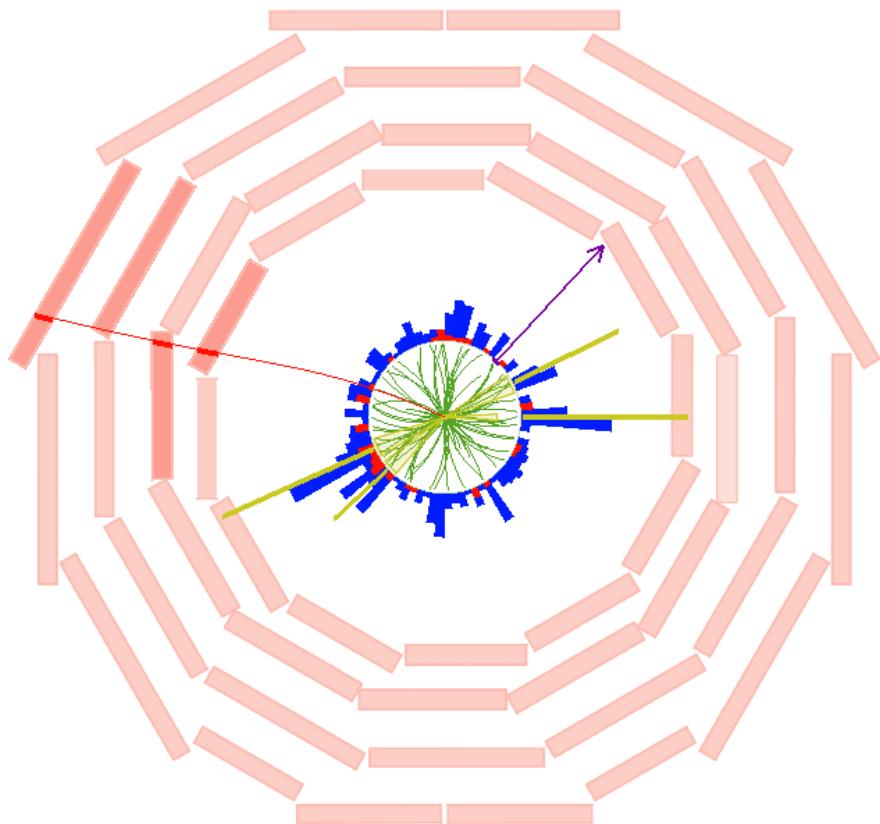


Search for a heavy Higgs boson decaying into a pair of top quarks
with the CMS 13 TeV dataset

Muhammad Gul

Supervisors: Dr. Michael Tytgat, Prof. Dr. Didar Dobur

A dissertation submitted to Ghent University in partial fulfilment of the requirements for
the degree of Doctor of Science: Physics
Academic year 2017-2018



Acknowledgements

My thesis is based on a joint project and I would like to thank all those who contributed directly and indirectly. I would like to especially thank my supervisors, colleagues, friends, and family members. Without their support, I would not have been able to perform this interesting Higgs research and make the DCS project.

First, I am extremely thankful to *Prof. Dr. Dirk Ryckbosch* and *Dr. Michael Tytgat* who gave me a chance to come out of a troubled area in Pakistan (FATA). This opened a door for me to participate in international collaborations and meet people all over the world. You always welcomed me whenever I faced a problem. Thank you so much for your support while sending me to CERN for long- and short-time visits. Thank you *Prof. Dr. Didar Dobur* for your support and always welcoming attitude.

I am thankful to Efe Yazgan and Bugra Bilin for all of their technical support at the beginning of the project. You introduced me to the big team of the heavy Higgs to ttbar searches, and we eventually made it public.

The members of the heavy Higgs to ttbar group have contributed immensely to my work during the data analysis. The group provided me support at every level when I was stuck. Andrey Popov, working with you always gave me a new dimension of knowledge. Every time I spoke to you, I felt like I learned something new. I am very thankful to you and all members of the group – Viola Sordini, Mauro Verzetti, Alexander Grohsjean, and Jan Steggemann – for your constant support. I had a wonderful and unforgettable time with you and learnt a lot.

I would like to thank all members of the GIF++ RPC team and the CERN engineering department for their help during the development of the DCS system. I would especially like to thank Nicolas Zaganidis, Gabriella Pugliese, and Marino Romano for their support and insight whenever I was stuck.

I wish to thank the members of my dissertation committee: Prof. Dr. Dirk Ryckbosch, Prof. Dr. Jan Ryckebusch, Prof. Dr. Natalie Jachowicz, Prof. Dr. Van Mechelen Pierre, Prof. Dr. Luc Van Hoorebeke, Prof. Dr. Andrea Giammanco for generously offering their time, support, guidance and good will throughout the preparation and review of this document.

I am grateful to all my INW and Gent friends. It was indeed a wonderful time that I spent with you. Thank you, Sam De Ridder, for helping me with the Dutch translation of the summary of this thesis. Thanks to all my Pakistani friends and colleagues. Prof. Hafeez, I am very thankful for your support throughout my career. I am grateful for the time I spent at CERN with Mehar, Qamar, Waqas, Waqar, Ammara, Bilal, and Ahmad and our memorable trips to the mountains. Thank you Muhammad Islam, Sherin Khan, and Nasar Ali for all your support during my life.

Lastly, I would like to thank my family for all their love and encouragement. For my parents, who raised me with a love of science and supported me in all my endeavours. And most of all, for my loving, supportive, encouraging, and patient wife, Kalsoom Gul and my cute children, Muhammad Shoraim and Muhammad Sudais, whose faithful support during the whole process is highly appreciated. I am sorry that I spent most of my time away from you, but I promise I will join you soon. Thank you.

Muhammad Gul
Dept. of Physics and Astronomy,
Ghent University,
Belgium

Table of contents

Acknowledgements	i
1 Introduction	1-1
2 The Standard Model and beyond	2-1
2.1 The Standard Model	2-1
2.1.1 The Top quark	2-5
2.1.2 The SM Higgs boson	2-8
2.2 Beyond the Standard Model	2-10
2.2.1 The Hierarchy problem	2-10
2.2.2 Supersymmetry	2-11
2.2.3 Minimal Supersymmetric Standard Model	2-11
2.2.4 Higgs sector of MSSM	2-12
2.2.5 Heavy Higgs production and decay at LHC	2-15
2.2.6 The $\Phi \rightarrow t\bar{t}$ channel with interference from the SM $gg \rightarrow t\bar{t}$	2-17
3 The CMS experiment at the Large Hadron Collider	3-1
3.1 The Large Hadron Collider at CERN	3-1
3.2 The CMS experiment at the LHC	3-3
3.2.1 Tracker	3-4
3.2.2 Electromagnetic calorimeter (ECAL)	3-5
3.2.3 Hadronic calorimeter (HCAL)	3-6
3.2.4 Magnet	3-7
3.2.5 Muon spectrometer	3-8
3.2.6 Trigger and data acquisition systems	3-9
4 Detector control system for RPCs at GIF++	4-1
4.1 The HL-LHC program and the CMS detector upgrade	4-1
4.2 WinCC-OA	4-5
4.3 The CMS RPC DCS project at GIF++	4-6
4.3.1 High and low voltage system	4-7
4.3.2 Environmental and gas parameters	4-7
4.3.3 High voltage scan and stability test	4-8
4.3.4 Database	4-8
4.4 CMS RPC longevity studies	4-9
4.4.1 RPCs setup at GIF++ for efficiency studies	4-10
4.4.2 Efficiency and cluster size results	4-11
4.5 Summary	4-11

5 Monte Carlo simulation and event reconstruction	5-1
5.1 Monte Carlo simulation	5-1
5.1.1 Parton distribution functions	5-1
5.1.2 Hard scattering	5-2
5.1.3 Particle physics generators	5-2
5.1.4 Parton showering	5-3
5.1.5 Hadronization	5-4
5.1.6 Event simulation	5-4
5.2 Event reconstruction	5-4
5.2.1 Track and vertex reconstruction	5-5
5.2.2 Primary vertex reconstruction	5-5
5.3 Physics objects reconstruction	5-5
5.3.1 Muon reconstruction	5-6
5.3.2 Electron reconstruction	5-6
5.3.3 Jet reconstruction	5-7
5.3.4 Heavy flavour jet identification	5-7
5.3.5 Missing transverse energy	5-8
5.4 Signal modelling	5-8
5.4.1 Analytical calculation	5-9
5.4.2 Event generation and validation	5-11
5.5 Cross section and BR calculation	5-12
5.5.1 2HDMC and SUSHI	5-12
5.5.2 TOP++	5-17
6 Search for a heavy higgs state in the semileptonic channel	6-1
6.1 Datasets	6-2
6.1.1 Collision data	6-2
6.1.2 Simulation	6-2
6.2 Corrections to simulation	6-5
6.2.1 Pile-up re-weighting	6-5
6.2.2 Jet Energy Scale and Resolution	6-5
6.2.3 B tagging	6-5
6.2.4 Trigger efficiency	6-6
6.2.5 Muon and electron efficiency scale factors	6-8
6.2.6 Generator level weights	6-8
6.2.7 Correction of p_T spectrum of top quarks in $t\bar{t}$ events	6-11
6.3 Physics objects selection	6-12
6.4 Event selection	6-15
6.5 Search variables	6-17
6.5.1 Reconstruction of the $t\bar{t}$ system	6-17
6.5.2 Search variables	6-18
6.6 Data-driven modelling of the multijet background	6-20
6.6.1 Normalization	6-21
6.6.2 Shape	6-26
6.7 Systematic uncertainties	6-29
6.7.1 Experimental uncertainties	6-30
6.7.2 Theory uncertainties	6-30
6.7.3 MC statistical uncertainty	6-39
6.8 Statistical analysis	6-39

6.9	Summary	6-42
7	Combination of semileptonic and dileptonic channels	7-1
7.1	Statistical evaluation and combination	7-1
7.1.1	Systematic uncertainties	7-1
7.2	Interpretation	7-2
7.2.1	Interpretation in the hMSSM	7-2
7.2.2	Higher-order cross sections	7-4
7.3	Interpolation and extrapolation to different masses and widths	7-6
7.3.1	Mass and width interpolation	7-6
7.3.2	Width extrapolation	7-9
7.4	Results	7-11
7.4.1	Model-independent limits	7-11
7.4.2	Limits in the hMSSM	7-13
7.5	Summary	7-13
8	Summary and Conclusion	8-1
9	Nederlandstalige samenvatting	9-1
A	Signal distributions	i
B	Validation of the generation procedure for interference	iii
C	Data MC comparison	xi
D	Impact of systematic uncertainties	xv
D.0.1	Single uncertainties	xv
D.0.2	Groups	xv
D.0.3	Nuisance parameter pulls, constraints, and correlations	xvi

List of acronyms

ADC	Analog to Digital Converter
ALICE	A Large Ion Collider Experiment
AOD	Analysis Objects Data
APD	Avalanche PhotoDiodes
ATLAS	A Toroidal LHC ApparatuS
BDT	Boosted Decision Tree
BE	Back-End
BEH	Brout-Englert-Higgs
BR	Branching Ratio
BTV POG	b-Tag & Vertexing Physics Object Group
BW	Breit-Wigner
CAEN	Costruzioni Apparecchiature Elettroniche Nucleari S.p.A.
CDF	Collider Detector at Fermilab
CERN	Conseil Européen pour la Recherche Nucléaire
CHS	Charged Hadron Subtraction
CEST	Central European Summer Time
CKM	Cabibbo–Kobayashi–Maskawa
CL	Confidence Level
CMS	Compact Muon Solenoid
cMVA	combining MutiVariate Analysis
CoM	Center of Mass
CP	Charge Parity
CPU	Central Processing Unit
CSC	Cathode Strip Chamber
CSV	Combined Secondary Vertex
CTF	Combinatorial Track Finder
DB	Data Base
DCS	Detector Control System
DE	Dark Energy
DGLAP	Dokshitzer–Gribov–Lipatov–Altarelli–Parisi
DIM	Distribute Information Management
DIP	Data Interchange Protocol
DM	Dark Matter
DoF	Degrees of Freedom
DT	Drift Tube
DY	Drell-Yan
EASY	Embedded Assembly SYstem
EB	ECAL Barrel
ECAL	Electromagnetic CALorimeter
EE	ECAL Endcaps

EG	Event Generators
EGM	Electron & Gamma (photon)
EM	Electromagnetic
ES	ECAL preShower
ETM	Company Name
EV	Event Manager
EWSB	Electroweak Symmetry Breaking
fb	femto-barn
FE	Front-End
FEB	Front-End Board
FH	FEYNHIGGS
FSM	Finite State Machine
FSR	Final State Radiations
GE	GEM detector at Endcap
GEANT	GEometry ANd Tracking - a series of software toolkit platforms developed by CERN
GEM	Gaseous Electron Multiplier
GeV	Giga electron Volts
gg	gluon-gluon
GIF	Gamma Irradiation Facility
GIF++	new Gamma Irradiation Facility
GRPC	Glass-RPCs
GSF	Gaussian-Sum Filter
GUI	Graphical User Interface
GWS	Glashow, Weinberg and Salam
HB	HCAL Barrel
HCAL	Hadronic CALorimeter
HD	HDECAY
2HDM	two Higgs Doublet Model
2HDMC	2HDM Calculator
HE	HCAL Endcaps
HEP	High Energy Physics
HERA	Hadron-Electron Ring Accelerator
HF	HCAL Forward
HL-LHC	High Luminosity LHC
HLT	High Level Trigger
HO	HCAL Outer
HV	High Voltage
ID	IDentification
IP	Internet Protocol
IP	Impact Parameter
iRPC	Improved RPC
ISR	Initial State Radiations
JCOP	Joint Control Project
JEC	Jet Energy Corrections
JER	Jet Energy Resolution
JP	Jet Probability
KF	Kalman Filter
LEP	Large Electron-Positron

LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LHE	LesHouches Events
LINAC	LINear ACcelerators
LO	Leading Order
LS1	First Long Shutdown
LS2	Second Long Shutdown
LS3	Third Long Shutdown
LSP	Lightest Supersymmetric Particle
LV	Low Voltage
MAD	Material Access Door
MB1	Muonsystem at Barrel (DTs)
MB	Mega Bites
MC	Monte Carlo
MCNP	Monte Carlo N-Particle
ME1	Muonsystem at Endcaps (CSCs)
MeV	Mega electron Volts
ML	Maximum Likelihood
MS	Minimal Subtraction
MSSM	Minimal Supersymmetric Standard Model
MV	Mega Volts
NLO	Next-to-Leading Order
NNLL	Next-to-Next-to-Leading Logarithm
NNLO	Next-to-Next-to-Leading Order
ns	nano second
OA	Open Architecture
OLE	Object Linking and Embedding
OPC	OLE for Process Control
ORACLE	Name of a Corporation
PAG	Physics Analysis Group
PCA	Principal Component Analysis
PDF	Parton Distribution Function
pdfs	probability density functions
PDG	Particle Data Group
PF	Particle Flow
PLC	Programmable Logic Controllers
POWHEG	PositiveWeight Hardest Emission Generator
pp	proton-proton
PPE	Personal Protection Entrance
PPG	Personal Protection Gate
PPX	Personal Protection Exit
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
PU	Pile Up
PV	Primary Vertex
p-value	probability value
PVSS	old name of WinCC-OA
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics

qq	quark-quark
RB	RPCs at Barrel
RDB	Relational DataBase
RE	RPCs at Endcap
RF	Radio Frequency
RMS	Root Mean Square
ROOT	a framework for data processing born at CERN
RPC	Resistive Plate Chamber
SC	Super Cluster
SCADA	Supervisory Control And Data Acquisition
SF	Scale Factors
SIMATIC	An automation system developed by Siemens
SMI	State Machine Interface
SPS	Super Proton Synchrotron
SQL	Structured Query Language
SUSHI	SUperSymmetric Higgs
SV	Secondary Vertex
TBq	Tera Becquerel
TCP	Transmission Control Protocol
TDC	Time-to-Digital Converter
TEC	Tracker EndCaps
TeV	Tera electron Volts
TIB	Tracker Inner Barrel
TID	Tracker Inner Disk
TOB	Tracker Outer Barrel
UE	Underlying Events
UTC	Coordinated Universal Time
VBF	Vector Boson Fusion
VPT	Vacuum PhotoTriodes
vev	vacuum expectation value
WWP	Weighted-Weighted Print (chi2, ndf, p-value, igood)

1

Introduction

Elementary particle physics aims to answers the most fundamental question asked by human kind – “What is matter made of?” The journey of particle physics began from atomic theory that considers the atom to be the building block of matter. With the passage of time, scientists realised that the atom can be further disintegrated into smaller parts. The last two centuries have been the bright era of particle physics, where the knowledge pertaining to the fundamental building blocks of nature has developed rapidly, leading to changing ideas of what really is fundamental. The experimental investigation of the particles, carried out using detectors and accelerators, started from the discovery of x-rays, followed by electrons, protons, neutrons, and a zoo of other particles till the recent discovery of the Higgs boson. The detectors and accelerators used in the field of particle physics originated from simple technology like Geiger Counter, Cloud Chamber, etc., which evolved with time. Now we have the world’s most powerful Large Hadron Collider (LHC) at the European Laboratory for Particle Physics (CERN) in Geneva. The LHC is going to be upgraded to High Luminosity LHC (HL-LHC) machine to increase the luminosity further by a factor of 10 in order to improve statistically marginal measurements.

Theory and experiment run parallelly, and in the last few decades, a theory has emerged that describes all the known elementary particle interactions, except gravity, known as the Standard Model (SM) of particle physics. The SM is not the final word on the subject and has great scope for being extended, but at least we now have (for the first time) a full deck of cards to play with. The discovery of the Higgs boson completed the SM but left some unanswered questions. Why is the mass of the Higgs boson 125 GeV? What is the origin of Dark Matter (DM) and Dark Energy? What is the origin of gravity? And so forth. Many beyond the SM (BSM) theories have been proposed to answer these questions and among them, supersymmetry (SUSY) is the leading possibility.

The LHC successfully began its operations in 2009 at low energy (1.18 TeV per beam); on 30th March, 2010 at 13:06 CEST, two beams were collided at the centre of mass (CoM) energy, $\sqrt{s} = 7 \text{ TeV}$ (3.5 TeV per beam), setting a record of high-energy pp collision. The CoM energy further increased to 8 TeV in April 2012 and the first new particle was discovered on 4th July, 2012 – the Higgs boson. The combined 7 and 8 TeV data collected during the LHC Run did not point to BSM physics. After the first long shutdown, the machine began operation in May, 2015 by colliding protons at the record-breaking collision CoM energy of 13 TeV (Run II), successfully delivering $\mathcal{L} \approx 80 \text{ fb}^{-1}$ integrated luminosity in 2016–17. The high luminosity might provide answers to some of the questions through the discovery of new particle(s), applying limits on lower masses in the kinematic plane [mass, $\tan \beta$]. This will allow for a more precise test of the properties of the Higgs boson and top quark pair and the exploration of its rare decay modes, where deviations from the SM might be found. It will also open a window for searching for particles using high-mass ranges up to TeV scale in the heaviest SM particles’ final state.

The extensions of the SM, such as 2HDM and hMSSM predict more than one Higgs boson. There are five Higgs states – a light SM-like scalar h , a heavy scalar H , a pseudoscalar A , and two charged Higgs bosons

H^\pm . The work presented in this thesis aims to search for heavy Higgs, \mathcal{CP} -even neutral H and \mathcal{CP} -odd neutral A, in the $t\bar{t}$ final state in the context of the so-called hMSSM that takes the existence of the 125 GeV Higgs boson directly into account. The large mass of the top quark naturally leads to a strong coupling between it and the additional Higgs bosons.

The High Luminosity Large Hadron Collider (HL-LHC) machine will induce higher background radiations as compared to the current operating conditions. It is important to study the performance and stability of the currently installed and future detectors in a high radiation environment. The CERN engineering and physics departments designed and built a dedicated Gamma Irradiation Facility (GIF++), where high-energy charged particle beams, mainly muon beams with a momentum up to 100 GeV/c, are combined with a 13.9 TBq ^{137}Cs source. The GIF++ cumulates doses equivalent to the HL-LHC experimental conditions within a reasonable time frame. The CMS muon group installed many RPC detectors to test for the HL-LHC conditions. To operate the installed detectors, I designed and applied a dedicated detector control system project for the CMS RPCs at the GIF++ facility.

I began my PhD at the beginning of 2014 using 8 TeV data for the charged Higgs searches, but we soon realized that the channel was not feasible owing to the small cross section. I moved towards the search for the neutral heavy Higgs in the $t\bar{t}$ final state and started from the generator level. I prepared the signal at 13 TeV using fast simulation and then full simulation for the 2015 data set, processed all the data and background samples, applied all corrections to MC simulations and synchronized with other groups. LHC delivered about 2 fb^{-1} data in 2015, which was not sufficient for this search. We further extended the search to analyse the full 2016 data set.

The analysis included in this thesis is a team effort, wherein I contributed with regard to the signal simulation, made synchronizations with other groups in our team, processed the entirety of the 2016 data and all simulated backgrounds, applied corrections to data and simulations, and compared data to backgrounds that showed good agreement. Further, my contribution involved making QCD templates for the final statistical analysis using data-driven techniques. I also applied top p_T correction scale factors from theory to our simulated $t\bar{t}$ sample and compared them with the correction taken from real data. I calculated the k-factors to scale the signal both in semileptonic and dileptonic final states to higher order cross sections in order to interpret the results in the context of hMSSM.

The thesis is structured in the following way: Chapter 2 describes the SM of physics, which mainly focuses on the top quark and SM Higgs boson. It further provides a detailed overview of Higgs physics in BSM, especially the heavy neutral Higgs in the $t\bar{t}$ final state with interference effects from the SM $t\bar{t}$. This provides the theoretical background for the search presented in the thesis. Chapter 3 briefly covers the LHC and CMS experiments. The High Luminosity LHC (HL-LHC), the Gamma Irradiation Facility (GIF++) setup designed to test the installed and new detectors for the HL-LHC conditions, the detector control system, based on the WinCC-OA framework, for the operation of the CMS Resistive Plate Chamber (RPC) installed in the GIF++ and the RPC performance results will be described in Chapter 4. The Monte Carlo simulation, particle physics generators, object reconstruction in the CMS detector, and signal modelling will be discussed in Chapter 5. Chapter 6 describes the full data analysis in the semileptonic state, including corrections to MC simulations, physics objects and event selection, search variables, multijet estimation from data driven techniques, experimental, theoretical and MC statistical uncertainties as well as the statistical analysis for semileptonic channel. The strategy used for combining semileptonic and dileptonic channels, masses and widths interpolation and extrapolation, and model independent and model dependent limits will be provided in Chapter 7. The paper concludes with a summary and outlook in Chapter 8.

2

The Standard Model and Beyond

This chapter describes the Standard Model (SM) of particle physics, the physics of top quark and the SM Higgs boson. The discussion follows the introduction to the SM extension, SUSY, MSSM, and model-dependent heavy Higgs searches. The last part is devoted to the search of neutral heavy Higgs decays into a pair of top quarks ($H \rightarrow t\bar{t}$) with QCD $t\bar{t}$ interference, observables for signal and background discrimination and sensitivity at higher luminosities.

2.1 The Standard Model

The SM theory describes matter in terms of elementary particles, fermions, all having a half-integer spin and the three fundamental forces with mediators and gauge bosons (integer spin). Fermions are further divided into two classes based on their properties – quarks and leptons. The three fundamental forces are electromagnetic (EM); strong; and weak; and they are mediated by photon, gluon, and gauge bosons (W^\pm , Z) respectively. The Higgs mechanism is an important and fundamental feature of the SM to describe symmetry breaking that is responsible for giving masses to all fundamental particles. The discovery of the SM Higgs boson by the CMS and ATLAS experiments [1, 2] in 2012 at the LHC at CERN solved the mystery. The most familiar force, “ gravity”, is not included in the SM because the current energy scale of the accelerators can’t cope with strength of gravity at a particle level, which is extremely weak as compared to the strong force (10^{38}). The SM of particle physics is a renormalizable quantum field theory based on a local unitary product group $SU(3) \times SU(2) \times U(1)$, where the subgroup $SU(3)$ represents the strong interaction and $SU(2) \times U(1)$ is the unification of the electromagnetic and weak interaction (electroweak). Table 2.1 provides a summary of the SM particles along with their properties; the three fundamental forces are described in the following section.

- **Strong Nuclear Force:** is the fundamental force experienced by quarks and gluons, where the colour quantum numbers (red, blue, green) of particles play an important role, such as a particle charge in electromagnetism. It is responsible for binding quarks inside the nucleus via exchange of gluons within a short range of 10^{-15} m. The force is independent of the charge of particles and is explained by the Quantum Chromodynamics (QCD) theory. QCD depends on the non-Abelian gauge group $SU(3)$; hence, gluons are self-interacting. The strength of the strong coupling constant is $\sqrt{4\pi\alpha_s(Q^2)}$ and the coupling constant is, in fact, not constant at all but depends on the separation distance between the particles, known as the “running” coupling constant. The strong coupling constant increases at a short distance (the size of proton) and decreases when the distance becomes large. This phenomenon is known as **asymptotic freedom**. QCD plays a key role in the high-energy physics regime that is present in the LHC

Leptons				
Particle	Spin	Charge	Interaction	Mass (GeV)
Electron (e)	$\frac{1}{2}$	-1	e, w	0.511×10^{-3}
Electron Neutrino (ν_e)	$\frac{1}{2}$	0	w	$< 2.20 \times 10^{-9}$
Muon (μ)	$\frac{1}{2}$	-1	e, w	0.105
Muon Neutrino (ν_μ)	$\frac{1}{2}$	0	w	$< 0.17 \times 10^{-6}$
Tau (τ)	$\frac{1}{2}$	-1	e, w	1.776
Tau Neutrino (ν_τ)	$\frac{1}{2}$	0	w	$< 15.5 \times 10^{-6}$
Quarks				
Particle	Spin	Charge	Interaction	Mass (GeV)
Up Quark (u)	$\frac{1}{2}$	$+\frac{2}{3}$	c, e, w	2.2×10^{-3}
Down Quark (d)	$\frac{1}{2}$	$-\frac{1}{3}$	c, e, w	4.7×10^{-3}
Charm Quark (c)	$\frac{1}{2}$	$+\frac{2}{3}$	c, e, w	1.27
Strange Quark (s)	$\frac{1}{2}$	$-\frac{1}{3}$	c, e, w	96.0×10^{-3}
Top Quark (t)	$\frac{1}{2}$	$+\frac{2}{3}$	c, e, w	173.2
Bottom Quark (b)	$\frac{1}{2}$	$-\frac{1}{3}$	c, e, w	4.18
Gauge Bosons				
Particle	Spin	Charge	Interaction	Mass (GeV)
Photon (γ)	1	0	e	0
W Boson (W)	1	± 1	e, w	80.3
Z Boson (Z)	1	0	w	91.1
Gluon (g)	1	0	c	0
Scalar Bosons				
Particle	Spin	Charge	Interaction	Mass (GeV)
Higgs Boson (h)	0	0	w	125.5

Table 2.1: Summary of the SM of particle physics. First column – particle name (symbol); second column – spin; third column – charge; fourth column – corresponding interaction, where c stands for colour, e for electromagnetic, and w for weak force. The last column shows masses of the particles. The values are taken from the Particle Data Group (PDG) [3].

where the colliding particles are hadrons (protons). The full Lagrangian of QCD is given in Eq. 2.1:

$$\begin{aligned} \mathcal{L}_{QCD} = & -\frac{1}{4}(\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a) + \sum_f q_f^{-\alpha}(i\gamma^\mu \partial_\mu - m_f)q_f^\alpha \\ & + g_s G_a^\mu \sum_f q_f^{-\alpha} \gamma_\mu \left(\frac{\lambda^a}{2} \right)_{\alpha\beta} q_f^\beta \\ & - \frac{g_s}{2} f^{abc} (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu) G_\mu^b G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade} G_b^\mu G^\nu c G^d \mu G_\nu^e \end{aligned} \quad (2.1)$$

The first line involves the correct kinetic terms for the different fields, which give rise to the corresponding propagators. The second line contains $SU(3)_C$ matrices and λ^a provides the colour interaction between gluons and quarks. The last line gives the self-interactions in gluons generated by $G_a^{\mu\nu} G_{\mu\nu}^a$; g_s is the strength of the coupling constant [4].

- **Electromagnetic Interaction:** is a fundamental long-range force that acts upon charge particles and is mediated by virtual photons. It is described by the Quantum Electrodynamics (QED) theory, based on Abelian gauge group $U(1)_{EM}$, where the coupling constant is $\alpha = \frac{e^2}{4\phi} \approx \frac{1}{137}$, much smaller than unity, where “e” shows the electric charge of particle. The coulomb force between the two static charged

particles, q_1 and q_2 , is given by Eq. 2.2:

$$F = k \frac{q_1 q_2}{r^2} \quad (2.2)$$

Where k is the proportionality constant and r is the distance between the centre of the charged particles. An example of QED is the Bhabha scattering of the $e^+e^- \rightarrow e^+e^-$, as shown in Fig.2.1.

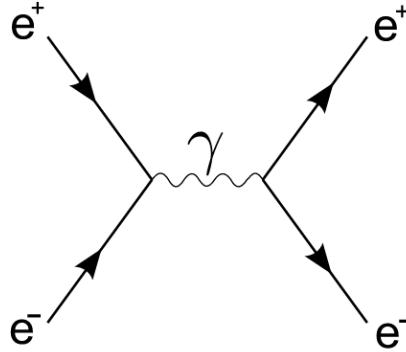


Figure 2.1: An overview of the Bhabha scattering in QED.

- **Weak Interaction:** All fundamental particles, except gluon (g) and photon (γ), experience weak interactions by exchanging or producing W^\pm or Z bosons. It is a short-range (10^{-17} to 10^{-16} m) interaction and responsible for many phenomena, such as radioactivity. Weak isospin, given in Eq. 2.5, conserves in weak interaction and plays a similar role in weak interactions as colour and charge in strong and electromagnetic interactions respectively. In case of leptons, the theory is very clear, like the decay of muon mediated by W^- boson ($\mu^- \rightarrow e^- + \nu_\mu + \bar{\nu}_e$) and electron-neutrino scattering via Z boson ($\nu_\mu + e^- \rightarrow \nu_\mu + e^-$). The lepton number is conserved in both processes. A similar model is applied to three generations of quarks, where a quark converts into the same generation; but in nature, cross-generation conversion is observed. The dilemma was solved by introducing the 3×3 **CKM** (Cabibbo–Kobayashi–Maskawa) matrix 2.3, with diagonal elements very near to unitary.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (2.3)$$

Down-type quarks (d' , s' , b') are a superposition of down-type quarks following Eq.2.3 and couple to up-type quarks via charged current weak interaction. The matrix elements provide the rate at which one quark converts to another, e.g. V_{ud} gives the transition rate of $d \rightarrow u$. All the nine CKM matrix values are given in Eq. 2.4.

$$V_{\text{CKM}} = \begin{pmatrix} 0.97434^{+0.00011}_{-0.00012} & 0.22506 \pm 0.0005 & 0.00357 \pm 0.00015 \\ 0.22492 \pm 0.0005 & 0.97351 \pm 0.00013 & 0.0411 \pm 0.0013 \\ 0.00875^{+0.00032}_{-0.00033} & 0.0403 \pm 0.0013 & 0.99915 \pm 0.00005 \end{pmatrix} \quad (2.4)$$

The diagonal elements are ≈ 1 , which shows that the transition in the same generation is dominant [5].

- **The Electroweak Symmetry:** The electromagnetic and weak forces were combined by Glashow, Weinberg, and Salam through a unique gauge electroweak theory (GWS) based on the $SU(2)_L \times U(1)_Y$ group [6]. The theory introduces three $SU(2)_L$ gauge bosons W_μ^i , $i = 1, 2, 3$, with coupling strength g_w , and one $U(1)_Y$ gauge boson, B_μ , with coupling constant $g'/2$. The GWS theory depends upon the chirality of particles, which means that it transforms left-handed particles as doublets with weak isospin (I) $\pm \frac{1}{2}$ and right-handed particles as singlet with $I = 0$. The hypercharge (Y) and the third component of weak isospin (I^3) are related to charge (Q in unit of e) as:

$$Q = I^3 + \frac{1}{2}Y \quad (2.5)$$

The wave functions of the charged fields W_μ^\pm , neutral field Z_μ , and photon A_μ are represented by:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp W_\mu^2) \\ Z_\mu &= (W_\mu^3 \cos \theta_w - B_\mu \sin \theta_w) \\ A_\mu &= (W_\mu^3 \sin \theta_w + B_\mu \cos \theta_w) \end{aligned} \quad (2.6)$$

Where θ_w is the weak mixing angle given in terms of electromagnetic coupling constant (g_e) as:

$$g_w \sin \theta_w = g' \cos \theta_w = g_e \quad (2.7)$$

In GWS theory, breaking of the underlying $SU(2)_L \times U(1)_Y$ symmetry predicts the existence of two charged gauge fields and two neutral gauge fields shown in Eq. 2.6.

- The Higgs Mechanism:** SM is based on gauge invariance that forbids introducing explicit mass terms in the Lagrangian of a chiral theory, as it breaks the symmetry. This results to massless SM particles – a contradiction to nature. A mechanism is needed for introducing the mass term in the SM without spoiling the gauge symmetry, known as the Higgs mechanism published by three groups independently and therefore also called the Brout-Englert-Higgs (BEH) mechanism [7, 8]. The GWS uses the BEH mechanism that introduces a new doublet of complex scalar field, known as the Higgs field, preserves symmetry under the gauge transformations, and acquires a non-zero vacuum expectation value, spontaneous electroweak symmetry breaking. The Higgs field contains a total of four additional degrees of freedom, where three of them become massive, given by the relation 2.6, after breaking the symmetry and one remains massless, the photon γ . The Higgs doublet introduced in the $SU(2)_L$ with hypercharge ($Y = \frac{1}{2}$) is in the form of:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix} \quad (2.8)$$

The SM electroweak Lagrangian density is given in term of

$$\mathcal{L}_{BEH} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi), \quad \text{with} \quad V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (2.9)$$

Where the first term of the Lagrangian D^μ corresponds to the covariant derivative of the Higgs field under the $SU(2) \times U(1)$ group, and the second term is the scalar potential known as “Mexican Hat” potential, shown graphically in Fig. 2.2. The Higgs potential has chosen to achieve the necessary spontaneous

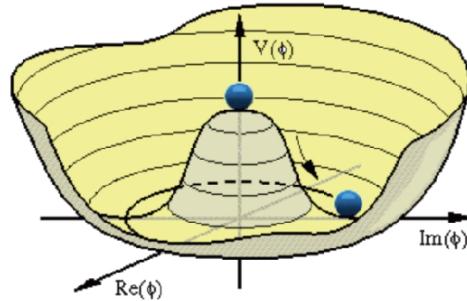


Figure 2.2: The scalar potential $V(\phi)$ with $\mu^2 < 0$ has a “Mexican hat shape”, with the minimum in a rim around the origin. Moving from the origin down to the minimum, the symmetry breaks and the scalar field acquires a vacuum expectation value (vev) in the process

symmetry breaking in the simplest form while being renormalizable. The Higgs potential $V(\phi)$ depends on the values of λ and μ^2 . For a stable vacuum, the parameter λ has to be a positive value. If $\mu^2 > 0$, the scalar potential has a global minimum, $\langle 0|\phi|0 \rangle = 0$, where no spontaneous symmetry breaking occurs. While for $\mu^2 < 0$, the symmetry breaks and the Higgs field acquires a non-zero vacuum expectation value with conditions $\langle 0|\phi|0 \rangle = \frac{-\mu^2}{2\lambda} = \frac{\nu}{\sqrt{2}}$ where $\nu = \sqrt{-\frac{\mu^2}{\lambda}}$ is the vev. The scale of electroweak

symmetry breaking is related to the W boson mass and to the Fermi constant G_F :

$$\nu = 2 \frac{m_w}{g_w} = (\sqrt{2}G_F)^{-\frac{1}{2}} \approx 240 \text{ GeV} \quad (2.10)$$

The Higgs field can be re-defined in terms of a perturbation around its non-zero vev by inserting a new scalar field $h(x)$:

$$\phi(x) = \frac{1}{2} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \quad (2.11)$$

The new scalar field corresponds to the Higgs particle with mass $m_h = \nu\sqrt{2\lambda}$ and no charge. Rearranging Eq. 2.9 by inserting ϕ from Eq. 2.11 and defining the bosonic states as shown in expression 2.6, we can see that the three bosons (W^\pm, Z^0) acquire masses while photon γ remains massless.

$$m_W^2 = \frac{\nu^2 g^2}{4}; m_Z^2 = \frac{\nu^2(g^2 + g'^2)}{4}; m_\gamma = 0 \quad (2.12)$$

The concept of the gauge boson mass terms can be further extended to the mass terms of the fermions in a different manner that would also break the local gauge invariance of the theory. This can be done by introducing the Yukawa coupling λ_f between the fermion fields and the Higgs boson with the coupling constant proportional to the respective fermion mass. By doing this, we can obtain the fermion mass:

$$m_f = \frac{\lambda_f}{\sqrt{2}}\nu \quad (2.13)$$

In SM, neutrinos are considered massless. The Higgs mechanism successfully serves in the GWS model to provide masses to weak bosons and predict a new scalar particle, the Higgs boson. In 2012, both CMS and ATLAS discovered the Higgs boson, which confirmed the symmetry breaking in nature.

2.1.1 The Top quark

Top quark belongs to the third generation of quarks, the heaviest known elementary particle in the SM. Its unique properties have long been considered as potentially carrying key information that may lead to answers of some of the basic open questions in particle physics. The CDF and D \emptyset collaborations at the Tevatron in 1995 bring the long quest of the sixth and last quark of the SM to an end with the discovery of top quark [9]. The mass of the top quark is measured to be 173.5 GeV, and it has very short life span $\approx 5 \times 10^{-25}$ s, less than the time of hadronization, which enables the top quark to decay before forming bound state. The LHC started collecting data in 2010 at centre-of-mass energy $\sqrt{s} = 7$ TeV. After three years of data collection by CMS and ATLAS, about a million of top quark events were recorded, labelling the LHC as “top factory”. All the current results of the top quark production, decay, and coupling are consistent with the SM prediction. Further, no hints of BSM physics have been observed in its properties. However, different studies are underway to search for new physics; one of them is the $t\bar{t}$ resonance and interference production considered in this thesis. The interference pattern comes into play when the SM $t\bar{t}$ background interferes with the signal resonance (decay of the scalar or pseudo-scalar into $t\bar{t}$). The focus of this thesis is the search for heavy scalar and pseudo-scalar decaying into $t\bar{t}$, where $t\bar{t}$ further decays leptonically. Both resonance and interference are studied in detail in Chapter 6.

Top quark production

In hadron colliders, two mechanisms are responsible for the production of top quark – $t\bar{t}$ pairs via strong interaction and single top production via electroweak process, where $t\bar{t}$ is the most abundant production in the high-energy regime. The production of $t\bar{t}$ is contributed by the two LO processes at tree level shown in Fig. 2.4: quark–antiquark ($q\bar{q}$) annihilation and gluon–gluon (gg) fusion in the s-, t-, and u-channel. In $p\bar{p}$ collider such as Tevatron, the dominant production mode for $t\bar{t}$ is the $q\bar{q}$ annihilation while for pp collider like LHC, gg fusion more likely takes place. The hadronic production cross section of the $t\bar{t}$ pair is successfully explained

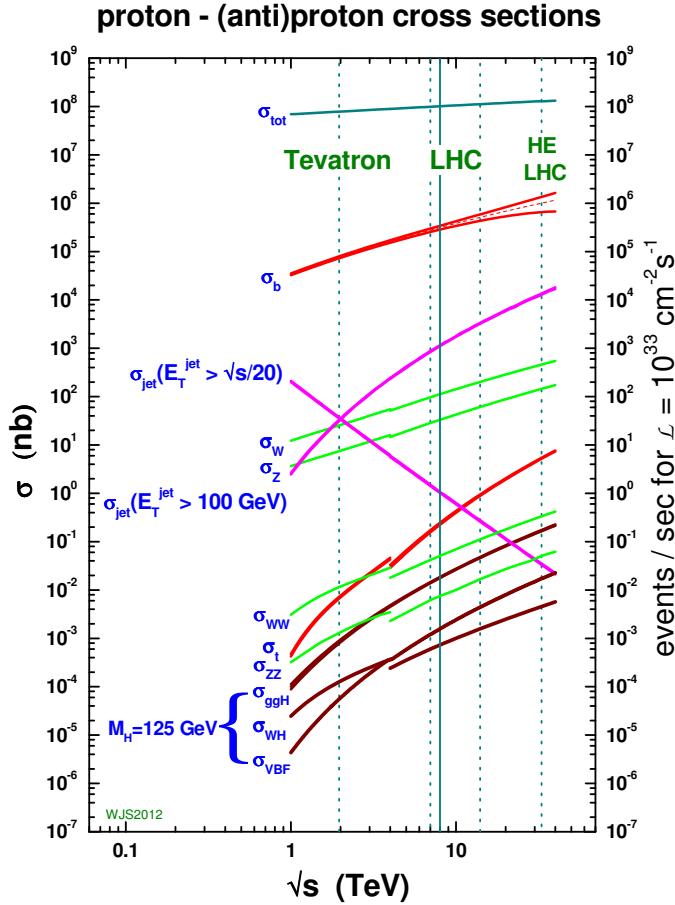


Figure 2.3: The Standard Model cross sections as a function of energy [10].

by the QCD and obtained from the proton-proton interaction using the factorization theorem [11] read as:

$$d\sigma_{pp \rightarrow t\bar{t}}(s, m_t) = \sum_{i,j=q,\bar{q},g} \int dx_i dx_j f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \cdot \hat{\sigma}_{ij \rightarrow t\bar{t}}(\hat{s}, m_t, \mu_f, \mu_r, \alpha_s)) \quad (2.14)$$

The cross sections are integrated over all partons participating in the scattering process with momentum fractions x_i and x_j with respect to the proton momenta and summed over all parton types i and j . The PDFs $f_i(x_i, \mu_f^2)$ are partonic distribution functions that describe the probability of finding a parton i inside a hadron in the momentum range x to $x+dx$. The PDFs absorbs all types of long-distance effects of partons inside the hadrons, whereas the hard scattering partonic cross section $\hat{\sigma}$ represents the actual hard process at small distances that can be computed in perturbative QCD. The long- and short-distance processes can be separated by introducing a new energy scale called the “factorization scale (μ_f)” upon which both PDFs and $\hat{\sigma}$ depend. The hard-scattering partonic cross section $\hat{\sigma}$ further depends on the partonic centre-of-mass energy $\hat{s} = x_i x_j s$ (s is the pp centre-of-mass energy squared), μ_r is the renormalization scale for ultraviolet divergence treatment, and the strong coupling constant α_s . Both energy scales, μ_f and μ_r , are set to the physical energy scale of the process, like for $t\bar{t}$ pair production cross section, they are set to the top quark mass; $\mu_f = \mu_r = m_t$. The recent measurement of $\sigma(t\bar{t})$ at CMS resulted $834.6 \pm 2.5(\text{stat}) \pm 19.1(\text{syst}) \pm 22.5(\text{lumi}) \pm 12.5(\text{extrapol.})$ pb [12], in agreement with the SM prediction.

Single-top can be produced using hadron colliders through several processes. Among them, three electroweak processes are prominent – t -channel (the dominant channel with 70% of the total cross section), the associated production of a top (anti-top) quark, a W boson (tW -channel with 25% cross section), and the s -channel (5% of the total cross section). The Feynman diagrams for the three processes of single-top quark production are shown in Fig. 2.5. The electroweak production vertex includes the CKM matrix element V_{tb} that makes the single-top processes more interesting for SM and BSM physics [13].

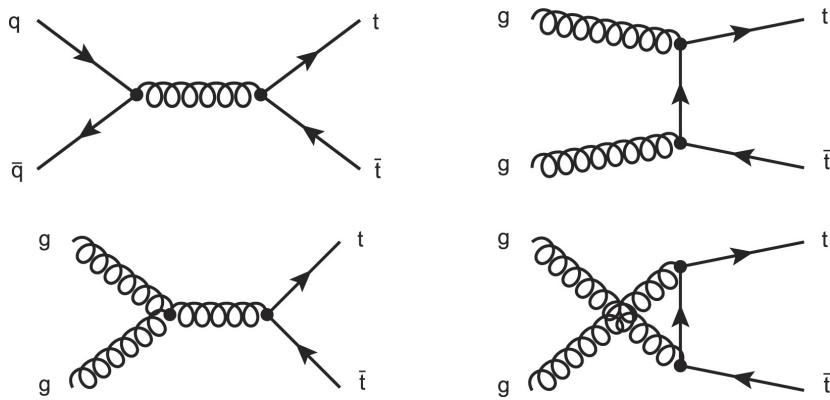


Figure 2.4: $t\bar{t}$ pair production in QCD at LO: top left is $q\bar{q}$ annihilation, bottom left is gg fusion in the s -channel, top right is gg fusion in the t -channel, and bottom right is gg fusion in the u -channel.

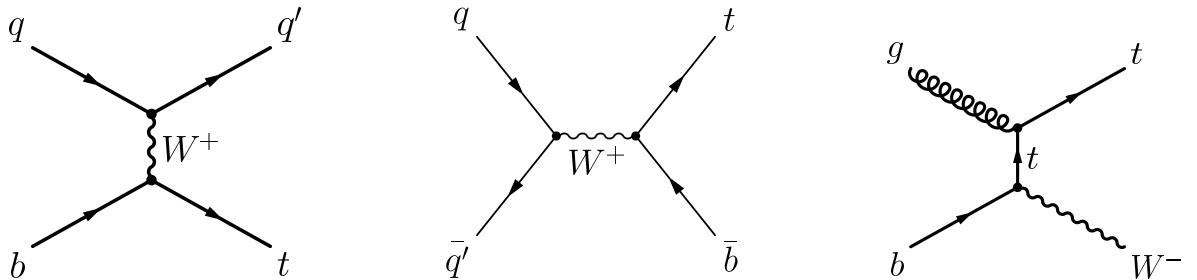


Figure 2.5: Feynman diagram for electroweak single top quark production in the standard model. From left to right: t -channel, s -channel and W -associated production.

Top quark decay

The top quark almost exclusively decays into a W boson and a b -quark via electroweak interaction due to the CKM matrix element $V_{tb} \approx 1$. As compared to other quarks of the SM, top quark has a very short life span ($\tau_t \approx 5 \times 10^{-25}$ s) and decays before hadronization ($\tau_{had} \approx 10^{-24}$ s). This makes it more interesting to study the top quark in its bare state along with its kinematic and dynamic information, such as spin correlation transfers to its decay products without being diluted by hadronization. The decay of $t\bar{t}$ system can be classified on the basis of W boson decay products, where the W boson can decay into a pair of light quarks, with $\text{BR}(W \rightarrow q\bar{q}') \simeq 67\%$, or charged leptons with corresponding neutrinos of the same generation with $\text{BR}(W \rightarrow l\nu_l) \simeq 33\%$. At first order, the possible decay modes of a $t\bar{t}$ event can be grouped into three main categories, as given in Table 2.2

Channel	Decay mode	BR
Hadronic	$t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}q\bar{q}'q''\bar{q}'''$	46%
Semileptonic	$t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}q\bar{q}'l\nu_l$	45%
Dileptonic	$t\bar{t} \rightarrow b\bar{b}W^+W^- \rightarrow b\bar{b}l'l'\nu_l\nu_{l'}$	9%

Table 2.2: Three categories of $t\bar{t}$ system decay with branching ratios[14].

- **Hadronic Channel:** Both W bosons coming from $t\bar{t}$ event decay into light quark–antiquark pairs (quark and antiquark further hadronized in the form of jets) with high-branching fraction, resulting in a large sample of selected events. In this channel, at least six jets are required – two of them stem from bjets and remaining from light jets. In hadron colliders like LHC, this channel suffers from an overwhelming QCD multijet background, whose cross section many magnitudes higher than the one for $t\bar{t}$ production. The QCD multijet background is difficult to model from simulation and requires data-driven techniques to estimate using collision data.
- **Dileptonic Channel:** In dileptonic channel, both W bosons decay leptonically, resulting two bjets, two charged leptons, and two corresponding neutrinos in $t\bar{t}$ event. It has the lowest branching ratio but a very clean signature at the experimental level due to the presence of two oppositely charged leptons, which

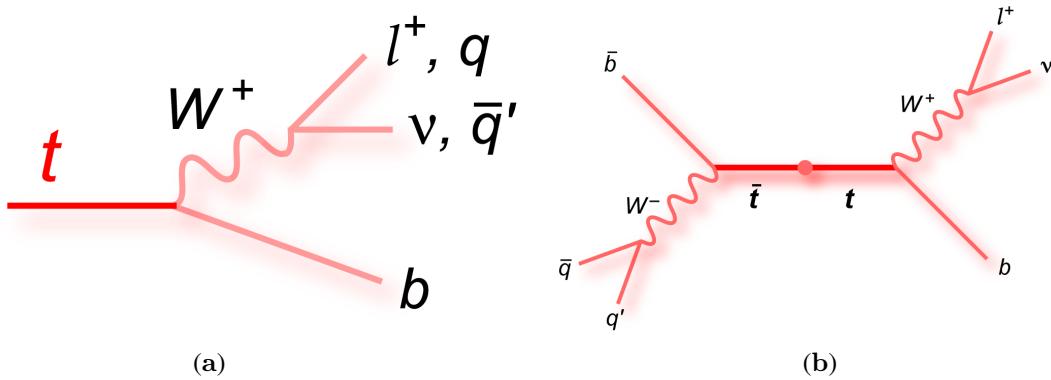


Figure 2.6: Feynman diagram for top quark decay (a) and the semileptonic decay of $t\bar{t}$ (b).

can be easily distinguished from QCD multijet events. The disadvantage of a dileptonic event is the presence of two neutrinos, considered as the missing transverse energy in the event, where $t\bar{t}$ can't be reconstructed perfectly.

- **Semileptonic Channel:** Finally, the lepton+jets channel, whose decay chain is shown in Fig. 2.6b, is a good compromise between a reasonable BR and a relatively clean experimental signature with reconstructible $t\bar{t}$ -event topology (especially when lepton = μ , e). The semileptonic channel is restricted to $\mu + \text{jets}$ and $e + \text{jets}$ as the τ lepton further decay into charged lepton and neutrino or jets that make the channel more complicated. Therefore, a dedicated study is needed to take into account the $\tau + \text{jets}$ channel.

With a clean signature and adequate BR ($\sim 30\%$), the $l(\mu/e) + \text{jets}$ channel was chosen for the analysis included in this thesis. In the resolved case, where the final products of each top quark ($p_T \leq 400 \text{ GeV}$) are reconstructed as separate objects, the $t\bar{t}$ final objects consists of four jets, one charged lepton (muon or electron) and neutrino. Two of the jets, tagged as bjets, arise directly from the decay of top quark, while two of the light jets come from one of the W bosons. The second W boson decays into a charged lepton and corresponding neutrino. As the neutrino escapes without detection, it is treated as missing transverse energy (E_T^{miss}) in the event. The longitudinal component of E_T^{miss} is unmeasured and can be reconstructed using the analytical solution explained in detail in [15]. The $t\bar{t}$ system is reconstructed using the Rochester Algorithm explained in Sec. 6.5.1.

2.1.2 The SM Higgs boson

In the early 60s, the existence of the Higgs boson was theoretically predicted by Robert Brout, François Englert, and Peter Higgs using the well-known BEH mechanism, as explained in Sec. 2.1. The main goal in the field of particle physics during the last few decades has been the discovery of the Higgs boson, where a number of searches have been conducted at the Large Electron-Positron Collider (LEP) at CERN and at Fermilab with the experiments CDF and D \emptyset at $\sqrt{s} = 1.96 \text{ TeV}$, using an integrated luminosity of 10.0 fb^{-1} . The mass range was later narrowed down by excluding two regions: $100 < m_H < 130 \text{ GeV}/c^2$ and $147 < m_H < 180 \text{ GeV}/c^2$, at the 95% C.L. Furthermore, they observed a significant excess of data events with respect to the background estimation in the mass range, $115 < m_H < 140 \text{ GeV}/c^2$ [16]. The official announcement of the Higgs boson discovery, with 5σ significance, was made by CERN experiments, CMS, and ATLAS in July 2012 by combining data of 7 and 8 TeV energy with integrated luminosity of 10.4 fb^{-1} (CMS) and 10.6 fb^{-1} (ATLAS) [1, 2]. The search was performed in five Higgs decay modes – $\gamma\gamma$, ZZ , W^+W^- , $\tau^+\tau^-$, and $b\bar{b}$, where the excess is most significant in the two decay modes with the best mass resolution, $\gamma\gamma$ and ZZ . By performing a fit to the signals, the mass of the Higgs boson is found to be $125.3 \pm 0.4(\text{stat.}) \pm 0.5(\text{syst.}) \text{ GeV}$. At 13 TeV, the study has been repeated and the mass of the Higgs boson was measured to be $124.50^{+0.48}_{-0.46} \text{ GeV}$ with a constraint on the width $\Gamma_H < 41 \text{ MeV}$ [17]. The measured cross section agrees well with the expectation of the SM. The signal strength μ , defined as the production cross section of the Higgs boson times its branching fraction to four leptons relative to the SM expectation, is measured as $\mu = 0.99^{+0.33}_{-0.26}$ at $m_H = 125.09 \text{ GeV}$. The properties of the Higgs boson, such as spin, parity, ratios of production rates for different production modes, ratios of couplings to fermions and vector bosons, etc. agree well with the SM prediction. The Higgs mass measured at

13 TeV using 12.9 fb^{-1} data in two channels $H \rightarrow \gamma\gamma$ (a) and $H \rightarrow ZZ \rightarrow 4l$ (b) is shown in Fig. 2.7, where the excess around 125 GeV in the data is visible.

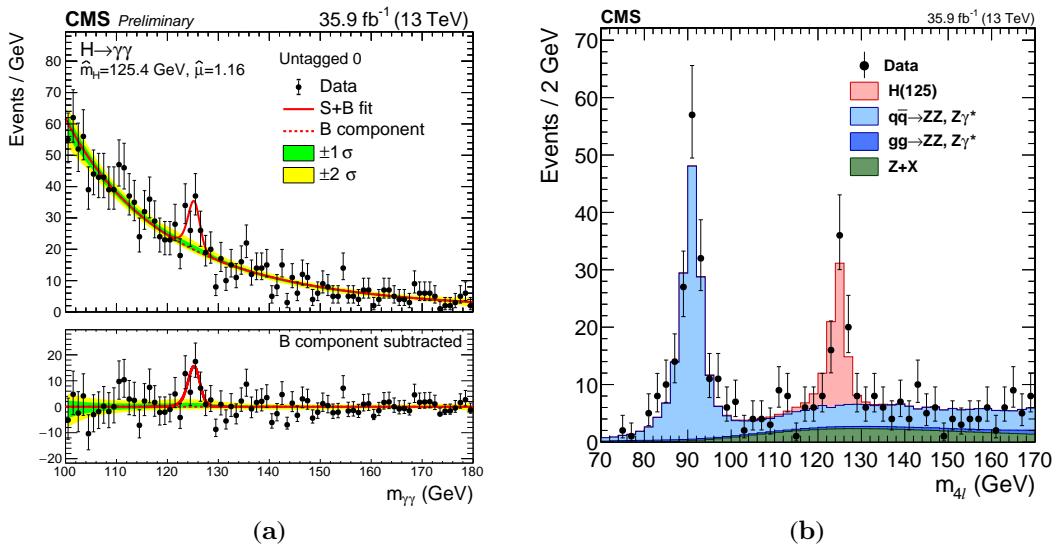


Figure 2.7: The SM Higgs mass spectra obtained in the 2016 CMS search using the two most sensitive decay channels, di-photon (left) and four-lepton (right). Both channels show more than 5 standard deviations significance of the observed signals around 125 GeV; the data corresponds to an integrated luminosity of 13 fb^{-1} , collected with the CMS detector at a centre-of-mass energy of 13 TeV [18, 19].

- **Production and decay of the SM Higgs boson:** The SM Higgs boson couples with every massive particle; hence, the production in hadron collider is possible in many ways. The following four are the most common.
 - Gluon-gluon fusion (ggF):** In this mode of production, two gluons interact to produce a Higgs boson mediated by heavy quarks (mainly top quark) loop as Higgs boson doesn't directly couple to massless gluon. This channel of production has a higher cross section throughout the Higgs boson mass range.
 - Vector boson fusion (VBF):** In the VBF mode, two quarks from colliding protons radiate massive vector bosons (W or Z) and its fusion results in the production of the Higgs boson. It is the second highest cross section process with 10 times less than the ggH but has a clear signature because of the pure electroweak process.
 - Associated production with W^\pm or Z^0 (VH):** This process is also known as Higgs-Strahlung, the third highest cross section, in which a massive vector boson is produced by the interaction of a valence quark with an antiquark from the sea and the massive vector boson further radiates a Higgs boson. This mode is helpful for testing the Higgs coupling with the massive vector boson.
 - Associated production with $t\bar{t}$ ($t\bar{t}H$):** In this mode of production, the Higgs is associated with heavy quark pairs, like the top quark pair, which can be initiated either by gluon-gluon fusion or $q\bar{q}$ annihilation. This channel has smaller cross sections but is more interesting because of the direct coupling with the top quark.

Higgs boson decays through many channels, either through direct coupling to the final state massive particles or indirectly via a boson or fermion loop in massless final state particles (photon, gluon). The decay channel $H \rightarrow b\bar{b}$ has the largest branching ratio (BR) but is diluted by the irreducible QCD background which mimics the signal. The same is true for the $H \rightarrow \tau\bar{\tau}$ channel, where τ decays hadronically. Despite the low BRs, the two channels $H \rightarrow ZZ^* \rightarrow 4l$ and $H \rightarrow \gamma\gamma$ show a very clear signature to the Higgs. Ultimately, the Higgs discovery became possible owing to these two channels. Figure 2.8 (a) shows the Higgs cross section as a function of centre-of-mass energy for $m_H = 125 \text{ GeV}$ with a band of theoretical uncertainty and (b) shows the BRs as a function of Higgs mass with a band of total uncertainty(%).

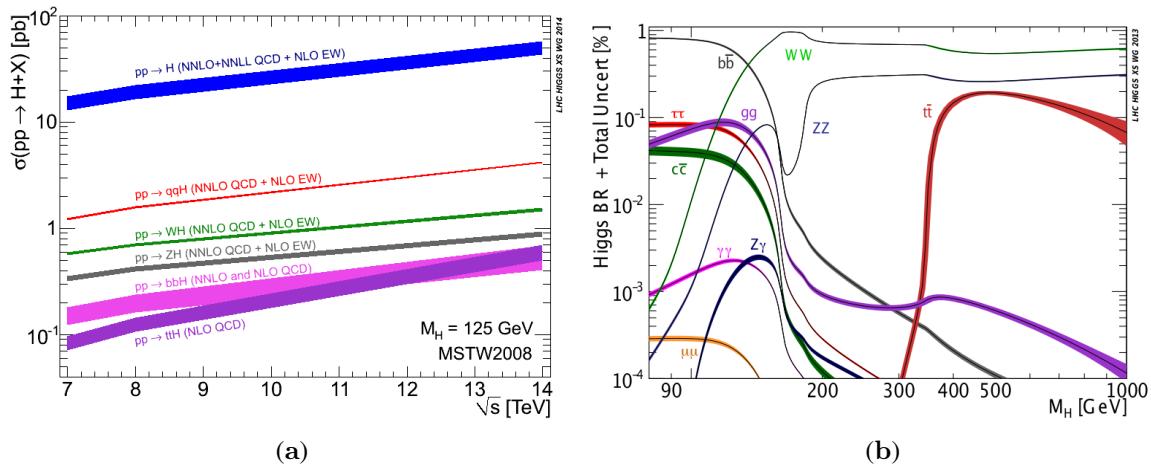


Figure 2.8: Higgs boson (a) production cross section for main modes as a function of centre-of-mass energy exploiting $m_H = 125\text{ GeV}$ for pp collider and (b) branching ratio (BR) with full uncertainty in almost all channels as a function of Higgs mass [20].

2.2 Beyond the Standard Model

Although the SM has been tested with great success with regard to predicting and explaining many physics processes, especially in its gauge and flavour sector, it is still not the ultimate theory. This is because there exist a few experimental pieces of evidence and theoretical issues that are not explained by the SM. The most prominent among them are gravity, matter–anti-matter asymmetry, existence of dark matter and dark energy, and non-vanishing neutrino mass. There are some fundamental characteristics of the SM that we can't explain. The most important among them are: Why is the mass of the SM Higgs boson 125 GeV when the Planck scale is $\mathcal{O}(10^{19})$ GeV? This phenomenon is commonly known as the Hierarchy Problem and is a basic motivation for this search, as explained in more detail in Sec. 2.2.1. Similarly, another question that arises is: Why do we have three families of fermion that have a wide mass range?

2.2.1 The Hierarchy problem

Physics beyond the Standard Model (SM) can be probed for many reasons, but the core issue before us today, which is relevant to Higgs boson physics and electroweak explorations at the Large Hadron Collider, is the Hierarchy Problem. The problem involves the large difference between the Planck mass $M_{pl} \approx 10^{19}$ GeV, an energy scale associated with gravity, and the electroweak symmetry breaking scale (10^2) of the SM. Higgs boson is the only fundamental scalar in the SM that can be subjected to higher-order loop corrections from spin 0, 1/2, and 1 particle with a quadratic dependence on the cut-off scale (Planck scale). These corrections may take the Higgs mass to the highest scale, but the Higgs mass ~ 125 GeV is in contrast. The masses of fermions and gauge bosons are protected by chiral and local gauge symmetry respectively, but for scalar Higgs, there is no symmetry banning its mass term $m^2 H^\dagger H$. In Fig. 2.9, the upper row shows the quantum loop corrections to the Higgs mass; left is the Higgs self-interaction, middle is the gauge boson loop, and right shows the fermionic loop. In mathematical language, consider massive N_f fermions with Youkawa coupling $\lambda = \sqrt{2}m_f/\nu$. Neglecting the external Higgs momentum squared, we obtain:

$$\Delta M_H^2 = N_f \frac{\lambda_f}{8\pi^2} \left[\Lambda^2 + 6m_f^2 \log \frac{\Lambda}{m_f} - 2m_f^2 \right] + \mathcal{O}\left(\frac{1}{\Lambda^2}\right) \quad (2.15)$$

Where $\Delta M_H^2 \propto \Lambda^2$. If we consider the cut-off scale to be the Planck scale $M_{pl} \approx 10^{19}$ GeV, the Higgs boson mass becomes very huge, which stands against the experimental results. A counter-term $\mathcal{O}(10^{-30})$ needs to be added to the mass squared so that the value matches with the experimental result, but this seems unnatural. This is commonly known as fine-tuning or naturalness [21]. Several theories are proposed to cancel out infinities and leave the mass around the Higgs mass (125 GeV). Supersymmetry (SUSY) is one of the leading possibilities for solving the naturalness and Hierarchy Problem by introducing “superpartner” to the SM Higgs particles. Such superpartners introduce additional Feynman diagrams to cancel out the infinities, as shown in the lower row in Fig. 2.9 for Higgs, W boson, and top quark.

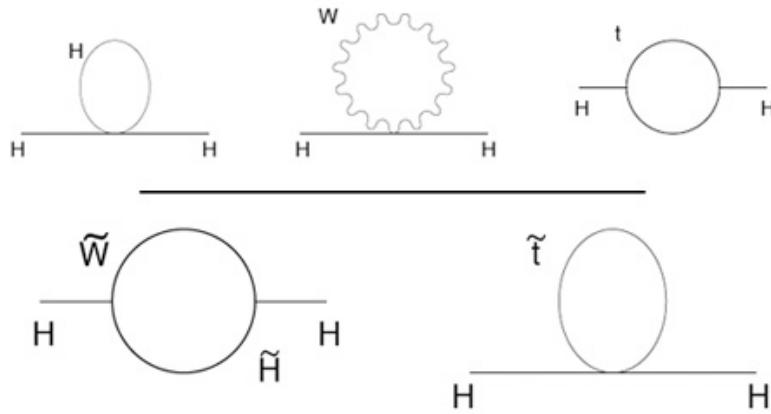


Figure 2.9: The three Feynman diagrams in the top row, from left to right, represent Higgs self-, W/Z boson, and fermion (top quark) coupling respectively – a quantum effects on the Higgs mass. These effects boost the Higgs mass to an infinite value that needs to be finite by using some unknown contributions. SUSY suggests supersymmetric partners of the SM particles, one must include additional diagrams, where superpartners cancel out the effects of the loops in the Higgs mass calculation, as shown in the lower row. Source of the Feynman diagrams is [22].

2.2.2 Supersymmetry

Supersymmetry (SUSY) is a space-time symmetry relating to two different classes of particles – integer spin-0 and spin-1 (bosons) and spin-1/2 (fermions) [23]. The SUSY generators \mathcal{Q} transform a fermionic state into a bosonic state and vice-versa:

$$\mathcal{Q}|Fermion\rangle = |Boson\rangle, \mathcal{Q}|Boson\rangle = |Fermion\rangle \quad (2.16)$$

When the symmetry is exact, bosons and fermions have the same mass and quantum number, except for the spin. But the question arises, why didn't we find superpartners of the SM particles? For example, the electron was discovered more than a century ago. If it has a superpartner with the same mass and charge with opposite spin, it should be discoverable through simple experiments. Similarly, no squark (where “s” stands for superpartner) has been observed in hadron-hadron colliders so far. This suggests that SUSY does not have exact symmetry and must be broken in a manner such that the superpartners acquire higher masses (higher than the top quark) but not too high as to reintroduce the Hierarchy Problem. This is commonly known as the “soft” breaking of the SUSY. Up till now, nobody has found a completely satisfactory dynamic way to break SUSY; however, there are many optional cases, where the SUSY can break. One possibility is to introduce the soft term to the SUSY Lagrangian by hand, which breaks the SUSY explicitly.

$$\mathcal{L} = \mathcal{L}_{SUSY} + \mathcal{L}_{soft} \quad (2.17)$$

This results in a low-energy effective SUSY theory and can be used for mass and decay predictions. Minimal Supersymmetric Standard Model (MSSM) is the most economical version that will be discussed in detail in the next section 2.2.3.

2.2.3 Minimal Supersymmetric Standard Model

The Minimal Supersymmetric Standard Model (MSSM) [24] is based on the group $SU(3)_C \times SU(2)_L \times U(1)_Y$ with a minimal amount of additional fields required with regard to the existing one in order to make the theory supersymmetric. In MSSM, each SM particle has its superpartner differ by spin-1/2. An overview of the gauge bosons (spin-1) and their spin-1/2 partners, the gauginos, is given in Table 2.3. In SM, there are three generations of spin-1/2 fermions (quarks and leptons without right-handed neutrino) that correspond to spin-0 SUSY superpartners, squarks, and sleptons: $\hat{Q}, \hat{U}_R, \hat{D}_R, \hat{L}, \hat{E}_R$. SUSY requires two Higgs doublets for anomaly cancellation and to give mass to all fermions. Conventionally, the two Higgs doublets are denoted by H_u with $Y = +1/2$ and H_d with $Y = -1/2$, where the first one gives mass to the up-type quarks and the second to the down-type and charged leptons. Table 2.4 summarizes the particles and their supermultiplets.

Unlike the SM, SUSY doesn't respect lepton and baryon number conservation. This can be enforced in a

Names	Superfield \tilde{S}	Spin- $\frac{1}{2}$	Spin-1	$SU(3)_C, SU(2)_L, U(1)_Y$
gluino, gluon	\hat{G}^a	\tilde{G}	g	(8, 1, 0)
winos, W bosons	\hat{W}	$\tilde{W}^\pm, \tilde{W}^0$	W^\pm, W^0	(1, 3, 0)
bino, B boson	\hat{B}	\tilde{B}^0	B^0	(1, 1, 0)

Table 2.3: Gauge supermultiplets in the MSSM with quantum numbers.

simple manner – by introducing a discrete and multiplicative symmetry called R-parity:

$$R_p = (-1)^{2s+3B+L} \quad (2.18)$$

Where s, B, and L represent spin, baryon, and lepton number respectively. For the ordinary particles (fermions, gauge bosons, and Higgs bosons), $R_p = +1$ and for their superpartner $R_p = -1$. SUSY allows R_p conservative terms in its Lagrangian. In practice, R-parity conservation has important effects on the MSSM phenomenology by allowing pair production of the SUSY particles in colliders. Further decay of the SUSY particles comprises an odd number of SUSY particles, where the stable lightest supersymmetric particle (LSP) is a strong candidate for dark matter if it is neutral and only weakly interacting. In light of the above-mentioned properties, it is easy

Names	Superfield \tilde{S}	Spin-0	Spin- $\frac{1}{2}$	$SU(3)_C, SU(2)_L, U(1)_Y$
squarks, quarks ($\times 3$ families)	\hat{Q}	$(\tilde{u}_L \quad \tilde{d}_L)$	$(u_L \quad d_L)$	$(3, 2, \frac{1}{6})$
	\hat{U}	\tilde{u}_R^*	u_R^\dagger	$(\bar{3}, 1, -\frac{2}{3})$
	\hat{D}	\tilde{d}_R^*	d_R^\dagger	$(\bar{3}, 1, \frac{1}{3})$
sleptons, leptons ($\times 3$ families)	\hat{L}	$(\tilde{\nu} \quad \tilde{e}_L)$	$(\nu \quad e_L)$	$(1, 2, -\frac{1}{2})$
	\hat{E}	\tilde{e}_R^*	e_R^\dagger	$(1, 1, 1)$
Higgs, Higgsinos	\hat{H}_u	$(H_u^+ \quad H_u^0)$	$(\tilde{H}_u^+ \quad \tilde{H}_u^0)$	$(1, 2, +\frac{1}{2})$
	\hat{H}_d	$(H_d^0 \quad H_d^-)$	$(\tilde{H}_d^0 \quad \tilde{H}_d^-)$	$(1, 2, -\frac{1}{2})$

Table 2.4: Chiral supermultiplets in the MSSM with quantum numbers.

to define a globally supersymmetric Lagrangian for MSSM. I will only discuss the final results here, without going into detail about the derivation because it is beyond the scope of this thesis. The kinetic part of the Lagrangian is the supersymmetric equivalent of the SM and is obtained by generalizing the notion of covariant derivative to the SUSY. The gauge invariant's general superpotential, also compatible with renormalizability and R-parity conservation rule, is written as:

$$W = \sum_{i,j=gen} -Y_{ij}^u \hat{u}_{Ri} \hat{H}_2 \cdot \hat{Q}_j + Y_{ij}^d \hat{d}_{Ri} \hat{H}_1 \cdot \hat{Q}_j + Y_{ij}^l \hat{l}_{Ri} \hat{H}_1 \cdot \hat{L}_j + \mu \hat{H}_2 \cdot \hat{H}_1 \quad (2.19)$$

Where i and j run over three generations and $\epsilon_{12} = -\epsilon_{21} = 1$ and $\epsilon_{11} = -\epsilon_{22} = 0$. $H.Q \equiv \epsilon_{12} H^1 Q^2$ is the SU(2) doublets products and $Y_{ij}^{u,d,l}$ represents the Youkawa couplings among generations. The last term represents a globally supersymmetric mass term for the Higgs doublets.

2.2.4 Higgs sector of MSSM

SUSY requires two Higgs doublets for anomaly cancellation and giving mass to all the fermions. This minimal version of SUSY is commonly referred to as the Two Higgs Doublets Model (2HDM). The SM fermions cancel the anomaly by themselves, where the superpartner of the Higgs boson is a fermion that contributes to triangle gauge anomalies. The anomaly can be cancelled out by introducing a second fermion that is basically the vector compliment of the first one. Therefore, MSSM introduces H_u and its vector compliment H_d [25]. Similarly, \hat{H}_u gives mass to up-type quarks and is unable to give mass to down-type quarks, as in superpotential \hat{H}_u^* is forbidden and \hat{H}_d gives mass to the down-type quarks. The Higgs doublets can be represented in terms of

complex field:

$$H_u = \begin{pmatrix} H_u^+ \\ H_u^0 \end{pmatrix} \quad \text{and} \quad H_d = \begin{pmatrix} H_d^0 \\ H_d^- \end{pmatrix} \quad (2.20)$$

We have not observed the supersymmetric partners of SM particles, which requires the SUSY to be broken at some higher scale. Soft-SUSY breaking mechanism does this job by introducing new terms to the Lagrangian in a manner such that it does not reintroduce quadratic sensitivities to the cut-off. The possibly soft-SUSY terms in the Higgs sector are:

$$\mathcal{V}_1 = m_{H_u}^2 \left(|H_u^+|^2 + |H_u^0|^2 \right) + m_{H_d}^2 \left(|H_d^-|^2 + |H_d^0|^2 \right) + \{ b (H_u^+ H_d^- - H_u^0 H_d^0) + h.c \} \quad (2.21)$$

Where the parameters $m_{H_u}^2$, $m_{H_d}^2$ and b are still arbitrary. The total scalar potential is written as [26]:

$$\begin{aligned} \mathcal{V} = & \left(|\mu|^2 + m_{H_u}^2 \right) \left(|H_u^+|^2 + |H_u^0|^2 \right) + \left(|\mu|^2 + m_{H_d}^2 \right) \left(|H_d^-|^2 + |H_d^0|^2 \right) \\ & + \{ b (H_u^+ H_d^- - H_u^0 H_d^0) + h.c \} + \frac{g^2}{2} |H_u^{+\ast} H_d^0 + H_u^{0\ast} H_d^-|^2 \\ & + \frac{g^2 + g'^2}{8} \left(|H_u^+|^2 + |H_u^0|^2 - |H_d^-|^2 - |H_d^0|^2 \right) \end{aligned} \quad (2.22)$$

The Higgs mechanism in MSSM can be used to obtain the minimum of this scalar potential in the same way as in SM, where the electromagnetism gauge group U(1) is not broken. We have freedom to suppose a vanishing vev for the electromagnetic part $\langle H_u^+ \rangle = 0$, at the minimum of the potential, where $\frac{\partial \mathcal{V}}{\partial H_u^+} \Big|_{H_u^+=0} \Rightarrow H_d^- = 0$.

In the same manner, we consider $\langle H_u^- \rangle = 0, \Rightarrow H_d^+ = 0$. Thus, electromagnetism is unbroken, as the charged directions cannot attain a vacuum expectation value. Consider the vev of the neutral part $\langle H_u^0 \rangle = v_u$ and $\langle H_d^0 \rangle = v_d$, hence at minimum of \mathcal{V} we obtain:

$$\begin{aligned} \left(|\mu|^2 + m_{H_u}^2 \right) v_u &= bv_d + \frac{g^2 + g'^2}{4} v_u (v_d^2 - v_u^2) \\ \left(|\mu|^2 + m_{H_d}^2 \right) v_d &= bv_u - \frac{g^2 + g'^2}{4} v_d (v_d^2 - v_u^2) \end{aligned} \quad (2.23)$$

Where the vevs are related to W/Z masses. Consider the kinetic terms for the Higgs field:

$$\mathcal{L}_{kin} = (D_\mu H_u)^\dagger + (D^\mu H_u) + (D_\mu H_d)^\dagger + (D^\mu H_d) \quad (2.24)$$

with $D_\mu = \partial_\mu + ig\tau_a W_\mu^a/2 + ig'yB_\mu/2$, the W/Z masses can be written as:

$$\begin{aligned} m_W^2 &= \frac{g^2}{2} (v_u^2 + v_d^2) \\ m_Z^2 &= \frac{g^2 + g'^2}{2} (v_u^2 + v_d^2) \end{aligned} \quad (2.25)$$

The two Higgs-doublets extension of the MSSM generates eight scalar degrees of freedom (DoF) before electroweak symmetry breaking (EWSB). After symmetry breaking, three of them are eaten by gauge bosons, given by 2.25, and the remaining five DoFs should manifest as physical states. Among five, two of the DoFs are charged (H^\pm), one is neutral pseudo scalar (A^0), and two are neutral \mathcal{CP} -even scalar h^0 and H^0 states, where conventionally $h^0 < H^0$. Their masses can be computed by expanding the doublet fields around their vev ($H = \langle H \rangle + \phi + i\varphi$) value and plug this expansion into the Higgs potential. It can be seen that the mass terms of the fields φ_u and φ_d only mix among themselves. Taking the ratio of vevs of the doublets as $\tan \beta = \frac{v_u}{v_d}$, we obtain the \mathcal{CP} -odd state A^0 :

$$A^0 = \sqrt{2} (\varphi_u \cos \beta + \varphi_d \sin \beta); \quad m_{A^0}^2 = \frac{b}{\sin 2\beta} \quad (2.26)$$

For ϕ_u and ϕ_d we obtain in the same manner the mass terms:

$$\begin{aligned} m_{h^0}^2 &= \frac{1}{2} \left[(m_{A^0}^2 + m_Z^2) - \sqrt{(m_{A^0}^2 + m_Z^2)^2 - 4m_{A^0}^2 m_Z^2 \cos^2 2\beta} \right] \\ m_{H^0}^2 &= \frac{1}{2} \left[(m_{A^0}^2 + m_Z^2) + \sqrt{(m_{A^0}^2 + m_Z^2)^2 - 4m_{A^0}^2 m_Z^2 \cos^2 2\beta} \right] \end{aligned} \quad (2.27)$$

Consider the possibility that $m_{A^0} \gg m_Z$, then the mass of the lightest Higgs h^0 is bounded above by $m_Z |\cos 2\beta|$. One can include higher-order corrections and large log resummations that push the Higgs mass to ≈ 130 GeV and expected to behave like SM Higgs boson for most of the MSSM parameter space. The discovery of the SM Higgs boson with mass ≈ 125 GeV is consistent with these predictions. The corresponding fields for h^0 and H^0 are:

$$\begin{aligned} h^0 &= \sqrt{2} (\phi_u \cos \alpha - \phi_d \sin \alpha) \\ H^0 &= \sqrt{2} (\phi_u \sin \alpha + \phi_d \cos \alpha) \end{aligned} \quad (2.28)$$

The charged Higgs field is:

$$H^\pm = \phi_d^\pm \sin \beta + \phi_u^\pm \cos \beta \quad (2.29)$$

We can also obtain the equation:

$$\tan \alpha = - \frac{\left[\frac{1}{2} (m_{H^0}^2 - m_{h^0}^2) + \frac{\cos 2\beta}{2} (m_{A^0}^2 - m_Z^2) \right]}{\frac{1}{2} (m_{A^0}^2 + m_Z^2) \sin 2\beta} \quad (2.30)$$

$m_{A^0} \gg m_Z$, $\tan \alpha = -\cot \beta \Rightarrow \alpha \rightarrow \beta - \frac{\pi}{2}$. This limit is commonly known as the decoupling limit, where the lightest scalar Higgs behaves like a SM Higgs boson. We use the same condition while calculating the k-factor from the SusHi generator, as described in 5.5.1. The corresponding tree-level squared Higgs masses are given by [27]:

$$\begin{aligned} m_h^2 &\simeq m_Z^2 \cos^2 2\beta \\ m_H^2 &\simeq m_A^2 + m_Z^2 \sin^2 2\beta \\ m_{H^\pm}^2 &\simeq m_A^2 + m_W^2 \end{aligned} \quad (2.31)$$

Table 2.5 provides an overview of the couplings of the neutral 2HDM/MSSM Higgs bosons to up-type quarks, down-type quarks, and vector bosons. Pseudoscalar decays to vector bosons are forbidden, and the coupling to up-type quarks is inversely proportional to $\tan \beta$, whereas the coupling to down-type quarks is proportional to $\tan \beta$. From the above discussion, it can be concluded that in the 2HDM/hMSSM¹ approach, we only need

Φ	$g_{\Phi \bar{u} u}$	$g_{\Phi \bar{d} d}$	$g_{\Phi VV}$
h	$\cos \alpha / \sin \alpha$	$-\sin \alpha / \cos \beta$	$\sin(\beta - \alpha)$
H	$\sin \alpha / \sin \beta$	$\cos \alpha / \cos \beta$	$\cos(\beta - \alpha)$
A	$\cot \beta$	$\tan \beta$	0

Table 2.5: Couplings in type-2 2HDM models, in particular the MSSM, of the three neutral Higgs bosons, denoted as Φ , to up-type quarks, down-type quarks, and vector bosons, in terms of the MSSM parameters α and β . The couplings are normalised to SM Higgs boson couplings.

two input parameters (M_A , $\tan \beta$) to describe the entire MSSM Higgs sector as well as the Higgs masses and couplings.

¹hMSSM is a post-SM Higgs discovery scenario, where the SM Higgs boson mass $M_h \approx 125$ GeV has been exploited to fully describe the Higgs sector of the MSSM using two parameters [m_A , $\tan \beta$]. This approach reduces the dependency of numerous SUSY parameters for dominant radiative corrections to the MSSM Higgs sector. It is a simple approach that reopens the search possibilities in the low $\tan \beta$ regime (near to unity).

2.2.5 Heavy Higgs production and decay at LHC

In this section, I am going to discuss the production and decay of the heavier A, H, and H^\pm particles at the LHC in the 2HDM/hMSSM. I will focus on the high, moderate, and low $\tan\beta$ regions as both production and decay are sensitive to $\tan\beta$.

Neutral Higgs production: The dominant production mode for heavy Higgs $\Phi = H/A$ at high energies involves a leading contribution from the two main processes – the gluon fusion mechanism $gg \rightarrow \Phi$, initiated by a heavy quark loop, and the associated Higgs production with b quark $gg/q\bar{q} \rightarrow b\bar{b} + \Phi$ or only b-quark fusion $b\bar{b} \rightarrow \Phi$ process [28]. The VBF, associated production with massive gauge bosons, and top-quark pair have low rates as the couplings are highly suppressed or the phase space is not favourable. The production cross sections of both \mathcal{CP} -even (H) and \mathcal{CP} -odd (A) in Eq. 2.32 depend on the magnitude of the Higgs couplings to quarks at the tree level, but the form factors for the two cases are different as shown in Eq. 2.33. The \mathcal{CP} -even lightest state h behaves as SM Higgs in the alignment limit approximation described in 2.31. In the case of heavy Higgs ($\Phi = H/A$), the rates are sensitive to the mass, M_Φ , and $\tan\beta$ values. The leading-order partonic cross sections $\hat{\sigma}$ for the two dominant processes $b\bar{b} \rightarrow \Phi$ and $gg \rightarrow \Phi$ can be expressed in terms of partonic centre-of-mass energy \hat{s} and mass M_Φ as:

$$\begin{aligned}\hat{\sigma}(b\bar{b} \rightarrow \Phi) &= \frac{\pi}{12} g_{\Phi bb}^2 \delta(\hat{s} - M_\Phi^2) \\ \hat{\sigma}(gg \rightarrow \Phi) &= \frac{G_F \alpha_s^2}{288\sqrt{2}\pi} M_\Phi^2 \delta(\hat{s} - M_\Phi^2) \left| \frac{3}{4} \sum_Q g_{\Phi QQ} A_{1/2}^\Phi(\tau_Q) \right|^2\end{aligned}\quad (2.32)$$

Where the reduce variable is $\tau_Q = M_\Phi^2/4m_Q^2$ and the quark in loop Q is a heavy bottom or top quark. For consistency, I use the label from reference [28]. The form factors for the two states (H/A) using fermion loops are given by:

$$\begin{aligned}A_{1/2}^H(\tau) &= 2[\tau + (\tau - 1)f(\tau)]\tau^{-2} \\ A_{1/2}^A(\tau) &= 2\tau^{-1}f(\tau),\end{aligned}\quad (2.33)$$

where function $f(\tau)$ is given by 2.34 for the two ranges of τ .

$$f(\tau) = \begin{cases} \arcsin^2(\sqrt{\tau}), & \text{for } \tau \leq 1. \\ -\frac{1}{4} \left(\log \frac{1+\sqrt{1-1/\tau}}{1-\sqrt{1-1/\tau}} - i\pi \right)^2, & \text{otherwise.} \end{cases}\quad (2.34)$$

In case of top quarks in the loop, the reduce variable $\tau_t = M_\Phi^2/4m_t^2 \approx 4$ at $M_\Phi = 750$ GeV and $m_t = 172$ GeV makes the form factors $|A_{1/2}^A(\tau)/H_{1/2}^A(\tau)|^2 \sim 2$ that increases the \mathcal{CP} -odd cross section two times compare to the \mathcal{CP} -even state [29].

$$\frac{\sigma(gg \rightarrow A)}{\sigma(gg \rightarrow H)} \approx 2\quad (2.35)$$

As stated above, the production rate is sensitive to $\tan\beta$ values. At low $\tan\beta$, the phenomenology is richer for heavy Higgs (H/A) production. In this limit, coupling for the top quark, $g_{\Phi t\bar{t}} \propto 1/\tan\beta$, increases drastically and the dominant contribution to the cross section coming from the gluon fusion process $gg \rightarrow \Phi$ via top quark loops. On the other hand, the contribution from $b\bar{b} \rightarrow \Phi$ is negligible to cross sections as the coupling, $g_{\Phi b\bar{b}}$, is very small. For higher mass, $M_\Phi \geq 2m_t$, the associated production with $t\bar{t}$ and massive gauge boson is suppressed owing to the small phase-space and negligible coupling g_{HVV} . The \mathcal{CP} -invariance forbids AVV couplings; as a result, the \mathcal{CP} -odd state is not produced.

In case of an intermediate value of $\tan\beta \approx 3 - 10$, both the couplings $g_{\Phi b\bar{b}}$ and $g_{\Phi t\bar{t}}$ have its minimum level; hence, this results in a smaller production cross section. For $\tan\beta \approx \sqrt{m_t/\bar{m}_b} \approx 7$, the cross section has its minimum value where top quarks mass $m_t = 173$ GeV and b-quark mass $\bar{m}_b \approx 3$ in the \overline{MS} scheme [30].

For high values of $\tan\beta \geq 10$, the top quark coupling $g_{\Phi t\bar{t}}$ decreases to a large scale while the b-quark coupling $g_{\Phi b\bar{b}}$ is enhanced, which dominates the b-quark contribution in the $gg \rightarrow \Phi$ loop. The production rates for both processes $gg \rightarrow \Phi$ and $b\bar{b} \rightarrow \Phi$ are the same but at $M_\Phi \gtrsim 2m_Q$, $gg \rightarrow \Phi$ is known to NLO while $b\bar{b} \rightarrow \Phi$ is known to NNLO corrections [31]. For the top-quark loop, we don't include the NNLO corrections

as it is only valid for $M_\Phi \lesssim 2m_Q$ [32]. Figure 2.10a shows the production cross sections for \mathcal{CP} -even and odd states at centre-of-mass energy 14 TeV and $\tan\beta = 2.5$ exploiting the mass of the SM Higgs to be 126 GeV. At $M_\Phi \approx 2m_t$ there is a sharp increase in cross sections, especially for the \mathcal{CP} -odd state A.

Charged Higgs production: The coupling of charged Higgs (H^\pm) is proportional to up- and down-type quarks by the following relation:

$$g_{H^\pm \bar{u}d} \propto m_d \tan\beta (1 + \gamma_5) + m_u \cot\beta (1 - \gamma_5) \quad (2.36)$$

The second part of the Eq. 2.36 makes a large contribution at low $\tan\beta$ values because of the heavy top quark. At very high $\tan\beta$, the contribution from the first part ($m_d \tan\beta$) increases and the second part is suppressed. This shows a similar behaviour as described for the neutral heavy Higgs section above. When the mass of the charged Higgs is less than the top mass, $M_{H^\pm} \lesssim 160$ GeV, it is produced with the decay of one top quark ($t \rightarrow bH^+$) in the process $gg + q\bar{q} \rightarrow t\bar{t}$ at the LHC. Figure 2.10b shows M_{H^\pm} versus the cross section for $\tan\beta = 1, 3, 10, 15, 35, 60$ at $\sqrt{s} = 14$ TeV, where the dominant contribution comes from the lowest $\tan\beta$ values. A detailed explanation of the plot can be found here [28].

