

<sup>1</sup> SEARCH FOR EXOTIC HIGGS DECAYS TO LIGHT  
<sup>2</sup> NEUTRAL SCALARS IN FINAL STATES WITH  
<sup>3</sup> BOTTOM QUARKS AND TAU LEPTONS

<sup>4</sup> KA YU STEPHANIE KWAN

<sup>5</sup> A DISSERTATION  
<sup>6</sup> PRESENTED TO THE FACULTY  
<sup>7</sup> OF PRINCETON UNIVERSITY  
<sup>8</sup> IN CANDIDACY FOR THE DEGREE  
<sup>9</sup> OF DOCTOR OF PHILOSOPHY

<sup>10</sup> NOT YET RECOMMENDED FOR ACCEPTANCE  
<sup>11</sup> BY THE DEPARTMENT OF PHYSICS  
<sup>12</sup> ADVISER: ISOBEL OJALVO

<sup>13</sup> MAY 2024

<sup>14</sup>

© Copyright by Ka Yu Stephanie Kwan, 2024.

<sup>15</sup>

All Rights Reserved

## Abstract

With the discovery of the Higgs boson with mass 125 GeV at the LHC in 2012, the Compact Muon Solenoid (CMS) experimental physics program has evolved to include the precise characterization of the Standard Model 125 GeV Higgs boson and searches for evidence of additional Higgs particles in an extended Higgs sector. Many properties of a potential extended Higgs sector remain unconstrained by current measurements, making direct searches of exotic Higgs decays a powerful probe of new physics.

The decay of the 125 GeV Higgs boson into two light neutral scalar particles ( $h \rightarrow aa$ ) is allowed in extensions of the Standard Model, such as Two Higgs Doublet Models extended with a scalar singlet (2HDM+S). We present a search at CMS for exotic decays of the 125 GeV Higgs boson to two light neutral scalars, which decay to two bottom quarks and two tau leptons ( $h \rightarrow aa \rightarrow bb\tau\tau$ ). This analysis is combined with a similar search in a different final state where the light scalars decay to two bottom quarks and two muons. The results from the  $bb\tau\tau$  analysis and the combined analyses are interpreted in 2HDM+S scenarios. In a different extension of the Standard Model, the Two Real Singlet Model (TRSM), the 125 GeV Higgs boson can decay to two light scalars with unequal mass ( $h \rightarrow a_1a_2$ ). This decay has not been searched for to date at CMS. We present ongoing work on a search for  $h \rightarrow a_1a_2$ , where the  $a_2$  decays into two  $a_1$ , resulting in four bottom quarks and two tau leptons in the final state, in the  $\mu\tau_h$  channel of the  $\tau\tau$  decay. Such searches for rare processes will directly benefit from the increased datasets that will be generated by the High-Luminosity LHC (HL-LHC), which is scheduled to increase the LHC's number of simultaneous proton-proton collisions by a factor of five to seven. To contribute to the performance of the CMS Level-1 Trigger in selecting collisions with interesting physics, this thesis presents an upgraded algorithm for reconstructing electrons and photons in the barrel calorimeter, which will use information with higher spatial granularity to distinguish genuine electrons and photons from background.

<sup>43</sup>

## Acknowledgements

<sup>44</sup> Placeholder acknowledgements.



# <sup>46</sup> Contents

<sup>47</sup>	Abstract . . . . .	iii
<sup>48</sup>	Acknowledgements . . . . .	iv
<sup>49</sup>	List of Tables . . . . .	xi
<sup>50</sup>	List of Figures . . . . .	xv
<sup>51</sup>	<b>1 Introduction</b>	<b>1</b>
<sup>52</sup>	1.1 History of the Standard Model . . . . .	1
<sup>53</sup>	1.2 The Standard Model as a gauge theory . . . . .	3
<sup>54</sup>	1.3 The Higgs mechanism . . . . .	6
<sup>55</sup>	1.4 Two-Higgs Doublet Models . . . . .	8
<sup>56</sup>	1.5 Two Real Singlet Model . . . . .	11
<sup>57</sup>	<b>2 The Large Hadron Collider and the CMS Experiment</b>	<b>15</b>
<sup>58</sup>	2.1 The Large Hadron Collider . . . . .	16
<sup>59</sup>	2.2 Luminosity and pile-up . . . . .	17
<sup>60</sup>	2.3 The High-Luminosity LHC . . . . .	20
<sup>61</sup>	2.4 The CMS detector . . . . .	21
<sup>62</sup>	2.5 Sub-detectors of CMS and data processing . . . . .	23
<sup>63</sup>	2.5.1 Inner tracking system . . . . .	23
<sup>64</sup>	2.5.2 ECAL . . . . .	24
<sup>65</sup>	2.5.3 HCAL . . . . .	25

66	2.5.4 Muon detectors . . . . .	27
67	2.5.5 The Level-1 Trigger . . . . .	28
68	2.5.6 The High-Level Trigger . . . . .	32
69	2.5.7 Particle reconstruction . . . . .	33
70	2.5.8 Data storage and computational infrastructure . . . . .	34
71	<b>3 The Phase-2 Upgrade of CMS</b>	<b>35</b>
72	3.1 The High-Luminosity LHC . . . . .	35
73	3.2 The Phase-2 Level-1 Trigger . . . . .	36
74	3.3 Standalone barrel calorimeter electron/photon reconstruction . . . . .	39
75	3.3.1 Electron/photon standalone barrel procedure . . . . .	39
76	3.3.2 Electron/photon standalone barrel results . . . . .	44
77	<b>4 Datasets and Monte Carlo samples</b>	<b>48</b>
78	4.1 Datasets used . . . . .	48
79	4.2 Monte Carlo samples . . . . .	49
80	4.3 Embedded samples . . . . .	50
81	<b>5 Object reconstruction and corrections applied</b>	<b>53</b>
82	5.1 Object reconstruction . . . . .	54
83	5.1.1 Taus . . . . .	54
84	5.1.2 Muons . . . . .	57
85	5.1.3 Electrons . . . . .	58
86	5.1.4 Jets . . . . .	60
87	5.1.5 B-flavored jets . . . . .	61
88	5.2 Reconstruction of the di-tau mass . . . . .	62
89	5.2.1 Original SVFit “standalone”: maximum likelihood . . . . .	63
90	5.2.2 “Classic SVFit” with matrix element . . . . .	64
91	5.2.3 FastMTT: optimized SVFit . . . . .	64

92	5.3 Corrections applied to simulation . . . . .	65
93	5.3.1 Tau energy scale . . . . .	66
94	5.3.2 Muon energy scale . . . . .	67
95	5.3.3 Electron energy scale . . . . .	67
96	5.3.4 $\tau_h$ identification efficiency . . . . .	67
97	5.3.5 Trigger efficiencies definition . . . . .	68
98	5.3.6 Tau trigger efficiencies . . . . .	69
99	5.3.7 Single muon trigger efficiencies . . . . .	70
100	5.3.8 Single electron trigger efficiencies . . . . .	71
101	5.3.9 $e\mu$ cross-trigger efficiencies . . . . .	72
102	5.3.10 Electrons and muons faking $\tau_h$ : energy scales . . . . .	73
103	5.3.11 Electrons and muons faking $\tau_h$ : misidentification efficiencies .	74
104	5.3.12 Electron ID and tracking efficiency . . . . .	75
105	5.3.13 Muon ID, isolation, and tracking efficiencies . . . . .	75
106	5.3.14 Recoil corrections . . . . .	77
107	5.3.15 Drell-Yan corrections . . . . .	78
108	5.3.16 Pile-up reweighing . . . . .	78
109	5.3.17 Pre-firing corrections . . . . .	79
110	5.3.18 Top $p_T$ spectrum reweighing . . . . .	79
111	5.3.19 B-tagging efficiency . . . . .	79
112	5.3.20 Jet energy resolution and jet energy smearing . . . . .	80
113	<b>6 Event selection</b>	<b>81</b>
114	6.1 General procedure for all channels . . . . .	81
115	6.2 Event selection in the $\mu\tau_h$ channel . . . . .	83
116	6.3 Event selection in the $e\tau_h$ channel . . . . .	85
117	6.4 Event selection in the $e\mu$ channel . . . . .	87
118	6.5 Extra lepton vetoes in all channels . . . . .	88

<sup>119</sup>	<b>7 Background estimation</b>	<b>91</b>
<sup>120</sup>	7.1 Z+jets . . . . .	91
<sup>121</sup>	7.2 W+jets . . . . .	92
<sup>122</sup>	7.3 $t\bar{t}$ + jets . . . . .	92
<sup>123</sup>	7.4 Single top . . . . .	93
<sup>124</sup>	7.5 Diboson . . . . .	93
<sup>125</sup>	7.6 Standard Model Higgs . . . . .	93
<sup>126</sup>	7.7 Jet faking $\tau_h$ . . . . .	94
<sup>127</sup>	7.8 QCD multijet background . . . . .	95
<sup>128</sup>	<b>8 Systematic uncertainties</b>	<b>97</b>
<sup>129</sup>	8.1 Uncertainties in the lepton energy scales . . . . .	98
<sup>130</sup>	8.2 Uncertainties from other lepton corrections . . . . .	99
<sup>131</sup>	8.3 Uncertainties from jet energy scale and resolution . . . . .	100
<sup>132</sup>	8.4 Uncertainties from b-tagging scale factors . . . . .	101
<sup>133</sup>	8.5 Uncertainties from MET . . . . .	101
<sup>134</sup>	8.6 Uncertainties associated with samples used . . . . .	102
<sup>135</sup>	8.7 Other uncertainties . . . . .	103
<sup>136</sup>	8.8 Pulls and impacts . . . . .	103
<sup>137</sup>	<b>9 Event categorization and signal extraction</b>	<b>105</b>
<sup>138</sup>	9.1 B-tag jet multiplicity . . . . .	105
<sup>139</sup>	9.2 DNN-based event categorization . . . . .	106
<sup>140</sup>	9.3 Methodology for signal extraction . . . . .	110
<sup>141</sup>	9.3.1 Model building and parameter estimation . . . . .	110
<sup>142</sup>	9.3.2 Hypothesis testing . . . . .	111
<sup>143</sup>	9.3.3 Confidence intervals . . . . .	113
<sup>144</sup>	9.3.4 Profile likelihood ratio . . . . .	115

<sup>145</sup>	9.3.5 Modified frequentist method: $CL_S$	116
<sup>146</sup>	<b>10 Results</b>	<b>117</b>
<sup>147</sup>	10.1 Results from $bb\tau\tau$	117
<sup>148</sup>	10.2 Combination with $bb\mu\mu$ final state	119
<sup>149</sup>	<b>11 Asymmetric exotic Higgs decays</b>	<b>130</b>
<sup>150</sup>	11.1 Signal masses	130
<sup>151</sup>	11.2 Cascade scenario signal studies	131
<sup>152</sup>	11.3 Control plots for $\mu\tau_h$ channel	133
<sup>153</sup>	<b>12 Conclusion and outlook</b>	<b>141</b>
<sup>154</sup>	<b>A Samples used</b>	<b>143</b>

# <sup>155</sup> List of Tables

<sup>156</sup>	4.1	Expected event composition after selecting two muons in the embedded technique, before additional cuts (i.e. inclusive), and after adding a requirement on the di-muon mass $m_{\mu\mu} > 70 \text{ GeV}$ , or a requirement on the number of b-tag jets in the event. . . . .	52
<sup>160</sup>	5.1	Energy scales applied to genuine hadronic tau decays $\tau_h$ by data-taking year/era and decay mode, along with systematic errors. . . . .	66
<sup>161</sup>	5.2	Energy scales and systematic errors applied to genuine muons. . . . .	67
<sup>162</sup>	5.3	Energy scales and systematic errors applied to electrons in embedded samples by data-taking year/era. . . . .	67
<sup>164</sup>	5.4	Tau ID efficiency for the DeepTau vs. jet medium working point, with central, up, and down values for 2018, binned in the tau $p_T$ . . . . .	68
<sup>165</sup>	5.5	Energy scales and up/down systematic uncertainties applied to electrons misidentified as hadronic taus. . . . .	73
<sup>166</sup>	5.6	Tau mis-identification efficiencies (in %) for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018. . . . .	74
<sup>169</sup>	5.7	Tau mis-identification efficiencies (in %) for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018. . . . .	74
<sup>170</sup>			
<sup>171</sup>			
<sup>172</sup>			

173	6.1	Trigger thresholds used for the leptons in the $bb\mu\mu$ analysis and the 174 $bb\tau\tau$ analysis (the focus of this work). The thresholds for the three $bb\tau\tau$ 175           channels ( $e\mu$ , $e\tau_h$ , and $\mu\tau_h$ ) are listed separately, with some channels 176           and years taking the logical OR of two triggers with different thresholds. . . . .	83
177	6.2	Summary of requirements applied to the leptons in the $bb\mu\mu$ analysis 178           and the $bb\tau\tau$ analysis (the focus of this work). $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ 179           is a measure of spatial separation. Relative isolation is defined in 180           Eqn. 5.2Muonsequation.5.1.2 and Section 5.1.2Muonssubsection.5.1.2. 181           The b-tag jets are required to pass the listed DeepFlavour working 182           points (WP), which are described in Section 5.1.5B-flavored jetssub- 183           section.5.1.5. In the $bb\tau\tau$ analysis, the required $ \eta $ of the hadronic 184           taus are listed for the single and cross triggers respectively. The $bb\mu\mu$ 185           analysis requires two b-tag jets in all events, while the $bb\tau\tau$ analysis 186           only requires one. . . . .	84
187	6.3	High-Level Trigger (HLT) paths used to select data and simulation 188           events in 2016 for the three $\tau\tau$ channels. . . . .	88
189	6.4	High-Level Trigger (HLT) paths used to select data and simulation 190           events in 2017 for the three $\tau\tau$ channels. . . . .	89
191	6.5	High-Level Trigger (HLT) paths used to select data and simulation 192           events in 2018 for the three $\tau\tau$ channels. In 2018 a HLT trigger path 193           using the hadron plus strips (HPS) tau reconstruction algorithm be- 194           came available. . . . .	90
195	9.1	Event categorization based on DNN scores for events with exactly 1 196           b-tag jet (1bNN), for the three $\tau\tau$ channels and three eras. . . . .	109
197	9.2	Event categorization based on DNN scores for events with 2 b-tag jets 198           (2bNN), for the three $\tau\tau$ channels and three eras. . . . .	109

199	A.1 Datasets used in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis for the 2016 era. . . . .	144
200	A.2 Datasets used in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis for the 2017 era. . . . .	145
201	A.3 Datasets used in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis for the 2018 eras. . . . .	145
202	A.4 Embedded samples used in the analysis for the 2016 era. . . . .	145
203	A.5 Embedded samples used in the analysis for the 2017 era. . . . .	146
204	A.6 Embedded samples used in the analysis for the 2018 era. . . . .	146
205	A.7 Background MC samples used in the analysis for the 2016 era. Samples marked with a $\dagger$ are generated with the powhegV2-madspin-pythia8 tag.	147
206		
207	A.8 Background MC samples used in the analysis for the 2017 era. All samples use powheg, except the DYJets and WJets samples, which use madgraphMLM. Samples marked with a $\dagger$ , $\ddagger$ , or $\ddagger\dagger$ were produced with Powheg2 and Pythia8, and Madspin, JHUGenV714, or jhugen724 respectively. . . . .	148
208		
209		
210		
211		
212	A.9 Background Monte Carlo samples used in the analysis for the 2018 era. All samples listed are generated for 13 TeV collisions and use pythia8. All samples use powheg, except the DYJets and WJets samples, which use madgraphMLM. Samples marked with a $\dagger$ , $\ddagger$ , or $\ddagger\dagger$ , were pro- duced with Powheg and Pythia8, and Madspin, JHUGenV714, and Jhugen724 respectively. . . . .	149
213		
214		
215		
216		
217		
218	A.10 Signal samples used in the analysis for the 2016 era. All belong to the RunIISummer16MiniAODv3 campaign and are produced with Mad- graph and Pythia8. The second column is the number of events after the generator-level filter is applied, and the third column is the filter efficiency (percentage of all events that pass the generator-level filter). . . . .	150
219		
220		
221		
222		

223 A.11 Signal samples used in the analysis for the 2017 era. All belong to  
224 the RunIIFall17MiniAODv2 campaign and are produced with Mad-  
225 graph and Pythia8. The second column is the number of events after  
226 the generator-level filter is applied, and the third column is the filter  
227 efficiency (percentage of all events that pass the generator-level filter). 150

228 A.12 Signal samples used in the analysis for the 2018 era. All belong to  
229 the RunIIAutumn18MiniAOD campaign and are produced with Mad-  
230 graph and Pythia8. The second column is the number of events after  
231 the generator-level filter is applied, and the third column is the filter  
232 efficiency (percentage of all events that pass the generator-level filter). 151

# <sup>233</sup> List of Figures

<sup>234</sup> 1.1	Table of Standard Model particles showing the grouping of the fermions into three generations of matter and the bosons, responsible for carrying the three fundamental forces in the Standard Model. The masses, charges, and spins of the particles are shown. The antimatter counter- parts of the fermions are not shown. The possible interactions between the fermions and gauge bosons are highlighted. . . . .	3
<sup>239</sup> 1.2	An illustration of the Higgs potential. . . . .	8
<sup>240</sup> 1.3	Branching ratios of a singlet-like pseudoscalar in Type II 2HDM+S for $\tan \beta = 0.5$ (left) and $\tan \beta = 5$ (right). . . . .	11
<sup>241</sup> 1.4	Benchmark plane BP1 for benchmark scenario 1, for the decay signa- ture $h_{125} \rightarrow h_1 h_2$ with $h_{125} \equiv h_3$ , defined in the $(M_1, M_2)$ plane. . . .	14
<sup>243</sup> 2.1	Aerial view of the Large Hadron Collider (LHC). . . . .	18
<sup>244</sup> 2.2	Distribution of the mean number of inelastic collisions per bunch cross- ing (pile-up) in data, for proton-proton collisions in 2016-2018 . . . .	20
<sup>245</sup> 2.3	Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector.	22
<sup>246</sup> 2.4	Cross section of the current Phase-1 CMS tracker. . . . .	24
<sup>247</sup> 2.5	Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters.	26

253	2.6	Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels. . . . .	28
254	2.7	Dataflow for the Phase-1 Level-1 Trigger. . . . .	29
255	2.8	Schematic of the calorimeter trigger after Long Shutdown 2. The Layer-1 calorimeter trigger is implemented in CTP7 cards, which send time-multiplexed outputs to the Layer-2 MP7 cards. The Layer-2 cards handle the data in a round-robin style and the outputs are de- multiplexed, producing one output data stream to the Global Trigger.	31
256			
257			
258			
259			
260			
261	3.1	Functional diagram of the CMS L1 Phase-2 upgraded trigger design. .	37
262	3.2	Summary of the links between the trigger primitives, the trigger ob- jects, the Level-1 algorithms, and the physics channels in the Phase-2 menu. . . . .	40
263			
264			
265	3.3	Schematic of the geometry of the Phase-2 ECAL barrel in the Regional Calorimeter Trigger (RCT), showing the division of the barrel region into 36 Regional Calorimeter Trigger (RCT) cards ( <i>red</i> ). Each card spans $17 \times 4$ towers in $\eta \times \phi$ ( <i>green</i> ), and each tower is $5 \times 5$ in single crystals in $\eta \times \phi$ . Towers in the overlap region ( <i>shaded yellow</i> ) are read out to both the barrel and endcap. . . . .	41
266			
267			
268			
269			
270			
271	3.4	Schematic of two example RCT cards in the negative eta ( <i>top left</i> ) and positive eta ( <i>bottom left</i> ) regions of the ECAL barrel. Each RCT card is divided into six regions: five regions are of size $3 \times 4$ towers in $\eta \times \phi$ ( <i>top right</i> ), and a sixth smaller overlap region of size $2 \times 4$ towers ( <i>bottom right</i> ). Each tower is $5 \times 5$ ( $\eta \times \phi$ ) in crystals. . . . .	42
272			
273			
274			
275			





323	5.3 Hadronic tau leg efficiency of the cross-triggers for $\mu\tau_h$ ( <i>left</i> ) and $e\tau_h$ ( <i>right</i> ) triggers as a function of offline tau $p_T$ for 2016, 2017, and 2018.	70
324		
325	5.4 Trigger efficiencies in data ( <i>top panels</i> ) and ratio of efficiencies af-	
326	ter/before a HLT muon reconstruction update ( <i>bottom panels</i> ) for the	
327	muon in the isolated single muon trigger with threshold $p_T > 24 \text{ GeV}$	
328	in the data-taking year 2018, as functions of the muon $p_T$ ( <i>left</i> ) and	
329	muon $ \eta $ ( <i>right</i> ). . . . .	71
330		
331	5.5 Trigger efficiencies in data and the data/MC ratio for the electron in	
332	the single electron trigger with threshold $p_T > 32 \text{ GeV}$ in the data-	
333	taking year 2018, as functions of the electron $p_T$ ( <i>left</i> ) and electron $ \eta $ ( <i>right</i> ). . . . .	72
334		
335	5.6 Efficiencies of the electron leg vs. $p_T$ ( <i>left</i> ) and the muon log vs. $\eta$ ( <i>right</i> ), for the HLT path with online thresholds of 12 GeV for the	
336	electron and 23 GeV for the muon, with the data-taking years 2016	
337	through 2018 overlaid. . . . .	73
338		
339	5.7 Efficiencies in data ( <i>top panels</i> ) and the ratio of efficiencies in data/MC ( <i>bottom panels</i> ), for the electron multivariate analysis (MVA) identifi-	
340	cation ( <i>left</i> ) and for the Gaussian-sum filter (GSF) tracking ( <i>right</i> ). . .	75
341		
342	5.8 Muon identification efficiencies in 2015 data and MC as a function of	
343	the muon $p_T$ for the loose ID ( <i>left</i> ) and tight ID ( <i>right</i> ) working points.	76
344		
345	5.9 Muon isolation efficiencies in Run-2 data as a function of the muon $p_T$ ( <i>left</i> ) and $ \eta $ ( <i>right</i> ). . . . .	77
346		
347	5.10 Muon tracking efficiencies as a function of $ \eta $ for standalone muons in	
	Run-2 data ( <i>black</i> ) and Drell-Yan ( <i>blue</i> ) MC simulation. . . . .	78
348	7.1 Leading-order Feynman diagrams of Higgs production. . . . .	94

348	8.1 Top sixty pulls and impacts for the combination of all channels and	
349	years. . . . .	104
350	9.1 Schematic of the Neyman construction for confidence intervals. . . . .	114
351	10.1 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $\mu\tau_h$ chan-	
352	nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions. . . . .	121
353	10.2 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\tau_h$ chan-	
354	nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions. . . . .	122
355	10.3 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\mu$ channel.	123
356	10.4 Observed 95% CL exclusion limits ( <i>black, solid lines</i> ) and expected 95%	
357	CL and 68% CL limits ( <i>shaded yellow and green</i> ) on the branching	
358	fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard	
359	Model production for the 125 GeV Higgs ( $h$ ). Limits are shown for the	
360	$\mu\tau_h$ channel ( <i>top left</i> ), the $e\tau_h$ channel ( <i>top right</i> ), and the $e\mu$ channel	
361	( <i>bottom left</i> ), and lastly the combination of all three channels ( <i>bottom</i>	
362	<i>right</i> ) The dataset corresponds to $138 \text{ fb}^{-1}$ of data collected in the	
363	years 2016-2018 at a center-of-mass energy 13 TeV. . . . .	124
364	10.5 Observed 95% CL upper limits on $B(h \rightarrow aa)$ in %, for the $bb\tau\tau$ final	
365	state ( <i>left</i> ) and $bb\mu\mu$ final state ( <i>right</i> ) using the full Run 2 integrated	
366	luminosity of $138 \text{ fb}^{-1}$ in 2HDM+S type I ( <i>blue</i> ), type II with $\tan\beta =$	
367	$2.0$ ( <i>orange dashed</i> ), type III with $\tan\beta = 2.0$ ( <i>dotted green</i> ), and type	
368	IV with $\tan\beta = 0.6$ ( <i>red dashed</i> ). . . . .	125

369	10.6 Observed 95% CL upper limits on the branching fraction of the 125	
370	GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$ , in percentages,	
371	as a function of the pseudoscalar mass $m_a$ , in 2HDM+S type I ( <i>blue</i> ),	
372	type II with $\tan\beta = 2.0$ ( <i>orange dashed</i> ), type III with $\tan\beta = 2.0$	
373	( <i>dotted green</i> ), and type IV with $\tan\beta = 0.6$ ( <i>red dashed</i> ), for the	
374	combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated	
375	luminosity of $138 \text{ fb}^{-1}$ . . . . .	126
376	10.7 Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination	
377	of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity	
378	of $138 \text{ fb}^{-1}$ for Type II ( <i>left</i> ), Type III ( <i>middle</i> ), and Type IV ( <i>right</i> )	
379	2HDM+S in the $\tan\beta$ vs. $m_a$ phase space. . . . .	127
380	10.8 Summary plot of current observed and expected 95% CL limits on the	
381	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
382	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
383	$B(h \rightarrow aa)$ , in the 2HDM+S type I scenario, obtained at CMS with	
384	data collected at 13 TeV. . . . .	127
385	10.9 Summary plot of current observed and expected 95% CL limits on the	
386	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
387	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
388	$B(h \rightarrow aa)$ , in the 2HDM+S type II scenario with $\tan\beta = 2.0$ , ob-	
389	tained at CMS with data collected at 13 TeV. . . . .	128
390	10.10 Summary plot of current observed and expected 95% CL limits on the	
391	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
392	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
393	$B(h \rightarrow aa)$ , in the 2HDM+S type III scenario with $\tan\beta = 2.0$ , ob-	
394	tained at CMS with data collected at 13 TeV. . . . .	129

395	11.1 Generator-level b-flavor jet transverse momenta $p_T$ , for $h \rightarrow a_1 a_2$ cas-	
396	cade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) =$	
397	$(100, 15)$ GeV ( <i>left</i> ) and $(40, 20)$ GeV ( <i>right</i> ). In each plot the generator-	
398	level $p_T$ of the leading ( <i>black</i> ), sub-leading ( <i>red</i> ), third ( <i>blue</i> ), and	
399	fourth ( <i>light green</i> ) are overlaid. . . . .	132
400	11.2 Distributions (arbitrary units) of transverse momentum $p_T$ resolution	
401	and $\Delta R$ between the two closest generator-level $b$ jets, treated as one	
402	object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow$	
403	$a_1 a_2$ mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV ( <i>top left, top right</i> )	
404	and $(40, 20)$ GeV ( <i>bottom left, bottom right</i> ) in the ggH production of	
405	the 125 GeV $h$ . In the $(40, 20)$ GeV mass point, the longer $p_T$ resolution	
406	tail ( <i>bottom left</i> ) indicates that the reconstructed jet underestimates	
407	the generator b-flavor jets' energy, and the significant fraction of events	
408	with larger $\Delta R$ values ( <i>bottom right</i> ) indicate worse matching. . . . .	133
409	11.3 Kinematic properties of the leading muon in the $\mu\tau_h$ channel using 2018	
410	samples: transverse momentum $p_T$ ( <i>top</i> ), $\eta$ ( <i>middle</i> ), and $\phi$ ( <i>bottom</i> ).	
411	The errors shown in the figures include statistical errors and only sev-	
412	eral of the full set of systematic errors (only those associated with the	
413	lepton energy scales and $\tau_h$ identification efficiency). . . . .	135
414	11.4 Kinematic properties of the leading hadronic tau ( $\tau_h$ ) in the $\mu\tau_h$ channel	
415	using 2018 samples: transverse momentum $p_T$ ( <i>top</i> ), $\eta$ ( <i>middle</i> ), and $\phi$	
416	( <i>bottom</i> ). The errors shown in the figures include statistical errors and	
417	only several of the full set of systematic errors (only those associated	
418	with the lepton energy scales and $\tau_h$ identification efficiency). . . . .	136

419	11.5 The visible di-tau mass $m_{\text{vis}}$ ( <i>top</i> ) and visible di-tau transverse momen-	
420	tum $p_{T,\text{vis}}$ ( <i>bottom</i> ), computed from the sum of the visible 4-momenta	
421	of the muon and $\tau_h$ in the $\mu\tau_h$ channel using 2018 samples. The er-	
422	rors shown in the figures include statistical errors and only several of	
423	the full set of systematic errors (only those associated with the lepton	
424	energy scales and $\tau_h$ identification efficiency). . . . .	137
425	11.6 Kinematic properties of the leading b-tag jet in the $\mu\tau_h$ final state	
426	using 2018 samples: transverse momentum $p_T$ ( <i>top</i> ), $\eta$ ( <i>middle</i> ), $\phi$	
427	( <i>bottom</i> ). The errors shown in the figures include statistical errors and	
428	only several of the full set of systematic errors (only those associated	
429	with the lepton energy scales and $\tau_h$ identification efficiency). . . . .	138
430	11.7 Kinematic properties of the sub-leading b-tag jet in the $\mu\tau_h$ final state	
431	using 2018 samples: transverse momentum $p_T$ ( <i>top</i> ), $\eta$ ( <i>middle</i> ), $\phi$	
432	( <i>bottom</i> ). The errors shown in the figures include statistical errors and	
433	only several of the full set of systematic errors (only those associated	
434	with the lepton energy scales and $\tau_h$ identification efficiency). . . . .	139
435	11.8 Missing transverse energy magnitude ( <i>top</i> ) and azimuthal direction	
436	( <i>bottom</i> ) in the $\mu\tau_h$ final state using 2018 samples. The errors shown	
437	in the figures include statistical errors and only several of the full set of	
438	systematic errors (only those associated with the lepton energy scales	
439	and $\tau_h$ identification efficiency). . . . .	140

# <sup>440</sup> Chapter 1

## <sup>441</sup> Introduction

<sup>442</sup> The Standard Model is the current prevailing theoretical framework that encompasses  
<sup>443</sup> all known elementary particles to date and describes their interactions, yet falls short  
<sup>444</sup> of describing open problems in physics. Here, we describe the history of the Standard  
<sup>445</sup> Model and its particle content (Section 1.1), and provide a mathematical motivation of  
<sup>446</sup> the SM as a gauge theory (Section 1.2). We introduce the Higgs mechanism (Section  
<sup>447</sup> 1.3), and outline two groups of theoretical extensions to the Standard Model that  
<sup>448</sup> feature extended Higgs sectors (Sections 1.4 and 1.5).

### <sup>449</sup> 1.1 History of the Standard Model

<sup>450</sup> The building blocks of our modern-day understanding of particle physics were estab-  
<sup>451</sup> lished over the course of many decades by experimental discoveries and theoretical  
<sup>452</sup> advances, culminating in the development of a theoretical framework known as the  
<sup>453</sup> Standard Model (SM). In the 1880s, the electron was the first subatomic particle to  
<sup>454</sup> be identified, through measurements of particles produced by ionizing gas. By the  
<sup>455</sup> 1930s, atoms were known to consist mostly of empty space, with protons and neutrons  
<sup>456</sup> concentrated at the center and orbited by electrons. Spurred by advances in parti-  
<sup>457</sup> cle accelerator technology, the experimental discoveries of the positron, the muon,

and the pion, painted an increasingly complicated picture of particle physics that could not be described solely with atomic physics [1]. Quantum field theory (QFT) began to be developed in the early 20th century as an extension of the conceptual framework of quantum mechanics to electromagnetic fields [2]. In 1927, Dirac coined the name quantum electrodynamics (QED), which was the first part of QFT that was developed. QED quantized the electromagnetic field and supplied a relativistic theory of the electron, and could be applied to concrete physical processes such as the scattering of high-frequency photons by free electrons (Compton scattering), and the production of electron-positron pairs by photons [2]. In the 1940s the QED-only picture was challenged by the realization that the four-fermion theory of weak interactions had infinities at higher orders of perturbation theory which could not be removed via the technique of renormalization [3], i.e. shifting divergences into parts of the theory that do not influence empirical measurements [2].

In the 1950s and 1960s, QFT was extended to describe not only the electromagnetic force, but also the strong and weak force, with the final picture forming the Standard Model. This took place in the development and maturation of three principles: the quark model, the idea of gauge (or local) symmetry, and spontaneously broken symmetry [3]. In the fully fledged QFT, Lagrangians had to be formed that contained new classes of quantum fields, or particles [2].

The particle content of the Standard Model is summarized in Fig. 1.1. Particles are grouped into fermions, which comprise all known matter, and bosons, which mediate the interactions between particles. Fermions consist of quarks and leptons, and are grouped into three generations. For example, the electron belongs to the first generation of leptons. The second and third generation counterparts of the electron are the muon and the tau lepton, and are over 200 and 30,000 times heavier than the electron respectively. The quarks are also organized into three generations (top and bottom quarks, charm and strange quarks, and up and down quarks), and

485 carry fractional electric charge. Bosons are force carriers; the interaction of fermions  
486 with bosons corresponds to fundamental forces. The Standard Model describes the  
487 electromagnetic force, the strong nuclear force, and the weak nuclear force. Through  
488 the strong force, quarks can form composite particles known as hadrons. Familiar  
489 examples of hadrons are the protons and neutrons in the nucleus of an atom.

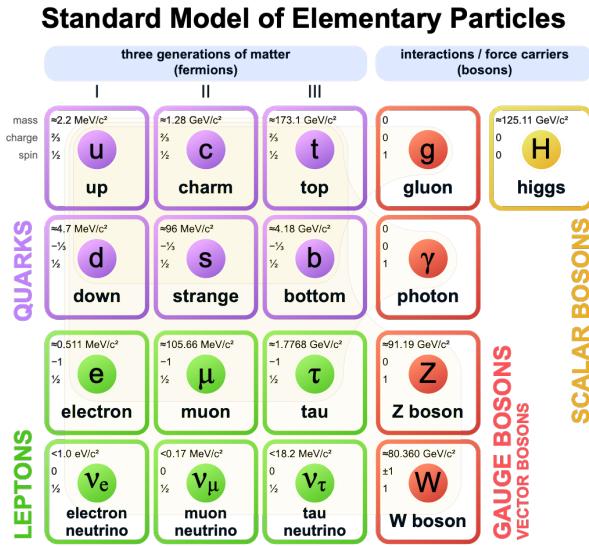


Figure 1.1: Table of Standard Model particles showing the grouping of the fermions into three generations of matter and the bosons, responsible for carrying the three fundamental forces in the Standard Model. The masses, charges, and spins of the particles are shown. The antimatter counterparts of the fermions are not shown. The possible interactions between the fermions and gauge bosons are highlighted.

## 1.2 The Standard Model as a gauge theory

<sup>491</sup> In this section we lay the theoretical foundations of the Standard Model as a gauge  
<sup>492</sup> theory, starting from the principle of gauge invariance (gauge symmetry), with local  
<sup>493</sup> gauge symmetries giving rise to interactions between particles.

Gauge theories of elementary particle interactions originate from a freedom of choice in the mathematical description of particle fields which has no effect on the particles' physical states [4]. The existence and form of the particles' interactions,

497 can be deduced from the existence of physically indeterminate, gaugable quantities.

498 An example of this gauge invariance in classical physics is the electromagnetic  
499 interaction, where the fundamental field is the four-vector potential  $A^\mu$  [4]. The  
500 physical electromagnetic fields and Maxwell's equations arise from the elements of  
501 the tensor  $F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x)$ . Any two choices of  $A^\mu$  that are related by a  
502 transformation of the form

$$A_\mu \rightarrow A_\mu + \partial_\mu \alpha \quad (1.1)$$

503 for any real, differentiable function  $\alpha(x)$ , describe the same physical configuration,  
504 and have no effect on Maxwell's equations. This "redundancy" in the choice of gauge  
505 in Eqn. 1.1 is called a gauge symmetry.

506 One important consequence of gauge symmetry comes from the application of  
507 Noether's theorem, which states that for every global transformation under which the  
508 Lagrangian density is invariant, there exists a conserved quantity. If  $\mathcal{L}(\Psi(x), \partial_\mu \Psi(x))$   
509 is invariant under the transformation of the wave function  $\Psi(x) \rightarrow \Psi'(x)$ , where  
510  $\Psi'(x) = \Psi(x) + \delta\Psi(x)$ , then there exists a conserved current

$$\partial_\mu \left( \frac{\partial \mathcal{L}(x)}{\partial(\partial_\mu \Psi(x))} \delta\Psi(x) \right) = 0 \quad (1.2)$$

511 In classical mechanics, the conservation of linear momentum, angular momentum,  
512 and energy follows from translational invariance, rotational variance, and invariance  
513 under translations in time [4]. Likewise, charge conservation can be shown to arise  
514 from the invariance of the Dirac Lagrangian density  $\mathcal{L}_{\text{Dirac}} = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi$  under the  
515 particle wavefunction's phase transformation,  $\Psi'(x) = \exp(i\epsilon\chi)\Psi(x)$ . Thus Noether's  
516 theorem establishes a correspondence between a gauge symmetry and a conserved  
517 internal property (e.g. charge or momentum).

518 Interactions between particles arise if we modify the wave function with a phase

519 transformation  $\Psi'(x) = \exp(ie\chi)\Psi(x)$ , and allow the phase  $\chi$  to be a function of  
520 spacetime [4]. A wave function of the form

$$\Psi'(x) = \exp(ie\chi(x))\Psi(x) \quad (1.3)$$

521 can be verified to *not* be a solution to the Dirac equation for free particles:  $(i\gamma^\mu\partial_\mu -$   
522  $m)\Psi(x) = 0$ . This necessitates a modified Dirac equation, where the derivative takes  
523 into account that the vector field  $V(x)$  needs to be compared at two displaced space-  
524 time points in a curvilinear coordinate system:

$$\mathcal{D}_\mu \equiv \lim_{\Delta x^\mu \rightarrow 0} \frac{V_{||}(x + \Delta x) - V(x)}{\Delta x^\mu} \quad (1.4)$$

525 We define a covariant derivative,

$$D_\mu = \partial_\mu + ieA_\mu \quad (1.5)$$

526 where  $A_\mu(x)$  is a 4-vector potential. Thus the modified Dirac equation reads:

$$(i\gamma^\mu\mathcal{D}_\mu - m)\Psi(x) = 0 \quad (1.6)$$

527 The simultaneous gauge transformation  $A'_\mu(x) = A_\mu(x) - \partial_\mu\chi(x)$  and wavefunction  
528 transformation  $\Psi'(x) = \exp(ie\chi(x))\Psi(x)$  leaves the covariant-derivative form of the  
529 Dirac equation (Eqn 1.1) invariant.

530 The generalization of this result is as follows: if a theory is invariant for unitary  
531 transformations  $U$  of the particle states according to

$$\Psi' = U\Psi \quad (1.7)$$

532 One must define a derivative of the form

$$D^\mu = \partial^\mu + igB^\mu \quad (1.8)$$

533 to keep the theory invariant under Eqn. 1.7. The four-potential  $B^\mu$  represents the  
534 interacting four-potential which must be added to keep the theory invariant.

535 In the case of the Standard Model, the theory is built around the gauge trans-  
536 formations  $G = SU(3) \times SU(2) \times U(1)$ .  $SU(3)$  is associated to the strong force  
537 (subscripted  $C$ );  $SU(2)$  is associated to the weak force (subscripted  $L$ ); and  $U(1)$  is  
538 hypercharge (subscripted  $Y$ ). The gauge-covariant derivative is

$$\mathcal{D}_\mu = \partial_\mu - ig'B_\mu \frac{Y}{2} - igW_\mu^\alpha \frac{\tau_a}{2} - ig_s G_\mu^k \frac{\lambda_k}{2} \quad (1.9)$$

539 • In the  $U(1)_Y$  term,  $B_\mu$  is the weak hypercharge field.

540 • In the  $SU(2)_L$  term,  $W_\mu(x) = (W_\mu^1(x), W_\mu^2(x), W_\mu^3(x))$  are a triplet of four-  
541 potentials.  $\tau/2$  are the Pauli matrices, generators of the  $SU(2)$  transformation.

542 • In the  $SU(3)_C$  term, the gluon (color) field is  $G_\mu$ .  $\lambda_k$  are the Gell-Man matrices,  
543 generators of the  $SU(3)$  transformation.

544 The invariance of the Standard Model under  $SU(3)_C \times SU(2)_L \times U(1)_Y$  requires  
545 massless fermions and massless force carriers.

### 546 1.3 The Higgs mechanism

547 To introduce mass into the theory, i.e. to change the propagation of the gauge par-  
548 ticles and all the fermions, the physical vacuum cannot have all the symmetries of  
549 the Standard Model Lagrangian [4]. The symmetries of the physical vacuum must  
550 be spontaneously broken, without affecting gauge invariance in the Lagrangian. The

551 Higgs mechanism proposes the existence of a scalar field, or fields, with nonzero vac-  
 552 um expectation values, which reduce the gauge symmetries of the physical vacuum  
 553 from  $SU(3)_C \times SU(2)_L \times U(1)_Y$  down to  $SU(3)_C \times U(1)_{EM}$ .

554 The Higgs field interacts with the gauge bosons and fermions throughout space,  
 555 impeding their free propagation. The resulting broken symmetry correctly predicts  
 556 the mass ratio of the neutral (Z) and charged (W) massive electroweak bosons, and  
 557 predicts that at least one physical degree of freedom in the Higgs field is a particle  
 558 degree of freedom, called the Higgs boson. The location of the minimum of the Higgs  
 559 potential can be constrained from previously measured Standard Model parameters,  
 560 but the shape of the mass distribution of the Higgs boson must be experimentally  
 561 measured.

562 The minimal choice of Higgs field comes from the breaking of  $SU(2)_L \times U(1)_Y$   
 563 down to  $U(1)_{EM}$ . The smallest  $SU(2)$  multiplet is the doublet. The existence of three  
 564 massive electroweak bosons leads the Higgs sector to have at least three degrees of  
 565 freedom. The minimal single-doublet complex scalar Higgs field is

$$\Phi(x) = \begin{pmatrix} \phi^+(x) \\ \phi^0(x) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1^+(x) + i\phi_2^+(x) \\ \phi_1^0(x) + i\phi_2^0(x) \end{pmatrix} \quad (1.10)$$

566 where  $\phi_1^+$ ,  $\phi_2^+$ ,  $\phi_1^0$ , and  $\phi_2^0$  are real (four degrees of freedom). By convention, the  
 567 nonzero vacuum expectation value is assigned to  $\phi_1^0$ .

568 The minimal self-interacting Higgs potential that is invariant under  $SU(2)_L \times$   
 569  $U(1)_Y$  is given by

$$V(\Phi^\dagger \Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \quad \mu^2 > 0, \lambda > 0 \quad (1.11)$$

570 where  $\lambda$  is the coupling strength of the four-point Higgs interaction. The potential

571 energy is minimized at

$$\Phi_{\min} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \text{ where } v = \sqrt{\mu^2/\lambda} \quad (1.12)$$

572 Choosing a fixed orientation of  $\langle \Phi \rangle$  out of a continuous set of possible ground states  
573 spontaneously breaks the symmetry of the physical vacuum, as illustrated in Fig 1.2.

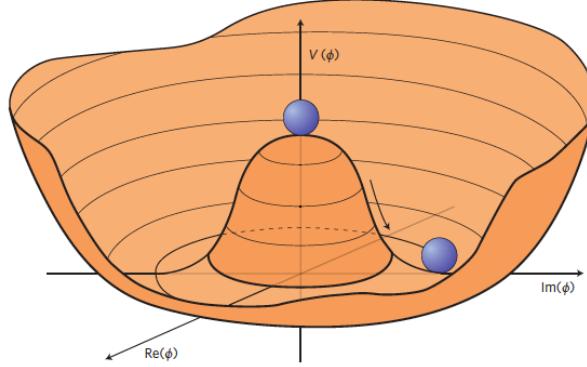


Figure 1.2: An illustration of the Higgs potential [5]. Choosing any of the points at the bottom of the potential breaks spontaneously the rotational  $U(1)$  symmetry.

574 The excitations of the Higgs field with respect to the minimum  $\Phi_{\min}$  are parame-  
575 terized by

$$\Phi(x) = \exp(i\xi(x) \cdot \tau) \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix} \quad (1.13)$$

576 Three degrees of freedom are coupled directly to the electroweak gauge bosons; this  
577 is often referred to as the gauge bosons “eating” the Goldstone bosons to form the  
578 longitudinal polarizations of the massive spin-1 boson states. The  $H(x)$  excitation is  
579 in the radial direction and corresponds to the free particle state of the Higgs boson.

## 580 1.4 Two-Higgs Doublet Models

581 One of the simplest possible extensions to the Standard Model is adding a doublet  
582 to the minimal Higgs sector of the Standard Model, which is a  $SU(2)_L$  doublet  $H$

583 with hypercharge  $Y = +\frac{1}{2}$ , denoted here as  $H \sim 2_{+1/2}$ . These extensions are found  
 584 in several theories such as supersymmetry. A general 2HDM can be extended with a  
 585 light scalar (2HDM+S) to obtain a rich set of exotic Higgs decays [6].

The charges of the Higgs fields are chosen to be  $H_1 \sim 2_{-1/2}$  and  $H_2 \sim 2_{+1/2}$ , which  
 acquire vacuum expectation values  $v_{1,2}$  which are assumed to be real and aligned [6].  
 Expanding about the minima yields two complex and four real degrees of freedom:

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} v_1 + H_{1,R}^0 + iH_{1,I}^0 \\ H_{1,R}^- + iH_{1,I}^- \end{pmatrix} \quad (1.14)$$

$$H_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} H_{2,R}^+ + iH_{2,I}^+ \\ v_2 + H_{2,R}^0 + iH_{2,I}^0 \end{pmatrix} \quad (1.15)$$

586 The charged scalar and pseudoscalar mass matrices are diagonalized by a rotation  
 587 angle  $\beta$ , defined as  $\tan \beta = v_2/v_1$ . One charged (complex) field and one neutral  
 588 pseudoscalar combination of  $H_{1,2,I}^0$  are eaten by the SM gauge bosons after electroweak  
 589 symmetry breaking [6]. The other complex field yields two charged mass eigenstates  
 590  $H^\pm$ , which are assumed to be heavy. The remaining three degrees of freedom yield  
 591 one neutral pseudoscalar mass eigenstate

$$A = H_{1,I}^0 \sin \beta - H_{2,I}^0 \cos \beta \quad (1.16)$$

592 and two neutral scalar mass eigenstates (where  $-\pi/2 \leq \alpha \leq \pi/2$ )

$$\begin{pmatrix} h \\ H^0 \end{pmatrix} = \begin{pmatrix} -\sin \alpha & \cos \alpha \\ \cos \alpha & \sin \alpha \end{pmatrix} \begin{pmatrix} H_{1,R}^0 \\ H_{2,R}^0 \end{pmatrix} \quad (1.17)$$

593 We assume that the 2HDM is near or in the decoupling limit:  $\alpha \rightarrow \pi/2 - \beta$ , where the  
 594 lightest state in the 2HDM is  $h$ , which we identify as the 125 GeV Higgs particle [6]. In  
 595 this limit, the fermion couplings of  $h$  become identical to the Standard Model Higgs,

596 while the gauge boson couplings are very close to Standard Model-like for  $\tan \beta \gtrsim 5$ .  
597 All of the properties of  $h$  can be determined by just two parameters:  $\tan \beta$  and  $\alpha$ ,  
598 and the fermion couplings to the two Higgs doublets.

599 2HDM can be extended by a scalar singlet (2HDM+S) [6]:

$$S = \frac{1}{\sqrt{2}}(S_R + iS_I) \quad (1.18)$$

600 If this singlet only couples to the Higgs doublets  $H_{1,2}$  and has no direct Yukawa  
601 couplings, all of its couplings to SM fermions result from mixing with  $H_{1,2}$ . Under  
602 these simple assumptions, exotic Higgs decays  $h \rightarrow ss \rightarrow X\bar{X}Y\bar{Y}$  or  $h \rightarrow aa \rightarrow$   
603  $X\bar{X}Y\bar{Y}$ , and  $h \rightarrow aZ \rightarrow X\bar{X}Y\bar{Y}$  are permitted, where  $s(a)$  is a (pseudo)scalar mass  
604 eigenstate mostly composed of  $S_R(S_I)$ , and  $X, Y$  are Standard Model fermions or  
605 gauge bosons. There are two pseudoscalars in the 2HDM+S, and the mostly singlet-  
606 like pseudoscalar can be chosen to be the one lighter than the SM-like Higgs. For  
607  $m_a < m_h - m_Z \sim 35$  GeV, the exotic Higgs decay  $h \rightarrow Za$  is possible, and for  
608  $m_a < m_h/2 \approx 63$  GeV, the exotic Higgs decay  $h \rightarrow aa$  is possible.

609 In 2HDM, and by extension 2HDM+S, there are four types of fermion couplings  
610 commonly discussed in the literature that forbid flavor-changing neutral currents at  
611 tree level [6]. These are referred to as Type I (all fermions couple to  $H_2$ ), Type II  
612 (MSSM-like,  $d_R$  and  $e_R$  couple to  $H_1$ ,  $u_R$  to  $H_2$ ), Type III (lepton-specific, leptons  
613 and quarks couple to  $H_1$  and  $H_2$  respectively) and Type IV (flipped, with  $u_R$ ,  $e_R$   
614 coupling to  $H_2$  and  $d_R$  to  $H_1$ ). The exact branching ratios of the pseudoscalars to  
615 Standard Model particles vary depending on the 2HDM+S model and the value of  
616  $\tan \beta$  (e.g. Fig. 1.3).

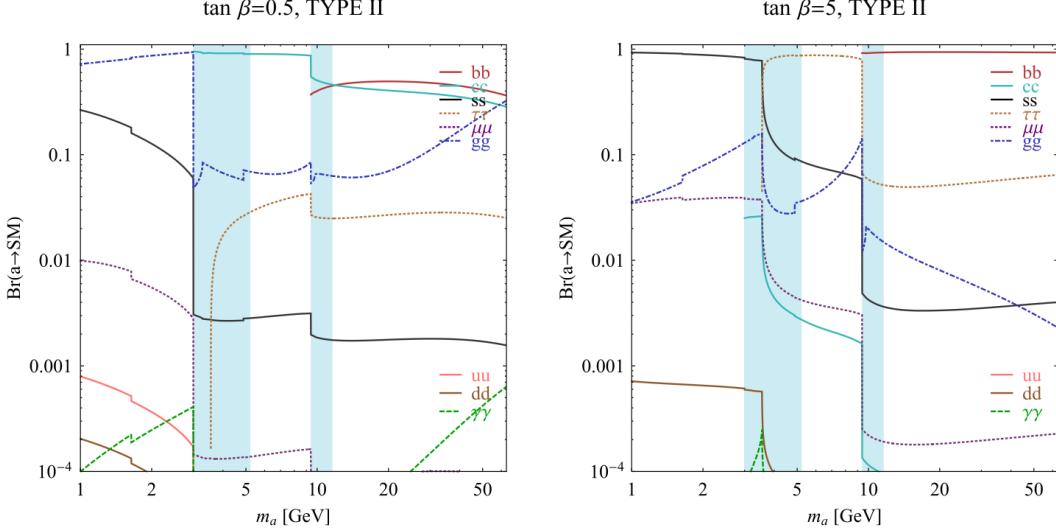


FIG. 7 (color online). Branching ratios of a singletlike pseudoscalar in the 2HDM + S for type-II Yukawa couplings. Decays to quarkonia likely invalidate our simple calculations in the shaded regions.

