

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

17 Open questions in particle physics may be addressed by the existence of an extended
18 Higgs sector beyond the Higgs boson with mass 125 GeV discovered at the Large
19 Hadron Collider (LHC) by the CMS and ATLAS experiments in 2012, may answer
20 remaining open questions in particle physics. Many properties of a potential extended
21 Higgs sector remain unconstrained by current measurements, making direct searches
22 of exotic Higgs decays are a powerful probe of new physics. We present a search
23 performed at the CMS experiment for exotic decays of a Higgs boson with mass 125
24 GeV to two light neutral pseudoscalars which respectively decay to two bottom quarks
25 and two tau leptons ($h \rightarrow aa \rightarrow bb\tau\tau$), which is combined with a different analysis
26 to the final state with two bottom quarks and two muons, with results interpreted
27 in Two Higgs Doublet Models extended with a singlet scalar (2HDM+S). We also
28 present preliminary work on a new search of 125 GeV Higgs boson decays to two
29 light pseudoscalars which can have different masses ($h \rightarrow a_1a_2$) with final states
30 containing bottom quarks and tau leptons. This “asymmetric” decay scenario has
31 not been studied to date at the CMS experiment and is predicted by Two Real
32 Singlet Models (TRSM).

³³

Acknowledgements

³⁴ Placeholder acknowledgements.

Contents

36	Abstract	iii
38	Acknowledgements	iv
39	List of Tables	xi
40	List of Figures	xiii
41	1 Introduction	1
42	1.1 History of the Standard Model	1
43	1.2 The Standard Model as a gauge theory	3
44	1.2.1 Gauge invariance	3
45	1.2.2 Local gauge symmetries	4
46	1.3 The Higgs Mechanism	6
47	1.4 Two-Higgs Doublet Models	8
48	1.5 Two Real Singlet Model	11
49	2 The Large Hadron Collider and the CMS Experiment	14
50	2.1 The Large Hadron Collider	14
51	2.2 Luminosity and pileup	16
52	2.3 The High-Luminosity LHC	17
53	2.4 The CMS Detector	18
54	2.5 Sub-detectors of CMS	20
55	2.5.1 Inner tracking system	20

56	2.5.2	ECAL	22
57	2.5.3	HCAL	23
58	2.5.4	Muon detectors	24
59	2.5.5	The Level-1 Trigger	26
60	2.5.6	The High-Level Trigger	29
61	2.5.7	Particle reconstruction	29
62	2.5.8	Data storage and computational infrastructure	30
63	3	The Phase-2 Upgrade of CMS	31
64	3.1	High-Luminosity LHC and CMS	31
65	3.2	The Phase-2 Level-1 Trigger	31
66	3.3	Standalone Barrel Calorimeter electron/photon reconstruction	34
67	3.3.1	Phase-2 geometry of the ECAL Barrel trigger	34
68	3.3.2	Phase-2 electron/photon reconstruction algorithm	35
69	4	Datasets and Monte Carlo samples	44
70	4.1	Datasets used	44
71	4.2	Monte Carlo samples	44
72	4.3	Embedded samples	46
73	5	Object reconstruction and corrections applied	49
74	5.1	Object reconstruction	49
75	5.1.1	Taus	49
76	5.1.2	Muons	53
77	5.1.3	Electrons	54
78	5.1.4	Jets	56
79	5.1.5	B-flavored jets	57
80	5.2	Reconstruction of the $\tau\tau$ mass	58
81	5.2.1	Original SVFit “standalone”: maximum likelihood	59

82	5.2.2 “Classic SVFit” with matrix element	60
83	5.2.3 FastMTT: optimized SVFit	61
84	5.3 Corrections applied to simulation	61
85	5.3.1 Tau energy scale	62
86	5.3.2 Muon energy scale	62
87	5.3.3 Electron energy scale	63
88	5.3.4 τ_h identification efficiency	64
89	5.3.5 Trigger efficiencies	64
90	5.3.6 Tau trigger efficiencies	64
91	5.3.7 Single muon trigger efficiencies	65
92	5.3.8 Single electron trigger efficiencies	66
93	5.3.9 $e\mu$ cross-trigger efficiencies	67
94	5.3.10 Electrons and muons faking τ_h : energy scales	68
95	5.3.11 Electrons and muons faking τ_h : misidentification efficiencies	69
96	5.3.12 Electron ID and tracking efficiency	70
97	5.3.13 Muon ID, isolation, and tracking efficiencies	71
98	5.3.14 Recoil corrections	72
99	5.3.15 Drell-Yan corrections	73
100	5.3.16 Pileup reweighing	73
101	5.3.17 Pre-firing corrections	74
102	5.3.18 Top p_T spectrum reweighing	75
103	5.3.19 B-tagging efficiency	75
104	5.3.20 Jet energy resolution and jet energy smearing	75
105	6 Event selection	77
106	6.1 General procedure for all channels	77
107	6.2 Event selection in the $\mu\tau_h$ channel	78
108	6.3 Event selection in the $e\tau_h$ channel	80

109	6.4	Event selection in the $e\mu$ channel	81
110	6.5	Extra lepton vetoes in all channels	82
111	7	Background estimation	86
112	7.1	Z+jets	86
113	7.2	W+jets	87
114	7.3	$t\bar{t}$ + jets	87
115	7.4	Single top	88
116	7.5	Diboson	88
117	7.6	Standard Model Higgs	88
118	7.7	Jet faking τ_h	89
119	7.8	QCD multijet background	90
120	8	Systematic uncertainties	92
121	8.1	Uncertainties associated with physics objects	93
122	8.1.1	Uncertainties in the lepton energy scales	93
123	8.1.2	Uncertainties from other lepton corrections	94
124	8.1.3	Uncertainties from jet energy scale and resolution	94
125	8.1.4	Uncertainties from b-tagging scale factors	95
126	8.1.5	Uncertainties from MET	96
127	8.2	Uncertainties associated with samples used	96
128	8.3	Other uncertainties	97
129	8.4	Pulls and impacts	97
130	9	Event categorization and signal extraction	99
131	9.1	B-tag jet multiplicity	99
132	9.2	DNN-based event categorization	99
133	9.3	Methodology for signal extraction	102
134	9.3.1	Model building and parameter estimation	103

135	9.3.2 Hypothesis testing	105
136	9.3.3 Confidence intervals	106
137	9.3.4 Profile likelihood ratio	108
138	9.3.5 Modified frequentist method: CL_S	109
139	10 Results	110
140	10.1 Results from $bb\tau\tau$	110
141	10.2 Combination with $bb\mu\mu$ final state	112
142	11 Asymmetric exotic Higgs decays	123
143	11.1 Signal masses	123
144	11.2 Cascade scenario signal studies	124
145	11.3 Current control plots for $\mu\tau_h$ channel	126
146	12 Conclusion and outlook	130

List of Tables

148	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.	48
152	5.1	Energy scales applied to genuine hadronic tau decays τ_h by data-taking year/era and decay mode, along with systematic errors.	62
154	5.2	Energy scales and systematic errors applied to genuine muons.	63
155	5.3	Energy scales and systematic errors applied to electrons in embedded samples by data-taking year/era.	64
157	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	64
159	5.5	Energy scales and up/down systematic uncertainties applied to electrons misidentified as hadronic taus.	68
161	5.6	Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. muons in 2018.	70
163	5.7	Tau mis-identification efficiency for the DeepTau Tight and Very Loose (VLoose) working points vs. electrons in 2018.	70

165	6.1	Trigger thresholds used for the leptons in the $bb\mu\mu$ analysis and the 166 $bb\tau\tau$ analysis (the focus of this work). The thresholds for the three $bb\tau\tau$ 167 channels ($e\mu$, $e\tau_h$, and $\mu\tau_h$) are listed separately, with some channels 168 and years taking the logical OR of two triggers with different thresholds.	78
169	6.2	High-Level Trigger (HLT) paths used to select data and simulation 170 events in 2016 for the three $\tau\tau$ channels.	83
171	6.3	High-Level Trigger (HLT) paths used to select data and simulation 172 events in 2017 for the three $\tau\tau$ channels.	83
173	6.4	High-Level Trigger (HLT) paths used to select data and simulation 174 events in 2018 for the three $\tau\tau$ channels. In 2018 a HLT trigger path 175 using the hadron plus strips (HPS) tau reconstruction algorithm be- 176 came available.	84
177	9.1	Event categorization based on DNN scores for events with exactly 1 178 b-tag jet (1bNN), for the three $\tau\tau$ channels and three eras.	102
179	9.2	Event categorization based on DNN scores for events with 2 b-tag jets 180 (2bNN), for the three $\tau\tau$ channels and three eras.	103

¹⁸¹ List of Figures

¹⁸² 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.	2
¹⁸⁷ 1.2	An illustration of the Higgs potential.	7
¹⁸⁸ 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).	10
¹⁹¹ 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. . . .	13
¹⁹³ 2.1	Aerial view of the Large Hadron Collider (LHC).	15
¹⁹⁴ 2.2	Distribution of the mean number of inelastic collisions per bunch cross- ing (pileup) in data, for proton-proton collisions in 2016-2018	18
¹⁹⁶ 2.3	Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector.	19
¹⁹⁸ 2.4	Cross section of the current Phase-1 CMS tracker.	21
¹⁹⁹ 2.5	Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters.	23

201	2.6	Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels.	25
202	2.7	Dataflow for the Phase-1 Level-1 Trigger.	27
203	3.1	Functional diagram of the CMS L1 Phase-2 upgraded trigger design. .	32
204	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.	35
205	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.	38
206	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.	39
207	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.	40

225	3.6 Illustration of an example electron/photon (e/γ) cluster in the Phase-	
226	2 Level-1 Trigger standalone barrel e/γ reconstruction, in a region of	
227	15×20 crystals (3×4 towers). Each small pink square is one crystal,	
228	the highest-granularity ECAL trigger primitives available to the L1	
229	Trigger in Phase-2. The core cluster consists of the energy sum in a	
230	3×5 window of crystals, (<i>shaded light blue</i>) centered around the seed	
231	crystal (<i>red</i>). Bremsstrahlung corrections are checked in the adjacent	
232	3×5 windows in the ϕ direction (<i>shaded light yellow</i>). The relative	
233	energies in windows of size 2×5 and 5×5 in crystals (<i>dashed dark blue</i>	
234	<i>and dark red</i>) are used to compute shower shape variables to identify	
235	true e/γ objects. Lastly, an isolation sum is computed in a window of	
236	size 7×7 in towers (not shown in figure).	41
237	3.7 Efficiency of the standalone barrel e/γ reconstruction, as a function of	
238	the true electron's transverse momentum p_T	42
239	3.8 Rates of the standalone barrel e/γ reconstruction measured as a func-	
240	tion of the minimum energy (E_T) required of the reconstructed e/γ	
241	object in each event.	43
242	4.1 Cumulative delivered and recorded luminosity versus time for 2015-	
243	2018 at CMS, in proton-proton collision data only, at nominal center-	
244	of-mass energy.	45
245	4.2 Schematic view of the four main steps of the embedding technique for	
246	τ leptons.	47
247	5.1 Distributions of $m_{\tau\tau}$ reconstructed by the classic SVFit algorithm, and	
248	masses of visible tau decay products (before SVFit).	60
249	5.2 Electron/photon energy scale factors and uncertainties for 2018. . . .	63

250	5.3 Hadronic tau leg efficiency of the cross-triggers for $\mu\tau_h$ (<i>left</i>) and $e\tau_h$ 251 (<i>right</i>) triggers as a function of offline tau p_T for 2016, 2017, and 2018.	66
252	5.4 Trigger efficiencies in data (<i>top panels</i>) and ratio of efficiencies af- 253 ter/before a HLT muon reconstruction update (<i>bottom panels</i>) for the 254 muon in the isolated single muon trigger with threshold $p_T > 24$ GeV 255 in the data-taking year 2018, as functions of the muon p_T (<i>left</i>) and 256 muon $ \eta $ (<i>right</i>).	67
257	5.5 Trigger efficiencies in data and the data/MC ratio for the electron in 258 the single electron trigger with threshold $p_T > 32$ GeV in the data- 259 taking year 2018, as functions of the electron p_T (<i>left</i>) and electron $ \eta $ 260 (<i>right</i>).	68
261	5.6 Efficiencies of the electron leg vs. p_T (<i>left</i>) and the muon log vs. η 262 (<i>right</i>), for the HLT path with online thresholds of 12 GeV for the 263 electron and 23 GeV for the muon, with the data-taking years 2016 264 through 2018 overlaid.	69
265	5.7 Efficiencies in data (<i>top panels</i>) and the ratio of efficiencies in data/MC 266 (<i>bottom panels</i>), for the electron multivariate analysis (MVA) identifi- 267 cation (<i>left</i>) and for the Gaussian-sum filter (GSF) tracking (<i>right</i>). . .	71
268	5.8 Muon identification efficiencies in 2015 data and MC as a function of 269 the muon p_T for the loose ID (<i>left</i>) and tight ID (<i>right</i>) working points.	72
270	5.9 Muon isolation efficiencies in Run-2 data as a function of the muon p_T 271 (<i>left</i>) and $ \eta $ (<i>right</i>).	73
272	5.10 Muon tracking efficiencies as a function of $ \eta $ for standalone muons in 273 Run-2 data (<i>black</i>) and Drell-Yan (<i>blue</i>) MC simulation.	74
274	7.1 Leading-order Feynman diagrams of Higgs production.	89
275	8.1 Top sixty impacts for the combination of all channels and years. . . .	98

276	9.1 Schematic of the Neyman construction for confidence intervals.	107
277	10.1 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $\mu\tau_h$ chan-	
278	nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions.	114
279	10.2 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\tau_h$ chan-	
280	nel, for the 1 b-tag jet and 2 b-tag jet signal and control regions.	115
281	10.3 Postfit final observed and expected $m_{\tau\tau}$ distributions in the $e\mu$ channel.	116
282	10.4 Observed 95% CL exclusion limits (<i>black, solid lines</i>) and expected 95%	
283	CL and 68% CL limits (<i>shaded yellow and green</i>) on the branching	
284	fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard	
285	Model production for the 125 GeV Higgs (h). Limits are shown for the	
286	$\mu\tau_h$ channel (<i>top left</i>), the $e\tau_h$ channel (<i>top right</i>), and the $e\mu$ channel	
287	(<i>bottom left</i>), and lastly the combination of all three channels (<i>bottom</i>	
288	<i>right</i>) The dataset corresponds to 138 fb^{-1} of data collected in the	
289	years 2016-2018 at a center-of-mass energy 13 TeV.	117
290	10.5 Observed 95% CL upper limits on $B(h \rightarrow aa)$ in %, for the $bb\tau\tau$ final	
291	state (<i>left</i>) and $bb\mu\mu$ final state (<i>right</i>) using the full Run 2 integrated	
292	luminosity of 138 fb^{-1} in 2HDM+S type I (<i>blue</i>), type II with $\tan\beta =$	
293	2.0 (<i>orange dashed</i>), type III with $\tan\beta = 2.0$ (<i>dotted green</i>), and type	
294	IV with $\tan\beta = 0.6$ (<i>red dashed</i>). .	118
295	10.6 Observed 95% CL upper limits on the branching fraction of the 125	
296	GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$, in percentages,	
297	as a function of the pseudoscalar mass m_a , in 2HDM+S type I (<i>blue</i>),	
298	type II with $\tan\beta = 2.0$ (<i>orange dashed</i>), type III with $\tan\beta = 2.0$	
299	(<i>dotted green</i>), and type IV with $\tan\beta = 0.6$ (<i>red dashed</i>), for the	
300	combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated	
301	luminosity of 138 fb^{-1} .	119

302	10.7 Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination	
303	of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity	
304	of 138 fb^{-1} for Type II (<i>left</i>), Type III (<i>middle</i>), and Type IV (<i>right</i>)	
305	2HDM+S in the $\tan\beta$ vs. m_a phase space.	120
306	10.8 Summary plot of current observed and expected 95% CL limits on the	
307	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
308	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
309	$B(h \rightarrow aa)$, in the 2HDM+S type I scenario, obtained at CMS with	
310	data collected at 13 TeV.	120
311	10.9 Summary plot of current observed and expected 95% CL limits on the	
312	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
313	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
314	$B(h \rightarrow aa)$, in the 2HDM+S type II scenario with $\tan\beta = 2.0$, ob-	
315	tained at CMS with data collected at 13 TeV.	121
316	10.10 Summary plot of current observed and expected 95% CL limits on the	
317	branching ratio of the 125 GeV Higgs boson to two pseudoscalars, nor-	
318	malized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times$	
319	$B(h \rightarrow aa)$, in the 2HDM+S type III scenario with $\tan\beta = 2.0$, ob-	
320	tained at CMS with data collected at 13 TeV.	122
321	11.1 Generator-level b-flavor jet transverse momenta p_T , for $h \rightarrow a_1 a_2$ cas-	
322	cade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) =$	
323	$(100, 15) \text{ GeV}$ (<i>left</i>) and $(40, 20) \text{ GeV}$ (<i>right</i>). In each plot the generator-	
324	level p_T of the leading (<i>black</i>), sub-leading (<i>red</i>), third (<i>blue</i>), and	
325	fourth (<i>light green</i>) are overlaid.	125

326	11.2 Distributions (arbitrary units) of transverse momentum p_T resolution	
327	and ΔR between the two closest generator-level b jets, treated as one	
328	object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow$	
329	$a_1 a_2$ mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV (<i>top left, top right</i>)	
330	and $(40, 20)$ GeV (<i>bottom left, bottom right</i>) in the ggH production of	
331	the 125 GeV h . In the $(40, 20)$ GeV mass point, the longer p_T resolution	
332	tail (<i>bottom left</i>) indicates that the reconstructed jet underestimates	
333	the generator b-flavor jets' energy, and the significant fraction of events	
334	with larger ΔR values (<i>bottom right</i>) indicate worse matching.	126
335	11.3 Kinematic properties of the leading muon and τ_h in the $\mu\tau_h$ channel: p_T	
336	(<i>top row</i>), η (<i>second row</i>), and ϕ (<i>third row</i>). The visible 4-momenta	
337	of the muon and τ_h are summed, giving the visible di-tau mass m_{vis}	
338	and transverse momentum $p_{T,\text{vis}}$. The errors shown in the figures only	
339	include statistical errors and only several of the full set of systematic	
340	errors (only those associated with the lepton energy scales and τ_h iden-	
341	tification efficiency).	128
342	11.4 Kinematic properties of the leading and sub-leading b-tag jets in the	
343	$\mu\tau_h$ final state: jet p_T (<i>top row</i>), η (<i>second row</i>), ϕ (<i>third row</i>), as well	
344	as the missing transverse energy magnitude and azimuthal direction	
345	(<i>bottom row</i>). The errors shown in the figures only include statistical	
346	errors and only several of the full set of systematic errors (only those	
347	associated with the lepton energy scales and τ_h identification efficiency).	129

³⁴⁸ Chapter 1

³⁴⁹ Introduction

³⁵⁰ The Standard Model is the current prevailing theoretical framework that encompasses
³⁵¹ all known elementary particles to date and describes their interactions, yet falls short
³⁵² of describing open problems in physics. Here, we introduce the Standard Model (Sec-
³⁵³ tion 1.1) and provide a mathematical motivation of the SM a gauge theory (Section
³⁵⁴ 1.2). We introduce the Higgs mechanism (Section 1.3), and outline two groups of
³⁵⁵ theoretical extensions to the Standard Model that feature extended Higgs sectors
³⁵⁶ (Sections 1.4 and 1.5).

³⁵⁷ 1.1 History of the Standard Model

³⁵⁸ The building blocks of our modern-day understanding of particle physics were estab-
³⁵⁹ lished over the course of decades by experimental discoveries and theoretical advances,
³⁶⁰ culminating in the development of a theoretical framework known as the Standard
³⁶¹ Model (SM). In the 1880s, the electron was the first subatomic particle to be iden-
³⁶² tified, through measurements of particles produced by ionizing gas. By the 1930s,
³⁶³ atoms were known to consist mostly of empty space, with protons and neutrons con-
³⁶⁴ centrated at the center and orbited by electrons. Spurred by advances in particle
³⁶⁵ accelerator technology, the experimental discoveries of the positron, the muon, and

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

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

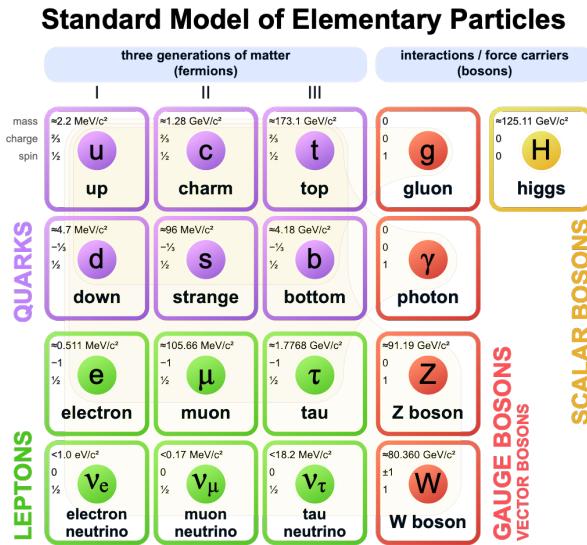


Figure 1.1: Table of Standard Model particles showing the grouping of the fermions into three generations of matter and the bosons, responsible for carrying the three fundamental forces in the Standard Model. The masses, charges, and spins of the particles are shown. The antimatter counterparts of the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

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

380 and the weak nuclear force.

381 1.2 The Standard Model as a gauge theory

382 1.2.1 Gauge invariance

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

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

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

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

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

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

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

407 1.2.2 Local gauge symmetries

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

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

411 can be verified to *not* be a solution to the Dirac equation for free particles: $(i\gamma^\mu \partial_\mu -$
 412 $m)\Psi(x) = 0$. This necessitates a modified Dirac equation, where the derivative takes
 413 into account that the vector field $V(x)$ needs to be compared at two displaced space-
 414 time points in a curvilinear coordinate system:

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

415 We define a covariant derivative,

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

⁴¹⁶ where $A_\mu(x)$ is a 4-vector potential. Thus the modified Dirac equation reads:

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

⁴¹⁷ The simultaneous gauge transformation $A'_\mu(x) = A_\mu(x) - \partial_\mu \chi(x)$ and wavefunction
⁴¹⁸ transformation $\Psi'(x) = \exp(ie\chi(x))\Psi(x)$ leaves the covariant-derivative form of the
⁴¹⁹ Dirac equation (Eqn 1.1) invariant.

⁴²⁰ The generalization of this result is as follows: if a theory is invariant for unitary
⁴²¹ transformations U of the particle states according to

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

⁴²² One must define a derivative of the form

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

⁴²³ to keep the theory invariant under Eqn. 1.7. The four-potential B^μ represents the
⁴²⁴ interacting four-potential which must be added to keep the theory invariant.

⁴²⁵ In the case of the Standard Model, the theory is built around the gauge trans-
⁴²⁶ formations $G = SU(3) \times SU(2) \times U(1)$. $SU(3)$ is associated to the strong force
⁴²⁷ (subscripted C); $SU(2)$ is associated to the weak force (subscripted L); and $U(1)$ is
⁴²⁸ hypercharge (subscripted Y). The gauge-covariant derivative is

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

⁴²⁹ • In the $U(1)_Y$ term, B_μ is the weak hypercharge field.

⁴³⁰ • 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-
⁴³¹ potentials. $\tau/2$ are the Pauli matrices, generators of the $SU(2)$ transformation.

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

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

436 1.3 The Higgs Mechanism

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

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

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

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

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

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

458 The minimal self-interacting Higgs potential that is invariant under $SU(2)_L \times$
459 $U(1)_Y$ is given by

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

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

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

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

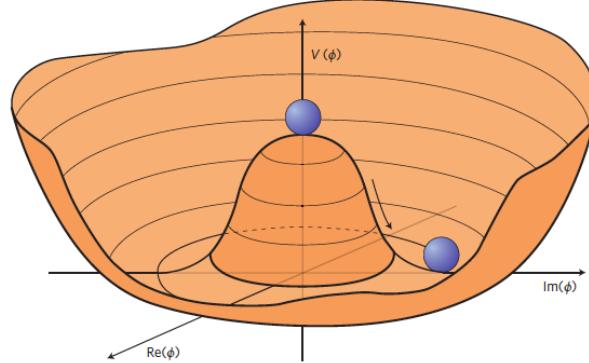


Figure 1.2: An illustration of the Higgs potential [3]. Choosing any of the points at the bottom of the potential breaks spontaneously the rotational $U(1)$ symmetry.

⁴⁶⁴ The excitations of the Higgs field with respect to the minimum Φ_{\min} are parame-
⁴⁶⁵ terized by

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

⁴⁶⁶ Three degrees of freedom are coupled directly to the electroweak gauge bosons; this
⁴⁶⁷ is often referred to as the gauge bosons “eating” the Goldstone bosons to form the
⁴⁶⁸ longitudinal polarizations of the massive spin-1 boson states. The $H(x)$ excitation is
⁴⁶⁹ in the radial direction and corresponds to the free particle state of the Higgs boson.

⁴⁷⁰ 1.4 Two-Higgs Doublet Models

⁴⁷¹ One of the simplest possible extensions to the Standard Model is adding a doublet
⁴⁷² to the minimal Higgs sector of the Standard Model, which is a $SU(2)_L$ doublet H
⁴⁷³ with hypercharge $Y = +\frac{1}{2}$, denoted here as $H \sim 2_{+1/2}$. These extensions are found
⁴⁷⁴ in several theories such as supersymmetry. A general 2HDM can be extended with a
⁴⁷⁵ light scalar (2HDM+S) to obtain a rich set of exotic Higgs decays [4].

The charges of the Higgs fields are chosen to be $H_1 \sim 2_{-1/2}$ and $H_2 \sim 2_{+1/2}$, which acquire vacuum expectation values $v_{1,2}$ which are assumed to be real and aligned [4]. Expanding about the minima yields two complex and four real degrees of freedom:

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

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

⁴⁷⁶ The charged scalar and pseudoscalar mass matrices are diagonalized by a rotation
⁴⁷⁷ angle β , defined as $\tan \beta = v_2/v_1$. One charged (complex) field and one neutral
⁴⁷⁸ pseudoscalar combination of $H_{1,2,I}^0$ are eaten by the SM gauge bosons after electroweak

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

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

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

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

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

489 2HDM can be extended by a scalar singlet (2HDM+S) [4]:

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

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

497 $m_a < m_h - m_Z \sim 35$ GeV, the exotic Higgs decay $h \rightarrow Za$ is possible, and for
 498 $m_a < m_h/2 \approx 63$ GeV, the exotic Higgs decay $h \rightarrow aa$ is possible.

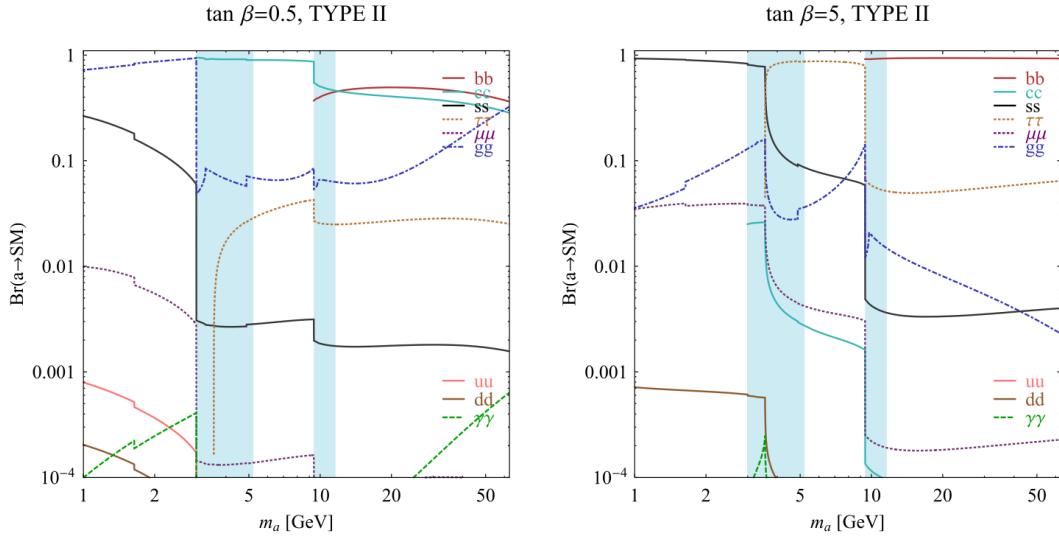


FIG. 7 (color online). Branching ratios of a singletlike pseudoscalar in the 2HDM + S for type-II Yukawa couplings. Decays to quarkonia likely invalidate our simple calculations in the shaded regions.

Figure 1.3: Branching ratios of a singlet-like pseudoscalar in Type II 2HDM+S for $\tan\beta = 0.5$ (*left*) and $\tan\beta = 5$ (*right*) from [4], showing the dependence of the branching ratios on $\tan\beta$, as well as the prominence of the branching ratios to bb and $\tau\tau$, the channels searched for in the analysis presented here.

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

507 1.5 Two Real Singlet Model

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

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

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

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

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

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

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

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

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

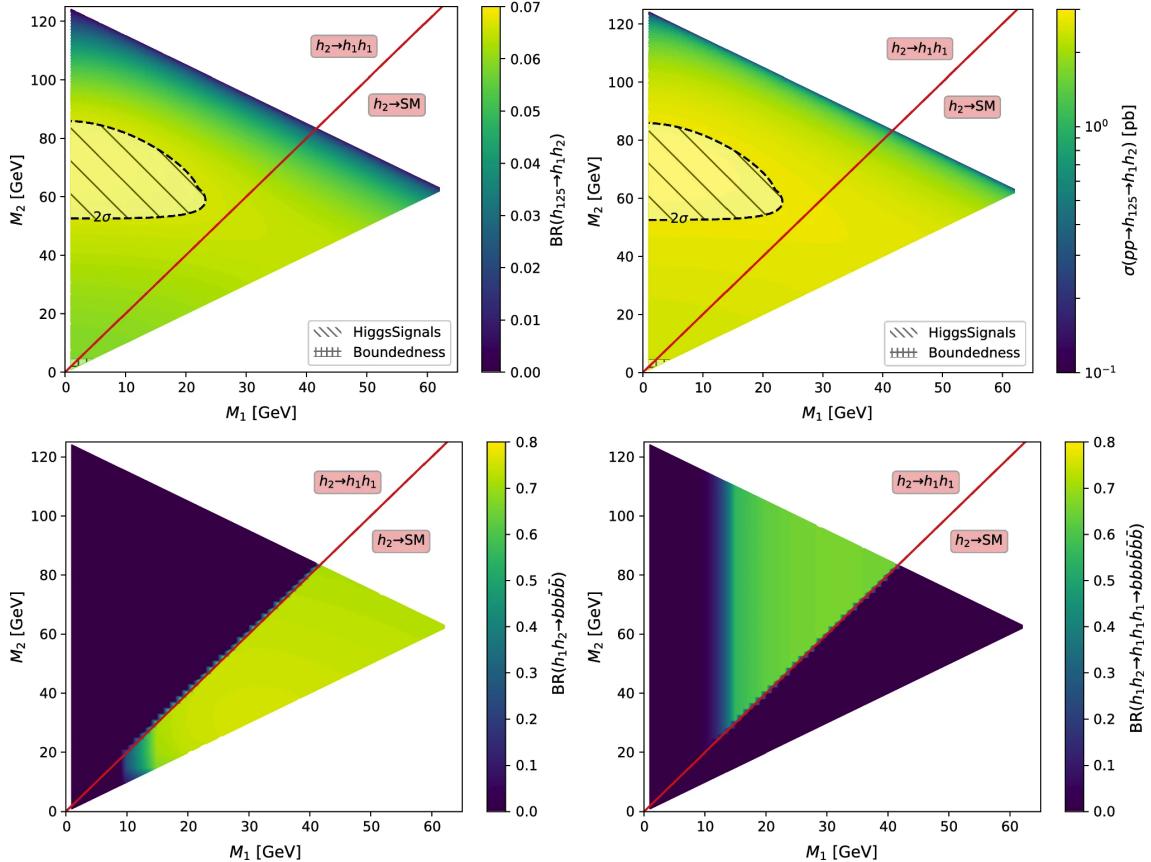


Figure 1.4: Benchmark plane BP1 for benchmark scenario 1 from [5], for the decay signature $h_{125} \rightarrow h_1 h_2$ with $h_{125} \equiv h_3$, defined in the (M_1, M_2) plane. The color code shows $\text{BR}(h_3 \rightarrow h_1 h_2)$ (*top left*) and the 13 TeV LHC signal rate for $pp \rightarrow h_3 \rightarrow h_1 h_2$ (*top right*). The red line separates the region $M_2 > 2M_1$, where $\text{BR}(h_2 \rightarrow h_1 h_1) \sim 100\%$, from the region $M_2 < 2M_1$, where $\text{BR}(h_2 \rightarrow F_{SM}) \sim 100\%$. The *bottom left* and *right* show the branching ratio of the $h_1 h_2$ into (respectively) $b\bar{b}b\bar{b}$, and through a $h_2 \rightarrow h_1 h_1$ cascade to $b\bar{b}b\bar{b}b\bar{b}$. The hatched region indicates where the decay rate slightly exceeds the 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.

546 **Chapter 2**

547 **The Large Hadron Collider and the**
548 **CMS Experiment**

549 **2.1 The Large Hadron Collider**

550 The CERN Large Hadron Collider (LHC) is an accelerator complex consisting of a
551 27-kilometer ring of superconducting magnets with accelerating structures to boost
552 the energy of particles, which collide at a center-of-mass energy of up to 14 TeV. The
553 beams inside the LHC are made to collide at four locations around the accelerator
554 ring, at the locations of four particle detectors: ATLAS, CMS, ALICE, and LHCb.
555 An aerial view of the four major experiments' locations is shown in Fig. 2.1 [6]. AT-
556 LAS and CMS are the two general-purpose detectors with broad physics programmes
557 spanning Standard Model measurements and searches for signatures of new physics
558 [7] [8]. The two experiments use different technical solutions and different magnet
559 system designs. ALICE is a general-purpose detector dedicated to measuring LHC
560 heavy-ion collisions, and is designed to address the physics of strongly interacting
561 matter, and the properties of quark-gluon plasma [9]. The LHCb experiment special-
562 izes in investigating CP violation through measuring the differences in matter and

563 antimatter, by using a series of subdetectors to detect mainly forward particles close
564 to the beam direction [10].

