

**2 Higgs pair production in the four bottom quarks final state**

**3**

**Ana Luísa Moreira de Carvalho**

**4**

Thesis to obtain the Master of Science Degree in

**5**

**Physics Engineering**

**6**

Supervisor(s): Dr. José Ricardo Gonçalo  
Prof. Pedro Abreu

**7**

**Examination Committee**

**8**

Chairperson: Prof. Full Name

Supervisor: Prof. Full Name 1 (or 2)

Member of the Committee: Prof. Full Name 3

**9**

**Month Year**



## 10 Acknowledgments

11 I would like to express my sincere thanks to my supervisor Ricardo Gonçalo for his support, invaluable  
12 scientific advise and inspirational enthusiasm, without which this work would not have been concluded.  
13 To my co-supervisor, Pedro Abreu, I thank for his thorough reading of this thesis and for taking the time  
14 to engage in interesting physics discussions.

15 I acknowledge the funding of this research by Fundação para a Ciência e Tecnologia - FCT, through  
16 CERN/FIS-PAR/0008/2017 grant.

17 I am very thankful to Patricia Conde and Rute Pedro for fruitful discussions and very helpful advise.  
18 To my master colleagues in the LIP ATLAS group, Aidan Kelly, António Costa and Ricardo Barrué, I  
19 thank for the companionship and motivation throughout this journey.

20 I am thankful to Duarte Azevedo, Pedro Ferreira and Rui Santos for being our on-call theoreticians,  
21 always available to help.

22 I would like to thank all my colleagues in the LIP ATLAS group for making me feel very welcome.  
23 I extend my gratitude to everybody at LIP for making it a fantastic place to work. I would also like to  
24 acknowledge the LIP IT and secretariat for the technical support.

25 To the people at CERN, namely Clement Helsens and Michele Selvaggi, I thank for the numerous  
26 occasions in which they helped me regarding the FCC software and sample generation. I would also  
27 like to acknowledge the CERN IT for the very efficient service.

28 I would like to thank my boyfriend, João Augusto, for his never ending patience and unconditional  
29 support. To my friends of Chafarica I express my most sincere thanks for always being a great support,  
30 source of joy and laughter.

31 To my sister I thank for always welcoming me home with a smile, for the numerous time she offered  
32 to help and for all the memes that lighten up some of my days.

33 To my parents, I will never be able to thank enough. I thank them for their love, unconditional support,  
34 inspiration, help in times of need, ... .

## **Resumo**

- <sup>35</sup> Inserir o resumo em Português aqui com o mximo de 250 palavras e acompanhado de 4 a 6 palavras-chave...

- <sup>38</sup> **Palavras-chave:** palavra-chave1, palavra-chave2,...

39 **Abstract**

40 The production of pairs of Higgs bosons is a key benchmark process for future colliders because it  
41 provides crucial insight into the Electroweak Symmetry Breaking mechanism but it is out of the current  
42 reach of the Large Hadron Collider.

43 CERN is currently leading a study that analyzes the feasibility of a 100 km circular collider located  
44 in the Geneva area. The hadronic Future Circular Collider (FCC-hh) is expected to work at a center of  
45 mass energy of  $\sqrt{s} = 100$  TeV and to collect a total integrated luminosity of  $O(30)$  ab $^{-1}$ .

46 This thesis describes a Monte Carlo study targeting the search for  $hh \rightarrow b\bar{b}$  in a boosted kinematic  
47 regime, using proton-proton collisions at center of mass energy of  $\sqrt{s} = 100$  TeV. The focus is on the im-  
48 pact of the granularity of hadronic calorimeter on the significance,  $S/\sqrt{B}$ , targeting detector optimization  
49 studies for future colliders. In addition to traditional kinematic variables, jet substructure observables are  
50 also explored.

51 For the FCC-hh, the achieved significance is  $S/\sqrt{B} = \dots$  for an integrated luminosity of  $\mathcal{L} = 30$  ab $^{-1}$ ,  
52 which is above the  $5\sigma$  threshold. When using particle flow jets, the significance changes very little over  
53 the range of detector configurations considered. The change is more accentuated when using pure  
54 calorimeter jets.

55 **Keywords:** keyword1, keyword2,...



# Contents

56	Acknowledgments . . . . .	iii
58	Resumo . . . . .	iv
59	Abstract . . . . .	v
60	List of Tables . . . . .	xi
61	List of Figures . . . . .	xiii
62	Glossary . . . . .	xvi
63	<b>1 Introduction</b>	<b>1</b>
64	<b>2 The standard model and beyond</b>	<b>4</b>
65	2.1 The Standard Model of Particle Physics . . . . .	4
66	2.1.1 Higgs pair production . . . . .	10
67	2.2 Going beyond . . . . .	13
68	<b>3 Collider experiments</b>	<b>17</b>
69	3.1 Experimental aspects . . . . .	18
70	3.2 The Large Hadron Collider . . . . .	18
71	3.2.1 The ATLAS detector . . . . .	20
72	3.3 Jet reconstruction . . . . .	23
73	3.3.1 Jet properties and substructure observables . . . . .	25
74	3.3.2 Jet grooming algorithms . . . . .	27
75	3.4 Future Colliders . . . . .	27
76	3.4.1 The hadronic Future Circular Collider . . . . .	28
77	3.4.2 FCC-hh baseline detector . . . . .	29
78	3.4.3 FCC-hh physics program . . . . .	29
79	<b>4 State of the art</b>	<b>32</b>
80	4.1 Searches for Higgs pair production at the LHC . . . . .	32
81	4.2 Feasibility studies for high-luminosity and future colliders . . . . .	34
82	4.3 Hadronic calorimeter granularity studies . . . . .	35

83	<b>5 Sample generation and analysis tools</b>	<b>39</b>
84	5.1 Fast simulation workflow . . . . .	39
85	5.1.1 Particle flow and calorimeter reconstruction in Delphes . . . . .	40
86	5.1.2 FCC-hh software . . . . .	41
87	5.2 Signal and background samples . . . . .	42
88	5.2.1 MadGraph . . . . .	42
89	5.2.2 Pythia . . . . .	44
90	5.2.3 Delphes . . . . .	45
91	<b>6 Analysis</b>	<b>48</b>
92	6.1 Overview of the $hh \rightarrow b\bar{b}b\bar{b}$ channel . . . . .	48
93	6.1.1 Event pre-selection . . . . .	48
94	6.1.2 Signal characterization . . . . .	49
95	6.1.3 Backgrounds . . . . .	50
96	6.2 Analysis strategy . . . . .	54
97	6.2.1 Implementation of b-tagging . . . . .	54
98	6.2.2 Baseline analysis . . . . .	55
99	6.2.3 Optimization . . . . .	56
100	6.2.4 Handling of uncertainties . . . . .	60
101	6.2.5 Trigger discussion . . . . .	61
102	<b>7 Results</b>	<b>63</b>
103	7.1 Di-Higgs discovery potential at the FCC-hh . . . . .	63
104	7.1.1 Comparing with the ATLAS detector . . . . .	64
105	7.2 Hadronic calorimeter granularity studies for future colliders . . . . .	65
106	7.2.1 Resolution of jet mass and N-subjetiness . . . . .	65
107	7.2.2 Signal efficiency and significance . . . . .	68
108	<b>8 Conclusions</b>	<b>73</b>
109	<b>Bibliography</b>	<b>75</b>
110	<b>A Jet radius discussion</b>	<b>83</b>
111	<b>B Additional background processes</b>	<b>84</b>
112	<b>C Softdrop mass</b>	<b>85</b>
113	<b>D Samples generation: parameters</b>	<b>86</b>
114	D.1 MadGraph . . . . .	86
115	D.2 Pythia . . . . .	86
116	<b>E Cutflow tables</b>	<b>88</b>





# List of Tables

118	3.1 ATLAS tile and liquid argon hadronic calorimeters: summary of the pseudorapidity coverages and transversal segmentation (granularity). . . . .	21
120		
121	3.2 Comparison between the working parameters of the LHC and of the FCC-hh. The values of the number bunches and of the number of events per bunch crossing are given assuming a bunch spacing of 25 ns. . . . .	28
123		
124	3.3 Hadronic calorimeter layout, granularity and energy resolution. . . . .	30
125		
126	5.1 Summary of the effective cross sections ( $\sigma \times BR$ ) and k factors of the signal and background samples used in the analysis. . . . .	46
127		
128	5.2 Summary of the benchmark granularity configurations of the HCAL. . . . .	47
129		
130	5.3 Summary of some key detector parameters for the FCC-hh and ATLAS detectors. . . . .	47
131		
132	6.1 b-Tagging (black), c (blue) and light (red) mistag probabilities as a function of $\eta$ and $p_T$ of the (sub)jet. The momentum dependent factor, $(1 - p_T/15000)$ , is common to the three probabilities. . . . .	55
133		
134		
135		
136		
137		
138	7.1 Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM and 2HDM). The absolute value of expected events after some key selection cuts is shown in curved brackets. The number of expected events is normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC-hh baseline detector design, as implemented in Delphes by the FCC-hh study group. . . . .	64
139		
140		
141		
142		
143	7.2 Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b + j$ , $jj + 0/1/2j$ and $t\bar{t}+0/1/2 j$ ). The absolute value of expected events after some key selection cuts is shown in curved brackets. The number of expected events is normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC-hh baseline detector design, as implemented in Delphes by the FCC-hh study group. . . . .	65

144	7.3 Cumulative efficiency, in percentage, of each event selection criterion for the signal sam-	
145	ples (SM and 2HDM). The absolute value of expected events after some key selection	
146	cuts is shown in curved brackets. The number of expected events is normalized to	
147	$\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results	
148	were obtained using the ATLAS detector design, as implemented in Delphes. . . . .	66
149	7.4 Cumulative efficiency, in percentage, of each event selection criterion for the background	
150	samples ( $4b + j$ , $jj + 0/1/2j$ and $t\bar{t}+0/1/2 j$ ). The absolute value of expected events after	
151	some key selection cuts is shown in curved brackets. The number of expected events	
152	is normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts.	
153	These results were obtained using the ATLAS detector design, as implemented in Delphes. 66	
154	7.5 Significances achieved with the ATLAS and FCC-hh detector configurations for $\mathcal{L} = 30 \text{ ab}^{-1}$	
155	for the three benchmark signal models: SM, 1 TeV DM mediator and 2HDM with $m_H = 900$	
156	GeV. . . . .	67
157	B.1 Effective cross section ( $\sigma \times \text{BR} \times \text{k-factor}$ ), efficiency and expected number of events for	
158	$\mathcal{L} = 30 \text{ ab}^{-1}$ for the $t\bar{t}h + 0/1 j(h \rightarrow b\bar{b})$ and $h + 0/1/2 j(h \rightarrow b\bar{b})$ backgrounds. . . . .	84
159	D.1 Generator (MadGraph5) level cuts for the signal and background samples. . . . .	87
160	D.2 Pythia settings for the signal and background samples. . . . .	87
161	E.1 oi . . . . .	88
162	E.2 oi . . . . .	88
163	E.3 oi . . . . .	89
164	E.4 oi . . . . .	90

# <sup>165</sup> List of Figures

<sup>166</sup> 2.1	Schematic representation of the Standard Model of elementary particles. Particle are represented inside squares. The electric and color charge are shown in the right upper corner of each square. The spin is shown in the right lower corner. The mass of the particles is also shown, in electron Volt. . . . .	<sup>167</sup> 5
<sup>170</sup> 2.2	Postulated shape of the Higgs potential for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right). . . . .	<sup>171</sup> 8
<sup>171</sup> 2.3	Feynman diagrams of Higgs pair production from gluon fusion. Triple vertex diagram (left) and box diagram (right). The minus sign between the diagrams indicates that they interfere destructively. . . . .	<sup>172</sup> 11
<sup>174</sup> 2.4	Total cross sections at the NLO in QCD for the six largest HH production channels at pp colliders. The thickness of the lines corresponds to the scale and PDF uncertainties added linearly. . . . .	<sup>175</sup> 13
<sup>177</sup> 2.5	Here, $\alpha_1$ , $\alpha_2$ and $\alpha_3$ represent the coupling constants of $U_Y(1)$ , $SU_L(2)$ and $SU_{color}(3)$ , respectively. What is shown is the variation of the inverse of the coupling with the energy: on the left for the SM and on the right for the MSSM. Plots from [23]. . . . .	<sup>178</sup> 14
<sup>180</sup> 2.6	BSM contributions to Higgs pair production from the spin-2 Kaluza-Klein graviton (a) and from the SUSY partner of the top quark (b). Figures from Refs. [26, 27]. . . . .	<sup>181</sup> 16
<sup>182</sup> 3.1	ATLAS detector. . . . .	<sup>183</sup> 22
<sup>183</sup> 3.2	FCC detector baseline concept. . . . .	<sup>184</sup> 30
<sup>184</sup> 4.1	Exclusion limits (at 95% CL) on the cross section of resonant Higgs pair production from the decay of heavy scalar particles. These results were obtained using proton-proton collision data collected with the CMS (a) and ATLAS (b) detector at a CM energy of $\sqrt{s} = 8$ TeV and result from the combination of different channels. Plots from [70] and [74]. . . . .	<sup>185</sup> 34
<sup>188</sup> 4.2	Azimuthal distribution of the energy deposits in the ECAL and HCAL for a pair of $K_L^0$ with $E = 100$ GeV for an hadronic calorimeter with $20 \text{ cm} \times 20 \text{ cm}$ (a) and $5 \text{ cm} \times 5 \text{ cm}$ (b). Figures from [79] (based on [78]). . . . .	<sup>189</sup> 36
<sup>191</sup> 4.3	Jet mass (a) and jet mass resolution (b) plots in $t\bar{t}$ events for eflow jets with $p_T > 3$ TeV for three different HCAL and ECAL configurations: HCAL(ECAL) $0.1(0.025) \eta \times 5.6(1.4)^\circ \phi$ , HCAL(ECAL) $0.05(0.012) \eta \times 2.8(0.7)^\circ \phi$ and HCAL(ECAL) $0.025(0.006) \eta \times 1.4(0.35)^\circ \phi$ . Plots from Ref. [79]. . . . .	<sup>192</sup> 37

195	4.4	$\tau_{32}$ resolution plots for QCD dijet events reconstructed using eflow jets (a) and tower jets (b). Plots from Ref. [80]. . . . .	38
196			
197	4.5	$\tau_{21}$ variable for 20 TeV $W$ (black) and QCD jets (red) for three HCAL granularities: $\Delta\eta \times$ $\Delta\phi = 0.1 \times 0.1$ (left), $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$ (middle) and $\Delta\eta \times \Delta\phi = 0.005 \times 0.005$ (right). Plots from Ref. [79]. . . . .	38
198			
199			
200	6.1	Higgs pairs branching ratios (left) and event topology targeted by this analysis (right). . . . .	49
201	6.2	$\Delta\phi$ between the Higgs candidates, $\Delta\phi(hh)$ , (left) and $\Delta R$ between the b quarks coming from the decay of the leading Higgs candidate, $h_1$ (right). To obtain this plot the cut $ M - 125  < 40$ GeV was applied in addition to the pre-selection cuts.. . . . .	51
202			
203	6.3	Example of diagrams that contribute to the QCD multijet background: five final state jets, four of which are b-jets (left), three final state jets (middle) and four final state jets (right). Here, $q$ stands for a light quark/jet. . . . .	52
204			
205	6.4	Dominant diagrams of $pp \rightarrow t\bar{t}$ at LO. . . . .	52
206			
207	6.5	$p_T$ distributions for the leading (left) and sub leading (right) Higgs candidates. The signal is the SM $hh \rightarrow b\bar{b}$ process. The histograms are normalized to unit area. . . . .	53
208			
209	6.6	$\eta$ distribution for the leading Higgs candidate (left) and softdrop mass distribution for the leading Higgs candidate (right). The signal is the SM $hh \rightarrow b\bar{b}$ process. The histograms are normalized to unit area. . . . .	53
210			
211	6.7	Distributions of the $\Delta\eta$ between the Higgs candidates (left) and of the $\tau_{21}$ variable for the leading Higgs candidate (right). The histograms are normalized to unit area. . . . .	53
212			
213	6.8	Minimun $\Delta R$ between b quarks and subjets of the $R = 0.8$ jets. . . . .	54
214			
215	6.9	$p_T$ distributions for the leading (a) and sub leading Higgs candidates (b). The histograms are normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	56
216			
217	6.10	(a) $p_T$ distribution for the Higgs pair, (b) $\tau_{21}$ variable for the leading Higgs candidate. The histograms are normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	57
218			
219	6.11	(a) $H_2$ variable for the leading Higgs candidate, (b) softdrop mass distribution for the leading Higgs candidate. The histograms are normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	57
220			
221	6.12	$S/\sqrt{B}$ as a function of the cut on the $p_T$ of the leading Higgs candidate, $p_T(h_1)$ (a) after the pre-selection cuts and of the sub leading Higgs candidate, $p_T(h_2)$ (b) after the cut on $p_T(h_1)$ . . . . .	59
222			
223	6.13	$S/\sqrt{B}$ as a function of the cut on the $p_T$ of the Higgs pair, $p_T(hh)$ (a) after the cut on $p_T(h_1)$ and as a function of the $\tau_{21}$ variable for leading Higgs candidate, $\tau_{21}(h_1)$ (b) after the cut on $p_T(hh)$ . . . . .	59
224			
225	6.14	$S/\sqrt{B}$ as a function of the cut on the $\Delta\eta$ between the Higgs candidates, $\Delta\eta(hh)$ (a) after the cuts on the transverse momenta and on the $\tau_{21}$ variables and as a function of the second Fox-Wolfram momentum for the leading Higgs candidate, $H_l(h_1)$ (b) after the cuts on the transverse momentums, $\tau_{21}$ and $\Delta\eta$ . . . . .	59
226			
227			
228			
229			
230			
231			

232	7.1	Leading Higgs candidate softdrop mass after the pre-selection cuts for particle flow (a)	69
233		and calorimeter (b) jets. The colors indicate the different detector configurations. The	
234		x axis range is from 80 GeV to 160 GeV in order to make the differences between the	
235		histograms more clear. . . . .	
236	7.2	Invariant mass of the Higgs pair for the SM signal. The plots include the events that pass	69
237		all the analysis cuts. The colors indicate the different detector configurations. . . . .	
238	7.3	(a) Invariant mass of the Higgs pair for the 1 TeV DM mediator, (b) Invariant mass of the	70
239		Higgs pair for the 2HDM with $m_H = 900$ GeV. The plots include the events that pass all	
240		the analysis cuts. The colors indicate the different detector configurations. . . . .	
241	7.4	(a) Leading Higgs candidate $\tau_{21}$ distribution for eflow jets, (b) Leading Higgs candidate $\tau_{21}$	
242		distribution for HCAL jets. The colors indicate the different detector configurations. The	
243		distributions are shown for the signal (filled lines) and for the $4b + j$ background (dashed	
244		lines). . . . .	70
245	7.5	(a) Signal efficiency as a function of the detector configuration for particle flow jets, (b)	
246		Signal efficiency as a function of the detector configuration for calorimeter jets. Three	
247		signal models are shown: SM (filled squares), 1 TeV DM mediator (empty squares) and	
248		type II 2HDM with $m_H = 900$ GeV (crosses). The error bars are drawn but are smaller	
249		than the markers. . . . .	71
250	7.6	Significance as a function of the detector configuration for the SM signal, for the baseline	
251		(a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and	
252		pure HCAL jets (triangles). . . . .	71
253	7.7	Significance as a function of the detector configuration for the DM mediator signal, for	
254		the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets	
255		(squares) and pure HCAL jets (triangles). . . . .	72
256	7.8	Significance as a function of the detector configuration for the 2HDM signal, for the base-	
257		line (a) and optimized (b) analyses. The significances are shown for eflow jets (squares)	
258		and pure HCAL jets (triangles). . . . .	72
259	A.1	$\Delta R(b, \bar{b})$ distributions for $p_T(h_1) > 200/300$ GeV (solid blue/black) and for $p_T(h_1, h_2) >$	
260		200/300 GeV (dashed blue/black). . . . .	83
261	C.1	Correlation between the maximum $\Delta R$ between the Higgs candidate jet and one of the b	
262		quarks (y axis) and the jet's mass (x axis). The horizontal line corresponds to $\Delta R = 0.8$ . .	85
263	F.1	. The histograms are normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	92

# Glossary

<b>2HDM</b>	Two Higgs Doublet Model
<b>ATLAS</b>	A Toroidal LHC Apparatus
<b>BR</b>	Branching Ratio
<b>BSM</b>	Beyond the Standard Model
<b>CERN</b>	European Organization for Nuclear Research
<b>CMS</b>	Compact Muon Solenoid
<b>ECAL</b>	Electromagnetic Calorimeter
<b>EM</b>	Electromagnetic
<b>HCAL</b>	Hadronic Calorimeter
<b>ID</b>	Inner Detector
<b>LHC</b>	Large Hadron Collider
<b>LO</b>	Leading Order
<b>NLO</b>	Next to Leading Order
<b>QCD</b>	Quantum Chromodynamics
<b>QED</b>	Quantum Electrodynamics
<b>QFT</b>	Quantum Field Theory
<b>SM</b>	Standard Model
<b>SUSY</b>	Super Symmetry
<b>TileCal</b>	Tile Calorimeter
<b>VEV</b>	Vacuum Expectation Value

# <sup>265</sup> Chapter 1

## <sup>266</sup> Introduction

<sup>267</sup> It is the ultimate goal of particle physics to discover and study all of Nature's fundamental particles and  
<sup>268</sup> to understand their interactions. Through a joint endeavor of theorists and experimentalists, models that  
<sup>269</sup> describe particle's dynamics and properties can be precisely probed at collider experiments such as the  
<sup>270</sup> Large Hadron Collider (LHC).

<sup>271</sup> We know today that matter particles interact by means of four fundamental forces: electromagnetic,  
<sup>272</sup> weak, strong and gravitational, of which the first three are associated with a mediator particle. We even  
<sup>273</sup> know that a very special particle, the Higgs boson, is linked to the mechanism that generates the mass  
<sup>274</sup> of the mediator particles. It is called the Electroweak Symmetry Breaking (EWSB) mechanism and it  
<sup>275</sup> involves the spontaneous breaking of a gauge symmetry. Through Yukawa couplings, the Higgs boson  
<sup>276</sup> is also responsible for the masses of fermions. This knowledge is beautifully summarized in the Standard  
<sup>277</sup> Model of Particle Physics (SM) that was developed in the 1960's and 1970's, long before many of the  
<sup>278</sup> particles it predicts were discovered. The extraordinary precision of the predictions it delivers make it a  
<sup>279</sup> very successful model. The Higgs boson, was the most recent elementary particle to be discovered in  
<sup>280</sup> 2012 at the LHC, which marks an important point in the history of particle physics: we have now found  
<sup>281</sup> all the particles predicted by the SM and yet we know that it cannot be the whole story. Mainly because  
<sup>282</sup> there are experimental evidences of physics that it cannot explain.

<sup>283</sup> From the theoretical point of view, this is enough motivation to construct models that extend the SM  
<sup>284</sup> but that can still deliver predictions that are compatible with experimental data. From the experimental  
<sup>285</sup> standpoint, this is an indication that we need to keep increasing the precision of our measurements and  
<sup>286</sup> probing new kinematic regimes in the hope of finding more discrepancies with the SM or some hint that  
<sup>287</sup> some new phenomenon is taking place.

<sup>288</sup> A higher precision requires a larger integrated luminosity and a larger center of mass energy opens  
<sup>289</sup> the door for the exploration of new kinematic regimes. Very recently, work towards the upgrade of the  
<sup>290</sup> LHC to its High-Luminosity (HL) version has began. It is expected to work for a period of ten years  
<sup>291</sup> between 2026 and 2036 and it will extend the experimental reach of the LHC. In order to keep extending  
<sup>292</sup> the physics reach of the LHC and HL-LHC, new colliders with unprecedentedly high center of mass  
<sup>293</sup> (CM) energies are currently being designed in the hope that they begin to deliver data shortly after the

294 HL-LHC has reached its full discovery potential. One of these projects is the hadronic Future Circular  
295 Collider (FCC-hh) that consists of a 100 km ring located in the Geneva area and it is expected to work  
296 at a CM energy of 100 TeV. The FCC-hh will deliver a peak luminosity of  $\mathcal{L} = 30 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$  in  
297 its ultimate phase. This will result in  $O(30) \text{ ab}^{-1}$  per experiment which corresponds to ten times the  
298 expected integrated luminosity by the end of the HL-LHC operation.

299 The next milestone for the FCC-hh project is the submission of a Conceptual Design Report by the  
300 end of 2018. This document will be used as input for the next meeting of the European Strategy for  
301 Particle Physics that will take place in the beginning of 2019. It should present a baseline design for the  
302 detector, a first cost estimate and preliminary analysis for physics benchmark processes demonstrating  
303 the physics reach of such a machine.

304 Both in the HL-LHC and in future colliders, one of the most important benchmark processes is the  
305 production of pairs of Higgs bosons. Firstly, this process is predicted by the SM but has not yet been  
306 measured which is due to its very small cross section and overwhelming backgrounds. Furthermore,  
307 it provides unique insight into the EWSB mechanism because it is sensitive to the shape of the Higgs  
308 potential and can also be used to probe physics beyond the SM (BSM).

309 In this work we study Higgs pair production at a center of mass energy of  $\sqrt{s} = 100 \text{ TeV}$  using the final  
310 state with four b quarks. We choose this final state because it benefits from the large branching fraction  
311 of the Higgs boson to a pair of b quarks. However, in this channel, the multijet production through  
312 Quantum Chromodynamics (QCD) interactions is extremely overwhelming. Nonetheless, it is a well  
313 known feature of QCD interactions that the partons (and jets) produced tend to have a low transverse  
314 momentum ( $p_T$ ). This indicates that exploring a high  $p_T$  region of the phase space could be the key  
315 to suppress this background. In this kinematic regime, traditional jet reconstruction techniques, that  
316 try to establish a one-to-one correspondence between partons and jets, begin to fail because the decay  
317 products of heavy and highly boosted particles become more collimated as the  $p_T$  of the mother particles  
318 increase. State of the art jet reconstruction techniques make use of jets with a larger radius parameter  
319 to reconstruct both decay products (two b quarks) as a single jet that can then be used as a proxy for the  
320 original particle (a Higgs boson). In order to recover as much information as possible from these jets, it  
321 is important to analyze their intrinsic structure, referred to as substructure.

322 Jet substructure techniques explore the existence of subjets inside the large- $R$  jets. For example, a  
323 jet containing the two b quarks from the decay of a Higgs boson is expected to be more consistent with  
324 having two subjets than a jet produced by a QCD process. From the standpoint of detector design for  
325 the FCC-hh, as well as future upgrades of existent detectors, the ability to resolve the substructure of  
326 large- $R$  jets is a key requirement for which can be helpful a high granularity hadronic calorimeter (HCAL).

327 The main goal of this thesis is to use boosted di-Higgs production in the four b quarks final state to  
328 study the influence of the granularity of the HCAL in significance of the detection that can be achieved.

329 Chapter 2 presents an overview of the SM. It summarizes its particle content and interactions and  
330 introduces the mathematical formulation of the EWSB breaking mechanism. The successes and short-  
331 comings of the SM are presented and several BSM models are introduced and their motivations dis-  
332 cussed. Finally, a theoretical description of the production of Higgs pairs is provided.

333        The FCC-hh baseline accelerator and detector were highly based on LHC and its current experiments,  
334        namely ATLAS and CMS. In chapter 3, after a brief discussion of the general features of particle  
335        accelerators, we introduce the LHC and the ATLAS experiment. A discussion of jet reconstruction is  
336        included. We then introduce the FCC-hh accelerator and its current baseline detector design. Some key  
337        detector parameters are compared to ATLAS.

338        In chapter 4 we present the state of the art of searches for Higgs pair production at the LHC and  
339        of feasibility studies targeting searches for this process at future colliders (High Luminosity LHC and  
340        FCC-hh). We also review previous studies that focus on the impact of the HCAL granularity on the jet  
341        mass and jet substructure observables resolution.

342        In chapter 5 we describe in detail the Monte Carlo samples that were used in this work. In addition,  
343        the different detector configurations that were tested are described. Chapter 6 entails the description  
344        of the analysis strategy and its optimization. A comprehensive characterization of the signal and back-  
345        grounds processes is provided.

346        In chapter 7 we present and discuss the main results. We focus on the resolution of the jet mass  
347        and of the ratio of N-subjetiness variables, namely,  $\tau_{21}$ . In addition, we analyze how the significance  
348        chances as the granularity of the HCAL is varied. The results obtained using particle flow jets and pure  
349        calorimeter jets are compared.

350        Finally, conclusions are drawn in chapter 8.

351 **Chapter 2**

352 **The standard model and beyond**

353 The Standard Model (SM) is the theoretical framework that summarizes our present knowledge of particle physics. In section 2.1, we provide an overview of this model, focusing on the Higgs mechanism. In 354 section 2.2, we motivate the need to explore models beyond the SM (BSM) and introduce some of the 355 most well known BSM models. In section 2.1.1, we provide a theoretical description of the production of 356 Higgs boson pairs which is the physical process that is under study throughout this work. 357

358 **2.1 The Standard Model of Particle Physics**

359 The Standard Model of particle physics summarizes our present knowledge of fundamental particles 360 and their interactions. It is formulated in the framework of Quantum Field Theory (QFT) and describes 361 the subatomic world in terms of fields whose excitations are the particles we can detect. The particle 362 content of the SM is summarized in figure 2.1. Each particle is represented inside a square. The electric 363 and color charge are shown in the right upper corner and the spin in the right lower corner. The mass 364 of the particles are given in electron-Volt (eV) on top of each square (M standing for Million eV and G 365 standing for Billion eV). There are two types of fundamental particles: matter particles and force carriers.

366 Matter particles are the building blocks of all the matter in our world. They come in two groups, 367 leptons and quarks. Quarks make up atomic nuclei; leptons, namely electrons and muons, can orbit 368 atomic nuclei forming atoms. Quarks and leptons are elementary fermions which means they have half-369 integer spin. There are six flavors of quarks: three of the 'up type' (up, charm and top represented by 370  $u$ ,  $d$  and  $t$ ) with electric charge of  $+2/3$  and three of 'down type' (down, strange and bottom represented 371 by  $d$ ,  $s$  and  $b$ ) with electric charge of  $-1/3$ . Similarly, we have three leptons with charge  $-1$  (electron, 372 muon and tau represented by  $e$ ,  $\mu$  and  $\tau$ ) and three neutral leptons (electron, muon and tau neutrinos 373 represented by  $\nu$  with the symbol of the corresponding charged lepton as subscript) that are, within the 374 SM, massless. We can classify quarks and leptons in three generations, each composed of an up type 375 and down type quark or of a charged lepton and the corresponding neutrino.

376 The force carriers, technically called gauge bosons, are particles associated with the fundamental

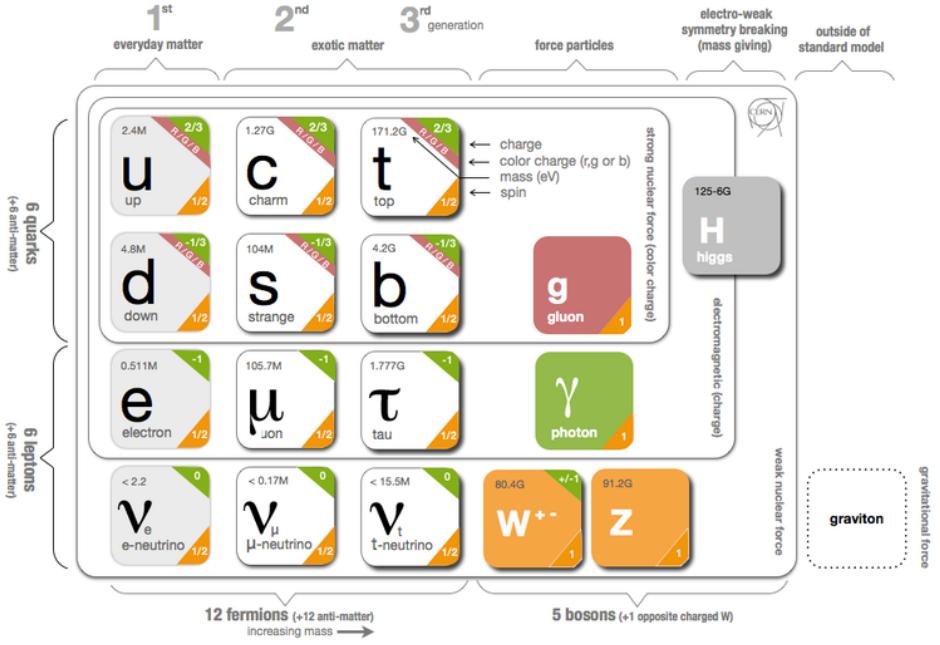


Figure 2.1: Schematic representation of the Standard Model of elementary particles. Particle are represented inside squares. The electric and color charge are shown in the right upper corner of each square. The spin is shown in the right lower corner. The mass of the particles is also shown, in electron Volt.

377 interactions: strong, electromagnetic, weak and gravitational <sup>1</sup>. Each interaction can be interpreted as  
 378 the result of the exchange of the corresponding gauge boson. Gluons ( $g$ ) and the photon ( $\gamma$ ) are the  
 379 mediators of the strong and electromagnetic interactions, respectively. They are massless, electrically  
 380 neutral and have spin 1. The  $W^\pm$  and  $Z$  bosons are the mediators of the weak interaction and have a  
 381 mass of 82 and 91 GeV, respectively. The  $W^+$  and  $W^-$  bosons have electric charges of +1 and -1,  
 382 respectively and spin 1. The  $Z$  boson is electrically neutral and also has spin 1. The gauge bosons can  
 383 also be referred to as vector bosons because they have spin equal to one.

384 In addition to matter particles and gauge bosons, the theoretical formulation of the SM rests on the  
 385 existence of the Higgs boson that is an electrically neutral and spin 0 particle. It has a mass of 125 GeV  
 386 and it interacts with every particle that has mass.

387 Historically, an empirically successful quantum theory of electromagnetism, Quantum Electrodynamics  
 388 (QED), was developed in the late 1940's. In the early 1950's there were high hopes that quantum  
 389 theories could also be formulated for the weak and strong interactions. This is the context in which  
 390 Yang-Mills theories emerged. They extend the concept of gauge theory from abelian groups, that lead  
 391 to the development of QED, to non-abelian gauge groups. However, the quanta of the fields predicted  
 392 by these theories must be massless in order to maintain gauge invariance. Therefore, they were set  
 393 aside until the 1960's when the idea of particles acquiring mass through symmetry breaking in massless  
 394 theories was put forward by Goldstone [1], Nambu and Jona-Lasinio [2]. In the following paragraphs we  
 395 discuss in more detail the caveats of Yang-Mills theories and the phenomenon of Spontaneous Symme-

<sup>1</sup> The gauge boson that corresponds to the gravitational force has not yet been found. In addition we still do not have a theory that successfully describes gravitation in the framework of QFT so we will not include the gravitational force or its gauge boson in any of the following discussions. Moreover, the gravitational interaction is not relevant for the energy ranges studied.

396 try Breaking (SSB) as the basis of the modern Higgs mechanism. We then describe this mechanism in  
 397 the framework of the SM.

398 On the one hand, if one takes a Yang-Mills theory, it becomes clear that it is not possible to include  
 399 in the Lagrangian a mass term for the gauge bosons because it is not invariant under a gauge transfor-  
 400 mation. This would not be a problem if we just wanted to describe electromagnetic or strong interactions  
 401 because the gauge bosons associated with these interactions, the photon and the gluon, are indeed  
 402 massless. However, for the weak interactions this is not the case. Even before the discovery of the  
 403  $Z$  and  $W^\pm$  bosons [3, 4] there was experimental evidence of the short range character of the weak  
 404 interactions which indicated that the corresponding gauge bosons should be massive.

405 On the other hand, spontaneous symmetry breaking (SSB) is a phenomenon through which the  
 406 invariance of a system under a certain symmetry group is destroyed [5]. The system may then be in-  
 407 variant under a subgroup of the initial symmetry but the invariance under the original symmetry group is  
 408 no longer present. In particle physics, this happens because the vacuum of the system (lowest energy  
 409 states) does not share the symmetry of the Lagrangian. The SSB mechanism predicts the existence  
 410 of scalar massless particles, the Nambu-Goldstone bosons, as a consequence of the Goldstone theo-  
 411 rem [1] (the number depends on the number of generators of the original and final symmetry groups).  
 412 Though, when considering this mechanism we get once again massless particles which does not seem  
 413 to be a step in the right direction if we wish to describe weak interactions.

414 However, the real breakthrough occurs when we combine a theory with local gauge invariance with  
 415 the mechanism of SSB. In this case the Nambu-Goldstone bosons do not appear and it is possible to  
 416 give mass to the gauge bosons. This is the Higgs mechanism, proposed independently by P.W. Higgs  
 417 [6], F. Englert and R. Brout [7] and by G. Guralnik, C. R. Hagen and T. Kibble [8] in 1964.

418 The SM is a non-abelian gauge theory with spontaneous symmetry breaking. It is locally invariant  
 419 under the following symmetry group:

$$SU_{color}(3) \times SU_L(2) \times U_Y(1), \quad (2.1)$$

420 where the  $SU_{color}(3)$  group describes the strong interactions (QCD) and the  $SU_L(2) \times U_Y(1)$  group  
 421 describes the electroweak interactions. Here,  $L$  stands for left and  $Y$  stands for hypercharge. In the SM  
 422 the Higgs mechanism, which we now describe, is realized in the  $SU_L(2) \times U_Y(1)$  group. The Lagrangian  
 423 corresponding to the Higgs and gauge sectors of this theory is given by:

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) - \frac{1}{4} W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}, \quad (2.2)$$

where the Higgs potential,  $V(\phi^\dagger \phi)$ , is given by:

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.3)$$

424  $W_{\mu\nu}^a$  and  $B_{\mu\nu}$  are the field tensors, defined as a function of the gauge fields of  $SU_L(2)$  and  $U_Y(1)$ ,  
 425 respectively,  $W_\mu^a$  ( $a = 1, 2, 3$ ) and  $B_\mu$ :

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{abc}W_\mu^b W_\nu^c \quad (2.4)$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (2.5)$$

426 where  $g$  is the coupling constant associated with the  $SU_L(2)$  group and  $\epsilon^{abc}$  is the completely anti-  
 427 symmetric tensor in 3 dimensions. The covariant derivative,  $D_\mu$ , is introduced to preserve local gauge  
 428 invariance and is given by:

$$D_\mu \phi = \left( \partial_\mu + igW_\mu^a T^a + i\frac{g'}{2} B_\mu \right) \phi. \quad (2.6)$$

429  $T^a = \frac{\tau^a}{2}$  (where  $\tau^a$  are the Pauli matrices) are the  $SU_L(2)$  group generators in the fundamental representation and  $g'$  is the coupling constant associated with the  $U_Y(1)$  group.

