SEARCH FOR AN EXOTIC DECAY OF THE HIGGS BOSON TO A PAIR OF LIGHT PSEUDOSCALARS IN THE FINAL STATE WITH TWO B QUARKS AND TWO TAU LEPTONS IN PROTON-PROTON COLLISIONS AT $\sqrt{s} = 13 \, \text{TeV}$
SEARCH FOR AN EXOTIC DECAY OF THE HIGGS BOSON TO A PAIR OF LIGHT PSEUDOSCALARS IN THE FINAL STATE WITH TWO B QUARKS AND TWO TAU LEPTONS IN PROTON-PROTON COLLISIONS AT $\sqrt{s} = 13 \text{TeV}$

TSIAKKOURI DEMETRA

A Dissertation Submitted to the University of Cyprus in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

MAY 2019
VALIDATION PAGE

Doctoral Candidate: Tsiakkouri Demetra

Doctoral Thesis Title: Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two b quarks and two tau leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV

The present Doctoral Dissertation was submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy at the Department of Physics and was approved on the 09/05/2019 by the members of the Examination Committee.

Examination Committee:

Chair: Andreas Othonos (Prof. Dep. Of Physics, UCY)

Research Supervisor: Panos A. Razis (Prof. Dep. Of Physics, UCY)

Committee Member: Fotios Ptochos (Associate Prof. Dep. Of Physics, UCY)

Committee Member: George Fanourakis (Director of Research, Demokritos, Athens-Greece)

Committee Member: Spyros Tzamarias (Prof. Dep. Of Physics, Aristotle University of Thessaloniki – Greece)
DECLARATION OF DOCTORAL CANDIDATE

The present doctoral dissertation was submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy of the University of Cyprus. It is a product of original work of my own, unless otherwise mentioned through references, notes, or any other statements.

Tsiakkouri Demetra
Abstract

This dissertation focuses on the search for an exotic decay of the Higgs boson $h$ to a pair of light pseudoscalar bosons $a$, which further decay to a pair of $b$ quarks and a pair of $\tau$ leptons. This is allowed, as the couplings and properties of the recently discovered Higgs boson leave room up to 34% on the Higgs branching ratio to Beyond Standard Model (BSM) particles at the 95% CL.

This is the first time that the $2b2\tau$ channel is been investigated as a final state of the $h \rightarrow aa$ decay. Three $\tau\tau$ final states, $e\mu$, $e\tau_h$ and $\mu\tau_h$ are considered. The motivation is derived from theoretical models Beyond the Standard Model, (BSM) and more specifically the two Higgs Doublet Model complemented by a scalar Singlet, 2HDM+S. The 2HDM+S is built on four different types in which the Next to Minimal Supersymmetric Standard Model (NMSSM) is a special case.

The search has been performed using data collected by the CMS Experiment at the LHC in 2016, at a center-of-mass energy 13TeV, and corresponding to an integrated luminosity of 35.9 fb$^{-1}$. The light pseudoscalar boson mass probed is in the range between 15GeV and 60GeV. No excess of events above the SM expectations is observed. Upper limits are set on the 95% CL production cross section over the Standard Model one, times the branching ratio $B(h \rightarrow aa \rightarrow bb\tau\tau)$. They range from 3% to 12%. This corresponds to upper limits on the branching fraction of the Higgs boson decay to a pair of light pseudoscalar bosons, between 6% and 20% in the most favorable 2HDM+S scenarios.
Περίληψη

Η διδακτορική αυτή διατριβή επικεντρώνεται στην αναζήτηση της εξωτικής διάσπασης του μποζονίου Higgs σε ζεύγος ψευδοβαθμωτών μποζονίων α, με περαιτέρω διάσπαση τους σε ζεύγος b κουάρκς και ζεύγος των λεπτονίων. Αυτή η διάσπαση είναι επιτρεπτή, καθώς οι συζεύξεις και οι ιδιότητες του μποζονίου Higgs που ανακαλύφθηκε το 2012 αφήνουν χώρο εως και 34% στο λόγο διάσπασης του Higgs σε σωματίδια πέρα από το Καθιερωμένο Πρότυπο με 95% όριο εμπιστοσύνης.

Αυτή είναι η πρώτη φορά που το κανάλι 2b2τ ερευνήθηκε ως τελική κατάσταση της διάσπασης \( h \rightarrow aa \). Λαμβάνονται υπόψη μόνο τρεις τελικές καταστάσεις του ζεύγους των ταυ λεπτονίων: \( e\mu \), \( e\tau \) και \( \tau \). Το κίνητρο για τη μελέτη του συγκεκριμένου καναλιού προέρχεται από θεωρητικά μοντέλα περα από το Καθιερωμένο Πρότυπο, πιο συγκεκριμένα από το μοντέλο 2HDM+S. Το συγκεκριμένο μοντέλο χωρίζεται σε τέσσερις διαφορετικούς τύπους, με το NMSSM να αποτελεί μια ειδική περίπτωση.

Στην ανάλυση χρησιμοποιήθηκαν δεδομένα που συλλέχθηκαν από το Πείραμα CMS στο LHC το 2016, με ενέργεια κέντρου μάζας 13 TeV, και τα οποία αντιστοιχούν σε ολοκληρωμένη φωτεινότητα 35.9fb\(^{-1}\). Το εύρος των μαζών του ψευδοβαθμωτού μποζονίου α μελετάται από 15GeV μέχρι 60GeV. Ανώτερα όρια τίθενται στο λόγο της ενεργού διατομής παραγωγής του Higgs ως προς αυτή του Καθιερωμένου Προτύπου επί το λόγο διάσπασης, \( \sigma_h/\sigma_{SM} \times B(h \rightarrow aa \rightarrow bb\tau\tau) \), τα οποία κυμαίνονται από 3% έως 12%. Αυτά μεταφράζονται ως ανώτερα όρια στο λόγο διάσπασης του Higgs σε ζεύγος ψευδοβαθμωτών μποζονίων α μεταξύ 6% και 20% για την ενδιάμεση μάζα \( m_a = 40\text{GeV} \) στους ευνοϊκότερους τύπους του μοντέλου 2HDM+S.
Table of Contents

List of Figures I
List of Tables VI

CHAPTER 1
1. Introduction 1
   References 6

CHAPTER 2
2. Theoretical Framework 8
   2.1 Standard Model 8
   2.2 Electroweak Symmetry Breaking 12
   2.3 Higgs Boson Production Modes and Decay Channels 16
      2.3.1 SM Production Modes 16
      2.3.2 SM Higgs Decay Modes 19
   2.4 Limitations of Standard Model 22
   References 26

CHAPTER 3
3. Beyond the Standard Model (BSM) 28
   3.1 Supersymmetry – SUSY 28
   3.2 Minimal Supersymmetric Extension of Standard Model (MSSM) 30
   3.3 Two Higgs Doublet Model – 2HDM 33
   3.4 Two Higgs Doublet Models Plus a Scalar Singlet – (2HDM+S) 40
      3.4.1 Light Pseudoscalar Boson $a$ 43
   References 47

CHAPTER 4
4. LHC and CMS Experiment 48
   4.1 Large Hadron Collider LHC 48
   4.2 CMS Experiment 54
      4.2.1 Overview of CMS Detectors 54
      4.2.2 CMS Coordinate System 55
      4.2.3 Tracker 56
      4.2.4 Electromagnetic Calorimeter ECAL 58
      4.2.5 Hadron Calorimeter HCAL 61
      4.2.6 Magnet 63
      4.2.7 Muon System 63
      4.2.8 Trigger System 66
   References 69

CHAPTER 5
5. Event Generation and Reconstruction 70
   5.1 Event Generation 71
      5.1.1 Signal Event Generation 73
      5.1.2 Background Event Generation 74
   5.2 Pile-up Events 77
   5.3 Simulation of the Detector 78
   5.4 Reconstruction and Identification 79
      5.4.1 Muon Reconstruction 80
      5.4.1.1 Muon Identification 81
List of figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Caption</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Combined Results from CMS and ATLAS.</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Feynman diagram for the signal topology.</td>
<td>4</td>
</tr>
<tr>
<td>2.1</td>
<td>Shape of Higgs potential if $\mu^2 &lt; 0$ and $\lambda &gt; 0$.</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Feynman Higgs LO production diagrams: ggF, VBF, VH.</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Production cross section for a SM Higgs boson at the LHC.</td>
<td>18</td>
</tr>
<tr>
<td>2.4</td>
<td>Branching ratios of Higgs bosons and their uncertainties.</td>
<td>20</td>
</tr>
<tr>
<td>2.5</td>
<td>The invariant mass distribution of the diphoton candidates.</td>
<td>21</td>
</tr>
<tr>
<td>2.6</td>
<td>The evolution of coupling constants in SM and the two MSSM scenarios.</td>
<td>24</td>
</tr>
<tr>
<td>2.7</td>
<td>Feynman diagrams for one – loop quantum corrections of the Higgs squared mass.</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>Representation of the two angle performing the diagonalization of the mass matrices of the Higgs fields in 2HDM.</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>The branching ratios of pseudoscalar $a$ for Type I 2HDM+S.</td>
<td>44</td>
</tr>
<tr>
<td>3.3</td>
<td>The branching ratios of pseudoscalar $a$ for Type II 2HDM+S.</td>
<td>44</td>
</tr>
<tr>
<td>3.4</td>
<td>The branching ratios of pseudoscalar $a$ for Type III 2HDM+S.</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>The branching ratios of pseudoscalar $a$ for Type IV 2HDM+S.</td>
<td>45</td>
</tr>
<tr>
<td>3.6</td>
<td>Predicted $B(aa \rightarrow bb\tau\tau)$ for $m_a=40\text{GeV}$ in the different models of 2HDM+S, as a function of $\tan\beta$.</td>
<td>46</td>
</tr>
<tr>
<td>4.1</td>
<td>The overview of LHC circular underground tunnel with four main experiments.</td>
<td>49</td>
</tr>
<tr>
<td>4.2</td>
<td>Overview of the CERN acceleration complex.</td>
<td>50</td>
</tr>
<tr>
<td>4.3</td>
<td>The schematic view of beam collision at the point where two bunches interact.</td>
<td>52</td>
</tr>
<tr>
<td>4.4</td>
<td>Total integrated luminosity of proton beam collisions delivered by the LHC.</td>
<td>53</td>
</tr>
<tr>
<td>4.5</td>
<td>Overview of LHC schedule until 2037.</td>
<td>53</td>
</tr>
<tr>
<td>4.6</td>
<td>A schematic representation of the CMS Detector and its subdetectors.</td>
<td>55</td>
</tr>
<tr>
<td>4.7</td>
<td>The CMS coordinate system.</td>
<td>57</td>
</tr>
<tr>
<td>4.8</td>
<td>Sectional view of the tracker.</td>
<td>57</td>
</tr>
<tr>
<td>4.9</td>
<td>The CMS Pixel detector.</td>
<td>58</td>
</tr>
<tr>
<td>4.10</td>
<td>Sectional view of ECAL: the Barrel ECAL (EB), the Preshower (ES) in front of the Endcap ECAL (EE).</td>
<td>59</td>
</tr>
<tr>
<td>4.11</td>
<td>The ECAL energy resolution</td>
<td>61</td>
</tr>
<tr>
<td>4.12</td>
<td>Longitudinal view of quarter of the CMS hadronic calorimeter HCAL.</td>
<td>64</td>
</tr>
<tr>
<td>4.13</td>
<td>Longitudinal view of a quarter of the CMS detector illustrating the layout of the Muon System.</td>
<td>66</td>
</tr>
<tr>
<td>4.14</td>
<td>Overview of the CMS Level 1 Trigger.</td>
<td>67</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.1</td>
<td>The event generation processes begging from proton - proton collision.</td>
<td>72</td>
</tr>
<tr>
<td>5.2</td>
<td>Shapes (normalized to unity) of the signal produced in ggH, WH, and ZH productions.</td>
<td>74</td>
</tr>
<tr>
<td>5.3</td>
<td>Feynman’s diagrams for signal processes.</td>
<td>75</td>
</tr>
<tr>
<td>5.4</td>
<td>Feynman’s diagrams for some of the background processes.</td>
<td>76</td>
</tr>
<tr>
<td>5.5</td>
<td>Invariant mass of the muons in a $Z \rightarrow \mu\mu$ region.</td>
<td>77</td>
</tr>
<tr>
<td>5.6</td>
<td>Mean number of interactions per bunch crossing for the 2016</td>
<td>79</td>
</tr>
<tr>
<td>5.7</td>
<td>Reconstruction and identification efficiency for loose and tight muons.</td>
<td>83</td>
</tr>
<tr>
<td>5.8</td>
<td>Isolation efficiency of muons that pass the isolation requirements on top of the identification criteria using the Tag and Probe method.</td>
<td>84</td>
</tr>
<tr>
<td>5.9</td>
<td>Electron identification efficiency in data collected in 2016.</td>
<td>87</td>
</tr>
<tr>
<td>5.10</td>
<td>Schematic view of a heavy-flavor jet with a secondary vertex (SV) from the $b$-quark decay.</td>
<td>90</td>
</tr>
<tr>
<td>5.11</td>
<td>Misidentification probability as a function of $b$ jet identification for different $b$-tagging algorithms.</td>
<td>92</td>
</tr>
<tr>
<td>5.12</td>
<td>Misidentification probability as a function of $\tau_h$ identification efficiencies estimated for $Z \rightarrow \tau\tau$ and for QCD multijet events.</td>
<td>97</td>
</tr>
<tr>
<td>5.13</td>
<td>Misidentification probability versus $\tau_h$ identification using the two different isolation methods.</td>
<td>98</td>
</tr>
<tr>
<td>5.14</td>
<td>Efficiencies curves of $\tau_h$ identification as a function of $p_T$ of the reconstructed hadronic tau for different working points using simulated $Z/\gamma^* \rightarrow \tau\tau$ events</td>
<td>99</td>
</tr>
<tr>
<td>5.15</td>
<td>Misidentification probability as a function of the $p_T$ of generated electron for different Wps using simulated $Z/\gamma^* \rightarrow ee$ events.</td>
<td>100</td>
</tr>
<tr>
<td>5.16</td>
<td>HLT trigger efficiencies of $E_T^{miss}$ as a function of generated $E_T^{miss}$.</td>
<td>101</td>
</tr>
<tr>
<td>6.1</td>
<td>Postfit plots of the visible mass distribution for the pass (left) and fail (right) category using tag and probe method</td>
<td>108</td>
</tr>
<tr>
<td>6.2</td>
<td>Transverse momentum of the dimuon system in $Z \rightarrow \mu\mu$ events + 1 $b$ jet before and after correction.</td>
<td>111</td>
</tr>
<tr>
<td>6.3</td>
<td>Invariant mass of the dimuon system in $Z \rightarrow \mu\mu$ events + 1 $b$ jet after the correction on the $p_T$ of the dimuon system.</td>
<td>111</td>
</tr>
<tr>
<td>6.4</td>
<td>Invariant mass of the dimuon system and leading $b$ jet in $Z \rightarrow \mu\mu$ events + 1 $b$ jet before correction of $m_{b\mu\mu}$ but after correction of the dimuon $p_T$ spectrum.</td>
<td>112</td>
</tr>
<tr>
<td>6.5</td>
<td>Primary vertex distribution before and after pile reweighting in the $\mu\tau_h$ final state.</td>
<td>113</td>
</tr>
<tr>
<td>7.1</td>
<td>Transverse momentum distribution plots for $m_a = 20$ GeV.</td>
<td>117</td>
</tr>
<tr>
<td>7.2</td>
<td>Transverse momentum distribution plots for $m_a = 40$ GeV.</td>
<td>118</td>
</tr>
<tr>
<td>7.3</td>
<td>Transverse momentum distribution plots for $m_a = 60$ GeV.</td>
<td>119</td>
</tr>
<tr>
<td>7.4</td>
<td>Pseudorapidity distribution plots for $m_a = 20$ GeV.</td>
<td>120</td>
</tr>
<tr>
<td>7.5</td>
<td>Pseudorapidity distribution plots for $m_a = 40$ GeV.</td>
<td>121</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.6</td>
<td>Pseudorapidity distribution plots for $m_a = 60$ GeV.</td>
<td>122</td>
</tr>
<tr>
<td>7.7</td>
<td>2D efficiency plots based on different visible tau $p_T$ thresholds, $m_a = 20$ GeV.</td>
<td>124</td>
</tr>
<tr>
<td>7.8</td>
<td>2D efficiency plots based on different visible tau $p_T$ thresholds, $m_a = 40$ GeV.</td>
<td>125</td>
</tr>
<tr>
<td>7.9</td>
<td>2D efficiency plots based on different visible tau $p_T$ thresholds, $m_a = 60$ GeV.</td>
<td>126</td>
</tr>
<tr>
<td>7.10</td>
<td>$p_T, \eta$ distributions plots for each $b$ jet as well as the $\Delta R$ between the two $b$ jets at generated level for pseudoscalar $m_a = 20$ GeV</td>
<td>127</td>
</tr>
<tr>
<td>7.11</td>
<td>$p_T, \eta$ distributions plots for each $b$ jet as well as the $\Delta R$ between the two $b$ jets at generated level for pseudoscalar $m_a = 40$ GeV</td>
<td>128</td>
</tr>
<tr>
<td>7.12</td>
<td>$p_T, \eta$ distributions plots for each $b$ jet as well as the $\Delta R$ between the two $b$ jets at generated level for pseudoscalar $m_a = 60$ GeV</td>
<td>129</td>
</tr>
<tr>
<td>7.13</td>
<td>Comparison of the signal acceptance when requiring at least 1 or 2 $b$-tagged jets</td>
<td>130</td>
</tr>
<tr>
<td>7.14</td>
<td>Visible invariant mass of the leptons and the leading $b$ jet, for different pseudoscalar masses.</td>
<td>132</td>
</tr>
<tr>
<td>7.15</td>
<td>Visible invariant mass of the leptons and the leading $b$ jet, for background samples</td>
<td>133</td>
</tr>
<tr>
<td>7.16</td>
<td>Distribution of $m_{\text{vis}}^{\mu\tau}$ in the $\mu\tau_h$ final state.</td>
<td>134</td>
</tr>
<tr>
<td>7.17</td>
<td>Correlation between the visible mass of the tau candidates, and the visible mass of the tau candidates plus a $b$ tagged jet, in the $e\mu$ final state.</td>
<td>135</td>
</tr>
<tr>
<td>7.18</td>
<td>Correlation between the visible mass of the tau candidates, and the visible mass of the tau candidates plus a leading $b$ tagged jet, in the $e\tau_h$ final state.</td>
<td>136</td>
</tr>
<tr>
<td>7.19</td>
<td>Correlation between the visible mass of the tau candidates, and the visible mass of the tau candidates plus a leading $b$ tagged jet, in the $\mu\tau_h$ final state.</td>
<td>137</td>
</tr>
<tr>
<td>7.20</td>
<td>Illustration of the definition of parameters used in the $\zeta$ cut.</td>
<td>139</td>
</tr>
<tr>
<td>7.21</td>
<td>Distributions of $m_T(\mu, \tilde{p}<em>T^{\text{miss}})$, $m_T(\tau_h, \tilde{p}<em>T^{\text{miss}})$, and $D</em>\zeta$ in the $\mu\tau_h$ final state before the $m</em>{\text{vis}}$-based categorization.</td>
<td>140</td>
</tr>
<tr>
<td>7.22</td>
<td>Distributions of $m_T(\mu, \tilde{p}<em>T^{\text{miss}})$, $m_T(e, \tilde{p}<em>T^{\text{miss}})$, and $D</em>\zeta$ in the $e\mu$ final state before the $m</em>{\text{vis}}$-based categorization.</td>
<td>141</td>
</tr>
<tr>
<td>7.23</td>
<td>Distributions of $m_T(\mu, \tilde{p}<em>T^{\text{miss}})$, $m_T(e, \tilde{p}<em>T^{\text{miss}})$, and $D</em>\zeta$ in the $e\tau_h$ final state before the $m</em>{\text{vis}}$-based categorization.</td>
<td>142</td>
</tr>
<tr>
<td>7.24</td>
<td>$P_\zeta$ in the different categories of the $e\tau_h$ final state.</td>
<td>143</td>
</tr>
<tr>
<td>7.25</td>
<td>Transverse mass between the electron and the MET in the different categories of the $e\tau_h$ final state.</td>
<td>144</td>
</tr>
<tr>
<td>7.26</td>
<td>Transverse mass between the $\tau_h$ and the MET in the different categories of the $e\tau_h$ final state.</td>
<td>145</td>
</tr>
<tr>
<td>7.27</td>
<td>Transverse mass between the muon and the MET, in the different categories of the $\mu\tau_h$ final state.</td>
<td>146</td>
</tr>
<tr>
<td>7.28</td>
<td>Transverse mass between the $\tau_h$ and the MET, in the different categories of the $\mu\tau_h$ final state.</td>
<td>147</td>
</tr>
<tr>
<td>7.29</td>
<td>$P_\zeta$ in the different categories of the $\mu\tau_h$ final state.</td>
<td>148</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>7.30</td>
<td>Transverse mass between the muon and the MET in the different categories of the $e\mu$ final state.</td>
<td></td>
</tr>
<tr>
<td>7.31</td>
<td>Transverse mass between the electron and the MET in the different categories of the $e\mu$ final state.</td>
<td></td>
</tr>
<tr>
<td>7.32</td>
<td>$p_T$ in the different categories of the $e\mu$ final state.</td>
<td></td>
</tr>
<tr>
<td>7.33</td>
<td>Expected limits obtained in the $e\mu$ final state for the three signal regions, for the mass hypotheses 15, 40, and 60 GeV.</td>
<td></td>
</tr>
<tr>
<td>7.34</td>
<td>Expected limits obtained in the $e\tau_h$ final state for the three signal regions, for the mass hypotheses 15, 40, and 60 GeV.</td>
<td></td>
</tr>
<tr>
<td>7.35</td>
<td>Expected limits obtained in the $\mu\tau_h$ final state for the three signal regions, for the mass hypotheses 15, 40, and 60 GeV.</td>
<td></td>
</tr>
<tr>
<td>8.1</td>
<td>Jet $\to \tau_h$ fake rate fit to 1 prong, 1 prong + $\pi^0$, and 3 prong tau decay modes.</td>
<td></td>
</tr>
<tr>
<td>8.2</td>
<td>Jet $\to \tau_h$ fake rate for jets passing the very loose MVA isolation (denominator) to pass the tight MVA isolation (numerator), in the 1 prong, 1 prong + $\pi^0$, and 3 prong decay modes.</td>
<td></td>
</tr>
<tr>
<td>8.3</td>
<td>Prefit plots, in the $e\tau_h$ channel, of the ditau mass distributions in the same sign region without requirements on $m_\tau$ nor $p_T$.</td>
<td></td>
</tr>
<tr>
<td>8.4</td>
<td>Prefit plots, in the $\mu\tau_h$ channel, of the ditau mass distributions in the same sign region without requirements on $m_\tau$ nor $p_T$.</td>
<td></td>
</tr>
<tr>
<td>8.5</td>
<td>Prefit plots, in the $\mu\tau_h$ channel, of the ditau mass distributions in two categories, for both background estimations methods.</td>
<td></td>
</tr>
<tr>
<td>8.6</td>
<td>Prefit plots, in the $e\tau_h$ channel, of the ditau mass distributions in two categories, for both background estimations methods.</td>
<td></td>
</tr>
<tr>
<td>9.1</td>
<td>Poissonian distributions for different $\lambda$ parameters.</td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>A symmetric log-likelihood function - asymmetric log-likelihood function.</td>
<td></td>
</tr>
<tr>
<td>9.3</td>
<td>Log-normal and gamma probability density functions.</td>
<td></td>
</tr>
<tr>
<td>9.4</td>
<td>Distributions of probability density function under the background and signal hypothesis.</td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>Distribution of a Guassian probability density function.</td>
<td></td>
</tr>
<tr>
<td>9.6</td>
<td>Cumulative probability of the signal strengths.</td>
<td></td>
</tr>
<tr>
<td>10.1</td>
<td>Distributions of $m_{\tau\tau}^{vis}$ in the four categories of the $e\mu$ channel.</td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>Distributions of $m_{\tau\tau}^{vis}$ in the four categories of the $e\tau_h$ channel.</td>
<td></td>
</tr>
<tr>
<td>10.3</td>
<td>Distributions of $m_{\tau\tau}^{vis}$ in the four categories of the $\mu\tau_h$ channel.</td>
<td></td>
</tr>
<tr>
<td>10.4</td>
<td>Expected and observed 95% CL limits on $\sigma(h)/\sigma_{SM} B(H \to aa \to 2b2\tau)$ in %.</td>
<td></td>
</tr>
<tr>
<td>10.5</td>
<td>Comparison by category of expected 95% CL limits on $\sigma(h)/\sigma_{SM} B(H \to aa \to 2b2\tau)$ in %.</td>
<td></td>
</tr>
<tr>
<td>10.6</td>
<td>Observed 95% CL on $B(H \to aa)$ in 2HDM+S type-1, type-2, type-3, and type-4.</td>
<td></td>
</tr>
<tr>
<td>10.7</td>
<td>Observed 95% CL limits on $\sigma(h)/\sigma_{SM} B(H \to aa)$ for various scenarios of 2HDM+S.</td>
<td></td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>10.8</td>
<td>Observed and expected limits of $B(h \rightarrow aa)$ for different channels and 2HDM+S types, as derived from recent CMS analysis in Run 1 and in Run 2.</td>
<td>197</td>
</tr>
<tr>
<td>10.9</td>
<td>Expected 95% CL limits $(\sigma(h)/\sigma_{SM})B(h \rightarrow aa \rightarrow 2\tau2b)$, comparing different integrated luminosities.</td>
<td>198</td>
</tr>
<tr>
<td>10.10</td>
<td>Projected expected limits $(\sigma(h)/\sigma_{SM})B(h \rightarrow aa \rightarrow 2\tau2b)$, comparing different scenarios for systematic uncertainties for an integrated luminosity of 3000fb$^{-1}$.</td>
<td>199</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Caption</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>List of channels probed by CMS and ATLAS collaborations.</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Fermion content of the Standard Model.</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>The production cross sections of the SM Higgs boson with ( m_H = 125 \text{GeV} ).</td>
<td>17</td>
</tr>
<tr>
<td>2.3</td>
<td>The most principal decay channels for mass SM Higgs boson of ( m_H = 125 \text{GeV} ).</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Chiral and Vector Supermultiplets of the MSSM.</td>
<td>31</td>
</tr>
<tr>
<td>3.2</td>
<td>The four different types of the 2HDM model.</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>The Yukawa coupling factors in Yukawa interaction terms.</td>
<td>39</td>
</tr>
<tr>
<td>4.1</td>
<td>Performance parameters of LHC.</td>
<td>52</td>
</tr>
<tr>
<td>4.2</td>
<td>Main CMS magnet parameters.</td>
<td>64</td>
</tr>
<tr>
<td>5.1</td>
<td>Signal samples</td>
<td>76</td>
</tr>
<tr>
<td>5.2</td>
<td>MC background samples</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Electron MVA working point</td>
<td>87</td>
</tr>
<tr>
<td>5.4</td>
<td>Variables set as input for CSV algorithm for Run1 and Run2.</td>
<td>92</td>
</tr>
<tr>
<td>5.5</td>
<td>Tau decay modes</td>
<td>94</td>
</tr>
<tr>
<td>6.1</td>
<td>List of data sets included in the analysis.</td>
<td>107</td>
</tr>
<tr>
<td>6.2</td>
<td>Scale factors used to correct the efficiency of electron/muon faking hadronic tau.</td>
<td>109</td>
</tr>
<tr>
<td>6.3</td>
<td>List of parameters applied to recoil corrections</td>
<td>110</td>
</tr>
<tr>
<td>7.1</td>
<td>Acceptance of the ( t\bar{t} ) processes in the different final states.</td>
<td>131</td>
</tr>
<tr>
<td>7.2</td>
<td>Baseline selection criteria on the objects selected in the various final states.</td>
<td>131</td>
</tr>
<tr>
<td>7.3</td>
<td>Optimized selection and categorization in the various final states.</td>
<td>144</td>
</tr>
<tr>
<td>8.1</td>
<td>Background predictions and observed data.</td>
<td>162</td>
</tr>
<tr>
<td>8.2</td>
<td>Prefit background predictions and observed data in the various categories.</td>
<td>166</td>
</tr>
<tr>
<td>8.3</td>
<td>List of systematic uncertainties related to the physics objects and other uncertainties</td>
<td>171</td>
</tr>
<tr>
<td>8.4</td>
<td>List of systematic uncertainties related to background estimation.</td>
<td>172</td>
</tr>
<tr>
<td>10.1</td>
<td>The postfit yield predictions for background and the observed data in various categories.</td>
<td>189</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

In 2012 the CMS and ATLAS collaboration [1-3], discovered a particle, compatible with the Higgs boson which predicted within the Standard Model (SM) by Brout-Englert-Higgs mechanism [4-9]. This was a landmark since the Higgs boson was the last piece of the puzzle that was missing for the Standard Model (SM) theory completion. The SM has been the most successful particle physics theory, since it has passed with success several experimental tests in the last 50 years. To be sure that the observed particle was the same particle which was predicted by the Standard Model theory and it was responsible for the spontaneous electroweak symmetry breaking, one need to check for example, its self-coupling to be proportional to its mass or its couplings with fermions and gauge bosons to be proportional to their masses. The ATLAS and CMS collaborations made their analyses using the full set of data collected in Run 1, at the center of mass energies 7TeV and 8TeV and integrated luminosities of $4.8\text{fb}^{-1}$ and $20.7\text{fb}^{-1}$, respectively. Their results confirmed that the observed particle has indeed the properties of the Higgs boson predicted by the Standard Model.

The Higgs boson discovery play an essential role in SM theory but in parallel open the way for new physics. The SM is not a fully completed theory, since it cannot give answers to very well defined problems in particle physics. If there is any possibility for the observed Higgs boson to decay to beyond the SM particles, this indicates the presence of new and more extensive theory. The measuring of the Higgs couplings to fermions and to gauge bosons from both experiments CMS and ATLAS, set an upper limit on the rate of invisible Higgs Decays between $~20 – 60\%$ depending on the assumption [10]. Based on the latest combined results from CMS and ATLAS shown in Figure 1.1, the branching fraction of exotic decays of the Higgs boson to particles beyond the SM, $B(h\rightarrow BSM)$, is constrained to lower than $34\%$, at 95% confidence level (CL) [11].
Nonetheless, the above percentage gives a non negligible room for exotic Higgs boson decays and also indicates the way for new theories in particle physics. A significant part of those exotic Higgs boson decays is the decay of the Higgs particle to a pair of light pseudoscalar bosons with further decay to a pair of SM particles. Such type of process is allowed in various models, like the next-to-minimal supersymmetric standard model (NMSSM, which solved the μ-problem [12]) and the two Higgs doublet models augmented by a scalar singlet (2HDM+S) in which the NMSSM is a particular case. Seven scalar and pseudoscalar particles are predicted in the 2HDM+S model. In the decoupling limit of this model, one of the scalar particles, $h$, can be compatible with the discovered particle of mass 125 GeV, and another, the pseudoscalar $a$, can be light enough so that $h \rightarrow aa$ decays are allowed [13].

![Figure 1.1: The black solid line represent the observed and the dash blue line the expected negative log-likelihood scan of the branching fraction of Higgs boson, under the assumption of additional decay modes BSM to the Higgs boson width ($B_{BSM} \geq 0$). The red horizontal line at 3.84 indicates the log-likelihood variation corresponding to the 95% CL [11].](image)

The 2HDM+S is separated in four types, depending on how the Standard Model fermions interact with the two Higgs doublets and its forbid flavor changing neutral currents (FCNC) at the tree level. In each type, SM particles, quarks and leptons, couple in different combinations with the two doublets [13, 14]. More specifically, in type I, all particles couple with the first Higgs doublet, $\phi_1$ while in type II, up-type quarks couple with $\phi_1$ and down-type quarks and leptons couple with $\phi_2$. In type III (lepton specific) leptons couple with $\phi_2$, and quarks couple with $\phi_1$. Finally, leptons and up-type quarks couple with $\phi_2$, while down-type quarks couple with $\phi_2$. The types of the model along
with the pseudoscalar mass $m_a$ and the $\tan\beta$, which is the ratio of the vacuum expectation values of the two doublets, affect the branching fractions of the light pseudoscalars to SM particles.

In this thesis we present a search for exotic decay of the 125GeV Higgs boson to a pair of light pseudoscalar bosons $a$ with further decay to a pair of b quarks and a pair of tau leptons, $h \rightarrow aa \rightarrow bb\tau\tau$. In many theoretical models and more specific in the 2HDM, the decay of the light pseudoscalar boson to a pair of b quarks as well as its decay to a pair of tau leptons, are the most prevalent. This final states benefits from large branching fractions because of the large masses of tau leptons and b quarks with respect to other leptons and quarks. Therefore, studying the $h \rightarrow aa \rightarrow bb\tau\tau$ channel, benefits from relatively large branching fractions, but suffers from a low signal acceptance because of high trigger thresholds, low efficiency for the identification of $\tau$ leptons and b jets, and a very large background.

Theoretical studies have been done for this channel, since it has attracted the attention from several research groups from Tevatron and LHC [15, 16, 17]. Studies performed in Reference [17] for LHC at 14TeV, claim that with $100fb^{-1}$ data, a significance of 2 is possible ($S/\sqrt{B} \sim 2$). In this study, only leptonically decays of di-tau leptons in the final state were considered. If hadronically decays had been also considered, the results would have been much better but there are no studies in this direction. Similar studies were performed concerning W/Z associated production modes, for LHC at 14TeV. The results gave a very small signal over background ratio at $100fb^{-1}$ [15]. Studies at Tevatron [16] found a very small cross section $\sigma(Wh) \times Br(h \rightarrow aa \rightarrow bb\tau\tau)$ with assuming $Br(h \rightarrow aa) = 1$ (more than we expected), which led to a very small number of signal events. So, in order to observe the $h \rightarrow aa \rightarrow bb\tau\tau$ channel, a high integrated luminosity and a large cross-sections are necessary. Therefore, this channel is expected to play an important role in studies of the LHC at 14TeV, with large integrated luminosity and by combining the cross-sections from all production modes. This was the reason why this channel has not been studied previously.

Several searches for exotic Higgs decays to a pair of light pseudoscalar bosons are been performed by CMS and ATLAS collaborations but not with two b quark and two tau leptons in the final state. They are based on data collected at a center-of-mass energy 8TeV and 13TeV and involved different final states in different pseudoscalar mass, $m_a$, ranges. None of them has observed an excess relatively to the SM background but they have set limits on exotic decays of the Higgs boson to a pair of light pseudoscalar bosons. Table 1.1, represent the list of channels that are been studied in both collaborations for the different mass ranges.
This thesis presents the results from the first attempt ever done to analyze the $h \rightarrow aa \rightarrow bb\tau\tau$ channel. The analysis is based on the data recorded with the CMS detector in 2016 in proton-proton collision at a center-of-mass energy 13 TeV, corresponding to an integrated luminosity of 35.9 fb$^{-1}$. The mass range between 15 and 60 GeV are probed, in steps of 5 GeV. For low $m_a$ values, between the $b\bar{b}$ threshold and 15 GeV, the decay products of each of the pseudoscalar bosons become collimated, which would necessitate the use of special reconstruction techniques. The presence of light leptons originating from the $\tau$ decays allows events to be triggered in the dominant gluon fusion production mode (see figure 1.2). All three major production modes, gluon fusion, vector boson fusion VBF and W/Z associated production were considered, as well as a large number of background samples: $t\bar{t}$, Drell-Yan, W+jets, Single top, Diboson and Standard Model Higgs decays.

![Feynman diagram for the signal topology of $ggH \rightarrow aa \rightarrow bb\tau\tau$.](image)

The two tau in the final state can decay leptonically and hadronically. In this analysis three of the six di-tau final states are studied: $e\tau_h$, $\mu\tau_h$, and $e\mu$. The term $\tau_h$ denotes $\tau$ leptons decaying hadronically. The $\tau_h\tau_h$ channel is discarded because the trigger $p_T$ thresholds are too high to keep sufficient signal acceptance (at least $p_T > 40$ GeV offline for each $\tau_h$), and the ee and $\mu\mu$ channels are discarded because of their low branching fraction and high contribution of Drell-Yan background. For each final state different triggers or a combination of triggers are used. In order to increase the sensitivity of the analysis the events in each final state are divided in four categories depending on the signal over background ratio and after that a set of cuts are applied to reduce the dominant background. Systematic uncertainties related to background estimation, to physics objects and others are also considered.
The structure of this dissertation is the following. Chapter 2 makes a brief presentation of the Standard Model along with its limitations, leading to the need for introducing new theories. The two Higgs doublet models augmented by a scalar singlet (2HDM+S) is introduced in chapter 3 as one of the theoretical models that allows the certain process and analysis. Chapter 4 describes the Large Hadron Collider (LHC) and the apparatus of the Compact Muon Solenoid (CMS) detector. Chapter 5, after a brief presentation of the event generation processes, is followed by a description of physical object’s reconstruction used in this analysis. The simulated samples used for signal and background reconstruction along with their correction factors and the event selection are presented in Chapter 6. These are followed by Chapter 7 and 8, which describe the methodology used for increasing the analysis’s sensitivity. Chapter 9, describe the statistic tools useful to interpret the results of physics analyses. Finally, Chapter 10 is presenting the analysis results, which are already been published in Physics Letter B [27].

Table 1.1: List of channels probed by the CMS and ATLAS collaborations at a center-of-mass energy of 8TeV and 13TeV.

<table>
<thead>
<tr>
<th>Collaboration</th>
<th>Channels</th>
<th>Center-of-mass energy (TeV)</th>
<th>Pseudoscalar mass range (GeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS</td>
<td>$h \rightarrow aa \rightarrow 2\mu 2b$</td>
<td>8</td>
<td>$25.0 &lt; m_a &lt; 62.5$ GeV [18]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 2\mu 2\tau$</td>
<td>8</td>
<td>$15.0 &lt; m_a &lt; 62.5$ GeV [18]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4\tau$</td>
<td>8</td>
<td>$4 &lt; m_a &lt; 8$ GeV [19]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4\mu$</td>
<td>8</td>
<td>$5 &lt; m_a &lt; 15$ GeV [18]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 2\mu 2\tau$</td>
<td>13</td>
<td>$0.25 &lt; m_a &lt; 3.50$ GeV [20]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4\mu$</td>
<td>8</td>
<td>$15.0 &lt; m_a &lt; 62.5$ GeV [21]</td>
</tr>
<tr>
<td>ATLAS</td>
<td>$h \rightarrow aa \rightarrow 4\mu$</td>
<td>8</td>
<td>$15 &lt; m_a &lt; 60$ GeV [22]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4e$</td>
<td>8</td>
<td>$15 &lt; m_a &lt; 60$ GeV [22]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 2e 2\mu$</td>
<td>8</td>
<td>$15 &lt; m_a &lt; 60$ GeV [22]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4\gamma$</td>
<td>8</td>
<td>$10 &lt; m_a &lt; 62$ GeV [23]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 2\mu 2\tau$</td>
<td>8</td>
<td>$3.7 &lt; m_a &lt; 50.0$ GeV [24]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4b$</td>
<td>13</td>
<td>$20 &lt; m_a &lt; 60$ GeV [25]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4\mu$</td>
<td>13</td>
<td>$1 &lt; m_a &lt; 60$ GeV [26]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 4e$</td>
<td>13</td>
<td>$1 &lt; m_a &lt; 60$ GeV [26]</td>
</tr>
<tr>
<td></td>
<td>$h \rightarrow aa \rightarrow 2e 2\mu$</td>
<td>13</td>
<td>$1 &lt; m_a &lt; 60$ GeV [26]</td>
</tr>
</tbody>
</table>
References


CHAPTER 2

THEORETICAL FRAMEWORK

2.1 STANDARD MODEL

The Standard Model, SM is a gauge theory and is invariant under the $\text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y$ symmetry group. $\text{SU}(2)_L \times \text{U}(1)_Y$ describes the electroweak interaction and $\text{SU}(3)_C$ describes the strong interactions between fermions. The $C$ subscript refers to the color-charge, $L$ to the weak interactions between the left handed particles and $Y$ stands for the weak hypercharge. In general, in a local gauge theory the Lagrangian remains invariant under a group of local transformations. Invariance is achieved by the introduction of one or more additional fields in the Lagrangian density with a consequence being the release of gauge bosons and their couplings.

The SM Lagrangian density is formed based on experimentally observed particles and their interactions. The strategy followed to obtain the SM Lagrangian is summarized below, beginning from the Dirac Lagrangian density which describes the constituents of matter by a fermionic field $\psi$:

$$\mathcal{L}_D = \bar{\psi} \left( i \gamma^\mu \partial_\mu - m \right) \psi$$ (2.1)

where $\bar{\psi}$ represents the anti-fermion field, and $\gamma^\mu$ are the Dirac matrices\(^1\).

Requiring the Lagrangian to be invariant under the local phase transformation $U(1)$, the $\partial_\mu$ derivative is replaced by the covariant derivative $D_\mu$:

$$D_\mu = \partial_\mu + ig \frac{\tau^a}{2} A^a_\mu$$ (2.2)

\(^1\) The Dirac matrices are defined by $\{ \gamma^\mu, \gamma^\nu \} = 2g^{\mu\nu}$ with the Minkowski metric $g^{\mu\nu}$ on space-time. The notation $\{a,b\} = ab + ba$, denotes the anti-commutator.
where $A_\mu$ is a vector gauge field that corresponds to a photon, the gauge boson which is the carrier of electromagnetic interaction. The Lagrangian then becomes:

$$\mathcal{L}_D = \bar{\psi} \left(i \gamma^\mu \partial_\mu - g \gamma^\mu \frac{\tau}{2} A_\mu - m \right) \psi$$

(2.3)

The term now with $A_\mu$ corresponds to the coupling between the interacting gauge field and the fermion field. The parameter $g$ is interpreted as the interaction strength or coupling, while $\tau$ are the generators that can be expressed as Pauli matrices$^2$.

