

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

⁴ KA YU STEPHANIE KWAN

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

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

¹³ MAY 2024

¹⁴

© Copyright by Ka Yu Stephanie Kwan, 2024.

¹⁵

All Rights Reserved

Abstract

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

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

⁴³

Acknowledgements

⁴⁴ Placeholder acknowledgements.

⁴⁶ Contents

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

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

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

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

¹⁴⁵	9.3.5 Modified frequentist method: CL_S	116
¹⁴⁶	10 Results	117
¹⁴⁷	10.1 Results from $bb\tau\tau$	117
¹⁴⁸	10.2 Combination with $bb\mu\mu$ final state	119
¹⁴⁹	11 Asymmetric exotic Higgs decays	130
¹⁵⁰	11.1 Signal masses	130
¹⁵¹	11.2 Cascade scenario signal studies	131
¹⁵²	11.3 Current control plots for $\mu\tau_h$ channel	133
¹⁵³	12 Conclusion and outlook	137
¹⁵⁴	A Samples used	139

¹⁵⁵ List of Tables

¹⁵⁶	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.	52
¹⁶⁰	5.1	Energy scales applied to genuine hadronic tau decays τ_h by data-taking year/era and decay mode, along with systematic errors.	66
¹⁶¹	5.2	Energy scales and systematic errors applied to genuine muons.	67
¹⁶²	5.3	Energy scales and systematic errors applied to electrons in embedded samples by data-taking year/era.	67
¹⁶⁴	5.4	Tau ID efficiency for the DeepTau vs. jet medium working point, with central, up, and down values for 2018, binned in the tau p_T	68
¹⁶⁵	5.5	Energy scales and up/down systematic uncertainties applied to electrons misidentified as hadronic taus.	73
¹⁶⁹	5.6	Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018.	74
¹⁷⁰	5.7	Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018.	74

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

199	A.1 Datasets used in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis for the 2016 era.	140
200	A.2 Datasets used in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis for the 2017 era.	141
201	A.3 Datasets used in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis for the 2018 eras.	141
202	A.4 Embedded samples used in the analysis for the 2016 era.	141
203	A.5 Embedded samples used in the analysis for the 2017 era.	142
204	A.6 Embedded samples used in the analysis for the 2018 era.	142
205	A.7 Background MC samples used in the analysis for the 2016 era. Samples marked with a \dagger are generated with the powhegV2-madspin-pythia8 tag.	143
206		
207	A.8 Background MC samples used in the analysis for the 2017 era. All samples use powheg, except the DYJets and WJets samples, which use madgraphMLM. Samples marked with a \dagger , \ddagger , or $\ddagger\dagger$ were produced with Powheg2 and Pythia8, and Madspin, JHUGenV714, or jhugen724 respectively.	144
208		
209		
210		
211		
212	A.9 Background Monte Carlo samples used in the analysis for the 2018 era. All samples listed are generated for 13 TeV collisions and use pythia8. All samples use powheg, except the DYJets and WJets samples, which use madgraphMLM. Samples marked with a \dagger , \ddagger , or $\ddagger\dagger$, were pro- duced with Powheg and Pythia8, and Madspin, JHUGenV714, and Jhugen724 respectively.	145
213		
214		
215		
216		
217		
218	A.10 Signal samples used in the analysis for the 2016 era. All belong to the RunIISummer16MiniAODv3 campaign and are produced with Mad- graph and Pythia8. The second column is the number of events after the generator-level filter is applied, and the third column is the filter efficiency (percentage of all events that pass the generator-level filter). .	146
219		
220		
221		
222		

223 A.11 Signal samples used in the analysis for the 2017 era. All belong to
224 the RunIIFall17MiniAODv2 campaign and are produced with Mad-
225 graph and Pythia8. The second column is the number of events after
226 the generator-level filter is applied, and the third column is the filter
227 efficiency (percentage of all events that pass the generator-level filter). 146

228 A.12 Signal samples used in the analysis for the 2018 era. All belong to
229 the RunIIAutumn18MiniAOD campaign and are produced with Mad-
230 graph and Pythia8. The second column is the number of events after
231 the generator-level filter is applied, and the third column is the filter
232 efficiency (percentage of all events that pass the generator-level filter). 147

²³³ List of Figures

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

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

297	3.7	Efficiencies of the current and previous emulators of the standalone barrel e/γ algorithm for the Phase-2 Level-1 Trigger, evaluated in a simulated sample containing electrons, as a function of the electron’s generator-level transverse momentum p_T . The standalone working point (WP) is defined as the logical OR of the isolation flag and shower shape flag. The efficiencies with and without requiring the standalone WP, are shown for the current emulator (labeled “Phase 2”, <i>black, red</i>) and the previous emulator (labeled “TDR”, <i>dark blue, grey</i>).	45
305	3.8	Rates in kHz of the current Phase-2 and previous (“TDR”) emulators of the standalone barrel e/γ algorithm for the Phase-2 Level-1 Trigger, evaluated on a minimum bias (MinBias) sample with 200 pile-up (PU), measured as a function of the minimum energy (E_T) required of the reconstructed e/γ object in each event. The standalone working point (standalone WP) is defined to be the logical OR of the isolation flag and the shower shape flag. The rates with and without requiring the standalone WP, are shown for the current emulator (labeled “Phase 2”, <i>black, red</i>) and the previous emulator (labeled “TDR”, <i>dark blue, grey</i>).	47
313	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.	49
319	4.2	Schematic view of the four main steps of the embedding technique for τ leptons.	51
320	5.1	Distributions of $m_{\tau\tau}$ reconstructed by the classic SVFit algorithm, and masses of visible tau decay products (before SVFit).	65
322	5.2	Electron/photon energy scale factors and uncertainties for 2018. . . .	68

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

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

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

395	11.1 Generator-level b-flavor jet transverse momenta p_T , for $h \rightarrow a_1 a_2$ cas-	
396	cade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) =$	
397	$(100, 15)$ GeV (<i>left</i>) and $(40, 20)$ GeV (<i>right</i>). In each plot the generator-	
398	level p_T of the leading (<i>black</i>), sub-leading (<i>red</i>), third (<i>blue</i>), and	
399	fourth (<i>light green</i>) are overlaid.	132
400	11.2 Distributions (arbitrary units) of transverse momentum p_T resolution	
401	and ΔR between the two closest generator-level b jets, treated as one	
402	object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow$	
403	$a_1 a_2$ mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV (<i>top left, top right</i>)	
404	and $(40, 20)$ GeV (<i>bottom left, bottom right</i>) in the ggH production of	
405	the 125 GeV h . In the $(40, 20)$ GeV mass point, the longer p_T resolution	
406	tail (<i>bottom left</i>) indicates that the reconstructed jet underestimates	
407	the generator b-flavor jets' energy, and the significant fraction of events	
408	with larger ΔR values (<i>bottom right</i>) indicate worse matching.	133
409	11.3 Kinematic properties of the leading muon and τ_h in the $\mu\tau_h$ channel: p_T	
410	(<i>top row</i>), η (<i>second row</i>), and ϕ (<i>third row</i>). The visible 4-momenta of	
411	the muon and τ_h are summed, giving the visible di-tau mass m_{vis} and	
412	transverse momentum $p_{T,\text{vis}}$. The errors shown in the figures include	
413	statistical errors and only several of the full set of systematic errors	
414	(only those associated with the lepton energy scales and τ_h identifica-	
415	tion efficiency).	135
416	11.4 Kinematic properties of the leading and sub-leading b-tag jets in the	
417	$\mu\tau_h$ final state: jet p_T (<i>top row</i>), η (<i>second row</i>), ϕ (<i>third row</i>), as well	
418	as the missing transverse energy magnitude and azimuthal direction	
419	(<i>bottom row</i>). The errors shown in the figures include statistical er-	
420	rors and only several of the full set of systematic errors (only those	
421	associated with the lepton energy scales and τ_h identification efficiency).	136

⁴²² Chapter 1

⁴²³ Introduction

⁴²⁴ The Standard Model is the current prevailing theoretical framework that encompasses
⁴²⁵ all known elementary particles to date and describes their interactions, yet falls short
⁴²⁶ of describing open problems in physics. Here, we describe the history of the Standard
⁴²⁷ Model and its particle content (Section 1.1), and provide a mathematical motivation of
⁴²⁸ the SM as a gauge theory (Section 1.2). We introduce the Higgs mechanism (Section
⁴²⁹ 1.3), and outline two groups of theoretical extensions to the Standard Model that
⁴³⁰ feature extended Higgs sectors (Sections 1.4 and 1.5).

⁴³¹ 1.1 History of the Standard Model

⁴³² The building blocks of our modern-day understanding of particle physics were estab-
⁴³³ lished over the course of many decades by experimental discoveries and theoretical
⁴³⁴ advances, culminating in the development of a theoretical framework known as the
⁴³⁵ Standard Model (SM). In the 1880s, the electron was the first subatomic particle to
⁴³⁶ be identified, through measurements of particles produced by ionizing gas. By the
⁴³⁷ 1930s, atoms were known to consist mostly of empty space, with protons and neutrons
⁴³⁸ concentrated at the center and orbited by electrons. Spurred by advances in parti-
⁴³⁹ cle accelerator technology, the experimental discoveries of the positron, the muon,

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

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

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

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

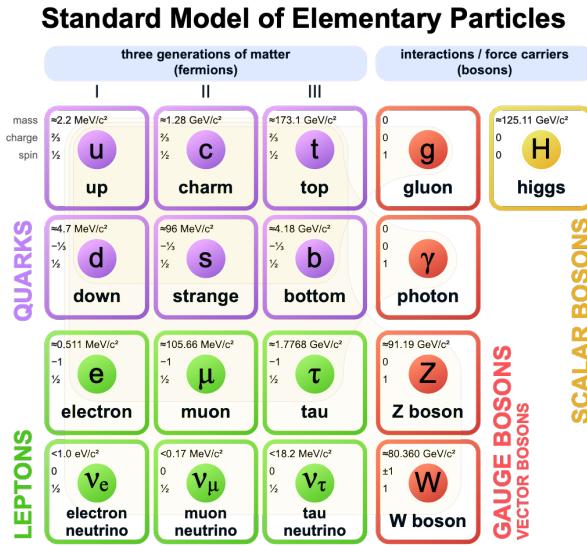


Figure 1.1: Table of Standard Model particles showing the grouping of the fermions into three generations of matter and the bosons, responsible for carrying the three fundamental forces in the Standard Model. The masses, charges, and spins of the particles are shown. The antimatter counterparts of the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

472 1.2 The Standard Model as a gauge theory

473 In this section we lay the theoretical foundations of the Standard Model as a gauge
 474 theory, starting from the principle of gauge invariance (gauge symmetry), with local
 475 gauge symmetries giving rise to interactions between particles.

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

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

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

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

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

488 One important consequence of gauge symmetry comes from the application of
489 Noether's theorem, which states that for every global transformation under which the
490 Lagrangian density is invariant, there exists a conserved quantity. If $\mathcal{L}(\Psi(x), \partial_\mu \Psi(x))$
491 is invariant under the transformation of the wave function $\Psi(x) \rightarrow \Psi'(x)$, where
492 $\Psi'(x) = \Psi(x) + \delta\Psi(x)$, then there exists a conserved current

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

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

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

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

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

503 can be verified to *not* be a solution to the Dirac equation for free particles: $(i\gamma^\mu\partial_\mu -$
 504 $m)\Psi(x) = 0$. This necessitates a modified Dirac equation, where the derivative takes
 505 into account that the vector field $V(x)$ needs to be compared at two displaced space-
 506 time points in a curvilinear coordinate system:

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

507 We define a covariant derivative,

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

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

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

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

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

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

514 One must define a derivative of the form

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

515 to keep the theory invariant under Eqn. 1.7. The four-potential B^μ represents the
516 interacting four-potential which must be added to keep the theory invariant.

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

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

521 • In the $U(1)_Y$ term, B_μ is the weak hypercharge field.

522 • 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-
523 potentials. $\tau/2$ are the Pauli matrices, generators of the $SU(2)$ transformation.

524 • In the $SU(3)_C$ term, the gluon (color) field is G_μ . λ_k are the Gell-Man matrices,
525 generators of the $SU(3)$ transformation.

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

528 1.3 The Higgs mechanism

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

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

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

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

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

548 where ϕ_1^+ , ϕ_2^+ , ϕ_1^0 , and ϕ_2^0 are real (four degrees of freedom). By convention, the
 549 nonzero vacuum expectation value is assigned to ϕ_1^0 .

550 The minimal self-interacting Higgs potential that is invariant under $SU(2)_L \times$
 551 $U(1)_Y$ is given by

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

552 where λ is the coupling strength of the four-point Higgs interaction. The potential

553 energy is minimized at

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

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

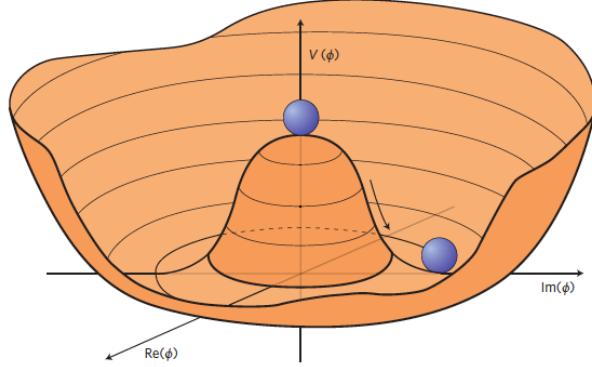


Figure 1.2: An illustration of the Higgs potential [5]. Choosing any of the points at the bottom of the potential breaks spontaneously the rotational $U(1)$ symmetry.

556 The excitations of the Higgs field with respect to the minimum Φ_{\min} are parame-
 557 terized by

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

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

562 1.4 Two-Higgs Doublet Models

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

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

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

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

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

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

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

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

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

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

578 Higgs, while the gauge boson couplings are very close to Standard Model-like for
579 $\tan \beta \gtrsim 5$. All of the properties of h can be determined by just two parameters: $\tan \beta$
580 and α , and the fermion couplings to the two Higgs doublets.

581 2HDM can be extended by a scalar singlet (2HDM+S) [6]:

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

582 If this singlet only couples to the Higgs doublets $H_{1,2}$ and has no direct Yukawa
583 couplings, all of its couplings to SM fermions result from mixing with $H_{1,2}$. Under
584 these simple assumptions, exotic Higgs decays $h \rightarrow ss \rightarrow X\bar{X}Y\bar{Y}$ or $h \rightarrow aa \rightarrow$
585 $X\bar{X}Y\bar{Y}$, and $h \rightarrow aZ \rightarrow X\bar{X}Y\bar{Y}$ are permitted, where $s(a)$ is a (pseudo)scalar mass
586 eigenstate mostly composed of $S_R(S_I)$, and X, Y are Standard Model fermions or
587 gauge bosons. There are two pseudoscalars in the 2HDM+S, and the mostly singlet-
588 like pseudoscalar can be chosen to be the one lighter than the SM-like Higgs. For
589 $m_a < m_h - m_Z \sim 35$ GeV, the exotic Higgs decay $h \rightarrow Za$ is possible, and for
590 $m_a < m_h/2 \approx 63$ GeV, the exotic Higgs decay $h \rightarrow aa$ is possible.

591 In 2HDM, and by extension 2HDM+S, there are four types of fermion couplings
592 commonly discussed in the literature that forbid flavor-changing neutral currents at
593 tree level [6]. These are referred to as Type I (all fermions couple to H_2), Type II
594 (MSSM-like, d_R and e_R couple to H_1 , u_R to H_2), Type III (lepton-specific, leptons
595 and quarks couple to H_1 and H_2 respectively) and Type IV (flipped, with u_R , e_R
596 coupling to H_2 and d_R to H_1). The exact branching ratios of the pseudoscalars to
597 Standard Model particles vary depending on the 2HDM+S model and the value of
598 $\tan \beta$ (e.g. Fig. 1.3).

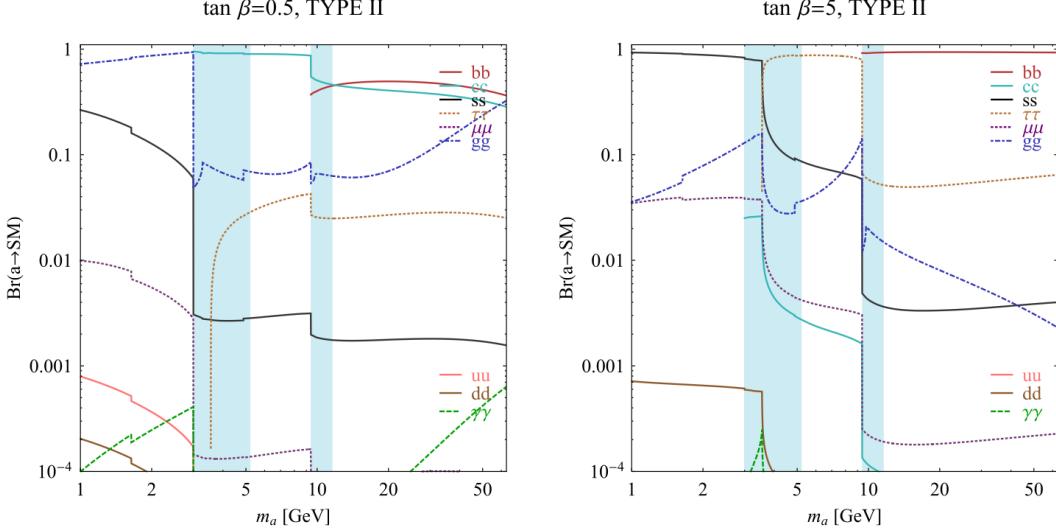


FIG. 7 (color online). Branching ratios of a singletlike pseudoscalar in the 2HDM + S for type-II Yukawa couplings. Decays to quarkonia likely invalidate our simple calculations in the shaded regions.

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

1.5 Two Real Singlet Model

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

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

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

609 The mass eigenstates $h_{1,2,3}$ are related to the fields $\phi_{h,S,X}$ through a 3×3 orthogonal
610 mixing matrix denoted R . The mass eigenstates are assumed to be ordered $M_1 \leq$
611 $M_2 \leq M_3$. R is parameterized by the three mixing angles θ_{hS} , θ_{hX} , θ_{SX} . The nine
612 parameters of the scalar potential can be expressed in terms of the three physical
613 Higgs masses, the three mixing angles, and the three vacuum expectation values.

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

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

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

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

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

₆₃₆ (*bottom left*), while at smaller masses, combinations with τ leptons and eventually
₆₃₇ final states with charm quarks and muons become relevant [7].

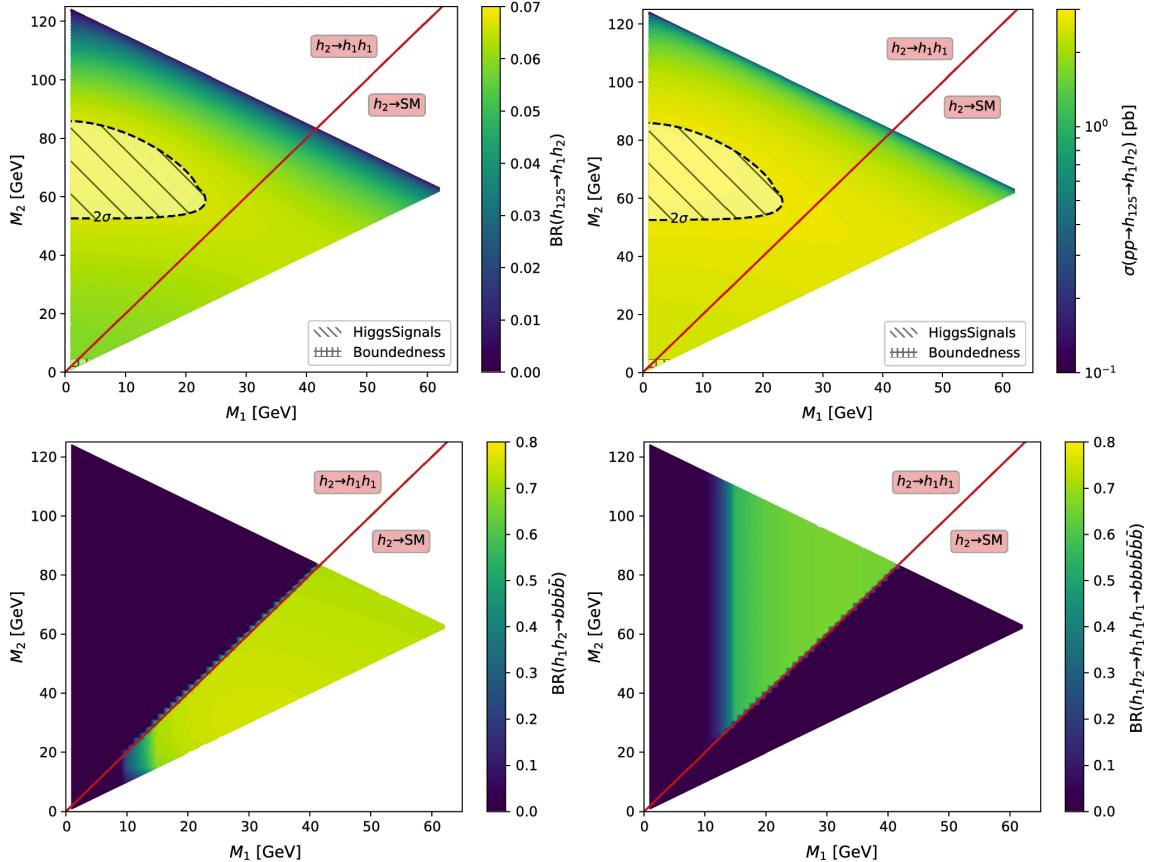


Figure 1.4: Benchmark plane BP1 for benchmark scenario 1 from [7], for the decay signature $h_{125} \rightarrow h_1 h_2$ with $h_{125} \equiv h_3$, defined in the (M_1, M_2) plane. The color code shows $\text{BR}(h_3 \rightarrow h_1 h_2)$ (*top left*) and the 13 TeV LHC signal rate for $pp \rightarrow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

638 **Chapter 2**

639 **The Large Hadron Collider and the**
640 **CMS Experiment**

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

653 2.1 The Large Hadron Collider

654 CERN, the European Organization for Nuclear Research, is an international organiza-
655 tion based in Meyrin, Switzerland which operates the world's largest particle physics
656 laboratory, and is the site of the Large Hadron Collider (LHC) [8]. The very first
657 accelerator built at CERN was the 600 MeV Synchrocyclotron (SC), which initially
658 provided beams for CERN's first experiments. The newer and more powerful Proton
659 Synchrotron (PS), which could accelerate particles to an energy of 28 GeV, began op-
660 erations in 1959 and is still in use today. The first hadron collider at CERN was the
661 Intersecting Storage Rings (ISR), which consisted of two interlaced rings each with a
662 diameter of 200. The ISR collided protons at a center-of-mass energy of 62 GeV and
663 began measuring collisions in 1971. In 1968 CERN began to accelerate heavy ions
664 in the Super Proton Synchrotron (SPS), which is 7 kilometers in circumference and
665 was the first of CERN's giant underground rings to be built. The SPS became the
666 forefront of CERN's particle physics program in 1976, and in 1981 was converted into
667 a proton-antiproton collider. The final and largest underground ring constructed at
668 CERN was the Large Electron-Positron (LEP) collider, which was commissioned in
669 July 1989 and hosted 5176 magnets and 128 accelerating cavities located around a
670 27-kilometer circumference. Over 11 years of research, four detectors, ALEPH, DEL-
671 PHI, L3, and OPAL measured the collisions, with collision energies reaching up to
672 209 GeV in the year 2000. In November 2000, LEP was closed down to make way for
673 the construction of the LHC in the same tunnel.

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

680 the Proton Synchrotron (PS), and 450 GeV in the Super Proton Synchrotron (SPS).
681 The protons are transferred to the two beam pipes of the Large Hadron Collider
682 (LHC). The LHC is a 27-kilometer ring of superconducting magnets, inside which
683 one beam circulates clockwise and the other counterclockwise. Each LHC ring takes
684 4 minutes and 20 seconds to fill, and it takes about 20 minutes for the protons to
685 reach their maximum energy. During normal operating conditions, beams circulate
686 for many hours inside the LHC ring.

687 The beams of particles in the LHC are made to collide at a center-of-mass energy
688 of up to 14 TeV, at four positions at particle detector experiments located around
689 the ring: ATLAS, CMS, ALICE, and LHCb. An aerial view of the four major
690 experiments' locations is shown in Fig. 2.1 [11]. ATLAS and CMS are the two
691 general-purpose detectors with broad physics programmes spanning Standard Model
692 measurements and searches for signatures of new physics [12] [13]. The two experi-
693 ments use different technical solutions and different magnet system designs. ALICE
694 is a general-purpose detector dedicated to measuring LHC heavy-ion collisions, and
695 is designed to address the physics of strongly interacting matter, and the properties
696 of quark-gluon plasma [14]. The LHCb experiment specializes in investigating CP vi-
697 olation through measuring the differences in matter and antimatter, by using a series
698 of subdetectors to detect mainly forward particles close to the beam direction [15].

699 2.2 Luminosity and pile-up

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

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

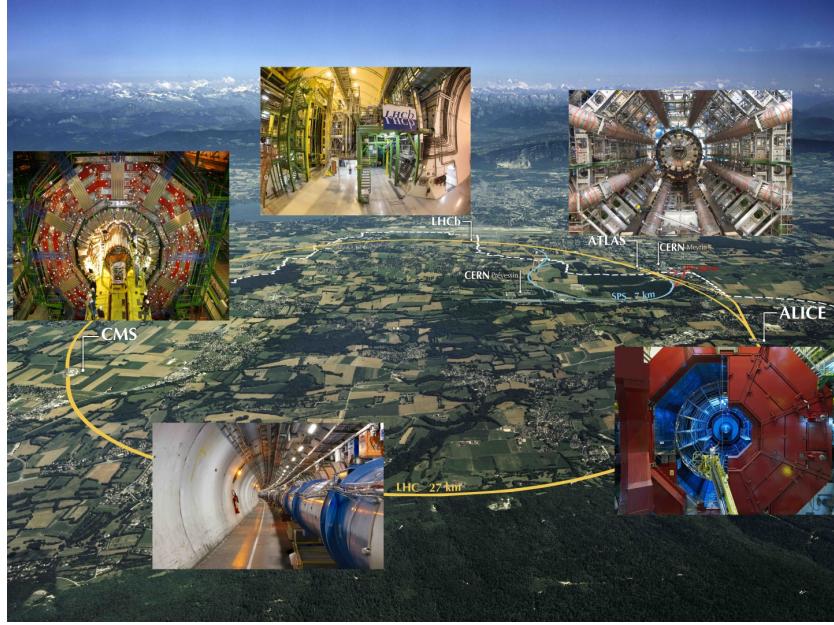


Figure 2.1: Aerial view of the Large Hadron Collider (LHC) spanning the border of France and Switzerland, and the four major experiments located around the ring: CMS (Compact Muon Solenoid), LHCb (LHC beauty), ATLAS (A Toroidal LHC Apparatus), and ALICE (A Large Ion Collider Experiment) [11].

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

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

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

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

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

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

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

The interaction vertex corresponding to the hard scattering of the protons is called the primary interaction vertex, or primary vertex (PV). The LHC's nominal beam luminosities are sufficiently large for multiple proton-proton collisions to occur in the same time window of 25 nanoseconds in which proton bunches collide [18]. To measure a proton-proton collision, the primary vertices must be separated from overlapping collisions, called “pile-up” collisions.

The pile-up is defined as the average number of pp collisions per bunch crossing, and can be estimated from the inelastic pp cross section of $\sigma_{\text{inel}} = 68.6$ millibarns at a center-of-mass energy of $\sqrt{s} = 13$ TeV [19]:

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

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

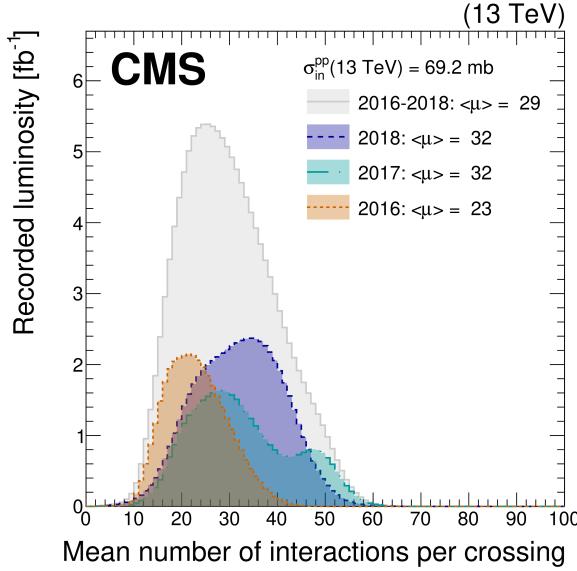


Figure 2.2: Distribution of the mean number of inelastic collisions per bunch crossing (pile-up) in data [18], for proton-proton collisions in 2016 (*dotted orange*), 2017 (*dotted light blue*), 2018 (*dotted dark blue*), and integrated over 2016-2018 (*solid grey*). A cross-section of inelastic proton-proton collisions of 69.2 mbarns is assumed. In the running conditions of the High-Luminosity LHC, pile-up will reach unprecedented levels of up to 200 per bunch crossing [20].

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

2.3 The High-Luminosity LHC

The High-Luminosity LHC (HL-LHC) is a major upgrade of the LHC scheduled to take place in the late 2020s, that will increase the instantaneous luminosity by a factor of five beyond the original design value, and the integrated luminosity by a factor of ten [20]. This will be accomplished through accelerator technological advances: for instance, reduction of the interaction point β^* from 0.55 m down to 0.15

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

751 2.4 The CMS detector

752 We give a brief overview of the Compact Muon Solenoid (CMS) experiment here
753 and discuss each of the subdetectors in more detail in the following sections. The
754 CMS experiment was conceived to study proton-proton and lead-lead collisions at
755 a center-of-mass energy of 14 TeV (5.5 TeV nucleon-nucleon) and at luminosities up
756 to $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ($10^{27} \text{ cm}^{-2} \text{ s}^{-1}$) [21] [22]. Starting from the beam interaction region
757 at the center of the CMS detector, particles first pass through a silicon pixel and
758 strip tracker, in which charged-particle trajectories (tracks) and origins (vertices)
759 are reconstructed from signals (hits) in the sensitive layers. The tracker, electro-
760 magnetic calorimeter (ECAL), and hadronic calorimeter (HCAL) are immersed in a
761 high-magnetic-field superconducting solenoid that bends the trajectories of charged
762 particles. After passing through the tracker, electrons and photons are then absorbed
763 in the electromagnetic calorimeter (ECAL) comprised of lead-tungstate scintillating-
764 crystals. The corresponding electromagnetic showers are detected as clusters of energy
765 recording in neighboring cells, from which the direction and energy of the particles can
766 be determined. Charged and neutral hadrons may initiate a hadronic shower in the
767 ECAL as well, which is then fully absorbed in the hadron calorimeter (HCAL). The
768 resulting clusters are used to estimate their direction and energies. Muons and neu-
769 trinos pass through the calorimeters with little to no interactions. Neutrinos escaped
770 undetected; muons produce hits in additional gas-ionization chamber muon detectors
771 housed in the iron yoke of the flux-return. A sketch of example particle interactions

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

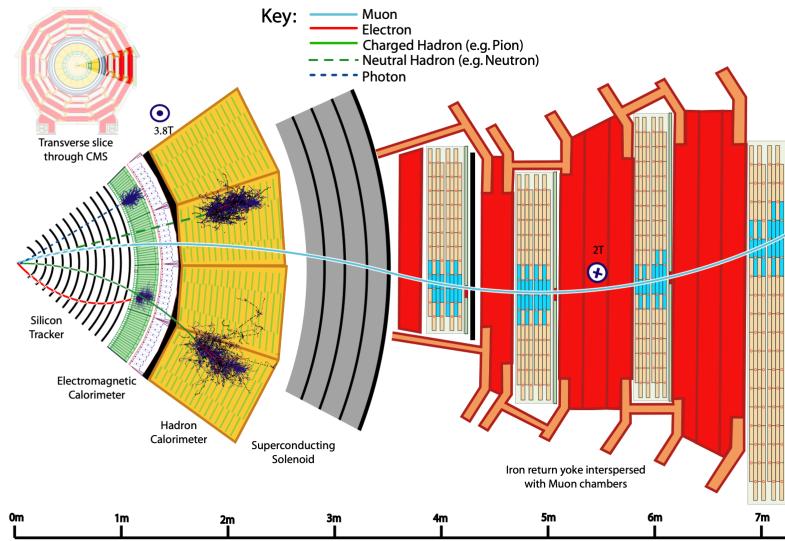


Figure 2.3: Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector [22].

CMS uses a right-handed coordinate system [21]. The origin is centered at the nominal collision point inside the experiment. The x axis points towards the center of the LHC, and the y axis points vertically upwards. The z axis points along the beam direction. The azimuthal angle, ϕ , is measured from the x axis in the x - y plane, and the radial coordinate in this plane is denoted by r . The polar angle, θ , is measured from the z axis. The pseudorapidity, η , is defined as $\eta = -\ln \tan(\theta/2)$. The momentum and energy transverse to the beam direction, denoted by p_T and E_T respectively, are computed from the x and y components. The momentum imbalance in the transverse plane is called the missing transverse momentum, and its magnitude is denoted by E_T^{miss} .

786 2.5 Sub-detectors of CMS and data processing

787 This section details the sub-detectors of CMS that operate to identify and precisely
788 measure muons, electrons, photons, and jets over a large energy range. The sections
789 are ordered starting from the innermost sub-detectors closest to the beam interaction
790 area: the tracker, the electromagnetic and hadronic calorimeters, and the muon de-
791 tectors. The two-stage trigger system is described, starting with the firmware-based
792 Level-1 Trigger and followed by the High-Level Trigger. Lastly, particle reconstruction
793 and data storage and computational infrastructure are discussed.

794 2.5.1 Inner tracking system

795 The CMS Tracker performs robust tracking and detailed vertex reconstruction in the
796 4 T magnetic field of the superconducting solenoidal magnet. The primary sensors
797 used in the tracker are p^+ on n -bulk devices, which allow high voltage operation and
798 are radiation-resistant [23] [24]. The active envelope of the CMS Tracker extends to a
799 radius of 115 cm, over a length of approximately 270 cm on each side of the interaction
800 point [23]. Charged particles in the region $|\eta| \lesssim 1.6$ benefit from the full momentum
801 measurement precision. In this region, a charged particle with p_T of 1000 GeV has a
802 sagitta of $\sim 195 \mu\text{m}$. The Tracker acceptance extends further to $|\eta| = 2.5$, with a
803 reduced radius of approximately 50 cm.

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

808 A schematic view of the current Phase-1 CMS tracker [25], including the pixel
809 detector, is shown in Fig. 2.4. The Phase-1 pixel detector consists of three barrel
810 layers (BPIX) at radii of 4.4 cm, 7.3 cm, and 10.2 cm, and two forward/backward disks

811 (FPIX) at longitudinal positions of ± 34.5 cm and ± 46.5 cm, and extending in radius
 812 from about 6 cm to 15 cm. These pixelated detectors produce 3D measurements along
 813 the paths of charged particles with single hit resolutions between 10-20 μm .

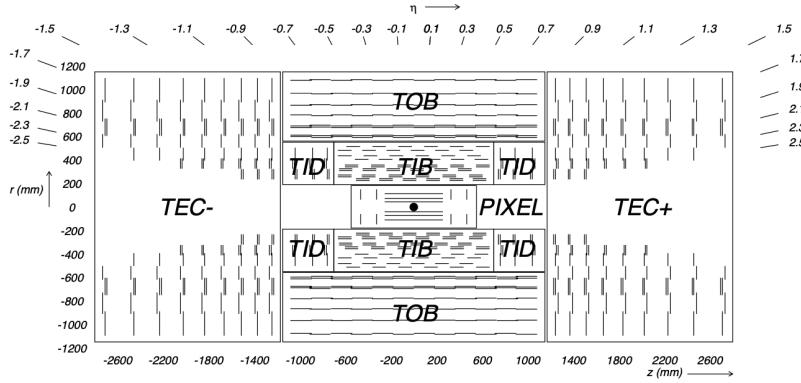


Figure 2.4: Cross section of the current Phase-1 CMS tracker [25]. Each line represents a detector module. Double lines indicate back-to-back modules which deliver two-dimensional (stereo) hits in the strip tracker.

