

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

Open questions in particle physics may be addressed by the existence of an extended Higgs sector beyond the Standard Model Higgs boson with mass 125 GeV, which was discovered in 2012 at the Large Hadron Collider (LHC) by the CMS and ATLAS experiments. 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 different search 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.

⁴²

Acknowledgements

⁴³ Placeholder acknowledgements.

⁴⁵ Contents

⁴⁶	Abstract	iii
⁴⁷	Acknowledgements	iv
⁴⁸	List of Tables	xi
⁴⁹	List of Figures	xiii
⁵⁰	1 Introduction	1
⁵¹	1.1 History of the Standard Model	1
⁵²	1.2 The Standard Model as a gauge theory	3
⁵³	1.2.1 Gauge invariance	3
⁵⁴	1.2.2 Local gauge symmetries	5
⁵⁵	1.3 The Higgs Mechanism	7
⁵⁶	1.4 Two-Higgs Doublet Models	9
⁵⁷	1.5 Two Real Singlet Model	11
⁵⁸	2 The Large Hadron Collider and the CMS Experiment	15
⁵⁹	2.1 The Large Hadron Collider	16
⁶⁰	2.2 Luminosity and pileup	17
⁶¹	2.3 The High-Luminosity LHC	20
⁶²	2.4 The CMS Detector	21
⁶³	2.5 Sub-detectors of CMS	22
⁶⁴	2.5.1 Inner tracking system	23

65	2.5.2	ECAL	24
66	2.5.3	HCAL	25
67	2.5.4	Muon detectors	27
68	2.5.5	The Level-1 Trigger	28
69	2.5.6	The High-Level Trigger	31
70	2.5.7	Particle reconstruction	32
71	2.5.8	Data storage and computational infrastructure	32
72	3	The Phase-2 Upgrade of CMS	33
73	3.1	High-Luminosity LHC and CMS	33
74	3.2	The Phase-2 Level-1 Trigger	33
75	3.3	Standalone Barrel Calorimeter electron/photon reconstruction	36
76	3.3.1	Phase-2 geometry of the ECAL Barrel trigger	36
77	3.3.2	Phase-2 electron/photon reconstruction algorithm	37
78	4	Datasets and Monte Carlo samples	46
79	4.1	Datasets used	46
80	4.2	Monte Carlo samples	46
81	4.3	Embedded samples	48
82	5	Object reconstruction and corrections applied	51
83	5.1	Object reconstruction	51
84	5.1.1	Taus	51
85	5.1.2	Muons	55
86	5.1.3	Electrons	56
87	5.1.4	Jets	58
88	5.1.5	B-flavored jets	59
89	5.2	Reconstruction of the $\tau\tau$ mass	60
90	5.2.1	Original SVFit “standalone”: maximum likelihood	61

91	5.2.2 “Classic SVFit” with matrix element	62
92	5.2.3 FastMTT: optimized SVFit	63
93	5.3 Corrections applied to simulation	63
94	5.3.1 Tau energy scale	64
95	5.3.2 Muon energy scale	64
96	5.3.3 Electron energy scale	65
97	5.3.4 τ_h identification efficiency	66
98	5.3.5 Trigger efficiencies	66
99	5.3.6 Tau trigger efficiencies	66
100	5.3.7 Single muon trigger efficiencies	67
101	5.3.8 Single electron trigger efficiencies	68
102	5.3.9 $e\mu$ cross-trigger efficiencies	69
103	5.3.10 Electrons and muons faking τ_h : energy scales	70
104	5.3.11 Electrons and muons faking τ_h : misidentification efficiencies . .	71
105	5.3.12 Electron ID and tracking efficiency	72
106	5.3.13 Muon ID, isolation, and tracking efficiencies	73
107	5.3.14 Recoil corrections	74
108	5.3.15 Drell-Yan corrections	75
109	5.3.16 Pileup reweighing	75
110	5.3.17 Pre-firing corrections	76
111	5.3.18 Top p_T spectrum reweighing	77
112	5.3.19 B-tagging efficiency	77
113	5.3.20 Jet energy resolution and jet energy smearing	77
114	6 Event selection	79
115	6.1 General procedure for all channels	79
116	6.2 Event selection in the $\mu\tau_h$ channel	80
117	6.3 Event selection in the $e\tau_h$ channel	82

118	6.4	Event selection in the $e\mu$ channel	83
119	6.5	Extra lepton vetoes in all channels	84
120	7	Background estimation	88
121	7.1	Z+jets	88
122	7.2	W+jets	89
123	7.3	$t\bar{t}$ + jets	89
124	7.4	Single top	90
125	7.5	Diboson	90
126	7.6	Standard Model Higgs	90
127	7.7	Jet faking τ_h	91
128	7.8	QCD multijet background	92
129	8	Systematic uncertainties	94
130	8.1	Uncertainties associated with physics objects	95
131	8.1.1	Uncertainties in the lepton energy scales	95
132	8.1.2	Uncertainties from other lepton corrections	96
133	8.1.3	Uncertainties from jet energy scale and resolution	96
134	8.1.4	Uncertainties from b-tagging scale factors	97
135	8.1.5	Uncertainties from MET	98
136	8.2	Uncertainties associated with samples used	98
137	8.3	Other uncertainties	99
138	8.4	Pulls and impacts	99
139	9	Event categorization and signal extraction	101
140	9.1	B-tag jet multiplicity	101
141	9.2	DNN-based event categorization	101
142	9.3	Methodology for signal extraction	104
143	9.3.1	Model building and parameter estimation	105

144	9.3.2 Hypothesis testing	107
145	9.3.3 Confidence intervals	108
146	9.3.4 Profile likelihood ratio	110
147	9.3.5 Modified frequentist method: CL_S	111
148	10 Results	112
149	10.1 Results from $bb\tau\tau$	112
150	10.2 Combination with $bb\mu\mu$ final state	114
151	11 Asymmetric exotic Higgs decays	125
152	11.1 Signal masses	125
153	11.2 Cascade scenario signal studies	126
154	11.3 Current control plots for $\mu\tau_h$ channel	128
155	12 Conclusion and outlook	132

¹⁵⁶ List of Tables

¹⁵⁷	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.	50
¹⁶¹	5.1	Energy scales applied to genuine hadronic tau decays τ_h by data-taking year/era and decay mode, along with systematic errors.	64
¹⁶²	5.2	Energy scales and systematic errors applied to genuine muons.	65
¹⁶³	5.3	Energy scales and systematic errors applied to electrons in embedded samples by data-taking year/era.	66
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	66
¹⁶⁶	5.5	Energy scales and up/down systematic uncertainties applied to electrons misidentified as hadronic taus.	70
¹⁶⁷	5.6	Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018.	72
¹⁶⁹	5.7	Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018.	72

174	6.1	Trigger thresholds used for the leptons in the $bb\mu\mu$ analysis and the	
175		$bb\tau\tau$ analysis (the focus of this work). The thresholds for the three $bb\tau\tau$	
176		channels ($e\mu$, $e\tau_h$, and $\mu\tau_h$) are listed separately, with some channels	
177		and years taking the logical OR of two triggers with different thresholds.	80
178	6.2	High-Level Trigger (HLT) paths used to select data and simulation	
179		events in 2016 for the three $\tau\tau$ channels.	85
180	6.3	High-Level Trigger (HLT) paths used to select data and simulation	
181		events in 2017 for the three $\tau\tau$ channels.	85
182	6.4	High-Level Trigger (HLT) paths used to select data and simulation	
183		events in 2018 for the three $\tau\tau$ channels. In 2018 a HLT trigger path	
184		using the hadron plus strips (HPS) tau reconstruction algorithm be-	
185		came available.	86
186	9.1	Event categorization based on DNN scores for events with exactly 1	
187		b-tag jet (1bNN), for the three $\tau\tau$ channels and three eras.	104
188	9.2	Event categorization based on DNN scores for events with 2 b-tag jets	
189		(2bNN), for the three $\tau\tau$ channels and three eras.	105

¹⁹⁰ List of Figures

¹⁹¹ 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
¹⁹⁶ 1.2	An illustration of the Higgs potential.	8
¹⁹⁷ 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
¹⁹⁸ 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
²⁰² 2.1	Aerial view of the Large Hadron Collider (LHC).	18
²⁰³ 2.2	Distribution of the mean number of inelastic collisions per bunch cross- ing (pileup) in data, for proton-proton collisions in 2016-2018	20
²⁰⁴ 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
²⁰⁶ 2.4	Cross section of the current Phase-1 CMS tracker.	24
²⁰⁷ 2.5	Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters.	26
²⁰⁹		

210	2.6	Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels.	28
211			
212	2.7	Dataflow for the Phase-1 Level-1 Trigger.	29
213			
214	3.1	Functional diagram of the CMS L1 Phase-2 upgraded trigger design. .	34
215	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.	37
216			
217	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.	40
218			
219	3.4	Schematic of two example RCT cards in the negative eta (<i>top</i>) and positive eta (<i>center</i>) regions of the ECAL barrel. Each RCT card is divided into five regions: four regions are of size 3×4 towers in $\eta \times \phi$ (<i>bottom left</i>), and a fifth smaller overlap region of size 2×4 towers (<i>bottom right</i>). Each tower is 5×5 ($\eta \times \phi$) in crystals.	41
220			
221	3.5	Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trig- ger (GCT), which will process the outputs of the Regional Calorimeter Trigger (RCT) in three cards (<i>magenta highlights</i>). Each card in the GCT processes the equivalent of sixteen RCT cards, with the center twelve being unique to that GCT card (<i>shaded pink</i>), and the remain- ing four processed in overlap with the other GCT cards.	42
222			
223			
224			
225			
226			
227			
228			
229			
230			
231			
232			
233			

234	3.6 Illustration of an example electron/photon (e/γ) cluster in the Phase-	
235	2 Level-1 Trigger standalone barrel e/γ reconstruction, in a region of	
236	15×20 crystals (3×4 towers). Each small pink square is one crystal,	
237	the highest-granularity ECAL trigger primitives available to the L1	
238	Trigger in Phase-2. The core cluster consists of the energy sum in a	
239	3×5 window of crystals, (<i>shaded light blue</i>) centered around the seed	
240	crystal (<i>red</i>). Bremsstrahlung corrections are checked in the adjacent	
241	3×5 windows in the ϕ direction (<i>shaded light yellow</i>). The relative	
242	energies in windows of size 2×5 and 5×5 in crystals (<i>dashed dark blue</i>	
243	<i>and dark red</i>) are used to compute shower shape variables to identify	
244	true e/γ objects. Lastly, an isolation sum is computed in a window of	
245	size 7×7 in towers (not shown in figure).	43
246	3.7 Efficiency of the standalone barrel e/γ reconstruction, as a function of	
247	the true electron's transverse momentum p_T	44
248	3.8 Rates of the standalone barrel e/γ reconstruction measured as a func-	
249	tion of the minimum energy (E_T) required of the reconstructed e/γ	
250	object in each event.	45
251	4.1 Cumulative delivered and recorded luminosity versus time for 2015-	
252	2018 at CMS, in proton-proton collision data only, at nominal center-	
253	of-mass energy.	47
254	4.2 Schematic view of the four main steps of the embedding technique for	
255	τ leptons.	49
256	5.1 Distributions of $m_{\tau\tau}$ reconstructed by the classic SVFit algorithm, and	
257	masses of visible tau decay products (before SVFit).	62
258	5.2 Electron/photon energy scale factors and uncertainties for 2018. . . .	65

259	5.3 Hadronic tau leg efficiency of the cross-triggers for $\mu\tau_h$ (<i>left</i>) and $e\tau_h$ 260 (<i>right</i>) triggers as a function of offline tau p_T for 2016, 2017, and 2018.	68
261	5.4 Trigger efficiencies in data (<i>top panels</i>) and ratio of efficiencies af- 262 ter/before a HLT muon reconstruction update (<i>bottom panels</i>) for the 263 muon in the isolated single muon trigger with threshold $p_T > 24$ GeV 264 in the data-taking year 2018, as functions of the muon p_T (<i>left</i>) and 265 muon $ \eta $ (<i>right</i>).	69
266	5.5 Trigger efficiencies in data and the data/MC ratio for the electron in 267 the single electron trigger with threshold $p_T > 32$ GeV in the data- 268 taking year 2018, as functions of the electron p_T (<i>left</i>) and electron $ \eta $ 269 (<i>right</i>).	70
270	5.6 Efficiencies of the electron leg vs. p_T (<i>left</i>) and the muon log vs. η 271 (<i>right</i>), for the HLT path with online thresholds of 12 GeV for the 272 electron and 23 GeV for the muon, with the data-taking years 2016 273 through 2018 overlaid.	71
274	5.7 Efficiencies in data (<i>top panels</i>) and the ratio of efficiencies in data/MC 275 (<i>bottom panels</i>), for the electron multivariate analysis (MVA) identifi- 276 cation (<i>left</i>) and for the Gaussian-sum filter (GSF) tracking (<i>right</i>). . .	73
277	5.8 Muon identification efficiencies in 2015 data and MC as a function of 278 the muon p_T for the loose ID (<i>left</i>) and tight ID (<i>right</i>) working points.	74
279	5.9 Muon isolation efficiencies in Run-2 data as a function of the muon p_T 280 (<i>left</i>) and $ \eta $ (<i>right</i>).	75
281	5.10 Muon tracking efficiencies as a function of $ \eta $ for standalone muons in 282 Run-2 data (<i>black</i>) and Drell-Yan (<i>blue</i>) MC simulation.	76
283	7.1 Leading-order Feynman diagrams of Higgs production.	91
284	8.1 Top sixty impacts for the combination of all channels and years. . . .	100

285	9.1 Schematic of the Neyman construction for confidence intervals.	109
286	10.1 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $\mu\tau_h$ chan-	
287	nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions.	116
288	10.2 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\tau_h$ chan-	
289	nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions.	117
290	10.3 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\mu$ channel.	118
291	10.4 Observed 95% CL exclusion limits (<i>black, solid lines</i>) and expected 95%	
292	CL and 68% CL limits (<i>shaded yellow and green</i>) on the branching	
293	fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard	
294	Model production for the 125 GeV Higgs (h). Limits are shown for the	
295	$\mu\tau_h$ channel (<i>top left</i>), the $e\tau_h$ channel (<i>top right</i>), and the $e\mu$ channel	
296	(<i>bottom left</i>), and lastly the combination of all three channels (<i>bottom</i>	
297	<i>right</i>) The dataset corresponds to 138 fb^{-1} of data collected in the	
298	years 2016-2018 at a center-of-mass energy 13 TeV.	119
299	10.5 Observed 95% CL upper limits on $B(h \rightarrow aa)$ in %, for the $bb\tau\tau$ final	
300	state (<i>left</i>) and $bb\mu\mu$ final state (<i>right</i>) using the full Run 2 integrated	
301	luminosity of 138 fb^{-1} in 2HDM+S type I (<i>blue</i>), type II with $\tan\beta =$	
302	2.0 (<i>orange dashed</i>), type III with $\tan\beta = 2.0$ (<i>dotted green</i>), and type	
303	IV with $\tan\beta = 0.6$ (<i>red dashed</i>). .	120
304	10.6 Observed 95% CL upper limits on the branching fraction of the 125	
305	GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$, in percentages,	
306	as a function of the pseudoscalar mass m_a , in 2HDM+S type I (<i>blue</i>),	
307	type II with $\tan\beta = 2.0$ (<i>orange dashed</i>), type III with $\tan\beta = 2.0$	
308	(<i>dotted green</i>), and type IV with $\tan\beta = 0.6$ (<i>red dashed</i>), for the	
309	combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated	
310	luminosity of 138 fb^{-1} .	121

311	10.7 Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination	
312	of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity	
313	of 138 fb^{-1} for Type II (<i>left</i>), Type III (<i>middle</i>), and Type IV (<i>right</i>)	
314	2HDM+S in the $\tan\beta$ vs. m_a phase space.	122
315	10.8 Summary plot of current observed and expected 95% CL limits on the	
316	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
317	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
318	$B(h \rightarrow aa)$, in the 2HDM+S type I scenario, obtained at CMS with	
319	data collected at 13 TeV.	122
320	10.9 Summary plot of current observed and expected 95% CL limits on the	
321	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
322	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
323	$B(h \rightarrow aa)$, in the 2HDM+S type II scenario with $\tan\beta = 2.0$, ob-	
324	tained at CMS with data collected at 13 TeV.	123
325	10.10 Summary plot of current observed and expected 95% CL limits on the	
326	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
327	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
328	$B(h \rightarrow aa)$, in the 2HDM+S type III scenario with $\tan\beta = 2.0$, ob-	
329	tained at CMS with data collected at 13 TeV.	124
330	11.1 Generator-level b-flavor jet transverse momenta p_T , for $h \rightarrow a_1 a_2$ cas-	
331	cade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) =$	
332	$(100, 15) \text{ GeV}$ (<i>left</i>) and $(40, 20) \text{ GeV}$ (<i>right</i>). In each plot the generator-	
333	level p_T of the leading (<i>black</i>), sub-leading (<i>red</i>), third (<i>blue</i>), and	
334	fourth (<i>light green</i>) are overlaid.	127

³⁵⁷ Chapter 1

³⁵⁸ Introduction

³⁵⁹ The Standard Model is the current prevailing theoretical framework that encompasses
³⁶⁰ all known elementary particles to date and describes their interactions, yet falls short
³⁶¹ of describing open problems in physics. Here, we describe the history of the Standard
³⁶² Model and its particle content (Section 1.1), and provide a mathematical motivation
³⁶³ of the SM a gauge theory (Section 1.2). We introduce the Higgs mechanism (Section
³⁶⁴ 1.3), and outline two groups of theoretical extensions to the Standard Model that
³⁶⁵ feature extended Higgs sectors (Sections 1.4 and 1.5).

³⁶⁶ 1.1 History of the Standard Model

³⁶⁷ The building blocks of our modern-day understanding of particle physics were estab-
³⁶⁸ lished over the course of many decades by experimental discoveries and theoretical
³⁶⁹ advances, culminating in the development of a theoretical framework known as the
³⁷⁰ Standard Model (SM). In the 1880s, the electron was the first subatomic particle to
³⁷¹ be identified, through measurements of particles produced by ionizing gas. By the
³⁷² 1930s, atoms were known to consist mostly of empty space, with protons and neutrons
³⁷³ concentrated at the center and orbited by electrons. Spurred by advances in parti-
³⁷⁴ 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

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

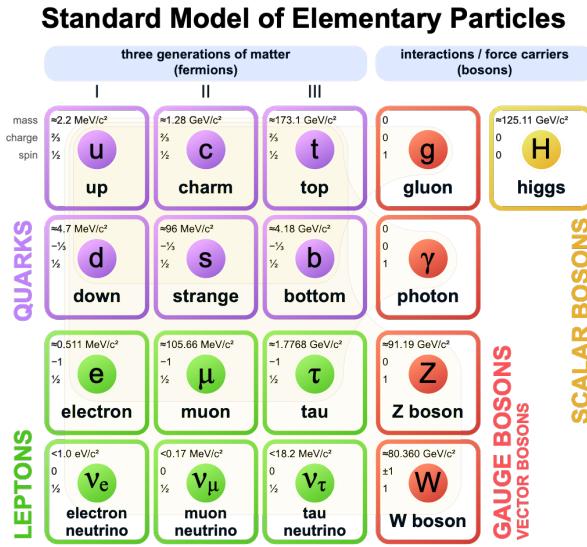


Figure 1.1: Table of Standard Model particles showing the grouping of the fermions into three generations of matter and the bosons, responsible for carrying the three fundamental forces in the Standard Model. The masses, charges, and spins of the particles are shown. The antimatter counterparts of the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

407 1.2 The Standard Model as a gauge theory

408 1.2.1 Gauge invariance

409 Gauge theories of elementary particle interactions originate from a freedom of choice
 410 in the mathematical description of particle fields which has no effect on the particles'
 411 physical states [4]. The existence and form of the particles' interactions, can be
 412 deduced from the existence of physically indeterminate, gaugable quantities.

413 An example of this gauge invariance in classical physics is the electromagnetic
 414 interaction, where the fundamental field is the four-vector potential A^μ [4]. The
 415 physical electromagnetic fields and Maxwell's equations arise from the elements of
 416 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
 417 transformation of the form

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

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

421 One important consequence of gauge symmetry comes from the application of
 422 Noether's theorem, which states that for every global transformation under which the
 423 Lagrangian density is invariant, there exists a conserved quantity. If $\mathcal{L}(\Psi(x), \partial_\mu \Psi(x))$
 424 is invariant under the transformation of the wave function $\Psi(x) \rightarrow \Psi'(x)$, where
 425 $\Psi'(x) = \Psi(x) + \delta\Psi(x)$, then there exists a conserved current

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

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

433 1.2.2 Local gauge symmetries

434 Interactions between particles arise if we modify the wave function with a phase
 435 transformation $\Psi'(x) = \exp(ie\chi)\Psi(x)$, and allow the phase χ to be a function of
 436 spacetime [4]. A wave function of the form

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

437 can be verified to *not* be a solution to the Dirac equation for free particles: $(i\gamma^\mu\partial_\mu -$
 438 $m)\Psi(x) = 0$. This necessitates a modified Dirac equation, where the derivative takes
 439 into account that the vector field $V(x)$ needs to be compared at two displaced space-
 440 time points in a curvilinear coordinate system:

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

441 We define a covariant derivative,

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

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

$$(i\gamma^\mu D_\mu - m)\Psi(x) = 0 \quad (1.6)$$

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

446 The generalization of this result is as follows: if a theory is invariant for unitary

⁴⁴⁷ transformations U of the particle states according to

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

⁴⁴⁸ One must define a derivative of the form

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

⁴⁴⁹ to keep the theory invariant under Eqn. 1.7. The four-potential B^μ represents the
⁴⁵⁰ interacting four-potential which must be added to keep the theory invariant.

⁴⁵¹ In the case of the Standard Model, the theory is built around the gauge trans-
⁴⁵² formations $G = SU(3) \times SU(2) \times U(1)$. $SU(3)$ is associated to the strong force
⁴⁵³ (subscripted C); $SU(2)$ is associated to the weak force (subscripted L); and $U(1)$ is
⁴⁵⁴ hypercharge (subscripted Y). The gauge-covariant derivative is

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

⁴⁵⁵ • In the $U(1)_Y$ term, B_μ is the weak hypercharge field.

⁴⁵⁶ • 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-
⁴⁵⁷ potentials. $\tau/2$ are the Pauli matrices, generators of the $SU(2)$ transformation.

⁴⁵⁸ • In the $SU(3)_C$ term, the gluon (color) field is G_μ . λ_k are the Gell-Man matrices,
⁴⁵⁹ generators of the $SU(3)$ transformation.

⁴⁶⁰ The invariance of the Standard Model under $SU(3)_C \times SU(2)_L \times U(1)_Y$ requires
⁴⁶¹ massless fermions and massless force carriers.

462 1.3 The Higgs Mechanism

463 To introduce mass into the theory, i.e. to change the propagation of the gauge par-
 464 ticles and all the fermions, the physical vacuum cannot have all the symmetries of
 465 the Standard Model Lagrangian [4]. The symmetries of the physical vacuum must
 466 be spontaneously broken, without affecting gauge invariance in the Lagrangian. The
 467 Higgs mechanism proposes the existence of a scalar field, or fields, with nonzero vac-
 468 uum expectation values, which reduce the gauge symmetries of the physical vacuum
 469 from $SU(3)_C \times SU(2)_L \times U(1)_Y$ down to $SU(3)_C \times U(1)_{EM}$.

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

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

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

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

484 The minimal self-interacting Higgs potential that is invariant under $SU(2)_L \times$

485 $U(1)_Y$ is given by

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

486 where λ is the coupling strength of the four-point Higgs interaction. The potential
487 energy is minimized at

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

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

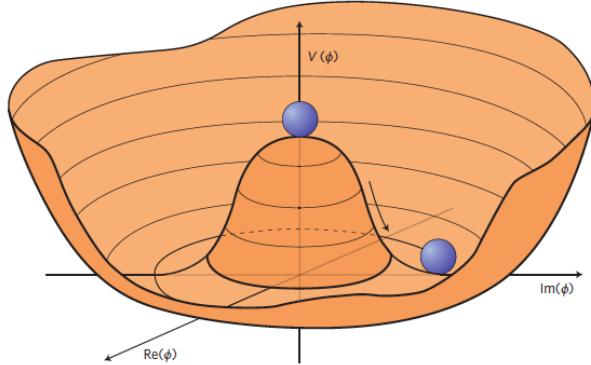


Figure 1.2: An illustration of the Higgs potential [5]. Choosing any of the points at the bottom of the potential breaks spontaneously the rotational $U(1)$ symmetry.

490 The excitations of the Higgs field with respect to the minimum Φ_{\min} are parameterized by
491

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

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

496 1.4 Two-Higgs Doublet Models

497 One of the simplest possible extensions to the Standard Model is adding a doublet
 498 to the minimal Higgs sector of the Standard Model, which is a $SU(2)_L$ doublet H
 499 with hypercharge $Y = +\frac{1}{2}$, denoted here as $H \sim 2_{+1/2}$. These extensions are found
 500 in several theories such as supersymmetry. A general 2HDM can be extended with a
 501 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:

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

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

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

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

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

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

509 We assume that the 2HDM is near or in the decoupling limit: $\alpha \rightarrow \pi/2 - \beta$, where
 510 the lightest state in the 2HDM is h , which we identify as the 125 GeV Higgs particle
 511 [6]. In this limit, the fermion couplings of h become identical to the Standard Model
 512 Higgs, while the gauge boson couplings are very close to Standard Model-like for
 513 $\tan \beta \gtrsim 5$. All of the properties of h are determined by just two parameters: $\tan \beta$
 514 and α , and the fermion couplings to the two Higgs doublets.

515 2HDM can be extended by a scalar singlet (2HDM+S) [6]:

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

516 If this singlet only couples to the Higgs doublets $H_{1,2}$ and has no direct Yukawa
 517 couplings, all of its couplings to SM fermions result from mixing with $H_{1,2}$. Under
 518 these simple assumptions, exotic Higgs decays $h \rightarrow ss \rightarrow X\bar{X}Y\bar{Y}$ or $h \rightarrow aa \rightarrow$
 519 $X\bar{X}Y\bar{Y}$, and $h \rightarrow aZ \rightarrow X\bar{X}Y\bar{Y}$ are permitted, where $s(a)$ is a (pseudo)scalar mass
 520 eigenstate mostly composed of $S_R(S_I)$, and X, Y are Standard Model fermions or
 521 gauge bosons. There are two pseudoscalars in the 2HDM+S, and the mostly singlet-
 522 like pseudoscalar can be chosen to be the one lighter than the SM-like Higgs. For
 523 $m_a < m_h - m_Z \sim 35$ GeV, the exotic Higgs decay $h \rightarrow Za$ is possible, and for
 524 $m_a < m_h/2 \approx 63$ GeV, the exotic Higgs decay $h \rightarrow aa$ is possible.

525 In 2HDM, and by extension 2HDM+S, there are four types of fermion couplings
 526 commonly discussed in the literature that forbid flavor-changing neutral currents at
 527 tree level [6]. These are referred to as Type I (all fermions couple to H_2), Type II
 528 (MSSM-like, d_R and e_R couple to H_1 , u_R to H_2), Type III (lepton-specific, leptons

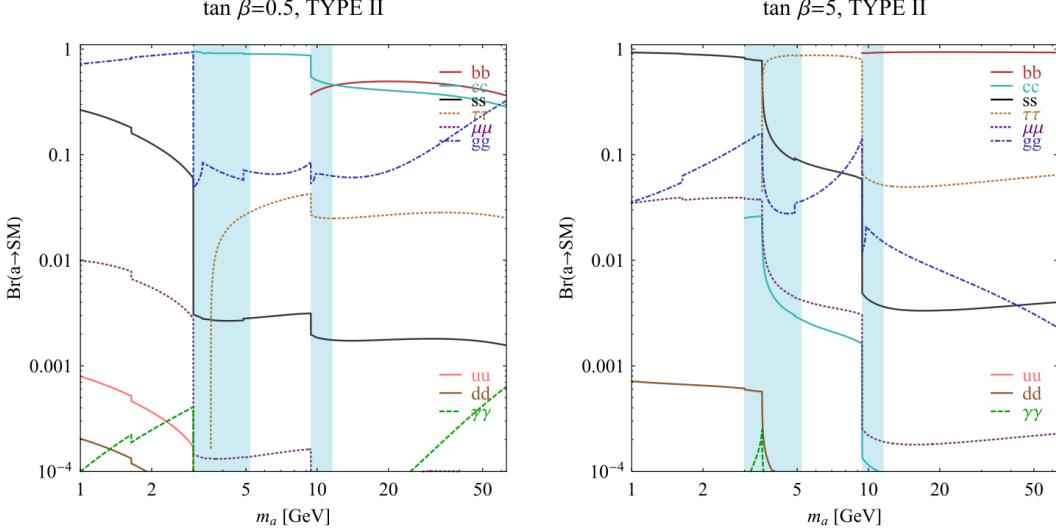


FIG. 7 (color online). Branching ratios of a singletlike pseudoscalar in the 2HDM + S for type-II Yukawa couplings. Decays to quarkonia likely invalidate our simple calculations in the shaded regions.

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.

and quarks couple to H_1 and H_2 respectively) and Type IV (flipped, with u_R , e_R coupling to H_2 and d_R to H_1). The exact branching ratios of the pseudoscalars to Standard Model particles vary depending on the 2HDM+S model and the value of $\tan\beta$ (e.g. Fig. 1.3).

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

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

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

543 The mass eigenstates $h_{1,2,3}$ are related to the fields $\phi_{h,S,X}$ through a 3×3 orthogonal
 544 mixing matrix denoted R . The mass eigenstates are assumed to be ordered $M_1 \leq$
 545 $M_2 \leq M_3$. R is parameterized by the three mixing angles $\theta_{hS}, \theta_{hX}, \theta_{SX}$. The nine
 546 parameters of the scalar potential can be expressed in terms of the three physical
 547 Higgs masses, the three mixing angles, and the three vacuum expectation values.

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

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

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

562 If the decay channel $h_2 \rightarrow h_1 h_1$ is kinematically open (i.e. $M_2 > 2M_1$), it is the

563 dominant decay mode leading to a significant rate for the $h_1 h_1 h_1$ final state, in a
564 “cascade” decay. In BP1, $BR(h_2 \rightarrow h_1 h_1) \simeq 100\%$ above the red line in Fig. 1.4. If,
565 in addition, $M_1 \gtrsim 10$ GeV, the h_1 decays dominantly to $b\bar{b}$ leading to a sizable rate
566 for the $b\bar{b} b\bar{b} b\bar{b}$ final state as shown in Fig. 1.4 (*bottom right*).

567 If the $h_2 \rightarrow h_1 h_1$ decay is kinematically closed (i.e. $M_2 < 2M_1$), both scalars decay
568 directly to Standard Model particles, with branching ratios identical to a Standard
569 Model-like Higgs boson, i.e. with the $b\bar{b} b\bar{b}$ final state dominating, as shown in Fig. 1.4
570 (*bottom left*), while at smaller masses, combinations with τ leptons and eventually
571 final states with charm quarks and muons become relevant [7].

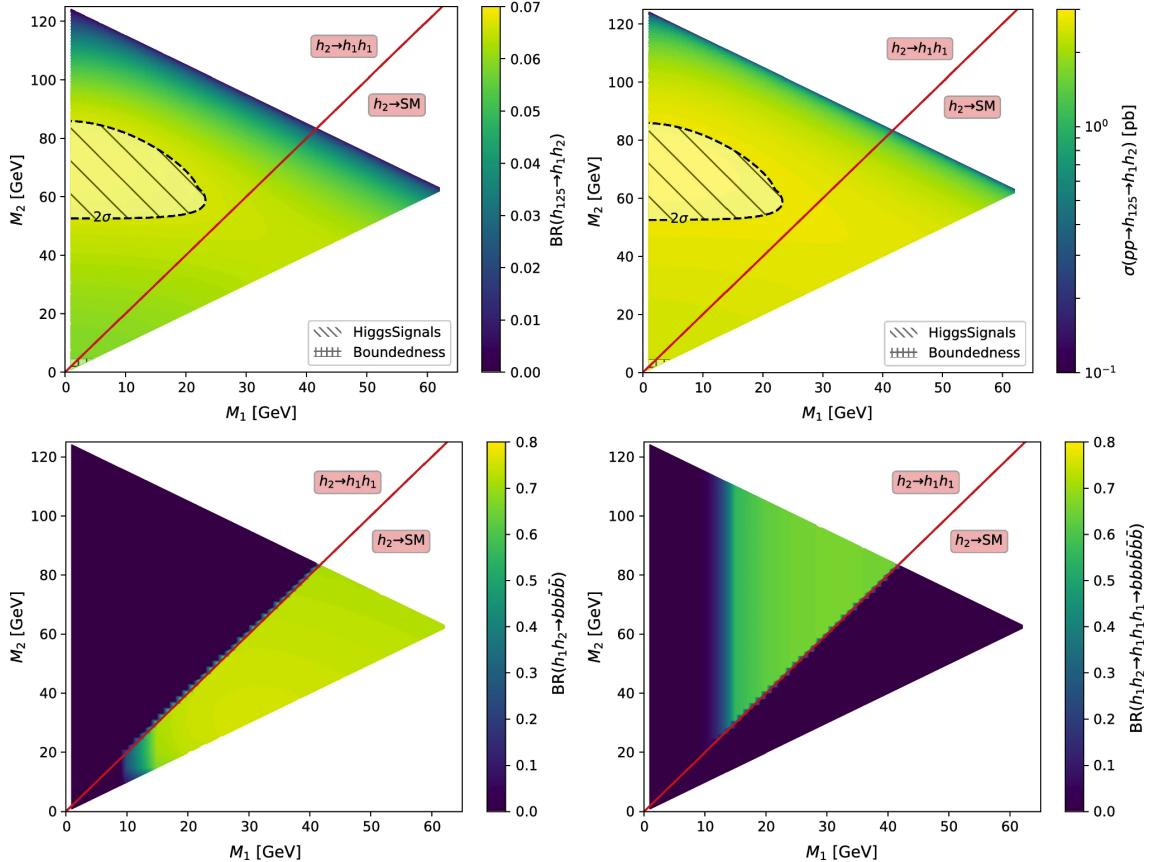


Figure 1.4: Benchmark plane BP1 for benchmark scenario 1 from [7], for the decay signature $h_{125} \rightarrow h_1 h_2$ with $h_{125} \equiv h_3$, defined in the (M_1, M_2) plane. The color code shows $\text{BR}(h_3 \rightarrow h_1 h_2)$ (*top left*) and the 13 TeV LHC signal rate for $pp \rightarrow h_3 \rightarrow h_1 h_2$ (*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.

