

## **Prospective studies of highly boosted Higgs pairs decaying to four b quarks**

**Ana Luísa Moreira de Carvalho**

Thesis to obtain the Master of Science Degree in

### **Engineering Physics**

Supervisor(s): José Ricardo Morais Silva Gonçalo  
Pedro Morais Salgueiro Teixeira de Abreu

### **Examination Committee**

Chairperson: Jorge Manuel Rodrigues Crispim Romão  
Supervisor: José Ricardo Morais Silva Gonçalo  
Member of the Committee: Patrícia Conde Muíño  
Member of the Committee: Kostas Nikolopoulos

**October 2018**



## Acknowledgments

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

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

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

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

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

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

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

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

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

## Resumo

A produção de pares de bosões de Higgs é um processo chave a ser estudado em aceleradores futuros porque fornece informação sobre o mecanismo de quebra da simetria eletrofraca mas está fora do alcance do LHC.

A construção de um acelerador circular com 100 km de perímetro na área de Genebra está a ser analisada pelo CERN. Espera-se que o futuro acelerador hadrónico (FCC-hh) trabalhe a uma energia do centro de massa (CM) de  $\sqrt{s} = 100$  TeV e que acumule uma luminosidade integrada de  $O(30)$   $\text{ab}^{-1}$ .

Esta tese descreve um estudo de Monte Carlo que analisa o processo  $hh \rightarrow b\bar{b}b\bar{b}$  numa região cinemática de alto momento transverso, usando colisões protão-protão a uma energia do CM de  $\sqrt{s} = 100$  TeV. O trabalho foca-se no impacto da granularidade do calorímetro hadrónico na significância,  $S/\sqrt{B}$ .

Para o FCC-hh, a significância atingida é  $S/\sqrt{B} = 8.8 \pm 1.6$  (stat.)  $^{+4.1}_{-3.0}$  (sys.) para uma luminosidade integrada de  $\mathcal{L} = 30$   $\text{ab}^{-1}$ . Usando jatos reconstruídos usando informação do sistema de traços e do calorímetro, a significância varia aproximadamente 54%. A variação é mais acentuada quando se usam jatos reconstruídos usando apenas a informação do calorímetro ( $\sim 71\%$ ).

**Palavras-chave:** Pares de bosões de Higgs, Futuro acelerador hadrónico circular, estudos de granularidade do calorímetro hadrónico, FCC-hh, estado final com quatro quarks b, upgrade do detetor ATLAS

## Abstract

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

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

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

For the FCC-hh, the achieved significance is  $S/\sqrt{B} = 8.8 \pm 1.6$  (stat.)  $^{+4.1}_{-3.0}$  (sys.) for an integrated luminosity of  $\mathcal{L} = 30$  ab $^{-1}$ , which is above the conventional  $5\sigma$  discovery threshold. When using particle flow jets, the significance changes by approximately 54% over the range of detector configurations considered. The change is more accentuated when using pure calorimeter jets ( $\sim 71\%$ ).

**Keywords:** Higgs pair production, Future Circular Collider, HCAL granularity studies, FCC-hh, four b-quarks final state, ATLAS upgrade



# Contents

Acknowledgments . . . . .	iii
Resumo . . . . .	iv
Abstract . . . . .	v
List of Tables . . . . .	xi
List of Figures . . . . .	xiv
Glossary . . . . .	xviii
<b>1 Introduction</b>	<b>1</b>
<b>2 The standard model and beyond</b>	<b>4</b>
2.1 The Standard Model of Particle Physics . . . . .	4
2.1.1 Higgs pair production . . . . .	10
2.2 Going beyond . . . . .	13
<b>3 Collider experiments</b>	<b>17</b>
3.1 Experimental aspects . . . . .	18
3.2 The Large Hadron Collider . . . . .	18
3.2.1 The ATLAS detector . . . . .	20
3.3 Jet reconstruction . . . . .	23
3.3.1 Jet properties and substructure observables . . . . .	25
3.3.2 Jet grooming algorithms . . . . .	26
3.4 Future Colliders . . . . .	27
3.4.1 The hadronic Future Circular Collider . . . . .	27
3.4.2 FCC-hh baseline detector . . . . .	28
3.4.3 FCC-hh physics program . . . . .	30
<b>4 State of the art</b>	<b>31</b>
4.1 Searches for Higgs pair production at the LHC . . . . .	31
4.2 Feasibility studies for high-luminosity and future colliders . . . . .	32
4.3 Hadronic calorimeter granularity studies . . . . .	34

<b>5 Sample generation and analysis tools</b>	<b>37</b>
5.1 Fast simulation workflow . . . . .	37
5.1.1 Particle flow and calorimeter reconstruction in Delphes . . . . .	38
5.1.2 FCC-hh software . . . . .	39
5.2 Signal and background samples . . . . .	40
5.2.1 MadGraph . . . . .	40
5.2.2 Pythia . . . . .	42
5.2.3 Delphes . . . . .	43
<b>6 Analysis</b>	<b>46</b>
6.1 Overview of the $hh \rightarrow b\bar{b}b\bar{b}$ channel . . . . .	46
6.1.1 Event pre-selection . . . . .	46
6.1.2 Signal characterization . . . . .	47
6.1.3 Backgrounds . . . . .	48
6.2 Analysis strategy . . . . .	52
6.2.1 Implementation of b-tagging . . . . .	52
6.2.2 Baseline analysis . . . . .	53
6.2.3 Optimization . . . . .	56
6.2.4 Handling of uncertainties . . . . .	59
6.2.5 Trigger discussion . . . . .	60
<b>7 Results</b>	<b>62</b>
7.1 Di-Higgs discovery potential at the FCC-hh . . . . .	62
7.1.1 Comparing with the ATLAS detector . . . . .	63
7.2 Hadronic calorimeter granularity studies for future colliders . . . . .	64
7.2.1 Resolution of jet mass and N-subjetiness . . . . .	66
7.2.2 Signal efficiency and significance . . . . .	67
<b>8 Conclusions</b>	<b>72</b>
<b>Bibliography</b>	<b>74</b>
<b>A Jet radius discussion</b>	<b>82</b>
<b>B Additional background processes</b>	<b>83</b>
<b>C Softdrop mass</b>	<b>84</b>
<b>D Samples generation: parameters</b>	<b>85</b>
D.1 MadGraph . . . . .	85
D.2 Pythia . . . . .	85
<b>E Cutflow tables</b>	<b>87</b>





# List of Tables

3.1	ATLAS tile and liquid argon hadronic calorimeters: summary of the pseudorapidity coverages and transversal segmentation (granularity). . . . .	21
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
3.3	Layout, granularity and energy resolution of the proposed hadronic calorimeter for the FCC-hh detector. . . . .	29
5.1	Summary of the effective cross sections ( $\sigma \times BR$ ) and k factors of the signal and background samples used in the analysis. . . . .	44
5.2	Summary of the benchmark granularity configurations of the HCAL. . . . .	45
5.3	Summary of some key detector parameters for the FCC-hh and ATLAS detectors. . . . .	45
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. . . . .	53
7.1	Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM, DM mediator 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. . . . .	63
7.2	Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b+j$ , $jj+0/1/2\ j$ and $t\bar{t}+0/1/2j$ ). 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

7.3	Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM, DM mediator 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. . . . .	65
7.4	Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b+j$ , $jj+0/1/2\,j$ 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. . . . .	65
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 $m_H = 900 \text{ GeV}$ . . . . .	66
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. . . . .	83
D.1	Generator (MadGraph5) level cuts for the signal and background samples. . . . .	86
D.2	Pythia settings for the signal and background samples. . . . .	86
E.1	Summary of the significances obtained with particle flow and calorimeter jets for the baseline and optimized analysis for the SM signal. . . . .	87
E.2	Summary of the significances obtained with particle flow and calorimeter jets for the baseline and optimized analysis for the DM mediator (black) and 2HDM (blue) signal processes. . . . .	87
E.3	Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS default detector, as implemented in Delphes. . . . .	88
E.4	Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity. . . . .	89
E.5	Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity with $\eta \times 4$ . . . . .	90
E.6	Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with $\phi/2$ . . . . .	91

E.7 Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC default detector design. . . . .	92
E.8 Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with $\eta, \phi \times 2$ . . . . .	93
E.9 Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS default detector, as implemented in Delphes. . . . .	94
E.10 Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity. . . . .	95
E.11 Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity with $\eta \times 4$ . . . . .	96
E.12 Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with $\phi/2$ . . . . .	97
E.13 Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC default detector. . . . .	98
E.14 Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with $\eta, \phi \times 2$ . . . . .	99

# List of Figures

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. . . . .	5
2.2 Postulated shape of the Higgs potential for $\mu^2 > 0$ (a) and $\mu^2 < 0$ (b). . . . .	8
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. . . . .	11
2.4 Total cross sections at 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. . . . .	13
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 Ref. [23]. . . . .	14
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], respectively. . . . .	16
3.1 ATLAS detector. . . . .	22
3.2 FCC detector baseline concept. . . . .	30
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 Refs. [70, 74], respectively. . . . .	33
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 Ref. [80] (based on Ref. [79]). . . . .	34

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. [80]. . . . .	36
4.4 $\tau_{32}$ resolution plots for QCD dijet events reconstructed using eflow jets (a) and tower jets (b). Plots from Ref. [81]. . . . .	36
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. [80]. . . . .	36
6.1 (a) Higgs pairs branching ratios, (b) event topology targeted by a boosted analysis. . . . .	47
6.2 (a) $\Delta\phi$ between the Higgs candidates, $\Delta\phi(hh)$ , (b) $\Delta R$ between the b quarks coming from the decay of the leading Higgs candidate, $h_1$ , versus $p_T(h_1)$ . To obtain this plot the cut $ M(h_1, h_2) - 125  < 40$ GeV was applied in addition to the pre-selection cuts. . . . .	49
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. . . . .	50
6.4 Dominant diagrams of $pp \rightarrow t\bar{t}$ at LO. . . . .	51
6.5 $p_T$ distributions for the leading (a) and sub leading (b) Higgs candidates. The signal is the SM $hh \rightarrow b\bar{b}$ process. The histograms are normalized to unit area and include all events that pass the pre-selection cuts. . . . .	51
6.6 $\eta$ (a) and softdrop mass (b) distributions for the leading Higgs candidate. The signal is the SM $hh \rightarrow b\bar{b}$ process. The histograms are normalized to unit area and include all events that pass the pre-selection cuts. . . . .	51
6.7 Distributions of the $\Delta\eta$ between the Higgs candidates (a) and of the $\tau_{21}$ variable for the leading Higgs candidate (b). The histograms are normalized to unit area and include all events that pass the pre-selection cuts. . . . .	52
6.8 Minimum $\Delta R$ between b quarks and subjets of the $R = 0.8$ jets in a $hh \rightarrow b\bar{b}b\bar{b}$ after requiring the existence of at least two large- $R$ jets. . . . .	53
6.9 $p_T$ distributions for the leading (a) and sub leading Higgs candidates (b). The histograms are normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	54
6.10 (a) $p_T$ distribution for the Higgs pair, (b) $\tau_{21}$ variable for the leading Higgs candidate. The histograms are normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	55
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 $\mathcal{L} = 30 \text{ ab}^{-1}$ . . . . .	55
6.12 Negative log-likelihood on $\mu$ . . . . .	56
6.13 $S/\sqrt{B}$ 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 $p_T(h_1)$ . . . . .	58

6.14 $S/\sqrt{B}$ 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 $p_T(hh)$ . . . . .	58
6.15 $S/\sqrt{B}$ 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 $\Delta\eta$ . . . . .	58
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. . . . .	69
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. . . . .	69
7.3 Invariant mass of the Higgs pair for the 1 TeV DM mediator (a) and for the 2HDM with $m_H = 900$ GeV (b). The plots include the events that pass all the analysis cuts. The colors indicate the different detector configurations. . . . .	69
7.4 Leading Higgs candidate $\tau_{21}$ distribution for eflow jets (a) and for HCAL jets (b). The colors indicate the different detector configurations. The distributions are shown for the signal (filled lines) and for the $4b + j$ background (dashed lines). . . . .	70
7.5 Signal efficiency as a function of the detector configuration for particle flow jets (a) and for calorimeter jets (b). 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 statistical error bars are drawn but are smaller than the markers. . . . .	70
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). . . . .	71
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). . . . .	71
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). . . . .	71
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). . . . .	82
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 $\Delta R = 0.8$ . . . . .	84

F.1	$\tau_{21}$ distribution for the sub leading Higgs candidate for the baseline analysis. The histogram is normalized to $\mathcal{L} = 30 \text{ ab}^{-1}$	100
F.2	Sub leading Higgs candidate $\tau_{21}$ distribution for eflow jets (a) and for HCAL jets (b). The colors indicate the different detector configurations. The distributions are shown for the signal (filled lines) and for the $jj + 0/1/2 j$ background (dashed lines).	100
F.3	Signal efficiency as a function of the detector configuration for particle flow jets (a) and for calorimeter jets (b). Three signal models are shown: SM (filled squares), 1 TeV DM mediator (empty squares) and type II 2HDM with $m_H = 900 \text{ GeV}$ (crosses). The statistical error bars are drawn but are smaller than the markers.	101

# 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

# Chapter 1

## Introduction

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

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

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

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

HL-LHC has reached its full discovery potential. One of these projects is the hadronic Future Circular Collider (FCC-hh) that consists of a 100 km ring located in the Geneva area and it is expected to work 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 its ultimate phase. This will result in  $O(30) \text{ ab}^{-1}$  per experiment which corresponds to ten times the expected integrated luminosity by the end of the HL-LHC operation.

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

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

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

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

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

Chapter 2 presents an overview of the SM. It summarizes its particle content and interactions and introduces the mathematical formulation of the EWSB breaking mechanism. The successes and shortcomings of the SM are presented and several BSM models are introduced and their motivations dis-

cussed. Finally, a theoretical description of the production of Higgs pairs is provided.

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

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

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

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

Finally, conclusions are drawn in chapter 8.

# Chapter 2

## The standard model and beyond

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 section 2.2, we motivate the need to explore models beyond the SM and introduce some of the most well known beyond the SM (BSM) models. In section 2.1.1, we provide a theoretical description of the production of Higgs boson pairs which is the physical process that is under study throughout this work.

### 2.1 The Standard Model of Particle Physics

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

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

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

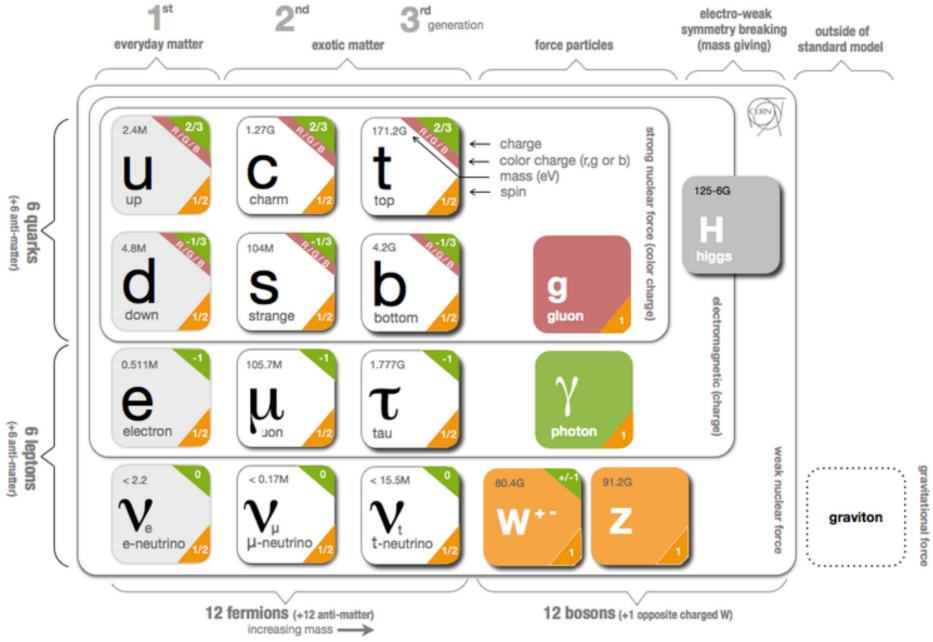


Figure 2.1: Schematic representation of the Standard Model of elementary particles. Particle are represented inside squares. The electric and color charge are shown in the right upper corner of each square. The spin is shown in the right lower corner. The mass of the particles is also shown, in electron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

interactions: strong, electromagnetic, weak and gravitational <sup>1</sup>. Each interaction can be interpreted as the result of the exchange of the corresponding gauge boson. Gluons ( $g$ ) and the photon ( $\gamma$ ) are the mediators of the strong and electromagnetic interactions, respectively. They are massless, electrically neutral and have spin 1. The  $W^\pm$  and  $Z$  bosons are the mediators of the weak interaction and have a mass of 80.4 and 91.2 GeV, respectively. The  $W^+$  and  $W^-$  bosons have electric charges of +1 and -1, respectively and spin 1. The  $Z$  boson is electrically neutral and also has spin 1. The gauge bosons can also be referred to as vector bosons because they have spin equal to one.

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

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

<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.

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

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

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

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

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

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

where the  $SU_{color}(3)$  group describes the strong interactions (QCD) and the  $SU_L(2) \times U_Y(1)$  group describes the electroweak interactions. Here,  $L$  stands for left and  $Y$  stands for hypercharge. In the SM the Higgs mechanism, which we now describe, is realized in the  $SU_L(2) \times U_Y(1)$  group. The Lagrangian 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:

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

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

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

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

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

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

$T^a = \frac{\tau^a}{2}$  (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:

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

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

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), respectively. For  $\mu^2 > 0$  we have a single minimum located at  $\langle \phi \rangle = 0$ . For  $\mu^2 < 0$  the potential has the shape of a 'Mexican hat'. There is an infinite number of minima located in a circumference centered at zero. In this case the minima occur for  $\langle \phi \rangle, \langle \phi^\dagger \rangle \neq 0$ . Therefore the field acquires 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):

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

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 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.

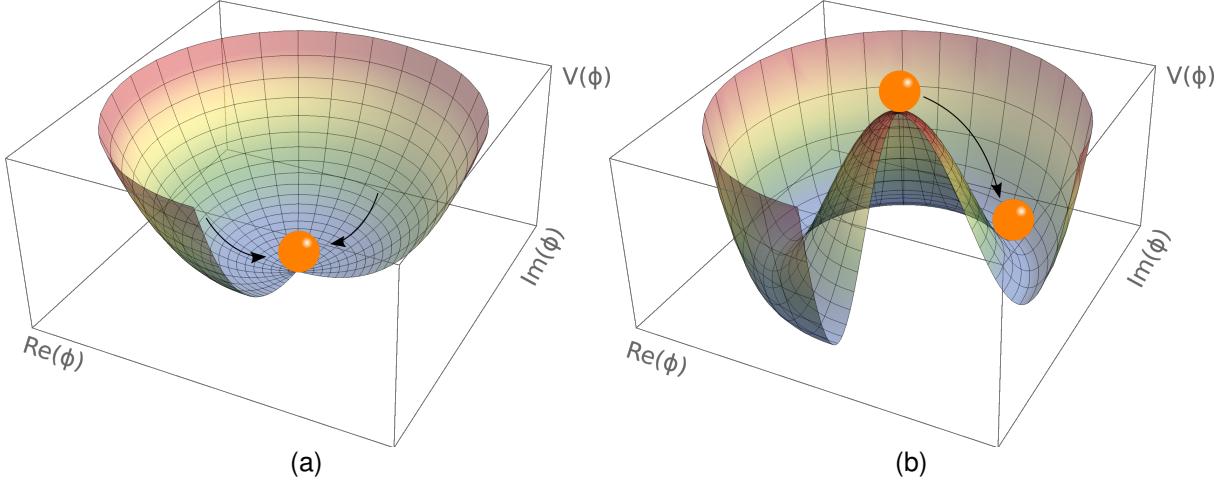


Figure 2.2: Postulated shape of the Higgs potential for  $\mu^2 > 0$  (a) and  $\mu^2 < 0$  (b).

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

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

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

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 fermions, 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 Ref. [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

and bottom quarks) and the second to up type fermions (up, charm and top quarks). In addition, these terms give rise to the interaction terms between the Higgs field and the fermions. Take, as an example,  $\bar{L} = (\bar{t}, \bar{b})_L$  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:

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

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

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

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

### 2.1.1 Higgs pair production

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

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

The diagram on the right has an off-shell (virtual) Higgs boson,  $h^*$ , that couples to gluons by the usual heavy quark triangle (same mechanism as in single Higgs production).  $h^*$  then decays to two on-shell 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).

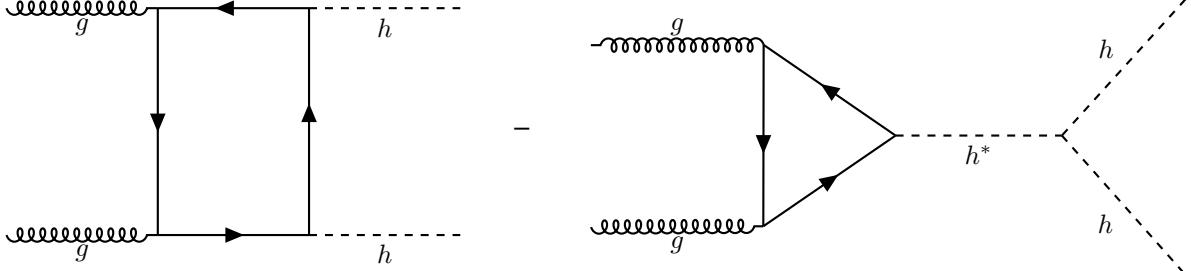


Figure 2.3: Feynman diagrams of Higgs pair production from gluon fusion. Triple vertex diagram (left) and box diagram (right). The minus sign between the diagrams indicates that they interfere destructively.

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 Ref. [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.

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

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

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

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

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

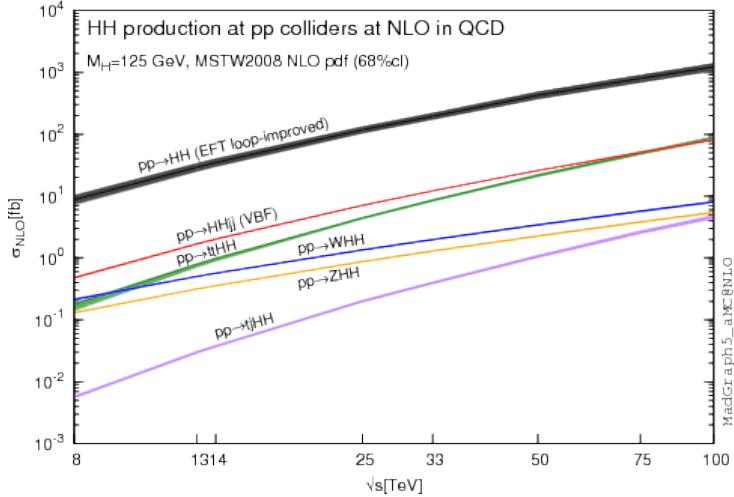


Figure 2.4: Total cross sections at NLO in QCD for the six largest HH production channels at pp colliders. The thickness of the lines corresponds to the scale and PDF uncertainties added linearly.

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

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

## 2.2 Going beyond

Despite the success of the SM, there is evidence that indicates that it cannot be the final theory of particle physics. This led to the development of alternative models that extend the SM but that can still reproduce its successful predictions. These are referred to as Beyond the Standard Model (BSM) models.

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

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

When faced with these weaknesses, or rather hints of incompleteness, the theoretical community put a great effort into the development of models that add new ingredients to the SM. In the following

---

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

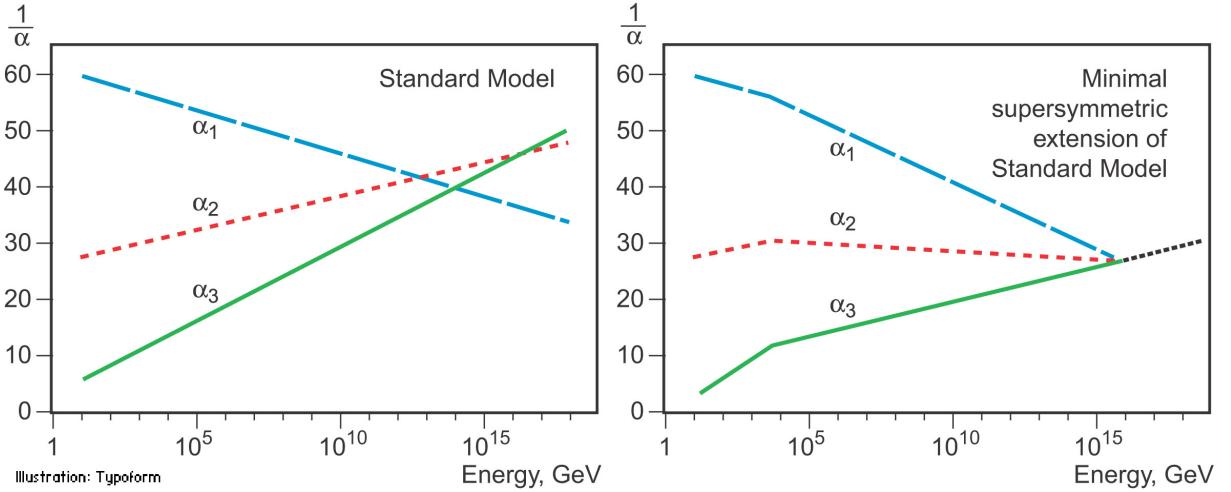


Figure 2.5: Here,  $\alpha_1$ ,  $\alpha_2$  and  $\alpha_3$  represent the coupling constants of  $U_Y(1)$ ,  $SU_L(2)$  and  $SU_{color}(3)$ , respectively. What is shown is the variation of the inverse of the coupling with the energy: on the left for the SM and on the right for the MSSM. Plots from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

paragraphs we introduce and briefly describe some of the most well studied (both theoretically and experimentally) BSM models. We follow the discussion presented in Ref. [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 are usually known as Kaluza-Klein gravitons (and their excited states). Nonetheless, the Kaluza-Klein

model does not provide a satisfactory explanation for the hierarchy problem. Therefore, in 1999, Lisa Randall and Raman Sundrum introduced a new model that does. This model introduces only two new particles: a spin 2 graviton (and its Kaluza-Klein excitations) and a radion, that is a spin 0 neutral particle.

