

The overlap between the two sub-samples has been determined to be  $3.95\% \pm 0.14\%$  of the events passing the initial selection. Since this number is relevant, and to avoid event double counting, events with combinations that would pass both sub-sample definitions should be selected into the smallest equivalent luminosity sub-sample.

### 8.2.5 Migration study

One concern when making cuts at steps below event reconstruction is the possibility of removing events that may pass analysis level event selections. This migration of events needs to be taken into account while defining the requirements at parton and generator levels. The signal region selection used during the 2012-13 parked data analysis selected events with a dijet passing  $\Delta\eta > 3.6$ ,  $M_{jj} > 1200$  GeV where the lead jet  $p_T > 50$  GeV and sub-lead jet  $p_T > 45$  GeV and both have  $|\eta| < 4.7$  (condition to guarantee the used AK5 jets are fully contained in the detector). It is unlikely that the Run II offline selection would be able to cut below jet  $p_T > 50$  GeV.

In order to study migration, a second MC sample with lower parton cuts was generated. MADGRAPH was used again to generate events with the same parameters with the only difference being the dijet cuts. Events are selected with at least a pair of outgoing partons with invariant mass of 600 GeV, where both partons are inside the detector volume with  $|\eta| < 5.0$  and  $p_T > 10$  GeV. Hadronization was then performed with the same procedure described in the previous section.

In order to compare generator jets to the partons that created them, matching is needed. For each parton, all generator jet, are selected which are located at  $\Delta R < 0.4$  and from those we select the generator jet closest in  $p_T$  to the selected parton. This procedure attempts to account for the situation where more than one jet is within the matching distance, but the best match in  $p_T$  is not the closest one in  $\Delta R$ . Using this procedure, a match for the di-parton passing the imposed cuts can be found for 73.24% of the events and the matched generator jet is not the closest one in  $\Delta R$  for 3.45% of the partons. A table of the matching efficiency discriminated by physics process can be found in table 8.8.

Partons are being simulated with fairly low  $p_T$ , two jets with  $p_T > 10$  GeV and up to two more with no restriction on energy. It is not a surprise that in significant number of events all partons cannot be matched to generator jets. This is due to the spread of energy over a larger area than the jet algorithm can cluster and due to the default AK4

$n_{match}$	Process			Total
	jj	jjj	jjjj	
0	22.04% $\pm$ 0.22%	2.18% $\pm$ 0.09%	0.14% $\pm$ 0.03%	11.62% $\pm$ 0.11%
1	38.60% $\pm$ 0.30%	17.82% $\pm$ 0.25%	3.02% $\pm$ 0.13%	25.27% $\pm$ 0.17%
2	39.35% $\pm$ 0.30%	42.35% $\pm$ 0.39%	16.91% $\pm$ 0.32%	35.99% $\pm$ 0.20%
3		37.65% $\pm$ 0.37%	41.88% $\pm$ 0.50%	19.83% $\pm$ 0.15%
4			38.05% $\pm$ 0.47%	7.29% $\pm$ 0.09%

**Table 8.8:** Table showing the percentage of partons successfully matched to a generator AK4 jets. Numbers obtained for a total of 88282 events over all 3 possible hard scattering processes and for events with at least one di-parton with  $m_{jj} > 600 \text{ GeV}$  where each parton has  $p_T < 10 \text{ GeV}$  and  $|\eta| < 5.0$

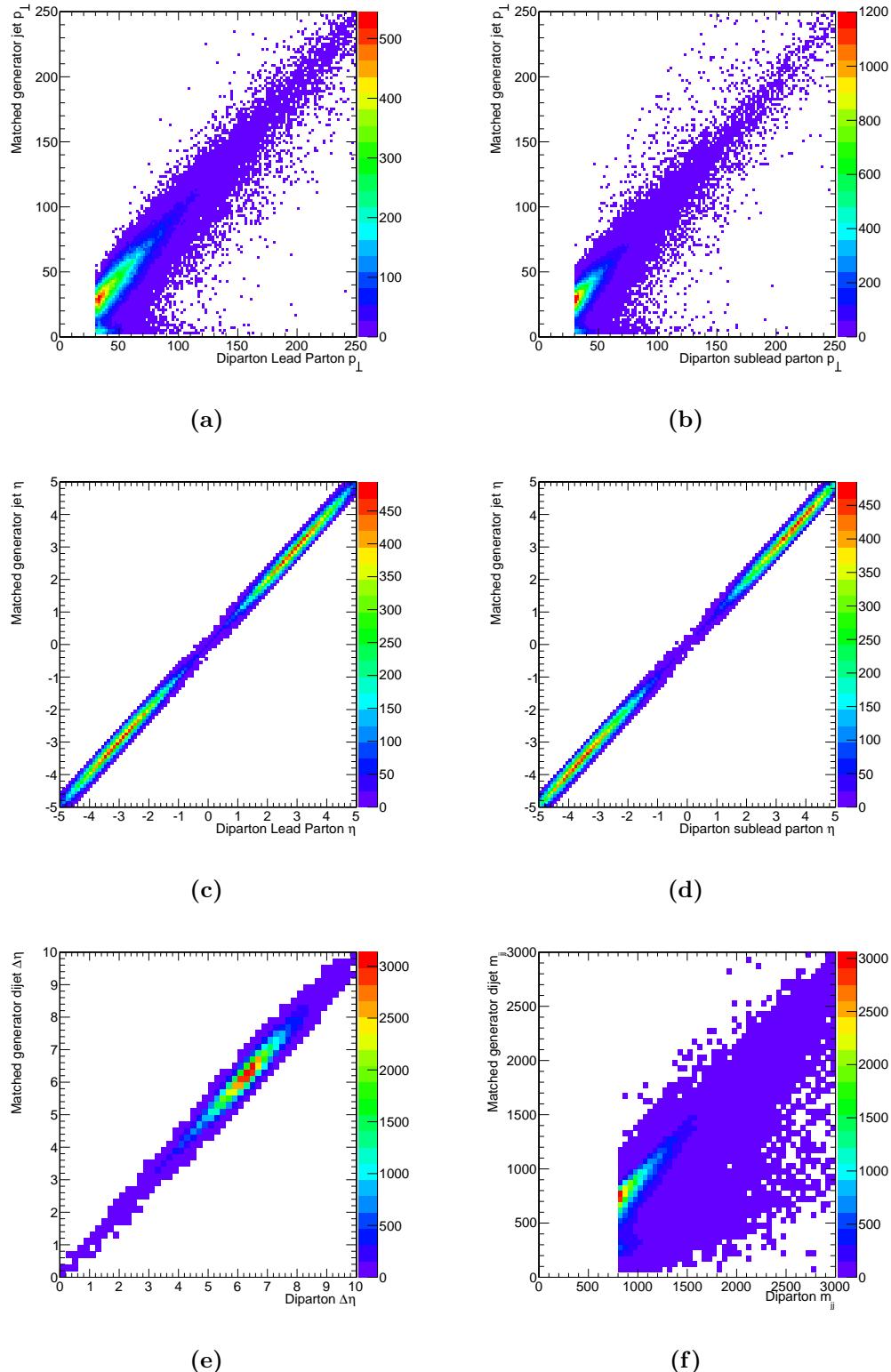
minimum  $p_T$  necessary to form a jet of 3 GeV. A set of plots of the relevant variables is shown in figure 8.10. The selected di-partons are taken and each variable is plotted against the matched dijet.