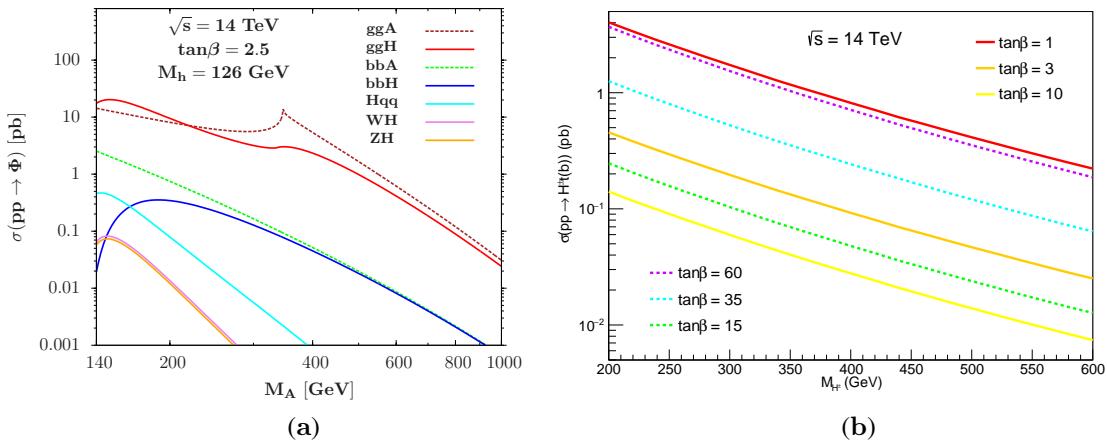


Figure 2.10: (a): Heavy Higgs (neutral) boson production cross sections in MSSM for main channels with $\tan\beta = 2.5$ and SM Higgs mass is 126 GeV at $\sqrt{s} = 14$ TeV w.r.t M_A . (b): Charged Higgs production cross sections for a range of $\tan\beta$ values w.r.t to M_{H^\pm} [28, 33].

Neutral Higgs decay: The decay of the neutral Higgs ($\Phi = H/A$) is dictated by the $\tan\beta$ value in a similar pattern as the production because of the coupling sensitivity to $\tan\beta$. In decoupling limits ($\sin(\alpha - \beta) = 1$), the coupling of the \mathcal{CP} -even state is $\frac{\cos\alpha}{\cos\beta}$; as shown in Table 2.5, it is explicitly equal to $\tan\beta$. For high $\tan\beta \gtrsim 10$, the couplings to b-quark and tau lepton enhanced and it decays exclusively to a $b\bar{b}$ or $\tau^+\tau^-$ pairs with branching ratios $\text{BR}(\Phi \rightarrow b\bar{b} \approx 90\%)$ and $\text{BR}(\Phi \rightarrow \tau^+\tau^- \approx 10\%)$ respectively. Decay to $t\bar{t}$ pair is highly suppressed for high $\tan\beta$ range as $g_{\Phi t\bar{t}} \propto 1/\tan\beta$. Other decay modes of \mathcal{CP} -even state H like in di-boson ($H \rightarrow VV; V = WW, ZZ$) and decay of \mathcal{CP} -odd state into lighter Higgs and Z bosons are strongly suppressed in the alignment limit approximation. For $H \rightarrow hh$ in the limit $M_H \gtrsim 2M_h$, the coupling g_{Hhh} vanished at high $\tan\beta$ [30]:

$$g_{Hhh} = 2 \sin 2\alpha \sin(\alpha + \beta) - \cos 2\alpha \cos(\alpha + \beta) + 3 \frac{\Delta M_{22}^2}{M_Z^2} \frac{\sin\alpha}{\sin\beta} \cos\alpha^2 \quad (2.37)$$

Becomes:

$$\begin{aligned} g_{Hhh} &\xrightarrow{M_A \gg M_Z} -3\Delta \frac{M_{22}^2}{M_Z^2} \times \sin 2\beta \\ \sin 2\beta &\propto \cot\beta \xrightarrow{\tan\beta \gtrsim 10} 0 \end{aligned} \quad (2.38)$$

For $\Phi \rightarrow \gamma\gamma$ channel, the coupling is very small and shows a constant behaviour for $0.3 \lesssim \tan\beta \lesssim 10$, which reduces abruptly for $\tan\beta \gtrsim 10$. The same pattern is followed by the $\Phi \rightarrow gg$ channel, but higher coupling compares to $\Phi \rightarrow \gamma\gamma$. For intermediate values of $\tan\beta \approx 5 - 10$ with $M_\Phi < 2m_t$, the bosonic decay of $H \rightarrow WW, ZZ$ has a significant contribution as its competition is only with the $H \rightarrow b\bar{b}$ decay. In this limit,

the $H \rightarrow t\bar{t}$ decay is suppressed as the phase-space is not feasible. At $\tan \beta = 2.5$ and $2M_h \lesssim M_\Phi < 2m_t$, $H \rightarrow hh$ and $A \rightarrow hZ$ have prominent branching ratios. Figure 2.11 shows branching ratios w.r.t $\tan \beta$ for a specific $M_\Phi = 750$ GeV. Figure 2.12a and b show mass versus branching ratios for \mathcal{CP} -odd and \mathcal{CP} -even states for $\tan \beta = 2.5$ and $M_h = 126$ GeV.

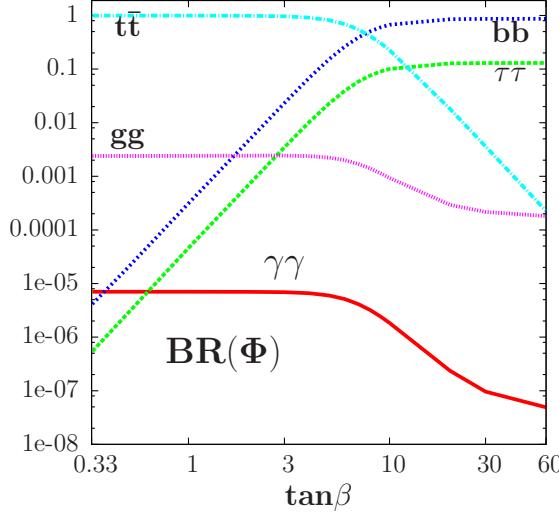


Figure 2.11: The branching ratios of $\Phi = H/A$ in different channels as a function of $\tan \beta$ exploiting the $M_\Phi = 750$ GeV [29].

Charged Higgs decay: In a high range of $\tan \beta \gtrsim 10$ with $M_{H^\pm} \lesssim m_t - m_b$, the dominant decay of H^\pm occurs in the $H^\pm \rightarrow \tau\nu$ channel with a small fraction in hadronic decay mode. For high mass range, $M_{H^\pm} > m_t - m_b$, the decay channel $M_{H^\pm} \rightarrow tb$ becomes the dominant one while $H^\pm \rightarrow \tau\nu$ branching ratio is suppressed. The same pattern for the mentioned mass range is followed for low $\tan \beta$ by these two channels; but this time, the $H^\pm \rightarrow \tau\nu$ channel suppressed is more comparable to high $\tan \beta$. For low $\tan \beta$ and $M_{H^\pm} \gtrsim 160$ GeV, another important channel, $H^\pm \rightarrow hW^\pm$, comes into play with a smaller branching ratio than the $H^\pm \rightarrow tb$ one. Figure 2.12c shows the branching ratios as a function of M_{H^\pm} for H^\pm at $\tan \beta = 2.5$. Until now, I have discussed all the main decay channels for neutral and charged heavy Higgs in different $\tan \beta$ ranges. However, there are many more that have small contributions. In order to present the complete picture, I have written all of them in the form of total decay width equations. The total decay widths of neutral and charged Higgs are given in expression 2.39 and 2.40 respectively. For completeness, I mentioned the SUSY decay widths in the formulae. FH is an abbreviation for FEYNHIGGS and HD stands for HDECAY programs commonly used for decay widths and branching-ratio calculation of the Higgs boson. More detail can be found here [34].

$$\begin{aligned} \Gamma_\Phi = & \Gamma_{\Phi \rightarrow \tau^+ \tau^-}^{\text{FH}} + \Gamma_{\Phi \rightarrow \mu^+ \mu^-}^{\text{FH}} + \Gamma_{\Phi \rightarrow W^* W^*}^{\text{FH/P4f}} + \Gamma_{\Phi \rightarrow Z^* Z^*}^{\text{FH/P4f}} \\ & + \Gamma_{\Phi \rightarrow b\bar{b}}^{\text{HD}} + \Gamma_{\Phi \rightarrow t\bar{t}}^{\text{HD}} + \Gamma_{\Phi \rightarrow c\bar{c}}^{\text{HD}} + \Gamma_{\Phi \rightarrow g\bar{g}}^{\text{HD}} + \Gamma_{\Phi \rightarrow \gamma\gamma}^{\text{HD}} + \Gamma_{\Phi \rightarrow Z\gamma}^{\text{HD}} \\ & + \Gamma_{\Phi \rightarrow Zh}^{\text{FH}} + \Gamma_{\Phi \rightarrow hh}^{\text{FH}} + \Gamma_{\Phi \rightarrow ZA}^{\text{FH}} + \Gamma_{\Phi \rightarrow AA}^{\text{FH}} + \Gamma_{\Phi \rightarrow H^\pm W^\mp}^{\text{HD}} + \Gamma_{\Phi \rightarrow \text{SUSY}}^{\text{FH}} \end{aligned} \quad (2.39)$$

$$\begin{aligned} \Gamma_{H^\pm} = & \Gamma_{H^\pm \rightarrow \tau\nu_\tau}^{\text{FH}} + \Gamma_{H^\pm \rightarrow \mu\nu_\mu}^{\text{FH}} + \Gamma_{H^\pm \rightarrow hW}^{\text{FH}} + \Gamma_{H^\pm \rightarrow HW}^{\text{FH}} + \Gamma_{H^\pm \rightarrow AW}^{\text{FH}} \\ & + \Gamma_{H^\pm \rightarrow tb}^{\text{HD}} + \Gamma_{H^\pm \rightarrow ts}^{\text{HD}} + \Gamma_{H^\pm \rightarrow td}^{\text{HD}} + \Gamma_{H^\pm \rightarrow cb}^{\text{HD}} + \Gamma_{H^\pm \rightarrow cs}^{\text{HD}} + \Gamma_{H^\pm \rightarrow cd}^{\text{HD}} \\ & + \Gamma_{H^\pm \rightarrow ub}^{\text{HD}} + \Gamma_{H^\pm \rightarrow us}^{\text{HD}} + \Gamma_{H^\pm \rightarrow ud}^{\text{HD}} + \Gamma_{H^\pm \rightarrow \text{SUSY}}^{\text{HD}} \end{aligned} \quad (2.40)$$

2.2.6 The $\Phi \rightarrow t\bar{t}$ channel with interference from the SM $gg \rightarrow t\bar{t}$

Introduction to interference: In particle physics, a process is represented by one or more Feynman diagrams and the resultant amplitude is expressed in terms of the matrix element. For a given process, interference is the contribution of the cross terms of different processes to the final amplitude that introduces a peak, dip, nothingness, or a more complex peak-dip or dip-peak structure. It can be considered to be a similar phenomenon as quantum wave interference. Interference depends on the physics model and affects different kinematics of the

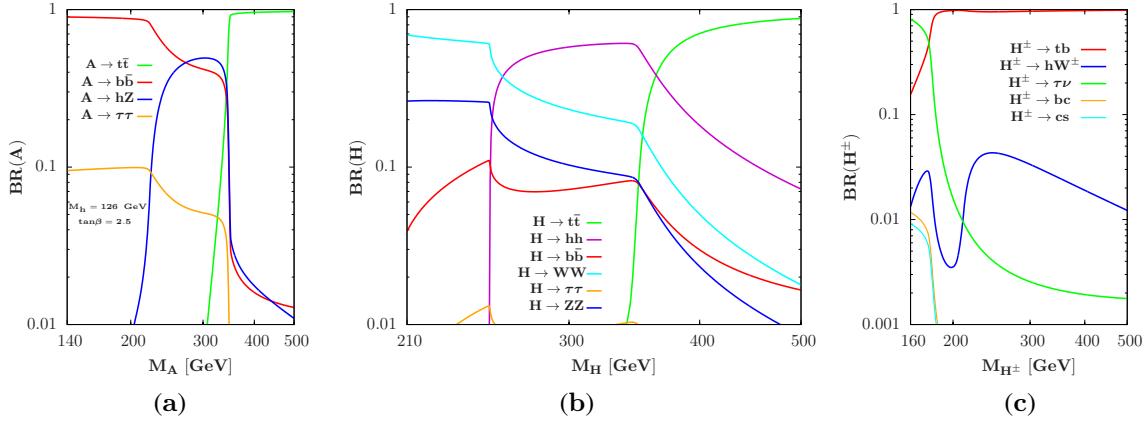


Figure 2.12: Heavy Higgs decay branching ratios in MSSM, (a) pseudo scalar Higgs A, (b) Scalar Higgs H and (c) charged Higgs (H^\pm) with $\tan\beta = 2.5$ and SM Higgs mass is 126 [33].

searched particle in different ways. Most particle discoveries are carried out by simply confirming a resonant peak in the invariant or transverse mass distribution above the continuum backgrounds, like J/Ψ meson, W/Z bosons, Higgs (h) boson, and top quark. Currently, most new LHC physics searches focus on the same search strategy by examining the access above the continuum background. But in some cases, the kinematics (mass, p_T) of the searching particle are not pure Breit-Wigner (BW) resonance peak but instead, a modified shape from the interfering background or other resonance. The complicated dip-peak or peak-dip structure is more sensitive to new physics than the simple resonant peak. This was first realised in the $gg \rightarrow \Phi \rightarrow t\bar{t}$ channel [35]. This channel, along with interference effect, is studied in this thesis and will be explained in detail after the general introduction to the interference is given.

Consider a two-body scattering $ab \rightarrow cd$ whose partonic differential cross section w.r.t angular variable $z \cos \theta^*$, where θ^* is the scattering angle in the centre-of-mass frame, can be expressed as [36]:

$$\frac{d\hat{\sigma}}{dz} = \frac{1}{32\pi\hat{s}} \sum \left| \mathcal{A}_{bg} e^{i\phi_{bg}} + \frac{M^2}{\hat{s} - M^2 + iM\Gamma} \cdot \mathcal{A}_{res} e^{i\phi_{res}} \right|^2. \quad (2.41)$$

In Eq. 2.41, the sum pertains to all spin and colour components; \hat{s} is the partonic energy and $\phi_{bg,res}$ is the complex phase of background and resonant part respectively. \mathcal{A}_{bg} shows the continuum background amplitude and the term \mathcal{A}_{res} is the resonance amplitude for the searching particle with mass M and width Γ . Expanding the square and re-arranging, we get Eq. 2.42 where we define new terms, σ_{int} , R , ω and $\phi = \phi_{res} - \phi_{bg}$ with definitions in Eq. 2.43.

$$\hat{\sigma} = \hat{\sigma}_{bg} + \frac{M^4}{(\hat{s} - M^2)^2 + M^4\omega^2} \times \left[\frac{2(\hat{s} - M^2)}{M^2} \hat{\sigma}_{int} c_\phi + \hat{\sigma}_{res} \left(1 + \frac{2\omega}{R} s_\phi \right) \right] \quad (2.42)$$

where $c_\phi = \cos\phi$ and $s_\phi = \sin\phi$.

$$\begin{aligned} \hat{\sigma}_{bg,res} &= \frac{1}{32\pi\hat{s}} \int dz \sum \mathcal{A}_{bg,res}^2, \\ \hat{\sigma}_{int} e^{i\phi} &= \frac{1}{32\pi\hat{s}} \int dz \sum \mathcal{A}_{bg} \mathcal{A}_{res} e^{i(\phi_{res} - \phi_{bg})}, \\ R &= \frac{\hat{\sigma}_{res}}{\hat{\sigma}_{int}}, \quad \omega \equiv \frac{\Gamma}{M}. \end{aligned} \quad (2.43)$$

ω , R and ϕ are parameters of interest, with the latter two being relative strength and phase difference between signal resonance and background continuum respectively. The second interference term in equation 2.42 further consists of two parts – the real with c_ϕ and imaginary s_ϕ , which decides the final interference pattern. The well-known peak-dip or dip-peak structure arises from the real interference part with $s_\phi = 0$ with a final shift in the resonance peak position. We focus on the pure resonance as well as the peak-dip structure of the signal in this analysis. The imaginary term also makes the shape interesting at specific conditions of ϕ and

factor C, where C is defined as:

$$C \equiv \left(1 + \frac{2\omega}{R} s_\phi \right). \quad (2.44)$$

At $\phi = -\pi/2$ and $C < 0$, the $gg \rightarrow \Phi \rightarrow AB$ line adopted a pure dip shape, as shown in Fig. 2.13 by the green line. On the contrary, the access can be achieved by $C > 0$ shown by the orange line in Fig. 2.13, where its amplitude can be increased or decreased relative to pure resonance (dashed-orange line) by varying the value of ϕ . At $\phi = -\pi/2$ and $C = 0$, the line shape becomes more interesting when both the real and imaginary part disappear and the signal line now runs parallel to the continuum background. These types of shapes are termed as “nothingness” and need more careful treatment. The interference pattern arising from the purely imaginary part is even around the resonance mass, Fig. 2.13 green and orange line. This type of interference significantly changes the total signal rate and doesn’t depend on the precise magnitude of the resonance width. All the above defined signal structures demand a new search strategy as compared to the simple resonance we seek.

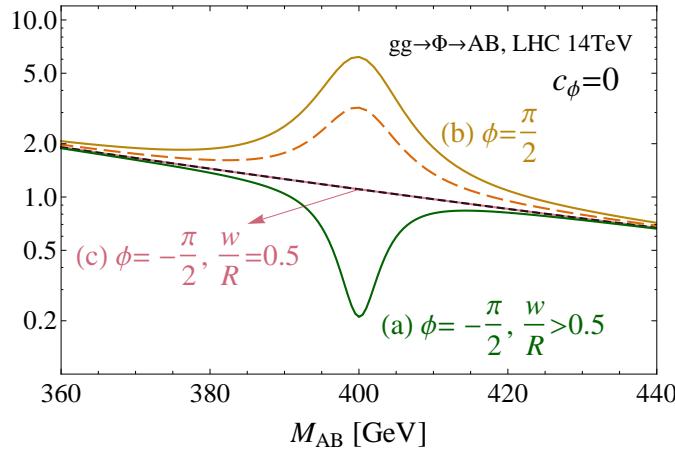


Figure 2.13: Resonance shapes after pure imaginary interference ($c_\phi = 0$) with mass = 400 GeV, $\Gamma = 10$ GeV and $R = 0.035$. Green line is pure dip, orange is access, orange-dashed is pure resonance, red is nothingness and solid black is continuum background. Arbitrary units are used on y-axis [36].

The dip-peak or peak-dip structure around the resonance mass is the most common signal shape studied in the literature. This is the standard case that doesn’t affect the total rate and instead only changes the signal line shape. It arises from the real part of the interference term, c_ϕ in equation 2.42 when $\phi = 0$ or π . From Eq. 2.42, it is evident that the real part (c_ϕ) is odd around the resonance mass compared to the imaginary part (s_ϕ), which is even. That’s why the real interference term doesn’t alter the signal rate and instead provides an arc-type profile around \sqrt{s} values close to the scalar mass. The schematic signal line shape corresponding to the real interference part is shown in Fig. 2.14 by an orange dashed line as a function of the centre-of-mass energy, \sqrt{s} .

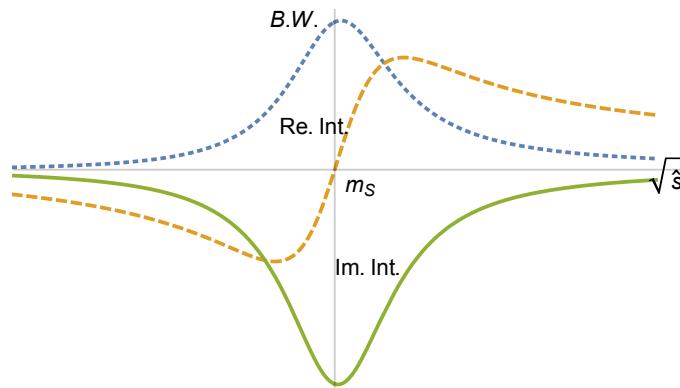


Figure 2.14: The blue dotted line is the signal component, namely the Breit-Wigners resonance; the orange dashed line is the interference between signal and continuum background produced by the real component of the propagator; and the solid green line is the interference proportional to the imaginary part of the propagator as a function of the centre-of-mass energy, \sqrt{s} [37].

Now specifically consider the analysis included in this thesis related to the search for heavy Higgs in the $t\bar{t}$

final state, $gg \rightarrow A/H \rightarrow t\bar{t}$, the heavy Higgs boson ($\phi_i = A/H$) is the additional s-channel contribution shown in leading order diagrams 2.15. Although all quarks can contribute to the fermionic loop, in low $\tan\beta$ region,

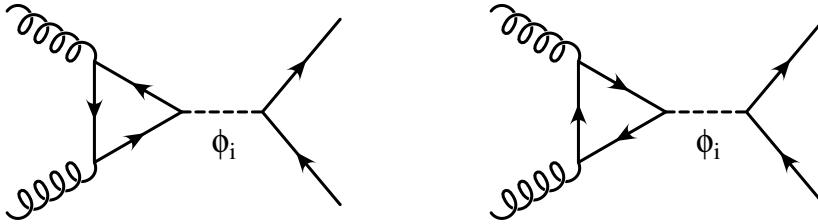


Figure 2.15: Tree-level Feynman diagrams for the heavy Higgs ($\phi_i = A/H$) production via gluon-gluon fusion and decay into fermion (top quarks) pair. The only difference between the two diagrams is the clock and anti-clock wise fermionic loop [38].

the bottom quark loop is strongly suppressed and the top quark is the dominant candidate in the loop. Other lighter quarks have less contributions than the b-quark and can safely be neglected. With higher mass than the $t\bar{t}$ threshold, $M_{A/H} \geq 2m_t$, and low $\tan\beta$, heavy Higgs decays exclusively into a top quark pair. The SM background $gg \rightarrow t\bar{t}$ diagrams are shown in Fig. 2.4, produced by hadron colliders at high energy, interfere with the signal. The signal amplitude for \mathcal{CP} -even and odd state can be written in terms of fermionic loop (top quark loop) function:

$$\mathcal{A}^{\mathcal{CP}-\text{even}} = y_t^2 A_{\frac{1}{2}}^H(\tau_t), \mathcal{A}^{\mathcal{CP}-\text{odd}} = \tilde{y}_t^2 A_{\frac{1}{2}}^A(\tau_t) \quad (2.45)$$

Where $A_{\frac{1}{2}}^{H/A}(\tau_t)$ is the loop function given by Eq. 2.33; the constants in the equation have been omitted for simplicity. I want to discuss the behaviour of the \mathcal{CP} -even and odd states based on the fermionic (top quark in our case) loop functions here, as they are the only source of additional phase ϕ . In Fig. 2.16a, the phase of the loop function is drawn as a function of the of $\sqrt{\tau} \equiv \sqrt{\hat{s}}/2m_t$ – the solid red line for scalar and blue-dashed line for pseudo-scalar particle. On the upper horizontal axis, centre-of-mass energy is drawn in GeV units for the top quark in the loop. At lower values of $\tau \ll 1$, the phase of both states is nearly constant, but just after crossing the threshold of on-shell top pair, $\sqrt{\hat{s}} \approx 2m_t$, at $\tau \approx 1$ the phase changes quickly and at larger τ the variation become slow. The fact is that at $\tau \geq 1$, the imaginary part of the loop function rises quickly and then decreases slowly for increasing values of τ . It is clear that the phase of \mathcal{CP} -odd always reaches faster as compared to \mathcal{CP} -even. In Fig. 2.16b, the phase is drawn as a function of the mass $M_{A/H}$ for the two states, \mathcal{CP} -even (H) blue band and \mathcal{CP} -odd (A) red band. Different $\tan\beta$ (vacuum expectation values for two states) values, 2, 5 and 10 are represented by solid, dashed, and dotted lines respectively. The upper-horizontal axis is labelled with τ values. It shows that \mathcal{CP} -odd state attains higher phase before the \mathcal{CP} -even for the whole mass range. At typical value of $\tan\beta = 2$, \mathcal{CP} -odd reaches to $-\pi/2$ at $M_A = 850$ GeV while \mathcal{CP} -even at $M_H = 1170$ GeV.

Until now, we studied the \mathcal{CP} -conserving scenario, where the \mathcal{CP} -even and odd state don't interfere and the resultant cross section is the sum of the two independent states. In case of nearly degenerate heavy Higgs states (A, H, H^\pm), as predicted by MSSM, the two \mathcal{CP} states interfere with each other. This interference is small in comparison to the SM $t\bar{t}$ background interference but still allows for a rich phenomenology in this channel. Interference between the two \mathcal{CP} states depends on the masses, separation between the masses, widths, and the phase of the two scalars. The new interference produces another minor peak-dip between the two masses, which needs special consideration [37]. We considered only the top quark in the loop and neglected light quarks because of their small contributions. If we consider some moderate value of $\tan\beta$, the bottom quark loop contribution scales as $\tan\beta$, while the top quark scales $1/\tan\beta$. The bottom and top quarks loops can also interfere constructively or destructively, depending upon the relative phase between them. Although the bottom-quark contribution only provides a small correction, it makes the study more interesting. Many BSM theories like composite Higgs models [39] and flavour models motivated the idea of vector-like quarks. These vector-like quarks contribute to the loop and change the phenomenology of this channel. SUSY predicts the scalar partners of the SM fermions which may also contribute to the gluon-gluon-scalar effective coupling. But for higher masses, like heavy stop, the current data is not sufficient to further probe and impose some relevant constraints.

The two observables, mass of $t\bar{t}$ and Collins Soper angle: The search for heavy Higgs boson in the $t\bar{t}$ final state exploits two variables – the mass of $t\bar{t}$ and angular variable $\cos\theta^*$ – for the final statistical evaluation. The top quark production and decay in different channels are discussed in Sec. 2.1.1 and the $t\bar{t}$ system reconstruction will be explained in detail in section 6.5.1. The angular variable, commonly known as Collins-Soper angle [40],

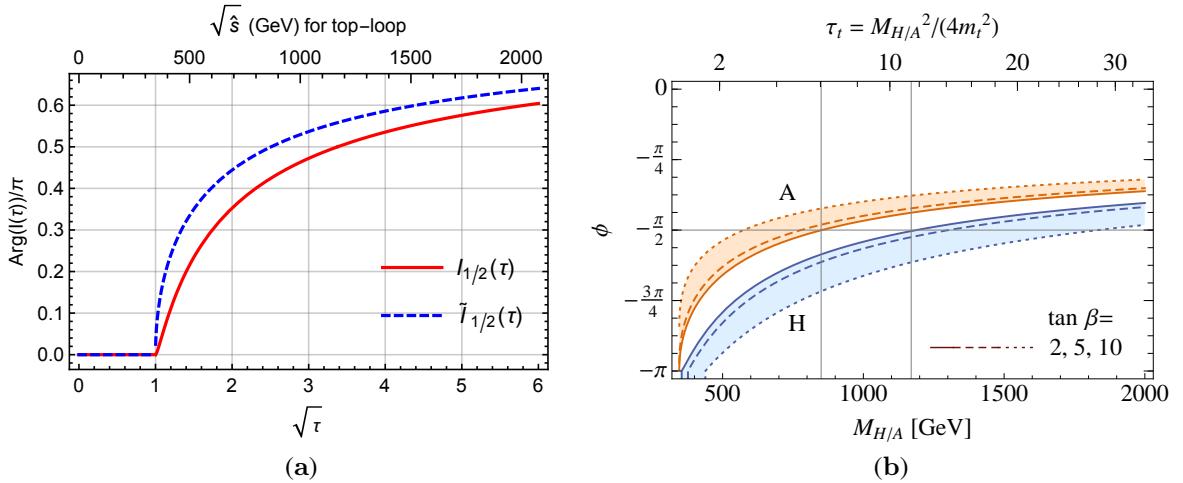


Figure 2.16: (a) The phase of the loop functions in units of π as a function of τ for scalar (red line) and pseudoscalar (blue-dashed line) respectively. The upper-horizontal axis is labelled with the corresponding centre-of-mass energy $\sqrt{\hat{s}}$ in units of GeV for the case of a top quark loop [37]. (b) shows the relative phase ϕ between continuum $gg \rightarrow t\bar{t}$ and resonance $gg \rightarrow \Phi \rightarrow t\bar{t}$ w.r.t $M_{H/A}$ at different $\tan\beta$ values (2, 5, 10) with $\tau_t = M_{H/A}^2/(4m_t^2)$ at the upper-horizontal axis. For all values of $\tan\beta$, the $-\pi/2$ condition achieved by \mathcal{CP} -odd state (A) much faster than the \mathcal{CP} -even state (H) shown by vertical lines [36].

is sensitive to the spin of the intermediate heavy Higgs boson decaying into $t\bar{t}$ and very less affected by the initial state radiation. Consider the collision of two non-collinear hadrons with momenta p_1 and p_2 in the rest mass frame of the $t\bar{t}$. The resultant top and anti-top quark will not be back-to-back and will have non-zero transverse momentum. The Collins-Soper angle θ is defined as being the angle between the axis that bisects the angle between p_1 and p_2 and the top quark momentum in the $t\bar{t}$ rest frame.

In the SM process, $gg \rightarrow t\bar{t}$, the squared amplitude is proportional to the cosine of θ which is shown by a complicated relation 2.46:

$$\begin{aligned} |\mathcal{M}(gg \rightarrow t\bar{t})|^2 &\sim \\ \frac{s(7 + 9\cos^2\theta) - 36m_t^2\cos^2\theta}{(sc_- + 4m_t^2\cos^2\theta)^2} \\ [s^2c_+c_- + 2sm_t^2(3c_-^2 + c_+^2) - 4m_t^2(3c_-^2 + c_+^2 + c_-)], \end{aligned} \quad (2.46)$$

where s is the centre-of-mass energy, m_t is mass of the top quark, $c_+ = 1 + \cos^2\theta$ and $c_- = 1 - \cos^2\theta$.

In case of spin-0 resonance, the amplitude squared is independent of the mass and parity of the resonance and given by 2.47:

$$|\mathcal{M}(gg \rightarrow A/H \rightarrow t\bar{t})|^2 \sim (|a_1|^2 + |a_2|^2)p_t \cdot p_{\bar{t}} - (|a_1|^2 - |a_2|^2)m_t^2, \quad (2.47)$$

Where a_1 and a_2 are the coupling constants for \mathcal{CP} -even and odd state, p_t and $p_{\bar{t}}$ are the top and anti-top quark momenta respectively.

From Eq. 2.46 it is clear that the SM $t\bar{t}$ is minimal in the central region and maximum in the forward and backward regions, while Eq. 2.47 is independent of $\cos\theta$ and gives a flat distribution. The distributions in Fig. 2.17 show resonance pseudo-scalar (A) with red and SM $t\bar{t}$ with blue histograms. The $\cos\theta$ has strong power to discriminate between signals and background, and in this analysis, it has been used as a second variable with the $t\bar{t}$ mass for final statistical evaluation.

Sensitivity in the $gg \rightarrow A/H \rightarrow t\bar{t}$ channel: In the previous section, production from the gluon-gluon fusion via top loop and decay to $t\bar{t}$ of the two \mathcal{CP} states has been discussed in detail along with the SM $t\bar{t}$ interference. A large parameter space in $[M_A, \tan\beta]$ is available to be probed at LHC high luminosity in this channel. A very simple analysis is performed in [28] in order to find the sensitivity in the $t\bar{t}$ final state of a spin-one resonance based on ATLAS and CMS results [41, 42]. In a short introduction, spin-one resonance is produced by $q\bar{q}$ annihilation and doesn't interfere with the (coloured) QCD $q\bar{q} \rightarrow t\bar{t}$. Thus, a simple resonance peak appeared above the continuum SM $t\bar{t}$ background. For signal (masses = 400, 600 and 800 GeV) and background $t\bar{t}$ processes, the events generated using MADGRAPH generator. Signal production cross section and branching

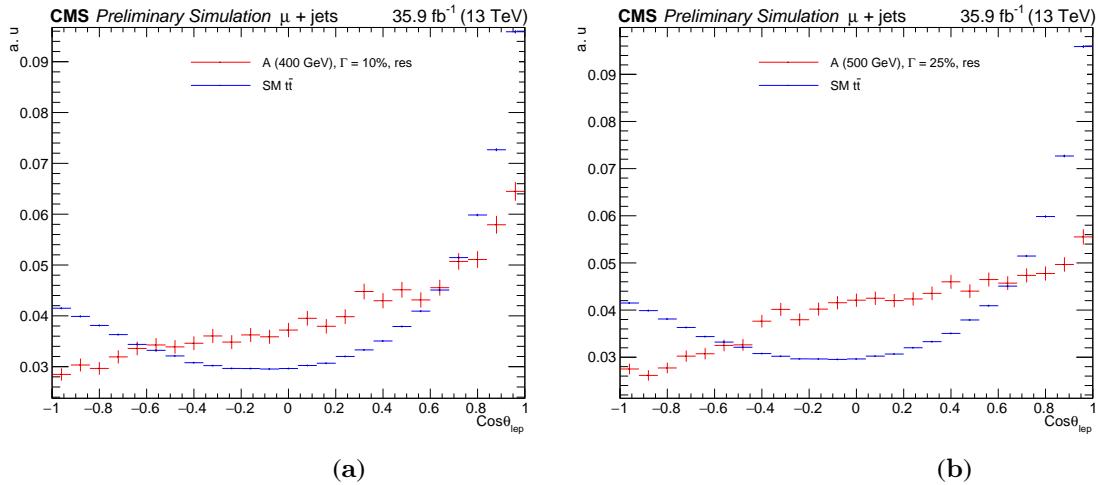


Figure 2.17: The $\cos \theta$ distribution, where θ is the leptonic top (the reconstruction of the $t\bar{t}$ system in the detector depends on the final state of the top quark. For the semileptonic decay mode, one top quark is reconstructed as leptonic and other one as hadronic, with more explanation provided in section 6.5.) quark 3-momentum in the rest frame of the $t\bar{t}$. The red histogram shows the resonance pseudo-scalar (A) with mass = 400 GeV and $\Gamma = 10\%$ in (a) and mass = 500 GeV and $\Gamma = 25\%$ in (b). The blue line shows the SM $t\bar{t}$ distribution in both (a) and (b). Both histograms are scaled to one for shape comparison.

ratios have been calculated using HIGLU and HDECAY programs respectively. The background QCD $t\bar{t}$ cross section has been calculated using TOP++ program with top quark mass $m_t = 173.2$ GeV, $\mu_R = \mu_F = m_t$ and at the order of NNLL including resummation scale.

The number of signal and background events are simply taken with cuts on basic kinematic variables to suppress the $t\bar{t}$ QCD background and probed that region in the parameter space of $[M_A, \tan \beta]$ in the hMSSM context where $N_{sig}/\sqrt{N_{bkg}} \geq s$. The results shown in Fig. 2.18 for $\sqrt{\hat{s}} = 14$ TeV use red, green, blue, and magenta lines for 2, 3, 4, and 5σ sensitivity respectively, which is scaled to 300 fb^{-1} luminosity. With increasing luminosity, the situation could be vastly improved for sensitivity and a 2σ ‘evidence’ or 95% CL exclusion limit can be achieved with $\tan \beta \approx 7$ with $M_A \approx 350$ GeV in the MSSM context. A 5σ discovery can be claimed at $\tan \beta \approx 6$ and $M_A \approx 350$ GeV and at $\tan \beta \approx 1$ and $M_A \approx 800$ GeV.

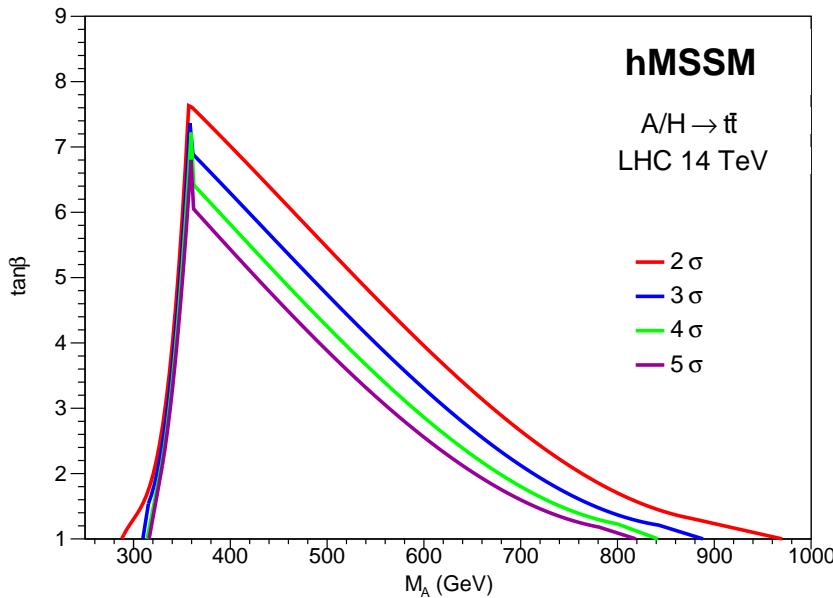


Figure 2.18: Sensitivity levels for $\Phi \rightarrow t\bar{t}$ channel at $\sqrt{s} = 14$ TeV with total integrated luminosity $\mathcal{L}=300\text{fb}^{-1}$ in hMSSM on $[M_\Phi, \tan \beta]$ plane. At $M_\Phi \approx 2m_t$, we have a large $\tan \beta$ space up to 7.5, which shows the sensitivity of the top quark to new physics [28]. In the absence of new physics, the area under the contours will be excluded with corresponding sensitivity.

Constraints on A/H/H $^\pm$ and projections at 0.3 and 3.0 ab $^{-1}$: CMS and ATLAS performed searches for heavy neutral and charged Higgs in different decay channels, fermionic decay channel $A/H \rightarrow \tau\tau$ and $H^\pm \rightarrow \tau\nu$ and bosonic decay channel $H \rightarrow WW, ZZ, hh$ and $A \rightarrow Zh$, that excluded a large parameter space in $[M_A, \tan \beta]$. For large $\tan \beta$ and high mass range, $A/H \rightarrow \tau\tau$ is the most sensitive channel that excludes a large parameter space. The bosonic channels are effective in lower $\tan \beta$ areas, while $H^\pm \rightarrow \tau\nu$ channel is sensitive to all $\tan \beta$ values at low masses. The virgin area in $[M_A, \tan \beta]$ for $M_A \gtrsim 350$ and $\tan \beta \lesssim 4$ is not covered by the aforementioned channels. The $gg \rightarrow A/H \rightarrow t\bar{t}$ channel perfectly occupies this area because of the high $t\bar{t}$ mass threshold, ≈ 350 GeV, demanding a heavier Higgs and coupling of top quark to Higgs proportional to $1/\tan \beta$ (a lower $\tan \beta$ is the best option). It is shown in Fig. 2.19 by a red-dashed contour in the parameter space of $[M_A, \tan \beta]$ in the context of hMSSM along with the shadow area excluded by the combined searches of the CMS and ATLAS collaborations at $\sqrt{s} = 7+8$ TeV at 25 fb^{-1} luminosity up to 2σ sensitivity.

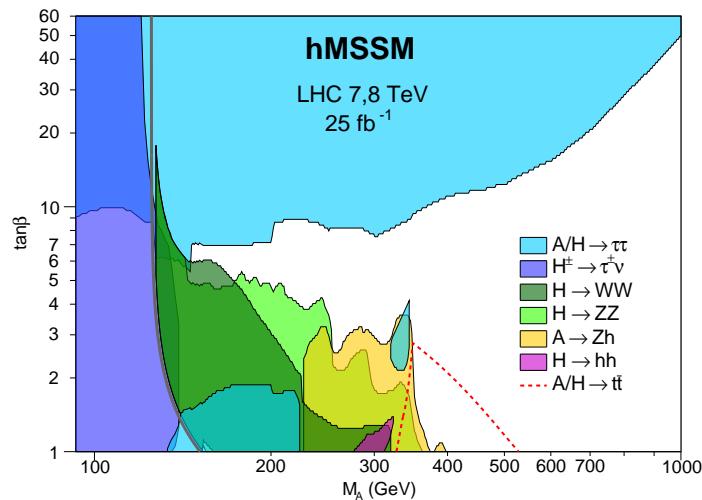


Figure 2.19: The combined constraints applied by the collaboration of CMS and ATLAS on the heavy Higgs, A, H, H^\pm , in the parameter space of $[M_A, \tan \beta]$ in the context of hMSSM, target fermionic and bosonic channels. All the searches are performed up to $\sqrt{s} = 8$ TeV using 25 fb^{-1} data. The $gg \rightarrow A/H \rightarrow t\bar{t}$ channel is shown by a red-dashed contour that covers the high mass and low $\tan \beta$ range where other channels are not reachable[28].

During the next phase of LHC when the integrated luminosity reaches 0.3 and 3.0 ab^{-1} , the exclusion limit will be vastly improved in the absence of new signal, especially the 2σ sensitivity will be drastically enhanced in the next LHC run. To extrapolate the 95% CL exclusion limit for the MSSM Higgs, put together by CMS and ATLAS searches at $\sqrt{s} = 8$ TeV and $\mathcal{L} \approx 20 \text{ fb}^{-1}$, to $\mathcal{L} \approx 0.3 \text{ ab}^{-1}$ at $\sqrt{s} = 14$ TeV, a naive approach has been adopted. The assumption made on the basis that “the results of the experimental analyses are limits on the signal cross section at a given centre-of-mass energy, \sqrt{s} , and fixed integrated luminosity value $\mathcal{L}_{\sqrt{s}}$ for a given resonance mass bin, $R_{\sqrt{s}}^S(M_A)$, for a channel that is subject to a given background rate $R_{\sqrt{s}}^B(M_A)$ at this mass bin.” Once we have knowledge about the $R_8^S(M_A)$ at 8 TeV, an extrapolation related to 14 TeV can be written as:

$$R_{14}^S(M_A) = \sqrt{\mathcal{L}_8/\mathcal{L}_{14}} \sqrt{R_{14}^B(M_A)/R_8^B(M_A)} R_8^S(M_A). \quad (2.48)$$

Further assumption is made for the unknown background rates based on the known signal cross section at certain points, “the background rate increasing linearly to the signal cross section”. This assumption is suitable in the $gg \rightarrow A/H \rightarrow t\bar{t}$ channel, where the main background is also produced by the gluon-gluon fusion, $gg \rightarrow t\bar{t}$, which increases with increasing in energy at LHC. However, this approach is conservative for those channels whose main background is from $q\bar{q}$ annihilation like $H \rightarrow WW, ZZ$ or $A/H \rightarrow \tau\tau$ because with an increase in energy, the rate increases very slowly. With these assumptions, Eq. 2.48 can be written as:

$$R_{14}^S(M_A) \approx \sqrt{\mathcal{L}_8/\mathcal{L}_{14}} \sqrt{\sigma_{14}^S(M_A)/\sigma_8^S(M_A)} R_8^S(M_A). \quad (2.49)$$

Using this approach, the combined CMS and ATLAS results at $\sqrt{s} = 8$ TeV and $\mathcal{L} \approx 20 \text{ fb}^{-1}$ are extrapolated to 0.3 and 3.0 ab^{-1} in 2.20(a) and (b) respectively on $[M_A, \tan \beta]$ plane in the context of hMSSM. In

Fig. 2.20 an extra channel, $H^\pm \rightarrow tb$, is added in comparison to Fig. 2.19, shown by the dark-blue area. At high luminosity, the results are optimistic and at 95% CL, the sensitivity of most channels is highly improved. The $A/H \rightarrow \tau\tau$ channel now covers the whole parameter space in $[M_A, \tan \beta]$ for $M_A \lesssim 350$ GeV. The sensitivity of $gg \rightarrow A/H \rightarrow t\bar{t}$ channel increases and at 0.3 ab^{-1} , it can probe $\tan \beta \lesssim 8$ and M_A 350-1000 GeV in $[M_A, \tan \beta]$ space, while for 3.0 ab^{-1} , its sensitivity is further enhanced to $\tan \beta \lesssim 12$ and M_A 350-1500 GeV. Examining the extrapolated results, especially for 3.0 ab^{-1} , the $gg \rightarrow A/H \rightarrow t\bar{t}$, the channel receives more attention and will play a pivotal role in the heavy Higgs searches.

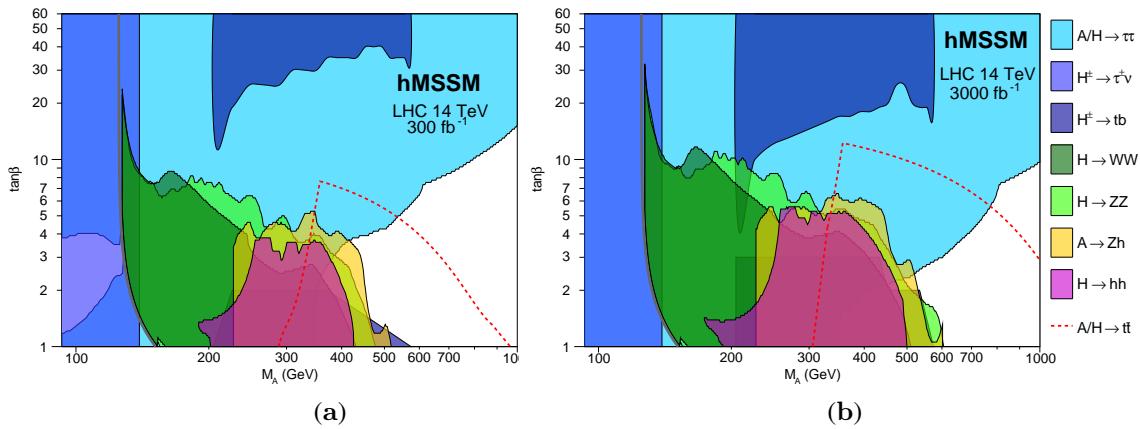


Figure 2.20: Extrapolated from 8 TeV and $\approx 20 \text{ fb}^{-1}$ data, the combined CMS and ATLAS results for neutral and charged Higgs searches with expectation of 2σ at $\sqrt{s} = 14 \text{ TeV}$ using hMSSM with total integrated luminosity $\mathcal{L} = 300 \text{ fb}^{-1}$ (a) and HL-LHC $\mathcal{L} = 3000 \text{ fb}^{-1}$ (b) in the $[\tan \beta, M_A]$ plane. The red-dashed contour shows the $\Phi \rightarrow t\bar{t}$ channel projection, where its sensitivity increases as compared to 8 TeV [28].

3

The CMS experiment at the Large Hadron Collider

This chapter briefly explains the Large Hadron Collider at the laboratories of the European Organization for Nuclear Research (CERN). It further focuses on Compact Muon Solenoid (CMS) detector and its subdetectors, because the data used in this analysis is collected by the CMS experiment. Each of the subsystems is discussed in a separate section; beginning at the centre of the CMS, these are as follows: Tracker, electromagnetic calorimeter, hadronic calorimeter, magnet, and muon system. In the last part, the CMS trigger and data acquisition systems are discussed.

3.1 The Large Hadron Collider at CERN

The Large Hadron Collider (LHC) is a circular protons and heavy ions collider with a circumference of ~ 27 km, 100 m underground, operated at the laboratories of the European Organization for Nuclear Research (CERN¹) in Geneva, Switzerland [43]. The LHC has been designed to operate at centre-of-mass energy $\sqrt{s} = 14$ TeV which makes it the world's most powerful accelerator. Proton-proton (pp) collision is the main part of the LHC physics program; a part of the machine's schedule is periodically dedicated to the delivery of heavy-ion collisions. The accelerator complex at CERN is a succession of machines that accelerate the particle beams in many steps before injecting them into the main ring. Protons are first accelerated in different linear accelerators (LINAC) to the energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), followed by the Proton Synchrotron (PS), which pushes the beam to 25 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV. Most of the pre-accelerators in the chain have their own experimental halls, where beams are used for various purposes, such as other particle-physics experiments, test beams or irradiation of detector material for radiation-hardness studies.

The proton beams are finally transferred to the largest ring of the LHC to attain an energy of 6.5 TeV per beam. This includes two adjacent beam pipes, each containing one of two colliding beams that travel in opposite directions in a collider ring. An ultra-high beam vacuum (10^{10} Torr) is created to avoid possible collision with gas molecules. The beams are guided and focused in a circular trajectory using superconducting (1.9-4 K temperature) dipole and quadrupole magnets. In the LHC, beams are accelerated/decelerated by the electromagnetic field generated by radio-frequency (RF) cavities (eight per beam) located along the collider ring. Each of these cavities also operates in a superconducting state, at a temperature of approximately 4.5 K, and can deliver a voltage of 2 MV at a frequency of 400 MHz. The main role of the RF cavities is to keep the 2808 proton bunches tightly bunched in order to ensure high luminosity at the collision points, hence maximizing the number of collisions (luminosity). The two beams cross each other at four interaction points

¹Conseil Européen pour la Recherche Nucléaire.

where the detectors are installed. The CERN accelerator complex with four interaction points is shown in Fig. 3.1.

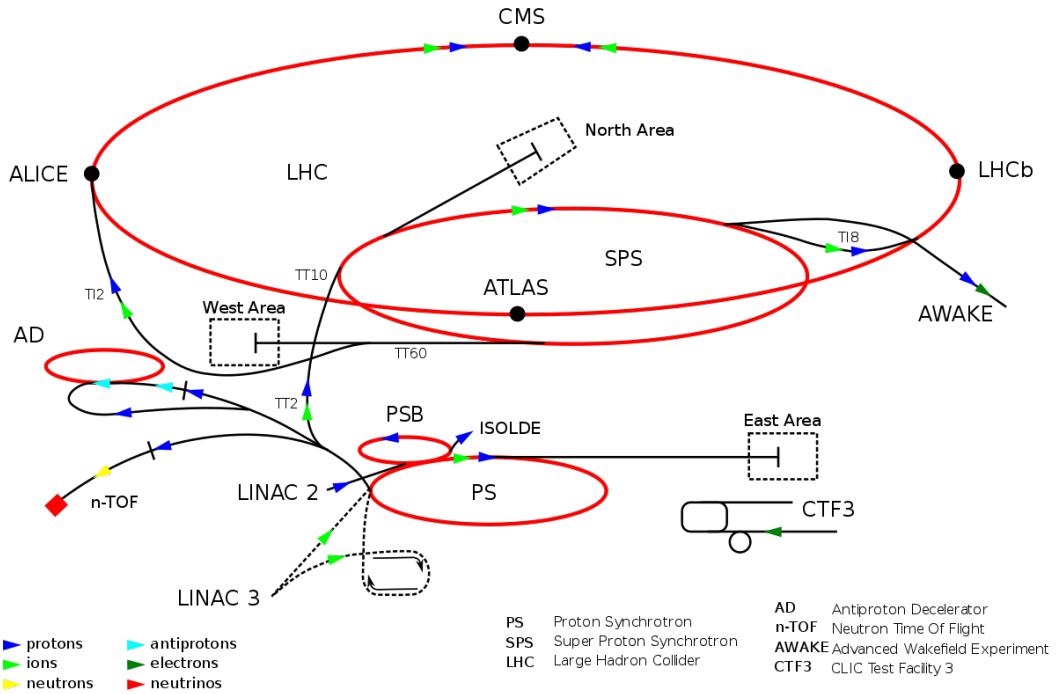


Figure 3.1: The CERN accelerator complex. The proton injection chain for the LHC starts from the LINAC2 and proceeds through the Booster, PS, and SPS.

The performance of an accelerator is measured in terms of instantaneous luminosity \mathcal{L} ; the most important characteristic of a collider is that it ties the event rate to the cross-section (σ) of a process:

$$\frac{dN_i}{dt} = \sigma_i \mathcal{L}(t) \quad (3.1)$$

Where i indicates any general process and $\mathcal{L}(t)$ is the instantaneous luminosity that depends on the number of interactions per unit of time. The machine's instantaneous luminosity is related to the parameters of the beam and can be written as follows:

$$\mathcal{L} = \frac{f_{rev} N_b^2 n_b \gamma_r}{4\pi \epsilon_n \beta^*} F \quad (3.2)$$

Where $f_{rev} = 11$ kHz is the revolution frequency; N_b is the number of particles per bunch; n_b is the number of bunches per beam; ϵ_n is the normalized transverse beam emittance; β^* is the beta function at the collision point, which measures the beam focalization and is corrected by the relativistic gamma factor γ_r ; and F is a geometric luminosity reduction factor that accounts for the crossing angle at the interaction point [44]. The amount of data delivered by a collider is measured in terms of the total integrated luminosity, $L = \int \mathcal{L} dt$, and measured in unit fb^{-1} . Since discoveries in particle physics depend on statistics, the higher the luminosity, the more chances physicists of discovering a particle or process.

During 2016, the LHC was running at centre-of-mass energy $\sqrt{s} = 13$ TeV with a bunch spacing of 25 ns. Reducing β^* parameter from 80 cm to 40 cm, the nominal LHC luminosity crossed the record value of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ in late summer 2016. The instantaneous luminosity delivered by LHC (blue) and recorded by CMS (yellow) is shown in Fig. 3.2(a) and the total integrated luminosity during 2016 in Fig. 3.2(b).

Among the four detectors, two (ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid)) are the largest multi-purpose experiments. Their main focus is the precise measurements of SM and BSM physics. ALICE (A Large Ion Collider Experiment) is designed for heavy-ion collisions to study the physics of strongly interacting matter at extreme energy densities as well as the quark-gluon plasma. The Hadron Collider beauty (LHCb) experiment is designed to study heavy flavour beauty hadron physics.

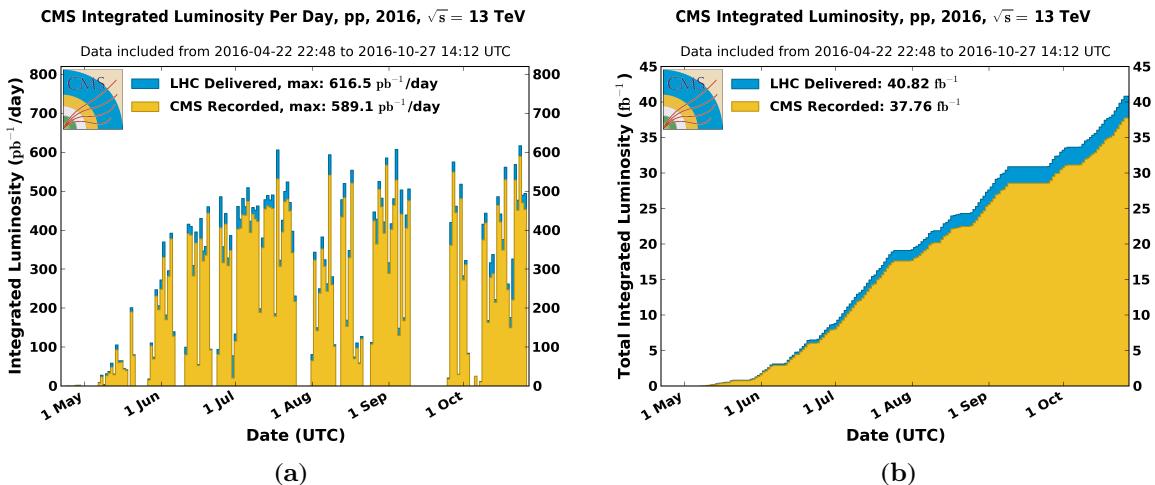


Figure 3.2: The evolution of the instantaneous luminosity per day (a) and the integrated luminosity (b) delivered by LHC (blue) and recorded by CMS (yellow) in 2016. The plots have been taken from [45].

3.2 The CMS experiment at the LHC

The CMS [46] is one of the two multi-purpose experiments installed at the LHC ring to study a wide range of particles and heavy-ion physics. The primary goals are to elucidate the nature of the electroweak symmetry breaking mechanism linked to the Higgs mechanism, search for physics Beyond the Standard Model (BSM), and obtain precise measurements of various SM processes. An overview of the CMS detector and its subsystems is shown in Fig. 3.3. The general concept driving the detector design is the configuration of the magnetic field needed to bend the trajectory of the charged particles, especially muons, to have a good resolution while measuring high momentum. The CMS detector has a superconducting solenoid magnetic producing a field of 3.8 T parallel to the beam axis, encapsulating a high-quality tracking system and calorimeters. The muon system is installed outside the magnet in four stations and interleaved with return yoke. The magnetic field is strong enough to saturate 1.5 m of iron. Each muon station consists of several layers of aluminium drift tubes (DT) in the barrel region (coaxial to the beam axis) and cathode strip chambers (CSCs) in the endcap region (perpendicular to the beam axis), complemented by resistive plate chambers (RPCs). The CMS detector is relatively compact in spite of its heavy weight, 12,500 Tons (as compared to ATLAS), with a length of 21.6 m and a diameter of 14.6 m.

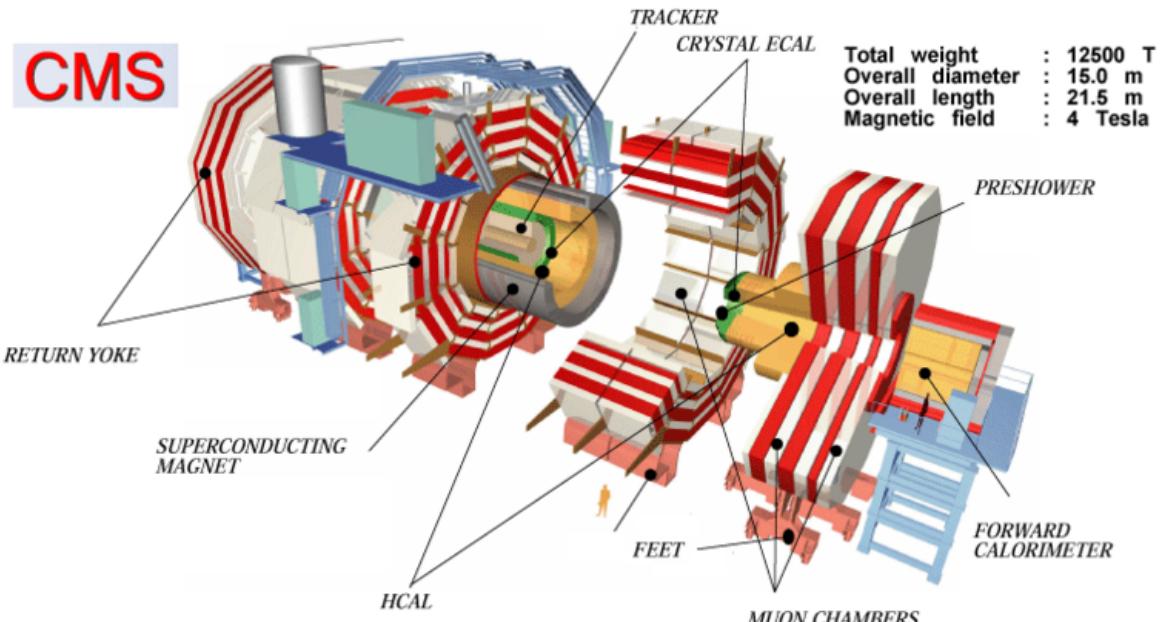


Figure 3.3: An overview of the CMS detector showing the major sub detectors.

The CMS experiment adopted a cylindrical coordinate system with its origin at the nominal interaction point at the centre of the detector. The z-axis is chosen along the anti-clockwise beam and is referred to as longitudinal. The x-axis points radially towards the centre of the LHC ring, and the y-axis is vertical and points upwards as shown in Fig 3.4. The azimuthal angle ϕ is measured on the x-axis in the x-y plane, where the

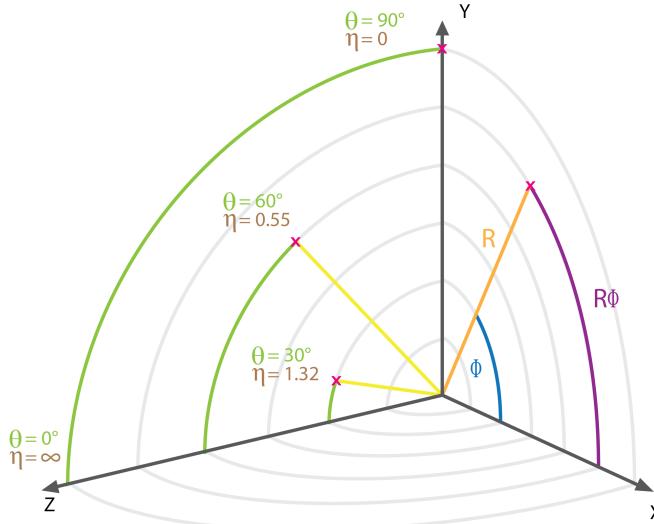


Figure 3.4: An overview of the CMS experiment coordinate system.

x-y plane is perpendicular to the beam axis and called the transverse plane. The polar angle θ is measured on the z-axis. Instead of θ , the angular position is preferentially expressed in terms of a kinematic quantity called pseudorapidity (η), defined as:

$$\eta \equiv -\ln(\tan(\frac{\theta}{2})) \quad (3.3)$$

Adopting the above coordinate system, a particle momentum p has two components, one transverse to the beam direction denoted by p_T and calculated using the x and y components. The second component is the longitudinal one, p_z , pointing along the z-axis. The missing transverse energy E_T^{miss} term is used for imbalance in total transverse energy of a collision. The angular separation between the two particles is usually expressed in the $\phi - \eta$ plane, expressed as:

$$\Delta R = \sqrt{\Delta\Phi^2 + \Delta\eta^2} \quad (3.4)$$

In the following section, a brief summary of all the subsystems of the CMS experiment is given.

3.2.1 Tracker

The tracker is the innermost and closest to the interaction point, sub-detector of CMS, immersed in the homogeneous magnetic field of 3.8 T provided by the solenoid. It is designed to measure the momentum and charge of charged particles emerging from the interaction point by determining the bending of their trajectories in their passage through the detector layers. Secondary vertices can also be reconstructed in association to the late decays of particles such as B hadrons. The trajectory deviation from the straight-line propagation is measured by the sagitta:

$$s \approx \frac{0.3BL^2}{8p_T} \quad (3.5)$$

Where s is sagitta, B is the solenoid magnetic field, L is the track length, and p_T is the transverse momentum. The transverse momentum resolution is mainly dependent on the geometric accuracy on the sagitta (σ_s) using:

$$\frac{\sigma_{p_T}}{p_T} \approx \frac{8p_T}{0.3BL^2} \cdot \sigma_s \quad (3.6)$$

At the LHC, design luminosity and energy (around 1 MHz/mm² at 4 cm from the beamline), several hundred particles go through the tracker during each bunch crossing; high granularity, fast response, and high-radiation tolerance are especially important to its design. In order to fulfil the aforementioned criteria, the semiconductor

silicon technology was chosen to build the inner tracking system [47]. Charged particles traversing a sensor produce electron-hole pairs that drift under an applied electric field, giving rise to a current pulse. Cooling is ensured by liquid perfluorohexane that maintains the sensor temperature at around -10°C to limit the noise due to radiation-induced leakage.

The tracker occupies a cylindrical volume of 5.8 m in length and 2.5 m in diameter with a surface area of 210 m^2 , extends in the region of $|\eta| < 2.5$, $r < 120\text{ cm}$, $|z| < 270\text{ cm}$. A schematic view of the CMS tracker is shown in Fig. 3.5.

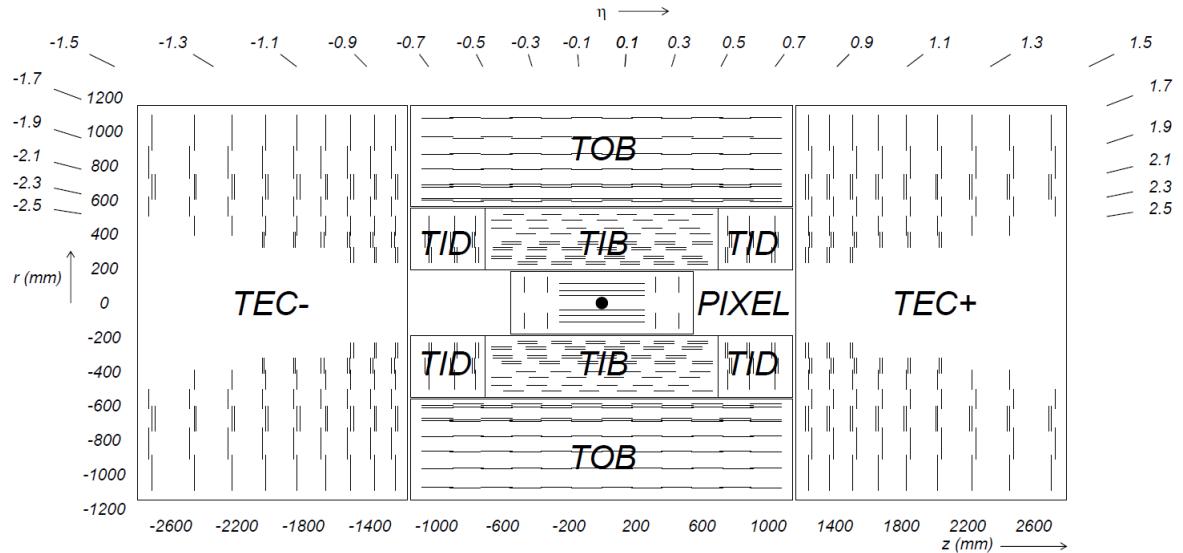


Figure 3.5: The CMS tracker system with its subsystems in the (r, z) plane.

The tracker has two sensors classes – pixel and strip detectors.

- **Pixel Tracker System:** is placed closest to the interaction point ($r \leq 10\text{ cm}$) because it has the largest track density. It is comprised of a total of approximately 66-million pixel cells grouped in 1440 modules with a cell size of $100 \times 150 \mu\text{m}^2$. They are arranged in three cylindrical barrel layers with a radii between 4.4 cm and 10.2 cm from the beam line and two endcap discs at each side of the barrel, approximately 34.5 cm and 46.5 cm from the interaction point. With a high reconstruction hit efficiency well above 99%, the pixel detector covers a pseudorapidity range $|\eta| < 2.5$, corresponding to the acceptance of the entire tracker. The typical spatial hit resolution is measured to be around $10 \mu\text{m}$ in the $r - \phi$ plane and $15 \mu\text{m}$ along the z -axis, while the third coordinate is given by the sensor plane position, allowing for a three-dimensional vertex reconstruction.
- **Strip Tracker System:** occupies the outer part of the tracker cope with reduced particle density. The silicon strip detector is made of two concentric sets of layers in the barrel (tracker inner barrel, TIB, and tracker outer barrel, TOB) occupying the region $20\text{ cm} < |r| < 55\text{ cm}$ and $|z| < 118\text{ cm}$ and two blocks of forward disks in the endcaps, called tracker endcaps (TEC) and tracker inner disk (TID) covering the region with $55\text{ cm} < |r| < 116\text{ cm}$ and $|z| < 118\text{ cm}$. The single-point resolution is about $30 \mu\text{m}$ in the $r - \phi$ plane and $300 \mu\text{m}$ in the z direction.

3.2.2 Electromagnetic calorimeter (ECAL)

The CMS Electromagnetic Calorimeter (ECAL) [48] is a hermetic, homogeneous, and high-granularity system made of inorganic scintillating crystals. ECAL precisely measures the position and energy of electrons and photons, inducing electromagnetic showers in the material. The physical process that drives the design of the ECAL is the low-mass Higgs decay into two photons – $H \rightarrow \gamma\gamma$ – one of the leading channels for the study of the Higgs properties. ECAL provides excellent energy resolution in order to maintain the advantage of a narrow width in the low Higgs mass range. Its role in the $H \rightarrow ZZ \rightarrow 4l$ channel is also important with regard to reconstructing the electron along with the tracker. ECAL crystals are made of lead tungstate (PbWO_4) scintillating crystals that act both as absorbers with a short radiation length ($X_0 = 0.89\text{ cm}$) and as scintillators with a very fast response (80% of the light being emitted within 25 ns), which deal with the LHC bunch spacing. Avalanche Photodiodes (APD) are used in barrel and Vacuum Phototriodes (VPT) in endcaps to

detect scintillation light with wavelengths around 420 nm. An overview of the CMS ECAL is shown in Fig. 3.6.

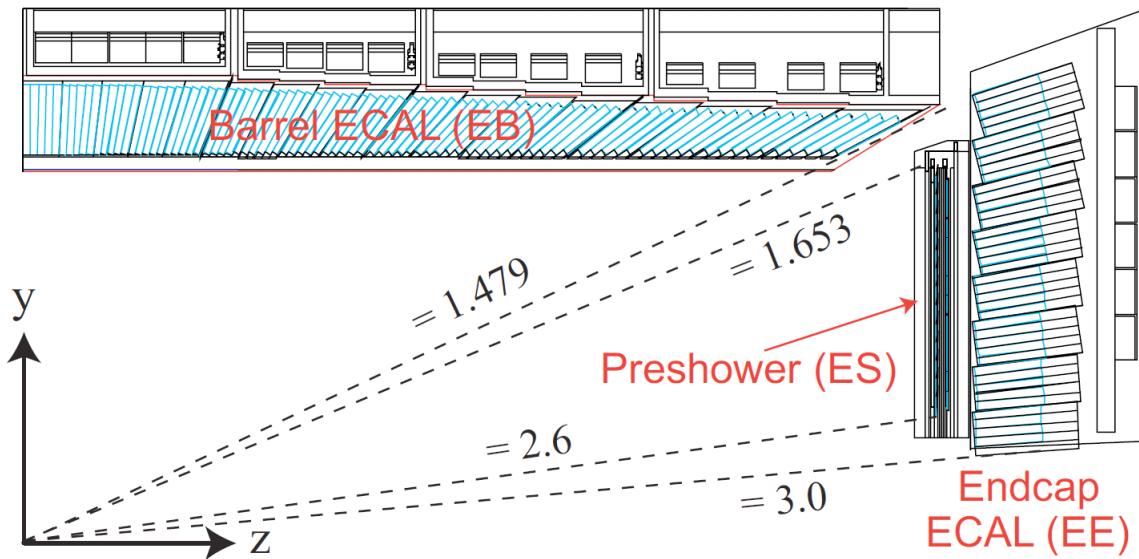


Figure 3.6: The ECAL longitudinal overview of a quadrant.

- **ECAL Barrel (EB):** is located at a distance $r = 129$ cm from the beam axis and covers a pseudorapidity range of $0 < |\eta| < 1.48$. It is equipped with 61,200 crystals with a high granularity of 0.0174×0.0174 rad in $\eta - \phi$ plane, each with a transverse section of 22×22 mm 2 at the front face and a length of 230 mm ($25.8X_0$). Crystals are grouped in arrays of 2×5 , contained in a very thin 200 μm alveolar structure, each corresponding to a sub-module. A group of 40/50 sub-modules are further assembled in modules, where four modules make a super-module. Finally, EB is divided into 36 super-modules, each subtending an angle of 20° in ϕ .
- **ECAL Endcaps (EE):** is located at a distance of $|z| = 315.4$ cm from the interaction point and covers a range of $1.479 < |\eta| < 3$ with identically shaped crystals, grouped in a carbon-fiber structure of 5×5 elements, called a super-crystal. The crystals have a rear face cross section of 30×30 mm 2 , a front face cross section of 28.62×28.62 mm 2 and a radiation length of 220 mm ($24.7X_0$). Each EE has 134 identical supercrystals with a further 18 sectioned supercrystals to complete the inner and outer perimeter.
- **ECAL Preshower (ES):** detectors are placed at each end of tracker, in front of EE and cover a pseudorapidity range $1.653 < |\eta| < 2.6$. It is a two-layer sampling calorimeter consisting of lead radiator layer with a total of $3X_0$ that initiates electromagnetic showers from incoming particles, and silicon strip sensors that measure the deposited energy and transverse shower profiles. The ES help distinguish $H \rightarrow \gamma\gamma$ decays from single photons, and to identify electrons against minimum ionizing particles.

Energy resolution of ECAL for energies below than 500 GeV is given by:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.7)$$

where S is the stochastic term arising due to fluctuations in the lateral shower containment, photostatistics and preshower energy deposition, the second term quantifies the effect of electronic noise and pile-up and a constant term is related to the non-uniformity of longitudinal light collection, calibration uncertainties and energy leakage from the back. ECAL barrel module's performance is studied using the test-beam data, typical values obtained are $S = 2.8\%$, $N = 12\%$ and $C = 0.30\%$ [49];

3.2.3 Hadronic calorimeter (HCAL)

The CMS hadronic calorimeter (HCAL) [50] is a hermetic sampling calorimeter which measures precisely the energy and position of charged and neutral hadrons in the form of jets. It also provides an indirect measurement

to non-interacting particles in terms of missing energy (E_T^{miss}). HCAL surrounds the ECAL and the design is restricted by geometrical dimensions of the ECAL and magnet systems. A schematic overview of the HCAL is shown in Fig 3.7 with its four sub-detectors. The HCAL Barrel (HB) comprise of two half-barrel sections

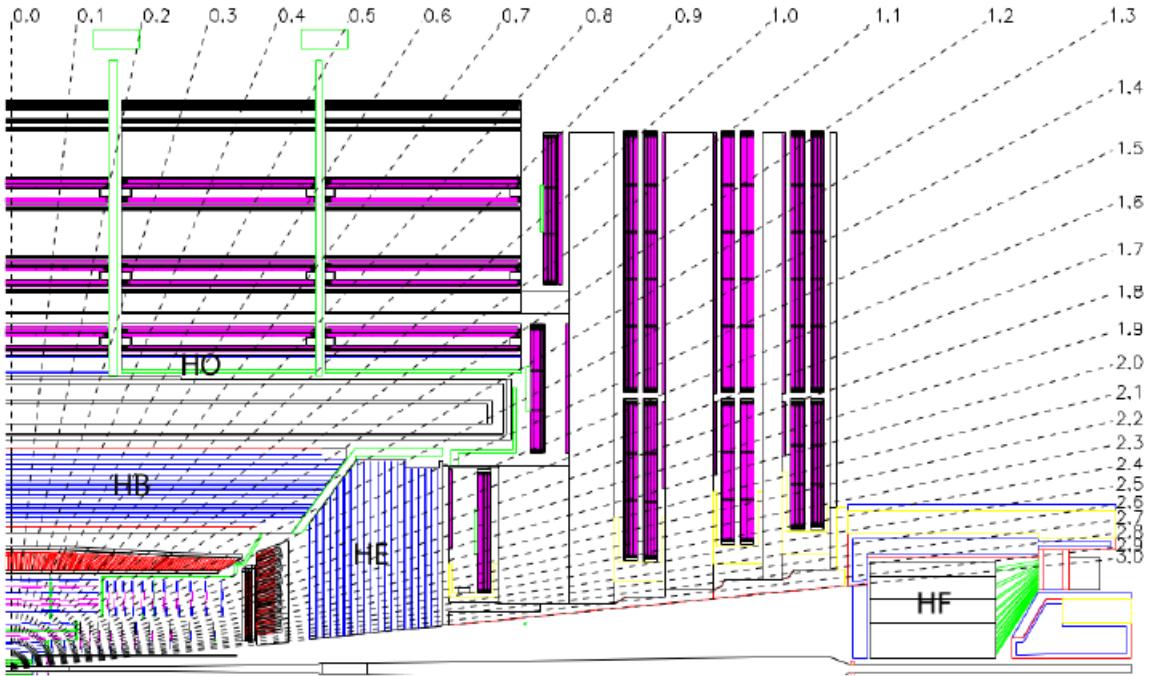


Figure 3.7: Longitudinal overview of the one quadrant of the CMS HCAL.

covering a total pseudorapidity range of $|\eta| < 1.3$, located between EB with $r = 1.77$ m and the inner extent of the magnet coil ($r = 2.95$ m), corresponds to interaction lengths of $5.82 \Lambda_I$ in the central region. Due to space restriction, the HB is complemented by an additional layer of scintillators outside the solenoid, referred to as the Hadronic Outer (HO) calorimeter. HO provides additional depth up to a minimum $11.8 \Lambda_I$ to the HCAL system in the radial direction and have same characteristics as HB.

The HCAL Endcap (HE) covers forward region in pseudorapidity range of $1.3 < |\eta| < 3.0$ and has sufficient depth around $10 \Lambda_I$. HB and HE use non-magnetic brass layers as absorber material, and are interspersed with plastic scintillator tiles which serve as the active medium. Two forward hadron calorimeters (HF) are installed which cover the very forward region in pseudorapidity given by $3.0 < |\eta| < 5.0$, positioned at $|z| = 11.2$ m from the interaction point, thus ensuring good hermeticity. HF detectors using steel as absorber and Cherenkov-light-emitting quartz fibres are as active medium, motivated by the high particle flux in this region. The energy resolution of the HCAL system can be expressed in stochastic term a and constant term b as:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + b^2 \quad (3.8)$$

3.2.4 Magnet

The CMS detector uses solenoidal magnetic field which is the core of detector and hence the name of experiment. A strong magnetic field is needed to achieve good resolution of the high momentum charged particle up to 1 TeV by measuring curvature while direction gives charge of particle [51]. The structure of superconducting magnet for CMS has 6 m diameter and 12.5 m length, able to generate a uniform magnetic field of 3.8 T with a stored energy of 2.3 GJ at full current of 19500 A. A graphical overview of the CMS magnetic field is shown in Fig. 3.8. The magnet mainly uses a superconducting coil (ensured by a helium cooling system at temperature 4 K), a vacuum tank (responsible to isolate it from the external environment) and a 10000 ton magnet yoke comprising five wheels and two endcaps (necessary to return magnetic flux, which otherwise would get lost, disturbing the surrounding environment).

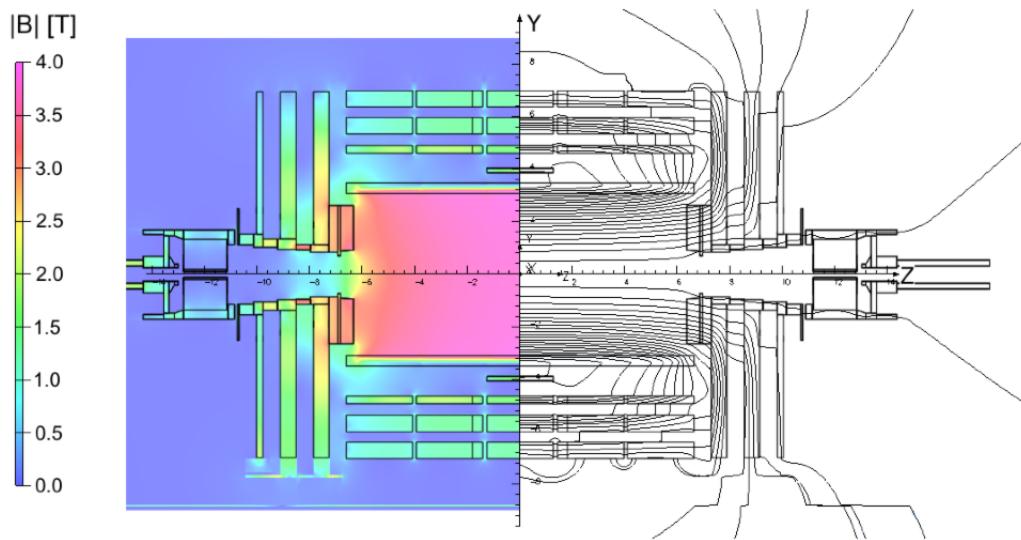


Figure 3.8: Longitudinal view of the CMS detector magnet where left side shows simulation of the magnetic field and the magnetic field lines are at the right side. At the heart of the detector, the magnetic field value is 3.8 T.

3.2.5 Muon spectrometer

Muon plays a key role in the search for many physics phenomena, ranging from precise measurement of the SM of particle physics to new searches, especially in the discovery of SM Higgs boson through golden channel ($H \rightarrow ZZ \rightarrow 4l$). Being a massive particle as compared to electrons, the muon is less effected by bremsstrahlung radiations and traverses through all sub detectors. Owing to these characteristics, the muon system is placed behind all the sub detectors. It is embedded in the return yoke of the solenoid. Muons bring a highly clean signature to the spectrometer because other particles are stopped by the calorimeters. The CMS muon spectrometer covers a pseudorapidity range to $|\eta| < 2.4$ and consists of three subsystems, Drift Tubes (DTs), Cathode Strips Chambers (CSCs), and Resistive Plate Chambers (RPCs) [52], as shown in Fig. 3.9.

- **Drift Tubes:** In the barrel region of the CMS detector, the magnetic field is mostly uniform and the muon rate is low. Four layers of muon stations are installed, consisting of drift tube (DTs) chambers covering a pseudorapidity region of $|\eta| < 1.2$. In each wheel, DTs are installed into 12 ϕ -segments, forming four stations in the radial direction and interleaved between the plates of the return yoke. In the longitudinal (except MB4) and bending plane, each station consists of four and eight layers of DTs respectively to enable the precise measurement of position.

DTs consist of an individual cell with grounded walls acting as cathode and a 50 μm -diameter gold-plated stainless-steel anode wire at the centre of the cell. The drift electric field is created by two electrode plates mounted at the two sides of the cell. The cells use a gas mixture of 85% of Ar and 15% CO₂ with HV settings: $V_{\text{wire}} = +3600\text{V}$, $V_{\text{cath}} = -1800\text{ V}$ and $V_{\text{strip}} = +1800\text{V}$ [53].

- **Cathode Strip Chambers:** Cathode Strip Chambers (CSCs) are installed in endcap regions of pseudo-rapidity coverage $0.9 < |\eta| < 2.4$. Due to the higher muon flux, background radiation, and non-uniform magnetic field in the endcap region, CSC detectors are designed to be robust, fast, radiation hard, and finely segmented. CSCs operate as standard multi-wire proportional counters, comprising six planes of anode wires interleaved among seven cathode strips. The anode wires run azimuthally and identify the radial component of a track hit. The cathode strips are oriented radially, almost perpendicular to the wires and provide a precise measurement in the $r-\phi$ bending plane. The CSCs system uses a nominal gas mixture of 40% Ar, 50% CO₂ and 10% CF₄.

- **Resistive Plate Chambers:** The muon system of the CMS experiment comprises of a third type of muon detectors, Resistive Plate Chambers (RPCs), installed both in the barrel and endcap regions along with DT and CSC, covering a pseudo rapidity range of $|\eta| < 1.6$. RPCs are gaseous parallel-plate detectors that operate in avalanche mode and are capable of tagging the time of an ionizing event in times less than 25 ns (bunch crossing time at the design luminosity of the LHC). This makes RPCs an ideal trigger system that provides correct bunch-crossing time information between two successive bunches with muons, even at the largest LHC luminosities. CMS RPC detector is a double-gap chamber where each gap separates two

parallel plates of bakelite with a bulk resistivity of 10^{10} – $10^{11} \Omega\text{cm}$. A gas mixture of 95.2% $\text{C}_2\text{H}_2\text{F}_4$, 4.5% i- C_4H_{10} , and 0.3% HF_6 is used in RPC operation, where $\text{C}_2\text{H}_2\text{F}_4$ is ionized by incident particles and the other two gases act as quenching agents, preventing the detector from streamer mode. The operational voltage of RPCs is lower than 10 kV with an efficiency of about 100%.

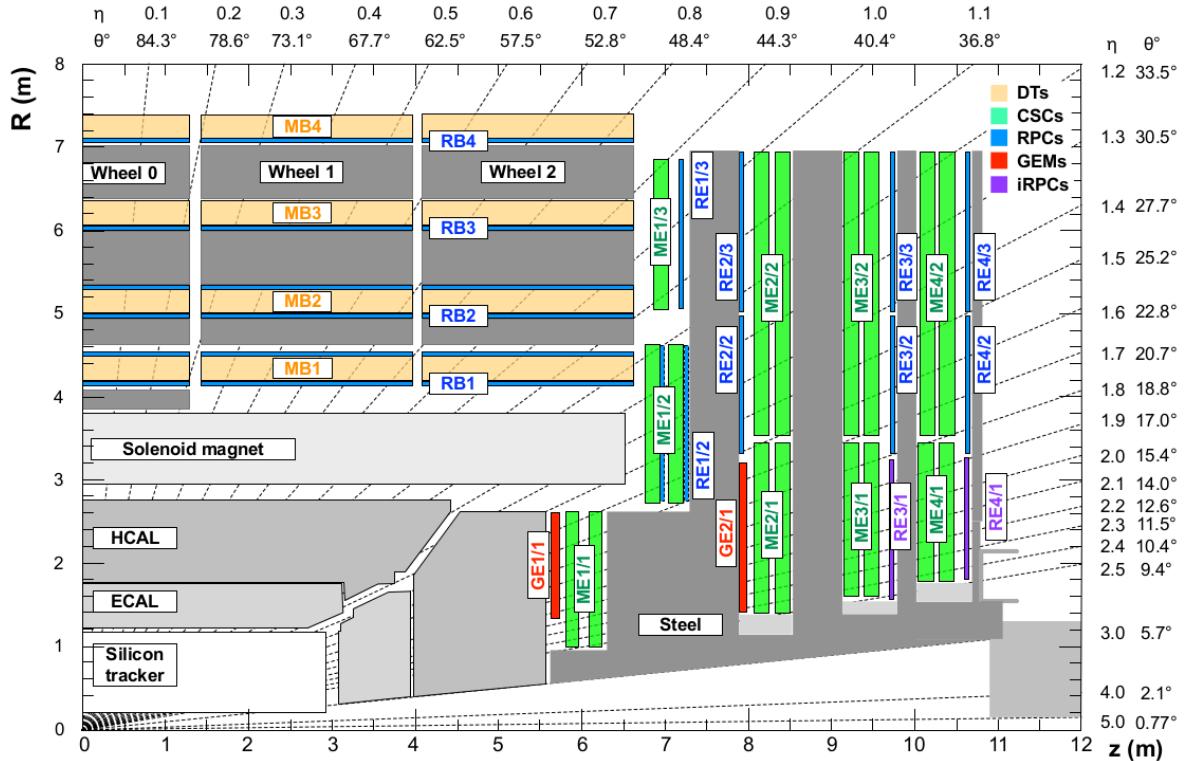


Figure 3.9: Quadrant of CMS muon system in (R, z) plane, where R is perpendicular to the beam axis and z is the distance in meters along beam axis. DTs are shown using light oranges, CSCs by green, and RPCs by blue.

3.2.6 Trigger and data acquisition systems

The protons in the LHC beam are 25 ns apart, which corresponds to 40 MHz. On an average, there are 20 proton-proton collisions per bunch crossing. This produces an enormous amount of data, and the average size of an LHC event is around 1 MB. Due to the limitations of detector performance and storage system, the event rate needs to be reduced to few hundred events per second. The CMS experiment uses a dedicated two-level (Level-1 and HLT) trigger system to control the event rate [54].

L1 Trigger: A hardware-based trigger system that reduces the event rate to 100 kHz with a maximum decision time of $4\ \mu\text{s}$ per event. Because of the short decision time, it only uses the calorimeter and muon system information with no tracker information. Local triggers reconstruct primitive particle candidates in each component of a given subdetector, which are combined using regional triggers to construct higher level L1 objects (muons, electrons, and jets) and are subsequently merged into a global trigger that decides whether to select or reject the event.