Due to the requirement of Lorentz invariance, only the scalar field,  $\phi$ , can have a vacuum expectation value (VEV),  $v$ , different from zero<sup>2</sup>. The values of  $v$  are determined by the minima of the potential:

$$v = 0 \quad \text{or} \quad v = \sqrt{\frac{-\mu^2}{2\lambda}}. \quad (2.7)$$

431 For the equation on the right (for which we get  $v \neq 0$ ) we only obtain a real value for  $v$  (which is a  
 432 requirement for the VEV of a theory) if  $\mu^2 < 0$ . Therefore we conclude that the equation on the right  
 433 corresponds to  $\mu^2 \leq 0$  while the equation on the left corresponds to  $\mu^2 \geq 0$ . In both cases  $\lambda$  has to  
 434 be larger than zero to guarantee that the energy is bounded from below<sup>3</sup> because in Eq. 2.3  $\lambda$  is the  
 435 coefficient of the term with the highest power in  $\phi$  and therefore determines the concavity of the potential.

436 The shapes of the Higgs potential for  $\mu^2 > 0$  and  $\mu^2 < 0$  are shown in Figures 2.2(a) and 2.2(b),  
 437 respectively. For  $\mu^2 > 0$  (left) we have a single minimum located at  $\langle\phi\rangle = 0$ . For  $\mu^2 < 0$  (right)  
 438 the potential has the shape of a 'Mexican hat'. There is an infinite number of minima located in a  
 439 circumference centered at zero. In this case the minima occur for  $\langle\phi\rangle, \langle\phi^\dagger\rangle \neq 0$ . Therefore the fields  
 440 acquire a VEV different than zero and this is what leads to the SSB.

We can now write the scalar field in terms of its minimum value,  $v$ , and of oscillations around that minimum,  $h$  (which corresponds to the Higgs field):

$$\phi = \begin{bmatrix} 0 \\ v + \frac{h}{\sqrt{2}} \end{bmatrix} \quad (\text{unitary gauge}). \quad (2.8)$$

441 If we expand the first term of the Lagrangian shown in Eq. 2.2 using Eq. 2.6 and Eq. 2.8 and taking into  
 442 consideration that  $W_\mu^a T^a$  represents a sum over all values of  $a$  we get

---

<sup>2</sup>The other fields that appear in Eq. 2.2 are vector fields. If they were to acquire a VEV different from zero that would break Lorentz invariance.

<sup>3</sup>In a purely mathematical formulation this means that the function that represents the Higgs potential is concave upwards.

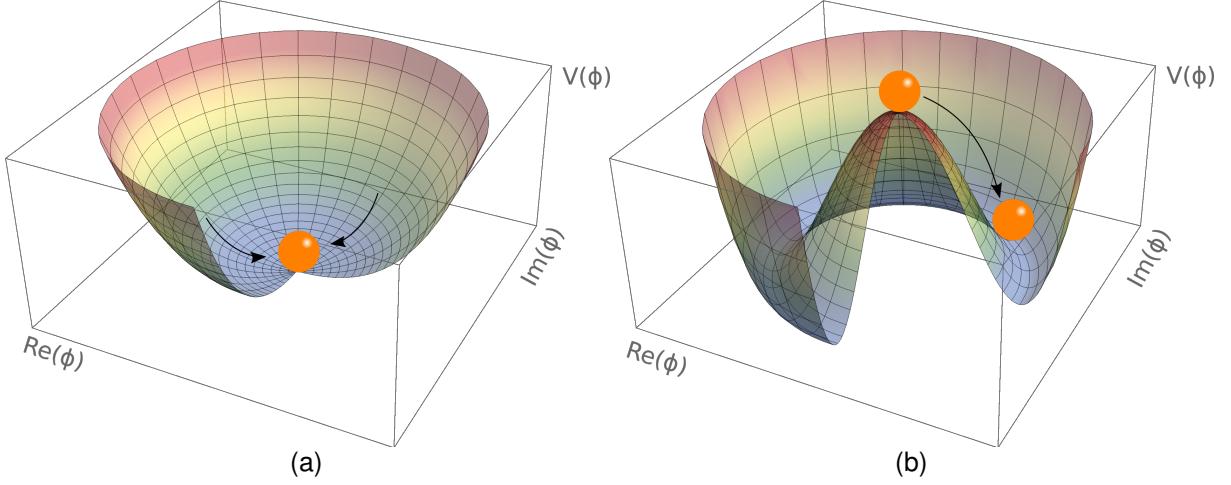


Figure 2.2: Postulated shape of the Higgs potential for  $\mu^2 > 0$  (left) and  $\mu^2 < 0$  (right).

$$\mathcal{L} = \frac{1}{4} \left( v^2 + \frac{h^2}{2} + \frac{2}{\sqrt{2}} vh \right) [g^2 (W_\mu^1 W^{1\mu} + W_\mu^2 W^{2\mu} + W_\mu^3 W^{3\mu}) - 2gg' B^\mu W_\mu^3 + g'^2 B_\mu B^\mu] + \dots . \quad (2.9)$$

We see that for the  $W_\mu^1$  and  $W_\mu^2$  fields we have only terms that are quadratic in these fields. These correspond to mass terms. However, for the  $W_\mu^3$  and  $B_\mu$  fields there is a term that mixes the two fields. To obtain the physical states of the theory we need to transform these fields in order to get rid of the mixing term which is not physical. We can start by writing the last three terms of Eq. 2.9 in a matrix form and diagonalize the corresponding matrix:

$$\begin{bmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{bmatrix} \begin{bmatrix} W_\mu^3 \\ B_\mu \end{bmatrix} \xrightarrow{\text{Diagonalization}} \begin{bmatrix} 0 & 0 \\ 0 & g^2 + g'^2 \end{bmatrix} \begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix}. \quad (2.10)$$

$A_\mu$  and  $Z_\mu$  are the physical fields that are related with  $W_\mu^3$  and  $B_\mu$  by means of a rotation matrix:

$$\begin{bmatrix} A_\mu \\ Z_\mu \end{bmatrix} = \begin{bmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{bmatrix} \begin{bmatrix} W_\mu^3 \\ B_\mu \end{bmatrix} \quad (2.11)$$

where  $\theta_W$  is the Weinberg angle. By inverting this relation we can write  $W_\mu^3$  and  $B_\mu$  as a function of  $A_\mu$  and  $Z_\mu$ . Replacing in Eq. 2.9 and imposing that the  $A_\mu$  field has zero mass we can determine  $\theta_W$ :  $\tan \theta_W = \frac{g'}{g}$ . The Lagrangian of Eq. 2.9 takes then the form

$$\mathcal{L} = \frac{1}{2} (v^2 g^2) (W_\mu^1 W^{1\mu} + W_\mu^2 W^{2\mu}) + \frac{1}{2} (v^2 [g^2 + g'^2]) Z_\mu Z^\mu + \dots \quad (2.12)$$

<sup>443</sup> where we show only the mass terms for the gauge bosons. Note that, by construction, there is no mass term for  $A_\mu$  which allows us to identify this field with the photon.  $W_\mu^1$  and  $W_\mu^2$  are related to the  $W^\pm$  boson and  $Z_\mu$  corresponds to the  $Z$  boson. We have shown that it is the fact that  $v \neq 0$  that allows for the existence of non-zero mass terms for the  $W^\pm$  and  $Z$  bosons.

If we now expand the second term of the Higgs potential (Eq. 2.3) using Eq. 2.8 we get, among other terms,

$$\mathcal{L} = -h^3 \sqrt{-\mu^2 \lambda} - h^4 \lambda + \dots . \quad (2.13)$$

These terms encode the Higgs self interactions and represent, respectively, the three and four point interactions. We see that the coupling constants of these interactions depend on the parameters of the Higgs potential,  $\mu^2$  and  $\lambda$ .

In addition to being responsible for giving mass to the gauge bosons, the Higgs field is also responsible for the mass of the fermions. However, the mechanism through which this occurs is fundamentally different. In the case of the leptons the mass terms are placed explicitly in the Lagrangian:

$$\mathcal{L}_{\text{fermions}} = G_1 \bar{L} \phi R + G_2 \bar{L} \phi_c R + \text{hermitian conjugate} \quad (2.14)$$

where  $L$  denotes a left-handed fermion doublet and  $R$  denotes a right-handed fermion singlet. Here, left and right refer to helicity states.  $G_1$  and  $G_2$  are arbitrary coupling constants that can be written in terms of the fermion's mass and the VEV.  $\phi$  is given by Eq. 2.8 and  $\phi_c$  is given by (after the spontaneous symmetry breaking and in the unitary gauge):

$$\phi_c = \begin{bmatrix} v + \frac{h}{\sqrt{2}} \\ 0 \end{bmatrix}. \quad (2.15)$$

We now take a quick detour to motivate why fermions are represented as chiral states (left and right) of the  $SU_L(2)$  symmetry. We base this discussion on [9]. In the context of the unification of the electromagnetic and weak forces, formalized by Weinberg, Glashow and Salam in 1961, both interactions are interpreted as manifestations of the electroweak force. Weak charged currents are axial vector currents which means they couple only to left handed fermions while weak neutral currents, as well as QED, couple to both helicity states. This suggested that fermions were better represented as left-handed doublets and right-handed singlets of the  $SU_L(2)$  symmetry group. The left handed doublets,  $L$ , are defined as:

$$L : \begin{pmatrix} \nu_l \\ l \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L \quad (2.16)$$

where  $l$  represents an electron, muon or tau,  $u$  is any quark of the up type and  $d'$  is a quark of the down type. The right-handed states,  $R$ , are singlets, define as:

$$R : l_R, u_R, d'_R. \quad (2.17)$$

In 1956, C. S. Wu *et al.* showed that the weak interaction violates parity conservation [10]. In 1958, M. Goldhaber *et al.* conducted an experiment that showed that neutrinos are left-handed and anti-neutrinos are right-handed [11] which is why the SM does not include a right-handed state for neutrinos. We can now continue the discussion of the mass generation mechanism for fermions.

The first term in Eq. 2.14 gives mass to down type fermions (electron, muon, tau, down, strange

462 and bottom quarks) and the second to up type fermions (up, charm and top quarks). In addition, these  
 463 terms give rise to the interaction terms between the Higgs field and the fermions. Take, as an example,  
 464  $\bar{L} = (\bar{t}, \bar{b})_L$  and  $R = b_R$ . For the first term of Eq. 2.14 we get:

$$G_1 \bar{L} \phi R = G_1 (\bar{t}, \bar{b})_L \begin{bmatrix} 0 \\ v + \frac{h}{\sqrt{2}} \end{bmatrix} b_R = G_1 v \bar{b}_L b_R + \frac{G_1}{\sqrt{2}} \bar{b}_L b_R h. \quad (2.18)$$

The first term is the mass term for b quarks. Therefore we can redefine  $G_1 v = m_b$  and obtain:

$$G_1 \bar{L} \phi R = m_b \bar{b}_L b_R + \frac{m_b}{v \sqrt{2}} \bar{b}_L b_R h. \quad (2.19)$$

465 The second term gives the interaction between the Higgs boson and the fermions, in this case, the b  
 466 quarks. The strength of this interaction is directly proportional to the mass of the corresponding fermion.

467 In the SM formalism, neutrinos as massless particles. However, there is no reason why they cannot  
 468 acquire mass through a mechanism similar to the one we just described. Nonetheless, the usual argu-  
 469 ment is that it would be unnatural for the same mechanism to produce the mass of very heavy particles,  
 470 such as the top quark, and the mass of very light particle, such as the neutrinos. Therefore, BSM models  
 471 that try to explain the mass generation for neutrinos usually resort to a different mechanism.

472 The SM has delivered extremely accurate predictions about the existence and properties of new  
 473 particles which make it a very successful theory. It predicted the existence of the W and Z bosons [12],  
 474 the gluon, the charm and top quarks and the Higgs boson [6–8]. In addition, the SM prediction for the  
 475 value of the anomalous magnetic dipole moment of the electron (calculated up to order  $\alpha^5$ ) agrees with  
 476 the measured value up to the 11<sup>th</sup> decimal place, making it the most precise measurement in science.

### 477 2.1.1 Higgs pair production

478 Within the SM there are still some processes that have not been measured. One of these is the pro-  
 479 duction of pairs of Higgs bosons. The experimental challenges and efforts related to this process are  
 480 discussed in section 4.1. Here we provide a theoretical description of the process.

481 At the Large Hadron Collider (LHC), the main production process of Higgs pairs is gluon-gluon fusion  
 482 ( $ggF$ ). Higgs pairs can also be produced through vector boson ( $V$ ) fusion (VBF), in association with a  
 483 pair of top quarks ( $t\bar{t}h$ ) or through Higgs strahlung ( $Vh$ ) <sup>4</sup>. At a center of mass (CM) energy of  $\sqrt{s} = 13$   
 484 TeV, the  $ggF$  production process has a cross section approximately seventeen times larger than the next  
 485 most common production process which is VBF (approximately 30 fb *versus* 1.6 fb [13]). Therefore it  
 486 is the dominant contribution when we study inclusive production. For this reason we focus the following  
 487 discussion on this production mode. The leading order Feynman diagrams for Higgs pair production via  
 488  $ggF$  are shown in figure 2.3.

489 The diagram on the right has an off-shell (virtual) Higgs boson,  $h^*$ , that couples to gluons by the usual  
 490 heavy quark triangle (same mechanism as in single Higgs production).  $h^*$  then decays to two on-shell  
 491 Higgs bosons. This diagram contains the three point interaction between Higgs bosons and therefore it

---

<sup>4</sup>In this process, at LO, a Higgs boson is radiated from a vector boson

is the one that allows us to probe this coupling. In the diagram on the left, the two Higgs bosons couple to the gluons by a box of heavy quarks and are directly radiated from a quark. The largest contributions for these quantum loops come from heavy quarks, such as the top and bottom, because the coupling constant of the Higgs boson to fermions is directly proportional to the fermions mass (see section 2.1).

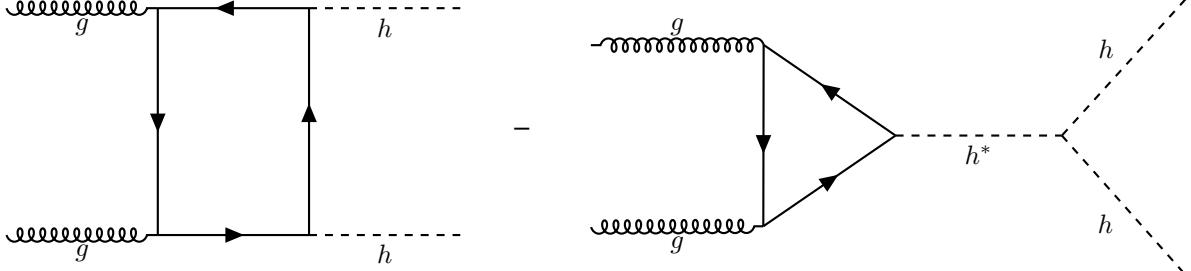


Figure 2.3: Feynman diagrams of Higgs pair production from gluon fusion. Triple vertex diagram (left) and box diagram (right). The minus sign between the diagrams indicates that they interfere destructively.

The amplitudes for the box,  $\mathcal{M}_\square$ , and triangle,  $\mathcal{M}_\Delta$ , diagrams scale as [14]:

$$\mathcal{M}_\square \sim \frac{\alpha_s}{4\pi} y_t^2, \quad \mathcal{M}_\Delta \sim \lambda_{hhh} \frac{\alpha_s}{4\pi} y_t^2 \frac{m_h^2}{\hat{s}} \left( \log \frac{m_t^2}{\hat{s}} + i\pi \right)^2 \quad (2.20)$$

where  $\hat{s}$  is the CM energy,  $y_t$  is the Yukawa coupling of the top quark,  $\alpha_s$  is the electroweak coupling constant,  $\lambda_{hhh}$  is the coupling constant of the Higgs boson three point self-interaction and  $m_h, m_t$  are the masses of the Higgs boson and top quark.

At a CM energy of  $\sqrt{s} = 13$  TeV the cross section for Higgs pair production, as predicted by NLO calculations, is very small, approximately 30 fb [13]. It is suppressed due to the destructive interference between the LO diagrams that leads to a  $\sim 50\%$  suppression of the total cross section [14]. Furthermore, the cross section of the triangle diagram is smaller than the one of the box diagram, approximately 4 fb compared to 30 fb<sup>5</sup>, and it is strongly suppressed for larger values of the CM energy which can be seen directly from the expression of the amplitude in Eq. 2.20. This means that the Higgs trilinear coupling mostly affects the Higgs pair production at threshold. A variable that is sensitive to these effects is the invariant mass of the Higgs pair. The tail of this distribution (high invariant mass of the Higgs pair), however, is mostly determined by the box diagram contribution [14].

The LO calculation for the cross section of Higgs pair production has been performed, for example, in [15]. A value of the order of 10 fb is reported. The NLO calculation is a theoretical challenge: several two-loop diagrams that take into account virtual and real radiation have to be considered. In addition, top quark mass effects can be included in various approximations. This leads to corrections with different signs which suggests that the uncertainty on the cross section due to top quark mass effects are of the order of  $\pm 10\%$  at NLO. Therefore, a calculation including the full top mass dependence was of the utmost importance. This result became available recently [16]:

$$\sigma_{gg \rightarrow hh}^{\text{NLO}} = 27.80^{+13.8\%}_{-12.8\%} (\text{scale}) \pm 0.3\% (\text{stat.}) \pm 0.1\% (\text{int.}) \text{ fb} \quad (2.21)$$

<sup>5</sup>These values are obtained using MadGraph5. They are shown here to give a rough estimate of the difference between the values of the cross sections of both diagrams.

508 where the dependence of the result on the variation of the scales by a factor of two around the central  
 509 scale, the statistical error coming from the limited number of phase space points evaluated and the error  
 510 coming from the numerical integration of the amplitude are shown.

511 This result shows that the introduction of NLO contributions produces a significantly different result.  
 512 Therefore the inclusion of such effects is necessary if we wish to obtain an accurate result that can be  
 513 compared to experimental values. The analytical expressions for the NLO cross section are long and  
 514 complex so we abstain from reproducing them here. Nonetheless, the LO cross section can be written in  
 515 a compact form and it allows us to discuss some key features of the process. Therefore, we will present  
 516 it here, based on [15] and [17].

The partonic LO cross section for  $gg \rightarrow hh$  can be written

$$\hat{\sigma}_{gg \rightarrow hh}^{\text{LO}}(\hat{s}) \sim \int_{\hat{t}_-}^{\hat{t}_+} d\hat{t} \left( |C_\Delta F_\Delta + C_\square F_\square|^2 + |C_\square G_\square|^2 \right) \quad (2.22)$$

where  $\hat{s}$  and  $\hat{t}$  are the Mandelstam variables and, in addition,  $\hat{s}$  can be identified with the square of the partonic CM energy of the process. The integration limits,  $\hat{t}_\pm$ , are derived from a momentum parametrization in the CM frame, leading to  $\hat{t}_\pm = m_h^2 - \frac{\hat{s}}{2}(1 \mp \beta_h)$ , where  $\beta_h^2 = 1 - 4\frac{m_h^2}{\hat{s}}$  and  $m_h$  is the mass of the Higgs boson [17].  $F_\Delta$ ,  $F_\square$  and  $G_\square$  are form factors whose full expressions can be found, for example, in [15].  $C_\Delta$  and  $C_\square$  can be interpreted as generalized couplings and are given by

$$C_\Delta = \lambda_{hhh} \frac{m_Z^2}{\hat{s} - m_h^2}, \quad C_\square = 1, \quad (2.23)$$

where  $m_Z$  is the mass of the  $Z$  boson. If we take the limit  $m_Q^2 \gg \hat{s} \sim m_h^2$  (where  $m_Q$  is the mass of the quarks that contribute to the quantum loops) we can get simple expressions for the remaining form factors:

$$F_\Delta = \frac{2}{3} + \mathcal{O}(\hat{s}/m_Q^2), \quad F_\square = -\frac{2}{3} + \mathcal{O}(\hat{s}/m_Q^2), \quad G_\square = \mathcal{O}(\hat{s}/m_Q^2). \quad (2.24)$$

In this limit, the partonic cross section is simply given by

$$\hat{\sigma}_{gg \rightarrow hh}^{\text{LO}}(\hat{s}) \sim \int_{\hat{t}_-}^{\hat{t}_+} d\hat{t} \left| \lambda_{hhh} \frac{m_Z^2}{\hat{s} - m_h^2} - 1 \right|^2. \quad (2.25)$$

517 There are two important points that are worth discussing. Firstly, the total cross section has terms  
 518 that are proportional to the Higgs triple coupling,  $\lambda_{hhh}$ , which can be read directly from Eq. 2.25. On the  
 519 one hand, this means that measuring this process gives us access to the value of  $\lambda_{hhh}$  and therefore  
 520 provides valuable insight into the shape of the Higgs potential and ultimately into the EWSB mechanism  
 521 in the SM. On the other hand, if  $\lambda_{hhh}$  has a value that is different from the one predicted by the SM, that  
 522 will affect the measured value of the cross section and can lead to hints of new physics.

523 Secondly, although this is not evident from Eq. 2.25, the cross section for di-Higgs production in-  
 524 creases with  $\hat{s}$ . This can be seen in figure 2.4 that shows the variation of the total (integrated) NLO cross  
 525 section with the CM energy for the six largest production channels. Note that increasing the CM energy  
 526 from 13 to 100 TeV increases the inclusive cross section by approximately two orders of magnitude which

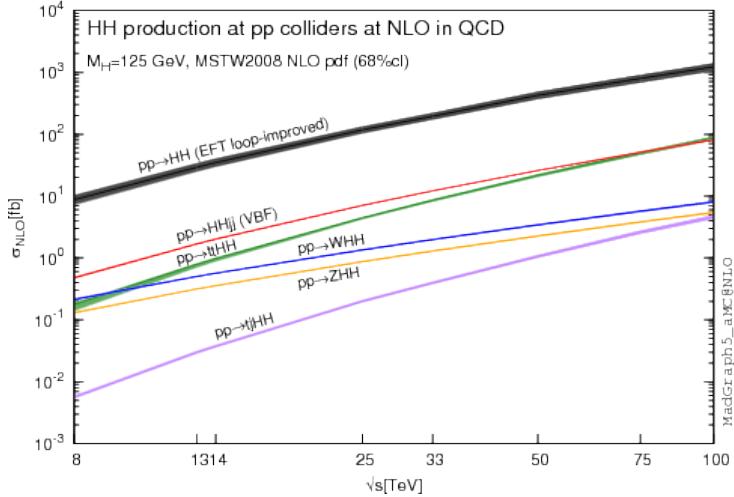


Figure 2.4: Total cross sections at the NLO in QCD for the six largest HH production channels at pp colliders. The thickness of the lines corresponds to the scale and PDF uncertainties added linearly.

527 is a consequence of the increased phase space that becomes available.

528 Therefore, the increase in the cross section of rare processes, such as Higgs pairs production, as  
529 the CM energy of collision experiments increases supports the claim that future colliders, with higher  
530 CM energies, might be our chance of discovering and precisely studying these processes.

## 531 2.2 Going beyond

532 Despite the success of the SM, there is evidence, both experimental and theoretical in nature, that  
533 indicate that it cannot be the final theory of particle physics. This led to the development of alternative  
534 models that extend the SM but that can still reproduce its successful predictions. These are referred to  
535 as Beyond the Standard Model (BSM) models.

536 On the one hand, there are several pieces of experimental evidence that the SM cannot explain.  
537 These include the nature of dark matter, postulated to explain the experimental observations of the  
538 velocity of far away galaxies [18], the asymmetry between matter and anti-matter in the present Universe  
539 <sup>6</sup> and the fact that neutrinos oscillate between flavors which implies that they have a non-zero mass.  
540 This phenomenon was measured independently by two collaborations, the Super-Kamiokande and the  
541 Sudbury Neutrino Observatory (SNO), in 1998 and 2001-2002, respectively, [19–21].

542 On the other hand, its theoretical formulation also has some weaknesses: it accurately describes  
543 particles interactions at the electroweak scale ( $\sim 246$  GeV) but it does not include gravity which means it  
544 cannot be valid at the Planck scale ( $\sim 10^{19}$  GeV) where gravity cannot be overlooked; it has a lot (over 20)  
545 of free parameters whose values have to be tuned to fit experimental observations, and there is a large  
546 discrepancy between the mass scales associated with the electroweak and gravitational interactions  
547 (this is one of the simplest formulations of what is known as the hierarchy problem).

548 When faced with these weaknesses, or rather hints of incompleteness, the theoretical community  
549 put a great effort into the development of models that add new ingredients to the SM but that simultane-

---

<sup>6</sup>Or why do we live in a local in an Universe made mostly out of matter?

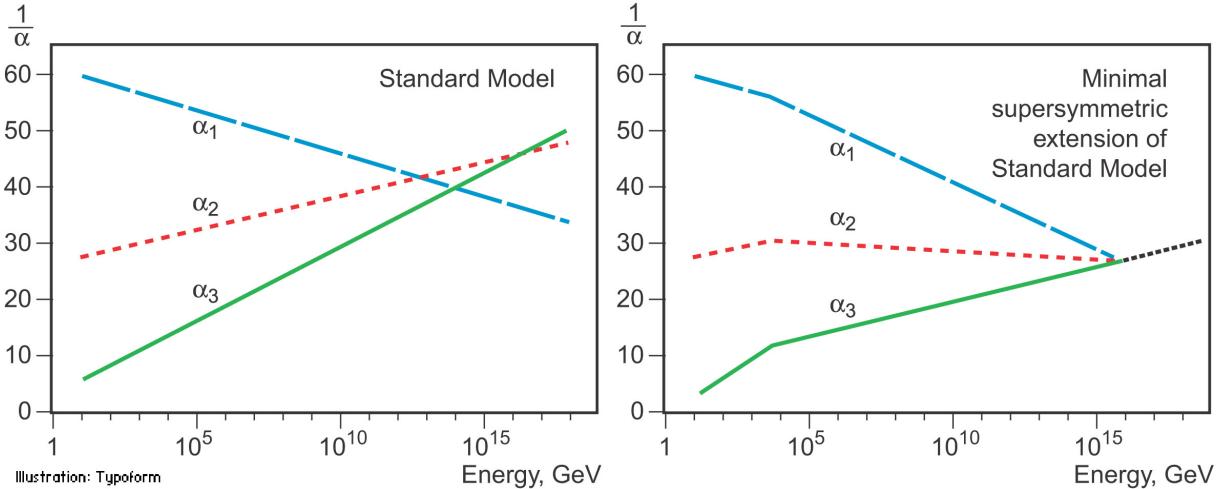


Figure 2.5: Here,  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  represent the coupling constants of  $U_Y(1)$ ,  $SU_L(2)$  and  $SU_{color}(3)$ , respectively. What is shown is the variation of the inverse of the coupling with the energy: on the left for the SM and on the right for the MSSM. Plots from [23].

ous lead to predictions that are compatible with existent experimental measurements. In the following paragraphs we will introduce and briefly describe some of the most well studied (both theoretically and experimentally) BSM models. We follow the discussion presented in [22] as a starting point.

It is a well known consequence of renormalization in QFT that the coupling constants become dependent on the energy scale at which the theory is probed. As the energy scale increases the  $U_Y(1)$  coupling constant gets larger while the  $SU_L(2)$  and  $SU_{color}(3)$  coupling constants get smaller. If one extrapolates far enough these become nearly equal at an energy scale of approximately  $10^{15}$  GeV. Although this matching is far from perfect, it sparked the idea that these three forces could be unified at an energy scale of  $10^{15}$  GeV. Grand Unification Theories (GUT) try to combine  $SU_{color}(3) \times SU_L(2) \times U_Y(1)$  into a larger symmetry group.

Supersymmetric (SUSY) models introduce a new symmetry that links fermions and bosons. For each boson(fermion) of the SM it introduces a fermionic(bosonic) partner. Apart from spin, the supersymmetric partners would share the same mass and quantum numbers. Since we have not found any supersymmetric particles in the LHC this means that supersymmetry is necessarily a broken symmetry and, if they exist, new particles should have a larger mass (outside of the present reach of the LHC) than their SM partners. SUSY models were introduced because they offer a natural fix for the hierarchy problem. In addition, the Minimal Supersymmetric extension of the SM (MSSM) also leads to a better convergence of the coupling constants as can be seen in figure 2.5. From the standpoint of GUT this is extremely appealing.

An early proposal of a theory that could unify gravity with electromagnetism was given by Theodore Kaluza in 1921. In particular, he showed that these two forces could stem from a single tensor with the introduction of an extra space dimension. In 1926, Oscar Klein offered an explanation for this extra dimension; he proposed that it had a circular topology such that at each point of the four dimensional space-time we would have a circle with a small radius. This theory has more degrees of freedom (because it is formulated in a higher dimensional space-time) and therefore it predicts new particles that

575 are usually known as Kaluza-Klein gravitons (and their excited states). Nonetheless, the Kaluza-Klein  
 576 model does not provide a satisfactory explanation for the hierarchy problem. Therefore, in 1999, Lisa  
 577 Randall and Raman Sundrum introduced a new model that does. This model introduces only two new  
 578 particles: a spin 2 graviton (and its Kaluza-Klein excitations) and a radion, that is a spin 0 neutral particle.

579 Models with two Higgs doublets (2HDM) are one of the simplest possible extensions of the Higgs  
 580 sector of the SM. They are appealing because while the fermionic sector is rather complex, having three  
 581 families, the scalar sector is quite simple, having a single particle, which seems unnatural. This type of  
 582 structure is realized in various new physics models including SUSY models. In addition, they provide  
 583 an additional source for CP violation which could help explain the matter-anti-matter asymmetry in the  
 584 Universe.

The most general renormalizable 2HDM scalar potential is written as [24]

$$V = m_{11}^2 |\Phi_1|^2 + m_{22}^2 |\Phi_2|^2 - (m_{12}^2 \Phi_1^\dagger \Phi_2 + h.c.) \\ + \frac{1}{2} \lambda_1 |\Phi_1|^4 + \frac{1}{2} \lambda_2 |\Phi_2|^4 + \lambda_3 |\Phi_1|^2 |\Phi_2|^2 + \lambda_4 |\Phi_1^\dagger \Phi_2|^2 \\ + \left[ \frac{1}{2} \lambda_5 (\Phi_1^\dagger \Phi_2)^2 + \lambda_6 |\Phi_1|^2 (\Phi_1^\dagger \Phi_2)^2 + \lambda_7 |\Phi_2|^2 (\Phi_1^\dagger \Phi_2)^2 + h.c. \right], \quad (2.26)$$

585 where  $\Phi_1$  and  $\Phi_2$  are hypercharge doublets and the coefficients  $m_{12}^2$  and  $\lambda_{5,6,7}$  can be complex. However,  
 586 when including the Yukawa interactions, the most general lagrangian leads to tree-level flavor changing  
 587 neutral currents (FCNC) in the Yukawa sector. These FCNC are very tightly constrained by experimental  
 588 data and should be avoided. They can be eliminated by imposing a  $\mathbb{Z}_2$  symmetry. Although, usually,  
 589 this symmetry is allowed to be softly broken in order to allow the theory to have a decoupling limit [25]  
 590 where the mass of all the scalars other than the SM-like one can be made very large. In addition to  
 591 the softly-broken  $\mathbb{Z}_2$  symmetry, which leads to  $\lambda_{6,7} = 0$ , we also impose CP conservation which makes  
 592 all possible complex phases vanish. In this case, we obtain five Higgs bosons that are CP eigenstates.  
 593 Three of them are neutral,  $h$ ,  $H$  and  $A$ , and the other two are charged,  $H^\pm$ .  $h$  and  $H$  are CP-even states  
 594 while  $A$  is CP-odd.  $h$  is usually taken to be the SM Higgs boson and its mass is set to 125 GeV.

595 There are several types of 2HDM classified according to their fermion-scalar interactions. We high-  
 596 light the type II (the one used in this work), where all right-handed up-type quarks couple to  $\Phi_2$  and  
 597 right-handed down-type quarks and charged leptons couple to  $\Phi_1$ . This type of couplings is analogous  
 598 to what happens in SUSY models.

599 Instead of the parameters in Eq. 2.26, we can describe the model in terms of the four physical  
 600 masses,  $m_h$ ,  $m_H$ ,  $m_A$  and  $m_{H^\pm}$ , the angles  $\alpha$  and  $\beta$ , the VEV  $v = 246$  GeV and a further parameter,  
 601 chosen to be  $m_{12}^2$  [24]. The angle  $\beta$  is defined as  $\tan(\beta) = v_2/v_1$ , where  $v_{1,2}$  are the VEVs of the two  
 602 Higgs doublets. The angle  $\alpha$  diagonalizes the square mass matrix of the CP-even states. The quartic  
 603 couplings of the potential can then be written in terms of these parameters, as can be found in Eq. 11 of  
 604 Ref. [24].

605 Simplified dark matter (DM) models are based on the exchange of a single particle between DM and  
 606 SM particles and try to explain the nature of DM and how it interacts with the SM. The particle exchanged  
 607 is called a mediator and, depending on the specific model, it can be neutral or electrically charged and

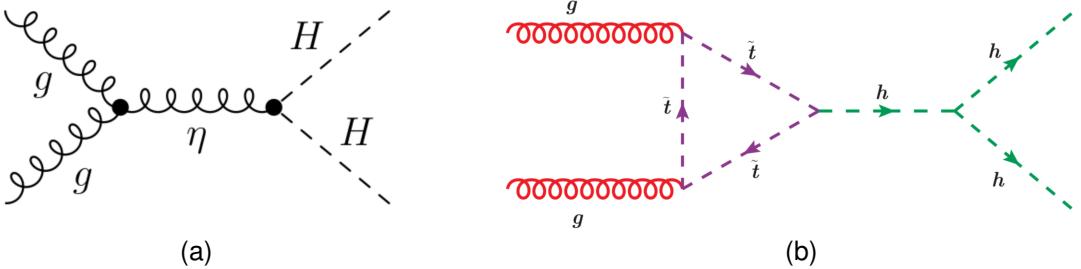


Figure 2.6: BSM contributions to Higgs pair production from the spin-2 Kaluza-Klein graviton (a) and from the SUSY partner of the top quark (b). Figures from Refs. [26, 27].

608 have spin 0, 1 or 2. The simplest possible scenario is a neutral scalar mediator.

609 The CP-conserving 2HDM and dark matter model with a spin 0 mediator are explored in this work  
 610 as sources of alternative di-Higgs production processes. For these models, the main modification with  
 611 respect to the SM occurs because new heavy particles, namely,  $H$  and the DM mediator can couple  
 612 to the Higgs bosons through the s-channel diagram. This corresponds to replacing the off-shell Higgs  
 613 boson by one of these particles in the Feynman diagram on the left in figure 2.3.

614 The spin-2 graviton predicted by Kaluza-Klein and Randal-Sundrum models can couple directly to  
 615 gluons and then decay producing a Higgs pair, as it illustrated in figure 2.6(a). SUSY particles can  
 616 contribute to the quantum loops in the Higgs production Feynmann diagrams. An example is shown in  
 617 figure 2.6(b), where the top quark is replaced by its supersymmetric partner is the s-channel diagram.  
 618 These are two examples of how BSM models can change the Higgs pair production process but there  
 619 are many other. The crucial point is that some BSM contributions can lead to an enhancement of the  
 620 cross section for Higgs pair production with respect to what is predicted by the SM. Experimentally, this  
 621 means that we would not need as much sensitivity and therefore the process could be measured with  
 622 less data and therefore sooner. This is the reason why a lot of the searches performed at the LHC focus  
 623 on this type of scenario. In addition, if a new heavy particle couples to the Higgs boson via an s-channel  
 624 diagram, it could lead to the existence of a peak in the Higgs pair invariant mass spectrum (assuming  
 625 that we have enough experimental resolution and that there are not other processes coming into play).

626 Moreover, BSM models introduce new free parameters (in addition to the SM ones) that can be  
 627 constrained using the experimental results obtained at the LHC (and other experiments). This reduces  
 628 the available parameter space of the models and may even reject some of them.

629 **Chapter 3**

630 **Collider experiments**

631 In this chapter we start by providing an overview of the goals and main challenges of modern collider  
632 experiments. The definitions of some key quantities that describe an accelerator are introduced and a  
633 brief discussion on how they influence the discovery potential of an accelerator is presented.

634 In section 3.2 we introduce the LHC and in section 3.2.1 we describe the ATLAS experiment, includ-  
635 ing brief discussions on b-tagging, trigger and data acquisition algorithms and systems. In section 3.3  
636 we introduce the concept of a hadronic jet and describe how these objects are reconstructed in a general  
637 collider experiment. Jet properties and substructure observables, as well as jet grooming algorithms,  
638 are introduced in sections 3.3.1 and 3.3.2, respectively.

639 In section 3.4 we shift the focus to future collider experiments and accelerators and motivate their  
640 need. In sections 3.4.1 and 3.4.2 we introduce the hadronic Future Circular Collider (FCC-hh) and  
641 describe the baseline detector design, respectively.

642 Collider experiments are the best tool we have to explore matters' most fundamental structure. When  
643 we accelerate a particle we increase its momentum. If we take into account the wave particle duality  
644 and the De Broglie expression,  $\lambda = h/p$ , where  $\lambda$  is the wavelength and  $p$  is the particle's momentum,  
645 we can see that a particle with a large  $p$  will have a small  $\lambda$ . The wavelength gives us the dimension  
646 scale of the objects we can probe with a given wave. If we want to probe very small particles (subatomic  
647 and smaller) we need very small  $\lambda$  and therefore very large  $p$ . Conceptually, this is the basic idea behind  
648 modern particle accelerators.