Unifying the electromagnetic and the weak interaction (Glashow, Salam and Weinberg in the 1960’s) led to an electroweak sector in the Lagrangian. The symmetry group of the unified electroweak force is the $SU(2)_L \times U(1)_Y$. First, requiring gauge invariance under the non-Abelian$^3$ group $SU(2)_L$, a new quantum number, namely the weak isospin $T_3$ and three gauge fields $W^1_\mu, W^2_\mu, W^3_\mu$ are introduced. The weak isospin $T_3$ is related to the electric charge and the weak hypercharge under the relation:

$$Q = T_3 + \frac{Y}{2}$$

(2.4)

Secondly, the gauge invariance under the Abelian group $U(1)_{EM}$ introduced another single field $B_\mu$. Based on the aforesaid new fields, the covariant derivative from (2.2) becomes:

$$D_\mu = \partial_\mu + ig \frac{\tau^a}{2} W^a_\mu + ig' \frac{Y}{2} B_\mu$$

(2.5)

where $g$ and $g'$ stand for the coupling constants of the $SU(2)_L$ and $U(1)_{EM}$ respectively. The $\tau^a/2$ are the generators that can be expressed as half of the Pauli matrices, with $\alpha = 1, 2, 3$ and satisfying the Lie algebra:

$$[\tau^a, \tau^b] = i \epsilon^{abc} \tau^c \quad \text{and} \quad [\tau^a, Y] = 0$$

(2.6)

$^2$ Pauli matrices: $\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

$^3$ A non-Abelian group is defined as a group in which the commutator of the generators does not vanish but is given by linear combination of the generators, while in the Abelian group the commutator of the generator vanishes (e.g.: (2.6))
The gauge fields $B_\mu$ and $W^{\alpha}_\mu$ are related to the experimentally observable massive $W^\pm$ and $Z^0$ bosons and the massless photon:

\[ W^\pm_\mu = \frac{1}{2} (W^1_\mu \mp i W^2_\mu) \]  

\[ Z_\mu = W^3_\mu \cos \theta_W - B_\mu \sin \theta_W \]  

\[ A_\mu = W^3_\mu \sin \theta_W + B_\mu \cos \theta_W \]

$A_\mu$ represents the photon field, while $\theta_W$ is the Weinberg mixing angle which is defined as:

\[ \tan \theta_W = \frac{g'}{g} \]  

or

\[ \sin \theta_W = \frac{g'}{\sqrt{g'^2 + g^2}} \]

The electron charge $e$ and the Fermi constant $G_F$ can be expressed as a function of the $W$ boson mass, the coupling constant $g$ and the Weinberg mixing angle as below:

\[ e = g \sin \theta_W \quad \text{or} \quad e = g' \cos \theta_W \]  

\[ G_F = \frac{\sqrt{2} g^2}{8 m^2_W} \]

Continuing with the unification of the electroweak with the strong interaction, the SM gauge group is completed. The strong interaction is defined by the SU(3)$_C$ gauge group and remains invariant under this group, with the introduction of 8 massless gauge bosons $G^b_\mu$. Those gauge bosons are known as gluons which are vector particles that carry a spin 1 and are the carriers of the strong interaction.
The covariant derivative of equation (2.5) with the new term formed is indicated below:

\[
D_\mu = \partial_\mu + ig \frac{T^a}{2} W^a_\mu + ig Y B_\mu + ig_3 \frac{\lambda^b}{2} G^b_\mu
\]  

(2.14)

where \( b = [1,2, \ldots, 8] \), \( g_3 \) is the coupling constant of the \( SU(3)_C \) and \( \lambda^b \) are the Gell-Mann matrices.

Table 2.1 below summarizes the fermion content together with their representation under the different groups of symmetry and their associated electroweak charges.

<table>
<thead>
<tr>
<th>Fields ((\psi))</th>
<th>SU(2)(_L) representation</th>
<th>SU(3)(_C) representation</th>
<th>( T_3 )</th>
<th>( Y/2 )</th>
<th>( Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_\ell )</td>
<td>([v_e, v_\mu, v_\tau]_L )</td>
<td>2</td>
<td>1</td>
<td>1/2</td>
<td>-1/2</td>
</tr>
<tr>
<td>( \ell_R )</td>
<td>([e_R, \mu_R, \tau_R] )</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>( v_\ell )</td>
<td>([v_e, v_\mu, v_\tau]_R )</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( u_q )</td>
<td>([u_d, c_L, t_L] )</td>
<td>2</td>
<td>3</td>
<td>1/2</td>
<td>1/6</td>
</tr>
<tr>
<td>( d_q )</td>
<td>([u_d, c_L, s_L] )</td>
<td>2</td>
<td>3</td>
<td>-1/2</td>
<td>1/6</td>
</tr>
<tr>
<td>( q_R )</td>
<td>([u_R, c_R, t_R] )</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>2/3</td>
</tr>
<tr>
<td>( q_R )</td>
<td>([d_R, s_R, b_R] )</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>-1/3</td>
</tr>
</tbody>
</table>

So far the SM Lagrangian density can be expressed as a sum of gauge and fermion Lagrangian densities describing the interactions between fermions and gauge bosons:

\[
\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{gauge}} + \mathcal{L}_f
\]  

(2.15)

The gauge Lagrangian density \( \mathcal{L}_{\text{gauge}} \) regroups the gauge fields of all three symmetries consisting of the kinetic energy terms of the gauge fields:
\[ \mathcal{L}_{\text{gauge}} = -\frac{1}{4} G^i_{\mu \nu} G^{i \mu \nu} - \frac{1}{4} W^i_{\mu \nu} W^{i \mu \nu} - \frac{1}{4} B^i_{\mu \nu} B^{i \mu \nu} \]  

(2.16)

where the tensors are:

\[ G^i_{\mu \nu} = \partial_{\mu} G^{i \nu}_{\nu} - \partial_{\nu} G^{i \mu}_{\mu} - g_3 f_{ijk} G^{j \mu}_{\mu} G^{k \nu}_{\nu} , \text{ with } i, j, k = 1, \ldots, 8 \]  

(2.17)

\[ W^i_{\mu \nu} = \partial_{\mu} W^{i \nu}_{\nu} - \partial_{\nu} W^{i \mu}_{\mu} - g f_{ijk} W^{j \mu}_{\mu} W^{k \nu}_{\nu} , \text{ with } i, j, k = 1, 2, 3 \]  

(2.18)

\[ B^i_{\mu \nu} = \partial_{\mu} B^{i \nu}_{\nu} - \partial_{\nu} B^{i \mu}_{\mu} \]  

(2.19)

The fermionic Lagrangian part \( \mathcal{L}_f \) consisting of the kinetic terms of fermions is:

\[ \mathcal{L}_f = i \bar{\psi}_L \gamma^\mu D^\mu \psi_L + i \bar{\psi}_R \gamma^\mu D^\mu \psi_R \]  

(2.20)

The subscripts L and R denote the left-handed and right-handed particles and the dirac spinors of particles fields \( \psi_i \), are represented by \( \ell, \ e, \ u, \ d, \ q \). The \( \gamma^\mu \) are matrices\(^4\) that can be represented as a function of Pauli matrices and the covariant derivatives \( D^\mu \) are defined as follows:

\[ D^\mu = \partial^\mu - ig^l \frac{\lambda^a}{2} G^a_{\mu} - ig \frac{\tau^b}{2} W^b_{\mu} - ig^l Y^a_{\mu} B^a \]  

(2.21)

### 2.2 Electroweak Symmetry Breaking

All particles described by the previous Lagrangian remain massless as there are no mass terms included. This happens due to the fact that the theory remains invariant under the group \( \text{SU}(3)_C \times \text{SU}(2)_L \times \text{U}(1)_Y \), which is considered as a global symmetry. Although, it is noted that it is possible to give particles their observed masses by breaking the symmetry locally with the theory remaining gauge invariant. The solution was proposed by R.Brout, F.Englert and P.W.Higgs via the Brout-Englert-Higgs (BEM) mechanism in 1964 [6-8]. The BEH mechanism allows for the generation of

---

\(^4\) The \( \gamma^\mu \) matrices are defined by the following two properties: \( [ \gamma^\mu, \gamma^\nu ] = 2 \eta_{\mu \nu} \), \( ( \gamma^\mu )^\dagger = \gamma^0 \gamma^\mu \gamma^0 \)
mass terms via the spontaneous breaking of the electroweak symmetry. For this, a complex scalar $SU(2)_L$ doublet is introduced with real fields:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i \phi_2 \\ \phi_3 + i \phi_4 \end{pmatrix} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$  \hspace{1cm} (2.22)$$

and lead to a new term in the Langrangian density:

$$L_{BEH} = (D^\mu \Phi)^\dagger D_\mu \Phi - V(\Phi, \Phi^*)$$  \hspace{1cm} (2.23)$$

with the potential to have the following form:

$$V(\Phi, \Phi^*) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$$  \hspace{1cm} (2.24)$$

The $\mu$ parameter is a real constant representing a mass parameter and $\lambda$ is a dimensionless parameter standing for the self-interaction strength.

The above potential has an infinite set of minimal or ground states. An arbitrary choice of the ground state breaks the $SU(2)_L \times U(1)_Y$ symmetry by keeping the gauge invariant of the Lagrangian. After the spontaneous electroweak symmetry breaking a random ground state has been taken:

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \upsilon \end{pmatrix}$$  \hspace{1cm} (2.25)$$

where $\upsilon$ is the vacuum expectation value obtained after minimizing the Higgs potential:

$$\upsilon = \sqrt{-\frac{\mu^2}{\lambda}}$$  \hspace{1cm} (2.26)$$

Figure (2.1) illustrates the shape of the potential if $\mu^2 < 0$ and $\lambda > 0$. The potential has a minimum for:

$$\Phi^\dagger \Phi = \frac{-\mu^2}{2\lambda}$$  \hspace{1cm} (2.27)$$
Figure 2.1: Shape of Higgs potential if $\mu^2 < 0$ and $\lambda > 0$.

The vacuum expectation value is not zero and according to the Goldstone theorem, fields that acquire non-zero vacuum expectation values should have an associated massless Goldstone boson. Keeping in mind that photon is massless, the symmetry should be taken in such a way as the fields with zero electric charge acquire a vacuum expectation value. After expanding around the minimum of the theory so as to keep the symmetry, the complex field $\Phi$ becomes equal to:

$$
\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix}
$$

(2.28)

The field $h(x)$ is gauge invariant and fluctuates away from the vacuum state and subconstituent in the Lagrangian density leading to the appearance of mass term for the gauge bosons. More specifically, the $h(x)$ as a complex scalar field carries four degrees of freedom. Three of them gave mass to the $W^\pm$ and the $Z$ bosons via the spontaneous symmetry breaking mechanism and the other gave mass to the new Higgs field. In this way, this mechanism gave masses for the gauge boson fields, while the photon remains massless:

$$
M_\gamma = 0, \quad M_H = \sqrt{2\lambda\nu^2}, \quad M_W = \frac{1}{2}\nu g, \quad M_Z = \frac{1}{2}\nu \sqrt{g'^2 + g^2} = \frac{M_W}{\cos \theta_W}
$$

(2.29)

From the above equation one can see that there is a relation between the mass of gauge bosons $W$ and $Z$. The mass ratio of those two bosons is equal to the cosine of the Weinberg angle.
This expression is often expressed as the so-called ρ-(Veltman) parameter:

\[
\rho = \frac{M_W^2}{M_Z^2 \cos^2 \theta_W} = 1
\]  

(2.30)

The current measurements of the weak W± and Z bosons masses, 80.39 GeV and 91.18 GeV respectively, confirm this relation [9].

Also, from equations (2.13) and (2.29) the vacuum expectation value is given by:

\[
v = \frac{1}{\sqrt{\sqrt{2} G_F}} \approx 246 \text{ GeV}
\]  

(2.31)

Additionally, the free parameter \( \lambda \) obtained from the recent measurements of the Higgs boson mass at the LHC experiments, implies that \( \lambda = 0.129 \) and \( |\mu|^2 = (88.4 \text{ GeV})^2 \).

The gauge bosons obtain mass by introducing in the Lagrangian a new Higgs field, but the fermions still remain massless. This happens because in the Lagrangian terms like \( m \bar{\psi} \psi \) are not allowed, as they are not invariant under \( \text{SU}(2)_L \times \text{U}(1)_Y \). However, the Higgs mechanism can give mass to fermions as to gauge boson in a similar way.

Using the complex Higgs doublet introduced previously in the Higgs mechanism, the mass terms for fermions can be expressed in the following way:

\[
\mathcal{L}_{Yuk} = -\lambda_f \left[ \bar{\psi}_L \Phi \psi_R + \bar{\psi}_R \Phi^* \psi_L \right]
\]  

(2.32)

where \( \lambda_f \) denotes a new series of coupling constants for each fermion \( f \), the so called Yukawa couplings.

The above Lagrangian describes an interaction between the Higgs field and the fermion, and implies that the fermions will acquire a finite mass if the \( \Phi \)-doublet has a non-zero expectation value, as in equation (2.28).
For example:

$$\mathcal{L}_e = -\frac{\lambda_e}{\sqrt{2}} \left[ (\bar{\nu}, \bar{\nu})_L \left( 0, v + h \right) e_R + \bar{e}_R (0, v + h) \left( \nu \right) e_L \right] = -\frac{\lambda_e (v + h)}{\sqrt{2}} \left[ e_R e_R + \bar{e}_R \bar{e}_L \right]$$

$$\mathcal{L}_e = -\frac{\lambda_e}{\sqrt{2}} \bar{e} e - \frac{\lambda_e}{\sqrt{2}} h \bar{e} e$$

(2.33)

The first term gives the electron mass, $$m_e = \frac{\lambda_e v}{\sqrt{2}}$$ and the second demonstrates the electron-higgs interaction ($$\frac{\lambda_e}{\sqrt{2}} m_e$$), which denotes that Yukawa couplings are proportional to the mass of the fermions. However, the fermions mass cannot be predicted since the Yukawa couplings are free parameters. Combining all of the above, the SM Lagrangian density can be decomposed as the sum of four different terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_{BEH} + \mathcal{L}_{Yuk}$$

(2.34)

describing the interaction between fermions and gauge bosons and predicting masses for them.

### 2.3 Higgs Boson Production Modes and Decay Channels

#### 2.3.1 SM Production Modes

The production mechanism of the Higgs boson at the hadron colliders are mainly four: the gluon–gluon fusion (ggF), the vector boson fusion (VBF), the associated production with a vector boson (VH) and associated production with heavy quarks like top and bottom quarks ($q\bar{q}H$). Figure (2.2) presents the Feynman diagrams of the three main production processes and figure (2.3) are showcases of the production cross sections of the Higgs boson as a function of the Higgs mass at center of mass energy of 14TeV. Table 2.2 from [21] presents the production cross sections for the SM Higgs boson with $m_{H}=125$GeV in proton-proton collisions as a function of the center of mass energy.
The gluon-gluon fusion (ggF) is the dominant production mechanism at the LHC. It is mediated by the exchange of a virtual quark loop (top or bottom quark) and has a production cross section of 49.47 pb for \( m_H = 125 \text{GeV} \) at center of mass energy 14 TeV\(^5\). Vector boson fusion (VBF) is the second most dominant production mode and has a cross section of one order of magnitude lower than ggF. However, at high values of the Higgs mass it becomes compatible with ggF. This process gives a very clear experimental signature due to the two jets in the final state with kinematic characteristics, such as the their large invariant mass.

**Table 2.2:** The production cross sections of the SM Higgs boson with \( m_H = 125 \text{GeV} \) in proton-proton collisions as a function of the center of the mass energy. The center of mass energy of 1.96 GeV refers to the Tevatron and 7, 8, 13 and 14 TeV to LHC energies [21].

<table>
<thead>
<tr>
<th>( \sqrt{s} ) (TeV)</th>
<th>ggF</th>
<th>VBF</th>
<th>WH</th>
<th>ZH</th>
<th>( ttH )</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.96</td>
<td>0.95(^{+17%}_{-17%})</td>
<td>0.065(^{+8%}_{-7%})</td>
<td>0.13(^{+8%}_{-8%})</td>
<td>0.079(^{+8%}_{-8%})</td>
<td>0.004(^{+10%}_{-10%})</td>
<td>1.23</td>
</tr>
<tr>
<td>7</td>
<td>15.3(^{+10%}_{-10%})</td>
<td>1.24(^{+2%}_{-2%})</td>
<td>0.58(^{+3%}_{-3%})</td>
<td>0.34(^{+4%}_{-4%})</td>
<td>0.09(^{+8%}_{-14%})</td>
<td>17.5</td>
</tr>
<tr>
<td>8</td>
<td>19.5(^{+10%}_{-11%})</td>
<td>1.60(^{+2%}_{-2%})</td>
<td>0.70(^{+3%}_{-3%})</td>
<td>0.42(^{+5%}_{-5%})</td>
<td>0.13(^{+8%}_{-13%})</td>
<td>22.3</td>
</tr>
<tr>
<td>13</td>
<td>44.1(^{+11%}_{-11%})</td>
<td>3.78(^{+2%}_{-2%})</td>
<td>1.37(^{+2%}_{-2%})</td>
<td>0.88(^{+5%}_{-5%})</td>
<td>0.51(^{+9%}_{-13%})</td>
<td>50.6</td>
</tr>
<tr>
<td>14</td>
<td>49.7(^{+11%}_{-11%})</td>
<td>4.28(^{+2%}_{-2%})</td>
<td>1.51(^{+2%}_{-2%})</td>
<td>0.99(^{+5%}_{-5%})</td>
<td>0.61(^{+9%}_{-13%})</td>
<td>57.1</td>
</tr>
</tbody>
</table>

The other two processes have very small cross sections. Despite its low cross section, the associated production of Higgs boson with a Vector boson (VH) (or Higgsstrahlung) is very promising for low Higgs masses (<135 GeV). This is due to leptons and quarks coming from the vector boson decays and are possible to trigger them and discriminate the VH signal from the QCD background. The associated production with the pair of quarks (top quarks) has the smallest cross section which makes it unaccessible experimentally in Run-1. On the other hand it is going to play an important role in the low Higgs mass range and is expecting to be seen in Run-2, where the luminosity and the center of mass energy are higher than in Run-1. It can provide a clear environment to identify the Higgs boson mass and the top/bottom quark and therefore provide information about the Yukawa couplings between quarks and the Higgs boson. In this dissertation, except from the two dominant

---

\(^5\) The cross section is computed at NNLO+NNLL QCD and NLO EW
processes the last two processes are also taken into account, in association with particular Higgs Decay modes for a clear signature and completion.

**Figure 2.2:** Feynman Higgs LO production diagrams: gluon-gluon fusion (left), Vector boson fusion (VBF) (middle) and Vector bosons (W, Z) associated production (Higgsstrahlung) (right) [10].

**Figure 2.3:** Production cross section for a SM Higgs boson at the LHC as a function of the Higgs mass.[11]
2.3.2 SM Higgs Decay Modes

For the interpretation of all the experimental results, it is essential to make the calculation of all the relevant Higgs decays widths, and estimate their uncertainties. Unstable particles like Higgs are short-lived particles which decays into several other particles via different decay modes. The rate of the decay into a mode $i$ is given by the partial width, $\Gamma_i$. The total width $\Gamma$ is given by the sum of all partial widths. The fraction of a partial width for a certain final state over the total width is known as branching ratio, BR.

The Higgs boson with mass of about 125GeV and a very short lifetime ($\tau_H \sim 10^{-22}\text{s}$) can only be observed through its decay products. The branching (BR) of the Higgs boson is given by the following equation:

$$B(H \rightarrow XX) = \frac{\Gamma(H \rightarrow XX)}{\sum_{Y \in SM} \Gamma(H \rightarrow YY)}$$

(2.35)

The numerator represents the decay width of the Higgs to an X, SM particle and the denominator is the sum of the decay widths of all the possible Higgs decay products.

The decay width takes a different form depending on the final decay product. If the Higgs decays to fermions is based on the Born approximation, the decay width is then given by:

$$\Gamma(H \rightarrow \bar{f} f) = \frac{N_C}{8\pi v^2} m_H m_f^2 \sqrt{1-x} \cdot \delta_v \cdot R(x), \text{ where } x = \frac{4 M_f^2}{m_H^2}$$

(2.36)

and if the Higgs decays to an off-shell gauge boson then it is given by:

$$\Gamma(H \rightarrow VV^*) = \frac{3 M_V^3}{32\pi v^4} m_H \delta'_V \cdot R(x)$$

(2.37)

---

6 The Higgs boson decay to gauge bosons is forbidden below the WW and ZZ threshold.
where $\delta' = 1, \delta'_z = \frac{7}{12} - \frac{10}{9} \sin^2 \theta_w + \frac{40}{27} \sin^4 \theta_w$ and

$$R(x) = \frac{3(1 - 8x + 20x^2)}{\sqrt{4x - 1}} \cos \left( \frac{3x - 1}{2x^{3/2}} \right) - \frac{1}{2x}(2 - 13x + 47x^2) - \frac{3}{2}(1 - 6x + 4x^2) \ln(x).$$

Comparing the above decay widths with the masses of gauge bosons and fermions predicted by the SM theory and Higgs mechanism and the recently measured Higgs mass, there is a very interesting variety of accessible decay branching ratios. Figure (2.4) shows the branching ratio (BR) of the SM Higgs boson as a function of its mass and the different decay channels that account for over 99% of the total width. The most principle decay channels for the SM Higgs boson as a function of the Higgs boson mass are listed in table 2.5.

From figure 2.4 one can see that the Higgs interactions are relative to the mass of the fermion or boson pairs. The decay modes to a pair of fermions (low Higgs mass) or bosons (high Higgs mass) are generally dominating the Higgs mass range. The Higgs main decay mode is the $H \rightarrow b \bar{b}$ in a mass range between 80 and 135 GeV. This channel could give a clear signature, however at LHC it suffers from a large background coming from the direct production of $pp \rightarrow b \bar{b} + X$. Though, there are many techniques used for b-tagging that help to reducing the different backgrounds.

**Figure 2.4:** Branching ratios of Higgs bosons and their uncertainties as a function of Higgs mass for low mass range (left plot) and for a larger range (right plot) [14].
The invariant mass distribution of the diphoton candidates with each event weighted by the ratio of S/(S+B) in each event category by CMS [22].

**Figure 2.5:** The invariant mass distribution of the diphoton candidates with each event weighted by the ratio of S/(S+B) in each event category by CMS [22].

The second largest BR decay channel is the $H \rightarrow \tau \tau$. This channel, due to its further decay to lighter leptons (leptonic decay) or to jets from hadronic decays can be considered as an interesting channel. The Higgs boson decay to a pair of c-quarks $H \rightarrow \bar{c}c$ is a very difficult channel to be distinguished from a large QCD dijets events and, on the other hand, the $H \rightarrow \mu \mu$ which is a very clean channel has a very small BR.

The Higgs to a pair of photons $H \rightarrow \gamma \gamma$ also gives a very clear signature due to the two isolated and highly energetic photons. It is one of the two channels that led to the discovery of the SM-like Higgs boson in 2012 by the CMS and ATLAS collaborations [12-13]. Figure 2.3 shows the invariant $m_{\gamma\gamma}$ distribution weighted by the ratio of signal-to-background in each category events for the combined data of 7TeV and 8TeV and for the integrated luminosities of 5.1 fb$^{-1}$ and 19.7 fb$^{-1}$ respectively. This figure shows a clear signal in the diphoton channel at the mass of 124.7 GeV, with
significance of $5.7\sigma$ and led to the experimental observation for the first time of the SM Higgs Boson [22], [24].

The diboson decay modes experimentally are very good channels as they provide a very good sensitivity because they give leptonic and hadronic final states. More specifically, the so-called golden channel $H \rightarrow ZZ^* \rightarrow 4\mu$ leads to a narrow peak over the relatively small background and it was the second channel that led to the Higgs boson discovery [23], [24].

Table 2.3: The most principal decay channels for mass SM Higgs boson of $m_H = 125\text{GeV}$ [21].

<table>
<thead>
<tr>
<th>Decay channel</th>
<th>Branching ratio</th>
<th>Rel. uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H \rightarrow \gamma\gamma$</td>
<td>$2.27 \times 10^{-3}$</td>
<td>$+5.0%$ $-4.9%$</td>
</tr>
<tr>
<td>$H \rightarrow ZZ$</td>
<td>$2.62 \times 10^{-2}$</td>
<td>$+4.3%$ $-4.1%$</td>
</tr>
<tr>
<td>$H \rightarrow W^+W^-$</td>
<td>$2.14 \times 10^{-1}$</td>
<td>$+4.3%$ $-4.2%$</td>
</tr>
<tr>
<td>$H \rightarrow \tau^+\tau^-$</td>
<td>$6.27 \times 10^{-2}$</td>
<td>$+5.7%$ $-5.7%$</td>
</tr>
<tr>
<td>$H \rightarrow b\bar{b}$</td>
<td>$5.84 \times 10^{-1}$</td>
<td>$+3.2%$ $-3.3%$</td>
</tr>
<tr>
<td>$H \rightarrow Z\gamma$</td>
<td>$1.53 \times 10^{-3}$</td>
<td>$+9.0%$ $-8.9%$</td>
</tr>
<tr>
<td>$H \rightarrow \mu^+\mu^-$</td>
<td>$2.18 \times 10^{-4}$</td>
<td>$+6.0%$ $-5.9%$</td>
</tr>
</tbody>
</table>

2.4 Limitations of the Standard Model

Despite its success, the Standard Model cannot provide answers to many open questions in particle physics. There are indications to believe that the Standard Model is not the ultimate theory of elementary particles but the first part of a more global theory. Those indications are derived either from direct observations or from conceptual problems of the SM. For instance, the SM contains 19 free parameters [15]. Could a fundamental theory contain a so large number of free parameters? Why are there three generations of matter particles and why are the masses of elementary particles so different? In addition, the SM cannot explain the baryon asymmetry or the existence of Dark matter and Dark Energy. Another issue is that the SM cannot incorporate gravity or unify the gauge couplings of the three fundamental forces at some larger scale (the Grand Unified Theory (GUT) scale).
In the early 20th century, many astrophysics observations and measurements of different galaxies led to the conclusion that the universe must be filled with large quantities of non-radiating materials. The latest predictions estimates that we are surrounding by matter that represents only the 4.5% of the mass content of the universe. All the other ~26% is dark matter and the remaining ~70% dark energy [16-17]. The SM does not predict any candidate for the Dark matter. Supersymmetry theory came to fill that part with the Lightest Supersymmetric Particle (LSP) as a suitable candidate.

Based on the SM predictions the neutrino are massless and interact weakly with other particles. However, observations by many collaborations, like the Super-Kamiokande, showed that neutrino oscillate from one flavor to another [18]. This implies that neutrino could not be massless. Until now, the masses of neutrino have not been accurately measured, but higher limits for their masses have been set and the difference between their masses has been measured.

Following the demonstration of electricity and magnetism to a unified electromagnetic force and similarly, the late 1960’s demonstration of electromagnetic and weak interactions as unified electroweak force, many people believed to a Grand Unified Theory (GUT). More specifically, they believed that there should be a unification of the strong and electroweak interaction at the very high energy scale of $\sim 10^{16}$ GeV. The SM arbitrary selection of the $\text{SU}(3)_c \times \text{SU}(2)_L \times \text{U}(1)_Y$ symmetry group implies this unification and it also makes precise predictions of the running coupling constants of the Strong, Electromagnetic and weak forces (2.39):

$$
\alpha_1 = \frac{5}{3} \frac{\alpha_{cm}}{\cos^2\theta_w}, \quad \alpha_2 = \frac{\alpha_{em}}{\sin^2\theta_w}, \quad \alpha_3 = a_s \approx 1, \quad \alpha_3^{\text{TeV}} = 137
$$

The strength of the three forces shows an apparent large disparity around the electroweak scale, however at higher energies (as shown in figure 2.5), the coupling constants tend to have compatible strengths. The electromagnetic ($\alpha_2$) and the weak forces ($\alpha_1$) can be unified in a so-called electroweak interaction, but the strong coupling ($\alpha_3$) does not meet the two others [19]. Those running couplings can be modified by the addition of a new particle as in some theoretical models, like MSSM (Minimal SuperSymmetry Model) and become the same at a very high energy scale (fig.2.6).
The Hierarchy problem is one of the main conceptual problems of the SM. This problem, is derived from the fact that radiation corrections of the Higgs boson mass are much larger than the Higgs mass. In more detail the radiation corrections arise from the Higgs boson couplings to fermions, to gauge bosons and self-couplings:

\[ m_H = m_{H}^{\text{bare}} + \Delta m_{H}^{\text{fermion}} + \Delta m_{H}^{\text{gauge}} + \Delta m_{H}^{\text{Higgs}} + \ldots \]  

and they are proportional to the square of ultraviolet cut off scale \( \Lambda_{\text{UV}} \):

\[ \left( \Delta m_H^2 \right)^f = -\frac{1}{8 \pi^2} \lambda_f^2 \left( \Lambda_{\text{UV}}^2 - 6 m_f^2 \ln \frac{\Lambda_{\text{UV}}}{m_f} \right) \]  

\[ \left( \Delta m_H^2 \right)^s = -\frac{1}{16 \pi^2} \lambda_s^2 \left( \Lambda_{\text{UV}}^2 - 2 m_s^2 \ln \frac{\Lambda_{\text{UV}}}{m_s} \right) \]  

\( \Lambda_{\text{UV}} \) is a high energy scale.

**Figure 2.6:** The dashed lines show the evolution of the coupling constants in the SM and the solid lines shows two MSSM scenarios in which the coupling constants are unified in a high energy scale [19].
Equation 2.41 and 2.42 represent the corrections to the mass of the scalar boson coming from fermions and from scalar particles respectively [20]. The same corrections are presented schematically via Feynman diagrams in figure 2.7.

**Figure 2.7:** Feynman diagrams for one – loop quantum corrections of the Higgs squared mass parameter $m_H^2$, due to fermions $f$ (left diagram) and due to scalar $S$ [19].

Consequently, if the SM theory is valid up to the Plank scale (~$10^{19}$ GeV) or up to a GUT scale of $10^{16}$ GeV, then the Higgs mass corrections is many decades of order larger than its bare mass:

\[
(\Delta m_H^2) \approx 10^{38} \text{GeV}^2 \Rightarrow m_H^2 - 10^{38} \approx 10^4 \text{GeV}^2
\]  

That implies about 34 orders of magnitude to be cancelled, since experimental data shows that the Higgs boson mass is ~ 125 GeV (~ $10^2$) and to do that it will need a lot of fine-tuning. Hence, the ultraviolet cut-off $\Lambda_{\text{UV}}$ should be much lower than the Planck scale, leading to Beyond the Standard Model (BSM) theories that will cancel the $\Lambda^2$ divergence in a way such as to avoid the fine-tuning.
References


CHAPTER 3

BEYOND THE STANDARD MODEL (BSM)

After the discovery of the Higgs boson from the CMS and ATLAS collaborations in 2012, the LHC experiments have studied the properties of this new particle. Each experiment of the LHC probed different channels and measured the mass, coupling constants and the spin parity so as to confirm that the new particle is compatible with the SM one. As mentioned in the previous chapter, the SM does not predict the Higgs boson mass since mass is a free parameter. Although, when the mass is specified, the cross section and branching ratios are predicted by the theory.

From the recent results of the ATLAS and CMS experiments and the limitations of the SM theory, the existence of new physics beyond the SM (BSM) is strongly suggested. There are many theories that have been probed, like Supersymmetry (SUSY), minimal Supersymmetric Standard Model (MSSM) and two Higgs Doublets Models (2HDM), which extend the SM by adding one or more multiplets. In this chapter, we make a brief reference to the theoretical models of SUSY and MSSM and we continue with the study of the 2HDM as a general form which will lead us to the extension of 2HDM augmented by complex scalar singlet (2HDM + S).

3.1 Supersymmetry – SUSY

Supersymmetry (SUSY) is one of the theoretical models proposed for extending the Standard model and providing answers to many questions that the Standard Model leaves open (see Section 2.5). SUSY is based on the assumption of an underlying symmetry between bosons and fermions [1],[2]. For each SM particle there is a superpartner, the so-called s-particle. All the observed particles are placed in supermultiplets, with their superpartners having all quantum numbers the same except the spin, which differs by $\frac{1}{2}$. More precisely, for each fermion there is a bosonic superpartner and for each SM boson there is a fermionic superpartner. Therefore, SUSY introduces an operator $Q$ which denotes the relationship between fermions and bosons:

$$Q |\text{boson}\rangle = |\text{fermion}\rangle \quad (3.1)$$

$$Q |\text{fermion}\rangle = |\text{boson}\rangle \quad (3.2)$$
SUSY should be a space-time symmetry and, in order to avoid the possibility of parity violation, it must satisfy some commutational and anti-commutational algebra:

\[
\{Q, Q^\dagger\} = p^\mu, \quad \{Q, Q\} = \{Q^\dagger, Q^\dagger\} = 0, \quad \{P^\mu, Q\} = \{P^\mu, Q^\dagger\} = 0
\]  

(3.3)

where \(P^\mu\) represents the four momentum generator of space-time.

SUSY is not yet experimentally confirmed, but if sometime its existence is verified, it is going to resolve many problems in particle physics. The introduction of superpartners solves theoretically the hierarchy problem. From equations 2.41 and 2.42, the corrections to the mass of the Higgs boson are coming from fermions which have an opposite sign than the corrections coming from the scalar particles. SUSY, by associating a new boson for each fermion and vice versa, cancels the quadratic divergences \(\Lambda_{UV}^2\) for each pair. It also predicts the unification of the gauge-couplings at a large scale, as shown in fig. 2.5.

SUSY has a side-effect, namely the violation of lepton and baryon numbers. For that reason a new quantum number \(R\) was introduced:

\[
R = (-1)^{2S + 3(B - L)}
\]  

(3.4)

where \(S\), \(B\) and \(L\) are the spin, baryon and lepton numbers respectively. For all SM particles \(R = +1\) and for s-particles \(R = -1\), meaning that the superparticles must be produced in pairs and implying that the lightest supersymmetric particle (LSP) must be stable. This LSP proposed by SUSY is a neutral particle which does not interact via the strong or electromagnetic force and this makes it an excellent dark matter candidate.

Although SUSY is a very good theory that closes gaps from SM, it does not predict the exact masses of the SUSY particles since after the symmetry breaks, a large number of free parameters are introduced. In order to constrain these parameters, simpler models were introduced, one of them being the Minimal Supersymmetric Standard Model (MSSM) \([3]\).
3.2 Minimal Supersymmetric Extension of the Standard Model (MSSM)

The Minimal Supersymmetric Standard model (MSSM) is a theoretical model extending the Standard Model (SM) by incorporating SUSY. It is an extension of $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge theory. All the physical particles in MSSM are separated in two types of supermultiplets. These are the chiral (matter) supermultiplet and the vector (gauge) supermultiplet [4]. In the chiral supermultiplets, one Weyl fermion with spin $\frac{1}{2}$ and a complex scalar field are included. In the vector supermultiplets, one vector boson with spin 1 and a Weyl fermion are included. As in the case of SUSY, for every single field in the SM there is a need for a superpartner field. The fermions partners are the scalar sleptons or squarks, the fermionic partners of the Higgs are the Higgsinos and the partners of the gauge bosons are the gauginos.

The Higgs boson, as it has a zero spin, its needs a different treatment. The MSSM introduces two extra doublets for each of the two Higgs doublet fields, the so-called Higgsinos. The Higgsinos are the fermionic superpartners with hypercharges $+1/2$ and $-1/2$. Table 3.1 shows the particle content of the MSSM.

After electroweak symmetry breaking, the $W^0$ and the $B^0$ gauge eigenstates are mixing together leading to two mass eigenstates, the $Z^0$ and $\gamma$. In the same way, their superpartners $\tilde{W}^0$ and $\tilde{B}^0$ mix to give the zino ($\tilde{Z}^0$) and photino ($\tilde{\gamma}^0$) mass eigenstates respectively.

Similarly, the bino $\tilde{B}^0$ and winos ($\tilde{W}^\pm, \tilde{W}^0$) eigenstates mix with the Higgsinos ($\tilde{H}^+_u \tilde{H}^0_u, \tilde{H}^+_d \tilde{H}^0_d$) in a linear combination that leads to six mass eigenstates. The two of them are results of the mixing with the charged component of Higgsinos, the so-called charginos $\tilde{\chi}^{\pm}$, and the other four are outcome of the mixing with the neutral components of Higgsinos, called neutralinos $\tilde{\chi}^0$.

The mass term for the neutralino $\tilde{\chi}^0$ is given by:

$$M_{\tilde{\chi}^0} = \begin{pmatrix} M_1 & 0 & M_2 \cos \beta \sin \theta_w & -M_2 \sin \beta \sin \theta_w \\ 0 & M_2 & -M_2 \cos \beta \cos \theta_w & M_2 \sin \beta \cos \theta_w \\ M_2 \cos \beta \sin \theta_w & -M_2 \cos \beta \cos \theta_w & 0 & -\mu \\ -M_2 \sin \beta \sin \theta_w & M_2 \sin \beta \cos \theta_w & -\mu & 0 \end{pmatrix}$$ (3.5)
where $M_1$ and $M_2$ are the $U(1)_Y$ and $SU(2)_L$ gauginos mass terms, respectively, and $\mu$ is the supersymmetric Higgsino mass parameter. The $M_w$ and $M_z$ are the $W$ and $Z$ boson masses respectively.

The neutralinos eigenstates are written as:

\[
\begin{pmatrix}
\tilde{\chi}^0_1 \\
\tilde{\chi}^0_2 \\
\tilde{\chi}^0_3 \\
\tilde{\chi}^0_4
\end{pmatrix}
= N
\begin{pmatrix}
\tilde{B}_L \\
\tilde{W}^3_L \\
\tilde{H}^+_1 \\
\tilde{H}^0_1
\end{pmatrix},
\begin{pmatrix}
\tilde{\chi}^0_1 \\
\tilde{\chi}^0_2 \\
\tilde{\chi}^0_3 \\
\tilde{\chi}^0_4
\end{pmatrix}
= N^* \begin{pmatrix}
\tilde{B}_R \\
\tilde{W}^3_R \\
\tilde{H}^+_1 \\
\tilde{H}^0_1
\end{pmatrix},
\tag{3.6}
\]

where $N$ is a unitary matrix which satisfies the relation:

\[
N^* M_n N^{-1} = \text{diag}(m_{\tilde{\chi}^0_1}, m_{\tilde{\chi}^0_2}, m_{\tilde{\chi}^0_3}, m_{\tilde{\chi}^0_4})
\tag{3.7}
\]

Table 3.1: Chiral and Vector Supermultiplets of the MSSM.
The lightest neutralino is the $\tilde{\chi}_4^0$ and is been proposed by the MSSM as an excellent candidate for Dark matter.

The chargino mass terms are given by:

$$M_{\chi_{\pm}} = \begin{pmatrix} M_2 & -M_w \sqrt{2}\sin \beta \\ -M_w \sqrt{2}\cos \beta & \mu \end{pmatrix}$$  \hspace{1cm} (3.8)$$

with eigenstates:

$$\begin{pmatrix} \tilde{\chi}_{1L}^+ \\ \tilde{\chi}_{2L}^+ \end{pmatrix} = Z_L \begin{pmatrix} \tilde{W}_L^+ \\ \tilde{H}_{2L}^+ \end{pmatrix}, \hspace{1cm} \begin{pmatrix} \tilde{\chi}_{1R}^+ \\ \tilde{\chi}_{2R}^+ \end{pmatrix} = Z_R \begin{pmatrix} \tilde{W}_R^+ \\ \tilde{H}_{2R}^+ \end{pmatrix}$$  \hspace{1cm} (3.9)$$

The $Z_L$ and $Z_R$ are such as to satisfy:

$$Z_R M C Z_L^\dagger = \begin{pmatrix} m_{\chi_1^\pm} & 0 \\ 0 & m_{\chi_2^\pm} \end{pmatrix}$$  \hspace{1cm} (3.10)$$

One has:

$$m_{\chi_{1,2}}^2 = \frac{1}{2} \left[ \mu^2 + M_2^2 + 2 M_w^2 \sqrt{\left( \mu^2 - M_Z^2 \right)^2 + 4 M_w^2 (\mu^2 + 2 \mu M_2 \sin \beta + M_2^2) + 4 M_w^4 \cos^2 2 \beta} \right]$$  \hspace{1cm} (3.11)$$

and

$$Z_{L,R} = \begin{pmatrix} \cos \phi_{L,R} & \sin \phi_{L,R} \\ -\sin \phi_{L,R} & \cos \phi_{L,R} \end{pmatrix}$$  \hspace{1cm} (3.12)$$

with

$$\tan 2\phi_L = \frac{2 M_w \sqrt{2} \mu \cos \beta + M_2 \sin \beta}{\mu^2 - M_2^2 - 2 M_w^2 \cos^2 2 \beta}$$  \hspace{1cm} (3.13)$$

$$\tan 2\phi_R = \frac{2 M_w \sqrt{2} \mu \sin \beta + M_2 \cos \beta}{\mu^2 - M_2^2 + 2 M_w^2 \cos^2 2 \beta}$$  \hspace{1cm} (3.14)$$

The scalar sector of the MSSM includes one CP-odd Higgs boson $A^0$, two CP-even Higgs bosons $H^0$, $h^0$ and the charged Higgs Bosons $H^\pm$. 
The masses of the neutral and charged scalars are given below:

\[ m_A^2 = m_{12}^2 \frac{\sin \beta}{\cos \beta} \]  
\[ m_{h, h}^2 = \frac{1}{2} \left( m_A^2 + m_Z^2 \pm \sqrt{(m_A^2 + m_Z^2)^2 - 4 m_A^2 m_Z^2 \cos^2 2\beta} \right) \]  
\[ m_{h^\pm}^2 = m_A^2 + m_W^2 \]

where \( m_{12} \) is a real coefficient, observed in (3.20).