Figure 1.3: Branching ratios of a singlet-like pseudoscalar in Type II 2HDM+S for  $\tan\beta = 0.5$  (left) and  $\tan\beta = 5$  (right) from [6], showing the dependence of the branching ratios on  $\tan\beta$ , as well as the prominence of the branching ratios to  $bb$  and  $\tau\tau$ , the channels searched for in the analysis presented here.

## 1.5 Two Real Singlet Model

The two real singlet model (TRSM) adds two real singlet degrees of freedom to the Standard Model. These are written as two real singlet fields  $S$  and  $X$ . Depending on the vacuum expectation values acquired by the scalars, different phases of the model can be realized [7]. To reduce the number of free parameters, two discrete  $\mathbb{Z}_2$  symmetries are introduced. The fields are decomposed as

$$\Phi = \begin{pmatrix} 0 \\ \frac{\phi_h + v}{\sqrt{2}} \end{pmatrix}, S = \frac{\phi_S + v_S}{\sqrt{2}}, X = \frac{\phi_X + v_X}{\sqrt{2}} \quad (1.19)$$

To achieve electroweak-breaking symmetry,  $v = v_{SM} \sim 246$  GeV is necessary. If the vacuum expectation values  $v_S, v_X \neq 0$  the  $\mathbb{Z}_2$  are spontaneously broken, and the fields  $\phi_{h,S,X}$  mix into three physical scalar states. This is called the broken phase and leads to the most interesting collider phenomenology.

627        The mass eigenstates  $h_{1,2,3}$  are related to the fields  $\phi_{h,S,X}$  through a  $3 \times 3$  orthogonal  
628        mixing matrix denoted  $R$ . The mass eigenstates are assumed to be ordered  $M_1 \leq$   
629         $M_2 \leq M_3$ .  $R$  is parameterized by the three mixing angles  $\theta_{hS}$ ,  $\theta_{hX}$ ,  $\theta_{SX}$ . The nine  
630        parameters of the scalar potential can be expressed in terms of the three physical  
631        Higgs masses, the three mixing angles, and the three vacuum expectation values.

632        After fixing one of the Higgs masses to the mass of the observed Higgs boson, and  
633        fixing the Higgs doublet vacuum expectation value to its Standard Model value, there  
634        are seven remaining free parameters of the TRSM [7].

635        In one benchmark scenario of TRSM [7], the heaviest scalar state  $h_3$  is identified  
636        with the 125 GeV Higgs,  $h_{125}$ , and it can decay asymmetrically  $h_{125} \rightarrow h_1 h_2$ , which  
637        we also denote  $h \rightarrow a_1 a_2$  to highlight the similarity with the symmetric decay  $h \rightarrow aa$   
638        typically interpreted in 2HDM+S as discussed. The parameter values in TRSM are  
639        chosen such that the coupling of  $h_3$  to Standard Model particles are nearly identical  
640        to the Standard Model predictions.

641        In benchmark scenario 1 (benchmark plane 1, or BP1) (Fig. 1.4) [7], the maximal  
642        branching ratios for  $h_3 \rightarrow h_1 h_2$  reach up to 7 – 8% which translates into a signal  
643        rate of around 3 pb. These maximal branching ratios are reached in the intermediate  
644        mass state for  $h_2$ ,  $M_2 \sim 60\text{--}80\text{ GeV}$ . For  $M_2 < 40\text{ GeV}$ , although phase space opens  
645        up significantly for light decay products, the branching ratio becomes smaller.

646        If the decay channel  $h_2 \rightarrow h_1 h_1$  is kinematically open (i.e.  $M_2 > 2M_1$ ), it is the  
647        dominant decay mode leading to a significant rate for the  $h_1 h_1 h_1$  final state, in a  
648        “cascade” decay. In BP1,  $BR(h_2 \rightarrow h_1 h_1) \simeq 100\%$  above the red line in Fig. 1.4. If,  
649        in addition,  $M_1 \gtrsim 10\text{ GeV}$ , the  $h_1$  decays dominantly to  $b\bar{b}$  leading to a sizable rate  
650        for the  $b\bar{b}b\bar{b}b\bar{b}$  final state as shown in Fig. 1.4 (*bottom right*).

651        If the  $h_2 \rightarrow h_1 h_1$  decay is kinematically closed (i.e.  $M_2 < 2M_1$ ), both scalars decay  
652        directly to Standard Model particles, with branching ratios identical to a Standard  
653        Model-like Higgs boson, i.e. with the  $b\bar{b}b\bar{b}$  final state dominating, as shown in Fig. 1.4

654 (*bottom left*), while at smaller masses, combinations with  $\tau$  leptons and eventually  
655 final states with charm quarks and muons become relevant [7].

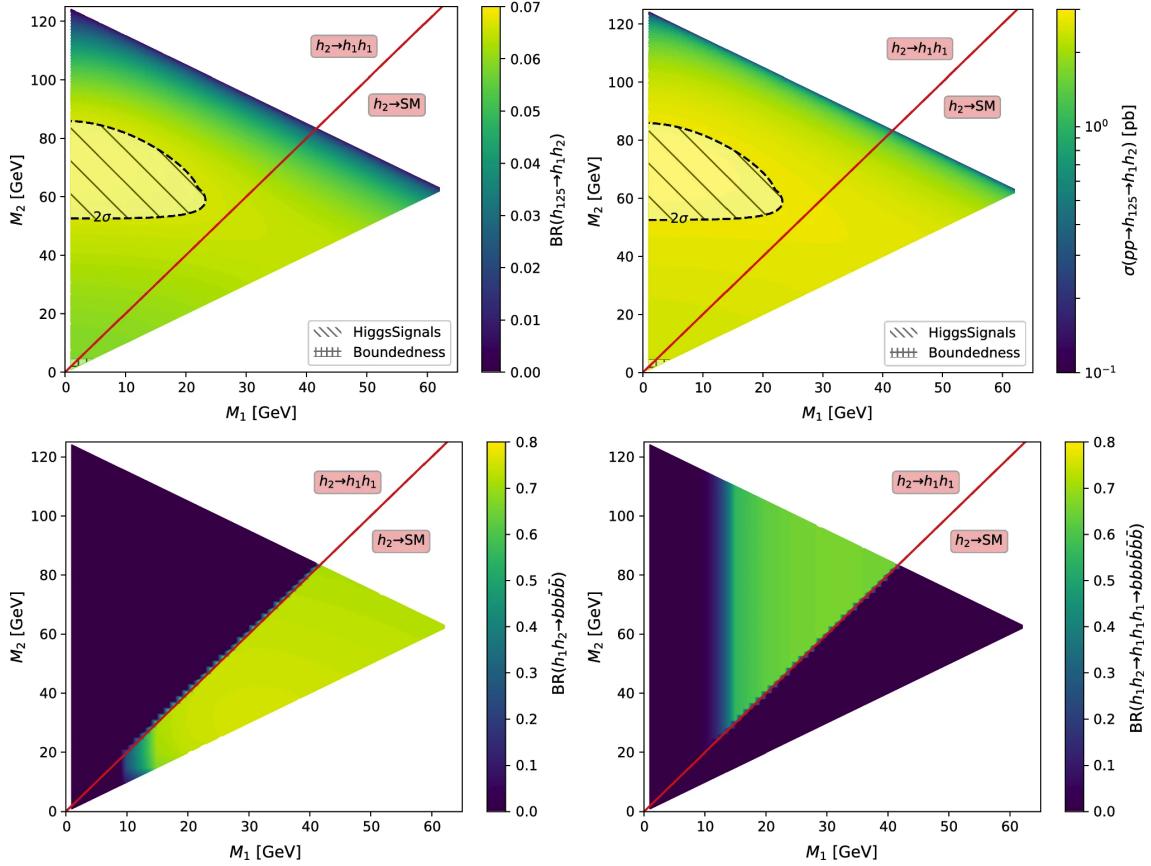


Figure 1.4: Benchmark plane BP1 for benchmark scenario 1 from [7], for the decay signature  $h_{125} \rightarrow h_1 h_2$  with  $h_{125} \equiv h_3$ , defined in the  $(M_1, M_2)$  plane. The color code shows  $\text{BR}(h_3 \rightarrow h_1 h_2)$  (*top left*) and the 13 TeV LHC signal rate for  $pp \rightarrow h_3 \rightarrow h_1 h_2$  (*top right*). The red line separates the region  $M_2 > 2M_1$ , where  $\text{BR}(h_2 \rightarrow h_1 h_1) \sim 100\%$ , from the region  $M_2 < 2M_1$ , where  $\text{BR}(h_2 \rightarrow F_{SM}) \sim 100\%$ . The *bottom left* and *right* show the branching ratio of the  $h_1 h_2$  into (respectively)  $b\bar{b}b\bar{b}$ , and through a  $h_2 \rightarrow h_1 h_1$  cascade to  $b\bar{b}b\bar{b}b\bar{b}$ . The hatched region indicates where the decay rate slightly exceeds the  $2\sigma$  upper limit inferred from the LHC Higgs rate measurements, though the region depends on the parameter choices and experimental searches should cover the whole mass range.

656 **Chapter 2**

657 **The Large Hadron Collider and the**  
658 **CMS Experiment**

659 This chapter introduces the key aspects of the CERN Large Hadron Collider (LHC)  
660 and the Compact Muon Solenoid (CMS) experiment where the work for this thesis was  
661 conducted. Section 2.1 describes the history of accelerator developments at CERN  
662 that led to the construction of the LHC, the current LHC configuration, and the  
663 largest experiments located at the LHC. The concepts of beam luminosity and pile-  
664 up, which are critical for understanding and measuring high-energy particle collisions,  
665 are described in Section 2.2 and discussed in the context of the High-Luminosity  
666 LHC (HL-LHC) upgrade in Section 2.3. Lastly, Section 2.4 describes the design  
667 and function of CMS and its subdetectors, and terminates in a description of data  
668 processing at CMS, beginning from online event filtering in the Level-1 Trigger, to  
669 processing in the High-Level Trigger, to offline particle reconstruction, and finally  
670 long-term storage and processing of measured events.

## 671 2.1 The Large Hadron Collider

672 CERN, the European Organization for Nuclear Research, is an international organiza-  
673 tion based in Meyrin, Switzerland which operates the world's largest particle physics  
674 laboratory, and is the site of the Large Hadron Collider (LHC) [8]. The very first  
675 accelerator built at CERN was the 600 MeV Synchrocyclotron (SC), which initially  
676 provided beams for CERN's first experiments. The newer and more powerful Proton  
677 Synchrotron (PS), which could accelerate particles to an energy of 28 GeV, began  
678 operations in 1959 and is still in use today. The first hadron collider at CERN was  
679 the Intersecting Storage Rings (ISR), which consisted of two interlaced rings each  
680 with a diameter of 300 meters. The ISR collided protons at a center-of-mass energy  
681 of 62 GeV and began measuring collisions in 1971. In 1968 CERN began to accel-  
682 erate heavy ions in the Super Proton Synchrotron (SPS), which is 7 kilometers in  
683 circumference and was the first of CERN's giant underground rings to be built. The  
684 SPS became the forefront of CERN's particle physics program in 1976, and in 1981  
685 was converted into a proton-antiproton collider. The final and largest underground  
686 ring constructed at CERN was the Large Electron-Positron (LEP) collider, which  
687 was commissioned in July 1989 and hosted 5176 magnets and 128 accelerating cav-  
688 ities located around a 27-kilometer circumference. Over 11 years of research, four  
689 detectors, ALEPH, DELPHI, L3, and OPAL measured the collisions, with collision  
690 energies reaching up to 209 GeV in the year 2000. In November 2000, LEP was closed  
691 down to make way for the construction of the LHC in the same tunnel.

692 In its current configuration, the LHC accelerator complex at CERN is a suc-  
693 cession of machines that accelerate particles in stages until they reach their final energy  
694 of 6.5 TeV per beam [9] [10]. In Linear accelerator 4 (Linac4), negative hydrogen  
695 ions (hydrogen atoms with an additional electron) are accelerated to 160 MeV, and  
696 stripped of their two electrons, leaving only protons, before entering the Proton Syn-  
697 chrotron Booster (PSB). These protons are accelerated to 2 GeV, then to 26 GeV in

698 the Proton Synchrotron (PS), and 450 GeV in the Super Proton Synchrotron (SPS).  
699 The protons are transferred to the two beam pipes of the Large Hadron Collider  
700 (LHC). The LHC is a 27-kilometer ring of superconducting magnets, inside which  
701 one beam circulates clockwise and the other counterclockwise. Each LHC ring takes  
702 4 minutes and 20 seconds to fill, and it takes about 20 minutes for the protons to  
703 reach their maximum energy. During normal operating conditions, beams circulate  
704 for many hours inside the LHC ring.

705 The beams of particles in the LHC are made to collide at a center-of-mass en-  
706 ergy of up to 13.6 TeV, at four positions at particle detector experiments located  
707 around the ring: ATLAS, CMS, ALICE, and LHCb. An aerial view of the four ma-  
708 jor experiments' locations is shown in Fig. 2.1 [11]. ATLAS and CMS are the two  
709 general-purpose detectors with broad physics programmes spanning Standard Model  
710 measurements and searches for signatures of new physics [12] [13]. The two experi-  
711 ments use different technical solutions and different magnet system designs. ALICE  
712 is a general-purpose detector dedicated to measuring LHC heavy-ion collisions, and  
713 is designed to address the physics of strongly interacting matter, and the properties  
714 of quark-gluon plasma [14]. The LHCb experiment specializes in investigating CP vi-  
715 olation through measuring the differences in matter and antimatter, by using a series  
716 of subdetectors to detect mainly forward particles close to the beam direction [15].

## 717 **2.2 Luminosity and pile-up**

718 In order to search for rare processes, such as those resulting from a Higgs, W, or Z  
719 boson, a large number of parton interactions per second are required at the LHC.  
720 The number of events generated per second by the LHC collisions is given by

$$n_{event} = \mathcal{L} \cdot \sigma_{event} \quad (2.1)$$

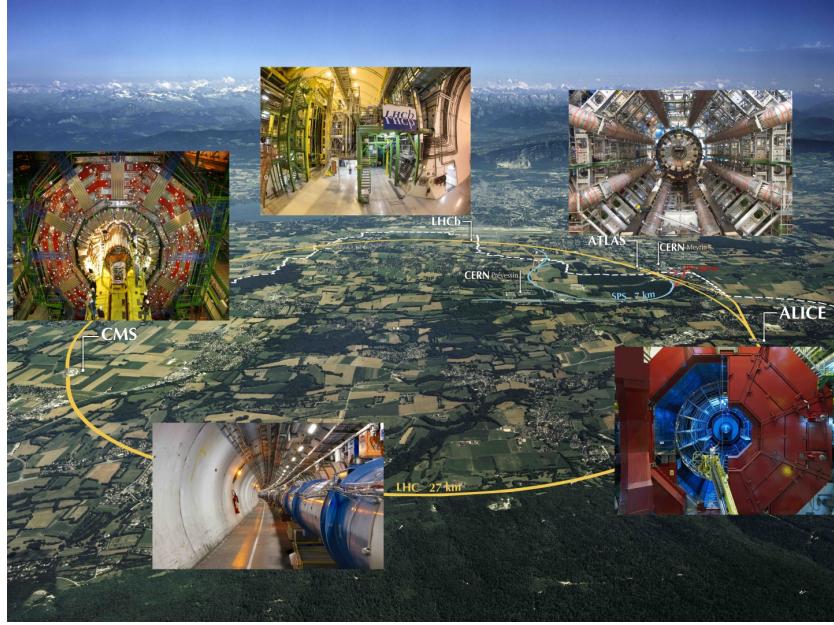


Figure 2.1: Aerial view of the Large Hadron Collider (LHC) spanning the border of France and Switzerland, and the four major experiments located around the ring: CMS (Compact Muon Solenoid), LHCb (LHC beauty), ATLAS (A Toroidal LHC Apparatus), and ALICE (A Large Ion Collider Experiment) [11].

where  $\sigma_{event}$  is the cross-section for the event under study, and  $\mathcal{L}$  the instantaneous luminosity. The instantaneous luminosity is measured in units of  $\text{cm}^{-2} \text{ s}^{-1}$ , and depends only on the beam parameters, and can be written for a Gaussian beam distribution as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.2)$$

where the parameters are as defined, along with some example typical nominal values in Phase-1 of the LHC [16] [17]:

- $N_b$  is the number of particles per bunch ( $N_b \approx 1.15 \times 10^{11}$  protons per bunch)
- $n_b$  is the number of bunches per beam (maximum 2808),
- $f_{rev}$  is the revolution frequency ( $\approx 11 \text{ kHz}$ ),
- $\gamma_r$  is the relativistic gamma factor,

- 731     •  $\epsilon_n$  is the normalized transverse beam emittance (area in a transverse plane  
 732                 occupied by the beam particles),  
 733     •  $\beta^*$  is the beta function at the collision point ( $\beta^* = 0.55$  m),  
 734     • and  $F$  is the geometric luminosity reduction factor due to the crossing angle at  
 735                 the interaction points ( $F \approx 0.84$  for Phase-1. Note that complete overlap would  
 736                 give  $F = 1$ ).

737 Peak luminosity at interaction points 1 and 5 reach values of  $\sim 1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ,  
 738 with peak luminosity per bunch crossing reaching  $\sim 3.56 \times 10^{30} \text{ cm}^{-2} \text{ s}^{-1}$  [18].

739 Per Eqn. 2.1, the integrated luminosity over time is proportional to the number  
 740 of events produced, and the size of LHC datasets is commonly presented in terms of  
 741 integrated luminosity. Collider operation aims to optimize the integrated luminosity.  
 742 Thus the exploration of rare events in the LHC collisions requires both high beam  
 743 energies and high beam intensities.

744 The interaction vertex corresponding to the hard scattering of the protons is called  
 745 the primary interaction vertex, or primary vertex (PV). The LHC's nominal beam  
 746 luminosities are sufficiently large for multiple proton-proton collisions to occur in the  
 747 same time window in which proton bunches collide [19]. To measure a proton-proton  
 748 collision, the primary vertices must be separated from overlapping collisions, called  
 749 "pile-up" collisions.

750 The pile-up is defined as the average number of  $pp$  collisions per bunch crossing,  
 751 and can be estimated from the inelastic  $pp$  cross section of  $\sigma_{\text{inel}} = 68.6$  millibarns at  
 752 a center-of-mass energy of  $\sqrt{s} = 13 \text{ TeV}$  [20]:

$$\text{Pile-up} = \frac{\mathcal{L} \times \sigma_{\text{inel}}}{n_b \cdot f} \sim 22 \quad (2.3)$$

753 A distribution of pile-up in the data-taking years 2016-2018 is shown in Fig. 2.2.

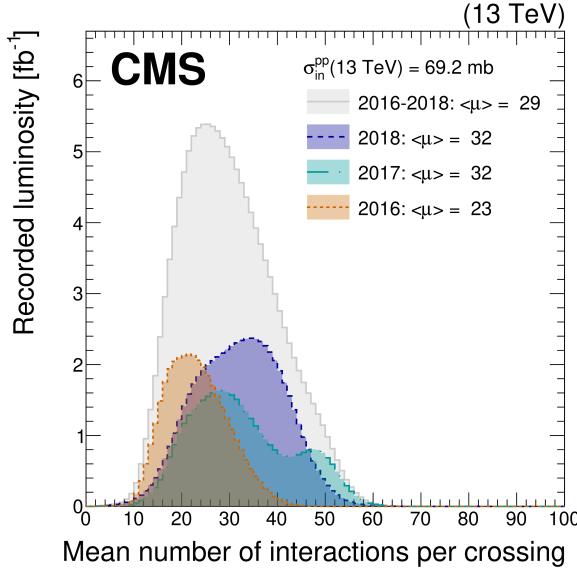


Figure 2.2: Distribution of the mean number of inelastic collisions per bunch crossing (pile-up) in data [19], for proton-proton collisions in 2016 (*dotted orange*), 2017 (*dotted light blue*), 2018 (*dotted dark blue*), and integrated over 2016-2018 (*solid grey*). A cross-section of inelastic proton-proton collisions of 69.2 mbarns is assumed. In the running conditions of the High-Luminosity LHC, pile-up will reach unprecedented levels of up to 200 per bunch crossing [21].

These multiple collisions will lead to higher occupancies in the detector, and particles originating from the pile-up interactions can be confused with those originating from the primary vertex. Thus, higher luminosities create more intense pile-up conditions, posing a greater challenge to detector performance and particle reconstruction and identification.

## 2.3 The High-Luminosity LHC

The High-Luminosity LHC (HL-LHC) is a major upgrade of the LHC scheduled to take place in the late 2020s, that will increase the instantaneous luminosity by a factor of five beyond the original design value, and the integrated luminosity by a factor of ten [21]. This will be accomplished through accelerator technological advances: for instance, reduction of the interaction point  $\beta^*$  from 0.55 m down to 0.15

765 m by installation of new final-focusing magnets, and improvements in the geometric  
766 luminosity loss factor  $F \approx 1$  through the installation of crab cavities that optimize  
767 the orientation of colliding bunches. A further discussion of the HL-LHC upgrades  
768 for the CMS detector follows in Chapter 3.

## 769 2.4 The CMS detector

770 We give a brief overview of the Compact Muon Solenoid (CMS) experiment here  
771 and discuss each of the subdetectors in more detail in the following sections. The  
772 CMS experiment was conceived to study proton-proton and lead-lead collisions at  
773 a center-of-mass energy of 14 TeV (5.5 TeV nucleon-nucleon) and at luminosities up  
774 to  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  ( $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ ) [22] [23]. Starting from the beam interaction region  
775 at the center of the CMS detector, particles first pass through a silicon pixel and  
776 strip tracker, in which charged-particle trajectories (tracks) and origins (vertices)  
777 are reconstructed from signals (hits) in the sensitive layers. The tracker, electro-  
778 magnetic calorimeter (ECAL), and hadronic calorimeter (HCAL) are immersed in a  
779 high-magnetic-field superconducting solenoid that bends the trajectories of charged  
780 particles. After passing through the tracker, electrons and photons are then absorbed  
781 in the electromagnetic calorimeter (ECAL) comprised of lead-tungstate scintillating-  
782 crystals. The corresponding electromagnetic showers are detected as clusters of energy  
783 recording in neighboring cells, from which the direction and energy of the particles can  
784 be determined. Charged and neutral hadrons may initiate a hadronic shower in the  
785 ECAL as well, which is then fully absorbed in the hadron calorimeter (HCAL). The  
786 resulting clusters are used to estimate their direction and energies. Muons and neu-  
787 trinos pass through the calorimeters with little to no interactions. Neutrinos escape  
788 undetected; muons produce hits in additional gas-ionization chamber muon detectors  
789 housed in the iron yoke of the flux-return. A sketch of example particle interactions

in a transverse slice of the CMS detector is shown in Fig. 2.3. The collision data is recorded with the use of the Level-1 (L1) trigger (discussed in greater detail in 2.5.5), the High-Level Trigger (HLT), and data acquisition systems ensuring high efficiency in selecting physics events of interest.

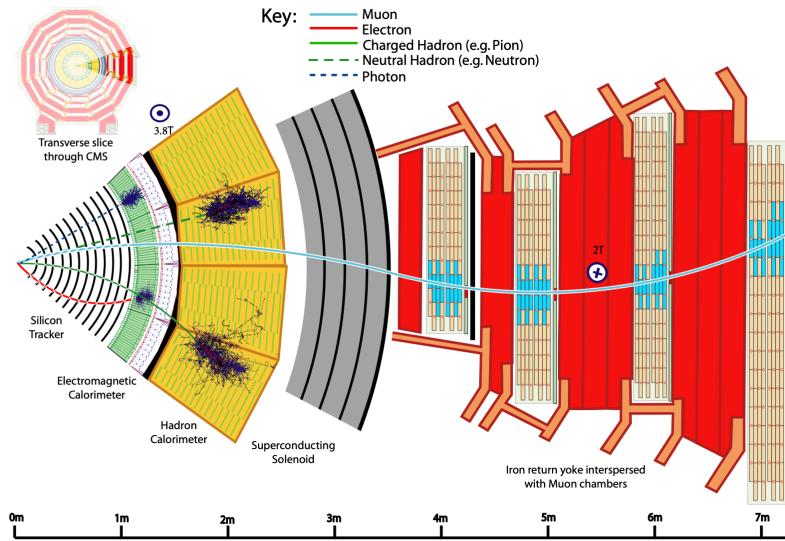


Figure 2.3: Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector [23].

CMS uses a right-handed coordinate system [22]. The origin is centered at the nominal collision point inside the experiment. The  $x$  axis points towards the center of the LHC, and the  $y$  axis points vertically upwards. The  $z$  axis points along the beam direction. The azimuthal angle,  $\phi$ , is measured from the  $x$  axis in the  $x$ - $y$  plane, and the radial coordinate in this plane is denoted by  $r$ . The polar angle,  $\theta$ , is measured from the  $z$  axis. The pseudorapidity,  $\eta$ , is defined as  $\eta = -\ln \tan(\theta/2)$ . The momentum and energy transverse to the beam direction, denoted by  $p_T$  and  $E_T$  respectively, are computed from the  $x$  and  $y$  components. The momentum imbalance in the transverse plane is called the missing transverse momentum, and its magnitude is denoted by  $E_T^{\text{miss}}$ .

## 804 2.5 Sub-detectors of CMS and data processing

805 This section details the sub-detectors of CMS that operate to identify and precisely  
806 measure muons, electrons, photons, and jets over a large energy range. The sections  
807 are ordered starting from the innermost sub-detectors closest to the beam interaction  
808 area: the tracker, the electromagnetic and hadronic calorimeters, and the muon de-  
809 tectors. The two-stage trigger system is described, starting with the hardware-based  
810 Level-1 Trigger and followed by the software-based High-Level Trigger. Lastly, parti-  
811 cle reconstruction and data storage and computational infrastructure are discussed.

### 812 2.5.1 Inner tracking system

813 The CMS Tracker performs robust tracking and detailed vertex reconstruction in the  
814 4 T magnetic field of the superconducting solenoidal magnet. The primary sensors  
815 used in the tracker are  $p^+$  on  $n$ -bulk devices, which allow high voltage operation and  
816 are radiation-resistant [24] [25]. The active envelope of the CMS Tracker extends to a  
817 radius of 115 cm, over a length of approximately 270 cm on each side of the interaction  
818 point [24]. Charged particles in the region  $|\eta| \lesssim 1.6$  benefit from the full momentum  
819 measurement precision. In this region, a charged particle with  $p_T$  of 1000 GeV has a  
820 sagitta of  $\sim 195 \mu\text{m}$ . The Tracker acceptance extends further to  $|\eta| = 2.5$ , with a  
821 reduced radius of approximately 50 cm.

822 The high magnetic field of CMS causes low  $p_T$  charged particles to travel in helical  
823 trajectories with small radii. The majority of events contain particles with a steeply  
824 falling  $p_T$  spectrum, resulting in a track density which rapidly decreases at higher  
825 radii.

826 A schematic view of the current Phase-1 CMS tracker [26], including the pixel  
827 detector, is shown in Fig. 2.4. The Phase-1 pixel detector consists of three barrel  
828 layers (BPIX) at radii of 4.4 cm, 7.3 cm, and 10.2 cm, and two forward/backward disks

829 (FPIX) at longitudinal positions of  $\pm 34.5$  cm and  $\pm 46.5$  cm, and extending in radius  
 830 from about 6 cm to 15 cm. These pixelated detectors produce 3D measurements along  
 831 the paths of charged particles with single hit resolutions between 10-20  $\mu\text{m}$ .

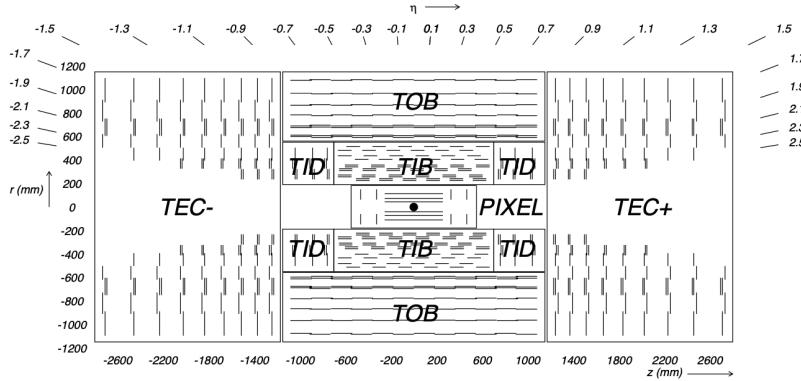


Figure 2.4: Cross section of the current Phase-1 CMS tracker [26]. Each line represents a detector module. Double lines indicate back-to-back modules which deliver two-dimensional (stereo) hits in the strip tracker.

832 After the pixel and on their way out of the tracker, particles pass through the  
 833 silicon strip tracker which reaches out to a radius of 130 cm (Fig. 2.4). The sensor el-  
 834 ements in the strip tracker are single-sided  $p$ -on- $n$  type silicon micro-strip sensors [22].  
 835 The silicon strip detector consists of four inner barrel (TIB) layers assembled in shells,  
 836 with two inner endcaps (TID), each composed of three small discs. The outer barrel  
 837 (TOB) consists of six concentric layers. Two endcaps (TEC) close off the tracker on  
 838 either end.

### 839 2.5.2 ECAL

840 The electromagnetic calorimeter (ECAL) of CMS measures electromagnetic energy  
 841 deposits with high granularity. One of the driving criteria in the design was the capa-  
 842 bility of detecting the Standard Model Higgs boson decay to two photons (in fact, the  
 843 channel in which the 125 GeV Higgs boson was discovered at CMS). ECAL is a her-  
 844 metic homogeneous calorimeter comprised of 61,200 lead tungstate ( $\text{PbWO}_4$ ) crystals

845 mounted in the central barrel, with 7,324 crystals in each of the two endcaps [22]. A  
846 preshower detector is located in front of the endcap crystals. Avalanche photodiodes  
847 (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in  
848 the endcaps.

849 The design of the ECAL is driven by the behaviour of high-energy electrons, which  
850 predominantly lose energy in matter via bremsstrahlung, and high-energy photons  
851 by  $e^+e^-$  pair production. The characteristic amount of matter traversed for these  
852 interactions is the radiation length  $X^0$ , usually measured in units of  $\text{g} \cdot \text{cm}^{-2}$ . The  
853 radiation length is also the mean distance over which a high-energy electron loses all  
854 but  $1/e$  of its energy via bremsstrahlung [27]. Thus high granularity in  $\eta$  and  $\phi$ , and  
855 the length of the ECAL crystals, is designed to capture the shower of  $e/\gamma$  produced  
856 by electrons and photons.

857 The barrel part of the ECAL (EB) covers the pseudorapidity range  $|\eta| < 1.479$  [22].  
858 The barrel granularity is 360-fold in  $\phi$  and  $(2 \times 85)$ -fold in  $\eta$ . The crystal cross-section  
859 corresponds to approximately  $0.0174 \times 0.0174$  in  $\eta - \phi$  or  $22 \times 22 \text{ mm}^2$  at the front  
860 face of the crystal, and  $26 \times 26 \text{ mm}^2$  at the rear face. The crystal length is 230 mm,  
861 corresponding to  $25.8 X_0$ .

862 The ECAL read-out acquires the signals of the photodetectors [22]. At each bunch  
863 crossing, digital sums representing the energy deposit in a trigger tower, comprising  
864  $5 \times 5$  crystals in  $\eta \times \phi$ , are generated and sent to the Level-1 trigger system (detailed  
865 in Section 2.5.5).

### 866 2.5.3 HCAL

867 The hadronic calorimeter (HCAL) of CMS measures hadronic energy, which is key to  
868 characterizing the presence of apparent missing transverse energy which could arise  
869 from hadron jets and neutrinos or exotic particles [22]. A schematic of the components  
870 of HCAL are shown in Fig. 2.5. The HCAL barrel (HB) and endcaps (HE) are located

outside of the tracker and the ECAL, spanning a radius of 1.77 m (outer extent of  
 ECAL) up to 2.95 m (inner extent of the magnet coil). An outer hadron calorimeter  
 (HO) is placed outside the solenoid to complement the barrel calorimeter. Beyond  
 $|\eta| = 3$ , the forward hadron calorimeter (HF) at 11.2 m from the interaction point  
 extend the pseudorapidity coverage to  $|\eta| = 5.2$ .

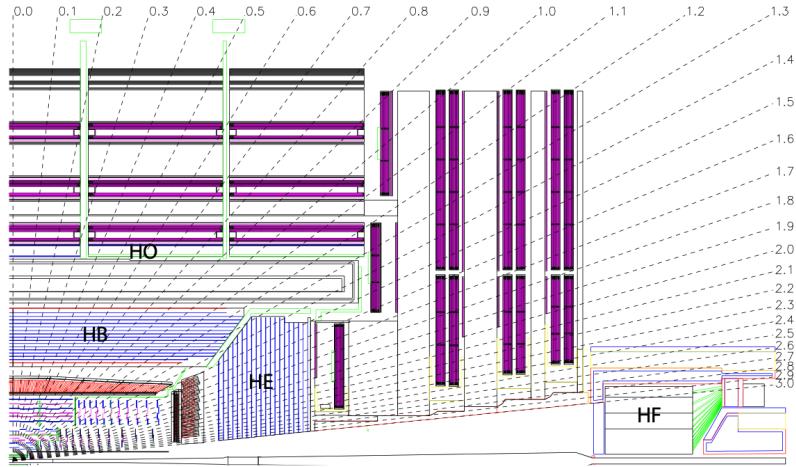


Figure 2.5: Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters from [22].

The HB is a sampling calorimeter covering the pseudorapidity range  $|\eta| < 1.3$  [22]. It consists of 36 identical azimuthal wedges which form two half-barrels (HB+ and HB-), with a segmentation of  $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$ . The HE covers pseudorapidity  $1.3 < |\eta| < 3$ . The HB and endcap HE calorimeters are sampling calorimeters which use brass as the absorber and plastic scintillator as the active material. Light from the plastic scintillator is wavelength-shifted and captured in optic fibers which are read out by front-end electronics [28].

The HF is a Cherenkov calorimeter based on a steel absorber and quartz fibers which run longitudinally through the absorber and collect Cherenkov light, primarily from the electromagnetic component of showers developed in the calorimeter [28]. Photomultiplier tubes are used to collect light from the quartz fibers. The HF is designed to survive in the harsh radiation conditions and high particle flux of the for-

ward region. On average, 760 GeV per proton-proton interaction is deposited into the two forward calorimeters, compared to only 100 GeV for the rest of the detector [22]. Furthermore, this energy has a pronounced maximum at the highest rapidities.

#### 2.5.4 Muon detectors

The CMS muon system is designed to have the capability of reconstructing the momentum and charge of muons over the kinematic range of the LHC, since muons are a powerful handle on signatures of interesting processes over the high background rate of the LHC [22]. For instance, the decay of the Standard Model Higgs boson into  $ZZ$ , which in turn decay to 4 leptons, can be reconstructed with high 4-particle mass resolution if all the leptons are muons, since muons are less affected than electrons by radiative losses in the tracker material.

The muon system consists of a cylindrical barrel section and two planar endcap regions [22]. The barrel muon detector consists of drift tube (DT) chambers covering the pseudorapidity region  $|\eta| < 1.2$  (Fig. 2.6). The DTs can be used as tracking detectors due to the barrel region’s characteristic low neutron-induced backgrounds, low muon rate, and relatively uniform 4T magnetic field contained in the steel yoke.

In the two endcap regions, the muon rates and background levels are high and the magnetic field is large and non-uniform [22]. Here, the muon system uses cathode strip chambers (CSCs) to identify muons between  $0.9 < |\eta| < 2.4$ . The cathode strips of each chamber run radially outwards and provide a precision measurement in the  $r - \phi$  bending plane. The anode wires run approximately perpendicular to the strips and are read out in order to measure  $\eta$  and the beam-crossing time of a muon.

In addition to the DT and CSC, a dedicated trigger system consisting of resistive plate chambers (RPCs) in the barrel and endcap regions provide a fast, independent, and highly-segmented trigger with a sharp  $p_T$  threshold over a large portion of the pseudorapidity range ( $|\eta| < 1.6$ ) of the muon system [22]. RPCs have good time

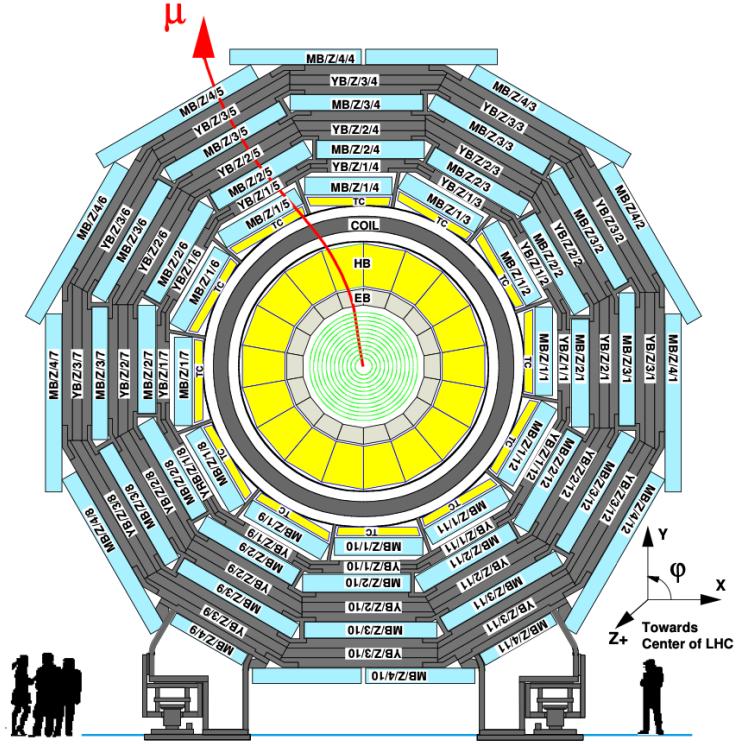


Figure 2.6: Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels from [22]. The DTs are organized in 12 sectors of the yoke barrel (YB). In each of the 12 sectors of the yoke, there are 4 muon chambers per wheel (MB1, MB2, MB3, and MB4).

resolution but coarser position resolution compared to the DTs or CSCs. The RPCs also play a role in resolving ambiguities in reconstructing tracks from multiple hits in a chamber.

### 2.5.5 The Level-1 Trigger

The design performance of the LHC corresponds to an instantaneous luminosity of  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with a 25 ns bunch crossing rate, giving an average pile-up (number of simultaneous events) of 25 per bunch crossing [29]. However, during Run 2, in 2017 and 2018 the LHC was able to surpass this goal with a mean number of 32 interactions per bunch crossing, and reaching over 50 interactions in short periods (Fig.

923 2.2). The large number of events from inelastic collisions (minimum bias events) per  
 924 bunch crossing, combined with the small cross-sections of possible physics discovery  
 925 signatures, necessitates a sophisticated event selection system for filtering this large  
 926 event rate, as it is impossible to save all events. This data filtering system is imple-  
 927 mented by CMS in two stages. The first stage is the Level-1 (L1) Trigger, which is  
 928 deployed in custom electronic hardware systems and is responsible for reducing the  
 929 event rate to around 100 kHz. The second stage is the High-Level Trigger (HLT)  
 930 which is described in Section 2.5.6. This section describes the Phase-1 configuration  
 931 of the Level-1 Trigger.

932 The L1 Trigger data flow of Phase-1 is shown in Fig. 2.7 [29], with organization  
 933 into the L1 calorimeter trigger, the L1 muon trigger, and the L1 Global Trigger (GT).

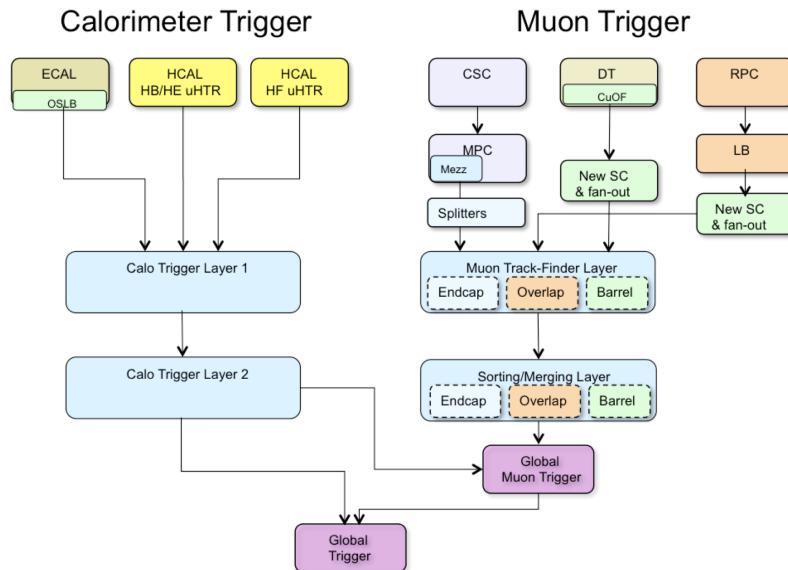


Figure 2.7: Dataflow for the Phase-1 Level-1 Trigger [29], which is implemented in custom hardware and is responsible for reducing the event rate from the LHC bunch crossing frequency of 400 MHz (bunch crossings every 25 ns) to a maximum rate of 100 kHz. In Phase-1, the Level-1 Trigger has access to information from the calorimeter and muon detectors.

934 The L1 calorimeter trigger begins with trigger tower energy sums formed by the  
 935 ECAL, HCAL, and HF Trigger Primitive Generator (TPG) circuits from the indi-

936 individual calorimeter cell energies. In the original configuration, the ECAL energies  
937 were accompanied by a bit indicating the transverse extent of the electromagnetic  
938 energy deposits, and the HCAL energies were accompanied by a bit indicating the  
939 presence of minimum ionizing energy [30]. During Long Shutdowns 1 and 2 (LS1  
940 and LS2), HF was upgraded to provide finer granularity information to the trigger,  
941 and the HCAL barrel and endcap front-end electronics were upgraded to provide  
942 high-precision timing information and depth segmentation information.

943 In the original design of the L1 calorimeter trigger, the trigger primitives are  
944 processed by the Regional Calorimeter Trigger (RCT) which finds isolated and non-  
945 isolated electron/photon candidates [29]. At this stage, electrons/photons candidates  
946 are treated together since they cannot be definitively distinguished at this stage due  
947 to lack of tracking information in the L1 trigger. The Global Calorimeter Trigger  
948 (GCT) sorts further the candidate electrons/photons, finds jets (classified as central,  
949 forward, and tau) using the  $E_T$  sums and performs calibration of the clustered jet  
950 energies, and calculates global quantities such as missing  $E_T$ . It sends the top four  
951 candidates of each type to the Global Trigger [29].

952 During LS2 and before Run-2, the legacy calorimeter trigger was upgraded to be  
953 more flexible, maintainable, and performant [31] [32] [33]. These upgrades included  
954 the replacement of legacy VME-based electronics with the MicroTCA ( $\mu$ TCA) mod-  
955 ern telecommunication standard, and system-wide usage of the latest generation of  
956 FPGAs, Xilinx Virtex 7. Parallel copper links were replaced in almost all cases with  
957 serial optical links, allowing link speeds to increase from 1 Gb/s to 10 Gb/s [31]. A  
958 schematic of the current calorimeter trigger is shown in Fig. 2.8. The calorimeter  
959 Layer-1 is implemented in 18 Calorimeter Trigger Processor (CTP7) boards, with  
960 each card spanning 4 out of 72 towers in  $\phi$  and all of  $\eta$ . Tower-level operations are  
961 performed in Layer-1, such as the sum of ECAL and HCAL energies, energy calibra-  
962 tion, and the computation of the ratio of HCAL to ECAL energies. The Layer-1 cards

963 each transmit 48 output links at 10 Gb/s to the nine Layer-2 Master Processor cards  
 964 (MP7) cards, which host calorimeter algorithms that find particle candidates and  
 965 compute global energy sums. Each MP7 takes 72 input links and has access to the  
 966 whole event at trigger tower granularity, such that the algorithms are fully pipelined  
 967 and start processing as soon as the minimum amount of data is received. The trigger  
 968 candidates are sent to a demultiplexer board (Demux), also a MP7, which formats  
 969 the data for the upgraded Global Trigger, also called the microGT ( $\mu$ GT).

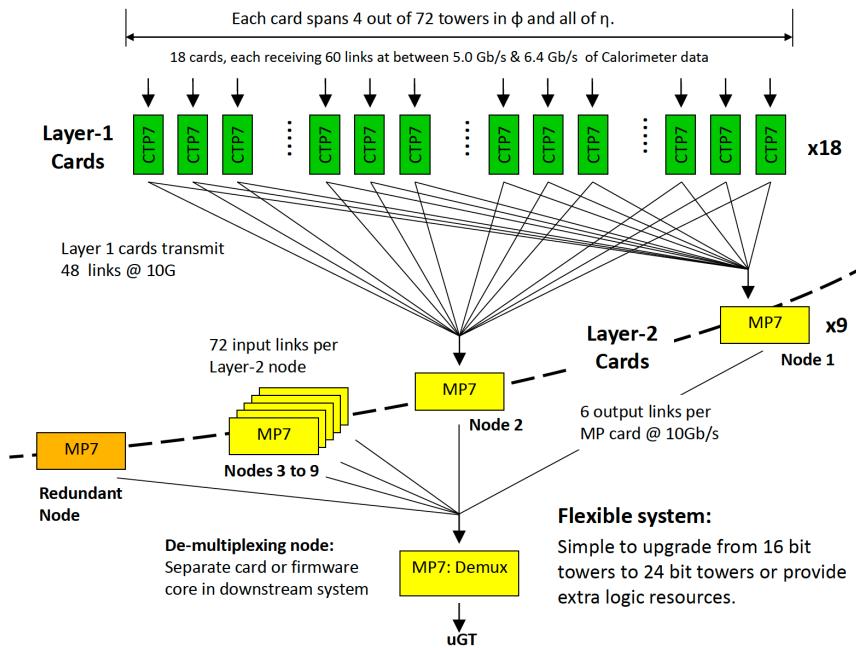


Figure 2.8: Schematic of the calorimeter trigger after Long Shutdown 2 [31]. The Layer-1 calorimeter trigger is implemented in CTP7 cards, which send time-multiplexed outputs to the Layer-2 MP7 cards. The Layer-2 cards handle the data in a round-robin style and the outputs are de-multiplexed, producing one output data stream to the Global Trigger.

970 Each of the L1 muon triggers has its own trigger logic [30]. The RPC strips are  
 971 connected to a Pattern Comparator Trigger (PACT), which forms trigger segments  
 972 that are used to build tracks and calculate  $p_T$ . The RPC logic also provides some  
 973 hit data to the CSC trigger system to resolve ambiguities caused by two muons in  
 974 the same CSC. The CSCs form local charged tracks (LCTs) from the cathode strips,

975 which are combined with the anode wire information. LCTs are combined into full  
976 muon tracks and assigned  $p_T$  values.

977 The Global Muon Trigger (GMT) sorts the RPC, DT, and CSC muon tracks,  
978 converts these tracks to the same  $\eta$ ,  $\phi$ , and  $p_T$  scale, and validates the muon sign [30].  
979 It improves the trigger efficiency by merging muon candidates that were detected  
980 in two complementary sub-systems (i.e. DT+RPC, or CSC+RPC). The GMT also  
981 contains logic to correlate the found muon tracks with an  $\eta-\phi$  grid of quiet calorimeter  
982 towers to determine if the muons are isolated, as well as logic to remove duplicate  
983 candidates originating in the overlap regions from both DT and CSC systems. The  
984 final collection of muons are sorted based on their initial quality, correlation, and  $p_T$ ,  
985 and the top four muons are sent to the Global Trigger [30].

986 The Global Trigger (GT) receives information from the Demux and GMT, and  
987 makes the Level-1 Accept (L1A) decision to either discard or accept the bunch cross-  
988 ing [30]. This is accomplished by sorting ranked trigger objects that are accompanied  
989 by positional information in  $\eta$  and  $\phi$ , permitting the trigger to apply criteria with  
990 thresholds that can vary based on the location of the trigger objects, and/or to re-  
991 quire trigger objects to be close to or opposite from each other. The GT L1A decision  
992 arrives at the detector front end with a  $3.8\ \mu\text{s}$  latency after the interaction at a rate  
993 which is required to be less than 100 kHz, and triggers a full readout of the detector  
994 for further processing.

### 995 2.5.6 The High-Level Trigger

996 The HLT is implemented in software running on a large computer farm of fast com-  
997 mercial processors [34] [35]. The algorithms in HLT have access to full data from  
998 all CMS sub-detectors, including the tracker, with full granularity and resolution.  
999 The HLT reconstruction software is similar to what is used offline for full CMS data  
1000 analysis. As a result, the HLT can calculate quantities with a resolution compara-

ble to the final detector resolution, compared to the L1 Trigger. The HLT performs more computationally-intensive algorithms, such as combining tau-jet candidates in the calorimeter with high- $p_T$  stubs in the tracker, to form a hadronic tau trigger. The maximum HLT input rate from the L1 Trigger is 100 kHz, and the HLT output rate is approximately 100 Hz.

The HLT contains trigger paths, each corresponding to a dedicated trigger [36]. A path consists of several steps implemented as software modules. Each HLT trigger path must be seeded by one or more L1 trigger bits: the first module always looks for a L1 seed, consisting of L1 bit(s) and L1 object(s). Each module performs a well-defined task such as unpacking (raw to digitized quantities), reconstruction of physics objects (electrons, muons, jet, missing transverse energy, etc.), making intermediate decisions that trigger more detailed reconstruction modules, and calculating the final decision for the trigger path. If an intermediate filter decision is negative, the rest of the path is not executed, and the trigger rejects the event.

### 2.5.7 Particle reconstruction

To build a description of the physics objects present in the particle collision, the basic elements from the detector layers (tracks and clusters of energy) are correlated to identify each particle in the final state. Measurements from different sub-detectors are combined to reconstruct the particle properties. This approach is called particle-flow (PF) reconstruction [23]. Key to the success of the PF reconstruction is the fine spatial granularity of the detector layers. Coarse-grained detectors can cause the signals from different particles to merge, especially within jets. However, if the subdetectors are sufficiently segmented to separate individual particles, it becomes possible to produce a global event description that identifies all physics objects with high efficiencies and resolution.

## **2.5.8 Data storage and computational infrastructure**

The LHC generates over 15 petabytes (15 million gigabytes) of data every year, necessitating a flexible computing system that can be accessed by researchers working at the four main LHC experiments: ALICE, ATLAS, CMS, and LHCb. The Worldwide LHC Computing Grid (WLCG) [37] is a global collaboration of computer centers that links thousands of computers and storage systems in over 170 centers across 41 countries. These centers are arranged in “tiers”, and provide near real-time access to users processing, analyzing, and storing LHC data. One of the final stages of data analysis at LHC experiments is large-scale data processing taking place over distributing computing, for instance, with the use of Condor [38], a distributed, scalable, flexible batch processing system which accepts a computing job, allocates a resource to it, executes it, and returns the result back to a user transparently.