⁵⁷² **Chapter 2**

⁵⁷³ **The Large Hadron Collider and the**
⁵⁷⁴ **CMS Experiment**

⁵⁷⁵ This chapter introduces the key aspects of the CERN Large Hadron Collider (LHC)
⁵⁷⁶ and the Compact Muon Solenoid (CMS) experiment where the work for this thesis was
⁵⁷⁷ conducted. Section 2.1 describes the history of accelerator developments at CERN
⁵⁷⁸ that led to the construction of the LHC, the current LHC configuration, and the
⁵⁷⁹ largest experiments located at the LHC. The concepts of beam luminosity and pileup,
⁵⁸⁰ which are critical for understanding and measuring high-energy particle collisions,
⁵⁸¹ are described in Section 2.2 and discussed in the context of the High-Luminosity
⁵⁸² LHC (HL-LHC) upgrade in Section 2.3. Lastly, Section 2.4 describes the design
⁵⁸³ and function of CMS and its subdetectors, and terminates in a description of data
⁵⁸⁴ processing at CMS, beginning from online event filtering in the Level-1 Trigger, to
⁵⁸⁵ processing in the High-Level Trigger, to offline particle reconstruction, and finally
⁵⁸⁶ long-term storage and processing of measured events.

587 2.1 The Large Hadron Collider

588 CERN, the European Organization for Nuclear Research, is an international organiza-
589 tion based in Meyrin, Switzerland which operates the world's largest particle physics
590 laboratory, and is the site of the Large Hadron Collider (LHC) [8]. The very first
591 accelerator built at CERN was the 600 MeV Synchrocyclotron (SC), which initially
592 provided beams for CERN's first experiments. The newer and more powerful Proton
593 Synchrotron (PS), which could accelerate particles to an energy of 28 GeV, began op-
594 erations in 1959 and is still in use today. The first hadron collider at CERN was the
595 Intersecting Storage Rings (ISR), which consisted of two interlaced rings each with a
596 diameter of 200. The ISR collided protons at a center-of-mass energy of 62 GeV and
597 began measuring collisions in 1971. In 1968 CERN began to accelerate heavy ions
598 in the Super Proton Synchrotron (SPS), which is 7 kilometers in circumference and
599 was the first of CERN's giant underground rings to be built. The SPS became the
600 forefront of CERN's particle physics program in 1976, and in 1981 was converted into
601 a proton-antiproton collider. The final and largest underground ring constructed at
602 CERN was the Large Electron-Positron (LEP) collider, which was commissioned in
603 July 1989 and hosted 5176 magnets and 128 accelerating cavities located around a
604 27-kilometer circumference. Over 11 years of research, four detectors, ALEPH, DEL-
605 PHI, L3, and OPAL measured the collisions, with collision energies reaching up to
606 209 GeV in the year 2000. In November 2000, LEP was closed down to make way for
607 the construction of the LHC in the same tunnel.

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

614 the Proton Synchrotron (PS), and 450 GeV in the Super Proton Synchrotron (SPS).
615 The protons are transferred to the two beam pipes of the Large Hadron Collider
616 (LHC). The LHC is a 27-kilometer ring of superconducting magnets, inside which
617 one beam circulates clockwise and the other counterclockwise. Each LHC ring takes
618 4 minutes and 20 seconds to fill, and it takes about 20 minutes for the protons to
619 reach their maximum energy. During normal operating conditions, beams circulate
620 for many hours inside the LHC ring.

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

633 2.2 Luminosity and pileup

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

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

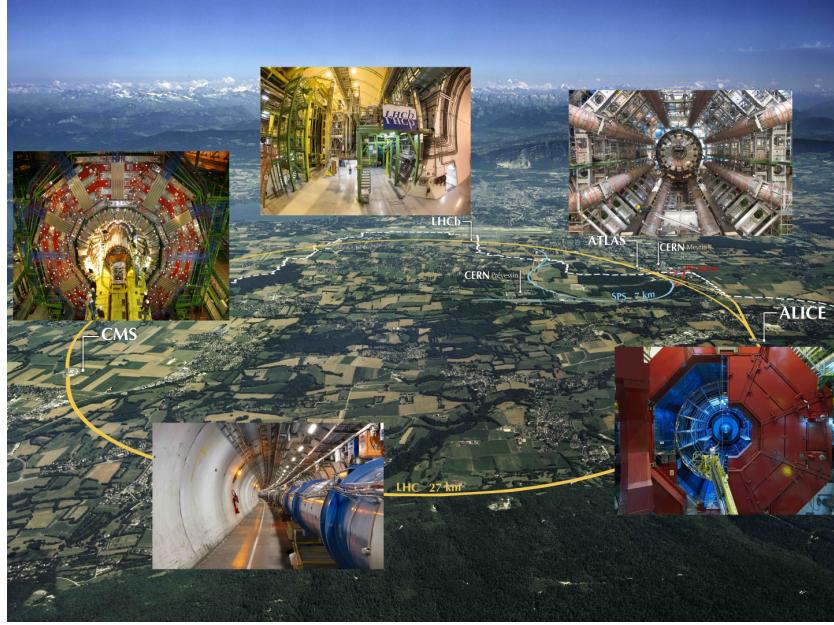


Figure 2.1: Aerial view of the Large Hadron Collider (LHC) spanning the border of France and Switzerland, and the four major experiments located around the ring: CMS (Compact Muon Solenoid), LHCb (LHC beauty), ATLAS (A Toroidal LHC Apparatus), and ALICE (A Large Ion Collider Experiment) [11].

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

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

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

- 643 • N_b is the number of particles per bunch ($N_b \approx 1.15 \times 10^{11}$ protons per bunch)
 644 • n_b is the number of bunches per beam (maximum 2808),
 645 • f_{rev} is the revolution frequency ($\approx 11 \text{ kHz}$),
 646 • 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 $\sim 1.0 \times 10^{34}$ cm $^{-2}$ s $^{-1}$, with peak luminosity per bunch crossing reaching $\sim 3.56 \times 10^{34}$ cm $^{-2}$ 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 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]. These multiple collisions will lead to particle interactions overlapping in the detector. To measure a proton-proton collision, the single collision must be separated from overlapping collisions, which are called “pileup” collisions. A distribution of pileup in the data-taking years 2016-2018 is shown in Fig. 2.2. The pileup is defined as the average number of pp collisions per bunch crossing.

CMS reports an inelastic pp cross section of $\sigma_{\text{inel}} = 68.6$ millibarns at a center-of-mass energy of $\sqrt{s} = 13$ TeV [19], which can be used to estimate pileup as follows:

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

With the example values above, pileup can be estimated to be ~ 22 .

670 While useful in the search for rare physics processes, higher luminosities create
 671 more intense pileup conditions, posing a greater challenge to detector performance
 672 and particle reconstruction and identification.

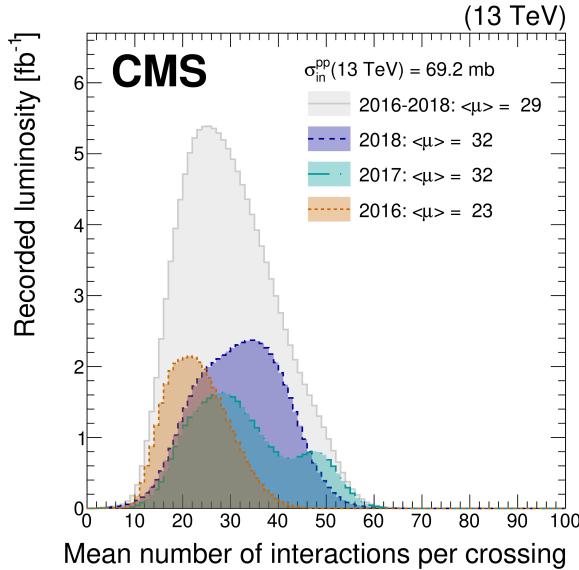


Figure 2.2: Distribution of the mean number of inelastic collisions per bunch crossing (pileup) in data [18], for proton-proton collisions in 2016 (*dotted orange*), 2017 (*dotted light blue*), 2018 (*dotted dark blue*), and integrated over 2016-2018 (*solid grey*). A cross-section of inelastic proton-proton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

673 2.3 The High-Luminosity LHC

674 The High-Luminosity LHC (HL-LHC) is a major upgrade of the LHC scheduled
 675 to take place in the late 2020s, that will increase the instantaneous luminosity by
 676 a factor of five beyond the original design value, and the integrated luminosity
 677 by a factor of ten [20]. This will be accomplished through accelerator technological
 678 advances: for instance, reduction of the interaction point β^* from 0.55 m down to 0.15
 679 m by installation of new final-focusing magnets, and improvements in the geometric
 680 luminosity loss factor $F \approx 1$ through the installation of crab cavities that optimize

681 the orientation of colliding bunches. A further discussion of the HL-LHC upgrades
682 for the CMS detector follows in Chapter 3.

683 2.4 The CMS Detector

684 The Compact Muon Solenoid (CMS) experiment was conceived to study proton-
685 proton and lead-lead collisions at a center-of-mass energy of 14 TeV (5.5 TeV nucleon-
686 nucleon) and at luminosities up to $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ($10^{27} \text{ cm}^{-2} \text{ s}^{-1}$) [21] [22]. Starting
687 from the beam interaction region at the center of the CMS detector, particles first
688 pass through a silicon pixel and strip tracker, in which charged-particle trajectories
689 (tracks) and origins (vertices) are reconstructed from signals (hits) in the sensitive
690 layers. The tracker is immersed in a high-magnetic-field superconducting solenoid
691 that bends the trajectories of charged particles, allowing the measurement of their
692 electric charge and momenta. Electrons and photons are then absorbed in an electro-
693 magnetic calorimeter (ECAL) comprised of lead-tungstate scintillating-crystals. The
694 corresponding electromagnetic showers are detected as clusters of energy recording in
695 neighboring cells, from which the direction and energy of the particles can be deter-
696 mined. Charged and neutral hadrons may initiate a hadronic shower in the ECAL
697 as well, which is then fully absorbed in the hadron calorimeter (HCAL). The result-
698 ing clusters are used to estimate their direction and energies. Muons and neutrinos
699 pass through the calorimeters with little to no interactions. Neutrinos escaped un-
700 detected; muons produce hits in additional gas-ionization chamber muon detectors
701 housed in the iron yoke of the flux-return. A sketch of example particle interactions
702 in a transverse slice of the CMS detector is shown in Fig. 2.3. The collision data is
703 recorded with the use of the Level-1 (L1) trigger (discussed in greater detail in 2.5.5),
704 the High-Level Trigger (HLT), and data acquisition systems ensuring high efficiency
705 in selecting physics events of interest.

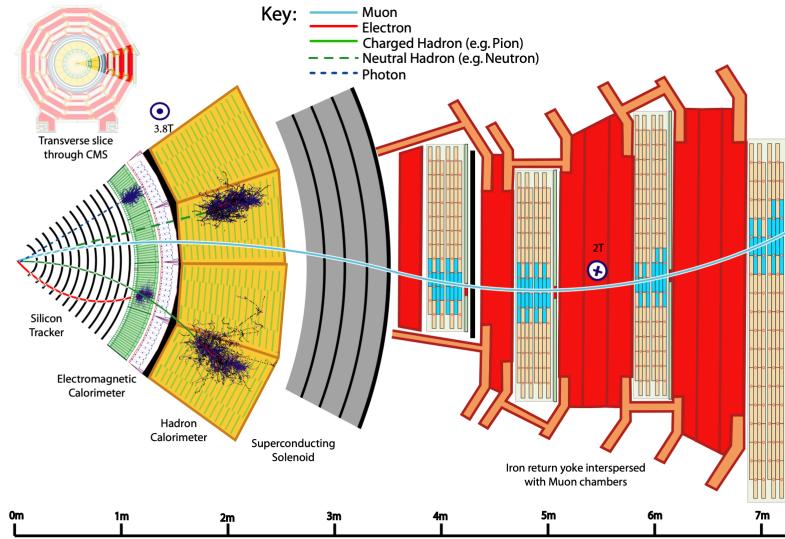


Figure 2.3: Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector [22].

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

716 2.5 Sub-detectors of CMS

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

719 **2.5.1 Inner tracking system**

720 The CMS Tracker performs robust tracking and detailed vertex reconstruction in the
721 4 T magnetic field of the superconducting solenoidal magnet. The primary sensors
722 used in the tracker are p^+ on n -bulk devices, which allow high voltage operation and
723 are radiation-resistant [23] [24]. The active envelope of the CMS Tracker extends to a
724 radius of 115 cm, over a length of approximately 270 cm on each side of the interaction
725 point [23]. Charged particles in the region $|\eta| \lesssim 1.6$ benefit from the full momentum
726 measurement precision. In this region, a charged particle with p_T of 1000 GeV has a
727 sagitta of $\sim 195 \mu\text{m}$. The Tracker acceptance extends further to $|\eta| = 2.5$, with a
728 reduced radius of approximately 50 cm.

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

733 A schematic view of the current Phase-1 CMS tracker [25], including the pixel
734 detector, is shown in Fig. 2.4. The Phase-1 pixel detector consists of three barrel
735 layers (BPIX) at radii of 4.4 cm, 7.3 cm, and 10.2 cm, and two forward/backward disks
736 (FPIX) at longitudinal positions of ± 34.5 cm and ± 46.5 cm, and extending in radius
737 from about 6 cm to 15 cm. These pixelated detectors produce 3D measurements along
738 the paths of charged particles with single hit resolutions between 10-20 μm .

739 After the pixel and on their way out of the tracker, particles pass through the
740 silicon strip tracker which reaches out to a radius of 130 cm (Fig. 2.4). The sensor
741 elements in the strip tracker are single-sided p -on- n type silicon micro-strip sensors
742 [21]. The silicon strip detector consists of four inner barrel (TIB) layers assembled
743 in shells, with two inner endcaps (TID), each composed of three small discs. The
744 outer barrel (TOB) consists of six concentric layers. Two endcaps (TEC) close off
745 the tracker on either end.

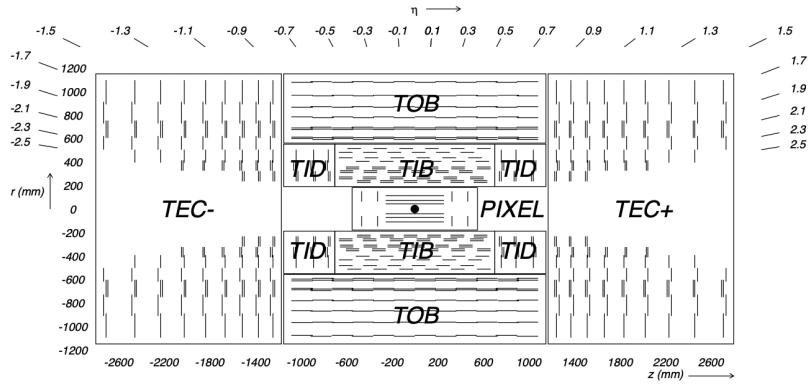


Figure 2.4: Cross section of the current Phase-1 CMS tracker [25]. Each line represents a detector module. Double lines indicate back-to-back modules which deliver two-dimensional (stereo) hits in the strip tracker.

2.5.2 ECAL

The electromagnetic calorimeter (ECAL) of CMS measures electromagnetic energy deposits with high granularity. One of the driving criteria in the design was the capability of detecting the Standard Model Higgs boson decay to two photons (in fact, the channel in which the 125 GeV Higgs boson was discovered at CMS). ECAL is a hermetic homogeneous calorimeter comprised of 61,200 lead tungstate (PbWO_4) crystals mounted in the central barrel, with 7,324 crystals in each of the two endcaps [21]. A preshower detector is located in front of the endcap crystals. Avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps.

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

763 by electrons and photons.

764 The barrel part of the ECAL (EB) covers the pseudorapidity range $|\eta| < 1.479$
765 [21]. The barrel granularity is 360-fold in ϕ and (2×85) -fold in η . The crystal cross-
766 section corresponds to approximately 0.0174×0.0174 in $\eta - \phi$ or 22×22 mm 2 at the
767 front face of the crystal, and 26×26 mm 2 at the rear face. The crystal length is 230
768 mm, corresponding to $25.8 X_0$.

769 The ECAL read-out acquires the signals of the photodetectors [21]. At each bunch
770 crossing, digital sums representing the energy deposit in a trigger tower, comprising
771 5×5 crystals in $\eta \times \phi$, are generated and sent to the Level-1 trigger system (detailed
772 in Section 2.5.5).

773 2.5.3 HCAL

774 The hadronic calorimeter (HCAL) of CMS measures hadronic energy, which is key to
775 characterizing the presence of apparent missing transverse energy which could arise
776 from hadron jets and neutrinos or exotic particles [21]. A schematic of the components
777 of HCAL are shown in Fig. 2.5. The HCAL barrel (HB) and endcaps (HE) are located
778 outside of the tracker and the ECAL, spanning a radius of 1.77 m (outer extent of
779 ECAL) up to 2.95 m (inner extent of the magnet coil). An outer hadron calorimeter
780 (HO) is placed outside the solenoid to complement the barrel calorimeter. Beyond
781 $|\eta| = 3$, the forward hadron calorimeter (HF) at 11.2 m from the interaction point
782 extend the pseudorapidity coverage to $|\eta| = 5.2$.

783 The HB is a sampling calorimeter covering the pseudorapidity range $|\eta| < 1.3$ [21].
784 It consists of 36 identical azimuthal wedges which form two half-barrels (HB+ and HB-
785), with a segmentation of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The HE covers pseudorapidity
786 $1.3 < |\eta| < 3$. The HB and endcap HE calorimeters are sampling calorimeters which
787 use brass as the absorber and plastic scintillator as the active material. Light from
788 the plastic scintillator is wavelength-shifted and captured in optic fibers which are

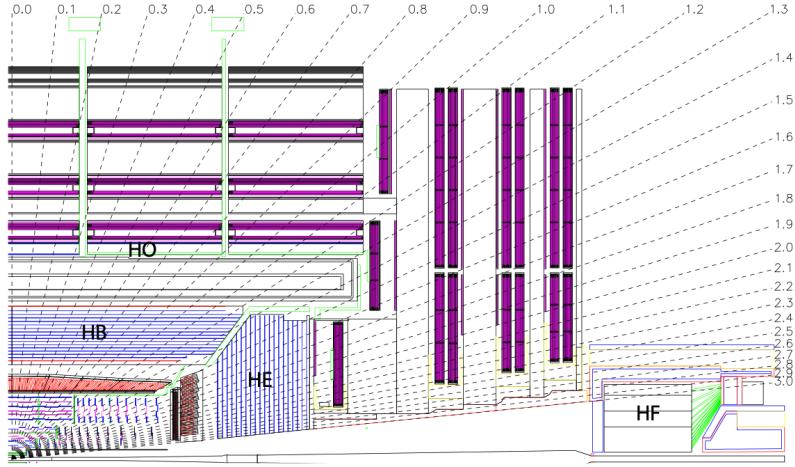


Figure 2.5: Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters from [21].

789 read out by front-end electronics [27].

790 In the central pseudorapidity region, the combined stopping power of EB plus the
 791 HB is insufficient to contain hadron showers [21]. To ensure adequate sampling depth,
 792 the hadron calorimeter is extended with a tail catcher, the HO. The size and position
 793 of the tiles are designed to roughly map the layers of the HB to make towers with
 794 the same granularity of 0.087×0.087 in η and ϕ . HO uses the same active material
 795 as the HB and HE calorimeters, but uses the steel return yoke and magnet material
 796 of CMS as absorbers [27].

797 The HF is a Cherenkov calorimeter based on a steel absorber and quartz fibers
 798 which run longitudinally through the absorber and collect Cherenkov light, primarily
 799 from the electromagnetic component of showers developed in the calorimeter [27].
 800 Photomultiplier tubes are used to collect light from the quartz fibers. The HF is
 801 designed to survive in the harsh radiation conditions and high particle flux of the
 802 forward region. On average, 760 GeV per proton-proton interaction is deposited into
 803 the two forward calorimeters, compared to only 100 GeV for the rest of the detector
 804 [21]. Furthermore, this energy has a pronounced maximum at the highest rapidities.

805 **2.5.4 Muon detectors**

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

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

818 In the two endcap regions, the muon rates and background levels are high and the
819 magnetic field is large and non-uniform [21]. Here, the muon system uses cathode
820 strip chambers (CSCs) to identify muons between $0.9 < |\eta| < 2.4$. The cathode strips
821 of each chamber run radially outwards and provide a precision measurement in the
822 $r - \phi$ bending plane. The anode wires run approximately perpendicular to the strips
823 and are read out in order to measure η and the beam-crossing time of a muon.

824 In addition to the DT and CSC, a dedicated trigger system consisting of resistive
825 plate chambers (RPCs) in the barrel and endcap regions provide a fast, independent,
826 and highly-segmented trigger with a sharp p_T threshold over a large portion of the
827 pseudorapidity range ($|\eta| < 1.6$) of the muon system [21]. RPCs have good time
828 resolution but coarser position resolution compared to the DTs or CSCs. The RPCs
829 also play a role in resolving ambiguities in reconstructing tracks from multiple hits in
830 a chamber.

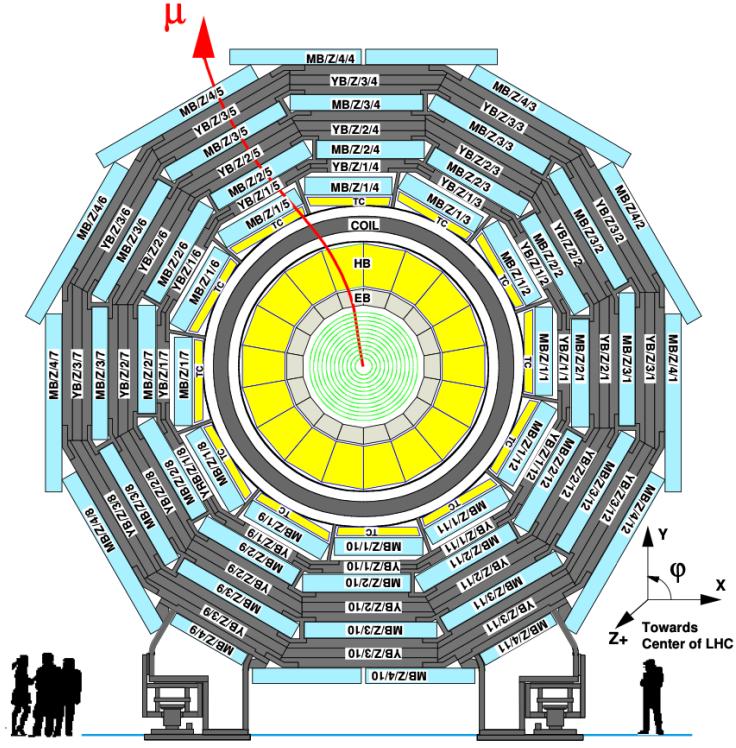


Figure 2.6: Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels from [21]. The DTs are organized in 12 sectors of the yoke barrel (YB). In each of the 12 sectors of the yoke, there are 4 muon chambers per wheel (MB1, MB2, MB3, and MB4).

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]. The large number of 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

841 Trigger (HLT) which is described in Section 2.5.6. This section describes the Phase-1
842 configuration of the Level-1 Trigger.

843 The L1 Trigger data flow of Phase-1 is shown in Fig. 2.7 [28], with organization
844 into the L1 calorimeter trigger, the L1 muon trigger, and the L1 global trigger.

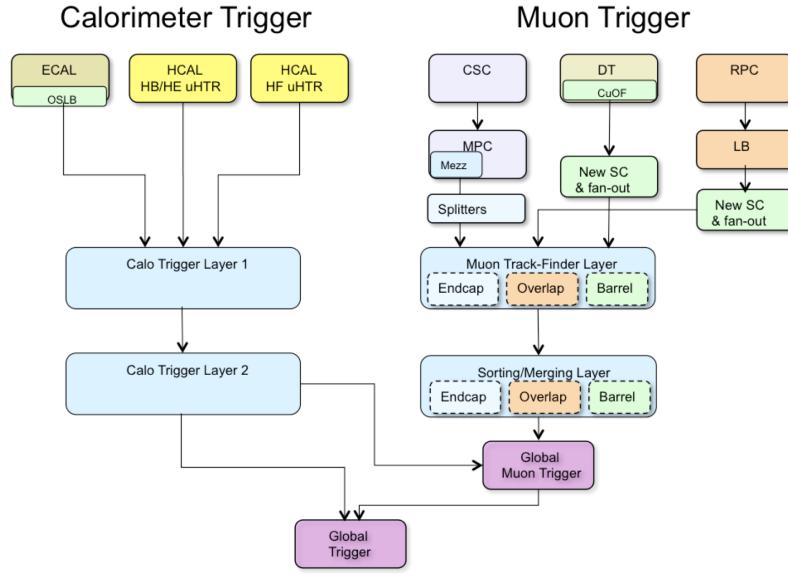


Figure 2.7: Dataflow for the Phase-1 Level-1 Trigger [28], which is implemented in custom hardware and is responsible for reducing the event rate from the LHC bunch crossing frequency of 400 MHz (bunch crossings every 25 ns) to a maximum rate of 100 kHz. In Phase-1, the Level-1 Trigger has access to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

845 The L1 calorimeter trigger begins with trigger tower energy sums formed by the
846 ECAL, HCAL, and HF Trigger Primitive Generator (TPG) circuits from the indi-
847 vidual calorimeter cell energies. In the original configuration, the ECAL energies
848 were accompanied by a bit indicating the transverse extent of the electromagnetic
849 energy deposits, and the HCAL energies were accompanied by a bit indicating the
850 presence of minimum ionizing energy [29]. Between Long Shutdowns 1 and 2 (LS1
851 and LS2), HF was upgraded to provide finer granularity information to the trigger,
852 and the HCAL barrel and endcap front-end electronics were upgraded to provide
853 high-precision timing information and depth segmentation information.

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

864 Each of the L1 muon triggers has its own trigger logic [29]. The RPC strips are
865 connected to a Pattern Comparator Trigger (PACT), which forms trigger segments
866 that are used to build tracks and calculate p_T . The RPC logic also provides some
867 hit data to the CSC trigger system to resolve ambiguities caused by two muons in
868 the same CSC. The CSCs form local charged tracks (LCTs) from the cathode strips,
869 which are combined with the anode wire information. LCTs are combined into full
870 muon tracks and assigned p_T values.

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

880 Information from the GCT and GT are sent to the Global Trigger (GT), which

881 makes the Level-1 Accept (L1A) decision to either discard or accept the bunch crossing
882 [29]. This is accomplished by sorting ranked trigger objects that are accompanied by
883 positional information in η and ϕ , permitting the trigger to applying criteria with
884 thresholds that can vary based on the location of the trigger objects, and/or to
885 require trigger objects to be close to or opposite from each other. The GT L1A
886 decision arrives at the detector front end with a $3.8\ \mu\text{s}$ latency after the interaction
887 at a rate which is required to be less than 100 kHz, and triggers a full readout of the
888 detector for further processing.

889 **2.5.6 The High-Level Trigger**

890 The HLT is implemented in software running on a large computer farm of fast com-
891 mercial processors [30] [31]. The algorithms in HLT have access to full data from
892 all CMS sub-detectors, including the tracker, with full granularity and resolution.
893 The HLT reconstruction software is similar to what is used offline for full CMS data
894 analysis. As a result, the HLT can calculate quantities with a resolution compara-
895 ble to the final detector resolution, compared to the L1 Trigger. The HLT performs
896 more computationally-intensive algorithms, such as combining tau-jet candidates in
897 the calorimeter with high- p_T stubs in the tracker, to form a hadronic tau trigger. The
898 maximum HLT input rate from the L1 Trigger is 100 kHz, and the HLT output rate
899 is approximately 100 Hz.

900 The HLT contains trigger paths, each corresponding to a dedicated trigger [32].
901 A path consists of several steps implemented as software modules. Each HLT trigger
902 path must be seeded by one or more L1 trigger bits: the first module always looks
903 for a L1 seed, consisting of L1 bit(s) and L1 object(s). Each module performs a well-
904 defined task such as unpacking (raw to digitized quantities), reconstruction of physics
905 objects (electrons, muons, jet, missing transverse energy, etc.), making intermediate
906 decisions that trigger more detailed reconstruction modules, and calculating the final

907 decision for the trigger path. If an intermediate filter decision is negative, the rest of
908 the path is not executed, and the trigger rejects the event.

909 **2.5.7 Particle reconstruction**

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

920 **2.5.8 Data storage and computational infrastructure**

921 The LHC generates over 15 petabytes (15 million gigabytes) of data every year, neces-
922 sitating a flexible computing system that can be accessed by researchers working at
923 the four main LHC experiments: ALICE, ATLAS, CMS, and LHCb. The Worldwide
924 LHC Computing Grid (WLCG) [33] is a global collaboration of computer centers that
925 links thousands of computers and storage systems in over 170 centers across 41 coun-
926 tries. These centers are arranged in “tiers”, and provide near real-time access to users
927 processing, analyzing, and storing LHC data. One of the final stages of data analy-
928 sis at LHC experiments is large-scale data processing taking place over distributing
929 computing, for instance, with the use of Condor [34], a distributed, scalable, flexible
930 batch processing system which accepts a computing job, allocates a resource to it,
931 executes it, and returns the result back to a user transparently.

932 Chapter 3

933 The Phase-2 Upgrade of CMS

934 3.1 High-Luminosity LHC and CMS

935 In order to sustain and extend the LHC’s physics discovery program and maintain
936 operability for a decade or more, the LHC is undergoing a major upgrade to the High-
937 Luminosity LHC (HL-LHC). In its final configuration, the HL-LHC will deliver a peak
938 luminosity of $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, potentially leading to total integrated luminosity
939 of 4000 fb^{-1} after ten years of operations, scheduled to begin in 2027 [35]. This
940 integrated luminosity is about ten times the predicted luminosity reach of the LHC
941 in its initial configuration. To maximize the discovery potential of this unprecedented
942 amount of data, the CMS detector is undergoing Phase-2 upgrades in order to perform
943 high-precision measurements and searches for physics beyond the Standard Model in
944 the intense running conditions of the HL-LHC.

945 3.2 The Phase-2 Level-1 Trigger

946 To achieve the goals of the HL-LHC program and to ensure the collection of information-
947 rich datasets in the HL-LHC, the Phase-2 upgrade of the CMS Level-1 Trigger [35]
948 must be upgraded in conjunction with the CMS sub-detectors and their readouts, to

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

952 To profit from the extended coverage and increased granularity of the upgraded
 953 CMS detector, the latency of the L1 trigger system (time available to produce a L1
 954 Accept signal) will be increased significantly from $3.8 \mu\text{s}$ to $12.5 \mu\text{s}$, with an increased
 955 maximum output bandwidth of 750 kHz [35]. With the increased latency, in addition
 956 to information from calorimeters and muon detectors (as in the Phase-1 system),
 957 information from the new tracker and high-granularity endcap calorimeter can also
 958 be included at L1 for the first time. This is illustrated in the functional diagram of
 959 the architecture of the Phase-2 trigger system in Fig. 3.1.

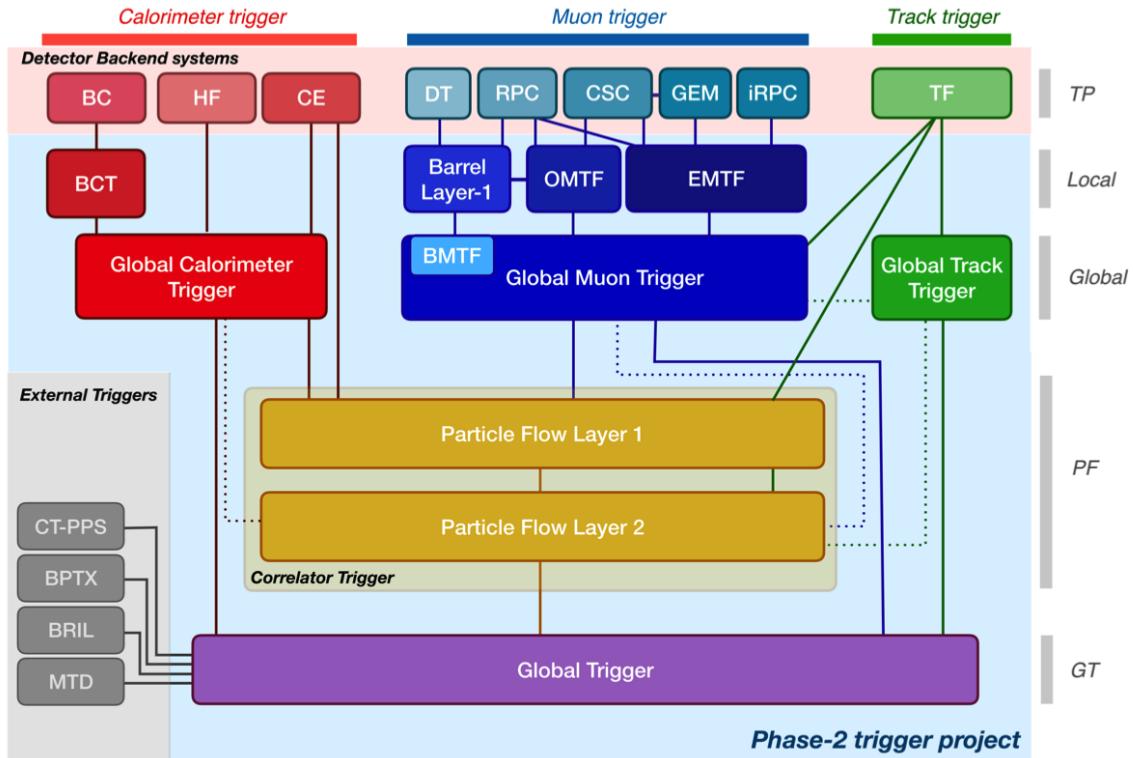


Figure 3.1: Functional diagram of the CMS L1 Phase-2 upgraded trigger design [35], showing the four trigger paths: calorimeter, muon, track, and Particle Flow. For the first time, tracking information will be available as early as the L1 Trigger.

