

<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

17 Open questions in particle physics may be addressed by the existence of an extended  
18 Higgs sector beyond the Higgs boson with mass 125 GeV discovered in 2012 at the  
19 Large Hadron Collider (LHC) by the CMS and ATLAS experiments. Many properties  
20 of a potential extended Higgs sector remain unconstrained by current measurements,  
21 making direct searches of exotic Higgs decays a powerful probe of new physics. In  
22 extensions of the Standard Model of particle physics, such as Two Higgs Doublet  
23 Models extended with a singlet scalar (2HDM+S), the decay of the 125 GeV Higgs  
24 boson into light neutral scalar particles is allowed. We present a search at CMS for  
25 exotic decays of a Higgs boson with mass 125 GeV to two light neutral scalars, which  
26 respectively decay to two bottom quarks and two tau leptons (denoted  $h \rightarrow aa \rightarrow$   
27  $bb\tau\tau$ ). This analysis is combined with a different search where the light scalars decay  
28 to two bottom quarks and two muons. Results are interpreted in various 2HDM+S  
29 scenarios. In Two Real Singlet Models (TRSMs), the 125 GeV Higgs boson can decay  
30 to two light neutral scalars with unequal mass, denoted  $h \rightarrow a_1a_2$  where  $m_{a_1} \neq m_{a_2}$ .  
31 This scenario has not been searched for to date at the CMS experiment. We present  
32 ongoing work on a search for  $h \rightarrow a_1a_2$ , where the  $a_2$  decays into two  $a_1$ , resulting in  
33 four bottom quarks and two tau leptons in the final state, in the  $\mu\tau_h$  channel of the  
34  $\tau\tau$  decay.

<sup>35</sup>

## Acknowledgements

<sup>36</sup> Placeholder acknowledgements.



# <sup>38</sup> Contents

|               |   |           |
|---------------|---|-----------|
| <sup>39</sup> | Abstract . . . . .  | iii       |
| <sup>40</sup> | Acknowledgements . . . . .                                | iv        |
| <sup>41</sup> | List of Tables . . . . .                                  | xi        |
| <sup>42</sup> | List of Figures . . . . .                                 | xiii      |
| <sup>43</sup> | <b>1 Introduction</b>                                     | <b>1</b>  |
| <sup>44</sup> | 1.1 History of the Standard Model . . . . .               | 1         |
| <sup>45</sup> | 1.2 The Standard Model as a gauge theory . . . . .        | 3         |
| <sup>46</sup> | 1.2.1 Gauge invariance . . . . .                          | 3         |
| <sup>47</sup> | 1.2.2 Local gauge symmetries . . . . .                    | 4         |
| <sup>48</sup> | 1.3 The Higgs Mechanism . . . . .                         | 6         |
| <sup>49</sup> | 1.4 Two-Higgs Doublet Models . . . . .                    | 8         |
| <sup>50</sup> | 1.5 Two Real Singlet Model . . . . .                      | 11        |
| <sup>51</sup> | <b>2 The Large Hadron Collider and the CMS Experiment</b> | <b>14</b> |
| <sup>52</sup> | 2.1 The Large Hadron Collider . . . . .                   | 15        |
| <sup>53</sup> | 2.2 Luminosity and pileup . . . . .                       | 16        |
| <sup>54</sup> | 2.3 The High-Luminosity LHC . . . . .                     | 19        |
| <sup>55</sup> | 2.4 The CMS Detector . . . . .                            | 20        |
| <sup>56</sup> | 2.5 Sub-detectors of CMS . . . . .                        | 21        |
| <sup>57</sup> | 2.5.1 Inner tracking system . . . . .                     | 21        |

|    |          |  |           |
|----|----------|--|-----------|
| 58 | 2.5.2    | ECAL . . . . .   | 23        |
| 59 | 2.5.3    | HCAL . . . . .   | 24        |
| 60 | 2.5.4    | Muon detectors . . . . .   | 25        |
| 61 | 2.5.5    | The Level-1 Trigger . . . . .  | 27        |
| 62 | 2.5.6    | The High-Level Trigger . . . . .                                       | 30        |
| 63 | 2.5.7    | Particle reconstruction . . . . .                                      | 31        |
| 64 | 2.5.8    | Data storage and computational infrastructure . . . . .                | 31        |
| 65 | <b>3</b> | <b>The Phase-2 Upgrade of CMS</b>                                      | <b>32</b> |
| 66 | 3.1      | High-Luminosity LHC and CMS . . . . .                                  | 32        |
| 67 | 3.2      | The Phase-2 Level-1 Trigger . . . . .                                  | 32        |
| 68 | 3.3      | Standalone Barrel Calorimeter electron/photon reconstruction . . . . . | 35        |
| 69 | 3.3.1    | Phase-2 geometry of the ECAL Barrel trigger . . . . .                  | 35        |
| 70 | 3.3.2    | Phase-2 electron/photon reconstruction algorithm . . . . .             | 36        |
| 71 | <b>4</b> | <b>Datasets and Monte Carlo samples</b>                                | <b>45</b> |
| 72 | 4.1      | Datasets used . . . . .  | 45        |
| 73 | 4.2      | Monte Carlo samples . . . . .  | 45        |
| 74 | 4.3      | Embedded samples . . . . .   | 47        |
| 75 | <b>5</b> | <b>Object reconstruction and corrections applied</b>                   | <b>50</b> |
| 76 | 5.1      | Object reconstruction . . . . .  | 50        |
| 77 | 5.1.1    | Taus . . . . .   | 50        |
| 78 | 5.1.2    | Muons . . . . .  | 54        |
| 79 | 5.1.3    | Electrons . . . . .  | 55        |
| 80 | 5.1.4    | Jets . . . . .   | 57        |
| 81 | 5.1.5    | B-flavored jets . . . . .  | 58        |
| 82 | 5.2      | Reconstruction of the $\tau\tau$ mass . . . . .                        | 59        |
| 83 | 5.2.1    | Original SVFit “standalone”: maximum likelihood . . . . .              | 60        |

|     |   |           |
|-----|---|-----------|
| 84  | 5.2.2 “Classic SVFit” with matrix element . . . . .                             | 61        |
| 85  | 5.2.3 FastMTT: optimized SVFit . . . . .  | 62        |
| 86  | 5.3 Corrections applied to simulation . . . . .                                 | 62        |
| 87  | 5.3.1 Tau energy scale . . . . .  | 63        |
| 88  | 5.3.2 Muon energy scale . . . . .   | 63        |
| 89  | 5.3.3 Electron energy scale . . . . .   | 64        |
| 90  | 5.3.4 $\tau_h$ identification efficiency . . . . .                              | 65        |
| 91  | 5.3.5 Trigger efficiencies . . . . .  | 65        |
| 92  | 5.3.6 Tau trigger efficiencies . . . . .  | 65        |
| 93  | 5.3.7 Single muon trigger efficiencies . . . . .                                | 66        |
| 94  | 5.3.8 Single electron trigger efficiencies . . . . .                            | 67        |
| 95  | 5.3.9 $e\mu$ cross-trigger efficiencies . . . . .                               | 68        |
| 96  | 5.3.10 Electrons and muons faking $\tau_h$ : energy scales . . . . .            | 69        |
| 97  | 5.3.11 Electrons and muons faking $\tau_h$ : misidentification efficiencies . . | 70        |
| 98  | 5.3.12 Electron ID and tracking efficiency . . . . .                            | 71        |
| 99  | 5.3.13 Muon ID, isolation, and tracking efficiencies . . . . .                  | 72        |
| 100 | 5.3.14 Recoil corrections . . . . .   | 73        |
| 101 | 5.3.15 Drell-Yan corrections . . . . .  | 74        |
| 102 | 5.3.16 Pileup reweighing . . . . .  | 74        |
| 103 | 5.3.17 Pre-firing corrections . . . . .   | 75        |
| 104 | 5.3.18 Top $p_T$ spectrum reweighing . . . . .                                  | 76        |
| 105 | 5.3.19 B-tagging efficiency . . . . .   | 76        |
| 106 | 5.3.20 Jet energy resolution and jet energy smearing . . . . .                  | 76        |
| 107 | <b>6 Event selection</b>  | <b>78</b> |
| 108 | 6.1 General procedure for all channels . . . . .                                | 78        |
| 109 | 6.2 Event selection in the $\mu\tau_h$ channel . . . . .                        | 79        |
| 110 | 6.3 Event selection in the $e\tau_h$ channel . . . . .                          | 81        |

|     |          |  |            |
|-----|----------|--|------------|
| 111 | 6.4      | Event selection in the $e\mu$ channel . . . . .              | 82         |
| 112 | 6.5      | Extra lepton vetoes in all channels . . . . .                | 83         |
| 113 | <b>7</b> | <b>Background estimation</b>                                 | <b>87</b>  |
| 114 | 7.1      | Z+jets . . . . .   | 87         |
| 115 | 7.2      | W+jets . . . . .   | 88         |
| 116 | 7.3      | $t\bar{t}$ + jets . . . . .                                  | 88         |
| 117 | 7.4      | Single top . . . . .   | 89         |
| 118 | 7.5      | Diboson . . . . .  | 89         |
| 119 | 7.6      | Standard Model Higgs . . . . .                               | 89         |
| 120 | 7.7      | Jet faking $\tau_h$ . . . . .                                | 90         |
| 121 | 7.8      | QCD multijet background . . . . .                            | 91         |
| 122 | <b>8</b> | <b>Systematic uncertainties</b>                              | <b>93</b>  |
| 123 | 8.1      | Uncertainties associated with physics objects . . . . .      | 94         |
| 124 | 8.1.1    | Uncertainties in the lepton energy scales . . . . .          | 94         |
| 125 | 8.1.2    | Uncertainties from other lepton corrections . . . . .        | 95         |
| 126 | 8.1.3    | Uncertainties from jet energy scale and resolution . . . . . | 95         |
| 127 | 8.1.4    | Uncertainties from b-tagging scale factors . . . . .         | 96         |
| 128 | 8.1.5    | Uncertainties from MET . . . . .                             | 97         |
| 129 | 8.2      | Uncertainties associated with samples used . . . . .         | 97         |
| 130 | 8.3      | Other uncertainties . . . . .                                | 98         |
| 131 | 8.4      | Pulls and impacts . . . . .                                  | 98         |
| 132 | <b>9</b> | <b>Event categorization and signal extraction</b>            | <b>100</b> |
| 133 | 9.1      | B-tag jet multiplicity . . . . .                             | 100        |
| 134 | 9.2      | DNN-based event categorization . . . . .                     | 100        |
| 135 | 9.3      | Methodology for signal extraction . . . . .                  | 103        |
| 136 | 9.3.1    | Model building and parameter estimation . . . . .            | 104        |

|     |  |            |
|-----|--|------------|
| 137 | 9.3.2 Hypothesis testing . . . . .                           | 106        |
| 138 | 9.3.3 Confidence intervals . . . . .                         | 107        |
| 139 | 9.3.4 Profile likelihood ratio . . . . .                     | 109        |
| 140 | 9.3.5 Modified frequentist method: $CL_S$ . . . . .          | 110        |
| 141 | <b>10 Results</b>  | <b>111</b> |
| 142 | 10.1 Results from $bb\tau\tau$ . . . . .                     | 111        |
| 143 | 10.2 Combination with $bb\mu\mu$ final state . . . . .       | 113        |
| 144 | <b>11 Asymmetric exotic Higgs decays</b>                     | <b>124</b> |
| 145 | 11.1 Signal masses . . . . .                                 | 124        |
| 146 | 11.2 Cascade scenario signal studies . . . . .               | 125        |
| 147 | 11.3 Current control plots for $\mu\tau_h$ channel . . . . . | 127        |
| 148 | <b>12 Conclusion and outlook</b>                             | <b>131</b> |

# List of Tables

|     |     |   |    |
|-----|-----|---|----|
| 150 | 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 $m_{\mu\mu} > 70$ GeV, or a requirement on the number of b-tag jets in the event. . . . . | 49 |
| 151 |     |   |    |
| 152 | 5.1 | Energy scales applied to genuine hadronic tau decays $\tau_h$ by data-taking year/era and decay mode, along with systematic errors. . . . .   | 63 |
| 153 |     |   |    |
| 154 | 5.2 | Energy scales and systematic errors applied to genuine muons. . . . .   | 64 |
| 155 |     |   |    |
| 156 | 5.3 | Energy scales and systematic errors applied to electrons in embedded samples by data-taking year/era. . . . .   | 65 |
| 157 |     |   |    |
| 158 | 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 $p_T$ . . . . .   | 65 |
| 159 |     |   |    |
| 160 | 5.5 | Energy scales and up/down systematic uncertainties applied to electrons misidentified as hadronic taus. . . . .   | 69 |
| 161 |     |   |    |
| 162 | 5.6 | Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018. . . . .   | 71 |
| 163 |     |   |    |
| 164 | 5.7 | Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018. . . . .   | 71 |
| 165 |     |   |    |
| 166 |     |   |    |

|     |     |   |     |
|-----|-----|---|-----|
| 167 | 6.1 | Trigger thresholds used for the leptons in the $bb\mu\mu$ analysis and the                  |     |
| 168 |     | $bb\tau\tau$ analysis (the focus of this work). The thresholds for the three $bb\tau\tau$   |     |
| 169 |     | channels ( $e\mu$ , $e\tau_h$ , and $\mu\tau_h$ ) are listed separately, with some channels |     |
| 170 |     | and years taking the logical OR of two triggers with different thresholds.                  | 79  |
| 171 | 6.2 | High-Level Trigger (HLT) paths used to select data and simulation                           |     |
| 172 |     | events in 2016 for the three $\tau\tau$ channels. . . . .                                   | 84  |
| 173 | 6.3 | High-Level Trigger (HLT) paths used to select data and simulation                           |     |
| 174 |     | events in 2017 for the three $\tau\tau$ channels. . . . .                                   | 84  |
| 175 | 6.4 | High-Level Trigger (HLT) paths used to select data and simulation                           |     |
| 176 |     | events in 2018 for the three $\tau\tau$ channels. In 2018 a HLT trigger path                |     |
| 177 |     | using the hadron plus strips (HPS) tau reconstruction algorithm be-                         |     |
| 178 |     | came available. . . . .   | 85  |
| 179 | 9.1 | Event categorization based on DNN scores for events with exactly 1                          |     |
| 180 |     | b-tag jet (1bNN), for the three $\tau\tau$ channels and three eras. . . . .                 | 103 |
| 181 | 9.2 | Event categorization based on DNN scores for events with 2 b-tag jets                       |     |
| 182 |     | (2bNN), for the three $\tau\tau$ channels and three eras. . . . .                           | 104 |

# <sup>183</sup> List of Figures

|                    |  |    |
|--------------------|--|----|
| <sup>184</sup> 1.1 | Table of Standard Model particles showing the grouping of the fermions<br>into three generations of matter and the bosons, responsible for carrying<br>the three fundamental forces in the Standard Model. The masses,<br>charges, and spins of the particles are shown. The antimatter counter-<br>parts of the fermions are not shown. The possible interactions between<br>the fermions and gauge bosons are highlighted. . . . . | 2  |
| <sup>185</sup> 1.2 | An illustration of the Higgs potential. . . . .  | 7  |
| <sup>186</sup> 1.3 | Branching ratios of a singlet-like pseudoscalar in Type II 2HDM+S for<br>$\tan \beta = 0.5$ (left) and $\tan \beta = 5$ (right). . . . .   | 10 |
| <sup>187</sup> 1.4 | Benchmark plane BP1 for benchmark scenario 1, for the decay signa-<br>ture $h_{125} \rightarrow h_1 h_2$ with $h_{125} \equiv h_3$ , defined in the $(M_1, M_2)$ plane. . . .  | 13 |
| <sup>188</sup> 2.1 | Aerial view of the Large Hadron Collider (LHC). . . . .  | 17 |
| <sup>189</sup> 2.2 | Distribution of the mean number of inelastic collisions per bunch cross-<br>ing (pileup) in data, for proton-proton collisions in 2016-2018 . . . .  | 19 |
| <sup>190</sup> 2.3 | Sketch of particle trajectories of muons, electrons, charged and neutral<br>hadrons, and photons in a transverse cross-section of the CMS detector.  | 21 |
| <sup>191</sup> 2.4 | Cross section of the current Phase-1 CMS tracker. . . . .  | 22 |
| <sup>192</sup> 2.5 | Longitudinal view of the CMS detector showing the hadron calorimeter<br>barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters.   | 24 |

|     |     |  |    |
|-----|-----|--|----|
| 203 | 2.6 | Layout of the CMS barrel muon drift tube (DT) chambers in one of<br>the five wheels. . . . .   | 26 |
| 204 | 2.7 | Dataflow for the Phase-1 Level-1 Trigger. . . . .  | 28 |
| 205 | 3.1 | Functional diagram of the CMS L1 Phase-2 upgraded trigger design. .  | 33 |
| 206 | 3.2 | Summary of the links between the trigger primitives, the trigger ob-<br>jects, the Level-1 algorithms, and the physics channels in the Phase-2<br>menu. . . . .  | 36 |
| 207 | 3.3 | Schematic of the geometry of the Phase-2 ECAL barrel in the Regional<br>Calorimeter Trigger (RCT), showing the division of the barrel region<br>into 36 Regional Calorimeter Trigger (RCT) cards ( <i>red</i> ). Each card<br>spans $17 \times 4$ towers in $\eta \times \phi$ ( <i>green</i> ), and each tower is $5 \times 5$ in single<br>crystals in $\eta \times \phi$ . Towers in the overlap region ( <i>shaded yellow</i> ) are read<br>out to both the barrel and endcap. . . . . | 39 |
| 208 | 3.4 | Schematic of two example RCT cards in the negative eta ( <i>top</i> ) and<br>positive eta ( <i>center</i> ) regions of the ECAL barrel. Each RCT card is<br>divided into five regions: four regions are of size $3 \times 4$ towers in $\eta \times \phi$<br>( <i>bottom left</i> ), and a fifth smaller overlap region of size $2 \times 4$ towers<br>( <i>bottom right</i> ). Each tower is $5 \times 5$ ( $\eta \times \phi$ ) in crystals. . . . .                                     | 40 |
| 209 | 3.5 | Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trig-<br>ger (GCT), which will process the outputs of the Regional Calorimeter<br>Trigger (RCT) in three cards ( <i>magenta highlights</i> ). Each card in the<br>GCT processes the equivalent of sixteen RCT cards, with the center<br>twelve being unique to that GCT card ( <i>shaded pink</i> ), and the remain-<br>ing four processed in overlap with the other GCT cards. . . . .                                     | 41 |

|     |   |    |
|-----|---|----|
| 227 | 3.6 Illustration of an example electron/photon ( $e/\gamma$ ) cluster in the Phase-             |    |
| 228 | 2 Level-1 Trigger standalone barrel $e/\gamma$ reconstruction, in a region of                   |    |
| 229 | $15 \times 20$ crystals ( $3 \times 4$ towers). Each small pink square is one crystal,          |    |
| 230 | the highest-granularity ECAL trigger primitives available to the L1                             |    |
| 231 | Trigger in Phase-2. The core cluster consists of the energy sum in a                            |    |
| 232 | $3 \times 5$ window of crystals, ( <i>shaded light blue</i> ) centered around the seed          |    |
| 233 | crystal ( <i>red</i> ). Bremsstrahlung corrections are checked in the adjacent                  |    |
| 234 | $3 \times 5$ windows in the $\phi$ direction ( <i>shaded light yellow</i> ). The relative       |    |
| 235 | energies in windows of size $2 \times 5$ and $5 \times 5$ in crystals ( <i>dashed dark blue</i> |    |
| 236 | <i>and dark red</i> ) are used to compute shower shape variables to identify                    |    |
| 237 | true $e/\gamma$ objects. Lastly, an isolation sum is computed in a window of                    |    |
| 238 | size $7 \times 7$ in towers (not shown in figure). . . . .                                      | 42 |
| 239 | 3.7 Efficiency of the standalone barrel $e/\gamma$ reconstruction, as a function of             |    |
| 240 | the true electron’s transverse momentum $p_T$ . . . . .   | 43 |
| 241 | 3.8 Rates of the standalone barrel $e/\gamma$ reconstruction measured as a func-                |    |
| 242 | tion of the minimum energy ( $E_T$ ) required of the reconstructed $e/\gamma$                   |    |
| 243 | object in each event. . . . .   | 44 |
| 244 | 4.1 Cumulative delivered and recorded luminosity versus time for 2015-                          |    |
| 245 | 2018 at CMS, in proton-proton collision data only, at nominal center-                           |    |
| 246 | of-mass energy. . . . .   | 46 |
| 247 | 4.2 Schematic view of the four main steps of the embedding technique for                        |    |
| 248 | $\tau$ leptons. . . . .   | 48 |
| 249 | 5.1 Distributions of $m_{\tau\tau}$ reconstructed by the classic SVFit algorithm, and           |    |
| 250 | masses of visible tau decay products (before SVFit). . . . .                                    | 61 |
| 251 | 5.2 Electron/photon energy scale factors and uncertainties for 2018. . . .                      | 64 |

|     |      |   |    |
|-----|------|---|----|
| 252 | 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.  | 67 |
| 253 |      |   |    |
| 254 | 5.4  | Trigger efficiencies in data ( <i>top panels</i> ) and ratio of efficiencies af-<br>ter/before a HLT muon reconstruction update ( <i>bottom panels</i> ) for the<br>muon in the isolated single muon trigger with threshold $p_T > 24$ GeV<br>in the data-taking year 2018, as functions of the muon $p_T$ ( <i>left</i> ) and<br>muon $ \eta $ ( <i>right</i> ). . . . . | 68 |
| 255 |      |   |    |
| 256 |      |   |    |
| 257 |      |   |    |
| 258 |      |   |    |
| 259 | 5.5  | Trigger efficiencies in data and the data/MC ratio for the electron in<br>the single electron trigger with threshold $p_T > 32$ GeV in the data-<br>taking year 2018, as functions of the electron $p_T$ ( <i>left</i> ) and electron $ \eta $<br>( <i>right</i> ). . . . .   | 69 |
| 260 |      |   |    |
| 261 |      |   |    |
| 262 |      |   |    |
| 263 | 5.6  | Efficiencies of the electron leg vs. $p_T$ ( <i>left</i> ) and the muon log vs. $\eta$<br>( <i>right</i> ), for the HLT path with online thresholds of 12 GeV for the<br>electron and 23 GeV for the muon, with the data-taking years 2016<br>through 2018 overlaid. . . . .  | 70 |
| 264 |      |   |    |
| 265 |      |   |    |
| 266 |      |   |    |
| 267 | 5.7  | Efficiencies in data ( <i>top panels</i> ) and the ratio of efficiencies in data/MC<br>( <i>bottom panels</i> ), for the electron multivariate analysis (MVA) identifi-<br>cation ( <i>left</i> ) and for the Gaussian-sum filter (GSF) tracking ( <i>right</i> ). . .  | 72 |
| 268 |      |   |    |
| 269 |      |   |    |
| 270 | 5.8  | Muon identification efficiencies in 2015 data and MC as a function of<br>the muon $p_T$ for the loose ID ( <i>left</i> ) and tight ID ( <i>right</i> ) working points.  | 73 |
| 271 |      |   |    |
| 272 | 5.9  | Muon isolation efficiencies in Run-2 data as a function of the muon $p_T$<br>( <i>left</i> ) and $ \eta $ ( <i>right</i> ). . . . .   | 74 |
| 273 |      |   |    |
| 274 | 5.10 | Muon tracking efficiencies as a function of $ \eta $ for standalone muons in<br>Run-2 data ( <i>black</i> ) and Drell-Yan ( <i>blue</i> ) MC simulation. . . . .  | 75 |
| 275 |      |   |    |
| 276 | 7.1  | Leading-order Feynman diagrams of Higgs production. . . . .   | 90 |
| 277 | 8.1  | Top sixty impacts for the combination of all channels and years. . . .  | 99 |

|     |   |     |
|-----|---|-----|
| 278 | 9.1 Schematic of the Neyman construction for confidence intervals. . . . .                                  | 108 |
| 279 | 10.1 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $\mu\tau_h$ chan-              |     |
| 280 | nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions. . . . .                                | 115 |
| 281 | 10.2 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\tau_h$ chan-                |     |
| 282 | nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions. . . . .                                | 116 |
| 283 | 10.3 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\mu$ channel.                | 117 |
| 284 | 10.4 Observed 95% CL exclusion limits ( <i>black, solid lines</i> ) and expected 95%                        |     |
| 285 | CL and 68% CL limits ( <i>shaded yellow and green</i> ) on the branching                                    |     |
| 286 | fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard                 |     |
| 287 | Model production for the 125 GeV Higgs ( $h$ ). Limits are shown for the                                    |     |
| 288 | $\mu\tau_h$ channel ( <i>top left</i> ), the $e\tau_h$ channel ( <i>top right</i> ), and the $e\mu$ channel |     |
| 289 | ( <i>bottom left</i> ), and lastly the combination of all three channels ( <i>bottom</i>                    |     |
| 290 | <i>right</i> ) The dataset corresponds to $138 \text{ fb}^{-1}$ of data collected in the                    |     |
| 291 | years 2016-2018 at a center-of-mass energy 13 TeV. . . . . . . . . . . . . . . .                            | 118 |
| 292 | 10.5 Observed 95% CL upper limits on $B(h \rightarrow aa)$ in %, for the $bb\tau\tau$ final                 |     |
| 293 | state ( <i>left</i> ) and $bb\mu\mu$ final state ( <i>right</i> ) using the full Run 2 integrated           |     |
| 294 | luminosity of $138 \text{ fb}^{-1}$ in 2HDM+S type I ( <i>blue</i> ), type II with $\tan\beta =$            |     |
| 295 | $2.0$ ( <i>orange dashed</i> ), type III with $\tan\beta = 2.0$ ( <i>dotted green</i> ), and type           |     |
| 296 | IV with $\tan\beta = 0.6$ ( <i>red dashed</i> ).            | 119 |
| 297 | 10.6 Observed 95% CL upper limits on the branching fraction of the 125                                      |     |
| 298 | GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$ , in percentages,                               |     |
| 299 | as a function of the pseudoscalar mass $m_a$ , in 2HDM+S type I ( <i>blue</i> ),                            |     |
| 300 | type II with $\tan\beta = 2.0$ ( <i>orange dashed</i> ), type III with $\tan\beta = 2.0$                    |     |
| 301 | ( <i>dotted green</i> ), and type IV with $\tan\beta = 0.6$ ( <i>red dashed</i> ), for the                  |     |
| 302 | combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated                         |     |
| 303 | luminosity of $138 \text{ fb}^{-1}$ .         | 120 |

|     |  |     |
|-----|--|-----|
| 304 | 10.7 Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination                   |     |
| 305 | of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity                               |     |
| 306 | of $138 \text{ fb}^{-1}$ for Type II ( <i>left</i> ), Type III ( <i>middle</i> ), and Type IV ( <i>right</i> )   |     |
| 307 | 2HDM+S in the $\tan\beta$ vs. $m_a$ phase space. . . . .   | 121 |
| 308 | 10.8 Summary plot of current observed and expected 95% CL limits on the  |     |
| 309 | branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-  |     |
| 310 | malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$      |     |
| 311 | $B(h \rightarrow aa)$ , in the 2HDM+S type I scenario, obtained at CMS with                                      |     |
| 312 | data collected at 13 TeV. . . . .  | 121 |
| 313 | 10.9 Summary plot of current observed and expected 95% CL limits on the  |     |
| 314 | branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-  |     |
| 315 | malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$      |     |
| 316 | $B(h \rightarrow aa)$ , in the 2HDM+S type II scenario with $\tan\beta = 2.0$ , ob-                              |     |
| 317 | tained at CMS with data collected at 13 TeV. . . . .   | 122 |
| 318 | 10.10 Summary plot of current observed and expected 95% CL limits on the   |     |
| 319 | branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-  |     |
| 320 | malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$      |     |
| 321 | $B(h \rightarrow aa)$ , in the 2HDM+S type III scenario with $\tan\beta = 2.0$ , ob-                             |     |
| 322 | tained at CMS with data collected at 13 TeV. . . . .   | 123 |
| 323 | 11.1 Generator-level b-flavor jet transverse momenta $p_T$ , for $h \rightarrow a_1 a_2$ cas-                    |     |
| 324 | cade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) =$                           |     |
| 325 | $(100, 15) \text{ GeV}$ ( <i>left</i> ) and $(40, 20) \text{ GeV}$ ( <i>right</i> ). In each plot the generator- |     |
| 326 | level $p_T$ of the leading ( <i>black</i> ), sub-leading ( <i>red</i> ), third ( <i>blue</i> ), and              |     |
| 327 | fourth ( <i>light green</i> ) are overlaid. . . . .  | 126 |

|     |   |     |
|-----|---|-----|
| 328 | 11.2 Distributions (arbitrary units) of transverse momentum $p_T$ resolution  |     |
| 329 | and $\Delta R$ between the two closest generator-level $b$ jets, treated as one   |     |
| 330 | object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow$  |     |
| 331 | $a_1 a_2$ mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV ( <i>top left, top right</i> )                             |     |
| 332 | and $(40, 20)$ GeV ( <i>bottom left, bottom right</i> ) in the ggH production of  |     |
| 333 | the 125 GeV $h$ . In the $(40, 20)$ GeV mass point, the longer $p_T$ resolution   |     |
| 334 | tail ( <i>bottom left</i> ) indicates that the reconstructed jet underestimates   |     |
| 335 | the generator b-flavor jets' energy, and the significant fraction of events   |     |
| 336 | with larger $\Delta R$ values ( <i>bottom right</i> ) indicate worse matching. . . . .                                    | 127 |
| 337 | 11.3 Kinematic properties of the leading muon and $\tau_h$ in the $\mu\tau_h$ channel: $p_T$                              |     |
| 338 | ( <i>top row</i> ), $\eta$ ( <i>second row</i> ), and $\phi$ ( <i>third row</i> ). The visible 4-momenta                  |     |
| 339 | of the muon and $\tau_h$ are summed, giving the visible di-tau mass $m_{\text{vis}}$                                      |     |
| 340 | and transverse momentum $p_{T,\text{vis}}$ . The errors shown in the figures only   |     |
| 341 | include statistical errors and only several of the full set of systematic   |     |
| 342 | errors (only those associated with the lepton energy scales and $\tau_h$ iden-  |     |
| 343 | tification efficiency). . . . .   | 129 |
| 344 | 11.4 Kinematic properties of the leading and sub-leading b-tag jets in the  |     |
| 345 | $\mu\tau_h$ final state: jet $p_T$ ( <i>top row</i> ), $\eta$ ( <i>second row</i> ), $\phi$ ( <i>third row</i> ), as well |     |
| 346 | as the missing transverse energy magnitude and azimuthal direction  |     |
| 347 | ( <i>bottom row</i> ). The errors shown in the figures only include statistical   |     |
| 348 | errors and only several of the full set of systematic errors (only those  |     |
| 349 | associated with the lepton energy scales and $\tau_h$ identification efficiency).   | 130 |

# <sup>350</sup> Chapter 1

## <sup>351</sup> Introduction

<sup>352</sup> The Standard Model is the current prevailing theoretical framework that encompasses  
<sup>353</sup> all known elementary particles to date and describes their interactions, yet falls short  
<sup>354</sup> of describing open problems in physics. Here, we introduce the Standard Model (Sec-  
<sup>355</sup> tion 1.1) and provide a mathematical motivation of the SM a gauge theory (Section  
<sup>356</sup> 1.2). We introduce the Higgs mechanism (Section 1.3), and outline two groups of  
<sup>357</sup> theoretical extensions to the Standard Model that feature extended Higgs sectors  
<sup>358</sup> (Sections 1.4 and 1.5).

### <sup>359</sup> 1.1 History of the Standard Model

<sup>360</sup> The building blocks of our modern-day understanding of particle physics were estab-  
<sup>361</sup> lished over the course of decades by experimental discoveries and theoretical advances,  
<sup>362</sup> culminating in the development of a theoretical framework known as the Standard  
<sup>363</sup> Model (SM). In the 1880s, the electron was the first subatomic particle to be iden-  
<sup>364</sup> tified, through measurements of particles produced by ionizing gas. By the 1930s,  
<sup>365</sup> atoms were known to consist mostly of empty space, with protons and neutrons con-  
<sup>366</sup> centrated at the center and orbited by electrons. Spurred by advances in particle  
<sup>367</sup> accelerator technology, the experimental discoveries of the positron, the muon, and

368 the pion, painted an increasingly complicated picture of particle physics that could  
 369 not be described solely with atomic physics [1].

370 In the absence of a theoretical framework describing these particles, in the 1960s  
 371 and 1970s physicists and mathematicians developed the Standard Model to describe  
 372 and encompass these fundamental particles and the forces that govern their interac-  
 373 tions. The particle content of the Standard Model is shown in Fig. 1.1: they are  
 374 grouped into fermions, which comprise all known matter, and bosons, which mediate  
 375 the interactions between particles.

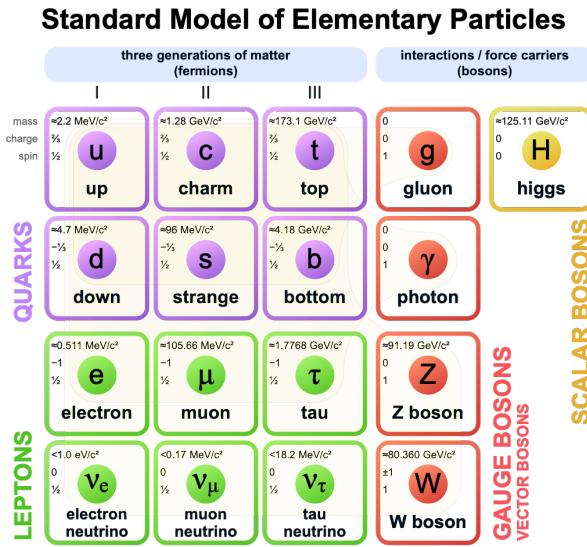


Figure 1.1: Table of Standard Model particles showing the grouping of the fermions into three generations of matter and the bosons, responsible for carrying the three fundamental forces in the Standard Model. The masses, charges, and spins of the particles are shown. The antimatter counterparts of the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

376 Fermions consist of quarks and leptons, and are grouped into three generations.  
 377 For example, the electron belongs to the first generation of leptons. The second and  
 378 third generation counterparts of the electron are the muon and the tau lepton, and  
 379 are over 200 and 30,000 times heavier than the electron respectively. Bosons are force  
 380 carriers; the interaction of fermions with bosons corresponds to fundamental forces.  
 381 The Standard Model describes the electromagnetic force, the strong nuclear force,

382 and the weak nuclear force.

## 383 1.2 The Standard Model as a gauge theory

### 384 1.2.1 Gauge invariance

385 Gauge theories of elementary particle interactions originate from a freedom of choice  
386 in the mathematical description of particle fields which has no effect on the particles'  
387 physical states [2]. The existence and form of the particles' interactions, can be  
388 deduced from the existence of physically indeterminate, gaugable quantities.

389 An example of this gauge invariance is classical physics is the electromagnetic  
390 interaction, where the fundamental field is the four-vector potential  $A^\mu$  [2]. The  
391 physical electromagnetic fields and Maxwell's equations arise from the elements of  
392 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  
393 transformation of the form

$$A_\mu \rightarrow A_\mu + \partial_\mu \alpha \tag{1.1}$$

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

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

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

402 In classical mechanics, the conservation of linear momentum, angular momentum,  
 403 and energy follows from translational invariance, rotational variance, and invariance  
 404 under translations in time [2]. Likewise, charge conservation can be shown to arise  
 405 from the invariance of the Dirac Lagrangian density  $\mathcal{L}_{\text{Dirac}} = \bar{\Psi}(i\gamma^\mu \partial_\mu - m)\Psi$  under the  
 406 particle wavefunction's phase transformation,  $\Psi'(x) = \exp(ie\chi)\Psi(x)$ . Thus Noether's  
 407 theorem establishes a correspondence between a gauge symmetry and a conserved  
 408 internal property (e.g. charge or momentum).

### 409 1.2.2 Local gauge symmetries

410 Interactions between particles arise if we modify the wave function with a phase  
 411 transformation  $\Psi'(x) = \exp(ie\chi)\Psi(x)$ , and allow the phase  $\chi$  to be a function of  
 412 spacetime [2]. A wave function of the form

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

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

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

417 We define a covariant derivative,

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

<sup>418</sup> where  $A_\mu(x)$  is a 4-vector potential. Thus the modified Dirac equation reads:

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

<sup>419</sup> The simultaneous gauge transformation  $A'_\mu(x) = A_\mu(x) - \partial_\mu \chi(x)$  and wavefunction  
<sup>420</sup> transformation  $\Psi'(x) = \exp(ie\chi(x))\Psi(x)$  leaves the covariant-derivative form of the  
<sup>421</sup> Dirac equation (Eqn 1.1) invariant.

<sup>422</sup> The generalization of this result is as follows: if a theory is invariant for unitary  
<sup>423</sup> transformations  $U$  of the particle states according to

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

<sup>424</sup> One must define a derivative of the form

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

<sup>425</sup> to keep the theory invariant under Eqn. 1.7. The four-potential  $B^\mu$  represents the  
<sup>426</sup> interacting four-potential which must be added to keep the theory invariant.

<sup>427</sup> In the case of the Standard Model, the theory is built around the gauge trans-  
<sup>428</sup> formations  $G = SU(3) \times SU(2) \times U(1)$ .  $SU(3)$  is associated to the strong force  
<sup>429</sup> (subscripted  $C$ );  $SU(2)$  is associated to the weak force (subscripted  $L$ ); and  $U(1)$  is  
<sup>430</sup> hypercharge (subscripted  $Y$ ). The gauge-covariant derivative is

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

<sup>431</sup> • In the  $U(1)_Y$  term,  $B_\mu$  is the weak hypercharge field.

<sup>432</sup> • 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-  
<sup>433</sup> potentials.  $\tau/2$  are the Pauli matrices, generators of the  $SU(2)$  transformation.

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

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

### 438        1.3 The Higgs Mechanism

439        To introduce mass into the theory, i.e. to change the propagation of the gauge par-  
440            ticles and all the fermions, the physical vacuum cannot have all the symmetries of  
441            the Standard Model Lagrangian [2]. The symmetries of the physical vacuum must  
442            be spontaneously broken, without affecting gauge invariance in the Lagrangian. The  
443            Higgs mechanism proposes the existence of a scalar field, or fields, with nonzero vac-  
444            uum expectation values, which reduce the gauge symmetries of the physical vacuum  
445            from  $SU(3)_C \times SU(2)_L \times U(1)_Y$  down to  $SU(3)_C \times U(1)_{EM}$ .

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

454        The minimal choice of Higgs field comes from the breaking of  $SU(2)_L \times U(1)_Y$   
455            down to  $U(1)_{EM}$ . The smallest  $SU(2)$  multiplet is the doublet. The existence of three  
456            massive electroweak bosons leads the Higgs sector to have at least three degrees of

457 freedom. The minimal single-doublet complex scalar Higgs field is

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

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

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

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

462 where  $\lambda$  is the coupling strength of the four-point Higgs interaction. The potential  
463 energy is minimized at

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

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

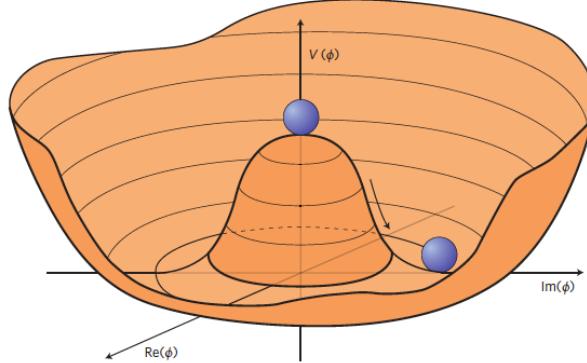


Figure 1.2: An illustration of the Higgs potential [3]. Choosing any of the points at the bottom of the potential breaks spontaneously the rotational  $U(1)$  symmetry.