814 After the pixel and on their way out of the tracker, particles pass through the
 815 silicon strip tracker which reaches out to a radius of 130 cm (Fig. 2.4). The sensor el-
 816 ements in the strip tracker are single-sided p -on- n type silicon micro-strip sensors [21].
 817 The silicon strip detector consists of four inner barrel (TIB) layers assembled in shells,
 818 with two inner endcaps (TID), each composed of three small discs. The outer barrel
 819 (TOB) consists of six concentric layers. Two endcaps (TEC) close off the tracker on
 820 either end.

821 2.5.2 ECAL

822 The electromagnetic calorimeter (ECAL) of CMS measures electromagnetic energy
 823 deposits with high granularity. One of the driving criteria in the design was the capa-
 824 bility of detecting the Standard Model Higgs boson decay to two photons (in fact, the
 825 channel in which the 125 GeV Higgs boson was discovered at CMS). ECAL is a her-
 826 metic homogeneous calorimeter comprised of 61,200 lead tungstate (PbWO_4) crystals

827 mounted in the central barrel, with 7,324 crystals in each of the two endcaps [21]. A
828 preshower detector is located in front of the endcap crystals. Avalanche photodiodes
829 (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in
830 the endcaps.

831 The design of the ECAL is driven by the behaviour of high-energy electrons, which
832 predominantly lose energy in matter via bremsstrahlung, and high-energy photons
833 by e^+e^- pair production. The characteristic amount of matter traversed for these
834 interactions is the radiation length X^0 , usually measured in units of $\text{g} \cdot \text{cm}^{-2}$. The
835 radiation length is also the mean distance over which a high-energy electron loses all
836 but $1/e$ of its energy via bremsstrahlung [26]. Thus high granularity in η and ϕ , and
837 the length of the ECAL crystals, is designed to capture the shower of e/γ produced
838 by electrons and photons.

839 The barrel part of the ECAL (EB) covers the pseudorapidity range $|\eta| < 1.479$ [21].
840 The barrel granularity is 360-fold in ϕ and (2×85) -fold in η . The crystal cross-section
841 corresponds to approximately 0.0174×0.0174 in $\eta - \phi$ or $22 \times 22 \text{ mm}^2$ at the front
842 face of the crystal, and $26 \times 26 \text{ mm}^2$ at the rear face. The crystal length is 230 mm,
843 corresponding to $25.8 X_0$.

844 The ECAL read-out acquires the signals of the photodetectors [21]. At each bunch
845 crossing, digital sums representing the energy deposit in a trigger tower, comprising
846 5×5 crystals in $\eta \times \phi$, are generated and sent to the Level-1 trigger system (detailed
847 in Section 2.5.5).

848 2.5.3 HCAL

849 The hadronic calorimeter (HCAL) of CMS measures hadronic energy, which is key to
850 characterizing the presence of apparent missing transverse energy which could arise
851 from hadron jets and neutrinos or exotic particles [21]. A schematic of the components
852 of HCAL are shown in Fig. 2.5. The HCAL barrel (HB) and endcaps (HE) are located

853 outside of the tracker and the ECAL, spanning a radius of 1.77 m (outer extent of
 854 ECAL) up to 2.95 m (inner extent of the magnet coil). An outer hadron calorimeter
 855 (HO) is placed outside the solenoid to complement the barrel calorimeter. Beyond
 856 $|\eta| = 3$, the forward hadron calorimeter (HF) at 11.2 m from the interaction point
 857 extend the pseudorapidity coverage to $|\eta| = 5.2$.

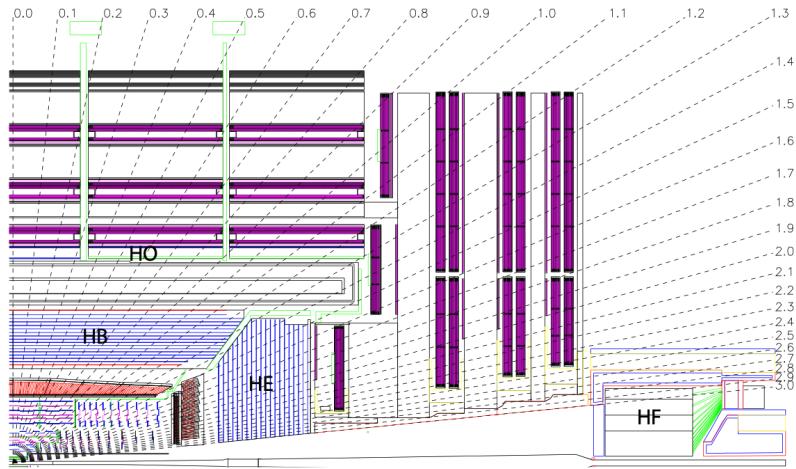


Figure 2.5: Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters from [21].

858 The HB is a sampling calorimeter covering the pseudorapidity range $|\eta| < 1.3$ [21].
 859 It consists of 36 identical azimuthal wedges which form two half-barrels (HB+ and HB-
 860), with a segmentation of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The HE covers pseudorapidity
 861 $1.3 < |\eta| < 3$. The HB and endcap HE calorimeters are sampling calorimeters which
 862 use brass as the absorber and plastic scintillator as the active material. Light from
 863 the plastic scintillator is wavelength-shifted and captured in optic fibers which are
 864 read out by front-end electronics [27].

865 The HF is a Cherenkov calorimeter based on a steel absorber and quartz fibers
 866 which run longitudinally through the absorber and collect Cherenkov light, primarily
 867 from the electromagnetic component of showers developed in the calorimeter [27].
 868 Photomultiplier tubes are used to collect light from the quartz fibers. The HF is
 869 designed to survive in the harsh radiation conditions and high particle flux of the for-

ward region. On average, 760 GeV per proton-proton interaction is deposited into the two forward calorimeters, compared to only 100 GeV for the rest of the detector [21]. Furthermore, this energy has a pronounced maximum at the highest rapidities.

2.5.4 Muon detectors

The CMS muon system is designed to have the capability of reconstructing the momentum and charge of muons over the kinematic range of the LHC, since muons are a powerful handle on signatures of interesting processes over the high background rate of the LHC [21]. For instance, the decay of the Standard Model Higgs boson into ZZ , which in turn decay to 4 leptons, can be reconstructed with high 4-particle mass resolution if all the leptons are muons, since muons are less affected than electrons by radiative losses in the tracker material.

The muon system consists of a cylindrical barrel section and two planar endcap regions [21]. The barrel muon detector consists of drift tube (DT) chambers covering the pseudorapidity region $|\eta| < 1.2$ (Fig. 2.6). The DTs can be used as tracking detectors due to the barrel region’s characteristic low neutron-induced backgrounds, low muon rate, and relatively uniform 4T magnetic field contained in the steel yoke.

In the two endcap regions, the muon rates and background levels are high and the magnetic field is large and non-uniform [21]. Here, the muon system uses cathode strip chambers (CSCs) to identify muons between $0.9 < |\eta| < 2.4$. The cathode strips of each chamber run radially outwards and provide a precision measurement in the $r - \phi$ bending plane. The anode wires run approximately perpendicular to the strips and are read out in order to measure η and the beam-crossing time of a muon.

In addition to the DT and CSC, a dedicated trigger system consisting of resistive plate chambers (RPCs) in the barrel and endcap regions provide a fast, independent, and highly-segmented trigger with a sharp p_T threshold over a large portion of the pseudorapidity range ($|\eta| < 1.6$) of the muon system [21]. RPCs have good time

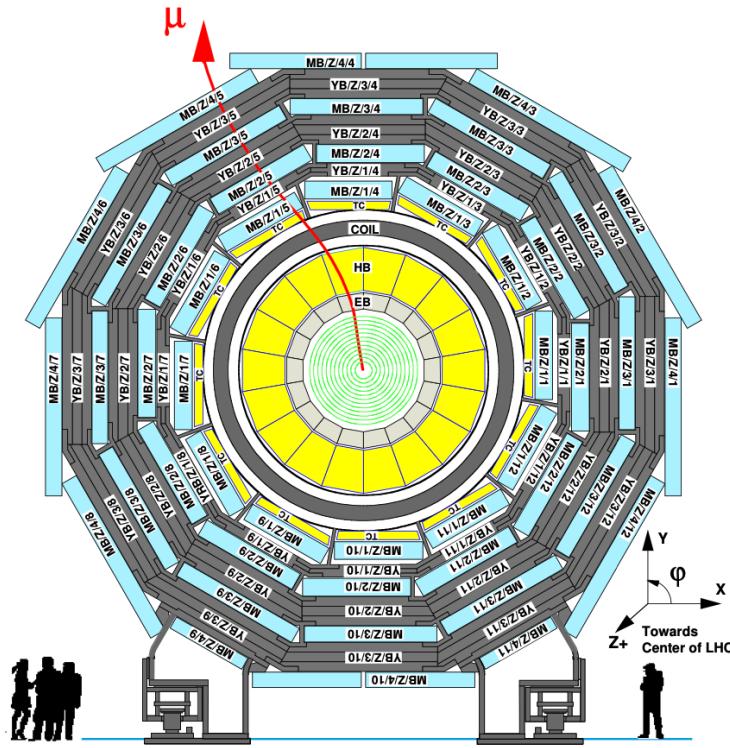


Figure 2.6: Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels from [21]. The DTs are organized in 12 sectors of the yoke barrel (YB). In each of the 12 sectors of the yoke, there are 4 muon chambers per wheel (MB1, MB2, MB3, and MB4).

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

899 2.5.5 The Level-1 Trigger

The design performance of the LHC corresponds to an instantaneous luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a 25 ns bunch crossing rate, giving an average pile-up (number of simultaneous events) of 25 per bunch crossing [28]. However, during Run 2, in 2017 and 2018 the LHC was able to surpass this goal with a mean number of 32 interactions per bunch crossing, and reaching over 50 interactions in short periods (Fig. 904)

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

The L1 Trigger data flow of Phase-1 is shown in Fig. 2.7 [28], with organization into the L1 calorimeter trigger, the L1 muon trigger, and the L1 Global Trigger (GT).

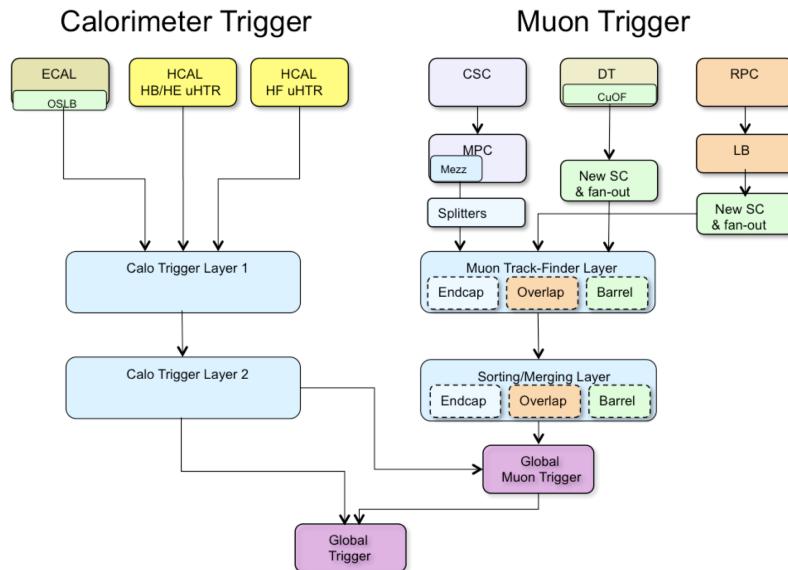


Figure 2.7: Dataflow for the Phase-1 Level-1 Trigger [28], which is implemented in custom hardware and is responsible for reducing the event rate from the LHC bunch crossing frequency of 400 MHz (bunch crossings every 25 ns) to a maximum rate of 100 kHz. In Phase-1, the Level-1 Trigger has access to information from the calorimeter and muon detectors.

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

918 individual calorimeter cell energies. In the original configuration, the ECAL energies
919 were accompanied by a bit indicating the transverse extent of the electromagnetic
920 energy deposits, and the HCAL energies were accompanied by a bit indicating the
921 presence of minimum ionizing energy [29]. During Long Shutdowns 1 and 2 (LS1
922 and LS2), HF was upgraded to provide finer granularity information to the trigger,
923 and the HCAL barrel and endcap front-end electronics were upgraded to provide
924 high-precision timing information and depth segmentation information.

925 In the original design of the L1 calorimeter trigger, the trigger primitives are pro-
926 cessed by the Regional Calorimeter Trigger (RCT, upgraded to Calo Layer 1 after
927 LS2) which finds isolated and non-isolated electron/photon candidates [28]. At this
928 stage, electrons/photons candidates are treated together since they cannot be defini-
929 tively distinguished at this stage due to lack of tracking information in the L1 trigger.
930 The Global Calorimeter Trigger (GCT, upgraded to Calo Layer 2 after LS2) sorts
931 further the candidate electrons/photons, finds jets (classified as central, forward, and
932 tau) using the E_T sums and performs calibration of the clustered jet energies, and
933 calculates global quantities such as missing E_T . It sends the top four candidates of
934 each type to the Global Trigger [28].

935 During LS2 and before Run-2, the legacy calorimeter trigger was upgraded to be
936 more flexible, maintainable, and performant [30] [31] [32]. These upgrades included
937 the replacement of legacy VME-based electronics with the MicroTCA (μ TCA) mod-
938 ern telecommunication standard, and system-wide usage of the latest generation of
939 FPGAs, Xilinx Virtex 7. Parallel copper links were replaced in almost all cases with
940 serial optical links, allowing link speeds to increase from 1 Gb/s to 10 Gb/s [30]. A
941 schematic of the current calorimeter trigger is shown in Fig. 2.8. The calorimeter
942 Layer-1 is implemented in 18 Calorimeter Trigger Processor (CTP7) boards, with
943 each card spanning 4 out of 72 towers in ϕ and all of η . Tower-level operations are
944 performed in Layer-1, such as the sum of ECAL and HCAL energies, energy cali-

bration, and the computation of the ratio of HCAL to ECAL energies. The Layer-1 cards each transmit 48 output links at 10 Gb/s to the nine Layer-2 Master Processor cards (MP7) cards, which host calorimeter algorithms that find particle candidates and compute global energy sums. Each MP7 takes 72 input links and has access to the whole event at trigger tower granularity, such that the algorithms are fully pipelined and start processing as soon as the minimum amount of data is received. The trigger candidates are sent to a demultiplexer (demux) board, also a MP7, which formats the data for the upgraded Global Trigger, also called the microGT (μ GT).

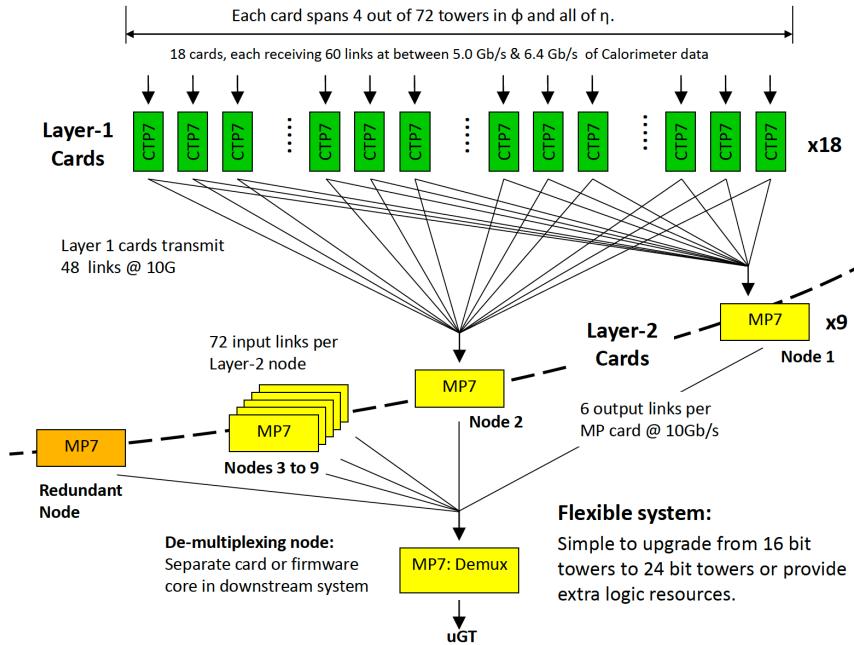


Figure 2.8: Schematic of the calorimeter trigger after Long Shutdown 2 [30]. The Layer-1 calorimeter trigger is implemented in CTP7 cards, which send time-multiplexed outputs to the Layer-2 MP7 cards. The Layer-2 cards handle the data in a round-robin style and the outputs are de-multiplexed, producing one output data stream to the Global Trigger.

Each of the L1 muon triggers has its own trigger logic [29]. 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

957 the same CSC. The CSCs form local charged tracks (LCTs) from the cathode strips,
958 which are combined with the anode wire information. LCTs are combined into full
959 muon tracks and assigned p_T values.

960 The Global Muon Trigger (GMT) sorts the RPC, DT, and CSC muon tracks,
961 converts these tracks to the same η , ϕ , and p_T scale, and validates the muon sign [29].
962 It improves the trigger efficiency by merging muon candidates that were detected
963 in two complementary sub-systems (i.e. DT+RPC, or CSC+RPC). The GMT also
964 contains logic to correlate the found muon tracks with an $\eta-\phi$ grid of quiet calorimeter
965 towers to determine if the muons are isolated, as well as logic to remove duplicate
966 candidates originating in the overlap regions from both DT and CSC systems. The
967 final collection of muons are sorted based on their initial quality, correlation, and p_T ,
968 and the top four muons are sent to the Global Trigger [29].

969 The Global Trigger (GT) receives information from the GCT and GMT, and
970 makes the Level-1 Accept (L1A) decision to either discard or accept the bunch cross-
971 ing [29]. This is accomplished by sorting ranked trigger objects that are accompanied
972 by positional information in η and ϕ , permitting the trigger to applying criteria with
973 thresholds that can vary based on the location of the trigger objects, and/or to re-
974 quire trigger objects to be close to or opposite from each other. The GT L1A decision
975 arrives at the detector front end with a $3.8\ \mu\text{s}$ latency after the interaction at a rate
976 which is required to be less than 100 kHz, and triggers a full readout of the detector
977 for further processing.

978 2.5.6 The High-Level Trigger

979 The HLT is implemented in software running on a large computer farm of fast com-
980 mercial processors [33] [34]. The algorithms in HLT have access to full data from
981 all CMS sub-detectors, including the tracker, with full granularity and resolution.
982 The HLT reconstruction software is similar to what is used offline for full CMS data

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

The HLT contains trigger paths, each corresponding to a dedicated trigger [35]. A path consists of several steps implemented as software modules. Each HLT trigger path must be seeded by one or more L1 trigger bits: the first module always looks for a L1 seed, consisting of L1 bit(s) and L1 object(s). Each module performs a well-defined task such as unpacking (raw to digitized quantities), reconstruction of physics objects (electrons, muons, jet, missing transverse energy, etc.), making intermediate decisions that trigger more detailed reconstruction modules, and calculating the final decision for the trigger path. If an intermediate filter decision is negative, the rest of the path is not executed, and the trigger rejects the event.

2.5.7 Particle reconstruction

To build a description of the physics objects present in the particle collision, the basic elements from the detector layers (tracks and clusters of energy) are correlated to identify each particle in the final state. Measurements from different sub-detectors are combined to reconstruct the particle properties. This approach is called particle-flow (PF) reconstruction [22]. Key to the success of the PF reconstruction is the fine spatial granularity of the detector layers. Coarse-grained detectors can cause the signals from different particles to merge, especially within jets. However, if the subdetectors are sufficiently segmented to separate individual particles, it becomes possible to produce a global event description that identifies all physics objects with high efficiencies and resolution.

1009 **2.5.8 Data storage and computational infrastructure**

1010 The LHC generates over 15 petabytes (15 million gigabytes) of data every year, neces-
1011 sitating a flexible computing system that can be accessed by researchers working at
1012 the four main LHC experiments: ALICE, ATLAS, CMS, and LHCb. The Worldwide
1013 LHC Computing Grid (WLCG) [36] is a global collaboration of computer centers that
1014 links thousands of computers and storage systems in over 170 centers across 41 coun-
1015 tries. These centers are arranged in “tiers”, and provide near real-time access to users
1016 processing, analyzing, and storing LHC data. One of the final stages of data analy-
1017 sis at LHC experiments is large-scale data processing taking place over distributing
1018 computing, for instance, with the use of Condor [37], a distributed, scalable, flexible
1019 batch processing system which accepts a computing job, allocates a resource to it,
1020 executes it, and returns the result back to a user transparently.

1021 **Chapter 3**

1022 **The Phase-2 Upgrade of CMS**

1023 This chapter gives an overview of the High-Luminosity LHC upgrade of the LHC in
1024 Section 3.1, and the upgrades for the Phase-2 CMS Level-1 (L1) Trigger in Section
1025 3.2. One of the major upgrades is the new availability of calorimeter crystal-level
1026 information to the L1 calorimeter trigger, compared to the current trigger which only
1027 has access to tower-level information (a tower being 5 by 5 in crystals). To capitalize
1028 on the increased spatial granularity of this information, an upgraded algorithm is
1029 presented which reconstructs and identifies electron and photon candidates in the the
1030 Layer-1 Calorimeter Trigger. A description of the algorithm and a validation of its
1031 performance in Phase-2 conditions is given in Section 3.3.

1032 **3.1 The High-Luminosity LHC**

1033 In order to sustain and extend the LHC’s physics discovery program and maintain
1034 operability for a decade or more, the LHC is undergoing a major upgrade to the High-
1035 Luminosity LHC (HL-LHC). In its final configuration, the HL-LHC will deliver a peak
1036 luminosity of $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, potentially leading to total integrated luminosity
1037 of 4000 fb^{-1} after ten years of operations, scheduled to begin in 2027 [38]. This
1038 integrated luminosity is about ten times the predicted luminosity reach of the LHC

1039 in its initial configuration. To enable the CMS experiment to continue operations and
1040 data-taking and to maximize the discovery potential of the unprecedented amount
1041 of data, the CMS detector is undergoing Phase-2 upgrades in order to perform high-
1042 precision measurements and searches for physics beyond the Standard Model in the
1043 intense running conditions of the HL-LHC.

1044 3.2 The Phase-2 Level-1 Trigger

1045 To achieve the goals of the HL-LHC program and to ensure the collection of information-
1046 rich datasets in the HL-LHC, the Phase-2 upgrade of the CMS Level-1 Trigger [38]
1047 must be upgraded in conjunction with the CMS sub-detectors and their readouts, to
1048 maintain physics selectivity. The HL-LHC will produce an intense hadronic environ-
1049 ment corresponding to 200 simultaneous collisions per beam crossing, necessitating
1050 comprehensive upgrades of the trigger system outlined below.

1051 In order to cope with the increased pile-up and high occupancies of the HL-LHC,
1052 the latency of the L1 trigger system (time available to produce a L1 Accept signal) will
1053 be increased significantly from $3.8 \mu\text{s}$ to $12.5 \mu\text{s}$, with an increased maximum output
1054 bandwidth of 750 kHz [38]. With the increased latency, in addition to information
1055 from calorimeters and muon detectors (as in the Phase-1 system), information from
1056 the new tracker and high-granularity endcap calorimeter can also be included at L1
1057 for the first time. This is illustrated in the functional diagram of the architecture of
1058 the Phase-2 trigger system in Fig. 3.1.

1059 The key feature of the Phase-2 L1 Trigger is the introduction of a correlator layer,
1060 where algorithms produce higher-level trigger objects by combining information from
1061 sub-detectors, with a selectivity approaching that of offline reconstruction in the
1062 HLT [38]. Four independent data processing paths (grouped together in Fig. 3.1) are
1063 implemented: tracking, calorimetry, muon systems, and particle-flow techniques:

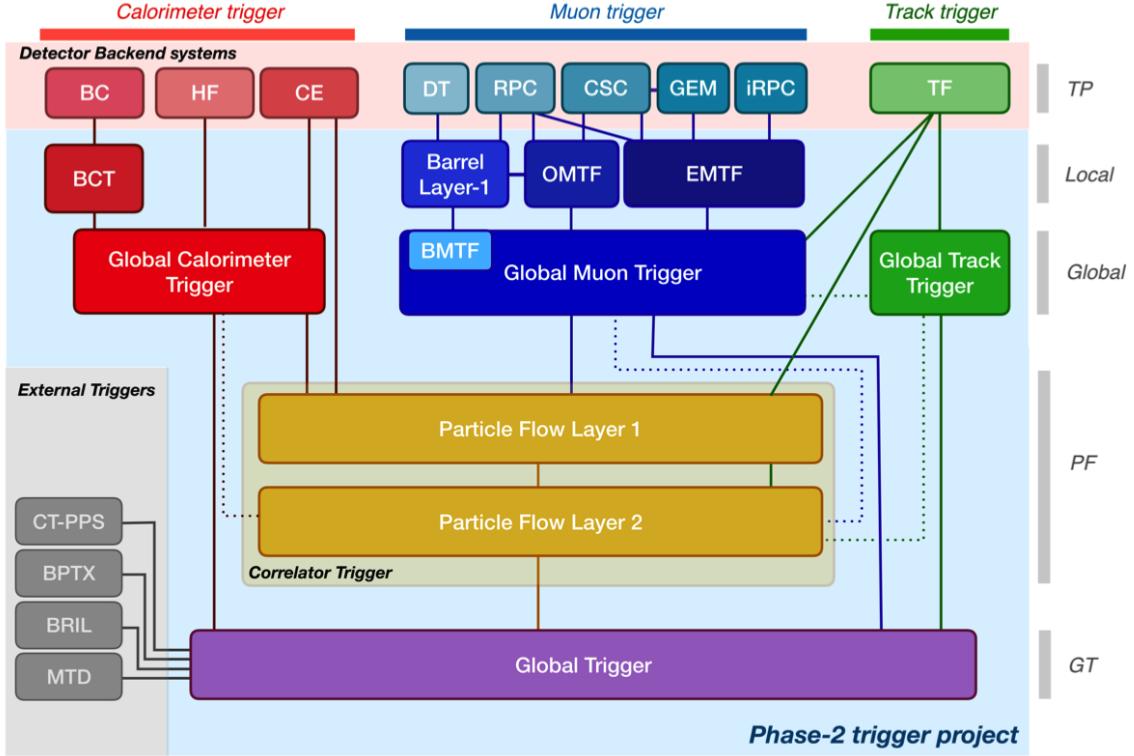


Figure 3.1: Functional diagram of the CMS L1 Phase-2 upgraded trigger design [38], showing the four trigger paths: calorimeter, muon, track, and Particle Flow. For the first time, tracking information will be available as early as the L1 Trigger.

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

- **Muon Trigger path:** (*blue*, Fig. 3.1) Trigger primitives are processed by muon track finder algorithms, again separated into the barrel (barrel muon track finder, BMTF), overlap (overlap muon track finder, OMTF), and endcap (endcap muon track finder, EMTF). Standalone muons and stubs containing information such as position, bend angle, and timing, as well as L1 tracks, are sent to the global muon trigger (GMT).
- **Particle-Flow Trigger path:** (*yellow*, Fig. 3.1) The correlator trigger (CT) aims to approach the performance of offline Particle Flow, and is implemented in two layers. “Layer-1” produces the particle-flow candidates from matching calorimeter clusters and tracks. “Layer 2” builds and sorts final trigger objects and applies additional identification and isolation criteria.

The outputs from the above trigger paths are combined in the Global Trigger (GT) (*purple*, Fig. 3.1), which calculates the final trigger decision (Level-1 Accept), transmitting it to the Trigger Control and Distribution System (TCDS), which distributes it to the detector backend systems, initiating the readout to the DAQ. The GT also provides the interface to external triggers (*grey*, Fig. 3.1), such as triggers for the precision proton spectrometer (PPS), beam position and timing monitors (BPTX), and luminosity and beam monitoring (BRIL) detectors [38]. The design of the Phase-2 Level-1 Trigger allows for future inclusion of triggering information, for instance information about minimum ionizing particles (MIPs) from the MIP Timing Detector (MTD) [39].

3.3 Standalone barrel calorimeter electron/photon reconstruction

The reconstruction and identification of electrons and photons (e/γ) begin with the trigger primitives of the barrel ECAL and HCAL detectors and endcap HGCAL calorimeters, covering the pseudorapidity region $|\eta| < 3$. The barrel and endcap regions of the detector are intrinsically different enough to warrant different approaches to e/γ reconstruction. This work presents a firmware-based emulator for the standalone e/γ reconstruction in the barrel calorimeter (Fig. 3.2). “Standalone” refers to the fact that the tracker information is not used in this particular reconstruction chain. This firmware-based emulator is based on the parallelized, computational logic that will be deployed in the firmware of the Phase-2 Level-1 trigger. The emulator uses fixed-precision integers to represent all values, such as in the computation of cluster energies, and closely mimics the firmware logic which uses arrays and performs computations in flattened loops. It represents an improved, more realistic understanding of the trigger, compared to the previous emulator which used idealized logic such as vector operations, and floats to represent all values [38].

3.3.1 Electron/photon standalone barrel procedure

In Phase-2, the upgrade of both on-detector and off-detector electronics of the barrel calorimeters’ trigger primitive generator (TPG) will enable the streaming of single crystal data from the on-detector to the backend electronics. Currently in Phase-1, the ECAL and HCAL TPGs is restricted to providing lower-granularity information of trigger tower sums of 5×5 crystals to the Level-1 Trigger [38]. A schematic of the geometry of the ECAL barrel in the Phase-2 Regional Calorimeter Trigger (RCT) is shown in Fig. 3.3. The barrel is spanned by 36 RCT cards, each spanning 17×4 towers in $\eta \times \phi$. Each RCT card is subdivided into five “regions” as shown in Fig.

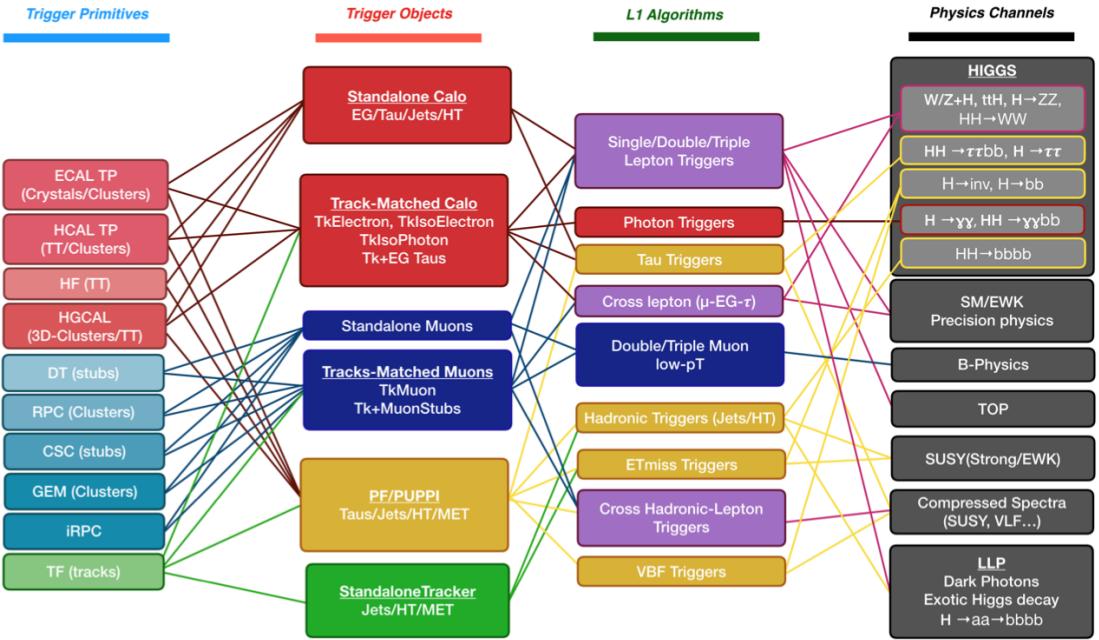


Figure 3.2: Summary of the links between the trigger primitives (*first column*), the trigger objects (*second column*), the Level-1 algorithms used in the menu (*3rd column*), and the physics channels (*4th column*), from [38], where a full description of the Phase-2 L1 algorithms can be found. This work focuses on developments for the Standalone Calorimeter electron and photon ("EG") reconstruction algorithm.

1122 3.4. After initial clustering and processing, the outputs of the RCT card are sent to
 1123 the Global Calorimeter (GCT) trigger, which is processed in three cards as shown in
 1124 Fig. 3.5. The reconstruction algorithm is detailed below.

1125 The standalone barrel algorithm for reconstructing and identifying electrons and
 1126 photons in the Phase-2 Level-1 Trigger takes as input the digitized response of each
 1127 crystal of the barrel ECAL, with a granularity 0.0175×0.0175 in $\eta \times \phi$, which is 25
 1128 times higher than the input to the Phase-1 trigger, which consisted of trigger towers
 1129 with a granularity of 0.0875×0.0875 . In HCAL the tower size of 0.0875×0.0875
 1130 is unchanged. The trigger algorithm is designed to closely reproduce the algorithm
 1131 used in the offline reconstruction, with limitations and simplifications due to trigger
 1132 latency.

1133 In the RCT, an initial requirement of $p_T > 0.5$ GeV is imposed on the input

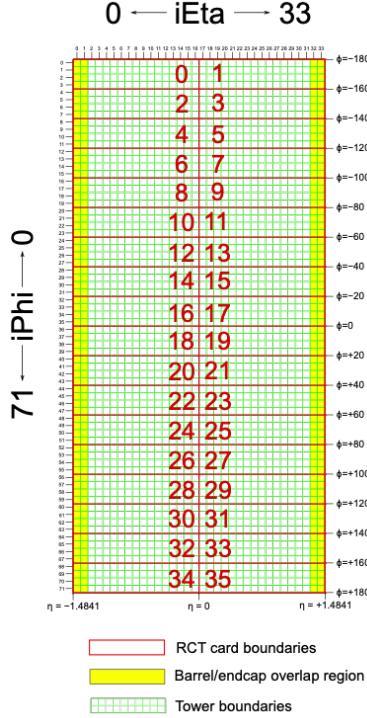


Figure 3.3: Schematic of the geometry of the Phase-2 ECAL barrel in the Regional Calorimeter Trigger (RCT), showing the division of the barrel region into 36 Regional Calorimeter Trigger (RCT) cards (*red*). Each card spans 17×4 towers in $\eta \times \phi$ (*green*), and each tower is 5×5 in single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

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

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

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

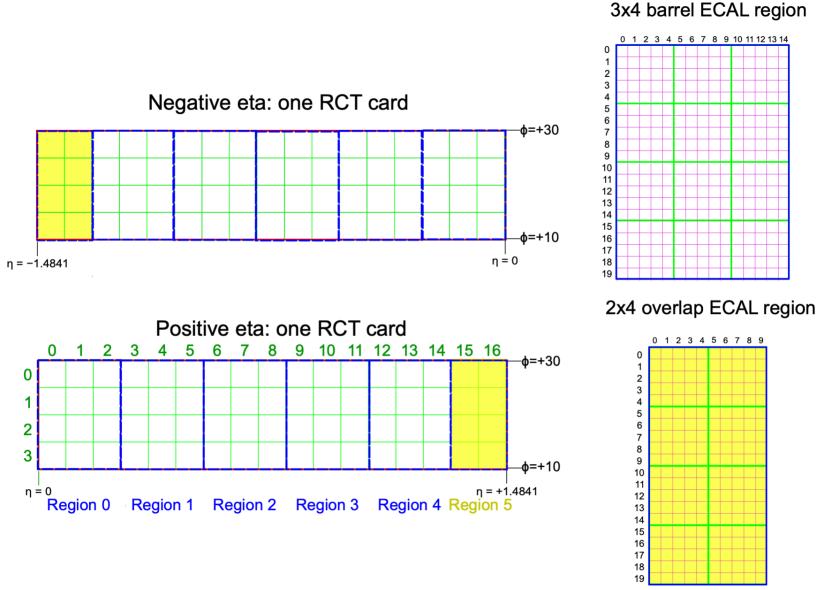


Figure 3.4: Schematic of two example RCT cards in the negative eta (*top left*) and positive eta (*bottom left*) regions of the ECAL barrel. Each RCT card is divided into six regions: five regions are of size 3×4 towers in $\eta \times \phi$ (*top right*), and a sixth smaller overlap region of size 2×4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

1145 tending to produce showers that deposit most of their energy in a 2×5 region.