1038 **Chapter 3**

1039 **The Phase-2 Upgrade of CMS**

1040 This chapter gives an overview of the High-Luminosity LHC upgrade of the LHC in  
1041 Section 3.1, and the upgrades for the Phase-2 CMS Level-1 (L1) Trigger in Section  
1042 3.2. One of the major upgrades is the new availability of calorimeter crystal-level  
1043 information to the L1 calorimeter trigger, compared to the current trigger which only  
1044 has access to tower-level information (a tower being 5 by 5 in crystals). To capitalize  
1045 on the increased spatial granularity of this information, an upgraded algorithm is  
1046 presented which reconstructs and identifies electron and photon candidates in the  
1047 Layer-1 Calorimeter Trigger. A description of the algorithm and a validation of its  
1048 performance in Phase-2 conditions is given in Section 3.3.

1049 **3.1 The High-Luminosity LHC**

1050 In order to sustain and extend the LHC’s physics discovery program and maintain  
1051 operability for a decade or more, the LHC is undergoing a major upgrade to the High-  
1052 Luminosity LHC (HL-LHC). In its final configuration, the HL-LHC will deliver a peak  
1053 luminosity of  $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , potentially leading to total integrated luminosity  
1054 of  $4000 \text{ fb}^{-1}$  after ten years of operations, scheduled to begin in 2027 [39]. This  
1055 integrated luminosity is about ten times the predicted luminosity reach of the LHC

1056 in its initial configuration. To enable the CMS experiment to continue operations and  
1057 data-taking and to maximize the discovery potential of the unprecedented amount  
1058 of data, the CMS detector is undergoing Phase-2 upgrades in order to perform high-  
1059 precision measurements and searches for physics beyond the Standard Model in the  
1060 intense running conditions of the HL-LHC.

## 1061 **3.2 The Phase-2 Level-1 Trigger**

1062 To achieve the goals of the HL-LHC program and to ensure the collection of information-  
1063 rich datasets in the HL-LHC, the Phase-2 upgrade of the CMS Level-1 Trigger [39]  
1064 must be upgraded in conjunction with the CMS sub-detectors and their readouts, to  
1065 maintain physics selectivity. The HL-LHC will produce an intense hadronic environ-  
1066 ment corresponding to 200 simultaneous collisions per beam crossing, necessitating  
1067 comprehensive upgrades of the trigger system outlined below.

1068 In order to cope with the increased pile-up and high occupancies of the HL-LHC,  
1069 the latency of the L1 trigger system (time available to produce a L1 Accept signal) will  
1070 be increased significantly from  $3.8 \mu\text{s}$  to  $12.5 \mu\text{s}$ , with an increased maximum output  
1071 bandwidth of 750 kHz [39]. With the increased latency, in addition to information  
1072 from calorimeters and muon detectors (as in the Phase-1 system), information from  
1073 the new tracker and high-granularity endcap calorimeter can also be included at L1  
1074 for the first time. This is illustrated in the functional diagram of the architecture of  
1075 the Phase-2 trigger system in Fig. 3.1.

1076 The key feature of the Phase-2 L1 Trigger is the introduction of a correlator layer,  
1077 where algorithms produce higher-level trigger objects by combining information from  
1078 sub-detectors, with a selectivity approaching that of offline reconstruction in the  
1079 HLT [39]. Four independent data processing paths (grouped together in Fig. 3.1) are  
1080 implemented: tracking, calorimetry, muon systems, and particle-flow techniques:

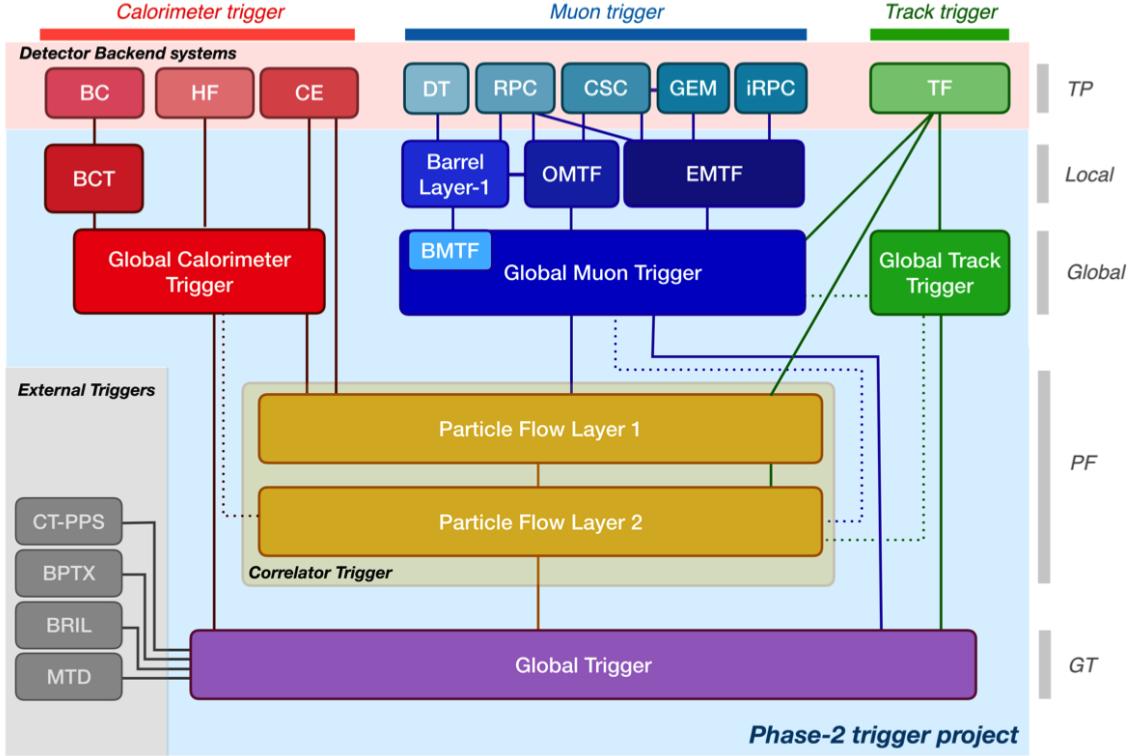


Figure 3.1: Functional diagram of the CMS L1 Phase-2 upgraded trigger design [39], showing the four trigger paths: calorimeter, muon, track, and Particle Flow. For the first time, tracking information will be available as early as the L1 Trigger.

- **Calorimeter Trigger path:** (red, Fig. 3.1) A barrel calorimeter trigger (BCT) and the HGCAL backend are used to process crystal-level information from the calorimeters to produce high-resolution clusters and identification variables used for later processing. Outputs from the BCT, HGCAL, and the HF are sent to a global calorimeter trigger (GCT), where calorimeter-only objects such as  $e/\gamma$  candidates, hadronically decaying tau lepton candidates, jets, and energy sums are built.
- **Track Trigger path:** (green, Fig. 3.1) Tracks from the Outer Tracker are reconstructed in the track finder (TF) processors as part of the detector backend. A global track trigger (GTT) will reconstruct the primary vertices of the event, along with tracker-only based objects, such as jets and missing transverse momentum.

1093     • **Muon Trigger path:** (*blue*, Fig. 3.1) Trigger primitives are processed by  
1094       muon track finder algorithms, again separated into the barrel (barrel muon  
1095       track finder, BMTF), overlap (overlap muon track finder, OMTF), and endcap  
1096       (endcap muon track finder, EMTF). Standalone muons and stubs containing  
1097       information such as position, bend angle, and timing, as well as L1 tracks, are  
1098       sent to the global muon trigger (GMT).

1099     • **Particle-Flow Trigger path:** (*yellow*, Fig. 3.1) The correlator trigger (CT)  
1100       aims to approach the performance of offline Particle Flow, and is implemented  
1101       in two layers. “Layer-1” produces the particle-flow candidates from matching  
1102       calorimeter clusters and tracks. “Layer 2” builds and sorts final trigger objects  
1103       and applies additional identification and isolation criteria.

1104       The outputs from the above trigger paths are combined in the Global Trigger  
1105       (GT) (*purple*, Fig. 3.1), which calculates the final trigger decision (Level-1 Accept),  
1106       transmitting it to the Trigger Control and Distribution System (TCDS), which dis-  
1107       tributes it to the detector backend systems, initiating the readout to the DAQ. The  
1108       GT also provides the interface to external triggers (*grey*, Fig. 3.1), such as trig-  
1109       gers for the precision proton spectrometer (PPS), beam position and timing monitors  
1110       (BPTX), and luminosity and beam monitoring (BRIL) detectors [39]. The design of  
1111       the Phase-2 Level-1 Trigger allows for future inclusion of triggering information, for  
1112       instance information about minimum ionizing particles (MIPs) from the MIP Timing  
1113       Detector (MTD) [40].

### **3.3 Standalone barrel calorimeter electron/photon reconstruction**

The reconstruction and identification of electrons and photons ( $e/\gamma$ ) begin with the trigger primitives of the barrel ECAL and HCAL detectors and endcap HGCAL calorimeters, covering the pseudorapidity region  $|\eta| < 3$ . The barrel and endcap regions of the detector are intrinsically different enough to warrant different approaches to  $e/\gamma$  reconstruction. This work presents a firmware-based emulator for the standalone  $e/\gamma$  reconstruction in the barrel calorimeter (Fig. 3.2). “Standalone” refers to the fact that the tracker information is not used in this particular reconstruction chain. This firmware-based emulator is based on the parallelized, computational logic that will be deployed in the firmware of the Phase-2 Level-1 trigger. The emulator uses fixed-precision integers to represent all values, such as in the computation of cluster energies, and closely mimics the firmware logic which uses arrays and performs computations in flattened loops. It represents an improved, more realistic understanding of the trigger, compared to the previous emulator which used idealized logic such as vector operations, and floats to represent all values [39].

#### **3.3.1 Electron/photon standalone barrel procedure**

In Phase-2, the upgrade of both on-detector and off-detector electronics of the barrel calorimeters’ trigger primitive generator (TPG) will enable the streaming of single crystal data from the on-detector to the backend electronics. Currently in Phase-1, the ECAL and HCAL TPGs is restricted to providing lower-granularity information of trigger tower sums of  $5 \times 5$  crystals to the Level-1 Trigger [39]. A schematic of the geometry of the ECAL barrel in the Phase-2 Regional Calorimeter Trigger (RCT) is shown in Fig. 3.3. The barrel is spanned by 36 RCT cards, each spanning  $17 \times 4$  towers in  $\eta \times \phi$ . Each RCT card is subdivided into five “regions” as shown in Fig.

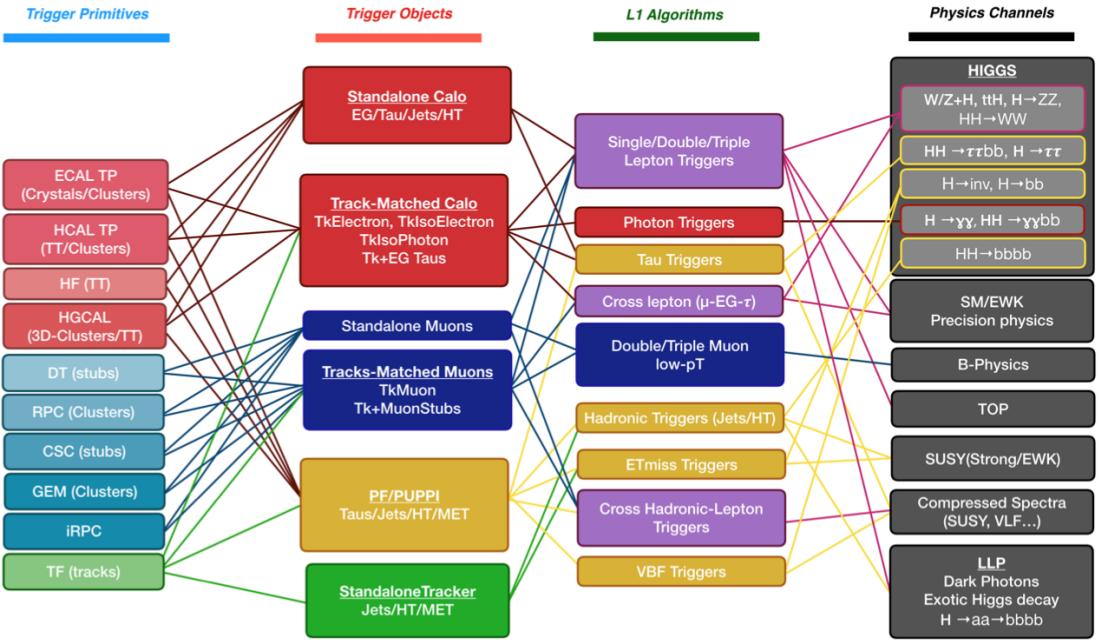


Figure 3.2: Summary of the links between the trigger primitives (*first column*), the trigger objects (*second column*), the Level-1 algorithms used in the menu (*3rd column*), and the physics channels (*4th column*), from [39], where a full description of the Phase-2 L1 algorithms can be found. This work focuses on developments for the Standalone Calorimeter electron and photon ("EG") reconstruction algorithm.

1139     3.4. After initial clustering and processing, the outputs of the RCT card are sent to  
 1140     the Global Calorimeter (GCT) trigger, which is processed in three cards as shown in  
 1141     Fig. 3.5. The reconstruction algorithm is detailed below.

1142     The standalone barrel algorithm for reconstructing and identifying electrons and  
 1143     photons in the Phase-2 Level-1 Trigger takes as input the digitized response of each  
 1144     crystal of the barrel ECAL, with a granularity  $0.0175 \times 0.0175$  in  $\eta \times \phi$ , which is 25  
 1145     times higher than the input to the Phase-1 trigger, which consisted of trigger towers  
 1146     with a granularity of  $0.0875 \times 0.0875$ . In HCAL the tower size of  $0.0875 \times 0.0875$   
 1147     is unchanged. The trigger algorithm is designed to closely reproduce the algorithm  
 1148     used in the offline reconstruction, with limitations and simplifications due to trigger  
 1149     latency.

1150     In the RCT, an initial requirement of  $p_T > 0.5$  GeV is imposed on the input

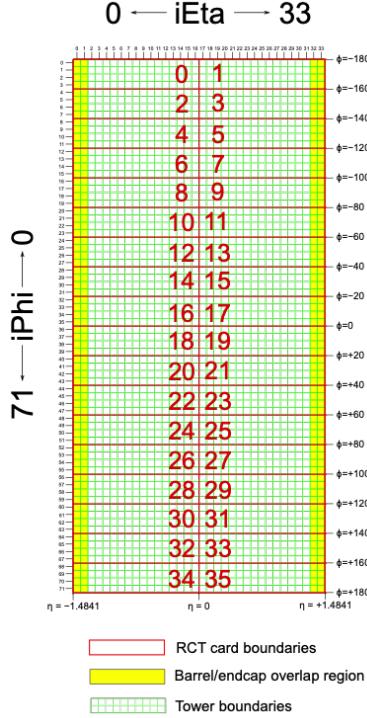


Figure 3.3: Schematic of the geometry of the Phase-2 ECAL barrel in the Regional Calorimeter Trigger (RCT), showing the division of the barrel region into 36 Regional Calorimeter Trigger (RCT) cards (*red*). Each card spans  $17 \times 4$  towers in  $\eta \times \phi$  (*green*), and each tower is  $5 \times 5$  in single crystals in  $\eta \times \phi$ . Towers in the overlap region (*shaded yellow*) are read out to both the barrel and endcap.

trigger primitives (i.e. energies from the ECAL crystals and HCAL towers) to reject contribution from pile-up. In one of the regions inside a RCT card (Fig. 3.4), the crystal containing the highest energy deposit is identified as the seed crystal, as shown in Fig. 3.6. The energy in the crystals in a window of size  $3 \times 5$  in  $\eta \times \phi$  around the seed cluster is added into a cluster. The energy is considered “clustered”. The process is repeated with the remaining “unclustered” energy, until up to four clusters are produced in the region.

To improve  $e/\gamma$  identification and to reduce background contributions, identification and reconstruction algorithms are implemented at this stage:

- Shower shape: The energy deposit sums around the seed crystal is computed in windows of size  $2 \times 5$  and  $5 \times 5$  (Fig. 3.6, *dashed lines*), with true  $e/\gamma$  clusters

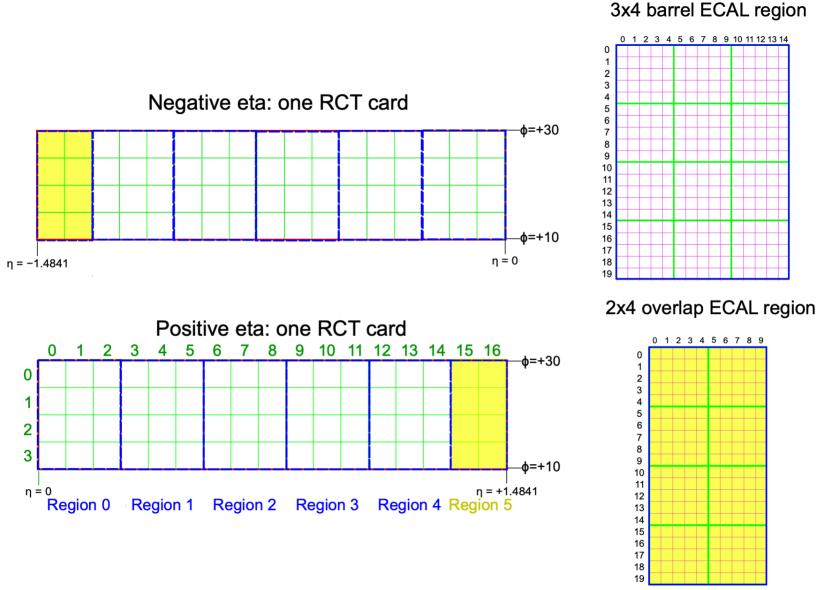


Figure 3.4: Schematic of two example RCT cards in the negative eta (*top left*) and positive eta (*bottom left*) regions of the ECAL barrel. Each RCT card is divided into six regions: five regions are of size  $3 \times 4$  towers in  $\eta \times \phi$  (*top right*), and a sixth smaller overlap region of size  $2 \times 4$  towers (*bottom right*). Each tower is  $5 \times 5$  ( $\eta \times \phi$ ) in crystals.

1162 tending to produce showers that deposit most of their energy in a  $2 \times 5$  region.

- 1163 • Bremsstrahlung recovery:  $e/\gamma$  tend to spread in the  $\phi$  direction due to charged  
 1164 particles being bent by the magnetic field of the CMS solenoid. If sufficient  
 1165 energy comparable to the core  $3 \times 5$  cluster is found in the adjacent  $3 \times 5$   
 1166 windows (Fig. 3.6, *shaded yellow*), the energy is added to the core cluster and  
 1167 no longer considered unclustered energy.

1168 After parallel processing in the regions, the clusters in a RCT card are stitched  
 1169 together if they are located directly along the borders of a region (Fig. 3.3). The  
 1170 remaining unclustered ECAL energy is summed into ECAL towers.

1171 From each RCT card, the twelve highest-energy clusters, as well as any remaining  
 1172 unclustered energy, are sent to the GCT. Since each GCT card has information from  
 1173 sixteen RCT cards (Fig. 3.5), final stitching across the boundaries of the RCT cards  
 1174 is performed. One more identification algorithm is performed at this stage:

## GCT/“Layer 2”

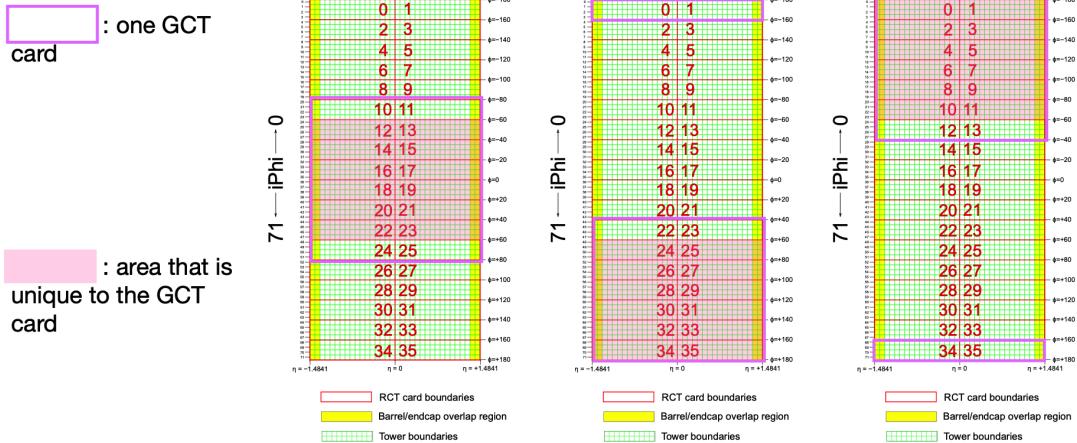


Figure 3.5: Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trigger (GCT), which will process the outputs of the Regional Calorimeter Trigger (RCT) in three GCT cards (*purple borders*). Each card in the GCT processes the equivalent of sixteen RCT cards, with the center twelve RCT cards being unique to that GCT card (*shaded pink*), and the remaining four RCT cards overlapping with one other GCT card.

1175     • Isolation: One handle to reject backgrounds from e.g. pile-up, comes from the  
 1176       tendency for background to be spread more uniformly across a large area in the  
 1177       detector, whereas genuine  $e/\gamma$  are expected to produce showers concentrated in  
 1178       the  $3 \times 5$  crystal window. The energy sum in a large window of  $7 \times 7$  in towers  
 1179       is computed and used to reject background.

1180     Flags that provide discrimination power between genuine  $e/\gamma$  and background, are  
 1181       computed using the relative isolation and shower shape quantities. The standalone  
 1182       working point (WP) is defined as the logical OR of the relative isolation and shower  
 1183       shape flags.

1184     The information of the clusters in the event, including their energies, crystal-level  
 1185       position, the relative isolation flag, the shower shape flag, the standalone WP, and  
 1186       the ratio of the HCAL over ECAL energies, are sent in 64 bits to the Correlator  
 1187       Trigger and the Global Trigger. The towers in the event are computed as the sum  
 1188       of all unclustered energy in the ECAL with the corresponding HCAL energy at each

### 3x4 barrel ECAL region

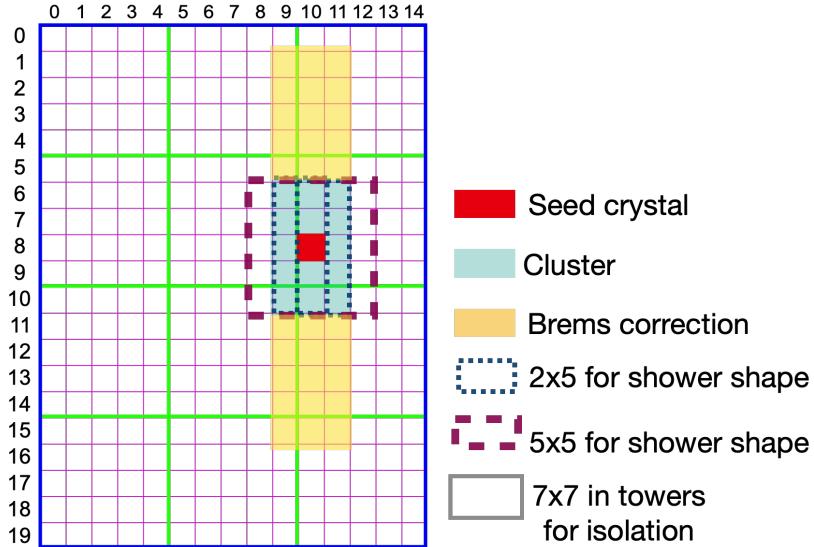


Figure 3.6: Illustration of an example electron/photon ( $e/\gamma$ ) cluster in the Phase-2 Level-1 Trigger standalone barrel  $e/\gamma$  reconstruction, in a region of  $15 \times 20$  crystals ( $3 \times 4$  towers) in  $\eta \times \phi$ . Each small pink square is one crystal, the highest-granularity ECAL trigger primitives available to the L1 Trigger in Phase-2. The core cluster consists of the energy sum in a  $3 \times 5$  window of crystals (*shaded light blue*), centered around the seed crystal (*red*). The presence of energy lost to bremsstrahlung radiation is checked in the adjacent  $3 \times 5$  windows in the  $\phi$  direction (*shaded light yellow*). The ratio of the total energies in windows of size  $2 \times 5$  and  $5 \times 5$  in crystals (*dashed dark blue and dark red*) around the seed crystal, is computed and compared to the core cluster energy to obtain shower shape flags. Lastly, the isolation, defined as the sum of the energy in a large window of size  $7 \times 7$  in towers (not shown in figure) is computed, and compared to the core cluster energy to obtain isolation flags.

<sup>1189</sup> tower location, and their energies are sent to the Correlator Trigger.

#### <sup>1190</sup> 3.3.2 Electron/photon standalone barrel results

<sup>1191</sup> The performance of the current emulator of the standalone barrel  $e/\gamma$  algorithm in  
<sup>1192</sup> Phase-2 conditions is quantified in efficiencies and rates. Efficiency is the fraction of  
<sup>1193</sup> true electrons that the algorithm can reconstruct and identify, and is evaluated in  
<sup>1194</sup> a Monte Carlo simulated sample containing electrons with transverse momentum  $p_T$   
<sup>1195</sup> ranging from 1 to 100 GeV. The efficiencies of the current and previous emulators as

1196 a function of the electron generator-level  $p_T$  are shown in Fig. 3.7.

1197 The rates are the event rates that this reconstruction and identification algorithm  
1198 would obtain if it were deployed in a trigger, assuming that proton-proton collisions  
1199 are occurring at the 40 MHz event rate of the HL-LHC. The rate is reported as a  
1200 function of the minimum energy threshold required by the trigger, and is estimated  
1201 using a simulated sample of minimum bias events, i.e. generic proton-proton colli-  
1202 sions without any specific physics selections. The rates for the current and previous  
1203 emulator are shown in Fig. 3.8.

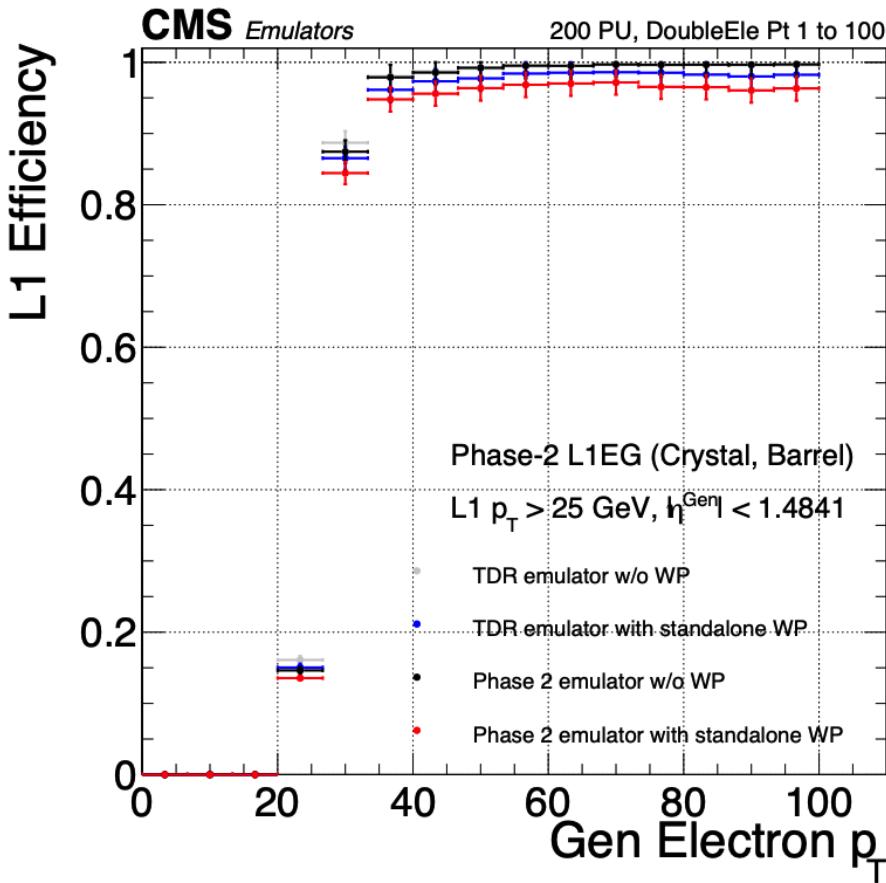


Figure 3.7: Efficiencies of the current and previous emulators of the standalone barrel  $e/\gamma$  algorithm for the Phase-2 Level-1 Trigger, evaluated in a simulated sample containing electrons, as a function of the electron's generator-level transverse momentum  $p_T$ . The standalone working point (WP) is defined as the logical OR of the isolation flag and shower shape flag. The efficiencies with and without requiring the standalone WP, are shown for the current emulator (labeled “Phase 2”, *black, red*) and the previous emulator (labeled “TDR”, *dark blue, grey*).

1204        The current emulator is incorporated into the full Phase-2 L1 menu, allowing an  
1205        estimate of the rates produced by the standalone  $e/\gamma$  barrel trigger path and all  
1206        other algorithms in the L1 Trigger. All rates are estimated with the assumption of  
1207        an average pile-up of 200 and event rate of 40 MHz. The standalone working point  
1208        single  $e/\gamma$  path with requirements on the  $e/\gamma$  candidate to have  $|\eta| < 2.4$ , offline  $p_T$   
1209        to be greater than 51 GeV, and online  $p_T$  to be greater than 41 GeV, is projected to  
1210        have a rate of around 23 kHz. The standalone working point double  $e/\gamma$  path with  
1211        requirements on the two  $e/\gamma$  candidates to have  $|\eta| < 2.4$ , offline  $p_T$  greater than 37  
1212        and 24 GeV, and online  $p_T$  greater than 29 and 18 GeV, is projected to give a rate of  
1213        around 6 kHz. For both paths, the objects efficiency plateau is 99%.

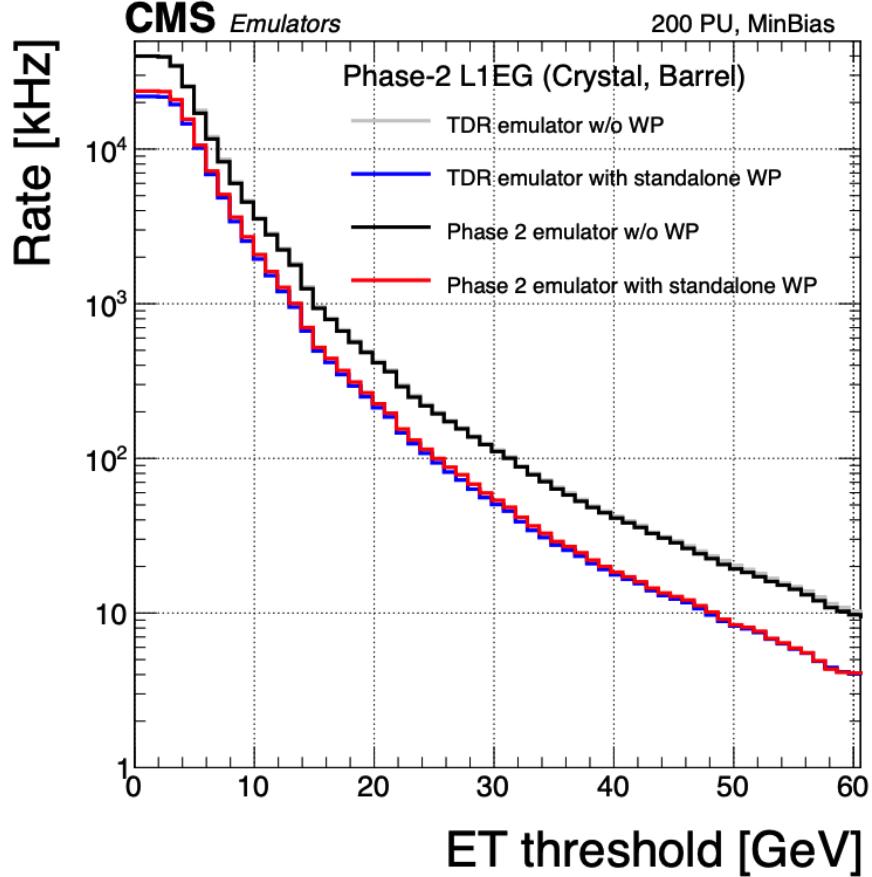


Figure 3.8: Rates in kHz of the current Phase-2 and previous (“TDR”) emulators of the standalone barrel  $e/\gamma$  algorithm for the Phase-2 Level-1 Trigger, evaluated on a minimum bias (MinBias) sample with 200 pile-up (PU), measured as a function of the minimum energy ( $E_T$ ) required of the reconstructed  $e/\gamma$  object in each event. The standalone working point (standalone WP) is defined to be the logical OR of the isolation flag and the shower shape flag. The rates with and without requiring the standalone WP, are shown for the current emulator (labeled “Phase 2”, *black, red*) and the previous emulator (labeled “TDR”, *dark blue, grey*).

<sub>1214</sub> **Chapter 4**

<sub>1215</sub> **Datasets and Monte Carlo samples**

<sub>1216</sub> This search for the exotic decay of the 125 GeV Higgs boson to two light neutral scalars  
<sub>1217</sub> decaying to a pair of bottom quarks and a pair of tau leptons ( $h \rightarrow aa \rightarrow bb\tau\tau$ ) is  
<sub>1218</sub> based on proton-proton collision data at a center-of-mass energy 13 TeV collected  
<sub>1219</sub> in Run-2 of data-taking, spanning the data-taking years 2016, 2017, and 2018. The  
<sub>1220</sub> datasets used and the triggers used to collect the data are described in Section 4.1.  
<sub>1221</sub> Section 4.2 describes the Monte Carlo simulated samples that are used to model the  
<sub>1222</sub>  $h \rightarrow aa \rightarrow bb\tau\tau$  signal and background Standard Model processes. Lastly, in order  
<sub>1223</sub> to obtain a better description of Standard Model backgrounds that contain two tau  
<sub>1224</sub> leptons, a data-Monte Carlo hybrid technique is used to generate embedded samples  
<sub>1225</sub> which model processes with genuine  $\tau\tau$  in the final state, as detailed in Section 4.3.  
<sub>1226</sub> All samples are listed in Appendix A.

<sub>1227</sub> **4.1 Datasets used**

<sub>1228</sub> The  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis [41] is based on proton-proton collision data at a center-  
<sub>1229</sub> of-mass energy of 13 TeV collected in full Run-2 (2016-18) with the CMS detector.  
<sub>1230</sub> The data analyzed corresponds to a total integrated luminosity of  $138 \text{ fb}^{-1}$  ( $36.33 \text{ fb}^{-1}$   
<sub>1231</sub> for 2016,  $41.53 \text{ fb}^{-1}$  for 2017, and  $59.74 \text{ fb}^{-1}$  for 2018) [42] [43] [44]. The cumulative

1232 delivered and recorded luminosity versus time for 2015-2018 is shown in Fig. 4.1.

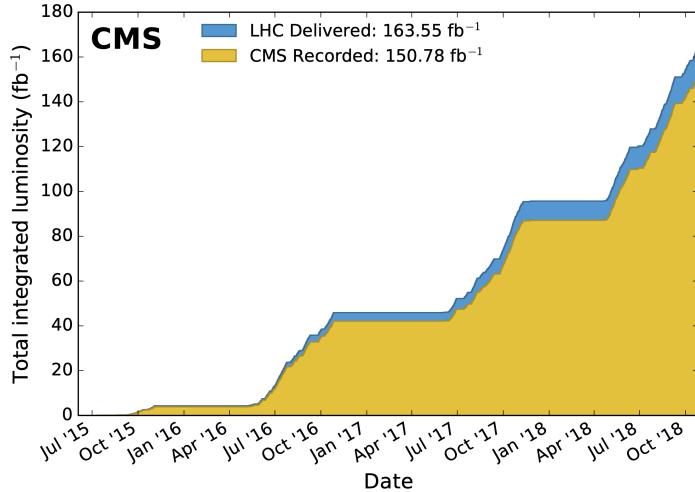


Figure 4.1: Cumulative delivered and recorded luminosity versus time for 2015-2018 at CMS, in proton-proton collision data only, at nominal center-of-mass energy [45].

1233 Data collected with the single muon trigger is used for the  $\mu\tau_h$  channel. For the  
1234  $e\tau_h$  channel, data collected with the single electron trigger is used; and for the  $e\mu$   
1235 channel, data collected with the electron + muon trigger is used. A more in-depth  
1236 discussion of the triggers used follows in a later section. The datasets are listed in  
1237 Appendix A in Tables A.1, A.2, and A.3.

## 1238 4.2 Monte Carlo samples

1239 Modeling and computing observables originating from arbitrary physics processes at  
1240 the tree level and at next-to-leading order (NLO) is performed by Monte Carlo (MC)  
1241 event generators, such as Powheg and MadGraph5\_amCNLO [46] [47]. The informa-  
1242 tion generated, e.g. the computation of the differential cross sections and kinematics  
1243 of the final state particles, is saved in a compressed file and used to generate MC sam-  
1244 ples that are used in physics analyses. The samples are digitized using GEANT4 [48],  
1245 a platform used at the LHC and other facilities to comprehensively simulate the

passage of particles through matter. The digitized samples are passed through the same detector reconstruction as real data events collected in the detector. The MC background samples used in this analysis for 2016-2018 are listed in Appendix A in Tables A.7, A.8, and A.9.

The Monte Carlo samples for modeling the signal ( $h \rightarrow aa \rightarrow 2b2\tau$ ) are generated at tree-level, for mass hypotheses of the  $a$  ranging from  $m_a = 12$  GeV to 60 GeV. The MC signal samples used in this analysis for 2016-2018 are listed in Appendix A in Tables A.10, A.11 and A.12.

### 4.3 Embedded samples

An important background for Higgs boson studies and searches for additional Higgs bosons is the decay of  $Z$  bosons into pairs of  $\tau$  leptons ( $Z \rightarrow \tau\tau$ ). An embedded technique was developed in the context of Standard Model Higgs to  $\tau\tau$  measurements, to model  $Z \rightarrow \tau\tau$  decays, and was expanded to also model all Standard Model processes that contain  $\tau\tau$  [49]. The embedded technique has since been used successfully at CMS for the Standard Model  $H \rightarrow \tau\tau$  measurement, as well as searches for minimal supersymmetric extensions to the Standard Model (MSSM) [50] [51].

The advantage of the embedded technique is that aspects of the event that are difficult to model and describe are directly taken from data, resulting in a better data description than can be achieved with only the  $Z \rightarrow \tau\tau$  simulation [49]. The simulation must be tuned extensively to accurately model aspects of the data, such as time-dependent pile-up profiles, the production of additional jets, e.g. in multijet and vector boson fusion topologies, the number of reconstructed primary interaction vertices, and the missing transverse momentum  $p_T^{\text{miss}}$ . Since all events with genuine  $\tau\tau$  are estimated with samples made with the embedded technique (referred to as embedded samples from here on), events in Monte Carlo simulation with genuine  $\tau\tau$

1271 are not used, in order to avoid double-counting.

1272 Fig. 4.2 shows a schematic of how embedded samples are produced. Data events  
1273 containing  $Z \rightarrow \mu\mu$  decays are selected. In these events, all energy deposits of the  
1274 recorded muons are removed, and are replaced with simulated tau leptons with the  
1275 same kinematic properties as the removed muons. This results in a hybrid data format  
1276 containing information from both observed and simulated events, as illustrated in Fig.  
1277 4.2 [49]. The embedded samples used for the years 2016-2018 are listed in Appendix  
1278 A in Tables A.4, A.5, and A.6.

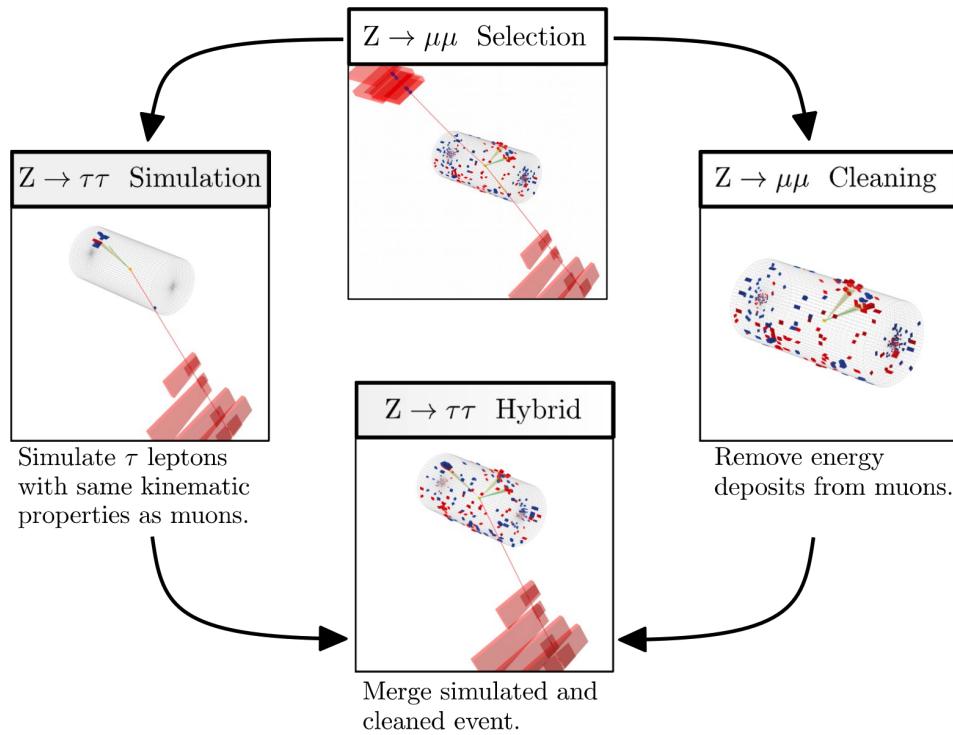


Figure 4.2: Schematic view of the four main steps of the embedding technique for  $\tau$  leptons, as described in Section 4.3 [49]. A  $Z \rightarrow \mu\mu$  event is selected in data ( $Z \rightarrow \mu\mu$  selection), all of the energy deposits associated with the muons are removed ( $Z \rightarrow \mu\mu$  cleaning), and two  $\tau$  leptons and their decays are simulated in an empty detector ( $Z \rightarrow \tau\tau$  simulation). Lastly, all energy deposits of the simulated  $\tau$  decays are combined with the data event ( $Z \rightarrow \tau\tau$  hybrid).

1279 In the selection step of the embedded technique, events are selected with at least  
1280 one of a set of di-muon trigger paths, which require  $p_T > 17(8)$  GeV for the leading  
1281 (sub-leading) muons, and a minimum requirement between 3.8 and 8.0 GeV on the

invariant di-muon mass  $m_{\mu\mu}$  [49]. The offline reconstructed muons must match the objects at trigger level and also have offline  $p_T > 17(8)$  GeV. They must have  $|\eta| < 2.4$  and be located at a distance  $|d_z| < 0.2$  cm to the primary vertex along the beam axis. To form a Z boson candidate, each muon is required to originate from a global muon track. The muon pairs must have opposite charges with an invariant mass of  $m_{\mu\mu} > 20$  GeV. If more than two di-muon pairs are found, the pair with the invariant mass closest to the Z boson mass (91.19 GeV) is chosen.

This selection is designed to be tight enough to ensure a high purity of genuine  $\mu\mu$  events, and also loose enough to minimize biases of the embedded event samples. Isolation requirements are avoided, since they would introduce a bias towards less hadronic activity in the vicinities of the embedded leptons that will appear more isolated than expected in data. The selection results in an expected mixture of events summarized in Table 4.1 from [49].  $Z \rightarrow \mu\mu$  is the dominant process modeled by the embedded technique, with  $t\bar{t}$ , QCD, and diboson and single top processes becoming more significant when considering events with b-tag jets.

Fraction (%)			
Process	Inclusive	$m_{\mu\mu} > 70$ GeV	N(b-tag jets) > 0
$Z \rightarrow \mu\mu$	97.36	99.11	69.25
QCD	0.84	0.10	2.08
$t\bar{t}$	0.78	0.55	25.61
$Z \rightarrow \tau\tau$	0.71	0.05	0.57
Diboson, single t	0.17	0.17	2.35
W+jets	0.08	0.02	0.14

Table 4.1: Expected event composition after selecting two muons in the embedded technique [49], before additional cuts (i.e. inclusive, *column 2*), and after adding a requirement on the di-muon mass  $m_{\mu\mu} > 70$  GeV (*column 3*), or a requirement on the number of b-tag jets in the event (*column 4*).

<sub>1297</sub> **Chapter 5**

<sub>1298</sub> **Object reconstruction and  
corrections applied**

<sub>1300</sub> In the data processing workflow, data events and simulated events are analyzed to  
<sub>1301</sub> reconstruct physics objects of interest, and algorithms for distinguishing genuine par-  
<sub>1302</sub> ticle candidates from background, are employed. Section 5.1 describes the physical  
<sub>1303</sub> properties of the most important objects in the  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis: taus,  
<sub>1304</sub> muons, electrons, jets, and jets originating from b-quarks (b-flavor jets), as well as  
<sub>1305</sub> their reconstruction and identification in CMS. In this analysis, the full energy and  
<sub>1306</sub> momentum of the two tau leptons ( $m_{\tau\tau}$ ) is estimated from the measured (i.e. visible)  
<sub>1307</sub> components of the tau leptons using the SVFit/FastMTT algorithm, which is de-  
<sub>1308</sub> scribed in Section 5.2. Corrections are applied to the simulated samples at the object  
<sub>1309</sub> level and the event level to account for known discrepancies between simulations and  
<sub>1310</sub> the data that the simulations are intended to model. These corrections are listed and  
<sub>1311</sub> detailed in Section 5.3.

<sub>1312</sub> **5.1 Object reconstruction**

<sub>1313</sub> **5.1.1 Taus**

<sub>1314</sub> The tau ( $\tau$ ) is the heaviest known lepton. With a rest mass of 1776.86 MeV, it can  
<sub>1315</sub> decay to not only electrons and muons, but also hadrons. In two thirds of the cases,  $\tau$   
<sub>1316</sub> leptons decay hadronically, typically into one or three charged mesons (predominantly  
<sub>1317</sub>  $\pi^+$ ,  $\pi^-$ ), often accompanied by neutral pions (that decay  $\pi^0 \rightarrow \gamma\gamma$ ), and a  $\nu_\tau$ . These  
<sub>1318</sub> hadronic decays are denoted  $\tau_h$ . In the remaining cases, the tau decays to the lighter  
<sub>1319</sub> leptons (electron or muon), termed leptonic decays. The mean lifetime of the  $\tau$  is  
<sub>1320</sub>  $\tau = 290 \times 10^{-15}$  seconds, corresponding to  $c\tau = 87.03 \mu\text{m}$ , which is short enough that  
<sub>1321</sub> taus decay in the CMS detector before reaching the detector elements, but also long  
<sub>1322</sub> enough that some decay length variables can help with hadronic tau identification.  
<sub>1323</sub> The tau's largest decay branching ratios (proportional to probability of decay) are  
<sub>1324</sub> listed below [27]:

- <sub>1325</sub> • 17.8% decay to  $e^- \bar{\nu}_e \nu_\tau$
- <sub>1326</sub> • 17.4% decay to  $\mu^- \bar{\nu}_\mu \nu_\tau$
- <sub>1327</sub> • 25.5% decay to  $\pi^- \pi^0 \nu_\tau$  ( $\rho^-$  resonance at 770 MeV)
- <sub>1328</sub> • 10.8% decay to  $\pi^- \nu_\tau$
- <sub>1329</sub> • 9.3% decay to  $\pi^- \pi^0 \pi^0 \nu_\tau$  ( $a_1^-$  resonance at 1200 MeV)
- <sub>1330</sub> • 9.0% decay to  $\pi^- \pi^- \pi^+ \nu_\tau$  ( $a_1^-$  resonance at 1200 MeV)

<sub>1331</sub> In all cases, at least one neutrino is produced. The neutrinos escape undetected  
<sub>1332</sub> from the CMS detector, resulting in missing transverse energy. Charged hadrons leave  
<sub>1333</sub> tracks in the tracking detector before being absorbed in the hadronic calorimeter; in  
<sub>1334</sub> CMS tau reconstruction terminology, they are often called “prongs”, i.e. the dominant

1335  $\tau_h$  decay modes are termed “1 prong” ( $\pi^\pm$ ), “1 prong +  $\pi^0$ (s)”, and “3-prong”. Neutral  
1336 pions decay to two photons which lose their energy in the electromagnetic calorimeter.  
1337 Taus that decay to electrons and muons, are typically triggered on and reconstructed  
1338 as electrons and muons respectively.

1339 **Hadron plus strips (HPS) reconstruction of  $\tau_h$**

1340 At CMS, hadronically decaying tau leptons are reconstructed with the hadron plus  
1341 strips (HPS) algorithm [52] [53]. The HPS algorithm capitalizes on photon conversions  
1342 in the CMS tracker material, which originate from the neutral pion ( $\pi^0$ ) decaying  
1343 to two photons. The bending of electron/positron tracks due to the CMS solenoid  
1344 magnetic field leads to a spread of the neutral pions’ calorimeter signatures in the  $\phi$   
1345 direction. This motivates the reconstruction of photons in “strips”: objects that are  
1346 built out of PF photons and electrons. The strip reconstruction starts with centering  
1347 a strip on the most energetic electromagnetic particle in a PF jet. Among other  
1348 electromagnetic particles located in a window of size  $\Delta\eta = 0.05$  and  $\Delta\phi = 0.20$   
1349 around the strip center, the most energetic one is associated with the strip and its  
1350 momentum is added to the strip momentum. This is repeated iteratively until no  
1351 further particles can be associated. Lastly, strips satisfying a requirement of  $p_T^{\text{strip}} >$   
1352 1 GeV are combined with charged hadrons to reconstruct individual  $\tau_h$  decay modes,  
1353 where  $h$  stands for both  $\pi$  and  $K$ :

- 1354 • *Single hadron:*  $h^- \nu_\tau$  and  $h^- \pi^0 \nu_\tau$  decay modes, in which the neutral pions have  
1355 too little energy to be reconstructed as strips.
- 1356 • *One hadron + one strip:*  $h^- \pi^0 \nu_\tau$  decay modes, where the photons from the  $\pi^0$   
1357 decay are close together in the calorimeter.
- 1358 • *One hadron + two strips:*  $h^- \pi^0 \nu_\tau$  decay modes, where the photons from the  $\pi^0$   
1359 decay are well separated.

- 1360       • *Three hadrons:*  $h^-h^+h^-\nu_\tau$  decay modes. The three charged hadrons are re-  
1361              quired to originate from the same secondary vertex.

1362     The  $h^-\pi^0\pi^0\nu_\tau$  and  $h^-h^+h^-\pi^0\nu_\tau$  decay modes do not have their own treatment, and  
1363     are reconstructed with the above topologies.

1364     In the HPS algorithm, the direction of the reconstructed tau momentum  $\vec{p}^{\tau_h}$   
1365     is required to fall within a distance of  $\Delta R = 0.1$  from the original PF jet. All  
1366     charged hadrons and strips are required to be contained within a cone of size  $\Delta R =$   
1367      $(2.8 \text{ GeV})/p_T^{\tau_h}$ , from the  $\tau_h$  as reconstructed by the HPS.