HLT: After L1 trigger, the event rate is further reduced to 40 Hz using high-level trigger (HLT), a software-based trigger that uses information from all the subsystems. The reconstruction and selection used by the HLT software is similar to the offline reconstruction and selection that takes place in two successive stages and follows the basic principle of time minimization for each event. In first stage (Level-2), HLT takes the input from L1 and reconstructs the basic objects from calorimeters and muon system. In the second step (Level-3) of the HLT, the complete information from the tracker is accessed for track reconstruction and vertices. The event is finally selected and permanently stored for offline analysis if the requirements of at least one HLT path are met.

4

Detector control system for RPCs at GIF++

This chapter summarizes the High Luminosity LHC (HL-LHC) program and the setup of the Gamma Irradiation Facility (GIF++), where the already-installed and future detectors are being tested. It further briefly explains the WinCC-OA program and the detector control system for the CMS RPCs project at the GIF++. The last section is devoted to the RPC performance study at the GIF++ setup. The work described in this chapter is documented in the publication [55].

4.1 The HL-LHC program and the CMS detector upgrade

The physics program using the Large Hadron Collider (LHC) began in 2009 at low energy (1.18 TeV per beam), which has increased gradually to 7 and 8 TeV in the following years. The two-year duration is termed as Run-I with a total delivered integrated luminosity of around 30 fb^{-1} for its major experiments, CMS and ATLAS, at 50 ns bunch spacing with peak instantaneously luminosity being $7 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. The highlight of Run-I has been the observation of the Higgs boson that lead the high-energy physics community to foresee a challenging upgrade of the accelerator, known as the High Luminosity LHC (HL-LHC). After successful operation of Run-I, LHC was long shut down (LS1) for a two-year period (2013–14) in order to increase the centre-of-mass energy to near its nominal value. In the next LS2 period’s (2019–20) phase-I upgrade, the injector chain will be improved and upgraded to deliver extremely bright bunches (high intensity and low emittance) into the LHC. In LS3 (2023–26), with some basic changes during LS2, the LHC itself will be upgraded with new HL-LHC conditions.

After LS1, the LHC started Run-II in 2015 at a record centre-of-mass energy of 13 TeV with bunch spacing of 50 ns, decreased to 25 ns in 2016. The LHC successfully run for two years (2016–17) and delivered about 80 fb^{-1} physics data and is expected to deliver more than 100 fb^{-1} physics data with nominal instantaneous luminosity being $5 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ at the end of Run-II. Run-III will last for about three years (2021–23) with an anticipated integrated luminosity of 300 fb^{-1} .

HL-LHC is foreseen to begin operations around 2026, following the so-called “Phase-II” machine upgrade at designed centre-of-mass energy $\sqrt{s} = 14 \text{ TeV}$ proton-proton collisions with instantaneous luminosities of between $5\text{--}7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ and expected 140 pile-up events on average. An integrated luminosity of 3 ab^{-1} is expected to be delivered during the ten-year run. This unprecedentedly large dataset will facilitate a wide spectrum of physics analyses, sensitive to high statistics, from precision tests of the Standard Model to New Physics searches with significantly enhanced discovery potential [56]. An overview of the LHC and HL-LHC schedule is shown in Fig. 4.1.

The HL-LHC conditions will be achieved using novel cutting-edge technology. As the first step, the inner triplet quadrupole magnets in the insertion regions will be replaced with advanced and more radiation

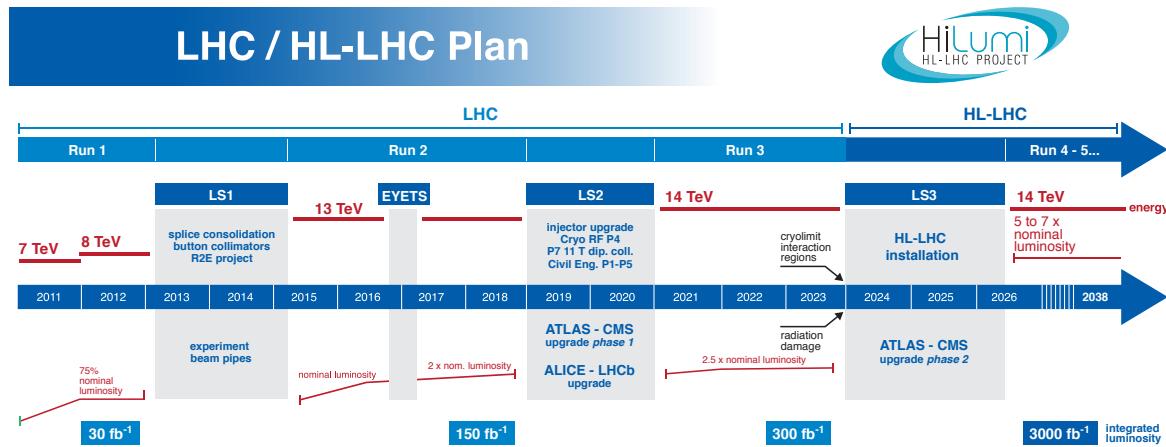


Figure 4.1: The LHC schedule summary and HL-LHC baseline plan with centre-of-mass energy in the upper thick-red line and the instantaneous luminosity lower red-thin line. Run-I started in 2011 at 7 TeV subsequently increased to 8 TeV in 2012, delivered up to 30 fb^{-1} in the two years. It was followed by a long shut down for two years (2013–14), upgraded to 13 TeV centre-of-mass energy and finally Run-II started in 2015 and will end in 2018 delivering more than 100 fb^{-1} data, which will be followed by LS2 for two years. Run-III will span for three years (2021-23) at the designed centre-of-mass, 14 TeV, with anticipated integrated luminosity of 300 fb^{-1} . The HL-LHC program will start from 2024 with three years detectors up-gradation (LS3) followed by ten years Run-IV and Run-V. The instantaneous luminosity is expected to increase up to 5-7 times the nominal value at 14 TeV centre-of-mass energy with expected total integrated luminosity of 3 ab^{-1} [57].

tolerant ones based on Nb₃Sn superconducting technology. To effectively collide the proton beams head on, special superconducting radiofrequency crab cavities will be installed in the interaction regions to rotate both beams longitudinally with half of the crossing angle. Further HL-LHC improvements are: new technology and physical processes for beam collimation and 300-metre-long high-power superconducting links with negligible energy dissipation [58].

The high radiation environment at the HL-LHC will also introduce challenges to the installed detectors. Therefore, it is planned to perform upgrades of several sub-detectors of the CMS to maintain the excellent performance of the Phase-I detector under challenging HL-LHC conditions. A short description of the upgrades of the CMS sub-detectors is listed here.

- **Tracker:** is installed near the interaction point and received a higher radiation dose. For HL-LHC, the tracker must have higher radiation tolerance to be able to operate efficiently up to an integrated luminosity of 3000 fb^{-1} . Other properties are increased granularity to incorporate high pile-up events, improved two-track separation produced by high energy jets, reduced material in the tracking volume, robust pattern recognition, compliance with the L1 trigger upgrade, and extended tracking acceptance.
- **Barrel calorimeters:** As compared to endcaps, the barrel region is not under severe pressure from high radiation. However, in order to maintain exactly the same physics performance, the FE electronics of both sub-detectors (ECAL and HCAL) must be improved to satisfy the new L1 trigger requirements.
- **Endcap calorimeters:** will suffer significant radiation damage during the LH-LHC conditions and needs a re-examination of the ability of the detector's active material and electronics to meet the requirements of 3000 fb^{-1} .
- **Muon detectors:** It is important for Phase-II physics to keep the efficiency of the L1 muon triggers high, while maintaining p_T thresholds low enough to collect a large fraction of Higgs, top quark, and electroweak bosons for conducting more sophisticated analysis. During Run-I and Run-II, no significant deterioration of key chamber parameters has been observed. Therefore, it is expected that the chambers themselves will provide excellent performance throughout the HL-LHC program. Some developments are instead foreseen for the FE electronics in the case of the DTs and the CSCs, which will be replaced with improved versions to increase radiation tolerance, readout speed and performance.

In the very forward regions, muon measurement can be made more efficient if a sufficient number of detector hits is measured for each muon track particular in the small regions not instrumented by the other muon sub-systems – CSC. It will enhance the redundancy of the muon system and enable obtaining

a robust track reconstruction, including rejection of wrongly reconstructed tracks already at the Level-1 trigger. The two endcap regions suffer from high-muon and background rates as well as reduced magnetic bending. The Phase-II muon detector upgrade must recover an effective muon system in the forward region by adding new forward muon detectors, GE1/1 and GE2/1, equipped with Gaseous Electron Multiplier (GEM) detectors and RE3/1 and RE4/1 with improved RPC (iRPC) detectors see Fig. 3.9, covering the pseudorapidity range up to 2.4. The ME0 chambers, also using GEM technology, extend the acceptance of the muon system further to $|\eta| = 2.8$, by providing hits for triggering and for offline muon reconstruction [59].

For RPC detectors, two important upgrades are planned – the RPC off-detector electronics (called link system) will be replaced and the RPC spatial coverage will be extended up to $|\eta| = 2.4$ by installing iRPCs. In addition, there is an ongoing study to find eco-friendly gas(es) to replace the current RPC gas mixture [60].

Based on the above-mentioned requirements, the CERN Engineering (EN) and Physics (PH) Department made a joint project – the Gamma Irradiation Facility (GIF++) [61]. GIF++ is the new CERN irradiation facility located north of the CERN Super Proton Synchrotron (SPS). It is a unique place where high energy (~ 100 GeV) charged particles (mainly muons) are combined with a high flux of gamma radiation (662 keV) produced by 13.9 TBq ^{137}Cs source. ^{137}Cs has been chosen due to its long half-life of 30.08 years compare to ^{60}Co (half-life = 5.2714 years), which makes the facility capable to have smaller decrease of the photon rate over time. GIF++ facility uses secondary particle beam from H4 line of the SPS north area that provides hadrons, electrons or muons. The primary SPS proton beam produces pions and kaons on a production target and the secondary muon beam is mainly generated by the decay of these particles. The spill structure of the muon beam for final users is about 30 seconds with an intensity of about 10^4 per spill for safety purpose[62]. A schematic overview of the GIF++ is shown in Fig. 4.2. The radiation field is uniformly distributed over the xy-plane as required

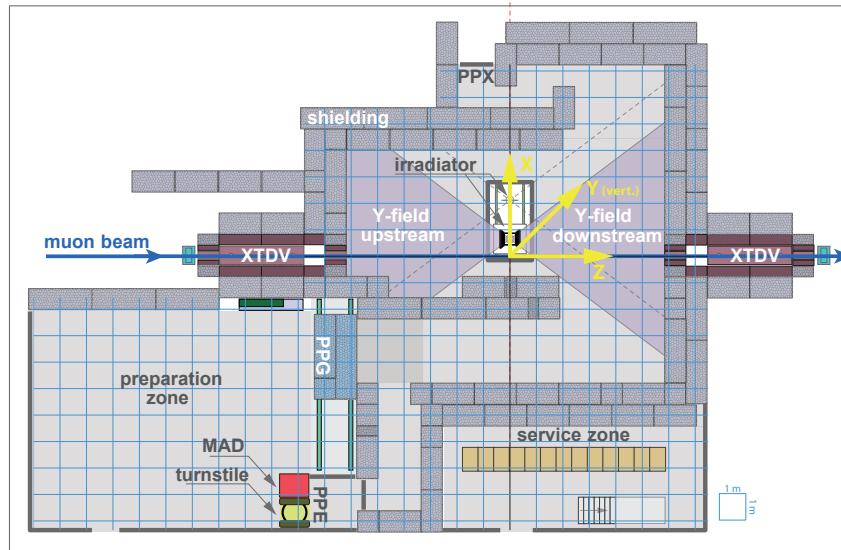


Figure 4.2: An overview of the GIF++ with entrance doors MAD (material access door), PPG (personal protection gate), PPE (personal protection entrance), PPX (personal protection exit).

for large-area flat detectors with the help of two $\pm 37^\circ$ angular correction filters both in the downstream and upstream regions. It is shown in Fig. 4.3a. The photon current is fine-tuned using two complete and independent attenuation systems. It consists of 3×3 convex lead filters and attenuates photons that have energy less than 662 keV to a higher degree. The attenuator system is shown in Fig. 4.3b with three planes (A, B, C) on either side of the source, and each plane further consists of three filters. Each filter possesses the nominal attenuation factors; 1 (A1,B1,C1), 1.5 (B2), 2.2 (C2), 4.6 (C3), 10 (A2), and 100 (A3, B3). A set of 24 nominal attenuation factors (nearly equidistant) have been selected from the 27 combination of 3×3 that varies from 1 to 46415 as shown in Fig. 4.3c.

Simulation of the 662 keV photon current is shown in Fig. 4.4 using an attenuator factor 1 (unattenuated). Using the angular correction filters, the current is uniformly distributed along the y-axis and varies from the source along the z-axis.

Table 4.1 displays the nominal attenuation of 662 keV photons and dose attenuation measured with the

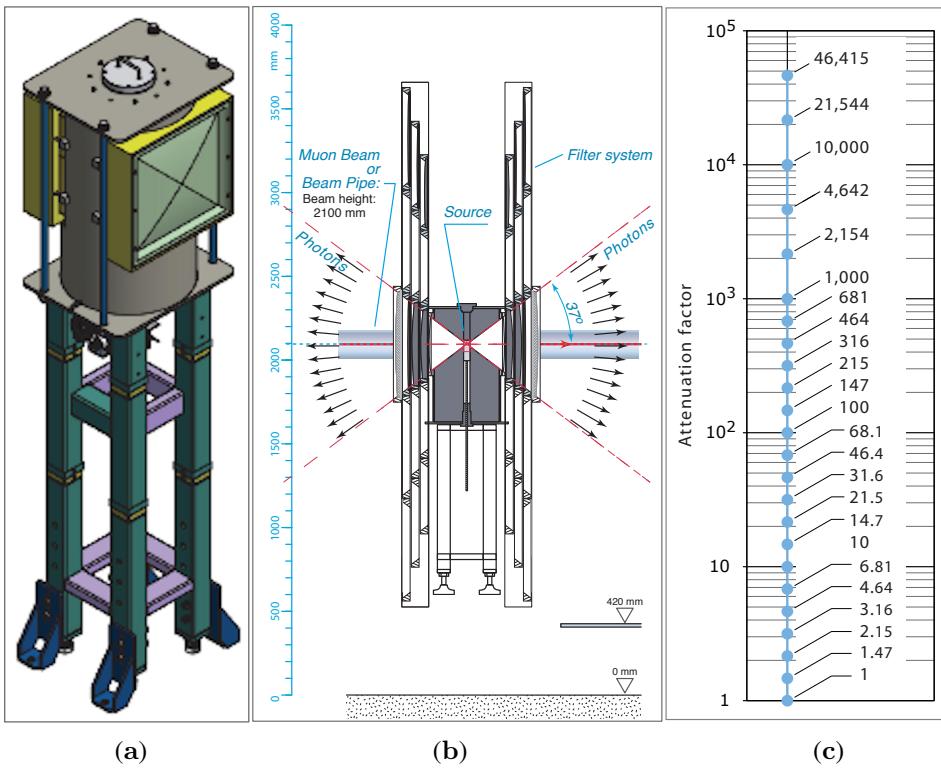


Figure 4.3: Irradiator with the attenuator system in the GIF++. (a): Irradiator with independent angular correction filters on both sides. (b): Irradiator with remotely controlled attenuation filters to vary the radiation intensity. (c): A set of 24 nominal attenuation factors are selected from the 27 combination (3×3) that varies from 1 to 46415.

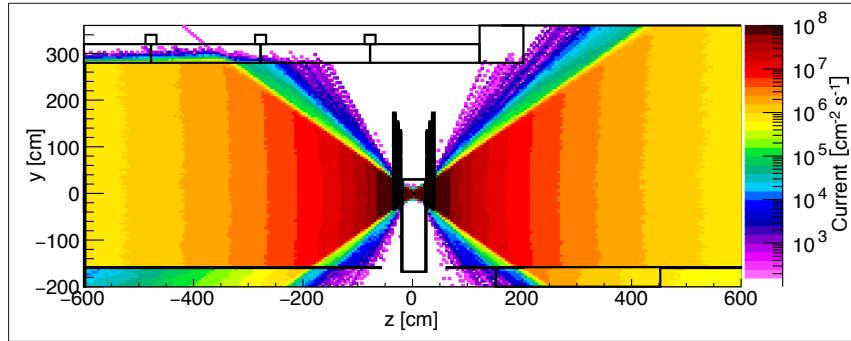


Figure 4.4: Simulation of the unattenuated 662 keV photon current in the yz-plane at $x = 0.65$ m; using the angular correction filter.

Automess gamma probe 6150AD-15 for different filter combinations. For factors less than 10, the nominal and effective dose attenuations are comparable while for greater than 10, the effective dose attenuation is considerably lower than the nominal attenuation factor, since scattered photons with an energy smaller than 662 keV contribute substantially.

At the GIF++, the RPC detectors setup consists of two endcap chambers of type RE2 and RE4 that are continuously irradiated. Two non-irradiated chambers of the same type are installed to be used as reference. They are accompanied by the new generation of Glass-RPC (GRPC) [63] and multi-gap RPCs. A dedicated control system has been built to control these detectors and archive the relevant parameters using the WinCC-OA (PVSS) Supervisory Control And Data Acquisition (SCADA) system [64]. The system controls high as well as low voltage supplies and monitors temperature, pressure, and humidity for both the RPC gas and the environment. The source status and attenuator values are accessed using the data interchange protocol (DIP), which is published centrally by the Engineering Department. The RPC gas supply is controlled and monitored by an external WinCC-OA project that shares relevant parameters with this project. All the relevant parameters are archived in a Structured Query Language (SQL) database (DB) for offline analysis.

One of the features of the GIF++ RPC DCS system is accessing the source status and attenuator values, as

Measured data			
Nominal Attenuation	Filter Combination	Dose Rate [mGy/h]	Dose Attenuation
1	A1 B1 C1	470.00	-
1.5	A1 B2 C1	400.00	1.2
2.2	A1 B1 C2	211.00	2.2
4.6	A1 B1 C3	105.00	4.5
10	A2 B1 C1	55.00	8.8
100	A3 B1 C1	6.50	72.3
100	A1 B3 C1	6.20	75.8
464	A1 B3 C3	1.59	295.6
4642	A2 B3 C3	0.22	2156.0
46415	A3 B3 C3	0.05	9400.0

Table 4.1: Nominal attenuation factors (attenuation of the 662 keV photons) of some filter settings and measured effective attenuation in position D1 ($x=0.65\text{ m}$, $y=0.00\text{ m}$, $z=1.10\text{ m}$)

shown in Fig. 4.3c. Based on this information, the RPC performance parameters (efficiency, working point, cluster size, and resistivity) are measured. To retrieve the data from the database, a specific algorithm has been developed to synchronize the detector parameters (current and voltage) with the external parameters (temperature, pressure, and humidity). It enables the precise monitoring of the effect the external parameters have on the detector.

4.2 WinCC-OA

The SCADA System SIMATIC WinCC-OA [65] is designed by ETM of the Siemens group and used extensively in large industries to supervise and control complex processes. Large experiments at CERN use the commercial ETM SCADA software, WinCC-OA, as a tool to develop control systems. WinCC-OA is consistently built on object-oriented structures, supported by both Windows and Linux systems, is flexible as well as distributed, and has an open architecture. WinCC-OA has the ability to connect hardware (or software) devices under a particular Detector Control System (DCS) and archive their data to observe the behaviour of the device under consideration. WinCC-OA describes a device in terms of a data point in a tree-like structure, with data point elements representing the device parameters. These can then be addressed directly to write to and read from the corresponding device. The WinCC-OA software runs different processes called “Managers”. Event Manager (EV) is the heart of a WinCC-OA system, which connects other specific tasks managers. Figure 4.5 shows these managers and their specific functions.

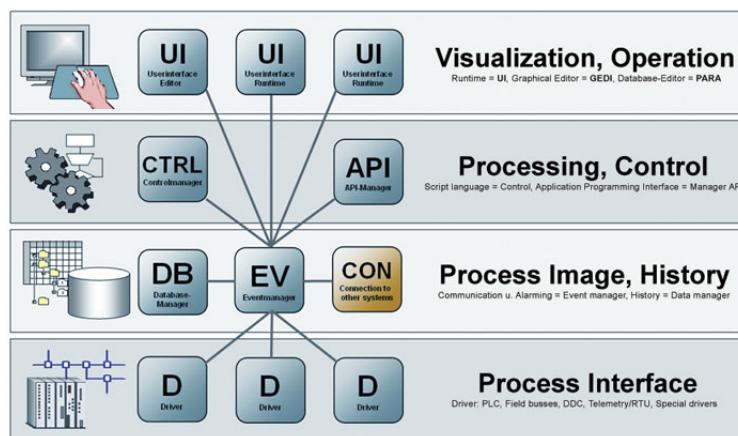


Figure 4.5: The concept of managers in WinCC-OA and it's functional layers. The event manager plays a key role in connecting all managers in a tree-like structure [66].

WinCC-OA is capable of building a distributed system in such a manner that different projects get interconnected and can exchange information remotely via the TCP/IP protocol using a “Distribution” Manager. WinCC-OA uses a number of Managers as “Drivers” to communicate with the Front End (FE) hardware for

data readout by mostly using the industry standard protocols, such as Profibus, CanBus, DIM, and Modbus, for communicating with the Programmable Logic Controllers (PLC) and OPC servers. In order for non-experts to easily and safely operate the system, WinCC-OA provides a user-friendly Graphic User Interface (GUI) panel – an intuitive tool for controlling, monitoring, and operating the detectors in the safe mode. It provides the flexibility to combine text, graphical objects, and synoptic diagrams. The GUI panels can be used to observe the online behaviour of the detector in the form of plots, tables, and histograms. An example of the GUI is shown in Fig. 4.6.

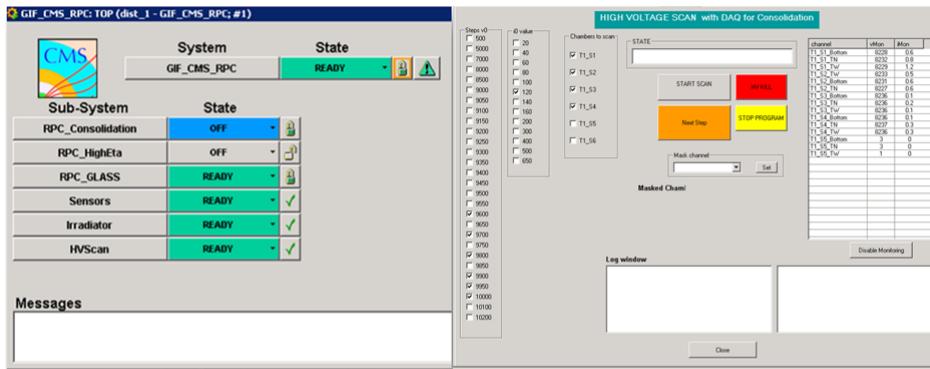


Figure 4.6: FSM main tree and high voltage scan panel using GUI.

All the LHC experiments have common tasks and requirements; therefore, it is necessary to have a general framework to provide all the required standard features and facilities. The Joint Controls Project (JCOP) [67] was developed to reduce the repetition of efforts by reusing common components and concealing the complexity of the underlying tools. The JCOP framework provides extra functions such as standardized Finite State Machine (FSM), the additional Graphical User Interface (GUI), the alarm handlers, and the ORACLE database interface [68]. The JCOP framework provides FSM toolkits in WinCC-OA based on State Machine Interface (SMI++). It offers an easy, robust, and safe way to control the full detector through the definition of a finite number of states, transitions, and actions (ON, OFF, STANDBY, Ramping Up, Ramping Down). A typical device state for an HV channel is implemented through the FSM mechanism, as shown in Fig. 4.7.

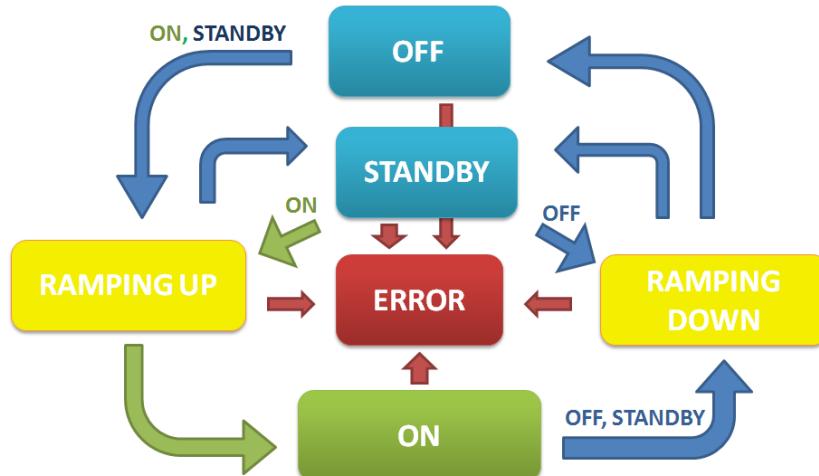


Figure 4.7: The DCS hierarchy tree for a typical high voltage channel using FSM. The tree shows a transition from one state to another.

4.3 The CMS RPC DCS project at GIF++

The CMS RPC DCS at GIF++ has been developed using WinCC-OA 3.11 and extended using the standard JCOP framework. It is designed in a tree-like structure with sub-systems of High Voltage (HV), Low Voltage (LV), environmental, and gas parameters (pressure, temperature, and humidity) as well as radiation levels. Each sub-system is mainly divided into two parts, the Front-End (FE) hardware located around the experimental area

(sensors, power supplies, etc.) and a Back-End (BE) computer network wherein the DCS is running. The HV and LV power supplies used in the RPC GIF++ setup consist of CAEN SY1527 mainframe modules as well as CAEN EASY modules, with additional ADC modules used to read gas and environmental sensors (pressure, temperature, and humidity). The project has access to the hardware registered through Object Linking and Embedding (OLE) for Process Control (OPC) server provided by CAEN using the OPC protocol [69]. The project controls the HV and LV system using the OPC protocol. The environmental and gas sensors (for pressure, temperature, and humidity) are also read out using the OPC protocol. The source status and attenuator values are available centrally via the DIP. The project has been designed to be distributed in order to enable communication with other projects and to read valuable information. Communication is established with the central GIF++ DCS in such a manner that the information from the gas system, such as flow rates, are readable. The FSM hierarchy of the project is based on the naming convention of the trolley, where the detectors are installed. Each trolley has six sections and each section accommodates one detector. Currently, three CMS RPCs trolleys are installed in the GIF++. Trolley 1 (RPC Consolidation) is equipped with spare RPCs, trolley 2 with small glass RPCs, and trolley 3 with prototypes of improved RPCs. Figure 4.13 shows the RPCs configuration in a trolley, and detailed information of the trolleys are provided here [70]. A schematic overview of the DCS project is shown in Fig. 4.8.

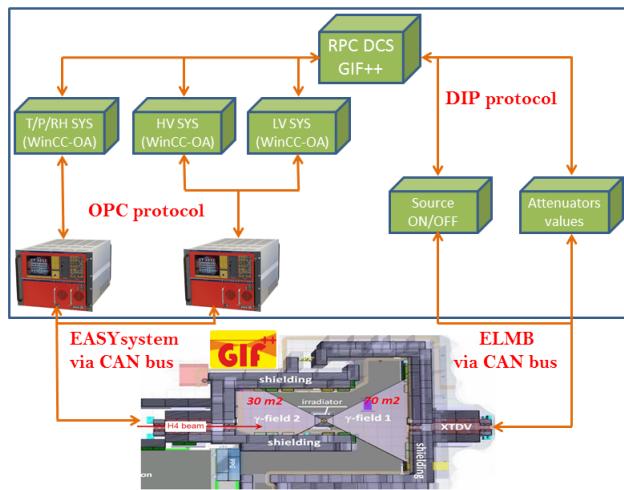


Figure 4.8: An overview of the CMS RPC DCS at the GIF++.

4.3.1 High and low voltage system

The power supply system provided both HV and LV to the installed RPCs in the GIF++. Owing to the simple setup, both the HV and LV boards were installed in the CAEN mainframe that are located outside the radiation zone. The high and low voltage system is controlled and monitored by the CAEN OPC server. Each gap in a chamber is independently connected to a single high-voltage channel, which improves the granularity of control. The RPC front-end electronics requires digital and analogue power supplies [71]. Each low-voltage line has been shared between the two front-end-boards (FEBs) for digital as well as for analogue power supply. The DCS communication with the CAEN mainframe is shown in Fig. 4.9

4.3.2 Environmental and gas parameters

The performance of the RPCs strongly depends on the temperature and pressure of the environment because the RPC's gas density is directly affected by these parameters. Hence, it is important to measure the environmental parameters (temperature, pressure, and humidity) at different locations where the RPCs are installed. The applied voltage is corrected for the environmental temperature and pressure in order to include their effects using Eq. 4.1:

$$HV_{app}(P, T) = HV_o \cdot \frac{P_o}{P} \frac{T}{T_o} \quad (4.1)$$

Where P_o , T_o are the environmental and P , T are the bunker (the radiation zone) pressures and temperatures respectively. HV_o is the user set-up high-voltage and HV_{app} is the corrected high voltage supplied by the

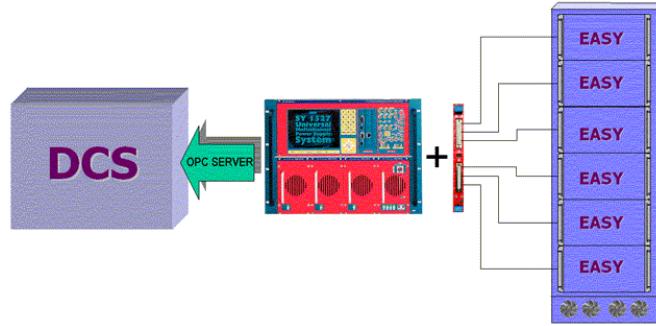


Figure 4.9: The HV and LV channels can be operated by the CAEN mainframe independently, and the DCS is connected to the CAEN via OPC server. An easy crate is used for environmental and gas parameter measurement.

DCS system. This procedure is described in detail in [72]. Figure 4.10a provides a plot for the environmental temperature, pressure, and humidity. The environmental and gas sensors (temperature, pressure and humidity) are readable through the ADC (analog-to-digital converter) board that is installed in the EASY crate. The JCOP framework presents the opportunity to convert the ADC counts into physical values online. The trending feature provides a comparison between different sensors located at different positions.

4.3.3 High voltage scan and stability test

The project has been designed for the R&D of detectors. Hence, it should be able to perform high-voltage scanning or stability tests. For high-voltage scanning, a separate branch has been incorporated in the FSM tree, wherein the user operates each detector independently. A typical UI panel used for HV scanning is shown in Fig. 4.6(right). Using this panel, a user can select different voltage values (left column), a single current value (second left column), and operate detectors independently (third column). The user has access to mask a single HV channel (connected to a specific part of the RPC detector) without making the whole detector. The stability test runs for a long period of time in order to expose the detectors to high radiation. Based on the requirements, a dedicated manager applies the stability script and restarts it automatically. An example of a HV scanning plot for one of the CMS RPC chambers (RE3) at GIFT++ is shown in Fig. 4.10b.

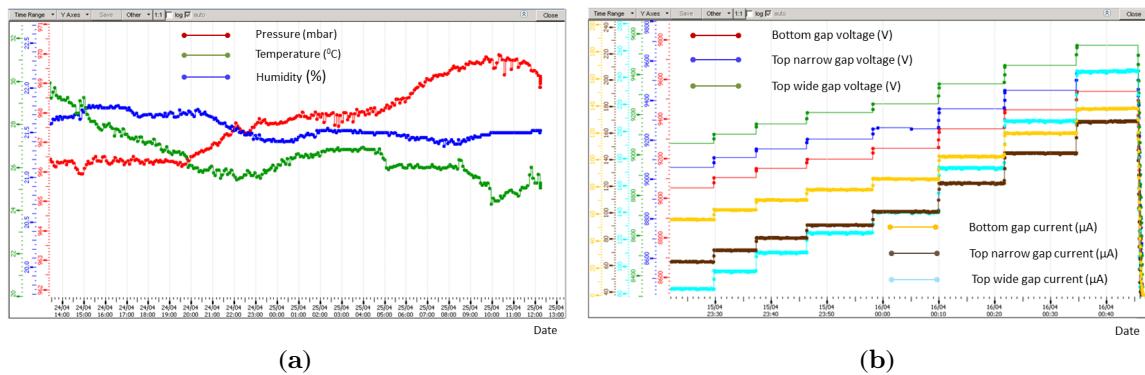


Figure 4.10: (a): Environmental temperature ($^{\circ}\text{C}$), pressure (mb), and humidity (%). Time is on the x-axis while the red, blue, and green lines represent the values of pressure, temperature, and humidity respectively on the y-axis. The pressure and temperature values used for operating voltage correction of the RPCs. (b): A HV scan of the CMS RE3 chamber. The x-axis shows time while the y-axis shows the voltage (V) and current (μA) values. The red, blue, and green lines are voltages while the cyan, brown, and orange lines are the corresponding current values for bottom, top-narrow, and top-wide gaps respectively.

4.3.4 Database

To study the behaviour of the detector over time and to perform an offline data analysis, it is necessary to store all the important parameters in a database. In particular, the HV, LV, environmental and gas parameters, and radiation levels (attenuator values) are kept for this purpose. This information can be utilized to constantly check

the online behaviour of a detector during operation and for the offline analysis to measure the ageing effects after it has absorbed a specific radiation dose. WinCC-OA's uses an internal database for small and simple projects, which stores data locally on the hard drive or on an external oracle database for more complex technical processes that generate large quantities of data. The Relational Database Manager (RDB) is specifically assigned to archiving data in the external oracle database.

Compared to the central CMS RPC system, the RPCs at the GIFT++ generate a small amount of data; hence, we used the internal built-in SQL database in this project. It doesn't need the RDB manager to run. The data point is archived in the database when a change occurs in its value. To suppress the noise fluctuation and reduce the size of data, a "deadband" is specified for each archiving parameter. Since the changes in environmental parameters and HV don't occur at the same time, the values stored in the database are not synchronized. A specific algorithm is applied to synchronize all the relevant stored parameters for analysis. The stored data is finally extracted to be used for the offline analysis using a GUI.

4.4 CMS RPC longevity studies

RPCs are gaseous detectors and in principle can suffer from ageing effects when exposed to a prolonged radiation that deteriorates the detector performance in the form of efficiency loss, dark rate, dark current etc. The deterioration is mainly caused by complicated chemical processes largely occurring in the hot plasma inside electron multiplication avalanches where gas molecules may form polymers growing on electrodes. The severity of ageing increases with integral of radiation exposure and depends on a myriad of factors, such as detector geometry, the materials used for the electrodes, operational gas gain, gas mixture, impurities in the gas itself, gas flow rate etc. For the life span of an experiment, the ageing effects of the installed detectors can be studied by subjecting the same detectors prototypes to accelerated ageing tests performed at a higher instantaneous radiation rate. Assuming the dependency of detector's ageing on the accumulated charge per unit area (in case of RPC), the obtained results from these tests can be projected towards many years of operation at the expected nominal radiation. Since this is approximation and quantitatively not productive, so a large safety margin factor of 3 is used to accumulate the total amount of radiation precisely. The performance of the installed CMS RPC system has already been certified for 10 LHC years in the GIFT facility at maximum background rate of $300 \text{ Hz} \cdot \text{cm}^{-2}$ and a total integrated charge of $50 \text{ mC} \cdot \text{cm}^{-2}$ [73].

The background rates showed linearly dependency on the instantaneous luminosity during the Run-I and Run-II data taking periods both in barrel as well as in endcap regions. Assuming the same linear relationship, the expected rates are extrapolated to the HL-LHC conditions reaching up to $600 \text{ Hz} \cdot \text{cm}^{-2}$ including a safety factor of 3, as shown in Fig. 4.11 for all barrel (left) and endcap (right) chambers. Using the same safety factor, the expected integrated charge in the hottest region of the current RPC system will be $840 \text{ mC} \cdot \text{cm}^{-2}$ at the end of the HL-LHC. The same amount of integrated charge will be accumulated in the GIFT++ facility to certify the detectors for HL-LHC conditions.

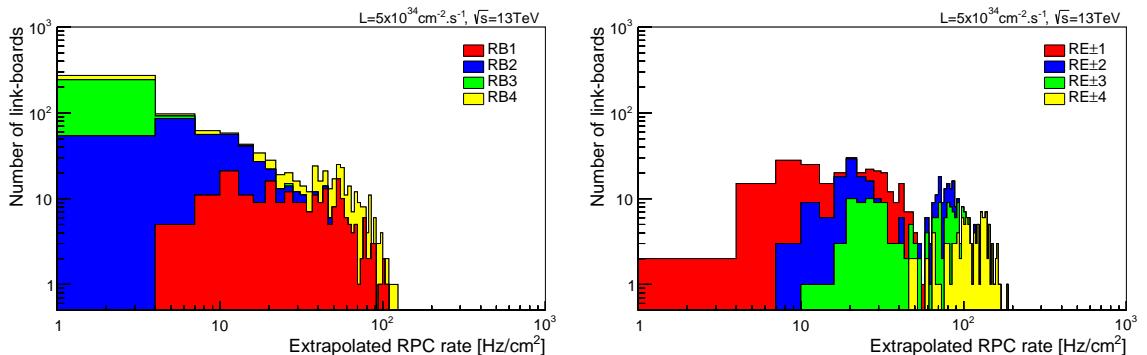


Figure 4.11: Extrapolation from 2016 data of single hit rate per unit area to HL-LHC conditions, in the barrel (left) and endcap (right) regions, for the present RPC system.

The longevity study has already been started in the GIFT++ facility by continuously irradiating the two turned on chambers, RE2/2 and RE4/2. The total integrated charge accumulated by the two chambers are shown in Fig. 4.12. At the end of 2017, the total integrated charge for RE2/2 and RE4/2 was about $292 \text{ mC} \cdot \text{cm}^{-2}$ and $119 \text{ mC} \cdot \text{cm}^{-2}$ respectively which corresponds to 34% and 14% of the expected integrated charge at the HL-LHC. Due to gas flow limitation, the RE4/2 was turned on later.

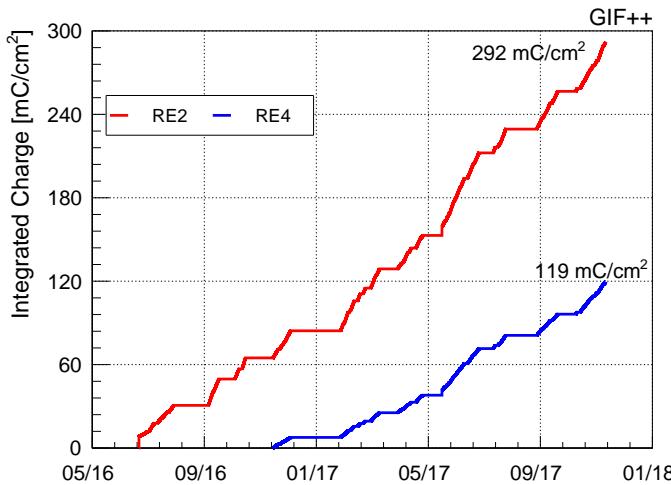


Figure 4.12: Integrated charge versus time accumulated during the Gif++ studies for the RE2/2 (red) and RE4/2 (blue) chambers. Because of total gas flow limitations, the RE4/2 chamber has been turned on few months later. Different slopes account for different irradiation conditions during data taking.

4.4.1 RPCs setup at Gif++ for efficiency studies

Four RPC chambers (two RE2 and two RE4 [71]) are placed parallel to each other in a vertical position, a few meters away from the source in the upstream area as shown in Fig. 4.13. Several high-voltage scans



Figure 4.13: CMS RPCs setup in the Gif++. Detectors are placed perpendicularly and a few meters away from the source.

were performed using different radiation levels (starting from the absence of radiation source) to define the optimal operating voltage of each chamber, which is called Working Point (WP). The efficiency (E) for different background radiation levels is calculated using the equation:

$$E = N_{RPC} / N_{TRACK}. \quad (4.2)$$

The muon track is reconstructed using three reference RPC planes and extrapolated to the RPC under test, examining the closest cluster (a strip or set of continuous strips). N_{TRACK} corresponds to the number of muons passing through the three reference RPC detectors at the same time. N_{RPC} corresponds to the number of fired clusters in the chamber under test. The dependence of the efficiency E on the effective high voltage HV_{eff} [74] can be fitted using the following sigmoidal curve:

$$\langle E \rangle = E_{max} / (1 + \exp(-\lambda(HV_{eff} - HV_{50\%}))), \quad (4.3)$$

where E_{max} is the maximum efficiency reached by the chambers at $HV \rightarrow \infty$, λ is proportional to the slope of sigmoid at flex point, and $HV_{50\%}$ is the high voltage at which a chamber reaches 50% of its maximum efficiency. The Working Point of a chamber is defined by:

$$HV_{wp} = HV_{knee} + 150V, \quad (4.4)$$

where HV_{knee} is the voltage at which the efficiency is 95% of the maximal one.

4.4.2 Efficiency and cluster size results

In Fig. 4.14a, the efficiency as a function of HV_{eff} for different radiation levels is presented. The maximum efficiency slightly decreases with the amount of radiation received by the detector. In Fig. 4.14b, the maximum efficiency as a function of the gamma hits rate is presented for four RPC chambers. The RPCs were placed parallel to each other, with RPC-1 being the closest to the source while RPC-4 was the furthest. The radiation dose depends on the distance between the RPCs and the gamma source. RPCs 3 and 4 received a smaller dose as compared to RPCs 1 and 2.

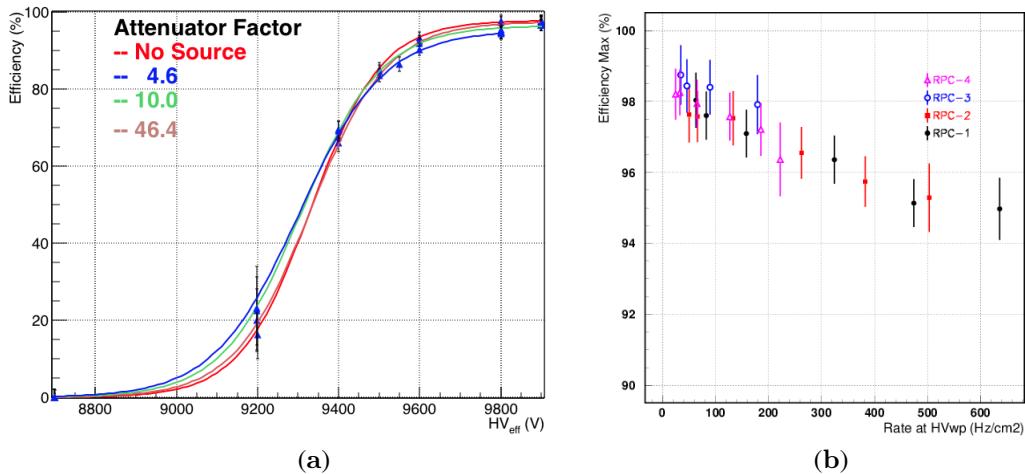


Figure 4.14: (a): Efficiency vs HV_{eff} for different gamma attenuator factors. (b): Maximum efficiency vs gamma rate at HV_{wp} for four RPCs.

Figure 4.14 represents the very first results obtained in the GIFT++ at the start of 2016. Afterwards, several measurements have been performed under different background radiation conditions using the dedicated attenuator system. Each time the detector performance has been studied. Figure 4.15 shows the hit efficiency of the RE2/2 chamber as a function of the effective HV at different irradiation stages, corresponding to an integrated charge of 0, 153 mC·cm⁻² and 257 mC·cm⁻². The background radiation rate is minimum (source-off) in the left plot and maximum (about 600 Hz·cm⁻²) in the right plot. The detector showed stable behaviour over time and no change in the efficiency and working point has been observed.

A similar study has been done to monitor the cluster size, defined as the number of fired strips per hit. Figure 4.16 shows the cluster size as a function of effective HV for RE2/2 chamber with integrated charge: 153 mC·cm⁻² (18%) and 257 mC·cm⁻² (31%) without background radiations (left) and background gamma irradiation of about 600 Hz·cm⁻² (right). No significant change has been observed in the cluster size.

4.5 Summary

The DCS project for the CMS RPCs has successfully been implemented and tested in the CERN GIFT++. Since June 2015, the project has been running in a stable manner. The detectors are being operated and the data being archived. The hardware integrated in the project fully controls the high-voltage scanning and stability tests. Environmental and gas sensors are included and used for temperature and pressure corrections. Gas flow-meters are read through the central DCS at GIFT++, and the data is used to study the behaviour of different gas mixtures. All useful parameters are archived in the internal database for offline analysis. As the project is designed for detector R&D studies, any new hardware can be added easily and safely.

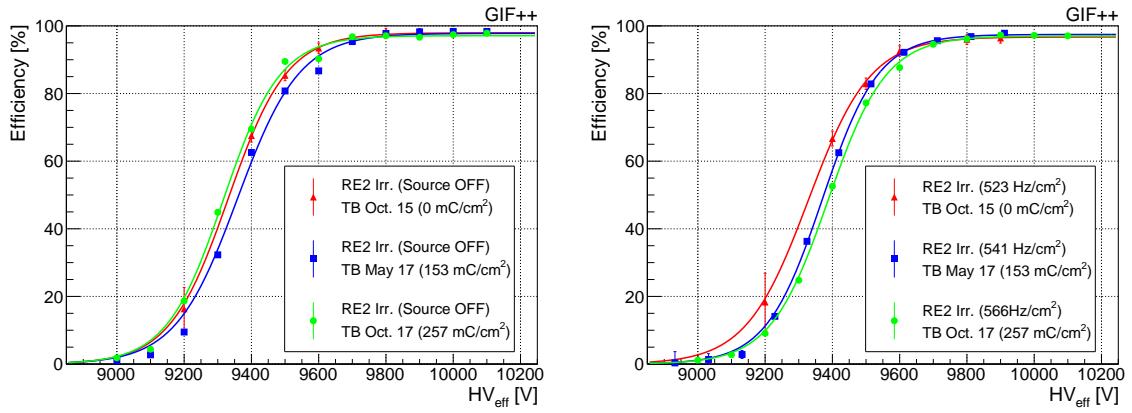


Figure 4.15: Hit efficiency of RE2/2 as a function of the effective HV, without irradiation (left) and under a gamma background rate of about $600 \text{ Hz}\cdot\text{cm}^{-2}$ (right). The measured efficiency of the RE2/2 chamber corresponds to different Test Beams (TB) with integrated charge: 0, $153 \text{ mC}\cdot\text{cm}^{-2}$ (18%) and $257 \text{ mC}\cdot\text{cm}^{-2}$ (31%). The detector performance is stable at high fraction of accumulated charge.

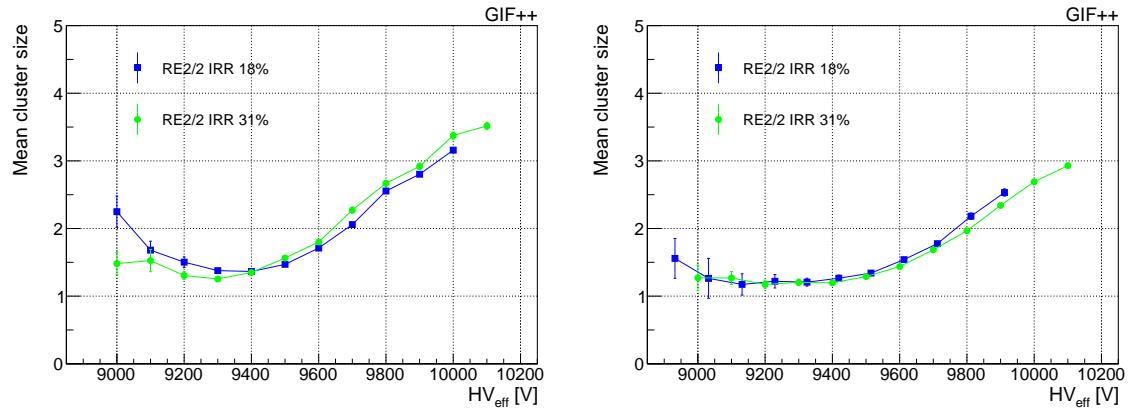


Figure 4.16: Cluster size of RE2/2 as a function of the effective HV, without irradiation (left) and under a gamma background rate of about $600 \text{ Hz}\cdot\text{cm}^{-2}$ (right). The measured cluster size of the RE2/2 chamber corresponds to different Test Beams (TB) with integrated charge: $153 \text{ mC}\cdot\text{cm}^{-2}$ (18%) and $257 \text{ mC}\cdot\text{cm}^{-2}$ (31%). The cluster size is above 2 at working point of the detector with high fraction of accumulated charge.

The performance of the CMS RPC chambers at GIF++ has been studied and compared at different radiation levels. At a rate of $600 \text{ Hz}\cdot\text{cm}^{-2}$, the Eff_{max} of the chamber was 95%. The detectors showed stable performance up to 34% of the expected integrated charge that will be accumulated at the end of the HL-LHC run.

5

Monte Carlo Simulation and Event Reconstruction

This chapter describes the Monte Carlo (MC) simulation methods used in the field of high-energy physics for signal and background process generation. It explains in detail the particle physics generators that are used in the analysis included in this thesis. Event and physics objects reconstruction in the CMS detector are described in the last two sections.

5.1 Monte Carlo simulation

In high-energy physics experiments, event modelling plays an important role to understand the collected data. A precisely modelled event maximizes the chance of discovering new physics and making precision measurements of the Standard Model (SM) processes. However, in hadron-hadron colliders where the colliding particles are composite objects, like proton-proton in LHC, event modelling is a challenging task. The same applies to particle interactions within the bulk of the detector volume. These tasks can be solved by employing Monte Carlo generation techniques, and incorporating the SM models of new physics as well as the detector effects while detecting the final-state particles in the interaction. An event occurs when two protons collide and produce a cascade of new particles as shown in Fig. 5.1. Because of the complexity of the event, the MC generators model the events in a simulation chain: Hard interaction, parton showering, hadronization, and decays. These processes are explained in the following sections.

5.1.1 Parton distribution functions

In hadron colliders such as the LHC, the colliding particles (protons) have a composite structure consisting of gluons and quarks. In this analysis, the SM $t\bar{t}$ and the heavy Higgs (H, A) states are produced by the gluon fusion. Therefore, it is necessary to understand the internal structure of proton, described by the *Parton Distributing Functions (PDFs)*. The parton density function $f_i(x, Q^2)$ gives the probability of finding a parton (quarks or gluons) of flavour i in the proton, carrying a fraction x of the proton momentum with Q being the energy scale of the hard interaction [76]. An accurate determination of PDFs, and their corresponding uncertainties is obtained from global fits to a variety of data from multiple experiments, such as HERA, Tevatron, and LHC, using the DGLAP evolution equation. These results show that for small values of Q^2 , a large fraction of the hadron momentum is carried by its valence quarks; whereas at high energies, most of the hadron momentum is carried by the so-called sea partons, i.e. gluons and virtual $q\bar{q}$ pairs.

This analysis uses the “NNPDF30” sets provided by the “NNPDF” collaboration [77]. “NNPDF” uses an unbiased modelling tool, Neural Networks, with trained genetic algorithms to construct a Monte Carlo representation of PDFs and their uncertainties – a probability distribution in a space of functions. As a final user, a

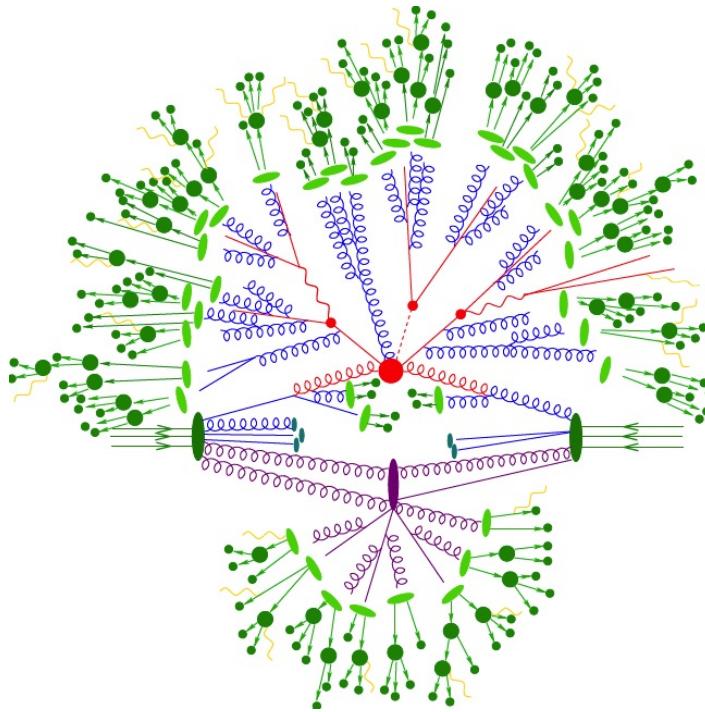


Figure 5.1: The pictorial representation of a pp collision event [75]. The hard interaction is represented by the big red blob. Additional hard QCD radiation is produced (red) and a secondary interaction takes place (purple blob) before the final-state partons hadronise (light-green blobs) and hadrons decay (dark green blobs). Photon radiation occurs at any stage (yellow)

set of 100 MC replica has been used to compute PDF-dependent quantities along with their uncertainties. The central value represent the “best fit” PDF, while the error members represent variations around the best result. The PDF uncertainties are discussed in detail in Sec. 6.7.2.

5.1.2 Hard scattering

The actual interaction of two protons occurs when partons from the two colliding protons interact and produce new particles with high p_T . These events are of keen interest for analysis and are referred to as hard scattering. If the colliding protons merely undergo soft collision, the resultant particles have low p_T , commonly known as soft scattering. PDFs describe the structure of the proton and contain the initial momentum distribution of the partons involved in the hard interaction. In hadron-hadron (in case of LHC, it is pp) collisions, a wide variety of hard scattering cross sections can be calculated using the QCD factorization theorem and weighting the subprocess cross section with the PDFs extracted from deep inelastic scattering [78]. Factorization theorems separate long- and short-distance physics when calculating these cross sections.

$$\sigma_{AB} = \int dx_a dx_b f_{a/A}(x_a, \alpha_s, \mu_F) \cdot f_{b/B}(x_b, \alpha_s, \mu_F) \cdot \hat{\sigma}(\hat{s}; \alpha_s, \mu_F, \mu_R) \quad (5.1)$$

Where, $f_{a/A}$ is the probability that a parton a inside a hadron A carries a momentum fraction x_a , \hat{s} is the parton centre-of-mass energy, α_s is the strong coupling constant, μ_F is the factorization, and μ_R is the renormalization scale.

5.1.3 Particle physics generators

In the current High Energy Physics (HEP) regime, the most challenging aspect is the multiparticle production, where the observed particle multiplicities extend to hundred. These multiplicities are expected to go upward in future HEP colliders. Event Generators (EG), based on computer programs, solve this problem by generating events as detailed as could be observed by a perfect detector. In High Energy Physics, a number of Monte Carlo (MC) generators are available to calculate the tree-level diagrams numerically and integrate them over the relevant phase space. A list of them are used in Monte Carlo production for this analysis is discussed below.

- **MADGRAPH:** now merged into **MG_AMC@NLO**, a general-purpose event generator [79] used for the generation of signal samples in this analysis. It is a pure-matrix element generator that generates hadron-hadron ($pp, pp\bar{p}, gg, q\bar{q}$) collision and in practice, produces eight particles (quarks, gluons, and leptons) in the final state without hadronization. It also provides an opportunity to produce samples using the interference effect. The interference comes into play when the signal and background have the same final state particles; in this analysis, the signal $gg \rightarrow H/A \rightarrow t\bar{t}$ and the SM $t\bar{t}$, the main background, have same final state particles. MADGRAPH is further interfaced with PYTHIA or HERWIG to generate showering and hadronization steps. The double counting in the showering is solved by applying the MLM scheme. To describe the parton structure, “NNPDF30” PDF [80] sets are used in this analysis.
- **PYTHIA:** is a multipurpose event generator used frequently in HEP. It provides the possibility to generate complete events in as much detail as experimentally observable within the bounds of our current understanding of the underlying physics. PYTHIA is used to model e^+e^- , ep and pp collisions and simulate a large variety (over 300) of hard $2 \rightarrow 2$ processes that include SM and many BSM processes up to Leading Order (LO) accuracy. PYTHIA uses the Lund string model [81] to describe hadronization and external PDFs for hard processes calculations. It uses a p_T -ordered showering technique for parton showering, where partons are ordered by their transverse momentum (p_T) and $Q^2 = p_T^2$ – the closer the parton is to the vicinity of the hard process, the higher p_T is assigned to it. More detail about the PYTHIA program is given in [82].
- **POWHEG:** PositiveWeight Hardest Emission Generator is an event generator [83] that provides modelling of the hardest interaction at NLO QCD accuracy and needs to be interfaced with PYTHIA or HERWIG for the parton showering and hadronization. The proton structure is described using the PDF sets NNPDF30 for the MC samples generated by POWHEG.
- **MC@NLO:** is an event generator [84] that generates hard emissions using NLO fixed-order QCD calculations with the (N)LL accuracy shower algorithm implemented in HERWIG. For showering hadronization and decays, the MC@NLO has to be interfaced with a general-purpose detector such as HERWIG or PYTHIA. The event generation strategy of the MC@NLO is a two-step process. In the first step (NLO), the hard-scattering events (un-weighted) are generated with NLO QCD accuracy. In the second step (MC), the resultant events are passed to HERWIG for further processing (parton-shower, hadronization, decays) and the final events are weighted. A specific fraction of events (20–30%) with negative weights are assigned in the matching procedure. The negative weights cancel the positive, leaving 40–60% positive weighted events. To cope with this situation, a large event sample needs to be generated. The PDF sets used by MC@NLO are NNPDF30.
- **HERWIG:** is a general-purpose event generator [85], which models hadron-hadron, lepton-lepton, and hadron-lepton collisions at LO. Like PYTHIA, it provides the description of all subprocesses of an event but uses different approaches and algorithms. The HERWIG tool implements an angular order showering, $Q^2 \sim 1^\circ \cos \theta$, where θ is the angle between the parent and emitted parton. HERWIG exploits the cluster model for showering.

5.1.4 Parton showering

The previous section illustrates the generation of a hard process according to lowest-order matrix elements, which results in a limited number of partons in the final state. These describe the momenta of the outgoing jets well, but any fixed order is insufficient to provide a complete picture of the overall process, including the internal structure of the jets and the distribution of the accompanying particles. The effect of all higher-order corrections, including additional ISR/FSR from the branching of the partons, can be simulated through the parton-shower (PS) algorithm. It describes the evolution in momentum transfer from the high-energy scales associated with the hard process down to the low scales of order 1 GeV associated with the confinement of the partons it describes into hadrons [86]. Multi-purpose event generators, like PYTHIA 5.1.3 and HERWIG 5.1.3, use parton showering algorithms with different implementations. Hard processes can be described well using matrix element calculations, where the partons are energetic and widely separated. It also incorporates the interference effects of amplitudes with the same final state topology. In this work, we produce the interference effect between the signal ($gg \rightarrow H/A \rightarrow t\bar{t}$) and SM $t\bar{t}$ ($gg \rightarrow t\bar{t}$) background using the MADGRAPH generator defined in Sec. 5.1.3. The matrix element and parton shower algorithms can be combined to fully describe an event with extra care to avoid double counting in the overlapping phase-space. Different schemes are used for

this purpose. CMS uses the MLM algorithm for matching interfaces MADGRAPH with PYTHIA, which vetoes the emission of partons by showering above a user-defined matching threshold [87].

5.1.5 Hadronization

The hadronization process starts after showering, where a set of coloured partons is transformed into a set of colour-singlet primary hadrons, which may decay further to secondary hadrons. Different models are used to describe the fragmentation of the partons after showering. The most commonly used one is the *Lund String Fragmentation Model*, implemented in PYTHIA and proposed to be universal, i.e. process independent. It is based on the observation that the colour potential of the sources, such as a heavy quark–antiquark pair, increases linearly with their separation [88]. The potential energy increases with the separation of quark–antiquark, and at an order of 1 fm, it collapses into colour-field strings between them. The original quark pair is now converted into two pairs of quark and antiquark $q\bar{q} \rightarrow q\bar{q}' + q'\bar{q}$; this process continues until only the hadrons remain [89]. Apart from the hard interaction, other constituents of the colliding proton can also interact, adding additional hadrons in the final state. This is usually referred to as the underlying event (UE) described by special tunes in the generators, such as PYTHIA uses CUETP8M2T4 in this analysis. During one bunch crossing, the average pp interaction goes up to 35, known as the pile-up, resulting in a relatively low number of p_T particles, which can, however, obscure the interesting hard process.

5.1.6 Event simulation

The newly generated particles from the event generator are passed through a chain of processes to simulate the detector’s material and magnetic field effects. CMS uses two types of MC event simulations – fast simulation (“FastSim”) and full simulation (“FullSim”) – based on the analysis requirement. The “FastSim” method reduces the CPU overhead time and is much faster as compared to “FullSim”. However, it uses a parametric approach to simulate and reconstruct events with the CMS detector and can be used for conducting specific analyses. The alternative “FullSim” approach is rather time consuming but more accurate and is based on GEANT4-based simulation [90]. The analysis included in this thesis benefits from the “FullSim” approach. GEANT4 is a toolkit that provides a comprehensive set of physics processes to model the interactions of particles with the detector materials, the resulting energy loss, and the detector’s electronic response. The algorithm incorporates information concerning the material budget, strength of the magnetic field, and the precise geometry of the CMS detector.

Pile-up events are also added at this stage, which are mostly soft QCD processes. They are simulated as minimum-bias beforehand in the form of a library and then used to overlay onto the signal event according to a specified pile-up scenario. “Out-of-time” pile-up information is also taken into account, where many bunch crossings take place before and after the central collision event.

Digitization is the next step in which information from the previous steps are converted into electronic signals including electronic noise. The L1 and HLT information is also included in the simulated MC samples. The simulated sample’s format at this level is similar to that of the real collision data and can pass the same reconstruction steps as the real collision data explained in Sec. 5.2.

5.2 Event reconstruction

The CMS event reconstruction procedure is interpreting the detector data as a set of physical objects, electrons, muons, photons, charged hadrons, and neutral hadrons. In this analysis, the final state objects of the $t\bar{t}$ system are generally reconstructed using the Particle Flow (PF) algorithm [91]. This provides a fully consistent picture of the event by reconstructing the particles and jets by taking information from all subdetectors of the CMS. It iterates the event to reconstruct particles and jets, starting from identifying the muons, as they have the most unambiguous particles. After identification, the muon signal is blinded and the charged hadrons are reconstructed. The electron is reconstructed in the next step while the remaining signals in the ECAL are assigned to the photons and the signals from HCAL to the neutral hadrons. Once all the particles and jets have been reconstructed in the event, the missing transverse energy (E_T^{miss}) is reconstructed on the basis of all the available information. PF algorithm is applied to data and simulation in the same manner, which generally leads to a good agreement between the data and MC. Sometimes, the quantities are not well modelled in simulation

and are further corrected by applying event reweighting techniques to simulation only without hiding the new physics.

5.2.1 Track and vertex reconstruction

Combinatorial Track Finder (CTF) [92] is used in the CMS experiment as a tracking algorithm – an extension of the Kalman filter [93]. From the inner tracking detectors, the neighbouring pixels and strips that produce signals are grouped to clusters, which provides the estimate hit of a passing particle from the detector material. A sequence of hits can be used for fitting in order to find the corresponding trajectory of the passing particle. A charged particle is bent inside the magnetic field, producing a curvature that determines the momentum of the particle. Four main steps are involved in track reconstruction.

- **Seed generation:** provides initial track candidates using 2–3 tracker hits.
- **Track finding:** the seed trajectory is extrapolated to the outer layers of the tracker along the expected flight paths, and further hits that are compatible with the original track are found. On each consecutive layer, all hits from a 3σ region around the seed trajectory are tried out and fitted with a Kalman filter.
- **Track fitting:** the final collection of hits found in all the tracker layers with the track findings; a track candidate is refitted using the Kalman filter.
- **Track selection:** checks if the track candidates satisfy a set of track-quality requirements.

The tracking procedure is repeated six times. On the first iteration, high p_T of the tracks and the smallest impact parameters are required. The tracks reconstructed in the first iteration are blinded, and the second iteration starts with slightly softer requirements on the seed tracks.

5.2.2 Primary vertex reconstruction

The primary vertex is the location of a proton-proton interaction in the CMS geometrical centre, which is determined separately for each event. The primary vertex reconstruction exploits information from the beam spot, which is the 3D-luminous region inside the CMS detector at the collision point. Its position is considered to be the average of collision points from many events. The primary vertex reconstruction procedure follows three steps [92]:

- **Track selection:** select tracks that are consistent with being produced in the primary interaction region, having hits in at least two-pixel layers and at least five pixel and strip layers associated with the track and having a χ^2 per degree of freedom for the track fit not higher than 20.
- **Track clustering:** the clustering algorithm prohibits the tracks from splitting into two and emerging from a single vertex while simultaneously not allowing tracks from different vertices to merge into a single track. The tracks are clustered according to their z-coordinate at the point of closest approach to the beam spot centre.
- **Fitting the vertex position:** tracks selected in the previous two steps are used for the fitting using the adaptive vertex filter [94].

The information from all primary vertices in the event is useful for the reconstruction of other objects as well as for separating the signal vertex with a hard interaction from one with a soft interaction. On average, 20 primary vertices can be reconstructed in one event, called a pile-up , corresponding to 20 pp interactions.

5.3 Physics objects reconstruction

The work presented in this thesis exploits the final state of the $t\bar{t}$ system that consists of leptons (electron, muon), jets, and b-jets originating from b-quarks and missing transverse energy. Their reconstructions in the CMS detector are described below.

5.3.1 Muon reconstruction

Compared to other physics objects that mostly stop in the calorimeters or magnet bulk, muons can be identified unambiguously as they travel through the entire detector. A muon can be reconstructed by combining the information from the inner tracking system with that of the muon spectrometers [95]. Muon reconstruction approaches used in CMS analyses follow three methods.

- **Standalone muon reconstruction:** relies only on the muon system; the Kalman filter (KF) fit is performed starting from the track segments in the innermost muon chambers. It is mainly used for the cosmic muon reconstruction because of the large volume of the muon spectrometer. For pp collision, the tracker information is added to increase the precision.
- **Global muon reconstruction:** muon tracks reconstructed from the muon system only (standalone muon) are fitted together with compatible tracks from the inner tracker using the Kalman filter technique.
- **Tracker muon reconstruction:** a tracker muon corresponds to a tracker track extrapolated to the muon detector region that is compatible with the position of at least one segment in the muon chambers. At a lower momentum, $p_T < 5 \text{ GeV}$, the tracker muon reconstruction is more efficient; at higher energies, the global muon shows more efficiency. In Fig. 5.2, a schematic view of the transverse plane of the CMS detector is shown with all detectable particles.

The work presented in this thesis uses a global muon reconstructed with the help of the Particle Flow (PF) algorithm. It also reconstructs a non-isolated muon within the jet cone, which is further vetoed during the final event selection. The detailed selection criteria for a tight muon and to veto a loose muon in an event will be described in Sec. 6.3.

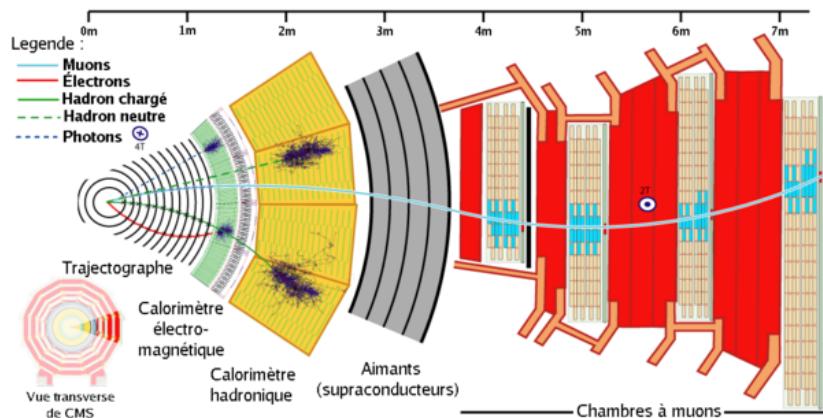


Figure 5.2: A sketch of the particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector.

5.3.2 Electron reconstruction

Electron reconstruction [96] begins with the clustering of ECAL energy deposits. In the absence of material interactions in the beam pipe or tracker, approximately 94% of the incident energy of a single electron is contained in 3×3 crystals and 97% in 5×5 crystals. Due to the strong magnetic field and the electrons undergoing Bremsstahlung, the energy deposited in the ECAL is spread in ϕ . This energy is clustered by building a group of clusters, a supercluster (SC), which is extended in ϕ . CMS employs a hybrid algorithm in the EB and an island algorithm in the EE.

Non-overlapping clusters are grouped into an SC. The procedure is seeded by searching for the most energetic cluster (seed cluster) and then by collecting other clusters in a fixed search area around the seed position. The clusters belonging to radiation from a single electron are aligned in η but spread in ϕ . By collecting all clusters in a narrow η window, whose size is dictated by the η resolution of the detector, it is possible to recover most of the radiated energy. The energy of the SC is corrected based on the number of crystals present in the seed cluster in order to remove any residual η dependence. The position of the shower is obtained by calculating the energy-weighted mean position of the crystals in the SC [2]. To complete the process of electron reconstruction, the SC needs to be associated with a track in the inner tracker. Electron tracking begins with the

formation of a pixel seed, which involves finding a pair of hits in the inner tracker consistent with the trajectory of the electron. The pixel seed itself is a vector located at the outer hit position, pointing in the direction of the electron’s trajectory, and serves as the starting point for tracking. The standard seed-finding process is referred to as pixel-matching since the hit pair is usually located in the pixel layers.

However, a major difficulty of electron reconstruction is that electrons can undergo bremsstrahlung in the tracker material. The radiation affects both energy and momentum measurements, and this effect depends on the material thickness. To account for bremsstrahlung losses, CMS employs a Gaussian-Sum Filter (GSF) track fit. This fit uses the Bethe-Heitler model of electron energy loss and approximates the energy loss distribution as a sum of Gaussian distributions. Different Gaussian models have different degrees of hardness of the bremsstrahlung in the layer under consideration. The GSF fit allows for good momentum resolution at the vertex while also providing a meaningful estimate of the momentum at the outermost part of the tracker.

Largely, there are four types of electron candidates – prompt, non-prompt, conversion, and fake. Prompt electrons are mainly formed by the decay of W and Z bosons. Non-prompt electrons arise from b or c quarks decaying to an electron. Although these electrons are usually not isolated within the quark jet, since there is a significant amount of nearby electromagnetic and/or hadronic activity, the kick from the quark decay might knock the electron out of the jet, enough for it to appear isolated. Conversion electrons come from a photon producing an electron-positron pair in the tracker. Fake electrons are a result of reconstruction error – a coincidence of a jet depositing a large amount of energy in the ECAL and a nearby (matched) single; high- p_T track is misinterpreted as an electron.

The electron used in this analysis is a PF candidate, passes trigger selection as explained in Sec. 6.2.4, and uses cut-based identification (ID) recommended for 2016 data samples with different selection criteria for barrel and endcap regions [97]. The selection criterion further uses tight and loose cuts according to the analysis requirements, as explained in Sec. 6.3.

5.3.3 Jet reconstruction

Jets containing cascades of hadrons, electrons, and photons are clustered to determine the original quark or gluon characteristics. The CMS particle-flow jet identification criteria [98] are used for the jets. This approach uses a jet-clustering algorithm to identify and cluster jets. In the CMS experiment, jets are clustered using anti- k_T clustering algorithm [99]. Particles with energy above a certain threshold are reconstructed using the PF algorithm within a jet cone size " ΔR " of 0.5. All PF jets below 10 GeV are considered to represent uncluttered energy. Jet energy corrections that include offset (L1FastJet with an active area calculation), relative (L2), and absolute (L3) are applied. The purpose of the offset term is to correct for the pile-up. The relative corrections smooth out any η dependence, and the absolute corrections relate to the overall energy scale. Additionally, jets in data have residual corrections applied to them in order to account for data-Monte Carlo simulation discrepancies. Jets are selected with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.4$ and are further required to satisfy loose quality criteria that suppresses noise and spurious energy deposits:

- at least two particles (with at least one being charged in a given jet);
- energy fraction of neutral hadrons < 0.99 ;
- contribution of both charged and neutral electromagnetic energy fractions < 0.99 .

Typical jet energy fractions carried by charged particles, photons, and neutral hadrons are 65%, 25%, and 10%, respectively. This means that 90% of the jet energy can be reconstructed quite precisely, both in magnitude and direction, by the PF algorithm; the remaining 10% fraction of energy of neutral hadrons is affected by the poor HCAL resolution and by calibration corrections of about 10 to 20% (a source of uncertainty that ECAL does not suffer from). Consequently, jets made of reconstructed particles are much closer to jets made of generated particles than jets reconstructed using only calorimeter information with respect to energy, direction, and content.

5.3.4 Heavy flavour jet identification

The accurate identification of heavy flavour jets is important for many searches concerning the LHC (top quark, Higgs boson, and SUSY), which includes light and heavy jets as final state objects. This thesis is based on the $t\bar{t}$ semileptonic final state search that contains at least four jets with two heavy flavour bjets and two light jets. bjets arise from b quark radiation and hadronization and are identified with the help of different algorithms,

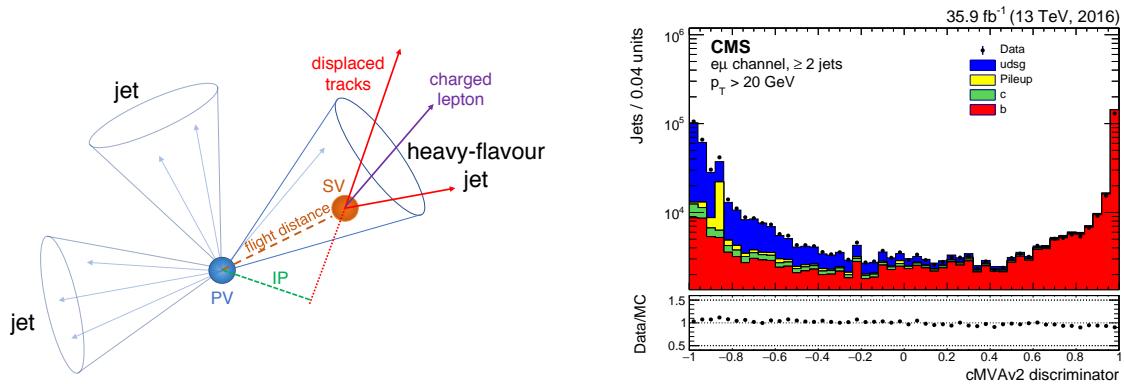


Figure 5.3: Left: A heavy-flavour jet (b/c) decays from a secondary vertex (SV) resulting in charged-particle tracks (including possibly a soft lepton) with a large impact parameter (IP) value. Right: The cMVAv2 distribution in data and MC.

commonly known as b-tagging. B-tagging algorithms use variables connected to the properties of heavy-flavour hadrons, e.g. the lifetime of hadrons containing b quarks is more than that of those with c quarks; hence, it decays after covering a few mm to 1 cm distance, making a displaced secondary vertex (SV). The left Fig. 5.3 shows the displaced SV originated from the decay of a heavy flavour jet. b-tagging algorithms exploit the impact parameter (IP) that characterizes the distance between the primary vertex and the displaced tracks at their points of closest approach. The algorithms further benefit from other measurable quantities, such as the masses of heavy hadrons and the presence of charged leptons in their decay.

During Run-I, CMS had been using the jet probability (JP) and combined secondary vertex (CSV) taggers; in Run-II, the CSV is optimized to CSVv2 and another version (DeepCSV) is introduced [100]. A new tagger cMVA is introduced in Run-II, based on the combined multivariate analysis, that combines the discriminator values of various taggers. We use cMVAv2 tagger at a medium value ($c\text{MVAv2} > 0.4432$) with a b-tagging efficiency of about 70%. The right plot in Fig. 5.3 is the cMVAv2 distribution in data and MC. The medium value ($c\text{MVAv2} > 0.4432$) clearly selects the region (red color) with high b -jets efficiency. A detailed description of the cMVA application in this analysis is given in Sec. 6.2.3.

5.3.5 Missing transverse energy

The E_T^{miss} can be defined as the imbalance in the transverse momentum of all particles that interact with detectors in an event. Owing to the momentum conservation, E_T^{miss} corresponds to the transverse momentum that is carried by weakly interacting particles, such as neutrinos. The E_T^{miss} is computed as the negative of the vectorial sum of transverse momenta of all PF particles and is also referred to as PF E_T^{miss} . This quantity is used in a majority of CMS analyses because of its high performance. Minimum energy thresholds in the calorimeters, inefficiencies in the tracker, and nonlinearity of the response of the calorimeter for hadronic particles could lead to overestimated or underestimated values of E_T^{miss} .

5.4 Signal modelling

Assuming that the sought-after heavy Higgs boson respects the usual hierarchy of couplings to fermions, its dominant production mode is the gluon fusion, as shown in Fig. 5.4 (left). Only the contribution with the top quark running in the loop is considered, while the subleading term with the bottom quark is neglected as $m_b \ll m_t$. The neglected contribution can become significant in models that enhance the coupling to bottom quarks and suppress the coupling to top quarks, such as type-II 2HDM in case of a large $\tan\beta$ value. Such models are not addressed in this study as they would also decrease the $\Phi \rightarrow t\bar{t}$ branching ratio.

The squared amplitude corresponding to the signal diagram gives rise to a resonant excess above the SM $gg \rightarrow t\bar{t}$ production, an example diagram for which is given in Fig. 5.4 (right). As with other BSM $t\bar{t}$ resonances, the excess has an approximately Breit–Wigner $m_{t\bar{t}}$ spectrum. Since the signal and the SM processes share the same initial and final states, there is also a contribution from the interference between the two. The background amplitude is much larger in value than the signal value; because of this, the overall BSM signature

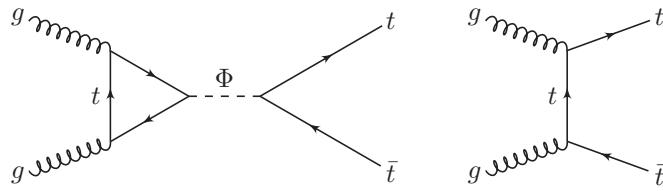


Figure 5.4: Feynman diagram for the signal process (left) and an example diagram for the SM $t\bar{t}$ production (right).

can be dominated by the interference. For the purpose of searching for the new particle, both the resonant Breit–Wigner and the interference are considered as parts of the signal process.

5.4.1 Analytical calculation

The leading-order cross section of the $gg \rightarrow t\bar{t}$ process, with the contribution from the heavy Higgs boson taken into account, was computed in Ref. [101]. The result obtained there for the \mathcal{CP} -even state can be adapted as follows:

$$\begin{aligned} \sigma_H(\hat{s}) - \sigma_{\text{QCD}}(\hat{s}) &= g_{Htt}^4 \cdot \frac{3\alpha_s^2 G_F^2 m_t^6 \beta^3}{1024\pi^3} \frac{16 + 8\beta^2 (\pi^2 - y^2) + \beta^4 (\pi^2 + y^2)^2}{(\hat{s} - m_\Phi^2)^2 + m_\Phi^2 \Gamma_\Phi^2} \\ &\quad - g_{Htt}^2 \cdot \frac{\alpha_s^2 G_F m_t^4 \beta^2 y}{32\pi\sqrt{2}\hat{s}} \frac{(\hat{s} - m_\Phi^2)(4 + \beta^2(\pi^2 - y^2)) + 2\pi\beta^2 m_\Phi \Gamma_\Phi y}{(\hat{s} - m_\Phi^2)^2 + m_\Phi^2 \Gamma_\Phi^2}, \end{aligned} \quad (5.2)$$

Where $\sqrt{\hat{s}}$ is the invariant mass of the incoming gluons; σ_{QCD} is the SM cross section; α_s and G_F are the strong coupling strength and the Fermi constant respectively; m_t is the mass of the top quark; and m_Φ and Γ_Φ are the mass and the total width of the heavy Higgs boson.

$$\beta = \sqrt{1 - \frac{4m_t^2}{\hat{s}}} \quad \text{and} \quad y = \ln\left(\frac{1 + \beta}{1 - \beta}\right) \quad (5.3)$$

can be interpreted, respectively, as the velocity and twice the rapidity of the top quark in the centre-of-mass frame. For simplicity, the energy-dependent width in Ref. [101] has been replaced by a constant one, effectively adopting the fixed-width scheme. The cross section for a pseudoscalar particle, obtained in a similar way, is:

$$\begin{aligned} \sigma_A(\hat{s}) - \sigma_{\text{QCD}}(\hat{s}) &= g_{Att}^4 \cdot \frac{3\alpha_s^2 G_F^2 m_t^6 \beta}{1024\pi^3} \frac{(\pi^2 + y^2)^2}{(\hat{s} - m_\Phi^2)^2 + m_\Phi^2 \Gamma_\Phi^2} \\ &\quad - g_{Att}^2 \cdot \frac{\alpha_s^2 G_F m_t^4 y}{32\pi\sqrt{2}\hat{s}} \frac{(\hat{s} - m_\Phi^2)(\pi^2 - y^2) + 2\pi m_\Phi \Gamma_\Phi y}{(\hat{s} - m_\Phi^2)^2 + m_\Phi^2 \Gamma_\Phi^2}. \end{aligned} \quad (5.4)$$

The terms proportional to g^4 in Eqs. 5.2 and 5.4 correspond to the square of the amplitude given by the signal Feynman diagram in Fig. 5.4. They produce the usual resonant peak in the $m_{t\bar{t}}$ spectrum. On the other hand, the interference terms, which are proportional to g^2 , result in a complex peak-dip structure. Both contributions, along with their sum, are shown in Fig. 5.5 for a width $\Gamma_\Phi = 0.1 \cdot m_\Phi$ and in App. A for other example values. The coupling scale factor is set to unity in these plots. As can be seen, the dip tends to dominate the combined line shape for larger masses and widths, which means that in certain cases, the presence of the heavy Higgs boson can manifest itself with a localized deficit in the $m_{t\bar{t}}$ spectrum.

In this search, the heavy Higgs boson is not required to decay exclusively to top quark pairs, which means that its total width is not fixed by the coupling scale factor but only bounded from below with the partial width $\Gamma_{\Phi \rightarrow t\bar{t}}$. The latter can be computed for the two \mathcal{CP} states as follows [101]:

$$\begin{aligned} \Gamma_{H \rightarrow t\bar{t}} &= g_{Htt}^2 \frac{3G_F m_t^2 m_\Phi}{4\pi\sqrt{2}} \left(1 - \frac{4m_t^2}{m_\Phi^2}\right)^{3/2}, \\ \Gamma_{A \rightarrow t\bar{t}} &= g_{Att}^2 \frac{3G_F m_t^2 m_\Phi}{4\pi\sqrt{2}} \left(1 - \frac{4m_t^2}{m_\Phi^2}\right)^{1/2}, \end{aligned} \quad (5.5)$$

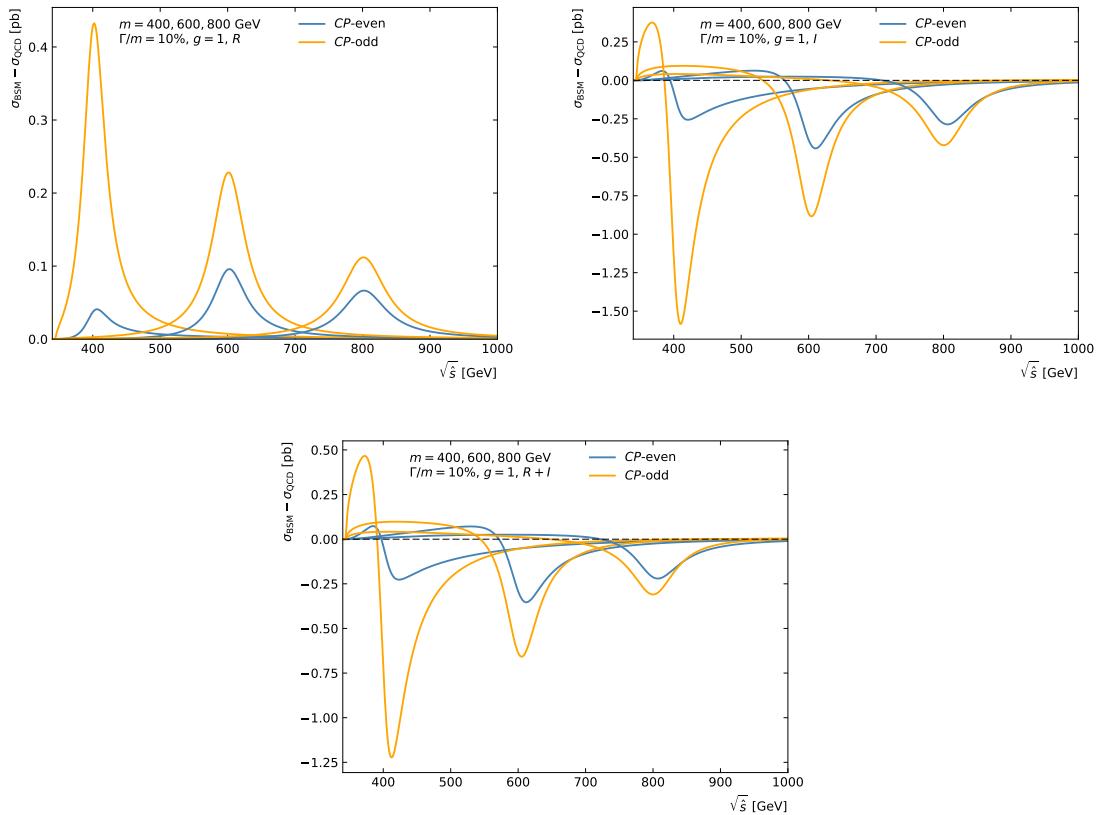


Figure 5.5: Example parton-level cross sections for the resonant part of the signal (upper left), the interference (upper right), and the sum of the two (bottom), with $g = 1$, shown as a function of $\sqrt{s} = m_{t\bar{t}}$. Computed using Eqs. 5.2 and 5.4. The total width is 10%.

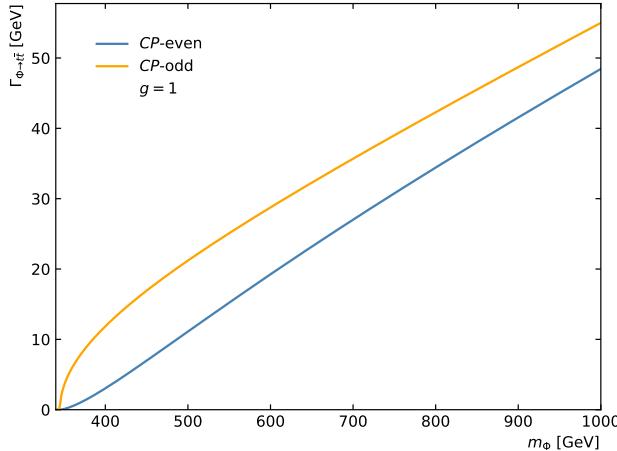


Figure 5.6: The partial width of the $\Phi \rightarrow t\bar{t}$ decay as a function of the mass of the heavy Higgs boson. Computed for $g = 1$.

Where the Higgs boson has been put on the mass shell. The dependence of the partial width on the mass of the particle is shown in Fig. 5.6. For $m_\Phi \gg 2m_t$ and $g = 1$ the relative partial width is about 6%.