- 1146 • Bremsstrahlung recovery: e/γ tend to spread in the ϕ direction due to charged
 1147 particles being bent by the magnetic field of the CMS solenoid. If sufficient
 1148 energy comparable to the core 3×5 cluster is found in the adjacent 3×5
 1149 windows (Fig. 3.6, *shaded yellow*), the energy is added to the core cluster and
 1150 no longer considered unclustered energy.

1151 After parallel processing in the regions, the clusters in a RCT card are stitched
 1152 together if they are located directly along the borders of a region (Fig. 3.3). The
 1153 remaining unclustered ECAL energy is summed into ECAL towers.

1154 From each RCT card, the twelve highest-energy clusters, as well as any remaining
 1155 unclustered energy, are sent to the GCT. Since each GCT card has information from
 1156 sixteen RCT cards (Fig. 3.5), final stitching across the boundaries of the RCT cards
 1157 is performed. One more identification algorithm is performed at this stage:

GCT/“Layer 2”

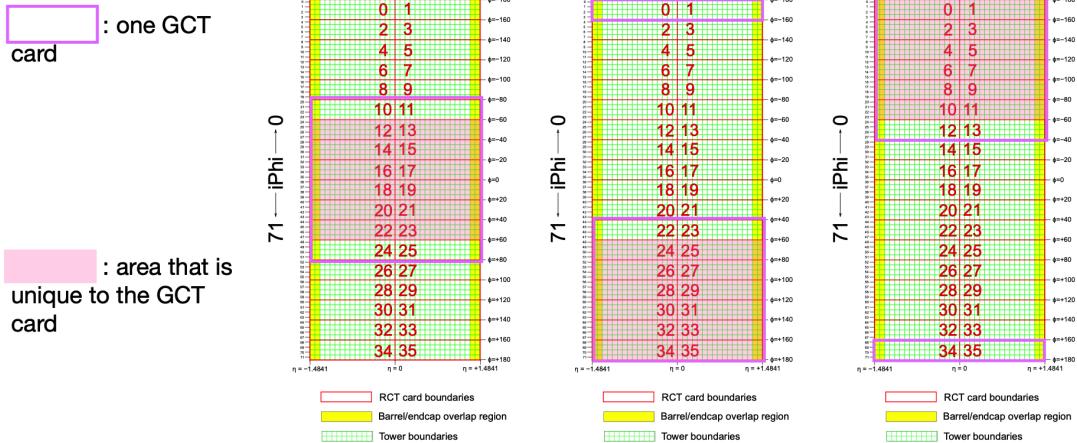


Figure 3.5: Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trigger (GCT), which will process the outputs of the Regional Calorimeter Trigger (RCT) in three GCT cards (*purple borders*). Each card in the GCT processes the equivalent of sixteen RCT cards, with the center twelve RCT cards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

1158 • Isolation: One handle to reject backgrounds from e.g. pile-up, comes from the
 1159 tendency for background to be spread more uniformly across a large area in the
 1160 detector, whereas genuine e/γ are expected to produce showers concentrated in
 1161 the 3×5 crystal window. The energy sum in a large window of 7×7 in towers
 1162 is computed and used to reject background.

1163 Flags that provide discrimination power between genuine e/γ and background, are
 1164 computed using the relative isolation and shower shape quantities. The standalone
 1165 working point (WP) is defined as the logical OR of the relative isolation and shower
 1166 shape flags.

1167 The information of the clusters in the event, including their energies, crystal-level
 1168 position, the relative isolation flag, the shower shape flag, the standalone WP, and
 1169 the ratio of the HCAL over ECAL energies, are sent in 64 bits to the Correlator
 1170 Trigger and the Global Trigger. The towers in the event are computed as the sum
 1171 of all unclustered energy in the ECAL with the corresponding HCAL energy at each

3x4 barrel ECAL region

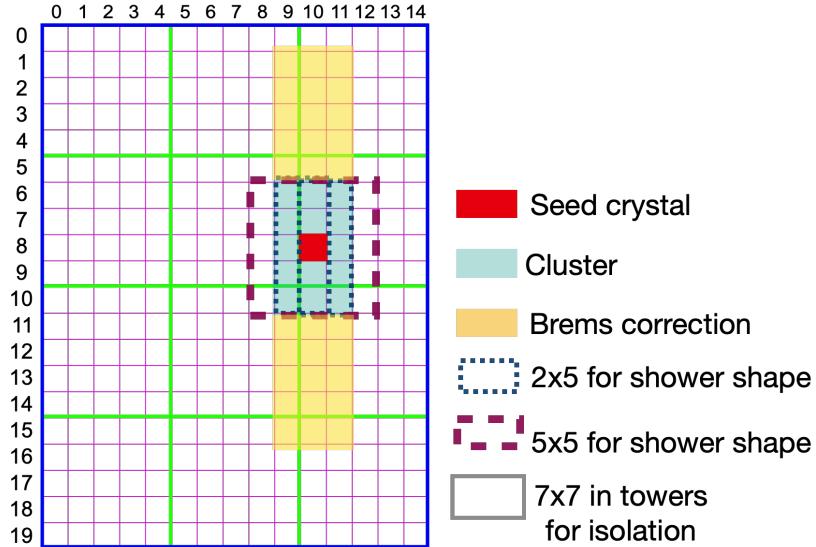


Figure 3.6: Illustration of an example electron/photon (e/γ) cluster in the Phase-2 Level-1 Trigger standalone barrel e/γ reconstruction, in a region of 15×20 crystals (3×4 towers) in $\eta \times \phi$. Each small pink square is one crystal, the highest-granularity ECAL trigger primitives available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

¹¹⁷² tower location, and their energies are sent to the Correlator Trigger.

¹¹⁷³ 3.3.2 Electron/photon standalone barrel results

¹¹⁷⁴ The performance of the current emulator of the standalone barrel e/γ algorithm in
¹¹⁷⁵ Phase-2 conditions is quantified in efficiencies and rates. Efficiency is the fraction of
¹¹⁷⁶ true electrons that the algorithm can reconstruct and identify, and is evaluated in
¹¹⁷⁷ a Monte Carlo simulated sample containing electrons with transverse momentum p_T
¹¹⁷⁸ ranging from 1 to 100 GeV. The efficiencies of the current and previous emulators as

1179 a function of the electron generator-level p_T are shown in Fig. 3.7.

1180 The rates are the event rates that this reconstruction and identification algorithm
1181 would obtain if it were deployed in a trigger, assuming that proton-proton collisions
1182 are occurring at the 40 MHz event rate of the HL-LHC. The rate is reported as a
1183 function of the minimum energy threshold required by the trigger, and is estimated
1184 using a simulated sample of minimum bias events, i.e. generic proton-proton colli-
1185 sions without any specific physics selections. The rates for the current and previous
1186 emulator are shown in Fig. 3.8.

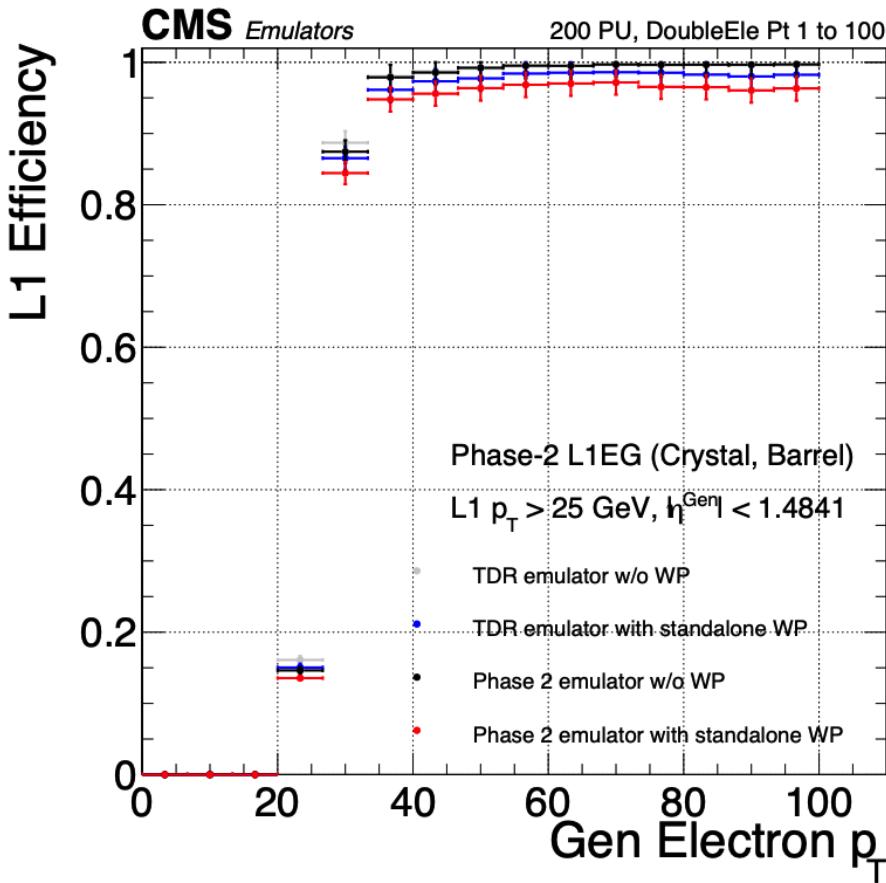


Figure 3.7: Efficiencies of the current and previous emulators of the standalone barrel e/γ algorithm for the Phase-2 Level-1 Trigger, evaluated in a simulated sample containing electrons, as a function of the electron's generator-level transverse momentum p_T . The standalone working point (WP) is defined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

1187 The current emulator is incorporated into the full Phase-2 L1 menu, allowing an
1188 estimate of the rates produced by the standalone e/γ barrel trigger path and all
1189 other algorithms in the L1 Trigger. All rates are estimated with the assumption of
1190 an average pile-up of 200 and event rate of 40 MHz. The standalone working point
1191 single e/γ path with requirements on the e/γ candidate to have $|\eta| < 2.4$, offline p_T
1192 to be greater than 51 GeV, and online p_T to be greater than 41 GeV, is projected to
1193 have a rate of around 23 kHz. The standalone working point double e/γ path with
1194 requirements on the two e/γ candidates to have $|\eta| < 2.4$, offline p_T greater than 37
1195 and 24 GeV, and online p_T greater than 29 and 18 GeV, is projected to give a rate of
1196 around 6 kHz. For both paths, the objects efficiency plateau is 99%.

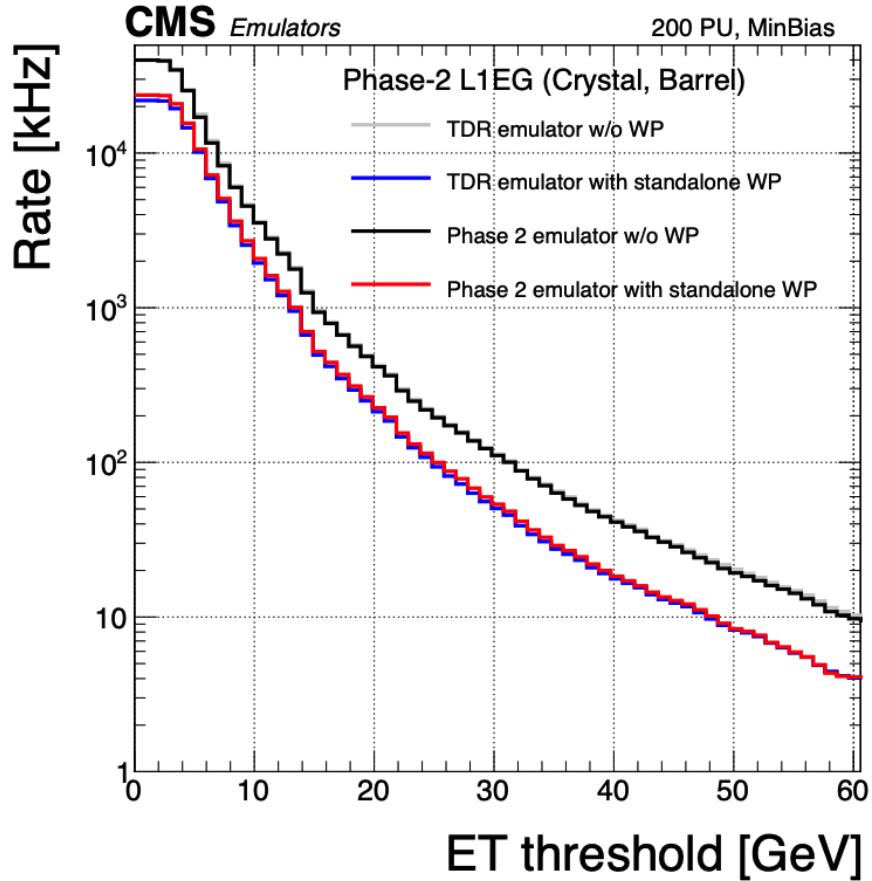


Figure 3.8: Rates in kHz of the current Phase-2 and previous (“TDR”) emulators of the standalone barrel e/γ algorithm for the Phase-2 Level-1 Trigger, evaluated on a minimum bias (MinBias) sample with 200 pile-up (PU), measured as a function of the minimum energy (E_T) required of the reconstructed e/γ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

₁₁₉₇ **Chapter 4**

₁₁₉₈ **Datasets and Monte Carlo samples**

₁₁₉₉ The search for the exotic decay of the 125 GeV Higgs boson to two light neutral scalars
₁₂₀₀ decaying to a pair of bottom quarks and a pair of tau leptons ($h \rightarrow aa \rightarrow bb\tau\tau$) is
₁₂₀₁ based on proton-proton collision data at a center-of-mass energy 13 TeV collected
₁₂₀₂ in Run-2 of data-taking, spanning the data-taking years 2016, 2017, and 2018. The
₁₂₀₃ datasets used and the triggers used to collect the data are described in Section 4.1.
₁₂₀₄ Section 4.2 describes the Monte Carlo simulated samples that are used to model the
₁₂₀₅ $h \rightarrow aa \rightarrow bb\tau\tau$ signal and background Standard Model processes. Lastly, in order
₁₂₀₆ to obtain a better description of Standard Model backgrounds that contain two tau
₁₂₀₇ leptons, a data-Monte Carlo hybrid technique is used to generate embedded samples
₁₂₀₈ which model processes with genuine $\tau\tau$ in the final state, as detailed in Section 4.3.
₁₂₀₉ All samples are listed in Appendix A.

₁₂₁₀ **4.1 Datasets used**

₁₂₁₁ The $h \rightarrow aa \rightarrow bb\tau\tau$ analysis [40] is based on proton-proton collision data at a center-
₁₂₁₂ of-mass energy of 13 TeV collected in full Run-2 (2016-18) with the CMS detector.
₁₂₁₃ The data analyzed corresponds to a total integrated luminosity of 138 fb^{-1} (36.33 fb^{-1}
₁₂₁₄ for 2016, 41.53 fb^{-1} for 2017, and 59.74 fb^{-1} for 2018) [41] [42] [43]. The cumulative

1215 delivered and recorded luminosity versus time for 2015-2018 is shown in Fig. 4.1.

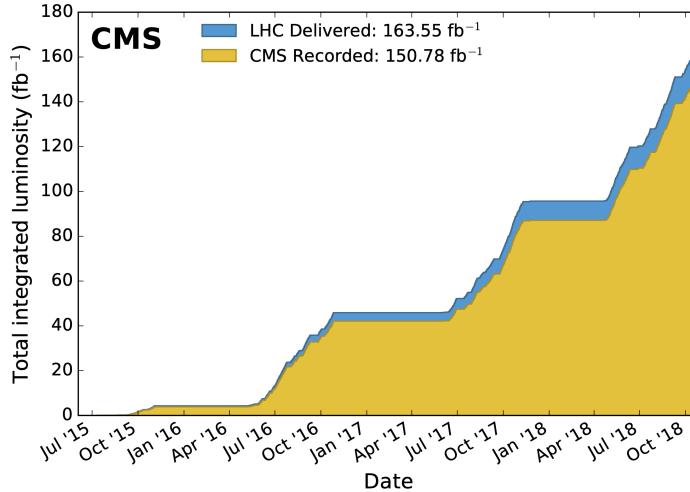


Figure 4.1: Cumulative delivered and recorded luminosity versus time for 2015-2018 at CMS, in proton-proton collision data only, at nominal center-of-mass energy [44].

1216 Data collected with the single muon trigger is used for the $\mu\tau_h$ channel. For the
1217 $e\tau_h$ channel, data collected with the single electron trigger is used; and for the $e\mu$
1218 channel, data collected with the electron + muon trigger is used. A more in-depth
1219 discussion of the triggers used follows in a later section. The datasets are listed in
1220 Appendix A in Tables A.1, A.2, and A.3.

1221 4.2 Monte Carlo samples

1222 Modeling and computing observables originating from arbitrary physics processes at
1223 the tree level and at next-to-leading order (NLO) is performed by Monte Carlo (MC)
1224 event generators, such as Powheg and MadGraph5_amCNLO [45] [46]. The informa-
1225 tion generated, e.g. the computation of the differential cross sections and kinematics
1226 of the final state particles, is saved in a compressed file and used to generate MC sam-
1227 ples that are used in physics analyses. The samples are digitized using GEANT4 [47],
1228 a platform used at the LHC and other facilities to comprehensively simulate the

1229 passage of particles through matter. The digitized samples are passed through the
1230 same detector reconstruction as real data events collected in the detector. The MC
1231 background samples used in this analysis for 2016-2018 are listed in Appendix A in
1232 Tables A.7, A.8, and A.9.

1233 The samples for modeling the signal ($h \rightarrow aa \rightarrow 2b2\tau$ and $h \rightarrow a_1a_2$) in the
1234 2HDM+S and TRSM are generated at tree-level, for a range of masses of the light
1235 neutral scalar a . For $h \rightarrow aa$, the mass hypotheses for the a range from $m_a =$
1236 (12 GeV, 62.5 GeV). For $h \rightarrow a_1a_2$, the mass hypotheses for the two light scalars span
1237 combinations of m_{a1} , m_{a2} ranging from (12 GeV, 62.5 GeV) for the two scalars. The
1238 MC signal samples used in this analysis for 2016-2018 are listed in Appendix A in
1239 Tables A.10, A.11 and A.12.

1240 4.3 Embedded samples

1241 An important background for Higgs boson studies and searches for additional Higgs
1242 bosons is the decay of Z bosons into pairs of τ leptons ($Z \rightarrow \tau\tau$). An embedded tech-
1243 nique was developed in the context of Standard Model Higgs to $\tau\tau$ measurements, to
1244 model $Z \rightarrow \tau\tau$ decays, and was expanded to also model all Standard Model processes
1245 that contain $\tau\tau$ [48]. The embedded technique has since been used successfully at
1246 CMS for the Standard Model $H \rightarrow \tau\tau$ measurement, as well as searches for minimal
1247 supersymmetric extensions to the Standard Model (MSSM) [49] [50].

1248 The advantage of the embedded technique is that aspects of the event that are
1249 difficult to model and describe are directly taken from data, resulting in a better
1250 data description than can be achieved with only the $Z \rightarrow \tau\tau$ simulation [48]. The
1251 simulation must be tuned extensively to accurately model aspects of the data, such
1252 as time-dependent pile-up profiles, the production of additional jets, e.g. in multijet
1253 and vector boson fusion topologies, the number of reconstructed primary interaction

1254 vertices, and the missing transverse momentum p_T^{miss} . Since all events with genuine
 1255 $\tau\tau$ are estimated with samples made with the embedded technique (referred to as
 1256 embedded samples from here on), events in Monte Carlo simulation with genuine $\tau\tau$
 1257 are not used, in order to avoid double-counting.

1258 Fig. 4.2 shows a schematic of how embedded samples are produced. Data events
 1259 containing $Z \rightarrow \mu\mu$ decays are selected. In these events, all energy deposits of the
 1260 recorded muons are removed, and are replaced with simulated tau leptons with the
 1261 same kinematic properties as the removed muons. This results in a hybrid data format
 1262 containing information from both observed and simulated events, as illustrated in Fig.
 1263 4.2 [48]. The embedded samples used for the years 2016-2018 are listed in Appendix
 1264 A in Tables A.4, A.5, and A.6.

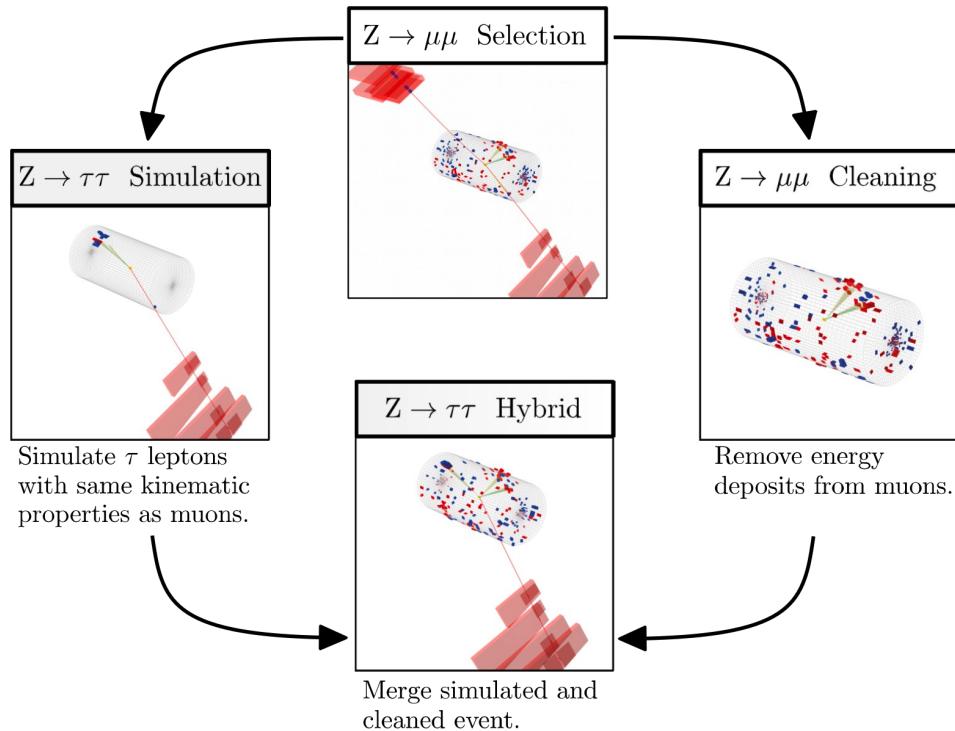


Figure 4.2: Schematic view of the four main steps of the embedding technique for τ leptons, as described in Section 4.3 [48]. A $Z \rightarrow \mu\mu$ event is selected in data ($Z \rightarrow \mu\mu$ selection), all of the energy deposits associated with the muons are removed ($Z \rightarrow \mu\mu$ cleaning), and two τ leptons and their decays are simulated in an empty detector ($Z \rightarrow \tau\tau$ simulation). Lastly, all energy deposits of the simulated τ decays are combined with the data event ($Z \rightarrow \tau\tau$ hybrid).

1265 In the selection step of the embedded technique, events are selected with at least
 1266 one of a set of $\mu\mu$ trigger paths, which require $p_T > 17(8)$ GeV for the leading
 1267 (sub-leading) muons, and a minimum requirement between 3.8 and 8.0 GeV on the
 1268 invariant di-muon mass $m_{\mu\mu}$ [48]. The offline reconstructed muons must match the
 1269 objects at trigger level and also have offline $p_T > 17(8)$ GeV. They must have $|\eta| < 2.4$
 1270 and be located at a distance $|d_z| < 0.2$ cm to the primary vertex along the beam
 1271 axis. To form a Z boson candidate, each muon is required to originate from a global
 1272 muon track. The muon pairs must have opposite charges with an invariant mass of
 1273 $m_{\mu\mu} > 20$ GeV. If more than two di-muon pairs are found, the pair with the invariant
 1274 mass closest to the Z boson mass (91.19 GeV) is chosen.

1275 This selection is designed to be tight enough to ensure a high purity of genuine
 1276 $\mu\mu$ events, and also loose enough to minimize biases of the embedded event samples.
 1277 Isolation requirements are avoided, since they would introduce a bias towards less
 1278 hadronic activity in the vicinities of the embedded leptons that will appear more
 1279 isolated than expected in data. The selection results in an expected mixture of events
 1280 summarized in Table 4.1 from [48]. $Z \rightarrow \mu\mu$ is the dominant process modeled by the
 1281 embedded technique, with $t\bar{t}$, QCD, and diboson and single top processes becoming
 1282 more significant when considering events with b-tag jets.

Process	Fraction (%)		
	Inclusive	$m_{\mu\mu} > 70$ GeV	$N(\text{b-tag jets}) > 0$
$Z \rightarrow \mu\mu$	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 [48], before additional cuts (i.e. inclusive, *column 2*), and after adding a requirement on the di-muon mass $m_{\mu\mu} > 70$ GeV (*column 3*), or a requirement on the number of b-tag jets in the event (*column 4*).

₁₂₈₃ **Chapter 5**

₁₂₈₄ **Object reconstruction and
1285 corrections applied**

₁₂₈₆ In the data processing workflow, data events and simulated events are analyzed to
₁₂₈₇ reconstruct physics objects of interest, and algorithms for distinguishing genuine par-
₁₂₈₈ ticle candidates from background, are employed. Section 5.1 describes the physical
₁₂₈₉ properties of the most important objects in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis: taus,
₁₂₉₀ muons, electrons, jets, and jets originating from b-quarks (b-flavor jets), as well as
₁₂₉₁ their reconstruction and identification in CMS. In this analysis, the full energy and
₁₂₉₂ momentum of the two tau leptons ($m_{\tau\tau}$) is estimated from the measured (i.e. visible)
₁₂₉₃ components of the tau leptons using the SVFit/FastMTT algorithm, which is de-
₁₂₉₄ scribed in Section 5.2. Corrections are applied to the simulated samples at the object
₁₂₉₅ level and the event level to account for known discrepancies between simulations and
₁₂₉₆ the data that the simulations are intended to model. These corrections are listed and
₁₂₉₇ detailed in Section 5.3.

₁₂₉₈ **5.1 Object reconstruction**

₁₂₉₉ **5.1.1 Taus**

₁₃₀₀ The tau (τ) is the heaviest known lepton. With a rest mass of 1776.86 MeV, it can
₁₃₀₁ decay to not only electrons and muons, but also hadrons. In two thirds of the cases, τ
₁₃₀₂ leptons decay hadronically, typically into one or three charged mesons (predominantly
₁₃₀₃ π^+ , π^-), often accompanied by neutral pions (that decay $\pi^0 \rightarrow \gamma\gamma$), and a ν_τ . These
₁₃₀₄ hadronic decays are denoted τ_h . In the remainder of the decays, the tau decays to
₁₃₀₅ the lighter leptons (electron or muon), termed leptonic decays. The mean lifetime of
₁₃₀₆ the τ is $\tau = 290 \times 10^{-15}$ seconds, corresponding to $c\tau = 87.03 \mu\text{m}$, which is short
₁₃₀₇ enough that taus decay in the CMS detector before reaching the detector elements,
₁₃₀₈ but also long enough that some decay length variables can help with hadronic tau
₁₃₀₉ identification. The tau's largest decay branching ratios (proportional to probability
₁₃₁₀ of decay) are listed below [26]:

- ₁₃₁₁ • 17.8% decay to $e^-\bar{\nu}_e\nu_\tau$
- ₁₃₁₂ • 17.4% decay to $\mu^-\bar{\nu}_\mu\nu_\tau$
- ₁₃₁₃ • 25.5% decay to $\pi^-\pi^0\nu_\tau$ (ρ^- resonance at 770 MeV)
- ₁₃₁₄ • 10.8% decay to $\pi^-\nu_\tau$
- ₁₃₁₅ • 9.3% decay to $\pi^-\pi^0\pi^0\nu_\tau$ (a_1^- resonance at 1200 MeV)
- ₁₃₁₆ • 9.0% decay to $\pi^-\pi^-\pi^+\nu_\tau$ (a_1^- resonance at 1200 MeV)

₁₃₁₇ In all cases, at least one neutrino is produced. The neutrinos escape undetected
₁₃₁₈ from the CMS detector, resulting in missing transverse energy. Charged hadrons leave
₁₃₁₉ tracks in the tracking detector before being absorbed in the hadronic calorimeter; in
₁₃₂₀ CMS tau reconstruction terminology, they are often called “prongs”, i.e. the dominant

1321 τ_h decay modes are termed “1 prong” (π^\pm), “1 prong + π^0 (s)”, and “3-prong”. Neutral
1322 pions decay to two photons which lose their energy in the electromagnetic calorimeter.
1323 Taus that decay to electrons and muons, are typically triggered on and reconstructed
1324 as electrons and muons respectively.

1325 **Hadron plus strips (HPS) reconstruction of τ_h**

1326 At CMS, hadronically decaying tau leptons are reconstructed with the hadron plus
1327 strips (HPS) algorithm [51] [52]. The HPS algorithm capitalizes on photon conversions
1328 in the CMS tracker material, which originate from the neutral pion (π^0) decaying
1329 to two photons. The bending of electron/positron tracks due to the CMS solenoid
1330 magnetic field leads to a spread of the neutral pions’ calorimeter signatures in the ϕ
1331 direction. This motivates the reconstruction of photons in “strips”: objects that are
1332 built out of PF photons and electrons. The strip reconstruction starts with centering
1333 a strip on the most energetic electromagnetic particle in a PF jet. Among other
1334 electromagnetic particles located in a window of size $\Delta\eta = 0.05$ and $\Delta\phi = 0.20$
1335 around the strip center, the most energetic one is associated with the strip and its
1336 momentum is added to the strip momentum. This is repeated iteratively until no
1337 further particles can be associated. Lastly, strips satisfying a requirement of $p_T^{\text{strip}} > 1$
1338 GeV are combined with charged hadrons to reconstruct individual τ_h decay modes,
1339 where h stands for both π and K :

- 1340 • *Single hadron:* $h^- \nu_\tau$ and $h^- \pi^0 \nu_\tau$ decay modes, in which the neutral pions have
1341 too little energy to be reconstructed as strips.
- 1342 • *One hadron + one strip:* $h^- \pi^0 \nu_\tau$ decay modes, where the photons from the π^0
1343 decay are close together in the calorimeter.
- 1344 • *One hadron + two strips:* $h^- \pi^0 \nu_\tau$ decay modes, where the photons from the π^0
1345 decay are well separated.

- 1346 • *Three hadrons:* $h^- h^+ h^- \nu_\tau$ decay modes. The three charged hadrons are re-
1347 quired to originate from the same secondary vertex.

1348 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
1349 reconstructed with the above topologies.

1350 In the HPS algorithm, the direction of the reconstructed tau momentum \vec{p}^{τ_h}
1351 is required to fall within a distance of $\Delta R = 0.1$ from the original PF jet. All
1352 charged hadrons and strips are required to be contained within a cone of size $\Delta R =$
1353 $(2.8 \text{ GeV})/p_T^{\tau_h}$, from the τ_h as reconstructed by the HPS.

1354 All charged hadrons are assumed to be pions, and they are required to be consis-
1355 tent with the masses of the intermediate meson resonances (if applicable), with the
1356 following allowed windows for candidates: 50-200 MeV for π^0 , 0.3-1.3 GeV for ρ , and
1357 0.8-1.5 GeV for a_1 . If the τ_h decay is compatible with more than one hypothesis, the
1358 one giving the highest $p_T^{\tau_h}$ is chosen. Lastly, an isolation requirement is applied: aside
1359 from the τ_h decay products, no charged hadrons or photons can be present within
1360 an isolation cone of size $\Delta R = 0.5$ around the direction of the τ_h . The outputs of
1361 the HPS algorithm are the reconstructed decay mode and the visible four-momentum
1362 (i.e. the four-momenta of all decay products excluding the neutrinos).

1363 DeepTau for identifying τ_h

1364 The identification of τ_h candidates in CMS has historically been divided into separate
1365 discriminators against jets, electrons, and muons. Discriminators versus jets and
1366 electrons use information from derived quantities, such as the p_T sum of particles
1367 near the τ_h axis. Building on the previous multivariate analysis (MVA) classifier [53]
1368 based on a boosted decision tree (BDT), DeepTau is a more recent classifier based on a
1369 deep neural network (DNN) that simultaneously discriminates against jets, electrons,
1370 and muons. The DNN uses a combination of high-level inputs, similar to previous
1371 algorithms, and also uses convolutional layers in η - ϕ space to process information

1372 from all reconstructed particles near the τ_h axis. Convolutional layers are based on
1373 the principle that an image can be processed independently of its position.

1374 The final DeepTau discriminators against jets, muons, and electrons are given by

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

1375 where y_τ (y_α) are estimates of the probabilities for the τ_h candidate to come from
1376 a genuine τ_h (jet, μ , e). Working points for each discriminator with different τ_h
1377 identification efficiencies are defined for D_e , D_μ , and D_{jet} , for usage in physics analyses
1378 and derivation of data-to-simulation corrections [54].

1379 5.1.2 Muons

1380 Muons are the next lightest lepton after taus, with a mass of 105.66 MeV and a
1381 mean lifetime of $\tau = 2.20 \times 10^{-6}$ seconds, or $c\tau = 658.64$ m. At CMS, muons are
1382 identified with requirements on the quality of the track reconstruction and on the
1383 number of measurements in the tracker and the muon systems [55]. In the standard
1384 CMS reconstruction, tracks are first reconstructed independently in the inner tracker
1385 (tracker track) and in the muon system (standalone-muon track). Next, these tracks
1386 are processed in two different methods.

1387 The first is Global Muon reconstruction (outside-in) [55], which fits combined hits
1388 from the tracker track and standalone-muon track, using the Kalman-filter technique.
1389 At large transverse momenta, $p_T \gtrsim 200$ GeV, the global-muon fit can improve the
1390 momentum resolution compared to the tracker-only fit.

1391 The second is Tracker Muon reconstruction (inside-out) [55], which starts with
1392 tracker tracks with $p_T > 0.5$ GeV and total momentum $p_T > 2.5$ GeV. These tracks
1393 are extrapolated outwards to the muon system and matched to detector segments
1394 there, taking into account the magnetic field, expected energy losses, and multiple

1395 Coulomb scattering in the detector material. Tracker Muon reconstruction is more
 1396 efficient than the Global Muon reconstruction at low momenta, $p \lesssim 5$ GeV, because
 1397 it only requires a single muon segment in the muon system, whereas Global Muon
 1398 reconstruction typically requires segments in at least two muon stations.

1399 To further suppress fake muons from decay in flight, isolation cuts are used. A
 1400 relative isolation variable is defined to quantify the energy flow of particles near the
 1401 muon trajectory. A relative isolation is defined similarly for muons and electrons:

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

1402 where $\sum_{\text{charged}} p_T$ is the scalar sum of the p_T of the charged particles originating from
 1403 the primary vertex and located in a cone of size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4(0.3)$
 1404 centered on the direction of the muon (electron). The sum $\sum_{\text{neutral}} p_T$ is the equivalent
 1405 for neutral particles. The sum $\sum_{\text{charged, PU}} p_T$ is the scalar sum of the p_T of the
 1406 charged hadrons in the cone originating from pile-up vertices. The factor 1/2 comes
 1407 from simulation estimations, which find that the ratio of neutral to charged hadron
 1408 production in the hadronization process of inelastic pp collisions is 1/2. Thus the
 1409 subtracted term is intended to subtract contribution from pile-up, from the neutral
 1410 particle contribution to the isolation sum. Finally, this is divided by the lepton
 1411 transverse momentum, p_T^ℓ .