Furthermore, one can notice that the neutral and charged scalar masses depend only on the mass of the CP-odd bosons and \( \tan \beta \). Therefore, the \( h \) mass can be parametrized as a function of \( m_A \) and \( \tan \beta \), leading to different regions of the parameter’s space. A large region of those parameters space is studied at LHC (7TeV and 8TeV), expected to be fully covered by the data at 14TeV with the very high luminosity expected to reach 300fb\(^{-1}\). More details about the formalism and phenomenology of MSSM can be found in [5] and [6].

### 3.3 Two Higgs Doublet Model - 2HDM

The 2HDM is a simple extension of the Higgs sector of the SM [7]. In this model, instead of one Higgs doublet, two SU(2) Higgs doublet fields \( \Phi_1 \) and \( \Phi_2 \), are introduced. Each field interacts with the fermions and also self interacts via a proper potential. The two Higgs doublet fields have hypercharge \( Y = 1 \) and satisfy the condition of \( \rho \) parameter, \( \rho = 1 \):

\[ \rho = \frac{\sum_{i=1}^{n} [I_i (I_i + 1) - \frac{1}{4} Y_i^2] v_i}{\sum_{i=1}^{n} \frac{1}{2} Y_i^2 v_i} \]  

(3.18)

The above equation is the same with equation 2.30. The value of \( \rho \)-Veltman parameter equal to 1, is been experimentally verified with more than 1% accuracy even without including one-loop corrections. The value of this parameter is a restriction for new models beyond the Standard Model.

The most general form of the scalar potential \( V \) which spontaneously breaks SU(2)\(_L\) \( \times \) U(1)\(_Y\) down to U(1)\(_{EM}\) is given by the following equation:
\[ V = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) \\
+ \frac{1}{2} \lambda_1 (\Phi_1^\dagger \Phi_1)^2 + \frac{1}{2} \lambda_2 (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) \\
+ \frac{1}{2} \lambda_5 (\Phi_2^\dagger \Phi_2)^2 + \lambda_6 (\Phi_1^\dagger \Phi_1) + \lambda_7 (\Phi_2^\dagger \Phi_2) | \Phi_1^\dagger \Phi_2 + h.c. \] (3.19)

Combined with the assumption that CP is conserved in the Higgs sector, the quartic terms of either of the two fields are absent, with all the remaining coefficient to be real

\[ \lambda_1 , \lambda_2 , \ldots \ldots , \lambda_6 , \]

\[ m_{11}^2 , m_{22}^2 , m_{12}^2 . \]

More specifically the symmetry enforced \( \lambda_6 = \lambda_7 = 0 \), and the form of the above potential can be simplified as follows:

\[ V = m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) + \frac{1}{2} \lambda_1 (\Phi_1^\dagger \Phi_1)^2 \\
+ \frac{1}{2} \lambda_2 (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_1^\dagger \Phi_1)(\Phi_2^\dagger \Phi_2) + \frac{\lambda_5}{2} [ (\Phi_1^\dagger \Phi_2)^2 + (\Phi_2^\dagger \Phi_1)^2 ] \] (3.20)

Both Higgs doublets acquire vacuum expectation values \( v_1 \) and \( v_2 \) respectively without affecting the SM gauge bosons mass, since the vacuum expectation value \( v \) is replaced by :

\[ u = \sqrt{v_1^2 + v_2^2} \] (3.21)

with the ratio of vacuum expectation values of the two doublets to be:

\[ \tan \beta = \frac{v_1}{v_2} \] (3.22)

The minimization from the potential gives two minima of the \( \Phi_1 \) and \( \Phi_2 \):

\[ \langle \Phi_1 \rangle_0 = \begin{pmatrix} 0 \\ \frac{v_1}{\sqrt{2}} \end{pmatrix}, \quad \langle \Phi_2 \rangle_0 = \begin{pmatrix} 0 \\ \frac{v_2}{\sqrt{2}} \end{pmatrix} \] (3.23)

---

1 If CP is not conserved then the parameters \( \lambda_5 \) and \( m_{12}^2 \) are complex.

2 By enforced \( \lambda_6 = \lambda_7 = 0 \), Higgs-fermion couplings are restrict and no FCNCs are present. This can be achieved by imposing a discrete symmetry \( (\Phi_1, \Phi_2) \rightarrow (-\Phi_1, \Phi_2) \) on the model.
Working in the same way as in the SM higgs mechanism, i.e. expanding around the minima, the above two complex scalar SU(2) doublets become:

$$\Phi_a = \left( \frac{\phi_a^+ + (v_a + \rho_a + i \eta_a)}{\sqrt{2}} \right), \quad a \in \{1, 2\} \quad (3.24)$$

There are eight fields, two complex and four real degrees of freedom. Three of those get “eaten” up by the $W^\pm_\mu$ and the $Z^0_\mu$ gauge bosons after electroweak symmetry breaking. The other five are “Higgs” fields. The second complex field yields the charged Higgs bosons $H^\pm$, and the other three reals yield a neutral pseudoscalar Higgs boson $A$ (CP-odd Higgs boson) and two neutral scalar Higgs bosons $h, H$ (CP-even Higgs bosons).

To determine the mass matrices, the covariant derivative over each field are set equal to zero:

$$\frac{\partial V}{\partial \Phi_1} = \frac{\partial V}{\partial \Phi_2} = 0 \quad (3.25)$$

The Charged scalar mass matrix is given below:

$$\begin{pmatrix} M_{\Phi} \end{pmatrix}_{ij} = \begin{pmatrix} \frac{\partial^2 V}{\partial \phi_1 \partial \phi_j} \end{pmatrix} = \begin{vmatrix} m_{12}^2 - (\lambda_4 + \lambda_5) v_1 v_2 & -1 \\ -1 & \frac{v_2}{v_1} \end{vmatrix} \quad (3.26)$$

This mass matrix has a zero eigenvalue corresponding to the charged Goldstone boson which gets eaten by the $W^\pm$ gauge bosons. The mass of the W bosons becomes:

$$M_W = \frac{g}{2} \sqrt{v_1^2 + v_2^2} \quad (3.27)$$

The mass of the charged Higgs is given by:

$$m_H^2 = \left[ m_{12}^2 / v_1 v_2 - (\lambda_4 + \lambda_5) \right] (v_1^2 + v_2^2) \quad (3.28)$$

One of the real degrees of freedom yields one neutral pseudoscalar mass eigenstate:
with the following neutral pseudoscalar Higgs mass matrix:

\[
(M_A^2) = \frac{\partial^2 V}{\partial \eta_i \partial \eta_j} = \left( \begin{array}{cc}
\frac{m_{12}^2}{\nu_1 \nu_2} - \lambda_5 & \nu_2 \\
-\nu_1 \nu_2 & \nu_1^2
\end{array} \right)
\] (3.30)

By diagonalizing this, one can find a massless pseudoscalar boson which gets eaten by the $Z^0$ boson giving a mass of:

\[
M_{Z^0} = \sqrt{\left( g'^2 + g^2 \right) \left( \nu_1^2 + \nu_2^2 \right)}
\] (3.31)

and one massive pseudoscalar with a mass of:

\[
m_A^2 = \left[ m_{12}^2 / (\nu_1 \nu_2) - \lambda_5 \right] (\nu_1^2 + \nu_2^2)
\] (3.32)

The last mass matrix is the matrix for the neutral scalar Higgs and is written as:

\[
(M_h^2) = \frac{\partial^2 V}{\partial \rho_i \partial \rho_j} = \left( \begin{array}{cc}
A & C \\
C & B
\end{array} \right)
\] (3.33)

where:

\[
A = m_{11}^2 + \frac{3 \lambda_1}{2} \nu_1^2 + \frac{\lambda_{345}}{2} \nu_2^2, \quad \lambda_{345} = \lambda_3 + \lambda_4 + \lambda_5
\] (3.34)

\[
B = m_{22}^2 + \frac{3 \lambda_2}{2} \nu_2^2 + \frac{\lambda_{345}}{2} \nu_1^2
\] (3.35)

\[
C = -m_{12}^2 + \lambda_{345} \nu_1 \nu_2
\] (3.36)

and this can be diagonalized by an angle $\alpha$, resulting in the physical states $h$ and $H$:  

\[
A = [\eta_1 \nu_2 - \eta_2 \nu_1] (\nu_1^2 + \nu_2^2)^{-\frac{1}{2}}
\] (3.29)
\[ h = (\rho_1 \sin \alpha - \rho_2 \cos \alpha) \quad \text{and} \quad H = (-\rho_1 \cos \alpha - \rho_2 \sin \alpha) \]  

(3.37)

Then, the masses of the two neutral scalar Higgs are:

\[ M_h = v^2 \left[ \lambda - \frac{\hat{\lambda} \cos (\beta - \alpha)}{\sin (\beta - \alpha)} \right] \]  

(3.38)

\[ M_H = v^2 \left[ \lambda + \frac{\hat{\lambda} \sin (\beta - \alpha)}{\cos (\beta - \alpha)} \right] \]  

(3.39)

where:

\[ \lambda = \lambda_1 \cos^4 (\beta) + \lambda_2 \sin^2 (\beta) + \frac{1}{2} \lambda_{345} \sin^2 (2\beta) \]  

(3.40)

\[ \hat{\lambda} = \frac{1}{2} \sin (2\beta) [\lambda_1 \cos^2 (\beta) - \lambda_2 \sin^2 (\beta) - \lambda_{345} \cos (2\beta)] \]  

(3.41)

The \( \alpha \) parameter is the rotation angle performing the diagonalization of the neutral scalar mass matrix and \( \beta \), as defined in equation (3.22), is the angle used for diagonalizing the charged scalar and the neutral pseudoscalar mass matrices. Based on [8], one can choose \( \beta \) in the first quadrant and the angle in the first or in the fourth quadrant \((-\pi/2 \leq \alpha \leq \pi/2)\) as illustrated in figure (3.1).

\[ \text{Figure 3.1:} \text{ The schematic representation of the two angles performing the diagonalization of the mass matrices of the Higgs fields in 2HDM.} \]
These two parameters, $\alpha$ and $\beta$, are constituent of the Yukawa couplings of the gauge bosons and fermions in the 2HDM model. If we allow the most general form of Yukawa couplings, the flavor changing neutral currents are enhanced, in contrast with the SM in which there are no tree-level FCNC\(^3\). To avoid the flavor changing neutral currents at the tree-level, the discrete $Z_2$ symmetry is introduced:

$$Z_2 : \quad \Phi_1 \rightarrow \Phi_1, \quad \Phi_2 \rightarrow -\Phi_2$$

(3.42)

which ensures that the fermions with the same quantum numbers couple only with one Higgs field as it was formalized by the Paschos-Glashow-Weinberg theorem [9].

By imposing the $Z_2$ symmetry, four types of fermion couplings result, being the four different types of the 2HDM. Each type, as shown in table 3.2, depends on how the leptons, up-type, and down-type quarks couple with the two Higgs fields $\Phi_1$ and $\Phi_2$. In type I (SM-like), all the fermions couple with $\Phi_2$ and in type II, the fermions and the down-type quarks couple with $\Phi_1$ while up-type quarks couple with $\Phi_2$. In type III, the so-called “lepton-specific”, the leptons couple with $\Phi_1$ whereas quarks couple with $\Phi_2$. Lastly, in type IV (flipped), the down type quarks couple with $\Phi_1$ and the leptons and up-type quarks couple with $\Phi_2$.

Table 3.2: Shows the four different types of couplings in the 2HDM model and the Higgs fields to which the fermions couple with.

<table>
<thead>
<tr>
<th>Fermions\Types</th>
<th>Type I</th>
<th>Type II</th>
<th>Type III (lepton specific)</th>
<th>Type IV (flipped)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ell$</td>
<td>$\Phi_2$</td>
<td>$\Phi_1$</td>
<td>$\Phi_1$</td>
<td>$\Phi_2$</td>
</tr>
<tr>
<td>$u$</td>
<td>$\Phi_2$</td>
<td>$\Phi_2$</td>
<td>$\Phi_2$</td>
<td>$\Phi_2$</td>
</tr>
<tr>
<td>$d$</td>
<td>$\Phi_2$</td>
<td>$\Phi_1$</td>
<td>$\Phi_2$</td>
<td>$\Phi_1$</td>
</tr>
</tbody>
</table>

The interactions between fermions, Higgs bosons and self-interactions are described by Yukawa Lagrangian terms. The Lagrangian below expresses the Yukawa interactions in terms of mass eigenstates of the Higgs bosons as:

$$L^{2\text{HDM}}_{\text{Yukawa}} = - \sum_{f=u,d,\ell} \left( \frac{m_f}{\sqrt{2}} \bar{\psi}_b \gamma^i \bar{f} h + \frac{m_f}{\sqrt{2}} \bar{\psi}_h \gamma^i \bar{f} H - i \frac{m_f}{\sqrt{2}} \bar{\psi}_A \gamma^5 \bar{f} Y_5 f A \right)$$

\(^3\) Diagonalizing the mass matrix in the SM automatically diagonalizes the Yukawa interactions and therefore there are no tree-level FCNC. Actually FCNC are suppressed due to the Glashow-Iliopoulos-Maiani mechanism,[10]. Presence of FCNC can cause phenomenological difficulties [11], [12].
\[-\left\{ \frac{\sqrt{2} V_{ud}}{\mu} (m_u \xi_u^d P_L + m_d \xi_d^d P_R) d H^+ + \frac{\sqrt{2} m_f \xi_f^f}{\sqrt{2} m_f} H^0 \ell_R H^+ + h.c. \right\} \] (3.43)

where the factors $\xi_\phi^f$ represent the Yukawa couplings and $P_{L/R}$ represent the projection operators\(^4\) for left and right handed fermions. Table (3.3) shows that the factors $\xi_\phi^f$ are functions of both $\alpha$ and $\beta$ parameters and depend on the types of 2HDM.

**Table 3.3**: The Yukawa coupling factors in Yukawa interaction terms of equation (3.43). The coefficient of the couplings of the charged scalar Higgs $H^+$ are the same as those of the pseudoscalar $A$.

<table>
<thead>
<tr>
<th></th>
<th>Type I</th>
<th>Type II</th>
<th>Type III (Lepton-specific)</th>
<th>Type IV (Flipped)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\xi_h^u$</td>
<td>$\cos \alpha / \sin \beta$</td>
<td>$\cos \alpha / \sin \beta$</td>
<td>$\cos \alpha / \sin \beta$</td>
<td>$\cos \alpha / \sin \beta$</td>
</tr>
<tr>
<td>$\xi_d^d$</td>
<td>$\cos \alpha / \sin \beta$</td>
<td>$-\sin \alpha / \cos \beta$</td>
<td>$\cos \alpha / \sin \beta$</td>
<td>$-\sin \alpha / \cos \beta$</td>
</tr>
<tr>
<td>$\xi_l^f$</td>
<td>$\cos \alpha / \sin \beta$</td>
<td>$-\sin \alpha / \cos \beta$</td>
<td>$-\sin \alpha / \cos \beta$</td>
<td>$\cos \alpha / \sin \beta$</td>
</tr>
<tr>
<td>$\xi_H^u$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\sin \alpha / \sin \beta$</td>
</tr>
<tr>
<td>$\xi_H^d$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\cos \alpha / \cos \beta$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\cos \alpha / \sin \beta$</td>
</tr>
<tr>
<td>$\xi_H^f$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\cos \alpha / \cos \beta$</td>
<td>$\sin \alpha / \sin \beta$</td>
<td>$\sin \alpha / \sin \beta$</td>
</tr>
<tr>
<td>$\xi_A^u$</td>
<td>$\cot \beta$</td>
<td>$\cot \beta$</td>
<td>$\cot \beta$</td>
<td>$\cot \beta$</td>
</tr>
<tr>
<td>$\xi_A^d$</td>
<td>$-\cot \beta$</td>
<td>$\tan \beta$</td>
<td>$-\cot \beta$</td>
<td>$\tan \beta$</td>
</tr>
<tr>
<td>$\xi_A^f$</td>
<td>$-\cot \beta$</td>
<td>$\tan \beta$</td>
<td>$\tan \beta$</td>
<td>$-\cot \beta$</td>
</tr>
</tbody>
</table>

The 2HDM phenomenology depends in detail on the various couplings of the Higgs bosons to fermions, gauge bosons and Higgs bosons. The Higgs couplings to gauge bosons normalized to those of the Standard Model are:

\[ g_{hVV} = \sin(\beta - \alpha) g_V \quad g_{hVV} = \cos(\beta - \alpha) g_V \] (3.44)

where V are for $W$ or $Z$ bosons and $g_V$ are the SM Higgs couplings:

\[ g_W = g \quad \text{and} \quad g_Z = g / \cos \theta_W \] (3.45)

Couplings of $A$ and $H^+$ to $VV$ are zero.

\(^4\) Where $P_{L/R} = \frac{1}{2} (1 \pm \gamma_5)$
Trilinear couplings of one Gauge boson to two Higgs bosons are fixed in any CP conserving 2HDM as:

\[
g_{hZA} = \frac{1}{2} g_Z \cos(\beta - \alpha), \quad g_{hZA} = -\frac{1}{2} g_Z \sin(\beta - \alpha)
\]  

(3.46)

Finalizing, the 2HDM model allows a large range of production and decay mode of an individual scalar or a cascade that involves more than one scalars. Therefore the 2HDM allows exotic decays of the SM-like Higgs which could be one of the neutral scalars h or H. Unfortunately, the 2HDM has been constrained from recent experimental data [13], [14] and restricts a large variety of exotic Higgs decay phenomenology.

To avoid these restrictions, two ways can be put forth. One way is to assume that the 2HDM is near or in the decoupling limits where \(\alpha \rightarrow \beta - \pi/2\); as a result, the lightest state of the 2HDM is the observed 125GeV Higgs state. At these limits, the couplings of the h to the fermions and to the gauge bosons are very close to those of SM-like Higgs and reach the exact SM limit if \(\alpha - \beta = \pi/2\).

The other way is to add in the existing model one complex scalar singlet such as:

\[
S = \frac{1}{\sqrt{2}} (S_R + i S_I)
\]  

(3.47)

leading to a new extended model, the 2HDM + S, which is the model that predicts the exotic decay channel of this analysis (\(H \rightarrow aa \rightarrow bb \tau\tau\)). This model is discussed below.

3.4 Two Higgs Doublet Models Plus a Scalar Singlet - (2HDM + S)

The 2HDM + S is a simple extension of the 2HDM by adding a complex scalar singlet (eq. 3.47). The addition of a complex scalar singlet has no direct Yukawa couplings and couples only to the two Higgs doublet fields \(\Phi_1\) and \(\Phi_2\) in the potential of eq. 3.20. This allows to keep its coupling to SM fermions avoiding the spoil of the Standard model-like state. Actually, none of the mass terms of the 2HDM is affected from the complex scalar singlet.
One of the key motivations for introducing the 2HDMS+S model is the fact that the next to minimal supersymmetric standard model (NMSSM), is a particular case of this model. The NMSSM is a simple extension of the MSSM developed to solve the so-called μ-problem [17]. The MSSM superpotential contains a μ mass parameter (μ Φ1 Φ2), with scale well below the Plank scale. In the NMSSM, by introducing a complex scalar singlet this problem disappears and it thus presents a very strong motivation.

The general form of the 2HDM+S potential is:

\[
V(\Phi_1, \Phi_2, S) = m_{11}^2 \Phi_1 \dagger \Phi_1 + m_{22}^2 \Phi_2 \dagger \Phi_2 - m_{12}^2 (\Phi_1 \dagger \Phi_2 + \Phi_2 \dagger \Phi_1) + \frac{1}{2} \lambda_1 (\Phi_1 \dagger \Phi_1)^2 \\
+ \frac{1}{2} \lambda_2 (\Phi_2 \dagger \Phi_2)^2 + \lambda_3 (\Phi_1 \dagger \Phi_1)(\Phi_2 \dagger \Phi_2) + \lambda_4 (\Phi_1 \dagger \Phi_2)(\Phi_2 \dagger \Phi_1) + \frac{\lambda_5}{2} [(\Phi_1 \dagger \Phi_2)^2 + (\Phi_2 \dagger \Phi_1)^2] + h.c. \\
+ \frac{1}{2} m_0^2 S^2 + \frac{1}{4!} \lambda_S S^4 + \kappa_1 S^2 (\Phi_1 \dagger \Phi_1) + \kappa_2 S^2 (\Phi_2 \dagger \Phi_2)
\] (3.48)

As can be seen from the above potential, the 2HDM sector can perform independently of the S, and the usual mass terms derived from the 2HDM are not changed due to the extra S field. The extra S field does not acquire a vacuum expectation value. However, the physical S particle mass term (S^2 mass terms) is affected from the two Higgs fields when they develop VEVs. The S-dependent part of the scalar potential in terms of the mass eigenstates, is written as:

\[
- V_S = - \frac{1}{2} m_S^2 S^2 - \lambda_h \nu h S^2 - \lambda_H \nu H S^2 - \frac{1}{2} (\lambda_{HH} HH + \lambda_{hh} hH + \lambda_{AA} AA + \lambda_{HH} H^+ H^-)
\] (3.49)

where the mass of the physical S particle and the trilinear couplings are the following:

\[
m_S^2 = m_0^2 + (\kappa_1 \cos^2 \beta + \kappa_2 \sin^2 \beta) \nu^2
\] (3.50)

\[
\lambda_h = - \kappa_1 \sin \alpha \cos \beta + \kappa_2 \cos \alpha \sin \beta
\] (3.51)

\[
\lambda_H = \kappa_1 \cos \alpha \cos \beta + \kappa_2 \sin \alpha \sin \beta
\] (3.52)
In the decoupling limits where \( \cos(\beta - \alpha) \ll 1 \) or \( \sin(\beta - \alpha) = 1 \) for which the lightest scalar state \( h \) has exactly the SM-like couplings to vector bosons (\( VV \)) and to fermions (\( \bar{f} f \)), the trilinear couplings take the form shown below:

\[
\lambda_h = \kappa_1 \cos^2 \beta + \kappa_2 \sin^2 \beta \quad (3.53)
\]

\[
\lambda_H = (\kappa_1 - \kappa_2) \sin \beta \cos \beta \quad (3.54)
\]

The set of \( m_0, \kappa_1, \text{ and } \kappa_2 \) constitute a complete set of extra free parameters for the scalar sector in the Lagrangian and are treated as independent free parameters associated only with the \( S \) sector.

The quadrilinear couplings of the scalar with the Higgs bosons can also be expressed by the parameters \( \kappa_1, \kappa_2, \alpha \) and \( \beta \):

\[
\lambda_{AA} = \frac{1}{2} (\kappa_1 \sin^2 \beta + \kappa_2 \cos^2 \beta) \quad , \quad \lambda_{hh} = \frac{1}{2} (\kappa_2 \cos^2 \alpha + \kappa_1 \sin^2 \alpha) \quad , \quad \lambda_{Hh} = \frac{1}{2} (\kappa_1 \cos^2 \alpha + \kappa_2 \sin^2 \alpha) 
\]

\[
\lambda_{hh} = \frac{1}{2} (\kappa_2 - \kappa_1) \sin^2 2 \alpha \quad (3.55)
\]

For non spoiling the SM-like nature of \( h \) (lightest Higgs state), the mixing between Higgs bosons and the single \( S \) need to be small. After a small mixing of singlet \( S \) with pseudoscalar \( A \) boson the imaginary part of the singlet gives rise to a light pseudoscalar \( \alpha \) and the real part after mixing with the two neutral scalar Higgs, gives rise to a light scalar \( s \). Therefore, the 2HDM+S allows exotic Higgs decays of the form:

\[
h \rightarrow \alpha \alpha \quad , \quad h \rightarrow s s \quad \text{or} \quad h \rightarrow Z \alpha \quad (3.56)
\]

The pseudo(scalar) \( \alpha \) (s) are the lightest case of the heavy \( A \) (S) states and so they decay further to SM particles as the heavy pseudo(scalar) bosons [15]. In the following content, we focus on the exotic SM-like Higgs decay to a pair of light pseudoscalar bosons, since it is the first part of the investigated channel in this document.
3.4.1 Light Pseudoscalar Boson $\alpha$

In the case where the light pseudoscalar $\alpha$ is lighter than the SM-like Higgs state, there are two possible exotic Higgs decays. More specifically, if the mass of the pseudoscalar $a$ is about 35GeV ($m_a < m_h - m_Z \approx 35\text{GeV}$), then the $h$ decays to a $Z$ boson and to a pseudoscalar $a$ ($h \rightarrow Za$). If the pseudoscalar mass $a$ is lower than the half of the SM-like higgs $h$ ($m_a < m_h/2 \approx 63\text{GeV}$), then the decay of the higgs to a couple of light pseudoscalar bosons $a$ ($h \rightarrow aa$) is allowed.

In the same way as in the 2HDM, the model is separated in four different types which forbid the FCNC. The decays of light pseudoscalar bosons to SM particles are yields from the couplings of the heavy pseudoscalar state $A$ (as in table 3.3) multiplied by $\sin \theta_a$. In this way, the Yukawa couplings are rescaled by a small singlet mixing angle $\theta_a$ between the doublet and the pseudoscalar single sector. The $\cos \theta_a$ is defined as:

$$\alpha = \cos \theta_a S_1 + \sin \theta_a A, \quad \theta_a \ll 1$$

(3.57)

Therefore, the light pseudoscalar boson $a$ further decay to SM particles is imposed by the couplings of $a$ to the SM fermions in the different types of the 2HDM+S. In Type I, as in the 2HDM, all fermions couple only to $\Phi_2$ (like in the SM with only one doublet) so the couplings are independent of $\tan \beta$. This means that the couplings are the same as in the SM and also that the branching ratios of the pseudoscalar bosons are very similar to those in the SM + S [16]. Figure (3.2) shows the Branching ratios of the light pseudoscalar boson of 2HDM + S Type I.

In Type II, all the leptons and the down type quarks couple to the same doublet $\Phi_1$ and up type quarks couple to $\Phi_2$, as in the NMSSM. Now the couplings of the down type fermions, unlike Type I, depend on $\tan \beta$. For $\tan \beta < 1$, the decay to down type fermions is suppressed and for $\tan \beta > 1$ it is enhanced. As figure 3.3 shows, for larger $\tan \beta$ and for $m_a > 10 \text{ GeV}$, the channels $h \rightarrow aa \rightarrow bbbb$ and $h \rightarrow aa \rightarrow bb\tau\tau$ seem to be the most promising ones.

In Type III (lepton-specific), only the leptons couple to $\Phi_1$ in contrast to quarks that couple with $\Phi_2$. Therefore, the increasing of $\tan \beta$ as shown in figure (3.4), increases the branching ratios of channels with leptons in the final state like, $h \rightarrow aa \rightarrow \mu\mu\tau\tau$ and $h \rightarrow aa \rightarrow \tau\tau\tau\tau$ and suppresses the decays to quarks.
Figure 3.2: The branching ratios of pseudoscalar $a$ for Type I 2HDM+S. The shaded regions represent the decays to quarkonia which may invalidate our calculations [17].

Figure 3.3: The branching ratios of pseudoscalar $a$ for Type II 2HDM+S. The left hand plot represents the branching ratio for $\tan \beta = 0.5$ and the right hand side the branching ratio for $\tan \beta = 5$. The shaded regions represent the decays to quarkonia which may invalidate our calculations.
Figure 3.4: The branching ratios of pseudoscalar $a$ for Type III 2HDM+S. The left hand plot represents the branching ratio for $\tan\beta = 0.5$ and the right hand side the branching ratios for $\tan\beta = 5$. The shaded regions represent the decays to quarkonia which may invalidate our calculations.

Figure 3.5: The branching ratios of pseudoscalar $a$ for Type IV 2HDM+S. The left hand plot represents the branching ratio for $\tan\beta = 0.5$ and the right hand side the branching ratios for $\tan\beta = 5$. The shaded regions represent the decays to quarkonia which may invalidate our calculations.
Finally in Type IV (flipped) the up-type fermions couple to $\Phi_2$ and the down type quarks to $\Phi_1$. Then the pseudoscalar decays to leptons and up-type quarks enhance for larger values of $\tan \beta$ in contrast with the Type II model. From figure (3.5), one can deduce that the most favorable channels are the $h \rightarrow aa \rightarrow \tau\tau bb$ and $h \rightarrow aa \rightarrow \tau\tau cc$.

The exotic decays of the 125 GeV Higgs boson to light pseudoscalars, with a pair of $b$ jets and a pair of tau leptons in the final state, seems to be very promising. The analysis of this channel is strongly encouraged from the 2HDM+S and its four different types. Figure 3.6 shows the branching fraction $B(aa \rightarrow bb\tau\tau)$ for few of the 2HDM+S scenarios. In type II and for $\tan\beta > 1$, (in which NMSSM is a spacial case [17]), the branching fraction is flat and above 10%, while in 2HDM+S type III with $\tan\beta \sim 2$ the branching fraction can reach up to about 45%.

*Figure 3.6:* Predicted $B(aa \rightarrow bb\tau\tau)$ for $m_a = 40$ GeV in the different models of 2HDM+S, as a function of $\tan \beta$. The picture is essentially the same for all $m_a$ hypotheses considered in this analysis. The branching fractions are computed following the formulas of Ref. [18].
References:


CHAPTER 4

LHC and CMS Experiment

4.1 Large Hadron Collider LHC

The Large Hadron Collider (LHC) is the largest and most powerful circular particle collider ever built. It was built in the tunnel previously used by the LEP, which was located about 100m underground on the border between France and Switzerland, and has a circumference of 26.7 km [1]. There are four main experiments located at different collision points of the tunnel (Figure 4.1): CMS, ATLAS, LHCb and ALICE. The CMS (Compact Muon Solenoid) [2] and ATLAS (A Toidal LHC Apparatus) [3] are multipurpose experiments with wide range of physics objectives such as the search of SM scalar bosons and particles from beyond the SM. The LHCb (Large Hadron Collider Beauty) is dedicated to the study of the physics of B-mesons and makes precision measurements of the CP-violation [4]. For the study of strongly interacting matter and quark gluon plasma at very high values of energy density, a heavy-ion detector was designed, namely ALICE (A Large Ion Collider Experiment)[5].

One of the main goals of the general purpose detectors as previously discussed was the search for the SM scalar boson called Higgs. Based on the knowledge and results derived from the LEP and Tevatron the mass of the scalar boson should be larger than 114GeV and lower than 1TeV [6], [7]. However, the aims of LHC are much wider, as it is expected to cover a much wider range of searches beyond the Standard model like Supersymmetry (SUSY), extra dimensions and Dark matter.

The LHC is designed for collisions of protons and it can also support lead-lead or lead-proton collisions. Protons originate from a single bottle of hydrogen and pushed into a cathode chamber of
duoplasmatron. There, a combination of electric and magnetic fields create an intense ionization and lead to the confinement of a plasma. Protons are extracted from the plasma by an electrode and form the proton beam. Protons are first accelerated in a linear accelerator, the LINAC2, which is the first accelerator in the chain of accelerations (figure 4.2). LINAC2 accelerates the protons up to the energy of 50 MeV and are then injected in a circular accelerator, namely the PS booster, where the beam energy reaches 1.4GeV. Proton Synchotron (PS) and Super Proton Synchrotron (SPS) follow and increase the protons energy from 25GeV to 450GeV.

![Figure 4.1: The overview of LHC circular underground tunnel with four LHC main experiments located at different collision points of the tunnel: CMS, ATLAS, LHCb and ALICE.](image)

The next step is the LHC, where the protons travel inside two separate beam pipes. It needs 4 minutes and 20 seconds to fill each LHC ring and an additional 20 minutes for the proton beam to reach the energy of 6.5 TeV. When the LHC is completely filled, the two beams are brought into collision inside the four detectors, with the center of mass energy at the collision point becoming equal to 13TeV.

Inside the LHC, protons are accelerated via 16 radiofrequency cavities by an electromagnetic field oscillating at a frequency of 400 MHz, while the beams are bent by the magnetic field of 1232
superconducting dipole magnets. The superconducting dipole magnets along with 392 quadrupole magnets ensure the deflection and the collimation of the beam, respectively. The opposite directions of the two beams in the LHC ring, require opposite magnetic fields and separate vacuum chambers. Due to the limited size of the tunnel, the LHC uses twin bore magnets. The superconducting magnets are kept at the temperature of 1.9K using 120t of superfluid helium at the pressure of 0.13MPa.

The LHC start was scheduled in 2008, but a leak of liquid helium at the SPS delayed its initiation to the end of 2009. On 30 March 2010 the first collisions were recorded at the center of mass energy of 7TeV, which was kept constant during the LHC operations for 2010 and 2011. In 2012, the center of mass energy was increased to 8TeV and in 2015, after the first long shut down (LS1), the center of mass energy reached 13TeV.

![Figure 4.2](image)

**Figure 4.2:** Overview of the CERN acceleration complex. To reach their final energy protons successively pass through the LINAC2, the PSB (Proton Synchrotron Booster), the PS (Proton Synchrotron), the SPS (Super Proton Synchrotron) and finally the LHC (Large Hadron Collider).
Except from the collision energy another important parameter in collider physics is the luminosity.

The interaction rate, \( \frac{dN}{dt} \), depends on the process cross section \( \sigma \), and on the luminosity \( \mathcal{L} \) with the following relation:

\[
\frac{dN}{dt} = \mathcal{L} \sigma
\]  

(4.1)

The instantaneous Luminosity of the beam collision in units of \( \text{s}^{-1} \text{cm}^{-2} \), can be expressed with the beam characteristic properties:

\[
\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma^*}{4\pi \Lambda \beta^*} F
\]  

(4.2)

where \( N_b \) is the number of protons per bunch, \( n_b \) denotes the number of bunches in LHC ring, \( f_{\text{rev}} \) is the revolution frequency, \( \gamma \) the relativistic factor of protons (Lorentz factor), \( \Lambda \) is the normalized emittance, \( * \) is the beta function at the collision point and \( F \) is the geometrical reduction factor due to the crossing angle \( \theta_c \):

\[
F = \left(1 + \left(\frac{\sigma_z}{\sigma_x}\right)^2 \left(\frac{\theta_c}{2}\right)^2 \right)^{-\frac{1}{2}} \quad \text{(4.3)}
\]

where \( \theta_c \) is the beam crossing angle and \( \sigma_x, \sigma_z \) are the beam cross-sectional sizes of the bunch in \( x \) and \( z \) directions at the interaction point as shown in figure (4.3). The \( \sigma_z \) is constant \( \sim 7.5 \text{ cm} \) and \( \sigma_x \) varies and takes the minimum value at the interaction point. In table (4.1), the values of the forementioned parameters are presented along with other performance parameters for Run-1 and Run-2.

The integrated luminosity \( L \), is the integral of the instantaneous luminosity over a given range of time. The increasing of the integrated luminosity and the collisions center-of-mass energies from the start until today is presented in figure 4.4.
During that period, the LHC stopped for two years (LS1) after Run-1, in order to undergo technical upgrades of the machine and its experiments. The second long shutdown (LS2) is planned to happen in 2019 – 2020 after Run-2 and Run-3 will extent up to 2023. Run-1, Run-2 and Run-3 constitute the Phase-1 of the program whereas Phase-2 will be extended up to year 2037. The integrated luminosity collected in the first phase is expected to reach the value of 300fb⁻¹. At Phase-2 it is expected to reach 10 times higher, close to 3000fb⁻¹. The schedule of LHC is presented in Figure 4.5.

![Beam collision schematic view](image)

**Figure 4.3**: The schematic view of beam collision at the point where two bunches interact.

<table>
<thead>
<tr>
<th>Table 4.1: Performance parameters of LHC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parameter</strong></td>
</tr>
<tr>
<td>Beam Energy (TeV)</td>
</tr>
<tr>
<td>$\beta^*$ (beta function)[m]</td>
</tr>
<tr>
<td>Bunch spacing [ns]</td>
</tr>
<tr>
<td>Number of Bunches, $n_b$</td>
</tr>
<tr>
<td>Revolution frequency [Hz]</td>
</tr>
<tr>
<td>Average bunch intensity, $N_b$</td>
</tr>
<tr>
<td>Normalized emittance, $\epsilon_n$</td>
</tr>
<tr>
<td>[mm.mrad]</td>
</tr>
<tr>
<td>Events/bunch crossing</td>
</tr>
</tbody>
</table>
Figure 4.4: Total integrated luminosity of proton beam collisions delivered by the LHC in 2010 (green line), 2011 (red line), 2012 (blue line), 2015 (purple line), 2016 (yellow line), 2017 (sky blue line) and 2018 (dark blue line).

Figure 4.5: Overview of the LHC schedule until 2037. The integrated luminosity collected in Phase-1 is expected to reach 300fb\(^{-1}\), while 3000fb\(^{-1}\) should be collected by the end of Phase-2.
4.2 CMS Experiment

The CMS (Compact Muon Solenoid) experiment is an international collaboration involving more than 4000 members from about 200 universities and institutes in Europe, America and Asia. The main goal of the experiment is the study of proton-proton collisions at 14TeV center-of-mass energy and at Luminosity levels (L = 10^{34} \text{cm}^{-2}\text{s}^{-1}) that had never been reached in previous experiments.

The CMS detector is one of the two general purpose experiments at LHC. It is designed to cover a broad physics program, from the search for the Standard Model Higgs boson to extra dimensions and dark matter candidates (BSM physics). The main characteristic which enables it to achieve these goals is the excellent energy resolution for charged particles, photons and electrons. The good missing traverse energy and di-jet mass resolution are also another notable characteristics of the CMS detector [2].

4.2.1 Overview of the CMS Detector

The CMS detector has a diameter of 15.0 m, is 28.7 m long and weights 14000 t (figure 4.6). It is located in the underground cavern near Cessy at point 5 of LHC. It is cylindrical in shape and is separated into two main parts, the center of the cylinder, the barrel and the two endcaps. It is composed of different subdetector layers and its geometrical shape surrounds the proton beamline and the interaction point, providing nearly 4\pi coverage of proton collisions. Traversing from the innermost to the outermost parts of the CMS detector, one encounters first an inner tracking system with highly efficient reconstruction of trajectories (tracks) of all charged particles and of the secondary vertices in proton-proton collisions. The tracker is surrounded from the electromagnetic calorimeter (ECAL), which provides an accurate measurement of the energy of photons and electrons by absorbing their energy. In front of the ECAL endcaps a preshower detector is positioned to measure the energy of photons coming from neutral pion decays and distinguish them from photons coming from Higgs boson decay. Next comes the hadron calorimeter (HCAL), which measures the energy of neutral and charged hadrons and records the missing transverse energy. All three subdetectors are located inside the superconducting solenoid, which generates a 3.8T magnetic field parallel to the beam axis to bend the tracks of charged particles. Outside the magnet there is a muon system with a high efficiency of muon’s identification. Combine information from the muon...
system and from the tracker system a good resolution of muon momentum measurement can be ensure. Finally all the informations coming from the muon system and subdetectors are passing to a trigger system which consists two levels, the L1 trigger (hardware) and the HLT (software) for selecting the interesting events and store them for subsequent analysis.

**Figure 4.6**: A schematic representation of the CMS Detector and its subdetectors.

### 4.2.2 CMS Coordinate System

The CMS coordinate system has its origin at the center of the detector at the point of the collisions. The CMS uses a right handed coordinate system origin at the center of the nominal collision point inside the experiment. The x-axis points radially to the center of the LHC, while the y-axis points vertically upwards. The z-axis is orthogonal to the x-y plane and points along the beam direction toward the Jura mountains from LHC Point 5. The azimuthal angle $\phi$ is measured from the x-axis to the x-y plane and the polar angle $\theta$ from the z-axis (fig: 4.7). Although, instead of the polar angle, in hadron colliders pseudorapidity is more widely used, which is defined as:

\[ \eta = -\ln \tan \left( \frac{\theta}{2} \right) \]  

(4.4)
The pseudorapidity can also be expressed as:

\[
\eta = \frac{1}{2} \ln \left( \frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right) = \operatorname{artanh} \left( \frac{p_z}{|\vec{p}|} \right)
\]  

(4.5)

For particles with \(|p| \gg m \Rightarrow E \approx |p| \Rightarrow y \approx \eta\), where \(y\) is the rapidity defined through the relation:

\[
y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right)
\]  

(4.6)

The transverse momentum \(p_T\) is defined as a projection of \(\vec{p}\) to the x-y plane:

\[
p_T = \sqrt{p_x^2 + p_y^2}
\]  

(4.7)

while the transverse energy is defined as:

\[E_T = E \sin \theta\]  

(4.8)

### 4.2.3 Tracker

The CMS tracking system (tracker) is of high importance to the experiment (figure 4.9). It is the closest subdetector to the beamline and it is exposed to a large flux of particles. It is designed to have a high granularity, fast response and be able to withstand severe irradiation. The data from the tracker is used to reconstruct trajectories of charged particles and to measure their momentum as well as reconstruct primary and secondary vertices from hadron collisions and subsequent decays of their products. The tracking takes place in a 5.8 m length and 2.5 m diameter cylinder surrounding the interaction point and consisting of two parts: silicon pixel and silicon strip detectors.