BSM models with an extended Higgs sector typically include additional scalars of each \mathcal{CP} state. However, there is no interference between the $gg \rightarrow H \rightarrow t\bar{t}$ and $gg \rightarrow A \rightarrow t\bar{t}$ processes; therefore, a prediction for the full $gg \rightarrow t\bar{t}$ production can be constructed in a straightforward manner. Although this thesis mainly focuses on one \mathcal{CP} state at a time as the main analysis, the combination [102] is added in chapter 7 also considers a case where both states are present.

5.4.2 Event generation and validation

Signal events are generated with the **MADGRAPH5_AMC@NLO** program [103], version 2.5.1, which is run in the leading-order mode. A custom **MADGRAPH** model [104] adds a heavy Higgs boson to the SM, with the couplings to top quarks defined in Eq. 6.1. The effective coupling to gluons is implemented at the leading order following Ref. [105]; only top quarks are included in the loop. This model has been cross-checked against the one used in Ref. [106]. Following the choice in Ref. [106], the factorization and renormalization scales are set on a per-event basis to $m_{t\bar{t}}/2$. The leading-order PDF set **NNPDF3.0** [107] is used. The produced top quarks are decayed in **MADGRAPH5_AMC@NLO**, which preserves spin correlations. Showering and hadronization are implemented with the **PYTHIA 8** program [108], using the tune **CUETP8M2T4** [109]. Given that the sought-after particle can only manifest itself in small deviations from the SM $t\bar{t}$ production, it is impractical to generate the full $gg \rightarrow t\bar{t}$ process with the BSM contribution. Instead, only the resonant part and interference are generated, which are then used along with the centrally produced SM $t\bar{t}$ sample. Samples corresponding to the two signal components are created separately. While event generation for the resonant part poses no technical difficulty; in the case of the interference, the situation is complicated by the fact that **MADGRAPH5_AMC@NLO** does not support decay chains with the squared order syntax [110], which would be required to keep only interference terms in the squared matrix element.

Production of interference samples involves modification of the Fortran code generated by **MADGRAPH5_AMC@NLO**. Let $C_{\Phi tt}$ be the coupling of the heavy Higgs boson to top quarks, which is utilized by the routine that evaluates the squared matrix element $|\mathcal{M}|^2$. The code is modified at the point where this routine is called in the following way. First, $|\mathcal{M}|^2$ is evaluated for the nominal value of $C_{\Phi tt}$ and saved. Then, the sign of $C_{\Phi tt}$ is flipped, and $|\mathcal{M}|^2$ is computed again, yielding value $|\mathcal{M}(-C_{\Phi tt})|^2$. The effective coupling of Φ to gluons is controlled by an independent parameter in the routine and is therefore not affected by the modification of $C_{\Phi tt}$. As a result, the signs of the interference terms in $|\mathcal{M}|^2$ are flipped, while the SM terms (independent of the Φ couplings) and the resonant BSM part (proportional to $C_{\Phi tt}^2$) are left unchanged. Finally, the value $(\mathcal{M}^2(C_{\Phi tt}) - \mathcal{M}^2(-C_{\Phi tt})) / 2$ is computed and then used in place of the original squared matrix element everywhere. All terms except the interference terms cancel out in this computation.

The part of $|\mathcal{M}|^2$ that represents the interference can be written as $2\Re(A_S A_B^\dagger)$, where A_S and A_B are the amplitude of the signal process and the sum of amplitudes for all SM diagrams respectively. Unlike the usual squared matrix element, this quantity can be negative in some regions of the phase space; this fact is reflected

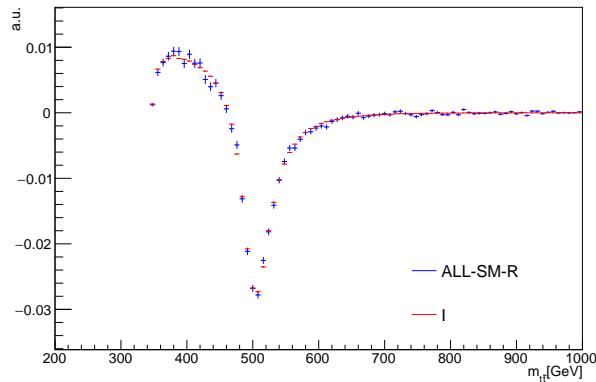


Figure 5.7: Spectra of m_{tt} in the interference obtained by the modification of the generated squared matrix element (red) and the validation procedure detailed in the text (blue).

in the sign of event weights assigned by the generator.

The generation procedure is validated for an example of a pseudoscalar particle with a mass of 500 GeV and 10% width. A sample of 40 M events is generated for the full $gg \rightarrow t\bar{t}$ process. The interference is then modelled by subtracting the SM-only production (40 M events) and the resonant Φ production (400 k events) from it, all normalized to the respective leading-order cross sections. Distributions of various observables are compared between this reference construction and the method described above. The comparison for the $t\bar{t}$ invariant mass is shown in Fig. 5.7, and App. B includes other observables along with additional details about the validation. Although some small discrepancies can be observed, they are consistent with statistical fluctuations.

A number of versions of signal samples have been produced to scan parameters of the heavy Higgs boson. The couplings (g_{Htt}, g_{Att}) are set either to (1, 0) or (0, 1). This results in pure \mathcal{CP} states, which, as discussed above, can be mixed in a straightforward manner if required. Since samples for the resonant part and the interference are generated independently, distributions for any values of the coupling can be reproduced by scaling the normalization factors with g^4 and g^2 respectively (cf. Eqs. 5.2 and 5.4). Four mass hypotheses are modelled for each \mathcal{CP} state: 400, 500, 600, and 750 GeV. For each mass point, the generation is performed for a total width of 2.5, 5, 10, 25, and 50%, where the % is with respect to the mass point. For example, 10% of mass 500 GeV is a 50 GeV width. Furthermore, samples with the two final states are produced. In the first group, one of the top quarks must decay, producing an electron or a muon (but not a tau lepton), while the other one must decay to quarks. In the second group, each of the top quarks must decay, producing a charged lepton of any generation. These samples are mostly intended for the complementary search in the dilepton channel but nonetheless account for a small fraction of selected events in the $l + jets$ channel.

Figures 5.8 to 5.11 show distributions of the $t\bar{t}$ invariant mass for different signal hypotheses. The resonant and interference parts are shown separately, all normalized to the same area.

5.5 Cross section and BR calculation

To scale the signal leading-order cross section to next-to-next leading order, a number of different types of cross section calculators have been used, as defined in the following section.

5.5.1 2HDMC and SUSHI

2HDMC: Two-Higgs-Doublet Model Calculator [111] is a general-purpose calculator based on C++ code, which can be used to study the phenomenology of a general (\mathcal{CP} -conserving) two-Higgs doublet model (2HDM), as described in Sec. 2.2. 2HDMC provided a user-friendly interface to implement its favourite parameters' space for the Higgs potential. A user has full control over the Yukawa sector through changing the coupling. The Higgs masses, type of the model (type I and type II), alignment limit condition $\sin(\alpha - \beta)$, the ratio of the vacuum expectation values of the doublet in 2HDM ($\tan \beta$), the scalar mass matrix, m_{12}^2 , etc. can be specified with full freedom. The output of the 2HDMC consists of a check on the theoretical properties of the model,

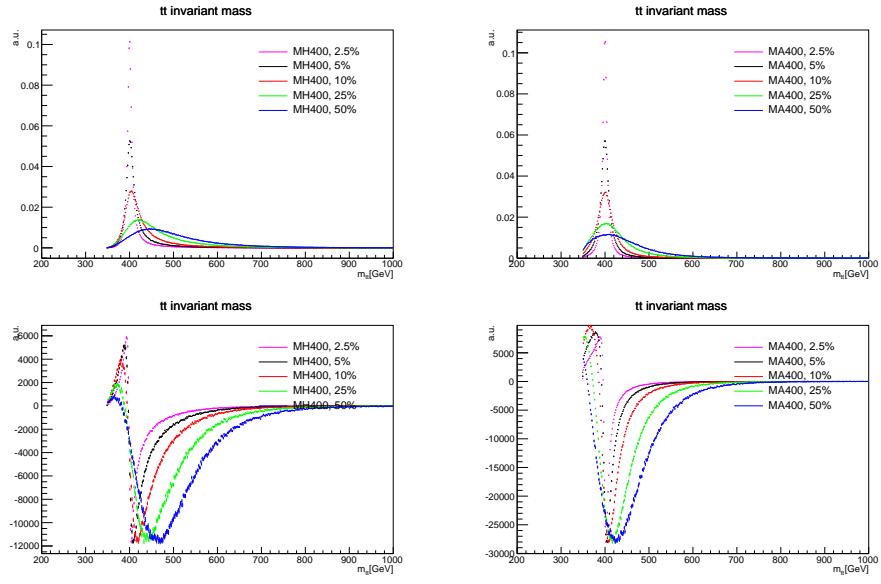


Figure 5.8: Distribution of the $t\bar{t}$ invariant mass at parton level for a \mathcal{CP} -even (left) and a \mathcal{CP} -odd (right) particle of mass 400 GeV for different width hypotheses. The plots correspond to the resonant BSM contribution (top) and interference (bottom).

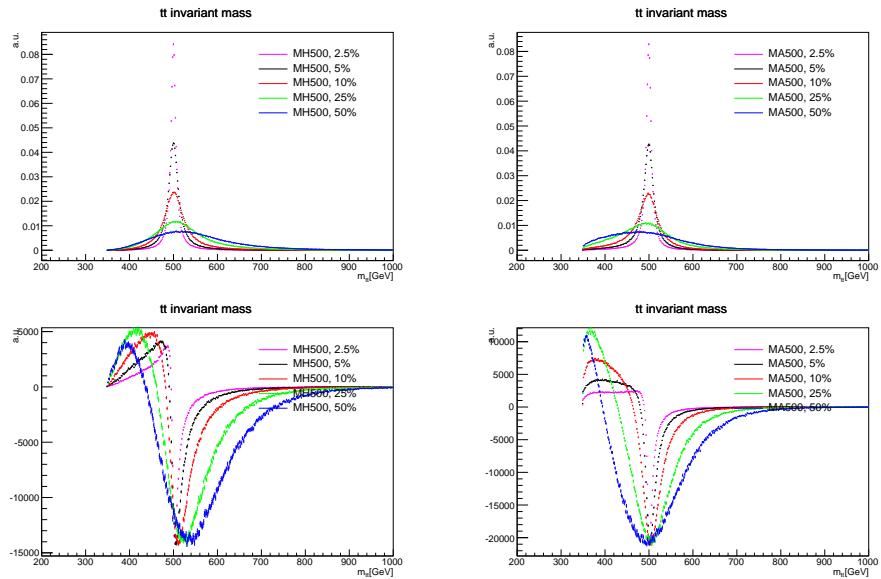


Figure 5.9: Distribution of the $t\bar{t}$ invariant mass at parton level for a \mathcal{CP} -even (left) and a \mathcal{CP} -odd (right) particle of mass 500 GeV for different width hypotheses. The plots correspond to the resonant BSM contribution (top) and interference (bottom).

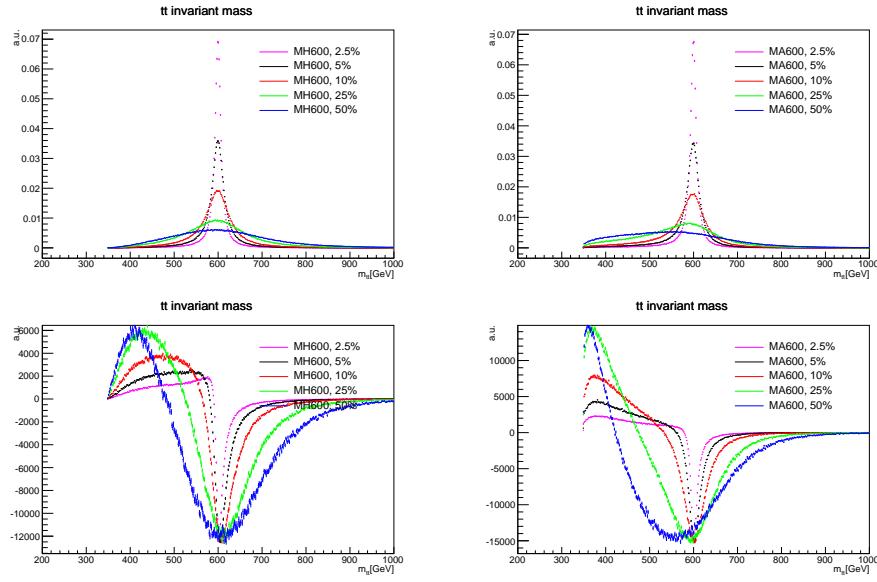


Figure 5.10: Distribution of the $t\bar{t}$ invariant mass at parton level for a \mathcal{CP} -even (left) and a \mathcal{CP} -odd (right) particle of mass 600 GeV for different width hypotheses. The plots are shown for the resonant BSM contribution (top) and interference (bottom).

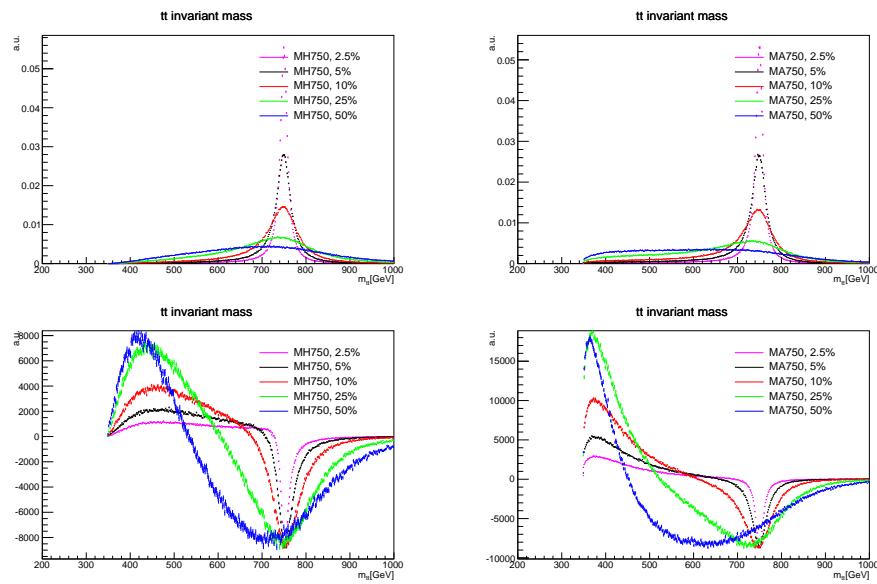


Figure 5.11: Distribution of the $t\bar{t}$ invariant mass at parton level for a \mathcal{CP} -even (left) and a \mathcal{CP} -odd (right) particle of mass 750 GeV for different width hypotheses. The plots are shown for the resonant BSM contribution (top) and interference (bottom).

such as \mathcal{CP} -conservation, positivity and stability of the potential, and the tree level unitarity and perturbativity. 2HDMC can be used to calculate the Higgs decay widths and branching ratios.

This work uses 2HDMC to calculate decay widths of the neutral Higgs sector, scalar (H), or pseudo-scalar (A), by making an iteration over the $\tan\beta$ values. The signal samples have been generated for heavy Higgs (scalar and pseudo-scalar) using MADGRAPH for fixed masses and widths. The considered masses are 400, 500, 600, and 750 GeV; for each mass, five width values (1, 5, 10, 25, 50)% are taking into account. 2HDMC uses fixed values for charged Higgs mass ($m_C = 600$ GeV), scalar or pseudo-scalar Higgs mass ($m_A/H = 400$ GeV), and $\sin(\alpha - \beta) = 1$, $\lambda_{6,7} = 0$ and $m_{12}^2 = m_A^2 \tan\beta / (1 + \tan^2\beta)$. An iteration of the $\tan\beta$ values is conducted and the one that provides the corresponding width considered for signal generation in MADGRAPH is selected. The same procedure is repeated for all masses. The selected $\tan\beta$ values are further used in the SUSHI as input.

SUSHI: Supersymmetric Higgs [112] is a Fortran-based program, designed to calculate inclusive Higgs boson production cross sections up to NNLO QCD through gluon fusion and bottom-quark annihilation in the Standard Model (SM), general Two-Higgs-Doublet Models (2HDM), the Minimal Supersymmetric Standard Model (MSSM), and its next-to-minimal extension (NMSSM). The program can also be used to calculate differential cross sections with respect to the Higgs transverse momentum p_T and (pseudo-)rapidity $y(\eta)$. SUSHI can be linked with the 2HDMC calculator for performing 2HDM calculation, with FeynHiggs for the MSSM Higgs masses calculations, and many more. In this work, SUSHI links with 2HDMC for neutral Higgs LO and NNLO cross section calculations in 2HDM/hMSSM and incorporates the PDF effects by using external PDF sets (NNPDF30). Factorization and normalization scales are fixed with respect to the neutral Higgs mass ($\mu_{F/R} = m_{A/H}/2$). The output card of the 2HDMC is used as the input in SUSHI with 2HDM model in the physical Higgs basis. The $\tan\beta$ value that corresponds to a certain width is used as a coupling modifier ($g_{A/H} \rightarrow t\bar{t}$) in the MADGRAPH parameter card. The output of LO cross sections from SUSHI is comparable with the MADGRAPH cross sections shown in table 5.1. The k-factor is calculated as a ratio of NNLO SUSHI cross sections to LO. The results are interpreted to weight the SM cross sections with these k-factors. NNLO SUSHI cross sections to LO MADGRAPH with errors is plotted in Fig. 5.12.

Mass (GeV)	Width (GeV)	$\tan \beta$	$\frac{1}{\tan \beta}$	LO $\sigma(MG)$ pb	LO $\sigma(\text{SUSHi})$ pb	NNLO $\sigma(\text{SUSHi})$ pb	BR ($A \rightarrow t\bar{t}$)	$K = \frac{NNLO\sigma(\text{SUSHi})}{\sigma(MG)}$
400	4.002	1.908	0.52410901467	3.641 ± 0.003587	3.63478 ± 0.00000	7.71736 ± 0.00795	9.88010^{-01}	2.12 ± 0.003
	10	1.2068	0.82863771958	9.096 ± 0.009417	9.13195 ± 0.0	19.26195 ± 0.01987	9.95010^{-01}	2.118 ± 0.009417
	19.49	0.8645	1.15673799884	17.76 ± 0.01694	17.82577 ± 0.00001	37.51884 ± 0.03872	9.96010^{-01}	2.113 ± 0.0030
	38.99	0.6113	1.63585800752	35.47 ± 0.03629	35.68351 ± 0.00002	75.01936 ± 0.07743	9.96310^{-01}	2.115 ± 0.0031
	97.52	0.3865	2.5873221216	88.86 ± 0.07161	89.31375 ± 0.00005	187.64029 ± 0.19371	9.96310^{-01}	2.112 ± 0.0028
	195	0.2733	3.65898280278	177.3 ± 0.1163	178.65641 ± 0.00009	375.25547 ± 0.38741	9.96310^{-01}	2.117 ± 0.0026
500	5.032	2.12	0.4716981132	0.9349 ± 0.0007767	0.94977 ± 0.00000	1.90812 ± 0.00263	9.62710^{-01}	2.041 ± 0.0033
	12.4	1.3506	0.73746312684	2.291 ± 0.002217	2.32963 ± 0.000	4.66077 ± 0.00648	9.85410^{-01}	2.034 ± 0.0034
	24.81	0.9549	1.04723007645	4.613 ± 0.00459	4.65371 ± 0.000	9.29670 ± 0.01297	9.91810^{-01}	2.015 ± 0.0035
	49.61	0.6752	1.48104265403	9.235 ± 0.009654	9.30136 ± 0.00001	18.56738 ± 0.02594	9.94710^{-01}	2.011 ± 0.0035
	124.0	0.4271	2.34137204402	23.08 ± 0.02179	23.23658 ± 0.00001	46.36393 ± 0.06484	9.96410^{-01}	2.009 ± 0.0034
	248.2	0.3019	3.31235508447	46.15 ± 0.041482	46.49921 ± 0.00003	92.76579 ± 0.12977	9.96910^{-01}	2.010 ± 0.0033
600	6.017	2.19	0.45662100456	0.3291 ± 0.0003056	0.33671 ± 0.000	0.65702 ± 0.00117	4.13310^{-01}	1.996 ± 0.0040
	15.02	1.3863	0.7213445863	0.8211 ± 0.0007007	0.83209 ± 0.000	1.61965 ± 0.00292	6.38010^{-01}	1.973 ± 0.0039
	30.03	0.9803	1.02009588901	1.641 ± 0.001376	1.65874 ± 0.000	3.22588 ± 0.00583	7.78610^{-01}	1.966 ± 0.0039
	60.05	0.6932	1.44258511252	3.28 ± 0.00259	3.31199 ± 0.000	6.43825 ± 0.01166	8.74810^{-01}	1.963 ± 0.0039
	150.1	0.4384	2.28102189781	8.213 ± 0.007184	8.27281 ± 0.000	16.07741 ± 0.02915	9.44810^{-01}	1.958 ± 0.0039
	300.3	0.31	3.22580645161	16.41 ± 0.01363	16.53994 ± 0.00001	32.14089 ± 0.05831	9.70610^{-01}	1.959 ± 0.0039
700	7.012	1.97	0.5076142132	0.1219 ± 0.0001041	0.11765 ± 0.00000	0.22728 ± 0.00052	8.64910^{-01}	1.864479 ± 0.005120
	17.7	1.24	0.8064516129	0.3078 ± 0.0002247	0.29830 ± 0.00000	0.57819 ± 0.00131	9.42710^{-01}	1.878460 ± 0.004986
	34.36	0.89	1.1235955056	0.5973 ± 0.0004875	0.57992 ± 0.00000	1.12524 ± 0.00254	9.69210^{-01}	1.883877 ± 0.005069
	70.80	0.62	1.6129032258	1.231 ± 0.001089	1.19600 ± 0.00000	2.32196 ± 0.00523	9.84110^{-01}	1.886239 ± 0.005133
	175.3	0.394	2.538071066	3.048 ± 0.002417	2.96298 ± 0.00000	5.75427 ± 0.01296	9.92610^{-01}	1.887884 ± 0.005045
	349.6	0.279	3.5842293907	6.08 ± 0.00438	5.90994 ± 0.00000	11.47866 ± 0.02585	9.95510^{-01}	1.887937 ± 0.004972

Table 5.1: Comparison of leading order cross section from MADGRAPH and SUSHI for pseudo-scalar resonance with masses 400, 500, and 600 GeV and scalar mass = 700 GeV. The typical k-factor shows universality for a single mass point that is calculated as the ratio of NNLO from SUSHI and MADGRAPH LO cross sections.

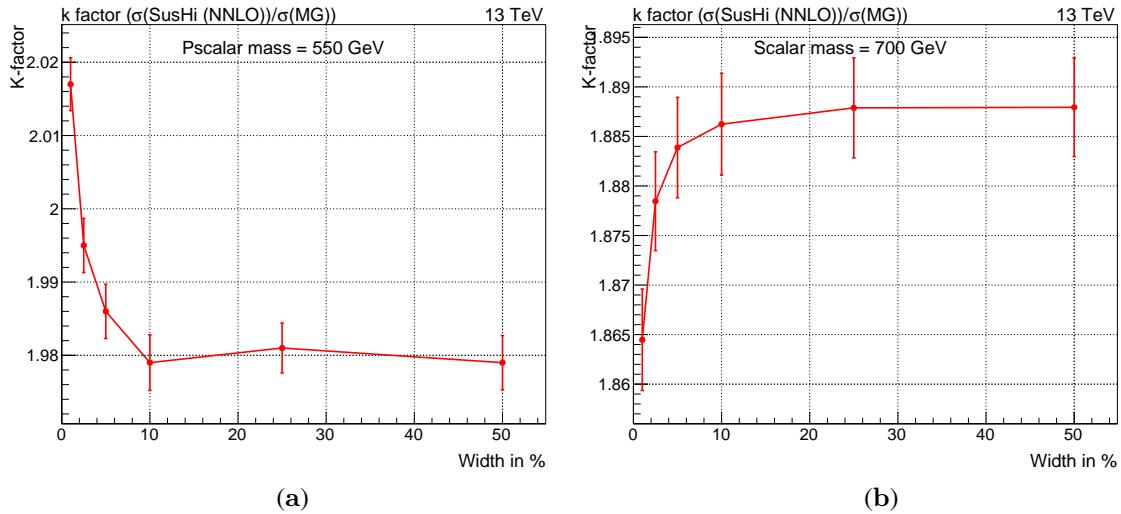


Figure 5.12: Typical k -factor for pseudo-scalar mass = 550 GeV (a) and scalar mass = 700 GeV (b), obtained from the ratio of $\sigma_{NNLO}(\text{SUSHI})$ to $\sigma(\text{MG})$. Both SUSHI and MG LO cross sections are comparable, which leads to the same results if we use $\sigma_{LO}(\text{SUSHI})$ instead of $\sigma_{LO}(\text{SUSHI})$. For the whole range of width, the k -factor shows universality, which is easy to apply for other width values .

5.5.2 TOP++

The program TOP++ [113] calculates the total inclusive cross-section for top pairs production in hadronic collisions using two different approaches a) fixed order with NNLO accuracy and b) including soft-gluon resummation in Mellin space with full next-to-next-to-leading logarithmic order (NNLL) accuracy matched through NNLO. TOP++ is the only publicly available program using soft-gluon resummation for the $t\bar{t}$ inclusive production cross section in hadron colliders. The TOP++ program is based on C++ coding in a modular and easy manner with a single configuration file (`top++.cfg`) for user interface. The configuration file consists of simple input parameters such as the type of collider, PDF set, pure fixed order calculation versus one with resummation, mass of top quark, and LO, NLO, and NNLO orders. In this work, we use TOP++ to calculate the $t\bar{t}$ hadronic cross section up to NNLO accuracy with soft-gluon resummation and LO cross section without soft-gluon resummation. From these two cross sections, the k -factor ($\sigma(\text{NNLO})/\sigma(\text{LO})$) has been calculated for further scaling of interference. The input parameters used for this study (PDF sets) are NNPDF30 for NO and NNLO; the top quark mass is 172.5 GeV, $\mu_{\text{R/F}}$ scale varied as 0.5, 1, and 2.

6

Search for a heavy higgs state in the semileptonic channel

This chapter gives a complete overview of the heavy Higgs search in the $t\bar{t}$ semileptonic final state. The chapter is organized sequentially as, starting from the signal modelling with analytical calculation and event generation and validation, the simulation and the data used in the analysis. It defines corrections to simulation, pile-up reweighting, jet energy scale and resolution, b-tagging, trigger efficiencies, lepton scale factors, generator level weights and top p_T correction. It also describes the final objects selection used in the analysis which consists of leptons (muon or electron) with corresponding neutrino, jets, b-tagged jets and the event selection in the semileptonic final state. The search variables utilized for the statistical evaluation, the data-driven modelling of the multijet background, systematic uncertainties and the final statistical analysis are described.

The work described in this chapter has been documented in the CMS analysis note in Ref. [114]. It is a team work and I contributed in specific parts, partially in the signal generation and simulation, synchronization of objects and event selection with other groups in our team, process full 2016 data and all simulated backgrounds, applied corrections to data and simulations and compared data to background. My further contribution is in making QCD templates for the final statistical analysis using data driven techniques. I also tried to apply top p_T correction scale factors from theory. I calculated the k-factors to scale the signal both in semileptonic and dileptonic final states to higher order cross section for interpretation in the context of hMSSM.

The discovery of the Higgs boson by CMS [1] and ATLAS [2] in July 2012 was not only the confirmation of the Higgs sector in the Standard Model, but also many consequences on BSM. These results strongly support the Higgs mechanism of electroweak symmetry breaking, but don't exclude the possibility of any additional Higgs state with mass range below or around 1 TeV. A number of searches are devoted to additional spin-zero particles at the LHC, but until now all of them are negative [38]. There are many extensions of the SM, like MSSM, 2HDM [115]. In the Standard Model the simplest possible scalar structure is assumed (one SU(2) doublet) that has only one Higgs boson; on the contrary, its extensions like 2HDM produces a spectrum of physical particles which consists of five spin 0 particles, two of them are charged H^\pm , two are \mathcal{CP} -even neutral H and one \mathcal{CP} -odd A .

Different types of possibilities must be take into account to search these particles. One possibility is to reduce the couplings of one of these Higgs particle to the weak vector bosons - or even zero as in the case of pseudo-scalar - where its couplings to fermions enhanced with respect to the SM Higgs couplings [116]. The Higgs sector in the above mentioned models have strong Youkawa couplings to the top quark in the small vacuum expectation value and higher mass regime. Hence decay of \mathcal{CP} -even neutral H to $t\bar{t}$ becomes dominant. At tree level the \mathcal{CP} -odd A can't decay into WW and ZZ and the only mode of decay is $t\bar{t}$. This makes the \mathcal{CP} -odd state A more interesting.

The large branching ratio of the heavy Higgs boson, Φ , in the $t\bar{t}$ final state is more attractive at the LHC but this channel faces challenges from the SM $t\bar{t}$ background which is copiously produced. The production of heavy Higgs from gluon fusion via a top quark loop and its subsequent decay into $t\bar{t}$ interferes with the SM $t\bar{t}$. A non-trivial phase involved in the signal and SM background amplitude which further augmented the large interference effects. As a result, a peak-dip, dip-peak, pure dip, pure peak or nothingness [36] can appear instead of normal resonance shape and doesn't vanish at the narrow width approximation. The only public experimental result in which the interference has been taken into account is the search for $H/A \rightarrow t\bar{t}$ performed by the ATLAS collaboration [117].

This chapter and the following are based on the search of a heavy Higgs boson Φ in the final state of $t\bar{t}$. The $t\bar{t}$ decays in semileptonic final state that consists of a lepton (electron or muon) and jets. The interference effects from the SM $t\bar{t}$ has been taken explicitly. The coupling of the new particle to top quarks is given by

$$\mathcal{L}_{\text{Yukawa}} \supset \frac{m_t}{v} (g_{H_{tt}} t\bar{t} + i g_{A_{tt}} t\gamma_5 \bar{t}) \Phi, \quad (6.1)$$

where m_t is the mass of the top quark, $v \approx 246$ GeV is the vacuum expectation value, and $g_{H_{tt}}$ and $g_{A_{tt}}$ are arbitrary real-valued scale factors. To consider a pure \mathcal{CP} (scalar, H) or \mathcal{CP} -odd (pseudoscalar, A), one of the two coupling scale factors is always zero.

The analysis included in this thesis is performed based on pp collision data collected by the CMS experiment during 2016, corresponding to integrated luminosity of 36 fb^{-1} . The semileptonic final state has been selected with a lepton (muon or electron), at least four jets, two of them must be b-tagged. The $t\bar{t}$ system is reconstructed by utilizing kinematic constraints imposed by the known masses of the top quark and the W boson and the interference effects coming from the SM $t\bar{t}$ is taken into account. All backgrounds are taken from the simulation except the QCD, which is estimated from the data using a data driven techniques. A two-dimensional distribution of the mass of the reconstructed $t\bar{t}$ system, $m_{t\bar{t}}$ and an angular variable $|\cos \theta^*|$ is investigated with a complete assessment of the main systematic uncertainties. With no indication of the presence of the BSM signal, upper limits on the coupling scale factors are computed for various signal hypotheses.

The dileptonic channel of the $t\bar{t}$ final state has been searched and separately documented in Ref. [118]. The semileptonic and dileptonic channels are combined in Ref. [102] which includes more detailed discussion of the statistical analysis and an interpretation of results in the hMSSM model [119]. Chapter 7 based on the combination of semileptonic and dileptonic channel.

6.1 Datasets

6.1.1 Collision data

The data used in the present search collected in 2016 in pp collisions at a centre-of-mass energy of 13 TeV. Two primary data sets, SingleMuon and SingleElectron are used, both of the data sets are divided into seven data taking periods (B, C, D, E, F, G and H) under different data conditions. The analysis uses the latest reprocessing, 03Feb2017, and only certified data selected by mask Cert_271036-284044_13TeV_23-Sep2016ReReco_Collisions16_JSON.txt are considered. Corresponding integrated luminosity, as recorded with an unprescaled trigger, totals to $35.9 \pm 0.9 \text{ fb}^{-1}$ [120]. The complete list of data sets is provided in Table 6.1.

6.1.2 Simulation

Signal and background characteristics can be investigated using simulated samples. Monte Carlo (MC) events for signal samples were produced using **MADGRAPH5_AMC@NLO** event generator as described in Sec. 5.4.2 and simulation is done privately with the same configuration for modelling of the detector response and for reconstruction as applied in central campaign RunIIISummer16MiniAODv2. Samples with the $l+jets$ final state are published under names that match the mask

```
/HToTT-semilep_{parity}-M{mass}-RelW{width}-{type}-v2_
13TeV-madgraph-pythia8/aapopov-MiniAOD-
28028af67189b3de7224b79195bd0e1d/USER
```

where the placeholders take the following values:

Data set	Integrated luminosity, fb^{-1}
/SingleMuon/Run2016B-03Feb2017_ver2-v2	5.8
/SingleMuon/Run2016C-03Feb2017-v1	2.6
/SingleMuon/Run2016D-03Feb2017-v1	4.2
/SingleMuon/Run2016E-03Feb2017-v1	4.0
/SingleMuon/Run2016F-03Feb2017-v1	3.1
/SingleMuon/Run2016G-03Feb2017-v1	7.5
/SingleMuon/Run2016H-03Feb2017_ver2-v1	8.4
/SingleMuon/Run2016H-03Feb2017_ver3-v1	0.2
/SingleElectron/Run2016B-03Feb2017_ver2-v2	5.8
/SingleElectron/Run2016C-03Feb2017-v1	2.6
/SingleElectron/Run2016D-03Feb2017-v1	4.2
/SingleElectron/Run2016E-03Feb2017-v1	4.0
/SingleElectron/Run2016F-03Feb2017-v1	3.1
/SingleElectron/Run2016G-03Feb2017-v1	7.5
/SingleElectron/Run2016H-03Feb2017_ver2-v1	8.4
/SingleElectron/Run2016H-03Feb2017_ver3-v1	0.2

Table 6.1: Data sets with collision data. Ending `/MINIAOD` in their names is omitted. Integrated luminosities are computed using the certification file referenced above.

```
{parity}: scalar, pseudoscalar
{mass}: 400, 500, 600, 750
{width}: 2p5, 5, 10, 25, 50
{type}: res, int
```

Each sample contains about 0.5 M events. Although the dilepton final state is not included in this thesis, it can give a small contribution when one of the leptons is not reconstructed or when it is a τ lepton reconstructed as a jet. For this reason corresponding samples are also included. They were not published in DAS, and files are stored at T2_DE_DESY under path `/store/user/afiqaze/samples80X/ahtt_maod_100117/`. There is about 0.4 M events generated for each combination of parameters.

Samples with simulation of background processes considered in this search are listed in Table 6.1.2. The dominant background is $t\bar{t}$ production, which is modelled with POWHEG v2 [121] and PYTHIA 8. The tune CUETP8M2T4, also referred to as CP0, is used, which improves description of the jet multiplicity and some other observables. The parton distribution functions are taken from NNPDF3.0. The cross section of $t\bar{t}$ production was computed with the TOP++ 2.0 program to next-to-next-to-leading order (NNLO) in perturbative QCD, including soft-gluon resummation to next-to-next-to-leading-log order, (see Ref. [113] and references therein). A top quark mass $m_t = 172.5$ GeV is assumed in both event generation and computation of the cross section.

Remaining minor backgrounds are normalized according to cross sections given in Ref. [122], with two exceptions. Leading-order cross sections are used for samples of events comprised of jets produced via the strong interaction (referred to as “QCD” in the table), and a special procedure is applied for W boson production, which is described with a set of samples with different parton multiplicities. These exclusive samples are normalized by scaling the NNLO inclusive cross section by the fraction of events with the given multiplicity found in the inclusive sample, in the same way as was done in Ref. [123]. Events with zero partons in the final state of the matrix element are neglected in this analysis as they constitute only about 3% of selected $W + \text{jets}$ events, and W boson production is already a small background.

Finally, Table 6.3 lists dedicated samples with systematic variations in $t\bar{t}$ production. As will be discussed in Sec. 6.3, they are assigned the same cross section as the nominal sample, i-e 831.76 pb. Samples with systematic variations for other backgrounds are not considered because of the small contribution from these processes.

Data set	Cross section, pb	Equivalent luminosity, fb^{-1}
TT_TuneCUETP8M2T4_13TeV-powheg-pythia8	831.76	186
ST_t-channel_top_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powhegV2-madspin	136.02	44
ST_t-channel_antitop_4f_inclusiveDecays_TuneCUETP8M2T4_13TeV-powhegV2-madspin	80.95	49
ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85	28
ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M2T4	35.85	29
ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1	3.36	298
W1JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	11917.5	3.8
W2JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	3850.4	7.8
W3JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1124.0	53
W4JetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	579.3	52
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5765.4	25
WW_TuneCUETP8M1_13TeV-pythia8	118.7	8.4
WZ_TuneCUETP8M1_13TeV-pythia8	47.13	21
ZZ_TuneCUETP8M1_13TeV-pythia8	16.523	60
TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.2043	11×10^3
TTWJetsToQQ_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	0.4062	2.1×10^3
TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.2529	7.9×10^3
TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.5297	1.4×10^3
QCD_Pt-*_MuEnrichedPt5_TuneCUETP8M1_13TeV_pythia8	—	—
QCD_Pt-*_EMEnriched_TuneCUETP8M1_13TeV_pythia8	—	—
QCD_Pt_*_bcToE_TuneCUETP8M1_13TeV_pythia8	—	—

Table 6.2: Nominal simulated samples used in the analysis. Only primary data set names are listed, while the condition string “RunIIISummer16MiniAODv2-PUMoriond17*_80X_mcRun2_asymptotic_2016_TrancheIV_v6*” and the data tier “MINIAODSIM” are omitted. Effective integrated luminosity is computed taking into account all available extension samples.

Data set	Equivalent luminosity, fb^{-1}
TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8	71
TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8	71
TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8	71
TT_TuneCUETP8M2T4_13TeV-powheg-fsrdown-pythia8	71
TT_hdampUP_TuneCUETP8M2T4_13TeV-powheg-pythia8	71
TT_hdampDOWN_TuneCUETP8M2T4_13TeV-powheg-pythia8	70
TT_TuneCUETP8M2T4_mtop1755_13TeV-powheg-pythia8	71
TT_TuneCUETP8M2T4_mtop1695_13TeV-powheg-pythia8	70
TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8	71
TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8	70

Table 6.3: Simulated samples with systematic variations. Only primary data set names are listed, while the condition string “RunIISummer16MiniAODv2-PUMoriond17*_80X_mcRun2_asymptotic_2016_TrancheIV_v6*” and the data tier “MINIAODSIM” are omitted. Effective integrated luminosity is computed taking into account all available extension samples.

6.2 Corrections to simulation

Following corrections to simulation are applied per-event basis to reproduce corresponding distributions in data.

6.2.1 Pile-up re-weighting

In the high luminosity pp collision in the LHC, in-time multiple interactions stemming from soft QCD and overlap to hard scattering, can occur in the same bunch crossing. Out-of-time pp collisions occurring in bunch crossings just before and after the collision of interest. Due to in-time and out-of-time pile-up in the simulated events, the pile-up profile imposed for the MC production does not agree exactly with the one observed in the data, as illustrated by Fig. 6.1. The profile in data is deduced from the distribution of the luminosity delivered per bunch crossing, assuming for the (effective) total cross section of inelastic pp scattering a value of 69.2 mb. In order to reproduce this profile, events in simulation are assigned weights based on the expected (“true”) number of pile-up interactions, as directed by the standard prescription [124, 125].

6.2.2 Jet Energy Scale and Resolution

Jet Energy Corrections (JEC) are applied to calibrate jets in order to have the correct energy scale like other experimental objects. Jet energy scale corrections are applied to the reconstructed jets in simulation using the so called Summer16_23Sep2016ReReco_V4 [126] to have the correct energy scale as observed in data. Jet Energy Resolution (JER) gives the spread of the response in the Gaussian core region and is different from JEC. In simulation the nominal jet energy resolution is smeared using a p_T, η -dependent parameterization [127]. This analysis uses the recommended “hybrid” method for JER.

6.2.3 B tagging

The identification of b-jet is very important for the study of top quark decays, Higgs decay and many new physics phenomena. b-tagging is a reconstruction technique that takes advantage of the b hadron properties and assigns to each jet a likelihood that contains a b hadron. Following are the b-jets characteristics which discriminate them from jets produced by light quarks and gluons.

- Long life time: $\tau \sim 1.5 \text{ ps}$, $c\tau \sim 450 \mu\text{m}$
- Large mass: mass $\sim 4.2 \text{ GeV}$
- High track multiplicity $\sim 4\text{-}5$
- Large semileptonic branching fraction

We use the combined multivariate algorithm (cMVAv2) that combines information from six different b-jet identification discriminators with a Boosted Decision Tree (BDT) [128]. cMVAv2 has three operating points,

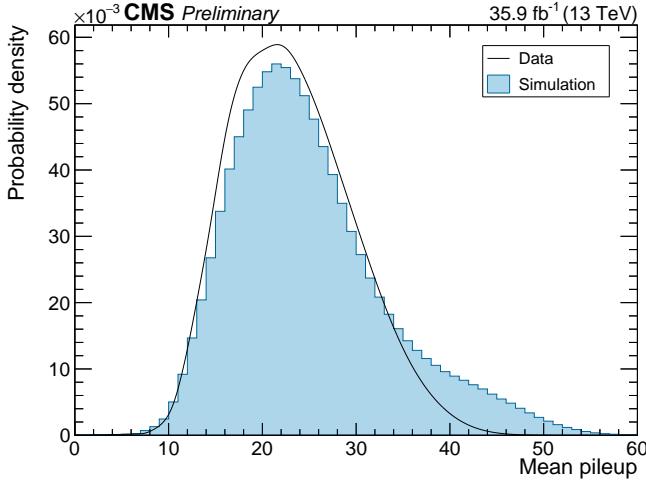


Figure 6.1: Pile-up profiles in data and simulation without corrections.

loose, medium and tight and this analysis uses the medium working point with $\text{cMVAv2} > 0.4432$. Performance of b-tagging algorithms in data is known to be slightly worse than expected from simulation [129]. In order to account for the difference, the b-tagging efficiencies in simulation are modified with the help of scale factors provided by the BTV POG [130]. There are several ways to perform the correction. This search utilizes method 1a from Ref. [131], which is summarized below.

For each simulated event the probability to reproduce the observed b-tagging configuration is given by

$$\mathcal{P}_{\text{Sim}} = \prod_{i \in \text{tagged}} \epsilon_i \cdot \prod_{j \notin \text{tagged}} (1 - \epsilon_j), \quad (6.2)$$

where ϵ_i is the b-tagging efficiency for jet i in simulation, and the first (second) product is calculated over b-tagged (not b-tagged) jets. The probability to reproduce the same b-tagging configuration with efficiencies as in data reads as

$$\mathcal{P}_{\text{Data}} = \prod_{i \in \text{tagged}} s_i \epsilon_i \cdot \prod_{j \notin \text{tagged}} (1 - s_j \epsilon_j), \quad (6.3)$$

where s_i is the scale factor for jet i , which is parameterized as a function of p_T , η , and flavour of the jet. In this analysis scale factors from the payload “ttbar” (“incl”) are used for b and c quark (light-flavour) jets. To account for the mismodelling of the b-tagging probabilities, the event is assigned a weight

$$w = \mathcal{P}_{\text{Data}} / \mathcal{P}_{\text{Sim}} = \prod_{i \in \text{tagged}} s_i \cdot \prod_{j \notin \text{tagged}} \frac{1 - s_j \epsilon_j}{1 - \epsilon_j}. \quad (6.4)$$

The b-tagging efficiencies in simulation, which are needed to determine the weight, depend on the physics process and the event selection and therefore cannot be provided centrally. They have been computed after applying the full analysis event selection (which will be described later) but b-tagging. The measurement is done with only the $t\bar{t}$ sample as this is by far the dominant background and also representative of jet properties in the signal process. The efficiencies are parameterized with jet p_T , η , and flavour, with the three usual flavour categories considered: b and c quark jets and all the rest.

The effect of b-tagging correction is shown on leading b-jet p_T in Fig. 6.2, (a) is without b-tagging correction while correction is applied in (b).

6.2.4 Trigger efficiency

Electron trigger efficiency: Corrections to efficiency of electron trigger `HLT_Ele27_WPTight_Gsf_v*`, which is used in this search, are not provided centrally. Instead, a dedicated measurement is performed in a sample of $Z \rightarrow e^+e^-$ events using the tag-and-probe method [132–134]. Results discussed below have been presented before EGM POG and approved [135].

The method exploits two sets of selections for electrons. A “tag” electron is required to have $p_T > 40$ GeV

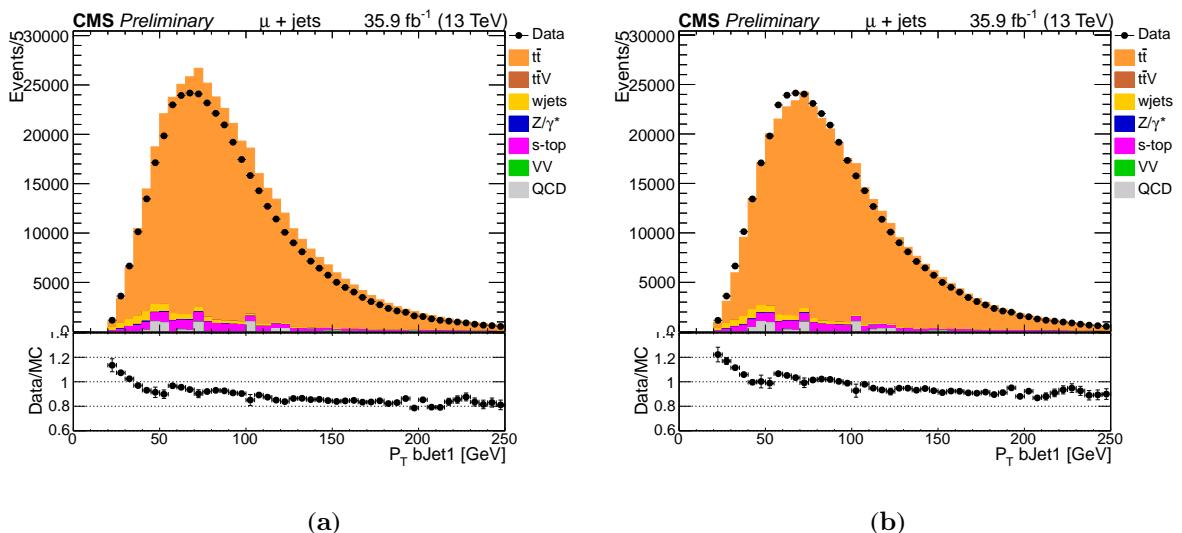


Figure 6.2: Transverse momentum of first b -tagged jet in data and MC pre (a) and post (b) b -tagging correction in the $t\bar{t}$ dominated region using muon channel. Data/MC comparison improves after applying jet energy correction.

and $|\eta_{\text{SC}}| < 2.1$, not fall in the transition region between the barrel and the endcaps of the ECAL ($1.4442 < |\eta_{\text{SC}}| < 1.5660$), and pass the tight working point of the cut-based identification algorithm and the recommended additional selection on impact parameters. It must also be matched, within $\Delta R < 0.3$, to a trigger object passing the last filter in trigger `HLT_Ele27_WPTight_Gsf_v*`. Finally, it is required to be matched, also within $\Delta R < 0.3$, to an L1T e/γ object that passes trigger `L1_SingleIsoEG34er`. This last condition is needed because all single-electron HLT paths in 2016 are seeded by an inclusive *or* of all available single- e/γ L1T seeds, most of which were prescaled during some periods of data taking. The lowest- p_T seed that was never prescaled is `L1_SingleIsoEG34er` (although `L1_SingleIsoEG32er` was unprescaled in the full data set but few pb⁻¹). The matching to the L1T seed guarantees that the sample of selected events is not biased in terms of prescales of L1T seeds with lower thresholds [136]. The other set of requirements defines a “probe” electron. It is subject to a looser selection, requiring $p_T > 25$ GeV and $|\eta_{\text{SC}}| < 2.5$ and asking for the tight working point of the identification algorithm and the additional selection on impact parameters. If a “probe” is additionally matched to one of the considered HLT objects, it is referred to as a “passing probe”. Efficiency of the trigger under study is determined as the probability for a “probe” electron to be also a “passing probe”. It is parameterized with the transverse momentum of the electron and the pseudorapidity of the associated ECAL supercluster, which allows to apply the measurement to electrons used in this search despite the fact that they are defined with a tighter selection on p_T and with the transition region between the barrel and the endcaps excluded.

Trigger efficiency is measured in data as well as in simulation of the Drell-Yan process using events selected by trigger `HLT_Ele27_WPTight_Gsf_v*` and containing exactly two “probe” electrons with opposite electric charges, at least one of which also satisfies definition of a “tag”. Invariant mass of the pair of electrons must satisfy $60 < m_{e^+e^-} < 120$ GeV. The tag-and-probe method allows to account for the presence of background with non-prompt leptons by fitting the $m_{e^+e^-}$ spectrum, taking an advantage of the non-peaking distributions for the background. The fit, however, is not required in this case because the selected region is already very pure in $Z \rightarrow e^+e^-$ events. The trigger efficiency is computed simply as

$$\epsilon = \frac{N_{\text{P}}}{N_{\text{P}} + N_{\text{E}}}, \quad (6.5)$$

where N_p (N_F) is the number of tag–probe pairs with a passing (failing) “probe”. It should be noted that if both “probes” in an event are also “tags”, this single event gives two tag–probe pairs, in which the two electrons are swapped. Example distributions of m_{ee} are shown in Fig. 6.3. Figure 6.4 shows one-dimensional trigger efficiencies as functions of different observables. Compared to simulation, certain degradation of the efficiency is observed for low- p_T electrons as well as in the forward region. In part, this is an effect of imperfect transparency corrections for ECAL crystals. A moderate dependence of the scale factors on the amount of pile-up is visible, which, however, will be absorbed in the final scale correction because simulation is reweighted to

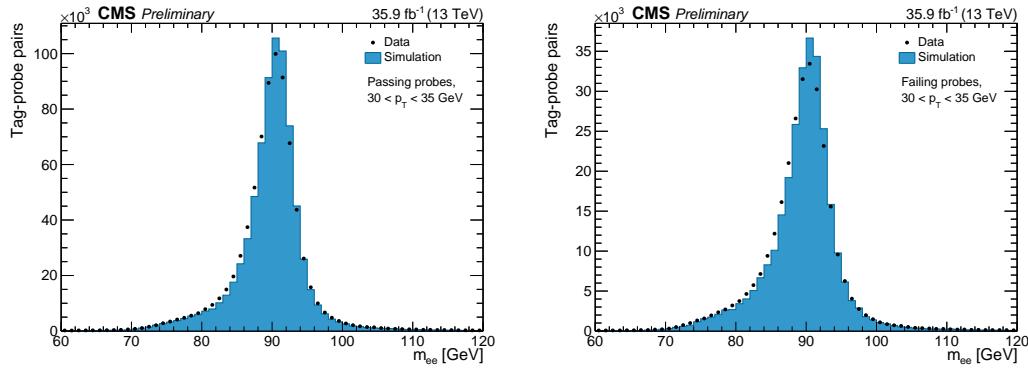


Figure 6.3: Example distributions of m_{ee} in tag–probe pairs with passing (left) and failing (right) “probes”. Shown for $30 < p_T^{probe} < 35 \text{ GeV}$.

match the pile-up profile observed in data. The dependence on the number of jets in the event will be covered by an additional uncertainty.

Two-dimensional scale factors used in this search are shown in Fig. 6.5, together with their full uncertainties. Main components of the uncertainties are shown in Fig. 6.6 and represent the statistical uncertainty (in both data and simulation) and two systematic shifts. The first systematic uncertainty accounts for the change in the scale factors when the mass window is tightened to $70 < m_{ee} < 110 \text{ GeV}$. The second one is induced by increasing the p_T threshold in the definition of the “tag” to $p_T > 45 \text{ GeV}$. These three sources are summed up in quadrature, together with a 2% uncertainty to cover for the dependence on jet multiplicity and an additional 1% uncertainty for unaccounted effects.

Muon trigger efficiency: An inclusive *or* trigger of `HLT_IsoMu24_v*` and `HLT_IsoTkMu24_v*` triggers are used for muon selection in this analysis. The efficiencies scale factors are provided centrally as a function of p_T and η of the muons using the tag-and-probe method as discussed in the above Sec. 6.2.4. A detail study is given in the presentation here [137] while Fig. 6.7 shows the trigger efficiencies as a function of p_T and η , left to right.

6.2.5 Muon and electron efficiency scale factors

Electrons and muons efficiencies scale factors for all other aspects of reconstruction are officially provided, except of the electron trigger scale factors discussed above. All these measurements have also been performed with the tag-and-probe method.

Corrections for muons are taken from Ref. [139]. They include data-to-simulation scale factors for efficiencies of inclusive *or* of triggers `HLT_IsoMu24_v*` and `HLT_IsoTkMu24_v*`, track reconstruction, tight working point of the identification algorithm, and tight selection on isolation. The correction for the efficiency of track reconstruction depends on the pseudorapidity of the muon, while all other scale factors are parameterized with p_T and $|\eta|$ of the muon. Whenever the scale factors are provided separately for different data-taking periods, they are combined assigning each measurement a weight that is proportional to the respective integrated luminosity. Plot 6.8(b) is corrected by muon ID and isolation scale factors and in plot 6.8(a) they are not applied.

In the electron channel this search makes use of scale factors for track reconstruction efficiency and the efficiency of the tight working point of the cut-based identification algorithm. They are provided in Ref. [140] as functions of η_{SC} and (p_T, η_{SC}) respectively. Although the scale factors for electron identification were computed with no selection on impact parameters, it has been demonstrated [141] that they are applicable also for electrons defined as in Sec. 6.3.

6.2.6 Generator level weights

Simulation processes are scaled according to the integrated luminosity using the following formula.

$$N = \mathcal{L} * \sigma \quad (6.6)$$

where \mathcal{L} is the integrated luminosity, σ is the theoretical prediction N is the generated number of events.

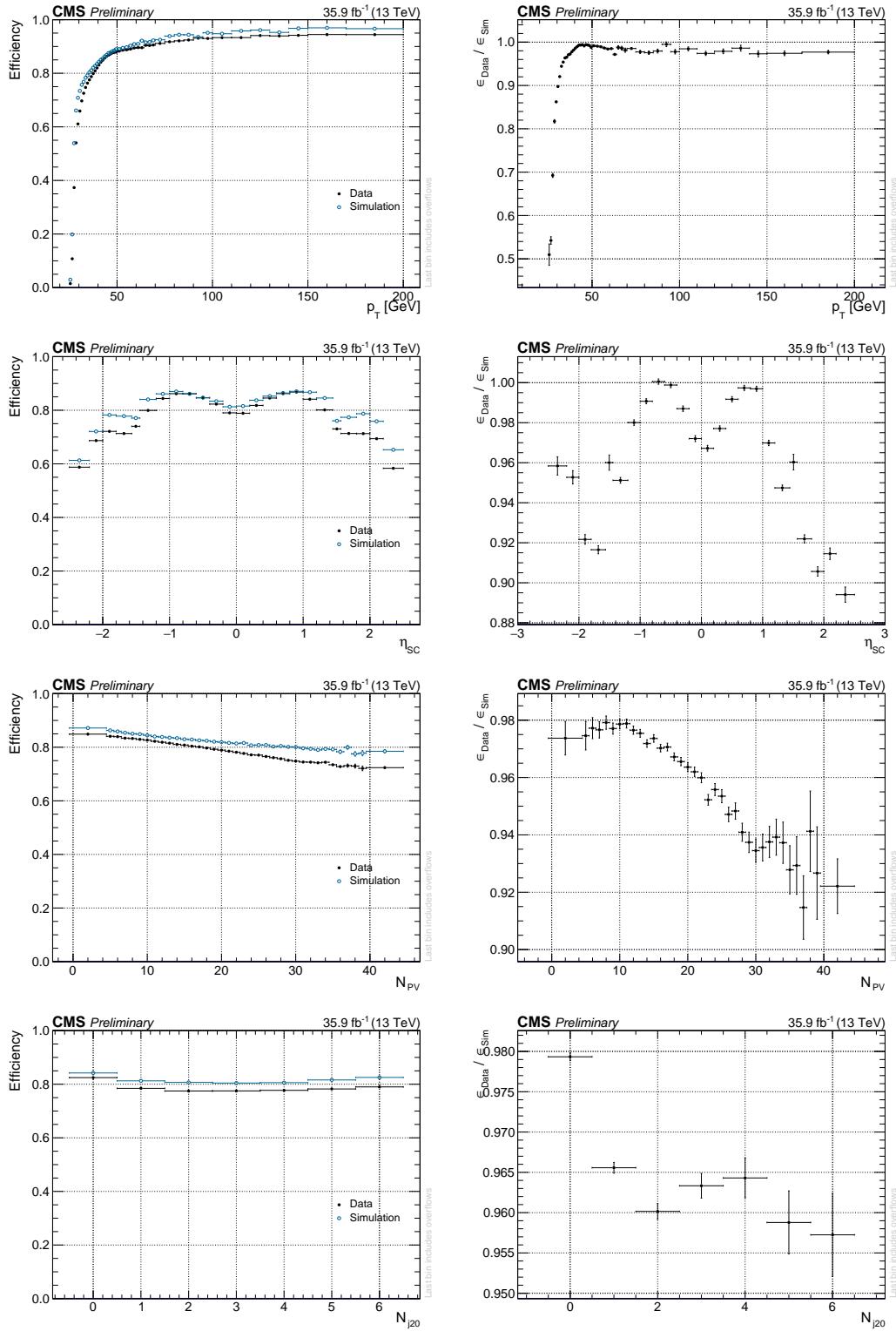


Figure 6.4: Trigger efficiencies in data and Drell-Yan simulation and their ratios as a function of p_T of the “probe”, pseudorapidity of the associated ECAL supercluster, number of reconstructed primary vertices, and number of jets with $p_T > 20$ GeV. Statistical uncertainties are shown.

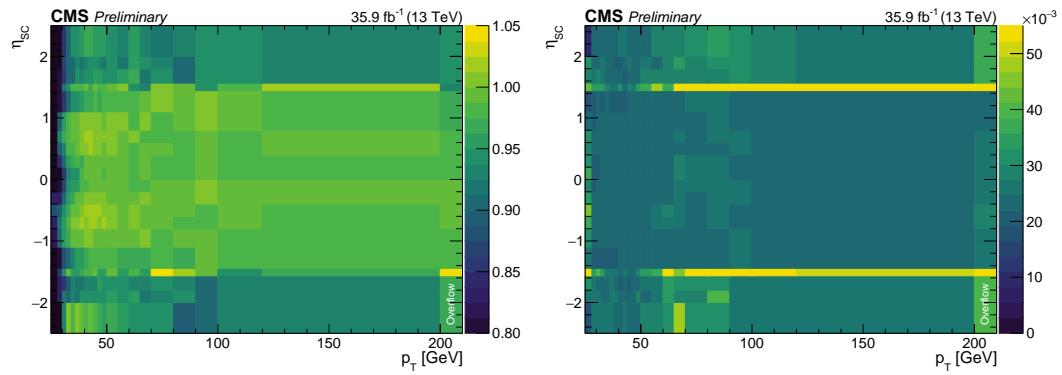


Figure 6.5: Trigger scale factors as a function of p_T and η_{SC} (left) and their full uncertainties (right).

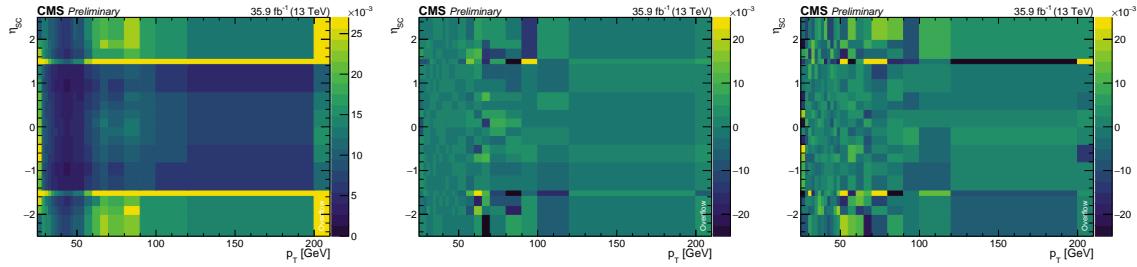


Figure 6.6: Components of uncertainties of the two-dimensional trigger scale factors: statistical uncertainty (left), absolute difference due to the variation of the m_{ee} window (middle) and the p_T of the “tag” (right).

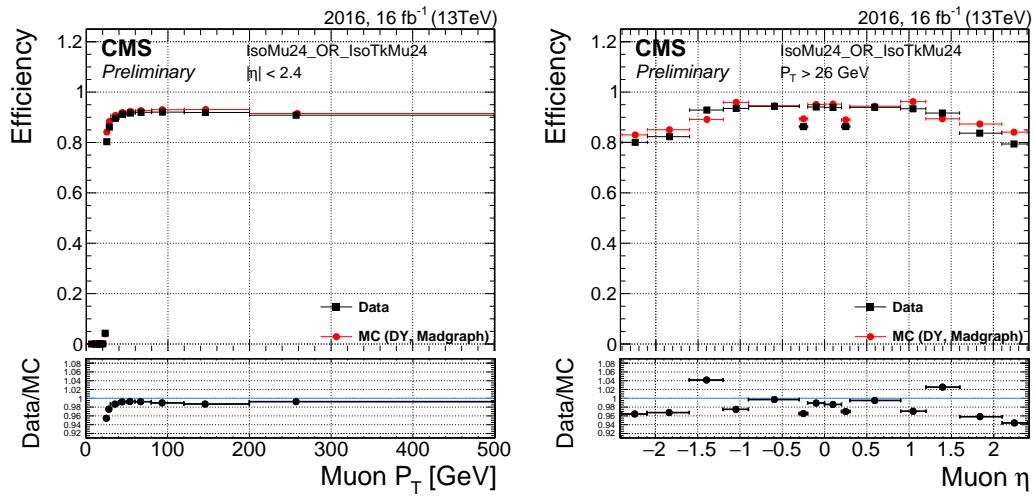


Figure 6.7: Efficiencies of the muon trigger with OR combination of triggers $HLT_IsoMu24_v*$ and $HLT_IsoTkMu24_v*$: as a function of muon p_T , left and η on right [138].

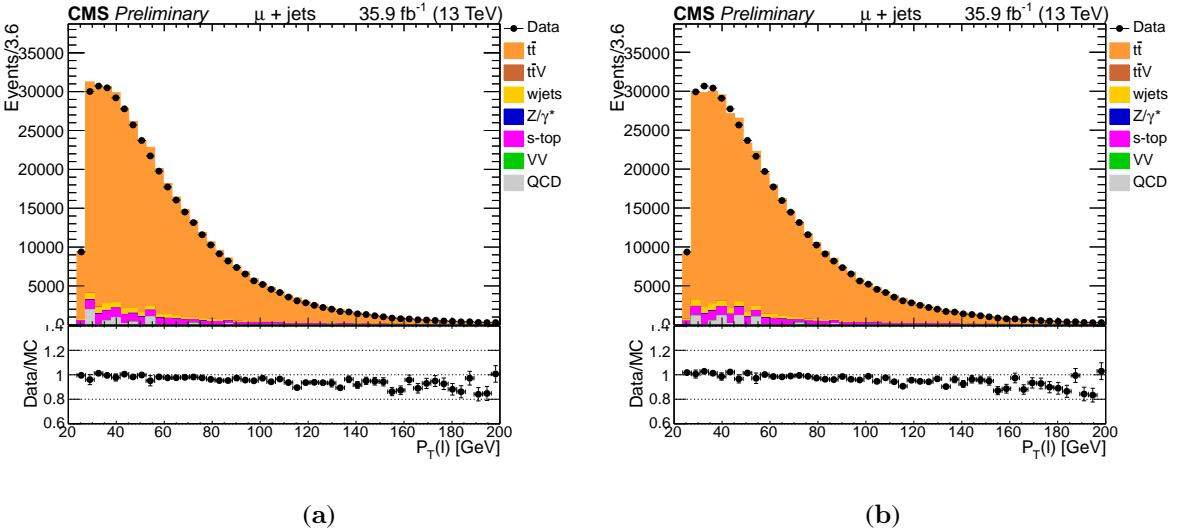


Figure 6.8: Transverse momentum of muon in data and MC pre (a) and post (b) ID and Isolation correction in the $t\bar{t}$ dominated region.

6.2.7 Correction of p_T spectrum of top quarks in $t\bar{t}$ events

It is known that available event generators do not describe distributions of transverse momenta of top quarks in the SM $t\bar{t}$ production [142–145]. In particular, the combination of POWHEG V2 and PYTHIA 8 used in this search predicts an overall harder spectrum than observed in data. The origin of this discrepancy is not understood, and uncovering it will likely require a substantial effort from the theory community. At the same time it has a significant impact on the majority of analyses involving SM $t\bar{t}$ and affects p_T distributions of various reconstructed objects, not only the top quarks but also their decay products.

In order to correct for the discrepancy, an empirical reweighting based on the observed p_T distributions of top quarks [146] is utilized. It is only applied to the SM $t\bar{t}$ production. Event weights are evaluated as a function of the transverse momenta of the last top quarks found in the event history (i.e. after the modelling of the final state radiation):

$$w(p_T^{(1)}, p_T^{(2)}) = \sqrt{\prod_{i=1}^2 \exp(p_0 + p_1 \cdot p_T^{(i)})} = \exp\left(p_0 + p_1 \cdot \frac{p_T^{(1)} + p_T^{(2)}}{2}\right). \quad (6.7)$$

Parameters of the reweighting are computed as

$$\begin{aligned} p_0 &= 6.15024 \times 10^{-2} + 3.243 \times 10^{-2} \cdot \nu_1 - 4.353 \times 10^{-7} \cdot \nu_2 \\ p_1 &= -5.17833 \times 10^{-4} - 1.404 \times 10^{-4} \cdot \nu_1 - 1.005 \times 10^{-4} \cdot \nu_2, \end{aligned} \quad (6.8)$$

where ν_1 and ν_2 are independent nuisance parameters distributed according to the standard normal distribution $\mathcal{N}(0, 1)$. These nuisance parameters reflect uncertainties of the fit of the unfolded top quark’s p_T spectrum. Factors in front of $\nu_{1,2}$ have been obtained by finding a linear transformation that transforms the covariance matrix of the fit [147] into an identity one. Nominal weights are computed by setting $(\nu_1, \nu_2) = (0, 0)$, while two pairs of independent systematic variations are given by values $(\pm 1, 0)$ and $(0, \pm 1)$.

The empirical reweighting has been constructed from the normalized differential cross section $1/\sigma \cdot d\sigma/dp_T$ and as such should only change the shape of the distribution but not the overall yield. To enforce this condition, weights given by Eq. 6.7 are additionally divided by their mean values computed over the full $t\bar{t}$ dataset with no event selection. These values are reported in Table 6.4 for all considered variations.

The effect of the reweighting is demonstrated in Fig. 6.9, which shows distributions of the p_T of the reconstructed top quarks, the mass of the $t\bar{t}$ system, and the angular observable used in this analysis, $\cos\theta_{t_l}^*$, in the signal region. Definition of the signal region and the reconstruction of the $t\bar{t}$ system will be described in dedicated sections below. The results are compared against a reweighting to differential distributions computed with NNLO QCD + NLO EW precision [148]. The corresponding event weights have been obtained as the

Nuisances (ν_1, ν_2)	Mean weight
(0, 0)	0.9985
(+1, 0)	1.0142
(-1, 0)	0.9832
(0, +1)	0.9865
(0, -1)	1.0107

Table 6.4: Mean values of weights given by Eqs. 6.7, 6.8 with no event selection.

ratio between the theoretical differential distributions and distributions of the last top quarks in the event history in the used $t\bar{t}$ datasets before any event selection, both normalized to the same cross section. Theoretical results are not available for the angular variable. As can be seen from the figure, both reweightings improve the modelling of the top quarks' p_T distributions, while for $m_{t\bar{t}}$ using the theoretical distribution leads to a significantly worse description. This does not necessarily indicate a problem in the theoretical computation because the matching between definitions of top quarks in the POWHEG + PYTHIA sample and the stable top quarks from Ref. [148] is not a well-posed problem, and using different definitions of top quarks might affect the distributions. The empirical reweighting, on the other hand, gives a sufficiently good description of all the four observables and, as a direct result of the change in the distributions of top quarks' p_T , also improves modelling of the transverse momenta of leptons, jets, and p_T^{miss} . Figure 6.9 shows a remaining discrepancy in the low-energy region, but, at least in part, it can be attributed to an underestimation of the multijet background. It is described with a data-driven method, as will be discussed in Sec. 6.6, and a conservative uncertainty of $+100/-50\%$ is assigned for its normalization. The impact of this uncertainty is also shown in the figure, and it covers for a part of the discrepancy. It should be noted that if the mismodelling of top quark's p_T is caused by unknown BSM effects, the empirical reweighting would hide them from the analysis. However, these are not the kind of BSM effects that are targeted by this search. Indeed, the sought-for signal would manifest itself with a peak-dip structure in the $m_{t\bar{t}}$ spectrum (including the peak-only and dip-only cases as extremes), while the reweighting implements mostly a change of the slope of the data-to-expectation ratio. As such, it cannot introduce nor absorb features in the $m_{t\bar{t}}$ distribution. Corrections to top p_T based on pure theoretical values and its uncertainties given in Ref. [148] are also plotted in Fig. 6.10. In Fig. 6.10 the leptonic top p_T is plotted and in lower pad, a: the PDF uncertainty band and b: the $\mu_{F,R}$ variation from 0 to ± 2 are shown. The band shows normalization effect and doesn't cover the whole space which is the reason for using an alternative method from data as described above.

6.3 Physics objects selection

For physics objects selection, we require first primary vertex (the one with the greatest sum of p_T^2 of associated physics objects) to be reconstructed in each event from at least four tracks and must not marked as “fake” by the reconstruction algorithm. It should also lie within a cylinder of radius 2 cm around the nominal beam axis, and its z coordinate must satisfy $|z| < 24$ cm, where $z = 0$ cm corresponds to the geometric centre of the CMS detector.

Vertices that satisfy the above requirements are exploited to mitigate the impact from additional pp interactions within the same bunch crossing (the pile-up) according to the so-called Charged Hadron Subtraction (CHS) scheme. If a PF candidate is identified as a charged hadron and it is associated to any but the first vertex, the candidate is removed from the event.

Muons

The analysis uses a PF candidate which is reconstructed outside-in as a Global Muon satisfying the tight working point of the identification algorithm [149]. The conditions for this selection requires a goodness-of-fit of its track with $\chi^2/\text{n. d. f.} < 10$ and at least one muon chamber hit included in the global-muon track fit. Muon segments in at least two muon stations must be associated with the candidate. The tracker track's transverse impact parameter with respect to the first primary vertex must satisfy $|d_{xy}| < 2$ mm, and the z position of the muon vertex must lie within 5 mm around the z position of the primary vertex. At least one hit in the pixel system and at least six tracker layers with measurements are required.

In addition to the above identification criteria, the muon must have $p_T > 26$ GeV and $|\eta| < 2.4$ and must

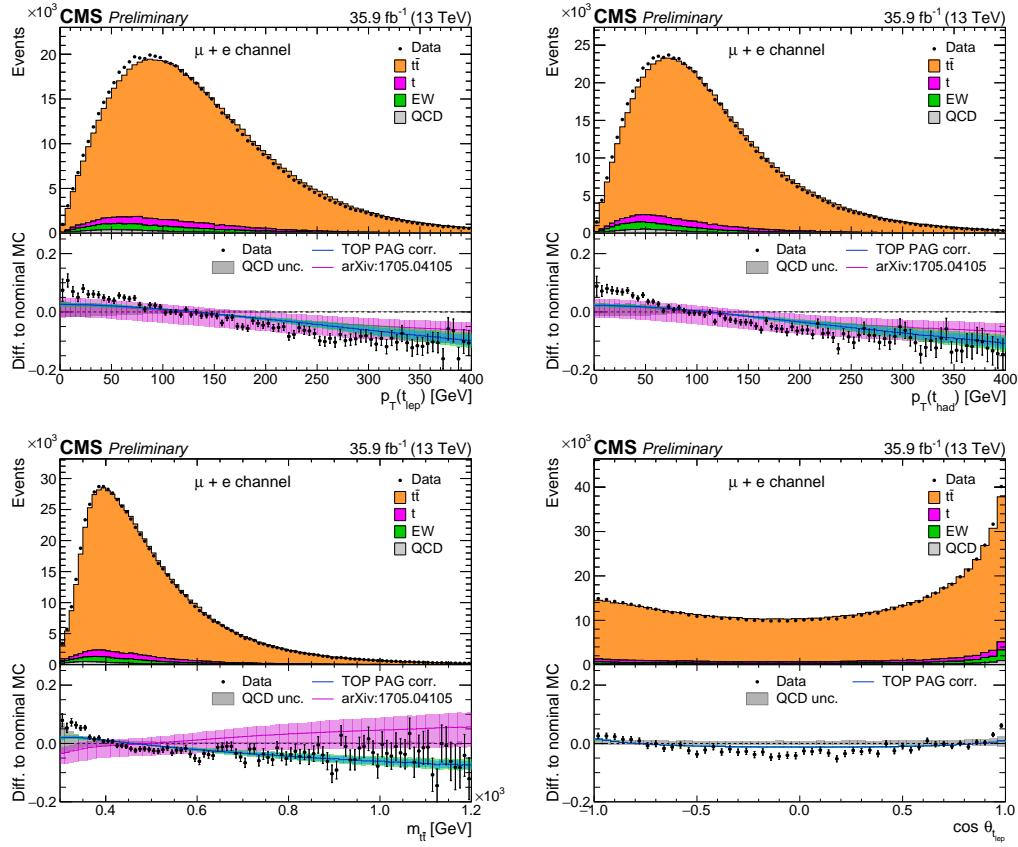


Figure 6.9: Effect of the reweighting of SM $t\bar{t}$. Shown for p_T of the semileptonically and hadronically decaying top quarks (upper row), mass of the $t\bar{t}$ system (lower left figure), and the angular observable $\cos \theta_{t_1}^$ (lower right figure). Simulation of $t\bar{t}$ production in the stacked plots in the upper parts of the figures does not have the reweighting applied. The lower parts show relative differences from the full expectation without the $t\bar{t}$ reweighting. The empirical reweighting discussed in the text is shown with (largely overlapping) blue and green bands, which correspond to the two systematic variations. For comparison, also included are the results of reweighting to differential distributions in Ref. [148], which are available for top p_T and $m_{t\bar{t}}$ (magenta bands). The grey bands show the effect of $+100/-50\%$ variation of the normalization of the data-driven multijet background.*

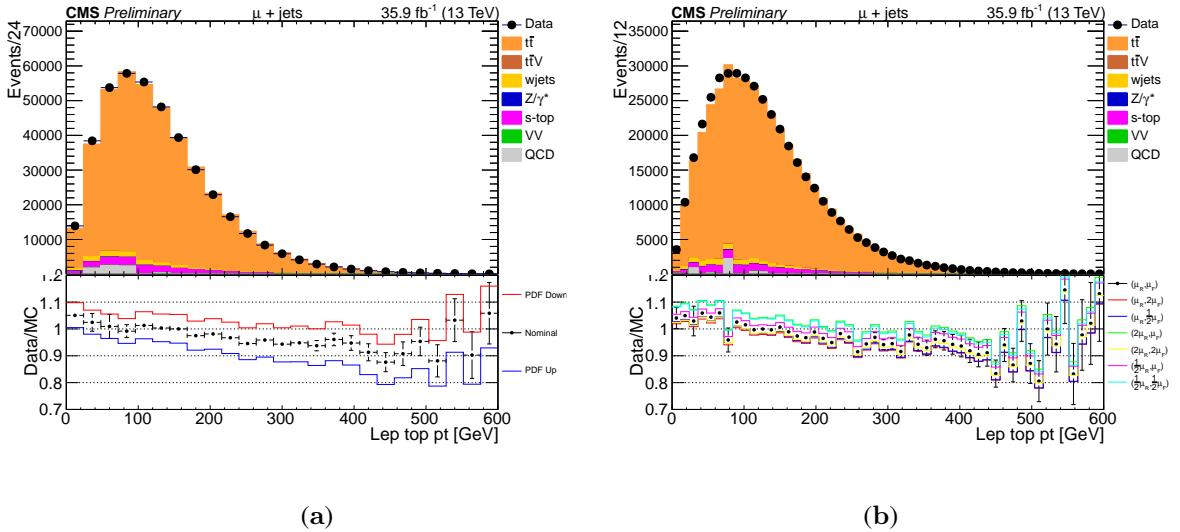


Figure 6.10: Corrections from theory to the leptonic top quark p_T , lower pad a: PDF uncertainty band and b: the $\mu_{F,R}$ variation from 0 to ± 2

be isolated satisfying $I_{\Delta\beta} < 0.15$, where the relative $\Delta\beta$ -corrected isolation is defined as

$$I_{\Delta\beta} = \frac{1}{p_T} \left(\mathcal{I}_{h^\pm} + \max \left(\mathcal{I}_{h^0} + \mathcal{I}_\gamma - 0.5 \cdot \mathcal{I}_{h^\pm}^{\text{PU}}, 0 \right) \right). \quad (6.9)$$

Here, values \mathcal{I} in the numerator are sums of the transverse momenta of charged hadron, neutral hadron, and photon PF candidates as well as charged hadron pile-up PF candidates inside a cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ around the muon, and p_T is the transverse momentum of the muon. In $\mathcal{I}_{h^\pm}^{\text{PU}}$ term, the contribution of neutral particles from pile-up interactions has been subtracted; the specific form of this correction can be either based on the PF algorithm ($\Delta\beta$ -correction) or on the average energy density measured in the event (effective-area correction). The kinematic selection is driven by exploited triggers, while isolation is defined as recommended in Ref. [149].

A muon satisfying the above requirements is referred to as a “tight” muon. For the purpose of event selection, “loose” muons are also defined. They are required to pass the loose working point of the identification algorithm [149], which is to be muon PF candidates that are at the same time global or tracker muons, and to satisfy $p_T > 10 \text{ GeV}$, $|\eta| < 2.5$, and $I_{\Delta\beta} < 0.25$.

Electron

Similarly to the case of muons, “tight” and “loose” electrons are exploited. They are defined according to the tight and the veto working points of the 2016 identification algorithm, respectively. In case of “tight” electrons also the recommended selection on transverse and longitudinal impact parameters, namely $|d_{xy}| < 0.5$ (1) mm and $|d_z| < 1$ (2) mm in the ECAL barrel (endcaps), is applied.

In addition, a “tight” electron must have $p_T > 30 \text{ GeV}$ and $|\eta_{\text{SC}}| < 2.5$, excluding the transition region between the barrel and the endcap of the ECAL, $1.4442 < |\eta_{\text{SC}}| < 1.5660$, where η_{SC} is the pseudorapidity of the associated ECAL supercluster, computed with respect to the geometrical centre of the CMS detector.

“Loose” electrons are only required to have $p_T > 20 \text{ GeV}$ and $|\eta_{\text{SC}}| < 2.5$ in addition to the criteria of the identification algorithm.

Jets

Jets are clustered using the anti- k_T algorithm [99] with a cone size of 0.4. Pile-up charged hadrons are removed from the source collection for clustering, as described above. Reconstructed jets must satisfy the loose working point of the PF jet identification algorithm [150], which applies selection on the number of charged and neutral PF candidates clustered in the jet and fractions of the total energy carried by different types of constituent PF candidates.

In order to account for non-linearity of the calorimeter response, impact of minimum energy thresholds, and non-perfect modelling of the detector, several corrections are applied to measured jet momenta [151, 152]. First, contribution to jet energy from pile-up is subtracted (on the average) using the L1FastJet corrections. Then the L2L3 corrections are applied in both data and simulation. They are derived in simulation and correct jet momentum to match transverse momentum of the corresponding particle-level jet. Finally, jets in data are subject to the L2L3Residual corrections, which account for mismodelling of the detector response. The version of jet energy corrections (JEC) used in this search is Summer16_23Sep2016ReReco_V4.

Jet energy resolution (JER) in data is known to be worse than predicted by simulation [151]. In order to account for this difference, jets in simulation are smeared following the recommended “hybrid” approach [153]. Namely, momenta of reconstructed jets that have matching particle-level jets are scaled with factors

$$c_{\text{JER}} = 1 + (s_{\text{JER}} - 1) \frac{p_T - p_T^{\text{ptcl}}}{p_T}, \quad (6.10)$$

where p_T and p_T^{ptcl} are transverse momenta of the reconstructed jet and the corresponding particle-level jet, and s_{JER} is the JER scale factor provided in Refs. [153, 154]. The matching particle-level jet is required to satisfy $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.2$, which is the half of the jet cone size, and $|p_T - p_T^{\text{ptcl}}| < 3 \sigma_{\text{JER}} p_T$, where σ_{JER} is the relative p_T resolution determined from simulation [155]. In rare cases when no particle-level jet is matched,

jet momentum is rescaled using a stochastic factor

$$c_{\text{JER}} = 1 + \mathcal{N}(0, \sigma_{\text{JER}}^2) \sqrt{\max(s_{\text{JER}}^2 - 1, 0)}, \quad (6.11)$$

where $\mathcal{N}(0, \sigma^2)$ denotes a random number sampled from a normal distribution with a zero mean and variance σ^2 . Jet energy resolution and scale factors of version Spring16_25nsV10 are utilized.

Jets with $p_T > 20$ GeV and $|\eta| < 2.4$ are considered in this analysis. Since jets are clustered from all non-pile-up PF candidates, leptons defined as described above can also be reconstructed as jets. To prevent the double counting, jets overlapping with a “loose” muon or electron within $\Delta R < 0.4$ are removed from the event.

As the signal signature involves b quarks, the b-tagging capability of the detector is exploited in the analysis. Jets produced by b quarks are identified with the help of the updated combined MVA (cMVAv2) algorithm [129], which combines several individual algorithms. Indirectly, it makes use of impact parameters of individual tracks in the jet, properties of reconstructed secondary vertices and also soft muons and electrons originating from decays of B hadrons, and other observables. The medium working point of the algorithm is utilized, which corresponds to the b-tagging discriminator value larger than 0.4432. Selection efficiency for b quark (light-flavour) jets is about 65% (1%).

Missing transverse momentum

Reconstruction of neutrino from the decay $t \rightarrow bl\nu$ exploits missing transverse momentum

$$\vec{p}_T^{miss} = - \sum_i \vec{p}_T(i) - \sum_j (\vec{p}_T(j) - \vec{p}_T^{\text{L1}}(j)). \quad (6.12)$$

The first sum runs over all PF candidates, including those that have been identified by the CHS procedure. The second sum represents the so-called type-1 correction [156] implemented following the L1L2L3 – L1FastJet scheme. Here $\vec{p}_T(j)$ is the fully corrected transverse momentum of jet j whereas $\vec{p}_T^{\text{L1}}(j)$ denotes its momentum with only the L1FastJet correction applied, and the sum includes all jets with $p_T > 15$ GeV that do not overlap with “loose” muons or electrons. JER smearing is not applied in this computation.

Two additional corrections are applied to \vec{p}_T^{miss} in data but not in simulation. A significant fraction of the 2016 data set was affected by the dynamic tracker inefficiency [157]. Its impact was mitigated in reconstruction, but as a side effect duplicate and fake muons were created in some events. These spurious muons are propagated into \vec{p}_T^{miss} and can thus produce an artificial tail at large values of \vec{p}_T^{miss} [158]. To correct for this bias, duplicate and fake muons are identified with dedicated filters [159] and removed from the collection of PF candidates that is used to compute \vec{p}_T^{miss} . This is done centrally when MiniAOD data are produced [160]. The second correction addresses a mismeasurement of ECAL energy deposition for high- p_T electrons and photons due to the so-called ECAL gain switch effect [161]. The correction that is applied to momenta of electrons and photons, is also propagated into \vec{p}_T^{miss} .

6.4 Event selection

This search targets semileptonic decays $\Phi \rightarrow t\bar{t} \rightarrow l 2b 2q$, where l denotes a muon or an electron. Decays involving a τ lepton are not considered explicitly. Final states with a muon or an electron are referred to as the muon and electron channels respectively.

In the muon channel the analysis is performed with data collected using an inclusive *or* of triggers `HLT_IsoMu24_v*` and `HLT_IsoTkMu24_v*`, while in the electron channel trigger `HLT_Ele27_WPTight_Gsf_v*` is exploited. These triggers were unprescaled throughout 2016 data taking. Simulated events are subjected to an emulation of the same triggers.

After the trigger selection, an event in the muon (electron) channel is required to contain exactly one “tight” muon (electron), as defined in Sec. 6.3. In order to suppress the contribution from Drell–Yan and other processes in which multiple prompt leptons are produced, an event is rejected if an additional “loose” muon or electron is found, regardless of the channel.

Since decay products of the heavy Higgs boson include two b quarks and two lighter quarks, an event must contain at least four jets, and at least two of them are required to be b-tagged. The jets are defined as described

μ (e) with $p_T > 26$ (30) GeV
no additional loose muons or electrons
≥ 4 jets with $p_T > 20$ GeV, $ \eta < 2.4$
≥ 2 b-tagged jets
$m_T^W > 50$ GeV

Table 6.5: Summary of the event selection.

Process	Muon channel	Electron channel
Data	483359	326346
$t\bar{t}$	$(435_{-25}^{+26}) 10^3$	$(288 \pm 17) 10^3$
Single top	$(24_{-2}^{+3}) 10^3$	$(16.1_{-1.7}^{+1.9}) 10^3$
W	$(17_{-6}^{+9}) 10^3$	$(11_{-4}^{+5}) 10^3$
Z/γ^*	$(2.2_{-0.7}^{+1.1}) 10^3$	$(2.1_{-0.7}^{+1.0}) 10^3$
VV	$(0.6_{-0.2}^{+0.3}) 10^3$	360_{-120}^{+180}
$t\bar{t}V$	$(1.2 \pm 0.3) 10^3$	810_{-190}^{+250}
Multijet	$(7_{-4}^{+8}) 10^3$	$(11_{-6}^{+12}) 10^3$
Total background	$(491_{-27}^{+28}) 10^3$	$(332_{-19}^{+22}) 10^3$
Signal (A, 500 GeV, 10%)	-570_{-370}^{+330}	-450_{-270}^{+240}

Table 6.6: Observed and expected event yields. Here V denotes a W or a Z boson. Numbers for backgrounds include only statistical uncertainties and uncertainties assigned for normalization of individual processes that will be described in Sec. 6.7. The entry for an example signal benchmark is the sum of contributions from the resonant production and the interference for a \mathcal{CP} -even state with mass 500 GeV, relative total width 10% of mass, and $g_{Att} = 1$. For signal, the statistical uncertainty and the combined theory uncertainty described in Sec. 6.7 are included.

in Sec. 6.3. The relatively low p_T threshold of 20 GeV is motivated by the spectrum of the subleading non-b quark from the hadronically decaying top quark.

To further suppress the QCD multijet background, only events with $m_T^W > 50$ GeV are selected. The variable m_T^W , which has a characteristic distribution in events containing a leptonically decaying W boson, is defined as

$$m_T^W = \sqrt{2p_T^l p_T^{miss} (1 - \cos \Delta\phi(\vec{p}_T^l, \vec{p}_T^{miss}))}, \quad (6.13)$$

where \vec{p}_T^l is the transverse momentum of the only “tight” muon or electron in the event.