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

961 where algorithms produce higher-level trigger objects by combining information from
962 sub-detectors, with a selectivity approaching that of offline reconstruction in the
963 HLT [35]. Four independent data processing paths (grouped together in Fig. 3.1) are
964 implemented: tracking, calorimetry, muon systems, and particle-flow techniques:

- 965 • **Calorimeter Trigger path:** (*red*, Fig. 3.1) A barrel calorimeter trigger (BCT)
966 and the HGCAL backend are used to produce high-granularity information from
967 the calorimeters to produce high-resolution clusters and identification variables
968 used for later processing. Outputs from the BCT, HGCAL, and the HF are sent
969 to a global calorimeter trigger (GCT), where calorimeter-only objects such as
970 e/γ candidates, hadronically decaying tau lepton candidates, jets, and energy
971 sums are built.
- 972 • **Track Trigger path:** (*green*, Fig. 3.1) Tracks from the Outer Tracker are
973 reconstructed in the track finder (TF) processors as part of the detector back-
974 end. A global track trigger (GTT) will reconstruct the primary vertices of the
975 event, along with tracker-only based objects, such as jets and missing transverse
976 momentum.
- 977 • **Muon Trigger path:** (*blue*, Fig. 3.1) Trigger primitives are processed by
978 muon track finder algorithms, again separated into the barrel (barrel muon
979 track finder, BMTF), overlap (overlap muon track finder, OMTF), and endcap
980 (endcap muon track finder, EMTF). Standalone muons and stubs containing
981 information such as position, bend angle, and timing, as well as L1 tracks, are
982 sent to the global muon trigger (GMT).
- 983 • **Particle-Flow Trigger path:** (*yellow*, Fig. 3.1) The correlator trigger (CT)
984 aims to approach the performance of offline Particle Flow, and is implemented
985 in two layers. “Layer-1” produces the particle-flow candidates from matching

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

988 The outputs from the above trigger paths are combined in the Global Trigger
989 (GT) (*purple*, Fig. 3.1), which calculates the final trigger decision (Level-1 Accept),
990 transmitting it to the Trigger Control and Distribution System (TCDS), which dis-
991 tributes it to the detector backend systems, initiating the readout to the DAQ. The
992 GT also provides the interface to external triggers (*grey*, Fig. 3.1), such as trig-
993 gers for the precision proton spectrometer (PPS), beam position and timing monitors
994 (BPTX), and luminosity and beam monitoring (BRIL) detectors [35]. The design of
995 the Phase-2 Level-1 Trigger allows for future inclusion of triggering information, for
996 instance information about minimum ionizing particles (MIPs) from the MIP Timing
997 Detector (MTD) [36].

998 **3.3 Standalone Barrel Calorimeter electron/photon 999 reconstruction**

1000 The reconstruction and identification of electrons and photons (e/γ) begin with the
1001 trigger primitives of the barrel ECAL and HCAL detectors and endcap HGCAL
1002 calorimeters, covering the pseudorapidity region $|\eta| < 3$. The barrel and endcap re-
1003 gions of the detector are intrinsically different enough to warrant different approaches
1004 to e/γ reconstruction. This work focuses on the Standalone Calorimeter e/γ recon-
1005 struction taking place in the barrel (Fig. 3.2).

1006 **3.3.1 Phase-2 geometry of the ECAL Barrel trigger**

1007 In Phase-2, the upgrade of both on-detector and off-detector electronics for the barrel
1008 calorimeters trigger primitive generator (TPG) will stream single crystal data from

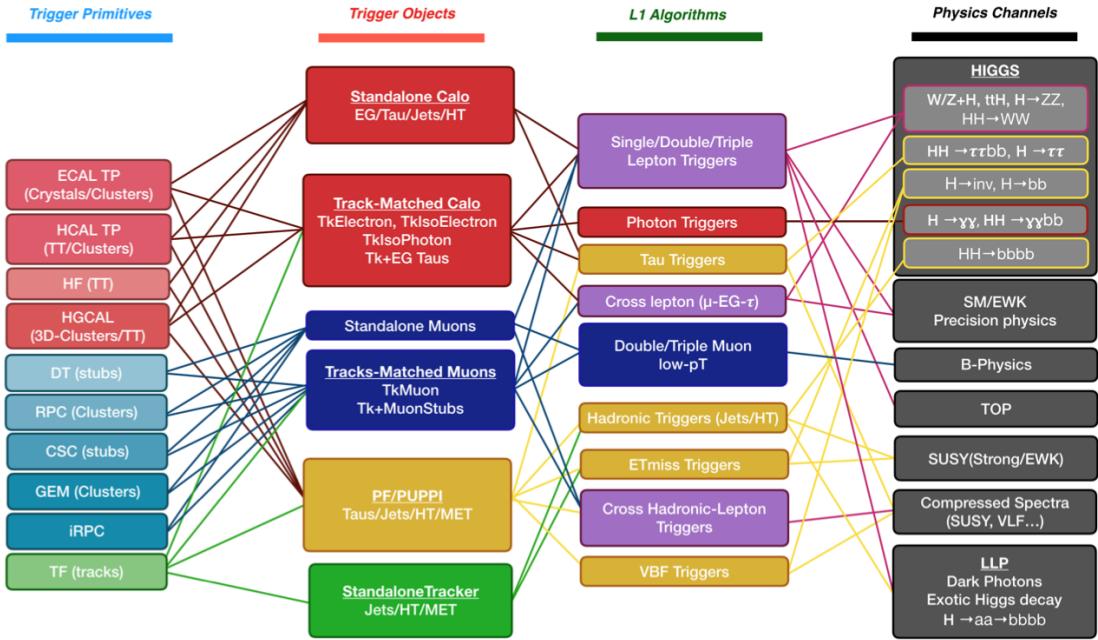


Figure 3.2: Summary of the links between the trigger primitives (*first column*), the trigger objects (*second column*), the Level-1 algorithms used in the menu (*3rd column*), and the physics channels (*4th column*), from [35], where a full description of the Phase-2 L1 algorithms can be found. This work focuses on developments for the Standalone Calorimeter electron and photon ("EG") reconstruction algorithm.

the on-detector to the backend electronics, in contrast to the lower-granularity output of the Phase-1 ECAL TPG that is restricted to providing trigger tower sums of 5×5 crystals [35]. A schematic representation of the geometry of the ECAL barrel in the 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. 3.4. After initial clustering and processing, the outputs of the RCT card are sent to the Global Calorimeter (GCT) trigger, which is processed in three cards as shown in Fig. 3.5.

3.3.2 Phase-2 electron/photon reconstruction algorithm

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

1020 of the barrel ECAL, with a granularity 0.0175×0.0175 in $\eta \times \phi$, which is 25 times
1021 higher than the input to the Phase-1 trigger, which consisted of trigger towers with
1022 a granularity of 0.0875×0.0875 . In HCAL the tower size of 0.0875×0.0875 is un-
1023 changed. The trigger algorithm is designed to closely reproduce the algorithm used in
1024 the offline reconstruction, with limitations and simplifications due to trigger latency.

1025 In the RCT, an initial requirement of $p_T > 0.5$ GeV is imposed on the input
1026 trigger primitives (i.e. energies from the ECAL crystals and HCAL towers) to reject
1027 contribution from pileup. In one of the regions inside a RCT card (Fig. 3.4), the
1028 crystal containing the highest energy deposit is identified as the seed crystal, as shown
1029 in Fig. 3.6. The energy in the crystals in a window of size 3×5 in $\eta \times \phi$ around
1030 the seed cluster is added into a cluster. The energy is considered “clustered”. The
1031 process is repeated with the remaining “unclustered” energy, until up to four clusters
1032 are produced in the region.

1033 To improve e/γ identification and to reduce background contributions, identifica-
1034 tion and reconstruction algorithms are implemented at this stage:

- 1035 • Shower shape: The energy deposit sums around the seed crystal is computed in
1036 windows of size 2×5 and 5×5 (Fig. 3.6, *dashed lines*), with true e/γ clusters
1037 tending to produce showers that deposit most of their energy in a 2×5 region.
- 1038 • Bremsstrahlung recovery: e/γ tend to spread in the ϕ direction due to charged
1039 particles being bent by the magnetic field of the CMS solenoid. If sufficient
1040 energy comparable to the core 3×5 cluster is found in the adjacent 3×5
1041 windows (Fig. 3.6, *shaded yellow*), the energy is added to the core cluster and
1042 no longer considered unclustered energy.

1043 After parallel processing in the regions, the clusters in a RCT card are stitched
1044 together if they are located directly along the borders of a region (Fig. 3.3). The
1045 remaining unclustered ECAL energy is summed into ECAL towers.

1046 From each RCT card, the twelve highest-energy clusters, as well as any remaining
1047 unclustered energy, are sent to the GCT. Since each GCT card has information from
1048 sixteen RCT cards (Fig. 3.5), final stitching across the boundaries of the RCT cards
1049 is performed. One more identification algorithm is performed at this stage:

- 1050 • Isolation: One handle to reject backgrounds from e.g. pileup, comes from the
1051 tendency for background to be spread more uniformly across a large area in the
1052 detector, whereas genuine e/γ are expected to produce showers concentrated in
1053 the 3×5 crystal window. The energy sum in a large window of 7×7 in towers
1054 is computed and used to reject background.

1055 The performance of the standalone barrel e/γ algorithm in Phase-2 conditions is
1056 summarized in the efficiency and rates. The efficiencies are measured with a simulated
1057 Monte Carlo sample containing electrons. The rates are measured with a simulated
1058 minimum bias sample intended to closely mimic generic proton-proton collisions in
1059 the CMS detector. The performance of the Phase-2 emulator discussed in this work,
1060 which closely mimics the firmware logic and uses fixed-precision integers, is shown to
1061 be comparable to the previous emulator which used floats and idealized logic.

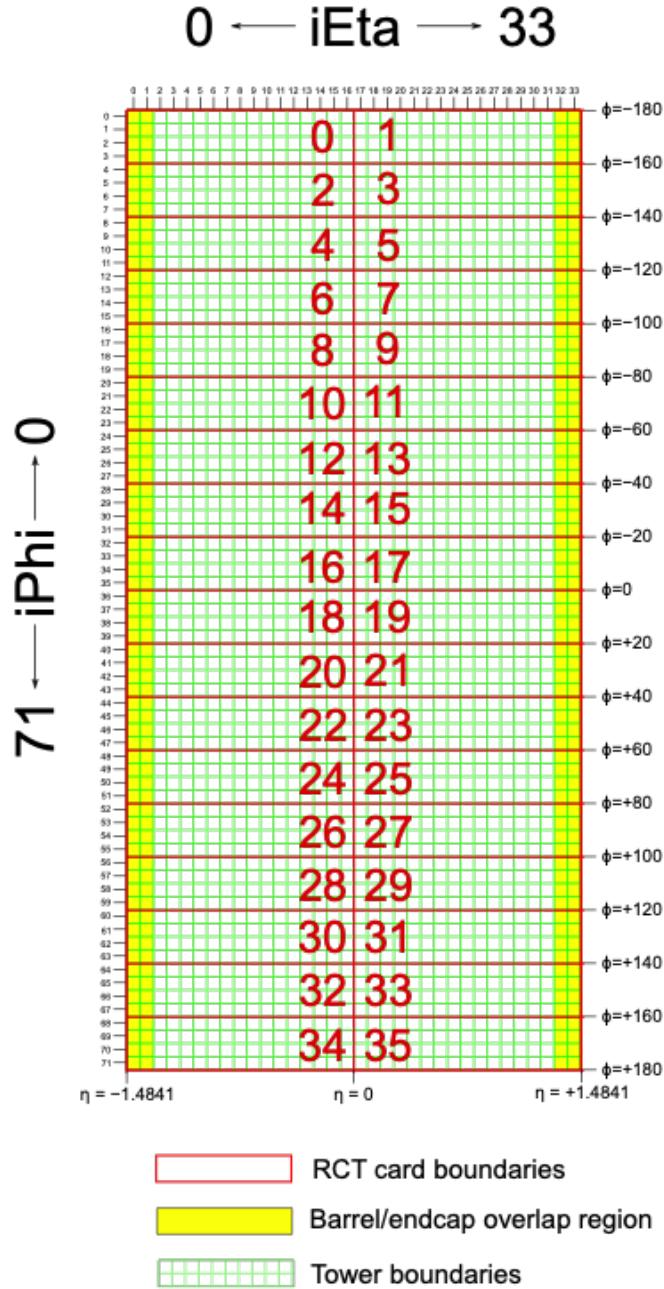


Figure 3.3: Schematic of the geometry of the Phase-2 ECAL barrel in the Regional Calorimeter Trigger (RCT), showing the division of the barrel region into 36 Regional Calorimeter Trigger (RCT) cards (red). Each card spans 17×4 towers in $\eta \times \phi$ (green), and each tower is 5×5 in single crystals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

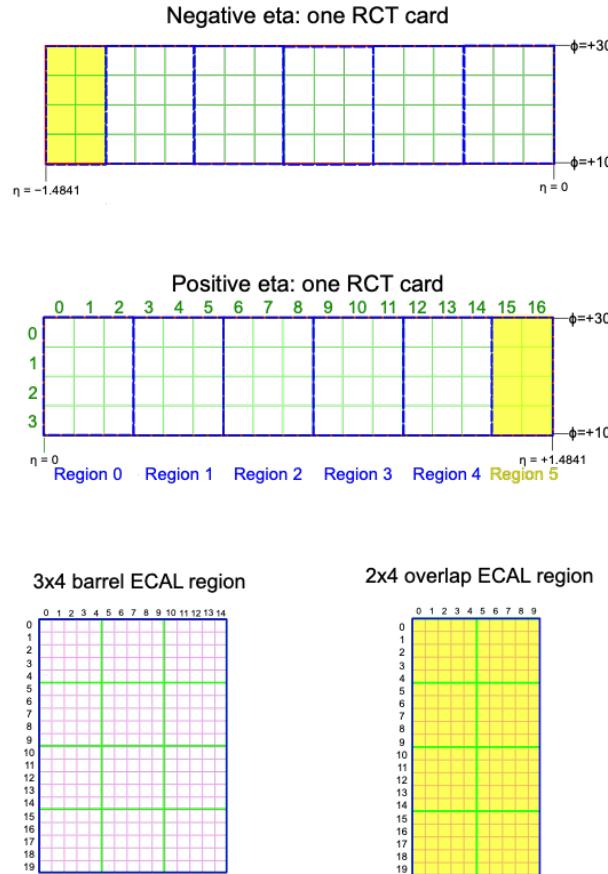


Figure 3.4: Schematic of two example RCT cards in the negative eta (*top*) and positive eta (*center*) regions of the ECAL barrel. Each RCT card is divided into five regions: four regions are of size 3×4 towers in $\eta \times \phi$ (*bottom left*), and a fifth smaller overlap region of size 2×4 towers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

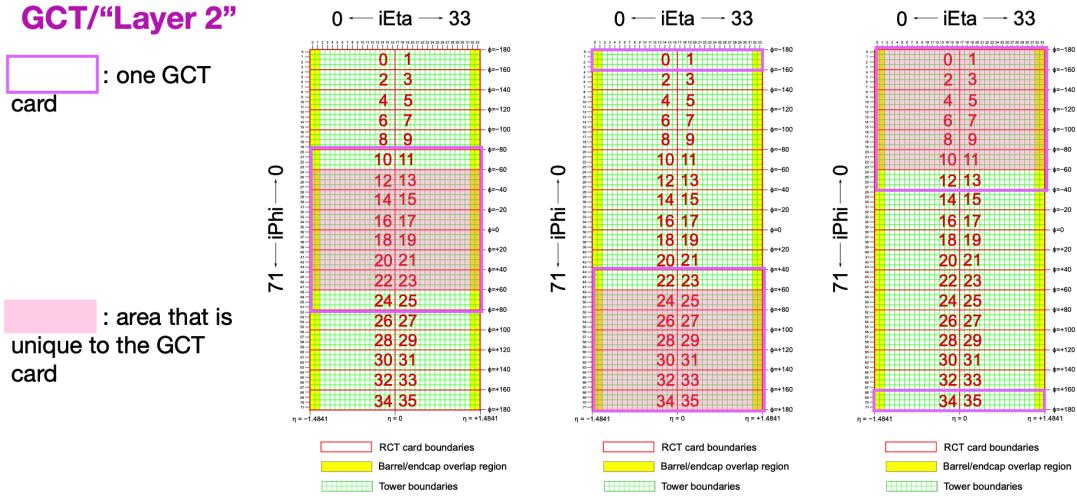


Figure 3.5: Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trigger (GCT), which will process the outputs of the Regional Calorimeter Trigger (RCT) in three cards (*magenta highlights*). Each card in the GCT processes the equivalent of sixteen RCT cards, with the center twelve being unique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

3x4 barrel ECAL region

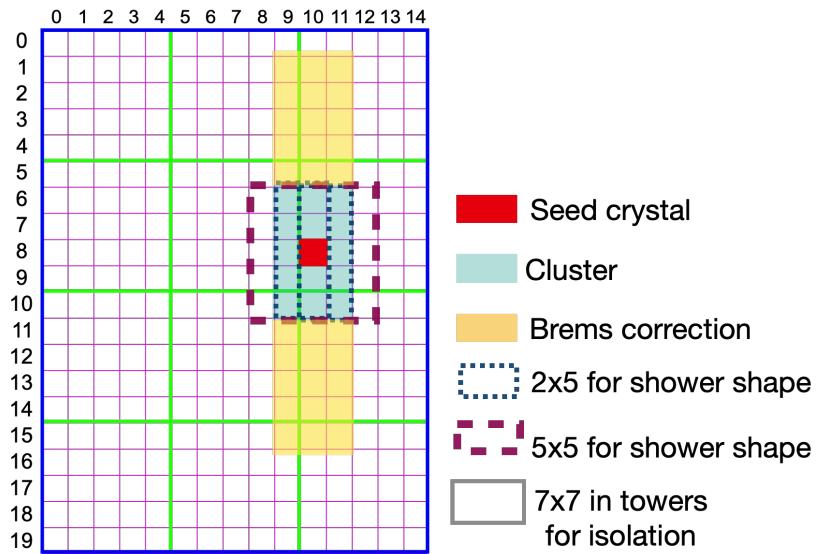


Figure 3.6: Illustration of an example electron/photon (e/γ) cluster in the Phase-2 Level-1 Trigger standalone barrel e/γ reconstruction, in a region of 15×20 crystals (3 \times 4 towers). Each small pink square is one crystal, the highest-granularity ECAL trigger primitives available to the L1 Trigger in Phase-2. The core cluster consists of the energy sum in a 3×5 window of crystals, (*shaded light blue*) centered around the seed crystal (*red*). Bremsstrahlung corrections are checked in the adjacent 3×5 windows in the ϕ direction (*shaded light yellow*). The relative energies in windows of size 2×5 and 5×5 in crystals (*dashed dark blue and dark red*) are used to compute shower shape variables to identify true e/γ objects. Lastly, an isolation sum is computed in a window of size 7×7 in towers (not shown in figure).

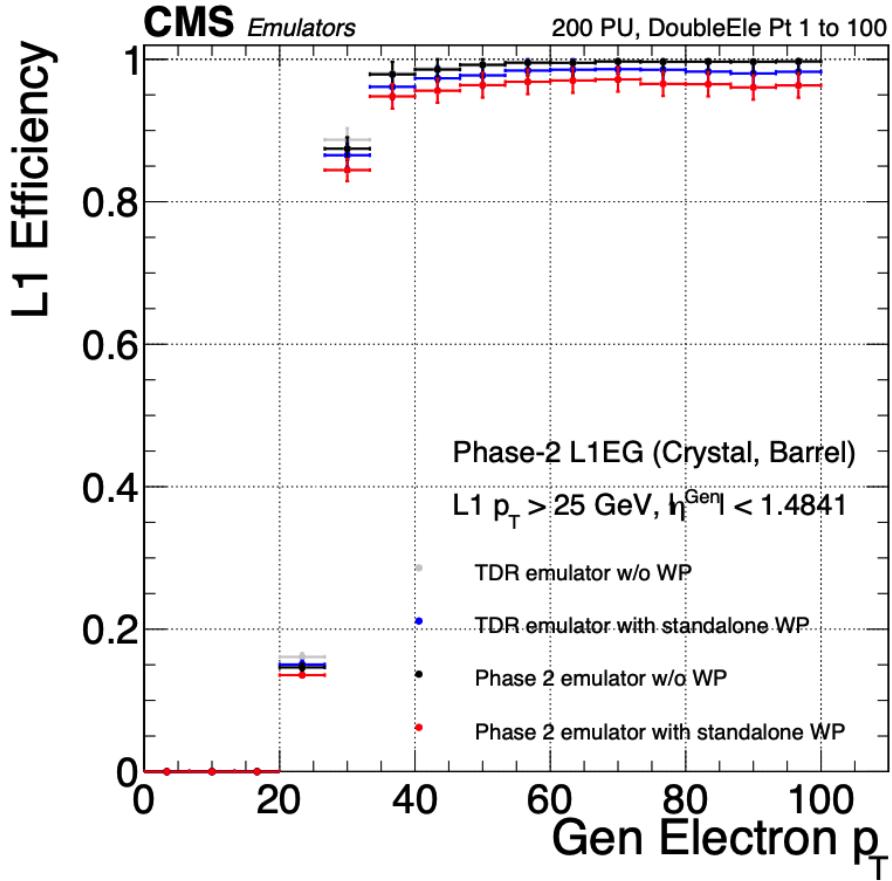


Figure 3.7: Efficiency of the standalone barrel e/γ reconstruction, measured in a simulated sample of electrons, as a function of the true electron’s transverse momentum p_T . The performance of the previous, idealized algorithm as shown in the 2021 Phase-2 TDR [35] with and without the isolation and shower shape discrimination variables (“standalone working point/ WP”) (*dark blue, grey*). The Phase-2 emulator discussed in this work with and without the same working point (*black, red*) is shown to have comparable performance.

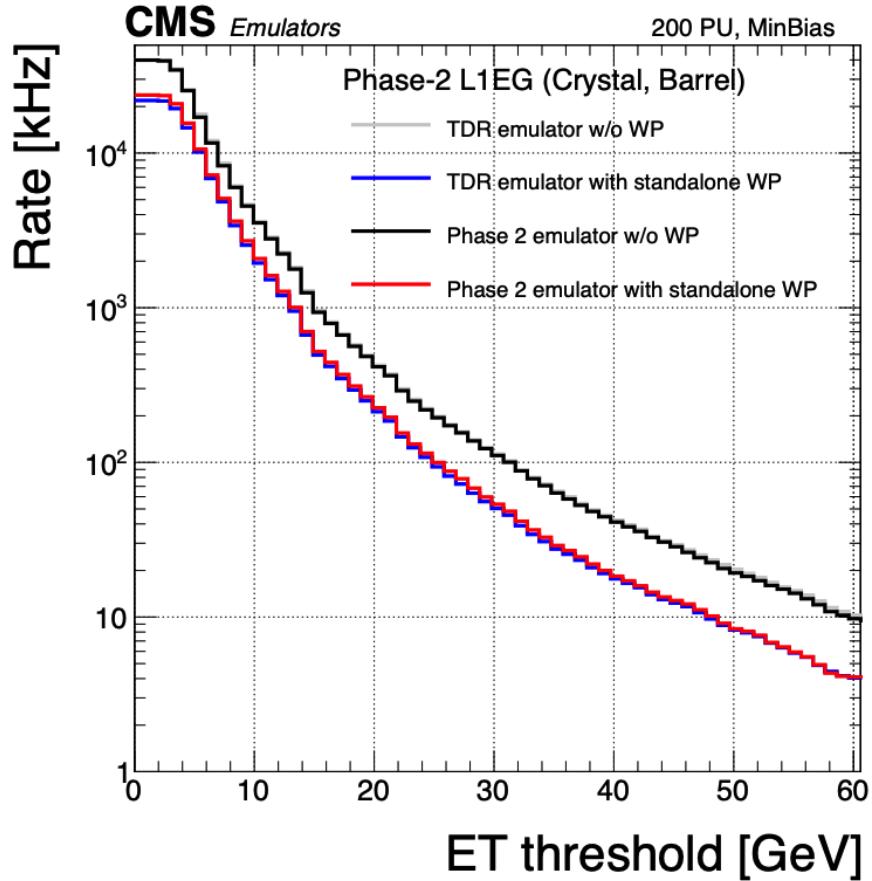


Figure 3.8: Rates of the standalone barrel e/γ reconstruction, evaluated on a minimum bias sample, measured as a function of the minimum energy (E_T) required of the reconstructed e/γ object in each event. The performance of the previous, idealized algorithm as shown in the 2021 Phase-2 TDR [35] with and without the isolation and shower shape discrimination variables (“standalone working point/ WP”) (*dark blue, grey*). The Phase-2 emulator discussed in this work with and without the same working point (*black, red*) is shown to have comparable performance.

1062

Chapter 4

1063

Datasets and Monte Carlo samples

1064

4.1 Datasets used

1065 The $h \rightarrow aa \rightarrow 2b2\tau$ analysis (CMS CADI line HIG-22-007) is based on proton-proton
1066 collision data at a center-of-mass energy of 13 TeV collected in full Run-2 (2016-
1067 18) with the CMS detector. The data analyzed corresponds to a total integrated
1068 luminosity of 138 fb^{-1} (36.33 fb^{-1} for 2016, 41.53 fb^{-1} for 2017, and 59.74 fb^{-1} for
1069 2018) [37] [38] [39]. The cumulative delivered and recorded luminosity versus time
1070 for 2015-2018 is shown in Fig. 4.1.

1071 Data collected with the single muon trigger is used for the $\mu\tau_h$ channel. For the
1072 $e\tau_h$ channel, data collected with the single electron trigger is used; and for the $e\mu$
1073 channel, data collected with the electron + muon trigger is used. A more in-depth
1074 discussion of the triggers used follows in a later section.

1075 A full list of samples used can be found in the full documentation [41] [42].

1076

4.2 Monte Carlo samples

1077 Modeling and computing observables originating from arbitrary physics processes at
1078 the tree level and at next-to-leading order (NLO) is performed by Monte Carlo (MC)

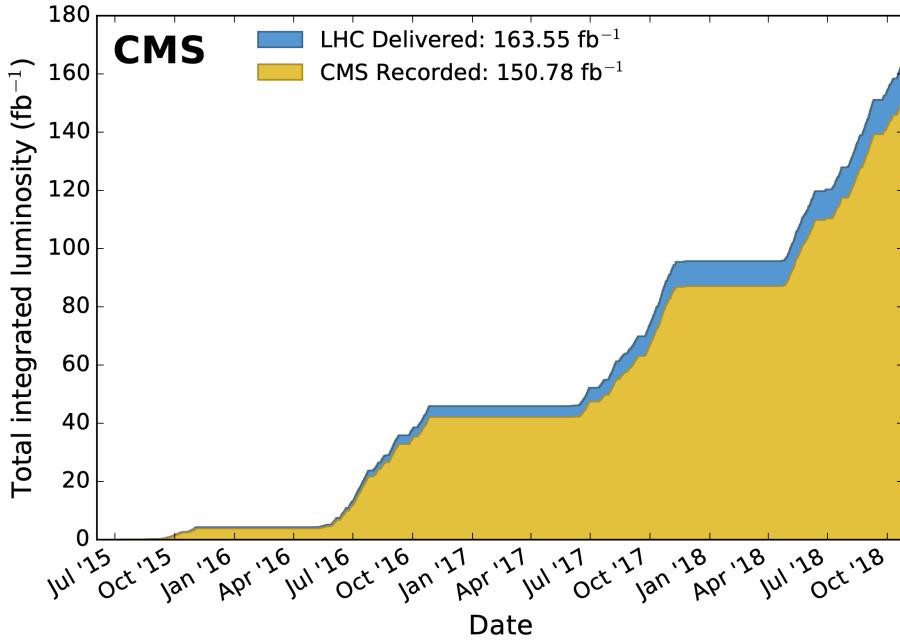


Figure 4.1: Cumulative delivered and recorded luminosity versus time for 2015-2018 at CMS, in proton-proton collision data only, at nominal center-of-mass energy [40].

event generators, such as Powheg and MadGraph5_amCNLO [43] [44]. The information generated, e.g. the computation of the differential cross sections and kinematics of the final state particles, is saved in a compressed file and used to generate MC samples that are used in physics analyses. The samples are digitized using GEANT4 [45], a platform used at the LHC and other facilities to comprehensively simulate the passage of particles through matter. The digitized samples are passed through the same detector reconstruction as real data events collected in the detector.

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

1091 4.3 Embedded samples

1092 An important background for Higgs boson studies and searches for additional Higgs
1093 bosons is the decay of Z bosons into pairs of τ leptons ($Z \rightarrow \tau\tau$). An embedded tech-
1094 nique was developed in the context of Standard Model Higgs to $\tau\tau$ measurements, to
1095 model $Z \rightarrow \tau\tau$ decays, and was expanded to also model all Standard Model processes
1096 that contain $\tau\tau$ [46]. The embedded technique has since been used successfully at
1097 CMS for the Standard Model $H \rightarrow \tau\tau$ measurement, as well as searches for minimal
1098 supersymmetric extensions to the Standard Model (MSSM) [47] [48].

1099 Fig. 4.2 shows a schematic of how embedded samples are produced. Data events
1100 containing $Z \rightarrow \mu\mu$ decays are selected. In these events, all energy deposits of the
1101 recorded muons are removed, and are replaced with simulated tau leptons with the
1102 same kinematic properties as the removed muons. This results in a hybrid data format
1103 containing information from both observed and simulated events, as illustrated in Fig.
1104 4.2 [46].

1105 In the selection step of the embedded technique, events are selected with at least
1106 one of a set of $\mu\mu$ trigger paths, which require $p_T > 17(8)$ GeV for the leading
1107 (sub-leading) muons, and a minimum requirement between 3.8 and 8.0 GeV on the
1108 invariant di-muon mass $m_{\mu\mu}$ [46]. The offline reconstructed muons must match the
1109 objects at trigger level and also have offline $p_T > 17(8)$ GeV. They must have $|\eta| < 2.4$
1110 and be located at a distance $|d_z| < 0.2$ cm to the primary vertex along the beam
1111 axis. To form a Z boson candidate, each muon is required to originate from a global
1112 muon track. The muon pairs must have opposite charges with an invariant mass of
1113 $m_{\mu\mu} > 20$ GeV. If more than two di-muon pairs are found, the pair with the invariant
1114 mass closest to the Z boson mass (91.19 GeV) is chosen.

1115 This selection is designed to be tight enough to ensure a high purity of genuine
1116 $\mu\mu$ events, and also loose enough to minimize biases of the embedded event samples.
1117 Isolation requirements are avoided, since they would introduce a bias towards less

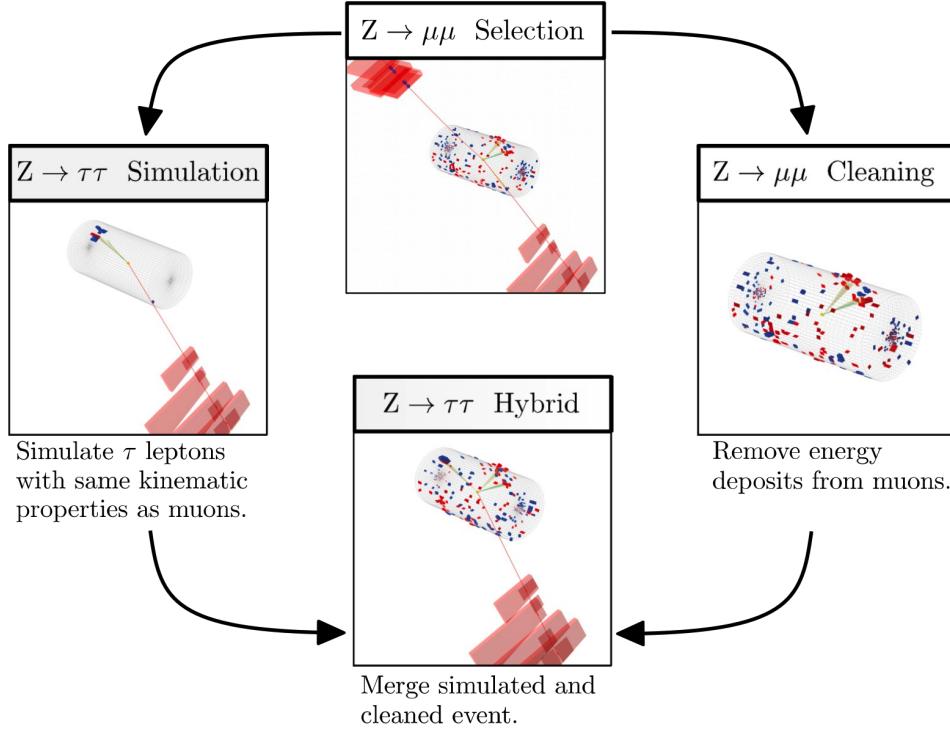


Figure 4.2: Schematic view of the four main steps of the embedding technique for τ leptons, as described in Section 4.3 [46]. A $Z \rightarrow \mu\mu$ event is selected in data ($Z \rightarrow \mu\mu$ selection), all of the energy deposits associated with the muons are removed ($Z \rightarrow \mu\mu$ 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).

hadronic activity in the vicinities of the embedded leptons that will appear more isolated than expected in data. The selection results in an expected mixture of events summarized in Table 4.1 from [46]. $Z \rightarrow \mu\mu$ is the dominant process modeled by the embedded technique, with $t\bar{t}$, QCD, and diboson and single top processes becoming more significant when considering events with b-tag jets.

The advantage of the embedded technique is that aspects of the event that are difficult to model and describe are directly taken from data, resulting in a better data description than can be achieved with only the $Z \rightarrow \tau\tau$ simulation [46]. The simulation must be tuned extensively to accurately model aspects of the data, such as time-dependent pileup profiles, the production of additional jets, e.g. in multijet and vector boson fusion topologies, the number of reconstructed primary interaction

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

Table 4.1: Expected event composition after selecting two muons in the embedded technique [46], before additional cuts (i.e. *inclusive*, *column 2*), and after adding a requirement on the di-muon mass $m_{\mu\mu} > 70 \text{ GeV}$ (*column 3*), or a requirement on the number of b-tag jets in the event (*column 4*).

1129 vertices, and the missing transverse momentum p_T^{miss} . Since all events with genuine
 1130 $\tau\tau$ are estimated with samples made with the embedded technique (referred to as
 1131 embedded samples from here on), events in Monte Carlo simulation with genuine $\tau\tau$
 1132 are not used, in order to avoid double-counting.

¹¹³³ **Chapter 5**

¹¹³⁴ **Object reconstruction and
corrections applied**

¹¹³⁶ In this chapter on object reconstruction and corrections, Section 5.1 reviews the
¹¹³⁷ physical properties of the objects most pertinent to the analyses presented in this
¹¹³⁸ work: taus (τ), muons (μ), electrons (e), and jets, with a focus on jets originating from
¹¹³⁹ b quarks (b-flavor jets), as well as the methodology used to reconstruct the particles
¹¹⁴⁰ from their characteristic signatures in the CMS detector. Section 5.2 describes the
¹¹⁴¹ method used to reconstruct the invariant full $\tau\tau$ mass which is used for the final signal
¹¹⁴² extraction. Lastly, Section 5.3 describes the corrections applied to the simulated
¹¹⁴³ samples which improve their modeling of data.