1368     All charged hadrons are assumed to be pions, and they are required to be consis-  
1369     tent with the masses of the intermediate meson resonances (if applicable), with the  
1370     following allowed windows for candidates: 50-200 MeV for  $\pi^0$ , 0.3-1.3 GeV for  $\rho$ , and  
1371     0.8-1.5 GeV for  $a_1$ . If the  $\tau_h$  decay is compatible with more than one hypothesis, the  
1372     one giving the highest  $p_T^{\tau_h}$  is chosen. Lastly, an isolation requirement is applied: aside  
1373     from the  $\tau_h$  decay products, no charged hadrons or photons can be present within  
1374     an isolation cone of size  $\Delta R = 0.5$  around the direction of the  $\tau_h$ . The outputs of  
1375     the HPS algorithm are the reconstructed decay mode and the visible four-momentum  
1376     (i.e. the four-momenta of all decay products excluding the neutrinos).

### 1377     **DeepTau for identifying $\tau_h$**

1378     The identification of  $\tau_h$  candidates in CMS has historically been divided into separate  
1379     discriminators against jets, electrons, and muons. Discriminators versus jets and  
1380     electrons use information from derived quantities, such as the  $p_T$  sum of particles  
1381     near the  $\tau_h$  axis. Building on the previous multivariate analysis (MVA) classifier [54]  
1382     based on a boosted decision tree (BDT), DeepTau is a more recent classifier based on a  
1383     deep neural network (DNN) that simultaneously discriminates against jets, electrons,  
1384     and muons. The DNN uses a combination of high-level inputs, similar to previous  
1385     algorithms, and also uses convolutional layers in  $\eta$ - $\phi$  space to process information

1386 from all reconstructed particles near the  $\tau_h$  axis. Convolutional layers are based on  
1387 the principle that an image can be processed independently of its position.

1388 The final DeepTau discriminators against jets, muons, and electrons are given by

$$D_\alpha(y) = \frac{y_\tau}{y_\tau + y_\alpha} \quad (5.1)$$

1389 where  $y_\tau$  ( $y_\alpha$ ) are estimates of the probabilities for the  $\tau_h$  candidate to come from  
1390 a genuine  $\tau_h$  (jet,  $\mu$ ,  $e$ ). Working points for each discriminator with different  $\tau_h$   
1391 identification efficiencies are defined for  $D_e$ ,  $D_\mu$ , and  $D_{\text{jet}}$ , for usage in physics analyses  
1392 and derivation of data-to-simulation corrections [55].

### 1393 5.1.2 Muons

1394 Muons are the next lightest lepton after taus, with a mass of 105.66 MeV and a mean  
1395 lifetime of  $\tau = 2.20 \times 10^{-6}$  seconds, or  $c\tau = 658.64$  meters. At CMS, muons are  
1396 identified with requirements on the quality of the track reconstruction and on the  
1397 number of measurements in the tracker and the muon systems [56]. In the standard  
1398 CMS reconstruction, tracks are first reconstructed independently in the inner tracker  
1399 (tracker track) and in the muon system (standalone-muon track). Next, these tracks  
1400 are processed in two different methods.

1401 The first is Global Muon reconstruction (outside-in) [56], which fits combined hits  
1402 from the tracker track and standalone-muon track, using the Kalman-filter technique.  
1403 At large transverse momenta,  $p_T \gtrsim 200$  GeV, the global-muon fit can improve the  
1404 momentum resolution compared to the tracker-only fit.

1405 The second is Tracker Muon reconstruction (inside-out) [56], which starts with  
1406 tracker tracks with  $p_T > 0.5$  GeV and total momentum  $p_T > 2.5$  GeV. These tracks  
1407 are extrapolated outwards to the muon system and matched to detector segments  
1408 there, taking into account the magnetic field, expected energy losses, and multiple

1409 Coulomb scattering in the detector material. Tracker Muon reconstruction is more  
1410 efficient than the Global Muon reconstruction at low momenta,  $p \lesssim 5 \text{ GeV}$ , because  
1411 it only requires a single muon segment in the muon system, whereas Global Muon  
1412 reconstruction typically requires segments in at least two muon stations.

1413 To further suppress fake muons from decay in flight, isolation cuts are used. A  
1414 relative isolation variable is defined to quantify the energy flow of particles near the  
1415 muon trajectory. A relative isolation is defined similarly for muons and electrons:

$$I^\ell \equiv \frac{\sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutral}} p_T - \frac{1}{2} \sum_{\text{charged, PU}} p_T\right)}{p_T^\ell} \quad (5.2)$$

1416 where  $\sum_{\text{charged}} p_T$  is the scalar sum of the  $p_T$  of the charged particles originating from  
1417 the primary vertex and located in a cone of size  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4(0.3)$   
1418 centered on the direction of the muon (electron). The sum  $\sum_{\text{neutral}} p_T$  is the equivalent  
1419 for neutral particles. The sum  $\sum_{\text{charged, PU}} p_T$  is the scalar sum of the  $p_T$  of the  
1420 charged hadrons in the cone originating from pile-up vertices. The factor 1/2 comes  
1421 from simulation estimations, which find that the ratio of neutral to charged hadron  
1422 production in the hadronization process of inelastic  $pp$  collisions is 1/2. Thus the  
1423 subtracted term is intended to subtract contribution from pile-up, from the neutral  
1424 particle contribution to the isolation sum. Finally, this is divided by the lepton  
1425 transverse momentum,  $p_T^\ell$ .

### 1426 5.1.3 Electrons

1427 Electrons are the lightest lepton with a mass of 0.511 MeV. At CMS, electrons are  
1428 reconstructed by associating a track reconstructed in the silicon tracking detector  
1429 with a cluster of energy in the ECAL. Performance is maximized via a combination  
1430 of a stand-alone approach and the complementary global particle-flow approach [57].

1431 In the stand-alone approach, the electron energy, which is typically spread over

1432 several crystals of the ECAL, is clustered with the “hybrid” algorithm in the barrel  
1433 and the “multi- $5 \times 5$ ” in the endcaps [57]. The hybrid algorithm collects energy in a  
1434 small window in  $\eta$  and an extended window in  $\phi$ . It identifies a seed crystal, and adds  
1435 arrays of  $5 \times 1$  crystals in  $\eta \times \phi$  in a range of  $N = 17$  crystals in both directions of  
1436  $\phi$ , if their energies exceed a minimum threshold, thus forming a supercluster (SC). In  
1437 the endcap, crystals are not arranged in an  $\eta \times \phi$  geometry; instead clusters are build  
1438 around seed crystals in clusters of  $5 \times 5$  crystals that can partly overlap. Nearby  
1439 clusters are grouped into a supercluster, and energy is recovered from associated  
1440 deposits in the preshower.

1441 In the PF reconstruction [57], PF clusters are reconstructed by aggregating around  
1442 a seed all contiguous crystals with energies two standard deviations above the elec-  
1443 tronic noise observed at the beginning of a data-taking run. The energy of a given  
1444 crystal can be shared among two or more clusters.

1445 The electron track reconstruction is performed in two ways [57]: the ECAL-based  
1446 seeding, which begins with the SC energy and positioning, and the tracker-based  
1447 seeding (part of the PF reconstruction algorithm), which uses tracks reconstructed  
1448 from the general algorithm for charged particles, extrapolated towards the ECAL and  
1449 matched to an SC. Kalman filter (KF) tracks with a small number of hits or that are  
1450 not well-fitted, are re-fitted with a dedicated Gaussian sum Filter (GSF).

1451 A global identification variable [57] is defined using a multivariate analysis (MVA)  
1452 technique that combines information on track observables (kinematics, quality of the  
1453 KF track and GSF track), the electron PF cluster observables (shape and pattern),  
1454 and the association between the two (geometric and kinematic observables). For  
1455 electrons seeded only through the tracker-based approach, a weak selection is applied  
1456 on this MVA variable. For electrons seeded through both approaches, a logical OR is  
1457 taken.

1458 Electron isolation, i.e. the presence of energy deposits near the electron trajectory,

1459 is a separate key handle in rejecting significant background. Compared to isolated  
1460 electrons, electrons from misidentified jets or genuine electrons within a jet resulting  
1461 from semileptonic decays of  $b$  or  $c$  quarks tend to have significant energy deposits  
1462 near the primary trajectory [57]. Offline analyses benefit from the PF technique  
1463 for defining isolation, which sums the PF candidates reconstructed located within a  
1464 specified isolation cone around the electron candidate, as in Eqn. 5.2.

#### 1465 5.1.4 Jets

1466 The vast majority of processes of interest at the LHC contains quarks or gluons in  
1467 the final state, but these particles cannot be observed directly. In a process called  
1468 hadronization, they fragment into spatially-grouped collections of particles called jets,  
1469 which can be detected in the tracking and calorimeter systems. Hadronization and  
1470 the subsequent decays of unstable hadrons can produce hundreds of nearby particles  
1471 in the CMS detector. Jets are reconstructed by the PF algorithm (PF jets), or from  
1472 the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo  
1473 jets). In PF jets, typically used in offline analyses, jets are built using the anti- $k_T$   
1474 (AK) clustering algorithm [58]. The anti- $k_T$  algorithm iterates over particle pairs and  
1475 finds the two that are closest in a distance measure  $d$ , and determines whether to  
1476 combine them:

$$d_{ij} = \min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta_{ij}^2}{R^2}, \text{ combine when } d_{ij} < p_{T,i}^{-2}; \text{ stop when } d_{ij} > p_{T,i}^{-2} \quad (5.3)$$

1477 where  $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$  and  $p_{T,i}$ ,  $\eta_i$ ,  $\phi_i$  are the transverse momentum, rapid-  
1478 ity, and azimuthal angle of particle  $i$ . The power  $-2$  means that higher-momentum  
1479 particles are clustered first, leading to jets that tend to be centered on the hardest  
1480 (highest  $p_T$ ) particle.

1481 There are several methods to remove contributions of pile-up collisions from jet

1482 clustering [59]:

- 1483 • Charged hadron subtraction (CHS), which removes all charged hadron candidates associated with a track that is not associated with the primary vertex.
- 1485 • PileUp Per Particle Identification (PUPPI), which weighs input particles based  
1486 on their likelihood of arising from pile-up. QCD particles tend to have a collinear  
1487 structure, compared to soft diffuse radiation coming from pile-up. The local  
1488 shape for charged pile-up, used as a proxy for all pile-up particles, is used on an  
1489 event-by-event basis to calculate a weight for each particle. PUPPI is deployed  
1490 in Run-2 and is more performant than CHS in high pile-up scenarios.

1491 **5.1.5 B-flavored jets**

1492 Jets that arise from bottom-quark hadronization (b-flavor jets) have overwhelming  
1493 background from processes involving jets from gluons (g) and light-flavor quarks (u, d,  
1494 s), and from c-quark fragmentation. The ability to identify b-flavor jets, or b-tagging,  
1495 exploits the b hadrons' relatively large masses, long lifetimes, and daughter particles  
1496 with hard momentum spectra [58].

1497 The impact parameter (IP) of a track is the 3-dimensional distance between the  
1498 track and the primary vertex (PV) at the point of closest approach. The IP is positive  
1499 if the track originates from the decay of particles travelling along the jet axis. The  
1500 resolution of the IP depends on the  $p_T$  and  $\eta$  of the track, motivating the use of the  
1501 impact parameter significance  $S_{\text{IP}}$  (ratio of the IP to its estimated uncertainty) as an  
1502 observable [58].

1503 Because of their large, but finite, lifetimes, b hadrons tend to travel a short dis-  
1504 tance before decaying at a secondary vertex (SV), which can be measured and recon-  
1505 structed separately from the primary vertex due to the excellent position resolution  
1506 of the pixel detector [58]. Previous b-tagging algorithms (e.g. CSV, cMVAv2, and

1507 DeepCSV) have capitalized on variables such as the presence of a SV, the flight  
1508 distance and direction (computed from the vector between the PV and the SV), and  
1509 kinematics of the system of associated secondary tracks (e.g. track multiplicity, mass,  
1510 and energy).

1511 The DeepJet (formerly known as DeepFlavour) algorithm [60] is a deep-neural-  
1512 network multi-classification algorithm, which uses 16 properties of up to 25 charged  
1513 and 6 properties of 25 neutral particle-flow jet constituents, as well as 17 properties  
1514 from up to 4 secondary vertices associate with the jet. Compared to the previous clas-  
1515 sifying algorithm DeepCSV, DeepJet has been demonstrated to have higher efficiency  
1516 with lower misidentification probability in Phase-1 data [61].

## 1517 5.2 Reconstruction of the di-tau mass

1518 The final signal extraction is done to the total di-tau ( $\tau\tau$ ) mass, which is estimated  
1519 from the visible  $\tau\tau$  mass using the FastMTT algorithm [62]. FastMTT is based on the  
1520 SVFit algorithm, originally developed for the Standard Model  $H \rightarrow \tau\tau$  analysis [63].  
1521 Both the SVFit algorithms, and the FastMTT algorithm, are described below, to give  
1522 a complete picture of how the algorithms attempt to reconstruct the true invariant  
1523 mass of a Higgs or Z boson decay.

1524 To specify a hadronic  $\tau$  decay, six parameters are needed [63]: the polar and  
1525 azimuthal angles of the visible decay product system in the  $\tau$  rest frame, the three  
1526 boost parameters from the  $\tau$  rest frame to the laboratory frame, and the invariant  
1527 mass  $m_{\text{vis}}$  of the visible decay products. For a leptonic  $\tau$  decay, two neutrinos are  
1528 produced, and a seventh parameter, the invariant mass of the two-neutrino system, is  
1529 necessary. The unknown parameters are constrained by four observables that are the  
1530 components of the four-momentum of the system formed by the visible decay products  
1531 of the  $\tau$  lepton, measured in the laboratory frame. The remaining unconstrained

1532 parameters for hadronic and leptonic  $\tau$  decays are thus:

1533 • The fraction of the  $\tau$  energy in the laboratory frame carried by the visible decay  
1534 products,

1535 •  $\phi$ , the azimuthal angle of the  $\tau$  direction in the laboratory frame,

1536 •  $m_{\nu\nu}$ , the invariant mass of the two-neutrino system in leptonic  $\tau$  decays (for  
1537 hadronic  $\tau$  decays,  $m_{\nu\nu}$  is set to 0).

1538  $E_x^{\text{miss}}$  and  $E_y^{\text{miss}}$ , the  $x$  and  $y$  components of the missing transverse energy  $E_T^{\text{miss}}$   
1539 provide two further constraints.

### 1540 5.2.1 Original SVFit ‘‘standalone’’: maximum likelihood

1541 In one of the original versions of SVFit, called ‘‘standalone’’ SVFit [63], a maximum  
1542 likelihood fit method is used to reconstruct the mass  $m_{\tau\tau}$  by combining the measured  
1543 observables  $E_x^{\text{miss}}$  and  $E_y^{\text{miss}}$  with a likelihood model that includes terms for the  $\tau$   
1544 decay kinematics and the  $E_T^{\text{miss}}$  resolution [63]. The likelihood function  $f(\vec{z}, \vec{y}, \vec{a}_1 \vec{a}_2)$   
1545 of the parameters  $\vec{z} = (E_x^{\text{miss}}, E_y^{\text{miss}})$  in an event is constructed, where the remaining  
1546 parameters are the kinematics of the two  $\tau$  decays, denoted  $\vec{a}_1 = (x_1, \phi_1, m_{\nu\nu,1})$  and  
1547  $\vec{a}_2 = (x_2, \phi_2, m_{\nu\nu,2})$ , and the four-momenta of the visible decay products with the  
1548 measured values  $\vec{y} = (p_1^{\text{vis}}, p_2^{\text{vis}})$ .

1549 The likelihood  $f$  is the product of three likelihood functions. The first two likeli-  
1550 hood functions model the decay parameters  $\vec{a}_1$  and  $\vec{a}_2$  of the two  $\tau$  leptons. For lep-  
1551 tonic decays, the likelihood function is modeled using matrix elements for  $\tau$  decays,  
1552 and integrated over the allowed phase space  $0 \leq x \leq 1$  and  $0 \leq m_{\nu\nu} \leq m_\tau \sqrt{1-x}$ . For  
1553 hadronic  $\tau$  decays, a model based on the two-body phase space is used and integrated  
1554 over  $m_{\text{vis}}^2/m_{\tau\tau}^2 \leq x \leq 1$ . The third likelihood function quantifies the compatibility of  
1555 a  $\tau$  decay hypothesis with the reconstructed  $\vec{E}_T^{\text{miss}}$  in an event, assuming the neutrini-  
1556 nos are the only source of missing transverse energy. The expected  $\vec{E}_T^{\text{miss}}$  resolution

1557 is represented by a covariant matrix, estimated on an event-by-event basis using a  
1558 significance algorithm [64].

### 1559 5.2.2 “Classic SVFit” with matrix element

1560 Classic SVFit is an improved algorithm of the original “standalone” SVFit using the  
1561 formalism of the matrix element (ME) method [62]. In the ME method, an estimate  
1562 for the unknown model parameter  $\Theta$  (here, the mass  $m_{\tau\tau}$ ) is obtained by maximizing  
1563 the probability density  $\mathcal{P}$ . The key ingredients of the probability density are the  
1564 squared modulus of the matrix element  $|\mathcal{M}(\mathbf{p}, \Theta)|^2$  and the transfer function  $W(\mathbf{y}|\mathbf{p})$   
1565 (probability density to observe the measured observables  $\mathbf{y}$  given the phase space  
1566 point  $\mathbf{p}$ ). The best estimate  $m_{\tau\tau}$  is obtained by computing the probability density  $\mathcal{P}$   
1567 for a range of mass hypotheses and finding the value of  $m_{\tau\tau}$  that maximizes  $\mathcal{P}$ .

1568 Distributions illustrating the performance of the classic matrix element SVFit  
1569 algorithm are shown in Fig. 5.1 from [62], showing the di-tau mass after and before  
1570 application of SVFit to recover energy lost to neutrinos. The SVFit algorithm is  
1571 found to improve the sensitivity of the Standard Model  $H \rightarrow \tau\tau$  analysis performed  
1572 by CMS by about 30%, compared to performing the same analysis using only the  
1573 visible mass  $m_{\text{vis}}$ .

### 1574 5.2.3 FastMTT: optimized SVFit

1575 FastMTT [65] is a further simplification to the matrix element method of Classic  
1576 SVFit which has comparable performance but is about 100 times faster. FastMTT  
1577 drops the matrix element component of the computation without significant impact  
1578 on the final mass resolution, and simplifies the computation of the transfer functions.  
1579 The opening angle of the  $\tau$  decay products with respect to the initial  $\tau$  momenta ap-  
1580 proaches 0 for  $\tau$  with high  $\gamma = E_\tau/m_\tau$ , with typical  $\tau$  decays from the Z boson decays  
1581 already satisfying this condition. In this collinear approximation, the dimensionality

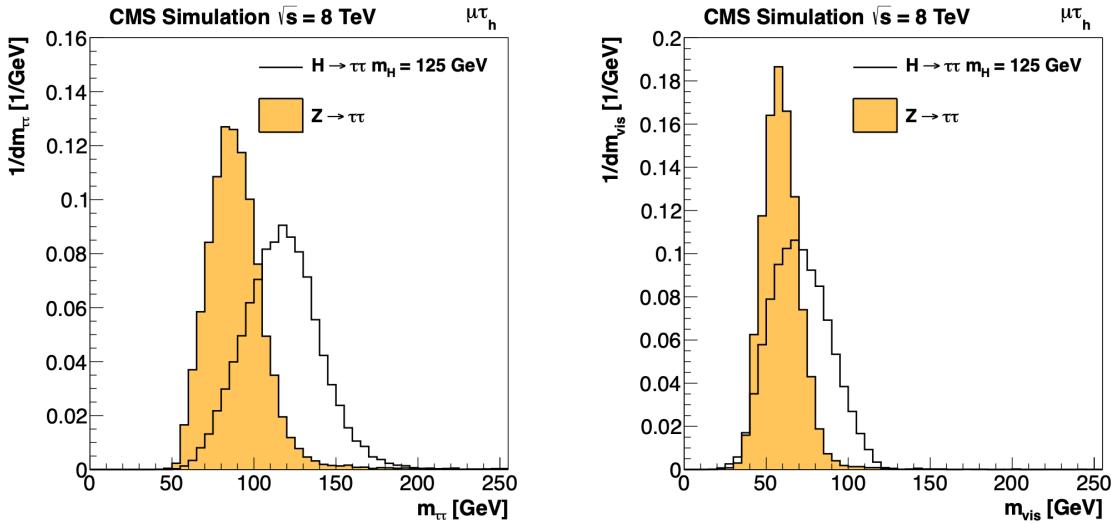


Figure 5.1: Distributions from [62], of  $m_{\tau\tau}$  after reconstruction with the original SVFit algorithm (*left*), and before SVFit with only the visible tau decay products (*right*), for  $H \rightarrow \tau\tau$  signal events of mass  $m_H = 125$  GeV (*black line*) and the  $Z/\gamma^* \rightarrow \tau\tau$  background (*orange, solid*), in the decay channel  $\tau\tau \rightarrow \mu\tau_h$ .

1582 of the transfer function can be reduced in the computation of FastMTT, while still  
1583 yielding similar results to Classic SVFit [65].

### 1584 5.3 Corrections applied to simulation

1585 Corrections are applied to simulated samples to account for known effects in the event  
1586 modeling and reconstruction and data-taking, and are intended to bring simulations  
1587 in closer agreement with data. Corrections fall into two broad categories: *energy*  
1588 *scale corrections* applied to physics objects, and *event-level corrections*. Energy scale  
1589 corrections are multiplicative factors applied to the energy and transverse momentum  
1590  $p_T$  of simulated objects (e.g. leptons or jets), and bring the average reconstructed  
1591 energies of simulated particles into better agreement with those of objects recon-  
1592 structed from data. Event-level corrections are applied as a per-event multiplicative  
1593 weight, and account for effects such as differences in object identification efficiencies  
1594 and trigger efficiencies between data and simulated samples, mis-modeling in simu-

1595 lations of the underlying physics process, or changing detector operating conditions  
 1596 during data-taking. Event-level corrections change the shapes of the distributions of  
 1597 all the physical observables.

1598 Uncertainties in scale factors and corrections are also sources of systematic errors  
 1599 in the analysis, detailed in Chapter 8. Systematic uncertainties in the tau, muon, and  
 1600 electron energy scales can shift the  $p_T$  of the leptons up or down, which can change  
 1601 whether events pass or fail the offline  $p_T$  thresholds for the trigger paths described in  
 1602 the previous section, i.e. change the number of events in the signal region.

### 1603 **5.3.1 Tau energy scale**

1604 An energy scale is applied to the transverse momentum  $p_T$  and mass of the hadronic  
 1605 tau  $\tau_h$  in the  $\mu\tau_h$  and  $e\tau_h$  channels, to correct for a deviation of the average recon-  
 1606 structed  $\tau_h$  energy from the generator-level energy of the visible  $\tau_h$  decay products.  
 1607 These correction factors are derived centrally [54], by fitting to events in  $e\tau_h$  and  $\mu\tau_h$   
 1608 final states in  $Z/\gamma^*$  events separately for the  $h^\pm$ ,  $h^\pm\pi^0$ , and  $h^\pm h^\mp h^\pm$  decays. The  
 1609 values used are shown in Table 5.1.

1610 When applying the energy scale to the  $\tau_h$ , the 4-momentum of the missing trans-  
 1611 verse energy (MET) is adjusted such that the total 4-momenta of the  $\tau_h$  and the MET  
 1612 remains unchanged [66].

Tau energy scale factor				
Decay mode	2018	2017	2016 pre-VFP	2016 post-VFP
1-prong	$0.991 \pm 0.008$	$0.986 \pm 0.009$	$0.987 \pm 0.01$	$0.993 \pm 0.009$
1-prong + $\pi^0$	$1.004 \pm 0.006$	$0.999 \pm 0.006$	$0.998 \pm 0.006$	$0.991 \pm 0.007$
3-prong	$0.998 \pm 0.007$	$0.999 \pm 0.007$	$0.984 \pm 0.008$	$1.001 \pm 0.007$
3-prong + $\pi^0$	$1.004 \pm 0.009$	$0.996 \pm 0.01$	$0.999 \pm 0.011$	$0.997 \pm 0.016$

Table 5.1: Energy scales applied to genuine hadronic tau decays  $\tau_h$  by data-taking year/era and decay mode, along with systematic errors.

### **5.3.2 Muon energy scale**

An energy scale is applied to the  $p_T$  and mass of genuine muons from  $\tau$  decays in the  $e\mu$  and  $\mu\tau_h$  channels [67]. The applied values are the same for MC and embedded samples and are shown in Table 5.2. Following the SM  $H \rightarrow \tau\tau$  analysis, Rochester corrections are not applied, and instead prescriptions from [68] are followed.

Muon energy scale factor	
Eta range	Value for all years
$ \eta  \in [0.0, 1.2)$	$1.0 \pm 0.004$
$ \eta  \in [1.2, 2.1)$	$1.0 \pm 0.009$
$ \eta  \in [2.1, 2.4)$	$1.0 \pm 0.027$

Table 5.2: Energy scales and systematic errors applied to genuine muons. The values are the same for MC and embedded for all years [69] [68].

### **5.3.3 Electron energy scale**

Corrections to the electron energy scale are applied to genuine  $e$  from  $\tau$  decays, and are binned in two dimensions by electron  $p_T$  and  $\eta$  for barrel vs. endcap [70]. The scale factors are binned in  $p_T$  and  $\eta$  for MC samples: e.g. values for 2018 are shown in Fig. 5.2 from [71]. For embedded samples the electron energy scale is taken as only binned in  $\eta$  (Table 5.3).

Electron energy scale factor for embedded samples			
Eta range	2018	2017	2016
$ \eta  \in [0.0, 1.479)$	$0.973 \pm 0.005$	$0.986 \pm 0.009$	$0.9976 \pm 0.0050$
$ \eta  \in [1.479, 2.4)$	$0.980 \pm 0.0125$	$0.887 \pm 0.0125$	$0.993 \pm 0.0125$

Table 5.3: Energy scales and systematic errors applied to electrons in embedded samples, binned in the electron  $\eta$ , by data-taking year [72] [73] [74].

### **5.3.4 $\tau_h$ identification efficiency**

The  $\tau_h$  identification efficiency can differ in data and MC [66]. Recommended corrections are provided by the Tau POG, and we use the medium DeepTau vs. jet working

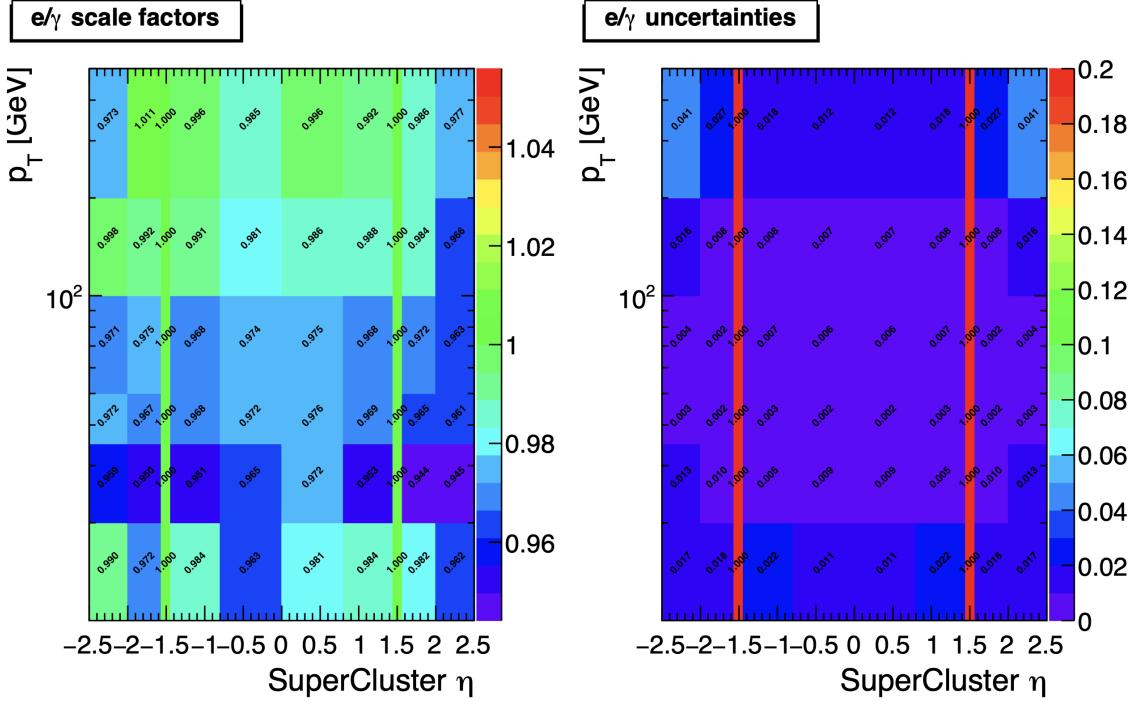


Figure 5.2: Electron/photon energy scale factors (*left*) and corresponding uncertainties (*right*) binned in the electron  $\eta$  and  $p_T$ , for the data-taking year 2018 [71].

1627 point values. The identification efficiency is measured in  $Z \rightarrow \tau\tau$  events in the  $\mu\tau_h$   
1628 final state, and is binned in  $p_T$  due to clear  $p_T$  dependence of the DeepTau ID.

Tau ID efficiency for DeepTau Medium vs. jet WP in 2018						
$p_T$ (GeV)	< 20	(20, 25]	(25, 30]	(30, 35]	(35, 40]	(40, 500]
Central value	0	0.945	0.946	0.916	0.921	1.005
Up value	0	1.001	0.981	0.946	0.950	1.035
Down value	0	0.888	0.981	0.883	0.893	0.953

Table 5.4: Tau ID efficiency for the DeepTau vs. jet medium working point, with central, up, and down values for 2018, binned in the tau  $p_T$  [66].

### 1629 5.3.5 Trigger efficiencies definition

1630 Scale factors are applied to correct for differences in trigger efficiencies between MC  
1631 and embedded vs. data, with values taken from tools provided by the Standard Model  
1632  $H \rightarrow \tau\tau$  working group which uses the same trigger paths [69]. In the following

1633 sections we review relevant trigger efficiencies in data, which form the basis of the  
1634 trigger efficiency corrections applied to MC and embedded.

1635 **5.3.6 Tau trigger efficiencies**

1636 The efficiencies in data of the single- $\tau_h$  leg in  $\mu\tau_h$ ,  $e\tau_h$ , and di- $\tau_h$  triggers is computed  
1637 centrally per using a Tag and Probe (TnP) method [75] which is outlined here. In  
1638 this method,  $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$  are selected in data and a Drell-Yan simulated sample  
1639 ( $Z \rightarrow \ell\ell, \ell = e, \mu, \tau_h$ ) with high purity. Cuts are applied to reject events not in this  
1640 final state, e.g. suppressing  $Z \rightarrow \mu\mu$  by vetoing events with a single loose ID muon.  
1641 An isolated muon candidate (the tag) with online  $p_T > 27 \text{ GeV}$  and  $|\eta| < 2.1$  is  
1642 identified and matched to an offline  $\mu$ . An offline  $\tau_h$  candidate (the probe) is selected,  
1643 which is separated from the tag  $\mu$ , and has  $p_T > 20 \text{ GeV}$  and  $|\eta| < 2.1$ . The probe  
1644  $\tau_h$  must pass anti-muon and anti-electron discriminators to avoid fakes from muons  
1645 and electrons, and must pass the medium MVA tau isolation to suppress fakes from  
1646 QCD jets. The trigger efficiency in the TnP method is calculated as

$$\text{Efficiency} = \frac{\text{Number of events passing the TnP selection with fires the HLT path}}{\text{Number of events passing the TnP selection}} \quad (5.4)$$

1647 The efficiencies for the hadronic tau legs in the relevant channels of this analyses  
1648 ( $\mu\tau_h$  and  $e\tau_h$ ) as a function of the offline tau  $p_T$  and  $\eta$ , are shown for data taken in  
1649 2016, 2017, and 2018 in Figures 5.3a and 5.3b [75] [76]. In both figures, the different  
1650 HLT thresholds and differences in the L1 seed result in higher efficiencies in 2016 and  
1651 differences in shapes of the 2016 efficiencies compared to 2017 and 2018. The low  
1652 pile-up in 2016 also leads to higher efficiencies in that year.

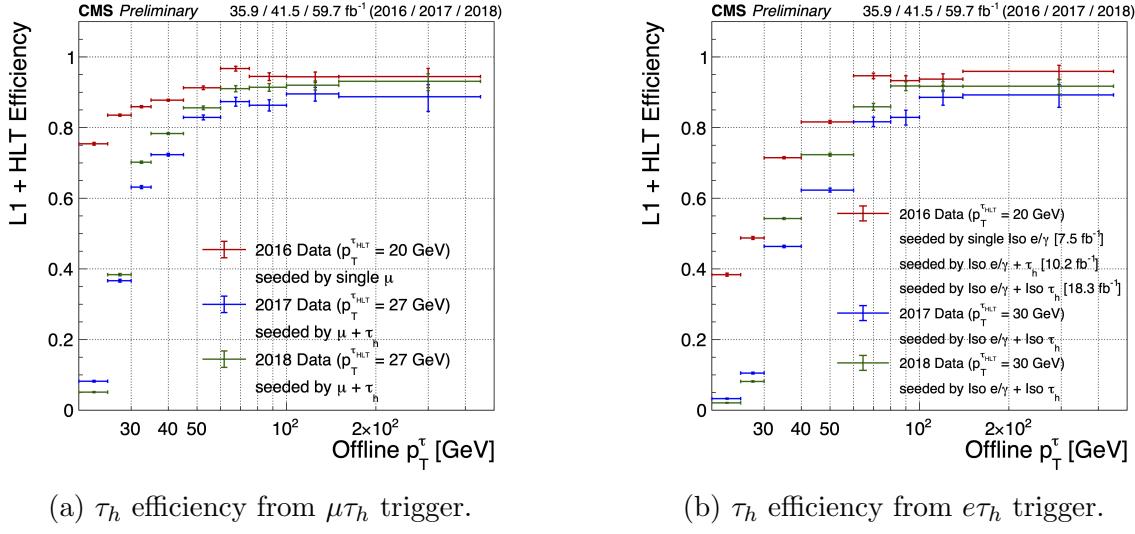
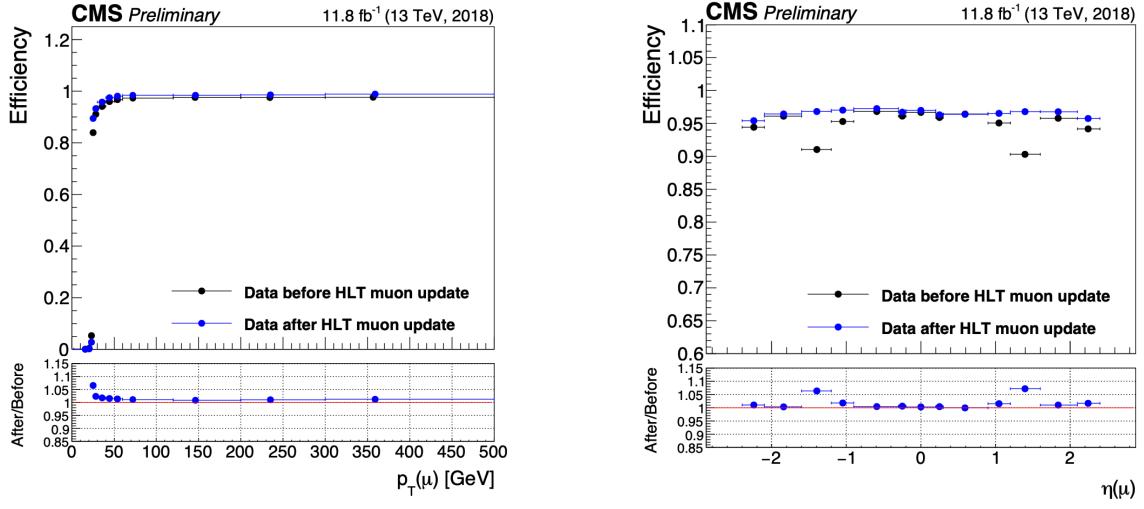


Figure 5.3: Hadronic tau leg efficiency of the cross-triggers for  $\mu\tau_h$  (*left*) and  $e\tau_h$  (*right*) triggers as a function of offline tau  $p_T$  for the years 2016 (red), 2017 (blue) and 2018 (green), from [76]. HLT  $p_T$  thresholds and L1 seeds are indicated in the legends.

### 1653 5.3.7 Single muon trigger efficiencies

1654 The efficiencies for the single isolated muon trigger with  $p_T > 24$  GeV used in this  
 1655 analysis, is shown for the data-taking year 2018 in Fig. 5.4a as a function of the muon  
 1656  $p_T$  and as a function of the muon  $|\eta|$  in Fig. 5.4b from [77]. The data is split with  
 1657 respect to a HLT muon reconstruction update that was deployed on 15/05/2018. A  
 1658 small asymmetry in efficiencies between negative and positive  $\eta$  in Fig. 5.4b is due to  
 1659 disabled muon chambers (CSCs). The efficiencies shown are estimated using a Tag  
 1660 and Probe method using  $Z \rightarrow \mu\mu$  events, with the tag being an offline muon with  
 1661  $p_T > 29$  GeV and  $|\eta| < 2.4$  passing a tight ID criteria, and the probe is an online (L1)  
 1662 trigger object with  $\Delta R < 0.3$  and passing tight ID and Particle Flow based isolation  
 1663 requirements with  $p_T > 26$  GeV.



(a) Muon efficiency vs  $p_T$  for SingleMuon.

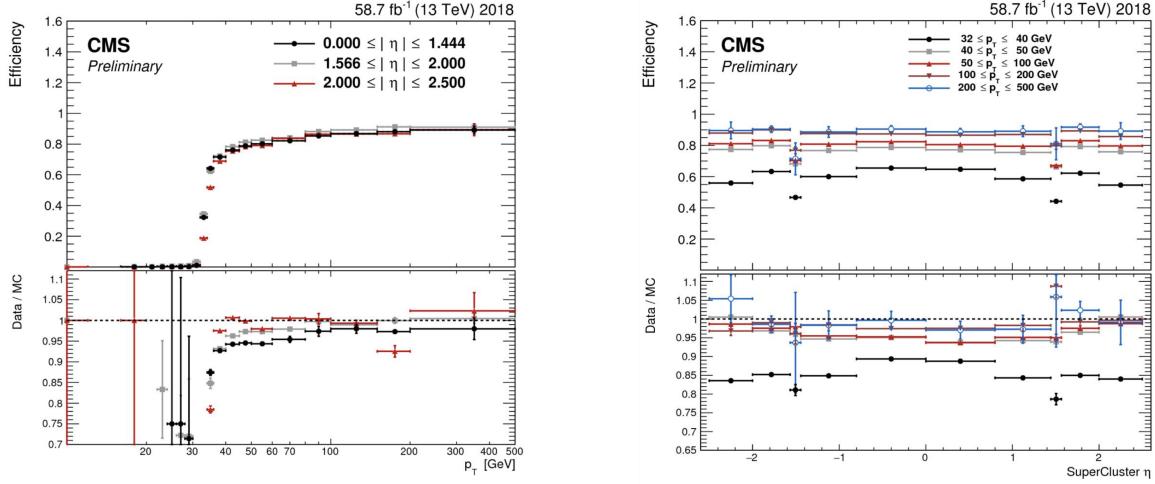
(b) Muon efficiency vs  $|\eta|$  for SingleMuon.

Figure 5.4: Trigger efficiencies in data (*top panels*) and ratio of efficiencies after/before a HLT muon reconstruction update (*bottom panels*) for the muon in the isolated single muon trigger with threshold  $p_T > 24$  GeV in the data-taking year 2018, as functions of the muon  $p_T$  (*left*) and muon  $|\eta|$  (*right*). Only statistical errors are shown [77].

### 1664 5.3.8 Single electron trigger efficiencies

1665 The efficiencies in data, and the ratio between data and MC, of the single electron  
 1666 HLT trigger with  $p_T$  threshold 32 GeV used in this analysis are shown for 2018, as  
 1667 a function of the electron  $p_T$  in Fig. 5.5a and of the electron  $|\eta|$  in Fig. 5.5b,  
 1668 from [78]. In the Tag and Probe method used for the 2018 dataset, the tag is an  
 1669 offline reconstructed electron with  $|\eta| \leq 2.1$  and not in the barrel and endcap overlap  
 1670 region, with  $p_T > 35$  GeV with tight isolation and shower shape requirements, firing  
 1671 the tag trigger. The probe is an offline reconstructed electron with  $|\eta| \leq 2.5$  with  
 1672  $E_T^{\text{ECAL}} > 5$  GeV with no extra identification criteria [78].

1673 The disagreement between data and MC, particularly at low transverse momentum,  
 1674 is in part due to detector effects that are difficult to simulate, such as crys-  
 1675 tal transparency losses in the ECAL and the evolution of dead regions in the pixel  
 1676 tracker [78].



(a) Electron efficiency vs  $p_T$  for single electron.

(b) Electron efficiency vs  $|\eta|$  for single electron.

Figure 5.5: Trigger efficiencies in data, and the data/MC ratio for the electron in the single electron trigger with threshold  $p_T > 32$  GeV in the data-taking year 2018, as functions of the electron  $p_T$  (*left*) and electron  $|\eta|$  (*right*) [78]. In the plot vs.  $p_T$ , the region  $1.442 \leq |\eta| \leq 1.566$  is not included as it corresponds to the transition between barrel and endcap parts of the ECAL.

### 1677 5.3.9 $e\mu$ cross-trigger efficiencies

1678 The efficiencies of the electron and muons for the cross-trigger with leading muon  
 1679 used in the  $e\mu$  channel are shown for data in 2016, 2017, and 2018 in Figures 5.6a and  
 1680 5.6b [79]. These efficiencies were measured centrally using a Tag and Probe in events  
 1681 with  $Z$  to dileptons with the same flavor and opposite charge, where the tags are an  
 1682 isolated muon or electron, and the probe (offline) candidate is required to satisfy the  
 1683 same lepton selection as that of the tag candidate, be matched within  $\Delta R < 0.1$  with  
 1684 a corresponding online trigger object, and also to pass the cross-trigger. The trigger  
 1685 efficiency is then:

$$\text{Efficiency} = \frac{\text{Events passing lepton pair selections and probe passing trigger}}{\text{Events passing lepton pair selections}} \quad (5.5)$$

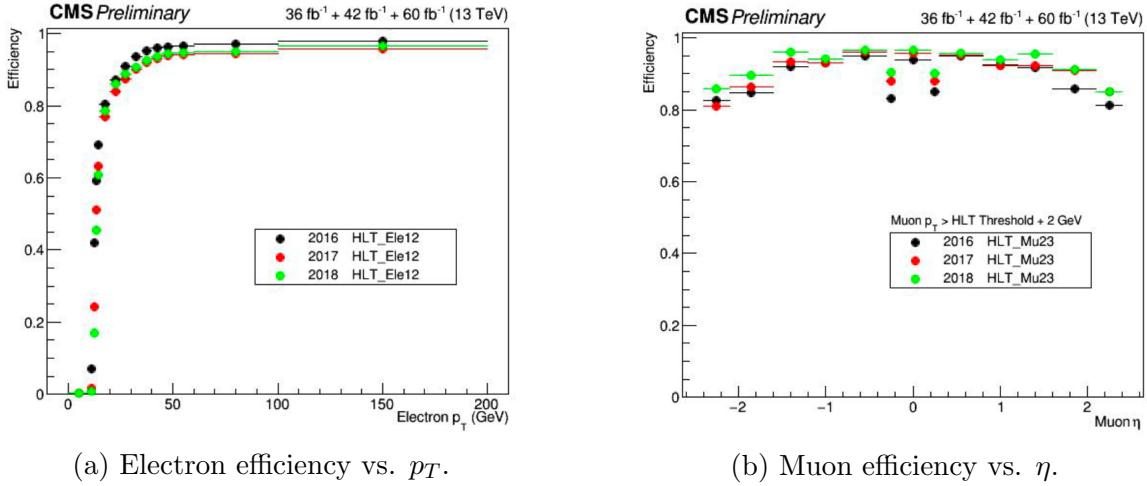


Figure 5.6: Efficiencies of the electron leg vs.  $p_T$  (*left*) and the muon log vs.  $\eta$  (*right*), for the HLT path with online thresholds of 12 GeV for the electron and 23 GeV for the muon, for the data-taking years 2016 (*black*), 2017 (*red*), and 2018 (*green*) [79].

### 1686 5.3.10 Electrons and muons faking $\tau_h$ : energy scales

1687 Energy scales for electrons misidentified as hadronic tau decays ( $e$  faking  $\tau_h$ ) are  
 1688 provided by the Tau POG, and were measured in the  $e\tau_h$  channel with the visible  
 1689 invariant mass of the electron and hadronic tau system [69]. This energy scale is  
 1690 applied for  $\tau_h$  with  $p_T > 20$  GeV regardless of which DeepTau vs. electron working  
 1691 point was used. Values for 2018 are shown in Table 5.5.

Electrons faking $\tau_h$ energy scale factor in 2018	
Reconstructed decay mode of the fake $\tau_h$	Central value and (up, down) shifts
1-prong	1.01362 (+0.00474, -0.00904)
1-prong + $\pi^0$	1.01945 (+0.01598, -0.01226)
3-prong	0.96903 (+0.0125, -0.03404)
3-prong + $\pi^0$	0.985 (+0.04309, -0.05499)

Table 5.5: Energy scales and up/down systematic uncertainties applied to electrons misidentified as hadronic taus for 2018, binned in decay mode of the fake  $\tau_h$  [69].

1692 No nominal energy scale is applied for muons mis-reconstructed as  $\tau_h$ , and the  
 1693 uncertainty is treated as  $\pm 1\%$  and uncorrelated in the reconstructed decay mode [69].

1694    **5.3.11 Electrons and muons faking  $\tau_h$ : misidentification effi-**  
 1695    **ciencies**

1696    Corrections on identification efficiencies are applied to genuine electrons and muons  
 1697    misidentified as  $\tau$  to account for differences in data and MC.

1698    The specific values depend on the vs. electron and vs. muon discriminator working  
 1699    points used. For misidentified  $\mu \rightarrow \tau_h$ , the scale factors are split into different  $|\eta|$   
 1700    regions, determined by the CMS muon and tracker detector geometries, as shown in  
 1701    Table 5.6 for 2018 [66].

Tau ID efficiency for DeepTau vs. muon WPs in 2018		
$ \eta $	Tight working point	VLoose working point
(0.0, 0.2)	$(0.767 \pm 0.127)\%$	$(0.954 \pm 0.069)\%$
(0.2, 0.6)	$(1.255 \pm 0.258)\%$	$(1.009 \pm 0.098)\%$
(0.6, 1.0)	$(0.902 \pm 0.203)\%$	$(1.029 \pm 0.075)\%$
(1.0, 1.45)	$(0.833 \pm 0.415)\%$	$(0.928 \pm 0.145)\%$
(1.45, 2.0)	$(4.436 \pm 0.814)\%$	$(5.000 \pm 0.377)\%$
(2.0, 2.53)	$(1.000 \pm 0.000)\%$	$(1.000 \pm 0.000)\%$

Table 5.6: Tau mis-identification efficiencies (in %) for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018, binned in the muon  $|\eta|$  [66].

1702    For misidentified  $e \rightarrow \tau_h$ , the scale factors are split into barrel and endcap regions,  
 1703    dictated by the ECAL detector geometry, as shown in Table 5.7 for 2018.

Tau ID efficiencies for DeepTau vs. electron WPs in 2018		
$ \eta $	Tight working point	VLoose working point
(0.0, 0.73)	$(1.47 \pm 0.27)\%$	$(0.95 \pm 0.07)\%$
(0.73, 1.509)	$(1.509 \pm 0.0)\%$	$(1.00 \pm 0.0)\%$
(1.509, 1.929)	$(1.929 \pm 0.2)\%$	$(0.86 \pm 0.1)\%$
(1.929, 2.683)	$(2.683 \pm 0.9)\%$	$(2.68 \pm 0.0)\%$

Table 5.7: Tau mis-identification efficiencies (in %) for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018, binned in the electron  $|\eta|$  [66].

### 1704 5.3.12 Electron ID and tracking efficiency

1705 Scale factors are applied to MC to correct for differences between MC and data in  
 1706 the performance of electron identification (ID) and tracking.

1707 Electron and photon identification, as discussed earlier, use variables with good  
 1708 signal vs. background discrimination power such as lateral shower shape and ratio  
 1709 of energy deposited in the HCAL to energy deposited in the ECAL at the position  
 1710 of the electron. The cut-based electron identification efficiencies in data and ratio of  
 1711 efficiencies in data to MC are shown in Fig. 5.7a for the multivariate analysis (MVA)  
 1712 identification working point.

1713 The tracking efficiencies in data and the data/MC ratio are shown in Fig. 5.7b  
 1714 for the Gaussian-sum filter (GSF) tracking [80].

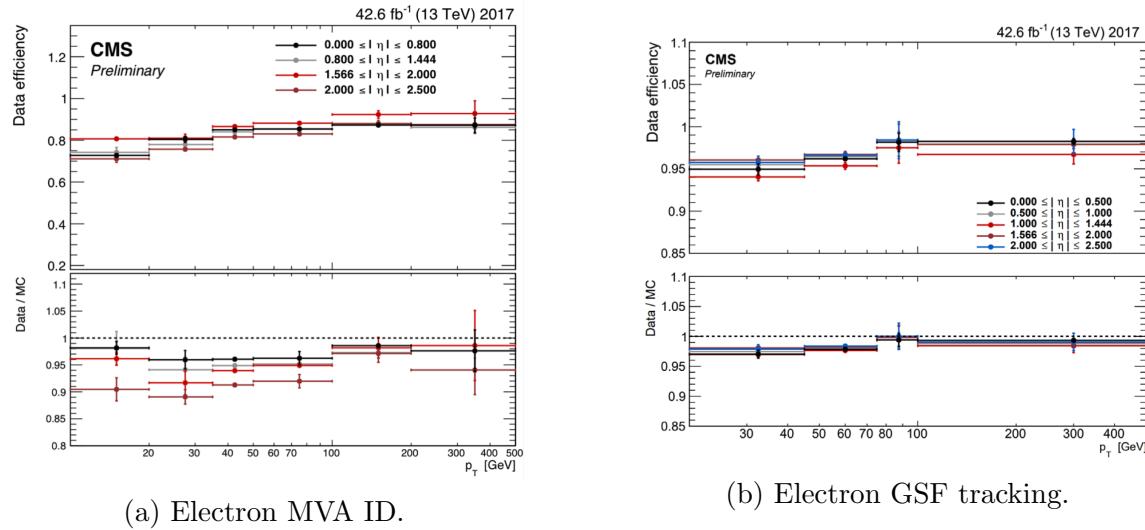


Figure 5.7: Efficiencies in data (*top panels*) and the ratio of efficiencies in data/MC (*bottom panels*), for the electron multivariate analysis (MVA) identification (*left*) and for the Gaussian-sum filter (GSF) tracking (*right*) [80]. Error bars represent statistical and systematic uncertainties.

### 1715 5.3.13 Muon ID, isolation, and tracking efficiencies

1716 Scale factors are applied to MC to correct for differences between MC and data in  
 1717 the performance of muon identification, isolation, and tracking, as detailed below.