1412 5.1.3 Electrons

1413 Electrons are the lightest lepton with a mass of 0.511 MeV. At CMS, electrons are
 1414 reconstructed by associating a track reconstructed in the silicon tracking detector
 1415 with a cluster of energy in the ECAL. Performance is maximized via a combination
 1416 of a stand-alone approach and the complementary global particle-flow approach [56].

1417 In the stand-alone approach, the electron energy, which is typically spread over

1418 several crystals of the ECAL, is clustered with the “hybrid” algorithm in the barrel
1419 and the “multi- 5×5 ” in the endcaps [56]. The hybrid algorithm collects energy in a
1420 small window in η and an extended window in ϕ . It identifies a seed crystal, and adds
1421 arrays of 5×1 crystals in $\eta \times \phi$ in a range of $N = 17$ crystals in both directions of
1422 ϕ , if their energies exceed a minimum threshold, thus forming a supercluster (SC). In
1423 the endcap, crystals are not arranged in an $\eta \times \phi$ geometry; instead clusters are build
1424 around seed crystals in clusters of 5×5 crystals that can partly overlap. Nearby
1425 clusters are grouped into a supercluster, and energy is recovered from associated
1426 deposits in the preshower.

1427 In the PF reconstruction [56], PF clusters are reconstructed by aggregating around
1428 a seed all contiguous crystals with energies two standard deviations above the elec-
1429 tronic noise observed at the beginning of a data-taking run. The energy of a given
1430 crystal can be shared among two or more clusters.

1431 The electron track reconstruction is performed in two ways [56]: the ECAL-based
1432 seeding, which begins with the SC energy and positioning, and the tracker-based
1433 seeding (part of the PF reconstruction algorithm), which uses tracks reconstructed
1434 from the general algorithm for charged particles, extrapolated towards the ECAL and
1435 matched to an SC. Kalman filter (KF) tracks with a small number of hits or that are
1436 not well-fitted, are re-fitted with a dedicated Gaussian sum Filter (GSF).

1437 A global identification variable [56] is defined using a multivariate analysis (MVA)
1438 technique that combines information on track observables (kinematics, quality of the
1439 KF track and GSF track), the electron PF cluster observables (shape and pattern),
1440 and the association between the two (geometric and kinematic observables). For
1441 electrons seeded only through the tracker-based approach, a weak selection is applied
1442 on this MVA variable. For electrons seeded through both approaches, a logical OR is
1443 taken.

1444 Electron isolation, i.e. the presence of energy deposits near the electron trajectory,

1445 is a separate key handle in rejecting significant background. Compared to isolated
1446 electrons, electrons from misidentified jets or genuine electrons within a jet resulting
1447 from semileptonic decays of b or c quarks tend to have significant energy deposits
1448 near the primary trajectory [56]. Offline analyses benefit from the PF technique
1449 for defining isolation, which sums the PF candidates reconstructed located within a
1450 specified isolation cone around the electron candidate, as in Eqn. 5.2.

1451 5.1.4 Jets

1452 The vast majority of processes of interest at the LHC contains quarks or gluons in
1453 the final state, but these particles cannot be observed directly. In a process called
1454 hadronization, they fragment into spatially-grouped collections of particles called jets,
1455 which can be detected in the tracking and calorimeter systems. Hadronization and
1456 the subsequent decays of unstable hadrons can produce hundreds of nearby particles
1457 in the CMS detector. Jets are reconstructed by the PF algorithm (PF jets), or from
1458 the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo
1459 jets). In PF jets, typically used in offline analyses, jets are built using the anti- k_T
1460 (AK) clustering algorithm [57]. The anti- k_T algorithm iterates over particle pairs and
1461 finds the two that are closest in a distance measure d , and determines whether to
1462 combine them:

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

1463 where $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ and $p_{T,i}$, η_i , ϕ_i are the transverse momentum, rapid-
1464 ity, and azimuthal angle of particle i . The power -2 means that higher-momentum
1465 particles are clustered first, leading to jets that tend to be centered on the hardest
1466 (highest p_T) particle.

1467 There are several methods to remove contributions of pile-up collisions from jet

1468 clustering [58]:

- 1469 • Charged hadron subtraction (CHS), which removes all charged hadron candi-
1470 dates associated with a track that is not associated with the primary vertex.
- 1471 • PileUp Per Particle Identification (PUPPI), which weighs input particles based
1472 on their likelihood of arising from pile-up. QCD particles tend to have a collinear
1473 structure, compared to soft diffuse radiation coming from pile-up. The local
1474 shape for charged pile-up, used as a proxy for all pile-up particles, is used on an
1475 event-by-event basis to calculate a weight for each particle. PUPPI is deployed
1476 in Run-2 and is more performant than CHS in high pile-up scenarios.

1477 **5.1.5 B-flavored jets**

1478 Jets that arise from bottom-quark hadronization (b-flavor jets) have overwhelming
1479 background from processes involving jets from gluons (g) and light-flavor quarks (u, d,
1480 s), and from c-quark fragmentation. The ability to identify b-flavor jets, or b-tagging,
1481 exploits the b hadrons' relatively large masses, long lifetimes, and daughter particles
1482 with hard momentum spectra [57].

1483 The impact parameter (IP) of a track is the 3-dimensional distance between the
1484 track and the primary vertex (PV) at the point of closest approach. The IP is positive
1485 if the track originates from the decay of particles travelling along the jet axis. The
1486 resolution of the IP depends on the p_T and η of the track, motivating the use of the
1487 impact parameter significance S_{IP} (ratio of the IP to its estimated uncertainty) as an
1488 observable [57].

1489 Because of the large but finite lifetimes of the b hadrons, b hadrons tend to
1490 travel a short distance before decaying at a secondary vertex (SV), which can be
1491 measured and reconstructed separately from the primary vertex due to the excellent
1492 position resolution of the pixel detector [57]. Previous b-tagging algorithms (e.g.

1493 CSV, cMVAv2, and DeepCSV) have capitalized on variables such as the presence of
1494 a SV, the flight distance and direction (computed from the vector between the PV
1495 and the SV), and kinematics of the system of associated secondary tracks (e.g. track
1496 multiplicity, mass, and energy).

1497 The DeepJet (formerly known as DeepFlavour) algorithm [59] is a deep-neural-
1498 network multi-classification algorithm, which uses 16 properties of up to 25 charged
1499 and 6 properties of 25 neutral particle-flow jet constituents, as well as 17 properties
1500 from up to 4 secondary vertices associate with the jet. Compared to the previous clas-
1501 sifying algorithm DeepCSV, DeepJet has been demonstrated to have higher efficiency
1502 with lower misidentification probability in Phase-1 data [60].

1503 5.2 Reconstruction of the di-tau mass

1504 The final signal extraction is done to the total di-tau ($\tau\tau$) mass, which is estimated
1505 from the visible $\tau\tau$ mass using the FastMTT algorithm [61]. FastMTT is based on the
1506 SVFit algorithm, originally developed for the Standard Model $H \rightarrow \tau\tau$ analysis [62].
1507 Both the SVFit algorithms, and the FastMTT algorithm, are described below, to give
1508 a complete picture of how the algorithms attempt to reconstruct the true invariant
1509 mass of a Higgs or Z boson decay.

1510 To specify a hadronic τ decay, six parameters are needed [62]: the polar and
1511 azimuthal angles of the visible decay product system in the τ rest frame, the three
1512 boost parameters from the τ rest frame to the laboratory frame, and the invariant
1513 mass m_{vis} of the visible decay products. For a leptonic τ decay, two neutrinos are
1514 produced, and a seventh parameter, the invariant mass of the two-neutrino system, is
1515 necessary. The unknown parameters are constrained by four observables that are the
1516 components of the four-momentum of the system formed by the visible decay products
1517 of the τ lepton, measured in the laboratory frame. The remaining unconstrained

1518 parameters for hadronic and leptonic τ decays are thus:

1519 • The fraction of the τ energy in the laboratory frame carried by the visible decay
1520 products,

1521 • ϕ , the azimuthal angle of the τ direction in the laboratory frame,

1522 • $m_{\nu\nu}$, the invariant mass of the two-neutrino system in leptonic τ decays (for
1523 hadronic τ decays, $m_{\nu\nu}$ is set to 0).

1524 E_x^{miss} and E_y^{miss} , the x and y components of the missing transverse energy E_T^{miss}
1525 provide two further constraints.

1526 5.2.1 Original SVFit ‘‘standalone’’: maximum likelihood

1527 In one of the original versions of SVFit, called ‘‘standalone’’ SVFit [62], a maximum
1528 likelihood fit method is used to reconstruct the mass $m_{\tau\tau}$ by combining the measured
1529 observables E_x^{miss} and E_y^{miss} with a likelihood model that includes terms for the τ
1530 decay kinematics and the E_T^{miss} resolution [62]. The likelihood function $f(\vec{z}, \vec{y}, \vec{a}_1 \vec{a}_2)$
1531 of the parameters $\vec{z} = (E_x^{\text{miss}}, E_y^{\text{miss}})$ in an event is constructed, where the remaining
1532 parameters are the kinematics of the two τ decays, denoted $\vec{a}_1 = (x_1, \phi_1, m_{\nu\nu,1})$ and
1533 $\vec{a}_2 = (x_2, \phi_2, m_{\nu\nu,2})$, and the four-momenta of the visible decay products with the
1534 measured values $\vec{y} = (p_1^{\text{vis}}, p_2^{\text{vis}})$.

1535 The likelihood f is the product of three likelihood functions. The first two likeli-
1536 hood functions model the decay parameters \vec{a}_1 and \vec{a}_2 of the two τ leptons. For lep-
1537 tonic decays, the likelihood function is modeled using matrix elements for τ decays,
1538 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
1539 hadronic τ decays, a model based on the two-body phase space is used and integrated
1540 over $m_{\text{vis}}^2/m_{\tau\tau}^2 \leq x \leq 1$. The third likelihood function quantifies the compatibility of
1541 a τ decay hypothesis with the reconstructed \vec{E}_T^{miss} in an event, assuming the neutrini-
1542 nos are the only source of missing transverse energy. The expected \vec{E}_T^{miss} resolution

1543 is represented by a covariant matrix, estimated on an event-by-event basis using a
1544 significance algorithm [63].

1545 5.2.2 “Classic SVFit” with matrix element

1546 Classic SVFit is an improved algorithm of the original “standalone” SVFit using the
1547 formalism of the matrix element (ME) method [61]. In the ME method, an estimate
1548 for the unknown model parameter Θ (here, the mass $m_{\tau\tau}$) is obtained by maximizing
1549 the probability density \mathcal{P} . The key ingredients of the probability density are the
1550 squared modulus of the matrix element $|\mathcal{M}(\mathbf{p}, \Theta)|^2$ and the transfer function $W(\mathbf{y}|\mathbf{p})$
1551 (probability density to observe the measured observables \mathbf{y} given the phase space
1552 point \mathbf{p}). The best estimate $m_{\tau\tau}$ is obtained by computing the probability density \mathcal{P}
1553 for a range of mass hypotheses and finding the value of $m_{\tau\tau}$ that maximizes \mathcal{P} .

1554 Distributions illustrating the performance of the classic matrix element SVFit
1555 algorithm are shown in Fig. 5.1 from [61], showing the di-tau mass after and before
1556 application of SVFit to recover energy lost to neutrinos. The SVFit algorithm is
1557 found to improve the sensitivity of the Standard Model $H \rightarrow \tau\tau$ analysis performed
1558 by CMS by about 30%, compared to performing the same analysis using only the
1559 visible mass m_{vis} .

1560 5.2.3 FastMTT: optimized SVFit

1561 FastMTT [64] is a further simplification to the matrix element method of Classic
1562 SVFit which has comparable performance but is about 100 times faster. FastMTT
1563 drops the matrix element component of the computation without significant impact
1564 on the final mass resolution, and simplifies the computation of the transfer functions.
1565 The opening angle of the τ decay products with respect to the initial τ momenta ap-
1566 proaches 0 for τ with high $\gamma = E_\tau/m_\tau$, with typical τ decays from the Z boson decays
1567 already satisfying this condition. In this collinear approximation, the dimensionality

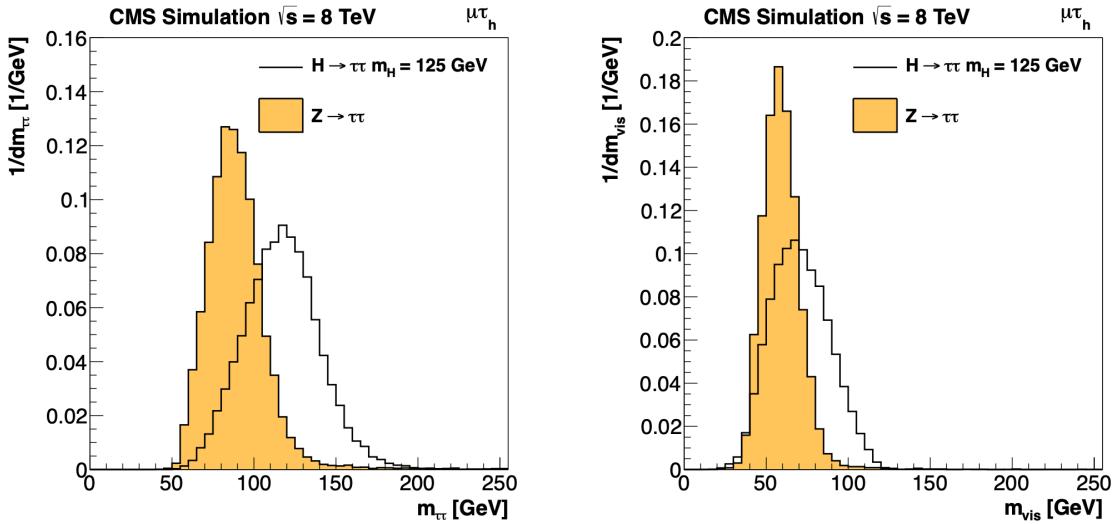


Figure 5.1: Distributions from [61], of $m_{\tau\tau}$ after reconstruction with the original SVFit algorithm (*left*), and before SVFit with only the visible tau decay products (*right*), for $H \rightarrow \tau\tau$ signal events of mass $m_H = 125$ GeV (*black line*) and the $Z/\gamma^* \rightarrow \tau\tau$ background (*orange, solid*), in the decay channel $\tau\tau \rightarrow \mu\tau_h$.

of the transfer function can be reduced in the computation of FastMTT, while still yielding similar results to Classic SVFit [64].

5.3 Corrections applied to simulation

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

1581 lations of the underlying physics process, or changing detector operating conditions
 1582 during data-taking. Event-level corrections change the shapes of the distributions of
 1583 all the physical observables.

1584 Uncertainties in scale factors and corrections are also sources of systematic errors
 1585 in the analysis, detailed in Chapter 8. Systematic uncertainties in the tau, muon, and
 1586 electron energy scales can shift the p_T of the leptons up or down, which can change
 1587 whether events pass or fail the offline p_T thresholds for the trigger paths described in
 1588 the previous section, i.e. change the number of events in the signal region.

1589 **5.3.1 Tau energy scale**

1590 An energy scale is applied to the transverse momentum p_T and mass of the hadronic
 1591 tau τ_h in the $\mu\tau_h$ and $e\tau_h$ channels, to correct for a deviation of the average recon-
 1592 structed τ_h energy from the generator-level energy of the visible τ_h decay products.
 1593 These correction factors are derived centrally [53], by fitting to events in $e\tau_h$ and $\mu\tau_h$
 1594 final states in Z/γ^* events separately for the h^\pm , $h^\pm\pi^0$, and $h^\pm h^\mp h^\pm$ decays. The
 1595 values used are shown in Table 5.1.

1596 When applying the energy scale to the τ_h , the 4-momentum of the missing trans-
 1597 verse energy (MET) is adjusted such that the total 4-momenta of the τ_h and the MET
 1598 remains unchanged [65].

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

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

1599 **5.3.2 Muon energy scale**

1600 An energy scale is applied to the p_T and mass of genuine muons from τ decays in the
1601 $e\mu$ and $\mu\tau_h$ channels [66]. The applied values are the same for MC and embedded
1602 samples and are shown in Table 5.2. Following the SM $H \rightarrow \tau\tau$ analysis, Rochester
1603 corrections are not applied, and instead prescriptions from [67] are followed.

Muon energy scale factor	
Eta range	Value for all years
$ \eta \in [0.0, 1.2)$	1.0 ± 0.004
$ \eta \in [1.2, 2.1)$	1.0 ± 0.009
$ \eta \in [2.1, 2.4)$	1.0 ± 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 [68] [67].

1604 **5.3.3 Electron energy scale**

1605 Corrections to the electron energy scale are applied to genuine e from τ decays, and
1606 are binned in two dimensions by electron p_T and η for barrel vs. endcap [69]. The
1607 scale factors are binned in p_T and η for MC samples: e.g. values for 2018 are shown
1608 in Fig. 5.2 from [70]. For embedded samples the electron energy scale is taken as
1609 only binned in η (Table 5.3).

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

Table 5.3: Energy scales and systematic errors applied to electrons in embedded samples, binned in the electron η , by data-taking year [71] [72] [73].

1610 **5.3.4 τ_h identification efficiency**

1611 The τ_h identification efficiency can differ in data and MC [65]. Recommended correc-
1612 tions are provided by the Tau POG, and we use the medium DeepTau vs. jet working

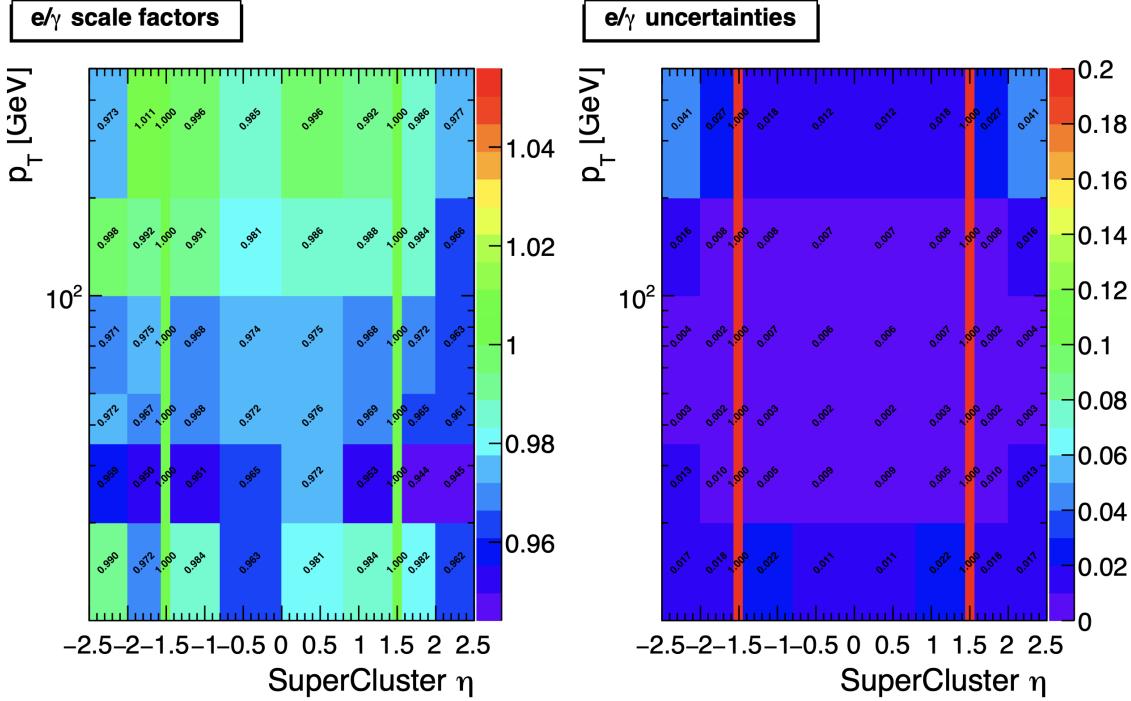


Figure 5.2: Electron/photon energy scale factors (*left*) and corresponding uncertainties (*right*) binned in the electron η and p_T , for the data-taking year 2018 [70].

₁₆₁₃ point values. The identification efficiency is measured in $Z \rightarrow \tau\tau$ events in the $\mu\tau_h$
₁₆₁₄ 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 [65].

₁₆₁₅ 5.3.5 Trigger efficiencies definition

₁₆₁₆ Scale factors are applied to correct for differences in trigger efficiencies between MC
₁₆₁₇ and embedded vs. data, with values taken from tools provided by the Standard Model
₁₆₁₈ $H \rightarrow \tau\tau$ working group which uses the same trigger paths [68]. In the following

1619 sections we review relevant trigger efficiencies in data, which form the basis of the
1620 trigger efficiency corrections applied to MC and embedded.

1621 **5.3.6 Tau trigger efficiencies**

1622 The efficiencies in data of the single- τ_h leg in $\mu\tau_h$, $e\tau_h$, and di- τ_h triggers is computed
1623 centrally per using a Tag and Probe (TnP) method [74] which is outlined here. In
1624 this method, $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$ are selected in data and a Drell-Yan simulated sample
1625 ($Z \rightarrow \ell\ell, \ell = e, \mu, \tau_h$) with high purity. Cuts are applied to reject events not in this
1626 final state, e.g. suppressing $Z \rightarrow \mu\mu$ by vetoing events with a single loose ID muon.
1627 An isolated muon candidate (the tag) with online $p_T > 27$ GeV and $|\eta| < 2.1$ is
1628 identified and matched to an offline μ . An offline τ_h candidate (the probe) is selected,
1629 which is separated from the tag μ , and has $p_T > 20$ GeV and $|\eta| < 2.1$. The probe
1630 τ_h must pass anti-muon and anti-electron discriminators to avoid fakes from muons
1631 and electrons, and must pass the medium MVA tau isolation to suppress fakes from
1632 QCD jets. The trigger efficiency in the TnP method is calculated as

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

1633 The efficiencies for the hadronic tau legs in the relevant channels of this analyses
1634 ($\mu\tau_h$ and $e\tau_h$) as a function of the offline tau p_T and η , are shown for data taken in
1635 2016, 2017, and 2018 in Figures 5.3a and 5.3b [74] [75]. In both figures, the different
1636 HLT thresholds and differences in the L1 seed result in higher efficiencies in 2016 and
1637 differences in shapes of the 2016 efficiencies compared to 2017 and 2018. The low
1638 pile-up in 2016 also leads to higher efficiencies in that year.

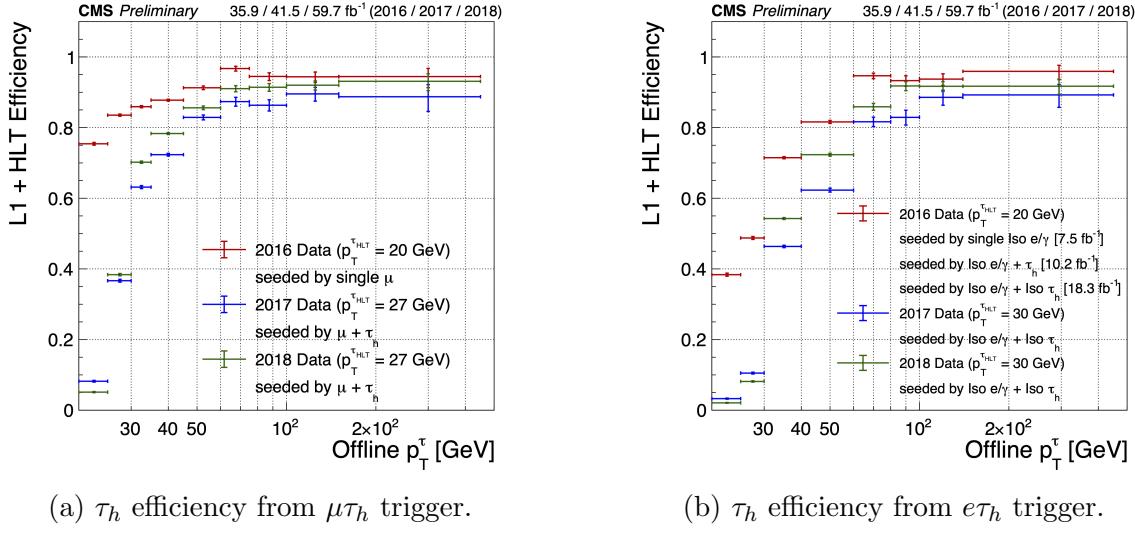
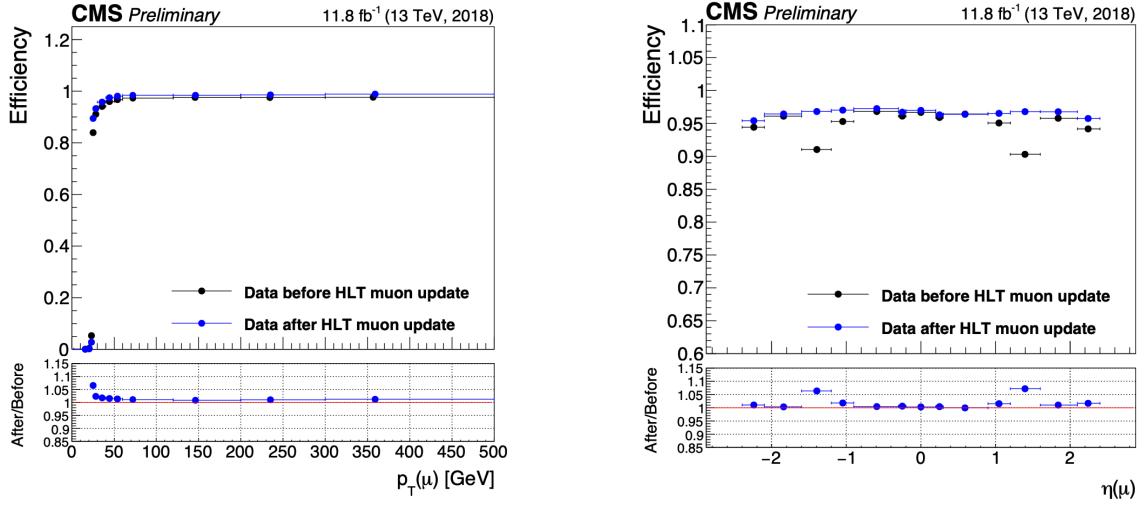


Figure 5.3: Hadronic tau leg efficiency of the cross-triggers for $\mu\tau_h$ (*left*) and $e\tau_h$ (*right*) triggers as a function of offline tau p_T for the years 2016 (red), 2017 (blue) and 2018 (green), from [75]. HLT p_T thresholds and L1 seeds are indicated in the legends.

5.3.7 Single muon trigger efficiencies

The efficiencies for the single isolated muon trigger with $p_T > 24$ GeV used in this analysis, is shown for the data-taking year 2018 in Fig. 5.4a as a function of the muon p_T and as a function of the muon $|\eta|$ in Fig. 5.4b from [76]. The data is split with respect to a HLT muon reconstruction update that was deployed on 15/05/2018. A small asymmetry in efficiencies between negative and positive η in Fig. 5.4b is due to disabled muon chambers (CSCs). The efficiencies shown are estimated using a Tag and Probe method using $Z \rightarrow \mu\mu$ events, with the tag being an offline muon with $p_T > 29$ GeV and $|\eta| < 2.4$ passing a tight ID criteria, and the probe is an online (L1) trigger object with $\Delta R < 0.3$ and passing tight ID and Particle Flow based isolation requirements with $p_T > 26$ GeV.



(a) Muon efficiency vs p_T for SingleMuon.

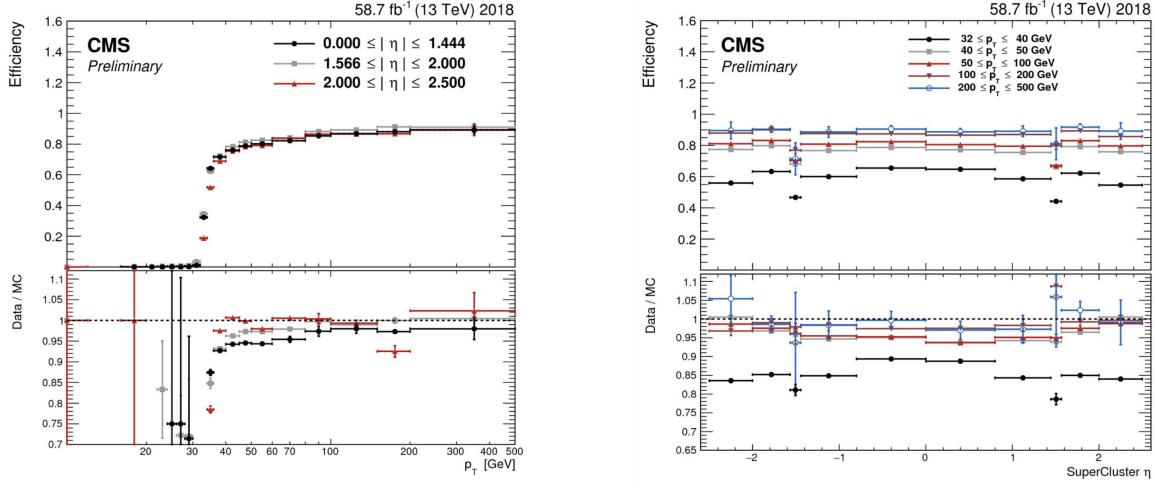
(b) Muon efficiency vs $|\eta|$ for SingleMuon.

Figure 5.4: Trigger efficiencies in data (*top panels*) and ratio of efficiencies after/before a HLT muon reconstruction update (*bottom panels*) for the muon in the isolated single muon trigger with threshold $p_T > 24$ GeV in the data-taking year 2018, as functions of the muon p_T (*left*) and muon $|\eta|$ (*right*). Only statistical errors are shown [76].

1650 5.3.8 Single electron trigger efficiencies

1651 The efficiencies in data, and the ratio between data and MC, of the single electron
 1652 HLT trigger with p_T threshold 32 GeV used in this analysis are shown for 2018,
 1653 as a function of the electron p_T in Fig. 5.5a and of the electron $|\eta|$ in Fig. 5.5b,
 1654 from [77]. In the Tag and Probe method used for the 2018 dataset, the tag is an
 1655 offline reconstructed electron with $|\eta| \leq 2.1$ and not in the barrel and endcap overlap
 1656 region, with $p_T > 35$ GeV with tight isolation and shower shape requirements, firing
 1657 the tag trigger. The probe is an offline reconstructed electron with $|\eta| \leq 2.5$ with
 1658 $E_T^{\text{ECAL}} > 5$ GeV with no extra identification criteria [77].

1659 The disagreement between data and MC, particularly at low transverse momentum,
 1660 is in part due to detector effects that are difficult to simulate, such as crys-
 1661 tal transparency losses in the ECAL and the evolution of dead regions in the pixel
 1662 tracker [77].



(a) Electron efficiency vs p_T for single electron.

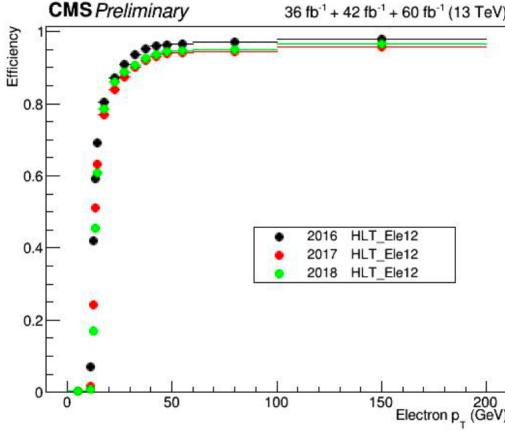
(b) Electron efficiency vs $|\eta|$ for single electron.

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*) [77]. 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.

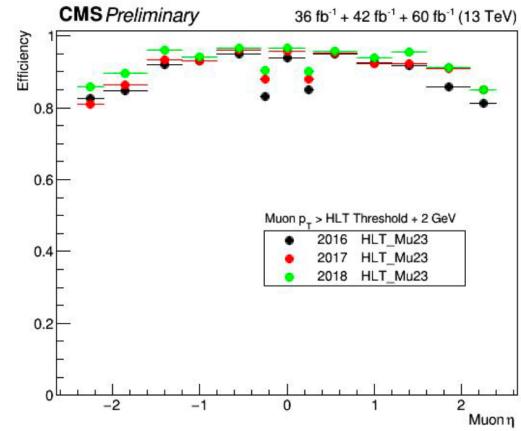
1663 5.3.9 $e\mu$ cross-trigger efficiencies

1664 The efficiencies of the electron and muons for the cross-trigger with leading muon
 1665 used in the $e\mu$ channel are shown for data in 2016, 2017, and 2018 in Figures 5.6a and
 1666 5.6b [78]. These efficiencies were measured centrally using a Tag and Probe in events
 1667 with Z to dileptons with the same flavor and opposite charge, where the tags are an
 1668 isolated muon or electron, and the probe (offline) candidate is required to satisfy the
 1669 same lepton selection as that of the tag candidate, be matched within $\Delta R < 0.1$ with
 1670 a corresponding online trigger object, and also to pass the cross-trigger. The trigger
 1671 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 .



(b) Muon efficiency vs. η .

Figure 5.6: Efficiencies of the electron leg vs. p_T (*left*) and the muon log vs. η (*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*) [78].

1672 5.3.10 Electrons and muons faking τ_h : energy scales

1673 Energy scales for electrons misidentified as hadronic tau decays (e faking τ_h) are
 1674 provided by the Tau POG, and were measured in the $e\tau_h$ channel with the visible
 1675 invariant mass of the electron and hadronic tau system [68]. This energy scale is
 1676 applied for τ_h with $p_T > 20$ GeV regardless of which DeepTau vs. electron working
 1677 point was used. Values for 2018 are shown in Table 5.5.

Electrons faking τ_h energy scale factor in 2018	
Reconstructed decay mode of the fake τ_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 τ_h [68].

1678 No nominal energy scale is applied for muons mis-reconstructed as τ_h , and the
 1679 uncertainty is treated as $\pm 1\%$ and uncorrelated in the reconstructed decay mode [68].

1680 **5.3.11 Electrons and muons faking τ_h : misidentification effi-**
 1681 **ciencies**

1682 Corrections on identification efficiencies are applied to genuine electrons and muons
 1683 misidentified as τ to account for differences in data and MC.

1684 The specific values depend on the vs. electron and vs. muon discriminator working
 1685 points used. For misidentified $\mu \rightarrow \tau_h$, the scale factors are split into different $|\eta|$
 1686 regions, determined by the CMS muon and tracker detector geometries, as shown in
 1687 Table 5.6 for 2018 [65].

Tau ID efficiency for DeepTau vs. muon WPs in 2018		
$ \eta $	Tight working point	VLoose working point
(0.0, 0.2)	0.767 ± 0.127	0.954 ± 0.069
(0.2, 0.6)	1.255 ± 0.258	1.009 ± 0.098
(0.6, 1.0)	0.902 ± 0.203	1.029 ± 0.075
(1.0, 1.45)	0.833 ± 0.415	0.928 ± 0.145
(1.45, 2.0)	4.436 ± 0.814	5.000 ± 0.377
(2.0, 2.53)	1.000 ± 0.000	1.000 ± 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 $|\eta|$ [65].

1688 For misidentified $e \rightarrow \tau_h$, the scale factors are split into barrel and endcap regions,
 1689 dictated by the ECAL detector geometry, as shown in Table 5.7 for 2018.

Tau ID efficiency for DeepTau vs. electron WPs in 2018		
$ \eta $	Tight working point	VLoose working point
(0.0, 0.73)	1.47 ± 0.27	0.95 ± 0.07
(0.73, 1.509)	1.509 ± 0.0	1.00 ± 0.0
(1.509, 1.929)	1.929 ± 0.2	0.86 ± 0.1
(1.929, 2.683)	2.683 ± 0.9	2.68 ± 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 $|\eta|$ [65].

1690 **5.3.12 Electron ID and tracking efficiency**

1691 Scale factors are applied to MC to correct for differences between MC and data in
 1692 the performance of electron identification (ID) and tracking.

1693 Electron and photon identification, as discussed earlier, use variables with good
 1694 signal vs. background discrimination power such as lateral shower shape and ratio
 1695 of energy deposited in the HCAL to energy deposited in the ECAL at the position
 1696 of the electron. The cut-based electron identification efficiencies in data and ratio of
 1697 efficiencies in data to MC are shown in Fig. 5.7a for the multivariate analysis (MVA)
 1698 identification working point.