<sup>466</sup> The excitations of the Higgs field with respect to the minimum  $\Phi_{\min}$  are parameterized by  
<sup>467</sup>

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

<sup>468</sup> Three degrees of freedom are coupled directly to the electroweak gauge bosons; this  
<sup>469</sup> is often referred to as the gauge bosons “eating” the Goldstone bosons to form the  
<sup>470</sup> longitudinal polarizations of the massive spin-1 boson states. The  $H(x)$  excitation is  
<sup>471</sup> in the radial direction and corresponds to the free particle state of the Higgs boson.

## <sup>472</sup> 1.4 Two-Higgs Doublet Models

<sup>473</sup> One of the simplest possible extensions to the Standard Model is adding a doublet  
<sup>474</sup> to the minimal Higgs sector of the Standard Model, which is a  $SU(2)_L$  doublet  $H$   
<sup>475</sup> with hypercharge  $Y = +\frac{1}{2}$ , denoted here as  $H \sim 2_{+1/2}$ . These extensions are found  
<sup>476</sup> in several theories such as supersymmetry. A general 2HDM can be extended with a  
<sup>477</sup> light scalar (2HDM+S) to obtain a rich set of exotic Higgs decays [4].

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 [4]. Expanding about the minima yields two complex and four real degrees of freedom:

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

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

<sup>478</sup> The charged scalar and pseudoscalar mass matrices are diagonalized by a rotation  
<sup>479</sup> angle  $\beta$ , defined as  $\tan \beta = v_2/v_1$ . One charged (complex) field and one neutral  
<sup>480</sup> pseudoscalar combination of  $H_{1,2,I}^0$  are eaten by the SM gauge bosons after electroweak

481 symmetry breaking [4]. The other complex field yields two charged mass eigenstates  
 482  $H^\pm$ , which are assumed to be heavy. The remaining three degrees of freedom yield  
 483 one neutral pseudoscalar mass eigenstate

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

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

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

485 We assume that the 2HDM is near or in the decoupling limit:  $\alpha \rightarrow \pi/2 - \beta$ , where  
 486 the lightest state in the 2HDM is  $h$ , which we identify as the 125 GeV Higgs particle  
 487 [4]. In this limit, the fermion couplings of  $h$  become identical to the Standard Model  
 488 Higgs, while the gauge boson couplings are very close to Standard Model-like for  
 489  $\tan \beta \gtrsim 5$ . All of the properties of  $h$  are determined by just two parameters:  $\tan \beta$   
 490 and  $\alpha$ , and the fermion couplings to the two Higgs doublets.

491 2HDM can be extended by a scalar singlet (2HDM+S) [4]:

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

492 If this singlet only couples to the Higgs doublets  $H_{1,2}$  and has no direct Yukawa  
 493 couplings, all of its couplings to SM fermions result from mixing with  $H_{1,2}$ . Under  
 494 these simple assumptions, exotic Higgs decays  $h \rightarrow ss \rightarrow X\bar{X}Y\bar{Y}$  or  $h \rightarrow aa \rightarrow$   
 495  $X\bar{X}Y\bar{Y}$ , and  $h \rightarrow aZ \rightarrow X\bar{X}Y\bar{Y}$  are permitted, where  $s(a)$  is a (pseudo)scalar mass  
 496 eigenstate mostly composed of  $S_R(S_I)$ , and  $X, Y$  are Standard Model fermions or  
 497 gauge bosons. There are two pseudoscalars in the 2HDM+S, and the mostly singlet-  
 498 like pseudoscalar can be chosen to be the one lighter than the SM-like Higgs. For

499     $m_a < m_h - m_Z \sim 35$  GeV, the exotic Higgs decay  $h \rightarrow Za$  is possible, and for  
500     $m_a < m_h/2 \approx 63$  GeV, the exotic Higgs decay  $h \rightarrow aa$  is possible.

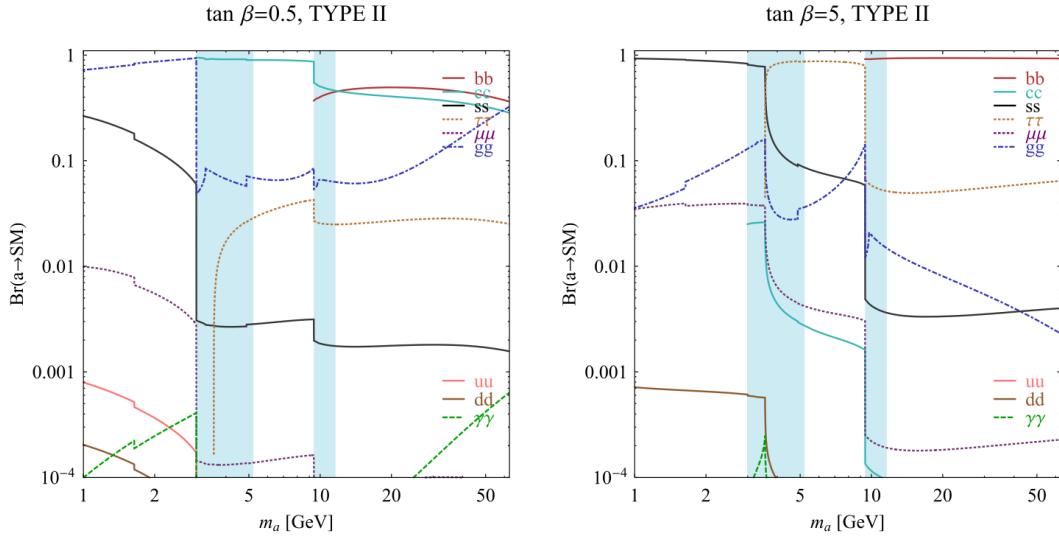


FIG. 7 (color online). Branching ratios of a singletlike pseudoscalar in the 2HDM + S for type-II Yukawa couplings. Decays to quarkonia likely invalidate our simple calculations in the shaded regions.

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 [4], 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.

501    In 2HDM, and by extension 2HDM+S, there are four types of fermion couplings  
502    commonly discussed in the literature that forbid flavor-changing neutral currents at  
503    tree level [4]. These are referred to as Type I (all fermions couple to  $H_2$ ), Type II  
504    (MSSM-like,  $d_R$  and  $e_R$  couple to  $H_1$ ,  $u_R$  to  $H_2$ ), Type III (lepton-specific, leptons  
505    and quarks couple to  $H_1$  and  $H_2$  respectively) and Type IV (flipped, with  $u_R$ ,  $e_R$   
506    coupling to  $H_2$  and  $d_R$  to  $H_1$ ). The exact branching ratios of the pseudoscalars to  
507    Standard Model particles vary depending on the 2HDM+S model and the value of  
508     $\tan\beta$  (e.g. Fig. 1.3).

## 509 1.5 Two Real Singlet Model

510 The two real singlet model (TRSM) adds two real singlet degrees of freedom to the  
 511 Standard Model. These are written as two real singlet fields  $S$  and  $X$ . Depending  
 512 on the vacuum expectation values acquired by the scalars, different phases of the  
 513 model can be realized [5]. To reduce the number of free parameters, two discrete  $\mathbb{Z}_2$   
 514 symmetries are introduced. The fields are decomposed as

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

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

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

524 After fixing one of the Higgs masses to the mass of the observed Higgs boson, and  
 525 fixing the Higgs doublet vacuum expectation value to its Standard Model value, there  
 526 are seven remaining free parameters of the TRSM [5].

527 In one benchmark scenario of TRSM [5], the heaviest scalar state  $h_3$  is identified  
 528 with the 125 GeV Higgs,  $h_{125}$ , and it can decay asymmetrically  $h_{125} \rightarrow h_1 h_2$ , which  
 529 we also denote  $h \rightarrow a_1 a_2$  to highlight the similarity with the symmetric decay  $h \rightarrow aa$   
 530 typically interpreted in 2HDM+S as discussed. The parameter values in TRSM are  
 531 chosen such that the coupling of  $h_3$  to Standard Model particles are nearly identical  
 532 to the Standard Model predictions.

533 In benchmark scenario 1 (benchmark plane 1, or BP1) (Fig. 1.4) [5], the maximal  
534 branching ratios for  $h_3 \rightarrow h_1 h_2$  reach up to 7 – 8% which translates into a signal  
535 rate of around 3 pb. These maximal branching ratios are reached in the intermediate  
536 mass state for  $h_2$ ,  $M_2 \sim 60 – 80$  GeV. For  $M_2 < 40$  GeV, although phase space opens  
537 up significantly for light decay products, the branching ratio becomes smaller.

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

543 If the  $h_2 \rightarrow h_1 h_1$  decay is kinematically closed (i.e.  $M_2 < 2M_1$ ), both scalars decay  
544 directly to Standard Model particles, with branching ratios identical to a Standard  
545 Model-like Higgs boson, i.e. with the  $b\bar{b} b\bar{b}$  final state dominating, as shown in Fig. 1.4  
546 (*bottom left*), while at smaller masses, combinations with  $\tau$  leptons and eventually  
547 final states with charm quarks and muons become relevant [5].

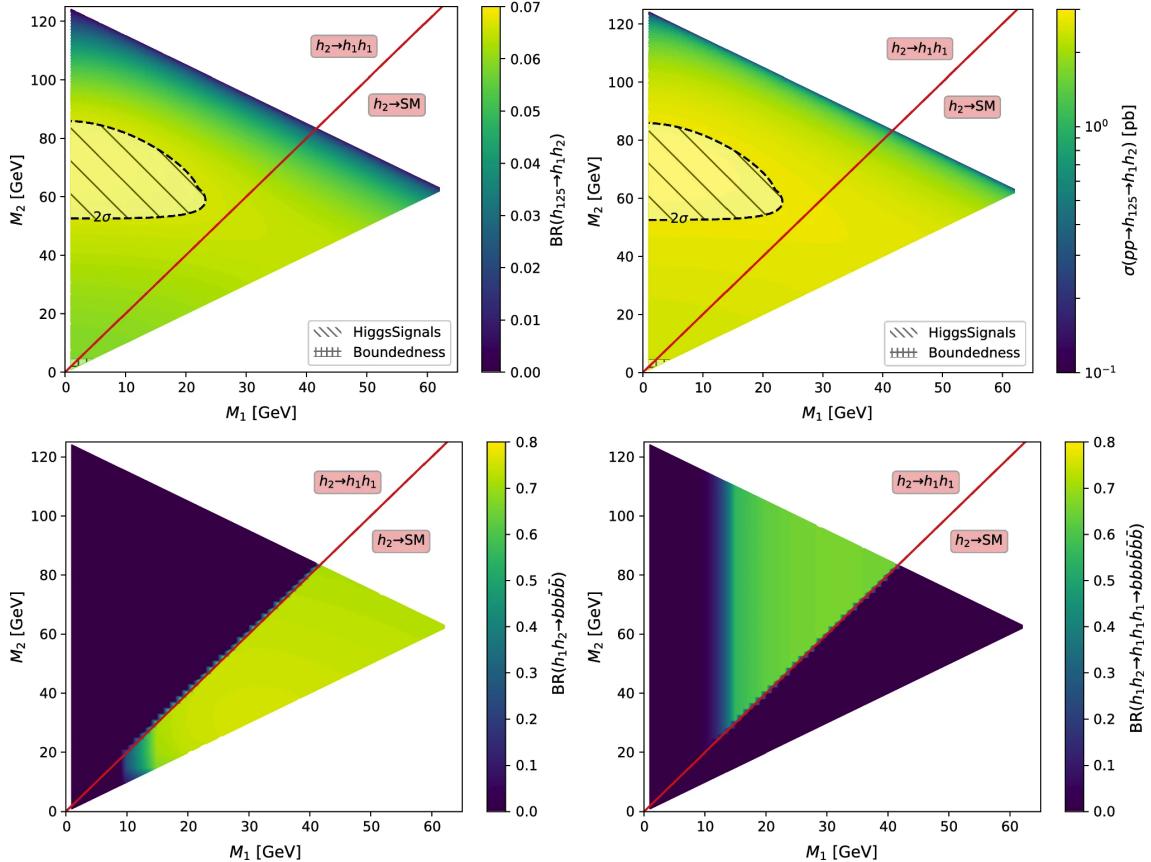


Figure 1.4: Benchmark plane BP1 for benchmark scenario 1 from [5], for the decay signature  $h_{125} \rightarrow h_1 h_2$  with  $h_{125} \equiv h_3$ , defined in the  $(M_1, M_2)$  plane. The color code shows  $\text{BR}(h_3 \rightarrow h_1 h_2)$  (*top left*) and the 13 TeV LHC signal rate for  $pp \rightarrow h_3 \rightarrow h_1 h_2$  (*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.

<sup>548</sup> **Chapter 2**

<sup>549</sup> **The Large Hadron Collider and the**  
<sup>550</sup> **CMS Experiment**

<sup>551</sup> This chapter introduces the key aspects of the CERN Large Hadron Collider (LHC)  
<sup>552</sup> and the Compact Muon Solenoid (CMS) experiment where the work for this thesis was  
<sup>553</sup> conducted. Section 2.1 describes the history of accelerator developments at CERN  
<sup>554</sup> that led to the construction of the LHC, the current LHC configuration, and the  
<sup>555</sup> largest experiments located at the LHC. The concepts of beam luminosity and pileup,  
<sup>556</sup> which are critical for understanding and measuring high-energy particle collisions,  
<sup>557</sup> are described in Section 2.2 and discussed in the context of the High-Luminosity  
<sup>558</sup> LHC (HL-LHC) upgrade in Section 2.3. Lastly, Section 2.4 describes the design  
<sup>559</sup> and function of CMS and its subdetectors, and terminates in a description of data  
<sup>560</sup> processing at CMS, beginning from online event filtering in the Level-1 Trigger, to  
<sup>561</sup> processing in the High-Level Trigger, to offline particle reconstruction, and finally  
<sup>562</sup> long-term storage and processing of measured events.

## 563    2.1   The Large Hadron Collider

564    CERN, the European Organization for Nuclear Research, is an international organiza-  
565    tion based in Meyrin, Switzerland which operates the world's largest particle physics  
566    laboratory, and is the site of the Large Hadron Collider (LHC) [6]. The very first  
567    accelerator built at CERN was the 600 MeV Synchrocyclotron (SC), which initially  
568    provided beams for CERN's first experiments. The newer and more powerful Proton  
569    Synchrotron (PS), which could accelerate particles to an energy of 28 GeV, began op-  
570    erations in 1959 and is still in use today. The first hadron collider at CERN was the  
571    Intersecting Storage Rings (ISR), which consisted of two interlaced rings each with a  
572    diameter of 200. The ISR collided protons at a center-of-mass energy of 62 GeV and  
573    began measuring collisions in 1971. In 1968 CERN began to accelerate heavy ions  
574    in the Super Proton Synchrotron (SPS), which is 7 kilometers in circumference and  
575    was the first of CERN's giant underground rings to be built. The SPS became the  
576    forefront of CERN's particle physics program in 1976, and in 1981 was converted into  
577    a proton-antiproton collider. The final and largest underground ring constructed at  
578    CERN was the Large Electron-Positron (LEP) collider, which was commissioned in  
579    July 1989 and hosted 5176 magnets and 128 accelerating cavities located around a  
580    27-kilometer circumference. Over 11 years of research, four detectors, ALEPH, DEL-  
581    PHI, L3, and OPAL measured the collisions, with collision energies reaching up to  
582    209 GeV in the year 2000. In November 2000, LEP was closed down to make way for  
583    the construction of the LHC in the same tunnel.

584    In its current configuration, the LHC accelerator complex at CERN is a succession  
585    of machines that accelerate particles in stages until they reach their final energy of 6.5  
586    TeV per beam [7] [8]. In Linear accelerator 4 (Linac4), negative hydrogen ions (hy-  
587    drogen atoms with an additional electron) are accelerated to 160 MeV, and stripped  
588    of their two electrons, leaving only protons, before entering the Proton Synchrotron  
589    Booster (PSB). These protons are accelerated to 2 GeV, then to 26 GeV in the Proton

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

597 The beams of particles in the LHC are made to collide at a center-of-mass en-  
598 ergy of up to 14 TeV, at four positions at particle detector experiments located  
599 around the ring: ATLAS, CMS, ALICE, and LHCb. An aerial view of the four ma-  
600 jor experiments' locations is shown in Fig. 2.1 [9]. ATLAS and CMS are the two  
601 general-purpose detectors with broad physics programmes spanning Standard Model  
602 measurements and searches for signatures of new physics [10] [11]. The two experi-  
603 ments use different technical solutions and different magnet system designs. ALICE  
604 is a general-purpose detector dedicated to measuring LHC heavy-ion collisions, and  
605 is designed to address the physics of strongly interacting matter, and the properties  
606 of quark-gluon plasma [12]. The LHCb experiment specializes in investigating CP vi-  
607 olation through measuring the differences in matter and antimatter, by using a series  
608 of subdetectors to detect mainly forward particles close to the beam direction [13].

## 609 2.2 Luminosity and pileup

610 The number of events generated per second by the LHC collisions is given by

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

611 where  $\sigma_{event}$  is the cross-section for the event under study, and  $\mathcal{L}$  the machine lumi-  
612 nosity. The machine luminosity is measured in units of  $\text{cm}^{-2} \text{ s}^{-1}$ , and depends only

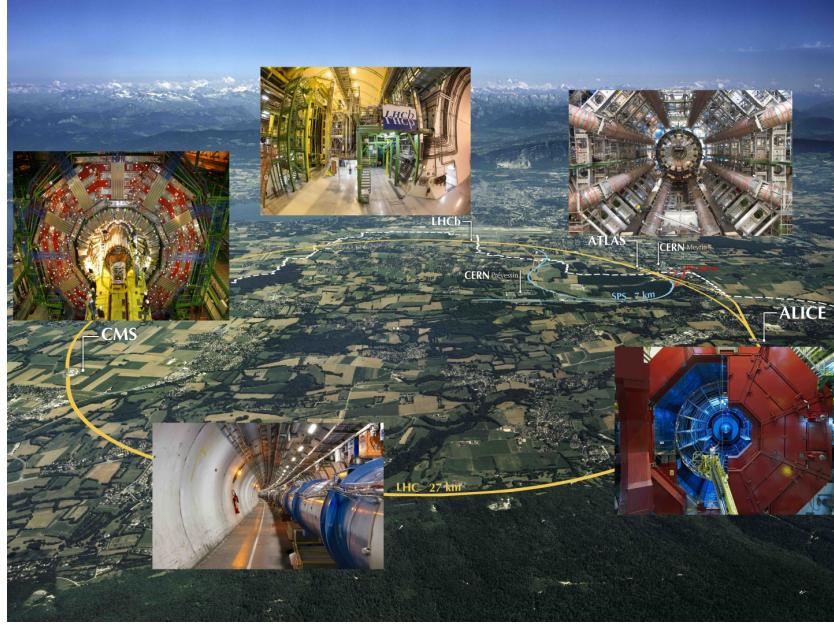


Figure 2.1: Aerial view of the Large Hadron Collider (LHC) spanning the border of France and Switzerland, and the four major experiments located around the ring: CMS (Compact Muon Solenoid), LHCb (LHC beauty), ATLAS (A Toroidal LHC Apparatus), and ALICE (A Large Ion Collider Experiment). [9]

613 on the beam parameters, and can be written for a Gaussian beam distribution as:

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

614 where the parameters are as defined, along with some example typical nominal values  
 615 in Phase-1 of the LHC [14] [15]:

616 •  $N_b$  is the number of particles per bunch ( $N_b \approx 1.15 \times 10^{11}$  protons per bunch)

617 •  $n_b$  is the number of bunches per beam (maximum 2808),

618 •  $f_{rev}$  is the revolution frequency ( $\approx 11$  kHz),

619 •  $\gamma_r$  is the relativistic gamma factor,

620 •  $\epsilon_n$  is the normalized transverse beam emittance (area in a transverse plane  
 621 occupied by the beam particles),

- 622     •  $\beta^*$  is the beta function at the collision point ( $\beta^* = 0.55$  m),  
 623     • and  $F$  is the geometric luminosity reduction factor due to the crossing angle at  
 624       the interaction points ( $F \approx 0.84$  for Phase-1. Note that complete overlap would  
 625       give  $F = 1$ ).

626   Peak luminosity at interaction points 1 and 5 reach values of  $\sim 1.0 \times 10^{34}$  cm $^{-2}$  s $^{-1}$ ,  
 627   with peak luminosity per bunch crossing reaching  $\sim 3.56 \times 10^{34}$  cm $^{-2}$  s $^{-1}$ .

628   Per Eqn. 2.1, the integrated luminosity over time is proportional to the number  
 629   of events produced, and the size of LHC datasets is commonly presented in terms of  
 630   integrated luminosity. Collider operation aims to optimize the integrated luminosity.  
 631   Thus the exploration of rare events in the LHC collisions requires both high beam  
 632   energies and high beam intensities.

633   The LHC’s nominal beam luminosities are sufficiently large for multiple proton-  
 634   proton collisions to occur in the same time window of 25 nanoseconds in which proton  
 635   bunches collide [16]. These multiple collisions will lead to particle interactions over-  
 636   lapping in the detector. To measure a proton-proton collision, the single collision  
 637   must be separated from overlapping collisions, which are called “pileup” collisions. A  
 638   distribution of pileup in the data-taking years 2016-2018 is shown in Fig. 2.2. The  
 639   pileup is defined as the average number of  $pp$  collisions per bunch crossing.

640   CMS reports an inelastic  $pp$  cross section of  $\sigma_{\text{inel}} = 68.6$  millibarns at a center-of-  
 641   mass energy of  $\sqrt{s} = 13$  TeV [17], which can be used to estimate pileup as follows:

$$\text{Pileup} = \frac{\mathcal{L} \times \sigma_{\text{inel}}}{n_b \cdot f} \quad (2.3)$$

642   With the example values above, pileup can be estimated to be  $\sim 22$ .  
 643   Thus, higher luminosities create more intense pileup conditions, posing a greater  
 644   challenge to detector performance and particle reconstruction and identification.

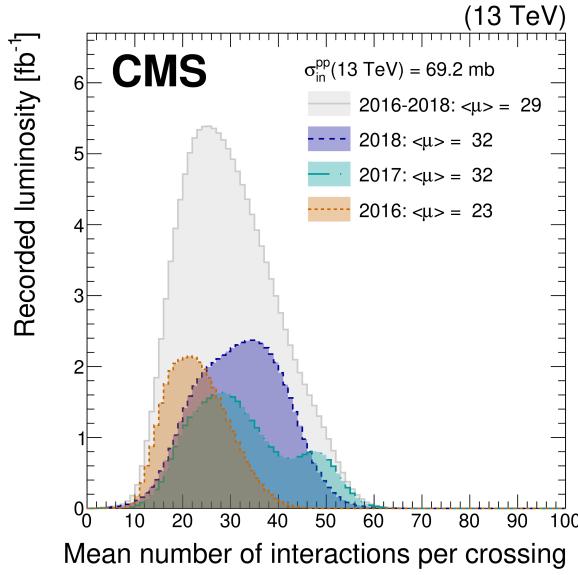


Figure 2.2: Distribution of the mean number of inelastic collisions per bunch crossing (pileup) in data [16], for proton-proton collisions in 2016 (*dotted orange*), 2017 (*dotted light blue*), 2018 (*dotted dark blue*), and integrated over 2016-2018 (*solid grey*). A cross-section of inelastic proton-proton collisions of 69.2 mbarns is assumed. In the running conditions of the High-Luminosity LHC, pileup will reach unprecedented levels of up to 200 per bunch crossing [18].

## 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 [18]. This will be accomplished through accelerator technological advances: for instance, reduction of the interaction point  $\beta^*$  from 0.55 m down to 0.15 m by installation of new final-focusing magnets, and improvements in the geometric luminosity loss factor  $F \approx 1$  through the installation of crab cavities that optimize the orientation of colliding bunches. A further discussion of the HL-LHC upgrades for the CMS detector follows in Chapter 3.

## 655 2.4 The CMS Detector

656 The Compact Muon Solenoid (CMS) experiment was conceived to study proton-  
657 proton and lead-lead collisions at a center-of-mass energy of 14 TeV (5.5 TeV nucleon-  
658 nucleon) and at luminosities up to  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  ( $10^{27} \text{ cm}^{-2} \text{ s}^{-1}$ ) [19] [20]. Starting  
659 from the beam interaction region at the center of the CMS detector, particles first  
660 pass through a silicon pixel and strip tracker, in which charged-particle trajectories  
661 (tracks) and origins (vertices) are reconstructed from signals (hits) in the sensitive  
662 layers. The tracker is immersed in a high-magnetic-field superconducting solenoid  
663 that bends the trajectories of charged particles, allowing the measurement of their  
664 electric charge and momenta. Electrons and photons are then absorbed in an electro-  
665 magnetic calorimeter (ECAL) comprised of lead-tungstate scintillating-crystals. The  
666 corresponding electromagnetic showers are detected as clusters of energy recording in  
667 neighboring cells, from which the direction and energy of the particles can be deter-  
668 mined. Charged and neutral hadrons may initiate a hadronic shower in the ECAL as  
669 well, which is then fully absorbed in the hadron calorimeter (HCAL). The resulting  
670 clusters are used to estimate their direction and energies. Muons and neutrinos pass  
671 through the calorimeters with little to no interactions. Neutrinos escaped undetected;  
672 muons produce hits in additional gas-ionization chamber muon detectors housed in  
673 the iron yoke of the flux-return. A sketch of example particle interactions in a trans-  
674 verse slice of the CMS detector is shown in Fig. 2.3. The collision data is recorded  
675 with the use of the Level-1 (L1) trigger (discussed separately in 2.5.5), high-level trig-  
676 ger (HLT), and data acquisition systems ensuring high efficiency in selecting physics  
677 events of interest.

678 CMS uses a right-handed coordinate system [19]. The origin is centered at the  
679 nominal collision point inside the experiment. The  $x$  axis points towards the center  
680 of the LHC, and the  $y$  axis points vertically upwards. The  $z$  axis points along the  
681 beam direction. The azimuthal angle,  $\phi$ , is measured from the  $x$  axis in the  $x$ - $y$

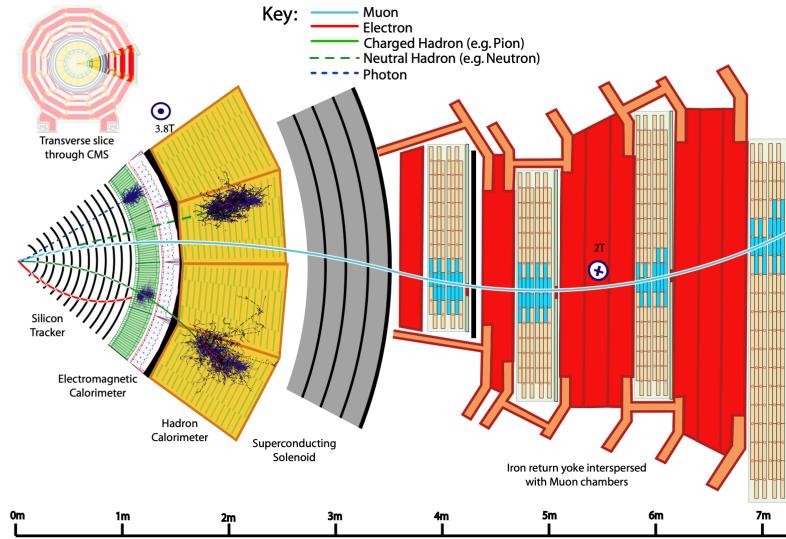


Figure 2.3: Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector [20].

682 plane, and the radial coordinate in this plane is denoted by  $r$ . The polar angle,  $\theta$ ,  
 683 is measured from the  $z$  axis. The pseudorapidity,  $\eta$ , is defined as  $\eta = -\ln \tan(\theta/2)$ .  
 684 The momentum and energy transverse to the beam direction, denoted by  $p_T$  and  $E_T$   
 685 respectively, are computed from the  $x$  and  $y$  components. The momentum imbalance  
 686 in the transverse plane is called the missing transverse momentum, and its magnitude  
 687 is denoted by  $E_T^{\text{miss}}$ .

## 688 2.5 Sub-detectors of CMS

689 This section details the sub-detectors of CMS that operate to identify and precisely  
 690 measure muons, electrons, photons, and jets over a large energy range.

### 691 2.5.1 Inner tracking system

692 The CMS Tracker performs robust tracking and detailed vertex reconstruction in the  
 693 4 T magnetic field of the superconducting solenoidal magnet. The primary sensors  
 694 used in the tracker are  $p^+$  on  $n$ -bulk devices, which allow high voltage operation and

are radiation-resistant [21] [22]. The active envelope of the CMS Tracker extends to a radius of 115 cm, over a length of approximately 270 cm on each side of the interaction point [21]. Charged particles in the region  $|\eta| \lesssim 1.6$  benefit from the full momentum measurement precision. In this region, a charged particle with  $p_T$  of 1000 GeV has a sagitta of  $\sim 195 \mu\text{m}$ . The Tracker acceptance extends further to  $|\eta| = 2.5$ , with a reduced radius of approximately 50 cm.

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

A schematic view of the current Phase-1 CMS tracker [23], including the pixel detector, is shown in Fig. 2.4. The Phase-1 pixel detector consists of three barrel layers (BPIX) at radii of 4.4 cm, 7.3 cm, and 10.2 cm, and two forward/backward disks (FPIX) at longitudinal positions of  $\pm 34.5$  cm and  $\pm 46.5$  cm, and extending in radius from about 6 cm to 15 cm. These pixelated detectors produce 3D measurements along the paths of charged particles with single hit resolutions between 10-20  $\mu\text{m}$ .

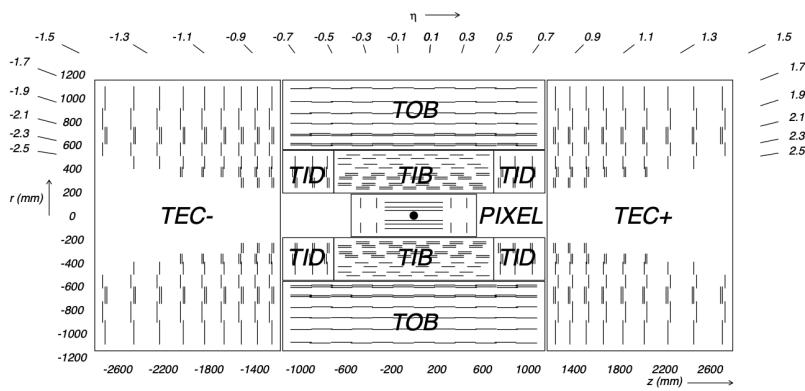


Figure 2.4: Cross section of the current Phase-1 CMS tracker [23]. Each line represents a detector module. Double lines indicate back-to-back modules which deliver two-dimensional (stereo) hits in the strip tracker.

After the pixel and on their way out of the tracker, particles pass through the silicon strip tracker which reaches out to a radius of 130 cm (Fig. 2.4). The sensor

elements in the strip tracker are single-sided *p*-on-*n* type silicon micro-strip sensors [19]. The silicon strip detector consists of four inner barrel (TIB) layers assembled in shells, with two inner endcaps (TID), each composed of three small discs. The outer barrel (TOB) consists of six concentric layers. Two endcaps (TEC) close off the tracker on either end.

### 2.5.2 ECAL

The electromagnetic calorimeter (ECAL) of CMS measures electromagnetic energy deposits with high granularity. One of the driving criteria in the design was the capability of detecting the Standard Model Higgs boson decay to two photons (in fact, the channel in which the 125 GeV Higgs boson was discovered at CMS). ECAL is a hermetic homogeneous calorimeter comprised of 61,200 lead tungstate ( $\text{PbWO}_4$ ) crystals mounted in the central barrel, with 7,324 crystals in each of the two endcaps [19]. A preshower detector is located in front of the endcap crystals. Avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps.

The design of the ECAL is driven by the behaviour of high-energy electrons, which predominantly lose energy in matter via bremsstrahlung, and high-energy photons by  $e^+e^-$  pair production. The characteristic amount of matter traversed for these interactions is the radiation length  $X^0$ , usually measured in units of  $\text{g cm}^{-2}$ . The radiation length is also the mean distance over which a high-energy electron loses all but  $1/e$  of its energy via bremsstrahlung [24]. Thus high granularity in  $\eta$  and  $\phi$ , and the length of the ECAL crystals, is designed to capture the shower of  $e/\gamma$  produced by electrons and photons.

The barrel part of the ECAL (EB) covers the pseudorapidity range  $|\eta| < 1.479$  [19]. The barrel granularity is 360-fold in  $\phi$  and  $(2 \times 85)$ -fold in  $\eta$ . The crystal cross-section corresponds to approximately  $0.0174 \times 0.0174$  in  $\eta - \phi$  or  $22 \times 22 \text{ mm}^2$  at the

739 front face of the crystal, and  $26 \times 26 \text{ mm}^2$  at the rear face. The crystal length is 230  
 740 mm, corresponding to  $25.8 X_0$ .

741 The ECAL read-out acquires the signals of the photodetectors [19]. At each bunch  
 742 crossing, digital sums representing the energy deposit in a trigger tower, comprising  
 743  $5 \times 5$  crystals in  $\eta \times \phi$ , are generated and sent to the Level-1 trigger system (detailed  
 744 in Section 2.5.5).

### 745 2.5.3 HCAL

746 The hadronic calorimeter (HCAL) of CMS measures hadronic energy, which is key to  
 747 characterizing the presence of apparent missing transverse energy which could arise  
 748 from hadron jets and neutrinos or exotic particles [19]. A schematic of the components  
 749 of HCAL are shown in Fig. 2.5. The HCAL barrel (HB) and endcaps (HE) are located  
 750 outside of the tracker and the ECAL, spanning a radius of 1.77 m (outer extent of  
 751 ECAL) up to 2.95 m (inner extent of the magnet coil). An outer hadron calorimeter  
 752 (HO) is placed outside the solenoid to complement the barrel calorimeter. Beyond  
 753  $|\eta| = 3$ , the forward hadron calorimeter (HF) at 11.2 m from the interaction point  
 754 extend the pseudorapidity coverage to  $|\eta| = 5.2$ .

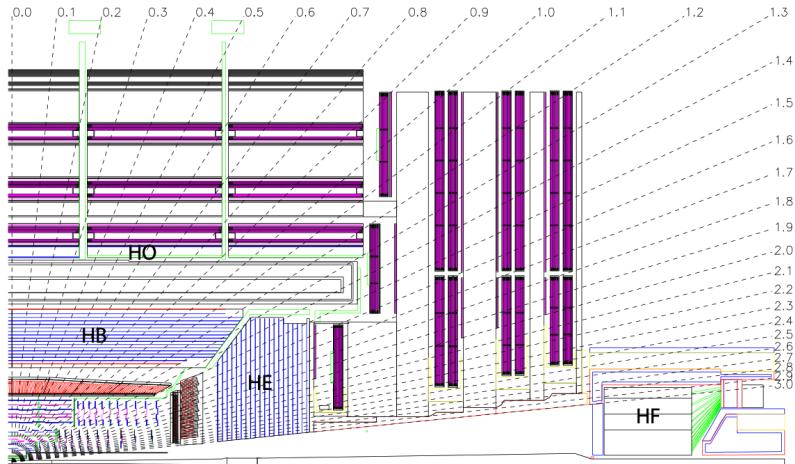


Figure 2.5: Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters from [19].

755        The HB is a sampling calorimeter covering the pseudorapidity range  $|\eta| < 1.3$  [19].  
756        It consists of 36 identical azimuthal wedges which form two half-barrels (HB+ and HB-  
757        ), with a segmentation of  $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$ . The HE covers pseudorapidity  
758         $1.3 < |\eta| < 3$ . The HB and endcap HE calorimeters are sampling calorimeters which  
759        use brass as the absorber and plastic scintillator as the active material. Light from  
760        the plastic scintillator is wavelength-shifted and captured in optic fibers which are  
761        read out by front-end electronics [25].

762        In the central pseudorapidity region, the combined stopping power of EB plus the  
763        HB is insufficient to contain hadron showers [19]. To ensure adequate sampling depth,  
764        the hadron calorimeter is extended with a tail catcher, the HO. The size and position  
765        of the tiles are designed to roughly map the layers of the HB to make towers with  
766        the same granularity of  $0.087 \times 0.087$  in  $\eta$  and  $\phi$ . HO uses the same active material  
767        as the HB and HE calorimeters, but uses the steel return yoke and magnet material  
768        of CMS as absorbers [25].

769        The HF is a Cherenkov calorimeter based on a steel absorber and quartz fibers  
770        which run longitudinally through the absorber and collect Cherenkov light, primarily  
771        from the electromagnetic component of showers developed in the calorimeter [25].  
772        Photomultiplier tubes are used to collect light from the quartz fibers. The HF is  
773        designed to survive in the harsh radiation conditions and high particle flux of the  
774        forward region. On average, 760 GeV per proton-proton interaction is deposited into  
775        the two forward calorimeters, compared to only 100 GeV for the rest of the detector  
776        [19]. Furthermore, this energy has a pronounced maximum at the highest rapidities.

#### 777        2.5.4 Muon detectors

778        The CMS muon system is designed to have the capability of reconstructing the mo-  
779        mentum and charge of muons over the kinematic range of the LHC, since muons are a  
780        powerful handle on signatures of interesting processes over the high background rate

781 of the LHC [19]. For instance, the decay of the Standard Model Higgs boson into  
 782  $ZZ$ , which in turn decay to 4 leptons, can be reconstructed with high 4-particle mass  
 783 resolution if all the leptons are muons, since muons are less affected than electrons  
 784 by radiative losses in the tracker material.

785 The muon system consists of a cylindrical barrel section and two planar endcap  
 786 regions [19]. The barrel muon detector consists of drift tube (DT) chambers covering  
 787 the pseudorapidity region  $|\eta| < 1.2$  (Fig. 2.6). The DTs can be used as tracking  
 788 detectors due to the barrel region's characteristic low neutron-induced backgrounds,  
 789 low muon rate, and relatively uniform 4T magnetic field contained in the steel yoke.

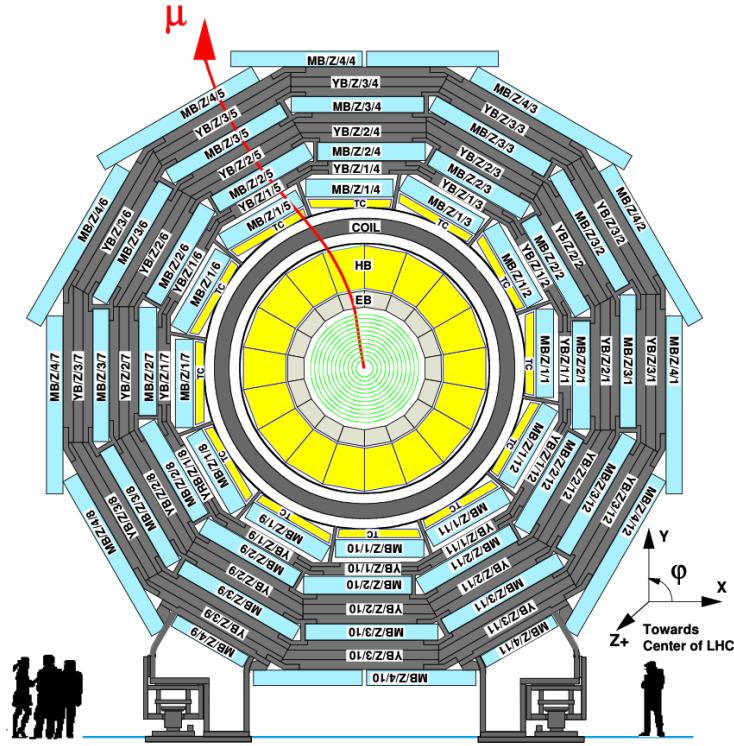


Figure 2.6: Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels from [19]. The DTs are organized in 12 sectors of the yoke barrel (YB). In each of the 12 sectors of the yoke, there are 4 muon chambers per wheel (MB1, MB2, MB3, and MB4).