Finally, dedicated filters are applied to reject events with anomalous noise or problems in reconstruction [162], as recommended for analyses that exploit \vec{p}_T^{miss} . They target interactions between beam halo muons and calorimeters, anomalous noise in HCAL, unknown and unrecoverable energy depositions in some cells of ECAL [163], large spontaneous signals from specific supercrystals in EE, poorly reconstructed muons. In addition, an event is rejected if its first primary vertex does not meet quality requirements listed in Sec. 6.3.

The event selection is briefly summarized in Table 6.5. Resulting observed event yields and expectations for background processes and an example signal benchmark are shown in Table 6.6. Combined signal acceptance for targeted decays changes from 7 to 10% approximately, depending on the mass of the heavy Higgs boson. The acceptance is computed for the resonant production and is defined as the ratio between the expected number of selected events, summed over the muon and electron channels, and the total expected number of produced events with the $e +$ jets and $\mu +$ jets final states.

After applying all cuts and corrections, a good data/MC comparison observed in distributions of basic observable, shown in appendix C.

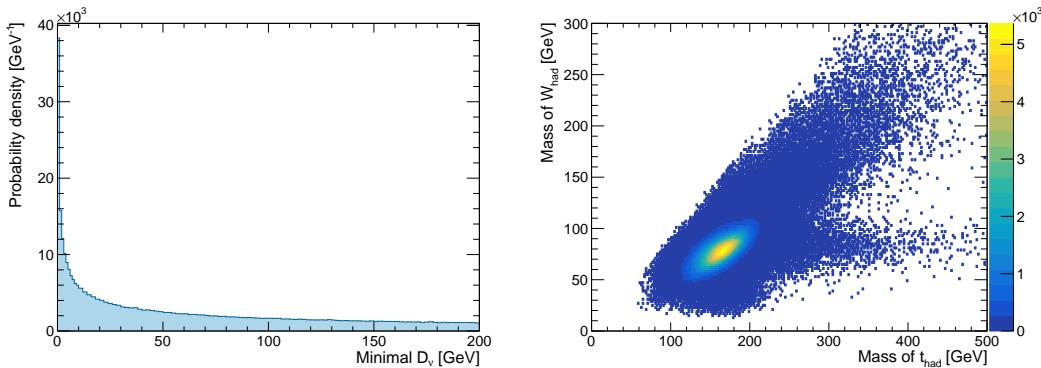


Figure 6.11: Components of the likelihood function for $t\bar{t}$ reconstruction: minimal D_ν (left) and masses involved in the hadronic branch of the decay (right).

6.5 Search variables

This section discusses the two search variables exploited in this analysis. They are defined based on reconstructed top quarks, which approximate the underlying parton-level system.

6.5.1 Reconstruction of the $t\bar{t}$ system

Each event that passes the selection described in the previous section is reconstructed under the assumption that it has been produced in the process $pp \rightarrow t\bar{t} \rightarrow l\nu 2b2q$. This is done using a variant of the algorithm adopted in Ref. [144]. All possible ways to assign four reconstructed jets to the four quarks in the final state are considered. In order to reduce the number of combinations, it is required that each of the two b quarks is associated with a b-tagged jet. Out of the remaining jet assignments, the algorithm chooses the one with the largest value of the likelihood function constructed as described below.

For each considered choice of b_l , the b quark jet stemming from the semileptonically decaying top quark, the neutrino is reconstructed following the procedure introduced in Ref. [15]. The z component of its momentum is not known, and while the transverse component can be approximated by the measured \vec{p}_T^{miss} , the accuracy of such an approximation is severely limited by the resolution of this observable. Because of this, the algorithm treats the whole neutrino's three-momentum as unknown. It is determined by utilizing mass constraints $(p(l) + p(\nu))^2 = m_W^2$ and $(p(l) + p(\nu) + p(b_l))^2 = m_t^2$, which typically result in a one-dimensional space of allowed solutions for $\vec{p}(\nu)$. A unique solution is then chosen by minimizing the Euclidean distance $D_\nu = |\vec{p}_T(\nu) - \vec{p}_T^{miss}|$.

The likelihood function is constructed using $t\bar{t}$ events with the correct jet assignment, which is performed according to the ΔR matching between reconstructed jets and generated partons. It is computed as a product of two components. The first one is the probability density of the minimal D_ν , and the second component is the two-dimensional probability density of reconstructed masses of the top quark and the W boson in the hadronic branch of the $t\bar{t}$ decay. The two probability densities are shown in Fig. 6.11. Among all considered jet assignments, the one with the greatest value of $L(D_\nu) \cdot L(m_t^{\text{reco}}, m_W^{\text{reco}})$ is selected. This jet assignment also fixes momentum of the neutrino, which allows reconstructing the semileptonically decaying top quark.

Under some circumstances the $t\bar{t}$ reconstruction can fail. The most common reason for the failure is the absence of a solution for neutrino's momentum when $(p(l) + p(b_l))^2 > m_t^2$ for all possible choices of b_l . This might point to a largely mismeasured momentum of the jet or indicate that the actual jet from the decay $t \rightarrow bl\nu$ has not been reconstructed or has not passed the b-tagging requirement. More rarely, the reconstruction fails if the likelihood is zero for all considered jet assignments. Events for which the $t\bar{t}$ reconstruction has not succeeded are excluded from the subsequent analysis. For SM $t\bar{t}$ this amounts to about 12% of events, some of which originate from other decays than the targeted $l + \text{jets}$ final state.

The performance of the $t\bar{t}$ reconstruction is studied below using SM $t\bar{t}$ events that pass the event selection. Additionally, only events with the targeted decays are considered, which further rejects about 15% of events. Kinematic requirements applied to jets as well as imperfection of the jet reconstruction and b-tagging limit the maximal achievable performance of the jet assignment algorithm, which is illustrated in Fig. 6.12. In the left panel it shows the probability that a quark from the $t\bar{t}$ decay is matched to a reconstructed jet within $\Delta R < 0.4$. It can be seen that even with a perfect $t\bar{t}$ reconstruction it would only be possible to identify all the four jets

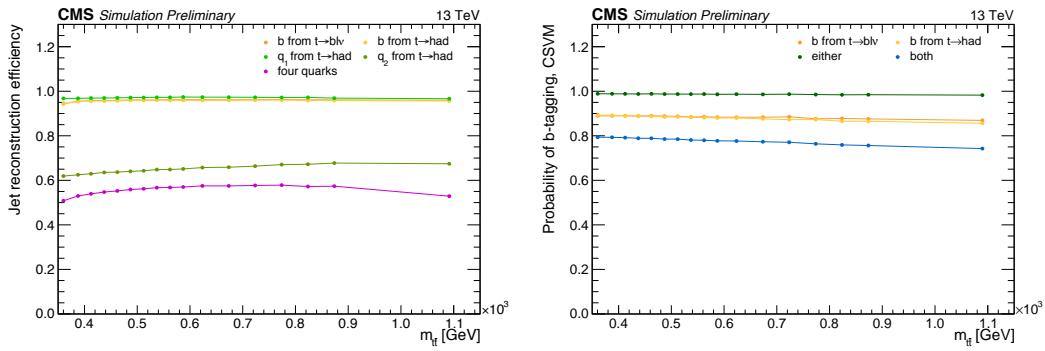


Figure 6.12: Efficiency of reconstruction of jets stemming from the four quarks in the final state of the $t\bar{t}$ decay (left) and the b-tagging probability for jets matched to the two b quarks (right). Shown as a function of parton-level $m_{t\bar{t}}$.

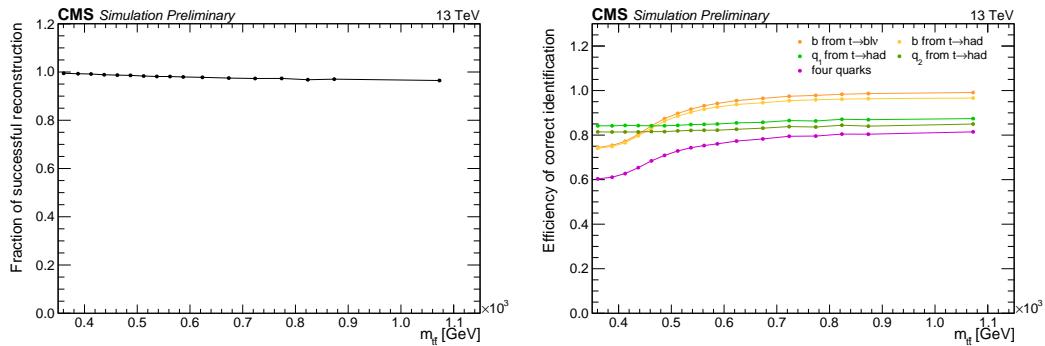


Figure 6.13: Success rate of reconstruction of the $t\bar{t}$ system (left) and efficiencies of correct identification of jets stemming from various quarks in the final state of the $t\bar{t}$ decay (right). Shown as a function of parton-level $m_{t\bar{t}}$. Only fully reconstructable events are considered.

correctly in about 55% of cases. In the remaining events the subleading non-b-quark jet from the decay often does not satisfy the kinematic requirements imposed in this analysis. The right-hand panel shows that even when both b quarks can be matched to reconstructed jets, in about 20% of cases one of these jets does not pass the b-tagging requirement, which can happen, for instance, when one of the two b-tagged jets required by the event selection stems from the hadronization of the c quark from the $W \rightarrow cs$ decay. This further reduces the maximal achievable efficiency of the $t\bar{t}$ reconstruction. In the following events in which each of the four quarks can be matched to reconstructed jets and the jets matched to the b quarks are b-tagged are referred to as fully reconstructable events.

The performance of the actual $t\bar{t}$ reconstruction algorithm is quantified over the fully reconstructable events, and the results are provided in Fig. 6.13. The algorithm declares the reconstruction successful in more than about 97% of cases, and these events are used to determine the probability of correct assignment of reconstructed jets to the four quarks in the final state. The probability that all the four jets are correctly identified varies from 60 to 80% depending on the value of $m_{t\bar{t}}$. The resolution and the bias of the measured value of $m_{t\bar{t}}$ are shown in Fig. 6.14. They are computed as the standard deviation and the mean of distribution of $(m_{t\bar{t}}^{\text{reco}} - m_{t\bar{t}}^{\text{parton}})/m_{t\bar{t}}^{\text{parton}}$ respectively. The mass resolution is found to be about 14%. It should be noted that if not only fully reconstructable events were considered, the apparent resolution would be worse.

Figure 6.15 shows reconstructed $m_{t\bar{t}}$ distribution in the signal process for two example mass points. Due to the experimental resolution, shapes for hypotheses with $\Gamma = 2.5$ and 5% are very similar.

6.5.2 Search variables

The search for $\Phi \rightarrow t\bar{t}$ is performed based on a joint distribution of two discriminative variables. The first variable is the reconstructed invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$, which serves as a proxy for the mass of the heavy Higgs boson. In the subsequent statistical analysis its distribution is described using the non-uniform binning shown in Fig. 6.16. It has been adjusted so that the bin width is significantly smaller than the $m_{t\bar{t}}$ resolution everywhere except for the edges of the range. The first and the last bins are inclusive.

The second variable is $|\cos \theta^*|$, where θ^* is the angle between the three-momentum of the leptonically

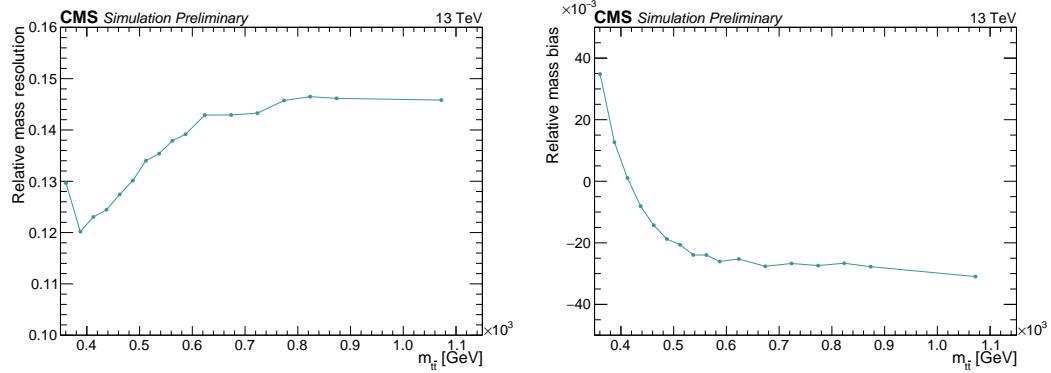


Figure 6.14: Relative $m_{t\bar{t}}$ resolution (left) and bias (right) in fully reconstructable events.

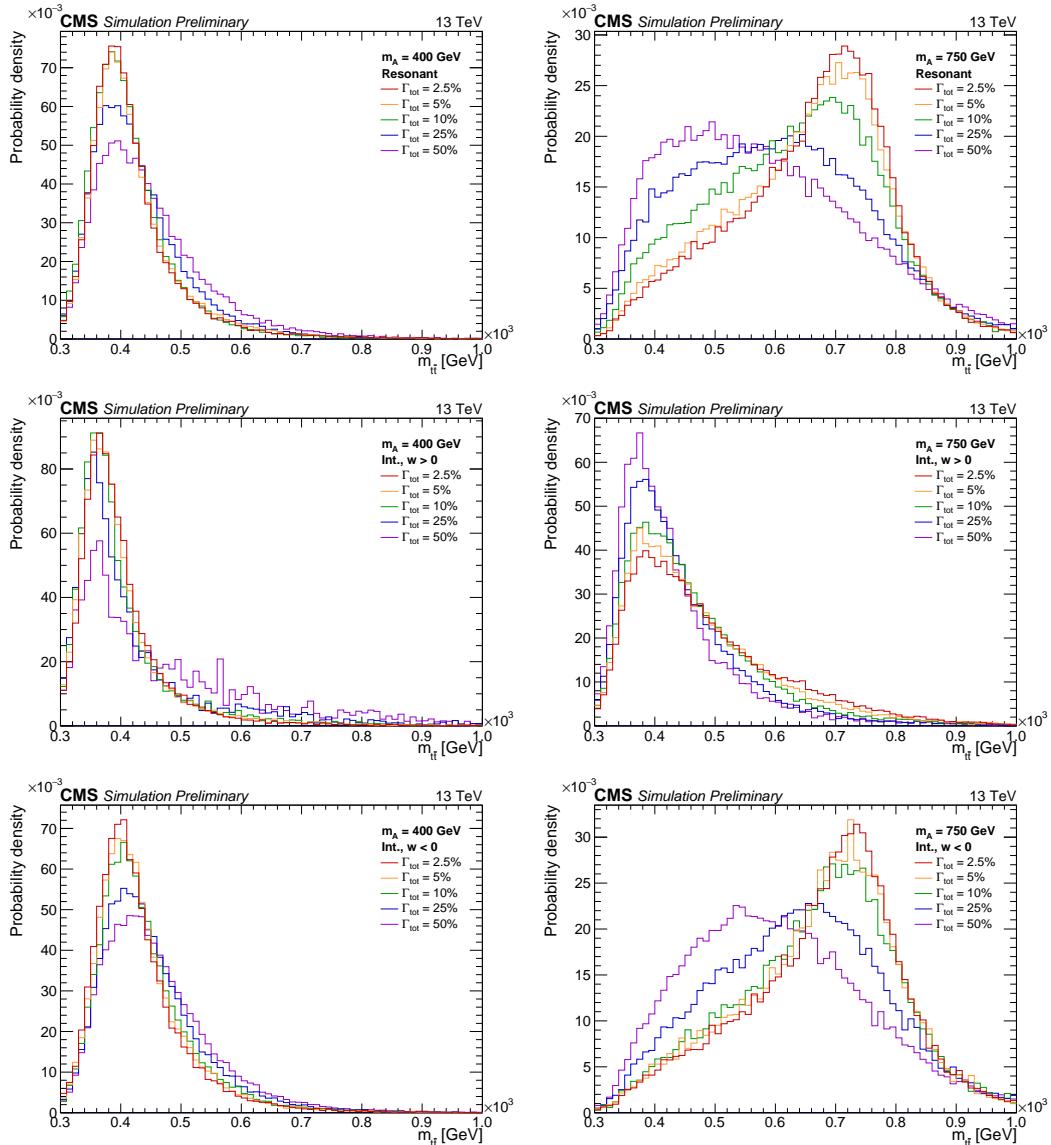


Figure 6.15: Reconstructed $m_{t\bar{t}}$ spectrum in resonant part of signal (top), interference with positive event weights (middle), and interference with negative weights (bottom). \mathcal{CP} -odd states with various total widths, $m = 400$ GeV (left) and 750 GeV (right). Distributions are normalized to unit area.

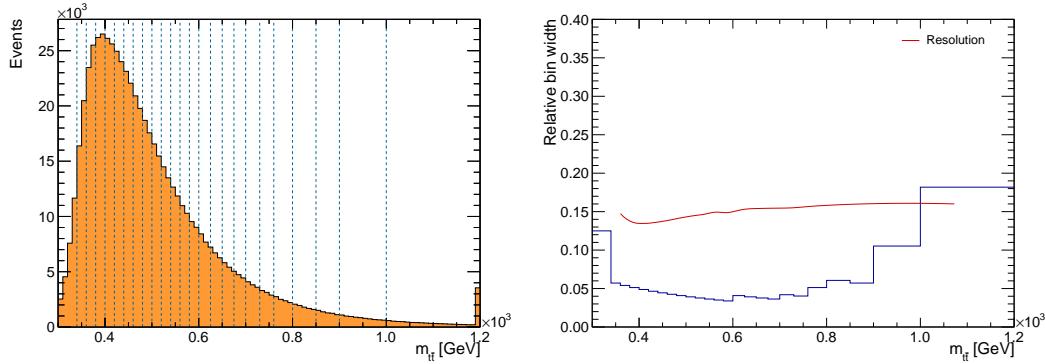


Figure 6.16: Optimized binning in $m_{t\bar{t}}$, superimposed over the distribution of SM $t\bar{t}$, (left) and relative bin width compared to the $m_{t\bar{t}}$ resolution (right).

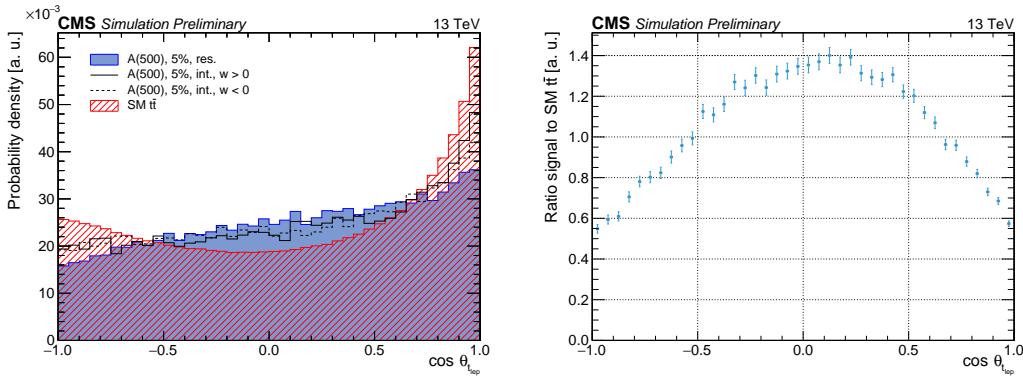


Figure 6.17: Distribution of $\cos \theta^*$ in an example signal process for resonance, positive and negative interference and the SM $t\bar{t}$ background (left) and the ratio between the pure resonance and the SM $t\bar{t}$ background (right).

decaying top quark in the $t\bar{t}$ rest frame and the three-momentum of the $t\bar{t}$ system in the lab frame. In the (resonant part of the) signal process the distribution of this variable reflects the fact that the Higgs boson is a scalar particle and therefore decays into top quarks isotropically. On the other hand, the distribution for the SM $t\bar{t}$ production has a non-trivial shape peaking at $|\cos \theta^*| = 1$ (see Ref. [101] and references therein), which in part is driven by the contribution with the s -channel gluon exchange.

As illustrated in Fig. 6.17, the reconstructed distribution of $\cos \theta^*$ is not symmetric. However, the ratio between distributions of the signal process and the SM $t\bar{t}$ background is approximately symmetric, which allows to consider in the analysis only the absolute value of the observable without a loss of sensitivity. The binning used for the statistical analysis is adjusted to minimize the variation of the ratio within each bin, so that regions over which the ratio changes slowly are covered by wider bins. The following bin edges have been chosen: 0., 0.4, 0.6, 0.75, 0.9, 1.

6.6 Data-driven modelling of the multijet background

One of the main background for semi-leptonic $t\bar{t}$ decays is the multijet QCD that has overwhelming cross section. Due to tight selection criteria, only a small fraction of these events mimics the semi-leptonic final state. This requires large simulation samples to have proper QCD modelling and enough statistics after the final selection that is not practical. We used data-driven technique to model our multijet QCD where we defined sideband regions in data that are enriched in QCD events. The number of events stemming from the multijet background and the distributions of the relevant kinematic variables for final fit are taken directly from these data. The mass of $t\bar{t}$ and $|\cos \theta^*|$ are used as normalization variables. In semi-leptonic decays of $t\bar{t}$, the multijet QCD background can mimic the signal via two main processes.

- Non-prompt and less isolated leptons from the decay of beauty and charm quarks. These are real leptons that gain significant p_T from the mother b and c hadrons and have more activities in the vicinity. Lepton isolation cut suppresses these events.

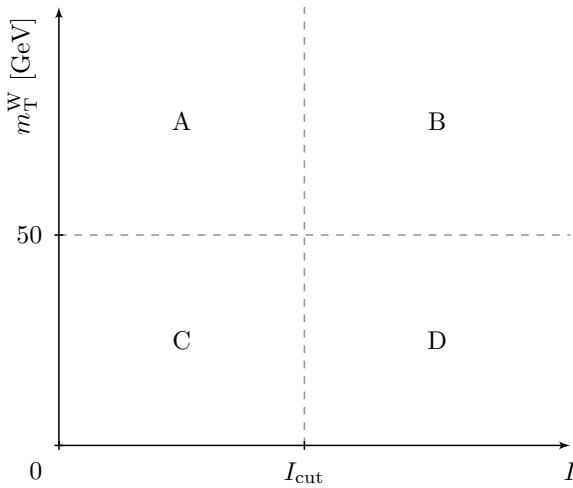


Figure 6.18: Schematic representation of the four regions used for the evaluation of the number of multijet events entering the signal region.

Process	A	B	C	D
Multijet	0.01	0.21	0.06	0.67
$t\bar{t}$	0.92	0.75	0.88	0.31
Single top	0.02	0.02	0.02	< 0.01
W	0.04	0.03	0.04	0.01
Other	< 0.01	< 0.01	< 0.01	< 0.01

Table 6.7: Fractions of events from each background source in the four regions used to evaluate the multijet background normalization. Muon channel.

- Some hadrons didn't absorb in the HCAL and make tracks in the muon system or jets with high electromagnetic fraction mimics electron signatures. They are treated as fake leptons.

6.6.1 Normalization

The number of events from the multijet background is estimated with the so-called ABCD method (see, for instance, a review in Ref. [164]), independently in the muon and electron channels. The method is applied for the M_T^W variable 6.13 and the lepton's relative isolation I , which in case of muons is defined by Eq. 6.9, while for electrons the ρ -corrected isolation included in the identification criterion is exploited. Four regions are defined based on these variables as sketched in Fig. 6.18. Region A is the signal region with an additional requirement of a successful $t\bar{t}$ reconstruction. Region C is built from it by inverting the M_T^W selection: $m_T^W < 50$ GeV. Complementarily, region B is defined by inverting the isolation requirement: $I > 0.15$ for muons and $I > 0.0588$ (0.0571) for electrons in the barrel (endcaps). This inversion, however, requires a change to the logic of the lepton step of the event selection. In the definition of tight leptons in Sec. 6.3 the isolation requirement is removed, and such a lepton with the highest p_T in an event must satisfy the aforementioned lower cut on the isolation. The event must also contain no additional loose leptons according to the standard definition. Jets are removed if they overlap with a loose lepton (of which there can be zero or one in an event, depending on the isolation of the selected lepton). Finally, region D is constructed from region A by inverting both M_T^W and isolation requirements. Since the selections on the isolation and M_T^W have been designed to suppress the multijet background, inversion of any of them increases its contribution. The fractions of events from each background source expected in the different regions are reported in Tables 6.7 and 6.8.

The ABCD method determines the number of events from the multijet background in the signal region, n_A^{QCD} , using an assumption that the two variables are not correlated for this background, which gives

$$n_A^{\text{QCD}} = \frac{n_B^{\text{QCD}} \cdot n_C^{\text{QCD}}}{n_D^{\text{QCD}}}. \quad (6.14)$$

The assumption of no correlation is verified with MC simulation and is demonstrated to hold within statistical

Process	A	B	C	D
Multijet	0.01	0.16	0.04	0.31
$t\bar{t}$	0.92	0.78	0.89	0.64
Single top	0.02	0.02	0.02	0.01
W	0.04	0.04	0.04	0.03
Other	< 0.01	< 0.01	< 0.01	< 0.01

Table 6.8: Fractions of events from each background source in the four regions used to evaluate the multijet background normalization. Electron channel.

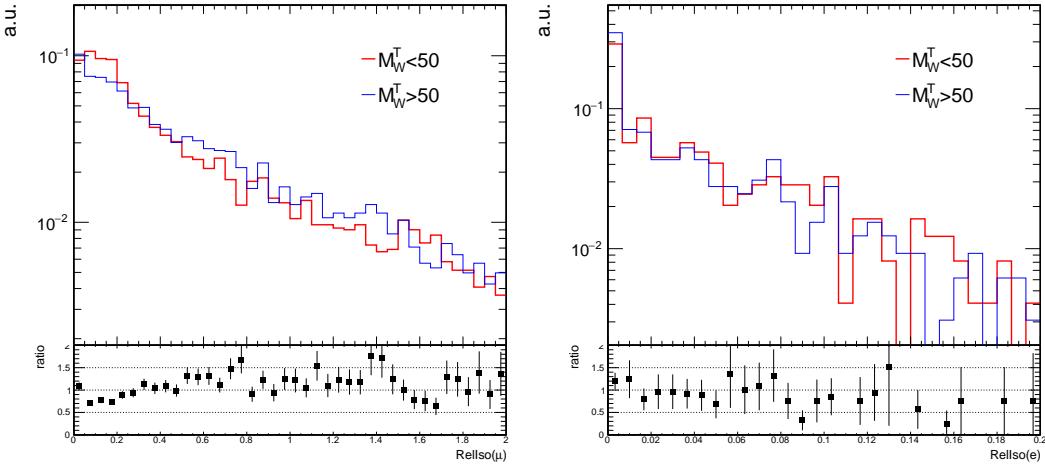


Figure 6.19: Distributions of the lepton's relative isolation, for events satisfying the nominal M_T^W selection and the inverted one, in the muon (left) and electron (right) channels using MC simulation.

uncertainties:

$$\frac{n_A^{\text{QCD}}}{n_C^{\text{QCD}}} = 0.35 \pm 0.13, \quad \frac{n_B^{\text{QCD}}}{n_D^{\text{QCD}}} = 0.25 \pm 0.07 \quad \text{in muon channel},$$

$$\frac{n_A^{\text{QCD}}}{n_C^{\text{QCD}}} = 0.53 \pm 0.22, \quad \frac{n_B^{\text{QCD}}}{n_D^{\text{QCD}}} = 0.93 \pm 0.45 \quad \text{in electron channel}.$$

This conclusion is further supported by Fig. 6.19, which shows that distributions of the relative isolation agree between the two regions in M_T^W .

Figure 6.20 shows distributions of variables I and M_T^W in data and simulation, and they demonstrate a reasonable agreement. The excess of data for large values of the isolation in case of the electron channel is attributed to the way the simulated samples for the multijet background have been produced. The samples are enriched in events with higher electromagnetic activity, and the corresponding generator-level filter (EMEnrichingFilter) includes a loose isolation requirement. This results in a deficit in simulation for large values of the isolation. In Figs. 6.21 and 6.22 distributions of an example observable (the transverse momentum of the hadronically decaying top quark) in the four regions are shown.

The sought-for number of events from the multijet background in the signal region is determined with a maximum-likelihood fit to the number of events observed in each of the four regions. The likelihood is defined as

$$L = \prod_R \frac{\lambda_R^{n_R^{\text{Data}}} e^{-\lambda_R}}{n_R^{\text{Data}}!}, \quad \lambda_R = s \cdot n_R^{\text{Prompt}} + n_R^{\text{QCD}}, \quad (6.15)$$

where n_R^{Data} is the number of events observed in region R and n_R^{Prompt} is the expected number of events from backgrounds with prompt leptons i-e those produced in decays of vector bosons. The overall normalization of backgrounds that produce prompt leptons is allowed to float in the fit, simultaneously in all the four regions, as implemented with the scale factor s . Eq. 6.14 is imposed as a constraint in the fit by solving it for n_D^{QCD} and substituting in the likelihood (6.15). The nuisance parameters of the fit are the numbers $n_{B,C}^{\text{QCD}}$ and the scale factor s .

Results of the fit are shown in Table 6.9. They are compared against a naive prediction from MC simulation

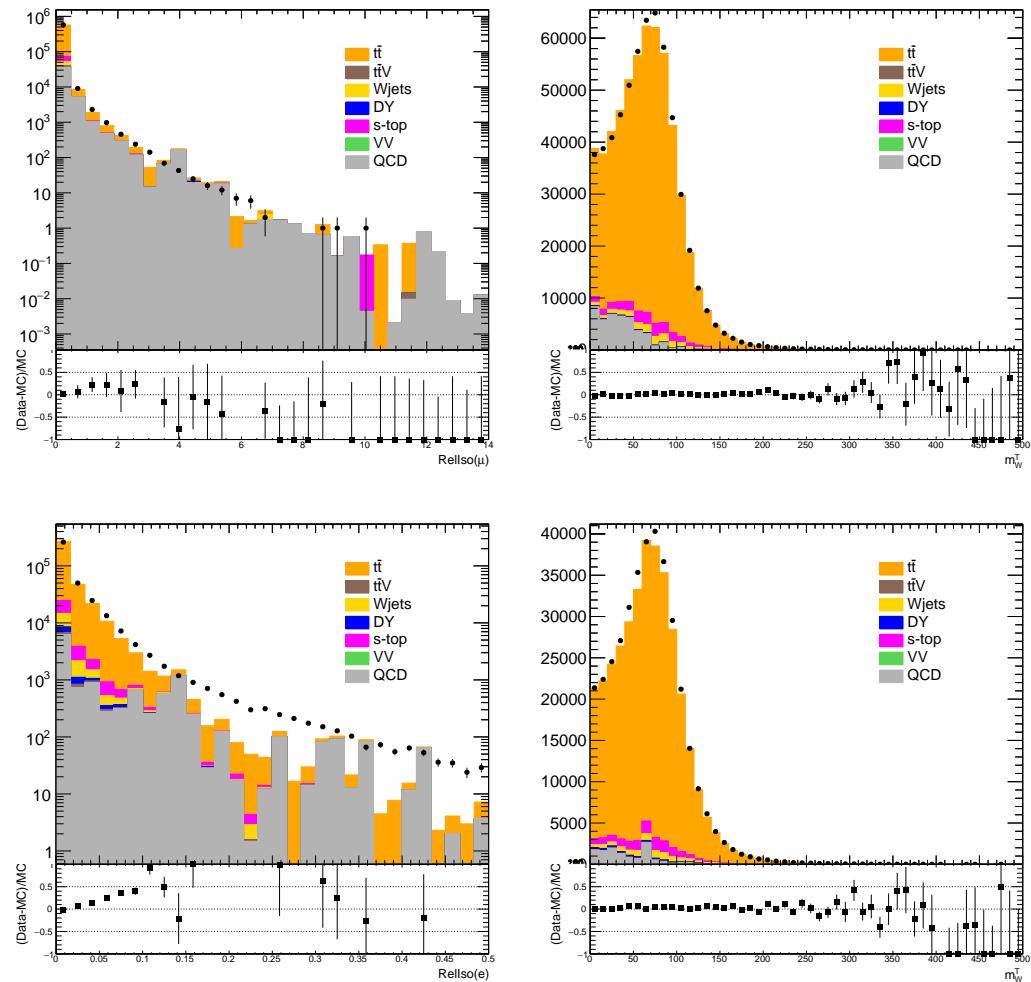


Figure 6.20: Distributions of I (left) and M_T^W (right) in the muon (top) and electron (bottom) channels.

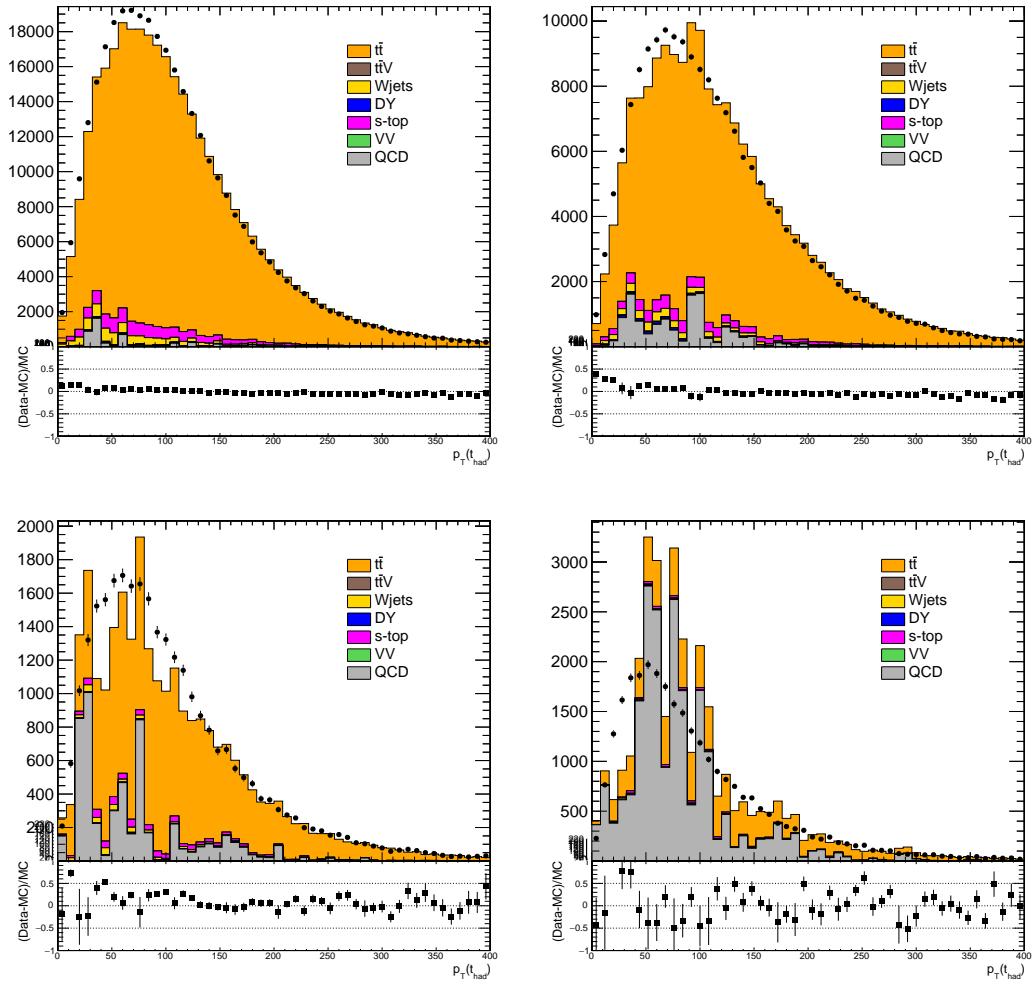


Figure 6.21: Distributions of the p_T of the hadronically decaying top quark in the four regions used to estimate the normalization of the multijet background (from left to right, top to bottom: A, B, C, and D). Muon channel.

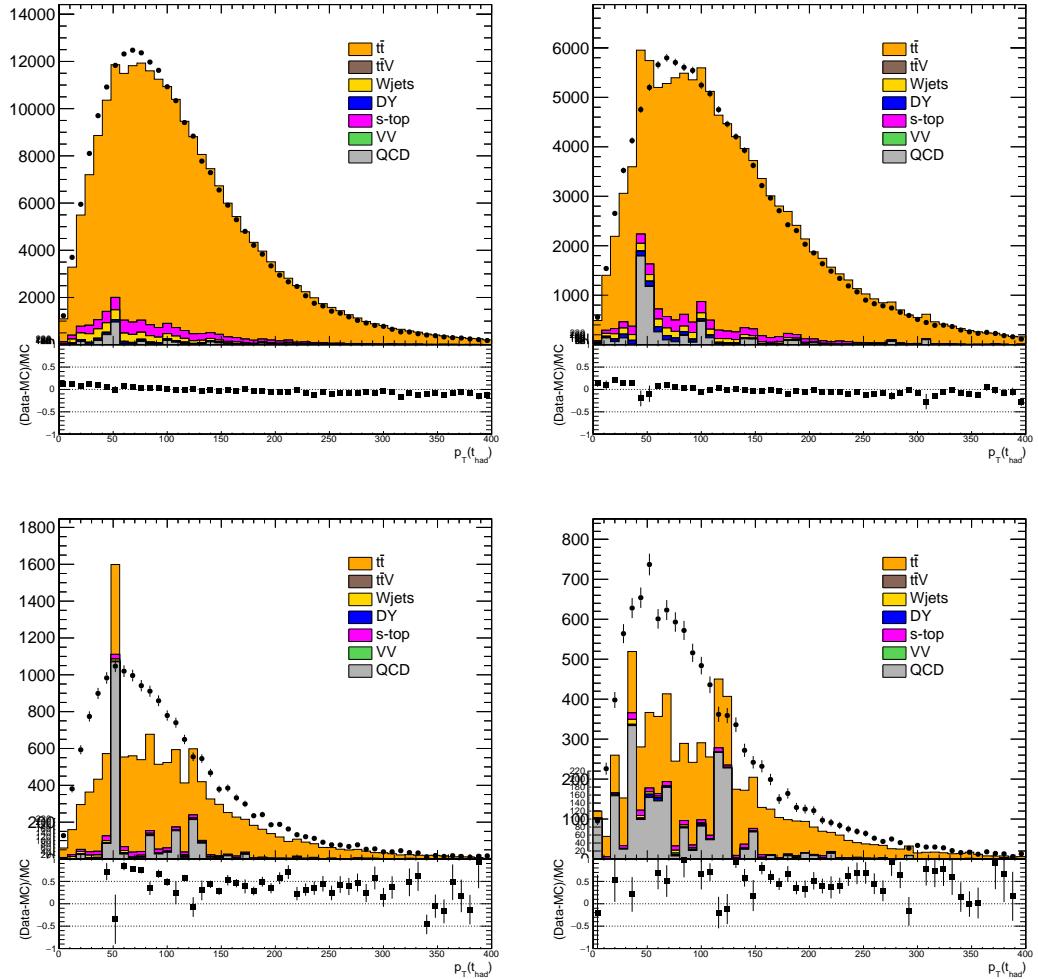


Figure 6.22: Distributions of the p_T of the hadronically decaying top quark in the four regions used to estimate the normalization of the multijet background (from left to right, top to bottom: A, B, C, and D). Electron channel.

Method	Muon channel	Electron channel
ML fit	4284^{+540}_{-520}	3682^{+1801}_{-1703}
Subtraction of prompt bkg	3913 ± 347	3539 ± 500
MC simulation	4762 ± 1528	3093 ± 1024

Table 6.9: Number of events from the multijet background computed using the maximum-likelihood fit, from Eq. 6.14 after backgrounds with prompt leptons have been subtracted, and from simulation directly.

Channel	Region 1	Region 2	Region 3
Muon	$0.15 \leq I < 0.24$	$0.24 \leq I < 0.43$	$I \geq 0.43$
Electron	$0.0588 \leq I < 0.083$ (EB) $0.0571 \leq I < 0.083$ (EE)	$0.083 \leq I < 0.13$	$I \geq 0.13$

Table 6.10: Regions in the lepton's relative isolation used in the construction of the data-driven multijet templates. In the electron channel the lower boundary of region 1 differs between the ECAL barrel and endcaps as it is aligned with the standard identification criterion.

and a simplified computation from Eq. 6.14 while setting $n_R^{\text{QCD}} = n_R^{\text{Data}} - n_R^{\text{Prompt}}$ for $R = B, C, D$. The uncertainties in the table are statistical only, and they are not used in the subsequent analysis. Instead, normalization of the multijet background is assigned a conservative uncertainty of $+100/-50\%$.

6.6.2 Shape

A similar method is adopted as described in region B above to extract the shape of multijet background from data for the search variables, $m_{t\bar{t}}$ and $|\cos \theta^*|$. The simulated backgrounds with prompt leptons are subtracted from the data and resultant shape is attributed to the multijet background, called the multijet template. The only difference compare to the normalization is, the lepton's relative isolation I region is divided into three sub-regions in such a way that each isolation bin has a similar number of events in data. The three bins in isolation have been selected by making a scan over the entire inverted isolation regions and calculated the number of events and the p -value between the two regions. The p -value is calculated using the Pearson's χ^2 test [165], $H_1 \rightarrow \text{Chi2Test}(H_2, "WWP")$, which provides a pairwise comparison between the two shapes.

For muon, the isolation region is divided according to the isolation variable $I_{\text{rel}}^\mu(\Delta\beta)$ as in table 6.10. Comparing shape of data and MC in the three regions, we see enough multijet QCD statistics from simulation shown in Fig. 6.23. Based on the p -values obtained in table 6.11 using, we use $0.15 \leq I_{\text{rel}}^\mu(\Delta\beta) < 0.43$ to obtain the final distributions of $t\bar{t}$ mass and the 2D distribution of the $t\bar{t}$ mass vs $|\cos \theta^*|$. Fig. 6.24, second row, show the data MC comparison in the first and second combined isolation regions for $t\bar{t}$ mass and the 2D distribution of the $t\bar{t}$ mass vs $|\cos \theta^*|$ respectively.

Electron channel is passed through the same procedure as muon but the first isolation region has different ECAL barrel (EB) and endcap (EE) lower limits shown in table 6.10. The p -values obtained from the three regions, table 6.11, show compatibility and the final template for electron channel is taken from the entire anti-isolation region. Data and MC comparison for $t\bar{t}$ mass and the 2D distribution of the $t\bar{t}$ mass vs $|\cos \theta^*|$ in the three regions are shown in Fig. 6.23 third and fourth row respectively. The final data driven multijet template is shown in Fig. 6.24k and l plots.

The multijet templates obtained from data can be effected by systematic variations in the dominant $t\bar{t}$ background with prompt leptons. Therefore a dedicated study has been done to see the effects of these variations. Normalization in $t\bar{t}$ is varied $\pm 6\%$, explain in detail in Sec. 6.7, and the resultant up and down data driven templates are compared using Pearson's χ^2 test as explained above. The two final data driven templates are

Channel	Observable	Region 1 vs 2	Region 1 vs 3	Region 2 vs 3
Muon	$m_{t\bar{t}}$	0.229	$< 10^{-3}$	$< 10^{-3}$
	$\cos \theta^*$	0.037	$< 10^{-3}$	$< 10^{-3}$
Electron	$m_{t\bar{t}}$	0.778	0.265	0.190
	$\cos \theta^*$	0.334	0.115	0.760

Table 6.11: Pairwise comparisons of the shapes of the data-driven multijet templates constructed from the different regions in isolation. Shown are the p -values for the Pearson's χ^2 test of homogeneity.

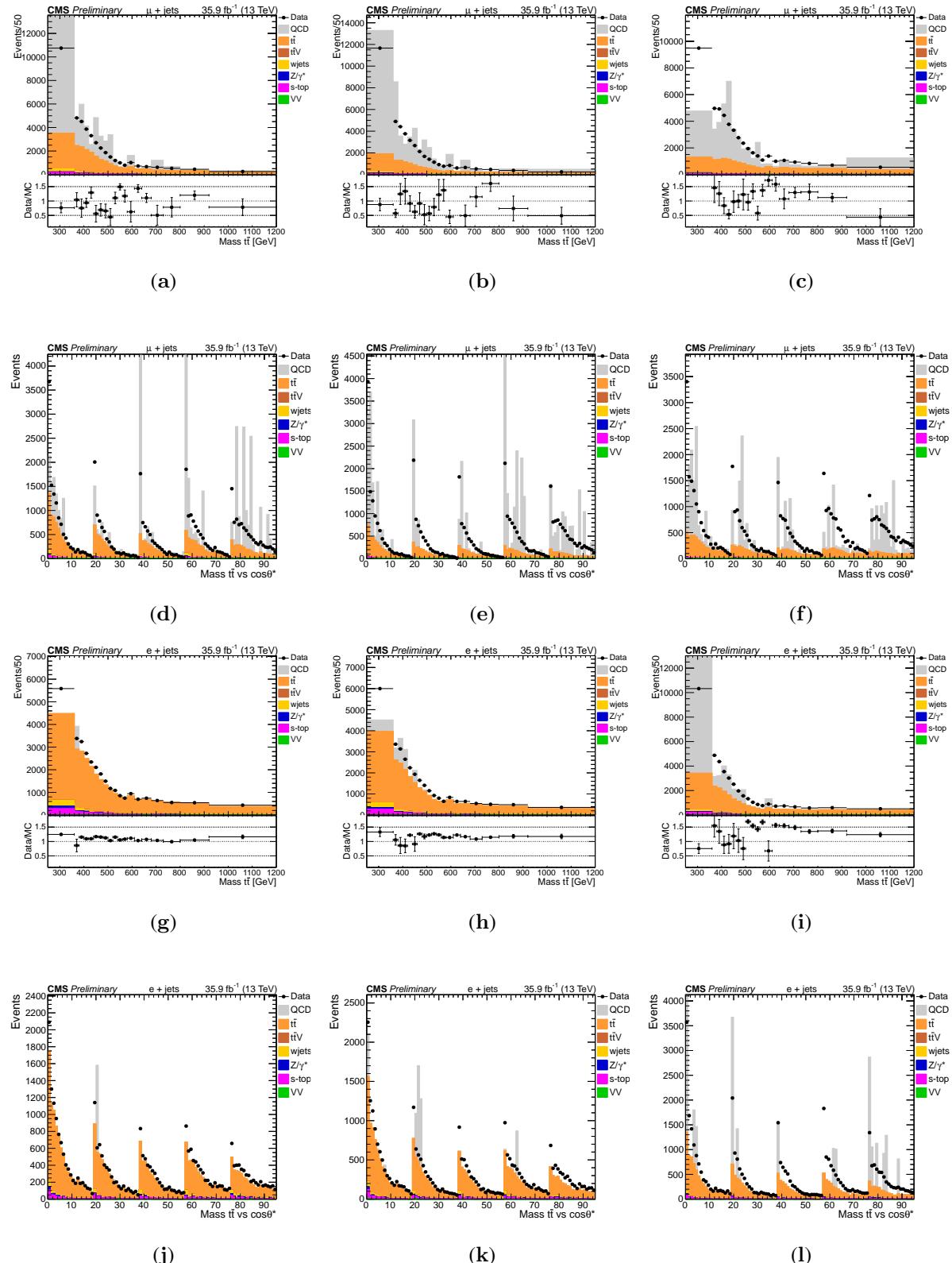


Figure 6.23: Shape comparison of $t\bar{t}$ mass (first row) and $t\bar{t}$ mass versus $|\cos \theta^*|$ (second row) for Data and MC in μ -channel using three isolation regions: $0.15 \leq I_{\text{rel}}^\mu(\Delta\beta) < 0.24$ (first column), $0.24 \leq I_{\text{rel}}^\mu(\Delta\beta) < 0.43$ (second column) and $0.43 \leq I_{\text{rel}}^\mu(\Delta\beta)$ (last column).

For electron channel, third row represents $t\bar{t}$ mass and fourth row $t\bar{t}$ mass versus $|\cos \theta^*|$ where the isolation regions are: $0.0588 \leq I(\rho) < 0.083$ for barrel and $0.0571 \leq I(\rho) < 0.083$ for endcap in the first column. $0.083 \leq I(\rho) < 0.13$ both for barrel and endcap in the second column and $0.13 \leq I(\rho)$ in third column.

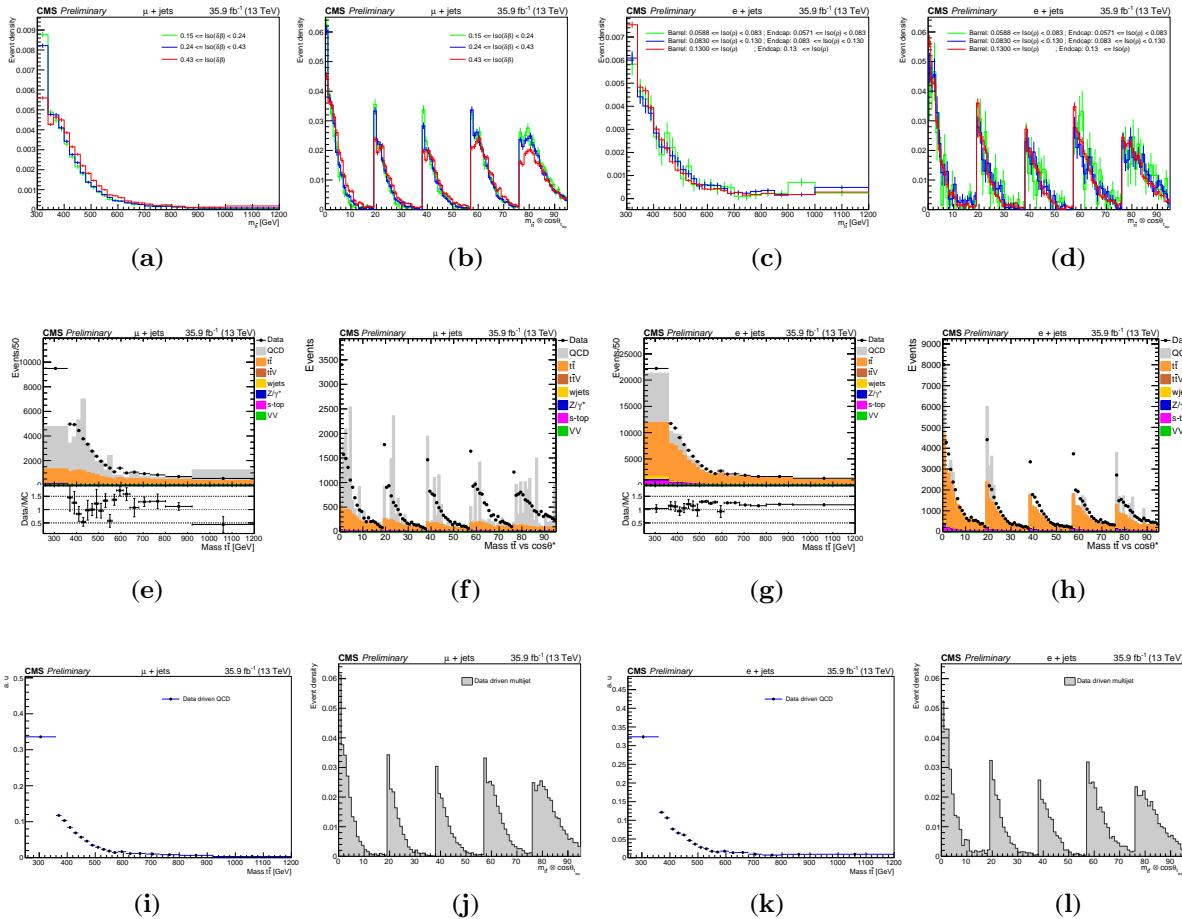


Figure 6.24: (a) Shows mass of $t\bar{t}$ and (b) the 2D distribution of $t\bar{t}$ mass vs $|\cos \theta^*|$ for the difference between Data and non-QCD MC in three isolation regions shown in table 6.10 for μ channel while (c) and (d) are the same kinematic distributions for electron channel. Distributions (e) and (f) show $|\cos \theta^*|$ and mass of $t\bar{t}$ vs $|\cos \theta^*|$ for the range $0.15 \leq I_{\text{rel}}^\mu(\Delta\beta) < 0.43$ respectively for muon channel and (g) and (h) for electron channel with $0.0588 \leq I(\rho)$ for barrel and $0.0571 \leq I(\rho)$ in endcap. Third row are the final data-driven multijet QCD distributions normalized to the area used in the statistical evaluation. (i) and (j) are mass of $t\bar{t}$ and mass of $t\bar{t}$ vs $|\cos \theta^*|$ distributions for μ channel and (k), (l) for electron channel.

Channel	Observable	Normalization	JES	B-efficiency	FSR
Muon	$m_{t\bar{t}}$	0.554	0.727	1	0.981
	2D	1	1	1	0.957
Electron	$m_{t\bar{t}}$	$< 10^{-2}$	0.013	1	0.867
	2D	0.860	0.990	1	0.999

Table 6.12: Pairwise shapes comparison of the data-driven multijet templates constructed from the up down variations of different systematics on $t\bar{t}$. Shown are the p-values for the Pearson's χ^2 test of homogeneity for $m_{t\bar{t}}$ and 2D distribution of $m_{t\bar{t}} \otimes \cos \theta^*$

consistent. Uncertainties from jet energy corrections, b-tagging efficiencies, and renormalization scale used for the final-state radiation are studied for $t\bar{t}$ background. No deviation has been observed in the up and down final templates. The p-values obtained from the Pearson's χ^2 test for different systematics are listed in the table 6.12 and the resultant shapes for $t\bar{t}$ normalization are shown in Fig. 6.25. Because of this, no uncertainty on the shapes is assigned, apart from the inherent statistical uncertainty.

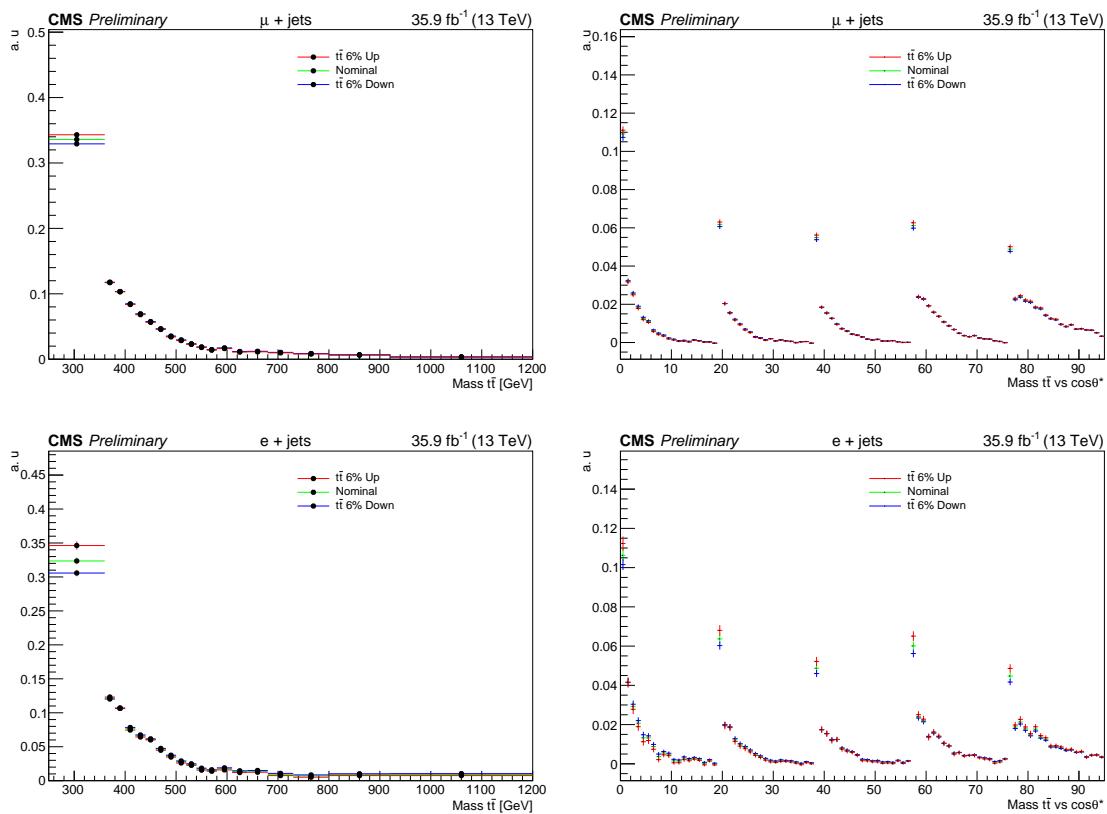


Figure 6.25: A 6% statistical variation on $t\bar{t}$ in the final data driven templates is shown for muon channel in upper and for electron channel in lower row. The up, down variations show consistency in the shape and as a result no systematics have been assigned.

6.7 Systematic uncertainties

Some assumptions and corrections are made because of the limited knowledge of detector features and theory predictions that leads to systematic deviations of the final results. The systematic uncertainties are calculated with changed assumptions or corrections from the nominal value and then compared the deviation to the nominal result. The majority of systematic uncertainties considered in this search can be classified into two categories. The first one includes various experimental effects related to imperfect description of the detector or the LHC machine. These uncertainties affect all considered processes. The second group consists of theoretical uncertainties, which are defined for each simulated process individually. All uncertainties are discussed below. Whenever applicable, their impacts on the distributions of the search variables are shown using the

SM $t\bar{t}$ production as an example.

6.7.1 Experimental uncertainties

One of the dominant uncertainties in this analysis originates from the jet calibration. It is evaluated by varying the multiplicative jet energy correction within its uncertainties. All recommended individual JEC variations [166] are considered, each controlled with a dedicated nuisance parameter in the statistical model. Variations of the same type are applied in a fully correlated manner to all jets, and they are propagated into the p_T^{miss} as well. Figure 6.26 demonstrates the impacts of JEC variations that affect central jets ($|\eta| < 2.4$). It shows ratios between distributions of the search variables in the SM $t\bar{t}$ with each nuisance parameter shifted by one standard deviation and the distribution with the nominal JEC. For a comparison, the impact of the combined JEC uncertainty is also plotted. Figure 6.27 shows the impacts of remaining seven JEC uncertainties that affect only jets with $|\eta| > 2.4$. Since these jets are not used in the analysis directly, the corresponding uncertainties only contribute through the type-1 p_T^{miss} correction. As a result, their impacts are small, and these uncertainties are neglected, leaving a total of 19 independent JEC variations.

JER data-to-simulation scale factors utilized in Eqs. 6.10 and 6.11 are also varied within their uncertainties, simultaneously for all jets. The impact of this variation is shown in Fig. 6.28, together with the variation of the unclustered part of p_T^{miss} . The latter uncertainty is computed by shifting energies of PF candidates not clustered into jets with $p_T > 15$ GeV (which is the threshold used for the type-1 correction) according to the energy resolution for each type of PF candidates [156].

Uncertainties due to variations of b-tagging scale factors are provided in Fig. 6.28 as well. They are evaluated by varying values of the scale factors in Eq. 6.4 within the respective uncertainties. The variations for b and c quark jets are fully correlated, with the uncertainty for c quark jets set to be twice the value for b quark jets. The scale factors for light-flavour jets are varied independently from the heavy-flavour ones.

Figure 6.29 shows impacts of uncertainties in trigger scale factors and scale factors of lepton identification. The latter combine all aspects of the lepton definition discussed in Sec. 6.2.5, which include also the isolation requirement and the track reconstruction. Uncertainties of individual contributing scale factors are summed up in quadrature.

The uncertainty on the integrated luminosity of 2.5% [120] is translated into a simultaneous variation of normalizations of all processes that are described with MC simulation. The effective pp cross section used for the pile-up reweighting as described in Sec. 6.2.1 is also affected by the imprecise knowledge of the luminosity and, additionally, the uncertainty of the physical cross section of the inelastic pp scattering [167]. To account for this uncertainty, the pile-up reweighting is repeated with the effective cross section shifted by 5% [168] in each direction. The resulting change in the distributions is shown in Fig. 6.30.

Finally, the normalization of the data-driven template for the multijet background is varied conservatively by $+100/-50\%$, independently in the muon and electron channels. As it has been demonstrated in Sec. 6.6, uncertainties on the shape of this template can be neglected.

6.7.2 Theory uncertainties

An extensive list of theory uncertainties is considered for the SM $t\bar{t}$ background as it dominates the signal region. Since the signal region includes the bulk of its kinematic phase space (as opposed of selecting an extreme corner), the overall normalization of this background is assigned the same 6% uncertainty as the uncertainty of the NNLO + NNLL QCD cross section used for the normalization. It is imposed that this uncertainty accounts for shifts in the inclusive cross section that can be introduced by any other theory uncertainty thus factorizing the latter one into a variation of the cross section and the impact on shapes of distributions as well as the selection efficiency. To implement this, templates that describe theory uncertainties in $t\bar{t}$ are always normalized to the nominal inclusive cross section. The corresponding variations of the normalization only incorporate effects of changes of the selection efficiency, including the acceptance.

The first set of uncertainties addresses changes of the renormalization and factorization scales, μ_R and μ_F , in the matrix element. The two scales are varied independently by a factor of two in each direction, and the impacts of these variations are shown in Fig. 6.31. In some analyses a simultaneous variation of the two scales, which is provided with a dedicated set of LHE weights, is also considered. In order to check if this is necessary, an attempt was made to decompose the simultaneous variation into a linear combination of independent variations.

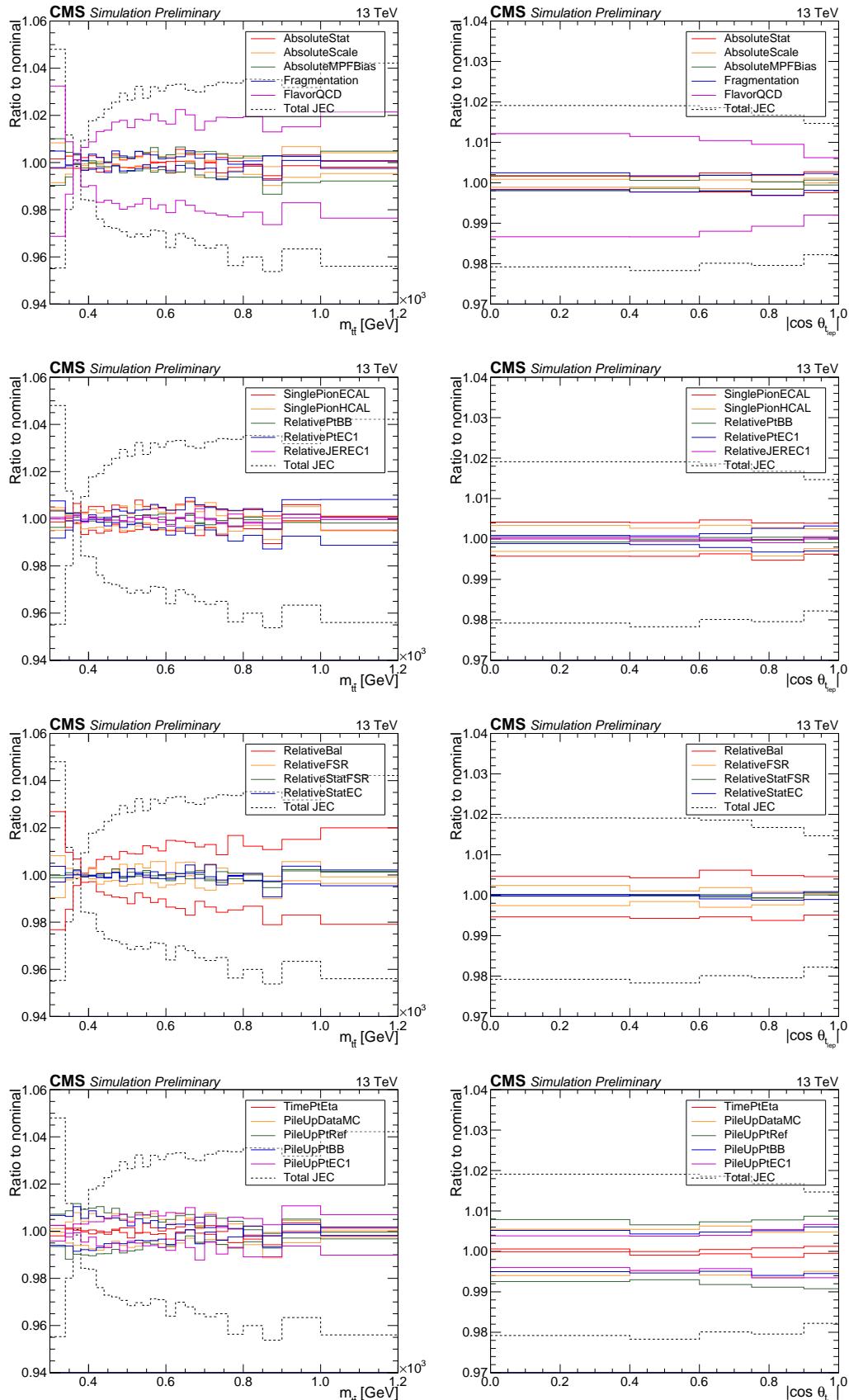


Figure 6.26: Impacts of individual JEC uncertainties for central jets. SM $t\bar{t}$, sum of muon and electron channels.

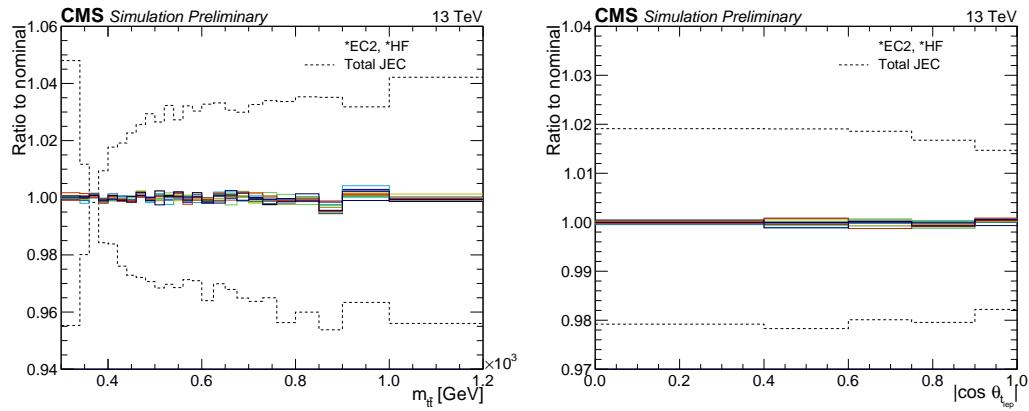


Figure 6.27: Impacts of individual JEC uncertainties for forward jets. SM $t\bar{t}$, sum of muon and electron channels.

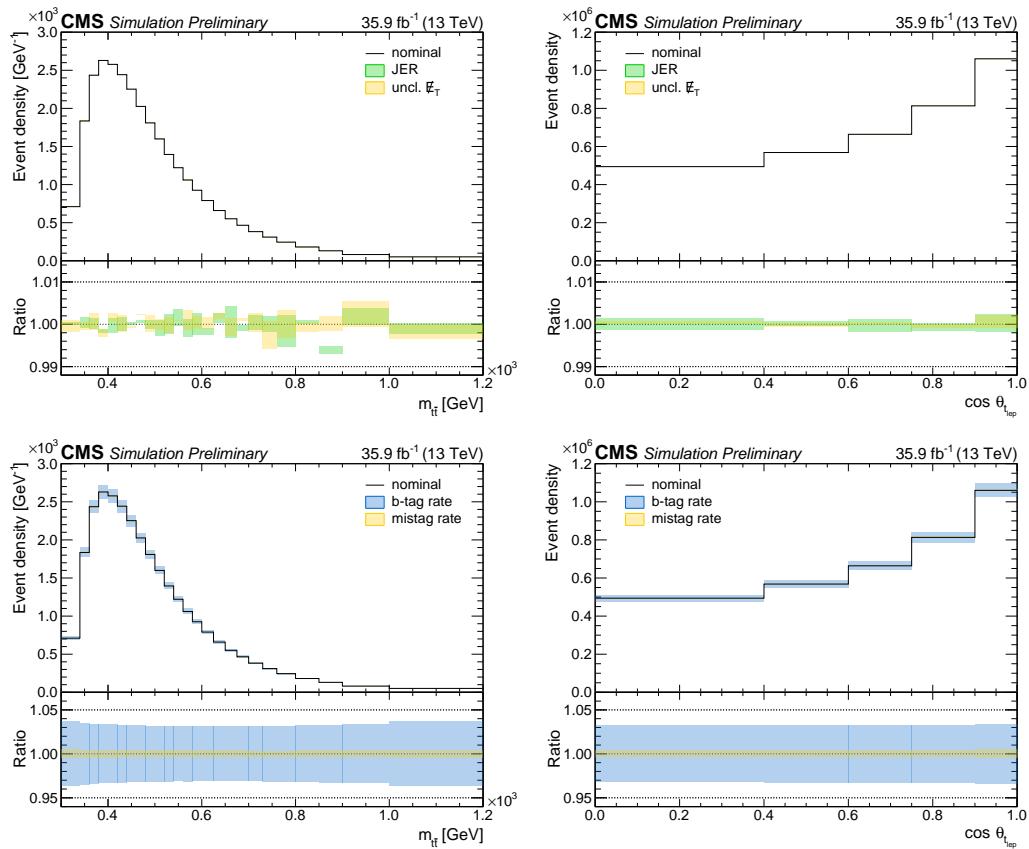


Figure 6.28: Impacts of uncertainties in JER and unclustered part of p_T^{miss} (upper row) and b-tagging (lower row). SM $t\bar{t}$, sum of muon and electron channels.

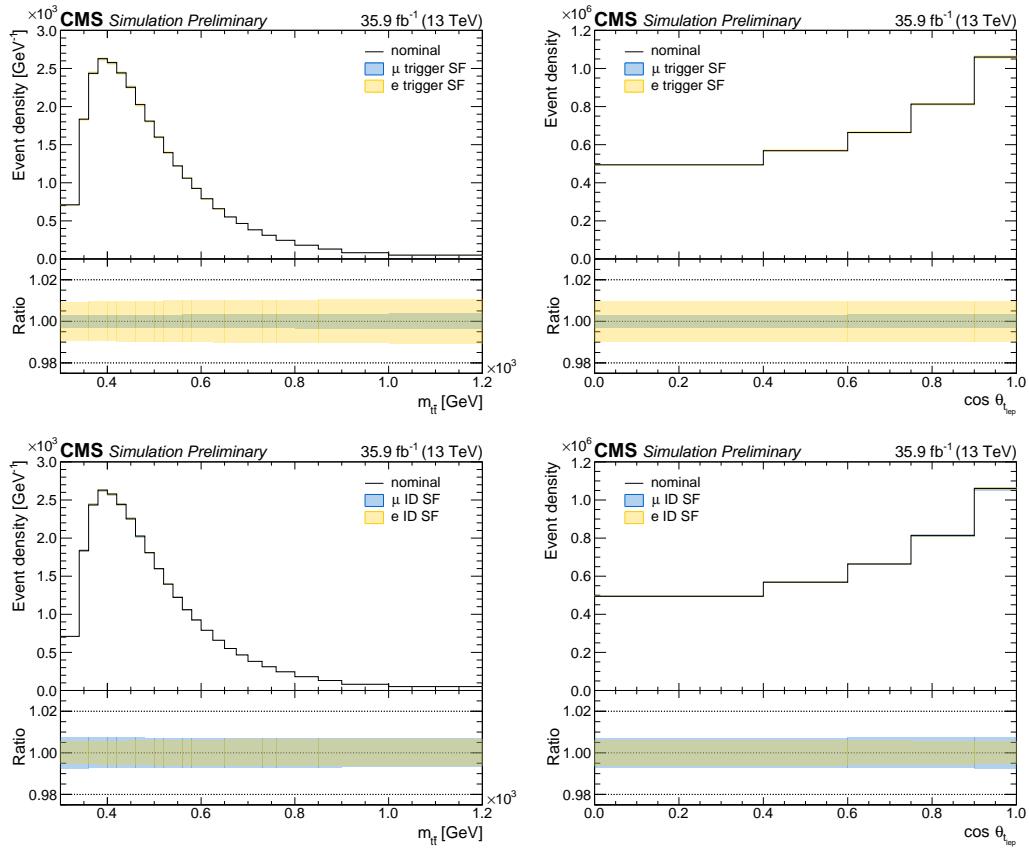


Figure 6.29: Impacts of uncertainties of trigger scale factors (upper row) and efficiencies of lepton identification (lower row). SM $t\bar{t}$, sum of muon and electron channels.

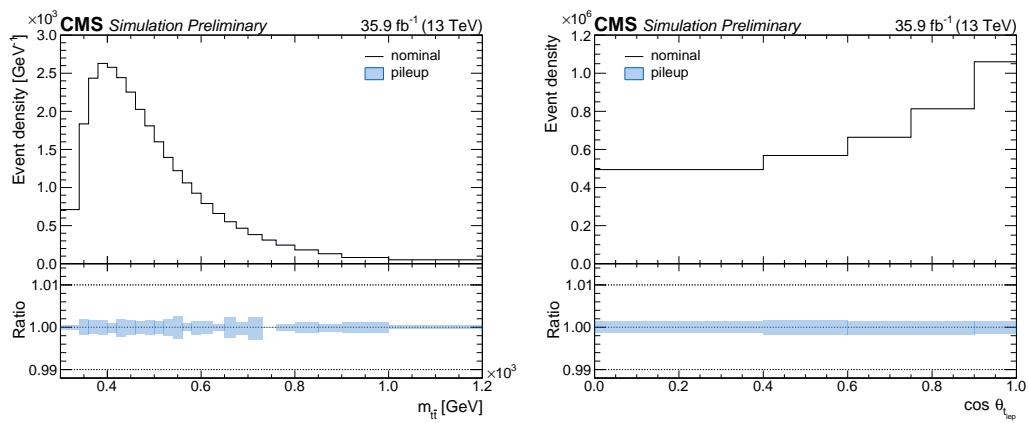


Figure 6.30: Impact of the uncertainty on the pile-up profile. SM $t\bar{t}$, sum of muon and electron channels.

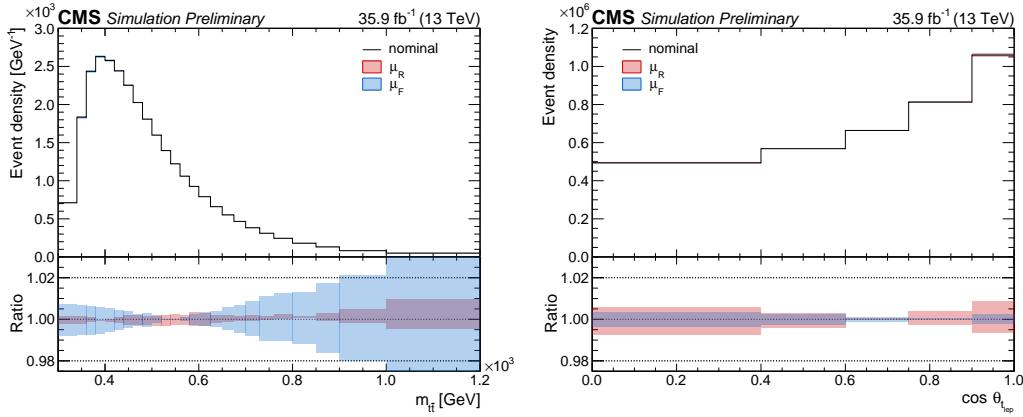


Figure 6.31: Impacts of variations of the renormalization and the factorization scales in the matrix element. SM $t\bar{t}$, sum of muon and electron channels.

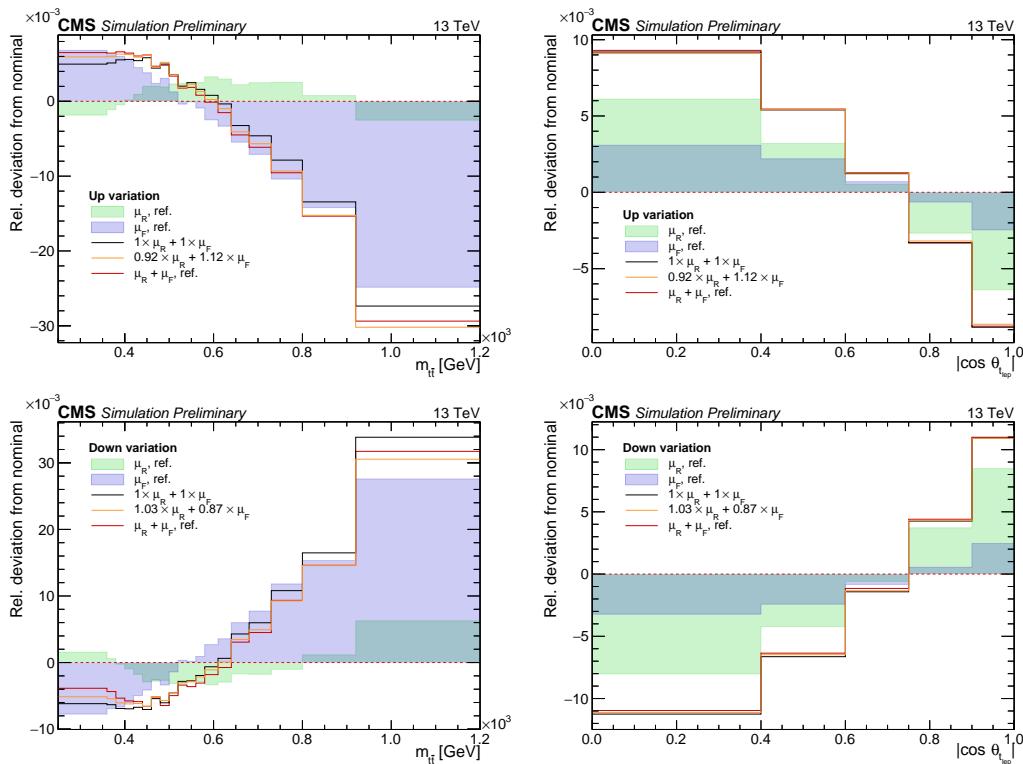


Figure 6.32: Decomposition of the simultaneous variation of the μ_R and μ_F scales (shown in red) into a linear combination of individual variations (filled histograms). The fitted combinations are shown with orange lines. For a comparison, a simple sum of the individual variations is also included (black). SM $t\bar{t}$, sum of muon and electron channels.

Using the two-dimensional distributions of $(m_{t\bar{t}}, |\cos \theta^*|)$, the decomposition was performed by minimizing

$$\sum_{i \in \text{bins}} \left(c_R \cdot h_R^{(i)} + c_F \cdot h_F^{(i)} - h_{R+F}^{(i)} \right)^2 \quad (6.16)$$

with respect to freely floating parameters $c_{R,F}$. Here $h^{(i)}$ is the relative deviation of the given template from the nominal one in bin i . The results are shown in Fig. 6.32 in projections onto each axis. The difference between the reference simultaneous variation and the fitted combination of individual variations is within 0.2% everywhere, and the overall shape is captured correctly. It is therefore concluded that there is no need to consider the simultaneous variation explicitly.

Several uncertainties in the $t\bar{t}$ production are described with dedicated simulated data sets, which were listed in Table 6.3. The renormalization scales used in simulation of initial- and final-state radiation (ISR and FSR), which are controlled in the PYTHIA program with parameters `SpaceShower:renormMultFac`

Variation	Shape compatibility			Rate change	
	Nominal vs up	Nominal vs down	Up vs down	Up	Down
ISR	0.150	0.655	0.487	+0.8%	-0.5%
FSR	$< 1 \times 10^{-3}$	$< 1 \times 10^{-3}$	$< 1 \times 10^{-3}$	-9.6%	+6.0%
m_t	$< 1 \times 10^{-3}$	$< 1 \times 10^{-3}$	$< 1 \times 10^{-3}$	+1.8%	-2.2%
h_{damp}	0.021	0.024	$< 1 \times 10^{-3}$	+0.3%	-0.9%
UE	0.921	0.279	0.721	-0.0%	-0.2%

Table 6.13: Systematic variations described by dedicated $t\bar{t}$ samples: p -values for pairwise homogeneity tests and impacts on the rate.

and `TimeShower:renormMultFac`, are varied independently by a factor of two in each direction. Two alternative values for the mass of the top quark $m_t = 169.5$ and 175.5 GeV are probed. The h_{damp} parameter in POWHEG, which controls the suppression of radiation of additional high- p_T jets, is changed from its nominal value of $1.58 m_t$ to $0.99 m_t$ and $2.24 m_t$ [109]. Finally, a variation of parameters of the underlying event (UE) tune determined in Ref. [109] is provided. The effective integrated luminosity in each of these alternative datasets is about factor 2.6 smaller than in the nominal one, and some of the variations, such as the UE tune, are expected to have a small impact in this search. Because of this, the two-dimensional $t\bar{t}$ distributions with the analysis binning defined in Sec. 6.5.2 are compared taking statistical uncertainties into account. The results of the comparison are reported in Table 6.13. Pairwise tests for compatibility between shapes of the up and down variations or of one of the variations and the nominal distribution are performed using the Pearson’s χ^2 test for homogeneity. The table also shows the impacts of the variations on the normalization. It is evident that the variations in ISR and UE do not produce statistically significant changes in the shape of the distribution, and the shifts in the normalization are small compared to the overall 6% uncertainty. Consequently, these two uncertainties are neglected in the analysis. Impacts of the three remaining uncertainties are shown in Fig. 6.33. For the mass of the top quark the provided variation is ± 3 GeV, while the full uncertainty in the latest CMS combination is about 0.5 GeV [169]. To account for the difference, the corresponding templates are mapped to a $\pm 6\sigma$ variation in the statistical analysis.

The PDF uncertainties are computed from the set NNPDF3.0. They are described by 100 alternative versions of the PDF, referred to as “MC replicas”, which are provided in the form of dedicated event weights. These inputs cannot be used directly in the adopted statistical analysis, which assigns a continuous variation in the distribution ($m_{t\bar{t}}, |\cos \theta^*|$) to each nuisance parameter. To overcome this obstacle, an attempt is made to construct a small number of base variations such that the deviation from the nominal distribution given by each MC replica can be described as a linear combination of said base variations. This is accomplished with the help of the Principal Component Analysis (PCA). For the purpose of the PCA, each MC replica represents a “measurement” and corresponding relative deviations from the nominal distribution in its individual bins identified as the “features”. In order to allow for a combination with the sister search in the dilepton final state, the relative deviations in bins of the distribution of its search variables are also included in the measurements. Each measurement is thus described by 240 features: $25 \times 5 = 125$ relative deviations in bins of the distribution ($m_{t\bar{t}}, |\cos \theta^*|$) in the $l + jets$ final state (muon and electron channel are combined together) plus $23 \times 5 = 115$ deviations in bins of the distribution ($m_{t\bar{t}}, c_{\text{hel}}$) used in the dilepton final state (c_{hel} is an angular variable sensitive to the spin of the $t\bar{t}$ resonance [102]). The PCA finds an orthogonal transformation of coordinates in the feature space that eliminates linear correlations between the (transformed) features. Technically, it diagonalizes the 240×240 covariance matrix constructed from the 100 measurements. The eigenvalues of the covariance matrix have the meaning of the variance of the set of measurements in the directions given by the corresponding eigenvectors. The largest eigenvalues are shown in the left panel of Fig. 6.34. It reveals that the first two eigenvalues are dominant, which can be interpreted that the dissimilarity between the 100 measurements is mostly restricted to the plane defined by the corresponding two eigenvectors, while along any perpendicular direction the measurements are clustered together. This hierarchy allows to approximate the measurements by their projections to this plane, i.e. by keeping only the two dominant components. The projection is shown in the right panel of Fig. 6.34. The distribution of the measurements is approximately Gaussian, with a zero correlation by construction, and therefore the variations given by the two eigenvectors can be mapped to independent normally distributed nuisance parameters in the statistical analysis. The two dominant components define the sought-for base variations, which technically can be constructed by transforming back to the original feature space eigenvectors $(\sqrt{\lambda_1}, 0, 0, \dots)^T$ and $(0, \sqrt{\lambda_2}, 0, \dots)^T$, where $\lambda_{1,2}$ are the first two eigenvalues. The resulting variations are shown in Fig. 6.35. When viewed as functions of $m_{t\bar{t}}$, they are qualitatively similar

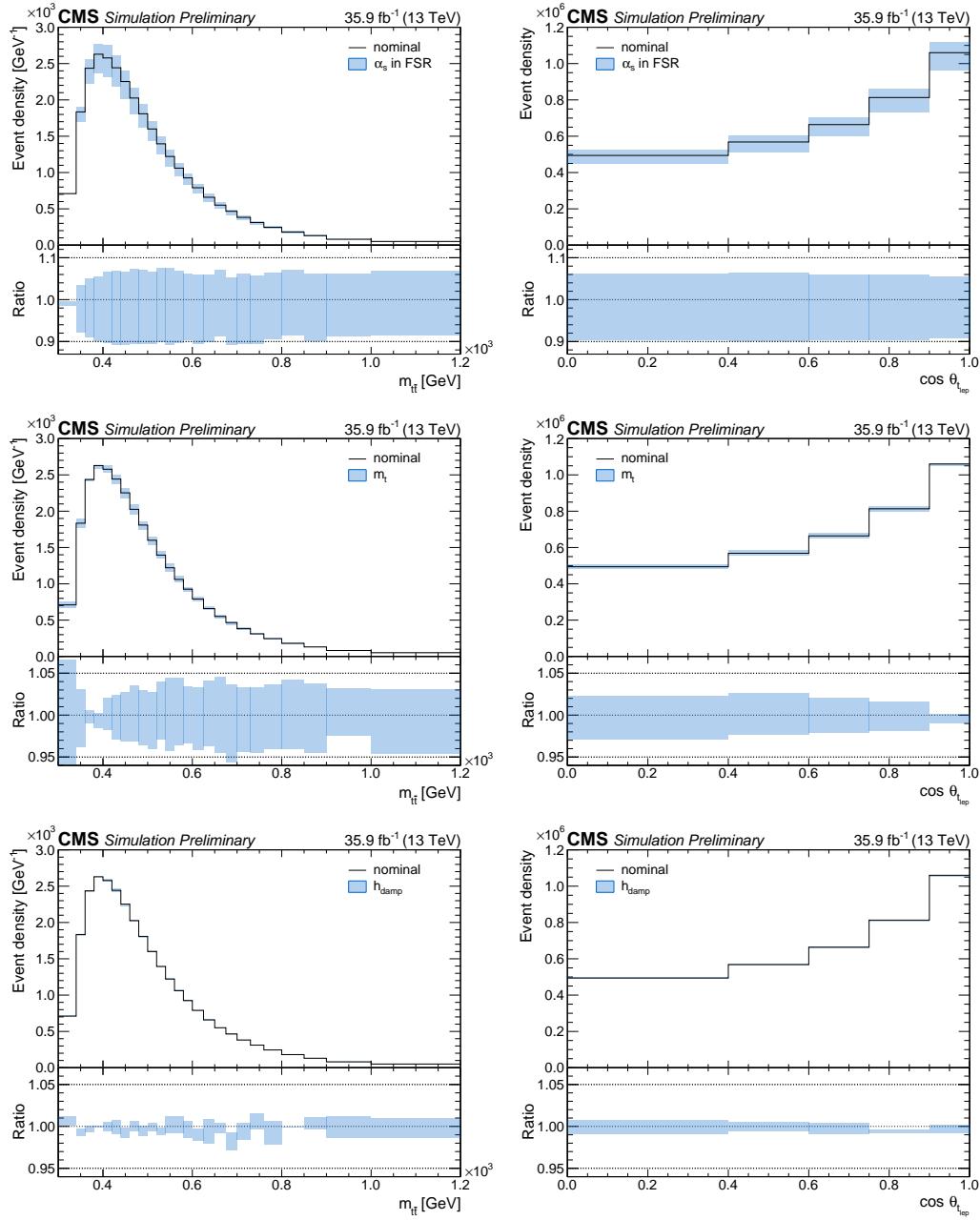


Figure 6.33: Impacts of variations of the renormalization scale in FSR (top), mass of the top quark (middle), and the h_{damp} parameter (bottom). SM $t\bar{t}$, sum of muon and electron channels.