1699 The tracking efficiencies in data and the data/MC ratio are shown in Fig. 5.7b
 1700 for the Gaussian-sum filter (GSF) tracking [79].

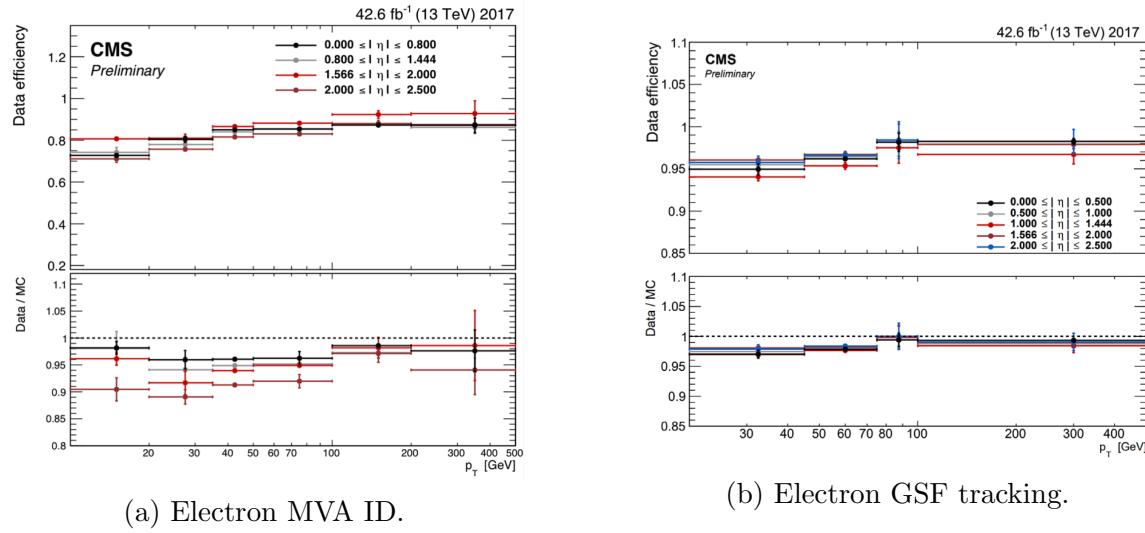


Figure 5.7: Efficiencies in data (*top panels*) and the ratio of efficiencies in data/MC (*bottom panels*), for the electron multivariate analysis (MVA) identification (*left*) and for the Gaussian-sum filter (GSF) tracking (*right*) [79]. Error bars represent statistical and systematic uncertainties.

1701 **5.3.13 Muon ID, isolation, and tracking efficiencies**

1702 Scale factors are applied to MC to correct for differences between MC and data in
 1703 the performance of muon identification, isolation, and tracking, as detailed below.

1704 The efficiencies for muon identification measured in 2015 data and MC simulation
 1705 are shown in Figures 5.8a and 5.8b for the loose ID and tight ID respectively [80]. The
 1706 loose ID is chosen such that efficiency exceeds 99% over the full η range, and the data
 1707 and simulation agree to within 1%. The tight ID is chosen such that efficiency varies
 1708 between 95% and 99% as a function of η , and the data and simulation agree to within
 1709 1-3%. The muon identification working point used in this analysis is the medium ID,
 1710 which has an efficiency of 98% for all η and an agreement within 1-2% [80].

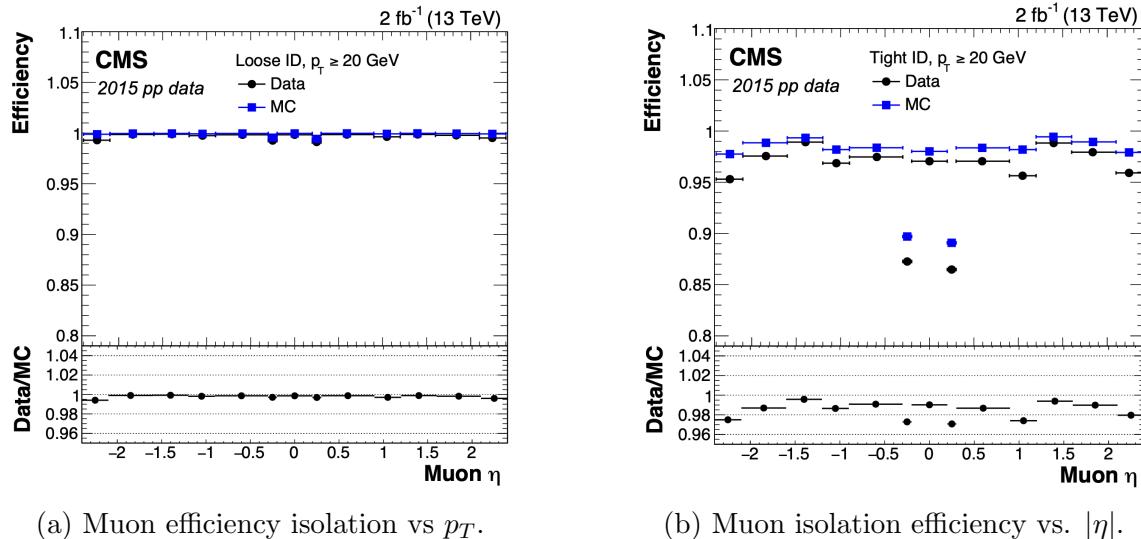
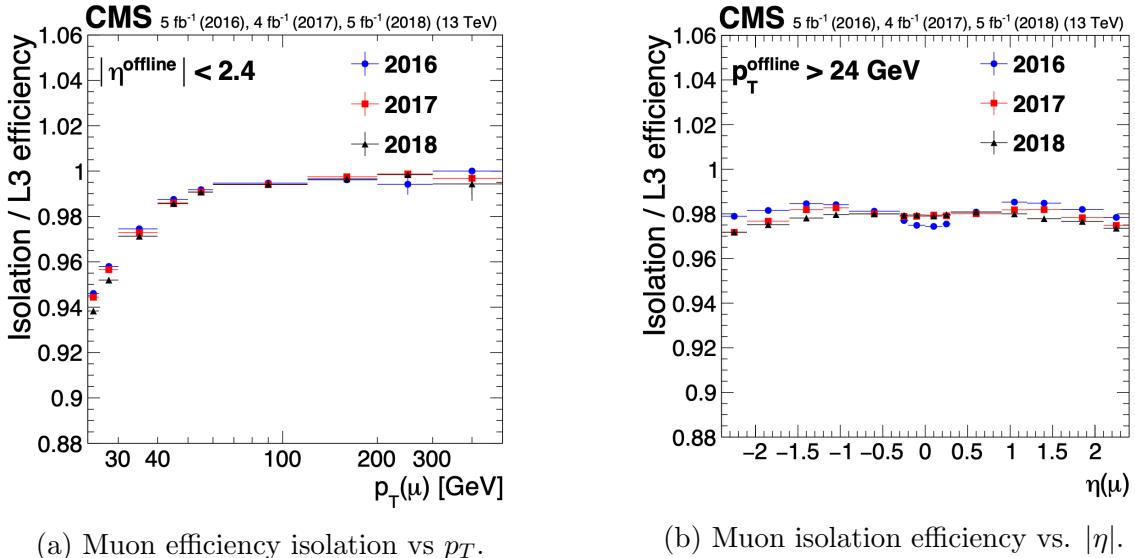


Figure 5.8: Muon identification efficiencies in 2015 data and MC as a function of the muon p_T for the loose ID (*left*) and tight ID (*right*) working points [80].

1711 The efficiencies in data for the muon isolation, as measured in Level-3 muons
 1712 (muons in one of the final stages of reconstruction in the HLT), as a function of the
 1713 muon p_T and $|\eta|$ are shown in Figures 5.9a and 5.9b [80]. The HLT muon reconstruc-
 1714 tion consists of two steps: Level-2 (L2), where the muon is reconstructed in the muon
 1715 subdetectors only, and Level-3 (L3) which is a global fit of tracker and muon hits (i.e.
 1716 the global muon reconstruction as described in Section 5.1.2) [81].

1717 The muon tracking efficiencies as a function of $|\eta|$ for standalone muons (i.e. tracks
 1718 from only the muon system, i.e. DT, CSC, and RPC, as discussed in Section 5.1.2),
 1719 is shown for data and simulated Drell-Yan samples in Fig. 5.10 [82].



(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*) [80].

1720 5.3.14 Recoil corrections

1721 In proton-proton collisions, W and Z bosons are predominantly produced through
1722 quark-antiquark annihilation. Higher-order processes can induce radiated quarks or
1723 gluons that recoil against the boson, imparting a non-zero transverse momentum to
1724 the boson [83]. Recoil corrections accounting for this effect are applied to samples
1725 with W+jets, Z+jets, and Higgs bosons [68]. The corrections are performed on the
1726 vectorial difference between the measured missing transverse momentum and the total
1727 transverse momentum of neutrinos originating from the decay of the W, Z, or Higgs
1728 boson. This vector is projected onto the axes parallel and orthogonal to the boson
1729 p_T . This vector, and the resulting correction to use, is measured in $Z \rightarrow \mu\mu$ events,
1730 since these events have leptonic recoil that do not contain neutrinos, allowing the
1731 4-vector of the Z boson to be measured precisely. The corrections are binned in
1732 generator-level p_T of the parent boson and also the number of jets in the event.

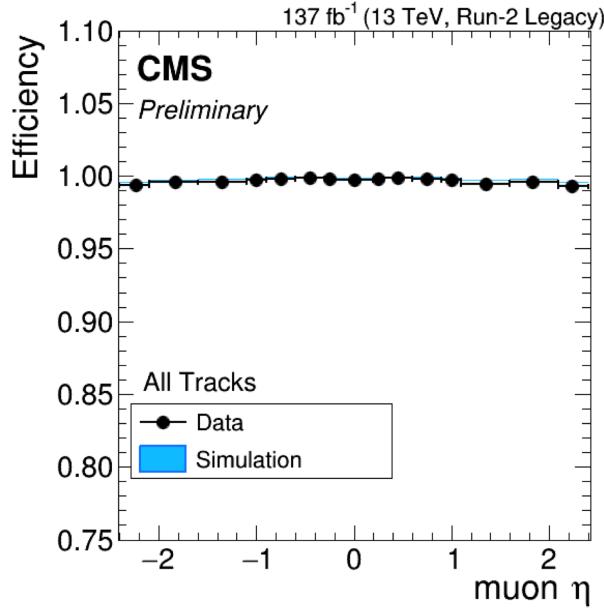


Figure 5.10: Muon tracking efficiencies as a function of $|\eta|$ for standalone muons in Run-2 data (*black*) and Drell-Yan MC simulation (*blue*) [82]. All Tracks refers to tracks which exploit the presence of muon candidates in the muon system to seed the track reconstruction in the inner tracker, in contrast to tracks that use tracker-only hits for seeding. Uncertainties shown are statistical.

5.3.15 Drell-Yan corrections

The Z boson transverse momentum distribution disagrees between leading-order (LO) simulations and data in a $Z \rightarrow \mu\mu$ control region with at least one b-tag jet [84]. Per-event weights derived by the 2016 data-only version of this analysis [84] are applied to $Z \rightarrow \tau\tau/\ell\ell$ events, as a function of the generator-level Z boson p_T to provide better matching of MC to data.

5.3.16 Pile-up reweighting

Reweighting is performed to rescale MC events to account for differences between MC and data, in the distribution of the pile-up (number of additional proton-proton interactions per bunch crossing). A tool for calculating the pile-up reweighting for the MC samples used is provided centrally by the Luminosity POG [85].

1744 **5.3.17 Pre-firing corrections**

1745 In 2016 and 2017 data-taking, a gradual timing shift of ECAL was not properly
1746 propagated to L1 trigger primitives (TPs), resulting in a large fraction of high η
1747 TPs being incorrectly associated with the previous bunch crossing. L1 trigger rules
1748 prevent two consecutive bunch crossings from firing, causing events to be rejected if
1749 significant ECAL energy was deposited in $2.0 < |\eta| < 3.0$. To account for this issue,
1750 MC simulations for 2016 and 2017 are corrected using an event-dependent weight.
1751 Embedded samples are not corrected [50].

1752 **5.3.18 Top p_T spectrum reweighing**

1753 In Run-1 and Run-2 it was observed that the p_T spectra of top quarks in $t\bar{t}$ data
1754 was significantly softer than those predicted by MC simulations [86]. Possible sources
1755 of this discrepancy are higher order QCD and/or electroweak corrections, and non-
1756 resonant production of $t\bar{t}$ -like final states. To account for this, corrections derived
1757 from Run-2 data by the Top Physics Analysis Group (PAG) are applied to the p_T
1758 of the top and anti-top quarks in MC simulations, computed as a function of their
1759 generator-level p_T [86].

1760 **5.3.19 B-tagging efficiency**

1761 In order to predict correct b-tagging discriminant distributions and event yields in
1762 data, the weight of selected MC events is reweighed according to recommendations by
1763 the BTV POG [87]. The reweighing depends on the jet p_T , η , and the b-tagging dis-
1764 criminant. In this method, there is no migration of events from one b-tag multiplicity
1765 bin to another.

1766 5.3.20 Jet energy resolution and jet energy smearing

1767 Calibration of jet energies, i.e. ensuring that the energy and momentum of the recon-
1768 structed jet matches that of the quark/gluon-initiated jet, is a challenging task due
1769 to time-dependent changes in the detector response and calibration and high pile-
1770 up [88] [89]. Jet calibration is done via jet energy corrections (JECs) applied to the
1771 p_T of jets in MC samples, accounting successively for the effects of pile-up, uniformity
1772 of the detector response, and residual data-simulation jet energy scale differences [90].
1773 Typical jet energy resolutions reported at $\sqrt{s} = 8$ TeV in the central rapidities are
1774 15-20% at 30 GeV and about 10% at 100 GeV [88]. Jet energy corrections are also
1775 propagated to the missing transverse energy.

1776 Measurements show that the jet energy resolution (JER) in data is worse than
1777 in simulation, and so the jets in MC need to be smeared to describe the data. JER
1778 corrections are applied after JEC on MC simulations, and adjust the width of the p_T
1779 distribution based on pile-up, jet size, and jet flavor [91]. Tools for applying JEC and
1780 JER are provided centrally by the JER Corrections group.

1781 Chapter 6

1782 Event selection

1783 This chapter describes how events in data and simulated samples are selected in the
1784 search for $h \rightarrow aa \rightarrow bb\tau\tau$. The event selection is motivated by optimization checks
1785 aimed at maximizing the final expected limit, and is also based on recommendations
1786 from CMS Physics Objects Groups. As described in the previous chapter, the tau
1787 lepton can decay to electrons (e), muons (μ), or hadronic states (τ_h). As a result,
1788 several different final states of the $\tau\tau$ system are possible, and are here referred to
1789 as “channels” since they are mutually exclusive. The three $\tau\tau$ final states studied in
1790 this analysis are muon and hadronic tau ($\mu\tau_h$), electron and hadronic tau ($e\tau_h$), and
1791 electron and muon ($e\mu$). The procedure for dividing events into these three channels
1792 begins with checking the High-Level Trigger paths passed by the events as detailed
1793 in Section 6.1. Events are further accepted or rejected based on criteria applied to
1794 the leptons in the event. These event selections are described for the $\mu\tau_h$ channel in
1795 Section 6.2, the $e\tau_h$ channel in Section 6.3, and the $e\mu$ channel in Section 6.4.

1796 6.1 General procedure for all channels

1797 For the search for $h \rightarrow aa \rightarrow bb\tau\tau$, three final states of the $\tau\tau$ system are considered:
1798 $\mu\tau_h$, $e\tau_h$, and $e\mu$. The $\tau_h\tau_h$ final state is not considered because signal events in the

1799 $\tau_h\tau_h$ channel would typically produce hadronic taus with momenta below data-taking
1800 trigger thresholds. In all three final states, events are required to have at least one
1801 b-tag jet passing the medium working point of the DeepFlavour tagger, with $p_T > 20$
1802 GeV, and $|\eta| < 2.4$. A second b-tag jet is not required because such a requirement
1803 would reduce signal acceptance by 80% compared to only requiring one b-tag jet.

1804 Events in MC samples are sorted into one of the three $\tau\tau$ channels if they pass the
1805 following trigger requirements and requirements on the offline reconstructed objects
1806 in the event, first checking the HLT paths for the $\mu\tau_h$ channel, then $e\tau_h$, and finally $e\mu$.
1807 The two leading leptons (e.g. muon and hadronic tau for the $\mu\tau_h$ channel) that were
1808 determined to have originated from the $\tau\tau$ decay, are called the $\tau\tau$ “legs”. For events
1809 in data and embedded samples, the HLT paths requirements for the corresponding
1810 channel are checked.

1811 After sorting events by HLT paths and identifying the leading tau legs in the offline
1812 reconstructed objects, the p_T of the offline objects is checked against the online trigger
1813 thresholds. Trigger matching is also performed, which checks the correspondence
1814 between each offline reconstructed object used in the analysis (e.g. a muon), and a
1815 trigger object in the HLT (e.g. a HLT muon). An offline object is considered to be
1816 matched, if it corresponds to a trigger object of the same object type, with $\Delta R < 0.5$.
1817 This matched trigger object is also required to pass the filter(s) of the HLT trigger.
1818 The trigger thresholds used for the $bb\mu\mu$ final state and the $bb\tau\tau$ final state (the focus
1819 of this work) are summarized in Tables 6.1.

1820 After checking the HLT paths and trigger objects in each channel, events are
1821 subject to further selection to ensure that they contain leptons and b-tag jet(s) of in-
1822 terest. These requirements are summarized in Table 6.2, and detailed in the following
1823 sections.

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

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

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

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

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

The muon is required to pass the medium identification (ID) working point [92], which is defined by the Muon POG as a loose muon (i.e. a Particle Flow muon that is either a global or a tracker muon - see Section 5.1.2) with additional requirements

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

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

1840 on track quality and muon quality. This identification criteria is designed to be
1841 highly efficiently for prompt muons and for muons from heavy quark decays. In
1842 addition to the ID, for prompt muons it is recommended to apply cuts on the impact
1843 parameter [92]: we apply $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$. A cut is applied on the
1844 muon relative isolation (defined in Section 5.1.2), to be less than 0.15 in a cone size of
1845 $\Delta R = 0.4$, which corresponds to the Tight Particle Flow isolation requirement [92].

1846 The τ_h is required to pass a cut on its impact parameter of $|\Delta(z)| < 0.2$. The τ_h
1847 is also required to pass the VLoose (Very Loose) DeepTau working point vs. elec-
1848 tron, the Tight DeepTau working point vs. muons, and the VVVLoose and Medium
1849 DeepTau working point vs. jets. Events with taus reconstructed in two of the decay
1850 modes (labeled 5 and 6) are rejected, since these decay modes are meant to recover
1851 3-prong taus, but are only recommended for use in analyses where the benefits in final
1852 significance outweigh the resulting increase in background [65]. Decays reconstructed
1853 with 2 prongs are not considered as they are only recommended for taus with a very
1854 high transverse momentum, where the prongs may overlap.

1855 For the estimation of the background from jets faking τ_h , which is described in Sec-
1856 tion 7.7, anti-isolated events are selected, by requiring events to pass all the selections
1857 described above, except failing the Medium DeepTau working point vs. jets.

1858 6.3 Event selection in the $e\tau_h$ channel

1859 The HLT trigger paths for the $e\tau_h$ channel are summarized in Tables 6.3, 6.4, and
1860 6.5. Similarly to the $\mu\tau_h$ channel, a single electron trigger is used if the electron has
1861 sufficiently high p_T in 2018 and 2017. For data taken in 2018 (2017), the OR of the
1862 single electron triggers with online p_T thresholds at 32 and 35 (27 and 32) GeV are
1863 used, with the corresponding offline electrons required to have p_T greater than 33
1864 (28) GeV. A $e + \tau_h$ cross-trigger is used for electrons with lower offline p_T between

1865 25 and 33 GeV (25 and 28 GeV). For the 2016 dataset, there is no cross trigger but
1866 only a single electron trigger with online p_T threshold at 25 GeV, which is used if the
1867 offline electron has p_T greater than 26 GeV.

1868 The electron and τ_h are required to have opposite charge and be separated by
1869 $\Delta R > 0.4$. The electron is required to be within $|\eta| < 2.3$ when no cross trigger is
1870 used, and $|\eta| < 2.1$ when the cross trigger is fired. The τ_h is required to have $|\eta| < 2.3$
1871 if no cross trigger is fired, and have $|\eta| < 2.1$ if the cross trigger is fired.

1872 The electron is required to have a relative isolation (same definition as in Section
1873 5.1.2) of less than 0.1 in a cone size of $\Delta R = 0.3$, which is the standard recommended
1874 cone size giving minimal pile-up dependence and reduced probability of other objects
1875 overlapping with the cone. The isolation quantity used includes an “effective area”
1876 (EA) correction to remove the effect of pile-up in the barrel and endcap parts of the
1877 detector [93]. The electron is also required to pass cuts on its impact parameter of
1878 $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$. It is also required to pass the non-isolated MVA
1879 working point corresponding to 90% efficiency. The electron’s number of missing hits,
1880 which are gaps in its trajectory through the inner tracker [93], must be less than or
1881 equal to 1. The electron must pass a conversion veto, which rejects electrons coming
1882 from photon conversions in the tracker, which should instead be reconstructed as part
1883 of the photon [93].

1884 The impact parameter cut for the τ_h is $|\Delta(z)| < 0.2$. In contrast to the $\mu\tau_h$ event
1885 selection, the vs. electron and vs. muon DeepTau working points are flipped, to
1886 reject muons faking the τ_h leg. The τ_h is required to pass the Tight DeepTau working
1887 point vs. electrons, the VLoose DeepTau working point vs. muons, and the Medium
1888 DeepTau working point vs. jets.

1889 As in the $\mu\tau_h$ channel, for the estimation of the background from jets faking τ_h ,
1890 which is described in Section 7.7, anti-isolated events are selected by requiring events
1891 to pass all the selections described above, except failing the Medium DeepTau working

1892 point vs. jets.

1893 6.4 Event selection in the $e\mu$ channel

1894 The HLT trigger paths for the $e\mu$ channel are summarized in Tables 6.3, 6.4, and
1895 6.5. Events are selected with the logical OR of two $e + \mu$ cross triggers, where either
1896 the electron or muon can have larger p_T : (1) leading electron, where the electron has
1897 online $p_T > 23$ GeV and muon has online $p_T > 8$ GeV, or (2) leading muon, where
1898 electron has online $p_T > 12$ GeV and muon has online $p_T > 23$ GeV.

1899 The leading and sub-leading leptons are required to have an offline p_T greater
1900 than 1 GeV above the online threshold (i.e. $p_T > 24$ GeV). If the sub-leading lepton
1901 is the electron, the offline p_T threshold is 1 GeV above the online threshold ($p_T > 13$
1902 GeV), but if it is a muon, the offline p_T threshold is required to be at least 5 GeV
1903 greater than the online threshold (i.e. $p_T > 13$ GeV). This is because of poor data
1904 and simulation agreement for low- p_T muons with p_T between 9 GeV and 13 GeV, and
1905 the higher probability of mis-identifying jets as muons at lower p_T . With no effect on
1906 the expected limits, the offline p_T threshold for muons is raised to 13 GeV instead of
1907 9 GeV, even though it may lead to loss in signal acceptance. Both the electron and
1908 muon are required to have $|\eta| < 2.4$.

1909 The electron and muon are required to have opposite charge and be separated
1910 by $\Delta R > 0.3$ (note the decreased separation requirement compared to the other
1911 two channels). The electron is required to pass the non-isolated MVA identification
1912 working point corresponding to 90% efficiency, and to have a relative isolation less
1913 than 0.1 for a cone size of $\Delta R = 0.3$ with the EA pile-up subtraction correction.
1914 The electron must have one or fewer missing hits and pass the conversion veto (both
1915 described previously in Section 6.3).

1916 The muon is required to pass the medium identification working point (described

1917 earlier in 6.2), and to have a relative isolation less than 0.15 for a cone size of $\Delta R =$
1918 0.4. The muon impact parameter is required to have $|\Delta(z)| > 0.2$ and $|\Delta(xy)| < 0.045$.

1919 For the QCD multijet background estimation described in Section 7.8, the same-
1920 sign region is selected by requiring all the above selections, except the legs are required
1921 to have the same electric charge rather than opposite.

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

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

1922 6.5 Extra lepton vetoes in all channels

1923 Events containing a third lepton (electron or muon) that is neither of the leading $\tau\tau$
1924 legs are rejected, and events with di-muons and di-electrons are vetoed, with criteria
1925 taken from the Standard Model $H \rightarrow \tau\tau$ working group [68]. These vetoes on extra
1926 leptons also ensure orthogonality of events to analyses such as the $bb\mu\mu$ final state,
1927 whose results are combined with this $bb\tau\tau$ final state as described in Section 10.2.

1928 The event is vetoed if a third electron is found with the following properties:
1929 $p_T > 10$ GeV, $|\eta| < 2.5$, impact parameter $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$, passing

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

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

1930 non-isolation MVA identification with 90% efficiency, conversion veto, ≤ 1 missing
 1931 hits, and relative isolation < 0.3 with cone size $\Delta R = 0.3$. The event is also vetoed if
 1932 a third muon is found with the following properties: $p_T > 10$ GeV, $|\eta| < 2.4$, impact
 1933 parameter $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$, medium ID, and isolation < 0.3 with
 1934 cone size $\Delta R = 0.4$.

1935 A di-muon veto is applied, which rejects events containing a pair of muons with
 1936 opposite charge and separation of $\Delta R > 0.15$, that both pass the following selections:
 1937 $p_T > 15$ GeV, $|\eta| < 2.4$, flag for global muons, flag for tracker muon, flag for Particle
 1938 Flow muon, $|\Delta(z)| < 0.2$, $|\Delta(xy)| < 0.045$, and isolation < 0.3 with cone size $\Delta R =$
 1939 0.4. A similar di-electron veto is applied to reject events containing a pair of electrons
 1940 with opposite charge and separation of $\Delta R > 0.15$, that both pass the following
 1941 selections: $p_T > 15$ GeV, $|\eta| < 2.5$, a dedicated electron ID (cut-based) for vetoing
 1942 third leptons, $|\Delta(z)| < 0.2$, $|\Delta(xy)| < 0.045$, with pile-up corrected relative isolation
 1943 < 0.3 with cone size $\Delta R = 0.3$.

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

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

1944 Chapter 7

1945 Background estimation

1946 This section describes methods used to estimate backgrounds from Standard Model
1947 processes in the search for $h \rightarrow aa \rightarrow bb\tau\tau$. The background contributions directly
1948 taken from MC are described in Sections 7.1 to 7.6. Section 7.7 describes the data-
1949 driven method for estimating backgrounds from jets faking hadronic tau decays (jet
1950 $\rightarrow \tau_h$), which is used in the $\mu\tau_h$ and $e\tau_h$ channels. Section 7.8 describes the data-driven
1951 method for estimating background from quantum chromodynamic (QCD) processes
1952 in the $e\mu$ channel.

1953 7.1 Z+jets

1954 A major source of background for $\tau\tau$ analyses is the Drell-Yan (DY) process (Z+jets).
1955 The Z boson decays to $\tau\tau/\mu\mu/ee$ with equal probability of 3.4% each, with the domi-
1956 nant decay modes being to hadrons (around 70%) and neutrinos (invisible) (20%) [26].
1957 The Drell-Yan contribution with genuine taus, $Z \rightarrow \tau\tau$, is estimated using embed-
1958 ded samples, described in Section 4.3. To avoid double-counting between embedded
1959 and MC samples, in all MC samples, events with legs that originated from genuine τ
1960 are discarded.

1961 The other decays of the Z, $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$, are estimated from MC simulation,

1962 and are hereafter referred to as simply the Drell-Yan background. These MC samples
1963 are generated to leading order (LO) with different numbers of jets (jet multiplicity) in
1964 the matrix element: Z+1 jet, Z+2jets, Z+3 jets, Z+4 jets, and inclusive Z+jets. The
1965 cross-sections of the samples with ≥ 1 jets are normalized to next-to-NLO (NNLO)
1966 in QCD. For the inclusive Drell-Yan sample, two samples are used with different
1967 thresholds for the di-lepton invariant mass ($m_{\ell\ell}$) at the generator level: one with
1968 $m_{\ell\ell} > 50$ GeV and the other with $10 < m_{\ell\ell} < 50$.

1969 **7.2 W+jets**

1970 The dominant W boson decay modes are to hadrons (67.4%), $e + \nu_e$ (10.7%), $\mu + \nu_\mu$
1971 (10.6%), and $\tau + \nu_\tau$ (11.4%) [26]. The W+jets background is estimated from MC
1972 simulation. Similarly to the Z+jets, the W+jets samples are generated with different
1973 jet multiplicities in the matrix element. LO samples are used for greater statistics
1974 and are normalized to NNLO cross sections.

1975 **7.3 $t\bar{t}$ + jets**

1976 In hadron collisions, top quarks are produced singly with the weak interaction, or in
1977 pairs via the strong interaction, with interference between these leading-order pro-
1978 cesses possible in higher orders of the perturbation theory. The top quark is the
1979 heaviest fermion in the Standard Model and has a short lifetime ($\sim 10^{-25}$ s), decay-
1980 ing without hadronization into a bottom quark and a W boson [26], with the decay
1981 modes of the W boson as listed in the previous section. With two top quarks, the
1982 final states of the two resulting W bosons can be described as fully leptonic, semilep-
1983 tonic, and fully hadronic. These three final states are modeled separately with MC
1984 simulation in 2018 and 2017, while for 2016 the sample used is inclusive.

7.4 Single top

There are three main production modes of the single top in pp collisions [94]: the exchange of a virtual W boson (t channel), the production and decay of a virtual W boson (s channel), and the associated production of a top quark and W boson (tW , or W-associated) channel. As the s channel process is rare and only 3% of the total production, the dominant production mode of the t -channel and the tW production are considered and modeled with MC.

7.5 Diboson

In pp collisions, the production of dibosons (pairs of electroweak gauge bosons, i.e. WW, WZ, and ZZ) is dominated by quark-antiquark annihilation, with a small contribution from gluon-gluon interaction [95]. MC is used to model the pair production and decays of VV to $2\ell 2\nu$, WZ to $2q 2\ell$ and $3\ell\nu$, and ZZ to 4ℓ and $2q 2\ell$ (q being quarks and ℓ being leptons).

7.6 Standard Model Higgs

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

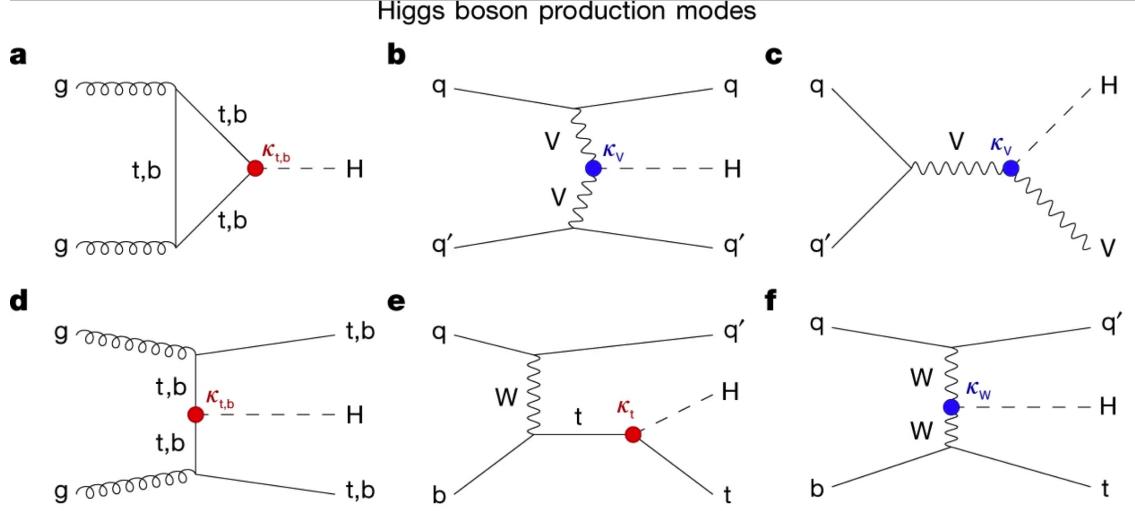


Figure 7.1: Leading-order Feynman diagrams of Higgs production from [96], in ggH (a) and vector boson fusion (VBF; b), associated production with a W or Z (V) boson (VH; c), associated production with a top or bottom quark pair (ttH or bbH); d, and associated production with a single top quark (tH; e, f).

2006 7.7 Jet faking τ_h

2007 Events with a jet mis-reconstructed as the hadronic tau leg τ_h are a major source of
 2008 background in the $\mu\tau_h$ and $e\tau_h$ channels. The main processes contributing to jet $\rightarrow \tau_h$
 2009 events are QCD multijet, W+jets, and $t\bar{t}$ production. These events are estimated
 2010 using a data-driven method adapted from past analyses [50] [84]. This background
 2011 includes contributions from W+jets, QCD multijets, and $t\bar{t}$ +jets. To estimate this
 2012 background, a sideband region is constructed, where events are required to pass all
 2013 baseline $\mu\tau_h/e\tau_h$ selection criteria, but fail the τ_h isolation criteria. The events in
 2014 this sideband region are reweighed with a factor $f/(1 - f)$, where f is the probability
 2015 for a jet to be misidentified as a τ_h . The jet $\rightarrow \tau_h$ background is the anti-isolated,
 2016 reweighed MC and embedded events subtracted from the anti-isolated, reweighted
 2017 data events.

2018 The fake factor is measured in $Z \rightarrow \mu\mu + \text{jets}$ events in data in the $\mu\mu\tau_h$ final
 2019 state, as any reconstructed τ_h in these events must originate from a jet. The two
 2020 muons are required to be isolated (< 0.15), have opposite electric charge, and have

2021 an invariant mass between 76 and 106 GeV (close to the Z mass). These events are
2022 selected with a double muon trigger, with the leading muon having offline $p_T > 20$
2023 GeV and the subleading muon $p_T > 10$ GeV. Simulated diboson (ZZ and WZ) events
2024 are subtracted to avoid contamination from events with real τ_h . The denominator of
2025 the fake rate corresponds to fake taus passing the VVVLoose working point of the
2026 discriminator vs. jets, while the numerator corresponds to those passing the Medium
2027 working point, i.e. $f = N_{\text{jet passing tight}} / N_{\text{jet passing loose}}$.

2028 f is measured as a function of the τ_h transverse momentum and is 8% - 10% in
2029 each of the data-taking years. f is derived separately for the $\mu\tau_h$ and $e\tau_h$ channels
2030 because the channels use different anti-lepton identification working points.

2031 7.8 QCD multijet background

2032 In the $e\mu$ channel, events with jets faking electrons or muons originating from QCD
2033 multijet, is estimated from data events with the same baseline selection as in the
2034 signal region, except with same-signed (SS) charged $e + \mu$, ensuring orthogonality
2035 with the signal region which requires opposite-sign (OS) $e\mu$ pairs. All same-sign MC
2036 events (both events with real and fake $e + \mu$) are subtracted from same-sign data
2037 events to remove contamination from other backgrounds. i.e. $\text{QCD}_{\text{SS}} = \text{Data}_{\text{SS}} -$
2038 MC_{SS} .

2039 Three scale factors are applied to the QCD_{SS} events to compute the QCD multijet
2040 background [84] [40]:

- 2041 • *OS-to-SS scale factor*: This scales the SS QCD to the OS region, and is mea-
2042 sured from an orthogonal region with an isolated electron and an anti-isolated
2043 muon. Only the muon is chosen to be anti-isolated because this scale factor was
2044 observed to depend more strongly on electron isolation than that of the muon.
2045 This scale factor is treated as a function of the ΔR separation of the trajectories

2046 of the electron and muon, and is measured separately for events with 0 jets, 1,
2047 jet, and greater than 1 jet.