790 In the two endcap regions, the muon rates and background levels are high and the

791 magnetic field is large and non-uniform [19]. Here, the muon system uses cathode  
792 strip chambers (CSCs) to identify muons between  $0.9 < |\eta| < 2.4$ . The cathode strips  
793 of each chamber run radially outwards and provide a precision measurement in the  
794  $r - \phi$  bending plane. The anode wires run approximately perpendicular to the strips  
795 and are read out in order to measure  $\eta$  and the beam-crossing time of a muon.

796 In addition to the DT and CSC, a dedicated trigger system consisting of resistive  
797 plate chambers (RPCs) in the barrel and endcap regions provide a fast, independent,  
798 and highly-segmented trigger with a sharp  $p_T$  threshold over a large portion of the  
799 pseudorapidity range ( $|\eta| < 1.6$ ) of the muon system [19]. RPCs have good time  
800 resolution but coarser position resolution compared to the DTs or CSCs. The RPCs  
801 also play a role in resolving ambiguities in reconstructing tracks from multiple hits in  
802 a chamber.

### 803 2.5.5 The Level-1 Trigger

804 The design performance of the LHC corresponds to an instantaneous luminosity of  
805  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  with a 25 ns bunch crossing rate, giving an average pile-up (number  
806 of simultaneous events) of 25 per bunch crossing [26]. The large number of minimum  
807 bias events per bunch crossing, combined with the small cross-sections of possible  
808 physics discovery signatures, necessitates a sophisticated event selection system for  
809 filtering this large event rate, as it is impossible to save all events. This data filtering  
810 system is implemented by CMS in two stages. The first stage is the Level-1 (L1)  
811 Trigger, which is deployed in custom electronic hardware systems and is responsible  
812 for reducing the event rate to around 100 kHz. The second stage is the High-Level  
813 Trigger (HLT) which is described in Section 2.5.6. This section describes the Phase-1  
814 configuration of the Level-1 Trigger.

815 The L1 Trigger data flow of Phase-1 is shown in Fig. 2.7 [26], with organization  
816 into the L1 calorimeter trigger, the L1 muon trigger, and the L1 global trigger.

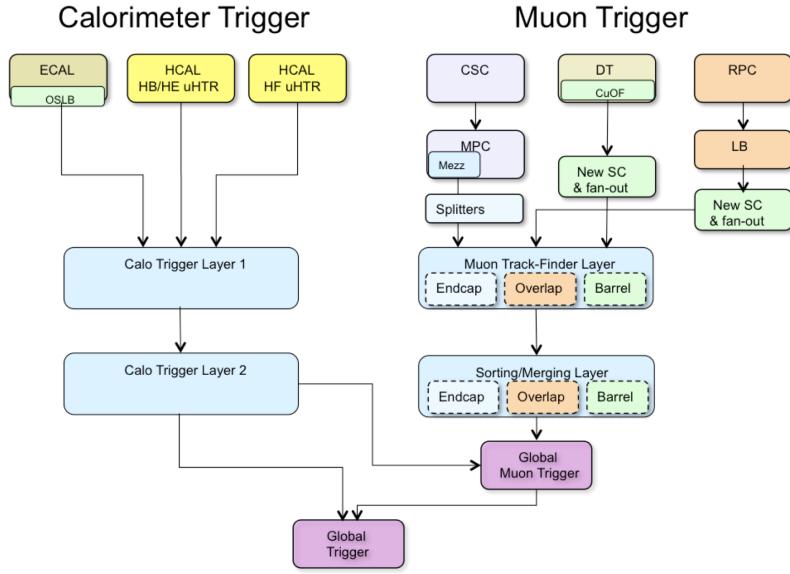


Figure 2.7: Dataflow for the Phase-1 Level-1 Trigger [26], which is implemented in custom hardware and is responsible for reducing the event rate from the LHC bunch crossing frequency of 400 MHz (bunch crossings every 25 ns) to a maximum rate of 100 kHz. In Phase-1, the Level-1 Trigger has access to information from the calorimeter and muon detectors.

817        The L1 calorimeter trigger begins with trigger tower energy sums formed by the  
 818        ECAL, HCAL, and HF Trigger Primitive Generator (TPG) circuits from the indi-  
 819        vidual calorimeter cell energies. In the original configuration, the ECAL energies  
 820        were accompanied by a bit indicating the transverse extent of the electromagnetic  
 821        energy deposits, and the HCAL energies were accompanied by a bit indicating the  
 822        presence of minimum ionizing energy [27]. Between Long Shutdowns 1 and 2 (LS1  
 823        and LS2), HF was upgraded to provide finer granularity information to the trigger,  
 824        and the HCAL barrel and endcap front-end electronics were upgraded to provide  
 825        high-precision timing information and depth segmentation information.

826        In the original design of the L1 calorimeter trigger, the trigger primitives are pro-  
 827        cessed by the Regional Calorimeter Trigger (RCT, upgraded to Calo Layer 1 after  
 828        LS2) which finds isolated and non-isolated electron/photon candidates [26]. At this  
 829        stage, electrons/photon candidates are treated together since they cannot be defini-

tively distinguished at this stage due to lack of tracking information in the L1 trigger.  
The Global Calorimeter Trigger (GCT, upgraded to Calo Layer 2 after LS2) sorts further the candidate electrons/photons, finds jets (classified as central, forward, and tau) using the  $E_T$  sums and performs calibration of the clustered jet energies, and calculates global quantities such as missing  $E_T$ . It sends the top four candidates of each type to the global trigger (GT) [26].

Each of the L1 muon triggers has its own trigger logic [27]. The RPC strips are connected to a Pattern Comparator Trigger (PACT), which forms trigger segments that are used to build tracks and calculate  $p_T$ . The RPC logic also provides some hit data to the CSC trigger system to resolve ambiguities caused by two muons in the same CSC. The CSCs form local charged tracks (LCTs) from the cathode strips, which are combined with the anode wire information. LCTs are combined into full muon tracks and assigned  $p_T$  values.

The Global Muon Trigger (GMT) sorts the RPC, DT, and CSC muon tracks, converts these tracks to the same  $\eta$ ,  $\phi$ , and  $p_T$  scale, and validates the muon sign [27]. It improves the trigger efficiency by merging muon candidates that were detected in two complementary sub-systems (i.e. DT+RPC, or CSC+RPC). The GMT also contains logic to correlate the found muon tracks with an  $\eta-\phi$  grid of quiet calorimeter towers to determine if the muons are isolated, as well as logic to remove duplicate candidates originating in the overlap regions from both DT and CSC systems. The final collection of muons are sorted based on their initial quality, correlation, and  $p_T$ , and the top four muons are sent to the Global Trigger [27].

Information from the GCT and GT are sent to the Global Trigger (GT), which makes the Level-1 Accept (L1A) decision to either discard or accept the bunch crossing [27]. This is accomplished by sorting ranked trigger objects that are accompanied by positional information in  $\eta$  and  $\phi$ , permitting the trigger to applying criteria with thresholds that can vary based on the location of the trigger objects, and/or to

857 require trigger objects to be close to or opposite from each other. The GT L1A  
858 decision arrives at the detector front end with a  $3.8 \mu\text{s}$  latency after the interaction  
859 at a rate which is required to be less than 100 kHz, and triggers a full readout of the  
860 detector for further processing.

### 861 2.5.6 The High-Level Trigger

862 The HLT is implemented in software running on a large computer farm of fast com-  
863 mercial processors [28] [29]. The algorithms in HLT have access to full data from  
864 all CMS sub-detectors, including the tracker, with full granularity and resolution.  
865 The HLT reconstruction software is similar to what is used offline for full CMS data  
866 analysis. As a result, the HLT can calculate quantities with a resolution compara-  
867 ble to the final detector resolution, compared to the L1 Trigger. The HLT performs  
868 more computationally-intensive algorithms, such as combining tau-jet candidates in  
869 the calorimeter with high- $p_T$  stubs in the tracker, to form a hadronic tau trigger. The  
870 maximum HLT input rate from the L1 Trigger is 100 kHz, and the HLT output rate  
871 is approximately 100 Hz.

872 The HLT contains trigger paths, each corresponding to a dedicated trigger [30].  
873 A path consists of several steps implemented as software modules. Each HLT trigger  
874 path must be seeded by one or more L1 trigger bits: the first module always looks  
875 for a L1 seed, consisting of L1 bit(s) and L1 object(s). Each module performs a well-  
876 defined task such as unpacking (raw to digitized quantities), reconstruction of physics  
877 objects (electrons, muons, jet, missing transverse energy, etc.), making intermediate  
878 decisions that trigger more detailed reconstruction modules, and calculating the final  
879 decision for the trigger path. If an intermediate filter decision is negative, the rest of  
880 the path is not executed, and the trigger rejects the event.

881 **2.5.7 Particle reconstruction**

882 To build a description of the physics objects present in the particle collision, the  
883 basic elements from the detector layers (tracks and clusters of energy) are correlated  
884 to identify each particle in the final state. Measurements from different sub-detectors  
885 are combined to reconstruct the particle properties. This approach is called particle-  
886 flow (PF) reconstruction [20]. Key to the success of the PF reconstruction is the  
887 fine spatial granularity of the detector layers. Coarse-grained detectors can cause  
888 the signals from different particles to merge, especially within jets. However, if the  
889 subdetectors are sufficiently segmented to separate individual particles, it becomes  
890 possible to produce a global event description that identifies all physics objects with  
891 high efficiencies and resolution.

892 **2.5.8 Data storage and computational infrastructure**

893 The LHC generates over 15 petabytes (15 million gigabytes) of data every year, neces-  
894 sitating a flexible computing system that can be accessed by researchers working at  
895 the four main LHC experiments: ALICE, ATLAS, CMS, and LHCb. The Worldwide  
896 LHC Computing Grid (WLCG) [31] is a global collaboration of computer centers that  
897 links thousands of computers and storage systems in over 170 centers across 41 coun-  
898 tries. These centers are arranged in “tiers”, and provide near real-time access to users  
899 processing, analyzing, and storing LHC data. One of the final stages of data analy-  
900 sis at LHC experiments is large-scale data processing taking place over distributing  
901 computing, for instance, with the use of Condor [32], a distributed, scalable, flexible  
902 batch processing system which accepts a computing job, allocates a resource to it,  
903 executes it, and returns the result back to a user transparently.

904 **Chapter 3**

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

906 **3.1 High-Luminosity LHC and CMS**

907 In order to sustain and extend the LHC’s physics discovery program and maintain  
908 operability for a decade or more, the LHC is undergoing a major upgrade to the High-  
909 Luminosity LHC (HL-LHC). In its final configuration, the HL-LHC will deliver a peak  
910 luminosity of  $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ , potentially leading to total integrated luminosity  
911 of  $4000 \text{ fb}^{-1}$  after ten years of operations, scheduled to begin in 2027 [33]. This  
912 integrated luminosity is about ten times the predicted luminosity reach of the LHC  
913 in its initial configuration. To maximize the discovery potential of this unprecedented  
914 amount of data, the CMS detector is undergoing Phase-2 upgrades in order to perform  
915 high-precision measurements and searches for physics beyond the Standard Model in  
916 the intense running conditions of the HL-LHC.

917 **3.2 The Phase-2 Level-1 Trigger**

918 To achieve the goals of the HL-LHC program and to ensure the collection of information-  
919 rich datasets in the HL-LHC, the Phase-2 upgrade of the CMS Level-1 Trigger [33]  
920 must be upgraded in conjunction with the CMS sub-detectors and their readouts, to

921 maintain physics selectivity. The HL-LHC will produce an intense hadronic environment  
 922 corresponding to 200 simultaneous collisions per beam crossing, necessitating  
 923 comprehensive upgrades of the trigger system outlined below.

924 To profit from the extended coverage and increased granularity of the upgraded  
 925 CMS detector, the latency of the L1 trigger system (time available to produce a L1  
 926 Accept signal) will be increased significantly from  $3.8 \mu\text{s}$  to  $12.5 \mu\text{s}$ , with an increased  
 927 maximum output bandwidth of 750 kHz [33]. With the increased latency, in addition  
 928 to information from calorimeters and muon detectors (as in the Phase-1 system),  
 929 information from the new tracker and high-granularity endcap calorimeter can also  
 930 be included at L1 for the first time. This is illustrated in the functional diagram of  
 931 the architecture of the Phase-2 trigger system in Fig. 3.1.

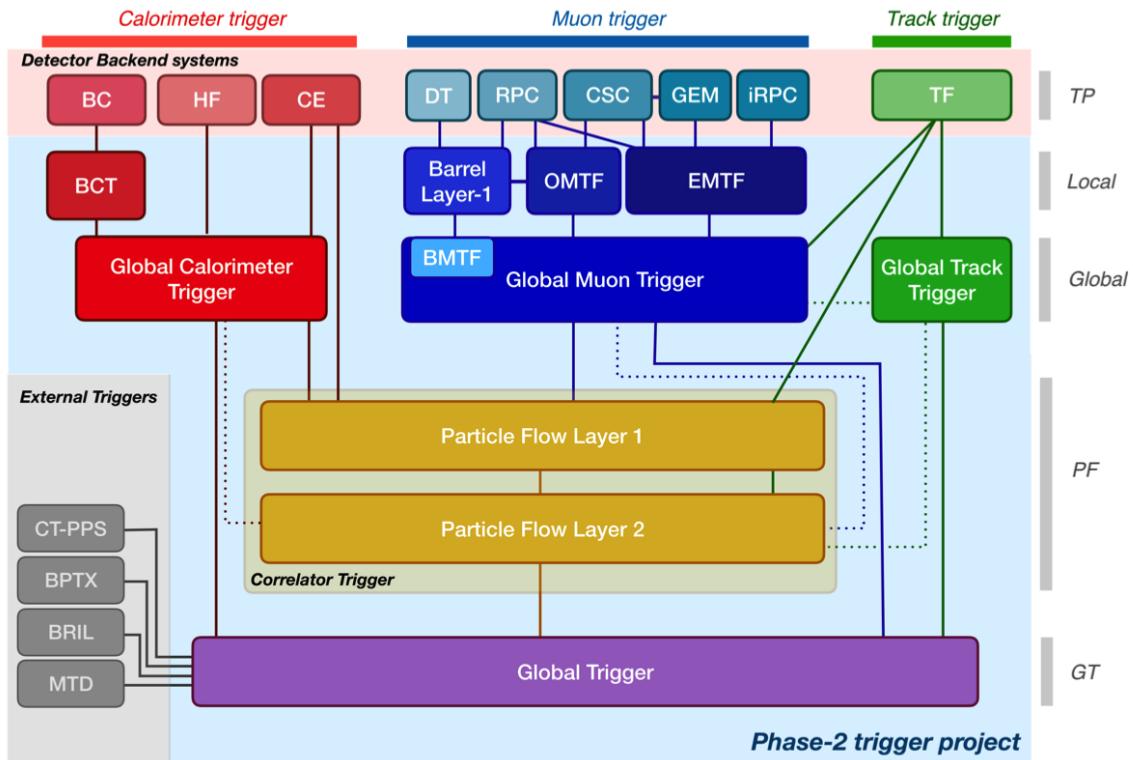


Figure 3.1: Functional diagram of the CMS L1 Phase-2 upgraded trigger design [33], showing the four trigger paths: calorimeter, muon, track, and Particle Flow. For the first time, tracking information will be available as early as the L1 Trigger.

932 The key feature of the Phase-2 L1 Trigger is the introduction of a correlator layer,

933 where algorithms produce higher-level trigger objects by combining information from  
934 sub-detectors, with a selectivity approaching that of offline reconstruction in the  
935 HLT [33]. Four independent data processing paths (grouped together in Fig. 3.1) are  
936 implemented: tracking, calorimetry, muon systems, and particle-flow techniques:

- 937     • **Calorimeter Trigger path:** (*red*, Fig. 3.1) A barrel calorimeter trigger (BCT)  
938         and the HGCAL backend are used to produce high-granularity information from  
939         the calorimeters to produce high-resolution clusters and identification variables  
940         used for later processing. Outputs from the BCT, HGCAL, and the HF are sent  
941         to a global calorimeter trigger (GCT), where calorimeter-only objects such as  
942          $e/\gamma$  candidates, hadronically decaying tau lepton candidates, jets, and energy  
943         sums are built.
- 944     • **Track Trigger path:** (*green*, Fig. 3.1) Tracks from the Outer Tracker are  
945         reconstructed in the track finder (TF) processors as part of the detector back-  
946         end. A global track trigger (GTT) will reconstruct the primary vertices of the  
947         event, along with tracker-only based objects, such as jets and missing transverse  
948         momentum.
- 949     • **Muon Trigger path:** (*blue*, Fig. 3.1) Trigger primitives are processed by  
950         muon track finder algorithms, again separated into the barrel (barrel muon  
951         track finder, BMTF), overlap (overlap muon track finder, OMTF), and endcap  
952         (endcap muon track finder, EMTF). Standalone muons and stubs containing  
953         information such as position, bend angle, and timing, as well as L1 tracks, are  
954         sent to the global muon trigger (GMT).
- 955     • **Particle-Flow Trigger path:** (*yellow*, Fig. 3.1) The correlator trigger (CT)  
956         aims to approach the performance of offline Particle Flow, and is implemented  
957         in two layers. “Layer-1” produces the particle-flow candidates from matching

958        calorimeter clusters and tracks. “Layer 2” builds and sorts final trigger objects  
959        and applies additional identification and isolation criteria.

960        The outputs from the above trigger paths are combined in the Global Trigger  
961        (GT) (*purple*, Fig. 3.1), which calculates the final trigger decision (Level-1 Accept),  
962        transmitting it to the Trigger Control and Distribution System (TCDS), which dis-  
963        tributes it to the detector backend systems, initiating the readout to the DAQ. The  
964        GT also provides the interface to external triggers (*grey*, Fig. 3.1), such as trig-  
965        gers for the precision proton spectrometer (PPS), beam position and timing monitors  
966        (BPTX), and luminosity and beam monitoring (BRIL) detectors [33]. The design of  
967        the Phase-2 Level-1 Trigger allows for future inclusion of triggering information, for  
968        instance information about minimum ionizing particles (MIPs) from the MIP Timing  
969        Detector (MTD) [34].

### 970        **3.3 Standalone Barrel Calorimeter electron/photon 971        reconstruction**

972        The reconstruction and identification of electrons and photons ( $e/\gamma$ ) begin with the  
973        trigger primitives of the barrel ECAL and HCAL detectors and endcap HGCAL  
974        calorimeters, covering the pseudorapidity region  $|\eta| < 3$ . The barrel and endcap re-  
975        gions of the detector are intrinsically different enough to warrant different approaches  
976        to  $e/\gamma$  reconstruction. This work focuses on the Standalone Calorimeter  $e/\gamma$  recon-  
977        struction taking place in the barrel (Fig. 3.2).

#### 978        **3.3.1 Phase-2 geometry of the ECAL Barrel trigger**

979        In Phase-2, the upgrade of both on-detector and off-detector electronics for the barrel  
980        calorimeters trigger primitive generator (TPG) will stream single crystal data from

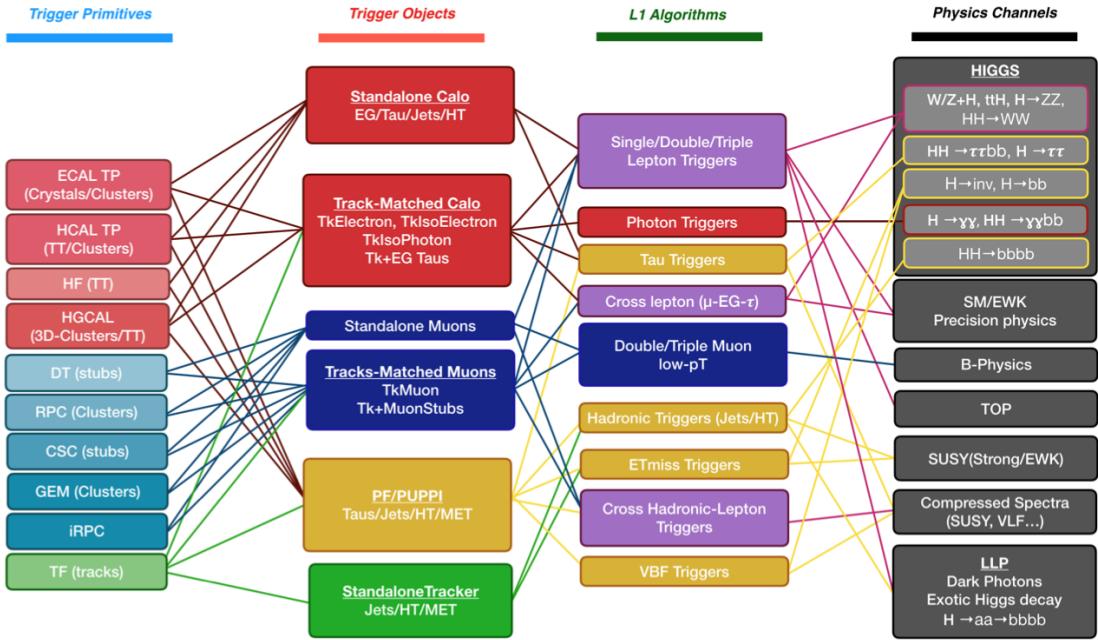


Figure 3.2: Summary of the links between the trigger primitives (*first column*), the trigger objects (*second column*), the Level-1 algorithms used in the menu (*3rd column*), and the physics channels (*4th column*), from [33], where a full description of the Phase-2 L1 algorithms can be found. This work focuses on developments for the Standalone Calorimeter electron and photon ("EG") reconstruction algorithm.

981 the on-detector to the backend electronics, in contrast to the lower-granularity output  
 982 of the Phase-1 ECAL TPG that is restricted to providing trigger tower sums of  $5 \times 5$   
 983 crystals [33]. A schematic representation of the geometry of the ECAL barrel in the  
 984 Regional Calorimeter Trigger (RCT) is shown in Fig. 3.3. The barrel is spanned by  
 985 36 RCT cards, each spanning  $17 \times 4$  towers in  $\eta \times \phi$ . Each RCT card is subdivided  
 986 into five “regions” as shown in Fig. 3.4. After initial clustering and processing, the  
 987 outputs of the RCT card are sent to the Global Calorimeter (GCT) trigger, which is  
 988 processed in three cards as shown in Fig. 3.5.

### 989 3.3.2 Phase-2 electron/photon reconstruction algorithm

990 The standalone barrel algorithm for reconstructing and identifying electrons and pho-  
 991 tons in the Phase-2 Level-1 Trigger takes as input the digitized response of each crystal

992 of the barrel ECAL, with a granularity  $0.0175 \times 0.0175$  in  $\eta \times \phi$ , which is 25 times  
993 higher than the input to the Phase-1 trigger, which consisted of trigger towers with  
994 a granularity of  $0.0875 \times 0.0875$ . In HCAL the tower size of  $0.0875 \times 0.0875$  is un-  
995 changed. The trigger algorithm is designed to closely reproduce the algorithm used in  
996 the offline reconstruction, with limitations and simplifications due to trigger latency.

997 In the RCT, an initial requirement of  $p_T > 0.5$  GeV is imposed on the input  
998 trigger primitives (i.e. energies from the ECAL crystals and HCAL towers) to reject  
999 contribution from pileup. In one of the regions inside a RCT card (Fig. 3.4), the  
1000 crystal containing the highest energy deposit is identified as the seed crystal, as shown  
1001 in Fig. 3.6. The energy in the crystals in a window of size  $3 \times 5$  in  $\eta \times \phi$  around  
1002 the seed cluster is added into a cluster. The energy is considered “clustered”. The  
1003 process is repeated with the remaining “unclustered” energy, until up to four clusters  
1004 are produced in the region.

1005 To improve  $e/\gamma$  identification and to reduce background contributions, identifica-  
1006 tion and reconstruction algorithms are implemented at this stage:

- 1007 • Shower shape: The energy deposit sums around the seed crystal is computed in  
1008 windows of size  $2 \times 5$  and  $5 \times 5$  (Fig. 3.6, *dashed lines*), with true  $e/\gamma$  clusters  
1009 tending to produce showers that deposit most of their energy in a  $2 \times 5$  region.
- 1010 • Bremsstrahlung recovery:  $e/\gamma$  tend to spread in the  $\phi$  direction due to charged  
1011 particles being bent by the magnetic field of the CMS solenoid. If sufficient  
1012 energy comparable to the core  $3 \times 5$  cluster is found in the adjacent  $3 \times 5$   
1013 windows (Fig. 3.6, *shaded yellow*), the energy is added to the core cluster and  
1014 no longer considered unclustered energy.

1015 After parallel processing in the regions, the clusters in a RCT card are stitched  
1016 together if they are located directly along the borders of a region (Fig. 3.3). The  
1017 remaining unclustered ECAL energy is summed into ECAL towers.

1018        From each RCT card, the twelve highest-energy clusters, as well as any remaining  
1019        unclustered energy, are sent to the GCT. Since each GCT card has information from  
1020        sixteen RCT cards (Fig. 3.5), final stitching across the boundaries of the RCT cards  
1021        is performed. One more identification algorithm is performed at this stage:

- 1022        • Isolation: One handle to reject backgrounds from e.g. pileup, comes from the  
1023        tendency for background to be spread more uniformly across a large area in the  
1024        detector, whereas genuine  $e/\gamma$  are expected to produce showers concentrated in  
1025        the  $3 \times 5$  crystal window. The energy sum in a large window of  $7 \times 7$  in towers  
1026        is computed and used to reject background.

1027        The performance of the standalone barrel  $e/\gamma$  algorithm in Phase-2 conditions is  
1028        summarized in the efficiency and rates. The efficiencies are measured with a simulated  
1029        Monte Carlo sample containing electrons. The rates are measured with a simulated  
1030        minimum bias sample intended to closely mimic generic proton-proton collisions in  
1031        the CMS detector. The performance of the Phase-2 emulator discussed in this work,  
1032        which closely mimics the firmware logic and uses fixed-precision integers, is shown to  
1033        be comparable to the previous emulator which used floats and idealized logic.

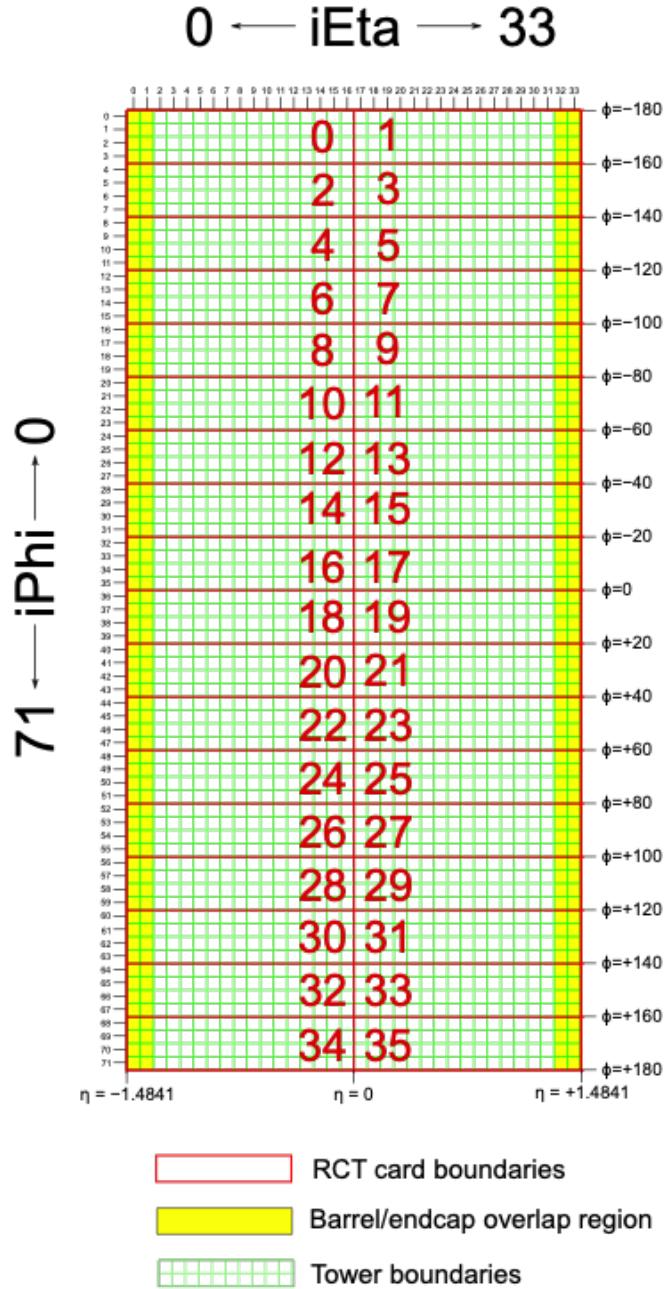


Figure 3.3: Schematic of the geometry of the Phase-2 ECAL barrel in the Regional Calorimeter Trigger (RCT), showing the division of the barrel region into 36 Regional Calorimeter Trigger (RCT) cards (red). Each card spans  $17 \times 4$  towers in  $\eta \times \phi$  (green), and each tower is  $5 \times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

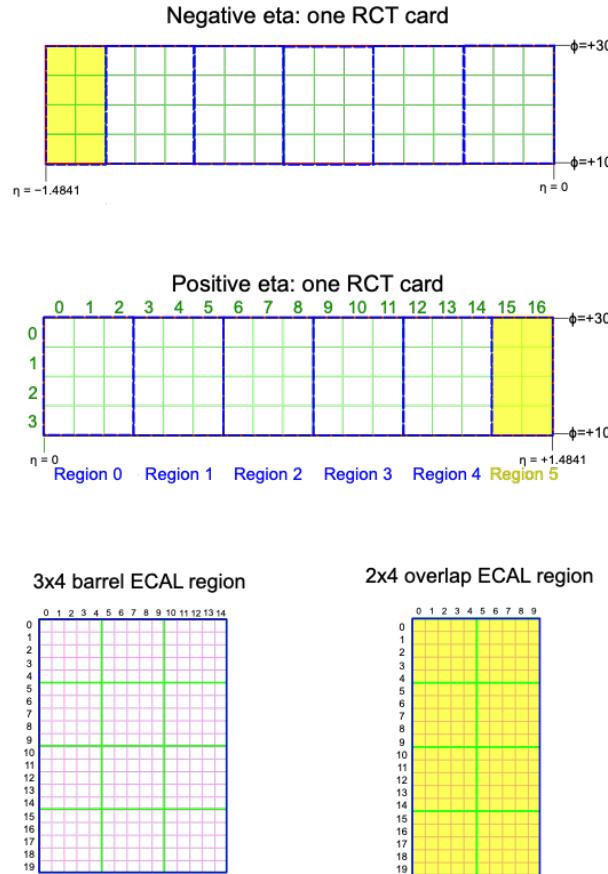


Figure 3.4: Schematic of two example RCT cards in the negative eta (*top*) and positive eta (*center*) regions of the ECAL barrel. Each RCT card is divided into five regions: four regions are of size  $3 \times 4$  towers in  $\eta \times \phi$  (*bottom left*), and a fifth smaller overlap region of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

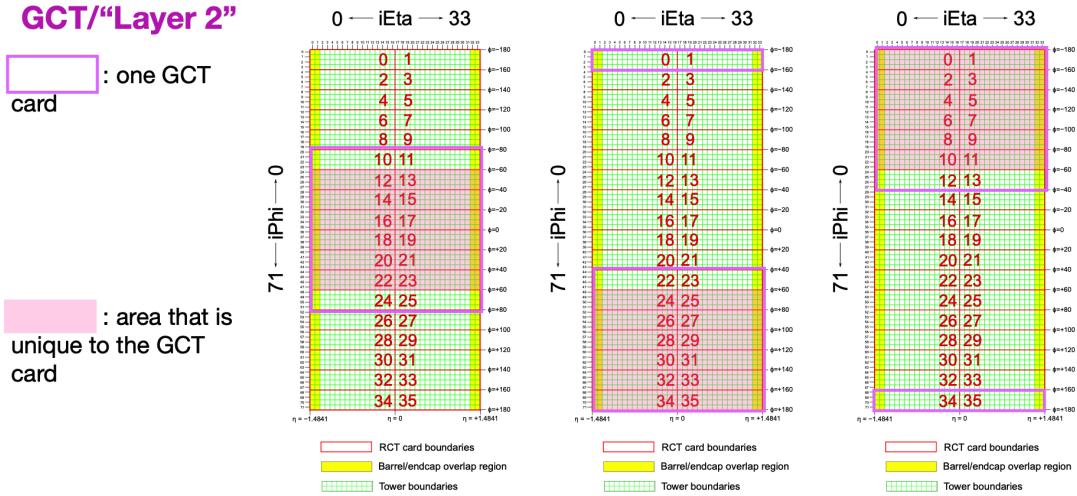


Figure 3.5: Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trigger (GCT), which will process the outputs of the Regional Calorimeter Trigger (RCT) in three cards (*magenta highlights*). Each card in the GCT processes the equivalent of sixteen RCT cards, with the center twelve being unique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

### 3x4 barrel ECAL region

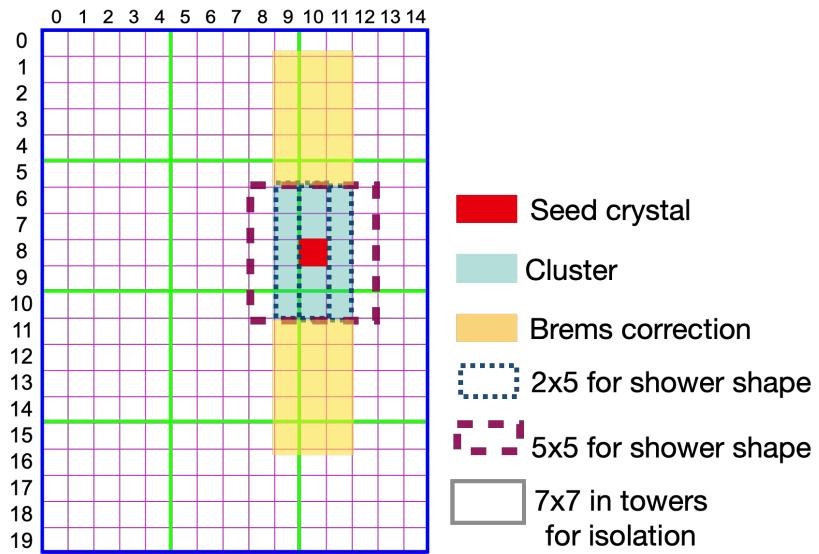


Figure 3.6: Illustration of an example electron/photon ( $e/\gamma$ ) cluster in the Phase-2 Level-1 Trigger standalone barrel  $e/\gamma$  reconstruction, in a region of  $15 \times 20$  crystals (3  $\times$  4 towers). Each small pink square is one crystal, the highest-granularity ECAL trigger primitives 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*). Bremsstrahlung corrections are checked in the adjacent  $3 \times 5$  windows in the  $\phi$  direction (*shaded light yellow*). The relative energies in windows of size  $2 \times 5$  and  $5 \times 5$  in crystals (*dashed dark blue and dark red*) are used to compute shower shape variables to identify true  $e/\gamma$  objects. Lastly, an isolation sum is computed in a window of size  $7 \times 7$  in towers (not shown in figure).

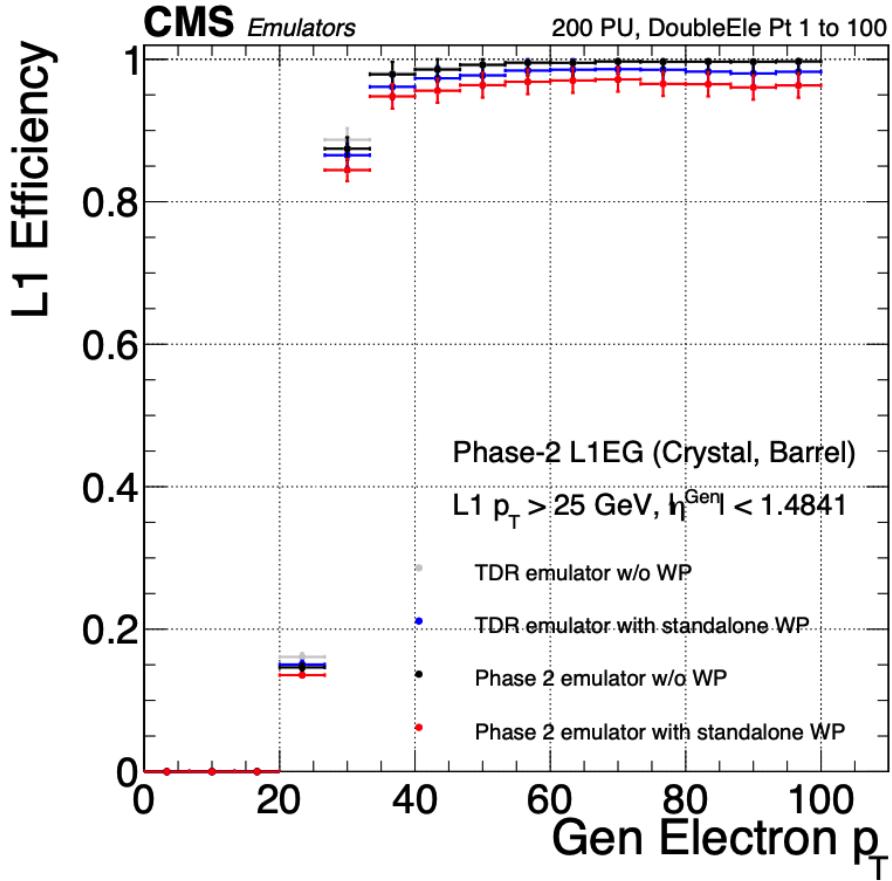


Figure 3.7: Efficiency of the standalone barrel  $e/\gamma$  reconstruction, measured in a simulated sample of electrons, as a function of the true electron’s transverse momentum  $p_T$ . The performance of the previous, idealized algorithm as shown in the 2021 Phase-2 TDR [33] with and without the isolation and shower shape discrimination variables (“standalone working point/ WP”) (*dark blue, grey*). The Phase-2 emulator discussed in this work with and without the same working point (*black, red*) is shown to have comparable performance.