649 In practice, charged particles can be accelerated and their trajectories controlled by means of electro-  
650 magnetic fields. However, this is not without numerous technical challenges. When a charged particle is  
651 subject to an acceleration perpendicular to its velocity (which is exactly what happens in circular acceler-  
652 ators) it emits electromagnetic radiation, called synchrotron radiation. The power emitted is proportional  
653 to the fourth power of the particle's energy and inversely proportional to the radius squared and to the  
654 fourth power of the particle's mass. This radiation limits the maximum energy that can be achieved in  
655 electron-positron colliders. In proton-proton colliders, however, the energy is limited by the maximum  
656 magnetic field that can be achieved. Therefore, there is also the need for extremely powerful magnets  
657 which are usually implemented using technology based in superconductivity. Using superconducting

658 magnets raises another challenge: they can only operate at very low temperatures, close to the ab-  
659 solute zero. In addition, in order to sustain a stable beam it is necessary that the beam pipe has an  
660 environment very close to absolute vacuum.

### 661 3.1 Experimental aspects

One of the most important parameters of a particle accelerator is the time integrated luminosity,  $\int \mathcal{L}(t)dt$ . For a given process with cross section  $\sigma$ , it determines the number of event that will be produced,  $N$ :

$$N = \sigma \int \mathcal{L}(t)dt, \quad (3.1)$$

where  $\mathcal{L}(t)$  is the instantaneous luminosity that is a measure of the number of collisions per bunch crossing. The instantaneous luminosity is given by:

$$\mathcal{L} = f_{coll} \frac{n_1 n_2}{4\pi \sigma_x \sigma_y}, \quad (3.2)$$

662 where  $f_{coll}$  is the collision frequency,  $n_1$  and  $n_2$  are the number of protons in each bunch and  $\sigma_x$  and  $\sigma_y$   
663 characterize the transverse beam size in the horizontal and vertical directions.

664 To increase the chances of measuring a rare process, or to increase the statistical significance of the  
665 measurement of an already discovered process, we want to increase  $N$  as much as possible. To do so  
666 we can either increase the integrated luminosity or change the CM energy in order to increase the cross  
667 section of the process.

668 While the cross sections of most physics processes increase when the CM energy goes from 13 to  
669 100 TeV, many BSM models predict new processes, or new contributions to existing processes, whose  
670 cross sections increase more rapidly than the SM backgrounds. In addition, by conservation of energy,  
671 a larger CM energy implies that particles with larger mass can be created. Based on Eq. 3.2, we can  
672 tune its parameters to obtain the highest possible luminosity. Nonetheless, because we are dealing with  
673 charged particles, there is a limit on how close the bunches can be and on how many protons we can  
674 pack in a bunch. Moreover, the beam's transverse dimensions cannot be infinitely reduced. A smarter  
675 way to increase the number of collisions that an accelerator can produce is to run for a longer time,  
676 therefore increasing the integration time in Eq. 3.1.

677 In conclusion, the CM energy and the integrated luminosity are two of the main parameters that drive  
678 the discovery potential of an accelerator.

### 679 3.2 The Large Hadron Collider

680 The LHC is the world's largest and most powerful particle accelerator. It is housed by the European  
681 Organization for Nuclear Research (CERN) which focuses on fundamental particle physics with the goal  
682 of probing matter's most elementary structure. Ever since its creation, in 1954, CERN has housed many

683 accelerators and experiments and played a key role in the development of fundamental and applied  
684 science.

685 The LHC consists of a 27-kilometer ring located beneath the Franco-Swiss border, near Geneva.  
686 Most of its running time is dedicated to accelerating protons in opposite directions up to a maximum center  
687 of mass energy of  $\sqrt{s} = 13$  TeV and colliding them at the center of the two general purpose experiments,  
688 ATLAS (which is described in section 3.2.1) and CMS. The LHCb experiment also records data from  
689 proton-proton collisions but it is dedicated to the study of beauty particles. The ALICE experiment is  
690 optimized to study heavy-ion collisions at a CM energy of 2.76 TeV per nucleon.

691 The acceleration of charged particles at the LHC is based on radio frequency (RF) cavities. These  
692 cavities are shaped to sustain a resonant electromagnetic field that oscillates at a frequency of 400 MHz.  
693 During the acceleration stage, charged particles passing through the cavities feel an overall force that  
694 propels them forward. When the LHC is running at full energy, a perfectly timed proton with exactly the  
695 right energy feels a zero net force when passing the cavities. Protons with a slightly different energies  
696 arriving slightly earlier or later are decelerated or accelerated in order to keep the beam sorted in discrete  
697 packages with the same energy. These are called bunches. There are 2808 bunches circulating at the  
698 same time, each containing approximately  $10^{11}$  protons. The bunches are spaced by 25 ns. Furthermore,  
699 the successful operation of the LHC also relies on superconducting magnets made of Niobium-Titanium  
700 filaments chilled to  $-271.3^\circ$  C and on an ultra high vacuum (of the order of  $10^{-10} - 10^{-11}$  mbar) inside  
701 the beam pipes. The magnets are placed along the LHC ring and produce dipole and quadrupole  
702 electromagnetic fields. The dipole magnets create a nominal field of 8.3 T and bend the beam along the  
703 tunnel. The quadrupole magnets focus the beam at the interaction points. The ultra high vacuum greatly  
704 reduces the probability that the beam interacts with any particle. It is crucial to keep a stable beam to  
705 continuously maintain collisions during long runs.

706 One of the main research goals of the LHC was to discover the Higgs boson. This was achieved  
707 in 2012 when ATLAS and CMS reported the discovery of a particle consistent with the boson predicted  
708 by the Higgs mechanism, with a mass of 125 GeV [28],[29]. Ever since, efforts have been directed  
709 to measuring its mass, couplings, spin-parity properties with increasing precision using different decay  
710 channels and production modes.

711 ATLAS and CMS reported the observation (measurement with a significance greater than five sigma)  
712 of the Higgs decaying to  $b\bar{b}$  [30, 31]. The searches targeted the  $VH$  production mode. It offers the  
713 best sensitivity to the  $hb\bar{b}$  Yukawa coupling because requiring a vector boson helps reduce the SM  
714 backgrounds, namely the ones from QCD interactions. The observation of  $h \rightarrow b\bar{b}$  means we have  
715 now observed most of the Higgs boson's decay modes predicted by the SM. CMS reported the first  
716 observation of the Higgs boson decaying to a pair of tau leptons [32]. In addition, the observation of  
717 Higgs boson production in association with a  $t\bar{t}$  pair was very recently reported by both collaborations  
718 [33, 34]. Moreover, precision measurements of the masses of the Higgs [35, 36] and  $W$  [37] bosons and  
719 of the top quark [38, 39] were also performed. So far, no conclusive signs of new physics were seen at  
720 the LHC.

721 Future prospects for the LHC include its upgrade to the High Luminosity-LHC (HL-LHC) after the

722 scheduled long shutdown of 2024-2026. This upgrade will increase the size of the dataset to  $3000 \text{ fb}^{-1}$   
723 over the course of ten years [40]. During the shutdown, the ATLAS detector will be upgraded.

724 In the Higgs sector, the high value of the integrated luminosity will improve the statistical precision of  
725 already measured channels and the discovery potential of rare processes [41].

### 726 3.2.1 The ATLAS detector

727 The ATLAS detector has a cylindrical geometry and a multi layered structure. Its dimensions are 25  
728 meters in height (diameter) and 44 meters in length and it weights approximately 7000 tonnes. In the fol-  
729 lowing paragraphs we describe the detector's layers and their functionalities. A schematic representation  
730 of the detector as well as the appropriate coordinate system can be found in figure 3.1.

731 A combination of cartesian and cylindrical coordinates is used to describe the detector. In both  
732 cases, the origin is defined to coincide with the interaction point. The Cartesian system is right-handed  
733 and the z axis is defined to be the direction of the beam. The x-axis points from the interaction point  
734 to the center of the LHC ring and the y-axis points upwards. The azimuthal angle,  $\phi$ , is measured  
735 around the beam axis and the polar angle,  $\theta$ , from the beam line. The pseudorapidity is defined as  
736  $\eta = -\ln \tan(\theta/2)$ . Another commonly used quantity is the rapidity,  $y$ , defined as a function of a particle's  
737 energy,  $E$ , and longitudinal momentum,  $p_L$ :  $y = \frac{1}{2} \ln \left( \frac{E+p_L}{E-p_L} \right)$ . In the limit where a particle's mass is  
738 negligible with respect to its momentum the pseudorapidity converges to the definition of rapidity. In  
739 addition, the angular distance between two points,  $\Delta R$ , is defined as  $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ , where  
740 rapidity can also be used instead of the pseudorapidity.

741 The detector consists of an inner detector (ID) or tracker, electromagnetic (EM) and hadronic calorime-  
742 ters and a muon spectrometer (MS). The magnet configuration consists of a thin superconducting  
743 solenoid that surrounds the ID cavity and three superconducting toroids (one barrel and two end-caps)  
744 arranged with an eight-fold azimuthal symmetry around the calorimeters.

745 The ID covers the pseudorapidity range  $|\eta| < 2.5$  and it makes up the innermost layer of the detector.  
746 It consists of silicon pixel, silicon micro-strip, and straw tube transition radiation tracking detectors. It is  
747 contained in a solenoid magnet with a central field of 2 T. The tracker provides precision measurements  
748 of the positions and momenta of charged particles. As a charged particle transverses the several layers  
749 of the ID it ionizes the medium creating electrical signals that can be read out. These individual electrical  
750 signals are then combined to reconstruct the trajectory of the particle.

751 Lead/Liquid-Argon (LAr) sampling electromagnetic (EM) calorimeters cover the pseudorapidity range  
752  $|\eta| < 3.2$ . The EM calorimeter has an accordion like structure with layers of showering material (lead)  
753 interleaved with layers of active material (liquid argon). These calorimeters provide measurements of  
754 the energy of electrons and photons. The interaction of these particles with the lead layers induces the  
755 production of an EM shower whose energy is measured in the liquid argon layers. The granularity of the  
756 EM calorimeter strongly depends on the longitudinal layer and on the pseudorapidity region.

757 The hadronic calorimetry in the pseudorapidity range  $|\eta| < 1.7$  is provided by a scintillator-tile  
758 calorimeter (TileCal) which is divided in a central barrel and two smaller end-cap barrels, one on each

Table 3.1: ATLAS tile and liquid argon hadronic calorimeters: summary of the pseudorapidity coverages and transversal segmentation (granularity).

<b>TileCal</b>	<b>Barrel</b>	<b>Extended barrel</b>
Coverage	$ \eta  < 1.0$	$0.8 <  \eta  < 1.7$
Granularity	$0.1 \times 0.1$	$0.1 \times 0.1$
<b>LAr calorimeter</b>	<b>End-cap</b>	<b>Forward</b>
Coverage	$1.5 <  \eta  < 3.2$	$3.2 <  \eta  < 4.9$
Granularity	$0.1 \times 0.1$ for $1.5 <  \eta  < 2.5$ $0.2 \times 0.2$ for $2.5 <  \eta  < 3.2$	$0.2 \times 0.2$

759 side of the central barrel. The active components are scintillator tiles made of polystyrene that are in-  
 760 terleaved with steel plates as the passive material. The scintillation light emitted by the tiles when an  
 761 ionising particle crosses the calorimeter is collected on both ends of the tiles by wavelength-shifting op-  
 762 tical fibers. The light signal emitted is proportional to the particle's energy. The TileCal is composed of  
 763 several cells with transverse segmentation  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ <sup>1</sup> [42].

764 For  $|\eta| > 1.5$  LAr calorimeters extend the pseudorapidity range to  $|\eta| = 4.9$ . The LAr calorimeter is  
 765 divided in end-cap and forward. These cover the pseudorapidity ranges  $1.5 < |\eta| < 3.2$  and  $3.2 < |\eta| <$   
 766 4.9, respectively. The active material is liquid-argon and the absorbers are copper and tungsten for the  
 767 end-cap and forward calorimeters, respectively. In the end-cap LAr calorimeters the segmentation is  
 768  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  for  $1.5 < |\eta| < 2.5$  and  $0.2 \times 0.2$  for  $2.5 < |\eta| < 3.2$ . In the forward LAr calorimeter  
 769 the segmentation is  $\Delta\eta \times \Delta\phi = 0.2 \times 0.2$ .

770 The granularity of the ATLAS hadronic calorimeters is summarized in table 3.1.

771 The hadronic calorimeters provide measurements of the energy of hadrons, jets; these allow the  
 772 estimation of the missing transverse energy ( $E_T^{miss}$ ). Approximately one third of the energy of jets is  
 773 deposited in this layer. In the TileCal, the jet energy resolution is given by  $\sigma/E \sim 50\%/\sqrt{E} + 3\%$  [42],  
 774 where the first term is the stochastic term that derives from sampling fluctuations and follows a Poisson  
 775 distribution and the second term is a constant that depends on the characteristics of the calorimeter. For  
 776 the LAr calorimeter, the jet energy resolution is given by  $\sigma/E \sim 60\%/\sqrt{E} + 2\%$  [43].

777 The MS is the outermost layer of the detector and it is dedicated to detecting muons that travel  
 778 through the previous layers almost without interacting. This layer provides measurements of the muons  
 779 transverse momenta. It is composed of Monitored Drift Tubes (MDT) and Cathode Strip Chambers  
 780 (CSC) that provide high precision measurements of the muons' momentum in the pseudorapidity range  
 781  $|\eta| < 2.7$  and of Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC) dedicated to triggering  
 782 purposes for  $|\eta| < 2.4$ .

---

<sup>1</sup>The TileCal is composed of three longitudinal layers. Only the first two have a segmentation equal to  $0.1 \times 0.1$ . In the third layer the segmentation is  $0.2 \times 0.1$ . However, most of the energy of hadronic showers is deposited in the first layers and therefore this detail is not very relevant for this work.

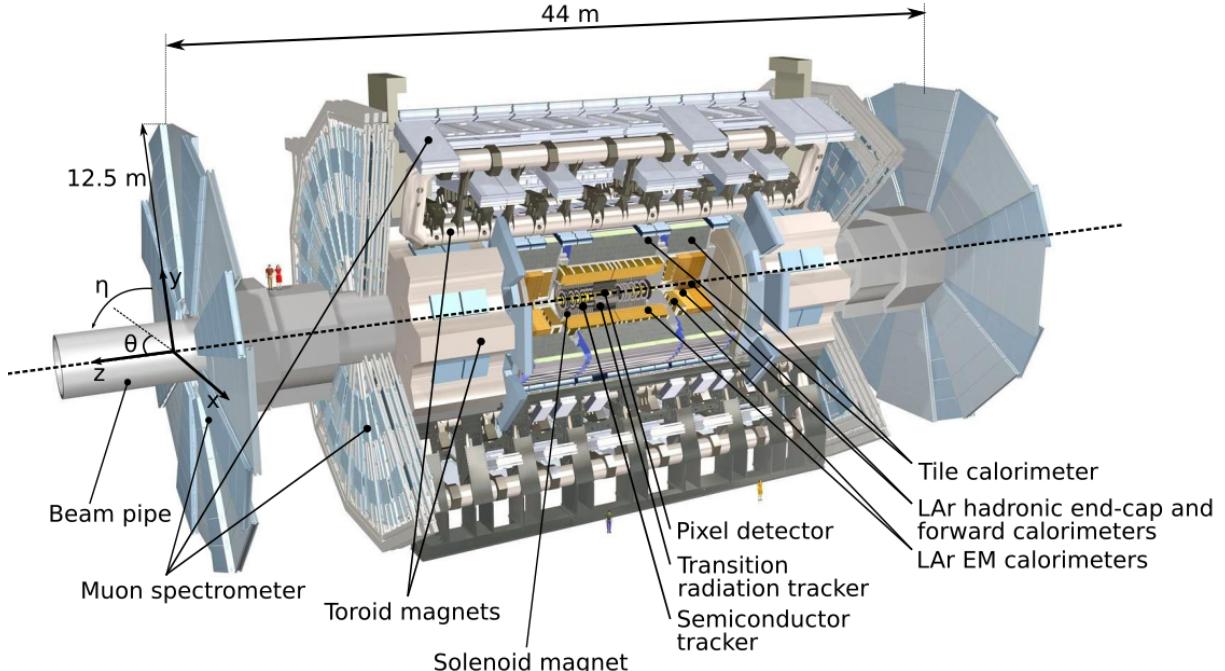


Figure 3.1: ATLAS detector.

### 783 **b-Tagging**

784 Each collision produces a large number of hadronic jets (we refer to section 3.3 for a detailed description  
 785 of jets and how they are reconstructed). For this work, jets initiated by a b quark (b jets) are particularly  
 786 important: we are looking for a Higgs pair decaying to four b quarks which leads to an experimental  
 787 signature that consists of four b jets. b-Tagging algorithms determine, with a given probability, if a jet  
 788 was originated by a b quark.

789 When a b quark is produced it hadronizes almost instantly, producing a B hadron. B hadrons have  
 790 a lifetime of  $\sim 1$  ps and can be highly relativistic meaning that they can travel a few millimeters to a few  
 791 centimeters inside the inner detector before decaying. When they decay there is often a reconstructible  
 792 secondary vertex that is slightly displaced from the primary vertex where the b-quark was produced.  
 793 The existence of a secondary vertex is used by b-tagging algorithms to identify, or tag, a jet as coming  
 794 from a b-quark. It is important to note that a complete b-tagging algorithm relies on the reconstruction of  
 795 a secondary vertex which can only be done using the information from the inner detector. This implies  
 796 that, in ATLAS, we can only b-tag jets that are produced in the region  $|\eta| < 2.5$ .

797 In ATLAS, b-tagging algorithms are applied to the sub-set of tracks that are associated with a given  
 798 jet. The matching between tracks and calorimeter-based jets is performed using the ghost association  
 799 technique [44]<sup>2</sup>. The identification of b-jets in ATLAS is based on distinct strategies encoded on three  
 800 b-tagging algorithms: impact parameter-based algorithms, an inclusive secondary vertex reconstruction  
 801 algorithm and a decay chain multi-vertex reconstruction algorithm. The output of these algorithms are

<sup>2</sup>This procedure works by introducing ghost versions of the measured tracks that have the same direction but infinitesimally small  $p_T$  such that they do not modify the properties of the calorimeter jets. The jets are then reclustered and a track is considered to be associated with a given jet if its ghost version is contained in the jet after reclustering.

802 combined in a multivariate discriminant based on a Boosted Decision Tree (BDT) which provides the  
803 best discrimination between the different jet flavors [45].

804 The impact parameter-based algorithms [46], IP2D and IP3D, use as discriminant variables the trans-  
805 verse impact parameter significance and the transverse and longitudinal impact parameter significance,  
806 respectively. The secondary vertex finding algorithm [46], SV, explicitly reconstructs a displaced sec-  
807 ondary vertex inside the jet by trying to find pairs of tracks with a common origin. The decay chain  
808 multi-vertex reconstruction algorithm [47], JetFitter, tries to reconstruct the full b-hadron decay chain.  
809 This approach allows to resolve b- and c-hadrons vertices even if there are not two tracks associated  
810 with them.

### 811 **Trigger and data acquisition**

812 The LHC delivers approximately 1000 million proton-proton collisions per second, which corresponds to  
813 an event rate of 1 GHz. On the one hand, only a small fraction of these events result in interesting physics  
814 processes. On the other hand, the detector does not have enough storage or read out capabilities to  
815 record all the collisions. The triggering and data acquisition systems are responsible for selecting a  
816 manageable rate of events for permanent storage and further analysis.

817 The trigger is responsible for selecting events with interesting experimental signatures. The trigger  
818 system in Run-2 consists of a hardware-based first level trigger (Level-1) and a software-based high  
819 level trigger (HLT). The Level-1 trigger takes as input coarse granularity calorimeter and muon detector  
820 information and reduces the event rate to 100 kHz. The HLT uses full granularity detector information  
821 and reduces the rate to approximately 1 kHz [48].

## 822 **3.3 Jet reconstruction**

823 A jet is a collimated spray of hadrons that is interpreted as coming from a single initial parton such  
824 that approximately retains information about its physical properties, namely 3-momentum, mass and  
825 charge. The existence of such objects is a direct consequence of the confinement property of QCD.  
826 Quarks and gluons, the fundamental degrees of freedom of QCD, are not asymptotically free. They are  
827 confined inside hadrons. Therefore, when one of these particles is produced it undergoes showering  
828 and hadronization processes that lead to the formation of hadrons.

829 At a particle detector we are interested in reconstructing jets. These are the objects that are used in  
830 the physics analysis. Working with jets instead of hadrons is an advantage because it greatly reduces the  
831 number of objects we need to analyze per event. In addition, they work as a *proxy* for the fundamental  
832 partons produced in the event.

833 Jets are obtained through jet finding algorithms. These are clustering algorithms that group together  
834 experimental quantities (energy deposits in the calorimeters, tracks or particle flow objects) using a se-  
835 quential recombination scheme. If the jets are reconstructed using only the information from the energy  
836 clusters in the calorimeter they are called calorimeter or tower jets. Jets can also be reconstructed using

837 particle flow objects that combine information from the tracking system and from the calorimeters. These  
838 jets are referred to as particle flow or energy flow jets (eflow in short).

839 A crucial property of jet finding algorithms is that they should be infrared and collinear safe, mean-  
840 ing that the outcome of the algorithm (namely the number of jets and their properties) should not be  
841 significantly modified by the emission of soft radiation or by a collinear splitting.

842 The most widely used jet finding algorithms are the  $k_T$  [49], the Cambridge/Aachen (C/A) [50] and  
843 the anti- $k_T$  [51]. They are based on measurements of the quantity:

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \quad d_{iB} = k_{Ti}^{2p}, \quad (3.3)$$

844 referred to as a distance between entities  $i$  and  $j$  with  $d_{iB}$  being the distance between entity  $i$  and the  
845 beam ( $B$ ).  $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$  and  $k_{Ti}$ ,  $y_i$  and  $\phi_i$  are the transverse momentum, rapidity and  
846 azimuth of entity  $i$ . Here,  $k_T$  stands for the transverse momentum. For  $p = 1, 0, -1$  we get the  $k_T$ , C/A  
847 and anti- $k_T$  algorithms, respectively.

848 The clustering algorithm starts by computing all  $d_{ij}$  and  $d_{iB}$  distances. If the smallest distance is a  
849  $d_{ij}$ , the four momenta of particle  $i$  and  $j$  are summed and the distances are updated. If the smallest  
850 distance is a  $d_{iB}$ , particle  $i$  is removed and called a jet. This procedure is iterated until all particles are  
851 clustered in jets.

852 It is worth discussing further the anti- $k_T$  algorithm since it is the default algorithm used in most  
853 analyses, including this one. In this algorithm, the  $\Delta_{ij}$  distance between constituents  $i$  and  $j$  is weighted  
854 by the inverse of the transverse momentum of the constituent with a largest  $k_T$ . This feature implies that  
855 particles with larger momenta will have a smaller  $d_{ij}$  distance and therefore will be clustered first. This  
856 prevents soft particles from being clustered among themselves before clustering the hardest particles.

857 In ATLAS, the transverse and longitudinal segmentation of the calorimeters allow for a three dimen-  
858 sional reconstruction of particle showers which is based in a topological clustering algorithm. Topo-  
859 clusters of calorimeters cells are seeded by cells whose absolute energy exceeds the electronic and  
860 pile-up noise by four standard deviations. The topo-clusters are then expanded by adding all adjacent  
861 cells with absolute energy two standard deviations above noise. Finally, all cells neighbouring the pre-  
862 vious set are also added. After energy calibration, the topo-clusters are fed as input to a jet finding  
863 algorithm.

## 864 **Boosted kinematic regime**

865 With the increase of the CM energy of particle colliders the production of particles with a transverse  
866 momentum much larger than their mass became a reality. In this kinematic regime (referred to as boosted  
867 due to the high Lorentz boost of the particles) traditional jet reconstruction algorithms, that rely on a one-  
868 to-one correspondence between jets and partons, begin to fail [52].

Due to the high Lorentz boosts, decay products of heavy resonances get more collimated. The

angular separation of the decay products is approximately [53]:

$$\Delta R \sim \frac{2m}{p_T} \quad (3.4)$$

where  $p_T$  and  $m$  are the transverse momentum and mass of the decaying particle. In addition to decaying particles with a larger  $p_T$ , another event topology that can produce highly collimated particles is the decay of a particle with a very large mass. This scenario is of particular interest for new physics searches because BSM models often predict the existence of heavy particles.

Take, as an example, the decay of a Higgs boson to two b quarks. Considering that the  $p_T$  of the Higgs is approximately 200 GeV (which is a reasonable and commonly used value in boosted Higgs bosons searches) we get  $\Delta R \sim 1$ . For resolved jets, the default jet radius parameter used in ATLAS is 0.4. We see that for a  $p_T \sim 200$  GeV the angular separation between the decay products of a Higgs boson is already similar to the default jet diameter, which can jeopardize the ability to resolve the individual decay products.

A possible workaround is to use a single jet with a larger  $R$  parameter to reconstruct both decay products. The problem with doing so is that we no longer have information about each individual decay product which means we are loosing some information about the event. We cannot, for example, compute angular variables between the decay products. This led to the development of techniques and observables that allow for the exploration of the intrinsic structure (or substructure) of these large- $R$  jets. Some of these techniques are introduced and discussed in the following section.

### 3.3.1 Jet properties and substructure observables

In general, jet substructure variables aim to quantify the existence of energy clusters insider a jet. Each cluster is interpreted as corresponding to an individual jet. These are called subjets because they are contained inside the large- $R$  jet that was reconstructed. Once they are identified, the subjets can be handled and used for the analysis like normal jets. However, a lot of the substructure techniques do not focus on reconstructing the subjets but rather on determining whether or not they exist, how many they are and how are they distributed inside the large- $R$  jet.

Heavy resonances decaying to two(three) particles will produce large- $R$  jets that are consistent with the existence of two(three) energy clusters. Examples of such topologies are the Higgs boson and top quark decays: the Higgs boson always decays to pairs of particles (leptons, quarks or bosons) and the top quark decays, with a probability close to 100%, to a  $b$  quark and a  $W$  boson that then decays to a pair of leptons or quarks. These topologies are usually referred to as two or three prong. In contrast, jets initiated by a gluon or quark splitting are not expected to have a meaningful substructure. The energy is expected to be concentrated around the jet axis following an isotropic distribution and to become less dense as we approach the jet's border. This is the signature of a one-prong topology. The previous discussion is valid at LO and captures the generic features of jets that are targeted by jet substructure techniques. Nonetheless, other effects may come into play. Take, for example, a highly virtual gluon with a high  $p_T$ . If it splits into two quarks the resulting jet may have two subjets and thus mimic the topology

903 of a heavy resonance decay.

The N-subjetiness variable [54],  $\tau_N$ , may be used to identify jets compatible with N subjets. It is given by

$$\tau_N = \frac{1}{d_0} \sum_k p_T^k \min(\Delta R_1^k, \dots, \Delta R_N^k), \quad d_0 = \sum_k p_T^k R_0 \quad (3.5)$$

904 where the index  $k$  runs over all particles in a jet, the indexes 1 to  $N$  identify the number of axis inside the  
905 jet and  $R_0$  is the radius of the jet. This variable will have a small value if the particles with the highest  $p_T$   
906 (the ones we are most interested in) are clustered around the axis (because  $\Delta R$  will be small) and will  
907 have a larger value otherwise. A jet with small  $\tau_N$  is considered to be consistent with having  $N$  or fewer  
908 subjets because all its constituents are aligned with the axis.

909 We usually use ratios of  $\tau_N$  variables ( $\tau_{MN} = \tau_M / \tau_N$ ). Of particular interest for this work are the  $\tau_{21}$ ,  
910  $\tau_{31}$  and  $\tau_{32}$  ratios. These observables can take values between zero and one. For  $\tau_{21}$ , for example, a  
911 small value (close to zero) indicates that the jet is more compatible with two subjets than with one and  
912 therefore it can help discriminate between two-prong and one-prong jets. The same applies to every  
913 other ratio.

The Fox-Wolfram moments,  $H_l$ , can be used to identify jets that have a structure of two back-to-back subjets in their rest frames [55]. They are given by:

$$H_l = \sum_{i,j} \frac{|\vec{p}_i| |\vec{p}_j|}{E^2} P_l(\cos(\theta_{ij})) \quad (3.6)$$

914 where  $\theta_{ij}$  is the opening angle between energy clusters  $i$  and  $j$  with respect to the interaction point,  $E$   
915 is the total energy of the clusters in the jet rest frame and  $P_l(x)$  are the Legendre polynomials. For a jet  
916 that has a structure of two back-to-back subjets in its rest frame,  $H_1 = 0$ ,  $H_l \sim 1$  for even  $l$ , and  $H_l \sim 0$   
917 for odd  $l$ . This is what we expect for jets coming from the decay of a boosted resonance. In this work we  
918 make use of  $H_2$ .

In addition to jet substructure observables, more standard variables, from which we highlight the jet mass, can also be used to discriminate between jets coming from QCD background and jets resulting from heavy resonance decays. In the latter, the jet's invariant mass should roughly correspond to the mass of the resonance. The invariant mass of a jet,  $M$ , is calculated from the energies and momenta of its constituents as follows:

$$M^2 = \left( \sum_i E_i \right)^2 - \left( \sum_i \vec{p}_i \right)^2 \quad (3.7)$$

919 where  $E_i$  and  $\vec{p}_i$  are the energy and three-momentum of the  $i^{th}$  constituent. However, the mass resolution  
920 is not expected to be very good for large-R jets because of all the extra QCD radiation that may be  
921 caught inside the jet. A possible workaround is the use of jet grooming algorithms which we describe in  
922 the following section.

### 923 3.3.2 Jet grooming algorithms

924 The main goal of jet grooming algorithms is to remove contamination of softer jet constituents from pileup  
925 or underlying event and to leave behind the hard substructure. The main advantage of such algorithms is  
926 that they improve the mass resolution of jets. These features are of particular interest in high luminosity  
927 environments such as the HL-LHC and future high energy colliders.

928 There are three main jet grooming algorithms: trimming, pruning and mass drop filtering. In this  
929 work we do not use pruning or trimming, although these techniques might be worth exploring in future,  
930 more comprehensive, studies. In particular, they can be useful to help reject pileup contributions. The  
931 mass drop filtering procedure isolates relatively symmetric subjets within a jet, each with a significantly  
932 smaller mass than the original jet [53]. This technique was developed and optimized using C/A jets for  
933 the search of Higgs decaying to  $b\bar{b}$  pairs [56]. It works as follows:

- The last step of the C/A clustering is undone such that the jet is split in two subjets,  $j_1$  and  $j_2$  with  $m_{j_1} > m_{j_2}$ . We require that there is a significant difference between the mass of the original jet,  $m_{\text{jet}}$ , and  $m_{j_1}$ :  $m_{j_1}/m_{\text{jet}} < \mu_{\text{frac}}$ , where  $\mu_{\text{frac}}$  is a parameter of the algorithm. In addition the splitting is required to be relatively symmetric:

$$\frac{\min[(p_T^{j_1})^2, (p_T^{j_2})^2]}{(m_{\text{jet}})^2} \times \Delta R_{j_1, j_2}^2 > y_{\text{cut}} \quad (3.8)$$

934 where  $y_{\text{cut}}$  is a parameter that defines the energy sharing between the subjets. It is usually taken  
935 to be  $\sim 0.09$ . If these two criteria are not met the jet is discarded.

- The subjets are clustered with the C/A algorithm with radius parameter  $R_{\text{filt}} = \min[0.3, \Delta R_{j_1, j_2}/2]$ . All jet's constituents that are outside of the three hardest subjets are discarded and we obtained the filtered jet and its subjets.

939 The subjets that are identified within a jet are interpreted as corresponding to the decay products of  
940 the particle that produced the original jet. There are usually only two decay products. A third subjet is  
941 allowed in order to account for extra QCD radiation.

## 942 3.4 Future Colliders

943 As we already argued in the beginning of this chapter, a larger CM energy is one of the factors driving  
944 the discovery potential of an accelerator. As far as we know today, proton-proton colliders are the main,  
945 and possibly only, man-made experimental tool available to explore particle physics in the energy range  
946 on tens of TeV, in a controlled way. With this in mind, new hadronic colliders with CM energies of the  
947 order of tens of TeV have been proposed. The main projects are the hadronic Future Circular Collider  
948 (FCC-hh) lead by CERN and the Super Proton-Proton Collider (SPPC) proposed by China.

949 In this work we focus on FCC-hh and use the established detector's baseline design as our starting  
950 point for this study. In the following section we describe in detail the FCC-hh accelerator and detector.

Table 3.2: Comparison between the working parameters of the LHC and of the FCC-hh. The values of the number bunches and of the number of events per bunch crossing are given assuming a bunch spacing of 25 ns.

Parameter	LHC	FCC-hh
Circumference [km]	27	100
CM energy [TeV]	13	100
Luminosity (peak) [ $10^{34} \text{cm}^{-2}\text{s}^{-1}$ ]	1	30
Dipole field [T]	8.33	16
Nb. of bunches	2808	10600
Nb. of events per bunch crossing	27	1026

### 951 3.4.1 The hadronic Future Circular Collider

952 CERN's FCC study group was launched as a result of a recommendation made in the 2013 update  
 953 of the European Strategy for Particle Physics that '*Europe needs to be in a position to propose an*  
 954 *ambitious post-LHC accelerator project at CERN by the time of the next Strategy update*'. It investigates  
 955 the technological challenges and physics opportunities of a future circular collider. The design and  
 956 infrastructure are driven by a proton-proton collider (FCC-hh) requirements. Electron-positron (FCC-ee)  
 957 and electron-proton (FCC-eh) colliders are also being analyzed. The main goal of this effort is to deliver  
 958 a Conceptual Design Report (CDR) by the end of 2018. This document will include a first cost estimate  
 959 to be submitted for the next update of the European Strategy for Particle Physics, foreseen by 2020.

960 The FCC-hh baseline design consists of a proton-proton circular collider with a maximum CM energy  
 961 of 100 TeV housed by a 100 km tunnel in the area of Geneva. This machine will extend the research  
 962 program of the LHC (and of the HL-LHC) after these have reached their full discovery potential, by  
 963 around 2040. In addition, it will allow for the exploration of an entirely new kinematic regime, probing  
 964 energy scales where new physics may come into play. A possible way of defining the target luminosity of  
 965 this machine is to require that within the first year of operation it surpasses the exploration potential of the  
 966 LHC [57]. Comprehensive studies [57, 58] indicate that this can be achieved with an integrated luminosity  
 967 of the order of  $10 \text{ ab}^{-1}$  per experiment. Considering a reasonable operation period of 10 years this  
 968 leads to integrated luminosity per experiment of the order of  $1 \text{ ab}^{-1}$  per year. In this work, we consider  
 969 a slightly more optimistic scenario that sets the target integrated luminosity at  $30 \text{ ab}^{-1}$  [76].

970 The FCC-hh is expected to work with a dipole field of 16 T and to provide a peak instantaneous lumi-  
 971 nosity thirty times larger than the LHC. The number of bunches is expected to be almost a factor of four  
 972 larger than for the LHC and the number of events per bunch crossing is expected to be approximately  
 973 1000. The latter brings a lot of technical challenges because it means that the mean pileup expected  
 974 for the FCC-hh is almost 40 times larger than for the LHC. This requires the development (or improve-  
 975 ment of already existing techniques) of techniques and algorithms that allow us to further reject pileup  
 976 contributions.

977 Some relevant parameters of the LHC and FCC-hh are summarized in table 3.2.

978 **3.4.2 FCC-hh baseline detector**

979 The design of the FCC-hh baseline detector, which we describe in detail in this section, has been greatly  
980 based on that of the ATLAS and CMS experiments, in particular the central barrel. The layers and sub  
981 detectors are arranged in the same order and perform very similar roles. The geometry is cylindrical  
982 and therefore we can use exactly the same coordinate system that was introduced in section 3.2.1. The  
983 dimensions are very close to the ones of ATLAS: 25 meters in height (diameter) and 48 m in length [59].  
984 A schematic representation of the FCC detector is shown in figure 3.2.

985 The detector consists of trackers, EM and hadronic calorimeters and MS. The magnet configuration  
986 consists of three solenoid magnets (one central-barrel and two forward) that surround the central bar-  
987 rel calorimeters and the forward trackers. The center solenoid delivers a magnetic field of 4 T at the  
988 interaction point [60].

989 The tracker covers the pseudorapidity range  $|\eta| < 6$  and is divided in three sub systems: inner,  
990 outer and forward. The inner and outer trackers and the forward tracker are expected to cover the  
991 pseudorapidity ranges  $|\eta| < 2.5$  and  $2.5 < |\eta| < 6.0$ , respectively. The inner tracker will be instrumented  
992 with pixel detectors while the outer and forward tracker will have layers of both pixel and strip detectors.

993 The EM calorimeter [61, 62] covers the pseudorapidity range  $|\eta| < 6$ . It is divided in barrel, end-cap  
994 and forward. These cover the pseudorapidity ranges  $|\eta| < 1.5$ ,  $1.4 < |\eta| < 2.5$  and  $2.3 < |\eta| < 6$ ,  
995 respectively. The proposed layout for the EM calorimeter is a LAr sampling configuration with lead, glue  
996 and steal plates as absorbers. The granularity is expected to be two to four times better than for the  
997 ATLAS ECAL. For the barrel calorimeter, the goal energy resolution is  $10\%/\sqrt{E} \oplus 1\%$ .

998 The hadronic calorimeter [63] covers the pseudorapidity range  $|\eta| < 6$ . It is also divided in barrel,  
999 end-cap and forward that cover the pseudorapidity ranges  $|\eta| < 1.3$ ,  $1.0 < |\eta| < 1.8$  and  $2.3 < |\eta| < 6.0$ ,  
1000 respectively. The proposed layout for the barrel calorimeter consists of scintillator tiles interleaved with  
1001 lead and stainless steel plates as absorbers. The end-cap and forward calorimeters are expected to be  
1002 based on liquid argon with copper plates as absorbers. For the barrel and end-cap calorimeters, the  
1003 expected segmentation  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  while for the forward calorimeter it is  $\Delta\eta \times \Delta\phi = 0.05 \times$   
1004 0.05. Overall, this corresponds to approximately four times the ATLAS HCAL granularity. The energy  
1005 resolution is expected to be  $40\%/\sqrt{E} \oplus 2.5\%$ ,  $50\%/\sqrt{E} \oplus 3\%$  and  $100\%/\sqrt{E} \oplus 5\%$ , for the barrel, end-  
1006 cap and forward calorimeters, respectively. The pseudorapidity coverage, layout, granularity and energy  
1007 resolution of the hadronic calorimeters are summarized in table 3.3.