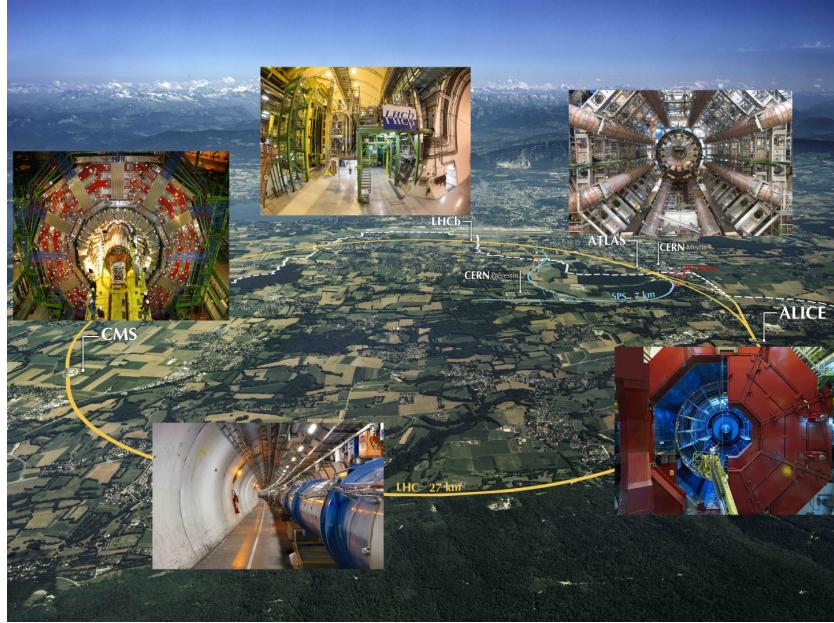


Figure 2.1: Aerial view of the Large Hadron Collider (LHC) spanning the border of France and Switzerland, and the four major experiments located around the ring: CMS (Compact Muon Solenoid), LHCb (LHC beauty), ATLAS (A Toroidal LHC Apparatus), and ALICE (A Large Ion Collider Experiment). [6]

565 The accelerator complex at CERN is a succession of machines that accelerate
566 particles in stages until they reach their final energy of 6.5 TeV per beam [11] [12].
567 In Linear accelerator 4 (Linac4), negative hydrogen ions (hydrogen atoms with an
568 additional electron) are accelerated to 160 MeV, and stripped of their two electrons,
569 leaving only protons, before entering the Proton Synchrotron Booster (PSB). These
570 protons are accelerated to 2 GeV, then to 26 GeV in the Proton Synchrotron (PS),
571 and 450 GeV in the Super Proton Synchrotron (SPS). The protons are transferred to
572 the two beam pipes of the LHC, where one beam circulates clockwise and the other
573 counterclockwise. Each LHC ring takes 4 minutes and 20 seconds to fill, and it takes
574 about 20 minutes for the protons to reach their maximum energy. During normal
575 operating conditions, beams circulate for many hours inside the LHC ring.

576 2.2 Luminosity and pileup

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

$$578 N_{event} = \mathcal{L} \cdot \sigma_{event} \quad (2.1)$$

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

$$581 \mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (2.2)$$

581 where the parameters are as defined, along with some example typical nominal values
582 in Phase-1 of the LHC [13] [14]:

- 583 • N_b is the number of particles per bunch ($N_b \approx 1.15 \times 10^{11}$ protons per bunch)
- 584 • n_b is the number of bunches per beam (maximum 2808),
- 585 • f_{rev} is the revolution frequency (≈ 11 kHz),
- 586 • γ_r is the relativistic gamma factor,
- 587 • ϵ_n is the normalized transverse beam emittance (area in a transverse plane
588 occupied by the beam particles),
- 589 • β^* is the beta function at the collision point ($\beta^* = 0.55$ m),
- 590 • and F is the geometric luminosity reduction factor due to the crossing angle at
591 the interaction points ($F \approx 0.84$ for Phase-1. Note that complete overlap would
592 give $F = 1$).

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

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

600 The LHC’s nominal beam luminosities are sufficiently large for multiple proton-
601 proton collisions to occur in the same time window of 25 nanoseconds in which proton
602 bunches collide [15]. These multiple collisions will lead to particle interactions over-
603 lapping in the detector. To measure a proton-proton collision, the single collision
604 must be separated from overlapping collisions, which are called “pileup” collisions. A
605 distribution of pileup in the data-taking years 2016-2018 is shown in Fig. 2.2. The
606 pileup is defined as the average number of pp collisions per bunch crossing.

607 CMS reports an inelastic pp cross section of $\sigma_{\text{inel}} = 68.6$ millibarns at a center-of-
608 mass energy of $\sqrt{s} = 13$ TeV [16], which can be used to estimate pileup as follows:

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

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

610 Thus, higher luminosities create more intense pileup conditions, posing a greater
611 challenge to detector performance and particle reconstruction and identification.

612 2.3 The High-Luminosity LHC

613 The High-Luminosity LHC (HL-LHC) is a major upgrade of the LHC scheduled
614 to take place in the late 2020s, that will increase the instantaneous luminosity by
615 a factor of five beyond the original design value, and the integrated luminosity
616 by a factor of ten [17]. This will be accomplished through accelerator technological
617 advances: for instance, reduction of the interaction point β^* from 0.55 m down to 0.15

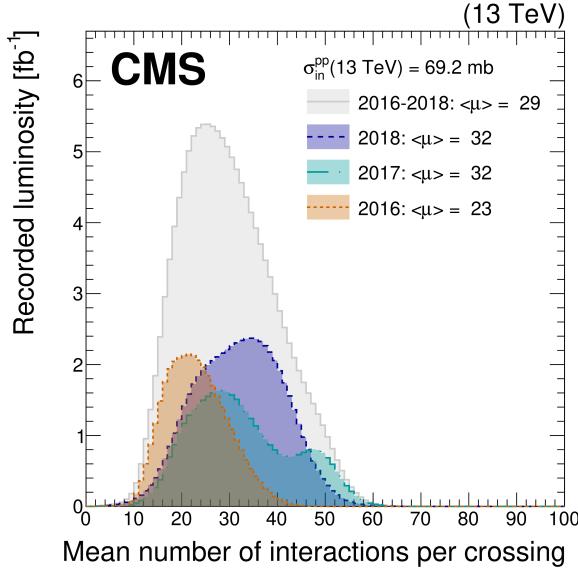


Figure 2.2: Distribution of the mean number of inelastic collisions per bunch crossing (pileup) in data [15], for proton-proton collisions in 2016 (*dotted orange*), 2017 (*dotted light blue*), 2018 (*dotted dark blue*), and integrated over 2016-2018 (*solid grey*). A cross-section of inelastic proton-proton 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 [17].

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

622 2.4 The CMS Detector

623 The Compact Muon Solenoid (CMS) experiment was conceived to study proton-
 624 proton and lead-lead collisions at a center-of-mass energy of 14 TeV (5.5 TeV nucleon-
 625 nucleon) and at luminosities up to $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ ($10^{27} \text{ cm}^{-2} \text{ s}^{-1}$) [18] [19]. Starting
 626 from the beam interaction region at the center of the CMS detector, particles first
 627 pass through a silicon pixel and strip tracker, in which charged-particle trajectories
 628 (tracks) and origins (vertices) are reconstructed from signals (hits) in the sensitive

layers. The tracker is immersed in a high-magnetic-field superconducting solenoid that bends the trajectories of charged particles, allowing the measurement of their electric charge and momenta. Electrons and photons are then absorbed in an electromagnetic calorimeter (ECAL) comprised of lead-tungstate scintillating-crystals. The corresponding electromagnetic showers are detected as clusters of energy recording in neighboring cells, from which the direction and energy of the particles can be determined. Charged and neutral hadrons may initiate a hadronic shower in the ECAL as well, which is then fully absorbed in the hadron calorimeter (HCAL). The resulting clusters are used to estimate their direction and energies. Muons and neutrinos pass through the calorimeters with little to no interactions. Neutrinos escaped undetected; muons produce hits in additional gas-ionization chamber muon detectors housed in the iron yoke of the flux-return. A sketch of example particle interactions in a transverse slice of the CMS detector is shown in Fig. 2.3. The collision data is recorded with the use of the Level-1 (L1) trigger (discussed separately in 2.5.5), high-level trigger (HLT), and data acquisition systems ensuring high efficiency in selecting physics events of interest.

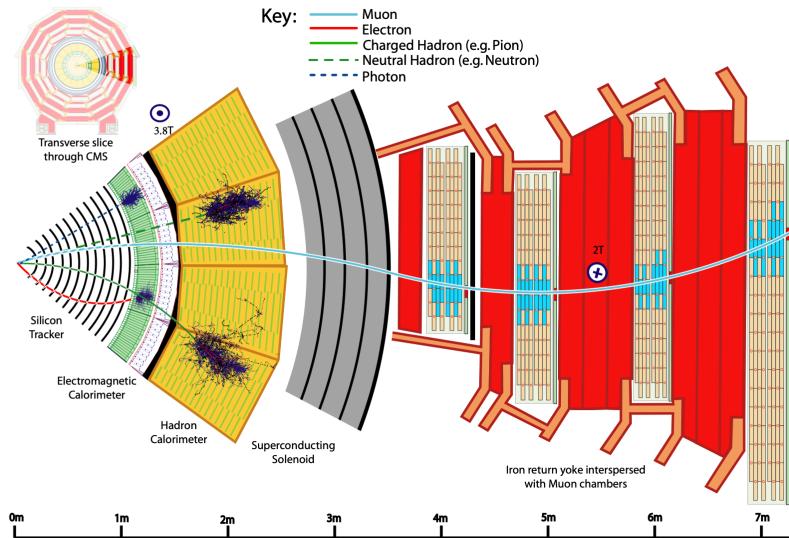


Figure 2.3: Sketch of particle trajectories of muons, electrons, charged and neutral hadrons, and photons in a transverse cross-section of the CMS detector [19].

645 CMS uses a right-handed coordinate system [18]. The origin is centered at the
646 nominal collision point inside the experiment. The x axis points towards the center
647 of the LHC, and the y axis points vertically upwards. The z axis points along the
648 beam direction. The azimuthal angle, ϕ , is measured from the x axis in the x - y
649 plane, and the radial coordinate in this plane is denoted by r . The polar angle, θ ,
650 is measured from the z axis. The pseudorapidity, η , is defined as $\eta = -\ln \tan(\theta/2)$.
651 The momentum and energy transverse to the beam direction, denoted by p_T and E_T
652 respectively, are computed from the x and y components. The momentum imbalance
653 in the transverse plane is called the missing transverse momentum, and its magnitude
654 is denoted by E_T^{miss} .

655 2.5 Sub-detectors of CMS

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

658 2.5.1 Inner tracking system

659 The CMS Tracker performs robust tracking and detailed vertex reconstruction in the
660 4 T magnetic field of the superconducting solenoidal magnet. The primary sensors
661 used in the tracker are p^+ on n -bulk devices, which allow high voltage operation and
662 are radiation-resistant [20] [21]. The active envelope of the CMS Tracker extends
663 to a radius of 115 cm, over a length of approximately 270 cm on each side of the
664 interaction point [20]. Charged particles in the region $|\eta| \lesssim 1.6$ benefit from the full
665 momentum measurement precision. In this region, a charged particle with p_T of 1000
666 GeV has a sagitta of $\sim 195 \mu\text{m}$. The Tracker acceptance extends further to $|\eta| = 2.5$,
667 with a reduced radius of approximately 50 cm.

668 The high magnetic field of CMS causes low p_T charged particles to travel in helical

669 trajectories with small radii. The majority of events contain particles with a steeply
 670 falling p_T spectrum, resulting in a track density which rapidly decreases at higher
 671 radii.

672 A schematic view of the current Phase-1 CMS tracker [22], including the pixel
 673 detector, is shown in Fig. 2.4. The Phase-1 pixel detector consists of three barrel
 674 layers (BPIX) at radii of 4.4 cm, 7.3 cm, and 10.2 cm, and two forward/backward disks
 675 (FPIX) at longitudinal positions of ± 34.5 cm and ± 46.5 cm, and extending in radius
 676 from about 6 cm to 15 cm. These pixelated detectors produce 3D measurements along
 677 the paths of charged particles with single hit resolutions between 10-20 μm .

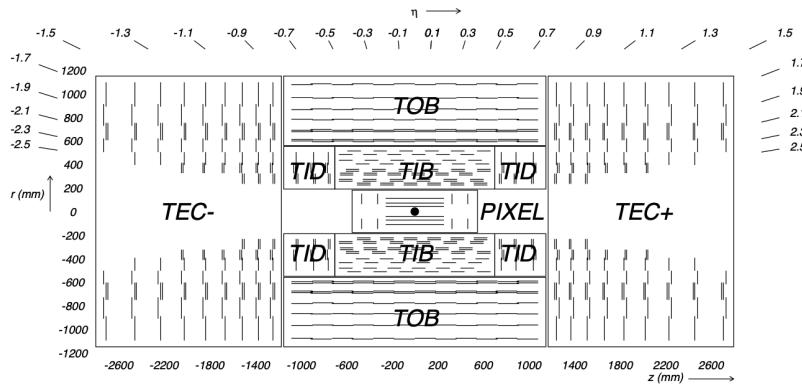


Figure 2.4: Cross section of the current Phase-1 CMS tracker [22]. Each line represents a detector module. Double lines indicate back-to-back modules which deliver two-dimensional (stereo) hits in the strip tracker.

678 After the pixel and on their way out of the tracker, particles pass through the
 679 silicon strip tracker which reaches out to a radius of 130 cm (Fig. 2.4). The sensor
 680 elements in the strip tracker are single-sided $p\text{-on-}n$ type silicon micro-strip sensors
 681 [18]. The silicon strip detector consists of four inner barrel (TIB) layers assembled
 682 in shells, with two inner endcaps (TID), each composed of three small discs. The
 683 outer barrel (TOB) consists of six concentric layers. Two endcaps (TEC) close off
 684 the tracker on either end.

685 **2.5.2 ECAL**

686 The electromagnetic calorimeter (ECAL) of CMS measures electromagnetic energy
687 deposits with high granularity. One of the driving criteria in the design was the ca-
688 pability of detecting the Standard Model Higgs boson decay to two photons (in fact,
689 the channel in which the 125 GeV Higgs boson was discovered at CMS). ECAL is
690 a hermetic homogeneous calorimeter comprised of 61,200 lead tungstate (PbWO_4)
691 crystals mounted in the central barrel, with 7,324 crystals in each of the two endcaps
692 [18]. A preshower detector is located in front of the endcap crystals. Avalanche pho-
693 todiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes
694 (VPTs) in the endcaps.

695 The design of the ECAL is driven by the behaviour of high-energy electrons, which
696 predominantly lose energy in matter via bremsstrahlung, and high-energy photons
697 by e^+e^- pair production. The characteristic amount of matter traversed for these
698 interactions is the radiation length X^0 , usually measured in units of g cm^{-2} . The
699 radiation length is also the mean distance over which a high-energy electron loses all
700 but $1/e$ of its energy via bremsstrahlung [23]. Thus high granularity in η and ϕ , and
701 the length of the ECAL crystals, is designed to capture the shower of e/γ produced
702 by electrons and photons.

703 The barrel part of the ECAL (EB) covers the pseudorapidity range $|\eta| < 1.479$
704 [18]. The barrel granularity is 360-fold in ϕ and (2×85) -fold in η . The crystal cross-
705 section corresponds to approximately 0.0174×0.0174 in $\eta - \phi$ or $22 \times 22 \text{ mm}^2$ at the
706 front face of the crystal, and $26 \times 26 \text{ mm}^2$ at the rear face. The crystal length is 230
707 mm, corresponding to $25.8 X_0$.

708 The ECAL read-out acquires the signals of the photodetectors [18]. At each bunch
709 crossing, digital sums representing the energy deposit in a trigger tower, comprising
710 5×5 crystals in $\eta \times \phi$, are generated and sent to the Level-1 trigger system (detailed
711 in Section 2.5.5).

712 2.5.3 HCAL

713 The hadronic calorimeter (HCAL) of CMS measures hadronic energy, which is key to
714 characterizing the presence of apparent missing transverse energy which could arise
715 from hadron jets and neutrinos or exotic particles [18]. A schematic of the components
716 of HCAL are shown in Fig. 2.5. The HCAL barrel (HB) and endcaps (HE) are located
717 outside of the tracker and the ECAL, spanning a radius of 1.77 m (outer extent of
718 ECAL) up to 2.95 m (inner extent of the magnet coil). An outer hadron calorimeter
719 (HO) is placed outside the solenoid to complement the barrel calorimeter. Beyond
720 $|\eta| = 3$, the forward hadron calorimeter (HF) at 11.2 m from the interaction point
721 extend the pseudorapidity coverage to $|\eta| = 5.2$.

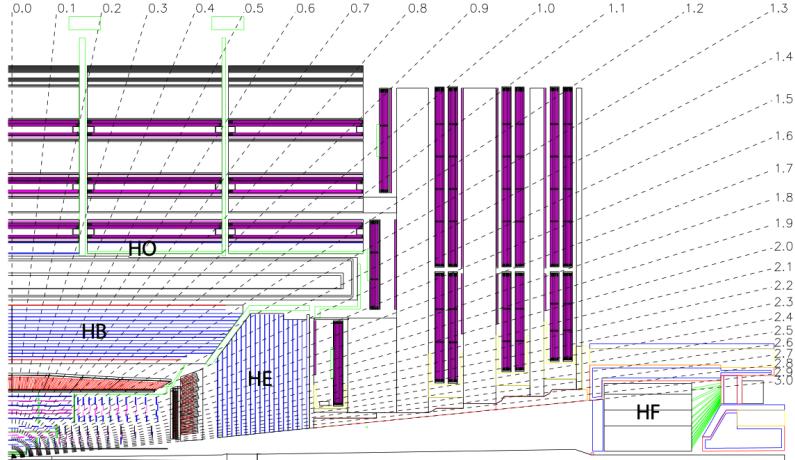


Figure 2.5: Longitudinal view of the CMS detector showing the hadron calorimeter barrel (HB), endcap (HE), outer (HO), and forward (HF) calorimeters from [18].

722 The HB is a sampling calorimeter covering the pseudorapidity range $|\eta| < 1.3$ [18].
723 It consists of 36 identical azimuthal wedges which form two half-barrels (HB+ and HB-
724), with a segmentation of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The HE covers pseudorapidity
725 $1.3 < |\eta| < 3$. The HB and endcap HE calorimeters are sampling calorimeters which
726 use brass as the absorber and plastic scintillator as the active material. Light from
727 the plastic scintillator is wavelength-shifted and captured in optic fibers which are

728 read out by front-end electronics [24].

729 In the central pseudorapidity region, the combined stopping power of EB plus the
730 HB is insufficient to contain hadron showers [18]. To ensure adequate sampling depth,
731 the hadron calorimeter is extended with a tail catcher, the HO. The size and position
732 of the tiles are designed to roughly map the layers of the HB to make towers with
733 the same granularity of 0.087×0.087 in η and ϕ . HO uses the same active material
734 as the HB and HE calorimeters, but uses the steel return yoke and magnet material
735 of CMS as absorbers [24].

736 The HF is a Cherenkov calorimeter based on a steel absorber and quartz fibers
737 which run longitudinally through the absorber and collect Cherenkov light, primarily
738 from the electromagnetic component of showers developed in the calorimeter [24].
739 Photomultiplier tubes are used to collect light from the quartz fibers. The HF is
740 designed to survive in the harsh radiation conditions and high particle flux of the
741 forward region. On average, 760 GeV per proton-proton interaction is deposited into
742 the two forward calorimeters, compared to only 100 GeV for the rest of the detector
743 [18]. Furthermore, this energy has a pronounced maximum at the highest rapidities.

744 2.5.4 Muon detectors

745 The CMS muon system is designed to have the capability of reconstructing the mo-
746 mentum and charge of muons over the kinematic range of the LHC, since muons are a
747 powerful handle on signatures of interesting processes over the high background rate
748 of the LHC [18]. For instance, the decay of the Standard Model Higgs boson into
749 ZZ , which in turn decay to 4 leptons, can be reconstructed with high 4-particle mass
750 resolution if all the leptons are muons, since muons are less affected than electrons
751 by radiative losses in the tracker material.

752 The muon system consists of a cylindrical barrel section and two planar endcap
753 regions [18]. The barrel muon detector consists of drift tube (DT) chambers covering

754 the pseudorapidity region $|\eta| < 1.2$ (Fig. 2.6). The DTs can be used as tracking
 755 detectors due to the barrel region's characteristic low neutron-induced backgrounds,
 756 low muon rate, and relatively uniform 4T magnetic field contained in the steel yoke.

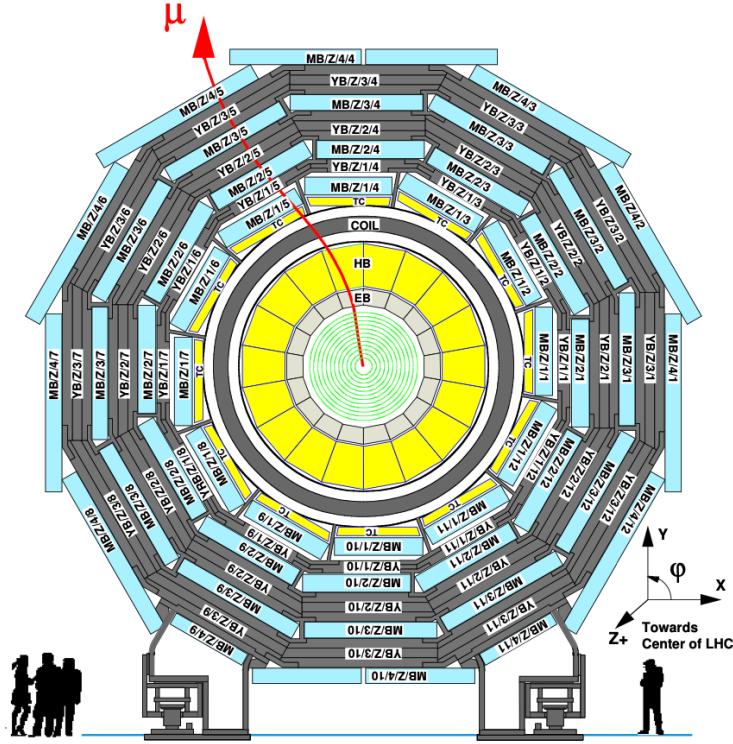


Figure 2.6: Layout of the CMS barrel muon drift tube (DT) chambers in one of the five wheels from [18]. The DTs are organized in 12 sectors of the yoke barrel (YB). In each of the 12 sectors of the yoke, there are 4 muon chambers per wheel (MB1, MB2, MB3, and MB4).

757 In the two endcap regions, the muon rates and background levels are high and the
 758 magnetic field is large and non-uniform [18]. Here, the muon system uses cathode
 759 strip chambers (CSCs) to identify muons between $0.9 < |\eta| < 2.4$. The cathode strips
 760 of each chamber run radially outwards and provide a precision measurement in the
 761 $r - \phi$ bending plane. The anode wires run approximately perpendicular to the strips
 762 and are read out in order to measure η and the beam-crossing time of a muon.

763 In addition to the DT and CSC, a dedicated trigger system consisting of resistive

764 plate chambers (RPCs) in the barrel and endcap regions provide a fast, independent,
765 and highly-segmented trigger with a sharp p_T threshold over a large portion of the
766 pseudorapidity range ($|\eta| < 1.6$) of the muon system [18]. RPCs have good time
767 resolution but coarser position resolution compared to the DTs or CSCs. The RPCs
768 also play a role in resolving ambiguities in reconstructing tracks from multiple hits in
769 a chamber.

770 2.5.5 The Level-1 Trigger

771 The design performance of the LHC corresponds to an instantaneous luminosity of
772 $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a 25 ns bunch crossing rate, giving an average pile-up (number
773 of simultaneous events) of 25 per bunch crossing [25]. The large number of minimum
774 bias events per bunch crossing, combined with the small cross-sections of possible
775 physics discovery signatures, necessitates a sophisticated event selection system for
776 filtering this large event rate, as it is impossible to save all events. This data filtering
777 system is implemented by CMS in two stages. The first stage is the Level-1 (L1)
778 Trigger, which is deployed in custom electronic hardware systems and is responsible
779 for reducing the event rate to around 100 kHz. The second stage is the High-Level
780 Trigger (HLT) which is described in Section 2.5.6. This section describes the Phase-1
781 configuration of the Level-1 Trigger.

782 The L1 Trigger data flow of Phase-1 is shown in Fig. 2.7 [25], with organization
783 into the L1 calorimeter trigger, the L1 muon trigger, and the L1 global trigger.

784 The L1 calorimeter trigger begins with trigger tower energy sums formed by the
785 ECAL, HCAL, and HF Trigger Primitive Generator (TPG) circuits from the indi-
786 vidual calorimeter cell energies. In the original configuration, the ECAL energies
787 were accompanied by a bit indicating the transverse extent of the electromagnetic
788 energy deposits, and the HCAL energies were accompanied by a bit indicating the
789 presence of minimum ionizing energy [26]. Between Long Shutdowns 1 and 2 (LS1

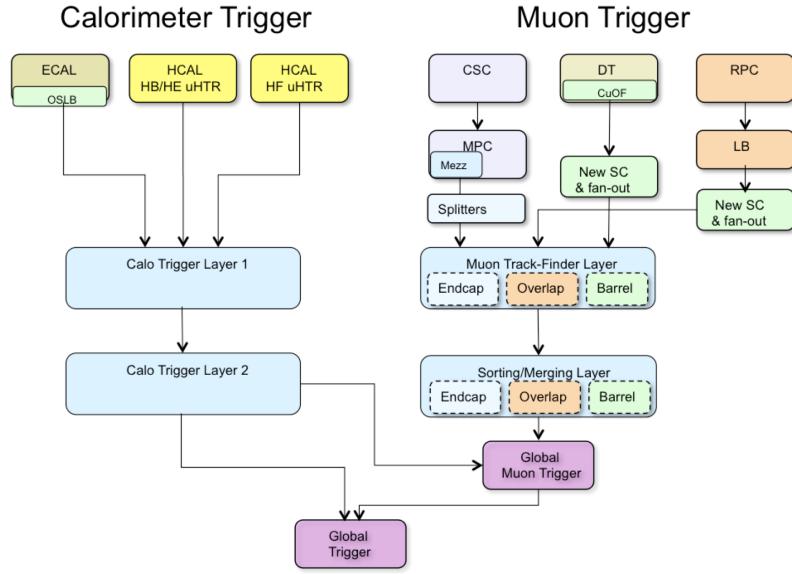


Figure 2.7: Dataflow for the Phase-1 Level-1 Trigger [25], which is implemented in custom hardware and is responsible for reducing the event rate from the LHC bunch crossing frequency of 400 MHz (bunch crossings every 25 ns) to a maximum rate of 100 kHz. In Phase-1, the Level-1 Trigger has access to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

and LS2), HF was upgraded to provide finer granularity information to the trigger, and the HCAL barrel and endcap front-end electronics were upgraded to provide high-precision timing information and depth segmentation information.

In the original design of the L1 calorimeter trigger, the trigger primitives are processed by the Regional Calorimeter Trigger (RCT, upgraded to Calo Layer 1 after LS2) which finds isolated and non-isolated electron/photon candidates [25]. At this stage, electrons/photons candidates are treated together since they cannot be definitively distinguished at this stage due to lack of tracking information in the L1 trigger. The Global Calorimeter Trigger (GCT, upgraded to Calo Layer 2 after LS2) sorts further the candidate electrons/photons, finds jets (classified as central, forward, and tau) using the E_T sums and performs calibration of the clustered jet energies, and calculates global quantities such as missing E_T . It sends the top four candidates of each type to the global trigger (GT) [25].

803 Each of the L1 muon triggers has its own trigger logic [26]. The RPC strips are
804 connected to a Pattern Comparator Trigger (PACT), which forms trigger segments
805 that are used to build tracks and calculate p_T . The RPC logic also provides some
806 hit data to the CSC trigger system to resolve ambiguities caused by two muons in
807 the same CSC. The CSCs form local charged tracks (LCTs) from the cathode strips,
808 which are combined with the anode wire information. LCTs are combined into full
809 muon tracks and assigned p_T values.

810 The Global Muon Trigger (GMT) sorts the RPC, DT, and CSC muon tracks,
811 converts these tracks to the same η , ϕ , and p_T scale, and validates the muon sign [26].
812 It improves the trigger efficiency by merging muon candidates that were detected
813 in two complementary sub-systems (i.e. DT+RPC, or CSC+RPC). The GMT also
814 contains logic to correlate the found muon tracks with an $\eta-\phi$ grid of quiet calorimeter
815 towers to determine if the muons are isolated, as well as logic to remove duplicate
816 candidates originating in the overlap regions from both DT and CSC systems. The
817 final collection of muons are sorted based on their initial quality, correlation, and p_T ,
818 and the top four muons are sent to the Global Trigger [26].

819 Information from the GCT and GT are sent to the Global Trigger (GT), which
820 makes the Level-1 Accept (L1A) decision to either discard or accept the bunch crossing
821 [26]. This is accomplished by sorting ranked trigger objects that are accompanied by
822 positional information in η and ϕ , permitting the trigger to applying criteria with
823 thresholds that can vary based on the location of the trigger objects, and/or to
824 require trigger objects to be close to or opposite from each other. The GT L1A
825 decision arrives at the detector front end with a $3.8 \mu\text{s}$ latency after the interaction
826 at a rate which is required to be less than 100 kHz, and triggers a full readout of the
827 detector for further processing.

828 **2.5.6 The High-Level Trigger**

829 The HLT is implemented in software running on a large computer farm of fast com-
830 mercial processors [27] [28]. The algorithms in HLT have access to full data from
831 all CMS sub-detectors, including the tracker, with full granularity and resolution.
832 The HLT reconstruction software is similar to what is used offline for full CMS data
833 analysis. As a result, the HLT can calculate quantities with a resolution compara-
834 ble to the final detector resolution, compared to the L1 Trigger. The HLT performs
835 more computationally-intensive algorithms, such as combining tau-jet candidates in
836 the calorimeter with high- p_T stubs in the tracker, to form a hadronic tau trigger. The
837 maximum HLT input rate from the L1 Trigger is 100 kHz, and the HLT output rate
838 is approximately 100 Hz.

839 The HLT contains trigger paths, each corresponding to a dedicated trigger [29].
840 A path consists of several steps implemented as software modules. Each HLT trigger
841 path must be seeded by one or more L1 trigger bits: the first module always looks
842 for a L1 seed, consisting of L1 bit(s) and L1 object(s). Each module performs a well-
843 defined task such as unpacking (raw to digitized quantities), reconstruction of physics
844 objects (electrons, muons, jet, missing transverse energy, etc.), making intermediate
845 decisions that trigger more detailed reconstruction modules, and calculating the final
846 decision for the trigger path. If an intermediate filter decision is negative, the rest of
847 the path is not executed, and the trigger rejects the event.

848 **2.5.7 Particle reconstruction**

849 To build a description of the physics objects present in the particle collision, the
850 basic elements from the detector layers (tracks and clusters of energy) are correlated
851 to identify each particle in the final state. Measurements from different sub-detectors
852 are combined to reconstruct the particle properties. This approach is called particle-
853 flow (PF) reconstruction [19]. Key to the success of the PF reconstruction is the

854 fine spatial granularity of the detector layers. Coarse-grained detectors can cause
855 the signals from different particles to merge, especially within jets. However, if the
856 subdetectors are sufficiently segmented to separate individual particles, it becomes
857 possible to produce a global event description that identifies all physics objects with
858 high efficiencies and resolution.

859 **2.5.8 Data storage and computational infrastructure**

860 The LHC generates over 15 petabytes (15 million gigabytes) of data every year, neces-
861 sitating a flexible computing system that can be accessed by researchers working at
862 the four main LHC experiments: ALICE, ATLAS, CMS, and LHCb. The Worldwide
863 LHC Computing Grid (WLCG) [30] is a global collaboration of computer centers that
864 links thousands of computers and storage systems in over 170 centers across 41 coun-
865 tries. These centers are arranged in “tiers”, and provide near real-time access to users
866 processing, analyzing, and storing LHC data. One of the final stages of data analy-
867 sis at LHC experiments is large-scale data processing taking place over distributing
868 computing, for instance, with the use of Condor [31], a distributed, scalable, flexible
869 batch processing system which accepts a computing job, allocates a resource to it,
870 executes it, and returns the result back to a user transparently.

871 **Chapter 3**

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

873 **3.1 High-Luminosity LHC and CMS**

874 In order to sustain and extend the LHC’s physics discovery program and maintain
875 operability for a decade or more, the LHC is undergoing a major upgrade to the High-
876 Luminosity LHC (HL-LHC). In its final configuration, the HL-LHC will deliver a peak
877 luminosity of $7.5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$, potentially leading to total integrated luminosity
878 of 4000 fb^{-1} after ten years of operations, scheduled to begin in 2027 [32]. This
879 integrated luminosity is about ten times the predicted luminosity reach of the LHC
880 in its initial configuration. To maximize the discovery potential of this unprecedented
881 amount of data, the CMS detector is undergoing Phase-2 upgrades in order to perform
882 high-precision measurements and searches for physics beyond the Standard Model in
883 the intense running conditions of the HL-LHC.

884 **3.2 The Phase-2 Level-1 Trigger**

885 To achieve the goals of the HL-LHC program and to ensure the collection of information-
886 rich datasets in the HL-LHC, the Phase-2 upgrade of the CMS Level-1 Trigger [32]
887 must be upgraded in conjunction with the CMS sub-detectors and their readouts, to

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

891 To profit from the extended coverage and increased granularity of the upgraded
 892 CMS detector, the latency of the L1 trigger system (time available to produce a L1
 893 Accept signal) will be increased significantly from $3.8 \mu\text{s}$ to $12.5 \mu\text{s}$, with an increased
 894 maximum output bandwidth of 750 kHz [32]. With the increased latency, in addition
 895 to information from calorimeters and muon detectors (as in the Phase-1 system),
 896 information from the new tracker and high-granularity endcap calorimeter can also
 897 be included at L1 for the first time. This is illustrated in the functional diagram of
 898 the architecture of the Phase-2 trigger system in Fig. 3.1.

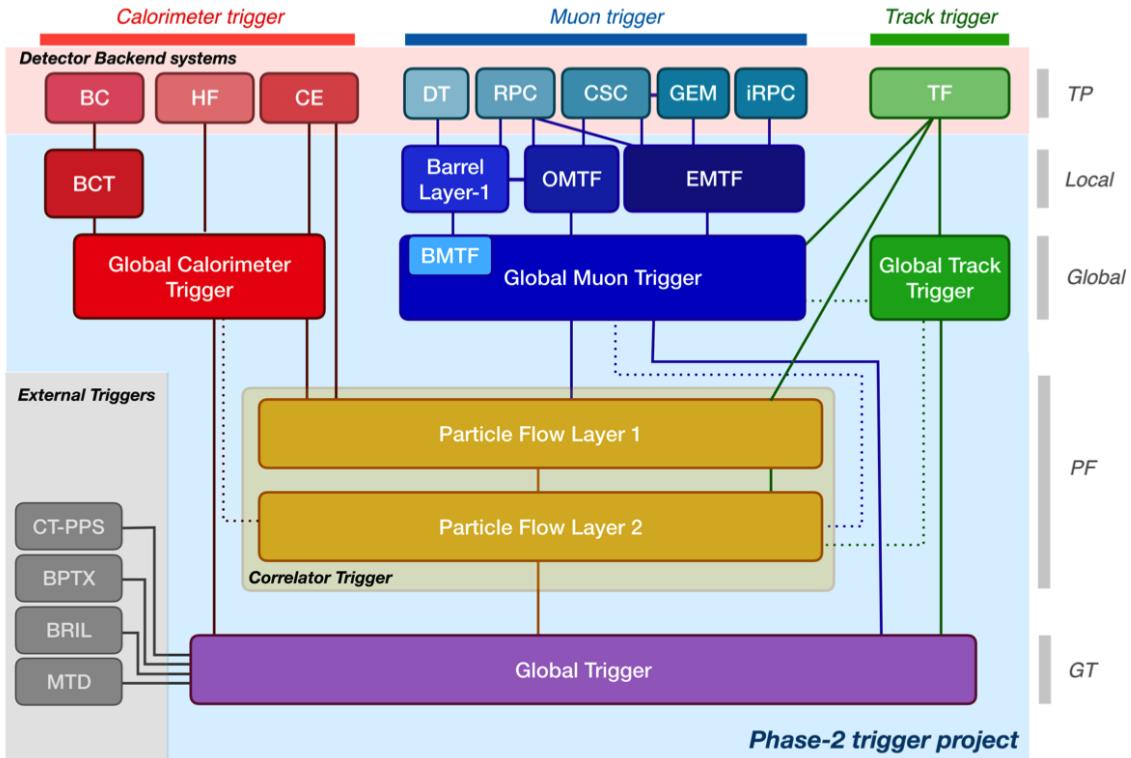


Figure 3.1: Functional diagram of the CMS L1 Phase-2 upgraded trigger design [32], showing the four trigger paths: calorimeter, muon, track, and Particle Flow. For the first time, tracking information will be available as early as the L1 Trigger.