On plots 8.10 a), b) and f) two populations can be seen. In the parton  $p_T$  plots they are along the diagonal and along the line of generator jet  $p_T$  equal to zero and in the  $M_{jj}$  plot along the diagonal and along the line of  $y = x/2$ . In both cases the non-diagonal population probably arises from mismatching a parton to a low  $p_T$  jet. In the  $M_{jj}$  plot, the second diagonal line is due to the fact that at parton level the partons are perfectly matched in energy and momentum but if they are matched to only one correct generator jet and the other jet has  $p_T$  close to zero, the system will have half the mass of the correctly assigned events.

Event migrations can now be calculated from the events that did not pass the parton event selection but could have passed the generator level selection. This effect can be from jet dispersion, overlap, or clustering artefacts. Let's first consider the migrations on each variable separately, lead jet  $p_T$  (eq. 8.1), sub-lead jet  $p_T$  (eq. 8.2) and dijet  $M_{jj}$  (eq. 8.3).

$$\frac{p_T^{Parton} < 30 \wedge p_T^{GenJet} \geq 40}{p_T^{GenJet} \geq 40} = 0.27\% \pm 0.04\% \quad (8.1)$$

$$\frac{p_T^{Parton} < 30 \wedge p_T^{GenJet} \geq 40}{p_T^{GenJet} \geq 40} = 0.56\% \pm 0.08\% \quad (8.2)$$



**Figure 8.10:** Plots for relevant variables of the selected di-parton against its matched dijet:  
 (a) lead parton  $p_T$  (b) sub-leading parton  $p_T$  (c) lead parton  $\eta$  (d) sub-lead parton  $\eta$  (e) di-parton  $\Delta\eta$  (d) di-parton  $M_{jj}$ .

$$\frac{M_{jj}^{Parton} < 800 \wedge M_{jj}^{GenJet} \geq 1000}{M_{jj}^{GenJet} \geq 800} = 0.13\% \pm 0.04\% \quad (8.3)$$

Migrations of events over all variables simultaneously can now be consider using equation 8.4.

$$\frac{(p_T^{GenJet} > 40 \wedge M_{jj}^{GenJet} > 1000) \wedge (p_T^{Parton} < 30 \vee M_{jj}^{Parton} < 800)}{p_T^{GenJet} > 40 \cup M_{jj}^{GenJet} > 1000} = 0.23\% \pm 0.13\% \quad (8.4)$$

Global migrations of events from below the selected parton level cuts to above the selected generator cuts are of  $0.23\% \pm 0.13\%$  of the total number events passing the generator filter. This is an acceptable value which should not bias in any relevant way the physics usage of this sample.

### 8.2.6 Summary

The production of a new QCD MC event sample with VBF characteristics was studied and all objectives were achieved. The MADGRAPH event generator is used and configured to produce proton-proton to two, three or four outgoing partons, where these partons can be gluons or quarks except the top quark. At this stage events are filtered by only accepting those that have at least one di-parton with  $M_{jj} > 800$  GeV where each parton has at least 30 GeV and is contained inside the detector acceptance of  $|\eta| < 5.0$ . This process has a cross section of  $1.029 \times 10^7 \pm 1.614 \times 10^4$  pb.

Event hadronization is performed using PYTHIA 8 event generator with the MLM jet matching scheme as traditionally done in the CMS experiment. The estimated efficiency of this step is  $9.45\% \pm 0.03\%$  which leads cross section of  $(9.718 \pm 0.029) \times 10^5$  pb. From those events, the ones containing at least one generator dijet passing  $\Delta\eta > 3.0$ ,  $M_{jj} > 1000$  GeV, where both jets pass  $p_T > 40$  GeV and  $|\eta| < 4.8$ , are kept. The sample is split into 2 sub-samples according if the dijet passing all cuts is below (sub-sample A) of above  $\Delta\phi = 2.15$  (sub-sample B). The filter efficiency for sub-sample A is  $(2.938 \pm 0.005) \times 10^{-1}$  and this sub-sample is aimed to have  $1 \text{ fb}^{-1}$  of equivalent integrated luminosity. Sub-sample B filter efficiency is  $(1.125 \pm 0.009) \times 10^{-1}$  and will have  $0.1 - 1.0 \text{ fb}^{-1}$  of equivalent integrated luminosity depending on available resources.