1008 The muon spectrometer is divided in barrel, end-cap and forward regions that cover the pseudora-  
1009 pidity ranges  $|\eta| < 1.0$ ,  $1.0 < |\eta| < 2.5$  and  $2.5 < |\eta| < 6.0$ , respectively. The muon's system layout is a  
1010 layered structure of gas chambers.

1011 **3.4.3 FCC-hh physics program**

1012 The FCC-hh physics program is vast and diverse. Within the SM, it includes the study of gauge bosons  
1013 pair production and heavy flavor production, the measurement of the top quark's properties and the  
1014 the study of the EWSB mechanism via multi-Higgs production. In addition, heavy ions collisions will

Table 3.3: Hadronic calorimeter layout, granularity and energy resolution.

Parameter	Barrel	End-cap	Forward
$\eta$ coverage	$ \eta  < 1.3$	$1.0 <  \eta  < 1.8$	$2.3 <  \eta  < 6.0$
Layout	Sci-Pb-Steel (1 : 1.3 : 3.3)	LAr-Cu (1 : 5)	LAr-Cu (1 : 200)
Granularity ( $\Delta\eta \times \Delta\phi$ )	$0.025 \times 0.025$	$0.025 \times 0.025$	$0.05 \times 0.05$
Energy resolution ( $\sigma_E/E$ )	$40\%/\sqrt{E} \oplus 2.5\%$	$50\%/\sqrt{E} \oplus 3\%$	$100\%/\sqrt{E} \oplus 5\%$

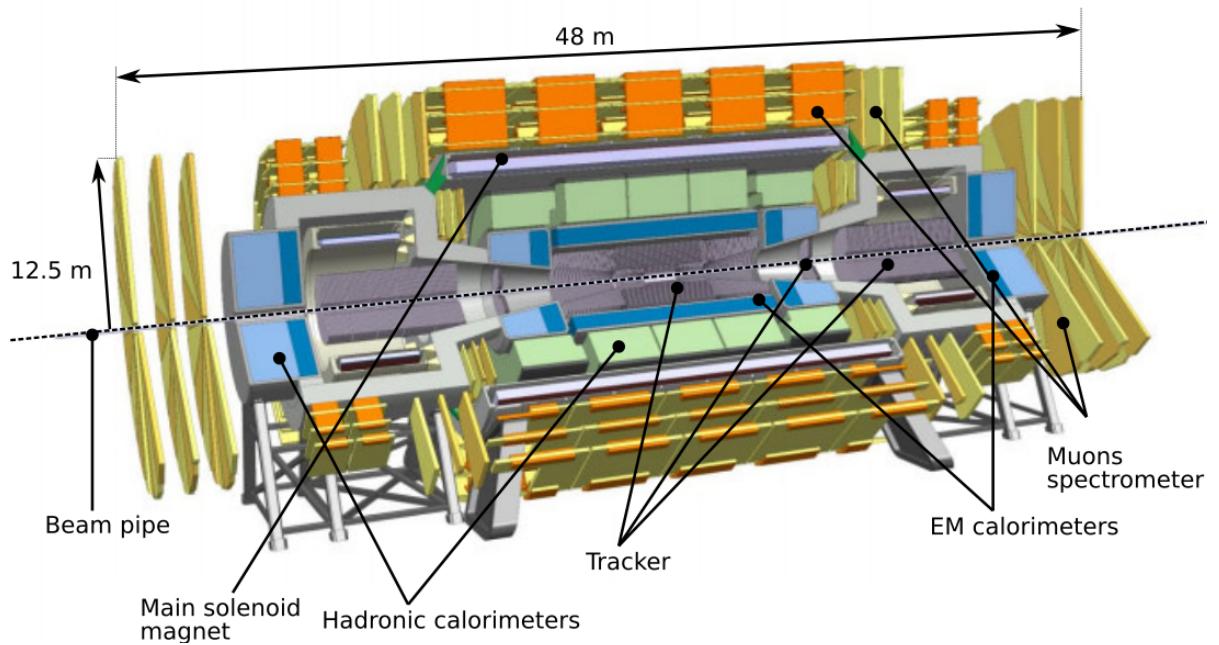


Figure 3.2: FCC detector baseline concept.

allow for a deeper understanding of the Quark-Gluon Plasma. From a BSM standpoint, searches for supersymmetric and dark matter particles in new kinematic regimes can be pursued. A comprehensive review of the physics potential of the FCC-hh can be found here [14]. This document collects the results of the many studies that have been carried out since the beginning of the FCC initiative, in 2014.

Regarding Higgs pair production, a maximum precision on the SM cross section of 3% is expected to be achieved using the  $b\bar{b}\gamma\gamma$  final state. This would allow to constraint the Higgs triple coupling to be  $\lambda_{hhh} \in [0.97, 1.03]$ . The  $hh \rightarrow b\bar{b}b\bar{b}$  would allow for a 5% precision on the SM cross section and to constraint the triple coupling to be  $\lambda_{hhh} \in [0.9, 1.5]$ . In spite of the larger background yield, the  $b\bar{b}b\bar{b}$  channel provides a reasonable number of events in the tail of the  $m_{hh}$  distribution. As discussed in section 2.1.1, the tail of this distribution does not have a large contribution from the triangle diagram. Therefore, the sensitivity to the Higgs triple coupling is expected to be smaller. However, the high energy regime can be more sensitive to new physics contributions which makes the  $b\bar{b}b\bar{b}$  channel a very interesting one.

The Higgs quartic coupling could be probed through triple Higgs production. In this case the most promising final state seems to be  $b\bar{b}b\bar{b}\gamma\gamma$ . This channel could constraint the Higgs quartic coupling to be  $\lambda_{hhhh} \in [-4, 16]$ .

# 1031 Chapter 4

## 1032 State of the art

1033 In this chapter we present the state of the art of searches for Higgs pairs production at the LHC and  
1034 feasibility studies targeting searches for this process at future colliders. In addition, previous studies  
1035 focusing on the granularity of the hadronic calorimeter as a key detector parameter are also reviewed.

1036 In section 4.1, we review the searches that have been conducted at the LHC by the ATLAS and CMS  
1037 experiments. We include discussions on the different final states that were targeted and report on the  
1038 constraints that were derived for the di-Higgs cross section and trilinear coupling. A brief overview of  
1039 the current constraints on some BSM models is also presented. In section 4.2, we present the results  
1040 obtained from feasibility studies that access the discovery potential for this process at the HL-LHC and  
1041 at the FCC-hh. In section 4.3, we present previous studies on the impact of the granularity of the  
1042 hadronic calorimeter on the spatial resolution of hadrons and on the resolution of the jet mass and of jet  
1043 substructure observables.

### 1044 4.1 Searches for Higgs pair production at the LHC

1045 The discovery of the Higgs boson is a strong evidence that the Higgs mechanism operates in Nature.  
1046 However, by itself, it does not guarantee that the shape of the Higgs potential is the one depicted in  
1047 figure 2.2, on the right. In order to reconstruct the Higgs potential and gain a deeper understanding of  
1048 the mechanism that leads to the breaking of the electroweak symmetry one must measure the Higgs  
1049 boson self-couplings, namely its three and four point interactions, whose strengths depend on the values  
1050 of the parameters of the Higgs potential, as shown in section 2.13.

1051 However, in the SM, the cross section for the production of Higgs pairs through ggF is extremely  
1052 small:  $\sim 30$  fb at the current CM energy achieved at the LHC (value computed at Next to Leading  
1053 Order (NLO) accuracy) [64]. This value is approximately three orders of magnitude smaller than the  
1054 production cross section of a single Higgs boson. In addition, this value has to be multiplied by the  
1055 branching fraction of the chosen decay channel which further reduces the effective cross section of the  
1056 full process. Nonetheless, ATLAS and CMS have conducted searches for this process whose results  
1057 are summarized in this section.

1058 The searches performed so far covered different decay channels and targeted not only the SM pro-  
 1059 cess but also some BSM scenarios where di-Higgs production is enhanced. Neither could achieve  
 1060 enough statistical significance to declare the measurement of this process in the SM nor have found any  
 1061 significant deviation from the expected values. These searches resulted in upper limits for the cross sec-  
 1062 tion of di-higgs production in the SM and for the values of the parameters of BSM benchmark theories.  
 1063 From the limits on the cross section it is also possible to constraint the values of the Higgs self coupling,  
 1064  $k_\lambda = \lambda_{hhh}/\lambda_{hhh}^{SM}$ .

1065 The  $hh \rightarrow b\bar{b}b\bar{b}$  channel [65, 66] benefits from the large branching fraction of  $h \rightarrow b\bar{b}$  ( $\sim 58\%$ ). In  
 1066 addition, ATLAS showed that this is the most sensitive channel to resonance masses over 500 GeV [67].  
 1067 However, this channel suffers with an overwhelming multijet background which drives the need for very  
 1068 tight trigger level cuts in order to bring the event rates down to manageable values.

1069 The  $hh \rightarrow b\bar{b}\tau\tau$  analysis [68, 69] benefits from a sizable branching fraction of  $h \rightarrow \tau^+\tau^-$  ( $\sim 7.3\%$ )  
 1070 and from a relatively small background contribution from other SM processes. These searches target  
 1071 the semi-leptonic decay of the  $\tau\tau$  pair to reduce contamination from QCD processes.

1072 The  $hh \rightarrow b\bar{b}\gamma\gamma$  [70, 71],  $WW^*\gamma\gamma$  [67, 72],  $ZZ^*\gamma\gamma$  [72] analysis can make use of very efficient dipho-  
 1073 ton triggers and isolation criteria that greatly reduce multijet background. In addition the excellent mass  
 1074 resolution of  $h \rightarrow \gamma\gamma$  can be exploited. The  $hh \rightarrow b\bar{b}WW^*$  channel also benefits from the large branching  
 1075 fraction of  $h \rightarrow W^+W^-$  ( $\sim 21\%$ ).

1076 The most stringent upper limit on the cross section of Higgs pair production comes from a combina-  
 1077 tion of three searches using up to  $36.1 \text{ fb}^{-1}$  of proton-proton collision data at a CM energy of  $\sqrt{s} = 13$   
 1078 TeV collected with the ATLAS detector [73]. The combination is based on the three most sensitive decay  
 1079 channels of Higgs boson pairs, i.e.,  $hh \rightarrow b\bar{b}\gamma\gamma$ ,  $hh \rightarrow b\bar{b}\tau^+\tau^-$  and  $hh \rightarrow b\bar{b}b\bar{b}$ . The combined observed  
 1080 limit on the non-resonant Higgs boson pair cross-section is  $0.223 \text{ pb}$  at 95% CL, which corresponds to  
 1081 6.7 times the cross-section predicted by the SM. The ratio of the Higgs boson self-coupling to its SM  
 1082 expectation,  $k_\lambda$ , is constrained, at 95% CL, to be  $-5.0 < k_\lambda < 12.1$ .

1083 The upper limits on the cross section (at 95% CL) of Higgs pair production at 8 TeV as a function of  
 1084 the mass of a spin 0 resonance are summarized in figure 4.1(a) [70]. The limits were obtained using  
 1085 data collected with the CMS detector and come from searches using different final states, namely,  $b\bar{b}\gamma\gamma$   
 1086 (blue),  $b\bar{b}b\bar{b}$  (red and pink) and  $b\bar{b}\tau\tau$  (green). Depending on the analysis, the corresponding integrated  
 1087 luminosity varies between  $17.9 \text{ fb}^{-1}$  and  $19.7 \text{ fb}^{-1}$ . The results are usually interpreted in the framework  
 1088 of the Randall-Sundrum models such that the spin 0 resonance corresponds to the radion that decays  
 1089 to a pair of Higgs bosons. For a mass of 1 TeV the upper limit on the cross section is approximately 12  
 1090 pb.

1091 The upper limits on the cross section (at 95% CL) of Higgs pair production as a function of the mass  
 1092 of a heavy scalar Higgs,  $m_H$ , are summarized in figure 4.1(b) [74]. The limits were obtained through a  
 1093 combination of searches performed in the  $hh \rightarrow b\bar{b}\gamma\gamma$ ,  $hh \rightarrow b\bar{b}b\bar{b}$ ,  $hh \rightarrow b\bar{b}\tau^+\tau^-$  and  $hh \rightarrow \gamma\gamma WW^*$   
 1094 channels using proton-proton collision data corresponding to an integrated luminosity of  $20.3 \text{ fb}^{-1}$  col-  
 1095 lected with the ATLAS detector at a CM energy of  $\sqrt{s} = 8 \text{ TeV}$ . The improvement above  $m_H = 500 \text{ GeV}$   
 1096 is due to the sensitivity of the  $hh \rightarrow b\bar{b}b\bar{b}$  analysis. For  $m_H = 900 \text{ GeV}$  the observed limit is  $0.015 \text{ pb}$ .

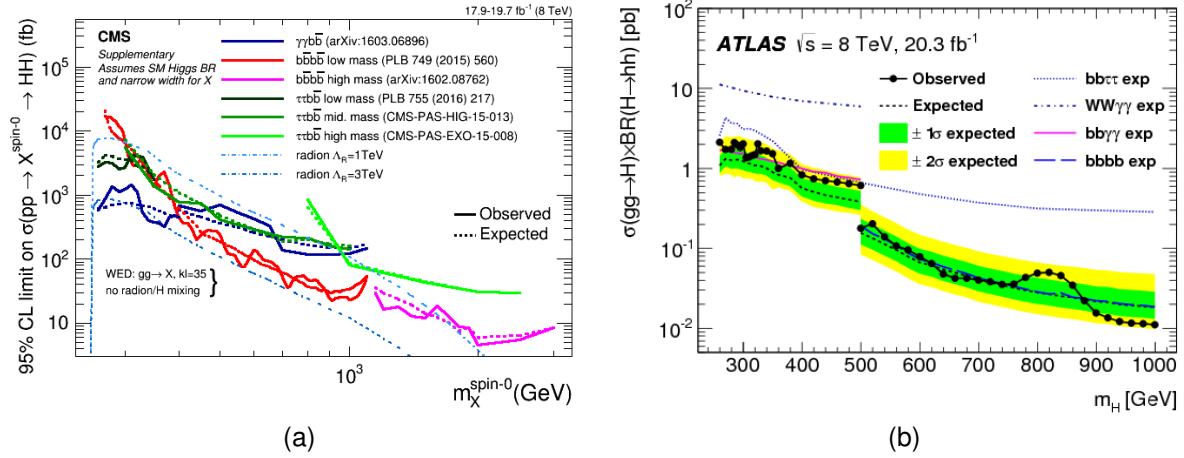


Figure 4.1: Exclusion limits (at 95% CL) on the cross section of resonant Higgs pair production from the decay of heavy scalar particles. These results were obtained using proton-proton collision data collected with the CMS (a) and ATLAS (b) detector at a CM energy of  $\sqrt{s} = 8$  TeV and result from the combination of different channels. Plots from [70] and [74].

## 4.2 Feasibility studies for high-luminosity and future colliders

Without any BSM contribution, the discovery of Higgs pairs production in the four b quarks final state at the LHC is highly unlikely even considering the total expected integrated luminosity of  $300 \text{ fb}^{-1}$ . There might be evidence for this process but retrieving useful information about the value of the Higgs trilinear coupling will remain out of reach. Nonetheless, the HL-LHC as well as future colliders pose a good opportunity for the discovery and precision studies of this process. Therefore, Monte Carlo studies assessing the feasibility of searches for  $hh \rightarrow b\bar{b}b\bar{b}$  at the HL-LHC and at the FCC-hh have been performed.

For the HL-LHC, a study including the  $pp \rightarrow b\bar{b}b\bar{b}$ ,  $pp \rightarrow b\bar{b}jj$ ,  $pp \rightarrow jjjj$  and  $pp \rightarrow t\bar{t} \rightarrow b\bar{b}jjjj$  backgrounds reports a significance ( $S/\sqrt{B}$ ) of 4 (1.3) for an integrated luminosity of 3000 (300)  $\text{fb}^{-1}$  [75]. The analysis is performed in three orthogonal regions (boosted, intermediate and resolved) and the reported significance is obtained from the combination of these three regions. The highest significance ( $S/\sqrt{B} = 2.9$ ) is achieved in the boosted category. In addition, the impact of pile up is evaluated. Considering a mean pile up of 80 and making use of jet grooming techniques, a significance of 3.1 (1.0) for an integrated luminosity of 3000 (300)  $\text{fb}^{-1}$  is reported (also considering the three analysis regions). This work makes use of artificial neural networks (ANN's) as well as of jet substructure observables to further increase the signal-background separation. Furthermore, single Higgs backgrounds, namely  $Z(\rightarrow b\bar{b})h(\rightarrow b\bar{b})$ ,  $t\bar{t}h(\rightarrow b\bar{b})$  and  $b\bar{b}h(\rightarrow b\bar{b})$ , are shown to be negligible for the analysis, when compared to the dominant QCD multijet background.

ATLAS and CMS also carried out preliminary studies for the sensitivity on the trilinear coupling at the HL-LHC. Several channels have been investigated:  $hh \rightarrow b\bar{b}\gamma\gamma$ ,  $hh \rightarrow b\bar{b}\tau^+\tau^-$ ,  $hh \rightarrow b\bar{b}WW^*$  and  $hh \rightarrow b\bar{b}b\bar{b}$ . The  $b\bar{b}\gamma\gamma$  final state is the most sensitive one. In this channel, ATLAS and CMS reported significances of  $1.05\sigma$  and  $1.6\sigma$ , respectively, for an integrated luminosity of  $3000 \text{ fb}^{-1}$ . Taking, as an example, the ATLAS result, the achieved significance translates to an upper limit on the total

1121 di-Higgs cross section of approximately twice the SM value. This corresponds to an exclusion limit of  
1122  $-0.8 < k_\lambda < 7.7$ . The analysis conducted by ATLAS in this channel is done at generator level. The  
1123 energy and momenta of the particles are smeared to simulate the detector's response. A mean pile up  
1124 of 200 is considered. No MVA techniques are employed in the analysis.

1125 For the  $hh \rightarrow b\bar{b}b\bar{b}$  channel, ATLAS states that a cross section 5.2 times larger than the SM value can  
1126 be excluded at 95% CL, with systematic uncertainties being taken into consideration. The Higgs trilinear  
1127 coupling is expected to be constrained to  $-3.5 < k_\lambda < 11$ . These results are based on the extrapolation  
1128 of the current results obtained with the 2016 dataset, comprising an integrated luminosity of  $10.1 \text{ fb}^{-1}$ .  
1129 This study is based only on the resolved analysis documented in [65].

1130 For the FCC-hh, a recent study that uses as signal sample  $pp \rightarrow hhj \rightarrow b\bar{b}b\bar{b}j$  and that includes  
1131 only the irreducible  $pp \rightarrow b\bar{b}b\bar{b}j$  reports a significance of 6.61 for an integrated luminosity of  $30 \text{ ab}^{-1}$ ,  
1132 considering an analysis that targets the boosted region [76]. The extra jet in the signal sample has  
1133  $p_T > 200 \text{ GeV}$  which provides the Higgs pair with a large Lorentz boost. This enhances the sensitivity  
1134 to the Higgs boson self coupling because it favors highly boosted virtual Higgs bosons decaying to a  
1135 pair of Higgs bosons. The analysis relies on the jet substructure observable  $\tau_{21}$  and on a tight mass cut  
1136 around the Higgs mass. No multivariate techniques are employed.

1137 A study comparing the feasibility of the search for di-Higgs production in the  $b\bar{b}b\bar{b}$  in the HL-LHC and at  
1138 the FCC-hh was presented in [77]. Only the irreducible background is considered. For a boosted region,  
1139 cut based analysis, a significance of 1.1 is reported for an integrated luminosity of  $3 \text{ ab}^{-1}$  at  $\sqrt{s} = 14 \text{ TeV}$   
1140 (HL-LHC). For an integrated luminosity of  $10 \text{ ab}^{-1}$  at  $\sqrt{s} = 100 \text{ TeV}$ , which corresponds to the FCC-hh,  
1141 this number is 5.7. While the significance is large for the FCC-hh, the reported signal to background ratio  
1142 ( $S/B$ ) is approximately one order of magnitude smaller, which means the measurement might be more  
1143 sensitive to systematic uncertainties on the backgrounds.

### 1144 4.3 Hadronic calorimeter granularity studies

1145 Even before the baseline design for the FCC-hh was established, there were studies regarding the  
1146 impact of the granularity of the calorimeters in the spatial resolving power of hadronic showers and on  
1147 the resolution of jet mass and substructure variables. These studies targeted the development of future  
1148 colliders and greatly influenced the baseline design of the FCC-hh.

1149 The granularity of hadronic calorimeters is a key parameter for future collider detectors because it  
1150 determines how well we can resolve energy deposits from pileup vertices and highly-boosted jet topolo-  
1151 gies. In Ref. [78], the use of smaller calorimeter cells (smaller than the ones that are used in currently  
1152 operating hadronic calorimeters) to resolve individual hadrons is investigated. For two Kaons ( $K_L^0$ ) with  
1153 an energy of 100 GeV each and with a truth level separation ( $\Delta R$ ) equal to 0.035 the energy deposited in  
1154 the ECAL (blue) and HCAL (red) is shown as a function of  $\Delta R$  for HCAL cells with sizes  $20 \text{ cm} \times 20 \text{ cm}$   
1155 ( $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ ) and  $5 \text{ cm} \times 5 \text{ cm}$  ( $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$ ) in figures 4.2(a) and 4.2(b),  
1156 respectively. The ECAL segmentation is equal to  $2 \text{ cm} \times 2 \text{ cm}$ . It is shown that for a granularity of  
1157  $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$  both hadrons can be resolved in the HCAL.

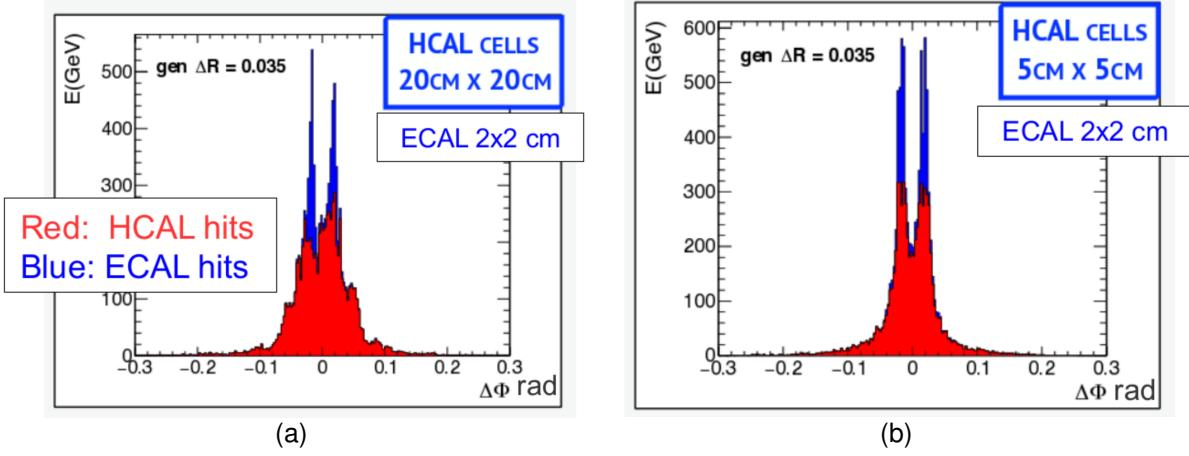


Figure 4.2: Azimuthal distribution of the energy deposits in the ECAL and HCAL for a pair of  $K_L^0$  with  $E = 100$  GeV for an hadronic calorimeter with  $20\text{ cm} \times 20\text{ cm}$  (a) and  $5\text{ cm} \times 5\text{ cm}$  (b). Figures from [79] (based on [78]).

Additional studies focusing on the jet mass resolution and on the resolution of jet substructure observables were performed for three calorimeter configurations (HCAL and ECAL). Some of these studies were presented in major conferences focused on future colliders, namely FCC week 2015 and 2016 and BOOST 2017. Here, we show results presented in Refs. [79–81]. The calorimeter configurations tested are: HCAL(ECAL)  $0.1(0.025) \eta \times 5.6(1.4)^\circ \phi$ , HCAL(ECAL)  $0.05(0.012) \eta \times 2.8(0.7)^\circ \phi$  and HCAL(ECAL)  $0.025(0.006) \eta \times 1.4(0.35)^\circ \phi$ . For the segmentation in  $\phi$  the numbers are give in degrees. For the HCAL, these correspond to approximately  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1, 0.05 \times 0.05, 0.025 \times 0.025$ .

For  $t\bar{t}$  events generated at NLO with MadGraph5 and passed through Delphes 3.2 to simulate detector response, the energy flow jet mass distribution for  $p_T(\text{jet}) > 3$  TeV is shown in figure 4.3(a) [80]. The jet mass resolution is shown in figure 4.3(b). For  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  cells the root mean square (RMS) of the distribution is 0.130. This value decreases to 0.090 and 0.088 for  $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$  and  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  cells, respectively. This indicates that the mass resolution increases as the granularity of the HCAL increases.

Regarding jet substructure, the resolution of the  $\tau_{32}$  variable in QCD dijet events generated with Pythia8 is shown in figure 4.4(a) for particle flow jets [80]. The same distribution is shown in figure 4.4(b) for calorimeter jets. Regardless of the type of jets used (eflow or tower jets), the  $\tau_{32}$  resolution increases as the HCAL resolution increases. In addition, it is important to note that the resolution is a lot worse when using tower jets which is an indication that exploiting tracking information is vital to achieve a good resolution in substructure variables.

Another interesting result, presented in Ref. [79], has to do with the overlap between the  $\tau_{21}$  distribution for jets resulting from QCD interactions and from the decay of a  $W$  boson. Jets with of 2.5 TeV show a reduction in overlap ( $80\% \rightarrow 60\%$ ) going from  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  to  $\Delta\eta \times \Delta\phi = 0.005 \times 0.005$  HCAL cells. For 5(10) TeV jets the overlap goes from 88(91)% to 78(85)%. However, for 20 TeV jets, the change in the HCAL granularity does not significantly modify the overlap, as is shown in figure 4.5.

In summary, previous studies focus on the impact of granularity in the resolution of jet mass and jet

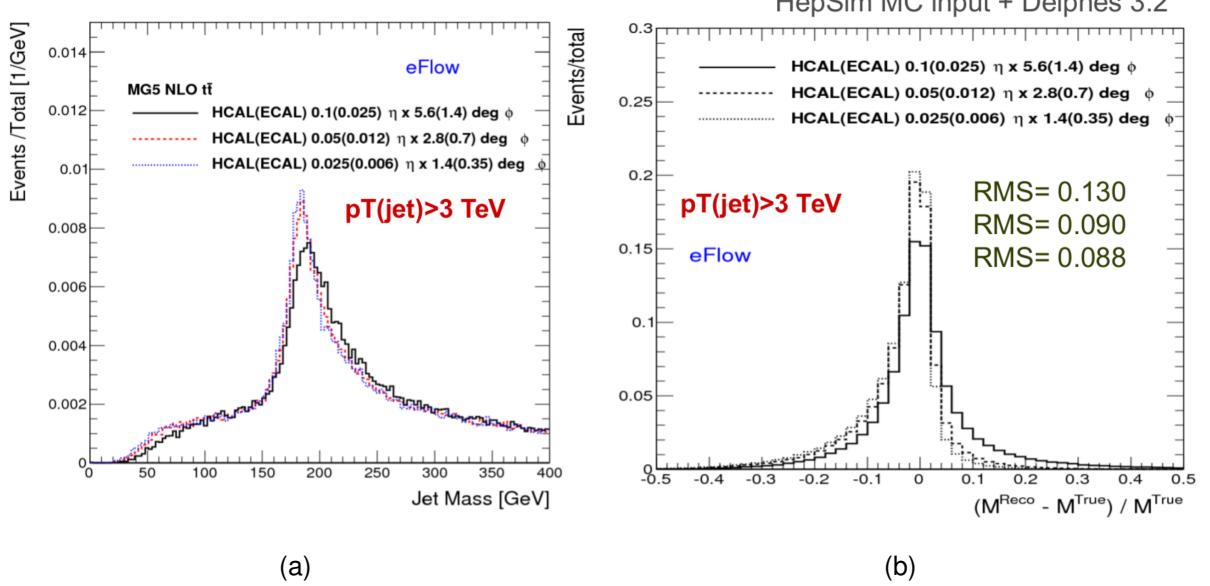


Figure 4.3: Jet mass (a) and jet mass resolution (b) plots in  $t\bar{t}$  events for eflow jets with  $p_T > 3$  TeV for three different HCAL and ECAL configurations: HCAL(ECAL) 0.1(0.025)  $\eta \times 5.6(1.4)$   $^\circ \phi$ , HCAL(ECAL) 0.05(0.012)  $\eta \times 2.8(0.7)$   $^\circ \phi$  and HCAL(ECAL) 0.025(0.006)  $\eta \times 1.4(0.35)$   $^\circ \phi$ . Plots from Ref. [79].

<sup>1183</sup> substructure observables. We did not find any studies that explored the change in the significance of a  
<sup>1184</sup> given analysis as a function of the granularity or of the detector configuration.

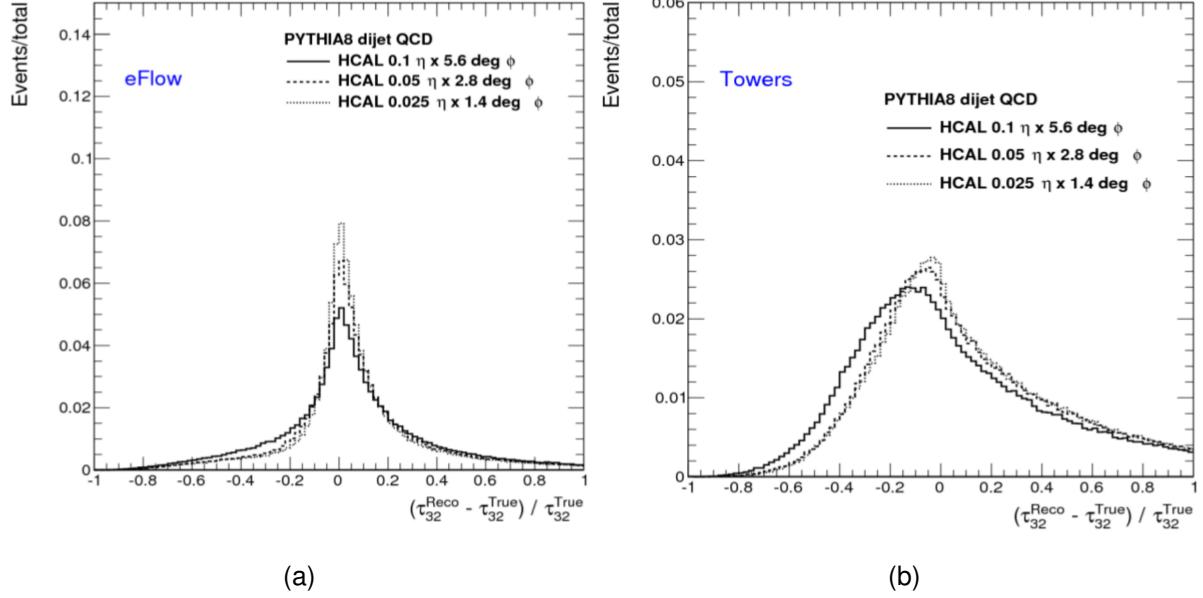


Figure 4.4:  $\tau_{32}$  resolution plots for QCD dijet events reconstructed using eflow jets (a) and tower jets (b). Plots from Ref. [80].

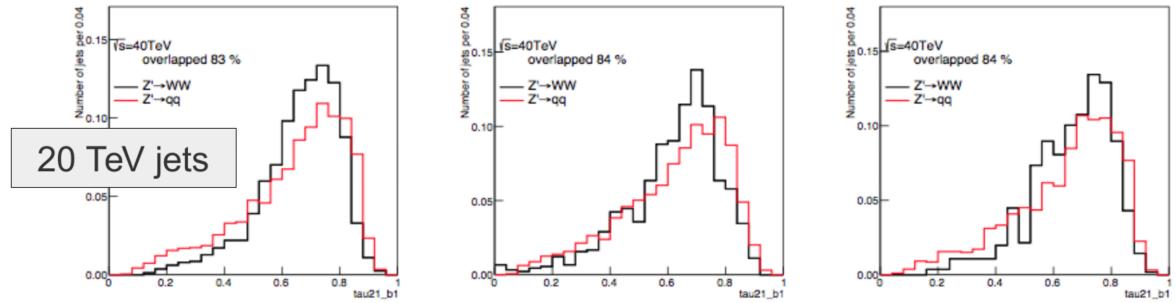


Figure 4.5:  $\tau_{21}$  variable for 20 TeV  $W$  (black) and QCD jets (red) for three HCAL granularities:  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  (left),  $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$  (middle) and  $\Delta\eta \times \Delta\phi = 0.005 \times 0.005$  (right). Plots from Ref. [79].

1185 **Chapter 5**

1186 **Sample generation and analysis tools**

1187 A crucial component of this work is the generation of the Monte Carlo samples that are used in the anal-  
1188 ysis. They are produced using fast simulation. We use the software and machinery that was developed  
1189 by the FCC study group at CERN. The simulation work flow and Monte Carlo samples are described in  
1190 this chapter.

1191 In section 5.1 we introduce the generators used to produce the Monte Carlo samples and briefly  
1192 describe their working principles and functionalities. We then focus on the FCC-hh software, in subsec-  
1193 tion 5.1.2, and explain how the previously described simulation work flow is implemented. A detailed  
1194 description of the technical settings used to produce the samples is provided in sections 5.2 and 5.2.3.

1195 **5.1 Fast simulation workflow**

1196 The Monte Carlo samples used in this study were generated using MadGraph5\_aMC@NLO [82]. The  
1197 showering and hadronization are simulated using Pythia 8 [83] and the detector response is parametrized  
1198 using Delphes 3 [84]. These samples are available at the CERN EOS storage.

1199 MadGraph5 is a matrix element generator for high energy physics processes, such as decays and  
1200 scatterings. The user specifies the desired process in terms of the initial and final states and can impose  
1201 additional constraints such as allowing for a number of refined criteria, including forced or forbidden s-  
1202 channel resonances, excluded internal particles, and forced decay chains of final state particles [82].  
1203 As a result, MadGraph automatically generates the corresponding Feynman diagrams and creates the  
1204 necessary code to compute the matrix element at a given point of the phase space. The output file is  
1205 written in the Les Houches Accord [85].

1206 Pythia8 is frequently used for event generation in high energy physics. In this work, however, we use  
1207 it only to simulate the showering and hadronization process and not the parton level hard process which  
1208 is simulated in MadGraph5. The Les Houches file that is produced by MadGraph5 is used as input to a  
1209 Pythia8 program that can decay unstable particles, simulate initial and final state showers as well as the  
1210 hadronization of coloured particles, such as quarks and gluons. The desired settings can be specified  
1211 in an additional file, a card, that is used as input for a Pythia8 run.

1212 Delphes3 allows for a quick and simple simulation of the detector's response [84]. Its goal is to allow  
1213 the fast simulation of a multipurpose detector for phenomenological studies. The simulation includes a  
1214 track propagation system, electromagnetic and hadronic calorimeters and a muon identification system.  
1215 Low level physics objects, such as tracks and energy deposits, and high level physics objects, such as  
1216 leptons and jets, are reconstructed from the detector's response and can be used to perform physics  
1217 analysis. In the following paragraphs we briefly describe how the detector response is simulated and  
1218 how jet reconstruction is performed.

1219 The magnetic field is uniform, axial and parallel to the beam direction. Charged particles follow a  
1220 helicoidal trajectory from the interaction point to the calorimeter while neutral particles have a straight  
1221 line trajectory. The probability of a charged particle being reconstructed as a track is defined by the  
1222 user. Only a smearing in the modulus of the transverse momentum is applied (not to the direction). The  
1223 tracking efficiency as well as energy and momentum resolutions are specified by the user and may in-  
1224 clude a dependency on the particle type, momentum and pseudorapidity. The calorimeters have a finite  
1225 segmentation in the  $(\eta, \phi)$  plane and the cell size can be defined in the configuration file. The amount  
1226 of energy deposited in the calorimeters by each particle type can be defined by the user. By default,  
1227 stable hadrons deposit all their energy in the HCAL although in a real detector a significant fraction of  
1228 their energy is deposited in the ECAL. The energy resolution of the ECAL and HCAL are parameter-  
1229 ized as a function of  $\eta$  and include stochastic, noise and constant terms:  $\sigma(E)/E = a/\sqrt{E} \oplus b$ . The  
1230 electromagnetic and hadronic energy deposits are independently smeared by log-normal distributions.

1231 Jets can be produced using generator level long-lived particles after showering and hadronization,  
1232 tracks, calorimeter towers or particle-flow tracks and towers. These are referred to as generator, track,  
1233 calorimeter or particle-flow jets, respectively. For generator level jets no detector simulation nor recon-  
1234 struction are taken into account. In spite of the type of jet the user can choose the jet clustering algorithm  
1235 and the values of its parameters as well as the minimum transverse momentum of the jets that are stored  
1236 in the final collection. Delphes integrates the FastJet package [86] and therefore allows jet reconstruc-  
1237 tion with the most popular jet reconstruction algorithms, namely, anti- $k_T$ ,  $k_T$  and Cambridge-Aachen  
1238 (C/A). Jets resulting from the hadronization of a b quark (known as b jets), are identified if a b quark is  
1239 found within a  $\Delta R$  distance from the jet's axis. The tagging efficiency and mis-tagging probabilities can  
1240 be defined by the user.