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

900 where algorithms produce higher-level trigger objects by combining information from
901 sub-detectors, with a selectivity approaching that of offline reconstruction in the
902 HLT [32]. Four independent data processing paths (grouped together in Fig. 3.1) are
903 implemented: tracking, calorimetry, muon systems, and particle-flow techniques:

- 904 • **Calorimeter Trigger path:** (*red*, Fig. 3.1) A barrel calorimeter trigger (BCT)
905 and the HGCAL backend are used to produce high-granularity information from
906 the calorimeters to produce high-resolution clusters and identification variables
907 used for later processing. Outputs from the BCT, HGCAL, and the HF are sent
908 to a global calorimeter trigger (GCT), where calorimeter-only objects such as
909 e/γ candidates, hadronically decaying tau lepton candidates, jets, and energy
910 sums are built.
- 911 • **Track Trigger path:** (*green*, Fig. 3.1) Tracks from the Outer Tracker are
912 reconstructed in the track finder (TF) processors as part of the detector back-
913 end. A global track trigger (GTT) will reconstruct the primary vertices of the
914 event, along with tracker-only based objects, such as jets and missing transverse
915 momentum.
- 916 • **Muon Trigger path:** (*blue*, Fig. 3.1) Trigger primitives are processed by
917 muon track finder algorithms, again separated into the barrel (barrel muon
918 track finder, BMTF), overlap (overlap muon track finder, OMTF), and endcap
919 (endcap muon track finder, EMTF). Standalone muons and stubs containing
920 information such as position, bend angle, and timing, as well as L1 tracks, are
921 sent to the global muon trigger (GMT).
- 922 • **Particle-Flow Trigger path:** (*yellow*, Fig. 3.1) The correlator trigger (CT)
923 aims to approach the performance of offline Particle Flow, and is implemented
924 in two layers. “Layer-1” produces the particle-flow candidates from matching

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

927 The outputs from the above trigger paths are combined in the Global Trigger
928 (GT) (*purple*, Fig. 3.1), which calculates the final trigger decision (Level-1 Accept),
929 transmitting it to the Trigger Control and Distribution System (TCDS), which dis-
930 tributes it to the detector backend systems, initiating the readout to the DAQ. The
931 GT also provides the interface to external triggers (*grey*, Fig. 3.1), such as trig-
932 gers for the precision proton spectrometer (PPS), beam position and timing monitors
933 (BPTX), and luminosity and beam monitoring (BRIL) detectors [32]. The design of
934 the Phase-2 Level-1 Trigger allows for future inclusion of triggering information, for
935 instance information about minimum ionizing particles (MIPs) from the MIP Timing
936 Detector (MTD) [33].

937 **3.3 Standalone Barrel Calorimeter electron/photon 938 reconstruction**

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

945 **3.3.1 Phase-2 geometry of the ECAL Barrel trigger**

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

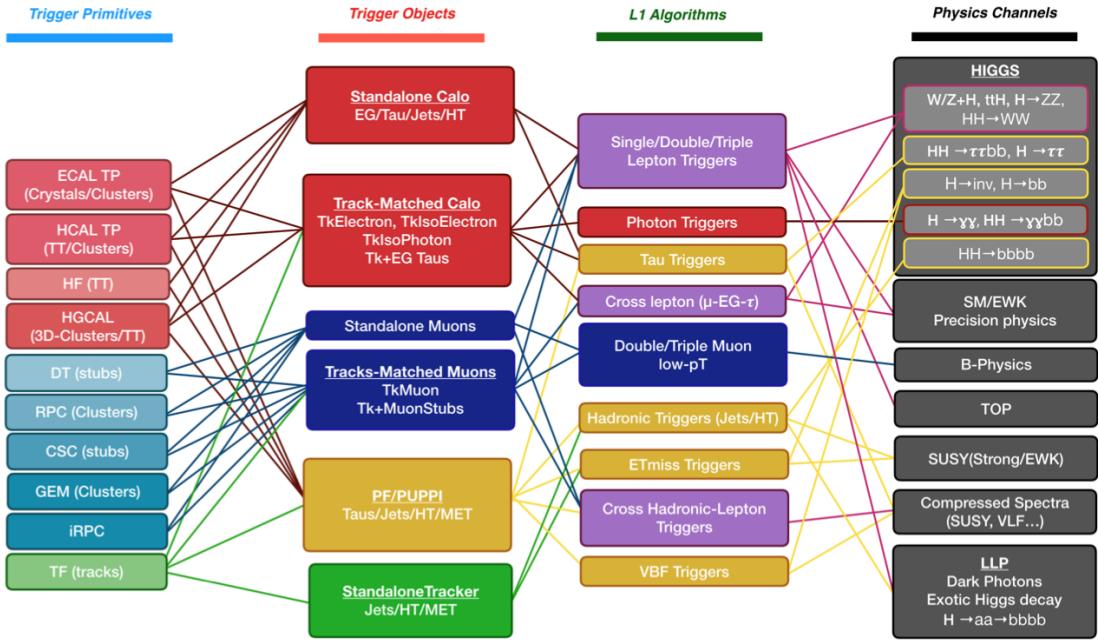


Figure 3.2: Summary of the links between the trigger primitives (*first column*), the trigger objects (*second column*), the Level-1 algorithms used in the menu (*3rd column*), and the physics channels (*4th column*), from [32], where a full description of the Phase-2 L1 algorithms can be found. This work focuses on developments for the Standalone Calorimeter electron and photon ("EG") reconstruction algorithm.

948 the on-detector to the backend electronics, in contrast to the lower-granularity output
 949 of the Phase-1 ECAL TPG that is restricted to providing trigger tower sums of 5×5
 950 crystals [32]. A schematic representation of the geometry of the ECAL barrel in the
 951 Regional Calorimeter Trigger (RCT) is shown in Fig. 3.3. The barrel is spanned by
 952 36 RCT cards, each spanning 17×4 towers in $\eta \times \phi$. Each RCT card is subdivided
 953 into five “regions” as shown in Fig. 3.4. After initial clustering and processing, the
 954 outputs of the RCT card are sent to the Global Calorimeter (GCT) trigger, which is
 955 processed in three cards as shown in Fig. 3.5.

956 3.3.2 Phase-2 electron/photon reconstruction algorithm

957 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
 958

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

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

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

- 974
 - Shower shape: The energy deposit sums around the seed crystal is computed in
975 windows of size 2×5 and 5×5 (Fig. 3.6, *dashed lines*), with true e/γ clusters
976 tending to produce showers that deposit most of their energy in a 2×5 region.
 - Bremsstrahlung recovery: e/γ tend to spread in the ϕ direction due to charged
977 particles being bent by the magnetic field of the CMS solenoid. If sufficient
978 energy comparable to the core 3×5 cluster is found in the adjacent 3×5
979 windows (Fig. 3.6, *shaded yellow*), the energy is added to the core cluster and
980 no longer considered unclustered energy.

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

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

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

994 The performance of the standalone barrel e/γ algorithm in Phase-2 conditions is
995 summarized in the efficiency and rates. The efficiencies are measured with a simulated
996 Monte Carlo sample containing electrons. The rates are measured with a simulated
997 minimum bias sample intended to closely mimic generic proton-proton collisions in
998 the CMS detector. The performance of the Phase-2 emulator discussed in this work,
999 which closely mimics the firmware logic and uses fixed-precision integers, is shown to
1000 be comparable to the previous emulator which used floats and idealized logic.

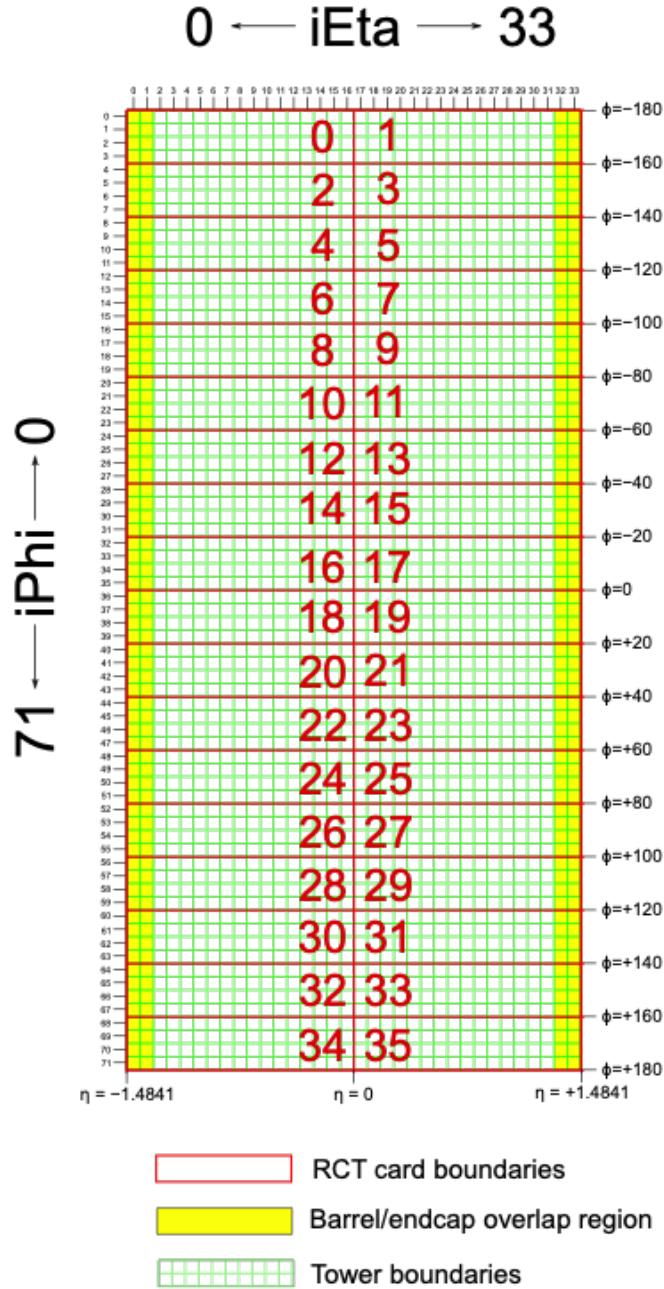


Figure 3.3: Schematic of the geometry of the Phase-2 ECAL barrel in the Regional Calorimeter Trigger (RCT), showing the division of the barrel region into 36 Regional Calorimeter Trigger (RCT) cards (*red*). Each card spans 17×4 towers in $\eta \times \phi$ (*green*), and each tower is 5×5 in single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

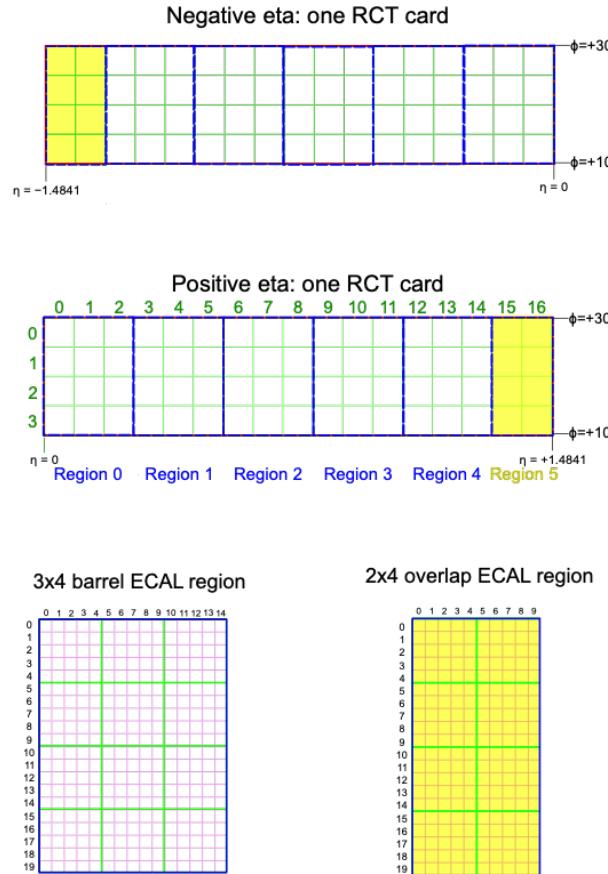


Figure 3.4: Schematic of two example RCT cards in the negative eta (*top*) and positive eta (*center*) regions of the ECAL barrel. Each RCT card is divided into five regions: four regions are of size 3×4 towers in $\eta \times \phi$ (*bottom left*), and a fifth smaller overlap region of size 2×4 towers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

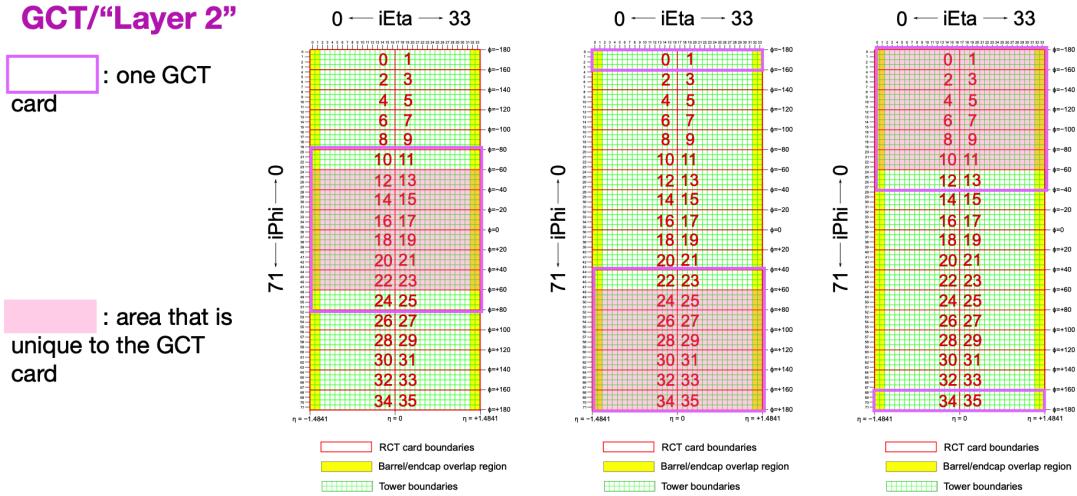


Figure 3.5: Schematic of the Phase-2 ECAL barrel in the Global Calorimeter Trigger (GCT), which will process the outputs of the Regional Calorimeter Trigger (RCT) in three cards (*magenta highlights*). Each card in the GCT processes the equivalent of sixteen RCT cards, with the center twelve being unique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

3x4 barrel ECAL region

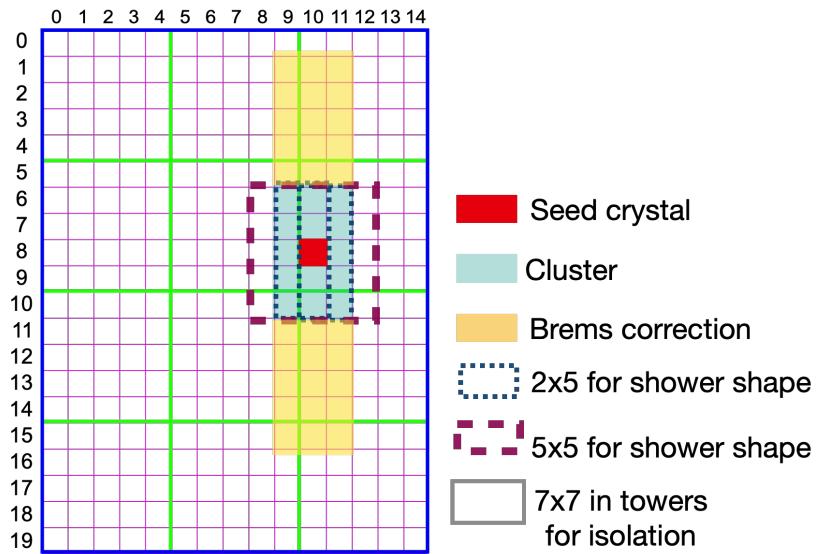


Figure 3.6: Illustration of an example electron/photon (e/γ) cluster in the Phase-2 Level-1 Trigger standalone barrel e/γ reconstruction, in a region of 15×20 crystals (3 \times 4 towers). Each small pink square is one crystal, the highest-granularity ECAL trigger primitives available to the L1 Trigger 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).

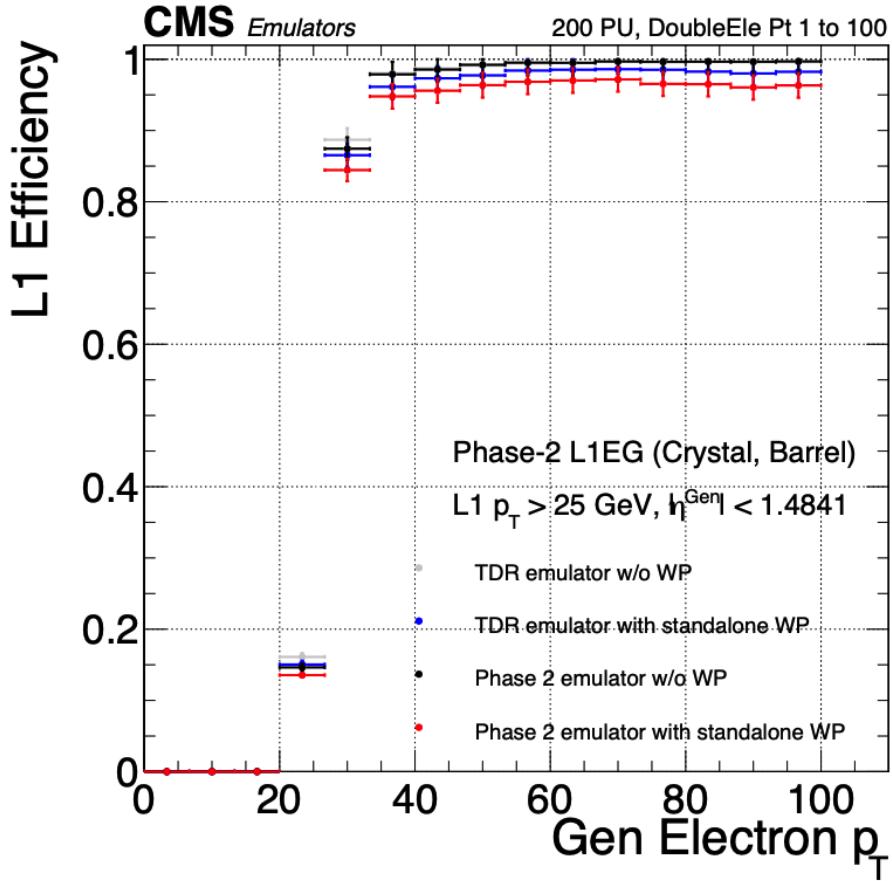


Figure 3.7: Efficiency of the standalone barrel e/γ reconstruction, measured in a simulated sample of electrons, as a function of the true electron's transverse momentum p_T . The performance of the previous, idealized algorithm as shown in the 2021 Phase-2 TDR [32] with and without the isolation and 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.

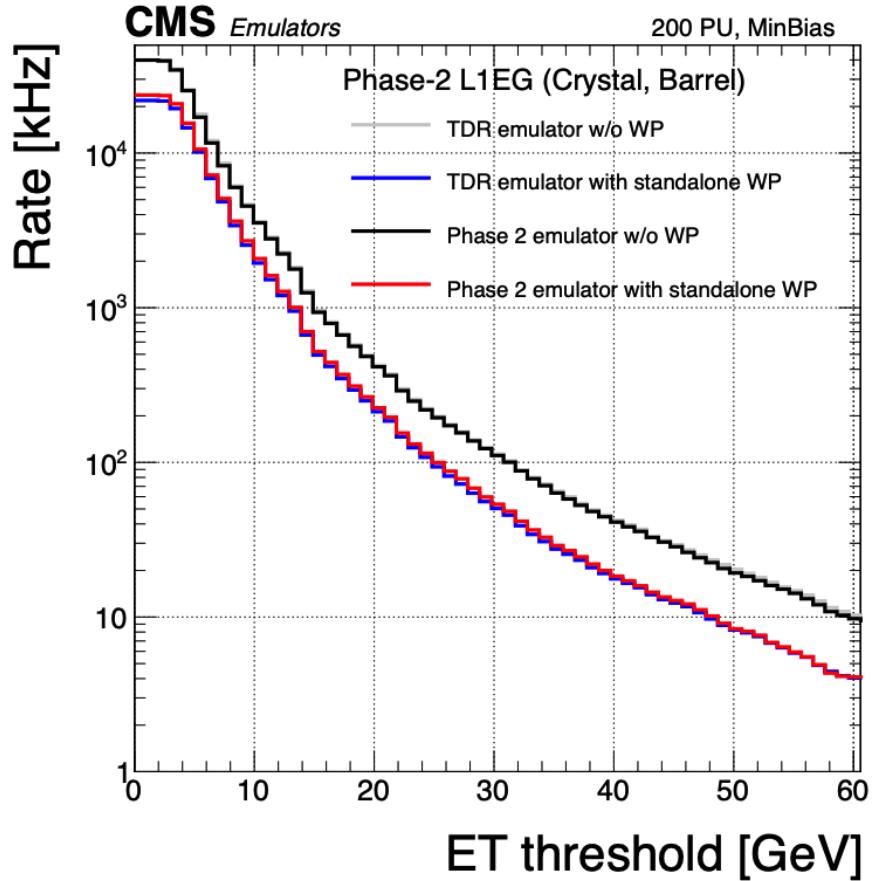


Figure 3.8: Rates of the standalone barrel e/γ reconstruction, evaluated on a minimum bias sample, measured as a function of the minimum energy (E_T) required of the reconstructed e/γ object in each event. The performance of the previous, idealized algorithm as shown in the 2021 Phase-2 TDR [32] with 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.

1001 **Chapter 4**

1002 **Datasets and Monte Carlo samples**

1003 **4.1 Datasets used**

1004 The $h \rightarrow aa \rightarrow 2b2\tau$ analysis (CMS CADI line HIG-22-007) is based on proton-proton
1005 collision data at a center-of-mass energy of 13 TeV collected in full Run-2 (2016-
1006 18) with the CMS detector. The data analyzed corresponds to a total integrated
1007 luminosity of 138 fb^{-1} (36.33 fb^{-1} for 2016, 41.53 fb^{-1} for 2017, and 59.74 fb^{-1} for
1008 2018) [34] [35] [36]. The cumulative delivered and recorded luminosity versus time
1009 for 2015-2018 is shown in Fig. 4.1.

1010 Data collected with the single muon trigger is used for the $\mu\tau_h$ channel. For the
1011 $e\tau_h$ channel, data collected with the single electron trigger is used; and for the $e\mu$
1012 channel, data collected with the electron + muon trigger is used. A more in-depth
1013 discussion of the triggers used follows in a later section.

1014 A full list of samples used can be found in the full documentation [38] [39].

1015 **4.2 Monte Carlo samples**

1016 Modeling and computing observables originating from arbitrary physics processes at
1017 the tree level and at next-to-leading order (NLO) is performed by Monte Carlo (MC)

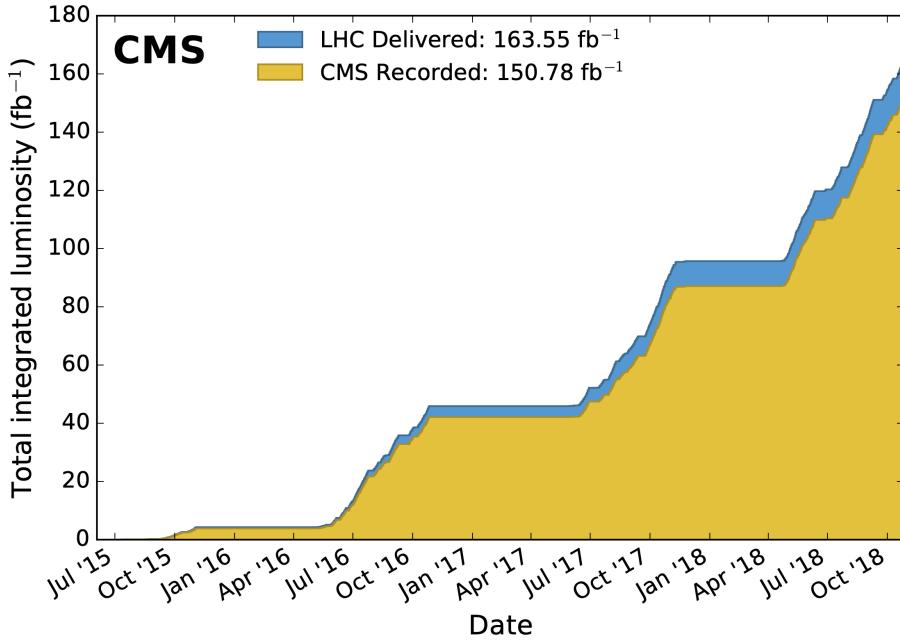


Figure 4.1: Cumulative delivered and recorded luminosity versus time for 2015-2018 at CMS, in proton-proton collision data only, at nominal center-of-mass energy [37].

1018 event generators, such as Powheg and MadGraph5_amCNLO [40] [41]. The information
 1019 generated, e.g. the computation of the differential cross sections and kinematics
 1020 of the final state particles, is saved in a compressed file and used to generate MC
 1021 samples that are used in physics analyses. The samples are digitized using GEANT4
 1022 [42], a platform used at the LHC and other facilities to comprehensively simulate the
 1023 passage of particles through matter. The digitized samples are passed through the
 1024 same detector reconstruction as real data events collected in the detector.

1025 The samples for modeling the signal ($h \rightarrow aa \rightarrow 2b2\tau$ and $h \rightarrow a_1a_2$) in the
 1026 2HDM+S and TRSM are generated at tree-level, for a range of masses of the light
 1027 neutral scalar a . For $h \rightarrow aa$, the mass hypotheses for the a range from $m_a =$
 1028 (12 GeV, 62.5 GeV). For $h \rightarrow a_1a_2$, the mass hypotheses for the two light scalars span
 1029 combinations of m_{a1} , m_{a2} ranging from (12 GeV, 62.5 GeV) for the two scalars.

1030 4.3 Embedded samples

1031 An important background for Higgs boson studies and searches for additional Higgs
1032 bosons is the decay of Z bosons into pairs of τ leptons ($Z \rightarrow \tau\tau$). An embedded tech-
1033 nique was developed in the context of Standard Model Higgs to $\tau\tau$ measurements, to
1034 model $Z \rightarrow \tau\tau$ decays, and was expanded to also model all Standard Model processes
1035 that contain $\tau\tau$ [43]. The embedded technique has since been used successfully at
1036 CMS for the Standard Model $H \rightarrow \tau\tau$ measurement, as well as searches for minimal
1037 supersymmetric extensions to the Standard Model (MSSM) [44] [45].

1038 Fig. 4.2 shows a schematic of how embedded samples are produced. Data events
1039 containing $Z \rightarrow \mu\mu$ decays are selected. In these events, all energy deposits of the
1040 recorded muons are removed, and are replaced with simulated tau leptons with the
1041 same kinematic properties as the removed muons. This results in a hybrid data format
1042 containing information from both observed and simulated events, as illustrated in Fig.
1043 4.2 [43].

1044 In the selection step of the embedded technique, events are selected with at least
1045 one of a set of $\mu\mu$ trigger paths, which require $p_T > 17(8)$ GeV for the leading
1046 (sub-leading) muons, and a minimum requirement between 3.8 and 8.0 GeV on the
1047 invariant di-muon mass $m_{\mu\mu}$ [43]. The offline reconstructed muons must match the
1048 objects at trigger level and also have offline $p_T > 17(8)$ GeV. They must have $|\eta| < 2.4$
1049 and be located at a distance $|d_z| < 0.2$ cm to the primary vertex along the beam
1050 axis. To form a Z boson candidate, each muon is required to originate from a global
1051 muon track. The muon pairs must have opposite charges with an invariant mass of
1052 $m_{\mu\mu} > 20$ GeV. If more than two di-muon pairs are found, the pair with the invariant
1053 mass closest to the Z boson mass (91.19 GeV) is chosen.

1054 This selection is designed to be tight enough to ensure a high purity of genuine
1055 $\mu\mu$ events, and also loose enough to minimize biases of the embedded event samples.
1056 Isolation requirements are avoided, since they would introduce a bias towards less

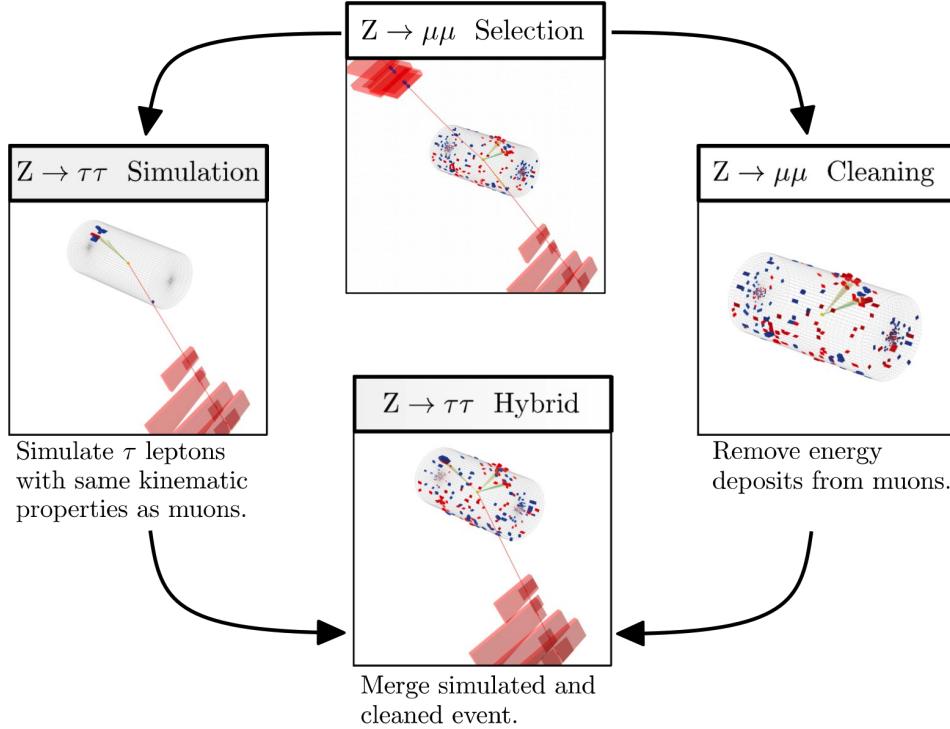


Figure 4.2: Schematic view of the four main steps of the embedding technique for τ leptons, as described in Section 4.3 [43]. A $Z \rightarrow \mu\mu$ event is selected in data ($Z \rightarrow \mu\mu$ selection), all of the energy deposits associated with the muons are removed ($Z \rightarrow \mu\mu$ 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 [43]. $Z \rightarrow \mu\mu$ 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 [43]. 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 [43], 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*).

1068 vertices, and the missing transverse momentum p_T^{miss} . Since all events with genuine
 1069 $\tau\tau$ are estimated with samples made with the embedded technique (referred to as
 1070 embedded samples from here on), events in Monte Carlo simulation with genuine $\tau\tau$
 1071 are not used, in order to avoid double-counting.

1072 **Chapter 5**

1073 **Object reconstruction and
1074 corrections applied**

1075 In this chapter on object reconstruction and corrections, Section 5.1 reviews the
1076 physical properties of the objects most pertinent to the analyses presented in this
1077 work: taus (τ), muons (μ), electrons (e), and jets, with a focus on jets originating from
1078 b quarks (b-flavor jets), as well as the methodology used to reconstruct the particles
1079 from their characteristic signatures in the CMS detector. Section 5.2 describes the
1080 method used to reconstruct the invariant full $\tau\tau$ mass which is used for the final signal
1081 extraction. Lastly, Section 5.3 describes the corrections applied to the simulated
1082 samples which improve their modeling of data.

1083 **5.1 Object reconstruction**

1084 **5.1.1 Taus**

1085 The tau (τ) is the heaviest known lepton. With a rest mass of 1776.86 MeV, it can
1086 decay to not only electrons and muons, but also hadrons. The mean lifetime of the τ
1087 is $\tau = 290 \times 10^{-15}$ seconds, corresponding to $c\tau = 87.03 \mu\text{m}$, which is short enough

1088 that taus decay in the CMS detector before reaching the detector elements.

1089 In two thirds of the cases, τ leptons decay hadronically, typically into one or three
1090 charged mesons (predominantly π^+ , π^-), often accompanied by neutral pions (that
1091 decay $\pi^0 \rightarrow \gamma\gamma$), and a ν_τ . These hadronic decays are denoted τ_h . In the remainder of
1092 the decays, the tau decays to the lighter leptons (electron or muon), termed leptonic
1093 decays. In all cases, at least one neutrino is produced, resulting in missing transverse
1094 energy in the CMS detector. The tau's largest decay branching ratios (proportional
1095 to probability of decay) are listed below [23]:

1096 • 17.8% decay to $e^- \bar{\nu}_e \nu_\tau$

1097 • 17.4% decay to $\mu^- \bar{\nu}_\mu \nu_\tau$

1098 • 25.5% decay to $\pi^- \pi^0 \nu_\tau$ (ρ^- resonance at 770 MeV)

1099 • 10.8% decay to $\pi^- \nu_\tau$

1100 • 9.3% decay to $\pi^- \pi^0 \pi^0 \nu_\tau$ (a_1^- resonance at 1200 MeV)

1101 • 9.0% decay to $\pi^- \pi^- \pi^+ \nu_\tau$ (a_1^- resonance at 1200 MeV)

1102 The neutrinos escape undetected from the CMS detector and are not considered
1103 in the reconstruction. Charged hadrons leave tracks in the tracking detector before
1104 being absorbed in the hadronic calorimeter; in CMS tau reconstruction terminology,
1105 they are often called “prongs”, i.e. the dominant τ_h decay modes are termed “1 prong”
1106 (π^\pm), “1 prong + $\pi^0(s)$ ”, and “3-prong”. Neutral pions decay to two photons which
1107 lose their energy in the electromagnetic calorimeter. Taus that decay to electrons
1108 and muons, are typically triggered on and reconstructed as electrons and muons
1109 respectively.