¹¹⁴⁴ **5.1 Object reconstruction**

¹¹⁴⁵ **5.1.1 Taus**

¹¹⁴⁶ The tau (τ) is the heaviest known lepton. With a rest mass of 1776.86 MeV, it can
¹¹⁴⁷ decay to not only electrons and muons, but also hadrons. The mean lifetime of the τ
¹¹⁴⁸ is $\tau = 290 \times 10^{-15}$ seconds, corresponding to $c\tau = 87.03 \mu\text{m}$, which is short enough

1149 that taus decay in the CMS detector before reaching the detector elements.

1150 In two thirds of the cases, τ leptons decay hadronically, typically into one or three
1151 charged mesons (predominantly π^+ , π^-), often accompanied by neutral pions (that
1152 decay $\pi^0 \rightarrow \gamma\gamma$), and a ν_τ . These hadronic decays are denoted τ_h . In the remainder of
1153 the decays, the tau decays to the lighter leptons (electron or muon), termed leptonic
1154 decays. In all cases, at least one neutrino is produced, resulting in missing transverse
1155 energy in the CMS detector. The tau's largest decay branching ratios (proportional
1156 to probability of decay) are listed below [26]:

1157 • 17.8% decay to $e^- \bar{\nu}_e \nu_\tau$

1158 • 17.4% decay to $\mu^- \bar{\nu}_\mu \nu_\tau$

1159 • 25.5% decay to $\pi^- \pi^0 \nu_\tau$ (ρ^- resonance at 770 MeV)

1160 • 10.8% decay to $\pi^- \nu_\tau$

1161 • 9.3% decay to $\pi^- \pi^0 \pi^0 \nu_\tau$ (a_1^- resonance at 1200 MeV)

1162 • 9.0% decay to $\pi^- \pi^- \pi^+ \nu_\tau$ (a_1^- resonance at 1200 MeV)

1163 The neutrinos escape undetected from the CMS detector and are not considered
1164 in the reconstruction. Charged hadrons leave tracks in the tracking detector before
1165 being absorbed in the hadronic calorimeter; in CMS tau reconstruction terminology,
1166 they are often called “prongs”, i.e. the dominant τ_h decay modes are termed “1 prong”
1167 (π^\pm), “1 prong + $\pi^0(s)$ ”, and “3-prong”. Neutral pions decay to two photons which
1168 lose their energy in the electromagnetic calorimeter. Taus that decay to electrons
1169 and muons, are typically triggered on and reconstructed as electrons and muons
1170 respectively.

1171 **Hadron plus strips (HPS) reconstruction of τ_h**

1172 At CMS, hadronically decaying tau leptons are reconstructed with the hadron plus
1173 strips (HPS) algorithm [49] [50]. The HPS algorithm capitalizes on photon conversions
1174 in the CMS tracker material, which originate from the neutral pion (π^0) decaying
1175 to two photons. The bending of electron/positron tracks due to the CMS solenoid
1176 magnetic field leads to a spread of the neutral pions' calorimeter signatures in the ϕ
1177 direction. This motivates the reconstruction of photons in “strips”: objects that are
1178 built out of PF photons and electrons. The strip reconstruction starts with centering
1179 a strip on the most energetic electromagnetic particle in a PF jet. Among other
1180 electromagnetic particles located in a window of size $\Delta\eta = 0.05$ and $\Delta\phi = 0.20$
1181 around the strip center, the most energetic one is associated with the strip and its
1182 momentum is added to the strip momentum. This is repeated iteratively until no
1183 further particles can be associated. Lastly, strips satisfying a requirement of $p_T^{\text{strip}} > 1$
1184 GeV are combined with charged hadrons to reconstruct individual τ_h decay modes,
1185 where h stands for both π and K :

1186 • *Single hadron:* $h^- \nu_\tau$ and $h^- \pi^0 \nu_\tau$ decay modes, in which the neutral pions have
1187 too little energy to be reconstructed as strips.

1188 • *One hadron + one strip:* $h^- \pi^0 \nu_\tau$ decay modes, where the photons from the π^0
1189 decay are close together in the calorimeter.

1190 • *One hadron + two strips:* $h^- \pi^0 \nu_\tau$ decay modes, where the photons from the π^0
1191 decay are well separated.

1192 • *Three hadrons:* $h^- h^+ h^- \nu_\tau$ decay modes. The three charged hadrons are re-
1193 quired to originate from the same secondary vertex.

1194 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
1195 reconstructed with the above topologies.

1196 In the HPS algorithm, the direction of the reconstructed tau momentum $\vec{p}_T^{\tau_h}$
1197 is required to fall within a distance of $\Delta R = 0.1$ from the original PF jet. All
1198 charged hadrons and strips are required to be contained within a cone of size $\Delta R =$
1199 $(2.8 \text{ GeV})/p_T^{\tau_h}$, from the τ_h as reconstructed by the HPS.

1200 All charged hadrons are assumed to be pions, and they are required to be consis-
1201 tent with the masses of the intermediate meson resonances (if applicable), with the
1202 following allowed windows for candidates: 50-200 MeV for π^0 , 0.3-1.3 GeV for ρ , and
1203 0.8-1.5 GeV for a_1 . If the τ_h decay is compatible with more than one hypothesis, the
1204 one giving the highest $p_T^{\tau_h}$ is chosen. Lastly, an isolation requirement is applied: aside
1205 from the τ_h decay products, no charged hadrons or photons can be present within
1206 an isolation cone of size $\Delta R = 0.5$ around the direction of the τ_h . The outputs of
1207 the HPS algorithm are the reconstructed decay mode and the visible four-momentum
1208 (i.e. the four-momenta of all decay products excluding the neutrinos).

1209 DeepTau for identifying τ_h

1210 The identification of τ_h candidates in CMS has historically been divided into separate
1211 discriminators against jets, electrons, and muons. Discriminators versus jets and
1212 electrons use information from derived quantities, such as the p_T sum of particles
1213 near the τ_h axis. Building on the previous multivariate analysis (MVA) classifier [51]
1214 based on a boosted decision tree (BDT), DeepTau is a more recent classifier based on a
1215 deep neural network (DNN) that simultaneously discriminates against jets, electrons,
1216 and muons. The DNN uses a combination of high-level inputs, similar to previous
1217 algorithms, and also uses convolutional layers in $\eta\text{-}\phi$ space to process information
1218 from all reconstructed particles near the τ_h axis. Convolutional layers are based on
1219 the principle that an image can be processed independently of its position.

1220 The final DeepTau discriminators against jets, muons, and electrons are given by

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

1221 where y_τ (y_α) are estimates of the probabilities for the τ_h candidate to come from
1222 a genuine τ_h (jet, μ , e). Working points for each discriminator with different τ_h
1223 identification efficiencies are defined for D_e , D_μ , and D_{jet} , for usage in physics analyses
1224 and derivation of data-to-simulation corrections [52].

1225 5.1.2 Muons

1226 Muons are the next lightest lepton after taus, with a mass of 105.66 MeV and a
1227 mean lifetime of $\tau = 2.20 \times 10^{-6}$ seconds, or $c\tau = 658.64$ m. At CMS, muons are
1228 identified with requirements on the quality of the track reconstruction and on the
1229 number of measurements in the tracker and the muon systems [53]. In the standard
1230 CMS reconstruction, tracks are first reconstructed independently in the inner tracker
1231 (tracker track) and in the muon system (standalone-muon track). Next, these tracks
1232 are processed in two different methods.

1233 The first is Global Muon reconstruction (outside-in) [53], which fits combined hits
1234 from the tracker track and standalone-muon track, using the Kalman-filter technique.
1235 At large transverse momenta, $p_T \gtrsim 200$ GeV, the global-muon fit can improve the
1236 momentum resolution compared to the tracker-only fit.

1237 The second is Tracker Muon reconstruction (inside-out) [53], which starts with
1238 tracker tracks with $p_T > 0.5$ GeV and total momentum $p_T > 2.5$ GeV. These tracks
1239 are extrapolated outwards to the muon system and matched to detector segments
1240 there, taking into account the magnetic field, expected energy losses, and multiple
1241 Coulomb scattering in the detector material. Tracker Muon reconstruction is more
1242 efficient than the Global Muon reconstruction at low momenta, $p \lesssim 5$ GeV, because

1243 it only requires a single muon segment in the muon system, whereas Global Muon
1244 reconstruction typically requires segments in at least two muon stations.

1245 To further suppress fake muons from decay in flight, isolation cuts are used. A
1246 relative isolation variable is defined to quantify the energy flow of particles near the
1247 muon trajectory. A relative isolation is defined similarly for muons and electrons:

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

1248 where $\sum_{\text{charged}} p_T$ is the scalar sum of the p_T of the charged particles originating from
1249 the primary vertex and located in a cone of size $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4(0.3)$
1250 centered on the direction of the muon (electron). The sum $\sum_{\text{neutral}} p_T$ is the equivalent
1251 for neutral particles. The sum $\sum_{\text{charged, PU}} p_T$ is the scalar sum of the p_T of the
1252 charged hadrons in the cone originating from pileup vertices. The factor 1/2 comes
1253 from simulation estimations, which find that the ratio of neutral to charged hadron
1254 production in the hadronization process of inelastic pp collisions is 1/2. Thus the
1255 subtracted term is intended to subtract contribution from pileup, from the neutral
1256 particle contribution to the isolation sum. Finally, this is divided by the lepton
1257 transverse momentum, p_T^ℓ .

1258 5.1.3 Electrons

1259 Electrons are the lightest lepton with a mass of 0.511 MeV. At CMS, electrons are
1260 reconstructed by associating a track reconstructed in the silicon tracking detector
1261 with a cluster of energy in the ECAL. Performance is maximized via a combination
1262 of a stand-alone approach and the complementary global particle-flow approach [54].

1263 In the stand-alone approach, the electron energy, which is typically spread over
1264 several crystals of the ECAL, is clustered with the “hybrid” algorithm in the barrel
1265 and the “multi- 5×5 ” in the endcaps [54]. The hybrid algorithm collects energy in a

1266 small window in η and an extended window in ϕ . It identifies a seed crystal, and adds
1267 arrays of 5×1 crystals in $\eta \times \phi$ in a range of $N = 17$ crystals in both directions of
1268 ϕ , if their energies exceed a minimum threshold, thus forming a supercluster (SC). In
1269 the endcap, crystals are not arranged in an $\eta \times \phi$ geometry; instead clusters are build
1270 around seed crystals in clusters of 5×5 crystals that can partly overlap. Nearby
1271 clusters are grouped into a supercluster, and energy is recovered from associated
1272 deposits in the preshower.

1273 In the PF reconstruction [54], PF clusters are reconstructed by aggregating around
1274 a seed all contiguous crystals with energies two standard deviations above the elec-
1275 tronic noise observed at the beginning of a data-taking run. The energy of a given
1276 crystal can be shared among two or more clusters.

1277 The electron track reconstruction is performed in two ways [54]: the ECAL-based
1278 seeding, which begins with the SC energy and positioning, and the tracker-based
1279 seeding (part of the PF reconstruction algorithm), which uses tracks reconstructed
1280 from the general algorithm for charged particles, extrapolated towards the ECAL and
1281 matched to an SC. Kalman filter (KF) tracks with a small number of hits or that are
1282 not well-fitted, are re-fitted with a dedicated Gaussian sum Filter (GSF).

1283 A global identification variable [54] is defined using a multivariate analysis (MVA)
1284 technique that combines information on track observables (kinematics, quality of the
1285 KF track and GSF track), the electron PF cluster observables (shape and pattern),
1286 and the association between the two (geometric and kinematic observables). For
1287 electrons seeded only through the tracker-based approach, a weak selection is applied
1288 on this MVA variable. For electrons seeded through both approaches, a logical OR is
1289 taken.

1290 Electron isolation, i.e. the presence of energy deposits near the electron trajectory,
1291 is a separate key handle in rejecting significant background. Compared to isolated
1292 electrons, electrons from misidentified jets or genuine electrons within a jet resulting

1293 from semileptonic decays of b or c quarks tend to have significant energy deposits
1294 near the primary trajectory [54]. Offline analyses benefit from the PF technique
1295 for defining isolation, which sums the PF candidates reconstructed located within a
1296 specified isolation cone around the electron candidate, as in Eqn. 5.2.

1297 5.1.4 Jets

1298 The vast majority of processes of interest at the LHC contains quarks or gluons in
1299 the final state, but these particles cannot be observed directly. In a process called
1300 hadronization, they fragment into spatially-grouped collections of particles called jets,
1301 which can be detected in the tracking and calorimeter systems. Hadronization and
1302 the subsequent decays of unstable hadrons can produce hundreds of nearby particles
1303 in the CMS detector. Jets are reconstructed by the PF algorithm (PF jets), or from
1304 the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo
1305 jets). In PF jets, typically used in offline analyses, jets are built using the anti- k_T
1306 (AK) clustering algorithm [55]. The anti- k_T algorithm iterates over particle pairs and
1307 finds the two that are closest in a distance measure d , and determines whether to
1308 combine them:

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

1309 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-
1310 ity, and azimuthal angle of particle i . The power -2 means that higher-momentum
1311 particles are clustered first, leading to jets that tend to be centered on the hardest
1312 (highest p_T) particle.

1313 There are several methods to remove contributions of pileup collisions from jet
1314 clustering [56]:

- 1315 • Charged hadron subtraction (CHS), which removes all charged hadron candi-

1316 dates associated with a track that is not associated with the primary vertex.

- 1317 • PileUp Per Particle Identification (PUPPI), which weighs input particles based
1318 on their likelihood of arising from pileup. QCD particles tend to have a collinear
1319 structure, compared to soft diffuse radiation coming from pileup. The local
1320 shape for charged pileup, used as a proxy for all pileup particles, is used on an
1321 event-by-event basis to calculate a weight for each particle. PUPPI is deployed
1322 in Run-2 and is more performant than CHS in high pileup scenarios.

1323 5.1.5 B-flavored jets

1324 Jets that arise from bottom-quark hadronization (b-flavor jets) have overwhelming
1325 background from processes involving jets from gluons (g) and light-flavor quarks (u, d,
1326 s), and from c-quark fragmentation. The ability to identify b-flavor jets, or b-tagging,
1327 exploits the b hadrons' relatively large masses, long lifetimes, and daughter particles
1328 with hard momentum spectra [55].

1329 The impact parameter (IP) of a track is the 3-dimensional distance between the
1330 track and the primary vertex (PV) at the point of closest approach. The IP is positive
1331 if the track originates from the decay of particles travelling along the jet axis. The
1332 resolution of the IP depends on the p_T and η of the track, motivating the use of the
1333 impact parameter significance S_{IP} (ratio of the IP to its estimated uncertainty) as an
1334 observable [55].

1335 Because of the large but finite lifetimes of the b hadrons, b hadrons tend to
1336 travel a short distance before decaying at a secondary vertex (SV), which can be
1337 measured and reconstructed separately from the primary vertex due to the excellent
1338 position resolution of the pixel detector [55]. Previous b-tagging algorithms (e.g.
1339 CSV, cMVAv2, and DeepCSV) have capitalized on variables such as the presence of
1340 a SV, the flight distance and direction (computed from the vector between the PV
1341 and the SV), and kinematics of the system of associated secondary tracks (e.g. track

1342 multiplicity, mass, and energy).

1343 The DeepJet (formerly known as DeepFlavour) algorithm [57] is a deep-neural-
1344 network multi-classification algorithm, which uses 16 properties of up to 25 charged
1345 and 6 properties of 25 neutral particle-flow jet constituents, as well as 17 properties
1346 from up to 4 secondary vertices associate with the jet. Compared to the previous clas-
1347 sifying algorithm DeepCSV, DeepJet has been demonstrated to have higher efficiency
1348 with lower misidentification probability in Phase-1 data [58].

1349 5.2 Reconstruction of the $\tau\tau$ mass

1350 The final signal extraction is done to the total $\tau\tau$ mass, which is estimated from the
1351 visible $\tau\tau$ mass using the FastMTT algorithm [59]. FastMTT is based on the SVFit
1352 algorithm, originally developed for the Standard Model $H \rightarrow \tau\tau$ analysis [60]. Both
1353 the SVFit algorithms, and the FastMTT algorithm, are described below, to give a
1354 complete picture of how tau decays are parameterized.

1355 To specify a hadronic τ decay, six parameters are needed [60]: the polar and
1356 azimuthal angles of the visible decay product system in the τ rest frame, the three
1357 boost parameters from the τ rest frame to the laboratory frame, and the invariant
1358 mass m_{vis} of the visible decay products. For a leptonic τ decay, two neutrinos are
1359 produced, and a seventh parameter, the invariant mass of the two-neutrino system, is
1360 necessary. The unknown parameters are constrained by four observables that are the
1361 components of the four-momentum of the system formed by the visible decay products
1362 of the τ lepton, measured in the laboratory frame. The remaining unconstrained
1363 parameters for hadronic and leptonic τ decays are thus:

- 1364 • The fraction of the τ energy in the laboratory frame carried by the visible decay
1365 products,
- 1366 • ϕ , the azimuthal angle of the τ direction in the laboratory frame,

- 1367 • $m_{\nu\nu}$, the invariant mass of the two-neutrino system in leptonic τ decays (for
 1368 hadronic τ decays, $m_{\nu\nu}$ is set to 0).

1369 E_x^{miss} and E_y^{miss} , the x and y components of the missing transverse energy \vec{E}_T^{miss}
 1370 provide two further constraints.

1371 **5.2.1 Original SVFit “standalone”: maximum likelihood**

1372 In one of the original versions of SVFit, called “standalone” SVFit [60], a maximum
 1373 likelihood fit method is used to reconstruct the mass $m_{\tau\tau}$ by combining the measured
 1374 observables E_x^{miss} and E_y^{miss} with a likelihood model that includes terms for the τ
 1375 decay kinematics and the \vec{E}_T^{miss} resolution [60]. The likelihood function $f(\vec{z}, \vec{y}, \vec{a}_1 \vec{a}_2)$
 1376 of the parameters $\vec{z} = (E_x^{\text{miss}}, E_y^{\text{miss}})$ in an event is constructed, where the remaining
 1377 parameters are the kinematics of the two τ decays, denoted $\vec{a}_1 = (x_1, \phi_1, m_{\nu\nu,1})$ and
 1378 $\vec{a}_2 = (x_2, \phi_2, m_{\nu\nu,2})$, and the four-momenta of the visible decay products with the
 1379 measured values $\vec{y} = (p_1^{\text{vis}}, p_2^{\text{vis}})$.

1380 The likelihood f is the product of three likelihood functions. The first two likelihood
 1381 functions model the decay parameters \vec{a}_1 and \vec{a}_2 of the two τ leptons. For leptonic
 1382 decays, the likelihood function is modeled using matrix elements for τ decays,
 1383 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
 1384 hadronic τ decays, a model based on the two-body phase space is used and integrated
 1385 over $m_{\text{vis}}^2/m_{\tau\tau}^2 \leq x \leq 1$. The third likelihood function quantifies the compatibility of
 1386 a τ decay hypothesis with the reconstructed \vec{E}_T^{miss} in an event, assuming the neutrinos
 1387 are the only source of missing transverse energy. The expected \vec{E}_T^{miss} resolution
 1388 is represented by a covariant matrix, estimated on an event-by-event basis using a
 1389 significance algorithm [61].

1390 5.2.2 “Classic SVFit” with matrix element

1391 Classic SVFit is an improved algorithm of the original “standalone” SVFit using the
 1392 formalism of the matrix element (ME) method [59]. In the ME method, an estimate
 1393 for the unknown model parameter Θ (here, the mass $m_{\tau\tau}$) is obtained by maximizing
 1394 the probability density \mathcal{P} . The key ingredients of the probability density are the
 1395 squared modulus of the matrix element $|\mathcal{M}(\mathbf{p}, \Theta)|^2$ and the transfer function $W(\mathbf{y}|\mathbf{p})$
 1396 (probability density to observe the measured observables \mathbf{y} given the phase space
 1397 point \mathbf{p}). The best estimate $m_{\tau\tau}$ is obtained by computing the probability density \mathcal{P}
 1398 for a range of mass hypotheses and finding the value of $m_{\tau\tau}$ that maximizes \mathcal{P} .

1399 Distributions illustrating the performance of the classic matrix element SVFit
 1400 algorithm are shown in Fig. 5.1 from [59], showing the di-tau mass after and before
 1401 application of SVFit to recover energy lost to neutrinos. The SVFit algorithm is
 1402 found to improve the sensitivity of the Standard Model $H \rightarrow \tau\tau$ analysis performed
 1403 by CMS by about 30%, compared to performing the same analysis using only the
 1404 visible mass m_{vis} .

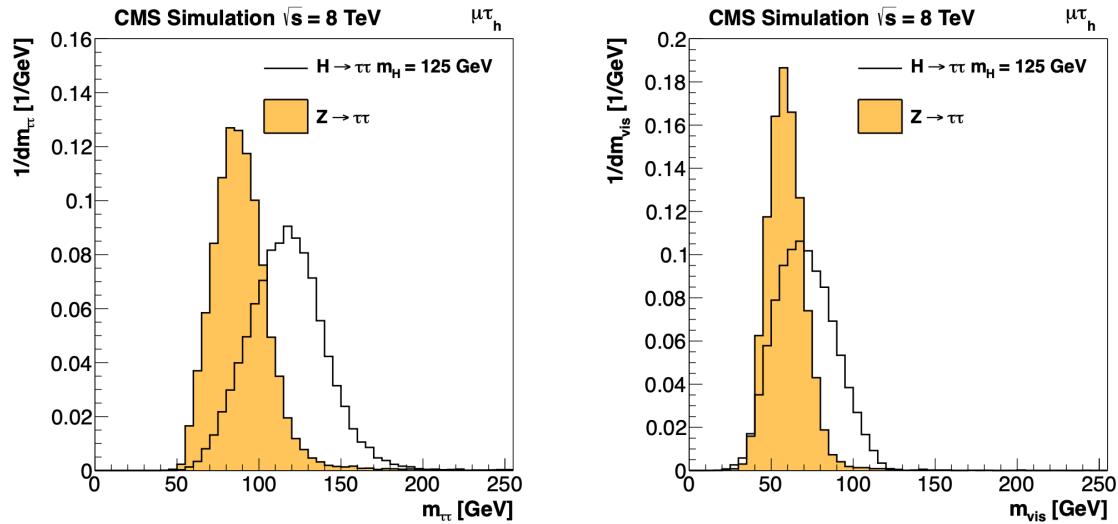


Figure 5.1: Distributions from [59], of $m_{\tau\tau}$ after reconstruction with the original SVFit algorithm (*left*), and before SVFit with only the visible tau decay products (*right*), for $H \rightarrow \tau\tau$ signal events of mass $m_H = 125$ GeV (*black line*) and the $Z/\gamma^* \rightarrow \tau\tau$ background (*orange, solid*), in the decay channel $\tau\tau \rightarrow \mu\tau_h$.

5.2.3 FastMTT: optimized SVFit

FastMTT [62] is a further simplification to the matrix element method of Classic SVFit which has comparable performance but is about 100 times faster. FastMTT drops the matrix element component of the computation without significant impact on the final mass resolution, and simplifies the computation of the transfer functions. The opening angle of the τ decay products with respect to the initial τ momenta approaches 0 for τ with high $\gamma = E_\tau/m_\tau$, with typical τ decays from the Z boson decays already satisfying this condition. In this collinear approximation, the dimensionality of the transfer function can be reduced in the computation of FastMTT, while still yielding similar results to Classic SVFit [62].

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 mis-modeling in simulations of the underlying physics process, or changing detector operating conditions during data-taking. Event-level corrections change the shapes of the distributions of all the physical observables.

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

1430 whether events pass or fail the offline p_T thresholds for the trigger paths described in
1431 the previous section, i.e. change the number of events in the signal region.

1432 5.3.1 Tau energy scale

1433 An energy scale is applied to the transverse momentum p_T and mass of the hadronic
1434 tau τ_h in the $\mu\tau_h$ and $e\tau_h$ channels, to correct for a deviation of the average recon-
1435 structed τ_h energy from the generator-level energy of the visible τ_h decay products.
1436 These correction factors are derived centrally [51], by fitting to events in $e\tau_h$ and $\mu\tau_h$
1437 final states in Z/γ^* events separately for the h^\pm , $h^\pm\pi^0$, and $h^\pm h^\mp h^\pm$ decays. The
1438 values used are shown in Table 5.1.

1439 When applying the energy scale to the τ_h , the 4-momentum of the missing trans-
1440 verse energy (MET) is adjusted such that the total 4-momenta of the τ_h and the MET
1441 remains unchanged [63].

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.

1442 5.3.2 Muon energy scale

1443 An energy scale is applied to the p_T and mass of genuine muons from τ decays in the
1444 $e\mu$ and $\mu\tau_h$ channels [64]. The applied values are the same for MC and embedded
1445 samples and are shown in Table 5.2. Following the SM $H \rightarrow \tau\tau$ analysis, Rochester
1446 corrections are not applied, and instead prescriptions from [65] 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 [66] [65].

1447 5.3.3 Electron energy scale

1448 Corrections to the electron energy scale are applied to genuine e from τ decays, and
 1449 are binned in two dimensions by electron p_T and η for barrel vs. endcap [67]. The
 1450 scale factors are binned in p_T and η for MC samples: e.g. values for 2018 are shown
 1451 in Fig. 5.2 from [68]. For embedded samples the electron energy scale is taken as
 1452 only binned in η (Table 5.3).

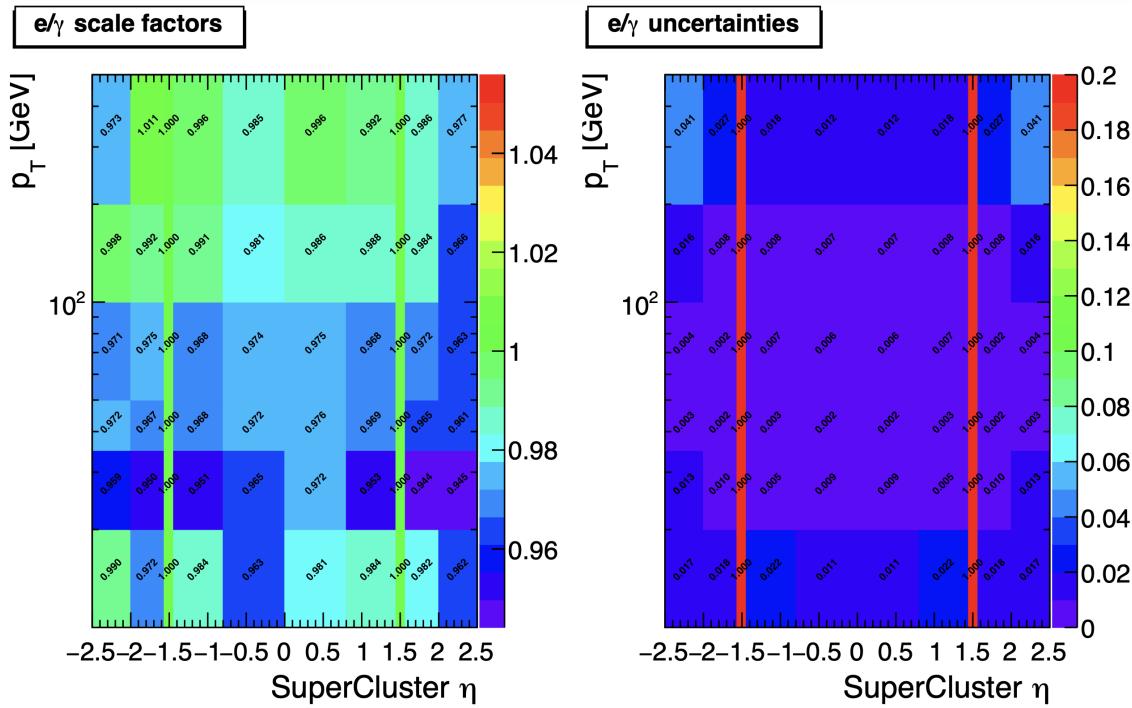


Figure 5.2: Electron/photon energy scale factors (*left*) and corresponding uncertainties (*right*) binned in the electron η and p_T , for the data-taking year 2018 [68].

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 [69] [70] [71].

5.3.4 τ_h identification efficiency

The τ_h identification efficiency can differ in data and MC [63]. Recommended corrections are provided by the Tau POG, and we use the medium DeepTau vs. jet working point values. The identification efficiency is measured in $Z \rightarrow \tau\tau$ events in the $\mu\tau_h$ 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 [63].

5.3.5 Trigger efficiencies

Scale factors are applied to correct for differences in trigger efficiencies between MC and embedded vs. data, with values taken from tools provided by the Standard Model $H \rightarrow \tau\tau$ working group which uses the same trigger paths [66]. In the following sections we review relevant trigger efficiencies in data, which form the basis of the trigger efficiency corrections applied to MC and embedded.

5.3.6 Tau trigger efficiencies

The efficiencies in data of the single- τ_h leg in $\mu\tau_h$, $e\tau_h$, and di- τ_h triggers is computed centrally per using a Tag and Probe (TnP) method [72] which is outlined here. In

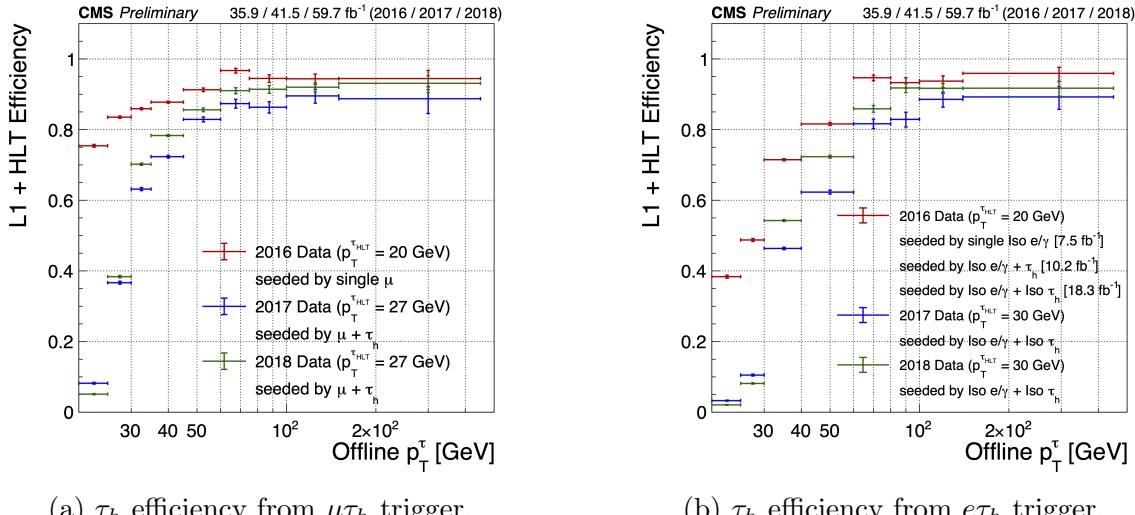
1467 this method, $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$ are selected in data and a Drell-Yan simulated sample
 1468 ($Z \rightarrow \ell\ell, \ell = e, \mu, \tau_h$) with high purity. Cuts are applied to reject events not in this
 1469 final state, e.g. suppressing $Z \rightarrow \mu\mu$ by vetoing events with a single loose ID muon.
 1470 An isolated muon candidate (the tag) with online $p_T > 27$ GeV and $|\eta| < 2.1$ is
 1471 identified and matched to an offline μ . An offline τ_h candidate (the probe) is selected,
 1472 which is separated from the tag μ , and has $p_T > 20$ GeV and $|\eta| < 2.1$. The probe
 1473 τ_h must pass anti-muon and anti-electron discriminators to avoid fakes from muons
 1474 and electrons, and must pass the medium MVA tau isolation to suppress fakes from
 1475 QCD jets. The trigger efficiency in the TnP method is calculated as

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

1476 The efficiencies for the hadronic tau legs in the relevant channels of this analyses
 1477 ($\mu\tau_h$ and $e\tau_h$) as a function of the offline tau p_T and η , are shown for data taken in
 1478 2016, 2017, and 2018 in Figures 5.3a and 5.3b [72] [73]. In both figures, the different
 1479 HLT thresholds and differences in the L1 seed result in higher efficiencies in 2016 and
 1480 differences in shapes of the 2016 efficiencies compared to 2017 and 2018. The low
 1481 pileup in 2016 also leads to higher efficiencies in that year.

1482 **5.3.7 Single muon trigger efficiencies**

1483 The efficiencies for the single isolated muon trigger with $p_T > 24$ GeV used in this
 1484 analysis, is shown for the data-taking year 2018 in Fig. 5.4a as a function of the muon
 1485 p_T and as a function of the muon $|\eta|$ in Fig. 5.4b from [74]. The data is split with
 1486 respect to a HLT muon reconstruction update that was deployed on 15/05/2018. A
 1487 small asymmetry in efficiencies between negative and positive η in Fig. 5.4b is due to
 1488 disabled muon chambers (CSCs). The efficiencies shown are estimated using a Tag
 1489 and Probe method using $Z \rightarrow \mu\mu$ events, with the tag being an offline muon with



(a) τ_h efficiency from $\mu\tau_h$ trigger.

(b) τ_h efficiency from $e\tau_h$ trigger.

Figure 5.3: Hadronic tau leg efficiency of the cross-triggers for $\mu\tau_h$ (*left*) and $e\tau_h$ (*right*) triggers as a function of offline tau p_T for the years 2016 (red), 2017 (blue) and 2018 (green), from [73]. HLT p_T thresholds and L1 seeds are indicated in the legends.

1490 $p_T > 29$ GeV and $|\eta| < 2.4$ passing a tight ID criteria, and the probe is an online (L1)
1491 trigger object with $\Delta R < 0.3$ and passing tight ID and Particle Flow based isolation
1492 requirements with $p_T > 26$ GeV.

1493 5.3.8 Single electron trigger efficiencies

1494 The efficiencies in data, and the ratio between data and MC, of the single electron
1495 HLT trigger with p_T threshold 32 GeV used in this analysis are shown for 2018, as
1496 a function of the electron p_T in Fig. 5.5a and of the electron $|\eta|$ in Fig. 5.5b, from
1497 [75]. In the Tag and Probe method used for the 2018 dataset, the tag is an offline
1498 reconstructed electron with $|\eta| \leq 2.1$ and not in the barrel and endcap overlap region,
1499 with $p_T > 35$ GeV with tight isolation and shower shape requirements, firing the tag
1500 trigger. The probe is an offline reconstructed electron with $|\eta| \leq 2.5$ with $E_T^{\text{ECAL}} > 5$
1501 GeV with no extra identification criteria [75].