### 1241 5.1.1 Particle flow and calorimeter reconstruction in Delphes

1242 In Delphes, hadronic jets can be reconstructed using only the information from the HCAL towers or using  
1243 a particle flow algorithm that combines information from the tracking system and from the HCAL towers.  
1244 These two approaches create jets that are referred to as calorimeter and particle flow jets, respectively.  
1245 The latter can also be referred to as energy flow jets (eflow jets in short). In this work we performed the  
1246 analysis using both sets of jets and compare the results. Therefore we briefly describe them here.

1247 Calorimeter jets are very simple. They are reconstructed using as input for the jet clustering algorithm  
1248 the 4-vectors associated with the calorimeter towers, after a cell energy smearing has been applied.

1249 Therefore the energy resolution is limited by the transverse segmentation of the calorimeters.

1250 The goal of the particle flow approach is to make use of all the available information provided by the  
1251 various sub-detectors for reconstructing an event [84]. This approach is used by some experimental  
1252 collaborations [87, 88] but the exact implementation depends on the specificities of the experiment. If  
1253 the momentum resolution of the tracking system is better than the energy resolution of the calorimeters  
1254 it might be convenient to use the tracking information to estimate the momentum of charged particles.  
1255 In real experiments, the tracking resolution is only better than the calorimeter's energy resolution up to  
1256 some energy threshold. However, in Delphes, it is assumed that it is always convenient to estimate the  
1257 momentum of charged particles via the tracker.

1258 The particle flow algorithm works as follows [84]. For each calorimeter tower it counts:

- 1259 • the total energy deposited in ECAL and HCAL,  $E_{\text{ECAL}}$  and  $E_{\text{HCAL}}$ , respectively;
- 1260 • the total energy deposited in ECAL and HCAL originating from charged particles for which a track  
1261 has been reconstructed,  $E_{\text{ECAL,trk}}$  and  $E_{\text{HCAL,trk}}$ , respectively.

Then it defines  $\Delta_{\text{ECAL}} = E_{\text{ECAL}} - E_{\text{ECAL,trk}}$ ,  $\Delta_{\text{HCAL}} = E_{\text{HCAL}} - E_{\text{HCAL,trk}}$  and computes  $E_{\text{Tower}}^{\text{eflow}}$  given by

$$E_{\text{Tower}}^{\text{eflow}} = \max(0, \Delta_{\text{ECAL}}) + \max(0, \Delta_{\text{HCAL}}). \quad (5.1)$$

1262 All reconstructed tracks result in a particle flow track. If  $E_{\text{Tower}}^{\text{eflow}} > 0$  a particle flow tower is created with  
1263 energy  $E_{\text{Tower}}^{\text{eflow}}$ . The particle flow tracks and jets are then used as input for the jet clustering algorithms.

## 1264 5.1.2 FCC-hh software

1265 FCC software (FCCSW) [89], common to all FCC studies (electron-electron, electron-hadron and hadron-  
1266 hadron) has been developed and is maintained by the FCC study group. The software is based on Gaudi  
1267 [90]. An FCC Event Data Model based on Podio [91] was also developed. It consists in specific classes  
1268 that encode the information about the events.

1269 The FCC-hh study group is responsible for the generation of Monte Carlo quick simulation samples  
1270 for the main benchmark processes for the FCC-hh. The samples are generated using the workflow  
1271 described in the previous section. CERN users can request rights to run the EventProducer package  
1272 [92] and produce samples for any desired process using the machinery that is already implemented. In  
1273 addition, the FCC Event Data Model classes are directly accessible and can be used to read the ROOT  
1274 files that are produced after the events are passed through Delphes.

1275 The machinery to submit jobs to CERN's batch system is also implemented for both the generation  
1276 (MadGraph5) and reconstruction (Pythia8 plus Delphes3) levels.

1277 In this work we make use of this software in order to produce the necessary samples.

1278 **5.2 Signal and background samples**

1279 In this work we focus on the  $hh \rightarrow b\bar{b}b\bar{b}$  channel and perform an analysis targeting the boosted kinematic  
1280 region. The main backgrounds are multijet and  $t\bar{t}$  production and the irreducible  $pp \rightarrow b\bar{b}b\bar{b}$  process. This  
1281 was introduced and motivated in chapter 1 and it is discussed in detail in chapter 6. Here we provide a  
1282 technical description of how the signal and background samples were generated.

1283 **5.2.1 MadGraph**

1284 The irreducible background is generated with an extra jet with a high  $p_T$  at generator level,  $4b+j$ , where  $j$   
1285 stands for a light jet (initiated by a gluon or  $u, d, c, s$  quarks). This is referred to as the four flavor scheme.  
1286 A high- $p_T$  extra jet forces the four b quarks to have a high Lorentz boost and therefore increases the  
1287 probability of the events being reconstructed with two large- $R$  jets with a two-prong substructure. For this  
1288 process, QCD and electroweak contributions are considered separately, i.e, we include three different  
1289 types of samples: one in which only QCD processes are considered, one in which only electroweak  
1290 processes are considered and one in which both process are considered simultaneously. The samples  
1291 do not overlap. The  $4b+j$  QCD sample is constituted by two independent samples that have a different  
1292 generator level cut in the minimum  $p_T$  of the light jets, namely,  $200 < p_{T,j}^{\min} < 500$  GeV and  $p_{T,j}^{\min} > 500$   
1293 GeV. This allows for a more efficient generation.

1294 The multijet background is simulated through  $jj+0/1/2 j$  where  $j$  stands for a light or b jet (five flavor  
1295 scheme). This background is divided into several individual samples that are produced in different  $H_T$   
1296 regions, where  $H_T$  is the scalar sum of the  $p_T$  of all partons at generator level. The minimum(maximum)  
1297 allowed  $H_T$  is 500(100,000) GeV. Since these backgrounds are QCD processes, the  $p_T$  distribution of  
1298 the final state jets falls very steeply as the  $p_T$  increases. Therefore, if one were to generate events for  
1299 these processes without restricting the phase space, most events would consist of jets with a very low  
1300  $p_T$  which are exactly the type of events that are rejected the most by a boosted analysis (see chapter  
1301 6 for more details on the event topology that is targeted and on the analysis strategy). As we move to  
1302 regions of the phase with a higher  $H_T$  the cross section decreases meaning we need fewer MC events  
1303 to properly simulate the background in that region.

1304 Note that for the  $4b+j$  (QCD) and  $jj+0/1/2 j$  samples we do not take into account the regions of the  
1305 phase space with  $p_{T,j} < 200$  GeV and with  $0 < H_T < 500$  GeV. We assume that we can reject most of  
1306 the events (if not all) with these kinematic characteristics by going to a sufficiently boosted region of the  
1307 phase space. In addition, note that after the showering procedure there could be some overlap between  
1308 the  $4b+j$  and  $jj+0/1/2 j$ . This is taken care of in our analysis code: if an event from the  $jj+0/1/2 j$   
1309 has four b quarks at truth level [GIVE PERCENTAGE] then we do not consider it because it will certainly  
1310 overlap with an event from the  $4b+j$  sample.

1311 The  $t\bar{t}$  background sample is generated with extra jets at generator level,  $t\bar{t}+0/1/2 j$ , using the five  
1312 flavor scheme. We consider an inclusive sample for this background, meaning that we do not force any  
1313 particular decay of the top quark or of the subsequent particles.

1314 For the  $hh$  SM signal sample, one of the Higgs is decayed to  $b\bar{b}$  in MadGraph. The reasoning behind

1315 this choice is that most searches for Higgs pair production make use of a final state that includes at  
 1316 least two b quarks in order to keep the cross section times BR of the process large enough. In addition,  
 1317 this method allows the same generator level samples to be used to perform different analysis, simply by  
 1318 choosing the decay channel of the remaining Higgs boson. The decay of the remaining Higgs boson  
 1319 can be implemented in Pythia, in the case of this work, to  $b\bar{b}$ .

1320 The generator level cuts for the signal and background samples are summarized in table D.1 in  
 1321 appendix D.

## 1322 Signal samples - BSM

1323 In addition to the SM di-Higgs signal, we also explore the signature and analysis sensitivity for di-Higgs  
 1324 signals produced by two benchmark BSM models: CP-conserving type II 2HDM and a simplified dark  
 1325 matter model (DM) with a spin 0 mediator. These models were described in section 2.2.

1326 Both models are readily available in FeynRules [93] (which is a Mathematica package for the imple-  
 1327 mentation of new physics models) model database and can be straightforwardly implemented in Mad-  
 1328 graph5. The parameters of the models, namely the masses of the new particles, can be changed by  
 1329 the user. In the case of this work, we want new particles to have a large mass so that they produce SM  
 1330 Higgs pairs with a high Lorentz boost.

1331 For the DM model [18], the spin 0 mediator's mass is set at 1 TeV. The cross section for the signal  
 1332 generated with this model is smaller than the cross section of the SM signal (approximately 0.2 pb *versus*  
 1333 0.7 pb) and therefore this model is not excluded by experimental data.

1334 For the 2HDM [94, 95], a finer tuning of the parameters is required because the model is very general  
 1335 and has a lot external free parameters. In addition, it is written in the Higgs basis such that we need to  
 1336 convert the parameters from the physical basis (introduced in 2.2) to this basis. The free parameters of  
 1337 the model that we are interested in are the masses of the neutral,  $m_{h_1}, m_{h_2}, m_{h_3}$ , and charged scalars,  
 1338  $m_{hc}$ , the neutral scalars mixing angles,  $mix_h, mix_{h_2}, mix_{h_3}$ , the real and imaginary parts of the up, down  
 1339 and charged lepton  $3 \times 3$  Yukawa matrices,  $GUR, GUI, GDR, GDI, GLR, GLI$  and the values (real and  
 1340 imaginary parts) of the second, third and seventh quartic couplings,  $l_{2,3,7,..}$ . In the CP-conserving model  
 1341  $h_{1,2,3}$  can be identified with the physical states  $h, H$  and  $A$ , respectively.

We start from the following parameters (in the physical basis):

$$\begin{aligned}
 m_h &= 125 \text{ GeV}, & m_H &= 900 \text{ GeV}, & m_A &= 850 \text{ GeV}, & m_{H^\pm} &= 800 \text{ GeV} \\
 \beta &= \frac{\pi}{4}, & \alpha &= -0.75 \\
 m_{12}^2 &= [(m_H^2 + m_A^2 + m_{H^\pm}^2)/3] \cos(\beta) \sin(\beta) \simeq 181041.6667. & (5.2)
 \end{aligned}$$

In the Higgs basis, the input parameters for the model are

$$\begin{aligned}
m_{h_1} &= 125 \text{ GeV}, & m_{h_2} &= 900 \text{ GeV}, & m_{h_3} &= 850 \text{ GeV}, & m_{hc} &= 800 \text{ GeV} \\
mix_{h_2} &= mix_{h_3} = 0 \\
mix_{h_1} &= \frac{\pi}{2} - (\beta - \alpha) \simeq 0.035 \\
l_2 &\simeq 0.27, & l_3 &\simeq 9.46, & l_7 &\simeq 0.46.
\end{aligned} \tag{5.3}$$

The masses of the scalars cannot change when going from one basis to the other. Therefore we set them to the exact same values.  $mix_{h_2}, mix_{h_3} = 0$  because we are considering the CP-conserving model. To obtain the values of  $l_{2,3,7}$  we use Eq. 11 from [24] we write the  $\lambda$  parameters of the scalar potential in terms of the parameters defined in Eq. 5.2. Then we use Eq. 47,48 and 50 from [96] to obtain the values of the quartic coupling in Higgs basis.

Regarding the Yukawa interactions, we work in the type II model. Following the type II restriction file found in [94] we set

$$\begin{aligned}
GLR &= GLI = GDI = GUI = 0 \\
GDR &= \text{diag} \left( 0, 0, \frac{m_b \sqrt{2} \tan(\beta)}{v} \right) \\
GDR &= \text{diag} \left( 0, 0, -\frac{m_t \sqrt{2}}{v \tan(\beta)} \right)
\end{aligned} \tag{5.4}$$

where  $m_b = 4.7$  GeV and  $m_t = 172$  GeV are the masses of the bottom and top quarks.

## 5.2.2 Pythia

For the signal samples (SM and BSM) we simply turn off all other decays except  $h \rightarrow b\bar{b}$  therefore forcing the Higgs to decay to a pair of b quarks leading to the desired final state with four b quarks. All other settings are not altered with respect to their default configuration. For the BSM samples, this implies that the coupling of the SM Higgs boson to the b quarks is set to its SM value.

For the  $jj + 0/1/2$  and  $t\bar{t} + 0/1/2 j$  samples we have to perform jet matching because we require additional jets at the level of the matrix element. In addition to the partons generated in MadGraph and that can produce a jet, Pythia may introduce extra jets that are usually soft and collinear (with the particle from which they were radiated) and result from the showering process. This could lead to the same process (meaning processes with the same final states) being counted twice (double counting). Take, for example, the processes  $jj$  and  $j jj$  at MadGraph level. It can happen that Pythia generates an extra jet for the first process but not for the second, leading to both processes having the same final state (three jets). Each process would then give its independent contribution to the total number of events but because they simply represent two distinct ways of achieving the same final state they should only be counted once. The goal of jet matching procedures is to avoid this problem.

The cross section times branching ratio (when applicable), the k-factors and the total number of

1364 events used in the analysis are summarized in table 5.1, for all samples. The k-factor multiplies the  
 1365 cross section times branching ratio in order to reproduce known higher order results. It corresponds  
 1366 only to the ratio between the total cross sections and it does not correct for possible differences that  
 1367 might exist between the differential cross sections.

For the SM signal,  $\sigma \times BR$  is given by  $\sigma(hh, h \rightarrow b\bar{b}) \times BR(h \rightarrow b\bar{b})$ , with  $BR(h \rightarrow b\bar{b}) = 0.5824$  for  $m_h = 125$  GeV [97].  $\sigma(hh, h \rightarrow b\bar{b})$  is given by:

$$\sigma(hh, h \rightarrow b\bar{b}) = \sigma(hh) \times 2 \times BR(h \rightarrow b\bar{b}) = 1.22 \text{ pb} \times 2 \times 0.5824 = 1.42 \text{ pb} \quad (5.5)$$

1368 where  $\sigma(hh) = 1.22$  pb follows from Ref. [98] and the factor of 2 is a combinatorial factor that indicates  
 1369 that both Higgs bosons can decay to a pair of b quarks. The k-factor is given by  $\sigma_{\text{NNLL}}/\sigma_{\text{NNLO}}$ , where  
 1370  $\sigma_{\text{NNLL}}$  is calculated according to the prescription given in equation I.7.8 of Ref. [97], with  $\delta_t = -0.315$ ,  
 1371 which yields  $\sigma_{\text{NNLL}} = 1.33$  pb.

For the BSM signal samples,  $\sigma \times BR$  is given by:

$$\sigma \times BR = \sigma(hh)_{\text{MG}} \times (BR(h \rightarrow b\bar{b}))^2 \quad (5.6)$$

1372 where  $\sigma(hh)_{\text{MG5}}$  is the cross section for Higgs pair production as given by MadGraph. For these sample  
 1373 we consider a k-factor of 1.0.

1374 For the remaining samples,  $\sigma \times BR$  are the values given by MadGraph. In these samples no decay  
 1375 mode is imposed, therefore the branching ratio is one. For the  $4b + j(\text{QCD})$  sample, a k-factor of two  
 1376 is applied [92] in order to parameterize our ignorance on the QCD irreducible background. For the  
 1377  $t\bar{t} + 0/1/2 j$ , a k-factor of 1.74 is applied, following Ref. [92].

1378 The parameters used in Pythia 8 can be found in table D.2 in appendix D.

### 1379 5.2.3 Delphes

1380 It is one of the main goals of this work to evaluate how the analysis sensitivity is influenced by the  
 1381 granularity of the hadronic calorimeter. We start from the same MadGraph level samples and pass them  
 1382 through Pythia and Delphes changing the settings of the Delphes card that correspond to the HCAL. All  
 1383 other detector's parameters were kept unchanged with respect to the FCC default Delphes card. We  
 1384 tested five benchmark granularity configurations:

- 1385 1. ATLAS HCAL granularity (as implemented in the standard ATLAS Delphes card);
- 1386 2. Starting from the ATLAS HCAL configuration we increase the granularity in  $|\eta|$  by a factor of four,  
 1387 in the pseudo rapidity range  $|\eta| < 1.7$  which corresponds to the TileCal region;
- 1388 3. Starting from the FCC HCAL configuration we decrease the granularity in  $\phi$  by a factor of two, in  
 1389 the entire pseudo rapidity range covered by the HCAL.
- 1390 4. FCC HCAL default granularity (as implemented in the standard FCC Delphes card);

Table 5.1: Summary of the effective cross sections ( $\sigma \times BR$ ) and k factors of the signal and background samples used in the analysis.

Sample	$\sigma \times BR$ [pb]	k-factor
$hh \rightarrow b\bar{b}b\bar{b}$ - <b>SM</b>	0.827	1.09
$hh \rightarrow b\bar{b}b\bar{b}$ - <b>DM mediator</b>	0.218	1.0
$hh \rightarrow b\bar{b}b\bar{b}$ - <b>2HDM type II</b>	1.85	1.0
$4b + j$ (QCD, $200 < p_T^j < 500$ )	756.4	2.0
$4b + j$ (QCD, $p_T^j > 500$ )	57.71	2.0
$4b + j$ (QCD+EWK)	6.204	1.0
$4b + j$ (EWK)	0.07206	1.0
$jj + 0/1/2 j$ ( $500 < H_T < 1000$ )	$1.64 \times 10^7$	1.0
$jj + 0/1/2 j$ ( $1000 < H_T < 2000$ )	$1.67 \times 10^6$	1.0
$jj + 0/1/2 j$ ( $2000 < H_T < 4000$ )	$1.32 \times 10^5$	1.0
$jj + 0/1/2 j$ ( $4000 < H_T < 7200$ )	$7.32 \times 10^3$	1.0
$jj + 0/1/2 j$ ( $7200 < H_T < 15000$ )	$4.75 \times 10^2$	1.0
$jj + 0/1/2 j$ ( $15000 < H_T < 25000$ )	7.35	1.0
$jj + 0/1/2 j$ ( $25000 < H_T < 35000$ )	0.176	1.0
$jj + 0/1/2 j$ ( $35000 < H_T < 100000$ )	0.00765	1.0
$t\bar{t} + 0/1/2 j$	$4.31 \times 10^4$	1.74

- 1391 5. Starting from the FCC HCAL configuration we increase the granularity in  $|\eta|$  and in  $\phi$  by a factor of  
 1392 two, in the entire pseudo rapidity range covered by the HCAL.

1393 The granularities of these five configurations are summarized in table 5.2. The values that are shown,  
 1394 as well as the corresponding pseudorapidity regions, are exactly what is implemented in Delphes.

1395 In addition, we also passed the same generator level samples through the default ATLAS detector  
 1396 simulation in Delphes. The HCAL granularity is the one that is indicated in table 5.2 but the other detector  
 1397 parameters, such as the radius, magnetic field, tracking resolutions are the ones that are implemented  
 1398 in the default ATLAS Delphes card. This additional detector configuration gives us an extra point in the  
 1399 space of parameters that we are trying to explore. Furthermore, in a first, very crude, approximation,  
 1400 it allow us to compare the results obtained at  $\sqrt{s} = 100$  TeV to the ones obtained at  $\sqrt{s} = 13$  TeV  
 1401 by ATLAS. For completion we summarize, in table 5.3, the values of some key detector parameters as  
 1402 implemented in Delphes for the ATLAS and FCC-hh detectors. The magnetic field is twice as strong for  
 1403 the FCC-hh. The charged hadrons tracking efficiency and the HCAL resolution are fairly similar between  
 1404 the two detector configurations.

Table 5.2: Summary of the benchmark granularity configurations of the HCAL.

<b>Configuration</b>	$\Delta\eta \times \Delta\phi$	$\eta$ range
1 (ATLAS HCAL)	$0.1 \times 0.1$	$ \eta  < 2.5$
	$0.2 \times 0.2$	$2.5 <  \eta  < 5.0$
2 (ATLAS HCAL $\eta \times 4$ )	$0.025 \times 0.1$	$ \eta  < 1.7$
	$0.1 \times 0.1$	$1.7 <  \eta  < 2.5$
3 (FCC HCAL $\phi/2$ )	$0.2 \times 0.2$	$2.5 <  \eta  < 5.0$
	$0.025 \times 0.05$	$ \eta  < 2.5$
4 (FCC HCAL)	$0.05 \times 0.1$	$2.5 <  \eta  < 6.0$
	$0.025 \times 0.025$	$ \eta  < 2.5$
5 (FCC HCAL $\eta, \phi \times 2$ )	$0.05 \times 0.05$	$2.5 <  \eta  < 6.0$
	$0.0125 \times 0.0125$	$ \eta  < 2.5$
	$0.025 \times 0.025$	$2.5 <  \eta  < 6.0$

Table 5.3: Summary of some key detector parameters for the FCC-hh and ATLAS detectors.

<b>Parameter</b>	FCC-hh	ATLAS
Radius of magnetic field coverage [m]	1.5	1.15
Half length [m]	5	3.51
Magnetic field [T]	4	2
Charged hadrons tracking efficiency	90% for $2.5 <  \eta  < 4.0$ , $p_T > 1.0$	85% for $1.5 <  \eta  < 2.5$ , $p_T > 1.0$
HCAL resolution	$\sqrt{E^2(0.03)^2 + E(0.60)^2}$ for $1.7 <  \eta  < 4.0$	$\sqrt{E^2(0.05)^2 + E(0.706)^2}$ for $1.7 <  \eta  < 3.2$

1405

# Chapter 6

1406

# Analysis

1407

## 6.1 Overview of the $hh \rightarrow b\bar{b}b\bar{b}$ channel

1408 For the SM Higgs, with a mass around 125 GeV, the branching ratio of the  $hh \rightarrow b\bar{b}b\bar{b}$  decay is ap-  
1409 proximately 33.6%, making it the most probable decay for Higgs pairs, as is illustrated in figure 6.1(a).  
1410 However, in this channel, the main background is QCD multijet production that has a cross section sev-  
1411 eral orders of magnitude larger than di-Higgs production in the SM, as table 5.1 shows. Nonetheless,  
1412 the jet  $p_T$  distributions in this background have a very large yield close to zero and then fall very steeply  
1413 while the signal has a much larger tail to high values of  $p_T$ . This indicates that searches targeting the  
1414 boosted kinematic regime may be the key to measure  $hh \rightarrow b\bar{b}b\bar{b}$  using inclusive production.

1415 In this work we target the boosted regime. In this kinematic region, the final state of the signal is  
1416 characterized by two jets with a large radius parameter. Each jet is expected to contain the two b quarks  
1417 originated from the decay of a Higgs boson, as figure 6.1(b) illustrates. Extra jets are also expected to  
1418 be reconstructed as a consequence of QCD radiation.

1419 The analysis presented here is performed using the Monte-Carlo samples described in chapter 5. All  
1420 samples assume  $m_h = 125$  GeV. The event selection criteria are designed to optimize the significance,  
1421 given by  $S/\sqrt{B}$ , where  $S$  and  $B$  represent the number of signal and background events, respectively.

1422

### 6.1.1 Event pre-selection

1423 We start by applying very simple and loose cuts that target events consistent with the boosted topol-  
1424 ogy. Jets are reconstructed with the anti- $k_T$  algorithm with an radius parameter  $R = 0.8$ . They are  
1425 reconstructed from particle flow objects or calorimeter towers. Both approaches were explored in this  
1426 analysis.

1427 We require at least two b-tagged  $R = 0.8$  jets which corresponds to at least four b-tagged subjets,  
1428 at least two in each  $R = 0.8$  jet. The subjets are found through the soft drop mass technique that was  
1429 described in section 3.3.2. The leading and sub leading jets must have  $p_T \geq 200$  GeV in order for the  
1430 event to be accepted. From Eq. 3.4, a Higgs boson with, e.g.,  $p_T = 200$  GeV leads to a pair of b quarks  
1431 with  $\Delta R \sim \frac{2m}{p_T} = 1.25$ . Therefore, considering jet with  $R = 0.8$  (diameter equal to 1.6) and  $p_T \geq 200$  GeV

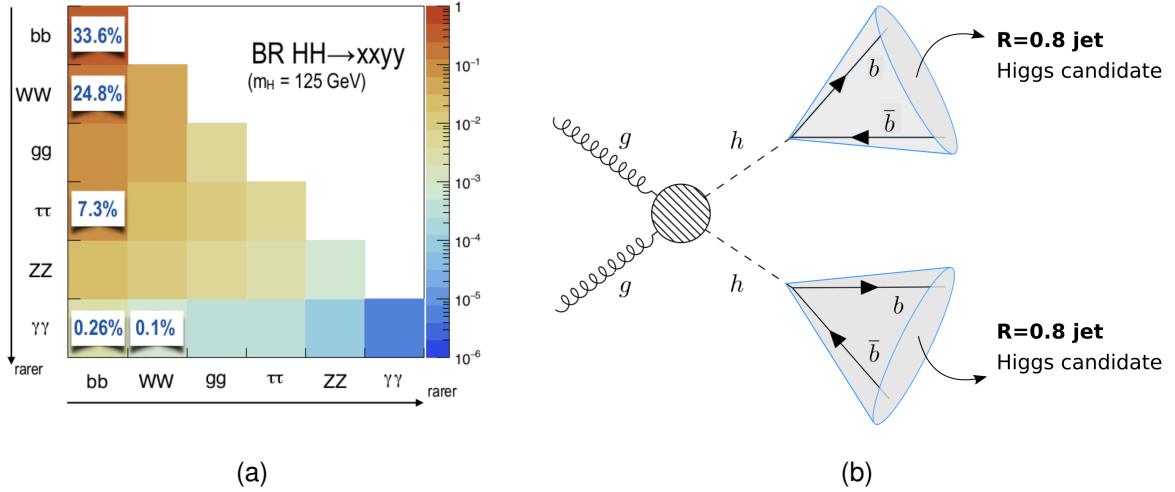


Figure 6.1: Higgs pairs branching ratios (left) and event topology targeted by this analysis (right).

is a reasonable choice. For a more complete description about the jet radius see appendix A. Due to the b-tagging parameterization (described in detail in section 6.2.1), there is a natural cutoff at  $|\eta| = 4$  so we do not place any additional cut in  $\eta$ . From now on, these cuts are referred to as pre-selection cuts.

After the pre-selection cuts we can plot all the relevant distributions that allow us to characterize the signal and background events. These are shown in the next two sections. In addition, these distributions provide a first insight into the discriminating power of the kinematic and substructure variables that are explored.

### 6.1.2 Signal characterization

The cross section for di-Higgs inclusive production from  $pp$  collisions at 100 TeV is 1.22 pb [98]. This value is computed at NNLO accuracy. The dominant process is gluon-gluon fusion that has been extensively discussed in section 2.1.1. The signal process comprehends the di-Higgs inclusive production and the  $hh \rightarrow b\bar{b}b\bar{b}$  subsequent decays. The total process cross section amounts to approximately 0.9 pb. This number includes the branching ratio and a k-factor to account for known NNLL results. This is described in detail in section 5.2.2. For an integrated luminosity of  $30 \text{ ab}^{-1}$  of proton-proton collisions at  $\sqrt{s} = 100$  TeV we expect  $2.7 \times 10^7$  events.

The Higgs candidates are reconstructed using large  $R$  jets that can be directly measured using information from the calorimeters and tracking systems. Figure 6.2(a) shows the  $\Delta\phi$  between the two leading jets (Higgs candidates),  $\Delta\phi(hh)$ . For a large fraction of the signal events (red curve)  $\Delta\phi(hh) \sim \pi$  which indicates that the jets are produced back-to-back in the transverse plane. There is also a small fraction of events for which the jets have  $\Delta\phi(hh) \sim 1$ . For these events there must be at least a third jet that balances the momentum of the jets pair.

Each wide jet is expected to be consistent with having two subjets associated with the two b quarks from the  $h \rightarrow b\bar{b}$  decay. In the Higgs rest frame, the two subjets are produced back-to-back conserving momentum. In the laboratory frame, the  $\Delta R$  between the subjets depends on the momentum of the

1456 Higgs boson, with larger  $\Delta R$  corresponding to Higgs bosons with a lower momentum. This correlation  
1457 between the  $\Delta R$  between the subjets of the leading Higgs candidate ( $\Delta R(b\bar{b})$ ) and its momentum  
1458 ( $p_T(h_1)$ ) is shown in figure 6.2(b).

1459 The  $p_T$  distributions of the leading and sub leading jets (Higgs candidates) are shown in figures  
1460 6.5(a) and 6.5(b), respectively. The histograms are normalized to unit area in order to allow for shape  
1461 comparison between them. For the leading jet, the signal  $p_T$  spectrum peaks at approximately 350 GeV  
1462 and has a long tail for large values of  $p_T$ , in particular, longer than any of the backgrounds, at least up to  
1463  $p_T > 1000$  GeV. For the sub leading jet the spectrum is softer, as expected, but the tail of the distribution  
1464 is still longer than for any of the backgrounds. These distributions show that the signal process produces  
1465 jets with larger transverse momenta than the jets produced by the background processes which indicates  
1466 that a boosted analysis might perform well.

1467 The  $\eta$  distribution of the leading jet is shown in figure 6.6(a). This distribution is limited to  $|\eta| < 4$   
1468 because the b-tagging efficiency goes to zero for  $|\eta| > 4$  (see section 6.2.1 for more details). The  
1469 number of events decreases as we go to larger values of  $\eta$  which is due to the detector's acceptance.  
1470 The  $\eta$  distribution of the sub leading jet is very similar to this one so we abstain from showing it here.

1471 The softdrop mass distribution of the leading jet is shown in figure 6.6(b). For the signal (red curve)  
1472 there is a clear peak at approximately 120 GeV which corresponds to the SM Higgs boson mass peak.  
1473 The peak is quite broad which means the mass resolution in this channel is not very good. On the one  
1474 hand, this is because we are using large- $R$  jets to reconstruct the Higgs candidates. These objects  
1475 have a worse mass resolution than other cleaner objects such as photons or electrons. On the other  
1476 hand, the cuts applied before plotting these distributions are very loose. As we place additional cuts the  
1477 mass peak can be made slightly narrower. Another interesting feature of the signal mass spectrum is  
1478 the existence of a peak close to zero. The explanation is the following: some jets do not contain both b  
1479 quarks from the Higgs decay such that when applying the mass drop procedure only one b quark with a  
1480 mass of approximately 5 GeV remains inside the jet, creating the peak at lower masses. A more detailed  
1481 discussion, as well as a plot that supports this interpretation, can be found in appendix C.

1482 In addition to the basic kinematic distributions that we have just described, there are a multitude of  
1483 other variables one can look at. Here we show the  $\Delta\eta$  between the Higgs pairs, figure 6.7(a), and the  
1484  $\tau_{21}$  variable for the leading Higgs candidate, figure 6.7(b). The first distribution shows that the Higgs pair  
1485 tends to have a  $\Delta\eta$  close to zero, which indicates that the pair is highly boosted in the longitudinal plane.  
1486 The  $\tau_{21}$  distribution shows that for the signal this variable takes values close to zero which means the jet  
1487 is consistent with having two subjets.

### 1488 6.1.3 Backgrounds

1489 The relevant backgrounds for this analysis are QCD multijet production,  $t\bar{t}$  and  $b\bar{b}b\bar{b}$ . Although the  $b\bar{b}b\bar{b}$   
1490 background is a particular case of a QCD multijet production process we consider it separately because  
1491 it constitutes the irreducible background and therefore will have a higher efficiency in the analysis. The  
1492 cross sections for these processes are several orders of magnitude larger than the cross section for the

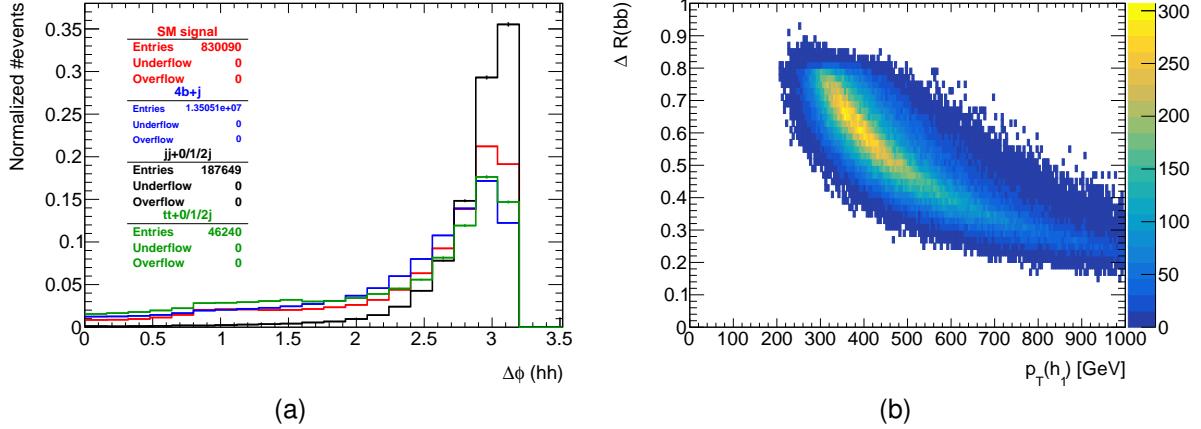


Figure 6.2:  $\Delta\phi$  between the Higgs candidates,  $\Delta\phi(hh)$ , (left) and  $\Delta R$  between the b quarks coming from the decay of the leading Higgs candidate,  $h_1$  (right). To obtain this plot the cut  $|M - 125| < 40$  GeV was applied in addition to the pre-selection cuts..

1493 signal, as table 5.1 shows. In addition, in the case of the  $t\bar{t}$  background, the event topology is expected  
 1494 to be similar to the signal because it also consists in the production of a pair of particles with the same  
 1495 mass.

1496 The assumption that QCD multijet production and  $t\bar{t}$  are the two main backgrounds is corroborated  
 1497 by the ATLAS di-Higgs search performed in the same channel, where these backgrounds are found to be  
 1498 the dominant ones. All other sources of backgrounds, including processes involving Higgs bosons, are  
 1499 found to be negligible [65]. In appendix B we discuss and evaluate the importance of some backgrounds  
 1500 that include Higgs bosons in our analysis.

1501 Figure 6.3 shows examples of LO Feynman diagrams that contribute to  $4b + j$  (left), three (middle)  
 1502 and four (right) light jets production. The  $b\bar{b}b\bar{b}$  background is generated with an extra light jet at generator  
 1503 level, as discussed in chapter 5. This jet boosts the four b quarks and has a minimum  $p_T$  of 200 GeV. This  
 1504 increases the probability of two b quarks being reconstructed as single large- $R$  jet therefore emulating  
 1505 the signal's final state signature, i.e, high  $p_T$  jets compatible with a two-prong substructure. The QCD  
 1506  $4b + j$  sample with  $(200 < p_T < 500)$  GeV has a cross section of approximately 756 pb. In addition, as  
 1507 explained in section 5.2.2, we multiply this cross section by a k-factor of 2, which yields a total cross  
 1508 section of approximately  $1.5 \times 10^3$  pb. The cross section is a lot smaller for the other  $4b + j$  QCD sample  
 1509 as well as for the samples that include both QCD and electroweak contributions, as expected.

1510 The QCD multijet background consists of two, three and four jet events (represented as  $jj + 0/1/2 j$ ),  
 1511 at generator level. The jets can originate from light and b quarks and from gluons. A jet matching pro-  
 1512 cedure is implemented in Pythia in order to avoid double counting. Due to mis-tagging probabilities, light  
 1513 and c jets can be identified as b jets. Although these probabilities are relatively small when compared  
 1514 to the b-tagging efficiency (see section 6.2.1) the cross section of multijet processes is very large such  
 1515 that this background becomes dominant. The  $jj + 0/1/2 j$  sample with  $(500 < H_T < 1000)$  GeV has  
 1516 a cross section of the order of  $10^7$  GeV. As anticipated, the  $p_T$  spectrum of the multijet background is a  
 1517 lot softer than for the signal and remaining backgrounds, as it is shown in figures 6.5(a) and 6.5(b). In

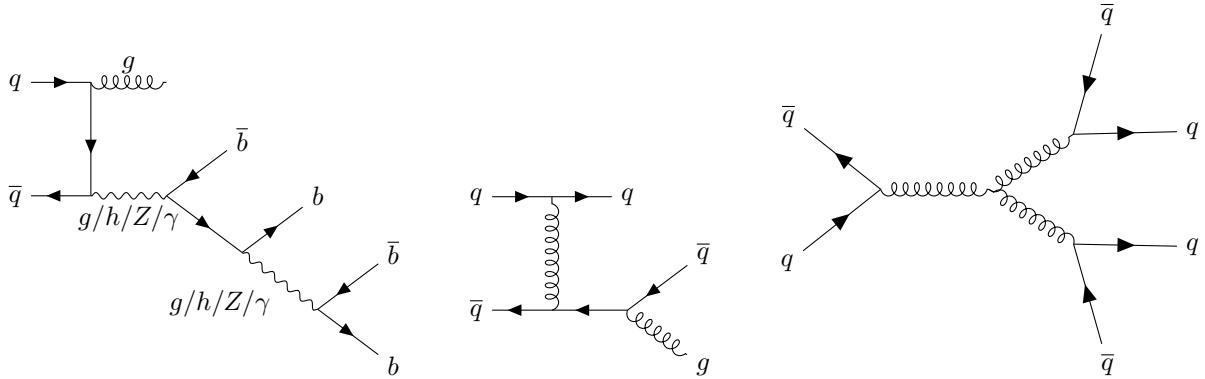


Figure 6.3: Example of diagrams that contribute to the QCD multijet background: five final state jets, four of which are b-jets (left), three final state jets (middle) and four final state jets (right). Here,  $q$  stands for a light quark/jet.