1110 **Hadron plus strips (HPS) reconstruction of τ_h**

1111 At CMS, hadronically decaying tau leptons are reconstructed with the hadron plus
1112 strips (HPS) algorithm [46] [47]. The HPS algorithm capitalizes on photon conversions
1113 in the CMS tracker material, which originate from the neutral pion (π^0) decaying
1114 to two photons. The bending of electron/positron tracks due to the CMS solenoid
1115 magnetic field leads to a spread of the neutral pions' calorimeter signatures in the ϕ
1116 direction. This motivates the reconstruction of photons in “strips”: objects that are
1117 built out of PF photons and electrons. The strip reconstruction starts with centering
1118 a strip on the most energetic electromagnetic particle in a PF jet. Among other
1119 electromagnetic particles located in a window of size $\Delta\eta = 0.05$ and $\Delta\phi = 0.20$
1120 around the strip center, the most energetic one is associated with the strip and its
1121 momentum is added to the strip momentum. This is repeated iteratively until no
1122 further particles can be associated. Lastly, strips satisfying a requirement of $p_T^{\text{strip}} > 1$
1123 GeV are combined with charged hadrons to reconstruct individual τ_h decay modes,
1124 where h stands for both π and K :

1125 • *Single hadron:* $h^- \nu_\tau$ and $h^- \pi^0 \nu_\tau$ decay modes, in which the neutral pions have
1126 too little energy to be reconstructed as strips.

1127 • *One hadron + one strip:* $h^- \pi^0 \nu_\tau$ decay modes, where the photons from the π^0
1128 decay are close together in the calorimeter.

1129 • *One hadron + two strips:* $h^- \pi^0 \nu_\tau$ decay modes, where the photons from the π^0
1130 decay are well separated.

1131 • *Three hadrons:* $h^- h^+ h^- \nu_\tau$ decay modes. The three charged hadrons are re-
1132 quired to originate from the same secondary vertex.

1133 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
1134 reconstructed with the above topologies.

1135 In the HPS algorithm, the direction of the reconstructed tau momentum \vec{p}^{τ_h}
1136 is required to fall within a distance of $\Delta R = 0.1$ from the original PF jet. All
1137 charged hadrons and strips are required to be contained within a cone of size $\Delta R =$
1138 $(2.8 \text{ GeV})/p_T^{\tau_h}$, from the τ_h as reconstructed by the HPS.

1139 All charged hadrons are assumed to be pions, and they are required to be consis-
1140 tent with the masses of the intermediate meson resonances (if applicable), with the
1141 following allowed windows for candidates: 50-200 MeV for π^0 , 0.3-1.3 GeV for ρ , and
1142 0.8-1.5 GeV for a_1 . If the τ_h decay is compatible with more than one hypothesis, the
1143 one giving the highest $p_T^{\tau_h}$ is chosen. Lastly, an isolation requirement is applied: aside
1144 from the τ_h decay products, no charged hadrons or photons can be present within
1145 an isolation cone of size $\Delta R = 0.5$ around the direction of the τ_h . The outputs of
1146 the HPS algorithm are the reconstructed decay mode and the visible four-momentum
1147 (i.e. the four-momenta of all decay products excluding the neutrinos).

1148 **DeepTau for identifying τ_h**

1149 The identification of τ_h candidates in CMS has historically been divided into separate
1150 discriminators against jets, electrons, and muons. Discriminators versus jets and
1151 electrons use information from derived quantities, such as the p_T sum of particles
1152 near the τ_h axis. Building on the previous multivariate analysis (MVA) classifier [48]
1153 based on a boosted decision tree (BDT), DeepTau is a more recent classifier based on a
1154 deep neural network (DNN) that simultaneously discriminates against jets, electrons,
1155 and muons. The DNN uses a combination of high-level inputs, similar to previous
1156 algorithms, and also uses convolutional layers in $\eta\text{-}\phi$ space to process information
1157 from all reconstructed particles near the τ_h axis. Convolutional layers are based on
1158 the principle that an image can be processed independently of its position.

1159 The final DeepTau discriminators against jets, muons, and electrons are given by

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

1160 where y_τ (y_α) are estimates of the probabilities for the τ_h candidate to come from
1161 a genuine τ_h (jet, μ , e). Working points for each discriminator with different τ_h
1162 identification efficiencies are defined for D_e , D_μ , and D_{jet} , for usage in physics analyses
1163 and derivation of data-to-simulation corrections [49].

1164 5.1.2 Muons

1165 Muons are the next lightest lepton after taus, with a mass of 105.66 MeV and a
1166 mean lifetime of $\tau = 2.20 \times 10^{-6}$ seconds, or $c\tau = 658.64$ m. At CMS, muons are
1167 identified with requirements on the quality of the track reconstruction and on the
1168 number of measurements in the tracker and the muon systems [50]. In the standard
1169 CMS reconstruction, tracks are first reconstructed independently in the inner tracker
1170 (tracker track) and in the muon system (standalone-muon track). Next, these tracks
1171 are processed in two different methods.

1172 The first is Global Muon reconstruction (outside-in) [50], which fits combined hits
1173 from the tracker track and standalone-muon track, using the Kalman-filter technique.
1174 At large transverse momenta, $p_T \gtrsim 200$ GeV, the global-muon fit can improve the
1175 momentum resolution compared to the tracker-only fit.

1176 The second is Tracker Muon reconstruction (inside-out) [50], which starts with
1177 tracker tracks with $p_T > 0.5$ GeV and total momentum $p_T > 2.5$ GeV. These tracks
1178 are extrapolated outwards to the muon system and matched to detector segments
1179 there, taking into account the magnetic field, expected energy losses, and multiple
1180 Coulomb scattering in the detector material. Tracker Muon reconstruction is more
1181 efficient than the Global Muon reconstruction at low momenta, $p \lesssim 5$ GeV, because

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

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

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

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

5.1.3 Electrons

Electrons are the lightest lepton with a mass of 0.511 MeV. At CMS, electrons are reconstructed by associating a track reconstructed in the silicon tracking detector with a cluster of energy in the ECAL. Performance is maximized via a combination of a stand-alone approach and the complementary global particle-flow approach [51].

In the stand-alone approach, the electron energy, which is typically spread over several crystals of the ECAL, is clustered with the “hybrid” algorithm in the barrel and the “multi- 5×5 ” in the endcaps [51]. The hybrid algorithm collects energy in a

1205 small window in η and an extended window in ϕ . It identifies a seed crystal, and adds
1206 arrays of 5×1 crystals in $\eta \times \phi$ in a range of $N = 17$ crystals in both directions of
1207 ϕ , if their energies exceed a minimum threshold, thus forming a supercluster (SC). In
1208 the endcap, crystals are not arranged in an $\eta \times \phi$ geometry; instead clusters are build
1209 around seed crystals in clusters of 5×5 crystals that can partly overlap. Nearby
1210 clusters are grouped into a supercluster, and energy is recovered from associated
1211 deposits in the preshower.

1212 In the PF reconstruction [51], PF clusters are reconstructed by aggregating around
1213 a seed all contiguous crystals with energies two standard deviations above the elec-
1214 tronic noise observed at the beginning of a data-taking run. The energy of a given
1215 crystal can be shared among two or more clusters.

1216 The electron track reconstruction is performed in two ways [51]: the ECAL-based
1217 seeding, which begins with the SC energy and positioning, and the tracker-based
1218 seeding (part of the PF reconstruction algorithm), which uses tracks reconstructed
1219 from the general algorithm for charged particles, extrapolated towards the ECAL and
1220 matched to an SC. Kalman filter (KF) tracks with a small number of hits or that are
1221 not well-fitted, are re-fitted with a dedicated Gaussian sum Filter (GSF).

1222 A global identification variable [51] is defined using a multivariate analysis (MVA)
1223 technique that combines information on track observables (kinematics, quality of the
1224 KF track and GSF track), the electron PF cluster observables (shape and pattern),
1225 and the association between the two (geometric and kinematic observables). For
1226 electrons seeded only through the tracker-based approach, a weak selection is applied
1227 on this MVA variable. For electrons seeded through both approaches, a logical OR is
1228 taken.

1229 Electron isolation, i.e. the presence of energy deposits near the electron trajectory,
1230 is a separate key handle in rejecting significant background. Compared to isolated
1231 electrons, electrons from misidentified jets or genuine electrons within a jet resulting

1232 from semileptonic decays of b or c quarks tend to have significant energy deposits
 1233 near the primary trajectory [51]. Offline analyses benefit from the PF technique
 1234 for defining isolation, which sums the PF candidates reconstructed located within a
 1235 specified isolation cone around the electron candidate, as in Eqn. 5.2.

1236 5.1.4 Jets

1237 The vast majority of processes of interest at the LHC contains quarks or gluons in
 1238 the final state, but these particles cannot be observed directly. In a process called
 1239 hadronization, they fragment into spatially-grouped collections of particles called jets,
 1240 which can be detected in the tracking and calorimeter systems. Hadronization and
 1241 the subsequent decays of unstable hadrons can produce hundreds of nearby particles
 1242 in the CMS detector. Jets are reconstructed by the PF algorithm (PF jets), or from
 1243 the sum of the ECAL and HCAL energies deposited in the calorimeter towers (Calo
 1244 jets). In PF jets, typically used in offline analyses, jets are built using the anti- k_T
 1245 (AK) clustering algorithm [52]. The anti- k_T algorithm iterates over particle pairs and
 1246 finds the two that are closest in a distance measure d , and determines whether to
 1247 combine them:

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

1248 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-
 1249 ity, and azimuthal angle of particle i . The power -2 means that higher-momentum
 1250 particles are clustered first, leading to jets that tend to be centered on the hardest
 1251 (highest p_T) particle.

1252 There are several methods to remove contributions of pileup collisions from jet
 1253 clustering [53]:

- 1254 • Charged hadron subtraction (CHS), which removes all charged hadron candi-

1255 dates associated with a track that is not associated with the primary vertex.

- 1256 • PileUp Per Particle Identification (PUPPI), which weighs input particles based
1257 on their likelihood of arising from pileup. QCD particles tend to have a collinear
1258 structure, compared to soft diffuse radiation coming from pileup. The local
1259 shape for charged pileup, used as a proxy for all pileup particles, is used on an
1260 event-by-event basis to calculate a weight for each particle. PUPPI is deployed
1261 in Run-2 and is more performant than CHS in high pileup scenarios.

1262 **5.1.5 B-flavored jets**

1263 Jets that arise from bottom-quark hadronization (b-flavor jets) have overwhelming
1264 background from processes involving jets from gluons (g) and light-flavor quarks (u, d,
1265 s), and from c-quark fragmentation. The ability to identify b-flavor jets, or b-tagging,
1266 exploits the b hadrons' relatively large masses, long lifetimes, and daughter particles
1267 with hard momentum spectra [52].

1268 The impact parameter (IP) of a track is the 3-dimensional distance between the
1269 track and the primary vertex (PV) at the point of closest approach. The IP is positive
1270 if the track originates from the decay of particles travelling along the jet axis. The
1271 resolution of the IP depends on the p_T and η of the track, motivating the use of the
1272 impact parameter significance S_{IP} (ratio of the IP to its estimated uncertainty) as an
1273 observable [52].

1274 Because of the large but finite lifetimes of the b hadrons, b hadrons tend to
1275 travel a short distance before decaying at a secondary vertex (SV), which can be
1276 measured and reconstructed separately from the primary vertex due to the excellent
1277 position resolution of the pixel detector [52]. Previous b-tagging algorithms (e.g.
1278 CSV, cMVAv2, and DeepCSV) have capitalized on variables such as the presence of
1279 a SV, the flight distance and direction (computed from the vector between the PV
1280 and the SV), and kinematics of the system of associated secondary tracks (e.g. track

1281 multiplicity, mass, and energy).

1282 The DeepJet (formerly known as DeepFlavour) algorithm [54] is a deep-neural-
1283 network multi-classification algorithm, which uses 16 properties of up to 25 charged
1284 and 6 properties of 25 neutral particle-flow jet constituents, as well as 17 properties
1285 from up to 4 secondary vertices associate with the jet. Compared to the previous clas-
1286 sifying algorithm DeepCSV, DeepJet has been demonstrated to have higher efficiency
1287 with lower misidentification probability in Phase-1 data [55].

1288 5.2 Reconstruction of the $\tau\tau$ mass

1289 The final signal extraction is done to the total $\tau\tau$ mass, which is estimated from the
1290 visible $\tau\tau$ mass using the FastMTT algorithm [56]. FastMTT is based on the SVFit
1291 algorithm, originally developed for the Standard Model $H \rightarrow \tau\tau$ analysis [57]. Both
1292 the SVFit algorithms, and the FastMTT algorithm, are described below, to give a
1293 complete picture of how tau decays are parameterized.

1294 To specify a hadronic τ decay, six parameters are needed [57]: the polar and
1295 azimuthal angles of the visible decay product system in the τ rest frame, the three
1296 boost parameters from the τ rest frame to the laboratory frame, and the invariant
1297 mass m_{vis} of the visible decay products. For a leptonic τ decay, two neutrinos are
1298 produced, and a seventh parameter, the invariant mass of the two-neutrino system, is
1299 necessary. The unknown parameters are constrained by four observables that are the
1300 components of the four-momentum of the system formed by the visible decay products
1301 of the τ lepton, measured in the laboratory frame. The remaining unconstrained
1302 parameters for hadronic and leptonic τ decays are thus:

- 1303 • The fraction of the τ energy in the laboratory frame carried by the visible decay
1304 products,
- 1305 • ϕ , the azimuthal angle of the τ direction in the laboratory frame,

- 1306 • $m_{\nu\nu}$, the invariant mass of the two-neutrino system in leptonic τ decays (for
 1307 hadronic τ decays, $m_{\nu\nu}$ is set to 0).

1308 E_x^{miss} and E_y^{miss} , the x and y components of the missing transverse energy \vec{E}_T^{miss}
 1309 provide two further constraints.

1310 **5.2.1 Original SVFit “standalone”: maximum likelihood**

1311 In one of the original versions of SVFit, called “standalone” SVFit [57], a maximum
 1312 likelihood fit method is used to reconstruct the mass $m_{\tau\tau}$ by combining the measured
 1313 observables E_x^{miss} and E_y^{miss} with a likelihood model that includes terms for the τ
 1314 decay kinematics and the \vec{E}_T^{miss} resolution [57]. The likelihood function $f(\vec{z}, \vec{y}, \vec{a}_1 \vec{a}_2)$
 1315 of the parameters $\vec{z} = (E_x^{\text{miss}}, E_y^{\text{miss}})$ in an event is constructed, where the remaining
 1316 parameters are the kinematics of the two τ decays, denoted $\vec{a}_1 = (x_1, \phi_1, m_{\nu\nu,1})$ and
 1317 $\vec{a}_2 = (x_2, \phi_2, m_{\nu\nu,2})$, and the four-momenta of the visible decay products with the
 1318 measured values $\vec{y} = (p_1^{\text{vis}}, p_2^{\text{vis}})$.

1319 The likelihood f is the product of three likelihood functions. The first two likelihood
 1320 functions model the decay parameters \vec{a}_1 and \vec{a}_2 of the two τ leptons. For leptonic
 1321 decays, the likelihood function is modeled using matrix elements for τ decays,
 1322 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
 1323 hadronic τ decays, a model based on the two-body phase space is used and integrated
 1324 over $m_{\text{vis}}^2/m_{\tau\tau}^2 \leq x \leq 1$. The third likelihood function quantifies the compatibility of
 1325 a τ decay hypothesis with the reconstructed \vec{E}_T^{miss} in an event, assuming the neutrinos
 1326 are the only source of missing transverse energy. The expected \vec{E}_T^{miss} resolution
 1327 is represented by a covariant matrix, estimated on an event-by-event basis using a
 1328 significance algorithm [58].

1329 5.2.2 “Classic SVFit” with matrix element

1330 Classic SVFit is an improved algorithm of the original “standalone” SVFit using the
 1331 formalism of the matrix element (ME) method [56]. In the ME method, an estimate
 1332 for the unknown model parameter Θ (here, the mass $m_{\tau\tau}$) is obtained by maximizing
 1333 the probability density \mathcal{P} . The key ingredients of the probability density are the
 1334 squared modulus of the matrix element $|\mathcal{M}(\mathbf{p}, \Theta)|^2$ and the transfer function $W(\mathbf{y}|\mathbf{p})$
 1335 (probability density to observe the measured observables \mathbf{y} given the phase space
 1336 point \mathbf{p}). The best estimate $m_{\tau\tau}$ is obtained by computing the probability density \mathcal{P}
 1337 for a range of mass hypotheses and finding the value of $m_{\tau\tau}$ that maximizes \mathcal{P} .

1338 Distributions illustrating the performance of the classic matrix element SVFit
 1339 algorithm are shown in Fig. 5.1 from [56], showing the di-tau mass after and before
 1340 application of SVFit to recover energy lost to neutrinos. The SVFit algorithm is
 1341 found to improve the sensitivity of the Standard Model $H \rightarrow \tau\tau$ analysis performed
 1342 by CMS by about 30%, compared to performing the same analysis using only the
 1343 visible mass m_{vis} .

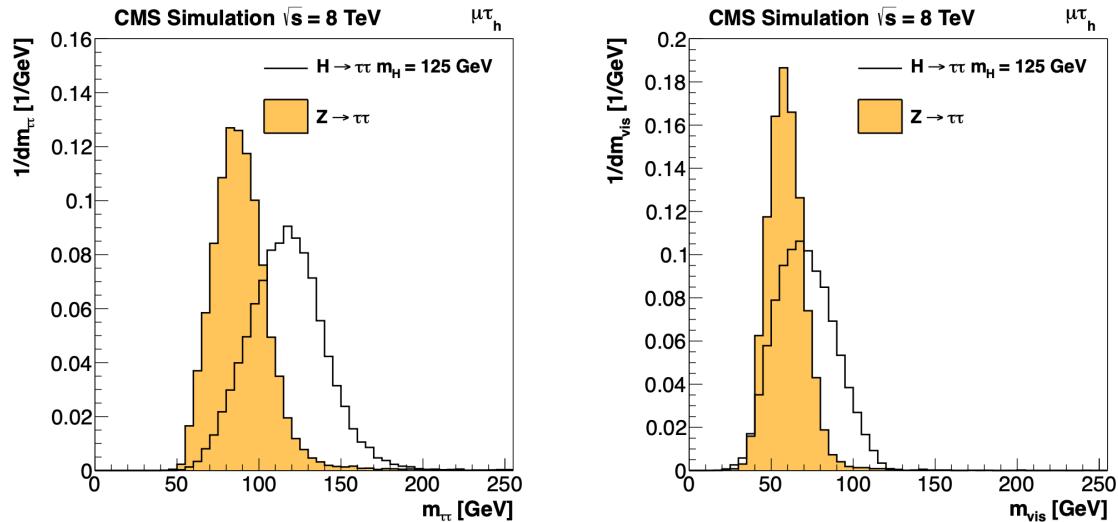


Figure 5.1: Distributions from [56], of $m_{\tau\tau}$ after reconstruction with the original SVFit algorithm (*left*), and before SVFit with only the visible tau decay products (*right*), for $H \rightarrow \tau\tau$ signal events of mass $m_H = 125$ GeV (*black line*) and the $Z/\gamma^* \rightarrow \tau\tau$ background (*orange, solid*), in the decay channel $\tau\tau \rightarrow \mu\tau_h$.

1344 5.2.3 FastMTT: optimized SVFit

1345 FastMTT [59] is a further simplification to the matrix element method of Classic
1346 SVFit which has comparable performance but is about 100 times faster. FastMTT
1347 drops the matrix element component of the computation without significant impact
1348 on the final mass resolution, and simplifies the computation of the transfer functions.
1349 The opening angle of the τ decay products with respect to the initial τ momenta ap-
1350 proaches 0 for τ with high $\gamma = E_\tau/m_\tau$, with typical τ decays from the Z boson decays
1351 already satisfying this condition. In this collinear approximation, the dimensionality
1352 of the transfer function can be reduced in the computation of FastMTT, while still
1353 yielding similar results to Classic SVFit [59].

1354 5.3 Corrections applied to simulation

1355 Corrections are applied to simulated samples to account for known effects in the event
1356 modeling and reconstruction and data-taking, and are intended to bring simulations
1357 in closer agreement with data. Corrections fall into two broad categories: *energy*
1358 *scale corrections* applied to physics objects, and *event-level corrections*. Energy scale
1359 corrections are multiplicative factors applied to the energy and transverse momentum
1360 p_T of simulated objects (e.g. leptons or jets), and bring the average reconstructed en-
1361 ergies of simulated particles into better agreement with those of objects reconstructed
1362 from data. Event-level corrections are applied as a per-event multiplicative weight,
1363 and account for effects such as mis-modeling in simulations of the underlying physics
1364 process, or changing detector operating conditions during data-taking. Event-level
1365 corrections change the shapes of the distributions of all the physical observables.

1366 Uncertainties in scale factors and corrections are also sources of systematic errors
1367 in the analysis, detailed in Chapter 8. Systematic uncertainties in the tau, muon, and
1368 electron energy scales can shift the p_T of the leptons up or down, which can change

1369 whether events pass or fail the offline p_T thresholds for the trigger paths described in
1370 the previous section, i.e. change the number of events in the signal region.

1371 5.3.1 Tau energy scale

1372 An energy scale is applied to the transverse momentum p_T and mass of the hadronic
1373 tau τ_h in the $\mu\tau_h$ and $e\tau_h$ channels, to correct for a deviation of the average recon-
1374 structed τ_h energy from the generator-level energy of the visible τ_h decay products.
1375 These correction factors are derived centrally [48], by fitting to events in $e\tau_h$ and $\mu\tau_h$
1376 final states in Z/γ^* events separately for the h^\pm , $h^\pm\pi^0$, and $h^\pm h^\mp h^\pm$ decays. The
1377 values used are shown in Table 5.1.

1378 When applying the energy scale to the τ_h , the 4-momentum of the missing trans-
1379 verse energy (MET) is adjusted such that the total 4-momenta of the τ_h and the MET
1380 remains unchanged [60].

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.

1381 5.3.2 Muon energy scale

1382 An energy scale is applied to the p_T and mass of genuine muons from τ decays in the
1383 $e\mu$ and $\mu\tau_h$ channels [61]. The applied values are the same for MC and embedded
1384 samples and are shown in Table 5.2. Following the SM $H \rightarrow \tau\tau$ analysis, Rochester
1385 corrections are not applied, and instead prescriptions from [62] 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 [63] [62].

1386 5.3.3 Electron energy scale

1387 Corrections to the electron energy scale are applied to genuine e from τ decays, and
 1388 are binned in two dimensions by electron p_T and η for barrel vs. endcap [64]. The
 1389 scale factors are binned in p_T and η for MC samples: e.g. values for 2018 are shown
 1390 in Fig. 5.2 from [65]. For embedded samples the electron energy scale is taken as
 1391 only binned in η (Table 5.3).

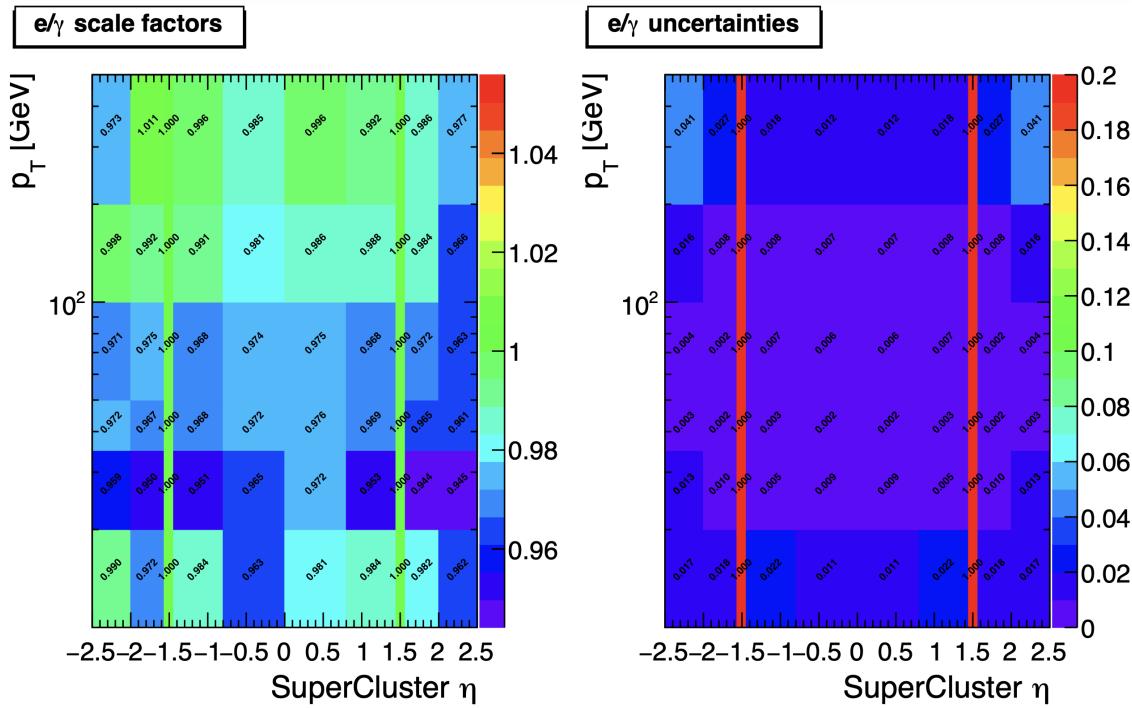


Figure 5.2: Electron/photon energy scale factors (*left*) and corresponding uncertainties (*right*) binned in the electron η and p_T , for the data-taking year 2018 [65].

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 [66] [67] [68].

1392 5.3.4 τ_h identification efficiency

1393 The τ_h identification efficiency can differ in data and MC [60]. Recommended correc-
1394 tions are provided by the Tau POG, and we use the medium DeepTau vs. jet working
1395 point values. The identification efficiency is measured in $Z \rightarrow \tau\tau$ events in the $\mu\tau_h$
1396 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 [60].

1397 5.3.5 Trigger efficiencies

1398 Scale factors are applied to correct for differences in trigger efficiencies between MC
1399 and embedded vs. data, with values taken from tools provided by the Standard Model
1400 $H \rightarrow \tau\tau$ working group which uses the same trigger paths [63]. In the following
1401 sections we review relevant trigger efficiencies in data, which form the basis of the
1402 trigger efficiency corrections applied to MC and embedded.

1403 5.3.6 Tau trigger efficiencies

1404 The efficiencies in data of the single- τ_h leg in $\mu\tau_h$, $e\tau_h$, and di- τ_h triggers is computed
1405 centrally per using a Tag and Probe (TnP) method [69] which is outlined here. In

1406 this method, $Z \rightarrow \tau\tau \rightarrow \mu\tau_h$ are selected in data and a Drell-Yan simulated sample
 1407 ($Z \rightarrow \ell\ell, \ell = e, \mu, \tau_h$) with high purity. Cuts are applied to reject events not in this
 1408 final state, e.g. suppressing $Z \rightarrow \mu\mu$ by vetoing events with a single loose ID muon.
 1409 An isolated muon candidate (the tag) with online $p_T > 27$ GeV and $|\eta| < 2.1$ is
 1410 identified and matched to an offline μ . An offline τ_h candidate (the probe) is selected,
 1411 which is separated from the tag μ , and has $p_T > 20$ GeV and $|\eta| < 2.1$. The probe
 1412 τ_h must pass anti-muon and anti-electron discriminators to avoid fakes from muons
 1413 and electrons, and must pass the medium MVA tau isolation to suppress fakes from
 1414 QCD jets. The trigger efficiency in the TnP method is calculated as

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

1415 The efficiencies for the hadronic tau legs in the relevant channels of this analyses
 1416 ($\mu\tau_h$ and $e\tau_h$) as a function of the offline tau p_T and η , are shown for data taken in
 1417 2016, 2017, and 2018 in Figures 5.3a and 5.3b [69] [70]. In both figures, the different
 1418 HLT thresholds and differences in the L1 seed result in higher efficiencies in 2016 and
 1419 differences in shapes of the 2016 efficiencies compared to 2017 and 2018. The low
 1420 pileup in 2016 also leads to higher efficiencies in that year.

1421 **5.3.7 Single muon trigger efficiencies**

1422 The efficiencies for the single isolated muon trigger with $p_T > 24$ GeV used in this
 1423 analysis, is shown for the data-taking year 2018 in Fig. 5.4a as a function of the muon
 1424 p_T and as a function of the muon $|\eta|$ in Fig. 5.4b from [71]. The data is split with
 1425 respect to a HLT muon reconstruction update that was deployed on 15/05/2018. A
 1426 small asymmetry in efficiencies between negative and positive η in Fig. 5.4b is due to
 1427 disabled muon chambers (CSCs). The efficiencies shown are estimated using a Tag
 1428 and Probe method using $Z \rightarrow \mu\mu$ events, with the tag being an offline muon with

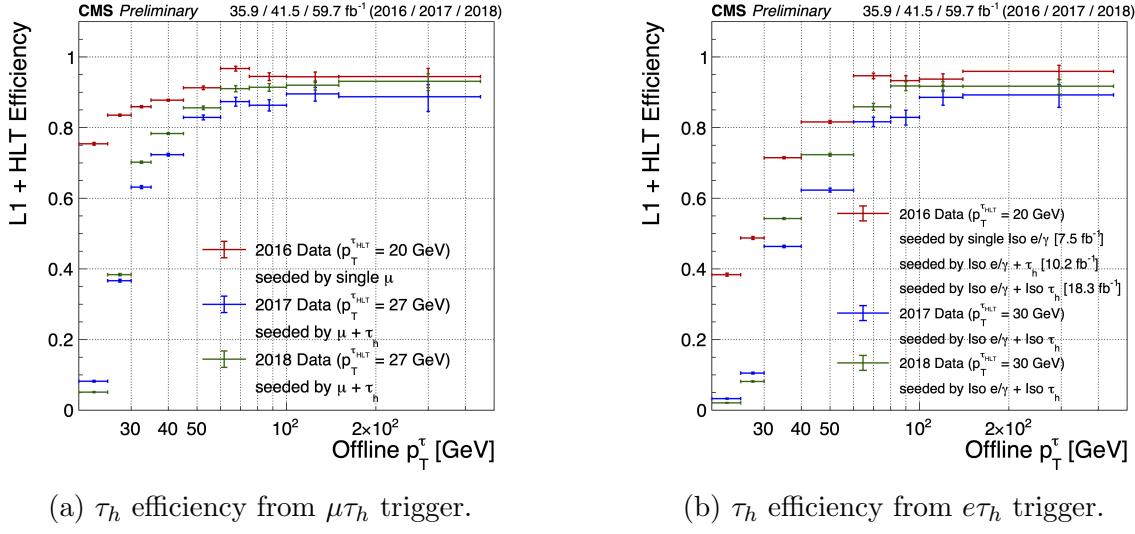


Figure 5.3: Hadronic tau leg efficiency of the cross-triggers for $\mu\tau_h$ (left) and $e\tau_h$ (right) triggers as a function of offline tau p_T for the years 2016 (red), 2017 (blue) and 2018 (green), from [70]. HLT p_T thresholds and L1 seeds are indicated in the legends.

1429 $p_T > 29$ GeV and $|\eta| < 2.4$ passing a tight ID criteria, and the probe is an online (L1)
1430 trigger object with $\Delta R < 0.3$ and passing tight ID and Particle Flow based isolation
1431 requirements with $p_T > 26$ GeV.

1432 5.3.8 Single electron trigger efficiencies

1433 The efficiencies in data, and the ratio between data and MC, of the single electron
1434 HLT trigger with p_T threshold 32 GeV used in this analysis are shown for 2018, as
1435 a function of the electron p_T in Fig. 5.5a and of the electron $|\eta|$ in Fig. 5.5b, from
1436 [72]. In the Tag and Probe method used for the 2018 dataset, the tag is an offline
1437 reconstructed electron with $|\eta| \leq 2.1$ and not in the barrel and endcap overlap region,
1438 with $p_T > 35$ GeV with tight isolation and shower shape requirements, firing the tag
1439 trigger. The probe is an offline reconstructed electron with $|\eta| \leq 2.5$ with $E_T^{\text{ECAL}} > 5$
1440 GeV with no extra identification criteria [72].

1441 The disagreement between data and MC, particularly at low transverse momen-
1442 tum, is in part due to detector effects that are difficult to simulate, such as crystal

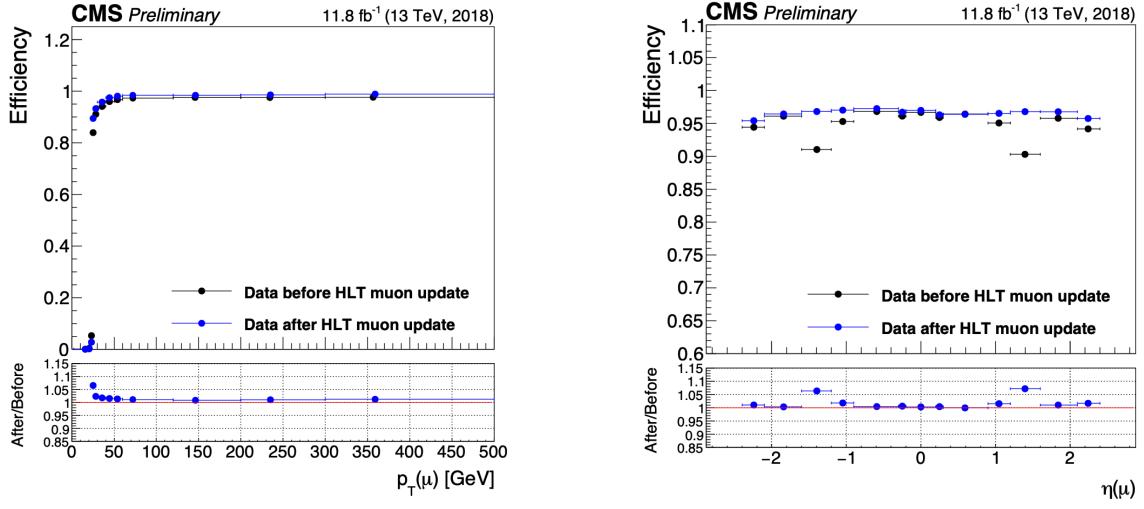
(a) Muon efficiency vs p_T for SingleMuon.(b) Muon efficiency vs $|\eta|$ for SingleMuon.