The overlap between the two sub-samples has been estimated to be  $3.95\% \pm 0.14\%$  thus requiring care in combining them.

Migrations from events below the parton level cuts to above the generator level cuts have been determined to be  $0.23\% \pm 0.13\%$  of the total number of events passing the generator filters.

The MADGRAPH code for event generation has been approved by the CMS MC production team. The additional code necessary for the generator level filtering has been also approved and is queued for integration in the experiment's software. Final approval of this sample production is under way.

# Chapter 9

## Conclusions

This thesis describes the study of a Higgs particle decaying invisibly. These studies start with the development of a dedicated trigger and continue to the cross checking of the final analysis on parked data. Finally, preparation for an updated analysis for Run II of the LHC was presented.

The search for VBF produced SM Higgs boson decaying invisibly was preformed using the full promptly reconstructed 8 TeV Run I data. A dedicated trigger was used to collect events over which a single bin counting experiment was optimised to select events containing a dijet with VBF characteristics and large MET. Control regions were also defined to normalise the main background processes which were extrapolated to the signal region with the help of MC simulation. In the signal region, 390 data events were observed, this yield is compatible with the background only prediction. Since no evidence of signal is observed 95% CL upper limits on the Higgs boson production cross section times branching were determined. Assuming the SM VBF production cross section and acceptance, corresponding to an observed (expected) upper limit on  $\mathcal{B}(H \rightarrow \text{inv})$  of 0.65 (0.49) for  $m_H = 125$  GeV.

Following the reconstruction of the lower trigger thresholds parked datasets, after the end of the Run I data taking, a new analysis was designed to take advantage of the higher signal acceptance provided by dedicated triggers in this new available data. New variables were introduced, as well as new control regions and background estimation methods. A new cross check analysis was also implemented to validate the obtained results. In the signal region 508 data events were observed, this yield is again compatible with the background only prediction. In the absence of signal, 95% CL upper limits, were once again determined for the Higgs boson production cross section times branching.

Under the assumption of the SM production cross sections and acceptances, the observed (expected) 95% C.L. limit on  $\mathcal{B}(H \rightarrow \text{inv})$  of a SM 125 GeV Higgs boson is 57% (40%).

In preparation for the Run II analysis new dedicated triggers, for both signal recording and systematics control, were developed and were successfully used during the 2015 Run II data taking. Additionally, a study leading to official production proposal of QCD MC datasets with signal like characteristics was performed. This new approach should for the first time allow the simulation of events with miss measured MET to be used in future searches.

The developed tools for the CMS L1T system, were used during Run I and provided the ability to monitor trigger objects production rate, synchronization with the LHC, assert the correct operation of the BPTX system, detect problematic regions of the detector and allow easy trouble shooting for the shift crew. These tools also played a crucial role in the data certification for physics usage allowing identification of problematic periods of data taking.

The LHC continues its ground breaking program exploring the TeV energy scale. During the LHC Run I the discovery of a Higgs boson with mass around 125 GeV by both the CMS and ATLAS collaborations has lead to the 2013 Nobel Prize in Physics being awarded to Higgs and Englert. With the start of Run II in the beginning of 2015, a significant increase in centre of mass energy to 13 TeV has been achieved and plans are in place to record even greater volumes of data than before. With these new data physicists will be able to probe the Standard Model even farther, opening the door for new physics discovery.



# Bibliography

- [1] CMS Collaboration, “Search for invisible decays of Higgs bosons in the vector boson fusion production mode”,.
- [2] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [3] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [4] D. Griffiths, “Introduction to Elementary Particles”. Wiley, 2nd edition, 2008.
- [5] F. Halzen and A. Martin, “Quarks and leptons: an introductory course in modern particle physics”. Wiley, 1984.
- [6] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.
- [7] CDF Collaboration, D0 Collaboration Collaboration, “Precise measurements of the  $W$  mass at the Tevatron and indirect constraints on the Higgs mass”, *ArXiv e-prints: 1204.3260* (2012) arXiv:1204.3260.
- [8] DELPHI Collaboration, “Cross-sections and leptonic forward backward asymmetries from the Z0 running of LEP”, *Eur. Phys. J. C* **16** (2000) 371–405, doi:10.1007/s100520000392.
- [9] E. Noether, “Invariant variation problems”, *Transport Theory and Statistical Physics* **1** (1971) 186–207, doi:10.1080/00411457108231446, arXiv:arXiv:physics/0503066.
- [10] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl.Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.