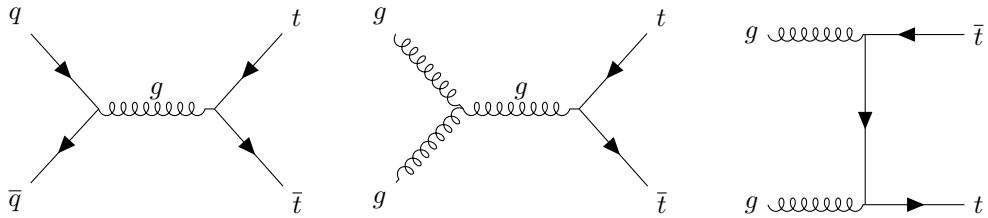


Figure 6.4: Dominant diagrams of  $pp \rightarrow t\bar{t}$  at LO.

addition, for this background the  $\tau_{21}$  variable takes values close to one, as it is shown in figure 6.7(b).  
This indicates that the jets are not consistent with two subjets.

Figure 6.4 shows examples of LO Feynman diagrams that contribute to  $t\bar{t}$  production through  $q\bar{q}$  (left) and  $gg$  (middle and right) fusion. The  $t\bar{t}$  background is simulated with additional zero, one or two jets (represented as  $t\bar{t} + 0/1/2 j$ ), at generator level. The extra jets can originate from light and b quarks and from gluons. A jet matching procedure is implemented in Pythia in order to avoid double counting. The top quark has a very short life time, predicted to be  $5 \times 10^{-25}$  s, such that it decays before it can hadronize. This sample is inclusive in the top quark decay modes. However, the most favoured decay of the top quark is  $t \rightarrow Wb$  with a branching ratio close to 96% [99]. Therefore,  $t\bar{t}$  events will, most of the times, result in the  $W^+W^-b\bar{b}$  final state. We do not specify any decay mode for the  $W$  such that the sample is also inclusive in the  $W$  decay modes. The  $W$  decays to hadrons ( $W^+ \rightarrow q\bar{q}$ ) and leptons ( $W^+ \rightarrow l^+\nu$ ) with  $BR(W^+ \rightarrow q\bar{q}) \sim 68\%$  and  $BR(W^+ \rightarrow l^+\nu) \sim 10\%$ . If one (or both)  $W$  bosons decays to hadrons then there will be additional jets in the final state. These can be b jets or can be misidentified as such. If both  $W$  bosons decay to leptons there will still be at least two b jets in the final state, coming from the  $t\bar{t}$  decay. The cross section of the  $t\bar{t} + 0/1/2 j$  is approximately  $4.3 \times 10^4$  pb. As explained in 5.2.2, k-factor of 1.74 is applied which yields a total cross section of approximately  $7.5 \times 10^4$  pb.

It is interesting to note that in the softdrop mass plot for the leading Higgs candidate (figure 6.6(b)) the  $t\bar{t}$  background (green line) shows a small peak around 170 GeV. In this region, all the decay products of the top quark are contained inside the  $R = 0.8$  jet such that the jet mass corresponds to the mass of the original top quark.

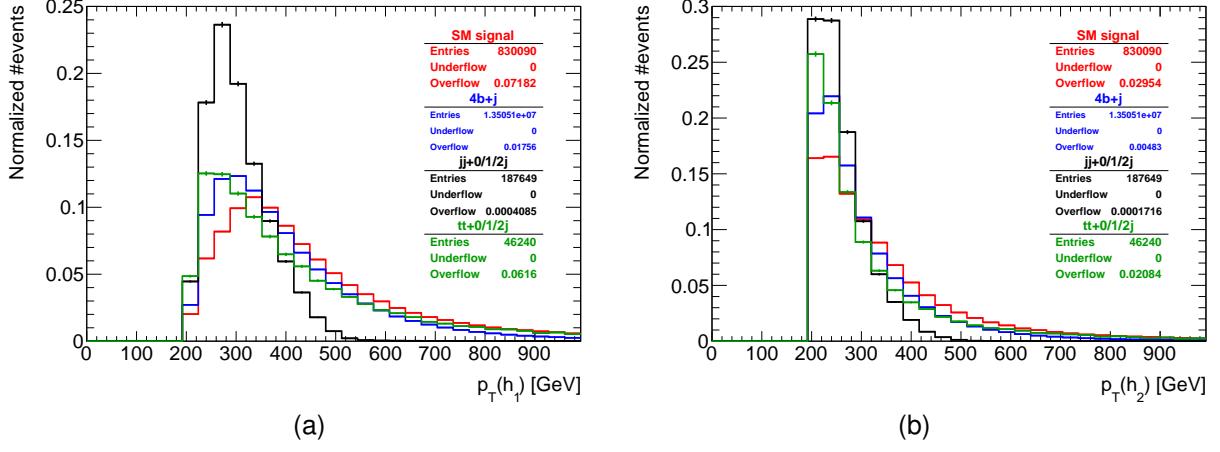


Figure 6.5:  $p_T$  distributions for the leading (left) and sub-leading (right) Higgs candidates. The signal is the SM  $hh \rightarrow b\bar{b}$  process. The histograms are normalized to unit area.

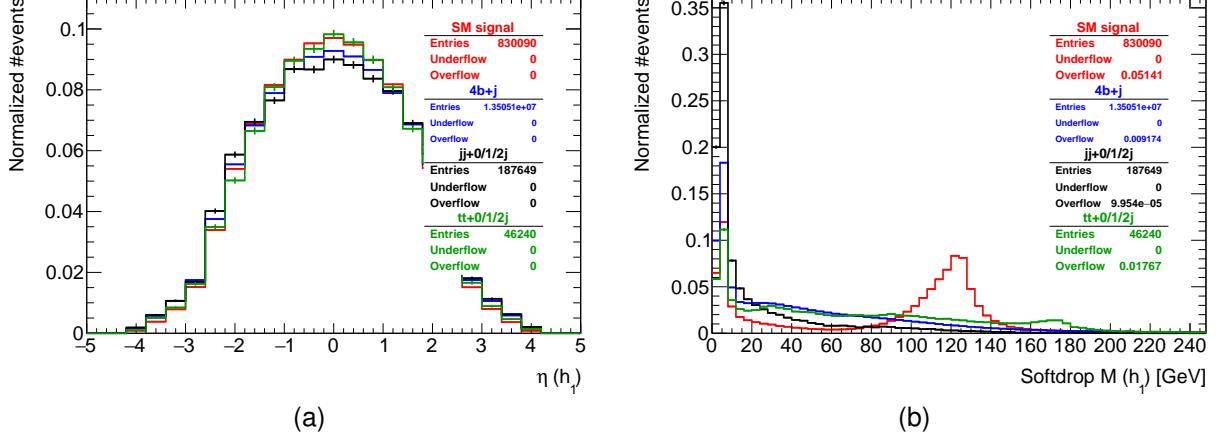


Figure 6.6:  $\eta$  distribution for the leading Higgs candidate (left) and softdrop mass distribution for the leading Higgs candidate (right). The signal is the SM  $hh \rightarrow b\bar{b}$  process. The histograms are normalized to unit area.

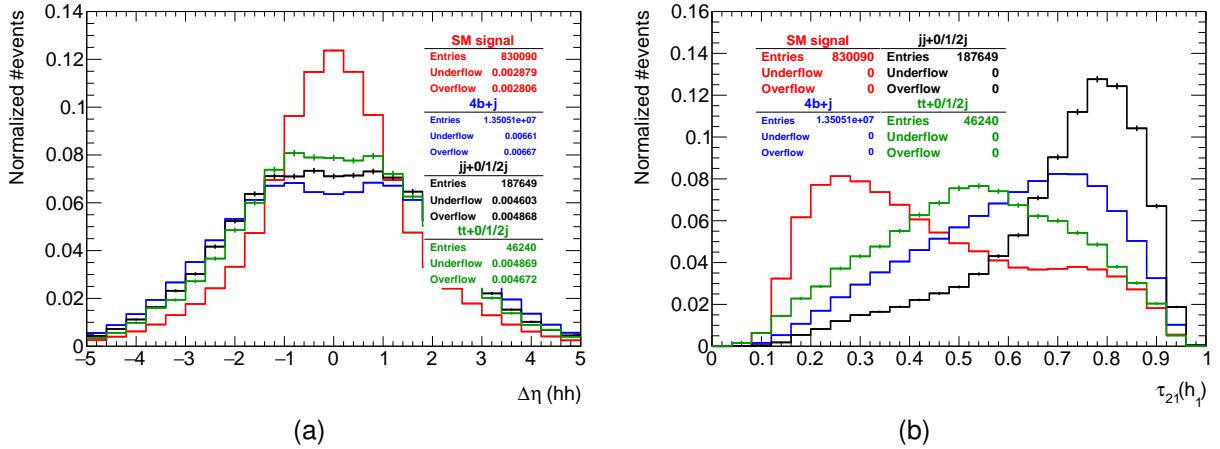


Figure 6.7: Distributions of the  $\Delta\eta$  between the Higgs candidates (left) and of the  $\tau_{21}$  variable for the leading Higgs candidate (right). The histograms are normalized to unit area.

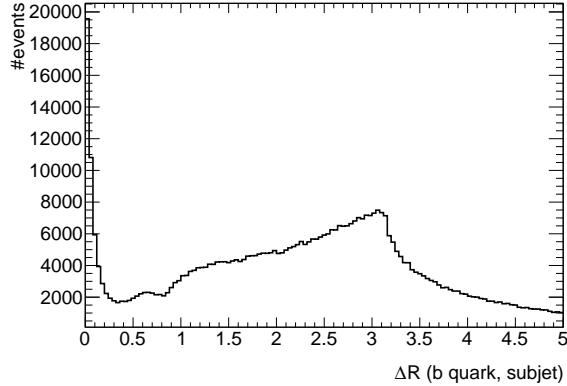


Figure 6.8: Minimum  $\Delta R$  between b quarks and subjets of the  $R = 0.8$  jets.

## 6.2 Analysis strategy

As already discussed, this analysis targets events in which at least two large- $R$  jets are reconstructed. The jet with the highest momentum is assumed to correspond to the leading Higgs candidate and the jet with the second highest momentum to the sub leading one. Both the leading and sub leading jets must be b-tagged in order for the event to be accepted.

The events are reconstructed using particle flow (or eflow) or pure calorimeter jets (we explored both approaches) with  $R = 0.8$ , clustered with the anti- $k_T$  algorithm. We perform the b-tagging of jets using truth level information as it is described in the following section. Jets with a large- $R$  parameter cannot be b-tagged using Delphes default algorithm because the tagging of large  $R$  jets is an ambiguous task that can be performed in several different ways. Therefore, we implemented our own b-tagging algorithm that is described in section 6.2.1.

In this section, we present the baseline analysis based on cuts on the kinematic and substructure variables. An optimized version of the analysis is also presented and compared to the baseline. The optimization is based on plots of the significance ( $S/\sqrt{B}$ ) as a function of the cut on a given variable. We also briefly describe how statistical uncertainties are taken into account.

### 6.2.1 Implementation of b-tagging

For each jet, the two hardest subjets are found using the mass drop procedure. It might happen that there are not two subjets because the algorithm's criteria are not met. In that case, the jet is rejected. We compute the  $\Delta R$  distance between all b and c quarks in the event with Pythia 8 status equal to 23 and with  $p_T > 10$  GeV and each subjet ( $\Delta R(\text{subjet,parton})$ ). According to the Pythia manual, particles with status 23 result directly from the hardest subprocess. We consider that a subjet is matched to a given quark if  $\Delta R(\text{subjet,parton}) < 0.3$ , as indicated by the plot in figure 6.8. If the subjet is matched to at least a b quark, we b-tag the subjet with a given probability. If the subjet is not matched to any b quark but it is matched to at least one c quark we apply a c mistag rate. If the subjet is not matched to any b or c quark we apply a light mis tag rate. The b-tag probability and mistag rates were obtained from the Delphes FCC-hh card. They depend on the momentum of the jet and on its  $\eta$  coordinate. They are

Table 6.1: b-Tagging (black), c (blue) and light (red) mistag probabilities as a function of  $\eta$  and  $p_T$  of the (sub)jet. The momentum dependent factor,  $(1 - p_T/15000)$ , is common to the three probabilities.

$\eta \backslash p_T$	$10 < p_T < 500$	$500 < p_T < 15000$
$ \eta  < 2.5$	0.85; <b>0.05</b> ; <b>0.01</b>	$(0.85; \textcolor{blue}{0.05}; \textcolor{red}{0.01}) \times (1 - p_T/15000)$
$2.5 <  \eta  < 4.0$	0.64; <b>0.03</b> ; <b>0.0075</b>	$(0.64; \textcolor{blue}{0.03}; \textcolor{red}{0.0075}) \times (1 - p_T/15000)$

1564 summarized in table 6.1. The b-tagging probabilities are given in black and the c and light mis tagging  
 1565 probabilities are given in blue and red, respectively. Note that a jet cannot be b-tagged if  $|\eta| > 4$  or if its  
 1566 momentum is smaller than 10 GeV or larger than 15000 GeV.

1567 In terms of the technical implementation, the b-tagging algorithm works as follows: for each subjet we  
 1568 look for a truth-level b quark within  $\Delta R = 0.3$  of the subjet and calculate the b-tagging and mis-tagging  
 1569 efficiencies using the expressions in table 6.1, where  $p_T$  and  $\eta$  refer to the subjet. If a b quark is found we  
 1570 generate a random number between 0 and 1. If it smaller than the b-tagging efficiency we consider the  
 1571 subjet to be b-tagged. If the subjet is not b-tagged we look for truth-level c quarks within  $\Delta R = 0.3$  of the  
 1572 subjet. If one is found we generate a new random number between 0 and 1 and if the number is smaller  
 1573 than the c mis-tag probability we consider the subjet to be b-tagged. If the subjet is not b-tagged we  
 1574 generate a random number between 0 and 1 and consider the subjet b-tagged if the number is smaller  
 1575 than the light mis-tag probability.

## 1576 6.2.2 Baseline analysis

1577 The analysis described in this section and later was developed using the sample simulated with the  
 1578 default FCC-hh detector implementation. The same analysis selection was then applied to the samples  
 1579 generated using the different detector configurations. This allows for a straightforward comparison of the  
 1580 results in terms of the significances.

As a first step, we implemented a baseline analysis based on rectangular cuts on kinematic and substructure variables. Firstly, we apply cuts on the transverse momenta of the leading and sub leading Higgs candidates,  $p_T(h_1)$  and  $p_T(h_2)$ , and of the Higgs pair,  $p_T(hh)$ :

$$1581 \quad p_T(h_1) > 400 \text{ GeV}, \quad p_T(h_2) > 350 \text{ GeV}, \quad p_T(hh) > 100 \text{ GeV}. \quad (6.1)$$

1582 These cuts follow from both distributions shown in figures 6.9(a), 6.9(b) and 6.10(a) and guarantee that  
 1583 we are choosing events for which the Higgs candidates are sufficiently boosted. In addition, a high  
 1584 threshold for the  $p_T$  of the jets suppresses  $t\bar{t}$  events because the decay products of the top quark are  
 1585 reconstructed in a single jet. In a more realistic analysis, these cuts would also be dictated by the trigger  
 threshold.

From figure 6.10(b) we then apply a cut on the  $\tau_{21}$  of the leading Higgs candidate to select jets that

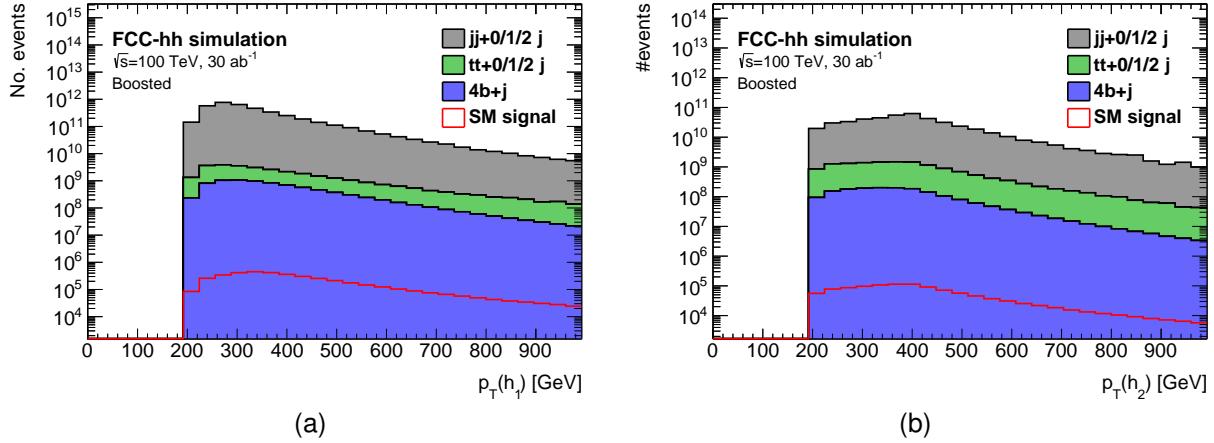


Figure 6.9:  $p_T$  distributions for the leading (a) and sub leading Higgs candidates (b). The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

are more compatible with a two-prong structure:

$$\tau_{21}(h_1) < 0.55. \quad (6.2)$$

The cut on this variable works as a Higgs tagging method. Therefore we apply the same cut on the  $\tau_{21}$  of the sub leading Higgs candidate although that does not necessarily follow from the distribution of  $\tau_{21}$  for the sub leading Higgs candidate (which can be found in figure F.1 in appendix F). Then, from the distribution in figure 6.11(a), we place a cut on the second Fox-Wolfram momentum (defined in section 3.3.1) of the leading Higgs candidate,  $H_2(h_1)$ :

$$H_2(h_1) > 0.2 \quad (6.3)$$

This substructure variable is particularly interesting because it helps suppress the  $t\bar{t}$  background.

Finally, based on the distribution shown in figure 6.11(b), we apply a cut on the softdrop mass of both Higgs candidates,  $M_{SD}(h_1, h_2)$ . This cut is placed in a window around the nominal SM Higgs mass:

$$(100 \leq M_{SD}(h_1, h_2) \leq 135) \text{ GeV.} \quad (6.4)$$

Using this analysis, we achieve a significance,  $S/\sqrt{B}$ , of  $5.1 \pm 0.5$  (stat.) $^{+2.6}_{-1.8}$  (sys.) (1.62 $\pm$ 0.16 (stat.) $^{+0.8}_{-0.6}$  (sys.))

for an integrated luminosity of 30 (3)  $\text{ab}^{-1}$ .

### 6.2.3 Optimization

The first approach to the optimization consists in placing successive cuts in the most relevant kinematic variables. The value of each cut is chosen in order to optimize the significance after that cut. As for the baseline analysis we start by looking at the transverse momenta of the leading and sub leading Higgs candidates. After the pre-selection cuts are applied, we scan the histograms of the  $p_T$  of the leading

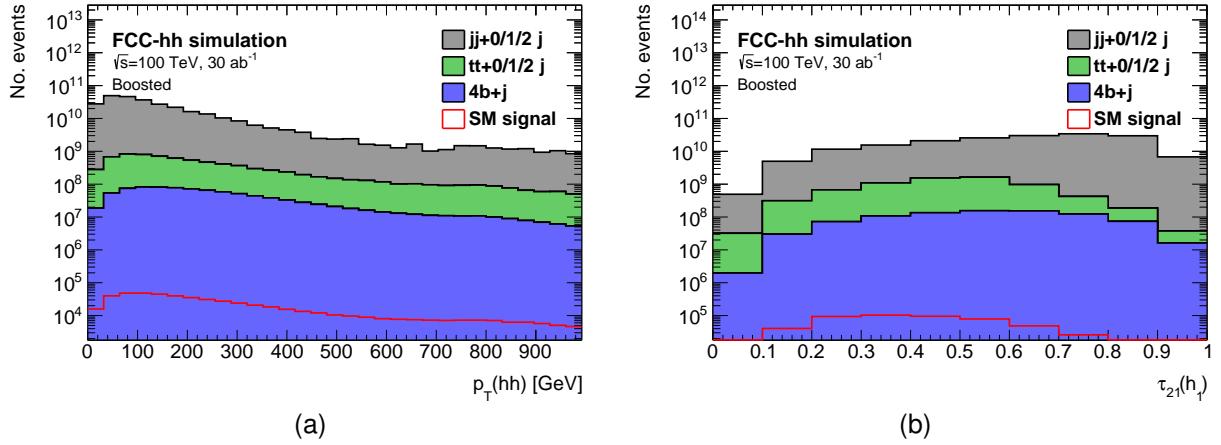


Figure 6.10: (a)  $p_T$  distribution for the Higgs pair, (b)  $\tau_{21}$  variable for the leading Higgs candidate. The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

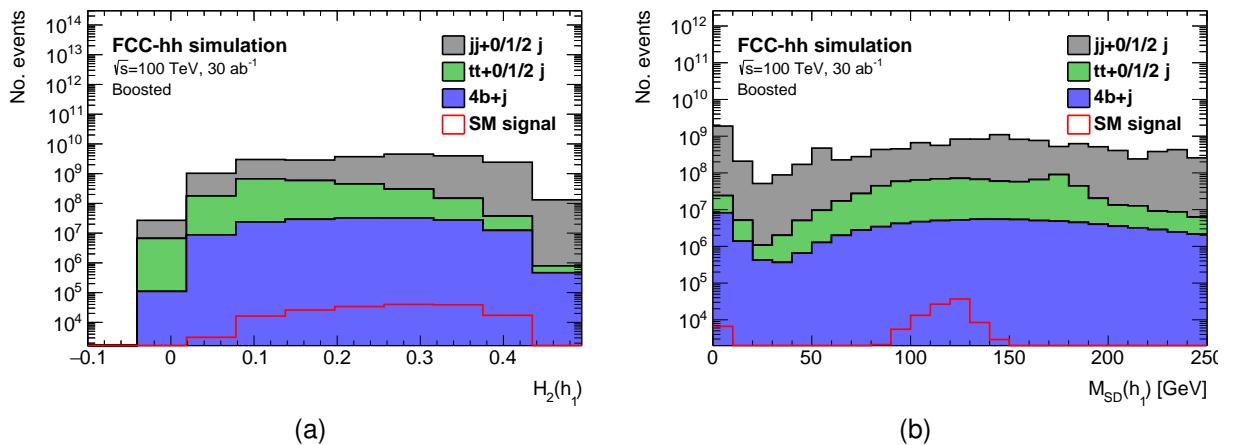


Figure 6.11: (a)  $H_2$  variable for the leading Higgs candidate, (b) softdrop mass distribution for the leading Higgs candidate. The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

Higgs candidate for signal and backgrounds by placing a lower cut on this variable. For each value of the cut we integrate upwards in order to obtain the expected number of signal and background events after the cut. Using these numbers we calculate the significance,  $S/\sqrt{B}$ . The significance as a function of the lower cut on  $p_T(h_1)$  is shown in figure 6.12(a). Based on this plot we choose the cut:

$$p_T(h_1) > 300 \text{ GeV}. \quad (6.5)$$

After placing the cut on  $p_T(h_1)$  we do the same plot for the  $p_T$  of the sub leading Higgs candidate and of the Higgs pair. These are shown in figures 6.12(b) and 6.13(a), respectively. From figure 6.12(b) we see that a cut on  $p_T(h_2)$  above 200 GeV is not favorable. Therefore, we do not apply any other cut on this variable. Based on the plot in figure 6.13(a) we choose the cut:

$$p_T(hh) > 100 \text{ GeV}. \quad (6.6)$$

Following the cuts on the momenta we place a cut on the  $\tau_{21}$  variable for both Higgs candidates. From all the variables considered during the optimization process this was the one that lead to the highest increase in the significance. From the definition of the  $\tau_{21}$  variable we expect the signal to take lower values than the background. Therefore we optimize the cut on this variable by placing an upper cut and integrating the distribution backwards. The plot is shown in figure 6.13(b). From this plot we place the following cut on the  $\tau_{21}$  of the leading Higgs candidate:

$$\tau_{21}(h_1) < 0.4. \quad (6.7)$$

1590 We apply exactly the same cut on the  $\tau_{21}$  of the sub leading Higgs candidate. This guarantees that both  
 1591 jets are consistent with having two subjets and therefore are more likely to originate from the decay of a  
 1592 Higgs boson.

Next we apply cuts on the  $\Delta\eta$  between the Higgs candidates and on the second Fox-Wolfram momentum of the leading Higgs candidates, in this order:

$$|\Delta\eta(hh)| < 1.5 \quad H_2(h_1) > 0.2. \quad (6.8)$$

1593 These cuts follow from the plots on figures 6.14(a) and 6.14(b). The optimization plot for  $\Delta\eta(hh)$  is  
 1594 obtained placing a cut in a window around zero which follows directly from the shape of the distributions  
 1595 in figure 6.7(a). For the  $H_2(h_1)$  variable the cut is placed on the lower value.

1596 Finally we apply the mass cuts on the leading and sub leading Higgs candidates. These are the  
 1597 same that were applied in the baseline analysis.

1598 Using the optimized analysis we obtain a significance,  $S/\sqrt{B}$ , of  $8.8 \pm 1.6$  (stat.)  $^{+4.4}_{-3.4}$  (sys.)  $(2.36 \pm$   
 1599  $0.32$  (stat.)  $^{+1.4}_{-1.1}$  (sys.)) for an integrated luminosity of  $30$  ( $3$ )  $\text{ab}^{-1}$ . This corresponds to an improvement of  
 1600 approximately 70% with respect to the baseline analysis. This percentage is calculated as  $((S/\sqrt{B})_{\text{opt.}} -$   
 1601  $(S/\sqrt{B})_{\text{base.}})/(S/\sqrt{B})_{\text{base.}}) \times 100$ .

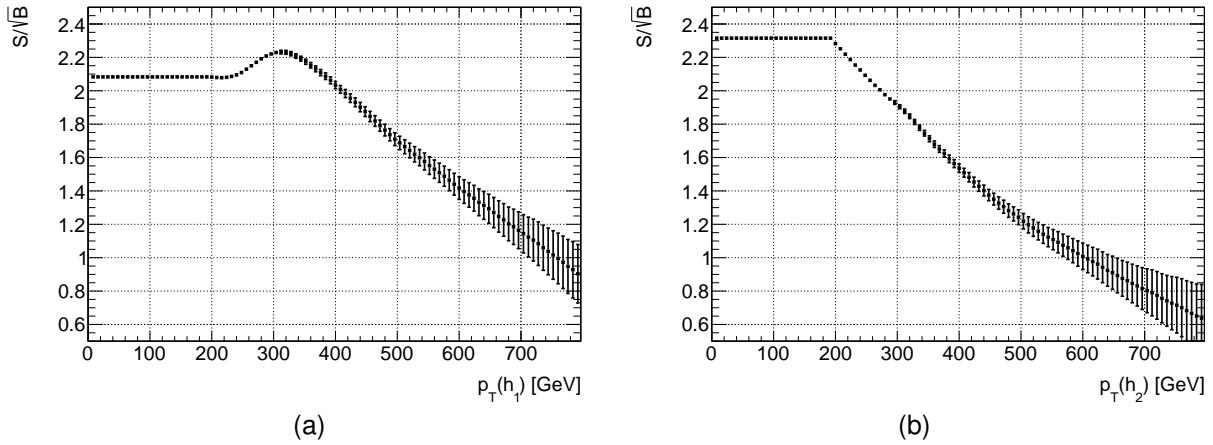


Figure 6.12:  $S/\sqrt{B}$  as a function of the cut on the  $p_T$  of the leading Higgs candidate,  $p_T(h_1)$  (a) after the pre-selection cuts and of the sub leading Higgs candidate,  $p_T(h_2)$  (b) after the cut on  $p_T(h_1)$ .

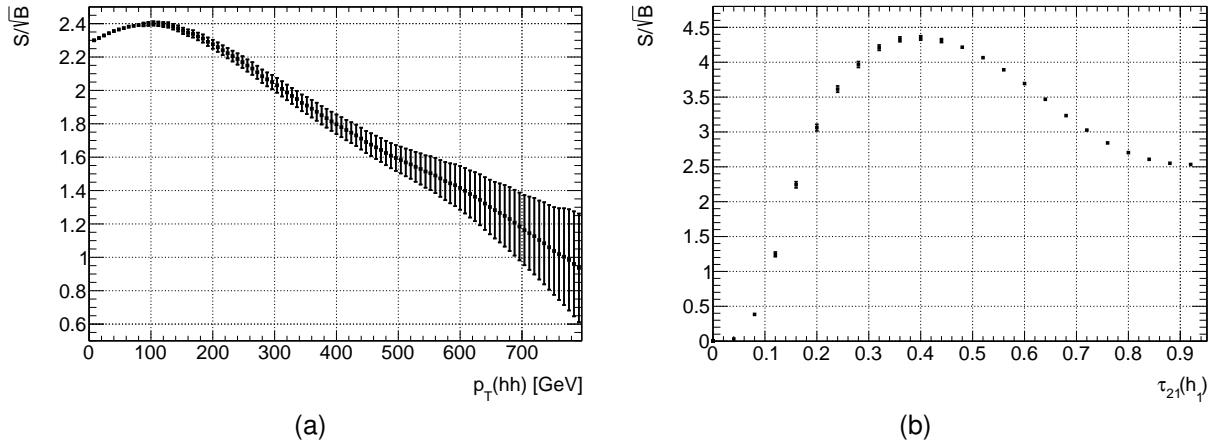


Figure 6.13:  $S/\sqrt{B}$  as a function of the cut on the  $p_T$  of the Higgs pair,  $p_T(hh)$  (a) after the cut on  $p_T(h_1)$  and as a function of the  $\tau_{21}$  variable for leading Higgs candidate,  $\tau_{21}(h_1)$  (b) after the cut on  $p_T(hh)$ .

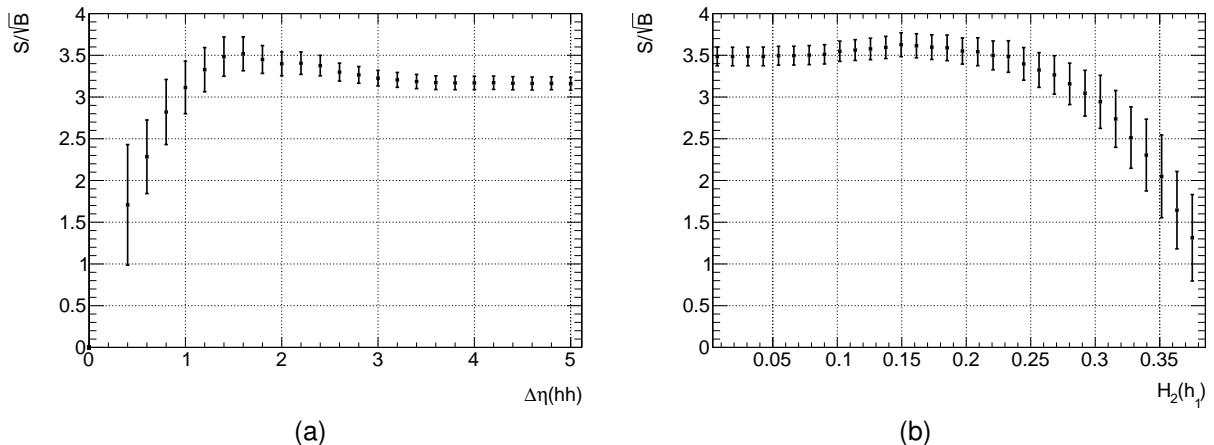


Figure 6.14:  $S/\sqrt{B}$  as a function of the cut on the  $\Delta\eta$  between the Higgs candidates,  $\Delta\eta(hh)$  (a) after the cuts on the transverse momenta and on the  $\tau_{21}$  variables and as a function of the second Fox-Wolfram momentum for the leading Higgs candidate,  $H_2(h_1)$  (b) after the cuts on the transverse momentums,  $\tau_{21}$  and  $\Delta\eta$ .

## 1602 6.2.4 Handling of uncertainties

### 1603 Statistical uncertainties

For challenging analyses such as this one, the backgrounds usually have a much larger cross section than the signal meaning that we need a lot of Monte Carlo events to properly simulate them. Most of the times it is not feasible to generate as much events as we would need for a given luminosity. This is the case with this work, namely because we are targeting very high luminosities (of the order of tens of  $\text{ab}^{-1}$ ). Take for example the multijet sample with  $500 < H_T < 1000$ . The cross section is approximately  $10^7 \text{ pb}$  which means that we expect a total of  $3 \times 10^{14}$  events for an integrated luminosity of  $30 \text{ ab}^{-1}$ . It is not feasible, within the time frame of this work and with the available computational resources, to generate this number of events. Therefore, we apply a weight,  $w$ , to each event given by:

$$w = \frac{\mathcal{L} \times \sigma}{N} \quad (6.9)$$

1604 where  $\mathcal{L}$  is the target integrated luminosity,  $\sigma$  is the cross section of the sample and  $N$  is the number of  
1605 MC events generated.

If we assume that the number of events follows a Poisson distribution then the standard deviation (or uncertainty) is given by  $\sqrt{N}$  where  $N$  is the mean number of events. In order to normalize the number of events to a given luminosity we need to multiply the mean value and the uncertainty by the respective weight such that the number of normalized events, including the uncertainty,  $N_{\text{norm}} \pm \Delta N_{\text{norm}}$ , is given by:

$$N_{\text{norm}} \pm \Delta N_{\text{norm}} = w \times (N \pm \sqrt{N}). \quad (6.10)$$

$\Delta N_{\text{norm}}$  can then be used in standard error propagation to compute the statistical uncertainty associated with any expression, in particular, with  $S/\sqrt{B}$ . The error associated with  $S/\sqrt{B}$  is given by:

$$\left( \Delta \frac{S}{\sqrt{B}} \right)_{\text{stat.}}^2 = \left| \frac{1}{\sqrt{B}} \right|^2 (\Delta S)^2 + \left| -\frac{S}{B^{3/2}} \right|^2 (\Delta B)^2 \quad (6.11)$$

where  $\Delta S$  and  $\Delta B$  are the uncertainties associated with the number of signal and background events.  $\Delta B$  is given by:

$$(\Delta B)^2 = \sum_i (\Delta B_i)^2 \quad (6.12)$$

1606 where the index  $i$  runs over all independent backgrounds.

### 1607 Systematic uncertainties

1608 Systematic uncertainties are usually related to the detector specificities and object reconstruction tech-  
1609 niques, which makes them difficult to list and quantify for a detector that is not yet built. Nonetheless,  
1610 we can use the numbers for the systematic uncertainties that are reported for existing searches for  
1611  $pp \rightarrow hh \rightarrow b\bar{b}b\bar{b}$  using the ATLAS detector.

1612 When requiring four b-tags, the largest uncertainty affecting the signal is the the b-tagging uncer-

1613 tainty (table 6 of Ref. [100]). It is of the order of 30%. We use this number to estimate the impact of the  
 1614 b-tagging uncertainty on the analysis. For the background, the largest uncertainty comes from the esti-  
 1615 mation of its yield and shape, according to the same reference. The reported value is 16%. However, a  
 1616 study performed by the FCC study group [101] considers more conservative values for the uncertainties  
 1617 on the  $t\bar{t}$  and QCD di-jet backgrounds normalizations. The reported values are 20% for  $t\bar{t}$  and 50% for di-  
 1618 jet production. In this study, we use these numbers to estimate the impact of the systematic uncertainty  
 1619 on the normalization of the backgrounds.

1620 To estimated the b-tagging systematic uncertainty, the number of signal events after all the analysis  
 1621 cuts is varied by 30% up and down. This leads to an upper (lower) bound in the significance of 6.6 (3.6).

1622 The background normalization uncertainties are implemented by increasing and decreasing the k-  
 1623 factor of each background by the corresponding percentage, with respect to what is given in table 5.1.  
 1624 We vary the k-factors of the QCD multijet ( $4b + j$  and  $jj + 0/1/2 j$ ) and of the  $t\bar{t}$  backgrounds individually.  
 1625 For the  $4b + j$  and  $jj + 0/1/2 j$  backgrounds we consider the same uncertainty of 50% and the variations  
 1626 are done simultaneously for both. This leads to an upper (lower) bound in the significance of 7.1 (4.2).  
 1627 For the  $t\bar{t}$  background we consider an uncertainty of 20% which leads to an upper (lower) bound in the  
 1628 significance of 5.13 (5.09).

We assume that the three sources of systematic uncertainty are uncorrelated and therefore the respective variations can be added in quadrature. The statistical error is also added in quadrature. Therefore, the total uncertainty that is given by:

$$\left( \Delta \frac{S}{\sqrt{B}} \right)_{\text{sys.+stat.}}^2 = (\delta_{\text{b-tag}})^2 + (\delta_{\text{QCD bkg.}})^2 + (\delta_{t\bar{t} \text{ bkg.}})^2 + (\delta_{\text{stat.}})^2 \quad (6.13)$$

1629 where  $\delta_{\text{b-tag}}$ ,  $\delta_{\text{QCD bkg.}}$  and  $\delta_{t\bar{t} \text{ bkg.}}$  are the variations with respect to the nominal value of the significance  
 1630 caused by the b-tagging uncertainty and by the uncertainties on the normalization of the QCD multijet  
 1631 and  $t\bar{t}$  backgrounds, respectively.  $\delta_{\text{stat.}}$  is the statistical uncertainty.

## 1632 6.2.5 Trigger discussion

1633 In the previous discussions regarding the baseline and optimized analysis no trigger requirements were  
 1634 taken into account. We now review the trigger requirements that had to be fulfilled in the analysis  
 1635 performed by ATLAS and discuss where do the analyses we designed stand with respect to those  
 1636 requirements.

1637 In Ref. [100], the event selection requires the existence of at least two anti- $k_T$   $R = 1.0$  jets with  
 1638  $p_T > 250$  GeV,  $|\eta| < 2.0$  and  $m_J > 50$  GeV. In addition, the leading jet must have  $p_T > 450$  GeV to  
 1639 guarantee 100% trigger efficiency.

1640 We require at least two anti- $k_T$   $R = 0.8$  jets with  $p_T > 200$  GeV. In the baseline analysis, we require  
 1641 that the leading jet has  $p_T > 400$  GeV. These  $p_T$  thresholds are compatible to the ones used in the  
 1642 ATLAS analysis mainly when we consider that the radius of the jets is slightly smaller and therefore the  
 1643  $p_T$  threshold can also be lower. We do not consider any cuts on the pseudorapidity or mass of the jets.  
 1644 However, we do require that the sub leading jet has  $p_T > 300$  GeV and that the Higgs pair has  $p_T > 100$

1645 GeV which would help lower the rate.

1646 In the optimized analysis, however, the  $p_T$  cuts are significantly looser: the leading jet must have  
1647  $p_T > 300$  GeV and the Higgs pair must have  $p_T > 100$  GeV. This could indicate that although the  
1648 optimized analysis gives a better significance, tighter cuts might have to be applied in order to fulfill the  
1649 trigger requirements, therefore reducing the significance.