The silicon pixel detector consists of three cylindrical barrel layers (TPB) with two endcap disks on each side (TPE). The three barrel layers are positioned at 4.4 cm, 7.3 cm and 10.2 cm, respectively, from the beamline. The two endcap disks extend from 6 to 15 cm radius and are placed at \(|z| = 34.5\) and 46.5 cm, with pixel cell size of \(100 \times 150 \mu \text{m}^2\) [8]. The schematic illustration of the silicon pixel detector is shown in figure 4.8.
The silicon pixel detector is surrounded by the silicon strip detector. It consists of four parts, two in the barrel region and two parts in the endcap region. The barrel parts are a TIB (Tracker Inner Barrel) and a TOB (Tracker Outer Barrel). The TIB has four layers and covers the volume defined by $20 < r < 55$ cm and $|z| < 65$ cm, using silicon sensors with thickness of 320 μm and a strip pitch which varies from 80 – 120 μm. The TOB comprises 6 layers with half-length of $|z| < 110$ cm. Due to the smaller level of radiation in this region, thicker silicon sensors of 500 μm are used as well as wider pitch which varies from 120 - 180 μm.

*Figure 4.7:* The CMS coordinate system.

*Figure 4.8:* Sectional view of the tracker [8].
The endcap parts are a TEC (Tracker End Cap) and a TID (Tracker Inner Disks). The TEC is comprised from nine disks that extend into the region defined by $120 \text{ cm} < |z| < 280 \text{ cm}$, while each TID comprises from three small disks that fill the gap between TIB and TEC. The sensors thickness is $320 \mu\text{m}$ for TID in the innermost disk of the TEC and $500 \mu\text{m}$ for the rest of the TEC.

\textbf{Figure 4.9:} The CMS Pixel detector [8].

In total, the inner tracker consists of 1440 silicon pixels and 15148 silicon strip detector modules. The resolution on the transverse momentum for 100 GeV charged particles is about 2%, while the impact parameter resolution from the inner tracker is about 15 $\mu\text{m}$.

\subsection*{4.2.4 Electromagnetic Calorimeter ECAL}

The Electromagnetic Calorimeter (ECAL) played a central role in the discovery of a new resonance with a mass 125 GeV, attributed to the Standard Model Higgs boson during 2011 and 2012 data taking [9, 10]. The design of the CMS electromagnetic calorimeter (ECAL) is based on the requirements imposed by the search for the Higgs boson decays to a pair of two photons in which its narrow peak has to be distinguished from the a large and continuous background. Its good energy resolution and its fine granularity led to a very good energy resolution of the di-photon system by improving the energy and the angle measurement of the two photons.
The electromagnetic calorimeter (ECAL) [11] covers a pseudorapidity region of $|\eta| < 3.0$ and measures and absorbs the energy of electrons and photons (figure 4.10). It is a hermetic, fine grained and homogenous calorimeter that surrounds the CMS tracking system. It is made of 75 848 lead-tungstate (PbWO$_4$) scintillating crystals 61200 in the barrel and 7324 in each endcap region. The choice of the lead-tungstate crystals was based on key characteristic features such as: their tolerance to high radiation levels, their short radiation length, their small Molière radius ($R_M = 2.2$ cm), which allows a more compact detector, and their density of $8.28 g/cm^3$. In addition, their fast response (80% of the scintillation light is emitted within 25 ns), is of great importance since it allow ECAL to identify electrons and photons in a very high pile-up environment. Their relatively low light yield of 30 photons/MeV, requires the use of photodetectors. Two types of photodetectors are used: the APDs in the barrel region and VPTs in the endcap region, which can both operate in a strong magnetic field.

**Figure 4.10:** Sectional view of ECAL: the Barrel ECAL (EB), the Preshower (ES) in front of the Endcap ECAL (EE) [11].

The Barrel ECAL (EB) has an inner radius of 129cm and covers a pseudorapidity range of $|\eta| < 1.479$. It is structured as 36 identical “supermodules”, each of them consisting of 1700 crystals. The crystals are truncated pyramids with a front face cross section of $22 \times 22$ mm$^2$, rear face cross section $26 \times 26$ mm$^2$ and length 230mm (25.8X$_0$). Further, crystals in the barrel are inclined by 3° in the $\eta$ - $\phi$ direction to prevent particles from passing through the intersections between two crystals.
The endcaps of ECAL (EE) cover a pseudorapidity range of $1.479 < |\eta| < 3.0$. Each of the endcaps is divided into two halves (“Dees”), which consist of “supercrystals”. Each supercrystal includes $5 \times 5$ crystals arranged in an x-y grid, with the total number of crystals in each Dee to be 3662. They are all identical and have a front face cross section of $28.6 \times 28.6$ mm$^2$, rear face cross section $30 \times 30$ mm$^2$ and a length of 220mm ($24.7X_0$).

Two different types of photodetectors are used to collect the scintillation light in the barrel (EB) and the endcap (EE). The barrel photodetectors are the APDs which have an active area cross section of $5 \times 5$ mm$^2$. Each EB crystal is glued with two silicon avalanche photodiodes, APDs, with a mean gain of 50. In contrast, the EE crystals are glued only to one vacuum phototriode, VPT, with 25mm diameter and with a mean gain of 10.2 at 0T. The vacuum phototriodes, VPTs, are position in the endcap region instead of APDs due to their better tolerance at high radiation levels, which dominate the endcaps. The signal from both photodetectors is amplified, digitized and immediately transported away by fiber optic cables for upper level readout.

In front of the endcap a sampling calorimeter is placed, the ECAL Preshower (ES). The preshower detector is used to identify neutral pions, decaying to photons pairs. It covers a pseudorapidity range $1.653 < |\eta| < 2.6$. It consists of two layers. The first layer consists of lead radiators that initiate electromagnetic showers of photons and electrons. The second layer is composed of silicon strip sensors. These silicon strip sensors have a thickness of 320 μm and measure the deposited energy and the transverse shower profile.

The ECAL energy resolution can be parametrized as a function of energy:

$$
\frac{\sigma}{E}^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + 0.3\%
$$

(4.9)

The 2.8% represents the statistical fluctuations on the number of secondary particles produced, 12% is the noise coming from the electronics and digitization and 0.3% is a constant term. These parameter values are estimated in an electron test beam and the results are obtained by reconstructing the shower energy within a $3 \times 3$ crystal matrix [12]. Figure 4.11 shows the ECAL
energy resolution as a function of the incident electron energy measured at beam test, as well as the values of terms obtained after fitting equation 4.9 with the data.

![Graph showing energy resolution as a function of electron energy.](image)

**Figure 4.11:** The ECAL energy resolution as a function of the electron energy measured during a test beam at 2006. The stochastic (S), noise (N), and constant (C) values are shown in the plot [13].

4.2.5 Hadron Calorimeter HCAL

The CMS Hadronic Calorimeter (HCAL) plays an important role in the search for new physics [14]. It is used to measure the energy of hadronic showers and the missing energy resulting from neutrino or exotic particles. The hadron calorimeter is located around the ECAL and extends between 1.77 < r < 2.95 m within the solenoid operating at 3.8 Tesla. This constrains the total amount of material which can be put in to absorb the hadronic shower. HCAL is made from materials with short interaction lengths. An outer hadron calorimeter is placed outside the solenoid complementing the barrel calorimeter. In addition to providing a good measurement of the transverse missing energy, it should be as hermetic as possible and extends to large pseudorapidity values. Therefore, the forward hadron calorimeter is placed at 11.2 m from the interaction point and extends the pseudorapidity coverage to |\( \eta \) | = 5.2.
Figure 4.12 shows the layout of the detector. The HCAL includes four distinct subsystems: the barrel (HB), endcap (HE), outer(HO) and forward (HF) calorimeter. The HB consists of 32 towers covering the pseudorapidity region -1.4 < η < 1.4, resulting in 2304 towers with segmentation Δη × Δφ = 0.087 × 0.087. It is made of sixteen absorber plates, 14 of them made from brass (8 with 50.5 mm thickness and 6 with thickness 56.5), and two steel plates to ensure mechanical strength.

The hadron outer (HO) detector is added around the magnet to complement the HB. It contains scintillators with a thickness of 10mm and covers a pseudorapidity range -1.26 < η < 1.26. In combination with the HB, it increases the effective thickness of the hadron calorimeter to 12 interaction lengths and thus reduces the tails in the energy resolution function.

Hadron endcaps (HE) cover the region 1.3 < |η| < 3.0 and consist of layers of absorbed brass plates (78mm thickness) and scintillation tiles ( 9mm thickness), which correspond to approximately ten interaction lengths. So, it provides granularity Δη × Δφ 0.087 × 0.087 in pseudorapidity range 1.3 < |η| < 1.6 and Δη × Δφ 0.17 × 0.17 in pseudorapidity range 1.6 < |η| < 3.0.

The forward hadron calorimeter (HF) front face is located at 11.2 m away from the interaction point and covers a high pseudorapidity region of 3.0 < |η| < 5.0. Due to its location, it has significantly higher particle flux passing through it. The radiation-hard quartz fibers that the HF consists of, collect the Cherenkov light which is then channeled by the fibers to photomultipliers.

The ECAL and the HCAL combine energy resolution can be parametrized by as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + C^2 \quad (4.10)$$

The S and C are constant values and have been measured to be S = 0.847 ± 0.016 GeV^{1/2} and C = 0.074 ± 0.008 [15]. This resolution obtain after installation of the hadron outer detector (HO) which improves the energy resolution for high energy pions. Before the correction the S and C constants values where S = 1.107 GeV^{1/2} and C = 0.074.
4.2.6 Magnet

The superconducting solenoid magnet is a highly importance system for the CMS detector. It is the largest magnet of its type ever constructed and surrounds all calorimeters. This allows all the calorimeters to be placed inside the coil, resulting a detector that is overall compact compared to similar detectors.

The magnet is 12.9 m in length, has an inner diameter of 5.9m and weights 12000 tonnes [2]. It provides 3.8T of magnetic field, which is approximately 100000 times stronger than the Earth’s. It is made of 2168 turns carrying a 19.5 kA current and its serves to curve the tracks of charged particles. Combining the excellent resolution in position measurement from the tracker and the muon detector, it allows the precise measurement of the momenta of high energy particles. The magnet is cooling down with liquid helium, which achieves a decrease in its temperature to -268.5 °C. The main magnet parameters are presented in Table 4.2.

4.2.7 Muon System

The muons as a low interacting particles are not stopped in any of the calorimeters and they can penetrate several meters of iron without interacting. Therefore, the muon system is placed at the very edge of the experiment where other particles will rarely reach and thus provides a natural discrimination against non-muon backgrounds. The CMS muon system provides an efficient reconstruction and identification of muons and together with the tracker, it provides measurement of muon momentum with 1 – 10% precision within a wide kinematic range [15].

The layout of the muon system is illustrated in Figure 4.13. It consists of the barrel region (|η|<1.2) and the two endcap regions (|η|<2.4). It’s composed of three types of gaseous detectors, located outside the magnetic solenoid: The Drift Tube chambers (DT) in the barrel region, the Cathode Strip Chambers (CSC) in the endcap regions and the Resistive Plate Chambers (RPC).
**Table 4.2**: The main CMS magnet parameters [2]

<table>
<thead>
<tr>
<th>General parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnetic length</td>
<td>12.5 m</td>
</tr>
<tr>
<td>Cold bore diameter</td>
<td>6.3 m</td>
</tr>
<tr>
<td>Central magnetic induction</td>
<td>4 T</td>
</tr>
<tr>
<td>Total Ampere-turns</td>
<td>41.7 MA-turns</td>
</tr>
<tr>
<td>Nominal current</td>
<td>19.14 kA</td>
</tr>
<tr>
<td>Inductance</td>
<td>14.2 H</td>
</tr>
<tr>
<td>Stored energy</td>
<td>2.6 GJ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cold mass</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Radial thickness of cold mass</td>
<td>312 mm</td>
</tr>
<tr>
<td>Radiation thickness of cold mass</td>
<td>3.9 $X_0$</td>
</tr>
<tr>
<td>Weight of cold mass</td>
<td>220 t</td>
</tr>
<tr>
<td>Maximum induction on conductor</td>
<td>4.6 T</td>
</tr>
<tr>
<td>Temperature margin wrt operating temperature</td>
<td>1.8 K</td>
</tr>
<tr>
<td>Stored energy/unit cold mass</td>
<td>11.6 kJ/kg</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Iron yoke</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer diameter of the iron flats</td>
<td>14 m</td>
</tr>
<tr>
<td>Length of barrel</td>
<td>13 m</td>
</tr>
<tr>
<td>Thickness of the iron layers in barrel</td>
<td>300, 630 and 630 mm</td>
</tr>
<tr>
<td>Mass of iron in barrel</td>
<td>6000 t</td>
</tr>
<tr>
<td>Thickness of iron disks in endcaps</td>
<td>250, 600 and 600 mm</td>
</tr>
<tr>
<td>Mass of iron in each endcap</td>
<td>2000 t</td>
</tr>
<tr>
<td>Total mass of iron in return yoke</td>
<td>10 000 t</td>
</tr>
</tbody>
</table>

**Figure 4.12**: Longitudinal view of a quarter of the CMS hadronic calorimeter HCAl. The location of hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters are indicated [2].
In the barrel region (MB), where the neutron induced background is small the drift tube (DT) chambers are used. The muon rate is low and the magnetic field contained in the iron yoke is almost uniform. The muon system in the barrel consists of 250 chambers organized in 4 concentric muon stations inside the magnet return yoke. Each of the first three stations has 60 chambers and the last station has 70 chambers. It is chosen in such a way so as to provide a good efficiency for reconstructing muon tracks from hits in different stations. The barrel muon system is further divided in five wheels around the beam axis and each of them is divided into 12 sectors, with each covering 30° azimuthal angle.

Since the muon rate and the neutron induced background rate is high in the endcaps. Cathode Strip Chambers (CSC) are used. The Muon Endcap (ME) system regroups 468 CSCs, divided in four stations per endcap. Each CSC has a trapezoidal shape and contains six layers. Each layers contains a plane of radial cathode strips and a plane of anode wires that run along her $r - \phi$ direction. The resolution provided by each chamber from the strips is typically around 100 μm and 1 mrad in $r - \phi$ direction.

In addition to CSCs and DTs, resistive plate chambers (RPC) are used in both the barrel and the endcap regions. Six layers of RPC are embedded in the barrel, whereas three layers of RCPs are part of each endcap. The RPCs are double gap gaseous chambers operating in the avalanche mode. They are composed of parallel anode and cathode plates with gas gap in between. The RCPs are introduced in the barrel and the endcaps as a trigger system, with fast response and excellent time resolution. Their time response is much shorter than the time between two consecutive bunch crossings in contrast with DT and CSC detectors which have much better $p_T$ efficiency but not so fast response.

For low transverse momentum muons of less than 100GeV, the best momentum resolution is dominated by the resolution obtained in the silicon tracker. The best momentum resolution for high transverse momentum muons ($p_T > 200$GeV) is driven by the combination of the inner tracker and the muon detector measurements. The combine muon momentum resolution is better than 5% for $p_T = 300$GeV and rapitity region $|\eta| < 2$, and about 10% for $p_T = 2$TeV. For lower muon momentum $p_T < 100$GeV the resolution is about 1.5% at rapitity region $|\eta| < 2$. 


4.2.8 Trigger System

The LHC nominal Luminosity is $10^{34}$ cm$^{-2}$s$^{-1}$, giving a pp interaction rate that exceeds 1GHz [16]. Due to the technical and computational limitations, it is not possible to register, process and store the data for every collision. Nevertheless, only a small fraction of these collisions contain interesting events based on the CMS physics program with a rate less than 10Hz. So, the trigger job is to select the most interesting events based on their characteristics, like the transverse momentum, and store them in an offline dataset for further analysis.

The trigger system is comprised of an L1 hardware trigger and an HLT array of computers running high-level physics algorithms. The L1 trigger is using the information from the calorimeters and the muon detectors and decides if an event should be accepted or rejected. It has a design output rate of 100kHz and a response time of 3.2 μs. Events, after passing the L1 trigger, pass the High Level Trigger (HLT), which performs more complex calculations based on the combination of information from the different subdetectors. In the period 2015-2016, the L1 system passed through a series of upgrades and now is powerful and fully programmable. This allows the tau hadron reconstruction
and isolation with more advanced algorithms. More details about the L1 algorithms based on 2016 data can be found in [17].

A functional diagram of L1 trigger is shown in figure 4.14. The L1 calorimeter trigger comprises two stages, a Regional Calorimeter Trigger (RCT) and a Global Calorimeter Trigger (GCT). Data from the forward (HF) and barrel hadronic calorimeters (HB) and from the electromagnetic calorimeter (ECAL) are processed first regionally (RCT) and then globally (GCT). The regional calorimeter trigger determines regional candidate electrons or photons, regional transverse energy sums and other information for muon as coming direct from the calorimeters. Then, the highest ranked calorimeter trigger objects (e/γ, muons, tau, jets and transverse missing energy) in the whole detector are determined by the global calorimeter trigger (GCT). The energy deposits from the three muon systems, Resistive Plate Chambers (RPC), Cathode Strip Chambers (CSC) and Drift Tubes (DT) are sent to the Global Muon Trigger (GMT). Then the Global Trigger (GT) combines the information from the GCT and the GMT and makes the decision if one event should be accepted or not.

![Figure 4.14: Overview of the CMS Level 1 Trigger](image-url)
The decision is then passed to the HLT which reduces the event rate from 100kHz to about 100Hz. The HLT includes reconstruction and filtering algorithms arranged into paths, where each path targets events of a particular final state. For example trigger paths for processes with electron/muon in the final state may require a single electron and a single muon trigger or a combination of triggers with a predefined transverse momentum threshold. The events that pass the HLT are then kept in storage.
References


CHAPTER 5

Event Generation and Reconstruction

5. Methodology

In the Large Hadron Collider, after a proton-proton collision, a large number of particles are produced, with momenta of several orders of magnitude. This leads to multiple scattering and rescattering effects. In this rough environment it is very difficult to identify any particle or to make accurate measurements. The observation of new particles or the verification of any theoretical model can be made possible only by comparing the experimental data with the expected results according to the SM processes. Monte Carlo (MC) event simulations is a toolkit that models the SM processes or some times, processes beyond the Standard Model (BSM).

The MC generated events once produced by different generators, are passing through several steps, beginning from the modeling of the hard scattering until the observed particles (hadrons, bosons and leptons) in the final states. In this way all the physical processes like interactions between particles and decays of unstable particles are been simulated. The events then pass through GEANT4 [1], which simulates the passage of the particles through the detector and models the multiple scattering of events with the detector materials as well as the detector response to them. Then all information coming from subdetectors (energy deposits in calorimeters, particle trajectories and muon tracks) need to be reconstructed into physical objects. The reconstructed physics objects are asked to passed some identification criteria. Some of the object passed correctly these criteria while others failed to pass or falsie passed them. For that reason a data-driven method is used for comparing the simulated events with data and find the disagreement in reconstruction efficiency, isolation and energy resolution. These disagreement is express with scale factors applied to simulated events in order to increase the goodness of the analysis.

The analysis of $H \rightarrow aa \rightarrow bb \tau \tau$ channel is very challenging since the final states include several physical objects like b jets, leptons, charged and neutral hadrons and the missing energy derived from leptonic and hadronic tau decays. All of these objects need to be reconstructed and identified
in high accuracy through a large background contribution. This chapter begins with event generations of all the background and signal samples, continue with the detector simulation and finish with the addressing of all the methods used to reconstruct, identified and isolated the final objects of this channel: Muon, electron, jets, tau and missing energy.

5.1 Event Generation

Many generators have been constructed over the past decades in order to simulate the pp collision at LHC. The MC generators simulate the proton – proton collision divided in separated tasks, as illustrated in figure 5.1, in order to fully describe an event. The event generation processes can be summarized chronologically by the following steps.

The two protons traveling in opposite directions and collide. The nominal center of mass energy is expected to reach the value of 14TeV. Each proton is a collection of quarks and gluons, the so-called partons. The partons from each side, collide and this compose the hard scattering. The exact momentum of each parton is required in order to describe the hard scattering but the partons are only carrying a fraction of the momentum that the original protons had. The patron distribution functions PDFs, \( f(x, Q^2) \), come to solve this problem by giving the probability that a parton carry a momentum fraction \( x \) in the proton at a given scale \( Q^2 \). Such PDFs are provided by CTEQ [2], NNPDF [3] or MRST/MSTW [4] groups.

During the proton-proton collision, different physical processes are taking place, leading to a chain of events that need to be generated (figure 5.1):

- Partons before hard interaction as well as partons produced after the hard interaction can branch into other partons, resulting in a shower of secondary hadrons (Parton showering) [5].
- The remaining parts of the proton that did not take place in the hard interaction since are not color neutral radiate and hadronize themselves. The events coming from this process are called underlying events.
- During the hard scattering there is a possibility a short-lived resonance like W boson or top quark to decay.
- In the case where the particles energy scales are lower than 1GeV, the perturbation theory breaks down and replaced by a non-perturbation theory leading to partons recombination to hadrons without any color change, the so called Hadronization.
- The short-live particles like taus decay giving rise to more hadrons and leptons. For the tau decay simulation the Tauola [6] generator was used as an additional generator.

![Diagram](image)

**Figure 5.1**: The event generation processes begging from proton-proton collision [7].

These physical processes resulting from hadron collisions lead to the need of more sophisticated software programs simulating all those different steps with a more accurate approach. Many such programs were develop based on Matrix Element (ME) computations. The calculation of ME are done by parton-level or Matrix Element generators. Such generators are the MadGraph5, MadGraph5_aMC@NLO [8] and PowHEG [9,10], which are capable of calculating the ME with leading order (LO) or next-to-LO (NLO) accuracy. For example, the package MadGraph5 calculate the processes cross section at LO accuracy and the MadGraph5_aMC@NLO can incorporate NLO computations. In this analysis all those different generators are used. In cases where the events were generated in LO a k-factor was multiplied for scaling the difference between the NLO and LO
cross section [5]. The NLO calculations are more complicated and not available, so in some cases the events are generated in LO and rescaled as mentioned above with a scaling ratio.

5.1.1 Signal Event Generation

The MADGRAPH5_aMC@NLO generator is used for the $h \rightarrow aa \rightarrow 2\tau 2b$ signal process, for pseudoscalar masses between 15 and 60 GeV, with steps of 5 GeV for the gluon fusion (ggF) process, and 20 GeV for the vector boson fusion (VBF) process. A generator-level filter requiring the presence of at least one electron with $p_T > 20$ GeV or one muon with $p_T > 18$ GeV is applied to produce the MC samples; the efficiency varies between 10 and 15% depending on the mass of the pseudoscalar boson. The filter has a 100% efficiency for events selected in the analysis because the reconstructed $p_T$ thresholds are several GeV higher than the generated level thresholds. The ditau mass distributions are identical between the ggF and VBF processes.

The signal contributions in the VBF mode for masses for which no simulation exists is taken as identical to the ggF distributions, and rescaled by the average ratio of the signal normalization in the VBF and ggF modes for masses where both signal samples exist.

Signal samples in WH and ZH productions are produced for one mass point, $m_a = 40$ GeV. The cross sections of those signal samples of about 1.37 pb and 0.88 pb for WH and ZH respectively, are actually very small compared to ggH production. Therefore it is expected that VH events contribute to the signal region with much less events than gluon-gluon fusion. The variable used in this analysis to extract the results is the invariant mass of the two taus. This variable is distributed in the same way for gluon-gluon fusion and for the VH events, because the two taus originate from Higgs boson in both cases (although the normalization is expected to be different because of the different cross sections). Figure 5.2, shows the the shapes of the signal produced in ggH, WH and ZH productions normalized to unity for the $e\tau_h$ final state. In the low mass regions (left plot) the visible di-tau mass between ggH and VH samples are compatible but for larger mass (right plot) of the di-tau and the leading b system this compatibility is destroyed leading to a high mass tail in visible di-tau mass distribution. This is due to the probability that one lepton from W/Z decay is considered wrongly as a tau decay product, increasing the di-tau mass distribution.
Figure 5.2: Shapes (normalized to unity) of the signal produced in ggH, WH, and ZH productions, for the $e\tau_h$ final state in a category with low invariant mass between the leading $b$ jet and the tau candidates, and in a category with large mass.

In practice, the VH contribution is taken as equivalent in shape with the ggH contribution and rescaled by the VH to ggH ratio. More plots concerning the compatibility between ggH and VBF as well as ggH with VH signal samples for the three different final states are presented in Appendix A2. Figure 5.3 show the Feynman diagrams for all signal processes and table 5.1 represent the signal samples included in the analysis along with their cross sections.

5.1.2 Background Event Generation

For the analysis of $H \rightarrow aa \rightarrow bb\tau\tau$ channel as mention before, the background contribution is large due to the final states products. Leptons, b-jets, charged and neutral hadrons are the decay product of several physical processes. So a large number of background events coming from Drell-Yan (Z+jets), t\(\bar{t}\), single-top, diboson, W+jets and standard model Higgs decays are need to be generated. Feynman diagrams for some of background processes are shown in Figure 5.4. The SM backgrounds contribute along with their cross-sections are shown in table 5.2.

The Z+jets and W+jets processes are generated with the \texttt{MadGraph5_aMC@NLO} generator at leading order (LO) in perturbative quantum chromodynamics (QCD) with the MLM jet matching and merging [11]. The $Z +$ jets simulation contains non-resonant Drell–Yan production, with a minimum dilepton mass threshold of 10 GeV. Also a k-factor of 1.16 is considered for the Z+jets
samples and 1.21 for the W+jets samples since they are generated in LO. These k-factors were applied for scaling the difference between LO and next to the leading order (NLO) in which the other samples were generated. The FxFx merging scheme [12] is used to generate diboson background with the MadGraph5_aMC@NLO generator at next-to-LO (NLO). The $t\bar{t}$ and single top quark processes are generated at NLO with the POWHEG 2.0 and 1.0 generator [13–18].

![Feynman's diagrams for signal processes: a) gluon fusion, b) vector boson fusion (VBF) and c) vector boson associated production (VH).](image)

**Figure 5.3**: Feynman’s diagrams for signal processes: a) gluon fusion, b) vector boson fusion (VBF) and c) vector boson associated production (VH).

Background from SM Higgs boson production are generated at NLO with the POWHEG 2.0 generator [19]. POWHEG 2.0 is also used for the Wh and Zh simulated samples. The generators are interfaced with PYTHIA 8.212 [20] to model the parton showering and fragmentation, as well as the decay of the $\tau$ leptons. The full next-to-NLO (NNLO) plus next-to-next to-leading logarithmic (NNLL) order calculation [21–26], performed with the Top++ 2.0 program [27], is used to compute a $t\bar{t}$ production cross section equal to $832^{+29}_{-28}(\text{scale})\pm35(\text{PDF}+\alpha_S)\,pb$ setting the top quark mass to 172.5 GeV. This cross section is used to normalize the $t\bar{t}$ background simulated with POWHEG.
Figure 5.4: Feynman’s diagrams for some of the background processes: a) Drell-Yan, b) $t\bar{t}$, c) single top and d) W+jets decays. All of these decays give lepton, hadrons and jets in their final states as the signal processes.

Table 5.1: Signal samples included in the analysis along with their cross sections.

<table>
<thead>
<tr>
<th>Signal samples</th>
<th>Cross section (pb)</th>
<th>Cross section (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ggH \rightarrow aa \rightarrow b\bar{b}\tau\tau$</td>
<td>48.58</td>
<td>0.4858</td>
</tr>
<tr>
<td>$VBH \rightarrow aa \rightarrow b\bar{b}\tau\tau$</td>
<td>3.782</td>
<td>0.0378</td>
</tr>
<tr>
<td>$WH \rightarrow aa \rightarrow b\bar{b}\tau\tau$</td>
<td>$(0.5328 + 0.840)$</td>
<td>0.0137</td>
</tr>
<tr>
<td>$ZH \rightarrow aa \rightarrow b\bar{b}\tau\tau$</td>
<td>0.8839</td>
<td>0.0088</td>
</tr>
</tbody>
</table>

Table 5.2: MC background samples included in the analysis along with their cross sections estimated in Leading Order (LO) accuracy or in Next-to-Leading Order (NLO) accuracy. All those samples are generated for p-p collision at 13TeV. More detailed table can be seen in Appendix A1

<table>
<thead>
<tr>
<th>Background MC Samples</th>
<th>Cross Section (pb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drell-Yan+jets $(Z \rightarrow \ell\ell)$</td>
<td>62292.557 (Leading Order, LO)</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>831.76 (Next to Leading Order, NLO)</td>
</tr>
<tr>
<td>Single top</td>
<td>288.17 (NLO)</td>
</tr>
<tr>
<td>W+jets</td>
<td>64409.4 (LO)</td>
</tr>
<tr>
<td>Diboson $(WZ, WW, ZZ, VV)$</td>
<td>41.657 (NLO)</td>
</tr>
<tr>
<td>SM Higgs decays $(ggH, VBF, VH \rightarrow \tau\tau, WW)$</td>
<td>Depend on the decay</td>
</tr>
<tr>
<td></td>
<td>(See Appendix A1)</td>
</tr>
</tbody>
</table>
The cross section (times k factor) for the low mass ($10 < m_{\ell\ell} < 50$ GeV) Drell-Yan samples, as well as the junction with the high mass Drell-Yan samples, are verified in a $Z \to \mu\mu$ region. As shown in Figure 5.5, the agreement in the low mass region is very good, and there is no discontinuity at 50 GeV. The agreement in the very high mass region is not very good because no shape correction has been applied to the Drell-Yan simulated samples at leading order accuracy.

\textbf{Figure 5.5:} Invariant mass of the muons in a $Z \to \mu\mu$ region. The data/MC agreement is good in the region where the low mass Drell-Yan samples contribute, and there is a good continuity at 50 GeV.

### 5.2 Pile-up Events

During the hard scattering more than one proton-proton collision taking place while the bunches cross. The additional proton-proton collisions are called pile-up. All simulated samples include additional proton-proton interactions per bunch crossing, pile-up, obtained by generating concurrent minimum bias collision events (events with soft partonic interactions since pile-up events are in general low energetic) using PYTHIA. In PYTHIA the number of pile-up events per signal collision is described by a Poisson distribution and its mean value in strongly depends from the luminosity. More specific the number of events per crossing is given by the following equation:

$$N = \sigma L \varepsilon$$  \hfill (5.1)
where \( N \) is the number of events per crossing, \( \sigma \) is the total cross section (elastic and inelastic) of the proton – proton collision and \( \varepsilon \) is the detector efficiency. The detector efficiency is given by the fraction of the number of events detected in the detector \( N \), over the number of events \( N_p \), produced in the LHC in each bunch:

\[
\varepsilon = \frac{N}{N_p}
\]  

(5.2)

In the LHC ring, during 2016 data taking period, 2554 bunches were fill and the design bunches were 2808. Considering the Luminosity of 2016, \( L = 1.5 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \) and the inelastic \( p-p \) cross section at 6.5 TeV about 60mb, the mean value of pile up events is calculated as below:

\[
N \approx 60 \times 10^{-27} \text{ cm}^2 \cdot 1.4 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1} \cdot \frac{2808}{2554} \cdot 25 \text{ ns} = 23
\]  

(5.3)

For more accurate measurements the mean value of the pile-up is derived from CMS analysts during 2016. Figure 5.6 shows the recorder luminosity as a function of the mean number of interactions per bunch crossing for the 2016 \( p-p \) collisions at 6.5TeV. For this plot a minimum bias cross section of 69.2 mb is used. This cross section for inelastic \( p-p \) collisions gives the best agreement between the data and the theoretical predictions from Pythia and is recommended for CMS analyses. So, all simulated samples need to be reweighted in order to agree with the pile–up distribution in data. More details about pile-up reweighting is follow in chapter 6.

5.3 Simulation of the Detector

For realistic and observable final states the generated events are processed through a simulation of the CMS detector based on GEANT4 [1, 28]. GEANT4 is an excellent toolkit in the hand of physicists, simulating the response of CMS detector. GEANT4 software covers a large variety of physics processes such as interactions and decays of a flight or at rest particles, electromagnetic interactions (e.g: compton scattering, photoelectric effect, bremsstrahlung pair production), multiple scattering, nuclear processes, Cherenkov radiation, scintillation etc. We can simulate all the interactions of particles with the material of subdetectors as well as the readout electronics and the trigger levels of the experiment as well as various deficiencies, for example noise, dead channels etc. The information from the electronic signals derived from the simulated interactions, as well as
the real data, event by event, hit by hit or energy clusters depositions are digitized before reconstruction.

CMS developed a software package interfacing most of the event generators to used directly in the production of Monte Carlo simulation events. The overall software developed by the CMS Experiment was called CMSSW and can be utilized depending on the needs of each analysis since it is written in Python/C++ programming languages. The CMSSW contains a full simulation code as well as a fast simulation code and it also contains the various software algorithms develop to select various categories of the interesting data events as well as simulated events. The CMSSW version used for the present analysis is the CMSSW_8_0_26.

![CMS Average Pileup, pp, 2016, \( \sqrt{s} = 13 \text{ TeV} \)](image)

**Figure 5.6:** Mean number of interactions per bunch crossing for the 2016 pp run at 13 TeV. The inelastic cross section for this plot is taken equal to 69.2 mb.

### 5.4 Reconstruction and Identification

For the reconstruction and identification of final states physical objects a Particle flow PF algorithm is used in CMS. This algorithm combines information from all the subdetectors allowing an optimal discrimination between particles emerging from pp collisions. The form of information taken as an
input in the algorithm are the tracks from charged particles, the clusters of energy deposits from the calorimeters and the hits from the muon system. PF associates each event with a particle candidate and reconstructs all stable particles like charged and neutral hadrons, electrons, muons, and photons. Combinations of these PF objects are used to reconstruct higher-level objects such as jets, $\tau_b$ candidates, and missing transverse momenta.

5.4.1 Muon Reconstruction

The muon’s mass is almost 200 times larger than the electron mass and do not loose energy from Bremsstrahlung like electrons. They penetrate the subdetectors material without losing much of their energy and that is the reason for placing the muon system in the outer part of the detector. Therefore, muons leave their “footsteps” in all subdetectors making their identification and reconstruction much easier compared to the other particles.

Muons reconstruction is performed by combining information from the muon system and the tracker. The reconstruction starts from the local reconstruction in the CMS muon detector (standalone muons tracks) and the tracks left by these muons in the tracker system (tracker muon tracks), independently. Then the information from those two are used as input for the muon track reconstruction.

The local reconstruction uses information from the hits\(^1\) in the detection layers of RPC, CSC and DT chambers. The CSC and DT are multi layer detectors and the reconstruction of hits in each layer is building a linear track, the so called “segment”. The segments are used to built the muon trajectories using the Kalman Filter technique [29]. The information from RPC is also used, and the resulting muons are referred to as standalone muons.

The tracker muon tracks are built “inside out”. Each tracker track with a traverse momentum of $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV is considered as a muon candidate. Subsequently, those tracks are extrapolated to the muon system and if there is a match with at least one muon segment the track is qualified as a tracker muon track.

Combining the information from standalone muons and tracker muon tracks a “Global muon track” is built. The Global muon reconstruction is built “outside-in”, meaning that a standalone muon

\(^1\) After a muon ionize the gas in muon chambers gives rise to an electric signal call “hits”.
reconstructed in the muon system is attempted to be matched with at least two tracker tracks from the inner tracker. This matching is performed using the Kalman filter which collects the information from both standalone track and tracker muon track.

The efficiency of reconstructing a global or a tracker muon or both of them is as high as 99% due to the high efficiency of the tracker track and muon segment reconstruction technique. Also better efficiency for a low $p_T$ muon candidate is performed by the tracker muon reconstruction.

**5.4.1.1 Muon Identification**

The reconstructed muons are then passed through the Particle Flow (PF) algorithm, which combines the information from all subdetectors to identify individual particles for each event. The identification of muons in PF is performed using a set of selection criteria to the reconstructed tracks in the muon system and in the inner tracker. Such criteria are for example the normalized $\chi^2$ of their global track, the segments compatibility probability between the tracker muon tracks and the muon segments or their compatibility with the primary vertex. Using these variables one can get different working points used in the physics analysis. In this dissertation for muon identification, a “Medium” Muon Identification working point was used.

The muon identification criteria are designed to be highly efficient for prompt muons originating at the primary vertex and the muons from heavy quark decays. For the medium muon identification working point, a set of requirements must be satisfied [30]:

- To be a “loose” muon, a global or a track muon.
- Fraction of valid tracker hits (inner tracker) > 80%.
- Position matched between tracker tracks and standalone muon (chi-square local position) <12.
- Kick Finder on the inner track $\chi^2_{\text{max}}< 20$.
- Segment compatibility between the tracker and muon tracks (if the muon is only reconstructed as a tracker muon) > 0.451.
- Segment compatibility between the tracker and muon tracks (if the muon is reconstructed as both tracker and a global muon) > 0.303.
- Goodness of fit $\chi^2/\text{dof}$ (if the muon is reconstructed as both a tracker and a global muon) < 3
5.4.1.2 Muon Isolation

The muon isolation over other objects is not a trivial task. The relative isolation of muon is evaluated as the ratio between its absolute isolation (numerator) and its transverse momentum $p_T$ as shown in equation 5.1. The relative isolation of muons is used to distinguish prompt muons from weak decays within jet (misidentified leptons). The equation below define the muon relative isolation:

$$I_\ell = \frac{\sum_{\text{charged}} p_T + \max(0, \sum_{\text{neutral}} p_T + \sum_{\gamma} p_T - \delta \beta \sum_{\text{charged,PU}} p_T)}{p_T}$$

(5.4)

For the computing of relative isolation, the sum of charged hadrons $p_T$ coming from the primary vertex within a cone size $\Delta R = \sqrt{(\Delta \phi)^2 + (\Delta \eta)^2}$ is summed with the neutral hadrons and photons $p_T$ sum $\sum_{\text{neutral}} p_T + \sum_{\gamma} p_T$ in the same cone size. The other term of eq. 5.4 $\sum_{\text{charged,PU}} p_T$ is the $p_T$ sum of charged hadrons originating from pileup vertices, estimated from simulation and is subtracted from the neutral particle $p_T$ sum, so as to correct the pile-up distribution to the neutral particles. The $\delta \beta$ correction term is estimated to be equal to 0.5, denoting that the neutral particles over the charged particles coming from pileup vertices is equal to $\frac{1}{2}$. The $p_T$ of the lepton is denoted $p_T^{\ell}$. The relative muon isolation for this analysis is taken lower than 0.15 within a cone size of $\Delta R = 0.4$.

5.4.1.3 Muon Reconstruction Identification and Isolation Efficiency

The efficiency for the muon reconstruction and identification is evaluated by applying the Tag and Probe technique to a pair of muons coming from $Z$ decays $Z \rightarrow \mu^+ \mu^-$ [31]. In more details, one of the two muons pass as the tight identification criteria (“tag” muon) and the other pass as more relaxed conditions (“probe” muon). The tag muon is a general global muon with a very small fake rate while the probe muon is selected by pairing the desired particle with the tag so as their invariant mass is consistent to the $Z$ resonance mass. Subsequently, the subtraction of the combinatorial background is eliminated by a separate fit on the invariant mass of the di-muon for passing and failing Tag and Probes with the signal and background shapes. Finally, the efficiency is measured.
from the ratio of the two signal shapes, which is actually the number of probes that pass the desired selection criteria:

\[
\varepsilon = \frac{P_{\text{pass}}}{P_{\text{all}}} \tag{5.5}
\]

where \(P_{\text{pass}}\) is the number of probe muons that pass the criteria and \(P_{\text{all}}\) is the number of all probe muons [32].

The efficiency for muon reconstruction and identification for “loose” and “tight” muons as a function of eta is shown in figure 5.7 for both data (2016) and simulation. The loose muon is a tracker or a global muon selected by the Particle Flow (PF) algorithm. The loose muon identification working point aims to distinguish prompt muons from light and heavy flavor decays and also reduce the misidentification rate of charged hadrons as muons. The tight muon is reconstructed in both as a global muon and as a tracker muon. The tight muon identification working point aims to reduce more the misidentification rate by suppress muons coming from flavor decays (muons from hadrons) and the remnants of a hadron shower (hadron punch-through) that penetrates the calorimeter detector. The reconstruction efficiency for loose working point exceed the 99%, while for the tight working point the average efficiency for all the eta range is about 97%.

![Figure 5.7: Reconstruction and identification efficiency for loose (left) and tight (right) muons with p_T > 20 GeV. The Tag and Probe method is used for data taken in 2015 and MC simulations [33].](image)
The muon isolation efficiency using also the Tag and Probe technique is presented in figure 5.8. The efficiency is calculated relatively to the identification criteria. The resulting plots show a very good agreement between the data and MC less than 0.5 for all the $p_T$ range (20 - 200 GeV). Also the probability of a muon with $p_T > 20$ GeV that satisfies the tight isolation requirements and passes the tight isolation criteria is about 5% in the barrel region ($|\eta| < 0.9$) and about 15% in the endcap region ($0.9 < |\eta| < 2.4$) [33].