- [11] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (Nov, 1967) 1264–1266, doi:[10.1103/PhysRevLett.19.1264](https://doi.org/10.1103/PhysRevLett.19.1264).
- [12] A. Salam, “Weak and Electromagnetic Interactions”, *Conf.Proc.* **C680519** (1968) 367–377.
- [13] Gargamelle Neutrino Collaboration, “Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment”, *Phys. Lett.* **B46** (1973) 138–140, doi:[10.1016/0370-2693\(73\)90499-1](https://doi.org/10.1016/0370-2693(73)90499-1).
- [14] UA1 Collaboration, “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $\sqrt{s} = 540$  GeV”, *Phys. Lett.* **B122** (1983) 103–116, doi:[10.1016/0370-2693\(83\)91177-2](https://doi.org/10.1016/0370-2693(83)91177-2).
- [15] 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) 476–485, doi:[10.1016/0370-2693\(83\)91605-2](https://doi.org/10.1016/0370-2693(83)91605-2).
- [16] UA1 Collaboration, “Experimental Observation of Lepton Pairs of Invariant Mass Around 95  $\text{GeV}/c^2$  at the CERN SPS Collider”, *Phys. Lett.* **B126** (1983) 398–410, doi:[10.1016/0370-2693\(83\)90188-0](https://doi.org/10.1016/0370-2693(83)90188-0).
- [17] UA2 Collaboration, “Evidence for  $Z^0 \rightarrow e^+ e^-$  at the CERN anti-p p Collider”, *Phys. Lett.* **B129** (1983) 130–140, doi:[10.1016/0370-2693\(83\)90744-X](https://doi.org/10.1016/0370-2693(83)90744-X).
- [18] I. J. R. Aitchison and A. J. G. Hey, “Gauge Theories in Particle Physics, 2 Volume Set”. Taylor & Francis, 3th edition, 2004.
- [19] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132, doi:[10.1016/0031-9163\(64\)91136-9](https://doi.org/10.1016/0031-9163(64)91136-9).
- [20] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Phys. Rev. Lett.* **13** (1964) 508, doi:[10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [21] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global conservation laws and massless particles”, *Phys. Rev. Lett.* **13** (1964) 585, doi:[10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585).
- [22] P. W. Higgs, “Spontaneous symmetry breakdown without massless bosons”, *Phys. Rev.* **145** (1966) 1156, doi:[10.1103/PhysRev.145.1156](https://doi.org/10.1103/PhysRev.145.1156).

- [23] T. W. B. Kibble, “Symmetry breaking in non-Abelian gauge theories”, *Phys. Rev.* **155** (1967) 1554, doi:10.1103/PhysRev.155.1554.
- [24] “The LEP Electroweak Working Group”, . <http://lepewwg.web.cern.ch>.
- [25] LEP Working Group for Higgs boson searches, ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, “Search for the standard model Higgs boson at LEP”, *Phys. Lett.* **B565** (2003) 61–75, doi:10.1016/S0370-2693(03)00614-2, arXiv:hep-ex/0306033.
- [26] CDF Collaboration, D0 Collaboration, “Higgs Boson Studies at the Tevatron”, *Phys. Rev.* **D88** (2013) 052014, doi:10.1103/PhysRevD.88.052014, arXiv:1303.6346.
- [27] LHC Higgs Cross Section Working Group, “Handbook of LHC Higgs Cross Sections: 3. Higgs Properties”, doi:10.5170/CERN-2013-004, arXiv:1307.1347.
- [28] ATLAS, CMS Collaboration, “Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.
- [29] CMS Collaboration, “Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV”, arXiv:1412.8662.
- [30] ATLAS Collaboration, “Measurements of Higgs boson production and couplings in the four-lepton channel in  $pp$  collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector”, *Phys. Rev.* **D91** (2015), no. 1, 012006, doi:10.1103/PhysRevD.91.012006, arXiv:1408.5191.
- [31] ATLAS Collaboration, “Measurements of fiducial and differential cross sections for Higgs boson production in the diphoton decay channel at  $\sqrt{s} = 8$  TeV with ATLAS”, *JHEP* **1409** (2014) 112, doi:10.1007/JHEP09(2014)112, arXiv:1407.4222.
- [32] ATLAS Collaboration, “Evidence for the Higgs-boson Yukawa coupling to tau leptons with the ATLAS detector”, arXiv:1501.04943.
- [33] CMS Collaboration, “Measurement of the properties of a Higgs boson in the four-lepton final state”, arXiv:1312.5353.

- [34] CMS Collaboration, “Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states”, *JHEP* **1401** (2014) 096, doi:10.1007/JHEP01(2014)096, arXiv:1312.1129.
- [35] ATLAS Collaboration, “Evidence for the spin-0 nature of the Higgs boson using ATLAS data”, *Phys. Lett. B* **726** (2013) 120, doi:10.1016/j.physletb.2013.08.026, arXiv:1307.1432.
- [36] G. Belanger, F. Boudjema, A. Cottrant et al., “The MSSM invisible Higgs in the light of dark matter and g-2”, *Phys. Lett.* **B519** (2001) 93–102, doi:10.1016/S0370-2693(01)00976-5, arXiv:hep-ph/0106275.
- [37] G. F. Giudice, R. Rattazzi, and J. D. Wells, “Graviscalars from higher dimensional metrics and curvature Higgs mixing”, *Nucl. Phys.* **B595** (2001) 250–276, doi:10.1016/S0550-3213(00)00686-6, arXiv:hep-ph/0002178.
- [38] M. Battaglia, D. Dominici, J. F. Gunion et al., “The Invisible Higgs decay width in the add model at the LHC”, in *Physics at TeV colliders. Proceedings, Workshop, Les Houches, France, May 26-June 3, 2003.* 2004. arXiv:hep-ph/0402062.
- [39] O. S. Bruning, P. Collier, P. Lebrun et al., “LHC Design Report Vol.1: The LHC Main Ring”,,
- [40] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [41] LEP Injector Study Group Collaboration, “LEP Design Report, Volume 1: The LEP Injector Chain”,.
- [42] P. Mouche, “CERN Document Server: Overall view of the LHC”, <http://cds.cern.ch/record/1708847> (Accessed Dec, 2015).
- [43] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation* **3** (2008), no. 08, S08003.
- [44] CMS Collaboration, “The CMS experiment at the CERN LHC”, *Journal of Instrumentation* **3** (2008), no. 08, S08004.
- [45] LHCb Collaboration, “The LHCb Detector at the LHC”, *Journal of Instrumentation* **3** (2008), no. 08, S08005.
- [46] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *Journal of Instrumentation* **3** (2008), no. 08, S08002.