1650

# Chapter 7

1651

## Results

1652 In this chapter we describe the main results of the search for  $hh \rightarrow b\bar{b}b\bar{b}$  at the FCC-hh using two  
1653 benchmark luminosities,  $30 \text{ ab}^{-1}$  and  $3 \text{ ab}^{-1}$  (section 7.1). In section 7.2 we show how the significance  
1654 of the analysis varies as a function of the granularity of the HCAL and/or the detector configuration. We  
1655 also compare the results obtained using particle flow and pure calorimeter jets.

1656

### 7.1 Di-Higgs discovery potential at the FCC-hh

1657 The event selection of the optimized analysis for the search for  $hh \rightarrow b\bar{b}b\bar{b}$  at the FCC with the baseline  
1658 detector design is summarized in tables 7.1 and 7.2 for the signal samples (SM, DM mediator and type  
1659 II 2HDM) and for the background samples ( $4b + j$ ,  $jj + 0/1/2j$  and  $t\bar{t}$ ), respectively.

1660 From table 7.1, we see that for the BSM models the signal efficiency is higher than for the SM. It  
1661 varies from 0.446 for the SM, to 0.487 for the DM mediator model and to 1.666 for the type II 2HDM. This  
1662 is due to the existence of heavy new particles in the BSM models that lead to the production of highly  
1663 boosted SM Higgs pairs.

Considering the SM production of Higgs pairs, the achieved significance is

$$S/\sqrt{B} = 8.8 \pm 1.6 \text{ (stat.)} {}^{+4.4}_{-3.4} \text{ (sys.)} \quad (2.36 \pm 0.32 \text{ (stat.)} {}^{+1.4}_{-1.1} \text{ (sys.)}) \quad (7.1)$$

1664 for an integrated luminosity of  $30 (3) \text{ ab}^{-1}$ . For  $\mathcal{L} = 30 \text{ ab}^{-1}$ , the significance is above the  $5\sigma$  threshold.  
1665 This result indicates that with the entire dataset that is expected to be accumulated by the FCC-hh  
1666 detector it should be possible to observe the production of Higgs pairs.

1667 For the signal model that includes a 1 TeV dark matter mediator that can decay to pairs of SM  
1668 Higgs bosons the achieve significance is  $2.3 \pm 0.4 \text{ (stat.)} {}^{+1.2}_{-0.9} \text{ (sys.)} \quad (0.73 \pm 0.13 \text{ (stat.)} {}^{+0.37}_{-0.28} \text{ (sys.)})$  for an  
1669 integrated luminosity of  $30 (3) \text{ ab}^{-1}$ . The significance is well bellow the  $3\sigma$  threshold for both luminosities.  
1670 Therefore, we do not expect to be able to detect or exclude this signal at the FCC-hh. In this model, the  
1671 coupling of the DM mediator to the Higgs pairs is small which means that the contribution from the box  
1672 diagram dominates over the resonant production (s-channel diagram), just like in the SM. In addition,

Table 7.1: Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM and 2HDM). The absolute value of expected events after some key selection cuts is shown in curved brackets. The number of expected events is normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC-hh baseline detector design, as implemented in Delphes by the FCC-hh study group.

Selection [FCC-hh]	SM	DM mediator	2HDM type II
Gen level	100 $(2704 \pm 12) \times 10^3$	100 $(65400 \pm 29) \times 10^2$	100 $(13977 \pm 7) \times 10^3$
$N(\text{b-tags}) \geq 4$	92.488	92.593	93.430
$p_T(j_1, j_2) \geq 200 \text{ GeV}$	16.602 $(4490 \pm 5) \times 10^3$	17.033 $(11140 \pm 12) \times 10^2$	33.975 $4734 \pm 4 \times 10^3$
<hr/>			
$p_T(j_1) \geq 300 \text{ GeV}$	13.521	14.007	20.869
$p_T(j_1 + j_2) \geq 100 \text{ GeV}$	10.933	11.301	22.863
$\tau_{21}(j_1, j_2) < 0.4$	1.309	1.410	3.952
$ \Delta\eta(hh)  < 1.5$	1.071	1.154	3.479
$FW2(j_1) > 0.2$	0.989	1.064	3.276
$(100 < M_{SD}(j_1, j_2) < 135) \text{ GeV}$	0.446 $(1207 \pm 8) \times 10^2$	0.487 $(3188 \pm 20) \times 10$	1.666 $(2328 \pm 8) \times 10^2$

<sup>1673</sup> the cross section is smaller than for the SM process which leads to a smaller significance even though  
<sup>1674</sup> the efficiency is higher.

For the type II 2HDM the achieved significance is

$$S/\sqrt{B} = 16.9 \pm 3.0 \text{ (stat.)} {}^{+8.5}_{-6.6} \text{ (sys.)} \quad (5.4 \pm 0.9 \text{ (stat.)} {}^{+2.7}_{-2.1} \text{ (sys.)}) \quad (7.2)$$

<sup>1675</sup> for an integrated luminosity of  $30 (3) \text{ ab}^{-1}$ . These results indicate that the contribution of this model to  
<sup>1676</sup> the production of Higgs pairs be observed (or excluded) at the FCC-hh. The high efficiency of this signal  
<sup>1677</sup> sample through the cuts, reflected in the high significance that is achieved, make it a very exciting and  
<sup>1678</sup> achievable benchmark.

### <sup>1679</sup> 7.1.1 Comparing with the ATLAS detector

<sup>1680</sup> The event selection of the search for  $hh \rightarrow b\bar{b}b\bar{b}$  with the ATLAS detector at a CM energy of  $\sqrt{s} = 100$   
<sup>1681</sup> TeV is summarized in tables 7.3 and 7.4 for the signal and background samples, respectively.

<sup>1682</sup> It is interesting to compare the results obtained with the FCC-hh default detector simulation with the  
<sup>1683</sup> ones obtained using the simulation of the ATLAS detector. These are summarized in table 7.5 in terms  
<sup>1684</sup> of the achieved significance for an integrated luminosity of  $30 \text{ ab}^{-1}$ , for the different signal models.

<sup>1685</sup> For all the signal models, the significance increases approximately 50% going from the ATLAS de-  
<sup>1686</sup> tector to the FCC-hh. This percentage is computed as  $((S/\sqrt{B})_{\text{FCC}} - (S/\sqrt{B})_{\text{ATLAS}}) / (S/\sqrt{B})_{\text{ATLAS}}$ . For  
<sup>1687</sup> the SM signal, nonetheless, using the ATLAS default detector configuration the achieved significance

Table 7.2: Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b + j$ ,  $jj + 0/1/2j$  and  $t\bar{t}+0/1/2 j$ ). The absolute value of expected events after some key selection cuts is shown in curved brackets. The number of expected events is normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC-hh baseline detector design, as implemented in Delphes by the FCC-hh study group.

Selection [FCC-hh]	$4b + j$	$jj + 0/1/2j$	$t\bar{t}$
Gen level	100 $(49035 \pm 12) \times 10^6$	100 $(54698 \pm 13) \times 10^{10}$	100 $(22503 \pm 11) \times 10^8$
$N(\text{b-tags}) \geq 4$	75.819	3.963	53.495
$p_T(j_1, j_2) \geq 200 \text{ GeV}$	17.811 $(8734 \pm 5) \times 10^6$	0.742 $(4058 \pm 11) \times 10^9$	1.056 $(2377 \pm 11) \times 10^7$
<hr/>			
$p_T(j_1) \geq 300 \text{ GeV}$	12.744	0.422	0.718
$p_T(j_1 + j_2) \geq 100 \text{ GeV}$	10.901	0.245	0.617
$\tau_{21}(j_1, j_2) < 0.4$	0.256	0.002	0.037
$ \Delta\eta(hh)  < 1.5$	0.130	0.001	0.024
$FW2(j_1) > 0.2$	0.105	0.001	0.014
<hr/>			
$(100 < M_{SD}(j_1, j_2) < 135) \text{ GeV}$	0.007 $(341 \pm 10) \times 10^4$	0.00003 $(17 \pm 11) \times 10^7$	0.0007 $(149 \pm 28) \times 10^5$

is already above  $5\sigma$ , for  $\mathcal{L} = 30 \text{ ab}^{-1}$ :  $S/\sqrt{B} = 5.7 \pm 1.3 \text{ (stat.)} {}^{+3.1}_{-2.4} \text{ (sys.)}$ . This indicates that the CM energy is the key factor determining the discovery potential of the accelerator.

## 7.2 Hadronic calorimeter granularity studies for future colliders

In this section we present the results that allow us to compare the different detector configurations. In section 7.2.1, we compare the resolution of the jet mass and of the  $\tau_{21}$  variable for the different detector configurations. In section 7.2.2, we show the results obtained with the baseline and optimized analyses. In the optimized analyses, the statistical fluctuations are higher because the  $jj + 0/1/2 j$  background with  $500 < H_T < 1000$ , which has the largest statistical weight, is not rejected as efficiency. Therefore, the shapes of the plots are easier to discern for the baseline analyses.

The cutflow tables for the baseline and optimized analysis for all detector configurations, for all signal models and using particle flow and calorimeter can be found in appendix E. In addition, the significances achieved for all possible combinations are summarized in tables E.1, E.2 and E.3.

### 7.2.1 Resolution of jet mass and N-subjetiness

The softdrop mass of the leading Higgs candidate for the SM signal sample is shown in figure 7.1(a) for the different detector configurations. The same plot is shown in figure 7.1(b) for calorimeter jets. It can be seen that the mass resolution increases as we increase the granularity, i.e, the mass peaks becomes

Table 7.3: Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM and 2HDM). The absolute value of expected events after some key selection cuts is shown in curved brackets. The number of expected events is normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS detector design, as implemented in Delphes.

<b>Selection [ATLAS]</b>	<b>SM</b>	<b>DM mediator</b>	<b>2HDM type II</b>
Gen level	100 $(27043 \pm 12) \times 10^3$	100 $(65400 \pm 29) \times 10^2$	100 $(13978 \pm 6) \times 10^3$
$N(\text{b-tags}) \geq 4$	88.690	88.787	89.643
$p_T(j_1, j_2) \geq 200 \text{ GeV}$	15.533 $(4201 \pm 5) \times 10^3$	15.941 32.181 $(10426 \pm 12) \times 10^2$	$(4498 \pm 4) \times 10^3$
<hr/>			
$p_T(j_1) \geq 300 \text{ GeV}$	12.599	13.061	29.141
$p_T(j_1 + j_2) \geq 100 \text{ GeV}$	10.185	10.526	21.523
$\tau_{21}(j_1, j_2) < 0.4$	1.139	1.220	3.411
$ \Delta\eta(hh)  < 1.5$	0.891	0.960	2.930
$FW2(j_1) > 0.2$	0.796	858	2.684
$(100 < M_{SD}(j_1, j_2) < 135) \text{ GeV}$	0.333 $(901 \pm 7) \times 10^2$	0.360 $(2358 \pm 18) \times 10$	1.266 $(1770 \pm 7) \times 10^2$

Table 7.4: Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b + j$ ,  $jj + 0/1/2j$  and  $t\bar{t}+0/1/2 j$ ). The absolute value of expected events after some key selection cuts is shown in curved brackets. The number of expected events is normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ . The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS detector design, as implemented in Delphes.

<b>Selection [ATLAS]</b>	<b><math>4b + j</math></b>	<b><math>jj + 0/1/2j</math></b>	<b><math>t\bar{t}</math></b>
Gen level	100 $(49035 \pm 15) \times 10^6$	100 $(54698 \pm 15) \times 10^{10}$	100 $(22503 \pm 9) \times 10^8$
$N(\text{b-tags}) \geq 4$	71.617	3.747	51.782
$p_T(j_1, j_2) \geq 200 \text{ GeV}$	16.299 $(7993 \pm 6) \times 10^6$	0.685 $(379 \pm 11) \times 10^9$	0.985 $(2215 \pm 9) \times 10^7$
<hr/>			
$p_T(j_1) \geq 300 \text{ GeV}$	11.627	0.390	0.669
$p_T(j_1 + j_2) \geq 100 \text{ GeV}$	9.932	0.227	0.574
$\tau_{21}(j_1, j_2) < 0.4$	0.234	0.003	0.031
$ \Delta\eta(hh)  < 1.5$	0.113	0.001	0.019
$FW2(j_1) > 0.2$	0.082	0.001	0.010
$(100 < M_{SD}(j_1, j_2) < 135) \text{ GeV}$	0.005 $(253 \pm 11) \times 10^4$	0.00004 $(23 \pm 15) \times 10^7$	0.0007 $(160 \pm 24) \times 10^5$

Table 7.5: Significances achieved with the ATLAS and FCC-hh detector configurations for  $\mathcal{L} = 30 \text{ ab}^{-1}$  for the three benchmark signal models: SM, 1 TeV DM mediator and 2HDM with  $m_H = 900 \text{ GeV}$ .

Signal sample	ATLAS	FCC-hh
SM	$5.7 \pm 1.3 \text{ (stat.)}$	$8.8 \pm 0.7 \text{ (stat.)}$
1 TeV DM mediator	$1.50 \pm 0.23 \text{ (stat.)}$	$2.3 \pm 0.4 \text{ (stat.)}$
2HDM type II	$11.3 \pm 1.7 \text{ (stat.)}$	$16.9 \pm 3.0 \text{ (stat.)}$

narrower. This effect is more pronounced when using pure calorimeter jets. In order to quantify this effect we use the full width at half maximum (FWHM). For particle flow (calorimeter) jets, the FWHM varies from 126 (130) for the ATLAS detector configuration to 124 (122) for the FCC granularity with  $\eta, \phi \times 2$  which corresponds to a 1% (6%) effect.

The invariant mass of the Higgs pair is shown in figure 7.2 for the SM signal when using particle flow jets. The same distribution is shown for the DM mediator and 2HDM signals in figures 7.3(a) and 7.3(b), respectively. For the SM, the mass spectrum of the Higgs pair does not show any resonance. On the one hand, the production is dominated by the box diagram. On the other hand, even if we only had the contribution from the triangle diagram, the Higgs that decays to a pair of Higgs bosons has to be off shell, therefore we expect a broad spectrum. For the DM mediator model, there is an enhancement with respect to the SM for  $M(hh) \sim 1 \text{ TeV}$ . This is due to the contribution of the 1 TeV DM mediator decaying to pairs of SM Higgs bosons. Nonetheless, the spectrum is still quite broad because the box diagram stills gives a very large contribution. For the 2HDM, there is a clear peak at  $M(hh) \sim 900 \text{ GeV}$  because resonant production dominates. The mass resolution of the resonances does not seem to increase as the granularity increases. This picture holds even when using HCAL jets.

The  $\tau_{21}$  variable for the leading Higgs candidate for the SM signal (filled lines) and for the  $4b + j$  background (dashed lines) is shown in figure 7.4(a) for the different detector configurations for particle flow jets. The same plot is show in figure 7.4(b) for calorimeter jets. The separation between signal and background increases as the granularity increases. This was expected because an increase in the granularity of the hadronic calorimeter should help resolve better the substructure of boosted jets. For particle flow jets, the distance between the maximum points of the two distributions varies from 0.4 for the ATLAS detector, ATLAS HCAL and ATLAS HCAL with  $\eta \times 4$  to 0.44 for the FCC HCAL with  $\phi/2$  and to 0.46 for the remaining configurations. For calorimeter jets the distance between the maximum is always smaller than when using particle flow jets, considering the same detector configuration. Therefore, the separation is always better when using particle flow jets. However, when using calorimeter jets, there is a larger difference in the shape of the distributions for the multiple detector configurations, particularly when going from the ATLAS granularity with  $\eta \times 4$  to the FCC granularity with  $\phi/2$ .

The separation is the largest between the signal and the  $jj + 0/1/2 j$  background, as expected. For this background, the distance between the maximums of the distributions varies from 0.48 (24) to 0.56 (0.48), for particle flow (calorimeter) jets. These distributions can be found in figures ... in appendix F.

1735 **7.2.2 Signal efficiency and significance**

1736 Figure 7.5(a) shows the signal efficiency for the three signal models: SM (filled squares), 1 TeV DM  
1737 mediator (empty squares) and type II 2HDM with  $m_H = 900$  GeV, for eflow jets. The same plot is  
1738 shown in figure 7.5(b) for calorimeter jets. For eflow jets the efficiency increases as we increase the  
1739 granularity, for all signal models. It varies approximately 30% between the lowest value (ATLAS default)  
1740 and the highest value (FCC granularity with  $\eta, \phi \times 2$ ). For HCAL jets, this tendency also exists with  
1741 the exception of the first point which corresponds to the ATLAS default detector configuration. [WHY?  
1742 OPTIMIZATION OF ATLAS?]. In addition, the efficiency changes more when using pure calorimeter jets.  
1743 It varies by approximately 98% between the ATLAS detector configuration and the configuration with the  
1744 FCC granularity and  $\eta, \phi \times 2$ . The efficiency is higher for both BSM models than for the SM. This is  
1745 because these models were chosen to have very heavy particles decaying to a pair of highly boosted  
1746 Higgs pairs.

1747 The significances achieved with the baseline analysis as a function of the detector configuration  
1748 for the SM signal, the DM mediator model and the type II 2HDM are shown in figures 7.6(a), 7.7(a)  
1749 and 7.8(a), respectively. The same plots are shown for the optimized analysis in figures 7.6(b), 7.7(b)  
1750 and 7.8(b). The square markers refer to analysis performed using particle flow jets. The statistical  
1751 uncertainty associated with each value of the significance is computed using standard error propagation  
1752 and is shown as error bars. The grey blocks represent the total uncertainty. For the baseline analysis,  
1753 regardless of the signal model considered, the significance changes by approximately 15% when going  
1754 from the ATLAS detector configuration to the configuration with the FCC granularity increased by a factor  
1755 of two in  $\eta$  and  $\phi$ . For the optimized analysis the change is of approximately 54%.

1756 Motivated by the small change in significance over the range of configurations that were tested,  
1757 we implemented exactly the same analysis but using HCAL jets instead of eflow jets. The results are  
1758 shown in the same plots using triangular markers and with green error blocks. On the one hand, the  
1759 achieved significance is always smaller when using HCAL jets because we are not making use of the  
1760 tracking information. On the other hand, when using HCAL jets, the significance changes more over  
1761 the configuration range. For the SM signal, it varies by approximately 80% when going from the ATLAS  
1762 detector configuration to the configuration with the FCC granularity increased by a factor of two in  $\eta$  and  
1763  $\phi$ . For the DM mediator model, this value is 77% and for the 2HDM it is 60%. These percentages refer  
1764 to the baseline analysis. For the optimized analysis the change is between 60% and 70% depending on  
1765 the signal model.

1766 All in all, given the large uncertainties, it is hard to conclude on the tendency followed by the plots  
1767 of the significance as a function of the detector configuration. Nonetheless, the significance does seem  
1768 to increase as the granularity increases, which what was expected. As previously discussed, this effect  
1769 is more accentuated when using pure calorimeter jets because we are only making use of the informa-  
1770 tion provided by the hadronic calorimeter and therefore the analysis is more sensitive to its transversal  
1771 segmentation (granularity).

1772 The small change in the significance when using eflow jets and the fact that the change increases  
1773 when using HCAL jets indicate that in the FCC-hh baseline detector design the resolution of the tracking

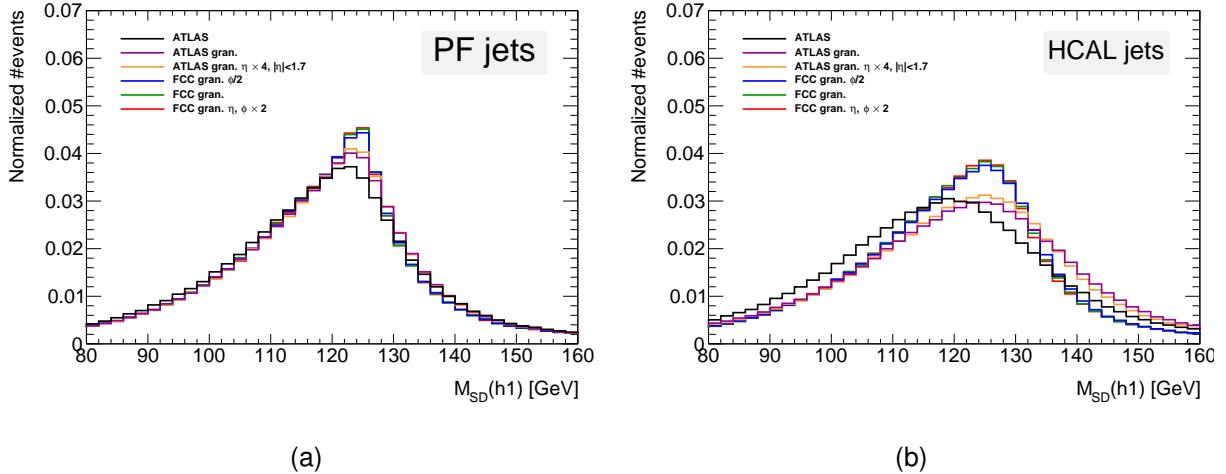


Figure 7.1: Leading Higgs candidate softdrop mass after the pre-selection cuts for particle flow (a) and calorimeter (b) jets. The colors indicate the different detector configurations. The x axis range is from 80 GeV to 160 GeV in order to make the differences between the histograms more clear.

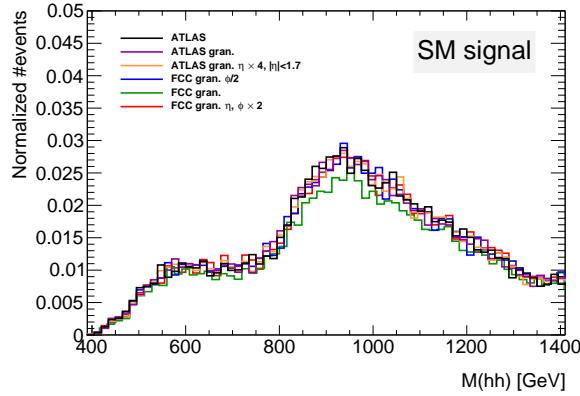


Figure 7.2: Invariant mass of the Higgs pair for the SM signal. The plots include the events that pass all the analysis cuts. The colors indicate the different detector configurations.

<sup>1774</sup> system is so good, in particular, so much better than the HCAL resolution that it becomes the limiting  
<sup>1775</sup> factor.

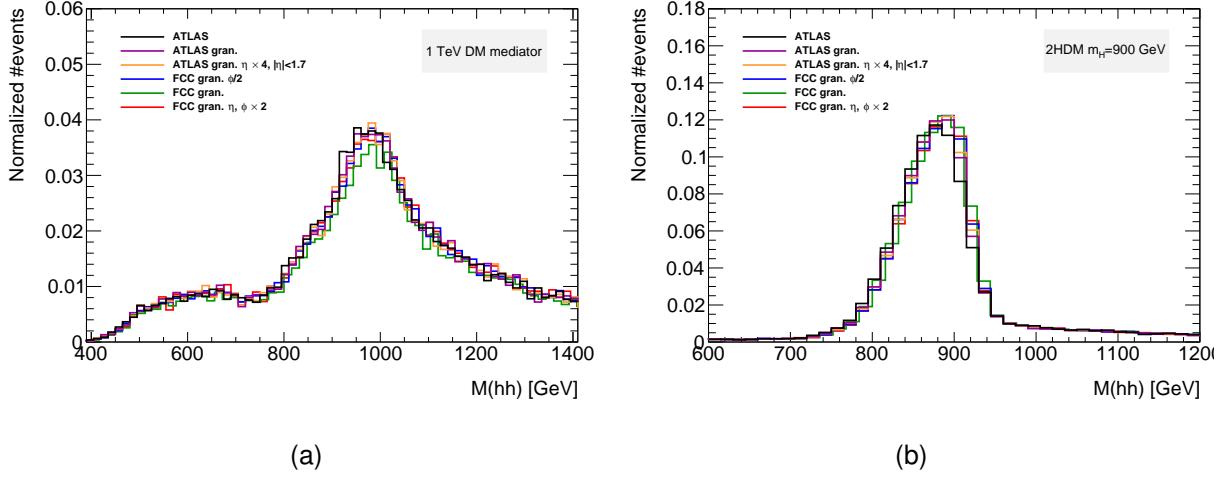


Figure 7.3: (a) Invariant mass of the Higgs pair for the 1 TeV DM mediator, (b) Invariant mass of the Higgs pair for the 2HDM with  $m_H = 900$  GeV. The plots include the events that pass all the analysis cuts. The colors indicate the different detector configurations.

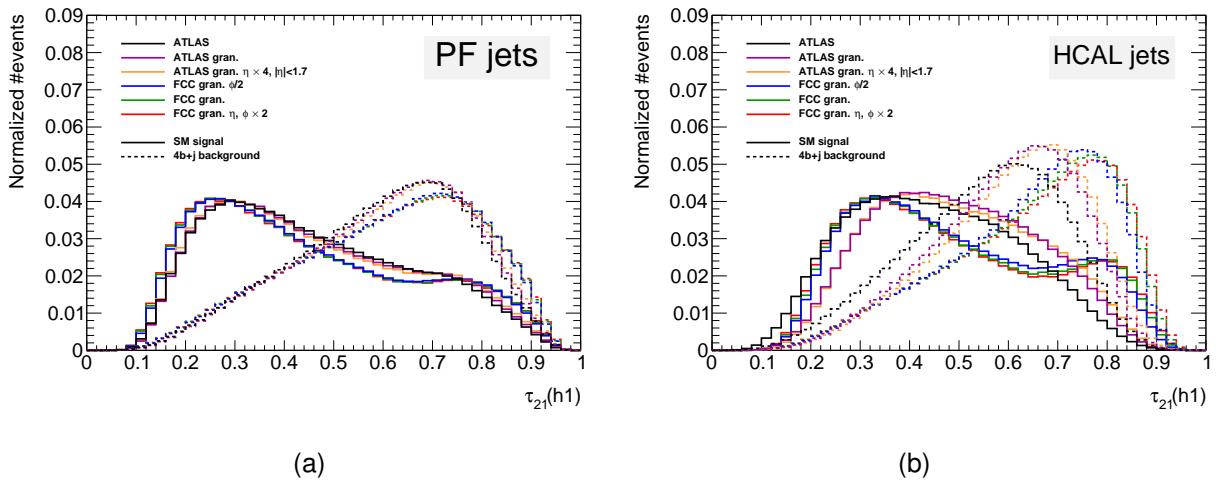


Figure 7.4: (a) Leading Higgs candidate  $\tau_{21}$  distribution for eflow jets, (b) Leading Higgs candidate  $\tau_{21}$  distribution for HCAL jets. The colors indicate the different detector configurations. The distributions are shown for the signal (filled lines) and for the  $4b + j$  background (dashed lines).

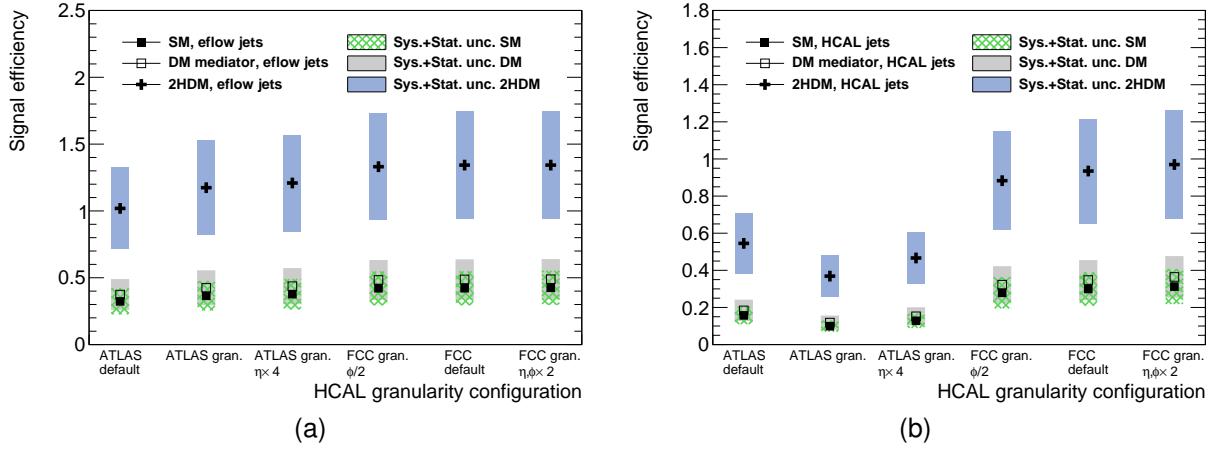


Figure 7.5: (a) Signal efficiency as a function of the detector configuration for particle flow jets, (b) Signal efficiency as a function of the detector configuration for calorimeter jets. Three signal models are shown: SM (filled squares), 1 TeV DM mediator (empty squares) and type II 2HDM with  $m_H = 900$  GeV (crosses). The error bars are drawn but are smaller than the markers.

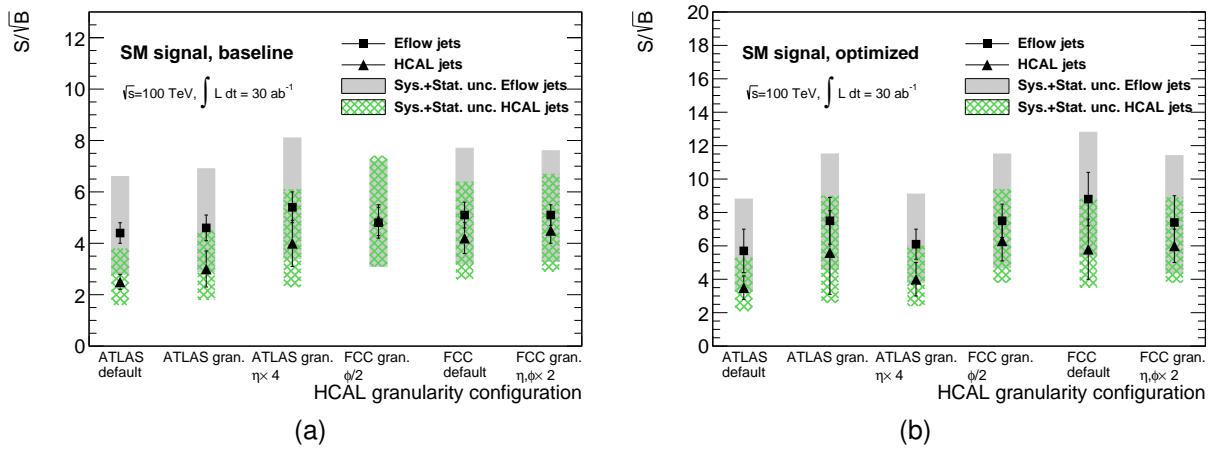


Figure 7.6: Significance as a function of the detector configuration for the SM signal, for the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and pure HCAL jets (triangles).

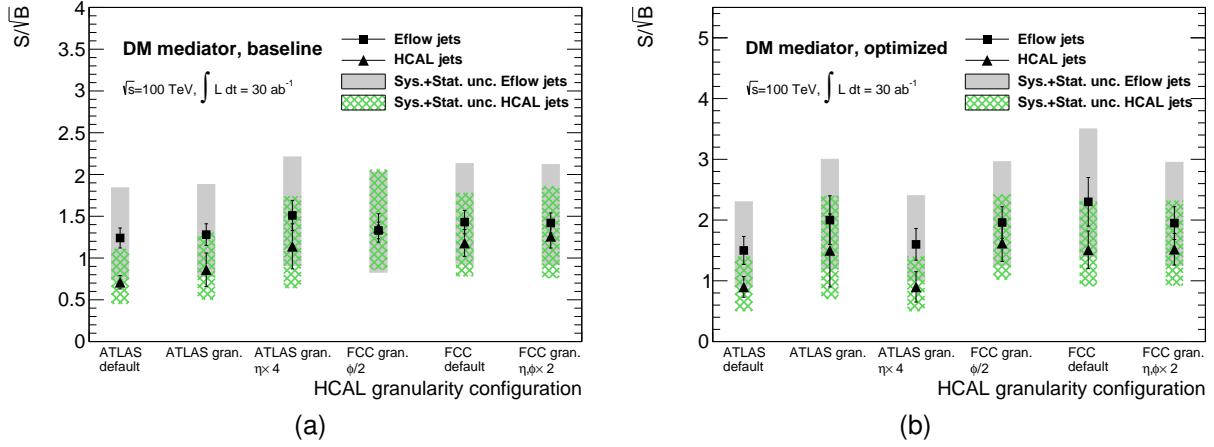


Figure 7.7: Significance as a function of the detector configuration for the DM mediator signal, for the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and pure HCAL jets (triangles).

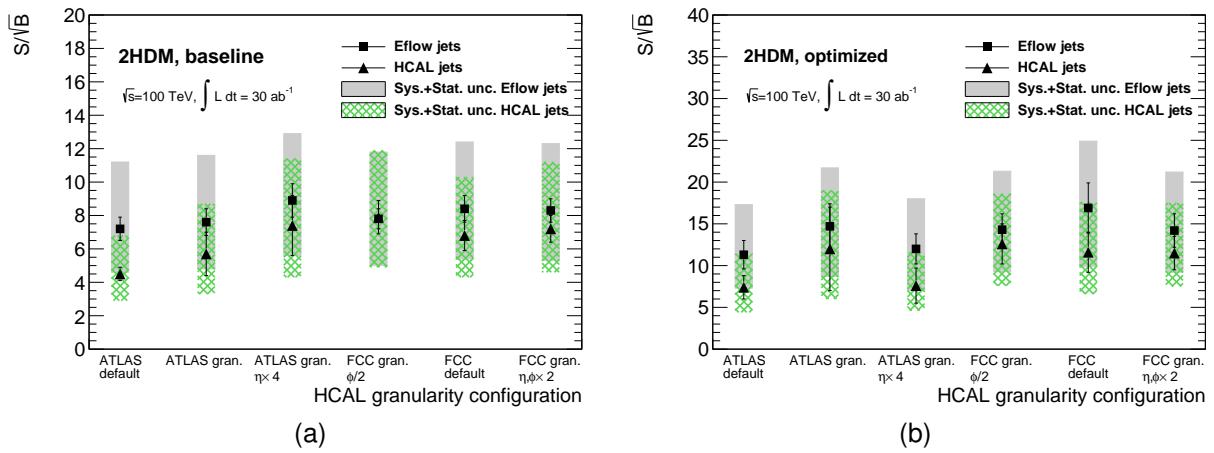


Figure 7.8: Significance as a function of the detector configuration for the 2HDM signal, for the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and pure HCAL jets (triangles).

1776 **Chapter 8**

1777 **Conclusions**

1778 This thesis presented a feasibility study targeting the search for Higgs pair production in the  $pp \rightarrow hh \rightarrow$   
1779  $b\bar{b}b\bar{b}$  channel at a center of mass energy of  $\sqrt{s} = 100$  TeV. The analysis targeted a boosted kinematic  
1780 regime. The impact of the granularity of the hadronic calorimeter in the achieved significance was  
1781 analyzed.

1782 The analysis consisted of a cut-based event selection designed to efficiently select signal events con-  
1783 sistent with a boosted topology. In addition to standard kinematic variables, jet substructure variables,  
1784 namely a ratio of N-subjetiness variables and a Fox-Wolfram momentum, are used to further suppress  
1785 the background. The analysis required the existence of at least two b-tagged large- $R$  jets consistent  
1786 with having two subjets and with  $p_T > 200$  GeV. Multijet production through QCD interactions and  $t\bar{t}$   
1787 production constitute the main backgrounds.

1788 In addition to SM Higgs pair production, two BSM benchmark processes were considered: a 1  
1789 TeV dark matter mediator and a heavy Higgs boson with  $m_H = 900$  GeV in the framework of the  
1790 CP-conserving type II 2HDM, both decaying to pairs of SM Higgs bosons. The existence of heavy  
1791 resonances is expected to enhance Higgs pair production with respect to the SM.

1792 For the hadronic Future Circular Collider (FCC-hh), the significance obtained was  $S/\sqrt{B} = 8.8 \pm$   
1793 1.6 (stat.) ( $2.8 \pm 0.5$  (stat.)) for an integrated luminosity of  $\mathcal{L} = 30$  (3)  $\text{ab}^{-1}$ . For  $\mathcal{L} = 30 \text{ ab}^{-1}$ , the  
1794 value is above the observation threshold ( $5\sigma$ ) which indicates that the full dataset that is expected to  
1795 be collected during the operation of the FCC-hh should be enough to claim the observation of Higgs  
1796 pair production, assuming SM production. For the CP-conserving type II 2HDM with  $m_H = 900$  GeV,  
1797 the achieved significance was  $S/\sqrt{B} = 16.9 \pm 3.0$  (stat.) ( $5.4 \pm 0.9$  (stat.)) for an integrated luminosity  
1798 of  $\mathcal{L} = 30$  (3)  $\text{ab}^{-1}$ , which makes it an accessible benchmark for the HL-LHC and future colliders.  
1799 The achieved significance for the 1 TeV DM mediator is of order 1 even for an integrated luminosity of  
1800  $\mathcal{L} = 30 \text{ ab}^{-1}$ , which makes it a very challenging benchmark and probably unaccessible even at future  
1801 high energy colliders.

1802 With the ATLAS detector simulation, the achieved significance for the SM signal is  $S/\sqrt{B} = 5.7 \pm$   
1803 1.3 (stat.) ( $1.8 \pm 0.4$  (stat.)) for an integrated luminosity of  $\mathcal{L} = 30$  (3)  $\text{ab}^{-1}$ , which is also above the ob-  
1804 servation threshold. This result indicates that the factor driving the discovery potential of the accelerator

1805 is the CM energy as opposed to the resolution of the detector.

1806 The resolution of the jet mass and the separation provided by the  $\tau_{21}$  variable are shown to increase  
1807 as the granularity of the HCAL increases. This effect is more evident when using pure calorimeter jets.