- 2048 • *2D closure correction for the lepton p_T :* This factor accounts for subleading
2049 dependencies of the first scale factor on the p_T of the two leptons. A 2D weight
2050 is derived in a similar fashion, as a ratio of QCD_{OS} events to QCD_{SS} events,
2051 but parameterized by both electron and muon p_T , where the SS events have the
2052 previous scale factor applied.
- 2053 • *Isolation correction for the muon:* The third and final factor is an isolation
2054 correction, which is a bias correction to account for the fact that the fake
2055 factor was determined for less-isolated muons. This factor is obtained as the
2056 ratio of the OS-to-SS scale factors measured in two other control regions: (1)
2057 events where the electron is anti-isolated ($0.15 < \text{iso} < 0.5$) and the muon is
2058 isolated, and (2) events where both leptons are anti-isolated.

2059 **Chapter 8**

2060 **Systematic uncertainties**

2061 Uncertainties in the measurement of a physical observable can be statistical or sys-
2062 tematic in nature. Statistical uncertainties originate from limitations on the number
2063 of events and experiments that can be performed. Systematic uncertainties arise
2064 from the dependence of the physical observable on quantities whose exact values are
2065 unknown and which can only be modeled imperfectly.

2066 The handling of systematic uncertainties is separated into normalization uncer-
2067 tainties (those that affect the total yield of a variables' distribution) and shape un-
2068 certainties (those that shift the distribution of events). Normalization uncertainties
2069 are expressed as multiplicative factors, while shape uncertainties are represented as
2070 up and down shifts of a variable's distribution.

2071 Up/down shifts of shape uncertainties can change the number of background
2072 events in a distribution. For instance, hadronic taus receive corrections from the
2073 nominal tau energy scale, with the nominal, up, and down energy scales provided
2074 centrally by CMS. For the $\mu\tau_h$ channel, an event could have a τ_h with p_T just below
2075 the offline threshold of 20 GeV (for instance, 19.5 GeV), so in the nominal distribution
2076 of $m_{\tau\tau}$ (or any other variable for this channel), the event is excluded. However, when
2077 we build our distributions with the tau energy scale “up” shift, the energy of this τ_h

2078 may be scaled up to, say, 20.5 GeV, and now the event passes the offline p_T threshold
2079 for the single muon trigger, leading to the event’s inclusion in the distributions made
2080 with the tau energy scale “up” shift.

2081 In evaluating the up and down shifts of a specific source of uncertainty, all other
2082 corrections and scale factors are held at their nominal values, and the full chain of
2083 object and event selection and event categorization is performed to obtain the observ-
2084 able distributions. Any “downstream” variables that depend on the shifted variable,
2085 e.g. the invariant di-tau mass $m_{\tau\tau}$, must be computed for the nominal case, and then
2086 re-computed separately for each up and down shift of the tau legs’ energy scale. The
2087 objective of this process is to quantify the effect of a single source of uncertainty on
2088 the resulting observable distributions. Each scale factor and correction described in
2089 Section 5.3 has an associated uncertainty. The binning of the uncertainties follows
2090 that of the nominal scale factor value.

2091 Sections 8.1 to 8.5 describe uncertainties associated with physics objects, and
2092 Sections 8.6 and 8.7 describe uncertainties associated with sample-level effects. The
2093 pulls and impacts for the top sixty most important systematics are shown in Section
2094 8.8.

2095 8.1 Uncertainties in the lepton energy scales

2096 The uncertainties in the tau energy scales [65] are binned by the tau decay mode and
2097 are taken as shape uncertainties treated as uncorrelated across the tau decay modes
2098 and years. Same as with the application of the nominal scale factor, when applying
2099 the up or down shifts, the missing transverse energy (p_T^{miss}) of the event is adjusted
2100 so that the 4-vector sum of the tau p_T^{miss} is unchanged.

2101 The uncertainties in the muon energy scale [66] are 0.4% for $|\eta| < 1.2$, 0.9% for
2102 $1.2 < |\eta| < 2.1$, and 2.7% for $2.1 < |\eta| < 2.4$, and are treated as shape uncertainties,

2103 fully uncorrelated between embedded and MC samples.

2104 The uncertainties in the electron energy scale [69] in MC are binned in the electron
2105 $|\eta|$ and p_T , and are shown in Fig. 5.2. The uncertainties range from 0.5% to 2.2% in
2106 the barrel, and 0.3% to 4.1% in the endcap, across the p_T range. The uncertainties
2107 for the embedded sample are binned only in $|\eta|$ and are on the order of 0.5% and
2108 1.25% for the barrel and endcap [73].

2109 There are also uncertainties in the energy scales for electrons and muons misiden-
2110 tified as τ_h . The uncertainty for muons misidentified as τ_h is 1% [65]. For electrons
2111 misidentified as τ_h , the uncertainty is binned in barrel/endcap η and by 1-prong and
2112 1-prong + π_0 decays. The probability for e/μ faking a 3-prong decay mode is much
2113 lower.

2114 8.2 Uncertainties from other lepton corrections

2115 Uncertainties associated with the τ_h identification efficiencies are treated as shapes,
2116 uncorrelated across the seven p_T bins and years. The shape uncertainties in the em-
2117 bedded samples are taken as 50% correlated with those of the MC samples. The
2118 uncertainties on electron and muon identification efficiencies are taken as normaliza-
2119 tion uncertainties of 2% each, with a 50% correlation between embedded and MC
2120 samples.

2121 In the $e\tau_h$ channel, there is an additional uncertainty for the vs. jet discrimination
2122 efficiency [65], because the analysis uses a looser anti-lepton working point (VLoose
2123 WP) than the working points used in the measurement of the efficiency (namely,
2124 VLoose WP vs e, and Tight WP vs mu). For nominal $\tau_h p_T < 100$ GeV, an additional
2125 uncertainty of 3% (5%) is used in MC (embedded), and for high p_T an uncertainty of
2126 15% is used for both.

2127 The uncertainties in trigger efficiencies are taken as shapes [65]. In the $e\tau_h$ and $\mu\tau_h$

2128 channels, there are uncertainties for the single and cross lepton triggers, and in the
2129 $e\mu$ channel there is one uncertainty each for the two $e + \mu$ triggers, and one combined
2130 uncertainty since their trigger phase spaces are not mutually exclusive.

2131 **8.3 Uncertainties from jet energy scale and reso- 2132 lution**

2133 The jet energy scale uncertainties are taken as shape uncertainties: there are eleven
2134 in total, with seven correlated across years (labeled “Year” below) and the remainder
2135 uncorrelated across years. They affect the b-tag jet p_T and mass, and hence the
2136 missing transverse energy p_T^{miss} . The shifts are propagated through the b-tagging
2137 scale factor calculation and b-tag jet counting.

2138 The uncertainties in the jet energy correction and resolution [88] [97] are as follows:

- 2139 • *Absolute, AbsoluteYear*: flat absolute scale uncertainties.
- 2140 • *BBEC1, BBEC1Year*: for sub-detector regions, with barrel “BB” in $|\eta| < 1.3$
2141 and endcap region 1 “EC1”: $1.3 < |\eta| < 2.5$.
- 2142 • *EC2, EC2 year*: for sub-detector regions, with endcap region 2 “EC2” in $2.5 <$
2143 $|\eta| < 3.0$.
- 2144 • *HF, HF year*: for sub-detector regions, with hadron forward “HF” in $|\eta| > 3$.
- 2145 • *FlavorQCD*: for uncertainty in jet flavor (uds/c/b-quark and gluon) estimates
2146 based on comparing Pythia and Herwig (different MC generator) predictions.
- 2147 • *RelativeBal*: account for difference between log-linear fits of the two methods
2148 used to study the jet energy response: MPF (missing transverse momentum
2149 projection fraction) and p_T balance.

- 2150 • *RelativeSample*: account for η -dependent uncertainty due to a difference be-
 2151 tween relative residuals, observed with dijet and Z+jets in Run D of 2018 data.
- 2152 • *JetResolution*: uncertainty in the jet energy resolution.

2153 8.4 Uncertainties from b-tagging scale factors

2154 The b-tagging scale factor has its own set of associated uncertainties (not to be
 2155 confused with shifts in the b-tagging scale factor due to the propagation of the jet
 2156 energy scale uncertainties described in the previous section 8.3). They are:

- 2157 • *hf*: contamination from heavy flavor (b+c) jets in the light flavor region.
- 2158 • *hfstats1, hfstats2*: linear and quadratic statistical fluctuations from b-flavor jets.
- 2159 • *lf*: contamination from light flavor (udsg+c jets) in the heavy flavor region.
- 2160 • *lfstats1, lfstats2*: linear and quadratic statistical fluctuations from udsg jets.
- 2161 • *cferr, cferr2*: uncertainty for charm jets.

2162 The variations for “lf, hf, hfstats1/2, lfstats1/2” are applied to both b and udsg jets.
 2163 For c-flavor jets, only “cferr1/2” is applied.

2164 8.5 Uncertainties from MET

2165 Samples where recoil corrections were applied (Z+jets, W+jets, and Standard Model
 2166 Higgs, as described in Section 5.3) have uncertainties from the response and resolution
 2167 of the hadronic recoil against the leptonic system. These are each binned in jet
 2168 multiplicity.

2169 8.6 Uncertainties associated with samples used

2170 Normalization uncertainties related to the samples used are:

- 2171 • *Cross-section uncertainties*: $\sigma(t\bar{t})$: 4.2%, $\sigma(\text{diboson})$: 5%, $\sigma(\text{single top})$: 5%,
2172 $\sigma(\text{ggH})$: 3.2%, $\sigma(\text{qqH})$: 2.1%, $\sigma(\text{WH})$: 1.9%, $\sigma(\text{ZH})$: 1.3%, $\sigma(\text{ttH})$: 3.6%
- 2173 • *Uncertainties in QCD renormalization scale*: QCD scale(qqH): +0.43%-0.33%,
2174 QCD scale(WH): +0.5%-0.7%, QCD scale(ttH): +5.8%-9.2%
- 2175 • *Branching ratio uncertainties*: $\text{BR}(\text{H} \rightarrow \tau\tau)$: 1.8%, and $\text{BR}(\text{H} \rightarrow \text{WW})$: 1.5%.
- 2176 • *Normalization uncertainties*: 2% for Drell-Yan, 4\$ for embedded, 20% pre-fit
2177 for the QCD multijet background in the $e\mu$ channel, 20% pre-fit for the jet
2178 faking background.

2179 The $t\bar{t}$ process has additional acceptance uncertainties from QCD scale variation
2180 and parton shower uncertainties [98]. Parton shower uncertainties originate from
2181 the modeling of perturbative and non-perturbative QCD effects handled in parton
2182 shower MC generators. The scale variations are determined from the envelope of the
2183 6 provided shapes due to variations in the factorization scale, renormalization scale,
2184 and their combined variation [98].

2185 The uncertainty in the Z p_T reweighting in Drell-Yan samples is taken as a shape
2186 uncertainty and the up and down values are 0.9 and 1.1 times the nominal reweighting.
2187 This 10% uncertainty is sufficient to cover uncertainties in the weights derived from
2188 the discrepancies between LO simulations and data in the di-muon mass in $Z \rightarrow \mu\mu$
2189 events.

2190 The weight applied to anti-isolated events in the $\mu\tau_h$ and $e\tau_h$ channels to estimate
2191 the background from jets faking τ_h , has shape uncertainties covering uncertainties in
2192 the derivation of the weight. There are six shape uncertainties corresponding to the
2193 binning of the fake rate in the τ_h transverse momentum. For the weight applied to

2194 scale up anti-isolated events in cross-trigger regions, 20% of the nominal weight is
2195 taken as a shape uncertainty.

2196 8.7 Other uncertainties

2197 A 3.6% yield uncertainty in the signal is used to cover uncertainties in the parton
2198 distribution functions (PDFs), knowledge of the α_s (fine structure constant), and
2199 QCD scale. The size of these uncertainties was estimated by a different analysis
2200 searching for two light scalars decaying to four muons, which compared the PDFs
2201 from different model libraries using recommendations from the PDF4LHC Working
2202 Group [99] [100].

2203 Uncertainties in the luminosity measurements can originate from uncertainties
2204 in the luminosity calibration in the van de Meer scan procedure and from detector
2205 operations [43]. Some effects are fully uncorrelated (e.g. if the systematic error is
2206 limited by the statistical uncertainty in the calibration scans taken independently in
2207 each year), and some are correlated, for example in the 2017 and 2018 measurements
2208 which used a method with the same systematic bias. The luminosity normalization
2209 uncertainties are applied all MC samples, divided into those uncorrelated across years
2210 (0.26% for 2016, 0.60% for 2017, and 0.65% for 2018), one correlated between 2017
2211 and 2018 (0.27%), and one correlated between all three years (1.30%) [41] [42] [43] [85].

2212 8.8 Pulls and impacts

2213 The top impacts and pulls computed for the combination of all channels and years is
2214 shown in Fig. 8.1. The top impacts are related to uncertainty in the signal sample and
2215 cross-section of the $t\bar{t}$ cross-section, and also the yields of the jet faking τ_h background,
2216 which is a major background in all channels and expected to be constrained due to
2217 the yield uncertainty which is taken to be 20% pre-fit.

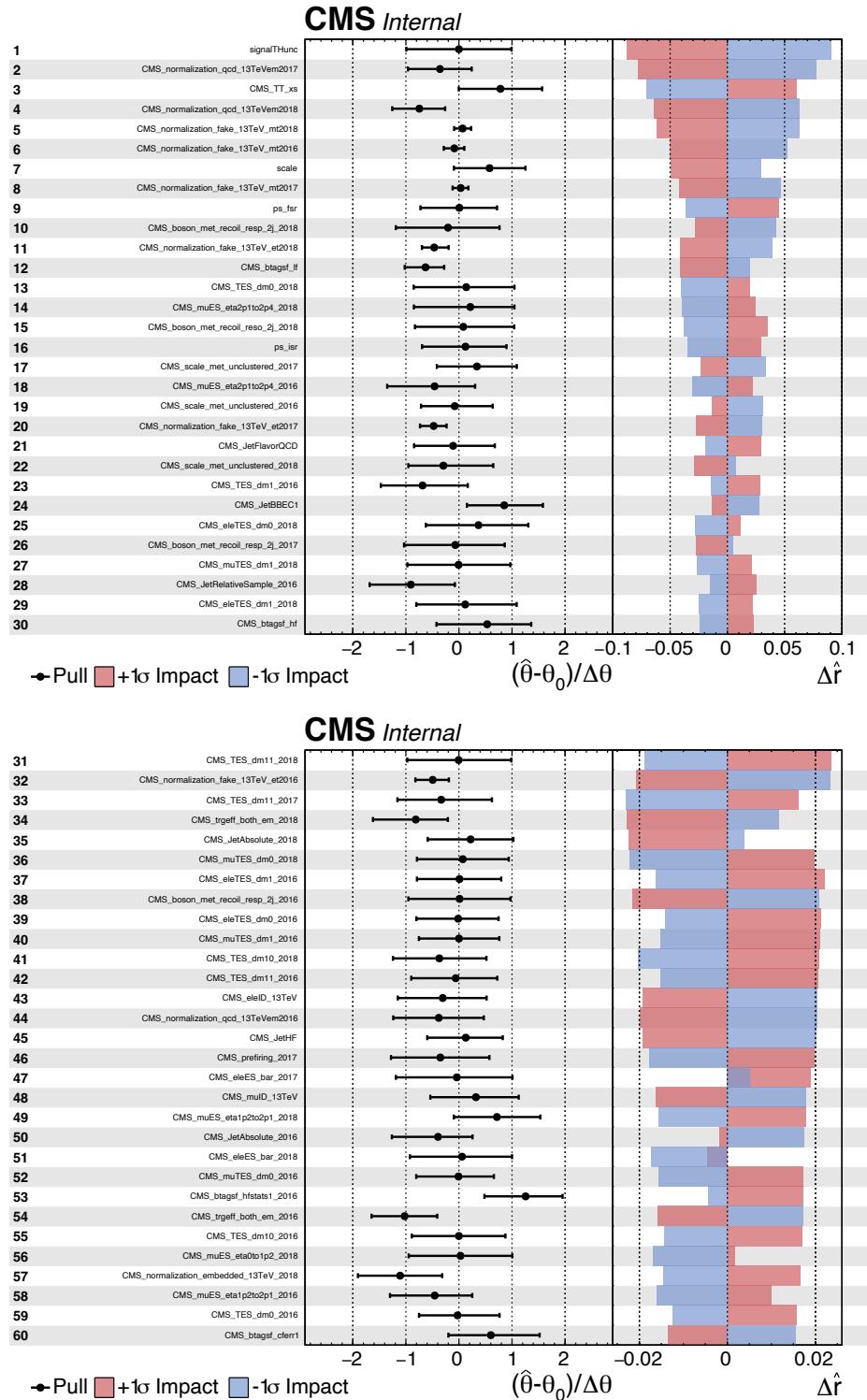


Figure 8.1: Top sixty pulls and impacts for the combination of all channels and years [101].

2218 Chapter 9

2219 Event categorization and signal 2220 extraction

2221 Measured events are divided into categories, based on cuts on values of observables
2222 in the event, or some derived quantity based on the observables in the event. The
2223 objective of event categorization is to divide events into signal regions, where the
2224 signal is enhanced and the background is suppressed, and control regions, which are
2225 signal-poor and used to check that the background estimation methods employed in
2226 the analysis in fact accurately models the data. In this analysis, events in each di-tau
2227 channel ($\mu\tau_h$, $e\tau_h$, and $e\mu$) are selected to contain one or more b-tag jets reconstructed
2228 in the event as described in Section 9.1. Events are further divided into signal and
2229 control regions using a deep learning-based approach described in Section 9.2. The
2230 signal is extracted from the di-tau mass distribution in the signal region using the
2231 statistical procedure described in Section 9.3.

2232 9.1 B-tag jet multiplicity

2233 Compared to the previous CMS $h \rightarrow aa \rightarrow bb\tau\tau$ analysis which used 2016 data corre-
2234 sponding to an integrated luminosity of 35.9 fb^{-1} [84], this analysis is performed on

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

9.2 DNN-based event categorization

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

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

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

- 2260 • The angular separation ΔR between pairs of the objects, where signal samples
 2261 peak at smaller ΔR values.
- 2262 • The transverse mass between the missing transverse energy p_T^{miss} and each of
 2263 the four objects [84], defined as

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

2264 where p_T^ℓ is the transverse momentum of the object ℓ , and $\Delta\phi$ is the difference
 2265 in azimuthal angle between the object and the p_T^{miss} . Events from $t\bar{t}$ +jets and
 2266 jets faking τ_h backgrounds have larger p_T^{miss} resulting in larger transverse mass
 2267 values compared to the signal, which tends to have smaller p_T^{miss} that is also
 2268 more aligned with the lepton legs.

- 2269 • The variable D_ζ [84], defined as

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

2270 where the ζ axis is the bisector of the transverse directions of the visible τ decay
 2271 products. p_ζ is the component of the p_T^{miss} along the ζ axis, and p_ζ^{vis} is the sum
 2272 of the components of the lepton p_T along the same axis. This variable captures
 2273 the fact that in signal the p_T^{miss} is small and approximately aligned with the $\tau\tau$.
 2274 In contrast, the $Z \rightarrow \tau\tau$ background tends towards large D_ζ values because the
 2275 p_T^{miss} is collinear to the $\tau\tau$, and the $t\bar{t}$ +jets events tend to have small D_ζ due to
 2276 a large p_T^{miss} not aligned with the $\tau\tau$.

- 2277 • For events with 2 b-tag jets, one additional variable is defined to capture the
 2278 difference in the invariant mass of the bb and the $\tau\tau$:

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

2279 This variable peaks at zero for the $h \rightarrow aa \rightarrow 2b2\tau$ signal.

2280 The DNN model consists of an input layer, two fully-connected hidden layers,
2281 and one output layer, which has only one output for this binary classification of
2282 signal versus background. Two hidden layers were used, as one hidden layer led
2283 to undertraining, and three hidden layers led to overtraining. One dropout layer
2284 was inserted after each of the two hidden layers, which set zero weights at nodes
2285 chosen at a random rate (the dropout rate) during training to reduce overfitting. The
2286 output node uses a sigmoid activation function to produce a probability-like output
2287 $0 < y < 1$, where background samples were assigned a score of 0 and signal samples
2288 were assigned a score of 1. The training datasets were shuffled and divided into
2289 training, validation, and test sets, with an equal number of signal and background
2290 events in each set. Models were trained on the training set, and the performance on
2291 the training set was compared to the performance on the validation set in order to
2292 guide the tuning of hyperparameters in the DNN models (e.g. the number of nodes
2293 in the hidden layers and the dropout rate). The test set was used only to perform an
2294 unbiased evaluation of the final training.

2295 Events in the data, Monte Carlo, and embedded samples are evaluated with the six
2296 trained DNNs and assigned a raw score between 0 and 1 (background-like and signal-
2297 like respectively). In order to flatten the distribution of the score and define score
2298 thresholds for categorizing events, the raw output scores are transformed with the
2299 function $\tilde{p}(n) = \text{arctanh}(p \times \tanh(n))/n$ where n is a positive integer. The thresholds
2300 of the DNN score used for signal/control region definition are determined using scans
2301 that optimize the signal sensitivity and are shown in Tables 9.1 and 9.2.

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

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

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

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

2302 9.3 Methodology for signal extraction

2303 After events are divided into categories, the data is compared to the expected back-
2304 grounds in the signal region categories. Here, we describe the fundamental concepts
2305 behind hypothesis testing in high-energy physics, as well as how exclusion limits
2306 can be set on parameters whose true values we cannot measure, culminating in the
2307 modified frequentist method CL_S which is used to perform signal extraction in this
2308 analysis.

2309 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 [102]. 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

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

2310 read “ f of x given α ”, is referred to as a probability model or model. The parameters α
2311 typically represent parameters of the theory or an unknown property of the detector’s
2312 response. The parameters are not frequentist in nature, unlike x . Out of all the
2313 parameters, typically only a few are of interest, and are called the parameters of
2314 interest (POI), labeled μ here. The remaining are referred to as nuisance parameters
2315 (NP) [102] and are labeled $\boldsymbol{\theta}$.

2316 $f(x)$ is the probability density for the observable in one event and we wish to

2317 describe the probability density for a dataset with many events, $\mathcal{D} = \{x_1, \dots, x_n\}$,
 2318 called the total probability model \mathbf{f} . For instance, if we also have a prediction for
 2319 the total number of events expected, called ν , we also account for the overall Poisson
 2320 probability for observing n events given ν 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 [102]. The likelihood function is not a probability density for α
 and is not normalized to unity:

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

2321 i.e. the likelihood function is the value of f as a function of α given a fixed value of
 2322 x .

2323 To estimate the parameter α we use an estimator, which is a function of the
 2324 data. Take for example the measurement of data distributed according to a Gaussian
 2325 probability density $f(x|\mu, \sigma) = \text{Gauss}(x|\mu, \sigma)$. One possible estimator of the mean μ ,
 2326 is the mean of the measured data points $\bar{x} = \sum_{i=1}^n x_i/n$ [102].

2327 A commonly used estimator in physics is the maximum likelihood estimator
 2328 (MLE), defined as the value $\hat{\alpha}$ which maximizes the likelihood function $L(\alpha)$. This
 2329 value, labeled $\hat{\alpha}$, also maximizes $\ln L(\alpha)$ and minimizes $-\ln L(\alpha)$. By convention the
 2330 $-\ln L(\alpha)$ is minimized, in a process called “fitting”, and the maximum likelihood
 2331 estimate is called the “best fit value”.

2332 9.3.2 Hypothesis testing

2333 In this section we next introduce concepts related to hypothesis testing such as the
 2334 test statistic constructed from the ratio of likelihood functions.

2335 The objective of a likelihood analysis is to distinguish different models repre-
2336 senting the various hypotheses, and determine the one that best explains the ex-
2337 perimental outcome. In a search for new physics, a signal is additive on top of the
2338 background. The background-only hypothesis is the null hypothesis, and the signal-
2339 plus-background hypothesis is the alternative.

2340 As a simple example, take the p -value test, for an experiment where we count
2341 events in the signal region, n_{SR} , and expect ν_B background events and ν_S events from
2342 the signal [102]. Then

- 2343 1. The null hypothesis (H_0), i.e. the background-only hypothesis in this experi-
2344 ment, with the probability modeled by $\text{Poisson}(n_{SR}|\nu_B)$.
- 2345 2. The alternate hypothesis (H_1), i.e. signal-plus-background hypothesis, with the
2346 probability modeled by $\text{Poisson}(n_{SR}|(\nu_B + \nu_S))$.

2347 The compatibility of the observed data ν_{SR}^0 and the null hypothesis, is quantified as
2348 the probability that the background-only hypothesis would produce at least as many
2349 events as was observed. This probability is the p -value:

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

2350 If the p -value is very small, we might reject the null hypothesis. The p -value is not the
2351 probability of the null hypothesis given the data; rather, it expresses the probability
2352 that data with a certain property was obtained, assuming the null hypothesis [102].

2353 The p -value is an example of a test statistic T , which maps the data to a single
2354 real number. The Neyman-Pearson lemma states that out of the infinite possibilities
2355 of choices of test statistic, the uniformly most powerful test statistic is the likelihood
2356 ratio T_{NP} [102]:

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

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

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

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

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

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

9.3.3 Confidence intervals

In an example of the measurement of the Standard Model Higgs boson, $\boldsymbol{\alpha}_{\text{POI}} = (\sigma/\sigma_{SM}, M_H)$, with σ/σ_{SM} is the ratio of the production cross-section for Higgs with respect to its value in the SM, and M_H is the unknown mass of the Higgs, values of these parameters outside specific bounds are said to be “excluded at the 95% confidence level”. These allowed regions are called confidence levels or confidence regions, and the parameter values outside of them are considered excluded [102]. A 95% confidence interval does not mean that there is a 95% chance that the true value of the parameter is inside the interval. Rather, a 95% confidence interval covers the

2371 true value 95% of the time (even though we do not know the true value).

2372 To construct a confidence interval for a parameter α , the Neyman Construction
2373 is used to invert a series of hypothesis tests; i.e. for each possible value of α , the null
2374 hypothesis is treated as α , and we perform a hypothesis test based on a test statistic.
2375 To construct a 95% confidence interval, we construct a series of hypothesis tests with
2376 size of 5%. The confidence interval $I(\mathcal{D})$ is constructed by taking the set of parameter
2377 values α where the null hypothesis is accepted:

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

2378 where $T(\mathcal{D})$ is the test statistic, and the last α (not bolded) and the subscript k_α
2379 refer to the size of the test. A schematic of the Neyman construction is shown in Fig.
2380 9.1. In a more generalized case, the x -axis is the test statistic T .

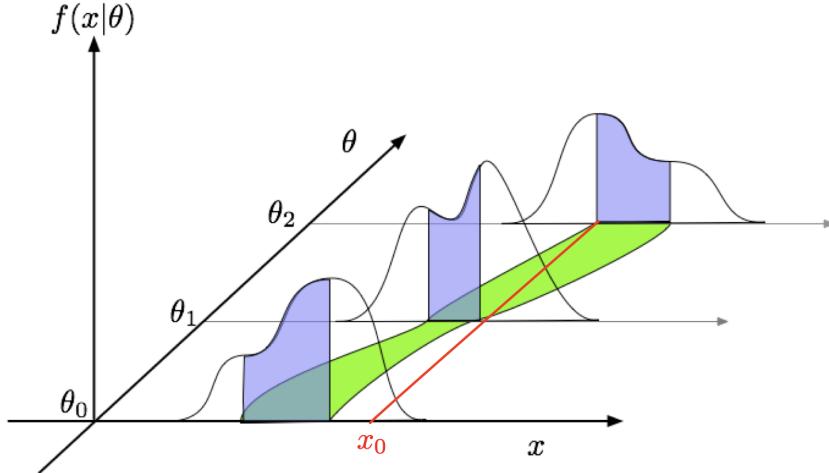


Figure 9.1: Schematic of the Neyman construction for confidence intervals [102]. For each value of θ , we find a region in x where $\int f(x|\theta)dx$ satisfies the size of the test (blue). These regions form a confidence belt (green). The intersection of the observation x_0 (red) with the confidence belt defines the confidence interval $[\theta_1, \theta_2]$ [102].

2381 9.3.4 Profile likelihood ratio

2382 In this section we describe a frequentist statistical procedure based on the profile
 2383 likelihood ratio test statistic, which is implemented using asymptotic distributions.

2384 With a multi-parameter likelihood function $L(\boldsymbol{\alpha})$, the maximum likelihood of
 2385 one specific parameter α_p with other parameters $\boldsymbol{\alpha}_o$ fixed, is called the conditional
 2386 maximum likelihood estimate and is denoted $\hat{\alpha}_p(\boldsymbol{\alpha}_0)$. The process of choosing specific
 2387 values of the nuisance parameters for a given value of μ , $\mathcal{D}_{\text{simulated}}$, and value of global
 2388 observables \mathcal{G} is called profiling. From the full list of parameters $\boldsymbol{\alpha}$, we denote the
 2389 parameter of interest μ , and the nuisance parameters $\boldsymbol{\theta}$.

2390 We construct the profile likelihood ratio,

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

2391 which depends explicitly on the parameter of interest μ , implicitly on the data \mathcal{D}_{sim}
 2392 and global observables \mathcal{G} , and is independent of the nuisance parameters $\boldsymbol{\theta}$, which
 2393 have been eliminated in profiling [102].

2394 The main conceptual reason for constructing the test statistic from the profile
 2395 likelihood ratio is that asymptotically (i.e. for measurements with many events) the
 2396 distribution of the profile likelihood ratio $\lambda(\mu = \mu_{\text{true}})$ is independent of the values of
 2397 the nuisance parameters [102].

2398 The following p -value is used to quantify the consistency with the hypothesis of a
 2399 signal strength of μ :

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

2400 9.3.5 Modified frequentist method: CL_S

2401 In the modified frequentist method called CL_S , to test a hypothesis with signal, we
2402 define p'_μ as a ratio of p -values [102]:

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

2403 where p_b is the p -value derived under the background-only hypothesis:

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

2404 The CL_S upper limit on μ , denoted μ_{up} , is obtained by solving for $p'_{\mu_{up}} = 5\%$.
2405 If testing the compatibility of the data with the background-only hypothesis, we
2406 consider the p_b value defined above and conventionally convert it into the quantile
2407 or “sigma” of a unit Gaussian. z standard deviations (e.g. $z = 5$ in “ 5σ ”) means
2408 that the probability of falling above these standard deviations, equals p_b (e.g. 3σ
2409 corresponds to $p_b = 2.7 \times 10^{-3}$ or 95.43%, and 5σ corresponds to $p_b = 5.7 \times 10^{-7}$ or
2410 99.999943%).

2411 **Chapter 10**

2412 **Results**

2413 In this chapter, Section 10.1 presents the results from the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis
2414 performed on 137 fb^{-1} of data from the full CMS Run-2 dataset in the years 2016 to
2415 2018, with interpretations provided for different 2HDM+S scenarios. This analysis
2416 was combined with a different search in the $h \rightarrow aa \rightarrow bb\mu\mu$ final state, which was
2417 also performed on the full Run-2 dataset. The combination procedure and results
2418 from the combined analyses ($h \rightarrow aa \rightarrow bb\ell\ell$, with $\ell = \mu, \tau$) are detailed in 10.2.
2419 The combined analysis places some of the most stringent limits to date at CMS for
2420 2HDM+S scenarios in the light scalar mass range $m_a = 12 \text{ GeV}$ to 60 GeV .

2421 **10.1 Results from $bb\tau\tau$**

2422 In each of the three $\tau\tau$ channels studied ($\mu\tau_h$, $e\tau_h$, and $e\mu$), events are divided based
2423 on whether they contain exactly 1 or 2 b-tag jets, and further divided into signal
2424 and control regions (SRs and CRs) using the DNN categorization score as described
2425 in Section 9.2. The control regions demonstrate good agreement between observed
2426 events in data, and the sum of the contributions from expected backgrounds that
2427 are modeled in simulated and embedded samples. The signal regions are defined to
2428 be sensitive to the $h \rightarrow aa \rightarrow bb\tau\tau$ signal. The postfit final observed and expected

2429 distributions of the di-tau invariant mass $m_{\tau\tau}$ reconstructed with SVFit (described
2430 in Section 5.2) are shown in Fig. 10.1 for the $\mu\tau_h$ channel, Fig. 10.2 for the $e\tau_h$
2431 channel, and Fig. 10.3 for the $e\mu$ channel. In all figures, the hypothesized yield for
2432 the $h \rightarrow aa \rightarrow bb\tau\tau$ signal is shown for the pseudoscalar mass $m_a = 35$ GeV and
2433 assuming a branching fraction $B(H \rightarrow aa \rightarrow bb\tau\tau) = 10\%$.

2434 The 95% CL expected and observed exclusion limits on the signal strength of the
2435 branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ as a function of the pseudoscalar mass m_a
2436 ranging from 12 GeV to 60 GeV, are shown for the three $\tau\tau$ channels and all three
2437 channels combined in Fig. 10.4. The limits are shown as percentages and normalized
2438 to the production cross-section of the Standard Model Higgs boson. No excess of
2439 events above the Standard Model expectations is observed. In the limits for the three
2440 $\tau\tau$ channels combined, expected (observed) limits range from 1.4 to 5.6% (1.7 to
2441 7.6%) for pseudoscalar masses between 12 and 60 GeV.

2442 The $e\mu$ channel is the only channel that has signal sensitivity to the $m_a = 12$
2443 GeV pseudoscalar mass hypothesis, because the minimum required spatial separation
2444 $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ between the two τ legs is smaller than the other two channels
2445 ($\Delta R < 0.3$ for $e\mu$, compared to $\Delta R < 0.4$ for the other two channels). This decreased
2446 ΔR requirement results in better signal acceptance for low mass signals for the $e\mu$
2447 channel. The $\mu\tau_h$ and $e\tau_h$ channels are most sensitive to the intermediate mass points
2448 studied, since the analysis targets a resolved signature: at low mass points, the tau
2449 legs are boosted, and at high mass points, the $m_{\tau\tau}$ distributions in signal have larger
2450 overlap with background distributions. In the combination of the three $\tau\tau$ channels,
2451 the limit for $m_a = 12$ GeV comes only from the $e\mu$ channel, and the best sensitivity
2452 is attained at intermediate mass points around $m_a = 20$ GeV to 45 GeV.

2453 To set limits on the branching fraction of the 125 GeV Higgs to the two pseu-
2454 doscalars, $B(h \rightarrow aa)$, we interpret the results in four types of 2HDM+S, which were
2455 introduced in Section 1.4. In 2HDM+S, the theorized branching fraction of the pseu-

2456 doscalars depends on the 2HDM+S model type, the pseudoscalar mass m_a , and the
2457 ratio of the two Higgs doublets' vacuum expectation values $\tan \beta$. In Type I models,
2458 the branching fraction is independent of $\tan \beta$, while in Types II, III, and IV, it is
2459 a function of m_a and $\tan \beta$. Limits for the $bb\tau\tau$ final state as a function of m_a for
2460 2HDM+S Type I (valid for all $\tan \beta$ values), Type II with $\tan \beta = 2.0$, Type III with
2461 $\tan \beta = 2.0$, and Type IV with $\tan \beta = 0.6$ are overlaid and shown in Fig. 10.5a.

2462 10.2 Combination with $bb\mu\mu$ final state

2463 Results from this analysis for the $h \rightarrow aa \rightarrow bb\tau\tau$ final state are combined with the
2464 analysis for the $h \rightarrow aa \rightarrow bb\mu\mu$ final state [103]. While the predicted branching ratio
2465 for $aa \rightarrow bb\mu\mu$ is comparatively small, the $bb\mu\mu$ final state has competitive results
2466 due to the excellent di-muon resolution measured by CMS. The $bb\mu\mu$ analysis uses
2467 an unbinned fit to the data using the di-muon mass $m_{\mu\mu}$ distribution. Details can be
2468 found in [103].

2469 Combining the results is possible since the $bb\tau\tau$ analysis explicitly rejects events
2470 with extra leptons, so there is no overlap between the events studied in the $bb\tau\tau$
2471 analysis and the $bb\mu\mu$ analysis. In the statistical combination, several systematic
2472 uncertainties are treated as correlated: the integrated luminosity normalization, the
2473 b-tagging scale factor, the scale factors related to muon reconstruction, identifica-
2474 tion, and trigger efficiencies, the inefficiency in the ECAL trigger readout, and the
2475 theoretical uncertainties related to signal modeling.