- [47] C. Lefèvre, “The CERN accelerator complex”, <https://cds.cern.ch/record/1260465> (Accessed Dec, 2015).
- [48] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”,,. There is an error on cover due to a technical problem for some items.
- [49] M. Bajko, F. Bertinelli, N. Catalan Lasheras et al., “Report of the task force on the incident of 19th September 2008 at the LHC”, *LHC Project Report LHC-PROJECT-REPORT-1168* (2009).
- [50] CMS Collaboration, “Multi-year Collisions Plots”,  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (Accessed Dec, 2015).
- [51] CMS Collaboration, “Average Pileup in 2012”,  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults> (Accessed Dec, 2015).
- [52] CMS Collaboration, “CMS Compact Muon Solenoid”,  
<http://public.web.cern.ch/public/Objects/LHC/CMSnc.jpg> (Accessed Nov, 2010).
- [53] CMS Collaboration, “CMS, the Compact Muon Solenoid: Technical proposal”,.
- [54] CMS Collaboration, “CMS, tracker technical design report”,.
- [55] CMS Collaboration, “Addendum to the CMS tracker TDR”,.
- [56] CMS Collaboration, “The CMS electromagnetic calorimeter project : Technical Design Report”,.
- [57] CMS Collaboration, “Changes to CMS ECAL electronics : addendum to the Technical Design Report”,.
- [58] USCMS, ECAL/HCAL Collaboration, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”, *Eur. Phys. J. C***60** (2009) 359–373, doi:10.1140/epjc/s10052-009-0959-5, 10.1140/epjc/s10052-009-1024-0. [Erratum: Eur. Phys. J.C61,353(2009)].
- [59] CMS Collaboration, “The CMS hadron calorimeter project : Technical Design Report”,.
- [60] CMS Collaboration, “The CMS magnet project : Technical Design Report”,.

- [61] CMS Collaboration, “The CMS muon project : Technical Design Report”,.
- [62] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2013) 30–61,  
[doi:10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021), [arXiv:1207.7235](https://arxiv.org/abs/1207.7235).
- [63] CMS Collaboration, “CMS The TriDAS Project : Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger”,.
- [64] CMS Collaboration, “CMS TriDAS project : Technical Design Report, Volume 1: The Trigger Systems”,.
- [65] CMS Collaboration Collaboration, “Data Parking and Data Scouting at the CMS Experiment”,.
- [66] CMS Collaboration, “CMS computing: Technical Design Report”,. Submitted on 31 May 2005.
- [67] CMS Collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade”,.
- [68] CMS Collaboration, “Technical proposal for the upgrade of the CMS detector through 2020”,.
- [69] W. Adam, B. Mangano, T. Speer et al., “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Geneva, (Dec, 2006).
- [70] S. Cucciarelli, M. Konecki, D. Kotlinski et al., “Track reconstruction, primary vertex finding and seed generation with the Pixel Detector”, Technical Report CMS-NOTE-2006-026, CERN, Geneva, (Jan, 2006).
- [71] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* **262** (1987) 444–450, [doi:10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [72] CMS Collaboration, “Measurement of Tracking Efficiency”,.
- [73] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression, and related Optimization Problems”, *Proceedings of the IEEE* **86** (1998) 2210.
- [74] W. Waltenberger, R. Frühwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 12, N343.

- [75] CMS Collaboration, “Tracking and Primary Vertex Results in First 7 TeV Collisions”,.
- [76] CMS Collaboration Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, Technical Report CMS-PAS-PFT-10-001, 2010.
- [77] CMS Collaboration Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, Technical Report CMS-PAS-PFT-09-001, CERN, 2009. Geneva, (Apr, 2009).
- [78] CMS Collaboration Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, Technical Report CMS-PAS-PFT-10-002, CERN, Geneva, (2010).
- [79] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV”, *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [80] T. C. collaboration, “Performance of CMS muon reconstruction in pp collision events at  $\sqrt{s}= 7$  TeV”, *Journal of Instrumentation* **7** (2012), no. 10, P10002.
- [81] S. Baffioni, C. Charlot, F. Ferri et al., “Electron reconstruction in CMS”, Technical Report CMS-NOTE-2006-040, CERN, Geneva, (Feb, 2006).
- [82] E. Meschi, T. Monteiro, C. Seez et al., “Electron Reconstruction in the CMS Electromagnetic Calorimeter”, Technical Report CMS-NOTE-2001-034, CERN, Geneva, (Jun, 2001).
- [83] W. Adam, R. Fröhwirth, A. Strandlie et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.
- [84] CMS Collaboration, “Electron Reconstruction and Identification at  $\sqrt{s} = 7$  TeV”,.
- [85] C. Collaboration, “Performance of CMS muon reconstruction in cosmic-ray events”, *Journal of Instrumentation* **5** (2010), no. 03, T03022.
- [86] CMS Collaboration, “Performance of muon identification in pp collisions at  $\sqrt{s} = 7$  TeV”,.