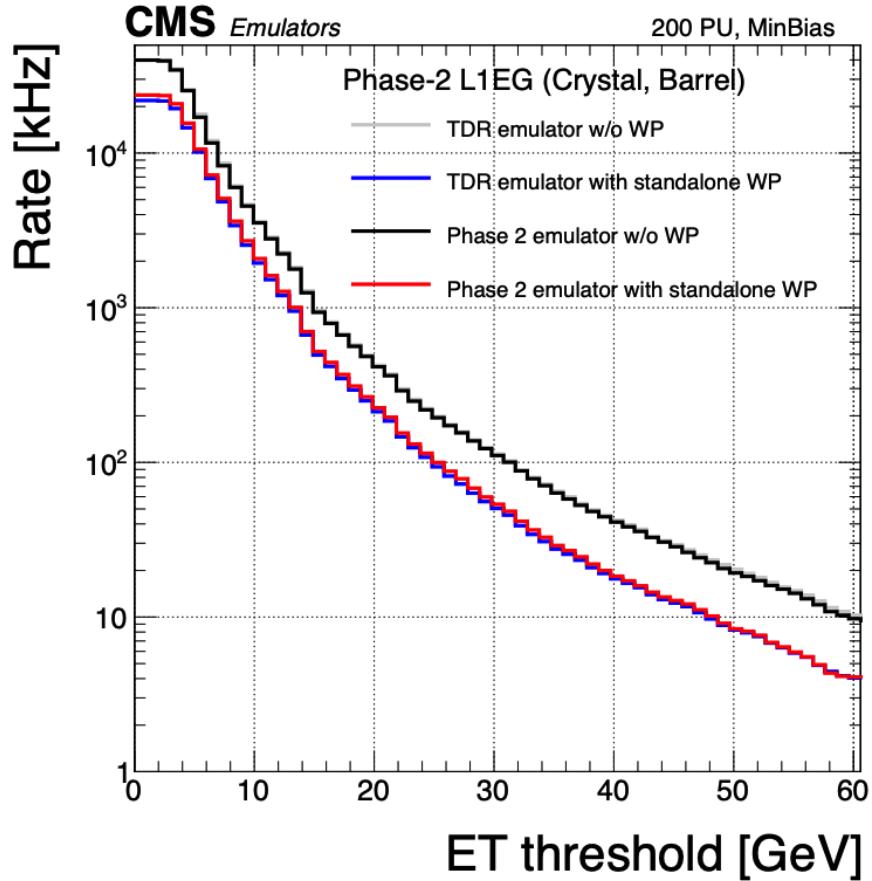


Figure 3.8: Rates of the standalone barrel  $e/\gamma$  reconstruction, evaluated on a minimum bias sample, measured as a function of the minimum energy ( $E_T$ ) required of the reconstructed  $e/\gamma$  object in each event. The performance of the previous, idealized algorithm as shown in the 2021 Phase-2 TDR [33] with and without the isolation and shower shape discrimination variables (“standalone working point/ WP”) (*dark blue, grey*). The Phase-2 emulator discussed in this work with and without the same working point (*black, red*) is shown to have comparable performance.

1034 **Chapter 4**

1035 **Datasets and Monte Carlo samples**

1036 **4.1 Datasets used**

1037 The  $h \rightarrow aa \rightarrow 2b2\tau$  analysis (CMS CADI line HIG-22-007) is based on proton-proton  
1038 collision data at a center-of-mass energy of 13 TeV collected in full Run-2 (2016-  
1039 18) with the CMS detector. The data analyzed corresponds to a total integrated  
1040 luminosity of  $138 \text{ fb}^{-1}$  ( $36.33 \text{ fb}^{-1}$  for 2016,  $41.53 \text{ fb}^{-1}$  for 2017, and  $59.74 \text{ fb}^{-1}$  for  
1041 2018) [35] [36] [37]. The cumulative delivered and recorded luminosity versus time  
1042 for 2015-2018 is shown in Fig. 4.1.

1043 Data collected with the single muon trigger is used for the  $\mu\tau_h$  channel. For the  
1044  $e\tau_h$  channel, data collected with the single electron trigger is used; and for the  $e\mu$   
1045 channel, data collected with the electron + muon trigger is used. A more in-depth  
1046 discussion of the triggers used follows in a later section.

1047 A full list of samples used can be found in the full documentation [39] [40].

1048 **4.2 Monte Carlo samples**

1049 Modeling and computing observables originating from arbitrary physics processes at  
1050 the tree level and at next-to-leading order (NLO) is performed by Monte Carlo (MC)

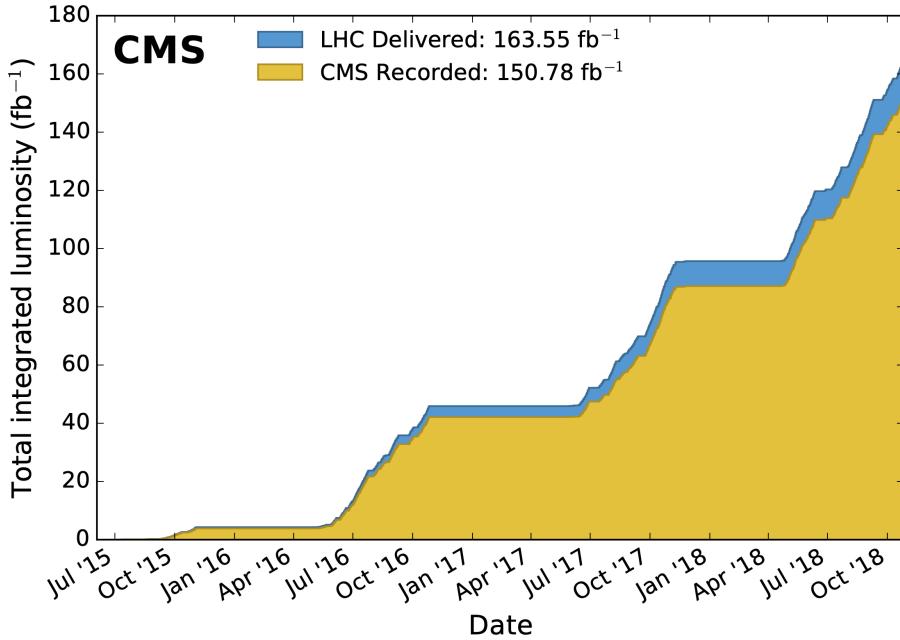


Figure 4.1: Cumulative delivered and recorded luminosity versus time for 2015-2018 at CMS, in proton-proton collision data only, at nominal center-of-mass energy [38].

1051 event generators, such as Powheg and MadGraph5\\_amCNLO [41] [42]. The informa-  
 1052 tion generated, e.g. the computation of the differential cross sections and kinematics  
 1053 of the final state particles, is saved in a compressed file and used to generate MC  
 1054 samples that are used in physics analyses. The samples are digitized using GEANT4  
 1055 [43], a platform used at the LHC and other facilities to comprehensively simulate the  
 1056 passage of particles through matter. The digitized samples are passed through the  
 1057 same detector reconstruction as real data events collected in the detector.

1058 The samples for modeling the signal ( $h \rightarrow aa \rightarrow 2b2\tau$  and  $h \rightarrow a_1a_2$ ) in the  
 1059 2HDM+S and TRSM are generated at tree-level, for a range of masses of the light  
 1060 neutral scalar  $a$ . For  $h \rightarrow aa$ , the mass hypotheses for the  $a$  range from  $m_a =$   
 1061 (12 GeV, 62.5 GeV). For  $h \rightarrow a_1a_2$ , the mass hypotheses for the two light scalars span  
 1062 combinations of  $m_{a1}$ ,  $m_{a2}$  ranging from (12 GeV, 62.5 GeV) for the two scalars.

## 1063 4.3 Embedded samples

1064 An important background for Higgs boson studies and searches for additional Higgs  
1065 bosons is the decay of  $Z$  bosons into pairs of  $\tau$  leptons ( $Z \rightarrow \tau\tau$ ). An embedded tech-  
1066 nique was developed in the context of Standard Model Higgs to  $\tau\tau$  measurements, to  
1067 model  $Z \rightarrow \tau\tau$  decays, and was expanded to also model all Standard Model processes  
1068 that contain  $\tau\tau$  [44]. The embedded technique has since been used successfully at  
1069 CMS for the Standard Model  $H \rightarrow \tau\tau$  measurement, as well as searches for minimal  
1070 supersymmetric extensions to the Standard Model (MSSM) [45] [46].

1071 Fig. 4.2 shows a schematic of how embedded samples are produced. Data events  
1072 containing  $Z \rightarrow \mu\mu$  decays are selected. In these events, all energy deposits of the  
1073 recorded muons are removed, and are replaced with simulated tau leptons with the  
1074 same kinematic properties as the removed muons. This results in a hybrid data format  
1075 containing information from both observed and simulated events, as illustrated in Fig.  
1076 4.2 [44].

1077 In the selection step of the embedded technique, events are selected with at least  
1078 one of a set of  $\mu\mu$  trigger paths, which require  $p_T > 17(8)$  GeV for the leading  
1079 (sub-leading) muons, and a minimum requirement between 3.8 and 8.0 GeV on the  
1080 invariant di-muon mass  $m_{\mu\mu}$  [44]. The offline reconstructed muons must match the  
1081 objects at trigger level and also have offline  $p_T > 17(8)$  GeV. They must have  $|\eta| < 2.4$   
1082 and be located at a distance  $|d_z| < 0.2$  cm to the primary vertex along the beam  
1083 axis. To form a  $Z$  boson candidate, each muon is required to originate from a global  
1084 muon track. The muon pairs must have opposite charges with an invariant mass of  
1085  $m_{\mu\mu} > 20$  GeV. If more than two di-muon pairs are found, the pair with the invariant  
1086 mass closest to the  $Z$  boson mass (91.19 GeV) is chosen.

1087 This selection is designed to be tight enough to ensure a high purity of genuine  
1088  $\mu\mu$  events, and also loose enough to minimize biases of the embedded event samples.  
1089 Isolation requirements are avoided, since they would introduce a bias towards less

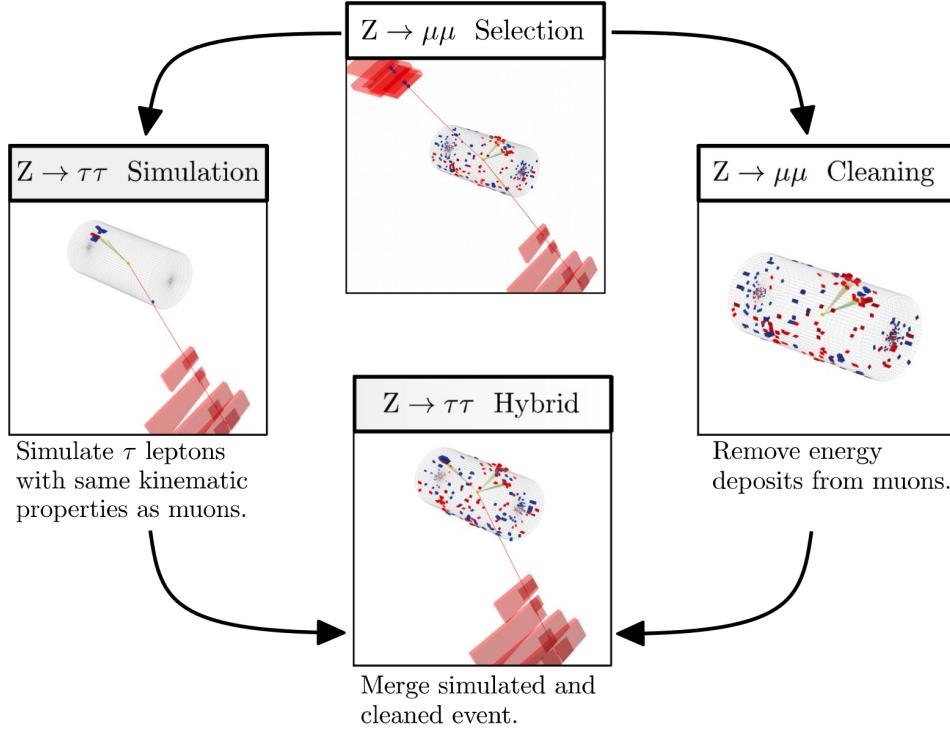


Figure 4.2: Schematic view of the four main steps of the embedding technique for  $\tau$  leptons, as described in Section 4.3 [44]. A  $Z \rightarrow \mu\mu$  event is selected in data ( $Z \rightarrow \mu\mu$  selection), all of the energy deposits associated with the muons are removed ( $Z \rightarrow \mu\mu$  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).

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 [44].  $Z \rightarrow \mu\mu$  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.

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 [44]. The simulation must be tuned extensively to accurately model aspects of the data, such as time-dependent pileup profiles, the production of additional jets, e.g. in multijet and vector boson fusion topologies, the number of reconstructed primary interaction

| Process                  | Fraction (%) |                               |                   |
|--------------------------|--------------|-------------------------------|-------------------|
|                          | Inclusive    | $m_{\mu\mu} > 70 \text{ GeV}$ | 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 [44], before additional cuts (i.e. inclusive, *column 2*), and after adding a requirement on the di-muon mass  $m_{\mu\mu} > 70 \text{ GeV}$  (*column 3*), or a requirement on the number of b-tag jets in the event (*column 4*).

1101 vertices, and the missing transverse momentum  $p_T^{\text{miss}}$ . Since all events with genuine  
 1102  $\tau\tau$  are estimated with samples made with the embedded technique (referred to as  
 1103 embedded samples from here on), events in Monte Carlo simulation with genuine  $\tau\tau$   
 1104 are not used, in order to avoid double-counting.

1105 **Chapter 5**

1106 **Object reconstruction and  
1107 corrections applied**

1108 In this chapter on object reconstruction and corrections, Section 5.1 reviews the  
1109 physical properties of the objects most pertinent to the analyses presented in this  
1110 work: taus ( $\tau$ ), muons ( $\mu$ ), electrons ( $e$ ), and jets, with a focus on jets originating from  
1111 b quarks (b-flavor jets), as well as the methodology used to reconstruct the particles  
1112 from their characteristic signatures in the CMS detector. Section 5.2 describes the  
1113 method used to reconstruct the invariant full  $\tau\tau$  mass which is used for the final signal  
1114 extraction. Lastly, Section 5.3 describes the corrections applied to the simulated  
1115 samples which improve their modeling of data.

1116 **5.1 Object reconstruction**

1117 **5.1.1 Taus**

1118 The tau ( $\tau$ ) is the heaviest known lepton. With a rest mass of 1776.86 MeV, it can  
1119 decay to not only electrons and muons, but also hadrons. The mean lifetime of the  $\tau$   
1120 is  $\tau = 290 \times 10^{-15}$  seconds, corresponding to  $c\tau = 87.03 \mu\text{m}$ , which is short enough

<sub>1121</sub> that taus decay in the CMS detector before reaching the detector elements.

<sub>1122</sub> In two thirds of the cases,  $\tau$  leptons decay hadronically, typically into one or three  
<sub>1123</sub> charged mesons (predominantly  $\pi^+$ ,  $\pi^-$ ), often accompanied by neutral pions (that  
<sub>1124</sub> decay  $\pi^0 \rightarrow \gamma\gamma$ ), and a  $\nu_\tau$ . These hadronic decays are denoted  $\tau_h$ . In the remainder of  
<sub>1125</sub> the decays, the tau decays to the lighter leptons (electron or muon), termed leptonic  
<sub>1126</sub> decays. In all cases, at least one neutrino is produced, resulting in missing transverse  
<sub>1127</sub> energy in the CMS detector. The tau's largest decay branching ratios (proportional  
<sub>1128</sub> to probability of decay) are listed below [24]:

<sub>1129</sub> • 17.8% decay to  $e^- \bar{\nu}_e \nu_\tau$

<sub>1130</sub> • 17.4% decay to  $\mu^- \bar{\nu}_\mu \nu_\tau$

<sub>1131</sub> • 25.5% decay to  $\pi^- \pi^0 \nu_\tau$  ( $\rho^-$  resonance at 770 MeV)

<sub>1132</sub> • 10.8% decay to  $\pi^- \nu_\tau$

<sub>1133</sub> • 9.3% decay to  $\pi^- \pi^0 \pi^0 \nu_\tau$  ( $a_1^-$  resonance at 1200 MeV)

<sub>1134</sub> • 9.0% decay to  $\pi^- \pi^- \pi^+ \nu_\tau$  ( $a_1^-$  resonance at 1200 MeV)

<sub>1135</sub> The neutrinos escape undetected from the CMS detector and are not considered  
<sub>1136</sub> in the reconstruction. Charged hadrons leave tracks in the tracking detector before  
<sub>1137</sub> being absorbed in the hadronic calorimeter; in CMS tau reconstruction terminology,  
<sub>1138</sub> they are often called “prongs”, i.e. the dominant  $\tau_h$  decay modes are termed “1 prong”  
<sub>1139</sub> ( $\pi^\pm$ ), “1 prong +  $\pi^0(s)$ ”, and “3-prong”. Neutral pions decay to two photons which  
<sub>1140</sub> lose their energy in the electromagnetic calorimeter. Taus that decay to electrons  
<sub>1141</sub> and muons, are typically triggered on and reconstructed as electrons and muons  
<sub>1142</sub> respectively.

1143 **Hadron plus strips (HPS) reconstruction of  $\tau_h$**

1144 At CMS, hadronically decaying tau leptons are reconstructed with the hadron plus  
1145 strips (HPS) algorithm [47] [48]. The HPS algorithm capitalizes on photon conversions  
1146 in the CMS tracker material, which originate from the neutral pion ( $\pi^0$ ) decaying  
1147 to two photons. The bending of electron/positron tracks due to the CMS solenoid  
1148 magnetic field leads to a spread of the neutral pions' calorimeter signatures in the  $\phi$   
1149 direction. This motivates the reconstruction of photons in “strips”: objects that are  
1150 built out of PF photons and electrons. The strip reconstruction starts with centering  
1151 a strip on the most energetic electromagnetic particle in a PF jet. Among other  
1152 electromagnetic particles located in a window of size  $\Delta\eta = 0.05$  and  $\Delta\phi = 0.20$   
1153 around the strip center, the most energetic one is associated with the strip and its  
1154 momentum is added to the strip momentum. This is repeated iteratively until no  
1155 further particles can be associated. Lastly, strips satisfying a requirement of  $p_T^{\text{strip}} > 1$   
1156 GeV are combined with charged hadrons to reconstruct individual  $\tau_h$  decay modes,  
1157 where  $h$  stands for both  $\pi$  and  $K$ :

1158 • *Single hadron:*  $h^-\nu_\tau$  and  $h^-\pi^0\nu_\tau$  decay modes, in which the neutral pions have  
1159 too little energy to be reconstructed as strips.

1160 • *One hadron + one strip:*  $h^-\pi^0\nu_\tau$  decay modes, where the photons from the  $\pi^0$   
1161 decay are close together in the calorimeter.

1162 • *One hadron + two strips:*  $h^-\pi^0\nu_\tau$  decay modes, where the photons from the  $\pi^0$   
1163 decay are well separated.

1164 • *Three hadrons:*  $h^-h^+h^-\nu_\tau$  decay modes. The three charged hadrons are re-  
1165 quired to originate from the same secondary vertex.

1166 The  $h^-\pi^0\pi^0\nu_\tau$  and  $h^-h^+h^-\pi^0\nu_\tau$  decay modes do not have their own treatment are  
1167 reconstructed with the above topologies.

1168 In the HPS algorithm, the direction of the reconstructed tau momentum  $\vec{p}_T^{\tau_h}$   
1169 is required to fall within a distance of  $\Delta R = 0.1$  from the original PF jet. All  
1170 charged hadrons and strips are required to be contained within a cone of size  $\Delta R =$   
1171  $(2.8 \text{ GeV})/p_T^{\tau_h}$ , from the  $\tau_h$  as reconstructed by the HPS.

1172 All charged hadrons are assumed to be pions, and they are required to be consis-  
1173 tent with the masses of the intermediate meson resonances (if applicable), with the  
1174 following allowed windows for candidates: 50-200 MeV for  $\pi^0$ , 0.3-1.3 GeV for  $\rho$ , and  
1175 0.8-1.5 GeV for  $a_1$ . If the  $\tau_h$  decay is compatible with more than one hypothesis, the  
1176 one giving the highest  $p_T^{\tau_h}$  is chosen. Lastly, an isolation requirement is applied: aside  
1177 from the  $\tau_h$  decay products, no charged hadrons or photons can be present within  
1178 an isolation cone of size  $\Delta R = 0.5$  around the direction of the  $\tau_h$ . The outputs of  
1179 the HPS algorithm are the reconstructed decay mode and the visible four-momentum  
1180 (i.e. the four-momenta of all decay products excluding the neutrinos).

### 1181 DeepTau for identifying $\tau_h$

1182 The identification of  $\tau_h$  candidates in CMS has historically been divided into separate  
1183 discriminators against jets, electrons, and muons. Discriminators versus jets and  
1184 electrons use information from derived quantities, such as the  $p_T$  sum of particles  
1185 near the  $\tau_h$  axis. Building on the previous multivariate analysis (MVA) classifier [49]  
1186 based on a boosted decision tree (BDT), DeepTau is a more recent classifier based on a  
1187 deep neural network (DNN) that simultaneously discriminates against jets, electrons,  
1188 and muons. The DNN uses a combination of high-level inputs, similar to previous  
1189 algorithms, and also uses convolutional layers in  $\eta\text{-}\phi$  space to process information  
1190 from all reconstructed particles near the  $\tau_h$  axis. Convolutional layers are based on  
1191 the principle that an image can be processed independently of its position.

1192 The final DeepTau discriminators against jets, muons, and electrons are given by

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

1193 where  $y_\tau$  ( $y_\alpha$ ) are estimates of the probabilities for the  $\tau_h$  candidate to come from  
1194 a genuine  $\tau_h$  (jet,  $\mu$ ,  $e$ ). Working points for each discriminator with different  $\tau_h$   
1195 identification efficiencies are defined for  $D_e$ ,  $D_\mu$ , and  $D_{\text{jet}}$ , for usage in physics analyses  
1196 and derivation of data-to-simulation corrections [50].

### 1197 5.1.2 Muons

1198 Muons are the next lightest lepton after taus, with a mass of 105.66 MeV and a  
1199 mean lifetime of  $\tau = 2.20 \times 10^{-6}$  seconds, or  $c\tau = 658.64$  m. At CMS, muons are  
1200 identified with requirements on the quality of the track reconstruction and on the  
1201 number of measurements in the tracker and the muon systems [51]. In the standard  
1202 CMS reconstruction, tracks are first reconstructed independently in the inner tracker  
1203 (tracker track) and in the muon system (standalone-muon track). Next, these tracks  
1204 are processed in two different methods.

1205 The first is Global Muon reconstruction (outside-in) [51], which fits combined hits  
1206 from the tracker track and standalone-muon track, using the Kalman-filter technique.  
1207 At large transverse momenta,  $p_T \gtrsim 200$  GeV, the global-muon fit can improve the  
1208 momentum resolution compared to the tracker-only fit.

1209 The second is Tracker Muon reconstruction (inside-out) [51], which starts with  
1210 tracker tracks with  $p_T > 0.5$  GeV and total momentum  $p_T > 2.5$  GeV. These tracks  
1211 are extrapolated outwards to the muon system and matched to detector segments  
1212 there, taking into account the magnetic field, expected energy losses, and multiple  
1213 Coulomb scattering in the detector material. Tracker Muon reconstruction is more  
1214 efficient than the Global Muon reconstruction at low momenta,  $p \lesssim 5$  GeV, because

it only requires a single muon segment in the muon system, whereas Global Muon reconstruction typically requires segments in at least two muon stations.

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

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

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

### 5.1.3 Electrons

Electrons are the lightest lepton with a mass of 0.511 MeV. At CMS, electrons are reconstructed by associating a track reconstructed in the silicon tracking detector with a cluster of energy in the ECAL. Performance is maximized via a combination of a stand-alone approach and the complementary global particle-flow approach [52].

In the stand-alone approach, the electron energy, which is typically spread over several crystals of the ECAL, is clustered with the “hybrid” algorithm in the barrel and the “multi- $5 \times 5$ ” in the endcaps [52]. The hybrid algorithm collects energy in a

1238 small window in  $\eta$  and an extended window in  $\phi$ . It identifies a seed crystal, and adds  
1239 arrays of  $5 \times 1$  crystals in  $\eta \times \phi$  in a range of  $N = 17$  crystals in both directions of  
1240  $\phi$ , if their energies exceed a minimum threshold, thus forming a supercluster (SC). In  
1241 the endcap, crystals are not arranged in an  $\eta \times \phi$  geometry; instead clusters are build  
1242 around seed crystals in clusters of  $5 \times 5$  crystals that can partly overlap. Nearby  
1243 clusters are grouped into a supercluster, and energy is recovered from associated  
1244 deposits in the preshower.

1245 In the PF reconstruction [52], PF clusters are reconstructed by aggregating around  
1246 a seed all contiguous crystals with energies two standard deviations above the elec-  
1247 tronic noise observed at the beginning of a data-taking run. The energy of a given  
1248 crystal can be shared among two or more clusters.

1249 The electron track reconstruction is performed in two ways [52]: the ECAL-based  
1250 seeding, which begins with the SC energy and positioning, and the tracker-based  
1251 seeding (part of the PF reconstruction algorithm), which uses tracks reconstructed  
1252 from the general algorithm for charged particles, extrapolated towards the ECAL and  
1253 matched to an SC. Kalman filter (KF) tracks with a small number of hits or that are  
1254 not well-fitted, are re-fitted with a dedicated Gaussian sum Filter (GSF).

1255 A global identification variable [52] is defined using a multivariate analysis (MVA)  
1256 technique that combines information on track observables (kinematics, quality of the  
1257 KF track and GSF track), the electron PF cluster observables (shape and pattern),  
1258 and the association between the two (geometric and kinematic observables). For  
1259 electrons seeded only through the tracker-based approach, a weak selection is applied  
1260 on this MVA variable. For electrons seeded through both approaches, a logical OR is  
1261 taken.

1262 Electron isolation, i.e. the presence of energy deposits near the electron trajectory,  
1263 is a separate key handle in rejecting significant background. Compared to isolated  
1264 electrons, electrons from misidentified jets or genuine electrons within a jet resulting

1265 from semileptonic decays of  $b$  or  $c$  quarks tend to have significant energy deposits  
1266 near the primary trajectory [52]. Offline analyses benefit from the PF technique  
1267 for defining isolation, which sums the PF candidates reconstructed located within a  
1268 specified isolation cone around the electron candidate, as in Eqn. 5.2.

### 1269 5.1.4 Jets

1270 The vast majority of processes of interest at the LHC contains quarks or gluons in  
1271 the final state, but these particles cannot be observed directly. In a process called  
1272 hadronization, they fragment into spatially-grouped collections of particles called jets,  
1273 which can be detected in the tracking and calorimeter systems. Hadronization and  
1274 the subsequent decays of unstable hadrons can produce hundreds of nearby particles  
1275 in the CMS detector. Jets are reconstructed by the PF algorithm (PF jets), or from  
1276 the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo  
1277 jets). In PF jets, typically used in offline analyses, jets are built using the anti- $k_T$   
1278 (AK) clustering algorithm [53]. The anti- $k_T$  algorithm iterates over particle pairs and  
1279 finds the two that are closest in a distance measure  $d$ , and determines whether to  
1280 combine them:

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

1281 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-  
1282 ity, and azimuthal angle of particle  $i$ . The power  $-2$  means that higher-momentum  
1283 particles are clustered first, leading to jets that tend to be centered on the hardest  
1284 (highest  $p_T$ ) particle.

1285 There are several methods to remove contributions of pileup collisions from jet  
1286 clustering [54]:

- 1287 • Charged hadron subtraction (CHS), which removes all charged hadron candi-

1288 dates associated with a track that is not associated with the primary vertex.

- 1289 • PileUp Per Particle Identification (PUPPI), which weighs input particles based  
1290 on their likelihood of arising from pileup. QCD particles tend to have a collinear  
1291 structure, compared to soft diffuse radiation coming from pileup. The local  
1292 shape for charged pileup, used as a proxy for all pileup particles, is used on an  
1293 event-by-event basis to calculate a weight for each particle. PUPPI is deployed  
1294 in Run-2 and is more performant than CHS in high pileup scenarios.

1295 **5.1.5 B-flavored jets**

1296 Jets that arise from bottom-quark hadronization (b-flavor jets) have overwhelming  
1297 background from processes involving jets from gluons (g) and light-flavor quarks (u, d,  
1298 s), and from c-quark fragmentation. The ability to identify b-flavor jets, or b-tagging,  
1299 exploits the b hadrons' relatively large masses, long lifetimes, and daughter particles  
1300 with hard momentum spectra [53].

1301 The impact parameter (IP) of a track is the 3-dimensional distance between the  
1302 track and the primary vertex (PV) at the point of closest approach. The IP is positive  
1303 if the track originates from the decay of particles travelling along the jet axis. The  
1304 resolution of the IP depends on the  $p_T$  and  $\eta$  of the track, motivating the use of the  
1305 impact parameter significance  $S_{\text{IP}}$  (ratio of the IP to its estimated uncertainty) as an  
1306 observable [53].

1307 Because of the large but finite lifetimes of the b hadrons, b hadrons tend to  
1308 travel a short distance before decaying at a secondary vertex (SV), which can be  
1309 measured and reconstructed separately from the primary vertex due to the excellent  
1310 position resolution of the pixel detector [53]. Previous b-tagging algorithms (e.g.  
1311 CSV, cMVAv2, and DeepCSV) have capitalized on variables such as the presence of  
1312 a SV, the flight distance and direction (computed from the vector between the PV  
1313 and the SV), and kinematics of the system of associated secondary tracks (e.g. track

1314 multiplicity, mass, and energy).

1315 The DeepJet (formerly known as DeepFlavour) algorithm [55] is a deep-neural-  
1316 network multi-classification algorithm, which uses 16 properties of up to 25 charged  
1317 and 6 properties of 25 neutral particle-flow jet constituents, as well as 17 properties  
1318 from up to 4 secondary vertices associate with the jet. Compared to the previous clas-  
1319 sifying algorithm DeepCSV, DeepJet has been demonstrated to have higher efficiency  
1320 with lower misidentification probability in Phase-1 data [56].

## 1321 5.2 Reconstruction of the $\tau\tau$ mass

1322 The final signal extraction is done to the total  $\tau\tau$  mass, which is estimated from the  
1323 visible  $\tau\tau$  mass using the FastMTT algorithm [57]. FastMTT is based on the SVFit  
1324 algorithm, originally developed for the Standard Model  $H \rightarrow \tau\tau$  analysis [58]. Both  
1325 the SVFit algorithms, and the FastMTT algorithm, are described below, to give a  
1326 complete picture of how tau decays are parameterized.

1327 To specify a hadronic  $\tau$  decay, six parameters are needed [58]: the polar and  
1328 azimuthal angles of the visible decay product system in the  $\tau$  rest frame, the three  
1329 boost parameters from the  $\tau$  rest frame to the laboratory frame, and the invariant  
1330 mass  $m_{\text{vis}}$  of the visible decay products. For a leptonic  $\tau$  decay, two neutrinos are  
1331 produced, and a seventh parameter, the invariant mass of the two-neutrino system, is  
1332 necessary. The unknown parameters are constrained by four observables that are the  
1333 components of the four-momentum of the system formed by the visible decay products  
1334 of the  $\tau$  lepton, measured in the laboratory frame. The remaining unconstrained  
1335 parameters for hadronic and leptonic  $\tau$  decays are thus:

- 1336 • The fraction of the  $\tau$  energy in the laboratory frame carried by the visible decay  
1337 products,
- 1338 •  $\phi$ , the azimuthal angle of the  $\tau$  direction in the laboratory frame,

- 1339       •  $m_{\nu\nu}$ , the invariant mass of the two-neutrino system in leptonic  $\tau$  decays (for  
 1340           hadronic  $\tau$  decays,  $m_{\nu\nu}$  is set to 0).

1341      $E_x^{\text{miss}}$  and  $E_y^{\text{miss}}$ , the  $x$  and  $y$  components of the missing transverse energy  $\vec{E}_T^{\text{miss}}$   
 1342     provide two further constraints.

### 1343     **5.2.1 Original SVFit “standalone”: maximum likelihood**

1344     In one of the original versions of SVFit, called “standalone” SVFit [58], a maximum  
 1345     likelihood fit method is used to reconstruct the mass  $m_{\tau\tau}$  by combining the measured  
 1346     observables  $E_x^{\text{miss}}$  and  $E_y^{\text{miss}}$  with a likelihood model that includes terms for the  $\tau$   
 1347     decay kinematics and the  $\vec{E}_T^{\text{miss}}$  resolution [58]. The likelihood function  $f(\vec{z}, \vec{y}, \vec{a}_1 \vec{a}_2)$   
 1348     of the parameters  $\vec{z} = (E_x^{\text{miss}}, E_y^{\text{miss}})$  in an event is constructed, where the remaining  
 1349     parameters are the kinematics of the two  $\tau$  decays, denoted  $\vec{a}_1 = (x_1, \phi_1, m_{\nu\nu,1})$  and  
 1350      $\vec{a}_2 = (x_2, \phi_2, m_{\nu\nu,2})$ , and the four-momenta of the visible decay products with the  
 1351     measured values  $\vec{y} = (p_1^{\text{vis}}, p_2^{\text{vis}})$ .

1352     The likelihood  $f$  is the product of three likelihood functions. The first two likelihood  
 1353     functions model the decay parameters  $\vec{a}_1$  and  $\vec{a}_2$  of the two  $\tau$  leptons. For leptonic  
 1354     decays, the likelihood function is modeled using matrix elements for  $\tau$  decays,  
 1355     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  
 1356     hadronic  $\tau$  decays, a model based on the two-body phase space is used and integrated  
 1357     over  $m_{\text{vis}}^2/m_{\tau\tau}^2 \leq x \leq 1$ . The third likelihood function quantifies the compatibility of  
 1358     a  $\tau$  decay hypothesis with the reconstructed  $\vec{E}_T^{\text{miss}}$  in an event, assuming the neutrinos  
 1359     are the only source of missing transverse energy. The expected  $\vec{E}_T^{\text{miss}}$  resolution  
 1360     is represented by a covariant matrix, estimated on an event-by-event basis using a  
 1361     significance algorithm [59].

## 1362 5.2.2 “Classic SVFit” with matrix element

1363 Classic SVFit is an improved algorithm of the original “standalone” SVFit using the  
 1364 formalism of the matrix element (ME) method [57]. In the ME method, an estimate  
 1365 for the unknown model parameter  $\Theta$  (here, the mass  $m_{\tau\tau}$ ) is obtained by maximizing  
 1366 the probability density  $\mathcal{P}$ . The key ingredients of the probability density are the  
 1367 squared modulus of the matrix element  $|\mathcal{M}(\mathbf{p}, \Theta)|^2$  and the transfer function  $W(\mathbf{y}|\mathbf{p})$   
 1368 (probability density to observe the measured observables  $\mathbf{y}$  given the phase space  
 1369 point  $\mathbf{p}$ ). The best estimate  $m_{\tau\tau}$  is obtained by computing the probability density  $\mathcal{P}$   
 1370 for a range of mass hypotheses and finding the value of  $m_{\tau\tau}$  that maximizes  $\mathcal{P}$ .

1371 Distributions illustrating the performance of the classic matrix element SVFit  
 1372 algorithm are shown in Fig. 5.1 from [57], showing the di-tau mass after and before  
 1373 application of SVFit to recover energy lost to neutrinos. The SVFit algorithm is  
 1374 found to improve the sensitivity of the Standard Model  $H \rightarrow \tau\tau$  analysis performed  
 1375 by CMS by about 30%, compared to performing the same analysis using only the  
 1376 visible mass  $m_{\text{vis}}$ .

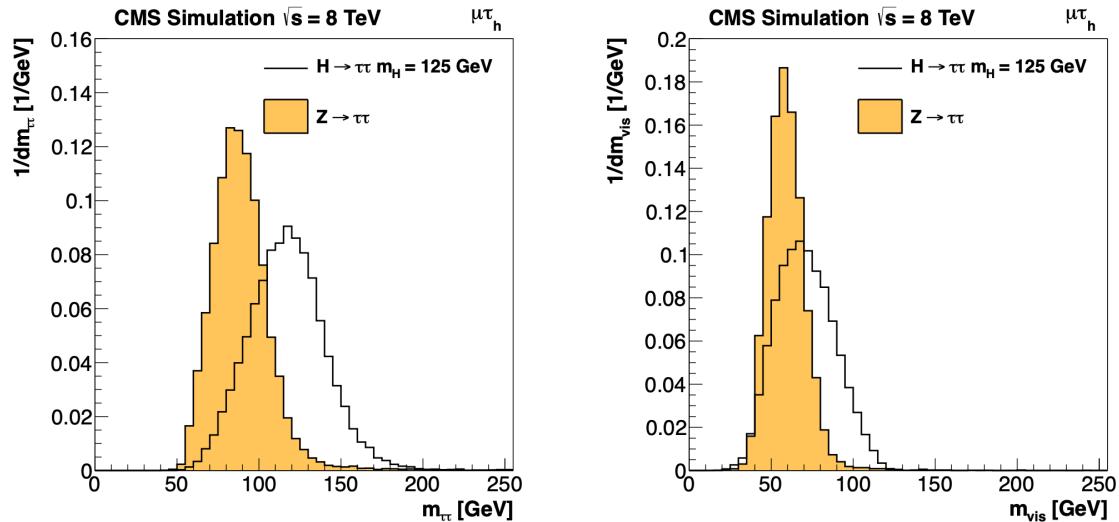


Figure 5.1: Distributions from [57], of  $m_{\tau\tau}$  after reconstruction with the original SVFit algorithm (*left*), and before SVFit with only the visible tau decay products (*right*), for  $H \rightarrow \tau\tau$  signal events of mass  $m_H = 125$  GeV (*black line*) and the  $Z/\gamma^* \rightarrow \tau\tau$  background (*orange, solid*), in the decay channel  $\tau\tau \rightarrow \mu\tau_h$ .

1377 **5.2.3 FastMTT: optimized SVFit**

1378 FastMTT [60] is a further simplification to the matrix element method of Classic  
1379 SVFit which has comparable performance but is about 100 times faster. FastMTT  
1380 drops the matrix element component of the computation without significant impact  
1381 on the final mass resolution, and simplifies the computation of the transfer functions.  
1382 The opening angle of the  $\tau$  decay products with respect to the initial  $\tau$  momenta ap-  
1383 proaches 0 for  $\tau$  with high  $\gamma = E_\tau/m_\tau$ , with typical  $\tau$  decays from the Z boson decays  
1384 already satisfying this condition. In this collinear approximation, the dimensionality  
1385 of the transfer function can be reduced in the computation of FastMTT, while still  
1386 yielding similar results to Classic SVFit [60].

1387 **5.3 Corrections applied to simulation**

1388 Corrections are applied to simulated samples to account for known effects in the event  
1389 modeling and reconstruction and data-taking, and are intended to bring simulations  
1390 in closer agreement with data. Corrections fall into two broad categories: *energy*  
1391 *scale corrections* applied to physics objects, and *event-level corrections*. Energy scale  
1392 corrections are multiplicative factors applied to the energy and transverse momentum  
1393  $p_T$  of simulated objects (e.g. leptons or jets), and bring the average reconstructed en-  
1394 ergies of simulated particles into better agreement with those of objects reconstructed  
1395 from data. Event-level corrections are applied as a per-event multiplicative weight,  
1396 and account for effects such as mis-modeling in simulations of the underlying physics  
1397 process, or changing detector operating conditions during data-taking. Event-level  
1398 corrections change the shapes of the distributions of all the physical observables.

1399 Uncertainties in scale factors and corrections are also sources of systematic errors  
1400 in the analysis, detailed in Chapter 8. Systematic uncertainties in the tau, muon, and  
1401 electron energy scales can shift the  $p_T$  of the leptons up or down, which can change

1402 whether events pass or fail the offline  $p_T$  thresholds for the trigger paths described in  
1403 the previous section, i.e. change the number of events in the signal region.

### 1404 5.3.1 Tau energy scale

1405 An energy scale is applied to the transverse momentum  $p_T$  and mass of the hadronic  
1406 tau  $\tau_h$  in the  $\mu\tau_h$  and  $e\tau_h$  channels, to correct for a deviation of the average recon-  
1407 structed  $\tau_h$  energy from the generator-level energy of the visible  $\tau_h$  decay products.  
1408 These correction factors are derived centrally [49], by fitting to events in  $e\tau_h$  and  $\mu\tau_h$   
1409 final states in  $Z/\gamma^*$  events separately for the  $h^\pm$ ,  $h^\pm\pi^0$ , and  $h^\pm h^\mp h^\pm$  decays. The  
1410 values used are shown in Table 5.1.