1502 The disagreement between data and MC, particularly at low transverse momen-
1503 tum, is in part due to detector effects that are difficult to simulate, such as crystal

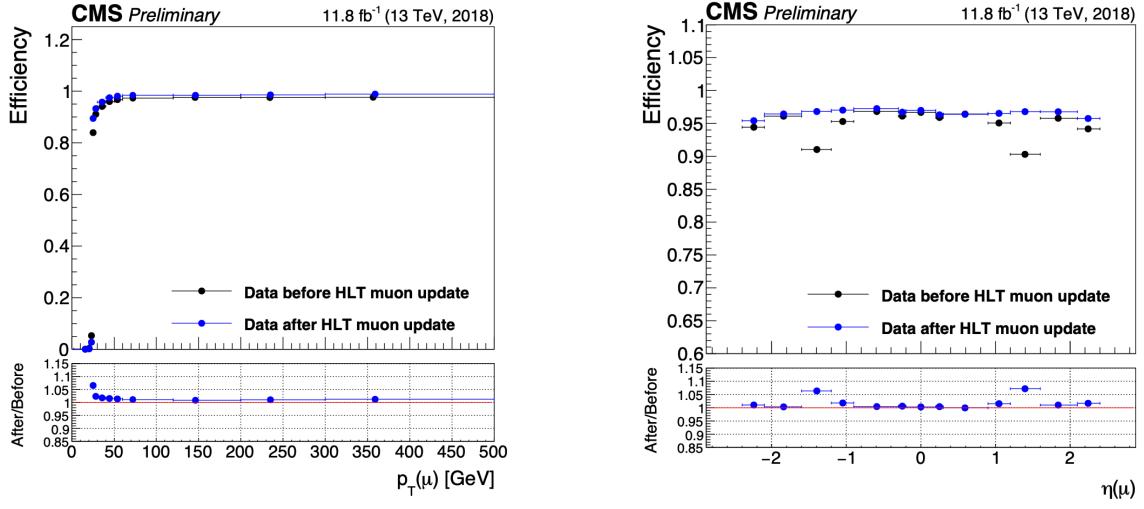
(a) Muon efficiency vs p_T for SingleMuon.(b) Muon efficiency vs $|\eta|$ for SingleMuon.

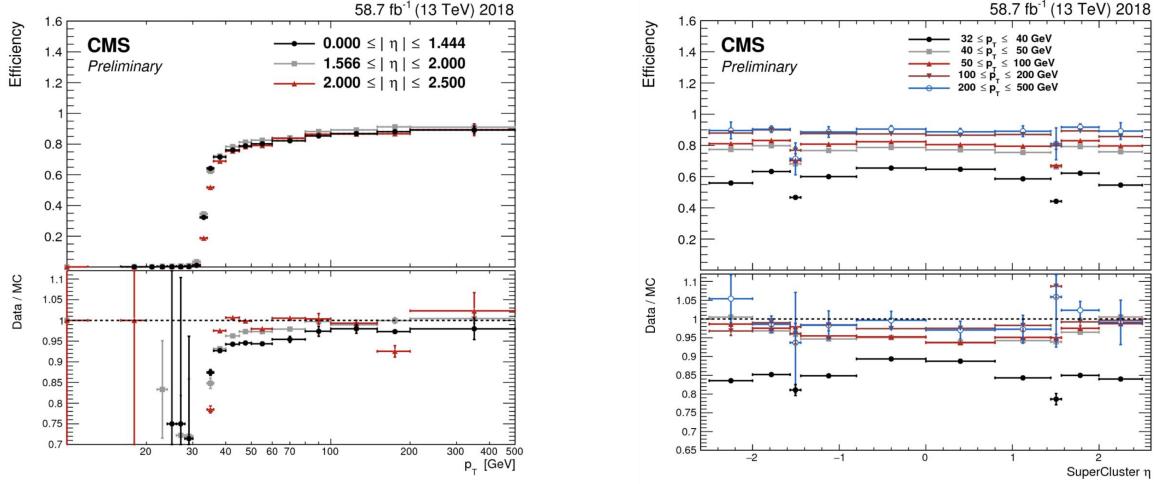
Figure 5.4: Trigger efficiencies in data (*top panels*) and ratio of efficiencies after/before a HLT muon reconstruction update (*bottom panels*) for the muon in the isolated single muon trigger with threshold $p_T > 24$ GeV in the data-taking year 2018, as functions of the muon p_T (*left*) and muon $|\eta|$ (*right*). Only statistical errors are shown [74].

transparency losses in the ECAL and the evolution of dead regions in the pixel tracker [75].

5.3.9 $e\mu$ cross-trigger efficiencies

The efficiencies of the electron and muons for the cross-trigger with leading muon used in the $e\mu$ channel are shown for data in 2016, 2017, and 2018 in Figures 5.6a and 5.6b [76]. These efficiencies were measured centrally using a Tag and Probe in events with Z to dileptons with the same flavor and opposite charge, where the tags are an isolated muon or electron, and the probe (offline) candidate is required to satisfy the same lepton selection as that of the tag candidate, be matched within $\Delta R < 0.1$ with a corresponding online trigger object, and also to pass the cross-trigger. The trigger efficiency is then:

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



(a) Electron efficiency vs p_T for single electron.

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

Figure 5.5: Trigger efficiencies in data, and the data/MC ratio for the electron in the single electron trigger with threshold $p_T > 32$ GeV in the data-taking year 2018, as functions of the electron p_T (left) and electron $|\eta|$ (right) [75]. 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.

1515 5.3.10 Electrons and muons faking τ_h : energy scales

1516 Energy scales for electrons misidentified as hadronic tau decays (e faking τ_h) are
 1517 provided by the Tau POG, and were measured in the $e\tau_h$ channel with the visible
 1518 invariant mass of the electron and hadronic tau system [66]. This energy scale is
 1519 applied for τ_h with $p_T > 20$ GeV regardless of which DeepTau vs. electron working
 1520 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 [66].

1521 No nominal energy scale is applied for muons mis-reconstructed as τ_h , and the

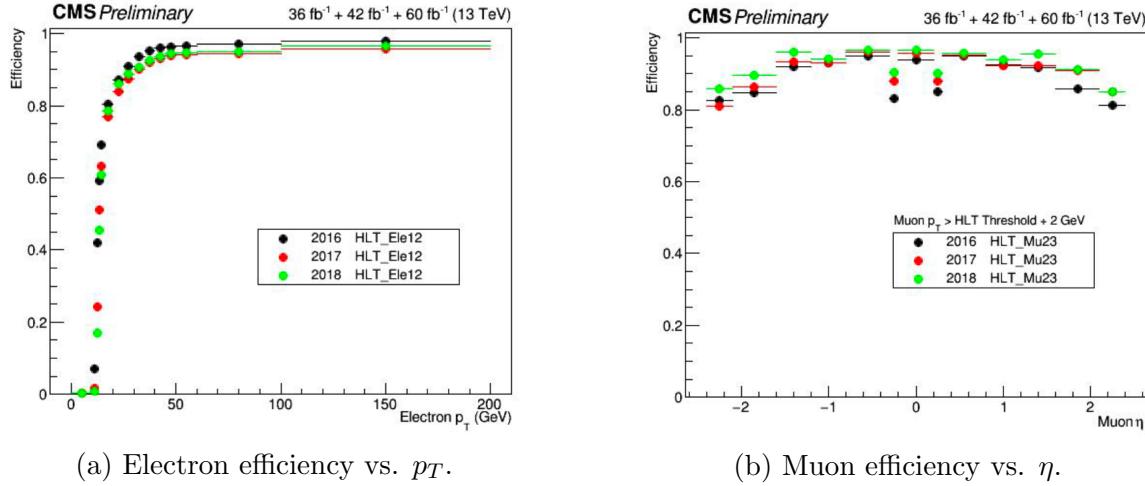


Figure 5.6: Efficiencies of the electron leg vs. p_T (*left*) and the muon leg vs. η (*right*), for the HLT path with online thresholds of 12 GeV for the electron and 23 GeV for the muon, for the data-taking years 2016 (*black*), 2017 (*red*), and 2018 (*green*) [76].

1522 uncertainty is treated as $\pm 1\%$ and uncorrelated in the reconstructed decay mode
1523 [66].

1524 5.3.11 Electrons and muons faking τ_h : misidentification effi- 1525 ciencies

1526 Corrections on identification efficiencies are applied to genuine electrons and muons
1527 misidentified as τ to account for differences in data and MC.

1528 The specific values depend on the vs. electron and vs. muon discriminator working
1529 points used. For misidentified $\mu \rightarrow \tau_h$, the scale factors are split into different $|\eta|$
1530 regions, determined by the CMS muon and tracker detector geometries, as shown in
1531 Table 5.6 for 2018 [63].

1532 For misidentified $e \rightarrow \tau_h$, the scale factors are split into barrel and endcap regions,
1533 dictated by the ECAL detector geometry, as shown in Table 5.7 for 2018.

Tau ID efficiency for DeepTau vs. muon WPs in 2018		
$ \eta $	Tight working point	VLoose working point
(0.0, 0.2)	0.767 ± 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 $|\eta|$ [63].

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 $|\eta|$ [63].

5.3.12 Electron ID and tracking efficiency

1534 Scale factors are applied to MC to correct for differences between MC and data in
 1535 the performance of electron identification (ID) and tracking.

1537 Electron and photon identification, as discussed earlier, use variables with good
 1538 signal vs. background discrimination power such as lateral shower shape and ratio
 1539 of energy deposited in the HCAL to energy deposited in the ECAL at the position
 1540 of the electron. The cut-based electron identification efficiencies in data and ratio of
 1541 efficiencies in data to MC are shown in Fig. 5.7a for the multivariate analysis (MVA)
 1542 identification working point.

1543 The tracking efficiencies in data and the data/MC ratio are shown in Fig. 5.7b
 1544 for the Gaussian-sum filter (GSF) tracking [77].

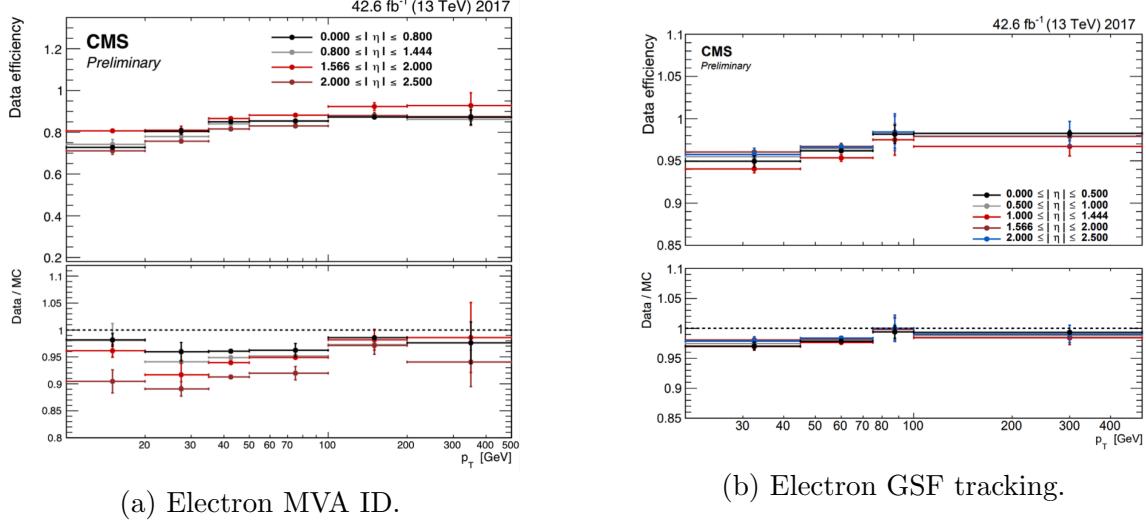


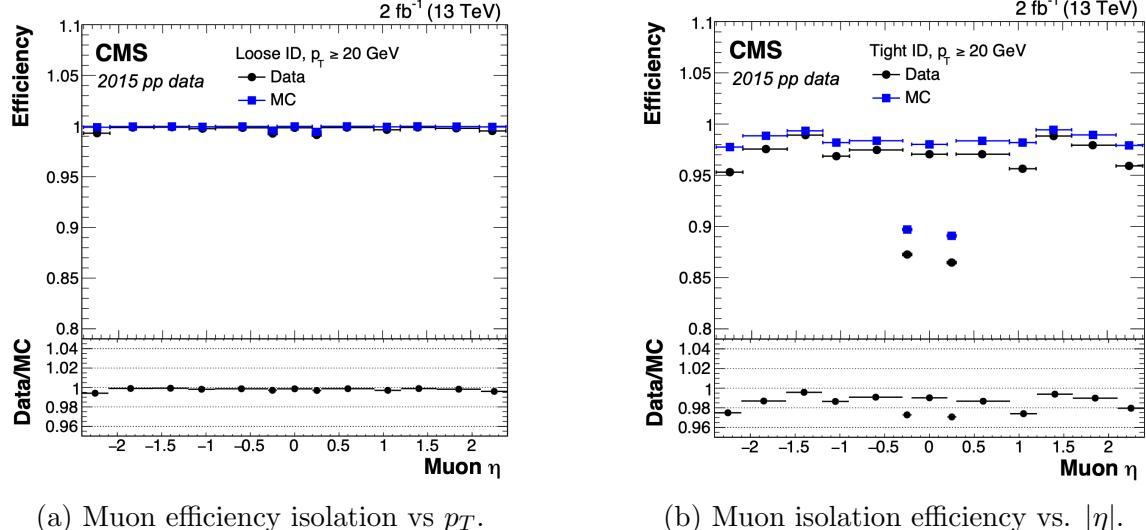
Figure 5.7: Efficiencies in data (*top panels*) and the ratio of efficiencies in data/MC (*bottom panels*), for the electron multivariate analysis (MVA) identification (*left*) and for the Gaussian-sum filter (GSF) tracking (*right*) [77]. Error bars represent statistical and systematic uncertainties.

1545 5.3.13 Muon ID, isolation, and tracking efficiencies

1546 Scale factors are applied to MC to correct for differences between MC and data in
 1547 the performance of muon identification, isolation, and tracking, as detailed below.

1548 The efficiencies for muon identification measured in 2015 data and MC simulation
 1549 are shown in Figures 5.8a and 5.8b for the loose ID and tight ID respectively [78].
 1550 The loose ID is chosen such that efficiency exceeds 99% over the full η range, and the
 1551 data and simulation agree to within 1%. The tight ID is chosen such that efficiency
 1552 varies between 95% and 99% as a function of η , and the data and simulation agree
 1553 to within 1-3%. The muon identification working point used in this analysis is the
 1554 medium ID, which has an efficiency of 98% for all η and an agreement within 1-2%
 1555 [78].

1556 The efficiencies in data for the muon isolation, as measured in Level-3 muons
 1557 (muons in one of the final stages of reconstruction in the HLT), as a function of the
 1558 muon p_T and $|\eta|$ are shown in Figures 5.9a and 5.9b [78]. The HLT muon reconstruc-
 1559 tion consists of two steps: Level-2 (L2), where the muon is reconstructed in the muon



(a) Muon efficiency isolation vs p_T .

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

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 [78].

1560 subdetectors only, and Level-3 (L3) which is a global fit of tracker and muon hits (i.e.
 1561 the global muon reconstruction as described in Section 5.1.2) [79].

1562 The muon tracking efficiencies as a function of $|\eta|$ for standalone muons (i.e. tracks
 1563 from only the muon system, i.e. DT, CSC, and RPC, as discussed in Section 5.1.2),
 1564 is shown for data and simulated Drell-Yan samples in Fig. 5.10 [80].

1565 5.3.14 Recoil corrections

1566 In proton-proton collisions, W and Z bosons are predominantly produced through
 1567 quark-antiquark annihilation. Higher-order processes can induce radiated quarks or
 1568 gluons that recoil against the boson, imparting a non-zero transverse momentum to
 1569 the boson [81]. Recoil corrections accounting for this effect are applied to samples
 1570 with W+jets, Z+jets, and Higgs bosons [66]. The corrections are performed on the
 1571 vectorial difference between the measured missing transverse momentum and the total
 1572 transverse momentum of neutrinos originating from the decay of the W, Z, or Higgs
 1573 boson. This vector is projected onto the axes parallel and orthogonal to the boson
 1574 p_T . This vector, and the resulting correction to use, is measured in $Z \rightarrow \mu\mu$ events,

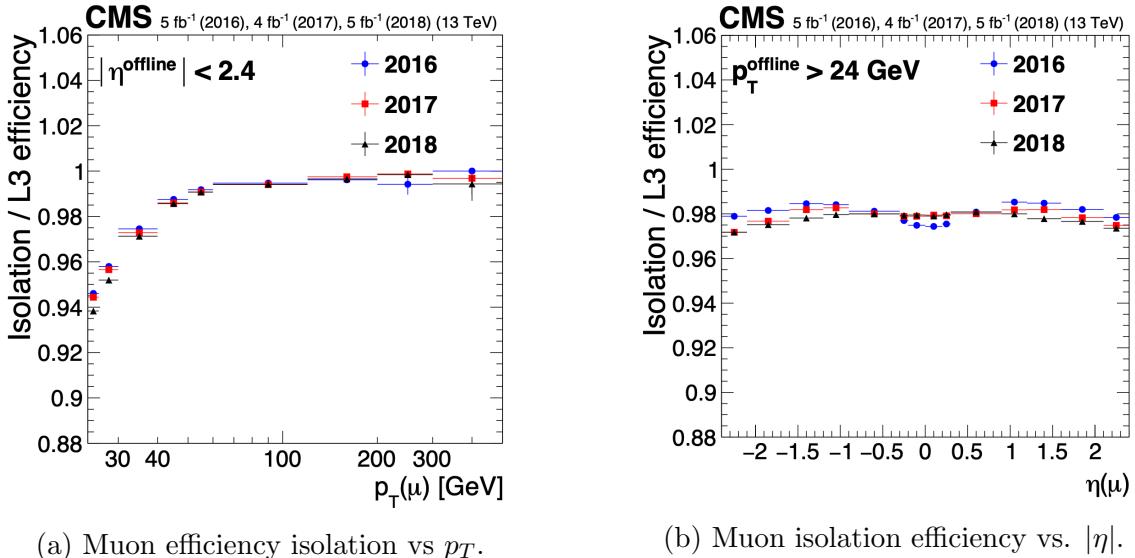


Figure 5.9: Muon isolation efficiencies in Run-2 data with respect to Level-3 muons (one of the final stages of HLT muon reconstruction) as a function of the muon p_T (*left*) and $|\eta|$ (*right*) [78].

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

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 [82]. Per-event weights derived by the 2016 data-only version of this analysis [82] 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 Pileup reweighting

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

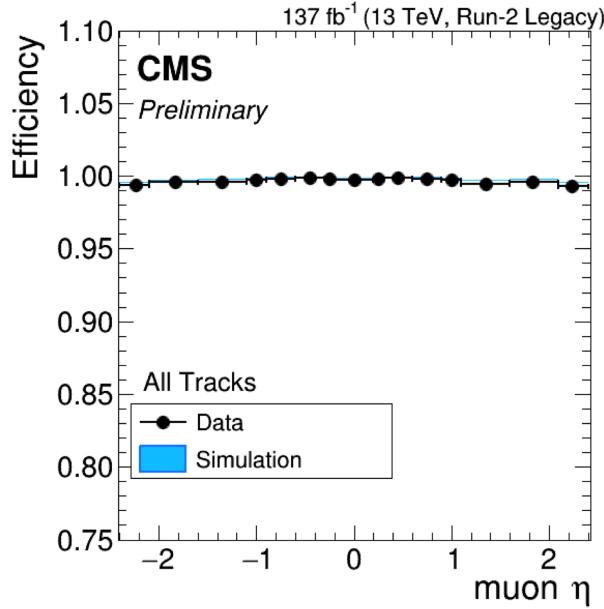


Figure 5.10: Muon tracking efficiencies as a function of $|\eta|$ for standalone muons in Run-2 data (*black*) and Drell-Yan MC simulation (*blue*) [80]. All Tracks refers to tracks which exploit the presence of muon candidates in the muon system to seed the track reconstruction in the inner tracker, in contrast to tracks that use tracker-only hits for seeding. Uncertainties shown are statistical.

1588 MC samples used is provided centrally by the Luminosity POG [83].

1589 5.3.17 Pre-firing corrections

1590 In 2016 and 2017 data-taking, a gradual timing shift of ECAL was not properly
1591 propagated to L1 trigger primitives (TPs), resulting in a large fraction of high η
1592 TPs being incorrectly associated with the previous bunch crossing. L1 trigger rules
1593 prevent two consecutive bunch crossings from firing, causing events to be rejected if
1594 significant ECAL energy was deposited in $2.0 < |\eta| < 3.0$. To account for this issue,
1595 MC simulations for 2016 and 2017 are corrected using an event-dependent weight.
1596 Embedded samples are not corrected [48].

1597 **5.3.18 Top p_T spectrum reweighting**

1598 In Run-1 and Run-2 it was observed that the p_T spectra of top quarks in $t\bar{t}$ data
1599 was significantly softer than those predicted by MC simulations [84]. Possible sources
1600 of this discrepancy are higher order QCD and/or electroweak corrections, and non-
1601 resonant production of $t\bar{t}$ -like final states. To account for this, corrections derived
1602 from Run-2 data by the Top Physics Analysis Group (PAG) are applied to the p_T
1603 of the top and anti-top quarks in MC simulations, computed as a function of their
1604 generator-level p_T [84].

1605 **5.3.19 B-tagging efficiency**

1606 In order to predict correct b-tagging discriminant distributions and event yields in
1607 data, the weight of selected MC events is reweighed according to recommendations by
1608 the BTV POG [85]. The reweighting depends on the jet p_T , η , and the b-tagging dis-
1609 criminant. In this method, there is no migration of events from one b-tag multiplicity
1610 bin to another.

1611 **5.3.20 Jet energy resolution and jet energy smearing**

1612 Calibration of jet energies, i.e. ensuring that the energy and momentum of the recon-
1613 structed jet matches that of the quark/gluon-initiated jet, is a challenging task due
1614 to time-dependent changes in the detector response and calibration and high pileup
1615 [86] [87]. Jet calibration is done via jet energy corrections (JECs) applied to the p_T
1616 of jets in MC samples, accounting successively for the effects of pileup, uniformity of
1617 the detector response, and residual data-simulation jet energy scale differences [88].
1618 Typical jet energy resolutions reported at $\sqrt{s} = 8$ TeV in the central rapidities are
1619 15-20% at 30 GeV and about 10% at 100 GeV [86]. Jet energy corrections are also
1620 propagated to the missing transverse energy.

1621 Measurements show that the jet energy resolution (JER) in data is worse than
1622 in simulation, and so the jets in MC need to be smeared to describe the data. JER
1623 corrections are applied after JEC on MC simulations, and adjust the width of the p_T
1624 distribution based on pileup, jet size, and jet flavor [89]. Tools for applying JEC and
1625 JER are provided centrally by the JER Corrections group.

1626 Chapter 6

1627 Event selection

1628 6.1 General procedure for all channels

1629 For the search for $h \rightarrow aa \rightarrow bb\tau\tau$, three final states of the $\tau\tau$ system are considered:
1630 $\mu\tau_h$, $e\tau_h$, and $e\mu$. The $\tau_h\tau_h$ final state is not considered because signal events in the
1631 $\tau_h\tau_h$ channel would typically produce hadronic taus with momenta below data-taking
1632 trigger thresholds.

1633 In all three final states, events are required to have at least one b-tag jet passing the
1634 medium working point of the DeepFlavour tagger, with $p_T > 20$ GeV, and $|\eta| < 2.4$.
1635 A second b-tag jet is not required because such a requirement would reduce signal
1636 acceptance by 80% compared to only requiring one b-tag jet.

1637 Events in MC samples are sorted into one of the three $\tau\tau$ channels if they pass the
1638 following trigger requirements and requirements on the offline reconstructed objects
1639 in the event, first checking the HLT paths for the $\mu\tau_h$ channel, then $e\tau_h$, and finally
1640 $e\mu$. The two leading leptons (e.g. muon and hadronic tau for the $\mu\tau_h$ channel) that
1641 were determined to have originated from the $\tau\tau$ decay, are called the $\tau\tau$ “legs” and
1642 are respectively subscripted 1 and 2 in this work. For events in data and embedded
1643 samples, the HLT paths requirements for the corresponding channel are checked.

1644 After sorting events by HLT paths and identifying the leading tau legs in the offline
 1645 reconstructed objects, the p_T of the offline objects is checked against the online trigger
 1646 thresholds. Trigger matching is also performed, which checks the correspondence
 1647 between each offline reconstructed object used in the analysis (e.g. a muon), and a
 1648 trigger object in the HLT (e.g. a HLT muon). An offline object is considered to be
 1649 matched, if it corresponds to a trigger object of the same object type, with $\Delta R < 0.5$.
 1650 This matched trigger object is also required to pass the filter(s) of the HLT trigger.
 1651 The trigger thresholds used for the $bb\tau\tau$ final state (the focus of this work) and the
 1652 $bb\mu\mu$ final state are summarized in Table 6.1 and detailed in the following sections.

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

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

1653 Further cuts are made on the offline objects in each channel to obtain the signal
 1654 region, or other data regions used to perform data-driven background estimations.

1655 6.2 Event selection in the $\mu\tau_h$ channel

1656 In all three years, a single muon trigger is used if the muon has sufficiently high p_T ,
 1657 otherwise a dilepton $\mu\tau_h$ cross-trigger is used (Tables 6.2, 6.3, and 6.4). For data

1658 taken in 2017-2018 (2016), the logical OR of the single muon triggers with online p_T
1659 thresholds 24 and 27 (23) GeV is used, with the corresponding offline muon required
1660 to have with p_T 1 GeV above the online threshold. For data taken in 2017-2018
1661 (2016), a dilepton $\mu + \tau_h$ cross-trigger with p_T thresholds of 20 (19) and 27 (20) GeV
1662 for the muon and tau respectively, is used. The τ_h is required to have $|\eta| < 2.3$ if the
1663 single trigger is fired, $|\eta| < 2.1$.

1664 The muon and τ_h are required to have opposite charge and be separated by $\Delta R >$
1665 0.4. The muon is required to have $|\eta| < 2.4$, and the τ_h is required to have $|\eta| < 2.3$
1666 unless a cross-trigger is required, in which case we require $|\eta| < 2.1$ as discussed
1667 above.

1668 The muon is required to pass the medium identification (ID) working point [90],
1669 which is defined by the Muon POG as a loose muon (i.e. a Particle Flow muon that is
1670 either a global or a tracker muon - see Section 5.1.2) with additional requirements on
1671 track quality and muon quality. This identification criteria is designed to be highly
1672 efficiently for prompt muons and for muons from heavy quark decays. In addition to
1673 the ID, for prompt muons it is recommended to apply cuts on the impact parameter
1674 [90]: we apply $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$.

1675 In addition, a cut is applied on the muon relative isolation (defined in Section
1676 5.1.2), to be less than 0.15 in a cone size of $\Delta R = 0.4$, which corresponds to the
1677 Tight Particle Flow isolation requirement [90].

1678 The τ_h is required to pass a cut on its impact parameter of $|\Delta(z)| < 0.2$. The τ_h
1679 is also required to pass the VLoose (Very Loose) DeepTau working point vs. elec-
1680 tron, the Tight DeepTau working point vs. muons, and the VVVLoose and Medium
1681 DeepTau working point vs. jets. Events with taus reconstructed in two of the decay
1682 modes (labeled 5 and 6) are rejected, since these decay modes are meant to recover
1683 3-prong taus, but are only recommended for use in analyses where the benefits in
1684 final significance outweigh the resulting increase in background [63].

1685 For the estimation of the background from jets faking τ_h , which is described in Sec-
1686 tion 7.7, anti-isolated events are selected, by requiring events to pass all the selections
1687 described above, except failing the Medium DeepTau working point vs. jets.

1688 6.3 Event selection in the $e\tau_h$ channel

1689 The HLT trigger paths for the $e\tau_h$ channel are summarized in Tables 6.2, 6.3, and
1690 6.4. Similarly to the $\mu\tau_h$ channel, a single electron trigger is used if the electron has
1691 sufficiently high p_T in 2018 and 2017. For data taken in 2018 (2017), the OR of the
1692 single electron triggers with online p_T thresholds at 32 and 35 (27 and 32) GeV are
1693 used, with the corresponding offline electrons required to have p_T greater than 33
1694 (28) GeV. A $e + \tau_h$ cross-trigger is used for electrons with lower offline p_T between
1695 25 and 33 GeV (25 and 28 GeV). For the 2016 dataset, there is no cross trigger but
1696 only a single electron trigger with online p_T threshold at 25 GeV, which is used if the
1697 offline electron has p_T greater than 26 GeV.

1698 The electron and τ_h are required to have opposite charge and be separated by
1699 $\Delta R > 0.4$. The electron is required to be within $|\eta| < 2.3$ when no cross trigger is
1700 used, and $|\eta| < 2.1$ when the cross trigger is fired. The τ_h is required to have $|\eta| < 2.3$
1701 if no cross trigger is fired, and have $|\eta| < 2.1$ if the cross trigger is fired.

1702 The electron is required to have a relative isolation (same definition as in Section
1703 5.1.2) of less than 0.1 in a cone size of $\Delta R = 0.3$, which is the standard recommended
1704 cone size giving minimal pileup dependence and reduced probability of other objects
1705 overlapping with the cone. The isolation quantity used includes an “effective area”
1706 (EA) correction to remove the effect of pileup in the barrel and endcap parts of the
1707 detector [91].

1708 The electron is also required to pass cuts on its impact parameter of $|\Delta(z)| < 0.2$
1709 and $|\Delta(xy)| < 0.045$. It is also required to pass the non-isolated MVA working point

1710 corresponding to 90% efficiency. The electron's number of missing hits, which are
1711 gaps in its trajectory through the inner tracker [91], must be less than or equal to
1712 1. The electron must pass a conversion veto, which rejects electrons coming from
1713 photon conversions in the tracker, which should instead be reconstructed as part of
1714 the photon [91].

1715 The impact parameter cut for the τ_h is $|\Delta(z)| < 0.2$. In contrast to the $\mu\tau_h$ event
1716 selection, the vs. electron and vs. muon DeepTau working points are flipped, to
1717 reject muons faking the τ_h leg. The τ_h is required to pass the Tight DeepTau working
1718 point vs. electrons, the VLoose DeepTau working point vs. muons, and the Medium
1719 DeepTau working point vs. jets.

1720 As in the $\mu\tau_h$ channel, for the estimation of the background from jets faking τ_h ,
1721 which is described in Section 7.7, anti-isolated events are selected, by requiring events
1722 to pass all the selections described above, except failing the Medium DeepTau working
1723 point vs. jets.

1724 6.4 Event selection in the $e\mu$ channel

1725 The HLT trigger paths for the $e\mu$ channel are summarized in Tables 6.2, 6.3, and
1726 6.4. Events are selected with the logical OR of two $e + \mu$ cross triggers, where either
1727 the electron or muon can have larger p_T : (1) leading electron, where the electron has
1728 online $p_T > 23$ GeV and muon has online $p_T > 8$ GeV, or (2) leading muon, where
1729 electron has online $p_T > 12$ GeV and muon has online $p_T > 23$ GeV.

1730 The leading and sub-leading leptons are required to have an offline p_T greater
1731 than 1 GeV above the online threshold (i.e. $p_T > 24$ GeV). If the sub-leading lepton
1732 is the electron, the offline p_T threshold is 1 GeV above the online threshold ($p_T > 13$
1733 GeV), but if it is a muon, the offline p_T threshold is required to be at least 5 GeV
1734 greater than the online threshold (i.e. $p_T > 13$ GeV). This is because of poor data

1735 and simulation agreement for low- p_T muons with p_T between 9 GeV and 13 GeV, and
1736 the higher probability of mis-identifying jets as muons at lower p_T . With no effect on
1737 the expected limits, the offline p_T threshold for muons is raised to 13 GeV instead of
1738 9 GeV, even though it may lead to loss in signal acceptance. Both the electron and
1739 muon are required to have $|\eta| < 2.4$.

1740 The electron and muon are required to have opposite charge and be separated
1741 by $\Delta R > 0.3$ (note the decreased separation requirement compared to the other
1742 two channels). The electron is required to pass the non-isolated MVA identification
1743 working point corresponding to 90% efficiency, and to have a relative isolation less
1744 than 0.1 for a cone size of $\Delta R = 0.3$ with the EA pileup subtraction correction.
1745 The electron must have one or fewer missing hits and pass the conversion veto (both
1746 described previously in Section 6.3).

1747 The muon is required to pass the medium identification working point (described
1748 earlier in 6.2), and to have a relative isolation less than 0.15 for a cone size of $\Delta R =$
1749 0.4. The muon impact parameter is required to have $|\Delta(z)| > 0.2$ and $|\Delta(xy)| < 0.045$.

1750 For the QCD multijet background estimation described in Section 7.8, the same-
1751 sign region is selected by requiring all the above selections, except the legs are required
1752 to have the same electric charge rather than opposite.

1753 6.5 Extra lepton vetoes in all channels

1754 Events containing a third lepton (electron or muon) that is neither of the leading $\tau\tau$
1755 legs are rejected, and events with di-muons and di-electrons are vetoed, with criteria
1756 taken from the Standard Model $H \rightarrow \tau\tau$ working group [66].

1757 The event is vetoed if a third electron is found with the following properties:
1758 $p_T > 10$ GeV, $|\eta| < 2.5$, impact parameter $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$, passing
1759 non-isolation MVA identification with 90% efficiency, conversion veto, ≤ 1 missing

2016 $\mu\tau_h$ trigger paths	
Notes	HLT Path
	HLT_IsoMu22_v
	HLT_IsoMu22_eta2p1_v
	HLT_IsoTkMu22_v
	HLT_IsoTkMu22_eta2p1_v
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_v
	HLT_IsoMu19_eta2p1_LooseIsoPFTau20_SingleL1_v

2016 $e\tau_h$ trigger paths	
Notes	HLT Path
	HLT_Ele25_eta2p1_WPTight_Gsf_v

2016 $e\mu$ trigger paths	
Notes	HLT Path
runs B-F and MC	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v
runs B-F and MC	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v
runs G-H	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
runs G-H	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v

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

2017 $\mu\tau_h$ trigger paths	
Notes	HLT Path
	HLT_IsoMu24_v
	HLT_IsoMu27_v
	HLT_IsoMu20_eta2p1_LooseChargedIso_PFTau27_eta2p1_CrossL1_v

2017 $e\tau_h$ trigger paths	
Notes	HLT Path
	HLT_Ele32_WPTight_Gsf_v
	HLT_Ele35_WPTight_Gsf_v
	HLT_Ele24_eta2p1_WPTight_Gsf_Loose_ChargedIsoPFTau30_eta2p1_CrossL1_v

2017 $e\mu$ trigger paths	
Notes	HLT Path
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v

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

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

Table 6.4: High-Level Trigger (HLT) paths used to select data and simulation events in 2018 for the three $\tau\tau$ channels. In 2018 a HLT trigger path using the hadron plus strips (HPS) tau reconstruction algorithm became available.