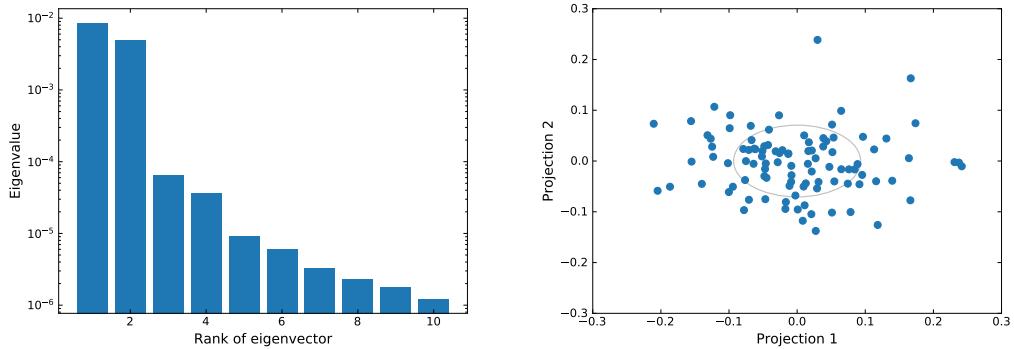


Figure 6.34: The PCA for PDF uncertainties. Ten largest eigenvalues (left) and the projection of the measurements onto the plane defined by the first two eigenvectors (right). Lengths of the semiaxes of the grey ellipse are equal to the standard deviations along the respective projections. SM $t\bar{t}$, sum of muon and electron channels.

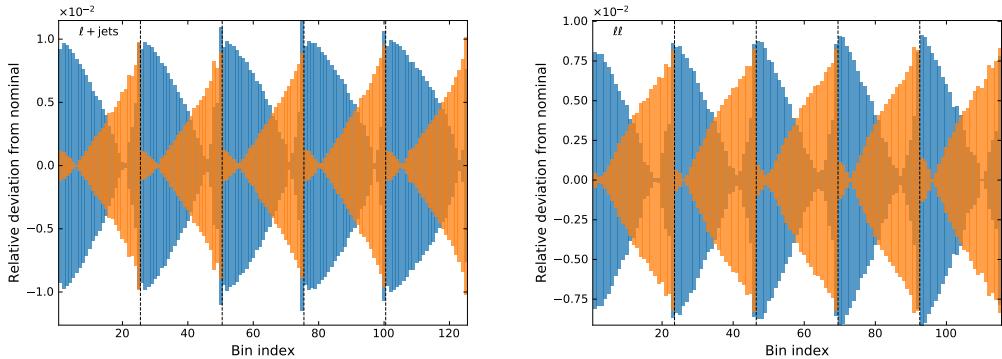


Figure 6.35: Base PDF variations constructed with the PCA, split by the final state. The dashed lines separate groups of ordered bins in $m_{t\bar{t}}$ between different bins in the angular variables.

between the two final states and different bins in the angular variables. The impact of the first (second) variation is pronounced most for smaller (larger) values of $m_{t\bar{t}}$. In the $l + \text{jets}$ final state the relative deviations given by the constructed base variations are applied to the nominal distributions in the muon and electron channels to obtain final templates used in the analysis. Their projections for the two search variables, together with the input deviations given by individual MC replicas, are shown in the upper row in Fig. 6.36.

Additionally, the uncertainty due to the choice of $\alpha_s(m_Z)$ in PDF is included. Its nominal value in the chosen PDF set is 0.118, and weights for a variation of $\pm 1 \times 10^{-3}$ are available. According to the PDF4LHC15 recommendations [170], these variations are rescaled to match a shift in α_s of $\pm 1.5 \times 10^{-3}$. They are shown in the lower row in Fig. 6.36.

The last uncertainty considered for the $t\bar{t}$ background accounts for mismodelling of the p_T spectrum of top quarks discussed in Sec. 6.2.7. Parameters of the involved reweighting are varied according to Eq. 6.8. Figure 6.37 demonstrates the resulting shifts in the distributions.

All other backgrounds are minor, and the only theory uncertainties included for them are conservative variations of their production rates. Their values are reported in Table 6.14, where SM $t\bar{t}$ is included for completeness. Each line in the table corresponds to an independent nuisance parameter. The variations are asymmetric because the nuisance parameters are assigned log-normal distributions.

In case of signal processes uncertainties due to factor two variations in the renormalization and the factorization scales are evaluated. Although the analysis utilizes the full variations, including changes in the shape of the distribution, these uncertainties mostly affect the overall normalization. The impacts of the two variations on the cross sections, together with the PDF uncertainty, are shown in Fig. 6.38 for an example set of signal hypotheses. The total uncertainty is dominated by the variation of the μ_R scale. This allows to ignore the PDF uncertainty, whose correlation with the uncertainty in the $t\bar{t}$ background would be highly non-trivial. The total uncertainty can reach values in excess of 30%, which is expected since the generation has been performed at leading order. As with the SM $t\bar{t}$, the two scales are varied independently. In addition, the variation of the same type of scale is done independently between the resonant part of the signal and the interference. This allows a

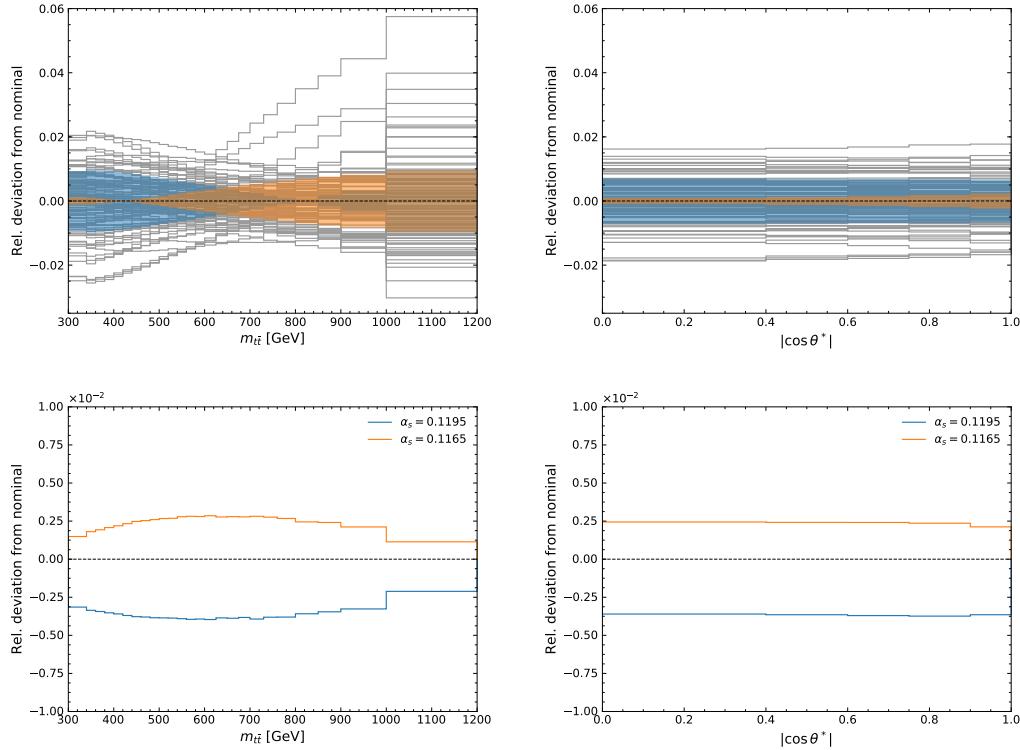


Figure 6.36: Variations due to PDF uncertainties. Individual PDF replicas and the two constructed combined variations (top) and variations in α_s in PDF (bottom). SM $t\bar{t}$, sum of muon and electron channels.

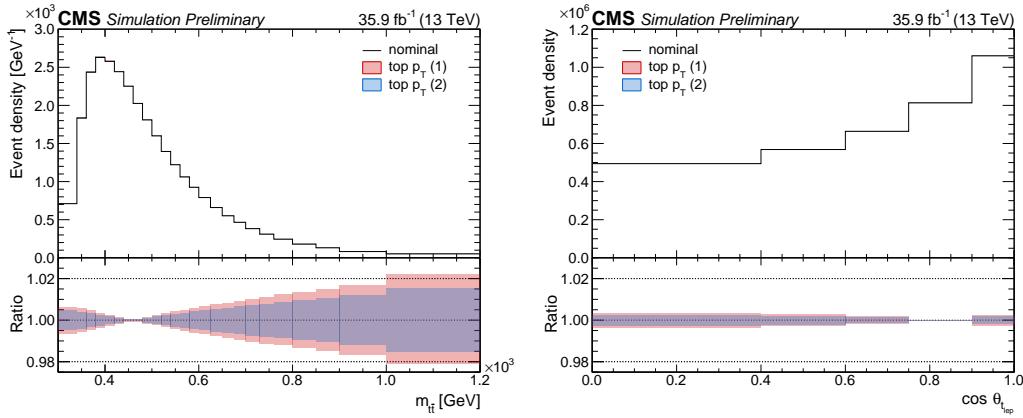


Figure 6.37: Variations due to uncertainties in the reweighting for the top quark's p_T spectrum. SM $t\bar{t}$, sum of muon and electron channels.

Process	Up/down uncertainties
$t\bar{t}$	+6/-6%
t, t channel	+20/-17%
tW	+15/-13%
t, s channel	+20/-17%
$W + \text{jets}$	+50/-33%
$Z/\gamma^* + \text{jets}$	+50/-33%
VV	+50/-33%
$t\bar{t}V$	+30/-23%

Table 6.14: Normalization uncertainties for backgrounds described with MC simulation.

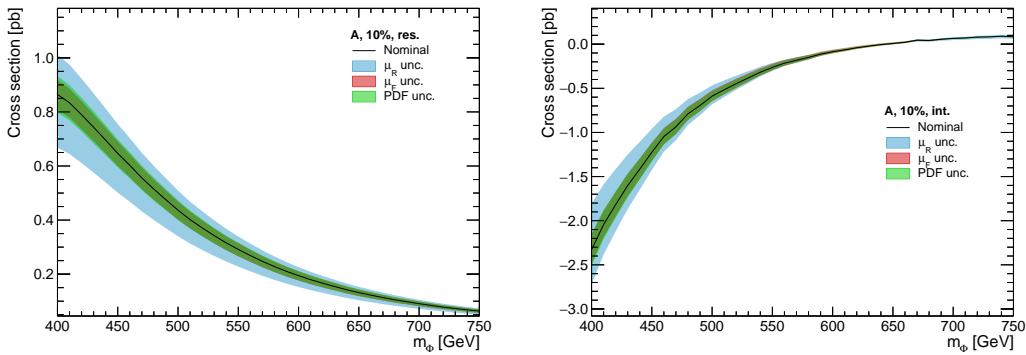


Figure 6.38: Cross sections of signal processes and their uncertainties, as functions of the mass of the heavy Higgs boson. \mathcal{CP} -odd state, 10% total width, resonant part (left) and interference (right), $l + \text{jets}$ final state.

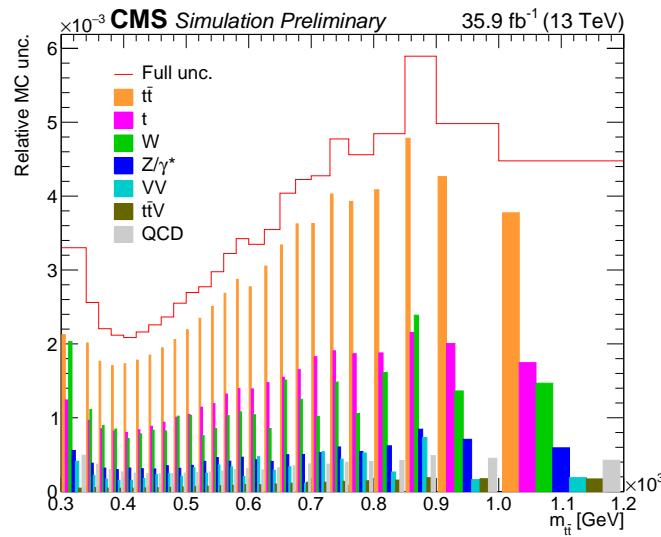


Figure 6.39: MC statistical uncertainty in bins of $m_{t\bar{t}}$, relative to the full expected background. For the multijet background the data-driven template is used.

greater flexibility in the signal model to mimic the SM background and therefore is more conservative than a simultaneous variation.

6.7.3 MC statistical uncertainty

The last category of considered systematic uncertainties accounts for the finite number of simulated events from which the expected distributions are constructed. Resulting statistical uncertainties, split by process, are shown in Fig. 6.39. The full uncertainty is dominated by the contribution from the SM $t\bar{t}$ background, and this allows to assign MC statistical uncertainties to this process only, thus limiting greatly the number of required nuisance parameters. A single uncertainty is computed per bin of the $(m_{t\bar{t}}, |\cos \theta^*|)$ distribution adding in quadrature uncertainties from $t\bar{t}$ and all other backgrounds. These uncertainties are attached to the $t\bar{t}$ process so that in the statistical model they are scaled together with the dominant background. In total, 260 nuisance parameters are added (two channels, 26 bins in $m_{t\bar{t}}$, 5 bins in the angle). For some results, the ‘‘light’’ Barlow–Beeston approach [171, 172] is utilized, which is largely equivalent to the above but profits from an analytical optimization with respect to the nuisance parameters that control the MC statistical uncertainties.

6.8 Statistical analysis

The analysis exploits the two-dimensional distribution of $(m_{t\bar{t}}, |\cos \theta^*|)$ for statistical evaluation. The binning defined in Sec. 6.5.2 and muon and electron channels are included separately. The statistical model is defined

by the likelihood function

$$L(g, \nu) = \prod_i \frac{\lambda_i^{n_i}(g, \nu)}{n_i!} e^{-\lambda_i(g, \nu)}, \quad \lambda_i(g, \nu) = g^4 \cdot s_{R,i}(\nu) + g^2 \cdot s_{I,i}(\nu) + b_i(\nu). \quad (6.17)$$

Here $g \geq 0$ is the coupling modifier from Eq. 6.1, which is treated as the parameter of interest, and ν is the vector of nuisance parameters that control systematic uncertainties described in the previous section. Index i enumerates bins of the distributions in each channel, and n_i is the number of events observed in the corresponding bin. The expected number of events, λ , is given by the sum of contributions from the resonant part of the signal, s_R , the interference s_I , and the SM background b . Although the interference template can contain negative values, the full expectation λ is always positive. When $g = 0$, the background-only model is reproduced.

The analysis is implemented with the help of the `combine` tool [173], which is based on the `RooStats` project [174]. Due to a limitation of the tool, which can only handle non-negative templates, and for the purpose of interpolation between reference values of m_Φ , the interference template is split into two parts according to the sign of the weight of contributing events. The sign of one of the templates is inverted so that both of them are non-negative. The two templates are scaled in the model by factors g^2 and $-g^2$ respectively, thus reproducing the desired distribution. The morphing of input templates for arbitrary values of nuisance parameters is done using the “shape” option, which results in an approximately quadratic interpolation in the vicinity of the nominal template and a linear extrapolation. Effects of simultaneous variations of multiple nuisance parameters are combined in a multiplicative manner.

To check the statistical model for potential pathological behaviour, a maximum-likelihood fit to an $s + b$ Asimov data set is performed, and resulting constraints on nuisance parameters are examined. In an Asimov data set the likelihood (6.17) is evaluated by setting the observed number of events in each bin, n_i , to its expectation λ_i obtained for nominal values of all nuisance parameters. As the signal benchmark, a \mathcal{CP} -odd Higgs boson with a mass of 500 GeV and a relative width of 5% of the mass has been chosen. The coupling modifier g has been set unity when defining the Asimov data set but allowed to float in the fit. The post-fit uncertainty on each parameter has been computed by scanning the likelihood profiled with respect to all other parameters. The resulting constraints are reported in Fig. 6.40 (left). It does not include the numerous nuisance parameters controlling MC statistical uncertainties; for all of them the uncertainties are around ± 0.9 . The majority of observed constraints are expected. The renormalization scale in FSR in the SM $t\bar{t}$ is one of the largest uncertainties, reaching around 10% (note that the scale is varied by a factor of 2, not $\sqrt{2}$ as done sometimes). Input uncertainties in the normalizations of W+jets and multijet backgrounds have been chosen conservatively, so as the uncertainty on the efficiency of the single-electron trigger. The two constrained JEC uncertainties, “FlavorQCD” and “RelativeBal” are the dominant JEC uncertainties, and they are conservative by construction. The constraint on the pile-up cross section is not expected, however. Figure 6.1 suggests that this variation is affected by statistical fluctuations and defined in a separate CMS Analysis Note [175]. But since this variation has a very small impact on the fitted coupling modifier g (see App. D), the constraint can be safely ignored.

Figure 6.40 (right) shows values of nuisance parameters obtained in an $s + b$ fit with the same signal hypothesis performed on data. Constraints on the parameters are similar to those observed with the Asimov data set. The largest deviation from the nominal value is found for the parameter controlling the renormalization scale μ_R in SM $t\bar{t}$. This is consistent with the observation that the down variation in μ_R (see Fig. 6.32) resembles the deviation of data from simulation seen in Fig. C.4.

The agreement between the data and the background-only model for the search variables, which is discussed in App. B on a qualitative ground, is quantified with two versions of goodness-of-fit tests. The tests are based on the “saturated model” [176, 177] and the Kolmogorov–Smirnov measure. They are performed independently for each search variable, utilizing one-dimensional distributions with a fine binning. The full list of systematic uncertainties is included, with the exception of individual JEC variations that are replaced with the total uncertainty. Distributions of the test statistic under the background-only hypothesis are constructed with the help of pseudo-experiments, where the central values of the nuisance parameters are set by the fit to data. The results are shown in Fig. 6.41. For both tests the reported p -values suggest compatibility with the background-only model.

The compatibility with the background-only model is further illustrated in Fig. 6.42. It shows distributions of the search variables after background-only maximum-likelihood fits to data have been performed, independently for each of the two variables.

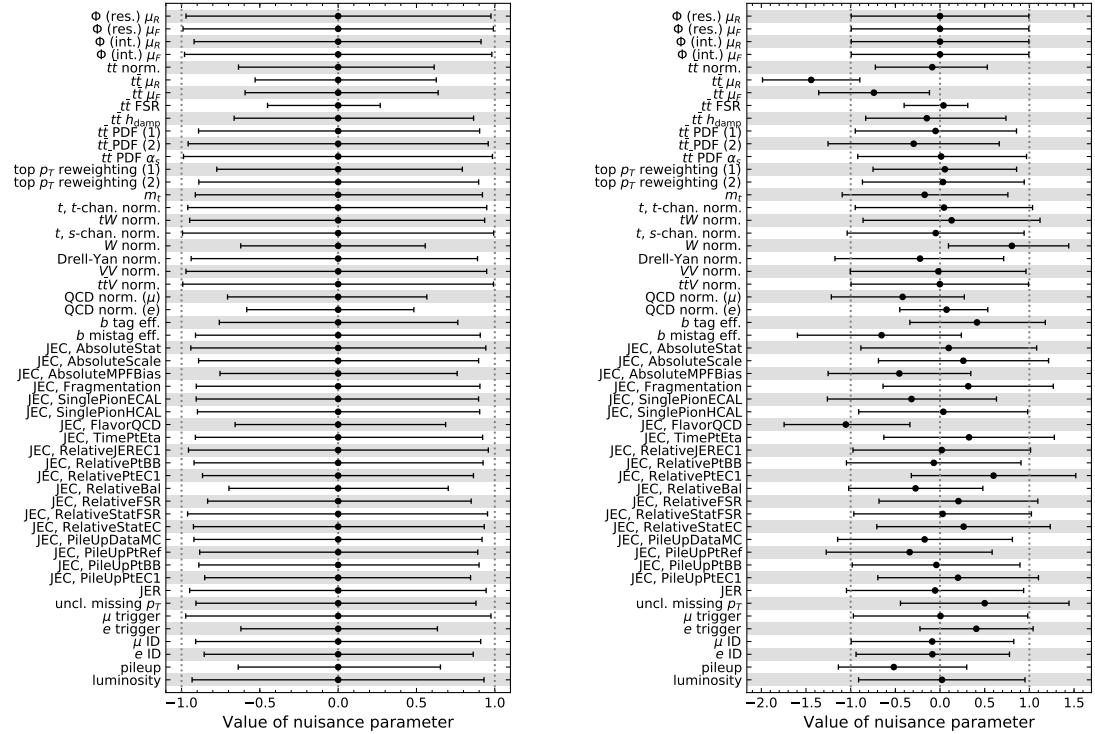


Figure 6.40: (left) Constraints on nuisance parameters obtained in a fit to an $s + b$ Asimov data set. (right) Constraints on nuisance parameters obtained in a fit to data.

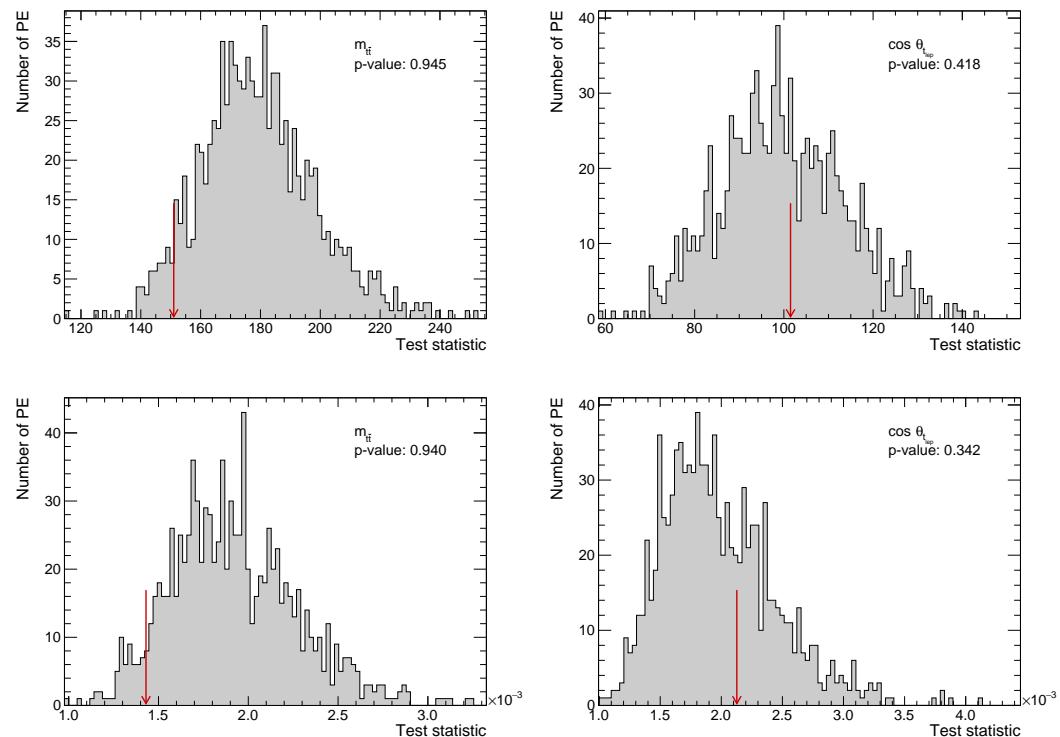


Figure 6.41: Goodness-of-fit tests for the search variables done with the saturated model (top) and the Kolmogorov-Smirnov measure (bottom). The red arrows mark values of the test statistics observed in data.

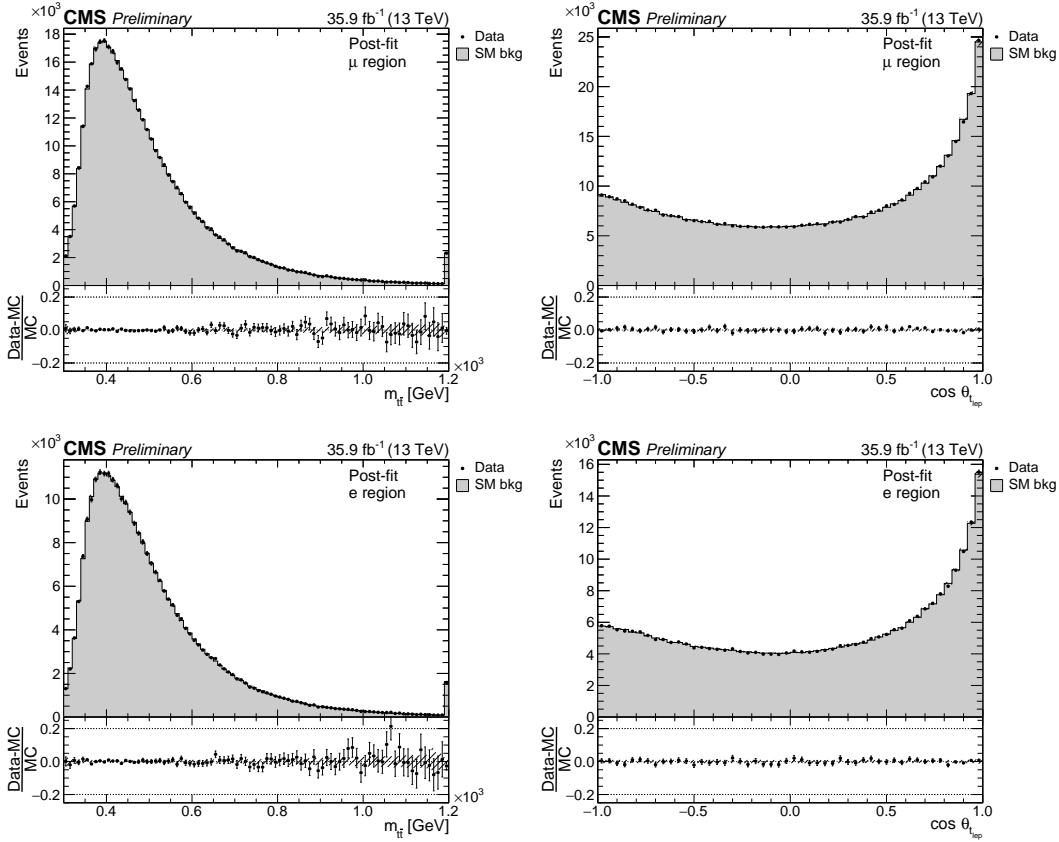


Figure 6.42: Distributions of the search variables after background-only fits. Post-fit uncertainty is shown.

Since no deviations from the background-only model have been observed, upper limits on the coupling modifier g are computed. The analysis follows the standard procedure outlined in Ref. [178]. The ‘‘LHC-type’’ test statistic [179] is utilized, and the critical region for the underlying statistical test is defined using the CL_s criterion [180–182]. Profiting from the sufficiently large number of events observed in each bin of the two-dimensional distributions, distributions of the test statistic are constructed using the asymptotic approximation [179]. The upper limits are computed for each considered signal hypothesis independently. In this chapter only results for the four reference mass points are reported, while the full computation with an interpolation for intermediate values of m_Φ is performed for the combination with the dilepton channel [102], shown in Chap. 7. The results are shown in Figs. 6.43 and 6.44. The red line in the plots marks lower boundaries of unphysical regions in which the partial width of the $\Phi \rightarrow t\bar{t}$ decay (5.5) would exceed the assumed total width of the particle.

6.9 Summary

We have studied the 2016 data using a l+jets selection to search for a heavy Higgs in the $t\bar{t}$ final state taking into account the interference effects with the SM $t\bar{t}$. A two-dimensional distribution of the mass of the reconstructed $t\bar{t}$ system, $m_{t\bar{t}}$ and an angular variable $|\cos \theta^*|$ is investigated with a complete assessment of the main systematic uncertainties. With no indication of the presence of any BSM signal, upper limits on the coupling scale factors are computed for various signal hypotheses.

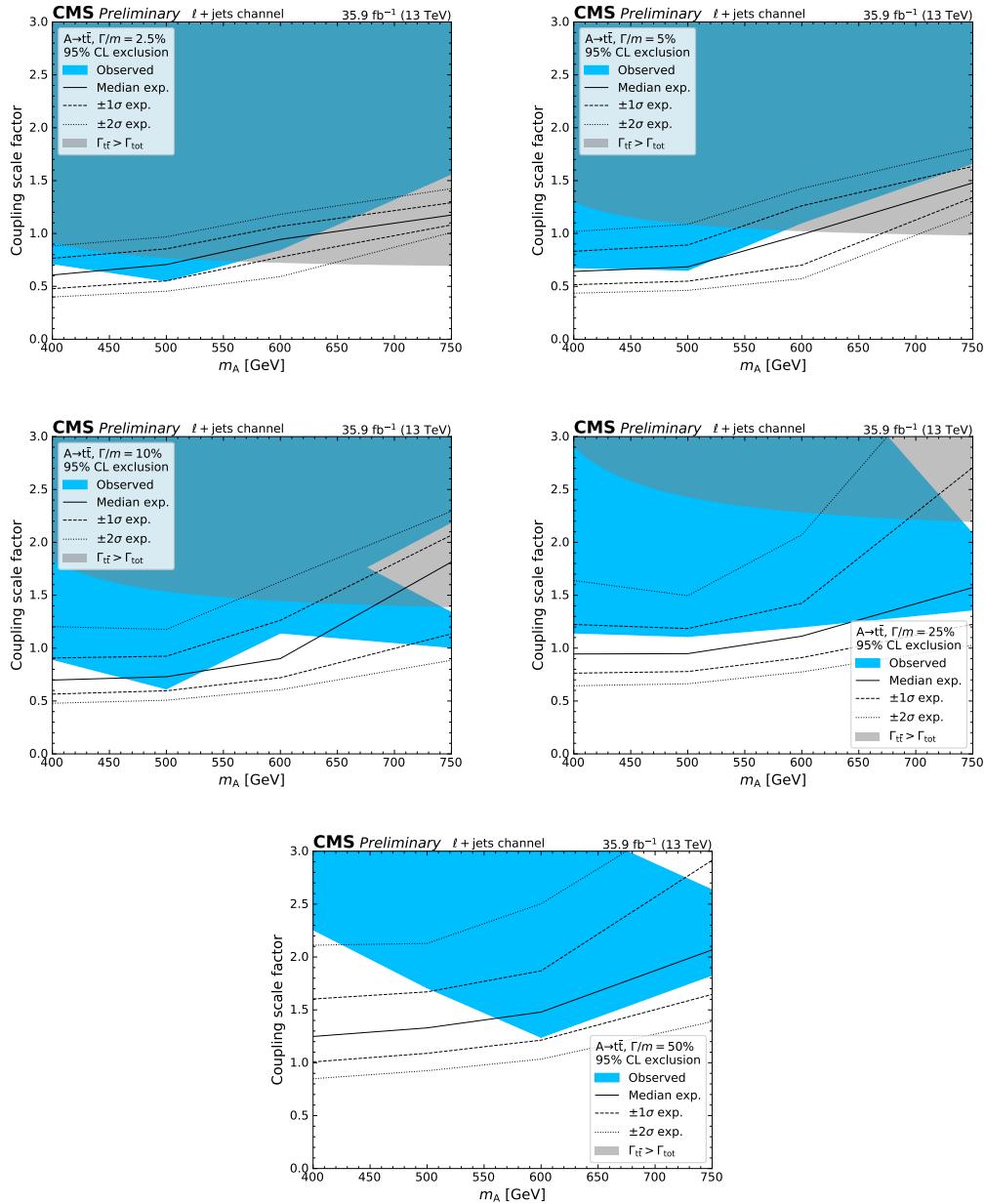


Figure 6.43: Observed and expected 95% CL exclusion for \mathcal{CP} -even heavy Higgs bosons with masses 400, 500, 600, and 750 GeV.

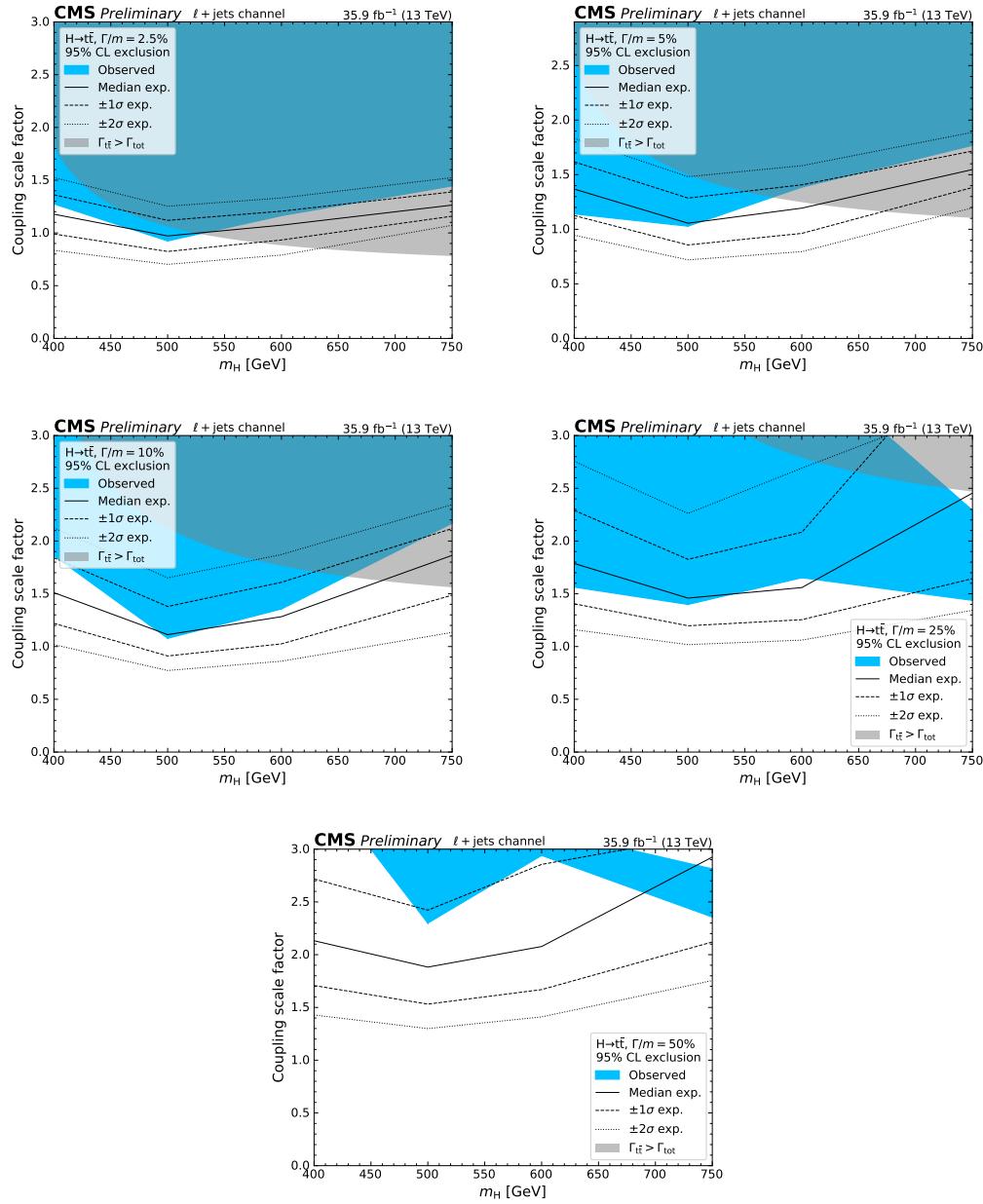


Figure 6.44: Observed and expected 95% CL exclusion for \mathcal{CP} -even heavy Higgs bosons with masses 400, 500, 600, and 750 GeV.

7

Combination of semileptonic and dileptonic channels

In this chapter, the semileptonic channel is combined with the dileptonic final state [118]. This work is also documented in the CMS analysis note in [102]. The two search channels are mutually exclusive because of the different lepton selection criteria (exactly one muon or electron in the first case, and two in the second case). The combined uncertainties of the two channels, hMSSM interpretation, high order cross sections calculation, interpolation and extrapolation to different masses and widths, model independent limits and limits in the context of the hMSSM are discussed. My contribution is the k-factors calculation, scaling the signal to a higher order cross section.

7.1 Statistical evaluation and combination

In both final states, a binned likelihood fit is performed, using as input the two-dimensional distributions of either $(m_{t\bar{t}}, |\cos \theta^*|)$ in the lepton-plus-jets final state and $(m_{t\bar{t}}, C_{\text{hel}})$ in the dilepton final state. The second observables $|\cos \theta^*|$ and C_{hel} are sensitive to the spin of the parent boson and are defined in references [114, 118]. For the lepton-plus-jets final state, muon-plus-jets and electron-plus-jets events are treated as two separate two-dimensional binned distributions, whereas all three dilepton final states are combined into a single two-dimensional binned distribution. The combined likelihood is defined in Eq. 6.17, where i runs over all bins of the three two-dimensional distribution and b_i denoting the combined background yield in a given bin i .

The statistical analysis is performed with the COMBINE toolkit [173], with additional scripts and datacard creation using the COMBINE HARVESTER toolkit [183]. Since separate processes are currently¹ implemented as independent probability density functions (pdfs), in the case at hand binned pdfs, bins with negative yield are not allowed and are forced to zero. To circumvent this issue, the two-dimensional binned distributions for the interference are split into two distributions based on the sign of the part of the squared matrix element that corresponds to interference², one only containing positive entries and one only containing negative entries as discussed in detail in Sec. 6.8.

7.1.1 Systematic uncertainties

Systematic uncertainties are discussed in detail in the analysis notes describing the lepton-plus-jets [114] and the dilepton [118] analyses. In the combination, uncertainties that have the same source are treated as 100%

¹This is not a fundamental limitation, and it may already be possible to implement things differently in the very latest version.

²The splitting by the generator-level sign instead of only splitting histograms with positive and negative bin contents after reconstruction has the advantage that the more physical distributions with larger event counts can be used for the mass morphing, which is discussed below.

correlated; all other uncertainties are treated as uncorrelated. In particular, the following sources of uncertainty pertain to both analyses:

- Experimental: different sources of jet energy scale uncertainties; jet energy resolution; unclustered missing p_T ; b-tagging, separately for light-flavoured jets and b+c jets; pile-up modelling; muon and electron identification/reconstruction efficiencies.
- Theory background: diboson background normalisation; $t\bar{t}$ background normalisation; tW background normalisation; W+jets background normalisation; DY background normalisation; vector-boson-associated $t\bar{t}$ normalisation.
- Theory $t\bar{t}$: PDF, FSR, renormalisation scale, factorisation scale, h_{damp} parameter, top p_T parametrisation (two parameters), top quark mass.
- Theory signal: renormalisation and factorisation scale uncertainties (two parameters each), separately for signal-only and interference and for scalar and pseudoscalar signal.

The uncertainties from the following sources are only considered in the lepton-plus-jets analysis:

- Normalisation of multijet background in muon-plus-jets and electron-plus-jets channel (2 separate parameters); normalisation of t-channel and s-channel single top quark production (2 separate parameters).
- Trigger efficiencies for single-electron and single-muon triggers.

The uncertainties from the following sources are only considered in the dilepton analysis:

- Trigger efficiency of triggers used in the dilepton channel.

Uncertainties that originate in the limited number of simulated events in each bin (so-called bin-by-bin uncertainties) are of statistical origin and are hence treated as uncorrelated across bins, and are hence also naturally uncorrelated for the two final states.

7.2 Interpretation

7.2.1 Interpretation in the hMSSM

The BSM higgs has already been discussed in detail in Sec. 2.2, I am going to give a short summary with emphasis on the hMSSM scenario, which is relevant to our search channel. The Higgs sector in the MSSM is a special case of a type-2 2HDM model³. It can be described by two parameters at tree level, $\tan \beta$ and m_A . In addition to a SM-like scalar h , 2HDM models introduce another scalar H , a pseudoscalar A , and two charged Higgs bosons H^\pm . After the discovery of a SM-like 125 GeV Higgs boson, and given the strict mass relations in the MSSM discussed below, the lighter of the two scalars is identified with the SM-like Higgs boson. While the observed mass of 125 GeV is well compatible with constraints in the MSSM, it implies that the supersymmetry scale must be relatively large for low values of $\tan \beta$ to be achievable. This constraint is perfectly compatible with the so far not successful searches for scalar partners of the top quark.

Various specific models of the MSSM Higgs sector were prepared before the discovery of the 125 GeV Higgs boson and made predictions at low $\tan \beta$ that are incompatible with this observation. To circumvent this issue, the so-called hMSSM was proposed [119], which takes the existence of the 125 GeV Higgs boson directly into account. Given the high SUSY scale, radiative SUSY corrections are not important, in particular at low $\tan \beta$, which is the region in which the presented search is sensitive. This implies in turn that, once the lighter scalar is identified with the 125 GeV Higgs boson, the hMSSM can again be fully described by the two parameters $\tan \beta$ and m_A . The importance of $t\bar{t}$ decays in searches for MSSM Higgs bosons is discussed in reference [28].

Table 7.1 gives an overview of the couplings of the neutral 2HDM/MSSM Higgs bosons to up-type quarks, down-type quarks, and vector bosons. Pseudoscalar decays to vector bosons are forbidden, and the coupling to

³Couplings in the type-2 2HDM are a subset of the MSSM couplings, hence extensively studied. The mixing angle in Eq. 2.30 and the tree level mass relation of the charged Higgs boson in Eq. 2.31 are enforced. The superparticles are assumed to be too heavy to have an impact on the Higgs sector. Coupling to the up-type quarks is same as in type-1 2HDM but for down-type quark the couplings are much smaller in the low $\tan \beta$ scenario compared to type-1 2HDM

Φ	$g_{\Phi \bar{u}u}$	$g_{\Phi \bar{d}d}$	$g_{\Phi VV}$
h	$\cos \alpha / \sin \alpha$	$-\sin \alpha / \cos \beta$	$\sin(\beta - \alpha)$
H	$\sin \alpha / \sin \beta$	$\cos \alpha / \cos \beta$	$\cos(\beta - \alpha)$
A	$\cot \beta$	$\tan \beta$	0

Table 7.1: Couplings in type-2 2HDM models, in particular also the MSSM, of the three neutral Higgs bosons, denoted as Φ , to up-type quarks, down-type quarks, and vector bosons, in terms of the MSSM parameters α and β . The couplings are normalised to SM Higgs boson couplings.

up-type quarks is inversely proportional to $\tan \beta$, whereas the coupling to down-type quarks is proportional to $\tan \beta$. The former implies that $gg \rightarrow A$ production via a top loop and subsequent decays to $t\bar{t}$ are dominant at low $\tan \beta$, whereas b-quark associated production is relevant at higher $\tan \beta$ as well as decays to either b quarks or tau leptons.

hMSSM mass relations

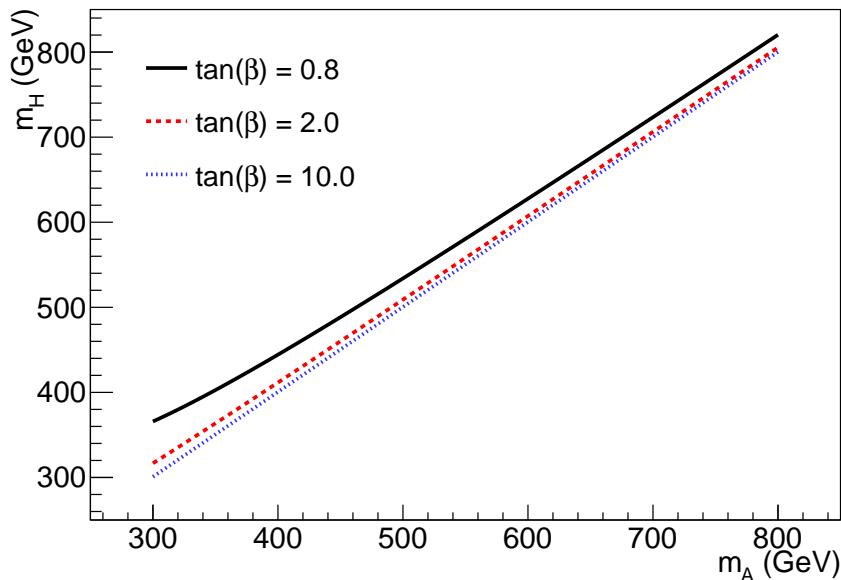


Figure 7.1: Scalar mass (m_H) as a function of pseudoscalar mass (m_A) for different values of $\tan \beta$ in the hMSSM.

Under the assumption that new particles do not affect the production and decay of the Higgs bosons considered, the discussion in this section applies equally to the hMSSM and to generic type-2 2HDM models. The major exception is that MSSM models lead to strict mass relations between the additional Higgs bosons for a given $\tan \beta$. Figure 7.1 shows the dependence of the scalar mass on the pseudoscalar mass for different values of $\tan \beta$ ⁴. While for high $\tan \beta$, the A and H masses are nearly degenerate, the H mass is significantly larger for $\tan \beta$ near 1 and below. This implies that the presented search cannot use the often-made assumption of mass degeneracy between the H and A , but needs to consider the H mass for each point in the $(m_A, \tan \beta)$ plane.

Figure 7.2 shows the A and H widths as a function of the corresponding mass for different values of $\tan \beta$.

The branching fractions for A and H decays to $t\bar{t}$ are given in Fig. 7.3. For A decays, the branching fraction is close to unity for small values of $\tan \beta$. For H decays, the branching fraction is still close to unity but a bit smaller because of additional open decay channels including bosons.

Finally, Fig. 7.4 shows the plane of pseudoscalar width vs. $1/(\tan \beta)^2$, with the latter being equivalent to the coupling squared, g^2 . The total width increases roughly linearly with g^2 where the decay to $t\bar{t}$ dominates. Also shown are the three produced samples at relative widths of 2.5, 5, and 10%, with the markers corresponding to $\tan \beta = 1$ or, equivalently, $g = 1$, and the attached lines showing the variation when the coupling is altered. For a given width, limits can be set on given points in the hMSSM by checking whether the observed upper limit on the coupling is smaller than the coupling that corresponds to the $\tan \beta$ at the given width. To cover sufficiently small variations in $\tan \beta$, the produced widths are evidently not sufficient. Therefore, an interpolation of the

⁴This plot and the following plots in this subsection are made using the publicly available hMSSM model files [184], as discussed in reference [185]. For the final interpretation, we use a finer grid for the predictions, which we needed to recalculate. The details are given in the next subsection.

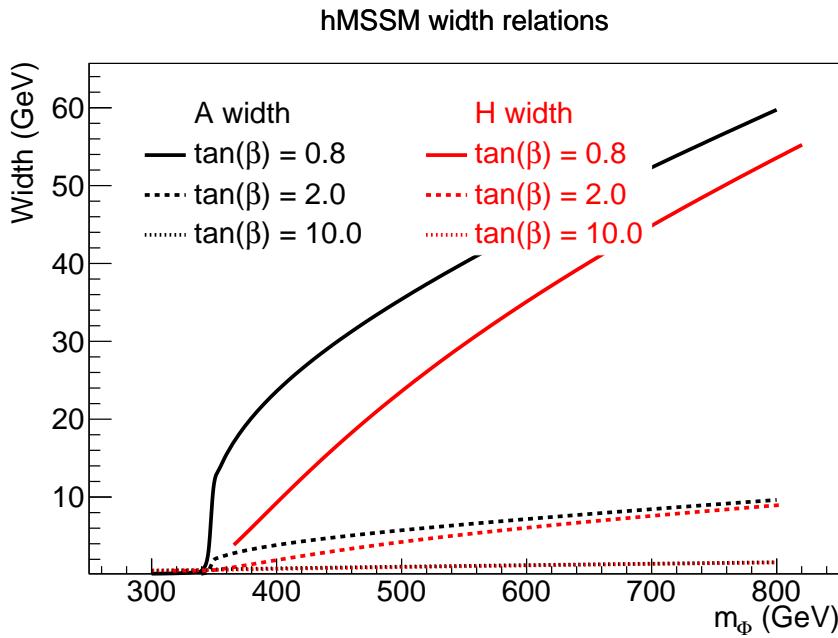


Figure 7.2: Scalar (red) and pseudoscalar (black) total widths as a function of corresponding mass (m_{Φ}) for different values of $\tan \beta$ in the hMSSM.

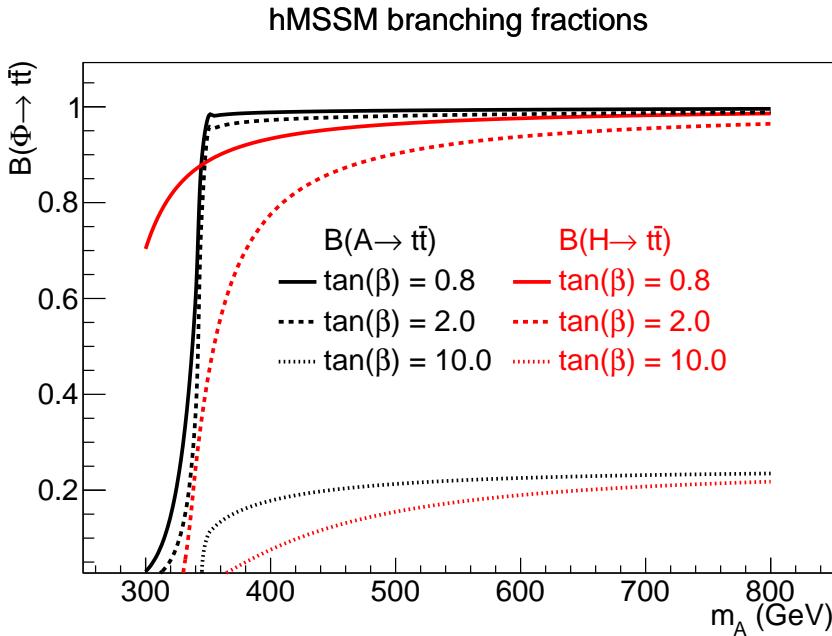


Figure 7.3: Branching fraction for scalar (red) and pseudoscalar (black) decays to $t\bar{t}$ as a function of pseudoscalar mass (m_A) for different values of $\tan \beta$ in the hMSSM.

produced samples is applied as well as an extrapolation to lower widths, using the narrow-width approximation. Details on these procedures are given in Sec. 7.3.

7.2.2 Higher-order cross sections

Cross section calculations with NNLO accuracy are available for $gg \rightarrow A$ and $gg \rightarrow H$ production, and recommended to be used by relevant LHC searches (see in particular the yellow report 4 [185]). We scale the resonant part of the signal with the ratio of NNLO to LO cross sections for $gg \rightarrow A$ or $gg \rightarrow H$ production, respectively. We define this ratio as r_S . In line with the 8 TeV ATLAS result [106, 186], we scale the yields of the interference between $gg \rightarrow A$ or $gg \rightarrow H$ signal and the SM $t\bar{t}$ background with

$$k = \sqrt{r_S r_{SMt\bar{t}}}, \quad (7.1)$$

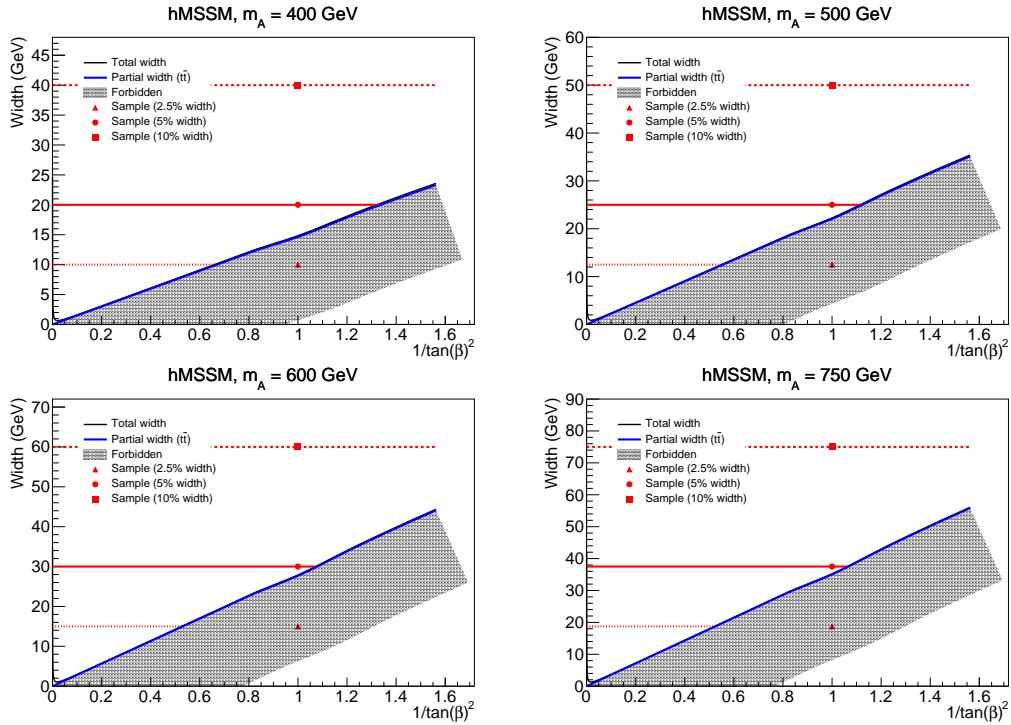


Figure 7.4: Total width of pseudoscalar in the hMSSM (black line) and partial width for decays to $t\bar{t}$ as a function of $1/\tan \beta^2$. In the forbidden region (shown in grey), the total width would be smaller than the partial width for decays to $t\bar{t}$. In addition, the three widths for which samples are produced are indicated as horizontal lines. In addition, their default configuration with $g = 1$, corresponding to $\tan \beta = 1$, is shown by the markers. The horizontal line indicates the phase space that is covered by varying the coupling g , and hence the rates of the signal and interference contributions. For a given sample, a specific point in the hMSSM, which can be obtained from the black curve, can be excluded if the observed limit on the coupling modifier g is larger than the corresponding $\tan \beta$ value. These relations are shown for masses of 400, 500, 600, and 750 GeV.

with $r_{SMt\bar{t}}$ denoting the ratio of the SM $t\bar{t}$ cross section at NNLO+NNLL and LO accuracy.

The signal cross sections are obtained in the following way. First, the hMSSM parameters (in particular masses and widths of the additional Higgs bosons) are obtained with the 2HDMC programme [111]. Then, the cross sections at NNLO accuracy are obtained with the SUSHI programme [112]. These cross sections are obtained with the same setup as the results in reference [185]. The numerical values of the parameters can be found in reference [184]. The only difference is that we use a different PDF set, NNPDF with version 3.0, to correspond to the PDF set that is used in the production of our signal and background samples.

The LO cross sections are calculated separately both with the MADGRAPH programme and with SUSHI. The obtained cross sections are found to be consistent within 1%. The ratios of NNLO and LO cross sections show a slight dependence on mass and width, and are typically of the order of 2. Finally, we scale the cross sections obtained by MADGRAPH with the ratio of the NNLO and LO cross sections obtained with SUSHI.

For $t\bar{t}$ production, we use the TOP++ [113] programme to calculate cross sections both at NNLO+NNLL and at LO accuracy. We also calculate the LO cross section with the MADGRAPH event generator used for our signal sample production. The LO cross sections obtained with TOP++ and MADGRAPH agree at the per mill level. The ratio of NNLO+NNLL and LO cross sections is found to be 1.57.

Figure 7.5 shows a summary of the obtained scaling factors. The scaling factors range from 2.12 for 400 GeV signal samples to 1.71 for 750 GeV interference samples. The factors vary by up to 7% as a function of mass, with the scaling factor being smaller for higher masses, and are consistent for different widths within about 1%.

Note that we currently do not assign uncertainties in the NNLO cross sections or in the scaling factors, but we intend to add the cross section uncertainties in the next iteration.

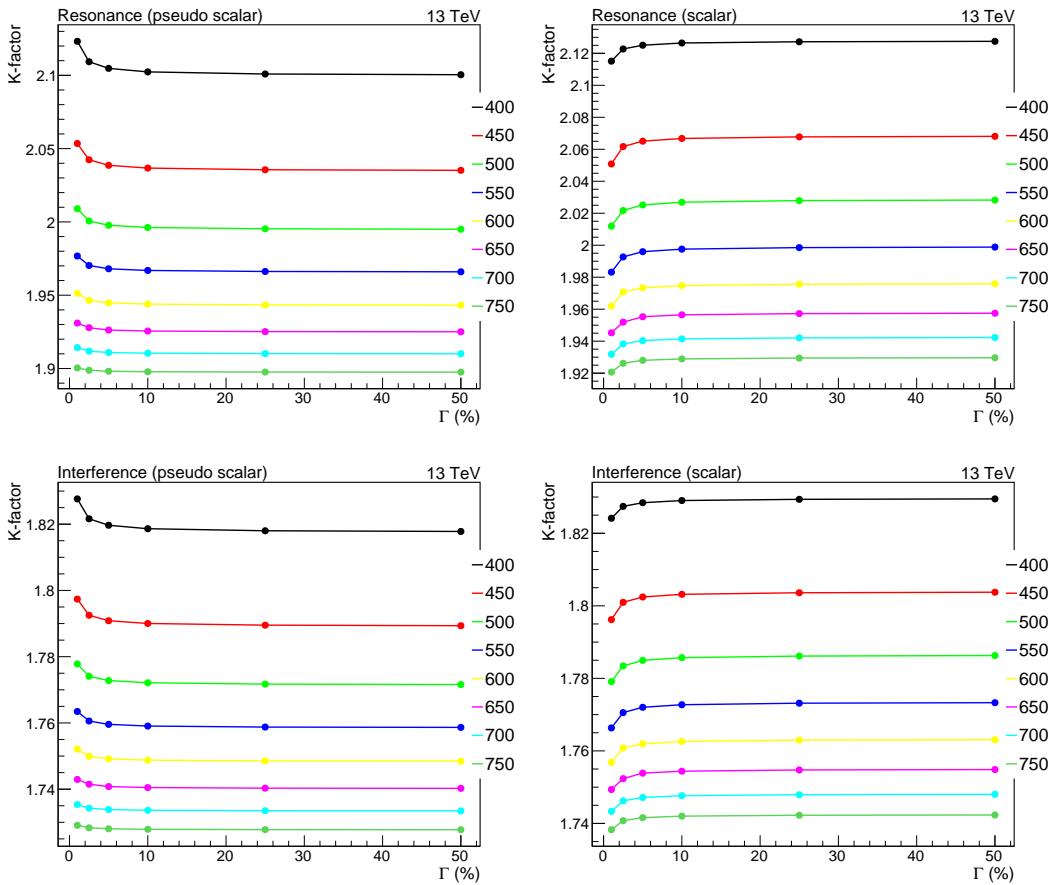


Figure 7.5: Scaling factors applied to the signal-only part of the pseudoscalar signal (top left), signal-only part of the scalar signal (top right), interference part of the pseudoscalar signal (bottom left), and interference part of the scalar signal (bottom right). The factors are shown as a function of resonance width for different signal masses. Note that the y axes only show a limited range, i.e. the relative variations are small.

7.3 Interpolation and extrapolation to different masses and widths

In the case of the hMSSM discussed in Sec. 7.2.1, it is impossible to fully cover the relevant part of the $(m_A, \tan \beta)$ spectrum with a limited set of samples in terms of fixed masses and widths, like the ones produced and discussed above. There are a number of reasons for this:

1. At low $\tan \beta$, where this search is sensitive, m_H and m_A are not mass-degenerate within the detector resolution. This implies that it is not possible to simply use the produced values of m_H and m_A and interpolate between the points scanned in this way.
2. Furthermore, the difference between the mass points (100–150 GeV) is larger than the mass resolution, which roughly corresponds to half this difference or worse.
3. In addition, the width depends strongly on $\tan \beta$, which implies that, even for a given m_A , a quasi-continuous dependence on the width is required to be able to set consistent limits.

This section therefore describes the algorithms with which the distributions are interpolated between different masses and widths, and in a second part how the distributions are also extrapolated to lower widths within the narrow-width approximation.

7.3.1 Mass and width interpolation

One of the input variables of the two-dimensional distributions used for the statistical evaluation is the estimated mass of the parent boson. This reconstructed mass distribution shifts along the mass axis if the mass of the parent boson changes. This implies that an interpolation between two parent boson masses requires a morphing of the two input mass distributions that changes (morphs) the distributions along the mass axis, and

in particular that a vertical interpolation (also known as bin-by-bin morphing algorithm) cannot adequately describe intermediate masses.

It was also found that simple linear horizontal morphing algorithms do not guarantee a sufficiently good interpolation between the mass points. Therefore, the so-called ROOMOMENTMORPH algorithm is used [187]. It was found that the setting NONLINEARPOSFRACTIONS gives the most consistent and overall best results.

The algorithm is applied in the following way:

- As input to the algorithm, individual mass distributions in bins of the angular variable are used.
- These mass distributions are produced with a finer binning than the templates used for the final statistical evaluation.
- In each bin of the angular variable, the ROOMOMENTMORPH algorithm is applied, with the fine-binned distributions of all other masses (400, 500, 600, and 750 GeV) as input.
- Beforehand, these fine-binned distributions are divided by the predicted cross sections (in case of the resonant contribution) or effective positive and negative contributions to the yields (in case of the interference contribution). While the morphing algorithm can technically also handle changes in normalisation, the differences in cross section between the masses are large. Dividing by these cross sections therefore improves the results and only interpolates changes in yield because of acceptance effects.
- The one-dimensional interpolated distributions in bins of the angular variable are then multiplied by the cross section at the target mass, stitched together to unrolled two-dimensional distributions to comply with the inputs to the statistical evaluation, and finally bins are merged to arrive at the chosen final binning.

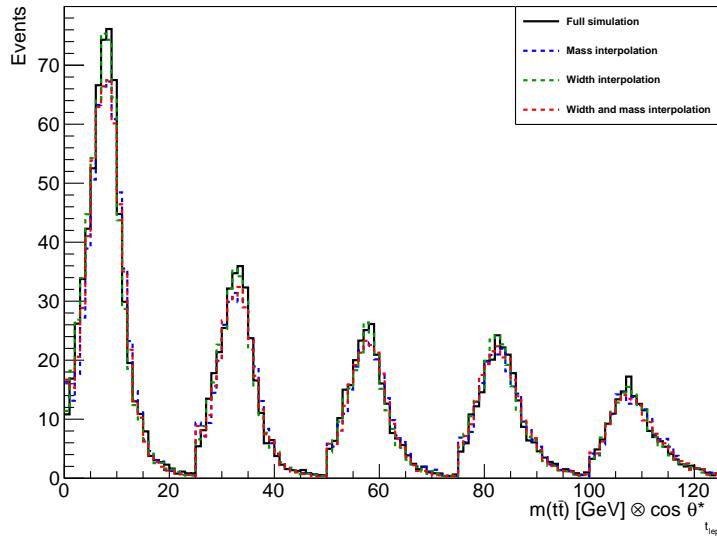


Figure 7.6: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the signal templates in the muon-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

Figures 7.6–7.14 show results of the validation performed, separately for the three different final states (muon-plus-jets, electron-plus-jets, and dilepton) and for different cases. In this validation, the distributions of the 500 GeV mass point are not used (“leave-one-out”), i.e. an interpolation between 400 and 600 GeV (with also the 750 GeV mass point as additional input) is performed. In all plots, the final unrolled distributions after applying the mass interpolation algorithm are compared to the distribution obtained after full simulation. There is overall good agreement between the interpolated and the fully simulated points. It should be noted that the interpolation that is used for the final results only interpolates between points that are 100 GeV apart (150 GeV between 600 and 750 GeV), whereas this validation is for a mass difference of 200 GeV.

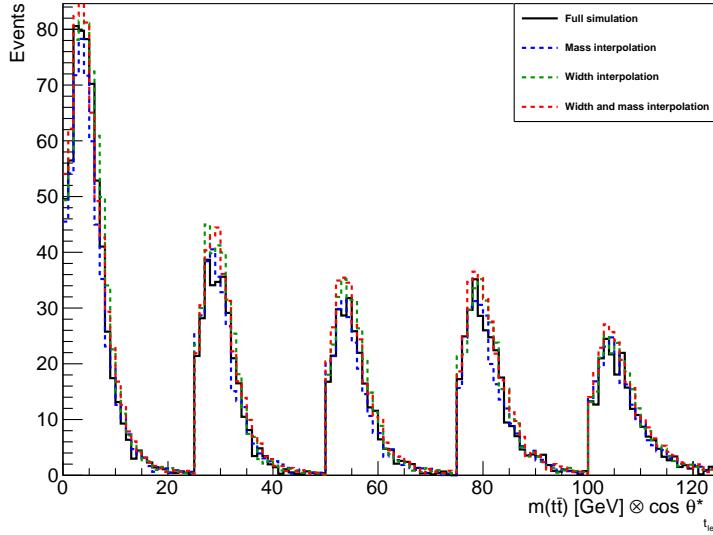


Figure 7.7: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the positive interference templates in the muon-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

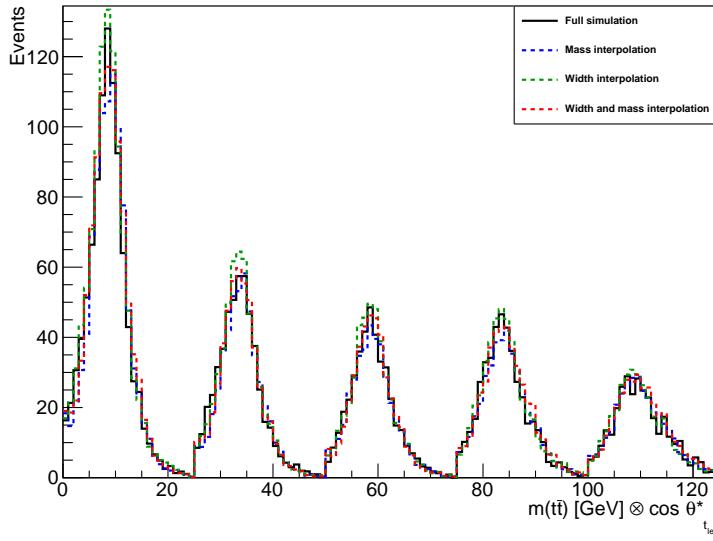


Figure 7.8: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the negative interference templates in the muon-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

Since changing the width does generally not lead to a shift in horizontal distribution but rather to a shift only of the width of the distribution, a bin-wise interpolation can be performed in this case. The cross section for the resonant part scales with $1/\text{width}$. Therefore, a hyperbolic interpolation is performed in this case. The yield of the interference part only exhibits a mild dependence on the resonance width. In this case, a linear extrapolation is performed.

Figures 7.6–7.14 also show results of the validation of the width interpolation. The validation is also performed with the “leave-one-out” strategy. Here, the points with 5% relative width are obtained by interpolating between the distributions with 2.5% and 10% relative width, and compared again to the fully simulated points with 5% relative width. Similar to the mass morphing, a good agreement is observed between the nominal

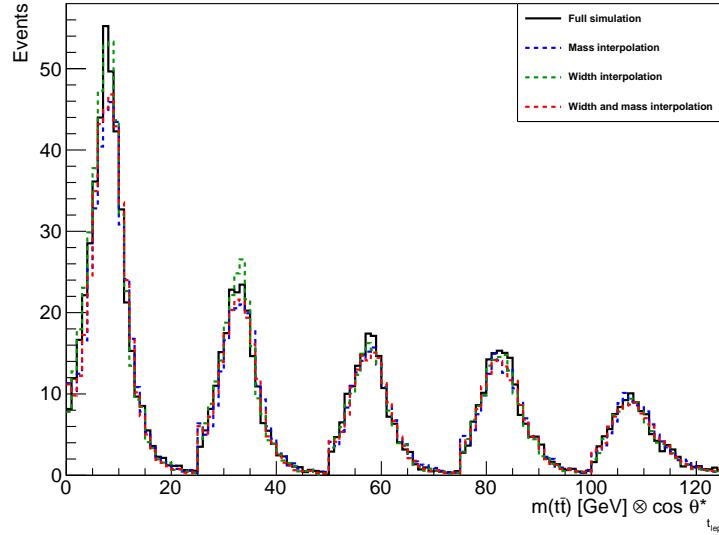


Figure 7.9: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the signal templates in the electron-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

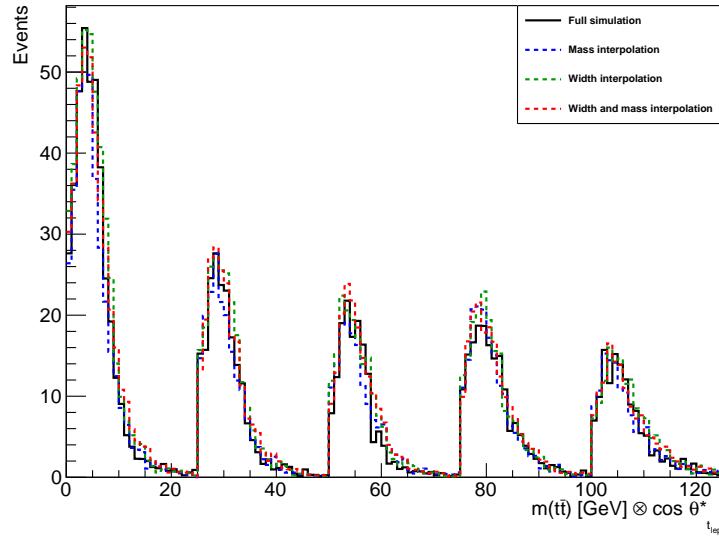


Figure 7.10: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the positive interference templates in the electron-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

templates and the ones obtained by the interpolation.

7.3.2 Width extrapolation

The analysis is sensitive to widths below the lowest produced relative width, which has a value of 2.5%. However, with the mass resolution being of order of 10–15% in the lepton-plus-jets-channel, and worse in the dilepton channel, it is expected that the signal lineshapes for widths with values of less than 2.5% are well reproduced by the lineshapes at 2.5% width. We validate this narrow-width approximation in the following using the samples with 2.5% and 5.0% relative width.

Figures 7.15–7.17 illustrate this by comparing the lineshapes for the different signal contributions (resonant,

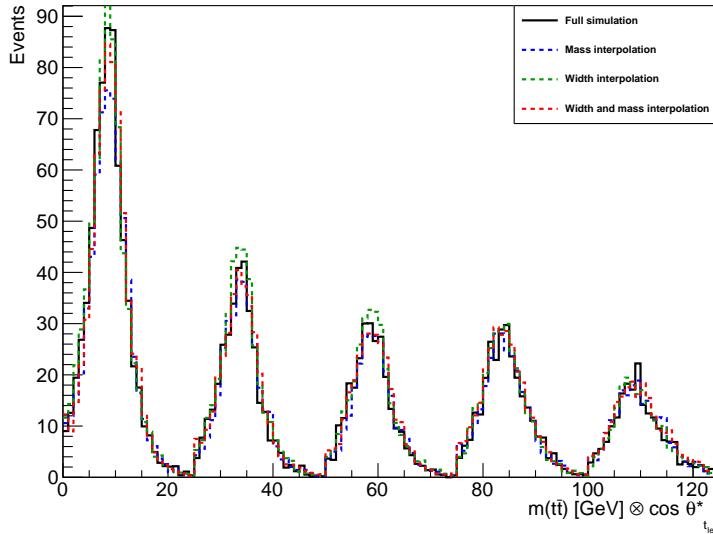


Figure 7.11: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the negative interference templates in the electron-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

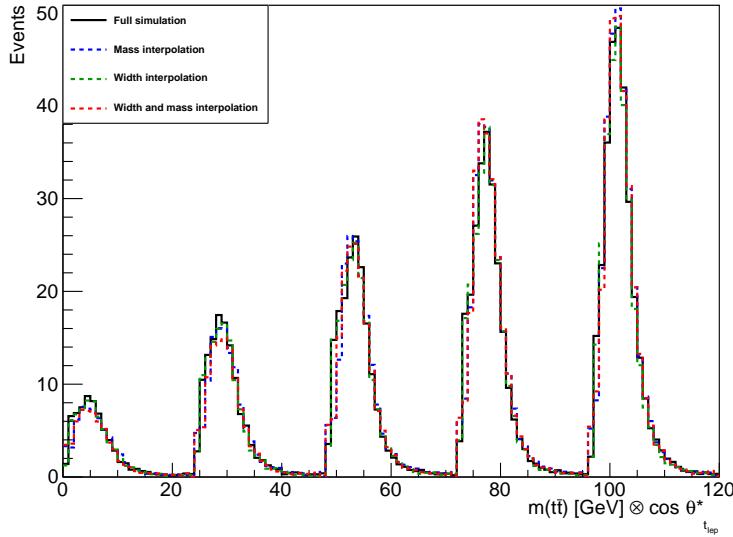


Figure 7.12: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the signal templates in the electron-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

positive interference, and negative interference) for different relative widths. The distributions for 2.5% and 5% relative width are very similar, whereas there are sizeable differences between the templates for larger widths.

To conclude, we construct additional templates at widths of 0.1 and 1%, and at intermediate widths if needed, by taking the lineshapes at 2.5% width and scaling the expected yield by the ratio of the expected cross sections (or expected yields for the interference) at the narrower width and at 2.5% width.

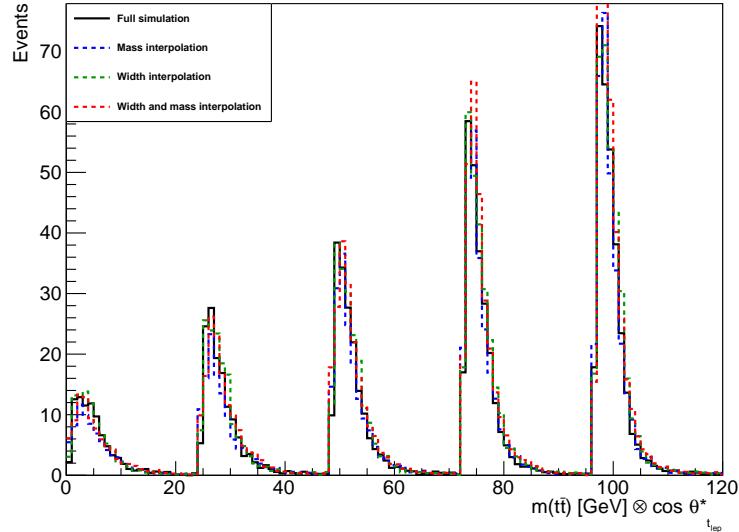


Figure 7.13: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the positive interference templates in the electron-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

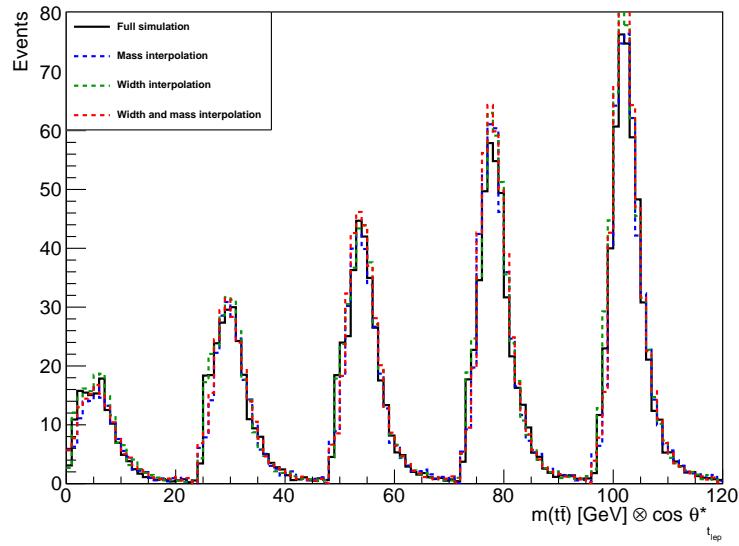


Figure 7.14: Comparison of unrolled $m_{t\bar{t}}$ distributions generated by mass morphing and width morphing with the leave-one-out strategy. The distributions are shown for the negative interference templates in the electron-plus-jets channel. The nominal sample at a mass of 500 GeV and a relative width of 5% (full simulation, red full line) is compared with the distribution obtained by interpolating between 400 and 600 GeV (blue dashed line) and the distribution obtained by interpolating between 2.5% and 10% relative width (green dashed line).

7.4 Results

The results in the following are obtained with the CL_S criterion, using the usual LHC test statistic and using the asymptotic approximation. The asymptotic approximation has been verified for single mass and width points by comparing to the limits obtained with the full CL_S method based on pseudoexperiments.

7.4.1 Model-independent limits

Model-independent upper limits on the coupling strength modifier are presented. The upper limits are derived as a function of width and mass. The results are given in two representations, either as function of mass in bins

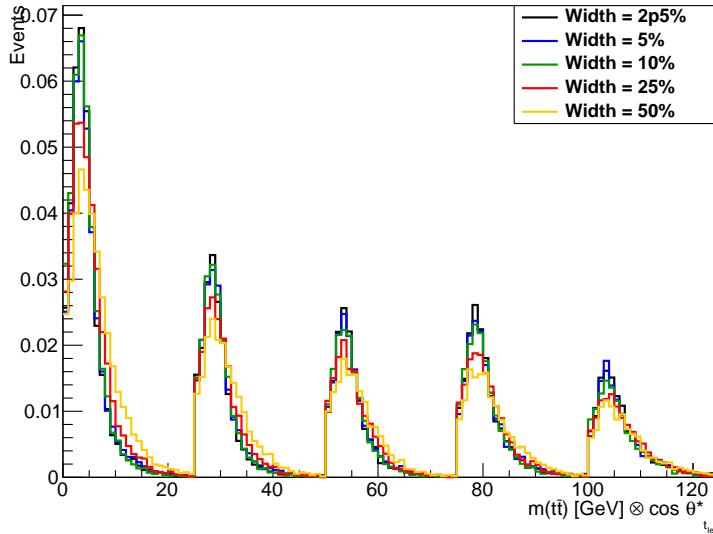


Figure 7.15: Comparison of unrolled $m_{t\bar{t}}$ distributions for different widths in the muon-plus-jets-channel and for a mass of 400 GeV, for the resonant signal contribution.

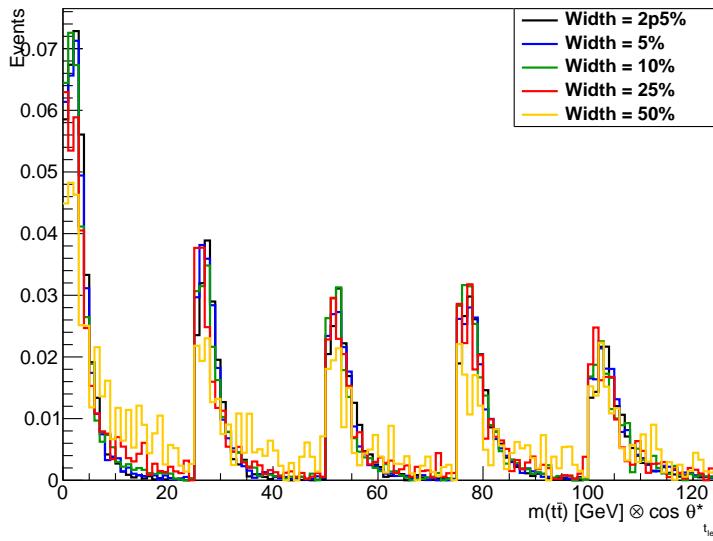


Figure 7.16: Comparison of unrolled $m_{t\bar{t}}$ distributions for different widths in the muon-plus-jets-channel and for a mass of 400 GeV, for the positive interference contribution.

of relative width, or vice versa. The upper limits on the coupling modifier are derived separately for the scalar and pseudoscalar signal models. All results in this section are for the combination of all final states. No scaling of the cross sections to higher orders is applied.

The results are shown in the following figures. All the results are consistent with the SM predictions and no new physics has been observed.

- Figure 7.18 shows the upper limits for pseudoscalar signal as a function of relative width in bins of mass.
- Figure 7.19 shows the upper limits for scalar signal as a function of relative width in bins of mass.
- Figure 7.20 shows the upper limits for pseudoscalar signal as a function of mass in bins of relative width.
- Figure 7.22 shows the upper limits for scalar signal as a function of mass in bins of relative width.

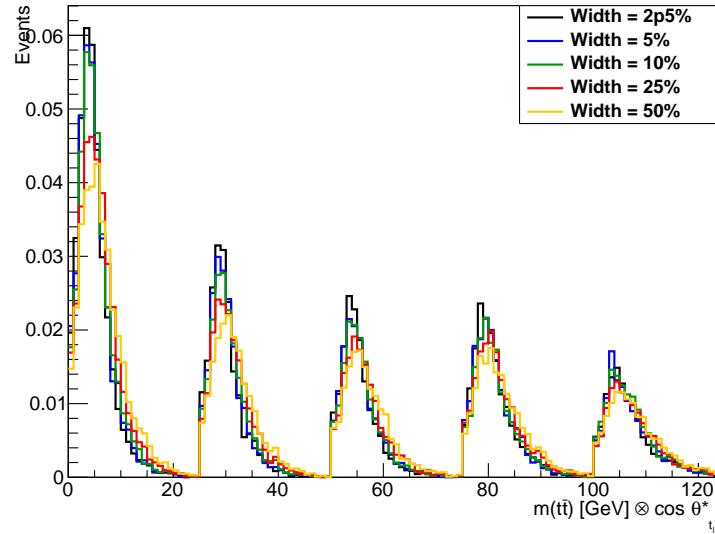


Figure 7.17: Comparison of unrolled $m_{t\bar{t}}$ distributions for different widths in the muon-plus-jets-channel and for a mass of 400 GeV, for the negative interference contribution.

7.4.2 Limits in the hMSSM

Limits in the hMSSM are produced with the following procedure. First, a grid of points in the $(m_A, \tan \beta)$ plane is produced. The current setup scans m_A between 400 and 750 GeV in steps of 50 GeV, and $\tan \beta$ between 0.4 and 10.0 in steps of 0.2. The lower limit on $\tan \beta$ of 0.4 is imposed to guarantee that the amplitudes preserve perturbative unitarity for all calculations.

For each such-produced $(m_A, \tan \beta)$ pair, the corresponding values of m_H and the widths of the A and H bosons are obtained with 2HDMC [111]. With the help of mass morphing and width interpolation/extrapolation where necessary, templates are constructed that correspond to the mass and width values obtained. With the such obtained templates, upper limits on the coupling strength modifier are calculated as outlined above. If the obtained limit on the coupling strength is smaller than the coupling strength that corresponds to the given $\tan \beta$, i.e. $g = 1/\tan \beta$, the given point is excluded. The ± 1 and $\pm 2\sigma$ bands of the expected limits are extracted in the same way.

Figure 7.23 shows the such-obtained limits in the $(m_A, \tan \beta)$ plane. The expected upper limits on $\tan \beta$ range from 2.3 at $m_A = 400$ GeV to 0.5 at $m_A = 700$ GeV. These are the first experimental limits in the region of low $\tan \beta$ and m_A beyond the $t\bar{t}$ production. The presented search hence gives access to a new region of phase space in searches for additional Higgs bosons in the MSSM context. While the presented results use the strict mass relations in the MSSM, they can easily be translated to more generic type-2 2HDM models.

7.5 Summary

Results of the combination of searches for additional Higgs bosons decaying to $t\bar{t}$ are presented. Upper limits are placed on the coupling modifier of these additional Higgs bosons to top quarks. These upper limits are given in a model-independent way in terms of the masses and widths of the either scalar or pseudoscalar new particles. In addition, the results are interpreted in the hMSSM (minimal supersymmetric standard model), which is an example of a two-Higgs doublet model of type 2 and considers the lightest additional scalar to correspond to the 125 GeV Higgs boson observed by the ATLAS and CMS experiments. Limits are given in the $(m_A, \tan \beta)$ plane. This analysis is sensitive in the low $\tan \beta$ region and has complementary sensitivity to previous searches, which are either sensitive at high $\tan \beta$ (di-tau final state) or for masses below the $t\bar{t}$ production threshold⁵.

⁵in particular searches for di-Higgs boson production and for pseudoscalars decaying to the 125 GeV Higgs boson and a Z boson, and constraints from the observed couplings of the 125 GeV Higgs boson

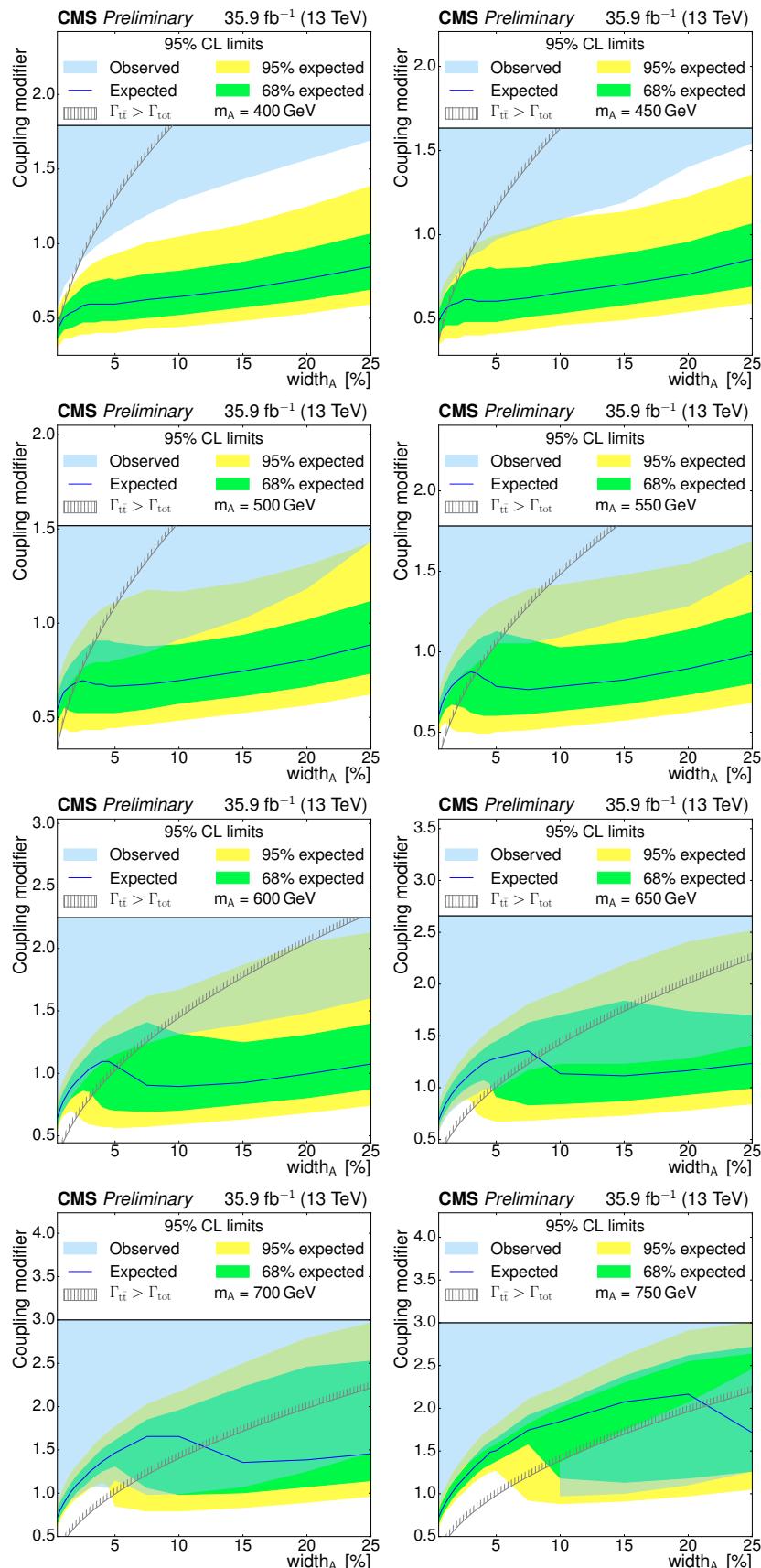


Figure 7.18: Model-independent combined limits on the coupling strength modifier as a function of relative width for masses between 400 and 750 GeV in steps of 50 GeV. The limits are derived for pseudoscalar signal only. The observed limits are shown by the blue shaded area. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. The region of phase space in which $\Gamma_{t\bar{t}} > \Gamma_{\text{tot}}$ is indicated by the hatched lines.