![Figure 5.8: Isolation efficiency of muons that pass the isolation requirements on top of the identification criteria using the Tag and Probe method. The left plot presents the certain efficiency as a function of muon transverse momentum and the right plot as a function of eta [33].](image)

5.4.2 Electron Reconstruction

The electron reconstruction and selection are very important for the CMS physics program such as SM precision physics, Higgs coupling measurements, and beyond the Standard Model searches. These analyses need a precise measurement of electron momentum and energy, an excellent reconstruction efficiency and a very low misidentification probability. All of these have been achieved by improving the performance of the online and offline electron reconstruction algorithms.

Electrons are reconstructed by matching the tracks from the silicon detector (tracker) with the energy deposits in the ECAL. In similar way as muon reconstruction, for an optimal performance, information from the electromagnetic calorimeter ECAL (stand-alone approach) is matched with the information from the inner tracking (tracker tracks) using a “global particle flow” (PF) algorithm.
The electron reconstruction is not so easy since electrons lose a part of their energy via bremsstrahlung radiation on the way to ECAL. For example, an average of 33% of an electron energy is lost before reaching the ECAL and about 86% in the intervening materials at |n|≈1.4. Therefore, to measure the total energy of an electron, it is necessary to collect all the energy lost, by photon radiation, which is mostly spread in the \( \phi \) direction due to the magnetic field (the magnetic field bending the electron trajectory).

The collecting of electron energy (clustering) in the electromagnetic calorimeter ECAL, is performed with two different “clustering” algorithms, the “hybrid” algorithm for the barrel and the “multi–5×5” algorithm for the encap regions. The two different algorithms are needed because of the different subdetectors geometry.

The hybrid algorithm takes into account the geometry of the EB and collects the energy in \( \eta \times \phi \) direction. In more detail, the algorithm starts from the seed crystal, which contains the largest energy deposit \( E_T > 1\text{GeV} \). Around the seed crystal, energy is collected from a set of crystal arrays of 5×1 (strips) in the \( \eta \times \phi \) direction, if their energy threshold is greater than 0.1GeV. Then all clusters from the contiguous crystal arrays are merged to build a supercluster (SC), if the energy of each cluster is at least 0.35GeV [34].

In the endcaps region the multi–5×5 algorithm starts from the seed crystals with energy greater than 0.18GeV. Around the seeds, the energy is collected from 5×5 crystals, which are then grouped into a supercluster if their total energy is greater than 1GeV, for \( \eta | < 0.07 \) and \( |\phi| < 0.3 \) rad from the seed crystal. Finally, the energy collected in the preshower is added also to the supercluster energy.

The electron track reconstruction, as in the muon case, is performed by the Kalman filter (KF) technique. Although a large fraction of energy lost due to bremsstrahlung in the tracker material leads to a poor estimation of electron tracks. For that reason, a dedicated tracking procedure is used and, for avoiding time consuming, this procedure begins from seeds which correspond to the electron track initiation.

For the seeding two algorithms are used. The results from both algorithms at the end are combined. The first, the ECAL-based seeding, begins from Superclustering (SC) the energy deposit (PF clustering algorithm) and then extrapolates the trajectories toward the collision vertex, for
estimating the electron trajectory. The second one, tracker-based, works from the opposite direction (inside-out). It begins from the track reconstruction in the tracker (KF algorithm), and is then matched with the SC after extrapolation toward the ECAL.

The KF algorithm reconstructs the electron trajectory with a very good accuracy if the bremsstrahlung is limited. So up to the ECAL, the PF clusters are matched with the KF track with a very good precision. In the presence of bremsstrahlung (tracker material) the KF algorithm cannot describe the electron trajectory and the Gaussian fit in the energy loss distribution (assumed by KF) gives a large value of chi-square, meaning a bad identification. Therefore, the energy loss distribution is refitted with a sum of Gaussian distributions using the Gaussian sum filter (GSF). The quality from both KF and GSF fit along with the energy matching of the ECAL and tracker information are used in the MVA (multivariate analysis) [35] which typically gives a better performance.

5.4.2.1 Electron Isolation and Identification

Isolation requirements are needed for isolating the prompt electron from misidentified jet or from genuine electrons resulting from semileptonic decays of b/c quarks. The electron isolation is computed exactly the same way as muon using equation 5.4. The first term represents the sum of the charged PF candidates \( p_T \) and the second the sum of neutral particles (hadrons and photons) \( p_T \), within a cone size \( \Delta R \). The \( \delta \beta \) correction which corrects the energy sum of neutral particles by subtracting the contribution from pileup vertices, is set to 0.5. For this analysis and more specifically for the final states \( e\mu \) and \( e\tau \), the electron relative isolation is taken lower than 0.10, within a cone size of \( \Delta R=0.3 \).

Information from both tracker and ECAL provides various choices of variables for the electron identification. For example, the observables that qualify the agreement between the ECAL and the tracker measurements, or the energy deposit of suppercluster based on calorimeter information only. Although, the ultimate and most efficient is the use of MVA techniques for the early dataset taken in 2016, with a center of mass energy of 13TeV, which gives the best performance.

This MVA has two working points depending on the electron transverse momentum and pseudorapidity. Table 5.2 indicates the two working points with 80% and 90% signal efficiency as measured on electron from Drell-Yan events.
Table 5.3: Two MVA working points as measured on electron DY events [35].

<table>
<thead>
<tr>
<th>Category</th>
<th>MVA_min cut (80% signal eff)</th>
<th>MVA_min cut (90% signal eff)</th>
</tr>
</thead>
<tbody>
<tr>
<td>barrel (eta&lt;0.8) pt above 10 GeV</td>
<td>0.941</td>
<td>0.837</td>
</tr>
<tr>
<td>barrel (eta&gt;0.8) pt above 10 GeV</td>
<td>0.899</td>
<td>0.715</td>
</tr>
<tr>
<td>endcap pt above GeV</td>
<td>0.758</td>
<td>0.357</td>
</tr>
</tbody>
</table>

5.4.2.2 Electron Isolation and Identification Efficiency

The isolation and identification efficiency is of primary importance for the analysis with an electron in the final state. The most constituent way is using the Tag and Probe method applied on $Z \rightarrow e^+ e^-$. In the tag and probe method we ask one of the two electrons to pass tight identification criteria (tag electron) and the other one just to give along with the first one the invariant mass of the $Z$ boson (probe electron).

![Electron identification efficiency](image)

*Figure 5.9*: Electron identification efficiency for the data collected in (2016) with an integrated luminosity of 35.9 fb$^{-1}$ and at the center of the mass energy of 13 TeV. The bottom plots show the ratio of data over MC efficiency [36].

The efficiency is estimated relatively to the identification criteria: Loose, Medium and Tight. In figure 5.9 the identification efficiency measured in MVA discriminator is presented, with an average
efficiency of 80%, as a function of the electron transverse momentum for different pseudorapidity ranges. This certain working point was used for the final states of $e\mu$ and $e\tau$, for electron identification in this analysis.

### 5.4.3 Jet Reconstruction

The quarks and gluons that are produced in pp collisions develop a parton shower and after hadronization jets are dominating the CMS detector. Jets are reconstructed using the PF algorithm. The PF algorithm collects information from all the CMS sudetectors, initially reconstructing stable particles like leptons and hadrons and then proceeds to the reconstruction of more complex objects like jets.

One of the main jet reconstruction algorithms used in the CMS collaboration is the anti-$k_T$ algorithm [37]. The anti-$k_T$ algorithm is a sequential clustering algorithm built on the idea that entities (“protojets”) with the smallest distance should be joined to form a new protojet.

Each protojet is characterized by its transverse energy $E_T = P_{i,T}$ (tranverse momenta), its azimuthal angle $\phi_i$ and its pseudorapidity, $\eta_i = -\ln(\tan(\theta_i/2))$. The distance between two protojets is defined as:

\[
d_{i,j} = \min(E_{T,i}^2, E_{T,j}^2) \left[ (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2 \right] / R^2
\]

(5.6)

\[
d_i = E_{T,i}^2
\]

(5.7)

where $E_{T,i}, E_{T,j}$ are the energies deposited of the two i and j protojets and $R$ is a cone parameter taken equal 0.4 in CMS Run-2, so as to handle high multiplicity final states. In Run-1 this parameter was set to 0.5 but for Run-2 it was found that with $R=0.4$ the algorithm is less sensitive to pileup. The algorithm as a next step finds the smallest distance of $d_{ij}$ or $d_i$, and if $d_{ij}$ is the smallest then the two protojets are joint to form a new protojet, labeled as Jet. The procedure is repeated until all the protojets become a part of a jet or a jet itself.

After jet reconstruction, corrections on their energy is implemented. Away from the theoretical definitions of a jet, the final output of a given jet-algorithm can be very different taking into account
the interaction of particles with the materials of the subdetectors. Thus, one needs to calibrate the reconstructed jet energy, by matching its energy with the energy of the generated jet (GenJet).

In the CMS experiment jet energy corrections derive in both simulation and data. A list of the main corrections are defined below:
- Offset corrections: An offset in the jet energy scale may be caused by electronic noise or pile-up (PU) events. For the subtraction of the extra energy deposits two different methods are used: the Average Offset (AO) and the Jet Area (JA) [38]. The Average Offset (AO) method is based on MC and data-driven techniques (matching simulated events with data events) and the Jet Area (JA) method is based on the FastJet algorithm [39]. Both methods show a linear combination of pile-up contribution with the number of primary vertices. To be able to control the PU, the response of the jets should be independent of primary vertex. Therefore, the response of the detector is expressed as the ratio of the $p_T$ of the PF jet (jets produced from PF candidates) to the $p_T$ of the PF jets matched GenJet (jets produced from generator level MC particles).
- Relative $\eta$ dependency: The jet energy response should be flat as a function of pseudorapidity and for that reason a calibration is applied to the relative energy scale.
- Absolute $p_T$ dependence: The energy should be uniform as a function of the transverse momentum so the absolute energy scale is calibrated.
- Electromagnetic fraction: The electromagnetic fraction of the jet energy deposited in the ECAL over the energy deposited at both ECAL and HCAL. The correction on the electromagnetic fraction reduce any differences in the jet energy scale.

Subsequently in this chapter, a special reference is made to the b-jet identification, since it plays a crucial role in the analysis presented in this thesis.

5.4.3.1 B Jet Reconstruction and Identification

B jets identification algorithms use variables based on the properties of heavy-flavour hadrons. The jets originate from b quarks hadronization and radiation and can be distinguished from other c-quark or light leptons (“uds”) using information from tracks, vertex and identified leptons.

The hadrons consisted of b quarks have a larger lifetime compared to hadrons made of c quarks and this leads to a typical displacement of tracks with respect to the primary vertex (flight distance). Thus gives rise to the existence of secondary vertices as illustrated in figure 5.10 [40]. The impact
parameter (IP), which is the distance of closest approach of an extrapolated track to the primary vertex is a powerful tool to discriminate b jets from light-flavour (uds) jets.

Additionally, b and c quarks have larger masses than light leptons and consequently when a heavy-flavor hadron decays its products have larger transverse momentum and wider jets relatively to the other jets originating from light quarks or gluons. Therefore, this is another way except from the tracks displacement and secondary vertices for separating b-jets from light leptons.

**Figure 5.10:** Schematic view of a heavy-flavor jet with a secondary vertex (SV) from the b-quark decay which is significantly displaced with respect to the primary vertex (PV) [40].

For the b-jet identification (b-tagging) three algorithms are used. The first one (impact parameter significance algorithm) uses the impact parameter of a track, the Secondary Vertex algorithm is based on the information from the secondary vertex and the Soft Lepton algorithm uses the properties of leptons present in the jet. The list below presents the common starting point for the three algorithms showing several quality cuts to ensure their power:

- b jet tracks are required to have a $p_T$ greater than 1 GeV (the well reconstructed tracks considered)
- chi-square normalized to the number of degrees of freedom (n.d.o.f) in the fit lower than 5
- number of hits in the tracker layers greater or equal to two.
- the point of closest approach from the primary vertex to be less than 5cm (reduce the tracks coming from long-lived $\kappa^0$ and $\Lambda$ hadrons)
- the absolute value of the transverse impact parameter of the track $|d_{xy}| < 0.2$ cm while the longitudinal impact parameter $|d_z| < 17$ cm
- the tracks are required to be within a cone size of $\Delta R < 0.3$ and have a track distance from the jet axis below 0.07cm
- the 2D flight distance should be more than 2.5 cm or less than 0.01 cm
- secondary vertex (defined as a vertex sharing less than 65% of its track with primary vertex) candidate with radial distance more than 2.5 cm and with mass greater than 6.5 GeV/c^2 is rejected (reducing the more the impact from long-lived mesons)
- 3D impact parameter value is at least 50 μm
- IP significance at least 1.2.

More details about the selection criteria and the different algorithms can be found in reference [40]. Only a brief reference will be made for the Combine Secondary Vertex algorithm which is used in this analysis for identification of the b jets.

The Combine Secondary Vertex (CSV) algorithm [41] is one of the two main algorithms based on the secondary vertex information used in CMS Run1 (CSV) and Run2 (CSVv2). It combines the information from secondary vertices along with the track-based lifetime information. Compared to the other algorithms it gives maximum efficiency because it can provide discrimination even if there is no secondary vertex. If no secondary vertex is found, the algorithm combines the tracks with significance IP greater than 2 with a “pseudo vertex”. When it is not possible to find any other pseudo-vertex it proceeds to track-based variables. Table 5.3 shows a set of variables used to identified b jets using the CSV algorithm and also make a comparison with the variables used in Run-1.

Based on the misidentification probability, the CSV algorithm established three different “working points” (WP). These WP are the “Loose”, the “Medium” and the “Tight” with their misidentification probability to be around 10%, 1% and 0.1% respectively. In this thesis the Medium-CSV working point was used. Figure 5.11 below shows the misidentification probability of c and light-flavour jets as a function of b jet identification probability for different b taggers in simulated t\bar{t} events. The requirements were that the jet transverse momentum p_T > 20 GeV and the absolute value of pseudorapidity |η| < 2.4.
Table 5.4: Variables set as input for CSV algorithm for Run1 and Run2. “x” means that the variable is used and “-” denotes that the variable is not used [40].

<table>
<thead>
<tr>
<th>Input variable</th>
<th>Run 1 CSV</th>
<th>CSVv2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV 2D flight distance significance</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Number of SV</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Track $\eta_{rel}$</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Corrected SV mass</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Number of tracks from SV</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SV energy ratio</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>$\Delta R(sv, jet)$</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>3D IP significance of the first four tracks</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Track $p_{T,rel}$</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>$\Delta R$ (track, jet)</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Track $p_{T,rel}$ ratio</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Track distance</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Track decay length</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Summed tracks $E_T$ ratio</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>$\Delta R$ (summed tracks, jet)</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>First track 2D IP significance above $c$ threshold</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Number of selected tracks</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Jet $p_T$</td>
<td>—</td>
<td>x</td>
</tr>
<tr>
<td>Jet $\eta$</td>
<td>—</td>
<td>x</td>
</tr>
</tbody>
</table>

Figure 5.11: Misidentification probability as a function of the b jet identification for different b-tagging algorithms. Simulated $t \bar{t}$ events are used with the jet $p_T > 20$ GeV and $|\eta| < 2.4$ [40].
The b jet identification efficiency for the medium working point using the CSVv2 algorithm is about $\sim 65\%$ for a misidentification rate for c-quark and light-flavors (udsg) of $\sim 10\%$ and $\sim 1\%$, respectively (fig. 5.6). The DeepCSV algorithm for the same working point improved the efficiency compared to CSVv2 almost 7%. In future analysis the DeepCSV tagger will replace the CSVv2 algorithm [40].

5.4.4 Tau Lepton Reconstruction

The decay of Higgs boson to a pair of taus is one of the most promising channels to the search for new phenomena in pp collisions. Recently, the decay of the SM Higgs to a pair of tau leptons was observed and the results are reported in reference [42, 43]. The tau lepton has the largest mass ($m_\tau = 1776.86$ MeV), compared to the other leptons and can decay to hadrons and leptons. About 35% of the tau decay leptonically and 65% hadronically. When a tau decays leptonically it means that it decays to an electron or a muon with a tau neutrino and an electron or a muon neutrino respectively. The hadronic decays of a tau lepton contain hadrons in different combinations:

- one prong: includes one charged hadron, (charged pioni or Kaon), $\tau \rightarrow h^- \nu_\tau$,
- three prong: includes three charged hadrons, $\tau \rightarrow h^+ h^- \nu_\tau$, $\tau \rightarrow h^+ h^- \pi^0 \nu_\tau$,
- one prong plus two strips: includes one charged hadrons and two neutral pions, $\tau \rightarrow h^- \pi^0 \pi^0 \nu_\tau$,
- one prong plus one strip: includes one charged hadrons and one neutral pion, $\tau \rightarrow h^- \pi^0 \nu_\tau$.

Some of the above decay modes include an intermediary resonance such as $\rho(770)$ and $\alpha_1(1260)$. Table 5.5 shows the tau decay modes along with their corresponding branching fractions in %.

The identification of electron and muon from tau decay are implemented by a lepton identification algorithm as mentioned previously. The charged candidates used in the reconstruction of $\tau_h$ (hadronic decays of tau) require to have transverse momentum $p_T > 0.5$ GeV, matched to a generator level tau (originating from the primary vertex) and a distance from primary vertex $|d_{xy}| < 0.1$ cm in the transverse plane. These requirements ensure the quality of the corresponding tracks and the minimal requirements so as the charged particles pass through the tracker layers. Taking into account all of them, the Hadron-plus-strips (HPS) algorithm is used to reconstruct and identify the hadronic decays of tau lepton.
Table 5.5: Hadronic and leptonic decay modes of tau lepton with the corresponding branching fraction in %. h denote a charged hadron, a pion or a Kaon [42].

<table>
<thead>
<tr>
<th>Decay mode</th>
<th>Resonance</th>
<th>$\mathcal{B}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leptonic decays</td>
<td>35.2</td>
<td></td>
</tr>
<tr>
<td>$\tau^- \rightarrow e^- \bar{\nu}<em>e \nu</em>\tau$</td>
<td></td>
<td>17.8</td>
</tr>
<tr>
<td>$\tau^- \rightarrow \mu^- \bar{\nu}<em>\mu \nu</em>\tau$</td>
<td></td>
<td>17.4</td>
</tr>
<tr>
<td>Hadronic decays</td>
<td>64.8</td>
<td></td>
</tr>
<tr>
<td>$\tau^- \rightarrow h^- \nu_\tau$</td>
<td></td>
<td>11.5</td>
</tr>
<tr>
<td>$\tau^- \rightarrow h^- \pi^0 \nu_\tau$</td>
<td>$\rho(770)$</td>
<td>25.9</td>
</tr>
<tr>
<td>$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$</td>
<td>$a_1(1260)$</td>
<td>9.5</td>
</tr>
<tr>
<td>$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$</td>
<td>$a_1(1260)$</td>
<td>9.8</td>
</tr>
<tr>
<td>$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$</td>
<td></td>
<td>4.8</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>3.3</td>
</tr>
</tbody>
</table>

5.4.4.1 The Hadron-Plus-Strips Algorithms

The hadrons-plus-strips (HPS) algorithm aims at reconstructing the different hadronic decay modes starting from constituents of reconstructed jets. These include charged hadrons as well as neutral pions. The latter promptly decay into pairs of photons, which have a high probability of converting into $e^+e^-$ pairs while traversing the tracker material. The large magnetic field from the CMS solenoid leads to a spatial separation of the $e^+e^-$ pairs in the plane generated by the azimuthal angle $\phi$ and the pseudorapidity $\eta$. In order to reconstruct the full energy of neutral pions, electron and photon candidates are clustered into “strips”. The clustering is done via an iterative procedure. The highest $p_T$ electron or photon ($e/\gamma$) not yet included into any strip is used to seed a new strip. The initial position of the strip in $\eta$ and $\phi$ is set to the $\eta$ and $\phi$ of the seed $e/\gamma$. The next highest $p_T e/\gamma$ that is within an $\eta \times \phi$ window centered on the strip location is merged into the strip. The construction of the strip ends in case no further $e/\gamma$ is within the $\eta \times \phi$ window. In this case the clustering proceeds by constructing a new strip, that is the seed by the next highest $p_T e/\gamma$.

In the version of the HPS algorithm used in Run-1 (“fixed-strip”), the size of the $\eta \times \phi$ window was set to a fixed value of 0.05 $\times$ 0.20 in the $\eta - \phi$ plane. However, the size of the strip in the Run-1 algorithm is sometimes not adequate to contain all electrons and photons originating from the hadronic tau decay. These decay products can be integrated as part of the signal by widening the strip size. Conversely, if the hadronic tau lepton has a large $p_T$ the decay product tends to be boosted.
in the $\tau_h$ flight direction. In this case, a smaller strip size than the one considered in Run-1 can reduce background contributions in the strip while accounting for all $\tau_h$ decay products.

Based on these considerations the dynamic strip clustering of the HPS algorithm (“dynamic-strip”) was introduced for Run-2. The change with respect to the fixed-strip algorithm is that the $\eta \times \phi$ window around the seed of the strip does now depend on the $p_T$ of the $e/\gamma$ candidate to be added to the strip and the $p_T$ of the strip itself.

\[
\Delta \eta = f(p_T^\gamma) + f(p_T^{strip}) \\
\Delta \phi = g(p_T^\gamma) + g(p_T^{strip})
\] (5.8) (5.9)

The function $f$ and $g$ are determined using a single $\tau$ gun MC sample such that 95% of all electrons and photons that are due to $\tau_h$ decay products are contained within a strip. The functional form is derived based on the distance in $\eta$ and in $\phi$ between $\tau_h$ and $e/\gamma$, as a function of $e/\gamma$ $p_T$.

\[
f(p_T) = 0.20 \cdot (p_T^{-0.66})
\] (5.10)

\[
g(p_T) = 0.35 \cdot (p_T^{-0.71})
\] (5.11)

The upper limit of the strip size is set to 0.3 in $\Delta \phi$ and 0.15 in $\Delta \eta$ and the lower limit is set to 0.05 for both. Then using the $p_T$-weighted average of all $e/\gamma$ in the strip the strip position is recomputed as below:

\[
\eta_{\text{strips}} = \frac{1}{p_T^{\text{strips}}} \sum p_T^{e/\gamma} \eta_{e/\gamma}
\] (5.12)

\[
\phi_{\text{strips}} = \frac{1}{p_T^{\text{strips}}} \sum p_T^{e/\gamma} \phi_{e/\gamma}
\] (5.13)

Finally, in Run2 the HPS algorithm changes the way of strip clustering and the strip size now does not depend on the cone size but the weighted $p_T$ of the strip is required to be within the signal cone while a part of it can lie outside the cone.
5.4.4.2 Tau Isolation and Identification

The reconstructed \( \tau_h \) candidates are required to be isolated. That means that no other hadrons or photons should be included in the certain isolation cone. The main aim is to reduce the misidentification probability of jet faking the \( \tau_h \). For that reason two isolation methods were used. The one is based on the isolation sum commonly for muon and electron isolation, namely “Isolation sum” discriminator and the other way is by using an “MVA-based” discriminator.

The “Isolation sum” \([\text{tauIso}]\) of \( \tau_h \) candidates is measured by summing the scalar values of the \( p_T \) of charged particles and photons with \( p_T >0.5 \text{ GeV} \) reconstructed with the PF algorithm within a cone size \( \Delta R<0.5 \), centered on the \( \tau_h \) direction and with \( |dz|<0.2 \text{ cm} \). The effect of pile up is reduced by subtracting the sum of charged candidates originating from the vertex of \( \tau_h \) with \( |dz|>0.2 \text{ cm} \) and \( \Delta R<0.8 \) around the \( \tau_h \) direction. The \( \Delta \beta \) correction factor is set to 0.2 instead of 0.46 in Run1 and is used to correct the difference between the two isolation cones \( \Delta R = 0.5 \) and \( \Delta R = 0.8 \). The equation below define the isolation as in equation 5.4:

\[
I_{\tau_h} = \sum_{dz<0.2 \text{ cm}, \Delta R<0.5, p_T>0.5 \text{ GeV}} p_T^{\text{charged}} + \max \left( 0, \sum_{\Delta R<0.5, p_T>0.5 \text{ GeV}} p_T^{Y}, \sum_{dz>0.2 \text{ cm}, \Delta R<0.8, p_T>0.5 \text{ GeV}} p_T^{charged}, -\Delta \beta \right) \tag{5.14}
\]

Be requiring the isolation to take different values, different working points are defined. The isolation working points defined for Run2 are the “loose WP”, with \( I_{\tau_h}<2.5 \text{ GeV} \), the “medium WP”, with \( I_{\tau_h}<1.5 \text{ GeV} \) and the “tight WP”, with \( I_{\tau_h}<0.8 \text{ GeV} \). Each WP point leads to different identification efficiency. The identification efficiency for previous and current version of the HPS algorithm is shown in figure 5.12. The results are evaluated for \( Z \to \tau \tau \) and for QCD multijet events. The \( \tau_h \) identification efficiency is estimated as the percentage of \( \tau_h \) decays with visible decay products with \( p_T>18 \text{ GeV} \) and \( |\eta|<2.3 \), that are matched to the generator level taus with \( p_T>20 \text{ GeV} \) and \( |\eta|<2.3 \), and pass the decay mode finding criterion (1 prong, 1prong + \( \pi^0 \), 1 prong + 2 \( \pi^0 \) and 3 prong) [42].
Figure 5.12: Misidentification probability as a function of $\tau_h$ identification efficiencies estimated for $Z \rightarrow \tau\tau$ and for QCD multijet events using four methods of reconstruction and isolation. The three different working points are indicated with three different markers for each method beginning with the loose working point from the left to the medium and then to the right [42].

The MVA-based discriminator [44] is used for the discrimination of $\tau_h$ decays from gluons and quark jets. It combines all the isolation criteria with the tau lifetime. Then, a boosted decision tree (BDT) is used to reduce more the misidentification probability [tauIso]. The brief list of variables that are used in the MVA-based discriminator as input in the BDT consist of the following elements:

- charged and neutral isolation sums (eq. 5.11)
- $\Delta\beta$ correction (eq: 5.11)
- $p_T$-weighted, $\Delta R$, $\Delta \eta$, $\Delta \phi$ (relative to the $\tau_h$ axis) and $p_T^{\text{strips}}$ (eq.: 5.5)
- 3D Impact parameter and its significance
- 2D Impact parameter and its significance
- the multiplicity of photons and electron candidates with $p_T>0.5\text{GeV}$ in the signal/isolation cone
- flight length and its significance

All the above information lead to the development of a powerful discriminator.

The different working points of the MVA-based discriminant are set through different requirements on the BDT in such a way so as to keep a uniform efficiency over $p_T$. Six working points are study, beginning from Vloose, Loose, Medium, Tight, Vtight, VVTight with 90%, 80%, 70%, 60%, 50% and 40% identification efficiency respectively. The misidentification probability as a function of
identification efficiency of $\tau_h$ is presented in figure 5.13. The results were estimated using $Z \rightarrow \tau\tau$ and QCD multijet events. The working points for each isolation method are shown with a different marker style.

![Diagram](image)

**Figure 5.13:** Misidentification probability versus $\tau_h$ identification using the two different isolation methods (Sum isolation and MVA_base isolation). For the MVA_discriminant six working points (very loose, loose, medium, tight, very tight and very very tight) are used and for the isolation-sum three working are taken: loose, medium and tight [42].

The MVA-based discriminant reduces the misidentification efficiencies almost by $\sim 3\%$ compared to the other method. In this analysis for final states including hadronic taus, it is chosen that the $\tau_h$ should pass the tight working point with efficiency almost 65% for less than 1% fake rate.

### 5.4.4.3 Tau Isolation Against Electron and Muon

Except from jets also muons and electrons have a high probability to be misidentified as a hadronic tau in different decay modes. The electron passing through the tracker material looses energy via bremsstrahlung radiation. The radiated photons, are mimicking the neutral pions in 1 prong or in 1 prong plus a neutral pion decay mode. The above two discriminators are also designed to separate muons and electrons from hadronic taus. They improve so as to reduce the electron/muon faking hadronic taus misidentification probability, keeping the high selection efficiency of genuine hadronic taus.
For the $\tau_h$ discrimination against muon, two working points corresponding to different $\tau_h$ identification efficiencies and misidentification rate are developed: “loose WP against muon” and “tight WP against muon”. The $\tau_h$ candidate fails to pass the loose working point against muon if the fraction of energy deposit in the calorimeters associated to the leading track is less that 20% of its track momentum or if there are track segments in at least two muon detector planes within a cone size $\Delta R = 0.3$ around the tau direction. The $\tau_h$ candidate fails to pass the tight working if its hits are found within a cone $\Delta R = 0.3$ around the tau direction in the CSC, DT or RPC chambers in the two outermost muon stations. For the final state $\mu\tau_h$ of this analysis the loose working point against muon is used so as to separate the muons from $\tau_h$ candidates.

For the discrimination of tau leptons against electrons the BDT is trained to discriminate electrons from hadronic taus using as input a list of variables described in reference [44]. For Run2 a couple of variables related to photons, were added. These variables are the number of photons in any of the strips associated to the track, the $p_T$-weighted RMS in the $\eta$ and $\phi$ directions of all photons included in the strips and the fraction of energy of $\tau_h$ candidates carried away by photons.

![Figure 5.14](image.png)

**Figure 5.14:** Efficiencies curves of $\tau_h$ identification as a function of $p_T$ of the reconstructed hadronic tau for different working points using simulated $Z/\gamma^* \rightarrow \tau \tau$ events [42].

The variables where the BDT is more sensitive are the fractions of energy deposited in the ECAL of the tau candidates over the energy deposited in both ECAL and HCAL, and the fractions of energy
deposited in each calorimeter separately over the transverse momentum of each one of the corresponding tracks. The efficiency curves of tau reconstruction as a function of $p_T$ of the reconstructed hadronic tau for different working points are shown in figure 5.14. The misidentification probability of electron faking hadronic tau as a function of the $p_T$ of the generated electron, taking into account the selected working points is presented in figure 5.15. The different working points of MVA-based discrimination algorithm used are the very loose, loose, medium, tight, and very tight.

![Figure 5.15](image)

*Figure 5.15:* Misidentification probability as a function of the $p_T$ of the generated electron for different Wps using simulated $Z/\gamma^* \rightarrow ee$ events [42].

### 5.4.5 Missing Transverse Energy ($E_T^{\text{miss}}$)

The accurate measurement of the missing transverse momentum plays a very important role in many searches of particle physics, especially in analysis with neutrino in the final states. The neutrino particle and the hypothetical neutral weakly interacting stable particles cannot be detected by CMS, since they do not interact with the detector materials. Their presence can only be inferred from the momentum imbalance in the plane of the beam direction. A proper reconstruction of the missing transverse momentum can be estimated via the precise measurement of the detectable particles such as $e, \mu, \tau$, charged and neutral hadrons, $\gamma$ and jets and alongside taking into account several reasons that can lead to a bias measurement. The bias measurement can be introduced via pileup
interactions, misidentification of physical objects, inefficiencies in the tracker, the energy thresholds in the calorimeters and the nonlinearity of its response.

The transverse missing energy \( \vec{E}_T^{\text{miss}} \), in the particle flow algorithm is given by the negative vectorial sum of the transverse momentum of all reconstructed particles:

\[
\vec{E}_T^{\text{miss}} = - \sum_i \vec{p}_i
\]  

(5.14)

while the bias on the missing \( \vec{E}_T^{\text{miss}} \) measurement can be reduced by correcting the \( p_T \) of the jets (\( \vec{p}_T^{\text{corr}_j} \)) as shown below:

\[
\vec{E}_T^{\text{miss corr}} = \vec{E}_T^{\text{miss}} - \sum_{\text{jets}} (\vec{p}_T^{\text{corr}_j} - \vec{p}_T^{\text{jet}})
\]  

(5.15)

This correction \( \vec{E}_T^{\text{miss corr}} \) is nominated as type-I correction and corrects all jets with \( p_T > 15 \text{GeV} \) which submit less than 90% of their energy in the ECAL calorimeter. The threshold on \( p_T \) was set to 15GeV because the uncertainties in the energy scale of jets for lower values of \( p_T \) is so large that gives also large uncertainties in \( \vec{E}_T^{\text{miss}} \).

**Figure 5.16:** HLT trigger efficiencies of \( \vec{E}_T^{\text{miss}} \) as a function of generated \( \vec{E}_T^{\text{miss}} \). The red line presents the corrected results and the green line the uncorrected [45].
In a similar way HLT corrects the $\vec{E}_{\text{miss}}$ with the difference that it takes into account the negative vectorial sum of all the jets in all the $p_T$ range. Figure 5.16 presents the HLT trigger efficiencies of the corrected and uncorrected $\vec{E}_{\text{miss}}$ as a function of $\vec{E}_{\text{miss}}$ at the generator level. More details about the performance of missing energy reconstruction in CMS using 13 TeV collision data can be found in reference [45].
References


[34] The CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at √s = 8 TeV”, (2015), JINST, 10(06):P06005.


CHAPTER 6

Dataset and Corrections to Simulated Samples

The aim of this analysis is to indicate the presence or not of a signal coming from the decay of the Standard Model Higgs boson (125GeV) to a pair of light pseudoscalars bosons and further decay to a pair of b jets and a pair of tau leptons. In this chapter, the full dataset used for this exotic Higgs boson decay is presented. The search is based on p-p collision data, collected at a center of mass energy 13 TeV, corresponding to an integrated luminosity of 35.9fb⁻¹ [1].

Comparing the simulated events with real data is the key for understanding the results coming from the detector and lead to the conclusion about whether or not this exotic decay could be observed. So, in order to achieve the best possible modeling of background and signal events over data, a series of corrections are applied to the simulated samples.

Through this analysis three of the six ττ final states are studied which are the eμ, eτh, and μτh. The τhτh final state is dropped because of the high trigger thresholds in that final state and final states ee and μμ are also discarded due to their very low branching fraction. In section 6.1 a brief reference is made for the dataset used. In section 6.2 a more extensive reference is made for all the corrections applied to simulated samples to achieve a better agreement between them and the data.

6.1 Data Samples

Three ditau final states are studied in this analysis: eτb, μτb, and eμ. The analysis is based on data collected in 2016, amounting to 35.9 fb⁻¹ collected at a center-of-mass energy of 13TeV. Masses of the pseudoscalar boson between 15 and 60 GeV in step of 5GeV are probed.

For each of the three ττ final state different triggers were used, which are the MuonEG for eμ, SingleMuon for μτb, and SingleElectron for the eτb final state. For all the data collected in 2016 the
miniAOD\(^1\) format was used. The trigger \(p_T\) thresholds depending on the \(|\eta|\) threshold are going to be discussed in the following Section 6.3. The table 6.1 show the list of dataset along with their integrated luminosity per data set are shown. More information about the dataset can be found in Appendix A1.

**Table 6.1**: List of data sets included in the analysis.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Data Set</th>
<th>Integrated Luminosity (fb(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>e(\mu)</td>
<td>MuonEG/Run2016B-03Feb2017/miniAOD</td>
<td>35.866</td>
</tr>
<tr>
<td>(\mu\tau_h)</td>
<td>SingleMuon/Run2016B-03Feb2017/miniAOD</td>
<td>35.866</td>
</tr>
<tr>
<td>e(\tau_h)</td>
<td>SingleElectron/Run2016B-03Feb2017/miniAOD</td>
<td>35.866</td>
</tr>
</tbody>
</table>

### 6.2 Corrections of Data/Simulation and Events Weights

A series of corrections are applied to simulated events due to the difference between them and the collected data. As described in Chapter 5 all the physical objects are reconstructed, isolated and identified using different algorithms. Each algorithm has an identification or misidentification rate which is applied to the simulated events to correct the efficiency of each process, so as to be compatible with real data. A brief reference on the different Scale Factors (SF) used to correct the triggers and particles identification efficiencies of simulated samples is discussed below.

#### 6.2.1 \(\tau_h\) Energy Scale

In the simulation, the hadronic tau energy scale is the fraction of the reconstructed \(\tau_h\) energy over the energy of visible tau decay products at the generator level. A common technique to control the energy scale of a simulation is to fit the simulated sample distributions to real data. In this case and according to reference [2] to control the simulated energy scale, a sample of \(Z \rightarrow \tau_\mu \tau_h\) in data is used. The visible mass (of hadronic tau + muon) distribution and hadronic tau mass distribution are fit to data, taking into account the different tau decay modes. The ratio of data /MC derived from these fits are used as Scale Factors to correct differences between the simulated and observed data.

---

\(^1\) The miniAOD is a high-level tier. The motivation for the Mini-AOD format is to have a small and quickly derived data format from which the majority of CMS analysis users can start their analysis work. Its includes informations about high level physics objects, a full list of reconstructed particles by the PF, MC truth information and trigger informations.
The Scale Factors (SF) used to correct the energy of $\tau_h$ in simulation as a function of the decay mode are the following: -1.8% for 1-prong, +1.0% for 1-prong + neutral pion and +0.4% for 3-prong. Note, that the correction is applied in this analysis to all reconstructed $\tau_h$, only if they are matched to $\tau_h$ in generator level and that the traverse missing energy (MET) is corrected subsequently.

6.2.2 $\tau_h$ Identification Efficiency

For correcting the tau identification between data and MC, different SF depending on the WP were estimated. The TAU POG (TAU Physics Object Group) used tag-and-probe method in $Z \rightarrow \tau_\mu \tau_h$ events and derived to a data/MC scale factor equal to 0.95. More specifically events with a muon and tau candidate, with the tau passing more relaxed conditions, are separated in two categories. In the first category the muon and tau pass the WP of the discriminator in contrast with the second category which failed to pass the same WP. Each region distribution is fit with a maximum likelihood fit and the ratio of those two postfit plots are considered as the tau identification efficiency scale factor. Figure 6.1 shows the postfit plots for the pass and fail categories. This SF corresponds to the tight MVA WP and is applied to all simulated events to correct the hadronic tau identification efficiency. The uncertainty based on the method used to estimate the certain scale factor is set to 5%.

![Figure 6.1: Postfit plots of the visible mass distribution for the pass (left) and fail (right) category using tag and probe method [3].](image-url)
6.2.3 $e/\mu \rightarrow \tau_h$ Identification Efficiency

There is a misidentification probability in all reconstructed objects. For example, the reconstructed tau can rise from a real hadronic decay as desirable, but can also rise from an electron, a muon or a jet. Therefore, the corrections must be applied from the origin of the reconstructed object. For that reason a correction is applied using the “MC matching” recipe, in which the reconstructed hadronic tau should be matched to the generated muon or electron to be corrected.

The correction factors, derived within the TAU POG (TAU Physics Object Group which its main aim is to reconstructed and identified the tau lepton), are applied. These corrections are substantial and strongly depend on the pseudorapidity of the fake $\tau_h$. The table 6.2 presents the scale factors applied based on different range of $|\eta|$ and MC matching recipe (if the reconstructed tau is coming from a prompt muon/electron).

6.2.4 $e/\mu \rightarrow \tau_h$ Energy Scale

Corrections to the energy scale of electrons and muons faking $\tau_h$ are applied. They depend on the reconstructed decay mode. The largest correction amounts to +9.5%, and corresponds to 1- prong + $\pi^0$ decays for electrons faking $\tau_h$ candidates. The reconstructed $\tau_h$ should be matched to generated muons or electrons to be corrected in such a way.

Table 6.2: Scale factors used to correct the efficiency of electron/moun faking hadronic tau.

| Type          | $|\eta|$ | Scale Factor (SF) |
|---------------|---------|------------------|
| Prompt muon   | $|\eta| < 0.4$ | 1.012            |
|               | $0.4 < |\eta| < 0.8$ | 1.007            |
|               | $0.8 < |\eta| < 1.2$ | 0.870            |
|               | $1.2 < |\eta| < 1.7$ | 1.154            |
|               | $|\eta| > 1.7$ | 2.281            |
| Prompt electron | $|\eta| < 1.460$ | 1.213            |
|                | $1.460 < |\eta| < 1.558$ | 1.000            |
|                | $|\eta| > 1.558$ | 1.375            |
6.2.5 $p_T^{\text{miss}}$ Recoil Corrections

Based on the HTT working group recent results, recoil corrections are applied to the Higgs, Drell Yan and W+jets MC samples and are not applied to the $t\bar{t}$, single-top and dibosson MC samples. The missing transverse momentum is estimated via the Type I Particle Flow algorithm and MVA discriminator as described in section 5.4.5. These corrections are applied on event by event basis to both of them and are strongly depended on the generated $p_T$ of the parent boson [4]. The list of corrected parameters are shown in table 6.3 [5].

6.2.6 Drell Yan Corrections

The $Z$ $p_T$ distribution is seen to disagree between LO simulations and data in a $Z \rightarrow \mu\mu$ control region with at least one b tagged jet. Weights are derived as a function of this variable to make the MC distributions match data (without correcting the overall yield), as shown in Figure 6.2. These weights are applied to the $Z \rightarrow \tau\tau/\ell\ell$ simulations in the signal region, as a function of the generated $Z$ $p_T$ (because the reconstructed variables are biased due to the neutrino from tau decays). In events with at least one b tagged jet, the invariant mass of the dilepton pair is in good agreement between data and MC (see Fig. 6.3), and no special reweighting is applied as a function of this variable.

Table 6.3: List of parameter in which recoil corrections are applied.