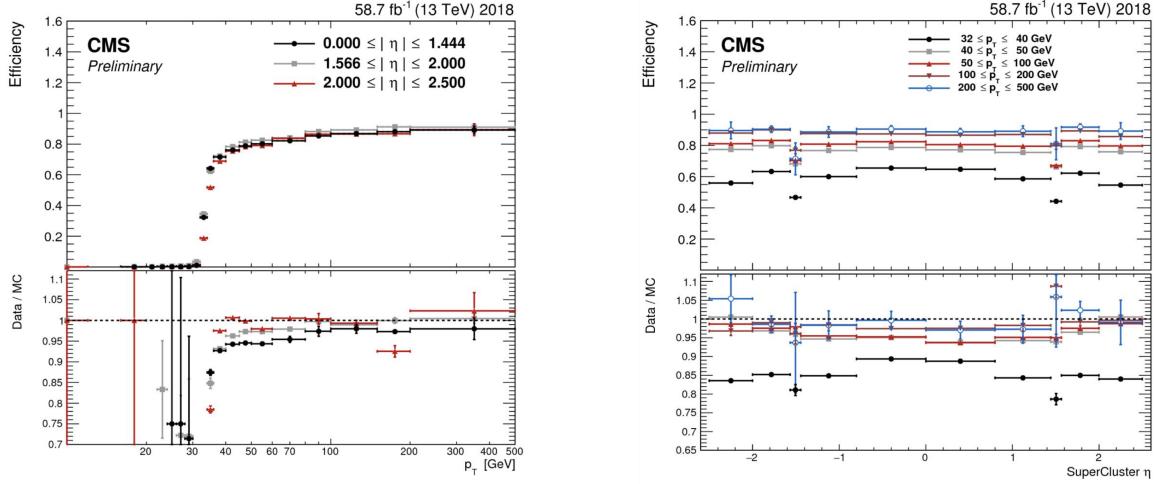
Figure 5.4: Trigger efficiencies in data (*top panels*) and ratio of efficiencies after/before a HLT muon reconstruction update (*bottom panels*) for the muon in the isolated single muon trigger with threshold $p_T > 24$ GeV in the data-taking year 2018, as functions of the muon p_T (*left*) and muon $|\eta|$ (*right*). Only statistical errors are shown [71].

1443 transparency losses in the ECAL and the evolution of dead regions in the pixel tracker
1444 [72].

1445 5.3.9 $e\mu$ cross-trigger efficiencies

1446 The efficiencies of the electron and muons for the cross-trigger with leading muon
1447 used in the $e\mu$ channel are shown for data in 2016, 2017, and 2018 in Figures 5.6a and
1448 5.6b [73]. These efficiencies were measured centrally using a Tag and Probe in events
1449 with Z to dileptons with the same flavor and opposite charge, where the tags are an
1450 isolated muon or electron, and the probe (offline) candidate is required to satisfy the
1451 same lepton selection as that of the tag candidate, be matched within $\Delta R < 0.1$ with
1452 a corresponding online trigger object, and also to pass the cross-trigger. The trigger
1453 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*) [72]. 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.

1454 5.3.10 Electrons and muons faking τ_h : energy scales

1455 Energy scales for electrons misidentified as hadronic tau decays (e faking τ_h) are
 1456 provided by the Tau POG, and were measured in the $e\tau_h$ channel with the visible
 1457 invariant mass of the electron and hadronic tau system [63]. This energy scale is
 1458 applied for τ_h with $p_T > 20$ GeV regardless of which DeepTau vs. electron working
 1459 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 [63].

1460 No nominal energy scale is applied for muons mis-reconstructed as τ_h , and the

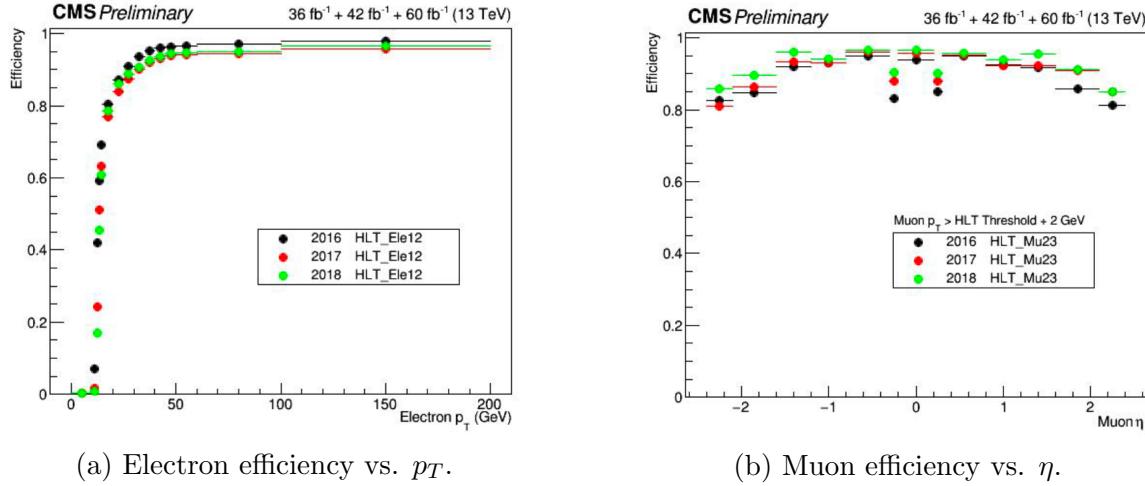


Figure 5.6: Efficiencies of the electron leg vs. p_T (*left*) and the muon leg vs. η (*right*), for the HLT path with online thresholds of 12 GeV for the electron and 23 GeV for the muon, for the data-taking years 2016 (*black*), 2017 (*red*), and 2018 (*green*) [73].

1461 uncertainty is treated as $\pm 1\%$ and uncorrelated in the reconstructed decay mode
1462 [63].

1463 5.3.11 Electrons and muons faking τ_h : misidentification effi- 1464 ciencies

1465 Corrections on identification efficiencies are applied to genuine electrons and muons
1466 misidentified as τ to account for differences in data and MC.

1467 The specific values depend on the vs. electron and vs. muon discriminator working
1468 points used. For misidentified $\mu \rightarrow \tau_h$, the scale factors are split into different $|\eta|$
1469 regions, determined by the CMS muon and tracker detector geometries, as shown in
1470 Table 5.6 for 2018 [60].

1471 For misidentified $e \rightarrow \tau_h$, the scale factors are split into barrel and endcap regions,
1472 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 $|\eta|$ [60].

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 $|\eta|$ [60].

5.3.12 Electron ID and tracking efficiency

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

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

The tracking efficiencies in data and the data/MC ratio are shown in Fig. 5.7b for the Gaussian-sum filter (GSF) tracking [74].

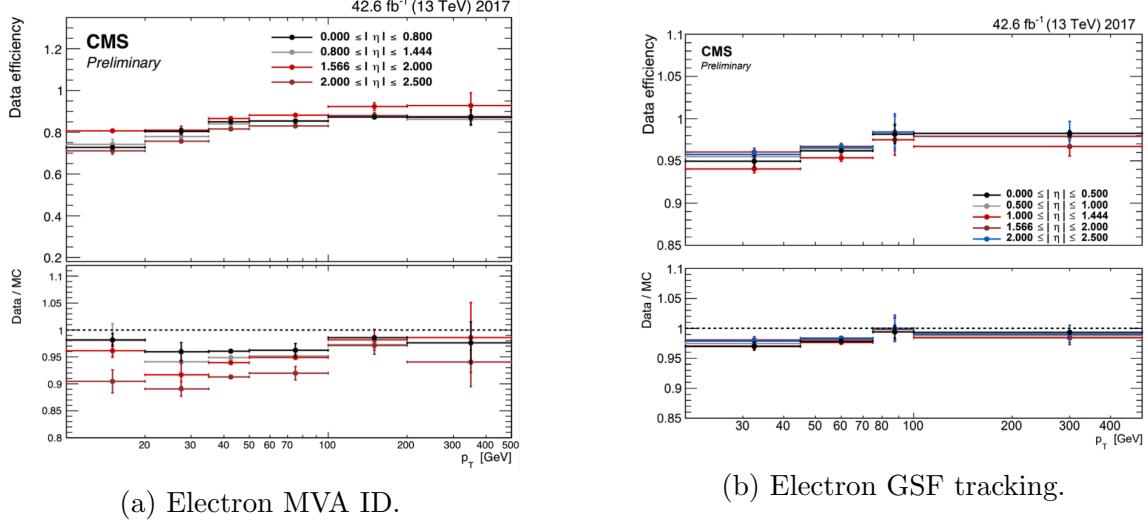


Figure 5.7: Efficiencies in data (*top panels*) and the ratio of efficiencies in data/MC (*bottom panels*), for the electron multivariate analysis (MVA) identification (*left*) and for the Gaussian-sum filter (GSF) tracking (*right*) [74]. Error bars represent statistical and systematic uncertainties.

5.3.13 Muon ID, isolation, and tracking efficiencies

1484 Scale factors are applied to MC to correct for differences between MC and data in
 1485 the performance of muon identification, isolation, and tracking, as detailed below.

1487 The efficiencies for muon identification measured in 2015 data and MC simulation
 1488 are shown in Figures 5.8a and 5.8b for the loose ID and tight ID respectively [75].
 1489 The loose ID is chosen such that efficiency exceeds 99% over the full η range, and the
 1490 data and simulation agree to within 1%. The tight ID is chosen such that efficiency
 1491 varies between 95% and 99% as a function of η , and the data and simulation agree
 1492 to within 1-3%. The muon identification working point used in this analysis is the
 1493 medium ID, which has an efficiency of 98% for all η and an agreement within 1-2%
 1494 [75].

1495 The efficiencies in data for the muon isolation, as measured in Level-3 muons
 1496 (muons in one of the final stages of reconstruction in the HLT), as a function of the
 1497 muon p_T and $|\eta|$ are shown in Figures 5.9a and 5.9b [75]. The HLT muon reconstruc-
 1498 tion consists of two steps: Level-2 (L2), where the muon is reconstructed in the muon

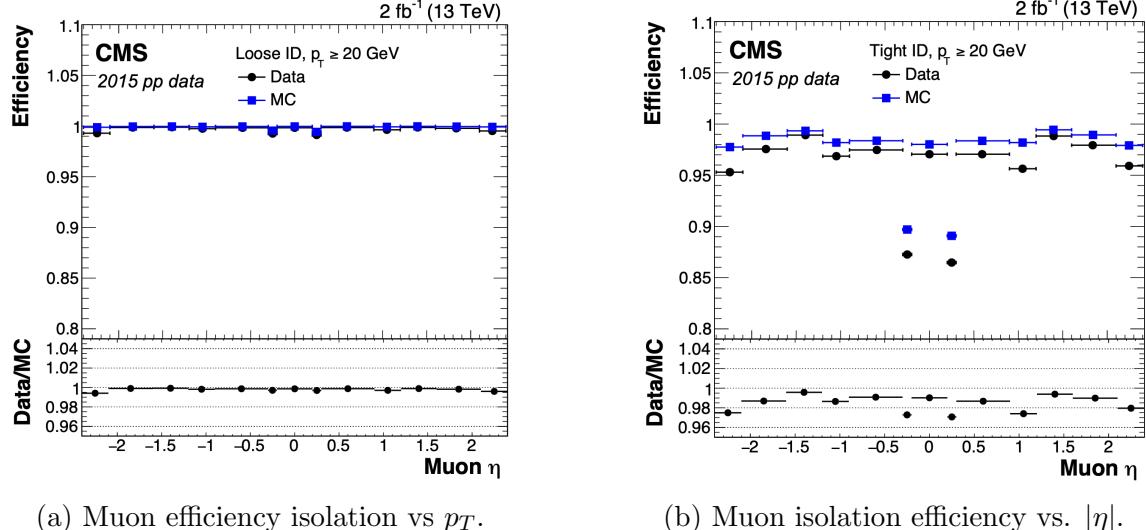


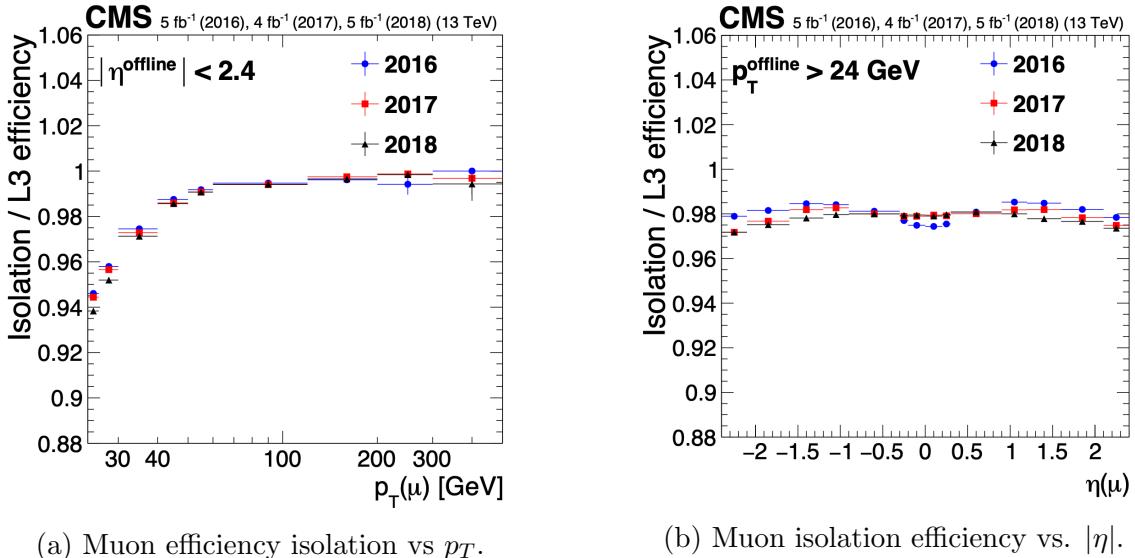
Figure 5.8: Muon identification efficiencies in 2015 data and MC as a function of the muon p_T for the loose ID (*left*) and tight ID (*right*) working points [75].

1499 subdetectors only, and Level-3 (L3) which is a global fit of tracker and muon hits (i.e.
 1500 the global muon reconstruction as described in Section 5.1.2) [76].

1501 The muon tracking efficiencies as a function of $|\eta|$ for standalone muons (i.e. tracks
 1502 from only the muon system, i.e. DT, CSC, and RPC, as discussed in Section 5.1.2),
 1503 is shown for data and simulated Drell-Yan samples in Fig. 5.10 [77].

1504 5.3.14 Recoil corrections

1505 In proton-proton collisions, W and Z bosons are predominantly produced through
 1506 quark-antiquark annihilation. Higher-order processes can induce radiated quarks or
 1507 gluons that recoil against the boson, imparting a non-zero transverse momentum to
 1508 the boson [78]. Recoil corrections accounting for this effect are applied to samples
 1509 with W+jets, Z+jets, and Higgs bosons [63]. The corrections are performed on the
 1510 vectorial difference between the measured missing transverse momentum and the total
 1511 transverse momentum of neutrinos originating from the decay of the W, Z, or Higgs
 1512 boson. This vector is projected onto the axes parallel and orthogonal to the boson
 1513 p_T . This vector, and the resulting correction to use, is measured in $Z \rightarrow \mu\mu$ events,



(a) Muon efficiency isolation vs p_T .

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

Figure 5.9: Muon isolation efficiencies in Run-2 data with respect to Level-3 muons (one of the final stages of HLT muon reconstruction) as a function of the muon p_T (*left*) and $|\eta|$ (*right*) [75].

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 [79]. Per-event weights derived by the 2016 data-only version of this analysis [79] 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

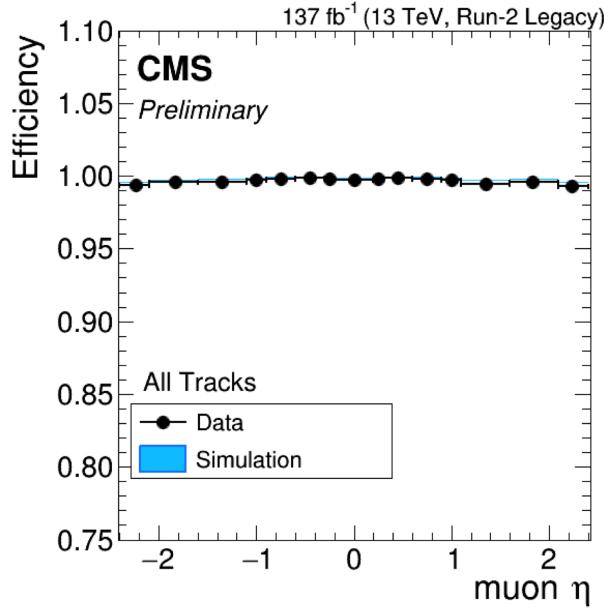


Figure 5.10: Muon tracking efficiencies as a function of $|\eta|$ for standalone muons in Run-2 data (*black*) and Drell-Yan MC simulation (*blue*) [77]. All Tracks refers to tracks which exploit the presence of muon candidates in the muon system to seed the track reconstruction in the inner tracker, in contrast to tracks that use tracker-only hits for seeding. Uncertainties shown are statistical.

1527 MC samples used is provided centrally by the Luminosity POG [80].

1528 5.3.17 Pre-firing corrections

1529 In 2016 and 2017 data-taking, a gradual timing shift of ECAL was not properly
1530 propagated to L1 trigger primitives (TPs), resulting in a large fraction of high η
1531 TPs being incorrectly associated with the previous bunch crossing. L1 trigger rules
1532 prevent two consecutive bunch crossings from firing, causing events to be rejected if
1533 significant ECAL energy was deposited in $2.0 < |\eta| < 3.0$. To account for this issue,
1534 MC simulations for 2016 and 2017 are corrected using an event-dependent weight.
1535 Embedded samples are not corrected [45].

1536 **5.3.18 Top p_T spectrum reweighting**

1537 In Run-1 and Run-2 it was observed that the p_T spectra of top quarks in $t\bar{t}$ data
1538 was significantly softer than those predicted by MC simulations [81]. Possible sources
1539 of this discrepancy are higher order QCD and/or electroweak corrections, and non-
1540 resonant production of $t\bar{t}$ -like final states. To account for this, corrections derived
1541 from Run-2 data by the Top Physics Analysis Group (PAG) are applied to the p_T
1542 of the top and anti-top quarks in MC simulations, computed as a function of their
1543 generator-level p_T [81].

1544 **5.3.19 B-tagging efficiency**

1545 In order to predict correct b-tagging discriminant distributions and event yields in
1546 data, the weight of selected MC events is reweighed according to recommendations by
1547 the BTV POG [82]. The reweighting depends on the jet p_T , η , and the b-tagging dis-
1548 criminant. In this method, there is no migration of events from one b-tag multiplicity
1549 bin to another.

1550 **5.3.20 Jet energy resolution and jet energy smearing**

1551 Calibration of jet energies, i.e. ensuring that the energy and momentum of the recon-
1552 structed jet matches that of the quark/gluon-initiated jet, is a challenging task due
1553 to time-dependent changes in the detector response and calibration and high pileup
1554 [83] [84]. Jet calibration is done via jet energy corrections (JECs) applied to the p_T
1555 of jets in MC samples, accounting successively for the effects of pileup, uniformity of
1556 the detector response, and residual data-simulation jet energy scale differences [85].
1557 Typical jet energy resolutions reported at $\sqrt{s} = 8$ TeV in the central rapidities are
1558 15-20% at 30 GeV and about 10% at 100 GeV [83]. Jet energy corrections are also
1559 propagated to the missing transverse energy.

1560 Measurements show that the jet energy resolution (JER) in data is worse than
1561 in simulation, and so the jets in MC need to be smeared to describe the data. JER
1562 corrections are applied after JEC on MC simulations, and adjust the width of the p_T
1563 distribution based on pileup, jet size, and jet flavor [86]. Tools for applying JEC and
1564 JER are provided centrally by the JER Corrections group.

1565 **Chapter 6**

1566 **Event selection**

1567 **6.1 General procedure for all channels**

1568 For the search for $h \rightarrow aa \rightarrow bb\tau\tau$, three final states of the $\tau\tau$ system are considered:
1569 $\mu\tau_h$, $e\tau_h$, and $e\mu$. The $\tau_h\tau_h$ final state is not considered because signal events in the
1570 $\tau_h\tau_h$ channel would typically produce hadronic taus with momenta below data-taking
1571 trigger thresholds.

1572 In all three final states, events are required to have at least one b-tag jet passing the
1573 medium working point of the DeepFlavour tagger, with $p_T > 20$ GeV, and $|\eta| < 2.4$.
1574 A second b-tag jet is not required because such a requirement would reduce signal
1575 acceptance by 80% compared to only requiring one b-tag jet.

1576 Events in MC samples are sorted into one of the three $\tau\tau$ channels if they pass the
1577 following trigger requirements and requirements on the offline reconstructed objects
1578 in the event, first checking the HLT paths for the $\mu\tau_h$ channel, then $e\tau_h$, and finally
1579 $e\mu$. The two leading leptons (e.g. muon and hadronic tau for the $\mu\tau_h$ channel) that
1580 were determined to have originated from the $\tau\tau$ decay, are called the $\tau\tau$ “legs” and
1581 are respectively subscripted 1 and 2 in this work. For events in data and embedded
1582 samples, the HLT paths requirements for the corresponding channel are checked.

1583 After sorting events by HLT paths and identifying the leading tau legs in the offline
 1584 reconstructed objects, the p_T of the offline objects is checked against the online trigger
 1585 thresholds. Trigger matching is also performed, which checks the correspondence
 1586 between each offline reconstructed object used in the analysis (e.g. a muon), and a
 1587 trigger object in the HLT (e.g. a HLT muon). An offline object is considered to be
 1588 matched, if it corresponds to a trigger object of the same object type, with $\Delta R < 0.5$.
 1589 This matched trigger object is also required to pass the filter(s) of the HLT trigger.
 1590 The trigger thresholds used for the $bb\tau\tau$ final state (the focus of this work) and the
 1591 $bb\mu\mu$ 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.

1592 Further cuts are made on the offline objects in each channel to obtain the signal
 1593 region, or other data regions used to perform data-driven background estimations.

1594 6.2 Event selection in the $\mu\tau_h$ channel

1595 In all three years, a single muon trigger is used if the muon has sufficiently high p_T ,
 1596 otherwise a dilepton $\mu\tau_h$ cross-trigger is used (Tables 6.2, 6.3, and 6.4). For data

1597 taken in 2017-2018 (2016), the logical OR of the single muon triggers with online p_T
1598 thresholds 24 and 27 (23) GeV is used, with the corresponding offline muon required
1599 to have with p_T 1 GeV above the online threshold. For data taken in 2017-2018
1600 (2016), a dilepton $\mu + \tau_h$ cross-trigger with p_T thresholds of 20 (19) and 27 (20) GeV
1601 for the muon and tau respectively, is used. The τ_h is required to have $|\eta| < 2.3$ if the
1602 single trigger is fired, $|\eta| < 2.1$.

1603 The muon and τ_h are required to have opposite charge and be separated by $\Delta R >$
1604 0.4. The muon is required to have $|\eta| < 2.4$, and the τ_h is required to have $|\eta| < 2.3$
1605 unless a cross-trigger is required, in which case we require $|\eta| < 2.1$ as discussed
1606 above.

1607 The muon is required to pass the medium identification (ID) working point [87],
1608 which is defined by the Muon POG as a loose muon (i.e. a Particle Flow muon that is
1609 either a global or a tracker muon - see Section 5.1.2) with additional requirements on
1610 track quality and muon quality. This identification criteria is designed to be highly
1611 efficiently for prompt muons and for muons from heavy quark decays. In addition to
1612 the ID, for prompt muons it is recommended to apply cuts on the impact parameter
1613 [87]: we apply $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$.

1614 In addition, a cut is applied on the muon relative isolation (defined in Section
1615 5.1.2), to be less than 0.15 in a cone size of $\Delta R = 0.4$, which corresponds to the
1616 Tight Particle Flow isolation requirement [87].

1617 The τ_h is required to pass a cut on its impact parameter of $|\Delta(z)| < 0.2$. The τ_h
1618 is also required to pass the VLoose (Very Loose) DeepTau working point vs. elec-
1619 tron, the Tight DeepTau working point vs. muons, and the VVVLoose and Medium
1620 DeepTau working point vs. jets. Events with taus reconstructed in two of the decay
1621 modes (labeled 5 and 6) are rejected, since these decay modes are meant to recover
1622 3-prong taus, but are only recommended for use in analyses where the benefits in
1623 final significance outweigh the resulting increase in background [60].

1624 For the estimation of the background from jets faking τ_h , which is described in Sec-
1625 tion 7.7, anti-isolated events are selected, by requiring events to pass all the selections
1626 described above, except failing the Medium DeepTau working point vs. jets.

1627 6.3 Event selection in the $e\tau_h$ channel

1628 The HLT trigger paths for the $e\tau_h$ channel are summarized in Tables 6.2, 6.3, and
1629 6.4. Similarly to the $\mu\tau_h$ channel, a single electron trigger is used if the electron has
1630 sufficiently high p_T in 2018 and 2017. For data taken in 2018 (2017), the OR of the
1631 single electron triggers with online p_T thresholds at 32 and 35 (27 and 32) GeV are
1632 used, with the corresponding offline electrons required to have p_T greater than 33
1633 (28) GeV. A $e + \tau_h$ cross-trigger is used for electrons with lower offline p_T between
1634 25 and 33 GeV (25 and 28 GeV). For the 2016 dataset, there is no cross trigger but
1635 only a single electron trigger with online p_T threshold at 25 GeV, which is used if the
1636 offline electron has p_T greater than 26 GeV.

1637 The electron and τ_h are required to have opposite charge and be separated by
1638 $\Delta R > 0.4$. The electron is required to be within $|\eta| < 2.3$ when no cross trigger is
1639 used, and $|\eta| < 2.1$ when the cross trigger is fired. The τ_h is required to have $|\eta| < 2.3$
1640 if no cross trigger is fired, and have $|\eta| < 2.1$ if the cross trigger is fired.

1641 The electron is required to have a relative isolation (same definition as in Section
1642 5.1.2) of less than 0.1 in a cone size of $\Delta R = 0.3$, which is the standard recommended
1643 cone size giving minimal pileup dependence and reduced probability of other objects
1644 overlapping with the cone. The isolation quantity used includes an “effective area”
1645 (EA) correction to remove the effect of pileup in the barrel and endcap parts of the
1646 detector [88].

1647 The electron is also required to pass cuts on its impact parameter of $|\Delta(z)| < 0.2$
1648 and $|\Delta(xy)| < 0.045$. It is also required to pass the non-isolated MVA working point

1649 corresponding to 90% efficiency. The electron's number of missing hits, which are
1650 gaps in its trajectory through the inner tracker [88], must be less than or equal to
1651 1. The electron must pass a conversion veto, which rejects electrons coming from
1652 photon conversions in the tracker, which should instead be reconstructed as part of
1653 the photon [88].

1654 The impact parameter cut for the τ_h is $|\Delta(z)| < 0.2$. In contrast to the $\mu\tau_h$ event
1655 selection, the vs. electron and vs. muon DeepTau working points are flipped, to
1656 reject muons faking the τ_h leg. The τ_h is required to pass the Tight DeepTau working
1657 point vs. electrons, the VLoose DeepTau working point vs. muons, and the Medium
1658 DeepTau working point vs. jets.

1659 As in the $\mu\tau_h$ channel, for the estimation of the background from jets faking τ_h ,
1660 which is described in Section 7.7, anti-isolated events are selected, by requiring events
1661 to pass all the selections described above, except failing the Medium DeepTau working
1662 point vs. jets.

1663 6.4 Event selection in the $e\mu$ channel

1664 The HLT trigger paths for the $e\mu$ channel are summarized in Tables 6.2, 6.3, and
1665 6.4. Events are selected with the logical OR of two $e + \mu$ cross triggers, where either
1666 the electron or muon can have larger p_T : (1) leading electron, where the electron has
1667 online $p_T > 23$ GeV and muon has online $p_T > 8$ GeV, or (2) leading muon, where
1668 electron has online $p_T > 12$ GeV and muon has online $p_T > 23$ GeV.

1669 The leading and sub-leading leptons are required to have an offline p_T greater
1670 than 1 GeV above the online threshold (i.e. $p_T > 24$ GeV). If the sub-leading lepton
1671 is the electron, the offline p_T threshold is 1 GeV above the online threshold ($p_T > 13$
1672 GeV), but if it is a muon, the offline p_T threshold is required to be at least 5 GeV
1673 greater than the online threshold (i.e. $p_T > 13$ GeV). This is because of poor data

1674 and simulation agreement for low- p_T muons with p_T between 9 GeV and 13 GeV, and
1675 the higher probability of mis-identifying jets as muons at lower p_T . With no effect on
1676 the expected limits, the offline p_T threshold for muons is raised to 13 GeV instead of
1677 9 GeV, even though it may lead to loss in signal acceptance. Both the electron and
1678 muon are required to have $|\eta| < 2.4$.

1679 The electron and muon are required to have opposite charge and be separated
1680 by $\Delta R > 0.3$ (note the decreased separation requirement compared to the other
1681 two channels). The electron is required to pass the non-isolated MVA identification
1682 working point corresponding to 90% efficiency, and to have a relative isolation less
1683 than 0.1 for a cone size of $\Delta R = 0.3$ with the EA pileup subtraction correction.
1684 The electron must have one or fewer missing hits and pass the conversion veto (both
1685 described previously in Section 6.3).

1686 The muon is required to pass the medium identification working point (described
1687 earlier in 6.2), and to have a relative isolation less than 0.15 for a cone size of $\Delta R =$
1688 0.4. The muon impact parameter is required to have $|\Delta(z)| > 0.2$ and $|\Delta(xy)| < 0.045$.

1689 For the QCD multijet background estimation described in Section 7.8, the same-
1690 sign region is selected by requiring all the above selections, except the legs are required
1691 to have the same electric charge rather than opposite.

1692 6.5 Extra lepton vetoes in all channels

1693 Events containing a third lepton (electron or muon) that is neither of the leading $\tau\tau$
1694 legs are rejected, and events with di-muons and di-electrons are vetoed, with criteria
1695 taken from the Standard Model $H \rightarrow \tau\tau$ working group [63].

1696 The event is vetoed if a third electron is found with the following properties:
1697 $p_T > 10$ GeV, $|\eta| < 2.5$, impact parameter $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$, passing
1698 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.

1699 hits, and relative isolation < 0.3 with cone size $\Delta R = 0.3$. The event is also vetoed if
1700 a third muon is found with the following properties: $p_T > 10$ GeV, $|\eta| < 2.4$, impact
1701 parameter $|\Delta(z)| < 0.2$ and $|\Delta(xy)| < 0.045$, medium ID, and isolation < 0.3 with
1702 cone size $\Delta R = 0.4$.

1703 A di-muon veto is applied, which rejects events containing a pair of muons with
1704 opposite charge and separation of $\Delta R > 0.15$, that both pass the following selections:
1705 $p_T > 15$ GeV, $|\eta| < 2.4$, flag for global muons, flag for tracker muon, flag for Particle
1706 Flow muon, $|\Delta(z)| < 0.2$, $|\Delta(xy)| < 0.045$, and isolation < 0.3 with cone size $\Delta R =$
1707 0.4 .

1708 A similar di-electron veto is applied to reject events containing a pair of electrons
1709 with opposite charge and separation of $\Delta R > 0.15$, that both pass the following
1710 selections: $p_T > 15$ GeV, $|\eta| < 2.5$, a dedicated electron ID (cut-based) for vetoing
1711 third leptons, $|\Delta(z)| < 0.2$, $|\Delta(xy)| < 0.045$, with pileup-corrected relative isolation
1712 < 0.3 with cone size $\Delta R = 0.3$.

1713 These vetoes on extra leptons also ensure orthogonality of events to analyses such
1714 as the $bb\mu\mu$ final state, whose results are combined with this $bb\tau\tau$ final state as
1715 described in Section ??.

₁₇₁₆

Chapter 7

₁₇₁₇

Background estimation

₁₇₁₈ This section describes methods used to estimate sources of background from Standard
₁₇₁₉ Model processes in the search for $h \rightarrow aa \rightarrow bb\tau\tau$. Similar background estimation
₁₇₂₀ methods are being used for the $h \rightarrow a_1a_2$ analysis. The background contributions
₁₇₂₁ directly taken from MC are described first, followed by backgrounds estimated from
₁₇₂₂ data-driven methods to produce sufficient statistics in the signal region.

₁₇₂₃

7.1 Z+jets

₁₇₂₄ A major source of background for $\tau\tau$ analyses is the Drell-Yan (DY) process (Z+jets).
₁₇₂₅ The Z boson decays to $\tau\tau/\mu\mu/ee$ with equal probability of 3.4% each, with the dom-
₁₇₂₆ inant decay modes being to hadrons (around 70%) and neutrinos (invisible) (20%)
₁₇₂₇ [23].

₁₇₂₈ The Drell-Yan contribution with genuine taus, $Z \rightarrow \tau\tau$, is estimated using embed-
₁₇₂₉ ded samples, described in Section 4.3. To avoid double-counting between embedded
₁₇₃₀ and MC samples, in all MC samples, events with legs that originated from genuine τ
₁₇₃₁ are discarded.

₁₇₃₂ The other decays of the Z, $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$, are estimated from MC simulation,
₁₇₃₃ and are hereafter referred to as simply the Drell-Yan background. These MC samples

1734 are generated to leading order (LO) with different numbers of jets (jet multiplicity) in
1735 the matrix element: Z+1 jet, Z+2jets, Z+3 jets, Z+4 jets, and inclusive Z+jets. The
1736 cross-sections of the samples with ≥ 1 jets are normalized to next-to-NLO (NNLO)
1737 in QCD.

1738 For the inclusive Drell-Yan sample, two samples are used with different thresholds
1739 for the di-lepton invariant mass ($m_{\ell\ell}$) at the generator level: one with $m_{\ell\ell} > 50$ GeV
1740 and the other with $10 < m_{\ell\ell} < 50$.

1741 7.2 W+jets

1742 The dominant W boson decay modes are to hadrons (67.4%), $e + \nu_e$ (10.7%), $\mu + \nu_\mu$
1743 (10.6%), and $\tau + \nu_\tau$ (11.4%) [23]. The W+jets background is estimated from MC
1744 simulation. Similarly to the Z+jets, the W+jets samples are generated with different
1745 jet multiplicities in the matrix element. LO samples are used for greater statistics
1746 and are normalized to NNLO cross sections.

1747 7.3 $t\bar{t}$ + jets

1748 In hadron collisions, top quarks are produced singly with the weak interaction, or in
1749 pairs via the strong interaction, with interference between these leading-order pro-
1750 cesses possible in higher orders of the perturbation theory. The top quark is the
1751 heaviest fermion in the Standard Model and has a short lifetime ($\sim 10^{-25}$ s), decay-
1752 ing without hadronization into a bottom quark and a W boson [23], with the decay
1753 modes of the W boson as listed in the previous section. With two top quarks, the
1754 final states of the two resulting W bosons can be described as fully leptonic, semilep-
1755 tonic, and fully hadronic. These three final states are modeled separately with MC
1756 simulation in 2018 and 2017, while for 2016 the sample used is inclusive.