- 
- [87] G. P. Salam, “Towards Jetography”, *Eur. Phys. J.* **C67** (2010) 637–686, doi:10.1140/epjc/s10052-010-1314-6, arXiv:0906.1833.
  - [88] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
  - [89] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J.* **C72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
  - [90] N. Saoulidou and E. Tziaferi, “Performance of the Particle-Flow jet identification criteria using proton-proton collisions at 8 TeV”, *CMS Analysis Note: AN-14-227* (2014).
  - [91] CMS Collaboration Collaboration, “Pileup Jet Identification”, Technical Report CMS-PAS-JME-13-005, CERN, Geneva, (2013).
  - [92] T. C. collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *Journal of Instrumentation* **6** (2011), no. 11, P11002.
  - [93] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett.* **B659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.
  - [94] CDF Collaboration, “The Dijet angular distribution in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.8$  TeV”, *Phys. Rev. Lett.* **69** (1992) 2896–2900, doi:10.1103/PhysRevLett.69.2896.
  - [95] Particle Data Group Collaboration, “Review of Particle Physics\*”, *Phys. Rev. D* **86** (Jul, 2012) 010001, doi:10.1103/PhysRevD.86.010001.
  - [96] CMS Collaboration, “Performance of tau-lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001, arXiv:1109.6034.
  - [97] CMS Collaboration, “Reconstruction and identification of  $\tau$  lepton decays to hadrons and  $\nu_\tau$  at CMS”, arXiv:1510.07488.
  - [98] “Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV with the CMS Detector”,

- [99] T. C. collaboration, “Missing transverse energy performance of the CMS detector”, *Journal of Instrumentation* **6** (2011), no. 09, P09001.
- [100] A. Buckley et al., “General-purpose event generators for LHC physics”, *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.
- [101] F. Krauss, “Phenomenology at collider experiments”, [http://www.ippp.dur.ac.uk/~krauss/Lectures/MonteCarlos/Pheno\\_5\\_08.pdf](http://www.ippp.dur.ac.uk/~krauss/Lectures/MonteCarlos/Pheno_5_08.pdf) (Accessed May, 2014).
- [102] T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [103] T. Sjostrand, S. Mrenna, and P. Z. Skands, “A Brief Introduction to PYTHIA 8.1”, *Comput. Phys. Commun.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- [104] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J.* **C58** (2008) 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [105] T. Gleisberg, S. Hoeche, F. Krauss et al., “Event generation with SHERPA 1.1”, *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.
- [106] J. Alwall, M. Herquet, F. Maltoni et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.
- [107] M. L. Mangano, M. Moretti, F. Piccinini et al., “ALPGEN, a generator for hard multiparton processes in hadronic collisions”, *JHEP* **07** (2003) 001, doi:10.1088/1126-6708/2003/07/001, arXiv:hep-ph/0206293.
- [108] J. Alwall, R. Frederix, S. Frixione et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [109] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

- [110] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.
- [111] S. Alioli, P. Nason, C. Oleari et al., “NLO single-top production matched with shower in POWHEG:  $s$ - and  $t$ -channel contributions”, *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: DOI10.1007/JHEP02(2010)011].
- [112] K. Hamilton, P. Richardson, and J. Tully, “A positive-weight Next-to-Leading Order Monte Carlo simulation for Higgs boson production”, *JHEP* **09** (2009) 116, doi:10.1088/1126-6708/2009/04/116, arXiv:0903.4345.
- [113] P. Nason and C. Oleari, “NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG”, *JHEP* **02** (2010) 037, doi:10.1007/JHEP02(2010)037, arXiv:0911.5299.
- [114] S. Alioli, P. Nason, C. Oleari et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [115] E. Re, “Single-top  $Wt$ -channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.
- [116] S. Alioli, K. Hamilton, P. Nason et al., “Jet pair production in POWHEG”, *JHEP* **04** (2011) 081, doi:10.1007/JHEP04(2011)081, arXiv:1012.3380.
- [117] S. Hoeche, F. Krauss, N. Lavesson et al., “Matching parton showers and matrix elements”, in *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*. 2006. arXiv:hep-ph/0602031.
- [118] S. Catani, F. Krauss, R. Kuhn et al., “QCD matrix elements + parton showers”, *JHEP* **11** (2001) 063, doi:10.1088/1126-6708/2001/11/063, arXiv:hep-ph/0109231.
- [119] F. Krauss, “Matrix elements and parton showers in hadronic interactions”, *JHEP* **08** (2002) 015, doi:10.1088/1126-6708/2002/08/015, arXiv:hep-ph/0205283.