1718        The efficiencies for muon identification measured in 2015 data and MC simulation  
 1719        are shown in Figures 5.8a and 5.8b for the loose ID and tight ID respectively [81]. The  
 1720        loose ID is chosen such that efficiency exceeds 99% over the full  $\eta$  range, and the data  
 1721        and simulation agree to within 1%. The tight ID is chosen such that efficiency varies  
 1722        between 95% and 99% as a function of  $\eta$ , and the data and simulation agree to within  
 1723        1-3%. The muon identification working point used in this analysis is the medium ID,  
 1724        which has an efficiency of 98% for all  $\eta$  and an agreement within 1-2% [81].

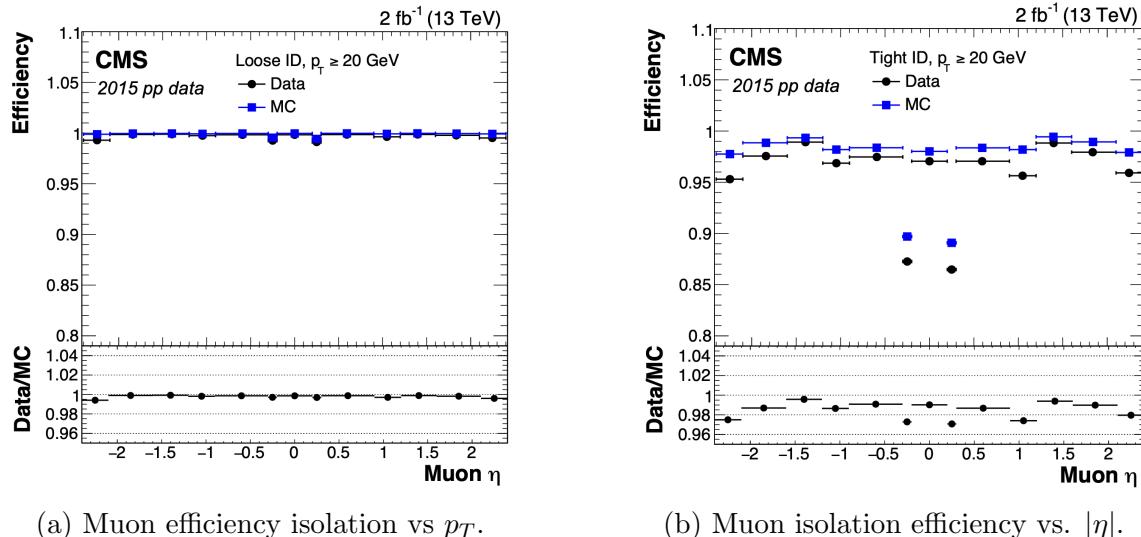
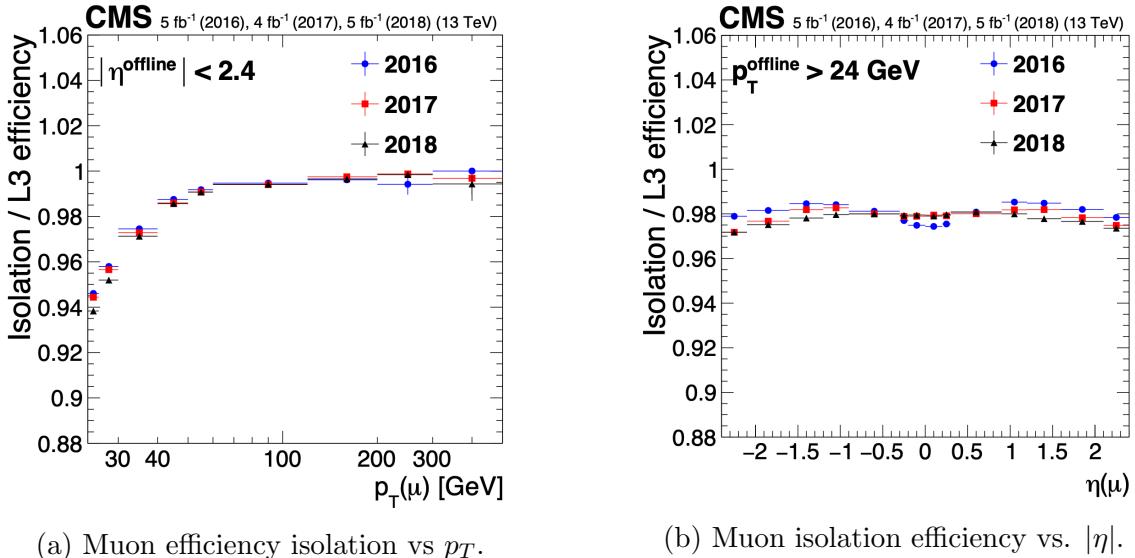


Figure 5.8: Muon identification efficiencies in 2015 data and MC as a function of the muon  $p_T$  for the loose ID (*left*) and tight ID (*right*) working points [81].

1725        The efficiencies in data for the muon isolation, as measured in Level-3 muons  
 1726        (muons in one of the final stages of reconstruction in the HLT), as a function of the  
 1727        muon  $p_T$  and  $|\eta|$  are shown in Figures 5.9a and 5.9b [81]. The HLT muon reconstruc-  
 1728        tion consists of two steps: Level-2 (L2), where the muon is reconstructed in the muon  
 1729        subdetectors only, and Level-3 (L3) which is a global fit of tracker and muon hits (i.e.  
 1730        the global muon reconstruction as described in Section 5.1.2) [82].

1731        The muon tracking efficiencies as a function of  $|\eta|$  for standalone muons (i.e. tracks  
 1732        from only the muon system, i.e. DT, CSC, and RPC, as discussed in Section 5.1.2),  
 1733        is shown for data and simulated Drell-Yan samples in Fig. 5.10 [83].



(a) Muon efficiency isolation vs  $p_T$ .

(b) Muon isolation efficiency vs.  $|\eta|$ .

Figure 5.9: Muon isolation efficiencies in Run-2 data with respect to Level-3 muons (one of the final stages of HLT muon reconstruction) as a function of the muon  $p_T$  (*left*) and  $|\eta|$  (*right*) [81].

### 1734 5.3.14 Recoil corrections

1735 In proton-proton collisions, W and Z bosons are predominantly produced through  
1736 quark-antiquark annihilation. Higher-order processes can induce radiated quarks or  
1737 gluons that recoil against the boson, imparting a non-zero transverse momentum to  
1738 the boson [84]. Recoil corrections accounting for this effect are applied to samples  
1739 with W+jets, Z+jets, and Higgs bosons [69]. The corrections are performed on the  
1740 vectorial difference between the measured missing transverse momentum and the total  
1741 transverse momentum of neutrinos originating from the decay of the W, Z, or Higgs  
1742 boson. This vector is projected onto the axes parallel and orthogonal to the boson  
1743  $p_T$ . This vector, and the resulting correction to use, is measured in  $Z \rightarrow \mu\mu$  events,  
1744 since these events have leptonic recoil that do not contain neutrinos, allowing the  
1745 4-vector of the Z boson to be measured precisely. The corrections are binned in  
1746 generator-level  $p_T$  of the parent boson and also the number of jets in the event.

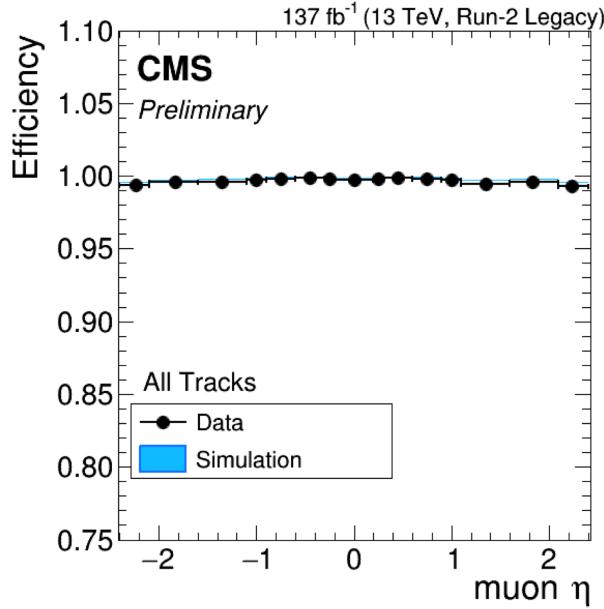


Figure 5.10: Muon tracking efficiencies as a function of  $|\eta|$  for standalone muons in Run-2 data (*black*) and Drell-Yan MC simulation (*blue*) [83]. All Tracks refers to tracks which exploit the presence of muon candidates in the muon system to seed the track reconstruction in the inner tracker, in contrast to tracks that use tracker-only hits for seeding. Uncertainties shown are statistical.

### 5.3.15 Drell-Yan corrections

The Z boson transverse momentum distribution disagrees between leading-order (LO) simulations and data in a  $Z \rightarrow \mu\mu$  control region with at least one b-tag jet [85]. Per-event weights derived by the 2016 data-only version of this analysis [85] are applied to  $Z \rightarrow \tau\tau/\ell\ell$  events, as a function of the generator-level Z boson  $p_T$  to provide better matching of MC to data.

### 5.3.16 Pile-up reweighting

Reweighting is performed to rescale MC events to account for differences between MC and data, in the distribution of the pile-up (number of additional proton-proton interactions per bunch crossing). A tool for calculating the pile-up reweighting for the MC samples used is provided centrally by the Luminosity POG [86].

1758 **5.3.17 Pre-firing corrections**

1759 In 2016 and 2017 data-taking, a gradual timing shift of ECAL was not properly  
1760 propagated to L1 trigger primitives (TPs), resulting in a large fraction of high  $\eta$   
1761 TPs being incorrectly associated with the previous bunch crossing. L1 trigger rules  
1762 prevent two consecutive bunch crossings from firing, causing events to be rejected if  
1763 significant ECAL energy was deposited in  $2.0 < |\eta| < 3.0$ . To account for this issue,  
1764 MC simulations for 2016 and 2017 are corrected using an event-dependent weight.  
1765 Embedded samples are not corrected [51].

1766 **5.3.18 Top  $p_T$  spectrum reweighing**

1767 In Run-1 and Run-2 it was observed that the  $p_T$  spectra of top quarks in  $t\bar{t}$  data  
1768 was significantly softer than those predicted by MC simulations [87]. Possible sources  
1769 of this discrepancy are higher order QCD and/or electroweak corrections, and non-  
1770 resonant production of  $t\bar{t}$ -like final states. To account for this, corrections derived  
1771 from Run-2 data by the Top Physics Analysis Group (PAG) are applied to the  $p_T$   
1772 of the top and anti-top quarks in MC simulations, computed as a function of their  
1773 generator-level  $p_T$  [87].

1774 **5.3.19 B-tagging efficiency**

1775 In order to predict correct b-tagging discriminant distributions and event yields in  
1776 data, the weight of selected MC events is reweighed according to recommendations by  
1777 the BTV POG [88]. The reweighing depends on the jet  $p_T$ ,  $\eta$ , and the b-tagging dis-  
1778 criminant. In this method, there is no migration of events from one b-tag multiplicity  
1779 bin to another.

### **5.3.20 Jet energy resolution and jet energy smearing**

Calibration of jet energies, i.e. ensuring that the energy and momentum of the reconstructed jet matches that of the quark/gluon-initiated jet, is a challenging task due to time-dependent changes in the detector response and calibration and high pile-up [89] [90]. Jet calibration is done via jet energy corrections (JECs) applied to the  $p_T$  of jets in MC samples, accounting successively for the effects of pile-up, uniformity of the detector response, and residual data-simulation jet energy scale differences [91]. Typical jet energy resolutions reported at  $\sqrt{s} = 8 \text{ TeV}$  in the central rapidities are 15-20% at 30 GeV and about 10% at 100 GeV [89]. Jet energy corrections are also propagated to the missing transverse energy.

Measurements show that the jet energy resolution (JER) in data is worse than in simulation, and so the jets in MC need to be smeared to describe the data. JER corrections are applied after JEC on MC simulations, and adjust the width of the  $p_T$  distribution based on pile-up, jet size, and jet flavor [92]. Tools for applying JEC and JER are provided centrally by the JER Corrections group.

1795 **Chapter 6**

1796 **Event selection**

1797 This chapter describes how events in data and simulated samples are selected in the  
1798 search for  $h \rightarrow aa \rightarrow bb\tau\tau$ . The event selection is motivated by optimization checks  
1799 aimed at maximizing the final expected limit, and is also based on recommendations  
1800 from CMS Physics Objects Groups. As described in the previous chapter, the tau  
1801 lepton can decay to electrons ( $e$ ), muons ( $\mu$ ), or hadronic states ( $\tau_h$ ). As a result,  
1802 several different final states of the  $\tau\tau$  system are possible, and are here referred to  
1803 as “channels” since they are mutually exclusive. The three  $\tau\tau$  final states studied in  
1804 this analysis are muon and hadronic tau ( $\mu\tau_h$ ), electron and hadronic tau ( $e\tau_h$ ), and  
1805 electron and muon ( $e\mu$ ). The procedure for dividing events into these three channels  
1806 begins with checking the High-Level Trigger paths passed by the events as detailed  
1807 in Section 6.1. Events are further accepted or rejected based on criteria applied to  
1808 the leptons in the event. These event selections are described for the  $\mu\tau_h$  channel in  
1809 Section 6.2, the  $e\tau_h$  channel in Section 6.3, and the  $e\mu$  channel in Section 6.4.

1810 **6.1 General procedure for all channels**

1811 For the search for  $h \rightarrow aa \rightarrow bb\tau\tau$ , three final states of the  $\tau\tau$  system are considered:  
1812  $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$ . The  $\tau_h\tau_h$  final state is not considered because signal events in

1813 the  $\tau_h\tau_h$  channel would typically produce hadronic taus with momenta below data-  
1814 taking trigger thresholds. In all three final states, events are required to have at  
1815 least one b-tag jet passing the medium working point of the DeepFlavour tagger,  
1816 with  $p_T > 20 \text{ GeV}$ , and  $|\eta| < 2.4$ . A second b-tag jet is not required because such a  
1817 requirement would reduce signal acceptance by 80% compared to only requiring one  
1818 b-tag jet.

1819 Events in MC samples are sorted into one of the three  $\tau\tau$  channels if they pass the  
1820 following trigger requirements and requirements on the offline reconstructed objects  
1821 in the event, first checking the HLT paths for the  $\mu\tau_h$  channel, then  $e\tau_h$ , and finally  $e\mu$ .  
1822 The two leading leptons (e.g. muon and hadronic tau for the  $\mu\tau_h$  channel) that were  
1823 determined to have originated from the  $\tau\tau$  decay, are called the  $\tau\tau$  “legs”. For events  
1824 in data and embedded samples, the HLT paths requirements for the corresponding  
1825 channel are checked.

1826 After sorting events by HLT paths and identifying the leading tau legs in the offline  
1827 reconstructed objects, the  $p_T$  of the offline objects is checked against the online trigger  
1828 thresholds. Trigger matching is also performed, which checks the correspondence  
1829 between each offline reconstructed object used in the analysis (e.g. a muon), and a  
1830 trigger object in the HLT (e.g. a HLT muon). An offline object is considered to be  
1831 matched, if it corresponds to a trigger object of the same object type, with  $\Delta R < 0.5$ .  
1832 This matched trigger object is also required to pass the filter(s) of the HLT trigger.  
1833 The trigger thresholds used for the  $bb\mu\mu$  final state and the  $bb\tau\tau$  final state (the focus  
1834 of this work) are summarized in Tables 6.1.

1835 After checking the HLT paths and trigger objects in each channel, events are  
1836 subject to further selection to ensure that they contain leptons and b-tag jet(s) of in-  
1837 terest. These requirements are summarized in Table 6.2, and detailed in the following  
1838 sections.

Year	Single/dilepton trigger $p_T$	$bb\mu\mu$	$bb\tau\tau$					
			$e\mu$		$e\tau_h$		$\mu\tau_h$	
		$\mu$	$e$	$\mu$	$e$	$\tau_h$	$\mu$	$\tau_h$
2016	Single lepton	24	–	–	25	–	22	–
	$p_T$ -leading lepton	17	23	23	–	–	–	20
	$p_T$ -subleading lepton	8	12	8	–	–	19	–
2017	Single lepton	24	–	–	27, 32	–	24, 27	–
	$p_T$ -leading lepton	17	23	23	–	30	–	27
	$p_T$ -subleading lepton	8	12	8	24	–	20	–
2018	Single lepton	24	–	–	32, 35	–	24, 27	–
	$p_T$ -leading lepton	17	23	23	–	30	–	27
	$p_T$ subleading lepton	8	12	8	24	–	20	–

Table 6.1: Trigger thresholds used for the leptons in the  $bb\mu\mu$  analysis and the  $bb\tau\tau$  analysis (the focus of this work). The thresholds for the three  $bb\tau\tau$  channels ( $e\mu$ ,  $e\tau_h$ , and  $\mu\tau_h$ ) are listed separately, with some channels and years taking the logical OR of two triggers with different thresholds.

## 6.2 Event selection in the $\mu\tau_h$ channel

In all three years, a single muon trigger is used if the muon has sufficiently high  $p_T$ , otherwise a dilepton  $\mu\tau_h$  cross-trigger is used (Tables 6.3, 6.4, and 6.5). For data taken in 2017-2018 (2016), the logical OR of the single muon triggers with online  $p_T$  thresholds 24 and 27 (23) GeV is used, with the corresponding offline muon required to have with  $p_T$  1 GeV above the online threshold. For data taken in 2017-2018 (2016), a dilepton  $\mu + \tau_h$  cross-trigger with  $p_T$  thresholds of 20 (19) and 27 (20) GeV for the muon and tau respectively, is used. The  $\tau_h$  is required to have  $|\eta| < 2.3$  if the single trigger is fired,  $|\eta| < 2.1$ .

The muon and  $\tau_h$  are required to have opposite charge and be separated by  $\Delta R > 0.4$ . The muon is required to have  $|\eta| < 2.4$ , and the  $\tau_h$  is required to have  $|\eta| < 2.3$  unless a cross-trigger is required, in which case we require  $|\eta| < 2.1$  as discussed above.

The muon is required to pass the medium identification (ID) working point [93], which is defined by the Muon POG as a loose muon (i.e. a Particle Flow muon that is either a global or a tracker muon - see Section 5.1.2) with additional requirements

All years (2016, 2017, 2018) and eras				
Kinematic variable	<i>bbμμ</i>	<i>bbττ</i>		
	$\mu$	$e\mu$	$e\tau_h$	$\mu\tau_h$
$\Delta R$ between leptons	>0.4	>0.3	>0.4	>0.4
$ \eta $ of electron	-	<2.4	<2.1	-
$ \eta $ of muon	<2.4	<2.4	-	<2.1
$ \eta $ of hadronic tau	-	-	<2.3/< 2.1	<2.3/< 2.1
Relative isolation of electron	-	<0.10	-	<0.15
Relative isolation of muon	<0.25	<0.15	-	<0.15
Leading b-tag jet $p_T$	>15 GeV		>20 GeV	
Leading b-tag jet $ \eta $	<2.4		<2.4	
Leading b-tag jet WP	Tight		Medium	
Sub-leading b-tag jet $p_T$	>15 GeV		-	
Sub-leading b-tag jet $ \eta $	<2.4		-	
Sub-leading b-tag jet WP	Loose		-	
$\Delta R$ between jet(s) and leptons	>0.4		>0.5	

Table 6.2: Summary of requirements applied to the leptons in the  $bb\mu\mu$  analysis and the  $bb\tau\tau$  analysis (the focus of this work).  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  is a measure of spatial separation. Relative isolation is defined in Eqn. 5.2 and Section 5.1.2. The b-tag jets are required to pass the listed DeepFlavour working points (WP), which are described in Section 5.1.5. In the  $bb\tau\tau$  analysis, the required  $|\eta|$  of the hadronic taus are listed for the single and cross triggers respectively. The  $bb\mu\mu$  analysis requires two b-tag jets in all events, while the  $bb\tau\tau$  analysis only requires one.

on track quality and muon quality. These identification criteria are designed to be highly efficiently for prompt muons and for muons from heavy quark decays. In addition to the ID, for prompt muons it is recommended to apply cuts on the impact parameter [93]: we apply  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ . A cut is applied on the muon relative isolation (defined in Section 5.1.2), to be less than 0.15 in a cone size of  $\Delta R = 0.4$ , which corresponds to the Tight Particle Flow isolation requirement [93].

The  $\tau_h$  is required to pass a cut on its impact parameter of  $|\Delta(z)| < 0.2$ . The  $\tau_h$  is also required to pass the VLoose (Very Loose) DeepTau working point vs. electron, the Tight DeepTau working point vs. muons, and the VVVLoose and Medium DeepTau working point vs. jets. Events with taus reconstructed in two of the decay modes (labeled 5 and 6) are rejected, since these decay modes are meant to recover 3-prong taus, but are only recommended for use in analyses where the benefits in final significance outweigh the resulting increase in background [66]. Decays reconstructed with 2 prongs are not considered as they are only recommended for taus with a very high transverse momentum, where the prongs may overlap.

For the estimation of the background from jets faking  $\tau_h$ , which is described in Section 7.7, anti-isolated events are selected, by requiring events to pass all the selections described above, except failing the Medium DeepTau working point vs. jets.

### 6.3 Event selection in the $e\tau_h$ channel

The HLT trigger paths for the  $e\tau_h$  channel are summarized in Tables 6.3, 6.4, and 6.5. Similarly to the  $\mu\tau_h$  channel, a single electron trigger is used if the electron has sufficiently high  $p_T$  in 2018 and 2017. For data taken in 2018 (2017), the OR of the single electron triggers with online  $p_T$  thresholds at 32 and 35 (27 and 32) GeV are used, with the corresponding offline electrons required to have  $p_T$  greater than 33 (28) GeV. A  $e + \tau_h$  cross-trigger is used for electrons with lower offline  $p_T$  between

1880 25 and 33 GeV (25 and 28 GeV). For the 2016 dataset, there is no cross trigger but  
1881 only a single electron trigger with online  $p_T$  threshold at 25 GeV, which is used if the  
1882 offline electron has  $p_T$  greater than 26 GeV.

1883 The electron and  $\tau_h$  are required to have opposite charge and be separated by  
1884  $\Delta R > 0.4$ . The electron is required to be within  $|\eta| < 2.3$  when no cross trigger is  
1885 used, and  $|\eta| < 2.1$  when the cross trigger is fired. The  $\tau_h$  is required to have  $|\eta| < 2.3$   
1886 if no cross trigger is fired, and have  $|\eta| < 2.1$  if the cross trigger is fired.

1887 The electron is required to have a relative isolation (same definition as in Section  
1888 5.1.2) of less than 0.1 in a cone size of  $\Delta R = 0.3$ , which is the standard recommended  
1889 cone size giving minimal pile-up dependence and reduced probability of other objects  
1890 overlapping with the cone. The isolation quantity used includes an “effective area”  
1891 (EA) correction to remove the effect of pile-up in the barrel and endcap parts of the  
1892 detector [94]. The electron is also required to pass cuts on its impact parameter of  
1893  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ . It is also required to pass the non-isolated MVA  
1894 working point corresponding to 90% efficiency. The electron’s number of missing hits,  
1895 which are gaps in its trajectory through the inner tracker [94], must be less than or  
1896 equal to 1. The electron must pass a conversion veto, which rejects electrons coming  
1897 from photon conversions in the tracker, which should instead be reconstructed as part  
1898 of the photon [94].

1899 The impact parameter cut for the  $\tau_h$  is  $|\Delta(z)| < 0.2$ . In contrast to the  $\mu\tau_h$  event  
1900 selection, the vs. electron and vs. muon DeepTau working points are flipped, to  
1901 reject muons faking the  $\tau_h$  leg. The  $\tau_h$  is required to pass the Tight DeepTau working  
1902 point vs. electrons, the VLoose DeepTau working point vs. muons, and the Medium  
1903 DeepTau working point vs. jets.

1904 As in the  $\mu\tau_h$  channel, for the estimation of the background from jets faking  $\tau_h$ ,  
1905 which is described in Section 7.7, anti-isolated events are selected by requiring events  
1906 to pass all the selections described above, except failing the Medium DeepTau working

1907 point vs. jets.

## 1908 6.4 Event selection in the $e\mu$ channel

1909 The HLT trigger paths for the  $e\mu$  channel are summarized in Tables 6.3, 6.4, and  
1910 6.5. Events are selected with the logical OR of two  $e + \mu$  cross triggers, where either  
1911 the electron or muon can have larger  $p_T$ : (1) leading electron, where the electron has  
1912 online  $p_T > 23\text{ GeV}$  and muon has online  $p_T > 8\text{ GeV}$ , or (2) leading muon, where  
1913 electron has online  $p_T > 12\text{ GeV}$  and muon has online  $p_T > 23\text{ GeV}$ .

1914 The leading and sub-leading leptons are required to have an offline  $p_T$  greater than  
1915 1 GeV above the online threshold (i.e.  $p_T > 24\text{ GeV}$ ). If the sub-leading lepton is the  
1916 electron, the offline  $p_T$  threshold is 1 GeV above the online threshold ( $p_T > 13\text{ GeV}$ ),  
1917 but if it is a muon, the offline  $p_T$  threshold is required to be at least 5 GeV greater  
1918 than the online threshold (i.e.  $p_T > 13\text{ GeV}$ ). This is because of poor data and  
1919 simulation agreement for low- $p_T$  muons with  $p_T$  between 9 GeV and 13 GeV, and the  
1920 higher probability of mis-identifying jets as muons at lower  $p_T$ . With no effect on  
1921 the expected limits, the offline  $p_T$  threshold for muons is raised to 13 GeV instead of  
1922 9 GeV, even though it may lead to loss in signal acceptance. Both the electron and  
1923 muon are required to have  $|\eta| < 2.4$ .

1924 The electron and muon are required to have opposite charge and be separated  
1925 by  $\Delta R > 0.3$  (note the decreased separation requirement compared to the other  
1926 two channels). The electron is required to pass the non-isolated MVA identification  
1927 working point corresponding to 90% efficiency, and to have a relative isolation less  
1928 than 0.1 for a cone size of  $\Delta R = 0.3$  with the EA pile-up subtraction correction.  
1929 The electron must have one or fewer missing hits and pass the conversion veto (both  
1930 described previously in Section 6.3).

1931 The muon is required to pass the medium identification working point (described

1932 earlier in 6.2), and to have a relative isolation less than 0.15 for a cone size of  $\Delta R =$   
1933 0.4. The muon impact parameter is required to have  $|\Delta(z)| > 0.2$  and  $|\Delta(xy)| < 0.045$ .

1934 For the QCD multijet background estimation described in Section 7.8, the same-  
1935 sign region is selected by requiring all the above selections, except the legs are required  
1936 to have the same electric charge rather than opposite.

2016 $\mu\tau_h$ trigger paths	
Notes	HLT Path
	HLT_IsoMu22_v
	HLT_IsoMu22_eta2p1_v
	HLT_IsoTkMu22_v
	HLT_IsoTkMu22_eta2p1_v
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_v
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1_v
2016 $e\tau_h$ trigger paths	
Notes	HLT Path
	HLT_Ele25_eta2p1_WPTight_Gsf_v
2016 $e\mu$ trigger paths	
Notes	HLT Path
runs B-F and MC	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v
runs B-F and MC	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v
runs G-H	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
runs G-H	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v

Table 6.3: High-Level Trigger (HLT) paths used to select data and simulation events in 2016 for the three  $\tau\tau$  channels.

## 6.5 Extra lepton vetoes in all channels

1938 Events containing a third lepton (electron or muon) that is neither of the leading  $\tau\tau$   
1939 legs are rejected, and events with di-muons and di-electrons are vetoed, with criteria  
1940 taken from the Standard Model  $H \rightarrow \tau\tau$  working group [69]. These vetoes on extra  
1941 leptons also ensure orthogonality of events to analyses such as the  $bb\mu\mu$  final state,  
1942 whose results are combined with this  $bb\tau\tau$  final state as described in Section 10.2.

1943 The event is vetoed if a third electron is found with the following properties:  
1944  $p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.5$ , impact parameter  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ , passing

2017 $\mu\tau_h$ trigger paths	
Notes	HLT Path
	HLT_IsoMu24_v
	HLT_IsoMu27_v
	HLT_IsoMu20_eta2p1_LooseChargedIso_PFTau27_eta2p1_CrossL1_v
2017 $e\tau_h$ trigger paths	
Notes	HLT Path
	HLT_Ele32_WPTight_Gsf_v
	HLT_Ele35_WPTight_Gsf_v
	HLT_Ele24_eta2p1_WPTight_Gsf_Loose_ChargedIsoPFTau30_eta2p1_CrossL1_v
2017 $e\mu$ trigger paths	
Notes	HLT Path
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v

Table 6.4: High-Level Trigger (HLT) paths used to select data and simulation events in 2017 for the three  $\tau\tau$  channels.

1945 non-isolation MVA identification with 90% efficiency, conversion veto,  $\leq 1$  missing  
 1946 hits, and relative isolation  $< 0.3$  with cone size  $\Delta R = 0.3$ . The event is also vetoed if  
 1947 a third muon is found with the following properties:  $p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.4$ , impact  
 1948 parameter  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ , medium ID, and isolation  $< 0.3$  with  
 1949 cone size  $\Delta R = 0.4$ .

1950 A di-muon veto is applied, which rejects events containing a pair of muons with  
 1951 opposite charge and separation of  $\Delta R > 0.15$ , that both pass the following selections:  
 1952  $p_T > 15 \text{ GeV}$ ,  $|\eta| < 2.4$ , flag for global muons, flag for tracker muon, flag for Particle  
 1953 Flow muon,  $|\Delta(z)| < 0.2$ ,  $|\Delta(xy)| < 0.045$ , and isolation  $< 0.3$  with cone size  $\Delta R =$   
 1954 0.4. A similar di-electron veto is applied to reject events containing a pair of electrons  
 1955 with opposite charge and separation of  $\Delta R > 0.15$ , that both pass the following  
 1956 selections:  $p_T > 15 \text{ GeV}$ ,  $|\eta| < 2.5$ , a dedicated electron ID (cut-based) for vetoing  
 1957 third leptons,  $|\Delta(z)| < 0.2$ ,  $|\Delta(xy)| < 0.045$ , with pile-up corrected relative isolation  
 1958  $< 0.3$  with cone size  $\Delta R = 0.3$ .

2018 $\mu\tau_h$ trigger paths	
Notes	HLT Path
	HLT_IsoMu24_v
	HLT_IsoMu27_v
only data run < 317509	HLT_IsoMu20_eta2p1_ (contd.)
	LooseChargedIsoPFTauHPS27_eta2p1_CrossL1_v
MC and data run $\geq$ 317509	HLT_IsoMu20_eta2p1_ (contd.)
	LooseChargedIsoPFTauHPS27_eta2p1_TightID_CrossL1_v
2018 $e\tau_h$ trigger paths	
Notes	HLT Path
	HLT_Ele32_WPTight_Gsf_v
	HLT_Ele35_WPTight_Gsf_v
only data run < 317509	HLT_Ele24_eta2p1_WPTight_Gsf_ (contd.)
	LooseChargedIsoPFTauHPS30_eta2p1_CrossL1_v
MC and data run $\geq$ 317509	HLT_Ele24_eta2p1_WPTight_Gsf_ (contd.)
	LooseChargedIsoPFTauHPS30_eta2p1_TightID_CrossL1_v
2018 $e\mu$ trigger paths	
Notes	HLT Path
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v

Table 6.5: High-Level Trigger (HLT) paths used to select data and simulation events in 2018 for the three  $\tau\tau$  channels. In 2018 a HLT trigger path using the hadron plus strips (HPS) tau reconstruction algorithm became available.

# <sup>1959</sup> Chapter 7

## <sup>1960</sup> Background estimation

<sup>1961</sup> This section describes methods used to estimate backgrounds from Standard Model  
<sup>1962</sup> processes in the search for  $h \rightarrow aa \rightarrow bb\tau\tau$ . The background contributions directly  
<sup>1963</sup> taken from MC are described in Sections 7.1 to 7.6. Section 7.7 describes the data-  
<sup>1964</sup> driven method for estimating backgrounds from jets faking hadronic tau decays (jet  
<sup>1965</sup>  $\rightarrow \tau_h$ ), which is used in the  $\mu\tau_h$  and  $e\tau_h$  channels. Section 7.8 describes the data-driven  
<sup>1966</sup> method for estimating background from quantum chromodynamic (QCD) processes  
<sup>1967</sup> in the  $e\mu$  channel.

### <sup>1968</sup> 7.1 Z+jets

<sup>1969</sup> A major source of background for  $\tau\tau$  analyses is the Drell-Yan (DY) process (Z+jets).  
<sup>1970</sup> The Z boson decays to  $\tau\tau/\mu\mu/ee$  with equal probability of 3.4% each, with the domi-  
<sup>1971</sup> nant decay modes being to hadrons (around 70%) and neutrinos (invisible) (20%) [27].

<sup>1972</sup> The Drell-Yan contribution with genuine taus,  $Z \rightarrow \tau\tau$ , is estimated using embed-  
<sup>1973</sup> ded samples, described in Section 4.3. To avoid double-counting between embedded  
<sup>1974</sup> and MC samples, in all MC samples, events with legs that originated from genuine  $\tau$   
<sup>1975</sup> are discarded.

<sup>1976</sup> The other decays of the Z,  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$ , are estimated from MC simulation,

and are hereafter referred to as simply the Drell-Yan background. These MC samples are generated to leading order (LO) with different numbers of jets (jet multiplicity) in the matrix element: Z+1 jet, Z+2jets, Z+3 jets, Z+4 jets, and inclusive Z+jets. The cross-sections of the samples with  $\geq 1$  jets are normalized to next-to-NLO (NNLO) in QCD. For the inclusive Drell-Yan sample, two samples are used with different thresholds for the di-lepton invariant mass ( $m_{\ell\ell}$ ) at the generator level: one with  $m_{\ell\ell} > 50$  GeV and the other with  $10 < m_{\ell\ell} < 50$ .

## 7.2 W+jets

The dominant W boson decay modes are to hadrons (67.4%),  $e + \nu_e$  (10.7%),  $\mu + \nu_\mu$  (10.6%), and  $\tau + \nu_\tau$  (11.4%) [27]. The W+jets background is estimated from MC simulation. Similarly to the Z+jets, the W+jets samples are generated with different jet multiplicities in the matrix element. LO samples are used for greater statistics and are normalized to NNLO cross sections.

## 7.3 $t\bar{t}$ + jets

In hadron collisions, top quarks are produced singly with the weak interaction, or in pairs via the strong interaction, with interference between these leading-order processes possible in higher orders of the perturbation theory. The top quark is the heaviest fermion in the Standard Model and has a short lifetime ( $\sim 10^{-25}$  s), decaying without hadronization into a bottom quark and a W boson [27], with the decay modes of the W boson as listed in the previous section. With two top quarks, the final states of the two resulting W bosons can be described as fully leptonic, semileptonic, and fully hadronic. These three final states are modeled separately with MC simulation in 2018 and 2017, while for 2016 the sample used is inclusive.

## 2000 7.4 Single top

2001 There are three main production modes of the single top in  $pp$  collisions [95]: the  
2002 exchange of a virtual W boson ( $t$  channel), the production and decay of a virtual W  
2003 boson ( $s$  channel), and the associated production of a top quark and W boson ( $tW$ ,  
2004 or W-associated) channel. As the  $s$  channel process is rare and only 3% of the total  
2005 production, the dominant production mode of the  $t$ -channel and the  $tW$  production  
2006 are considered and modeled with MC.

## 2007 7.5 Diboson

2008 In  $pp$  collisions, the production of dibosons (pairs of electroweak gauge bosons, i.e.  
2009 WW, WZ, and ZZ) is dominated by quark-antiquark annihilation, with a small con-  
2010 tribution from gluon-gluon interactions [96]. MC is used to model the pair production  
2011 and decays of VV to  $2\ell 2\nu$ , WZ to  $2q 2\ell$  and  $3\ell\nu$ , and ZZ to  $4\ell$  and  $2q 2\ell$  ( $q$  being quarks  
2012 and  $\ell$  being leptons).

## 2013 7.6 Standard Model Higgs

2014 MC is used to simulate backgrounds from major production modes of the Standard  
2015 Model 125 GeV Higgs boson: gluon-gluon fusion (ggH), vector boson fusion (VBF),  
2016 associated production with a W or Z (WH, ZH), and associated production with a  
2017 top pair (ttH) (see Fig. 7.1 for leading-order diagrams). For these production modes,  
2018 samples with the Higgs decaying to  $\tau\tau$  or to  $WW$  are used. Samples made with  
2019 higher-order diagrams for WH and ZH that include the production of a jet, with the  
2020 Higgs decaying to WW, are also used.

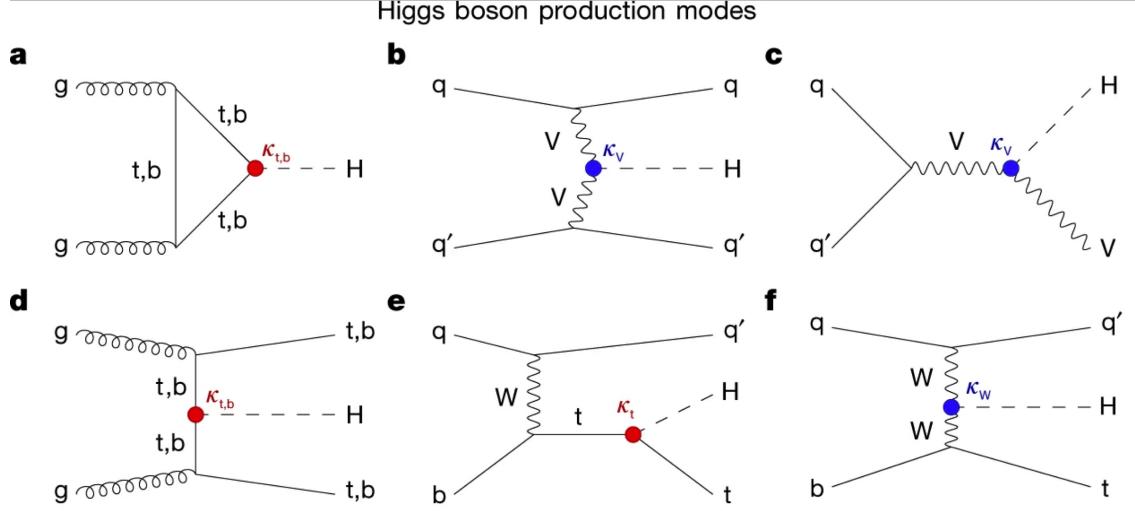


Figure 7.1: Leading-order Feynman diagrams of Higgs production from [97], in ggH (a) and vector boson fusion (VBF; b), associated production with a W or Z (V) boson (VH; c), associated production with a top or bottom quark pair (ttH or bbH); d, and associated production with a single top quark (tH; e, f).

## 2021 7.7 Jet faking $\tau_h$

2022 Events with a jet mis-reconstructed as the hadronic tau leg  $\tau_h$  are a major source of  
 2023 background in the  $\mu\tau_h$  and  $e\tau_h$  channels. The main processes contributing to jet  $\rightarrow \tau_h$   
 2024 events are QCD multijet, W+jets, and  $t\bar{t}$  production. These events are estimated  
 2025 using a data-driven method adapted from past analyses [51] [85]. This background  
 2026 includes contributions from W+jets, QCD multijets, and  $t\bar{t}$ +jets. To estimate this  
 2027 background, a sideband region is constructed, where events are required to pass all  
 2028 baseline  $\mu\tau_h/e\tau_h$  selection criteria, but fail the  $\tau_h$  isolation criteria. The events in  
 2029 this sideband region are reweighed with a factor  $f/(1 - f)$ , where  $f$  is the probability  
 2030 for a jet to be misidentified as a  $\tau_h$ . The jet  $\rightarrow \tau_h$  background is the anti-isolated,  
 2031 reweighed MC and embedded events subtracted from the anti-isolated, reweighted  
 2032 data events.

2033 The fake factor is measured in  $Z \rightarrow \mu\mu + \text{jets}$  events in data in the  $\mu\mu\tau_h$  final state,  
 2034 as any reconstructed  $\tau_h$  in these events must originate from a jet. The two muons are  
 2035 required to be isolated ( $< 0.15$ ), have opposite electric charge, and have an invariant

2036 mass between 76 and 106 GeV (close to the Z mass). These events are selected with  
2037 a double muon trigger, with the leading muon having offline  $p_T > 20$  GeV and the  
2038 subleading muon  $p_T > 10$  GeV. Simulated diboson (ZZ and WZ) events are subtracted  
2039 to avoid contamination from events with real  $\tau_h$ . The denominator of the fake rate  
2040 corresponds to fake taus passing the VVVLoose working point of the discriminator vs.  
2041 jets, while the numerator corresponds to those passing the Medium working point,  
2042 i.e.  $f = N_{\text{jet passing tight}} / N_{\text{jet passing loose}}$ .

2043  $f$  is measured as a function of the  $\tau_h$  transverse momentum and is 8% - 10% in  
2044 each of the data-taking years.  $f$  is derived separately for the  $\mu\tau_h$  and  $e\tau_h$  channels  
2045 because the channels use different anti-lepton identification working points.

## 2046 7.8 QCD multijet background

2047 In the  $e\mu$  channel, events with jets faking electrons or muons originating from QCD  
2048 multijet, is estimated from data events with the same baseline selection as in the  
2049 signal region, except with same-signed (SS) charged  $e + \mu$ , ensuring orthogonality  
2050 with the signal region which requires opposite-sign (OS)  $e\mu$  pairs. All same-sign MC  
2051 events (both events with real and fake  $e + \mu$ ) are subtracted from same-sign data  
2052 events to remove contamination from other backgrounds. i.e.  $\text{QCD}_{\text{SS}} = \text{Data}_{\text{SS}} -$   
2053  $\text{MC}_{\text{SS}}$ .

2054 Three scale factors are applied to the  $\text{QCD}_{\text{SS}}$  events to compute the QCD multijet  
2055 background [85] [41]:

- 2056 • *OS-to-SS scale factor*: This scales the SS QCD to the OS region, and is mea-  
2057 sured from an orthogonal region with an isolated electron and an anti-isolated  
2058 muon. Only the muon is chosen to be anti-isolated because this scale factor was  
2059 observed to depend more strongly on electron isolation than that of the muon.  
2060 This scale factor is treated as a function of the  $\Delta R$  separation of the trajectories

2061 of the electron and muon, and is measured separately for events with 0 jets, 1,  
2062 jet, and greater than 1 jet.

- 2063 • *2D closure correction for the lepton  $p_T$ :* This factor accounts for subleading  
2064 dependencies of the first scale factor on the  $p_T$  of the two leptons. A 2D weight  
2065 is derived in a similar fashion, as a ratio of  $\text{QCD}_{OS}$  events to  $\text{QCD}_{SS}$  events,  
2066 but parameterized by both electron and muon  $p_T$ , where the SS events have the  
2067 previous scale factor applied.
- 2068 • *Isolation correction for the muon:* The third and final factor is an isolation  
2069 correction, which is a bias correction to account for the fact that the fake  
2070 factor was determined for less-isolated muons. This factor is obtained as the  
2071 ratio of the OS-to-SS scale factors measured in two other control regions: (1)  
2072 events where the electron is anti-isolated ( $0.15 < \text{iso} < 0.5$ ) and the muon is  
2073 isolated, and (2) events where both leptons are anti-isolated.

2074

# Chapter 8

2075

## Systematic uncertainties

2076 Uncertainties in the measurement of a physical observable can be statistical or sys-  
2077 tematic in nature. Statistical uncertainties originate from limitations on the number  
2078 of events and experiments that can be performed. Systematic uncertainties arise  
2079 from the dependence of the physical observable on quantities whose exact values are  
2080 unknown and which can only be modeled imperfectly.

2081 The handling of systematic uncertainties is separated into normalization uncer-  
2082 tainties (those that affect the total yield of a variables' distribution) and shape un-  
2083 certainties (those that shift the distribution of events). Normalization uncertainties  
2084 are expressed as multiplicative factors, while shape uncertainties are represented as  
2085 up and down shifts of a variable's distribution.

2086 Up/down shifts of shape uncertainties can change the number of background  
2087 events in a distribution. For instance, hadronic taus receive corrections from the  
2088 nominal tau energy scale, with the nominal, up, and down energy scales provided  
2089 centrally by CMS. For the  $\mu\tau_h$  channel, an event could have a  $\tau_h$  with  $p_T$  just below  
2090 the offline threshold of 20 GeV (for instance, 19.5 GeV), so in the nominal distribution  
2091 of  $m_{\tau\tau}$  (or any other variable for this channel), the event is excluded. However, when  
2092 we build our distributions with the tau energy scale “up” shift, the energy of this  $\tau_h$

2093 may be scaled up to, say, 20.5 GeV, and now the event passes the offline  $p_T$  threshold  
2094 for the single muon trigger, leading to the event’s inclusion in the distributions made  
2095 with the tau energy scale “up” shift.

2096 In evaluating the up and down shifts of a specific source of uncertainty, all other  
2097 corrections and scale factors are held at their nominal values, and the full chain of  
2098 object and event selection and event categorization is performed to obtain the observ-  
2099 able distributions. Any “downstream” variables that depend on the shifted variable,  
2100 e.g. the invariant di-tau mass  $m_{\tau\tau}$ , must be computed for the nominal case, and then  
2101 re-computed separately for each up and down shift of the tau legs’ energy scale. The  
2102 objective of this process is to quantify the effect of a single source of uncertainty on  
2103 the resulting observable distributions. Each scale factor and correction described in  
2104 Section 5.3 has an associated uncertainty. The binning of the uncertainties follows  
2105 that of the nominal scale factor value.

2106 Sections 8.1 to 8.5 describe uncertainties associated with physics objects, and  
2107 Sections 8.6 and 8.7 describe uncertainties associated with sample-level effects. The  
2108 pulls and impacts for the top sixty most important systematics are shown in Section  
2109 8.8.

## 2110 8.1 Uncertainties in the lepton energy scales

2111 The uncertainties in the tau energy scales [66] are binned by the tau decay mode and  
2112 are taken as shape uncertainties treated as uncorrelated across the tau decay modes  
2113 and years. Same as with the application of the nominal scale factor, when applying  
2114 the up or down shifts, the missing transverse energy ( $p_T^{\text{miss}}$ ) of the event is adjusted  
2115 so that the 4-vector sum of the tau  $p_T^{\text{miss}}$  is unchanged.

2116 The uncertainties in the muon energy scale [67] are 0.4% for  $|\eta| < 1.2$ , 0.9% for  
2117  $1.2 < |\eta| < 2.1$ , and 2.7% for  $2.1 < |\eta| < 2.4$ , and are treated as shape uncertainties,

2118 fully uncorrelated between embedded and MC samples.

2119 The uncertainties in the electron energy scale [70] in MC are binned in the electron  
2120  $|\eta|$  and  $p_T$ , and are shown in Fig. 5.2. The uncertainties range from 0.5% to 2.2% in  
2121 the barrel, and 0.3% to 4.1% in the endcap, across the  $p_T$  range. The uncertainties  
2122 for the embedded sample are binned only in  $|\eta|$  and are on the order of 0.5% and  
2123 1.25% for the barrel and endcap [74].

2124 There are also uncertainties in the energy scales for electrons and muons misiden-  
2125 tified as  $\tau_h$ . The uncertainty for muons misidentified as  $\tau_h$  is 1% [66]. For electrons  
2126 misidentified as  $\tau_h$ , the uncertainty is binned in barrel/endcap  $\eta$  and by 1-prong and  
2127 1-prong +  $\pi_0$  decays. The probability for  $e/\mu$  faking a 3-prong decay mode is much  
2128 lower.

## 2129 8.2 Uncertainties from other lepton corrections

2130 Uncertainties associated with the  $\tau_h$  identification efficiencies are treated as shapes,  
2131 uncorrelated across the seven  $p_T$  bins and years. The shape uncertainties in the em-  
2132 bedded samples are taken as 50% correlated with those of the MC samples. The  
2133 uncertainties on electron and muon identification efficiencies are taken as normaliza-  
2134 tion uncertainties of 2% each, with a 50% correlation between embedded and MC  
2135 samples.

2136 In the  $e\tau_h$  channel, there is an additional uncertainty for the vs. jet discrimination  
2137 efficiency [66], because the analysis uses a looser anti-lepton working point (VLoose  
2138 WP) than the working points used in the measurement of the efficiency (namely,  
2139 VLoose WP vs e, and Tight WP vs mu). For nominal  $\tau_h p_T < 100$  GeV, an additional  
2140 uncertainty of 3% (5%) is used in MC (embedded), and for high  $p_T$  an uncertainty of  
2141 15% is used for both.

2142 The uncertainties in trigger efficiencies are taken as shapes [66]. In the  $e\tau_h$  and  $\mu\tau_h$

2143 channels, there are uncertainties for the single and cross lepton triggers, and in the  
2144  $e\mu$  channel there is one uncertainty each for the two  $e + \mu$  triggers, and one combined  
2145 uncertainty since their trigger phase spaces are not mutually exclusive.

## 2146 8.3 Uncertainties from jet energy scale and resolution

2147

2148 The jet energy scale uncertainties are taken as shape uncertainties: there are eleven  
2149 in total, with seven correlated across years (labeled “Year” below) and the remainder  
2150 uncorrelated across years. They affect the b-tag jet  $p_T$  and mass, and hence the  
2151 missing transverse energy  $p_T^{\text{miss}}$ . The shifts are propagated through the b-tagging  
2152 scale factor calculation and b-tag jet counting.

2153 The uncertainties in the jet energy correction and resolution [89] [98] are as follows:

- 2154 • *Absolute, AbsoluteYear*: flat absolute scale uncertainties.
- 2155 • *BBEC1, BBEC1Year*: for sub-detector regions, with barrel “BB” in  $|\eta| < 1.3$   
2156 and endcap region 1 “EC1”:  $1.3 < |\eta| < 2.5$ .
- 2157 • *EC2, EC2 year*: for sub-detector regions, with endcap region 2 “EC2” in  $2.5 <$   
2158  $|\eta| < 3.0$ .
- 2159 • *HF, HF year*: for sub-detector regions, with hadron forward “HF” in  $|\eta| > 3$ .
- 2160 • *FlavorQCD*: for uncertainty in jet flavor (uds/c/b-quark and gluon) estimates  
2161 based on comparing Pythia and Herwig (different MC generator) predictions.
- 2162 • *RelativeBal*: account for difference between log-linear fits of the two methods  
2163 used to study the jet energy response: MPF (missing transverse momentum  
2164 projection fraction) and  $p_T$  balance.

2165     • *RelativeSample*: account for  $\eta$ -dependent uncertainty due to a difference be-  
2166       tween relative residuals, observed with dijet and Z+jets in Run D of 2018 data.

2167     • *JetResolution*: uncertainty in the jet energy resolution.

## 2168 8.4 Uncertainties from b-tagging scale factors

2169 The b-tagging scale factor has its own set of associated uncertainties (not to be  
2170 confused with shifts in the b-tagging scale factor due to the propagation of the jet  
2171 energy scale uncertainties described in the previous section 8.3). They are:

- 2172     • *hf*: contamination from heavy flavor (b+c) jets in the light flavor region.
- 2173     • *hfstats1, hfstats2*: linear and quadratic statistical fluctuations from b-flavor jets.
- 2174     • *lf*: contamination from light flavor (udsg+c jets) in the heavy flavor region.
- 2175     • *lfstats1, lfstats2*: linear and quadratic statistical fluctuations from udsg jets.
- 2176     • *cferr, cferr2*: uncertainty for charm jets.