2476 Since the results in both final states are statistically limited, the combination ben-
2477 efits from the additional data. For $m_a = 35$ GeV, all systematic uncertainties amount
2478 to around 6% of the total uncertainty, with the dominant systematic uncertainties
2479 coming from jet energy systematics in the $bb\mu\mu$ final state, theoretical uncertainties
2480 in the signal, and uncertainties in the QCD multijet backgrounds in the $e\mu$ channel

2481 of the $bb\tau\tau$ final state.

2482 The mass distributions of the di-muon and di-tau objects ($m_{\mu\mu}$ and $m_{\tau\tau}$) are
2483 compared to the data in a combined maximum likelihood fit to derive upper limits
2484 on $B(h \rightarrow aa)$. The observed limits at 95% CL on $B(h \rightarrow aa)$ for different 2HDM+S
2485 scenarios, are shown for the search for $h \rightarrow aa \rightarrow bb\mu\mu$ in Fig. 10.5b, and the
2486 combined analyses $h \rightarrow aa \rightarrow bb\ell\ell$ in Fig. 10.6.

2487 Exclusion limits in a two-dimensional plane as a function of $\tan\beta$ and m_a are
2488 set for 2HDM+S Types II, III, and IV in Fig. 10.7. The most stringent constraints
2489 are observed for 2HDM+S type III because of large branching fractions predicted in
2490 theory, with predicted branching fractions between 0.47 and 0.42 for $\tan\beta = 2.0$ and
2491 values of m_a between 15 and 60 GeV, compared to the observed 95% CL upper limits
2492 which are between 0.08 and 0.03. For 2HDM+S type IV, the predicted branching
2493 fractions from theory are between 0.26 and 0.20 for $\tan\beta = 0.6$ for values of m_a
2494 between 15 and 60 GeV, and the 95% CL observed upper limits are between 0.12 and
2495 0.05.

2496 The combined results from $h \rightarrow aa \rightarrow bb\ell\ell$ are compared with CMS results in
2497 other final states as a function of the pseudoscalar mass m_a : for 2HDM+S type I in
2498 Fig. 10.8, type II with $\tan\beta = 2.0$ in Fig. 10.9, and type III with $\tan\beta = 2.0$ in Fig.
2499 10.10. In other scenarios, e.g. type III with $\tan\beta = 5.0$, more stringent limits are set
2500 by analyses in other final states, $\mu\mu\tau\tau$ in this case. Other summary plots for other
2501 model types and $\tan\beta$ values can be found at [104].

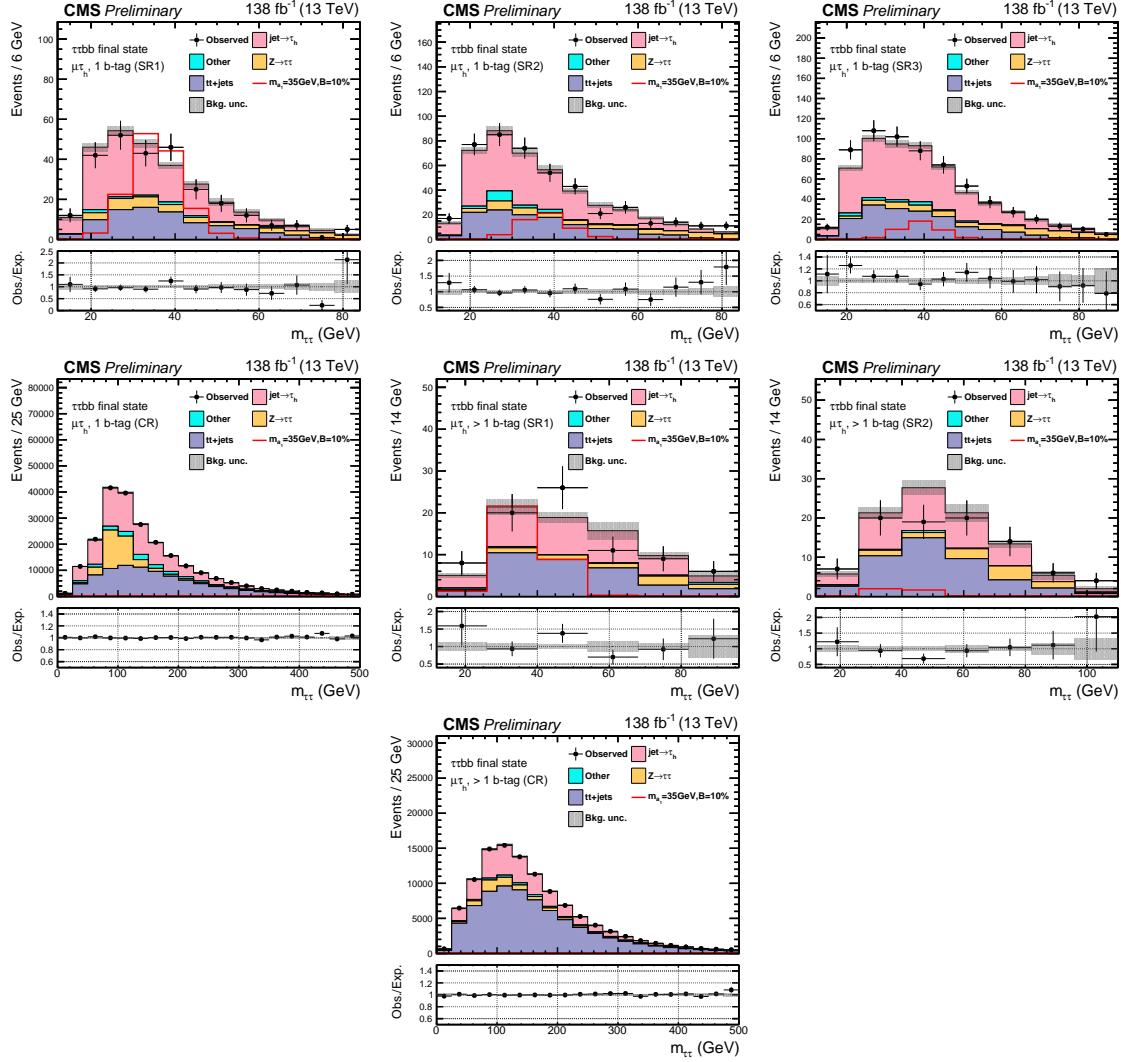


Figure 10.1: Postfit final $m_{\tau\tau}$ observed and expected distributions, and the observed/expected ratios, in the $\mu\tau_h$ channel [101]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), 1 b-tag jet control region (*middle row*), 2 b-tag jet signal regions (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 τ_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*).

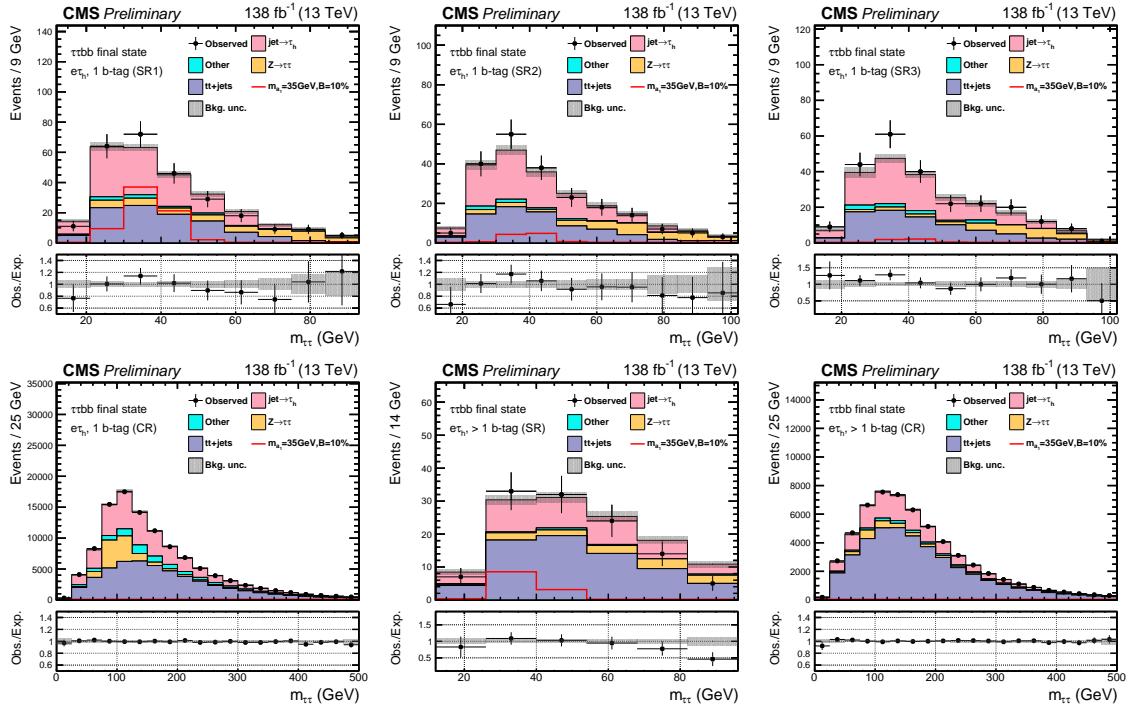


Figure 10.2: Postfit final observed and expected $m_{\tau\tau}$ distributions, and the observed/expected ratios, in the $e\tau_h$ channel [101]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), the 1 b-tag jet control region (CR) (*bottom row*), and 2 b-tag jet signal 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 τ_h leg (*pink*), $Z \rightarrow \tau\tau$ (*orange*), and $t\bar{t}+{\rm jets}$ (*purple*). For illustrative purposes, the beyond-Standard Model signal yield from $h \rightarrow aabb\tau\tau$ is shown for the pseudoscalar mass hypothesis $m_a = 35$ GeV, assuming a branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau) = 10\%$ (*red line*).

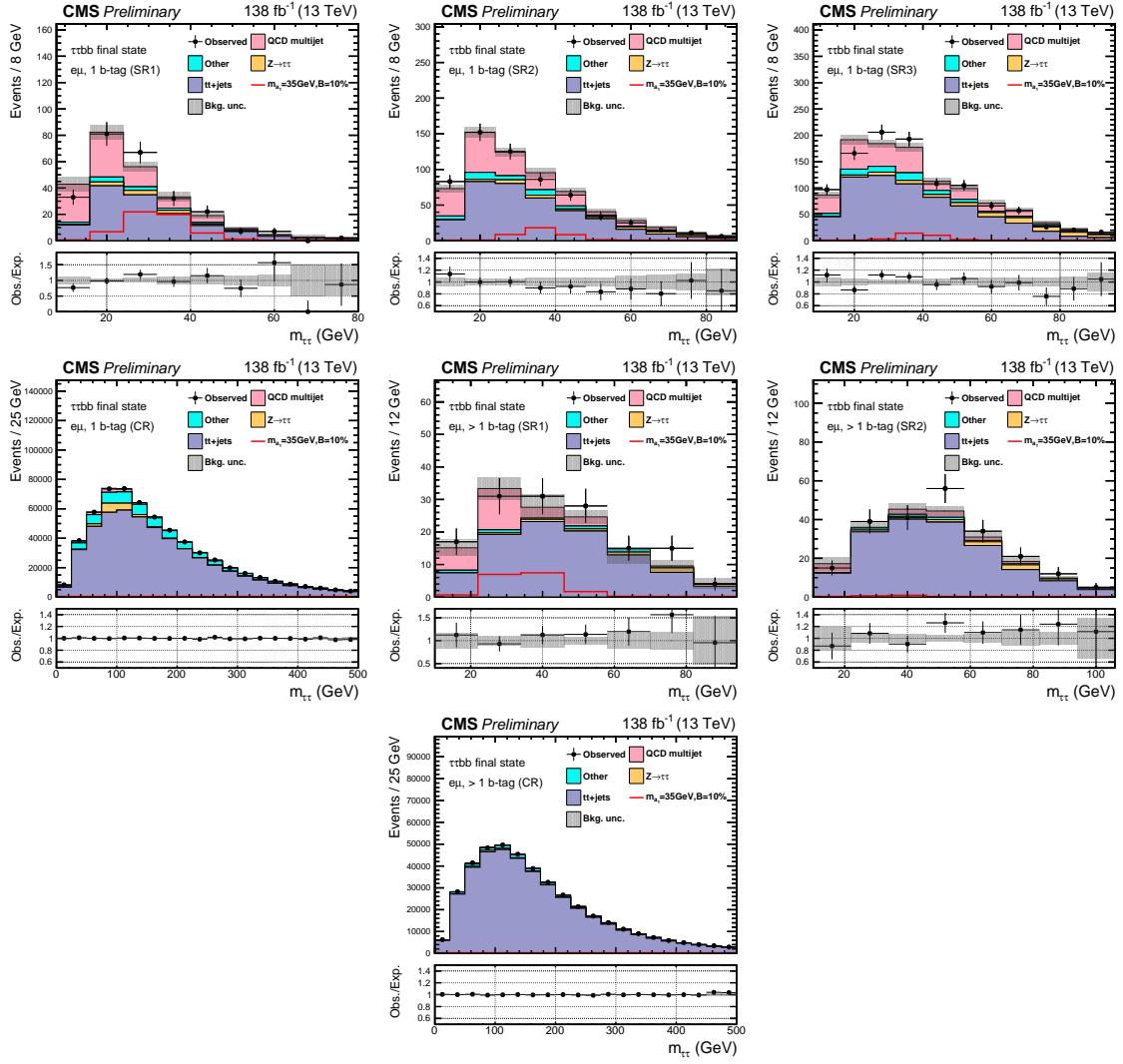


Figure 10.3: Postfit final observed and expected $m_{\tau\tau}$ distributions, and the observed/expected ratios, in the $e\mu$ channel [101]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, and SR3) (*top row*), 1 b-tag jet control region (CR) (*middle row*), 2 b-tag jet signal regions (SR1 and SR2) (*middle row*), and 2 b-tag jet control region (CR) (*bottom row*). Statistical and systematic sources of uncertainties in the expected events are added in quadrature and labeled “Bkg. unc” (*shaded gray*). The $t\bar{t}+j$ process (*purple*) is a major background, and in the signal regions the QCD multijet (*pink*) is also a major background. For illustrative purposes, the 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*).

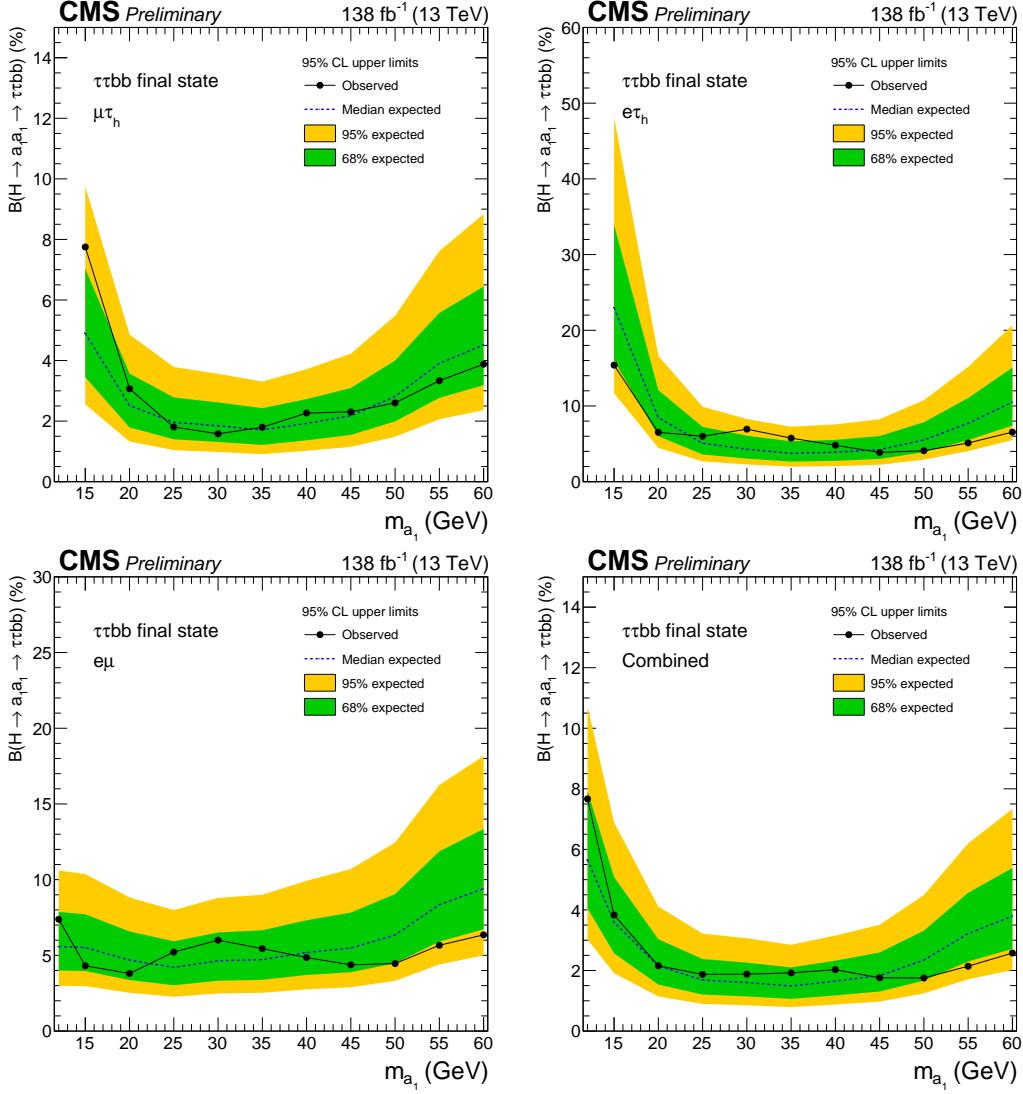
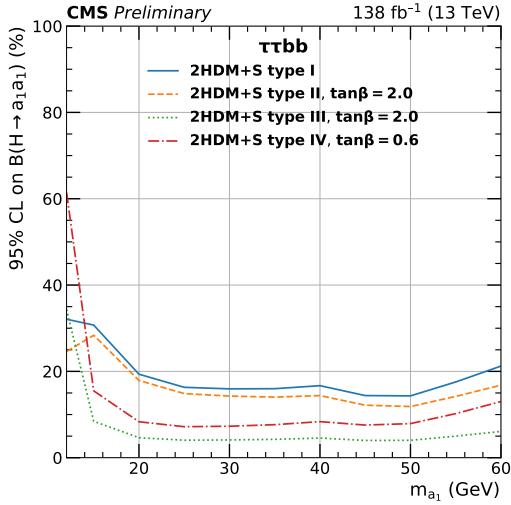
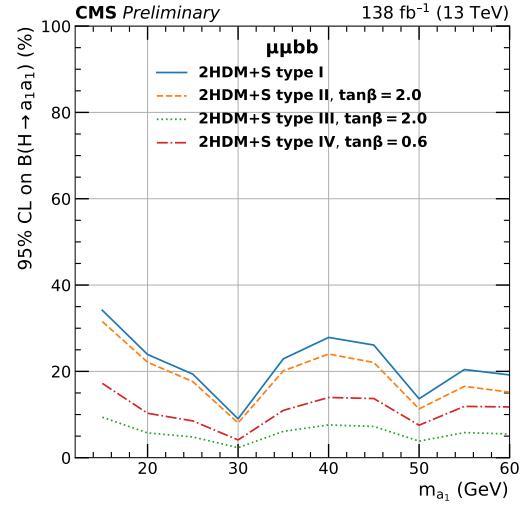


Figure 10.4: Observed 95% CL exclusion limits (*black, solid lines*) and expected 95% CL and 68% CL limits (*shaded yellow and green*) on the branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard Model production for the 125 GeV Higgs (h). Limits are shown 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*) [101]. The dataset corresponds to 138 fb^{-1} of data collected in the years 2016-2018 at a center-of-mass energy 13 TeV. Only the $e\mu$ channel has sensitivity to the mass hypothesis $m_a = 12 \text{ GeV}$. The best sensitivity is attained at intermediate mass points.



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



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

Figure 10.5: Observed 95% CL upper limits on $B(h \rightarrow aa)$ in %, for the $bb\tau\tau$ final state (*left*) and $bb\mu\mu$ final state (*right*) using the full Run 2 integrated luminosity of 138 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) [101]. Linear interpolation is used between points in the graphs. The $\tan\beta$ values chosen here correspond to the most stringent limits in each model.

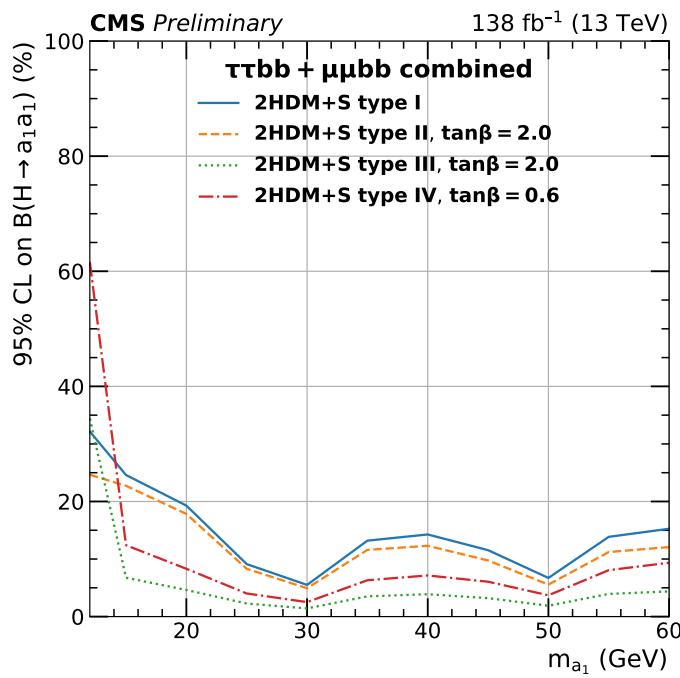


Figure 10.6: Observed 95% CL upper limits on the branching fraction of the 125 GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$, in percentages, as a function of the pseudoscalar mass m_a , in 2HDM+S type I (blue), type II with $\tan\beta = 2.0$ (orange dashed), type III with $\tan\beta = 2.0$ (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 fb^{-1} [101].

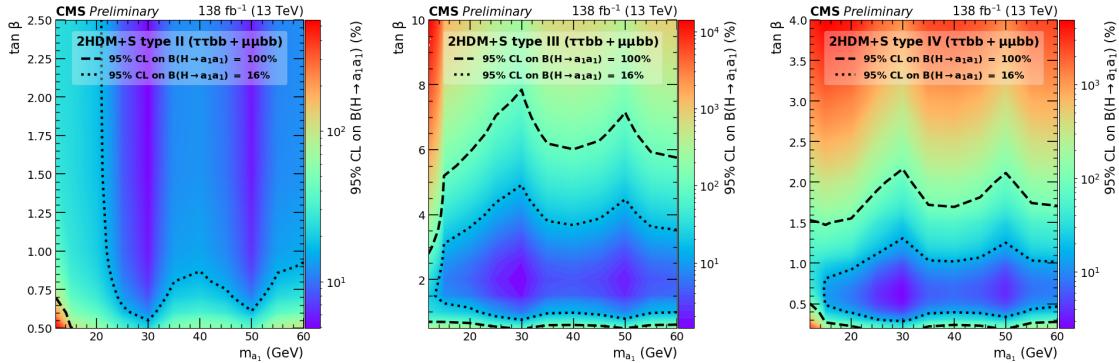


Figure 10.7: Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity of 138 fb^{-1} for Type II (*left*), Type III (*middle*), and Type IV (*right*) 2HDM+S in the $\tan \beta$ vs. m_a 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 [101].

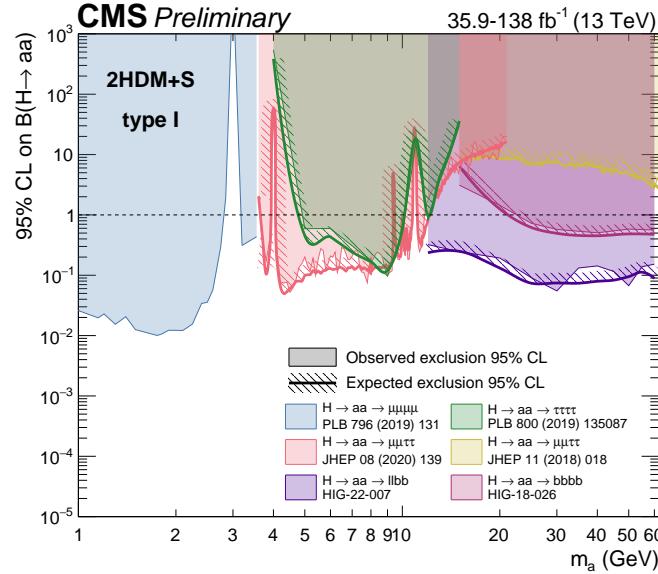


Figure 10.8: Summary plot of current 95% limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$ in the 2HDM+S type I scenario performed with data collected at 13 TeV [104]. Results from different final states studied at CMS are overlaid on this figure: $\mu\mu\mu\mu$ (blue), $\tau\tau\tau\tau$ (green), boosted $2\mu 2\tau$ (red), resolved $2\mu 2\tau$ (yellow), $bbbb$ (magenta), and the combined result for $\ell\ell bb$ ($\ell = \mu, \tau$) (purple).

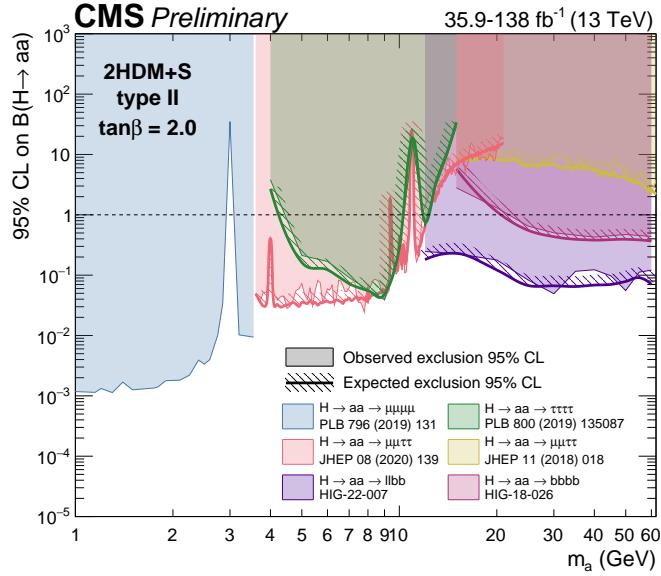


Figure 10.9: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$, in the 2HDM+S type II scenario with $\tan\beta = 2.0$, obtained at CMS with data collected at 13 TeV [104]. 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).

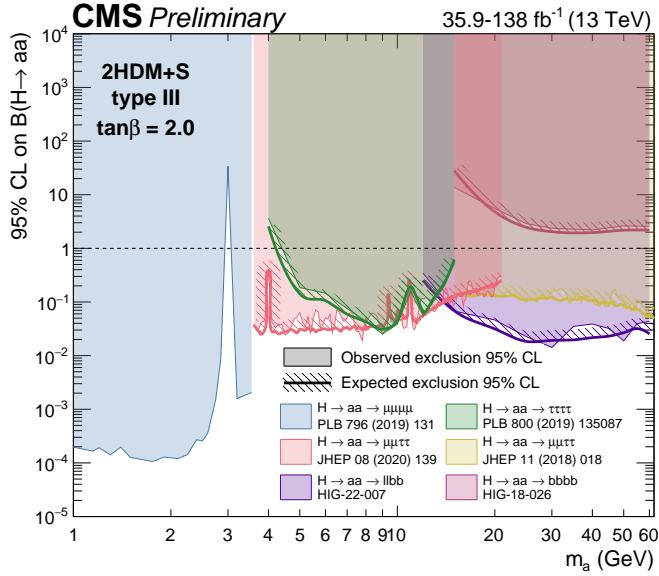


Figure 10.10: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross section, $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$ in the 2HDM+S type-III scenario with $\tan \beta = 2.0$, obtained at CMS with data collected at 13 TeV [104]. 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 $llbb$ ($\ell = \mu, \tau$) (purple).

2502 **Chapter 11**

2503 **Asymmetric exotic Higgs decays**

2504 This chapter presents progress towards a search for exotic Higgs decays to two light
2505 scalars with unequal mass ($h \rightarrow a_1 a_2$) to final states with bottom quarks and τ leptons,
2506 which has interpretations in Two Real Singlet Models (TRSMs) described in
2507 Section 1.5. Compared to the symmetric decay scenario $h \rightarrow aa$, which has been studied
2508 in multiple final states at CMS with stringent limits set on the various 2HDM+S
2509 scenarios, this asymmetric decay scenario has not been directly searched for at the
2510 CMS experiment. Section 11.1 lists the mass hypotheses of the new particles a_1 and
2511 a_2 considered in this search. Section 11.2 describes the studies performed on the simulated
2512 signal samples to determine which channels are viable for the analysis. Section
2513 11.3 shows the control plots produced using the framework for this analysis.

2514 **11.1 Signal masses**

2515 As discussed in Section 1.5, $h \rightarrow a_1 a_2$ can result in a “cascade” decay if one of the
2516 scalars, a_2 is sufficiently heavy ($m_{a_2} > 2m_{a_1}$). The “non-cascade” case is where the
2517 light scalars decay directly to Standard Model particles.

2518 The mass hypotheses (mass points) (m_{a_1}, m_{a_2}) studied here are:

- 2519 • *Cascade mass points:* (15, 30), (15, 40), (15, 50), (15, 60), (15, 70), (15, 80),
 2520 (15, 90), (15, 100), (15, 110), (20, 40), (20, 50), (20, 60), (20, 70), (20, 80), (20,
 2521 90), (20, 100), (30, 60), (30, 70), (30, 80), and (30, 90) GeV

- 2522 • *Non-cascade mass points:* (15, 20), (15, 30), (20, 30), (20, 40), (30, 40), (30,
 2523 50), (30, 60), (40, 50), (40, 60), (40, 70), (40, 80), (50, 60), and (50, 70) GeV

2524 Samples were produced using the MadGraph5_aMCatNLO event generator, for each
 2525 signal mass point in the gluon-gluon fusion (ggF) and vector boson fusion (VBF)
 2526 production modes of the 125 GeV Higgs boson. In the sample generation, the decays
 2527 of a to Standard Model particles were specified to be decays to bottom quarks or τ
 2528 leptons.

2529 11.2 Cascade scenario signal studies

2530 Studies of the signal phenomenology in the cascade scenario were performed to de-
 2531 termine the viability of the $4b2\tau$ and/or $2b4\tau$ channels.

2532 Cross sections and branching fractions of the $4b2\tau$ and $2b4\tau$ final states were
 2533 compared using cross-section predictions provided by the authors of [7]. For an
 2534 example mass point $m_{a_2} = 80$ GeV, $m_{a_1} = 30$ GeV, the branching fractions to
 2535 $4b2\tau$ is ten times larger than $2b4\tau$: $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 4b2\tau) = 0.00857$, vs.
 2536 $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 2b4\tau) = 0.00068$. The $4b2\tau$ final state is chosen for this
 2537 analysis.

2538 In general the four b-flavor jets have low p_T at generator level, as illustrated for
 2539 example mass points (100, 15) GeV and (40, 20) GeV in Fig. 11.1. The p_T distribution
 2540 of the sub-leading jet peaks at an energy below 20 GeV, with the third and fourth
 2541 jets tending to have even softer energies.

2542 An event category with three or more b-tag jets was determined to be infeasible
 2543 due to low statistics in this category, due to the difficulties in reconstructing the third

2544 and fourth b-flavor jets which have very low transverse momenta p_T . Event categories
 2545 with exactly 1 b-tag jet and ≥ 2 b-tag jets are used.

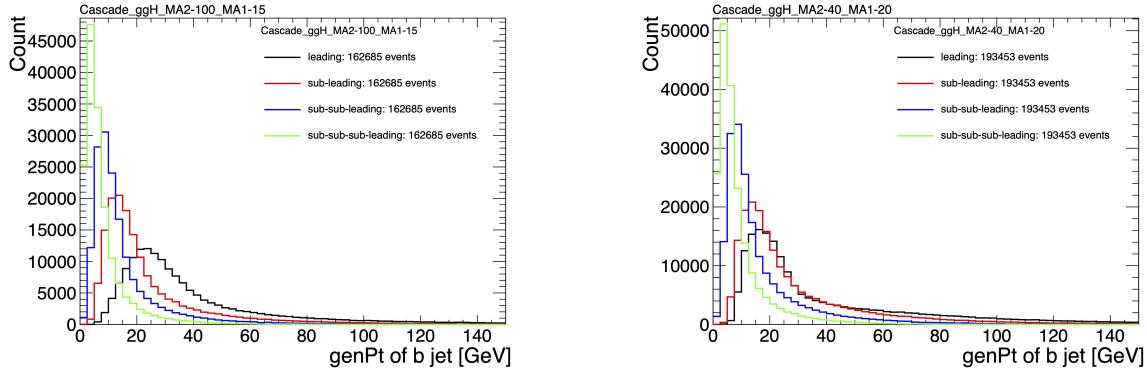


Figure 11.1: Generator-level b-flavor jet transverse momenta p_T , for $h \rightarrow a_1 a_2$ cascade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV (left) and $(40, 20)$ GeV (right). In each plot the generator-level p_T of the leading (black), sub-leading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

2546 In the $4b2\tau$ final state, the possibility of the leading and sub-leading b-tag jets
 2547 being sufficiently close in ΔR to require boosted jet reconstruction techniques was
 2548 explored. In the $4b2\tau$ case, the two b-flavor-jets in the generated event that were
 2549 spatially closest in ΔR were considered as one object. This two b-flavor jet object was
 2550 spatially matched in ΔR to the jets reconstructed with the standard AK4 algorithm
 2551 which uses a cone size of $\Delta R = 0.4$. The quality of the p_T resolution (computed as
 2552 $(p_{T,\text{reconstructed}} - p_{T,\text{gen}})/p_{T,\text{gen}}$) and closeness in distance ΔR of the reconstructed jet
 2553 to the nearest generator-level jets, was seen to depend on the absolute and relative
 2554 masses of the light scalars. The best (worst) performance occurred in samples with
 2555 large (small) mass differences between the heavier scalar a_2 and the lighter scalar a_1 ,
 2556 as illustrated for the mass hypotheses (m_{a_1}, m_{a_2}) $(100, 15)$ GeV and $(40, 20)$ GeV in
 2557 Fig. 11.2.