- [120] S. Mrenna and P. Richardson, “Matching matrix elements and parton showers with HERWIG and PYTHIA”, *JHEP* **05** (2004) 040,  
[doi:10.1088/1126-6708/2004/05/040](https://doi.org/10.1088/1126-6708/2004/05/040), [arXiv:hep-ph/0312274](https://arxiv.org/abs/hep-ph/0312274).
- [121] M. Mangano, “Exploring theoretical systematics in the ME-to-shower MC merging for multijet process”, in *E-proceedings of Matrix Element/Monte Carlo Tuning Working Group, Fermilab*. 2002.
- [122] CMS Collaboration, “Measurement of the Underlying Event Activity at the LHC with  $\sqrt{s} = 7$  TeV and Comparison with  $\sqrt{s} = 0.9$  TeV”, *JHEP* **09** (2011) 109,  
[doi:10.1007/JHEP09\(2011\)109](https://doi.org/10.1007/JHEP09(2011)109), [arXiv:1107.0330](https://arxiv.org/abs/1107.0330).
- [123] GEANT4 Collaboration, “GEANT4: A Simulation toolkit”, *Nucl. Instrum. Meth.* **A506** (2003) 250–303, [doi:10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [124] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (2006) 270, [doi:10.1109/TNS.2006.869826](https://doi.org/10.1109/TNS.2006.869826).
- [125] CMS Collaboration, “The CMS data quality monitoring software: experience and future prospects”, *J. Phys. Conf. Ser.* **513** (2014) 032024,  
[doi:10.1088/1742-6596/513/3/032024](https://doi.org/10.1088/1742-6596/513/3/032024).
- [126] D. Montgomery and G. Runger, “Applied Statistics and Probability for Engineers”. John Wiley & Sons, 2010.
- [127] ALEPH, DELPHI, L3, and OPAL Collaborations, LEP Higgs Working Group, “Searches for invisible Higgs bosons: Preliminary combined results using LEP data collected at energies up to 209-GeV”, (2001). [arXiv:hep-ex/0107032](https://arxiv.org/abs/hep-ex/0107032).
- [128] DELPHI Collaboration, “Searches for invisibly decaying Higgs bosons with the DELPHI detector at LEP”, *Eur. Phys. J. C* **32** (2004) 475,  
[doi:10.1140/epjc/s2003-01469-8](https://doi.org/10.1140/epjc/s2003-01469-8), [arXiv:hep-ex/0401022](https://arxiv.org/abs/hep-ex/0401022).
- [129] OPAL Collaboration, “Search for invisibly decaying Higgs bosons with large decay width using the OPAL detector at LEP”, *Eur. Phys. J. C* **49** (2007) 457,  
[doi:10.1140/epjc/s10052-006-0170-x](https://doi.org/10.1140/epjc/s10052-006-0170-x), [arXiv:hep-ex/0610056](https://arxiv.org/abs/hep-ex/0610056).
- [130] ATLAS Collaboration, “Search for Invisible Decays of a Higgs Boson Produced in Association with a Z Boson in ATLAS”, *Phys. Rev. Lett.* **112** (2014) 201802,  
[doi:10.1103/PhysRevLett.112.201802](https://doi.org/10.1103/PhysRevLett.112.201802), [arXiv:1402.3244](https://arxiv.org/abs/1402.3244).
- [131] ATLAS Collaboration, “Search for dark matter in events with a hadronically decaying W or Z boson and missing transverse momentum in pp collisions at

- $\sqrt{s} = 8$  TeV with the ATLAS detector”, *Phys. Rev. Lett.* **112** (2014) 041802, doi:10.1103/PhysRevLett.112.041802, arXiv:1309.4017.
- [132] ATLAS Collaboration, “Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at  $\sqrt{s} = 8$  TeV with the ATLAS detector”, arXiv:1502.01518. *sub. to Eur. Phys. J. C.*
- [133] “Search for an Invisibly Decaying Higgs Boson Produced via Vector Boson Fusion in  $pp$  Collisions at  $\sqrt{s} = 8$  TeV using the ATLAS Detector at the LHC”, ATLAS Conference Report ATLAS-CONF-2015-004, CERN, (Mar, 2015).
- [134] CMS Collaboration, “Search for invisible decays of Higgs bosons in the vector boson fusion and associated ZH production modes”, *Eur. Phys. J.* **C74** (2014) 2980, doi:10.1140/epjc/s10052-014-2980-6, arXiv:1404.1344.
- [135] O. J. P. Eboli and D. Zeppenfeld, “Observing an invisible Higgs boson”, *Phys. Lett.* **B495** (2000) 147–154, doi:10.1016/S0370-2693(00)01213-2, arXiv:hep-ph/0009158.
- [136] J. M. Campbell and R. K. Ellis, “MCFM for the Tevatron and the LHC”, *Nucl. Phys. Proc. Suppl.* **205** (2010) 10, doi:10.1016/j.nuclphysbps.2010.08.011.
- [137] LHC Higgs Cross Section Working Group, S. Dittmaier, et al., “Handbook of LHC Higgs Cross Sections: 2. Differential Distributions”, CERN Report CERN-2012-002, (2012).
- [138] A. L. Read, “Presentation of search results: the  $CL_s$  technique”, *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [139] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2.
- [140] G. Cowan, K. Cranmer, E. Gross et al., “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1, doi:10.1140/epjc/s10052-011-1554-0.
- [141] ATLAS and CMS Collaborations, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report ATL-PHYS-PUB-2011-11, CMS NOTE 2011/005, (2011).

- [142] LHC Higgs Cross Section Working Group, S. Dittmaier, et al., “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”, CERN Report CERN-2011-002, (2011).
- [143] G. Kane, “Perspectives on Higgs Physics II”. Number v. 2 in Advanced series on directions in high energy physics. World Scientific, 1997.
- [144] P. Dunne, “Internal CMS Higgs to Invisible Presentation (9 September 2014)”,.
- [145] CMS Collaboration Collaboration, “Analysis of diffractive jets”, Technical Report CMS-PAS-FWD-10-004, CERN, Geneva, (2012).
- [146] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting – Summer 2013 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, (2013).
- [147] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.
- [148] CMS Collaboration, “Measurement of  $W^+W^-$  and ZZ production cross sections in pp collisions at  $\sqrt{s} = 8$  TeV”, *Phys. Lett. B* **721** (2013) 190, doi:10.1016/j.physletb.2013.03.027, arXiv:1301.4698.
- [149] P. Dunne, “LHC End-of-Year Jamboree 2015”,  
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/HigPagDec2015> (Accessed Dec, 2015).