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

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

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

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

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

Instead of the parameters in Eq. 2.26, we can describe the model in terms of the four physical 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, 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 Higgs doublets. The angle  $\alpha$  diagonalizes the square mass matrix of the CP-even states. The quartic couplings of the potential can then be written in terms of these parameters, as can be found in Eq. 11 of Ref. [24].

Simplified dark matter (DM) models are based on the exchange of a single particle between DM and SM particles and try to explain the nature of DM and how it interacts with the SM. The particle exchanged is called a mediator and, depending on the specific model, it can be neutral or electrically charged and have spin 0, 1 or 2. The simplest possible scenario is a neutral scalar mediator.

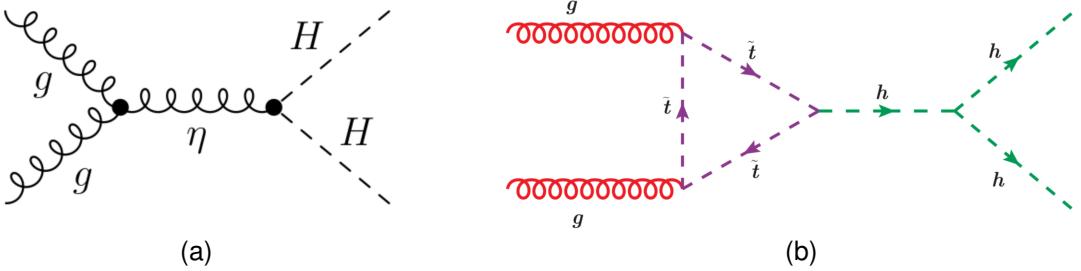


Figure 2.6: BSM contributions to Higgs pair production from the spin-2 Kaluza-Klein graviton (a) and from the SUSY partner of the top quark (b). Figures from Refs. [26, 27], respectively.

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

The spin-2 graviton predicted by Kaluza-Klein and Randal-Sundrum models can couple directly to gluons and then decay producing a Higgs pair, as it illustrated in figure 2.6(a). SUSY particles can contribute to the quantum loops in the Higgs production Feynmann diagrams. An example is shown in figure 2.6(b), where the top quark is replaced by its supersymmetric partner is the s-channel diagram. These are two other examples of how BSM models can change the Higgs pair production process with respect to the SM.

The crucial point is that some BSM contributions can lead to an enhancement of the cross section for Higgs pair production with respect to what is predicted by the SM. Experimentally, this means that we would not need as much sensitivity and therefore the process could be measured with less data and therefore sooner. This is the reason why a lot of the searches performed at the LHC focus on this type of scenario. In addition, if a new heavy particle couples to the Higgs boson via an s-channel diagram, it could lead to the existence of a peak in the Higgs pair invariant mass spectrum (assuming that we have enough experimental resolution and that there are not other processes coming into play).

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

# Chapter 3

## Collider experiments

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

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

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

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

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

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

### 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:

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

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

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

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

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

### 3.2 The Large Hadron Collider

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

1954, CERN has housed many accelerators and experiments and played a key role in the development of fundamental and applied science.

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

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

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

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

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

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

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

### 3.2.1 The ATLAS detector

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

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

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

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

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

The hadronic calorimetry in the pseudorapidity range  $|\eta| < 1.7$  is provided by a scintillator-tile 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 $2.5 <  \eta  < 3.2$	$0.2 \times 0.2$

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

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

The granularity of the ATLAS hadronic calorimeters is summarized in table 3.1. The hadronic calorimeters provide measurements of the energy of hadrons and jets; these allow the estimation of the missing transverse energy ( $E_T^{miss}$ ). Approximately one third of the energy of jets is deposited in this layer. In the TileCal, the jet energy resolution is given by  $\sigma/E \sim 50\%/\sqrt{E} + 3\%$  [42], where the first term is the stochastic term that derives from sampling fluctuations and follows a Poisson distribution and the second term is a constant that depends on the characteristics of the calorimeter. For the LAr calorimeter, the jet energy resolution is given by  $\sigma/E \sim 60\%/\sqrt{E} + 2\%$  [43].

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

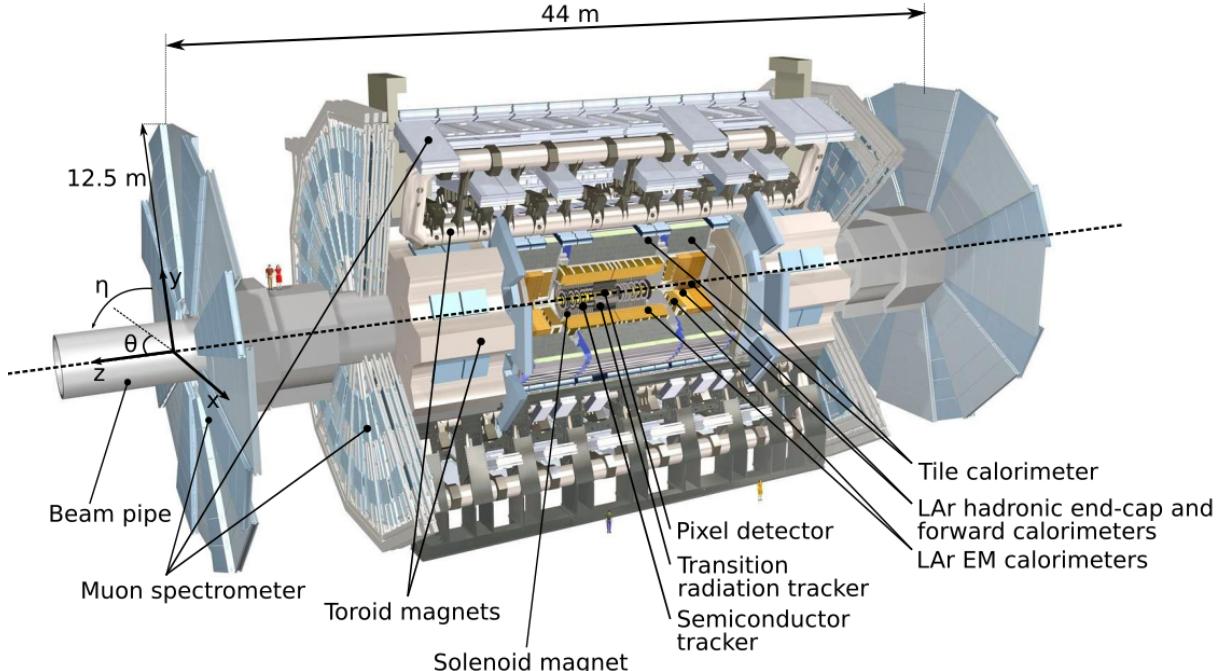


Figure 3.1: ATLAS detector.

### b-Tagging

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

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

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

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

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

### Trigger and data acquisition

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

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

## 3.3 Jet reconstruction

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

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

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

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

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

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

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

referred to as a distance between entities  $i$  and  $j$  with  $d_{iB}$  being the distance between entity  $i$  and the 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 azimuth of entity  $i$ . Here,  $k_T$  stands for the transverse momentum. For  $p = 1, 0, -1$  we get the  $k_T$ , C/A and anti- $k_T$  algorithms, respectively.

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

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

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

## Boosted kinematic regime

With the increase of the CM energy of particle colliders the production of particles with a transverse momentum much larger than their mass became a reality. In this kinematic regime, referred to as boosted due to the high Lorentz boost of the particles, traditional jet reconstruction algorithms, that rely on a one-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 96%, 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 a quark 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

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

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

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

We usually use ratios of  $\tau_N$  variables ( $\tau_{MN} = \tau_M / \tau_N$ ). Of particular interest for this work is the  $\tau_{21}$  ratio. This observables can take values between zero and one. A small value (close to zero) indicates that the jet is more compatible with two subjets than with one and therefore it can help discriminate between two-prong and one-prong jets.

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:

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

where  $\theta_{ij}$  is the opening angle between energy clusters  $i$  and  $j$  with respect to the interaction point,  $E$  is the total energy of the clusters in the jet rest frame and  $P_l(x)$  are the Legendre polynomials. For a jet 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  $H_l \sim 0$  for odd  $l$ . This is what we expect for jets coming from the decay of a boosted resonance. In this work we 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:

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

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

### 3.3.2 Jet grooming algorithms

The main goal of jet grooming algorithms is to remove contamination of softer jet constituents from pileup or underlying event and to leave behind the hard substructure. The main advantage of such algorithms is

that they improve the mass resolution of jets. These features are of particular interest in high luminosity environments such as the HL-LHC and future high energy colliders.

The three main jet grooming algorithms are trimming, pruning and mass drop filtering. In this work we do not use pruning or trimming, although these techniques might be worth exploring in future, more comprehensive, studies. In particular, they can be useful to help reject pileup contributions. The mass drop filtering procedure isolates relatively symmetric subjets within a jet, each with a significantly smaller mass than the original jet [53]. This technique was developed and optimized using C/A jets for 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:

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

where  $y_{\text{cut}}$  is a parameter that defines the energy sharing between the subjets. It is usually taken 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.

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

## 3.4 Future Colliders

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

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

### 3.4.1 The hadronic Future Circular Collider

CERN's FCC study group was launched as a result of a recommendation made in the 2013 update of the European Strategy for Particle Physics that '*Europe needs to be in a position to propose an*

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

*ambitious post-LHC accelerator project at CERN by the time of the next Strategy update’.* It investigates the technological challenges and physics opportunities of a future circular collider. The design and infrastructure are driven by a proton-proton collider (FCC-hh) requirements. Electron-positron (FCC-ee) and electron-proton (FCC-eh) colliders are also being analyzed. The main goal of this effort is to deliver a Conceptual Design Report (CDR) by the end of 2018. This document will include a first cost estimate to be submitted for the next update of the European Strategy for Particle Physics, foreseen by 2019.

The FCC-hh baseline design consists of a proton-proton circular collider with a maximum CM energy of  $\sqrt{s} = 100$  TeV housed by a 100 km tunnel in the area of Geneva. This machine will extend the research program of the LHC (and of the HL-LHC) after these have reached their full discovery potential, by around 2040. In addition, it will allow for the exploration of an entirely new kinematic regime, probing energy scales where new physics may come into play. A possible way of defining the target luminosity of this machine is to require that within the first year of operation it surpasses the exploration potential of the LHC [57]. Comprehensive studies [57, 58] indicate that this can be achieved with an integrated luminosity of the order of  $10 \text{ ab}^{-1}$  per experiment. Considering a reasonable operation period of 10 years this leads to integrated luminosity per experiment of the order of  $1 \text{ ab}^{-1}$  per year. In this work, we consider a slightly more optimistic scenario that sets the target integrated luminosity at  $30 \text{ ab}^{-1}$  [59].

The FCC-hh is expected to work with a dipole field of 16 T and to provide a peak instantaneous luminosity thirty times larger than the LHC. The number of bunches is expected to be almost a factor of four larger than for the LHC and the number of events per bunch crossing is expected to be approximately 1000. The latter brings a lot of technical challenges because it means that the mean pileup expected for the FCC-hh is almost 40 times larger than for the LHC. This requires the development of techniques and algorithms that allow us to further reject pileup contributions (or improvement of already existing techniques). Some relevant parameters of the LHC and FCC-hh are summarized in table 3.2.

### 3.4.2 FCC-hh baseline detector

The design of the FCC-hh baseline detector, which we describe in detail in this section, has been greatly based on that of the ATLAS and CMS experiments, in particular the central barrel. The layers and sub detectors are arranged in the same order and perform very similar roles. The geometry is cylindrical and therefore we can use exactly the same coordinate system that was introduced in section 3.2.1. The

Table 3.3: Layout, granularity and energy resolution of the proposed hadronic calorimeter for the FCC-hh detector.

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 ( $\sigma_E/E$ )	$40\%/\sqrt{E} \oplus 2.5\%$	$50\%/\sqrt{E} \oplus 3\%$	$100\%/\sqrt{E} \oplus 5\%$

dimensions are very close to the ones of ATLAS: 25 meters in height (diameter) and 48 m in length [60]. A schematic representation of the FCC detector is shown in figure 3.2.

The detector consists of trackers, EM and hadronic calorimeters and MS. The magnet configuration consists of three solenoid magnets (one central-barrel and two forward) that surround the central barrel calorimeters and the forward trackers. The center solenoid delivers a magnetic field of 4 T at the interaction point [61].

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

The EM calorimeter [62, 63] covers the pseudorapidity range  $|\eta| < 6$ . It is divided in barrel, end-cap and forward. These cover the pseudorapidity ranges  $|\eta| < 1.5$ ,  $1.4 < |\eta| < 2.5$  and  $2.3 < |\eta| < 6$ , respectively. The proposed layout for the EM calorimeter is a LAr sampling configuration with lead, glue and steal plates as absorbers. The granularity is expected to be two to four times better than for the ATLAS ECAL. For the barrel calorimeter, the goal energy resolution is  $10\%/\sqrt{E} \oplus 1\%$ .

The hadronic calorimeter [64] covers the pseudorapidity range  $|\eta| < 6$ . It is also divided in barrel, end-cap and forward that cover the pseudorapidity ranges  $|\eta| < 1.3$ ,  $1.0 < |\eta| < 1.8$  and  $2.3 < |\eta| < 6.0$ , respectively. The proposed layout for the barrel calorimeter consists of scintillator tiles interleaved with lead and stainless steel plates as absorbers. The end-cap and forward calorimeters are expected to be based on liquid argon with copper plates as absorbers. For the barrel and end-cap calorimeters, the expected segmentation  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  while for the forward calorimeter it is  $\Delta\eta \times \Delta\phi = 0.05 \times 0.05$ . Overall, this corresponds to approximately four times the ATLAS HCAL granularity. The energy 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-cap and forward calorimeters, respectively. The pseudorapidity coverage, layout, granularity and energy resolution of the hadronic calorimeters are summarized in table 3.3.

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

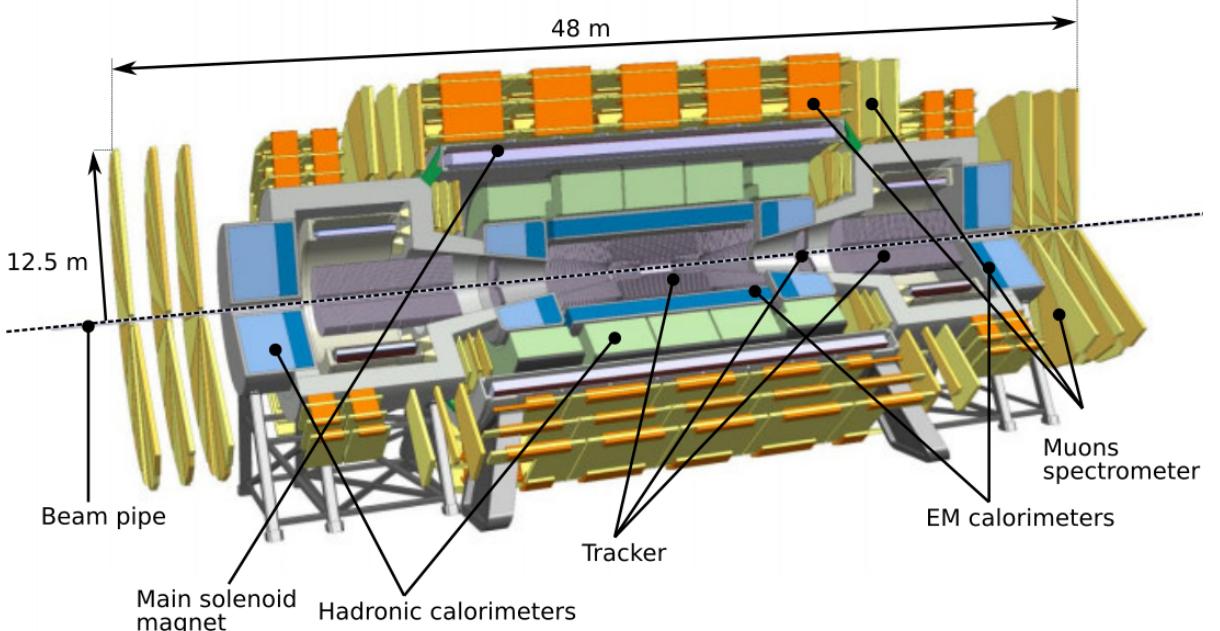


Figure 3.2: FCC detector baseline concept.

### 3.4.3 FCC-hh physics program

The FCC-hh physics program is vast and diverse. Within the SM, it includes the study of gauge bosons pair production and heavy flavor production, the measurement of the top quark's properties and the the study of the EWSB mechanism via multi-Higgs production. In addition, heavy ions collisions will 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 in Ref. [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 achieve using the  $b\bar{b}\gamma\gamma$  final state. This would allow to constraint the Higgs triple coupling to be  $\lambda_{hh} \in [0.97, 1.03]$ . 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  $\lambda_{hh} \in [0.9, 1.5]$ . 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  $\lambda_{hhh} \in [-4, 16]$ .

# Chapter 4

## State of the art

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

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

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

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

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 addition, ATLAS showed that this is the most sensitive channel to resonance masses over 500 GeV [67]. However, this channel suffers with an overwhelming multijet background which drives the need for very tight trigger level cuts in order to bring the event rates down to manageable values.

The  $hh \rightarrow b\bar{b}\tau^+\tau^-$  analysis [68, 69] benefits from a sizable branching fraction of  $h \rightarrow \tau^+\tau^-$  ( $\sim 7.3\%$ ) and from a relatively small background contribution from other SM processes. These searches target

the semi-leptonic decay of the  $\tau\tau$  pair to reduce contamination from QCD processes.

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 diphoton triggers and isolation criteria that greatly reduce multijet background. In addition the excellent mass resolution of  $h \rightarrow \gamma\gamma$  can be exploited. The  $hh \rightarrow b\bar{b}WW^*$  channel also benefits from the large branching fraction of  $h \rightarrow W^+W^-$  ( $\sim 21\%$ ).

The most stringent upper limit on the cross section of Higgs pair production comes from a combination of three searches using up to  $36.1 \text{ fb}^{-1}$  of proton-proton collision data at a CM energy of  $\sqrt{s} = 13 \text{ TeV}$  collected with the ATLAS detector [73]. The combination is based on the three most sensitive decay 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 limit on the non-resonant Higgs boson pair cross-section is  $0.223 \text{ pb}$  at  $95\% \text{ CL}$ , which corresponds to 6.7 times the cross-section predicted by the SM. The ratio of the Higgs boson self-coupling to its SM expectation,  $k_\lambda$ , is constrained, at  $95\% \text{ CL}$ , to be  $-5.0 < k_\lambda < 12.1$ .

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

The upper limits on the cross section (at  $95\% \text{ CL}$ ) of Higgs pair production as a function of the mass of a heavy scalar Higgs,  $m_H$ , are summarized in figure 4.1(b) [74]. The limits were obtained through a 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  $hh \rightarrow \gamma\gamma WW^*$  channels using proton-proton collision data corresponding to an integrated luminosity of  $20.3 \text{ fb}^{-1}$  collected with the ATLAS detector at a CM energy of  $\sqrt{s} = 8 \text{ TeV}$ . The improvement above  $m_H = 500 \text{ GeV}$  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  $0.015 \text{ pb}$ .

## 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

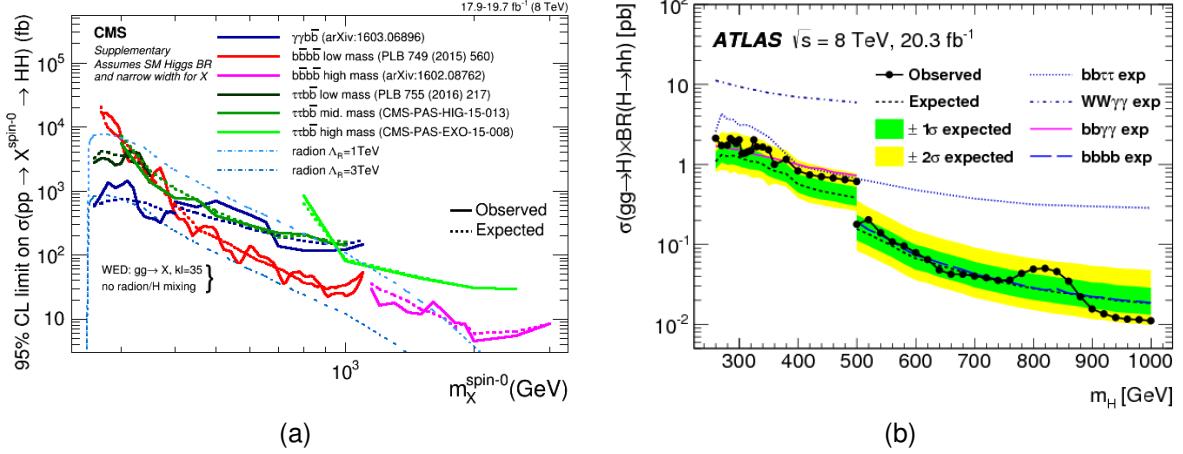


Figure 4.1: Exclusion limits (at 95% CL) on the cross section of resonant Higgs pair production from the decay of heavy scalar particles. These results were obtained using proton-proton collision data collected with the CMS (a) and ATLAS (b) detector at a CM energy of  $\sqrt{s} = 8$  TeV and result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

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  $Z \rightarrow b\bar{b} h \rightarrow b\bar{b} b\bar{b}$ ,  $t\bar{t} h \rightarrow b\bar{b} b\bar{b}$  and  $b\bar{b} h \rightarrow b\bar{b} b\bar{b}$ , 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  $1.05\sigma$  [76] and  $1.6\sigma$  [77], 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 di-Higgs cross section of approximately twice the SM value. This corresponds to an exclusion limit of  $-0.8 < k_\lambda < 7.7$ . The analysis conducted by ATLAS in this channel is done at generator level. The energy and momenta of the particles are smeared to simulate the detector's response. A mean pile up of 200 is considered. No MVA techniques are employed in the analysis.

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 be excluded at 95% CL, with systematic uncertainties being taken into consideration. The Higgs trilinear coupling is expected to be constrained to  $-3.5 < k_\lambda < 11$ . These results are based on the extrapolation of the current results obtained with the 2016 dataset, comprising an integrated luminosity of  $10.1 \text{ fb}^{-1}$ . This study is based only on the resolved analysis documented in Ref. [65].

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 only the irreducible background  $pp \rightarrow b\bar{b}b\bar{b}j$  reports a significance of 6.61 for an integrated luminosity

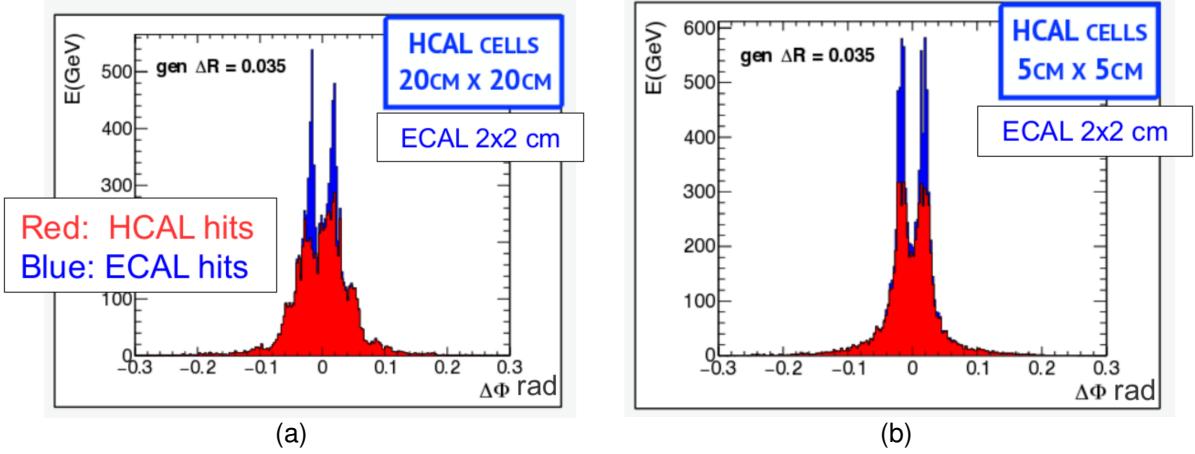


Figure 4.2: Azimuthal distribution of the energy deposits in the ECAL and HCAL for a pair of  $K_L^0$  with  $E = 100$  GeV for an hadronic calorimeter with  $20\text{ cm} \times 20\text{ cm}$  (a) and  $5\text{ cm} \times 5\text{ cm}$  (b). Figures from Ref. [80] (based on Ref. [79]).

of  $30\text{ ab}^{-1}$ , considering an analysis that targets the boosted region [59]. The extra jet in the signal sample has  $p_T > 200$  GeV which provides the Higgs pair with a large Lorentz boost. This enhances the sensitivity to the Higgs boson self coupling because it favors highly boosted virtual Higgs bosons decaying to a pair of Higgs bosons. The analysis relies on the jet substructure observable  $\tau_{21}$  and on a tight mass cut around the Higgs mass. No multivariate techniques are employed.