2177 The variations for “lf, hf, hfstats1/2, lfstats1/2” are applied to both b and udsg jets.

2178 For c-flavor jets, only “cferr1/2” is applied.

## 2179 8.5 Uncertainties from MET

2180 Samples where recoil corrections were applied (Z+jets, W+jets, and Standard Model  
2181 Higgs, as described in Section 5.3) have uncertainties from the response and resolution  
2182 of the hadronic recoil against the leptonic system. These are each binned in jet  
2183 multiplicity.

## 2184 8.6 Uncertainties associated with samples used

2185 Normalization uncertainties related to the samples used are:

2186 • *Cross-section uncertainties*:  $\sigma(t\bar{t})$ : 4.2%,  $\sigma(\text{diboson})$ : 5%,  $\sigma(\text{single top})$ : 5%,  
2187  $\sigma(\text{ggH})$ : 3.2%,  $\sigma(\text{qqH})$ : 2.1%,  $\sigma(\text{WH})$ : 1.9%,  $\sigma(\text{ZH})$ : 1.3%,  $\sigma(\text{ttH})$ : 3.6%

2188 • *Uncertainties in QCD renormalization scale*: QCD scale(qqH): +0.43%-0.33%,  
2189 QCD scale(WH): +0.5%-0.7%, QCD scale(ttH): +5.8%-9.2%

2190 • *Branching ratio uncertainties*:  $\text{BR}(\text{H} \rightarrow \tau\tau)$ : 1.8%, and  $\text{BR}(\text{H} \rightarrow \text{WW})$ : 1.5%.

2191 • *Normalization uncertainties*: 2% for Drell-Yan, 4\$ for embedded, 20% pre-fit  
2192 for the QCD multijet background in the  $e\mu$  channel, 20% pre-fit for the jet  
2193 faking background.

2194 The  $t\bar{t}$  process has additional acceptance uncertainties from QCD scale variation  
2195 and parton shower uncertainties [99]. Parton shower uncertainties originate from  
2196 the modeling of perturbative and non-perturbative QCD effects handled in parton  
2197 shower MC generators. The scale variations are determined from the envelope of the  
2198 6 provided shapes due to variations in the factorization scale, renormalization scale,  
2199 and their combined variation [99].

2200 The uncertainty in the Z  $p_T$  reweighting in Drell-Yan samples is taken as a shape  
2201 uncertainty and the up and down values are 0.9 and 1.1 times the nominal reweighting.  
2202 This 10% uncertainty is sufficient to cover uncertainties in the weights derived from  
2203 the discrepancies between LO simulations and data in the di-muon mass in  $Z \rightarrow \mu\mu$   
2204 events.

2205 The weight applied to anti-isolated events in the  $\mu\tau_h$  and  $e\tau_h$  channels to estimate  
2206 the background from jets faking  $\tau_h$ , has shape uncertainties covering uncertainties in  
2207 the derivation of the weight. There are six shape uncertainties corresponding to the  
2208 binning of the fake rate in the  $\tau_h$  transverse momentum. For the weight applied to

2209 scale up anti-isolated events in cross-trigger regions, 20% of the nominal weight is  
2210 taken as a shape uncertainty.

## 2211 8.7 Other uncertainties

2212 A 3.6% yield uncertainty in the signal is used to cover uncertainties in the parton  
2213 distribution functions (PDFs), knowledge of the  $\alpha_s$  (fine structure constant), and  
2214 QCD scale. The size of these uncertainties was estimated by a different analysis  
2215 searching for two light scalars decaying to four muons, which compared the PDFs  
2216 from different model libraries using recommendations from the PDF4LHC Working  
2217 Group [100] [101].

2218 Uncertainties in the luminosity measurements can originate from uncertainties  
2219 in the luminosity calibration in the van de Meer scan procedure and from detector  
2220 operations [44]. Some effects are fully uncorrelated (e.g. if the systematic error is  
2221 limited by the statistical uncertainty in the calibration scans taken independently in  
2222 each year), and some are correlated, for example in the 2017 and 2018 measurements  
2223 which used a method with the same systematic bias. The luminosity normalization  
2224 uncertainties are applied all MC samples, divided into those uncorrelated across years  
2225 (0.26% for 2016, 0.60% for 2017, and 0.65% for 2018), one correlated between 2017  
2226 and 2018 (0.27%), and one correlated between all three years (1.30%) [42] [43] [44] [86].

## 2227 8.8 Pulls and impacts

2228 The top impacts and pulls computed for the combination of all channels and years is  
2229 shown in Fig. 8.1. The top impacts are related to uncertainty in the signal sample and  
2230 cross-section of the  $t\bar{t}$  cross-section, and also the yields of the jet faking  $\tau_h$  background,  
2231 which is a major background in all channels and expected to be constrained due to  
2232 the yield uncertainty which is taken to be 20% pre-fit.

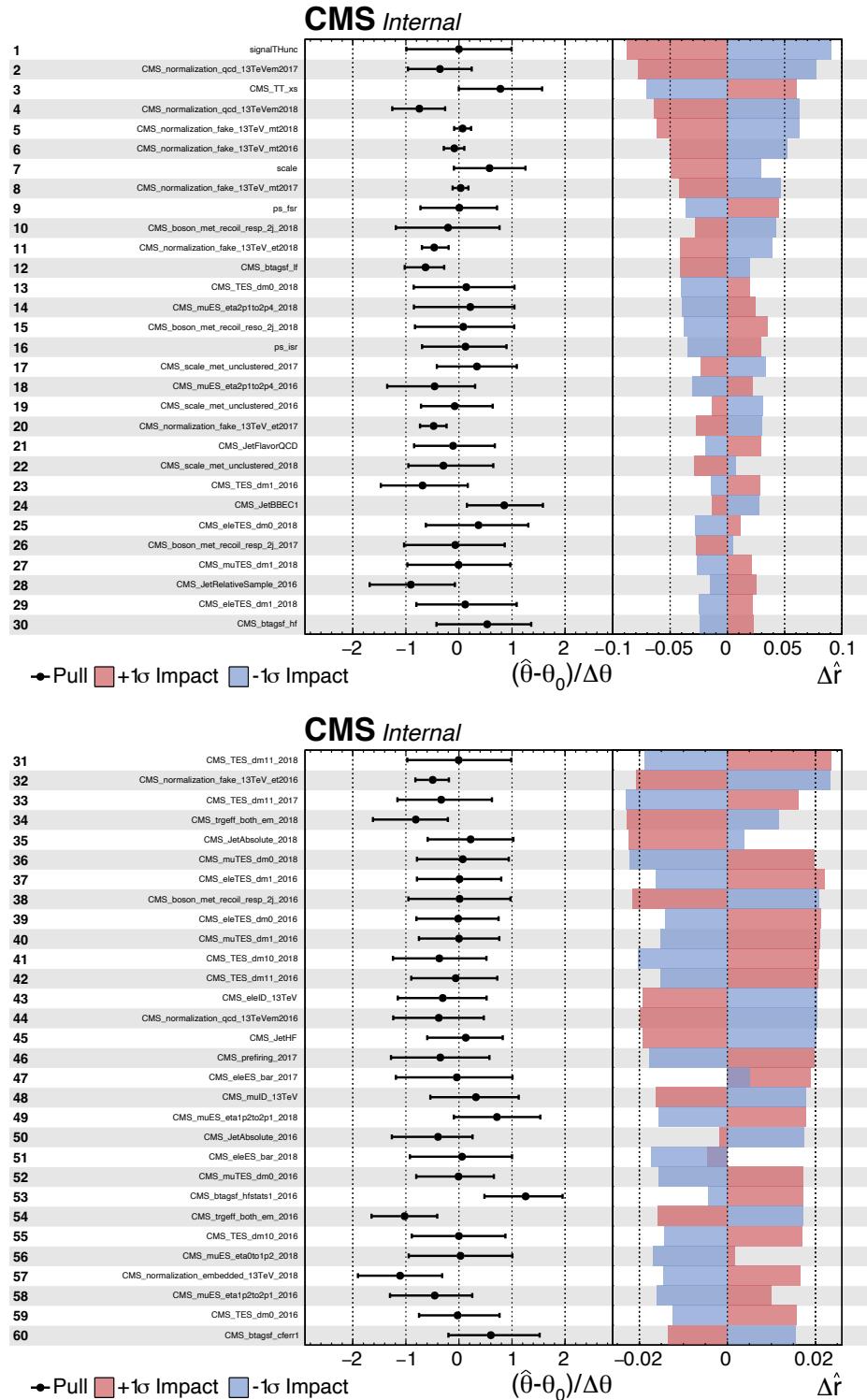


Figure 8.1: Top sixty pulls and impacts for the combination of all channels and years [102].

2233 **Chapter 9**

2234 **Event categorization and signal  
extraction**  
2235

2236 Measured events are divided into categories, based on cuts on values of observables  
2237 in the event, or some derived quantity based on the observables in the event. The  
2238 objective of event categorization is to divide events into signal regions, where the  
2239 signal is enhanced and the background is suppressed, and control regions, which are  
2240 signal-poor and used to check that the background estimation methods employed in  
2241 the analysis in fact accurately models the data. In this analysis, events in each di-tau  
2242 channel ( $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$ ) are selected to contain one or more b-tag jets reconstructed  
2243 in the event as described in Section 9.1. Events are further divided into signal and  
2244 control regions using a deep learning-based approach described in Section 9.2. The  
2245 signal is extracted from the di-tau mass distribution in the signal region using the  
2246 statistical procedure described in Section 9.3.

2247 **9.1 B-tag jet multiplicity**

2248 Compared to the previous CMS  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis which used 2016 data corre-  
2249 sponding to an integrated luminosity of  $35.9 \text{ fb}^{-1}$  [85], this analysis is performed on

the full Run-2 dataset corresponding to an integrated luminosity of  $138 \text{ fb}^{-1}$ . The increased statistics enables the separation of events into events with exactly 1 b-tag jet and events with greater than 1 b-tag jet, which was not possible in the previous analysis. Further event categorization is performed with deep neural networks (DNNs) described below. The DNNs are used only for separating events into signal and control regions in the 1 b-tag and 2 b-tag jets scenarios, and the final results are extracted from the di-tau mass.

## 9.2 DNN-based event categorization

Neural networks for event categorization are trained for each of the  $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$  channels, for 1 and 2 b-tag jets, giving  $3 \times 2 = 6$  networks in total for each year. In the training, the signal is taken to be all of the possible pseudoscalar mass  $m_a$  hypotheses added together. The backgrounds for each DNN are taken to be a representative combination of the three major backgrounds:  $Z \rightarrow \tau\tau$ ,  $t\bar{t}+\text{jets}$ , and jet faking  $\tau_h$  backgrounds. The proportions of each background for each channel and b-tag jet multiplicity are taken from the yields in the  $m_{\tau\tau}$  distribution. For instance, in the  $\mu\tau_h$  1 b-tag jet category, the composition of the background for training is 17.4% from  $Z \rightarrow \tau\tau$ , 42.4% from  $t\bar{t}+\text{jets}$ , and 40.2% jet faking  $\tau_h$ .

The input variables capture the key differences between the signal and the background:

- Transverse momentum  $p_T$  of the electron and muon in the  $e\tau_h$  and  $\mu\tau_h$  channels, where the signal tends to have a softer  $p_T$  spectrum (lower energy) than the background.
- $p_T$  of the b-tag jet(s). The signal sample b-tag jet(s) tend to have softer  $p_T$ .
- Invariant masses of the various objects ( $\tau\tau$  legs and the b-tag jet(s)), which tend to be smaller for the signal samples.

- 2275     • The angular separation  $\Delta R$  between pairs of the objects, where signal samples  
 2276       peak at smaller  $\Delta R$  values.
- 2277     • The transverse mass between the missing transverse energy  $p_T^{\text{miss}}$  and each of  
 2278       the four objects [85], defined as

$$m_T(\ell, p_T^{\text{miss}}) \equiv \sqrt{2p_T^\ell \cdot p_T^{\text{miss}}[1 - \cos(\Delta\phi)]} \quad (9.1)$$

2279     where  $p_T^\ell$  is the transverse momentum of the object  $\ell$ , and  $\Delta\phi$  is the difference  
 2280       in azimuthal angle between the object and the  $p_T^{\text{miss}}$ . Events from  $t\bar{t}$ +jets and  
 2281       jets faking  $\tau_h$  backgrounds have larger  $p_T^{\text{miss}}$  resulting in larger transverse mass  
 2282       values compared to the signal, which tends to have smaller  $p_T^{\text{miss}}$  that is also  
 2283       more aligned with the lepton legs.

- 2284     • The variable  $D_\zeta$  [85], defined as

$$D_\zeta \equiv p_\zeta - 0.85p_\zeta^{\text{vis}} \quad (9.2)$$

2285     where the  $\zeta$  axis is the bisector of the transverse directions of the visible  $\tau$  decay  
 2286       products.  $p_\zeta$  is the component of the  $p_T^{\text{miss}}$  along the  $\zeta$  axis, and  $p_\zeta^{\text{vis}}$  is the sum  
 2287       of the components of the lepton  $p_T$  along the same axis. This variable captures  
 2288       the fact that in signal the  $p_T^{\text{miss}}$  is small and approximately aligned with the  $\tau\tau$ .  
 2289       In contrast, the  $Z \rightarrow \tau\tau$  background tends towards large  $D_\zeta$  values because the  
 2290        $p_T^{\text{miss}}$  is collinear to the  $\tau\tau$ , and the  $t\bar{t}$ +jets events tend to have small  $D_\zeta$  due to  
 2291       a large  $p_T^{\text{miss}}$  not aligned with the  $\tau\tau$ .

- 2292     • For events with 2 b-tag jets, one additional variable is defined to capture the  
 2293       difference in the invariant mass of the  $bb$  and the  $\tau\tau$ :

$$\Delta m_{a_1} \equiv (m_{bb} - m_{\tau\tau})/m_{\tau\tau} \quad (9.3)$$

2294        This variable peaks at zero for the  $h \rightarrow aa \rightarrow 2b2\tau$  signal.

2295        The DNN model consists of an input layer, two fully-connected hidden layers,  
2296        and one output layer, which has only one output for this binary classification of  
2297        signal versus background. Two hidden layers were used, as one hidden layer led  
2298        to undertraining, and three hidden layers led to overtraining. One dropout layer  
2299        was inserted after each of the two hidden layers, which set zero weights at nodes  
2300        chosen at a random rate (the dropout rate) during training to reduce overfitting. The  
2301        output node uses a sigmoid activation function to produce a probability-like output  
2302         $0 < y < 1$ , where background samples were assigned a score of 0 and signal samples  
2303        were assigned a score of 1. The training datasets were shuffled and divided into  
2304        training, validation, and test sets, with an equal number of signal and background  
2305        events in each set. Models were trained on the training set, and the performance on  
2306        the training set was compared to the performance on the validation set in order to  
2307        guide the tuning of hyperparameters in the DNN models (e.g. the number of nodes  
2308        in the hidden layers and the dropout rate). The test set was used only to perform an  
2309        unbiased evaluation of the final training.

2310        Events in the data, Monte Carlo, and embedded samples are evaluated with the six  
2311        trained DNNs and assigned a raw score between 0 and 1 (background-like and signal-  
2312        like respectively). In order to flatten the distribution of the score and define score  
2313        thresholds for categorizing events, the raw output scores are transformed with the  
2314        function  $\tilde{p}(n) = \text{arctanh}(p \times \tanh(n))/n$  where  $n$  is a positive integer. The thresholds  
2315        of the DNN score used for signal/control region definition are determined using scans  
2316        that optimize the signal sensitivity and are shown in Tables 9.1 and 9.2.

	1bNN $\tilde{p}(n = 1.5)$			
	SR1	SR2	SR3	CR
$\mu\tau_h$ 2018	$> 0.98$	$\in [0.95, 0.98]$	$\in [0.90, 0.95]$	$< 0.90$
$\mu\tau_h$ 2017	$> 0.97$	$\in [0.94, 0.97]$	$\in [0.90, 0.94]$	$< 0.90$
$\mu\tau_h$ 2016	$> 0.97$	$\in [0.94, 0.97]$	$\in [0.89, 0.94]$	$< 0.89$
	1bNN $\tilde{p}(n = 1.5)$			
	SR1	SR2	SR3	CR
$e\tau_h$ 2018	$> 0.97$	$\in [0.945, 0.97]$	$\in [0.90, 0.945]$	$< 0.90$
$e\tau_h$ 2017	$> 0.985$	$\in [0.965, 0.985]$	$\in [0.93, 0.965]$	$< 0.93$
$e\tau_h$ 2016	$> 0.985$	$\in [0.965, 0.985]$	$\in [0.93, 0.965]$	$< 0.93$
	1bNN $\tilde{p}(n = 2.5)$			
	SR1	SR2	SR3	CR
$e\mu$ 2018	$> 0.99$	$\in [0.95, 0.99]$	$\in [0.85, 0.95]$	$< 0.85$
$e\mu$ 2017	$> 0.985$	$\in [0.95, 0.985]$	$\in [0.85, 0.95]$	$< 0.85$
$e\mu$ 2016	$> 0.99$	$\in [0.95, 0.99]$	$\in [0.85, 0.95]$	$< 0.85$

Table 9.1: Event categorization based on DNN scores for events with exactly 1 b-tag jet (1bNN), for the three  $\tau\tau$  channels and three eras.

	2bNN $\tilde{p}(n = 1.5)$		
	SR1	SR2	CR
$\mu\tau_h$ 2018	$> 0.99$	$\in [0.96, 0.99]$	$< 0.96$
$\mu\tau_h$ 2017	$> 0.98$	$\in [0.94, 0.98]$	$< 0.94$
$\mu\tau_h$ 2016	$> 0.97$	$\in [0.93, 0.97]$	$< 0.93$
	2bNN $\tilde{p}(n = 1.5)$		
	SR1	SR2	CR
$e\tau_h$ 2018	$> 0.96$	NA	$< 0.96$
$e\tau_h$ 2017	$> 0.985$	NA	$< 0.985$
$e\tau_h$ 2016	$> 0.96$	NA	$< 0.96$
	2bNN $\tilde{p}(n = 2.5)$		
	SR1	SR2	CR
$e\mu$ 2018	$> 0.98$	$\in [0.94, 0.98]$	$< 0.94$
$e\mu$ 2017	$> 0.97$	$\in [0.93, 0.97]$	$< 0.93$
$e\mu$ 2016	$> 0.98$	$\in [0.94, 0.98]$	$< 0.94$

Table 9.2: Event categorization based on DNN scores for events with 2 b-tag jets (2bNN), for the three  $\tau\tau$  channels and three eras.

## 2317 9.3 Methodology for signal extraction

2318 After events are divided into categories, the data is compared to the expected back-  
2319 grounds in the signal region categories. Here, we describe the fundamental concepts  
2320 behind hypothesis testing in high-energy physics, as well as how exclusion limits  
2321 can be set on parameters whose true values we cannot measure, culminating in the  
2322 modified frequentist method  $CL_S$  which is used to perform signal extraction in this  
2323 analysis.

### 2324 9.3.1 Model building and parameter estimation

In the frequentist interpretation of probability, an experiment measuring an observable can be repeated, resulting in different values of the observable, e.g. the invariant mass of a candidate Higgs boson in a search for the Higgs [103]. The ensemble of values of the observable  $x$  gives rise to the probability density function (PDF)  $f(x)$ , which has the important property that it is normalized to unity:

$$\int f(x) dx = 1 .$$

A parametric family of PDFs

$$f(x|\alpha) ,$$

2325 read “ $f$  of  $x$  given  $\alpha$ ”, is referred to as a probability model or model. The parameters  $\alpha$   
2326 typically represent parameters of the theory or an unknown property of the detector’s  
2327 response. The parameters are not frequentist in nature, unlike  $x$ . Out of all the  
2328 parameters, typically only a few are of interest, and are called the parameters of  
2329 interest (POI), labeled  $\mu$  here. The remaining are referred to as nuisance parameters  
2330 (NP) [103] and are labeled  $\boldsymbol{\theta}$ .

2331  $f(x)$  is the probability density for the observable in one event and we wish to

2332 describe the probability density for a dataset with many events,  $\mathcal{D} = \{x_1, \dots, x_n\}$ ,  
 2333 called the total probability model  $\mathbf{f}$ . For instance, if we also have a prediction for  
 2334 the total number of events expected, called  $\nu$ , we also account for the overall Poisson  
 2335 probability for observing  $n$  events given  $\nu$  expected:

$$\mathbf{f}(\mathcal{D}|\nu, \alpha) = \text{Poisson}(n|\nu) \prod_{e=1}^n f(x_e|\alpha) \quad (9.4)$$

The likelihood function  $L(\alpha)$  is numerically equivalent to  $f(x|\alpha)$  for fixed  $x$ , or  
 $\mathbf{f}(\mathcal{D}|\alpha)$  with  $\mathcal{D}$  fixed [103]. The likelihood function is not a probability density for  $\alpha$   
 and is not normalized to unity:

$$\int L(\alpha) d(\alpha) \neq 1.$$

2336 i.e. the likelihood function is the value of  $f$  as a function of  $\alpha$  given a fixed value of  
 2337  $x$ .

2338 To estimate the parameter  $\alpha$  we use an estimator, which is a function of the  
 2339 data. Take for example the measurement of data distributed according to a Gaussian  
 2340 probability density  $f(x|\mu, \sigma) = \text{Gauss}(x|\mu, \sigma)$ . One possible estimator of the mean  $\mu$ ,  
 2341 is the mean of the measured data points  $\bar{x} = \sum_{i=1}^n x_i/n$  [103].

2342 A commonly used estimator in physics is the maximum likelihood estimator  
 2343 (MLE), defined as the value  $\hat{\alpha}$  which maximizes the likelihood function  $L(\alpha)$ . This  
 2344 value, labeled  $\hat{\alpha}$ , also maximizes  $\ln L(\alpha)$  and minimizes  $-\ln L(\alpha)$ . By convention the  
 2345  $-\ln L(\alpha)$  is minimized, in a process called “fitting”, and the maximum likelihood  
 2346 estimate is called the “best fit value”.

### 2347 9.3.2 Hypothesis testing

2348 In this section we next introduce concepts related to hypothesis testing such as the  
 2349 test statistic constructed from the ratio of likelihood functions.

2350        The objective of a likelihood analysis is to distinguish different models repre-  
2351        senting the various hypotheses, and determine the one that best explains the ex-  
2352        perimental outcome. In a search for new physics, a signal is additive on top of the  
2353        background. The background-only hypothesis is the null hypothesis, and the signal-  
2354        plus-background hypothesis is the alternative.

2355        As a simple example, take the  $p$ -value test, for an experiment where we count  
2356        events in the signal region,  $n_{SR}$ , and expect  $\nu_B$  background events and  $\nu_S$  events from  
2357        the signal [103]. Then

- 2358        1. The null hypothesis ( $H_0$ ), i.e. the background-only hypothesis in this experi-  
2359        ment, with the probability modeled by  $\text{Poisson}(n_{SR}|\nu_B)$ .
- 2360        2. The alternate hypothesis ( $H_1$ ), i.e. signal-plus-background hypothesis, with the  
2361        probability modeled by  $\text{Poisson}(n_{SR}|(\nu_B + \nu_S))$ .

2362        The compatibility of the observed data  $\nu_{SR}^0$  and the null hypothesis, is quantified as  
2363        the probability that the background-only hypothesis would produce at least as many  
2364        events as was observed. This probability is the  $p$ -value:

$$p = \sum_{n=n_{SR}^0}^{\infty} \text{Poisson}(n|\nu_B). \quad (9.5)$$

2365        If the  $p$ -value is very small, we might reject the null hypothesis. The  $p$ -value is not the  
2366        probability of the null hypothesis given the data; rather, it expresses the probability  
2367        that data with a certain property was obtained, assuming the null hypothesis [103].

2368        The  $p$ -value is an example of a test statistic  $T$ , which maps the data to a single  
2369        real number. The Neyman-Pearson lemma states that out of the infinite possibilities  
2370        of choices of test statistic, the uniformly most powerful test statistic is the likelihood  
2371        ratio  $T_{NP}$  [103]:

$$T_{NP}(\mathcal{D}) = \frac{L(\mathcal{D}|H_1)}{L(\mathcal{D}|H_0)} \quad (9.6)$$

To reiterate, the test statistic  $T$  is a real-valued function of the data, implying that a particular probability model  $\mathbf{f}(\mathcal{D}|\boldsymbol{\alpha})$  implies a distribution of the test statistic,  $f(T|\boldsymbol{\alpha})$ , which depends on the value of  $\boldsymbol{\alpha}$ . With this distribution in hand, the  $p$ -value can be evaluated in the following equivalent formulations:

$$p(\boldsymbol{\alpha}) = \int_{T_0}^{\infty} f(T|\boldsymbol{\alpha}) dT \quad (9.7)$$

$$= \int \mathbf{f}(\mathcal{D}|\boldsymbol{\alpha}) \theta(T(\mathcal{D}) - T_0) d\mathcal{D} \quad (9.8)$$

$$= P(T \geq T_0|\boldsymbol{\alpha}) \quad (9.9)$$

where  $T_0$  is the value of  $T$  based on the observed data, and  $\theta()$  is the Heaviside function. The size of the test is conventionally chosen to be 10%, 5%, or 1%. As the  $p$ -value depends on  $\boldsymbol{\alpha}$  (both the POI and NP), the null hypothesis should not be rejected if the  $p$ -value is larger than the size of the test for any value of the nuisance parameters.

### 9.3.3 Confidence intervals

In an example of the measurement of the Standard Model Higgs boson,  $\boldsymbol{\alpha}_{\text{POI}} = (\sigma/\sigma_{SM}, M_H)$ , with  $\sigma/\sigma_{SM}$  is the ratio of the production cross-section for Higgs with respect to its value in the SM, and  $M_H$  is the unknown mass of the Higgs, values of these parameters outside specific bounds are said to be “excluded at the 95% confidence level”. These allowed regions are called confidence levels or confidence regions, and the parameter values outside of them are considered excluded [103]. A 95% confidence interval does not mean that there is a 95% chance that the true value of the parameter is inside the interval. Rather, a 95% confidence interval covers the

2386 true value 95% of the time (even though we do not know the true value).

2387 To construct a confidence interval for a parameter  $\alpha$ , the Neyman Construction  
2388 is used to invert a series of hypothesis tests; i.e. for each possible value of  $\alpha$ , the null  
2389 hypothesis is treated as  $\alpha$ , and we perform a hypothesis test based on a test statistic.  
2390 To construct a 95% confidence interval, we construct a series of hypothesis tests with  
2391 size of 5%. The confidence interval  $I(\mathcal{D})$  is constructed by taking the set of parameter  
2392 values  $\alpha$  where the null hypothesis is accepted:

$$I(\mathcal{D}) = \{\alpha | P(T(\mathcal{D}) > k_\alpha | \alpha) < \alpha\}, \quad (9.10)$$

2393 where  $T(\mathcal{D})$  is the test statistic, and the last  $\alpha$  (not bolded) and the subscript  $k_\alpha$   
2394 refer to the size of the test. A schematic of the Neyman construction is shown in Fig.  
2395 9.1. In a more generalized case, the  $x$ -axis is the test statistic  $T$ .

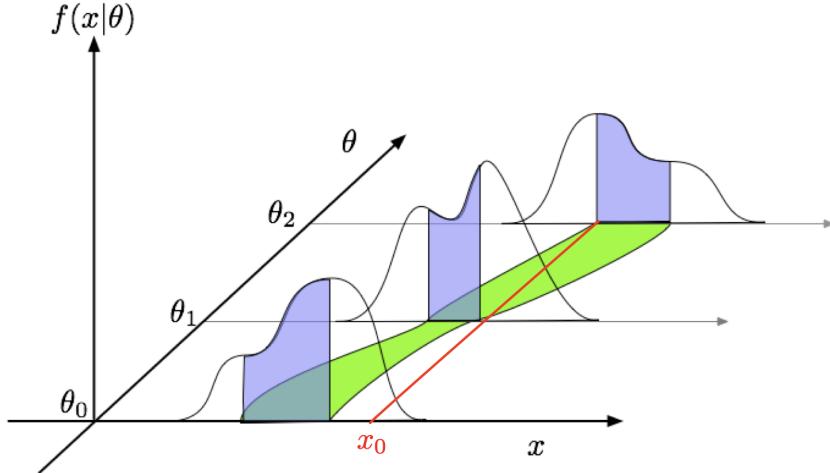


Figure 9.1: Schematic of the Neyman construction for confidence intervals [103]. For each value of  $\theta$ , we find a region in  $x$  where  $\int f(x|\theta)dx$  satisfies the size of the test (blue). These regions form a confidence belt (green). The intersection of the observation  $x_0$  (red) with the confidence belt defines the confidence interval  $[\theta_1, \theta_2]$  [103].

### 2396 9.3.4 Profile likelihood ratio

2397 In this section we describe a frequentist statistical procedure based on the profile  
 2398 likelihood ratio test statistic, which is implemented using asymptotic distributions.

2399 With a multi-parameter likelihood function  $L(\boldsymbol{\alpha})$ , the maximum likelihood of  
 2400 one specific parameter  $\alpha_p$  with other parameters  $\boldsymbol{\alpha}_o$  fixed, is called the conditional  
 2401 maximum likelihood estimate and is denoted  $\hat{\alpha}_p(\boldsymbol{\alpha}_0)$ . The process of choosing specific  
 2402 values of the nuisance parameters for a given value of  $\mu$ ,  $\mathcal{D}_{\text{simulated}}$ , and value of global  
 2403 observables  $\mathcal{G}$  is called profiling. From the full list of parameters  $\boldsymbol{\alpha}$ , we denote the  
 2404 parameter of interest  $\mu$ , and the nuisance parameters  $\boldsymbol{\theta}$ .

2405 We construct the profile likelihood ratio,

$$\lambda(\mu) = \frac{L(\mu, \hat{\boldsymbol{\theta}}(\mu))}{L(\mu, \hat{\boldsymbol{\theta}})} \quad (9.11)$$

2406 which depends explicitly on the parameter of interest  $\mu$ , implicitly on the data  $\mathcal{D}_{\text{sim}}$   
 2407 and global observables  $\mathcal{G}$ , and is independent of the nuisance parameters  $\boldsymbol{\theta}$ , which  
 2408 have been eliminated in profiling [103].

2409 The main conceptual reason for constructing the test statistic from the profile  
 2410 likelihood ratio is that asymptotically (i.e. for measurements with many events) the  
 2411 distribution of the profile likelihood ratio  $\lambda(\mu = \mu_{\text{true}})$  is independent of the values of  
 2412 the nuisance parameters [103].

2413 The following  $p$ -value is used to quantify the consistency with the hypothesis of a  
 2414 signal strength of  $\mu$ :

$$p_\mu = \int_{\tilde{q}_{\mu, \text{obs}}}^{\infty} f(\tilde{q}_\mu | \mu, \hat{\boldsymbol{\theta}}(\mu, \text{obs})) d\tilde{q}_\mu \quad (9.12)$$

### 2415 9.3.5 Modified frequentist method: $CL_S$

2416 In the modified frequentist method called  $CL_S$ , to test a hypothesis with signal, we  
 2417 define  $p'_\mu$  as a ratio of  $p$ -values [103]:

$$2418 p'_\mu = \frac{p_\mu}{1 - p_b} \quad (9.13)$$

2418 where  $p_b$  is the  $p$ -value derived under the background-only hypothesis:

$$2419 p_b = 1 - p_0 \equiv 1 - \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu | 0, \hat{\theta}(\mu = 0, \text{obs})) d\tilde{q}_\mu. \quad (9.14)$$

2419 The  $CL_S$  upper limit on  $\mu$ , denoted  $\mu_{up}$ , is obtained by solving for  $p'_{\mu_{up}} = 5\%$ .  
 2420 If testing the compatibility of the data with the background-only hypothesis, we  
 2421 consider the  $p_b$  value defined above and conventionally convert it into the quantile  
 2422 or “sigma” of a unit Gaussian.  $z$  standard deviations (e.g.  $z = 5$  in “ $5\sigma$ ”) means  
 2423 that the probability of falling above these standard deviations, equals  $p_b$  (e.g.  $3\sigma$   
 2424 corresponds to  $p_b = 2.7 \times 10^{-3}$  or 95.43%, and  $5\sigma$  corresponds to  $p_b = 5.7 \times 10^{-7}$  or  
 2425 99.999943%).

2426 **Chapter 10**

2427 **Results**

2428 In this chapter, Section 10.1 presents the results from the  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis  
2429 performed on  $137 \text{ fb}^{-1}$  of data from the full CMS Run-2 dataset in the years 2016 to  
2430 2018, with interpretations provided for different 2HDM+S scenarios. This analysis  
2431 was combined with a different search in the  $h \rightarrow aa \rightarrow bb\mu\mu$  final state, which was  
2432 also performed on the full Run-2 dataset. The combination procedure and results  
2433 from the combined analyses ( $h \rightarrow aa \rightarrow bb\ell\ell$ , with  $\ell = \mu, \tau$ ) are detailed in 10.2.  
2434 The combined analysis places some of the most stringent limits to date at CMS for  
2435 2HDM+S scenarios in the light scalar mass range  $m_a = 12 \text{ GeV}$  to  $60 \text{ GeV}$ .

2436 **10.1 Results from  $bb\tau\tau$**

2437 In each of the three  $\tau\tau$  channels studied ( $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$ ), events are divided based  
2438 on whether they contain exactly 1 or 2 b-tag jets, and further divided into signal  
2439 and control regions (SRs and CRs) using the DNN categorization score as described  
2440 in Section 9.2. The control regions demonstrate good agreement between observed  
2441 events in data, and the sum of the contributions from expected backgrounds that  
2442 are modeled in simulated and embedded samples. The signal regions are defined to  
2443 be sensitive to the  $h \rightarrow aa \rightarrow bb\tau\tau$  signal. The postfit final observed and expected

2444 distributions of the di-tau invariant mass  $m_{\tau\tau}$  reconstructed with SVFit (described  
2445 in Section 5.2) are shown in Fig. 10.1 for the  $\mu\tau_h$  channel, Fig. 10.2 for the  $e\tau_h$   
2446 channel, and Fig. 10.3 for the  $e\mu$  channel. In all figures, the hypothesized yield for  
2447 the  $h \rightarrow aa \rightarrow bb\tau\tau$  signal is shown for the pseudoscalar mass  $m_a = 35$  GeV and  
2448 assuming a branching fraction  $B(H \rightarrow aa \rightarrow bb\tau\tau) = 10\%$ .

2449 The 95% CL expected and observed exclusion limits on the signal strength of the  
2450 branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau)$  as a function of the pseudoscalar mass  $m_a$   
2451 ranging from 12 GeV to 60 GeV, are shown for the three  $\tau\tau$  channels and all three  
2452 channels combined in Fig. 10.4. The limits are shown as percentages and normalized  
2453 to the production cross-section of the Standard Model Higgs boson. No excess of  
2454 events above the Standard Model expectations is observed. In the limits for the three  
2455  $\tau\tau$  channels combined, expected (observed) limits range from 1.4 to 5.6% (1.7 to  
2456 7.6%) for pseudoscalar masses between 12 and 60 GeV.

2457 The  $e\mu$  channel is the only channel that has signal sensitivity to the  $m_a = 12$   
2458 GeV pseudoscalar mass hypothesis, because the minimum required spatial separation  
2459  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  between the two  $\tau$  legs is smaller than the other two channels  
2460 ( $\Delta R < 0.3$  for  $e\mu$ , compared to  $\Delta R < 0.4$  for the other two channels). This decreased  
2461  $\Delta R$  requirement results in better signal acceptance for low mass signals for the  $e\mu$   
2462 channel. The  $\mu\tau_h$  and  $e\tau_h$  channels are most sensitive to the intermediate mass points  
2463 studied, since the analysis targets a resolved signature: at low mass points, the tau  
2464 legs are boosted, and at high mass points, the  $m_{\tau\tau}$  distributions in signal have larger  
2465 overlap with background distributions. In the combination of the three  $\tau\tau$  channels,  
2466 the limit for  $m_a = 12$  GeV comes only from the  $e\mu$  channel, and the best sensitivity  
2467 is attained at intermediate mass points around  $m_a = 20$  GeV to 45 GeV.

2468 To set limits on the branching fraction of the 125 GeV Higgs to the two pseu-  
2469 doscalars,  $B(h \rightarrow aa)$ , we interpret the results in four types of 2HDM+S, which were  
2470 introduced in Section 1.4. In 2HDM+S, the theorized branching fraction of the pseu-

2471 doscalars depends on the 2HDM+S model type, the pseudoscalar mass  $m_a$ , and the  
2472 ratio of the two Higgs doublets' vacuum expectation values  $\tan \beta$ . In Type I models,  
2473 the branching fraction is independent of  $\tan \beta$ , while in Types II, III, and IV, it is  
2474 a function of  $m_a$  and  $\tan \beta$ . Limits for the  $bb\tau\tau$  final state as a function of  $m_a$  for  
2475 2HDM+S Type I (valid for all  $\tan \beta$  values), Type II with  $\tan \beta = 2.0$ , Type III with  
2476  $\tan \beta = 2.0$ , and Type IV with  $\tan \beta = 0.6$  are overlaid and shown in Fig. 10.5a.

## 2477 10.2 Combination with $bb\mu\mu$ final state

2478 Results from this analysis for the  $h \rightarrow aa \rightarrow bb\tau\tau$  final state are combined with the  
2479 analysis for the  $h \rightarrow aa \rightarrow bb\mu\mu$  final state [104]. While the predicted branching ratio  
2480 for  $aa \rightarrow bb\mu\mu$  is comparatively small, the  $bb\mu\mu$  final state has competitive results  
2481 due to the excellent di-muon resolution measured by CMS. The  $bb\mu\mu$  analysis uses  
2482 an unbinned fit to the data using the di-muon mass  $m_{\mu\mu}$  distribution. Details can be  
2483 found in [104].

2484 Combining the results is possible since the  $bb\tau\tau$  analysis explicitly rejects events  
2485 with extra leptons, so there is no overlap between the events studied in the  $bb\tau\tau$   
2486 analysis and the  $bb\mu\mu$  analysis. In the statistical combination, several systematic  
2487 uncertainties are treated as correlated: the integrated luminosity normalization, the  
2488 b-tagging scale factor, the scale factors related to muon reconstruction, identifica-  
2489 tion, and trigger efficiencies, the inefficiency in the ECAL trigger readout, and the  
2490 theoretical uncertainties related to signal modeling.

2491 Since the results in both final states are statistically limited, the combination ben-  
2492 efits from the additional data. For  $m_a = 35$  GeV, all systematic uncertainties amount  
2493 to around 6% of the total uncertainty, with the dominant systematic uncertainties  
2494 coming from jet energy systematics in the  $bb\mu\mu$  final state, theoretical uncertainties  
2495 in the signal, and uncertainties in the QCD multijet backgrounds in the  $e\mu$  channel

2496 of the  $bb\tau\tau$  final state.

2497 The mass distributions of the di-muon and di-tau objects ( $m_{\mu\mu}$  and  $m_{\tau\tau}$ ) are  
2498 compared to the data in a combined maximum likelihood fit to derive upper limits  
2499 on  $B(h \rightarrow aa)$ . The observed limits at 95% CL on  $B(h \rightarrow aa)$  for different 2HDM+S  
2500 scenarios, are shown for the search for  $h \rightarrow aa \rightarrow bb\mu\mu$  in Fig. 10.5b, and the  
2501 combined analyses  $h \rightarrow aa \rightarrow bb\ell\ell$  in Fig. 10.6.

2502 Exclusion limits in a two-dimensional plane as a function of  $\tan\beta$  and  $m_a$  are  
2503 set for 2HDM+S Types II, III, and IV in Fig. 10.7. The most stringent constraints  
2504 are observed for 2HDM+S type III because of large branching fractions predicted in  
2505 theory, with predicted branching fractions between 0.47 and 0.42 for  $\tan\beta = 2.0$  and  
2506 values of  $m_a$  between 15 and 60 GeV, compared to the observed 95% CL upper limits  
2507 which are between 0.08 and 0.03. For 2HDM+S type IV, the predicted branching  
2508 fractions from theory are between 0.26 and 0.20 for  $\tan\beta = 0.6$  for values of  $m_a$   
2509 between 15 and 60 GeV, and the 95% CL observed upper limits are between 0.12 and  
2510 0.05.

2511 The combined results from  $h \rightarrow aa \rightarrow bb\ell\ell$  are compared with CMS results in  
2512 other final states as a function of the pseudoscalar mass  $m_a$ : for 2HDM+S type I in  
2513 Fig. 10.8, type II with  $\tan\beta = 2.0$  in Fig. 10.9, and type III with  $\tan\beta = 2.0$  in Fig.  
2514 10.10. In other scenarios, e.g. type III with  $\tan\beta = 5.0$ , more stringent limits are set  
2515 by analyses in other final states,  $\mu\mu\tau\tau$  in this case. Other summary plots for other  
2516 model types and  $\tan\beta$  values can be found at [105].

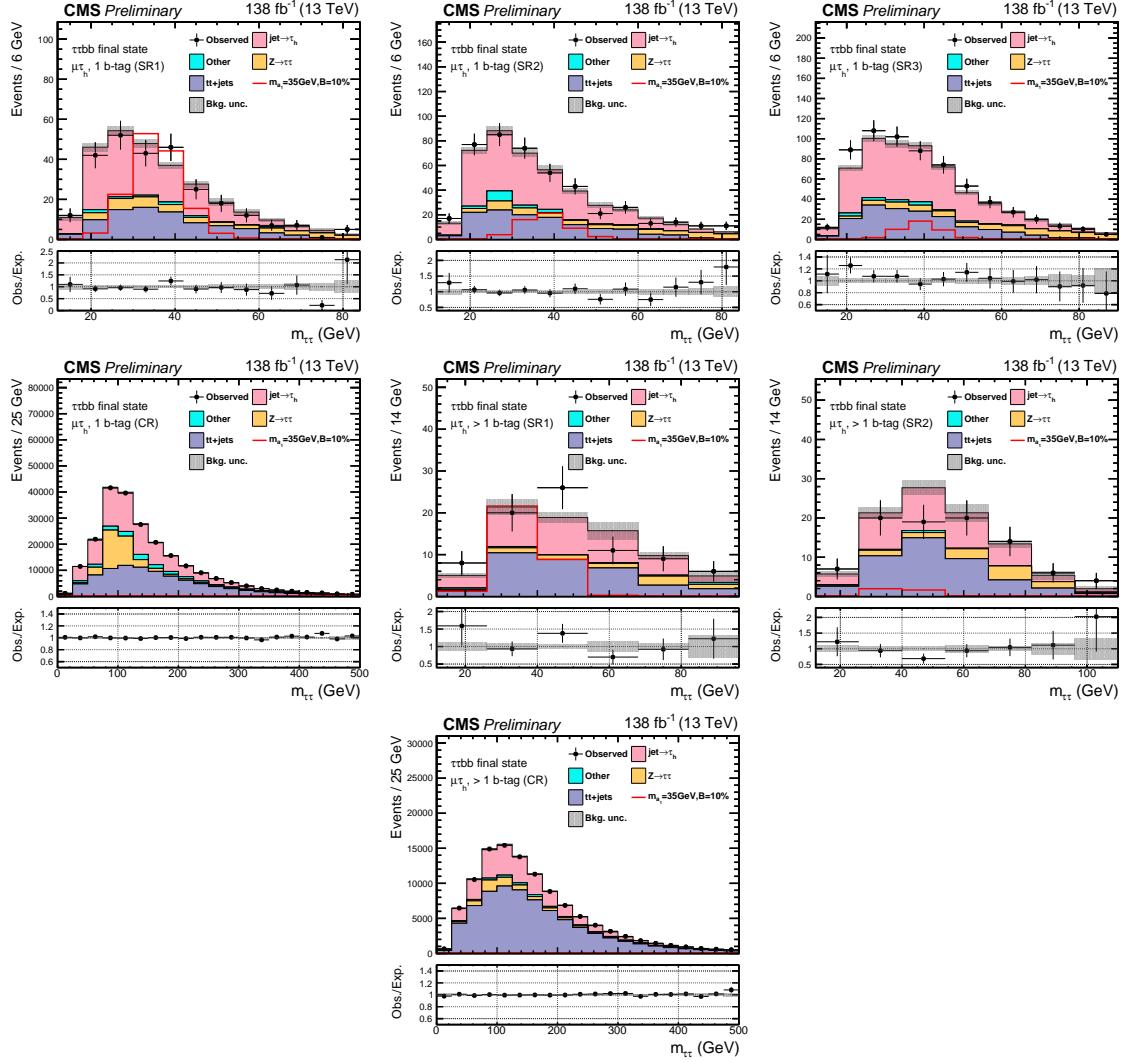


Figure 10.1: Postfit final  $m_{\tau\tau}$  observed and expected distributions, and the observed/expected ratios, in the  $\mu\tau_h$  channel [102]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), 1 b-tag jet control region (*middle row*), 2 b-tag jet signal regions (SR1, SR2) (*middle row*), and lastly the 2 b-tag jet control region (CR) (*bottom*). Statistical and systematic sources of uncertainties in the expected events are added in quadrature and labeled “Bkg. unc” (*shaded gray*). The dominant backgrounds in all categories are jets faking the  $\tau_h$  leg (*pink*),  $Z \rightarrow \tau\tau$  (*orange*), and  $t\bar{t}+j$ ets (*purple*). For illustrative purposes, the beyond-Standard Model signal yield from  $h \rightarrow aabb\tau\tau$  is shown for the pseudoscalar mass hypothesis  $m_a = 35$  GeV, assuming a branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau) = 10\%$  (*red line*).

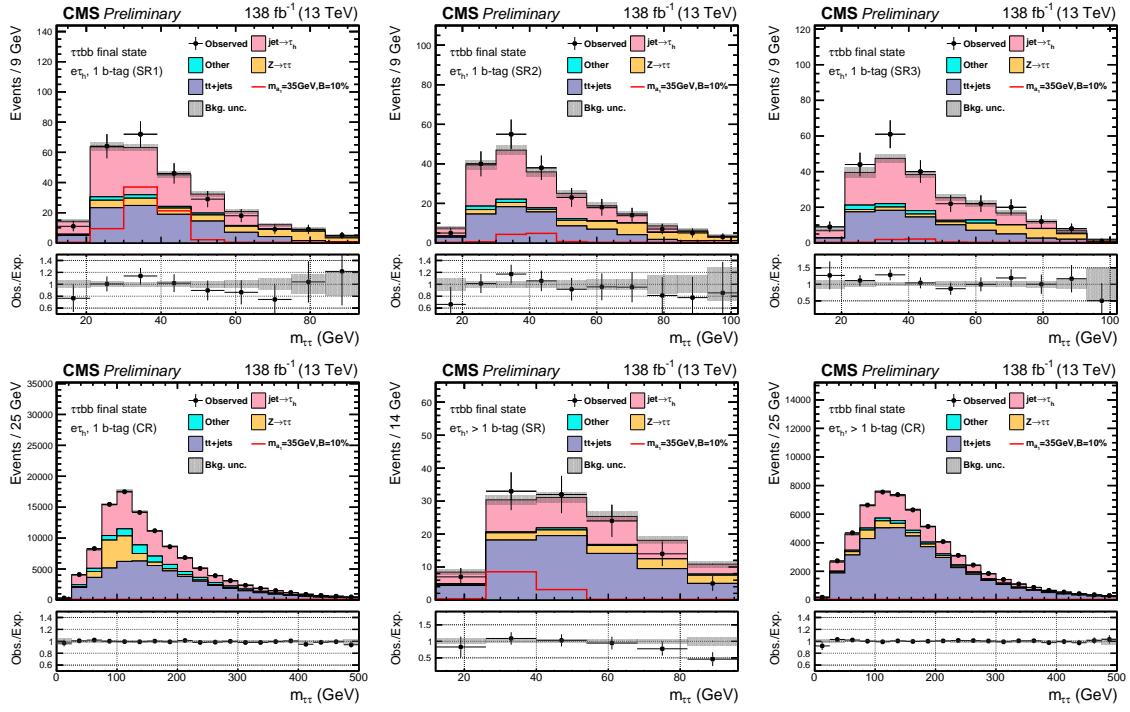


Figure 10.2: Postfit final observed and expected  $m_{\tau\tau}$  distributions, and the observed/expected ratios, in the  $e\tau_h$  channel [102]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), the 1 b-tag jet control region (CR) (*bottom row*), and 2 b-tag jet signal region (SR) and control region (CR) (*bottom row*). Statistical and systematic sources of uncertainties in the expected events are added in quadrature and labeled “Bkg. unc” (*shaded gray*). In this channel, the dominant backgrounds are jets faking the  $\tau_h$  leg (*pink*),  $Z \rightarrow \tau\tau$  (*orange*), and  $t\bar{t}+{\rm jets}$  (*purple*). For illustrative purposes, the beyond-Standard Model signal yield from  $h \rightarrow aabb\tau\tau$  is shown for the pseudoscalar mass hypothesis  $m_a = 35$  GeV, assuming a branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau) = 10\%$  (*red line*).

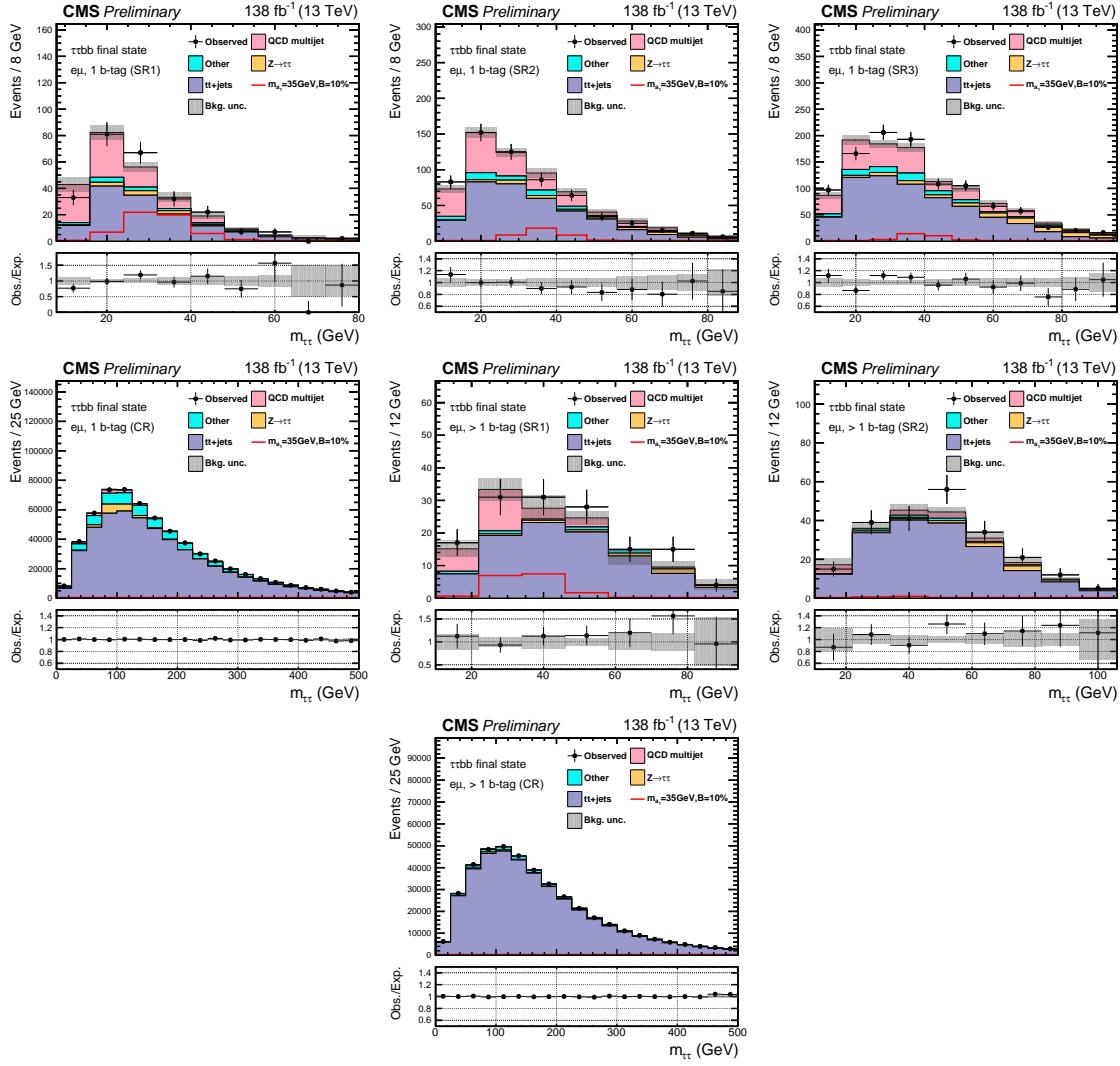


Figure 10.3: Postfit final observed and expected  $m_{\tau\tau}$  distributions, and the observed/expected ratios, in the  $e\mu$  channel [102]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, and SR3) (*top row*), 1 b-tag jet control region (CR) (*middle row*), 2 b-tag jet signal regions (SR1 and SR2) (*middle row*), and 2 b-tag jet control region (CR) (*bottom row*). Statistical and systematic sources of uncertainties in the expected events are added in quadrature and labeled “Bkg. unc” (*shaded gray*). The  $t\bar{t}+j$  process (*purple*) is a major background, and in the signal regions the QCD multijet (*pink*) is also a major background. For illustrative purposes, the pseudoscalar mass hypothesis  $m_a = 35$  GeV, assuming a branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau) = 10\%$  (*red line*).