¹⁷⁵⁷ 7.4 Single top

¹⁷⁵⁸ There are three main production modes of the single top in pp collisions [89]: the
¹⁷⁵⁹ exchange of a virtual W boson (t channel), the production and decay of a virtual W
¹⁷⁶⁰ boson (s channel), and the associated production of a top quark and W boson (tW ,
¹⁷⁶¹ or W-associated) channel. As the s channel process is rare and only 3% of the total
¹⁷⁶² production, the dominant production mode of the t -channel and the tW production
¹⁷⁶³ are considered and modeled with MC.

¹⁷⁶⁴ 7.5 Diboson

¹⁷⁶⁵ In pp collisions, the production of dibosons (pairs of electroweak gauge bosons, i.e.
¹⁷⁶⁶ WW, WZ, and ZZ) is dominated by quark-antiquark annihilation, with a small con-
¹⁷⁶⁷ tribution from gluon-gluon interaction [90]. MC is used to model the pair production
¹⁷⁶⁸ 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
¹⁷⁶⁹ quarks and ℓ being leptons).

¹⁷⁷⁰ 7.6 Standard Model Higgs

¹⁷⁷¹ MC is used to simulate backgrounds from major production modes of the Standard
¹⁷⁷² Model 125 GeV Higgs boson: gluon-gluon fusion (ggH), vector boson fusion (VBF),
¹⁷⁷³ associated production with a W or Z (WH, ZH), and associated production with a
¹⁷⁷⁴ top pair (ttH) (see Fig. 7.1 for leading-order diagrams). For these production modes,
¹⁷⁷⁵ samples with the Higgs decaying to $\tau\tau$ or to WW are used. Samples made with
¹⁷⁷⁶ higher-order diagrams for WH and ZH that include the production of a jet, with the
¹⁷⁷⁷ Higgs decaying to WW, are also used.

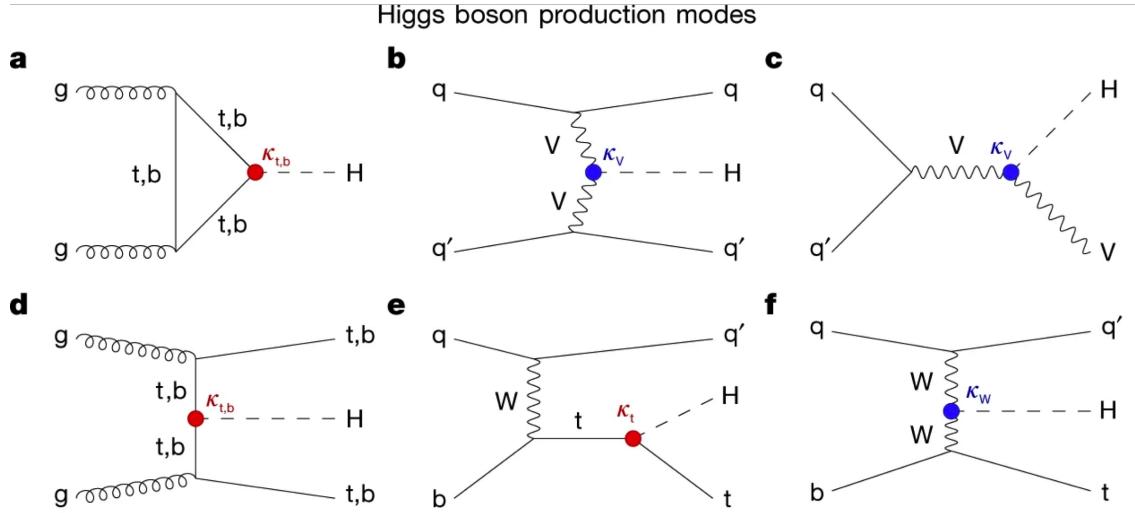


Figure 7.1: Leading-order Feynman diagrams of Higgs production from [91], in ggH (a) and vector boson fusion (VBF; b), associated production with a W or Z (V) boson (VH; c), associated production with a top or bottom quark pair (ttH or bbH); d, and associated production with a single top quark (tH; e, f).

1778 7.7 Jet faking τ_h

1779 Events with a jet mis-reconstructed as the hadronic tau leg τ_h are a major source of
1780 background in the $\mu\tau_h$ and $e\tau_h$ channels. The main processes contributing to jet $\rightarrow \tau_h$
1781 events are QCD multijet, W+jets, and $t\bar{t}$ production. These events are estimated
1782 using a data-driven method adapted from past analyses [45] [79]. This background
1783 includes contributions from W+jets, QCD multijets, and $t\bar{t}$ +jets. To estimate this
1784 background, a sideband region is constructed, where events are required to pass all
1785 baseline $\mu\tau_h/e\tau_h$ selection criteria, but fail the τ_h isolation criteria. The events in
1786 this sideband region are reweighed with a factor $f/(1-f)$, where f is the probability
1787 for a jet to be misidentified as a τ_h . The jet $\rightarrow \tau_h$ background is the anti-isolated,
1788 reweighed MC and embedded events subtracted from the anti-isolated, reweighted
1789 data events.

1790 The fake factor is measured in $Z \rightarrow \mu\mu + \text{jets}$ events in data in the $\mu\mu\tau_h$ final
1791 state, as any reconstructed τ_h in these events must originate from a jet. The two
1792 muons are required to be isolated (< 0.15), have opposite electric charge, and have

1793 an invariant mass between 76 and 106 GeV (close to the Z mass). These events are
1794 selected with a double muon trigger, with the leading muon having offline $p_T > 20$
1795 GeV and the subleading muon $p_T > 10$ GeV. Simulated diboson (ZZ and WZ) events
1796 are subtracted to avoid contamination from events with real τ_h . The denominator of
1797 the fake rate corresponds to fake taus passing the VVVLoose working point of the
1798 discriminator vs. jets, while the numerator corresponds to those passing the Medium
1799 working point, i.e. $f = N_{\text{jet passing tight}} / N_{\text{jet passing loose}}$.

1800 f is measured as a function of the τ_h transverse momentum and is 8% - 10% in
1801 each of the data-taking years. f is derived separately for the $\mu\tau_h$ and $e\tau_h$ channels
1802 because the channels use different anti-lepton identification working points.

1803 7.8 QCD multijet background

1804 In the $e\mu$ channel, events with jets faking electrons or muons originating from QCD
1805 multijet, is estimated from data events with the same baseline selection as in the
1806 signal region, except with same-signed (SS) charged $e + \mu$, ensuring orthogonality
1807 with the signal region which requires opposite-sign (OS) $e\mu$ pairs. All same-sign MC
1808 events (both events with real and fake $e + \mu$) are subtracted from same-sign data
1809 events to remove contamination from other backgrounds. i.e. $\text{QCD}_{\text{SS}} = \text{Data}_{\text{SS}} -$
1810 MC_{SS} .

1811 Three scale factors are applied to the QCD_{SS} events to compute the QCD multijet
1812 background [79] [39]:

- 1813 • *OS-to-SS scale factor*: This scales the SS QCD to the OS region, and is mea-
1814 sured from an orthogonal region with an isolated electron and an anti-isolated
1815 muon. Only the muon is chosen to be anti-isolated because this scale factor was
1816 observed to depend more strongly on electron isolation than that of the muon.
1817 This scale factor is treated as a function of the ΔR separation of the trajectories

1818 of the electron and muon, and is measured separately for events with 0 jets, 1,
1819 jet, and greater than 1 jet.

- 1820 • *2D closure correction for the lepton p_T :* This factor accounts for subleading
1821 dependencies of the first scale factor on the p_T of the two leptons. A 2D weight
1822 is derived in a similar fashion, as a ratio of QCD_{OS} events to QCD_{SS} events,
1823 but parameterized by both electron and muon p_T , where the SS events have the
1824 previous scale factor applied.
- 1825 • *Isolation correction for the muon:* The third and final factor is an isolation
1826 correction, which is a bias correction to account for the fact that the fake
1827 factor was determined for less-isolated muons. This factor is obtained as the
1828 ratio of the OS-to-SS scale factors measured in two other control regions: (1)
1829 events where the electron is anti-isolated ($0.15 < \text{iso} < 0.5$) and the muon is
1830 isolated, and (2) events where both leptons are anti-isolated.

¹⁸³¹ Chapter 8

¹⁸³² Systematic uncertainties

¹⁸³³ The handling of systematic uncertainties is separated into normalization uncertainties
¹⁸³⁴ (those that affect the total yield of a variables' distribution) and shape uncertainties
¹⁸³⁵ (those that shift the distribution of events). Normalization uncertainties are expressed
¹⁸³⁶ as multiplicative factors, while shape uncertainties are represented as up and down
¹⁸³⁷ shifts of a variable's distribution.

¹⁸³⁸ Up/down shifts of shape uncertainties can change the number of background
¹⁸³⁹ events in a distribution. For instance, hadronic taus receive corrections from the
¹⁸⁴⁰ nominal tau energy scale, with the nominal, up, and down energy scales provided
¹⁸⁴¹ centrally by CMS. For the $\mu\tau_h$ channel, an event could have a τ_h with p_T just below
¹⁸⁴² the offline threshold of 20 GeV (for instance, 19.5 GeV), so in the nominal distribution
¹⁸⁴³ of $m_{\tau\tau}$ (or any other variable for this channel), the event is excluded. However, when
¹⁸⁴⁴ we build our distributions with the tau energy scale “up” shift, the energy of this τ_h
¹⁸⁴⁵ may be scaled up to, say, 20.5 GeV, and now the event passes the offline p_T threshold
¹⁸⁴⁶ for the single muon trigger, leading to the event's inclusion in the distributions made
¹⁸⁴⁷ with the tau energy scale “up” shift.

¹⁸⁴⁸ In evaluating the up and down shifts of a specific source of uncertainty, all other
¹⁸⁴⁹ corrections and scale factors are held at their nominal values, and the full chain

1850 of object and event selection and event categorization is performed to obtain the
1851 observable distributions. Any “downstream” variables that depend on the shifted
1852 variable, e.g. the invariant di-tau mass $m_{\tau\tau}$, must be computed for the nominal case,
1853 and then re-computed separately for each up and down shift of the tau legs’ energy
1854 scale. The objective of this process is to quantify the effect of a single source of
1855 uncertainty on the resulting observable distributions.

1856 8.1 Uncertainties associated with physics objects

1857 Each scale factor and correction described in Section 5.3 has an associated uncertainty.
1858 The binning of the uncertainties follows that of the nominal scale factor value.

1859 8.1.1 Uncertainties in the lepton energy scales

1860 The uncertainties in the tau energy scales [60] are binned by the tau decay mode and
1861 are taken as shape uncertainties treated as uncorrelated across the tau decay modes
1862 and years. Same as with the application of the nominal scale factor, when applying
1863 the up or down shifts, the missing transverse energy (p_T^{miss}) of the event is adjusted
1864 so that the 4-vector sum of the tau p_T^{miss} is unchanged.

1865 The uncertainties in the muon energy scale [61] are 0.4% for $|\eta| < 1.2$, 0.9% for
1866 $1.2 < |\eta| < 2.1$, and 2.7% for $2.1 < |\eta| < 2.4$, and are treated as shape uncertainties,
1867 fully uncorrelated between embedded and MC samples.

1868 The uncertainties in the electron energy scale [64] in MC are binned in the electron
1869 $|\eta|$ and p_T , and are shown in Fig. 5.2. The uncertainties range from 0.5% to 2.2% in
1870 the barrel, and 0.3% to 4.1% in the endcap, across the p_T range. The uncertainties
1871 for the embedded sample are binned only in $|\eta|$ and are on the order of 0.5% and
1872 1.25% for the barrel and endcap [68].

1873 There are also uncertainties in the energy scales for electrons and muons misiden-

1874 tified as τ_h . The uncertainty for muons misidentified as τ_h is 1% [60]. For electrons
1875 misidentified as τ_h , the uncertainty is binned in barrel/endcap η and by 1-prong and
1876 1-prong + π_0 decays. The probability for e/μ faking a 3-prong decay mode is much
1877 lower.

1878 8.1.2 Uncertainties from other lepton corrections

1879 Uncertainties associated with the τ_h identification efficiencies are treated as shapes,
1880 uncorrelated across the seven p_T bins and years. The shape uncertainties in the
1881 embedded samples are taken as 50% correlated with those of the MC samples.

1882 The uncertainties on electron and muon identification efficiencies are taken as
1883 normalization uncertainties of 2% each, with a 50% correlation between embedded
1884 and MC samples.

1885 In the $e\tau_h$ channel, there is an additional uncertainty for the vs. jet discrimination
1886 efficiency [60], because the analysis uses a looser anti-lepton working point (VLoose
1887 WP) than the working points used in the measurement of the efficiency (namely,
1888 VLoose WP vs e, and Tight WP vs mu). For nominal $\tau_h p_T < 100$ GeV, an additional
1889 uncertainty of 3% (5%) is used in MC (embedded), and for high p_T an uncertainty of
1890 15% is used for both.

1891 The uncertainties in trigger efficiencies are taken as shapes [60]. In the $e\tau_h$ and $\mu\tau_h$
1892 channels, there are uncertainties for the single and cross lepton triggers, and in the
1893 $e\mu$ channel there is one uncertainty each for the two $e + \mu$ triggers, and one combined
1894 uncertainty since their trigger phase spaces are not mutually exclusive.

1895 8.1.3 Uncertainties from jet energy scale and resolution

1896 The jet energy scale uncertainties are taken as shape uncertainties: there are eleven
1897 in total, with seven correlated across years (labeled “Year” below) and the remainder
1898 uncorrelated across years. They affect the b-tag jet p_T and mass, and hence the

1899 missing transverse energy p_T^{miss} . The shifts are propagated through the b-tagging
1900 scale factor calculation and b-tag jet counting.

1901 The uncertainties in the jet energy correction and resolution [83] [92] are as follows:

1902 • *Absolute, AbsoluteYear*: flat absolute scale uncertainties.

1903 • *BBEC1, BBEC1Year*: for sub-detector regions, with barrel “BB” in $|\eta| < 1.3$
1904 and endcap region 1 “EC1”: $1.3 < |\eta| < 2.5$.

1905 • *EC2, EC2 year*: for sub-detector regions, with endcap region 2 “EC2” in $2.5 <$
1906 $|\eta| < 3.0$.

1907 • *HF, HF year*: for sub-detector regions, with hadron forward “HF” in $|\eta| > 3$.

1908 • *FlavorQCD*: for uncertainty in jet flavor (uds/c/b-quark and gluon) estimates
1909 based on comparing Pythia and Herwig (different MC generator) predictions.

1910 • *RelativeBal*: account for difference between log-linear fits of the two methods
1911 used to study the jet energy response: MPF (missing transverse momentum
1912 projection fraction) and p_T balance.

1913 • *RelativeSample*: account for η -dependent uncertainty due to a difference be-
1914 tween relative residuals, observed with dijet and Z+jets in Run D of 2018 data.

1915 • *JetResolution*: uncertainty in the jet energy resolution.

1916 8.1.4 Uncertainties from b-tagging scale factors

1917 The b-tagging scale factor has its own set of associated uncertainties (not to be
1918 confused with shifts in the b-tagging scale factor due to the propagation of the jet
1919 energy scale uncertainties described in the previous section 8.1.3). They are:

1920 • *hf*: contamination from heavy flavor (b+c) jets in the light flavor region.

1921 • *hfstats1, hfstats2*: linear and quadratic statistical fluctuations from b-flavor jets.

1922 • *lf*: contamination from light flavor (udsg+c jets) in the heavy flavor region.

1923 • *lfstats1, lfstats2*: linear and quadratic statistical fluctuations from udsg jets.

1924 • *cferr, cferr2*: uncertainty for charm jets.

1925 The variations for “lf, hf, hfstats1/2, lfstats1/2” are applied to both b and udsg jets.

1926 For c-flavor jets, only “cferr1/2” is applied.

1927 8.1.5 Uncertainties from MET

1928 Samples where recoil corrections were applied (Z+jets, W+jets, and Standard Model
1929 Higgs, as described in Section 5.3) have uncertainties from the response and resolution
1930 of the hadronic recoil against the leptonic system. These are each binned in jet
1931 multiplicity.

1932 8.2 Uncertainties associated with samples used

1933 Normalization uncertainties related to the samples used are:

1934 • *Cross-section uncertainties*: $\sigma(t\bar{t})$: 4.2%, $\sigma(\text{diboson})$: 5%, $\sigma(\text{single top})$: 5%,
1935 $\sigma(\text{ggH})$: 3.2%, $\sigma(\text{qqH})$: 2.1%, $\sigma(\text{WH})$: 1.9%, $\sigma(\text{ZH})$: 1.3%, $\sigma(\text{ttH})$: 3.6%

1936 • *Uncertainties in QCD renormalization scale*: QCD scale(qqH): +0.43%-0.33%,
1937 QCD scale(WH): +0.5%-0.7%, QCD scale(ttH): +5.8%-9.2%

1938 • *Branching ratio uncertainties*: $\text{BR}(\text{H} \rightarrow \tau\tau)$: 1.8%, and $\text{BR}(\text{H} \rightarrow \text{WW})$: 1.5%.

1939 • *Normalization uncertainties*: 2% for Drell-Yan, 4\$ for embedded, 20% pre-fit
1940 for the QCD multijet background in the $e\mu$ channel, 20% pre-fit for the jet
1941 faking background.

1942 The $t\bar{t}$ process has additional acceptance uncertainties from QCD scale variation
1943 and parton shower uncertainties [93]. Parton shower uncertainties originate from
1944 the modeling of perturbative and non-perturbative QCD effects handled in parton
1945 shower MC generators. The scale variations are determined from the envelope of the
1946 6 provided shapes due to variations in the factorization scale, renormalization scale,
1947 and their combined variation [93].

1948 The Z p_T reweighing uncertainty in Drell-Yan samples is taken to be 10% of the
1949 nominal value, taken as a shape uncertainty.

1950 The fake rate uncertainties are taken as shape uncertainties. For the weight ap-
1951 plied to scale up anti-isolated events in cross-trigger regions, 20% of the nominal
1952 weight is taken as a shape uncertainty.

1953 8.3 Other uncertainties

1954 A 3.6% yield uncertainty in the signal is used to cover uncertainties in the parton
1955 distribution functions, α_s (fine structure constant), and QCD scale.

1956 Normalization uncertainties from luminosity are applied to all MC samples, di-
1957 vided into those uncorrelated across years, those correlated between 2017 and 2018,
1958 and one for 2018 [80].

1959 8.4 Pulls and impacts

1960 The top impacts and pulls computed for the combination of all channels and years is
1961 shown in Fig. 8.1. The top impacts are related to uncertainty in the signal sample and
1962 cross-section of the $t\bar{t}$ cross-section, and also the yields of the jet faking τ_h background,
1963 which is a major background in all channels and expected to be constrained due to
1964 the yield uncertainty which is taken to be 20% pre-fit.

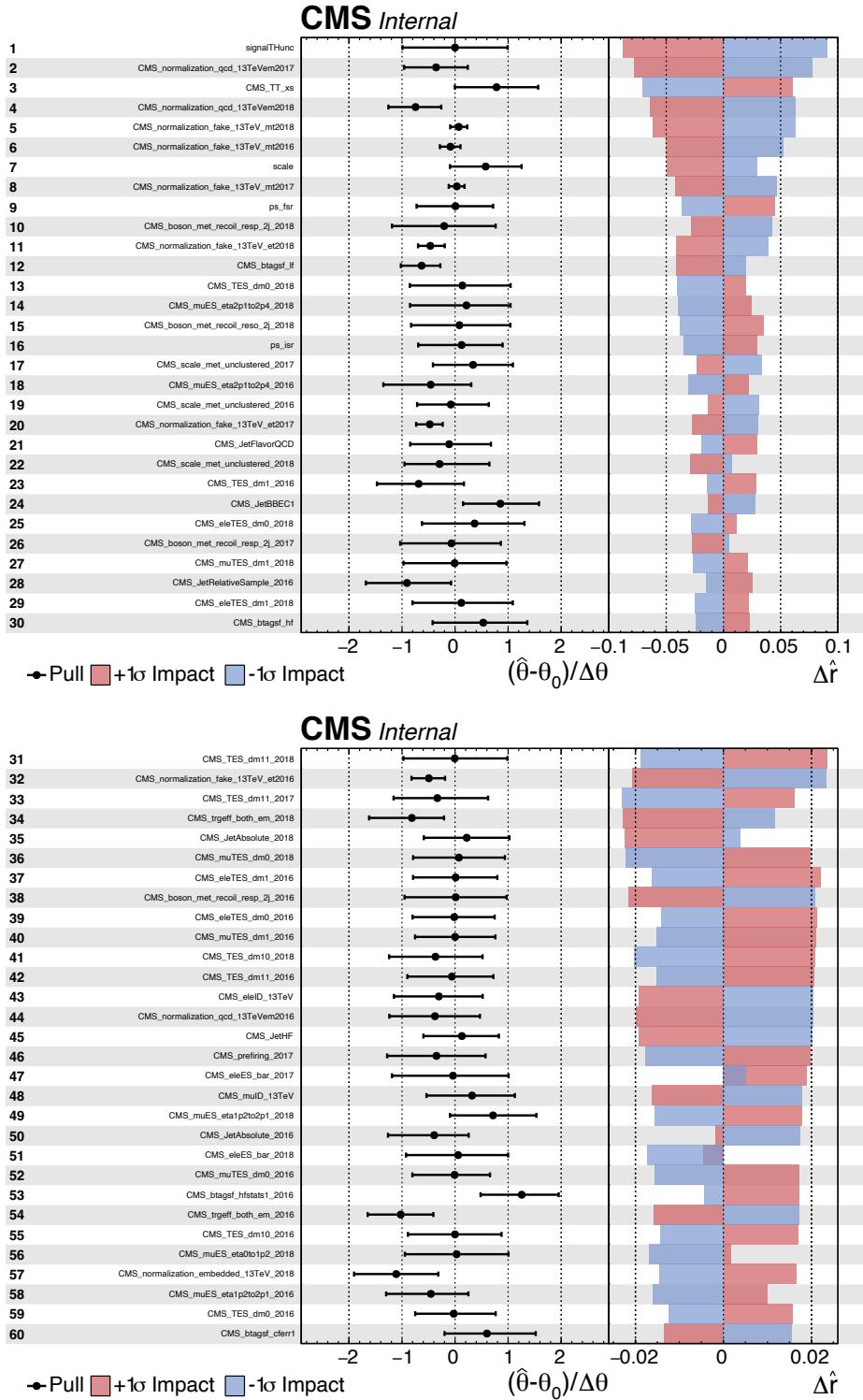


Figure 8.1: Top sixty impacts for the combination of all channels and years [38].

¹⁹⁶⁵ Chapter 9

¹⁹⁶⁶ Event categorization and signal ¹⁹⁶⁷ extraction

¹⁹⁶⁸ 9.1 B-tag jet multiplicity

¹⁹⁶⁹ The increased statistics of the full Run-2 dataset enables the separation of events into
¹⁹⁷⁰ events with exactly 1 b-tag jet and events with greater than 1 b-tag jet. Further event
¹⁹⁷¹ categorization is performed with deep neural networks (DNNs) described below. The
¹⁹⁷² DNNs are used only for separating events into signal and control regions in the 1
¹⁹⁷³ b-tag and 2 b-tag jets scenarios. The final results are extracted from the statistical
¹⁹⁷⁴ fitting to the mass of the $\tau\tau$, $m_{\tau\tau}$.

¹⁹⁷⁵ 9.2 DNN-based event categorization

¹⁹⁷⁶ A brief overview of the DNN-based event categorization is given below with a focus
¹⁹⁷⁷ on the physics aspects, with full details of the machine learning training in [39] and
¹⁹⁷⁸ associated documentation.

1979 **Training samples**

1980 Neural networks for event categorization are trained for each of the $\mu\tau_h$, $e\tau_h$, and $e\mu$
1981 channels, for 1 and 2 b-tag jets, giving $3 \times 2 = 6$ networks in total. In the training,
1982 the signal is taken to be all of the possible pseudoscalar mass m_a hypotheses together.
1983 The backgrounds for each DNN are taken to be a representative combination of the
1984 three major backgrounds: $Z \rightarrow \tau\tau$, $t\bar{t}$ +jets, and fake backgrounds. The proportions of
1985 each background for each channel and b-tag jet multiplicity are taken from the yields
1986 in the $m_{\tau\tau}$ distribution. For instance, in the $\mu\tau_h$ 1 b-tag jet category, the composition
1987 of the background for training is 17.4% from $Z \rightarrow \tau\tau$, 42.4% from $t\bar{t}$ +jets, and 40.2%
1988 fakes.

1989 **Input variables**

1990 The input variables capture the key differences between the signal and the back-
1991 ground:

- 1992 • Transverse momentum p_T of the electron and muon in the $e\tau_h$ and $\mu\tau_h$ channels,
1993 where the signal tends to have a softer p_T spectrum (lower energy) than the
1994 background.
- 1995 • p_T of the b-tag jet(s). The signal sample b-tag jet(s) tend to have softer p_T .
- 1996 • Invariant masses of the various objects ($\tau\tau$ legs and the b-tag jet(s)), which
1997 tend to be smaller for the signal samples.
- 1998 • The angular separation ΔR between pairs of the objects, where signal samples
1999 peak at smaller ΔR values.
- 2000 • The transverse mass between the missing transverse energy p_T^{miss} and each of

2001 the four objects [79], defined as

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

2002 where p_T^ℓ is the transverse momentum of the object ℓ , and $\Delta\phi$ is the difference
 2003 in azimuthal angle between the object and the p_T^{miss} . Events from $t\bar{t}$ +jets and
 2004 jets faking τ_h backgrounds have larger p_T^{miss} resulting in larger transverse mass
 2005 values compared to the signal, which tends to have smaller p_T^{miss} that is also
 2006 more aligned with the lepton legs.

- 2007 • The variable D_ζ [79], defined as

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

2008 where the ζ axis is the bisector of the transverse directions of the visible τ decay
 2009 products. p_ζ is the component of the p_T^{miss} along the ζ axis, and p_ζ^{vis} is the sum
 2010 of the components of the lepton p_T along the same axis. This variable captures
 2011 the fact that in signal the p_T^{miss} is small and approximately aligned with the $\tau\tau$.
 2012 In contrast, the $Z \rightarrow \tau\tau$ background tends towards large D_ζ values because the
 2013 p_T^{miss} is collinear to the $\tau\tau$, and the $t\bar{t}$ +jets events tend to have small D_ζ due to
 2014 a large p_T^{miss} not aligned with the $\tau\tau$.

- 2015 • For events with 2 b-tag jets, one additional variable is defined to capture the
 2016 difference in the invariant mass of the bb and the $\tau\tau$:

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

2017 This variable peaks at zero for the $h \rightarrow aa \rightarrow 2b2\tau$ signal.

2018 **Categorization using the DNN score**

2019 After training, events in data, MC, and embedded are evaluated with the six DNNs
2020 and assigned a raw score between 0 and 1 (background-like or signal-like). In order
2021 to flatten the distribution of the score and define score thresholds for categorizing
2022 events, the raw output scores are transformed with the function $\tilde{p}(n) = \text{arctanh}(p \times$
2023 $\tanh(n))/n$ where n is a positive integer. The thresholds of the DNN score used for
2024 signal/control region definition are determined using scans that optimize the signal
2025 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.

2026 **9.3 Methodology for signal extraction**

2027 In this section we outline the statistics terminology and concepts underlying the
2028 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.

2029

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 [94]. 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

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

2030

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) [94] 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 [94]. 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 $\bar{x} = \sum_{i=1}^n x_i/n$ [94].

A commonly used estimator in physics is the maximum likelihood estimator (MLE), defined as the value α 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”.

2052 9.3.2 Hypothesis testing

2053 In this section we next introduce concepts related to hypothesis testing such as the
2054 test statistic constructed from the ratio of likelihood functions.

2055 The objective of a likelihood analysis is to distinguish different models repre-
2056 senting the various hypotheses, and determine the one that best explains the ex-
2057 perimental outcome. In a search for new physics, a signal is additive on top of the
2058 background. The background-only hypothesis is the null hypothesis, and the signal-
2059 plus-background hypothesis is the alternative.

2060 As a simple example, take the p -value test, for an experiment where we count
2061 events in the signal region, n_{SR} , and expect ν_B background events and ν_S events from
2062 the signal [94]. Then

- 2063 1. The null hypothesis (H_0), i.e. the background-only hypothesis in this experi-
2064 ment, with the probability modeled by $\text{Poisson}(n_{SR}|\nu_B)$.
- 2065 2. The alternate hypothesis (H_1), i.e. signal-plus-background hypothesis, with the
2066 probability modeled by $\text{Poisson}(n_{SR}|(\nu_B + \nu_S))$.

2067 The compatibility of the observed data ν_{SR}^0 and the null hypothesis, is quantified as
2068 the probability that the background-only hypothesis would produce at least as many
2069 events as was observed. This probability is the p -value:

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

2070 If the p -value is very small, we might reject the null hypothesis. The p -value is not the
2071 probability of the null hypothesis given the data; rather, it expresses the probability
2072 that data with a certain property was obtained, assuming the null hypothesis [94].

2073 The p -value is an example of a test statistic T , which maps the data to a single
2074 real number. The Neyman-Pearson lemma states that out of the infinite possibilities

2075 of choices of test statistic, the uniformly most powerful test statistic is the likelihood
 2076 ratio T_{NP} [94]:

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

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:

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

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

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

2077 where T_0 is the value of T based on the observed data, and $\theta()$ is the Heaviside
 2078 function. The size of the test is conventionally chosen to be 10%, 5%, or 1%. As
 2079 the p -value depends on $\boldsymbol{\alpha}$ (both the POI and NP), the null hypothesis should not be
 2080 rejected if the p -value is larger than the size of the test for any value of the nuisance
 2081 parameters.

2082 9.3.3 Confidence intervals

2083 In an example of the measurement of the Standard Model Higgs boson, $\boldsymbol{\alpha}_{\text{POI}} =$
 2084 $(\sigma/\sigma_{SM}, M_H)$, with σ/σ_{SM} is the ratio of the production cross-section for Higgs with
 2085 respect to its value in the SM, and M_H is the unknown mass of the Higgs, values
 2086 of these parameters outside specific bounds are said to be “excluded at the 95%
 2087 confidence level”. These allowed regions are called confidence levels or confidence
 2088 regions, and the parameter values outside of them are considered excluded [94]. A

2089 95% confidence interval does not mean that there is a 95% chance that the true value
 2090 of the parameter is inside the interval. Rather, a 95% confidence interval covers the
 2091 true value 95% of the time (even though we do not know the true value).

2092 To construct a confidence interval for a parameter α , the Neyman Construction
 2093 is used to invert a series of hypothesis tests; i.e. for each possible value of α , the null
 2094 hypothesis is treated as α , and we perform a hypothesis test based on a test statistic.
 2095 To construct a 95% confidence interval, we construct a series of hypothesis tests with
 2096 size of 5%. The confidence interval $I(\mathcal{D})$ is constructed by taking the set of parameter
 2097 values $\boldsymbol{\alpha}$ where the null hypothesis is accepted:

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

2098 where $T(\mathcal{D})$ is the test statistic, and the last α (not bolded) and the subscript k_α
 2099 refer to the size of the test. A schematic of the Neyman construction is shown in Fig.
 2100 9.1. In a more generalized case, the x -axis is the test statistic T .

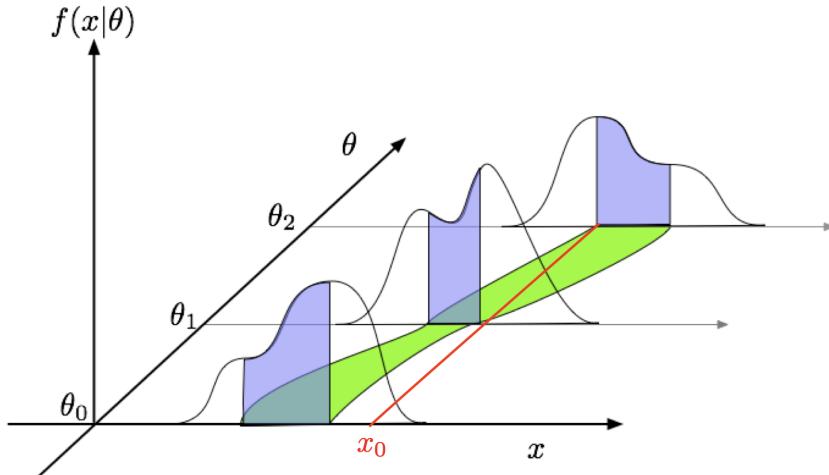


Figure 9.1: Schematic of the Neyman construction for confidence intervals [94]. For each value of θ , we find a region in x where $\int f(x|\theta)dx$ satisfies the size of the test (blue). These regions form a confidence belt (green). The intersection of the observation x_0 (red) with the confidence belt defines the confidence interval $[\theta_1, \theta_2]$ [94].

2101 9.3.4 Profile likelihood ratio

2102 In this section we describe a frequentist statistical procedure based on the profile
 2103 likelihood ratio test statistic, which is implemented using asymptotic distributions.

2104 With a multi-parameter likelihood function $L(\boldsymbol{\alpha})$, the maximum likelihood of
 2105 one specific parameter α_p with other parameters $\boldsymbol{\alpha}_o$ fixed, is called the conditional
 2106 maximum likelihood estimate and is denoted $\hat{\alpha}_p(\boldsymbol{\alpha}_0)$. The process of choosing specific
 2107 values of the nuisance parameters for a given value of μ , $\mathcal{D}_{\text{simulated}}$, and value of global
 2108 observables \mathcal{G} is called profiling. From the full list of parameters $\boldsymbol{\alpha}$, we denote the
 2109 parameter of interest μ , and the nuisance parameters $\boldsymbol{\theta}$.

2110 We construct the profile likelihood ratio,

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

2111 which depends explicitly on the parameter of interest μ , implicitly on the data \mathcal{D}_{sim}
 2112 and global observables \mathcal{G} , and is independent of the nuisance parameters $\boldsymbol{\theta}$, which
 2113 have been eliminated in profiling [94].

2114 The main conceptual reason for constructing the test statistic from the profile
 2115 likelihood ratio is that asymptotically (i.e. for measurements with many events) the
 2116 distribution of the profile likelihood ratio $\lambda(\mu = \mu_{\text{true}})$ is independent of the values of
 2117 the nuisance parameters [94].

2118 The following p -value is used to quantify the consistency with the hypothesis of a
 2119 signal strength of μ :

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

2120 **9.3.5 Modified frequentist method: CL_S**

2121 In the modified frequentist method called CL_S , to test a hypothesis with signal, we
2122 define p'_μ as a ratio of p -values [94]:

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

2123 where p_b is the p -value derived under the background-only hypothesis:

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

2124 The CL_S upper limit on μ , denoted μ_{up} , is obtained by solving for $p'_{\mu_{up}} = 5\%$.
2125 If testing the compatibility of the data with the background-only hypothesis, we
2126 consider the p_b value defined above and conventionally convert it into the quantile
2127 or “sigma” of a unit Gaussian. z standard deviations (e.g. $z = 5$ in “ 5σ ”) means
2128 that the probability of falling above these standard deviations, equals p_b (e.g. 3σ
2129 corresponds to $p_b = 2.7 \times 10^{-3}$ or 95.43%, and 5σ corresponds to $p_b = 5.7 \times 10^{-7}$ or
2130 99.999943%).