# Acronyms

**ALICE** A Large Ion Collider Experiment. 15

**APD** Avalanche photo-diodes. 23

**ATLAS** A Toroidal LHC ApparatuS. 15, 69

**AVF** Adaptive Vertex Fitter. 35

**BDT** Boosted Decision Tree. 46, 49, 88, 89, 92

**BPTX** Beam Pickup for Timing for the eXperiments. 59–61, 147

**BSM** Beyond the Standard Model. 15, 80

**CERN** European Organization for Nuclear Research. 14–16, 32

**CJV** Central Jet Veto. 70, 71, 73–75, 95, 96

**CL** Confidence Level. 10, 77, 78, 117, 146

**CMS** Compact Muon Solenoid. v, xiii, 12, 15, 18–22, 25–30, 32, 34, 37, 38, 41–46, 50–61, 69, 76, 77, 79, 80, 86, 95, 97–99, 103, 113, 117, 119, 121, 126, 133–135, 137, 144, 145, 147

**CPU** Central Processing Unit. 31

**CSC** Cathode Strip Chamber. 28, 29, 43

**CTF** Combinatorial Track Finder. 34, 39

**DA** Deterministic Annealing. 35

**DAQ** Data Acquisition. 30, 54, 55

**DCS** Detector Control System. 55

**DIP** Data Interchange Protocol. 58

**DPG** Detector Performance Group. 59, 60

**DQM** Data Quality Monitoring. 54–56

**DT** Drift Tube. 28, 29, 43

**EB** ECAL Barrel. 24

**ECAL** Electromagnetic Calorimeter. 23–27, 37–40, 51

**EE** ECAL Endcaps. 24

**EM** Electromagnetic. 46, 66

**EWK** Electroweak. 100, 105, 120

**FCT** Fundação para a Ciência e a Tecnologia. vii

**GCT** Global Calorimeter Trigger. 33, 66

**GSF** Gaussian Sum Filter. 39

**GT** Global Trigger. 58, 60

**HB** HCAL Barrel. 25, 26

**HCAL** Hadronic Calorimeter. 25–27, 37, 38, 51

**HE** HCAL Endcap. 26

**HF** HCAL Forward. 26, 27

**HLT** High Level Trigger. 31, 32, 51, 55, 58, 82–85, 120, 121, 123–126, 128–133

**HO** HCAL Outer. 26

**HPS** Hadron Plus Strips. 48

**HT** Hadronic Total. 83, 85, 92, 120

**ID** Identification. 41, 42, 46

**ISR** Initial State Radiation. 71, 77

**L1T** Level 1 Trigger. v, 25, 31–33, 54–62, 65–67, 80–85, 99, 120–124, 126, 128–133, 147

**LEP** Large Electron Positron collider. 10, 14, 69

- LHC** Large Hadron Collider. 9, 10, 12–23, 30, 32, 35, 36, 51, 53, 58–60, 69, 79, 80, 86, 99, 113, 119–121, 133, 146, 147
- LHCb** Large Hadron Collider beauty. 15
- LINAC2** Linear Particle Accelerator 2. 15
- LO** Leading Order. 51
- LS1** Long Shutdown 1. 98, 119
- MC** Monte Carlo. 34, 51, 53, 63, 64, 70–77, 79, 86–90, 92, 93, 95, 99, 101, 102, 104, 105, 107, 108, 111–114, 118–121, 134, 135, 141, 144–147
- ME** Matrix Element. 52, 135, 137
- MET** Missing Transverse Energy. xviii, xix, 48, 50, 51, 69–71, 73–76, 80–82, 84–90, 92, 97, 99, 100, 103, 105–107, 109–114, 116, 120–134, 146, 147
- MHT** Missing Hadronic Total. 122
- MPF** Missing Transverse Energy Projection Fraction. 47
- MSSM** Minimal Supersymmetric Standard Model. 99
- MVA** Multivariate Analysis. 95, 96
- NLO** Next to Leading Order. 51, 101
- NNLO** Next to Next to Leading Order. 101
- OMDS** Online Master Database System. 58
- ORCON** Offline Reconstruction Condition DB ONline. 58
- PDF** Parton Distribution Function. 76, 113
- PF** Particle Flow. 37, 38, 44–46, 48–51, 87, 88, 99, 102, 103, 121, 123–126, 128–133
- POG** Particle Object Group. xiii, 38, 41, 42, 44, 46, 50, 71, 103
- PS** Proton Synchrotron. 15
- PSB** Proton Synchrotron Booster. 15
- PU** Pile-Up. 19, 32, 38, 41, 43, 46, 47, 49, 70, 80–82, 86, 94–97, 102, 113, 119–121

**PV** Primary Vertex. 35, 36, 38, 46, 94, 95

**QCD** Quantum Chromodynamics. xviii, 4, 38, 49, 70, 75, 76, 79, 80, 84, 86–90, 92–97, 100, 103, 104, 107, 111–114, 117, 119–121, 134, 135, 144, 147

**QED** Quantum Electrodynamics. 4

**QFT** Quantum Field Theory. 2, 4

**RF** Radio Frequency. 16

**RPC** Resistive Plate Chamber. 28, 29

**SLD** SLAC Large Detector. 10

**SM** Standard Model. 2, 3, 5, 6, 8–10, 12, 13, 16, 27, 69, 71, 75, 78, 82, 84, 85, 88, 99, 118, 133, 146, 147

**SPS** Super Proton Synchrotron. 15, 18

**SUSY** Supersymmetry. 28, 135

**TEC** Tracker Endcaps. 23

**TIB** Tracker Inner Barrel. 23

**TID** Tracker Inner Disks. 23

**TOB** Tracker Outer Barrel. 23

**TSG** Trigger Studies Group. 59, 81, 119, 121–126, 128–133

**UE** Underlying Event. 47, 52

**VBF** Vector Boson Fusion. xiv, xviii, 10, 34, 45, 51, 69–71, 76–80, 82–86, 88, 92–95, 98–100, 106, 112, 115, 116, 119, 121, 124, 126–131, 133–135, 144, 146

**VPT** Vacuum Photo-Triodes. 23

**WbM** Web base Monitoring. 56