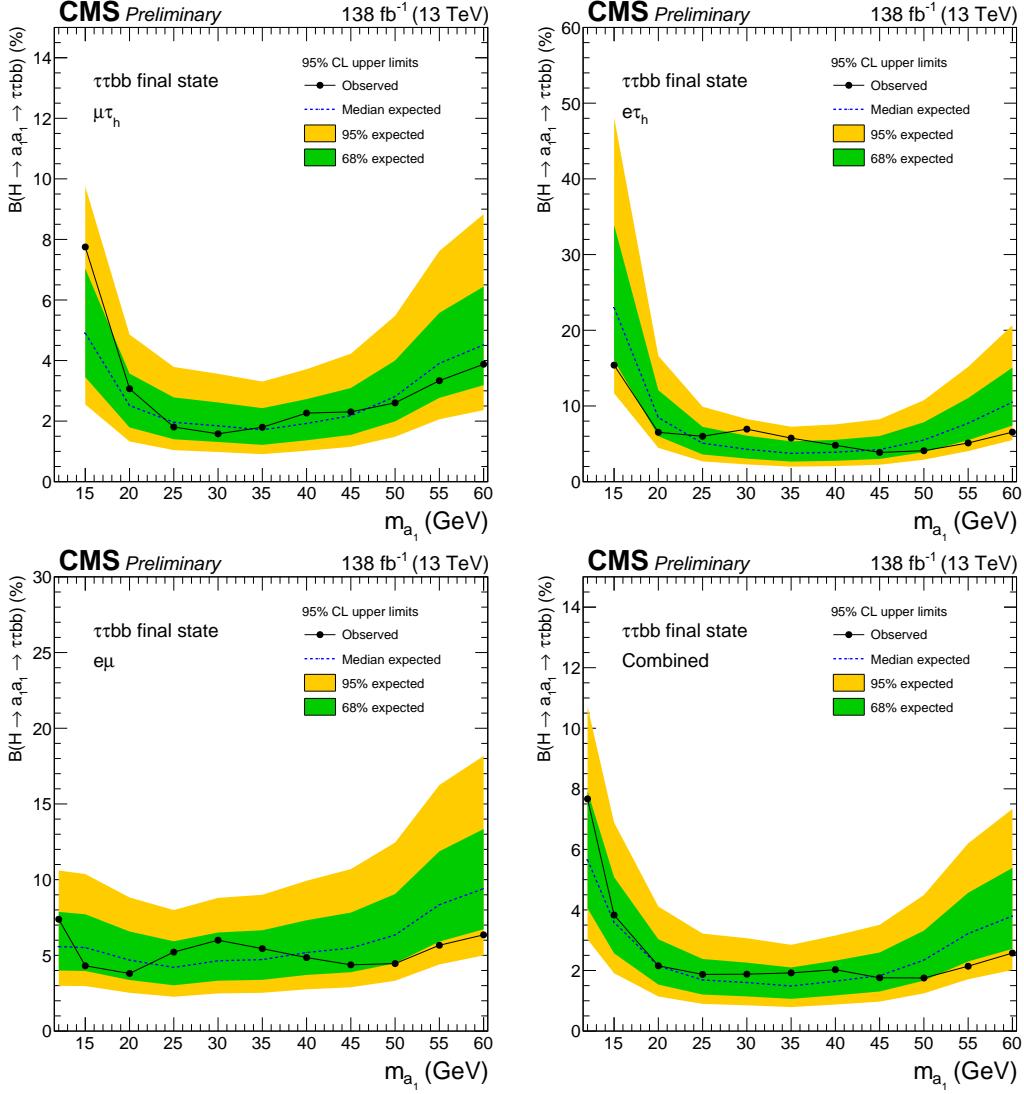
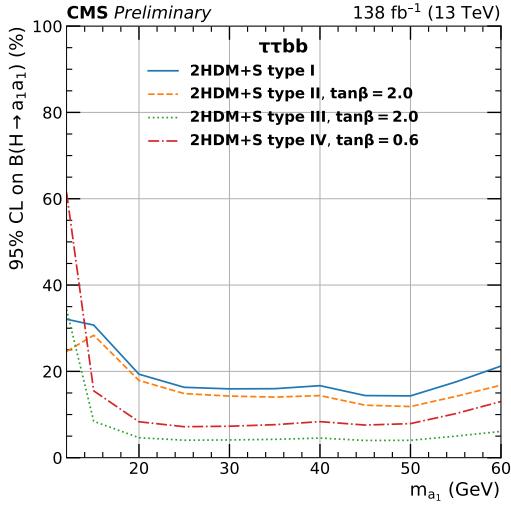
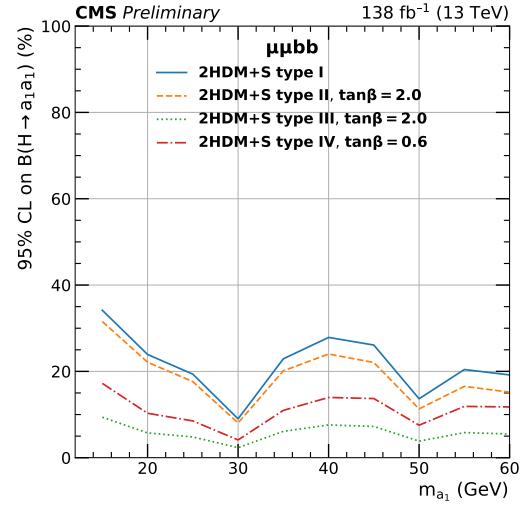


Figure 10.4: Observed 95% CL exclusion limits (*black, solid lines*) and expected 95% CL and 68% CL limits (*shaded yellow and green*) on the branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau)$  in percentages, assuming the Standard Model production for the 125 GeV Higgs ( $h$ ). Limits are shown for the  $\mu\tau_h$  channel (*top left*), the  $e\tau_h$  channel (*top right*), and the  $e\mu$  channel (*bottom left*), and lastly the combination of all three channels (*bottom right*) [102]. The dataset corresponds to  $138 \text{ fb}^{-1}$  of data collected in the years 2016-2018 at a center-of-mass energy 13 TeV. Only the  $e\mu$  channel has sensitivity to the mass hypothesis  $m_a = 12 \text{ GeV}$ . The best sensitivity is attained at intermediate mass points.



(a)  $bb\tau\tau$  final state.



(b)  $bb\mu\mu$  final state.

Figure 10.5: Observed 95% CL upper limits on  $B(h \rightarrow aa)$  in %, for the  $bb\tau\tau$  final state (*left*) and  $bb\mu\mu$  final state (*right*) using the full Run 2 integrated luminosity of  $138 \text{ fb}^{-1}$  in 2HDM+S type I (blue), type II with  $\tan\beta = 2.0$  (orange dashed), type III with  $\tan\beta = 2.0$  (dotted green), and type IV with  $\tan\beta = 0.6$  (red dashed) [102]. Linear interpolation is used between points in the graphs. The  $\tan\beta$  values chosen here correspond to the most stringent limits in each model.

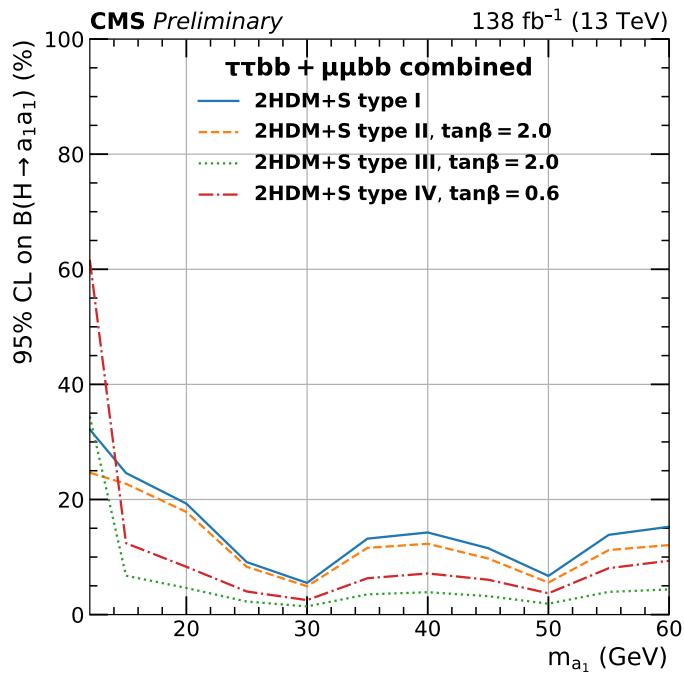


Figure 10.6: Observed 95% CL upper limits on the branching fraction of the 125 GeV Higgs boson to two pseudoscalars,  $B(h \rightarrow aa)$ , in percentages, as a function of the pseudoscalar mass  $m_a$ , in 2HDM+S type I (blue), type II with  $\tan\beta = 2.0$  (orange dashed), type III with  $\tan\beta = 2.0$  (dotted green), and type IV with  $\tan\beta = 0.6$  (red dashed), for the combination of  $bb\mu\mu$  and  $bb\tau\tau$  channels using the full Run 2 integrated luminosity of  $138 \text{ fb}^{-1}$  [102].

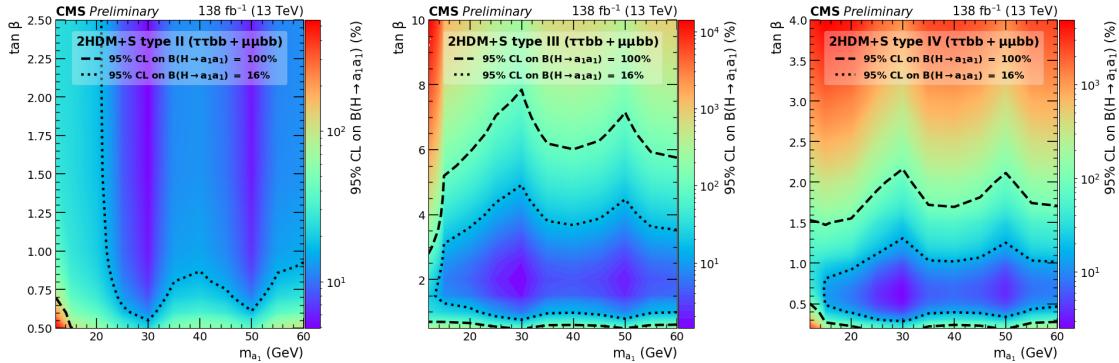


Figure 10.7: Observed 95% CL upper limits on  $\mathcal{B}(h \rightarrow aa)$  in %, for the combination of  $bb\mu\mu$  and  $bb\tau\tau$  channels using the full Run 2 integrated luminosity of  $138 \text{ fb}^{-1}$  for Type II (*left*), Type III (*middle*), and Type IV (*right*) 2HDM+S in the  $\tan \beta$  vs.  $m_a$  phase space. The contours (*dashed black*) correspond to branching fractions of 100% and 16%, where 16% is the combined upper limit on Higgs boson to undetected particle decays from previous Run-2 results. All points inside the contour are allowed within that upper limit. Linear extrapolation has been used between different points on the figures [102].

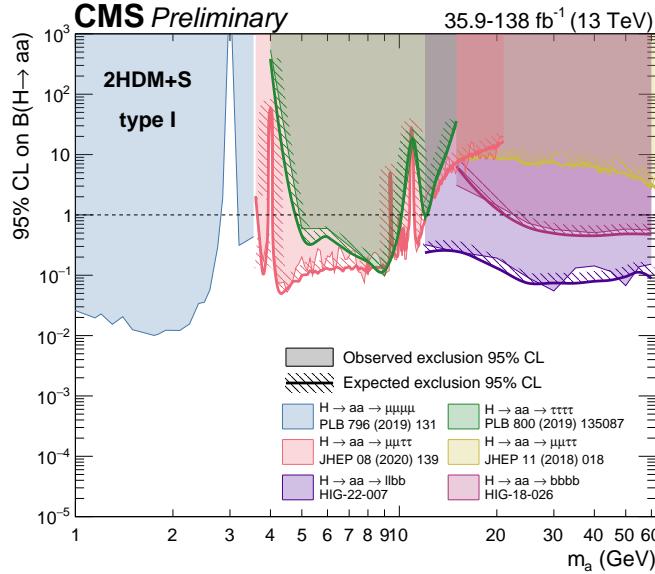


Figure 10.8: Summary plot of current 95% limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section,  $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$  in the 2HDM+S type I scenario performed with data collected at 13 TeV [105]. Results from different final states studied at CMS are overlaid on this figure:  $\mu\mu\mu\mu$  (blue),  $\tau\tau\tau\tau$  (green), boosted  $2\mu 2\tau$  (red), resolved  $2\mu 2\tau$  (yellow),  $bbbb$  (magenta), and the combined result for  $\ell\ell bb$  ( $\ell = \mu, \tau$ ) (purple).

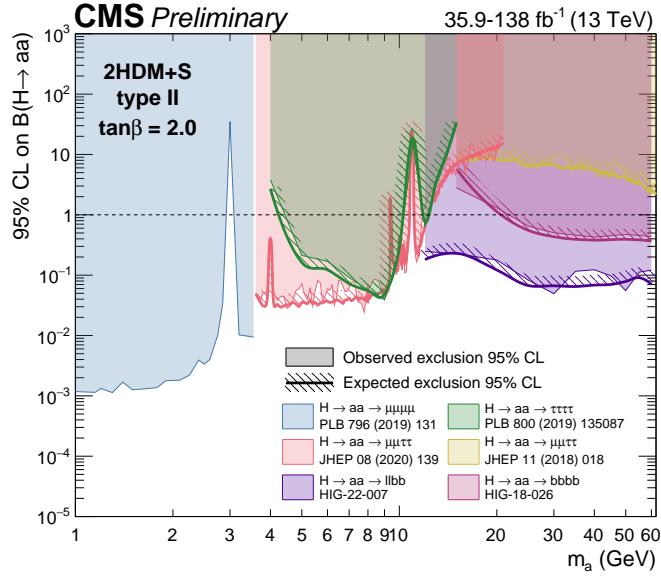


Figure 10.9: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section,  $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$ , in the 2HDM+S type II scenario with  $\tan \beta = 2.0$ , obtained at CMS with data collected at 13 TeV [105]. Results from different final states studied at CMS are overlaid on this figure:  $\mu\mu\mu\mu$  (blue),  $\tau\tau\tau\tau$  (green), boosted  $2\mu 2\tau$  (red), resolved  $2\mu 2\tau$  (yellow),  $bbbb$  (magenta), and the combined result for  $\ell\ell bb$  ( $\ell = \mu, \tau$ ) (purple).

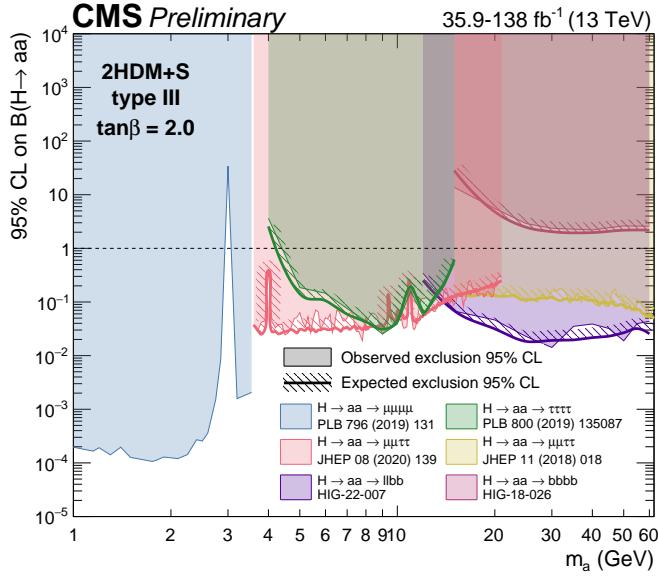


Figure 10.10: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross section,  $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$  in the 2HDM+S type-III scenario with  $\tan \beta = 2.0$ , obtained at CMS with data collected at 13 TeV [105]. Results from different final states studied at CMS are overlaid on this figure:  $\mu\mu\mu\mu$  (blue),  $\tau\tau\tau\tau$  (green), boosted  $2\mu 2\tau$  (red), resolved  $2\mu 2\tau$  (yellow),  $bbbb$  (magenta), and the combined result for  $\ell\ell bb$  ( $\ell = \mu, \tau$ ) (purple).

2517 **Chapter 11**

2518 **Asymmetric exotic Higgs decays**

2519 This chapter presents progress towards a search for exotic Higgs decays to two light  
2520 scalars with unequal mass ( $h \rightarrow a_1 a_2$ ) to final states with bottom quarks and  $\tau$  leptons,  
2521 which has interpretations in Two Real Singlet Models (TRSMs) described in  
2522 Section 1.5. Compared to the symmetric decay scenario  $h \rightarrow aa$ , which has been studied  
2523 in multiple final states at CMS with stringent limits set on the various 2HDM+S  
2524 scenarios, this asymmetric decay scenario has not been directly searched for at the  
2525 CMS experiment. Section 11.1 lists the mass hypotheses of the new particles  $a_1$  and  
2526  $a_2$  considered in this search. Section 11.2 describes the studies performed on the simulated  
2527 signal samples to determine which channels are viable for the analysis. Section  
2528 11.3 shows the control plots produced using the framework for this analysis.

2529 **11.1 Signal masses**

2530 As discussed in Section 1.5,  $h \rightarrow a_1 a_2$  can result in a “cascade” decay if one of the  
2531 scalars,  $a_2$  is sufficiently heavy ( $m_{a_2} > 2m_{a_1}$ ). The “non-cascade” case is where the  
2532 light scalars decay directly to Standard Model particles.

2533 The mass hypotheses (mass points) ( $m_{a_1}, m_{a_2}$ ) studied here are:

- 2534 • *Cascade mass points:* (15, 30), (15, 40), (15, 50), (15, 60), (15, 70), (15, 80),  
2535 (15, 90), (15, 100), (15, 110), (20, 40), (20, 50), (20, 60), (20, 70), (20, 80), (20,  
2536 90), (20, 100), (30, 60), (30, 70), (30, 80), and (30, 90) GeV

- 2537 • *Non-cascade mass points:* (15, 20), (15, 30), (20, 30), (20, 40), (30, 40), (30,  
2538 50), (30, 60), (40, 50), (40, 60), (40, 70), (40, 80), (50, 60), and (50, 70) GeV

2539 Samples were produced using the MadGraph5\_aMCatNLO event generator, for each  
2540 signal mass point in the gluon-gluon fusion (ggF) and vector boson fusion (VBF)  
2541 production modes of the 125 GeV Higgs boson. In the sample generation, the decays  
2542 of  $a$  to Standard Model particles were specified to be decays to bottom quarks or  $\tau$   
2543 leptons.

## 2544 11.2 Cascade scenario signal studies

2545 Generator-level studies of the  $h \rightarrow a_1 a_2$  cascade decay were performed to determine  
2546 the viability of the  $4b2\tau$  and/or  $2b4\tau$  channels.

2547 Cross sections and branching fractions of the  $4b2\tau$  and  $2b4\tau$  final states were  
2548 compared using cross-section predictions provided by the authors of [7]. For an  
2549 example mass point  $m_{a_2} = 80$  GeV,  $m_{a_1} = 30$  GeV, the branching fractions to  
2550  $4b2\tau$  is ten times larger than  $2b4\tau$ :  $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 4b2\tau) = 0.00857$ , vs.  
2551  $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 2b4\tau) = 0.00068$ . The  $4b2\tau$  final state is chosen for this  
2552 analysis.

2553 In general the four b-flavor jets have low  $p_T$  at generator level, as illustrated for  
2554 example mass points (100, 15) GeV and (40, 20) GeV in Fig. 11.1. The  $p_T$  distribution  
2555 of the sub-leading jet peaks at an energy below 20 GeV, with the third and fourth  
2556 jets tending to have even softer energies.

2557 An event category with three or more b-tag jets was determined to be infeasible  
2558 due to low statistics in this category, due to the difficulties in reconstructing the third

2559 and fourth b-flavor jets which have very low transverse momenta  $p_T$ . Event categories  
 2560 with exactly 1 b-tag jet and  $\geq 2$  b-tag jets are used.

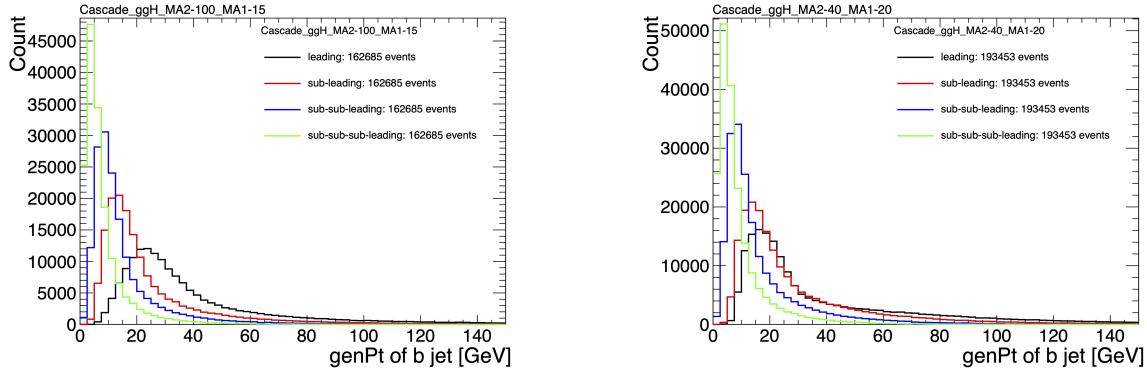


Figure 11.1: Generator-level b-flavor jet transverse momenta  $p_T$ , for  $h \rightarrow a_1 a_2$  cascade scenario in the  $4b2\tau$  final state, for mass hypotheses  $(m_{a_1}, m_{a_2}) = (100, 15)$  GeV (left) and  $(40, 20)$  GeV (right). In each plot the generator-level  $p_T$  of the leading (black), sub-leading (red), third (blue), and fourth (light green) are overlaid.

2561 In the  $4b2\tau$  final state, the possibility of the leading and sub-leading b-tag jets  
 2562 being sufficiently close in  $\Delta R$  to require boosted jet reconstruction techniques was  
 2563 explored. In the  $4b2\tau$  case, the two b-flavor-jets in the generated event that were  
 2564 spatially closest in  $\Delta R$  were considered as one object. This two b-flavor jet object was  
 2565 spatially matched in  $\Delta R$  to the jets reconstructed with the standard AK4 algorithm  
 2566 which uses a cone size of  $\Delta R = 0.4$ . The quality of the  $p_T$  resolution (computed as  
 2567  $(p_{T,\text{reconstructed}} - p_{T,\text{gen}})/p_{T,\text{gen}}$ ) and closeness in distance  $\Delta R$  of the reconstructed jet  
 2568 to the nearest generator-level jets, was seen to depend on the absolute and relative  
 2569 masses of the light scalars. The best (worst) performance occurred in samples with  
 2570 large (small) mass differences between the heavier scalar  $a_2$  and the lighter scalar  $a_1$ ,  
 2571 as illustrated for the mass hypotheses  $(m_{a_1}, m_{a_2})$   $(100, 15)$  GeV and  $(40, 20)$  GeV in  
 2572 Fig. 11.2.

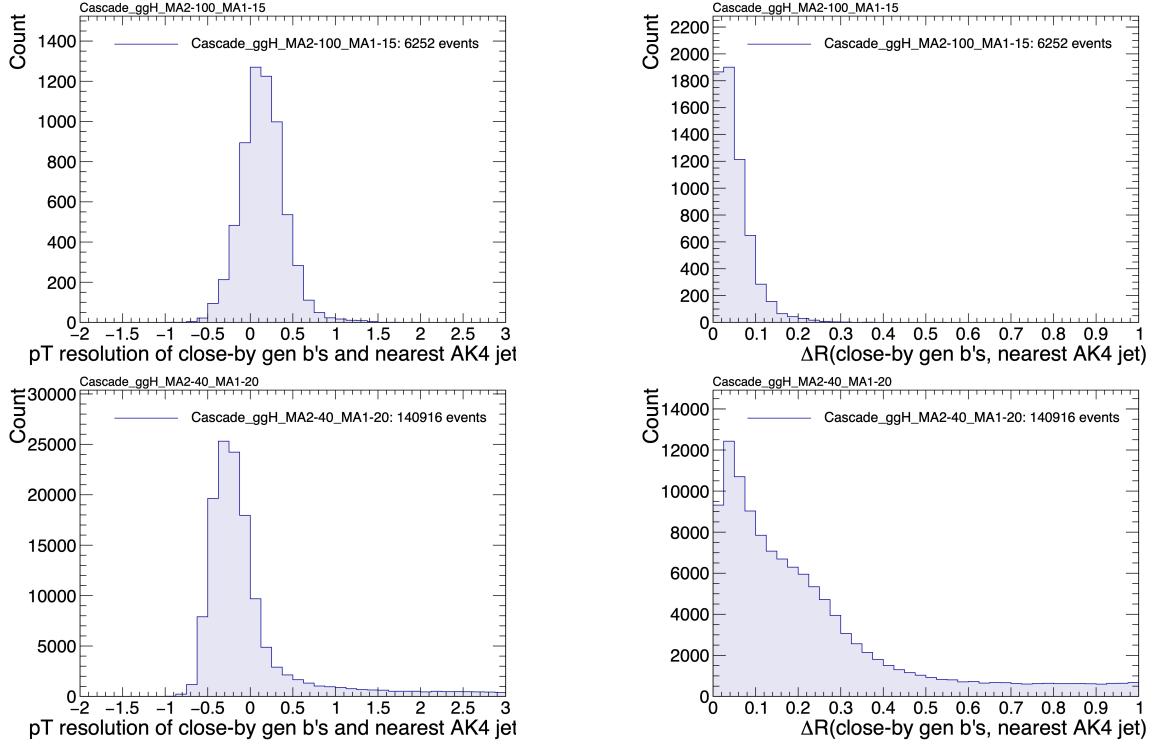


Figure 11.2: Distributions (arbitrary units) of transverse momentum  $p_T$  resolution and  $\Delta R$  between the two closest generator-level  $b$  jets, treated as one object, and the nearest reconstructed AK4 jet, for two different  $h \rightarrow a_1 a_2$  mass hypotheses ( $m_{a_1}, m_{a_2}$ ) = (100, 15) GeV (top left, top right) and (40, 20) GeV (bottom left, bottom right) in the ggH production of the 125 GeV  $h$ . In the (40, 20) GeV mass point, the longer  $p_T$  resolution tail (bottom left) indicates that the reconstructed jet underestimates the generator  $b$ -flavor jets' energy, and the significant fraction of events with larger  $\Delta R$  values (bottom right) indicate worse matching.

### 11.3 Control plots for $\mu\tau_h$ channel

2573     The  $\tau\tau$  states for the  $h \rightarrow a_1 a_2$  to  $4b2\tau$  analysis are similar to those in the  $h \rightarrow aa \rightarrow$   
2574      $bb\tau\tau$  analysis. For the  $\mu\tau_h$  channel, histograms of the key kinematic variables are  
2575     made for data and the sum of the expected backgrounds, which are estimated from  
2576     Monte Carlo samples, embedded samples, and the data-driven method for estimating  
2577     jets faking  $\tau_h$  as described in Chapter 7. Nominal values of the scale factors and event  
2578     reweighting are applied, as described in Chapter 5.3. The errors shown in the figures  
2579     include statistical errors and several of the full set of systematic errors (only those  
2580     associated with the lepton energy scales and  $\tau_h$  identification efficiency, described in  
2581     associated with the lepton energy scales and  $\tau_h$  identification efficiency, described in

2582 Sections 5.3.1, 5.3.2, and 5.3.4).

2583 The  $p_T$ ,  $\eta$ , and  $\phi$  of the leading muon and hadronic tau  $\tau_h$ , and the di-tau visible  
2584 mass  $m_{\text{vis}}$  and momentum  $p_{T,\text{vis}}$ , are shown in Figures 11.3, 11.4, and 11.5. The  $p_T$ ,  
2585  $\eta$ , and  $\phi$  of the the leading and sub-leading b-tag jets, and the missing transverse  
2586 energy magnitude and azimuthal direction, are shown in Figures 11.6, 11.7, and 11.8.

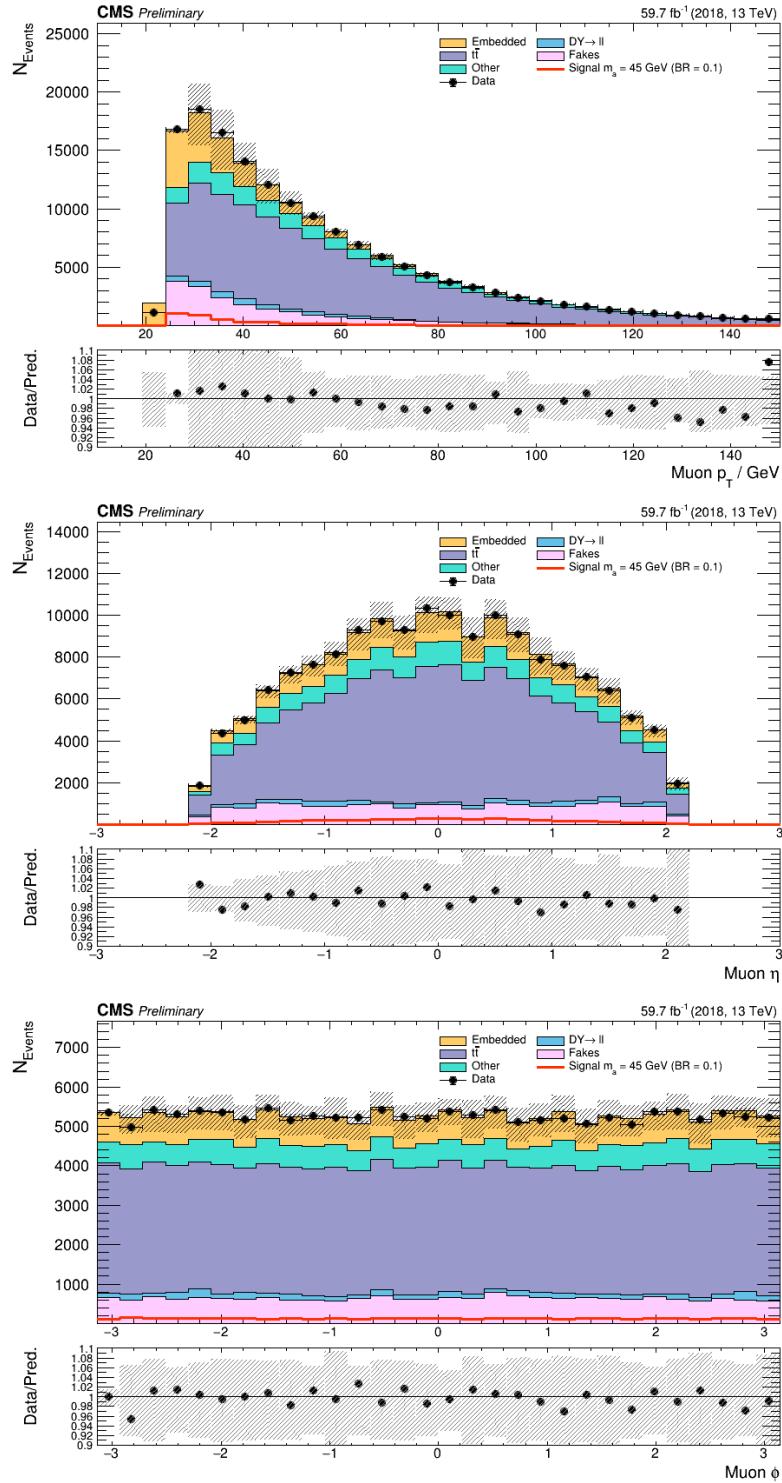


Figure 11.3: Kinematic properties of the leading muon in the  $\mu\tau_h$  channel using 2018 samples: transverse momentum  $p_T$  (top),  $\eta$  (middle), and  $\phi$  (bottom). The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and  $\tau_h$  identification efficiency).

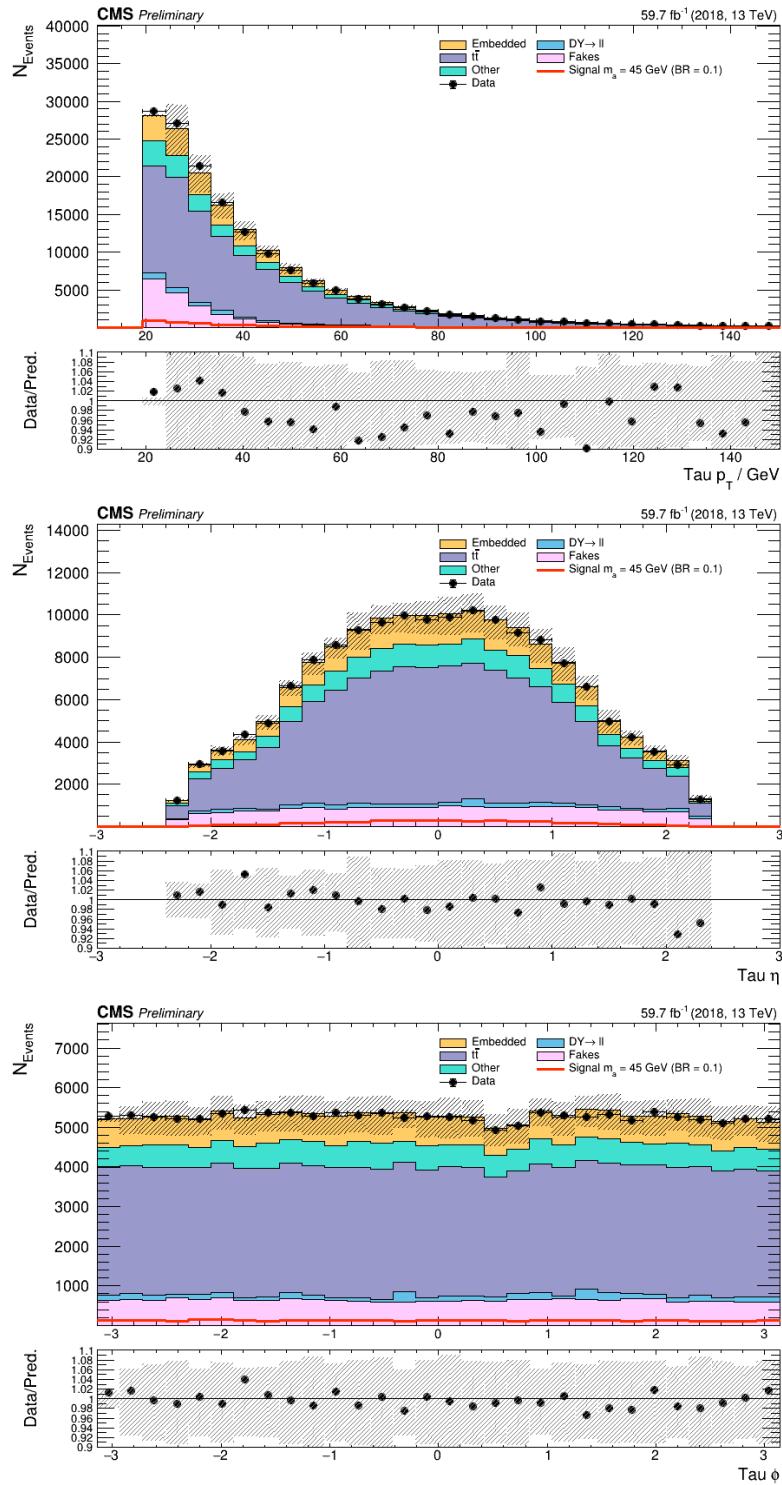


Figure 11.4: Kinematic properties of the leading hadronic tau ( $\tau_h$ ) in the  $\mu\tau_h$  channel using 2018 samples: transverse momentum  $p_T$  (top),  $\eta$  (middle), and  $\phi$  (bottom). The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and  $\tau_h$  identification efficiency).

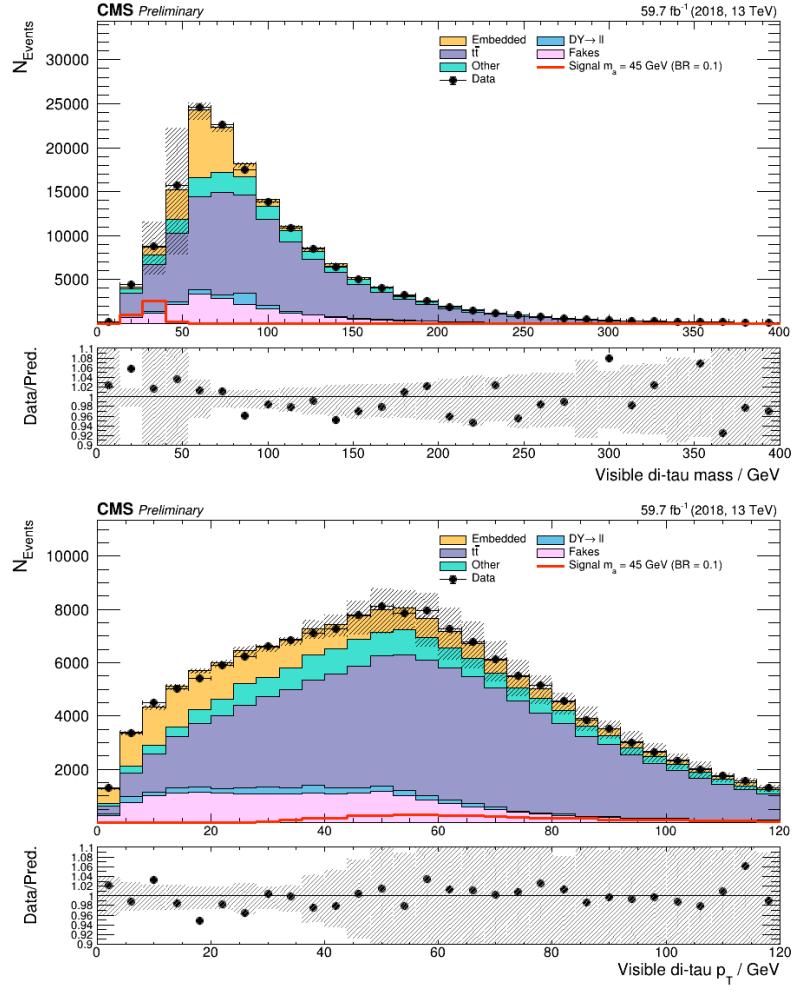


Figure 11.5: The visible di-tau mass  $m_{\text{vis}}$  (*top*) and visible di-tau transverse momentum  $p_{T,\text{vis}}$  (*bottom*), computed from the sum of the visible 4-momenta of the muon and  $\tau_h$  in the  $\mu\tau_h$  channel using 2018 samples. The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and  $\tau_h$  identification efficiency).

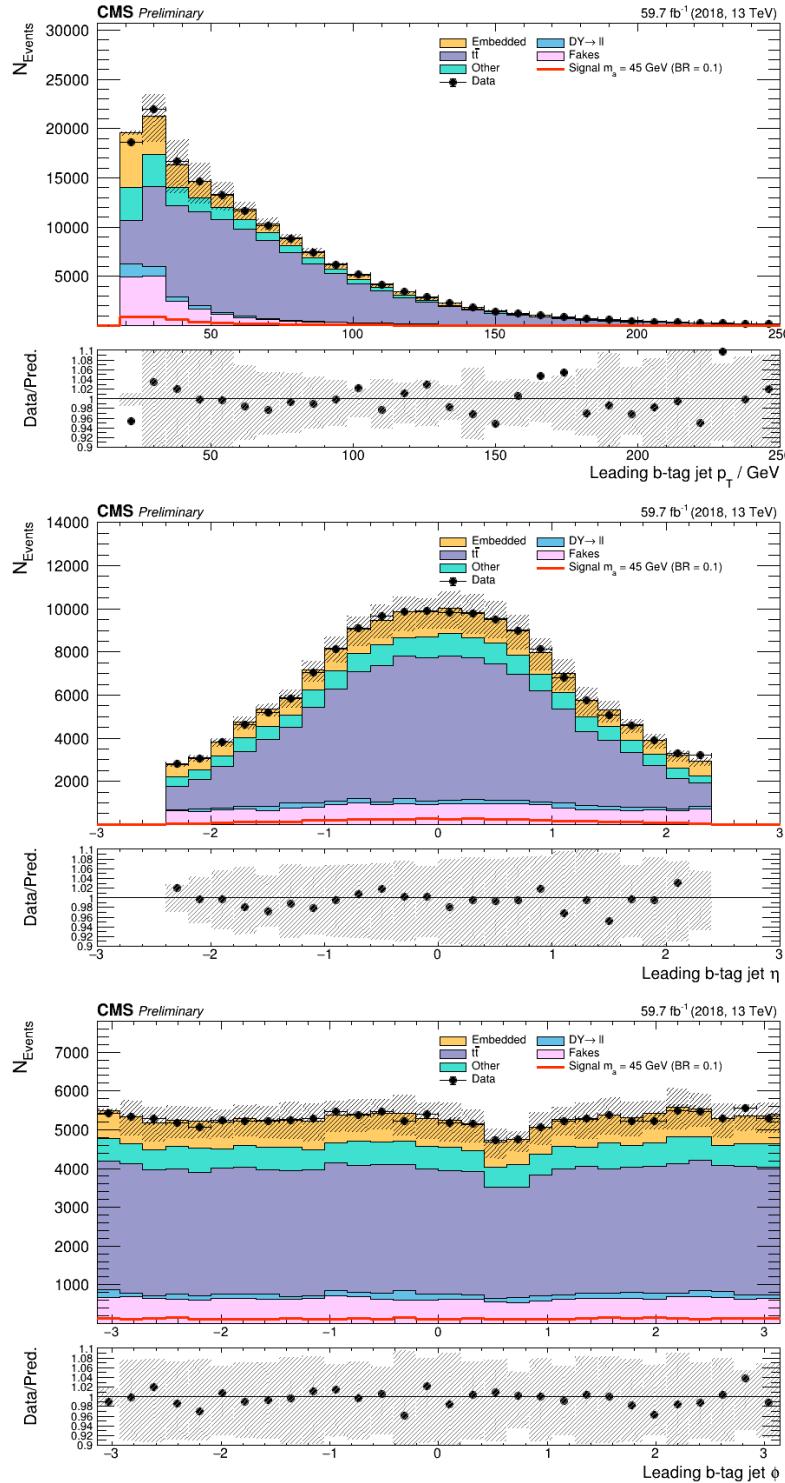


Figure 11.6: Kinematic properties of the leading b-tag jet in the  $\mu\tau_h$  final state using 2018 samples: transverse momentum  $p_T$  (top),  $\eta$  (middle),  $\phi$  (bottom). The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and  $\tau_h$  identification efficiency).

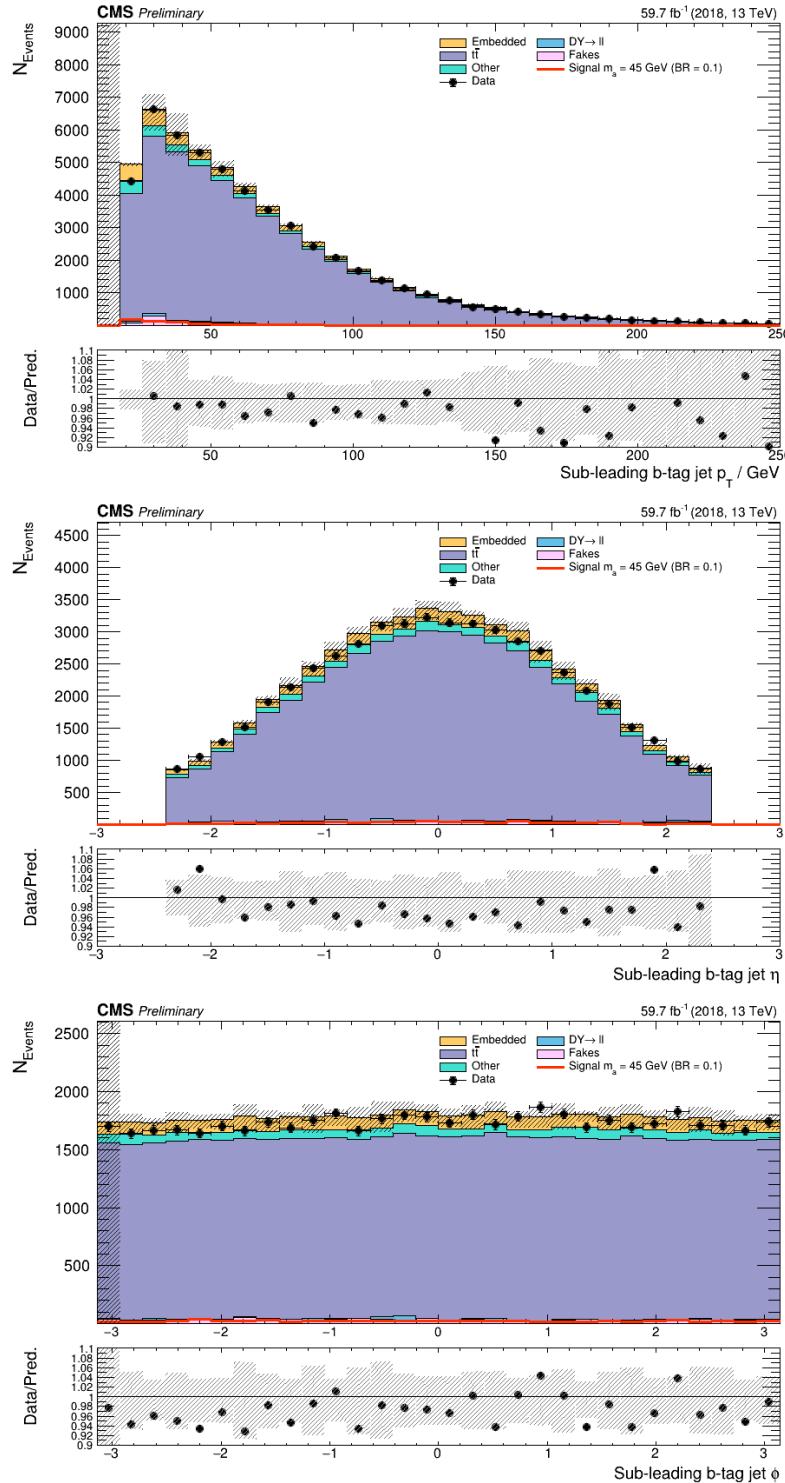


Figure 11.7: Kinematic properties of the sub-leading b-tag jet in the  $\mu\tau_h$  final state using 2018 samples: transverse momentum  $p_T$  (top),  $\eta$  (middle),  $\phi$  (bottom). The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and  $\tau_h$  identification efficiency).

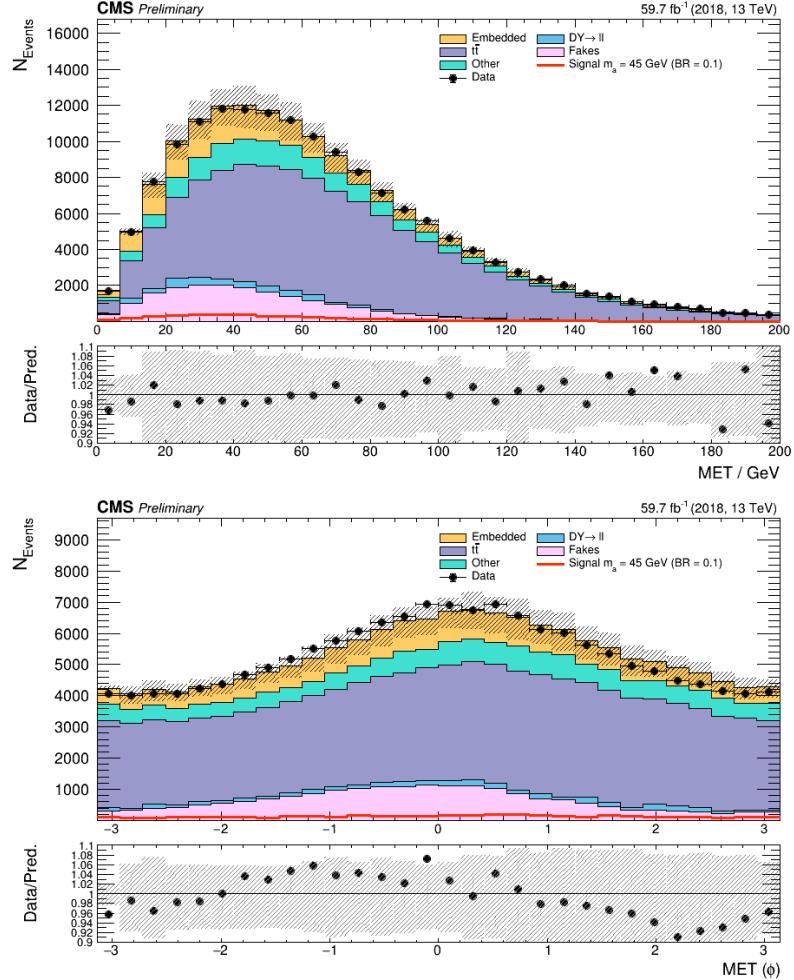


Figure 11.8: Missing transverse energy magnitude (*top*) and azimuthal direction (*bottom*) in the  $\mu\tau_h$  final state using 2018 samples. The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and  $\tau_h$  identification efficiency).