<table>
<thead>
<tr>
<th>For Type I PF (MET)</th>
<th>For MVA (MET)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- uncorrected PF MET $p_x$</td>
<td>- uncorrected MVA MET $p_x$</td>
</tr>
<tr>
<td>- uncorrected PF MET $p_y$</td>
<td>- uncorrected MVA MET $p_y$</td>
</tr>
<tr>
<td>- generator Z/W/Higgs $p_x$</td>
<td>- generator Z/W/Higgs $p_x$</td>
</tr>
<tr>
<td>- generator Z/W/Higgs $p_y$</td>
<td>- generator Z/W/Higgs $p_y$</td>
</tr>
<tr>
<td>- visible Z/W/Higgs $p_x$</td>
<td>- visible Z/W/Higgs $p_x$</td>
</tr>
<tr>
<td>- visible Z/W/Higgs $p_y$</td>
<td>- visible Z/W/Higgs $p_y$</td>
</tr>
<tr>
<td>- hadronic jet multiplicity</td>
<td>- hadronic jet multiplicity</td>
</tr>
<tr>
<td>- corrected PF MET $p_x$</td>
<td>- corrected MVA MET $p_x$</td>
</tr>
<tr>
<td>- corrected PF MET $p_y$</td>
<td>- corrected MVA MET $p_y$</td>
</tr>
</tbody>
</table>

After the dilepton $p_T$ correction, most distributions agree well between data and simulation, including the distribution of the leading b jet $p_T$. However, the distribution of the invariant mass of
the dilepton and b jet system shows some disagreements, as seen in Figure 6.4. An additional correction is applied as a function of the mass of the dilepton + b jet system, to correct for this trend.

**Figure 6.2:** Transverse momentum of the dimuon system in $Z \rightarrow \mu\mu$ events + 1 b jet before (left) and after (right) correction. The agreement in the second plot is by construction. The correction does not modify the overall yield of the Drell-Yan background predicted from simulation.

**Figure 6.3:** Invariant mass of the dimuon system in $Z \rightarrow \mu\mu$ events + 1 b jet after the correction on the $p_T$ of the dimuon system.
6.2.7 Pileup Reweighting

The pileup in CMS is estimated using the luminosity-based way, in which the instantaneous luminosity for all of the colliding bunches and the total inelastic cross section is taken into account [6]. For example if one knows the revolution frequency, the instantaneous luminosity can be estimated. Then, by multiplying it with the total number of inelastic cross section, one can calculate the number of interactions.

On the other hand, the MC samples are generated with a pileup distribution which is meant to cover the expected pileup for a certain data-taking period. Although the primary vertex reconstruction used is efficient and behaves well at high level pileup, the pileup distribution is very sensitive to the details of the primary vertex reconstruction, and can be biased by the offline selection criteria. Therefore, the pile-up (PU) distribution from MC and Data, present some differences.

These differences between Data and MC pileup are corrected by reweighting the MC distribution of the number of primary vertices so as to match the number of pileup interactions in data. The data distribution is obtain assuming a minimum bias cross section of 69.2mb (recommended by CMS analyses in order to have the best agreement with data) and correspond to an integrated luminosity
of 35.9 fb\(^{-1}\). The ratio derived from this comparison is used as a correction weight for the simulated events. Figure 6.5 shows the data/MC distribution before and after reweighting.

![Figure 6.5: Primery vertex distribution before (left) and after (right) pile-up reweighting in the \(\mu\tau\) final state [7].](image)

### 6.2.8 Electron and Muon Identification Efficiencies

The electron and muon identification efficiency were measured in both data and MC samples. This was done using different techniques like the tag and probe method and different triggers depending on the physical objects in the final state. Fitting the data over the MC sample distribution, a ratio of Data/MC is derived. This ratio is used as a Scale Factor and was applied to all simulated events to correct their identification efficiencies. In Appendix A3 all the efficiency plots for MC and data used are presented, to estimate the scale factors for correcting the identification/isolation efficiency or to correct the HLT efficiencies. All the scale factors strongly depend on the pseudorapidity and the transverse momentum of the leptons and are dependent on the final state objects.

### 6.2.9 Top \(p_T\) Reweighting

Corrections to the \(t\bar{t}\) simulation depending on the generated \(p_T\) of the top and antitop quarks are applied. The reason for this weight (scale factor) is the disagreement observed in the \(p_T\) spectra of the top quark between the collected data from Run 1 and Run2 and the various MC samples based
The Scale Factor is meant to correct only the shape of the $p_T$ distribution of the top quark and the measurement covers only the range of $p_T$ lower and equal to 400 GeV [8].

The weight is derived as a function of the generated $p_T$ of both the top and antitop quarks:

$$weight = \sqrt{SF(\text{top}) \cdot SF(\text{antitop})}$$

(6.1)

with the following Scale Factor function:

$$SF(x) = \exp(0.0615 - 0.0005x)$$

(6.2)

The numbers 0.0615 and -0.0005 were derived in Run2 for the 13 TeV center of mass energy.

6.2.10 Trigger Efficiency

The trigger efficiencies for single muons, single electron and the muon-electron/gamma (MuonEG) triggers are measured in data from the HTT group using the tag and probe method. The triggers used for this analysis are strongly dependent on the final state. The Scale Factor applied to correct the MC samples are derived from the trigger efficiency over the efficiency on MC samples, which is considered equal to one, since there are no trigger requirements applied to the MC process.

6.2.11 B Tagging Efficiency

The results from the methods described in section 5.4.3.1, for estimating the identification efficiency of a b jet or the misidentification probability of fake b jets, do not match with the data. For that reason the B Tag and Vertices POG (BTV Physics Object Group) measured the b-tagging efficiency scale factors. These scale factors are depend on the Working Points (WP) of the Combine Secondary Vertex (CSV) algorithm, on the hadron flavor of the jets and on the $p_T$ of the jets. More details can be found in reference [9].


References


CHAPTER 7

Event Selection Algorithm and Optimization

For the analysis of the $H \to aa \to bb\tau\tau$ channel, three of the six final states are study due to high $p_T$ threshold or low branching fractions [1]. Events are selected based on different criteria depending on the final state. Then, in order to increase the sensitivity of the analysis a certain method is followed. The events in each final state are divided into four different categories depending on the signal over background ratio. A set of cuts are applied to the four categories so as to reduce the various background and optimize the expected limits on the production cross section and branching fractions.

7.1 Events Selection

Three ditau final states are studied in this analysis: $e\tau_h$, $\mu\tau_h$, and $e\mu$. The $\tau_h\tau_h$ channel is discarded because the trigger thresholds for $p_T$ are too high to keep sufficient signal acceptance (at least $p_T > 40$ GeV offline for each $\tau_h$), and the ee and $\mu\mu$ channels are discarded because of their low branching fraction and high contributions to the Drell-Yan background. Some $p_T$ and $\eta$ distributions plots for the three final states ($e\tau_h$, $\mu\tau_h$, and $e\mu$) at the generated level are shown in Figures 7.1 – 7.6.

The $p_T$ thresholds, dictated by the triggers for electrons and muons, or by the reconstruction for visible hadronic taus, are clearly in the tails (especially for electrons and muons). The efficiency for selecting just one electron or muon compatible with trigger thresholds is a few percent only. Figures 7.7 – 7.9, shows the efficiency of these cuts based for different visible tau $p_T$ thresholds and similar for the b-jet selection.

Events in the $\mu\tau_h$ channel are required to pass a combination of $\eta$-restricted single muon triggers with online threshold 22 GeV, or a combination of muon+tau cross triggers (online muon $p_T > 19$ GeV, online $\tau_h p_T > 20$ GeV). If the offline muon $p_T$ is between 20 and 23 GeV, we require the cross trigger to be fired. If the muon $p_T$ is above 23 GeV, the single muon trigger should be fired.
Figure 7.1: Transverse momentum distribution plots at generated level for the three final state channels for pseudoscalar mass $m_a=20\text{GeV}$. 
Figure 7.2: Transverse momentum distribution plots at generated level for the three final state channels for pseudoscalar mass $m_a=40\text{GeV}$.
Figure 7.3: Transverse momentum distribution plots at generated level for the three final state channels for pseudoscalar mass $m_a=60\text{GeV}$. 
Figure 7.4: Pseudorapidity distribution plots at generated level for the three final state channels for pseudoscalar mass $m_a = 20\text{GeV}$.
Figure 7.5: Pseudorapidity distribution plots at generated level for the three final state channels for pseudoscalar mass $m_a = 40\text{ GeV}$. 
Figure 7.6: Pseudorapidity distribution plots at generated level for the three final state channels for pseudoscalar mass $m_a = 60 \text{GeV}$.
The muon is required to pass the medium identification working point and to have a relative isolation (within cone size of 0.4) less than 0.15. The \( \tau_h \) candidate should pass the Tight working point of the MVA isolation (about 65% efficiency for less than 1% fake rate), the very loose MVA discriminator against electrons, and the loose discriminator against muons. The loose discriminator against muons is chosen over a tighter working point to keep a high signal efficiency, and because the \( Z \rightarrow \mu \mu \) background does not peak where the signal is expected (due to lower \( \mu \tau_h \) invariant mass). The \( \tau_h \) should satisfy the requirements \( p_T > 25 \text{ GeV} \) and \( | \eta | < 2.3 \). The \( p_T \) threshold of \( \tau_h \) has been optimized to yield the best expected limits. The muon and the \( \tau_h \) are separated by at least \( \Delta R = 0.4 \), and have opposite signs.

Events in the \( e\tau_h \) channel are selected with a single electron trigger with an online threshold 25 GeV. Offline, the electron is required to have \( p_T > 26 \text{ GeV} \) and \( | \eta | < 2.1 \), whereas the \( \tau_h \) candidate should have \( p_T > 25 \text{ GeV} \) and \( | \eta | < 2.3 \). The \( p_T \) threshold of \( \tau_h \) has been optimized to yield the best expected limits. The electron should pass the medium working point with 80% MVA identification efficiency and have relative isolation (cone size of 0.3) less than 0.10. The \( \tau_h \) candidate should pass the tight working point of the MVA isolation (about 65% efficiency for less than 1% fake rate), the very loose MVA discriminator against electrons and the loose discriminator against muons. The very loose discriminator against electrons is chosen over a tighter working point to keep a high signal efficiency, and because the \( Z \rightarrow ee \) background does not peak where the signal is expected (lower \( e\tau_h \) invariant mass); this increases the signal acceptance by more than 20% compared to the case were the usual working points are used. The electron and the \( \tau_h \) are separated by at least \( \Delta R = 0.4 \), and have opposite signs.

Events in the \( e\mu \) channel are selected with a combination of asymmetric electron+muon triggers. The leading leptons are required to have \( p_T > 24 \text{ GeV} \) (1 GeV above online trigger threshold), and the subleading one \( p_T > 13 \text{ GeV} \). The \( p_T \) of the subleading electron is dictated by trigger thresholds. The \( p_T \) of the subleading muon was taken a bit higher than what is permitted by the trigger thresholds in order to minimize the expected limits. Better limits are obtained with higher subleading muon \( p_T \), despite the loss in signal acceptance, because the fake rate for jets to be identified as a muon is much larger at low \( p_T \), leading to a large QCD multijet background. This also explains why the single muon triggers are not used as an OR with the electron+muon triggers even if this would increase the signal acceptance for events with soft electrons. The electron should pass the medium MVA identification working point (80% identification efficiency), and have relative
isolation (cone size of 0.3) less than 0.10. The electron and the muon are separated by at least $\Delta R = 0.3$, and have opposite signs.

For all final states, events that have an additional isolated and identified electron or muon are discarded. The events should all contain at least one b-tagged jet (medium Combine Secondary Vertex, CSV) with $p_T > 20$ GeV and $| \eta | < 2.4$. Figures 7.10 – 7.12 show the $p_T$ and $\eta$ distributions for each of the two b jets as well the $\Delta R$ between the two b jets, at the generated level. The two b jets in signal events are typically soft ($p_T < 20$ GeV).

**Figure 7.7:** The 2D plots shows the efficiency for the cuts based on different visible tau $p_T$ thresholds for each channel for $\mu\tau$ (top left), for $e\tau$ (top right), for $e\mu$ (bottom left) and similar for the bjet selection (bottom right). The $m_a$ mass set to 20 GeV.
Selecting 2 b jets would lead to a very low signal acceptance and to a high rate of light flavor jets to be misidentified as b jets. Figure 7.13 compares the signal acceptance for the different final states when requiring one or two b jets; the acceptance is about one order of magnitude larger if only one b jet is required. Requiring exactly one b jet would reduce the signal acceptance by about 10% (it is a combined effect of the acceptance and b tagging efficiency), and the $\tau\bar{\tau}$ acceptance by about 35% for every final state (see Table 7.1). This would lead to worse limits for all $m_a$ masses except for very low $m_a$. At $m_a = 15$ GeV, for which the dominant background is $\tau\bar{\tau}$, the limits by requiring exactly one b jet would be better by only $0 – 5\%$ depending on the final state. All the selection criteria are summarized in table 7.2.

\textbf{Figure 7.8:} The 2D plots shows the efficiency for the cuts based on different visible tau $p_T$ thresholds for each channel for $\mu\tau$ (top left), for $e\tau$ (top right), for $e\mu$ (bottom left) and similar for the bjet selection (bottom right). The $m_a$ mass set to 40 GeV.
**Figure 7.9:** The 2D plota shows the efficiency for the cuts based on different visible tau $p_T$ thresholds for each channel for $\mu \tau_h$ (top left), for $e \tau_h$ (top right), for $e\mu$ (bottom left) and similar for the bjet selection (bottom right). The $m_a$ mass set to 60 GeV.
Figure 7.10: $p_T$ and $\eta$ distribution plots for each b jet as well as the $\Delta R$ between the two b jets at generated level for pseudoscalar $m_a = 20$ GeV.
Figure 7.11: $p_T$ and $\eta$ distribution plots for each b jet as well as the $\Delta R$ between the two b jets at generated level for pseudoscalar $m_a = 40$ GeV.
Figure 7.12: $p_t$ and $\eta$ distribution plots for each b jet as well as the $\Delta R$ between the two b jets at generated level for pseudoscalar $m_a = 60$ GeV.
Table 7.1: Acceptance (in %) of the tt process with genuine leptons in the different final states before the categorization and the optimized selection.

<table>
<thead>
<tr>
<th></th>
<th>≥ 1 b jet</th>
<th>Exactly 1 b jet</th>
<th>≥ 2 b jet</th>
</tr>
</thead>
<tbody>
<tr>
<td>eμ</td>
<td>0.156</td>
<td>0.100</td>
<td>0.056</td>
</tr>
<tr>
<td>eτ&lt;sub&gt;h&lt;/sub&gt;</td>
<td>0.0357</td>
<td>0.0087</td>
<td>0.0049</td>
</tr>
<tr>
<td>μτ&lt;sub&gt;h&lt;/sub&gt;</td>
<td>0.0266</td>
<td>0.0170</td>
<td>0.0099</td>
</tr>
</tbody>
</table>

Figure 7.13: Comparison of the signal acceptance when requiring at least 1 or 2 b-tagged jets (medium CSV) for the eμ (bottom left), eτ<sub>h</sub> (top right), μτ<sub>h</sub> (top left) and for all (bottom right) final states.
Table 7.2: Baseline selection criteria for the objects selected in the various final states.

<table>
<thead>
<tr>
<th></th>
<th>$\mu\tau$</th>
<th>$e\tau$</th>
<th>$e\mu$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T(\tau_h)$</td>
<td>&gt;25 GeV</td>
<td>&gt;25 GeV</td>
<td>-</td>
</tr>
<tr>
<td>$p_T(\mu)$</td>
<td>&gt;20 GeV</td>
<td>-</td>
<td>&gt;24/10 GeV</td>
</tr>
<tr>
<td>$p_T(e)$</td>
<td>-</td>
<td>&gt;26 GeV</td>
<td>&gt;23/24 GeV</td>
</tr>
<tr>
<td>$p_T(b)$</td>
<td>&gt;20 GeV</td>
<td>&gt;20 GeV</td>
<td>&gt;20 GeV</td>
</tr>
<tr>
<td>$</td>
<td>\eta(\tau_h)</td>
<td>$</td>
<td>&lt;2.3</td>
</tr>
<tr>
<td>$</td>
<td>\eta(\mu)</td>
<td>$</td>
<td>&lt;2.1</td>
</tr>
<tr>
<td>$</td>
<td>\eta(e)</td>
<td>$</td>
<td>-</td>
</tr>
<tr>
<td>$</td>
<td>\eta(b)</td>
<td>$</td>
<td>&lt;2.4</td>
</tr>
<tr>
<td>Isolation ($\tau_h$)</td>
<td>MVA Tight</td>
<td>MVA Tight</td>
<td>-</td>
</tr>
<tr>
<td>Isolation ($\mu$)</td>
<td>&lt;0.15</td>
<td>-</td>
<td>&lt;0.15</td>
</tr>
<tr>
<td>Isolation ($e$)</td>
<td>-</td>
<td>&lt;0.10</td>
<td>&lt;0.10</td>
</tr>
<tr>
<td>ID ($\mu$)</td>
<td>Medium</td>
<td>-</td>
<td>Medium</td>
</tr>
<tr>
<td>ID ($e$)</td>
<td>-</td>
<td>MVA 80%</td>
<td>MVA 80%</td>
</tr>
<tr>
<td>ID ($b$)</td>
<td>Medium CSV</td>
<td>Medium CSV</td>
<td>Medium CSV</td>
</tr>
</tbody>
</table>

7.2 Categorization

We separate the events in categories based on the visible invariant mass of the leptons and of the leading b-tagged jet. The masses are computed from the visible decay products of the taus, and only one b-tagged jet is considered, therefore the reconstructed Higgs boson mass is expected to peak below 125 GeV.

The mass for the $e\mu$ final state is expected to be lower when compared to final states with $\tau_h$ because of the larger number of neutrinos from tau decays. The distributions of the invariant mass of the leptons and of the b jet for different pseudoscalar masses are shown in Figure 7.14 and for the background samples are shown in Figure 7.15. There is a strong dependence of the distributions on the pseudoscalar mass hypothesis, which leads to larger reconstructed mass of the leptons and b jet with a lower $m_a$. The mass of the 3 objects is higher when the pseudoscalar $a$ is light. This can be explained by the fact that the missing b-tag jet in the mass calculation would be closer to the
reconstructed b jet for a signal with lower $m_\tau$ because of the boost of the pseudoscalar bosons, leading to a larger reconstructed mass.

The SVfit algorithm [2] that has been developed for the search of $H \rightarrow \tau \tau$ and more specifically for tau pair reconstruction is not used. The algorithm is highly efficient in reconstructing the mass of the ditau system and it would lead to a higher background acceptance for a same signal efficiency. As shown in Figure 7.16 the signal is distributed to lower invariant masses of the lepton and the b jet, compared to the main background distribution. Using the SVfit, the signal will be shifted to higher masses, where there is more background.

*Figure 7.14:* Visible invariant mass of the leptons and the leading b jet, after the baseline selection, in the $e\mu$ (top left), $e\tau_h$ (top right), and $\mu\tau_h$ (bottom) final states. The uncertainty is statistical only.
Each final state is separated in four categories: the Low-mass signal region, the Medium-mass signal region, the High-mass signal region and the High-mass control region. The four regions have different signal-to-background ratios. These ratios also strongly depend on the pseudoscalar mass hypothesis. The thresholds to define the categories vary depending on the final state and are driven by the background and signal acceptances.

**Figure 7.15:** Visible invariant mass of the leptons and the leading b jet, after the baseline selection, in the $e\mu$ (top left), $e\tau_h$ (top right), and $\mu\tau_h$ (bottom) final states. The ratio plots below show the disagreement between the observed and expected events. The disagreement between observed and expected events for the $\mu\tau_h$ and the $e\tau_h$ is lower than 20%, where for the $e\mu$ is less than 10%.
The $e\mu$ thresholds are lower because $m_{b\tau\tau}$ is lower in signal events due to the large number of neutrinos from $\tau$ decays. The $e\tau_h$ thresholds are higher than the $\mu\tau_h$ ones because the signal and background acceptances at low $m_{b\tau\tau}$ are very low due to higher lepton $p_T$ thresholds at the trigger level. The most sensitive category depends on the signal mass, with the first categories being sensitive for heavy resonances and the last ones for light resonances.

The Low-mass signal region has a large signal purity for most of $m_a$ hypotheses but rather low signal acceptance. It is characterized by the large sensitivity for most $m_a$ values, because of the low background contamination. The thresholds of this category are $m_{b\tau\tau}<65\,\text{GeV}$ for the $e\mu$ final state, $m_{b\tau\tau}<80\,\text{GeV}$ for the $e\tau_h$ final state and $m_{b\tau\tau}<75\,\text{GeV}$ for the $\mu\tau_h$ final state.

The Medium-mass signal region thresholds are: $65 < m_{b\tau\tau} < 80\,\text{GeV} (e\mu)$, $80 < m_{b\tau\tau} < 100\,\text{GeV} (e\tau_h)$ and $75 < m_{b\tau\tau} < 95\,\text{GeV} (\mu\tau_h)$. This region contains a larger fraction of the signal (the fraction depends on the pseudoscalar mass hypothesis) but has a larger background contamination than compared to the Low-mass region.

![Figure 7.16: Distribution of $m_{b\tau\tau}^{\text{vis}}$ in the $\mu\tau_h$ final state. The signal histogram corresponds to 10 times the SM production cross section for ggh, VBF, and Vh processes, and assumes $B(h \rightarrow aa \rightarrow 2\tau2b) = 100\%$. The ratio plot below show a small disagreement between observed and expected events < 20\%. Only statistical uncertainties are shown.](image)
The High-mass signal region thresholds are: \(80 < m_{b\tau\tau} < 95 \text{ GeV (e}\mu\text{)}, 100 < m_{b\tau\tau} < 120 \text{ GeV (e}\tau_h\text{)}\) and \(95 < m_{b\tau\tau} < 115 \text{ GeV (}\mu\tau_h\text{)}\). This region contains the last signal events, most from low \(m_a\) and suffers from larger background contaminations.

The High-mass control region thresholds are: \(m_{b\tau\tau} > 95 \text{ GeV (e}\mu\text{)}, m_{b\tau\tau} > 120 \text{ GeV (e}\tau_h\text{)}\) and \(m_{b\tau\tau} > 115 \text{ GeV (}\mu\tau_h\text{)}\). This category is almost signal-free (with the exception of a small fraction of the signal for low \(a\) boson masses) and helps to constrain the backgrounds with higher statistics.

The visible mass of the tau candidates is used as observable in this analysis. As shown in Figures 7.17 – 7.19, the visible ditau mass in signal events is partially correlated with \(m_{\tau\tau}\), which is used to define the categories. The backgrounds remain more or less flat in \(m_{\tau\tau}\) in the individual categories. Therefore, the results of the search are extracted from a fit of the visible ditau mass distributions in each of the categories.

![Figure 7.17](image)

**Figure 7.17:** Correlation between the visible mass of the tau candidates \((m_{\tau\tau})\), and the visible combine mass of the tau candidates and the leading b tagged jet \((m_{b\tau\tau})\), in the e\(\mu\) final state, for \(m_a= 15 \text{ GeV (top left), 40GeV (top right) and 60 (bottom) GeV.}\)
Figure 7.18: Correlation between the visible mass of the tau candidates ($m_{\tau\tau}$), and the visible combine mass of the tau candidates and the leading $b$ tagged jet ($m_{_b\tau\tau}$), in the $e\tau_h$ final state, for $m_a = 15$ GeV (top left), 40 GeV (top right) and 60 (bottom) GeV.
Figures 7.19: Correlation between the visible mass of the tau candidates ($m_\tau$), and the visible combine mass of the tau candidates and the leading b tagged jet ($m_{b\tau}$), in the $\mu\tau$ final state, for $m_a = 15$ GeV (top left), 40 GeV (top right) and 60 (bottom) GeV.
7.3 Selection Optimization

A set of cuts is applied to the four regions to reduce the various background and lead to the best expected limits. Looking at the background distribution of the ditau mass and the b jet (Figure 7.15) for all the final states, the dominant background is the \( \bar{t}t \) especially in \( e\mu \) final state. The second larger background contribution in \( e\tau_h \) and \( \mu\tau_h \) comes from the “\( \text{j}et \to \tau_h \)” which includes events with a jet misidentified as a \( \tau_h \) candidate, mainly from W+jets and QCD multijet. A more detail discussion about the provenance of the jet \( \to \tau_h \) background follows in chapter 8.

The selection criteria are two, the transverse mass between a lepton and the transverse missing momentum and the other based on the variable \( D_\zeta \). The first cut is applied in order to reject the events coming from the W+jets background. In all the final states the transverse mass of the electron or muon and the transverse missing momentum is defined as:

\[
m_T(\ell, \vec{p}_T^{\text{miss}}) \equiv \sqrt{2 p_T^{\ell} p_T^{\text{miss}} \left[ 1 - \cos(\Delta \phi) \right]} \tag{7.1}
\]

where \( p_T^{\ell} \) is the transverse momentum of the lepton \( \ell \), and \( \Delta \phi \) is the azimuthal angle between the lepton momentum and \( \vec{p}_T^{\text{miss}} \). Selecting events with low \( m_T \) strongly reduces the backgrounds from W + jets and \( \bar{t}t \) events, which are characterized by a larger \( \vec{p}_T^{\text{miss}} \).

The other selection criterion is based on the variable \( D_\zeta \) which is used as a discriminator to reject the dominant \( \bar{t}t \) events [3]. It is defined by the bisection \( \zeta \) axis in the transverse plane of the direction of the visible tau decay products (Figure 7.20). By projecting the visible decay products momenta and transverse missing energy into the \( \zeta \) axis two values are formed:

\[
p_\zeta = |\vec{p}_{T,1} + \vec{p}_{T,2} + \vec{p}_T^{\text{miss}}| \cdot \frac{\zeta}{|\zeta|} \tag{7.3}
\]

\[
p_\zeta^{\text{vis}} = |\vec{p}_{T,1} + \vec{p}_{T,2}| \cdot \frac{\zeta}{|\zeta|} \tag{7.4}
\]

where \( \vec{p}_{T,1} \) and \( \vec{p}_{T,2} \) denote the transverse momentum of the two reconstructed leptons. The variable \( D_\zeta \) is defined as:
The distributions of the \( m_T(\ell, \vec{p}_T^{\text{miss}}) \) and \( D_\zeta \) in the three final states, before the categorization based on \( m_{\text{vis}} \) cuts, are shown in Figures 7.21 – 7.23.

![Figure 7.20: Illustration of the definition of parameters used in the \( \zeta \) cut.](image)

The \( D_\zeta \) distributions in the three final states have different significance. More specifically, looking at the e\( \tau_h \) final state the \( t\bar{t} \) background contributes to the intermediate region where the signal also contributes. Therefore no cuts on \( D_\zeta \) are applied for this final state. On the other hand, the contribution of the \( t\bar{t} \) background in \( D_\zeta \) distributions in the case of the e\( \mu \) final state is concentrated at lower \( D_\zeta \) values, typically due to the large \( \vec{p}_T^{\text{miss}} \) (eq. 7.3), and not in the range of the signal distribution, allowing therefore a cut in \( D_\zeta \).

The contribution of the dominant backgrounds in the \( m_T(\ell, \vec{p}_T^{\text{miss}}) \) distribution in almost all cases is not in the range of the signal contribution. Thus, offering the potential to apply different cuts depending on the final state and the lepton \( \ell \), further reduces the background and increases the sensitivity of the analysis.

The selection is optimized to yield the best expected limits for a signal with \( m_a = 40 \) GeV in the low-mass and medium-mass SR, whereas the optimization in the high-mass SR is driven by the \( m_a = 15 \) GeV signal sample, which is the only one to have sensitivity in this region. The optimized working points in the low-mass and medium-mass SR are also optimal for higher \( m_a \) hypotheses, as it has been verified with the \( m_a = 60 \) GeV signal sample.
Figure 7.21: Distributions of $m_T(\mu, \vec{p}_T^{\text{miss}})$ (top left), $m_T(\tau_h, \vec{p}_T^{\text{miss}})$ (top right), and $D_c$ (bottom) in the $\mu\tau_h$ final state before the $m_{\text{vis}}$-based categorization. The “jet→$\tau_h$” contribution includes all events with a jet misidentified as a $\tau_h$ candidate. The “Other” contribution includes events from single top quark, diboson, and SM Higgs boson processes. The signal histogram corresponds to 10 times the SM production cross section for ggh, VBF, and Vh processes, and assumes $B(h\rightarrow a a \rightarrow 2\tau 2b)=100\%$. The ratio plots below shows a good agreement between observed and expected events within systematic uncertainties.
Figure 7.22: Distributions of $m_T(\mu, \vec{p}_T^{\text{miss}})$ (top left), $m_T(e, \vec{p}_T^{\text{miss}})$ (top right), and $D_\zeta$ (bottom) in the $\mu\tau_b$ final state before the $m_{\text{vis}}$-based categorization. The “Other” contribution includes events from single top quark, diboson, and SM Higgs boson processes. The signal histogram corresponds to 10 times the SM production cross section for ggh, VBF, and Vh processes, and assumes $B(h \rightarrow a \rightarrow 2\tau 2b)=100\%$. The ratio plots below shows a good agreement between observed and expected events within systematic uncertainties.
Figure 7.23: Distributions of $m_T(e, \vec{P}_T^{\text{miss}})$ (top left), $m_T(\tau_h, \vec{P}_T^{\text{miss}})$ (top right), and $D_\zeta$ (bottom) in the $\mu\tau_b$ final state before the $m_{\text{vis}}$-based categorization. The “jet→$\tau_b$” contribution includes all events with a jet misidentified as a $\tau_b$ candidate. The “Other” contribution includes events from single top quark, diboson, and SM Higgs boson processes. The signal histogram corresponds to 10 times the SM production cross section for ggh, VBF, and Vh processes, and assumes $B(h\rightarrow aa\rightarrow 2\tau 2b)=100\%$. The ratio plots below shows a good agreement between observed and expected events within systematic uncertainties.
This mass point is chosen because it is in the middle of the probed mass range and has the largest expected sensitivity (less backgrounds than at high mass, and larger acceptance than at low mass). It has been verified that optimizing the limits with a different mass point has a minimal effect on the selection thresholds.

For each category the optimization cuts applied are different, since they are dependent on the contribution of the dominant backgrounds. In the table 7.3 the precise cuts are given and their threshold for the three final states. Additionally, Figures 7.24 – 7.32 show the distributions of these variables in the four categories before the cuts in some of the final states. All the plots are prefit and partially blinded meaning that the observed events are blinded in bins where a signal with ma = 20, 40, 60 GeV is presented. Only the statistical uncertainties are shown. Also one can notice that the agreement between the observed and expected events are not very good in the first category, low-mass signal region, in all final states in contrast with the high-mass control region. This is due to the very small contamination of background in the low-mass signal region.

Substantially, all the above cuts and thresholds are being tested in order to justify the choice of the cuts. Figures 7.33 – 7.35 show the limits obtained for different thresholds on the transverse mass for ma = 15, 30, 60 GeV in different categories and final states.

For the final state μτh, the scenarios tested are mT (μ, MET) < 40 GeV, mT (μ, MET) < 50 GeV, and mT (μ, MET) < 60 GeV, together with mT (τh, MET) < 40 GeV, mT (τh, MET) < 50 GeV, and mT (τh, MET) < 60 GeV. The choice of the thresholds in the low-mass SR are driven by the limits for ma = 40 and 60 GeV, in the medium-mass SR by the limits for ma = 40 GeV, and in the high-mass SR by the limits for ma = 15 GeV. In the low-mass SR we choose mT (μ, MET) < 40 GeV and mT (τh, MET) < 60 GeV, in the medium-mass SR mT (μ, MET) < 40 GeV and mT (τh, MET) < 60 GeV, and in the high-mass SR mT (μ, MET) < 50 GeV and mT (τh, MET) < 60 GeV. Additionally requiring pζ < 0 GeV improves the results by 10% in the medium-mass SR.

For the final state eτh, the scenarios tested are mT (e, MET) < 40 GeV, mT (e, MET) < 50 GeV, and mT (e, MET) < 60 GeV, together with mT (τh, MET) < 40 GeV, mT (τh, MET) < 50 GeV, and mT (τh, MET) < 60 GeV. The choice of the thresholds in the low-mass SR are driven by the limits for ma = 40 and 60 GeV, in the medium-mass SR by the limits for ma = 40 GeV, and in the high-mass SR by the limits for ma = 15 GeV. In the low-mass SR we choose mT (e, MET) < 40 GeV and mT (τh, MET) < 60 GeV, in the medium-mass SR mT (e, MET) < 40 GeV and mT (τh, MET) < 60 GeV, and mT (τh, MET) < 60 GeV, and mT (τh, MET) < 60 GeV, and
in the high-mass SR $m_T(e, \text{MET}) < 60 \text{ GeV}$ and $m_T(\tau_h, \text{MET}) < 60 \text{ GeV}$. Applying a cut on $p_\zeta$ does not improve the results further.

For the final state $e\mu$, the scenarios tested are $m_T(\mu, \text{MET} ) < 40 \text{ GeV}$ and $m_T(e, \text{MET} ) < 40 \text{ GeV}$, $m_T(\mu, \text{MET} ) < 50 \text{ GeV}$ and $m_T(e, \text{MET} ) < 50 \text{ GeV}$, $m_T(\mu, \text{MET} ) < 60 \text{ GeV}$ and $m_T(e,\text{MET} ) < 60 \text{ GeV}$, without $p_\zeta$ cut or with $p_\zeta > -30 \text{ GeV}$. The choice of the thresholds in the low-mass SR are driven by the limits for $m_a = 40$ and $60 \text{ GeV}$, in the medium-mass SR by the limits for $m_a = 40 \text{ GeV}$, and in the high-mass SR by the limits for $m_a = 15 \text{ GeV}$. For all categories, we choose $m_T(\mu, \text{MET}) < 40 \text{ GeV}$ and $m_T(e, \text{MET}) < 40 \text{ GeV}$, as well as $p_\zeta > -30 \text{ GeV}$.

Table 7.3: Optimized selection and categorization in the various final states.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
<th>Category 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e\mu$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{e\mu,b}^{\text{vis}}$</td>
<td>$&lt; 65 \text{ GeV}$</td>
<td>$\in [65, 80] \text{ GeV}$</td>
<td>$\in [80, 95] \text{ GeV}$</td>
<td>$&gt; 95 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T(e, \vec{p}_T^{\text{miss}})$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T(\mu, \vec{p}_T^{\text{miss}})$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
</tr>
<tr>
<td>$D_\zeta$</td>
<td>$&gt; -30 \text{ GeV}$</td>
<td>$&gt; -30 \text{ GeV}$</td>
<td>$&gt; -30 \text{ GeV}$</td>
<td>$&gt; -30 \text{ GeV}$</td>
</tr>
<tr>
<td>$e\tau_h$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{e\tau_h,b}^{\text{vis}}$</td>
<td>$&lt; 80 \text{ GeV}$</td>
<td>$\in [80, 100] \text{ GeV}$</td>
<td>$\in [100, 120] \text{ GeV}$</td>
<td>$&gt; 120 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T(e, \vec{p}_T^{\text{miss}})$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 50 \text{ GeV}$</td>
<td>$&lt; 50 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T(\tau_h, \vec{p}_T^{\text{miss}})$</td>
<td>$&lt; 60 \text{ GeV}$</td>
<td>$&lt; 60 \text{ GeV}$</td>
<td>$&lt; 60 \text{ GeV}$</td>
<td>$&lt; 60 \text{ GeV}$</td>
</tr>
<tr>
<td>$D_\zeta$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\mu\tau_h$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_{\mu\tau_h,b}^{\text{vis}}$</td>
<td>$&lt; 75 \text{ GeV}$</td>
<td>$\in [75, 95] \text{ GeV}$</td>
<td>$\in [95, 115] \text{ GeV}$</td>
<td>$&gt; 115 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T(\mu, \vec{p}_T^{\text{miss}})$</td>
<td>$&lt; 40 \text{ GeV}$</td>
<td>$&lt; 50 \text{ GeV}$</td>
<td>$&lt; 50 \text{ GeV}$</td>
<td>$&lt; 40 \text{ GeV}$</td>
</tr>
<tr>
<td>$m_T(\tau_h, \vec{p}_T^{\text{miss}})$</td>
<td>$&lt; 60 \text{ GeV}$</td>
<td>$&lt; 60 \text{ GeV}$</td>
<td>$&lt; 60 \text{ GeV}$</td>
<td>$&lt; 60 \text{ GeV}$</td>
</tr>
<tr>
<td>$D_\zeta$</td>
<td>$-$</td>
<td>$&lt; 0 \text{ GeV}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>
Figure 7.24: $P_\zeta$ in the different categories of the $e\tau_\nu$ final state. The cuts on the transverse mass between the leptons and the MET have been applied. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.25: Transverse mass between the electron and the MET in the different categories of the $e\tau_b$ final state. The cuts on the transverse mass between the leptons and the MET have been applied. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.26: Transverse mass between the $\tau_h$ and the MET in the different categories of the $e\tau_h$ final state. The cuts on the transverse mass between the leptons and the MET have been applied. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.27: Transverse mass between the muon and the MET, in the different categories of the $\mu\tau_h$ final state. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.28: Transverse mass between the $\tau_h$ and the MET, in the different categories of the $\mu\tau_h$ final state. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.29: $P_T\tau_h$ in the different categories of the $\mu\tau_h$ final state. Loose cuts on the transverse mass between the $\tau_h$ and the MET, and the transverse mass between the muon and the MET have been applied. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.30: Transverse mass between the muon and the MET in the different categories of the $e\mu$ final state. The cuts on the transverse mass between the leptons and the MET have been applied. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.31: Transverse mass between the electron and the MET in the different categories of the $e\mu$ final state. The cuts on the transverse mass between the leptons and the MET have been applied. The plots areprefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.32: $p_T$ in the different categories of the $e\mu$ final state. The cuts on the transverse mass between the leptons and the MET have been applied. The plots are prefit and partially blinded, and only the statistical uncertainty is shown. The ratio plots below show a good agreement between observed and expected in the fourth category (right bottom plot). In the other categories and where the plots are unblinded, the agreement is good within systematic uncertainties.
Figure 7.33: Expected limits obtained in the $e\mu$ final state for the three signal regions, for the mass hypotheses 15, 40, and 60 GeV.
Figure 7.34: Expected limits obtained in the $e\tau$ final state for the three signal regions, for the mass hypotheses 15, 40, and 60 GeV.
Figure 7.35: Expected limits obtained in the $\mu\tau_b$ final state for the three signal regions, for the mass hypotheses 15, 40, and 60 GeV.
References


CHAPTER 8

Background Estimation and Systematic Uncertainties

The main part of the background originating from the Standard Model processes in this analysis is estimated from MC simulations. The QCD multijet background and the background containing a jet misidentified as a $\tau_h$ candidate are fully or partially estimated from data. The methods used for the estimation of these two sources are described in detail in subchapters 8.1.2 and 8.1.3 respectively. The second part of this chapter includes all the uncertainties related to the background estimation or the physical objects identification methods.

8.1 Background Estimation

8.1.1 Background Estimated from MC Simulations

The Drell-Yan background is estimated from MC simulations. As detailed in Chapter 6, the simulation is reweighted event-by-event as a function of the generated mass and $p_T$ of the parent boson, in order to cover data/MC disagreements observed in a $Z \rightarrow \mu\mu$ control region. The simulations with different generated jet multiplicities are stitched together. Simulations cover generated $Z$ boson masses between 10 and 50 GeV, and above 50 GeV.

The Drell-Yan background is separated between $Z \rightarrow \tau\tau$ decay, where the reconstructed $\tau_h$ is matched at the generator level to a real $\tau_h$ and the $Z \rightarrow ee/\mu\mu$ decays, in which at least one of the two leptons are misidentified as a hadronic tau candidate. The $Z + b$ jets cross section is not well known theoretically, but it can be controlled in the high mass regions fitted simultaneously as is done in the low-mass signal-enriched regions.
The dominant background $t\bar{t} + \text{jets}$, especially for the $e\mu$ final state, as well as the single top and diboson backgrounds are estimated from MC samples or in combination with MC simulation, respectively. It is separated between events where the reconstructed $\tau_h$ is matched at the generated level with a real $\tau_h$, to a light lepton, or to a jet.

The $W + \text{jets}$ background is estimated from MC simulation only in the $e\mu$ final state and corresponds to different generated jet multiplicities (0, 1, 2, 3, 4 jets). Events from $W + \text{jets}$ decays commonly have a genuine lepton and a jet misidentified as a second lepton. In the $e\tau_h$ and $\mu\tau_h$ final states, the $W + \text{jets}$ background is estimated from data as described in the following section.

8.1.2 QCD Multijet

The QCD multijet background is fully estimated from data. It is extracted by subtracting all MC simulations in the same-sign region from data in the same-sign region. The yield is corrected for differences between the same-sign and opposite-sign regions. The extrapolation factor is 1.06 in the $e\tau_h$ and $\mu\tau_h$ channels, and 2.27 in the $e\mu$ final state.

To improve the statistical precision of the templates (smoother distribution), the isolation conditions on the leptons are relaxed. The isolation conditions to extract the QCD multijet background distributions are:

- $\mu\tau_h$: medium MVA $\tau_h$ isolation, relative muon isolation < 0.3.
- $e\tau_h$: medium MVA $\tau_h$ isolation, relative electron isolation < 0.3.
- $e\mu$: relative electron isolation < 0.3, relative muon isolation < 0.3.

This background estimation method in the $e\tau_h$ and $\mu\tau_h$ channels is used only as a cross-check of the main method, which is described in the next section.