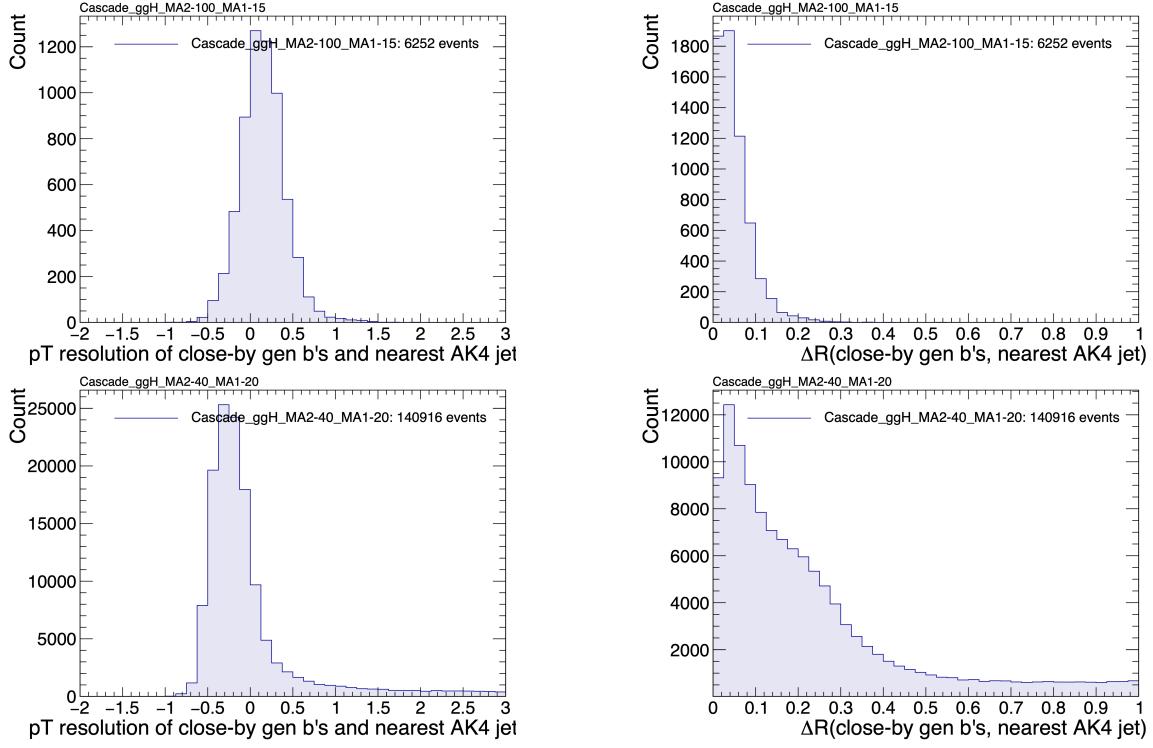


Figure 11.2: Distributions (arbitrary units) of transverse momentum p_T resolution and ΔR between the two closest generator-level b jets, treated as one object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow a_1 a_2$ mass hypotheses (m_{a_1}, m_{a_2}) = (100, 15) GeV (top left, 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 ΔR 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 are similar to those in the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis. 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 τ_h as described in Chapter 7. Nominal values of the scale factors and event reweighting are applied, as described in Chapter 5.3. The errors shown in the figures include statistical errors and several of the full set of systematic errors (only those associated with the lepton energy scales and τ_h identification efficiency, described in

2567 Sections 5.3.1, 5.3.2, and 5.3.4).

2568 The p_T , η , and ϕ of the leading muon and hadronic tau τ_h , and the di-tau visible
2569 mass m_{vis} and momentum $p_{T,\text{vis}}$, are shown in Fig. 11.3. The p_T , η , and ϕ of the the
2570 leading and sub-leading b-tag jets, and the missing transverse energy magnitude and
2571 azimuthal direction, are shown in Fig. 11.4.

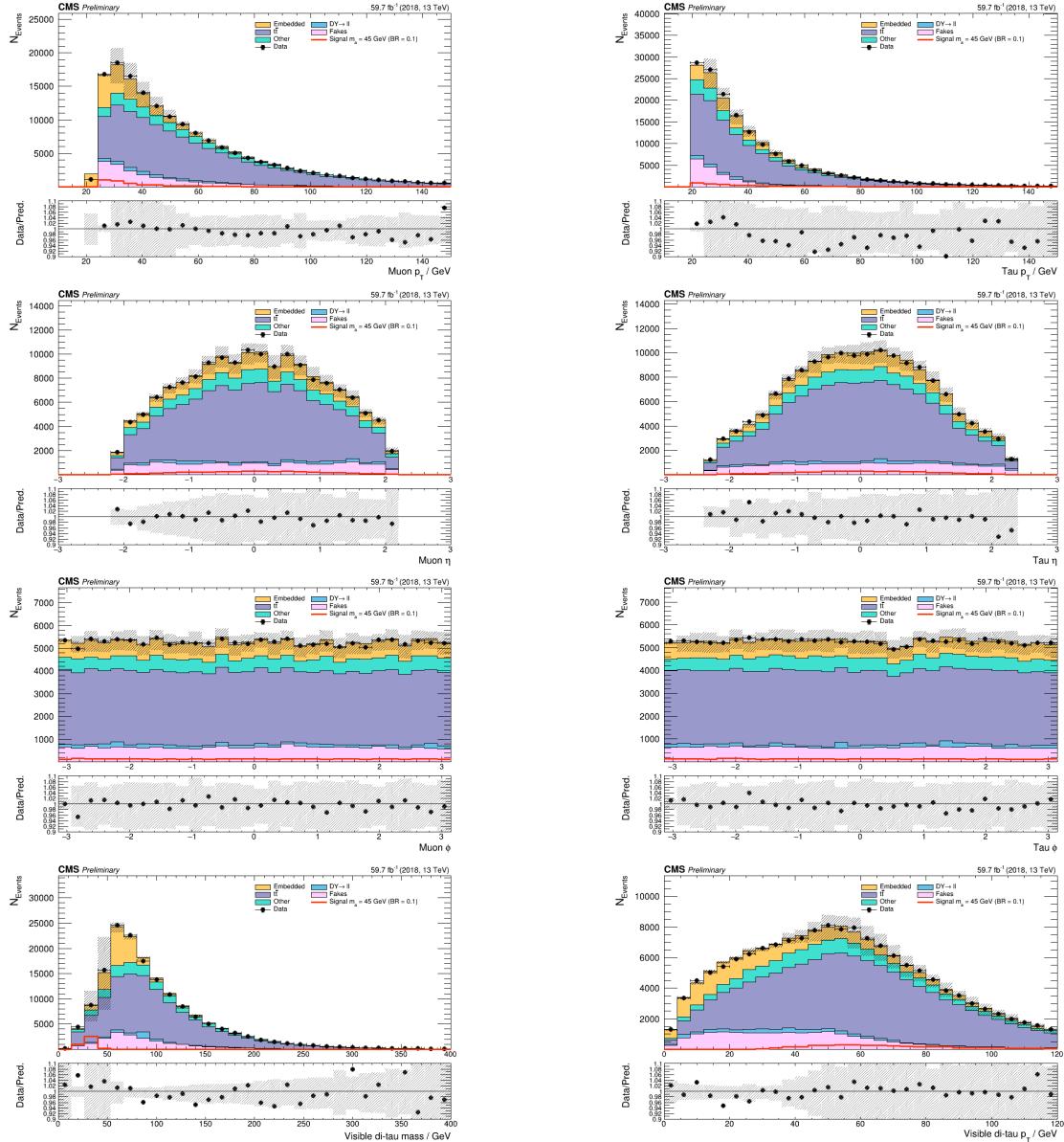


Figure 11.3: Kinematic properties of the leading muon and τ_h in the $\mu\tau_h$ channel: p_T (*top row*), η (*second row*), and ϕ (*third row*). The visible 4-momenta of the muon and τ_h are summed, giving the visible di-tau mass m_{vis} and transverse momentum $p_{T,\text{vis}}$. The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and τ_h identification efficiency).

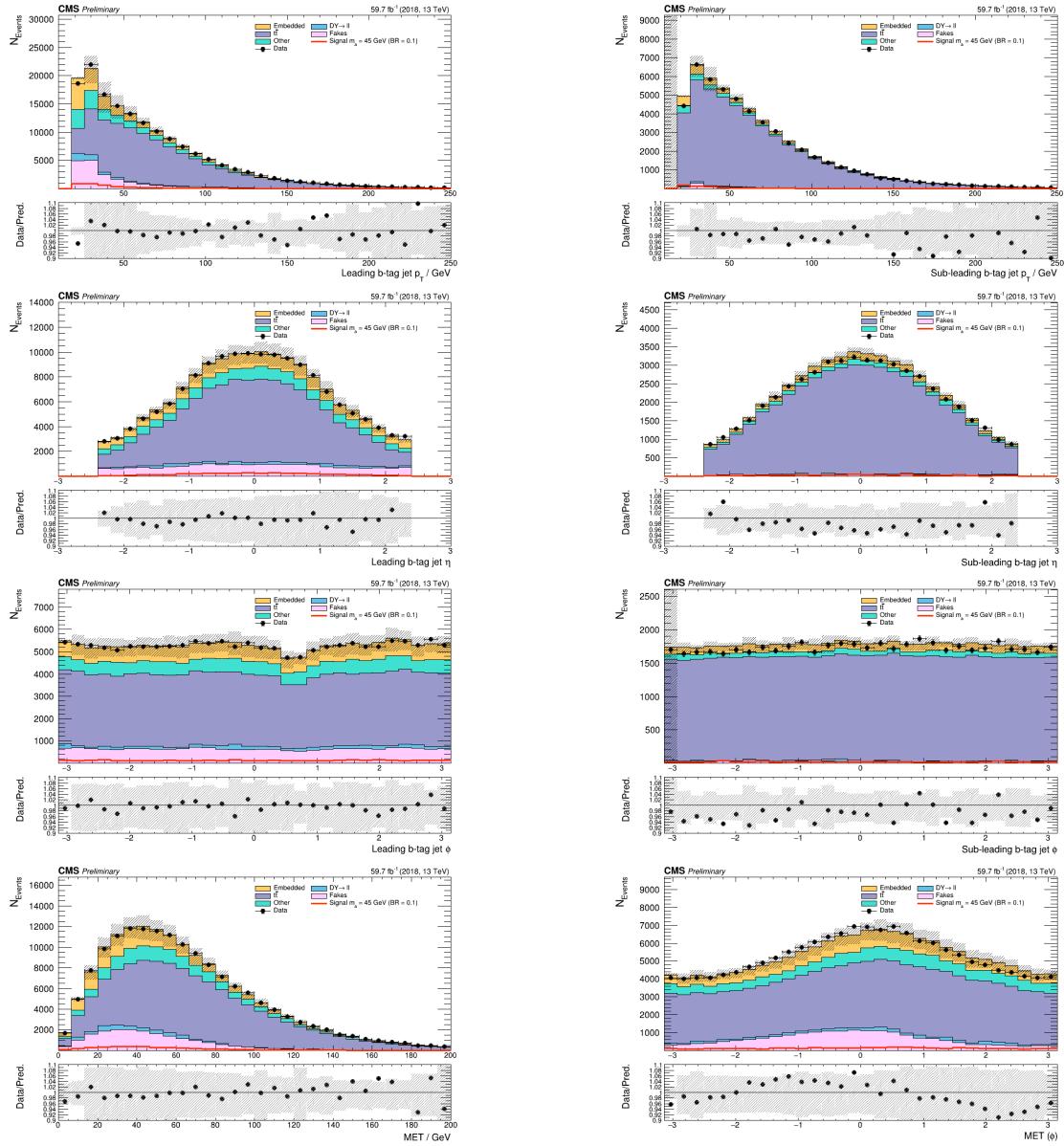


Figure 11.4: Kinematic properties of the leading and sub-leading b-tag jets in the $\mu\tau_h$ final state: jet p_T (top row), η (second row), ϕ (third row), as well as the missing transverse energy magnitude and azimuthal direction (bottom row). The errors shown in the figures include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and τ_h identification efficiency).

2572 **Chapter 12**

2573 **Conclusion and outlook**

2574 This thesis presents a direct search at the CMS experiment for exotic decays of the
2575 Higgs boson with mass 125 GeV in data collected in the years 2016-2018 in proton-
2576 proton collisions at center-of-mass energy 13 TeV, to two light neutral scalar particles
2577 that decay to two bottom quarks and two tau leptons ($h \rightarrow aa \rightarrow bb\tau\tau$). The results
2578 are combined with another search that was performed in the $h \rightarrow aa \rightarrow bb\mu\mu$ final
2579 state, giving the most stringent limits to date for theories with Two Higgs Doublet
2580 Models extended with a singlet scalar (2HDM+S), for pseudoscalar masses m_a ranging
2581 from 15 GeV to 60 GeV, in a number of 2HDM+S scenarios such as type II and III
2582 with $\tan\beta = 2.0$.

2583 As the rich physics program of CMS has set stringent limits on the exotic decay
2584 $h \rightarrow aa$, we turn our attention to direct searches for decays to light neutral scalars
2585 with potentially unequal mass, $h \rightarrow a_1a_2$, which has not been performed at CMS
2586 to date. Preliminary studies on $h \rightarrow a_1a_2$ signals in the Two Real Singlet Model
2587 (TRSM) are shown, and work is ongoing to develop the analysis for $h \rightarrow a_1a_2$ in final
2588 states with bottom quarks and tau leptons.

2589 To ensure the continued performance of the CMS detector and to enhance its
2590 data-taking capabilities in the intense pile-up conditions of the Phase-2 upgrade of

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

2607

Appendix A

2608

Samples used

2609 The datasets used in the MiniAOD-based framework for the $h \rightarrow aa \rightarrow bb\tau\tau$ analysis
2610 are listed in this appendix. The NanoAOD-based framework uses the NanoAOD ver-
2611 sions of these datasets. The data used for the years 2016-2018 are listed in Tables A.1,
2612 A.2, and A.3 respectively. The embedded samples used for the years 2016-2018 are
2613 listed in Tables A.4, A.5, and A.6 respectively. The Monte Carlo simulated samples
2614 used to estimate backgrounds for the years 2016-2018 are listed in Tables A.7, A.8,
2615 and A.9 respectively.

2616 The $h \rightarrow aa \rightarrow bb\tau\tau$ signal samples are generated for 11 psuedoscalar masses
2617 between 12 GeV and 60 GeV for gluon fusion (ggF) and vector boson fusion (VBF)
2618 Higgs production. The 2016-2018 signal samples are listed in Tables A.10, A.11 and
2619 A.12 respectively. A filter is applied at the generator level for each $\tau\tau$ final state:

- 2620 • ee final state: $p_T(e_1) > 22$ GeV, $p_T(e_2) > 10$ GeV, $|\eta(e_1)| < 2.6$, and $|\eta(e_2)| <$
2621 2.6.

- 2622 • $e\tau_h$ final state: $p_T(e) > 22$ GeV, $p_T(\tau_h) > 16$ GeV, $|\eta(e)| < 2.6$, and $|\eta(\tau_h)| < 2.7$.

- 2623 • $e\mu$ final state: $p_T(e) > 11$ GeV, $p_T(\mu) > 7$ GeV, $|\eta(e)| < 2.6$, and $|\eta(\mu)| < 2.5$.

- 2624 • $\tau_h\tau_h$ final state: $p_T(\tau_{h1}) > 28$ GeV, $p_T(\tau_{h2}) > 28$ GeV, $|\eta(\tau_{h1})| < 2.5$, and

2625 $|\eta(\tau_{h2})| < 2.5$.

2626 • $\mu\tau_h$ final state: $p_T(\mu) > 19 \text{ GeV}$, $p_T(\tau_h) > 16 \text{ GeV}$, $|\eta(\mu)| < 2.5$, and $|\eta(\tau_h)| <$
 2627 2.7 .

2628 • $\mu\mu$ final state: $p_T(\mu_1) > 17 \text{ GeV}$, $p_T(\mu_2) > 8 \text{ GeV}$, $|\eta(\mu_1)| < 2.5$, and $|\eta(\mu_2)| <$
 2629 2.5 .

2630 The tables also show for each sample the filter efficiencies, which is the percentage
 2631 of events that pass the above filters, and the number of events that were generated
 2632 after applying the filters.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Process	Simulated background samples (2018)	Cross section (pb)
DY	DY1JetsToLL_M-50_TuneCP5	877.8 (LO)
	DY2JetsToLL_M-50_TuneCP5	304.4 (LO)
	DY3JetsToLL_M-50_TuneCP5	111.5 (LO)
	DY4JetsToLL_M-50_TuneCP5	44.0 (LO)
	DYJetsToLL_M-50_TuneCP5	5343.0 (LO)
	DYJetsToLL_M-10to50_TuneCP5	15810.0 (LO)
Top	TTTo2L2Nu_TuneCP5	88.29
	TTToHadronic_TuneCP5	377.96
	TTToSemiLeptonic_TuneCP5	365.35
	ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5 [†]	80.94
	ST_t-channel_top_5f_TuneCP5 [†]	136.02
	ST_tW_antitop_5f_inclusiveDecays_TuneCP5	35.85
VV	ST_tW_top_5f_inclusiveDecays	35.85
	VVTo2L2Nu_13TeV_amcatnloFXFX_madspin	13.84
	WZTo2L2Q_13TeV_amcatnloFXFX_madspin	5.52
	WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8	4.43
	ZZTo2L2Q_13TeV_amcatnloFXFX_madspin	3.38
W	ZZTo4L_TuneCP5_13TeV-amcatnloFXFX-pythia8	1.212
	W1JetsToLNu_TuneCP5	8104.0 (LO)
	W2JetsToLNu_TuneCP5	2793.0 (LO)
	W3JetsToLNu_TuneCP5	992.5 (LO)
	W4JetsToLNu_TuneCP5	544.3 (LO)
H	WJetsToLNu_TuneCP5	52940.0 (LO)
	GluGluHToTauTau_M125	48.58*0.0627
	GluGluHToWWTo2L2Nu_M125 ^{††}	48.58*0.2137*0.3258*0.3258
	GluGluZH_HToWW_M125	0.1227*0.2137
	HWminusJ_HToWW_M125 ^{†††}	0.5328*0.2137
	HWplusJ_HToWW_M125 ^{†††}	0.840*0.2137
	HZJ_HToWW_M125 ^{††}	0.7612*0.2137
	VBFHToTauTau_M125	3.782*0.0627
	VBFHToWWTo2L2Nu_M125 ^{†††}	3.782*0.2137*0.3258*0.3258
	WminusHToTauTau_M125	0.5328*0.0627
	WplusHToTauTau_M125	0.840*0.0627
	ZHToTauTau_M125	0.7612*0.0627
	ggZH_HToTauTau_ZToLL_M125	0.1227*0.0627*3*0.033658
	ggZH_HToTauTau_ZToNuNu_M125	0.1227*0.0627*0.2000
	ggZH_HToTauTau_ZToQQ_M125	0.1227*0.0627*0.6991
	ttHTobb_M125_TuneCP5	0.5071*(1-0.5824)
	ttHTobb_M125_TuneCP5	0.5071*0.5824

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

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

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

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

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

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

Table A.12: Signal samples used in the analysis for the 2018 era. All belong to the RunIIIAutumn18MiniAOD campaign and are produced with Madgraph and Pythia8. The second column is the number of events after the generator-level filter is applied, and the third column is the filter efficiency (percentage of all events that pass the generator-level filter).

²⁶³³ Bibliography

- ²⁶³⁴ [1] Paul H. Frampton. Journeys Beyond the Standard Model. 54(1):52–
²⁶³⁵ 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.
- ²⁶³⁸ [2] Meinard Kuhlmann. Quantum Field Theory. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research
²⁶³⁹ Lab, Stanford University, Summer 2023 edition, 2023.
- ²⁶⁴¹ [3] Steven Weinberg. The Making of the Standard Model. *Eur. Phys. J. C*, 34:
²⁶⁴² 5–13, 2004. URL <https://cds.cern.ch/record/799984>.
- ²⁶⁴³ [4] Christopher G. Tully. *Elementary Particle Physics in a Nutshell*. Princeton
²⁶⁴⁴ University Press, Princeton, 2012. ISBN 9781400839353. doi: doi:10.1515/
²⁶⁴⁵ 9781400839353. URL <https://doi.org/10.1515/9781400839353>.
- ²⁶⁴⁶ [5] John Ellis. Higgs Physics. In *2013 European School of High-Energy Physics*,
²⁶⁴⁷ pages 117–168, 2015. doi: 10.5170/CERN-2015-004.117.
- ²⁶⁴⁸ [6] David Curtin, Rouven Essig, Stefania Gori, and Others. Exotic decays of the 125
²⁶⁴⁹ GeV Higgs boson. *Phys. Rev. D*, 90:075004, Oct 2014. doi: 10.1103/PhysRevD.
²⁶⁵⁰ 90.075004. URL <https://link.aps.org/doi/10.1103/PhysRevD.90.075004>.
- ²⁶⁵¹ [7] Tania Robens, Tim Stefański, and Jonas Wittbrodt. Two-real-scalar-singlet

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

- 2676 [17] Olivier S. Brning and Frank Zimmerman. Parameter space for the LHC lu-
2677 minosity upgrade. ISBN 978-3-95450-115-1. URL <https://accelconf.web.cern.ch/IPAC2012/papers/moppc005.pdf>.
- 2679 [18] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09):
2680 P09018, 2020. doi: 10.1088/1748-0221/15/09/P09018.
- 2681 [19] CMS Collaboration. Measurement of the inelastic proton-proton cross section
2682 at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018. doi: 10.1007/JHEP07(2018)161.
- 2683 [20] CMS Collaboration. High-Luminosity Large Hadron Collider (HL-LHC): Tech-
2684 nical design report. 10/2020, 12 2020. doi: 10.23731/CYRM-2020-0010.
- 2685 [21] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:
2686 S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004.
- 2687 [22] CMS Collaboration. Particle-flow reconstruction and global event description
2688 with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/
2689 12/10/P10003.
- 2690 [23] V Karimki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudens-
2691 reich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker sys-
2692 tem project: Technical Design Report*. Technical design report. CMS. CERN,
2693 Geneva, 1997. URL <https://cds.cern.ch/record/368412>.
- 2694 [24] The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva,
2695 2017. URL <https://cds.cern.ch/record/2272264>.
- 2696 [25] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Up-
2697 grade. 9 2012. doi: 10.2172/1151650.
- 2698 [26] R. L. Workman and Others. Review of particle physics. 2022:083C01. doi:
2699 10.1093/ptep/ptac097.

- 2700 [27] CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter. 9 2012. doi: 10.2172/1151651.
- 2701
- 2702 [28] CMS Technical Design Report for the Level-1 Trigger Upgrade. 6 2013.
- 2703
- 2704 [29] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems. 12 2000.
- 2705
- 2706 [30] Alex Tapper. The CMS Level-1 Trigger for LHC Run II. *PoS*, ICHEP2016:242, 2016. doi: 10.22323/1.282.0242.
- 2707
- 2708 [31] A. Zabi, F. Beaudette, L. Cadamuro, O. Davignon, T. Romantreau, T. Strebler, M. Cepeda, J.B. Sauvan, N. Wardle, R. Aggleton, F. Ball, J. Brooke, D. Newbold, S. Paramesvaran, D. Smith, J. Taylor, C. Foudas, M. Baber, A. Bundred, S. Breeze, M. Citron, A. Elwood, G. Hall, G. Iles, C. Laner, B. Penning, A. Rose, A. Shtipliyski, A. Tapper, I. Ojalvo, T. Durkin, K. Harder, S. Harper, C. Shepherd-Themistocleous, A. Thea, T. Williams, S. Dasu, L. Dodd, R. Forbes, T. Gorski, P. Klabbers, A. Levine, T. Ruggles, N. Smith, W. Smith, A. Svetek, J. Tikalsky, and M. Vicente. The cms level-1 calorimeter trigger for the lhc run ii. *Journal of Instrumentation*, 12(01):C01065, jan 2017. doi: 10.1088/1748-0221/12/01/C01065. URL <https://dx.doi.org/10.1088/1748-0221/12/01/C01065>.
- 2711
- 2712
- 2713
- 2714
- 2715
- 2716
- 2717
- 2718
- 2719
- 2720
- 2721
- 2722
- 2723
- [32] P. Klabbers et al. CMS level-1 upgrade calorimeter trigger prototype development. *JINST*, 8:C02013, 2013. doi: 10.1088/1748-0221/8/02/C02013.
- [33] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger. Technical report, CERN, Geneva, 2021. URL <https://cds.cern.ch/record/2759072>. This is the final version of the document, approved by the LHCC.

- 2724 [34] C. Foudas. The CMS Level-1 Trigger at LHC and Super-LHC. In *34th Inter-*
2725 *national Conference on High Energy Physics*, 10 2008.
- 2726 [35] CMS Software Guide. High Level Trigger (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHighLevelTrigger>.
- 2728 [36] The Worldwide LHC Computing Grid. 2012. URL <https://cds.cern.ch/record/1997398>.
- 2730 [37] Douglas Thain, Todd Tannenbaum, and Miron Livny. Distributed computing
2731 in practice: the Condor experience. *Concurrency and Computation: Practice*
2732 and *Experience*, 17(2-4):323–356, 2005. doi: <https://doi.org/10.1002/cpe.938>.
2733 URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.938>.
- 2734 [38] Alexandre Zabi, Jeffrey Wayne Berryhill, Emmanuelle Perez, and Alexander D.
2735 Tapper. The Phase-2 Upgrade of the CMS Level-1 Trigger. 2020.
- 2736 [39] Technical proposal for a MIP timing detector in the CMS experiment Phase 2
2737 upgrade. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2296612>.
- 2739 [40] CMS Collaboration. Search for exotic decays of the Higgs boson to a pair of
2740 pseudoscalars in the $\mu\mu bb$ and $\tau\tau bb$ final states. *European Physical Journal C*,
2741 2 2024.
- 2742 [41] CMS Collaboration. CMS luminosity measurement for the 2016 data-taking
2743 period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL
2744 <https://cds.cern.ch/record/2257069>.
- 2745 [42] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking
2746 period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-
2747 004, 2018. URL <https://cds.cern.ch/record/2621960>.

- 2748 [43] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking
2749 period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-
2750 002, 2019. URL <https://cds.cern.ch/record/2676164>.
- 2751 [44] CMS LUMI Group. CMS Luminosity Public Results (TWiki), 2024. URL
2752 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- 2753 [45] J. Alwall, R. Frederix, S. Frixione, et al. The automated computation of tree-
2754 level and next-to-leading order differential cross sections, and their matching
2755 to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July
2756 2014. ISSN 1029-8479. doi: 10.1007/jhep07(2014)079. URL [http://dx.doi.org/10.1007/JHEP07\(2014\)079](http://dx.doi.org/10.1007/JHEP07(2014)079).
- 2758 [46] R. Frederix, S. Frixione, V. Hirschi, et al. The automation of next-to-leading
2759 order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July
2760 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).
- 2762 [47] S. Agostinelli, J. Allison, K. Amako, et al. Geant4 - a simulation toolkit. 506(3):
2763 250–303. ISSN 0168-9002. doi: 10.1016/S0168-9002(03)01368-8. URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- 2765 [48] CMS Collaboration. An embedding technique to determine $\tau\tau$ backgrounds
2766 in proton-proton collision data. *JINST*, 14(06):P06032, 2019. doi: 10.1088/
2767 1748-0221/14/06/P06032.
- 2768 [49] CMS Collaboration. Search for neutral MSSM Higgs bosons decaying to a pair
2769 of tau leptons in pp collisions. *JHEP*, 10:160, 2014. doi: 10.1007/JHEP10(2014)
2770 160.
- 2771 [50] CMS Collaboration. Measurements of Higgs boson production in the decay

channel with a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV.

Eur. Phys. J. C, 83(7):562, 2023. doi: 10.1140/epjc/s10052-023-11452-8.

- [51] CMS Collaboration. Reconstruction and identification of tau lepton decays to hadrons and tau neutrinos at CMS. *Journal of Instrumentation*, 11(01):P01019–P01019, January 2016. ISSN 1748-0221. doi: 10.1088/1748-0221/11/01/p01019. URL <http://dx.doi.org/10.1088/1748-0221/11/01/P01019>.

- [52] CMS Collaboration. Performance of τ -lepton reconstruction and identification in CMS. *Journal of Instrumentation*, 7(01):P01001, jan 2012. doi: 10.1088/1748-0221/7/01/P01001. URL <https://dx.doi.org/10.1088/1748-0221/7/01/P01001>.

- [53] CMS Collaboration. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):P10005, 2018. doi: 10.1088/1748-0221/13/10/P10005.

- [54] CMS Collaboration. Identification of hadronic tau lepton decays using a deep neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/P07023.

- [55] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 7(10):P10002–P10002, October 2012. ISSN 1748-0221. doi: 10.1088/1748-0221/7/10/p10002. URL <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.

- [56] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(06):P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005. URL <https://dx.doi.org/10.1088/1748-0221/10/06/P06005>.

- 2796 [57] CMS Collaboration. Identification of b-quark jets with the CMS experiment.
2797 *Journal of Instrumentation*, 8(04):P04013–P04013, April 2013. ISSN 1748-
2798 0221. doi: 10.1088/1748-0221/8/04/p04013. URL <http://dx.doi.org/10.1088/1748-0221/8/04/P04013>.
- 2800 [58] CMS Collaboration. Pileup Removal Algorithms. Technical report, CERN,
2801 Geneva, 2014. URL <https://cds.cern.ch/record/1751454>.
- 2802 [59] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and
2803 developments. 2017. URL <https://cds.cern.ch/record/2263802>.
- 2804 [60] CMS Collaboration. Performance of the DeepJet b tagging algorithm using
2805 41.9 fb^{-1} of data from proton-proton collisions at 13 TeV with Phase 1 CMS
2806 detector. 2018. URL <https://cds.cern.ch/record/2646773>.
- 2807 [61] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Re-
2808 construction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood
2809 techniques. *Journal of Physics: Conference Series*, 513(2):022035, jun 2014.
2810 doi: 10.1088/1742-6596/513/2/022035. URL <https://dx.doi.org/10.1088/1742-6596/513/2/022035>.
- 2812 [62] CMS Collaboration. Evidence for the 125 GeV Higgs boson decaying to a pair
2813 of τ leptons. *JHEP*, 05:104, 2014. doi: 10.1007/JHEP05(2014)104.
- 2814 [63] CMS Collaboration. Missing transverse energy performance of the CMS detec-
2815 tor. *JINST*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.
- 2816 [64] Artur Kalinowski. CMS AN-19-032 (internal): Reconstruction of a τ pair in-
2817 variant mass with a simplified likelihood scan, 2019.
- 2818 [65] CMS TAU POG. Tau Physics Object Group: Tau ID Recommendation

- 2819 For Run 2, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2>.
- 2821 [66] CMS MUO POG. Muon Physics Object Group: Recommendations, 2024. URL
2822 <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>.
- 2823 [67] CMS MUO POG. Muon Physics Object Group: Reference guidelines and results
2824 for muon momentum scale and resolution in Run II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceScaleResolRun2>.
- 2826 [68] CMS HTT working group. Higgs To Tau Tau Working TWiki for the full Run-
2827 2 legacy analysis, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/HiggsToTauTauWorkingLegacyRun2>.
- 2829 [69] CMS ELE POG. Electron Physics Object Group: Recommenda-
2830 tions, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations>.
- 2832 [70] CMS ELE POG. Electron Physics Object Group: Recommendations for
2833 2016 to 2018 UL, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018>.
- 2835 [71] CMS TAU Embedding Group. Tau embedded samples using 2016
2836 data, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauTauEmbeddingSamples2016Legacy>.
- 2838 [72] CMS TAU Embedding Group. Tau embedded samples using 2017
2839 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2017>.
- 2841 [73] CMS TAU Embedding Group. Tau embedded samples using 2018

- 2842 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/>
2843 TauTauEmbeddingSamples2018.
- 2844 [74] Tau Lepton Run 2 Trigger Performance. 2019. URL <https://cds.cern.ch/record/2678958>.
- 2846 [75] CMS Taus High Level Trigger Studies. Tau Lepton Run 2 Trigger Performance
2847 (CMS DP-2019/012), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTauAllRun2>.
- 2849 [76] Muon HLT Performance with 2018 Data. 2018. URL <https://cds.cern.ch/record/2627469>.
- 2851 [77] Single and Double Electron Trigger Efficiencies using the full Run 2 dataset.
2852 2020. URL <https://cds.cern.ch/record/2888577>.
- 2853 [78] Run II Trigger Performance For $e\mu$ Triggers. 2019. URL <https://cds.cern.ch/record/2687013>.
- 2855 [79] Performance of electron and photon reconstruction in Run 2 with the CMS
2856 experiment. 2020. URL <https://cds.cern.ch/record/2725004>.
- 2857 [80] CMS Collaboration. Performance of the CMS muon detector and muon recon-
2858 struction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015,
2859 2018. doi: 10.1088/1748-0221/13/06/P06015.
- 2860 [81] Piet Verwilligen. Muons in the cms high level trigger system. *Nuclear
2861 and Particle Physics Proceedings*, 273-275:2509–2511, 2016. ISSN 2405-6014.
2862 doi: <https://doi.org/10.1016/j.nuclphysbps.2015.09.441>. URL <https://www.sciencedirect.com/science/article/pii/S240560141500930X>. 37th Inter-
2863 national Conference on High Energy Physics (ICHEP).
- 2864

- 2865 [82] Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-
2866 probe technique. 2020. URL <https://cds.cern.ch/record/2724492>.
- 2867 [83] V.M. Abazov, B. Abbott, M. Abolins, et al. A novel method for modeling
2868 the recoil in W boson events at hadron colliders. *Nuclear Instruments and*
2869 *Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors*
2870 *and Associated Equipment*, 609(2):250–262, 2009. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2009.08.056>. URL <https://www.sciencedirect.com/science/article/pii/S0168900209016623>.
- 2873 [84] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair of
2874 light pseudoscalars in the final state with two b quarks and two τ leptons in
2875 proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 785:462, 2018. doi:
2876 [10.1016/j.physletb.2018.08.057](https://doi.org/10.1016/j.physletb.2018.08.057).
- 2877 [85] CMS LUMI POG. Luminosity Physics Object Group: Recommendations, 2024.
2878 URL <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.
- 2879 [86] The modeling of the top quark p_T (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting>.
- 2881 [87] CMS BTV group. Methods to apply b-tagging efficiency scale fac-
2882 tors (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration>.
- 2884 [88] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in
2885 pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. doi: [10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014).
- 2887 [89] Garvita Agarwal. Jet Energy Scale and Resolution Measurements in CMS. *PoS*,
2888 ICHEP2022:652, 2022. doi: [10.22323/1.414.0652](https://doi.org/10.22323/1.414.0652).

- 2889 [90] CMS JERC group. Jet Energy Corrections (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- 2890
2891 [91] CMS JERC group. Jet Energy Resolution (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- 2892
2893 [92] CMS MUO POG. Muon Physics Object Group: Baseline muon selec-
2894 tions for Run-II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/>
2895 [SWGuideMuonIdRun2](#).
- 2896 [93] CMS ELE POG. Electron Identification Based on Simple Cuts, 2024. URL
2897 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaPublicData>.
- 2898 [94] Jose Enrique Palencia Cortezon. Single top quark production at CMS. Technical
2899 report, CERN, Geneva, 2018. URL <https://cds.cern.ch/record/2640578>.
- 2900 [95] CMS Collaboration. Measurements of the electroweak diboson production cross
2901 sections in proton-proton collisions at $\sqrt{s} = 5.02$ TeV using leptonic decays.
2902 *Phys. Rev. Lett.*, 127(19):191801, 2021. doi: 10.1103/PhysRevLett.127.191801.
- 2903 [96] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment
2904 ten years after the discovery. *Nature*, 607(7917):60–68, 2022. doi: 10.1038/
2905 s41586-022-04892-x.
- 2906 [97] CMS JERC group. Jet energy scale uncertainty sources (TWiki), 2024. URL
2907 <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources>.
- 2908 [98] TOP Systematic Uncertainties (Run 2) (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics>.
- 2909
2910 [99] Search for a non-standard-model Higgs boson decaying to a pair of new light
2911 bosons in four-muon final states. Technical report, CERN, Geneva, 2013. URL
2912 <https://cds.cern.ch/record/1563546>.

- 2913 [100] Michiel Botje et al. The PDF4LHC Working Group Interim Recommendations.
2914 1 2011.
- 2915 [101] Cécile Caillol, Pallabi Das, Sridhara Dasu, Pieter Everaerts, Stephanie Kwan,
2916 Isobel Ojalvo, and Ho-Fung Tsoi. CMS AN-20-213 (internal): Search for an
2917 exotic decay of the 125 GeV Higgs boson to light pseudoscalars, with a pair of
2918 b jets and a pair of tau leptons in the final state, 2020.
- 2919 [102] Kyle Cranmer. Practical Statistics for the LHC. In *2011 European School of*
2920 *High-Energy Physics*, pages 267–308, 2014. doi: 10.5170/CERN-2014-003.267.
- 2921 [103] Elham Khazaie, Maryam Zeinali, Hamed Bakhshiansohi, and Abideh Jafari.
2922 CMS AN-21-058 (internal): Search for exotic decays of the Higgs boson to a
2923 pair of new light bosons with two muons and two b jets in the final states at
2924 $\sqrt{s} = 13$ TeV, 2021.
- 2925 [104] CMS Higgs Physics Analysis Group. Summary of 2HDM+S searches at 13 TeV
2926 (Run 2), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Summary2HDMRun2>.