1760 hits, and relative isolation < 0.3 with cone size $\Delta R = 0.3$. The event is also vetoed if
1761 a third muon is found with the following properties: $p_T > 10$ GeV, $|\eta| < 2.4$, impact
1762 parameter $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$, medium ID, and isolation < 0.3 with
1763 cone size $\Delta R = 0.4$.

1764 A di-muon veto is applied, which rejects events containing a pair of muons with
1765 opposite charge and separation of $\Delta R > 0.15$, that both pass the following selections:
1766 $p_T > 15$ GeV, $|\eta| < 2.4$, flag for global muons, flag for tracker muon, flag for Particle
1767 Flow muon, $|\Delta(z)| < 0.2$, $|\Delta(xy)| < 0.045$, and isolation < 0.3 with cone size $\Delta R =$
1768 0.4.

1769 A similar di-electron veto is applied to reject events containing a pair of electrons
1770 with opposite charge and separation of $\Delta R > 0.15$, that both pass the following
1771 selections: $p_T > 15$ GeV, $|\eta| < 2.5$, a dedicated electron ID (cut-based) for vetoing
1772 third leptons, $|\Delta(z)| < 0.2$, $|\Delta(xy)| < 0.045$, with pileup-corrected relative isolation
1773 < 0.3 with cone size $\Delta R = 0.3$.

1774 These vetoes on extra leptons also ensure orthogonality of events to analyses such
1775 as the $bb\mu\mu$ final state, whose results are combined with this $bb\tau\tau$ final state as
1776 described in Section ??.

₁₇₇₇ **Chapter 7**

₁₇₇₈ **Background estimation**

₁₇₇₉ This section describes methods used to estimate sources of background from Standard
₁₇₈₀ Model processes in the search for $h \rightarrow aa \rightarrow bb\tau\tau$. Similar background estimation
₁₇₈₁ methods are being used for the $h \rightarrow a_1a_2$ analysis. The background contributions
₁₇₈₂ directly taken from MC are described first, followed by backgrounds estimated from
₁₇₈₃ data-driven methods to produce sufficient statistics in the signal region.

₁₇₈₄ **7.1 Z+jets**

₁₇₈₅ A major source of background for $\tau\tau$ analyses is the Drell-Yan (DY) process (Z+jets).
₁₇₈₆ The Z boson decays to $\tau\tau/\mu\mu/ee$ with equal probability of 3.4% each, with the dom-
₁₇₈₇ inant decay modes being to hadrons (around 70%) and neutrinos (invisible) (20%)
₁₇₈₈ [26].

₁₇₈₉ The Drell-Yan contribution with genuine taus, $Z \rightarrow \tau\tau$, is estimated using embed-
₁₇₉₀ ded samples, described in Section 4.3. To avoid double-counting between embedded
₁₇₉₁ and MC samples, in all MC samples, events with legs that originated from genuine τ
₁₇₉₂ are discarded.

₁₇₉₃ The other decays of the Z, $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$, are estimated from MC simulation,
₁₇₉₄ and are hereafter referred to as simply the Drell-Yan background. These MC samples

1795 are generated to leading order (LO) with different numbers of jets (jet multiplicity) in
1796 the matrix element: Z+1 jet, Z+2jets, Z+3 jets, Z+4 jets, and inclusive Z+jets. The
1797 cross-sections of the samples with ≥ 1 jets are normalized to next-to-NLO (NNLO)
1798 in QCD.

1799 For the inclusive Drell-Yan sample, two samples are used with different thresholds
1800 for the di-lepton invariant mass ($m_{\ell\ell}$) at the generator level: one with $m_{\ell\ell} > 50$ GeV
1801 and the other with $10 < m_{\ell\ell} < 50$.

1802 7.2 W+jets

1803 The dominant W boson decay modes are to hadrons (67.4%), $e + \nu_e$ (10.7%), $\mu + \nu_\mu$
1804 (10.6%), and $\tau + \nu_\tau$ (11.4%) [26]. The W+jets background is estimated from MC
1805 simulation. Similarly to the Z+jets, the W+jets samples are generated with different
1806 jet multiplicities in the matrix element. LO samples are used for greater statistics
1807 and are normalized to NNLO cross sections.

1808 7.3 $t\bar{t}$ + jets

1809 In hadron collisions, top quarks are produced singly with the weak interaction, or in
1810 pairs via the strong interaction, with interference between these leading-order pro-
1811 cesses possible in higher orders of the perturbation theory. The top quark is the
1812 heaviest fermion in the Standard Model and has a short lifetime ($\sim 10^{-25}$ s), decay-
1813 ing without hadronization into a bottom quark and a W boson [26], with the decay
1814 modes of the W boson as listed in the previous section. With two top quarks, the
1815 final states of the two resulting W bosons can be described as fully leptonic, semilep-
1816 tonic, and fully hadronic. These three final states are modeled separately with MC
1817 simulation in 2018 and 2017, while for 2016 the sample used is inclusive.

¹⁸¹⁸ 7.4 Single top

¹⁸¹⁹ There are three main production modes of the single top in pp collisions [92]: the
¹⁸²⁰ exchange of a virtual W boson (t channel), the production and decay of a virtual W
¹⁸²¹ boson (s channel), and the associated production of a top quark and W boson (tW ,
¹⁸²² or W-associated) channel. As the s channel process is rare and only 3% of the total
¹⁸²³ production, the dominant production mode of the t -channel and the tW production
¹⁸²⁴ are considered and modeled with MC.

¹⁸²⁵ 7.5 Diboson

¹⁸²⁶ In pp collisions, the production of dibosons (pairs of electroweak gauge bosons, i.e.
¹⁸²⁷ WW, WZ, and ZZ) is dominated by quark-antiquark annihilation, with a small con-
¹⁸²⁸ tribution from gluon-gluon interaction [93]. MC is used to model the pair production
¹⁸²⁹ 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
¹⁸³⁰ quarks and ℓ being leptons).

¹⁸³¹ 7.6 Standard Model Higgs

¹⁸³² MC is used to simulate backgrounds from major production modes of the Standard
¹⁸³³ Model 125 GeV Higgs boson: gluon-gluon fusion (ggH), vector boson fusion (VBF),
¹⁸³⁴ associated production with a W or Z (WH, ZH), and associated production with a
¹⁸³⁵ top pair (ttH) (see Fig. 7.1 for leading-order diagrams). For these production modes,
¹⁸³⁶ samples with the Higgs decaying to $\tau\tau$ or to WW are used. Samples made with
¹⁸³⁷ higher-order diagrams for WH and ZH that include the production of a jet, with the
¹⁸³⁸ Higgs decaying to WW, are also used.

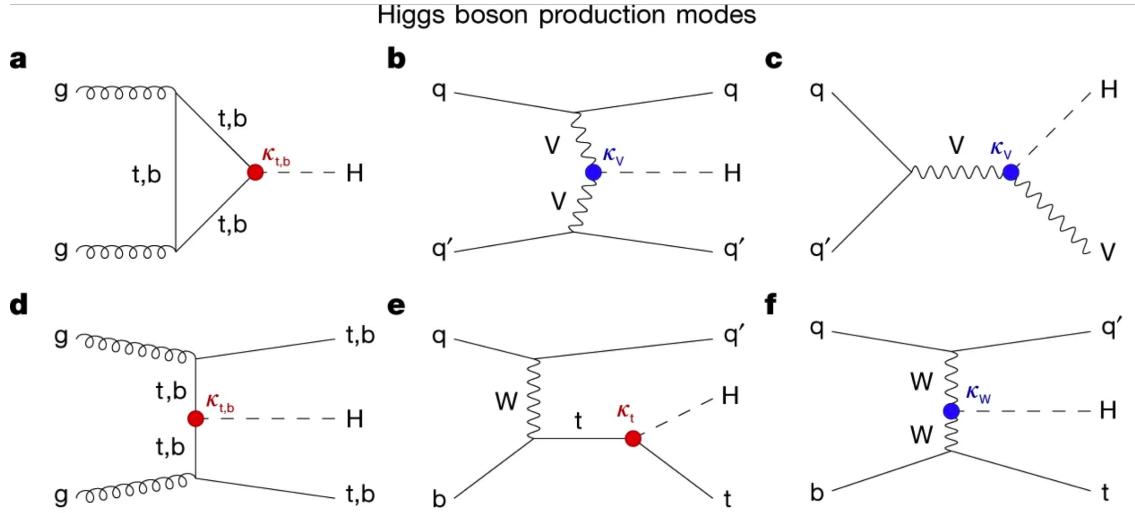


Figure 7.1: Leading-order Feynman diagrams of Higgs production from [94], in ggH (a) and vector boson fusion (VBF; b), associated production with a W or Z (V) boson (VH; c), associated production with a top or bottom quark pair (ttH or bbH); d, and associated production with a single top quark (tH; e, f).

1839 7.7 Jet faking τ_h

1840 Events with a jet mis-reconstructed as the hadronic tau leg τ_h are a major source of
 1841 background in the $\mu\tau_h$ and $e\tau_h$ channels. The main processes contributing to jet $\rightarrow \tau_h$
 1842 events are QCD multijet, W+jets, and $t\bar{t}$ production. These events are estimated
 1843 using a data-driven method adapted from past analyses [48] [82]. This background
 1844 includes contributions from W+jets, QCD multijets, and $t\bar{t}$ +jets. To estimate this
 1845 background, a sideband region is constructed, where events are required to pass all
 1846 baseline $\mu\tau_h/e\tau_h$ selection criteria, but fail the τ_h isolation criteria. The events in
 1847 this sideband region are reweighed with a factor $f/(1-f)$, where f is the probability
 1848 for a jet to be misidentified as a τ_h . The jet $\rightarrow \tau_h$ background is the anti-isolated,
 1849 reweighed MC and embedded events subtracted from the anti-isolated, reweighted
 1850 data events.

1851 The fake factor is measured in $Z \rightarrow \mu\mu + \text{jets}$ events in data in the $\mu\mu\tau_h$ final
 1852 state, as any reconstructed τ_h in these events must originate from a jet. The two
 1853 muons are required to be isolated (< 0.15), have opposite electric charge, and have

1854 an invariant mass between 76 and 106 GeV (close to the Z mass). These events are
1855 selected with a double muon trigger, with the leading muon having offline $p_T > 20$
1856 GeV and the subleading muon $p_T > 10$ GeV. Simulated diboson (ZZ and WZ) events
1857 are subtracted to avoid contamination from events with real τ_h . The denominator of
1858 the fake rate corresponds to fake taus passing the VVVLoose working point of the
1859 discriminator vs. jets, while the numerator corresponds to those passing the Medium
1860 working point, i.e. $f = N_{\text{jet passing tight}} / N_{\text{jet passing loose}}$.

1861 f is measured as a function of the τ_h transverse momentum and is 8% - 10% in
1862 each of the data-taking years. f is derived separately for the $\mu\tau_h$ and $e\tau_h$ channels
1863 because the channels use different anti-lepton identification working points.

1864 7.8 QCD multijet background

1865 In the $e\mu$ channel, events with jets faking electrons or muons originating from QCD
1866 multijet, is estimated from data events with the same baseline selection as in the
1867 signal region, except with same-signed (SS) charged $e + \mu$, ensuring orthogonality
1868 with the signal region which requires opposite-sign (OS) $e\mu$ pairs. All same-sign MC
1869 events (both events with real and fake $e + \mu$) are subtracted from same-sign data
1870 events to remove contamination from other backgrounds. i.e. $\text{QCD}_{\text{SS}} = \text{Data}_{\text{SS}} -$
1871 MC_{SS} .

1872 Three scale factors are applied to the QCD_{SS} events to compute the QCD multijet
1873 background [82] [42]:

- 1874 • *OS-to-SS scale factor*: This scales the SS QCD to the OS region, and is mea-
1875 sured from an orthogonal region with an isolated electron and an anti-isolated
1876 muon. Only the muon is chosen to be anti-isolated because this scale factor was
1877 observed to depend more strongly on electron isolation than that of the muon.
1878 This scale factor is treated as a function of the ΔR separation of the trajectories

1879 of the electron and muon, and is measured separately for events with 0 jets, 1,
1880 jet, and greater than 1 jet.

- 1881 • *2D closure correction for the lepton p_T :* This factor accounts for subleading
1882 dependencies of the first scale factor on the p_T of the two leptons. A 2D weight
1883 is derived in a similar fashion, as a ratio of QCD_{OS} events to QCD_{SS} events,
1884 but parameterized by both electron and muon p_T , where the SS events have the
1885 previous scale factor applied.
- 1886 • *Isolation correction for the muon:* The third and final factor is an isolation
1887 correction, which is a bias correction to account for the fact that the fake
1888 factor was determined for less-isolated muons. This factor is obtained as the
1889 ratio of the OS-to-SS scale factors measured in two other control regions: (1)
1890 events where the electron is anti-isolated ($0.15 < \text{iso} < 0.5$) and the muon is
1891 isolated, and (2) events where both leptons are anti-isolated.

¹⁸⁹² Chapter 8

¹⁸⁹³ Systematic uncertainties

¹⁸⁹⁴ The handling of systematic uncertainties is separated into normalization uncertainties
¹⁸⁹⁵ (those that affect the total yield of a variables' distribution) and shape uncertainties
¹⁸⁹⁶ (those that shift the distribution of events). Normalization uncertainties are expressed
¹⁸⁹⁷ as multiplicative factors, while shape uncertainties are represented as up and down
¹⁸⁹⁸ shifts of a variable's distribution.

¹⁸⁹⁹ Up/down shifts of shape uncertainties can change the number of background
¹⁹⁰⁰ events in a distribution. For instance, hadronic taus receive corrections from the
¹⁹⁰¹ nominal tau energy scale, with the nominal, up, and down energy scales provided
¹⁹⁰² centrally by CMS. For the $\mu\tau_h$ channel, an event could have a τ_h with p_T just below
¹⁹⁰³ the offline threshold of 20 GeV (for instance, 19.5 GeV), so in the nominal distribution
¹⁹⁰⁴ of $m_{\tau\tau}$ (or any other variable for this channel), the event is excluded. However, when
¹⁹⁰⁵ we build our distributions with the tau energy scale “up” shift, the energy of this τ_h
¹⁹⁰⁶ may be scaled up to, say, 20.5 GeV, and now the event passes the offline p_T threshold
¹⁹⁰⁷ for the single muon trigger, leading to the event's inclusion in the distributions made
¹⁹⁰⁸ with the tau energy scale “up” shift.

¹⁹⁰⁹ In evaluating the up and down shifts of a specific source of uncertainty, all other
¹⁹¹⁰ corrections and scale factors are held at their nominal values, and the full chain

1911 of object and event selection and event categorization is performed to obtain the
1912 observable distributions. Any “downstream” variables that depend on the shifted
1913 variable, e.g. the invariant di-tau mass $m_{\tau\tau}$, must be computed for the nominal case,
1914 and then re-computed separately for each up and down shift of the tau legs’ energy
1915 scale. The objective of this process is to quantify the effect of a single source of
1916 uncertainty on the resulting observable distributions.

1917 8.1 Uncertainties associated with physics objects

1918 Each scale factor and correction described in Section 5.3 has an associated uncertainty.
1919 The binning of the uncertainties follows that of the nominal scale factor value.

1920 8.1.1 Uncertainties in the lepton energy scales

1921 The uncertainties in the tau energy scales [63] are binned by the tau decay mode and
1922 are taken as shape uncertainties treated as uncorrelated across the tau decay modes
1923 and years. Same as with the application of the nominal scale factor, when applying
1924 the up or down shifts, the missing transverse energy (p_T^{miss}) of the event is adjusted
1925 so that the 4-vector sum of the tau p_T^{miss} is unchanged.

1926 The uncertainties in the muon energy scale [64] are 0.4% for $|\eta| < 1.2$, 0.9% for
1927 $1.2 < |\eta| < 2.1$, and 2.7% for $2.1 < |\eta| < 2.4$, and are treated as shape uncertainties,
1928 fully uncorrelated between embedded and MC samples.

1929 The uncertainties in the electron energy scale [67] in MC are binned in the electron
1930 $|\eta|$ and p_T , and are shown in Fig. 5.2. The uncertainties range from 0.5% to 2.2% in
1931 the barrel, and 0.3% to 4.1% in the endcap, across the p_T range. The uncertainties
1932 for the embedded sample are binned only in $|\eta|$ and are on the order of 0.5% and
1933 1.25% for the barrel and endcap [71].

1934 There are also uncertainties in the energy scales for electrons and muons misiden-

1935 tified as τ_h . The uncertainty for muons misidentified as τ_h is 1% [63]. For electrons
1936 misidentified as τ_h , the uncertainty is binned in barrel/endcap η and by 1-prong and
1937 1-prong + π_0 decays. The probability for e/μ faking a 3-prong decay mode is much
1938 lower.

1939 8.1.2 Uncertainties from other lepton corrections

1940 Uncertainties associated with the τ_h identification efficiencies are treated as shapes,
1941 uncorrelated across the seven p_T bins and years. The shape uncertainties in the
1942 embedded samples are taken as 50% correlated with those of the MC samples.

1943 The uncertainties on electron and muon identification efficiencies are taken as
1944 normalization uncertainties of 2% each, with a 50% correlation between embedded
1945 and MC samples.

1946 In the $e\tau_h$ channel, there is an additional uncertainty for the vs. jet discrimination
1947 efficiency [63], because the analysis uses a looser anti-lepton working point (VLoose
1948 WP) than the working points used in the measurement of the efficiency (namely,
1949 VLoose WP vs e, and Tight WP vs mu). For nominal $\tau_h p_T < 100$ GeV, an additional
1950 uncertainty of 3% (5%) is used in MC (embedded), and for high p_T an uncertainty of
1951 15% is used for both.

1952 The uncertainties in trigger efficiencies are taken as shapes [63]. In the $e\tau_h$ and $\mu\tau_h$
1953 channels, there are uncertainties for the single and cross lepton triggers, and in the
1954 $e\mu$ channel there is one uncertainty each for the two $e + \mu$ triggers, and one combined
1955 uncertainty since their trigger phase spaces are not mutually exclusive.

1956 8.1.3 Uncertainties from jet energy scale and resolution

1957 The jet energy scale uncertainties are taken as shape uncertainties: there are eleven
1958 in total, with seven correlated across years (labeled “Year” below) and the remainder
1959 uncorrelated across years. They affect the b-tag jet p_T and mass, and hence the

1960 missing transverse energy p_T^{miss} . The shifts are propagated through the b-tagging
1961 scale factor calculation and b-tag jet counting.

1962 The uncertainties in the jet energy correction and resolution [86] [95] are as follows:

1963 • *Absolute, AbsoluteYear*: flat absolute scale uncertainties.

1964 • *BBEC1, BBEC1Year*: for sub-detector regions, with barrel “BB” in $|\eta| < 1.3$
1965 and endcap region 1 “EC1”: $1.3 < |\eta| < 2.5$.

1966 • *EC2, EC2 year*: for sub-detector regions, with endcap region 2 “EC2” in $2.5 <$
1967 $|\eta| < 3.0$.

1968 • *HF, HF year*: for sub-detector regions, with hadron forward “HF” in $|\eta| > 3$.

1969 • *FlavorQCD*: for uncertainty in jet flavor (uds/c/b-quark and gluon) estimates
1970 based on comparing Pythia and Herwig (different MC generator) predictions.

1971 • *RelativeBal*: account for difference between log-linear fits of the two methods
1972 used to study the jet energy response: MPF (missing transverse momentum
1973 projection fraction) and p_T balance.

1974 • *RelativeSample*: account for η -dependent uncertainty due to a difference be-
1975 tween relative residuals, observed with dijet and Z+jets in Run D of 2018 data.

1976 • *JetResolution*: uncertainty in the jet energy resolution.

1977 8.1.4 Uncertainties from b-tagging scale factors

1978 The b-tagging scale factor has its own set of associated uncertainties (not to be
1979 confused with shifts in the b-tagging scale factor due to the propagation of the jet
1980 energy scale uncertainties described in the previous section 8.1.3). They are:

1981 • *hf*: contamination from heavy flavor (b+c) jets in the light flavor region.

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

8.1.5 Uncertainties from MET

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

8.2 Uncertainties associated with samples used

Normalization uncertainties related to the samples used are:

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

2003 The $t\bar{t}$ process has additional acceptance uncertainties from QCD scale variation
2004 and parton shower uncertainties [96]. Parton shower uncertainties originate from
2005 the modeling of perturbative and non-perturbative QCD effects handled in parton
2006 shower MC generators. The scale variations are determined from the envelope of the
2007 6 provided shapes due to variations in the factorization scale, renormalization scale,
2008 and their combined variation [96].

2009 The Z p_T reweighing uncertainty in Drell-Yan samples is taken to be 10% of the
2010 nominal value, taken as a shape uncertainty.

2011 The fake rate uncertainties are taken as shape uncertainties. For the weight ap-
2012 plied to scale up anti-isolated events in cross-trigger regions, 20% of the nominal
2013 weight is taken as a shape uncertainty.

2014 **8.3 Other uncertainties**

2015 A 3.6% yield uncertainty in the signal is used to cover uncertainties in the parton
2016 distribution functions, α_s (fine structure constant), and QCD scale.

2017 Normalization uncertainties from luminosity are applied to all MC samples, di-
2018 vided into those uncorrelated across years, those correlated between 2017 and 2018,
2019 and one for 2018 [83].

2020 **8.4 Pulls and impacts**

2021 The top impacts and pulls computed for the combination of all channels and years is
2022 shown in Fig. 8.1. The top impacts are related to uncertainty in the signal sample and
2023 cross-section of the $t\bar{t}$ cross-section, and also the yields of the jet faking τ_h background,
2024 which is a major background in all channels and expected to be constrained due to
2025 the yield uncertainty which is taken to be 20% pre-fit.

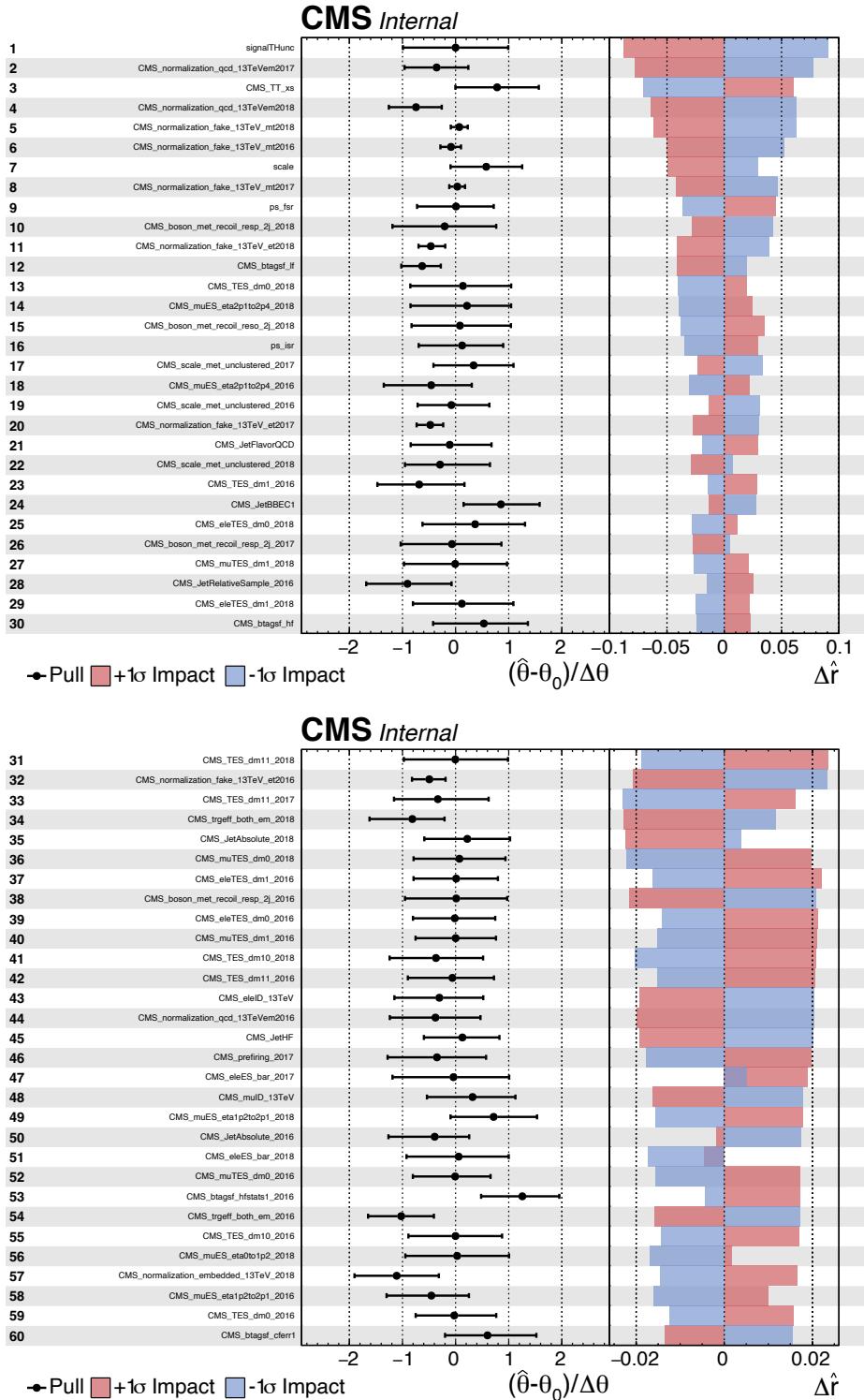


Figure 8.1: Top sixty impacts for the combination of all channels and years [41].

2026 Chapter 9

2027 Event categorization and signal 2028 extraction

2029 9.1 B-tag jet multiplicity

2030 The increased statistics of the full Run-2 dataset enables the separation of events into
2031 events with exactly 1 b-tag jet and events with greater than 1 b-tag jet. Further event
2032 categorization is performed with deep neural networks (DNNs) described below. The
2033 DNNs are used only for separating events into signal and control regions in the 1
2034 b-tag and 2 b-tag jets scenarios. The final results are extracted from the statistical
2035 fitting to the mass of the $\tau\tau$, $m_{\tau\tau}$.

2036 9.2 DNN-based event categorization

2037 A brief overview of the DNN-based event categorization is given below with a focus
2038 on the physics aspects, with full details of the machine learning training in [42] and
2039 associated documentation.

2040 **Training samples**

2041 Neural networks for event categorization are trained for each of the $\mu\tau_h$, $e\tau_h$, and $e\mu$
2042 channels, for 1 and 2 b-tag jets, giving $3 \times 2 = 6$ networks in total. In the training,
2043 the signal is taken to be all of the possible pseudoscalar mass m_a hypotheses together.
2044 The backgrounds for each DNN are taken to be a representative combination of the
2045 three major backgrounds: $Z \rightarrow \tau\tau$, $t\bar{t}$ +jets, and fake backgrounds. The proportions of
2046 each background for each channel and b-tag jet multiplicity are taken from the yields
2047 in the $m_{\tau\tau}$ distribution. For instance, in the $\mu\tau_h$ 1 b-tag jet category, the composition
2048 of the background for training is 17.4% from $Z \rightarrow \tau\tau$, 42.4% from $t\bar{t}$ +jets, and 40.2%
2049 fakes.

2050 **Input variables**

2051 The input variables capture the key differences between the signal and the back-
2052 ground:

- 2053 • Transverse momentum p_T of the electron and muon in the $e\tau_h$ and $\mu\tau_h$ channels,
2054 where the signal tends to have a softer p_T spectrum (lower energy) than the
2055 background.
- 2056 • p_T of the b-tag jet(s). The signal sample b-tag jet(s) tend to have softer p_T .
- 2057 • Invariant masses of the various objects ($\tau\tau$ legs and the b-tag jet(s)), which
2058 tend to be smaller for the signal samples.
- 2059 • The angular separation ΔR between pairs of the objects, where signal samples
2060 peak at smaller ΔR values.
- 2061 • The transverse mass between the missing transverse energy p_T^{miss} and each of

2062 the four objects [82], defined as

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

2063 where p_T^ℓ is the transverse momentum of the object ℓ , and $\Delta\phi$ is the difference
 2064 in azimuthal angle between the object and the p_T^{miss} . Events from $t\bar{t}$ +jets and
 2065 jets faking τ_h backgrounds have larger p_T^{miss} resulting in larger transverse mass
 2066 values compared to the signal, which tends to have smaller p_T^{miss} that is also
 2067 more aligned with the lepton legs.

- 2068 • The variable D_ζ [82], defined as

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

2069 where the ζ axis is the bisector of the transverse directions of the visible τ decay
 2070 products. p_ζ is the component of the p_T^{miss} along the ζ axis, and p_ζ^{vis} is the sum
 2071 of the components of the lepton p_T along the same axis. This variable captures
 2072 the fact that in signal the p_T^{miss} is small and approximately aligned with the $\tau\tau$.
 2073 In contrast, the $Z \rightarrow \tau\tau$ background tends towards large D_ζ values because the
 2074 p_T^{miss} is collinear to the $\tau\tau$, and the $t\bar{t}$ +jets events tend to have small D_ζ due to
 2075 a large p_T^{miss} not aligned with the $\tau\tau$.

- 2076 • For events with 2 b-tag jets, one additional variable is defined to capture the
 2077 difference in the invariant mass of the bb and the $\tau\tau$:

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

2078 This variable peaks at zero for the $h \rightarrow aa \rightarrow 2b2\tau$ signal.

2079 **Categorization using the DNN score**

2080 After training, events in data, MC, and embedded are evaluated with the six DNNs
 2081 and assigned a raw score between 0 and 1 (background-like or signal-like). In order
 2082 to flatten the distribution of the score and define score thresholds for categorizing
 2083 events, the raw output scores are transformed with the function $\tilde{p}(n) = \text{arctanh}(p \times$
 2084 $\tanh(n))/n$ where n is a positive integer. The thresholds of the DNN score used for
 2085 signal/control region definition are determined using scans that optimize the signal
 2086 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.

2087 **9.3 Methodology for signal extraction**

2088 In this section we outline the statistics terminology and concepts underlying the
 2089 modified frequentist method CL_S used to perform signal extraction.

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

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

2090

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 [97]. 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

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

2091

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

parameters, typically only a few are of interest, and are called the parameters of interest (POI), labeled μ here. The remaining are referred to as nuisance parameters (NP) [97] and are labeled $\boldsymbol{\theta}$.

$f(x)$ is the probability density for the observable in one event and we wish to describe the probability density for a dataset with many events, $\mathcal{D} = \{x_1, \dots, x_n\}$, called the total probability model \mathbf{f} . For instance, if we also have a prediction for the total number of events expected, called ν , we also account for the overall Poisson 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 [97]. The likelihood function is not a probability density for α and is not normalized to unity:

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

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

To estimate the parameter α we use an estimator, which is a function of the data. Take for example the measurement of data distributed according to a Gaussian probability density $f(x|\mu, \sigma) = \text{Gauss}(x|\mu, \sigma)$. One possible estimator of the mean μ , is the mean of the measured data points $\bar{x} = \sum_{i=1}^n x_i/n$ [97].

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

2113 9.3.2 Hypothesis testing

2114 In this section we next introduce concepts related to hypothesis testing such as the
2115 test statistic constructed from the ratio of likelihood functions.

2116 The objective of a likelihood analysis is to distinguish different models repre-
2117 senting the various hypotheses, and determine the one that best explains the ex-
2118 perimental outcome. In a search for new physics, a signal is additive on top of the
2119 background. The background-only hypothesis is the null hypothesis, and the signal-
2120 plus-background hypothesis is the alternative.

2121 As a simple example, take the p -value test, for an experiment where we count
2122 events in the signal region, n_{SR} , and expect ν_B background events and ν_S events from
2123 the signal [97]. Then

- 2124 1. The null hypothesis (H_0), i.e. the background-only hypothesis in this experi-
2125 ment, with the probability modeled by $\text{Poisson}(n_{SR}|\nu_B)$.
- 2126 2. The alternate hypothesis (H_1), i.e. signal-plus-background hypothesis, with the
2127 probability modeled by $\text{Poisson}(n_{SR}|(\nu_B + \nu_S))$.

2128 The compatibility of the observed data ν_{SR}^0 and the null hypothesis, is quantified as
2129 the probability that the background-only hypothesis would produce at least as many
2130 events as was observed. This probability is the p -value:

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

2131 If the p -value is very small, we might reject the null hypothesis. The p -value is not the
2132 probability of the null hypothesis given the data; rather, it expresses the probability
2133 that data with a certain property was obtained, assuming the null hypothesis [97].

2134 The p -value is an example of a test statistic T , which maps the data to a single
2135 real number. The Neyman-Pearson lemma states that out of the infinite possibilities

2136 of choices of test statistic, the uniformly most powerful test statistic is the likelihood
2137 ratio T_{NP} [97]:

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

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:

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

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

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

2138 where T_0 is the value of T based on the observed data, and $\theta()$ is the Heaviside
2139 function. The size of the test is conventionally chosen to be 10%, 5%, or 1%. As
2140 the p -value depends on $\boldsymbol{\alpha}$ (both the POI and NP), the null hypothesis should not be
2141 rejected if the p -value is larger than the size of the test for any value of the nuisance
2142 parameters.

2143 9.3.3 Confidence intervals

2144 In an example of the measurement of the Standard Model Higgs boson, $\boldsymbol{\alpha}_{\text{POI}} =$
2145 $(\sigma/\sigma_{SM}, M_H)$, with σ/σ_{SM} is the ratio of the production cross-section for Higgs with
2146 respect to its value in the SM, and M_H is the unknown mass of the Higgs, values
2147 of these parameters outside specific bounds are said to be “excluded at the 95%
2148 confidence level”. These allowed regions are called confidence levels or confidence
2149 regions, and the parameter values outside of them are considered excluded [97]. A

2150 95% confidence interval does not mean that there is a 95% chance that the true value
 2151 of the parameter is inside the interval. Rather, a 95% confidence interval covers the
 2152 true value 95% of the time (even though we do not know the true value).

2153 To construct a confidence interval for a parameter α , the Neyman Construction
 2154 is used to invert a series of hypothesis tests; i.e. for each possible value of α , the null
 2155 hypothesis is treated as α , and we perform a hypothesis test based on a test statistic.
 2156 To construct a 95% confidence interval, we construct a series of hypothesis tests with
 2157 size of 5%. The confidence interval $I(\mathcal{D})$ is constructed by taking the set of parameter
 2158 values $\boldsymbol{\alpha}$ where the null hypothesis is accepted:

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

2159 where $T(\mathcal{D})$ is the test statistic, and the last α (not bolded) and the subscript k_α
 2160 refer to the size of the test. A schematic of the Neyman construction is shown in Fig.
 2161 9.1. In a more generalized case, the x -axis is the test statistic T .