8.1.3 Jet Misidentified as a $\tau_h$ Candidate, Jet → $\tau_h$

The estimation of the QCD multijets and $W + \text{jets}$ backgrounds, as described above, leads to sizable statistical and systematic uncertainties. In the $e\tau_h$ and $\mu\tau_h$ channels, an alternative method is used by default to estimate processes where a jet is misidentified as a $\tau_h$ from data. The $\tau_h$ can come from misidentified jets (jet → $\tau_h$) in several processes include $W + \text{jets}$ and QCD multijet, but also a fraction of $t\bar{t}$, diboson, single top, and $Z + \text{jets}$ events where the reconstructed $\tau_h$ is actually a generated jet.
The method relies on the estimation of the jet → $\tau_h$ misidentification rate $f_\tau$ from an independent data sample $Z \rightarrow \mu\mu + \text{jet}$:

$$f_\tau = \frac{N_{\text{events}}(Z \rightarrow \mu\mu + \text{jet} = (\tau_{ID+\text{tight}}))}{N_{\text{events}}(Z \rightarrow \mu\mu + \text{jet} = (\tau_{ID+\text{loose}}))} \quad (8.1)$$

The denominator corresponds to jets passing the very loose $\tau_h$ MVA identification (anti-isolation region), whereas jets in the numerator pass the tight working point of the MVA identification (isolated region). The misidentification rates are separated by $\tau_h$ reconstructed decay mode (1-prong, 1-prong+$\pi^0$, or 3-prong), and are parametrized with a Landau function of the $\tau_h$ transverse momentum, $p_T$. The fake rates are shown in Figure 8.1, together with their uncertainties. The uncertainties are based on the uncorrelated uncertainties returned by the fit. The fake rate is smaller in the 3-prong decay mode.

To measure the fake rates, events with two muons and one $\tau_h$ candidate are selected. The two muons should be isolated with relative isolation, $I_{rel} < 0.15$, identified passing the medium ID working point, have opposite sign, and have an invariant mass between 76 and 106 GeV. They are selected with a double muon trigger with the following offline requirement: the leading muon should have $p_T > 20$ GeV and the subleading muon $p_T > 10$ GeV. The $\tau_h$ candidate should pass the very loose MVA ID, the very loose discriminator against electrons, and the tight discriminator against muon. The contamination from diboson background ZZ and WZ in the $Z \rightarrow \mu\mu + \text{jet}$ sample is estimated from simulations and subtracted from data. The impact of the diboson background is largest at high $\tau_h$ $p_T$.

![Figure 8.1: Jet → $\tau_h$ fake rate for jets passing the very loose MVA isolation (denominator) to pass the tight MVA isolation (numerator), in the 1 prong (left), 1 prong + $\pi^0$ (center), and 3 prong (right) decay modes. The fake rates are fitted with Landau functions.](image-url)
The fake rates have been re-measured in $Z \rightarrow \mu\mu$ events, where at least one jet is tagged as coming from a $b$ quarks. As shown in Figure 8.2, the fake rates measured in a $Z \rightarrow \mu\mu + \text{jet}$ (signal region), with at least 1 $b$-tagged jet, are well compatible with the fake rates measured previously, but with larger uncertainties. Therefore, the default measurement of the fake rates, without requiring a $b$ tagged jet in the region where they are measured are kept.

Data events that pass all selection criteria except from $\tau_\text{h}$ that does not pass the tight isolation but passed the very loose working point, are reweighted by $f_\tau/(1 - f_\tau)$, where $f_\tau$ is the misidentification rate, dependent on the $p_T$ of the $\tau_\text{h}$ and on its reconstructed decay mode. The re-weighted events represents the number of jets that faking $\tau_\text{h}$ in the isolated signal region. In more detail, the number of events in data that includes jets and enter the signal region with tight isolation over the number of events that enter the anti-isolated region with loose but not tight isolation is given by the following equation:

$$\frac{N_{\text{events}}(\tau_{\text{ID+tight\_isolated}})}{N_{\text{events}}(\tau_{\text{ID+loose\_isolated \& not tight\_isolated}})} = \frac{f_\tau N_{\text{events}}(\tau_{\text{ID+loose\_isolated}})}{(1 - f_\tau) N_{\text{events}}(\tau_{\text{ID+loose\_isolated}})}$$

(8.2)

So, the background from jets that faking $\tau_\text{h}$ in the signal region is given by:

$$N_{\text{events}}(\tau_{\text{ID+tight\_isolated}}) = \frac{f_\tau}{(1 - f_\tau)} N_{\text{events}}(\tau_{\text{ID+loose\_isolated \& not tight\_isolated}})$$

(8.3)
In order for the background estimate to include only jets faking taus (events with leptons faking taus are already estimated with MC simulations), the contribution of events with leptons is estimated from MC and subtracted from data. In practice, reweighted MC events that pass the same selection and whose reconstructed $\tau_h$ candidate does not correspond to a jet at generated level, are subtracted from the fake background estimate. The fake background estimate includes all events with a jet faking a $\tau_h$, among events from QCD multijet, W+jets, Z+jets, $t\bar{t}$, and other processes.

The reducible background estimation is verified in regions identical to the signal regions in the $\mu\tau$ final state, except that the tau candidates have the same charge. The yields of data and background predictions are indicated in Table 8.1. The background prediction is in fair agreement with the data within two statistical deviations for the data. In the signal region a yield uncertainty of 20% is considered for the reducible background to account for possible differences in the background composition.

The statistics in the same-sign region are increased by removing all requirements on $m_T$ and $p_\tau$. As shown in Figures 8.3 – 8.4 for the $e\tau$ and $\mu\tau$ final states, the agreement between data and predicted backgrounds is generally good, within the 20% systematic associated with the reducible background estimation. The 20% systematic derived from the ratio between observed and expected events in same sign control region (see bottom right ratio plot in figures 8.3 and 8.4).

Table 8.1: Background predictions and observed data in categories similar to those of the signal region except that the charged requirement is inverted for the $\mu\tau_h$ final state.

<table>
<thead>
<tr>
<th></th>
<th>Low-mass SR</th>
<th>Medium-mass SR</th>
<th>High-mass SR</th>
<th>High-mass CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$j\rightarrow\tau_h$ background</td>
<td>12.63</td>
<td>49.23</td>
<td>89.45</td>
<td>1125.00</td>
</tr>
<tr>
<td>Other backgrounds</td>
<td>0.35</td>
<td>0.43</td>
<td>2.79</td>
<td>30.91</td>
</tr>
<tr>
<td>Data</td>
<td>19</td>
<td>62</td>
<td>110</td>
<td>1079</td>
</tr>
</tbody>
</table>

The estimations obtained with this method, and with the background methods described above, (including QCD and W+jets) are in good agreement, with an increased statistical precision using the fully data-driven method. Visible ditau mass distributions after the optimized selection are shown in Figures 8.5 and 8.6 for several categories with both background estimation methods. The prefit yields of data and background predictions are indicated in Table 8.2.
Figure 8.3: Prefit plots, in the $e\tau_h$ channel, of the ditau mass distributions in the same sign region without requirements on $m_T$ nor $p\not{E}_T$. Only the statistical uncertainty is shown. The agreement between observed and expected events is generally good. In the control region the systematic uncertainties related to the reducible background estimation are up to 20%.
Figure 8.4: Prefit plots, in the $\mu\tau_b$ channel, of the ditau mass distributions in the same sign region without requirements on $m_{\tau}$ nor $p_\tau$. Only the statistical uncertainty is shown. The agreement between observed and expected events is generally good. In the control region the systematic uncertainties related to the reducible background estimation are up to 20%.
8.2 Systematic Uncertainties

The systematic uncertainties in this analysis are represented by nuisance parameters that vary in the fit according to their probability density functions. Sources for systematic uncertainties are uncertainties in the identification and reconstruction of the physical objects, due to energy scale fluctuations and the mismodelling of signal and background with event generators. Additional uncertainties are derived from the integrated luminosity and the production cross section.

**Figure 8.5:** Prefit plots, in the $\mu\tau_h$ channel, of the ditau mass distributions in two categories, for both background estimations methods (left: fake rate method, right: $W+\text{jets}$ from MC and QCD multijet from same-sign region). The plots are partially blinded according to the expected significance in each bin. The signal corresponds to $\text{B}(h \to aa \to \tau\tau bb) = 10\%$. Only the statistical uncertainty is shown. The agreement between observed and expected is very good within systematic uncertainties.
Table 8.2: Prefit background predictions and observed data in the various categories.

<table>
<thead>
<tr>
<th>Process</th>
<th>Low-mass SR</th>
<th>Medium-mass SR</th>
<th>High-mass SR</th>
<th>High-mass CR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>eμ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QCD multijet</td>
<td>4.86 ± 1.04</td>
<td>17.40 ± 3.65</td>
<td>53.67 ± 9.99</td>
<td>321.26 ± 61.53</td>
</tr>
<tr>
<td>W + jets</td>
<td>0.96 ± 0.18</td>
<td>0.92 ± 0.88</td>
<td>0.53 ± 0.10</td>
<td>13.01 ± 2.45</td>
</tr>
<tr>
<td>t̅ + t</td>
<td>24.33 ± 3.48</td>
<td>71.16 ± 8.09</td>
<td>170.79 ± 16.64</td>
<td>4438.66 ± 337.34</td>
</tr>
<tr>
<td>Z → ττ</td>
<td>9.77 ± 3.93</td>
<td>57.39 ± 10.67</td>
<td>83.02 ± 16.31</td>
<td>548.68 ± 55.95</td>
</tr>
<tr>
<td>Z → ℓℓ</td>
<td>0.33 ± 0.03</td>
<td>2.20 ± 0.18</td>
<td>2.58 ± 2.98</td>
<td>17.81 ± 1.49</td>
</tr>
<tr>
<td>Single top</td>
<td>1.77 ± 0.35</td>
<td>5.19 ± 0.78</td>
<td>7.10 ± 1.10</td>
<td>302.93 ± 42.51</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.34 ± 0.04</td>
<td>0.85 ± 0.19</td>
<td>2.50 ± 0.32</td>
<td>21.21 ± 2.09</td>
</tr>
<tr>
<td>SM H</td>
<td>0.22 ± 0.04</td>
<td>0.59 ± 0.07</td>
<td>0.70 ± 0.06</td>
<td>11.06 ± 0.83</td>
</tr>
<tr>
<td><strong>Total expected</strong></td>
<td><strong>42.60 ± 5.58</strong></td>
<td><strong>155.71 ± 15.08</strong></td>
<td><strong>320.89 ± 27.89</strong></td>
<td><strong>5674.61 ± 385.11</strong></td>
</tr>
<tr>
<td><strong>Observed data</strong></td>
<td>33</td>
<td>149</td>
<td>298</td>
<td>5565</td>
</tr>
<tr>
<td></td>
<td>etb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jet → τb bkg.</td>
<td>14.55 ± 3.38</td>
<td>63.78 ± 12.13</td>
<td>130.72 ± 25.02</td>
<td>1001.03 ± 185.22</td>
</tr>
<tr>
<td>t̅ + t</td>
<td>9.05 ± 1.57</td>
<td>37.00 ± 5.17</td>
<td>102.10 ± 10.65</td>
<td>1187.25 ± 99.76</td>
</tr>
<tr>
<td>Z → ττ</td>
<td>10.54 ± 3.63</td>
<td>86.07 ± 15.83</td>
<td>137.77 ± 21.63</td>
<td>529.95 ± 59.51</td>
</tr>
<tr>
<td>Z → ℓℓ</td>
<td>0.00 ± 0.00</td>
<td>12.96 ± 8.21</td>
<td>93.05 ± 14.45</td>
<td>560.57 ± 52.48</td>
</tr>
<tr>
<td>Single top</td>
<td>0.51 ± 0.18</td>
<td>3.89 ± 0.70</td>
<td>13.48 ± 2.09</td>
<td>144.34 ± 20.87</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.27 ± 0.05</td>
<td>1.13 ± 0.16</td>
<td>3.22 ± 0.33</td>
<td>27.21 ± 2.74</td>
</tr>
<tr>
<td>SM H</td>
<td>0.07 ± 0.01</td>
<td>0.50 ± 0.08</td>
<td>1.00 ± 0.12</td>
<td>9.86 ± 0.86</td>
</tr>
<tr>
<td><strong>Total expected</strong></td>
<td><strong>34.98 ± 5.30</strong></td>
<td><strong>205.33 ± 23.37</strong></td>
<td><strong>481.34 ± 41.22</strong></td>
<td><strong>3460.21 ± 253.26</strong></td>
</tr>
<tr>
<td><strong>Observed data</strong></td>
<td>30</td>
<td>190</td>
<td>452</td>
<td>3411</td>
</tr>
<tr>
<td></td>
<td>mτb</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jet → τb bkg.</td>
<td>42.58 ± 8.29</td>
<td>160.66 ± 30.21</td>
<td>384.53 ± 71.37</td>
<td>2138.51 ± 392.77</td>
</tr>
<tr>
<td>t̅ + t</td>
<td>16.69 ± 1.90</td>
<td>56.66 ± 6.49</td>
<td>264.23 ± 25.53</td>
<td>2446.68 ± 202.58</td>
</tr>
<tr>
<td>Z → ττ</td>
<td>33.09 ± 14.02</td>
<td>138.49 ± 23.71</td>
<td>387.01 ± 75.22</td>
<td>1313.00 ± 142.31</td>
</tr>
<tr>
<td>Z → ℓℓ</td>
<td>0.40 ± 0.06</td>
<td>0.44 ± 0.05</td>
<td>24.07 ± 4.52</td>
<td>171.79 ± 22.99</td>
</tr>
<tr>
<td>Single top</td>
<td>0.47 ± 0.17</td>
<td>4.05 ± 0.74</td>
<td>20.63 ± 3.17</td>
<td>211.52 ± 30.40</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.30 ± 0.04</td>
<td>1.94 ± 0.23</td>
<td>4.35 ± 0.54</td>
<td>30.03 ± 2.96</td>
</tr>
<tr>
<td>SM H</td>
<td>0.21 ± 0.03</td>
<td>0.98 ± 0.10</td>
<td>1.76 ± 0.21</td>
<td>21.21 ± 1.60</td>
</tr>
<tr>
<td><strong>Total expected</strong></td>
<td><strong>93.73 ± 16.60</strong></td>
<td><strong>363.21 ± 39.98</strong></td>
<td><strong>1086.58 ± 113.12</strong></td>
<td><strong>6332.74 ± 506.93</strong></td>
</tr>
<tr>
<td><strong>Observed data</strong></td>
<td>99</td>
<td>372</td>
<td>1079</td>
<td>6289</td>
</tr>
</tbody>
</table>
8.2.1 Uncertainties Related to Physics Objects

The uncertainty related to the identification of muons derived from data using tag and probe methods [1] amounts to 2%. It is applied to all processes fully estimated from MC samples as a function of lepton \( p_T \) and \( \eta \). Similarly, 2% uncertainty is assigned to the yield of MC processes with real electrons in the e\( \mu \) and e\( \tau \) final states, to account for the uncertainty in the electron isolation and identification efficiency.

The uncertainty in the data/MC scale factor related to the identification of hadronically decaying taus is 5% and is also derived using tag-and-probe method (recommended by Tau Physics Object Group (TAU POG)). This uncertainty is applied to events with real hadronic taus, determined from generator-level matching. In particular the \( t \bar{t} \) process is split into a component with real hadronic taus, and a component with other generated objects, so that the uncertainty is applied only to the first component (other uncertainties, described later in the text, are assigned to the other component).

The nominal energy scale is corrected based on the decay mode: -1.8% for 1 prong taus, +1.0% for 1 prong + \( \pi^0 \), and +0.4% for 3 prong taus. These corrections come from a measurement within the TAU POG, made in \( Z \rightarrow \mu\tau_b \) events and based on a fit to the visible mass distributions. The uncertainty on the correction is 1.2% per decay mode. The energy scale uncertainty is applied by shifting the four-vectors up and down by 1.2% and recomputing composite variables in the analysis. The change in the shape of the final fitted variables is then included as a shape systematic uncertainty. For one-prong taus the mass is constant and set equal to the pion mass for all shifts.

The uncertainty in the energy scale of muons is 0.2%. It is considered as a shape systematic in the \( \mu\tau_b \) and e\( \mu \) final states. The uncertainty in the energy scale of electrons is also a shape uncertainty, and the shifts of the energy scale, event-per-event, are taken from the regression prescribed by the EGamma Physics Object Group (EGamma POG).

The probability of a \( \mu \) to fake a \( \tau_b \) is measured by the TAU POG in a dedicated tag and probe study using \( Z \rightarrow \mu\mu \) events which are reconstructed as \( Z \rightarrow \mu\tau_b \). For different working points of the anti-muon discriminator (Loose, Medium, Tight), events are split into pass and fail categories. Additionally, they are fitted simultaneously between the pass and fail regions to extract a data to MC scale factor and uncertainty for events with a generator matched \( \mu \) which fakes a \( \tau_b \) in bins of \( \eta \).
In practice, the scale factors are applied as a function of the pseudorapidity of the muon that fakes the hadronic tau. A flat 5% uncertainty, corresponding to the loose working point of the discriminator, is considered on processes where a muon fakes a hadronic tau (mostly $Z \rightarrow \mu\mu$ in the $\mu\tau_h$ final state) [2]. It is uncorrelated with the tau decay modes. The corresponding uncertainty for electrons is also 5%, for the very loose working point of the corresponding discriminator [2].

**Figure 8.6:** Prefit plots, in the $e\tau_h$ channel, of the ditau mass distributions in two categories, for both background estimations methods (left: fake rate method, right: W+jets from MC and QCD multijet from same-sign region). The plots are partially blinded according to the expected significance in each bin. The signal corresponds to $B(h \rightarrow aa \rightarrow \tau\tau bb) = 10\%$. Only the statistical uncertainty is shown.
The uncertainties related to the b-tagging of b jets and light flavour jets are considered as shape systematics by varying the scale factors within their uncertainties, using the scale factors and uncertainties provided by the b Tag & Vertexing Physics Object Group (BTV POG). These are shape uncertainties because the b-tagging scale factors depend on the $p_T$ of the b jet, and the category definitions depend on the b jet four-momentum. They amount to 1.5% for the jet originating from a b quark, 5% from a c quark and 10% from a light-flavor parton [3].

The uncertainty on the transverse missing energy, MET, related to unclustered energy deposits is considered as a shape uncertainty. The values are accessed from miniAOD (a high-level tier) by reiterating the MET sequence. The uncertainty in the jet energy scale [4] is also taken into account as a shape uncertainty. It affects the choice of the b jet and the categorization as a function of the invariant mass of the leptons and the b jet. It is not split by sources because the b jets are in the same region in $\eta$ of the detector.

8.2.2 Uncertainties Related to Background Estimation

The uncertainty on the single top background, estimated purely from MC samples, amounts to 13%, according to a recent CMS measurement [5]. The 6% uncertainty is also considered for the various diboson processes [6].

The shape uncertainty of the Drell-Yan background is computed by applying the factor 0.9 or 1.1 times the correction on the dilepton $p_T$. The uncertainty in this correction is equal to 10% of the size of the correction itself. It is fully correlated between channels and categories (low-mass signal region, medium signal region, high-mass signal region and high-mass control region). The normalization of the Drell-Yan background has an uncertainty of 7% (taken from a CMS measurement of the differential $Z + \geq 1$ b cross section [7]); it can be constrained in the high-statistics signal-free region with high mass between the leptons and the leading b jet. Additionally, a shape uncertainty equal to the correction on $m_{b\tau\tau}$ is applied, and fully correlated between final states and categories. Its effect is largest in the low $m_{b\tau\tau}$ categories, because the $m_{b\tau\tau}$ distributions are steeply rising there and is considered as a shape uncertainty.

In the $e\mu$ channel, the uncertainty in the QCD multijet background is 20%. This arises from the extrapolation factor of the same sign to opposite sign scale factor [8]. The uncertainty in the W+jets
background is also 20%, due to uncertainties stemming from the mismodeling in simulation of the misidentification rate of jets as electrons or muons.

The uncertainty in the jet → τ_h fake background in the eτ_h and μτ_h channels amounts to 20%. It’s derived from the misidentification rates in Z + jets events, in W + jets and QCD multijet events which are dominating the contribution of the reducible background in the same-sign region. This uncertainty is reduced to 7% by fitting with the maximum likelihood the contribution of the invariant mass m_{bτ}, in the last category (High-massCR). Additionally, several shape uncertainties due to the parametrization of the fake rate as a function of hadronic tau p_T and the different tau decay modes (figures 8.1, 8.2), are also considered and affect the jet → τ_h fake background.

The t̅t background is reweighted according to the generated transverse momentum of the top quarks, as explained in section 6.2.9. A shape uncertainty corresponding to the size of the correction is applied to the background, and considered as fully correlated between the different final states. The normalization uncertainty of the t̅t background is 4%, as determined by a recent CMS measurement at 13 TeV [9].

Uncertainties in the Higgs boson cross sections and decay modes are taken into account for the SM Higgs background [10], as listed in Tables 8.3 – 8.4, along with uncertainties previously mentioned and others.

8.2.3 Other Uncertainties

The uncertainty in the collected integrated luminosity in 2016 amounts to 2.5%. It is applied to all processes whose normalization is taken from MC simulations, and it is fully correlated between categories and final states. Trigger scale factors uncertainties in the yields of simulated processes amount to 1 – 2 %, depending in the final state.

The procedure used to extrapolate the VBF signal templates to non-simulated masses has an uncertainty of 10%, determined by checking the closure between the extrapolation and a generated signal sample for existing masses (see Appendix A2). It is uncorrelated between final states and categories because the origin of the uncertainty is related to the statistics available in simulation for each final state. Similarly, the uncertainty in the VH produced signal events is 10%. 
Finally, bin-by-bin uncertainties are considered for all processes (signals and backgrounds), for all bins for which the statistical uncertainty is larger than 5%.

Table 8.3: List of systematic uncertainties related to the physical objects and other uncertainties.

<table>
<thead>
<tr>
<th>Uncertainties</th>
<th>Affected processes</th>
<th>$e_{\tau_b}$</th>
<th>$e_{\mu}$</th>
<th>$\mu_{\tau_b}$</th>
<th>Method used/POG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Muon ID</td>
<td>All MC</td>
<td>-</td>
<td>1%</td>
<td>1%</td>
<td>Tag and probe</td>
</tr>
<tr>
<td>Electron ID</td>
<td>All MC</td>
<td>2%</td>
<td>2%</td>
<td>-</td>
<td>Tag and probe</td>
</tr>
<tr>
<td>Tau ID</td>
<td>All MC</td>
<td>5%</td>
<td>-</td>
<td>5%</td>
<td>Tag and probe / TAU POG</td>
</tr>
<tr>
<td>$\mu$ misidentified $\tau_b$</td>
<td>$Z \rightarrow \mu\mu$</td>
<td>-</td>
<td>-</td>
<td>5%</td>
<td>Tag and probe / TAU POG</td>
</tr>
<tr>
<td>$e$ misidentified $\tau_b$</td>
<td>$Z \rightarrow ee$</td>
<td>5%</td>
<td>-</td>
<td>-</td>
<td>Tag and probe / TAU POG</td>
</tr>
<tr>
<td>$\tau_b$ energy scale</td>
<td>All MC</td>
<td>1.2% for each decay mode</td>
<td></td>
<td>Fit on $m_{vis}$ distribution of $Z \rightarrow \mu\tau_b$ events / TAU POG</td>
<td></td>
</tr>
<tr>
<td>$e$ energy scale</td>
<td>All MC</td>
<td>Shape uncertainties</td>
<td>-</td>
<td>Egamma POG regression</td>
<td></td>
</tr>
<tr>
<td>$\mu$ energy scale</td>
<td>All MC</td>
<td>-</td>
<td>-</td>
<td>0.2%</td>
<td>MUON POG</td>
</tr>
<tr>
<td>$b$ quark id. $\rightarrow b$ jet</td>
<td>All MC</td>
<td>~1.5 % (shape)</td>
<td></td>
<td>BTV POG</td>
<td></td>
</tr>
<tr>
<td>$c$ quark misid. $\rightarrow b$ jet</td>
<td>All MC</td>
<td>~0.5 % (shape)</td>
<td></td>
<td>BTV POG</td>
<td></td>
</tr>
<tr>
<td>Soft quark misid. $\rightarrow b$ jet</td>
<td>All MC</td>
<td>~10% (shape)</td>
<td></td>
<td>BTV POG</td>
<td></td>
</tr>
<tr>
<td>MET (unclustered energy deposits)</td>
<td>All MC</td>
<td>shape</td>
<td></td>
<td>From MiniAOD by rerunning MET sequence</td>
<td></td>
</tr>
<tr>
<td>Integrated Luminosity</td>
<td>All MC</td>
<td>2.5%</td>
<td></td>
<td>CMS Luminosity measurements for 2016 data-taking period</td>
<td></td>
</tr>
<tr>
<td>Trigger scale factors</td>
<td>All MC</td>
<td>1 – 2%</td>
<td></td>
<td>Tag-and-probe</td>
<td></td>
</tr>
<tr>
<td>VBF / VH signal</td>
<td>VBF</td>
<td>10%</td>
<td></td>
<td>Rescaling for all bins with stat. Uncertainty&gt; 5%</td>
<td></td>
</tr>
<tr>
<td>Bin-by-bin</td>
<td>ALL</td>
<td>Shape 5%</td>
<td></td>
<td>Applied for all bins with stat. Uncertainties &gt; 5%</td>
<td></td>
</tr>
</tbody>
</table>
Table 8.4: List of systematic uncertainties related to background estimation.

<table>
<thead>
<tr>
<th>Uncertainties</th>
<th>Affected processes</th>
<th>$\tau_{\ell}$</th>
<th>$\mu_{\ell}$</th>
<th>$\mu_{\ell_{h}}$</th>
<th>Derived from</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single top normal.</strong></td>
<td>Single top</td>
<td>-</td>
<td>13%</td>
<td>-</td>
<td>CMS measurements at 13 TeV</td>
</tr>
<tr>
<td><strong>Diboson normal.</strong></td>
<td>Diboson</td>
<td>-</td>
<td>6%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$t\bar{t}$ normal.</td>
<td>$t\bar{t}$</td>
<td>-</td>
<td>4%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>QCD multijet</strong></td>
<td>QCD mult.</td>
<td>-</td>
<td>20%</td>
<td>-</td>
<td>Extrapolation factor SS/OS</td>
</tr>
<tr>
<td><strong>$W + \text{jets normal.}$</strong></td>
<td>$W + \text{jets}$</td>
<td>-</td>
<td>20%</td>
<td>-</td>
<td>Misidentified jets $\rightarrow e/\mu$</td>
</tr>
<tr>
<td>Jet $\rightarrow \tau_{h}$ (yield)</td>
<td>Jet $\rightarrow \tau_{h}$</td>
<td>20%</td>
<td>-</td>
<td>20%</td>
<td>Misidentified rates in $Z +\text{jets, W+jets, QCD multijets events}$</td>
</tr>
<tr>
<td>Jet $\rightarrow \tau_{h}$ (shape)</td>
<td>Jet $\rightarrow \tau_{h}$</td>
<td>shape</td>
<td>shape</td>
<td>-</td>
<td>Fake rate fit functions</td>
</tr>
<tr>
<td><strong>$Z + \text{jets normal.}$</strong></td>
<td>$Z + \text{jets}$</td>
<td>-</td>
<td>7%</td>
<td>-</td>
<td>CMS measurements at 13 TeV</td>
</tr>
<tr>
<td>$Z + \text{jets}$</td>
<td>$Z + \text{jets}$</td>
<td>5% (uncorrelated between categories)</td>
<td>-</td>
<td>-</td>
<td>Mismodeling of $m_{\text{inv}}$ in LO MC</td>
</tr>
<tr>
<td>BR ($H \rightarrow \tau\tau$)</td>
<td>$H \rightarrow \tau\tau$</td>
<td>-</td>
<td>10%</td>
<td>-</td>
<td>Dilepton $p_T$ reweighting</td>
</tr>
<tr>
<td>BR ($H \rightarrow WW$)</td>
<td>$H \rightarrow WW$</td>
<td>-</td>
<td>1%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma (ggH)$</td>
<td>ggH</td>
<td>-</td>
<td>4%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma (VBF)$</td>
<td>VBF</td>
<td>-</td>
<td>2%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma (WH)$</td>
<td>WH</td>
<td>-</td>
<td>2%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma (ZH)$</td>
<td>ZH</td>
<td>-</td>
<td>4%</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>$\sigma (ttH)$</td>
<td>ttH</td>
<td>-</td>
<td>8%</td>
<td>-</td>
<td>LHC Higgs Cross Section Working Group</td>
</tr>
</tbody>
</table>
References


CHAPTER 9

General Statistic Tools Used

9.1 Statistic

The observation of the recently discovered SM Higgs boson relied on statistical tools, a common practice followed in all physics research. For the observation of a new physical object or for setting the limits on the production cross section, a likelihood function is required. This function includes all the information concerning the theoretical model of interest as well as the exact experimental measurements accompanied by their systematic uncertainties.

9.1.1 Likelihood

In physics searches, the analysis of event descriptions in an experiment, is one of the main parameters of interest. By comparing the expectation distributions from theory with the observed ones, information about the parameter of interest can be extracted. The method used is call “fit” and follows three consecutive steps. Firstly, a theoretical function \( p(x/\theta) \) is set, to define the hypothesis, with \( x \) being the variable of interest or a set of variables and \( \theta \) a set of parameters. Secondly, if the hypothesis is correct, a probability density function (PDF) is introduced which follows the histogram of the variable in the data. The last step is to compare the experimental data with the theory. The parameters of the theoretical model are adjusted for better agreement between the experimental data and the theory. The result of this technique is the evaluation of the variable of interest and if the probability density function of this value is of high probability, then the theory is verified. Else, we can't say that the theory is false but that the range of parameters employed didn’t allow for an acceptable agreement [1].
The distribution of the number of observed events in a simple counting experiment for a given amount of time and which are uncorrelated between them is described by a Poisson probability density function pdf:

\[ p(n) = \frac{e^{-\lambda} \lambda^n}{n!} \]  

(9.1)

where \( n \) is the number of observed events and \( \lambda \) is a Poisson variable. The mean value of this parameter and its variance are a good interpretation of the poissonian variable. Figure 9.1 shows an example of poissonian distributions for different values of the parameter \( \lambda \) [1].

In the case of binned data (fit to histogram), the \( N \) bins are considered as independent from each other and the number of events in each bin \( n_i \) follows the poissonian distribution:

\[ p(n_i, \ldots n_N) = \prod_{i=1}^{N} \frac{\lambda_i^n e^{-\lambda_i}}{n_i!} \]  

(9.2)

where \( \lambda_i \) is the expected number of events in each bin.

![Poissonian distributions for different \( \lambda \) parameters: \( \lambda = 0.5 \) (red line), \( \lambda = 5 \) (green line) and \( \lambda = 20 \) (blue line) [1].](image)

To check the agreement between observation and expectation, the notation likelihood is introduced. The likelihood is the product of probability density function (PDF) in each bin and for the poissonian case where no bin correlations occur is defined as follows:
\[ \mathcal{L}(a/b) = \prod_{i=1}^{N} \frac{b_i^n e^{-b_i}}{n_i!} \]  \hspace{1cm} (9.3)\\

where \( b_i \) are the events expected in the bin and \( n \) is the number of observed events in each bin. Typically the number of expected events in each bin can be write as:

\[ b_i = \sum_{j=1}^{n_{\text{source}}} L \sigma_j \epsilon_{ij} \]  \hspace{1cm} (9.4)\\

for integrated luminosity \( L \), cross section \( \sigma_j \) for source \( j \), and efficiency \( \epsilon_{ij} \) for source \( j \) in bin \( i \), obtain from MC simulation of the process. The source include the signal and background processes.

For the study of a functional dependence between two variables \( x \) and \( z \):

\[ z = f(x/\theta) \]  \hspace{1cm} (9.5)\\

with \( \theta \) a set of parameters, a gaussian likelihood can be introduced as below:

\[ \mathcal{L}(z/\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi} \sigma_i} e^{-\frac{(z_i-f(x/\theta))^2}{2\sigma_i^2}} \]  \hspace{1cm} (9.6)\\

where \( \sigma_i \) are the uncertainties from the gaussian fluctuations which are characterized from \( z_i \) measurements. In this analysis a gaussian likelihood fit is used to interpreted the nuisance parameters affecting the event yields of MC samples.

The likelihood in both previous cases is characterized by exponential functions but, for the sake of simplicity, the log-likelihood function is preferred. By taking the logarithm of a function, the product becomes a sum, and for an exponential function, one can have a linear function fit and be parametrized. For example, the Gaussian likelihood becomes:

\[ -2 \ln \mathcal{L}(z/\theta) = -2 \sum_{i=1}^{N} \ln \sqrt{2\pi} \sigma_i + \frac{(z_i-f(x/\theta))^2}{\sigma_i^2} \]  \hspace{1cm} (9.7)
The Poissonian and the gaussian probability density function (PDF) are set as an example, but in order to be able to use the likelihood function, the PDF should be known. The PDF in general depends on the considering problem and can be estimated by the “toy Monte Carlo” method. The toy MC method is an excellent tool in the hands of a physicist, since it helps to check, if the likelihood function is the proper one or not. It actually generates a pseudo-data events using different parameters which can help to define the likelihood if the probability density function is known [2].

9.1.2 Maximum Likelihood Fit

A maximum likelihood fit is performed in order to infer the real value of the parameter of interest. The true value of a parameter \( \alpha \) can be estimated by maximizing the likelihood function. For simplicity, instead of maximizing the likelihood it is preferred to maximize the logarithm of likelihood \( \ln \mathcal{L} \).

In order to find the value of parameter \( \alpha \), the derivative of the logarithm likelihood is set equal to zero:

\[
\frac{d(\ln \mathcal{L})}{d \alpha} = 0
\]

or by covariant derivative for more than one parameter denoted as \( \alpha_j \):

\[
\frac{\partial (\ln \mathcal{L})}{\partial \alpha_j} = 0
\]  

As an example, a Gaussian likelihood is considered for a variable \( M_B \):

\[
\ln \mathcal{L} = -N \ln \sqrt{2 \pi \sigma^2} - \frac{1}{2} \sum_{i}^{N} \frac{(M_i - M_B)^2}{\sigma^2}
\]  

Taking the covariant derivative as equal to zero:

\[
\frac{\partial (\ln \mathcal{L})}{\partial M_i} = 0
\]

The logarithm function is monotonic and the values that maximize the likelihood maximize also the logarithm likelihood.
the parameters that maximize the log-likelihood are:

\[
\hat{M}_b = \frac{1}{N} \sum_{i} M_i
\]  
(9.12)

\[
\hat{\sigma}^2 = \frac{1}{N} \sum_{i} (M_i - \hat{M}_b)^2
\]  
(9.13)

In this way one can estimate the mean value of the parameter of interest and its variance [2]. However, in order to be sure about the “goodness” of the maximum likelihood estimator, the toy MC method also is used in parallel.

Besides a good estimation of the parameter’s value, it is also important to know the validity of this value compared to the real one. This uncertainty between the estimation and the real value is sometimes expressed by the confidence interval and sometimes by the standard deviation. For example, if the uncertainty is set to one standard deviation, 1\(\sigma\), that means 68% confidence interval. In other words, a confidence interval of 68% CL has 32% probability to contain the true value of a parameter.

The maximum likelihood method also provides an estimation of this uncertainty. Recalling the Central Limit Theorem, the likelihood function for one parameter with a large number of events \(N\) become’s a Gaussian function. Thus, the log-likelihood function is a parabolic function of \(\alpha\) parameter, as shown below:

\[
\ln \mathcal{L}(\alpha) = \ln \mathcal{L}(\hat{\alpha}) + \frac{1}{2} \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} \bigg|_{\alpha = \hat{\alpha}} (\alpha - \hat{\alpha})^2
\]  
(9.14)

The standard deviation of Maximum likelihood estimators for the parameter \(\alpha\) is equivalent to:

\[
\Delta \alpha = \sqrt{(\alpha - \alpha_0)^2} = \frac{1}{\frac{\partial^2 \ln \mathcal{L}}{\partial \alpha^2} \bigg|_{\alpha = \hat{\alpha}}}
\]  
(9.15)
with the parameter $\alpha$ to be written as $\alpha = \hat{\alpha} \pm \Delta \alpha$. Figure 9.2 (the left part) shows an example of a symmetric log-likelihood function and its parabolic form.

In the case of small number of events, the log-likelihood is not symmetric or parabolic. Figure 9.2 (right part) shows an example of such asymmetric log-likelihood function with $\alpha^-$ and $\alpha^+$ representing the $\alpha$ confidence interval of 68% (1\(\sigma\) standard deviation).

Figure 9.2: A symmetric log-likelihood function on the left hand side plot and an asymmetric log-likelihood function on the right hand side. The notation $L^*$ and $\alpha^*$ are $L(\hat{\alpha})$ and $\hat{\alpha}$ respectively [3].

9.2 Systematic Uncertainty Probability Density Functions

The uncertainties are usually derived from experimental uncertainties, statistical uncertainties in the background control region and theoretical uncertainties affecting the signal or the background processes. All of them are embedded in the likelihood [4] by introducing nuisance parameters $\theta$, with a probability density function:

$$
\rho(\theta/\bar{\theta}) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(\theta - \bar{\theta})^2}{2\sigma^2}}
$$

(9.16)

with $\bar{\theta}$ the best estimation of the real value of the $\theta$ parameter.

The Gaussian PDF, is not the proper one for positively defined observables, such as the cross section. For that reason instead of the PDF, a log-normal PDF is used as follows:
\[
\rho(\theta/\tilde{\theta}) = \frac{1}{\sqrt{2\pi \ln(\kappa)}} \exp \left( -\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln \kappa)^2} \right) \frac{1}{\theta}
\]

(9.17)

In log-normal, the width \(\sigma\) is replaced by \(\ln(\kappa)\), with \(\kappa\) to characterizing the log-normal distribution which has a longer tail than the Gaussian one. Figure 9.3 shows the log-normal distributions for different \(\kappa\) values.

Additionally, in the case where the number of observed events in the control region is limited, leading to statistical uncertainties, a gamma distribution is used instead of the log-normal or Gaussian probability density function. The gamma distribution, describing the uncertainties in the signal rate \(n\), if \(N\) events are observed, is given from the equation below:

\[
\rho(n) = \frac{1}{\alpha \alpha^n} \frac{1}{N!} \exp \left( -\frac{n}{\alpha} \right)
\]

(9.18)

with \(\alpha = N/n\). The right side of figure 9.3, shows the gamma probability density function distributions for different numbers of control samples \(N\).
9.3 Observation – Upper Limits

Statistical methods have been used in LHC for the discovery of new physics or at least in order to define meaningful exclusion limits in signal models which have not been verified. The Hypothesis test is a common tool in statistical methods which can determine the compatibility of the observed data with one or more theoretical models.

Two hypotheses are considered in this statistical test: a simple hypothesis $H_0$ and an alternative hypothesis $H_1$. In the search of new processes, in the null Hypothesis $H_0$, samples contain only background events. For example, the data samples collected consist only of events from a known type (SM events). On the other hand, in the $H_1$ hypothesis a sample contains background and signal events whose existence is not known. The selection of events in both hypotheses is based on a discriminate variable $x$ or a set of variables $[5]$.

The main goal of this test is to reject the background hypothesis in order to establish or exclude the signal model. The establishment of a new phenomenon is not a simple endeavor. Strong evidence is required, to avoid a false discovery. The probability of a false discovery should be very small and the “significance” of the test should be large.

The significance level and the power of the test are two quantities that can give information about the test. The significance level $\alpha$ is the probability to reject the background hypothesis $H_0$ if it is true:

$$\alpha = \int_{y_{\text{cut}}}^{\infty} p(y(\vec{x})|b) \, dt \quad (9.19)$$

where $p(y(\vec{x})|b)$ is the probability density function to be observed in the data variable $x$ or a set of variables under the background hypothesis. $\alpha$ is a constant defined before the test and it is usually a small number (e.g.: 5%), is actually representing the events that are being rejected as background and which can be accepted as signal. In other words, it is the background efficiency. The $y_{\text{cut}}$ defines the critical region for which the background is rejected as shown in Figure 9.4.

The probability to accept the signal events as signal is called the power of the test, which is also representing the signal efficiency:
\[
\int_{y_{\text{cut}}}^{\infty} p(y|s) \, dt = 1 - \beta
\]  \hspace{1cm} (9.20)

where \( \beta \) is the misidentification probability.

**Figure 9.4:** Distributions of probability density function under the background and signal hypothesis.

The remaining question is how can someone choose the critical region in such a way as to get the optimal solution for this two specific hypotheses background and signal. The highest power for a given significance level can be obtained for a certain critical region if the ratio of the probability density functions for signal and background are greater than a given constant, \( c \), in the critical region only (Neyman – Pearson lemma states):

\[
\frac{p(y|s)}{p(y|b)} > c
\]  \hspace{1cm} (9.21)

In the case of Monte Carlo processes, where the data contain a variety of variables and the functions \( y(\vec{x}) \) are more complicated, the Neyman – Pearson lemma states [6] that is not the best way to discriminate the signal region from the background region. There are multivariate statistical methods, like neutral networks, boosted decision trees and support vector machines [7-10] that are being developed and used, although the most common software used in HEP is the TMVA [11].
9.3.1 Observation

When a variable $x$ is observed it is necessary to quantify the level of agreement between data and a hypothesis $H$. This is expressed with the so called p-value. P-value is the probability under an assumption $H$ to observe data with equal or less compatibility in this region [12]. The p-value is considered for different models and if its value is lower or equal to $\alpha$, the theoretical model is rejected.