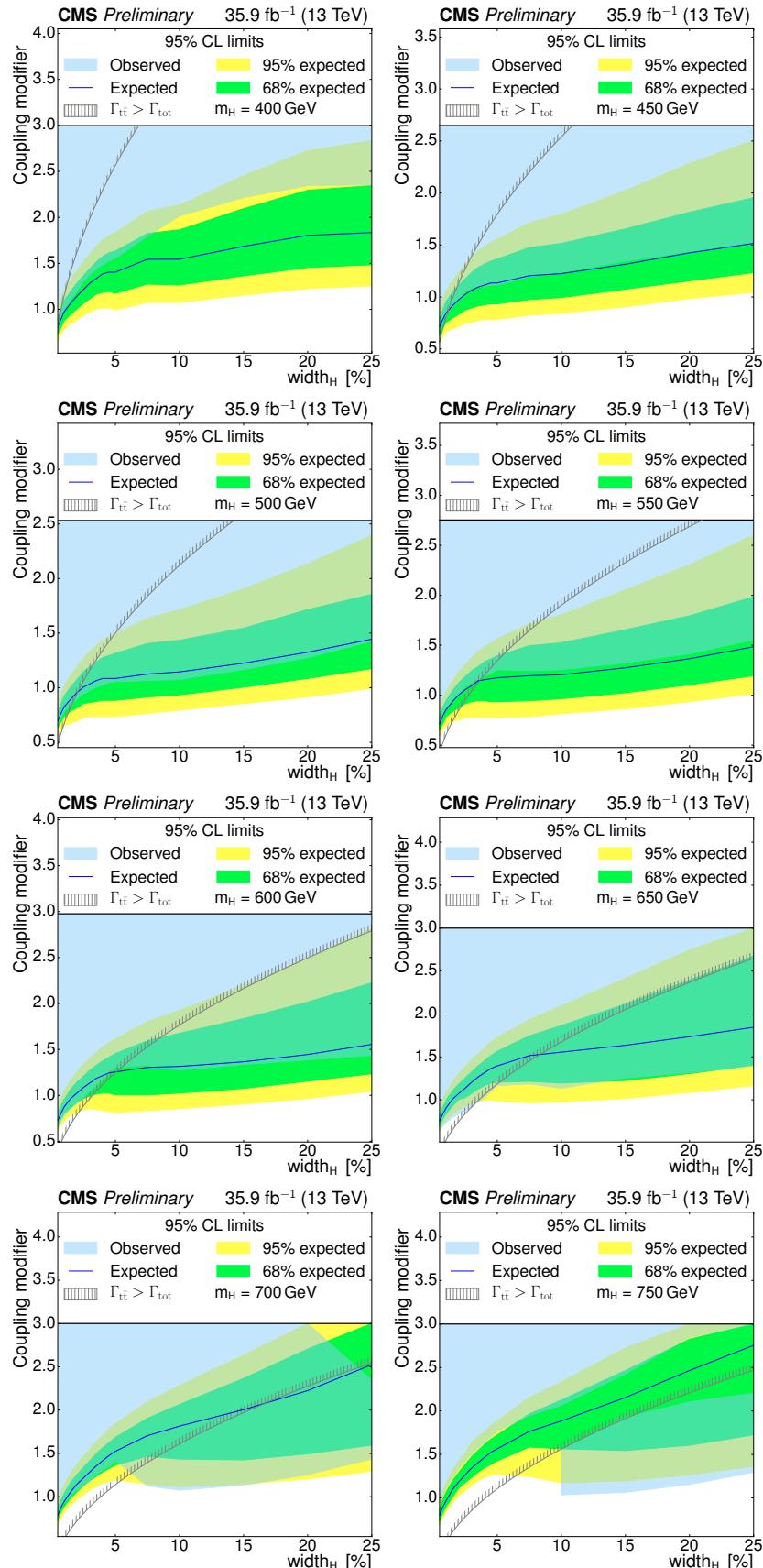


Figure 7.19: Model-independent combined limits on the coupling strength modifier as a function of relative width for masses between 400 and 750 GeV in steps of 50 GeV. The limits are derived for scalar signal only. The observed limits are shown by the blue shaded area. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. The region of phase space in which $\Gamma_{t\bar{t}} > \Gamma_{\text{tot}}$ is indicated by the hatched lines.

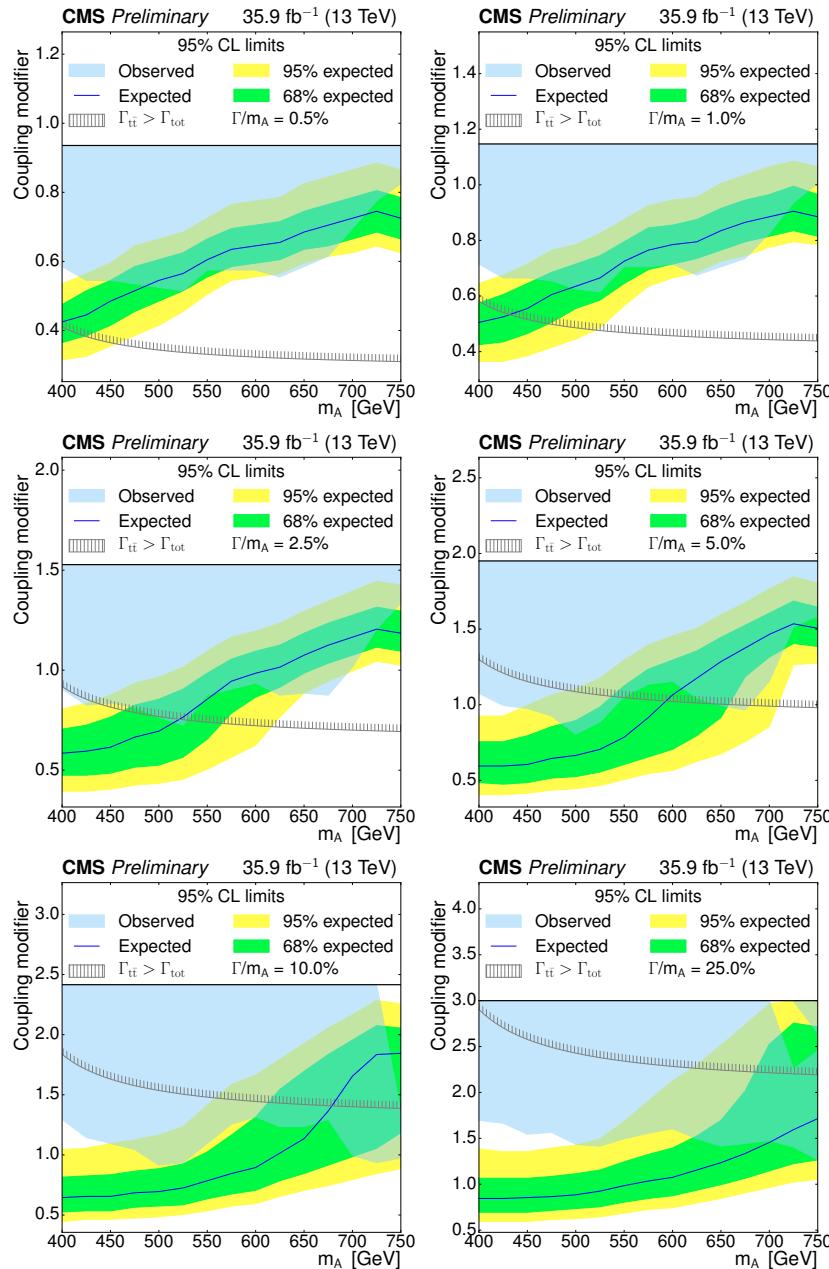


Figure 7.20: Model-independent combined limits on the coupling strength modifier as a function of mass for relative widths of 0.5, 1, 2.5, 5, 10, and 25%. The limits are derived for pseudoscalar signal only. The observed limits are shown by the blue shaded area. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. The region of phase space in which $\Gamma_{t\bar{t}} > \Gamma_{\text{tot}}$ is indicated by the hatched lines.

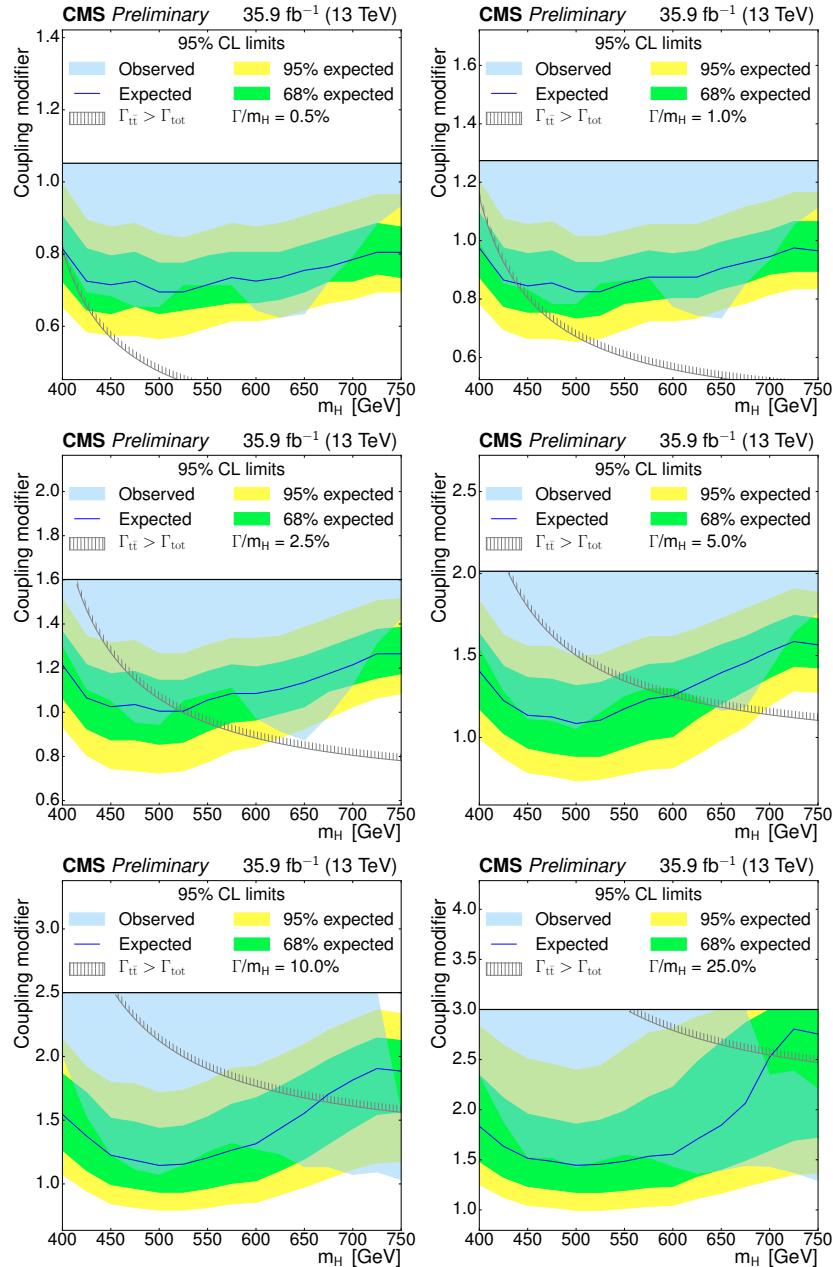


Figure 7.21: Model-independent combined limits on the coupling strength modifier as a function of mass for relative widths of 0.5, 1, 2.5, 5, 10, and 25%. The limits are derived for scalar signal only. The observed limits are shown by the blue shaded area. The inner (green) band and the outer (yellow) band indicate the regions containing 68 and 95%, respectively, of the distribution of limits expected under the background-only hypothesis. The region of phase space in which $\Gamma_{t\bar{t}} > \Gamma_{tot}$ is indicated by the hatched lines.

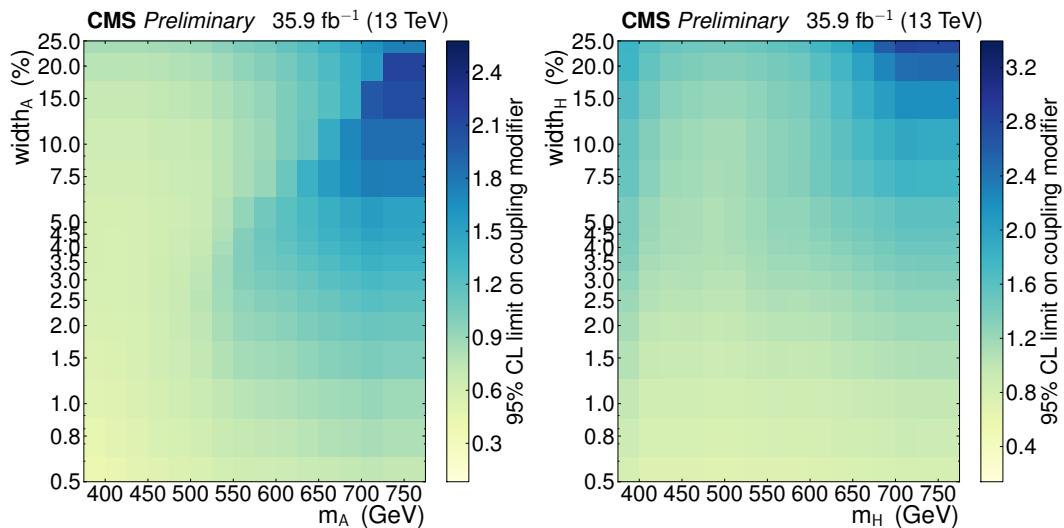


Figure 7.22: Model-independent combined upper limits on the coupling strength modifier in the plane of width and either pseudoscalar mass (left) or scalar mass (right). The limits are obtained separately for A production only (left) and for H production only (right).

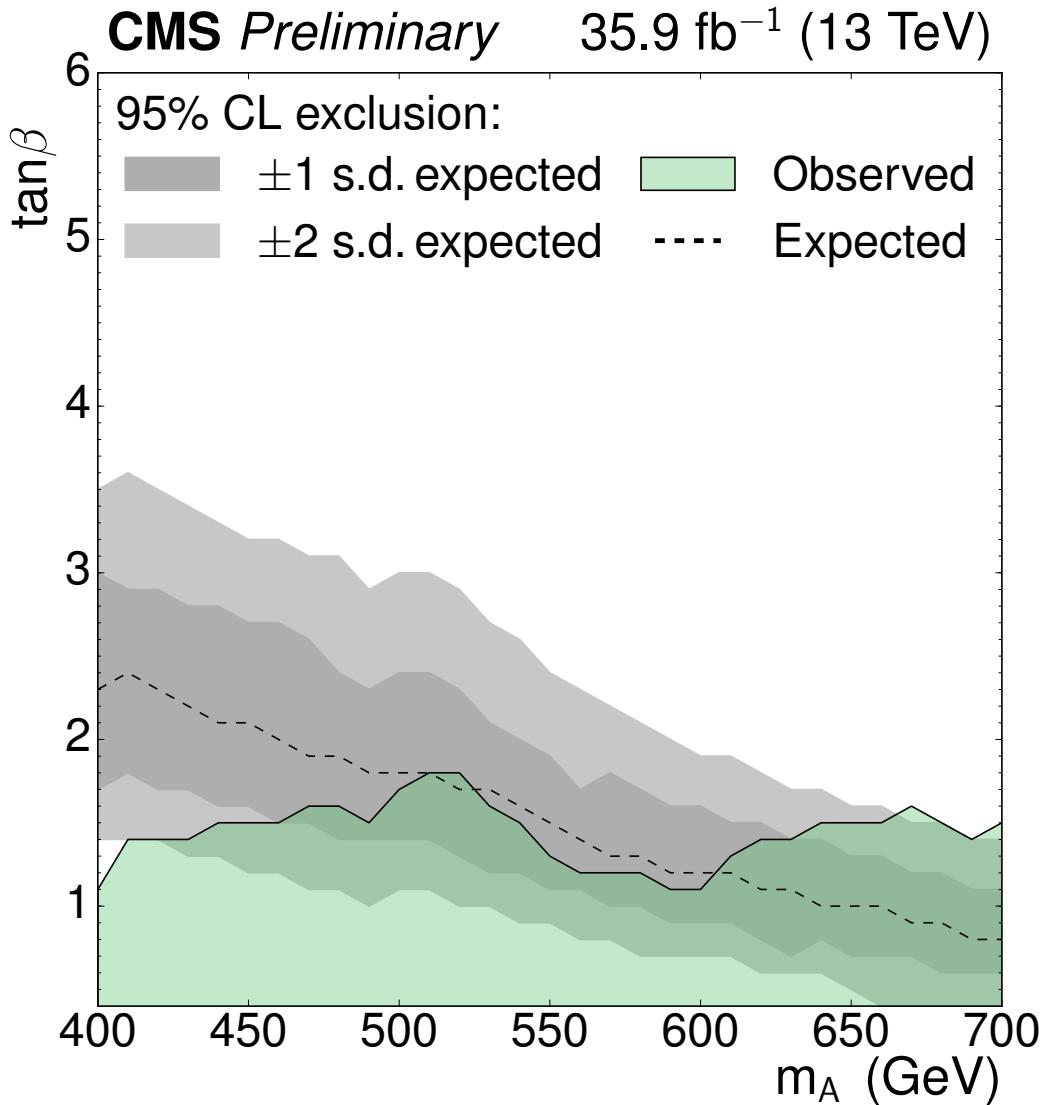


Figure 7.23: 95% confidence level limits on $(m_A, \tan \beta)$ in the hMSSM, using the combination of all channels and including both A and H signals with the masses and widths that correspond to a given point in the $(m_A, \tan \beta)$ plane.

8

Summary and Conclusion

Particle physics aims to describe the ultimate structure of matter at the level of its smallest constituents and the interactions among them. The Standard Model (SM) of particle physics describes all fundamental particles – leptons (electron, muon, tau, and their corresponding neutrinos), quarks (up, down, charm, strange, top, and bottom), gauge bosons (W^\pm , Z , γ), and a Higgs boson. The SM includes the interactions (electromagnetic, weak, and strong) and successfully explains how they act on matter particles by exchanging gauge bosons. However, the most familiar force in our everyday lives, gravity, is not part of the Standard Model. The SM is based on quantum field theory that was developed over the past sixty years and has been tested successfully by many different collider and non-collider experiments. Until recently, the last missing piece of the SM was the Higgs boson that was finally discovered in 2012 by the CMS and ATLAS experiments at the CERN Large Hadron Collider. This discovery represented the ultimate victory of the SM with the experimental verification of the mechanism of spontaneous electroweak symmetry breaking.

Particle detectors and accelerators have allowed for rapid progress in experimental particle physics research. The technology involved has evolved enormously over the last century, from the early Geiger counters and cloud chambers to today's most powerful Large Hadron Collider (LHC) accelerator and its state-of-the-art particle detectors at the European Laboratory for Particle Physics (CERN) in Geneva. The LHC began operation in 2009, and in spring 2010, the energy reached 3.5 TeV per beam, delivering 4.8 fb^{-1} by the end of 2011. In 2012, the per-beam energy increased to 4.0 TeV with an instantaneous luminosity of $6.5 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$. The machine delivered 23 fb^{-1} till the end of 2012, and this era is named as Run-I. In early 2013, the CERN accelerator complex shut down for two years of planned maintenance and consolidation, known as the first LHC Long Shutdown (LS1). Finally, the machine began its operations once again in 2015 at a record center-of-mass energy of 13 TeV and nominal instantaneous luminosity $5 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ using 50 ns bunch spacing. The bunch spacing decreased to 25 ns and the LHC machine delivered more than 80 fb^{-1} in 2016–17 successfully. The LHC will operate in 2018 to complete Run-II (2015–2018) with an expected integrated luminosity of more than 100 fb^{-1} . It will be followed by another Long Shutdown (LS2) in 2019–20 with essential upgrades to the LHC and the experiments. Run-III will span 2021–23 with an expected delivered luminosity of 300 fb^{-1} , which will be followed by the third Long Shutdown (LS3), allowing final upgrades of the machine and experiments before the High Luminosity LHC (HL-LHC) program. Starting in 2026, the HL-LHC is expected to provide an instantaneous luminosity of $5 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a potential peak value of $2 \times 10^{35} \text{ cm}^{-2} \text{ s}^{-1}$ at the beginning of fills.

The HL-LHC machine will induce a higher background radiation as compared to the present operating conditions, which will challenge the detectors. It is important to study the performance and stability of currently installed and future detectors in a high-radiation environment. With a focus on these requirements, the CERN Engineering- (EN) and Physics Department (PH) created a joint project – Gamma Irradiation Facility (GIF++), which is a new irradiation facility located north of the CERN Super Proton Synchrotron (SPS). It is a unique

place where high energy (≈ 100 GeV) charged particles (mainly muons) are combined with a high flux of gamma radiation (662 keV) produced by a 13.9 TBq ^{137}Cs source. It allows the cumulation of doses equivalent to HL-LHC experimental conditions within a reasonable time period.

The Gent group has been working on CMS RPCs since 2007 and is also involved in the upgrade of the CMS RPCs system. At HL-LHC, RPCs and associated electronics will operate at a much higher luminosity that will produce ageing effects. In order to study the ageing effects of the RPCs and its electronics for the HL-LHC upgrade program, the CMS RPC community installed several different types of RPC detectors at the GIF++. A dedicated Detector Control System (DCS) has been developed using the WINCC-OA tool to control and monitor these detectors and store the measured parameters' data. WINCC-OA is an open architecture adopted by all complicated experiments at the LHC and relies consistently on object orientation to process images and the database structure. This enables efficient and simple mass engineering and swift creation of projects with a number of parallel developments. The performance of the CMS RPC chambers at GIF++ has been studied and compared at different radiation levels. At a rate of $600 \text{ Hz}\cdot\text{cm}^{-2}$, i.e. the maximum rate expected at the HL-LHC, the maximum efficiency of the chamber is 95%. The detector performance is stable at 34% of the total accumulated charge expected at the HL-LHC.

The discovery of the Higgs boson (spin-zero resonance) with a mass of 125 GeV by the CMS and ATLAS experiments at the Large Hadron Collider (LHC) completes the SM of particle physics. However, it doesn't eliminate the possibility that additional Higgs states with masses above or below 1 TeV exist. Higgs mass hierarchy is one of the well-known motivations for considering the existence of an extended Higgs sector in SM extensions, such as the MSSM and 2HDM. These models predict more than one Higgs state; for example, in the MSSM, there are five Higgs states – two of them are charged (H^\pm) and three are neutral with one SM like (h), one heavy scalar (H), and one pseudoscalar (A). The fact that the properties of the discovered particle seem to be compatible with the expectations of the SM demands that any such model reproduces the SM Higgs boson, which is usually identified using the lightest state available. Such a configuration is referred to as the alignment limit.

The top quark is one of the heaviest particles in SM and is produced mainly in a pair at the LHC. It decays exclusively into a W boson and a b quark, where the W boson further decays into two jets or leptons and its neutrino. The decay mode (leptonic or hadronic) of the W bosons from the two top quark characterizes the top pair decay channels as semileptonic, dileptonic, or fully hadronic. Its large mass naturally leads to a strong coupling between the top quark and the additional Higgs bosons. As a result, the decay into a pair of top quarks can have a large branching ratio when kinematically allowed. This decay channel is especially favourable for a \mathcal{CP} -odd state, which cannot couple to the weak bosons. However, in the 2HDM model, couplings of the heavy \mathcal{CP} -even state to W and Z bosons also vanish in the alignment limit; in the MSSM, they are suppressed when $m_A \gg m_Z$ (the decoupling limit). Owing to this and other factors, various theories beyond the standard model (BSM) can manifest themselves in the $t\bar{t}$ final state. Studying it experimentally, however, is a challenging task because of the large SM $t\bar{t}$ background and the need to reconstruct decays of the top quarks with high precision. A number of searches for resonances in the $m_{t\bar{t}}$ spectrum have been performed, but no positive results have been obtained so far. An important feature of the $pp \rightarrow t\bar{t}$ process in the presence of a heavy Higgs boson is the interference with the SM continuous background. Because of a non-trivial phase in the involved loop-induced coupling of the Higgs boson to gluons, the $m_{t\bar{t}}$ lineshape can show a peak-dip, dip-peak, or dip-only structure instead of the usual resonance peak. The spectrum of possible deviations becomes even richer when one includes BSM particles in the loop, such as top squarks, or allows for a violation of the \mathcal{CP} symmetry in the Higgs sector. It should be noted that even in the limit of a narrow resonance, the impact of the interference is not negligible as it changes the size of the peak.

The analysis presented in this thesis focuses on the search for a heavy Higgs boson decaying into a pair of top quarks in the semileptonic final state, using around 36 fb^{-1} of pp collision data collected in 2016. The selected events contain a muon or an electron and at least four jets – at least two of which must be b-tagged. The search targets two pure \mathcal{CP} states with masses 400, 500, 600, and 750 GeV, where each mass point has five different widths (2.5, 5, 10, 25, 50)% . In total, 40 signal samples ($2 \times 4 \times 5$) for resonance and 40 for interference are produced. The interference with the QCD $t\bar{t}$ production is found to play an important role and is explicitly taken into account.

The production of interference samples involves the modification of Fortran code generated by MADGRAPH 5_A-MC@NLO. Let $C_{\Phi tt}$ be the coupling of the heavy Higgs boson to top quarks, which is utilized by the routine that evaluates the squared matrix element $|\mathcal{M}|^2$. The code is modified at the point where this routine is called, in the following manner. First, $|\mathcal{M}|^2$ is evaluated for the nominal value of $C_{\Phi tt}$ and saved. Then, the $C_{\Phi tt}$ sign is flipped and $|\mathcal{M}|^2$ is computed again, yielding the value of $|\mathcal{M}(-C_{\Phi tt})|^2$. The effective coupling

of Φ to gluons is controlled by an independent parameter in the routine and therefore is not affected by the modification of $C_{\Phi tt}$. As a result, the interference terms in $|\mathcal{M}|^2$ flip the sign, while the SM terms (independent of the Φ couplings) and the resonant BSM part (proportional to $C_{\Phi tt}^2$) are left unchanged. Finally, the value of $(\mathcal{M}^2(C_{\Phi tt}) - \mathcal{M}^2(-C_{\Phi tt})) / 2$ is computed and used replacing the original squared matrix element everywhere. Everything but the interference terms are cancelled out in this computation.

The $t\bar{t}$ system is reconstructed by utilizing kinematic constraints imposed by the known masses of the top quark and the W boson. The dominant background is $t\bar{t}$ production, which is modelled with POWHEG v2 and PYTHIA 8. The remaining minor backgrounds are modelled as single top tw-channel with POWHEG and PYTHIA 8; single top t-channel with POWHEG v2 and MADSPIN; single top s-channel and TTZ with AM-CATNLO and PYTHIA 8; wjets and Drell-Yan with MADGRAPH and PYTHIA 8; di-boson with PYTHIA 8; TTZ and TTWjets with AM-CATNLO; and MADSPIN and PYTHIA 8.

The multijet background is impossible to model from simulation and therefore has been determined from data using data-driven techniques. The number of events from the multijet background is estimated using the so-called ABCD method. The method is applied for the M_T^W variable 6.13 and the lepton's relative isolation I . The shape of the (two-dimensional) distribution of the search variables for the multijet background is modelled from a region with an inverted selection on the lepton's relative isolation I . The expected contribution from backgrounds that produce prompt leptons is subtracted from the observed data; the resulting distribution is then attributed to the multijet background.

Corrections to simulations, i.e pile-up re-weighting, jet energy scale and resolution, b-tagging, triggers and lepton efficiencies, and generator level weights are applied using CMS standard procedures. Top p_T correction is applied using an empirical reweighting based on the observed p_T distributions of top quarks. The empirical procedure is further compared with the theory prediction; top p_T uncertainty is considered. One of the dominant uncertainties in this analysis originates from the jet calibration. This is evaluated by varying the multiplicative jet energy correction within its uncertainties. Jet Energy Resolution (JER) data-to-simulation scale factors are also simultaneously varied within their uncertainties for all jets. Uncertainties related to b-tagging are evaluated by varying the values of the scale factors within the respective uncertainties. Uncertainties related to trigger scale factors and scale factors of lepton identification are taken into account. Integrated-luminosity uncertainty is considered as 2.5%. The normalization of the data-driven template for the multijet background is varied conservatively by $+100/-50\%$, independently in the muon and electron channels. Renormalization and factorization scale (μ_R and μ_F) uncertainties are taken from the independent variations of the two scales by a factor of two in each direction. Uncertainties related to Initial and Final State Radiation (ISR, FSR) and Underlying Event (UE) are taken from dedicated samples. For top quark mass uncertainties, two simulated samples of $t\bar{t}$ are used with top quark mass $m_t = 169.5$ and 175.5 GeV. The h_{damp} parameter in POWHEG, which controls the suppression of radiation of additional high- p_T jets, is changed from its nominal value of $1.58m_t$ to $0.99m_t$ and $2.24m_t$. Statistical uncertainties on the dominant $t\bar{t}$ background are taken into account. The variations in ISR and UE do not produce statistically significant changes in the shape of the distribution and are consequently dropped. The Probability Density Functions (PDF) uncertainties are computed from the set NNPDF3.0. They are described by 100 alternative versions of the PDF, referred to as “MC replicas”, which are provided in the form of dedicated event weights.

To probe for the presence of the hypothesized $\Phi(\text{H/A})$ particle, a two-dimensional distribution of the mass of the reconstructed $t\bar{t}$ system, $m_{t\bar{t}}$ and an angular variable $|\cos \theta^*|$ that reflects the spin of the s -channel resonance are investigated. The first variable is the reconstructed invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$, which serves as a proxy for the mass of the heavy Higgs boson. The second variable is $|\cos \theta^*|$, where θ^* is the angle between the three-momentum of the leptonically decaying top quark in the $t\bar{t}$ rest frame and the three-momentum of the $t\bar{t}$ system in the lab frame.

For the statistical analysis, the Higgs combine tool, based on RooFit/RooStat-based software tools, is used. RooFit is a toolkit integrated with ROOT, which performs different kinds of fits and toy Monte Carlo generations based on user-defined models of the PDF. The interference template, the 2D distribution of the $m_{t\bar{t}}$, and $|\cos \theta^*|$, has both positive- and negative-weighted events. Due to a limitation of the Higgs combine tool, which can only handle non-negative templates, and for the purpose of interpolation between reference values of m_Φ , the interference template is split into two parts according to the sign of the weight of the contributing events. The sign of one of the templates is inverted so that both are non-negative. The two templates are scaled in the model by factors g^2 and $-g^2$ respectively (g is the coupling scale factor and for interference, it is proportional to g^2), thus reproducing the desired distribution. With no indication of the presence of the BSM signal, upper limits on the coupling scale factors are computed for various signal hypotheses in the semileptonic channel.

To include the interference effect, leading order (LO) MADGRAPH is used for signal generation; thereafter,

the signal samples are scaled for higher order (NNLO) cross sections. We use the SUSHI program interfaced with a 2HDMC calculator to calculate scale factors for the resonance part using the hMSSM context; for the interference part, the TOP++ calculator is used for performing SM $t\bar{t}$ cross-section calculation. The k-factor is calculated as the ratio of NNLO to LO cross section and applied to the signal cross sections calculated by MADGRAPH. The interference k-factor is taken as the geometric mean of the resonance and the SM $t\bar{t}$ k-factors. For a single mass point, the k-factors show a constant behaviour for all the considered width hypotheses.

The presented search is sensitive to the high mass of Φ and low $\tan \beta$, where hMSSM is proposed as it takes the existence of the 125 GeV Higgs boson directly into account. In the case of the hMSSM, it is impossible to fully cover the relevant part of the $(m_A, \tan \beta)$ spectrum with a limited set of samples in terms of fixed masses and widths, such as the ones we produced for this analysis. We used the ROOMOMENTMORPH algorithm with which the distributions are interpolated between different masses and widths. The distributions are also extrapolated to lower widths within the narrow-width approximation.

The semileptonic final state has been combined with the dileptonic channel and upper limits are placed on the coupling modifier of these additional Higgs bosons to top quarks using the independent and combined channels. In addition, the results are interpreted in the context of the hMSSM in the $(m_A, \tan \beta)$ plane for the combined channels. The results are compatible with the theory predictions within systematic uncertainties. This analysis is sensitive in the low $\tan \beta$ region and has complementary sensitivity to previous searches, which are either sensitive at high $\tan \beta$ (2τ final state) or for masses below the $t\bar{t}$ production threshold¹.

So far, no new physics has been found and the results are consistent with the SM theory prediction within systematic uncertainties. The channel could benefit from higher statistics and precise measurement of the $t\bar{t}$ mass. Using full Run-II data will further increase the sensitivity in this channel while at 300 fb^{-1} and 3.0 ab^{-1} – data collected during Run-III and HL-LHC respectively – the $gg \rightarrow H \rightarrow t\bar{t}$ channel will play a pivotal role in heavy Higgs research. Because of the same final state topology, SM $t\bar{t}$ is the major background in this channel. To increase the sensitivity, we need to properly reduce the SM $t\bar{t}$ background using more signal-to-background discriminants. This will be done by using MVA techniques and properly taking into account the negative weights associated with the interference. The k-factors for interference are calculated as the geometric mean of the resonance and SM $t\bar{t}$ k-factors. In the next iteration, our aim is to calculate NLO k-factors for interference in a manner such that the line shape of the k-factors includes the interference effect. Different signal line shapes, such as pure dip, bump, or even nothingness, will also help to explore this channel.

¹In particular, searches for di-Higgs boson production and pseudoscalars decaying to the 125 GeV Higgs boson and a Z boson, and constraints from the observed couplings of the 125 GeV Higgs boson

9

Nederlandstalige samenvatting

Binnen de deeltjesfysica worden de kleinste bouwstenen van materie evenals hun onderlinge interacties bestudeerd. Het Standaardmodel (SM) van de deeltjesfysica beschrijft alle fundamentele deeltjes die we tot op heden kennen: de leptonen (electron, muon, tau en hun respectievelijke neutrino's), de quarks (up, down, charm, strange, bottom en top), de ijkbosonen (W^\pm , Z , γ) en het Higgs boson. Alle zichtbare materie in het Universum is uit quarks en leptonen uitgebouwd, terwijl de interacties tussen deze deeltjes succesvol in het SM beschreven wordt door middel van uitwisseling van ijkbosonen. Het SM omvat de electromagnetische, zwakke en sterke wisselwerking, maar geeft geen beschrijving van de meest gekende wisselwerking uit ons dagelijks leven, nl. de zwaartekracht. Het SM is gebaseerd op een kwantumveldentheorie die tijdens de voorbije 60 jaar ontwikkeld werd. Dit model werd gedurende al die jaren succesvol getoetst aan de resultaten van talrijke experimenten, zowel aan deelversnellers als daarbuiten. Tot voor kort was het laatste ontbrekende stukje van het SM het zogenaamde Higgs boson, waarvan het bestaan uiteindelijk bevestigd werd in 2012 door de CMS en ATLAS experimenten aan de Large Hadron Collider in het CERN. Deze ontdekking betekende het ultieme succes van het SM met de experimentele verificatie van het mechanisme voor spontane electrozwakke symmetriebreking.

De opkomst en ontwikkeling van de deeltjesversnellers en -detectoren zorgde voor een enorme vooruitgang in de experimentele deeltjesfysica. Gedurende de voorbije eeuw evolueerde de gebruikte technologie uitermate snel, gaande van de eerste geigertellers en nevelkamers tot de hedendaagse ultrakrachtige Large Hadron Collider (LHC) en bijhorende ultra geavanceerde detectoren in het CERN, het Europese laboratorium voor deeltjesfysica te Genève in Zwitserland. De LHC werd in 2009 opgestart en in de lente van 2010 werd reeds een energie van 3.5 TeV per protonenbundel gehaald, d.w.z. een totale botsingsenergie van 7 TeV, waarmee tegen eind 2011 een geïntegreerde luminositeit van 4.8 fb^{-1} aan de experimenten geleverd werd. In 2013 werd de bundelenergie verhoogd naar 4 TeV met een ogenblikkelijke luminositeit van $6.5 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, waarbij tegen eind 2013, op het einde van LHC Run-I 23 fb^{-1} afgeleverd werd. Na Run-I werd er een periode van 2 jaar ingelast voor onderhoud en opwaarderingen van de experimenten (Long shutdown 1, LS1). Toen de LHC bij de start van Run-II in 2015 opnieuw in werking trad, werd voor het eerst een record massamiddelpuntssenergie van 13 TeV bereikt. De ogenblikkelijke luminositeit bedroeg $5 \cdot 10^{33} \text{ cm}^{-2}\text{s}^{-1}$, waarbij de tijd tussen opeenvolgende protonpakketjes in de LHC deeltjesbundels 50 ns was. Deze tijd werd verlaagd naar 25 ns in 2016-2017 waarmee een totale luminositeit van 80 fb^{-1} behaald werd. De momenteel aan de gang zijnde Run-II loopt nog door tot eind 2018, waarbij een totale luminositeit van meer dan 100 fb^{-1} verwacht wordt. Tijdens de volgende, tweede lange periode waarbij de LHC in 2019-2020 stilgelegd zal worden (LS2), zullen er belangrijke opwaarderingen zowel aan de versneller als aan de experimenten gebeuren (fase-I upgrade). Vervolgens zal LHC Run-III starten in 2021 en duren tot 2023, waarbij er een geïntegreerde luminositeit van 300 fb^{-1} verwacht wordt. Hierna zal de LHC opnieuw afgezet worden voor een tweede fase van upgrades om tot de zogenaamde High Luminosity fase van het LHC programma (HL-LHC) te komen. De HL-LHC die in 2023

van start zal gaan zou een ogenblikkelijke luminositeit van $5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ met pieken tot $2 \cdot 10^{35} \text{ cm}^{-2}\text{s}^{-1}$ bereiken.

De hogere luminositeit van de HL-LHC in vergelijking met de huidige LHC condities vormt een uitdaging voor de detectoren die blootgesteld zullen worden aan hogere hoeveelheden straling. Het is daarom dan ook belangrijk om het functioneren en de stabiliteit van de detectoren te bestuderen onder dergelijke extreme omstandigheden. Dit kan o.a. gedaan worden aan de zogenaamde Gamma Irradiation Facility (GIF++), ingericht in het CERN, in het noordelijke deel van het Super Proton Synchrotron (SPS). In de GIF++ kunnen detectoren blootgesteld worden aan een bundel van hoogenergetische ($\approx 100 \text{ GeV}$) geladen deeltjes, voornamelijk muonen, in combinatie met een achtergrondstraling bestaande uit fotonen met energie 662 keV (afkomstig van een 13.9 TBq ^{137}Cs bron). Hiermee kan men de omstandigheden nabootsen waaraan de detectoren blootgesteld zullen worden tijdens de HL-LHC fase. De Gentse onderzoeks groep Experimentele Deeltjesfysica is sinds 2007 actief betrokken bij de ontwikkeling en opwaardering van de zogenaamde Resistive Plate Chambers (RPCs) in het CMS Muon systeem. Om de kwaliteit alsook het verouderingsproces van deze RPCs zoals verwacht wordt tijdens de HL-LHC te testen, worden deze onderzocht in de GIF++. Tijdens dit doctoraatswerk werd een controlesysteem (Dedicated Control System, DCS) gebruik makende van WINCC-OA ontwikkeld voor de CMS RPC opstelling in de GIF++. Dit systeem laat toe om op een efficiënte en snelle manier alle detectoren vanop afstand te besturen en op te volgen. Het gedrag van CMS RPCs, zowel bestaande als nieuw ontworpen detectoren voor de HL-LHC, werd op deze manier bestudeerd, waarbij een vergelijking van de detectorprestaties bij verschillende stralingsniveau's gemaakt werd. Bij een deeltjesflux van 600 Hz/cm^2 haalden de RPCs nog steeds een maximum detectorefficiëntie van 95 %. Met een tot op heden geabsorbeerde stralingsdosis die overeenkomt met 34 % van de totale dosis verwacht met de HL-LHC, blijven de detectorprestaties blijven nog steeds stabiel.

De ontdekking door de CMS en ATLAS experimenten aan het LHC van het Higgsdeeltje, een boson met spin 0 en een massa van $\approx 125 \text{ GeV}$, zorgde voor de ultieme bevestiging van de geldigheid van het Standaardmodel van de deeltjesfysica. Het is echter niet uitgesloten dat er naast dat ene Higgsdeeltje nog meer, zwaardere Higgsbosonen bestaan. De extensie van de Higgs sector is één van de belangrijkste mogelijkheden tot uitbreidings van het Standaardmodel, in bijvoorbeeld de MSSM en 2HDM modellen. In het MSSM model worden er b.v. vijf Higgsbosonen voorspeld, waarvan twee geladen (H^\pm) en drie neutraal zijn. Één van de neutrale deeltjes zou het SM Higgs boson kunnen zijn (h), één een zwaar scalair deeltje en het derde een pseudoscalair deeltje. De eigenschappen van het huidig waargenomen (SM) Higgsdeeltje moeten dus ook gereproduceerd worden in elk van de modellen die uitbreidingen van het Standaardmodel voorstellen; dit is de zogenaamde *alignment limit*.

Het topquark (t) is één van de zwaarste deeltjes van het SM en wordt voornamelijk geproduceerd als onderdeel van een top-antitop quarkpaar. Het vervalt voornamelijk naar een W boson en een bottomquark (b). Het W boson vervalt op zijn beurt verder ofwel hadronisch naar twee deeltjesbundels, zogenaamde jets, of leptonisch naar een lepton en het bijhorende neutrino. Het vervalkanaal van de W bosonen (leptonisch of hadronisch) gecreëerd door de twee topquarks zorgt voor volgende klasseringen van vervalmethoden: bi-leptonisch, semi-leptonisch of hadronisch. Omwille van de grote massa van het topquark is er een grote koppeling tussen het topquark en de Higgs bosonen. Dit zorgt ervoor dat een significante fractie van de Higgsdeeltjes vervalt naar twee topquarks als dit kinematisch toegelaten is. Dit vervalkanaal is voornamelijk belangrijk bij een \mathcal{CP} -oneven deeltje waarbij er niet kan gekoppeld worden aan de zwakke ijkbosonen. In het 2HDM model echter verdwijnen de koppelingen van een zwaar \mathcal{CP} -even deeltje in de "alignment limit", en in het MSSM model zijn ze onderdrukt wanneer $m_A \gg m_Z$ (de ontkoppelingslimiet). Omwille hiervan, maar ook om nog andere redenen, zouden bepaalde theoriën over de uitbreiding van het Standaardmodel (zogenaamde *Beyond the Standard Model*, BSM theoriën) zichtbaar kunnen worden bij de productie van $t\bar{t}$ paren. In dit $t\bar{t}$ vervalkanaal is er echter een enorme achtergrond omwille van andere SM productieprocessen, waardoor dit experimenteel waarnemen valt en de topquarks met hoge precisie gereconstrueerd dienen te worden. Er werden al een aantal analyses uitgevoerd waarbij gezocht werd naar resonanties in het $m_{t\bar{t}}$ spectrum, maar dat leverde tot nu toe nog geen positief resultaat. Een belangrijk kenmerk van het $pp \rightarrow t\bar{t}$ proces in aanwezigheid van een zwaar Higgs boson is de interferentie met de SM achtergrond. Omwille van een niet-triviale fase in de betrokken "loop"-geïnduceerde koppeling tussen het Higgs boson en gluon en kan het $m_{t\bar{t}}$ spectrum in plaats van de typische resonantie een piek-dip, dip-piek of enkel een dip patroon vertonen. Dit spectrum wordt zelfs nog verder uitgebreid wanneer BSM deeltes, zoals top squarks, in de loop voorkomen of wanneer \mathcal{CP} symmetriebreking in de Higgs sector toegelaten wordt. Hierbij dient opgemerkt te worden dat zelfs in de limiet van een smalle resonantie de impact van interferentie niet verwaarloosbaar is, aangezien dit de hoogte van de piek beïnvloedt.

De analyse die in deze thesis voorgesteld wordt ging op zoek naar een zwaar Higgsdeeltje dat vervalt naar

een topquarkpaar in het semileptonisch vervalkanaal. Er werd hierbij gebruik gemaakt van de CMS 2016 proton-proton botsingsdataset met een totale luminositeit van 36 fb^{-1} . De geselecteerde botsingsevenementen bevatten een muon of een electron en ten minste vier jets. Hiervan dienden er twee positief geselecteerd te zijn door het *b-tagging* algoritme. De analyse beschouwde twee pure \mathcal{CP} toestanden met massa's 400, 500, 600 and 750 GeV, elk met verschillende breedtes (2.5, 5, 10, 25, 50)%. In totaal werden 40 ($2 \times 4 \times 5$) signaal datasets voor de resonantie en 40 voor de interferentie geproduceerd. De interferentie met de QCD $t\bar{t}$ productie blijkt een belangrijke rol te spelen en werd hier expliciet in rekening gebracht.

De productie van de interferentie datasets vereiste dat de FORTRAN code gegenereerd door MADGRAPH5-A-MCNLO aangepast werd. $C_{\Phi tt}$ stelt hier de koppeling van het zware Higgs boson met de topquarks voor, die gebruikt wordt door het deel van de code dat het gekwadrateerde matrix element $|\mathcal{M}|^2$ evalueert. Deze code werd dan als volgt aangepast. Eerst wordt $|\mathcal{M}|^2$ geëvalueerd voor de nominale waarde van $C_{\Phi tt}$. Daarna wordt het teken van $C_{\Phi tt}$ omgekeerd, en wordt $|\mathcal{M}|^2$ nogmaals berekend, wat $|\mathcal{M}(-C_{\Phi tt})|^2$ geeft. De effectieve koppeling van Φ met gluonen wordt gecontroleerd door een onafhankelijke parameter in de code en wordt daardoor niet beïnvloed door de aanpassing van $C_{\Phi tt}$. Het resultaat is dat de interferentietermen in $|\mathcal{M}|^2$ hun teken omwisselen, terwijl de SM termen (onafhankelijk van de Φ koppelingen) en het resonante BSM deel (evenredig met $C_{\Phi tt}^2$) onveranderd blijven. Uiteindelijk wordt de waarde $(\mathcal{M}^2(C_{\Phi tt}) - \mathcal{M}^2(-C_{\Phi tt})) / 2$ berekend, wat het originele kwadratisch matrix element vervangt. Alle termen uitgezonderd de interferentietermen vallen weg uit deze berekening.

Het $t\bar{t}$ systeem werd gereconstrueerd door gebruik te maken van de kinematische beperkingen te wijten aan de kennis van de massa van het topquark en het W boson. De belangrijkste achtergrond is de $t\bar{t}$ productie, wat gemodelleerd werd met POWHEG v2 en PYTHIA 8. De overblijvende (kleinere) achtergronden werden dan als volgt gemodelleerd: enkelvoudig top tw-kanaal met POWHEG en PYTHIA 8, enkelvoudig top t-kanaal met POWHEG v2 en MADSPIN, enkelvoudig top s-kanaal en TTZ AMCATNLO en PYTHIA 8, W jets en Drell-Yan met MADGRAPH en PYTHIA 8, di-boson met PYTHIA 8, TTZ en TTW jets met AMCATNLO, MADSPIN en PYTHIA 8.

De multijet achtergrond is onmogelijk om te modelleren via simulaties en werd hier bepaald via technieken vertrekende van de proton-proton dataset zelf. Het aantal evenementen te wijten aan multijet achtergrond werd geschat met de zogenoemde ABCD methode. De methode werd toegepast voor de M_T^W variabele en de relatieve leptonisolatie I . De vorm van de (tweedimensionale) verdeling van de onderzochte variabelen voor de multijet achtergrond werd gemodelleerd aan de hand van de vorm die verkregen werd bij inversie van het relatieve leptonisolatie criterium. De verwachte bijdrage van achtergrond te wijten aan prompte leptonen werd van de verdeling van de data afgetrokken en de uiteindelijk resterende verdeling werd toegewezen aan de multijet achtergrond.

Correcties aan de simulaties, voor *pile-up*, jet energieniveau en resolutie, *b-tagging*, trigger en lepton efficiënties en simulatiegewichten werden doorgevoerd volgens de standaard CMS procedures. De top transversaal moment p_T correctie gebeurde met een empirische herschaling gebaseerd op de geobserveerde distributies van topquarks. De empirische procedure werd voorts vergeleken met de theoretische verwachting, waarbij de onzekerheid op de p_T van het topquark in rekening werd gebracht. Eén van de dominante onzekerheden in deze analyse is te wijten aan de energiecalibratie van de jets. Deze onzekerheid werd geëvalueerd door de multiplicatieve jet energiecorrectie te variëren binnen zijn eigen onzekerheden. JER data naar simulatie schaalfactoren werden ook gevarieerd binnen hun onzekerheden, gelijktijdig voor alle jets. Onzekerheden gerelateerd aan b-tagging werden geëvalueerd door de waarden van de schaalfactoren te laten variëren binnen hun onzekerheden. Ook onzekerheden te wijten aan schaalfactoren van de trigger en de leptonidentificatie werden in rekening gebracht. Voor de geïntegreerde luminositeit werd een onzekerheid van 2.5 % meegenomen. De normalisatie van de data-gedreven verdeling voor de multijet achtergrond werd (conservatief) gekozen op +100 % en -50 %, onafhankelijk voor het muon- en electronkanaal. Onzekerheden gerelateerd aan de renormalisatie- en factorisatieschaal (μ_R en μ_F) zijn verkregen via onafhankelijke variaties van de twee schalen met een factor twee in beide richtingen. Onzekerheden te wijten aan de initiële en finale toestand straling (ISR, FSR) en aan het onderliggend evenement (underlying event, UE) zijn bepaald via speciaal hiervoor gecreëerde sets. Topquark massa onzekerheden werden verkregen via twee gesimuleerde datasets van $t\bar{t}$ met top massa's m_t van 169.5 and 175.5 GeV. De h_{damp} parameter in POWHEG, die de onderdrukking van extra hoge p_T jets bepaalt, werd gevarieerd van zijn nominale waarde $1.58m_t$ tot $0.99m_t$ en $2.24m_t$. Ook de statistische onzekerheden van de dominante $t\bar{t}$ achtergrond werden meegerekend. De geïntroduceerde variaties van de ISR en UE bleken statistisch niet significant en werden hierom uiteindelijk niet verder in rekening genomen. De PDF onzekerheden werden berekend van de set NNPDF3.0. Ze werden beschreven door 100 alternatieve versies van de PDF (genoemd "MC replicas"), en werden verder gebruikt als gewicht waarmee de gebeurtenissen kunnen gewogen

worden.

Om de aanwezigheid van het hypothetische $\Phi(H/A)$ deeltje te onderzoeken werd een twee-dimensionale verdeling van de massa van het gereconstrueerde $t\bar{t}$ systeem $m_{t\bar{t}}$ en een hoekvariabele die de spin van de s-kanaal resonantie weergeeft geconstrueerd. De eerste variabele $m_{t\bar{t}}$ geeft een benadering weer voor de massa van het zware Higgs deeltje. De tweede variabele is $|\cos \theta^*|$, waarbij θ^* de hoek is tussen de impuls van het top quark dat leptonisch vervalt in het $t\bar{t}$ ruststelsel en de impuls van het $t\bar{t}$ systeem in het lab stelsel.

Voor de statistische analyse werd het zogenaamde *higgs combine tool* gebruikt, gebaseerd op RooFit/RooStat. RooFit is een toolkit in het ROOT softwarepakket dat verschillende types fits kan uitvoeren en ook toy monte-carlo datasets kan genereren gebaseerd op gebruikergedefinieerde modellen van PDFs. Het interferentie sjabloon, de 2D distributie van $m_{t\bar{t}}$ en $\cos \theta^*$, heeft zowel positief als negatief gewogen evenementen. Aangezien het higgs combine tool echter enkel met positieve gewichten kan werken, werd het sjabloon in twee opgedeeld naargelang het teken van het gewicht van de evenementen. Het teken van de negatieve gewichten werd vervolgens omgedraaid. De twee sjablonen werden dan geschaald in het model met respectievelijk g^2 and $-g^2$ (waarbij g de koppeling schaalfactor is, en we voor de interferentie een evenredigheid met g^2 verkrijgen), waarmee de uiteindelijk gewenste verdeling geproduceerd werd. Gezien er geen indicatie voor een BSM signaal te vinden was, werden voor diverse signaalhypotheses bovenlimieten voor de koppeling schaalfactoren bepaald in het semileptonisch kanaal.

Om het interferentie-effect in rekening te brengen werd LO (leading order) MADGRAPH gebruikt om de signaal datasets te genereren, waarna een herschaling uitgevoerd werd naar een hogere orde (NNLO) werkzame doorsnede toe. De SUSHI code samen met de 2HDMC calculator werden gebruikt om schaalfactoren voor het resonantiedeel in een hMSSM context te berekenen. Voor het interferentiedeel werd de TOP++ calculator gebruikt voor de berekening van de SM $t\bar{t}$ werkzame doorsnede. De k-factor werd berekend als de verhouding van de NNLO tot LO werkzame doorsnede en werd daarna toegepast op de MADGRAPH werkzame doorsnede voor het signaal. Voor eenzelfde massapunt in deze analyse blijven de k-factoren constant, d.w.z. onafhankelijk van alle beschouwde breedtes. De hier voorgestelde analyse is gevoelig voor hoge waarden van de Φ massa en lage $\tan \beta$, een gebied dat bestreken wordt door het hMSSM waarin rekening gehouden met het reeds ontdekte 125 GeV Higgs boson. Aangezien het onmogelijk is om het hele relevante ($m_A, \tan \beta$) gebied te bestrijken met een gelimiteerde set van massa's en breedtes zoals in deze analyse gebruikt, werd het ROOMENTMORPH algoritme aangewend om te interpoleren tussen de massa's en breedtes. Voorts is er ook een extrapolatie uitgevoerd naar kleinere breedtes onder de *narrow-width* benadering.

De resultaten voor de semileptonische finale toestand werden gecombineerd met die van de dileptonische kanaal, waarbij bovenlimieten bepaald werden voor de coupling modifier van de extra Higgs bosonen met topquarks. De resultaten werden binnen de hMSSM context geïnterpreteerd in het ($m_A, \tan \beta$) vlak. Binnen de systematische onzekerheden zijn de resultaten compatibel met de theoretische voorspellingen. Deze analyse is gevoelig aan het lage $\tan \beta$ gebied en is dus complementair aan andere analyses die ofwel gevoelig zijn voor hoge $\tan \beta$ (2τ finale toestand) of voor massa's onder de $t\bar{t}$ productiedrempel¹.

Tot nu toe werd er in deze analyse geen nieuwe fysica gevonden en zijn de resultaten in overeenkomst met de voorspellingen van het Standaardmodel. Een beter resultaat zou kunnen verkregen worden met meer statistiek in de data en met een meer preciese meting van de $t\bar{t}$ massa. De volledige Run-II dataset kan daarom de gevoeligheid in dit kanaal al verhogen. Met de 300 fb^{-1} en 3 ab^{-1} aan data die voorzien wordt gedurende respectievelijk Run-III en de HL-LHC zal het $gg \rightarrow H \rightarrow t\bar{t}$ kanaal een centrale rol spelen in de zoektocht naar zware Higgs deeltjes. Omwille van zijn zelfde topologie is in deze analyse de SM $t\bar{t}$ productie de voornaamste achtergrond. Om de gevoeligheid van deze analyse verder op te drijven dient deze achtergrond dan ook verder onderdrukt worden door middel van een verbeterde signaal-tot-achtergrond discriminerende technieken. Dit kan in principe gebeuren door gebruik te maken van MVA technieken waarbij de negatieve gewichten voor de interferentie in rekening gebracht worden. De k-factoren voor interferentie zijn momenteel berekend als het geometrisch gemiddelde van de resonantie en de SM $t\bar{t}$ k-factoren. Bij de volgende iteratie van deze analyse is het de bedoeling om de NLO k-factoren voor de interferentie te berekenen op een manier waarbij de curves van de k-factoren het interferentie-effect reeds in rekening brengen. Verschillende signaal curves zoals een zuivere *dip*, *bump* of zelfs *nothingness* zullen ook helpen om dit kanaal verder te onderzoeken.

¹In het bijzonder analyses die zoeken naar bi-Higgs bosonen en pseudoscalairen die vervallen naar het 125 GeV Higgs boson en een Z boson, alsook restricties komende van de geobserveerde koppelingen met het 125 GeV Higgs boson.

Bibliography

- [1] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Letts. B* 716 (2012). DOI: 10.1016/j.physletb.2012.08.021.
- [2] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Letts. B* 716 (2012). DOI: 10.1016/j.physletb.2012.08.020.
- [3] C. Patrignani et al. (Particle Data Group). “2016 Review of Particle Physics”. In: *Chin. Phys. C* 40.100001 (2016). URL: http://pdg.lbl.gov/2016/tables/contents_tables.html.
- [4] A. Pich. “Quantum Chromodynamics”. In: (5/1995). Lectures at the 1994 European School of High-Energy Physics (Sorrento, 1994). arXiv: hep-ph/9505231.
- [5] A. Ceccucci et. al. “The CKM quark-mixing matrix”. In: 12 (1/2016). URL: <http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-ckm-matrix.pdf>.
- [6] S. M. Bilenky and J. Hosek. “Glashow-Weinberg-Salam theory of electroweak interactions and the neutral currents”. In: *Joint Institute for Nuclear Research* 90 (10/1982), pp. 73–157. DOI: 10.1016/0370-1573(82)90016-3.
- [7] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (8/1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.
- [8] P.W. Higgs. “Broken symmetries, massless particles and gauge fields”. In: *Phys. Lett.* 12 (9/1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.
- [9] CDF Collaboration. “Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab”. In: *Phys. Rev. Lett.* 47 (4/1995). URL: <https://journals.aps.org/prl/pdf/10.1103/PhysRevLett.74.2626>.
- [10] W. J. Stirling. *Parton luminosity and cross section plots*. 2012. URL: <http://www.hep.ph.ic.ac.uk/~wstirlin/plots/plots.html>.
- [11] Frank-Peter Schillin. “Top quark physics at the LHC: A review of the first two years”. In: *Int. J. Mod. Phy. A* 27 (2012). DOI: 10.1142/S0217751X12300165.
- [12] CMS Collaboration. *Measurement of the $t\bar{t}$ production cross section in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ using final states with an isolated lepton and at least one jet*. Analysis Note (AN) AN-2015/233. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2015/233.
- [13] Andrea Giammanco. “Single top quark production at the LHC”. In: *Rev. Phys.* 1 (2016), pp. 1–12. DOI: 10.1016/j.revip.2015.12.001. arXiv: 1511.06748 [hep-ex].
- [14] *Useful Diagrams of Top Signals and Backgrounds*. https://wwwd0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html.
- [15] Burton A. Betchart et al. “Analytic solutions for neutrino momenta in decay of top quarks”. In: *Nucl. Instrum. Meth. A* 736 (2014), pp. 169–178. DOI: 10.1016/j.nima.2013.10.039. arXiv: 1305.1878 [hep-ph].

- [16] “Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb^{-1} of Data”. In: 2012. arXiv: 1207.0449 [hep-ex]. URL: http://lss.fnal.gov/cgi-bin/find_paper.pl?conf=12-318.
- [17] CMS Collaboration. *Measurements of properties of the Higgs boson and search for an additional resonance in the four-lepton final state at $\sqrt{s} = 13 \text{ TeV}$* . Physics Analysis Summary (PAS) HIG-16-033. 2016. URL: <https://cds.cern.ch/record/2204926/files/HIG-16-033-pas.pdf>.
- [18] Albert M Sirunyan et al. “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ ”. In: (2017). arXiv: 1706.09936 [hep-ex].
- [19] CMS Collaboration. *$H \rightarrow \gamma\gamma$ at $\sqrt{s} = 13 \text{ TeV}$ with 2016 dataset*. Analysis Note (AN) AN -2017/036. 2017. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/036.
- [20] Michael Duehrssen et. al. *LHC Higgs Cross Section*. 2017. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHSWG#SM_Higgs.
- [21] Abdelhak Djouadi. “The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model”. In: *Phys. Rept.* 459 (2008), pp. 1–241. DOI: [10.1016/j.physrep.2007.10.005](https://doi.org/10.1016/j.physrep.2007.10.005). arXiv: hep-ph/0503173 [hep-ph].
- [22] David Kaplan. *Beyond the Standard Model: New Forces of Nature*. 2017. URL: <https://www.learner.org/courses/physics/unit/text.html?unit=2&secNum=11>.
- [23] Roger H. K. Kadala. “Topics in supersymmetry phenomenology at the Large Hadron Collider”. PhD thesis. Hawaii U., 2012. arXiv: 1205.1267 [hep-ph]. URL: <https://inspirehep.net/record/1113772/files/arXiv:1205.1267.pdf>.
- [24] Stephen P. Martin. “A Supersymmetry primer”. In: (1997). *Adv. Ser. Direct. High Energy Phys.* 18, 1 (1998). DOI: [10.1142/9789812839657_0001](https://doi.org/10.1142/9789812839657_0001), [10.1142/9789814307505_0001](https://doi.org/10.1142/9789814307505_0001). arXiv: hep-ph/9709356 [hep-ph].
- [25] James D. Wells. “Lectures on Higgs Boson Physics in the Standard Model and Beyond”. In: *39th British Universities Summer School in Theoretical Elementary Particle Physics (BUSSTEPP 2009) Liverpool, United Kingdom, August 24-September 4, 2009*. 2009. arXiv: 0909.4541 [hep-ph]. URL: <https://inspirehep.net/record/832122/files/arXiv:0909.4541.pdf>.
- [26] Prerit Jaiswal. “Higgs Physics in Supersymmetric Models”. PhD thesis. Stony Brook University, 2012. URL: <http://graduate.physics.sunysb.edu/announ/theses/jaiswal-prerit-aug-2012.pdf>.
- [27] D. M. Asner et al. “ILC Higgs White Paper”. In: *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*. 2013. arXiv: 1310.0763 [hep-ph]. URL: <https://inspirehep.net/record/1256491/files/arXiv:1310.0763.pdf>.
- [28] A. Djouadi et al. “Fully covering the MSSM Higgs sector at the LHC”. In: *JHEP* 06 (2015), p. 168. DOI: [10.1007/JHEP06\(2015\)168](https://doi.org/10.1007/JHEP06(2015)168). arXiv: 1502.05653 [hep-ph].
- [29] Abdelhak Djouadi et al. “Future Collider Signatures of the Possible 750 GeV State”. In: *JHEP* 03 (2016), p. 205. DOI: [10.1007/JHEP03\(2016\)205](https://doi.org/10.1007/JHEP03(2016)205). arXiv: 1601.03696 [hep-ph].
- [30] Julien Baglio et al. “Prospects for Higgs physics at energies up to 100 TeV”. In: *Rept. Prog. Phys.* 79, 11 (2016), p. 116201. DOI: [10.1088/0034-4885/79/11/116201](https://doi.org/10.1088/0034-4885/79/11/116201). arXiv: 1511.07853 [hep-ph].
- [31] Robert V. Harlander and William B. Kilgore. “Higgs boson production in bottom quark fusion at next-to-next-to leading order”. In: *Phys. Rev. D* 68 (2003), p. 013001. DOI: [10.1103/PhysRevD.68.013001](https://doi.org/10.1103/PhysRevD.68.013001). arXiv: hep-ph/0304035 [hep-ph].

- [32] S. Dittmaier et al. “Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables”. In: (2011). DOI: 10.5170/CERN-2011-002. arXiv: 1101.0593 [hep-ph].
- [33] Abdelhak Djouadi and Jeremie Quevillon. “The MSSM Higgs sector at a high M_{SUSY} : reopening the low $\tan\beta$ regime and heavy Higgs searches”. In: *JHEP* 10 (2013), p. 028. DOI: 10.1007/JHEP10(2013)028. arXiv: 1304.1787 [hep-ph].
- [34] S. Dittmaier et al. “Handbook of LHC Higgs Cross Sections: 2. Differential Distributions”. In: (2012). DOI: 10.5170/CERN-2012-002. arXiv: 1201.3084 [hep-ph].
- [35] F. Gaemers K.J.F. Hoogeveen. “Higgs production and decay into heavy flavours with the gluon fusion mechanism”. In: *Phys. Lett. B* 146 (1984). DOI: 10.1016/0370-2693(84)91711-8.
- [36] Sungsoon Jung et al. “Dip or nothingness of a Higgs resonance from the interference with a complex phase”. In: *Phys. Rev.* D92.5 (2015), p. 055009. DOI: 10.1103/PhysRevD.92.055009. arXiv: 1505.00291 [hep-ph].
- [37] Marcela Carena and Zhen Liu. “Challenges and opportunities for heavy scalar searches in the $t\bar{t}$ channel at the LHC”. In: *JHEP* 11 (2016), p. 159. DOI: 10.1007/JHEP11(2016)159. arXiv: 1608.07282 [hep-ph].
- [38] W. Bernreuther et al. “Production of heavy Higgs bosons and decay into top quarks at the LHC”. In: *Phys. Rev.* D93.3 (2016), p. 034032. DOI: 10.1103/PhysRevD.93.034032. arXiv: 1511.05584 [hep-ph].
- [39] Gero von Gersdorff et al. “The Dynamical Composite Higgs”. In: *JHEP* 06 (2015), p. 119. DOI: 10.1007/JHEP06(2015)119. arXiv: 1502.07340 [hep-ph].
- [40] John C. Collins and Davison E. Soper. “Angular distribution of dileptons in high-energy hadron collisions”. In: *Phys. Rev. D* 16 (7 1977), pp. 2219–2225. DOI: 10.1103/PhysRevD.16.2219. URL: <https://link.aps.org/doi/10.1103/PhysRevD.16.2219>.
- [41] ATLAS Collaboration. “Search for $t\bar{t}$ resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s}=7 \text{ TeV}$ ”. In: *Phys. Rev. D* 88 (1 2013), p. 012004. DOI: 10.1103/PhysRevD.88.012004. URL: <https://link.aps.org/doi/10.1103/PhysRevD.88.012004>.
- [42] CMS Collaboration. “Searches for new physics using the $t\bar{t}$ invariant mass distribution in pp collisions at $\sqrt{s}=8 \text{ TeV}$ ”. In: *Phys. Rev. Lett.* 111 (21 2013), p. 211804. DOI: 10.1103/PhysRevLett.111.211804. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.111.211804>.
- [43] Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1.
- [44] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (2008), S08001. URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>.
- [45] CMS Luminosity. 2017. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [46] CMS Collaboration. *CMS physics: Technical design report*. Tech. rep. CERN-LHCC-2006-001. 2006.
- [47] CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Tech. rep. CERN-LHCC-98-006. 1997.
- [48] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-033. 1997.
- [49] Q. Ingram et. al. “Energy resolution of the barrel of the CMS Electromagnetic Calorimeter”. In: *Journal of Instrumentation* 2.04 (2007), P04004. URL: <http://stacks.iop.org/1748-0221/2/i=04/a=P04004>.

- [50] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-031. 1997.
- [51] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.
- [52] CMS Collaboration. *The CMS muon project: Technical Design Report*. Tech. rep. CERN-LHCC-97-032. 1997.
- [53] CMS Collaboration. “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC”. In: *JINST* P11002 (2013). DOI: 10.1088/1748-0221/8/11/P11002.
- [54] CMS Collaboration. “The CMS trigger system”. In: *Journal of Instrumentation* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020.
- [55] M. Gul et. al. “Detector control system and efficiency performance for CMS RPC at GIF++”. In: *Journal of Instrumentation* 11.10 (2016), p. C10013. URL: <http://stacks.iop.org/1748-0221/11/i=10/a=C10013>.
- [56] P La Rocca and F Riggi. “The upgrade programme of the major experiments at the Large Hadron Col- lide”. In: *Journal of Physics: Conference Series* 515.1 (2014), p. 012012. URL: <http://stacks.iop.org/1742-6596/515/i=1/a=012012>.
- [57] *Industry Relations and Procurement Website for the HL-LHC project*. URL: <https://project-hl-lhc-industry.web.cern.ch/content/project-schedule>.
- [58] G Apollinari et al. *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2015. URL: <http://cds.cern.ch/record/2116337>.
- [59] CMS Collaboration. *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. CERN-LHCC-2017-012. CMS-TDR-016. This is the final version, approved by the LHCC. Geneva: CERN, 2017. URL: <http://cds.cern.ch/record/2283189>.
- [60] CMS Collaboration. *Technical Proposal for the Phase-II Upgrade of the CMS Detector*. Tech. rep. CERN-LHCC-2015-010. CERN. Geneva. LHC Experiments Committee, 2015.
- [61] M.R. Jäkel et al. “CERN GIF ++ : A new irradiation facility to test large-area particle detectors for the high-luminosity LHC program”. In: vol. TIPP2014. PoS, 2014, pp. 102–109.
- [62] Dorothea Pfeiffer et al. “The radiation field in the Gamma Irradiation Facility GIF++ at CERN”. In: *Nucl. Instrum. Meth.* A866 (2017), pp. 91–103. DOI: 10.1016/j.nima.2017.05.045. arXiv: 1611.00299 [physics.ins-det].
- [63] F. Lagarde et al. “High rate, fast timing Glass RPC for the high eta CMS muon detectors”. In: *Journal of Instrumentation* 11.09 (2016), p. C09006. URL: <http://stacks.iop.org/1748-0221/11/i=09/a=C09006>.
- [64] CERN EN-Department. *WinCC Open Architecture (OA) Service*. 2017. URL: <https://readthedocs.web.cern.ch/display/ICKB/WinCC-OA+Service>.
- [65] *SIMATIC WinCC Open Architecture*. URL: https://www.automation.siemens.com/salesmaterial-as/brochure/en/brochure_simatic-wincc_oa_en.pdf.
- [66] *SIMATIC WinCC Open Architecture*. URL: http://www.etm.at/index_e.asp?id=2&sb1=25&sb2=31.
- [67] O. Holme et al. “THE JCOP FRAMEWORK”. In: ICALEPCS, 2005.
- [68] G. Polese et al. “The Detector Control Systems for the CMS Resistive Plate Chamber”. In: 219 (2010) 022019. *J. Phys.*, 2010. DOI: 10.1088/1742-6596/219/2/022019.

- [69] Passalacqua et al. “Integrating OPC Data into GSN Infrastructures”. In: (8/2008). IADIS International Conference APPLIED COMPUTING 2008, Algarve : Portugal (2008). arXiv: 0808.0055v1.
- [70] Salvador Carrillo. “CMS RPC preliminary results for aging studies at new CERN GIF++ facility”. In: XIII RPC worksop and related detectors, UGent Belgium. February, 2016.
- [71] CERN. Geneva. LHC Experiments Committee. *The Muon Project Technical Design Report*. Tech. rep. CERN/LHC 97-32, CMS TDR 3, 15 December, 1997. CMS Collaboration, 1997.
- [72] M. Abbrescia et al. “Operation, performance and upgrade of the CMS Resistive Plate Chamber system at LHC”. In: *Nucl. Instr. and Meth.* 732, 195 (2013) (2013). doi: 10.1016/j.nima.2013.05.150.
- [73] M. Abbrescia et al. “Study of long-term performance of CMS RPC under irradiation at the CERN GIF”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 533.1 (2004). Proceedings of the Seventh International Workshop on Resistive Plate Chambers and Related Detectors, pp. 102 –106. doi: <https://doi.org/10.1016/j.nima.2004.07.009>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900204014135>.
- [74] F. Thyssen et at. “Performance of the Resistive Plate Chambers in the CMS experiment”. In: *JINST* 7 (2012) C01104 (2012). doi: 10.1088/1748-0221/7/01/C01104.
- [75] T. Gleisberg et al. “Event generation with SHERPA 1.1”. In: *JHEP* 02 (2009), p. 007. doi: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph].
- [76] Ringaile Placakyte. “Parton Distribution Functions”. In: *Proceedings, 31st International Conference on Physics in collisions (PIC 2011): Vancouver, Canada, August 28-September 1, 2011*. 2011. arXiv: 1111.5452 [hep-ph]. URL: <https://inspirehep.net/record/954990/files/arXiv:1111.5452.pdf>.
- [77] NNPDF Collaboration. *NNPDF sets*. 2017. URL: <http://nnpdf.mi.infn.it/>.
- [78] John M. Campbell et al. “Hard Interactions of Quarks and Gluons: A Primer for LHC Physics”. In: *Rept. Prog. Phys.* 70 (2007), p. 89. doi: 10.1088/0034-4885/70/1/R02. arXiv: hep-ph/0611148 [hep-ph].
- [79] Fabio Maltoni and Tim Stelzer. “MadEvent: Automatic event generation with MadGraph”. In: *JHEP* 02 (2003), p. 027. doi: 10.1088/1126-6708/2003/02/027. arXiv: hep-ph/0208156 [hep-ph].
- [80] Andy Buckley et al. “LHAPDF6: parton density access in the LHC precision era”. In: *Eur. Phys. J.* C75 (2015), p. 132. doi: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].
- [81] B Andersson. “Revisiting the Lund model”. In: *Journal of Physics G: Nuclear and Particle Physics* 17.10 (1991), p. 1507. URL: <http://stacks.iop.org/0954-3899/17/i=10/a=005>.
- [82] Torbjorn Sjostrand et al. “PYTHIA 6.4 Physics and Manual”. In: *JHEP* 05 (2006), p. 026. doi: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].
- [83] Stefano Frixione et al. “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. doi: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- [84] Stefano Frixione and Bryan R. Webber. “Matching NLO QCD computations and parton shower simulations”. In: *JHEP* 06 (2002), p. 029. doi: 10.1088/1126-6708/2002/06/029. arXiv: hep-ph/0204244 [hep-ph].
- [85] G. Corcella et al. “HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)”. In: *JHEP* 01 (2001), p. 010. doi: 10.1088/1126-6708/2001/01/010. arXiv: hep-ph/0011363 [hep-ph].

- [86] Andy Buckley et al. “General-purpose event generators for LHC physics”. In: *Phys. Rept.* 504 (2011), pp. 145–233. DOI: 10.1016/j.physrep.2011.03.005. arXiv: 1101.2599 [hep-ph].
- [87] Stefan Hoeche et al. “Matching parton showers and matrix elements”. In: *HERA and the LHC: A Workshop on the implications of HERA for LHC physics: Proceedings Part A*. 2005, pp. 288–289. DOI: 10.5170/CERN-2005-014.288. arXiv: hep-ph/0602031 [hep-ph]. URL: https://inspirehep.net/record/709818/files/arXiv:hep-ph_0602031.pdf.
- [88] C. Mena and L. F. Palhares. “Quark-antiquark potentials in non-perturbative models”. In: *14th International Workshop on Hadron Physics (Hadron Physics 2018) Florianopolis, Santa Catarina, Brazil, March 18-23, 2018*. 2018. arXiv: 1804.09564 [hep-ph]. URL: <https://inspirehep.net/record/1670037/files/1804.09564.pdf>.
- [89] E. Norrbin and T. Sjostrand. “Production and hadronization of heavy quarks”. In: *Eur. Phys. J.* C17 (2000), pp. 137–161. DOI: 10.1007/s100520000460. arXiv: hep-ph/0005110 [hep-ph].
- [90] S. Agostinelli et al. “GEANT4: A Simulation toolkit”. In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.
- [91] Florian Beaudette. “The CMS Particle Flow Algorithm”. In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, pp. 295–304. arXiv: 1401.8155 [hep-ex]. URL: <https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf>.
- [92] Serguei Chatrchyan et al. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].
- [93] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Meth.* A262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.
- [94] Wolfgang Waltenberger et al. “Adaptive vertex fitting”. In: *Journal of Physics G: Nuclear and Particle Physics* 34.12 (2007), N343. URL: <http://stacks.iop.org/0954-3899/34/i=12/a=N01>.
- [95] Serguei Chatrchyan et al. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”. In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].
- [96] Vardan Khachatryan et al. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”. In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].
- [97] Ilya Kravchenko. *Cut Based Electron ID for Run 2*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2?rev=41>.
- [98] “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”. In: *CMS Physics Analysis Summary CMS-PAS-PFT-10-002* (2010).
- [99] Matteo Cacciari et al. “The anti- k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].
- [100] A. M. Sirunyan et al. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].
- [101] D. Dicus et al. “Higgs decay to top quarks at hadron colliders”. In: *Phys. Lett. B* 333 (1994), pp. 126–131. DOI: 10.1016/0370-2693(94)91017-0. arXiv: hep-ph/9404359 [hep-ph].

- [102] Alkaid Cheng et al. *Statistical evaluation, combination, and interpretation of searches for additional heavy Higgs bosons in top quark pair final states*. CMS Analysis Note AN-17-202. 2017. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/202.
- [103] Johan Alwall et al. “*MadGraph 5: Going beyond*”. In: *JHEP* 1106 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].
- [104] Andrey Popov. *HTTMadgraphDocumentation*. URL: https://github.com/IPNL-CMS/HTTMadgraphDocumentation/tree/master/model/Massive_Higgs_UFO.
- [105] M. Spira et al. “Higgs boson production at the LHC”. In: *Nucl. Phys. B* 453 (1995), pp. 17–82. DOI: 10.1016/0550-3213(95)00379-7. arXiv: hep-ph/9504378 [hep-ph].
- [106] B. Hespel et al. “Signal background interference effects in heavy scalar production and decay to a top-anti-top pair”. In: *JHEP* 10 (2016), p. 016. DOI: 10.1007/JHEP10(2016)016. arXiv: 1606.04149 [hep-ph].
- [107] Richard D. Ball et al. “Parton distributions for the LHC Run II”. In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].
- [108] Torbjörn Sjöstrand et al. “A brief introduction to PYTHIA 8.1”. In: *Comp. Phys. Comm.* 178 (2008), p. 852. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- [109] CMS Collaboration. *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV*. CMS Physics Analysis Summary TOP-16-021. 2016. URL: <http://cds.cern.ch/record/2235192>.
- [110] Sébastien Brochet. *Decay processes cannot specify squared orders constraints*. <https://bugs.launchpad.net/mg5amcnlo/+bug/1511378>.
- [111] David Eriksson et al. “2HDMC: Two-Higgs-Doublet Model Calculator Physics and Manual”. In: *Comput. Phys. Commun.* 181 (2010), pp. 189–205. DOI: 10.1016/j.cpc.2009.09.011. arXiv: 0902.0851 [hep-ph].
- [112] Robert V. Harlander et al. “SusHi: A program for the calculation of Higgs production in gluon fusion and bottom-quark annihilation in the Standard Model and the MSSM”. In: *Comput. Phys. Commun.* 184 (2013), pp. 1605–1617. DOI: 10.1016/j.cpc.2013.02.006. arXiv: 1212.3249 [hep-ph].
- [113] Michal Czakon and Alexander Mitov. “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”. In: *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: 10.1016/j.cpc.2014.06.021. arXiv: 1112.5675 [hep-ph].
- [114] Muhammad Gul et al. *Search for a heavy Higgs boson decaying into a pair of top quarks in the semileptonic final state with the 2016 pp dataset*. CMS Analysis Note AN-16-272. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2017/202.
- [115] G. C. Branco et al. “Theory and phenomenology of two-Higgs-doublet models”. In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034 [hep-ph].
- [116] W. Bernreuther et al. “Signatures of Higgs bosons in the top quark decay channel at hadron colliders”. In: *Phys. Rev. D* 58 (11 1998), p. 114031. DOI: 10.1103/PhysRevD.58.114031. URL: <https://link.aps.org/doi/10.1103/PhysRevD.58.114031>.
- [117] Morad Aaboud et al. “Search for heavy Higgs bosons A/H decaying to a top quark pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”. In: *Phys. Rev. Lett.* 119.19 (2017), p. 191803. DOI: 10.1103/PhysRevLett.119.191803. arXiv: 1707.06025 [hep-ex].

- [118] Afiq Anuar et al. *Search for a heavy Higgs boson decaying into dilepton $t\bar{t}$ final states at 13 TeV*. CMS Analysis Note AN-16-164. 2017. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2016/164.
- [119] A. Djouadi et al. “The post-Higgs MSSM scenario: Habemus MSSM?” In: *Eur. Phys. J. C* 73 (2013), p. 2650. DOI: 10.1140/epjc/s10052-013-2650-0. arXiv: 1307.5205 [hep-ph].
- [120] CMS Collaboration. *CMS luminosity measurements for the 2016 data-taking period*. CMS Physics Analysis Summary LUM-17-001. 2017. URL: <https://cds.cern.ch/record/2257069>.
- [121] Paolo Nason. “A new method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146 [hep-ph].
- [122] Henry Yee-Shian Tong Tomislav Seva. *Standard Model Cross Sections for CMS at 13 TeV*. <https://twiki.cern.ch/twiki/bin/view/CMS/StandardModelCrossSectionsat13TeV?rev=24>.
- [123] A. et. al Anuar. *Search for $H \rightarrow b\bar{b}$ in association with single top quarks as a test of Higgs boson couplings*. CMS Analysis Note AN-13-113. 2014. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2013/113.
- [124] Michael Hildreth. *Pileup Reweighting Utilities*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities?rev=29>.
- [125] MichaelHildreth. *Utilities for Accessing Pileup Information for Data*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData?rev=28>.
- [126] Henning Kirschenmann et. al. 2016. URL: <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>.
- [127] Andrey et. al. Popov. 2016. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution>.
- [128] CMS Collaboration. *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Physics Analysis Summary (PAS) CMS PAS BTV-15-001. 2016. URL: <https://cds.cern.ch/record/2138504/files/BTV-15-001-pas.pdf>.
- [129] CMS Collaboration. *Identification of b quark jets at the CMS experiment in the LHC Run 2*. CMS Physics Analysis Summary BTV-15-001. 2016. URL: <https://cds.cern.ch/record/2138504>.
- [130] <https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco?rev=14>.
- [131] <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMMethods?rev=27>.
- [132] N. Adam et al. *Generic tag and probe tool for measuring efficiency at CMS with early data*. CMS Analysis Note AN-09-111. 2009. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2009/111.
- [133] Vardan Khachatryan et al. “Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV”. In: *JHEP* 01 (2011), p. 080. DOI: 10.1007/JHEP01(2011)080. arXiv: 1012.2466 [hep-ex].
- [134] Egamma ID group. *Tag and probe methodology for analyses using electrons and photons*. CMS Analysis Note AN-12-116. 2012. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2012/116.
- [135] Andrey Popov. *Egamma identification*. <https://indico.cern.ch/event/604912/contributions/2490011/>.

- [136] Michal Bluj et al. <https://indico.cern.ch/event/598433/contributions/2475199/>.
- [137] *Single Muon Trigger Efficiencies of 80X Data and MC using T&P Method.* https://indico.cern.ch/event/608200/contributions/2452382/attachments/1401679/2156728/2017.02.05_MuonPOGTriggerSF_KPLee_v1.pdf. 2016.
- [138] *Summary performance Plots for 2016 data.* <https://twiki.cern.ch/twiki/bin/view/CMSPublic/HLTplotsSummary2016>. 2016.
- [139] *Work in progress (evolving results) and dedicated studies from PAGs.* <https://twiki.cern.ch/twiki/bin/view/CMS/MuonWorkInProgressAndPagResults?rev=27>. 2016.
- [140] Ilya Kravchenko. *Instructions for applying electron and photon ID.* <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2?rev=40>.
- [141] Georgios Krintiras and Till Arndt. *IP cuts for electron ID - Scale factor study.* <https://indico.cern.ch/event/604947/contributions/2467335/>. 2017.
- [142] Vardan Khachatryan et al. “Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ ”. In: *Eur. Phys. J. C* 75.11 (2015), p. 542. DOI: [10.1140/epjc/s10052-015-3709-x](https://doi.org/10.1140/epjc/s10052-015-3709-x). arXiv: [1505.04480 \[hep-ex\]](https://arxiv.org/abs/1505.04480).
- [143] Vardan Khachatryan et al. “Measurement of the $t\bar{t}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ ”. In: *Eur. Phys. J. C* 76.3 (2016), p. 128. DOI: [10.1140/epjc/s10052-016-3956-5](https://doi.org/10.1140/epjc/s10052-016-3956-5). arXiv: [1509.06076 \[hep-ex\]](https://arxiv.org/abs/1509.06076).
- [144] Vardan Khachatryan et al. “Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV”. In: *Phys. Rev. D* 95.9 (2017), p. 092001. DOI: [10.1103/PhysRevD.95.092001](https://doi.org/10.1103/PhysRevD.95.092001). arXiv: [1610.04191 \[hep-ex\]](https://arxiv.org/abs/1610.04191).
- [145] CMS Collaboration. *Measurement of particle level differential $t\bar{t}$ cross sections in the dilepton channel at $\sqrt{s} = 13 \text{ TeV}$.* CMS Physics Analysis Summary TOP-16-007. 2016. URL: <http://cds.cern.ch/record/2204931>.
- [146] Maria Aldaya. *pt (top-quark) based reweighting of ttbar MC.* <https://twiki.cern.ch/twiki/bin/viewauth/CMS/TopPtReweighting?rev=20>.
- [147] Andrey Popov. *Effect of top pT uncertainty?* <https://hypernews.cern.ch/HyperNews/CMS/get/top/2460/1/1/2/1/1/1.html>.
- [148] Michal Czakon et al. “Top-pair production at the LHC through NNLO QCD and NLO EW”. In: (2017). arXiv: [1705.04105 \[hep-ph\]](https://arxiv.org/abs/1705.04105).
- [149] <https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2?rev=28>.
- [150] <https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016?rev=7>.
- [151] Vardan Khachatryan et al. “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”. In: *JINST* 12.02 (2017), P02014. DOI: [10.1088/1748-0221/12/02/P02014](https://doi.org/10.1088/1748-0221/12/02/P02014). arXiv: [1607.03663 \[hep-ex\]](https://arxiv.org/abs/1607.03663).
- [152] CMS Collaboration. *Jet energy scale and resolution performances with 13 TeV data.* CMS Performance Note DP-2016-020. 2016. URL: <http://cds.cern.ch/record/2160347>.
- [153] <https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution?rev=56>.
- [154] Anastasia Karavdina et al. *Jet transverse momentum resolution measurement using dijet events at $\sqrt{s} = 13 \text{ TeV}$.* CMS Analysis Note AN-16-351. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2016/351.
- [155] Ia Iashvili et al. *Jet energy and angular resolutions in 13 TeV simulation.* CMS Analysis Note AN-16-116. 2016. URL: http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2016/116.