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 the FCC-hh was presented in Ref. [78]. Only the irreducible background is considered. For a boosted region, cut based analysis, a significance of 1.1 is reported for an integrated luminosity of  $3\text{ ab}^{-1}$  at  $\sqrt{s} = 14$  TeV (HL-LHC). For an integrated luminosity of  $10\text{ ab}^{-1}$  at  $\sqrt{s} = 100$  TeV, which corresponds to the FCC-hh, this number is 5.7. While the significance is large for the FCC-hh, the reported signal to background ratio ( $S/B$ ) is approximately one order of magnitude smaller, which means the measurement might be more sensitive to systematic uncertainties on the backgrounds.

### 4.3 Hadronic calorimeter granularity studies

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

The granularity of hadronic calorimeters is a key parameter for future collider detectors because it determines how well we can resolve energy deposits from pileup vertices and highly-boosted jet topologies. In Ref. [79], the use of smaller calorimeter cells (smaller than the ones that are used in currently operating hadronic calorimeters) to resolve individual hadrons is investigated. For two Kaons ( $K_L^0$ ) with an energy of 100 GeV each and with a truth level separation ( $\Delta R$ ) equal to 0.035 the energy deposited in the ECAL (blue) and HCAL (red) is shown as a function of  $\Delta R$  for HCAL cells with sizes  $20\text{ cm} \times 20\text{ cm}$

$(\Delta\eta \times \Delta\phi = 0.1 \times 0.1)$  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), respectively. The ECAL segmentation is equal to  $2 \text{ cm} \times 2 \text{ cm}$ . It is shown that for a granularity of  $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$  both hadrons can be resolved in the HCAL.

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. [80–82]. 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 given 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  $p_T(\text{jet}) > 3 \text{ TeV}$  is shown in figure 4.3(a) [81]. 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 [81]. 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. [80], 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  $p_T$  of  $2.5 \text{ TeV}$  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  $5(10) \text{ TeV}$  jets the overlap varies from  $88(91)\%$  to  $78(85)\%$ . However, for  $20 \text{ TeV}$  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 substructure observables. We did not find any studies that explored the change in the significance of a given analysis as a function of the granularity or of the detector configuration.

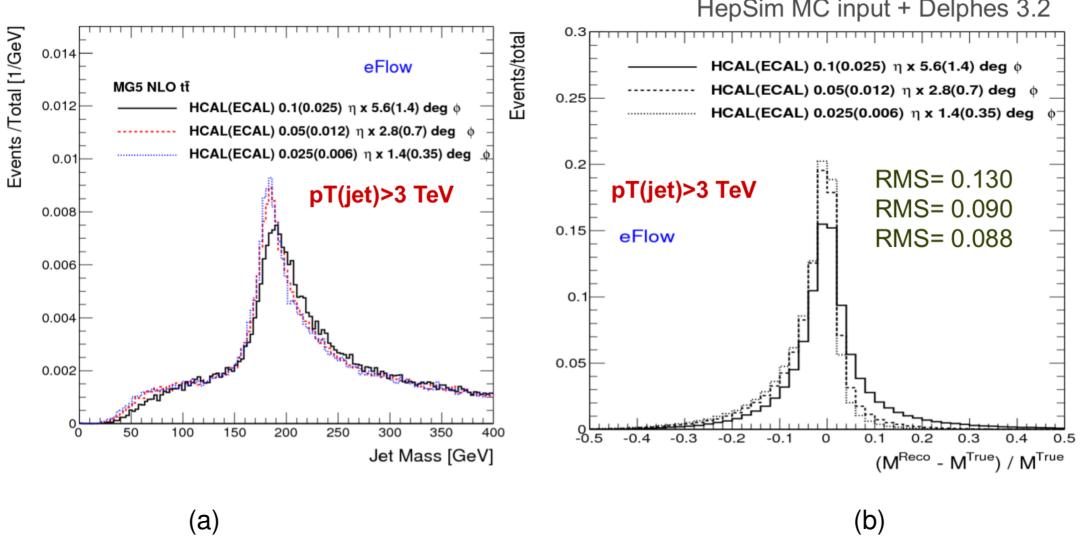


Figure 4.3: Jet mass (a) and jet mass resolution (b) plots in  $t\bar{t}$  events for eflow jets with  $p_T > 3$  TeV for three different HCAL and ECAL configurations: HCAL(ECAL) 0.1(0.025)  $\eta \times 5.6(1.4)$   $\circ \phi$ , HCAL(ECAL) 0.05(0.012)  $\eta \times 2.8(0.7)$   $\circ \phi$  and HCAL(ECAL) 0.025(0.006)  $\eta \times 1.4(0.35)$   $\circ \phi$ . Plots from Ref. [80].

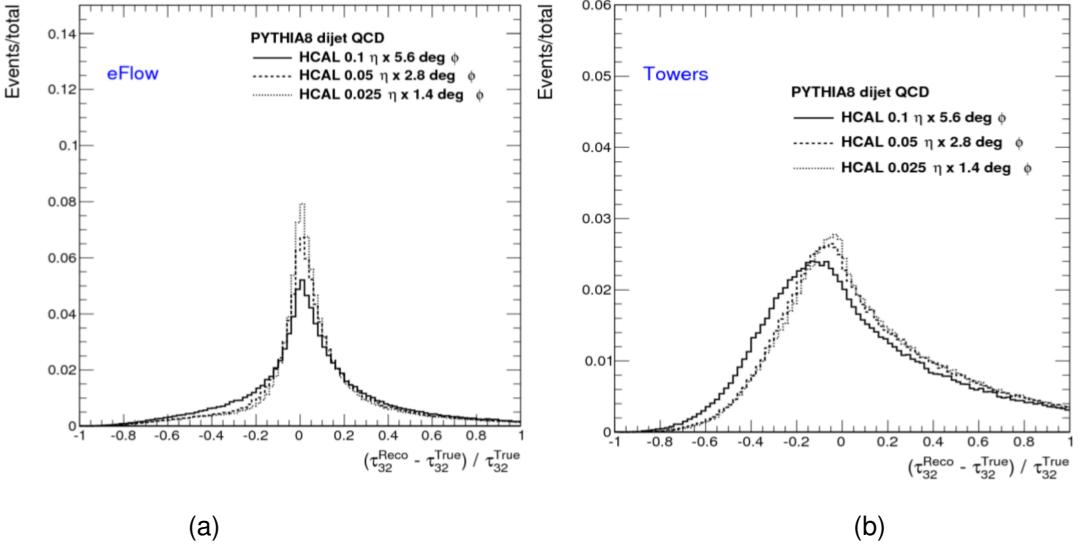


Figure 4.4:  $\tau_{32}$  resolution plots for QCD dijet events reconstructed using eflow jets (a) and tower jets (b). Plots from Ref. [81].

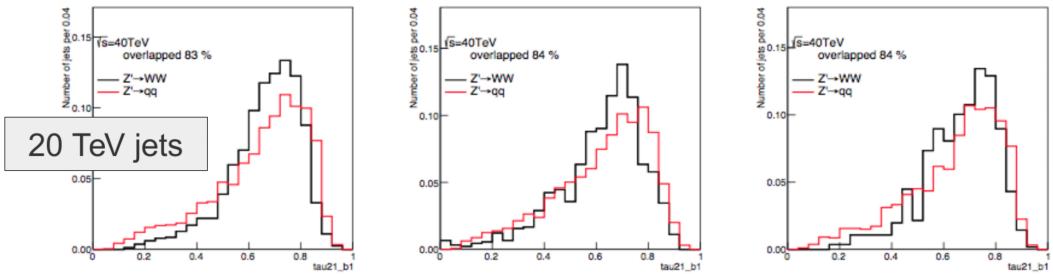


Figure 4.5:  $\tau_{21}$  variable for 20 TeV  $W$  (black) and QCD jets (red) for three HCAL granularities:  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$  (left),  $\Delta\eta \times \Delta\phi = 0.022 \times 0.022$  (middle) and  $\Delta\eta \times \Delta\phi = 0.005 \times 0.005$  (right). Plots from Ref. [80].

# Chapter 5

## Sample generation and analysis tools

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

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

### 5.1 Fast simulation workflow

The Monte Carlo samples used in this study were generated using MadGraph5\_aMC@NLO [83]. The showering and hadronization are simulated using Pythia 8 [84] and the detector response is parametrized using Delphes 3 [85]. These samples are available at the CERN EOS storage.

MadGraph5 is a matrix element generator for high energy physics processes, such as decays and scatterings. The user specifies the desired process in terms of the initial and final states and can impose additional constraints such as allowing for a number of refined criteria, including forced or forbidden s-channel resonances, excluded internal particles, and forced decay chains of final state particles [83]. As a result, MadGraph automatically generates the corresponding Feynman diagrams and creates the necessary code to compute the matrix element at a given point of the phase space. The output file is written in the Les Houches Accord [86].

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

Delphes3 allows for a quick and simple simulation of the detector's response [85]. Its goal is to allow the fast simulation of a multipurpose detector for phenomenological studies. The simulation includes a track propagation system, electromagnetic and hadronic calorimeters and a muon identification system. Low level physics objects, such as tracks and energy deposits, and high level physics objects, such as leptons and jets, are reconstructed from the detector's response and can be used to perform physics analysis. In the following paragraphs we briefly describe how the detector response is simulated and how jet reconstruction is performed.

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

Jets can be produced using generator level long-lived particles after showering and hadronization, tracks, calorimeter towers or particle-flow tracks and towers. These are referred to as generator, track, calorimeter or particle-flow jets, respectively. For generator level jets no detector simulation nor reconstruction are taken into account. In spite of the type of jet, the user can choose the jet clustering algorithm and the values of its parameters as well as the minimum transverse momentum of the jets that are stored in the final collection. Delphes integrates the FastJet package [87] and therefore allows jet reconstruction with the most popular jet reconstruction algorithms, namely, anti- $k_T$ ,  $k_T$  and Cambridge-Aachen (C/A). Jets resulting from the hadronization of a b quark (known as b jets), are identified if a b quark is found within a  $\Delta R$  distance from the jet's axis. The tagging efficiency and mis-tagging probabilities can be defined by the user.

### 5.1.1 Particle flow and calorimeter reconstruction in Delphes

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

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

Therefore the spatial resolution is limited by the transverse segmentation of the calorimeters.

The goal of the particle flow approach is to make use of all the available information provided by the various sub-detectors for reconstructing an event [85]. This approach is used by some experimental collaborations [88, 89] but the exact implementation depends on the specificities of the experiment. If the momentum resolution of the tracking system is better than the energy resolution of the calorimeters it might be convenient to use the tracking information to estimate the momentum of charged particles. In real experiments, the tracking resolution is only better than the calorimeter's energy resolution up to some energy threshold. However, in Delphes, it is assumed that it is always convenient to estimate the momentum of charged particles via the tracker.

The particle flow algorithm works as follows [85]. For each calorimeter tower it counts:

- the total energy deposited in ECAL and HCAL,  $E_{\text{ECAL}}$  and  $E_{\text{HCAL}}$ , respectively;
- the total energy deposited in ECAL and HCAL originating from charged particles for which a track 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

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

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

For example [85], a single charged pion is reconstructed as a track with energy  $E_{\text{HCAL,trk}}$  and deposits some energy  $E_{\text{HCAL}}$  in the hadronic calorimeter. If  $E_{\text{HCAL}} \leq E_{\text{HCAL,trk}}$ , only a particle flow track with energy  $E_{\text{HCAL,trk}}$  is produced. If  $E_{\text{HCAL}} > E_{\text{HCAL,trk}}$ , a particle flow track with energy  $E_{\text{HCAL,trk}}$  and a particle flow tower with energy  $E_{\text{HCAL}}$  are produced.

With this algorithm [85], particle-flow tracks contain charged particles estimated with a good resolution, while the particle-flow towers contain in general a combination of neutral particles, charged particles with no corresponding reconstructed track and additional excess deposits induced by the positive smearing of the calorimeters, and are characterized by a lower resolution. The particle flow approach produces high resolution inputs for jet clustering algorithms and can also be useful to reduce pileup contributions.

### 5.1.2 FCC-hh software

FCC software (FCCSW) [90], common to all FCC studies (electron-electron, electron-hadron and hadron-hadron) has been developed and is maintained by the FCC study group. The software is based on Gaudi [91]. An FCC Event Data Model based on Podio [92] was also developed. It consists in specific classes that encode the information about the events.

The FCC-hh study group is responsible for the generation of Monte Carlo quick simulation samples for the main benchmark processes for the FCC-hh and corresponding backgrounds. The samples are generated using the workflow described in the previous section. CERN users can request rights to run

the EventProducer package [93] and produce samples for any desired process using the machinery that is already implemented. In addition, the FCC Event Data Model classes are directly accessible and can be used to read the ROOT files that are produced after the events are passed through Delphes.

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

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

## 5.2 Signal and background samples

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

### 5.2.1 MadGraph

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

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

Note that for the  $4b+j$  (QCD) and  $jj+0/1/2 j$  samples we do not take into account the regions of the phase space with  $p_{T,j} < 200$  GeV and with  $0 < H_T < 500$  GeV. We assume that we can reject most of the events (if not all) with these kinematic characteristics by going to a sufficiently boosted region of the

phase space. In addition, note that after the showering procedure there could be some overlap between the  $4b + j$  and  $jj + 0/1/2 j$ . This is taken care of in our analysis code: if an event from the  $jj + 0/1/2 j$  sample has four b quarks at truth level (which happens for approximately 0.01% of the events) then we do not consider it because it will certainly overlap with an event from the  $4b + j$  sample.

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

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

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

## Signal samples - BSM

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

Both models are readily available in FeynRules [94] model database and can be straightforwardly implemented in Madgraph5. The parameters of the models, namely the masses of the new particles, can be changed by the user. In the case of this work, we want new particles to have a large mass so that they produce SM Higgs pairs with a high Lorentz boost.

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

For the 2HDM [95, 96], a finer tuning of the parameters is required because the model is very general and has many free parameters. In addition, it is written in the Higgs basis such that we need to convert the parameters from the physical basis (introduced in section 2.2) to this basis. The free parameters of the model are the masses of the neutral,  $m_{h_1}, m_{h_2}, m_{h_3}$ , and charged scalars,  $m_{hc}$ , the neutral scalars mixing angles,  $mix_h, mix_{h_2}, mix_{h_3}$ , the real and imaginary parts of the up, down and charged lepton  $3 \times 3$  Yukawa matrices,  $GUR, GUI, GDR, GDI, GLR, GLI$  and the values (real and imaginary parts) of the second, third and seventh quartic couplings,  $l_{2,3,7,.}$ . In the CP-conserving model  $h_{1,2,3}$  can be identified with the physical states  $h, H$  and  $A$ , respectively.

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

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

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

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

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 Ref. [24] and 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 Ref. [97] to obtain the values of the quartic coupling in the Higgs basis.

Regarding the Yukawa interactions, we work in the type II model. Following the type II restriction file found in Ref. [95] we set

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

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 (with the same final states) being counted twice (double counting). Take, for example,

the processes  $j\,j$  and  $j\,j\,j$  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 approximate number of events for the samples used in the analysis are summarized in table 5.1. The k-factor multiplies the cross section times branching ratio in order to reproduce known higher order results. It corresponds only to the ratio between the total cross sections and it does not correct for possible differences that might exist between the differential cross sections. The number of events changes according to the granularity configuration and to the type of jets that were used. The numbers shown should be taken only as an order of magnitude.

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 [98].  $\sigma(hh, h \rightarrow b\bar{b})$  is given by:

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

where  $\sigma_{NNLO}(hh) = 1.22$  pb follows from Ref. [99] and the factor of 2 is a combinatorial factor that indicates that both Higgs bosons can decay to a pair of b quarks. The k-factor is given by  $\sigma_{NNLL}/\sigma_{NNLO}$ , where  $\sigma_{NNLL}$  is calculated according to the prescription given in equation I.7.8 of Ref. [98], with  $\delta_t = -0.315$ , which yields  $\sigma_{NNLL} = 1.33$  pb. Therefore, the k-factor has the value 1.09.

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

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

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

For the remaining samples,  $\sigma \times BR$  are the values given by MadGraph. In these samples no decay mode is imposed, therefore the branching ratio is one. For the  $4b + j$ (QCD) samples, a k-factor of two is applied [93] in order to parameterize our ignorance on the QCD irreducible background. For the  $t\bar{t} + 0/1/2\,j$ , a k-factor of 1.74 is applied, following Ref. [93].

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

### 5.2.3 Delphes

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

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	$\sim$ no. events
$hh \rightarrow b\bar{b}b\bar{b}$ - <b>SM</b>	0.827	1.09	$5 \times 10^6$
$hh \rightarrow b\bar{b}b\bar{b}$ - <b>DM mediator</b>	0.218	1.0	$5 \times 10^6$
$hh \rightarrow b\bar{b}b\bar{b}$ - <b>2HDM type II</b>	0.466	1.0	$5 \times 10^6$
$4b + j$ (QCD, $200 < p_T^j < 500$ )	756.4	2.0	$10 \times 10^6$
$4b + j$ (QCD, $p_T^j > 500$ )	57.71	2.0	$10 \times 10^6$
$4b + j$ (QCD+EWK)	6.204	1.0	$10 \times 10^6$
$4b + j$ (EWK)	0.07206	1.0	$5 \times 10^6$
$jj + 0/1/2 j$ ( $500 < H_T < 1000$ )	$1.64 \times 10^7$	1.0	$10 \times 10^6$
$jj + 0/1/2 j$ ( $1000 < H_T < 2000$ )	$1.67 \times 10^6$	1.0	$1 \times 10^6$
$jj + 0/1/2 j$ ( $2000 < H_T < 4000$ )	$1.32 \times 10^5$	1.0	$1 \times 10^6$
$jj + 0/1/2 j$ ( $4000 < H_T < 7200$ )	$7.32 \times 10^3$	1.0	$1 \times 10^6$
$jj + 0/1/2 j$ ( $7200 < H_T < 15000$ )	$4.75 \times 10^2$	1.0	$1 \times 10^6$
$jj + 0/1/2 j$ ( $15000 < H_T < 25000$ )	7.35	1.0	$1 \times 10^6$
$jj + 0/1/2 j$ ( $25000 < H_T < 35000$ )	0.176	1.0	$5 \times 10^5$
$jj + 0/1/2 j$ ( $35000 < H_T < 100000$ )	0.00765	1.0	$3 \times 10^5$
$t\bar{t} + 0/1/2 j$	$4.31 \times 10^4$	1.74	$5 \times 10^6$

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

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

In addition, we also passed the same generator level samples through the default ATLAS detector simulation in Delphes. The HCAL granularity is the one that is indicated in the second row of table 5.2 but the other detector parameters, such as the radius, magnetic field, tracking resolutions are the ones that are implemented in the default ATLAS Delphes card. This additional detector configuration gives us an extra point in the space of parameters that we are trying to explore. Furthermore, in a first, very crude, approximation, it allows us to compare the results obtained at  $\sqrt{s} = 100$  TeV to the ones obtained

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 $1.7 <  \eta  < 3.2$

at  $\sqrt{s} = 13$  TeV by ATLAS. For completion we summarize, in table 5.3, the values of some key detector parameters as implemented in Delphes for the ATLAS and FCC-hh detectors. The magnetic field is twice as strong for the FCC-hh. The charged hadrons tracking efficiency and the HCAL resolution are fairly similar between the two detector configurations.

# Chapter 6

## Analysis

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

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 approximately 33.6%, making it the most probable decay for Higgs pairs, as is illustrated in figure 6.1(a). However, in this channel, the main background is QCD multijet production that has a cross section several orders of magnitude larger than di-Higgs production in the SM, as table 5.1 shows. Nonetheless, the jet  $p_T$  distributions in this background have a very large yield close to zero and then fall very steeply while the signal has a much larger tail to high values of  $p_T$  (see figure 6.5(a)). This indicates that searches targeting the boosted kinematic regime may be the key to measure  $hh \rightarrow b\bar{b}b\bar{b}$  using inclusive production.

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

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

#### 6.1.1 Event pre-selection

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

We require at least two b-tagged  $R = 0.8$  jets. A  $R = 0.8$  jet is considered b-tagged if it has at least two b-tagged subjets. The subjets are found through the soft drop mass technique that was described in section 3.3.2. The leading and sub leading jets must have  $p_T \geq 200$  GeV in order for the event

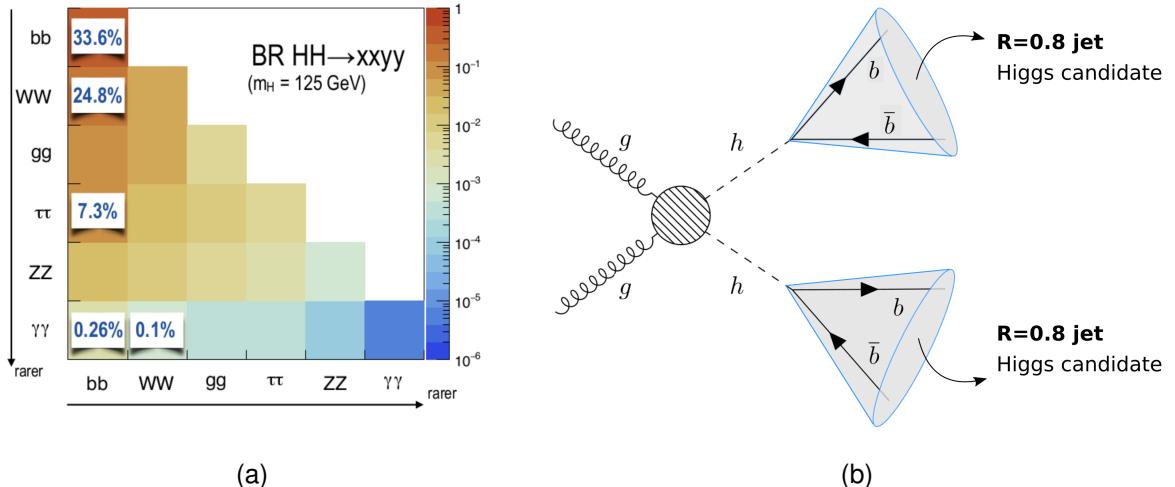


Figure 6.1: (a) Higgs pairs branching ratios, (b) event topology targeted by a boosted analysis.

to be accepted. From Eq. 3.4, a Higgs boson with, e.g.,  $p_T = 200$  GeV leads to a pair of b quarks with  $\Delta R \sim \frac{2m}{p_T} = 1.25$ . Therefore, considering jets with  $R = 0.8$  and  $p_T \geq 200$  GeV 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  $\sqrt{s} = 100$  TeV is 1.22 pb [99]. This value is computed at NNLO accuracy. The dominant production 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 times branching ratio 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 system. 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 jet 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 Higgs boson, with larger  $\Delta R$  corresponding to Higgs bosons with a lower momentum. This correlation between the  $\Delta R$  between the subjets of the leading Higgs candidate ( $\Delta R(b\bar{b})$ ) and its momentum ( $p_T(h_1)$ ) is shown in figure 6.2(b), for the signal sample, after the pre-selection cuts and after an additional mass cut of  $|M - 125| < 40$  GeV on both Higgs candidates.

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

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

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

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

### 6.1.3 Backgrounds

The relevant backgrounds for this analysis are QCD multijet production,  $t\bar{t}$  and  $b\bar{b}b\bar{b}$ . Although the  $b\bar{b}b\bar{b}$  background is a particular case of a QCD multijet production process we consider it separately because

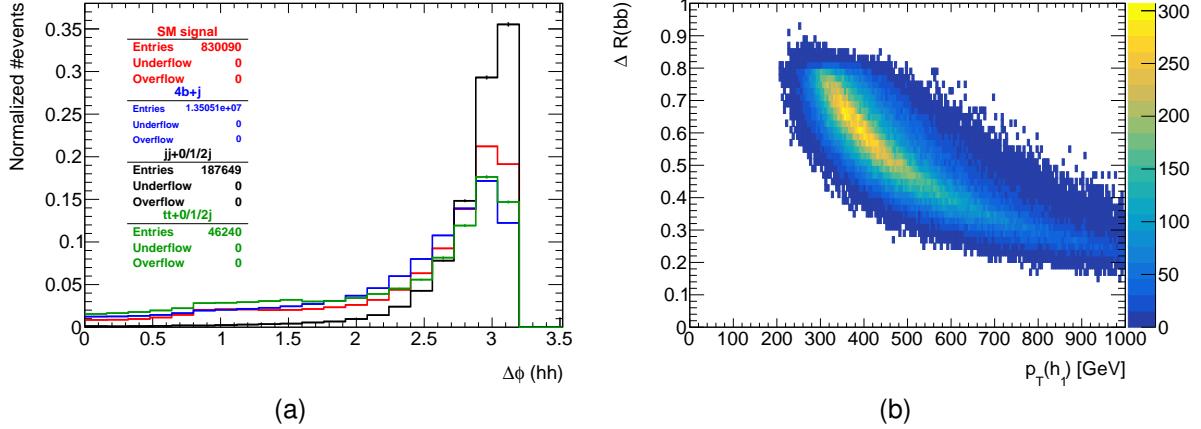


Figure 6.2: (a)  $\Delta\phi$  between the Higgs candidates,  $\Delta\phi(hh)$ , (b)  $\Delta R$  between the b quarks coming from the decay of the leading Higgs candidate,  $h_1$ , versus  $p_T(h_1)$ . To obtain this plot the cut  $|M(h_1, h_2) - 125| < 40$  GeV was applied in addition to the pre-selection cuts.

it constitutes the irreducible background and therefore will have a higher efficiency in the analysis. The cross sections for these processes are several orders of magnitude larger than the cross section for the signal, as shown in table 5.1. In addition, in the case of the  $t\bar{t}$  background, the event topology is expected to be similar to the signal because it also consists in the production of a pair of particles with the same mass.

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

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

The QCD multijet background consists of two, three and four jet events (represented as  $jj+0/1/2j$ ), at generator level. The 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. Due to non-zero mistagging probabilities, light and c jets can be identified as b jets. Although these probabilities are relatively small when compared to the b-tagging efficiency (see section 6.2.1) the cross section of multijet processes is

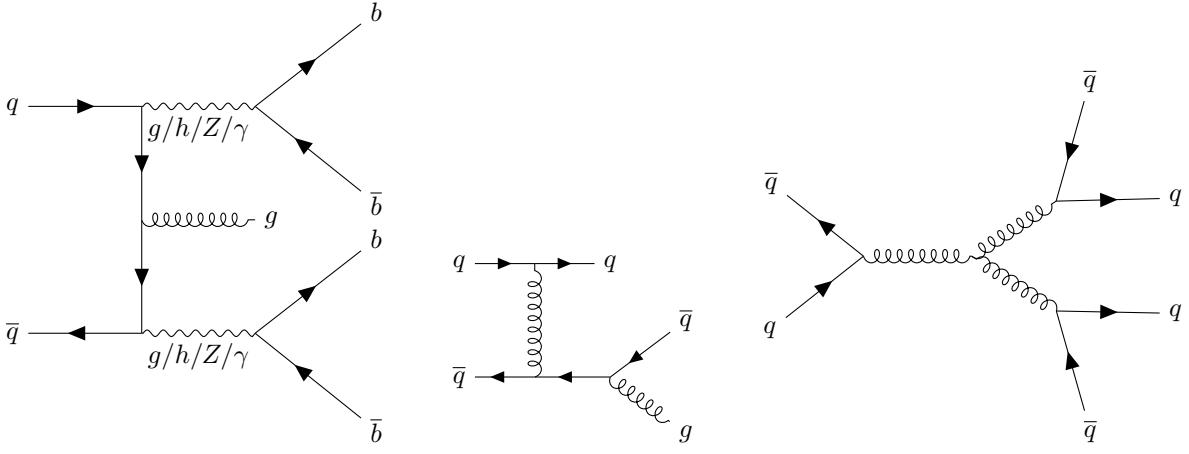


Figure 6.3: Example of diagrams that contribute to the QCD multijet background: five final state jets, four of which are b-jets (left), three final state jets (middle) and four final state jets (right). Here,  $q$  stands for a light quark/jet.

very large, such that this background becomes dominant. The  $jj + 0/1/2 j$  sample with ( $500 < H_T < 1000$ ) GeV has a cross section of the order of  $10^7$  pb. As anticipated, the  $p_T$  spectrum of the multijet background is a lot softer than that of the signal and remaining backgrounds, as is shown in figures 6.5(a) and 6.5(b). In 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 containing 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, c or b quarks or from gluons. A jet matching procedure is implemented in Pythia in order to avoid double counting. The top quark has a very short lifetime, 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% [100]. Therefore,  $t\bar{t}$  events will, most of the time, 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  $\text{BR}(W^+ \rightarrow q\bar{q}) \sim 68\%$  and  $\text{BR}(W^+ \rightarrow l^+\nu) \sim 10\%$ . If one (or both)  $W$  bosons decays to hadrons then there will be additional jets in the final state. These can be b initiated jets (although the probability is very small) or can be light jets misidentified as b jets. 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, a 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 distribution for the leading Higgs candidate (figure 6.6(b)) the  $t\bar{t}$  background (green line) shows a small peak at 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.

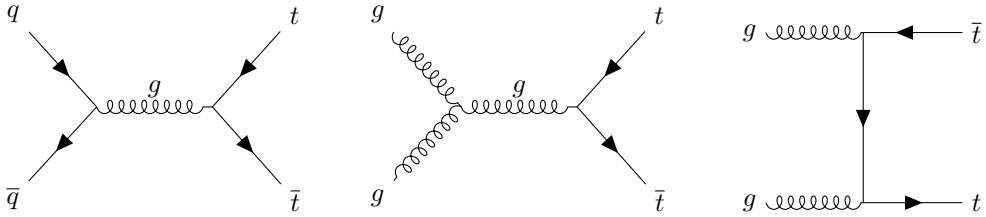


Figure 6.4: Dominant diagrams of  $pp \rightarrow t\bar{t}$  at LO.

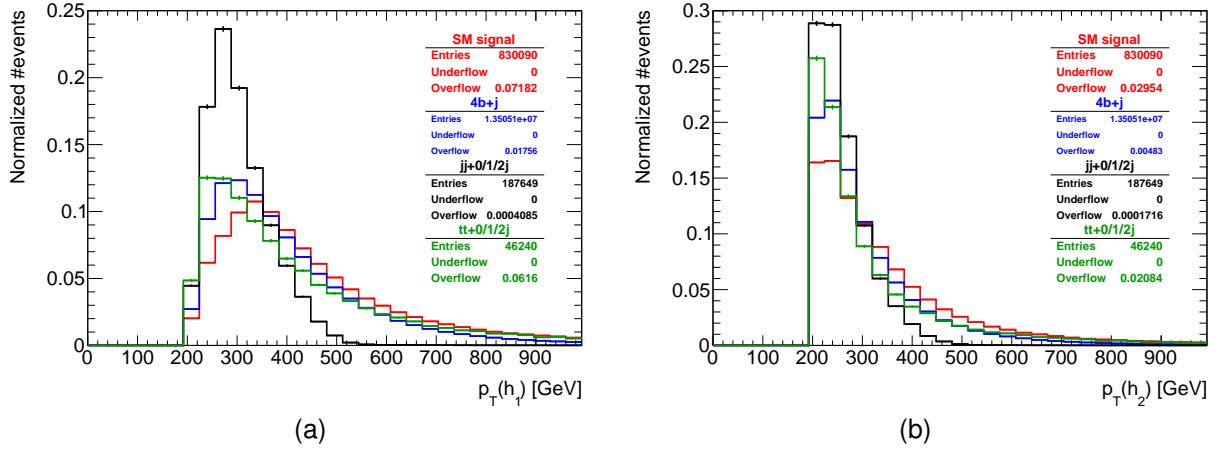


Figure 6.5:  $p_T$  distributions for the leading (a) and sub-leading (b) Higgs candidates. The signal is the SM  $hh \rightarrow b\bar{b}$  process. The histograms are normalized to unit area and include all events that pass the pre-selection cuts.

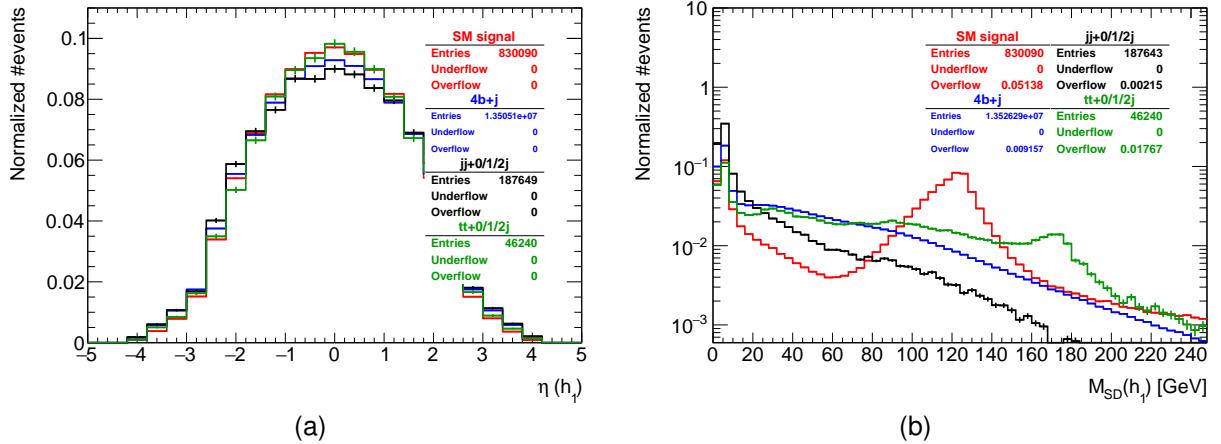


Figure 6.6:  $\eta$  (a) and softdrop mass (b) distributions for the leading Higgs candidate. The signal is the SM  $hh \rightarrow b\bar{b}$  process. The histograms are normalized to unit area and include all events that pass the pre-selection cuts.

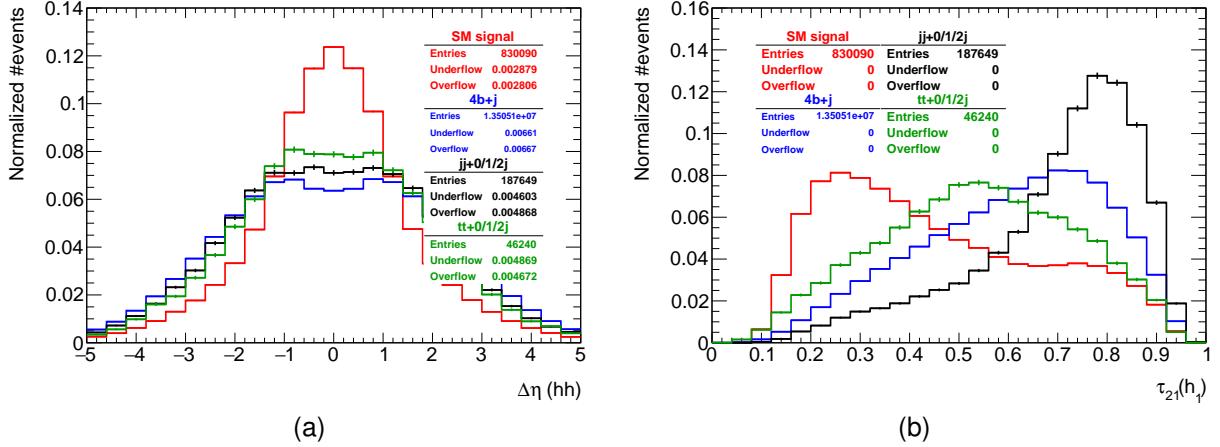


Figure 6.7: Distributions of the  $\Delta\eta$  between the Higgs candidates (a) and of the  $\tau_{21}$  variable for the leading Higgs candidate (b). The histograms are normalized to unit area and include all events that pass the pre-selection cuts.

## 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 pure calorimeter jets with  $R = 0.8$ , clustered with the anti- $k_T$  algorithm. We perform the b-tagging of jets using truth level information. Jets with a large- $R$  parameter cannot be b-tagged using the Delphes default algorithm. 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 kinematic and substructure variables. An optimized version of the analysis is also presented and compared to the baseline. The optimization is based on the significance ( $S/\sqrt{B}$ ) as a function of the cut threshold for a given variable. We also describe how statistical and systematic uncertainties are estimated.

### 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. For each event, we compute the  $\Delta R$  distance between all b and c quarks with Pythia status equal to 23 [101] and with  $p_T > 10$  GeV and each subjet ( $\Delta R(\text{subjet,parton})$ ). Particles with status 23 result directly from the hardest subprocess. We consider that a subjet is matched to a given quark if  $\Delta R(\text{subjet,parton}) < 0.3$ , based on figure 6.8. If the subjet is matched to at least one 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-jet mistag rate. This

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; 0.05; 0.01	$(0.85; 0.05; 0.01) \times (1 - p_T/15000)$
$2.5 <  \eta  < 4.0$	0.64; 0.03; 0.0075	$(0.64; 0.03; 0.0075) \times (1 - p_T/15000)$

allows for a more realistic modeling of the imperfect b-tagging performance, where light quark or c quark initiated jets may fake b jets. 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 summarized in table 6.1. The b-tagging probabilities are given in black and the c and light mis tagging probabilities are given in blue and red, respectively. Note that a jet cannot be b-tagged if  $|\eta| > 4$  or if its momentum is smaller than 10 GeV or larger than 15000 GeV.

In terms of the technical implementation, the b-tagging algorithm works as follows: for each subjet we look for a truth-level b quark within  $\Delta R = 0.3$  of the subjet and calculate the b-tagging and mistagging efficiencies using the expressions in table 6.1, where  $p_T$  and  $\eta$  refer to the subjet. If a b quark is found we generate a random number between 0 and 1. If this number is smaller than the b-tagging efficiency, we consider the 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 subjet. If one is found we generate a new random number between 0 and 1 and if the number is smaller than the c mistag probability we consider the subjet to be b-tagged. If the subjet is still not b-tagged we generate a random number between 0 and 1 and consider the subjet b-tagged if the number is smaller than the light-jet mistag probability.

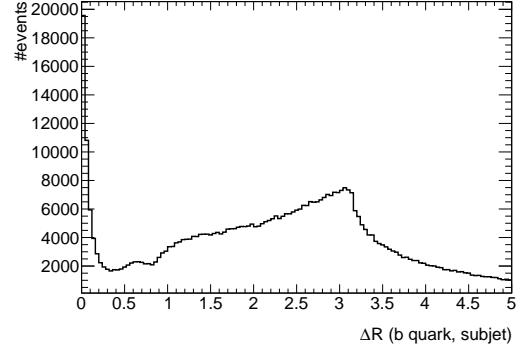


Figure 6.8: Minimum  $\Delta R$  between b quarks and subjets of the  $R = 0.8$  jets in a  $hh \rightarrow bb\bar{b}\bar{b}$  after requiring the existence of at least two large- $R$  jets.

## 6.2.2 Baseline analysis

The analysis described in this section and in section 6.2.3 were developed using the samples simulated with the default FCC-hh detector implementation. The same event selection was then applied to the samples generated using the different detector configurations.

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,  $p_T(hh)$ :

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

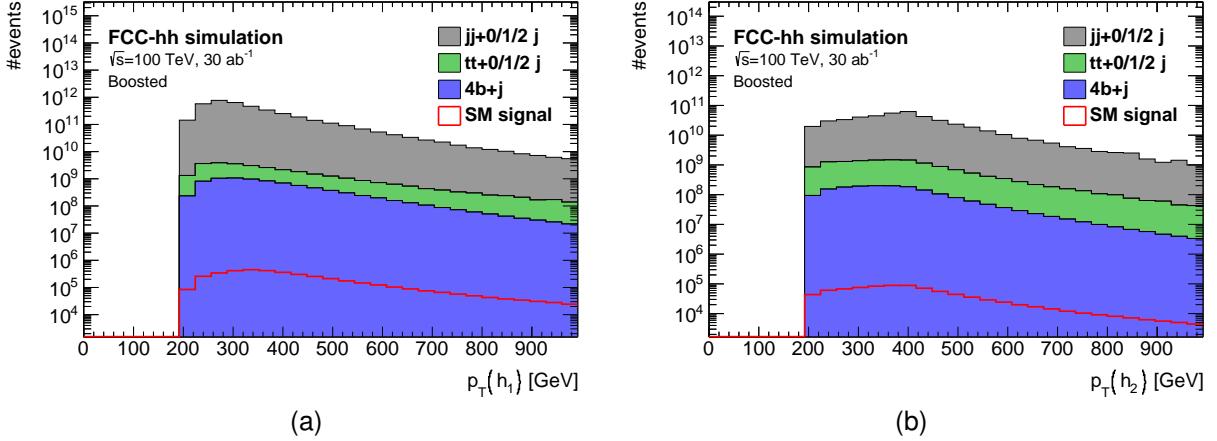


Figure 6.9:  $p_T$  distributions for the leading (a) and sub leading Higgs candidates (b). The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

The effect of these cuts may be appreciated from the distributions shown in figures 6.9(a), 6.9(b) and 6.10(a) and guarantee that we are choosing events for which the Higgs candidates are sufficiently boosted. In addition, a high threshold for the  $p_T$  of the jets suppresses  $t\bar{t}$  events because the decay products of the top quark are reconstructed as a single jet.

We then apply a cut on the  $\tau_{21}$  of the leading Higgs candidate, illustrated in figure 6.10(b), to select jets that 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,  $M_{SD}(h_1, h_2)$ . The distribution of  $M_{SD}(h_2)$  is similar to this one so we do not show it here. 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  $5.1 \pm 0.5 \text{ (stat.)}^{+2.5}_{-1.8} \text{ (sys.)}$  ( $1.62 \pm 0.16 \text{ (stat.)}^{+0.8}_{-0.6} \text{ (sys.)}$ ) for an integrated luminosity of  $30 (3) \text{ ab}^{-1}$ .

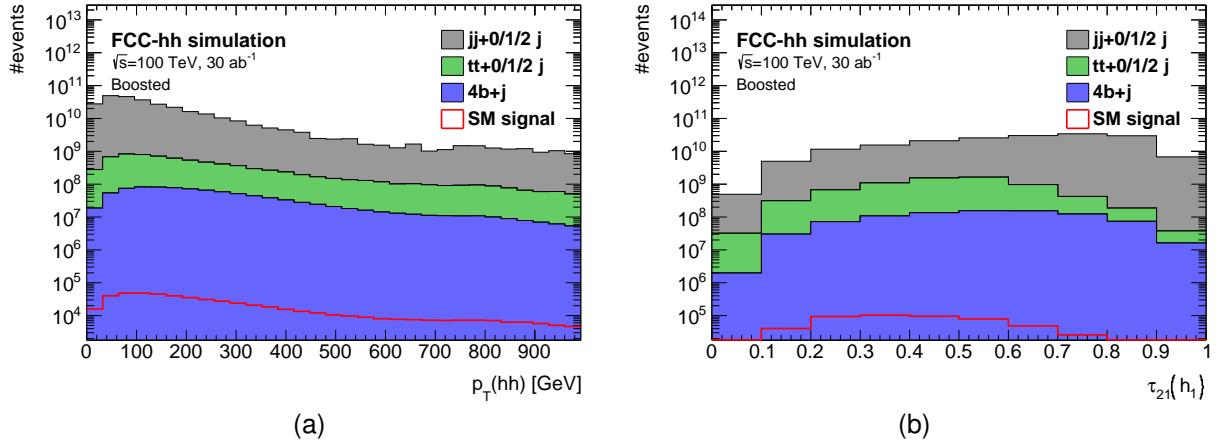


Figure 6.10: (a)  $p_T$  distribution for the Higgs pair, (b)  $\tau_{21}$  variable for the leading Higgs candidate. The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

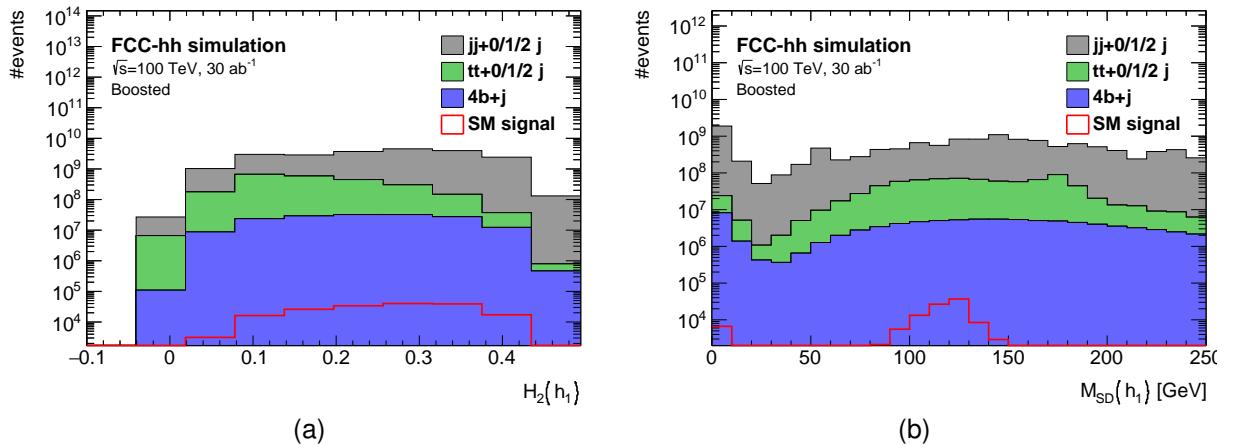


Figure 6.11: (a)  $H_2$  variable for the leading Higgs candidate, (b) softdrop mass distribution for the leading Higgs candidate. The histograms are normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

### Signal strength extraction

To extract the value of the expected signal strength we perform a negative log-likelihood fit to the invariant mass spectrum of the di-Higgs pair after the full event selection of the baseline analysis. Following Eq. 39.16 from Ref. [100] we minimize the function:

$$-2 \ln \lambda(\mu) = -2 \ln \frac{\mathcal{L}(\mu)}{\mathcal{L}(\hat{\mu})} = 2 \sum_{i=1}^{N_{\text{bins}}} \left[ (\mu_i) - n_i + n_i \ln \left( \frac{n_i}{\mu_i} \right) \right] \quad (6.5)$$

where  $\mathcal{L}(\mu) = \prod_{i=1}^{N_{\text{bins}}} \frac{\mu_i^{n_i} e^{-\mu_i}}{n_i!}$  is the likelihood function, calculated as product over all bins of Poisson probabilities with mean  $\mu_i$ .  $\mathcal{L}(\hat{\mu})$  is the likelihood function calculated for the values of  $\mu_i$  that maximize it, which corresponds to setting  $\mu_i = n_i$  in its definition. In addition, we take  $\mu_i = \mu \times s_i + b_i$ , where  $\mu$  is the fit parameter and corresponds to the signal strength,  $s_i$  and  $b_i$  are the expected number of signal and background events extracted from the MC distributions;  $n_i$  is the number of 'observed' events. Since we are not working with real data, the values of  $n_i$  are extracted from a distribution of pseudo data, generated from the signal and background MC distributions. For each bin, a random number is generated following a Poisson distribution with mean equal to the number of events in that bin. This number is taken to be the number of events in the same bin of the pseudo data distribution. This is done for all bins in the distribution and it corresponds to one toy experiment. The results presented are obtained from  $10k$  toy experiments.

The negative log-likelihood on the  $\mu$  parameter is shown in figure 6.12 assuming nominal background yield (blue line) and varying the background normalization up (red line) and down (green) according to the systematic uncertainties described in section 6.2.4. These plots were obtained considering an integrated luminosity of  $30 \text{ ab}^{-1}$ . For the nominal background yield, the expected signal strength is  $\mu = 1.0 \pm 0.4$  which corresponds to a precision of 40%. The impact of the systematic uncertainties on the precision of the signal strength is  $\sim 10\%$ .

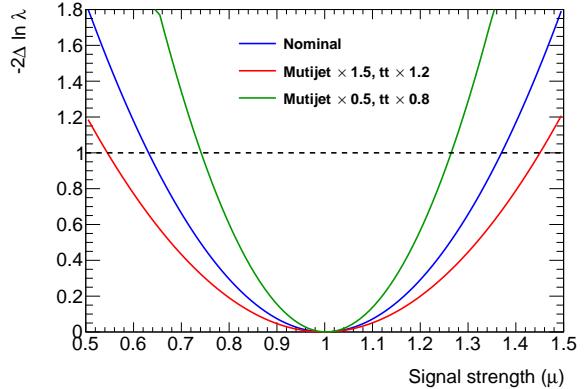


Figure 6.12: Negative log-likelihood on  $\mu$ .

### 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 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 6.13(a). Based on this plot we choose the cut:

$$p_T(h_1) > 300 \text{ GeV}. \quad (6.6)$$

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 6.13(b) and 6.14(a), respectively. From figure 6.13(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 figure 6.14(a) we choose the cut:

$$p_T(hh) > 100 \text{ GeV}. \quad (6.7)$$

Following the cuts on the momentums 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 below that cut. The plot is shown in figure 6.14(b). From the figure we place the following cut on the  $\tau_{21}$  of the leading Higgs candidate:

$$\tau_{21}(h_1) < 0.4. \quad (6.8)$$

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

$$|\Delta\eta(hh)| < 1.5 \quad H_2(h_1) > 0.2. \quad (6.9)$$

These cuts follow from the plots on figures 6.15(a) and 6.15(b). The optimization plot for  $\Delta\eta(hh)$  is obtained by placing a cut in a window around zero which follows directly from the shape of the distributions in figure 6.7(a). For the  $H_2(h_1)$  variable the cut is placed on the lower value.

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

Using the optimized analysis we obtain a significance,  $S/\sqrt{B}$ , of  $8.8 \pm 1.6$  (stat.)  $^{+4.1}_{-3.0}$  (sys.) ( $2.8 \pm 0.5$  (stat.)  $^{+1.3}_{-1.0}$  (sys.)) for an integrated luminosity of  $30$  ( $3$ )  $\text{ab}^{-1}$ . This corresponds to an improvement of approximately 70% with respect to the baseline analysis. This percentage is calculated as  $\left( ((S/\sqrt{B})_{\text{opt.}} - (S/\sqrt{B})_{\text{base.}}) / (S/\sqrt{B})_{\text{base.}} \right) \times 100$ .

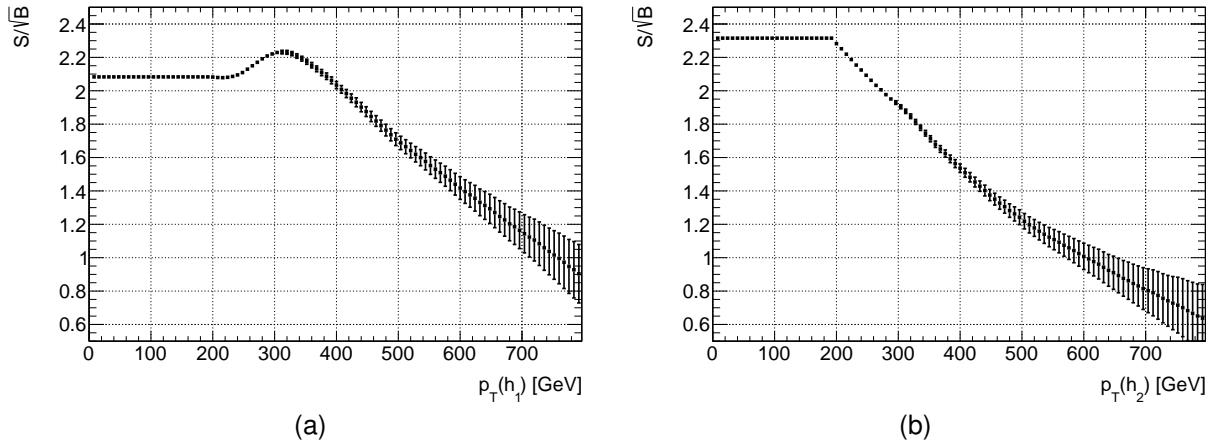


Figure 6.13:  $S/\sqrt{B}$  as a function of the cut on the  $p_T$  of the leading Higgs candidate,  $p_T(h_1)$  (a) after the pre-selection cuts and of the sub leading Higgs candidate,  $p_T(h_2)$  (b) after the cut on  $p_T(h_1)$ .

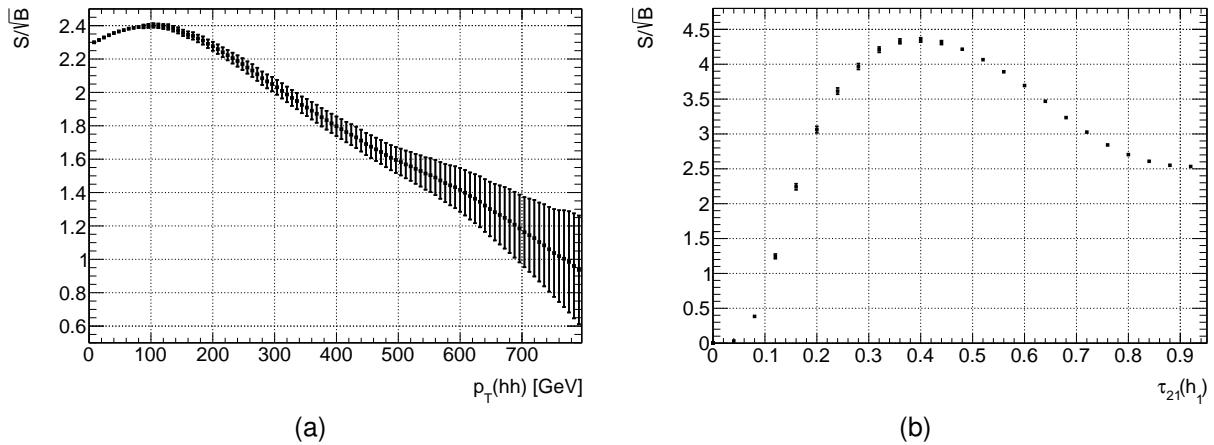


Figure 6.14:  $S/\sqrt{B}$  as a function of the cut on the  $p_T$  of the Higgs pair,  $p_T(hh)$  (a) after the cut on  $p_T(h_1)$  and as a function of the  $\tau_{21}$  variable for leading Higgs candidate,  $\tau_{21}(h_1)$  (b) after the cut on  $p_T(hh)$ .

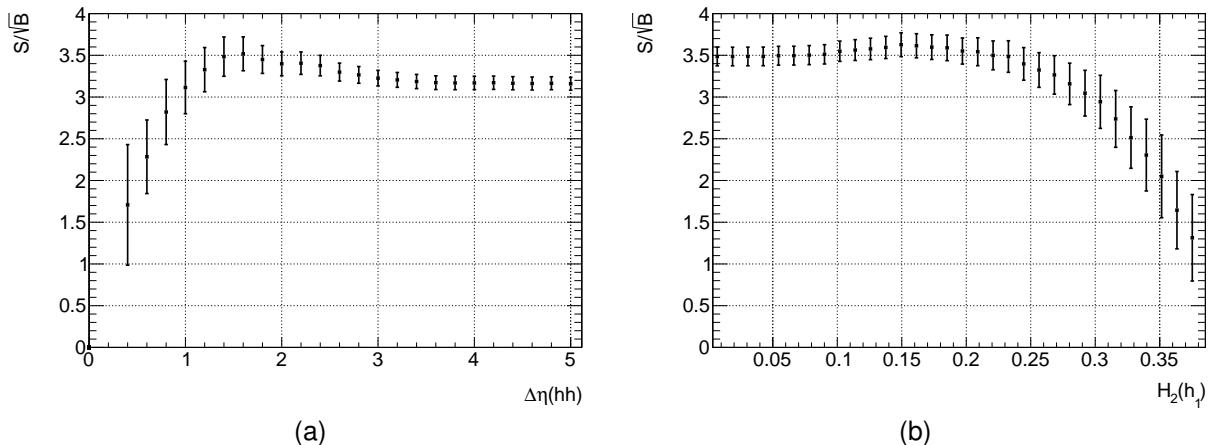


Figure 6.15:  $S/\sqrt{B}$  as a function of the cut on the  $\Delta\eta$  between the Higgs candidates,  $\Delta\eta(hh)$  (a) after the cuts on the transverse momenta and on the  $\tau_{21}$  variables and as a function of the second Fox-Wolfram momentum for the leading Higgs candidate,  $H_l(h_1)$  (b) after the cuts on the transverse momentums,  $\tau_{21}$  and  $\Delta\eta$ .

## 6.2.4 Handling of uncertainties

### Statistical uncertainties

For challenging analyses such as this one, the backgrounds usually have a much larger cross section than the signal, implying that we need large of Monte Carlo samples to properly simulate them. Most of the times it is not feasible to generate as much events as one would need for a given luminosity. This is the case with this work, in particular because we are targeting very high luminosities (of the order of tens of  $\text{ab}^{-1}$ ). Take for example the multijet sample with  $500 < H_T < 1000$ , where  $H_T$  is the scalar sum of the  $p_T$  of all partons at generator level. 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:

$$w = \frac{\mathcal{L} \times \sigma}{N} \quad (6.10)$$

where  $\mathcal{L}$  is the target integrated luminosity,  $\sigma$  is the cross section of the sample and  $N$  is the number of 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:

$$N_{\text{norm}} \pm \Delta N_{\text{norm}} = w \times (N \pm \sqrt{N}) . \quad (6.11)$$

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

$$\left( \Delta \frac{S}{\sqrt{B}} \right)_{\text{stat.}}^2 = \left| \frac{1}{\sqrt{B}} \right|^2 (\Delta S)^2 + \left| -\frac{S}{B^{3/2}} \right|^2 (\Delta B)^2 \quad (6.12)$$

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

$\Delta B$  is given by:

$$(\Delta B)^2 = \sum_i (\Delta B_i)^2 \quad (6.13)$$

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

### Systematic uncertainties

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

When requiring four b-tags, the largest uncertainty affecting the signal is the the b-tagging uncertainty (table 6 of Ref. [102]). It is of the order of 30%. For the background, the largest uncertainty comes from the estimation of its yield and shape, according to the same reference. The reported value is 16%. However, a study performed by the FCC study group [103] considers more conservative values for the uncertainties on the  $t\bar{t}$  and QCD di-jet backgrounds normalizations. The reported values are 20% for  $t\bar{t}$  and 50% for di-jet production. In this study we use these numbers to estimate the impact of the systematic uncertainties on the normalization of the backgrounds, since they give a more conservative estimate.

To estimate the b-tagging systematic uncertainty, the number of signal events after all the analysis cuts is varied by 30% up and down. This leads to an upper (lower) bound in the significance of 6.6 (3.6), considering the baseline analysis. This uncertainty is applied to signal only. According to Ref. [102], the b-tagging systematic uncertainty for the background is a lot smaller than for the signal ( $\sim 1\%$  compared to  $\sim 30\%$ ). Nonetheless, it could become relevant for the  $4b + j$  background. However, for the multijet background, the dominant systematic uncertainty is on its normalization. We consider a value of 50% which clearly dominates over the b-tagging systematic uncertainty even if we consider that it takes a value of 30% for the  $4b + j$  background.

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

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 is given by:

$$\left( \Delta \frac{S}{\sqrt{B}} \right)_{\text{sys.+stat.}}^2 = (\delta_{\text{b-tag}})^2 + (\delta_{\text{QCD bkg.}})^2 + (\delta_{t\bar{t} \text{ bkg.}})^2 + (\delta_{\text{stat.}})^2 \quad (6.14)$$

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 caused by the b-tagging uncertainty and by the uncertainties on the normalization of the QCD multijet and  $t\bar{t}$  backgrounds, respectively.  $\delta_{\text{stat.}}$  is the statistical uncertainty, given by Eq. 6.12.

### 6.2.5 Trigger discussion

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

In Ref. [102], the event selection requires the existence of at least two anti- $k_T$   $R = 1.0$  jets with

$p_T > 250$  GeV,  $|\eta| < 2.0$  and  $m_J > 50$  GeV. In addition, the leading jet must have  $p_T > 450$  GeV to guarantee 100% trigger efficiency.

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

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

# Chapter 7

## Results

In this chapter we describe the main results of the search for  $hh \rightarrow b\bar{b}b\bar{b}$  at the FCC-hh and compare them with the results obtained using the ATLAS default detector simulation (section 7.1). In section 7.2 we show how the significance of the analysis varies as a function of the granularity of the HCAL and/or the detector configuration. We also compare the results obtained using particle flow and pure calorimeter jets.

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

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

From table 7.1, we see that for the BSM models the signal efficiency is higher than for the SM. It 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 is due to the existence of heavy new particles in the BSM models that lead to the production of highly boosted SM Higgs pairs.

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

$$S/\sqrt{B} = 8.8 \pm 1.6 \text{ (stat.) } {}^{+4.1}_{-3.0} \text{ (sys.) } \quad (2.8 \pm 0.5 \text{ (stat.) } {}^{+1.3}_{-1.0} \text{ (sys.)}) \quad (7.1)$$

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. This result indicates that with the entire dataset that is expected to be accumulated by the FCC-hh detector it should be possible to observe the production of Higgs pairs, from this channel alone.

For the signal model that includes a 1 TeV dark matter mediator that can decay to pairs of SM Higgs bosons the achieve significance is  $2.3 \pm 0.4 \text{ (stat.) } {}^{+1.1}_{-0.8} \text{ (sys.) } (0.73 \pm 0.13 \text{ (stat.) } {}^{+0.34}_{-0.25} \text{ (sys.)})$  for an integrated luminosity of 30 (3)  $\text{ab}^{-1}$ . The significance is well bellow the  $5\sigma$  threshold for both luminosities. Therefore, we do not expect to be able to detect or exclude this challenging signal at the FCC-hh. In this model, the coupling of the DM mediator to the Higgs pairs is small, which means that the contribution

Table 7.1: Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM, DM mediator 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.

<b>Selection [FCC-hh]</b>	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	93.438
$p_T(j_1, j_2) \geq 200 \text{ GeV}$	16.602 $(4490 \pm 5) \times 10^3$	17.033 $(11140 \pm 12) \times 10^2$	33.802 $4725 \pm 4 \times 10^3$
<hr/>			
$p_T(j_1) \geq 300 \text{ GeV}$	13.521	14.007	30.691
$p_T(j_1 + j_2) \geq 100 \text{ GeV}$	10.933	11.301	22.736
$\tau_{21}(j_1, j_2) < 0.4$	1.309	1.410	3.921
$ \Delta\eta(hh)  < 1.5$	1.071	1.154	3.455
$FW2(j_1) > 0.2$	0.989	1.064	3.254
$(100 < M_{SD}(j_1, j_2) < 135) \text{ GeV}$	0.446 $(1207 \pm 8) \times 10^2$	0.487 $(3188 \pm 20) \times 10$	1.642 $(2296 \pm 8) \times 10^2$

from the box diagram dominates over the resonant production (s-channel diagram), just like in the SM. In addition, the cross section is smaller than that of the SM process which leads to a smaller significance even though the efficiency is higher.

For the type II 2HDM the achieved significance is

$$S/\sqrt{B} = 16.7 \pm 3.0 \text{ (stat.)} {}^{+7.8}_{-5.8} \text{ (sys.)} \quad (5.3 \pm 0.9 \text{ (stat.)} {}^{+2.5}_{-1.8} \text{ (sys.)}) \quad (7.2)$$

for an integrated luminosity of  $30 (3) \text{ ab}^{-1}$ . These results indicate that the contribution of this model to the production of Higgs pairs can be observed (or excluded) at the FCC-hh. When compared to the SM and DM mediator signal samples, this sample has a higher efficiency which translates to a higher significance, therefore making it a very exciting and achievable benchmark. It should also be noted that, in a real situation, several different final states will be used to observe and measure  $hh$  production.

### 7.1.1 Comparing with the ATLAS detector

The event selection of the search for  $hh \rightarrow b\bar{b}b\bar{b}$  with an ATLAS-like detector operating in a collider at a CM energy of  $\sqrt{s} = 100 \text{ TeV}$  is summarized in tables 7.3 and 7.4 for the signal and background samples, respectively.

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

Table 7.2: Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b + j$ ,  $jj + 0/1/2 j$  and  $t\bar{t} + 0/1/2j$ ). 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 $(149 \pm 28) \times 10^5$

For all the signal models, the significance increases approximately 50% going from the ATLAS detector to the FCC-hh. This percentage is computed as  $\left( ((S/\sqrt{B})_{\text{FCC}} - (S/\sqrt{B})_{\text{ATLAS}}) / (S/\sqrt{B})_{\text{ATLAS}} \right) \times 100$ . For the SM signal, nonetheless, using the ATLAS default detector configuration the achieved significance is already above  $5\sigma$ , for  $\mathcal{L} = 30 \text{ ab}^{-1}$ :  $S/\sqrt{B} = 5.7 \pm 1.3 \text{ (stat.)}^{+3.1}_{-2.4} \text{ (sys.)}$ .

The ATLAS detector has a momentum and energy resolution that is a lot worse than that of the FCC-hh. However, the results we obtained showed that even with an ATLAS-like detector operating at a collider at a CM energy of  $\sqrt{s} = 100 \text{ TeV}$ , the achieved significance is still above the  $5\sigma$  threshold. Nonetheless, a high resolution could be very useful to resolve the individual pileup vertices and ultimately help with pileup subtraction that is going to be crucial at any future 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 and compare them in terms of the achieved significances. 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 efficiently. 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 sig-

Table 7.3: Cumulative efficiency, in percentage, of each event selection criterion for the signal samples (SM, DM mediator 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>	SM	DM mediator	2HDM type II
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	89.624
$p_T(j_1, j_2) \geq 200 \text{ GeV}$	15.533 $(4201 \pm 5) \times 10^3$	15.941 32.001 $(10426 \pm 12) \times 10^2$	$(4473 \pm 4) \times 10^3$
<hr/>			
$p_T(j_1) \geq 300 \text{ GeV}$	12.599	13.061	28.960
$p_T(j_1 + j_2) \geq 100 \text{ GeV}$	10.185	10.526	21.414
$\tau_{21}(j_1, j_2) < 0.4$	1.139	1.220	3.394
$ \Delta\eta(hh)  < 1.5$	0.891	0.960	2.920
$FW2(j_1) > 0.2$	0.796	858	2.675
$(100 < M_{SD}(j_1, j_2) < 135) \text{ GeV}$	0.333 $(901 \pm 7) \times 10^2$	0.360 $(2358 \pm 18) \times 10$	1.245 $(1741 \pm 8) \times 10^2$

Table 7.4: Cumulative efficiency, in percentage, of each event selection criterion for the background samples ( $4b + j$ ,  $jj + 0/1/2j$  and  $t\bar{t} + 0/1/2j$ ). 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>	$4b + j$	$jj + 0/1/2j$	$t\bar{t}$
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	$5.7 \pm 1.3 \text{ (stat.)} {}^{+2.7}_{-2.0} \text{ (sys.)}$	$8.8 \pm 1.6 \text{ (stat.)} {}^{+4.1}_{-3.1} \text{ (sys.)}$
1 TeV DM mediator	$1.50 \pm 0.23 \text{ (stat.)} {}^{+0.7}_{-0.5} \text{ (sys.)}$	$2.3 \pm 0.4 \text{ (stat.)} {}^{+1.1}_{-0.8} \text{ (sys.)}$
2HDM type II	$11.1 \pm 1.7 \text{ (stat.)} {}^{+5.3}_{-3.9} \text{ (sys.)}$	$16.7 \pm 3.0 \text{ (stat.)} {}^{+7.8}_{-5.8} \text{ (sys.)}$

nal models and using particle flow and calorimeter jets can be found in appendix E. In addition, the significances achieved for all possible combinations are summarized in tables E.1 and E.2.

### 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 improves as we increase the granularity, i.e, the mass peaks becomes 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 26 (34) for the ATLAS detector configuration to 8 (6) for the FCC granularity with  $\eta, \phi \times 2$  which corresponds to a 70% (82%) 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 a resonance peak, confirming that the production is dominated by the box diagram. Even if we only had the contribution from the triangle diagram, the Higgs that decays to a pair of Higgs bosons would have 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 improve significantly as the granularity increases. This picture holds even when using calorimeter 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 (option 1 in Table 5.2) and ATLAS HCAL with  $\eta \times 4$  (option 2 in Table 5.2) to 0.44 for the FCC HCAL with  $\phi/2$  (option 3 in Table 5.2) 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 maxima of the distributions varies from 0.48 (0.24) to 0.56 (0.48), for particle flow (calorimeter) jets. These distributions can be found in figures F.2(a) and F.2(b) in appendix F.

### 7.2.2 Signal efficiency and significance

Figure 7.5(a) shows the signal efficiency for the three signal models: SM (filled squares), 1 TeV DM mediator (empty squares) and type II 2HDM with  $m_H = 900$  GeV (crosses), for particle flow jets, for the baseline analysis. The same plot is shown in figure 7.5(b) for calorimeter jets. For particle flow jets the efficiency increases as we increase the granularity, for all signal models. It varies approximately by 30% between the lowest value (ATLAS default) and the highest value (FCC granularity with  $\eta, \phi \times 2$ ). For HCAL jets, this tendency also exists with the exception of the first point which corresponds to the ATLAS default detector configuration. In addition, the efficiency changes more when using pure calorimeter jets. It varies by approximately 98% between the ATLAS detector configuration and the configuration with the FCC granularity and  $\eta, \phi \times 2$  (option 5 in Table 5.2). The efficiency is higher for both BSM models than for the SM. This is because these models were chosen to have very heavy particles decaying to a pair of highly boosted Higgs pairs. For the optimized analysis, these plots follow a very similar tendency. The absolute values of the efficiencies are slightly higher, as expected. These plots can be found in figures F.3(a) and F.3(b) in appendix F, for particle flow and calorimeter jets, respectively.

The significances achieved with the baseline analysis as a function of the detector configuration for the SM signal, the DM mediator model and the type II 2HDM are shown in figures 7.6(a), 7.7(a) and 7.8(a), respectively. The same plots are shown for the optimized analysis in figures 7.6(b), 7.7(b) and 7.8(b). The square markers refer to analysis performed using particle flow jets. The statistical uncertainty associated with each value of the significance is computed using standard error propagation and is shown as error bars. The grey blocks represent the total uncertainty. For the baseline analysis, regardless of the signal model considered, the significance changes by approximately 15% when going from the ATLAS detector configuration to the configuration with the FCC granularity increased by a factor of two in  $\eta$  and  $\phi$  (option 5 in Table 5.2). For the optimized analysis, the change is of approximately 54%. These percentages, as well as the ones given in the following paragraph, are calculated as  $\left( ((S/\sqrt{B})_{\text{FCC}} - (S/\sqrt{B})_{\text{ATLAS}}) / (S/\sqrt{B})_{\text{ATLAS}} \right) \times 100$ .

Motivated by the small change in significance over the range of configurations that were tested, we implemented exactly the same analysis but using HCAL jets instead of eflow jets. The results are shown in the same plots using triangular markers and with green error blocks. On the one hand, the achieved significance is always smaller when using HCAL jets because we are not making use of the

tracking information. On the other hand, when using HCAL jets, the significance changes more over the configuration range. For the SM signal, it varies by approximately 80% when going from the ATLAS detector configuration to the configuration with the FCC granularity increased by a factor of two in  $\eta$  and  $\phi$ . For the DM mediator model, this value is 77% and for the 2HDM it is 60%. These percentages refer to the baseline analysis. For the optimized analysis the change is of 71%, 69% and 55%, for the SM, DM mediator and 2HDM signal models, respectively.

All in all, given the large uncertainties, it is hard to conclude on the tendency followed by the plots of the significance as a function of the detector configuration. Nonetheless, the significance does seem to increase as the granularity increases, which is what was expected. As previously discussed, this effect is more accentuated when using pure calorimeter jets because we are only making use of the information provided by the hadronic calorimeter and therefore the analysis is more sensitive to its transverse segmentation (granularity), which is exactly the parameter that is varied in this study.

The small change in the significance when using eflow jets and the fact that the change increases when using HCAL jets indicate that in the FCC-hh baseline detector design the resolution of the tracking system is so good, in particular, so much better than the HCAL spatial resolution (given by its transversal segmentation), that it becomes the limiting factor and dominates the performance of jet reconstruction.

The discussion of some caveats is now in order. In a real experiment, the energy resolution of the tracking system is only better than that of the calorimeter up to a given energy threshold. In Delphes, however, it is assumed that it is always convenient to estimate the momentum of charged particles via the tracker [85]. Also, the tracking system is assumed to have an infinitely small transverse segmentation which is a good approximation but does not exactly reproduce what happens in a real experiment. In addition, in Delphes, the reconstruction of calorimeter jets is a lot simpler than in real experiments. For example, the calorimeters do not have longitudinal segmentation and no topological algorithm is used, as in ATLAS. Nonetheless, Ref. [85], shows that the energy and momentum resolution of the jets in Delphes reproduces quite well the results obtained by ATLAS and CMS. However, this might not be the case for the other detector configurations, for which no cross check can be performed.

All in all, in this work, the particle flow jets might be leading to an overly optimistic result because we are always making use of the tracking information even for energy/momenta ranges where that would not be the most favorable thing in a real experiment. Conversely, calorimeter jets might be leading to an overly pessimistic result because for these jets the jet reconstruction algorithm is overly simplified with respect to what would be used in a real experiment.

Therefore, studies using full detector simulation are of the utmost importance in order to obtain more realistic results that can be used as input for detector optimization studies for future colliders.

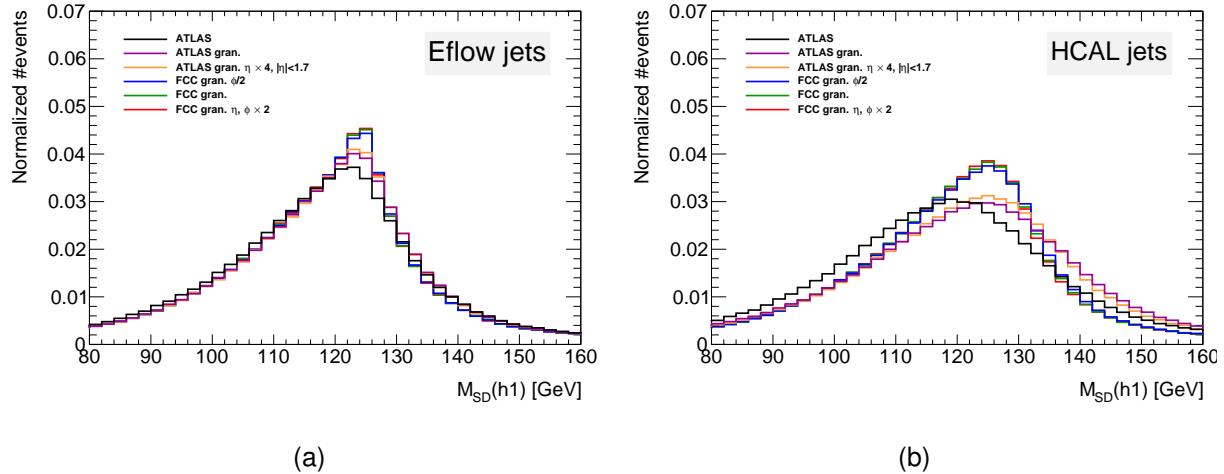


Figure 7.1: Leading Higgs candidate softdrop mass after the pre-selection cuts for particle flow (a) and calorimeter (b) jets. The colors indicate the different detector configurations. The x axis range is from 80 GeV to 160 GeV in order to make the differences between the histograms more clear.

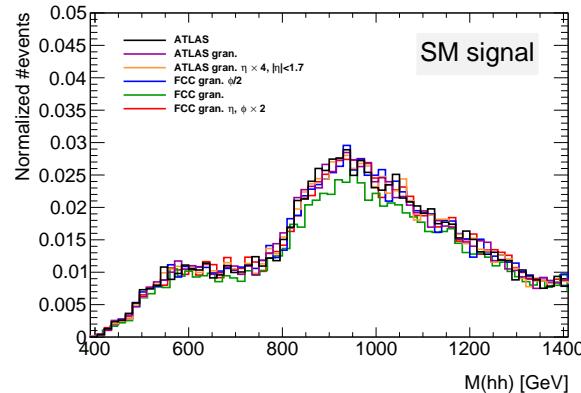


Figure 7.2: Invariant mass of the Higgs pair for the SM signal. The plots include the events that pass all the analysis cuts. The colors indicate the different detector configurations.

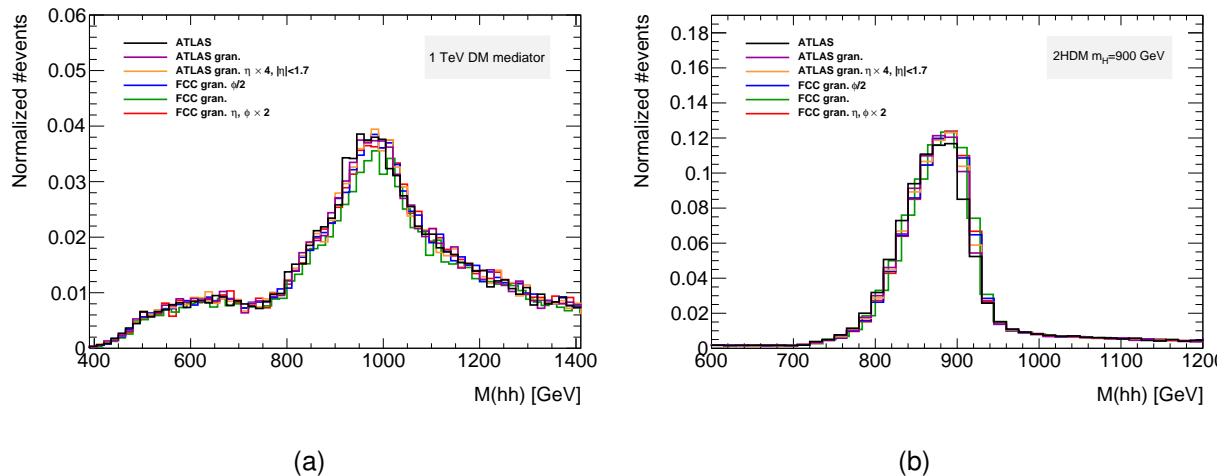


Figure 7.3: Invariant mass of the Higgs pair for the 1 TeV DM mediator (a) and for the 2HDM with  $m_H = 900$  GeV (b). The plots include the events that pass all the analysis cuts. The colors indicate the different detector configurations.

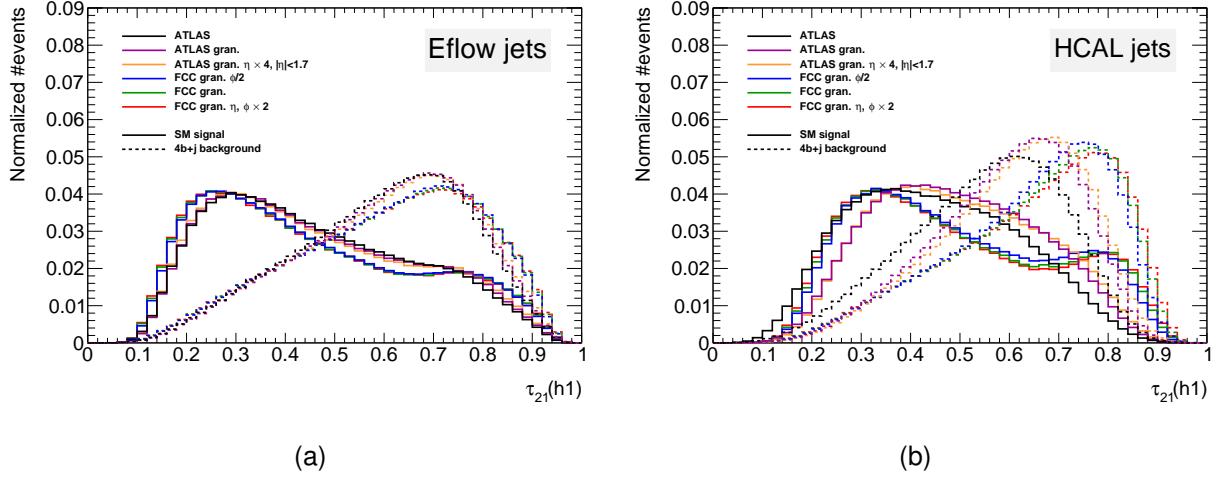


Figure 7.4: Leading Higgs candidate  $\tau_{21}$  distribution for eflow jets (a) and for HCAL jets (b). The colors indicate the different detector configurations. The distributions are shown for the signal (filled lines) and for the  $4b + j$  background (dashed lines).

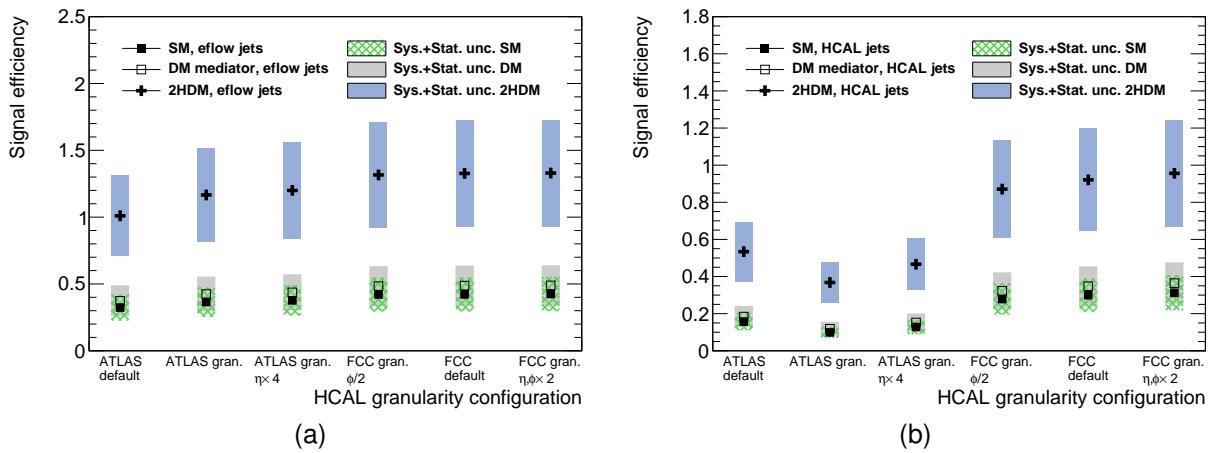


Figure 7.5: Signal efficiency as a function of the detector configuration for particle flow jets (a) and for calorimeter jets (b). Three signal models are shown: SM (filled squares), 1 TeV DM mediator (empty squares) and type II 2HDM with  $m_H = 900$  GeV (crosses). The statistical error bars are drawn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

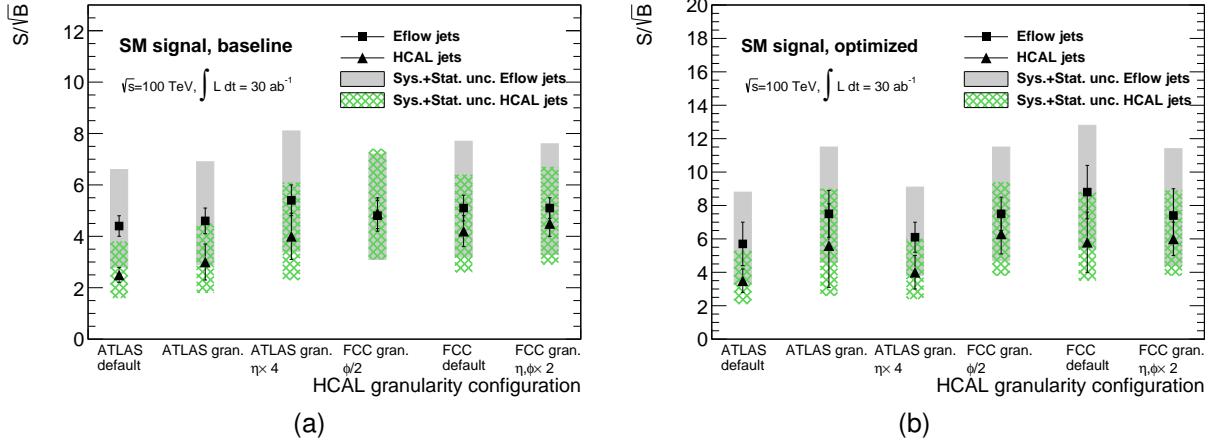


Figure 7.6: Significance as a function of the detector configuration for the SM signal, for the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and pure HCAL jets (triangles).

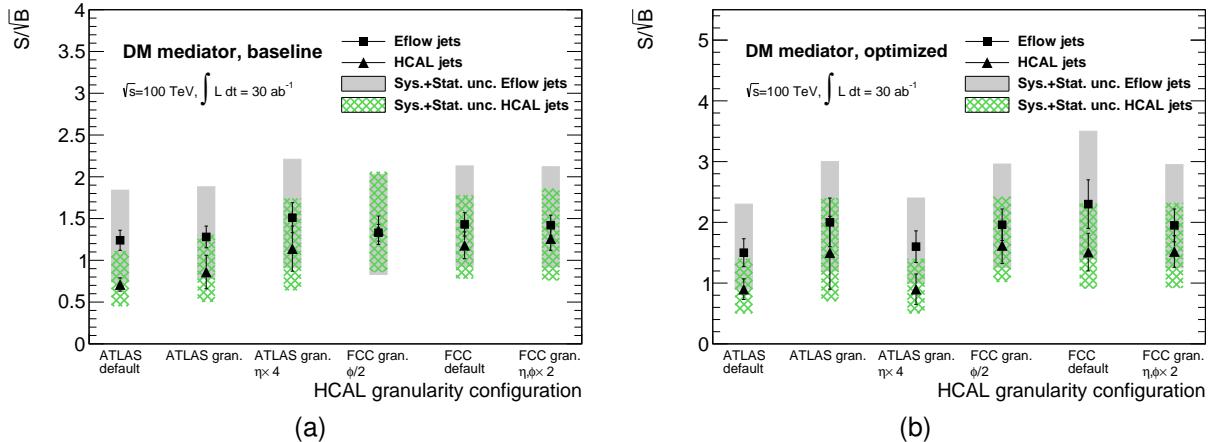


Figure 7.7: Significance as a function of the detector configuration for the DM mediator signal, for the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and pure HCAL jets (triangles).

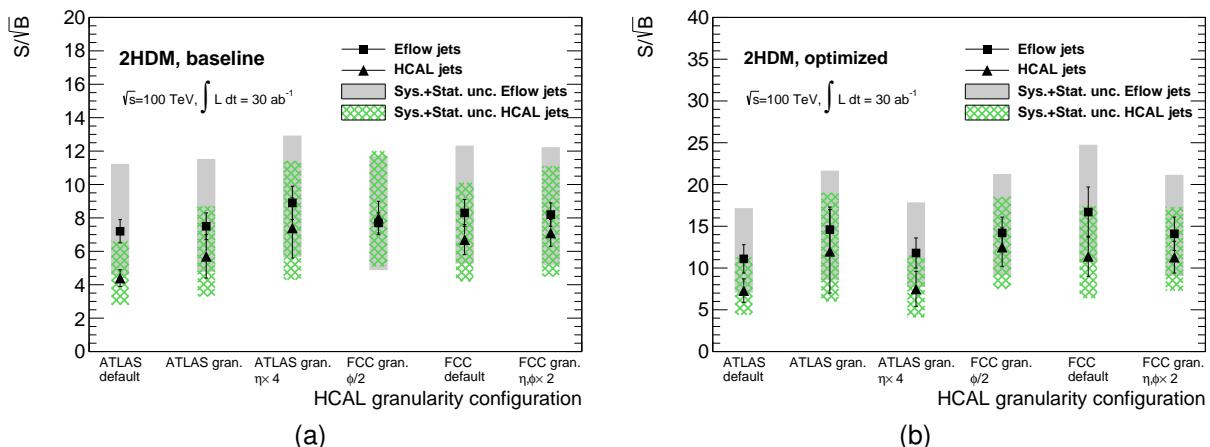


Figure 7.8: Significance as a function of the detector configuration for the 2HDM signal, for the baseline (a) and optimized (b) analyses. The significances are shown for eflow jets (squares) and pure HCAL jets (triangles).

# Chapter 8

## Conclusions

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

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

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

For the hadronic Future Circular Collider (FCC-hh), the significance obtained was  $S/\sqrt{B} = 8.8 \pm 1.6$  (stat.)  $^{+4.1}_{-3.0}$  (sys.) ( $2.8 \pm 0.5$  (stat.)  $^{+1.3}_{-1.0}$  (sys.)) for an integrated luminosity of  $\mathcal{L} = 30$  (3)  $\text{ab}^{-1}$ . For  $\mathcal{L} = 30 \text{ ab}^{-1}$ , the value is above the observation threshold ( $5\sigma$ ) which indicates that the full dataset that is expected to be collected during the operation of the FCC-hh should be enough to claim the observation of Higgs pair production from this channel alone, assuming SM production. For the CP-conserving type II 2HDM with  $m_H = 900$  GeV, the achieved significance was  $S/\sqrt{B} = 16.7 \pm 3.0$  (stat.)  $^{+7.8}_{-5.8}$  (sys.) ( $5.4 \pm 0.9$  (stat.)  $^{+2.5}_{-1.9}$  (sys.)) for an integrated luminosity of  $\mathcal{L} = 30$  (3)  $\text{ab}^{-1}$ , which makes it an accessible benchmark for the HL-LHC and future colliders. The achieved significance for the 1 TeV DM mediator is of order 2 even for an integrated luminosity of  $\mathcal{L} = 30 \text{ ab}^{-1}$ , which makes it a very challenging benchmark and probably unaccessible even at future high energy colliders.

With the ATLAS detector simulation, the achieved significance for the SM signal is  $S/\sqrt{B} = 5.7 \pm 1.3$  (stat.)  $^{+3.1}_{-2.4}$  (sys.) ( $1.8 \pm 0.4$  (stat.)  $^{+0.9}_{-0.6}$  (sys.)) for an integrated luminosity of  $\mathcal{L} = 30$  (3)  $\text{ab}^{-1}$ , for

which the nominal value is also above the observation threshold. This result indicates that the factors driving the discovery potential of the accelerator are the CM energy and high luminosity as opposed to the resolution of the detector.

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

When using particle flow jets, the change in the significance over the range of detector configurations tested is small, for all signal models. For the baseline (optimized) analysis, it varies by approximately 15% (54%), for all signal models, between the ATLAS default detector configuration and the FCC-hh detector with twice the granularity in  $\eta$  and  $\phi$ . Using pure calorimeter jets, the change in significance is more accentuated. For the SM signal, it varies by approximately 80% (71%), considering the baseline (optimized) analysis. For the same detector configuration, the significance is always smaller when using calorimeter jets. The same qualitative conclusions hold for the BSM models. The effect of the granularity is of the same order of magnitude as for the SM. These results lead to the conclusion that the jet reconstruction performance is dominated by the tracking system.

As future work, performing studies similar to the one presented here but using other benchmark processes as well as full detector simulation would be of the utmost importance to the optimization of the design of particle detectors for future high energy hadronic colliders. In addition, the impact of pileup should be assessed and multivariate techniques could be implemented to further optimize the event selection.

# 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>.

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

- [24] V. Branchina, F. Contino, and P. M. Ferreira. Electroweak vacuum lifetime in two Higgs doublet models. 2018.
- [25] J. F. Gunion and H. E. Haber. The CP conserving two Higgs doublet model: The Approach to the decoupling limit. *Phys. Rev.*, D67:075019, 2003. doi: 10.1103/PhysRevD.67.075019.
- [26] H. Sun, Y.-J. Zhou, and H. Chen. Constraints on large-extra-dimensions model through 125-GeV Higgs pair production at the LHC. *Eur. Phys. J.*, C72:2011, 2012. doi: 10.1140/epjc/s10052-012-2011-4.
- [27] P. Huang, A. Joglekar, M. Li, and C. E. M. Wagner. Corrections to di-Higgs boson production with light stops and modified Higgs couplings. *Phys. Rev.*, D97(7):075001, 2018. doi: 10.1103/PhysRevD.97.075001.
- [28] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.
- [29] S. Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.
- [30] Observation of  $H \rightarrow b\bar{b}$  decays and  $VH$  production with the ATLAS detector. Technical Report ATLAS-CONF-2018-036, CERN, Geneva, Jul 2018. URL <http://cds.cern.ch/record/2630338>.
- [31] Observation of Higgs boson decay to bottom quarks. Technical Report arXiv:1808.08242. CMS-HIG-18-016-003, CERN, Geneva, Aug 2018. URL <https://cds.cern.ch/record/2636067>. \* Temporary entry \*.
- [32] A. M. Sirunyan et al. Observation of the Higgs boson decay to a pair of  $\tau$  leptons with the CMS detector. *Phys. Lett.*, B779:283–316, 2018. doi: 10.1016/j.physletb.2018.02.004.
- [33] A. M. Sirunyan et al. Observation of  $t\bar{t}H$  production. *Phys. Rev. Lett.*, 120:231801, 2018. doi: 10.1103/PhysRevLett.120.231801.
- [34] M. Aaboud et al. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. 2018.
- [35] Measurement of the Higgs boson mass in the  $H \rightarrow ZZ^* \rightarrow 4\ell$  and  $H \rightarrow \gamma\gamma$  channels with  $\sqrt{s}=13\text{TeV}$   $pp$  collisions using the ATLAS detector. Technical Report ATLAS-CONF-2017-046, CERN, Geneva, Jul 2017. URL <http://cds.cern.ch/record/2273853>.
- [36] G. Aad et al. Combined Measurement of the Higgs Boson Mass in  $pp$  Collisions at  $\sqrt{s} = 7$  and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015. doi: 10.1103/PhysRevLett.114.191803.
- [37] M. Aaboud et al. Measurement of the  $W$ -boson mass in  $pp$  collisions at  $\sqrt{s} = 7$  TeV with the ATLAS detector. *Eur. Phys. J.*, C78(2):110, 2018. doi: 10.1140/epjc/s10052-017-5475-4.

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

- [53] G. A. et al. Performance of jet substructure techniques for large- $r$  jets in proton-proton collisions at  $\sqrt{s} = 7$  tev using the atlas detector. *Journal of High Energy Physics*, 2013(9), Sep 2013. doi: 10.1007/JHEP09(2013)076. URL [https://doi.org/10.1007/JHEP09\(2013\)076](https://doi.org/10.1007/JHEP09(2013)076).
- [54] J. Thaler and K. Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011. doi: 10.1007/JHEP03(2011)015.
- [55] C. Chen. New approach to identifying boosted hadronically-decaying particle using jet substructure in its center-of-mass frame. *Phys. Rev.*, D85:034007, 2012. doi: 10.1103/PhysRevD.85.034007.
- [56] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008. doi: 10.1103/PhysRevLett.100.242001.
- [57] I. Hinchliffe, A. Kotwal, M. L. Mangano, C. Quigg, and L.-T. Wang. Luminosity goals for a 100-TeV pp collider. *Int. J. Mod. Phys.*, A30(23):1544002, 2015. doi: 10.1142/S0217751X15440029.
- [58] N. Arkani-Hamed, T. Han, M. Mangano, and L.-T. Wang. Physics opportunities of a 100 TeV protonproton collider. *Phys. Rept.*, 652:1–49, 2016. doi: 10.1016/j.physrep.2016.07.004.
- [59] S. Banerjee, C. Englert, M. L. Mangano, M. Selvaggi, and M. Spannowsky.  $hh + \text{jet}$  production at 100 TeV. *Eur. Phys. J.*, C78(4):322, 2018. doi: 10.1140/epjc/s10052-018-5788-y.
- [60] 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.
- [61] T. an FCC-hh Detector Magnet System. E. bielert, 2018. URL <https://ep-news.web.cern.ch/content/towards-fcc-hh-detector-magnet-system>.
- [62] 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.
- [63] A. Zaborowska. Electromagnetic calorimetry based on liquid argon for the fcc-hh experiments. 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.
- [64] C. Neubser. Performance studies and requirements on the calorimeters for a fcc-hh experiment. URL: <https://indico.ihep.ac.cn/event/6387/session/14/contribution/117/material/slides/0.pdf>, 2017.
- [65] M. Aaboud et al. Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton–proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. *Phys. Rev.*, D94(5):052002, 2016. doi: 10.1103/PhysRevD.94.052002.

- [66] A. M. Sirunyan et al. Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- [67] G. Aad et al. Searches for Higgs boson pair production in the  $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$  channels with the ATLAS detector. *Phys. Rev.*, D92:092004, 2015. doi: 10.1103/PhysRevD.92.092004.
- [68] A. M. Sirunyan et al. Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- [69] M. Aaboud et al. A search for resonant and non-resonant Higgs boson pair production in the  $b\bar{b}\tau^+\tau^-$  decay channel in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector . Technical Report arXiv:1808.00336, CERN, Geneva, Aug 2018. URL <https://cds.cern.ch/record/2632994>. \* Temporary entry \*.
- [70] V. Khachatryan et al. Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV. *Phys. Rev.*, D94(5):052012, 2016. doi: 10.1103/PhysRevD.94.052012.
- [71] G. Aad et al. Search For Higgs Boson Pair Production in the  $\gamma\gamma b\bar{b}$  Final State using  $pp$  Collision Data at  $\sqrt{s} = 8$  TeV from the ATLAS Detector. *Phys. Rev. Lett.*, 114(8):081802, 2015. doi: 10.1103/PhysRevLett.114.081802.
- [72] A. M. Sirunyan et al. Search for resonant and nonresonant Higgs boson pair production in the  $b\bar{b}lnlnu$  final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV. 2017.
- [73] M. Aaboud et al. Combination of searches for heavy resonances decaying into bosonic and leptonic final states using  $36\text{ fb}^{-1}$  of proton-proton collision data at  $\sqrt{s} = 13$  TeV with the ATLAS detector. 2018.
- [74] G. Aad and t. . others.
- [75] J. K. Behr, D. Bortoletto, J. A. Frost, N. P. Hartland, C. Issever, and J. Rojo. Boosting Higgs pair production in the  $b\bar{b}b\bar{b}$  final state with multivariate techniques. *Eur. Phys. J.*, C76(7):386, 2016. doi: 10.1140/epjc/s10052-016-4215-5.
- [76] Study of the double Higgs production channel  $H(\rightarrow b\bar{b})H(\rightarrow \gamma\gamma)$  with the ATLAS experiment at the HL-LHC. Technical Report ATL-PHYS-PUB-2017-001, CERN, Geneva, Jan 2017. URL <https://cds.cern.ch/record/2243387>.
- [77] Higgs pair production at the High Luminosity LHC. Technical Report CMS-PAS-FTR-15-002, CERN, Geneva, 2015. URL <https://cds.cern.ch/record/2063038>.
- [78] N. P. Hartland. Double Higgs production from  $HH \rightarrow (b\bar{b})(b\bar{b})$  at a 100 TeV hadron collider. *PoS*, DIS2016:115, 2016.