2131 **Chapter 10**

2132 **Results**

2133 **10.1 Results from $bb\tau\tau$**

2134 In each of the three $\tau\tau$ channels studied ($\mu\tau_h$, $e\tau_h$, and $e\mu$), events are divided based
2135 on whether they contain exactly 1 or 2 b-tag jets, and further divided into signal
2136 and control regions (SRs and CRs) using the DNN categorization score as described
2137 in Section 9.2. The control regions demonstrate good agreement between observed
2138 events in data, and the sum of the contributions from expected backgrounds that
2139 are modeled in simulated and embedded samples. The signal regions are defined to
2140 be sensitive to the $h \rightarrow aa \rightarrow bb\tau\tau$ signal. The postfit final observed and expected
2141 distributions of the di-tau invariant mass $m_{\tau\tau}$ reconstructed with SVFit (described
2142 in Section 5.2) are shown in Fig. 10.1 for the $\mu\tau_h$ channel, Fig. 10.2 for the $e\tau_h$
2143 channel, and Fig. 10.3 for the $e\mu$ channel. In all figures, the hypothesized yield for
2144 the $h \rightarrow aa \rightarrow bb\tau\tau$ signal is shown for the pseudoscalar mass $m_a = 35$ GeV and
2145 assuming a branching fraction $B(H \rightarrow aa \rightarrow bb\tau\tau) = 10\%$.

2146 The 95% CL expected and observed exclusion limits on the signal strength of the
2147 branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ as a function of the pseudoscalar mass m_a
2148 ranging from 12 GeV to 60 GeV, are shown for the three $\tau\tau$ channels and all three

2149 channels combined in Fig. 10.4. The limits are shown as percentages and normalized
2150 to the production cross-section of the Standard Model Higgs boson. No excess of
2151 events above the Standard Model expectations is observed. In the limits for the three
2152 $\tau\tau$ channels combined, expected (observed) limits range from 1.4 to 5.6% (1.7 to
2153 7.6%) for pseudoscalar masses between 12 and 60 GeV.

2154 The $e\mu$ channel is the only channel that has signal sensitivity to the $m_a = 12$
2155 GeV pseudoscalar mass hypothesis, because the minimum required spatial separation
2156 $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ between the two τ legs is smaller than the other two channels
2157 ($\Delta R < 0.3$ for $e\mu$, compared to $\Delta R < 0.4$ for the other two channels). This decreased
2158 ΔR requirement results in better signal acceptance for low mass signals for the $e\mu$
2159 channel. The $\mu\tau_h$ and $e\tau_h$ channels are most sensitive to the intermediate mass points
2160 studied, since the analysis targets a resolved signature: at low mass points, the tau
2161 legs are boosted, and at high mass points, the $m_{\tau\tau}$ distributions in signal have larger
2162 overlap with background distributions. In the combination of the three $\tau\tau$ channels,
2163 the limit for $m_a = 12$ GeV comes only from the $e\mu$ channel, and the best sensitivity
2164 is attained at intermediate mass points around $m_a = 20$ GeV to 45 GeV.

2165 To set limits on the branching fraction of the 125 GeV Higgs to the two pseu-
2166 doscalars, $B(h \rightarrow aa)$, we interpret the results in four types of 2HDM+S, which were
2167 introduced in Section 1.4. In 2HDM+S, the theorized branching fraction of the pseu-
2168 doscalars depends on the 2HDM+S model type, the pseudoscalar mass m_a , and the
2169 ratio of the two Higgs doublets' vacuum expectation values $\tan\beta$. In Type I models,
2170 the branching fraction is independent of $\tan\beta$, while in Types II, III, and IV, it is
2171 a function of m_a and $\tan\beta$. Limits for the $bb\tau\tau$ final state as a function of m_a for
2172 2HDM+S Type I (valid for all $\tan\beta$ values), Type II with $\tan\beta = 2.0$, Type III with
2173 $\tan\beta = 2.0$, and Type IV with $\tan\beta = 0.6$ are overlaid and shown in Fig. 10.5a.

2174 10.2 Combination with $bb\mu\mu$ final state

2175 Results from this analysis for the $h \rightarrow aa \rightarrow bb\tau\tau$ final state are combined with the
2176 analysis for the $h \rightarrow aa \rightarrow bb\mu\mu$ final state [95]. While the predicted branching ratio
2177 for $aa \rightarrow bb\mu\mu$ is comparatively small, the $bb\mu\mu$ final state has competitive results
2178 due to the excellent di-muon resolution measured by CMS. The $bb\mu\mu$ analysis uses
2179 an unbinned fit to the data using the di-muon mass $m_{\mu\mu}$ distribution. Details can be
2180 found in [95].

2181 Combining the results is possible since the $bb\tau\tau$ analysis explicitly rejects events
2182 with extra leptons, so there is no overlap between the events studied in the $bb\tau\tau$
2183 analysis and the $bb\mu\mu$ analysis. In the statistical combination, several systematic
2184 uncertainties are treated as correlated: the integrated luminosity normalization, the
2185 b-tagging scale factor, the scale factors related to muon reconstruction, identifica-
2186 tion, and trigger efficiencies, the inefficiency in the ECAL trigger readout, and the
2187 theoretical uncertainties related to signal modeling.

2188 Since the results in both final states are statistically limited, the combination ben-
2189 efits from the additional data. For $m_a = 35$ GeV, all systematic uncertainties amount
2190 to around 6% of the total uncertainty, with the dominant systematic uncertainties
2191 coming from jet energy systematics in the $bb\mu\mu$ final state, theoretical uncertainties
2192 in the signal, and uncertainties in the QCD multijet backgrounds in the $e\mu$ channel
2193 of the $bb\tau\tau$ final state.

2194 The mass distributions of the di-muon and di-tau objects ($m_{\mu\mu}$ and $m_{\tau\tau}$) are
2195 compared to the data in a combined maximum likelihood fit to derive upper limits
2196 on $B(h \rightarrow aa)$. The observed limits at 95% CL on $B(h \rightarrow aa)$ for different 2HDM+S
2197 scenarios, are shown for the search for $h \rightarrow aa \rightarrow bb\mu\mu$ in Fig. 10.5b, and the
2198 combined analyses $h \rightarrow aa \rightarrow bb\ell\ell$ in Fig. 10.6.

2199 Exclusion limits in a two-dimensional plane as a function of $\tan\beta$ and m_a are
2200 set for 2HDM+S Types II, III, and IV in Fig. 10.7. The most stringent constraints

2201 are observed for 2HDM+S type III because of large branching fractions predicted in
2202 theory, with predicted branching fractions between 0.47 and 0.42 for $\tan \beta = 2.0$ and
2203 values of m_a between 15 and 60 GeV, compared to the observed 95% CL upper limits
2204 which are between 0.08 and 0.03. For 2HDM+S type IV, the predicted branching
2205 fractions from theory are between 0.26 and 0.20 for $\tan \beta = 0.6$ for values of m_a
2206 between 15 and 60 GeV, and the 95% CL observed upper limits are between 0.12 and
2207 0.05.

2208 The combined results from $h \rightarrow aa \rightarrow bb\ell\ell$ are compared with CMS results in
2209 other final states as a function of the pseudoscalar mass m_a : for 2HDM+S type I in
2210 Fig. 10.8, type II with $\tan \beta = 2.0$ in Fig. 10.9, and type III with $\tan \beta = 2.0$ in Fig.
2211 10.10. In other scenarios, e.g. type III with $\tan \beta = 5.0$, more stringent limits are set
2212 by analyses in other final states, $\mu\mu\tau\tau$ in this case. Other summary plots for other
2213 model types and $\tan \beta$ values can be found at [96].

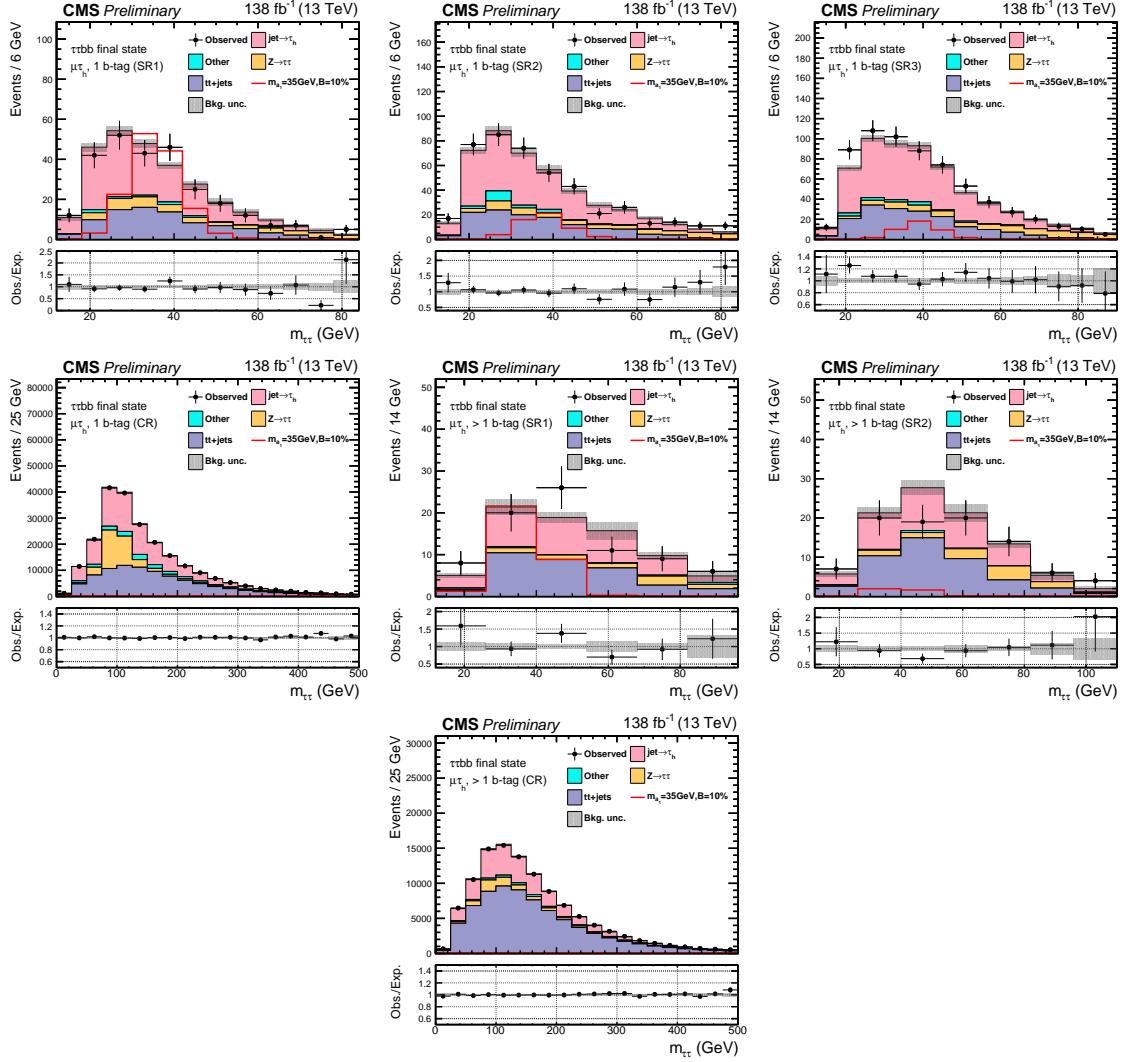


Figure 10.1: Postfit final $m_{\tau\tau}$ observed and expected distributions, and the observed/expected ratios, in the $\mu\tau_h$ channel [38]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), 1 b-tag jet control region (*middle row*), 2 b-tag jet signal regions (SR1, 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*).

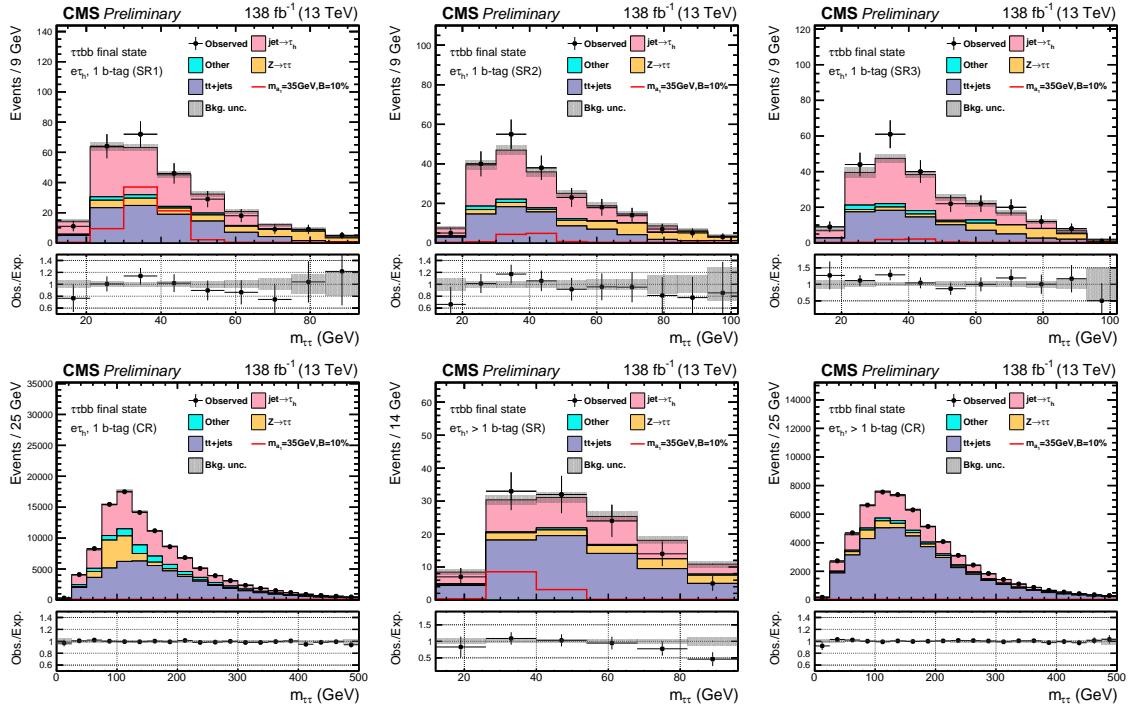


Figure 10.2: Postfit final observed and expected $m_{\tau\tau}$ distributions, and the observed/expected ratios, in the $e\tau_h$ channel [38]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, SR3) (*top row*), the 1 b-tag jet control region (CR) (*bottom row*), and 2 b-tag jet signal 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*).

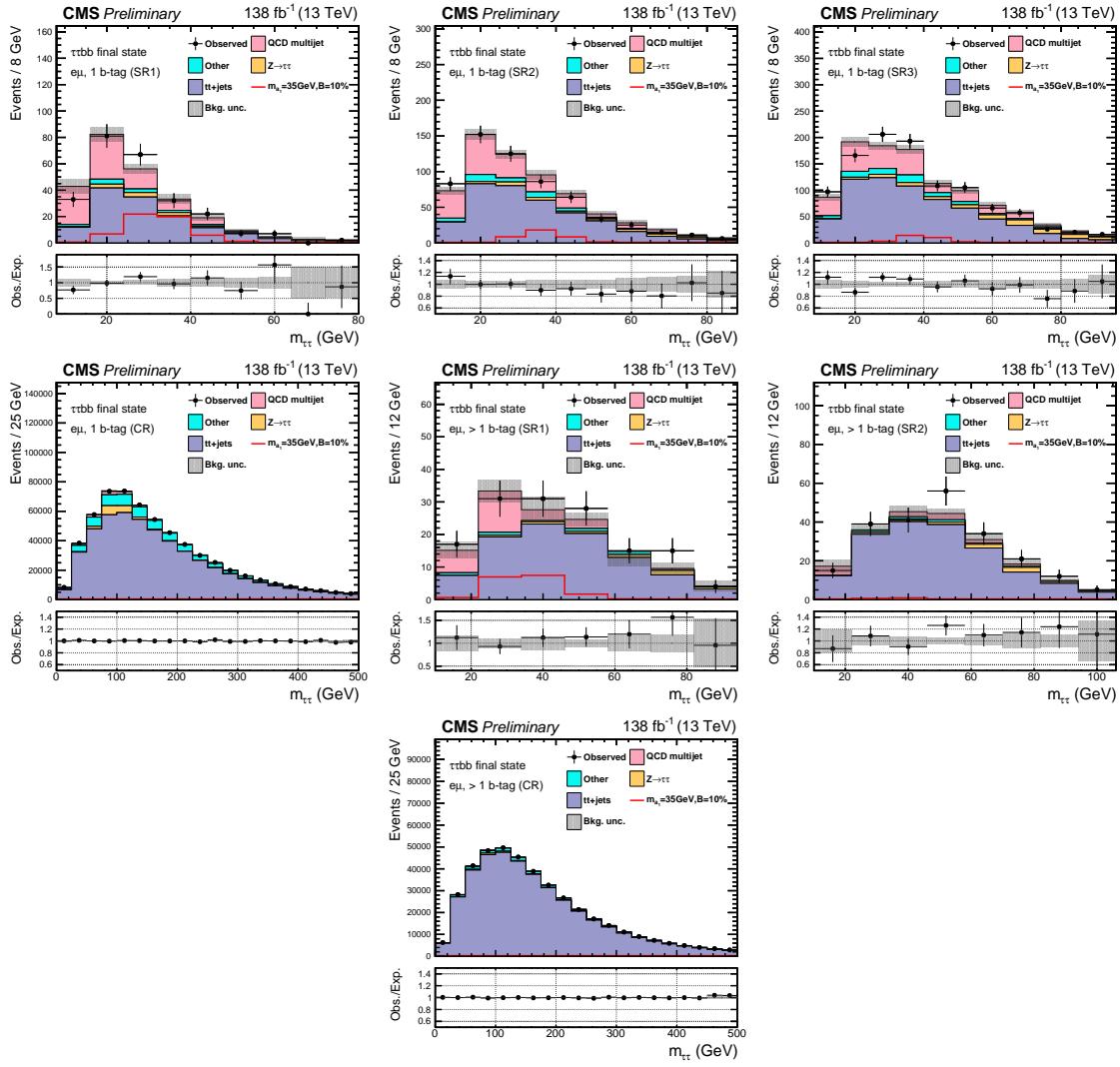


Figure 10.3: Postfit final observed and expected $m_{\tau\tau}$ distributions, and the observed/expected ratios, in the $e\mu$ channel [38]. Events are divided into the 1 b-tag jet signal regions (SR1, SR2, and SR3) (*top row*), 1 b-tag jet control region (CR) (*middle row*), 2 b-tag jet signal regions (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*).

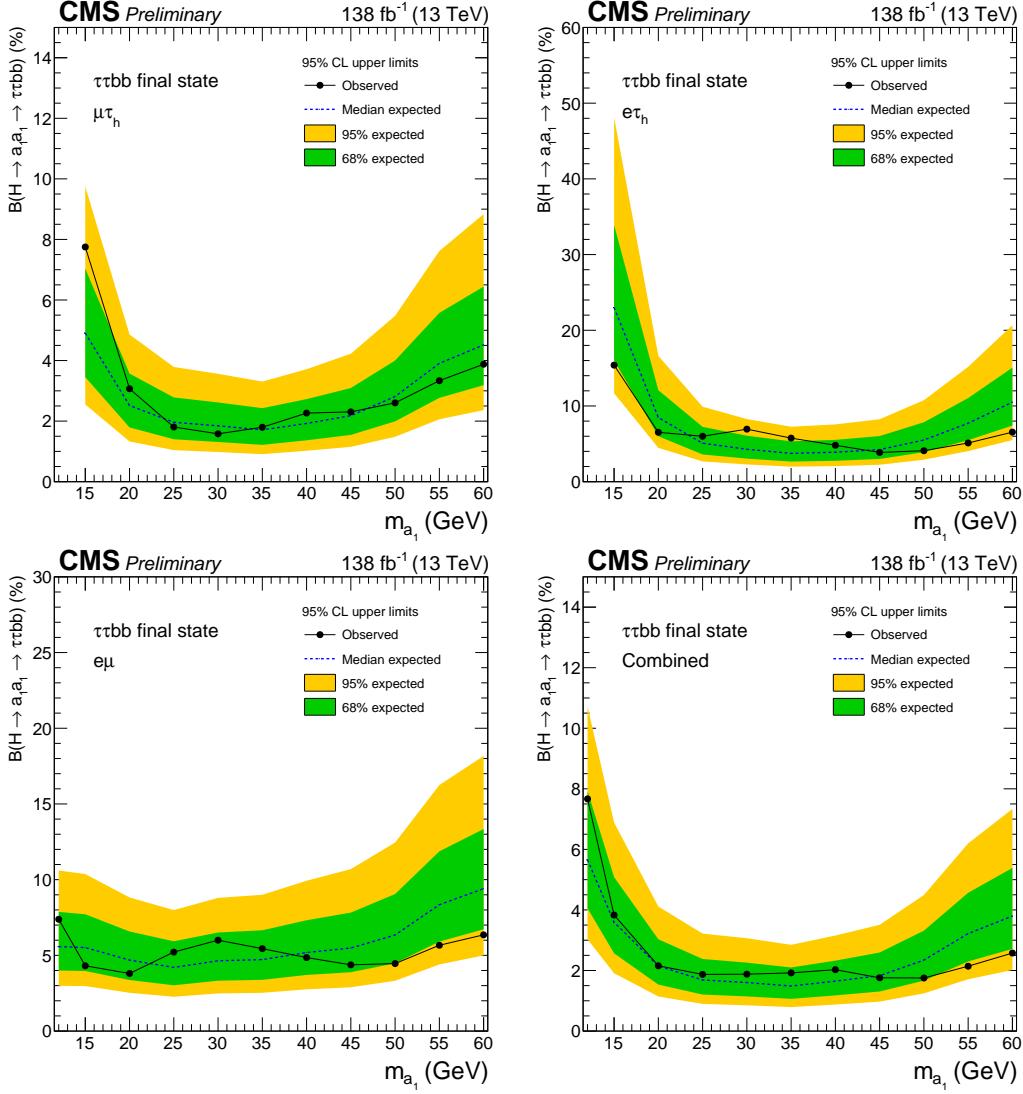
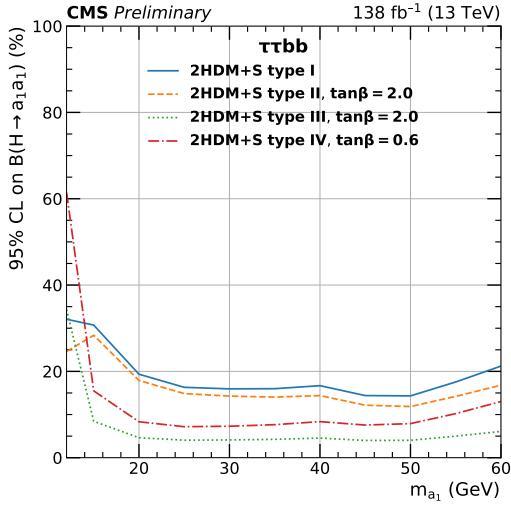
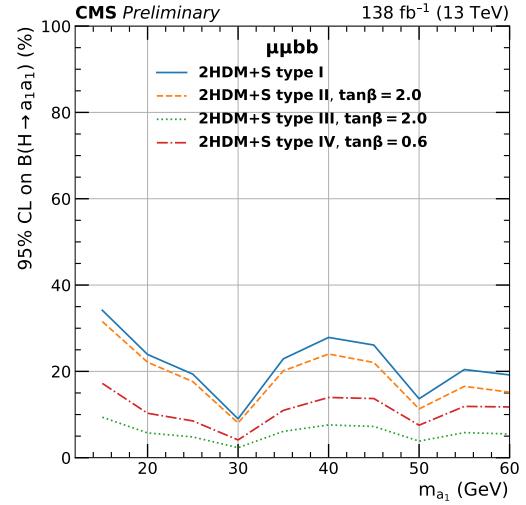


Figure 10.4: Observed 95% CL exclusion limits (*black, solid lines*) and expected 95% CL and 68% CL limits (*shaded yellow and green*) on the branching fraction $B(h \rightarrow aa \rightarrow bb\tau\tau)$ in percentages, assuming the Standard Model production for the 125 GeV Higgs (h). Limits are shown 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*) [38]. 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 $m_a = 12$ 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) [38]. Linear interpolation is used between points in the graphs. The $\tan\beta$ values chosen here correspond to the most stringent limits in each model.

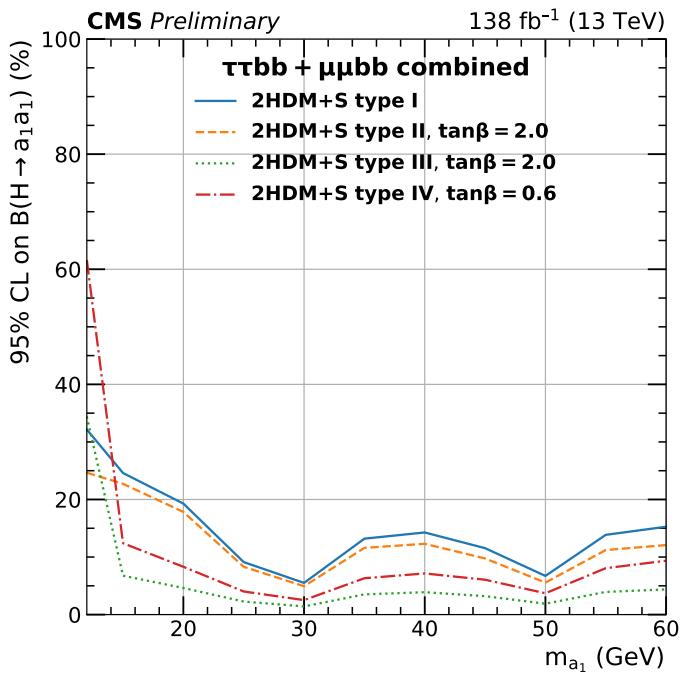


Figure 10.6: Observed 95% CL upper limits on the branching fraction of the 125 GeV Higgs boson to two pseudoscalars, $B(h \rightarrow aa)$, in percentages, as a function of the pseudoscalar mass m_a , in 2HDM+S type I (blue), type II with $\tan\beta = 2.0$ (orange dashed), type III with $\tan\beta = 2.0$ (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} [38].

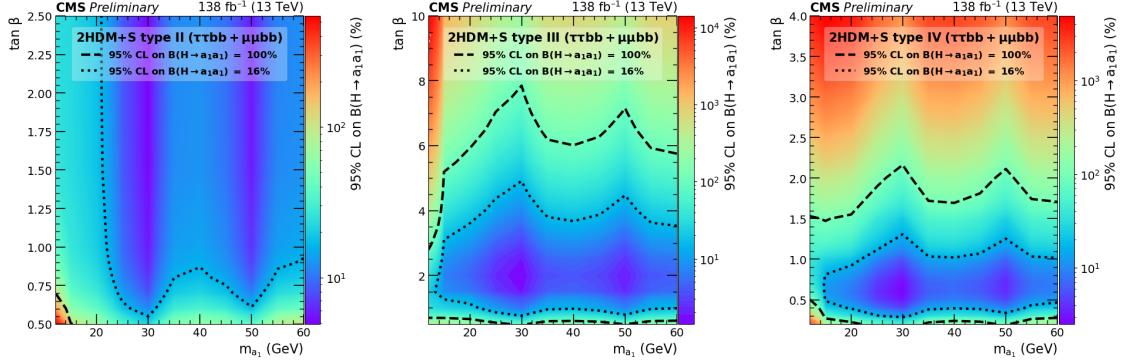


Figure 10.7: Observed 95% CL upper limits on $\mathcal{B}(h \rightarrow aa)$ in %, for the combination of $bb\mu\mu$ and $bb\tau\tau$ channels using the full Run 2 integrated luminosity of 138 fb^{-1} for Type II (*left*), Type III (*middle*), and Type IV (*right*) 2HDM+S in the $\tan \beta$ vs. m_a 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 [38].

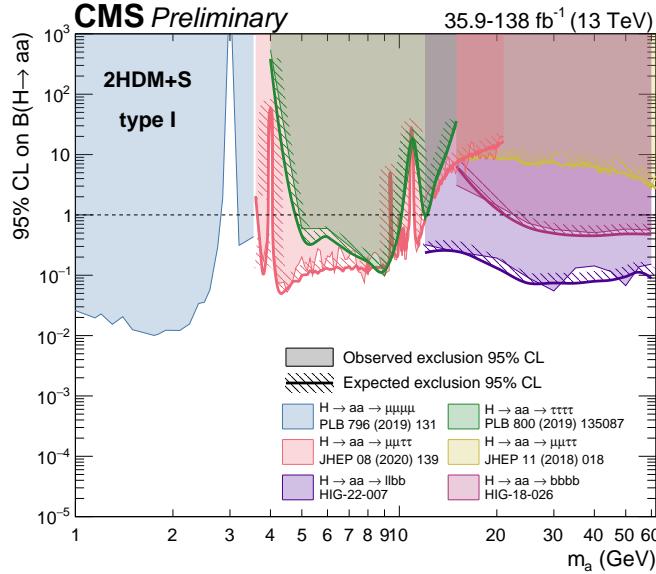


Figure 10.8: Summary plot of current 95% limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$ in the 2HDM+S type I scenario performed with data collected at 13 TeV [96]. 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).

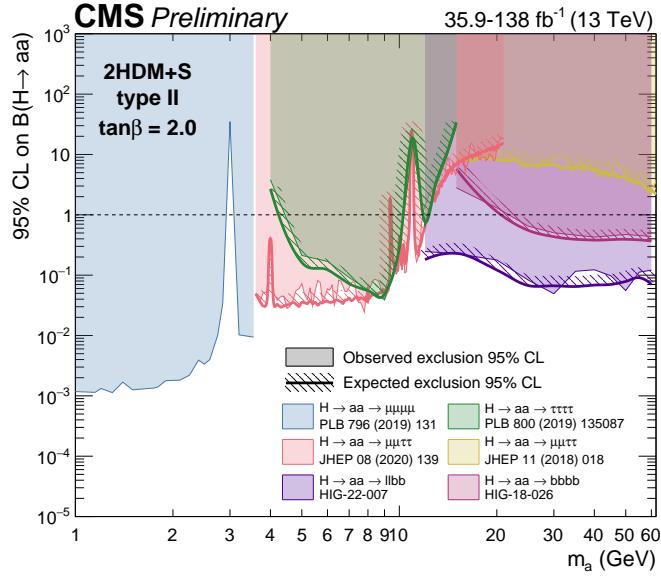


Figure 10.9: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross-section, $\frac{\sigma(h)}{\sigma_{\text{SM}}} \times B(h \rightarrow aa)$, in the 2HDM+S type II scenario with $\tan \beta = 2.0$, obtained at CMS with data collected at 13 TeV [96]. 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).

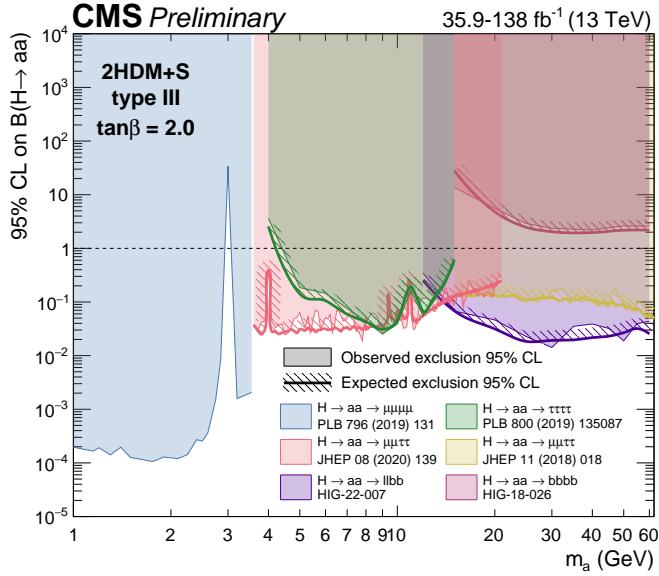


Figure 10.10: Summary plot of current observed and expected 95% CL limits on the branching ratio of the 125 GeV Higgs boson to two pseudoscalars, normalized to the Standard Model Higgs production cross section, $\frac{\sigma(h)}{\sigma_{SM}} \times B(h \rightarrow aa)$ in the 2HDM+S type-III scenario with $\tan \beta = 2.0$, obtained at CMS with data collected at 13 TeV [96]. 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).

2214 **Chapter 11**

2215 **Asymmetric exotic Higgs decays**

2216 This chapter presents progress towards a search for exotic Higgs decays to two light
2217 scalars with unequal mass ($h \rightarrow a_1 a_2$) final states with bottom quarks and τ leptons,
2218 with plans to interpret the results in the context of Two Real Singlet Models (TRSMs),
2219 described in Section 1.5. Compared to the symmetric decay scenario $h \rightarrow aa$ which
2220 has been studied in multiple final states at CMS with stringent limits set on the
2221 various 2HDM+S scenarios, this asymmetric decay scenario has not been directly
2222 searched for at the CMS experiment. Section 11.1 lists the mass hypotheses of the
2223 new particles a_1 and a_2 that will be studied. Section 11.2 describes the studies on
2224 which channels the analysis will be carried out in. Section 11.3 shows the control
2225 plots produced using the analysis framework that will be used for this analysis.

2226 **11.1 Signal masses**

2227 As discussed in Section 1.5, $h \rightarrow a_1 a_2$ can result in a “cascade” decay if one of the
2228 scalars, a_2 is sufficiently heavy ($m_{a_2} > 2m_{a_1}$). The “non-cascade” case is where the
2229 light scalars decay directly to Standard Model particles.

2230 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 [5]. For an example mass point $m_{a_2} = 80$ GeV, $m_{a_1} = 30$ GeV, the branching fractions to $4b2\tau$ is ten times larger than $2b4\tau$: $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 4b2\tau) = 0.00857$, vs. $B(h \rightarrow a_1 a_2 \rightarrow 3a_1 \rightarrow 2b4\tau) = 0.00068$. The $4b2\tau$ final state is chosen for this analysis.

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

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

2256 and fourth b-flavor jets which have very low transverse momenta p_T . Event categories
 2257 with exactly 1 b-tag jet and ≥ 2 b-tag jets will be used.