Instead of a p-value, a physicist often defines the significance $Z$, as the number of standard deviations that a Gaussian variable fluctuates given the p-value as illustrated in Figure 9.5.

![Figure 9.5: Distribution of a Gaussian probability density function with mean equal to zero and standard deviation equal to 1. Significance $Z$ definition is illustrated [12].](image)

The significance $Z$ value is often used to support a possible discovery. In High Energy Physics, the $Z$ should be equal to 5 (meaning $5\sigma$), which corresponds to a p-value considering only background hypothesis of $2.9 \times 10^{-7}$ in order to claim a new discovery.

9.3.2 Limits

In the case of a theoretical model which doesn’t pass the goodness-of-fit test and the alternative hypothesis $H_1$ is excluded by the data, meaning that there is no possible discovery, upper limits can be set on the procedure rate or cross section of the hypothetical signal. The upper limits denote that if there is a possibility the particle of interest to exist, it can be possible only if the cross section or its production rate are lower than the value of this limit, with a certain probability [1].
A modified frequentist approach or CL method, is the latest method developed and has been applied in many LHC analyses. Feldman and Cousins [13] introduced a method based on the classical frequentist approach, the \( q_\mu \) test statistic:

\[
\tilde{q}_\mu = -2 \ln \frac{L(data|\mu s+b)}{L(data|\mu s+b)} , \quad 0 \leq \mu \leq \mu
\] (9.22)

The \( \mu \) is the signal strength that maximizes the likelihood \( L(data|\mu s+b) \), with the constraint to be greater or equal to zero, \( \mu \geq 0 \) and lower or equal than \( \mu \) (where \( \mu \) is the signal strength of statistic test, a scale factor for all the production cross sections of the signal, so sometimes is written as a scale factor for all the production cross sections of the signal, \( \mu = \sigma/\sigma_{SM} \)), avoiding the negative signal rate and obtaining one-sided confidence interval.

The systematic uncertainties on the signal \( s(\theta) \) and background \( b(\theta) \) are also introduced to test statistic. The effect of systematic uncertainties is introduced by extending the Poissonian-likelihoods in such a way, such that the probability function of the nuisance parameter \( \rho(\theta/\tilde{\theta}) \) to be included:

\[
L(data|\mu s+b,\theta) = \text{Poisson}(data|\mu s+b(\theta)).\rho(\theta/\tilde{\theta})
\] (9.23)

The modified test statistic including the systematic uncertainties is defined below:

\[
\tilde{q}_\mu = -2 \ln \frac{L(data|\mu s(\theta)+b(\theta),\hat{\theta}_0)}{L(data|\mu s(\theta)+b(\theta),\hat{\theta})} , \quad 0 \leq \mu \leq \mu
\] (9.24)

From equation 9.24 one can easily compute the signal strength \( \mu \) of the observed test statistic \( q_\mu^{obs} \) as well as the nuisance parameters \( \hat{\theta}_0^{obs} \), \( \hat{\theta}_\mu^{obs} \) that maximize the likelihood for the background only (\( \mu = 0 \)) and signal plus background hypothesis (\( \mu = 1 \)), respectively. The nuisance parameters \( \hat{\theta}_0^{obs} \), \( \hat{\theta}_\mu^{obs} \) are the optimal values of the nuisance parameter and are obtained by fitting on the observed data.

The toy Monte-Carlo pseudo-dataset is used to estimate the probability density functions of the test statistic \( f(q_\mu|b(\theta),\hat{\theta}_0^{obs}) \) and \( f(q_\mu|\mu s(\theta)+b(\theta),\hat{\theta}_\mu^{obs}) \), which describe the test statistic.
distribution in the background and in signal plus background hypothesis, respectively. Then, one can estimate the p-values of both hypotheses as below:

\[
p_{\mu s+b} = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | s+b) = \int_{\tilde{q}_\mu}^{\infty} f(\tilde{q}_\mu | \mu, s(\theta) + b(\theta), \hat{\theta}_\mu^{\text{obs}}) \, d\tilde{q}_\mu
\]

(9.25)

\[
1-p_b = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{\text{obs}} | b) = \int_{\tilde{q}_\mu}^{\infty} f(\tilde{q}_\mu | b(\theta), \hat{\theta}_\mu^{\text{obs}}) \, d\tilde{q}_\mu
\]

(9.26)

The ratio of these two probabilities is giving the confidence level \(CL_s\) for a given signal strength:

\[
CL_s = \frac{CL_{\mu s+b}}{CL_b} = \frac{p_{\mu s+b}}{1-p_b}
\]

(9.27)

For example, if the \(CL_s \leq 0.05\) (5%) for the signal plus background hypothesis (\(\mu = 1\)) then a particle X is excluded with 95% CL.

A large number of toy MC pseudo-data are generated with a signal strength \(\mu\) that gives \(CL_s = 5\%\) \((\mu^{95\%CL})\). The median of the expected limits (50% quantile) and the uncertainty bands of \(\pm 1\) and \(\pm 2\) standard deviations for the expected value of \(\mu\), only for background hypothesis are shown in Figure 9.6.

**Figure 9.6:** Cumulative probability of the signal strengths with the median expected limit and the \(\pm 1\sigma\) (68%) and \(\pm 2\sigma\) (95%) bands as a function of the signal strength \((\mu^{95\%CL})\) [13].
References


CHAPTER 10

Results and Future Prospects

10.1 Results

The search for an excess of signal events over the expected background involves a global maximum likelihood fit, based on the mass distributions in the different channels and categories. Systematic uncertainties are usually derived from experimental, statistical and theoretical uncertainties. The search for the existence of the exotic decay of the 125GeV Higgs to a pair of light pseudoscalar bosons and further decay to a pair of b jets and a pair of tau leptons \((h \rightarrow aa \rightarrow bb\tau\tau)\), is largely dominated by statistical uncertainties over systematic uncertainties.

The systematic uncertainties are represented by nuisance parameters that are varied in the fit according to their probability density function. A log-normal probability density function (eq. 9.17) is assumed for the nuisance parameters affecting the event yields of the various background contributions, whereas systematic uncertainties that affect the distributions are represented by nuisance parameters whose variation results in a continuous perturbation of the spectrum and which are assumed to have a Gaussian probability density function (eq. 9.16). To take into account the limited size of simulated samples and of data in the control regions used to estimate some of the background processes, statistical uncertainties in individual bins of the \(m_{\text{vis}_{\tau\tau}}\) distributions are considered as Poissonian nuisance parameters. The combined effect of all these uncertainties is the dominant systematic uncertainty in this search. All of these systematics are been embedded in a binned Poissonian-likelihood function by introducing nuisance parameters \(\theta\), with a log-normal probability density function or a Gaussian probability density function. A general function for this extended likelihood is shown below:

\[
\mathcal{L}(n|s+b) = \prod_{i=1}^{N} \frac{e^{-(s+b)}(s+b)^{n_i}}{n_i!} \prod_{j=1}^{L} p(\tilde{\theta}_j|\theta_j) \tag{10.1}
\]

The first product is basically the product of Poisson functions over the bins and the second is the product for all nuisance parameters. Where \(n\) is the number of observed events in each bin and \(s+b\) is the sum of signal and background events expected. In formula 10.1 the \(i\) runs over the number of bins and \(j\) over the number of nuisances parameters \(\theta\).
The visible ditau mass distributions in the three categories of the $e\mu$, $e\tau_h$, and $\mu\tau_h$ final states, after the global binned maximum likelihood fit are shown in Figures 10.1, 10.2, and 10.3, respectively. In the most statistically limited regions, which are the low mass categories, the bin-by-bin uncertainty for one bin summing all processes is up to 40%. The uncertainty bands in all plots are postfit and the postfit reducible background uncertainty is 7%, because of the constrains from high mass categories (control region). The postfit yield predictions for background and the observed data in various categories is shown in table 10.1. The impacts related to bin-by-bin uncertainties of backgrounds in the poorly populated signal regions can be found in Appendix A4.

Figure 10.1: Distributions of $m_{\tau\tau}^{\text{vis}}$ in the four categories of the $e\mu$ channel. The “Other” contribution includes events from single top quark, diboson, SM Higgs boson, and $W +$ jets productions. The signal histogram corresponds to the SM production cross section for $ggh$, VBF, and Vh processes, and assumes $B( h \rightarrow aa \rightarrow 2\tau2b ) = 10\%$. The normalizations of the predicted background distributions correspond to the result of the global fit. There is a very good agreement between observed and expected events.
No excess is observed relatively to the SM background prediction. Upper limits at 95% CL are set on the signal strength using the modified frequentist construction CLs [1, 2], for pseudoscalar masses between 15 and 60 GeV. The limits per channel and the combination of the three channels are shown in Figure 10.4.

### Table 10.1: The postfit yield predictions for background and the observed data in various categories.

<table>
<thead>
<tr>
<th>Process</th>
<th>Low-mass SR</th>
<th>Medium-mass SR</th>
<th>High-mass SR</th>
<th>High-mass CR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>eμ</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QCD multijet</td>
<td>4.45 ± 0.72</td>
<td>15.91 ± 2.78</td>
<td>47.15 ± 7.59</td>
<td>285.97 ± 47.25</td>
</tr>
<tr>
<td>W ± jets</td>
<td>0.95 ± 0.17</td>
<td>1.07 ± 0.60</td>
<td>0.52 ± 0.10</td>
<td>12.89 ± 2.33</td>
</tr>
<tr>
<td>t̄</td>
<td>22.82 ± 2.25</td>
<td>70.64 ± 4.52</td>
<td>165.76 ± 7.92</td>
<td>4380.14 ± 84.94</td>
</tr>
<tr>
<td>Z → ττ</td>
<td>7.46 ± 2.10</td>
<td>51.47 ± 5.13</td>
<td>77.94 ± 6.74</td>
<td>542.61 ± 28.36</td>
</tr>
<tr>
<td>Z → ℓℓ</td>
<td>0.33 ± 0.02</td>
<td>2.16 ± 0.10</td>
<td>4.92 ± 2.88</td>
<td>17.39 ± 0.84</td>
</tr>
<tr>
<td>Single top</td>
<td>1.63 ± 0.22</td>
<td>4.90 ± 0.60</td>
<td>7.11 ± 0.93</td>
<td>297.31 ± 36.07</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.34 ± 0.03</td>
<td>0.85 ± 0.08</td>
<td>2.48 ± 0.18</td>
<td>21.36 ± 1.50</td>
</tr>
<tr>
<td>SM H</td>
<td>0.24 ± 0.03</td>
<td>0.56 ± 0.03</td>
<td>0.67 ± 0.03</td>
<td>10.95 ± 0.47</td>
</tr>
<tr>
<td><strong>Total expected</strong></td>
<td>38.21 ± 3.15</td>
<td>147.55 ± 6.74</td>
<td>306.55 ± 10.47</td>
<td>5568.62 ± 71.36</td>
</tr>
<tr>
<td><strong>Observed data</strong></td>
<td>33</td>
<td>149</td>
<td>298</td>
<td>5565</td>
</tr>
<tr>
<td><strong>etb</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jet → τb bkg.</td>
<td>14.57 ± 2.02</td>
<td>64.31 ± 5.42</td>
<td>128.16 ± 9.83</td>
<td>975.91 ± 65.34</td>
</tr>
<tr>
<td>t̄</td>
<td>8.86 ± 1.14</td>
<td>37.49 ± 3.63</td>
<td>101.41 ± 5.26</td>
<td>1183.56 ± 51.61</td>
</tr>
<tr>
<td>Z → ττ</td>
<td>8.47 ± 2.15</td>
<td>76.35 ± 10.24</td>
<td>123.00 ± 21.50</td>
<td>512.35 ± 32.93</td>
</tr>
<tr>
<td>Z → ℓℓ</td>
<td>0.00 ± 0.00</td>
<td>7.83 ± 3.76</td>
<td>88.02 ± 8.99</td>
<td>553.46 ± 29.84</td>
</tr>
<tr>
<td>Single top</td>
<td>0.44 ± 0.08</td>
<td>4.08 ± 0.57</td>
<td>13.08 ± 1.60</td>
<td>143.15 ± 18.18</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.27 ± 0.03</td>
<td>1.10 ± 0.10</td>
<td>3.24 ± 0.24</td>
<td>26.72 ± 2.08</td>
</tr>
<tr>
<td>SM H</td>
<td>0.07 ± 0.00</td>
<td>0.45 ± 0.04</td>
<td>0.95 ± 0.07</td>
<td>10.21 ± 0.53</td>
</tr>
<tr>
<td><strong>Total expected</strong></td>
<td>34.98 ± 5.30</td>
<td>191.61 ± 9.68</td>
<td>457.86 ± 23.05</td>
<td>3405.35 ± 54.25</td>
</tr>
<tr>
<td><strong>Observed data</strong></td>
<td>30</td>
<td>190</td>
<td>452</td>
<td>3411</td>
</tr>
<tr>
<td><strong>μtb</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jet → τb bkg.</td>
<td>41.85 ± 3.66</td>
<td>161.52 ± 11.55</td>
<td>380.25 ± 24.87</td>
<td>2097.29 ± 135.25</td>
</tr>
<tr>
<td>t̄</td>
<td>17.49 ± 1.38</td>
<td>58.79 ± 3.56</td>
<td>270.31 ± 13.48</td>
<td>2454.14 ± 99.05</td>
</tr>
<tr>
<td>Z → ττ</td>
<td>36.28 ± 6.07</td>
<td>138.14 ± 10.62</td>
<td>372.10 ± 22.79</td>
<td>1323.82 ± 60.29</td>
</tr>
<tr>
<td>Z → ℓℓ</td>
<td>0.38 ± 0.03</td>
<td>0.43 ± 0.03</td>
<td>23.32 ± 3.36</td>
<td>162.03 ± 15.07</td>
</tr>
<tr>
<td>Single top</td>
<td>0.46 ± 0.08</td>
<td>3.88 ± 0.56</td>
<td>20.09 ± 2.48</td>
<td>210.45 ± 26.09</td>
</tr>
<tr>
<td>Diboson</td>
<td>0.30 ± 0.03</td>
<td>1.95 ± 0.16</td>
<td>4.49 ± 0.33</td>
<td>30.79 ± 2.38</td>
</tr>
<tr>
<td>SM H</td>
<td>0.21 ± 0.01</td>
<td>0.97 ± 0.06</td>
<td>1.96 ± 0.17</td>
<td>21.31 ± 0.97</td>
</tr>
<tr>
<td><strong>Total expected</strong></td>
<td>96.97 ± 6.53</td>
<td>365.68 ± 12.22</td>
<td>1072.53 ± 24.71</td>
<td>6299.83 ± 70.22</td>
</tr>
<tr>
<td><strong>Observed data</strong></td>
<td>99</td>
<td>372</td>
<td>1079</td>
<td>6289</td>
</tr>
</tbody>
</table>
The relative sensitivity of the different channels can be understood by looking at the acceptance after the selection of the leptons and of the b jet, which is illustrated in Figure 6.11. As shown in Figure 10.5, the most sensitive final state is $\mu\tau$.

**Figure 10.2:** Distributions of $m_{\text{vis}}$ in the four categories of the $e\tau_h$ channel. The “jet→$\tau_h$” contribution includes all events with a jet misidentified as a $\tau_h$ candidate, whereas the rest of background contributions only include events where the reconstructed $\tau_h$ corresponds to a $\tau_h$, a muon, or an electron, at the generator level. The “Other” contribution includes events from single top quark, diboson, and SM Higgs boson processes. The signal histogram corresponds to the SM production cross section for ggh, VBF, and Vh processes, and assumes $B(h \to aa \to 2\tau 2b) = 10\%$. The normalizations of the predicted background distributions correspond to the result of the global fit. There is a very good agreement between observed and expected events.
The sensitivity of the $e\tau_h$ and $e\mu$ channels is approximately equivalent; the first channel suffers from higher trigger thresholds and lower object identification efficiency than $\mu\tau_h$, and the second one suffers from a lower branching fraction than $\mu\tau_h$. At low mass the $e\mu$ final state has a higher signal acceptance than the other final states, especially $e\tau_h$. The limits are best in the intermediary mass range because the analysis has been optimized for $m_s = 40$ GeV.

**Figure 10.3:** Distributions of $m^{\text{vis}}_{\tau\tau}$ in the four categories of the $\mu\tau_h$ channel. The “jet $\rightarrow \tau_h$” contribution includes all events with a jet misidentified as a $\tau_h$ candidate, whereas the rest of background contributions only include events where the reconstructed $\tau_h$ corresponds to a $\tau_h$, a muon, or an electron, at the generator level. The “Other” contribution includes events from single top quark, diboson, and SM Higgs boson processes. The signal histogram corresponds to the SM production cross section for $\text{ggh}$, VBF, and Vh processes, and assumes $B(h \rightarrow aa \rightarrow 2\tau2b) = 10\%$. The normalizations of the predicted background distributions correspond to the result of the global fit. There is a very good agreement between observed and expected events.
In addition, the low mass signals have a lower acceptance because of the \( p_T \) thresholds and of the overlap of the leptons due to the boost of the pseudoscalar bosons. The high mass signals lie in a region where more backgrounds, including Drell-Yan production contribute. The categories are complementary over the probed mass range, with the low-mass signal regions being more sensitive for heavy resonances, and the heavy-mass regions for light resonances.

**Figure 10.4:** Expected and observed 95% CL limits on \( \sigma(h)/\sigma_{\text{SM}} B(h \rightarrow aa \rightarrow 2b2\tau) \) in %. The \( e\mu \) results are shown in the top left panel, \( e\tau_h \) in the top right, \( \mu\tau_h \) in the bottom left, and the combination in the bottom right.
The combined limit at intermediate mass is about 3% on \( B(\rightarrow aa \rightarrow 2b2\tau) \), assuming the Standard Model production cross section for the Higgs boson in gluon fusion and VBF production modes. This translates to limits on \( B(h \rightarrow aa) \) as low as about 20–23% in type-1 and type-2 2HDM+S (including NMSSM). In the scenario with the highest branching fraction, 2HDM+S type-3 with \( \tan\beta = 2 \), the expected limit is as low as 6% for \( m_a = 40 \) GeV. Figure 10.6 shows the expected limits at 95% CL on \( B(h \rightarrow aa) \) for all types of 2HDM+S, for various values of \( \tan\beta \) and of the resonance mass. Figure 10.7 shows the expected limits at 95% CL on \( B(h \rightarrow aa) \) for a few scenarios of 2HDM+S.

**Figure 10.5:** Comparison by category of expected 95% CL limits on \( \sigma(h)/\sigma_{SM} B(H \rightarrow aa \rightarrow 2b2\tau) \) in %. The e\(\mu\) results are shown in the top left panel, e\(\tau\) in the top right, \(\mu\tau\) in the bottom left, and the combination in the bottom right. The sensitivity of the various categories is complementary over the probed mass range.
10.2 Summary

A search for exotic decays of the 125 Higgs boson, with two b jets and two taus in the final state, was performed with 35.9 fb$^{-1}$ of data, collected at 13 TeV center-of-mass energy in 2016, with the CMS Experiment at CERN. This channel has never been searched before due to its small cross section. Nevertheless, it has a large branching fraction in many theoretical models and can be triggered with high efficiency due to the presence of light leptons from leptonic tau decays.

![Figure 10.6: Observed 95% CL on B(h → aa) in 2HDM+S type-1 (top left), type-2 (top right), type-3 (bottom left), and type-4 (bottom right).](image-url)
The signal and background processes were modeled with Monte Carlo simulations. For the $h \rightarrow aa \rightarrow bb\tau\tau$ signal process three main production modes were considered: gluon-gluon fusion, vector boson fusion and vector boson associated production. With these objects in the final state the background distribution from several processes were considered: Drell-Yan, $W$+jets, $t\bar{t}$, single top, diboson and Standard Model Higgs boson decays. The dominant backgrounds were the $t\bar{t}$ and $Z \rightarrow \tau\tau$ productions. Another large background consists of events with jets that were faking $\tau$s, among events from QCD multijet events, $W$+jets, $Z$+jets and $t\bar{t}$ processes.

In order to compare the simulated events with data a series of corrections were applied to the simulated samples. These are the energy scale corrections of the leptons, lepton identification efficiency corrections, leptons $p_T$ corrections and trigger efficiency corrections. Also all simulated events were reweighted in order to have the same pileup distributions as the real data.

Three different $\tau\tau$ final states were considered: $e\mu$, $e\tau$, and $\mu\tau$. The events in each final state were selected with triggers or a combination of triggers depending on the final state leptons. A set of selection criteria were applied at the reconstruction level in order to identify and isolate the final products in each channel. All events were required to have at least one b-tagged jet.

To increase the sensitivity of the analysis, the events in each final state were separated into four categories depending on the signal over background ratios. The categories were defined based on the visible invariant mass of the leptons and of the leading b-tagged jet. The thresholds that define the categories vary depending on the $\tau\tau$ final state. Selection criteria depending on the category and final state were applied to optimize the expected limits on the product of the signal cross section times the corresponding branching fractions. The cut applied were based on the transverse mass between leptons, and the MET and on $P_\zeta$ parameter.

A global binned Maximum likelihood fit was performed on the visible di-tau mass distribution in the different channels and categories. Systematic uncertainties affected the analysis and which were related to physics objects, to background estimations and others, were represented by nuisance parameters. No excess of events is found on top of the expected SM background.

Upper limits are set on $B(h \rightarrow aa \rightarrow bb\tau\tau)$. They range from 3 to 12% over the pseudoscalar mass range between 15 and 60 GeV. This corresponds to upper limits on $B(h \rightarrow aa)$ between 6% and 20% in the most favorable 2HDM+S scenarios. Figure 10.8 shows the observed and expected limits
of $B(h \rightarrow aa)$ for different channels and the four 2HDM+S types, as derived from recent CMS analysis in Run 1 (8TeV, 19.6 fb$^{-1}$) and in Run 2 (13TeV, 35.9fb$^{-1}$). The results show a large improvement on the sensitivity for searching of the exotic Higgs decay to a pair of light pseudoscalar bosons compared to the other channels. More specifically, in the NMSSM model (a special case of 2HDM+S typeII) the sensitivity is improved by more than one order of magnitude compared to $2\mu2\tau$ final state in the mass range of 15 -- 25 GeV [3] and up to 5 times in the $2\mu2b$ final state in the mass ranges 25 -- 60 GeV [4].

![Figure 10.7: Observed 95% CL limits on $\sigma(h)/\sigma_{SM} B(H \rightarrow aa)$ for various scenarios of 2HDM+S.](image)

### 10.3 Future Prospects

The increasing on the integrated luminosity entering the High-Luminosity LHC period is expected to reduce the upper limits predicted. When the integrated luminosity reach the 1000fb$^{-1}$ the limits expected to be reduced by a factor of ~2. The plot in figure 10.9 represent the upper limits on $\sigma(h)/\sigma_{SM} B(h \rightarrow aa \rightarrow bb\tau\tau)$ after a hard code calculations. Nevertheless, with the increasing of the integrated luminosity, which the limits set which are currently mostly statistically dominated, are expected to play progressively an important role.
Figure 10.8: Observed and expected limits of $\text{B}(h \rightarrow aa)$ for different channels and 2HDM+S types, as derived from recent CMS analysis in Run 1 (center-of-mass-energy 8TeV, 19.6 fb$^{-1}$) and in Run 2 (center-of-mass-energy 13TeV, 35.9 fb$^{-1}$). Grey shaded regions correspond to regions where theoretical predictions for the branching fractions of the pseudoscalar boson to SM particles are not reliable. The 2HDM+S type I is independent of $\tan\beta$ [5].
Preliminary studies for the High-Luminosity LHC conditions for this channel have already begun [6]. Two different scenarios for the extrapolation of the systematic uncertainties are currently under focus. The first scenario $S_1$ assumes that nothing will be changed with respect to LHC Run 2 period and there will not be any improvement on the theoretical distributions of relevant physics effects. All the theoretical and experimental systematics uncertainties are assumed unchanged (as in Run 2) despite the increasing of the integrated luminosity. The second scenario $S_2$ assumed that there will be a change on the theoretical distributions of relevant physics effects (compare to Run 2). More specifically the theoretical uncertainties are assumed to be reduced by a factor of two and experimental systematic uncertainties will scale with the square root of the integrated luminosity. We should also expect to reach a defined lower limit based on the accuracy of the upgraded detector.

The results from these two scenarios and for an integrated luminosity of $3000 fb^{-1}$ are shown in figure 10.10. The difference between the limits in the $S_1$ and $S_2$ scenarios is of the order of 5%, and
we are expecting to do another 5% better if all systematic uncertainties can be mastered. Thus, with the integrated luminosity of 3000 fb$^{-1}$ will improve the sensitivity by about an order of magnitude for the search in the $h \rightarrow aa \rightarrow bb\tau\tau$ channel.

**Figure 10.10:** Projected expected limits ($\sigma_{\text{h}}/\sigma_{\text{SM}})B(h \rightarrow aa \rightarrow 2\tau 2b)$, comparing different scenarios for systematic uncertainties for an integrated luminosity of 3000 fb$^{-1}$ [6].
References


# APPENDIX A1

**Table A1.1:** List of data sets included in the analysis. Mini-AOD is used for data.

<table>
<thead>
<tr>
<th>Channel</th>
<th>Data set</th>
<th>Run range</th>
<th>Integrated Luminosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016B-03Feb2017-ver2-v2/MINIAOD</td>
<td>272007–275376</td>
<td>5.788 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016C-03Feb2017-v1/MINIAOD</td>
<td>275657–276283</td>
<td>2.573 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016D-03Feb2017-v1/MINIAOD</td>
<td>276315–276811</td>
<td>4.248 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016E-03Feb2017-v1/MINIAOD</td>
<td>276831–277420</td>
<td>4.009 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016F-03Feb2017-v1/MINIAOD</td>
<td>277772–278808</td>
<td>3.102 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016G-03Feb2017-v1/MINIAOD</td>
<td>278820–280385</td>
<td>7.540 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016H-03Feb2017-ver2-v1/MINIAOD</td>
<td>280919–284044</td>
<td>8.606 /fb</td>
</tr>
<tr>
<td>eμ</td>
<td>/MuonEG/Run2016H-03Feb2017-ver3-v1/MINIAOD</td>
<td>280919–284044</td>
<td>see above</td>
</tr>
<tr>
<td>μτb</td>
<td>/SingleMuon/Run2016B-03Feb2017-ver2-v2/MINIAOD</td>
<td>272007–275376</td>
<td>5.788 /fb</td>
</tr>
<tr>
<td>μτb</td>
<td>/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD</td>
<td>275657–276283</td>
<td>2.573 /fb</td>
</tr>
<tr>
<td>μτb</td>
<td>/SingleMuon/Run2016H-03Feb2017-ver3-v1/MINIAOD</td>
<td>280919–284044</td>
<td>see above</td>
</tr>
<tr>
<td>eτb</td>
<td>/SingleElectron/Run2016C-03Feb2017-v1/MINIAOD</td>
<td>275657–276283</td>
<td>2.573 /fb</td>
</tr>
<tr>
<td>eτb</td>
<td>/SingleElectron/Run2016H-03Feb2017-ver3-v1/MINIAOD</td>
<td>280919–284044</td>
<td>see above</td>
</tr>
</tbody>
</table>
Table A1.2: MC background samples included in the analysis. All samples are generated for p–p collisions at a center-of-mass energy of 13 TeV. Samples used in this analysis are reconstructed and stored in miniAOD format. A k-factor of 1.16 is considered for the Z+jets samples, and 1.21 for the W+jets samples. These MC samples all belong to the Summer16 production, with Moriond Premix conditions. When available, all sample extensions are used.

<table>
<thead>
<tr>
<th>Background MC simulations</th>
<th>Cross section</th>
</tr>
</thead>
<tbody>
<tr>
<td>/DYJetsToLL M-50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>4954.0 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>1012.5 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>332.8 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>101.8 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>54.8 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-150 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>6.657 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-10to50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>18610 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-10to50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>18610 pb (LO)</td>
</tr>
<tr>
<td>/DYJetsToLL M-10to50 TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>18610 pb (LO)</td>
</tr>
<tr>
<td>/TT TuneCUETP8M1 13TeV-powheg-pythia8</td>
<td>831.76 pb</td>
</tr>
<tr>
<td>/ST tW antitop 5f inclusiveDecays 13TeV-powheg-pythia8 TuneCUETP8M1</td>
<td>35.6 pb</td>
</tr>
<tr>
<td>/ST tW top 5f inclusiveDecays 13TeV-powheg-pythia8 TuneCUETP8M1</td>
<td>35.6 pb</td>
</tr>
<tr>
<td>/ST t-channel antitop 4f leptonDecays 13TeV-powheg-pythia8 TuneCUETP8M1</td>
<td>80.95 pb</td>
</tr>
<tr>
<td>/ST t-channel top 4f leptonDecays 13TeV-powheg-pythia8 TuneCUETP8M1</td>
<td>136.02 pb</td>
</tr>
<tr>
<td>/WJetsToLNu TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>50380 pb (LO)</td>
</tr>
<tr>
<td>/W1JetsToLNu TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>9644.5 pb (LO)</td>
</tr>
<tr>
<td>/W2JetsToLNu TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>3144.5 pb (LO)</td>
</tr>
<tr>
<td>/W3JetsToLNu TuneCUETP8M1 13TeV-madgraphMLM-pythia8</td>
<td>954.8 pb (LO)</td>
</tr>
<tr>
<td>/W4JetsToLNu TuneCUETP8M1 13TeV-madgraphMLM-pythia8/</td>
<td>485.6 pb (LO)</td>
</tr>
<tr>
<td>/WGToLNuG TuneCUETP8M1 13TeV-amcatnloFXFX-pythia8</td>
<td>489.0 pb</td>
</tr>
<tr>
<td>/WGstarToLNuNuMuMu 012Jets 13TeV-madgraph</td>
<td>2.793 pb</td>
</tr>
<tr>
<td>/WGstarToLNuEE 012Jets 13TeV-madgraph</td>
<td>3.526 pb</td>
</tr>
<tr>
<td>/WZToL3Nu 13TeV amcatnloFXFX madspin pythia8</td>
<td>3.05 pb</td>
</tr>
<tr>
<td>/WZToL1L1Nu2q 13TeV amcatnloFXFX madspin pythia8</td>
<td>10.71 pb</td>
</tr>
<tr>
<td>/WZTo2L2Q 13TeV amcatnloFXFX madspin pythia8</td>
<td>5.595 pb</td>
</tr>
<tr>
<td>/WWToL1L1Nu2q 13TeV amcatnloFXFX madspin pythia8</td>
<td>1.212 pb</td>
</tr>
<tr>
<td>/ZZTo2L2Q 13TeV amcatnloFXFX madspin pythia8</td>
<td>3.22 pb</td>
</tr>
<tr>
<td>/VVTo2L2Nu 13TeV amcatnloFXFX madspin pythia8</td>
<td>11.95 pb</td>
</tr>
<tr>
<td>/ZZTo4L 13TeV-amcatnloFXFX-pythia8</td>
<td>4.708 pb</td>
</tr>
<tr>
<td>/ZZTo4L 13TeV-amcatnloFXFX-pythia8</td>
<td>1.212 pb</td>
</tr>
<tr>
<td>/GluGluHToWWTo2L2Nu M125 13TeV powheg pythia8</td>
<td>48.580.2137.32582 pb</td>
</tr>
<tr>
<td>/VBFHToWWTo2L2Nu M125 13TeV powheg pythia8</td>
<td>3.7820.2137.32582 pb</td>
</tr>
<tr>
<td>/HWminusJ HToWW M125 13TeV powheg pythia8</td>
<td>0.60920.2137.32582 pb</td>
</tr>
<tr>
<td>/HWplusJ HToWW M125 13TeV powheg pythia8</td>
<td>0.8400.2137.32582 pb</td>
</tr>
<tr>
<td>/HZJ HToWW M125 13TeV powheg pythia8</td>
<td>0.88390.2137.32582 pb</td>
</tr>
<tr>
<td>/tHTomonbb M125 TuneCUETP8M2 tHtranche3 13TeV-powheg-pythia8</td>
<td>0.5085*(1-0.577)</td>
</tr>
<tr>
<td>/GluGluHToTautau M125 13TeV powheg pythia8</td>
<td>48.580.0627 pb</td>
</tr>
<tr>
<td>/VBFHToTautau M125 13TeV powheg pythia8</td>
<td>3.7820.0627 pb</td>
</tr>
<tr>
<td>/ZHToTautau M125 13TeV powheg pythia8</td>
<td>0.88390.0627 pb</td>
</tr>
<tr>
<td>/WplusHToTautau M125 13TeV powheg pythia8</td>
<td>0.8400.0627 pb</td>
</tr>
<tr>
<td>/WminusHToTautau M125 13TeV powheg pythia8</td>
<td>0.60920.0627 pb</td>
</tr>
</tbody>
</table>
**Table A1.3:** MC signal samples included in the analysis. These MC samples all belong to the Summer16 production, with Moriond Premix conditions. The efficiency of the generated-level filter is indicated. For each mass point 3 million events have been generated before filtering.

<table>
<thead>
<tr>
<th>Signal MC Simulation</th>
<th>Filter Eff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-15 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1475</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-20 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1426</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-25 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1381</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-30 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1358</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-35 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1338</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-40 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1322</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-45 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1282</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-50 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1232</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-55 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1193</td>
</tr>
<tr>
<td>SUSYGluGluToHToAA AToBB AToTauTau M-60 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1133</td>
</tr>
<tr>
<td>SUSYVBFToHToAA AToBB AToTauTau M-20 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1983</td>
</tr>
<tr>
<td>SUSYVBFToHToAA AToBB AToTauTau M-40 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1908</td>
</tr>
<tr>
<td>SUSYVBFToHToAA AToBB AToTauTau M-60 TuneCUETP8M1 13TeV madgraph pythia8</td>
<td>0.1799</td>
</tr>
</tbody>
</table>

**Table A1.4:** Trigger used in each final state

<table>
<thead>
<tr>
<th>Final States</th>
<th>Triggers</th>
</tr>
</thead>
<tbody>
<tr>
<td>eμ</td>
<td>HTL_Mu8_TrkIsoVVL_Ele23_CaloldL_TrackldL_IsoVL_v,</td>
</tr>
<tr>
<td></td>
<td>HTL_Mu23_TrkIsoVVL_Ele12_CaloldL_TrackldL_IsoVL_v,</td>
</tr>
<tr>
<td></td>
<td>HTL_Mu8_TrkIsoVVL_Ele23_CaloldL_TrackldL_IsoVL_DZ_v,</td>
</tr>
<tr>
<td></td>
<td>HTL_Mu23_TrkIsoVVL_Ele12_CaloldL_TrackldL_IsoVL_DZ_v,</td>
</tr>
<tr>
<td>μτ h</td>
<td>HLT_IsoMu22_v*, HLT_IsoTrkMu22_v*, HLT_IsoMu22_eta2p1_v*,</td>
</tr>
<tr>
<td></td>
<td>HLT_IsoTrkMu22_eta2p1_v*, HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1_v*,</td>
</tr>
<tr>
<td></td>
<td>HLT_IsoMu19_eta2p1_LooseIsoPFTau20_v*</td>
</tr>
<tr>
<td>eτ h</td>
<td>HLT_Ele25_eta2p1_WPTight_Gsf</td>
</tr>
</tbody>
</table>
APPENDIX A2

For this analysis the VBF and VH contributions are taken as equivalent in shape to the ggF contributions and rescaled by the VBF and VH to ggH ratio. The plots concerning the compatibility between ggH and VBF as well as ggH with VH signal samples for the three different final states are presented.

Figure A2.1: Ratio between the normalization of the VBF and ggH signal samples for $m_a = 20, 40, 60$ GeV, in the $e\mu$ final state. The average ratios, shown with a full red line, are determined per category (see Chapter 7 for the definition of the categories). The low-mass signal region is on the top left, the medium-mass signal region on the top right, the high-mass signal region on the bottom left, and the control region on the bottom right. The dashed lines indicate the 10% uncertainty considered for this procedure.
Figure A.2.2: Ratio between the normalization of the VBF and ggH signal samples for $m_a = 20, 40, 60$ GeV, in the $e\tau_h$ final state. The average ratios, shown with a full red line, are determined per category (see Chapter 7 for the definition of the categories). The low-mass signal region is on the top left, the medium-mass signal region on the top right, the high-mass signal region on the bottom left, and the control region on the bottom right. The dashed lines indicate the 10% uncertainty considered for this procedure.
Figure A2.3: Ratio between the normalization of the VBF and ggH signal samples for $m_a = 20, 40, 60$ GeV, in the $e\tau_h$ final state. The average ratios, shown with a full red line, are determined per category (see Chapter 7 for the definition of the categories). The low-mass signal region is on the top left, the medium-mass signal region on the top right, the high-mass signal region on the bottom left, and the control region on the bottom right. The dashed lines indicate the 10% uncertainty considered for this procedure.
Figure A2.4: Ratio between the normalization of the VH and ggH signal samples for $m_a = 40$ GeV, in the $e\mu$ (top left), $e\tau_b$ (top right), $\mu\tau_b$ (bottom) final states. The normalization of the VH samples is measured only in the ditau mass peak, neglecting tails due to selecting a lepton from V decays. The average ratios over the three signal regions, shown with a full red line, are determined per final state. The dashed lines indicate the 10% uncertainty considered for this procedure.
Figure A2.5: Shapes (normalized to unity) of the signal produced in ggH, WH, and ZH productions, for the e_\tau_\tau final state in a category with medium invariant mass (80 – 100 GeV) between the leading b jet and the tau candidates, and in a category with high-mass (100 – 120 GeV). The plots for invariant mass lower than 80 GeV and greater than 110 GeV can be seen in section 6.2.
APPENDIX A3

In this appendix are presented all the efficiency plots used for the three different final states. The Efficiency plots are used for final state objects identification as well as for trigger identification. For these results a Tag and Probe methodology was used by applied MC simulated samples different for each particle identification. The scale factors was driven from the comparison of MC simulations with the initial dataset taken in 2016 at center of the mass energy of 13TeV.

The $e\mu$ Final State

Electron identification efficiency measurement in data recorded with single electron trigger with the tag and probe method method using $Z \rightarrow ee$. The number of probe electrons obtain from a fit to the di-electron invariant mass and signal modeled by $Z/\gamma^* \rightarrow ee$. Scale factors correcting the identification, isolation and trigger efficiencies between data nad MC are applied to all samples. An interface and instructions how those scale factors was applied can be found in this page [1].
Efficiencies are computed by means of a tag-and-probe method exploiting the $Z \rightarrow \mu^+\mu^-$ resonance. Data correspond to an integrated luminosity of 16.3 fb$^{-1}$ of 2016 to pp collision at 13 TeV. The events are collected using single muon trigger. The MC samples was the DY + jets with LO accuracy. The tag-and-probe method used to extract the data/MC scale factors for the differences in identification and isolation. An interface and instructions how those scale factors was applied can by found in this page [2].
The $\mu \tau_b$ final state

The efficiencies have been measured from $Z \rightarrow \mu\mu$ events using the Tag-and-Probe method, where the tag muon is matched to the isolated trigger muon object. Data correspond to an integrated luminosity of 16.3 fb$^{-1}$ of 2016 pp collision at 13 TeV. The events are collected using single muon trigger. The MC samples was the DY + jets with LO accuracy. The tag-and-probe method used to extract the data/MC scale factors for the differences in identification and isolation. An interface and instructions how those scale factors was applied can by found in this page [3].

![Efficiency plots](image1.png)

![Efficiency plots](image2.png)

![Efficiency plots](image3.png)
The $\tau$ final state

The efficiencies have been measured from $Z \rightarrow ee$ events using the Tag-and-Probe method, where the tag muon is matched to the isolated trigger muon object. Data correspond to an integrated luminosity of $16.3\text{fb}^{-1}$ of 2016 pp collision at 13TeV. The events are collected using single muon trigger. The MC samples was the DY+ jets with LO accuracy. The tag-and-probe method used to extract the data/MC scale factors for the differences in identification and isolation. An interface and instructions how those scale factors was applied can by found in this page [1].
APPENDIX A4

IMPACTS

The impacts are computed for the background-only scenario (impacts in the S+B scenario are very similar). The fit is performed for the three final states, including the four categories for each final state. The dominant impacts are related to bin-by-bin uncertainties of backgrounds in the poorly populated signal regions. Some of the uncertainties get constrained by the fit:

- **Fakesnorm**: The yield uncertainty for the reducible background is 10%. This number accounts for limited non-closure in the same-sign region, and for potential differences in the fake rates for $Z \rightarrow \mu\mu$ events (where the fake rates are evaluated), and W+jets and QCD multijet events (dominant in the signal region). This uncertainty is expected to be constrained in the category with large invariant mass between the leptons and the leading b jet.

- **DY norm**: The yield uncertainty for the DY background is 7%. It is large because the process is $Z \rightarrow \ell\ell + >= 1$ b jet. It gets constrained to about 6% by the fit to the high mass categories.

- **CMS scale jet**: Uncertainty in the jet energy scale, propagated to the MET.

- **CMS scale met unclustered**: The unclustered energy related to the MET energy scale. It is always constrained in $Z \rightarrow \tau\tau$ events (SM HTT, LFV Higgs analyses) because the uncertainties are evaluated in $t\bar{t}$ events, which have higher real MET.

- **CMS eff tau**: Uncertainty related to the efficiency of real taus, equal to 5%. It is expected to be constrained because the TAU POG is conservative (accounts also for a tracking uncertainty of 3.9%).