1808 When using particle flow jets, the change in the significance over the range of detector configurations  
1809 tested is small, for all signal models. For the baseline (optimized) analysis, it varies by approximately  
1810 15% (54%), for all signal models. Using pure calorimeter jets, the change in significance is more accent-  
1811 tuated. For the SM signal, it varies by approximately 80% (65%), considering the baseline (optimized)  
1812 analysis. For the same detector configuration, the significance is always smaller when using calorimeter  
1813 jets. The same qualitative conclusions hold for the BSM models. The effect of the granularity is of the  
1814 same order of magnitude as for the SM.

1815 These results lead to the conclusion that the jet reconstruction performance is dominated by the  
1816 tracking system.

# Bibliography

- [1] J. Goldstone. Field Theories with Superconductor Solutions. *Nuovo Cim.*, 19:154–164, 1961. doi: 10.1007/BF02812722.
- [2] Y. Nambu and G. Jona-Lasinio. Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. 1. *Phys. Rev.*, 122:345–358, 1961. doi: 10.1103/PhysRev.122.345. [,127(1961)].
- [3] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c\*\*2 at the CERN SPS Collider. *Phys. Lett.*, B126:398–410, 1983. doi: 10.1016/0370-2693(83)90188-0. [,7.55(1983)].
- [4] G. Arnison et al. Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at  $s^{**}(1/2) = 540\text{-GeV}$ . *Phys. Lett.*, B122:103–116, 1983. doi: 10.1016/0370-2693(83)91177-2. [,611(1983)].
- [5] F. Wilczek. Spontaneous symmetry breaking: General. <http://web.mit.edu/8.701/www/LectureNotes/8.701spontaneousSymmetryBreaking-GeneralFA13.pdf>, 2013. Accessed 2018-01-11.
- [6] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964. doi: 10.1103/PhysRevLett.13.508. [,160(1964)].
- [7] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. doi: 10.1103/PhysRevLett.13.321. [,157(1964)].
- [8] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, 1964. doi: 10.1103/PhysRevLett.13.585. [,162(1964)].
- [9] R. Costa Batalha Pedro, J. Maneira, and P. Conde Muio. Search for the Higgs boson at ATLAS/LHC in WH associated production and decay to b quark pairs, Apr 2017. URL <https://cds.cern.ch/record/2304102>. Presented 09 Nov 2017.
- [10] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi: 10.1103/PhysRev.105.1413. URL <https://link.aps.org/doi/10.1103/PhysRev.105.1413>.

- 1844 [11] M. Goldhaber, L. Grodzins, and A. W. Sunyar. Helicity of neutrinos. *Phys. Rev.*, 109:  
 1845 1015–1017, Feb 1958. doi: 10.1103/PhysRev.109.1015. URL <https://link.aps.org/doi/10.1103/PhysRev.109.1015>.
- 1846  
 1847 [12] S. M. Bilenky and J. Hosek. GLASHOW-WEINBERG-SALAM THEORY OF ELECTROWEAK  
 1848 INTERACTIONS AND THE NEUTRAL CURRENTS. *Phys. Rept.*, 90:73–157, 1982. doi: 10.  
 1849 1016/0370-1573(82)90016-3.
- 1850 [13] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, E. Vryonidou, and M. Zaro.  
 1851 Higgs pair production at the LHC with NLO and parton-shower effects. *Phys. Lett.*, B732:142–149,  
 1852 2014. doi: 10.1016/j.physletb.2014.03.026.
- 1853 [14] M. Mangano. *Physics at the FCC-hh, a 100 TeV pp collider*. CERN Yellow Reports: Monographs.  
 1854 CERN, Geneva, 2017. URL <https://cds.cern.ch/record/2270978>.
- 1855 [15] T. Plehn, M. Spira, and P. M. Zerwas. Pair production of neutral Higgs particles in gluon-  
 1856 gluon collisions. *Nucl. Phys.*, B479:46–64, 1996. doi: 10.1016/0550-3213(96)00418-X, 10.1016/  
 1857 S0550-3213(98)00406-4. [Erratum: Nucl. Phys.B531,655(1998)].
- 1858 [16] S. Borowka, N. Greiner, G. Heinrich, S. Jones, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke.  
 1859 Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading Order with Full Top-Quark Mass  
 1860 Dependence. *Phys. Rev. Lett.*, 117(1):012001, 2016. doi: 10.1103/PhysRevLett.117.079901, 10.  
 1861 1103/PhysRevLett.117.012001. [Erratum: Phys. Rev. Lett.117,no.7,079901(2016)].
- 1862 [17] S. Borowka, N. Greiner, G. Heinrich, S. P. Jones, M. Kerner, J. Schlenk, and T. Zirke. Full top  
 1863 quark mass dependence in Higgs boson pair production at NLO. *JHEP*, 10:107, 2016. doi:  
 1864 10.1007/JHEP10(2016)107.
- 1865 [18] H. C. van de Hulst, E. Raimond, and H. van Woerden. Rotation and density distribution of the  
 1866 Andromeda nebula derived from observations of the 21-cm line. 14:1, Nov. 1957.
- 1867 [19] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567,  
 1868 1998. doi: 10.1103/PhysRevLett.81.1562.
- 1869 [20] Q. R. Ahmad et al. Measurement of the rate of  $\nu_e + d \rightarrow p + p + e^-$  interactions produced by  
 1870  $^8B$  solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001. doi:  
 1871 10.1103/PhysRevLett.87.071301.
- 1872 [21] Q. R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral current  
 1873 interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, 2002. doi:  
 1874 10.1103/PhysRevLett.89.011301.
- 1875 [22] B. Gripaios. Lectures on Physics Beyond the Standard Model. 2015.
- 1876 [23] N. M. A. 2018. Popular information, 2004. URL <https://www.nobelprize.org/prizes/physics/2004/popular-information/>.

- 1878 [24] V. Branchina, F. Contino, and P. M. Ferreira. Electroweak vacuum lifetime in two Higgs doublet  
1879 models. 2018.
- 1880 [25] J. F. Gunion and H. E. Haber. The CP conserving two Higgs doublet model: The Approach to the  
1881 decoupling limit. *Phys. Rev.*, D67:075019, 2003. doi: 10.1103/PhysRevD.67.075019.
- 1882 [26] H. Sun, Y.-J. Zhou, and H. Chen. Constraints on large-extra-dimensions model through 125-  
1883 GeV Higgs pair production at the LHC. *Eur. Phys. J.*, C72:2011, 2012. doi: 10.1140/epjc/  
1884 s10052-012-2011-4.
- 1885 [27] P. Huang, A. Joglekar, M. Li, and C. E. M. Wagner. Corrections to di-Higgs boson production  
1886 with light stops and modified Higgs couplings. *Phys. Rev.*, D97(7):075001, 2018. doi: 10.1103/  
1887 PhysRevD.97.075001.
- 1888 [28] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with  
1889 the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.  
1890 020.
- 1891 [29] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment  
1892 at the LHC. *Phys. Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- 1893 [30] Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector. Technical Report  
1894 ATLAS-CONF-2018-036, CERN, Geneva, Jul 2018. URL <http://cds.cern.ch/record/2630338>.
- 1895 [31] Observation of Higgs boson decay to bottom quarks. Technical Report arXiv:1808.08242. CMS-  
1896 HIG-18-016-003, CERN, Geneva, Aug 2018. URL <https://cds.cern.ch/record/2636067>. \*  
1897 Temporary entry \*.
- 1898 [32] A. M. Sirunyan et al. Observation of the Higgs boson decay to a pair of  $\tau$  leptons with the CMS  
1899 detector. *Phys. Lett.*, B779:283–316, 2018. doi: 10.1016/j.physletb.2018.02.004.
- 1900 [33] A. M. Sirunyan et al. Observation of  $t\bar{t}H$  production. *Phys. Rev. Lett.*, 120:231801, 2018. doi:  
1901 10.1103/PhysRevLett.120.231801.
- 1902 [34] M. Aaboud et al. Observation of Higgs boson production in association with a top quark pair at  
1903 the LHC with the ATLAS detector. 2018.
- 1904 [35] Measurement of the Higgs boson mass in the  $H \rightarrow ZZ^* \rightarrow 4\ell$  and  $H \rightarrow \gamma\gamma$  channels with  
1905  $\sqrt{s}=13\text{TeV}$   $pp$  collisions using the ATLAS detector. Technical Report ATLAS-CONF-2017-046,  
1906 CERN, Geneva, Jul 2017. URL <http://cds.cern.ch/record/2273853>.
- 1907 [36] G. Aad et al. Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$   
1908 and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015. doi:  
1909 10.1103/PhysRevLett.114.191803.
- 1910 [37] M. Aaboud et al. Measurement of the  $W$ -boson mass in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the  
1911 ATLAS detector. *Eur. Phys. J.*, C78(2):110, 2018. doi: 10.1140/epjc/s10052-017-5475-4.

- 1912 [38] R. Nisius. Measurements of the top quark mass with the ATLAS detector. *PoS*, EPS-HEP2017:  
1913 453, 2017. doi: 10.22323/1.314.0453.
- 1914 [39] Measurement of the top quark mass with muon+jets final states in pp collisions at  $\sqrt{s} = 13$  TeV.  
1915 Technical Report CMS-PAS-TOP-16-022, CERN, Geneva, 2017. URL <http://cds.cern.ch/record/2255834>.
- 1916 [40] CERN. The high-luminosity lhc. <http://cds.cern.ch/record/2114693>, 2015. Accessed  
1917 2018/01/11.
- 1918 [41] T. Koffas. ATLAS Higgs physics prospects at the high luminosity LHC. *PoS*, ICHEP2016:426,  
1919 2017.
- 1920 [42] A. M. Henriques Correia. The ATLAS Tile Calorimeter. Technical Report ATL-TILECAL-PROC-  
1921 2015-002, CERN, Geneva, Mar 2015. URL <https://cds.cern.ch/record/2004868>.
- 1922 [43] ATLAS liquid argon calorimeter: Technical design report. 1996.
- 1923 [44] M. Cacciari, G. P. Salam, and G. Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2008. doi:  
1924 10.1088/1126-6708/2008/04/005.
- 1925 [45] Optimisation of the ATLAS  $b$ -tagging performance for the 2016 LHC Run. Technical Report ATL-  
1926 PHYS-PUB-2016-012, CERN, Geneva, Jun 2016. URL <https://cds.cern.ch/record/2160731>.
- 1927 [46] Commissioning of the ATLAS high-performance  $b$ -tagging algorithms in the 7 TeV collision data.  
1928 Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011. URL <https://cds.cern.ch/record/1369219>.
- 1929 [47] G. Aad et al. Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics.  
1930 2009.
- 1931 [48] A. Ruiz-Martinez and A. Collaboration. The Run-2 ATLAS Trigger System. Technical Report ATL-  
1932 DAQ-PROC-2016-003, CERN, Geneva, Feb 2016. URL <https://cds.cern.ch/record/2133909>.
- 1933 [49] S. D. Ellis and D. E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*,  
1934 D48:3160–3166, 1993. doi: 10.1103/PhysRevD.48.3160.
- 1935 [50] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic  
1936 scattering. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg,*  
1937 *Germany, 1998-1999*, pages 270–279, 1998. URL [https://inspirehep.net/record/484872/files/arXiv:hep-ph\\_9907280.pdf](https://inspirehep.net/record/484872/files/arXiv:hep-ph_9907280.pdf).
- 1938 [51] M. Cacciari, G. P. Salam, and G. Soyez. The Anti- $k(t)$  jet clustering algorithm. *JHEP*, 04:063,  
1939 2008. doi: 10.1088/1126-6708/2008/04/063.
- 1940 [52] A. Altheimer et al. Boosted objects and jet substructure at the LHC. Report of BOOST2012, held  
1941 at IFIC Valencia, 23rd-27th of July 2012. *Eur. Phys. J.*, C74(3):2792, 2014. doi: 10.1140/epjc/s10052-014-2792-8.

- 1946 [53] G. A. et al. Performance of jet substructure techniques for large- $r$  jets in proton-proton collisions  
 1947 at  $\sqrt{s} = 7$  tev using the atlas detector. *Journal of High Energy Physics*, 2013(9), Sep 2013. doi:  
 1948 10.1007/JHEP09(2013)076. URL [https://doi.org/10.1007/JHEP09\(2013\)076](https://doi.org/10.1007/JHEP09(2013)076).
- 1949 [54] J. Thaler and K. Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015,  
 1950 2011. doi: 10.1007/JHEP03(2011)015.
- 1951 [55] C. Chen. New approach to identifying boosted hadronically-decaying particle using jet substruc-  
 1952 ture in its center-of-mass frame. *Phys. Rev.*, D85:034007, 2012. doi: 10.1103/PhysRevD.85.  
 1953 034007.
- 1954 [56] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam. Jet substructure as a new Higgs  
 1955 search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008. doi: 10.1103/PhysRevLett.100.  
 1956 242001.
- 1957 [57] I. Hinchliffe, A. Kotwal, M. L. Mangano, C. Quigg, and L.-T. Wang. Luminosity goals for a 100-TeV  
 1958 pp collider. *Int. J. Mod. Phys.*, A30(23):1544002, 2015. doi: 10.1142/S0217751X15440029.
- 1959 [58] N. Arkani-Hamed, T. Han, M. Mangano, and L.-T. Wang. Physics opportunities of a 100 TeV  
 1960 protonproton collider. *Phys. Rept.*, 652:1–49, 2016. doi: 10.1016/j.physrep.2016.07.004.
- 1961 [59] C. Helsens. Fcc/he-lhc physics/performance studies towards the cdr. URL: [https://indico.cern.ch/event/675588/contributions/2765045/attachments/1572548/2484309/Clement\\_FCC.pdf](https://indico.cern.ch/event/675588/contributions/2765045/attachments/1572548/2484309/Clement_FCC.pdf), 2017.
- 1964 [60] T. an FCC-hh Detector Magnet System. E. bielert, 2018. URL <https://ep-news.web.cern.ch/content/towards-fcc-hh-detector-magnet-system>.
- 1966 [61] J. Faltova. Electromagnetic calorimeter using lar technology for fcc-hh. URL: [https://indico.cern.ch/event/556692/contributions/2465166/attachments/1468150/2272205/ecal\\_fccBerlin.pdf](https://indico.cern.ch/event/556692/contributions/2465166/attachments/1468150/2272205/ecal_fccBerlin.pdf), 2017.
- 1969 [62] A. Zaborowska. Electromagnetic calorimetry based on liquid argon for the fcc-hh experi-  
 1970 ments. URL: [https://indico.cern.ch/event/656491/contributions/2923535/attachments/1631506/2601159/FCChh\\_LArCalo\\_Zaborowska.pdf](https://indico.cern.ch/event/656491/contributions/2923535/attachments/1631506/2601159/FCChh_LArCalo_Zaborowska.pdf), 2018.
- 1972 [63] C. Neubser. Performance studies and requirements on the calorimeters for a fcc-hh ex-  
 1973 periment. URL: <https://indico.ihep.ac.cn/event/6387/session/14/contribution/117/material/slides/0.pdf>, 2017.
- 1975 [64] J. Baglio, A. Djouadi, R. Grber, M. M. Mhlleitner, J. Quevillon, and M. Spira. The measurement  
 1976 of the Higgs self-coupling at the LHC: theoretical status. *JHEP*, 04:151, 2013. doi: 10.1007/  
 1977 JHEP04(2013)151.
- 1978 [65] M. Aaboud et al. Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton–  
 1979 proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev.*, D94(5):052002, 2016. doi:  
 1980 10.1103/PhysRevD.94.052002.

- 1981 [66] A. M. Sirunyan et al. Search for a massive resonance decaying to a pair of Higgs bosons in the  
 1982 four b quark final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- 1983 [67] G. Aad et al. Searches for Higgs boson pair production in the  $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$   
 1984 channels with the ATLAS detector. *Phys. Rev.*, D92:092004, 2015. doi: 10.1103/PhysRevD.92.  
 1985 092004.
- 1986 [68] A. M. Sirunyan et al. Search for Higgs boson pair production in events with two bottom quarks and  
 1987 two tau leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- 1988 [69] M. Aaboud et al. A search for resonant and non-resonant Higgs boson pair production in the  
 1989  $b\bar{b}\tau^+\tau^-$  decay channel in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector . Technical Report  
 1990 arXiv:1808.00336, CERN, Geneva, Aug 2018. URL <https://cds.cern.ch/record/2632994>. \*  
 1991 Temporary entry \*.
- 1992 [70] V. Khachatryan et al. Search for two Higgs bosons in final states containing two photons and  
 1993 two bottom quarks in proton-proton collisions at 8 TeV. *Phys. Rev.*, D94(5):052012, 2016. doi:  
 1994 10.1103/PhysRevD.94.052012.
- 1995 [71] G. Aad et al. Search For Higgs Boson Pair Production in the  $\gamma\gamma b\bar{b}$  Final State using  $pp$  Collision  
 1996 Data at  $\sqrt{s} = 8$  TeV from the ATLAS Detector. *Phys. Rev. Lett.*, 114(8):081802, 2015. doi:  
 1997 10.1103/PhysRevLett.114.081802.
- 1998 [72] A. M. Sirunyan et al. Search for resonant and nonresonant Higgs boson pair production in the  
 1999  $bb\lnu\lnu$  final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- 2000 [73] M. Aaboud et al. Combination of searches for heavy resonances decaying into bosonic and  
 2001 leptonic final states using  $36\text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV with the ATLAS  
 2002 detector. 2018.
- 2003 [74] G. Aad and t. . others.
- 2004 [75] J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever, and J. Rojo. Boosting Higgs pair  
 2005 production in the  $b\bar{b}b\bar{b}$  final state with multivariate techniques. *Eur. Phys. J.*, C76(7):386, 2016. doi:  
 2006 10.1140/epjc/s10052-016-4215-5.
- 2007 [76] S. Banerjee, C. Englert, M. L. Mangano, M. Selvaggi, and M. Spannowsky.  $hh + \text{jet}$  production at  
 2008 100 TeV. *Eur. Phys. J.*, C78(4):322, 2018. doi: 10.1140/epjc/s10052-018-5788-y.
- 2009 [77] N. P. Hartland. Double Higgs production from  $HH \rightarrow (b\bar{b})(b\bar{b})$  at a 100 TeV hadron collider. *PoS*,  
 2010 DIS2016:115, 2016.
- 2011 [78] S. V. Chekanov, M. Beydler, A. V. Kotwal, L. Gray, S. Sen, N. V. Tran, S. S. Yu, and J. Zuzelski.  
 2012 Initial performance studies of a general-purpose detector for multi-TeV physics at a 100 TeV pp  
 2013 collider. *JINST*, 12(06):P06009, 2017. doi: 10.1088/1748-0221/12/06/P06009.

- 2014 [79] S. Chekanov. Boosted jets in high-granularity calorimeter at a 100 TeV pp collider. BOOST 2017,  
2015 2017.
- 2016 [80] S. Chekanov. Physics requirements for a Hadron Calorimeter for a 100 TeV proton-proton collider.  
2017 FCC week 2015, 2015.
- 2018 [81] S. Chekanov. Simulation of a high-granular hadronic calorimeter for multi-TeV physics. FCC week  
2019 2016, 2016.
- 2020 [82] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer.
- 2021 [83] T. Sjostrand, S. Mrenna, and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys.*  
2022 *Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.
- 2023 [84] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi.  
2024 DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:  
2025 057, 2014. doi: 10.1007/JHEP02(2014)057.
- 2026 [85] E. Boos et al. Generic user process interface for event generators. In *Physics at TeV colliders.*  
2027 *Proceedings, Euro Summer School, Les Houches, France, May 21-June 1, 2001*, 2001. URL  
2028 <http://lss.fnal.gov/archive/preprint/fermilab-conf-01-496-t.shtml>.
- 2029 [86] M. Cacciari, G. P. Salam, and G. Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012. doi:  
2030 10.1140/epjc/s10052-012-1896-2.
- 2031 [87] D. Buskulic et al. Performance of the ALEPH detector at LEP. *Nucl. Instrum. Meth.*, A360:481–506,  
2032 1995. doi: 10.1016/0168-9002(95)00138-7.
- 2033 [88] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical  
2034 Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. URL <http://cds.cern.ch/record/1194487>.
- 2036 [89] FCC study group. Fccsw. <https://github.com/HEP-FCC/FCCSW>, 2018.
- 2037 [90] G. Barrand et al. GAUDI - A software architecture and framework for building HEP data processing  
2038 applications. *Comput. Phys. Commun.*, 140:45–55, 2001. doi: 10.1016/S0010-4655(01)00254-5.
- 2039 [91] F. Gaede, B. Hegner, and P. Mato. Podio: An event-data-model toolkit for high energy physics  
2040 experiments. *Journal of Physics: Conference Series*, 898(7):072039, 2017. URL <http://stacks.iop.org/1742-6596/898/i=7/a=072039>.
- 2042 [92] FCC-hh study group. Eventproducer. <https://github.com/FCC-hh-framework/EventProducer>,  
2043 2018.
- 2044 [93] N. D. Christensen and C. Duhr. FeynRules - Feynman rules made easy. *Comput. Phys. Commun.*,  
2045 180:1614–1641, 2009. doi: 10.1016/j.cpc.2009.02.018.

- 2046 [94] T. general Two-Higgs-Doublet Model. C. degrande, c. duhr, m. herquet, 2018. URL <http://feynrules.irmpl.ac.be/wiki/2HDM>.
- 2047
- 2048 [95] C. Degrande. Automatic evaluation of UV and R2 terms for beyond the Standard Model La-  
2049 grangians: a proof-of-principle. *Comput. Phys. Commun.*, 197:239–262, 2015. doi: 10.1016/j.  
2050 cpc.2015.08.015.
- 2051 [96] H. E. Haber and O. Stl. New LHC benchmarks for the  $\mathcal{CP}$ -conserving two-Higgs-doublet  
2052 model. *Eur. Phys. J.*, C75(10):491, 2015. doi: 10.1140/epjc/s10052-015-3697-x,10.1140/epjc/  
2053 s10052-016-4151-4. [Erratum: Eur. Phys. J.C76,no.6,312(2016)].
- 2054 [97] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the  
2055 Higgs Sector. 2016. doi: 10.23731/CYRM-2017-002.
- 2056 [98] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, and J. Mazzitelli. Higgs  
2057 boson pair production at NNLO with top quark mass effects. *JHEP*, 05:059, 2018. doi: 10.1007/  
2058 JHEP05(2018)059.
- 2059 [99] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi: 10.1088/  
2060 1674-1137/40/10/100001.
- 2061 [100] M. Aaboud et al. Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton-  
2062 proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. 2018.
- 2063 [101] C. Helsens. Heavy resonances at 100tev, 2018. URL [https://indico.cern.ch/event/704428/contributions/2957537/attachments/1627728/2594323/Resonances\\_FCC\\_hh\\_helsens.pdf](https://indico.cern.ch/event/704428/contributions/2957537/attachments/1627728/2594323/Resonances_FCC_hh_helsens.pdf).
- 2064

2065 **Appendix A**

2066 **Jet radius discussion**

2067 It is important to discuss if the radius of the jet that is used to reconstruct the boosted Higgs boson  
 2068 candidates ( $R = 0.8$ ) is the appropriate one. The question arises because in ATLAS boosted objects  
 2069 with  $p_T \sim 200$  GeV are reconstructed using  $R = 1.0$  jets. In this work, we work with a CM energy of 100  
 2070 TeV and require that the two leading jets have  $p_T > 200$  GeV and use jets with  $R = 0.8$ . It is necessary  
 2071 to understand if these jets are large enough to fully reconstruct the Higgs candidates.

2072 As a first approximation we compute the angle between the b quarks produced by the decay of  
 2073 a Higgs boson. We assume that the b quarks are massless and that the Higgs moves only in the  
 2074 transverse plane (perpendicular to the beam pipe) such that it has no longitudinal momentum and the  
 2075 angle between the b quarks is given by  $\Delta\phi$ . For  $p_T(\text{Higgs}) = 200$  GeV, we get  $\Delta\phi(b, \bar{b}) = 1.1$  which is  
 2076 smaller than the jet's diameter (1.6) and therefore the two b quarks can both be contained inside the jet  
 2077 and the Higgs boson fully reconstructed.

2078 Another test we can make is to compute the  
 2079  $\Delta R$  between the b quarks coming from the lead-  
 2080 ing Higgs candidate with  $p_T$  larger than a given  
 2081 value using truth level information. In figure A.1  
 2082 we show the distribution of  $\Delta R(b, \bar{b})$  for  $p_T$  of  
 2083 the leading Higgs candidate larger than 200 GeV  
 2084 (solid blue) and 300 GeV (solid black). To obtain  
 2085 the dashed histograms we apply the  $p_T$  to both  
 2086 Higgs candidates. The integral of the histograms  
 2087 between 0 and 1.6 gives an estimate of the frac-  
 2088 tion of signal we keep if we apply these  $p_T$  cuts.  
 2089 For  $p_T(h_1) > 200(300)$  GeV we get that 93(98) %  
 2090 of the signal has b quarks with  $\Delta R < 1.6$  and  
 2091 therefore can be fully reconstructed using a jet with  $R = 0.8$ .

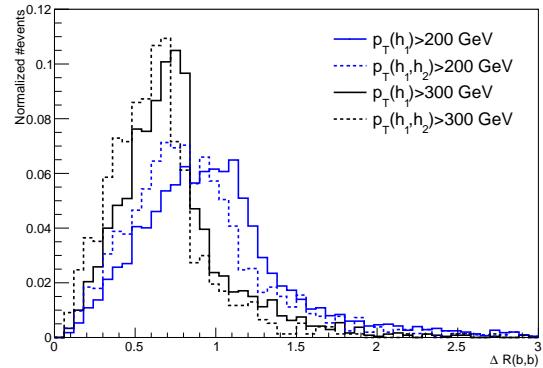


Figure A.1:  $\Delta R(b, \bar{b})$  distributions for  $p_T(h_1) > 200/300$  GeV (solid blue/black) and for  $p_T(h_1, h_2) > 200/300$  GeV (dashed blue/black).

2092 **Appendix B**

2093 **Additional background processes**

2094 In this section we discuss the importance of additional background processes and estimate how they  
2095 influence the analysis. In particular, we investigate backgrounds involving Higgs bosons. We consider  
2096 the  $t\bar{t}h$  and  $h + 0/1/2 j$  processes. The effective cross section, efficiency and expected number of events  
2097 for an integrated luminosity of  $\mathcal{L} = 30 \text{ ab}^{-1}$  are summarized in table B.1. Including these backgrounds  
2098 in the analysis leads to a decrease in the significance of 0.26%, which we consider to be a very small  
2099 effect. Therefore, neglecting these backgrounds is a safe assumption.

Table B.1: Effective cross section ( $\sigma \times \text{BR} \times \text{k-factor}$ ), efficiency and expected number of events for  $\mathcal{L} = 30 \text{ ab}^{-1}$  for the  $t\bar{t}h + 0/1 j(h \rightarrow b\bar{b})$  and  $h + 0/1/2 j(h \rightarrow b\bar{b})$  backgrounds.

Process	$\sigma \times \text{BR} \times \text{k-factor} [\text{pb}]$	Efficiency [%]	Expected nb. events ( $\mathcal{L} = 30 \text{ ab}^{-1}$ )
$t\bar{t}h + 0/1 j(h \rightarrow b\bar{b})$	31.86	0.089	$8.5 \times 10^5$
$h + 0/1/2 j(h \rightarrow b\bar{b})$	1286.52	0.0041	$1.6 \times 10^6$

2100 In addition, we also estimate the error associated with not considering the a sample of  $4b + j$  with  
2101  $p_T(j) < 200 \text{ GeV}$ . In order to do so, we generate a sample with 110k events of  $4b + j$  with  $(30 < p_T(j) <$   
2102  $200) \text{ GeV}$ . The cross section is  $\sigma = 7450 \text{ pb}$ . No event goes through the cuts  $p_T(\text{leading jet}) > 300$   
2103  $\text{GeV}$  and  $p_T(\text{Higgs pair}) > 100 \text{ GeV}$ . However, if we assume that 1 events goes through the cuts (which  
2104 corresponds to a very conservative  $9.09 \times 10^{-4}\%$  efficiency) we have an expected number of events of  
2105  $2.03 \times 10^6$  for an integrated luminosity of  $\mathcal{L} = 30 \text{ ab}^{-1}$ . If we include this background, the significance  
2106 decreases by 0.22% which is a very small effect. Therefore, it is safe to neglect this background.

2107 **Appendix C**

2108 **Softdrop mass**

2109 It is noticeable in the softdrop mass spectrum of both Higgs candidates (for signal and backgrounds) the  
2110 existence an atypical peak close to zero. In this section we explain the origin of this feature.

2111 We believe that the peak close to zero corresponds to Higgs candidates ( $R = 0.8$  jets) that do not  
2112 contain both b quarks from the Higgs decay. The plot that support this conclusion is shown in figure  
2113 C.1. It shows the correlation between the maximum  $\Delta R$  between the Higgs candidate jet and one of  
2114 the b quarks (y axis) and the jet's mass (x axis). The horizontal line corresponds to  $\Delta R = 0.8$ . For  
2115 small masses ( $< 80$  GeV) the maximum  $\Delta R$  is usually larger than 0.8 which means that at least one  
2116 of the b quarks is not contained in the jet's cone. When applying the soft drop procedure to these jets,  
2117 soft radiation is removed and we are left with a single b quark. The mass of b quarks is  $\sim 5$  GeV and  
2118 therefore we get a peak at this mass.

2119 In practice, this does not affect our analysis because we place a mass window cut around the Higgs  
2120 boson mass and therefore the low mass peak is removed. Nonetheless, the study presented here is  
2121 extremely important because it helps rule out possible malfunctions of the soft drop algorithm and gives  
2122 us confidence that we understand exactly what is happening.

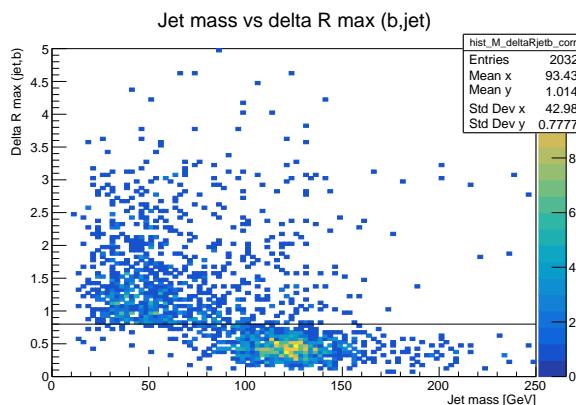


Figure C.1: Correlation between the maximum  $\Delta R$  between the Higgs candidate jet and one of the b quarks (y axis) and the jet's mass (x axis). The horizontal line corresponds to  $\Delta R = 0.8$ .

2123 **Appendix D**

2124 **Samples generation: parameters**

2125 In this appendix we provide we list the values of parameters used in MadGraph (section D.1) and Pythia  
2126 (section D.2) to generate the samples used in this work.

2127 **D.1 MadGraph**

2128 The MadGraph5 level cuts are summarized in table D.1. We show only the most relevant cuts for this  
2129 analysis: the minimum  $p_T$  of light and b quarks,  $p_{T,j}^{\min}$  and  $p_{T,b}^{\min}$ , the maximum pseudorapidity range for  
2130 light and b quarks,  $\eta_j^{\max}$  and  $\eta_b^{\max}$  and the  $\Delta R$  separation between two light quarks,  $\Delta R(jj)$ , two b  
2131 quarks,  $\Delta R(bb)$ , and between a light and b quarks,  $\Delta R(jb)$ . The  $xqcut$  parameter is a measure of the  
2132 required parton separation at Madgraph level. Whenever MadGraph produces two partons,  $i$  and  $j$ ,  
2133 we define the distance between them as  $\sqrt{2 * \min(p_{T,i}, p_{T,j}) [\cosh(\eta_i - \eta_j) - \cos(\phi_i - \phi_j)]}$ . If the value  
2134 of this expression is smaller than the specified value of  $xqcut$  then we do not generate the event. The  
2135  $bwcutoff$  parameter defines what is considered to be on-shell s-channel resonances. The  $H_T$  variable  
2136 is the scalar sum of the  $p_T$  of all truth level partons, including b quarks.

2137 **D.2 Pythia**

2138 The settings for jet matching can be found in table D.2 under the corresponding samples' columns. We  
2139 perform the jet matching procedure (merge=on) using the MLM matching scheme and the appropriate  
2140 algorithm for a parton level process generated in MadGraph (scheme=1). We do not read the matching  
2141 parameters from the MadGraph file (setMad=off) because this option is not available for these files. The  
2142 size of the cone drawn around the jet's center, the maximum pseudorapidity and the maximum number  
2143 of jets to be matched are given by coneRadius, etaJetMax and nJetMax, respectively. The cone radius  
2144 is set to one. The maximum allowed pseudorapidity of jets is ten which is a much loser cut than the  
2145 acceptance of any current detector. The maximum number of jets is set to four for the jj+0/1/2 j and to  
2146 two for the t̄t+0/1/2 j. The qCut parameter defines the  $k_T$  scale for merging shower products into jets.

Table D.1: Generator (MadGraph5) level cuts for the signal and background samples.

<b>MadGraph5</b>	SM ( $h \rightarrow b\bar{b}$ ) & BSM hh	hh & 4b+j (QCD)	4b+j (QCD+EWK)	4b+j (EWK)	jj+0/1/2 j	tt+0/1/2 j
$p_{T,j}^{\min}, p_{T,b}^{\min}$ [GeV]	0	200; 30 500; 30	20; 15	20; 15	20; 5	5; 5
$\eta_j^{\max}, \eta_b^{\max}$	—	5; 5	5; 3	5; 3	8; 8	8; 8
$\Delta R(jj), \Delta R(bb), \Delta R(jb)$	0.001	0.4; 0.1; 0.3	0.4; 0.2; 0.4	0.4; 0.2; 0.4	0; 0.001; 0.001	0.001
xqcut [GeV]	0	0	0	0	20	60
bwcutoff [GeV]	30	15	15	15	30	30
$H_T$	—	—	—	—	0 – 500 500 – 1k 1k – 2k 2k – 4k 4k – 7.2k 7.2k – 15k 15k – 25k 25k – 35k 35k – 100k	—

Table D.2: Pythia settings for the signal and background samples.

<b>Pythia</b>	hh ( $h \rightarrow b\bar{b}$ )	4b+j	jj+0/1/2 j	tt+0/1/2 j
Relevant settings	25:onMode=off 25:onIfAny= 5 -5	—	<b>Jet matching:</b> merge=on scheme=1 setMad=off coneRadius=1.0 etaJetMax=10 nJetMax=4 qCut=30	<b>Jet matching:</b> merge=on scheme=1 setMad=off coneRadius=1.0 etaJetMax=10 nJetMax=2 qCut=60
Description	Turn on the $h \rightarrow b\bar{b}$ decay for the undecayed Higgs, in the case of the SM sample, or for both Higgs in the case of the BSM samples.	—	Set the parameters for jet matching (a detailed description can be found in the text).	

2147 **Appendix E**

2148 **Cutflow tables**

2149 In this appendix we collect the cutflow tables for the baseline and optimized analysis, for particle flow  
2150 and calorimeter jets and for all detector configurations that were tested. The efficiencies of the cuts are  
2151 shown for all cuts. The expected number of events for an integrated luminosity of  $\mathcal{L} = 30 \text{ ab}^{-1}$  is also  
2152 shown for some key cuts.

Table E.1: oi

<b>Configuration</b>	PF jets		HCAL jets	
	Baseline	Optimized	Baseline	Optimized
ATLAS				
ATLAS gran.				
ATLAS gran. $\eta \times 4$				
FCC gran. $\phi/2$				
FCC				
FCC gran. $\eta, \phi \times 2$				

Table E.2: oi

<b>Configuration</b>	PF jets		HCAL jets	
	Baseline	Optimized	Baseline	Optimized
ATLAS				
ATLAS gran.				
ATLAS gran. $\eta \times 4$				
FCC gran. $\phi/2$				
FCC				
FCC gran. $\eta, \phi \times 2$				

Table E.3: oi

<b>Configuration</b>	PF jets		HCAL jets	
	Baseline	Optimized	Baseline	Optimized
<b>ATLAS</b>				
ATLAS gran.				
ATLAS gran. $\eta \times 4$				
FCC gran. $\phi/2$				
<b>FCC</b>				
FCC gran. $\eta, \phi \times 2$				

Table E.4:  $\alpha_i$ 

<b>Selection [ATLAS]</b>	$hh \rightarrow b\bar{b}b\bar{b}$ (SM)	$hh \rightarrow b\bar{b}b\bar{b}$ (DM mediator)	$hh \rightarrow b\bar{b}b\bar{b}$ (2HDM)	$4b + j$	$jj + 0/1/2j$	$t\bar{t}$
Gen level	100	100	100	100	100	100
N(b-tags)						
$p_T(j_1, j_2) \geq 200$ GeV						
$p_T(j_1) \geq 400$ GeV						
$p_T(j_2) \geq 350$ GeV						
$p_T(j_1 + j_2) \geq 100$ GeV						
$\tau_{21}(j_1) < 0.55$						
$\tau_{21}(j_2) < 0.55$						
$FW2(j_1) > 0.2$						
$(100 < M_{SD}(j1) < 135)$ GeV						
$(100 < M_{SD}(j2) < 135)$ GeV						

2153 **Appendix F**

2154 **Extra plots**

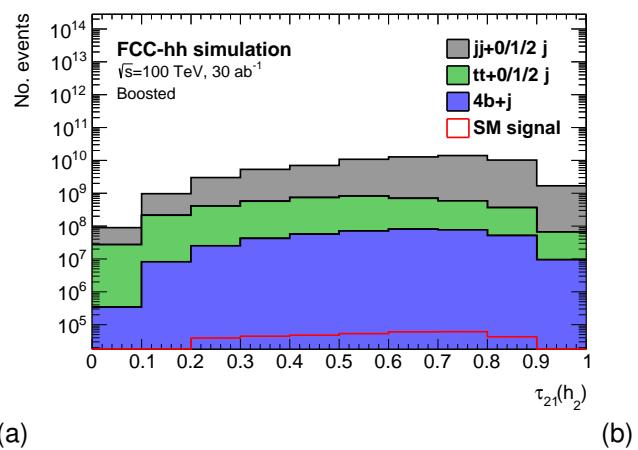


Figure F.1: . The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .