IFT - UNESP
INSTITUTO DE FÍSICA TEÓRICA

unesp
UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"

PhD Thesis IFT–T.002/20

Exploring the Higgs sector beyond the Standard Model with the Top Yukawa coupling: a phenomenological and experimental search

Leônidas Augusto Fernandes do Prado

Advisor
Ricardo D'Elia Matheus

Co-Advisor
Aurore Savoy-Navarro

July, 2020

Prado, Leônidas Augusto Fernandes do.
P896e Exploring the Higgs sector beyond the standard model with the top Yukawa coupling: a phenomenological and experimental search / Leônidas Augusto Fernandes do Prado. – São Paulo, 2020
128 f. : il.

Tese (doutorado) – Universidade Estadual Paulista (Unesp), Instituto de Física Teórica (IFT), São Paulo
Orientador: Ricardo D'Elia Matheus
Coorientador: Aurore Savoy-Navarro

1. Modelo padrão (Física nuclear). 2. Grande colisor de hádrons (França e Suiça). 3. Higgs, Bosons de. I. Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Física Teórica (IFT), São Paulo. Dados fornecidos pelo autor(a).

Resumo

A produção de $t\bar{t}H$ no LHC permite medir diretamente o acoplamento de Yukawa do quark top, previsto pelo Modelo Padrão (MP) como $\sqrt{2}m_t/v \approx 1$. Esta tese mostra como nova física pode fornecer diferentes valores para esse acoplamento para quatro modelos teóricos específicos, dentre diversas possíveis extensões do MP, e usa medidas experimentais para colocar limites no espaço de parâmetro dos modelos. Os primeiros dois são extensões simples do MP e contêm somente férmions *vector-like*, que compartilham os mesmos números quânticos do setor de quarks top-bottom, levando a misturas entre esses estados. Esses modelos prevêem uma diminuição na seção de choque de $t\bar{t}H$ e estão atualmente muito limitados por medidas experimentais. Os últimos dois modelos são realizações mínimas de modelos de Higgs composto, baseados na quebra de simetria de $SO(5) \rightarrow SO(4)$. No setor do top, duas possibilidades são exploradas: ressonâncias fermiônicas na representação fundamental **5** de $SO(5)$ e a representação simétrica **14** de $SO(5)$. Os efeitos desses modelos no processo de produção de $t\bar{t}H$ e $t\bar{t}HH$ são apresentados para energia de 14 TeV da fase *High Luminosity* do LHC, e para futuros colisores de hadrons de alta energia. O trabalho fenomenológico, feito na primeira parte desta tese, mostrou o interesse em se estudar o processo de produção de $t\bar{t}HH$ além da produção de $t\bar{t}H$ estudada até o momento.

A segunda parte desta tese apresenta resultados da busca experimental da produção de $t\bar{t}HH$ usando dados de colisões próton-próton correspondendo a uma luminosidade integrada de 41.5 fb^{-1} gravado pelo experimento CMS para uma energia de centro de massa de 13 TeV no ano de 2017. Essa foi a primeira vez que essa busca foi feita com dados reais. Os candidatos a eventos de $t\bar{t}HH$ são selecionados baseando-se no critério de realçar ambos os canais de decaimento de léptons mais jatos do sistema $t\bar{t}$, e o canal de decaimento do par de bósons de Higgs em dois pares de quark bottom-anti-bottom. Para aumentar a sensibilidade da busca, eventos selecionados são separados em diferentes categorias, em que são esperadas diferentes taxas de eventos de sinal e fundos. Um ajuste combidado, de *templates* de análise discriminante multivariada, aos dados, através de todas as diferentes categorias é feito para extrair os resultados. O valor obtido para a intensidade do sinal de $t\bar{t}HH$, relativo à seção de choque do MP, $\mu = \sigma/\sigma_{\texttt{SM}}$, para essa assinatura específica, é: $4.9^{+14.4}_{-12.8}$, obtida após o ajuste aos dados coletados pelo CMS. A estrutura detalhada da análise desenvolvida aqui servirá para uma futura análise com dados coletados durante todo o período (*Run 2*) do LHC, referente aos anos de 2016 a 2018, correspondendo a uma luminosidade total integrada de 140 fb^{-1}, a 13 TeV no LHC. Esse valor é cerca de 3 vezes maior que a luminosidade usada na presente análise e corresponde ao maior volume de dados coletados até o momento no LHC. Isso pode permitir a extração preliminar de um sinal de $t\bar{t}HH$. Esse trabalho continuará com dados a serem coletados em *Runs* futuras do LHC (*Run 3*) e no *High Luminosity* LHC (HL-LHC).

Palavras-Chave: Fenomenologia; Modelo Padrão; LHC; CMS; Acoplamentos do Higgs; Modelos de Higgs Composto.

Abstract

The $t\bar{t}H$ production at the LHC allows the direct measurement of the top-Higgs Yukawa coupling, predicted by the Standard Model (SM) to be $\sqrt{2}m_t/v \approx 1$. This thesis shows how new physics can provide different values of this coupling in four specific theoretical models within extensions of the SM and uses experimental measurements to constrain the parameter space of such models. The first two are simple extensions of the SM containing only vector-like fermions that share the same quantum numbers of the top-bottom quark sector, leading to mixing between these states. These models predict a decrease in the $t\bar{t}H$ cross sections and are currently very constrained by experimental measurements. The last two models are minimal realizations of composite Higgs models, based on the symmetry breaking pattern $SO(5) \to SO(4)$. In the top sector, two possibilities are considered: fermion resonances in the fundamental **5** representation of $SO(5)$ and in the symmetric **14** representation. The effects of these models on the $t\bar{t}H$ and $t\bar{t}HH$ production processes are presented for the 14 TeV High Luminosity phase of the LHC, and future high energy hadron colliders. The phenomenological work carried out in the first part of this thesis work led to pointing the interest of the $t\bar{t}HH$ production process in addition to the $t\bar{t}H$ production studied so far.

The second part of this thesis also presents results of the experimental search for the production of $t\bar{t}HH$ using proton-proton collision data corresponding to an integrated luminosity of 41.5 fb^{-1} recorded with the CMS experiment at a center-of-mass energy of 13 TeV in the year 2017. This is the first time this search is performed with real data. The candidate $t\bar{t}HH$ events are selected with criteria enhancing the lepton plus jets decay channels of the $t\bar{t}$ system and the decay of the double Higgs bosons into two bottom quark-antiquark pairs. In order to increase the sensitivity of the search, selected events are split into several categories with different expected signal and background rates. A combined fit of multivariate discriminant templates across all categories to data is performed to extract the result. The achieved value of the $t\bar{t}HH$ signal strength relative to the standard model cross section, $\mu = \sigma/\sigma_{\texttt{SM}}$, for this specific signature is: $4.9^{+14.4}_{-12.8}$, after a fit to CMS data. The detailed analysis framework developed will serve for the analysis with the data running period (Run 2) from 2016 to 2018 with a total integrated luminosity of about 140 fb^{-1} at 13 TeV at LHC. This is about 3 times more luminosity than what is used in this analysis. It corresponds to the largest amount of data collected so far at LHC. This could allow extracting a preliminary $t\bar{t}HH$ signal. This work will be carried on with data also to be collected in future Runs of the LHC (Run 3) and at the High Luminosity LHC (HL-LHC).

Keywords: Phenomenology; Standard Model; LHC; CMS; Higgs Couplings; Composite Higgs Models.

ACKNOWLEDGMENTS

I would like to thank the “Instituto de Física Teórica” (IFT), that provided, via CAPES, support during the beginning of my PhD from July of 2015 untill July of 2016. I thank as well the São Paulo Research Foundation (FAPESP) under Grants No. 2013/01907-0 (SPRACE thematic project), No. 2015/26624-6 and No. 2018/11505-0 for support of this PhD from March 2017 do December of 2019.

In particular, me and A. Savoy-Navarro would like to thank FAPESP for the excellence Fellowship under grant N0 2018/11505-0, obtained for one year from September 1^{st} 2018 to August 31^{st} 2019. It gave me the possibility to spend one year abroad for completing my PhD and extending the phenomenology topic I had so far developed to an experimental analysis at the LHC at CERN, within the IRFU-CEA, University Paris Saclay and especially within the CMS team of this research Center in France. Both also thank the Department of Particle Physics, at IRFU-CEA which hosted and provided additional support and working base during this overall year.

I would also like to thank Aurore Savoy-Navarro, who was awarded with the CAPES (Coordination for the Improvement of Higher Education Personnel) and the Ministry of Education in Brazil for three years grant in the “Programma Pesquisador Visitante Especial, PPVE" with Universidade Paulista Júlio de Mesquita Filho, UNESP at São Paulo, Brazil from 01/03/2014 to 28/02/2017 under the Grant 88887.090208/2014-00. This SWB grant also provided my full support for staying at the Laboratory of Astroparticles and Cosmology, APC, University Paris-Diderot and CNRS/IN2P3, in France, from August 2016 to March 2017 under the grant No: 88887.116917/2016-00. I would like to thank the APC Laboratory and CNRS/IN2P3 for the hospitality and support for providing all the needed working conditions during my stay and beyond.

I give a special thanks to my supervisors Ricardo D’Elia Matheus and again to Aurore Savoy-Navarro for all the support without which this work wouldn’t be possible.

I would like to also thank SPRACE for providing me access to the infrastructure of CMS for the period of July of 2017 to November of 2018 and from January of 2020 up to the end of this thesis.

I also thank all my collaborators at KIT, that together with Aurore Savoy-Navarro are the co-authors of my experimental analysis. they provided me all the support that made this analysis possible. I also thank KIT for the three day hospitality during which I could learn the basics of the analysis and do the first step of this one year and a half work. I give a special thanks to Matthias Schröder, Karim El Morabit, Jan van der Linden and Philip Keicher, with whom I was more closely in contact and were always helpful throughout the whole period of my analysis work. Their feedback was important for being able to achieve the current results.

I thank my colaborators in Brazil (Leonardo de Lima, Carlos Bautista and Eduardo Pontón) that, together with my supervisors made this thesis possible. A special thanks to the late Eduardo Pontón, with whom I collaborated since the beginning of my PhD, and who had the patience to teach me very complicated aspects of quantum field theory

and was always available for answering my questions, even during the last of his days. The work of this thesis has his touch almost everywhere.

I also thank all my colleagues and friends for making this thesis possible, each in their own way. A special thanks to Andresa Campos, Vinícius Terra, Henrique Flores, Ana Retore and Nathaly Spano for all the coffee breaks in IFT and beyond. I also thank Benedetta Nodari (grazie mille), Chenfan Zhang, Sergey Lapin and Anton Bogachev for the good company in the APC office. I thank Jelena Mijušković (hvala!) and Yunus Vilkkavaara Cankoçak (👾) for the coffees at Saclay and the company during the long trips of RER B. I also can not forget Diogo Borges, Aline Passos, Marcela Silva, Priscila Rothier, Roberto Pina, Luis Virgílio and Damily Diniz for the good times that started in Paris and continues in Brazil. Thanks to Allan Covacevice an Vinícius Amaro for all the data-science discussions while adventuring over Pandora. I can't wait to work more with you! George Takeuchi, you took me to work so many times. I thank you for all the trips. Michel Haddad, Luiz Leitão, and the whole "Alto Fluxo" band, thanks for all the tubaínas. At last I thank my family and my fiancée Camila Coelho for all the patience and support.

And a final thanks to all friends and family that provided financial support during this last semester.

TABLE OF CONTENTS

Title	Page
RESUMO	I
ABSTRACT	III
ACKNOWLEDGMENT	V
TABLE OF CONTENTS	VII
LIST OF FIGURES	XI
LIST OF TABLES	XVII
Chapter One: Introduction	1
Chapter Two: Impact of vector-like fermion hypothesis in the production of $t\bar{t}H$ at the LHC and the measurement of the top-Higgs Yukawa coupling	5
2.1 Theoretical and Phenomenological Motivations	5
2.1.1 Deviations in the Top Yukawa Coupling Without Deviations in the Higgs Gluon-Fusion Rate	7
2.2 Simple Vector-Like Quark Models	8
2.2.1 The Vector-like Singlet Model	10
2.2.2 The Vector-like Doublet Model	15
2.3 Experimental Constraints for the VLF Singlet and Doublet Models	19
2.3.1 Top Quark Width	20
2.3.2 Z Pole Measurements	20

Title Page

2.3.3 Higgs decay width measurements 22

2.3.4 Experimental Searches for $t\bar{t}H$ at the LHC 24

2.3.5 Experimental Searches for vector-like resonances 24

2.4 Numerical Analysis 25

2.4.1 Parameter Space scan 25

2.4.2 Monte Carlo Simulations 26

2.4.3 Results of this overall phenomenology study 26

2.5 Chapter Conclusions and Remarks 29

Chapter Three: Impact of Composite Higgs Models on the $t\bar{t}H$ and $t\bar{t}HH$ production processes at hadron colliders 31

3.1 Theoretical introduction to Composite Higgs Models 33

3.1.1 The Fermion Sector of the $MCHM_5$ 34

3.1.2 The Fermion Sector of the $MCHM_{14}$ 37

3.1.3 Higgs couplings to the gluons and light fermions 39

3.2 Numerical Analysis of the MCHMs 41

3.2.1 Definition of the Parameter Space 41

3.2.2 The mass spectrum and the physical top-Higgs Yukawa coupling 44

3.2.3 The production of $t\bar{t}H$ in pp collisions 45

3.2.4 The production of $t\bar{t}HH$ in pp collisions 46

3.2.5 The scan over the parameter space 47

3.3 Results of the MCHM parameter space scan 49

3.3.1 Results on $t\bar{t}H$ production 49

3.3.2 Results on $t\bar{t}HH$ production 51

3.4 Selection of example parameter space points 51

3.4.1 Selection of parameter space example points for MCHM$_5$ 52

3.4.2 Selection of parameter space example points for MCHM$_{14}$ 54

3.4.2.1 NR-$t\bar{t}$HH contributions in MCHM$_5$ and MCHM$_{14}$ 55

3.5 Chapter Conclusions and Remarks 57

Chapter Four: Experimental search for the $t\bar{t}$HH production at the CMS experiment at the LHC at CERN 59

4.1 The CMS experiment at the LHC as the experimental framework for a first exploration of the $t\bar{t}$HH process 60

4.1.1 The Large Hadron Collider pp collider, LHC, at CERN 60

4.1.2 The CMS experiment: the detector and the key detector systems for this search 62

4.1.3 Trigger system, Data flow, CMS software and analysis framework 68

4.2 Physics motivations for the $t\bar{t}$HH search 71

4.3 Data and simulation samples and event reconstruction 73

4.3.1 Data and simulation samples 73

4.3.2 Data samples corresponding to specific High Level Triggers and Run periods 76

4.3.3 Event reconstruction 76

4.4 Final Event selection 80

4.5 Analysis strategy 82

4.5.1 Description of the $t\bar{t}$HH (4b) Analysis based on Deep Neural Network tools 82

4.6 Systematic uncertainties and Results 88

4.6.1 Study of systematic uncertainties 88

4.6.2 Analysis results with Asimov data 90

4.6.3 Final results with real data 95

4.7 Perspectives 96

4.7.1 Overall Plan for the $t\bar{t}HH$ search: from now to High Luminosity LHC 100

4.8 Chapter concluding remarks 101

Chapter Five: Conclusions 103

Appendix A: Representations of SO(5) 105

Appendix B: Embeddings of SO(4) into SO(5) 109

Appendix C: Explicit form of gauge and Higgs interactions 111

Appendix D: The Deep Neural Network Input Variables 115

References 117

LIST OF FIGURES

Figure	Description	Page
1	The physical mass of the vector-like top partner (M^T_{phys}) as a function of: (a) the mass parameter M, for fixed $y = 1$, and (b) the parameter y, for fixed $M = 750$ GeV. For $y = 0$ there is no mixing and $M^T_{phys} = M$.	12
2	Diagrams, generated by Madgraph [39], corresponding to $gg \to t\bar{t}H$.	13
3	The physical top-Higgs Yukawa coupling (y_t^{phys}) as a function of (a) the parameter y, for fixed $M = 750$ GeV, and (b) the parameter M, for fixed $y = 1$.	14
4	The physical top partner-Higgs Yukawa coupling (y_T^{phys}) as a function of (a) the parameter M, for fixed $y = 1$, and (b) the parameter y, for fixed $M = 750$ GeV.	14
5	Contours for the partial decay width of the Higgs into bottom pairs, normalized to the SM, $\Gamma_{H\to b\bar{b}}/\Gamma^{SM}_{H\to b\bar{b}}$, as a function of M^B_{phys} and y_b^{phys}/y_b^{SM}. The red lines show the upper bound by ATLAS ($\sqrt{0.99}$) and lower bounds by ATLAS ($\sqrt{0.25}$) and CMS ($\sqrt{0.38}$) [2], after using Eq. (2.63). The upper bound by CMS is weaker than the ATLAS one and is not shown. The yellow line gives the limit as obtained from LEP/SLC [41], showing that the bottom-Higgs Yukawa coupling must be very close to the SM. The 'non-perturbative' region indicates when $y_1 > 4$, where the tree-level analysis has to be interpreted carefully.	23
6	The top width normalized to the SM as a function of the physical vector-like top partner mass, M^T_{phys} (in GeV), and the $t\bar{t}H$ production cross-section via gluon fusion, for (a) the doublet model and (b) the singlet model. The region below the dashed yellow line is excluded by the top width measurement. In both plots, the tree-level analysis is not reliable in the 'non-perturbative' region with $y > 4$. The small gray areas marked as "unphysical" correspond to regions where the top mass of 150.7 GeV cannot be reproduced.	27

7 The top width normalized by the SM as a function of the physical vector-like top partner mass, M^T_{phys} (in GeV) and the physical top-Higgs Yukawa coupling y_t^{phys}, for (a) the doublet model and (b) the singlet model. The region below the dashed yellow line is excluded by the top width measurement. In both plots, the tree-level analysis is not reliable in the 'non-perturbative' region with $y > 4$. The small gray areas marked as "unphysical" correspond to regions where the top mass of 150.7 GeV cannot be reproduced. 28

8 Contour plot of the top Yukawa coupling, $\mu_{\rm top} \equiv y_{\rm top}/y^{\rm SM}_{\rm top}$ (blue and red colors), as a function of M_1 and M_4. Blue(red) indicates $\mu_{\rm top} < 1$ ($\mu_{\rm top} > 1$). Curves of masses of the lightest $Q = 2/3$ resonance, $M_{T^{(1)}}$, is shown. Darker bands show direct exclusion limits for the resonances, assuming they decay only in the bW, tZ and tH channels [48, 49]. The top mass can not be reproduced in the white region, and limits on the ggH coupling is also shown (relevant only for a region in MCHM$_{14}$). 44

9 Representative diagrams of the non-resonant $t\bar{t}$HH process. There are three distinct processes containing vertices that involve only the Yukawa vertex (left), the Higgs trilinear self-coupling (middle) and the "double Higgs" Yukawa vertex (right), this last one being a new diagram arising in the MCHMs. 47

10 Invariant mass of the top quark and the hardest Higgs boson for a particular set of parameters for MCHM$_5$ ($M_1 = -2.5$ TeV, $M_4 = 2$ TeV, $f = 1.8$ TeV, $y_L = 1$). The blue curve shows the full $t\bar{t}$HH process in the MCHM$_5$, while the NR-$t\bar{t}$HH distribution is shown in red. The green curve corresponds to the process in the SM. A center of mass energy of 14 Tev (27 TeV) is used for the generation of events (MadGraph 5) for the upper (lower) plots, which are produced with MadAnalysis 5 [82]. 48

11 The $\mu(t\bar{t}H)$ is shown as a function of f, and is independent of the pp collision energy. The lightest vector-like mass is color coded. The upper left (right) plot corresponds to the Regions I (II) of MCHM$_5$. The lower left, central and right plots correspond to the Regions I, II + IV and III of MCHM$_{14}$, respectively. The blue arrow in the upper right plot indicates that the example point P_4 is outside the horizontal range of the plot with $f = 2450$ GeV. Also in this Figure are example points to be defined in Section 3.4. 50

12 Value of $\mu(t\bar{t}HH)$ as a function of the mass of the lightest $Q = 2/3$ vector-like fermion, for 14 (blue) and 27 TeV (red) center of mass energies. The left (right) plots correspond to the $MCHM_5$ ($MCHM_{14}$). 52

13 Ratio between the NR-$t\bar{t}HH$ cross section and the total $t\bar{t}HH$ cross section as a function of the lightest $Q = 2/3$ vector-like mass, for 14 and 27 TeV center of mass energies. The left (right) plots correspond to the $MCHM_5$ ($MCHM_{14}$). 52

14 Correlation between $\mu(NR - t\bar{t}HH)$ and $\mu(t\bar{t}H)$, for 14 and 27 TeV center of mass energies. The left (right) plots correspond to $MCHM_5$ ($MCHM_{14}$). 53

15 The CERN accelerator complex [84]. 61

16 The plan for the LHC and High Luminosity LHC (HL-LHC). It shows the activities of the machine for each year, showing the energy and luminosity during the running time. On the bottom it is also shown the HL-LHC technical equipment phases starting from the LHC Run 1 period. 62

17 Mean number of interactions per bunch crossing for the 2017 pp run at 13 TeV. The plot uses the "CMS recommended" value of 69.2 mb pp inelastic cross section, which is determined by finding the best agreement with data and is recommended for CMS analyses. 63

18 Cumulative measured luminosity versus day delivered. This plot uses the approved physics calibration for the 2017 pp Run [85]. 63

19 Overview of the CMS detector showing the main detector systems. The inner tracker, HCAL and ECAL are inside the superconducting coil and the muon detector chambers are interlayered between iron return yokes [87]. 64

20 A transverse slice of the CMS detector showing the paths of different kinds of particles through the different sub detectors [88]. 65

21 Layout of the different layers and disks in the old (current in the image) and current (upgrade in the image) pixel detectors. It shows the new added pixel barrel layer and the extra third endcap disk [90] 66

22 Architecture of the CMS trigger, showing the path events take starting from the detector, passing through the level 1 trigger (L1), the high level trigger (HLT), after which the offline analysis is made. Taken from Ref. [100]. 68

23 Architecture of the CMS L1 trigger, hardware based, that receives information from the muon and calorimeter detector systems. The signals are stored in pipeline memories in the front end electronics. The accepted events have a rate of 100 kHz out of the 40 MHz collision event rate. Taken form Ref. [86]. 69

24 Cartoon schema and the detailed data flow in CMS from raw data (as provided directly from the detectors) up to the final stage of data analysis, passing through triggers, offline computing (data archiving, reconstruction, and extraction of analysis object data (AOD) files), and, ending on the data analysis. Taken from Ref. [100]. 70

25 Summary of the path of data files, along different steps of offline computing, including file sizes per event, from RAW data to analysis object data (AOD) files, which are further used by physics groups to perform analyses [103, 104]. 70

26 Total cross sections at the LO and NLO in QCD for HH production channels, as a function of the self-interaction coupling λ, taken from Ref [105]. λ is varied for the self-interaction couplings, but the mass of the Higgs boson is fixed to be $m_H = 125$ GeV. The dashed (solid) lines and light-(dark) colored bands correspond to the LO (NLO) results and to the scale of the PDF uncertainties added linearly. The SM values of the cross-section are obtained at $\lambda/\lambda_{SM} = 1$. 72

27 Representative diagrams for the $t\bar{t}HH$ process, illustrating the two distinct physical subprocesses: the Yukawa vertex and the Higgs trilinear self-coupling as expected in the SM. 73

28 Representative diagrams for the $t\bar{t}HH$ process in the semileptonic decay mode. It illustrates the two distinct physical subprocesses containing only Yukawa vertices (left) and containing the Higgs trilinear self-coupling (right). 81

29 Data/MC comparisons of selected observables used as inputs to DNN before the fit to data. Each of the expected backgrounds is normalized to the corresponding SM cross section: the average mass of all jets ($\mathtt{m}_{\mathtt{j}}^{\mathtt{avg}}$); the average b-tagging discriminant value of all b-tagged jets($\mathtt{d}_{\mathtt{b}}^{\mathtt{avg}}$); the likelihood ratio discriminating between events with 4 b quark jets and 2 b quark jets (BLR), presented as ln[BLR/(1 − BLR)] ($\mathrm{BLR}^{\mathtt{trans}}$); the scalar sum of p_{T} of all b-tagged jets ($\mathtt{H}_{\mathtt{T}}^{\mathtt{b}}$); The difference in shapes between signal($\mathrm{t\bar{t}HH}$) and background processes can be seen. Error bands correspond to all systematics and are added in quadrature. Data points are included and a ratio of the data yields and MC background expectation is shown at the bottom. 85

30 Data/MC comparisons of selected observables used as inputs to DNN before the fit to data. Each of the expected backgrounds is normalized to the corresponding SM cross section: the number of b-tagged jets with a DeepCSV b-tagging discriminant value higher than 0.8001, called tight working point ($\mathtt{N}_{\mathtt{b}}^{(\mathtt{tight})}$). Note that the selection of events for this analysis requires events with DeepCSV discriminant higher than 0.4941 (medium working point); The number of selected jets ($\mathtt{N}_{\mathtt{Jets}}$). The difference in shapes between signal($\mathrm{t\bar{t}HH}$) and background processes can be seen. Error bands correspond to all systematics and are added in quadrature. Data points are included and a ratio of the data yields and MC background expectation is shown at the bottom. 86

31 The confusion matrix (left) shows the percentage of true class events that are classified in each one of the six predicted classes (output nodes). A good performance corresponding to a ROC-AUC=0.830 is observed. The loss function (right) is monitored at each epoch. Both training (blue) and validation (red) curves follow the same trend and don't cross, thus no overtraining is observed. The loss function gets close to the minimum very early in the training (around the first 50 epochs). 87

32 Comparison of the DNN discriminant distributions expected for the signal and background processes, constructed with a subset of 20% of the events listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The discriminant value shown in the horizontal axis is the probability relative to the most probable process (p_i) that each sample is classified as. 89

33 Discriminant shapes before the fit to data. They are constructed with signal and background processes listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The data corresponds to CMS collected data in the year of 2017. 92

34 Nuisance parameter pulls and impacts, sorted by the largest impact, for the top 30 most impacting uncertainties, obtained from the fit to Asimov data (blind analysis), for 41.5 fb^{-1} integrated luminosity. The pulls of the nuisance parameters (black markers) are computed relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$. The impact $\Delta\hat{r}$ is computed as the difference of the nominal best fit value of μ and the best fit value obtained when varying the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus/minus (1-σ) its post-fit uncertainty (colored areas). 94

35 Discriminant shapes after the fit to data. They are constructed with signal and background processes listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The data corresponds to CMS collected data in the year of 2017. 97

36 Discriminant shapes after the fit to data for the signal node. On the right it is shown a zoom of the plot on the left in the region of less than one expected event. The plots are constructed with signal and background processes listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The data corresponds to CMS collected data in the year of 2017. 98

37 Nuisance parameter pulls and impacts, sorted by the largest impact, for the top 30 most impacting uncertainties, obtained from the fit to CMS data collected in 2017, with 41.5 fb^{-1} of integrated luminosity. The pulls of the nuisance parameters (black markers) are computed relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$. The impact $\Delta\hat{r}$ is computed as the difference of the nominal best fit value of μ and the best fit value obtained when varying the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus/minus (1-σ) its post-fit uncertainty (colored areas). 98

LIST OF TABLES

Table	Description	Page
1	Relation between mass and gauge eigenstates in the fermion sector of the singlet model. The mixing angles are functions of the free parameters y and M of Eq. (2.8).	11
2	Relation between mass and gauge eigenstates in the fermion sector of the doublet model. The mixing angles are functions of the parameters y_1 and M in the bottom sector, and of y_2 and M in the top sector, as defined in Eq. (2.28).	16
3	A summary of the differences between the vector-like singlet and doublet models.	19
4	Range of each free parameter considered for the MadGraph 5 scan, for the two MCHMs considered here. A grid of points is generated in these ranges, which is further fed to the event generator.	47
5	Properties of selected example points in the MCHM_5. P_2 and P_5 belong to Region I ($M_1, M_4 > 0$); P_1, P_3 and P_4 belong to Region II ($M_1 < 0, M_4 > 0$).	53
6	Properties of selected parameter space example points in MCHM_{14}. Points P'$_1$ through P'$_4$ belong to Region III, while Q_1, Q_2 and Q_3 belong to Regions I, II and IV respectively.	55
7	Study of NR-t$\bar{\text{t}}$HH for MCHM_5 example points in Table 5. The cross sections $\sigma_{\not{h}\not{t}}$ and σ_{Yuk} are obtained by disregarding the classes of diagrams on the last column and $\sigma_{\text{NR}}^{\text{t}\bar{\text{t}}\text{HH}}$ is the total NR-ttHH cross section. The cross section σ_{Yuk}^{SM} means the SM trilinear Higgs coupling is turned off. The top Yukawas in the MCHM and SM are y_t and y_t^{SM} respectively.	56
8	Study of NR-t$\bar{\text{t}}$HH for the MCHM_{14} points in Table 6. The cross sections $\sigma_{\not{h}\not{t}}$ and σ_{Yuk} are obtained by disregarding the diagrams shown on Table 7 and $\sigma_{\text{NR}}^{\text{t}\bar{\text{t}}\text{HH}}$ is the total NR-t$\bar{\text{t}}$HH cross section. The cross section $\sigma_{\text{Yuk}}^{\text{SM}}$ means the SM trilinear Higgs coupling is turned off. The top Yukawas in the MCHM and SM are y_t and y_t^{SM} respectively.	56

9 Cross sections computed at NLO QCD for ZHH, WHH, VBF HH, $t\bar{t}$HH and tjHH at 14 TeV center of mass energy. 71

10 Signal and Background Simulation samples used in the analysis. 76

11 List of Files that translate the High Level Trigger (HLT) selection to which correspond these data (refined single electron and single muon triggers) with the corresponding certified Run periods. 77

12 Summary of muon identification (ID) requirements in the single muon channel as well as for the leading and sub-leading muon (ordered by p_{T}) for the dilepton channel. 78

13 Summary of electron identification (ID) requirements in the single electron channel as well as for the leading and sub-leading electron (ordered by p_{T}) for the dilepton channel. 78

14 Jet kinematic selection requirements for the single-lepton channel and for the first two leading jets (ordered in p_{T}) and the next leading jets for the dilepton channel. 80

15 Baseline event selection criteria for selecting single-lepton (SL) events in this $t\bar{t}$HH search. The final selection for this analysis requires one additional b-tagged jet. 81

16 Number of selected events used for fitting (see Section 4.6) (middle column). Number of selected events used for DNN modeling (see Section 4.5) (right column). The baseline selection of Table 15 is applied together with requiring ≥ 3 b-tagged jets. 81

17 Summary of the hyperparameters of the neural network, implemented with their default definitions in Keras. 84

18 Top 20 ranked Input Variables for this analysis, obtained after training the DNN with all variables and ranking them by training weight. 87

19 Inclusive cross section (rate) uncertainties used in the analysis. They are theoretical uncertainties taken from Ref. [106]. Note that an extra 50% rate uncertainty is assigned separately to each of the four considered $t\bar{t} + \mathrm{HF}$ processes. 90

20 Systematic uncertainties considered in this analysis. 91

21 Fit Results for Asimov data (blind analysis). The upper limits are obtained with asymptotic likelihood methods according to formulas of Ref [144]. Also shown is the best fit value of the signal strength modifier $\mu = \sigma/\sigma_{\mathtt{SM}} = \hat{\mu}$. 93

22 Fit Results for CMS data. The upper limits are obtained with asymptotic likelihood methods according to formulas of Ref [144]. Also shown is the best fit value of the signal strength modifier $\mu = \sigma/\sigma_{\mathtt{SM}} = \hat{\mu}$. 96

23 Luminosity scaling for a MC only analysis. Data events for this case are a sum of all the MC simulated events, thus having and expected $\mu = 1$. The expected uncertainties don't take into account the MC bin by bin statistics, and is checked to not influence much this discussion (this is needed for technical reasons). 99

24 The best-fit values of the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for each $t\bar{t}H$ channel at $m_{\mathrm{H}} = 125.6$ GeV. The signal strength in the four-lepton final state is not allowed to be below approximately -6 by the requirement that the expected signal-plus-background event yield must not be negative in either of the two jet multiplicity bins. The observed and expected 95% CL upper limits on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for each $t\bar{t}H$ channel at $m_{\mathrm{H}} = 125.6$ GeV are also shown. Taken from Table 8 of Ref. [147]. 100

25 Summary of the results of this analysis after a fit to CMS data. The upper limits are obtained with asymptotic likelihood methods according to formulas of Ref [144]. Also shown is the best fit value of the signal strength modifier $\mu = \sigma/\sigma_{\mathtt{SM}} = \hat{\mu}$. 102

26 Input variables used in the DNNs (Continued in Table 27.) 115

27 Continued from Table 26. 116

Chapter 1: Introduction

The $t\bar{t}H$ measurement at the Large Hadron Collider (LHC) allows measuring directly the top-Higgs Yukawa coupling which has a value of $\sqrt{2}m_t/v \approx 1$ in the Standard Model (SM). Different kinds of models such as Composite Models and Extra Dimensional Models provide different values for this coupling and together with other experimental measurements it is possible to constrain the parameter space of different Beyond Standard Model (BSM) scenarios, giving valuable information on the search for new physics. This is why, the search and measurement of the $t\bar{t}H$ signal strength, directly connected to the top-Higgs Yukawa coupling, is actively carried on by the two multipurpose experiments, ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), at the LHC since the beginning. This work focuses on models that contain vector-like fermions, where the new fields share the same quantum numbers as some SM fields leading to a mixing between these states. The main consequences are the modification of the top-Higgs Yukawa coupling and the introduction of new resonances. These modifications motivate the study of the production of $t\bar{t}H$ and $t\bar{t}HH$ in the LHC. The first is directly related to the measurement of the top-Higgs Yukawa coupling, while the later, related as well to this coupling, depends on the mass of new resonances and the triple Higgs coupling. The experimental search and study of the $t\bar{t}HH$ production process is important in the light of both new physics and SM studies. For new physics, it is an important channel leading to possible observable deviations from the SM. The cross-section of the $t\bar{t}HH$ process at 13 TeV in the SM is of 0.775 fb at Next Leading Order (NLO).

The increase in total luminosity at the LHC is allowing to explore production processes with cross-sections of few tens of femtobarns (as for example the double Higgs production through gluon fusion) and at the femtobarn level as the Vector Boson Fusion (VBF) double Higgs production that are actively studied now. This present work points out, for the first time, the unique capability of hadron colliders to search for the $t\bar{t}HH$ production. This process offers in addition to its own interest in SM and for exploring the BSM Physics, an access to the triple Higgs coupling, complementary to the double Higgs production. This work presents, for the first time, the search for the $t\bar{t}HH$ process, in the SM framework with CMS collected data in the year of 2017 with 13 TeV center of mass energy at the LHC at CERN

Chapter 2 shows how new vector-like fermions provide different values for the top-Higgs Yukawa coupling and focuses on the $gg \to t\bar{t}H$ production process at parton level. Two simple extensions of the SM are considered, involving a $SU(2)$ singlet vector-like fermion and a $SU(2)$ doublet vector-like fermion. The mixing with the SM fields are described in detail and experimental constraints coming from $Z \to b\bar{b}$, $H \to b\bar{b}$, and $t \to W^+b$ are taken into account.

Chapter 3 considers the production of $t\bar{t}H$ and $t\bar{t}HH$ in the context of two scnarios of Minimal Composite Higgs Models (MCHM). The focus is on minimal models with the symmetry breaking pattern $SO(5) \to SO(4)$. These models are richer in phenomenology when compared to the models presented in Chapter 2 because together with the introduction of vector-like fermions, the Higgs is also composite. Two models are considered explicitly, the first one containing fermion resonances in the fundamental **5** representation of $SO(5)$ and the second one in the symmetric **14** representation. Both $t\bar{t}H$ and $t\bar{t}HH$ processes are simulated in MadGraph considering center of mass energies of 14 TeV, relevant for the High Luminosity LHC (HL-LHC) and also with a major upgrade scenario (in fact total rebuilding of all the magnets) of the LHC machine to reach 27 TeV center of mass energy as addressed in the CERN Yellow report on the HL-HE/LHC [1].

Besides the theoretical and phenomenological study developed in this PhD Thesis, an experimental research work has been performed directly linked to and indeed motivated by the theoretical part of this work. This makes the overall topic of this thesis rather unique, as it embraces the related aspects from the theory to a detailed experimental analysis work, through the most phenomenological aspects.

The manifold interest of the $t\bar{t}HH$ production process at hadron colliders as stressed in the theoretical part of the thesis, has led to exploring for the first time this production process with real data at the CMS experiment at CERN. This challenging search has been performed in these last 2 years, by launching a new analysis line with the data collected by CMS in 2017 at 13 TeV center of mass energy at the LHC with an integrated total luminosity of 41.5 fb^{-1}. This is the subject of Chapter 4 in this thesis. For combining at best the search sensibility while preserving the signal rate, it was chosen to search for candidate $t\bar{t}HH$ events in the final state signature provided by the semi leptonic decay of the $t\bar{t}$ system with the Higgses decaying both to $b\bar{b}$ pairs. The data are selected by the Level-1 lepton (electron or muon) trigger further refined at the High Level Trigger (HLT) of the CMS experiment. The final selection of the events is optimized at the analysis level by requiring at least one lepton (electron or muon), at least 4 jets of which at least 3 are b-tagged jets, within the relevant detector acceptance. The dominant background is the QCD production of a pair of top-quarks. As in the $t\bar{t}H$ ($H \to b\bar{b}$) analysis, a Deep

Neural Network, DNN, analysis is performed in order to categorize selected events in either signal or background and discriminating templates are constructed to be further fit to data. The detailed analysis framework developed will serve as the basis of the overall Run 2 data where the first $t\bar{t}HH$ events are expected to be evidenced. The study of the $t\bar{t}HH$ production process will be further carried on at the Run 3 of the LHC starting February 2022 at 14 TeV with about 300 fb^{-1} expected total integrated luminosity and at the HL-LHC when it could be possible, with more statistics, to constrain the parameter space of BSM models such as the MCHMs of this work.

Final conclusions are presented on Chapter 5.

Chapter 2: Impact of vector-like fermion hypothesis in the production of $t\bar{t}H$ at the LHC and the measurement of the top-Higgs Yukawa coupling

This chapter introduces vector-like fermions (VLF) and their impact on the $pp \to t\bar{t}H$ process, in particular in the gluon fusion sub-process at 13-14 TeV LHC collisions. The aim is to determine if insights on new physics, based on the prediction of the models to be presented here, could be obtained with data collected at the LHC. The addition of VLF that mix with the top quark sector of the SM introduces deviations in the top-Higgs Yukawa coupling without affecting the $gg \to H$ production process. This indirect measurement of the top-Higgs Yukawa coupling agrees well with the SM expectation. This class of models are thus compatible with current Higgs measurements [2]. This chapter starts with theoretical and phenomenological motivations in Section 2.1. The two studied models are described in Section 2.2. The vector-like singlet (doublet) model is described in Section 2.2.1 (Section 2.2.2). Experimental measurements already constrain the models and are summarized in Section 2.3, and the results are presented in Section 2.4. Conclusions and perspectives are presented in Section 2.5. This Chapter reports on the work done in collaboration with R. D'Elia Matheus, E. Pontón and A. Savoy-Navarro, also published as an Internal CMS Note, AN-2017-248 [3].

2.1 Theoretical and Phenomenological Motivations

In the SM there are three generations of fermion families. A fourth generation of quarks is strongly constrained by the observation of the Higgs boson production through the gluon fusion process (see, e.g. [4]). These constraints can be relaxed if new fermions receive a dominant contribution to their masses from physics that respects the electroweak (EW) symmetry. This happens if the left handed (LH) and right handed (RH) chiralities of these new fermions both have the same SM quantum numbers. Such states are called "vector-like" fermions [5, 6]. While a new generation of quarks ("chiral fermions") are essentially excluded, these vector-like fermions are still a possibility.

Several well-motivated BSM theories have vector-like fermions as natural ingredients. For example, the introduction of new compact extra dimensions [7, 8, 9], which allows addressing the SM hierarchy problem [10], naturally lead to heavy Kaluza-Klein (KK) fermions that are vector-like. Further theoretical progress allowed understanding that many such higher dimensional constructions correspond to a dual description of four dimensional (4D) strongly interacting theories [11, 12]. In the 4D dual language, the KK modes can be interpreted as bound states of appropriate fundamental degrees of freedom [13, 14]. Hence, vector-like fermions are also a natural occurrence in such 4D constructions.

The above duality observation has led to the construction of models where the SM degrees of freedom are understood as composite to a certain degree (a scenario called "partial compositeness" [15]). In this language, the Higgs is understood as a purely composite state that would dissolve into its constituents at a "compositeness scale" not far above the EW scale. This scale acts as a physical cutoff on loop corrections to the Higgs mass parameter, thereby addressing the hierarchy problem (i.e. how to understand that the weak scale is so much below other high scales that can be present in nature, such as the Grand Unification or Planck scales). The observation of the Higgs boson at 125 GeV [16, 17], together with the fact that it is a weakly interacting (narrow) resonance, as well as the non-observation of additional nearby resonances, has led to further refinements of the generic "Composite Higgs" scenario, where the Higgs boson is understood as a pseudo-Nambu-Goldstone boson (pNGB) [18].

The top quark is the heaviest SM fermion, which means it has the strongest interactions to the electroweak symmetry breaking sector. This means that the new set of resonances associated to the top quark in the above mentioned BSM scenarios also has a significant coupling to the Higgs boson. Moreover these resonances are expected to be lighter than the ones associated with lighter SM fermions and would be the first ones observed experimentally. These light top partners are also essential in explaining the Higgs potential in these pNGB constructions.

In summary, vector-like fermions connected to the top quark are a natural ingredient in a number of realistic proposals for BSM physics (extra-dimensions, strongly interacting theories with or without a composite Higgs, theories where the Higgs boson is a pNGB and the EW symmetry is broken dynamically). Furthermore, a subset of the top partners often has the same SM quantum numbers as the (t,b) $SU(2)_L$ doublet or the top $SU(2)_L$ singlet. While top partners with exotic quantum numbers can also be present (and represent an exciting possibility that has been widely studied in the literature both by phenomenologists [19, 20, 21] and by the ATLAS [22] and CMS [23] collaborations),

those vector-like fermions with SM quantum numbers possess an interesting characteristic: they can mix with the SM fermions, and therefore affect observables even at energies below the threshold for direct production of the heavy states. This motivates the focus of this chapter on simplified models that capture these features and expect that the results can be interpreted in larger sets of BSM physics. In Chapter 3, one of these cases, the Composite Higgs Model, is considered in detail in minimal implementations.

The next sections of this chapter will focus on the effects of new vector-like resonances in the associated production of the Higgs with a top quark pair ($t\bar{t}H$). Other modified SM couplings and new couplings can be interesting subjects of future work. This study is expanded for Composite Higgs in the next chapter with the addition of processes involving top quark pairs in association with double Higgs production ($t\bar{t}HH$).

Here, two simple models are considered. The first adds to the SM a fermion that is a SU(2) vector-like singlet, while the second adds to the SM a SU(2) vector-like doublet and are introduced below in Section 2.2. The results describe one ingredient of more complete BSM theories and serve as an introduction to the work presented in Chapter 3.

2.1.1 Deviations in the Top Yukawa Coupling Without Deviations in the Higgs Gluon-Fusion Rate

As mentioned above, an interesting feature of vector-like fermion extensions of the top sector is that they easily allow for interesting deviations from the SM value of the top Yukawa coupling *without* appreciable deviations in the Higgs production cross-section by the gluon fusion process, which is sensitive to the top Yukawa coupling.

The argument is based on well-known Higgs low-energy theorems (see e.g. [24, 25]) and runs as follows (see, e.g. [26, 27, 28]). The ggH coupling is induced at 1-loop order through a triangle diagram involving colored fermions [29, 30, 31, 32]. In the SM, the effect is largely dominated by the top quark (with a subleading contribution from the bottom quark which can safely be neglected in the following argument). The contribution from each (mass eigenstate) fermion to the ggH amplitude is proportional to $\frac{Y_f v}{M_f} A_{\frac{1}{2}}(m_h^2/4M_f^2)$, where M_f and Y_f are the mass and coupling to the Higgs of fermion f, respectively, v is the Higgs vacuum expectation value (v.e.v), m_h is the Higgs boson mass, and $A_{1/2}(\tau)$ is the fermionic loop function, which can be found in [33]. The precise form of the loop function is not needed since the argument only relies on the fact that when $m_h^2 \ll 4M_f^2$, this function reaches a constant asymptotic value, independent of M_f. As a result, the

contribution to ggH due to *heavy* colored fermions is proportional to

$$\sum_f \frac{Y_f v}{M_f} \;=\; \sum_f \frac{v}{M_f}\frac{dM_f}{dv} \;=\; v\frac{d}{dv}\sum_f \ln M_f \;=\; \frac{v}{2}\frac{d}{dv}\ln\det(\mathcal{M}_f^\dagger \mathcal{M}_f)\;, \tag{2.1}$$

where it is used that the Yukawa coupling of the fermion f to the Higgs boson can be obtained as $Y_f = dM_f/dv$, as well as simple properties of logarithms. Writing the last expression in terms of the determinant of the fermion mass matrix, $\mathcal{M}_f$, shows that the effect can be obtained without ever finding the fermion mass eigenstates explicitly: one only needs the fermion mass matrix in any basis, e.g. in the gauge eigenbasis which can be derived from the Lagrangian in a straightforward manner.

It often happens (in particular in the models to be presented here, see Eqs. (2.12) and (2.32) below) that this determinant is simply proportional to v: $\det\mathcal{M}_f = F(\lambda_i)\,v$, where $F(\lambda_i)$ is a function that depends on all other model parameters λ_i (mass parameters and couplings), but not on v. In such a case, it follows that Eq. (2.1) is equal to 1. Since the dominant contributions to the ggH amplitude are from those fermions that couple most strongly to the Higgs boson, which here are precisely the top quark and the new vector-like fermions, the above observation is equivalent to having no deviations from the SM in the ggH amplitude. This corresponds to an interesting cancellation, where any modification of the top quark Yukawa coupling (typically a suppression) is exactly compensated by the additional contributions from the new vector-like fermions. Including the smaller contribution from the bottom quark and related vector-like partners can lead to small, although potentially interesting, deviations of the above picture [34]. More importantly, in other UV completions of the SM, such as the composite Higgs scenarios mentioned in the previous section, there can be additional deviations coming from the composite nature of the Higgs boson. The focus here is on the effects of vector-like fermions which could exist on their own and results should be interpreted as capturing part of the effects of more complicated scenarios. Similar conclusions hold for the loop-induced couplings of the Higgs to photons [35, 36].

2.2 Simple Vector-Like Quark Models

The simplified models under study are explored in [37], based on the SM gauge symmetry:

$$SU(3)_C \times SU(2)_L \times U(1)_Y\;. \tag{2.2}$$

The notation used for the SM-like third generation of fermions is:

$$\Psi_L^1 = \begin{pmatrix} \mathcal{T}_L^1 \\ \mathcal{B}_L^1 \end{pmatrix} \ , \quad \mathcal{T}_R^1 \ , \ \ \mathcal{B}_R^1 \ , \tag{2.3}$$

where Ψ_L^1 is a chiral $SU(2)_L$ doublet with hypercharge 1/6, and $\mathcal{T}_R^1$,$\mathcal{B}_R^1$ are $SU(2)_L$ singlets with hypercharge 2/3 and $-1/3$, respectively. The notation is chosen to make clear that these are not exactly the observed SM states. Those (SM) states are composed of these previous chiral states and the new vector-like fermions, denoted as $\mathcal{T}^2$, with charge 2/3, and $\mathcal{B}^2$, with charge $-1/3$, and are called top-quark and bottom-quark partners, respectively.

The most general renormalizable Lagrangian based on these degrees of freedom, plus the Higgs doublet

$$H = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{2.4}$$

takes the following SM-like form:

$$\mathcal{L}_{SM} = i\overline{\Psi_L^1}\not{D}\Psi_L^1 + i\overline{\mathcal{T}_R^1}\not{D}\mathcal{T}_R^1 + i\overline{\mathcal{B}_R^1}\not{D}\mathcal{B}_R^1 - \left(y_b\overline{\Psi_L^1}H\mathcal{B}_R^1 + y_t\overline{\Psi_L^1}\widetilde{H}\mathcal{T}_R^1 + h.c.\right), \tag{2.5}$$

which is common to both models considered here. The standard notation $\widetilde{H} = i\sigma_2 H^*$ is used. An important reminder is that if there were no new fields, the observed top and bottom quark fields would be $\mathcal{T}_L^1 \equiv t_L$, $\mathcal{B}_L^1 \equiv b_L$, $\mathcal{T}_R^1 \equiv t_R$, $\mathcal{B}_R^1 \equiv b_R$, and the Yukawa couplings y_b and y_t would have the SM values.

The $\mathcal{T}^2$ and $\mathcal{B}^2$ fields are added to the models as summarized below:

- *Singlet Model*: addition of a vector-like $SU(2)_L$ singlet with hypercharge 2/3 (a "top partner") denoted by $\mathcal{T}^2 = \mathcal{T}_L^2 + \mathcal{T}_R^2$.
- *Doublet Model*: addition of a vector-like $SU(2)_L$ doublet with hypercharge 1/6 (containing "top and bottom partners") denoted by $\Psi^2 = \Psi_L^2 + \Psi_R^2$, where

$$\Psi_L^2 = \begin{pmatrix} \mathcal{T}_L^2 \\ \mathcal{B}_L^2 \end{pmatrix} \ , \quad \Psi_R^2 = \begin{pmatrix} \mathcal{T}_R^2 \\ \mathcal{B}_R^2 \end{pmatrix} \ . \tag{2.6}$$

Both left and right hand chiralities of these fermions have the same $SU(3)_C \times SU(2)_L \times U(1)_Y$ quantum numbers, which define them as vector-like quarks. Details of each model are presented in the next sections.

2.2.1 The Vector-like Singlet Model

The Lagrangian of the model is defined by:

$$\mathcal{L} = \mathcal{L}_{SM} + \mathcal{L}_{NP} \ , \tag{2.7}$$

where $\mathcal{L}_{SM}$ is the SM-like Lagrangian defined by Eq. (2.5), and

$$\mathcal{L}_{NP} \ = \ i\overline{\mathcal{T}_L^2}\not{D}\mathcal{T}_L^2 + i\overline{\mathcal{T}_R^2}\not{D}\mathcal{T}_R^2 - (M\overline{\mathcal{T}_L^2}\mathcal{T}_R^2 + \text{h.c.}) - (y\overline{\Psi_L^1}\widetilde{H}\mathcal{T}_R^2 + \text{h.c.}) \ , \tag{2.8}$$

which introduces the vector-like mass parameter M and a Yukawa coupling y involving Ψ_L^1 and $\mathcal{T}_R^2$. These two new Lagrangian parameters determine the physical properties of the mass eigenstates after electroweak symmetry breaking (EWSB): the physical masses and the Yukawa interactions with the Higgs boson.

In the unitary gauge, the field $\widetilde{H}$ is replaced as

$$\widetilde{H} \to \frac{1}{\sqrt{2}} \begin{pmatrix} v+h \\ 0 \end{pmatrix}, \tag{2.9}$$

where h is the physical Higgs boson and $v \approx 246$ GeV is the v.e.v. of the Higgs field. The non-vanishing v.e.v. leads to mass terms that are not diagonal in the gauge eigenbasis, and a rotation to the mass eigenbasis must be performed.

The mass and Yukawa terms from $\mathcal{L}$ can be written as

$$\mathcal{L}_{\text{mass}} = -\bar{\chi}_L^t M_{(1)}^t \chi_R^t - \bar{\chi}_L^t H_{(1)}^t \chi_R^t + \text{h.c.} \ , \tag{2.10}$$

where

$$\chi_L^t = \begin{pmatrix} \mathcal{T}_L^1 \\ \mathcal{T}_L^2 \end{pmatrix}, \quad \chi_R^t = \begin{pmatrix} \mathcal{T}_R^1 \\ \mathcal{T}_R^2 \end{pmatrix}, \tag{2.11}$$

and

$$M_{(1)}^t = \begin{pmatrix} \frac{y_t v}{\sqrt{2}} & \frac{yv}{\sqrt{2}} \\ 0 & M \end{pmatrix} \ , \quad H_{(1)}^t = \begin{pmatrix} \frac{y_t}{\sqrt{2}} & \frac{y}{\sqrt{2}} \\ 0 & 0 \end{pmatrix} h \ . \tag{2.12}$$

The mass matrix is diagonalized in the following basis (see Table 1):

$$\chi_L^t \to \chi_L^{t\,\prime} = \begin{pmatrix} t_L \\ T_L \end{pmatrix} = U_L^t \chi_L^t \ , \quad \chi_R^t \to \chi_R^{t\,\prime} = \begin{pmatrix} t_R \\ T_R \end{pmatrix} = U_R^t \chi_R^t \ . \tag{2.13}$$

The matrices U_L^t and U_R^t are unitary and can be parameterized as:

$$U_L^t = \begin{pmatrix} \cos\theta_L & -\sin\theta_L \\ \sin\theta_L & \cos\theta_L \end{pmatrix}, \quad U_R^t = \begin{pmatrix} \cos\theta_R & -\sin\theta_R \\ \sin\theta_R & \cos\theta_R \end{pmatrix}, \tag{2.14}$$

where the mixing angles θ_R and θ_L depend on M, y and y_t (as well as v). It is, however,

Field	Charge	In terms of the mass eigenbasis
$\mathcal{T}_L^1$	2/3	$\mathcal{T}_L^1 = t_L \cos\theta_L + T_L \sin\theta_L$
$\mathcal{T}_L^2$		$\mathcal{T}_L^2 = -t_L \sin\theta_L + T_L \cos\theta_L$
$\mathcal{T}_R^1$		$\mathcal{T}_R^1 = t_R \cos\theta_R + T_R \sin\theta_R$
$\mathcal{T}_R^2$		$\mathcal{T}_R^2 = -t_R \sin\theta_R + T_R \cos\theta_R$

Table 1: Relation between mass and gauge eigenstates in the fermion sector of the singlet model. The mixing angles are functions of the free parameters y and M of Eq. (2.8).

convenient to eliminate y_t in favor of the smallest mass eigenvalue, which is identified as the running top quark mass at 1 TeV [38], $m_t = 150.7$ GeV, via

$$y_t = \frac{m_t\sqrt{2M^2 + y^2v^2 - 2m_t^2}}{v\sqrt{M^2 - m_t^2}} . \tag{2.15}$$

One then finds that the mixing angle in the LH sector is given by

$$\cos^2\theta_L = \frac{2(M^2 - m_t^2)^2}{2(M^2 - m_t^2)^2 + y^2v^2M^2} , \tag{2.16}$$

$$\sin^2\theta_L = \frac{y^2v^2M^2}{2(M^2 - m_t^2)^2 + y^2v^2M^2} , \tag{2.17}$$

while in the RH sector one finds

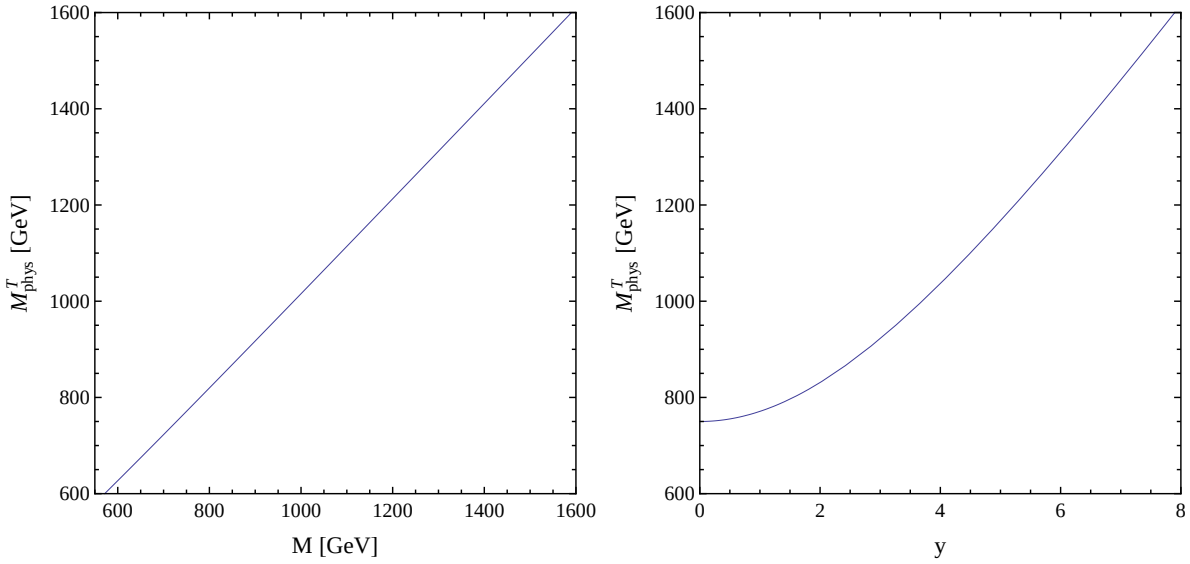
$$\cos^2\theta_R = \frac{2(M^2 - m_t^2)^2 + y^2v^2(M^2 - m_t^2)}{2(M^2 - m_t^2)^2 + y^2v^2M^2} , \tag{2.18}$$

$$\sin^2\theta_R = \frac{y^2v^2m_t^2}{2(M^2 - m_t^2)^2 + y^2v^2M^2} . \tag{2.19}$$

The physical mass of the new fermion can then be written as

$$M_{phys}^T = M\cos\theta_L\cos\theta_R + \frac{v\sin\theta_L(y\cos\theta_R + y_t\sin\theta_R)}{\sqrt{2}} . \tag{2.20}$$

Fig. 1 shows that M_{phys}^T increases as both M and y increase.



(a) **(b)**

Figure 1: The physical mass of the vector-like top partner (M^T_{phys}) as a function of: (a) the mass parameter M, for fixed $y = 1$, and (b) the parameter y, for fixed $M = 750$ GeV. For $y = 0$ there is no mixing and $M^T_{phys} = M$.

Rewriting the Lagrangian in terms of the mass eigenstates (t and T), one can also read the corresponding Yukawa couplings to the Higgs boson:

$$y_t^{phys} = \cos\theta_L \left(y_t \cos\theta_R - y \sin\theta_R\right) \ , \tag{2.21}$$

$$y_T^{phys} = \sin\theta_L \left(y \cos\theta_R + y_t \sin\theta_R\right) \ . \tag{2.22}$$

In terms of the same eigenstates, there is also the coupling that mixes t and T with the Higgs boson. The coupling of $\bar{t}_L T_R H$ is:

$$y_{tT}^{phys} = \cos\theta_L \left(y \cos\theta_R + y_t \sin\theta_R\right) \ , \tag{2.23}$$

while the coupling of $\bar{T}_L t_R H$ is:

$$y_{Tt}^{phys} = \sin\theta_L \left(y_t \cos\theta_R - y \sin\theta_R\right) \ . \tag{2.24}$$

These new couplings are the entries of the $H^t_{(1)}$ matrix in the new basis:

$$H^t_{(1)}{}' = U^t_L H^t_{(1)} U^{t\,\dagger}_R = \begin{pmatrix} \frac{y_t^{phys}}{\sqrt{2}} & \frac{y_{tT}^{phys}}{\sqrt{2}} \\ \frac{y_{Tt}^{phys}}{\sqrt{2}} & \frac{y_T^{phys}}{\sqrt{2}} \end{pmatrix} h \ . \tag{2.25}$$

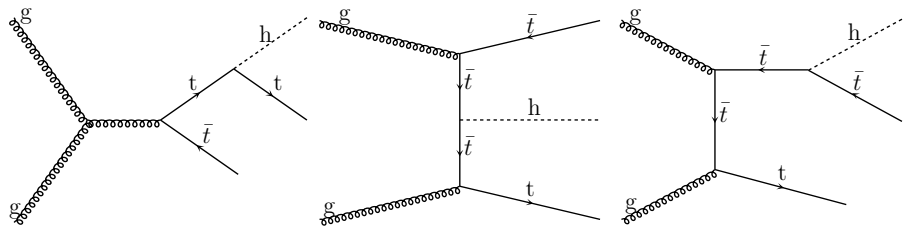


Figure 2: Diagrams, generated by Madgraph [39], corresponding to $gg \to t\bar{t}H$.

An important quantity is the top Yukawa coupling, which is directly related to the production of $t\bar{t}H$, one of the main subjects of this thesis and the focus of this chapter. The relevant Feynman diagrams are identical to those in the SM[1] (see Fig. 2), the only difference being the value of the top Yukawa coupling. This is because QCD can only produce the heavy fermion states, T, in pairs and therefore one cannot reach the desired $t\bar{t}H$ final state when the new physics is explicitly involved in the diagram.[2] Thus, one can immediately relate the $t\bar{t}H$ production cross-section to the SM one:

$$\mu_{t\bar{t}H} = \frac{\sigma^{t\bar{t}H}_{\text{Singlet}}}{\sigma^{t\bar{t}H}_{\text{SM}}} \quad = \quad \left(\frac{y_t^{phys}}{y_t^{\text{SM}}}\right)^2 , \tag{2.26}$$

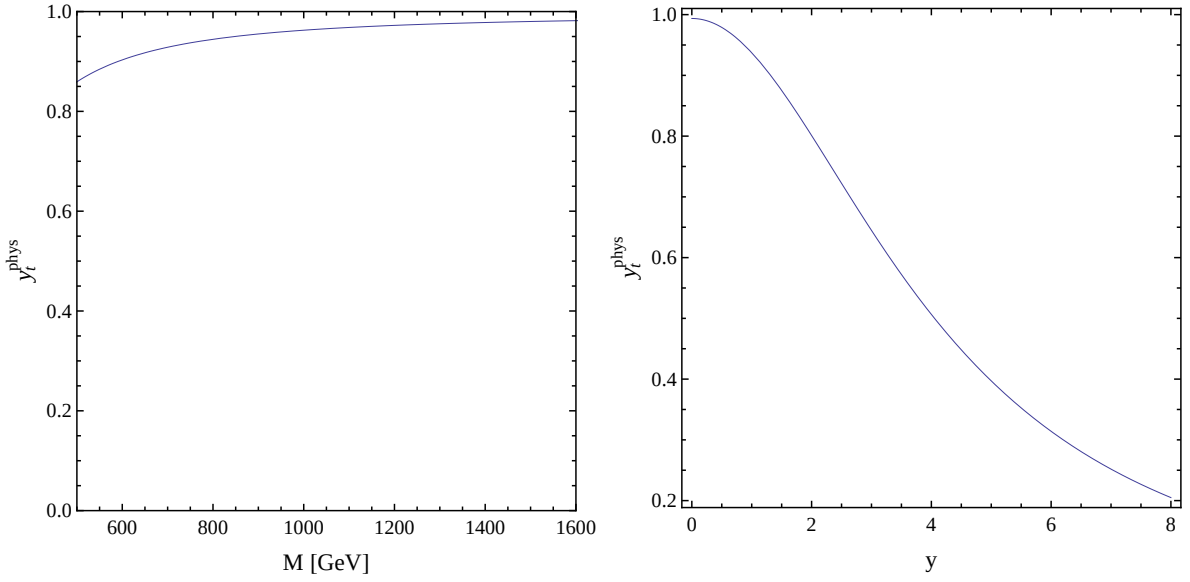
where $\sigma^{t\bar{t}H}_{\text{Singlet}}$ is the $t\bar{t}H$ production cross-section in the singlet model. This ratio $\mu_{t\bar{t}H}$ is with respect to the SM throughout this (and the next) chapter.[3]

Fig. 3 shows the top Yukawa coupling behavior as a function of M and y. Fig. 3a shows that as $M \to \infty$, the top Yukawa coupling goes back to its SM value, reflecting the decoupling behavior of the new physics. Fig. 3b shows that for $y = 0$ (no mixing) the SM top Yukawa is recovered. It also shows that as y increases, the physical top Yukawa decreases. This can be understood from the expression for y_t^{phys} in Eq. (2.21) as follows. Since y_t is fixed by Eq. (2.15), as a function of y it increases monotonically from its SM value of $\sqrt{2}m_t/v$ (at $y = 0$), and reaches the asymptotic linear behavior $y_t \propto y$ for $y \to \infty$. On the other hand, the various mixing angles are also monotonic functions of y: from Eqs. (2.18) and (2.19), $\cos\theta_R$ ($\sin\theta_R$) decreases (increases) from 1 (0) at $y = 0$ to a constant value as $y \to \infty$. Thus, it would seem that the parenthesis in Eq. (2.21) behaves linearly with y for large y. However, a more detailed analysis of the asymptotic region shows that the leading order behavior of each separate term ($= ym_t/M$) cancels out between the two terms, and that, in fact, the parenthesis decreases like $1/y$. This,

[1] Other diagrams, for example those including the vertex $t\bar{T}H$, do not contribute to the $t\bar{t}H$ final state at leading order. However, they will contribute to the $t\bar{t}HH$ production process. Such possibility is discussed in the next chapter in the context of Composite Higgs Models.

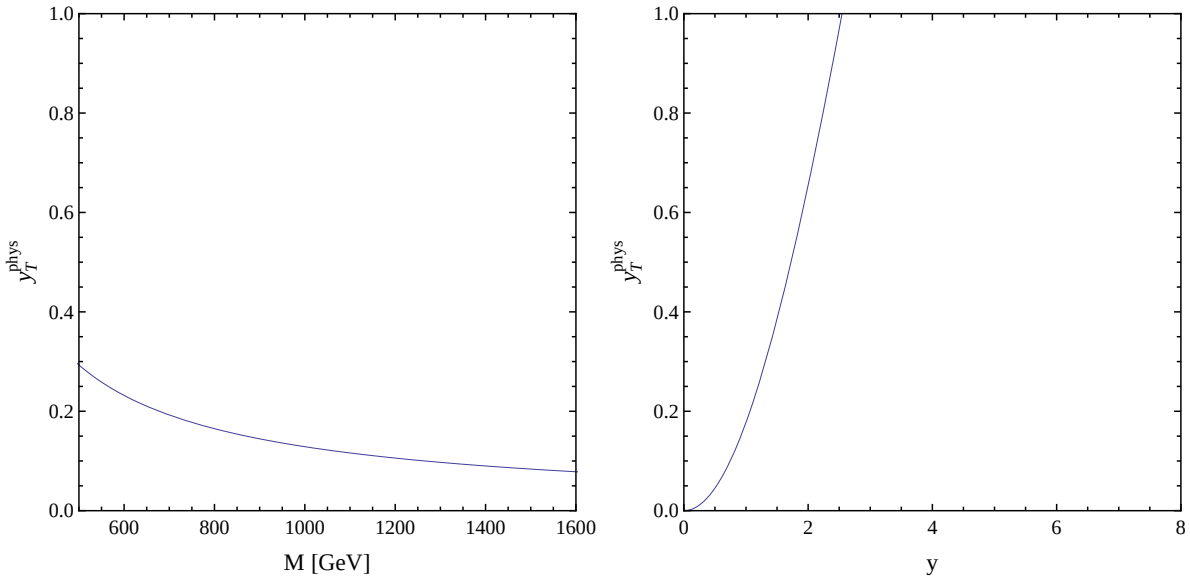
[2] Virtual or real T heavy fermion states can contribute through weak processes and at NLO

[3] The ratio μ is going to be related to the signal signal strength in the experimental part of this work.



(a) **(b)**

Figure 3: The physical top-Higgs Yukawa coupling (y_t^{phys}) as a function of (a) the parameter y, for fixed $M = 750$ GeV, and (b) the parameter M, for fixed $y = 1$.



(a) **(b)**

Figure 4: The physical top partner-Higgs Yukawa coupling (y_T^{phys}) as a function of (a) the parameter M, for fixed $y = 1$, and (b) the parameter y, for fixed $M = 750$ GeV.

together with the monotonic decreasing behavior of $\cos\theta_L$ exhibited in Eq. (2.16), shows that y_t^{phys} decreases like $1/y^2$ at large y, as shown in Fig. 3b.

It is also interesting to see that whenever the top Yukawa increases, the top partner Yukawa decreases, and vice versa, as seen by comparing Figs. 3a and 4a, on the one hand, and 3b and 4b, on the other.

Finally, for later reference, the coupling of the W gauge boson to the physical top and bottom quarks is modified with respect to the SM due to the mixing in the top sector:

$$\frac{g_w}{\sqrt{2}}\left[(\bar{t}\cos\theta_L+\bar{T}\sin\theta_L)\gamma^\mu P_L b\right]W_\mu+\text{h.c.}\ , \tag{2.27}$$

i.e. the Wtb coupling is suppressed by $\cos\theta_L$. This information is used in Subsection 2.3.1 to constrain the parameter space of the model.

2.2.2 The Vector-like Doublet Model

As in the singlet model, the Lagrangian of the doublet model is obtained by adding the new physics Lagrangian to the SM-like Lagrangian of Eq. (2.5), thus obtaining $\mathcal{L} = \mathcal{L}_{SM}+\mathcal{L}_{NP}$, where:

$$\mathcal{L}_{NP} = i\overline{\Psi_L^2}\not{D}\Psi_L^2+i\overline{\Psi_R^2}\not{D}\Psi_R^2-\left(y_1\overline{\Psi_L^2}H\mathcal{B}_R^1+y_2\overline{\Psi_L^2}\widetilde{H}\mathcal{T}_R^1+M\overline{\Psi_L^2}\Psi_R^2+\text{h.c.}\right) \tag{2.28}$$

The dimensionless parameters y_1 and y_2 couple the new physics doublet to the SM-like right-handed bottom and top, respectively, and the notation for the vector-like mass parameter is the same as in the singlet model. Spontaneous symmetry breaking leads to the substitutions (in the unitary gauge)

$$H\to\frac{1}{\sqrt{2}}\begin{pmatrix}0\\v+h\end{pmatrix},\quad \widetilde{H}\to\frac{1}{\sqrt{2}}\begin{pmatrix}v+h\\0\end{pmatrix}, \tag{2.29}$$

and the Lagrangian becomes

$$\begin{aligned}\mathcal{L} &= i\overline{\Psi_L^1}\not{D}\Psi_L^1+i\overline{\Psi_L^2}\not{D}\Psi_L^2+i\overline{\Psi_R^2}\not{D}\Psi_R^2+i\overline{\mathcal{T}_R^1}\not{D}\mathcal{T}_R^1+i\overline{\mathcal{B}_R^1}\not{D}\mathcal{B}_R^1\\ &\quad-\left(y_b\overline{\mathcal{B}_L^1}\mathcal{B}_R^1\frac{1}{\sqrt{2}}(h+v)+y_t\overline{\mathcal{T}_L^1}\mathcal{T}_R^1\frac{1}{\sqrt{2}}(h+v)+\text{h.c.}\right)\\ &\quad-\left(y_1\overline{\mathcal{B}_L^2}\mathcal{B}_R^1\frac{1}{\sqrt{2}}(h+v)+y_2\overline{\mathcal{T}_L^2}\mathcal{T}_R^1\frac{1}{\sqrt{2}}(h+v)+M\overline{\mathcal{T}_L^2}\mathcal{T}_R^2+M\overline{\mathcal{B}_L^2}\mathcal{B}_R^2+\text{h.c.}\right)\ .\end{aligned}$$

Field	Charge	In terms of the mass basis
$\mathcal{T}_L^1$	2/3	$\mathcal{T}_L^1 = t_L \cos\theta_L^t + T_L \sin\theta_L^t$
$\mathcal{T}_L^2$		$\mathcal{T}_L^2 = -t_L \sin\theta_L^t + T_L \cos\theta_L^t$
$\mathcal{T}_R^1$		$\mathcal{T}_R^1 = t_R \cos\theta_R^t + T_R \sin\theta_R^t$
$\mathcal{T}_R^2$		$\mathcal{T}_L^2 = -t_R \sin\theta_R^t + T_R \cos\theta_R^t$
$\mathcal{B}_L^1$	-1/3	$\mathcal{B}_L^1 = b_L \cos\theta_L^b + B_L \sin\theta_L^b$
$\mathcal{B}_L^2$		$\mathcal{B}_L^2 = -b_L \sin\theta_L^b + B_L \cos\theta_L^b$
$\mathcal{B}_R^1$		$\mathcal{B}_R^1 = b_R \cos\theta_R^b + B_R \sin\theta_R^b$
$\mathcal{B}_R^2$		$\mathcal{B}_R^2 = -b_R \sin\theta_R^b + B_R \cos\theta_R^b$

Table 2: Relation between mass and gauge eigenstates in the fermion sector of the doublet model. The mixing angles are functions of the parameters y_1 and M in the bottom sector, and of y_2 and M in the top sector, as defined in Eq. (2.28).

Unlike the singlet model, the EWSB induces mixing in both the top and bottom sectors. The mass and Yukawa terms can be written as

$$\mathcal{L}_{\text{mass}} = -\bar{\chi}_L^t M_{(2)}^t \chi_R^t - \bar{\chi}_L^t H_{(2)}^t \chi_R^t - \bar{\chi}_L^b M_{(2)}^b \chi_R^b - \bar{\chi}_L^b H_{(2)}^b \chi_R^b \ , \tag{2.30}$$

where

$$\chi_L^t = \begin{pmatrix} \mathcal{T}_L^1 \\ \mathcal{T}_L^2 \end{pmatrix}, \quad \chi_R^t = \begin{pmatrix} \mathcal{T}_R^1 \\ \mathcal{T}_R^2 \end{pmatrix}, \quad \chi_L^b = \begin{pmatrix} \mathcal{B}_L^1 \\ \mathcal{B}_L^2 \end{pmatrix}, \quad \chi_R^b = \begin{pmatrix} \mathcal{B}_R^1 \\ \mathcal{B}_R^2 \end{pmatrix}, \tag{2.31}$$

and

$$\begin{aligned} M_{(2)}^t &= \begin{pmatrix} \frac{y_t v}{\sqrt{2}} & 0 \\ \frac{y_2 v}{\sqrt{2}} & M \end{pmatrix}, \quad H_{(2)}^t = \begin{pmatrix} \frac{y_t}{\sqrt{2}} & 0 \\ \frac{y_2}{\sqrt{2}} & 0 \end{pmatrix} h \ , \\ M_{(2)}^b &= \begin{pmatrix} \frac{y_b v}{\sqrt{2}} & 0 \\ \frac{y_1 v}{\sqrt{2}} & M \end{pmatrix}, \quad H_{(2)}^b = \begin{pmatrix} \frac{y_b}{\sqrt{2}} & 0 \\ \frac{y_1}{\sqrt{2}} & 0 \end{pmatrix} h \ . \end{aligned} \tag{2.32}$$

The field redefinitions that diagonalize the mass matrices are (see Table 2):

$$\begin{aligned} \chi_L^t \to \chi_L^{t\,\prime} &= \begin{pmatrix} t_L \\ T_L \end{pmatrix} = U_L^t \chi_L^t \ , \qquad \chi_R^t \to \chi_R^{t\,\prime} = \begin{pmatrix} t_R \\ T_R \end{pmatrix} = U_R^t \chi_R^t \ , \\ \chi_L^b \to \chi_L^{b\,\prime} &= \begin{pmatrix} b_L \\ B_L \end{pmatrix} = U_L^b \chi_L^b \ , \qquad \chi_R^b \to \chi_R^{b\,\prime} = \begin{pmatrix} b_R \\ B_R \end{pmatrix} = U_R^b \chi_R^b \ , \end{aligned} \tag{2.33}$$

where the rotation matrices are unitary and parameterized as

$$
\begin{aligned}
U_L^t &= \begin{pmatrix} \cos\theta_L^t & -\sin\theta_L^t \\ \sin\theta_L^t & \cos\theta_L^t \end{pmatrix}, \quad U_R^t = \begin{pmatrix} \cos\theta_R^t & -\sin\theta_R^t \\ \sin\theta_R^t & \cos\theta_R^t \end{pmatrix}, \\
U_L^b &= \begin{pmatrix} \cos\theta_L^b & -\sin\theta_L^b \\ \sin\theta_L^b & \cos\theta_L^b \end{pmatrix}, \quad U_R^b = \begin{pmatrix} \cos\theta_R^b & -\sin\theta_R^b \\ \sin\theta_R^b & \cos\theta_R^b \end{pmatrix}.
\end{aligned}
\tag{2.34}
$$

The mixing angles depend on y_1, y_2, y_t, y_b and M, while both y_t and y_b are eliminated in favor of the lightest mass eigenvalues in the top and bottom sectors, to be identified with the running top and bottom masses at 1 TeV [38], respectively, using the equations

$$
y_t = \frac{m_t\sqrt{2M^2+y_2^2v^2-2m_t^2}}{v\sqrt{M^2-m_t^2}}\ , \tag{2.35}
$$

$$
y_b = \frac{m_b\sqrt{2M^2+y_1^2v^2-2m_b^2}}{v\sqrt{M^2-m_b^2}}\ . \tag{2.36}
$$

The expressions for the mixing angles are similar to the ones for the singlet case, with the difference that θ_L and θ_R are interchanged and that there are also mixing angles for the bottom sector, independent of the ones of the to sector. The expressions for the top sector are

$$
\begin{aligned}
\cos^2\theta_R^t &= \frac{2(M^2-m_t^2)^2}{2(M^2-m_t^2)^2+y_2^2v^2M^2}\ , \\
\sin^2\theta_R^t &= \frac{y_2^2v^2M^2}{2(M^2-m_t^2)^2+y_2^2v^2M^2}\ , \\
\cos^2\theta_L^t &= \frac{2(M^2-m_t^2)^2-y_2^2v^2m_t^2+y_2^2v^2M^2}{2(M^2-m_t^2)^2+y_2^2v^2M^2}\ , \\
\sin^2\theta_L^t &= \frac{y_2^2v^2m_t^2}{2(M^2-m_t^2)^2+y_2^2v^2M^2}\ ,
\end{aligned}
\tag{2.37}
$$

while for the bottom sector one has

$$
\begin{aligned}
\cos^2\theta_R^b &= \frac{2(M^2-m_b^2)^2}{2(M^2-m_b^2)^2+y_1^2v^2M^2}\ , \\
\sin^2\theta_R^b &= \frac{y_1^2v^2M^2}{2(M^2-m_b^2)^2+y_1^2v^2M^2}\ , \\
\cos^2\theta_L^b &= \frac{2(M^2-m_b^2)^2-y_1^2v^2m_b^2+y_1^2v^2M^2}{2(M^2-m_b^2)^2+y_1^2v^2M^2}\ , \\
\sin^2\theta_L^b &= \frac{y_1^2v^2m_b^2}{2(M^2-m_b^2)^2+y_1^2v^2M^2}\ .
\end{aligned}
\tag{2.38}
$$

The physical masses of the heavy states are then

$$\begin{aligned} M_T &= M\cos\theta_L^t\cos\theta_R^t + \frac{v\sin\theta_R^t(y_2\cos\theta_L^t + y_t\sin\theta_L^t)}{\sqrt{2}} \ , \\ M_B &= M\cos\theta_L^b\cos\theta_R^b + \frac{v\sin\theta_R^b(y_1\cos\theta_L^b + y_b\sin\theta_L^b)}{\sqrt{2}} \ . \end{aligned} \tag{2.39}$$

Expressing the Lagrangian in terms of the mass eigenstates, one can read the top sector Yukawa couplings to the Higgs:

$$\begin{aligned} y_t^{phys} &= \cos\theta_R^t\,(y_t\cos\theta_L^t - y_2\sin\theta_L^t) \ , \\ y_T &= \sin\theta_R^t\,(y_2\cos\theta_L^t + y_t\sin\theta_L^t) \ , \end{aligned} \tag{2.40}$$

as well as those of the bottom sector

$$\begin{aligned} y_b^{phys} &= \cos\theta_R^b\,(y_b\cos\theta_L^b - y_1\sin\theta_L^b) \ , \\ y_B &= \sin\theta_R^b\,(y_1\cos\theta_L^b + y_b\sin\theta_L^b) \ . \end{aligned} \tag{2.41}$$

For the top sector, the mixed couplings for the vertices $\bar{t}_L T_R H$ and $\bar{T}_L t_R H$ are, respectively:

$$y_{tT}^{phys} = \sin\theta_R^t\left(y_t\cos\theta_L^t - y_2\sin\theta_L^t\right), \tag{2.42}$$

$$y_{Tt}^{phys} = \cos\theta_R^t\left(y_2\cos\theta_L^t + y_t\sin\theta_L^t\right). \tag{2.43}$$

For the bottom sector, the mixed couplings for the vertices $\bar{b}_L B_R H$ and $\bar{B}_L b_R H$ are, respectively:

$$y_{bB}^{phys} = \sin\theta_R^b\left(y_b\cos\theta_L^b - y_1\sin\theta_L^b\right), \tag{2.44}$$

$$y_{Bb}^{phys} = \cos\theta_R^b\left(y_1\cos\theta_L^b + y_b\sin\theta_L^b\right). \tag{2.45}$$

In terms of the physical couplings, the Higgs matrices in the new basis are:

$$H_{(2)}^{t\,\prime} = \begin{pmatrix} \frac{y_t^{phys}}{\sqrt{2}} & \frac{y_{tT}^{phys}}{\sqrt{2}} \\ \frac{y_{Tt}^{phys}}{\sqrt{2}} & \frac{y_T^{phys}}{\sqrt{2}} \end{pmatrix} h \ , \quad H_{(2)}^{b\,\prime} = \begin{pmatrix} \frac{y_b^{phys}}{\sqrt{2}} & \frac{y_{bB}^{phys}}{\sqrt{2}} \\ \frac{y_{Bb}^{phys}}{\sqrt{2}} & \frac{y_B^{phys}}{\sqrt{2}} \end{pmatrix} h \ . \tag{2.46}$$

Since the mass matrices in the doublet model are very similar to the one analyzed in the singlet case, one can infer that the behavior of the various physical quantities as functions of the underlying parameters are as already presented in the figures and discussion of the previous subsection, and will not be repeated here.

Important differences between the singlet and doublet models involve the bottom quark. For example, in the doublet model the W gauge boson acquires a tree-level coupling to the RH top and bottom, as follows:

$$\Delta\mathcal{L} = \frac{g_w}{\sqrt{2}}\left[(\bar{T}\cos\theta_R^t - \bar{t}\sin\theta_R^t)\gamma^\mu P_R(B\cos\theta_R^b - b\sin\theta_R^b)\right]W_\mu + h.c.\ , \tag{2.47}$$

while in the singlet model the tree-level couplings of the W are modified compared to the SM only due to the mixing in the (LH) top sector (see Eq. (2.27)). Also, due to mixing in the bottom sector, the Z decays into bottom pairs, as well as the dominant decay of the Higgs into $b\bar{b}$ lead to constraints in the doublet model, while the singlet model does not present deviations from the SM in those observables. In connection to the Higgs, it is worth to emphasize again that neither model deviates significantly from the SM prediction in the loop-induced couplings to gluons and photons, as explained in Section 2.1.1 (see also Eqs. (2.12) and (2.32), which exhibit the mass matrices relevant for the argument). Table 3 summarizes the differences between the doublet and singlet models.

Singlet Model	Doublet Model
Only the top sector is modified.	The top and bottom sectors are modified.
The interaction of the W with the left-handed top is modified.	The interaction of the W with the left handed top is not modified.
There is no interaction of the W with the right handed top (as in the SM).	There is an interaction of the W with the right handed top.
The new parameters are y and M.	The new parameters are y_1, y_2 and M.

Table 3: A summary of the differences between the vector-like singlet and doublet models.

2.3 Experimental Constraints for the VLF Singlet and Doublet Models

Experimental measurements from LEP/SLC, ATLAS and CMS are used to constrain the parameter space of the VLF singlet and doublet models. The top width measurement [40] constrains both models, while the $Z \to b\bar{b}$ [41] and $\mathrm{H} \to \mathrm{b\bar{b}}$ [2] measurements constrain the bottom sector of the doublet model. The $\mathrm{t\bar{t}H}$ production measurements, as well as direct searches for new resonances will be discussed in the end of the chapter.

2.3.1 Top Quark Width

As already stated, the W couplings to the top and bottom quarks are modified. This affects the partial decay width of $t \to Wb$, the main decay mode for the top quark. In the singlet model, this coupling has the same structure of the SM with a suppression by the mixing angle in the LH top sector, as shown in Eq. (2.27). The partial decay width can be written in terms of the SM in a simple way as

$$\Gamma(t \to Wb)^{\text{Singlet}} \;=\; \cos^2\theta_L \times \Gamma(t \to Wb)^{\text{SM}} \; . \tag{2.48}$$

In the doublet model, the modification is more complicated because in addition to the SM couplings to the LH top and bottom quarks, there is a new coupling to the RH quarks, as shown in Eq. (2.47), while the interaction of the W with the left handed top quark is not modified when compared to the SM. There is an additional mixing in the bottom sector that is well constrained by the LEP measurements (next section). The expression for the tree level top decay width in the doublet model is not shown here because it is somewhat complicated and not particularly illuminating. One interesting feature of the new couplings to the RH top is that, when compared to the SM, it can result in an enhancement in the top decay width, while in the singlet case there is always a suppression (see Eq. (2.48)).

The total top width has been measured by the CMS collaboration [40]:

$$\Gamma_t = 1.36 \pm 0.02 \text{ (stat)}^{+0.14}_{-0.11} \text{ (syst) GeV} \; , \tag{2.49}$$

to be compared to the SM prediction of $\Gamma_t^{\text{SM}} \approx 1.3$ GeV. This is used in Section 2.4 to constrain the parameter space of both these models.

2.3.2 Z Pole Measurements

The Z pole precision measurements strongly constrain the couplings of the Z to the SM fermions. In particular, the Z$\to$ b$\bar{\text{b}}$ partial width is constrained by LEP and SLC [41]. The couplings of the Z to the bottom quark can be written as

$$\mathcal{L} \;=\; \frac{g_w}{c_w} Z_\mu \bar{b} \gamma^\mu (g_L^b P_L + g_R^b P_R) b \; , \tag{2.50}$$

where P_L and P_R are the chirality projectors and, in the SM (at tree-level):

$$(g_L^b)^{\rm SM} = -\frac{1}{2}+\frac{1}{3}s_w^2 \ , \tag{2.51}$$

$$(g_R^b)^{\rm SM} = \frac{1}{3}s_w^2 \ , \tag{2.52}$$

with $s_w^2 \approx 0.231$. The experimental collaborations at LEP and SLC measured

$$R_b = \frac{\Gamma(Z\to b\bar{b})}{\Gamma(Z\to \text{hadrons})} \approx \frac{(g_L^b)^2+(g_R^b)^2}{\sum_q[(g_L^q)^2+(g_R^q)^2]} \ , \tag{2.53}$$

as well as the forward-backward asymmetry of the bottom, which gives information on the asymmetry parameter

$$A_b = \frac{(g_L^b)^2-(g_R^b)^2}{(g_L^b)^2+(g_R^b)^2} \ . \tag{2.54}$$

A global fit to the Z-pole precision measurements gives [41]

$$g_L^b = -0.4182 \pm 0.0015 \ , \tag{2.55}$$

$$g_R^b = +0.0962 \pm 0.0063 \ , \tag{2.56}$$

with a strong correlation (see Fig. 7.4a of [41]). It is well-known that the SM prediction (including radiative corrections) is in some tension with these measurements [42], with the SM point lying close to the border of the 99.5% C.L. ellipse. [4] In order to extract numerical values from this ellipse, the ellipse plot was recreated by using the values of Table 7.9a of [41] and by constructing the ellipse equation (containing the correlation matrix). The curves achieved by doing this are close enough to the original ones, which are obtained from the fit to the measurements [43].

In the doublet model, at tree level, the expressions for these couplings are:

$$(g_L^b)^{\rm Doublet} = (g_L^b)^{\rm SM} \ , \tag{2.57}$$

$$(g_R^b)^{\rm Doublet} = (g_R^b)^{\rm SM}\cos^2\theta_R^b + (g_L^b)^{\rm SM}\sin^2\theta_R^b \ , \tag{2.58}$$

showing that only the RH coupling is modified with respect to the SM. By considering the ratio

$$\frac{(g_R^b)^{\rm Doublet}}{(g_R^b)^{\rm SM}} = \cos^2\theta_R^b + \frac{(g_L^b)^{\rm SM}}{(g_R^b)^{\rm SM}}\sin^2\theta_R^b \ , \tag{2.59}$$

[4] Fig. 7.4a of [41] shows the ellipse in the g_A^b-g_V^b plane, where $g_V^b = g_L^b + g_R^b$ and $g_A^b = g_L^b - g_R^b$. The central value is at $g_V^b = -0.322$ and $g_A^b = -0.5144$, while the SM prediction is $g_V^b = -0.34372^{+0.00041}_{-0.00020}$, $g_A^b = -0.49844^{+0.00013}_{-0.00010}$.

where $(g_L^b)^{\rm SM}/(g_R^b)^{\rm SM} \approx -5.44$ at tree level, it is observed that the RH coupling decreases when compared to the SM. Small θ_R^b, when $(g_R^b)^{\rm Doublet}$ is still positive, corresponds to a decrease in g_V^b (thus becoming more negative), and an increase in g_A^b (thus becoming less negative). Comparing to the ellipse in Fig. 7.4a of [41], it is derived a lower bound on the RH bottom mixing angle, $\cos\theta_R^b \geq 0.998$. It is also noted that for a sufficiently large RH mixing angle, θ_R^b, the sign of g_R^b can be flipped w.r.t. the SM. It has been pointed out that such a situation can lead to a better fit to the EW precision measurements [44], but this possibility is not pursued here.

2.3.3 Higgs decay width measurements

The production of $\rm t\bar{t}H$ before the Higgs decay is modulated only by the top Yukawa coupling. In this respect, both doublet and singlet models behave similarly. The Higgs decay, however, is different in both models since in the doublet case the bottom mixes with the vector-like bottom quark partner and the main decay channel of the Higgs ($\rm H \to b\bar{b}$) now depends on the parameters of the model (M and y_1).

In terms of the partial widths, the total Higgs width can be written as

$$\Gamma^{\rm total} = \Gamma_{gg} + \Gamma_{\gamma\gamma} + \Gamma_{b\bar{b}} + \Gamma_{c\bar{c}} + \Gamma_{\tau^+\tau^-} + \Gamma_{WW^*} + \Gamma_{ZZ^*} + \Gamma_{Z\gamma} + \Gamma_{\mu^+\mu^-}. \tag{2.60}$$

The branching fractions are then

$$\mathrm{BR}(H \to XX) = \Gamma_{XX}/\Gamma^{\rm total}, \tag{2.61}$$

where XX can be any of the particle pairs in Eq. (2.60). These widths can be compared to the experimental data in order to put bounds on the parameters of the new physics model. Deviations from the SM, on the models under study, depend on the mixing angles and affect mainly the bottom channel. The couplings of the Higgs to the massive vector bosons are not modified at tree level by the addition of vector-like quarks, while the effects on Γ_{gg} and $\Gamma_{\gamma\gamma}$ are small, as explained in Section 2.1.1. There are also no modifications in the couplings of the Higgs to the lighter fermions in the models under study. Therefore the focus here is on $\Gamma_{b\bar{b}}$, the dominant Higgs decay channel.

The partial width $\Gamma_{\rm H\to b\bar{b}}$ in the doublet model can be obtained from the SM one by the simple replacement (see Eq. (2.41)):

$$y_b^{\rm SM} \quad \to \quad y_b^{phys} = \cos\theta_R^b \left(y_b \cos\theta_L^b - y_1 \sin\theta_L^b\right) , \tag{2.62}$$

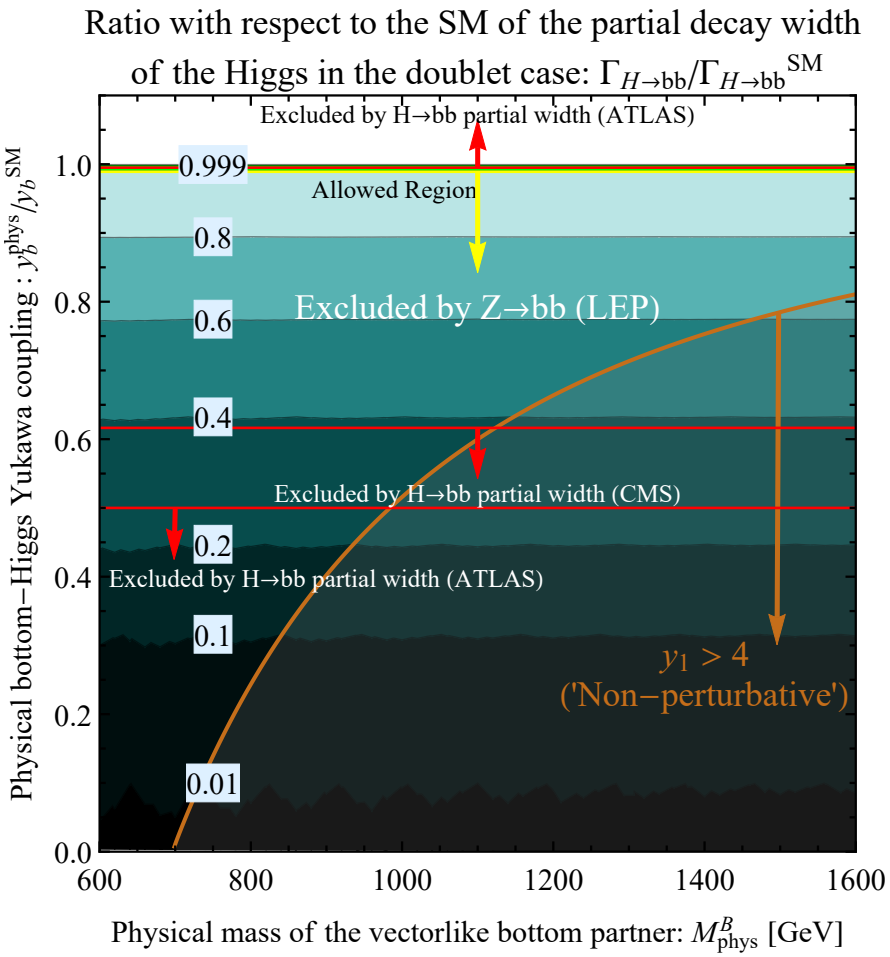


Figure 5: Contours for the partial decay width of the Higgs into bottom pairs, normalized to the SM, $\Gamma_{H\to b\bar{b}}/\Gamma^{SM}_{H\to b\bar{b}}$, as a function of M^B_{phys} and y_b^{phys}/y_b^{SM}. The red lines show the upper bound by ATLAS ($\sqrt{0.99}$) and lower bounds by ATLAS ($\sqrt{0.25}$) and CMS ($\sqrt{0.38}$) [2], after using Eq. (2.63). The upper bound by CMS is weaker than the ATLAS one and is not shown. The yellow line gives the limit as obtained from LEP/SLC [41], showing that the bottom-Higgs Yukawa coupling must be very close to the SM. The 'non-perturbative' region indicates when $y_1 > 4$, where the tree-level analysis has to be interpreted carefully.

so that

$$\Gamma^{\text{Doublet}}_{H\to b\bar{b}} = \left(\frac{y_b^{phys}}{y_b^{SM}}\right)^2 \Gamma^{SM}_{H\to b\bar{b}} \,. \tag{2.63}$$

In the singlet model, there is no deviation from the SM prediction.

Fig. 5 shows contours of $\Gamma^{\text{Doublet}}_{H\to b\bar{b}}$ (normalized to the SM), as a function of the physical mass of the heavy bottom partner in the doublet model, and the Yukawa coupling of the bottom quark to the Higgs boson (normalized to the SM). [5] As expected, the tree-level decay width of the Higgs into $b\bar{b}$ depends only on the physical bottom Yukawa coupling.

[5] Numerical details of the scan and numerical approach is described in Section 2.4.

Note that as the heavy fermion mass increases, to keep a fixed bottom Yukawa coupling one must increase the new coupling y_1, introduced in Eq. (2.28). If this coupling gets close to the non-perturbative value of about 4π, the tree-level treatment is expected to break down and the corresponding results are not reliable. The region of $y_1 = 4$ is thus indicated as a "semi-perturbative" boundary below which the tree-level treatment is expected to capture the physics correctly. For values above $y_1 = 4$, the figure must be interpreted with care.

The bounds on $y_b^{phys}/y_b^{\rm SM}$ are also indicated as inferred from the signal strengths of $H \to b\bar{b}$ [6] presented in Ref. [2]. These are $0.62^{+0.37}_{-0.37}$ (from ATLAS), and $0.81^{+0.45}_{-0.43}$ (from CMS), which exclude some of the regions of Fig. 5 (at the 95% confidence level) as shown by the red lines. The strongest constrain, however, comes from LEP/SLC, as described in the previous subsection, and is indicated by the yellow horizontal line. The upshot is that the bottom Yukawa coupling is constrained to be very close to the SM value, essentially fixing y_1 as a function of M^B_{phys}.

2.3.4 Experimental Searches for $t\bar{t}H$ at the LHC

Since the beginning of the LHC, the search for the $t\bar{t}H$ process is subject of very active analyses efforts. This process allows measuring the top Yukawa coupling, one of the main subjects of this thesis. It is within the LHC Run 2 period, with data collected in 2016, 2017 and 2018 that real progress is being achieved due to the increase in the center of mass energy, in luminosity, and with detector and trigger system upgrades as well as with further improvements in the overall data processing.

Both CMS and ATLAS collaborations report these results in references [45, 46]. For ATLAS, the combined Run 1 and Run 2 (containing 2016 and 2017 data) gives a signal strength of $1.32^{+0.28}_{-0.26}$. CMS, with the full Run 1 and Run 2 containing data collected in 2016, obtains a signal strength of $1.26^{+0.31}_{-0.26}$. Both experiments take into account the Higgs decays into $b\bar{b}$, $\gamma\gamma$, ZZ^*, WW^* and $\tau^+\tau^-$. The $t\bar{t}H$ measurements by ATLAS and CMS are currently given with a precision of about 20% and new results reducing this uncertainty are underway.

2.3.5 Experimental Searches for vector-like resonances

The lower bounds on the masses of the new vector-like fermions considered in this work come from direct searches by both the ATLAS and CMS collaborations [47, 48, 49]. These

[6] The signal strength is defined as $\mu = \frac{\mathrm{BR}(\mathrm{H}\to\mathrm{b\bar{b}})}{\mathrm{BR}_{\mathrm{SM}}(\mathrm{H}\to\mathrm{b\bar{b}})}$.

searches consider heavy top quark vector-like partner resonances decaying exclusively in the bW, tZ and tH channels. The bounds depend mildly on the decay branching fractions of the heavy state and are roughly around 1.3 TeV. These experimental results allow constraining the parameter space of theoretical Models including vector-like quarks.

2.4 Numerical Analysis

This section describes the numerical analysis implementation and the results for the two models described in this chapter.

2.4.1 Parameter Space scan

The description of the strategy of scanning over the parameter space of the models is presented focusing on the singlet model. The doublet models is treated in an analogous way with the same strategy and tools.

Once experimental inputs such as the top quark mass are taken into account, the singlet model is defined by two parameters: a Yukawa coupling y and a vector-like mass parameter M. In Section 2.2.1, the explicit analytic expressions for various physical quantities are provided, such as the physical heavy top partner mass M^T_{phys} or the strength of its coupling to the Higgs boson y_T^{phys}, etc. as a function of these two parameters. From these, other observables of interest can be computed, as described in the following Subsection 2.4.3. The observables are presented as a function of the physical parameters rather than the Lagrangian input parameters above.

For technical reasons in generating the plots, it would be useful to have a regular grid over the pairs (y_T^{phys}, M^T_{phys}) rather than (y, M). While the analytic formulas allow inverting the functional dependence, the relation is rather non-linear due to the presence of various trigonometric functions. An intermediate step found useful is to exchange M for M^T_{phys}, and work with the pairs (y, M^T_{phys}). This can be done by using Eq. (2.20) together with the expressions for the mixing angles in Eqs. (2.16)-(2.19), taking into account also Eq. (2.15) for y_t. Then, given y, M^T_{phys} and m_t, the required value for M can be found numerically. However, not all values of y allow for a given spectrum (m_t, M^T_{phys}): only values below

$$y_{\text{limit}} = \frac{\sqrt{2}}{v}\sqrt{(M_{phys} - m_t)^2}\ , \tag{2.64}$$

which defines the branch point in Eq. (2.15), are physically acceptable. This simple and explicit relation is used to define a physical boundary in the presented plots.

Given a pair (y, M^T_{phys}), a direct numerical diagonalization of the appropriate fermion mass matrices, is used to check that the desired spectrum is obtained. Such a numerical diagonalization is also the most straightforward way to choose the mixing angles in an appropriate quadrant: it is required that both eigenvalues are positive, and also that the $(1,1)$ entry of the diagonal mass matrix correspond to the smallest eigenvalue. Thus, for each such point the LH and RH diagonalization matrices are determined fully, which is then used to compute numerically other quantities, such as the physical Yukawa couplings y_t^{phys}, y_T^{phys}, y_{Tt}^{phys} and y_{tT}^{phys}. All such data are generated over a grid of M^T_{phys} ranging from 600 to 1600 GeV and y ranging from 0 to 9, excluding points that do not satisfy Eq. (2.64).

2.4.2 Monte Carlo Simulations

While most of the results presented in this chapter do not require Monte Carlo simulation, both models were implemented in FeynRules(v2.3) [50], a Mathematica [51] package used to implement models at the Lagrangian level, and UFO files that can be used for simulations in MadGraph 5 (MG5) (v2.6.2) [39] were generated. This allows to cross check the validity of this analysis framework using this MadGraph based overall simulation chain performed at tree level. In the next chapter, the results rely mostly on this simulation framework for studying Composite Higgs models.

2.4.3 Results of this overall phenomenology study

This chapter focuses on deviations from the SM in the $t\bar{t}H$ process. As previously explained, in the two considered models the modification in the total cross section comes from the modification of the Yukawa coupling, at least at leading order. The total cross section is given by Eq. (2.26) in the singlet model, with an analogous expression for the doublet model. The cross section is presented normalized to the SM prediction. This allows taking easily into account the state of the art radiative corrections that enter through $\sigma^{ttH}_{\rm SM}$, as well as any collider related effects. The tree-level results obtained are, however, also checked against Monte Carlo simulation of the models.

It is summarized in Section 2.3 how other measurements can constrain the parameter space of the models. These fall into two categories: those that affect the top sector directly, and therefore relevant to both the singlet and doublet models (this is the top decay width), and those that enter through the bottom sector and therefore affect models with bottom-like partners such as the doublet model, but not the singlet model. The first category is particularly interesting since it gives a rather direct correlation between top

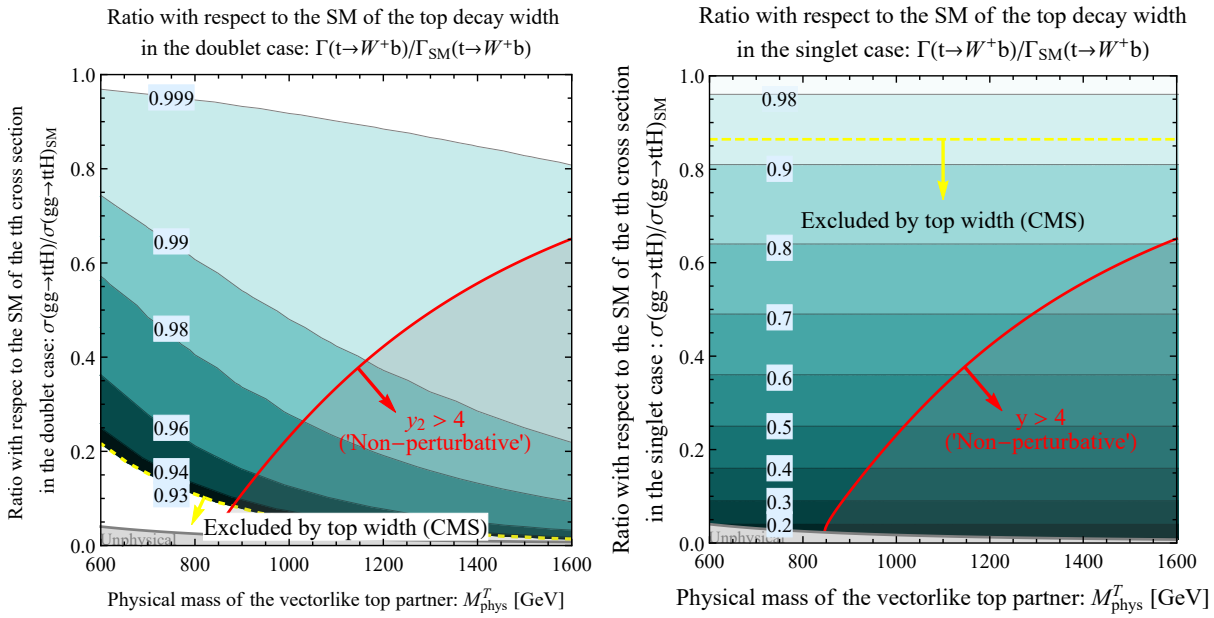


Figure 6: The top width normalized to the SM as a function of the physical vector-like top partner mass, M^T_{phys} (in GeV), and the $t\bar{t}H$ production cross-section via gluon fusion, for (a) the doublet model and (b) the singlet model. The region below the dashed yellow line is excluded by the top width measurement. In both plots, the tree-level analysis is not reliable in the 'non-perturbative' region with $y > 4$. The small gray areas marked as "unphysical" correspond to regions where the top mass of 150.7 GeV cannot be reproduced.

decay width and the $t\bar{t}H$ production cross section. The relation is particularly transparent in the singlet model, and relates the deviations in the top Yukawa coupling from its SM value to the LH mixing angle [see Eqs. (2.26) and (2.48)]. In the doublet model the relation is more complicated but it also gives a constraint that relates the top quark Yukawa coupling to the mixing angles in the top sector.

In Figs. 6a and 6b such a correlation is shown by plotting contours of constant top decay width (normalized to the SM) as a function of the total $t\bar{t}H$ cross section (normalized to the SM) and of the physical new fermion mass M^T_{phys}. Fig. 6a corresponds to the doublet model and Fig. 6b to the singlet model. The region below the dashed yellow line, corresponding to the CMS constraint of Eq. (2.49), is excluded. A "non-perturbative" region is also exhibited and has to be interpreted with care, as already explained in Subsection 2.3.3. In both models a suppression in the $t\bar{t}H$ production cross-section as compared to the SM is expected, and the suppression can be quite significant in the doublet model. The minimum $\sigma^{t\bar{t}H}_{\text{doublet}} \sim \sigma^{t\bar{t}H}_{\text{SM}}/10$, consistent with perturbativity and the top width measurement, occurs at $M^T_{phys} \approx 900$ GeV. The singlet model can present a suppression of about 14% of the SM expectation.

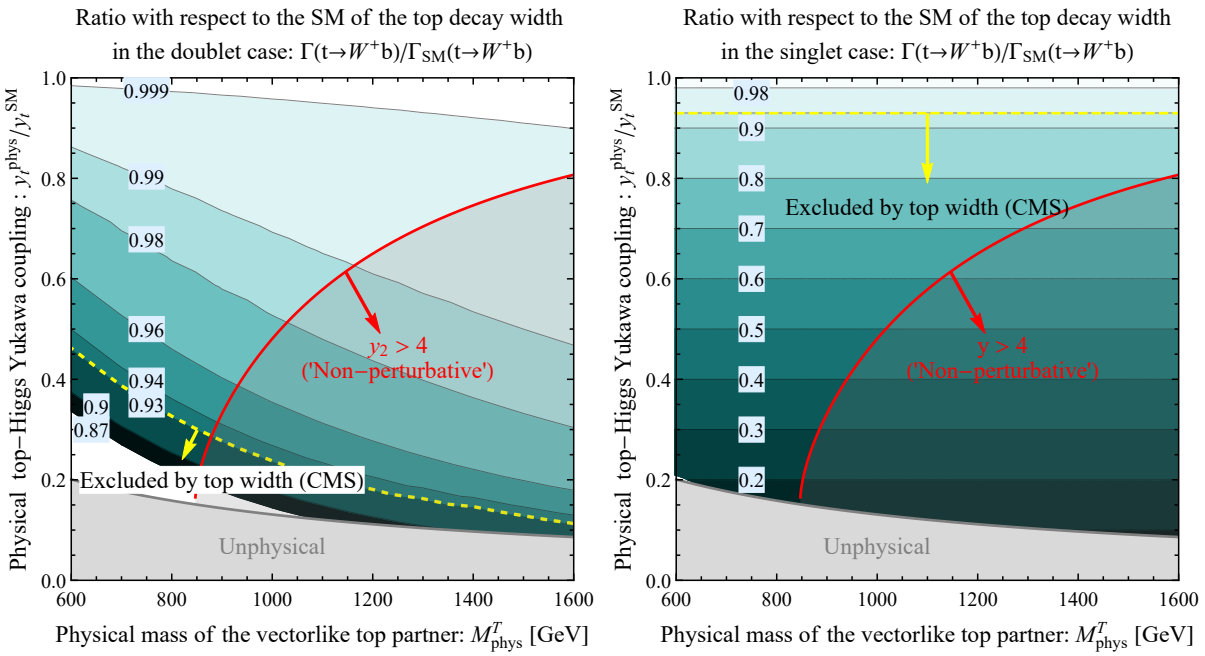


Figure 7: The top width normalized by the SM as a function of the physical vector-like top partner mass, M^T_{phys} (in GeV) and the physical top-Higgs Yukawa coupling y_t^{phys}, for (a) the doublet model and (b) the singlet model. The region below the dashed yellow line is excluded by the top width measurement. In both plots, the tree-level analysis is not reliable in the 'non-perturbative' region with $y > 4$. The small gray areas marked as "unphysical" correspond to regions where the top mass of 150.7 GeV cannot be reproduced.

Note that in the singlet model, the normalized top decay width ($= \cos^2\theta_L$) turns out to be independent of the top partner mass. This surprising result is also checked analytically. Note also that this is not in contradiction with a proper decoupling behavior, since to keep $\cos\theta_L$ fixed, as M^T_{phys} is increased, one must increase y. Keeping y fixed would result in the expected decoupling as the heavy top partner mass tends to infinity, i.e. $\cos\theta_L \to 1$. In contrast, for the doublet model there is a non-trivial dependence of the top width on the top partner mass.

An interpretation can also be given directly in terms of the physical top Yukawa coupling in Fig. 7, which otherwise contains the same information as Figs. 6a and 6b. For the singlet case, the top-Higgs Yukawa coupling ranges from the SM value down to 93% of the SM value. For the doublet case the top-Higgs Yukawa coupling ranges from the SM value down to 30% of the SM value.

The $t\bar{t}H$ measurements (Section 2.3.4) together with the search for top partners (Section 2.3.5) make the parameter space of the Singlet and Doublet models very constrained and thus makes them less interesting to pursue. It is important to remark that current

results to be published soon with new data collected at the LHC are going in the direction of constraining even more these models. The results of Fig. 6 and Fig. 7 show the constraints of these models in light of the experimental results. It is thus more interesting to study more complete extensions of the simplified models presented here and that are richer in phenomenology. Such models are considered in the next chapter.

Conclusions of this chapter remain true and help in understanding more complicated expressions or numerical analyses to be presented later in this Thesis. The calculations done analytically with the simple Single and Doublet models are less apparent in the analyses to come, but have a great didactic value in understanding details of these calculations, in getting trained to use these tools and in being introduced to the experimental world, subject of the other part of this analysis. [7]

2.5 Chapter Conclusions and Remarks

The focus here is on the $t\bar{t}H$ signal at leading order (which is dominated by QCD production). Both SM extensions studied in this chapter present a decrease in the total $\mu_{t\bar{t}H}$. The maximum possible suppression is constrained mainly by the existing measurement of the top quark width, and can reach $\mu_{t\bar{t}H}^{\text{doublet}} \sim 1/10$, and $\mu_{t\bar{t}H}^{\text{singlet}} \sim 0.86$, if not taking the current and to come $t\bar{t}H$ measurements into account. When $t\bar{t}H$ measurements are taken into account, the constraints on the doublet model look similar than the ones on the singlet model, with a deficit of about 20% (see Section 2.3.4). This makes the prediction of the two models rather similar. Possible extensions of the bottom sector, as illustrated by the doublet model, are also significantly constrained (mostly by precision measurements such as the $Z \to b\bar{b}$ partial decay width, although increasingly by Higgs measurements), but the corresponding region of parameter space is in general orthogonal to the top sector. The observation of $t\bar{t}H$ at the LHC is on the way to put stronger constraints on these two models.

As it was emphasized, at leading order the heavy top partners do not enter explicitly in the diagrams relevant for $t\bar{t}H$, except indirectly through mixing angles. This leads to the result of the only effect being $\mu_{t\bar{t}H} < 1$ due to a suppression of the top Yukawa coupling. Subdominant contributions can potentially be directly sensitive to the heavy physics, and other processes can be also affected. It is of interest to evaluate the feasibility of accessing other handles at the LHC and beyond. In this work, the measurement of $t\bar{t}HH$ is explored and is the main subject of the remaining chapters.

[7] This phenomenological study is reported in the CMS Internal Analysis Note[3] and was presented as such in internal CMS WG meetings. The basic ingredients of this model were introduced in the CMS generator framework.

One possible direction to take, not explored in this thesis, in the context of studying vector-like quarks, is the search for the top quark partners directly. This has been considered both in phenomenological studies [19, 20, 21] and by the ATLAS [22] and CMS [23] collaborations. It is well known that such fermionic resonances can be pair-produced (by QCD), or produced singly (through the weak interactions), and that the relative importance of the two production processes changes as a function of the heavy fermion mass. At low mass, pair production dominates, while at higher masses single production becomes more important. The work reported here and in the next chapters does not focus on the study of resonances; it instead addresses other measurements of interest for the study of the Top Higgs Yukawa domain.

In particular, the decay of the top partner into Ht suggests a contribution in an interesting manner to processes involving the top quark and the Higgs, even away from the resonance itself. This depends on the width of the resonance. In the simple cases considered here the resonance is relatively narrow, but it is possible to imagine the existence of new physics decay channels that could make it broader. Assessing whether the relevant processes can give a sufficiently large effect, and then developing strategies to extract such information from the LHC measurements will require more detailed simulation. Processes that, although likely challenging, are of interest include not only $t\bar{t}H$ itself, but also the production of Higgs pairs in association with top quarks, $t\bar{t}HH$. Such effects present interesting features in the distributions of appropriate kinematic variables, and are considered in detail in Chapter 3 in Composite Higgs models and also in Chapter 4, where a first search of this process is done in the CMS experiment at the LHC.

Chapter 3: Impact of Composite Higgs Models on the $t\bar{t}H$ and $t\bar{t}HH$ production processes at hadron colliders

The study of BSM physics involving vector-like top partners, as discussed in the previous chapter, is motivated by their presence in many proposed extensions of the SM. One such idea is Higgs compositeness models, in which the Higgs is introduced as a composite particle, and brings additional modifications apart from the ones coming from extra fermionic resonances, as discussed in Section 2.1. It is therefore of interest to consider the presence of both effects and see how each of them, whether separately or together, alter the SM predictions.

The absence of new physics near the Higgs mass scale has made popular the notion that the Higgs is not only composite but also a pseudo-Nambu Goldstone boson (pNGB) arising from the breaking of a symmetry at scales higher than the electroweak scale. Many models have been proposed in the literature, and the ones explored here are based on the $SO(5) \to SO(4)$ symmetry breaking, called Minimal Composite Higgs Models (MCHMs). There are many possibilities for the embedding of the fermions on the $SO(5)$ representation, and here two cases are considered: the MCHM_5, where the fermions are on the **5** of $SO(5)$, and the MCHM_{14}, where the fermions are on the symmetric **14** of $SO(5)$. These modifications will specially affect the top sector, since the top quark has the strongest interaction with the Higgs boson, as in the models discussed in Chapter 2.

In contrast to the models of Chapter 2, in MCHMs the gluon-gluon-Higgs coupling is generally suppressed with respect to the SM [34, 52]. The Higgs branching fractions, on the other hand, can be increased or decreased depending on the embeddings and parameters of the model. The top-top-Higgs coupling can also be increased, compensating or even offsetting the suppression shown in the previous chapter, deserving thus careful consideration.

The focus here is on the production of $t\bar{t}H$ and $t\bar{t}HH$. As already stated in Section 2.3.4, the current measurements of the $t\bar{t}H$ cross section is consistent with the SM expectation, within uncertainties of the order of 20% [45, 46]. The $t\bar{t}HH$ process has not been observed, and the first search for this process is presented in this Thesis (Chapter 4). Such a

process is of particular interest for the MCHMs discussed here, since new heavy vector-like fermions of charge 2/3 can be produced, and decay in the tH channel, thus a pair production of such resonances can lead to the $t\bar{t}HH$ production. Resonance searches focusing on this decay channel have been presented in [47], and the constraints in such vector-like resonances coming from the combined searches considering the bW, tZ and tH channels are taken into account [48, 49].

If resonances are heavy, the effects of the new fermions can be reduced. It is of interest to study the non-resonant part of the $t\bar{t}HH$ process, which is defined as the part of the $t\bar{t}HH$ production that does not come from the pair production of heavy resonances. Due to current experimental bounds, this non-resonant production is already accounting for a large fraction of the total cross section. The non-resonant $t\bar{t}HH$ process is closely connected to the $t\bar{t}H$ process, as both are sensitive to the top Yukawa coupling, but carries additional information coming from the triple Higgs interaction diagrams.

This chapter presents a first step in the analysis of such processes in the context of the 14 TeV High Luminosity phase of the LHC (HL-LHC), and of future high energy hadron colliders, such as a possible High Energy LHC upgrade at 27 TeV (HE-LHC) [53]. Other future machines, in the range of 100 and 150 TeV as in the FCC-hh project at CERN and the SppC project in China [54, 55, 56] are the subject of a paper in preparation [57]. Monte Carlo simulations at different center of mass energies are considered for studying the phenomenology of the models, but this Thesis does not consider details of new detectors.

Section 3.1 introduces the theoretical aspects of the MCHM construction and defines two models differing in the fermion implementation. The first scenario can be considered a minimal embedding of fermions, while the second one displays features that are distinct from the minimal case. In Section 3.1.3, it is explained how the rest of the SM sector is treated, as this is important when taking into account the Higgs decays in event generator programs. The analysis strategy is described in Section 3.2, where it is also introduced details of the $t\bar{t}H$ and $t\bar{t}HH$ production processes in the MCHM models. Results are presented in Section 3.3, while Section 3.5 contains conclusions and final remarks for this chapter. Some technical theoretical details are collected in appendices.

Most of the analyses in this chapter will be included in a paper in preparation [57] with extended work, to be submitted soon. Other results here included, specifically the 27 TeV simulations, are taken from the work published in Section 9.7 of Ref. [1], where the focus is on the HL-LHC as well as the HE-LHC.

3.1 Theoretical introduction to Composite Higgs Models

Relevant aspects of the Composite Higgs models that are used in this thesis are presented here. In particular, the spin-1 sector is not described, which is fixed by the pattern of symmetry breaking that here is $SO(5) \to SO(4)$. For the general construction and results for the MCHM bosonic sector see the complete review of Ref. [58]. One of the key points is that, in the unitary gauge, the electroweak symmetry can be parameterized as the following orthogonal matrix:

$$U = \begin{pmatrix} \mathbb{1}_{3\times3} & \vec{0} & \vec{0} \\ \vec{0}^T & \cos\frac{h_0+h}{f} & \sin\frac{h_0+h}{f} \\ \vec{0}^T & -\sin\frac{h_0+h}{f} & \cos\frac{h_0+h}{f}, \end{pmatrix} , \tag{3.1}$$

where h is the Higgs boson, $h_0 = \langle h \rangle$ is its expectation value, and f is the scale at which the breaking $SO(5) \to SO(4)$ occurs. In Eq. (3.1), $\mathbb{1}_{3\times3}$ is the 3×3 identity matrix and $\vec{0}$ is a 3-dimensional null vector. The mass of the W boson in this model is

$$M_W^2 = \frac{1}{4} g^2 f^2 \sin^2 \frac{h_0}{f} \equiv \frac{1}{4} g^2 v^2 \ , \tag{3.2}$$

where $v = 246$ GeV is the electroweak vev. An useful variable can be defined to characterize the deviations from a SM Higgs due to compositeness:

$$\xi = \frac{v^2}{f^2} = \sin^2 \frac{h_0}{f}. \tag{3.3}$$

Consistency with current Higgs measurements results in $\xi \lesssim 0.1$, or $f \gtrsim 800$ GeV [34, 52, 59, 60, 61, 62, 63].

The processes of interest here, i.e. $t\bar{t}H$ and $t\bar{t}HH$, depend on the Higgs compositeness and on the fermionic resonances that are introduced, and the way to introduce them is model dependent. They are introduced in representations of $SO(4)$, and in this Thesis, two representations are discussed. The first one is when the fermions are in the **5** representation of $SO(5)$, which is split into $SO(4)$ multiplets as $\mathbf{5} = \mathbf{4} + \mathbf{1}$. The second model considers fermions embedded in the symmetric **14** of $SO(5)$, which is split into $SO(4)$ multiplets as $\mathbf{14} = \mathbf{9} + \mathbf{4} + \mathbf{1}$. Thus there are 5 new resonances in the first model, denoted MCHM_5 and an extra nine resonances, leading to 14 new particles on the second model, denoted MCHM_{14}. The first case (MCHM_5) contains the minimal number of new fermions, while still respecting custodial symmetry [64], which is important for EW precision measurements [65, 66, 67, 68, 69, 70, 71] The second model has been considered in [34, 52, 72, 73, 74, 75, 76, 77]. It has the interesting property of being able to enhance

the Higgs couplings with respect to the SM[52, 74], instead of the obligatory decrease already observed in Chapter 2 that also applies to the MCHM$_5$ as we shall see. This feature will be explored in this chapter, with a goal to use the production of $t\bar{t}H$ and $t\bar{t}HH$ to differentiate these two models, in the context of the HL-LHC and the HE-LHC.

3.1.1 The Fermion Sector of the MCHM$_5$

As in Chapter 2, the focus is on the top sector. Here the notation is slightly different and is defined along the text. The top sector consists of the elementary (SM-like) quarks $q_L = (t_L, b_L)$ and t_R, as well as the new fermionic resonances being introduced. The elementary fields have the $SU(3)_C \times SU(2)_L \times U(1)_Y$ quantum numbers of the SM LH top-bottom $SU(2)_L$ doublet and of the SM RH top $SU(2)_L$ singlet, respectively. The new vector-like fermions, for the case of MCHM$_5$, are introduced in a **5** of $SO(5)$, which is split into a $SO(4)$ 4-plet, Ψ_4, and a $SO(4)$ singlet, Ψ_1. These states are defined as:

$$\begin{aligned} \Psi_4 &\sim (X_{5/3}, X_{2/3}, T, B) \ , \\ \Psi_1 &\sim \tilde{T} \ , \end{aligned} \tag{3.4}$$

where the subindex is the electric charge, and the states T and $\tilde{T}$ have charge $2/3$ and B has charge $-1/3$. The states $(X_{5/3}, X_{2/3})$ transform as a $SU(2)_L$ doublet with exotic hypercharge $Y = 7/6$. The states (T, B) also transform as a $SU(2)_L$ doublet and have hypercharge $Y = 1/6$, the same as the SM elementary fields, and the singlet $\tilde{T}$ has the SM quantum numbers of t_R. This is similar to the vector-like Singlet and Doublet models of Chapter 2, which are now combined, already pointing out a richer phenomenology. Note here a higher multiplicity of vector-like doublets, and also a new exotic doublet containing another charge $2/3$ particle that mixes with the top quark. Although more complicated than before, the effects of the mixing of the vector-like quarks with top and bottom quarks are very similar to Chapter 2, as detailed below, bringing modifications on the top Yukawa. New effects of course arise from Higgs compositeness and are controlled by the scale f.

The Lagrangian for this model is:

$$\mathcal{L} = \mathcal{L}_{\text{elem}} + \mathcal{L}^{\mathbf{5}}_{\text{comp}} + \mathcal{L}^{\mathbf{5}}_{\text{mix}} + \mathcal{L}^{\mathbf{5}}_{\text{int}} \ , \tag{3.5}$$

where $\mathcal{L}_{\text{elem}}$ contains interactions between elementary fermions, $\mathcal{L}^{\mathbf{5}}_{\text{comp}}$ contains the interactions between the new composite fermions, and the term $\mathcal{L}^{\mathbf{5}}_{\text{mix}}$ accounts for mixing

terms involving elementary-composite interactions. $\mathcal{L}^{\mathbf{5}}_{\text{int}}$ accounts for additional interactions involving the d_μ symbol (detailed in Appendix C).

The elementary Lagrangian is simply described by:

$$\mathcal{L}_{\text{elem}} = \bar{q}_L i \not{D} q_L + \bar{t}_R i \not{D} t_R \ , \tag{3.6}$$

where D stands for the $SU(3)_C \times SU(2)_L \times U(1)_Y$ covariant derivative.

The Lagrangian that describes the interactions between the new fermions can be written in terms of the $SO(4)$ multiplets:

$$\mathcal{L}^{\mathbf{5}}_{\text{comp}} = \bar{\Psi}_{\mathbf{4}} i(\not{D} - i\not{e})\Psi_{\mathbf{4}} - M_4 \bar{\Psi}_{\mathbf{4}}\Psi_{\mathbf{4}} + \bar{\Psi}_{\mathbf{1}} i \not{D} \Psi_{\mathbf{1}} - M_1 \bar{\Psi}_{\mathbf{1}}\Psi_{\mathbf{1}} \ , \tag{3.7}$$

where e_μ symbol is defined as a function of the matrix of Eq.(3.1) and $D_\mu = \partial_\mu + i g_s G_\mu + i\frac{2}{3} g' B_\mu$. The remaining electroweak interactions are inside the d_μ and e_μ symbols:

$$iU^{-1}(\partial_\mu + i g_a A^a_\mu T^a)U = d_{\mu,\hat{a}} T^{\hat{a}} + e_{\mu,a} T^a \equiv d_\mu + e_\mu \ , \tag{3.8}$$

where the generators of the unbroken and broken $SO(4)$ are, respectively, T^a and $T^{\hat{a}}$. The gauge fields are introduced by gauging the $SU(2)_L \times U(1)_Y$ subgroup of $SO(4)$, with the W fields gauging $SU(2)_L$ while B_μ gauges T^3_R. The remaining generators are ungauged. The hypercharge is then given by $Y = 2/3 + T^3_R$, where the $2/3$ arises from the introduction of an extra $U(1)_X$ factor, under which $\Psi_{\mathbf{5}}$ has charge $X = 2/3$, and is needed to reproduce the correct fermion hypercharges [58].

The $SO(5)$ symmetry is non-linearly realized in the kinetic terms by the introduction of this covariant derivative, even though the Lagrangian (3.7) exhibits explicitly only the $SO(4)$ symmetry [78, 79]. This breaks the $SO(5)$ symmetry explicitly (since only a part of the $SO(4)$ group is gauged. Appendix A lists the $SO(5)$ generators and Appendix C shows the form of the e_μ symbol, which contains the electroweak interactions. $\mathcal{L}^{\mathbf{5}}_{\text{comp}}$ also contains mass terms for the fermions $\Psi_{\mathbf{4}}$ and $\Psi_{\mathbf{1}}$, which are allowed because of their vector-like nature.

The Lagrangian $\mathcal{L}^{\mathbf{5}}_{\text{mix}}$ contains the terms that mix elementary and composite fermions. This can happen since they have the same quantum numbers, and it is introduced by the partial compositeness hypothesis [15]. In order to write this Lagrangian, it is necessary to first define how the embedding of the fermions is realized. For MCHM$_5$, they are embedded in the $\mathbf{5}$ of $SO(5)$, and the multiplets are defined as follows. The elementary

states are fiveplets of $SO(5)$:

$$Q_L^{\mathbf{5}} = \frac{1}{\sqrt{2}}\begin{bmatrix} -ib_L \\ -b_L \\ -it_L \\ t_L \\ 0 \end{bmatrix}, \quad T_R^{\mathbf{5}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ t_R \end{bmatrix}. \tag{3.9}$$

The new resonances($\Psi_{\mathbf{4}}$ and $\Psi_{\mathbf{1}}$) are written as the following multiplets, also in a 5-plet notation (more details in Appendix B):

$$\Psi_{\mathbf{4}} = \frac{1}{\sqrt{2}}\begin{bmatrix} -iB + iX_{5/3} \\ -B - X_{5/3} \\ -iT - iX_{2/3} \\ T - X_{2/3} \\ 0 \end{bmatrix}, \quad \Psi_{\mathbf{1}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \tilde{T} \end{bmatrix}. \tag{3.10}$$

With the multiplets defined above, the mixing Lagrangian can be written as:

$$\begin{aligned} \mathcal{L}_{\text{mix}}^{\mathbf{5}} &= f\,\bar{Q}_L^{\mathbf{5}} U\left[y_{L4}\Psi_{\mathbf{4}} + y_{L1}\Psi_{\mathbf{1}}\right] + \text{h.c.} \\ &+ f\,\bar{T}_R^{\mathbf{5}} U\left[y_{R4}\Psi_{\mathbf{4}} + y_{R1}\Psi_{\mathbf{1}}\right] + \text{h.c.}\,. \end{aligned} \tag{3.11}$$

The additional interactions involving the d_μ symbol are allowed by the symmetries at the lowest order of derivatives [58] and the corresponding Lagrangian can be written as:

$$\mathcal{L}_{\text{int}}^{\mathbf{5}} = -i\,c_L\,\bar{\Psi}_{\mathbf{4}} P_L \not{d}\,\Psi_{\mathbf{1}} - i\,c_R\,\bar{\Psi}_{\mathbf{4}} P_R \not{d}\,\Psi_{\mathbf{1}} + \text{h.c.}, \tag{3.12}$$

where $P_{L,R} = (1 \mp \gamma_5)/2$ are the left and right projectors and c_L and c_R are couplings, expected to be of order one. These terms contain extra Higgs interactions (see Appendix C). The part of d_μ containing the Higgs is $\frac{\partial_\mu h}{f}$, and will lead to amplitudes proportional to the Higgs momentum over the f scale, and are thus suppressed in relation to the interactions of $\mathcal{L}_{\text{mix}}^{\mathbf{5}}$. This is expected to be a small correction and is out of the scope of this Thesis. It is studied in Ref.[57].

The charge 2/3 mass matrix in the $\{\bar{t}_L,\ \bar{T}_L,\ \bar{X}_{2/3,L},\ \bar{\tilde{T}}_L\}$ vs $\{t_R,\ T_R,\ X_{2/3,R},\ \tilde{T}_R\}$ basis is then given by

$$\mathcal{M}^{\mathbf{5}}_{2/3} = \begin{bmatrix} 0 & \frac{1}{2}y_{L4}f(1+\sqrt{1-\xi}) & \frac{1}{2}y_{L4}f(1-\sqrt{1-\xi}) & \frac{1}{\sqrt{2}}y_{L1}f\sqrt{\xi} \\ -\frac{1}{\sqrt{2}}y_{R4}f\sqrt{\xi} & -M_4 & 0 & 0 \\ \frac{1}{\sqrt{2}}y_{R4}f\sqrt{\xi} & 0 & -M_4 & 0 \\ y_{R1}f\sqrt{1-\xi} & 0 & 0 & -M_1 \end{bmatrix} . \quad (3.13)$$

This matrix plays the same role as the matrix of Eq.(2.12) and of Eq.(2.32) of the simple models of last chapter. It needs to be diagonalized for obtaining the physical fermion eigenstates. This leads to the mixing between the SM-like fermions and the new fermions and the lightest fermion is identified with the observed top quark. Instead of the analytical analysis of Chapter 2, all calculations are done numerically here and follow the same principles. As described in subsequent sections, most of the analysis follow from this matrix. The remaining resonances do not mix, and their masses follow directly from the Lagrangian:

$$X_{5/3}: \qquad M_{X_{5/3}} = M_4 \ , \quad (3.14)$$

$$B: \qquad M_B = \sqrt{M_4^2 + y_{L4}^2 f^2} \ . \quad (3.15)$$

3.1.2 The Fermion Sector of the MCHM$_{14}$

What was discussed for MCHM$_5$ also applies to the MCHM$_{14}$ model. The difference comes from the embedding of the fermions, which instead of being introduced in the **5** of $SO(5)$, are introduced in the **14** of $SO(5)$. Under $SO(4)$, in addition to a 4-plet and a singlet, a nonet $\Psi_{\mathbf{9}}$ is added as:

$$\Psi_{\mathbf{9}} \ \sim \ (U_{8/3}, U_{5/3}, U_{2/3}, Y_{5/3}, Y_{2/3}, Y_{-1/3}, Z_{2/3}, Z_{-1/3}, Z_{-4/3}) \ . \quad (3.16)$$

The fermions U, Y and Z transform as $SU(2)_L$ triplets with hypercharges $Y = 5/3$, $2/3$ and $-1/3$, respectively.

The Lagrangian that involves interactions between the new composite fermions is just the sum of Eq. (3.7) to new terms involving Ψ_9 (see Appendix B):

$$\mathcal{L}^{\mathbf{14}}_{\rm comp} \ = \ \mathcal{L}^{\mathbf{5}}_{\rm comp} + {\rm Tr}\left[\bar{\Psi}_{\mathbf{9}}\left(i\not{D}\Psi_{\mathbf{9}} - i\left[\not{e}, \Psi_{\mathbf{9}}\right]\right)\right] - M_9 {\rm Tr}\left[\bar{\Psi}_{\mathbf{9}}\Psi_{\mathbf{9}}\right] \ . \quad (3.17)$$

In analogy to the MCHM$_5$ case, instead of 5-plets, the elementary fermions are embedded as $Q^{\mathbf{14}}_L$ and $T^{\mathbf{14}}_R$, while the $\Psi_{\mathbf{9}}$, $\Psi_{\mathbf{4}}$ and $\Psi_{\mathbf{1}}$ notation is kept for the composite states (see

Appendix B). The mixing Lagrangian can then be written as

$$\begin{aligned}\mathcal{L}^{\mathbf{14}}_{\rm mix} &= f\,{\rm Tr}\left[U^{\mathsf{T}}\bar{Q}^{\mathbf{14}}_L U\left(y_{L9}\Psi_{\mathbf{9}}+y_{L4}\Psi_{\mathbf{4}}+y_{L1}\Psi_{\mathbf{1}}\right)\right]+{\rm h.c.}\\ &+ f\,{\rm Tr}\left[U^{\mathsf{T}}\bar{T}^{\mathbf{14}}_R U\left(y_{R9}\Psi_{\mathbf{9}}+y_{R4}\Psi_{\mathbf{4}}+y_{R1}\Psi_{\mathbf{1}}\right)\right]+{\rm h.c.}\ .\end{aligned} \tag{3.18}$$

The complete Lagrangian is then given by:

$$\mathcal{L}=\mathcal{L}_{\rm elem}+\mathcal{L}^{\mathbf{14}}_{\rm comp}+\mathcal{L}^{\mathbf{14}}_{\rm mix}+\mathcal{L}^{\mathbf{14}}_{\rm int}\ , \tag{3.19}$$

where the term $\mathcal{L}^{\mathbf{14}}_{\rm int}$ contains d_μ interactions, with the complete forms reported in Appendix C, and are not included in the work of this thesis for the same reason that it is not included for MCHM$_5$.

This model than has a charge 2/3 mass matrix in the $\{\bar{t}_L,\,\bar{T}_L,\,\bar{X}_{2/3,L},\,\bar{\tilde{T}}_L,\,\bar{U}_{2/3,L},\,\bar{Y}_{2/3,L},\,\bar{Z}_{2/3,L}\}$ vs $\{t_R,\,T_R,\,X_{2/3,R},\,\tilde{T}_R,\,U_{2/3,R},\,Y_{2/3,R},\,Z_{2/3,R}\}$ basis, that is given by:

$$\mathcal{M}^{\mathbf{14}}_{2/3}=\begin{bmatrix}
0 & \frac{1}{2}y_{L4}fa_+ & -\frac{1}{2}y_{L4}fa_- & -\frac{\sqrt{5}}{4}y_{L1}fs_{2h} & -\frac{1}{2}y_{L9}fb_- & -\frac{1}{2}y_{L9}fs_{2h} & \frac{1}{4}y_{L9}fb_+\\
\frac{\sqrt{5}}{4}y_{R4}fs_{2h} & -M_4 & 0 & 0 & 0 & 0 & 0\\
-\frac{\sqrt{5}}{4}y_{R4}fs_{2h} & 0 & -M_4 & 0 & 0 & 0 & 0\\
y_{R1}f\left(1-\frac{5}{4}s_h^2\right) & 0 & 0 & -M_1 & 0 & 0 & 0\\
\frac{\sqrt{5}}{4}y_{R9}fs_h^2 & 0 & 0 & 0 & -M_9 & 0 & 0\\
-\frac{\sqrt{5}}{4}y_{R9}fs_h^2 & 0 & 0 & 0 & 0 & -M_9 & 0\\
\frac{\sqrt{5}}{4}y_{R9}fs_h^2 & 0 & 0 & 0 & 0 & 0 & -M_9
\end{bmatrix}, \tag{3.20}$$

where:

$$s_h^2=\xi\ ,\quad s_{2h}=2\sqrt{\xi}\sqrt{1-\xi}\ ,\quad a_\pm=1\pm\sqrt{1-\xi}-2\xi\ ,\quad b_\pm=\sqrt{\xi}\left(1\pm\sqrt{1-\xi}\right)\ .$$

The charge $-1/3$ mass matrix in the $\{\bar{b}_L,\,\bar{B}_L,\,\bar{Y}_{-1/3,L},\,\bar{Z}_{-1/3,L}\}$ vs $\{B_R,\,Y_{-1/3,R},\,Z_{-1/3,R}\}$ basis is:

$$\mathcal{M}^{\mathbf{14}}_{-1/3}=\begin{bmatrix}
y_{L4}f\sqrt{1-\xi} & -\frac{1}{\sqrt{2}}y_{L9}f\sqrt{\xi} & \frac{1}{\sqrt{2}}y_{L9}f\sqrt{\xi}\\
-M_4 & 0 & 0\\
0 & -M_9 & 0\\
0 & 0 & -M_9
\end{bmatrix}. \tag{3.21}$$

The remaining states have masses

$$X_{5/3}: \qquad M_{X_{5/3}}=M_4\ , \tag{3.22}$$

$$U_{8/3},U_{5/3},Y_{5/3},Z_{-4/3}: \qquad M_{U_{8/3}}=M_{U_{5/3}}=M_{Y_{5/3}}=M_{Z_{-4/3}}=M_9\ . \tag{3.23}$$

As in the **5**-plet case, the previous mass matrices form the fundamental input to the following phenomenological analysis. It is worth reminding again that the philosophy of the calculations follow closely the simpler diagonalization of mass matrices described in Chapter 2. Considerations on reproducing correctly the measured top quark mass are also taken here for both this MCHM$_{14}$ model, as well as the MCHM$_5$.

3.1.3 Higgs couplings to the gluons and light fermions

The composite Higgs Models, together with the partial compositeness hypothesis has an effect also on the coupling of the Higgs to gluons (ggH), which can be given by, in a way analogous to what was explained in Section 2.1.1:

$$c_g \;=\; \frac{v}{2}\frac{d}{dh_0}\log\det\left(\mathcal{M}^\dagger\mathcal{M}\right)\,, \tag{3.24}$$

where $\mathcal{M}$ is the fermion mass matrix, assuming all states are much heavier than the Higgs boson (in this work, light state contributions can be neglected). In the simple vector-like fermions case, $\det\left(\mathcal{M}^\dagger\mathcal{M}\right)$ is proportional to v^2, in such a way that the ggH amplitude did not differ from the SM one. For the models of this chapter, this is not the case.

For the MCHM$_5$, using $\mathcal{M}=\mathcal{M}^{\mathbf{5}}_{2/3}$ from Eq. (3.13) in (3.24) leads to:

$$c_g^{\mathbf{5}} \;=\; \frac{1-2\xi}{\sqrt{1-\xi}}\,. \tag{3.25}$$

For the MCHM$_{14}$, one gets

$$c_g^{\mathbf{14}} \;=\; c_g^t + c_g^b\,, \tag{3.26}$$

where

$$c_g^t \;=\; \frac{4\left(1-r_1\right)r_9-\left(9r_1+23r_9-32r_1r_9\right)\xi+4\left(3r_1+5r_9-8r_1r_9\right)\xi^2}{\sqrt{1-\xi}\left[4\left(1-r_1\right)r_9-\left(3r_1+5r_9-8r_1r_9\right)\xi\right]}\,, \tag{3.27}$$

arises from the charge 2/3 sector, $\mathcal{M}^{\mathbf{14}}_{2/3}$ in Eq. (3.20), and

$$c_g^b \;=\; \xi\sqrt{1-\xi}\,\frac{y_L^2f^2\left(1-r_9^2\right)}{r_9^2\left(M_4^2+y_L^2f^2\right)+y_L^2f^2\left(1-r_9^2\right)\xi}\,, \tag{3.28}$$

arises from the charge $-1/3$ sector, $\mathcal{M}^{\mathbf{14}}_{-1/3}$ in Eq. (3.21). For the latter, the physical bottom quark is explicitly removed and its contribution to the ggh coupling is neglected.

Global constraints on the gluon fusion process from Higgs measurements, which allow for about 20% deviations from unity in c_g at the 95% C.L. [2], are taken into account for the analysis performed in the next sections, leading to constraints on allowed deviations of the top Yukawa coupling from the SM prediction.

The embedding of light families of the SM in the MCHM has to be decided, but new effects are expected to be small, since the observed Higgs boson couples to the light fermions in a SM-like way. The dominant decay of the Higgs are thus as in the SM: $h \to b\bar{b}$, W^+W^-, gg, $\tau^+\tau^-$, $c\bar{c}$, ZZ. Other decays, which have branching fractions smaller than 1% can be neglected here.

The $b\bar{b}$ channel is the most important, since in the SM it has a branching ratio of about 60%. The left handed bottom quark is implemented in the MCHMs together with the left handed bottom quark. However, the implementation of the right handed bottom quark is yet to be specified. Out of several ways to proceed, it was chosen to embed the RH bottom quark in a **10** of $SO(5)$. These models, called $\text{MCHM}_{5,5,10}$ and $\text{MCHM}_{14,14,10}$, are introduced in [34]. The remaining lighter fermions are introduced the same way as the top and bottom quarks, assuming that the leptons follow the same quark implementation.

Under these assumptions, the couplings of the Higgs boson to vector bosons and light fermion pairs are controlled by model dependent functions that depend only on ξ. For MCHM_5, couplings of the Higgs to a gluon pair depend only on ξ, and in MCHM_{14}, it depends also on $c_g^{\mathbf{14}}$ (see Eq. 3.26). The partial widths can be expressed by the expressions below, which are rescalings of the corresponding SM widths. For more details see Ref. [34].

For MCHM_5, the widths are rescaled as:

$$\begin{aligned}
\Gamma(h \to b\bar{b}) &= F_2(\xi)^2 \Gamma_{\text{SM}}(h \to b\bar{b}) \,, \\
\Gamma(h \to c\bar{c}) &= F_1(\xi)^2 \Gamma_{\text{SM}}(h \to c\bar{c}) \,, \\
\Gamma(h \to \tau^+\tau^-) &= F_2(\xi)^2 \Gamma_{\text{SM}}(h \to \tau^+\tau^-) \,, \\
\Gamma(h \to VV) &= F_2(\xi)^2 \Gamma_{\text{SM}}(h \to VV) \,, \\
\Gamma(h \to gg) &= F_1(\xi)^2 \Gamma_{\text{SM}}(h \to gg) \,,
\end{aligned} \tag{3.29}$$

where

$$F_1(\xi) = \frac{1-2\xi}{\sqrt{1-\xi}} \,, \qquad F_2(\xi) = \sqrt{1-\xi} \,. \tag{3.30}$$

While for the MCHM_{14}, the bottom channel is controlled by F_1 instead of F_2, and the ggh coupling is controlled by $c_g^{\mathbf{14}}$ of Eq. (3.26) instead of $F_1(\xi)$. The total width is obtained

as the sum of all above partial widths. Branching fractions can be obtained as the ratio of partial widths to the total width. These formulas for taking into account the Higgs decays are useful for implementing the models in Monte Carlo generators. In this Thesis, the analysis comes from results obtained at parton level, with no decays implemented. Subject of a future work, the implementation of simulations of processes considering these models are expected to go through hadronization and detector simulation, where these decay widths are to be applied.

3.2 Numerical Analysis of the MCHMs

This section describes the analysis strategy and can be compared with Section 2.4, with the remark that here the parameter space is scanned in a different way, with cross sections being calculated point by point with MadGraph 5 in contrast with the simpler scan of the previous chapter.

3.2.1 Definition of the Parameter Space

The top sector of the models described in Sections 3.1.1 and 3.1.2 is controlled by the (vector-like) mass parameters, M_i, and by several dimensionless couplings, y_{Li}, y_{Ri}, which are a priori complex. In order to identify the number of physical phases the following procedure is done. The phase of each M_i is absorbed (thus making it real and positive) by redefining the phases of $\Psi_{i,L}$ or $\Psi_{i,R}$. This leaves one free phase in each such pair, say in $\Psi_{i,R}$, that can be adjusted to absorb the phase of the corresponding y_{Li} (thus making all of the y_{Li} real and positive). Finally, the phase of one of the y_{Ri} is absorbed into T^5_R or T^{14}_R. In conclusion there is one physical phase in the MCHM$_5$ and two in MCHM$_{14}$. This procedure can also be done by starting with the y_{Li} and y_{Ri} set to be real and positive, so all physical phases are included in M_1 for the MCHM$_5$, and in M_1, M_4 for the MCHM$_{14}$.

In this work, for simplicity, it is assumed that all parameters are real, which amounts to imposing CP conservation in the strong sector. This leaves one physical sign in the case of the MCHM$_5$ and two signs in the MCHM$_{14}$. These signs are chosen to be sign(M_1) in the first case and sign(M_1), sign(M_4) in the second.

While the above parameters respect the $SO(4)$ symmetry, in general they violate the $SO(5)$ symmetry. The $SO(5)$ symmetric limit corresponds to $M_1 = M_4$ and $y_{L1} = y_{L4}$, $y_{R1} = y_{R4}$ for the MCHM$_5$, and $M_1 = M_4 = M_9$ and $y_{L1} = y_{L4} = y_{L9}$, $y_{R1} = y_{R4} = y_{R9}$ for the MCHM$_{14}$. Deviations from the $SO(5)$ symmetric limit in the above couplings lead to a "hard" breaking of the symmetry, in the sense that the Higgs effective potential is

finite when the $SO(5)$ symmetry relations are satisfied, but becomes UV sensitive when not [1]. Deviations from the $SO(5)$ symmetric limit in the M_i, on the contrary, correspond to a "soft" breaking, and the Higgs potential remains IR dominated in that case. This motivates the focus on the case where

$$y_{Li} \equiv y_L \ , \qquad y_{Ri} \equiv y_R \ , \tag{3.31}$$

for all i, as a way to reduce the size of the parameter space. However, small deviations from this limit mean that one can expect additional UV dependent contributions to the Higgs potential. Here the mass of the Higgs is taken as an independent parameter. The calculation of the Higgs potential in order to check if, for a given point of the parameter space, the Higgs mass is correctly reproduced, is left for future work.

The free parameters of this model are the compositeness scale f, the mass parameters $|M_1|$ and $|M_4|$, sign(M_1) and the couplings y_L and y_R, for MCHM$_5$. For MCHM$_{14}$, added to these same parameters are the mass parameter $|M_9|$ and sign(M_4). However, as already discussed in Section 2.2.1, one of those parameters, for each model, can be fixed by requiring that the top mass is reproduced. Here y_R is fixed requiring that:

$$\mathrm{Det}(\mathcal{M}_{2/3}\mathcal{M}_{2/3}^T - \bar{m}_t^2 \mathbb{1}) = 0 \ , \tag{3.32}$$

which is solved numerically for y_R for each choice of the remaining free parameters. When there are no real solutions to this equation, the parameter point is discarded.

For $\bar{m}_t$ in Eq. (3.32), the running top mass of 150 GeV is also used, for the same reasons as for the vector-like Singlet and Doublet cases. For the MC simulation of the $t\bar{t}H$, and also for the parts of $t\bar{t}HH$, in pp collisions with MadGraph 5, to be described in the following section, the relevant scales are of the order of a couple hundred GeV. Different scales are then used in the implementation of the models on the event generator, i.e. for the diagonalization of the mass matrix, the high scale top mass (150 GeV) is used, and for the pp collision simulation, a low scale mass is used, calculated at the top pole (173 GeV). This is an approximation of the effect of the running between the two scales.

The strategy to extract the physical quantities is the same than in Section 2.2: given values for the M_i and for y_L, y_R is found from Eq. (3.32). The fermion mass matrix is diagonalized numerically to obtain the spectrum and the unitary transformations U_L and

[1] This can be seen from the trace of $\mathcal{M}_{2/3}^\dagger \mathcal{M}_{2/3}$, which becomes independent of the Higgs in this limit, signifying the cancellation of quadratic divergences to the potential.

U_R such that

$$U_L \mathcal{M}_{2/3} U_R^\dagger = \text{diag}(m_t, M_{T^{(1)}}, , M_{T^{(2)}}, \ldots) \ , \tag{3.33}$$

with all the physical masses real and positive. This is done with Mathematica [51]. The charge $-1/3$ sector in the MCHM$_{14}$ is also treated numerically. The effects of the bottom mass in the diagonalization are negligible, and the correct couplings between the bottom quark and the Higgs can be incorporated based on the description of Section 3.1.3. In what follows, the bottom coupling is not used (given the interest in $t\bar{t}H$ and $t\bar{t}HH$ at parton level). The physical spectrum and the rotation matrices are the main input to the rest of the numerical analysis, and they are always calculated numerically as well.

Other quantities of interest, such as the Yukawa matrix in the mass eigenbasis, can be obtained with

$$Y_{2/3}^{\text{mass}} = U_L Y_{2/3}^{\text{gauge}} U_R^\dagger \ , \tag{3.34}$$

where

$$Y_{2/3}^{\text{gauge}} \quad = \quad \frac{d}{dh_0} \mathcal{M}_{2/3} \tag{3.35}$$

is the Yukawa matrix in the gauge eigenbasis. The most relevant quantity will be the $(1,1)$ entry in $Y_{2/3}^{\text{mass}}$, which corresponds to the top Yukawa coupling. It includes exactly all tree-level effects arising from the Higgs compositeness (the dependence through $\sqrt{\xi} = \sin \frac{h_0}{f}$), as well as the mixing with the vector-like resonances.

The non-linear dependence on h in composite Higgs models leads to interactions between top pairs and a number of Higgs bosons. The $tthh$ vertex, whose Feynman rule is given by i times the (1,1) entry of $d^2 Y_{2/3}^{\text{gauge}}/dh_0^2$, after rotating to the mass eigenbasis, also enters in the analysis.

The way the models are implemented in the event generators is by using FeynRules, thus implementing the full Lagrangian. UFO files are the output of FeynRules which is fed into MadGraph 5. The diagonalization is done point-by-point of the parameter space. This means that, in principle it is possible to generate one UFO file for each point. A Python script, however, is used to interface the results of the diagonalization, which means all the spectrum plus Yukawa couplings, as well as all the entries of the associated rotation matrices (both U_L and U_R), and also the coupling y_R, written in function of the top mass (note that this changes for every parameter space point). The script calls MadGraph, sets up the program running parameters, such as beam energy, process definition (e.g.

proton proton producing top, anti-top and a Higgs). After that, it updates the default param_card.dat of MG5 with the parameters of a given point, and collects the results, which are the cross sections at 14 and 27 TeV center of mass energies of each process (to be defined below), the physical masses of all heavy resonances as well as the decay widths (tree level) of all particles. A loop goes over all the parameter space points, which are given beforehand. The collection of parameter space points is generated separately with a Mathematica script, where a grid of points is generated, already discarding points that either don't reproduce the top mass, or have resonances with masses too light when compared to current experimental bounds for direct searches of vector-like quarks (see Section 2.3.5). After collecting the results, if a point deviates from the ∼20% measured values for the signal strengths of μ_{ggH} and $\mu_{t\bar{t}H}$, it is also discarded. The top quark width also has small deviations, given the current relatively high lower bound on the possible new particle masses.

The MadGraph 5 implementation used here considers the default definition of the proton, which contains gluons and up, down, charm, strange quarks (and the corresponding anti-quarks). The default MadGraph 5 PDF function is also used (nn23lo1) [80].

3.2.2 The mass spectrum and the physical top-Higgs Yukawa coupling

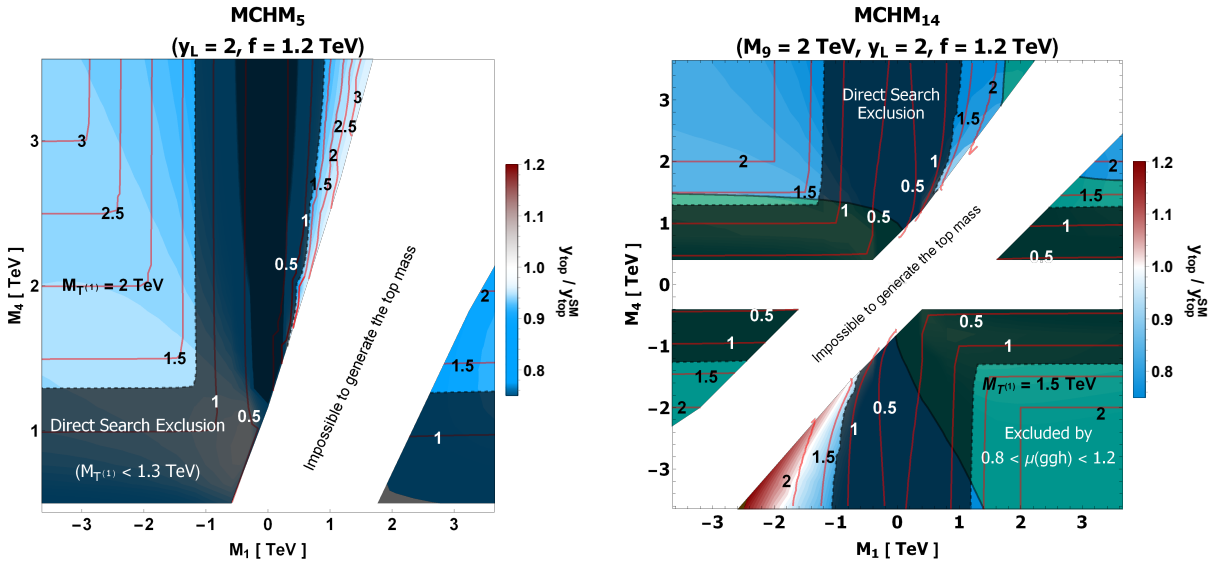


Figure 8: Contour plot of the top Yukawa coupling, $\mu_{\rm top} \equiv y_{\rm top}/y_{\rm top}^{\rm SM}$ (blue and red colors), as a function of M_1 and M_4. Blue(red) indicates $\mu_{\rm top} < 1$ ($\mu_{\rm top} > 1$). Curves of masses of the lightest $Q = 2/3$ resonance, $M_{T^{(1)}}$, is shown. Darker bands show direct exclusion limits for the resonances, assuming they decay only in the bW, tZ and tH channels [48, 49]. The top mass can not be reproduced in the white region, and limits on the ggH coupling is also shown (relevant only for a region in MCHM$_{14}$).

Before going through the analysis with the event generator, it is possible to already extract some features of the two composite Higgs models with the information of the mass matrices. The diagonalization of the $Q = 2/3$ matrix is done with Mathematica, already leading to the mass spectrum and the rotation matrices, both obtained numerically. When the rotation is applied to the Yukawa matrix, the physical Yukawa interactions are obtained. Up to small EW symmetry breaking effects, the mass spectrum of the charge $Q = 2/3$ resonances, which is of interest because of the mixing with the third generation of quarks, can be approximated to $\left[M_4\ , \sqrt{M_4^2 + y_L^2 f^2}, \sqrt{M_1^2 + y_R^2 f^2}\right]$ for MCHM$_5$ and to $\Big[M_4\ , \sqrt{M_4^2 + y_L^2 f^2}, \sqrt{M_1^2 + y_R^2 f^2}, M_9\ (\text{degeneracy} = 3)\Big)]$ for MCHM$_{14}$. These approximations, while useful for a first visualization of the mass spectrum, is not used in any of the analyses done here.

The lightest resonance, defined here as $T^{(1)}$, deserves special care, since its mass has to be within the bounds of direct searches (see Section 2.3.5). As shown later for some examples, some decay channels, apart from the standard bW, tZ and tH channels considered in the experimental searches, are observed, such as three body decays, which are not negligible. The current bounds must be thus taken with care.

The top Yukawa coupling is one of the most important quantities for this chapter as well. Figure 8 shows how $\mu_{\text{top}} \equiv y_{\text{top}}/y_{\text{top}}^{\text{SM}}$ behaves within a slice of the parameter space in the M_1-M_4 plane defined by a constant $y_L = 2$, $f = 1200$ GeV and $M_9 = 2$ TeV (for MCHM$_{14}$). Other slices display similar qualitative behaviours. For MCHM$_5$ it is observed, as already expected, that $\mu_{\text{top}} < 1$ everywhere. For MCHM$_{14}$, $\mu_{\text{top}} < 1$ in most regions, but it is observed that certain regions allow for $\mu_{\text{top}} > 1$ (colored in red), as pointed out in Ref. [52]. Curves of the mass of the lightest resonance, $M_{T^{(1)}}$, are shown and the approximate direct exclusion from direct vector-like resonance searches are shown in darker bands. The mass of the top can not be reproduced in the white area and the regions where the coupling ggH [Eqs. (3.25) and (3.26)], deviates by more than 20% (which applies mostly to MCHM$_{14}$) are also shown.

3.2.3 The production of $t\bar{t}H$ in pp collisions

The $t\bar{t}H$ process in the MCHM scenarios, although in a richer model environment when compared to the simple vector-like fermion models of the previous chapter, is still related in a similar simple way to the SM production of this process. The Feynman diagrams (at tree level) are identical in all the models, involving top/anti-top pair production associate to a Higgs boson (Fig. 2 applies here as well). As the simple models of Chapter 2, the diagrams involving initial states of $q\bar{q}$ can be neglected, the amplitude is proportional

to the top Yukawa coupling, and the cross section, for the MCHM models, can also be expressed as:

$$\sigma_{\mathrm{MCHM}}(\mathrm{t\bar{t}H}) \;=\; \left(\frac{y_t}{y_t^{\mathrm{SM}}}\right)^2 \sigma_{\mathrm{SM}}(\mathrm{t\bar{t}H})\ . \tag{3.36}$$

All the modifications due to Higgs compositeness or mixing with vector-like fermions enter only through the top Yukawa coupling. The dominant effect comes from compositeness and depends on ξ, while vector-like mixing is subdominant, due to the already heavy mass constraints. Therefore, as in other BSM cases, a modification in the total rate with respect to the SM is expected, but not in kinematic distributions. This means that the BSM effects for this case depend on the precision of the $\mathrm{t\bar{t}H}$ measurements only. (see Section 3.3)

3.2.4 The production of $\mathrm{t\bar{t}HH}$ in pp collisions

The top quark-anti-quark pair production in association to double Higgs production ($\mathrm{t\bar{t}HH}$) in pp collisions has a richer phenomenology when compared to $\mathrm{t\bar{t}H}$. As mentioned before, this process has two main contributions. The first one involves the production of charge 2/3 vector-like, and is largely dominated by a pair production of these resonances that further decay into the tH channel. This process depends mainly on the mass of the resonances.

The second main contribution is the production of $\mathrm{t\bar{t}HH}$ via Feynman diagrams that do not involve the resonances, defined here as non resonant $\mathrm{t\bar{t}HH}$, or NR-$\mathrm{t\bar{t}HH}$, and is illustrated in Fig. 9. This Figure shows representative tree level Feynman diagrams, involving only the Yukawa vertex (left), the Higgs trilinear self-coupling (middle)[2], and the "double Higgs" Yukawa vertex [81] (right). Apart from this rightmost diagram, the other two are just the SM diagrams (with of course the corresponding $\mathrm{MCHM_5}$ or $\mathrm{MCHM_{14}}$ modified couplings).

Fig. 10 shows the plot of the tH invariant mass distribution for for the full $\mathrm{t\bar{t}HH}$ process (blue curve) and the non-resonant (NR-$\mathrm{t\bar{t}HH}$) process (red curve) for a particular parameter set in the $\mathrm{MCHM_5}$. It is also compared to the SM simulation of the $\mathrm{t\bar{t}HH}$ process for the corresponding energies. It is noticeable that the NR-$\mathrm{t\bar{t}HH}$ follows the SM distribution. This can be understood by looking at the diagrams of Fig. 9 if the rightmost diagram, which is not present in the SM, is not contributing much to the process. The NR-$\mathrm{t\bar{t}HH}$ is suppressed with respect to the SM and follows mostly from the changes in

[2]The triple Higgs coupling changes as $\lambda = \left[(1-2\xi)/\sqrt{1-\xi}\right]\lambda_{\mathrm{SM}}$.

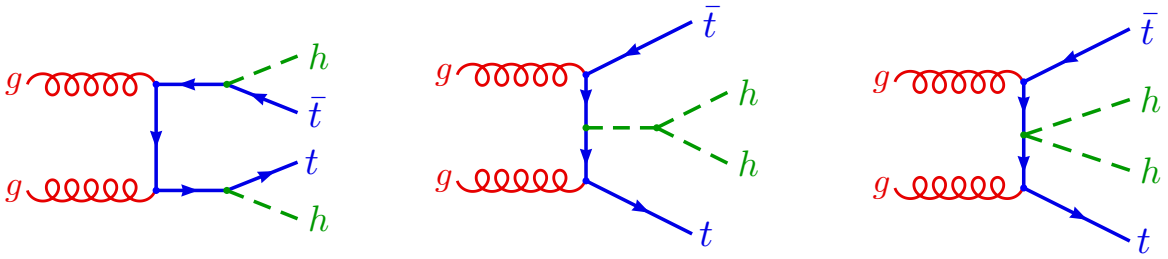


Figure 9: Representative diagrams of the non-resonant $t\bar{t}HH$ process. There are three distinct processes containing vertices that involve only the Yukawa vertex (left), the Higgs trilinear self-coupling (middle) and the "double Higgs" Yukawa vertex (right), this last one being a new diagram arising in the MCHMs.

the top Yukawa (μ_{top}<1). When increasing the center of mass energy, all cross sections increase, and the relative importance of the resonant $t\bar{t}HH$ production increases.

3.2.5 The scan over the parameter space

The main results of this chapter are obtained after a cross section scan for the $t\bar{t}H$ and $t\bar{t}HH$ processes. Differently from Chapter 2, where $\mu_{t\bar{t}H}$ is the only studied quantity and can be easily be parameterized as a ratio of the top Yukawas as shown in Eq.(2.26), here the $t\bar{t}HH$ process requires a numerical approach, and is calculated point by point on the parameter space with the diagonalization described in the previous section also performed point by point.

The parameter space of the $MCHM_5$ and the $MCHM_{14}$ studied in this Thesis considers a rather low scale, focusing more in in the range of reach at the HL-LHC. They are also studied for energies of the HE-LHC for comparison. It is, however, expected by the time that the HE-LHC (or any of the successors of the LHC) starts taking data, a number of regions of the parameter space will already have been excluded by the HL-LHC, or, even before, by current results being produced with the LHC Run 2 collected data and the upcoming Run 3.

	Range	
	$MCHM_5$	$MCHM_{14}$
$\lvert M_1\rvert$ (TeV)	[0.8, 3.0] TeV	[0.8, 3.0]
M_4 (TeV)	[1.2, 3.0] TeV	[-3.0, -1.2]∪[1.2, 3.0]
M_9 (TeV)	–	[1.3, 4.0]
f (TeV)	[0.8, 2.0] TeV	[0.8, 2.0]
y_L	[0.5, 3.0]	[0.5, 3.0]

Table 4: Range of each free parameter considered for the MadGraph 5 scan, for the two MCHMs considered here. A grid of points is generated in these ranges, which is further fed to the event generator.

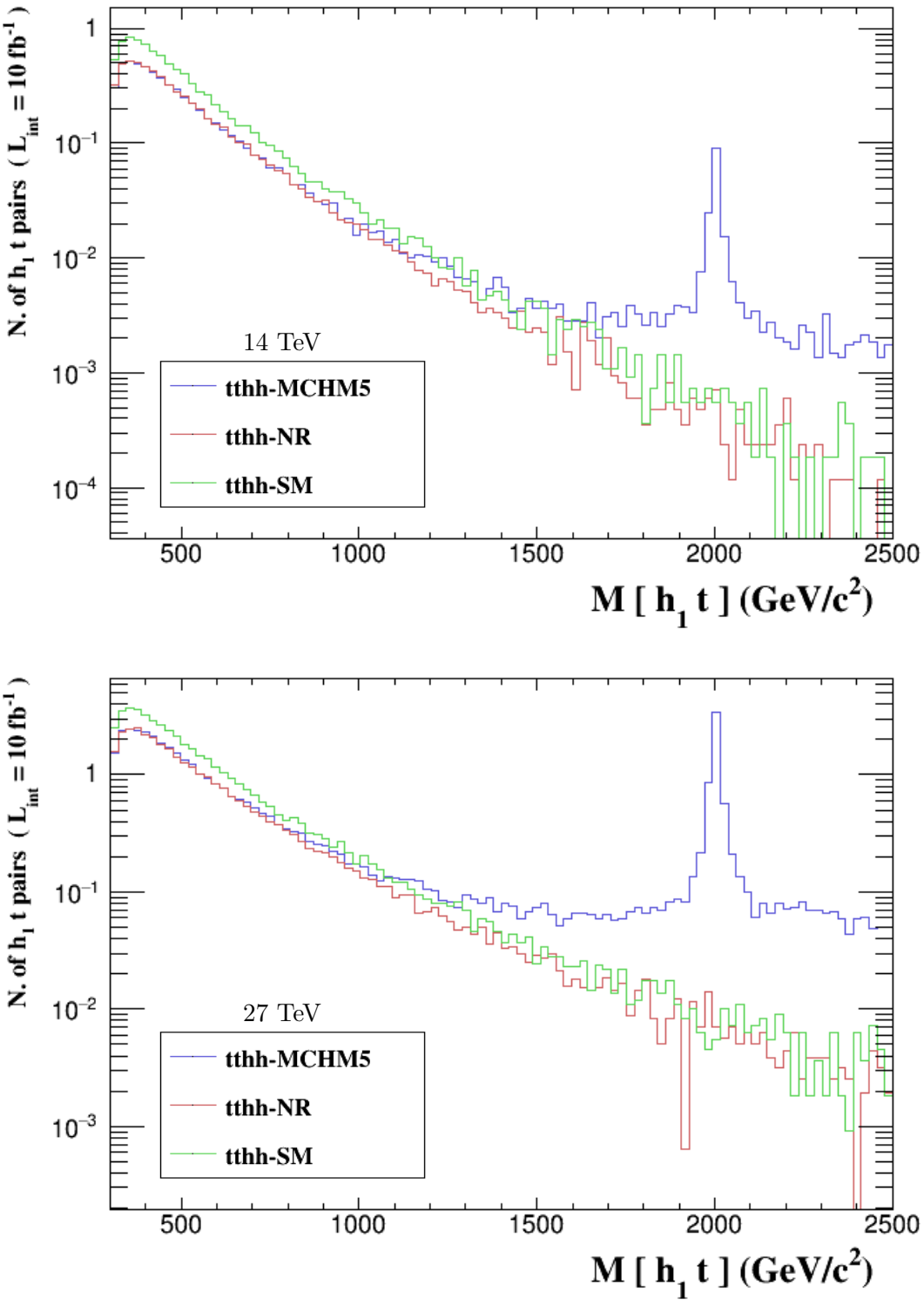


Figure 10: Invariant mass of the top quark and the hardest Higgs boson for a particular set of parameters for MCHM$_5$ ($M_1 = -2.5$ TeV, $M_4 = 2$ TeV, $f = 1.8$ TeV, $y_L = 1$). The blue curve shows the full $t\bar{t}HH$ process in the MCHM$_5$, while the NR-$t\bar{t}HH$ distribution is shown in red. The green curve corresponds to the process in the SM. A center of mass energy of 14 Tev (27 TeV) is used for the generation of events (MadGraph 5) for the upper (lower) plots, which are produced with MadAnalysis 5 [82].

Table 4 shows the range of each free parameter that is considered for each model. Note the upper limit on $y_L < 3$, taken in order to remain in the (semi-) perturbative regime, thus justifying the present at tree-level analysis. For the same reason, it is also checked that y_R, as determined by the top mass, is below 4. The distribution of points within those ranges is not uniform due to computing constraints, but it covers most of the parameter space.

3.3 Results of the MCHM parameter space scan

The results of the scan over the parameters, in the range described above, is presented here. The M_1-M_4 plane is divided in regions based on the physical sign of the mass parameters. For MCHM_5, which has only one sign as free parameter, these regions are defined as below, and are used in the presentation of the results in this section.

$$\text{Region I: } M_1 > 0 \qquad \text{Region II: } M_1 < 0 \ .$$

For the MCHM_{14}, four regions are defined as:

$$\begin{aligned} &\text{Region I: } M_1, M_4 > 0 && \text{Region II: } M_1 < 0,\ M_4 > 0 \\ &\text{Region III: } M_1, M_4 < 0 && \text{Region IV: } M_1 > 0,\ M_4 < 0 \ . \end{aligned}$$

3.3.1 Results on $t\bar{t}H$ production

The $t\bar{t}H$ cross section (see Eq. (3.36), is simply related to the SM one by a rescaling of the top Yukawa coupling. As already discussed, the effects due to compositeness dominate, and are controlled by ξ (or f). This is illustrated in Fig. 11. The signal strength $\mu(t\bar{t}H)$ of the production cross section does not depend on the pp collision energy at tree level. For generating this Figure, a 14 TeV center of mass energy is chosen in MadGraph. The CMS measurement of $\mu(t\bar{t}H)$ (see Section 2.3.4) already excludes points of this plot for $\mu(t\bar{t}H)<0.6$ at 2σ level.

For MCHM_5 the points of Region I follow two distinct behaviours. Most of the points colored in blue fall in a curve around a region that is near the experimental bounds on $t\bar{t}H$. They are also the points corresponding to the lowest mass of the lightest resonance, and are near exclusion from this point of view. The other points, mostly brown dots, correspond to points with a heavier mass spectrum and are closer to the SM limit of $\mu(t\bar{t}H)=1$. This is not observed in Region II, where all points follow a same pattern, independently of the mass. All points have $\mu(t\bar{t}H) < 1$, as expected for MCHM_5.

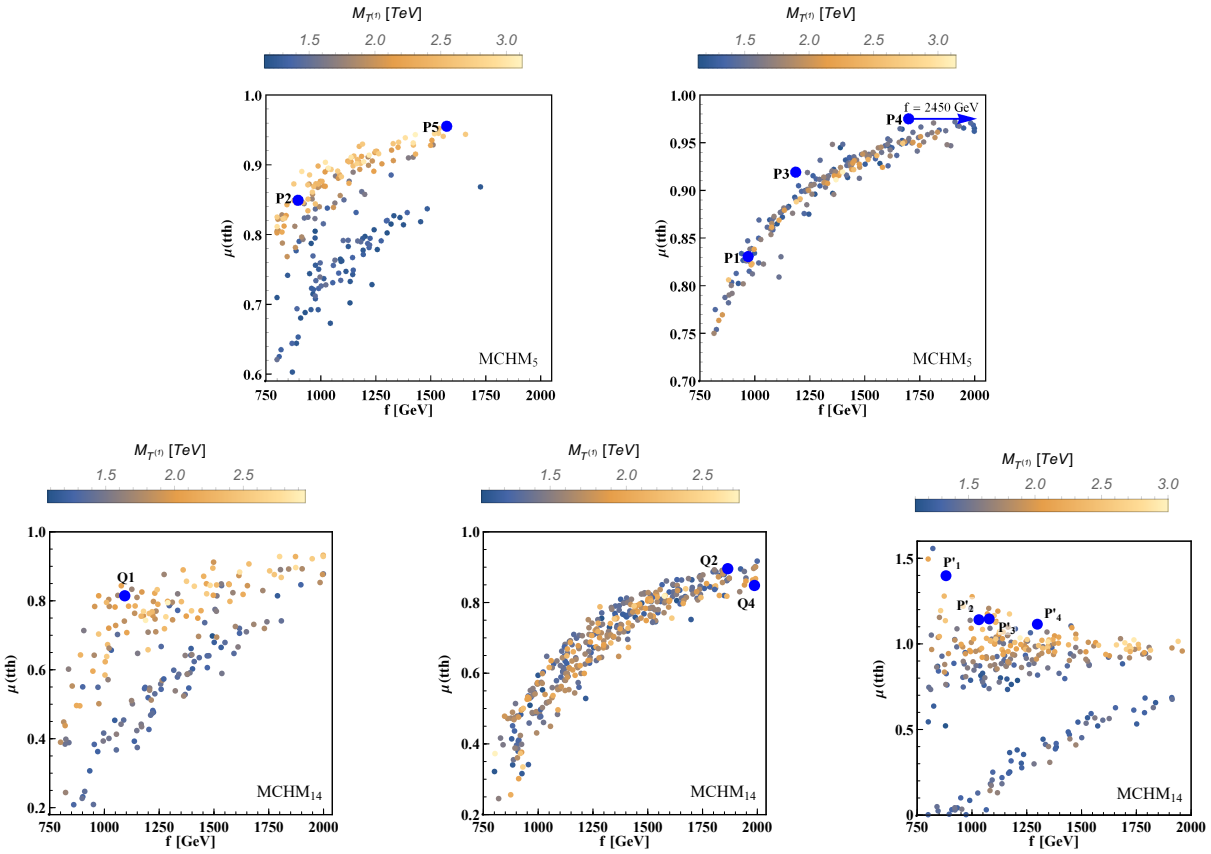


Figure 11: The $\mu(t\bar{t}H)$ is shown as a function of f, and is independent of the pp collision energy. The lightest vector-like mass is color coded. The upper left (right) plot corresponds to the Regions I (II) of $MCHM_5$. The lower left, central and right plots correspond to the Regions I, II + IV and III of $MCHM_{14}$, respectively. The blue arrow in the upper right plot indicates that the example point P_4 is outside the horizontal range of the plot with $f = 2450$ GeV. Also in this Figure are example points to be defined in Section 3.4.

The $MCHM_{14}$ model has three different distinct regions. Region I behaves qualitatively in the same way as the Region I of $MCHM_5$, with the difference of some points going as low as around $\mu(t\bar{t}H)$=0.2, and are surely in tension with experiments. Regions II and IV of $MCHM_{14}$ are similar, and behave as Region II of $MCHM_5$, with points also going to values of $\mu(t\bar{t}H)$ as low as 0.2.

Region III of $MCHM_{14}$ shows a different behaviour, not existent in $MCHM_5$. This is because it is precisely in this region that some points in parameter space can lead to $\mu(t\bar{t}H) > 1$. Most of the points fall along the SM limit of $\mu(t\bar{t}H) = 1$, while some points can go to $\mu(t\bar{t}H)$ values as high as 1.5. There is also some points of low mass (blue dots) that follow a curve around lower values of $\mu(t\bar{t}H)$ and are in tension with the experimental limits on $t\bar{t}H$ production.

These two models show similar and distinct behaviours in some regions. The parameter space of both models can be further constrained with improved $t\bar{t}H$ experimental searches.

This measurement can also disentangle between the two models. If $\mu(t\bar{t}H) < 1$, both MCHM scenarios are possible, and the distinction between them will depend on a more detailed phenomenology. Now if a measurement of $\mu(t\bar{t}H) > 1$ is observed, $MCHM_5$ will be excluded.

3.3.2 Results on $t\bar{t}HH$ production

This section discusses the results for the $t\bar{t}HH$ production process. Fig. 12 shows how the signal strength $\mu(t\bar{t}HH)$ depends on the mass of the lightest $Q = 2/3$ resonance for both MCHM models. The same points are simulated in 14 and 27 TeV center of mass energies. The points for $MCHM_{14}$ are more spread (See the different scales of the plot), going to values of $\mu(t\bar{t}HH)$ as high as 12. It should be noted that more points are generated in the $MCHM_{14}$ case, reflected in the high density of points in the plot. These plots show, as expected, that for lighter resonances the resonant production can result in a significant enhancement of the total $t\bar{t}HH$ cross section. This effect increases with center of mass energies, and for such points the pair production of heavy resonances that decay into the tH channel is dominating. Points on the right hand side of the plots (higher masses) have a dominating NR-$t\bar{t}HH$ process. Fig.13 shows exactly this behaviour, i.e. points with higher mass correspond to a higher percentage of NR-$t\bar{t}HH$ cross section. Looking at the same point, it is seen that increasing from 14 to 27 TeV, the resonant part contributes a bit more, thus the red points are always below the blue points. However, because of Higgs compositeness effects, the cross sections are not reaching SM values just by increasing the mass. So points with heavy masses, even around 3 TeV, show a deviation from the SM. To achieve the SM limit, one must also increase the values of f.

As already discussed in the early sections, there is a correlation between the $t\bar{t}H$ and the $t\bar{t}HH$ production, since both are controlled by the top-Yukawa coupling. To better see this, Fig. 14 shows the relation between $\mu(t\bar{t}H)$ and $\mu^{NR}(t\bar{t}HH)$ (thus taking out resonance effects). There is a clear correlation between the two quantities. The curves are not perfectly correlated because the NR-$t\bar{t}HH$ process receives contributions from the diagram involving the triple Higgs coupling, while the $t\bar{t}H$ production does not.

3.4 Selection of example parameter space points

In this section, some parameter space points are selected, which will be called example points, and many of their properties are shown. They complement the exploration of the

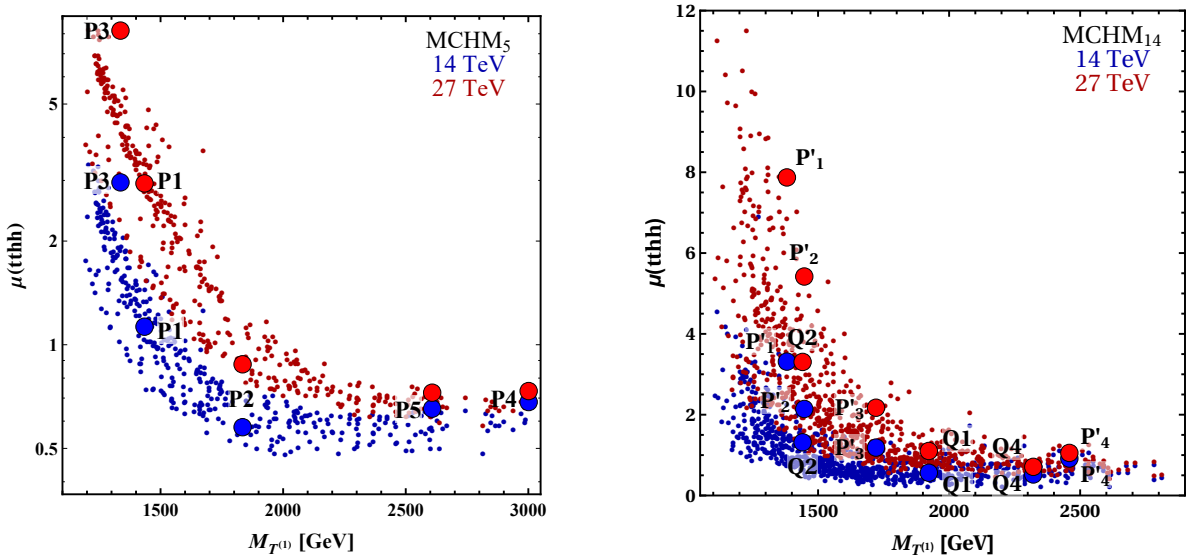


Figure 12: Value of $\mu(t\bar{t}HH)$ as a function of the mass of the lightest $Q = 2/3$ vector-like fermion, for 14 (blue) and 27 TeV (red) center of mass energies. The left (right) plots correspond to the MCHM$_5$ (MCHM$_{14}$).

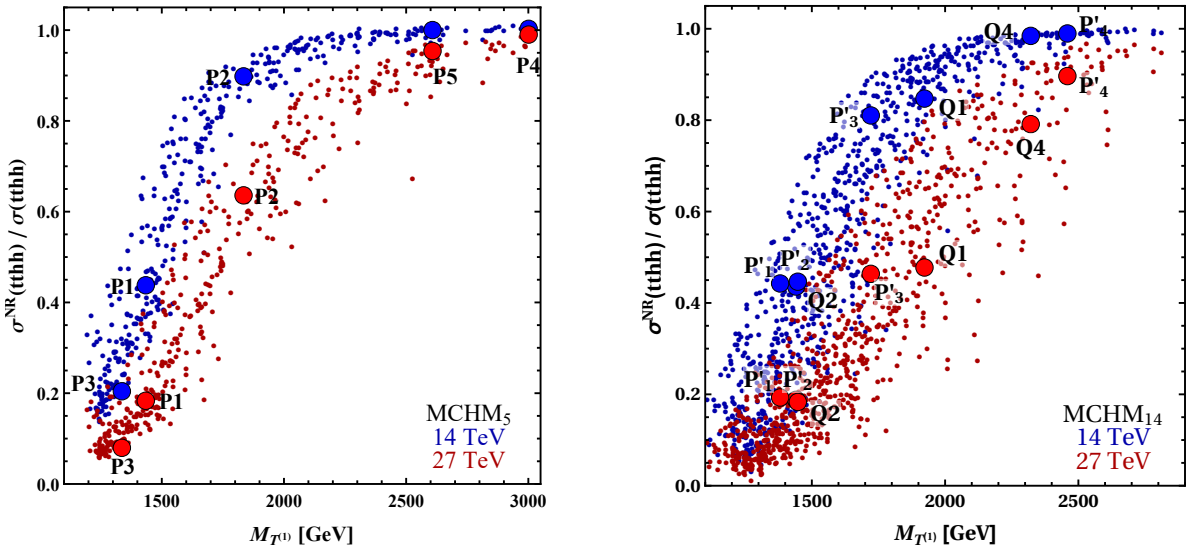


Figure 13: Ratio between the NR-$t\bar{t}HH$ cross section and the total $t\bar{t}HH$ cross section as a function of the lightest $Q = 2/3$ vector-like mass, for 14 and 27 TeV center of mass energies. The left (right) plots correspond to the MCHM$_5$ (MCHM$_{14}$).

parameter space, and some can also be selected for undergoing a full detector simulation, which is left for future work.

3.4.1 Selection of parameter space example points for MCHM$_5$

The Table 5 shows selected example points for MCHM$_5$ to illustrate some properties of the model. The free parameters are shown, together with the cross sections that come out of the Monte Carlo simulation, the mass spectrum, and branching ratios (obtained

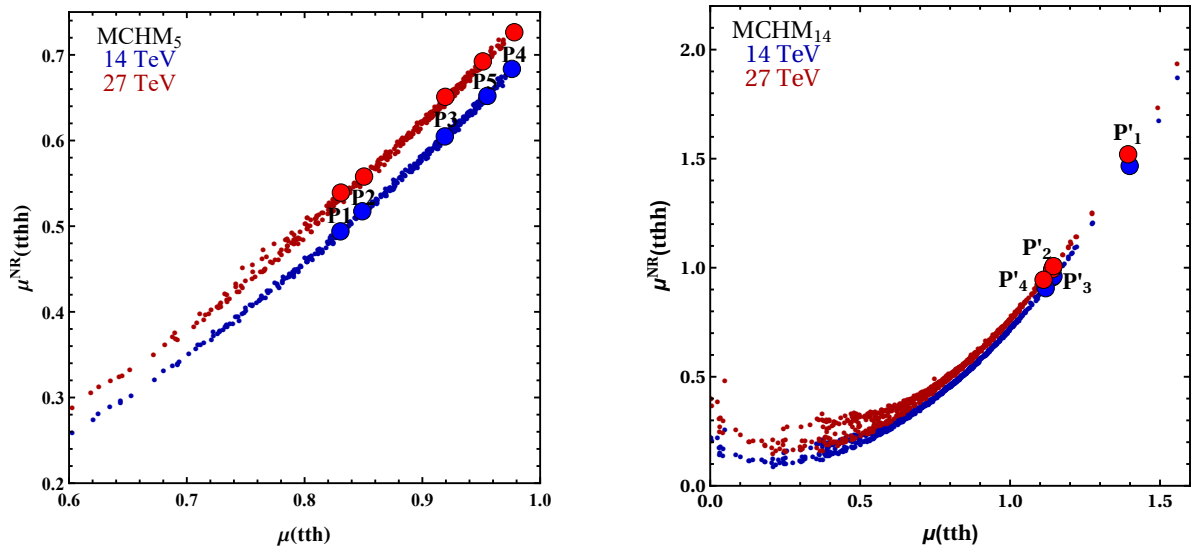


Figure 14: Correlation between $\mu(NR-\mathrm{t\bar{t}HH})$ and $\mu(\mathrm{t\bar{t}H})$, for 14 and 27 TeV center of mass energies. The left (right) plots correspond to $\mathrm{MCHM_5}$ ($\mathrm{MCHM_{14}}$).

numerically at tree level). The value of y_R, which fixes the top quark mass, is also shown. Points from different Region (I and II) are selected, but no difference between regions can be seen in the individual points.

		P_1	P_2	P_3	P_4	P_5
parameters	M_1(GeV)	-1317	800	-960	-3550	914
	M_4(GeV)	1580	2311	1400	3000	2632
	f(GeV)	969	896	1186	2450	1573
	y_L	1.66	1.80	0.88	1.00	2.36
	y_R	0.62	1.95	0.87	0.85	2.41
$\mu(\mathrm{t\bar{t}H})$ (All Energies)		0.83	0.85	0.92	0.97_5	0.95
$\mu(\mathrm{t\bar{t}HH})$ (14 TeV)		1.30	0.71	3.40	0.93	0.88
$\mu(\mathrm{t\bar{t}HH})$ (27 TeV)		3.18	0.96	9.12	0.95	0.91
NR-ttHH/ttHH (14 TeV)		0.50	0.97	0.24	1.01	1.02
NR-$\mathrm{t\bar{t}}$HH/$\mathrm{t\bar{t}}$HH (27 TeV)		0.21	0.72	0.09	1.00	0.99
$M_{T^{(1)}}$ (TeV)		1.44	1.83	1.34	3.00	2.61
$M_{T^{(2)}}$ (TeV)		1.59	2.37	1.45	3.82	3.91
$M_{T^{(3)}}$ (TeV)		2.25	2.83	1.76	3.99	4.56
$M_{B^{(1)}}$ (TeV)		2.25	2.82	1.75	3.87	4.56
$M_{X_{5/3}}$ (TeV)		1.58	2.31	1.40	3.06	2.63
BR($\mathrm{T^{(1)}} \to$tH)		0.33	0.30	0.61	0.31	0.33
BR($\mathrm{T^{(1)}} \to \mathrm{W^+b}$)		0.46	0.46	0.06	0	0.16
BR($\mathrm{T^{(1)}} \to$tZ)		0.22	0.20	0.26	0.30	0.26
BR($\mathrm{T^{(1)}} \to \mathrm{W^+W^-t}$)		0	0.04	0.07	0.39	0.25

Table 5: Properties of selected example points in the $\mathrm{MCHM_5}$. P_2 and P_5 belong to Region I ($M_1, M_4 > 0$); P_1, P_3 and P_4 belong to Region II ($M_1 < 0, M_4 > 0$).

The first point in Table 5 (P_1), with a $\mathrm{t\bar{t}H}$ cross section about 20% below the SM, has some interesting features. It has a 50% NR-$\mathrm{t\bar{t}}$HH contribution at 14 TeV and a rather low $M_{T^{(1)}}$ mass. When going to 27 TeV, the NR-$\mathrm{t\bar{t}}$HH contribution goes to about 20%. Points like this can be searched for by direct resonance searches. The branching ratios are in the standard channels considered in these searches.

Point P_2, on the other hand, is an example of when the NR-$t\bar{t}$HH cross section dominates. The $t\bar{t}$H cross section is similar to that of P_1, but the dominance of the non-resonant cross section leads to a decrease of the $t\bar{t}$HH overall cross section. Point P_2 also has a small value of f, thus leading to strong compositeness effects.

Point P_3 has one of the smallest resonance masses of the entire scan, and an f value higher than the previous points. The resonance production cross section dominates, and is reflected in the $t\bar{t}$HH cross section. Even if $t\bar{t}$H is closer to the SM, the BSM effects can be seen via resonance searches.

If BSM effects like the ones of points P_1 to P_3 exist in nature, they can be accessed already with the HL-LHC. Points P_4 and P_5, however, have larger masses and $t\bar{t}$H cross sections closer to the SM. The NR-$t\bar{t}$HH is dominating, and very close to SM values as well. Those are examples of points that need higher energy to be accessed. Another striking feature of these points is that the three body decay of the resonances is not negligible, and even dominates other channels as in P_4.

These different points show some specific features that complement the information already shown in the plots of last sections. A full detector simulation with some of these example points could allow for investigating BSM effects in future hadron colliders.

3.4.2 Selection of parameter space example points for MCHM$_{14}$

The example points for MCHM$_{14}$ are shown in Table 6. Of the four distinct regions, Region III deserves special attention since points there can show an enhancement on the top-Higgs Yukawa coupling. The first four points (P'$_1$ to P'$_4$) in Table 6 are the selected example points for this region. The other points are selected one for each of the remaining regions and show a similar behaviour to the points of MCHM$_5$, with a remark on the non negligible three body decay of the lightest resonance.

The points of Region III are all selected because of the increase in the cross section of $t\bar{t}$H. This means that even for points with a dominating NR-$t\bar{t}$HH cross section, such as points P'$_3$ and P'$_4$, both $t\bar{t}$H and $t\bar{t}$HH cross sections can be bigger than the SM ones, without the contribution of pair production of resonances, which is normally how the $t\bar{t}$HH cross section increases so much (see Figs.12 and 13)

One general behaviour is a very sharp decrease of the NR-$t\bar{t}$HH contribution when increasing the center of mass energy to 27 TeV. This means resonances that were too heavy to be found in lower center of mass energies become accessible already with doubling the center of mass energies. It is also, in general, very hard to distinguish between the two

		P'$_1$	P'$_2$	P'$_3$	P'$_4$	Q$_1$	Q$_2$	Q$_3$
parameters	M$_1$(GeV)	-1173	-1054	-1084	-1579	976	-1387	2998
	M$_4$(GeV)	-1823	-1826	-1767	-2512	1991	1443	-2318
	M$_9$(GeV)	1382	1448	2036	2714	3096	3115	2875
	f(GeV)	882	1032	1078	1298	1093	1865	1987
	y_L	1.98	1.93	2.95	2.71	1.49	1.52	0.94
	y_R	3.90	2.78	2.67	2.46	3.04	0.34	0.54
$\mu(t\bar{t}H)$ (All Energies)		1.40	1.14	1.15	1.11	0.82	0.89	0.85
$\mu(t\bar{t}HH)$ (14 TeV)		4.27	2.66	1.60	1.29	0.66	1.55	0.67
$\mu(t\bar{t}HH)$ (27 TeV)		9.74	6.37	2.68	1.43	1.17	3.60	0.83
NR-$t\bar{t}HH$/$t\bar{t}HH$ (14 TeV)		0.46	0.49	0.81	0.96	0.95	0.50	1.03
NR-$t\bar{t}HH$/$t\bar{t}HH$ (27 TeV)		0.20	0.20	0.49	0.87	0.54	0.21	0.84
$M_{T^{(1)}}$ (TeV)		1.38	1.45	1.72	2.46	1.92	1.44	2.32
$M_{T^{(2)}}$ (TeV)		1.38	1.45	2.01	2.70	2.47	1.52	2.82
$M_{T^{(3)}}$ (TeV)		1.41	1.46	2.04	2.71	3.09	2.96	2.87
$M_{B^{(1)}}$ (TeV)		1.38	1.45	2.02	2.70	2.53	2.98	2.84
$M_{X_{5/3}}$ (TeV)		1.82	1.83	1.77	2.51	1.99	1.44	2.32
BR(T$^{(1)}$ $\rightarrow$tH)		0.39	0.28	0.44	0.38	0.50	0.42	0.37
BR(T$^{(1)}$ $\rightarrow$W$^+$b)		0.35	0.48	0.14	0.13	0.12	0.15	0.02
BR(T$^{(1)}$ $\rightarrow$tZ)		0.16	0.13	0.30	0.27	0.23	0.34	0.35
BR(T$^{(1)}$ $\rightarrow$W$^+$W$^-$t)		0.09	0.09	0.12	0.22	0.14	0.08	0.27

Table 6: Properties of selected parameter space example points in MCHM$_{14}$. Points P'$_1$ through P'$_4$ belong to Region III, while Q_1, Q_2 and Q_3 belong to Regions I, II and IV respectively.

MCHM scenarios presented here, apart from when there is an increase in the top Yukawa, which if observed would exclude the MCHM$_5$.

3.4.2.1 NR-$t\bar{t}HH$ contributions in MCHM$_5$ and MCHM$_{14}$

As already mentioned before, the NR-$t\bar{t}HH$ cross section depends on the triple Higgs coupling. In the cases that this process dominates, these effects can be explored. In order to see how this coupling contributes to this process, a simulation was done by turning off, in turn, the "double Yukawa" coupling and then the triple Higgs coupling, being left with a cross section depending only on the top Yukawa. For this last case, the amplitude, instead of going as the square of the top Yukawa, as in the $t\bar{t}H$ case, will go as the 4^{th} power of this coupling. This can be seen in Table 7 for MCHM$_5$ and in Table 8 for the MCHM$_{14}$. In the first Table, the diagrams that are turned off are shown. The same diagrams are turned off in the second Table, following the same pattern.

The $\sigma_{\not{hh}}/\sigma_{\rm NR}^{t\bar{t}HH}$ ratios show that the effects of the double Higgs Yukawa coupling are typically at the couple to few percent level in the MCHM$_5$ and the MCHM$_{14}$ cross section, and don't change much with center of mass energies. These diagrams are negligible for all the studies of this work.

When the trilinear Higgs self coupling is turned off, the cross section decreases in a more significant way. This shows that diagrams that involve this coupling account for about 15% of the total NR-$t\bar{t}HH$ cross section at tree level. This value is of about 20% in the SM. Almost all kinematic quantities of the NR-$t\bar{t}HH$ are expected to behave like the

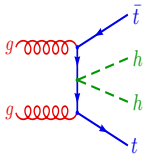
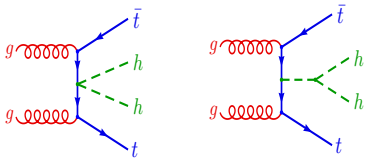
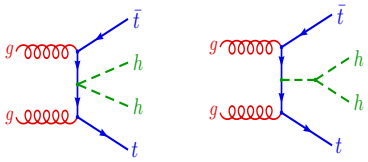
		P$_1$	P$_2$	P$_3$	P$_4$	P$_5$	Disregarded diagrams
$\sigma_{\not{hh}}/\sigma^{t\bar{t}HH}_{NR}$	(14 TeV)	1.05	1.04	1.03	1.01	1.01	
$\sigma_{\not{hh}}/\sigma^{t\bar{t}HH}_{NR}$	(27 TeV)	1.05	1.03	1.03	1.00	1.01	
$\sigma_{Yuk}/\sigma^{t\bar{t}HH}_{NR}$	(14 TeV)	0.86	0.85	0.84	0.82	0.82	
$\sigma_{Yuk}/\sigma^{t\bar{t}HH}_{NR}$	(27 TeV)	0.86	0.86	0.85	0.83	0.84	
$(y_t/y_t^{SM})^4$ $\sigma_{Yuk}/\sigma^{SM}_{Yuk}$	All Energies	0.69 0.69	0.72 0.72	0.84 0.84	0.95 0.94	0.91 0.90	

Table 7: Study of NR-t$\bar{t}$HH for MCHM$_5$ example points in Table 5. The cross sections $\sigma_{\not{hh}}$ and σ_{Yuk} are obtained by disregarding the classes of diagrams on the last column and $\sigma^{t\bar{t}HH}_{NR}$ is the total NR-ttHH cross section. The cross section σ^{SM}_{Yuk} means the SM trilinear Higgs coupling is turned off. The top Yukawas in the MCHM and SM are y_t and y_t^{SM} respectively.

		P'$_1$	P'$_2$	P'$_3$	P'$_4$	Q$_1$	Q$_2$	Q$_3$
$\sigma_{\not{hh}}/\sigma^{t\bar{t}HH}_{NR}$	(14 TeV)	0.95	0.97	0.96	0.98	1.06	1.03	1.05
$\sigma_{\not{hh}}/\sigma^{t\bar{t}HH}_{NR}$	(27 TeV)	0.94	0.97	0.96	0.97	1.06	1.03	1.05
$\sigma_{Yuk}/\sigma^{t\bar{t}HH}_{NR}$	(14 TeV)	0.81	0.82	0.81	0.81	0.86	0.84	0.85
$\sigma_{Yuk}/\sigma^{t\bar{t}HH}_{NR}$	(27 TeV)	0.82	0.83	0.82	0.82	0.86	0.85	0.86
$(y_t/y_t^{SM})^4$ $\sigma_{Yuk}/\sigma^{SM}_{Yuk}$	All Energies	1.94 1.93	1.30 1.29	1.31 1.30	1.24 1.24	0.67 0.66	0.80 0.80	0.72 0.72

Table 8: Study of NR-t$\bar{t}$HH for the MCHM$_{14}$ points in Table 6. The cross sections $\sigma_{\not{hh}}$ and σ_{Yuk} are obtained by disregarding the diagrams shown on Table 7 and $\sigma^{t\bar{t}HH}_{NR}$ is the total NR-t$\bar{t}$HH cross section. The cross section σ^{SM}_{Yuk} means the SM trilinear Higgs coupling is turned off. The top Yukawas in the MCHM and SM are y_t and y_t^{SM} respectively.

SM, and measurements with precision are needed to disentangle them. The NR-t$\bar{t}$HH is the dominant channel if the resonances are heavy, which is proving to be the case with the current experiments. Effects of Higgs compositeness will benefit from more luminosity of the HL-LHC and also from higher energy and luminosity of future machines, unless a resonance is seen directly, which could happen for some of the points in the parameter space of these models.

To access these BSM effects experimentally, if they prove to be small as some of the parameter regions of these models show, a combination of the t$\bar{t}$H, t$\bar{t}$HH and double Higgs processes is necessary. In particular, double Higgs can give complementary information

to $t\bar{t}HH$ since both processes depend on the triple Higgs coupling, and, while double Higgs has a bigger cross section, $t\bar{t}HH$ has a cleaner signal. More on this is discussed in Section 4.2

3.5 Chapter Conclusions and Remarks

An implementation of the Minimal Composite Higgs Models ($MCHM_5$ and $MCHM_{14}$) in event generators was performed, in energy conditions of future machines such as the HL and HE-LHC. A parton level phenomenological analysis is done, and the properties of the allowed parameter space of the two models are explored. The implemented simulation framework is ready to be interfaced with more sophisticated simulations procedures, such as parton showers and detector simulations. It is also possible to implement such a production chain in computer grids in such a way that simulations of order of million of events can be done in a faster way. This was already implemented for the simple Singlet and Doublet models for running in the CMS grid, and it is straightforward to apply this for these models in a similar manner. This is to be explored in future work.

The $t\bar{t}H$ and $t\bar{t}HH$ processes were studied as they are greatly affected by the Higgs Compositeness and vector-like fermion properties. The production of $t\bar{t}H$ behaves in a similar way as in the simple models of the last chapter, but additional effects of compositeness, controlled by the f parameter were explored. While the kinetic distributions of $t\bar{t}H$ are not affected by these models, the $MCHM_{14}$ brings an interesting feature, already expected and observed in simulation, of increasing the top-Higgs Yukawa coupling, thus also increasing the $t\bar{t}H$ cross section. The $MCHM_5$, on the other hand, behaves like the Singlet and Doublet models and provides only $\mu(t\bar{t}H) < 1$. Current measurements are already putting strong constraints in this process, and an accuracy of 5% (Sections 2.5, 2.6 and 2.7 of Ref. [1]) is expected to be achieved already at the HL-LHC.

The $t\bar{t}HH$ process, on the other hand, has all the modifications that $t\bar{t}H$ also goes through, but also has a direct dependence on the new vector-like resonances. If they are light enough, this channel can be used for searching the resonances directly. If they are heavy, the non resonant-$t\bar{t}HH$ production process becomes more important or even dominates. This process, like $t\bar{t}H$ has a dependence on the top Yukawa, but also depends, with about 15% of the cross section, on channels controlled by the triple Higgs coupling. Its production also depends on a new vertex, $tthh$, at tree level, but it is small when compared to the other diagrams. This process can be used for accessing information on the triple Higgs coupling if measurements are precise on the 10% level.

The $t\bar{t}HH$ production process can be important for the MCHM searches. An increase or decrease in cross section is predicted by both MCHM models. The cross section at 14 TeV center of mass energy is ≈ 1 fb at tree level (see Section 3.1 of Ref. [1]), while $t\bar{t}H$ is 500 times higher (section 2.2 of Ref. [1]). The HL-LHC can be used to find a first evidence for this process, or even a discovery if there is a strong deviation from the SM. Higher energy machines will be a key to study this process, which can be used to further explore Higgs compositeness, or even the simple vector-like fermion modes of Chapter 2. This motivates, together with the fact that this process has not been searched for experimentaly before, a dedicated experimental search and is the subject of Chapter 4.

Chapter 4: Experimental search for the $t\bar{t}HH$ production at the CMS experiment at the LHC at CERN

The search for the production of a top quark-antiquark pair associated to a pair of Higgs bosons is performed, for the first time, with real data taken by the CMS experiment at the Large Hadron Collider at CERN in 2017, during the so-called Run 2 period (which includes 3 years of LHC running from 2016 to 2018). This work is fully implemented in the CMS overall analysis framework. It is indeed involved in the overall Higgs Physics Analysis Group (PAG) of CMS and directly connected to the $t\bar{t}H(b\bar{b})$ analysis Working Group. This is one of the Higgs PAG subgroups. The candidate $t\bar{t}HH$ events are selected with criteria optimizing both the signal extraction while keeping as much as possible the produced events. For this reason, this experimental search looks for events signatures coming from the semi-leptonic decays (i.e. one top decay into b plus leptonic decay of the associated W and the other top decay into b plus hadronic decay from the corresponding W) of the top quark antiquark pair and the associated double Higgs each decaying into a pair of b-quarks. The final signature this search is looking for is thus characterized by events with one lepton (either an electron or muon are considered here) a large number of b-tagged jets (6 maximum), additional jets from light quarks and a relatively small amount of total missing transverse energy (essentially due to the neutrino from the one W decaying in leptonic mode). The interesting feature of these $t\bar{t}HH$ events are the 2 additional tops, when comparing to the double Higgs production process (especially when looking to the 4 b-quark decays of the Higgs pair produced) and the additional 2 b-quarks due to the second Higgs when comparing with the $t\bar{t}H(b\bar{b})$ case. The analysis described here will emphasize these assets.

The section 4.1 presents the overall experimental framework in which this search is taking place, i.e. the LHC machine, the CMS experiment with the main detector systems of interest for this work, the overall data processing and analysis chain which are the essential tools to perform the presented here data analysis. The Section 4.2 stresses the main Physics motivations for launching this new research line. The Section 4.3 presents the data and Monte Carlo simulated data samples used for this research work. Section 4.4 summarizes the event selection especially mentioning the main differences with the one used for the $t\bar{t}H$ analysis [83]. Section 4.5 focuses on the analysis strategy. It details

the multivariate analysis performed using Deep Neural Network tools and gives the final event classification and sensitivity. Section 4.6 presents the final result after a detailed study of the systematic uncertainties.

The short and long term perspectives opened by this experimental work are presented in Section 4.7. Final remarks are made in Section 4.8.

4.1 The CMS experiment at the LHC as the experimental framework for a first exploration of the $t\bar{t}HH$ process

4.1.1 The Large Hadron Collider pp collider, LHC, at CERN

The Large Hadron Collider (LHC) is a particle accelerator built in the tunnel of the Large Electron-Positron Collider (LEP) accelerator at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. This collider accelerates two proton beams in opposite directions along a 27 km ring. It is the largest particle accelerator in the world. When protons are accelerated, the parton carries a fraction of the total proton-proton momentum, leading to interactions with a variable center of mass energy.

Before arriving at the LHC, the protons pass through a succession of machines, as seen in Fig.15 of the CERN accelerator complex. The proton acceleration starts at the Linac 2, achieving an energy of 5 MeV. The beam then goes to the Proton Synchrotron Booster (PSB), that boosts the protons up to 1.4 GeV. Then the protons go to the Proton Synchrotron (PS), raising the energy of the beam to 25 GeV, and are afterwards sent to the Super Proton Synchrotron (SPS), where they achieve the energy of 450 GeV, and are finally injected into the LHC, where the beam of protons gets the current LHC energy per proton beam of 6.5 TeV. Each of these beams are running in opposite direction to each other in different pipes of the accelerator ring. They thus achieve collisions at 13 TeV center of mass energy. The nominal center of mass energy of the LHC is at 14 TeV, which will be reached in the forthcoming Run period (Run 3) as seen in Fig. 16. The circular motion of the protons in the LHC at 13 TeV is possible thanks to 1232 superconducting dipole magnets that produce an 8 T magnetic field. For the HL-LHC, new quadrupole magnets, up to a 12 T magnetic field (compared to the current 8 T in the LHC), will be installed close to either side of the ATLAS and CMS experiments. They are made of a superconducting intermetallic compound of niobium and tin (Nb3Sn), used for the first time.

There are 4 experiments installed around the LHC ring. Two of them are multi purpose large experiments (ATLAS and CMS). The two others are dedicated to specific physics

domains: LHCb for b-physics and ALICE for heavy ion physics. Fig. 15 shows the four collision sites where the experiments are installed.

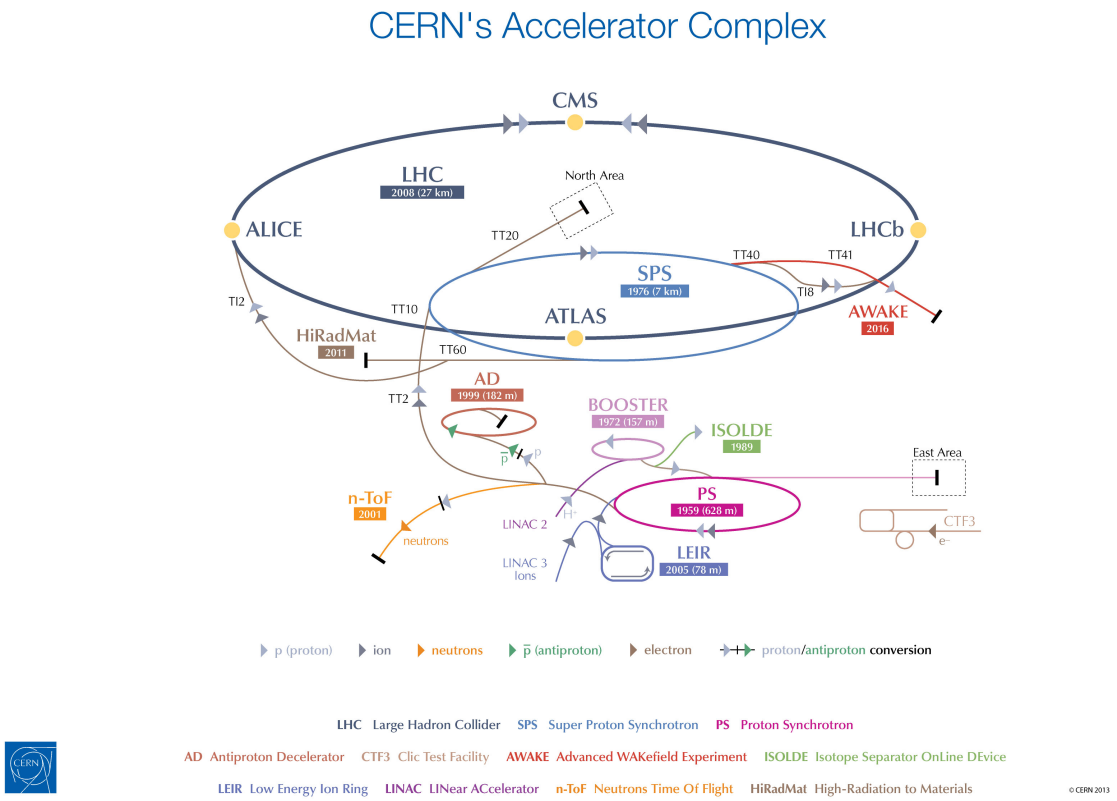


Figure 15: The CERN accelerator complex [84].

The beam of protons are made of about 2500 bunches each containing 1.15×10^{11} protons and separated by 25 ns; this corresponds to a frequency of 40 MHz. Collisions happen at each bunch crossing, where several proton-proton collisions happen simultaneously. These simultaneous collisions at each bunch crossing is denominated Pile Up (PU). Added to the Physics events produced in the collisions, are superimposed a number of simultaneous soft collisions. During the 2017 data taking, there was on average 32 interactions produced per bunch crossing. Fig.17 shows the mean number of interactions per crossing, with a proton-proton inelastic cross section of 69.2 mb, which is the recommended value for CMS for that year, and is used on the $t\bar{t}HH$ analysis of this Thesis.

A key parameter of a particle collider is its instantaneous luminosity, $L^{i}_{\texttt{inst}}$. The luminosity is directly linked to the total rate of events for any physics process by multiplying the luminosity by the cross section of this process. The instantaneous luminosity for a single colliding bunch can be expressed as a function of beam parameters. If there is no crossing angle between the beams and the beams collide with zero relative separation

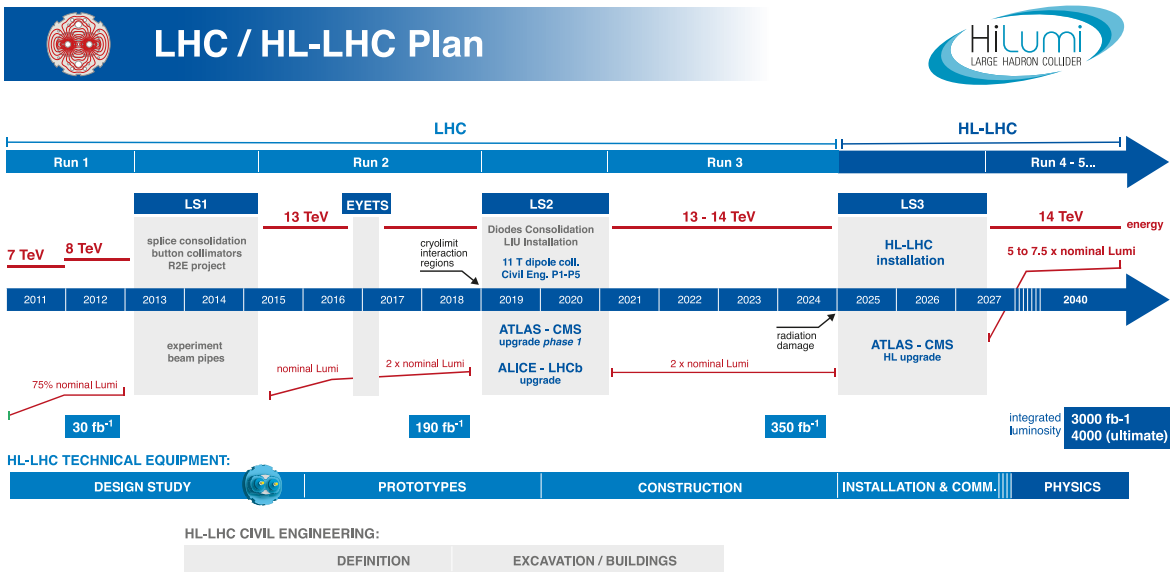


Figure 16: The plan for the LHC and High Luminosity LHC (HL-LHC). It shows the activities of the machine for each year, showing the energy and luminosity during the running time. On the bottom it is also shown the HL-LHC technical equipment phases starting from the LHC Run 1 period.

("head-on"), the instantaneous luminosity is given by the formula:

$$\mathcal{L}^i_{\texttt{inst}} = N^i_1 N^i_2 f \int \rho_1(x,y)\rho_2(x,y)\texttt{dxdy} \ , \tag{4.1}$$

where N^i_1 and N^i_2 are the number of protons for each of the two individual beams of the bunch i, and $f = 11246$ Hz is the LHC orbit frequency, and ρ_j is the normalized particle density for the bunch in beam j [85]. Therefore the instantaneous luminosity corresponds to the potential number of collisions per second. At the end of 2017 Run, the instantaneous luminosity record was smashed, reaching 2.06×10^{34} cm^{-2} s^{-1}, i.e. twice the nominal value. The summary of 2017 luminosity delivered by LHC and recorded by CMS is presented in Fig. 18.

4.1.2 The CMS experiment: the detector and the key detector systems for this search

The Compact Muon Solenoid (CMS) [86] is a multi purpose experiment located in one of the four collision points of LHC at CERN, 100 meters below the surface of the earth. It is built around a 3.8 T superconducting solenoid magnet, which takes a cylindrical form around the beam pipe comprising of a barrel and two endcaps (disks closing the barrel region) providing a 4π coverage of the collisions. This magnet is a unique case and

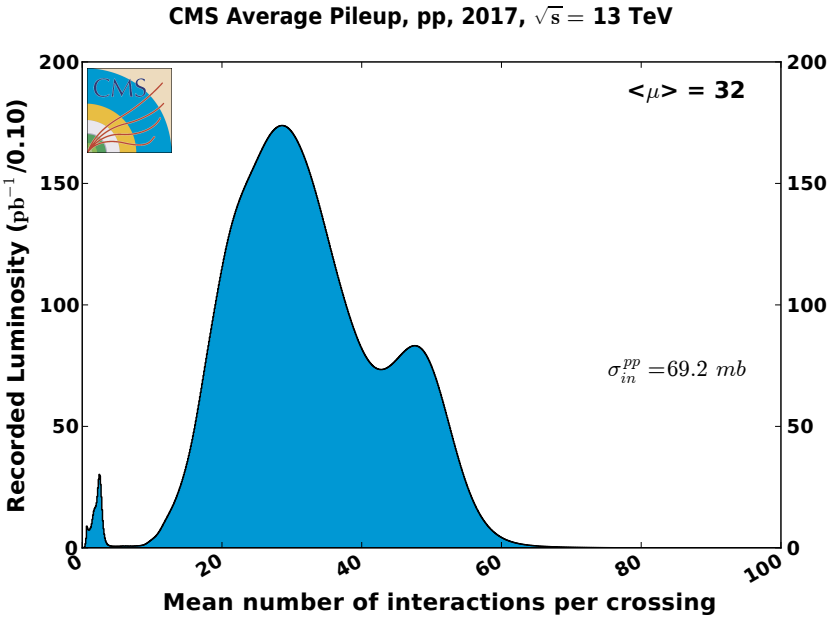


Figure 17: Mean number of interactions per bunch crossing for the 2017 pp run at 13 TeV. The plot uses the "CMS recommended" value of 69.2 mb pp inelastic cross section, which is determined by finding the best agreement with data and is recommended for CMS analyses.

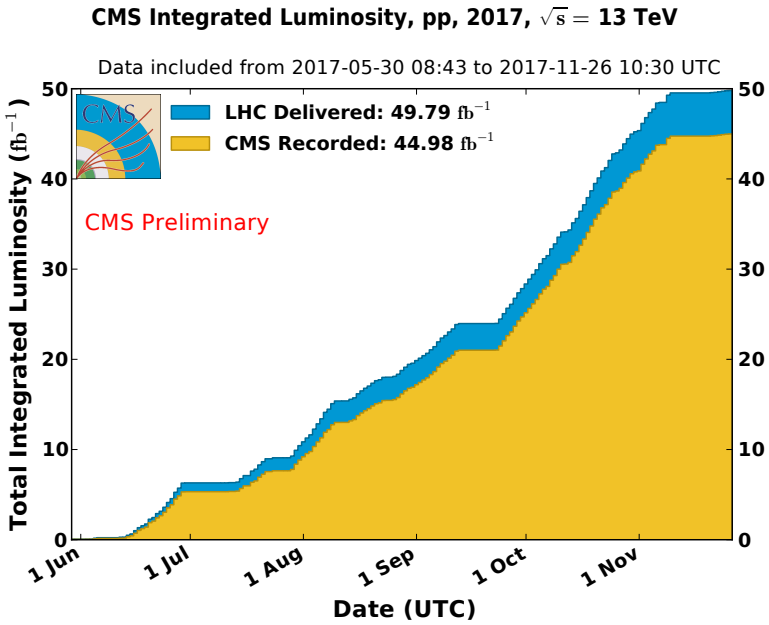


Figure 18: Cumulative measured luminosity versus day delivered. This plot uses the approved physics calibration for the 2017 pp Run [85].

innovative aspect of this detector: such a high magnetic field (3.8T) built with so large dimensions; it is the largest superconducting magnet ever built, weighs 12,000 tonnes and 7 m diameter. It is filled with multiple layers of very compact detector systems. Fig. 19 [87] shows an overview of the CMS detector and its main components, while Fig. 20 [88] shows the path of different kinds of particle in the detector system.

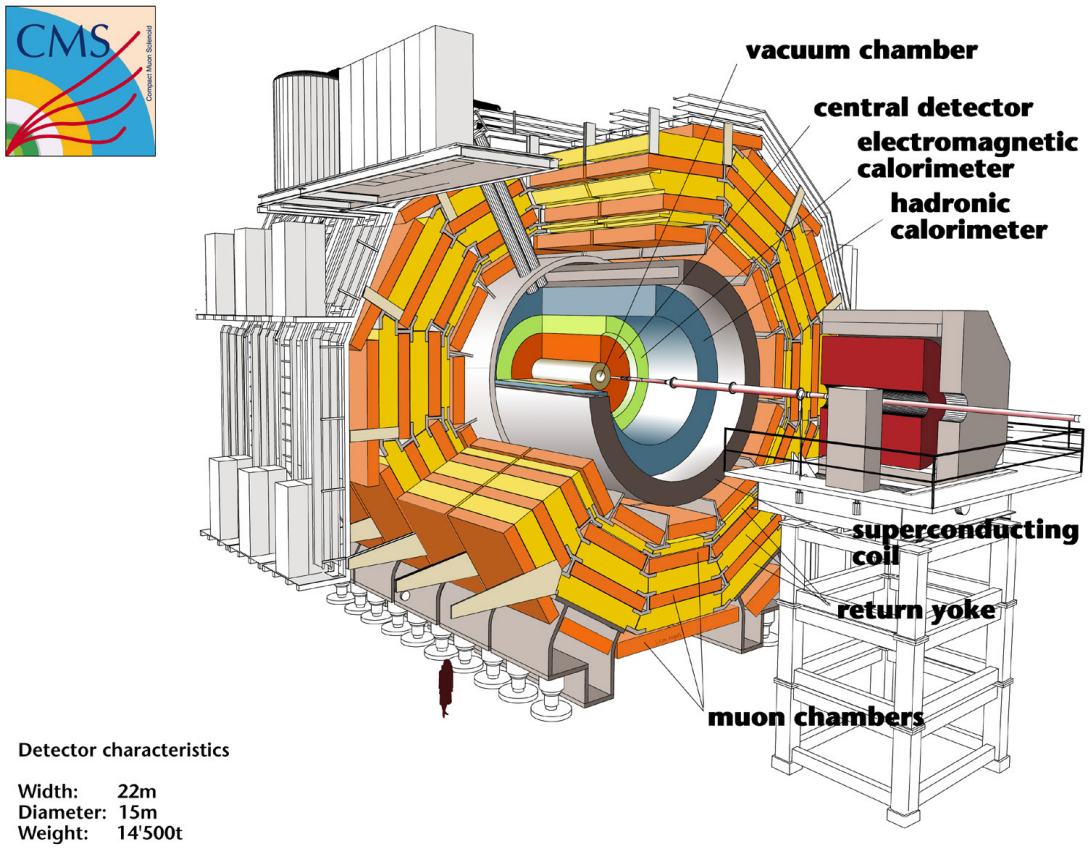


Figure 19: Overview of the CMS detector showing the main detector systems. The inner tracker, HCAL and ECAL are inside the superconducting coil and the muon detector chambers are interlayered between iron return yokes [87].

The inner part is filled with the overall tracking system for detecting and measuring the tracks (paths through this detector) of all the charged particles produced in the considered collision. This is a whole silicon (Si) based tracking system composed of Si pixels in its innermost part (microvertex) and of Si strips in the outer part of the tracking system. It is the first time that such an overall Silicon tracking system has been built. It successfully operated since the start of the LHC. It covers a total surface of about 200 m^2 and goes down to 2.4 in pseudorapidity (η). The η coverage of the tracking system, defines the acceptance in η of a few crucial Physics objects used in this search. Thanks to its high granularity and the High bending field magnet, it allows both an excellent spatial of about 20 to 40 μm and momentum resolution of around 1 to 10% [89, 90]. It is indeed a crucial detector element in this reported work.

It is followed by the Electromagnetic Calorimeter (ECAL) system made of almost 80,000 crystals of lead tungstate (PbWO4), a material with high density that produces scintillation light in fast, small, well-defined photon showers. The ECAL system is made of two parts, the central barrel, covering a region in $|\eta|$ up to 1.48. It is extended by two end caps calorimeters (Backward and Forward) that complete the η range covering of the

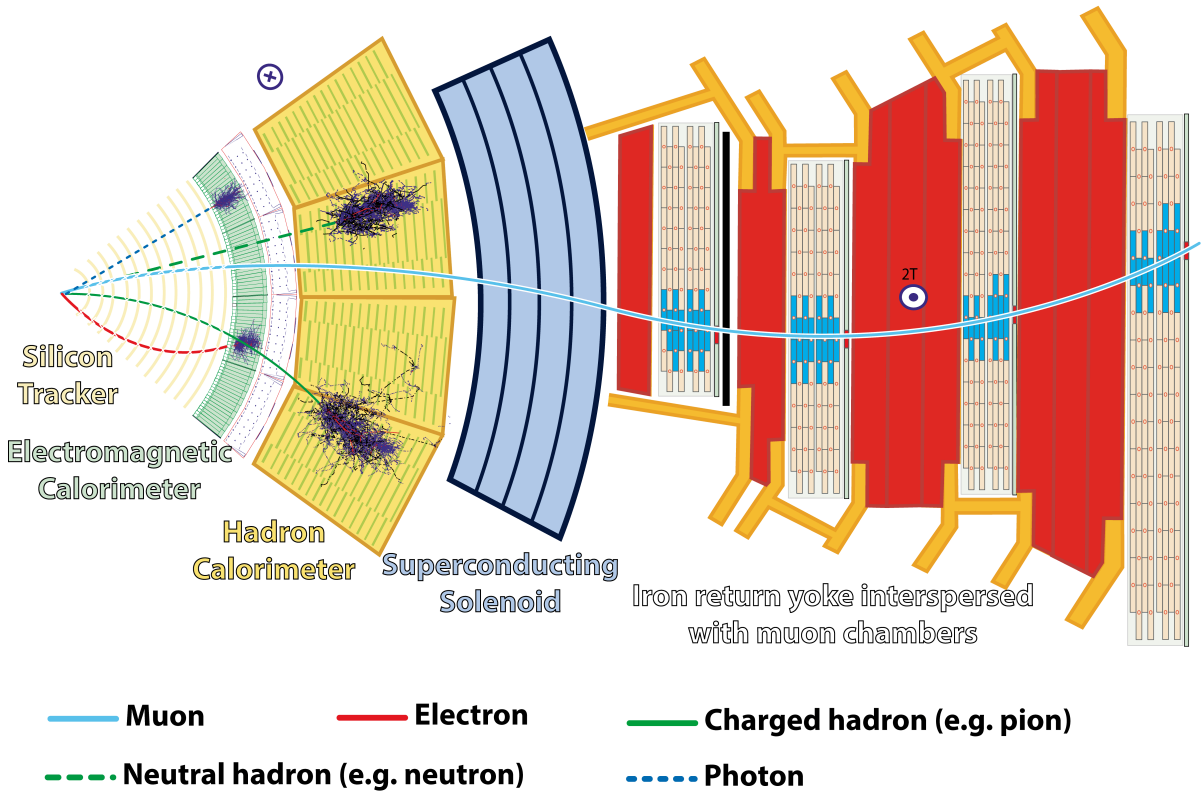


Figure 20: A transverse slice of the CMS detector showing the paths of different kinds of particles through the different sub detectors [88].

barrel down to $|\eta|$ of 3.0. The ECAL crystals have a projective geometry, pointing to the interaction point. They are surrounded by a preshower detector, in the endcaps, that takes the form of a disk made of two layers of lead and silicon sensors. This is a compact (thanks to the crystals), highly granular, hermetic and fast response detector (i.e. with a signal processing able to cope with the 40 MHz machine frequency). It is with the muon detector and the hadron calorimetry a key component of the Level-1 CMS trigger. It is another essential detector in this search for the $t\bar{t}HH$ events [91].

The ECAL is surrounded by the second main component of the calorimetry, i.e. the hadronic calorimeter (HCAL) [92] which aims to identify and measure the energy of the hadrons produced in the interactions (these are mainly pions and kaons). The HCAL includes four distinct subsystems: the barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. HB and HE, which are inside the cryostat of the CMS superconducting solenoid are sampling calorimeters where the absorber material is brass and the active material is plastic scintillator. The sampling fraction is about 7%. HB covers the range in η from zero to approximately 1.4; the η range between 1.3 and 1.4 is shared by both HB and HE; HE covers η from 1.3 to 3.0 and is made of brass disks, interleaved with scintillator wedges which cover 20 degrees in ϕ, which in turn are divided in four 5 degree

sectors. Because of the space constraint within the magnet cryostat, the HB thickness is limited to 5.8 hadronic interaction lengths at $\eta = 0$ and increases to 10 interaction lengths at $\eta = 1.2$. To catch the energy leakage from HB, layers of scintillators are placed outside the solenoid cryostat: they constitute HO. About 5% of all hadrons above 100 GeV deposit energy in HO. The calorimeter tower segmentation in η and ϕ of HB, HE and HO subsystems is 0.087×0.087 except in HE for η above 1.74, where the η segmentation ranges from 0.09 to 0.35 and the ϕ segmentation is 0.175. The HF segmentation is 0.175×0.175 except for η above 4.7, where the segmentation is 0.175×0.35. These tower segmentation constitute are essential pieces to build the Level 1 trigger. The hadronic calorimeter is another essential detector system in this analysis especially for the jets determination (particle flow) and missing transverse energy (hermeticity) component.

Located outside of the Solenoid magnet, four muon stations are placed between the iron return yoke plates of the magnet system. The barrel region stations consist of layers of aluminum drift tubes, while endcap stations contain cathode strip chambers. Stations of both regions also contain resistive plate chambers (RPC) [93].

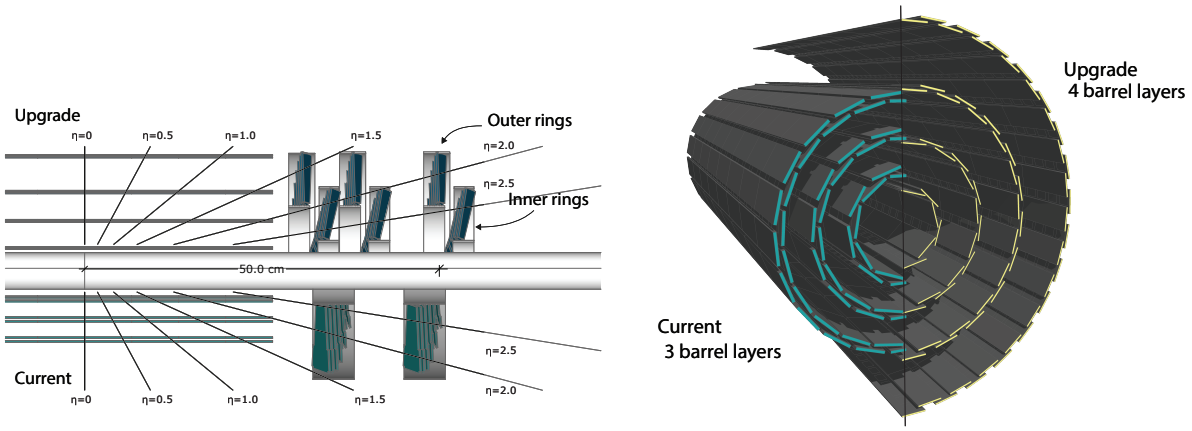


Figure 21: Layout of the different layers and disks in the old (current in the image) and current (upgrade in the image) pixel detectors. It shows the new added pixel barrel layer and the extra third endcap disk [90]

The innermost part of the tracking system, close to the interaction point, deserves special attention for this Thesis, where the final signatures include 6 b-tagged jets. An important remark here is that this search, based on the 2017 Run, has benefited from an upgraded microvertex detector with a $4^{\mathtt{th}}$ additional layer in the central part and an additional third disk on each End Cap side of this device (Fig. 21). Besides, the pipeline was changed and the first innermost layer of the microvertex was set closer to the beam pipe. On the off-line side an improved b-tagging algorithm was developed accordingly. The inner tracker provides a high charged particle momentum resolution. The B-Hadron has a relatively large lifetime of about 1.5 ps, corresponding to a flight distance of 450 μm that can be observed with the inner tracker microvertex and its high spatial resolution

provided by the very high granularity (Si pixels), thus allowing the secondary vertex reconstruction. The secondary vertex combined with the tracks of all charged particles in the B-Hadron decay are the ingredients for building the b-tagging algorithm , one of the key components of the $t\bar{t}HH$ analysis [94, 95].

The signatures of the events that are searched for here are characterized by a single electron or muon from one of the top decay, jets from light quarks from the other top decay, and b-tagged jets (from the double Higgs decays and the two tops decay) as well as a relatively small amount of transverse missing energy. This is why all the different detector systems listed above are of major importance for this search.

To complete this brief introduction to the CMS detector, major tool to perform the search reported in this Thesis, it is important to remind here some basics about the used coordinates system. This part here below, is directly borrowed from Refs. [96, 97].

The coordinate system of CMS is defined with the z-axis along the beam pipe with an angle θ defined relative to the beam axis, with $\theta = 0$ along the axis. An azimuthal angle, ϕ, is defined around the beam axis (z), and is thus invariant under Lorentz boosts along z. Different from ϕ, θ is not Lorentz invariant along the direction of the beam. It is then used, instead of the angle θ, the rapidity (y), which has the property that the difference of rapidity of two particles is invariant under boosts along z.

Given a particle with momentum $\mathbf{p}$ and energy E, and p_z being the component of the momentum along the beam axis, the rapidity is defined (see e.g. Section 47 of [96] or Chapter 11 of [97]), in natural units, as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} \ . \tag{4.2}$$

For a highly relativistic particle, i.e. $|\mathbf{p}| >> m$, where m is the mass of the particle, and reminding that in this coordinate system $p_z/|\mathbf{p}| = \cos\theta$, the rapidity can be written as

$$y \simeq -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \ . \tag{4.3}$$

Now, the right hand side of Eq.(4.3) is precisely the definition of pseudorapidity, η. Rapidity is useful to define distance measures that are invariant under Lorentz boots along the z direction such as $\Delta\mathtt{R} = \sqrt{(\Delta y)^2 + (\Delta\phi)^2}$. In the LHC, where the above approximation applies, this means that distances measures can be written as $\Delta\mathtt{R} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, i.e. entirely in terms of angles, proving to be a good coordinate system to work with LHC collisions in the center of mass frame.

4.1.3 Trigger system, Data flow, CMS software and analysis framework

A two levels only (new) Trigger [98] (and references therein) architecture, summarized in Fig. 22, is built to select a large scope of different physics topics (Flavour Physics, QCD, EW, BSM...), that this multipurpose experiment allows exploring with detailed associated measurements, out of the one billion interactions per second produced at each bunch crossing at the LHC at 40MHz. The Level 1 trigger (L1) [99] is the first stage, and uses hardware to select or reject events based on information from the overall calorimetry (ECAL and HCAL) and from the overall Muon detector system, and its architecture is summarized in Fig. 23. At this stage the event rate is reduced to 100kHz (from 40MHz). An event that passes this L1 trigger goes to the next level of triggering, the High Level Trigger (HLT), which is software based and performs a more sophisticated analysis of the event, already using track information. After this stage, the event rate is reduced to 1 kHz and is stored for later offline analysis. In the present analysis, only single lepton are taken (electron or muon trigger). Fig. 24 shows the data flow from the raw data, provided directly from the detector, passing through the L1 and HLT triggers. The next step is the offline computing, which is also software based and is where data are reconstructed and prepared for the last step (Users), the data analysis.

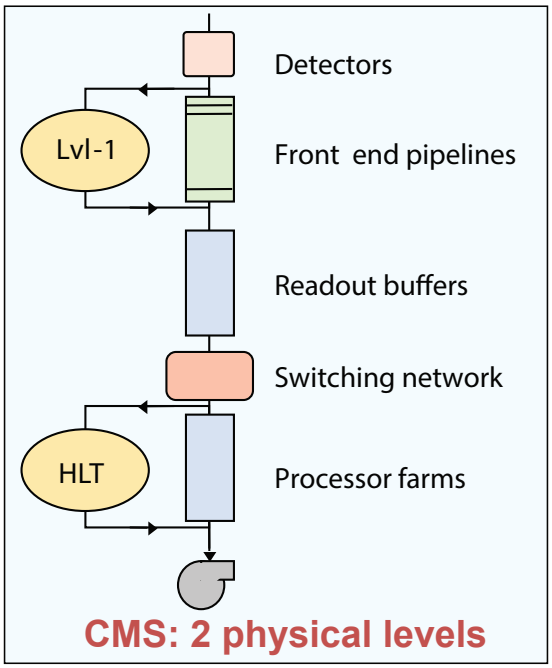


Figure 22: Architecture of the CMS trigger, showing the path events take starting from the detector, passing through the level 1 trigger (L1), the high level trigger (HLT), after which the offline analysis is made. Taken from Ref. [100].

The CMS Computing Model is organized in a grid of computers that function as a single coherent system. The user submits jobs to the grid, and the processing can occur

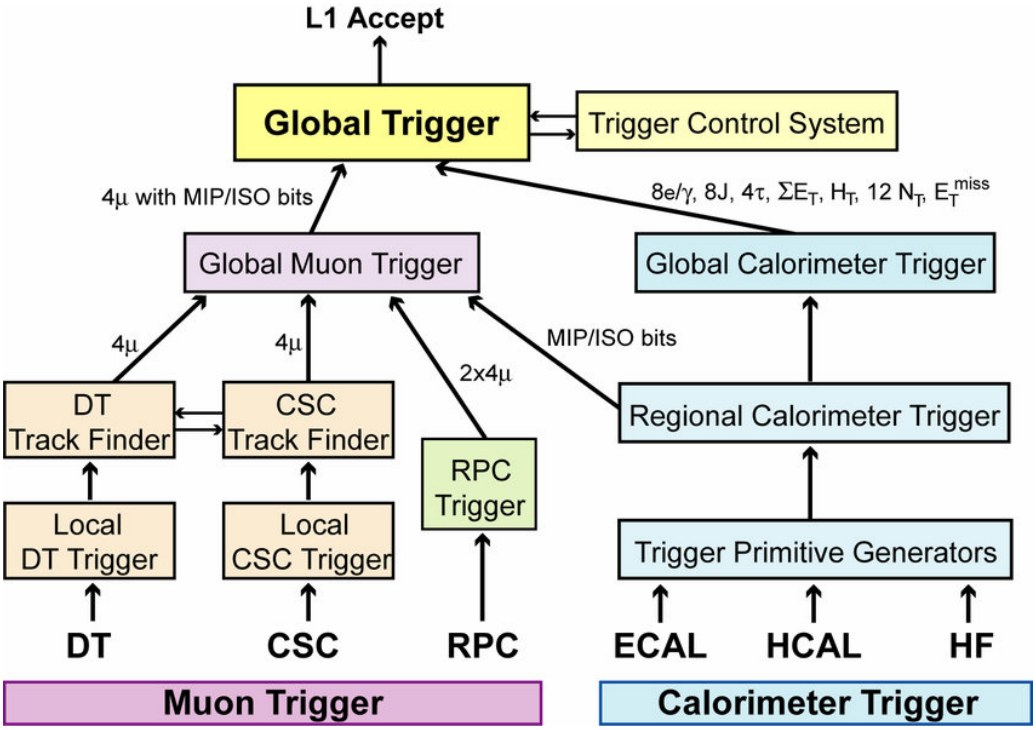


Figure 23: Architecture of the CMS L1 trigger, hardware based, that receives information from the muon and calorimeter detector systems. The signals are stored in pipeline memories in the front end electronics. The accepted events have a rate of 100 kHz out of the 40 MHz collision event rate. Taken form Ref. [86].

in any machine around the world, thus not overloading a single machine. This procedure makes possible a lot of jobs to run in parallel, increasing computing efficiency. The offline analysis is done within the CMS computing environment using the CMS Software (CMSSW). Many of the CMSSW components are publicly accessible in GitHub (see Ref. [101]), and consists of over a thousand sub-packages written in C++ or Python, which can be easily accessed and modified. This extensive toolkit is used by users to do their tasks in CMS. Also integrated to CMSSW is the Root package [102], which contains additional useful tools.

With the CMS computer grid and the tools described above, the offline analysis proceeds to further process the data as it comes out of the detectors. As illustrated in Fig. 25, the offline computing starts with RAW data, as provided directly by the detector. The reconstruction (RECO) data contain a number of physics objects created in an event, such as tracks, vertices and jets. Analysis object data (AOD) files have a reduced number of physics objects, suiting the needs of data analysis (it was used by analyses groups during Run 1). MiniAOD is further reduced in content with a more condensed number of physic objects, and is directly used by the physics analyses groups during the Run 2 period, and is currently on way to be substituted by an even smaller file format, the NanoAOD, containing flat ntuples and is used in some analyses [103, 104].

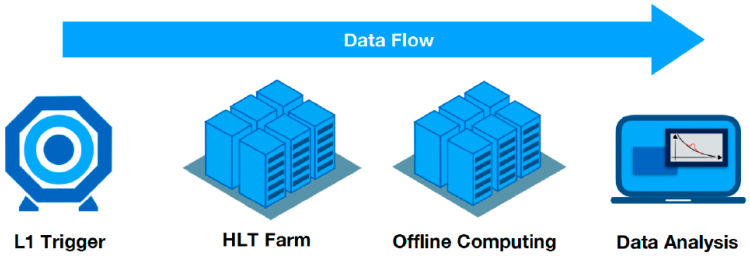


Figure 24: Cartoon schema and the detailed data flow in CMS from raw data (as provided directly from the detectors) up to the final stage of data analysis, passing through triggers, offline computing (data archiving, reconstruction, and extraction of analysis object data (AOD) files), and, ending on the data analysis. Taken from Ref. [100].

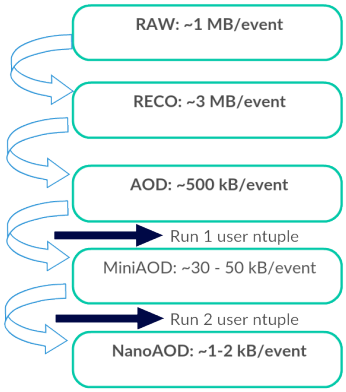


Figure 25: Summary of the path of data files, along different steps of offline computing, including file sizes per event, from RAW data to analysis object data (AOD) files, which are further used by physics groups to perform analyses [103, 104].

The final analysis, while already possible with samples in the mini-AOD format, is normally further processed with tools developed within each analysis group. In this step, further cleaning and processing can be done in order to store only what is specifically needed. The result of this is called an ntuple, greatly reduced in size when compared to the mini-AOD files, which also reduces the computing time. The ntuples can directly be compared with the nano-AOD files in terms of file format (flat ntuples).

The CMSSW framework also is used to apply to MC simulated events, following all the data processing that was described here for real data. After generation of events and the full detector simulation, implemented within CMSSW modules, the steps of reconstruction and creation of AOD files is done in the same way as with real data. Further MiniAOD, NanoAOD or ntupling processing is also done in the same way. This means that the tools developed for performing a data analysis, which is in fact first prepared and applied to MC simulated events, can then be directly applied to real data files.

4.2 Physics motivations for the $t\bar{t}HH$ search

The phenomenological work reported in Chapters 2 and 3 first initiated and focused on the $t\bar{t}H$ production process in pp collisions has in addition stressed the interest of including the $t\bar{t}HH$ production process in this exploration of the Top Higgs Yukawa sector. Introducing vector-like fermions brings a change in the top Yukawa coupling value, leading to $y_{\text{top}}/y_{\text{top}}^{\text{SM}} < 1$ in general. Different scenarios can arise when introducing Higgs compositeness, as discussed in Chapter 3. The MCHM_5 only provides a decrease in the top Yukawa, while MCHM_{14} can provide either an increase or decrease. These changes affect the $t\bar{t}H$ production only in its total production rate, and can be parameterized by the change on the Yukawa coupling only. The $t\bar{t}HH$, on the other hand, is influenced not only by these top Yukawa effects, but also by the production of new fermionic resonances. A deviation from the $t\bar{t}HH$ SM predictions can bring important information to BSM searches, and complements deviations that can be observed in the $t\bar{t}H$ searches. The LHC experiments, ATLAS and CMS after the discovery of the Higgs boson are dedicating over the last years a lot of work to explore in details the Higgs properties. With the increase in luminosity at 13 TeV they undertake a vast research program on the $t\bar{t}H$ production (about 500 fb cross-section) and on the HH production both in the gluon fusion mechanism (about 40 fb cross section) and on the Vector Boson Fusion double Higgs production (about 1.5 fb cross-section). Motivated in part by the studies reported especially in Chapter 3, the work in the CMS experiment reported here is launching, for the first time at the LHC, the experimental search for the $t\bar{t}HH$ production, in interplay with the $t\bar{t}H$ and HH production processes.

Besides its BSM interest, the study of this process is motivated by the fact that this is a production process expected in the SM with a cross-section of order of the femtobarn. The plot in Fig. 26 shows the series of processes that are produced in pp collisions that involve the production of double Higgs with a cross section that spans from few tens of fb's to one or a few fb's. They allow accessing a new region in the exploration of the Higgs sector.

The cross-sections computed at NLO QCD [106] for $t\bar{t}HH$ and compared to the ones of VBF(HH) and tjHH are listed in the Table 9.

$\sqrt{s}$(TeV)	ZHH	WHH	VBF HH	$t\bar{t}HH$	tjHH
14	$0.359^{+1.9\%}_{-1.3\%} \pm 1.7\%$	$0.573^{+2.0\%}_{-1.4\%} \pm 1.9\%$	$1.95^{+1.1\%}_{-1.5\%} \pm 2.0\%$	$0.948^{+3.9\%}_{-13.5\%} \pm 3.2\%$	$0.0383^{+5.2\%}_{-3.3\%} \pm 4.7\%$

Table 9: Cross sections computed at NLO QCD for ZHH, WHH, VBF HH, $t\bar{t}HH$ and tjHH at 14 TeV center of mass energy.

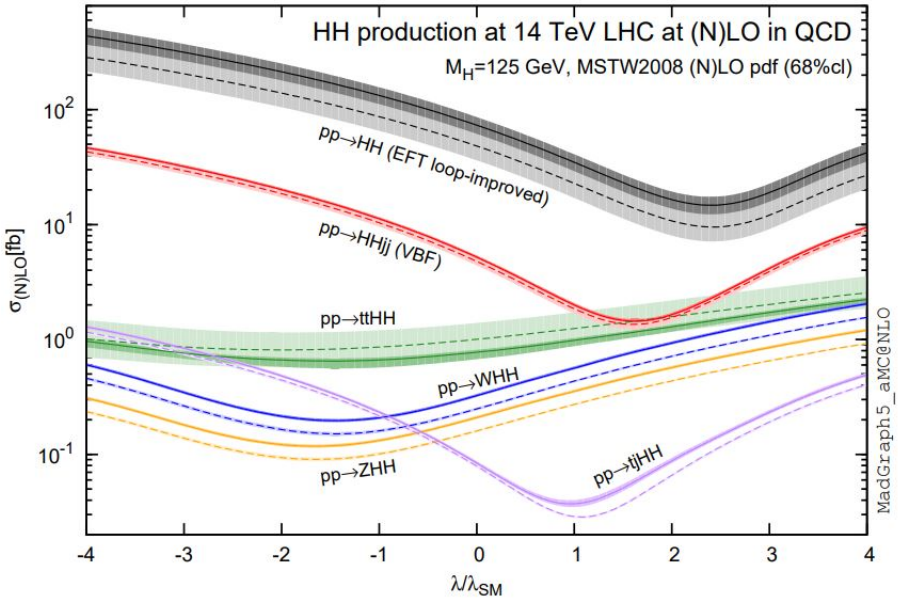


Figure 26: Total cross sections at the LO and NLO in QCD for HH production channels, as a function of the self-interaction coupling λ, taken from Ref [105]. λ is varied for the self-interaction couplings, but the mass of the Higgs boson is fixed to be $m_H = 125$ GeV. The dashed (solid) lines and light-(dark) colored bands correspond to the LO (NLO) results and to the scale of the PDF uncertainties added linearly. The SM values of the cross-section are obtained at $\lambda/\lambda_{SM} = 1$.

It is interesting to add the following $t\bar{t}H$ and $t\bar{t}HH$ production cross-sections listed here below to complete the above Table (values taken from Ref. [106]):

- $\sigma_{t\bar{t}H,13\ \mathrm{TeV}}$: 507.1 fb +5.8%-9.2% (QCD scale) ± 3.6% (PDF + α_s),
- $\sigma_{t\bar{t}H,14\ \mathrm{TeV}}$: 613.7 fb +6.0%-9.2% (QCD scale) ± 3.5% (PDF + α_s),
- $\sigma_{t\bar{t}HH,13\ \mathrm{TeV}}$: 0.775 fb +1.5%-4.3% (QCD scale) ± 3.2% (PDF + α_s),
- $\sigma_{t\bar{t}HH,14\ \mathrm{TeV}}$: 0.949 fb +1.7%-4.5% (QCD scale) ± 3.1% (PDF + α_s).

The double Higgs production through the direct process or vector boson fusion are currently under study both in ATLAS and CMS [107, 108]. The aim here is to complement these studies by searching for the double Higgs production in association with a top anti-top quark pair.

In the Standard Model, the mechanisms of production of a double Higgs production in association with a top anti-top quark pair at LO includes the Yukawa Vertex (80% of the total cross-section) and the trilinear Higgs coupling as shown in Fig. 27.

Different from the $t\bar{t}H$ process, the $t\bar{t}HH$ production process allows accessing to the triple Higgs coupling at 20% level in the total cross-section. Unlike the double Higgs production the $t\bar{t}HH$ does not include interference terms in the access to the triple Higgs

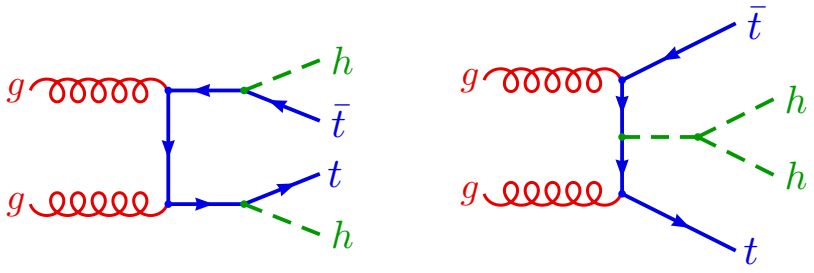


Figure 27: Representative diagrams for the $t\bar{t}HH$ process, illustrating the two distinct physical subprocesses: the Yukawa vertex and the Higgs trilinear self-coupling as expected in the SM.

coupling. These two facts emphasize the interest of the $t\bar{t}HH$ process in the SM, apart from its interest per se.

4.3 Data and simulation samples and event reconstruction

The main features of the data and simulation samples and the event reconstruction used in this analysis, similar in some way to the one developed for the $t\bar{t}H(b\bar{b})$ search [83, 109], are reviewed stressing the main differences.

4.3.1 Data and simulation samples

The data samples are those corresponding to the semi-leptonic case only in [83, 109]. These data sets collected during the 2017 LHC Run at 13 TeV are selected using trigger paths that require the presence of one electron or one muon, and correspond to a total integrated luminosity of 41.5 fb^{-1}. The MC simulation samples corresponding to the signal and to the background events are modeled according to the specific running conditions and data processing parameters for the 2017 data taking and recording. The main difference here is of course the $t\bar{t}HH$ data that are specially simulated within the Standard Model framework for this analysis. The samples are listed in Table 10 (both MC simulated signal and backgrounds). For the full Run 2 analysis (not in the scope of this thesis) and after, a $t\bar{t}ZZ$ sample is also going to be simulated.

The events are generated at next-to-leading order of perturbation theory (NLO) with POWHEG (v. 2) [110] or MadGraph 5 aMC@NLO (v.2.4.2) [39], or at LO with PYTHIA (v. 8.230) [111], depending on the process. The value of the Higgs boson mass is assumed to be 125 GeV, while the top quark mass value is set to 172.5 GeV.

The proton structure is described by the parton distribution functions (PDF) NNPDF3.1 [112]. Parton showering and hadronization are simulated with PYTHIA and the parameters for the underlying event description correspond to the CP5 tune [113], where CP stands for "CMS PYTHIA8", for all signal and background processes. This tune uses the NNPDF3.1 PDF set at NNLO and the strong coupling set to 0.118. As specified in Section 6.5 of the Ref. [113], the POWHEG generator scales the cross section by a damping function that controls the merging of perturbative matrix elements and parton showers, and that also regulates the high p_{T} radiation. The damping parameter (hdamp) in POWHEG is set to 1.379 times the mass of the top (hdamp = 237.9 GeV), and is turned on for the POWHEG samples. For comparison with the observed distributions, the events in the simulated samples are normalized to the same integrated luminosity of the data sample, according to their predicted cross sections.

All samples are reconstructed with the same CMSSW version 94X as the data samples. The Pile Up distribution in all MC samples is reweighted individually, using the standard procedure in CMS, so that the MC PU distribution matches the one expected for data.

Similarly to the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis the $\mathrm{t\bar{t}}$ background is the dominant one. A detailed study has been performed in the case of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis. This is the basis of the background study in the $\mathrm{t\bar{t}HH}$ case as well. Likewise in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis [83] the $\mathrm{t\bar{t}}$ + jets QCD background is divided into different classes based on the flavor of the additional jets that do not come from the top quark decays in the event. The flavour of those additional jets is determined using a tool in CMSSW [1] that identifies heavy-flavor (bottom, charm) jets at generator level and finds originating partons to which they correspond. The considered generator-level jets have $p_{\mathrm{T}} > 20$ GeV and abs(η) < 2.4 and are categorized as:

- $\mathrm{t\bar{t}}$ + $\mathrm{b\bar{b}}$: the event has at least two extra bottom jets, each of which originates from one or more overlapping b hadrons,
- $\mathrm{t\bar{t}}$ + b: the event has only one extra bottom jet which originates from a single b hadron,
- $\mathrm{t\bar{t}}$ + 2b: the event has only one extra bottom jet which originates from two or more overlapping b hadrons,
- $\mathrm{t\bar{t}}$ + $\mathrm{c\bar{c}}$: the event has at least one extra charm jet which originates from one or more overlapping c hadrons,
- $\mathrm{t\bar{t}}$ + LF[2]: the event does not belong to any of the above classes.

[1] This tool, called "GenHFHadronMatcher" can be found in GitHub in Ref.[114]
[2] Light Flavor

This study takes into account the $t\bar{t}$ + jets background, which is by far the dominant one because of the relatively small signal we are looking for. The single top background process, although not as important as $t\bar{t}$ is also taken into account. The $t\bar{t}$ background samples used for this analysis are listed in Table 10.

A special case is the contribution of the $t\bar{t}$ + 4b's component of the $t\bar{t}$ background. A dedicated work is launched and underway to estimate as precisely as possible this background. The status of the ongoing work is documented in Ref. [115]. As a first step, a MC simulation of $t\bar{t}$ +4b's at LO is being produced and will be used in the full Run 2 analysis. Beyond this first step, a theoretical description of this process at NLO is needed, in a similar manner to the dedicated $t\bar{t}$ +$b\bar{b}$ theoretical study at NLO performed [116] for the $t\bar{t}H$ case and with a full implementation on the MC generators. The full Run 2 $t\bar{t}H(b\bar{b})$ analysis is in fact using a dedicated $t\bar{t}+b\bar{b}$ sample that is merged with the overall $t\bar{t}$ sample for a better modeling of the $t\bar{t}$ +jets background [117, 118].

The additional 4b's are playing a similar role to the 2b's in the $t\bar{t}H(b\bar{b})$ case. The current available MC samples of $t\bar{t}$ + jets have in fact a few $t\bar{t}$ + 4b's events. However they represent a too small sample, statistically speaking. This means that this sample can not be considered as an independent sample. This was checked by using the GenHFHadronMatcher and implementing new $t\bar{t}$ categories and redefining $t\bar{t}+b\bar{b}$ as:

- $t\bar{t}+b\bar{b}$: the event has exactly two extra bottom jets, each of which originates from one or more overlapping b hadrons,
- $t\bar{t}+$bbb: the event has more than two extra bottom jets, each of which originates from one or more overlapping b hadrons (excluding the case of exactly 4 extra b's),
- $t\bar{t}+$4b: the event has exactly four extra bottom jets, each of which originates from one or more overlapping b hadrons.

With this implementation, and with the dedicated semi-leptonic $t\bar{t}$ sample used for this analysis (see Table 10), about 200 $t\bar{t}+$4b events get selected for the analysis (event selection is described further in Section 4.4), out of about 110 million total simulated events. When taking the dedicated $t\bar{t}+b\bar{b}$ MC sample, simulated at NLO and used for the current $t\bar{t}H(b\bar{b})$ analysis, this number rises to about 500 selected events, coming from about 5 million total simulated events. The small statistics of $t\bar{t}+$4b events presently available in the $t\bar{t}$ samples at disposal does not provide a good description of this background component as well as it does not allow getting well enough defined shapes. This is not at all critical in the present stage of this analysis but will be taken into consideration already with the full Run 2 data analysis. The full Run 2 analysis will benefit from the dedicated $t\bar{t}+$4b MC simulation at LO being produced [115]. A theoretical description of this process at

NLO will be important already for Run 3 and will benefit not only this analysis, but also analyses that are in the few tens of fb's range such as VBF(HH), HH+jets, ZHH, WHH, $t\bar{t}H$ and 4 tops, as well as a number of BSM processes.

For the current analysis, these $t\bar{t}$ + 4b's events fall into the $t\bar{t}+b\bar{b}$ background category listed above, containing $t\bar{t}$ plus two or more b-jets originating from one or more b hadrons and no distinction is done between events of this category.

Note that the smaller backgrounds as those taken into account for the $t\bar{t}H$ analysis (W, Z+jets, ttW, $t\bar{t}Z$, Diboson) are not considered at this stage in this analysis because there is still too small statistics.

Sample type	Number of MC events
$t\bar{t}HH$, $HH \to b\bar{b}b\bar{b}$ (all $t\bar{t}$ channels)	9,800,000
$t\bar{t}H$, $H \to b\bar{b}$ (all $t\bar{t}$ channels)	8,000,000
$t\bar{t}$ + jets (Semileptonic channel)	110,014,744
$t\bar{t}$ + jets (Semileptonic channel) (dedicated for DNN training)	43,732,445
$t\bar{t}$ + jets (Hadronic channel)	130,262,440
$t\bar{t}$ + jets (Dileptonic channel, i.e. electron/muon)	66,979,742
Single top (including single antitop, all channels)	35,313,395

Table 10: Signal and Background Simulation samples used in the analysis.

4.3.2 Data samples corresponding to specific High Level Triggers and Run periods

This analysis considers only the semi-leptonic decay of the top pair thus only the single lepton triggers are of interest here. The events are selected based on single lepton Level-1 triggers, which require either one electron or one muon. This translates at the HLT into the samples of events contained in the HLT files listed in Table 11. The single-lepton trigger performance in MC simulation has been adjusted based on the performance in data. The overall final trigger efficiency in data and MC are evaluated, and a scale factor (SF), which is the ratio of efficiency in data to that in MC, is applied to MC events in single-electron and single-muon event selection.

4.3.3 Event reconstruction

The software used for the ntupling, the baseline selection and for selecting the single lepton events are the same used in the $t\bar{t}H(b\bar{b})$ analysis [83, 109]. The MC simulation and data samples are the same than for the $t\bar{t}H(b\bar{b})$ case, apart from the new $t\bar{t}HH$ MC

Channel	Trigger Name	Run2017 Era
μ	HLT_IsoMu24_eta2p1_v*	B–D
	HLT_IsoMu27_v*	B–F
e	HLT_Ele35_WPTight_Gsf_v*	B–F
	HLT_Ele28_eta2p1_WPTight_Gsf_HT150_v*	B–F

Table 11: List of Files that translate the High Level Trigger (HLT) selection to which correspond these data (refined single electron and single muon triggers) with the corresponding certified Run periods.

simulation sample, requested at LO in the same conditions of the other samples, i.e. same MadGraph 5 and PYTHIA 8 versions, with the inclusion of the CP5 tune. The CMSSW version 9.4.10 is used for the ntuple processing.

All events are reconstructed using the particle-flow (PF) technique [119], which increases performance by identifying and reconstructing individually the particle candidates produced in the pp collision with the combined information provided by all the relevant detectors of the CMS experiment. Event cleaning is performed by requiring that data and MC events contain at least one primary vertex (PV) [120] passing the selection here below, with additional event cleaning using missing transverse energy (MET) filters:

- the number of degrees of freedom used to find the PV must be larger than 4,
- the absolute value of the z-coordinate (i.e. along the beamline direction starting from the detector centre) of the PV must be smaller than 24 cm,
- the absolute value of the ρ-coordinate (i.e. the distance from the beamline in the transverse plane) of the PV must be smaller than 2 cm.

Effects from additional pp interactions in the same bunch crossings due the Pile-up events are modelled in the simulation by adding simulated minimum-bias events to all the simulated processes. This PU modeling does not exactly describe the real PU present in the real data. The differences in the distribution of reconstructed primary vertices in MC is corrected by re-weighting simulated events to match the PU distribution in data (see Fig.17).

Electrons and muons are classified into three types. One definition is designated for the single lepton channel while the other two define leading and sub-leading leptons in the di-lepton channel. The latter is also applied in the semileptonic channel in order to veto extra leptons. It is worth reminding that this analysis is done only on the single lepton channel.

Both lepton candidates are required to pass the cuts on the kinematic variables of p_{T} and η and they have to be sufficiently isolated from nearby jet activity, following the respective relative isolation ($I_{\mathtt{rel}}$) definition for muons and electrons within a cone of

$\Delta R = 0.4$ for muons and 0.3 for electrons (see [83, 109]). The isolation is defined as follows. For each lepton, within the respective ΔR cone, excluding the contribution from the lepton candidate it is computed the sum of p_T of all particle candidates inside that cone and that is consistent with the PV selection described above. The neutral component from PU interactions is subtracted event-by-event. The relative isolation $I_{\texttt{rel}}$ is then defined as the ratio of this sum to the p_T of the lepton candidate and for the electrons, the required value for this discriminant is $I_{\texttt{rel}} < 0.06$. Furthermore, quality criteria [121] are used to reject misidentified electrons or electrons from identified photon conversions. Finally the electron candidates in the transition region between the barrel and endcap calorimeters, $1.4442 < |\eta| < 1.5560$, are excluded. Muon candidates are required to fulfil $I_{\texttt{rel}} < 0.15$ in the single-lepton channels and $I_{\texttt{rel}} < 0.25$ in the dilepton channel (this analysis, even if in the single lepton channel, uses this dilepton information to veto events). Additional quality criteria [122] are also applied to the case of the muons in order to increase purity and suppress misidentification. These isolation criteria are the same as the ones applied in the $t\bar{t}H(b\bar{b})$ analysis with 2017 collected data and can be found in Section 3 of Ref.[83]. Tables 12 and 13 summarize the muon and electron identification requirements.

	muon ID for single muon channel	leading muon ID for dilepton channel	sub-leading muon ID for dilepton channel
p_T [GeV] $>$	29	25	15
$\lvert\eta\rvert <$	2.4	2.4	2.4
$I_{\texttt{rel}} <$	0.15	0.25	0.25

Table 12: Summary of muon identification (ID) requirements in the single muon channel as well as for the leading and sub-leading muon (ordered by p_T) for the dilepton channel.

	electron ID for single electron channel	leading electron ID for dilepton channel	sub-leading electron ID for dilepton channel
p_T [GeV] $>$	30	25	15
$\lvert\eta\rvert <$	2.4	2.4	2.4
$I_{\texttt{rel}} <$	0.06	0.06	0.06

Table 13: Summary of electron identification (ID) requirements in the single electron channel as well as for the leading and sub-leading electron (ordered by p_T) for the dilepton channel.

Dedicated scale factors are applied to the MC events in order to improve the agreement with the data, following the recommendations of Particle Object Groups (POGs) in CMS. The Electron/Photon (EGamma) POG recommendations are found in Ref. [123] for electrons and the Muon POG recommendations are found in Ref. [124] for muons.

Jets are reconstructed from PF particle candidates clustered with the anti-kt algorithm [125] with the distance parameter R of 0.4, as implemented in FastJet [126] and is summarized as follows. A distance measure d_{ij} is defined between entities (particles,

pseudojets) i and j as

$$d_{ij} = d_{ji} = \mathtt{min}(p_{\mathrm{T}i}^{-2}, p_{\mathrm{T}j}^{-2}) \frac{\Delta \mathtt{R}_{ij}^2}{R^2} \ , \tag{4.4}$$

where $\Delta \mathtt{R}_{i,}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, with y_i and ϕ_i respectively the rapidity and azimuthal angle. The $p_{\mathrm{T}i}$ is the transverse momentum of the entity i with respect to the beam (z direction). A distance measure is also defined between every entity and the beam (B) by:

$$d_{iB} = p_{\mathrm{T}i}^2 \ . \tag{4.5}$$

The distance R in Eq.(4.4) is usually called the jet radius and determines the minimum distance between any final pair of jets. The clustering proceeds by identifying the smallest of the two distances and if the smallest is d_{ij}, it combines the entities if into one new object with momentum $p_i + p_j$, which is often called a "pseudojet" in the intermediate steps of the clustering. If the smallest distance is d_{iB}, the entity is removed from the list of particles/pseudojets and calls it an "inclusive jet". This procedure is repeated until there are no entities left. Of the final jets, only those that are above the p_{T} threshold defined below are used. Computational optimizations are also applied in this algorithm as defined in the FastJet manual of Ref. [126].

Charged hadrons tracks in the cluster which are associated with non-primary vertices are subtracted to mitigate the impact of pile-up collisions [127]. Furthermore, the jet energy is corrected for the remaining neutral-hadron pileup component in a manner similar to that used to find the energy within the lepton isolation cone [128]. Jet energy corrections (JEC), related to jet energy scale (JES) and jet energy resolution (JER) calibration, are also applied as a function of jet p_{T} and η [129] to data and simulation. All selected jets are required to satisfy $\eta < 2.4$. In addition, if any charged lepton passing the selection criteria described above is found within the distance of 0.4 in $\eta - \phi$ space from a jet, then the jet is removed from the analysis. The kinematic requirements for the jets are summarized in Table 14.

The CMS recommendations are given by the JetMET Physics Object Group (JME POG) and can be found in Ref. [130] for the jet identification criteria and in Ref. [131] for the loose jet pile up removal to suppress jets originating from PU events. The JER correction factors are found in [132]. All JEC scale factors are the same as used in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis and implemented within the same software framework [83, 109].

The deep combined secondary vertex algorithm (DeepCSV) [95] is used to identify jets that originate from b-quark decays (referred to as b-tagged jets). It provides a b-tagging discriminant by combining identified secondary vertices and track-based lifetime

	single-lepton channel	dilepton channel	
	jets	leading 2 jets	further jets
min p_{T} [GeV]	30	30	20
max $\lvert\eta\rvert$	2.4	2.4	2.4

Table 14: Jet kinematic selection requirements for the single-lepton channel and for the first two leading jets (ordered in p_{T}) and the next leading jets for the dilepton channel.

information. For this analysis, jets are defined to be b-tagged if their b-tagging discriminator is larger than 0.4941 (medium working point), in which the probability of tagging jets originating from light-flavour quarks (u, d, or s) or gluons is approximately 1%, and the corresponding efficiency for tagging jets from b (c) quarks is approximately 76% (18%) [133]. Differences in the b-tagging efficiency and mis-tag rate between data and simulated events are taken into account via scale factors that are derived separately for light-flavour and heavy flavour jets as documented in Section 8.5 of Ref. [95]. The scale factors are provided within the b Tag & Vertexing (BTV) POG and taken from Ref. [134]. Here it is used the same framework and DeepCSV version used in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis and described in [109], which lead to a better b-tagging efficiency when compared to the previous version of the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis[135]. The DeepCSV scale factors are, however, updated when compared to this $\mathrm{t\bar{t}H(b\bar{b})}$ analysis with 2017 data, and uses the latest version given by the BTV POG as described above.

4.4 Final Event selection

The search for $\mathrm{t\bar{t}HH}$ presented here considers the semi-leptonic channel of the top-antitop quark system and both Higgses decaying to a pair of bottom-antibottom quark pairs. This final state corresponding to this analysis is illustrated in Fig.28 that shows the two LO representative diagrams for the $\mathrm{t\bar{t}HH}$ production process.

The software used to select this semi-leptonic decay mode for top pairs and double Higgs decay into b-quarks is the same as the one used in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis, with a baseline selection as shown in Table 15. After this selection, events are required to have a missing transverse energy of at least 20 GeV and an extra b-tagged jet is required leading to at least 3 b-tagged jets with the DeepCSV b-tagging discriminator value greater than 0.4941 (medium working point).

Table 16 shows the selected number of events used for the DNN modeling and the fitting steps of the analysis. For DNN training it is only included the $\mathrm{t\bar{t}}$ events with a semileptonic decay. For the fitting, the $\mathrm{t\bar{t}}$ events contain dedicated samples of $\mathrm{t\bar{t}}$ decaying in the semileptonic mode (584,233 events), in the hadronic mode (697 events) and in the

	SL channel
Number of leptons	1
p_{T} of leptons (e/μ) [GeV]	$> 30/29$
p_{T} of additional leptons [GeV]	< 15
$\lvert\eta\rvert$ of leptons	< 2.4
Number of jets	≥ 4
p_{T} of jets [GeV]	> 30
$\lvert\eta\rvert$ of jets	< 2.4
Number of b-tagged jets	≥ 2

Table 15: Baseline event selection criteria for selecting single-lepton (SL) events in this $\mathrm{t\bar{t}HH}$ search. The final selection for this analysis requires one additional b-tagged jet.

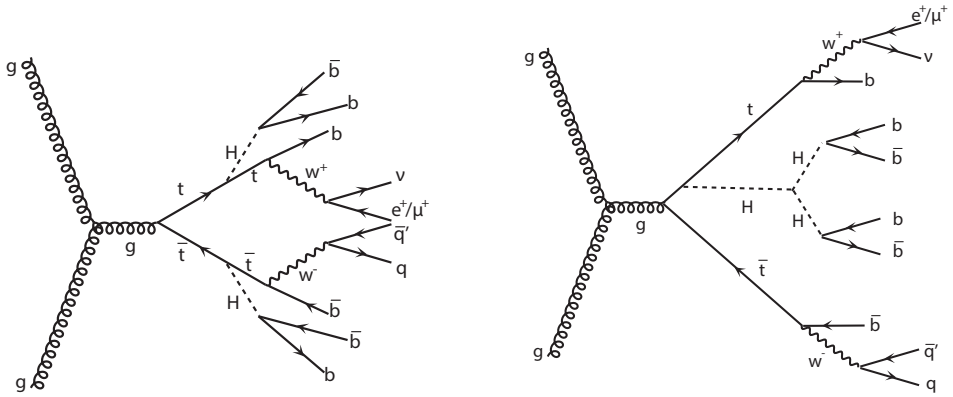


Figure 28: Representative diagrams for the $\mathrm{t\bar{t}HH}$ process in the semileptonic decay mode. It illustrates the two distinct physical subprocesses containing only Yukawa vertices (left) and containing the Higgs trilinear self-coupling (right).

	Total number of selected events for:	
Process	Fitting	DNN training
$\mathrm{t\bar{t}} + \mathrm{LF}$	490,653	143,517
$\mathrm{t\bar{t}} + \mathrm{c\bar{c}}$	130,718	37,298
$\mathrm{t\bar{t}} + \mathrm{b}$	85,849	23,417
$\mathrm{t\bar{t}} + 2\mathrm{b}$	41,800	11,057
$\mathrm{t\bar{t}} + \mathrm{b\bar{b}}$	67,780	17,021
Total $\mathrm{t\bar{t}}$	755,800	232,310
$\mathrm{t\bar{t}H}$	477,417	–
Single Top	17,332	–
$\mathrm{t\bar{t}HH}$	569,169	568,817

Table 16: Number of selected events used for fitting (see Section 4.6) (middle column). Number of selected events used for DNN modeling (see Section 4.5) (right column). The baseline selection of Table 15 is applied together with requiring ≥ 3 b-tagged jets.

dileptonic mode (170,870 events). It is clear that for the semileptonic selection applied in this analysis the full hadronic decay of $t\bar{t}$ is negligible, while the dileptonic decay events correspond to about 20% of the total $t\bar{t}$ events. These events are further weighted (cross section weights, DeepCSV weights, among other described in the previous section) to find the final event yields corresponding to 41.5 fb^{-1} luminosity to be used for fitting.

It is also worth to note that requesting higher number of b-tagged jets and higher DeepCSV b-tagging discriminator values did not improve the selection because too few events would pass the analysis selections.

4.5 Analysis strategy

The strategy followed in this analysis is based on Deep Neural Network tools allowing for an optimized event classification and analysis sensitivity.

4.5.1 Description of the $t\bar{t}HH$ (4b) Analysis based on Deep Neural Network tools

The $t\bar{t}H(b\bar{b})$ analysis of 2016 data studied both the boosted decision tree (BDT) plus the matrix element method (MEM) categorization approach and compared it with a DNN based approach [136, 137, 138]. Based on this study the 2017 data $t\bar{t}H(b\bar{b})$ analysis was carried on with an even optimized DNN based architecture. Following this result, this present $t\bar{t}HH$ study is based on the DNN categorization following the event selection described in Section 4.4. A DNN is used as a classifier in order to separate signal ($t\bar{t}HH$) events and each of the five $t\bar{t}$ + jets background processes, i.e. $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, or $t\bar{t}+LF$. Each of these six categories are implemented as output nodes in the DNN modeling. Thus the DNN is used to categorize events into the most probable process (mpp) and further construct the final discriminants that will be used for the final fit, e.g. if the mpp is $t\bar{t}HH$, the event is classified as such, with a probability p_i, defined in the following paragraphs.

The DNN architecture and implementation are similar to what is used in the $t\bar{t}H$ analysis [109]. It is implemented in Keras [139] as a feedforward neural network with three hidden layers of 100 nodes each. The Exponential Linear Unit (ELU) activation function is applied to the neurons of the hidden layers. The outputs numbers of the DNN are s_i (also referred to as scores), with i going from 1 to 6 (six output nodes). Different from the hidden layers, they are normalized to unity with a "soft-max" function [140],

$f(s_i)$, so the outputs numbers of the DNN, for each event, can be written as:

$$p_i = f(s_i) = \frac{e^{s_i}}{\sum_j^6 e^{s_j}} \ , \tag{4.6}$$

which allows the probabilistic interpretation. The mpp is then defined as the category with the highest associated p_i, i.e. the category (or output node) that can be either $t\bar{t}HH$ or one of five background categories. This output value, p_i, is then used subsequently as a discriminant value.

The loss function that is minimised during the training is the categorical cross entropy, which is the same as the cross entropy when the "soft-max" activation is applied to the outputs s_i. In this model, the cross entropy (CE) loss can be written as:

$$\texttt{CE} = -\sum_i^6 t_i \log(p_i) \ , \tag{4.7}$$

where t_i is an element of the "truth vector" with one entry equal to one, corresponding to the correct classification of the event under consideration, all other entries equal to zero. Note that this training is an example of supervised learning where it is known beforehand the correct category that the event should belong to. The training is done over a maximum of 500 epochs, where each epoch corresponds to a pass over the full training data. Early stopping is included to stop the training if there is no improvement on the DNN performance after 50 epochs or if the difference between train and validation loss function exceeds 2%. Before training all events are shuffled and training is performed in batches of 5000 events (batchsize) until all events are used (one epoch). The Adam [141] optimiser, which is used to minimize the loss function, is used with an initial learning rate of 10^{-4} as implemented in Keras. The loss functions is monitored at the end of each epoch (this is shown in Fig.31 for this DNN model). For minimizing overtraining, both L2 regularisation and dropout with a percentage of 0.5 are used as implemented in Keras. These and other hyperparameters of the neural network are summarized in Table 17.

Each of the training samples of Table 16 is further split into a subset "training" for the actual training (60 %) and an independent subset "test" for immediate cross-validation and hyperparameter optimisation (20 %) as well as a further independent subset "validation" (20%), which is used to study the performance of the DNNs. These samples used for training are independent from the events used for fitting in the sense that after the DNN modeling, they are never used again for the remaining steps of the analysis.

All the input variables used by the DNN are listed in Appendix D. These variables are related to kinematic properties of individual objects, event shape and the b-tagging

	(≥ 4 jets, ≥ 3 b-tags)
nodes per hidden layer	3×100
loss function	cross-entropy
dropout percentage	0.5
L2 regularization	10^{-5}
batchsize	5000
optimizer	Adam(10^{-4})
activation function	ELU
last activation	softmax
earlystopping percentage	2%
earlystopping min epochs	50

Table 17: Summary of the hyperparameters of the neural network, implemented with their default definitions in Keras.

discriminant. A first training of the DNN is performed using all the variables listed in this appendix, but the final model contains only the 20 most important variables ranked by training weight. The performance of the model in this last case when compared to the one containing all variables is similar within 1%, but computation time is greatly reduced. Table 18 shows the top 20 variables for this analysis, ranked automatically during the DNN model construction.

The distributions of some representative input variables are shown in Figs. 29 and 30. Also seen in these distribution is the good agreement between real data and the MC expectation. These are checked for all the quantities listed in Appendix D. The variables shown in Figs. 29 and 30 show a very distinct shape between the signal and the background distributions. These differences in shape, as seen in the plot of the sum of p_{T} of all b-tagged jets (also observed for all jets, both b-tagged or not) is an effect of the two extra b jets present in $\mathrm{t\bar{t}HH}$ as compared with the $\mathrm{t\bar{t}H}$ process. This is also reflected in several kinematic quantities (e.g. higher number of jets when compared o background processes in Fig. 30), and contributes to the good performance of classification of the built DNN model.

The error bands of the plots correspond to all shape changing systematic uncertainties and are added in quadrature. These are discussed further in Section 4.6.

Even if the $\mathrm{t\bar{t}HH}$ (4b's) process includes more jets and b-tags than the $\mathrm{t\bar{t}H}$ (2b's) process, the smaller rate (cross-section) impedes to use these facts with the presently used luminosity in this analysis (2017 data only). It gives too few events in the additional explored categories. This is why, at the present stage of this analysis, the (≥ 4 jets, $\geq$ 3 b-tags) selection is chosen and gives the best performances. Other selections based on different jets multiplicity and/or number of b-tags (the same applies by changing the DeepCSV discriminant selection threshold) do not improve the performances

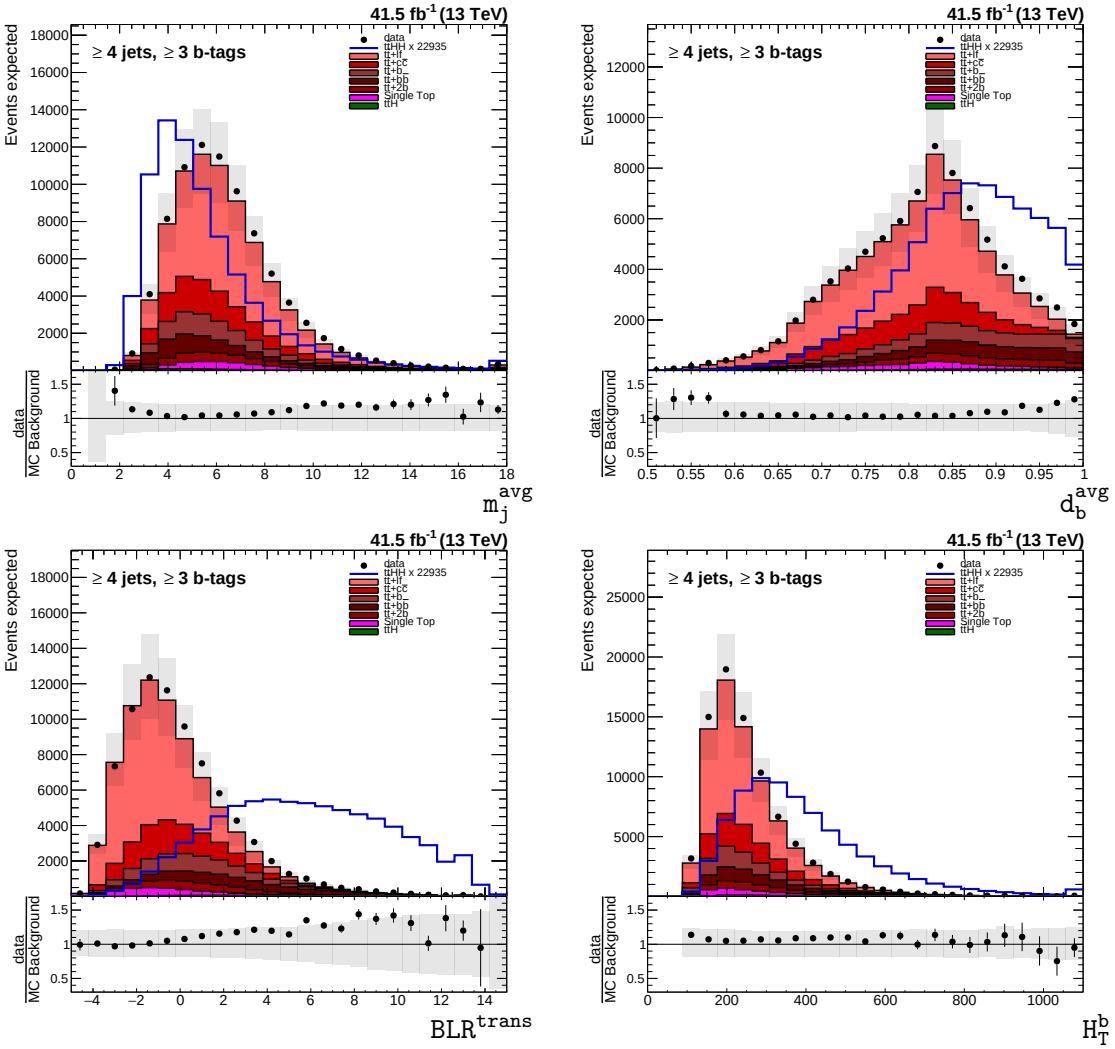


Figure 29: Data/MC comparisons of selected observables used as inputs to DNN before the fit to data. Each of the expected backgrounds is normalized to the corresponding SM cross section: the average mass of all jets (m_j^{avg}); the average b-tagging discriminant value of all b-tagged jets(d_b^{avg}); the likelihood ratio discriminating between events with 4 b quark jets and 2 b quark jets (BLR), presented as ln[BLR/(1 − BLR)] (BLR^{trans}); the scalar sum of p_T of all b-tagged jets (H_T^b); The difference in shapes between signal($t\bar{t}HH$) and background processes can be seen. Error bands correspond to all systematics and are added in quadrature. Data points are included and a ratio of the data yields and MC background expectation is shown at the bottom.

or even give worse results, thus there is no further separation of events by jet multiplicity or number of b-tags. Of course these will be options to play with when getting higher luminosity and thus higher statistics. The number of b-tagged jets passing the loose, medium or tight working points (lower to higher thresholds of DeepCSV discriminant values) are included in the DNN as input variables. The effect of adding $t\bar{t}H$ as an output node in the DNN leads to a drop in the performance of $t\bar{t}HH$ selection of about 10%. Adding $t\bar{t}H$ as a node is also subject of future studies, with higher statistics.

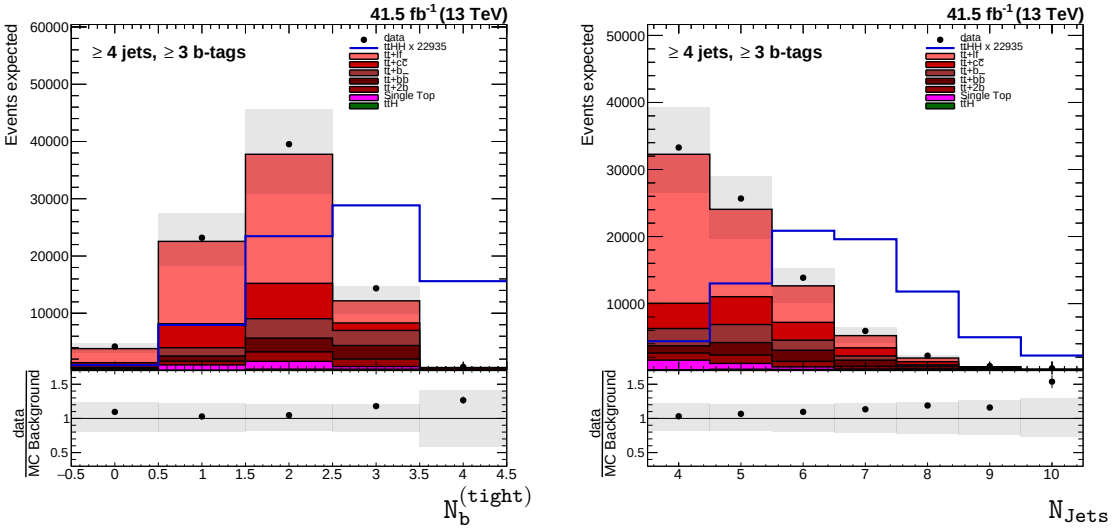


Figure 30: Data/MC comparisons of selected observables used as inputs to DNN before the fit to data. Each of the expected backgrounds is normalized to the corresponding SM cross section: the number of b-tagged jets with a DeepCSV b-tagging discriminant value higher than 0.8001, called tight working point ($N_b^{(tight)}$). Note that the selection of events for this analysis requires events with DeepCSV discriminant higher than 0.4941 (medium working point); The number of selected jets (N_{Jets}). The difference in shapes between signal($t\bar{t}HH$) and background processes can be seen. Error bands correspond to all systematics and are added in quadrature. Data points are included and a ratio of the data yields and MC background expectation is shown at the bottom.

Fig. 31 shows the final confusion matrix, where it is shown, for each event class (true class), the probability of it falling into each of the six output nodes (predicted class), remembering that it is known beforehand the label of each event. The confusion matrix shows a classification efficiency of 71.8% and is quite high thanks to the two additional b quarks when compared to $t\bar{t}H(b\bar{b})$, which has a performance of about 40% in the same conditions. The receiver operating characteristic curve (ROC) area under curve (AUC) is also calculated, and is a function of the true class labels and the predicted class labels for all the events. The closer this number is to one, the best the performance, and ROC-AUC=1 means there is no false positives and no false negatives. Fig. 31 also shows the loss function of this model, monitored at the end of each epoch, and it is already getting near its minimum within these 500 epochs after just a few epochs and no overtraining is observed (train and validation samples follow the same trend and the lines (red and blue) never cross.

Fig. 32 shows, for the subset of 20% of the events saved for testing, the events that are classified in each output node, with the probability p_i, as defined above. The expected number of events are normalized to 41.5 fb^{-1} luminosity and the signal ($t\bar{t}HH$) events are re-scaled to be visible on the plots. These distributions show the difference in shape

$\mathbf{d_b^{min}}$: minimal b-tagging discriminant value of all b-tagged jets	$\mathbf{m_j^{avg}}$: average mass of all jets
$\Delta R_{\mathbf{lep,b}}^{\mathbf{min}}$: minimal ΔR between lepton and b-tagged jet	$p_{\mathbf{T}}$ **(jet 4)**: p_{T} of the 4^{th} jet, ranked in jet p_{T}
$\Sigma_{\mathbf{j}}(\mathbf{d}-\mathbf{d_j^{avg}})^2$: squared difference between the b-tagged discriminant value of a jet and the average b-tagging discriminant value of all jets, averaged over all jets	$\mathbf{m_{b,b}^{min\Delta R}}$: mass of pair of b-tagged jets closest in ΔR
$\mathbf{N_b^{(loose)}}$: number of b-tagged jets with DeepCSV > 0.1522	**BLR**: likelihood ratio discriminating between events with 4 b-quark jets and 2 b-quark jets
$p_{\mathbf{T}}$ **(jet 1)**: p_{T} of the 1^{st} jet, ranked in jet p_{T}	$\mathbf{d_j^{avg}}$: average b-tagging discriminant value of all jets
$\mathbf{d}_3$: 3^{rd} highest b-tagging discriminant value of all jets	$(\mathbf{m}^2)_{\mathbf{b}}^{\mathbf{avg}}$: average squared mass of all b-tagged jets
$\mathbf{d_b^{avg}}$: average b-tagging discriminant value of all b-tagged jets	$\mathbf{N_b^{(medium)}}$: number of b-tagged jets with DeepCSV > 0.4941
$\Delta\eta_{\mathbf{j,j}}^{\mathbf{avg}}$: average $\Delta\eta$ between any two jets	$\mathbf{d}_4$: 4^{th} highest b-tagging discriminant value of all jets
$\mathbf{H_T^b}$: scalar sum of p_{T} of b-tagged jets	$\mathbf{BLR^{trans}}$: $\ln[\mathrm{BLR}/(1-\mathrm{BLR})]$
$\mathbf{H_T}$:scalar sum of p_{T} of all jets	$\mathbf{m_b^{avg}}$: average mass of all b-tagged jets

Table 18: Top 20 ranked Input Variables for this analysis, obtained after training the DNN with all variables and ranking them by training weight.

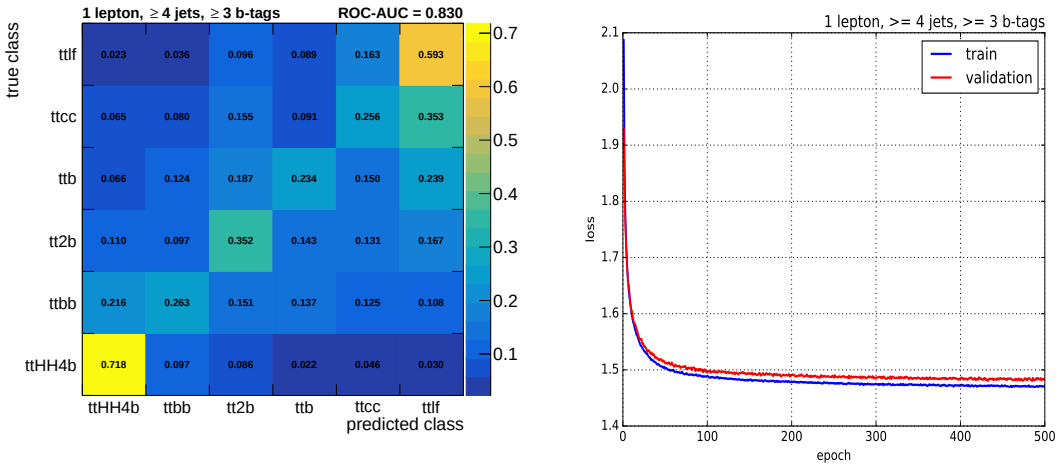


Figure 31: The confusion matrix (left) shows the percentage of true class events that are classified in each one of the six predicted classes (output nodes). A good performance corresponding to a ROC-AUC=0.830 is observed. The loss function (right) is monitored at each epoch. Both training (blue) and validation (red) curves follow the same trend and don't cross, thus no overtraining is observed. The loss function gets close to the minimum very early in the training (around the first 50 epochs).

obtained for the $t\bar{t}HH$ event distribution when compared to this specific sample of semi-leptonic $t\bar{t}$ background events used for this training. Each output node is a discriminant for this analysis.

The sub-sample of events listed in Table 16 and that are separated for Fitting, are classified in one of the output nodes, according to the constructed and trained DNN model. The output value of the DNN (p_i) for the mpp is used as the final discriminant value. The final discriminant distribution histograms are then constructed and are inputs for the final fit. In this step, several studies are made. It is checked that there are no empty bins in any histogram, and a minimum of 10 expected background events in each bin of the signal node are required. If less events than that are observed, bins are merged or bin widths are changed. For the background nodes, the same study is performed, with a threshold of 30 expected background events in each bin, at least. The shapes of the background histograms are observed to not bring a difference for the final fit, but the shapes are kept and will be studied in the future for the final Run 2 analysis and beyond.

The results of the binning choice is presented in the pre-fit plots of Fig. 33. Before presenting and discussing these results, the systematic uncertainties are presented in Section 4.6. Shape changing systematics have corresponding plots constructed with the same criteria explained in the last paragraph, as well as with the same trained DNN model. The error bands of the pre-fit distribution plots take into account all systematic uncertainties, subject of the following section.

4.6 Systematic uncertainties and Results

The implementation of the systematic in this analysis is discussed in Section 4.6.1. Afterwards in Sections 4.6.2 and 4.6.3, the fit to Asimov data and real CMS data is presented, respectively.

4.6.1 Study of systematic uncertainties

This section shows how the systematic uncertainties are included in this analysis and are described in Table 20. Each rate systematic is added as a number in the statistical tools, while for shape systematics, varied templates with up/down variation are generated. For jet energy resolution (JER) and jet energy scale (JES) systematics, the full analysis is redone starting from the event selection step, and events are fed to the already trained DNN for creating the final discriminants. The other shape systematics don't require doing the full analysis, and are included at the moment of plotting the final discriminants, taking as inputs the nominal templates and corresponding weights, event-by event, corresponding to "up" or "down" variations. The statistical MC uncertainty are included at fitting time, by the CMS fitting tools. In summary, for each output node of the histogram, there

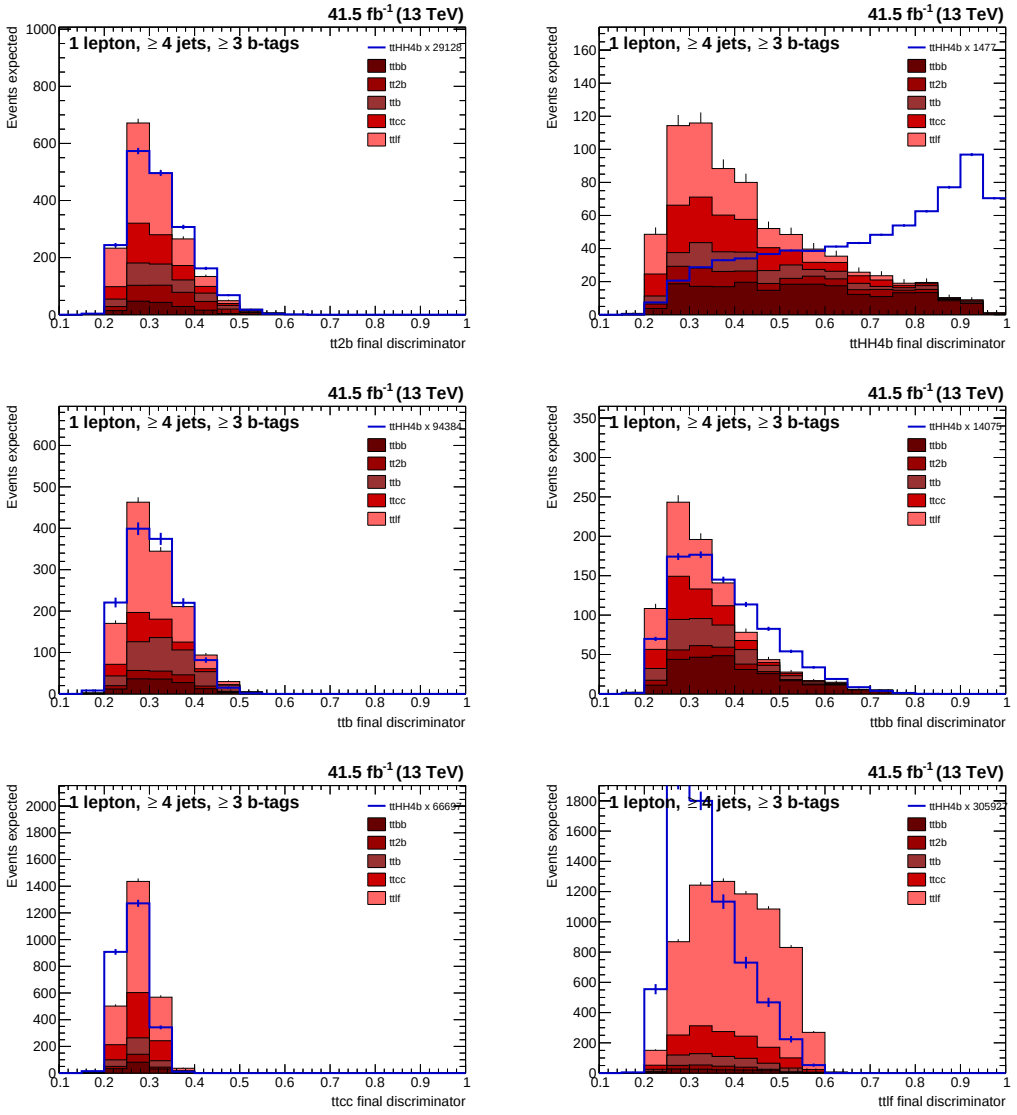


Figure 32: Comparison of the DNN discriminant distributions expected for the signal and background processes, constructed with a subset of 20% of the events listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The discriminant value shown in the horizontal axis is the probability relative to the most probable process (p_i) that each sample is classified as.

is one "nominal" plot, one set of "up" variations and one set of "down" variations, that are produced. All these histograms (about 400) are inputs for the fit to data. The rate inclusive cross section uncertainties are summarized in Table 19. The systematics uncertainties, reported in this Table for the $t\bar{t}HH$ (4b) process are similar to those for the $t\bar{t}H(b\bar{b})$ process [109], except for the addition of the ones related to $t\bar{t}HH$. We remind that the still small background processes (see Section 4.3.1), currently negligible with respect to present $t\bar{t}HH$ rate, are not included in the analysis.

Process	PDF				Renorm./fact. scales			
	$gg_{t\bar{t}H}$	$gg_{t\bar{t}HH}$	gg	qg	$t\bar{t}$	t	$t\bar{t}H$	$t\bar{t}HH$
$t\bar{t}H$	3.6%						-9.2%/+5.8%	
$t\bar{t}HH$		3.2%						-4.3%/+1.5%
$t\bar{t}$ +jets			4%		-4%/+2%			
Single Top				3%		-2%/+3%		

Table 19: Inclusive cross section (rate) uncertainties used in the analysis. They are theoretical uncertainties taken from Ref. [106]. Note that an extra 50% rate uncertainty is assigned separately to each of the four considered $t\bar{t}+$HF processes.

4.6.2 Analysis results with Asimov data

The fitting[3] is performed with the discriminators presented in Fig. 33 as inputs. They include simulated events of $t\bar{t}HH$, $t\bar{t}+$LF, $t\bar{t}+c\bar{c}$, $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}H$ as well as single top. Each event of each sample is classified into one output node of the neural network with the trained DNN model, thus constructing these discriminators. The systematics, as explained in the previous section, are also included either as rates (just a number, e.g. theoretical cross section uncertainties) or shapes, with "up"/"down" variations plotted in each output node after applying the DNN model, again for each sample. Finally, the data events also goes through the DNN model and thus classified in each of the 6 possible outcomes. These histograms are fed into the `combine` statistical tool of CMS, and the plots of Fig. 33 are obtained, with error bands containing all systematics, which include the MC statistics uncertainties (bin by bin) as well as the correlation between all such systematics. Data points are displayed on the plots and at the bottom it is plotted the ratio of the data yields divided by the yields of all MC samples. As already observed in the input variable plots, the MC simulation has good agreement with data.

Results are obtained for the signal strength modifier $\mu(t\bar{t}HH) = \sigma/\sigma_{SM}$ after a fit to data ($\hat{\mu}$) and is the outcome of maximum likelihood fits (see e.g. Ref. [143]).

[3] All the statistical analyses are performed with the Higgs `combine` tool version v8.0.1 in the CMSSW v. 10.2.13 environment. The software is available in GitHub (Ref. [142]). Details of the statistical methods contained in `combine` and used for this analysis can be found in Refs. [143, 144, 145, 146]

Source	Type	Remarks
Integrated luminosity	rate	Signal and all backgrounds
Lepton identification/isolation	shape	Signal and all backgrounds
Overall final trigger efficiency	shape	Signal and all backgrounds
Pileup	shape	Signal and all backgrounds
Jet energy scale	shape	Signal and all backgrounds
Jet energy resolution	shape	Signal and all backgrounds
b tag hf fraction	shape	Signal and all backgrounds
b tag hf stats (linear)	shape	Signal and all backgrounds
b tag hf stats (quadratic)	shape	Signal and all backgrounds
b tag lf fraction	shape	Signal and all backgrounds
b tag lf stats (linear)	shape	Signal and all backgrounds
b tag lf stats (quadratic)	shape	Signal and all backgrounds
b tag charm (linear)	shape	Signal and all backgrounds
b tag charm (quadratic)	shape	Signal and all backgrounds
Renorm./fact. scales ($t\bar{t}H$)	rate	Scale uncertainty of NLO $t\bar{t}H$ prediction
Renorm./fact. scales ($t\bar{t}$)	rate	Scale uncertainty of NNLO $t\bar{t}$ prediction
Renorm./fact. scales ($t\bar{t}HH$)	rate	Scale uncertainty of NLO $t\bar{t}HH$ prediction
Renorm./fact. scales (single top)	rate	Scale uncertainty of NLO single top prediction
$t\bar{t}+$HF cross sections	rate	Additional 50% rate uncertainty of $t\bar{t}+$HF predictions
PDF (gg)	rate	PDF uncertainty for gg initiated processes except $t\bar{t}H$ and $t\bar{t}HH$
PDF (gg $t\bar{t}H$)	rate	PDF uncertainty for $t\bar{t}H$
PDF (gg $t\bar{t}HH$)	rate	PDF uncertainty for $t\bar{t}HH$
PDF (qg)	rate	PDF uncertainty for qg initiated processes (single top)
PDF shape variations ($t\bar{t}H$, $t\bar{t}HH$, $t\bar{t}$)	shape	Based on the NNPDF replicas, same for $t\bar{t}H$, $t\bar{t}HH$ and additional jet flavours
μ_R scale ($t\bar{t}$)	shape	Renormalisation scale uncertainty of the $t\bar{t}$ ME generator (POWHEG), same for additional jet flavours
μ_F scale ($t\bar{t}$)	shape	Factorisation scale uncertainty of the $t\bar{t}$ ME generator (POWHEG), same for additional jet flavours
PS scale: ISR ($t\bar{t}$)	shape	Initial state radiation uncertainty of the PS (for $t\bar{t}$ events), same for additional jet flavours
PS scale: FSR ($t\bar{t}$)	shape	Final state radiation uncertainty of the PS (for $t\bar{t}$ events), same for additional jet flavours
Bin-by-bin event count	shape	Statistical uncertainty of the signal and background prediction due to the limited sample size

Table 20: Systematic uncertainties considered in this analysis.

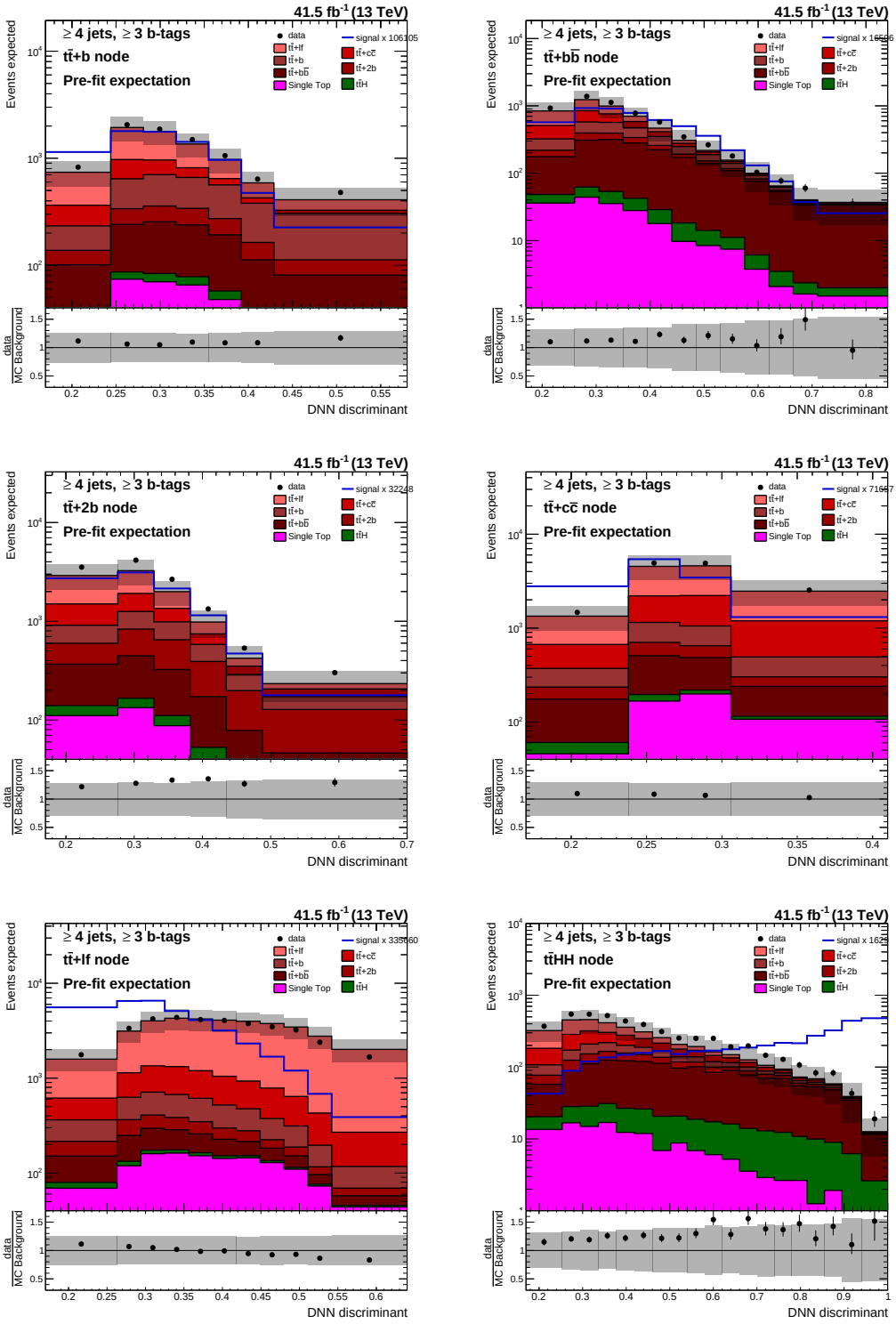


Figure 33: Discriminant shapes before the fit to data. They are constructed with signal and background processes listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The data corresponds to CMS collected data in the year of 2017.

For the first part of the analysis, a "blind analysis" is performed, i.e., an analysis with "pseudo-data", only provided by MC simulated samples, which are called Asimov data[144]. The Asimov data can be understood, up to a very good approximation, as just the sum of all signal plus background events. Although a very good approximation, in this analysis we don't use the sum of MC events as the pseudo-data, but the full blind data are generated with the `combine` tool. The pre-fit plots in this case look exactly like the plots of Fig. 33, where the only change is that data points fall exactly where the MC simulation is located, or in other words, the Asimov data/MC ratio is exactly one. All systematics (thus error bands in the plots, remain the same). The complete fit procedure is then developed and achieved with the Asimov data.

$\hat{\mu}$	Observed (Asimov)	$1^{+12.4}_{-11.2}$
95% CL upper limits on μ	Expected (Median)	25.6
	Expected (68% CL range)	[17.9,37.4]
	Expected (95% CL range)	[13.2,52.7

Table 21: Fit Results for Asimov data (blind analysis). The upper limits are obtained with asymptotic likelihood methods according to formulas of Ref [144]. Also shown is the best fit value of the signal strength modifier $\mu = \sigma/\sigma_{\texttt{SM}} = \hat{\mu}$.

Summarized in Table 21 are the observed best fit value, $\hat{\mu}$, as well as expected upper limits for μ with corresponding expected 68% and 95% confidence level (CL) ranges and are produced according to Ref. [144] and its implementation on the `combine` tool. Since the fit is done to Asimov data, expected and observed upper limits are the same, as expected, and are shown in Table 21 as "Expected(Median)". For the same reasons, the value for $\hat{\mu}$ is equal to one. The uncertainties for this analysis are dominated by statistics, but some systematic uncertainties are starting to play also a role. The pointed out systematic uncertainties will play an increasing role with the next steps of this experimental search which will include a larger number of data (increase in overall recorded luminosity, new decay channels, detector and analysis improvements). The analysis presented here is thus a starting point, and the complete systematics framework will be more relevant for these future studies involving the full Run 2 (currently being launched), the upcoming Run 3 as well as the High Luminosity LHC.

Fig. 34 shows the impact parameters and pulls sorted from the highest impact for the top 30 most impacting uncertainties. The impact of a nuisance parameter (θ) on a parameter of interest (μ) is defined as the shift $\Delta\hat{r} = \mu - \hat{\mu}$ (where $\hat{\mu}$ is the best fit value of μ) that is induced as the post fit value of $\hat{\theta}$ is fixed and brought to its $+1\sigma$ or -1σ post-fit uncertainties, with all other parameters profiled as normal. This is effectively a measure of the correlation between θ and μ, and is useful for determining which nuisance parameters have the largest effect on the parameter of interest (μ for this analysis) uncertainty.

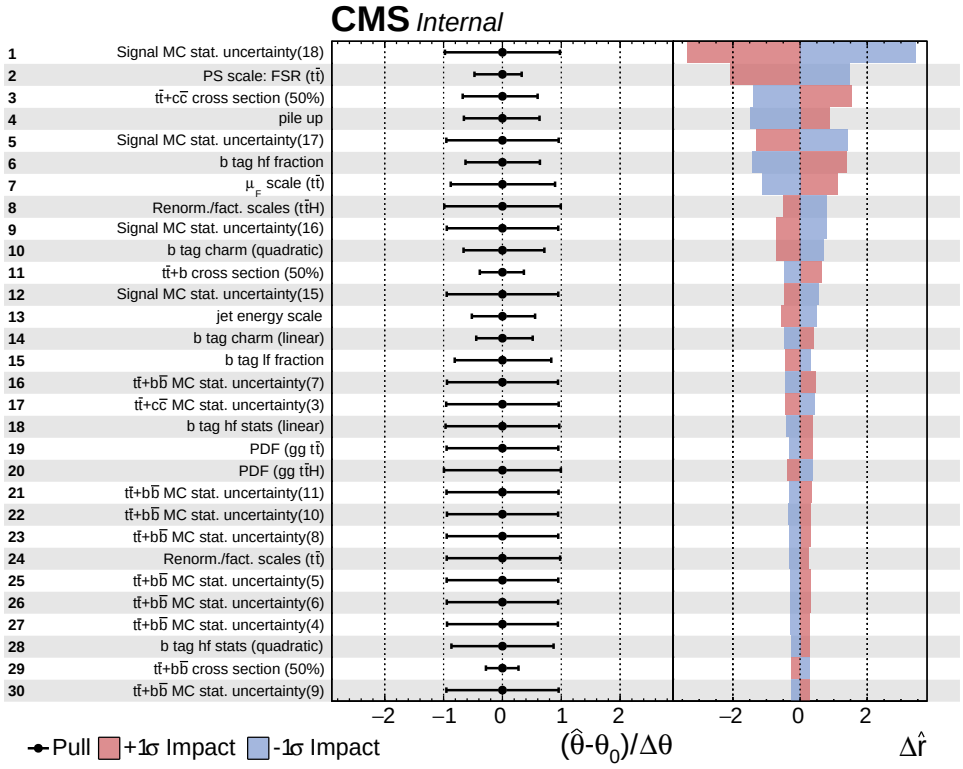


Figure 34: Nuisance parameter pulls and impacts, sorted by the largest impact, for the top 30 most impacting uncertainties, obtained from the fit to Asimov data (blind analysis), for 41.5 fb^{-1} integrated luminosity. The pulls of the nuisance parameters (black markers) are computed relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$. The impact $\Delta\hat{r}$ is computed as the difference of the nominal best fit value of μ and the best fit value obtained when varying the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus/minus (1-σ) its post-fit uncertainty (colored areas).

The pulls are defined as follows. The best fit value for each nuisance parameter ($\hat{\theta}$) and the uncertainties for all the floating parameters are computed. For all the constrained nuisance parameters a convention is used by which the nominal value θ_0 is zero, corresponding to the mean of a Gaussian constraint PDF with width 1.0, such that the parameter values ± 1.0 correspond to the $\pm\ 1\sigma$ input uncertainties. Thus in the plot of Fig. 34, one can see, marked with dots, the difference between post-fit ($\hat{\theta}$) and pre-fit values (θ_0). The size of the lines correspond to $\pm\ 1\sigma$ uncertainties (bigger uncertainties meaning smaller tighter lines). Here it is seen that these dots fall exactly on zero, as expected for Asimov data (pre-fit and post-fit values for θ are the same).

Apart from the low statistics, the most constraining systematics, as observed by looking at the pulls (i.e. the bigger $\Delta\theta$, the smaller the line range in the figure), are: b-tag charm (linear), PS scale-FSR($\text{t}\bar{\text{t}}$), additional 50 % rate uncertainty on $\text{t}\bar{\text{t}}+\text{HF}$ cross sections, and can be seen in Fig. 34. The jet energy scale systematic, although not as constraining as others, deserves special care and will be taken into account in more details in the full Run 2 analysis and beyond. The most impacting uncertainty is the MC

statistical uncertainty of the last bin (bin 18) of the $t\bar{t}HH$ node. This is expected as this is the most sensitive bin, i.e. where most of the MC simulated signal events are lying. This is also the bin with less background events, thus big uncertainties are expected and reflected in this impact parameter plot.

The blind analysis is done in order to avoid bias in the development of the analysis strategy. In this way, no optimizations are done based on real observed data. The Asimov results give thus a fully blind expected uncertainties for the analysis. After the approval by CMS, the first plots of basic kinematic quantities are produced and an agreement between data and MC simulation is checked. As already shown in Figs. 29 and 30, good agreement (within the uncertainty bands) is observed, and this was checked for all input variables relevant for this DNN based analysis.

4.6.3 Final results with real data

All the analysis procedure and corresponding checks developed for this $t\bar{t}HH$ search analysis, as performed with the Asimov data (Section 4.6.2) are presented for endorsement by the CMS HIGGS Physics Analysis Group. This public (within CMS) presentation is subject to questions and eventually further tests and work requested by the HIGGS PAG. When this stage of the analysis is endorsed, the analysis can proceed on with the "data unveiling" procedure, it is to say to apply all the developed analysis to real data to get the final results. This "blind analysis" procedure is a quite successful and nowadays the standard way to obtain results that are as less biased as possible by real data.

The pre-fit plots of Fig. 33 are then used, with the real data, as inputs to the statistical tools. In practice, everything is run the same way, by just changing the inputs of the already prepared scripts, replacing Asimov with real data, or by turning off the Asimov data generation routines. Summarized in Table 22 is the observed best fit value, $\hat{\mu}$, as well as the observed and expected upper limits for μ with corresponding expected 68% and 95% confidence level (CL) ranges. Here the expected limits are obtained with pseudodata. A fit is performed with the real data as a first step to generate the corresponding pseudodata. The nuisance parameters are set to the values obtained in the fit. This explains the small difference between the expected upper limits obtained here with real data and the ones obtained with the fully blind analysis [144].

The post-fit plots for the discriminants are shown in Fig. 35. A zoom of the signal node post-fit plot is done for the region of less than one expected event and is shown on Fig. 36, where the evolution of the signal along the bins can be observed.

$\hat{\mu}$	Observed	$4.9^{+14.4}_{-12.8}$
95% CL upper limits on μ	Observed	32.9
	Expected (Median)	28.9
	Expected (68% CL range)	[20.1, 42.2]
	Expected (95% CL range)	[14.9, 59.6]

Table 22: Fit Results for CMS data. The upper limits are obtained with asymptotic likelihood methods according to formulas of Ref [144]. Also shown is the best fit value of the signal strength modifier $\mu = \sigma/\sigma_{\texttt{SM}} = \hat{\mu}$.

The uncertainties for this analysis, based on data collected during the year 2017 with 41.5 fb^{-1} integrated luminosity is compatible with the expected ones, in both the blind and unblind analysis. Fig. 37 shows the impact parameters and pulls sorted from the highest impact for the top 30 most impacting uncertainties. All quantities are as defined for the blind analysis. Here the pulls can be observed, as expected, because the pre-fit and post-fit values for θ are not the same. The same conclusions as the ones of the blind results are valid here, with the most constraining systematics being again: b-tag charm (linear), PS scale-FSR($\mathrm{t\bar{t}}$), additional 50 % rate uncertainty on $\mathrm{t\bar{t}}+$HF cross sections, and can be seen in Fig. 37. Not shown here because lower in impact is the jet energy scale, but the pulls for this systematic uncertainty look similar to what is shown in the blind version of the impact plot. The most impacting uncertainty is again the Monte Carlo statistical uncertainty, due to limited sample size, of the last bin (bin 18) of the $\mathrm{t\bar{t}}$HH node.

4.7 Perspectives

A very rough and back of the envelope extrapolation of the $\mathrm{t\bar{t}}$(SL)HH(4b) result obtained here with 41.5 fb^{-1} is done with the CMS `combine` tool, for the full Run 2 (3 times more luminosity) and full Run 3 (10 times more luminosity) for this $\mathrm{t\bar{t}}$HH signature. This is done only with MC data, thus the expected μ is equal to one. The study in on the expected uncertainty change by increasing luminosity. The μ-values indicated in Table 23 are to be taken with a lot of caution and are surely underestimated.

A number of improvements besides the increase in luminosity will be applied already in the full Run 2 analysis for this specific signature, such as the new b-tagging, the improved $\mathrm{t\bar{t}}$-background, especially $\mathrm{t\bar{t}}+4$b and hopefully larger statistics on the $\mathrm{t\bar{t}}$ overall background, including $\mathrm{t\bar{t}}$H and also an improvement on the systematics uncertainties and how they are included.

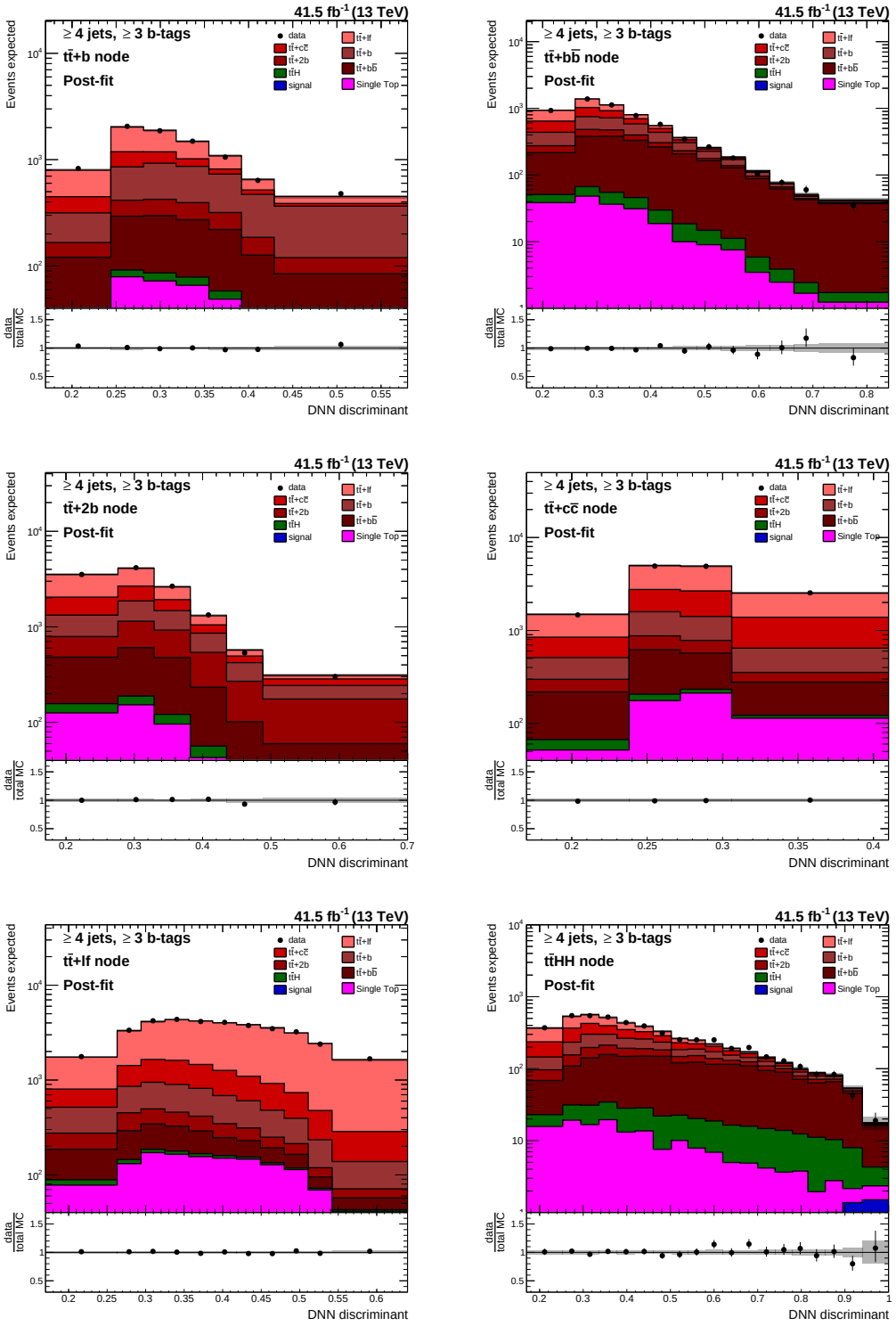


Figure 35: Discriminant shapes after the fit to data. They are constructed with signal and background processes listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The data corresponds to CMS collected data in the year of 2017.

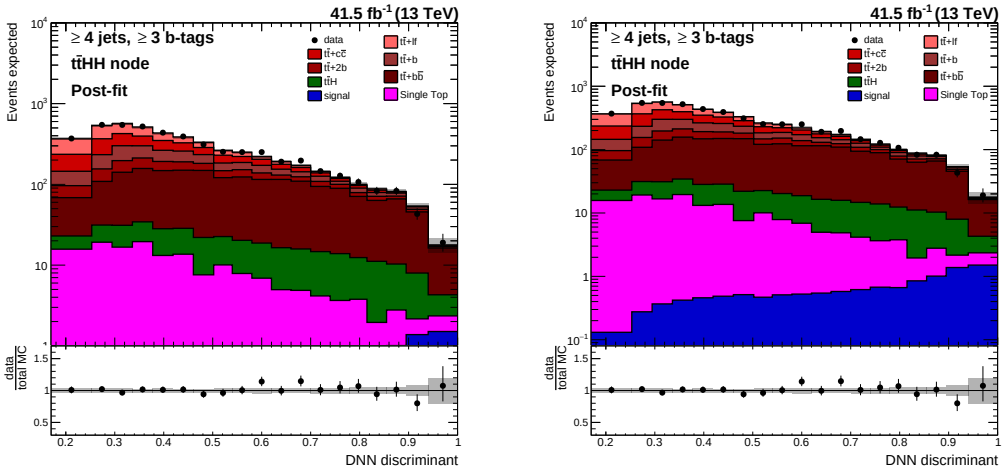


Figure 36: Discriminant shapes after the fit to data for the signal node. On the right it is shown a zoom of the plot on the left in the region of less than one expected event. The plots are constructed with signal and background processes listed in Table 16 and normalized to 41.5 fb^{-1} luminosity. The data corresponds to CMS collected data in the year of 2017.

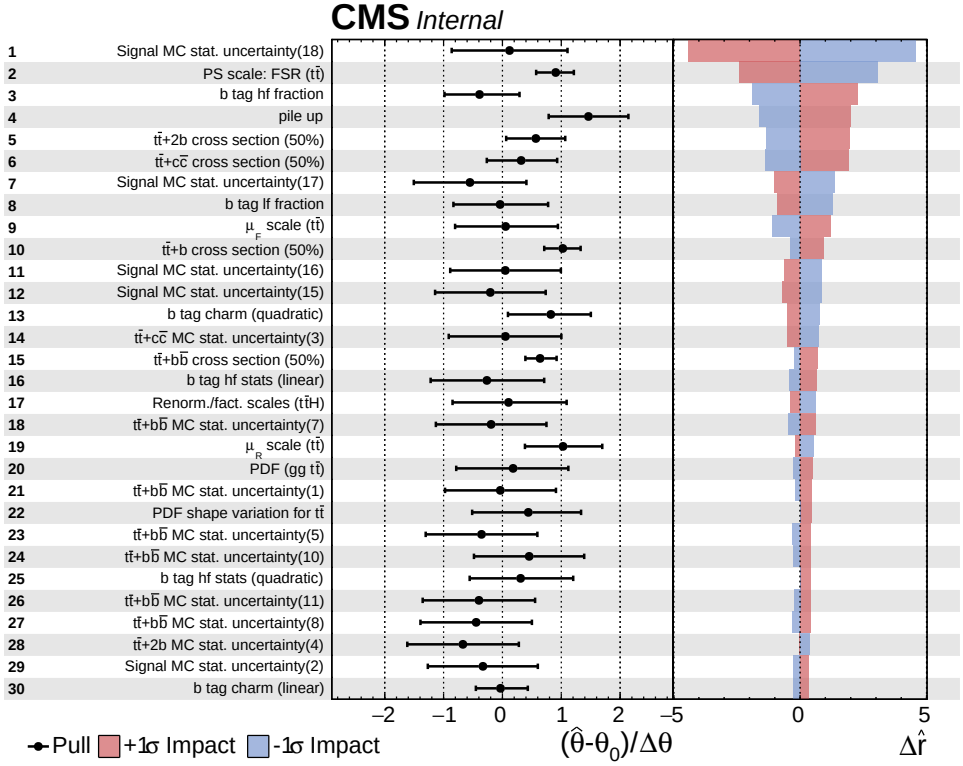


Figure 37: Nuisance parameter pulls and impacts, sorted by the largest impact, for the top 30 most impacting uncertainties, obtained from the fit to CMS data collected in 2017, with 41.5 fb^{-1} of integrated luminosity. The pulls of the nuisance parameters (black markers) are computed relative to their pre-fit values θ_0 and uncertainties $\Delta\theta$. The impact $\Delta\hat{r}$ is computed as the difference of the nominal best fit value of μ and the best fit value obtained when varying the nuisance parameter under scrutiny to its best fit value $\hat{\theta}$ plus/minus (1-σ) its post-fit uncertainty (colored areas).

Moreover note that the case of Run 3 is computed with a center of mass energy of 13 TeV instead of 14 TeV, which has about a 10% increase in the $t\bar{t}HH$ cross section when compared to 13 TeV. The 2017 total luminosity (i.e. 41.5 fb^{-1}) is multiplied by 10, giving a slightly higher luminosity than assuming Run 3 with 400 fb^{-1}. But let's stress here that this back of the envelope calculation does not take into account all the facts and corrections for a correct handling of the data. This is included in the software for preparing the events for the analysis (see event selection), but of course not taken at all into account here. Furthermore with the Run 2 and Run 3 data, other signatures should be investigated to further increase the statistics and refine the results already before the start the HL-LHC as briefly summarized in the next section.

Luminosity	Best-fit μ	95% CL limits on μ		
		Expected		
		Median	68% CL Range	95% CL Range
41.5	$1^{+11.8}_{-10.4}$	24.5	[17.0,35.9]	[12.5,50.8]
41.5∗3	$1^{+6.9}_{-6.3}$	14.0	[9.9,20.3]	[7.3,28.2]
41.5∗10	$1^{+3.9}_{-3.7}$	7.9	[5.6,11.3]	[4.1,15.6]

Table 23: Luminosity scaling for a MC only analysis. Data events for this case are a sum of all the MC simulated events, thus having and expected $\mu = 1$. The expected uncertainties don't take into account the MC bin by bin statistics, and is checked to not influence much this discussion (this is needed for technical reasons).

It looks interesting at this stage to compare with the status of the $t\bar{t}H$ analysis in Run 1. This allows situating the current results obtained on $t\bar{t}$(SL)HH(4b) with 41.5 fb^{-1} and at 13 TeV. It gives also a somehow better guess (than shown on Table 23) on the sensitivity reachable for this process by the end of Run 3, i.e. before the start of HL-LHC.

The first results on $t\bar{t}H$ have been obtained with Run 1 data, i.e. 5 fb^{-1} at 7 TeV and 19.5 fb^{-1} at 8 TeV. This is about 7 years ago.

For easy refference, Table 8 taken from [147] is shown here on Table 24. The results are shown for each Higgs final state and with all the 3 top quark pair signatures included for each final Higgs state. At 8 TeV the $t\bar{t}H$ cross-section is about 4 times smaller than at 13 TeV, i.e. about 100 fb at Leading Order as computed with MG5. The Run 1 has thus roughly provided a total of $\mathcal{O}(2500)$ events. This can be compared to $\mathcal{O}(500)$ $t\bar{t}HH$ events produced with the full Run 2 and Run 3. Besides, the progress in the detector (tracking, especially b-tagging, calorimetry, jet tagging), triggering and in the analysis tools with new DNN and statistical tools as well as in a deeper knowledge of the background (especially $t\bar{t}$ + jets and $t\bar{t}$ + b-jets) as well as of $t\bar{t}H$, will enhance the power of the $t\bar{t}HH$ analysis as compared to the $t\bar{t}H$ at Run 1. Thus provided it is put the needed

effort to explore it, the results on the sensitivity of the $t\bar{t}HH$ process by end of Run 3 will even surpass the one reminded in Table 24 for $t\bar{t}H$ on Run 1.

$t\bar{t}H$ channel	Best-fit μ	95% CL upper limits on $\mu = \sigma/\sigma_{SM}$ ($m_H = 125.6$ GeV)				
			Expected			
	Observed	Observed	Median signal-injected	Median	68% CL range	95% CL range
$\gamma\gamma$	$+2.7^{+2.6}_{-1.8}$	7.4	5.7	4.7	[3.1, 7.6]	[2.2, 11.7]
$b\bar{b}$	$+0.7^{+1.9}_{-1.9}$	4.1	5.0	3.5	[2.5, 5.0]	[1.9, 6.7]
$\tau_h\tau_h$	$-1.3^{+6.3}_{-5.5}$	13.0	16.2	14.2	[9.5, 21.7]	[6.9, 32.5]
4l	$-4.7^{+5.0}_{-1.3}$	6.8	11.9	8.8	[5.7, 14.3]	[4.0, 22.5]
3l	$+3.1^{+2.4}_{-2.0}$	7.5	5.0	4.1	[2.8, 6.3]	[2.0, 9.5]
Same-sign 2l	$+5.3^{+2.1}_{-1.8}$	9.0	3.6	3.4	[2.3, 5.0]	[1.7, 7.2]
Combined	$+2.8^{+1.0}_{-0.9}$	4.5	2.7	1.7	[1.2, 2.5]	[0.9, 3.5]

Table 24: The best-fit values of the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for each $t\bar{t}H$ channel at $m_H = 125.6$ GeV. The signal strength in the four-lepton final state is not allowed to be below approximately -6 by the requirement that the expected signal-plus-background event yield must not be negative in either of the two jet multiplicity bins. The observed and expected 95% CL upper limits on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ for each $t\bar{t}H$ channel at $m_H = 125.6$ GeV are also shown. Taken from Table 8 of Ref. [147].

4.7.1 Overall Plan for the $t\bar{t}HH$ search: from now to High Luminosity LHC

At the dawn of Run 3 and of HL-LHC, it is important to stress once more here the physics interest of the $t\bar{t}HH$ process (see Section 4.2, as well as the BSM motivations from the previous chapters). This implies to pursue the current exploratory study of $t\bar{t}$(SL)HH(4b) with the full Run 2 data and furthermore to include the other top pair channels as well as double Higgs decay especially where one Higgs decays into b-quark pair and the other one into two photons. The aim being to explore at best the $t\bar{t}HH$ process with Run 2, Run 3 and HL-LHC overall data.

Here below are summarized the milestones of the work-plan corresponding to reach this ambitious goal.

- Step1: The 2017 Run data provides $\mathcal{O}(40)$ total events. The $t\bar{t}$(SL)HH(4b) analysis analysed here correspond to $\mathcal{O}(5)$ of these expected events. The work reported here and in [148] corresponds to this first step.
- Step 2: The whole Run 2 is expected to provide $\mathcal{O}(140)$ $t\bar{t}HH$ events. The specific $t\bar{t}$(SL)HH(4b) process is expected to provide $\mathcal{O}(18)$ events. The second step now starting will perform the $t\bar{t}$(SL)HH(4b) analysis with this complete set of Run 2 data. It will follow the new full Run 2 analysis framework developed for the $t\bar{t}$(SL)H($b\bar{b}$) case. It will include improved DNN tools, updates on b-tagging, jets algorithm, on

the $t\bar{t}$ + b's backgrounds (especially the $t\bar{t}$ + 4b, $t\bar{t}H$ and $t\bar{t}Z$) and eventually $t\bar{t}ZZ$, jet energy scalde asystematics, among others.

Furhermore if enough means in people, other $t\bar{t}$ channel decays will be considered, starting with the all hadronic top-pair decay channel. This has an interesting decay rate. And even if this decay channel for $t\bar{t}H(b\bar{b})$ is giving the worst result so far [83], adding 2 b's should enhance the sensitivity to this signature.

– Step 3: with the overall Run 3 at 14 TeV, $\mathcal{O}(400)$ events are expected, as well as an analysis considering all $t\bar{t}$ signatures, taking into account the $t\bar{t}$ + 4b at NLO background. A first evidence for the $t\bar{t}HH$ process is expected and maybe even a possible deviation from the SM.

– Step 4: with the overall Run at HL-LHC, at 14 TeV, $\mathcal{O}(4000)$ $t\bar{t}HH$ events are expected. Here all signatures of $t\bar{t}$ will be studied as well with the $t\bar{t}$ + 4b at NLO background included as expected for the Run 3 analysis together with new $t\bar{t}H$, $t\bar{t}Z$ and $t\bar{t}ZZ$ backgrounds. Here an evidence of the $t\bar{t}HH$ process is expected as well as a possible deviation from the SM. If an increase in cross section is observed, this would be the indication of a BSM effect and, for instance, $MCHM_5$ would be excluded. The study will be combined with heavy resonance searches. The triple Higgs search with complementary inputs to double Higgs will also be addressed.

4.8 Chapter concluding remarks

A first search for $t\bar{t}HH$ is presented with the $t\bar{t}$ decay in the SL mode and both Higgses decaying to bottom quark-antiquark pairs and was endorsed by the CMS collaboration in the CMS PAG meeting of June 9 of 2020. The search is performed with 2017 CMS collected data corresponding to 41.5 fb^{-1} integrated luminosity. A DNN model is used to categorize events and construct discriminants for separating the $t\bar{t}HH$ signal from the background and a fit to data is performed. The best fit value for the signal strength modifier is $4.9^{+14.4}_{-12.8}$. The results are summarized in Table 25.

Finally, as a remark, the study of $t\bar{t}HH$ as the VBF(HH) processes, are both at the femtobarn cross-section level, and are becoming quite essential and complementary to the HH process. The results that can be achieved now with the full Run 2 and Run 3, and followed at the HL-LHC, will be really important in the full exploration of the Top-Higgs sector.

$\hat{\mu}$	Observed	$4.9^{+14.4}_{-12.8}$
	Expected (blind)	$1^{+12.4}_{-11.2}$
95% CL upper limits on μ	Observed	32.9
	Expected (Median)	28.9
	Expected (68% CL range)	[20.1, 42.2]
	Expected (95% CL range)	[14.9, 59.6]

Table 25: Summary of the results of this analysis after a fit to CMS data. The upper limits are obtained with asymptotic likelihood methods according to formulas of Ref [144]. Also shown is the best fit value of the signal strength modifier $\mu = \sigma/\sigma_{\mathtt{SM}} = \hat{\mu}$.

The detailed analysis framework developed will serve for the analysis of the overall Run 2 data and could allow extracting a preliminary $t\bar{t}HH$ signal. This work will be carried on with the Run 3 of the LHC and at the HL-LHC.

The author and his supervisor A. Savoy-Navarro are also co-authoring the $t\bar{t}H(b\bar{b})$ publication by CMS with the overall Run 2 data [117], in preparation and to be published in 2020. A related publication of this $t\bar{t}HH$ first result endorsed by CMS could also give a dedicated publication within this year or with all Run 2 data in 2021.

Chapter 5: Conclusions

This thesis explored both the phenomenology of BSM particle physics that have impacts on the $t\bar{t}H$ and $t\bar{t}HH$ production processes, as well as an experimental analysis, exploring the search for $t\bar{t}HH$ in the SM.

The vector-like fermions, partners of the SM top/bottom quarks were introduced in Chapter 2 with simple Singlet and Doublet models. These new heavy particles mix with the top/bottom quarks, leading to a decrease in the top-Higgs Yukawa coupling, for both models. This decrease also happens for the cross section of $t\bar{t}H$ production, which can be written entirely in terms of the modified top-Yukawa and the SM $t\bar{t}H$ cross section. Experimental measurements were used to put constraints in the parameter space of the models, leading to a very constrained parameter space.

Composite Higgs Models are more interesting BSM implementations, when compared to the simple Singlet and Doublet vector-like fermion models. They are studied in Chapter 3 with the $MCHM_5$ and $MCHM_{14}$ models. Both models also contain vector-like fermions, and mix to the SM top/bottom quark sector in the same way as in Chapter 2, but with a bigger spectrum of heavy fermionic resonances (5 for $MCHM_5$ and 14 for $MCHM_{14}$). They also have Higgs compositeness effects, that are controlled by a scale f. Both effects work together in altering the top Yukawa value that is predicted by the SM, as well as the $t\bar{t}H$ cross section, depending only on this Yukawa. The $t\bar{t}HH$ process was introduced and proved to be an interesting channel to be explored. Both models were implemented in MadGraph 5, at LO, for both 14 TeV and 100 TeV center of mass collision energies. A scan procedure of the parameter space was developed for simulating cross sections. The $MCHM_5$ model has a similar behaviour as the simple Singlet and Doublet models, as it leads only to a decrease in the $t\bar{t}H$ cross section (or the top Yukawa). The $MCHM_{14}$ has the interesting feature of allowing an increase in these two quantities, for some regions of the parameter space. The $t\bar{t}HH$ process is altered by these mentioned top-Yukawa effects, as well as by the vector-like resonances themselves, leading to a change in cross section that can go either in the direction of increasing or decreasing when compared to the SM. This process also has a contribution, at LO, of about 20% of the cross section, from the triple Higgs coupling, which can be further explored as a complementary channel for studying the Higgs potential.

Chapter 4 introduces, for the first time, an experimental search for the $t\bar{t}HH$ production with the CMS experiment at the LHC at CERN. The analysis was done with Run 2 data collected in 2017 with 41.5 fb^{-1} integrated luminosity. The analysis was performed for the semi-leptonic decay channel of the top quark pair, and for the Higgs pair decaying to a bottom quark-antiquark pair. An event selection of at least four jets, of which at least three of them are b-tagged, together with single electron or single muon triggers and a small amount of missing transverse energy, is used to select semi-leptonic events in both data and MC samples. A shape analysis, based on Deep Neural Networks, is used to first classify events in signal or background categories, and further construct discriminant distributions that is further used for fitting to data. A complete set of systematic uncertainties was included in the analysis, and a good agreement of data and MC simulation, within the uncertainties, is observed. A fit to pseudodata is first performed, and an expected value of $\mu = \sigma/\sigma_{\texttt{SM}} = 1^{+12.4}_{-11.2}$, is obtained together with an expected upper limit on μ of 28.9. After a fit to real data, an observed value for μ of $4.9^{+14.4}_{-12.8}$ and an observed upper limit on μ of 32.9 is obtained, and are compatible with the expected limits. The analysis is currently dominated by statistical uncertainties, with systematics starting to play a role. This analysis, fully performed and with the presented here results fully endorsed by the CMS collaboration, will be further developed to be applied to the full Run 2 CMS collected data, and will further consider additional top pair decay channels, different decays for the Higgs pair, and will be continued for the Run 3 data as well as for the HL-LHC.

Appendix A: Representations of SO(5)

The following 5×5 matrix representation of the generators T^B of SO(5) are used:

$$T^1_L = \begin{pmatrix} References: Some 0 & 0 & 0 & -\frac{i}{2} & 0 \\ 0 & 0 & -\frac{i}{2} & 0 & 0 \\ 0 & \frac{i}{2} & 0 & 0 & 0 \\ \frac{i}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \; T^2_L = \begin{pmatrix} 0 & 0 & \frac{i}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{i}{2} & 0 \\ -\frac{i}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{i}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \; T^3_L = \begin{pmatrix} 0 & -\frac{i}{2} & 0 & 0 & 0 \\ \frac{i}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{i}{2} & 0 \\ 0 & 0 & \frac{i}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T^1_R = \begin{pmatrix} 0 & 0 & 0 & \frac{i}{2} & 0 \\ 0 & 0 & -\frac{i}{2} & 0 & 0 \\ 0 & \frac{i}{2} & 0 & 0 & 0 \\ -\frac{i}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \; T^2_R = \begin{pmatrix} 0 & 0 & \frac{i}{2} & 0 & 0 \\ 0 & 0 & 0 & \frac{i}{2} & 0 \\ -\frac{i}{2} & 0 & 0 & 0 & 0 \\ 0 & -\frac{i}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \; T^3_R = \begin{pmatrix} 0 & -\frac{i}{2} & 0 & 0 & 0 \\ \frac{i}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{i}{2} & 0 \\ 0 & 0 & -\frac{i}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$T^{\hat{1}} = \begin{pmatrix} 0 & 0 & 0 & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \frac{i}{\sqrt{2}} & 0 & 0 & 0 & 0 \end{pmatrix}, \; T^{\hat{2}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{i}{\sqrt{2}} & 0 & 0 & 0 \end{pmatrix},$$

$$T^{\hat{3}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{i}{\sqrt{2}} & 0 & 0 \end{pmatrix}, \; T^{\hat{4}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{i}{\sqrt{2}} \\ 0 & 0 & 0 & \frac{i}{\sqrt{2}} & 0 \end{pmatrix}, \tag{A.1}$$

which act on the fundamental representation $\mathbf{5}$ of SO(5) as $\delta\Psi_{\mathbf{5}} = T^B \Psi_{\mathbf{5}}$. We write the $\mathbf{5}$ representation (using a notation appropriate for states with $U(1)_X$ charge $X = 2/3$) as

$$\Psi^{(2/3)}_{\mathbf{5}} \;=\; X_{5/3}\, v_{\frac{1}{2},\frac{1}{2}} + X_{2/3}\, v_{-\frac{1}{2},\frac{1}{2}} + T\, v_{\frac{1}{2},-\frac{1}{2}} + B\, v_{-\frac{1}{2},-\frac{1}{2}} + \tilde{T}\, v_0 \;, \tag{A.2}$$

where the normalized basis is used

$$v_{\frac{1}{2},\frac{1}{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} i \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix},\quad v_{-\frac{1}{2},\frac{1}{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ -i \\ -1 \\ 0 \end{pmatrix},\quad v_{\frac{1}{2},-\frac{1}{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 0 \\ -i \\ 1 \\ 0 \end{pmatrix},\quad v_{-\frac{1}{2},-\frac{1}{2}} = \frac{1}{\sqrt{2}}\begin{pmatrix} -i \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and

$$v_0^T = (0,0,0,0,1)\ .$$

The notation is such that the subindices a,b denote the $T^3_{L,R}$ eigenvalues: $T^3_L\, v_{a,b} = a\, v_{a,b}$, $T^3_R\, v_{a,b} = b\, v_{a,b}$, while v_0 denotes the complete singlet, $T^i_L\, v_0 = T^i_R\, v_0 = 0$. Furthermore, one has $T^+_L\, v_{\frac{1}{2},\pm\frac{1}{2}} = 0$, $T^+_L\, v_{-\frac{1}{2},\pm\frac{1}{2}} = v_{\frac{1}{2},\pm\frac{1}{2}}$, $T^-_L\, v_{\frac{1}{2},\pm\frac{1}{2}} = v_{-\frac{1}{2},\pm\frac{1}{2}}$, $T^-_L\, v_{-\frac{1}{2},\pm\frac{1}{2}} = 0$, and similarly $T^+_R\, v_{\pm\frac{1}{2},\frac{1}{2}} = 0$, $T^+_R\, v_{\pm\frac{1}{2},-\frac{1}{2}} = v_{\pm\frac{1}{2},\frac{1}{2}}$, $T^-_R\, v_{\pm\frac{1}{2},\frac{1}{2}} = v_{\pm\frac{1}{2},-\frac{1}{2}}$, $T^-_R\, v_{\pm\frac{1}{2},-\frac{1}{2}} = 0$, where $T^\pm_{L,R} \equiv T^1_{L,R} \pm i T^2_{L,R}$ are the standard $SU(2)$ raising and lowering operators, satisfying

$$\begin{aligned} [T^3_L, T^\pm_L] &= \pm T^\pm_L\ , \qquad [T^3_R, T^\pm_R] = \pm T^\pm_R\ , \\ [T^+_L, T^-_L] &= 2\,T^3_L\ , \qquad [T^+_R, T^-_R] = 2\,T^3_R\ , \end{aligned}$$

and $[T^i_L, T^j_R] = 0$. The states satisfy the standard normalization

$$\begin{aligned} T^\pm_L |s_L, m_L; s_R, m_R\rangle &= \sqrt{s_L(s_L+1) - m_L(m_L \mp 1)}\, |s_L, m_L \pm 1; s_R, m_R\rangle\ , \\ T^\pm_R |s_L, m_L; s_R, m_R\rangle &= \sqrt{s_R(s_R+1) - m_R(m_R \mp 1)}\, |s_L, m_L; s_R, m_R \pm 1\rangle\ . \end{aligned}$$

Thus, $\Psi_{\mathbf{4}} \sim (X_{5/3}, X_{2/3}, T, B)$ is a bi-doublet $(\mathbf{2},\mathbf{2})$ of $SU(2)_L \times SU(2)_R$ with $(s_L, s_R) = (1/2, 1/2)$, while $\Psi_{\mathbf{1}} \sim \tilde{T}$ is a singlet with $(s_L, s_R) = (0,0)$.

The **14** representation of SO(5) can be written in terms of a 5×5 symmetric and traceless matrix, such that the $SO(5)$ generators (A.1) act as $\delta\Psi_{\mathbf{14}} = [T^B, \Psi_{\mathbf{14}}]$. Using again the notation for states with $X = 2/3$:

$$\begin{aligned} \Psi^{(2/3)}_{\mathbf{14}} &= \tilde{T}\, S_0 + (X_{\frac{5}{3}}\, S_{\frac{1}{2},\frac{1}{2}} + X_{\frac{2}{3}}\, S_{-\frac{1}{2},\frac{1}{2}} + T\, S_{\frac{1}{2},-\frac{1}{2}} + B\, S_{-\frac{1}{2},-\frac{1}{2}}) \\ &\quad + [U_{\frac{8}{3}} S_{1,1} + U_{\frac{5}{3}} S_{0,1} + U_{\frac{2}{3}} S_{-1,1} \\ &\quad +\ Y_{\frac{5}{3}}\, S_{1,0} + Y_{\frac{2}{3}}\, S_{0,0} + Y_{-\frac{1}{3}}\, S_{-1,0} \\ &\quad +\ Z_{\frac{2}{3}}\, S_{1,-1} + Z_{-\frac{1}{3}}\, S_{0,-1} + Z_{-\frac{4}{3}}\, S_{-1,-1}]\ , \end{aligned} \tag{A.3}$$

which exhibits the decomposition $\mathbf{14} \sim (\mathbf{3},\mathbf{3}) + (\mathbf{2},\mathbf{2}) + (\mathbf{1},\mathbf{1})$ under $\mathrm{SU}(2)_L \times \mathrm{SU}(2)_R$.

The notation for the basis $S_{a,b}$ is the same as explained above, e.g. $[T_L^3, S_{a,b}] = a\, S_{a,b}$, $[T_R^3, S_{a,b}] = b\, S_{a,b}$, while S_0 denotes the complete singlet, $[T_L^i, S_0] = [T_R^i, S_0] = 0$. The states $\Psi_{\mathbf{9}} \sim (U_{8/3}, U_{5/3}, U_{2/3}, Y_{5/3}, Y_{2/3}, Y_{-1/3}, Z_{2/3}, Z_{-1/3}, Z_{-4/3})$ transform as a bi-triplet of $SU(2)_L \times SU(2)_R$ with $(s_L, s_R) = (1,1)$, and the bi-doublet $\Psi_{\mathbf{4}}$ and singlet $\Psi_{\mathbf{1}}$ follow the same notation used for the $\mathbf{5}$ of $SO(5)$ above. The subindices on the fields denote the electric charge, given by

$$Q = T_L^3 + T_R^3 + X \ . \tag{A.4}$$

Explicitly, the $S_{a,b}$ and S_0 matrices are given by:

$$S_{1,1} = \tfrac{1}{2}\begin{pmatrix} 1 & i & 0 & 0 & 0 \\ i & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, S_{1,0} = \tfrac{1}{2\sqrt{2}}\begin{pmatrix} 0 & 0 & -1 & -i & 0 \\ 0 & 0 & -i & 1 & 0 \\ -1 & -i & 0 & 0 & 0 \\ -i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, S_{1,-1} = \tfrac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & i & 0 \\ 0 & 0 & i & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S_{0,1} = \tfrac{1}{2\sqrt{2}}\begin{pmatrix} 0 & 0 & -1 & i & 0 \\ 0 & 0 & -i & -1 & 0 \\ -1 & -i & 0 & 0 & 0 \\ i & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, S_{0,0} = \tfrac{1}{2}\begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, S_{0,-1} = \tfrac{1}{2\sqrt{2}}\begin{pmatrix} 0 & 0 & 1 & i & 0 \\ 0 & 0 & -i & 1 & 0 \\ 1 & -i & 0 & 0 & 0 \\ i & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$S_{-1,1} = \tfrac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -i & 0 \\ 0 & 0 & -i & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, S_{-1,0} = \tfrac{1}{2\sqrt{2}}\begin{pmatrix} 0 & 0 & 1 & -i & 0 \\ 0 & 0 & -i & -1 & 0 \\ 1 & -i & 0 & 0 & 0 \\ -i & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, S_{-1,-1} = \tfrac{1}{2}\begin{pmatrix} 1 & -i & 0 & 0 & 0 \\ -i & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

for the **9** of $SO(4) = SU(2)_L \times SU(2)_R$,

$$S_{\frac{1}{2},\frac{1}{2}} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ i & -1 & 0 & 0 & 0 \end{pmatrix}, \quad S_{\frac{1}{2},-\frac{1}{2}} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -i & 1 & 0 \end{pmatrix},$$

$$S_{-\frac{1}{2},\frac{1}{2}} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & -i & -1 & 0 \end{pmatrix}, \quad S_{-\frac{1}{2},-\frac{1}{2}} = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 & -i \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -i & -1 & 0 & 0 & 0 \end{pmatrix},$$

for the **4** of $SO(4)$, and

$$S_0 = \frac{1}{2\sqrt{5}}\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -4 \end{pmatrix}.$$

for the $SO(4)$ singlet. These matrices form an orthonormal basis: $\mathrm{Tr}(S_A^* S_B) = \delta_{AB}$.

Appendix B: Embeddings of SO(4) into SO(5)

The four NGBs resulting from the spontaneous breaking of

$$SO(5) \to SO(4) = SU(2)_L \times SU(2)_R$$

can be parametrized as

$$U = e^{i\frac{\sqrt{2}}{f}h^{\hat{a}}T^{\hat{a}}} , \tag{B.1}$$

where $T^{\hat{a}}$ are the (four) broken generators in $SO(5)/SO(4)$ given in Eq (A.1) and f is the scale of spontaneous symmetry breaking. The $h^{\hat{a}}$ transform as a 4-plet of $SO(4)$, and can be arranged in a doublet of $SU(2)_L$ as:

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} h_2 + ih_1 \\ h_4 - ih_3 \end{pmatrix} . \tag{B.2}$$

One can assume that EWSB proceeds through a non-vanishing vev $h_0 = \langle h_4 \rangle$, with the vev's of the other components vanishing. In unitary gauge, $h_1 = h_2 = h_3 = 0$ and $h_4 = h_0 + h$, where h is the physical Higgs boson. Using the explicit form of the broken generators results in the matrix given in Eq. (3.1).

It is also easy to embed the various fermion $SO(4)$ multiplets used in the main text into (incomplete) representations of $SO(5)$ that simplify the writing of the Lagrangian. For example,

- For the MCHM$_5$:

  The elementary fermions are written as

$$\begin{aligned} Q_L^{\mathbf{5}} &= t_L\, v_{\frac{1}{2},-\frac{1}{2}} + b_L\, v_{-\frac{1}{2},-\frac{1}{2}} , \\ T_R^{\mathbf{5}} &= t_R\, v_0 , \end{aligned}$$

and the composite fermions are written as

$$\begin{aligned}\Psi_{\mathbf{4}} &= X_{5/3}\, v_{\frac{1}{2},\frac{1}{2}} + X_{2/3}\, v_{-\frac{1}{2},\frac{1}{2}} + T\, v_{\frac{1}{2},-\frac{1}{2}} + B\, v_{-\frac{1}{2},-\frac{1}{2}} \ , \\ \Psi_{\mathbf{1}} &= \tilde{T}\, v_0 \ .\end{aligned}$$

These result in Eqs. (3.9) and (3.10).

- For the MCHM$_{14}$:

  The elementary fermions are written as

$$\begin{aligned}Q_L^{\mathbf{14}} &= t_L\, S_{\frac{1}{2},-\frac{1}{2}} + b_L\, S_{-\frac{1}{2},-\frac{1}{2}} \ , \\ T_R^{\mathbf{14}} &= t_R\, S_0 \ ,\end{aligned}$$

  and the composite fermions are written as

$$\begin{aligned}\Psi_{\mathbf{9}} &= [U_{\frac{8}{3}} S_{1,1} + U_{\frac{5}{3}} S_{0,1} + U_{\frac{2}{3}} S_{-1,1} \\ &\quad + \ Y_{\frac{5}{3}}\, S_{1,0} + Y_{\frac{2}{3}}\, S_{0,0} + Y_{-\frac{1}{3}}\, S_{-1,0} \\ &\quad + \ Z_{\frac{2}{3}}\, S_{1,-1} + Z_{-\frac{1}{3}}\, S_{0,-1} + Z_{-\frac{4}{3}}\, S_{-1,-1}] \\ \Psi_{\mathbf{4}} &= X_{5/3}\, S_{\frac{1}{2},\frac{1}{2}} + X_{2/3}\, S_{-\frac{1}{2},\frac{1}{2}} + T\, S_{\frac{1}{2},-\frac{1}{2}} + B\, S_{-\frac{1}{2},-\frac{1}{2}} \ , \\ \Psi_{\mathbf{1}} &= \tilde{T}\, S_0 \ .\end{aligned}$$

  All of these are traceless, symmetric 5×5 matrices. It is then straightforward to form a complete **14** of $SO(5)$ as

$$\Psi_{\mathbf{14}} = \Psi_{\mathbf{9}} + \Psi_{\mathbf{4}} + \Psi_{\mathbf{1}} \ .$$

Appendix C: Explicit form of gauge and Higgs interactions

In unitary gauge, the gauged d_μ and e_μ symbols are given by

$$
\begin{aligned}
d_\mu &= \left\{ \frac{g}{\sqrt{2}} W_\mu^1 s_h,\ \frac{g}{\sqrt{2}} W_\mu^2 s_h,\ \frac{g W_\mu^3 - g' B_\mu}{\sqrt{2}} s_h, \frac{\partial_\mu h}{f} \right\} \\
e_\mu &= \frac{i}{2} \begin{pmatrix} 0 & -g' B_\mu - g W_\mu^3 & g W_\mu^2 & -g W_\mu^1 c_h \\ g' B_\mu + g W_\mu^3 & 0 & -g W_\mu^1 & -g W_\mu^2 c_h \\ -g W_\mu^2 & g W_\mu^1 & 0 & (g' B_\mu - g W_\mu^3) c_h \\ g W_\mu^1 c_h & g W_\mu^2 c_h & -(g' B_\mu - g W_\mu^3) c_h & 0 \end{pmatrix},
\end{aligned}
$$

where $c_h = \cos \frac{h_0+h}{f}$ and $s_h = \sin \frac{h_0+h}{f}$. Under $SO(4)$, d_μ^i transforms in the fundamental and e_μ^a in the adjoint. In terms of these, as well as the fermion embeddings from the previous appendices, we may write explicitly the vertices from the Lagrangians $\mathcal{L}_{\mathbf{int}}^{5,14}$ in the interaction basis as

$$
\begin{aligned}
&\mathcal{L}_{\text{int}}^{\mathbf{5}} = \\
&\quad c_R \Bigg\{ g\sqrt{\xi} \left[\frac{1}{\sqrt{2}} \overline{\left(X_{5/3}\right)_R} \not{W}^+ \tilde{T}_R - \frac{1}{\sqrt{2}} \overline{B_R} \not{W}^- \tilde{T}_R - \frac{1}{2c_w} \overline{T_R} \not{Z} \tilde{T}_R \ - \frac{1}{2c_w} \overline{\left(X_{2/3}\right)_R} \not{Z} \tilde{T}_R \right] \\
&\qquad + i \left[\overline{\left(X_{2/3}\right)_R} - \overline{T}_R \right] \frac{\not{\partial} h}{f} \tilde{T}_R \Bigg\} + (R \to L) + \text{h.c.}\ ,
\end{aligned}
\tag{C.1}
$$

$$
\begin{aligned}
\mathcal{L}_{\rm int}^{\mathbf{14}} \;=\;
& c_4\left[g\sqrt{\xi}\left(\frac{1}{\sqrt{2}}\overline{X}_{5/3}\not{W}^{+}\tilde{T}-\frac{1}{\sqrt{2}}\overline{B}\not{W}^{-}\tilde{T}-\frac{1}{2c_w}\overline{T}\not{Z}\tilde{T}\;-\frac{1}{2c_w}\overline{X}_{2/3}\not{Z}\tilde{T}\right)\right.\\
& \left.+\,i\left(\overline{X}_{2/3}-\overline{T}\right)\frac{\not{\partial}h}{f}\tilde{T}\right]\\
- \;& \frac{1}{4}c_9\frac{g_2}{c_w}\sqrt{\xi}\left\{\sqrt{2}\left[\overline{Y}_{-1/3}+\overline{Z}_{-1/3}\right]\not{Z}B+\left[\overline{Y}_{2/3}+2\overline{U}_{2/3}\right]\not{Z}X_{2/3}\right.\\
+ \;& \left.\left[\overline{Y}_{2/3}+2\overline{Z}_{2/3}\right]\not{Z}T+\sqrt{2}\left[\overline{U}_{5/3}+\overline{Y}_{5/3}\right]\not{Z}X_{5/3}\right\}\\
- \;& \frac{1}{4}c_9g_2\sqrt{\xi}\left\{\sqrt{2}\overline{Y}_{2/3}\not{W}^{-}\overline{X}_{5/3}+2\sqrt{2}\overline{Z}_{-4/3}\not{W}^{-}B+2\overline{Z}_{-1/3}\not{W}^{-}T+2\overline{Y}_{-1/3}\not{W}^{-}X_{2/3}\right.\\
- \;& \left.\sqrt{2}\overline{Y}_{2/3}\not{W}^{+}B+2\sqrt{2}\overline{U}_{8/3}\not{W}^{+}X_{5/3}+2\overline{Y}_{5/3}\not{W}^{+}T+2\overline{U}_{5/3}\not{W}^{+}X_{2/3}\right\}\\
+ \;& \frac{i}{2f}c_9\left\{\sqrt{2}\left[\overline{Y}_{-1/3}-\overline{Z}_{-1/3}\right]\not{\partial}hB+\left[\overline{Y}_{2/3}-2\overline{Z}_{2/3}\right]\not{\partial}hT+\right.\\
& \left.\left[2\overline{U}_{2/3}-\overline{Y}_{2/3}\right]\not{\partial}hX_{2/3}+\sqrt{2}\left[\overline{U}_{5/3}-\overline{Y}_{5/3}\right]\not{\partial}hX_{5/3}\right\}\\
+ \;& \frac{ic_{T9}}{4\pi f^3}\partial_\mu h\,\partial^\mu h\left[\overline{U}_{2/3}-\overline{Y}_{2/3}+\overline{Z}_{2/3}\right]\tilde{T}\\
+ \;& \frac{c_{T9}g_2\sqrt{\xi}}{4\pi f^2c_w}Z_\mu\left[\partial^\mu h\left(\overline{Z}_{2/3}-\overline{U}_{2/3}\right)+W_\mu^{+}\left(\overline{U}_{5/3}-\overline{Y}_{5/3}\right)+W_\mu^{-}\left(\overline{Z}_{-1/3}-\overline{Y}_{-1/3}\right)\right]\tilde{T}\\
+ \;& \frac{ic_{T9}g_2^2\xi}{8\pi f}\left[-\frac{Z_\mu Z^\mu}{2c_w}\left(\overline{U}_{2/3}+\overline{Y}_{2/3}+\overline{Z}_{2/3}\right)-W_\mu^{+}W^{+\mu}\overline{U}_{8/3}-\;W_\mu^{-}W^{-\mu}\overline{Z}_{-4/3}\right.\\
+ \;& \left.W_\mu^{+}W^{-\mu}\overline{Y}_{2/3}-\frac{Z_\mu W^{-\mu}}{c_w}\left(\overline{Y}_{-1/3}+\overline{Z}_{-1/3}\right)+\frac{Z_\mu W^{+\mu}}{c_w}\left(\overline{U}_{5/3}+\overline{Y}_{5/3}\right)\right]\tilde{T}+\text{h. c. ,}
\end{aligned}
$$

where c_w is the cosine of the Weinberg angle.

Furthermore, the e_μ symbol in the covariant derivatives lead to compositeness corrected electroweak interactions for the resonances in the fourplet and nonet, given by

$$
\begin{aligned}
\mathcal{L}^{\mathbf{5}}_{\text{gauge}} &= \bar{\Psi}_{\mathbf{4}}\left(\frac{2}{3}g'\not{B}-\not{\phi}\right)\Psi_{\mathbf{4}} \\
&= \frac{g}{c_w}\left[\left(-\frac{1}{2}+\frac{1}{3}s_w^2\right)\bar{B}\not{Z}B+\left(\frac{1}{2}-\frac{5}{3}s_w^2\right)\bar{X}_{5/3}\not{Z}X_{5/3}\right. \\
&+ \left.\left(\frac{\sqrt{1-\xi}}{2}-\frac{2}{3}s_w^2\right)\bar{T}\not{Z}T+\left(-\frac{\sqrt{1-\xi}}{2}-\frac{2}{3}s_w^2\right)\bar{X}_{2/3}\not{Z}X_{2/3}\right] \\
&+ \frac{g}{\sqrt{2}}\frac{1+\sqrt{1-\xi}}{2}\left[\bar{B}\not{W}^-T+\bar{X}_{5/3}\not{W}^+X_{2/3}+\text{h. c.}\right] \\
&+ \frac{g}{\sqrt{2}}\frac{1-\sqrt{1-\xi}}{2}\left[\bar{X}_{5/3}\not{W}^+T+\bar{B}\not{W}^-X_{2/3}+\text{h. c.}\right], \qquad \text{(C.2)}
\end{aligned}
$$

plus standard photon couplings, for the fourplet (these are the same as in reference [149], reported here again for completeness), and

$$
\begin{aligned}
\mathcal{L}^{\mathbf{14}}_{\text{gauge}} &= \bar{\Psi}_{\mathbf{9}}\left(\frac{2}{3}g'\not{B}\Psi_{\mathbf{9}}-[\not{\phi},\Psi_{\mathbf{9}}]\right) \\
&= \frac{g}{c_w}\left[\left(-\sqrt{1-\xi}+-\frac{2}{3}s_w^2\right)\bar{U}_{2/3}\not{Z}U_{2/3}+\left(\frac{1-\sqrt{1-\xi}}{2}-\frac{5}{3}s_w^2\right)\bar{U}_{5/3}\not{Z}U_{5/3}\right. \\
&+ \left(\frac{1}{2}-\frac{8}{3}s_w^2\right)\bar{U}_{8/3}\not{Z}U_{8/3}+\left(-\frac{1+\sqrt{1-\xi}}{2}+\frac{1}{3}s_w^2\right)\bar{Y}_{-1/3}\not{Z}Y_{-1/3} \\
&+ -\frac{2}{3}s_w^2\bar{Y}_{2/3}\not{Z}Y_{2/3}+\left(\frac{1+\sqrt{1-\xi}}{2}+\frac{1}{3}s_w^2\right)\bar{Y}_{5/3}\not{Z}Y_{5/3} \\
&+ \left(-1+\frac{4}{3}s_w^2\right)\bar{Z}_{-4/3}\not{Z}Z_{-4/3}+\left(-\frac{1-\sqrt{1-\xi}}{2}+\frac{1}{3}s_w^2\right)\bar{Z}_{-1/3}\not{Z}Z_{-1/3} \\
&+ \left.\left(\sqrt{1-\xi}-\frac{2}{3}s_w^2\right)\bar{Z}_{2/3}\not{Z}Z_{2/3}\right] \\
&+ \frac{g}{2}(1+\sqrt{1-\xi})\left[\bar{Y}_{5/3}\not{W}^+Y_{2/3}+\bar{Y}_{2/3}\not{W}^+Y_{-1/3}\right. \\
&+ \bar{Z}_{-1/3}\not{W}^+Z_{-4/3}+\bar{Z}_{2/3}\not{W}^+Z_{-1/3} \\
&+ \left.\bar{U}_{8/3}\not{W}^+U_{5/3}+\bar{U}_{5/3}\not{W}^+U_{2/3}+\text{h. c.}\right] \\
&+ \frac{g}{2}(1-\sqrt{1-\xi})\left[\bar{Y}_{-1/3}\not{W}^+Z_{-4/3}+\bar{Y}_{2/3}\not{W}^+Z_{-1/3}\right. \\
&+ \bar{Y}_{5/3}\not{W}^+Z_{2/3}+\bar{U}_{5/3}\not{W}^+Y_{2/3} \\
&+ \left.\bar{U}_{8/3}\not{W}^+Y_{5/3}+\bar{U}_{2/3}\not{W}^+Y_{-1/3}+\text{h. c.}\right], \qquad \text{(C.3)}
\end{aligned}
$$

plus standard photon couplings, for the nonet.

Appendix D: The Deep Neural Network Input Variables

Tables 26 and 27 list all the kinematic quantities used for this analysis as well as their definition. These input variables enter as input variables for the DNN model.

Table 26: Input variables used in the DNNs (Continued in Table 27.)

Variable	Definition
BLR	likelihood ratio discriminating between events with 4 b-quark jets and 2 b-quark jets
$\mathrm{BLR}^{\mathrm{trans}}$	$\ln[\mathrm{BLR}/(1-\mathrm{BLR})]$
p_{T} (jet 1)	p_{T} of the 1^{st} jet, ranked in jet p_{T}
p_{T} (jet 2)	p_{T} of the 2^{nd} jet, ranked in jet p_{T}
p_{T} (jet 3)	p_{T} of the 3^{rd} jet, ranked in jet p_{T}
p_{T} (jet 4)	p_{T} of the 4^{th} jet, ranked in jet p_{T}
H_T^b	scalar sum of p_{T} of b-tagged jets
H_T	scalar sum of p_{T} of all jets
$\mathrm{N}_b^{\mathrm{loose}}$	number of b-tagged jets with DeepCSV > 0.1522
$\mathrm{N}_b^{\mathrm{medium}}$	number of b-tagged jets with DeepCSV > 0.4941
$\mathrm{N}_b^{\mathrm{tight}}$	number of b-tagged jets with DeepCSV > 0.8001
$\mathrm{N}_{\mathrm{jets}}$	number of jets
d(jet 1)	b-tagging discriminant value of 1^{st} jet, ranked in jet p_{T}
d(jet 2)	b-tagging discriminant value of 2^{nd} jet, ranked in jet p_{T}
d(jet 3)	b-tagging discriminant value of 3^{rd} jet, ranked in jet p_{T}
d(jet 4)	b-tagging discriminant value of 4^{th} jet, ranked in jet p_{T}
d_1	1^{st} highest b-tagging discriminant value of all jets
d_2	2^{nd} highest b-tagging discriminant value of all jets
d_3	3^{rd} highest b-tagging discriminant value of all jets
d_4	4^{th} highest b-tagging discriminant value of all jets
$\mathrm{d}_j^{\mathrm{avg}}$	average b-tagging discriminant value of all jets
$\mathrm{d}_b^{\mathrm{avg}}$	average b-tagging discriminant value of all b-tagged jets
$\mathrm{d}_b^{\mathrm{min}}$	minimal b-tagging discriminant value of all b-tagged jets
$\mathrm{d}_j^{\mathrm{min}}$	minimal b-tagging discriminant value of all jets

Table 27: Continued from Table 26.

Variable	Definition
$\frac{1}{N_b}\Sigma_b^{N_b}(d-d_b^{\text{avg}})^2$	squared difference between the b-tagged discriminant value of a b-tagged jet and the average b-tagging discriminant value of all b-tagged jets, averaged over all b-tagged jets
$\frac{1}{N_j}\Sigma_j^{N_j}(d-d_j^{\text{avg}})^2$	squared difference between the b-tagged discriminant value of a jet and the average b-tagging discriminant value of all jets, averaged over all jets
m_j^{avg}	average mass of all jets
$\text{m}_{b,b}^{\text{closest to 125}}$	mass of pair of b-tagged jets closest to 125 GeV
$\text{m}_{\text{lep},b}^{\text{min}\Delta R}$	mass of pair of lepton and b-tagged jet closest in ΔR
$\text{m}_{j,j}^{\text{min}\Delta R}$	mass of pair of jets closest in ΔR
$\text{m}_{b,b}^{\text{min}\Delta R}$	mass of pair of b-tagged jets closest in ΔR
$\text{m}_{\text{lep},j}^{\text{min}\Delta R}$	mass of pair of lepton and jet closest in ΔR
m_b^{avg}	average mass of all b-tagged jets
$(\text{m}^2)_b^{\text{avg}}$	average squared mass of all b-tagged jets
$\Delta R_{\text{b,b}}^{\text{avg}}$	average ΔR between b-tagged jets
$\Delta R_{\text{j,j}}^{\text{avg}}$	average ΔR between two jets
$\Delta R_{\text{j,j}}^{\text{min}}$	minimal ΔR between any two jets
$\Delta R_{\text{b,b}}^{\text{min}}$	minimal ΔR between any two b-tagged jets
$\Delta R_{\text{lep,b}}^{\text{min}}$	minimal ΔR between lepton and b-tagged jet
$\Delta R_{\text{lep,j}}^{\text{min}}$	minimal ΔR between lepton and jet
$\Delta\eta_{\text{j,j}}^{\text{avg}}$	average $\Delta\eta$ between any two jets
$\Delta\eta_{\text{b,b}}^{\text{avg}}$	average $\Delta\eta$ between any two b-tagged jets
$\Delta\eta_{\text{b,b}}^{\text{max}}$	largest $\Delta\eta$ between any two b-tagged jets
$p_{\text{T}}{}_{\text{j}}^{\text{avg}}/\text{E}_{\text{j}}^{\text{avg}}$	average jet p_{T} over average jet E of all jets
H_0	0th Fox–Wolfram moment [150] computed with all jets
H_1	1st Fox–Wolfram moment [150] computed with all jets

References:

Some references listed here include Internal Notes or documents available only to CMS or LHC related users. They are highlighted as blue.

1. M. Cepeda et al., "Report from Working Group 2." *CERN Yellow Rep. Monogr.* **7** (2019) 221 [1902.00134].

2. ATLAS, CMS collaboration, G. Aad et al., "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV." *JHEP* **08** (2016) 045 [1606.02266].

3. Ricardo D'Elia Matheus, E. Pontón, L. A. F. do Prado and A. Savoy-Navarro, "Impact of vector-like fermion hypothesis in the production of $t\bar{t}H$ at the LHC and the measurement of the top-Higgs Yukawa coupling." CMS Analysis Note 2017/258, CERN, 2017. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/258.

4. A. Lenz, "Constraints on a fourth generation of fermions from Higgs Boson searches." *Adv. High Energy Phys.* **2013** (2013) 910275.

5. P. Ramond, "Lectures on grand unification." in *NATO Advanced Study Institute: Chaotic Behavior of Deterministic Systems Les Houches, France, June 29-July 31, 1981*, pp. 65–108, 1981.

6. G. Branco and L. Lavoura, "On the addition of vector-like quarks to the standard model." *Nuclear Physics B* **278** (1986) 738 .

7. T. Kaluza, "On the Problem of Unity in Physics." *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1921** (1921) 966.

8. O. Klein, "Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)." *Z. Phys.* **37** (1926) 895.

9. E. Witten, "Search for a Realistic Kaluza-Klein Theory." *Nucl. Phys.* **B186** (1981) 412.

10. L. Randall and R. Sundrum, "A Large mass hierarchy from a small extra dimension." *Phys. Rev. Lett.* **83** (1999) 3370 [hep-ph/9905221].

11. J. M. Maldacena, "The Large N limit of superconformal field theories and supergravity." *Int. J. Theor. Phys.* **38** (1999) 1113 [hep-th/9711200].

12. E. Witten, "Anti-de Sitter space and holography." *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150].

13. N. Arkani-Hamed, M. Porrati and L. Randall, "Holography and phenomenology."*JHEP* **08** (2001) 017 [hep-th/0012148].

14. R. Contino and A. Pomarol, "Holography for fermions."*JHEP* **11** (2004) 058 [hep-th/0406257].

15. D. B. Kaplan, "Flavor at SSC energies: A new mechanism for dynamically generated fermion masses."*Nuclear Physics B* **365** (1991) 259 .

16. ATLAS collaboration, 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** (2012) 1 [1207.7214].

17. CMS collaboration, 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** (2012) 30 [1207.7235].

18. K. Agashe, R. Contino and A. Pomarol, "The Minimal composite Higgs model."*Nucl. Phys.* **B719** (2005) 165 [hep-ph/0412089].

19. R. Contino and G. Servant, "Discovering the top partners at the LHC using same-sign dilepton final states."*JHEP* **06** (2008) 026 [0801.1679].

20. J. Mrazek and A. Wulzer, "A Strong Sector at the LHC: Top Partners in Same-Sign Dileptons."*Phys. Rev.* **D81** (2010) 075006 [0909.3977].

21. G. Dissertori, E. Furlan, F. Moortgat and P. Nef, "Discovery potential of top-partners in a realistic composite Higgs model with early LHC data."*JHEP* **09** (2010) 019 [1005.4414].

22. T. A. collaboration, "Search for anomalous production of events with same-sign dileptons and b jets in 14.3 fb^{-1} of pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector." Tech. Rep. ATLAS-CONF-2013-051, CERN, Geneva, May, 2013. https://cds.cern.ch/record/1547567.

23. CMS collaboration, S. Chatrchyan et al., "Search for top-quark partners with charge 5/3 in the same-sign dilepton final state."*Phys. Rev. Lett.* **112** (2014) 171801 [1312.2391].

24. J. F. Gunion, H. E. Haber, G. L. Kane and S. Dawson, "The Higgs Hunter's Guide."*Front. Phys.* **80** (2000) 1.

25. M. Gillioz, R. Grober, C. Grojean, M. Muhlleitner and E. Salvioni, "Higgs Low-Energy Theorem (and its corrections) in Composite Models."*JHEP* **10** (2012) 004 [1206.7120].

26. A. Falkowski, "Pseudo-goldstone Higgs production via gluon fusion."*Phys. Rev.* **D77** (2008) 055018 [0711.0828].

27. I. Low and A. Vichi, "On the production of a composite Higgs boson."*Phys. Rev.* **D84** (2011) 045019 [1010.2753].

28. A. Azatov and J. Galloway, "Light Custodians and Higgs Physics in Composite Models."*Phys. Rev.* **D85** (2012) 055013 [1110.5646].

29. F. Wilczek, "Decays of Heavy Vector Mesons Into Higgs Particles."*Phys. Rev. Lett.* **39** (1977) 1304.

30. H. M. Georgi, S. L. Glashow, M. E. Machacek and D. V. Nanopoulos, "Higgs Bosons from Two Gluon Annihilation in Proton Proton Collisions."*Phys. Rev. Lett.* **40** (1978) 692.

31. J. R. Ellis, M. K. Gaillard, D. V. Nanopoulos and C. T. Sachrajda, "Is the Mass of the Higgs Boson About 10-GeV?."*Phys. Lett.* **B83** (1979) 339.

32. T. G. Rizzo, "Decays of Heavy Higgs Bosons."*Phys. Rev.* **D22** (1980) 722.

33. A. Djouadi, "The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model."*Phys. Rept.* **457** (2008) 1 [hep-ph/0503172].

34. M. Carena, L. Da Rold and E. Pontón, "Minimal Composite Higgs Models at the LHC."*JHEP* **06** (2014) 159 [1402.2987].

35. J. R. Ellis, M. K. Gaillard and D. V. Nanopoulos, "A Phenomenological Profile of the Higgs Boson."*Nucl. Phys.* **B106** (1976) 292.

36. M. A. Shifman, A. I. Vainshtein, M. B. Voloshin and V. I. Zakharov, "Low-Energy Theorems for Higgs Boson Couplings to Photons."*Sov. J. Nucl. Phys.* **30** (1979) 711.

37. S. Dawson and E. Furlan, "A higgs conundrum with vector fermions."*Phys. Rev. D* **86** (2012) 015021.

38. Z.-z. Xing, H. Zhang and S. Zhou, "Updated Values of Running Quark and Lepton Masses."*Phys. Rev.* **D77** (2008) 113016 [0712.1419].

39. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations."*JHEP* **07** (2014) 079 [1405.0301].

40. V. Khachatryan et al., "Measurement of the ratio $\lfloor(t \to wb)/\lfloor(t \to wq)$ in pp collisions at $\sqrt{s} = 8$ tev."*Physics Letters B* **736** (2014) 33 .

41. SLD Electroweak Group, DELPHI, ALEPH, SLD, SLD Heavy Flavour Group, OPAL, LEP Electroweak Working Group, L3 collaboration, S. Schael et al., "Precision electroweak measurements on the Z resonance."*Phys. Rept.* **427** (2006) 257 [hep-ex/0509008].

42. M. S. Chanowitz, "The $Z \to \bar{b}b$ Decay Asymmetry: Lose-Lose For The Standard Model."*Phys. Rev. Lett.* **87** (2001) 231802 [hep-ph/0104024].

43. "Martin Grunewald, contact person of the Electroweak Working Group, private discussion." 2017.

44. D. Choudhury, T. M. P. Tait and C. E. M. Wagner, "Beautiful mirrors and precision electroweak data."*Phys. Rev.* **D65** (2002) 053002 [hep-ph/0109097].

45. ATLAS collaboration, M. Aaboud et al., "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector."*Phys. Lett. B* **784** (2018) 173 [1806.00425].

46. CMS collaboration, A. M. Sirunyan et al., "Observation of $t\bar{t}H$ production."*Phys. Rev. Lett.* **120** (2018) 231801 [1804.02610].

47. ATLAS collaboration, M. Aaboud et al., "Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector."*JHEP* **07** (2018) 089 [1803.09678].

48. ATLAS collaboration, M. Aaboud et al., "Combination of the searches for pair-produced vector-like partners of the third-generation quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector."*Phys. Rev. Lett.* **121** (2018) 211801 [1808.02343].

49. CMS collaboration, A. M. Sirunyan et al., "Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV."*JHEP* **08** (2018) 177 [1805.04758].

50. A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, "FeynRules 2.0 - A complete toolbox for tree-level phenomenology."*Comput. Phys. Commun.* **185** (2014) 2250 [1310.1921].

51. Wolfram Research, Inc., "Mathematica, Version 8.0." Champaign, IL (2010), https://www.wolfram.com/mathematica.

52. D. Liu, I. Low and C. E. M. Wagner, "Modification of Higgs Couplings in Minimal Composite Models."*Phys. Rev.* **D96** (2017) 035013 [1703.07791].

53. F. Zimmermann, M. Benedikt, M. Capeans Garrido, F. Cerutti, B. Goddard, J. Gutleber et al., "Future Circular Collider - European Strategy Update Documents." Tech. Rep. CERN-ACC-2019-0006, CERN, Geneva, Jan, 2019. https://cds.cern.ch/record/2653676.

54. FCC collaboration, A. Abada et al., "FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3."*Eur. Phys. J. ST* **228** (2019) 755.

55. CEPC Study Group, "CEPC Conceptual Design Report: Volume 1 - Accelerator." 1809.00285.

56. J. Gao, "CEPC and SPPC Status from the completion of the CDR towards TDR."*Journal of Modern Physics A* (2020) to be published.

57. C. Bautista, L. de Lima, R. Matheus, E. Pontón, L. Fernandes do Prado and A. Savoy-Navarro, "Probing the Top-Higgs Sector with Composite Higgs Models at Present and Future Hadron Colliders." 2020. to be submitted.

58. G. Panico and A. Wulzer, "The Composite Nambu-Goldstone Higgs."*Lect. Notes Phys.* **913** (2016) pp.1 [1506.01961].

59. A. Falkowski, F. Riva and A. Urbano, "Higgs at last."*JHEP* **11** (2013) 111 [1303.1812].

60. G. Buchalla, O. Cata and C. Krause, "A Systematic Approach to the SILH Lagrangian."*Nucl. Phys.* **B894** (2015) 602 [1412.6356].

61. V. Sanz and J. Setford, "Composite Higgs Models after Run 2."*Adv. High Energy Phys.* **2018** (2018) 7168480 [1703.10190].

62. A. Banerjee, G. Bhattacharyya, N. Kumar and T. S. Ray, "Constraining Composite Higgs Models using LHC data."*JHEP* **03** (2018) 062 [1712.07494].

63. J. de Blas, O. Eberhardt and C. Krause, "Current and Future Constraints on Higgs Couplings in the Nonlinear Effective Theory."*JHEP* **07** (2018) 048 [1803.00939].

64. K. Agashe, R. Contino, L. Da Rold and A. Pomarol, "A Custodial symmetry for $Zb\bar{b}$."*Phys. Lett.* **B641** (2006) 62 [hep-ph/0605341].

65. M. Carena, E. Pontón, J. Santiago and C. E. M. Wagner, "Light Kaluza Klein States in Randall-Sundrum Models with Custodial SU(2)."*Nucl. Phys.* **B759** (2006) 202 [hep-ph/0607106].

66. C. Anastasiou, E. Furlan and J. Santiago, "Realistic Composite Higgs Models."*Phys. Rev.* **D79** (2009) 075003 [0901.2117].

67. A. Orgogozo and S. Rychkov, "The S parameter for a Light Composite Higgs: a Dispersion Relation Approach."*JHEP* **06** (2013) 014 [1211.5543].

68. C. Grojean, O. Matsedonskyi and G. Panico, "Light top partners and precision physics."*JHEP* **10** (2013) 160 [1306.4655].

69. A. Pich, I. Rosell and J. J. Sanz-Cillero, "Oblique S and T Constraints on Electroweak Strongly-Coupled Models with a Light Higgs."*JHEP* **01** (2014) 157 [1310.3121].

70. C. Englert, A. Freitas, M. M. Muhlleitner, T. Plehn, M. Rauch, M. Spira et al., "Precision Measurements of Higgs Couplings: Implications for New Physics Scales."*J. Phys.* **G41** (2014) 113001 [1403.7191].

71. D. Croon, B. M. Dillon, S. J. Huber and V. Sanz, "Exploring holographic Composite Higgs models."*JHEP* **07** (2016) 072 [1510.08482].

72. A. Pomarol and F. Riva, "The Composite Higgs and Light Resonance Connection."*JHEP* **08** (2012) 135 [1205.6434].

73. G. Panico, M. Redi, A. Tesi and A. Wulzer, "On the Tuning and the Mass of the Composite Higgs."*JHEP* **03** (2013) 051 [1210.7114].

74. M. Montull, F. Riva, E. Salvioni and R. Torre, "Higgs Couplings in Composite Models."*Phys. Rev.* **D88** (2013) 095006 [1308.0559].

75. A. Carmona and F. Goertz, "A naturally light Higgs without light Top Partners."*JHEP* **05** (2015) 002 [1410.8555].

76. S. Kanemura, K. Kaneta, N. Machida, S. Odori and T. Shindou, "Single and double production of the Higgs boson at hadron and lepton colliders in minimal composite Higgs models."*Phys. Rev.* **D94** (2016) 015028 [1603.05588].

77. M. B. Gavela, K. Kanshin, P. A. N. Machado and S. Saa, "The linear-non-linear frontier for the Goldstone Higgs."*Eur. Phys. J.* **C76** (2016) 690 [1610.08083].

78. S. R. Coleman, J. Wess and B. Zumino, "Structure of phenomenological Lagrangians. 1.."*Phys. Rev.* **177** (1969) 2239.

79. C. G. Callan, Jr., S. R. Coleman, J. Wess and B. Zumino, "Structure of phenomenological Lagrangians. 2.."*Phys. Rev.* **177** (1969) 2247.

80. R. D. Ball et al., "Parton distributions with LHC data."*Nucl. Phys. B* **867** (2013) 244 [1207.1303].

81. R. Contino, M. Ghezzi, M. Moretti, G. Panico, F. Piccinini and A. Wulzer, "Anomalous Couplings in Double Higgs Production."*JHEP* **08** (2012) 154 [1205.5444].

82. E. Conte, B. Fuks and G. Serret, "MadAnalysis 5, A User-Friendly Framework for Collider Phenomenology."*Comput. Phys. Commun.* **184** (2013) 222 [1206.1599].

83. CMS Collaboration, "Measurement of $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel in $41.5\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$." Tech. Rep. CMS-PAS-HIG-18-030, CERN, Geneva, 5, 2019. https://cds.cern.ch/record/2675023.

84. F. Marcastel, "CERN's Accelerator Complex. La chaîne des accélérateurs du CERN.".

85. CMS collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV.".

86. CMS collaboration, S. Chatrchyan et al., "The CMS Experiment at the CERN LHC."*JINST* **3** (2008) S08004.

87. CERN AC, "Layout of CMS." Mar, 1998. `https://cds.cern.ch/record/39040`.

88. D. Barney, "CMS Detector Slice." Jan, 2016. CMS Collection, `https://cds.cern.ch/record/2120661`.

89. CMS collaboration, V. Karimäki, "The CMS tracker system project: Technical Design Report.".

90. A. Dominguez, D. Abbaneo, K. Arndt, N. Bacchetta, A. Ball, E. Bartz et al., "CMS Technical Design Report for the Pixel Detector Upgrade." Tech. Rep. CERN-LHCC-2012-016. CMS-TDR-11, CERN, Sep, 2012. https://cds.cern.ch/record/1481838.

91. C. Biino, "The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections."*J. Phys. Conf. Ser.* **587** (2015) 012001.

92. CMS collaboration, *The CMS hadron calorimeter project: Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 1997.

93. CMS Collaboration collaboration, J. G. Layter, *The CMS muon project: Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 1997.

94. C. Weiser, "A combined secondary vertex based B-tagging algorithm in CMS.".

95. CMS collaboration, A. Sirunyan et al., "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV."*JINST* **13** (2018) P05011 [1712.07158].

96. Particle Data Group collaboration, M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi et al., "Review of particle physics."*Phys. Rev. D* **98** (2018) 030001.

97. J. D. Jackson, *Classical Electrodynamics*. Wiley, 1998.

98. CMS Collaboration collaboration, M. Tosi, "The CMS trigger in Run 2." Tech. Rep. CMS-CR-2017-340, CERN, Geneva, Oct, 2017. 10.22323/1.314.0523, https://cds.cern.ch/record/2290106.

99. CMS collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV." Tech. Rep. CERN-EP-2020-065. CMS-TRG-17-001-003, CERN, Geneva, Jun, 2020. https://cds.cern.ch/record/2721198.

100. Z. Wu, M. Krohn, A. Henckel and Y. Gershtein, "Trigger cmsdas 2020." January, 2020, Lectures from the CMS Data Analysis School, Fermilab. `https://indico.cern.ch/event/843368/timetable/#10-trigger-exercise`.

101. "CMSSW on GitHub." `https://cms-sw.github.io/`.

102. R. Brun and F. Rademakers, "Root — an object oriented data analysis framework."*Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389** (1997) 81 .

103. CMS Collaboration collaboration, G. L. Bayatyan, M. Della Negra, Foà, A. Hervé and A. Petrilli, *CMS computing: Technical Design Report*, Technical Design Report CMS. CERN, Geneva, 2005.

104. CMS Collaboration collaboration, A. Rizzi, G. Petrucciani and M. Peruzzi, "A further reduction in CMS event data for analysis: the NANOAOD format."*EPJ Web Conf.* **214** (2019) 06021. 6 p.

105. R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli et al., "Higgs pair production at the LHC with NLO and parton-shower effects."*Phys. Lett.* **B732** (2014) 142 [1401.7340].

106. LHC Higgs Cross Section Working Group collaboration, D. de Florian et al., "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector." 1610.07922.

107. ATLAS collaboration, G. Aad et al., "Combination of searches for Higgs boson pairs in *pp* collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector."*Phys. Lett.* **B800** (2020) 135103 [1906.02025].

108. CMS collaboration, A. M. Sirunyan et al., "Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV."*Phys. Rev. Lett.* **122** (2019) 121803 [1811.09689].

109. CMS Collaboration, "Search for $\mathrm{t\bar{t}H}$, $\mathrm{h} \to \mathrm{b\bar{b}}$ production using the 2017 data sample." CMS Analysis Note 2018/235, CERN, 2018. `http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2018/235`.

110. S. Alioli et al., "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX." *JHEP* **06** (2010) 043 [1002.2581].

111. T. Sjostrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten et al., "An Introduction to PYTHIA 8.2." *Comput. Phys. Commun.* **191** (2015) 159 [1410.3012].

112. NNPDF collaboration, R. D. Ball et al., "Parton distributions from high-precision collider data." *Eur. Phys. J. C* **77** (2017) 663 [1706.00428].

113. CMS collaboration, A. M. Sirunyan et al., "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements." *Eur. Phys. J. C* **80** (2020) 4 [1903.12179].

114. N. Bartosik, "GenHFHadronMatcher tool." GitHub. https://github.com/cms-sw/cmssw/blob/CMSSW_9_4_X/PhysicsTools/JetMCAlgos/plugins/GenHFHadronMatcher.cc.

115. CMS Collaboration, "Study of the top quark and antiquark pair plus four b-quarks production as a background in the top-higgs sector searches." CMS Analysis Note 2020/053, CERN, 2020. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2020/053.

116. T. Ježo, J. M. Lindert, N. Moretti and S. Pozzorini, "New NLOPS predictions for $t\bar{t}+b$ -jet production at the LHC." *Eur. Phys. J.* **C78** (2018) 502 [1802.00426].

117. CMS Collaboration, "Measurement of tth and th in the h→bb channel with the full run 2 data sample." CMS Analysis Note 2019/094, CERN, 2019. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/094.

118. M. M. Horzela, "Merging of $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}+b\bar{b}}$ simulations for an improved background modeling for $\mathrm{t\bar{t}H(b\bar{b})}$ measurements." MSc thesis, Karlsruher Institut für Technologie (KIT), 2019. ETP-KA/2019-13.

119. CMS collaboration, A. Sirunyan et al., "Particle-flow reconstruction and global event description with the CMS detector." *JINST* **12** (2017) P10003 [1706.04965].

120. CMS collaboration, S. Chatrchyan et al., "Description and performance of track and primary-vertex reconstruction with the CMS tracker." *JINST* **9** (2014) P10009 [1405.6569].

121. CMS collaboration, V. Khachatryan et al., "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV." *JINST* **10** (2015) P06005 [1502.02701].

122. CMS collaboration, A. Sirunyan et al., "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV." *JINST* **13** (2018) P06015 [1804.04528].

123. "instructions for applying electron and photon id." https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Electron_efficiencies_and_scale.

124. "reference measurements and calibrations for run-ii." https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffs2017.

125. M. Cacciari, G. P. Salam and G. Soyez, "The anti-k_t jet clustering algorithm." *JHEP* **04** (2008) 063 [0802.1189].

126. M. Cacciari, G. P. Salam and G. Soyez, "FastJet User Manual." *Eur. Phys. J. C* **72** (2012) 1896 [1111.6097].

127. CMS Collaboration, "Pileup Removal Algorithms." Tech. Rep. CMS-PAS-JME-14-001, CERN, Geneva, 2014. https://cds.cern.ch/record/1751454.

128. M. Cacciari, G. P. Salam and G. Soyez, "The Catchment Area of Jets." *JHEP* **04** (2008) 005 [0802.1188].

129. CMS collaboration, V. Khachatryan et al., "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV." *JINST* **12** (2017) P02014 [1607.03663].

130. "jet identification." https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID13TeVRun2017.

131. "jet identification in high pile-up environment." https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJetID#Information_for_13_TeV_data_anal.

132. "jet energy resolution." https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Scaling_factors_and_Uncertai.

133. CMS collaboration, "Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector.".

134. "b-tagging discriminant shape calibration using event weights with a tag-and-probe method." https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagShapeCalibration.

135. CMS Collaboration collaboration, C. collaboration, "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with 2016 pp collision data at $\sqrt{s} = 13$ TeV." Tech. Rep. CMS-PAS-HIG-16-038, CERN, Geneva, 2016. http://cds.cern.ch/record/2231510.

136. CMS Collaboration, "Search for $t\bar{t}H$, $h \to b\bar{b}$ decays using the full 2016 data sample." CMS Analysis Note 2017/063, CERN, 2017. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/063.

137. CMS collaboration, A. M. Sirunyan et al., "Search for $t\bar{t}H$ production in the $H \to b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV." *JHEP* **03** (2019) 026 [1804.03682].

138. CMS collaboration, A. M. Sirunyan et al., "Search for $t\bar{t}H$ production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13$ TeV." *JHEP* **06** (2018) 101 [1803.06986].

139. F. Chollet et al., "Keras." 2015, GitHub. https://github.com/fchollet/keras.

140. I. Goodfellow, Y. Bengio and A. Courville, "Deep learning." 2016, MIT Press. http://www.deeplearningbook.org.

141. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization." *arXiv e-prints* (2014) arXiv:1412.6980 [1412.6980].

142. "Combine." GitHub. https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/.

143. ATLAS, CMS, LHC Higgs Combination Group collaboration, "Procedure for the LHC Higgs boson search combination in summer 2011.".

144. G. Cowan, K. Cranmer, E. Gross and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics." *Eur. Phys. J. C* **71** (2011) 1554 [1007.1727].

145. R. Barlow and C. Beeston, "Fitting using finite monte carlo samples." *Computer Physics Communications* **77** (1993) 219 .

146. J. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra." in *PHYSTAT 2011*, pp. 115–120, 2011, 1103.0354, DOI.

147. CMS collaboration, V. Khachatryan et al., "Search for the associated production of the Higgs boson with a top-quark pair." *JHEP* **09** (2014) 087 [1408.1682].

148. CMS Collaboration, "Search for $t\bar{t}HH$ in the semileptonic decay of the top pair and the higgs pair decay into b-quarks, using the 2017 data sample." CMS Analysis Note 2019/173, CERN, 2019. http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2019/173.

149. A. De Simone, O. Matsedonskyi, R. Rattazzi and A. Wulzer, "A first top partner hunter's guide." *Journal of High Energy Physics* **2013** (2013) .

150. G. Fox and S. Wolfram, "Event shapes in e+e- annihilation." *Nuclear Physics B* **157** (1979) 543 .