- [79] S. V. Chekanov, M. Beydler, A. V. Kotwal, L. Gray, S. Sen, N. V. Tran, S. S. Yu, and J. Zuzelski. Initial performance studies of a general-purpose detector for multi-TeV physics at a 100 TeV pp collider. *JINST*, 12(06):P06009, 2017. doi: 10.1088/1748-0221/12/06/P06009.
- [80] S. Chekanov. Boosted jets in high-granularity calorimeter at a 100 TeV pp collider. BOOST 2017, 2017.
- [81] S. Chekanov. Physics requirements for a Hadron Calorimeter for a 100 TeV proton-proton collider. FCC week 2015, 2015.
- [82] S. Chekanov. Simulation of a high-granular hadronic calorimeter for multi-TeV physics. FCC week 2016, 2016.
- [83] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer.
- [84] T. Sjostrand, S. Mrenna, and P. Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.
- [85] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lematre, A. Mertens, and M. Selvaggi. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014. doi: 10.1007/JHEP02(2014)057.
- [86] E. Boos et al. Generic user process interface for event generators. In *Physics at TeV colliders. Proceedings, Euro Summer School, Les Houches, France, May 21-June 1, 2001*, 2001. URL <http://lss.fnal.gov/archive/preprint/fermilab-conf-01-496-t.shtml>.
- [87] M. Cacciari, G. P. Salam, and G. Soyez. FastJet User Manual. *Eur. Phys. J.*, C72:1896, 2012. doi: 10.1140/epjc/s10052-012-1896-2.
- [88] D. Buskulic et al. Performance of the ALEPH detector at LEP. *Nucl. Instrum. Meth.*, A360:481–506, 1995. doi: 10.1016/0168-9002(95)00138-7.
- [89] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. URL <http://cds.cern.ch/record/1194487>.
- [90] FCC study group. Fccsw. <https://github.com/HEP-FCC/FCCSW>, 2018.
- [91] G. Barrand et al. GAUDI - A software architecture and framework for building HEP data processing applications. *Comput. Phys. Commun.*, 140:45–55, 2001. doi: 10.1016/S0010-4655(01)00254-5.
- [92] F. Gaede, B. Hegner, and P. Mato. Podio: An event-data-model toolkit for high energy physics experiments. *Journal of Physics: Conference Series*, 898(7):072039, 2017. URL <http://stacks.iop.org/1742-6596/898/i=7/a=072039>.
- [93] FCC-hh study group. Eventproducer. <https://github.com/FCC-hh-framework/EventProducer>, 2018.

- [94] N. D. Christensen and C. Duhr. FeynRules - Feynman rules made easy. *Comput. Phys. Commun.*, 180:1614–1641, 2009. doi: 10.1016/j.cpc.2009.02.018.
- [95] T. general Two-Higgs-Doublet Model. C. degrande, c. duhr, m. herquet, 2018. URL <http://feynrules.irmp.ucl.ac.be/wiki/2HDM>.
- [96] C. Degrande. Automatic evaluation of UV and R2 terms for beyond the Standard Model Lagrangians: a proof-of-principle. *Comput. Phys. Commun.*, 197:239–262, 2015. doi: 10.1016/j.cpc.2015.08.015.
- [97] H. E. Haber and O. Stl. New LHC benchmarks for the  $\mathcal{CP}$ -conserving two-Higgs-doublet model. *Eur. Phys. J.*, C75(10):491, 2015. doi: 10.1140/epjc/s10052-015-3697-x, 10.1140/epjc/s10052-016-4151-4. [Erratum: Eur. Phys. J.C76,no.6,312(2016)].
- [98] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. doi: 10.23731/CYRM-2017-002.
- [99] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, and J. Mazzitelli. Higgs boson pair production at NNLO with top quark mass effects. *JHEP*, 05:059, 2018. doi: 10.1007/JHEP05(2018)059.
- [100] C. Patrignani et al. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016. doi: 10.1088/1674-1137/40/10/100001.
- [101] T. Sjostrand, S. Mrenna, and P. Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026.
- [102] M. Aaboud et al. Search for pair production of Higgs bosons in the  $b\bar{b}b\bar{b}$  final state using proton-proton collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector. 2018.
- [103] 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).

## Appendix A

# Jet radius discussion

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

As a first approximation we compute the angle between the b quarks produced by the decay of a Higgs boson. We assume that the b quarks are massless and that the Higgs moves only in the transverse plane (perpendicular to the beam pipe) such that it has no longitudinal momentum and the 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 smaller than the jet's diameter (1.6) and therefore the two b quarks can both be contained inside the jet and the Higgs boson fully reconstructed.

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

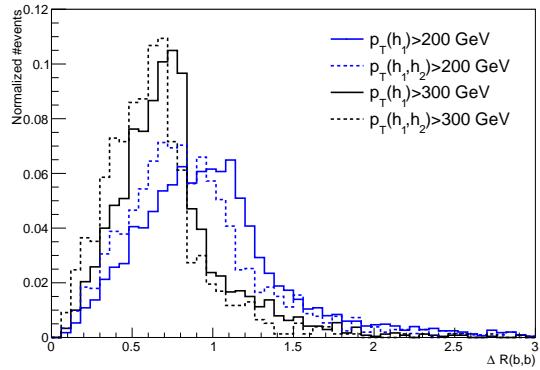


Figure A.1:  $\Delta R(b, \bar{b})$  distributions for  $p_T(h_1) > 200/300$  GeV (solid blue/black) and for  $p_T(h_1, h_2) > 200/300$  GeV (dashed blue/black).

## Appendix B

# Additional background processes

In this section we discuss the importance of additional background processes and estimate how they influence the analysis. In particular, we investigate backgrounds involving Higgs bosons. We consider the  $t\bar{t}h$  and  $h+0/1/2\ j$  processes. The effective cross section, efficiency and expected number of events for an integrated luminosity of  $\mathcal{L} = 30 \text{ ab}^{-1}$  are summarized in table B.1. Including these backgrounds in the analysis leads to a decrease in the significance of 0.26%, which we consider to be a very small 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	$1.6 \times 10^6$

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

## Appendix C

# Softdrop mass

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

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

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

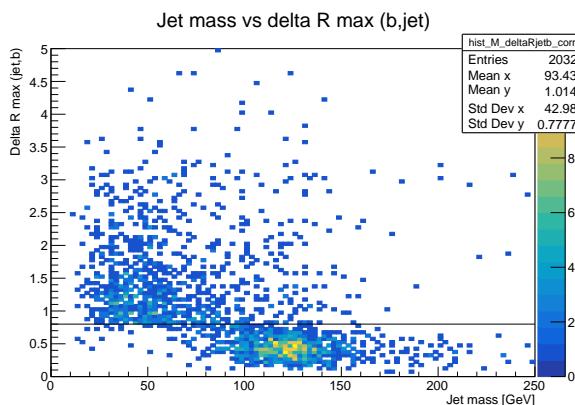


Figure C.1: Correlation between the maximum  $\Delta R$  between the Higgs candidate jet and one of the b quarks (y axis) and the jet's mass (x axis). The horizontal line corresponds to  $\Delta R = 0.8$ .

## Appendix D

# Samples generation: parameters

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

### D.1 MadGraph

The MadGraph5 level cuts are summarized in table D.1. We show only the most relevant cuts for this analysis: the minimum  $p_T$  of light and b quarks,  $p_{T,j}^{\min}$  and  $p_{T,b}^{\min}$ , the maximum pseudorapidity range for 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 quarks,  $\Delta R(bb)$ , and between a light and b quarks,  $\Delta R(jb)$ . The  $xqcut$  parameter is a measure of the required parton separation at Madgraph level. Whenever MadGraph produces two partons,  $i$  and  $j$ , it defines 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 of this expression is smaller than the specified value of  $xqcut$  then we do not generate the event. The  $bwcutoff$  parameter defines what is considered to be on-shell s-channel resonances. The  $H_T$  variable is the scalar sum of the  $p_T$  of all truth level partons, including b quarks.

### D.2 Pythia

The settings for jet matching can be found in table D.2 under the corresponding samples' columns. We perform the jet matching procedure (merge=on) using the MLM matching scheme and the appropriate algorithm for a parton level process generated in MadGraph (scheme=1). We do not read the matching parameters from the MadGraph file (setMad=off) because this option is not available for these files. The size of the cone drawn around the jet's center, the maximum pseudorapidity and the maximum number of jets to be matched are given by coneRadius, etaJetMax and nJetMax, respectively. The cone radius is set to one. The maximum allowed pseudorapidity of jets is ten which is a much loser cut than the acceptance of any current or future detector. The maximum number of jets is set to four for the  $jj + 0/1/2 j$  and to two for the  $t\bar{t} + 0/1/2 j$ . 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).	

# Appendix E

## Cutflow tables

In this appendix we summarize the significances achieved with the different detector configurations and for the SM (table E.1), DM mediator (in black) and 2HDM (in blue) (table E.2) signal samples. In addition, we collect the cutflow tables for the baseline and optimized analysis, for particle flow and calorimeter jets and for all detector configurations that were tested. The efficiencies of the cuts are shown for all cuts.

Table E.1: Summary of the significances obtained with particle flow and calorimeter jets for the baseline and optimized analysis for the SM signal.

Configuration	Eflow jets		HCAL jets	
	Baseline	Optimized	Baseline	Optimized
ATLAS	$4.4^{+2.1}_{-1.6}$ ; <b>7.2<sup>+3.6</sup><sub>-2.6</sub></b>	$5.7^{+3.1}_{-2.4}$	$2.50^{+1.3}_{-0.9}$	$3.5^{+1.8}_{-1.4}$
ATLAS gran.	$4.6^{+2.3}_{-1.6}$	$7.5^{+3.7}_{-2.9}$	$3.0^{+1.5}_{-1.2}$	$5.6^{+3.4}_{-3.0}$
ATLAS gran. $\eta \times 4$	$5.4^{+2.7}_{-2.0}$	$6.1^{+3.0}_{-2.3}$	$4.0^{+2.1}_{-1.7}$	$4^{+2.0}_{-1.6}$
FCC gran. $\phi/2$	$4.8^{+2.4}_{-1.7}$	$7.5^{+3.7}_{-2.8}$	$4.9^{+2.5}_{-1.8}$	$6.3^{+3.0}_{-2.2}$
FCC	$5.1^{+2.6}_{-1.9}$	$8.8^{+4.4}_{-3.4}$	$4.2^{+2.2}_{-1.6}$	$5.8^{+3.0}_{-2.3}$
FCC gran. $\eta, \phi \times 2$	$5.1^{+2.5}_{-1.8}$	$7.4^{+3.9}_{-3.0}$	$4.5^{+2.2}_{-1.6}$	$6^{+3}_{-3}$

Table E.2: Summary of the significances obtained with particle flow and calorimeter jets for the baseline and optimized analysis for the DM mediator (black) and 2HDM (blue) signal processes.

Configuration	Eflow jets		HCAL jets	
	Baseline	Optimized	Baseline	Optimized
ATLAS	$1.24^{+0.62}_{-0.45}$ ; <b>7.2<sup>+3.6</sup><sub>-2.6</sub></b>	$1.50^{+0.75}_{-0.57}$ ; <b>11.1<sup>+5.6</sup><sub>-4.2</sub></b>	$0.71^{+0.36}_{-0.26}$ ; <b>4.4<sup>+2.2</sup><sub>-1.6</sub></b>	$0.90^{+0.47}_{-0.36}$ ; <b>7.3<sup>+3.8</sup><sub>-2.9</sub></b>
ATLAS gran.	$1.28^{+0.64}_{-0.47}$ ; <b>7.5<sup>+3.8</sup><sub>-2.7</sub></b>	$2.0^{+1.0}_{-0.8}$ ; <b>14.6<sup>+7.3</sup><sub>-5.7</sub></b>	$0.86^{+0.45}_{-0.36}$ ; <b>5.7<sup>+3.0</sup><sub>-2.4</sub></b>	$1.5^{+0.9}_{-0.8}$ ; <b>12.1<sup>+7.3</sup><sub>-6.4</sub></b>
ATLAS gran. $\eta \times 4$	$1.51^{+0.75}_{-0.55}$ ; <b>8.9<sup>+4.4</sup><sub>-3.2</sub></b>	$1.60^{+0.80}_{-0.61}$ ; <b>11.8<sup>+5.9</sup><sub>-4.5</sub></b>	$1.14^{+0.61}_{-0.48}$ ; <b>7.4<sup>+4.0</sup><sub>-3.1</sub></b>	$0.90^{+0.50}_{-0.40}$ ; <b>7.5<sup>+4.2</sup><sub>-3.4</sub></b>
FCC gran. $\phi/2$	$1.33^{+0.66}_{-0.48}$ ; <b>7.7<sup>+3.8</sup><sub>-2.8</sub></b>	$1.96^{+0.96}_{-0.73}$ ; <b>14.2<sup>+7.0</sup><sub>-5.3</sub></b>	$1.36^{+0.69}_{-0.51}$ ; <b>7.8<sup>+4.0</sup><sub>-2.9</sub></b>	$1.62^{+0.81}_{-0.64}$ ; <b>12.5<sup>+6.3</sup><sub>-4.9</sub></b>
FCC	$1.43^{+0.72}_{-0.52}$ ; <b>8.3<sup>+4.2</sup><sub>-3.0</sub></b>	$2.3^{+1.2}_{-0.9}$ ; <b>16.7<sup>+8.4</sup><sub>-6.5</sub></b>	$1.18^{+0.61}_{-0.44}$ ; <b>6.7<sup>+3.4</sup><sub>-2.5</sub></b>	$1.51^{+0.78}_{-0.59}$ ; <b>11.4<sup>+5.9</sup><sub>-4.6</sub></b>
FCC gran. $\eta, \phi \times 2$	$1.42^{+0.71}_{-0.51}$ ; <b>8.2<sup>+4.1</sup><sub>-3.0</sub></b>	$1.95^{+0.96}_{-0.73}$ ; <b>14.1<sup>+6.9</sup><sub>-5.3</sub></b>	$1.26^{+0.63}_{-0.46}$ ; <b>7.1<sup>+3.6</sup><sub>-2.6</sub></b>	$1.52^{+0.78}_{-0.59}$ ; <b>11.3<sup>+5.8</sup><sub>-4.4</sub></b>

Table E.3: Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS default detector, as implemented in Delphes.

<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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	88.690; <b>83.646</b>	88.787; <b>83.734</b>	89.624; <b>84.504</b>	71.617; <b>66.487</b>	3.749; <b>3.354</b>	51.782; <b>46.516</b>
$p_T(j_1, j_2) \geq 200$ GeV	15.533; <b>13.966</b>	15.941; <b>14.367</b>	32.001; <b>29.581</b>	16.299; <b>14.290</b>	0.685; <b>0.601</b>	0.985; <b>0.862</b>
<hr/>						
$p_T(j_1) \geq 400$ GeV	7.997; <b>7.136</b>	8.499; <b>7.615</b>	19.297; <b>17.294</b>	6.378; <b>5.540</b>	0.170; <b>0.148</b>	0.416; <b>0.370</b>
$p_T(j_2) \geq 350$ GeV	5.282; <b>4.707</b>	5.704; <b>5.104</b>	11.841; <b>10.306</b>	3.560; <b>3.075</b>	0.112; <b>0.097</b>	0.245; <b>0.218</b>
$p_T(j_1 + j_2) \geq 100$ GeV	4.304; <b>3.833</b>	4.604; <b>4.117</b>	8.709; <b>7.628</b>	3.000; <b>2.580</b>	0.065; <b>0.057</b>	0.207; <b>0.182</b>
$\tau_{21}(j_1) < 0.55$	3.257; <b>2.805</b>	3.490; <b>3.022</b>	6.718; <b>5.762</b>	1.288; <b>1.380</b>	0.021; <b>0.032</b>	0.122; <b>0.107</b>
$\tau_{21}(j_2) < 0.55$	1.648; <b>1.619</b>	1.809; <b>1.764</b>	3.737; <b>3.525</b>	0.545; <b>0.685</b>	0.008; <b>0.015</b>	0.064; <b>0.060</b>
$FW2(j_1) > 0.2$	1.125; <b>0.724</b>	1.243; <b>0.795</b>	2.768; <b>1.816</b>	0.272; <b>0.189</b>	0.003; <b>0.003</b>	0.016; <b>0.008</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.724; <b>0.406</b>	0.815; <b>0.453</b>	2.012; <b>1.170</b>	0.047; <b>0.028</b>	0.000; <b>0.0003</b>	0.003; <b>0.002</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.323; <b>0.157</b>	0.374; <b>0.184</b>	1.009; <b>0.534</b>	0.010; <b>0.005</b>	0.00007; <b>0.00005</b>	0.00007; <b>0.00005</b>

Table E.4: Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity.

<b>Selection [ATLAS gran.]</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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	92.226; <b>85.810</b>	92.322; <b>85.896</b>	93.185; <b>86.656</b>	75.464; <b>67.886</b>	3.930; <b>3.448</b>	53.144; <b>45.851</b>
$p_T(j_1, j_2) \geq 200$ GeV	16.504; <b>14.820</b>	16.918; <b>15.204</b>	33.632; <b>30.867</b>	17.580; <b>15.152</b>	0.732; <b>0.644</b>	1.043; <b>0.917</b>
<hr/>						
$p_T(j_1) \geq 400$ GeV	8.550; <b>7.527</b>	9.069; <b>8.007</b>	20.647; <b>18.372</b>	6.890; <b>5.900</b>	0.181; <b>0.157</b>	0.443; <b>0.392</b>
$p_T(j_2) \geq 350$ GeV	5.662; <b>4.977</b>	6.104; <b>5.390</b>	12.889; <b>11.218</b>	3.867; <b>3.295</b>	0.120; <b>0.105</b>	0.261; <b>0.232</b>
$p_T(j_1 + j_2) \geq 100$ GeV	4.610; <b>4.037</b>	4.928; <b>4.336</b>	9.486; <b>8.206</b>	3.249; <b>2.761</b>	0.069; <b>0.061</b>	0.221; <b>0.195</b>
$\tau_{21}(j_1) < 0.55$	3.482; <b>2.653</b>	3.729; <b>2.851</b>	7.283; <b>5.587</b>	1.362; <b>1.193</b>	0.021; <b>0.027</b>	0.130; <b>0.097</b>
$\tau_{21}(j_2) < 0.55$	1.649; <b>1.300</b>	1.813; <b>1.419</b>	3.869; <b>2.962</b>	0.535; <b>0.483</b>	0.007; <b>0.010</b>	0.063; <b>0.047</b>
$FW2(j_1) > 0.2$	1.157; <b>0.468</b>	1.280; <b>0.509</b>	2.949; <b>1.247</b>	0.276; <b>0.117</b>	0.003; <b>0.001</b>	0.016; <b>0.005</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.761; <b>0.235</b>	0.859; <b>0.267</b>	2.188; <b>0.755</b>	0.048; <b>0.013</b>	0.0004; <b>0.0001</b>	0.004; <b>0.001</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.365; <b>0.099</b>	0.424; <b>0.118</b>	1.166; <b>0.368</b>	0.011; <b>0.002</b>	0.00008; <b>0.00001</b>	0.0008; <b>0.0003</b>

Table E.5: Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity with  $\eta \times 4$ .

<b>Selection [ATLAS gran. <math>\eta \times 4</math>]</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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	92.259; <b>87.076</b>	92.358; <b>87.167</b>	92.259; <b>88.043</b>	70.510; <b>69.147</b>	3.935; <b>3.523</b>	53.198; <b>46.560</b>
$p_T(j_1, j_2) \geq 200$ GeV	16.544; <b>15.131</b>	16.965; <b>15.566</b>	33.674; <b>31.481</b>	17.626; <b>15.567</b>	0.734; <b>0.665</b>	1.044; <b>0.938</b>
<hr/>						
$p_T(j_1) \geq 400$ GeV	8.577; <b>7.697</b>	9.096; 8.203	20.737; <b>18.751</b>	6.918; <b>6.079</b>	0.181; <b>0.162</b>	0.443; <b>0.400</b>
$p_T(j_2) \geq 350$ GeV	5.682; <b>5.091</b>	6.126; 5.524	12.971; <b>11.469</b>	3.875; <b>3.399</b>	0.120; <b>0.108</b>	0.261; <b>0.237</b>
$p_T(j_1 + j_2) \geq 100$ GeV	4.630; <b>4.139</b>	4.944; 4.442	9.486; 8.407	3.263; <b>2.852</b>	0.069; <b>0.062</b>	0.222; <b>0.200</b>
$\tau_{21}(j_1) < 0.55$	3.518; <b>2.683</b>	3.758; 2.889	7.339; <b>5.649</b>	1.359; <b>1.138</b>	0.020; <b>0.024</b>	0.131; <b>0.097</b>
$\tau_{21}(j_2) < 0.55$	1.661; <b>1.249</b>	1.825; 1.371	3.896; <b>2.896</b>	0.530; <b>0.432</b>	0.006; <b>0.008</b>	0.065; <b>0.044</b>
$FW2(j_1) > 0.2$	1.182; <b>0.538</b>	1.309; 0.595	3.005; <b>1.448</b>	0.281; <b>0.127</b>	0.003; <b>0.001</b>	0.017; <b>0.005</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.781; <b>0.289</b>	0.886; 0.329	2.243; <b>0.923</b>	0.049; <b>0.016</b>	0.0003; <b>0.0001</b>	0.004; <b>0.001</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.377; <b>0.128</b>	0.437; 0.152	1.200; <b>0.466</b>	0.011; <b>0.003</b>	0.00006; <b>0.00001</b>	0.0009; <b>0.0002</b>

Table E.6: Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with  $\phi/2$ .

<b>Selection [FCC gran. <math>\phi/2</math>]</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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	92.473; <b>90.243</b>	92.567; <b>90.345</b>	93.4220; <b>91.220</b>	75.798; <b>72.741</b>	3.961; <b>3.732</b>	53.474; <b>48.883</b>
$p_T(j_1, j_2) \geq 200$ GeV	16.603; <b>15.838</b>	17.025; <b>16.255</b>	33.788; <b>32.732</b>	17.812; <b>16.706</b>	0.740; <b>0.705</b>	1.054; <b>0.983</b>
<hr/>						
$p_T(j_1) \geq 400$ GeV	8.620; <b>8.091</b>	9.149; <b>8.606</b>	20.880; <b>19.619</b>	7.004; <b>6.534</b>	0.182; <b>0.172</b>	0.301; <b>0.419</b>
$p_T(j_2) \geq 350$ GeV	5.707; <b>5.346</b>	6.160; <b>5.786</b>	13.090; <b>12.024</b>	3.922; <b>3.644</b>	0.120; <b>0.113</b>	0.263; <b>0.248</b>
$p_T(j_1 + j_2) \geq 100$ GeV	4.644; <b>4.336</b>	4.966; <b>4.651</b>	9.559; <b>8.803</b>	3.307; <b>3.062</b>	0.069; <b>0.064</b>	0.222; <b>0.209</b>
$\tau_{21}(j_1) < 0.55$	3.630; <b>3.044</b>	3.886; <b>3.271</b>	7.567; <b>6.304</b>	1.371; <b>1.027</b>	0.018; <b>0.014</b>	0.138; <b>0.109</b>
$\tau_{21}(j_2) < 0.55$	1.692; <b>1.238</b>	1.861; <b>1.380</b>	3.986; <b>2.940</b>	0.534; <b>0.330</b>	0.005; <b>0.003</b>	0.069; <b>0.048</b>
$FW2(j_1) > 0.2$	1.266; <b>0.783</b>	1.403; <b>0.873</b>	3.205; <b>2.055</b>	0.314; <b>0.163</b>	0.003; <b>0.002</b>	0.020; <b>0.009</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.853; <b>0.528</b>	0.961; <b>0.601</b>	2.418; <b>1.546</b>	0.055; <b>0.026</b>	0.0004; <b>0.0002</b>	0.004; <b>0.002</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.421; <b>0.279</b>	0.483; <b>0.323</b>	1.316; <b>0.871</b>	0.014; <b>0.007</b>	0.00001; <b>0.00004</b>	0.0009; <b>0.0003</b>

Table E.7: Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC default detector design.