2587

# Chapter 12

2588

## Conclusion and outlook

2589 This thesis presents a direct search at the CMS experiment for exotic decays of the  
2590 Higgs boson with mass 125 GeV in data collected in the years 2016-2018 in proton-  
2591 proton collisions at center-of-mass energy 13 TeV, to two light neutral scalar particles  
2592 that decay to two bottom quarks and two tau leptons ( $h \rightarrow aa \rightarrow bb\tau\tau$ ). The results  
2593 are combined with another search that was performed in the  $h \rightarrow aa \rightarrow bb\mu\mu$  final  
2594 state, giving the most stringent limits to date for theories with Two Higgs Doublet  
2595 Models extended with a singlet scalar (2HDM+S), for pseudoscalar masses  $m_a$  ranging  
2596 from 15 GeV to 60 GeV, in a number of 2HDM+S scenarios such as type II and III  
2597 with  $\tan\beta = 2.0$ .

2598 As the rich physics program of CMS has set stringent limits on the exotic decay  
2599  $h \rightarrow aa$ , we turn our attention to direct searches for decays to light neutral scalars  
2600 with potentially unequal mass,  $h \rightarrow a_1a_2$ , which has not been performed at CMS  
2601 to date. Preliminary studies on  $h \rightarrow a_1a_2$  signals in the Two Real Singlet Model  
2602 (TRSM) are shown, and work is ongoing to develop the analysis for  $h \rightarrow a_1a_2$  in final  
2603 states with bottom quarks and tau leptons.

2604 To ensure the continued performance of the CMS detector and to enhance its  
2605 data-taking capabilities in the intense pile-up conditions of the Phase-2 upgrade of

2606 the High-Luminosity LHC, upgrades of the Level-1 Trigger are paramount for filtering  
2607 the increased data rate of the HL-LHC. This thesis also presents work on the stan-  
2608 dalone barrel calorimeter algorithm for reconstructing and identifying electron and  
2609 photon candidates, using high granularity crystal-level information from the ECAL  
2610 subdetector. For Phase-2, the increase in the granularity of information sent from  
2611 the electromagnetic calorimeter to the Level-1 trigger, from energy sums over towers  
2612 (which are  $5 \times 5$  in crystals) to crystal-level information, allows for the implementation  
2613 of a more sophisticated clustering algorithm that can exploit the fact that genuine  
2614 electrons and photons tend to leave energies concentrated a  $3 \times 5$  window in crystals,  
2615 and use shape and isolation information to distinguish genuine electrons and photons  
2616 from noise. Electrons and photons are key to characterizing Standard Model pro-  
2617 cesses and performing searches for new physics, and this represents one of the many  
2618 upgrades of the CMS detector in preparation for Phase-2. With the ongoing Run-3  
2619 data collecting period, and wealth of ongoing and scheduled upgrades, there remains  
2620 an abundance of directions for detector development and physics at CMS heading  
2621 into Phase-2 of the LHC.

# <sup>2622</sup> Appendix A

## <sup>2623</sup> Samples used

<sup>2624</sup> The datasets used in the MiniAOD-based framework for the  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis  
<sup>2625</sup> are listed in this appendix. The NanoAOD-based framework uses the NanoAOD ver-  
<sup>2626</sup> sions of these datasets. The data used for the years 2016-2018 are listed in Tables A.1,  
<sup>2627</sup> A.2, and A.3 respectively. The embedded samples used for the years 2016-2018 are  
<sup>2628</sup> listed in Tables A.4, A.5, and A.6 respectively. The Monte Carlo simulated samples  
<sup>2629</sup> used to estimate backgrounds for the years 2016-2018 are listed in Tables A.7, A.8,  
<sup>2630</sup> and A.9 respectively.

<sup>2631</sup> The  $h \rightarrow aa \rightarrow bb\tau\tau$  signal samples are generated for 11 pseudoscalar masses  
<sup>2632</sup> between 12 GeV and 60 GeV for gluon fusion (ggF) and vector boson fusion (VBF)  
<sup>2633</sup> Higgs production. The 2016-2018 signal samples are listed in Tables A.10, A.11 and  
<sup>2634</sup> A.12 respectively. A filter is applied at the generator level for each  $\tau\tau$  final state:

- <sup>2635</sup> •  $ee$  final state:  $p_T(e_1) > 22$  GeV,  $p_T(e_2) > 10$  GeV,  $|\eta(e_1)| < 2.6$ , and  $|\eta(e_2)| <$   
<sup>2636</sup> 2.6.
- <sup>2637</sup> •  $e\tau_h$  final state:  $p_T(e) > 22$  GeV,  $p_T(\tau_h) > 16$  GeV,  $|\eta(e)| < 2.6$ , and  $|\eta(\tau_h)| < 2.7$ .
- <sup>2638</sup> •  $e\mu$  final state:  $p_T(e) > 11$  GeV,  $p_T(\mu) > 7$  GeV,  $|\eta(e)| < 2.6$ , and  $|\eta(\mu)| < 2.5$ .
- <sup>2639</sup> •  $\tau_h\tau_h$  final state:  $p_T(\tau_{h1}) > 28$  GeV,  $p_T(\tau_{h2}) > 28$  GeV,  $|\eta(\tau_{h1})| < 2.5$ , and

2640  $|\eta(\tau_{h2})| < 2.5$ .

2641 •  $\mu\tau_h$  final state:  $p_T(\mu) > 19 \text{ GeV}$ ,  $p_T(\tau_h) > 16 \text{ GeV}$ ,  $|\eta(\mu)| < 2.5$ , and  $|\eta(\tau_h)| <$   
 2642  $2.7$ .

2643 •  $\mu\mu$  final state:  $p_T(\mu_1) > 17 \text{ GeV}$ ,  $p_T(\mu_2) > 8 \text{ GeV}$ ,  $|\eta(\mu_1)| < 2.5$ , and  $|\eta(\mu_2)| <$   
 2644  $2.5$ .

2645 The tables also show for each sample the filter efficiencies, which is the percentage  
 2646 of events that pass the above filters, and the number of events that were generated  
 2647 after applying the filters.

Channel	Datasets (2016)	Run range
$e\mu$	/MuonEG/Run2016B-17Jul2018_ver1-v1/MINIAOD	272760-273017
	/MuonEG/Run2016B-17Jul2018_ver2-v1/MINIAOD	273150-275376
	/MuonEG/Run2016C-17Jul2018-v1/MINIAOD	275656-276283
	/MuonEG/Run2016D-17Jul2018-v1/MINIAOD	276315-276811
	/MuonEG/Run2016E-17Jul2018-v2/MINIAOD	276831-277420
	/MuonEG/Run2016F-17Jul2018-v1/MINIAOD	277932-278808
	/MuonEG/Run2016G-17Jul2018-v1/MINIAOD	278820-280385
	/MuonEG/Run2016H-17Jul2018-v1/MINIAOD	281613-284044
$e\tau_h$	/SingleElectron/Run2016B-17Jul2018_ver1-v1/MINIAOD	272760-273017
	/SingleElectron/Run2016B-17Jul2018_ver2-v1/MINIAOD	273150-275376
	/SingleElectron/Run2016C-17Jul2018-v1/MINIAOD	275656-276283
	/SingleElectron/Run2016D-17Jul2018-v1/MINIAOD	276315-276811
	/SingleElectron/Run2016E-17Jul2018-v1/MINIAOD	276831-277420
	/SingleElectron/Run2016F-17Jul2018-v1/MINIAOD	277932-278808
	/SingleElectron/Run2016G-17Jul2018-v1/MINIAOD	278820-280385
	/SingleElectron/Run2016H-17Jul2018-v1/MINIAOD	281613-284044
$\mu\tau_h$	/SingleMuon/Run2016B-17Jul2018_ver1-v1/MINIAOD	272760-273017
	/SingleMuon/Run2016B-17Jul2018_ver2-v1/MINIAOD	273150-275376
	/SingleMuon/Run2016C-17Jul2018-v1/MINIAOD	275656-276283
	/SingleMuon/Run2016D-17Jul2018-v1/MINIAOD	276315-276811
	/SingleMuon/Run2016E-17Jul2018-v1/MINIAOD	276831-277420
	/SingleMuon/Run2016F-17Jul2018-v1/MINIAOD	277932-278808
	/SingleMuon/Run2016G-17Jul2018-v1/MINIAOD	278820-280385
	/SingleMuon/Run2016H-17Jul2018-v1/MINIAOD	281613-284044

Table A.1: Datasets used in the  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis for the 2016 era.

Channel	Datasets (2017)	Run range
$e\mu$	/MuonEG/Run2017B-31Mar2018-v1/MINIAOD	297047-299329
	/MuonEG/Run2017C-31Mar2018-v1/MINIAOD	299368-302029
	/MuonEG/Run2017D-31Mar2018-v1/MINIAOD	302031-302663
	/MuonEG/Run2017E-31Mar2018-v1/MINIAOD	303824-304797
	/MuonEG/Run2017F-31Mar2018-v1/MINIAOD	305040-306460
$e\tau_h$	/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD	297047-299329
	/SingleElectron/Run2017C-31Mar2018-v1/MINIAOD	299368-302029
	/SingleElectron/Run2017D-31Mar2018-v1/MINIAOD	302031-302663
	/SingleElectron/Run2017E-31Mar2018-v1/MINIAOD	303824-304797
	/SingleElectron/Run2017F-31Mar2018-v1/MINIAOD	305040-306460
$\mu\tau_h$	/SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	297047-299329
	/SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368-302029
	/SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	302031-302663
	/SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824-304797
	/SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040-306460

Table A.2: Datasets used in the  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis for the 2017 era.

Channel	Datasets (2018)	Run range
$e\mu$	/MuonEG/Run2018A-17Sep2018-v1/MINIAOD	315257-316995
	/MuonEG/Run2018B-17Sep2018-v1/MINIAOD	317080-319310
	/MuonEG/Run2018C-17Sep2018-v1/MINIAOD	319337-320065
	/MuonEG/Run2018D-PromptReco-v2/MINIAOD	320500-325175
$e\tau_h$	/EGamma/Run2018A-17Sep2018-v2/MINIAOD	315257-316995
	/EGamma/Run2018B-17Sep2018-v1/MINIAOD	317080-319310
	/EGamma/Run2018C-17Sep2018-v1/MINIAOD	319337-320065
	/EGamma/Run2018D-PromptReco-v2/MINIAOD	320497-325175
$\mu\tau_h$	/SingleMuon/Run2018A-17Sep2018-v2/MINIAOD	315257-316995
	/SingleMuon/Run2018B-17Sep2018-v1/MINIAOD	317080-319310
	/SingleMuon/Run2018C-17Sep2018-v1/MINIAOD	319337-320065
	/SingleMuon/Run2018D-PromptReco-v2/MINIAOD	320500-325175

Table A.3: Datasets used in the  $h \rightarrow aa \rightarrow bb\tau\tau$  analysis for the 2018 eras.

Channel	Embedded samples (2016)
$e\mu$	/EmbeddingRun2016B/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016C/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016D/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016E/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016F/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016G/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016H/ElMuFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
$e\tau_h$	/EmbeddingRun2016B/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016C/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016D/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016E/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016F/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016G/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016H/ElTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
$\mu\tau_h$	/EmbeddingRun2016B/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016C/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016D/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016E/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016F/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016G/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5
	/EmbeddingRun2016H/MuTauFinalState-inputDoubleMu_94X_Legacy_miniAOD-v5

Table A.4: Embedded samples used in the analysis for the 2016 era.

Channel	Embedded samples (2017)
$e\mu$	/EmbeddingRun2017B/ElMuFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017C/ElMuFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017D/ElMuFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017E/ElMuFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017F/ElMuFinalState-inputDoubleMu_94X_miniAOD-v2
$e\tau_h$	/EmbeddingRun2017B/ElTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017C/ElTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017D/ElTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017E/ElTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017F/ElTauFinalState-inputDoubleMu_94X_miniAOD-v2
$\mu\tau_h$	/EmbeddingRun2017B/MuTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017C/MuTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017D/MuTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017E/MuTauFinalState-inputDoubleMu_94X_miniAOD-v2 /EmbeddingRun2017F/MuTauFinalState-inputDoubleMu_94X_miniAOD-v2

Table A.5: Embedded samples used in the analysis for the 2017 era.

Channel	Embedded samples (2018)
$e\mu$	/EmbeddingRun2018A/ElMuFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018B/ElMuFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018C/ElMuFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018D/ElMuFinalState-inputDoubleMu_102X_miniAOD-v1
$e\tau_h$	/EmbeddingRun2018A/ElTauFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018B/ElTauFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018C/ElTauFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018D/ElTauFinalState-inputDoubleMu_102X_miniAOD-v1
$\mu\tau_h$	/EmbeddingRun2018A/MuTauFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018B/MuTauFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018C/MuTauFinalState-inputDoubleMu_102X_miniAOD-v1 /EmbeddingRun2018D/MuTauFinalState-inputDoubleMu_102X_miniAOD-v1

Table A.6: Embedded samples used in the analysis for the 2018 era.

Process	Simulated background samples (2016)	Cross section (pb)
DY	/DY1JetsToLL_M-50_TuneCUETP8M1	1012.5 (LO)
	/DY2JetsToLL_M-50_TuneCUETP8M1	332.8 (LO)
	/DY3JetsToLL_M-50_TuneCUETP8M1	101.8 (LO)
	/DY4JetsToLL_M-50_TuneCUETP8M1	54.8 (LO)
	/DYJetsToLL_M-50_TuneCUETP8M1	4963.0 (LO)
	/DY1JetsToLL_M-10to50_TuneCUETP8M1	730.3 (LO)
	/DY2JetsToLL_M-10to50_TuneCUETP8M1	387.4 (LO)
	/DY3JetsToLL_M-10to50_TuneCUETP8M1	95.0 (LO)
	/DY4JetsToLL_M-10to50_TuneCUETP8M1	36.7 (LO)
	/DYJetsToLL_M-10to50_TuneCUETP8M1	16290.0 (LO)
Top	/TTTo2L2Nu_TuneCP5_PSweights	88.29
	/TTToHadronic_TuneCP5_PSweights	377.96
	/TTToSemiLeptonic_TuneCP5_PSweights	365.35
	/ST_t-channel_antitop_4f_inclusiveDecays <sup>†</sup>	26.23
	/ST_t-channel_top_4f_inclusiveDecays <sup>†</sup>	44.07
	/ST_tW_antitop_5f_inclusiveDecays_TuneCUETP8M1	35.6
	/ST_tW_top_5f_inclusiveDecays_TuneCUETP8M1	35.6
VV	/VVTTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8	13.84
	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.52
	/WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	4.43
	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.38
	/ZZTo4L_13TeV-amcatnloFXFX-pythia8	1.212
W	/W1JetsToLNu_TuneCUETP8M1	8104.0 (LO)
	/W2JetsToLNu_TuneCUETP8M1	2793.0 (LO)
	/W3JetsToLNu_TuneCUETP8M1	992.5 (LO)
	/W4JetsToLNu_TuneCUETP8M1	544.3 (LO)
	/WJetsToLNu_TuneCUETP8M1	52940.0 (LO)
H	/GluGluHToTauTau_M125	48.58*0.0627
	/GluGluHToWWTo2L2Nu_M125	48.58*0.2137*0.3258*0.3258
	/GluGluZH_HToWW_M125	0.1227*0.2137
	/HWminusJ_HToWW_M125	0.5328*0.2137
	/HWplusJ_HToWW_M125	0.840*0.2137
	/HZJ_HToWW_M125	0.7612*0.2137
	/VBFHToTauTau_M125	3.782*0.0627
	/VBFHToWWTo2L2Nu_M125	3.782*0.2137*0.3258*0.3258
	/WminusHToTauTau_M125	0.5328*0.0627
	/WplusHToTauTau_M125	0.840*0.0627
	/ZHToTauTau_M125	0.7612*0.0627
	/ggZH_HToTauTau_ZToLL_M125	0.1227*0.0627*3*0.033658
	/ggZH_HToTauTau_ZToNuNu_M125	0.1227*0.0627*0.2000
	/ggZH_HToTauTau_ZToQQ_M125	0.1227*0.0627*0.6991
	/ttHToNonbb_M125_TuneCUETP8M2_ttHtranche3	0.5071*(1-0.5824)
	/ttHTobb_M125_TuneCP5	0.5071*0.5824

Table A.7: Background MC samples used in the analysis for the 2016 era. Samples marked with a <sup>†</sup> are generated with the powhegV2-madspin-pythia8 tag.

Process	Simulated background samples (2017)	Cross section (pb)
DY	DY1JetsToLL_M-50_TuneCP5	877.8 (LO)
	DY2JetsToLL_M-50_TuneCP5	304.4 (LO)
	DY3JetsToLL_M-50_TuneCP5	111.5 (LO)
	DY4JetsToLL_M-50_TuneCP5	44.0 (LO)
	DYJetsToLL_M-50_TuneCP5	5343.0 (LO)
	DYJetsToLL_M-10to50_TuneCP5	15810.0 (LO)
Top	TTTo2L2Nu_TuneCP5	88.29
	TTToHadronic_TuneCP5	377.96
	TTToSemileptonic_TuneCP5	365.35
	ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5 <sup>†</sup>	80.94
	ST_t-channel_top_4f_inclusiveDecays_TuneCP5 <sup>†</sup>	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5	35.85
	ST_tW_top_5f_inclusiveDecays_TuneCP5	35.85
VV	VVTo2L2Nu_13TeV_amcatnloFXFX_madspin_pythia8	13.84
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	5.52
	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	4.43
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8	3.38
	ZZTo4L_TuneCP5_13TeV-amcatnloFXFX-pythia8	1.212
W	W1JetsToLNu_TuneCP5	8104.0 (LO)
	W2JetsToLNu_TuneCP5	2793.0 (LO)
	W3JetsToLNu_TuneCP5	992.5 (LO)
	W4JetsToLNu_TuneCP5	544.3 (LO)
	WJetsToLNu_TuneCP5	52940.0 (LO)
H	GluGluHToTauTau_M125	48.58*0.0627
	GluGluHToWWTo2L2Nu_M125 <sup>††</sup>	48.58*0.2137*0.3258*0.3258
	GluGluZH_HToWW_M125	0.1227*0.2137
	HWminusJ_HToWW_M125	0.5328*0.2137
	HWplusJ_HToWW_M125	0.840*0.2137
	HZJ_HToWW_M125 <sup>††</sup>	0.7612*0.2137
	VBFHToTauTau_M125	3.782*0.0627
	VBFHToWWTo2L2Nu_M125 <sup>††</sup>	3.782*0.2137*0.3258*0.3258
	WminusHToTauTau_M125	0.5328*0.0627
	WplusHToTauTau_M125	0.840*0.0627
	ZHToTauTau_M125	0.7612*0.0627
	ggZH_HToTauTau_ZToLL_M125	0.1227*0.0627*3*0.033658
	ggZH_HToTauTau_ZToNuNu_M125	0.1227*0.0627*0.2000
	ggZH_HToTauTau_ZToQQ_M125	0.1227*0.0627*0.6991
	ttHToNonbb_M125_TuneCP5	0.5071*(1-0.5824)
	ttHTobb_M125_TuneCP5	0.5071*0.5824

Table A.8: Background MC samples used in the analysis for the 2017 era. All samples use powheg, except the DYJets and WJets samples, which use madgraphMLM. Samples marked with a <sup>†</sup>, <sup>††</sup>, or <sup>†††</sup> were produced with Powheg2 and Pythia8, and Madspin, JHUGenV714, or jhugen724 respectively.

Process	Simulated background samples (2018)	Cross section (pb)
DY	DY1JetsToLL_M-50_TuneCP5	877.8 (LO)
	DY2JetsToLL_M-50_TuneCP5	304.4 (LO)
	DY3JetsToLL_M-50_TuneCP5	111.5 (LO)
	DY4JetsToLL_M-50_TuneCP5	44.0 (LO)
	DYJetsToLL_M-50_TuneCP5	5343.0 (LO)
	DYJetsToLL_M-10to50_TuneCP5	15810.0 (LO)
Top	TTTo2L2Nu_TuneCP5	88.29
	TTToHadronic_TuneCP5	377.96
	TTToSemiLeptonic_TuneCP5	365.35
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5 <sup>†</sup>	80.94
	ST_t-channel_top_5f_TuneCP5 <sup>†</sup>	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5	35.85
VV	ST_tW_top_5f_inclusiveDecays	35.85
	VVTTo2L2Nu_13TeV_amcatnloFXFX_madspin	13.84
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin	5.52
	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	4.43
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin	3.38
W	ZZTo4L_TuneCP5_13TeV-amcatnloFXFX-pythia8	1.212
	W1JetsToLNu_TuneCP5	8104.0 (LO)
	W2JetsToLNu_TuneCP5	2793.0 (LO)
	W3JetsToLNu_TuneCP5	992.5 (LO)
	W4JetsToLNu_TuneCP5	544.3 (LO)
H	WJetsToLNu_TuneCP5	52940.0 (LO)
	GluGluHToTauTau_M125	48.58*0.0627
	GluGluHToWWTo2L2Nu_M125 <sup>††</sup>	48.58*0.2137*0.3258*0.3258
	GluGluZH_HToWW_M125	0.1227*0.2137
	HWminusJ_HToWW_M125 <sup>†††</sup>	0.5328*0.2137
	HWplusJ_HToWW_M125 <sup>†††</sup>	0.840*0.2137
	HZJ_HToWW_M125 <sup>††</sup>	0.7612*0.2137
	VBFHToTauTau_M125	3.782*0.0627
	VBFHToWWTo2L2Nu_M125 <sup>†††</sup>	3.782*0.2137*0.3258*0.3258
	WminusHToTauTau_M125	0.5328*0.0627
	WplusHToTauTau_M125	0.840*0.0627
	ZHToTauTau_M125	0.7612*0.0627
	ggZH_HToTauTau_ZToLL_M125	0.1227*0.0627*3*0.033658
	ggZH_HToTauTau_ZToNuNu_M125	0.1227*0.0627*0.2000
	ggZH_HToTauTau_ZToQQ_M125	0.1227*0.0627*0.6991
	ttHTobb_M125_TuneCP5	0.5071*(1-0.5824)
	ttHTobb_M125_TuneCP5	0.5071*0.5824

Table A.9: Background Monte Carlo samples used in the analysis for the 2018 era. All samples listed are generated for 13 TeV collisions and use pythia8. All samples use powheg, except the DYJets and WJets samples, which use madgraphMLM. Samples marked with a <sup>†</sup>, <sup>††</sup>, or <sup>†††</sup>, were produced with Powheg and Pythia8, and Madspin, JHUGenV714, and Jhugen724 respectively.

Signal samples (2016)	# events	Filter eff.
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-12_FilterTauTauTrigger	0.4M	3.81%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-15_FilterTauTauTrigger	0.4M	3.54%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-20_FilterTauTauTrigger	1M	3.37
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-25_FilterTauTauTrigger	0.2M	3.56%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-30_FilterTauTauTrigger	0.2M	3.16%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-35_FilterTauTauTrigger	0.2M	3.30%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-40_FilterTauTauTrigger	1M	3.30%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-45_FilterTauTauTrigger	0.2M	3.23%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-50_FilterTauTauTrigger	0.2M	3.42%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-55_FilterTauTauTrigger	0.2M	3.65%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-60_FilterTauTauTrigger	1M	3.73
/SUSYVBFHToAA_AToBB_AToTauTau_M-12_FilterTauTauTrigger	0.2M	7.94%
/SUSYVBFHToAA_AToBB_AToTauTau_M-15_FilterTauTauTrigger	0.2M	7.38%
/SUSYVBFHToAA_AToBB_AToTauTau_M-20_FilterTauTauTrigger	0.2M	7.27%
/SUSYVBFHToAA_AToBB_AToTauTau_M-25_FilterTauTauTrigger	0.2M	7.21%
/SUSYVBFHToAA_AToBB_AToTauTau_M-30_FilterTauTauTrigger	0.2M	6.87%
/SUSYVBFHToAA_AToBB_AToTauTau_M-35_FilterTauTauTrigger	0.2M	6.80%
/SUSYVBFHToAA_AToBB_AToTauTau_M-40_FilterTauTauTrigger	0.2M	6.78%
/SUSYVBFHToAA_AToBB_AToTauTau_M-45_FilterTauTauTrigger	0.2M	6.56%
/SUSYVBFHToAA_AToBB_AToTauTau_M-50_FilterTauTauTrigger	0.2M	6.40%
/SUSYVBFHToAA_AToBB_AToTauTau_M-55_FilterTauTauTrigger	0.2M	6.54%
/SUSYVBFHToAA_AToBB_AToTauTau_M-60_FilterTauTauTrigger	0.2M	6.55%

Table A.10: Signal samples used in the analysis for the 2016 era. All belong to the RunIISummer16MiniAODv3 campaign and are produced with Madgraph and Pythia8. The second column is the number of events after the generator-level filter is applied, and the third column is the filter efficiency (percentage of all events that pass the generator-level filter).

Signal samples (2017)	# events	Filter eff.
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-12_FilterTauTauTrigger	0.4M	3.78%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-15_FilterTauTauTrigger	0.4M	3.55%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-20_FilterTauTauTrigger	1M	3.40%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-25_FilterTauTauTrigger	0.2M	3.32%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-30_FilterTauTauTrigger	0.2M	3.36%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-35_FilterTauTauTrigger	0.2M	3.27%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-40_FilterTauTauTrigger	1M	3.03%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-45_FilterTauTauTrigger	0.2M	3.03%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-50_FilterTauTauTrigger	0.2M	3.31%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-55_FilterTauTauTrigger	0.2M	3.56%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-60_FilterTauTauTrigger	1M	3.95%
/SUSYVBFHToAA_AToBB_AToTauTau_M-12_FilterTauTauTrigger	0.2M	7.73%
/SUSYVBFHToAA_AToBB_AToTauTau_M-15_FilterTauTauTrigger	0.2M	7.35%
/SUSYVBFHToAA_AToBB_AToTauTau_M-20_FilterTauTauTrigger	0.2M	7.33%
/SUSYVBFHToAA_AToBB_AToTauTau_M-25_FilterTauTauTrigger	0.2M	7.23%
/SUSYVBFHToAA_AToBB_AToTauTau_M-30_FilterTauTauTrigger	0.2M	6.84%
/SUSYVBFHToAA_AToBB_AToTauTau_M-35_FilterTauTauTrigger	0.2M	6.97%
/SUSYVBFHToAA_AToBB_AToTauTau_M-40_FilterTauTauTrigger	0.2M	6.17%
/SUSYVBFHToAA_AToBB_AToTauTau_M-45_FilterTauTauTrigger	0.2M	6.67%
/SUSYVBFHToAA_AToBB_AToTauTau_M-50_FilterTauTauTrigger	0.2M	6.61%
/SUSYVBFHToAA_AToBB_AToTauTau_M-55_FilterTauTauTrigger	0.2M	6.51%
/SUSYVBFHToAA_AToBB_AToTauTau_M-60_FilterTauTauTrigger	0.2M	6.71%

Table A.11: Signal samples used in the analysis for the 2017 era. All belong to the RunIIFall17MiniAODv2 campaign and are produced with Madgraph and Pythia8. The second column is the number of events after the generator-level filter is applied, and the third column is the filter efficiency (percentage of all events that pass the generator-level filter).

Signal samples (2018)	# events	Filter eff.
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-12_FilterTauTauTrigger	0.4M	3.78%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-15_FilterTauTauTrigger	0.4M	3.49%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-20_FilterTauTauTrigger	1M	3.36%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-25_FilterTauTauTrigger	0.2M	3.46%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-30_FilterTauTauTrigger	0.2M	3.18%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-35_FilterTauTauTrigger	0.2M	3.28%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-40_FilterTauTauTrigger	1M	3.10%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-45_FilterTauTauTrigger	0.2M	3.21%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-50_FilterTauTauTrigger	0.2M	3.14%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-55_FilterTauTauTrigger	0.2M	3.56%
/SUSYGluGluToHToAA_AToBB_AToTauTau_M-60_FilterTauTauTrigger	1M	3.38%
/SUSYVBFHToAA_AToBB_AToTauTau_M-12_FilterTauTauTrigger	0.2M	7.78%
/SUSYVBFHToAA_AToBB_AToTauTau_M-15_FilterTauTauTrigger	0.2M	7.52%
/SUSYVBFHToAA_AToBB_AToTauTau_M-20_FilterTauTauTrigger	0.2M	6.87%
/SUSYVBFHToAA_AToBB_AToTauTau_M-25_FilterTauTauTrigger	0.2M	7.21%
/SUSYVBFHToAA_AToBB_AToTauTau_M-30_FilterTauTauTrigger	0.2M	6.51%
/SUSYVBFHToAA_AToBB_AToTauTau_M-35_FilterTauTauTrigger	0.2M	6.95%
/SUSYVBFHToAA_AToBB_AToTauTau_M-40_FilterTauTauTrigger	0.2M	6.81%
/SUSYVBFHToAA_AToBB_AToTauTau_M-45_FilterTauTauTrigger	0.2M	6.62%
/SUSYVBFHToAA_AToBB_AToTauTau_M-50_FilterTauTauTrigger	0.2M	6.56%
/SUSYVBFHToAA_AToBB_AToTauTau_M-55_FilterTauTauTrigger	0.2M	6.64%
/SUSYVBFHToAA_AToBB_AToTauTau_M-60_FilterTauTauTrigger	0.2M	6.75%

Table A.12: Signal samples used in the analysis for the 2018 era. All belong to the RunIIAutumn18MiniAOD campaign and are produced with Madgraph and Pythia8. The second column is the number of events after the generator-level filter is applied, and the third column is the filter efficiency (percentage of all events that pass the generator-level filter).

# <sup>2648</sup> Bibliography

- <sup>2649</sup> [1] Paul H. Frampton. Journeys Beyond the Standard Model. 54(1):52–  
<sup>2650</sup> 52. ISSN 0031-9228. doi: 10.1063/1.1349615. URL <https://doi.org/10.1063/1.1349615>. eprint: [https://pubs.aip.org/physicstoday/article-pdf/54/1/52/11109432/52\\_1\\_online.pdf](https://pubs.aip.org/physicstoday/article-pdf/54/1/52/11109432/52_1_online.pdf).
- <sup>2653</sup> [2] Meinard Kuhlmann. Quantum Field Theory. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research  
<sup>2654</sup> Lab, Stanford University, Summer 2023 edition, 2023.
- <sup>2655</sup> [3] Steven Weinberg. The Making of the Standard Model. *Eur. Phys. J. C*, 34:  
<sup>2656</sup> 5–13, 2004. URL <https://cds.cern.ch/record/799984>.
- <sup>2658</sup> [4] Christopher G. Tully. *Elementary Particle Physics in a Nutshell*. Princeton  
<sup>2659</sup> University Press, Princeton, 2012. ISBN 9781400839353. doi: doi:10.1515/  
<sup>2660</sup> 9781400839353. URL <https://doi.org/10.1515/9781400839353>.
- <sup>2661</sup> [5] John Ellis. Higgs Physics. In *2013 European School of High-Energy Physics*,  
<sup>2662</sup> pages 117–168, 2015. doi: 10.5170/CERN-2015-004.117.
- <sup>2663</sup> [6] David Curtin, Rouven Essig, Stefania Gori, and Others. Exotic decays of the 125  
<sup>2664</sup> GeV Higgs boson. *Phys. Rev. D*, 90:075004, Oct 2014. doi: 10.1103/PhysRevD.  
<sup>2665</sup> 90.075004. URL <https://link.aps.org/doi/10.1103/PhysRevD.90.075004>.
- <sup>2666</sup> [7] Tania Robens, Tim Stefański, and Jonas Wittbrodt. Two-real-scalar-singlet

- extension of the SM: LHC phenomenology and benchmark scenarios. *Eur. Phys. J. C*, 80(2):151, 2020. doi: 10.1140/epjc/s10052-020-7655-x.
- [8] CERN. The history of CERN, 2024. URL <https://timeline.web.cern.ch/timeline-header/89>.
- [9] R. Schmidt. Accelerator physics and technology of the LHC. In *ROXIE: Routine for the Optimizazation of Magnet X-Sections, Inverse Field Calculation and Coil End Design*, pages 7–17, 1998.
- [10] J. Vollaire et al. *Linac4 design report*, volume 6/2020 of *CERN Yellow Reports: Monographs*. CERN, Geneva, 9 2020. ISBN 978-92-9083-579-0, 978-92-9083-580-6. doi: 10.23731/CYRM-2020-006.
- [11] Antonella Del Rosso. Aerial view of the LHC and the four major experiments. 2017. URL <https://cds.cern.ch/record/2253966>. General Photo.
- [12] ATLAS Collaboration. ATLAS: Detector and physics performance technical design report. Volume 1. 5 1999.
- [13] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006.
- [14] L Musa. Conceptual Design Report for the Upgrade of the ALICE ITS. Technical report, CERN, Geneva, 2012. URL <https://cds.cern.ch/record/1431539>.
- [15] S. Amato et al. LHCb technical proposal: A Large Hadron Collider Beauty Experiment for Precision Measurements of CP Violation and Rare Decays. 2 1998.
- [16] Werner Herr and B Muratori. Concept of luminosity. 2006. doi: 10.5170/CERN-2006-002.361. URL <https://cds.cern.ch/record/941318>.

- 2691 [17] Olivier S. Brning and Frank Zimmerman. Parameter space for the LHC lu-  
2692 minosity upgrade. ISBN 978-3-95450-115-1. URL <https://accelconf.web.cern.ch/IPAC2012/papers/moppc005.pdf>.  
2693
- 2694 [18] LHC Design Report Vol.1: The LHC Main Ring. 6 2004. doi: 10.5170/  
2695 CERN-2004-003-V-1.
- 2696 [19] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09):  
2697 P09018, 2020. doi: 10.1088/1748-0221/15/09/P09018.
- 2698 [20] CMS Collaboration. Measurement of the inelastic proton-proton cross section  
2699 at  $\sqrt{s} = 13$  TeV. *JHEP*, 07:161, 2018. doi: 10.1007/JHEP07(2018)161.
- 2700 [21] CMS Collaboration. High-Luminosity Large Hadron Collider (HL-LHC): Tech-  
2701 nical design report. 10/2020, 12 2020. doi: 10.23731/CYRM-2020-0010.
- 2702 [22] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:  
2703 S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.
- 2704 [23] CMS Collaboration. Particle-flow reconstruction and global event description  
2705 with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/  
2706 12/10/P10003.
- 2707 [24] V Karimki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freuden-  
2708 reich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker sys-  
2709 tem project: Technical Design Report*. Technical design report. CMS. CERN,  
2710 Geneva, 1997. URL <https://cds.cern.ch/record/368412>.
- 2711 [25] The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva,  
2712 2017. URL <https://cds.cern.ch/record/2272264>.
- 2713 [26] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Up-  
2714 grade. 9 2012. doi: 10.2172/1151650.

- 2715 [27] R. L. Workman and Others. Review of particle physics. 2022:083C01. doi:  
2716 10.1093/ptep/ptac097.
- 2717 [28] CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorime-  
2718 ter. 9 2012. doi: 10.2172/1151651.
- 2719 [29] CMS Technical Design Report for the Level-1 Trigger Upgrade. 6 2013.
- 2720 [30] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The  
2721 trigger systems. 12 2000.
- 2722 [31] Alex Tapper. The CMS Level-1 Trigger for LHC Run II. *PoS*, ICHEP2016:242,  
2723 2016. doi: 10.22323/1.282.0242.
- 2724 [32] A. Zabi, F. Beaudette, L. Cadamuro, O. Davignon, T. Romantreau, T. Strebler,  
2725 M. Cepeda, J.B. Sauvan, N. Wardle, R. Aggleton, F. Ball, J. Brooke, D. New-  
2726 bold, S. Paramesvaran, D. Smith, J. Taylor, C. Foudas, M. Baber, A. Bun-  
2727 dock, S. Breeze, M. Citron, A. Elwood, G. Hall, G. Iles, C. Laner, B. Pen-  
2728 ning, A. Rose, A. Shtipliyski, A. Tapper, I. Ojalvo, T. Durkin, K. Harder,  
2729 S. Harper, C. Shepherd-Themistocleous, A. Thea, T. Williams, S. Dasu,  
2730 L. Dodd, R. Forbes, T. Gorski, P. Klabbers, A. Levine, T. Ruggles, N. Smith,  
2731 W. Smith, A. Svetek, J. Tikalsky, and M. Vicente. The cms level-1 calorimeter  
2732 trigger for the lhc run ii. *Journal of Instrumentation*, 12(01):C01065, jan 2017.  
2733 doi: 10.1088/1748-0221/12/01/C01065. URL <https://dx.doi.org/10.1088/1748-0221/12/01/C01065>.
- 2735 [33] P. Klabbers et al. CMS level-1 upgrade calorimeter trigger prototype develop-  
2736 ment. *JINST*, 8:C02013, 2013. doi: 10.1088/1748-0221/8/02/C02013.
- 2737 [34] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and  
2738 High Level Trigger. Technical report, CERN, Geneva, 2021. URL [https:](https://)

- 2739           `//cds.cern.ch/record/2759072`. This is the final version of the document,  
2740           approved by the LHCC.
- 2741 [35] C. Foudas. The CMS Level-1 Trigger at LHC and Super-LHC. In *34th Inter-*  
2742           *national Conference on High Energy Physics*, 10 2008.
- 2743 [36] CMS Software Guide. High Level Trigger (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHighLevelTrigger>.
- 2744 [37] The Worldwide LHC Computing Grid. 2012. URL <https://cds.cern.ch/record/1997398>.
- 2745 [38] Douglas Thain, Todd Tannenbaum, and Miron Livny. Distributed computing  
2746           in practice: the Condor experience. *Concurrency and Computation: Practice*  
2747           *and Experience*, 17(2-4):323–356, 2005. doi: <https://doi.org/10.1002/cpe.938>.  
2748           URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.938>.
- 2749 [39] Alexandre Zabi, Jeffrey Wayne Berryhill, Emmanuelle Perez, and Alexander D.  
2750           Tapper. The Phase-2 Upgrade of the CMS Level-1 Trigger. 2020.
- 2751 [40] Technical proposal for a MIP timing detector in the CMS experiment Phase 2  
2752           upgrade. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2296612>.
- 2753 [41] CMS Collaboration. Search for exotic decays of the Higgs boson to a pair of  
2754           pseudoscalars in the  $\mu\mu bb$  and  $\tau\tau bb$  final states. *European Physical Journal C*,  
2755           2 2024.
- 2756 [42] CMS Collaboration. CMS luminosity measurement for the 2016 data-taking  
2757           period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL  
2758           <https://cds.cern.ch/record/2257069>.

- 2762 [43] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking  
2763 period at  $\sqrt{s} = 13$  TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-  
2764 004, 2018. URL <https://cds.cern.ch/record/2621960>.
- 2765 [44] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking  
2766 period at  $\sqrt{s} = 13$  TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-  
2767 002, 2019. URL <https://cds.cern.ch/record/2676164>.
- 2768 [45] CMS LUMI Group. CMS Luminosity Public Results (TWiki), 2024. URL  
2769 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- 2770 [46] J. Alwall, R. Frederix, S. Frixione, et al. The automated computation of tree-  
2771 level and next-to-leading order differential cross sections, and their matching  
2772 to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July  
2773 2014. ISSN 1029-8479. doi: 10.1007/jhep07(2014)079. URL [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- 2774 [47] R. Frederix, S. Frixione, V. Hirschi, et al. The automation of next-to-leading  
2775 order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July  
2776 2018. ISSN 1029-8479. doi: 10.1007/jhep07(2018)185. URL [http://dx.doi.org/10.1007/JHEP07\(2018\)185](http://dx.doi.org/10.1007/JHEP07(2018)185).
- 2777 [48] S. Agostinelli, J. Allison, K. Amako, et al. Geant4 - a simulation toolkit. 506(3):  
2778 250–303. ISSN 0168-9002. doi: 10.1016/S0168-9002(03)01368-8. URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- 2779 [49] CMS Collaboration. An embedding technique to determine  $\tau\tau$  backgrounds  
2780 in proton-proton collision data. *JINST*, 14(06):P06032, 2019. doi: 10.1088/  
2781 1748-0221/14/06/P06032.
- 2782 [50] CMS Collaboration. Search for neutral MSSM Higgs bosons decaying to a pair

- 2786 of tau leptons in pp collisions. *JHEP*, 10:160, 2014. doi: 10.1007/JHEP10(2014)  
2787 160.
- 2788 [51] CMS Collaboration. Measurements of Higgs boson production in the decay  
2789 channel with a pair of  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV.  
2790 *Eur. Phys. J. C*, 83(7):562, 2023. doi: 10.1140/epjc/s10052-023-11452-8.
- 2791 [52] CMS Collaboration. Reconstruction and identification of tau lepton decays to  
2792 hadrons and tau neutrinos at CMS. *Journal of Instrumentation*, 11(01):P01019–  
2793 P01019, January 2016. ISSN 1748-0221. doi: 10.1088/1748-0221/11/01/p01019.  
2794 URL <http://dx.doi.org/10.1088/1748-0221/11/01/P01019>.
- 2795 [53] CMS Collaboration. Performance of  $\tau$ -lepton reconstruction and identification  
2796 in CMS. *Journal of Instrumentation*, 7(01):P01001, jan 2012. doi: 10.1088/  
2797 1748-0221/7/01/P01001. URL <https://dx.doi.org/10.1088/1748-0221/7/01/P01001>.
- 2798 [54] CMS Collaboration. Performance of reconstruction and identification of  $\tau$  lep-  
2799 tons decaying to hadrons and  $\nu_\tau$  in pp collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 13  
2800 (10):P10005, 2018. doi: 10.1088/1748-0221/13/10/P10005.
- 2801 [55] CMS Collaboration. Identification of hadronic tau lepton decays using a deep  
2802 neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/  
2803 P07023.
- 2804 [56] CMS Collaboration. Performance of CMS muon reconstruction in pp collision  
2805 events at  $\sqrt{s} = 7$  TeV. *Journal of Instrumentation*, 7(10):P10002–P10002,  
2806 October 2012. ISSN 1748-0221. doi: 10.1088/1748-0221/7/10/p10002. URL  
2807 <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- 2808 [57] CMS Collaboration. Performance of electron reconstruction and selection with  
2809 the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Journal of Instru-*

- 2811        *mentation*, 10(06):P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005. URL  
2812        <https://dx.doi.org/10.1088/1748-0221/10/06/P06005>.
- 2813        [58] CMS Collaboration. Identification of b-quark jets with the CMS experiment.  
2814        *Journal of Instrumentation*, 8(04):P04013–P04013, April 2013. ISSN 1748-  
2815        0221. doi: 10.1088/1748-0221/8/04/p04013. URL <http://dx.doi.org/10.1088/1748-0221/8/04/P04013>.  
2816
- 2817        [59] CMS Collaboration. Pileup Removal Algorithms. Technical report, CERN,  
2818        Geneva, 2014. URL <https://cds.cern.ch/record/1751454>.
- 2819        [60] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and  
2820        developments. 2017. URL <https://cds.cern.ch/record/2263802>.
- 2821        [61] CMS Collaboration. Performance of the DeepJet b tagging algorithm using  
2822         $41.9 \text{ fb}^{-1}$  of data from proton-proton collisions at 13 TeV with Phase 1 CMS  
2823        detector. 2018. URL <https://cds.cern.ch/record/2646773>.
- 2824        [62] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Re-  
2825        construction of the Higgs mass in  $H \rightarrow \tau\tau$  Events by Dynamical Likelihood  
2826        techniques. *Journal of Physics: Conference Series*, 513(2):022035, jun 2014.  
2827        doi: 10.1088/1742-6596/513/2/022035. URL <https://dx.doi.org/10.1088/1742-6596/513/2/022035>.  
2828
- 2829        [63] CMS Collaboration. Evidence for the 125 GeV Higgs boson decaying to a pair  
2830        of  $\tau$  leptons. *JHEP*, 05:104, 2014. doi: 10.1007/JHEP05(2014)104.
- 2831        [64] CMS Collaboration. Missing transverse energy performance of the CMS detec-  
2832        tor. *JINST*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.
- 2833        [65] Artur Kalinowski. CMS AN-19-032 (internal): Reconstruction of a  $\tau$  pair in-  
2834        variant mass with a simplified likelihood scan, 2019.

- 2835 [66] CMS TAU POG. Tau Physics Object Group: Tau ID Recommendation  
2836 For Run 2, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2>.
- 2838 [67] CMS MUO POG. Muon Physics Object Group: Recommendations, 2024. URL  
2839 <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>.
- 2840 [68] CMS MUO POG. Muon Physics Object Group: Reference guidelines and results  
2841 for muon momentum scale and resolution in Run II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceScaleResolRun2>.
- 2843 [69] CMS HTT working group. Higgs To Tau Tau Working TWiki for the full Run-  
2844 2 legacy analysis, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/HiggsToTauTauWorkingLegacyRun2>.
- 2846 [70] CMS ELE POG. Electron Physics Object Group: Recommendations  
2847 , 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations>.
- 2849 [71] CMS ELE POG. Electron Physics Object Group: Recommendations for  
2850 2016 to 2018 UL, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018>.
- 2852 [72] CMS TAU Embedding Group. Tau embedded samples using 2016  
2853 data, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauTauEmbeddingSamples2016Legacy>.
- 2855 [73] CMS TAU Embedding Group. Tau embedded samples using 2017  
2856 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2017>.

- 2858 [74] CMS TAU Embedding Group. Tau embedded samples using 2018  
2859 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/>  
2860 [TauTauEmbeddingSamples2018](#).
- 2861 [75] Tau Lepton Run 2 Trigger Performance. 2019. URL <https://cds.cern.ch/record/2678958>.
- 2863 [76] CMS Taus High Level Trigger Studies. Tau Lepton Run 2 Trigger Performance  
2864 (CMS DP-2019/012), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTauAllRun2>.
- 2866 [77] Muon HLT Performance with 2018 Data. 2018. URL <https://cds.cern.ch/record/2627469>.
- 2868 [78] Single and Double Electron Trigger Efficiencies using the full Run 2 dataset.  
2869 2020. URL <https://cds.cern.ch/record/2888577>.
- 2870 [79] Run II Trigger Performance For  $e\mu$  Triggers. 2019. URL <https://cds.cern.ch/record/2687013>.
- 2872 [80] Performance of electron and photon reconstruction in Run 2 with the CMS  
2873 experiment. 2020. URL <https://cds.cern.ch/record/2725004>.
- 2874 [81] CMS Collaboration. Performance of the CMS muon detector and muon recon-  
2875 struction with proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 13(06):P06015,  
2876 2018. doi: 10.1088/1748-0221/13/06/P06015.
- 2877 [82] Piet Verwilligen. Muons in the cms high level trigger system. *Nuclear  
2878 and Particle Physics Proceedings*, 273-275:2509–2511, 2016. ISSN 2405-6014.  
2879 doi: <https://doi.org/10.1016/j.nuclphysbps.2015.09.441>. URL <https://www.sciencedirect.com/science/article/pii/S240560141500930X>. 37th Inter-  
2880 national Conference on High Energy Physics (ICHEP).

- 2882 [83] Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-  
2883 probe technique. 2020. URL <https://cds.cern.ch/record/2724492>.
- 2884 [84] V.M. Abazov, B. Abbott, M. Abolins, et al. A novel method for modeling  
2885 the recoil in W boson events at hadron colliders. *Nuclear Instruments and*  
2886 *Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors*  
2887 *and Associated Equipment*, 609(2):250–262, 2009. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2009.08.056>. URL <https://www.sciencedirect.com/science/article/pii/S0168900209016623>.
- 2890 [85] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair of  
2891 light pseudoscalars in the final state with two b quarks and two  $\tau$  leptons in  
2892 proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Phys. Lett. B*, 785:462, 2018. doi:  
2893 [10.1016/j.physletb.2018.08.057](https://doi.org/10.1016/j.physletb.2018.08.057).
- 2894 [86] CMS LUMI POG. Luminosity Physics Object Group: Recommendations, 2024.  
2895 URL <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.
- 2896 [87] The modeling of the top quark  $p_T$  (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting>.
- 2897 [88] CMS BTV group. Methods to apply b-tagging efficiency scale factors (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration>.
- 2898 [89] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in  
2899 pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. doi: [10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014).
- 2900 [90] Garvita Agarwal. Jet Energy Scale and Resolution Measurements in CMS. *PoS*,  
2901 ICHEP2022:652, 2022. doi: [10.22323/1.414.0652](https://doi.org/10.22323/1.414.0652).

- 2906 [91] CMS JERC group. Jet Energy Corrections (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- 2907
- 2908 [92] CMS JERC group. Jet Energy Resolution (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- 2909
- 2910 [93] CMS MUO POG. Muon Physics Object Group: Baseline muon selec-
- 2911 tions for Run-II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2>.
- 2912
- 2913 [94] CMS ELE POG. Electron Identification Based on Simple Cuts, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaPublicData>.
- 2914
- 2915 [95] Jose Enrique Palencia Cortezon. Single top quark production at CMS. Technical
- 2916 report, CERN, Geneva, 2018. URL <https://cds.cern.ch/record/2640578>.
- 2917 [96] CMS Collaboration. Measurements of the electroweak diboson production cross
- 2918 sections in proton-proton collisions at  $\sqrt{s} = 5.02$  TeV using leptonic decays.
- 2919 *Phys. Rev. Lett.*, 127(19):191801, 2021. doi: 10.1103/PhysRevLett.127.191801.
- 2920 [97] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment
- 2921 ten years after the discovery. *Nature*, 607(7917):60–68, 2022. doi: 10.1038/s41586-022-04892-x.
- 2922
- 2923 [98] CMS JERC group. Jet energy scale uncertainty sources (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources>.
- 2924
- 2925 [99] TOP Systematic Uncertainties (Run 2) (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics>.
- 2926
- 2927 [100] Search for a non-standard-model Higgs boson decaying to a pair of new light
- 2928 bosons in four-muon final states. Technical report, CERN, Geneva, 2013. URL <https://cds.cern.ch/record/1563546>.
- 2929

- 2930 [101] Michiel Botje et al. The PDF4LHC Working Group Interim Recommendations.  
2931 1 2011.
- 2932 [102] Cécile Caillol, Pallabi Das, Sridhara Dasu, Pieter Everaerts, Stephanie Kwan,  
2933 Isobel Ojalvo, and Ho-Fung Tsoi. CMS AN-20-213 (internal): Search for an  
2934 exotic decay of the 125 GeV Higgs boson to light pseudoscalars, with a pair of  
2935 b jets and a pair of tau leptons in the final state, 2020.
- 2936 [103] Kyle Cranmer. Practical Statistics for the LHC. In *2011 European School of*  
2937 *High-Energy Physics*, pages 267–308, 2014. doi: 10.5170/CERN-2014-003.267.
- 2938 [104] Elham Khazaie, Maryam Zeinali, Hamed Bakhshiansohi, and Abideh Jafari.  
2939 CMS AN-21-058 (internal): Search for exotic decays of the Higgs boson to a  
2940 pair of new light bosons with two muons and two b jets in the final states at  
2941  $\sqrt{s} = 13$  TeV, 2021.
- 2942 [105] CMS Higgs Physics Analysis Group. Summary of 2HDM+S searches at 13 TeV  
2943 (Run 2), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Summary2HDMRun2>.