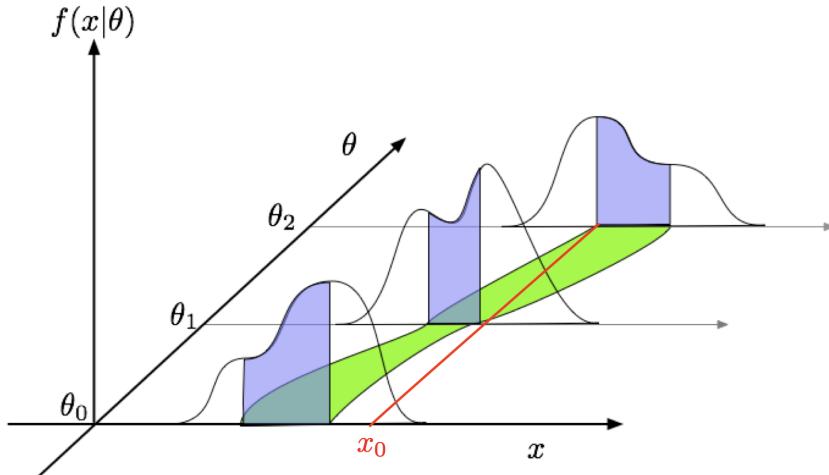


Figure 9.1: Schematic of the Neyman construction for confidence intervals [97]. For each value of θ , we find a region in x where $\int f(x|\theta)dx$ satisfies the size of the test (blue). These regions form a confidence belt (green). The intersection of the observation x_0 (red) with the confidence belt defines the confidence interval $[\theta_1, \theta_2]$ [97].

2162 **9.3.4 Profile likelihood ratio**

2163 In this section we describe a frequentist statistical procedure based on the profile
2164 likelihood ratio test statistic, which is implemented using asymptotic distributions.

2165 With a multi-parameter likelihood function $L(\boldsymbol{\alpha})$, the maximum likelihood of
2166 one specific parameter α_p with other parameters $\boldsymbol{\alpha}_o$ fixed, is called the conditional
2167 maximum likelihood estimate and is denoted $\hat{\alpha}_p(\boldsymbol{\alpha}_0)$. The process of choosing specific
2168 values of the nuisance parameters for a given value of μ , $\mathcal{D}_{\text{simulated}}$, and value of global
2169 observables \mathcal{G} is called profiling. From the full list of parameters $\boldsymbol{\alpha}$, we denote the
2170 parameter of interest μ , and the nuisance parameters $\boldsymbol{\theta}$.

2171 We construct the profile likelihood ratio,

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

2172 which depends explicitly on the parameter of interest μ , implicitly on the data \mathcal{D}_{sim}
2173 and global observables \mathcal{G} , and is independent of the nuisance parameters $\boldsymbol{\theta}$, which
2174 have been eliminated in profiling [97].

2175 The main conceptual reason for constructing the test statistic from the profile
2176 likelihood ratio is that asymptotically (i.e. for measurements with many events) the
2177 distribution of the profile likelihood ratio $\lambda(\mu = \mu_{\text{true}})$ is independent of the values of
2178 the nuisance parameters [97].

2179 The following p -value is used to quantify the consistency with the hypothesis of a
2180 signal strength of μ :

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

2181 9.3.5 Modified frequentist method: CL_S

2182 In the modified frequentist method called CL_S , to test a hypothesis with signal, we
 2183 define p'_μ as a ratio of p -values [97]:

$$2184 p'_\mu = \frac{p_\mu}{1 - p_b} \quad (9.13)$$

2184 where p_b is the p -value derived under the background-only hypothesis:

$$2185 p_b = 1 - p_0 \equiv 1 - \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu | 0, \hat{\theta}(\mu = 0, \text{obs})) d\tilde{q}_\mu. \quad (9.14)$$

2185 The CL_S upper limit on μ , denoted μ_{up} , is obtained by solving for $p'_{\mu_{up}} = 5\%$.
 2186 If testing the compatibility of the data with the background-only hypothesis, we
 2187 consider the p_b value defined above and conventionally convert it into the quantile
 2188 or “sigma” of a unit Gaussian. z standard deviations (e.g. $z = 5$ in “ 5σ ”) means
 2189 that the probability of falling above these standard deviations, equals p_b (e.g. 3σ
 2190 corresponds to $p_b = 2.7 \times 10^{-3}$ or 95.43%, and 5σ corresponds to $p_b = 5.7 \times 10^{-7}$ or
 2191 99.999943%).

2192 **Chapter 10**

2193 **Results**

2194 **10.1 Results from $bb\tau\tau$**

2195 In each of the three $\tau\tau$ channels studied ($\mu\tau_h$, $e\tau_h$, and $e\mu$), events are divided based
2196 on whether they contain exactly 1 or 2 b-tag jets, and further divided into signal
2197 and control regions (SRs and CRs) using the DNN categorization score as described
2198 in Section 9.2. The control regions demonstrate good agreement between observed
2199 events in data, and the sum of the contributions from expected backgrounds that
2200 are modeled in simulated and embedded samples. The signal regions are defined to
2201 be sensitive to the $h \rightarrow aa \rightarrow bb\tau\tau$ signal. The postfit final observed and expected
2202 distributions of the di-tau invariant mass $m_{\tau\tau}$ reconstructed with SVFit (described
2203 in Section 5.2) are shown in Fig. 10.1 for the $\mu\tau_h$ channel, Fig. 10.2 for the $e\tau_h$
2204 channel, and Fig. 10.3 for the $e\mu$ channel. In all figures, the hypothesized yield for
2205 the $h \rightarrow aa \rightarrow bb\tau\tau$ signal is shown for the pseudoscalar mass $m_a = 35$ GeV and
2206 assuming a branching fraction $B(H \rightarrow aa \rightarrow bb\tau\tau) = 10\%$.

2207 The 95% CL expected and observed exclusion limits on the signal strength of the
2208 branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ as a function of the pseudoscalar mass m_a
2209 ranging from 12 GeV to 60 GeV, are shown for the three $\tau\tau$ channels and all three

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

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

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

2235 10.2 Combination with $bb\mu\mu$ final state

2236 Results from this analysis for the $h \rightarrow aa \rightarrow bb\tau\tau$ final state are combined with the
2237 analysis for the $h \rightarrow aa \rightarrow bb\mu\mu$ final state [98]. While the predicted branching ratio
2238 for $aa \rightarrow bb\mu\mu$ is comparatively small, the $bb\mu\mu$ final state has competitive results
2239 due to the excellent di-muon resolution measured by CMS. The $bb\mu\mu$ analysis uses
2240 an unbinned fit to the data using the di-muon mass $m_{\mu\mu}$ distribution. Details can be
2241 found in [98].

2242 Combining the results is possible since the $bb\tau\tau$ analysis explicitly rejects events
2243 with extra leptons, so there is no overlap between the events studied in the $bb\tau\tau$
2244 analysis and the $bb\mu\mu$ analysis. In the statistical combination, several systematic
2245 uncertainties are treated as correlated: the integrated luminosity normalization, the
2246 b-tagging scale factor, the scale factors related to muon reconstruction, identifica-
2247 tion, and trigger efficiencies, the inefficiency in the ECAL trigger readout, and the
2248 theoretical uncertainties related to signal modeling.

2249 Since the results in both final states are statistically limited, the combination ben-
2250 efits from the additional data. For $m_a = 35$ GeV, all systematic uncertainties amount
2251 to around 6% of the total uncertainty, with the dominant systematic uncertainties
2252 coming from jet energy systematics in the $bb\mu\mu$ final state, theoretical uncertainties
2253 in the signal, and uncertainties in the QCD multijet backgrounds in the $e\mu$ channel
2254 of the $bb\tau\tau$ final state.

2255 The mass distributions of the di-muon and di-tau objects ($m_{\mu\mu}$ and $m_{\tau\tau}$) are
2256 compared to the data in a combined maximum likelihood fit to derive upper limits
2257 on $B(h \rightarrow aa)$. The observed limits at 95% CL on $B(h \rightarrow aa)$ for different 2HDM+S
2258 scenarios, are shown for the search for $h \rightarrow aa \rightarrow bb\mu\mu$ in Fig. 10.5b, and the
2259 combined analyses $h \rightarrow aa \rightarrow bb\ell\ell$ in Fig. 10.6.

2260 Exclusion limits in a two-dimensional plane as a function of $\tan\beta$ and m_a are
2261 set for 2HDM+S Types II, III, and IV in Fig. 10.7. The most stringent constraints

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

The combined results from $h \rightarrow aa \rightarrow bb\ell\ell$ are compared with CMS results in other final states as a function of the pseudoscalar mass m_a : for 2HDM+S type I in Fig. 10.8, type II with $\tan \beta = 2.0$ in Fig. 10.9, and type III with $\tan \beta = 2.0$ in Fig. 10.10. In other scenarios, e.g. type III with $\tan \beta = 5.0$, more stringent limits are set by analyses in other final states, $\mu\mu\tau\tau$ in this case. Other summary plots for other model types and $\tan \beta$ values can be found at [99].

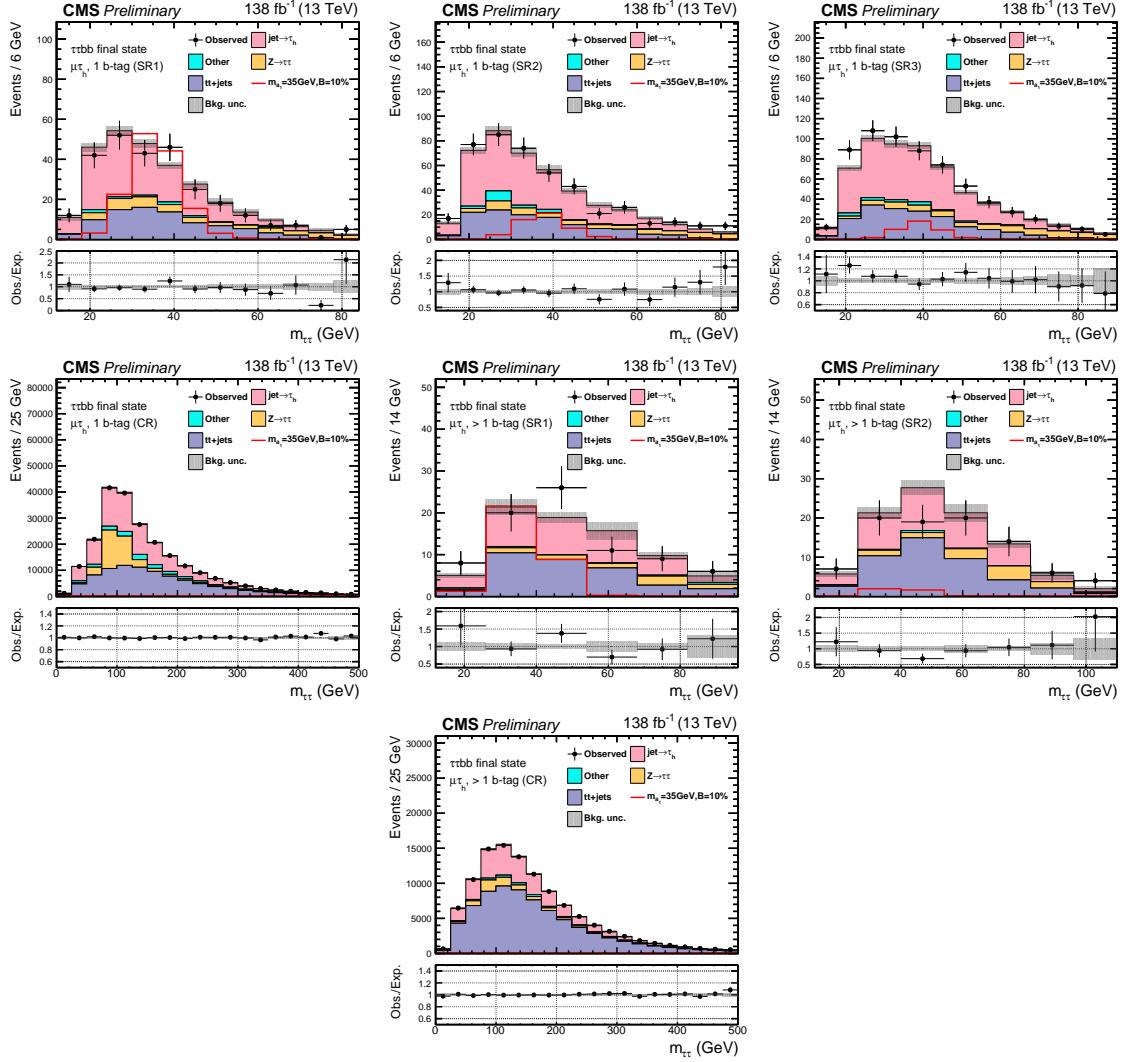


Figure 10.1: Postfit final $m_{\tau\tau}$ observed and expected distributions, and the observed/expected ratios, in the $\mu\tau_h$ channel [41]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), 1 b-tag jet control region (*middle row*), 2 b-tag jet signal regions (SR1, 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*).

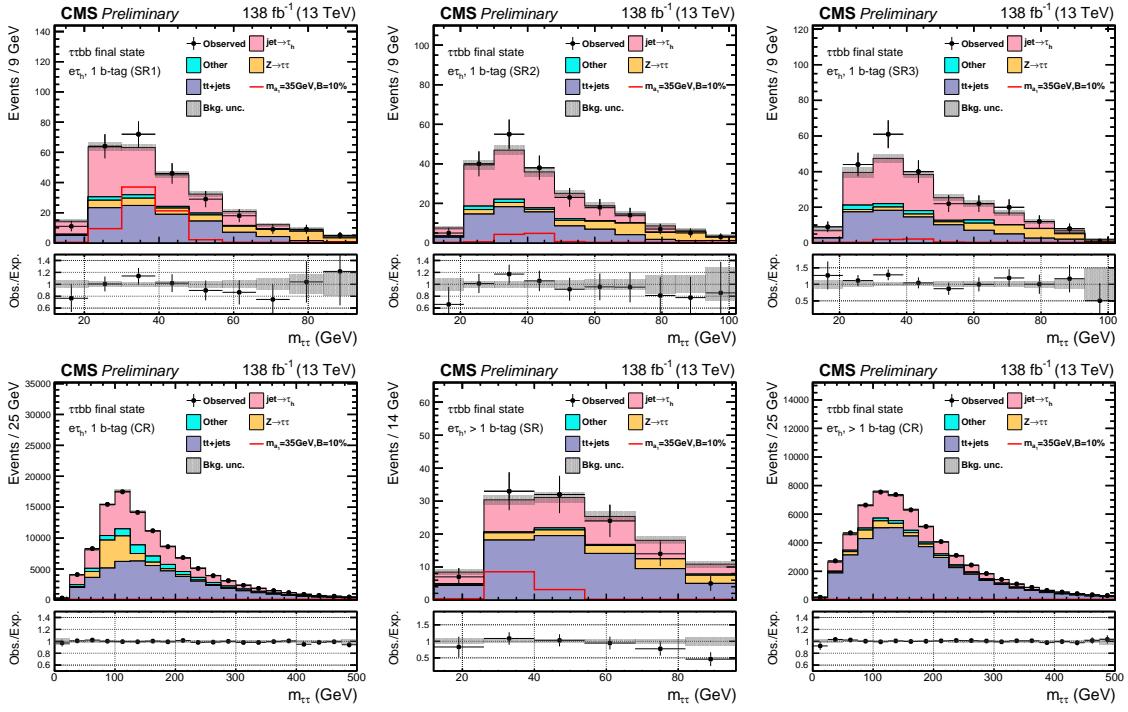


Figure 10.2: Postfit final observed and expected $m_{\tau\tau}$ distributions, and the observed/expected ratios, in the $e\tau_h$ channel [41]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), the 1 b-tag jet control region (CR) (*bottom row*), and 2 b-tag jet signal 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 $t\bar{t}+jets$ (*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*).

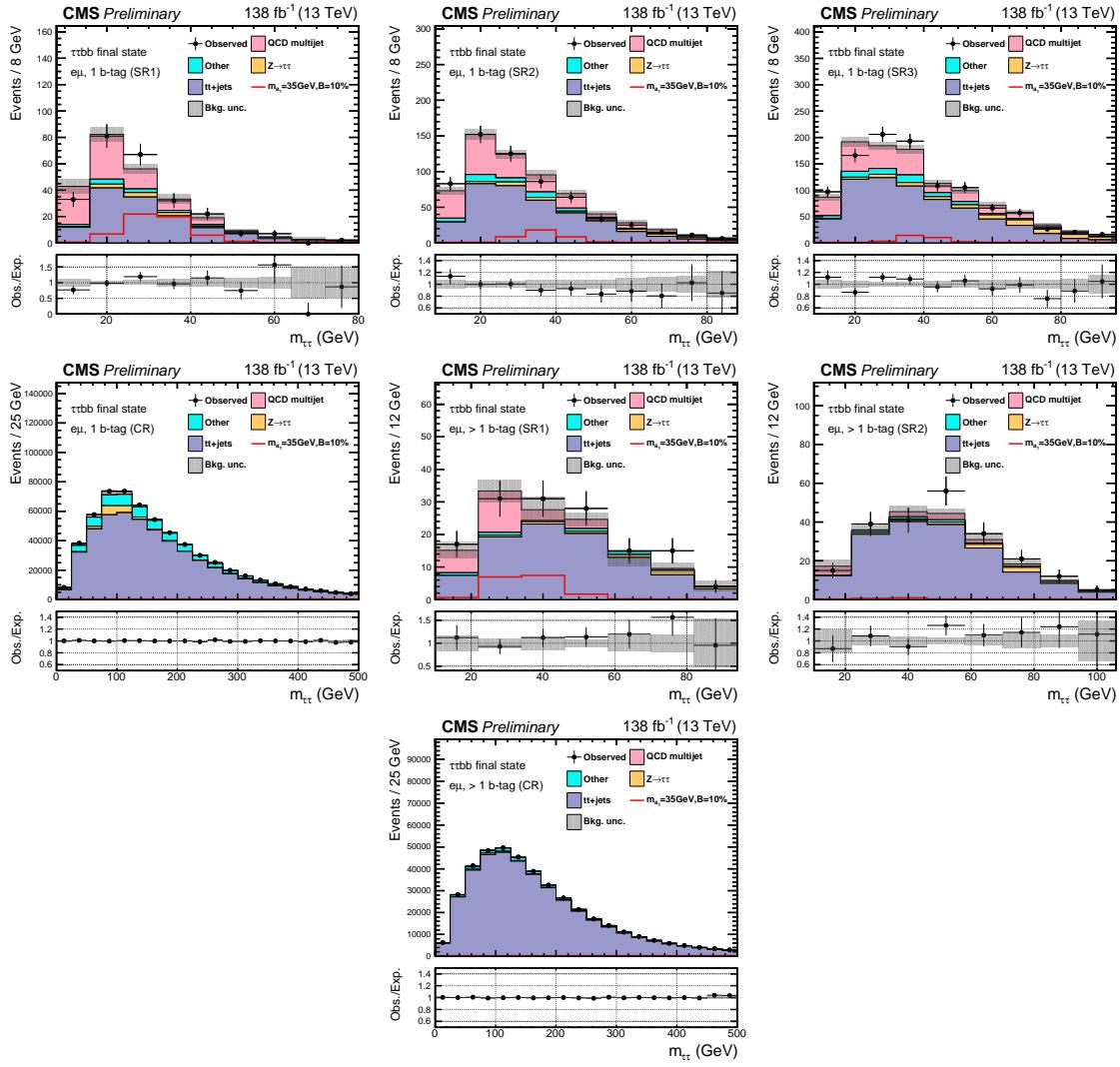


Figure 10.3: Postfit final observed and expected $m_{\tau\tau}$ distributions, and the observed/expected ratios, in the $e\mu$ channel [41]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, and SR3) (*top row*), 1 b-tag jet control region (CR) (*middle row*), 2 b-tag jet signal regions (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*).

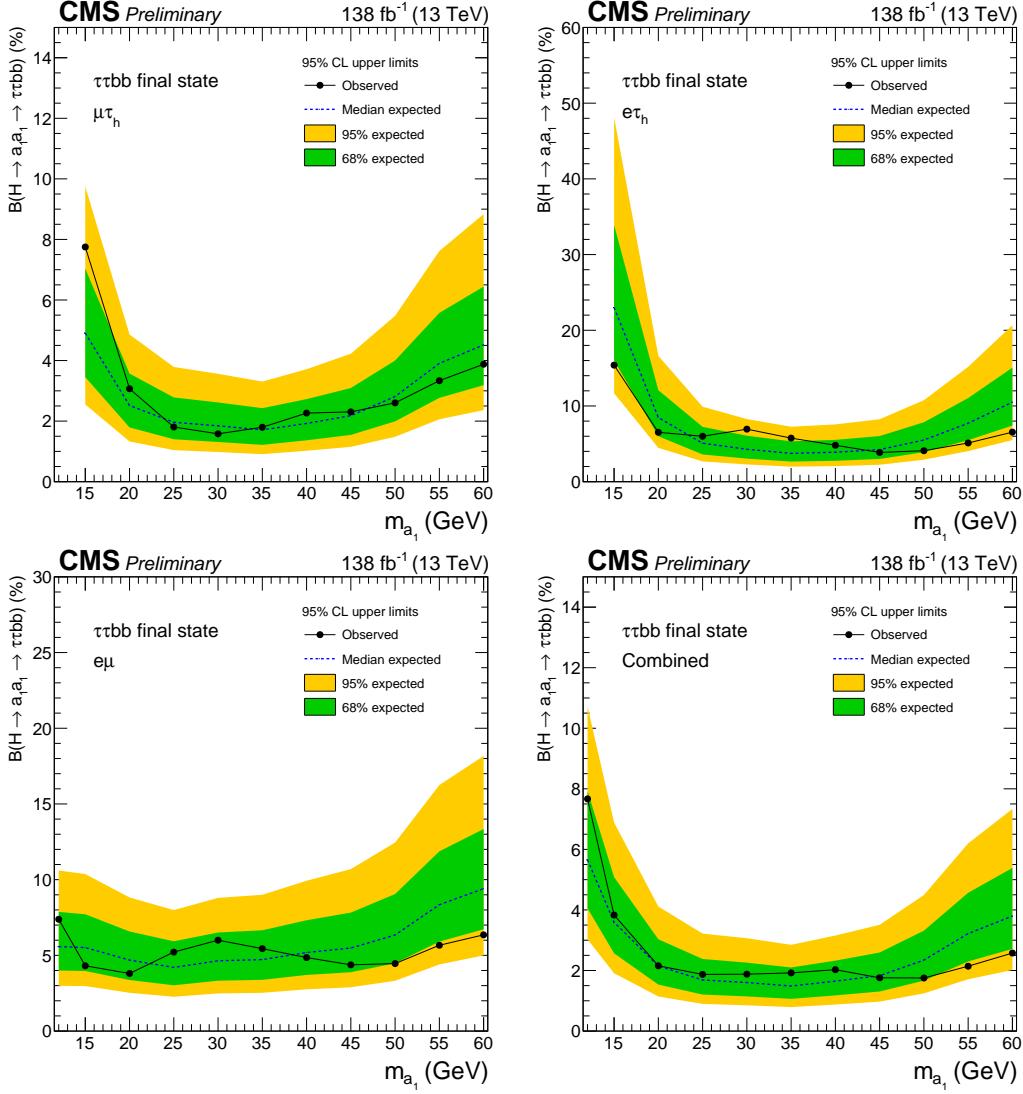
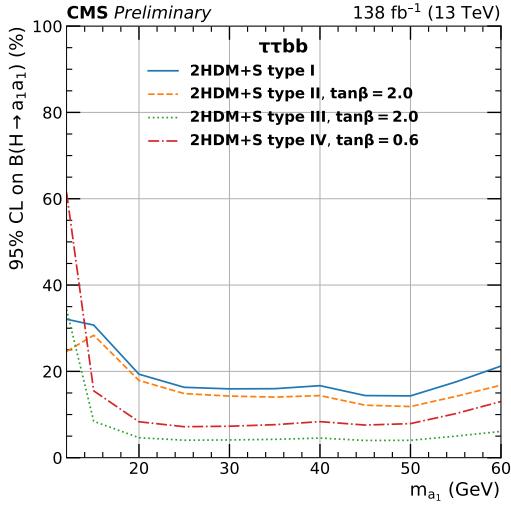
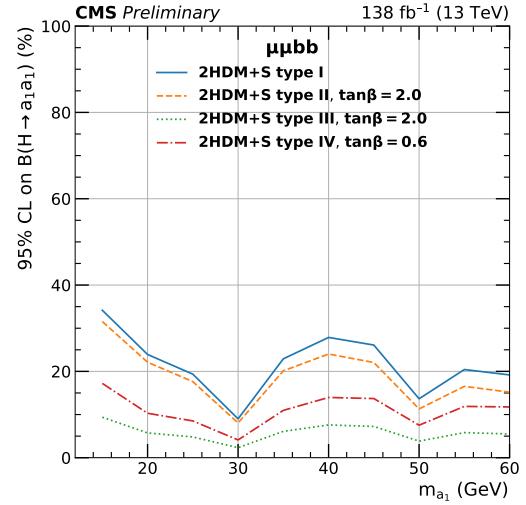


Figure 10.4: Observed 95% CL exclusion limits (*black, solid lines*) and expected 95% CL and 68% CL limits (*shaded yellow and green*) on the branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard Model production for the 125 GeV Higgs (h). Limits are shown 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*) [41]. 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) [41]. Linear interpolation is used between points in the graphs. The $\tan\beta$ values chosen here correspond to the most stringent limits in each model.

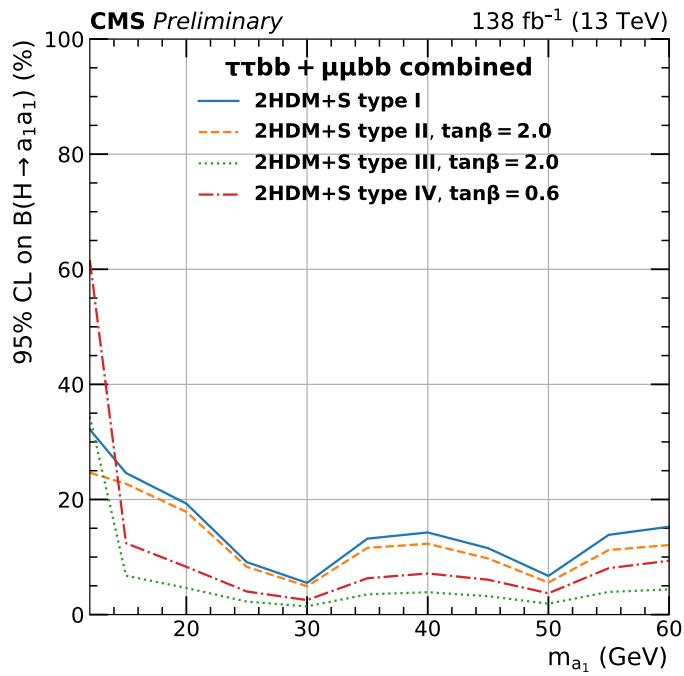


Figure 10.6: Observed 95% CL upper limits on the branching fraction of the 125 GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$, in percentages, as a function of the pseudoscalar mass m_a , in 2HDM+S type I (blue), type II with $\tan\beta = 2.0$ (orange dashed), type III with $\tan\beta = 2.0$ (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} [41].

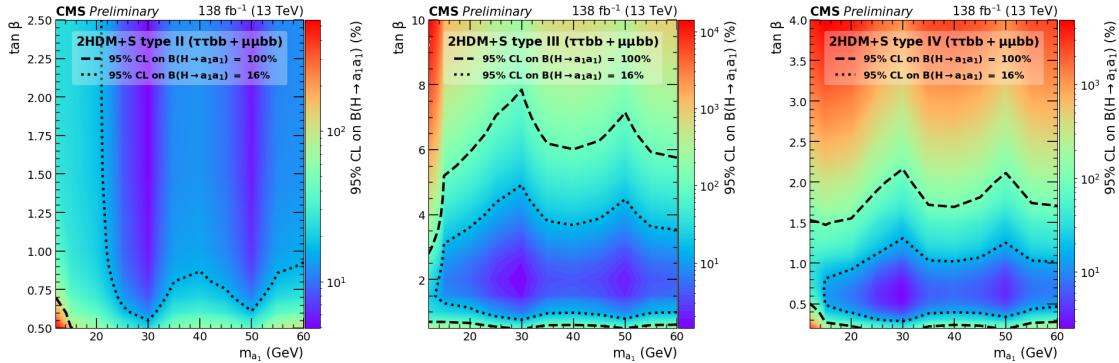


Figure 10.7: Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity of 138 fb^{-1} for Type II (*left*), Type III (*middle*), and Type IV (*right*) 2HDM+S in the $\tan \beta$ vs. m_a 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 [41].

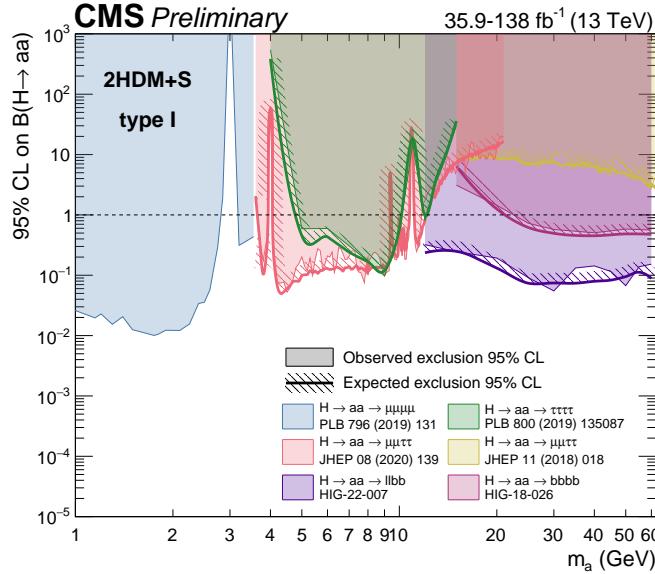


Figure 10.8: Summary plot of current 95% limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$ in the 2HDM+S type I scenario performed with data collected at 13 TeV [99]. Results from different final states studied at CMS are overlaid on this figure: $\mu\mu\mu\mu$ (blue), $\tau\tau\tau\tau$ (green), boosted $2\mu 2\tau$ (red), resolved $2\mu 2\tau$ (yellow), $bbbb$ (magenta), and the combined result for $\ell\ell bb$ ($\ell = \mu, \tau$) (purple).

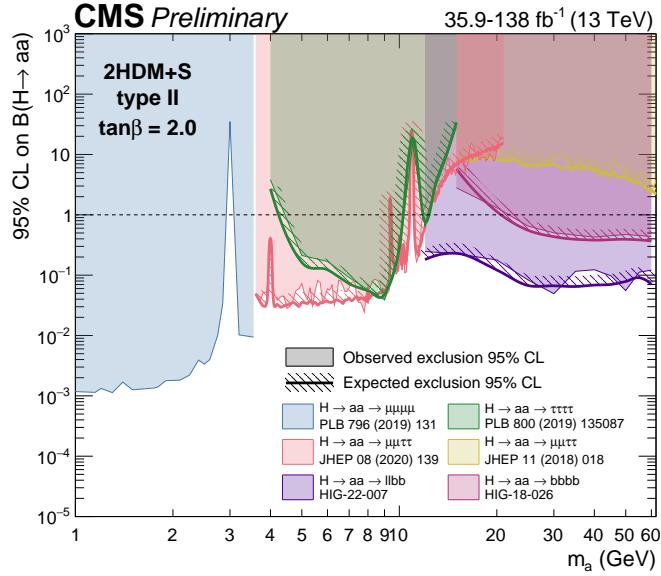


Figure 10.9: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$, in the 2HDM+S type II scenario with $\tan \beta = 2.0$, obtained at CMS with data collected at 13 TeV [99]. 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).

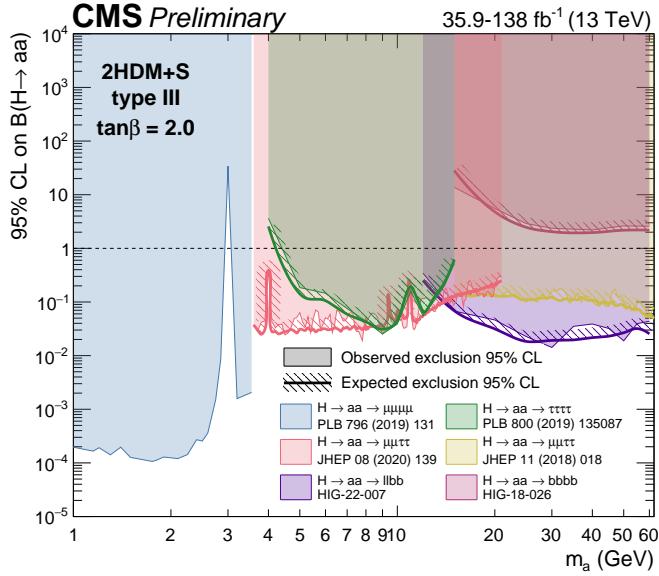


Figure 10.10: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross section, $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$ in the 2HDM+S type-III scenario with $\tan \beta = 2.0$, obtained at CMS with data collected at 13 TeV [99]. 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).

2275 **Chapter 11**

2276 **Asymmetric exotic Higgs decays**

2277 This chapter presents progress towards a search for exotic Higgs decays to two light
2278 scalars with unequal mass ($h \rightarrow a_1 a_2$) final states with bottom quarks and τ leptons,
2279 with plans to interpret the results in the context of Two Real Singlet Models (TRSMs),
2280 described in Section 1.5. Compared to the symmetric decay scenario $h \rightarrow aa$ which
2281 has been studied in multiple final states at CMS with stringent limits set on the
2282 various 2HDM+S scenarios, this asymmetric decay scenario has not been directly
2283 searched for at the CMS experiment. Section 11.1 lists the mass hypotheses of the
2284 new particles a_1 and a_2 that will be studied. Section 11.2 describes the studies on
2285 which channels the analysis will be carried out in. Section 11.3 shows the control
2286 plots produced using the analysis framework that will be used for this analysis.

2287 **11.1 Signal masses**

2288 As discussed in Section 1.5, $h \rightarrow a_1 a_2$ can result in a “cascade” decay if one of the
2289 scalars, a_2 is sufficiently heavy ($m_{a_2} > 2m_{a_1}$). The “non-cascade” case is where the
2290 light scalars decay directly to Standard Model particles.

2291 The mass hypotheses (mass points) (m_{a_1}, m_{a_2}) studied here are:

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

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

11.2 Cascade scenario signal studies

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

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

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

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

2317 and fourth b-flavor jets which have very low transverse momenta p_T . Event categories
 2318 with exactly 1 b-tag jet and ≥ 2 b-tag jets will be used.

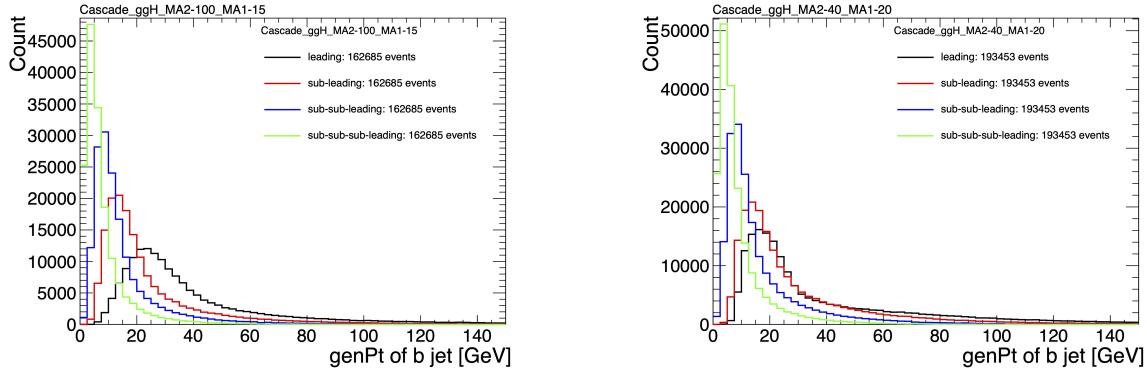


Figure 11.1: Generator-level b-flavor jet transverse momenta p_T , for $h \rightarrow a_1 a_2$ cascade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV (*left*) and $(40, 20)$ GeV (*right*). In each plot the generator-level p_T of the leading (*black*),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

2319 In the $4b2\tau$ final state, the possibility of the leading and sub-leading b-tag jets
 2320 being sufficiently close in ΔR to require boosted jet reconstruction techniques was
 2321 explored. In the $4b2\tau$ case, the two b-flavor-jets in the generated event that were
 2322 spatially closest in ΔR were considered as one object. This two b-flavor jet object was
 2323 spatially matched in ΔR to the jets reconstructed with the standard AK4 algorithm
 2324 which uses a cone size of $\Delta R = 0.4$. The quality of the p_T resolution (computed as
 2325 $(p_{T,\text{reconstructed}} - p_{T,\text{gen}})/p_{T,\text{gen}}$) and closeness in distance ΔR of the reconstructed jet
 2326 to the nearest generator-level jets, was seen to depend on the absolute and relative
 2327 masses of the light scalars. The best (worst) performance occurred in samples with
 2328 large (small) mass differences between the heavier scalar a_2 and the lighter scalar a_1 ,
 2329 as illustrated for the mass hypotheses (m_{a_1}, m_{a_2}) (100, 15) GeV and (40, 20) GeV in
 2330 Fig. 11.2.

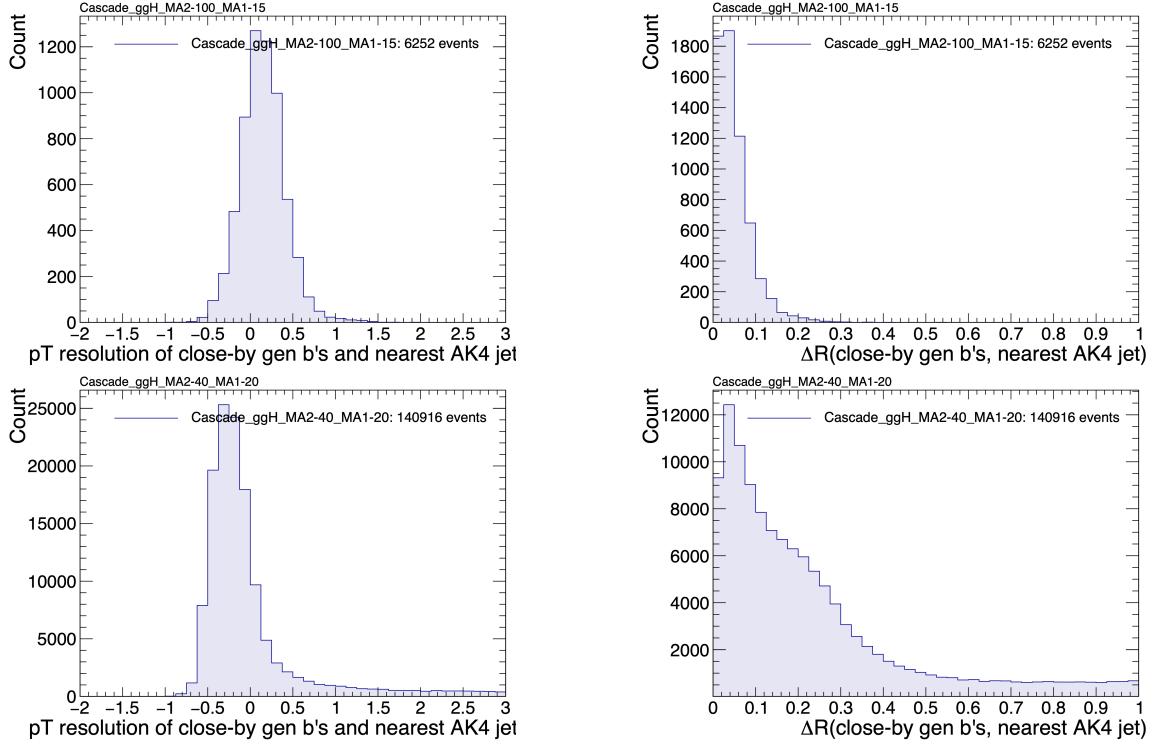


Figure 11.2: Distributions (arbitrary units) of transverse momentum p_T resolution and ΔR between the two closest generator-level b jets, treated as one object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow a_1 a_2$ mass hypotheses (m_{a_1}, m_{a_2}) = (100, 15) GeV (top left, 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 will be similar to those studied in $h \rightarrow aa \rightarrow bb\tau\tau$. For the $\mu\tau_h$ channel, histograms of the key kinematic variables are made for data and the sum of the expected backgrounds, which are estimated from Monte Carlo samples, embedded samples, and the data-driven method for estimating jets faking τ_h as described in Chapter 7. Nominal values of the scale factors and event reweighting are applied, as described in Chapter ???. The errors shown in the figures only include statistical errors and only several of the full set of systematic errors (only those associated with the lepton energy scales and τ_h identification efficiency,

2340 described in Sections 5.3.1, 5.3.2, and 5.3.4).

2341 The p_T , η , and ϕ of the leading muon and hadronic tau τ_h , and the di-tau visible
2342 mass m_{vis} and momentum $p_{T,\text{vis}}$, are shown in Fig. 11.3. The p_T , η , and ϕ of the the
2343 leading and sub-leading b-tag jets, and the missing transverse energy magnitude and
2344 azimuthal direction, are shown in Fig. 11.4.

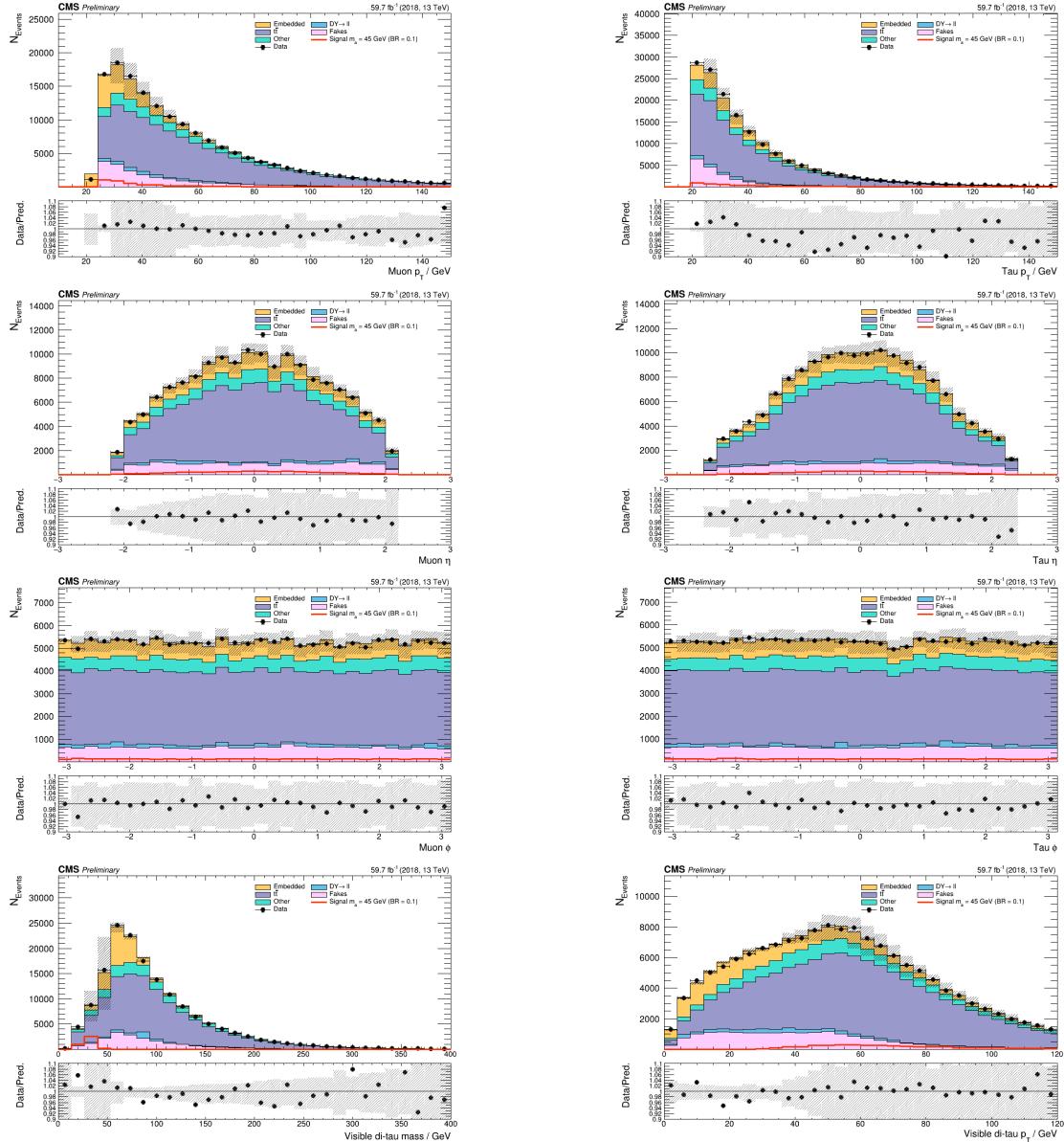


Figure 11.3: Kinematic properties of the leading muon and τ_h in the $\mu\tau_h$ channel: p_T (top row), η (second row), and ϕ (third row). The visible 4-momenta of the muon and τ_h are summed, giving the visible di-tau mass m_{vis} and transverse momentum $p_{T,\text{vis}}$. The errors shown in the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

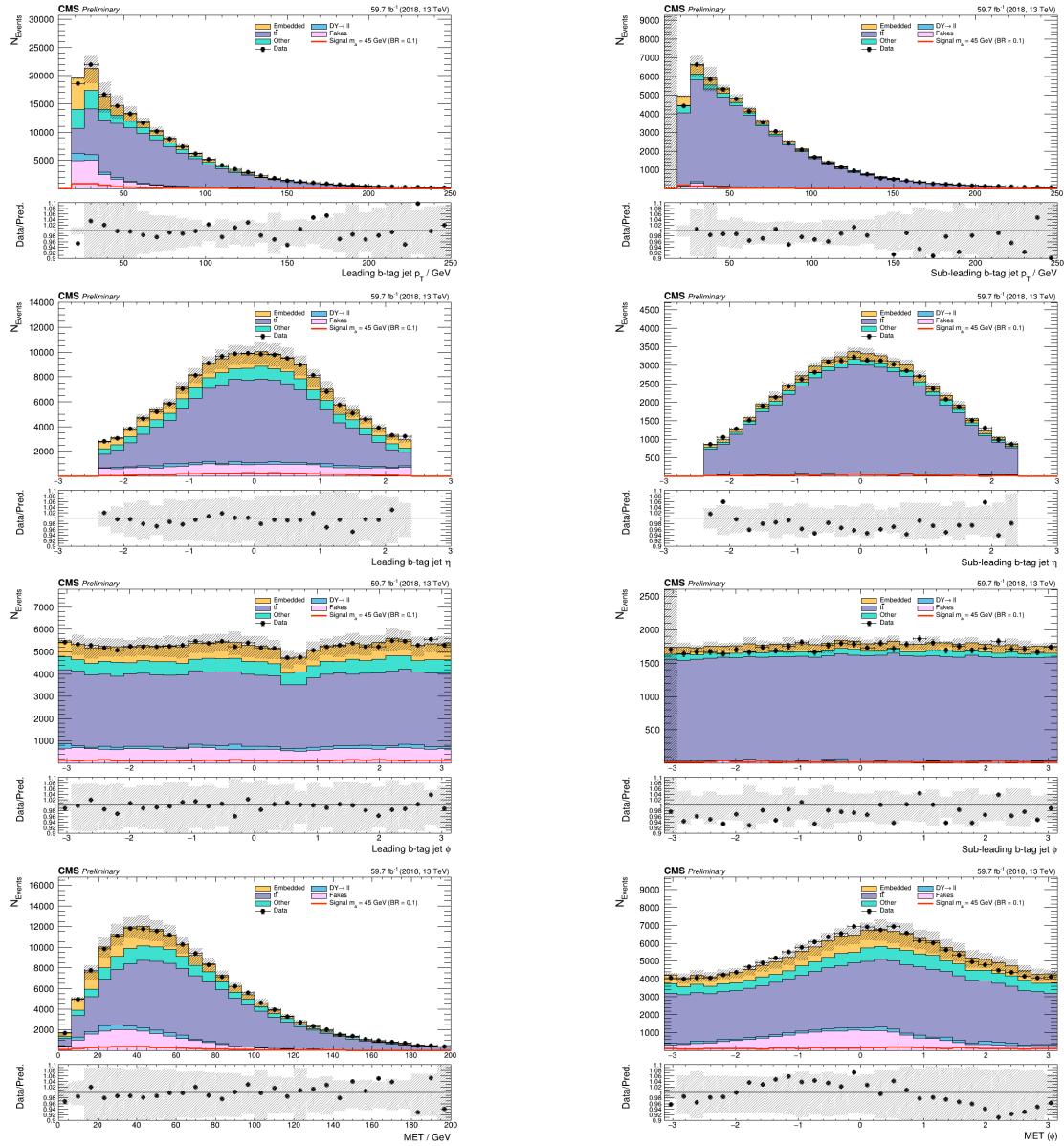


Figure 11.4: Kinematic properties of the leading and sub-leading b-tag jets in the $\mu\tau_h$ final state: jet p_T (*top row*), η (*second row*), ϕ (*third row*), as well as the missing transverse energy magnitude and azimuthal direction (*bottom row*). The errors shown in the figures only include statistical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

²³⁴⁵ **Chapter 12**

²³⁴⁶ **Conclusion and outlook**

²³⁴⁷ With the discovery of a Higgs boson with mass 125 GeV at the LHC in 2012, the LHC
²³⁴⁸ and CMS physics program has evolved to include the precise characterization of the
²³⁴⁹ 125 GeV Higgs boson and searching for evidence of additional Higgs particles in an
²³⁵⁰ extended Higgs sector. This thesis presents a direct search at CMS for exotic decays
²³⁵¹ of the Higgs boson with mass 125 GeV in data collected in the years 2016-2018 in
²³⁵² proton-proton collisions at center-of-mass energy 13 TeV, to two light neutral scalar
²³⁵³ particles that decay to two bottom quarks and two tau leptons ($h \rightarrow aa \rightarrow bb\tau\tau$). The
²³⁵⁴ results are combined with another search that was performed in the $h \rightarrow aa \rightarrow bb\mu\mu$
²³⁵⁵ final state, giving the most stringent limits to date for theories with Two Higgs
²³⁵⁶ Doublet Models extended with a singlet scalar (2HDM+S), for pseudoscalar masses
²³⁵⁷ m_a ranging from 15 GeV to 60 GeV, in a number of 2HDM+S scenarios such as type
²³⁵⁸ II and III with $\tan\beta = 2.0$.

²³⁵⁹ As the rich physics program of CMS has set stringent limits on the exotic decay
²³⁶⁰ $h \rightarrow aa$, we turn our attention to direct searches for decays to light neutral scalars
²³⁶¹ with potentially unequal mass, $h \rightarrow a_1a_2$, which has not been performed at CMS
²³⁶² to date. Preliminary studies on $h \rightarrow a_1a_2$ signals in the Two Real Singlet Model
²³⁶³ (TRSM) are shown, and work is ongoing to develop the analysis for $h \rightarrow a_1a_2$ in final

2364 states with bottom quarks and tau leptons.

2365 To ensure the continued performance of the CMS detector and to enhance its
2366 data-taking capabilities in the intense pileup conditions of the Phase-2 upgrade of
2367 the High-Luminosity LHC, upgrades of the Level-1 Trigger are paramount for filter-
2368 ing the increased data rate of the HL-LHC. This thesis presents work on the stan-
2369 dalone barrel calorimeter algorithm for reconstructing and identifying electron and
2370 photon candidates, using high granularity crystal-level information from the ECAL
2371 subdetector. For Phase-2, the increase in the granularity of information sent from
2372 the electromagnetic calorimeter to the Level-1 trigger, from energy sums over towers
2373 (which are 5×5 in crystals) to crystal-level information, allows for the implementation
2374 of a more sophisticated clustering algorithm that can exploit the fact that genuine
2375 electrons and photons tend to leave energies concentrated a 3×5 window in crystals,
2376 and use shape and isolation information to distinguish genuine electrons and photons
2377 from noise. Electrons and photons are key to characterizing Standard Model pro-
2378 cesses and performing searches for new physics, and this represents one of the many
2379 upgrades of the CMS detector in preparation for Phase-2. With the ongoing Run-3
2380 data collecting period, and wealth of ongoing and scheduled upgrades, there remains
2381 an abundance of directions for detector development and physics at CMS heading
2382 into Phase-2 of the LHC.

²³⁸³ Bibliography

- ²³⁸⁴ [1] Paul H. Frampton. Journeys Beyond the Standard Model. 54(1):52–
²³⁸⁵ 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.
- ²³⁸⁶ [2] Meinard Kuhlmann. Quantum Field Theory. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2023 edition, 2023.
- ²³⁸⁷ [3] Steven Weinberg. The Making of the Standard Model. *Eur. Phys. J. C*, 34:5–13, 2004. URL <https://cds.cern.ch/record/799984>.
- ²³⁸⁸ [4] Christopher G. Tully. *Elementary Particle Physics in a Nutshell*. Princeton University Press, Princeton, 2012. ISBN 9781400839353. doi: doi:10.1515/9781400839353. URL <https://doi.org/10.1515/9781400839353>.
- ²³⁸⁹ [5] John Ellis. Higgs Physics. In *2013 European School of High-Energy Physics*, pages 117–168, 2015. doi: 10.5170/CERN-2015-004.117.
- ²³⁹⁰ [6] David Curtin, Rouven Essig, Stefania Gori, and Others. Exotic decays of the 125 GeV Higgs boson. *Phys. Rev. D*, 90:075004, Oct 2014. doi: 10.1103/PhysRevD.90.075004. URL <https://link.aps.org/doi/10.1103/PhysRevD.90.075004>.
- ²³⁹¹ [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>.

- 2425 [17] Olivier S. Brning and Frank Zimmerman. Parameter space for the LHC lumi-
2426 nosity upgrade. ISBN 978-3-95450-115-1. URL <https://accelconf.web.cern.ch/IPAC2012/papers/moppc005.pdf>.
- 2428 [18] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09):
2429 P09018, 2020. doi: 10.1088/1748-0221/15/09/P09018.
- 2430 [19] CMS Collaboration. Measurement of the inelastic proton-proton cross section at
2431 $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018. doi: 10.1007/JHEP07(2018)161.
- 2432 [20] CMS Collaboration. High-Luminosity Large Hadron Collider (HL-LHC): Tech-
2433 nical design report. 10/2020, 12 2020. doi: 10.23731/CYRM-2020-0010.
- 2434 [21] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004,
2435 2008. doi: 10.1088/1748-0221/3/08/S08004.
- 2436 [22] CMS Collaboration. Particle-flow reconstruction and global event description
2437 with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/
2438 12/10/P10003.
- 2439 [23] V Karimki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudens-
2440 reich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker*
2441 *system project: Technical Design Report*. Technical design report. CMS. CERN,
2442 Geneva, 1997. URL <https://cds.cern.ch/record/368412>.
- 2443 [24] The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva,
2444 2017. URL <https://cds.cern.ch/record/2272264>.
- 2445 [25] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Up-
2446 grade. 9 2012. doi: 10.2172/1151650.
- 2447 [26] R. L. Workman and Others. Review of particle physics. 2022:083C01. doi:
2448 10.1093/ptep/ptac097.

- 2449 [27] CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter. 9 2012. doi: 10.2172/1151651.
- 2450
- 2451 [28] CMS Technical Design Report for the Level-1 Trigger Upgrade. 6 2013.
- 2452
- 2453 [29] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The trigger systems. 12 2000.
- 2454
- 2455 [30] 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.
- 2456
- 2457
- 2458 [31] C. Foudas. The CMS Level-1 Trigger at LHC and Super-LHC. In *34th International Conference on High Energy Physics*, 10 2008.
- 2459
- 2460 [32] CMS Software Guide. High Level Trigger (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHighLevelTrigger>.
- 2461
- 2462 [33] The Worldwide LHC Computing Grid. 2012. URL <https://cds.cern.ch/record/1997398>.
- 2463
- 2464 [34] Douglas Thain, Todd Tannenbaum, and Miron Livny. Distributed computing in practice: the Condor experience. *Concurrency and Computation: Practice and Experience*, 17(2-4):323–356, 2005. doi: <https://doi.org/10.1002/cpe.938>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.938>.
- 2465
- 2466
- 2467
- 2468 [35] Alexandre Zabi, Jeffrey Wayne Berryhill, Emmanuelle Perez, and Alexander D. Tapper. The Phase-2 Upgrade of the CMS Level-1 Trigger. 2020.
- 2469
- 2470 [36] Technical proposal for a MIP timing detector in the CMS experiment Phase 2 upgrade. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2296612>.
- 2471
- 2472

- 2473 [37] CMS Collaboration. CMS luminosity measurement for the 2016 data-taking
2474 period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL
2475 <https://cds.cern.ch/record/2257069>.
- 2476 [38] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking
2477 period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-
2478 004, 2018. URL <https://cds.cern.ch/record/2621960>.
- 2479 [39] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking
2480 period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-
2481 002, 2019. URL <https://cds.cern.ch/record/2676164>.
- 2482 [40] CMS LUMI Group. CMS Luminosity Public Results (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- 2483 [41] Cécile Caillol, Pallabi Das, Sridhara Dasu, Pieter Everaerts, Stephanie Kwan,
2484 Isobel Ojalvo, and Ho-Fung Tsoi. CMS AN-20-213 (internal): Search for an
2485 exotic decay of the 125 GeV Higgs boson to light pseudoscalars, with a pair of b
2486 jets and a pair of tau leptons in the final state, 2020.
- 2487 [42] CMS Collaboration. Search for exotic decays of the Higgs boson to a pair of
2488 pseudoscalars in the $\mu\mu bb$ and $\tau\tau bb$ final states. *European Physical Journal C*,
2489 2 2024.
- 2490 [43] J. Alwall, R. Frederix, S. Frixione, et al. The automated computation of tree-
2491 level and next-to-leading order differential cross sections, and their matching to
2492 parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.
2493 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).
- 2494 [44] R. Frederix, S. Frixione, V. Hirschi, et al. The automation of next-to-leading
2495 order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July

- 2498 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).
- 2500 [45] S. Agostinelli, J. Allison, K. Amako, et al. Geant4 - a simulation toolkit. 506(3):
2501 250–303. ISSN 0168-9002. doi: 10.1016/S0168-9002(03)01368-8. URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- 2502
2503 [46] CMS Collaboration. An embedding technique to determine $\tau\tau$ backgrounds
2504 in proton-proton collision data. *JINST*, 14(06):P06032, 2019. doi: 10.1088/
2505 1748-0221/14/06/P06032.
- 2506 [47] CMS Collaboration. Search for neutral MSSM Higgs bosons decaying to a pair of
2507 tau leptons in pp collisions. *JHEP*, 10:160, 2014. doi: 10.1007/JHEP10(2014)160.
- 2508 [48] CMS Collaboration. Measurements of Higgs boson production in the decay chan-
2509 nel with a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur.*
2510 *Phys. J. C*, 83(7):562, 2023. doi: 10.1140/epjc/s10052-023-11452-8.
- 2511 [49] CMS Collaboration. Reconstruction and identification of tau lepton decays to
2512 hadrons and tau neutrinos at CMS. *Journal of Instrumentation*, 11(01):P01019–
2513 P01019, January 2016. ISSN 1748-0221. doi: 10.1088/1748-0221/11/01/p01019.
2514 URL <http://dx.doi.org/10.1088/1748-0221/11/01/P01019>.
- 2515 [50] CMS Collaboration. Performance of τ -lepton reconstruction and identification
2516 in CMS. *Journal of Instrumentation*, 7(01):P01001, jan 2012. doi: 10.1088/
2517 1748-0221/7/01/P01001. URL <https://dx.doi.org/10.1088/1748-0221/7/01/P01001>.
- 2518
2519 [51] CMS Collaboration. Performance of reconstruction and identification of τ leptons
2520 decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):
2521 P10005, 2018. doi: 10.1088/1748-0221/13/10/P10005.

- 2522 [52] CMS Collaboration. Identification of hadronic tau lepton decays using a deep
2523 neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/
2524 P07023.
- 2525 [53] CMS Collaboration. Performance of CMS muon reconstruction in pp collision
2526 events at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 7(10):P10002–P10002,
2527 October 2012. ISSN 1748-0221. doi: 10.1088/1748-0221/7/10/p10002. URL
2528 <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- 2529 [54] CMS Collaboration. Performance of electron reconstruction and selection with
2530 the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instru-*
2531 *mentation*, 10(06):P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005. URL
2532 <https://dx.doi.org/10.1088/1748-0221/10/06/P06005>.
- 2533 [55] CMS Collaboration. Identification of b-quark jets with the CMS experiment.
2534 *Journal of Instrumentation*, 8(04):P04013–P04013, April 2013. ISSN 1748-
2535 0221. doi: 10.1088/1748-0221/8/04/p04013. URL <http://dx.doi.org/10.1088/1748-0221/8/04/P04013>.
- 2537 [56] CMS Collaboration. Pileup Removal Algorithms. Technical report, CERN,
2538 Geneva, 2014. URL <https://cds.cern.ch/record/1751454>.
- 2539 [57] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and
2540 developments. 2017. URL <https://cds.cern.ch/record/2263802>.
- 2541 [58] CMS Collaboration. Performance of the DeepJet b tagging algorithm using 41.9
2542 fb^{-1} of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector.
2543 2018. URL <https://cds.cern.ch/record/2646773>.
- 2544 [59] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Re-
2545 construction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood

- 2546 techniques. *Journal of Physics: Conference Series*, 513(2):022035, jun 2014.
2547 doi: 10.1088/1742-6596/513/2/022035. URL <https://dx.doi.org/10.1088/1742-6596/513/2/022035>.
- 2549 [60] CMS Collaboration. Evidence for the 125 GeV Higgs boson decaying to a pair
2550 of τ leptons. *JHEP*, 05:104, 2014. doi: 10.1007/JHEP05(2014)104.
- 2551 [61] CMS Collaboration. Missing transverse energy performance of the CMS detector.
2552 *JINST*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.
- 2553 [62] Artur Kalinowski. CMS AN-19-032 (internal): Reconstruction of a τ pair invariant
2554 mass with a simplified likelihood scan, 2019.
- 2555 [63] CMS TAU POG. Tau Physics Object Group: Tau ID Recommendation For Run 2, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2>.
- 2558 [64] CMS MUO POG. Muon Physics Object Group: Recommendations, 2024. URL
2559 <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>.
- 2560 [65] CMS MUO POG. Muon Physics Object Group: Reference guidelines and results
2561 for muon momentum scale and resolution in Run II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceScaleResolRun2>.
- 2563 [66] CMS HTT working group. Higgs To Tau Tau Working TWiki for the full Run-
2564 2 legacy analysis, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/HiggsToTauTauWorkingLegacyRun2>.
- 2566 [67] CMS ELE POG. Electron Physics Object Group: Recommendations,
2567 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations>.

- 2569 [68] CMS ELE POG. Electron Physics Object Group: Recommendations for
2570 2016 to 2018 UL, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018>.
2571
- 2572 [69] CMS TAU Embedding Group. Tau embedded samples using 2016
2573 data, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauTauEmbeddingSamples2016Legacy>.
2574
- 2575 [70] CMS TAU Embedding Group. Tau embedded samples using 2017
2576 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2017>.
2577
- 2578 [71] CMS TAU Embedding Group. Tau embedded samples using 2018
2579 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2018>.
2580
- 2581 [72] Tau Lepton Run 2 Trigger Performance. 2019. URL <https://cds.cern.ch/record/2678958>.
2582
- 2583 [73] CMS Taus High Level Trigger Studies. Tau Lepton Run 2 Trigger Performance
2584 (CMS DP-2019/012), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTTauAllRun2>.
2585
- 2586 [74] Muon HLT Performance with 2018 Data. 2018. URL <https://cds.cern.ch/record/2627469>.
2587
- 2588 [75] Single and Double Electron Trigger Efficiencies using the full Run 2 dataset.
2589 2020. URL <https://cds.cern.ch/record/2888577>.
2590
- 2591 [76] Run II Trigger Performance For $e\mu$ Triggers. 2019. URL <https://cds.cern.ch/record/2687013>.

- 2592 [77] Performance of electron and photon reconstruction in Run 2 with the CMS ex-
2593 periment. 2020. URL <https://cds.cern.ch/record/2725004>.
- 2594 [78] CMS Collaboration. Performance of the CMS muon detector and muon recon-
2595 struction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015,
2596 2018. doi: 10.1088/1748-0221/13/06/P06015.
- 2597 [79] Piet Verwilligen. Muons in the cms high level trigger system. *Nuclear
2598 and Particle Physics Proceedings*, 273-275:2509–2511, 2016. ISSN 2405-6014.
2599 doi: <https://doi.org/10.1016/j.nuclphysbps.2015.09.441>. URL <https://www.sciencedirect.com/science/article/pii/S240560141500930X>. 37th Inter-
2600 national Conference on High Energy Physics (ICHEP).
2601
- 2602 [80] Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-
2603 probe technique. 2020. URL <https://cds.cern.ch/record/2724492>.
- 2604 [81] V.M. Abazov, B. Abbott, M. Abolins, et al. A novel method for modeling the
2605 recoil in W boson events at hadron colliders. *Nuclear Instruments and Methods
2606 in Physics Research Section A: Accelerators, Spectrometers, Detectors and Asso-
2607 ciated Equipment*, 609(2):250–262, 2009. ISSN 0168-9002. doi: <https://doi.org/10.1016/j.nima.2009.08.056>. URL <https://www.sciencedirect.com/science/article/pii/S0168900209016623>.
- 2608 [82] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair
2609 of light pseudoscalars in the final state with two b quarks and two τ leptons in
2610 proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 785:462, 2018. doi:
2611 10.1016/j.physletb.2018.08.057.
- 2612 [83] CMS LUMI POG. Luminosity Physics Object Group: Recommendations, 2024.
2613 URL <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.

- 2616 [84] The modeling of the top quark p_T (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting>.
- 2617
- 2618 [85] CMS BTV group. Methods to apply b-tagging efficiency scale fac-
2619 tors (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration>.
- 2620
- 2621 [86] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in
2622 pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. doi: 10.1088/1748-0221/
2623 12/02/P02014.
- 2624 [87] Garvita Agarwal. Jet Energy Scale and Resolution Measurements in CMS. *PoS*,
2625 ICHEP2022:652, 2022. doi: 10.22323/1.414.0652.
- 2626 [88] CMS JERC group. Jet Energy Corrections (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- 2627
- 2628 [89] CMS JERC group. Jet Energy Resolution (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- 2629
- 2630 [90] CMS MUO POG. Muon Physics Object Group: Baseline muon selec-
2631 tions for Run-II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2>.
- 2632
- 2633 [91] CMS ELE POG. Electron Identification Based on Simple Cuts, 2024. URL
2634 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaPublicData>.
- 2635
- 2636 [92] Jose Enrique Palencia Cortezon. Single top quark production at CMS. Technical
report, CERN, Geneva, 2018. URL <https://cds.cern.ch/record/2640578>.
- 2637
- 2638 [93] CMS Collaboration. Measurements of the electroweak diboson production cross
sections in proton-proton collisions at $\sqrt{s} = 5.02$ TeV using leptonic decays.
2639 *Phys. Rev. Lett.*, 127(19):191801, 2021. doi: 10.1103/PhysRevLett.127.191801.

- 2640 [94] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment
2641 ten years after the discovery. *Nature*, 607(7917):60–68, 2022. doi: 10.1038/
2642 s41586-022-04892-x.
- 2643 [95] CMS JERC group. Jet energy scale uncertainty sources (TWiki), 2024. URL
2644 <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources>.
- 2645 [96] TOP Systematic Uncertainties (Run 2) (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics>.
- 2647 [97] Kyle Cranmer. Practical Statistics for the LHC. In *2011 European School of
2648 High-Energy Physics*, pages 267–308, 2014. doi: 10.5170/CERN-2014-003.267.
- 2649 [98] Elham Khazaie, Maryam Zeinali, Hamed Bakhshiansohi, and Abideh Jafari.
2650 CMS AN-21-058 (internal): Search for exotic decays of the Higgs boson to a
2651 pair of new light bosons with two muons and two b jets in the final states at
2652 $\sqrt{s} = 13$ TeV, 2021.
- 2653 [99] CMS Higgs Physics Analysis Group. Summary of 2HDM+S searches at 13 TeV
2654 (Run 2), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Summary2HDMSRun2>.