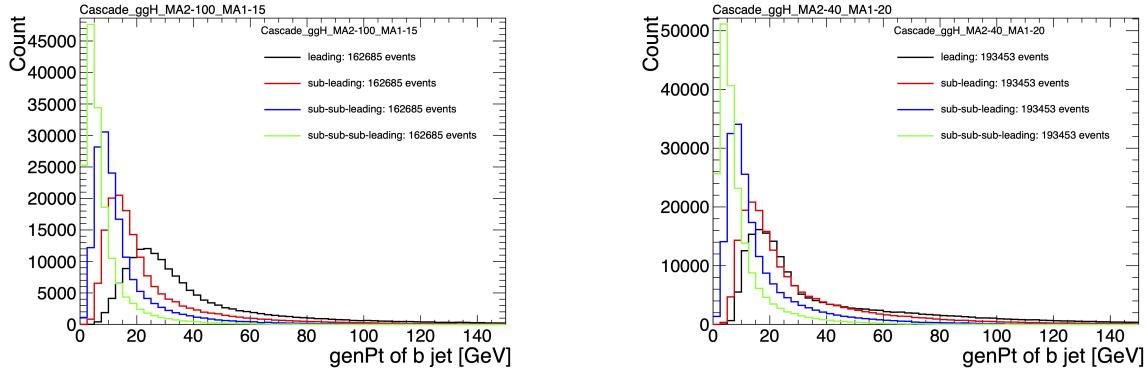


Figure 11.1: Generator-level b-flavor jet transverse momenta p_T , for $h \rightarrow a_1 a_2$ cascade scenario in the $4b2\tau$ final state, for mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV (*left*) and $(40, 20)$ GeV (*right*). In each plot the generator-level p_T of the leading (*black*),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

2258 In the $4b2\tau$ final state, the possibility of the leading and sub-leading b-tag jets
 2259 being sufficiently close in ΔR to require boosted jet reconstruction techniques was
 2260 explored. In the $4b2\tau$ case, the two b-flavor-jets in the generated event that were
 2261 spatially closest in ΔR were considered as one object. This two b-flavor jet object was
 2262 spatially matched in ΔR to the jets reconstructed with the standard AK4 algorithm
 2263 which uses a cone size of $\Delta R = 0.4$. The quality of the p_T resolution (computed as
 2264 $(p_{T,\text{reconstructed}} - p_{T,\text{gen}})/p_{T,\text{gen}}$) and closeness in distance ΔR of the reconstructed jet
 2265 to the nearest generator-level jets, was seen to depend on the absolute and relative
 2266 masses of the light scalars. The best (worst) performance occurred in samples with
 2267 large (small) mass differences between the heavier scalar a_2 and the lighter scalar a_1 ,
 2268 as illustrated for the mass hypotheses (m_{a_1}, m_{a_2}) (100, 15) GeV and (40, 20) GeV in
 2269 Fig. 11.2.

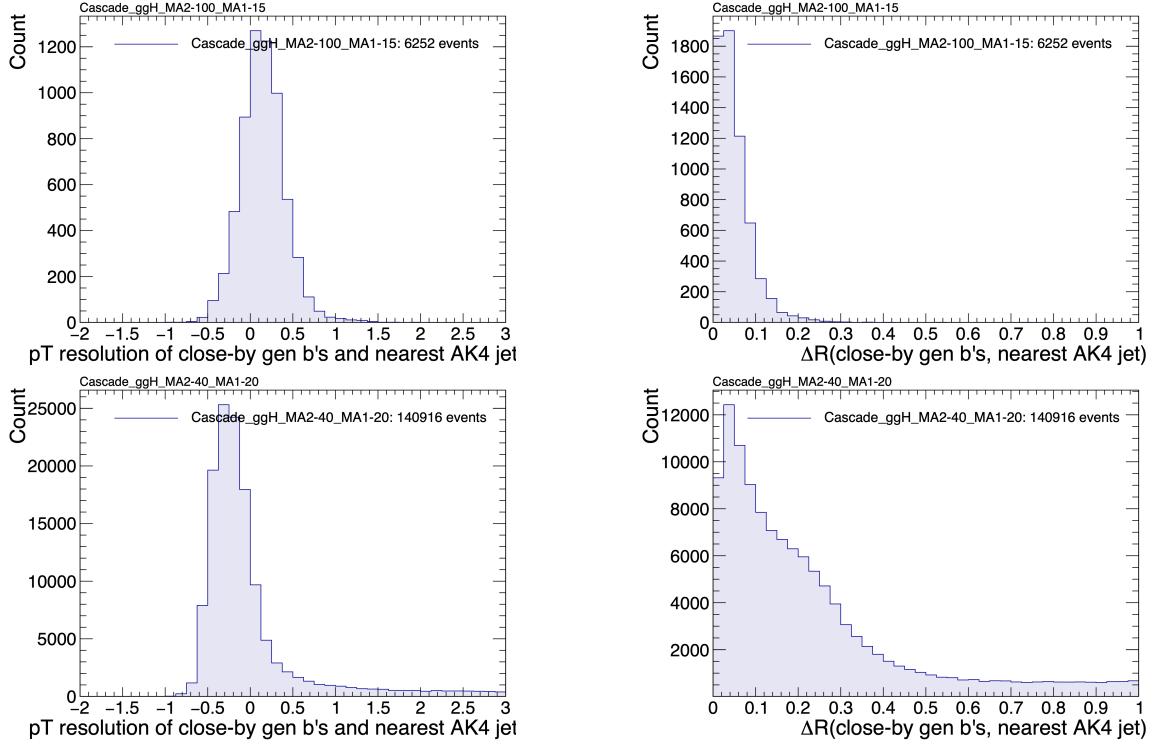


Figure 11.2: Distributions (arbitrary units) of transverse momentum p_T resolution and ΔR between the two closest generator-level b jets, treated as one object, and the nearest reconstructed AK4 jet, for two different $h \rightarrow a_1 a_2$ mass hypotheses $(m_{a_1}, m_{a_2}) = (100, 15)$ GeV (top left, 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,

2279 described in Sections 5.3.1, 5.3.2, and 5.3.4).

2280 The p_T , η , and ϕ of the leading muon and hadronic tau τ_h , and the di-tau visible
2281 mass m_{vis} and momentum $p_{T,\text{vis}}$, are shown in Fig. 11.3. The p_T , η , and ϕ of the the
2282 leading and sub-leading b-tag jets, and the missing transverse energy magnitude and
2283 azimuthal direction, are shown in Fig. 11.4.

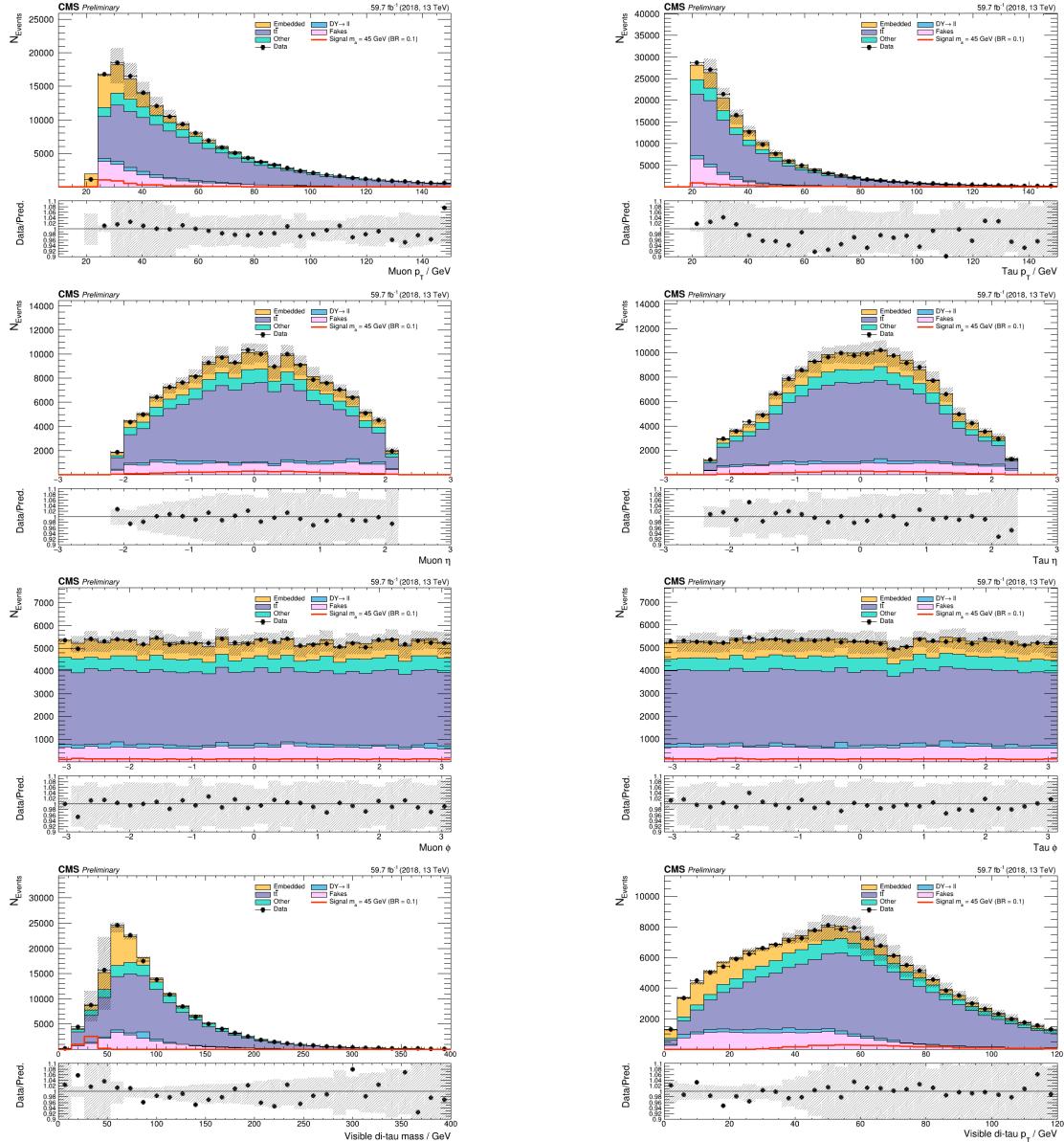


Figure 11.3: Kinematic properties of the leading muon and τ_h in the $\mu\tau_h$ channel: p_T (top row), η (second row), and ϕ (third row). The visible 4-momenta of the muon and τ_h are summed, giving the visible di-tau mass m_{vis} and transverse momentum $p_{T,\text{vis}}$. The errors shown in the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

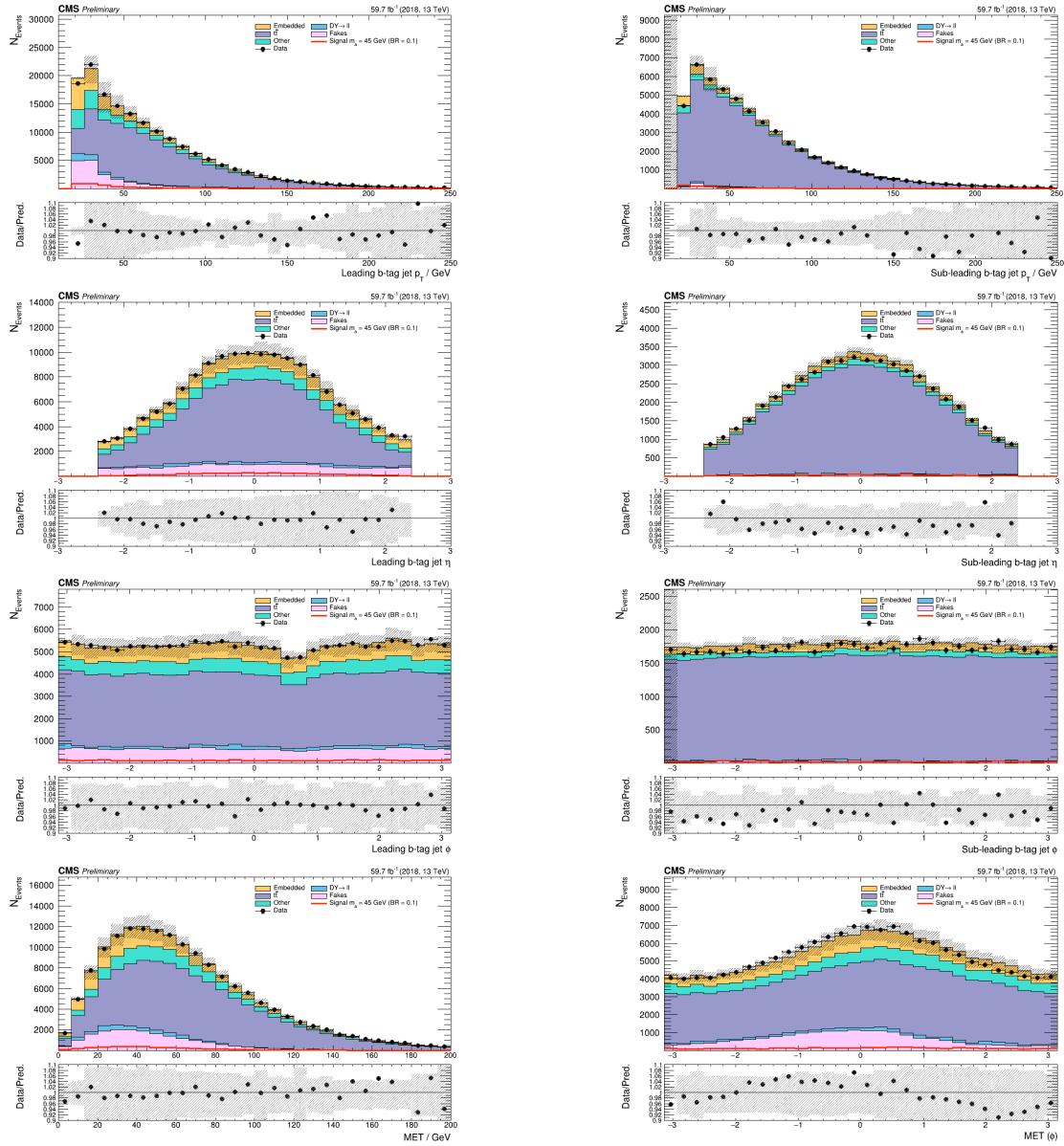


Figure 11.4: Kinematic properties of the leading and sub-leading b-tag jets in the $\mu\tau_h$ final state: jet p_T (*top row*), η (*second row*), ϕ (*third row*), as well as the missing transverse energy magnitude and azimuthal direction (*bottom row*). The errors shown in the figures only include statistical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

2284

Chapter 12

2285

Conclusion and outlook

2286 With the discovery of a Higgs boson with mass 125 GeV at the LHC in 2012, the LHC
2287 and CMS physics program has evolved to include the precise characterization of the
2288 125 GeV Higgs boson and searching for evidence of additional Higgs particles in an
2289 extended Higgs sector. This thesis presents a direct search at CMS for exotic decays
2290 of the Higgs boson with mass 125 GeV in data collected in the years 2016-2018 in
2291 proton-proton collisions at center-of-mass energy 13 TeV, to two light neutral scalar
2292 particles that decay to two bottom quarks and two tau leptons ($h \rightarrow aa \rightarrow bb\tau\tau$). The
2293 results are combined with another search that was performed in the $h \rightarrow aa \rightarrow bb\mu\mu$
2294 final state, giving the most stringent limits to date for theories with Two Higgs
2295 Doublet Models extended with a singlet scalar (2HDM+S), for pseudoscalar masses
2296 m_a ranging from 15 GeV to 60 GeV, in a number of 2HDM+S scenarios such as type
2297 II and III with $\tan\beta = 2.0$.

2298 As the rich physics program of CMS has set stringent limits on the exotic decay
2299 $h \rightarrow aa$, we turn our attention to direct searches for decays to light neutral scalars
2300 with potentially unequal mass, $h \rightarrow a_1a_2$, which has not been performed at CMS
2301 to date. Preliminary studies on $h \rightarrow a_1a_2$ signals in the Two Real Singlet Model
2302 (TRSM) are shown, and work is ongoing to develop the analysis for $h \rightarrow a_1a_2$ in final

2303 states with bottom quarks and tau leptons.

2304 To ensure the continued performance of the CMS detector and to enhance its
2305 data-taking capabilities in the intense pileup conditions of the Phase-2 upgrade of
2306 the High-Luminosity LHC, upgrades of the Level-1 Trigger are paramount for filter-
2307 ing the increased data rate of the HL-LHC. This thesis presents work on the stan-
2308 dalone barrel calorimeter algorithm for reconstructing and identifying electron and
2309 photon candidates, using high granularity crystal-level information from the ECAL
2310 subdetector. For Phase-2, the increase in the granularity of information sent from
2311 the electromagnetic calorimeter to the Level-1 trigger, from energy sums over towers
2312 (which are 5×5 in crystals) to crystal-level information, allows for the implementa-
2313 tion of a more sophisticated clustering algorithm that can exploit the fact that genuine
2314 electrons and photons tend to leave energies concentrated a 3×5 window in crystals,
2315 and use shape and isolation information to distinguish genuine electrons and photons
2316 from noise. Electrons and photons are key to characterizing Standard Model pro-
2317 cesses and performing searches for new physics, and this represents one of the many
2318 upgrades of the CMS detector in preparation for Phase-2. With the ongoing Run-3
2319 data collecting period, and wealth of ongoing and scheduled upgrades, there remains
2320 an abundance of directions for detector development and physics at CMS heading
2321 into Phase-2 of the LHC.

²³²² Bibliography

- ²³²³ [1] Paul H. Frampton. Journeys Beyond the Standard Model. 54(1):52–
²³²⁴ 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.
- ²³²⁷ [2] Christopher G. Tully. *Elementary Particle Physics in a Nutshell*. Princeton
²³²⁸ University Press, Princeton, 2012. ISBN 9781400839353. doi: doi:10.1515/
²³²⁹ 9781400839353. URL <https://doi.org/10.1515/9781400839353>.
- ²³³⁰ [3] John Ellis. Higgs Physics. In *2013 European School of High-Energy Physics*,
²³³¹ pages 117–168, 2015. doi: 10.5170/CERN-2015-004.117.
- ²³³² [4] David Curtin, Rouven Essig, Stefania Gori, and Others. Exotic decays of the 125
²³³³ GeV Higgs boson. *Phys. Rev. D*, 90:075004, Oct 2014. doi: 10.1103/PhysRevD.
²³³⁴ 90.075004. URL <https://link.aps.org/doi/10.1103/PhysRevD.90.075004>.
- ²³³⁵ [5] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt. Two-real-scalar-singlet
²³³⁶ 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.
- ²³³⁸ [6] Antonella Del Rosso. Aerial view of the LHC and the four major experiments.
²³³⁹ 2017. URL <https://cds.cern.ch/record/2253966>. General Photo.

- 2340 [7] ATLAS Collaboration. ATLAS: Detector and physics performance technical
2341 design report. Volume 1. 5 1999.
- 2342 [8] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector
2343 Performance and Software. 2006.
- 2344 [9] L Musa. Conceptual Design Report for the Upgrade of the ALICE ITS. Technical
2345 report, CERN, Geneva, 2012. URL <https://cds.cern.ch/record/1431539>.
- 2346 [10] S. Amato et al. LHCb technical proposal: A Large Hadron Collider Beauty
2347 Experiment for Precision Measurements of CP Violation and Rare Decays. 2
2348 1998.
- 2349 [11] R. Schmidt. Accelerator physics and technology of the LHC. In *ROXIE: Routine*
2350 *for the Optimizazation of Magnet X-Sections, Inverse Field Calculation and Coil*
2351 *End Design*, pages 7–17, 1998.
- 2352 [12] J. Vollaire et al. *Linac4 design report*, volume 6/2020 of *CERN Yellow Reports:*
2353 *Monographs*. CERN, Geneva, 9 2020. ISBN 978-92-9083-579-0, 978-92-9083-580-
2354 6. doi: 10.23731/CYRM-2020-006.
- 2355 [13] Werner Herr and B Muratori. Concept of luminosity. 2006. doi: 10.5170/
2356 CERN-2006-002.361. URL <https://cds.cern.ch/record/941318>.
- 2357 [14] Olivier S. Brning and Frank Zimmerman. Parameter space for the LHC lumi-
2358 nosity upgrade. ISBN 978-3-95450-115-1. URL [https://accelconf.web.cern.](https://accelconf.web.cern.ch/IPAC2012/papers/moppc005.pdf)
2359 [ch/IPAC2012/papers/moppc005.pdf](https://accelconf.web.cern.ch/IPAC2012/papers/moppc005.pdf).
- 2360 [15] CMS Collaboration. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15(09):
2361 P09018, 2020. doi: 10.1088/1748-0221/15/09/P09018.
- 2362 [16] CMS Collaboration. Measurement of the inelastic proton-proton cross section at
2363 $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018. doi: 10.1007/JHEP07(2018)161.

- 2364 [17] CMS Collaboration. High-Luminosity Large Hadron Collider (HL-LHC): Tech-
2365 nical design report. 10/2020, 12 2020. doi: 10.23731/CYRM-2020-0010.
- 2366 [18] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004,
2367 2008. doi: 10.1088/1748-0221/3/08/S08004.
- 2368 [19] CMS Collaboration. Particle-flow reconstruction and global event description
2369 with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/
2370 12/10/P10003.
- 2371 [20] V Karimki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freuden-
2372 reich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker*
2373 *system project: Technical Design Report*. Technical design report. CMS. CERN,
2374 Geneva, 1997. URL <https://cds.cern.ch/record/368412>.
- 2375 [21] The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva,
2376 2017. URL <https://cds.cern.ch/record/2272264>.
- 2377 [22] CMS Collaboration. CMS Technical Design Report for the Pixel Detector Up-
2378 grade. 9 2012. doi: 10.2172/1151650.
- 2379 [23] R. L. Workman and Others. Review of particle physics. 2022:083C01. doi:
2380 10.1093/ptep/ptac097.
- 2381 [24] CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorime-
2382 ter. 9 2012. doi: 10.2172/1151651.
- 2383 [25] CMS Technical Design Report for the Level-1 Trigger Upgrade. 6 2013.
- 2384 [26] S. Dasu et al. CMS. The TriDAS project. Technical design report, vol. 1: The
2385 trigger systems. 12 2000.
- 2386 [27] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and
2387 High Level Trigger. Technical report, CERN, Geneva, 2021. URL <https://cds>.

- 2388 cern.ch/record/2759072. This is the final version of the document, approved
2389 by the LHCC.
- 2390 [28] C. Foudas. The CMS Level-1 Trigger at LHC and Super-LHC. In *34th Interna-*
2391 *tional Conference on High Energy Physics*, 10 2008.
- 2392 [29] CMS Software Guide. High Level Trigger (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHighLevelTrigger>.
- 2394 [30] The Worldwide LHC Computing Grid. 2012. URL <https://cds.cern.ch/record/1997398>.
- 2396 [31] Douglas Thain, Todd Tannenbaum, and Miron Livny. Distributed computing in
2397 practice: the Condor experience. *Concurrency and Computation: Practice and*
2398 *Experience*, 17(2-4):323–356, 2005. doi: <https://doi.org/10.1002/cpe.938>. URL
2399 <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.938>.
- 2400 [32] Alexandre Zabi, Jeffrey Wayne Berryhill, Emmanuelle Perez, and Alexander D.
2401 Tapper. The Phase-2 Upgrade of the CMS Level-1 Trigger. 2020.
- 2402 [33] Technical proposal for a MIP timing detector in the CMS experiment Phase 2
2403 upgrade. Technical report, CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2296612>.
- 2405 [34] CMS Collaboration. CMS luminosity measurement for the 2016 data-taking
2406 period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. URL
2407 <https://cds.cern.ch/record/2257069>.
- 2408 [35] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking
2409 period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-
2410 004, 2018. URL <https://cds.cern.ch/record/2621960>.

- 2411 [36] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking
2412 period at $\sqrt{s} = 13$ TeV. CMS Physics Analysis Summary CMS-PAS-LUM-18-
2413 002, 2019. URL <https://cds.cern.ch/record/2676164>.
- 2414 [37] CMS LUMI Group. CMS Luminosity Public Results (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- 2416 [38] Cécile Caillol, Pallabi Das, Sridhara Dasu, Pieter Everaerts, Stephanie Kwan,
2417 Isobel Ojalvo, and Ho-Fung Tsoi. CMS AN-20-213 (internal): Search for an
2418 exotic decay of the 125 GeV Higgs boson to light pseudoscalars, with a pair of b
2419 jets and a pair of tau leptons in the final state, 2020.
- 2420 [39] CMS Collaboration. Search for exotic decays of the Higgs boson to a pair of
2421 pseudoscalars in the $\mu\mu bb$ and $\tau\tau bb$ final states. *European Physical Journal C*,
2422 2 2024.
- 2423 [40] J. Alwall, R. Frederix, S. Frixione, et al. The automated computation of tree-
2424 level and next-to-leading order differential cross sections, and their matching to
2425 parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.
2426 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).
- 2428 [41] R. Frederix, S. Frixione, V. Hirschi, et al. The automation of next-to-leading
2429 order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July
2430 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).
- 2432 [42] S. Agostinelli, J. Allison, K. Amako, et al. Geant4 - a simulation toolkit. 506(3):
2433 250–303. ISSN 0168-9002. doi: 10.1016/S0168-9002(03)01368-8. URL <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.

- 2435 [43] CMS Collaboration. An embedding technique to determine $\tau\tau$ backgrounds
2436 in proton-proton collision data. *JINST*, 14(06):P06032, 2019. doi: 10.1088/
2437 1748-0221/14/06/P06032.
- 2438 [44] CMS Collaboration. Search for neutral MSSM Higgs bosons decaying to a pair of
2439 tau leptons in pp collisions. *JHEP*, 10:160, 2014. doi: 10.1007/JHEP10(2014)160.
- 2440 [45] CMS Collaboration. Measurements of Higgs boson production in the decay chan-
2441 nel with a pair of τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur.*
2442 *Phys. J. C*, 83(7):562, 2023. doi: 10.1140/epjc/s10052-023-11452-8.
- 2443 [46] CMS Collaboration. Reconstruction and identification of tau lepton decays to
2444 hadrons and tau neutrinos at CMS. *Journal of Instrumentation*, 11(01):P01019–
2445 P01019, January 2016. ISSN 1748-0221. doi: 10.1088/1748-0221/11/01/p01019.
2446 URL <http://dx.doi.org/10.1088/1748-0221/11/01/P01019>.
- 2447 [47] CMS Collaboration. Performance of τ -lepton reconstruction and identification
2448 in CMS. *Journal of Instrumentation*, 7(01):P01001, jan 2012. doi: 10.1088/
2449 1748-0221/7/01/P01001. URL <https://dx.doi.org/10.1088/1748-0221/7/01/P01001>.
- 2451 [48] CMS Collaboration. Performance of reconstruction and identification of τ leptons
2452 decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):
2453 P10005, 2018. doi: 10.1088/1748-0221/13/10/P10005.
- 2454 [49] CMS Collaboration. Identification of hadronic tau lepton decays using a deep
2455 neural network. *JINST*, 17:P07023, 2022. doi: 10.1088/1748-0221/17/07/
2456 P07023.
- 2457 [50] CMS Collaboration. Performance of CMS muon reconstruction in pp collision
2458 events at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 7(10):P10002–P10002,

- 2459 October 2012. ISSN 1748-0221. doi: 10.1088/1748-0221/7/10/p10002. URL
2460 <http://dx.doi.org/10.1088/1748-0221/7/10/P10002>.
- 2461 [51] CMS Collaboration. Performance of electron reconstruction and selection with
2462 the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instru-*
2463 *mentation*, 10(06):P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005. URL
2464 <https://dx.doi.org/10.1088/1748-0221/10/06/P06005>.
- 2465 [52] CMS Collaboration. Identification of b-quark jets with the CMS experiment.
2466 *Journal of Instrumentation*, 8(04):P04013–P04013, April 2013. ISSN 1748-
2467 0221. doi: 10.1088/1748-0221/8/04/p04013. URL <http://dx.doi.org/10.1088/1748-0221/8/04/P04013>.
- 2469 [53] CMS Collaboration. Pileup Removal Algorithms. Technical report, CERN,
2470 Geneva, 2014. URL <https://cds.cern.ch/record/1751454>.
- 2471 [54] CMS Collaboration. CMS Phase 1 heavy flavour identification performance and
2472 developments. 2017. URL <https://cds.cern.ch/record/2263802>.
- 2473 [55] CMS Collaboration. Performance of the DeepJet b tagging algorithm using 41.9
2474 fb⁻¹ of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector.
2475 2018. URL <https://cds.cern.ch/record/2646773>.
- 2476 [56] Lorenzo Bianchini, John Conway, Evan Klose Friis, and Christian Veelken. Re-
2477 construction of the Higgs mass in $H \rightarrow \tau\tau$ Events by Dynamical Likelihood
2478 techniques. *Journal of Physics: Conference Series*, 513(2):022035, jun 2014.
2479 doi: 10.1088/1742-6596/513/2/022035. URL <https://dx.doi.org/10.1088/1742-6596/513/2/022035>.
- 2481 [57] CMS Collaboration. Evidence for the 125 GeV Higgs boson decaying to a pair
2482 of τ leptons. *JHEP*, 05:104, 2014. doi: 10.1007/JHEP05(2014)104.

- 2483 [58] CMS Collaboration. Missing transverse energy performance of the CMS detector.
2484 *JINST*, 6:P09001, 2011. doi: 10.1088/1748-0221/6/09/P09001.
- 2485 [59] Artur Kalinowski. CMS AN-19-032 (internal): Reconstruction of a τ pair invariant
2486 mass with a simplified likelihood scan, 2019.
- 2487 [60] CMS TAU POG. Tau Physics Object Group: Tau ID Recommendation For Run 2, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendationForRun2>.
- 2490 [61] CMS MUO POG. Muon Physics Object Group: Recommendations, 2024. URL
2491 <https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG>.
- 2492 [62] CMS MUO POG. Muon Physics Object Group: Reference guidelines and results
2493 for muon momentum scale and resolution in Run II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceScaleResolRun2>.
- 2495 [63] CMS HTT working group. Higgs To Tau Tau Working TWiki for the full Run-
2496 2 legacy analysis, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/HiggsToTauTauWorkingLegacyRun2>.
- 2498 [64] CMS ELE POG. Electron Physics Object Group: Recommendations
2499 for 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaRunIIRecommendations>.
- 2501 [65] CMS ELE POG. Electron Physics Object Group: Recommendations for
2502 2016 to 2018 UL, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaUL2016To2018>.
- 2504 [66] CMS TAU Embedding Group. Tau embedded samples using 2016
2505 data, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TauTauEmbeddingSamples2016Legacy>.

- 2507 [67] CMS TAU Embedding Group. Tau embedded samples using 2017
2508 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2017>.
- 2510 [68] CMS TAU Embedding Group. Tau embedded samples using 2018
2511 data, 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauTauEmbeddingSamples2018>.
- 2513 [69] Tau Lepton Run 2 Trigger Performance. 2019. URL <https://cds.cern.ch/record/2678958>.
- 2515 [70] CMS Taus High Level Trigger Studies. Tau Lepton Run 2 Trigger Performance
2516 (CMS DP-2019/012), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTauAllRun2>.
- 2518 [71] Muon HLT Performance with 2018 Data. 2018. URL <https://cds.cern.ch/record/2627469>.
- 2520 [72] Single and Double Electron Trigger Efficiencies using the full Run 2 dataset.
2521 2020. URL <https://cds.cern.ch/record/2888577>.
- 2522 [73] Run II Trigger Performance For $e\mu$ Triggers. 2019. URL <https://cds.cern.ch/record/2687013>.
- 2524 [74] Performance of electron and photon reconstruction in Run 2 with the CMS ex-
2525 periment. 2020. URL <https://cds.cern.ch/record/2725004>.
- 2526 [75] CMS Collaboration. Performance of the CMS muon detector and muon recon-
2527 struction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015,
2528 2018. doi: 10.1088/1748-0221/13/06/P06015.
- 2529 [76] Piet Verwilligen. Muons in the cms high level trigger system. *Nuclear
2530 and Particle Physics Proceedings*, 273-275:2509–2511, 2016. ISSN 2405-6014.

- 2531 doi: <https://doi.org/10.1016/j.nuclphysbps.2015.09.441>. URL <https://www.sciencedirect.com/science/article/pii/S240560141500930X>. 37th International Conference on High Energy Physics (ICHEP).
- 2534 [77] Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-
2535 probe technique. 2020. URL <https://cds.cern.ch/record/2724492>.
- 2536 [78] V.M. Abazov, B. Abbott, M. Abolins, et al. A novel method for modeling the
2537 recoil in W boson events at hadron colliders. *Nuclear Instruments and Methods*
2538 *in Physics Research Section A: Accelerators, Spectrometers, Detectors and Asso-*
2539 *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>.
- 2542 [79] CMS Collaboration. Search for an exotic decay of the Higgs boson to a pair
2543 of light pseudoscalars in the final state with two b quarks and two τ leptons in
2544 proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 785:462, 2018. doi:
2545 [10.1016/j.physletb.2018.08.057](https://doi.org/10.1016/j.physletb.2018.08.057).
- 2546 [80] CMS LUMI POG. Luminosity Physics Object Group: Recommendations, 2024.
2547 URL <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM>.
- 2548 [81] The modeling of the top quark p_T (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting>.
- 2550 [82] CMS BTV group. Methods to apply b-tagging efficiency scale fac-
2551 tors (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration>.
- 2553 [83] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in
2554 pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017. doi: [10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014).

- 2556 [84] Garvita Agarwal. Jet Energy Scale and Resolution Measurements in CMS. *PoS*,
2557 ICHEP2022:652, 2022. doi: 10.22323/1.414.0652.
- 2558 [85] CMS JERC group. Jet Energy Corrections (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- 2560 [86] CMS JERC group. Jet Energy Resolution (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution>.
- 2562 [87] CMS MUO POG. Muon Physics Object Group: Baseline muon selections for Run-II, 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2>.
- 2565 [88] CMS ELE POG. Electron Identification Based on Simple Cuts, 2024. URL
2566 <https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaPublicData>.
- 2567 [89] Jose Enrique Palencia Cortezon. Single top quark production at CMS. Technical
2568 report, CERN, Geneva, 2018. URL <https://cds.cern.ch/record/2640578>.
- 2569 [90] CMS Collaboration. Measurements of the electroweak diboson production cross
2570 sections in proton-proton collisions at $\sqrt{s} = 5.02$ TeV using leptonic decays.
2571 *Phys. Rev. Lett.*, 127(19):191801, 2021. doi: 10.1103/PhysRevLett.127.191801.
- 2572 [91] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment
2573 ten years after the discovery. *Nature*, 607(7917):60–68, 2022. doi: 10.1038/s41586-022-04892-x.
- 2575 [92] CMS JERC group. Jet energy scale uncertainty sources (TWiki), 2024. URL
2576 <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources>.
- 2577 [93] TOP Systematic Uncertainties (Run 2) (TWiki), 2024. URL <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopSystematics>.

- 2579 [94] Kyle Cranmer. Practical Statistics for the LHC. In *2011 European School of*
2580 *High-Energy Physics*, pages 267–308, 2014. doi: 10.5170/CERN-2014-003.267.
- 2581 [95] Elham Khazaie, Maryam Zeinali, Hamed Bakhshiansohi, and Abideh Jafari.
2582 CMS AN-21-058 (internal): Search for exotic decays of the Higgs boson to a
2583 pair of new light bosons with two muons and two b jets in the final states at
2584 $\sqrt{s} = 13$ TeV, 2021.
- 2585 [96] CMS Higgs Physics Analysis Group. Summary of 2HDM+S searches at 13 TeV
2586 (Run 2), 2024. URL <https://twiki.cern.ch/twiki/bin/view/CMSPublic/Summary2HDMSRun2>.