<b>Selection [FCC]</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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	92.488; <b>90.713</b>	92.593; <b>90.805</b>	93.438; <b>91.774</b>	75.819; <b>73.318</b>	3.960; <b>3.764</b>	53.495; <b>49.121</b>
$p_T(j_1, j_2) \geq 200$ GeV	16.56915.985	17.033; <b>16.397</b>	33.802; <b>33.005</b>	17.811; <b>16.902</b>	0.742; <b>0.711</b>	1.056; <b>0.991</b>
<hr/>						
$p_T(j_1) \geq 400$ GeV	8.606; <b>8.187</b>	9.156; <b>8.696</b>	20.910; <b>19.807</b>	7.008; <b>6.628</b>	0.183; <b>0.174</b>	0.446; <b>0.422</b>
$p_T(j_2) \geq 350$ GeV	5.698; <b>5.419</b>	6.161; <b>5.850</b>	13.119; <b>12.147</b>	3.927; <b>3.704</b>	0.121; <b>0.115</b>	0.263; <b>0.250</b>
$p_T(j_1 + j_2) \geq 100$ GeV	4.639; <b>4.399</b>	4.968; <b>4.707</b>	9.576; <b>8.908</b>	3.311; <b>3.114</b>	0.070; <b>0.056</b>	0.223; <b>0.211</b>
$\tau_{21}(j_1) < 0.55$	3.638; <b>3.119</b>	3.904; <b>3.340</b>	7.606; <b>6.417</b>	1.385; <b>1.031</b>	0.018; <b>0.013</b>	0.140; <b>0.113</b>
$\tau_{21}(j_2) < 0.55$	1.702; <b>1.256</b>	1.878; <b>1.391</b>	4.020; <b>2.965</b>	0.540; <b>0.327</b>	0.005; <b>0.003</b>	0.069; <b>0.048</b>
$FW2(j_1) > 0.2$	1.279; <b>0.831</b>	1.421; <b>0.922</b>	3.241; <b>2.149</b>	0.320; <b>0.174</b>	0.003; <b>0.002</b>	0.020; <b>0.009</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.860; <b>0.566</b>	0.973; <b>0.644</b>	2.440; <b>1.633</b>	0.056; <b>0.029</b>	0.0004; <b>0.00002</b>	0.004; <b>0.002</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.421; <b>0.299</b>	0.487; <b>0.348</b>	1.327; <b>0.921</b>	0.014; <b>0.008</b>	0.00009; <b>0.00007</b>	0.0008; <b>0.0005</b>

Table E.8: Cumulative efficiency, in percentage, of each event selection criterion of the baseline analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with  $\eta, \phi \times 2$ .

<b>Selection [FCC gran.]</b>	$\eta, \phi \times 2]$	$hh \rightarrow b\bar{b}b\bar{b}$ (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	100; 100	100; 100	100; 100
No. b-tags $\geq 4$	92.542; 91.202	92.621; 91.306	93.473; 92.346	75.847; 73.910	3.964; 3.794	53.516; 49.359	
$p_T(j_1, j_2) \geq 200$ GeV	16.607; 16.139	17.048; 16.568	33.826; 33.292	17.836; 17.085	0.742; 0.718	1.053; 0.997	
<hr/>							
$p_T(j_1) \geq 400$ GeV	8.631; 8.294	9.159; 8.806	20.945; 20.028	7.020; 6.717	0.183; 0.177	0.446; 0.425	
$p_T(j_2) \geq 350$ GeV	5.715; 5.485	4.121; 5.924	13.141; 12.293	3.937; 3.759	0.121; 0.117	0.263; 0.253	
$p_T(j_1 + j_2) \geq 100$ GeV	4.651; 4.461	4.966; 4.760	9.591; 9.040	3.318; 3.161	0.069; 0.067	0.222; 0.212	
$\tau_{21}(j_1) < 0.55$	3.662; 3.192	3.914; 3.413	7.639; 6.561	1.396; 1.053	0.018; 0.013	0.140; 0.115	
$\tau_{21}(j_2) < 0.55$	1.721; 1.279	1.890; 1.419	4.047; 3.028	0.550; 0.334	0.005; 0.003	0.070; 0.050	
$FW2(j_1) > 0.2$	1.297; 0.872	1.432; 0.949	3.268; 2.239	0.326; 0.183	0.003; 0.002	0.021; 0.011	
$(100 < M_{SD}(j1) < 135)$ GeV	0.871; 0.593	0.978; 0.676	2.456; 1.699	0.056; 0.031	0.0005; 0.0002	0.004; 0.002	
$(100 < M_{SD}(j2) < 135)$ GeV	0.425; 0.312	0.489; 0.364	1.330; 0.956	0.014; 0.008	0.00009; 0.00006	0.0009; 0.0006	

Table E.9: Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS default detector, as implemented in Delphes.

Selection [ATLAS]	$hh \rightarrow b\bar{b}b\bar{b}$ (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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	88.690; <b>83.646</b>	88.787; <b>83.734</b>	89.624; <b>84.504</b>	71.617; <b>66.487</b>	3.749; <b>3.354</b>	51.782; <b>46.516</b>
$p_T(j_1, j_2) \geq 200$ GeV	15.533; <b>13.966</b>	15.941; <b>14.367</b>	32.001; <b>29.581</b>	16.299; <b>14.299</b>	0.685; <b>0.601</b>	0.985; <b>0.862</b>
<hr/>						
$p_T(j_1) \geq 300$ GeV	12.599; <b>11.276</b>	13.061; <b>11.730</b>	28.960; <b>26.663</b>	11.627; <b>10.146</b>	0.390; <b>0.338</b>	0.669; <b>0.594</b>
$p_T(j_1 + j_2) \geq 100$ GeV	10.185; <b>9.081</b>	10.526; <b>9.423</b>	21.414; <b>19.582</b>	9.932; <b>8.632</b>	0.227; <b>0.197</b>	0.574; <b>0.508</b>
$\tau_{21}(j_1) < 0.4$	4.980; <b>4.004</b>	5.152; <b>4.149</b>	11.251; <b>9.303</b>	1.692; <b>1.738</b>	0.026; <b>0.040</b>	0.157; <b>0.136</b>
$\tau_{21}(j_2) < 0.4$	1.139; <b>0.974</b>	1.220; <b>1.035</b>	3.394; <b>2.909</b>	0.234; <b>0.276</b>	0.003; <b>0.006</b>	0.031; <b>0.029</b>
$ \Delta\eta(h_h)  < 1.5$	0.891; <b>0.753</b>	0.960; <b>0.803</b>	2.920; <b>2.480</b>	0.113; <b>0.138</b>	0.001; <b>0.003</b>	0.019; <b>0.017</b>
$FW2(j_1) > 0.2$	0.796; <b>0.567</b>	0.858; <b>0.601</b>	2.675; <b>1.955</b>	0.082; <b>0.064</b>	0.001; <b>0.001</b>	0.010; <b>0.007</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.610; <b>0.396</b>	0.658; <b>0.422</b>	2.134; <b>1.442</b>	0.022; <b>0.016</b>	0.0002; <b>0.0002</b>	0.003; <b>0.003</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.333; <b>0.182</b>	0.360; <b>0.196</b>	1.245; <b>0.738</b>	0.005; <b>0.003</b>	0.00004; <b>0.00003</b>	0.0007; <b>0.0005</b>

Table E.10: Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity.

Selection [ATLAS gran.]	$hh \rightarrow b\bar{b}b\bar{b}$ (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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; 100
No. b-tags $\geq 4$	92.226; <b>85.896</b>	92.322; <b>85.896</b>	93.185; <b>86.656</b>	75.464; <b>67.886</b>	3.930; <b>3.448</b>	53.144; <b>45.851</b>
$p_T(j_1, j_2) \geq 200$ GeV	16.504; <b>14.820</b>	16.918; <b>15.204</b>	33.632; <b>30.867</b>	17.580; <b>15.151</b>	0.732; <b>0.644</b>	1.043; <b>0.917</b>
<hr/>						
$p_T(j_1) \geq 300$ GeV	13.43; <b>11.950</b>	13.902; <b>12.392</b>	30.516; <b>27.876</b>	12.552; <b>10.792</b>	0.417; <b>0.364</b>	0.709; <b>0.630</b>
$p_T(j_1 + j_2) \geq 100$ GeV	10.860; <b>9.597</b>	11.212; <b>9.929</b>	22.587; <b>20.401</b>	10.731; <b>9.185</b>	0.241; <b>0.211</b>	0.609; <b>0.538</b>
$\tau_{21}(j_1) < 0.4$	5.324; <b>3.286</b>	5.506; <b>3.391</b>	11.920; <b>7.502</b>	1.772; <b>1.209</b>	0.025; <b>0.027</b>	0.167; <b>0.108</b>
$\tau_{21}(j_2) < 0.4$	1.131; <b>0.571</b>	1.213; <b>0.604</b>	3.448; <b>1.746</b>	0.217; <b>0.126</b>	0.002; <b>0.003</b>	0.031; <b>0.016</b>
$ \Delta\eta(hh)  < 1.5$	0.920; <b>0.439</b>	0.990; <b>0.465</b>	3.031; <b>1.491</b>	0.110; <b>0.1058</b>	0.001; <b>0.001</b>	0.019; <b>0.009</b>
$FW2(j_1) > 0.2$	0.841; <b>0.304</b>	0.903; <b>0.320</b>	2.827; <b>1.092</b>	0.084; <b>0.027</b>	0.001; <b>0.0003</b>	0.011; <b>0.003</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.658; <b>0.202</b>	0.710; <b>0.217</b>	2.295; <b>0.770</b>	0.022; <b>0.005</b>	0.0001; <b>0.00001</b>	0.004; <b>0.001</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.375; <b>0.107</b>	0.408; <b>0.117</b>	1.409; <b>0.446</b>	0.006; <b>0.001</b>	0.00003; <b>0.000004</b>	0.0007; <b>0.0002</b>

Table E.11: Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the ATLAS granularity with  $\eta \times 4$ .

<b>Selection [ATLAS gran. <math>\eta \times 4</math>]</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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$		92.259; <b>87.076</b>	92.358; <b>87.167</b>	93.228; <b>88.015</b>	70.510; <b>69.147</b>	3.935; <b>3.523</b>	53.198; <b>46.560</b>
$p_T(j_1, j_2) \geq 200$ GeV		16.544; <b>15.131</b>	16.965; <b>15.566</b>	33.674; <b>31.481</b>	17.626; <b>15.567</b>	0.734; <b>0.665</b>	1.044; <b>0.938</b>
<hr/>							
$p_T(j_1) \geq 300$ GeV		13.486; <b>12.229</b>	13.938; <b>12.696</b>	30.569; <b>28.428</b>	12.600; <b>11.100</b>	0.418; <b>0.375</b>	0.711; <b>0.645</b>
$p_T(j_1 + j_2) \geq 100$ GeV		10.889; <b>9.827</b>	11.250; <b>10.179</b>	22.634; <b>20.830</b>	10.774; <b>9.456</b>	0.242; <b>0.217</b>	0.612; <b>0.550</b>
$\tau_{21}(j_1) < 0.4$		5.407; <b>3.336</b>	5.589; <b>3.439</b>	12.079; <b>7.598</b>	1.789; <b>1.139</b>	0.024; <b>0.023</b>	0.169; <b>0.106</b>
$\tau_{21}(j_2) < 0.4$		1.158; <b>0.556</b>	1.241; <b>0.591</b>	3.515; <b>1.737</b>	0.219; <b>0.111</b>	0.002; <b>0.002</b>	0.032; <b>0.015</b>
$ \Delta\eta(hh)  < 1.5$		0.944; <b>0.430</b>	1.013; <b>0.455</b>	3.092; <b>1.490</b>	0.111; <b>0.051</b>	0.001; <b>0.001</b>	0.020; <b>0.009</b>
$FW2(j_1) > 0.2$		0.863; <b>0.331</b>	0.924; <b>0.348</b>	2.890; <b>1.202</b>	0.085; <b>0.027</b>	0.001; <b>0.0003</b>	0.011; <b>0.003</b>
$(100 < M_{SD}(j1) < 135)$ GeV		0.676; <b>0.231</b>	0.726; <b>0.242</b>	2.358; <b>0.882</b>	0.022; <b>0.006</b>	0.001; <b>0.0001</b>	0.004; <b>0.002</b>
$(100 < M_{SD}(j2) < 135)$ GeV		0.386; <b>0.129</b>	0.418; <b>0.136</b>	1.449; <b>0.533</b>	0.006; <b>0.002</b>	0.00005; <b>0.00002</b>	0.0009; <b>0.0002</b>

Table E.12: Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with  $\phi/2$ .

<b>Selection [FCC gran. <math>\phi/2</math>]</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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	92.473; <b>90.243</b>	92.567; <b>90.345</b>	93.420; <b>91.220</b>	75.798; <b>72.741</b>	3.961; <b>3.732</b>	53.474; <b>48.883</b>
$p_T(j_1, j_2) \geq 200$ GeV	16.603; <b>15.838</b>	17.025; <b>16.255</b>	33.788; <b>32.732</b>	17.812; <b>16.706</b>	0.741; <b>0.705</b>	1.054; <b>0.983</b>
<hr/>						
$p_T(j_1) \geq 300$ GeV	13.526; <b>12.827</b>	14.222; <b>13.292</b>	30.676; <b>29.595</b>	12.739; <b>11.917</b>	0.421; <b>0.396</b>	0.718; <b>0.674</b>
$p_T(j_1 + j_2) \geq 100$ GeV	10.936; <b>10.308</b>	11.291; <b>10.664</b>	22.722; <b>21.716</b>	10.902; <b>10.164</b>	0.244; <b>0.228</b>	0.617; <b>0.575</b>
$\tau_{21}(j_1) < 0.4$	5.792; <b>4.306</b>	5.987; <b>4.451</b>	12.899; <b>9.769</b>	1.928; <b>0.390</b>	0.023; <b>0.015</b>	0.186; <b>0.129</b>
$\tau_{21}(j_2) < 0.4$	1.286; <b>0.738</b>	1.385; <b>0.792</b>	3.874; <b>2.377</b>	0.248; <b>0.106</b>	0.002; <b>0.001</b>	0.037; <b>0.019</b>
$ \Delta\eta(hh)  < 1.5$	1.054; <b>0.608</b>	1.135; <b>0.652</b>	3.414; <b>2.119</b>	0.127; <b>0.052</b>	0.001; <b>0.0004</b>	0.024; <b>0.012</b>
$FW2(j_1) > 0.2$	0.974; <b>0.534</b>	1.046; <b>0.568</b>	3.216; <b>1.925</b>	0.102; <b>0.040</b>	0.001; <b>0.0004</b>	0.013; <b>0.005</b>
$(100 < M_{SD}(j1) < 135)$ GeV	0.770; <b>0.423</b>	0.830; <b>0.453</b>	2.643; <b>1.582</b>	0.026; <b>0.010</b>	0.0002; <b>0.0001</b>	0.004; <b>0.002</b>
$(100 < M_{SD}(j2) < 135)$ GeV	0.442; <b>0.263</b>	0.480; <b>0.281</b>	1.626; <b>1.016</b>	0.007; <b>0.003</b>	0.00004; <b>0.000002</b>	0.0009; <b>0.0005</b>

Table E.13: Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC default detector.

	<b>Selection [FCC]</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)	$4b + j$	$jj + 0/1/2j$	$t\bar{t}$
98	Gen level	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; 100
	No. b-tags $\geq 4$	92.488; <b>90.713</b>	92.593; <b>90.805</b>	93.438; <b>91.774</b>	75.819; <b>73.318</b>	3.960; <b>3.764</b>	53.495; 49.121
	$p_T(j_1, j_2) \geq 200$ GeV	16.602; <b>15.985</b>	17.033; <b>16.363</b>	33.802; <b>33.005</b>	17.811; <b>16.902</b>	0.742; <b>0.711</b>	1.056; <b>0.991</b>
	$p_T(j_1) \geq 300$ GeV	13.521; <b>12.943</b>	14.007; <b>13.388</b>	30.691; <b>39.849</b>	12.744; <b>12.06</b>	0.422; <b>0.401</b>	0.718; <b>0.677</b>
	$p_T(j_1 + j_2) \geq 100$ GeV	10.933; <b>10.409</b>	11.301; <b>10.748</b>	22.736; <b>21.927</b>	10.901; <b>10.291</b>	0.245; <b>0.232</b>	0.617; <b>0.580</b>
	$\tau_{21}(j_1) < 0.4$	5.838; <b>4.444</b>	6.043; <b>4.585</b>	12.986; <b>10.064</b>	1.953; <b>1.252</b>	0.023; <b>0.015</b>	0.186; <b>0.133</b>
	$\tau_{21}(j_2) < 0.4$	1.309; <b>0.771</b>	1.410; <b>0.826</b>	3.921; <b>2.466</b>	0.256; <b>0.111</b>	0.002; <b>0.001</b>	0.037; <b>0.020</b>
	$ \Delta\eta(hh)  < 1.5$	1.071; <b>0.636</b>	1.154; <b>0.680</b>	3.455; <b>2.199</b>	0.130; <b>0.055</b>	0.001; <b>0.0005</b>	0.024; <b>0.012</b>
	$FW2(j_1) > 0.2$	0.989; <b>0.566</b>	1.064; <b>0.603</b>	3.254; <b>2.018</b>	0.105; <b>0.044</b>	0.001; <b>0.0004</b>	0.014; <b>0.006</b>
(100 < $M_{SD}(j1) < 135$ ) GeV	(100 < $M_{SD}(j1) < 135$ ) GeV	0.783; <b>0.452</b>	0.844; <b>0.486</b>	2.673; <b>1.672</b>	0.026; <b>0.011</b>	0.0002; <b>0.0001</b>	0.004; <b>0.002</b>
	(100 < $M_{SD}(j2) < 135$ ) GeV	0.446; <b>0.281</b>	0.486; <b>0.303</b>	1.642; <b>1.074</b>	0.007; <b>0.003</b>	0.00003; <b>0.000003</b>	0.0007; <b>0.0005</b>

Table E.14: Cumulative efficiency, in percentage, of each event selection criterion of the optimized analysis for the signal background samples, for particle flow jets (black) and calorimeter jets (blue). The double horizontal line marks the pre-selection cuts. These results were obtained using the FCC granularity with  $\eta, \phi \times 2$ .

	<b>Selection [FCC gran.]</b> $\eta, \phi \times 2]$	$hh \rightarrow b\bar{b}b\bar{b}$ (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; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>	100; <b>100</b>
No. b-tags $\geq 4$	92.542; 91.202	92.621; 91.306	93.473; <b>92.346</b>	75.847; <b>73.910</b>	3.964; <b>3.794</b>	51.782; 49.359	
$p_T(j_1, j_2) \geq 200$ GeV	16.607; <b>16.139</b>	17.048; <b>16.534</b>	33.826; <b>33.292</b>	17.836; <b>17.085</b>	0.742; <b>0.718</b>	1.053; <b>0.997</b>	
$p_T(j_1) \geq 300$ GeV	13.529; <b>13.087</b>	14.012; <b>13.536</b>	30.720; <b>30.121</b>	12.762; <b>12.207</b>	0.422; <b>0.406</b>	0.716; <b>0.684</b>	
$p_T(j_1 + j_2) \geq 100$ GeV	10.941; <b>10.542</b>	11.305; <b>10.875</b>	22.762; <b>22.170</b>	10.919; <b>10.291</b>	0.244; <b>0.235</b>	0.615; <b>0.583</b>	
$\tau_{21}(j_1) < 0.4$	5.869; <b>4.571</b>	6.075; <b>4.716</b>	13.069; <b>10.340</b>	1.974; <b>1.296</b>	0.024; <b>0.016</b>	0.190; <b>0.137</b>	
$\tau_{21}(j_2) < 0.4$	1.321; <b>0.800</b>	1.429; <b>0.861</b>	3.962; <b>2.554</b>	0.262; <b>0.117</b>	0.002; <b>0.001</b>	0.039; <b>0.021</b>	
$ \Delta\eta(hh)  < 1.5$	1.078; <b>0.659</b>	1.169; <b>0.710</b>	3.483; <b>2.274</b>	0.134; <b>0.058</b>	0.001; <b>0.001</b>	0.024; <b>0.013</b>	
$FW2(j_1) > 0.2$	0.997; <b>0.592</b>	1.077; <b>0.635</b>	3.280; <b>2.100</b>	0.107; <b>0.047</b>	0.001; <b>0.0004</b>	0.013; <b>0.006</b>	
$(100 < M_{SD}(j1) < 135)$ GeV	0.787; <b>0.472</b>	0.851; <b>0.512</b>	2.691; <b>1.741</b>	0.027; <b>0.011</b>	0.0002; <b>0.0001</b>	0.004; <b>0.002</b>	
$(100 < M_{SD}(j2) < 135)$ GeV	0.451; <b>0.292</b>	0.490; <b>0.320</b>	1.650; <b>1.114</b>	0.007; <b>0.003</b>	0.00004; <b>0.00003</b>	0.0009; <b>0.0004</b>	

## Appendix F

### Extra plots

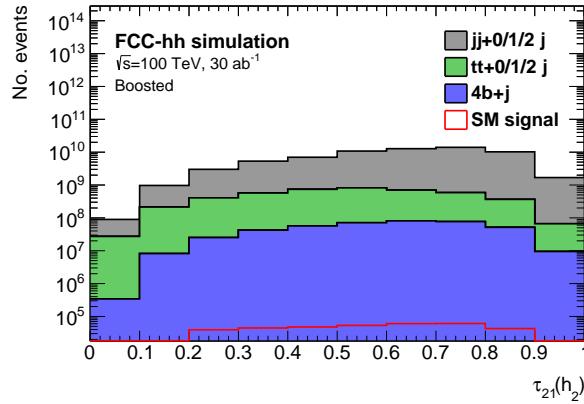


Figure F.1:  $\tau_{21}$  distribution for the sub leading Higgs candidate for the baseline analysis. The histogram is normalized to  $\mathcal{L} = 30 \text{ ab}^{-1}$ .

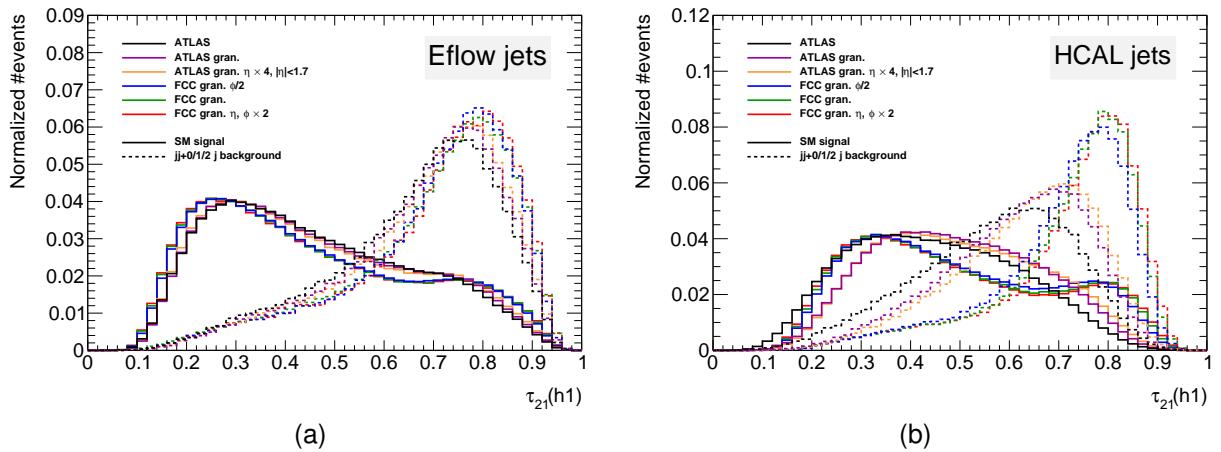


Figure F.2: Sub leading Higgs candidate  $\tau_{21}$  distribution for eflow jets (a) and for HCAL jets (b). The colors indicate the different detector configurations. The distributions are shown for the signal (filled lines) and for the  $jj + 0/1/2 j$  background (dashed lines).

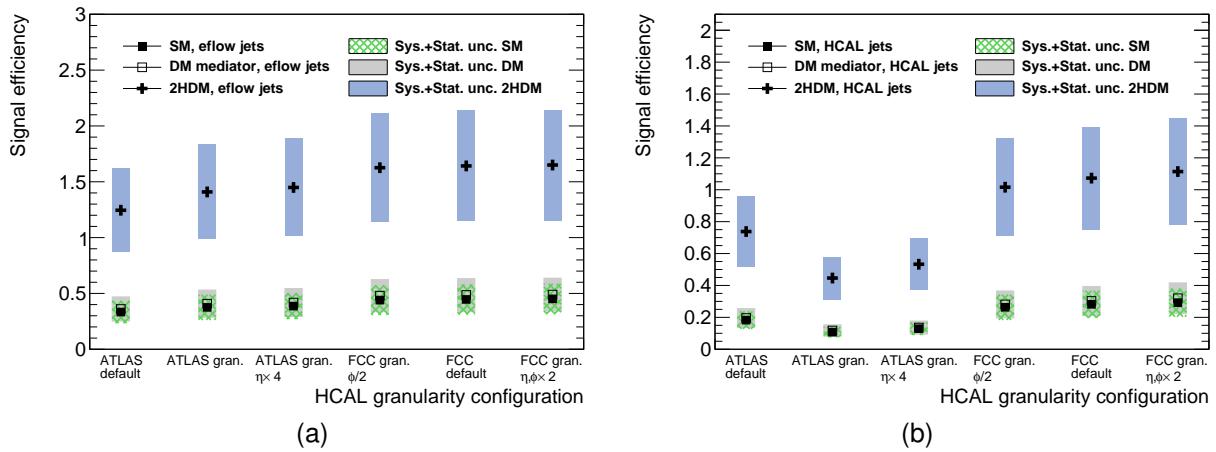


Figure F.3: Signal efficiency as a function of the detector configuration for particle flow jets (a) and for calorimeter jets (b). Three signal models are shown: SM (filled squares), 1 TeV DM mediator (empty squares) and type II 2HDM with  $m_H = 900$  GeV (crosses). The statistical error bars are drawn but are smaller than the markers.