1411 When applying the energy scale to the  $\tau_h$ , the 4-momentum of the missing trans-  
1412 verse energy (MET) is adjusted such that the total 4-momenta of the  $\tau_h$  and the MET  
1413 remains unchanged [61].

| Tau energy scale factor |                   |                   |                   |                   |
|-------------------------|-------------------|-------------------|-------------------|-------------------|
| Decay mode              | 2018              | 2017              | 2016 pre-VFP      | 2016 post-VFP     |
| 0                       | $0.991 \pm 0.008$ | $0.986 \pm 0.009$ | $0.987 \pm 0.01$  | $0.993 \pm 0.009$ |
| 1                       | $1.004 \pm 0.006$ | $0.999 \pm 0.006$ | $0.998 \pm 0.006$ | $0.991 \pm 0.007$ |
| 10                      | $0.998 \pm 0.007$ | $0.999 \pm 0.007$ | $0.984 \pm 0.008$ | $1.001 \pm 0.007$ |
| 11                      | $1.004 \pm 0.009$ | $0.996 \pm 0.01$  | $0.999 \pm 0.011$ | $0.997 \pm 0.016$ |

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.

### 1414 5.3.2 Muon energy scale

1415 An energy scale is applied to the  $p_T$  and mass of genuine muons from  $\tau$  decays in the  
1416  $e\mu$  and  $\mu\tau_h$  channels [62]. The applied values are the same for MC and embedded  
1417 samples and are shown in Table 5.2. Following the SM  $H \rightarrow \tau\tau$  analysis, Rochester  
1418 corrections are not applied, and instead prescriptions from [63] 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 [64] [63].

### 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 [65]. 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 [66]. For embedded samples the electron energy scale is taken as only binned in  $\eta$  (Table 5.3).

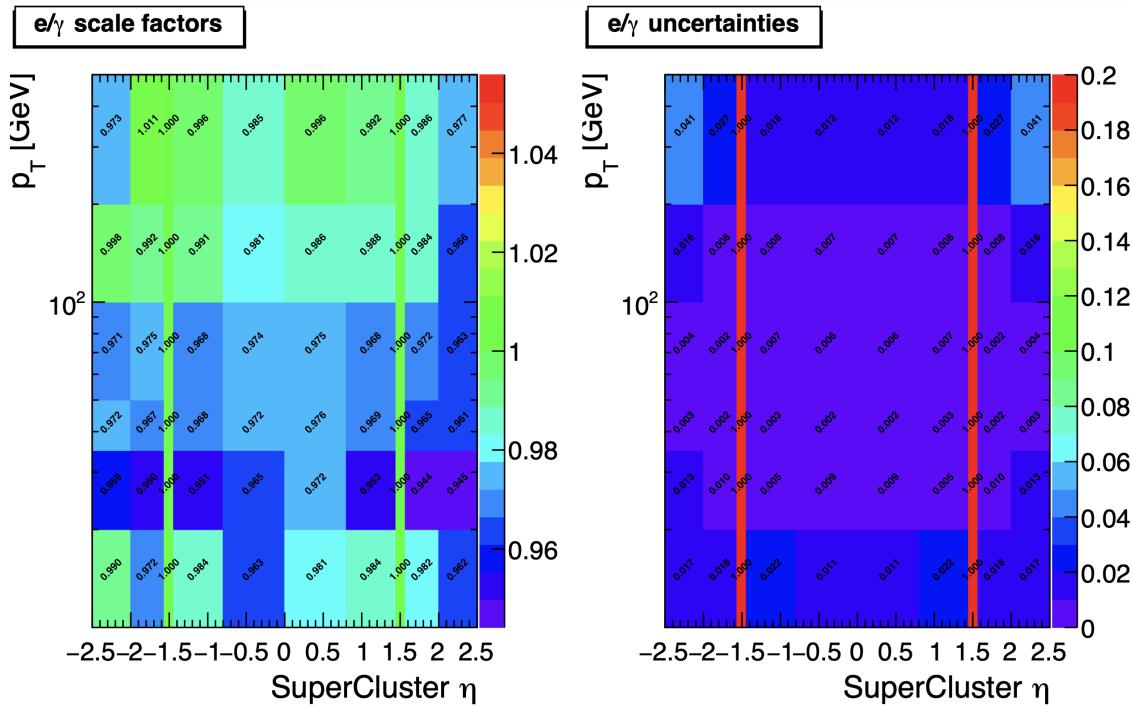


Figure 5.2: Electron/photon energy scale factors (*left*) and corresponding uncertainties (*right*) binned in the electron  $\eta$  and  $p_T$ , for the data-taking year 2018 [66].

| 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 [67] [68] [69].

### 1425 5.3.4 $\tau_h$ identification efficiency

1426 The  $\tau_h$  identification efficiency can differ in data and MC [61]. Recommended correc-  
 1427 tions are provided by the Tau POG, and we use the medium DeepTau vs. jet working  
 1428 point values. The identification efficiency is measured in  $Z \rightarrow \tau\tau$  events in the  $\mu\tau_h$   
 1429 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  $p_T$  [61].

### 1430 5.3.5 Trigger efficiencies

1431 Scale factors are applied to correct for differences in trigger efficiencies between MC  
 1432 and embedded vs. data, with values taken from tools provided by the Standard Model  
 1433  $H \rightarrow \tau\tau$  working group which uses the same trigger paths [64]. In the following  
 1434 sections we review relevant trigger efficiencies in data, which form the basis of the  
 1435 trigger efficiency corrections applied to MC and embedded.

### 1436 5.3.6 Tau trigger efficiencies

1437 The efficiencies in data of the single- $\tau_h$  leg in  $\mu\tau_h$ ,  $e\tau_h$ , and di- $\tau_h$  triggers is computed  
 1438 centrally per using a Tag and Probe (TnP) method [70] which is outlined here. In

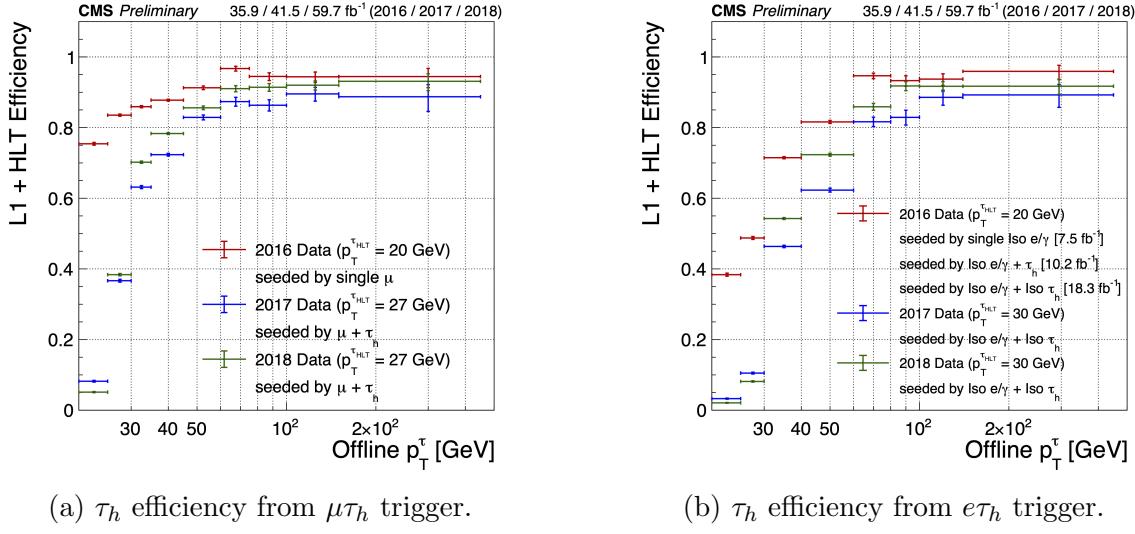
1439 this method,  $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$  are selected in data and a Drell-Yan simulated sample  
 1440 ( $Z \rightarrow \ell\ell, \ell = e, \mu, \tau_h$ ) with high purity. Cuts are applied to reject events not in this  
 1441 final state, e.g. suppressing  $Z \rightarrow \mu\mu$  by vetoing events with a single loose ID muon.  
 1442 An isolated muon candidate (the tag) with online  $p_T > 27$  GeV and  $|\eta| < 2.1$  is  
 1443 identified and matched to an offline  $\mu$ . An offline  $\tau_h$  candidate (the probe) is selected,  
 1444 which is separated from the tag  $\mu$ , and has  $p_T > 20$  GeV and  $|\eta| < 2.1$ . The probe  
 1445  $\tau_h$  must pass anti-muon and anti-electron discriminators to avoid fakes from muons  
 1446 and electrons, and must pass the medium MVA tau isolation to suppress fakes from  
 1447 QCD jets. The trigger efficiency in the TnP method is calculated as

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

1448 The efficiencies for the hadronic tau legs in the relevant channels of this analyses  
 1449 ( $\mu\tau_h$  and  $e\tau_h$ ) as a function of the offline tau  $p_T$  and  $\eta$ , are shown for data taken in  
 1450 2016, 2017, and 2018 in Figures 5.3a and 5.3b [70] [71]. In both figures, the different  
 1451 HLT thresholds and differences in the L1 seed result in higher efficiencies in 2016 and  
 1452 differences in shapes of the 2016 efficiencies compared to 2017 and 2018. The low  
 1453 pileup in 2016 also leads to higher efficiencies in that year.

### 1454 5.3.7 Single muon trigger efficiencies

1455 The efficiencies for the single isolated muon trigger with  $p_T > 24$  GeV used in this  
 1456 analysis, is shown for the data-taking year 2018 in Fig. 5.4a as a function of the muon  
 1457  $p_T$  and as a function of the muon  $|\eta|$  in Fig. 5.4b from [72]. The data is split with  
 1458 respect to a HLT muon reconstruction update that was deployed on 15/05/2018. A  
 1459 small asymmetry in efficiencies between negative and positive  $\eta$  in Fig. 5.4b is due to  
 1460 disabled muon chambers (CSCs). The efficiencies shown are estimated using a Tag  
 1461 and Probe method using  $Z \rightarrow \mu\mu$  events, with the tag being an offline muon with



(a)  $\tau_h$  efficiency from  $\mu\tau_h$  trigger. (b)  $\tau_h$  efficiency from  $e\tau_h$  trigger.

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 [71]. HLT  $p_T$  thresholds and L1 seeds are indicated in the legends.

<sup>1462</sup>  $p_T > 29$  GeV and  $|\eta| < 2.4$  passing a tight ID criteria, and the probe is an online (L1)  
<sup>1463</sup> trigger object with  $\Delta R < 0.3$  and passing tight ID and Particle Flow based isolation  
<sup>1464</sup> requirements with  $p_T > 26$  GeV.

### <sup>1465</sup> 5.3.8 Single electron trigger efficiencies

<sup>1466</sup> The efficiencies in data, and the ratio between data and MC, of the single electron  
<sup>1467</sup> HLT trigger with  $p_T$  threshold 32 GeV used in this analysis are shown for 2018, as  
<sup>1468</sup> a function of the electron  $p_T$  in Fig. 5.5a and of the electron  $|\eta|$  in Fig. 5.5b, from  
<sup>1469</sup> [73]. In the Tag and Probe method used for the 2018 dataset, the tag is an offline  
<sup>1470</sup> reconstructed electron with  $|\eta| \leq 2.1$  and not in the barrel and endcap overlap region,  
<sup>1471</sup> with  $p_T > 35$  GeV with tight isolation and shower shape requirements, firing the tag  
<sup>1472</sup> trigger. The probe is an offline reconstructed electron with  $|\eta| \leq 2.5$  with  $E_T^{\text{ECAL}} > 5$   
<sup>1473</sup> GeV with no extra identification criteria [73].

<sup>1474</sup> The disagreement between data and MC, particularly at low transverse momen-  
<sup>1475</sup> tum, is in part due to detector effects that are difficult to simulate, such as crystal

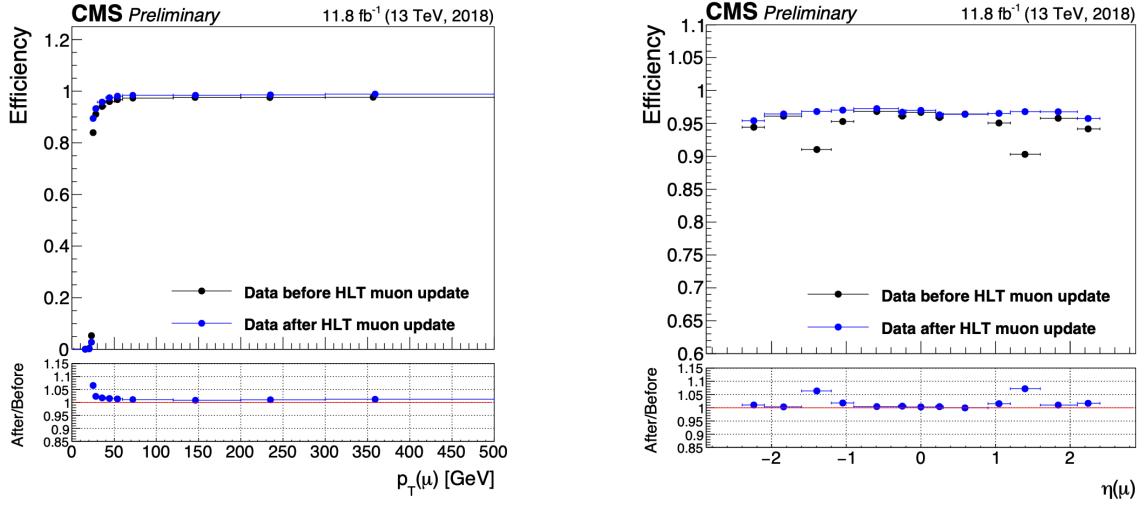
(a) Muon efficiency vs  $p_T$  for SingleMuon.(b) Muon efficiency vs  $|\eta|$  for SingleMuon.

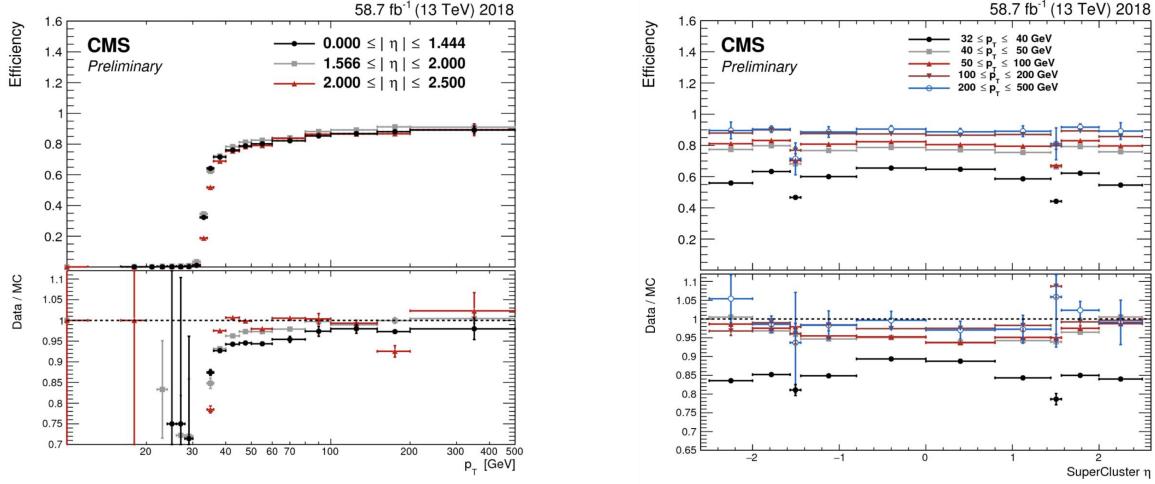
Figure 5.4: Trigger efficiencies in data (*top panels*) and ratio of efficiencies after/before a HLT muon reconstruction update (*bottom panels*) for the muon in the isolated single muon trigger with threshold  $p_T > 24$  GeV in the data-taking year 2018, as functions of the muon  $p_T$  (*left*) and muon  $|\eta|$  (*right*). Only statistical errors are shown [72].

<sup>1476</sup> transparency losses in the ECAL and the evolution of dead regions in the pixel tracker  
<sup>1477</sup> [73].

### <sup>1478</sup> 5.3.9 $e\mu$ cross-trigger efficiencies

<sup>1479</sup> The efficiencies of the electron and muons for the cross-trigger with leading muon  
<sup>1480</sup> used in the  $e\mu$  channel are shown for data in 2016, 2017, and 2018 in Figures 5.6a and  
<sup>1481</sup> 5.6b [74]. These efficiencies were measured centrally using a Tag and Probe in events  
<sup>1482</sup> with  $Z$  to dileptons with the same flavor and opposite charge, where the tags are an  
<sup>1483</sup> isolated muon or electron, and the probe (offline) candidate is required to satisfy the  
<sup>1484</sup> same lepton selection as that of the tag candidate, be matched within  $\Delta R < 0.1$  with  
<sup>1485</sup> a corresponding online trigger object, and also to pass the cross-trigger. The trigger  
<sup>1486</sup> efficiency is then:

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



(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) [73]. 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.

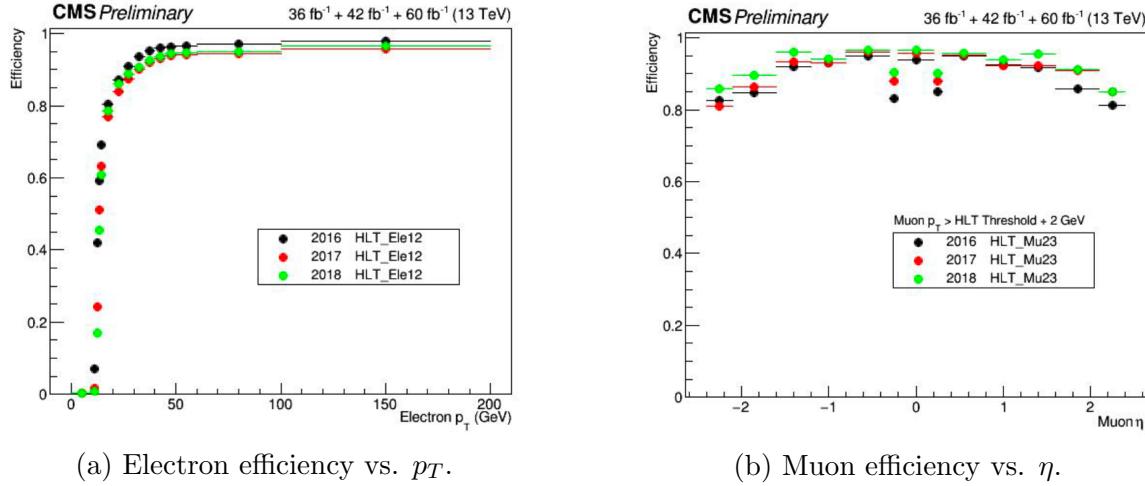
### 1487 5.3.10 Electrons and muons faking $\tau_h$ : energy scales

1488 Energy scales for electrons misidentified as hadronic tau decays ( $e$  faking  $\tau_h$ ) are  
 1489 provided by the Tau POG, and were measured in the  $e\tau_h$  channel with the visible  
 1490 invariant mass of the electron and hadronic tau system [64]. This energy scale is  
 1491 applied for  $\tau_h$  with  $p_T > 20$  GeV regardless of which DeepTau vs. electron working  
 1492 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 |
| 0   | 1.01362 (+0.00474, -0.00904)        |
| 1   | 1.01945 (+0.01598, -0.01226)        |
| 10  | 0.96903 (+0.0125, -0.03404)         |
| 11  | 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  $\tau_h$  [64].

1493 No nominal energy scale is applied for muons mis-reconstructed as  $\tau_h$ , and the



(a) Electron efficiency vs.  $p_T$ .

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

Figure 5.6: Efficiencies of the electron leg vs.  $p_T$  (left) and the muon leg 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) [74].

uncertainty is treated as  $\pm 1\%$  and uncorrelated in the reconstructed decay mode [64].

### 5.3.11 Electrons and muons faking $\tau_h$ : misidentification efficiencies

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

The specific values depend on the vs. electron and vs. muon discriminator working points used. For misidentified  $\mu \rightarrow \tau_h$ , the scale factors are split into different  $|\eta|$  regions, determined by the CMS muon and tracker detector geometries, as shown in Table 5.6 for 2018 [61].

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

| 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 efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018, binned in the muon  $|\eta|$  [61].

| Tau ID efficiency 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 efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018, binned in the electron  $|\eta|$  [61].

### 5.3.12 Electron ID and tracking efficiency

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

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

The tracking efficiencies in data and the data/MC ratio are shown in Fig. 5.7b for the Gaussian-sum filter (GSF) tracking [75].

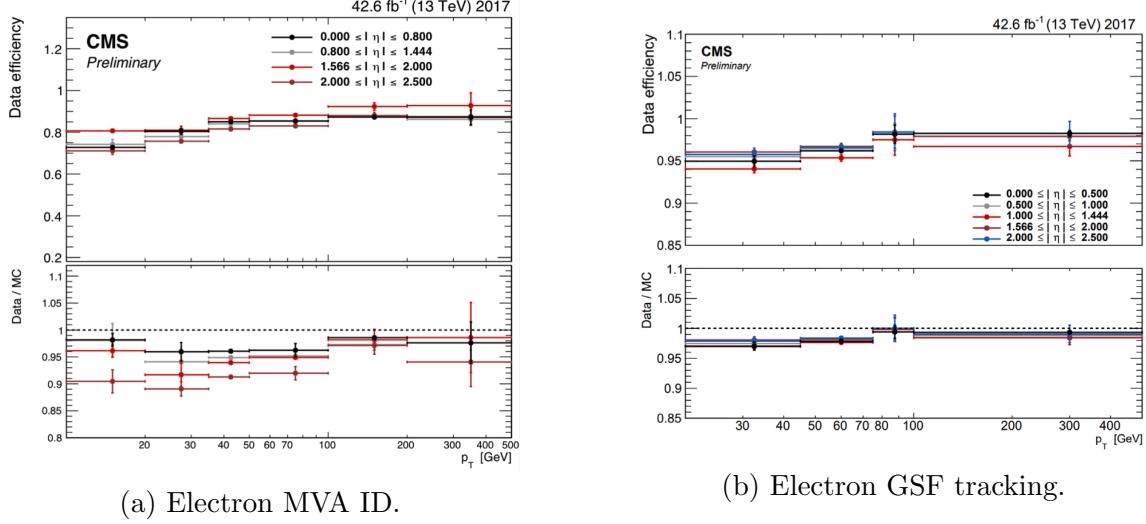


Figure 5.7: Efficiencies in data (*top panels*) and the ratio of efficiencies in data/MC (*bottom panels*), for the electron multivariate analysis (MVA) identification (*left*) and for the Gaussian-sum filter (GSF) tracking (*right*) [75]. Error bars represent statistical and systematic uncertainties.

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

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

The efficiencies for muon identification measured in 2015 data and MC simulation are shown in Figures 5.8a and 5.8b for the loose ID and tight ID respectively [76]. The loose ID is chosen such that efficiency exceeds 99% over the full  $\eta$  range, and the data and simulation agree to within 1%. The tight ID is chosen such that efficiency varies between 95% and 99% as a function of  $\eta$ , and the data and simulation agree to within 1-3%. The muon identification working point used in this analysis is the medium ID, which has an efficiency of 98% for all  $\eta$  and an agreement within 1-2% [76].

The efficiencies in data for the muon isolation, as measured in Level-3 muons (muons in one of the final stages of reconstruction in the HLT), as a function of the muon  $p_T$  and  $|\eta|$  are shown in Figures 5.9a and 5.9b [76]. The HLT muon reconstruction consists of two steps: Level-2 (L2), where the muon is reconstructed in the muon

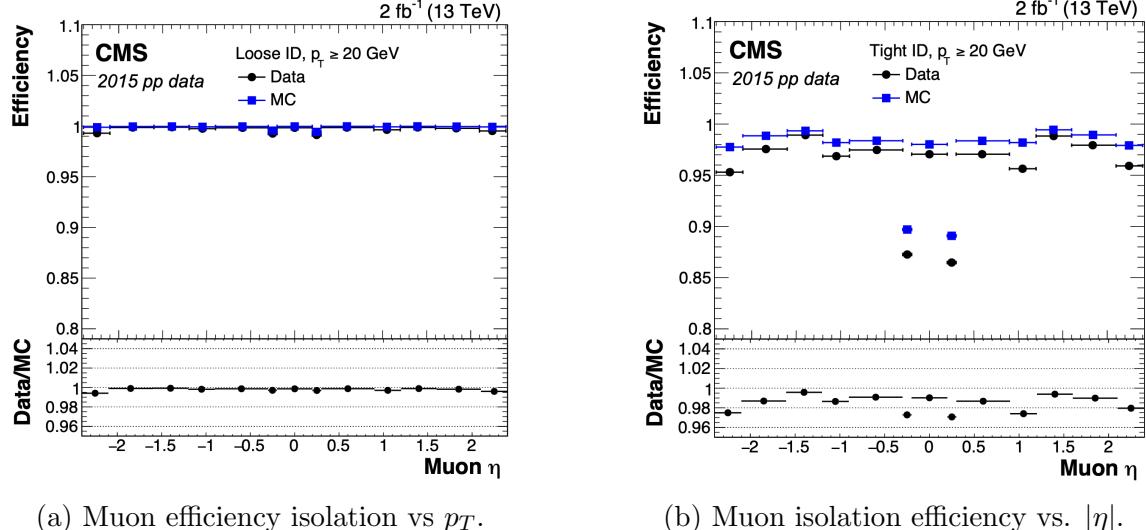


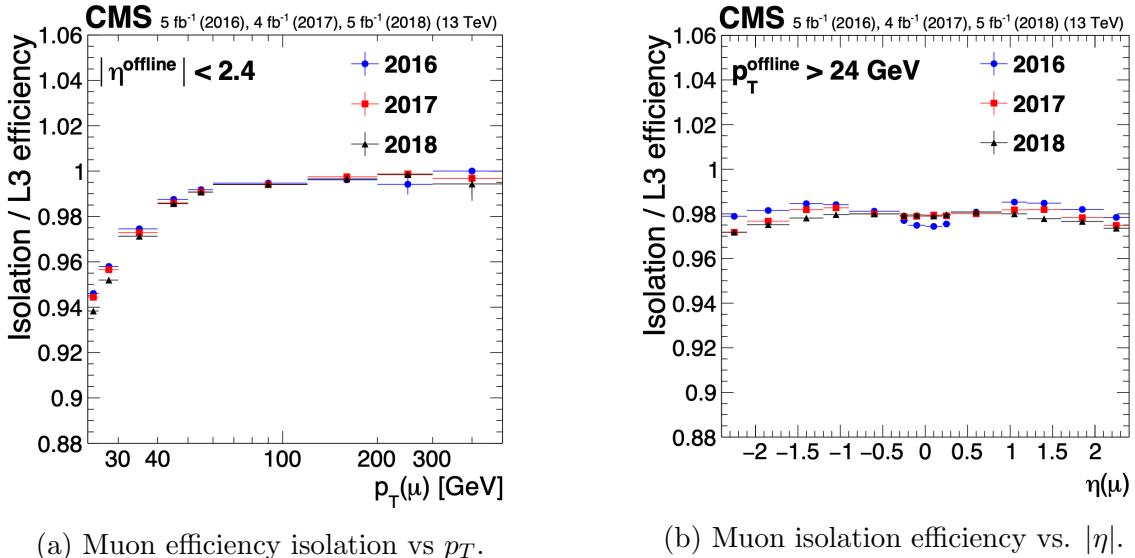
Figure 5.8: Muon identification efficiencies in 2015 data and MC as a function of the muon  $p_T$  for the loose ID (*left*) and tight ID (*right*) working points [76].

subdetectors only, and Level-3 (L3) which is a global fit of tracker and muon hits (i.e. the global muon reconstruction as described in Section 5.1.2) [77].

The muon tracking efficiencies as a function of  $|\eta|$  for standalone muons (i.e. tracks from only the muon system, i.e. DT, CSC, and RPC, as discussed in Section 5.1.2), is shown for data and simulated Drell-Yan samples in Fig. 5.10 [78].

### 5.3.14 Recoil corrections

In proton-proton collisions, W and Z bosons are predominantly produced through quark-antiquark annihilation. Higher-order processes can induce radiated quarks or gluons that recoil against the boson, imparting a non-zero transverse momentum to the boson [79]. Recoil corrections accounting for this effect are applied to samples with W+jets, Z+jets, and Higgs bosons [64]. The corrections are performed on the vectorial difference between the measured missing transverse momentum and the total transverse momentum of neutrinos originating from the decay of the W, Z, or Higgs boson. This vector is projected onto the axes parallel and orthogonal to the boson  $p_T$ . This vector, and the resulting correction to use, is measured in  $Z \rightarrow \mu\mu$  events,



(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*) [76].

1547 since these events have leptonic recoil that do not contain neutrinos, allowing the  
 1548 4-vector of the Z boson to be measured precisely. The corrections are binned in  
 1549 generator-level  $p_T$  of the parent boson and also the number of jets in the event.

### 1550 5.3.15 Drell-Yan corrections

1551 The Z boson transverse momentum distribution disagrees between leading-order (LO)  
 1552 simulations and data in a  $Z \rightarrow \mu\mu$  control region with at least one b-tag jet [80]. Per-  
 1553 event weights derived by the 2016 data-only version of this analysis [80] are applied to  
 1554  $Z \rightarrow \tau\tau/\ell\ell$  events, as a function of the generator-level Z boson  $p_T$  to provide better  
 1555 matching of MC to data.

### 1556 5.3.16 Pileup reweighting

1557 Reweighting is performed to rescale MC events to account for differences between  
 1558 MC and data, in the distribution of the pileup (number of additional proton-proton  
 1559 interactions per bunch crossing). A tool for calculating the pileup reweighting for the

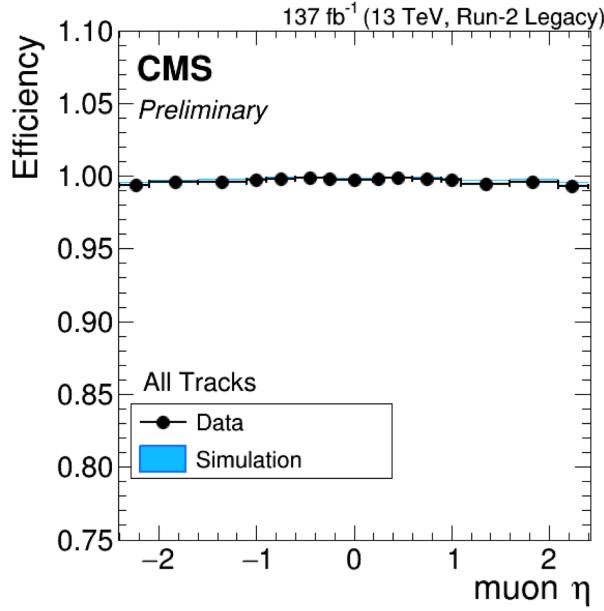


Figure 5.10: Muon tracking efficiencies as a function of  $|\eta|$  for standalone muons in Run-2 data (*black*) and Drell-Yan MC simulation (*blue*) [78]. All Tracks refers to tracks which exploit the presence of muon candidates in the muon system to seed the track reconstruction in the inner tracker, in contrast to tracks that use tracker-only hits for seeding. Uncertainties shown are statistical.

1560 MC samples used is provided centrally by the Luminosity POG [81].

### 1561 5.3.17 Pre-firing corrections

1562 In 2016 and 2017 data-taking, a gradual timing shift of ECAL was not properly  
1563 propagated to L1 trigger primitives (TPs), resulting in a large fraction of high  $\eta$   
1564 TPs being incorrectly associated with the previous bunch crossing. L1 trigger rules  
1565 prevent two consecutive bunch crossings from firing, causing events to be rejected if  
1566 significant ECAL energy was deposited in  $2.0 < |\eta| < 3.0$ . To account for this issue,  
1567 MC simulations for 2016 and 2017 are corrected using an event-dependent weight.  
1568 Embedded samples are not corrected [46].

1569 **5.3.18 Top  $p_T$  spectrum reweighting**

1570 In Run-1 and Run-2 it was observed that the  $p_T$  spectra of top quarks in  $t\bar{t}$  data  
1571 was significantly softer than those predicted by MC simulations [82]. Possible sources  
1572 of this discrepancy are higher order QCD and/or electroweak corrections, and non-  
1573 resonant production of  $t\bar{t}$ -like final states. To account for this, corrections derived  
1574 from Run-2 data by the Top Physics Analysis Group (PAG) are applied to the  $p_T$   
1575 of the top and anti-top quarks in MC simulations, computed as a function of their  
1576 generator-level  $p_T$  [82].

1577 **5.3.19 B-tagging efficiency**

1578 In order to predict correct b-tagging discriminant distributions and event yields in  
1579 data, the weight of selected MC events is reweighed according to recommendations by  
1580 the BTV POG [83]. The reweighting depends on the jet  $p_T$ ,  $\eta$ , and the b-tagging dis-  
1581 criminant. In this method, there is no migration of events from one b-tag multiplicity  
1582 bin to another.

1583 **5.3.20 Jet energy resolution and jet energy smearing**

1584 Calibration of jet energies, i.e. ensuring that the energy and momentum of the recon-  
1585 structed jet matches that of the quark/gluon-initiated jet, is a challenging task due  
1586 to time-dependent changes in the detector response and calibration and high pileup  
1587 [84] [85]. Jet calibration is done via jet energy corrections (JECs) applied to the  $p_T$   
1588 of jets in MC samples, accounting successively for the effects of pileup, uniformity of  
1589 the detector response, and residual data-simulation jet energy scale differences [86].  
1590 Typical jet energy resolutions reported at  $\sqrt{s} = 8$  TeV in the central rapidities are  
1591 15-20% at 30 GeV and about 10% at 100 GeV [84]. Jet energy corrections are also  
1592 propagated to the missing transverse energy.

1593 Measurements show that the jet energy resolution (JER) in data is worse than  
1594 in simulation, and so the jets in MC need to be smeared to describe the data. JER  
1595 corrections are applied after JEC on MC simulations, and adjust the width of the  $p_T$   
1596 distribution based on pileup, jet size, and jet flavor [87]. Tools for applying JEC and  
1597 JER are provided centrally by the JER Corrections group.

1598 **Chapter 6**

1599 **Event selection**

1600 **6.1 General procedure for all channels**

1601 For the search for  $h \rightarrow aa \rightarrow bb\tau\tau$ , three final states of the  $\tau\tau$  system are considered:  
1602  $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$ . The  $\tau_h\tau_h$  final state is not considered because signal events in the  
1603  $\tau_h\tau_h$  channel would typically produce hadronic taus with momenta below data-taking  
1604 trigger thresholds.

1605 In all three final states, events are required to have at least one b-tag jet passing the  
1606 medium working point of the DeepFlavour tagger, with  $p_T > 20$  GeV, and  $|\eta| < 2.4$ .  
1607 A second b-tag jet is not required because such a requirement would reduce signal  
1608 acceptance by 80% compared to only requiring one b-tag jet.

1609 Events in MC samples are sorted into one of the three  $\tau\tau$  channels if they pass the  
1610 following trigger requirements and requirements on the offline reconstructed objects  
1611 in the event, first checking the HLT paths for the  $\mu\tau_h$  channel, then  $e\tau_h$ , and finally  
1612  $e\mu$ . The two leading leptons (e.g. muon and hadronic tau for the  $\mu\tau_h$  channel) that  
1613 were determined to have originated from the  $\tau\tau$  decay, are called the  $\tau\tau$  “legs” and  
1614 are respectively subscripted 1 and 2 in this work. For events in data and embedded  
1615 samples, the HLT paths requirements for the corresponding channel are checked.

1616 After sorting events by HLT paths and identifying the leading tau legs in the offline  
 1617 reconstructed objects, the  $p_T$  of the offline objects is checked against the online trigger  
 1618 thresholds. Trigger matching is also performed, which checks the correspondence  
 1619 between each offline reconstructed object used in the analysis (e.g. a muon), and a  
 1620 trigger object in the HLT (e.g. a HLT muon). An offline object is considered to be  
 1621 matched, if it corresponds to a trigger object of the same object type, with  $\Delta R < 0.5$ .  
 1622 This matched trigger object is also required to pass the filter(s) of the HLT trigger.  
 1623 The trigger thresholds used for the  $bb\tau\tau$  final state (the focus of this work) and the  
 1624  $bb\mu\mu$  final state are summarized in Table 6.1 and detailed in the following sections.

| Year | Single/dilepton trigger $p_T$ | $bb\mu\mu$ |        | $bb\tau\tau$ |             |        |       |
|------|-------------------------------|------------|--------|--------------|-------------|--------|-------|
|      |                               | $\mu$      | $e\mu$ | $e\tau_h$    | $\mu\tau_h$ | $e$    | $\mu$ |
| 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     | —     |
|      | $p_T$ -subleading lepton      | 8          | 12     | 8            | 24          | —      | 20    |
| 2018 | Single lepton                 | 24         | —      | 32, 35       | —           | 24, 27 | —     |
|      | $p_T$ -leading lepton         | 17         | 23     | 23           | —           | 30     | —     |
|      | $p_T$ 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.

1625 Further cuts are made on the offline objects in each channel to obtain the signal  
 1626 region, or other data regions used to perform data-driven background estimations.

## 1627 6.2 Event selection in the $\mu\tau_h$ channel

1628 In all three years, a single muon trigger is used if the muon has sufficiently high  $p_T$ ,  
 1629 otherwise a dilepton  $\mu\tau_h$  cross-trigger is used (Tables 6.2, 6.3, and 6.4). For data

1630 taken in 2017-2018 (2016), the logical OR of the single muon triggers with online  $p_T$   
1631 thresholds 24 and 27 (23) GeV is used, with the corresponding offline muon required  
1632 to have with  $p_T$  1 GeV above the online threshold. For data taken in 2017-2018  
1633 (2016), a dilepton  $\mu + \tau_h$  cross-trigger with  $p_T$  thresholds of 20 (19) and 27 (20) GeV  
1634 for the muon and tau respectively, is used. The  $\tau_h$  is required to have  $|\eta| < 2.3$  if the  
1635 single trigger is fired,  $|\eta| < 2.1$ .

1636 The muon and  $\tau_h$  are required to have opposite charge and be separated by  $\Delta R >$   
1637 0.4. The muon is required to have  $|\eta| < 2.4$ , and the  $\tau_h$  is required to have  $|\eta| < 2.3$   
1638 unless a cross-trigger is required, in which case we require  $|\eta| < 2.1$  as discussed  
1639 above.

1640 The muon is required to pass the medium identification (ID) working point [88],  
1641 which is defined by the Muon POG as a loose muon (i.e. a Particle Flow muon that is  
1642 either a global or a tracker muon - see Section 5.1.2) with additional requirements on  
1643 track quality and muon quality. This identification criteria is designed to be highly  
1644 efficiently for prompt muons and for muons from heavy quark decays. In addition to  
1645 the ID, for prompt muons it is recommended to apply cuts on the impact parameter  
1646 [88]: we apply  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ .

1647 In addition, a cut is applied on the muon relative isolation (defined in Section  
1648 5.1.2), to be less than 0.15 in a cone size of  $\Delta R = 0.4$ , which corresponds to the  
1649 Tight Particle Flow isolation requirement [88].

1650 The  $\tau_h$  is required to pass a cut on its impact parameter of  $|\Delta(z)| < 0.2$ . The  $\tau_h$   
1651 is also required to pass the VLoose (Very Loose) DeepTau working point vs. elec-  
1652 tron, the Tight DeepTau working point vs. muons, and the VVVLoose and Medium  
1653 DeepTau working point vs. jets. Events with taus reconstructed in two of the decay  
1654 modes (labeled 5 and 6) are rejected, since these decay modes are meant to recover  
1655 3-prong taus, but are only recommended for use in analyses where the benefits in  
1656 final significance outweigh the resulting increase in background [61].

1657        For the estimation of the background from jets faking  $\tau_h$ , which is described in Sec-  
1658        tion 7.7, anti-isolated events are selected, by requiring events to pass all the selections  
1659        described above, except failing the Medium DeepTau working point vs. jets.

1660        **6.3 Event selection in the  $e\tau_h$  channel**

1661        The HLT trigger paths for the  $e\tau_h$  channel are summarized in Tables 6.2, 6.3, and  
1662        6.4. Similarly to the  $\mu\tau_h$  channel, a single electron trigger is used if the electron has  
1663        sufficiently high  $p_T$  in 2018 and 2017. For data taken in 2018 (2017), the OR of the  
1664        single electron triggers with online  $p_T$  thresholds at 32 and 35 (27 and 32) GeV are  
1665        used, with the corresponding offline electrons required to have  $p_T$  greater than 33  
1666        (28) GeV. A  $e + \tau_h$  cross-trigger is used for electrons with lower offline  $p_T$  between  
1667        25 and 33 GeV (25 and 28 GeV). For the 2016 dataset, there is no cross trigger but  
1668        only a single electron trigger with online  $p_T$  threshold at 25 GeV, which is used if the  
1669        offline electron has  $p_T$  greater than 26 GeV.

1670        The electron and  $\tau_h$  are required to have opposite charge and be separated by  
1671         $\Delta R > 0.4$ . The electron is required to be within  $|\eta| < 2.3$  when no cross trigger is  
1672        used, and  $|\eta| < 2.1$  when the cross trigger is fired. The  $\tau_h$  is required to have  $|\eta| < 2.3$   
1673        if no cross trigger is fired, and have  $|\eta| < 2.1$  if the cross trigger is fired.

1674        The electron is required to have a relative isolation (same definition as in Section  
1675        5.1.2) of less than 0.1 in a cone size of  $\Delta R = 0.3$ , which is the standard recommended  
1676        cone size giving minimal pileup dependence and reduced probability of other objects  
1677        overlapping with the cone. The isolation quantity used includes an “effective area”  
1678        (EA) correction to remove the effect of pileup in the barrel and endcap parts of the  
1679        detector [89].

1680        The electron is also required to pass cuts on its impact parameter of  $|\Delta(z)| < 0.2$   
1681        and  $|\Delta(xy)| < 0.045$ . It is also required to pass the non-isolated MVA working point

1682 corresponding to 90% efficiency. The electron's number of missing hits, which are  
1683 gaps in its trajectory through the inner tracker [89], must be less than or equal to  
1684 1. The electron must pass a conversion veto, which rejects electrons coming from  
1685 photon conversions in the tracker, which should instead be reconstructed as part of  
1686 the photon [89].

1687 The impact parameter cut for the  $\tau_h$  is  $|\Delta(z)| < 0.2$ . In contrast to the  $\mu\tau_h$  event  
1688 selection, the vs. electron and vs. muon DeepTau working points are flipped, to  
1689 reject muons faking the  $\tau_h$  leg. The  $\tau_h$  is required to pass the Tight DeepTau working  
1690 point vs. electrons, the VLoose DeepTau working point vs. muons, and the Medium  
1691 DeepTau working point vs. jets.

1692 As in the  $\mu\tau_h$  channel, for the estimation of the background from jets faking  $\tau_h$ ,  
1693 which is described in Section 7.7, anti-isolated events are selected, by requiring events  
1694 to pass all the selections described above, except failing the Medium DeepTau working  
1695 point vs. jets.

## 1696 6.4 Event selection in the $e\mu$ channel

1697 The HLT trigger paths for the  $e\mu$  channel are summarized in Tables 6.2, 6.3, and  
1698 6.4. Events are selected with the logical OR of two  $e + \mu$  cross triggers, where either  
1699 the electron or muon can have larger  $p_T$ : (1) leading electron, where the electron has  
1700 online  $p_T > 23$  GeV and muon has online  $p_T > 8$  GeV, or (2) leading muon, where  
1701 electron has online  $p_T > 12$  GeV and muon has online  $p_T > 23$  GeV.

1702 The leading and sub-leading leptons are required to have an offline  $p_T$  greater  
1703 than 1 GeV above the online threshold (i.e.  $p_T > 24$  GeV). If the sub-leading lepton  
1704 is the electron, the offline  $p_T$  threshold is 1 GeV above the online threshold ( $p_T > 13$   
1705 GeV), but if it is a muon, the offline  $p_T$  threshold is required to be at least 5 GeV  
1706 greater than the online threshold (i.e.  $p_T > 13$  GeV). This is because of poor data

1707 and simulation agreement for low- $p_T$  muons with  $p_T$  between 9 GeV and 13 GeV, and  
1708 the higher probability of mis-identifying jets as muons at lower  $p_T$ . With no effect on  
1709 the expected limits, the offline  $p_T$  threshold for muons is raised to 13 GeV instead of  
1710 9 GeV, even though it may lead to loss in signal acceptance. Both the electron and  
1711 muon are required to have  $|\eta| < 2.4$ .

1712 The electron and muon are required to have opposite charge and be separated  
1713 by  $\Delta R > 0.3$  (note the decreased separation requirement compared to the other  
1714 two channels). The electron is required to pass the non-isolated MVA identification  
1715 working point corresponding to 90% efficiency, and to have a relative isolation less  
1716 than 0.1 for a cone size of  $\Delta R = 0.3$  with the EA pileup subtraction correction.  
1717 The electron must have one or fewer missing hits and pass the conversion veto (both  
1718 described previously in Section 6.3).

1719 The muon is required to pass the medium identification working point (described  
1720 earlier in 6.2), and to have a relative isolation less than 0.15 for a cone size of  $\Delta R =$   
1721 0.4. The muon impact parameter is required to have  $|\Delta(z)| > 0.2$  and  $|\Delta(xy)| < 0.045$ .

1722 For the QCD multijet background estimation described in Section 7.8, the same-  
1723 sign region is selected by requiring all the above selections, except the legs are required  
1724 to have the same electric charge rather than opposite.

## 1725 6.5 Extra lepton vetoes in all channels

1726 Events containing a third lepton (electron or muon) that is neither of the leading  $\tau\tau$   
1727 legs are rejected, and events with di-muons and di-electrons are vetoed, with criteria  
1728 taken from the Standard Model  $H \rightarrow \tau\tau$  working group [64].

1729 The event is vetoed if a third electron is found with the following properties:  
1730  $p_T > 10$  GeV,  $|\eta| < 2.5$ , impact parameter  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ , passing  
1731 non-isolation MVA identification with 90% efficiency, conversion veto,  $\leq 1$  missing

| 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 6.2: High-Level Trigger (HLT) paths used to select data and simulation events in 2016 for the three  $\tau\tau$  channels.

| 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 6.3: High-Level Trigger (HLT) paths used to select data and simulation events in 2017 for the three  $\tau\tau$  channels.

| 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 6.4: 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.

1732 hits, and relative isolation  $< 0.3$  with cone size  $\Delta R = 0.3$ . The event is also vetoed if  
1733 a third muon is found with the following properties:  $p_T > 10$  GeV,  $|\eta| < 2.4$ , impact  
1734 parameter  $|\Delta(z)| < 0.2$  and  $|\Delta(xy)| < 0.045$ , medium ID, and isolation  $< 0.3$  with  
1735 cone size  $\Delta R = 0.4$ .

1736 A di-muon veto is applied, which rejects events containing a pair of muons with  
1737 opposite charge and separation of  $\Delta R > 0.15$ , that both pass the following selections:  
1738  $p_T > 15$  GeV,  $|\eta| < 2.4$ , flag for global muons, flag for tracker muon, flag for Particle  
1739 Flow muon,  $|\Delta(z)| < 0.2$ ,  $|\Delta(xy)| < 0.045$ , and isolation  $< 0.3$  with cone size  $\Delta R =$   
1740  $0.4$ .

1741 A similar di-electron veto is applied to reject events containing a pair of electrons  
1742 with opposite charge and separation of  $\Delta R > 0.15$ , that both pass the following  
1743 selections:  $p_T > 15$  GeV,  $|\eta| < 2.5$ , a dedicated electron ID (cut-based) for vetoing  
1744 third leptons,  $|\Delta(z)| < 0.2$ ,  $|\Delta(xy)| < 0.045$ , with pileup-corrected relative isolation  
1745  $< 0.3$  with cone size  $\Delta R = 0.3$ .

1746 These vetoes on extra leptons also ensure orthogonality of events to analyses such  
1747 as the  $bb\mu\mu$  final state, whose results are combined with this  $bb\tau\tau$  final state as  
1748 described in Section ??.

1749

# Chapter 7

1750

## Background estimation

1751 This section describes methods used to estimate sources of background from Standard  
1752 Model processes in the search for  $h \rightarrow aa \rightarrow bb\tau\tau$ . Similar background estimation  
1753 methods are being used for the  $h \rightarrow a_1a_2$  analysis. The background contributions  
1754 directly taken from MC are described first, followed by backgrounds estimated from  
1755 data-driven methods to produce sufficient statistics in the signal region.

1756

### 7.1 Z+jets

1757 A major source of background for  $\tau\tau$  analyses is the Drell-Yan (DY) process (Z+jets).  
1758 The Z boson decays to  $\tau\tau/\mu\mu/ee$  with equal probability of 3.4% each, with the dom-  
1759 inant decay modes being to hadrons (around 70%) and neutrinos (invisible) (20%)  
1760 [24].

1761 The Drell-Yan contribution with genuine taus,  $Z \rightarrow \tau\tau$ , is estimated using embed-  
1762 ded samples, described in Section 4.3. To avoid double-counting between embedded  
1763 and MC samples, in all MC samples, events with legs that originated from genuine  $\tau$   
1764 are discarded.

1765 The other decays of the Z,  $Z \rightarrow ee$  and  $Z \rightarrow \mu\mu$ , are estimated from MC simulation,  
1766 and are hereafter referred to as simply the Drell-Yan background. These MC samples

1767 are generated to leading order (LO) with different numbers of jets (jet multiplicity) in  
1768 the matrix element: Z+1 jet, Z+2jets, Z+3 jets, Z+4 jets, and inclusive Z+jets. The  
1769 cross-sections of the samples with  $\geq 1$  jets are normalized to next-to-NLO (NNLO)  
1770 in QCD.

1771 For the inclusive Drell-Yan sample, two samples are used with different thresholds  
1772 for the di-lepton invariant mass ( $m_{\ell\ell}$ ) at the generator level: one with  $m_{\ell\ell} > 50$  GeV  
1773 and the other with  $10 < m_{\ell\ell} < 50$ .

## 1774 7.2 W+jets

1775 The dominant W boson decay modes are to hadrons (67.4%),  $e + \nu_e$  (10.7%),  $\mu + \nu_\mu$   
1776 (10.6%), and  $\tau + \nu_\tau$  (11.4%) [24]. The W+jets background is estimated from MC  
1777 simulation. Similarly to the Z+jets, the W+jets samples are generated with different  
1778 jet multiplicities in the matrix element. LO samples are used for greater statistics  
1779 and are normalized to NNLO cross sections.

## 1780 7.3 $t\bar{t}$ + jets

1781 In hadron collisions, top quarks are produced singly with the weak interaction, or in  
1782 pairs via the strong interaction, with interference between these leading-order pro-  
1783 cesses possible in higher orders of the perturbation theory. The top quark is the  
1784 heaviest fermion in the Standard Model and has a short lifetime ( $\sim 10^{-25}$  s), decay-  
1785 ing without hadronization into a bottom quark and a W boson [24], with the decay  
1786 modes of the W boson as listed in the previous section. With two top quarks, the  
1787 final states of the two resulting W bosons can be described as fully leptonic, semilep-  
1788 tonic, and fully hadronic. These three final states are modeled separately with MC  
1789 simulation in 2018 and 2017, while for 2016 the sample used is inclusive.

## 1790 7.4 Single top

1791 There are three main production modes of the single top in  $pp$  collisions [90]: the  
1792 exchange of a virtual W boson ( $t$  channel), the production and decay of a virtual W  
1793 boson ( $s$  channel), and the associated production of a top quark and W boson ( $tW$ ,  
1794 or W-associated) channel. As the  $s$  channel process is rare and only 3% of the total  
1795 production, the dominant production mode of the  $t$ -channel and the  $tW$  production  
1796 are considered and modeled with MC.

## 1797 7.5 Diboson

1798 In  $pp$  collisions, the production of dibosons (pairs of electroweak gauge bosons, i.e.  
1799 WW, WZ, and ZZ) is dominated by quark-antiquark annihilation, with a small con-  
1800 tribution from gluon-gluon interaction [91]. MC is used to model the pair production  
1801 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  
1802 quarks and  $\ell$  being leptons).

## 1803 7.6 Standard Model Higgs

1804 MC is used to simulate backgrounds from major production modes of the Standard  
1805 Model 125 GeV Higgs boson: gluon-gluon fusion (ggH), vector boson fusion (VBF),  
1806 associated production with a W or Z (WH, ZH), and associated production with a  
1807 top pair (ttH) (see Fig. 7.1 for leading-order diagrams). For these production modes,  
1808 samples with the Higgs decaying to  $\tau\tau$  or to  $WW$  are used. Samples made with  
1809 higher-order diagrams for WH and ZH that include the production of a jet, with the  
1810 Higgs decaying to WW, are also used.

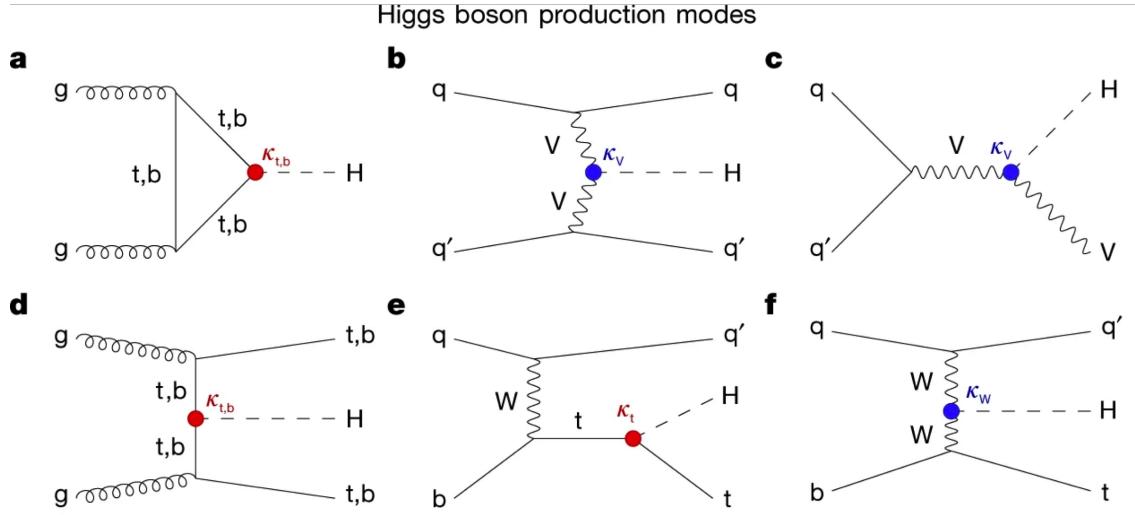


Figure 7.1: Leading-order Feynman diagrams of Higgs production from [92], in ggH (a) and vector boson fusion (VBF; b), associated production with a W or Z (V) boson (VH; c), associated production with a top or bottom quark pair (ttH or bbH); d, and associated production with a single top quark (tH; e, f).

## 1811 7.7 Jet faking $\tau_h$

1812 Events with a jet mis-reconstructed as the hadronic tau leg  $\tau_h$  are a major source of  
 1813 background in the  $\mu\tau_h$  and  $e\tau_h$  channels. The main processes contributing to jet  $\rightarrow \tau_h$   
 1814 events are QCD multijet, W+jets, and  $t\bar{t}$  production. These events are estimated  
 1815 using a data-driven method adapted from past analyses [46] [80]. This background  
 1816 includes contributions from W+jets, QCD multijets, and  $t\bar{t}$ +jets. To estimate this  
 1817 background, a sideband region is constructed, where events are required to pass all  
 1818 baseline  $\mu\tau_h/e\tau_h$  selection criteria, but fail the  $\tau_h$  isolation criteria. The events in  
 1819 this sideband region is reweighed with a factor  $f/(1 - f)$ , where  $f$  is the probability  
 1820 for a jet to be misidentified as a  $\tau_h$ . The jet  $\rightarrow \tau_h$  background is the anti-isolated,  
 1821 reweighed MC and embedded events subtracted from the anti-isolated, reweighted  
 1822 data events.

1823 The fake factor is measured in  $Z \rightarrow \mu\mu + \text{jets}$  events in data in the  $\mu\mu\tau_h$  final  
 1824 state, as any reconstructed  $\tau_h$  in these events must originate from a jet. The two  
 1825 muons are required to be isolated ( $< 0.15$ ), have opposite electric charge, and have

1826 an invariant mass between 76 and 106 GeV (close to the Z mass). These events are  
1827 selected with a double muon trigger, with the leading muon having offline  $p_T > 20$   
1828 GeV and the subleading muon  $p_T > 10$  GeV. Simulated diboson (ZZ and WZ) events  
1829 are subtracted to avoid contamination from events with real  $\tau_h$ . The denominator of  
1830 the fake rate corresponds to fake taus passing the VVVLoose working point of the  
1831 discriminator vs. jets, while the numerator corresponds to those passing the Medium  
1832 working point, i.e.  $f = N_{\text{jet passing tight}} / N_{\text{jet passing loose}}$ .

1833  $f$  is measured as a function of the  $\tau_h$  transverse momentum and is 8% - 10% in  
1834 each of the data-taking years.  $f$  is derived separately for the  $\mu\tau_h$  and  $e\tau_h$  channels  
1835 because the channels use different anti-lepton identification working points.

## 1836 7.8 QCD multijet background

1837 In the  $e\mu$  channel, events with jets faking electrons or muons originating from QCD  
1838 multijet, is estimated from data events with the same baseline selection as in the  
1839 signal region, except with same-signed (SS) charged  $e + \mu$ , ensuring orthogonality  
1840 with the signal region which requires opposite-sign (OS)  $e\mu$  pairs. All same-sign MC  
1841 events (both events with real and fake  $e + \mu$ ) are subtracted from same-sign data  
1842 events to remove contamination from other backgrounds. i.e.  $\text{QCD}_{\text{SS}} = \text{Data}_{\text{SS}} -$   
1843  $\text{MC}_{\text{SS}}$ .

1844 Three scale factors are applied to the  $\text{QCD}_{\text{SS}}$  events to compute the QCD multijet  
1845 background [80] [40]:

- 1846 • *OS-to-SS scale factor*: This scales the SS QCD to the OS region, and is mea-  
1847 sured from an orthogonal region with an isolated electron and an anti-isolated  
1848 muon. Only the muon is chosen to be anti-isolated because this scale factor was  
1849 observed to depend more strongly on electron isolation than that of the muon.  
1850 This scale factor is treated as a function of the  $\Delta R$  separation of the trajectories

1851 of the electron and muon, and is measured separately for events with 0 jets, 1,  
1852 jet, and greater than 1 jet.

- 1853 • *2D closure correction for the lepton  $p_T$ :* This factor accounts for subleading  
1854 dependencies of the first scale factor on the  $p_T$  of the two leptons. A 2D weight  
1855 is derived in a similar fashion, as a ratio of  $\text{QCD}_{OS}$  events to  $\text{QCD}_{SS}$  events,  
1856 but parameterized by both electron and muon  $p_T$ , where the SS events have the  
1857 previous scale factor applied.
- 1858 • *Isolation correction for the muon:* The third and final factor is an isolation  
1859 correction, which is a bias correction to account for the fact that the fake  
1860 factor was determined for less-isolated muons. This factor is obtained as the  
1861 ratio of the OS-to-SS scale factors measured in two other control regions: (1)  
1862 events where the electron is anti-isolated ( $0.15 < \text{iso} < 0.5$ ) and the muon is  
1863 isolated, and (2) events where both leptons are anti-isolated.

1864

# Chapter 8

1865

## Systematic uncertainties

1866 The handling of systematic uncertainties is separated into normalization uncertainties  
1867 (those that affect the total yield of a variables' distribution) and shape uncertainties  
1868 (those that shift the distribution of events). Normalization uncertainties are expressed  
1869 as multiplicative factors, while shape uncertainties are represented as up and down  
1870 shifts of a variable's distribution.

1871 Up/down shifts of shape uncertainties can change the number of background  
1872 events in a distribution. For instance, hadronic taus receive corrections from the  
1873 nominal tau energy scale, with the nominal, up, and down energy scales provided  
1874 centrally by CMS. For the  $\mu\tau_h$  channel, an event could have a  $\tau_h$  with  $p_T$  just below  
1875 the offline threshold of 20 GeV (for instance, 19.5 GeV), so in the nominal distribution  
1876 of  $m_{\tau\tau}$  (or any other variable for this channel), the event is excluded. However, when  
1877 we build our distributions with the tau energy scale “up” shift, the energy of this  $\tau_h$   
1878 may be scaled up to, say, 20.5 GeV, and now the event passes the offline  $p_T$  threshold  
1879 for the single muon trigger, leading to the event's inclusion in the distributions made  
1880 with the tau energy scale “up” shift.

1881 In evaluating the up and down shifts of a specific source of uncertainty, all other  
1882 corrections and scale factors are held at their nominal values, and the full chain

1883 of object and event selection and event categorization is performed to obtain the  
1884 observable distributions. Any “downstream” variables that depend on the shifted  
1885 variable, e.g. the invariant di-tau mass  $m_{\tau\tau}$ , must be computed for the nominal case,  
1886 and then re-computed separately for each up and down shift of the tau legs’ energy  
1887 scale. The objective of this process is to quantify the effect of a single source of  
1888 uncertainty on the resulting observable distributions.

## 1889 **8.1 Uncertainties associated with physics objects**

1890 Each scale factor and correction described in Section 5.3 has an associated uncertainty.  
1891 The binning of the uncertainties follows that of the nominal scale factor value.

### 1892 **8.1.1 Uncertainties in the lepton energy scales**

1893 The uncertainties in the tau energy scales [61] are binned by the tau decay mode and  
1894 are taken as shape uncertainties treated as uncorrelated across the tau decay modes  
1895 and years. Same as with the application of the nominal scale factor, when applying  
1896 the up or down shifts, the missing transverse energy ( $p_T^{\text{miss}}$ ) of the event is adjusted  
1897 so that the 4-vector sum of the tau  $p_T^{\text{miss}}$  is unchanged.

1898 The uncertainties in the muon energy scale [62] are 0.4% for  $|\eta| < 1.2$ , 0.9% for  
1899  $1.2 < |\eta| < 2.1$ , and 2.7% for  $2.1 < |\eta| < 2.4$ , and are treated as shape uncertainties,  
1900 fully uncorrelated between embedded and MC samples.

1901 The uncertainties in the electron energy scale [65] in MC are binned in the electron  
1902  $|\eta|$  and  $p_T$ , and are shown in Fig. 5.2. The uncertainties range from 0.5% to 2.2% in  
1903 the barrel, and 0.3% to 4.1% in the endcap, across the  $p_T$  range. The uncertainties  
1904 for the embedded sample are binned only in  $|\eta|$  and are on the order of 0.5% and  
1905 1.25% for the barrel and endcap [69].

1906 There are also uncertainties in the energy scales for electrons and muons misiden-

tified as  $\tau_h$ . The uncertainty for muons misidentified as  $\tau_h$  is 1% [61]. For electrons misidentified as  $\tau_h$ , the uncertainty is binned in barrel/endcap  $\eta$  and by 1-prong and 1-prong +  $\pi_0$  decays. The probability for  $e/\mu$  faking a 3-prong decay mode is much lower.

### 8.1.2 Uncertainties from other lepton corrections

Uncertainties associated with the  $\tau_h$  identification efficiencies are treated as shapes, uncorrelated across the seven  $p_T$  bins and years. The shape uncertainties in the embedded samples are taken as 50% correlated with those of the MC samples.

The uncertainties on electron and muon identification efficiencies are taken as normalization uncertainties of 2% each, with a 50% correlation between embedded and MC samples.

In the  $e\tau_h$  channel, there is an additional uncertainty for the vs. jet discrimination efficiency [61], because the analysis uses a looser anti-lepton working point (VLoose WP) than the working points used in the measurement of the efficiency (namely, VLoose WP vs e, and Tight WP vs mu). For nominal  $\tau_h p_T < 100$  GeV, an additional uncertainty of 3% (5%) is used in MC (embedded), and for high  $p_T$  an uncertainty of 15% is used for both.

The uncertainties in trigger efficiencies are taken as shapes [61]. In the  $e\tau_h$  and  $\mu\tau_h$  channels, there are uncertainties for the single and cross lepton triggers, and in the  $e\mu$  channel there is one uncertainty each for the two  $e + \mu$  triggers, and one combined uncertainty since their trigger phase spaces are not mutually exclusive.

### 8.1.3 Uncertainties from jet energy scale and resolution

The jet energy scale uncertainties are taken as shape uncertainties: there are eleven in total, with seven correlated across years (labeled “Year” below) and the remainder uncorrelated across years. They affect the b-tag jet  $p_T$  and mass, and hence the

<sub>1932</sub> missing transverse energy  $p_T^{\text{miss}}$ . The shifts are propagated through the b-tagging  
<sub>1933</sub> scale factor calculation and b-tag jet counting.

<sub>1934</sub> The uncertainties in the jet energy correction and resolution [84] [93] are as follows:

<sub>1935</sub> • *Absolute, AbsoluteYear*: flat absolute scale uncertainties.

<sub>1936</sub> • *BBEC1, BBEC1Year*: for sub-detector regions, with barrel “BB” in  $|\eta| < 1.3$   
<sub>1937</sub> and endcap region 1 “EC1”:  $1.3 < |\eta| < 2.5$ .

<sub>1938</sub> • *EC2, EC2 year*: for sub-detector regions, with endcap region 2 “EC2” in  $2.5 <$   
<sub>1939</sub>  $|\eta| < 3.0$ .

<sub>1940</sub> • *HF, HF year*: for sub-detector regions, with hadron forward “HF” in  $|\eta| > 3$ .

<sub>1941</sub> • *FlavorQCD*: for uncertainty in jet flavor (uds/c/b-quark and gluon) estimates  
<sub>1942</sub> based on comparing Pythia and Herwig (different MC generator) predictions.

<sub>1943</sub> • *RelativeBal*: account for difference between log-linear fits of the two methods  
<sub>1944</sub> used to study the jet energy response: MPF (missing transverse momentum  
<sub>1945</sub> projection fraction) and  $p_T$  balance.

<sub>1946</sub> • *RelativeSample*: account for  $\eta$ -dependent uncertainty due to a difference be-  
<sub>1947</sub> tween relative residuals, observed with dijet and Z+jets in Run D of 2018 data.

<sub>1948</sub> • *JetResolution*: uncertainty in the jet energy resolution.

#### <sub>1949</sub> 8.1.4 Uncertainties from b-tagging scale factors

<sub>1950</sub> The b-tagging scale factor has its own set of associated uncertainties (not to be  
<sub>1951</sub> confused with shifts in the b-tagging scale factor due to the propagation of the jet  
<sub>1952</sub> energy scale uncertainties described in the previous section 8.1.3). They are:

<sub>1953</sub> • *hf*: contamination from heavy flavor (b+c) jets in the light flavor region.

- $hfstats1$ ,  $hfstats2$ : linear and quadratic statistical fluctuations from b-flavor jets.
  - $lf$ : contamination from light flavor (udsg+c jets) in the heavy flavor region.
  - $lfstats1$ ,  $lfstats2$ : linear and quadratic statistical fluctuations from udsg jets.
  - $cferr$ ,  $cferr2$ : uncertainty for charm jets.
- The variations for “lf, hf, hfstats1/2, lfstats1/2” are applied to both b and udsg jets.  
For c-flavor jets, only “cferr1/2” is applied.

### 8.1.5 Uncertainties from MET

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

## 8.2 Uncertainties associated with samples used

- Normalization uncertainties related to the samples used are:
- *Cross-section uncertainties*:  $\sigma(t\bar{t})$ : 4.2%,  $\sigma(\text{diboson})$ : 5%,  $\sigma(\text{single top})$ : 5%,  $\sigma(\text{ggH})$ : 3.2%,  $\sigma(\text{qqH})$ : 2.1%,  $\sigma(\text{WH})$ : 1.9%,  $\sigma(\text{ZH})$ : 1.3%,  $\sigma(\text{ttH})$ : 3.6%
  - *Uncertainties in QCD renormalization scale*: QCD scale(qqH): +0.43%-0.33%, QCD scale(WH): +0.5%-0.7%, QCD scale(ttH): +5.8%-9.2%
  - *Branching ratio uncertainties*:  $\text{BR}(\text{H} \rightarrow \tau\tau)$ : 1.8%, and  $\text{BR}(\text{H} \rightarrow \text{WW})$ : 1.5%
  - *Normalization uncertainties*: 2% for Drell-Yan, 4\$ for embedded, 20% pre-fit for the QCD multijet background in the  $e\mu$  channel, 20% pre-fit for the jet faking background.

1975        The  $t\bar{t}$  process has additional acceptance uncertainties from QCD scale variation  
1976        and parton shower uncertainties [94]. Parton shower uncertainties originate from  
1977        the modeling of perturbative and non-perturbative QCD effects handled in parton  
1978        shower MC generators. The scale variations are determined from the envelope of the  
1979        6 provided shapes due to variations in the factorization scale, renormalization scale,  
1980        and their combined variation [94].

1981        The Z  $p_T$  reweighing uncertainty in Drell-Yan samples is taken to be 10% of the  
1982        nominal value, taken as a shape uncertainty.

1983        The fake rate uncertainties are taken as shape uncertainties. For the weight ap-  
1984        plied to scale up anti-isolated events in cross-trigger regions, 20% of the nominal  
1985        weight is taken as a shape uncertainty.

### 1986        **8.3 Other uncertainties**

1987        A 3.6% yield uncertainty in the signal is used to cover uncertainties in the parton  
1988        distribution functions,  $\alpha_s$  (fine structure constant), and QCD scale.

1989        Normalization uncertainties from luminosity are applied to all MC samples, di-  
1990        vided into those uncorrelated across years, those correlated between 2017 and 2018,  
1991        and one for 2018 [81].

### 1992        **8.4 Pulls and impacts**

1993        The top impacts and pulls computed for the combination of all channels and years is  
1994        shown in Fig. 8.1. The top impacts are related to uncertainty in the signal sample and  
1995        cross-section of the  $t\bar{t}$  cross-section, and also the yields of the jet faking  $\tau_h$  background,  
1996        which is a major background in all channels and expected to be constrained due to  
1997        the yield uncertainty which is taken to be 20% pre-fit.

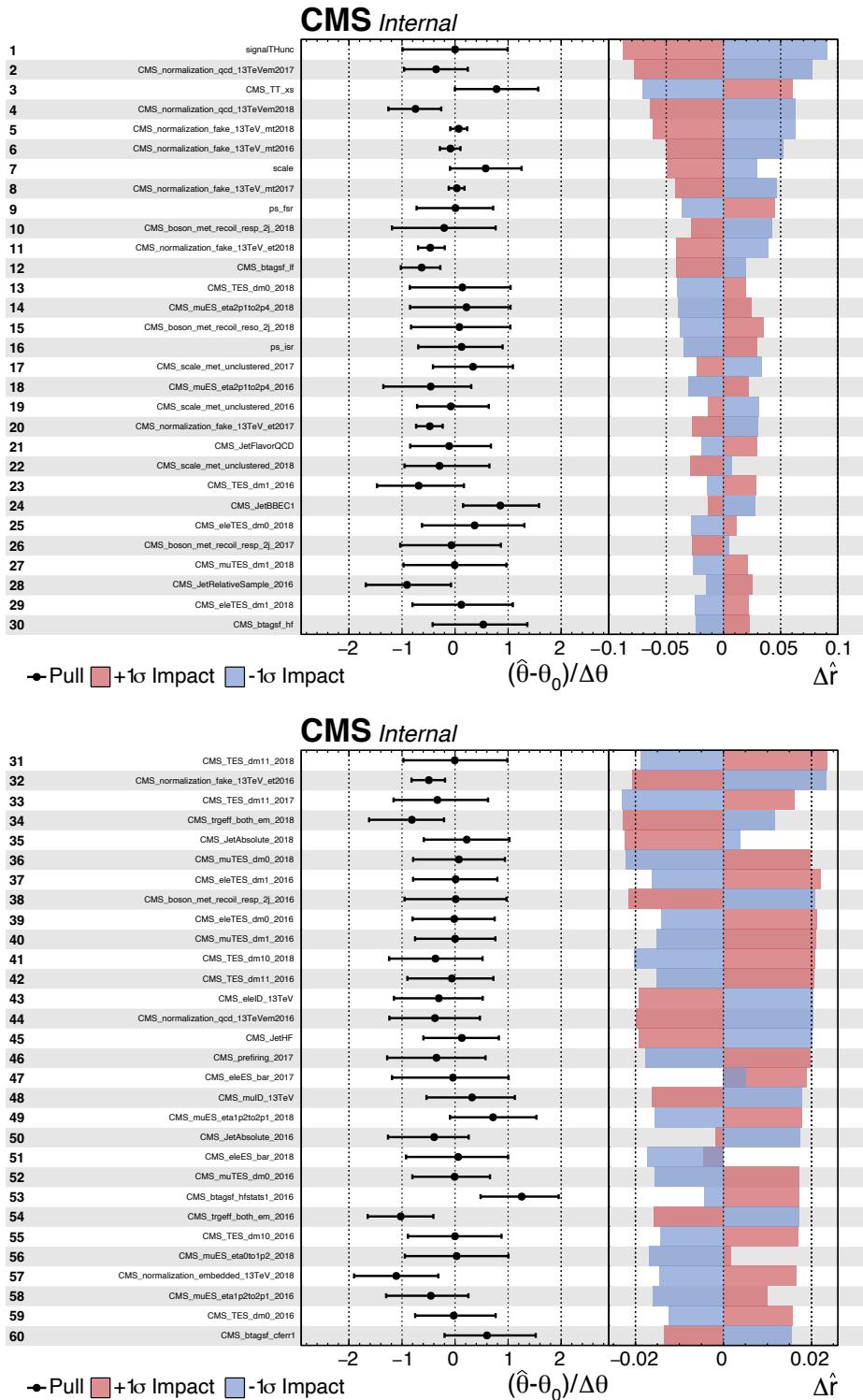


Figure 8.1: Top sixty impacts for the combination of all channels and years [39].

# 1998 Chapter 9

## 1999 Event categorization and signal 2000 extraction

### 2001 9.1 B-tag jet multiplicity

2002 The increased statistics of the full Run-2 dataset enables the separation of events into  
2003 events with exactly 1 b-tag jet and events with greater than 1 b-tag jet. Further event  
2004 categorization is performed with deep neural networks (DNNs) described below. The  
2005 DNNs are used only for separating events into signal and control regions in the 1  
2006 b-tag and 2 b-tag jets scenarios. The final results are extracted from the statistical  
2007 fitting to the mass of the  $\tau\tau$ ,  $m_{\tau\tau}$ .

### 2008 9.2 DNN-based event categorization

2009 A brief overview of the DNN-based event categorization is given below with a focus  
2010 on the physics aspects, with full details of the machine learning training in [40] and  
2011 associated documentation.

2012 **Training samples**

2013 Neural networks for event categorization are trained for each of the  $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$   
2014 channels, for 1 and 2 b-tag jets, giving  $3 \times 2 = 6$  networks in total. In the training,  
2015 the signal is taken to be all of the possible pseudoscalar mass  $m_a$  hypotheses together.  
2016 The backgrounds for each DNN are taken to be a representative combination of the  
2017 three major backgrounds:  $Z \rightarrow \tau\tau$ ,  $t\bar{t}$ +jets, and fake backgrounds. The proportions of  
2018 each background for each channel and b-tag jet multiplicity are taken from the yields  
2019 in the  $m_{\tau\tau}$  distribution. For instance, in the  $\mu\tau_h$  1 b-tag jet category, the composition  
2020 of the background for training is 17.4% from  $Z \rightarrow \tau\tau$ , 42.4% from  $t\bar{t}$ +jets, and 40.2%  
2021 fakes.

2022 **Input variables**

2023 The input variables capture the key differences between the signal and the back-  
2024 ground:

- 2025 • Transverse momentum  $p_T$  of the electron and muon in the  $e\tau_h$  and  $\mu\tau_h$  channels,  
2026 where the signal tends to have a softer  $p_T$  spectrum (lower energy) than the  
2027 background.
- 2028 •  $p_T$  of the b-tag jet(s). The signal sample b-tag jet(s) tend to have softer  $p_T$ .
- 2029 • Invariant masses of the various objects ( $\tau\tau$  legs and the b-tag jet(s)), which  
2030 tend to be smaller for the signal samples.
- 2031 • The angular separation  $\Delta R$  between pairs of the objects, where signal samples  
2032 peak at smaller  $\Delta R$  values.
- 2033 • The transverse mass between the missing transverse energy  $p_T^{\text{miss}}$  and each of

the four objects [80], defined as

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

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

- The variable  $D_\zeta$  [80], defined as

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

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

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

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

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

2051 **Categorization using the DNN score**

2052 After training, events in data, MC, and embedded are evaluated with the six DNNs  
 2053 and assigned a raw score between 0 and 1 (background-like or signal-like). In order  
 2054 to flatten the distribution of the score and define score thresholds for categorizing  
 2055 events, the raw output scores are transformed with the function  $\tilde{p}(n) = \text{arctanh}(p \times$   
 2056  $\tanh(n))/n$  where  $n$  is a positive integer. The thresholds of the DNN score used for  
 2057 signal/control region definition are determined using scans that optimize the signal  
 2058 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.

2059 **9.3 Methodology for signal extraction**

2060 In this section we outline the statistics terminology and concepts underlying the  
 2061 modified frequentist method  $CL_S$  used to perform signal extraction.

|                  | 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.

### 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 [95]. 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

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

read “ $f$  of  $x$  given  $\alpha$ ”, is referred to as a probability model or model. The parameters  $\alpha$  typically represent parameters of the theory or an unknown property of the detector’s response. The parameters are not frequentist in nature, unlike  $x$ . Out of all the

parameters, typically only a few are of interest, and are called the parameters of interest (POI), labeled  $\mu$  here. The remaining are referred to as nuisance parameters (NP) [95] and are labeled  $\boldsymbol{\theta}$ .

$f(x)$  is the probability density for the observable in one event and we wish to describe the probability density for a dataset with many events,  $\mathcal{D} = \{x_1, \dots, x_n\}$ , called the total probability model  $\mathbf{f}$ . For instance, if we also have a prediction for the total number of events expected, called  $\nu$ , we also account for the overall Poisson 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 [95]. The likelihood function is not a probability density for  $\alpha$  and is not normalized to unity:

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

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

To estimate the parameter  $\alpha$  we use an estimator, which is a function of the data. Take for example the measurement of data distributed according to a Gaussian probability density  $f(x|\mu, \sigma) = \text{Gauss}(x|\mu, \sigma)$ . One possible estimator of the mean  $\mu$ , is the mean of the measured data points  $\bar{x} = \sum_{i=1}^n x_i/n$  [95].

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

### 2085 9.3.2 Hypothesis testing

2086 In this section we next introduce concepts related to hypothesis testing such as the  
2087 test statistic constructed from the ratio of likelihood functions.

2088 The objective of a likelihood analysis is to distinguish different models repre-  
2089 senting the various hypotheses, and determine the one that best explains the ex-  
2090 perimental outcome. In a search for new physics, a signal is additive on top of the  
2091 background. The background-only hypothesis is the null hypothesis, and the signal-  
2092 plus-background hypothesis is the alternative.

2093 As a simple example, take the  $p$ -value test, for an experiment where we count  
2094 events in the signal region,  $n_{SR}$ , and expect  $\nu_B$  background events and  $\nu_S$  events from  
2095 the signal [95]. Then

- 2096 1. The null hypothesis ( $H_0$ ), i.e. the background-only hypothesis in this experi-  
2097       ment, with the probability modeled by  $\text{Poisson}(n_{SR}|\nu_B)$ .
- 2098 2. The alternate hypothesis ( $H_1$ ), i.e. signal-plus-background hypothesis, with the  
2099       probability modeled by  $\text{Poisson}(n_{SR}|(\nu_B + \nu_S))$ .

2100 The compatibility of the observed data  $\nu_{SR}^0$  and the null hypothesis, is quantified as  
2101 the probability that the background-only hypothesis would produce at least as many  
2102 events as was observed. This probability is the  $p$ -value:

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

2103 If the  $p$ -value is very small, we might reject the null hypothesis. The  $p$ -value is not the  
2104 probability of the null hypothesis given the data; rather, it expresses the probability  
2105 that data with a certain property was obtained, assuming the null hypothesis [95].

2106 The  $p$ -value is an example of a test statistic  $T$ , which maps the data to a single  
2107 real number. The Neyman-Pearson lemma states that out of the infinite possibilities

2108 of choices of test statistic, the uniformly most powerful test statistic is the likelihood  
2109 ratio  $T_{NP}$  [95]:

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

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:

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

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

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

2110 where  $T_0$  is the value of  $T$  based on the observed data, and  $\theta()$  is the Heaviside  
2111 function. The size of the test is conventionally chosen to be 10%, 5%, or 1%. As  
2112 the  $p$ -value depends on  $\boldsymbol{\alpha}$  (both the POI and NP), the null hypothesis should not be  
2113 rejected if the  $p$ -value is larger than the size of the test for any value of the nuisance  
2114 parameters.

### 2115 9.3.3 Confidence intervals

2116 In an example of the measurement of the Standard Model Higgs boson,  $\boldsymbol{\alpha}_{\text{POI}} =$   
2117  $(\sigma/\sigma_{SM}, M_H)$ , with  $\sigma/\sigma_{SM}$  is the ratio of the production cross-section for Higgs with  
2118 respect to its value in the SM, and  $M_H$  is the unknown mass of the Higgs, values  
2119 of these parameters outside specific bounds are said to be “excluded at the 95%  
2120 confidence level”. These allowed regions are called confidence levels or confidence  
2121 regions, and the parameter values outside of them are considered excluded [95]. A

2122 95% confidence interval does not mean that there is a 95% chance that the true value  
 2123 of the parameter is inside the interval. Rather, a 95% confidence interval covers the  
 2124 true value 95% of the time (even though we do not know the true value).

2125 To construct a confidence interval for a parameter  $\alpha$ , the Neyman Construction  
 2126 is used to invert a series of hypothesis tests; i.e. for each possible value of  $\alpha$ , the null  
 2127 hypothesis is treated as  $\alpha$ , and we perform a hypothesis test based on a test statistic.  
 2128 To construct a 95% confidence interval, we construct a series of hypothesis tests with  
 2129 size of 5%. The confidence interval  $I(\mathcal{D})$  is constructed by taking the set of parameter  
 2130 values  $\boldsymbol{\alpha}$  where the null hypothesis is accepted:

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

2131 where  $T(\mathcal{D})$  is the test statistic, and the last  $\alpha$  (not bolded) and the subscript  $k_\alpha$   
 2132 refer to the size of the test. A schematic of the Neyman construction is shown in Fig.  
 2133 9.1. In a more generalized case, the  $x$ -axis is the test statistic  $T$ .

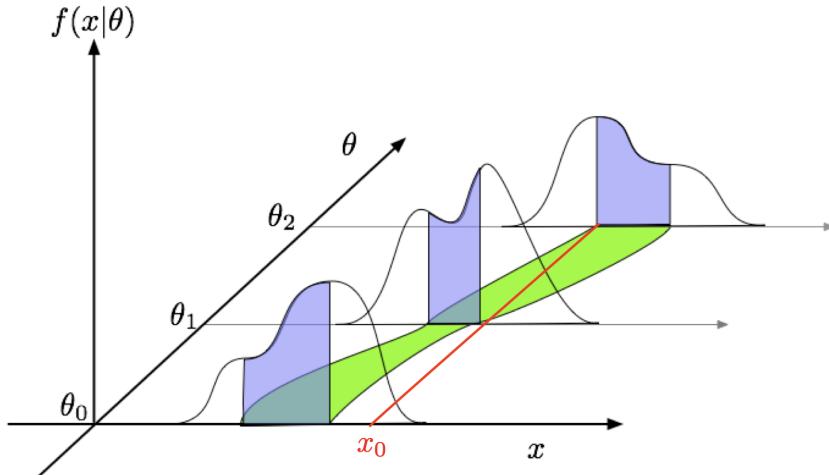


Figure 9.1: Schematic of the Neyman construction for confidence intervals [95]. For each value of  $\theta$ , we find a region in  $x$  where  $\int f(x|\theta)dx$  satisfies the size of the test (blue). These regions form a confidence belt (green). The intersection of the observation  $x_0$  (red) with the confidence belt defines the confidence interval  $[\theta_1, \theta_2]$  [95].

### 2134 9.3.4 Profile likelihood ratio

2135 In this section we describe a frequentist statistical procedure based on the profile  
 2136 likelihood ratio test statistic, which is implemented using asymptotic distributions.

2137 With a multi-parameter likelihood function  $L(\boldsymbol{\alpha})$ , the maximum likelihood of  
 2138 one specific parameter  $\alpha_p$  with other parameters  $\boldsymbol{\alpha}_o$  fixed, is called the conditional  
 2139 maximum likelihood estimate and is denoted  $\hat{\alpha}_p(\boldsymbol{\alpha}_0)$ . The process of choosing specific  
 2140 values of the nuisance parameters for a given value of  $\mu$ ,  $\mathcal{D}_{\text{simulated}}$ , and value of global  
 2141 observables  $\mathcal{G}$  is called profiling. From the full list of parameters  $\boldsymbol{\alpha}$ , we denote the  
 2142 parameter of interest  $\mu$ , and the nuisance parameters  $\boldsymbol{\theta}$ .

2143 We construct the profile likelihood ratio,

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

2144 which depends explicitly on the parameter of interest  $\mu$ , implicitly on the data  $\mathcal{D}_{\text{sim}}$   
 2145 and global observables  $\mathcal{G}$ , and is independent of the nuisance parameters  $\boldsymbol{\theta}$ , which  
 2146 have been eliminated in profiling [95].

2147 The main conceptual reason for constructing the test statistic from the profile  
 2148 likelihood ratio is that asymptotically (i.e. for measurements with many events) the  
 2149 distribution of the profile likelihood ratio  $\lambda(\mu = \mu_{\text{true}})$  is independent of the values of  
 2150 the nuisance parameters [95].

2151 The following  $p$ -value is used to quantify the consistency with the hypothesis of a  
 2152 signal strength of  $\mu$ :

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

2153 **9.3.5 Modified frequentist method:  $CL_S$**

2154 In the modified frequentist method called  $CL_S$ , to test a hypothesis with signal, we  
2155 define  $p'_\mu$  as a ratio of  $p$ -values [95]:

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

2156 where  $p_b$  is the  $p$ -value derived under the background-only hypothesis:

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

2157 The  $CL_S$  upper limit on  $\mu$ , denoted  $\mu_{up}$ , is obtained by solving for  $p'_{\mu_{up}} = 5\%$ .  
2158 If testing the compatibility of the data with the background-only hypothesis, we  
2159 consider the  $p_b$  value defined above and conventionally convert it into the quantile  
2160 or “sigma” of a unit Gaussian.  $z$  standard deviations (e.g.  $z = 5$  in “ $5\sigma$ ”) means  
2161 that the probability of falling above these standard deviations, equals  $p_b$  (e.g.  $3\sigma$   
2162 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  
2163 99.999943%).

2164 **Chapter 10**

2165 **Results**

2166 **10.1 Results from  $bb\tau\tau$**

2167 In each of the three  $\tau\tau$  channels studied ( $\mu\tau_h$ ,  $e\tau_h$ , and  $e\mu$ ), events are divided based  
2168 on whether they contain exactly 1 or 2 b-tag jets, and further divided into signal  
2169 and control regions (SRs and CRs) using the DNN categorization score as described  
2170 in Section 9.2. The control regions demonstrate good agreement between observed  
2171 events in data, and the sum of the contributions from expected backgrounds that  
2172 are modeled in simulated and embedded samples. The signal regions are defined to  
2173 be sensitive to the  $h \rightarrow aa \rightarrow bb\tau\tau$  signal. The postfit final observed and expected  
2174 distributions of the di-tau invariant mass  $m_{\tau\tau}$  reconstructed with SVFit (described  
2175 in Section 5.2) are shown in Fig. 10.1 for the  $\mu\tau_h$  channel, Fig. 10.2 for the  $e\tau_h$   
2176 channel, and Fig. 10.3 for the  $e\mu$  channel. In all figures, the hypothesized yield for  
2177 the  $h \rightarrow aa \rightarrow bb\tau\tau$  signal is shown for the pseudoscalar mass  $m_a = 35$  GeV and  
2178 assuming a branching fraction  $B(H \rightarrow aa \rightarrow bb\tau\tau) = 10\%$ .

2179 The 95% CL expected and observed exclusion limits on the signal strength of the  
2180 branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau)$  as a function of the pseudoscalar mass  $m_a$   
2181 ranging from 12 GeV to 60 GeV, are shown for the three  $\tau\tau$  channels and all three

2182 channels combined in Fig. 10.4. The limits are shown as percentages and normalized  
2183 to the production cross-section of the Standard Model Higgs boson. No excess of  
2184 events above the Standard Model expectations is observed. In the limits for the three  
2185  $\tau\tau$  channels combined, expected (observed) limits range from 1.4 to 5.6% (1.7 to  
2186 7.6%) for pseudoscalar masses between 12 and 60 GeV.

2187 The  $e\mu$  channel is the only channel that has signal sensitivity to the  $m_a = 12$   
2188 GeV pseudoscalar mass hypothesis, because the minimum required spatial separation  
2189  $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  between the two  $\tau$  legs is smaller than the other two channels  
2190 ( $\Delta R < 0.3$  for  $e\mu$ , compared to  $\Delta R < 0.4$  for the other two channels). This decreased  
2191  $\Delta R$  requirement results in better signal acceptance for low mass signals for the  $e\mu$   
2192 channel. The  $\mu\tau_h$  and  $e\tau_h$  channels are most sensitive to the intermediate mass points  
2193 studied, since the analysis targets a resolved signature: at low mass points, the tau  
2194 legs are boosted, and at high mass points, the  $m_{\tau\tau}$  distributions in signal have larger  
2195 overlap with background distributions. In the combination of the three  $\tau\tau$  channels,  
2196 the limit for  $m_a = 12$  GeV comes only from the  $e\mu$  channel, and the best sensitivity  
2197 is attained at intermediate mass points around  $m_a = 20$  GeV to 45 GeV.

2198 To set limits on the branching fraction of the 125 GeV Higgs to the two pseu-  
2199 doscalars,  $B(h \rightarrow aa)$ , we interpret the results in four types of 2HDM+S, which were  
2200 introduced in Section 1.4. In 2HDM+S, the theorized branching fraction of the pseu-  
2201 doscalars depends on the 2HDM+S model type, the pseudoscalar mass  $m_a$ , and the  
2202 ratio of the two Higgs doublets' vacuum expectation values  $\tan\beta$ . In Type I models,  
2203 the branching fraction is independent of  $\tan\beta$ , while in Types II, III, and IV, it is  
2204 a function of  $m_a$  and  $\tan\beta$ . Limits for the  $bb\tau\tau$  final state as a function of  $m_a$  for  
2205 2HDM+S Type I (valid for all  $\tan\beta$  values), Type II with  $\tan\beta = 2.0$ , Type III with  
2206  $\tan\beta = 2.0$ , and Type IV with  $\tan\beta = 0.6$  are overlaid and shown in Fig. 10.5a.

## 2207 10.2 Combination with $bb\mu\mu$ final state

2208 Results from this analysis for the  $h \rightarrow aa \rightarrow bb\tau\tau$  final state are combined with the  
2209 analysis for the  $h \rightarrow aa \rightarrow bb\mu\mu$  final state [96]. While the predicted branching ratio  
2210 for  $aa \rightarrow bb\mu\mu$  is comparatively small, the  $bb\mu\mu$  final state has competitive results  
2211 due to the excellent di-muon resolution measured by CMS. The  $bb\mu\mu$  analysis uses  
2212 an unbinned fit to the data using the di-muon mass  $m_{\mu\mu}$  distribution. Details can be  
2213 found in [96].

2214 Combining the results is possible since the  $bb\tau\tau$  analysis explicitly rejects events  
2215 with extra leptons, so there is no overlap between the events studied in the  $bb\tau\tau$   
2216 analysis and the  $bb\mu\mu$  analysis. In the statistical combination, several systematic  
2217 uncertainties are treated as correlated: the integrated luminosity normalization, the  
2218 b-tagging scale factor, the scale factors related to muon reconstruction, identifica-  
2219 tion, and trigger efficiencies, the inefficiency in the ECAL trigger readout, and the  
2220 theoretical uncertainties related to signal modeling.

2221 Since the results in both final states are statistically limited, the combination ben-  
2222 efits from the additional data. For  $m_a = 35$  GeV, all systematic uncertainties amount  
2223 to around 6% of the total uncertainty, with the dominant systematic uncertainties  
2224 coming from jet energy systematics in the  $bb\mu\mu$  final state, theoretical uncertainties  
2225 in the signal, and uncertainties in the QCD multijet backgrounds in the  $e\mu$  channel  
2226 of the  $bb\tau\tau$  final state.

2227 The mass distributions of the di-muon and di-tau objects ( $m_{\mu\mu}$  and  $m_{\tau\tau}$ ) are  
2228 compared to the data in a combined maximum likelihood fit to derive upper limits  
2229 on  $B(h \rightarrow aa)$ . The observed limits at 95% CL on  $B(h \rightarrow aa)$  for different 2HDM+S  
2230 scenarios, are shown for the search for  $h \rightarrow aa \rightarrow bb\mu\mu$  in Fig. 10.5b, and the  
2231 combined analyses  $h \rightarrow aa \rightarrow bb\ell\ell$  in Fig. 10.6.

2232 Exclusion limits in a two-dimensional plane as a function of  $\tan\beta$  and  $m_a$  are  
2233 set for 2HDM+S Types II, III, and IV in Fig. 10.7. The most stringent constraints

are observed for 2HDM+S type III because of large branching fractions predicted in theory, with predicted branching fractions between 0.47 and 0.42 for  $\tan \beta = 2.0$  and values of  $m_a$  between 15 and 60 GeV, compared to the observed 95% CL upper limits which are between 0.08 and 0.03. For 2HDM+S type IV, the predicted branching fractions from theory are between 0.26 and 0.20 for  $\tan \beta = 0.6$  for values of  $m_a$  between 15 and 60 GeV, and the 95% CL observed upper limits are between 0.12 and 0.05.

The combined results from  $h \rightarrow aa \rightarrow bb\ell\ell$  are compared with CMS results in other final states as a function of the pseudoscalar mass  $m_a$ : for 2HDM+S type I in Fig. 10.8, type II with  $\tan \beta = 2.0$  in Fig. 10.9, and type III with  $\tan \beta = 2.0$  in Fig. 10.10. In other scenarios, e.g. type III with  $\tan \beta = 5.0$ , more stringent limits are set by analyses in other final states,  $\mu\mu\tau\tau$  in this case. Other summary plots for other model types and  $\tan \beta$  values can be found at [97].

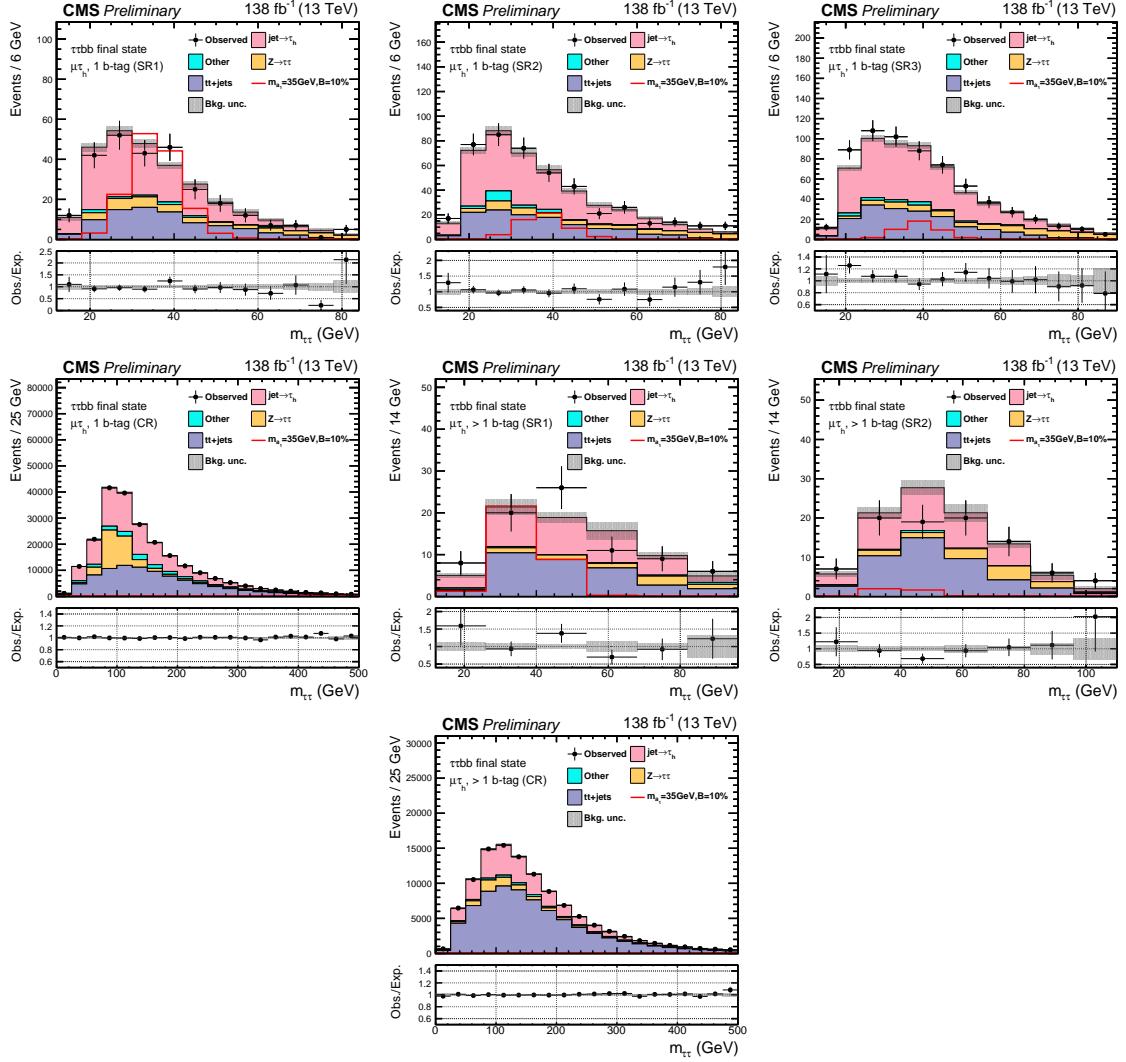


Figure 10.1: Postfit final  $m_{\tau\tau}$  observed and expected distributions, and the observed/expected ratios, in the  $\mu\tau_h$  channel [39]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), 1 b-tag jet control region (*middle row*), 2 b-tag jet signal regions (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*).

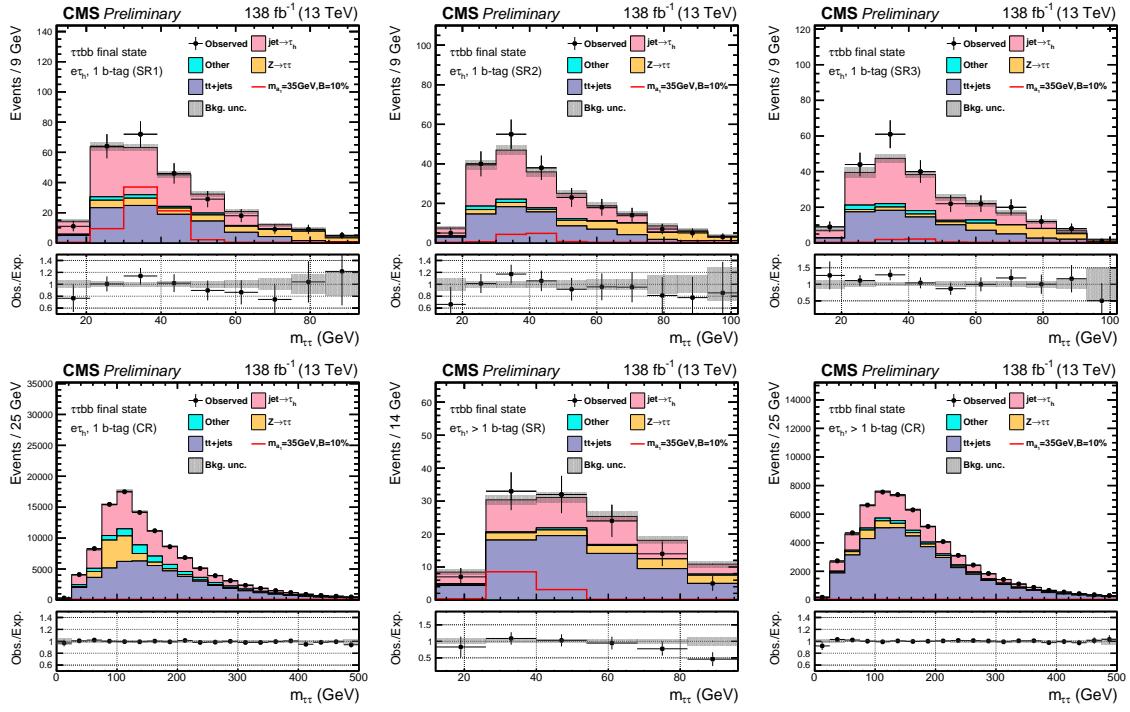


Figure 10.2: Postfit final observed and expected  $m_{\tau\tau}$  distributions, and the observed/expected ratios, in the  $e\tau_h$  channel [39]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), the 1 b-tag jet control region (CR) (*bottom row*), and 2 b-tag jet signal 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  $t\bar{t}+j$  (*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*).

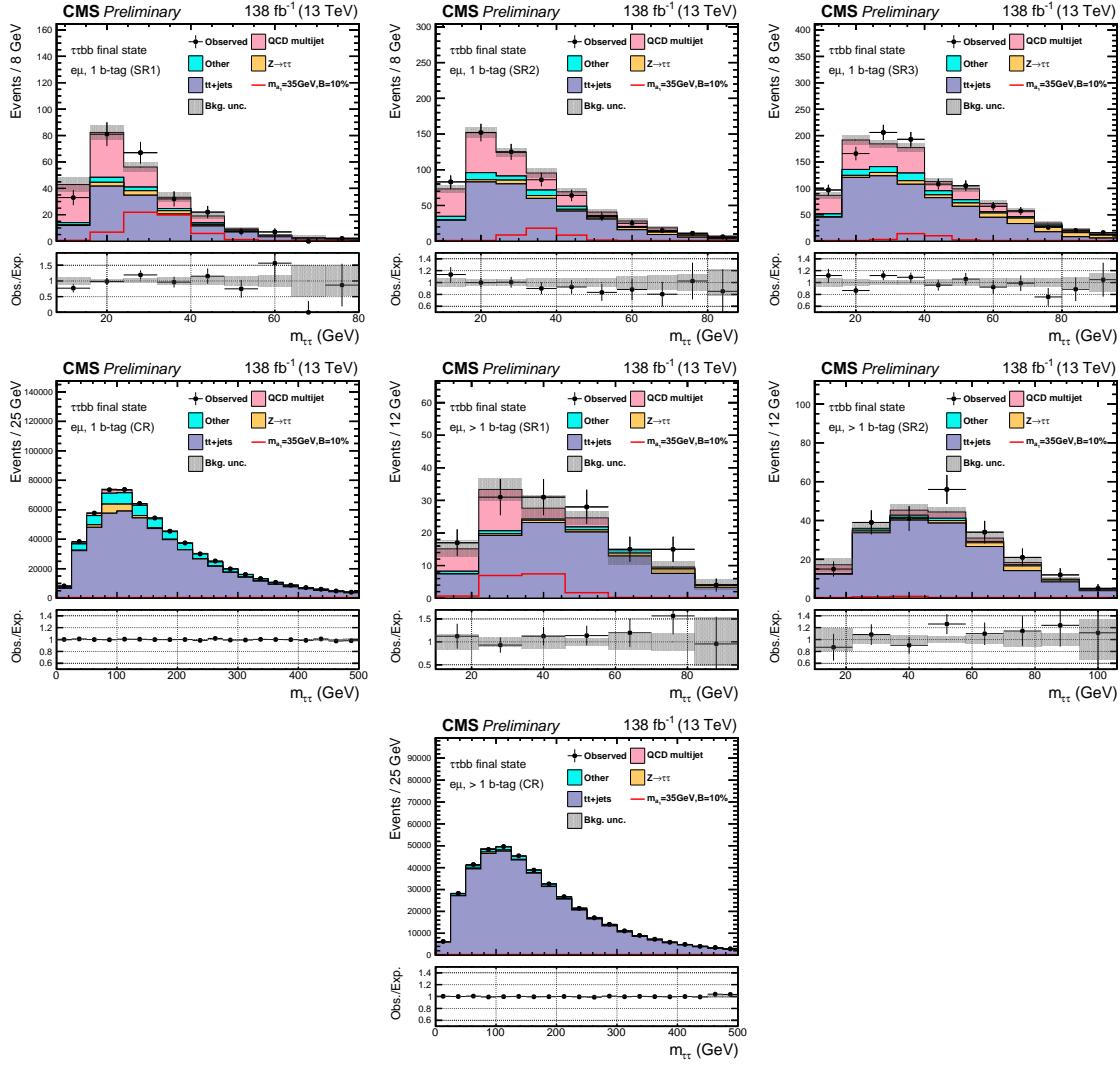


Figure 10.3: Postfit final observed and expected  $m_{\tau\tau}$  distributions, and the observed/expected ratios, in the  $e\mu$  channel [39]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, and SR3) (*top row*), 1 b-tag jet control region (CR) (*middle row*), 2 b-tag jet signal regions (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. TFor 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*).

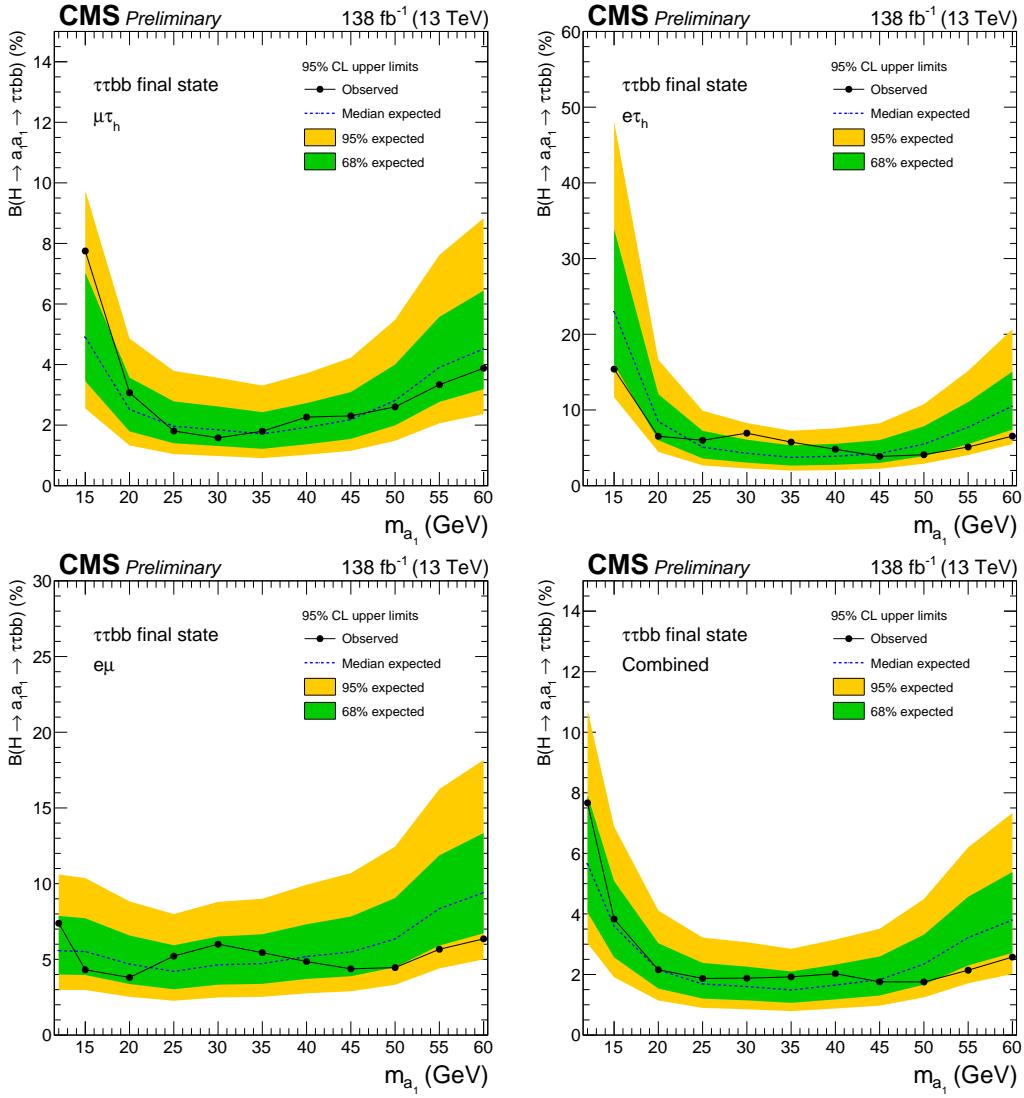
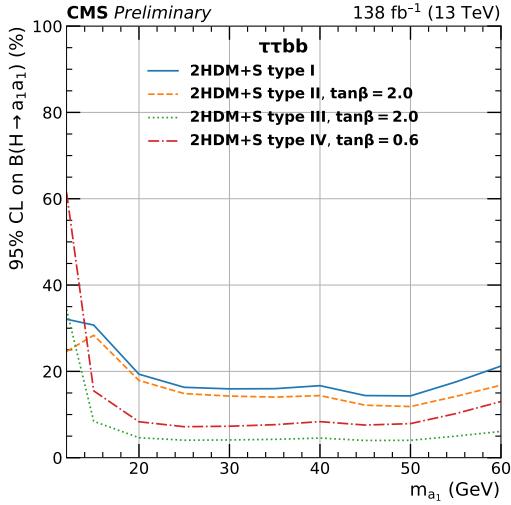
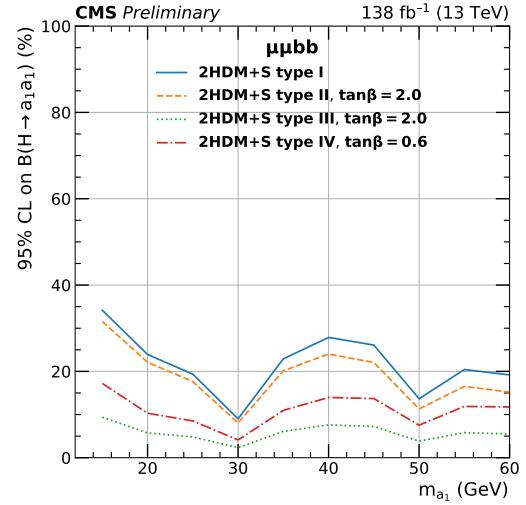


Figure 10.4: Observed 95% CL exclusion limits (*black, solid lines*) and expected 95% CL and 68% CL limits (*shaded yellow and green*) on the branching fraction  $B(h \rightarrow aa \rightarrow bb\tau\tau)$  in percentages, assuming the Standard Model production for the 125 GeV Higgs ( $h$ ). Limits 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*) [39]. 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  $m_a = 12$  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) [39]. Linear interpolation is used between points in the graphs. The  $\tan\beta$  values chosen here correspond to the most stringent limits in each model.

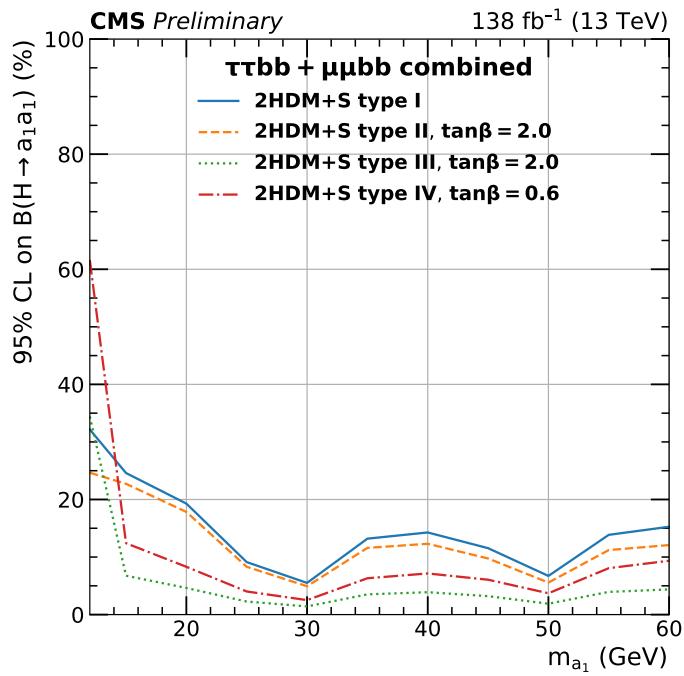


Figure 10.6: Observed 95% CL upper limits on the branching fraction of the 125 GeV Higgs boson to two pseudoscalars,  $B(h \rightarrow aa)$ , in percentages, as a function of the pseudoscalar mass  $m_a$ , in 2HDM+S type I (blue), type II with  $\tan\beta = 2.0$  (orange dashed), type III with  $\tan\beta = 2.0$  (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  $138 \text{ fb}^{-1}$  [39].

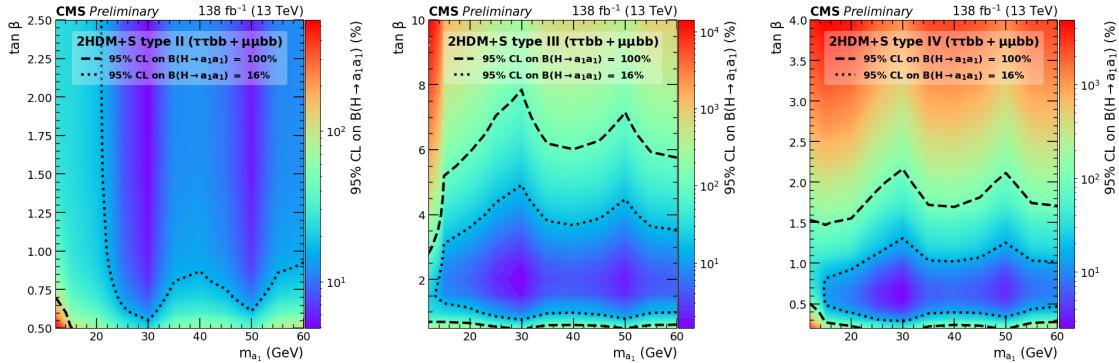


Figure 10.7: Observed 95% CL upper limits on  $\mathcal{B}(h \rightarrow aa)$  in %, for the combination of  $bb\mu\mu$  and  $bb\tau\tau$  channels using the full Run 2 integrated luminosity of  $138 \text{ fb}^{-1}$  for Type II (*left*), Type III (*middle*), and Type IV (*right*) 2HDM+S in the  $\tan \beta$  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 [39].

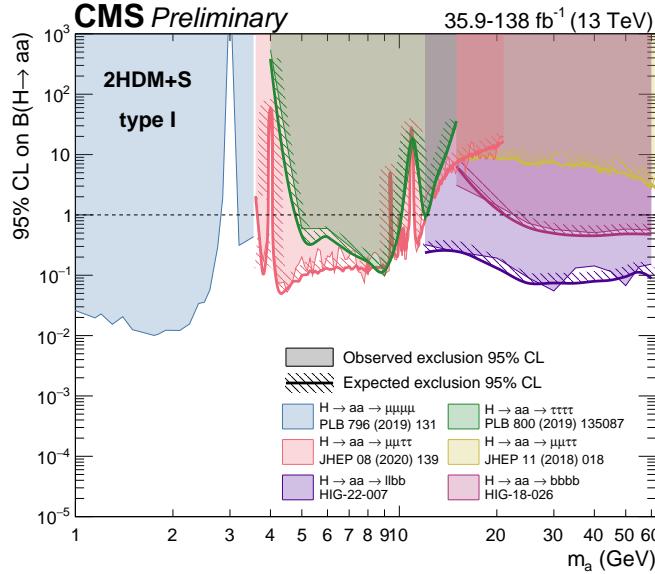


Figure 10.8: Summary plot of current 95% limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section,  $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$  in the 2HDM+S type I scenario performed with data collected at 13 TeV [97]. 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).

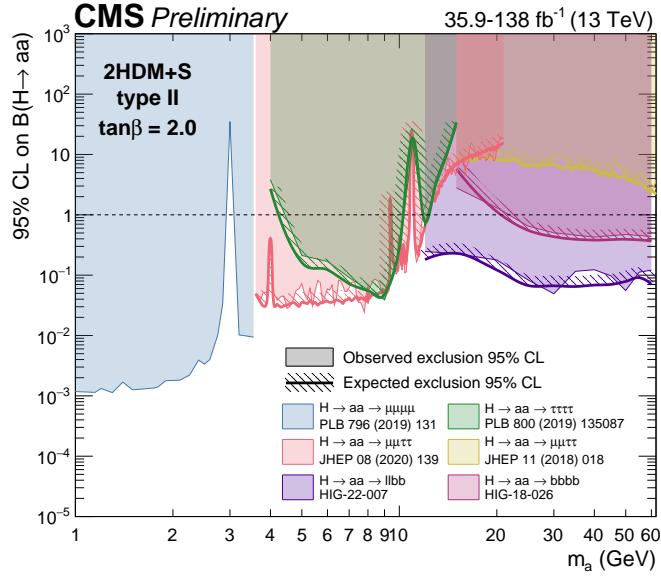


Figure 10.9: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section,  $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$ , in the 2HDM+S type II scenario with  $\tan \beta = 2.0$ , obtained at CMS with data collected at 13 TeV [97]. 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).

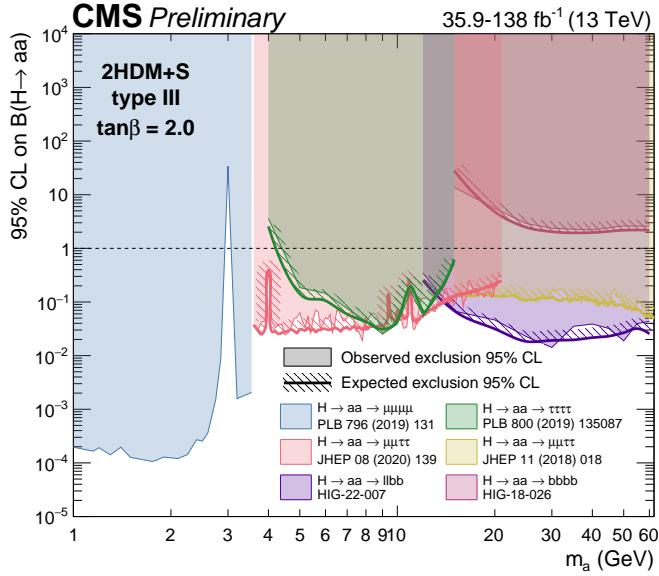


Figure 10.10: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross section,  $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$  in the 2HDM+S type-III scenario with  $\tan \beta = 2.0$ , obtained at CMS with data collected at 13 TeV [97]. 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).

2247 **Chapter 11**

2248 **Asymmetric exotic Higgs decays**

2249 This chapter presents progress towards a search for exotic Higgs decays to two light  
2250 scalars with unequal mass ( $h \rightarrow a_1 a_2$ ) final states with bottom quarks and  $\tau$  leptons,  
2251 with plans to interpret the results in the context of Two Real Singlet Models (TRSMs),  
2252 described in Section 1.5. Compared to the symmetric decay scenario  $h \rightarrow aa$  which  
2253 has been studied in multiple final states at CMS with stringent limits set on the  
2254 various 2HDM+S scenarios, this asymmetric decay scenario has not been directly  
2255 searched for at the CMS experiment. Section 11.1 lists the mass hypotheses of the  
2256 new particles  $a_1$  and  $a_2$  that will be studied. Section 11.2 describes the studies on  
2257 which channels the analysis will be carried out in. Section 11.3 shows the control  
2258 plots produced using the analysis framework that will be used for this analysis.

2259 **11.1 Signal masses**

2260 As discussed in Section 1.5,  $h \rightarrow a_1 a_2$  can result in a “cascade” decay if one of the  
2261 scalars,  $a_2$  is sufficiently heavy ( $m_{a_2} > 2m_{a_1}$ ). The “non-cascade” case is where the  
2262 light scalars decay directly to Standard Model particles.

2263 The mass hypotheses (mass points) ( $m_{a_1}, m_{a_2}$ ) studied here are:

- *Cascade mass points:* (15, 30), (15, 40), (15, 50), (15, 60), (15, 70), (15, 80), (15, 90), (15, 100), (15, 110), (20, 40), (20, 50), (20, 60), (20, 70), (20, 80), (20, 90), (20, 100), (30, 60), (30, 70), (30, 80), and (30, 90) GeV
- *Non-cascade mass points:* (15, 20), (15, 30), (20, 30), (20, 40), (30, 40), (30, 50), (30, 60), (40, 50), (40, 60), (40, 70), (40, 80), (50, 60), and (50, 70) GeV

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

## 11.2 Cascade scenario signal studies

Studies of the signal phenomenology in the cascade scenario were performed to determine the viability of the  $4b2\tau$  and/or  $2b4\tau$  channels.

Cross sections and branching fractions of the  $4b2\tau$  and  $2b4\tau$  final states were compared using cross-section predictions provided by the authors of [5]. For an example mass point  $m_{a_2} = 80$  GeV,  $m_{a_1} = 30$  GeV, the branching fractions to  $4b2\tau$  is ten times larger than  $2b4\tau$ :  $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 4b2\tau) = 0.00857$ , vs.  $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 2b4\tau) = 0.00068$ . The  $4b2\tau$  final state is chosen for this analysis.

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

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

2289 and fourth b-flavor jets which have very low transverse momenta  $p_T$ . Event categories  
 2290 with exactly 1 b-tag jet and  $\geq 2$  b-tag jets will be used.

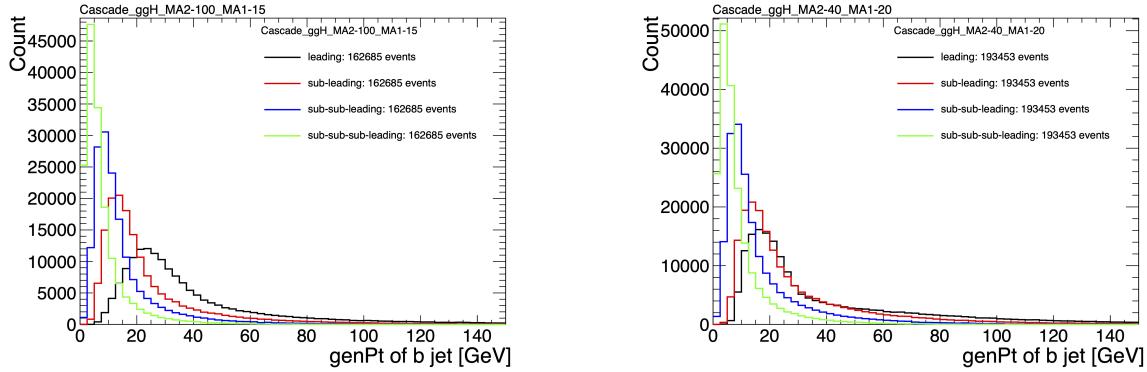


Figure 11.1: Generator-level b-flavor jet transverse momenta  $p_T$ , for  $h \rightarrow a_1 a_2$  cascade scenario in the  $4b2\tau$  final state, for mass hypotheses  $(m_{a_1}, m_{a_2}) = (100, 15)$  GeV (*left*) and  $(40, 20)$  GeV (*right*). In each plot the generator-level  $p_T$  of the leading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

2291 In the  $4b2\tau$  final state, the possibility of the leading and sub-leading b-tag jets  
 2292 being sufficiently close in  $\Delta R$  to require boosted jet reconstruction techniques was  
 2293 explored. In the  $4b2\tau$  case, the two b-flavor-jets in the generated event that were  
 2294 spatially closest in  $\Delta R$  were considered as one object. This two b-flavor jet object was  
 2295 spatially matched in  $\Delta R$  to the jets reconstructed with the standard AK4 algorithm  
 2296 which uses a cone size of  $\Delta R = 0.4$ . The quality of the  $p_T$  resolution (computed as  
 2297  $(p_{T,\text{reconstructed}} - p_{T,\text{gen}})/p_{T,\text{gen}}$ ) and closeness in distance  $\Delta R$  of the reconstructed jet  
 2298 to the nearest generator-level jets, was seen to depend on the absolute and relative  
 2299 masses of the light scalars. The best (worst) performance occurred in samples with  
 2300 large (small) mass differences between the heavier scalar  $a_2$  and the lighter scalar  $a_1$ ,  
 2301 as illustrated for the mass hypotheses  $(m_{a_1}, m_{a_2})$  (100, 15) GeV and (40, 20) GeV in  
 2302 Fig. 11.2.

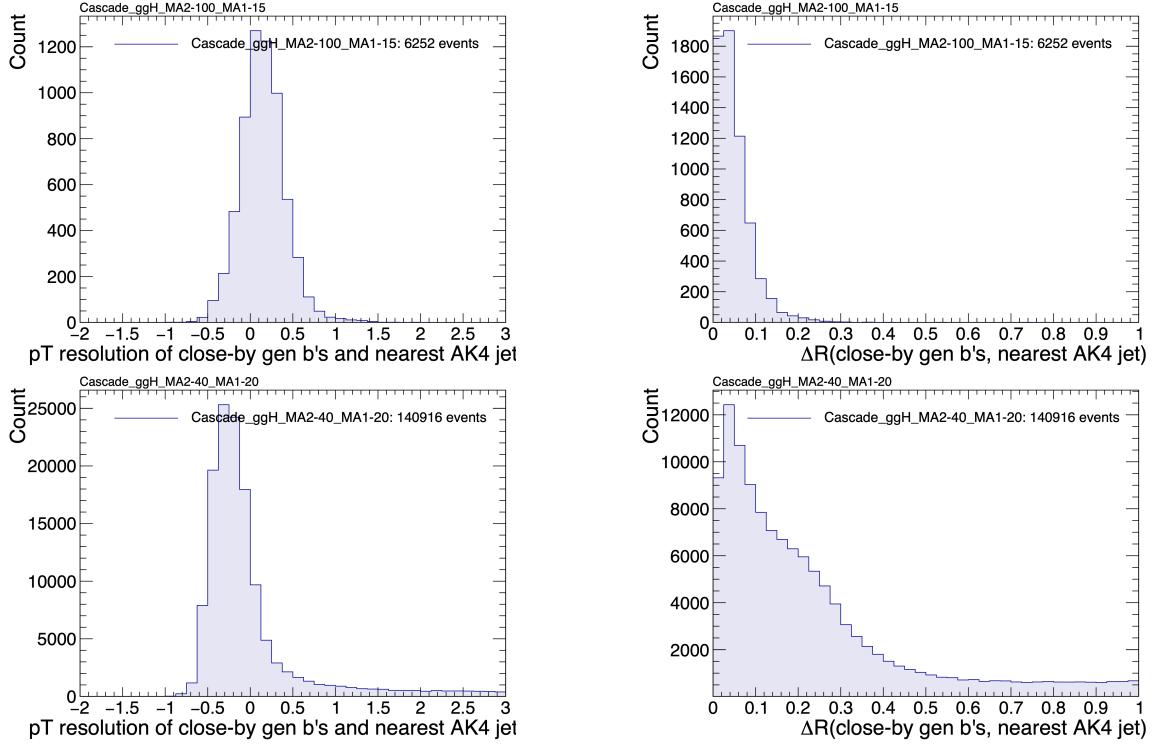


Figure 11.2: Distributions (arbitrary units) of transverse momentum  $p_T$  resolution and  $\Delta R$  between the two closest generator-level  $b$  jets, treated as one object, and the nearest reconstructed AK4 jet, for two different  $h \rightarrow a_1 a_2$  mass hypotheses ( $m_{a_1}, m_{a_2}$ ) = (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 Current control plots for $\mu\tau_h$ channel

The  $\tau\tau$  states for the  $h \rightarrow a_1 a_2$  to  $4b2\tau$  analysis will be similar to those studied in  $h \rightarrow aa \rightarrow bb\tau\tau$ . For the  $\mu\tau_h$  channel, histograms of the key kinematic variables are made for data and the sum of the expected backgrounds, which are estimated from Monte Carlo samples, embedded samples, and the data-driven method for estimating jets faking  $\tau_h$  as described in Chapter 7. Nominal values of the scale factors and event reweighting are applied, as described in Chapter ???. The errors shown in the figures only 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,

2312 described in Sections 5.3.1, 5.3.2, and 5.3.4).

2313 The  $p_T$ ,  $\eta$ , and  $\phi$  of the leading muon and hadronic tau  $\tau_h$ , and the di-tau visible  
2314 mass  $m_{\text{vis}}$  and momentum  $p_{T,\text{vis}}$ , are shown in Fig. 11.3. The  $p_T$ ,  $\eta$ , and  $\phi$  of the the  
2315 leading and sub-leading b-tag jets, and the missing transverse energy magnitude and  
2316 azimuthal direction, are shown in Fig. 11.4.

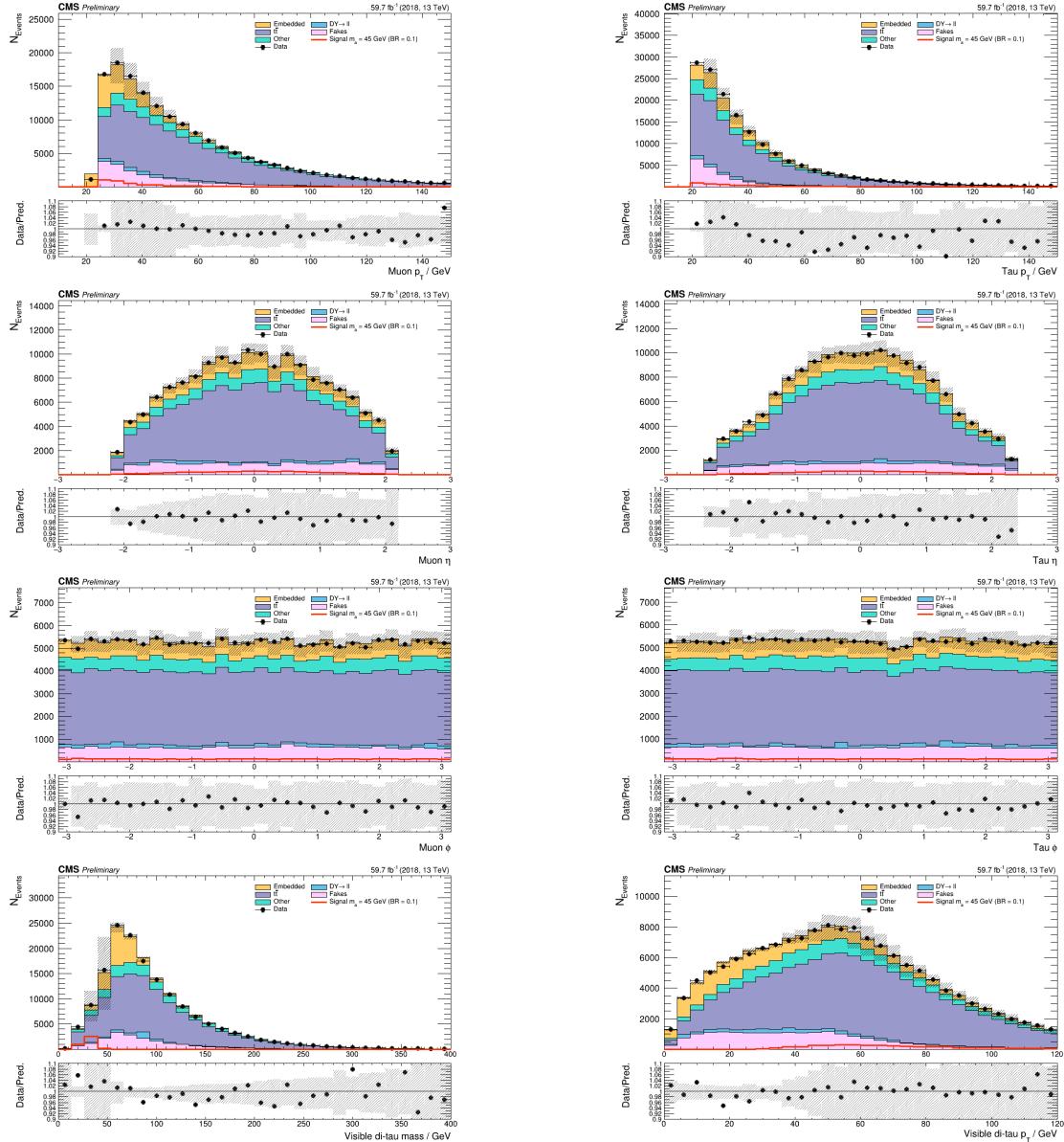


Figure 11.3: Kinematic properties of the leading muon and  $\tau_h$  in the  $\mu\tau_h$  channel:  $p_T$  (top row),  $\eta$  (second row), and  $\phi$  (third row). The visible 4-momenta of the muon and  $\tau_h$  are summed, giving the visible di-tau mass  $m_{\text{vis}}$  and transverse momentum  $p_{T,\text{vis}}$ . The errors shown in the figures only 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).

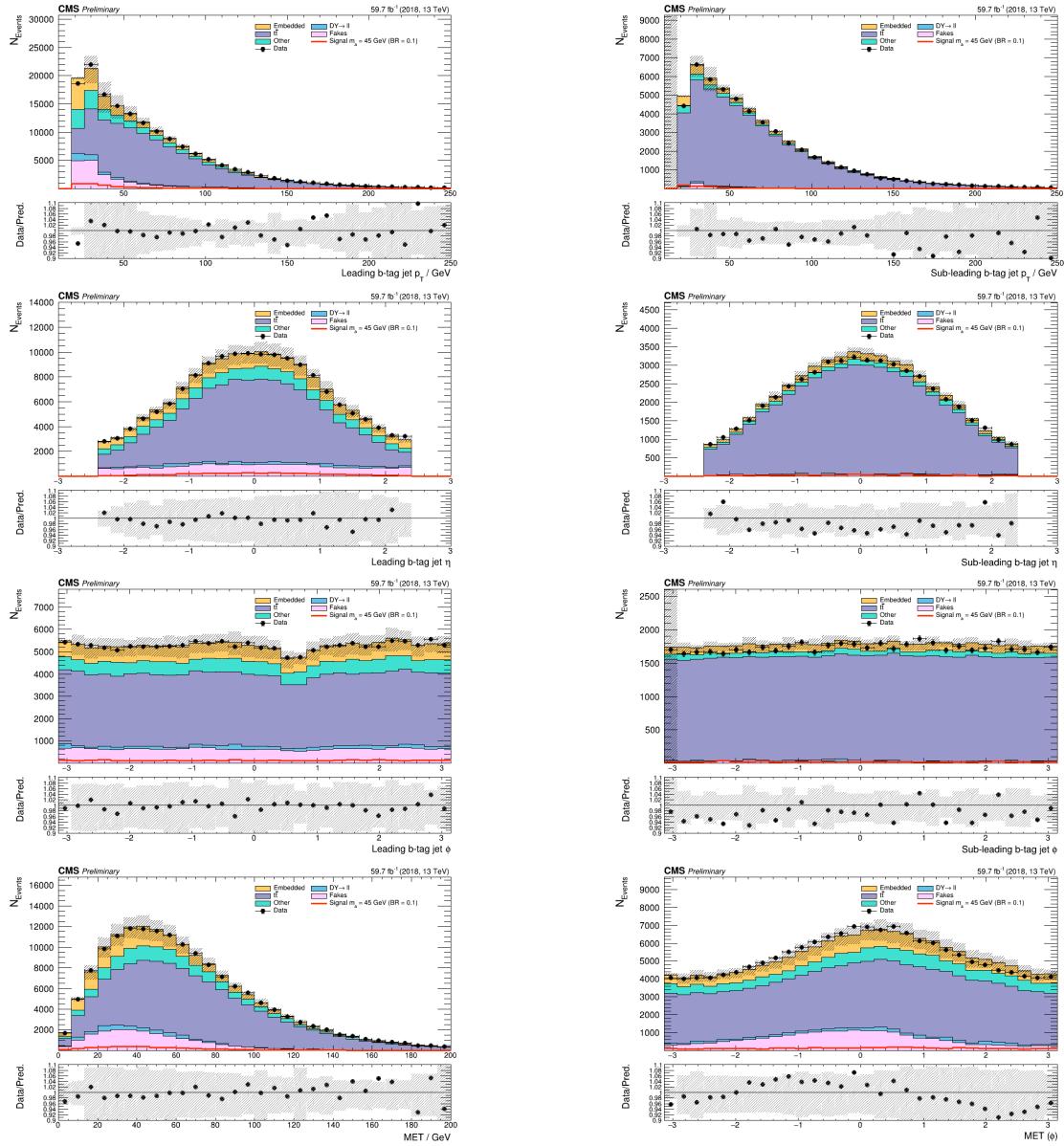


Figure 11.4: Kinematic properties of the leading and sub-leading b-tag jets in the  $\mu\tau_h$  final state: jet  $p_T$  (*top row*),  $\eta$  (*second row*),  $\phi$  (*third row*), as well as the missing transverse energy magnitude and azimuthal direction (*bottom row*). The errors shown in the figures only 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).

<sup>2317</sup> **Chapter 12**

<sup>2318</sup> **Conclusion and outlook**

<sup>2319</sup> With the discovery of a Higgs boson with mass 125 GeV at the LHC in 2012, the LHC  
<sup>2320</sup> and CMS physics program has evolved to include the precise characterization of the  
<sup>2321</sup> 125 GeV Higgs boson and searching for evidence of additional Higgs particles in an  
<sup>2322</sup> extended Higgs sector. This thesis presents a direct search at CMS for exotic decays  
<sup>2323</sup> of the Higgs boson with mass 125 GeV in data collected in the years 2016-2018 in  
<sup>2324</sup> proton-proton collisions at center-of-mass energy 13 TeV, to two light neutral scalar  
<sup>2325</sup> particles that decay to two bottom quarks and two tau leptons ( $h \rightarrow aa \rightarrow bb\tau\tau$ ). The  
<sup>2326</sup> results are combined with another search that was performed in the  $h \rightarrow aa \rightarrow bb\mu\mu$   
<sup>2327</sup> final state, giving the most stringent limits to date for theories with Two Higgs  
<sup>2328</sup> Doublet Models extended with a singlet scalar (2HDM+S), for pseudoscalar masses  
<sup>2329</sup>  $m_a$  ranging from 15 GeV to 60 GeV, in a number of 2HDM+S scenarios such as type  
<sup>2330</sup> II and III with  $\tan\beta = 2.0$ .

<sup>2331</sup> As the rich physics program of CMS has set stringent limits on the exotic decay  
<sup>2332</sup>  $h \rightarrow aa$ , we turn our attention to direct searches for decays to light neutral scalars  
<sup>2333</sup> with potentially unequal mass,  $h \rightarrow a_1a_2$ , which has not been performed at CMS  
<sup>2334</sup> to date. Preliminary studies on  $h \rightarrow a_1a_2$  signals in the Two Real Singlet Model  
<sup>2335</sup> (TRSM) are shown, and work is ongoing to develop the analysis for  $h \rightarrow a_1a_2$  in final

2336 states with bottom quarks and tau leptons.

2337 To ensure the continued performance of the CMS detector and to enhance its  
2338 data-taking capabilities in the intense pileup conditions of the Phase-2 upgrade of  
2339 the High-Luminosity LHC, upgrades of the Level-1 Trigger are paramount for filter-  
2340 ing the increased data rate of the HL-LHC. This thesis presents work on the stan-  
2341 dalone barrel calorimeter algorithm for reconstructing and identifying electron and  
2342 photon candidates, using high granularity crystal-level information from the ECAL  
2343 subdetector. For Phase-2, the increase in the granularity of information sent from  
2344 the electromagnetic calorimeter to the Level-1 trigger, from energy sums over towers  
2345 (which are  $5 \times 5$  in crystals) to crystal-level information, allows for the implementation  
2346 of a more sophisticated clustering algorithm that can exploit the fact that genuine  
2347 electrons and photons tend to leave energies concentrated a  $3 \times 5$  window in crystals,  
2348 and use shape and isolation information to distinguish genuine electrons and photons  
2349 from noise. Electrons and photons are key to characterizing Standard Model pro-  
2350 cesses and performing searches for new physics, and this represents one of the many  
2351 upgrades of the CMS detector in preparation for Phase-2. With the ongoing Run-3  
2352 data collecting period, and wealth of ongoing and scheduled upgrades, there remains  
2353 an abundance of directions for detector development and physics at CMS heading  
2354 into Phase-2 of the LHC.

- 2355 **Bibliography**
- 2356 [1] Paul H. Frampton. Journeys Beyond the Standard Model. 54(1):52–  
2357 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).
- 2358
- 2359
- 2360 [2] Christopher G. Tully. *Elementary Particle Physics in a Nutshell*. Princeton  
2361 University Press, Princeton, 2012. ISBN 9781400839353. doi: doi:10.1515/  
2362 9781400839353. URL <https://doi.org/10.1515/9781400839353>.
- 2363 [3] John Ellis. Higgs Physics. In *2013 European School of High-Energy Physics*,  
2364 pages 117–168, 2015. doi: 10.5170/CERN-2015-004.117.
- 2365 [4] David Curtin, Rouven Essig, Stefania Gori, and Others. Exotic decays of the 125  
2366 GeV Higgs boson. *Phys. Rev. D*, 90:075004, Oct 2014. doi: 10.1103/PhysRevD.  
2367 90.075004. URL <https://link.aps.org/doi/10.1103/PhysRevD.90.075004>.
- 2368 [5] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt. Two-real-scalar-singlet  
2369 extension of the SM: LHC phenomenology and benchmark scenarios. *Eur. Phys.  
2370 J. C*, 80(2):151, 2020. doi: 10.1140/epjc/s10052-020-7655-x.
- 2371 [6] CERN. The history of CERN, 2024. URL <https://timeline.web.cern.ch/timeline-header/89>.
- 2372
- 2373 [7] R. Schmidt. Accelerator physics and technology of the LHC. In *ROXIE: Routine*

2374       *for the Optimizazation of Magnet X-Sections, Inverse Field Calculation and Coil*  
2375       *End Design*, pages 7–17, 1998.

2376       [8] 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.

2379       [9] Antonella Del Rosso. Aerial view of the LHC and the four major experiments.  
2380       2017. URL <https://cds.cern.ch/record/2253966>. General Photo.

2381       [10] ATLAS Collaboration. ATLAS: Detector and physics performance technical  
2382       design report. Volume 1. 5 1999.

2383       [11] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector  
2384       Performance and Software. 2006.

2385       [12] L Musa. Conceptual Design Report for the Upgrade of the ALICE ITS. Technical  
2386       report, CERN, Geneva, 2012. URL <https://cds.cern.ch/record/1431539>.

2387       [13] S. Amato et al. LHCb technical proposal: A Large Hadron Collider Beauty  
2388       Experiment for Precision Measurements of CP Violation and Rare Decays. 2  
2389       1998.

2390       [14] Werner Herr and B Muratori. Concept of luminosity. 2006. doi: 10.5170/  
2391       CERN-2006-002.361. URL <https://cds.cern.ch/record/941318>.

2392       [15] Olivier S. Brning and Frank Zimmerman. Parameter space for the LHC lumi-  
2393       nosity upgrade. ISBN 978-3-95450-115-1. URL <https://accelconf.web.cern.ch/IPAC2012/papers/moppc005.pdf>.

2395       [16] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09):  
2396       P09018, 2020. doi: 10.1088/1748-0221/15/09/P09018.

- 2397 [17] CMS Collaboration. Measurement of the inelastic proton-proton cross section at  
2398        $\sqrt{s} = 13$  TeV. *JHEP*, 07:161, 2018. doi: 10.1007/JHEP07(2018)161.
- 2399 [18] CMS Collaboration. High-Luminosity Large Hadron Collider (HL-LHC): Tech-  
2400       nical design report. 10/2020, 12 2020. doi: 10.23731/CYRM-2020-0010.
- 2401 [19] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004,  
2402       2008. doi: 10.1088/1748-0221/3/08/S08004.
- 2403 [20] CMS Collaboration. Particle-flow reconstruction and global event description  
2404       with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/  
2405       12/10/P10003.
- 2406 [21] V Karimki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freuden-  
2407       reich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker*  
2408       *system project: Technical Design Report*. Technical design report. CMS. CERN,  
2409       Geneva, 1997. URL <https://cds.cern.ch/record/368412>.
- 2410 [22] The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva,  
2411       2017. URL <https://cds.cern.ch/record/2272264>.
- 2412 [23] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Up-  
2413       grade. 9 2012. doi: 10.2172/1151650.
- 2414 [24] R. L. Workman and Others. Review of particle physics. 2022:083C01. doi:  
2415       10.1093/ptep/ptac097.
- 2416 [25] CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorime-  
2417       ter. 9 2012. doi: 10.2172/1151651.
- 2418 [26] CMS Technical Design Report for the Level-1 Trigger Upgrade. 6 2013.
- 2419 [27] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The  
2420       trigger systems. 12 2000.

- 2421 [28] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and  
2422 High Level Trigger. Technical report, CERN, Geneva, 2021. URL <https://cds.cern.ch/record/2759072>. This is the final version of the document, approved  
2423 by the LHCC.
- 2425 [29] C. Foudas. The CMS Level-1 Trigger at LHC and Super-LHC. In *34th Interna-*  
2426 *tional Conference on High Energy Physics*, 10 2008.
- 2427 [30] CMS Software Guide. High Level Trigger (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHighLevelTrigger>.
- 2429 [31] The Worldwide LHC Computing Grid. 2012. URL <https://cds.cern.ch/record/1997398>.
- 2431 [32] Douglas Thain, Todd Tannenbaum, and Miron Livny. Distributed computing in  
2432 practice: the Condor experience. *Concurrency and Computation: Practice and*  
2433 *Experience*, 17(2-4):323–356, 2005. doi: <https://doi.org/10.1002/cpe.938>. URL  
2434 <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.938>.
- 2435 [33] Alexandre Zabi, Jeffrey Wayne Berryhill, Emmanuelle Perez, and Alexander D.  
2436 Tapper. The Phase-2 Upgrade of the CMS Level-1 Trigger. 2020.
- 2437 [34] Technical proposal for a MIP timing detector in the CMS experiment Phase 2  
2438 upgrade. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2296612>.
- 2440 [35] CMS Collaboration. CMS luminosity measurement for the 2016 data-taking  
2441 period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL  
2442 <https://cds.cern.ch/record/2257069>.
- 2443 [36] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking

- 2444 period at  $\sqrt{s} = 13$  TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-  
2445 004, 2018. URL <https://cds.cern.ch/record/2621960>.
- 2446 [37] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking  
2447 period at  $\sqrt{s} = 13$  TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-  
2448 002, 2019. URL <https://cds.cern.ch/record/2676164>.
- 2449 [38] CMS LUMI Group. CMS Luminosity Public Results (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- 2450 [39] Cécile Caillol, Pallabi Das, Sridhara Dasu, Pieter Everaerts, Stephanie Kwan,  
2451 Isobel Ojalvo, and Ho-Fung Tsoi. CMS AN-20-213 (internal): Search for an  
2452 exotic decay of the 125 GeV Higgs boson to light pseudoscalars, with a pair of b  
2453 jets and a pair of tau leptons in the final state, 2020.
- 2454 [40] CMS Collaboration. Search for exotic decays of the Higgs boson to a pair of  
2455 pseudoscalars in the  $\mu\mu bb$  and  $\tau\tau bb$  final states. *European Physical Journal C*,  
2456 2 2024.
- 2457 [41] J. Alwall, R. Frederix, S. Frixione, et al. The automated computation of tree-  
2458 level and next-to-leading order differential cross sections, and their matching to  
2459 parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.  
2460 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).
- 2461 [42] R. Frederix, S. Frixione, V. Hirschi, et al. The automation of next-to-leading  
2462 order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July  
2463 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).
- 2464 [43] S. Agostinelli, J. Allison, K. Amako, et al. Geant4 - a simulation toolkit. 506(3):

- 2468 250–303. ISSN 0168-9002. doi: 10.1016/S0168-9002(03)01368-8. URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- 2469
- 2470 [44] CMS Collaboration. An embedding technique to determine  $\tau\tau$  backgrounds  
2471 in proton-proton collision data. *JINST*, 14(06):P06032, 2019. doi: 10.1088/  
2472 1748-0221/14/06/P06032.
- 2473 [45] CMS Collaboration. Search for neutral MSSM Higgs bosons decaying to a pair of  
2474 tau leptons in pp collisions. *JHEP*, 10:160, 2014. doi: 10.1007/JHEP10(2014)160.
- 2475 [46] CMS Collaboration. Measurements of Higgs boson production in the decay chan-  
2476 nel with a pair of  $\tau$  leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Eur.*  
2477 *Phys. J. C*, 83(7):562, 2023. doi: 10.1140/epjc/s10052-023-11452-8.
- 2478 [47] CMS Collaboration. Reconstruction and identification of tau lepton decays to  
2479 hadrons and tau neutrinos at CMS. *Journal of Instrumentation*, 11(01):P01019–  
2480 P01019, January 2016. ISSN 1748-0221. doi: 10.1088/1748-0221/11/01/p01019.  
2481 URL <http://dx.doi.org/10.1088/1748-0221/11/01/P01019>.
- 2482 [48] CMS Collaboration. Performance of  $\tau$ -lepton reconstruction and identification  
2483 in CMS. *Journal of Instrumentation*, 7(01):P01001, jan 2012. doi: 10.1088/  
2484 1748-0221/7/01/P01001. URL <https://dx.doi.org/10.1088/1748-0221/7/01/P01001>.
- 2485
- 2486 [49] CMS Collaboration. Performance of reconstruction and identification of  $\tau$  leptons  
2487 decaying to hadrons and  $\nu_\tau$  in pp collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 13(10):  
2488 P10005, 2018. doi: 10.1088/1748-0221/13/10/P10005.
- 2489 [50] CMS Collaboration. Identification of hadronic tau lepton decays using a deep  
2490 neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/  
2491 P07023.

- 2492 [51] CMS Collaboration. Performance of CMS muon reconstruction in pp collision  
2493 events at  $\sqrt{s} = 7$  TeV. *Journal of Instrumentation*, 7(10):P10002–P10002,  
2494 October 2012. ISSN 1748-0221. doi: 10.1088/1748-0221/7/10/p10002. URL  
2495 <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- 2496 [52] CMS Collaboration. Performance of electron reconstruction and selection with  
2497 the CMS detector in proton-proton collisions at  $\sqrt{s} = 8$  TeV. *Journal of Instru-*  
2498 *mentation*, 10(06):P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005. URL  
2499 <https://dx.doi.org/10.1088/1748-0221/10/06/P06005>.
- 2500 [53] CMS Collaboration. Identification of b-quark jets with the CMS experiment.  
2501 *Journal of Instrumentation*, 8(04):P04013–P04013, April 2013. ISSN 1748-  
2502 0221. doi: 10.1088/1748-0221/8/04/p04013. URL <http://dx.doi.org/10.1088/1748-0221/8/04/P04013>.
- 2504 [54] CMS Collaboration. Pileup Removal Algorithms. Technical report, CERN,  
2505 Geneva, 2014. URL <https://cds.cern.ch/record/1751454>.
- 2506 [55] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and  
2507 developments. 2017. URL <https://cds.cern.ch/record/2263802>.
- 2508 [56] CMS Collaboration. Performance of the DeepJet b tagging algorithm using 41.9  
2509  $\text{fb}^{-1}$  of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector.  
2510 2018. URL <https://cds.cern.ch/record/2646773>.
- 2511 [57] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Re-  
2512 construction of the Higgs mass in  $H \rightarrow \tau\tau$  Events by Dynamical Likelihood  
2513 techniques. *Journal of Physics: Conference Series*, 513(2):022035, jun 2014.  
2514 doi: 10.1088/1742-6596/513/2/022035. URL <https://dx.doi.org/10.1088/1742-6596/513/2/022035>.

- 2516 [58] CMS Collaboration. Evidence for the 125 GeV Higgs boson decaying to a pair  
2517 of  $\tau$  leptons. *JHEP*, 05:104, 2014. doi: 10.1007/JHEP05(2014)104.
- 2518 [59] CMS Collaboration. Missing transverse energy performance of the CMS detector.  
2519 *JINST*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.
- 2520 [60] Artur Kalinowski. CMS AN-19-032 (internal): Reconstruction of a  $\tau$  pair invariant  
2521 mass with a simplified likelihood scan, 2019.
- 2522 [61] CMS TAU POG. Tau Physics Object Group: Tau ID Recommendation For Run 2, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2>.
- 2525 [62] CMS MUO POG. Muon Physics Object Group: Recommendations, 2024. URL  
2526 <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>.
- 2527 [63] CMS MUO POG. Muon Physics Object Group: Reference guidelines and results  
2528 for muon momentum scale and resolution in Run II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceScaleResolRun2>.
- 2530 [64] CMS HTT working group. Higgs To Tau Tau Working TWiki for the full Run-  
2531 2 legacy analysis, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/HiggsToTauTauWorkingLegacyRun2>.
- 2533 [65] CMS ELE POG. Electron Physics Object Group: Recommendations,  
2534 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations>.
- 2536 [66] CMS ELE POG. Electron Physics Object Group: Recommendations for  
2537 2016 to 2018 UL, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018>.

- 2539 [67] CMS TAU Embedding Group. Tau embedded samples using 2016  
2540 data, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauTauEmbeddingSamples2016Legacy>.
- 2542 [68] CMS TAU Embedding Group. Tau embedded samples using 2017  
2543 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2017>.
- 2545 [69] CMS TAU Embedding Group. Tau embedded samples using 2018  
2546 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2018>.
- 2548 [70] Tau Lepton Run 2 Trigger Performance. 2019. URL <https://cds.cern.ch/record/2678958>.
- 2550 [71] CMS Taus High Level Trigger Studies. Tau Lepton Run 2 Trigger Performance  
2551 (CMS DP-2019/012), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTauAllRun2>.
- 2553 [72] Muon HLT Performance with 2018 Data. 2018. URL <https://cds.cern.ch/record/2627469>.
- 2555 [73] Single and Double Electron Trigger Efficiencies using the full Run 2 dataset.  
2556 2020. URL <https://cds.cern.ch/record/2888577>.
- 2557 [74] Run II Trigger Performance For  $e\mu$  Triggers. 2019. URL <https://cds.cern.ch/record/2687013>.
- 2559 [75] Performance of electron and photon reconstruction in Run 2 with the CMS ex-  
2560 periment. 2020. URL <https://cds.cern.ch/record/2725004>.

- 2561 [76] CMS Collaboration. Performance of the CMS muon detector and muon recon-  
2562 struction with proton-proton collisions at  $\sqrt{s} = 13$  TeV. *JINST*, 13(06):P06015,  
2563 2018. doi: 10.1088/1748-0221/13/06/P06015.
- 2564 [77] Piet Verwilligen. Muons in the cms high level trigger system. *Nuclear  
2565 and Particle Physics Proceedings*, 273-275:2509–2511, 2016. ISSN 2405-6014.  
2566 doi: <https://doi.org/10.1016/j.nuclphysbps.2015.09.441>. URL <https://www.sciencedirect.com/science/article/pii/S240560141500930X>. 37th Interna-  
2567 tional Conference on High Energy Physics (ICHEP).  
2568
- 2569 [78] Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-  
2570 probe technique. 2020. URL <https://cds.cern.ch/record/2724492>.
- 2571 [79] V.M. Abazov, B. Abbott, M. Abolins, et al. A novel method for modeling the  
2572 recoil in W boson events at hadron colliders. *Nuclear Instruments and Methods  
2573 in Physics Research Section A: Accelerators, Spectrometers, Detectors and Asso-  
2574 ciated 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>.
- 2575 [80] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair  
2576 of light pseudoscalars in the final state with two b quarks and two  $\tau$  leptons in  
2577 proton-proton collisions at  $\sqrt{s} = 13$  TeV. *Phys. Lett. B*, 785:462, 2018. doi:  
2578 10.1016/j.physletb.2018.08.057.
- 2579 [81] CMS LUMI POG. Luminosity Physics Object Group: Recommendations, 2024.  
2580 URL <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.
- 2581 [82] The modeling of the top quark  $p_T$  (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting>.

- 2585 [83] CMS BTV group. Methods to apply b-tagging efficiency scale fac-  
2586 tors (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration>.
- 2588 [84] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in  
2589 pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. doi: 10.1088/1748-0221/  
2590 12/02/P02014.
- 2591 [85] Garvita Agarwal. Jet Energy Scale and Resolution Measurements in CMS. *PoS*,  
2592 ICHEP2022:652, 2022. doi: 10.22323/1.414.0652.
- 2593 [86] CMS JERC group. Jet Energy Corrections (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- 2595 [87] CMS JERC group. Jet Energy Resolution (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- 2597 [88] CMS MUO POG. Muon Physics Object Group: Baseline muon selec-  
2598 tions for Run-II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2>.
- 2600 [89] CMS ELE POG. Electron Identification Based on Simple Cuts, 2024. URL  
2601 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaPublicData>.
- 2602 [90] Jose Enrique Palencia Cortezon. Single top quark production at CMS. Technical  
2603 report, CERN, Geneva, 2018. URL <https://cds.cern.ch/record/2640578>.
- 2604 [91] CMS Collaboration. Measurements of the electroweak diboson production cross  
2605 sections in proton-proton collisions at  $\sqrt{s} = 5.02$  TeV using leptonic decays.  
2606 *Phys. Rev. Lett.*, 127(19):191801, 2021. doi: 10.1103/PhysRevLett.127.191801.
- 2607 [92] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment

- ten years after the discovery. *Nature*, 607(7917):60–68, 2022. doi: 10.1038/s41586-022-04892-x.
- [93] CMS JERC group. Jet energy scale uncertainty sources (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources>.
- [94] TOP Systematic Uncertainties (Run 2) (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics>.
- [95] Kyle Cranmer. Practical Statistics for the LHC. In *2011 European School of High-Energy Physics*, pages 267–308, 2014. doi: 10.5170/CERN-2014-003.267.
- [96] Elham Khazaie, Maryam Zeinali, Hamed Bakhshiansohi, and Abideh Jafari. CMS AN-21-058 (internal): Search for exotic decays of the Higgs boson to a pair of new light bosons with two muons and two b jets in the final states at  $\sqrt{s} = 13$  TeV, 2021.
- [97] CMS Higgs Physics Analysis Group. Summary of 2HDM+S searches at 13 TeV (Run 2), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Summary2HDMSSRun2>.