- [156] CMS Collaboration. *Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector*. CMS Physics Analysis Summary JME-16-004. 2016. URL: <https://cds.cern.ch/record/2205284>.
- [157] Erik Butz. <https://indico.cern.ch/event/560226/contributions/2277448/>.
- [158] <https://indico.cern.ch/event/598195/contributions/2429716/>.
- [159] <https://hypernews.cern.ch/HyperNews/CMS/get/physics-validation/2786.html>.
- [160] <https://indico.cern.ch/event/602633/contributions/2462363/>.
- [161] <https://indico.cern.ch/event/598195/contributions/2432685/>.
- [162] Zeynep Demiragli et. al. cms twiki. https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2?rev=106#Moriond_2017.
- [163] Kenneth Remington Call et al. <https://indico.cern.ch/event/401529/contributions/1845064/>.
- [164] A. Loginov. “Strategies of data-driven estimations of $t\bar{t}$ backgrounds in ATLAS”. In: *Nuovo Cim. C* 033.4 (2010), pp. 175–182. DOI: 10.1393/ncc/i2010-10657-2. URL: <https://cds.cern.ch/record/1287126>.
- [165] URL: <https://root.cern.ch/root/html/tutorials/math/chi2test.C.html>.
- [166] Hartmut Stadie. *Jet energy scale uncertainty sources*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources?rev=41#Main_uncertainties_2016_80X.
- [167] Michael Hildreth. *Estimating Systematic Errors Due to Pileup Modeling*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors?rev=11>.
- [168] URL: <https://hypernews.cern.ch/HyperNews/CMS/get/luminosity/613/2/1/1.html>.
- [169] Vardan Khachatryan et al. “Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 TeV”. In: *Phys. Rev. D* 93.7 (2016), p. 072004. DOI: 10.1103/PhysRevD.93.072004. arXiv: 1509.04044 [hep-ex].
- [170] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].
- [171] Roger J. Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.
- [172] J. S. Conway. “Incorporating nuisance parameters in likelihoods for multisource spectra”. In: *PHYSTAT11: Statistical issues related to discovery claims in search experiments and unfolding*. 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].
- [173] *Documentation of the RooStats -based statistics tools for Higgs PAG*. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideHiggsAnalysisCombinedLimit>. 2018.
- [174] L. Moneta et al. “The RooStats Project”. In: *13th International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT2010)*. PoS(ACAT2010)057. SISSA, 2010. eprint: 1009.1003 (physics.data-an). URL: http://pos.sissa.it/archive/conferences/093/057/ACAT2010_057.pdf.
- [175] Andrey Popov et al. *Artificial constraints from statistical fluctuations in systematic variations*. CMS Analysis Note AN-18-077. 2018. URL: http://cms.cern.ch/icMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS\%20AN-2018/077.

- [176] Steve Baker and Robert D. Cousins. “Clarification of the use of χ^2 and likelihood functions in fits to histograms”. In: *Nucl. Instrum. Meth.* 221 (1984), pp. 437–442. DOI: 10.1016/0167-5087(84)90016-4.
- [177] <https://twiki.cern.ch/twiki/bin/view/CMS/StatComGOF?rev=13>.
- [178] ATLAS and CMS collaborations. *Procedure for the LHC Higgs boson search combination in summer 2011*. Tech. rep. ATL-PHYS-PUB-2011-11, CMS-NOTE-2011-005. 2011. URL: <http://inspirehep.net/record/1196797>.
- [179] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an]. URL: <http://inspirehep.net/record/860907>.
- [180] Alexander L. Read. *Modified frequentist analysis of search results (The CL_s method)*. CERN OPEN-2000-205. 2000. DOI: 10.5170/CERN-2000-005.81. URL: <http://inspirehep.net/record/532313>.
- [181] Alexander L. Read. “Presentation of search results: The CL_s technique”. In: *J. Phys. G* 28 (2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: <http://inspirehep.net/record/599622>.
- [182] Thomas Junk. “Confidence level computation for combining searches with small statistics”. In: *Nucl. Instrum. Meth. A* 434.CARLETON-OPAL-PHYS-99-01, CERN-EP-99-041 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex]. URL: <http://inspirehep.net/record/494953>.
- [183] <http://cms-analysis.github.io/CombineHarvester/index.html>, based on the following branch https://github.com/steggema/CombineHarvester/tree/httbar_74.
- [184] Allison Mc Carn et. al. *MSSM Neutral Higgs*. 2017. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWGSSMNeutral#ROOT_histograms_MSSM_benchmark_s,_revision104.
- [185] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: (2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].
- [186] M. Aaboud et al. “Search for Heavy Higgs Bosons A/H Decaying to a Top Quark Pair in pp Collisions at $\sqrt{s} = 8$ TeV with the ATLAS Detector”. In: *Phys. Rev. Lett.* 119 (19 2017), p. 191803. DOI: 10.1103/PhysRevLett.119.191803. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.119.191803>.
- [187] Max Baak et al. “Interpolation between multi-dimensional histograms using a new non-linear moment morphing method”. In: *Nucl. Instrum. Meth. A* 771 (2015), pp. 39–48. DOI: 10.1016/j.nima.2014.10.033. arXiv: 1410.7388 [physics.data-an].
- [188] Andreas Hinzmann et al. *Measurement of the Top Quark Mass from the Invariant Mass of the Top-Antitop Quark System*. CMS Analysis Note 2008/021. 2008. URL: <http://cms.cern.ch/iCMS/user/noteinfo?cmsnoteid=CMS%20AN-2008/021>.

A

Signal distributions

Figures A.1 and A.2 show parton-level cross sections for the resonant part of the signal, the interference, and the sum of the two for different choices of properties of the heavy Higgs boson. The cross sections are computed according to Eqs. 5.2 and 5.4 with $g = 1$. Both \mathcal{CP} states and masses 400, 600, and 800 GeV are probed.

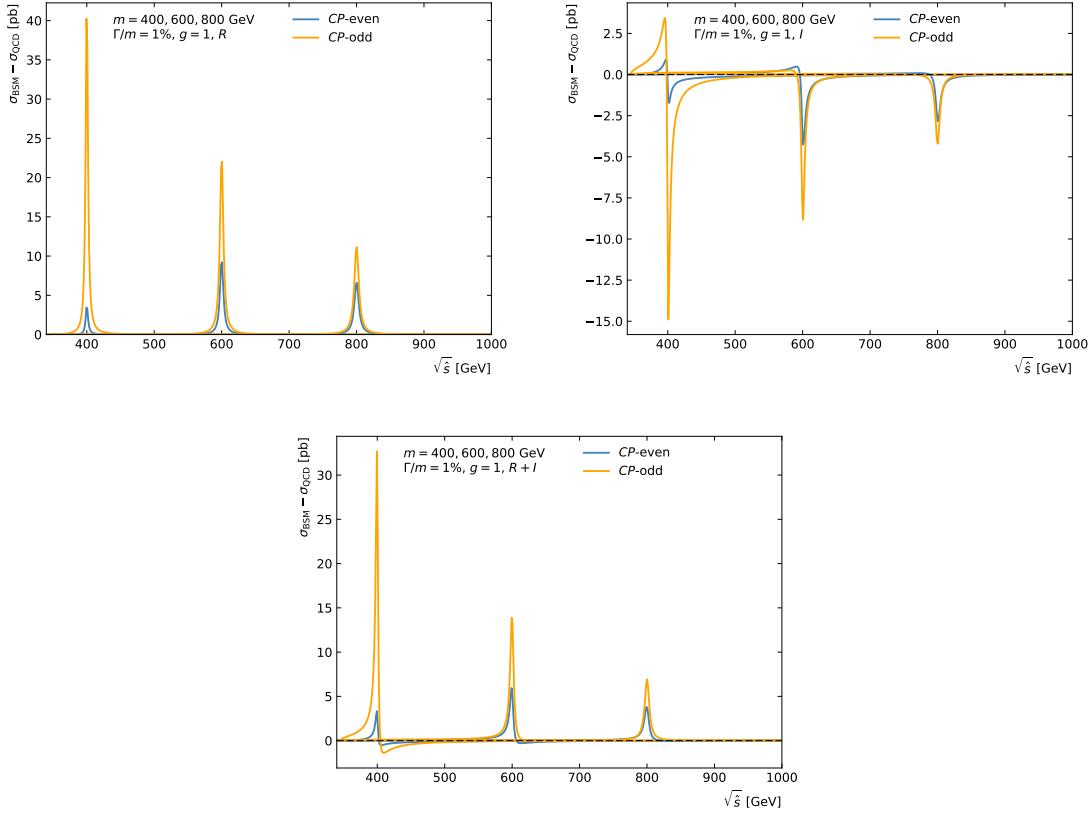


Figure A.1: Parton-level cross sections for the resonant part of the signal (upper left), the interference (upper right), and the sum of the two (bottom) for a total width of 1%.

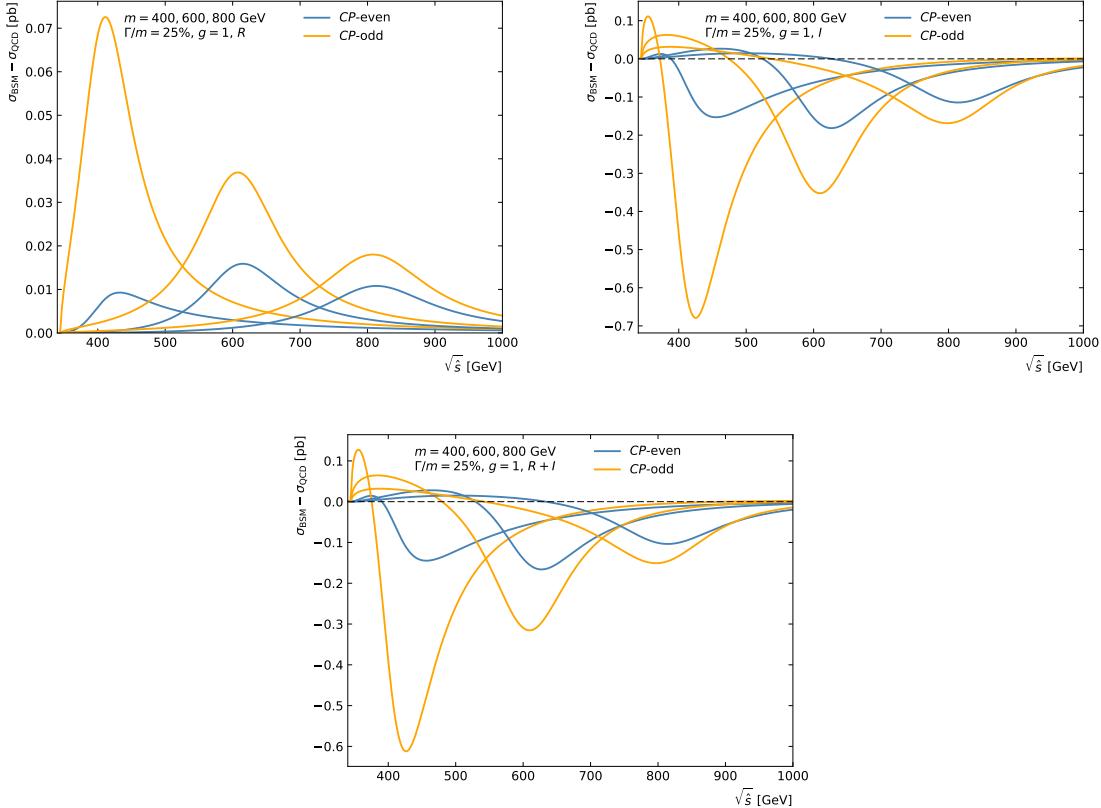


Figure A.2: Parton-level cross sections for the resonant part of the signal (upper left), the interference (upper right), and the sum of the two (bottom) for a total width of 25%.

B

Validation of the generation procedure for interference

The generation procedure for the interference contribution, described in 5.4.2, is validated for the case of a pseudoscalar of mass 500 GeV and partial width 10%. Both top quarks in the final state are required to decay semileptonically. For this study, the factorization and renormalization scales are fixed to half of the mass of the particle, i.e 250 GeV. Three different samples are generated:

- full $gg \rightarrow t\bar{t}$ production, including SM and BSM processes (40 M events, referred to as “ALL” in the plots),
- the SM contribution only (40 M events, referred to as “SM” in the plots),
- the resonant BSM process (400 k events, referred to as “R” in the plots).

A set of kinematic variables are scrutinized with the aim to validate the generation procedure. For each variable, histograms are filled from the three samples, taking the respective leading-order cross sections into account. Histograms for the latter two samples are then subtracted from the distribution for the full process, resulting in the histogram denoted as ALL-SM-R in the plots legend. In Figs B.1 to B.7 the resulting distributions for the interference (in blue) are compared to the ones obtained as described in 5.4.2 (in red, denoted as I in the plots legend). The plots show variables at the generator level, before any reconstruction simulation. Apart from some fluctuations, no evident and significant discrepancy is observed and the method is considered validated.

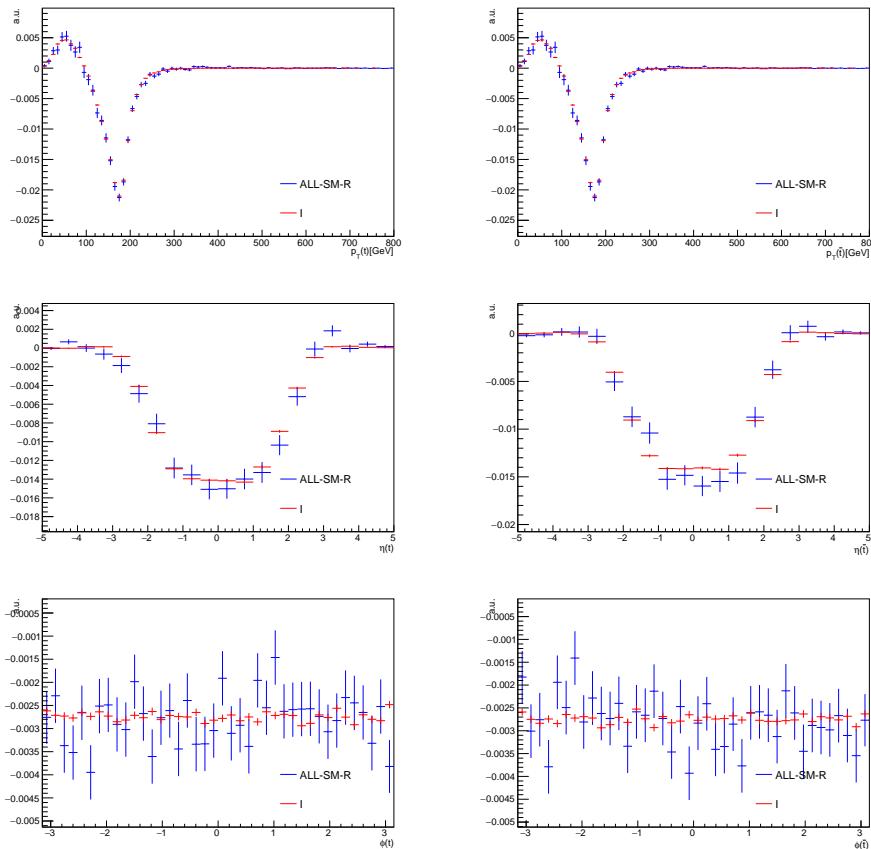


Figure B.1: Distributions of the transverse momenta, pseudorapidities, and azimuthal angles for top quark and antiquark. Comparison of the two approaches to model the interference.

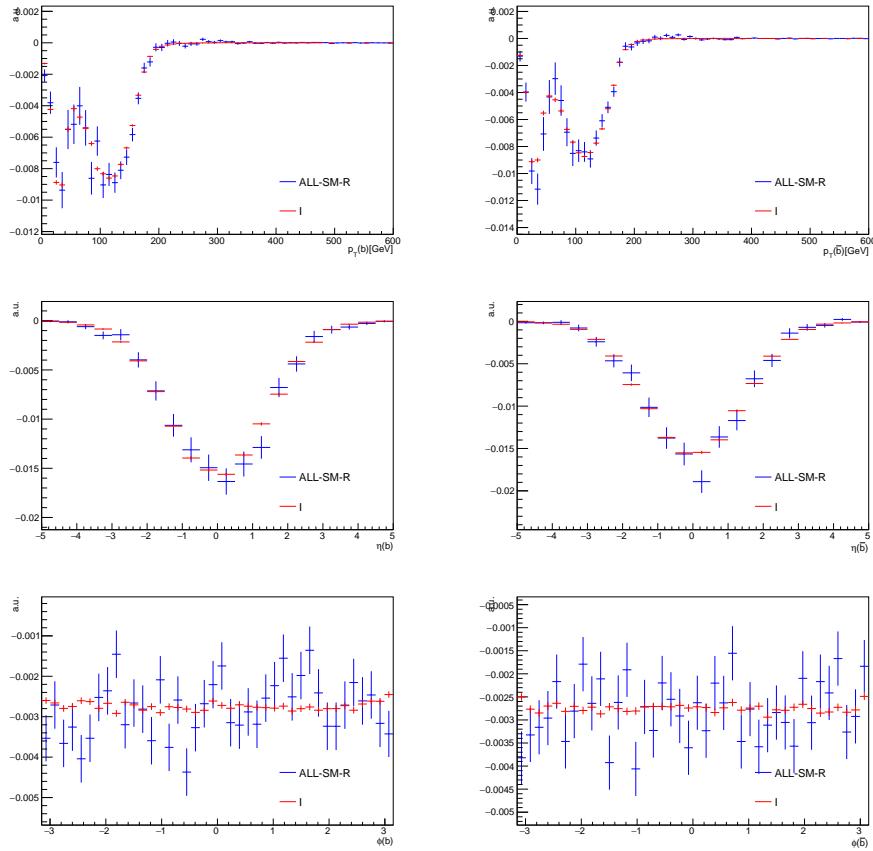


Figure B.2: Distributions of the transverse momenta, pseudorapidities, and azimuthal angles for bottom quark and anti-quark from the decays of the top quarks. Comparison of the two approaches to model the interference.

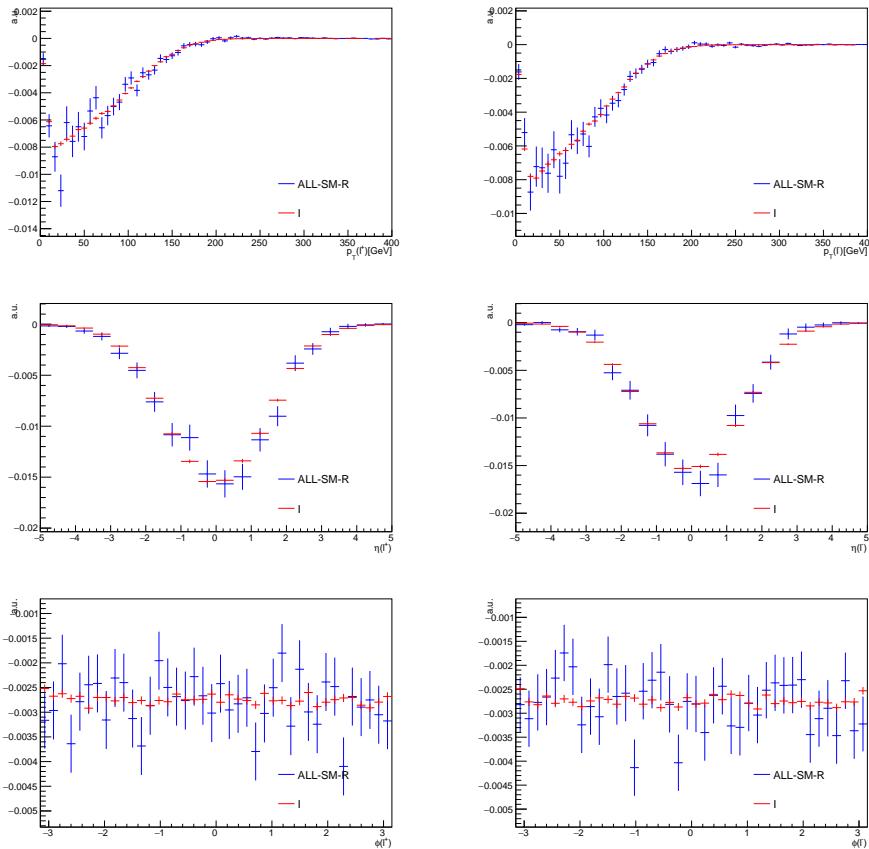


Figure B.3: Distributions of the transverse momenta, pseudorapidities, and azimuthal angles for changed leptons stemming from the decays of the top quarks. Comparison of the two approaches to model the interference.

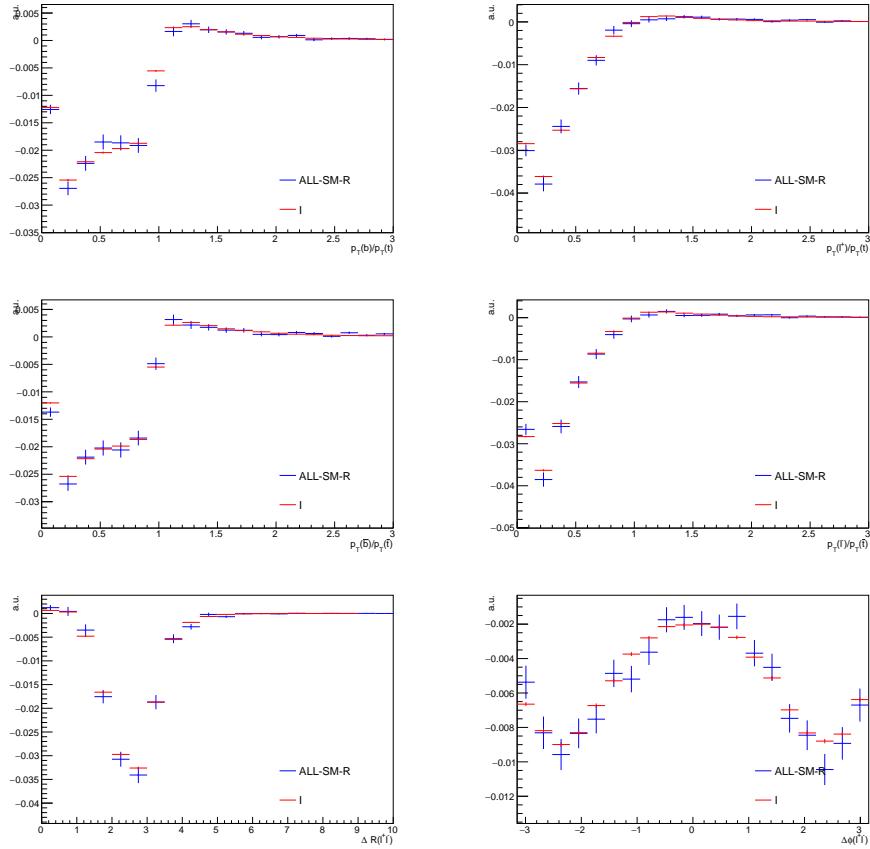


Figure B.4: Distributions of the fractions of the top quark's transverse momentum carried by the b quark and the charged lepton from its decay (top), the similar fractions from top antiquark (middle), and the ΔR and $\Delta\phi$ distances between the two leptons (bottom). Comparison of the two approaches to model the interference.

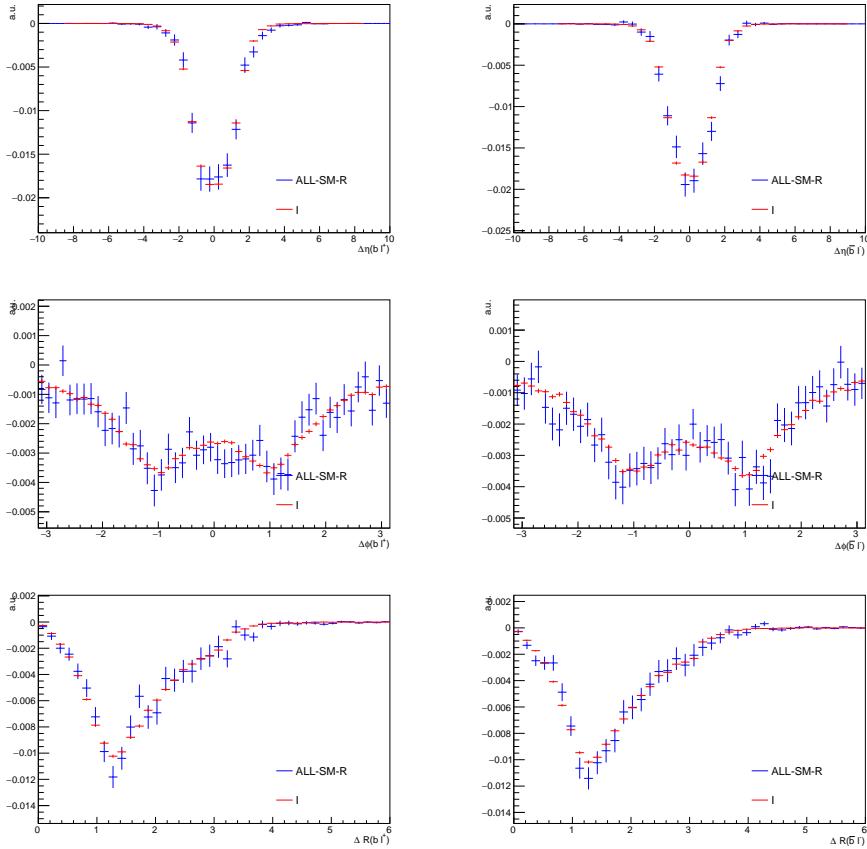


Figure B.5: Distributions of the $\Delta\eta$ (top), $\Delta\phi$ (middle) and ΔR (bottom) between the b quark/antiquark and the charged lepton coming from the decay of the top quark (left) or the top antiquark (right). Comparison of the two approaches to model the interference.

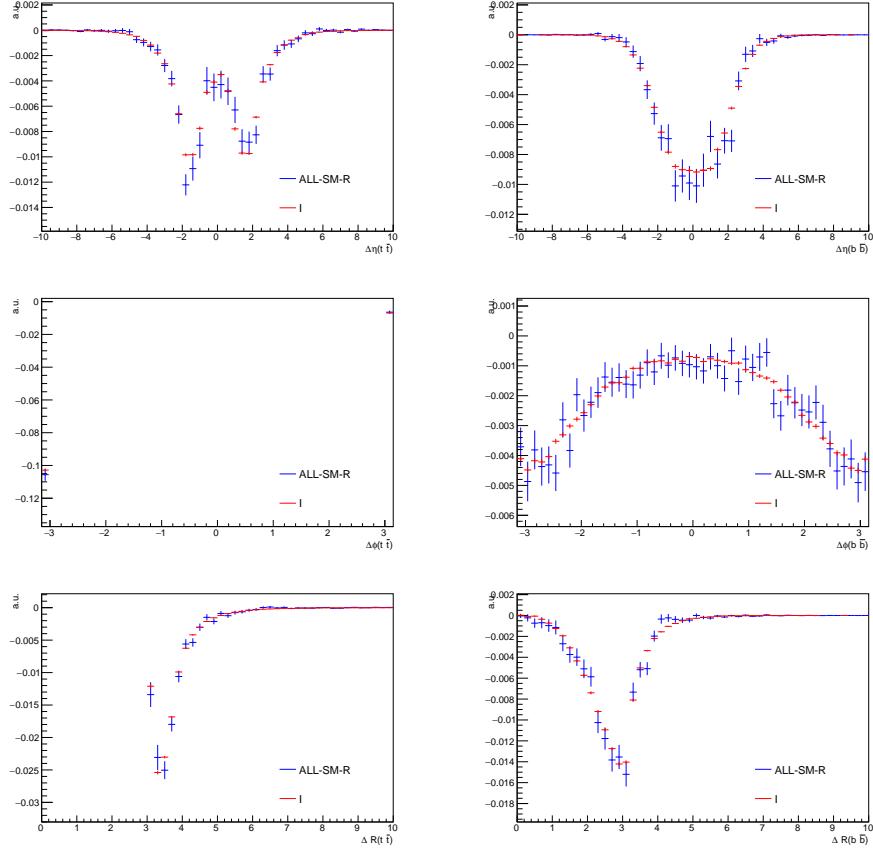


Figure B.6: Distributions of the $\Delta\eta$ (top), $\Delta\phi$ (middle) and ΔR (bottom) between the top quark and antiquark (left) and the b quark and antiquark (right) coming from decays of the top quarks. Comparison of the two approaches to model the interference.

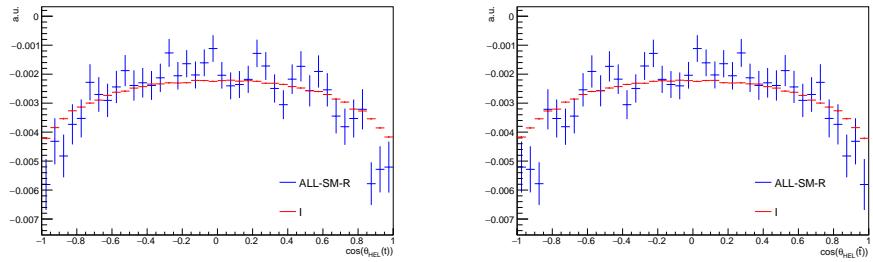


Figure B.7: Distributions of the cosine of the angle between the three-momentum of the top quark (left) and antiquark (right) in the center-of-mass frame of the $t\bar{t}$ system, and the three-momentum of the $t\bar{t}$ system in the laboratory frame. Comparison of the two approaches to model the interference.

C

Data MC comparison

Figures below show distributions of a number of basic observables after the full event selection. All corrections are applied and an overall good description of data is observed, with the notable exception of variables related to pileup, Primary Vertices (PV). This mismodelling is a known result of the dynamic tracker inefficiency and its mitigation in reconstruction algorithms, and it is absorbed in the calibration of physics objects and in corresponding scale factors. The figures also show large statistical fluctuations in the multijet background. This is a direct result of the small effective integrated luminosities of some of the simulated samples. This problem has been addressed in a data-driven description of the multijet background constructed in Section 6.6.

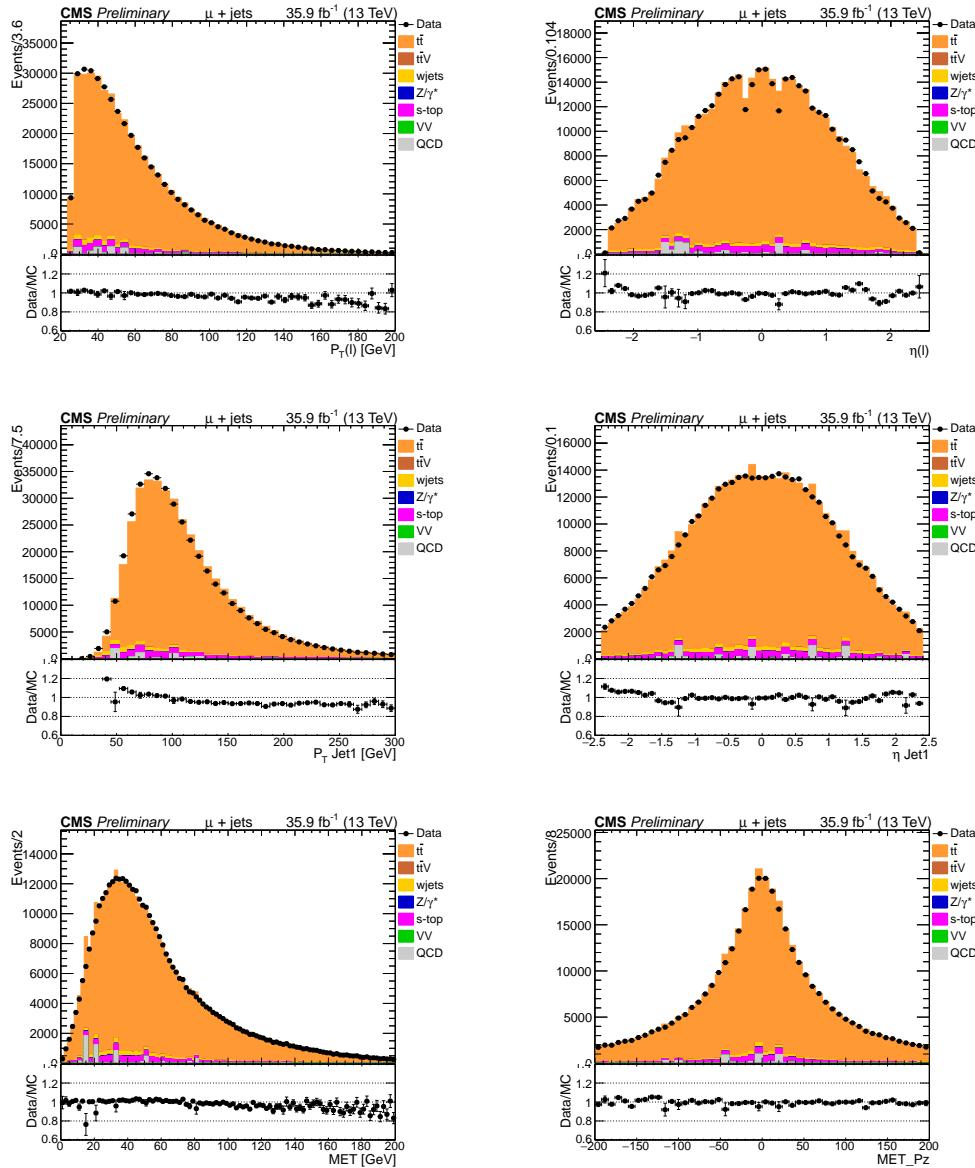


Figure C.1: Basic variables distributions in muon channel, from left to right, transverse momentum, p_T , and pseudorapidity, η , of muon, p_T and η of leading p_T jets, missing transverse energy, p_T^{miss} , and the corresponding reconstructed p_z of p_T^{miss} .

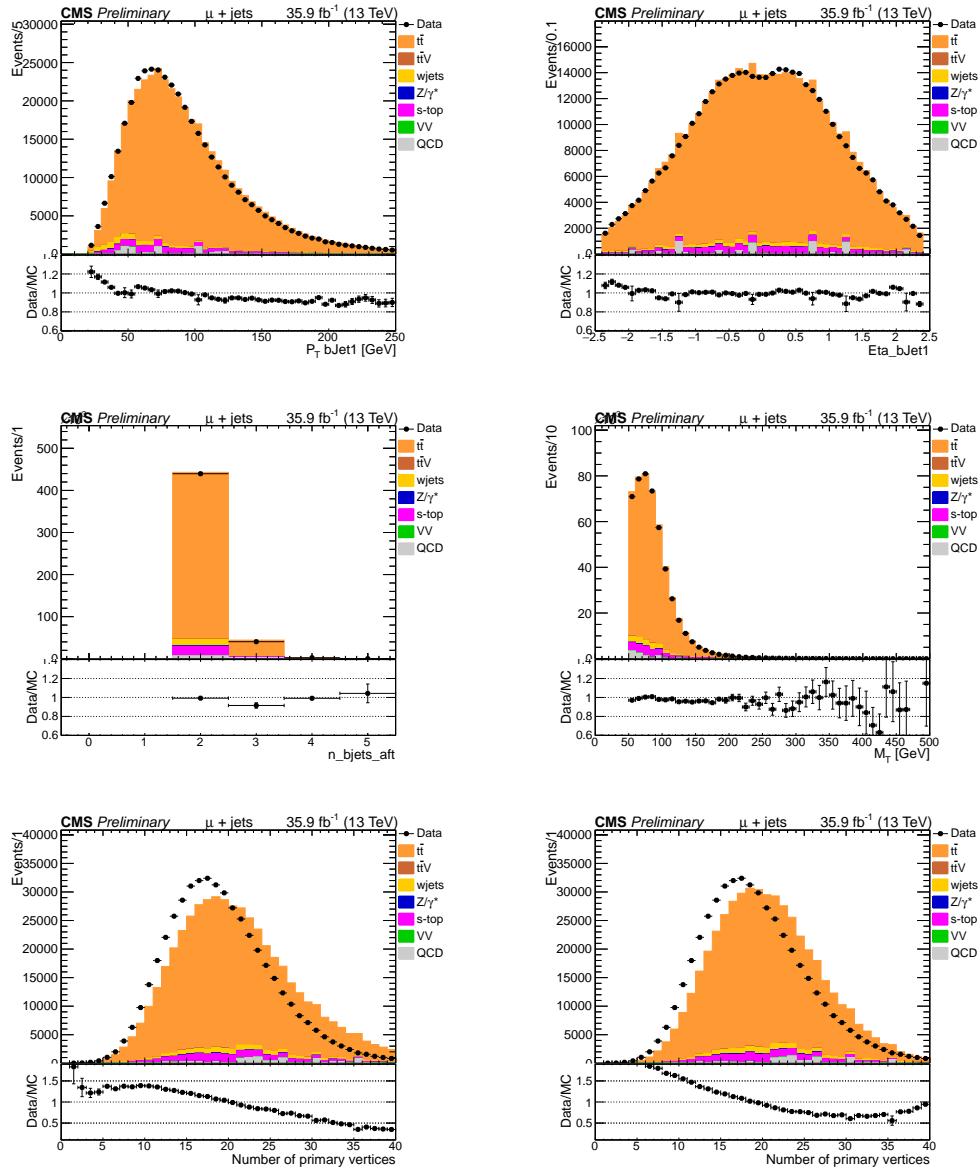


Figure C.2: Basic variables distributions in muon channel, from left to right, transverse momentum, p_T , and pseudorapidity, η , of leading b-jets, number of the selected b-jets and m_T^W , number of primary vertices without and with PU reweighting.

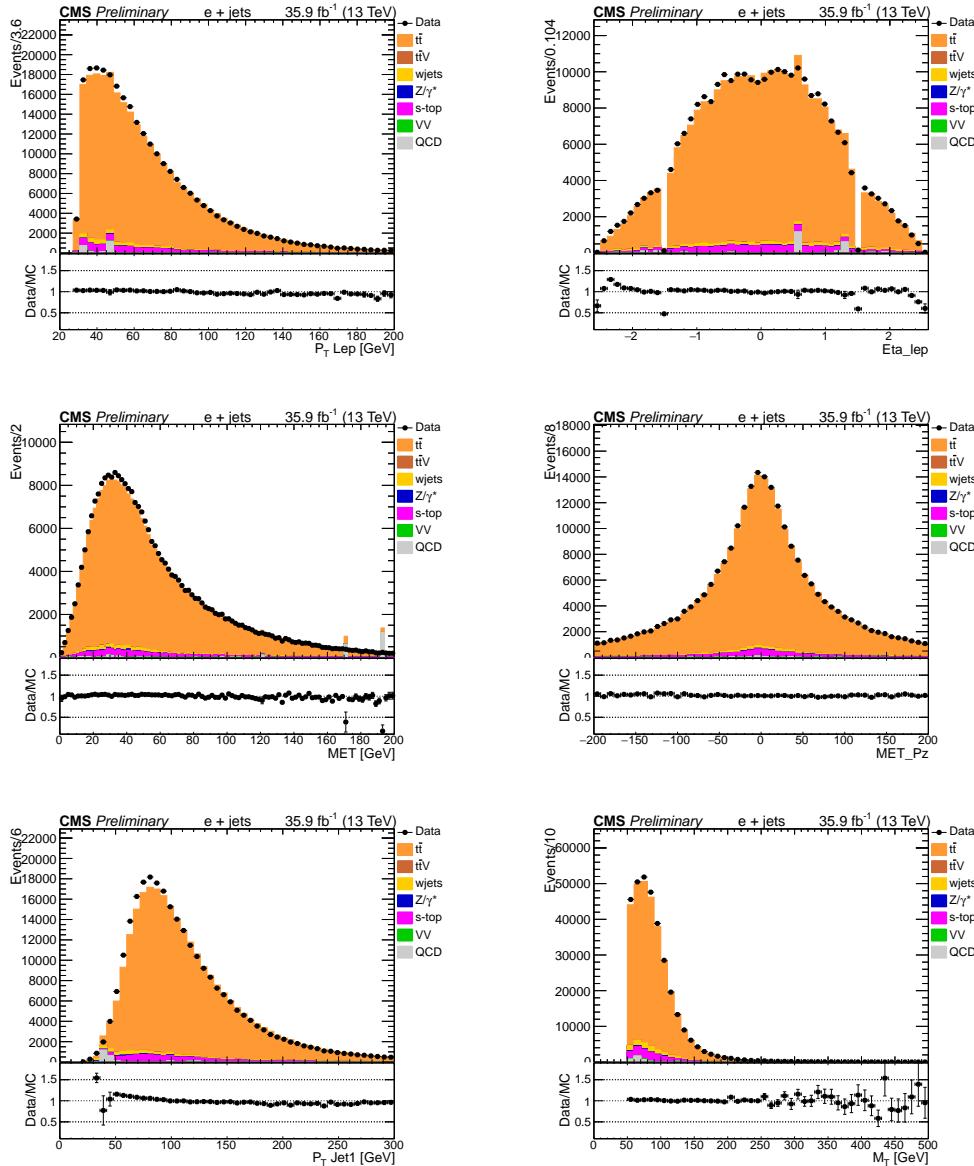


Figure C.3: Basic variables distributions in electron channel, from left to right, transverse momentum, p_T , and pseudorapidity, η , of electron, missing transverse energy, p_T^{miss} , and the corresponding reconstructed p_Z of p_T^{miss} , leading p_T of jets and m_T^W .

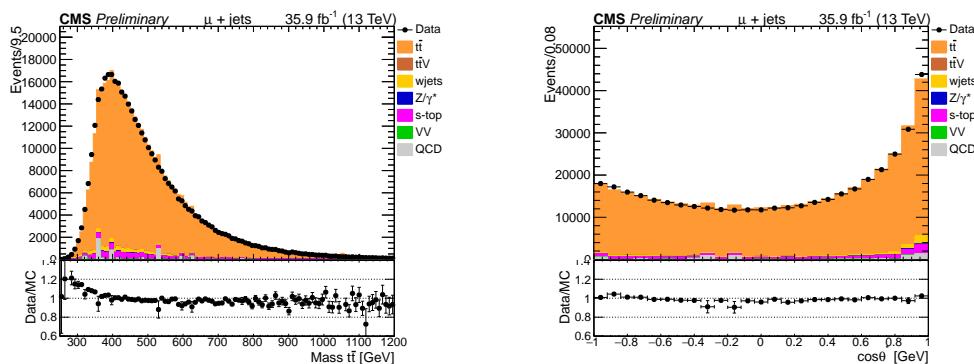


Figure C.4: From left to right, mass of $t\bar{t}$ and $\cos\theta^*$ distributions in muon channel. Both of the variables used in the final statistical evaluation.

D

Impact of systematic uncertainties

This appendix contains plots that show the impact of systematic uncertainties on the results. Section D.0.1 shows the impact that single uncertainties have on the combined obtained signal strength, whereas section D.0.2 shows how different groups of systematic uncertainties contribute to the expected upper limits on the coupling strength.

D.0.1 Single uncertainties

Figures D.1-D.4 (expected) and Figs. D.5-D.8 (observed) give a summary of the 30 most important systematic uncertainties for 4 different scenarios that range from low to high mass, with the widths near the exclusion in the hMSSM scenario, and the value of the coupling modifier adjusted to the central expected limit. Like for the fits carried out to obtain the limits, the coupling modifier is fixed in the fit, and only the signal strength modifier is kept floating. Impacts are hence calculated on the signal strength modifier, with an injected signal strength of unity for the expected impacts. The left parts of the figures contain the post-fit values of these nuisance parameters and their post-fit constraint. The right parts of the figures show the impact of each such uncertainty on the observed signal strength, which can also be understood as the contribution of each systematic uncertainty to the overall uncertainty. It can be noted that, in all cases, theory-related systematic uncertainties constitute the most important sources of systematic uncertainty, in particular also those related to the dominant $t\bar{t}$ background. At lower mass, the uncertainty in the value of the top quark mass is more important, which can be understood from the $m_{t\bar{t}}$ distribution being sensitive because of threshold effects, whereas missing higher orders are more relevant at higher masses.

It can also be noted that some uncertainties are subject to significant constraints. A few of the theoretical uncertainties pertaining to $t\bar{t}$ simulation are expected to be constrained. For example, the variation of FSR in $t\bar{t}$ simulated events and other variations were found to be larger than the uncertainties also in dedicated $t\bar{t}$ studies [109]. Similarly, it has been investigated previously that the $m_{t\bar{t}}$ distribution, in particular the turn-on region, is sensitive to the top quark mass [188]. In terms of experimental uncertainties, we note that there are moderate to tight constraints on the variations from all sources of uncertainties in the jet energy corrections.

D.0.2 Groups

The results presented in this section answer the question “How would the expected limits change if certain groups of systematic uncertainties were irrelevant?”, which can also be understood as the impact the according systematic uncertainties have on the values of the expected upper limits on the coupling strength. Figures D.9-D.12 show the upper limits with either experimental, theoretical cross sections, other theory-related, or bin-by-bin uncertainties removed. Uncertainties in the cross sections of background process have a negligible

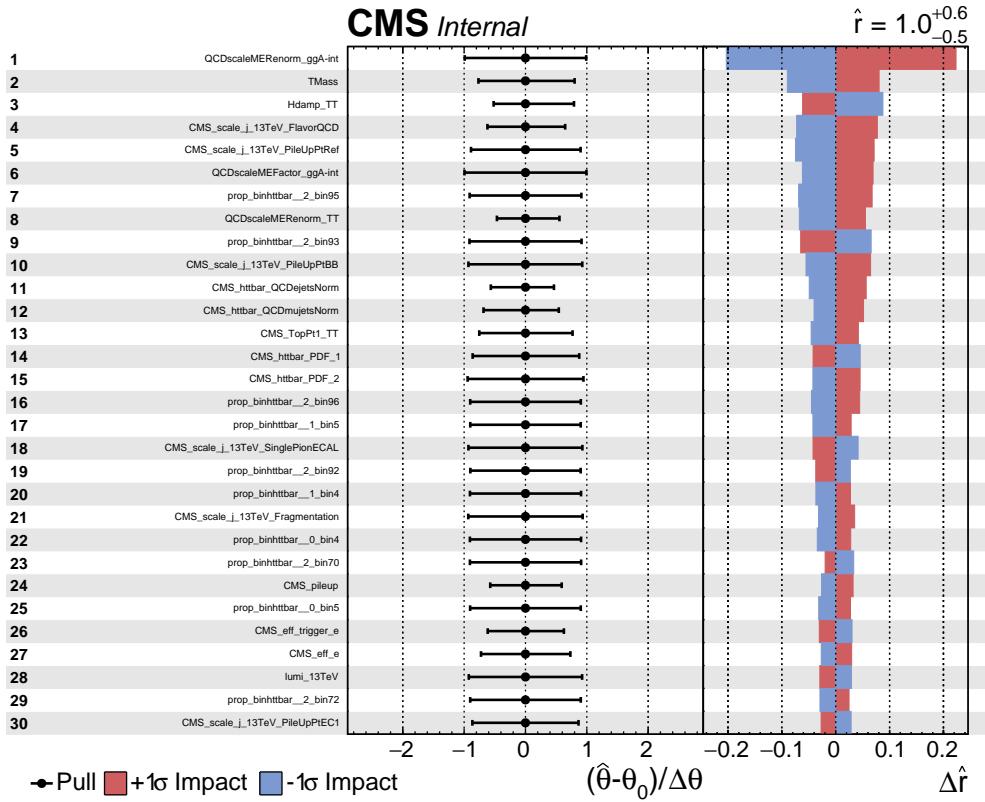


Figure D.1: Impacts of systematic uncertainties on obtained coupling strength modifier for a 400 GeV pseudoscalar with a relative width of 2.5%. The impacts are shown for the Asimov dataset with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

overall impact, and experimental uncertainties also have a minor overall impact. However, both the expected upper limits with bin-by-bin uncertainties and the other theory-related uncertainties removed show a significant improvement with respect to the nominal upper limits. For higher widths, the theory-related uncertainties become yet more important.

To understand the impact of the theory-related uncertainties in more detail, figures D.13-D.16 show the combined impacts of cross-section related uncertainties, missing higher orders as obtained from scale variations, top quark mass, and PDF uncertainties. For $m_\Phi = 400$ GeV and low to moderate width, the top quark mass uncertainty has the largest impact, with very similar impacts from PDF uncertainties and the other two combined sources. Starting from $m_\Phi = 500$ GeV, scale variations become most important, with all other combined sources giving similar additional contributions. The according relative variations are shown in Ref. [114].

D.0.3 Nuisance parameter pulls, constraints, and correlations

Figures D.17 and D.18 show the expected and observed constraints on the nuisance parameters obtained from a fit to the Asimov dataset. MC statistical uncertainties are not included.

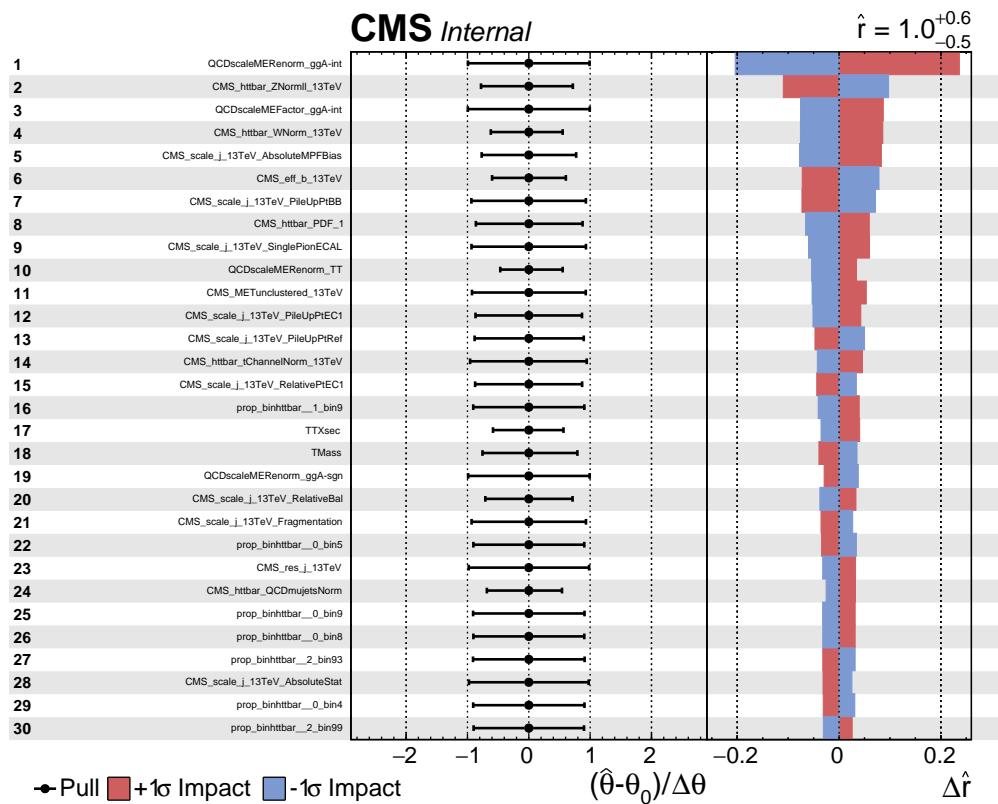


Figure D.2: Impacts of systematic uncertainties on obtained coupling strength modifier for a 500 GeV pseudoscalar with a relative width of 5%. The impacts are shown for the Asimov dataset with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

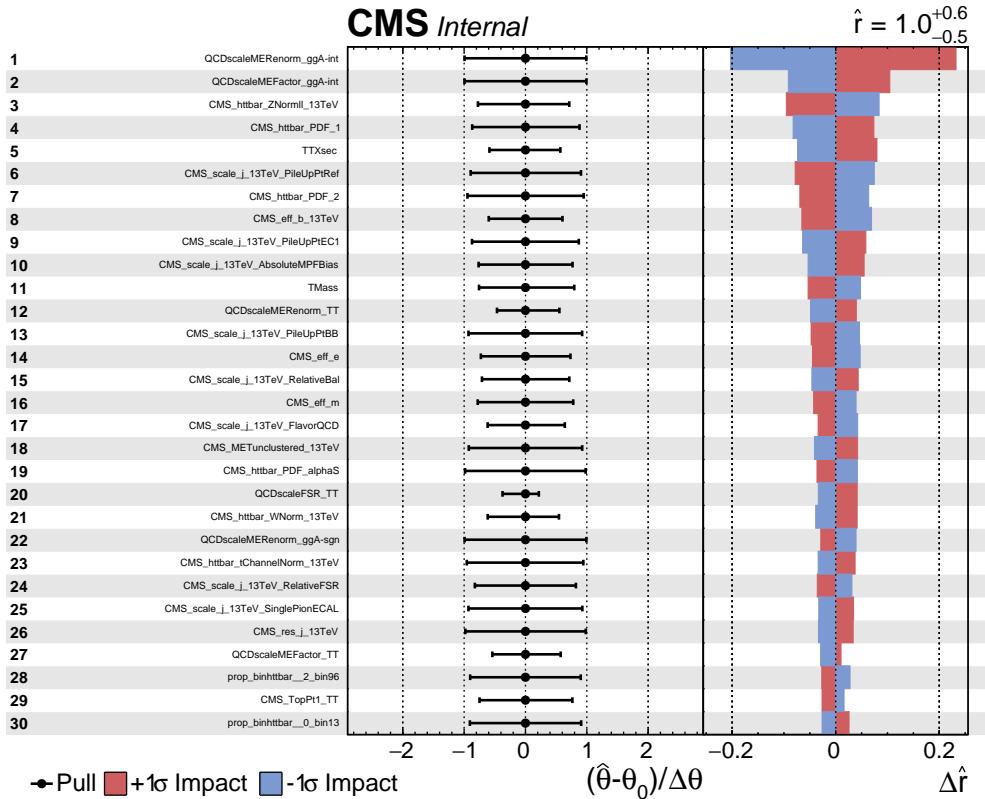


Figure D.3: Impacts of systematic uncertainties on obtained coupling strength modifier for a 600 GeV pseudoscalar with a relative width of 10%. The impacts are shown for the Asimov dataset with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

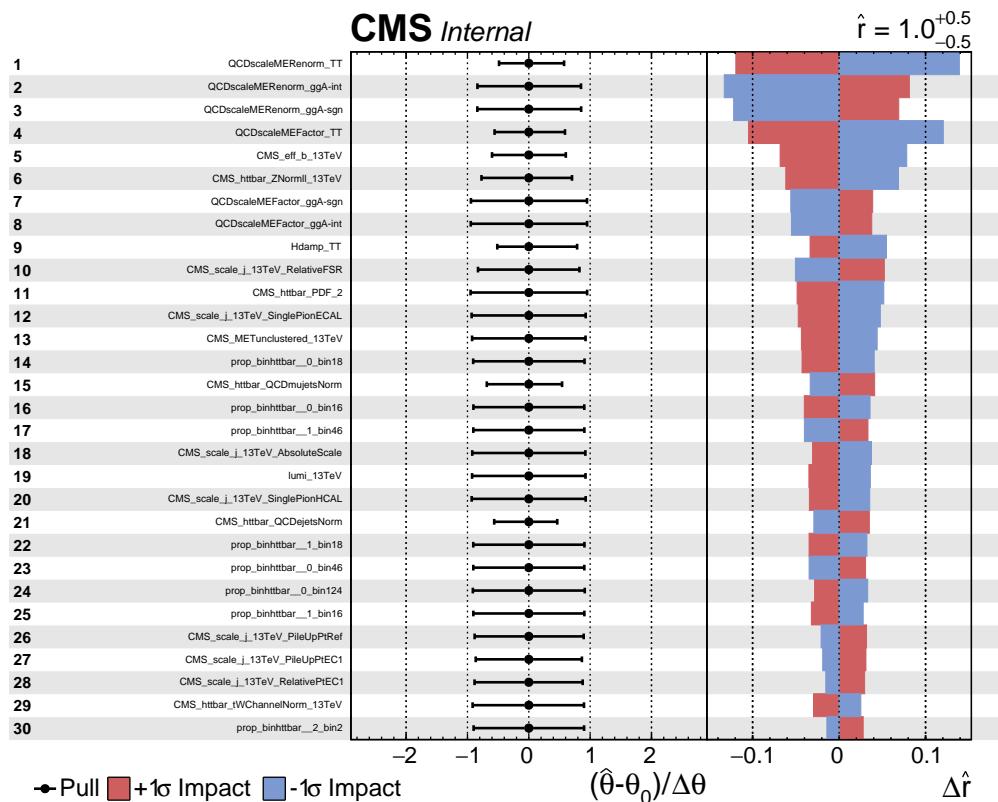


Figure D.4: Impacts of systematic uncertainties on obtained coupling strength modifier for a 750 GeV pseudoscalar with a relative width of 10%. The impacts are shown for the Asimov dataset with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

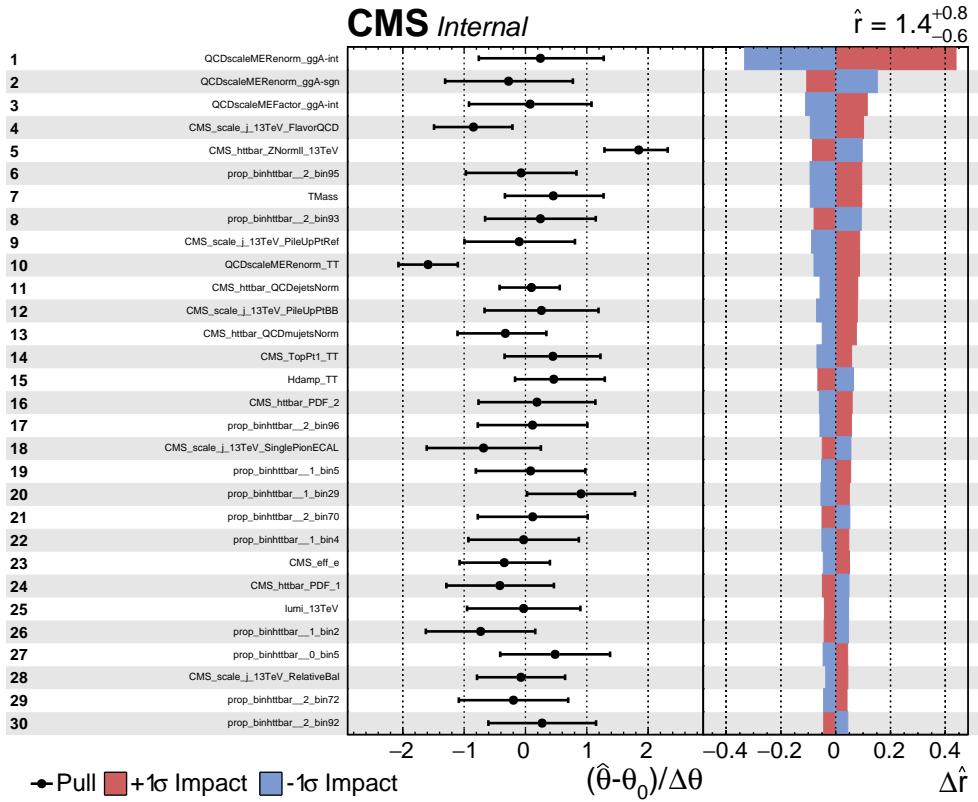


Figure D.5: Impacts of systematic uncertainties on obtained coupling strength modifier for a 400 GeV pseudoscalar with a relative width of 2.5%. The impacts are shown for the observed data with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

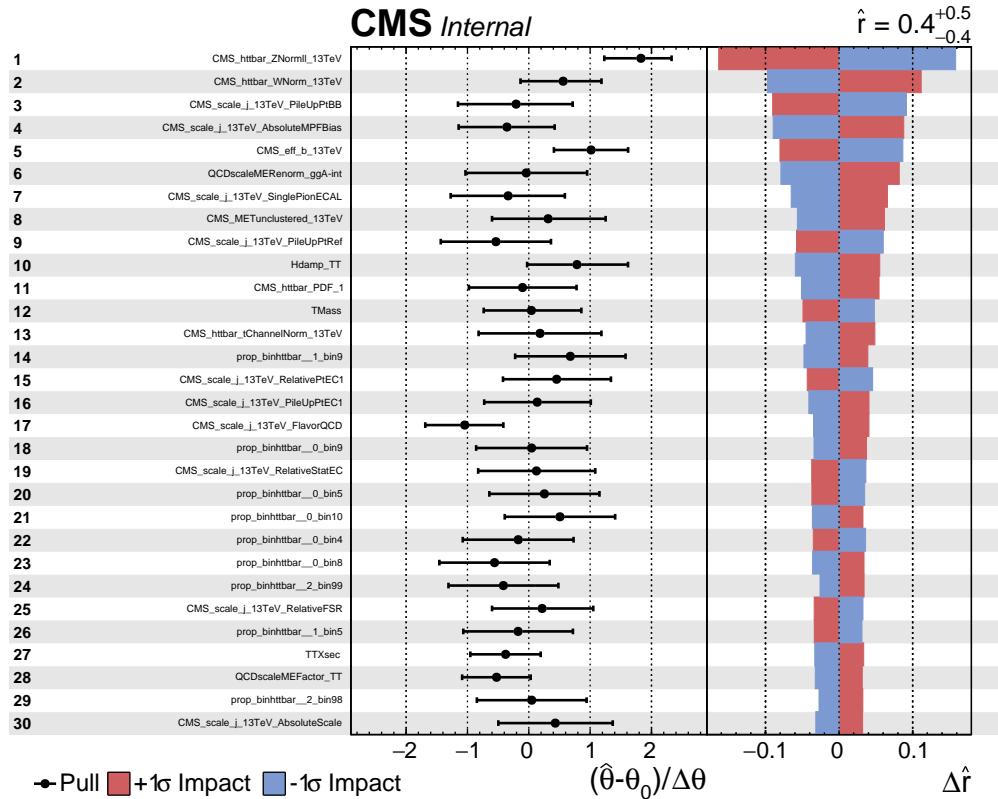


Figure D.6: Impacts of systematic uncertainties on obtained coupling strength modifier for a 500 GeV pseudoscalar with a relative width of 5%. The impacts are shown for the observed data with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

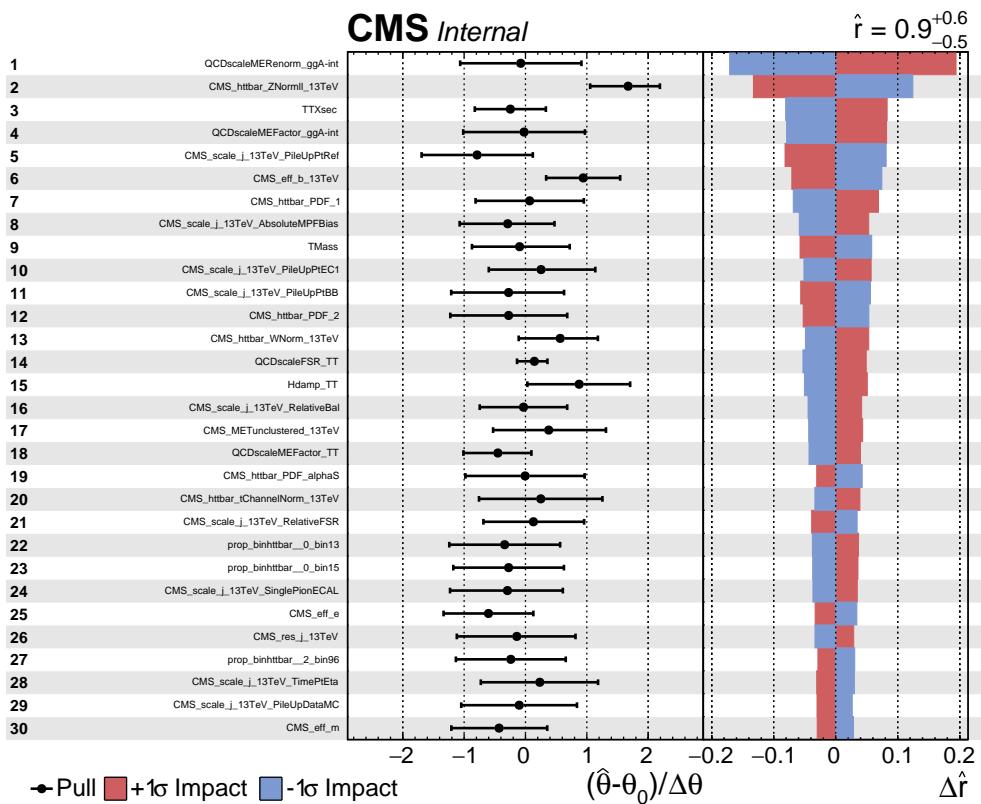


Figure D.7: Impacts of systematic uncertainties on obtained coupling strength modifier for a 600 GeV pseudoscalar with a relative width of 10%. The impacts are shown for the observed data with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

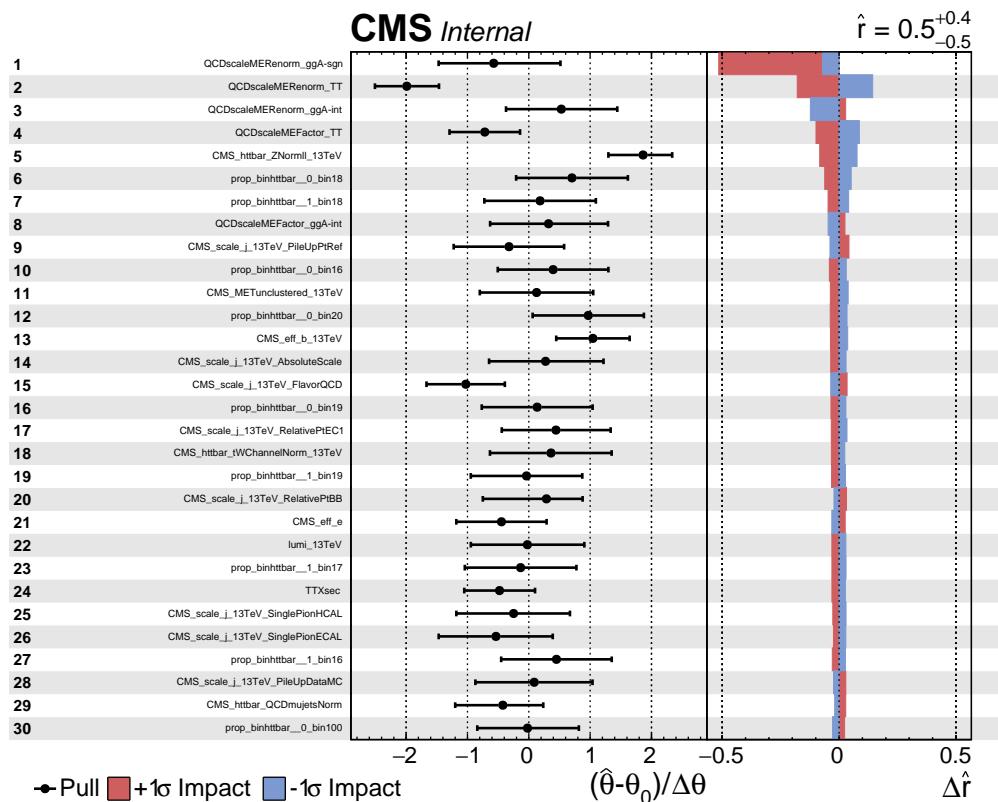


Figure D.8: Impacts of systematic uncertainties on obtained coupling strength modifier for a 750 GeV pseudoscalar with a relative width of 10%. The impacts are shown for the observed data with a coupling strength modifier corresponding to the central expected limit. The right-hand side of the plot shows the relative variation of the obtained coupling strength under variations of the considered nuisance parameter within ± 1 s.d. of its post-fit uncertainty (the so-called impact). The 30 uncertainties with the largest impact are shown. In addition, the post-fit value of the nuisance parameter and its uncertainty is shown on the left-hand side.

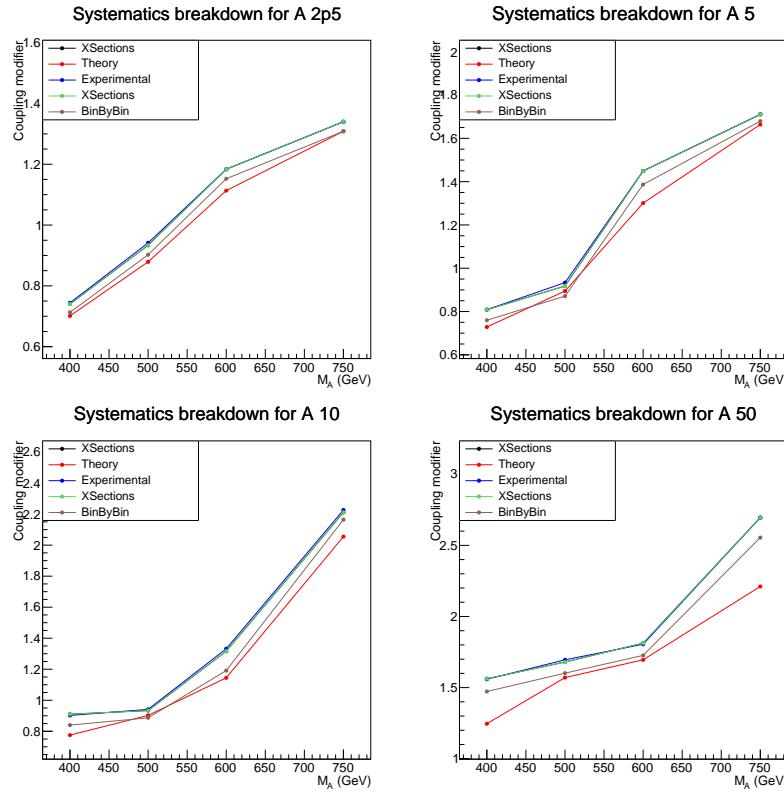


Figure D.9: Expected upper limits on the coupling modifier with different groups of systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for pseudoscalar signal as a function of mass for different values of the relative width, ranging from 2.5 to 50%.

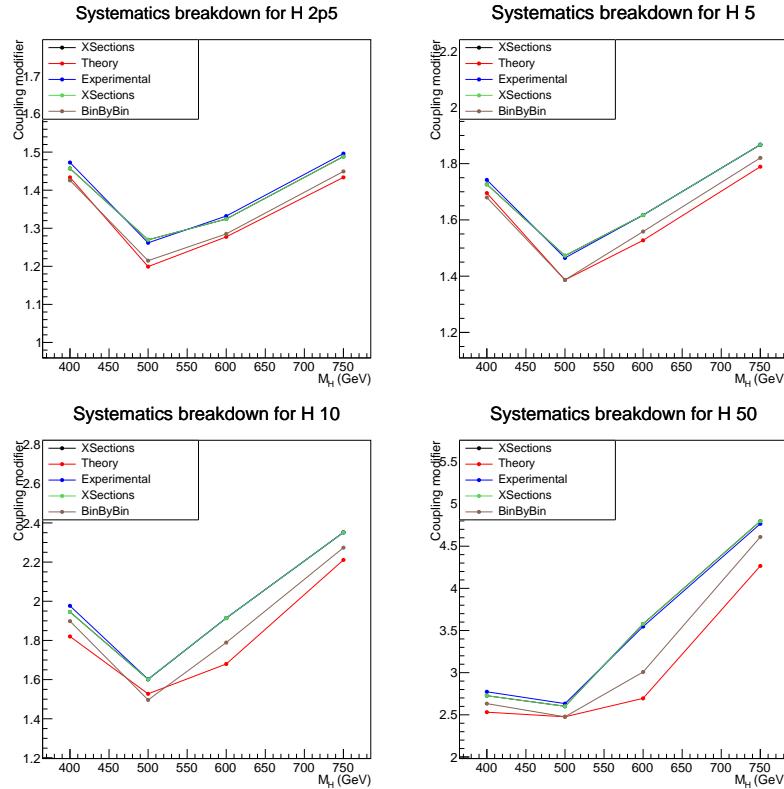


Figure D.10: Expected upper limits on the coupling modifier with different groups of systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for scalar signal as a function of mass for different values of the relative width, ranging from 2.5 to 50%.

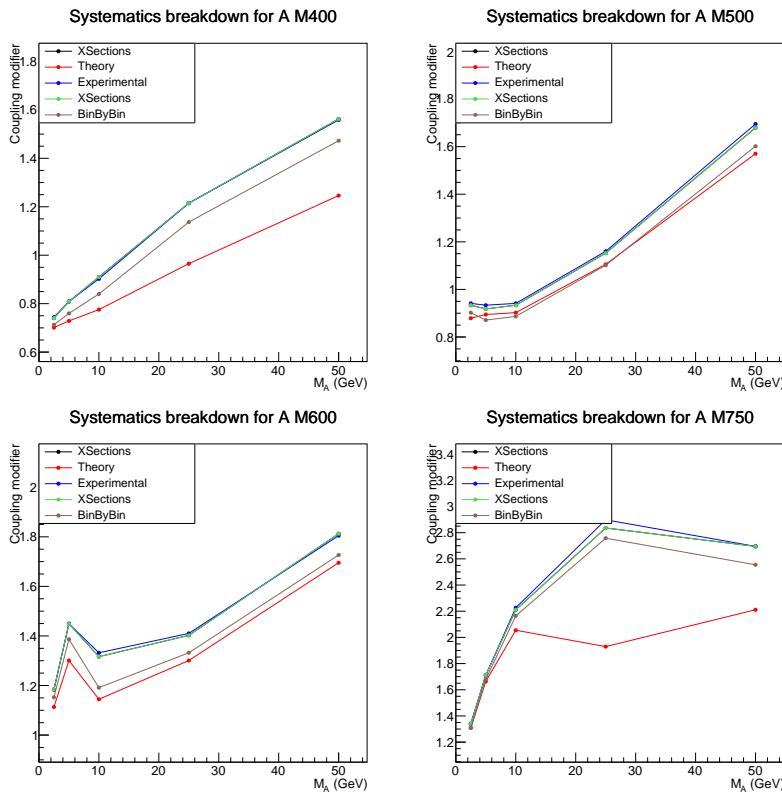


Figure D.11: Expected upper limits on the coupling modifier with different groups of systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for pseudoscalar signal as a function of relative width for different values of the mass, ranging from 400 to 750 GeV.

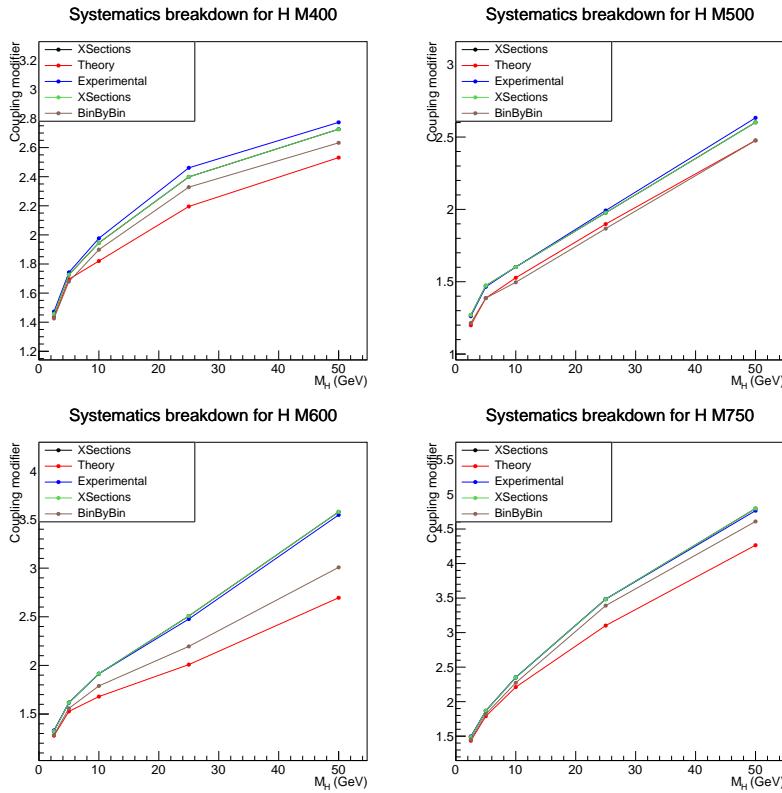


Figure D.12: Expected upper limits on the coupling modifier with different groups of systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for pseudoscalar signal as a function of relative width for different values of the mass, ranging from 400 to 750 GeV.

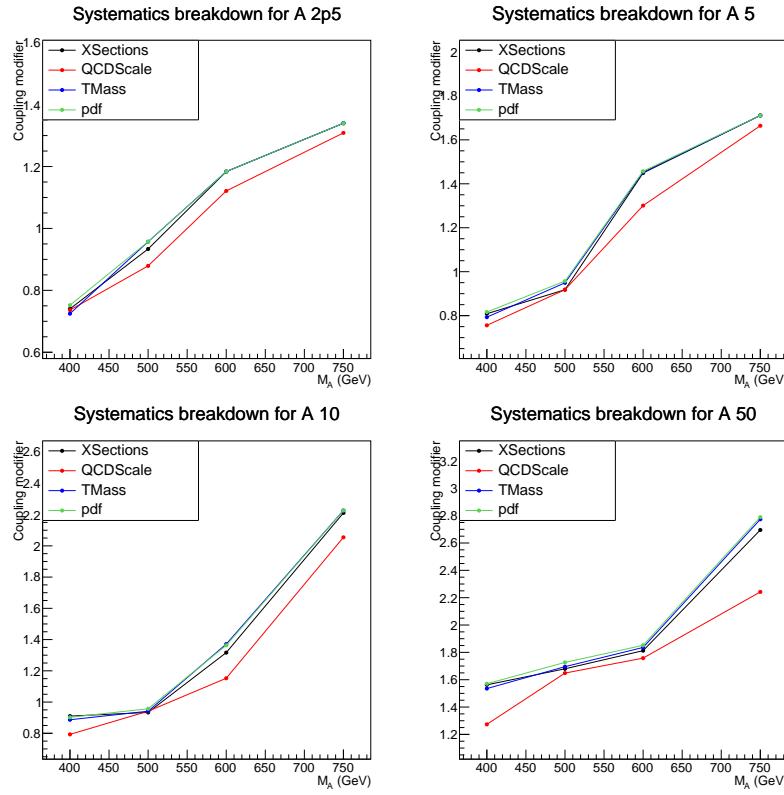


Figure D.13: Expected upper limits on the coupling modifier with different groups of theory-related systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for pseudoscalar signal as a function of mass for different values of the relative width, ranging from 2.5 to 50%.

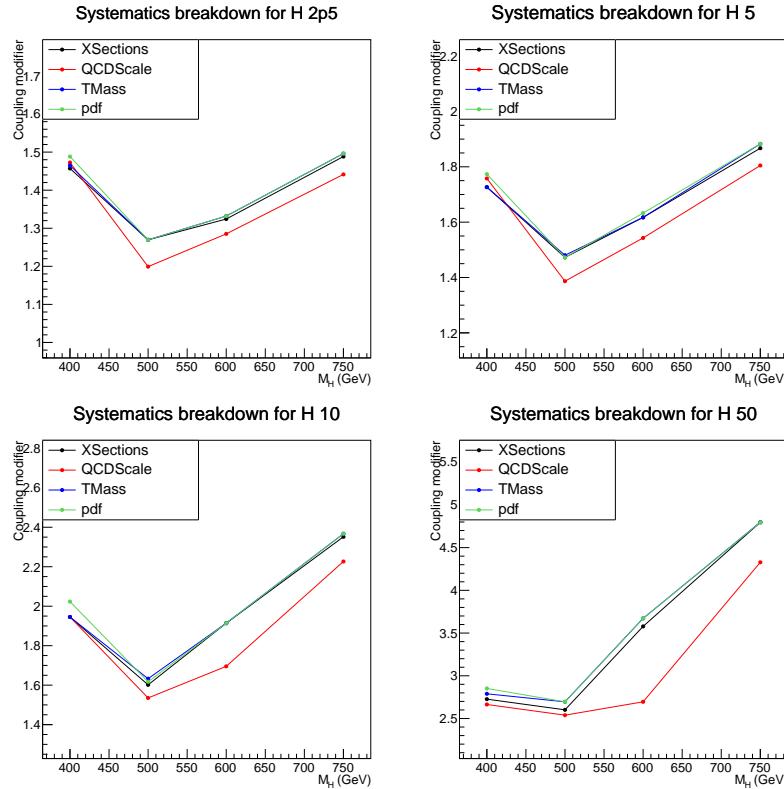


Figure D.14: Expected upper limits on the coupling modifier with different groups of theory-related systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for scalar signal as a function of mass for different values of the relative width, ranging from 2.5 to 50%.

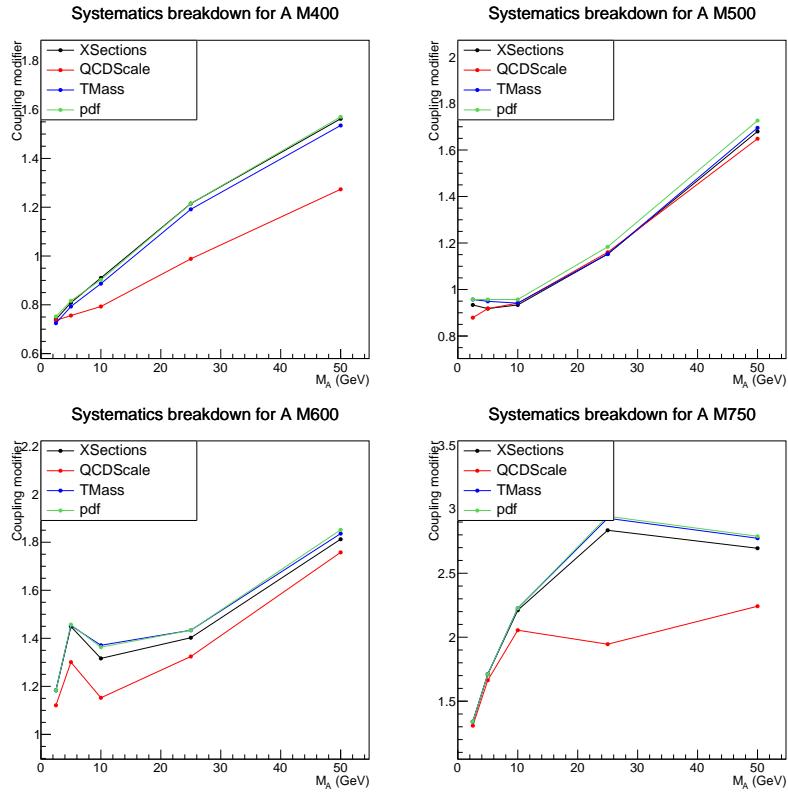


Figure D.15: Expected upper limits on the coupling modifier with different groups of theory-related systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for pseudoscalar signal as a function of relative width for different values of the mass, ranging from 400 to 750 GeV.

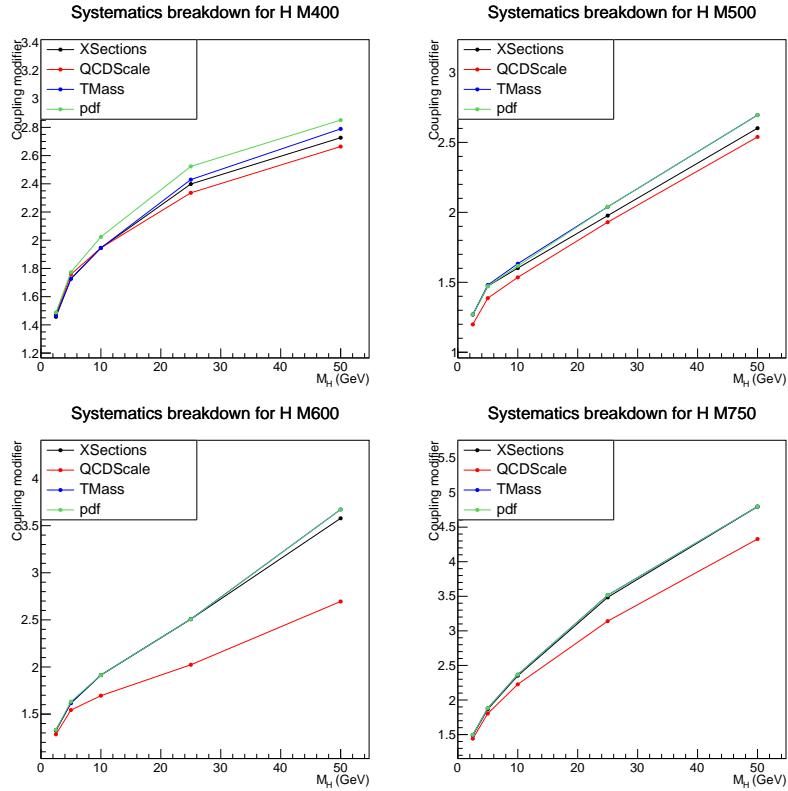


Figure D.16: Expected upper limits on the coupling modifier with different groups of theory-related systematic uncertainties removed. The nominal limits are shown as the blue curve. The expected upper limits are shown for pseudoscalar signal as a function of relative width for different values of the mass, ranging from 400 to 750 GeV.

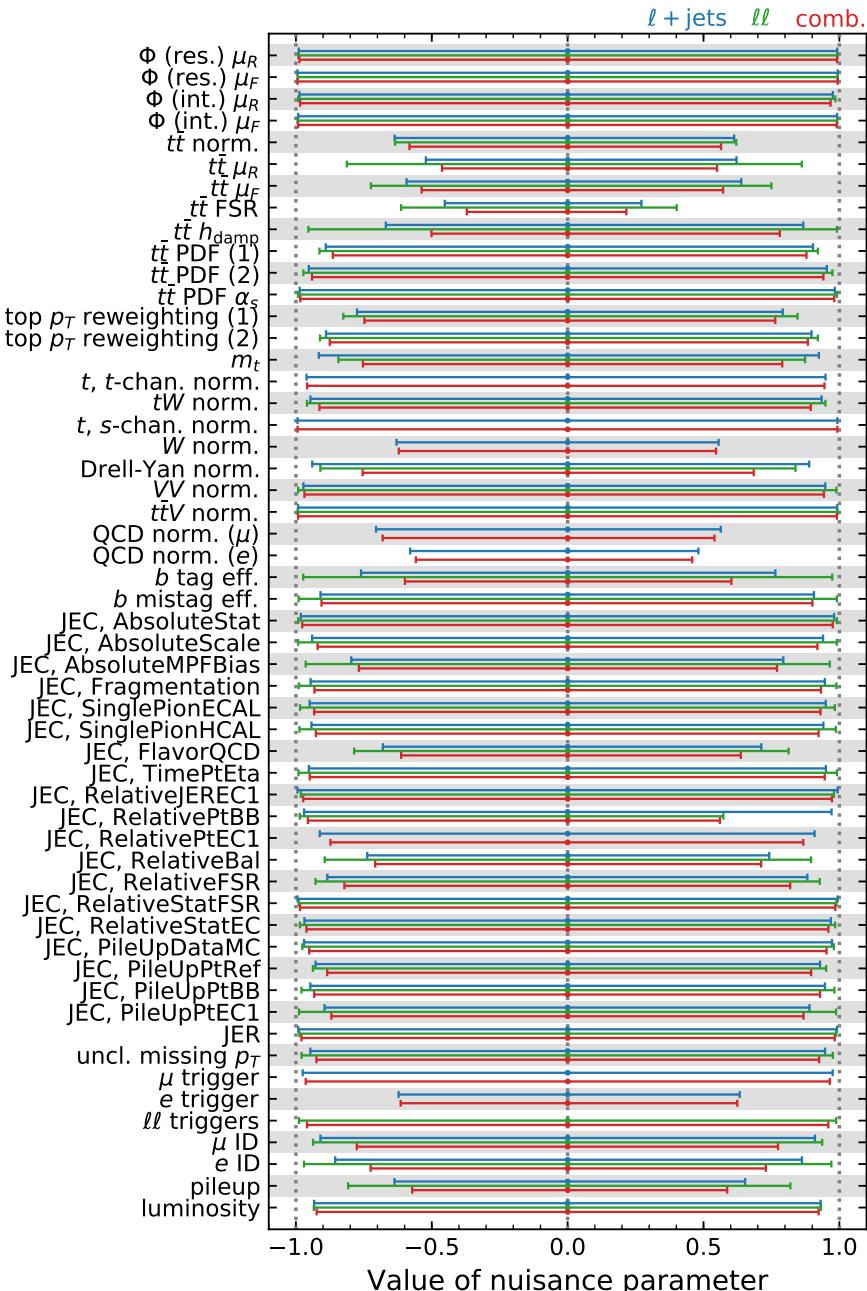


Figure D.17: Constraints on nuisance parameters from a fit to the Asimov dataset, shown separately for the lepton-plus-jets (green) and the dilepton channel (red), and their combination (black).

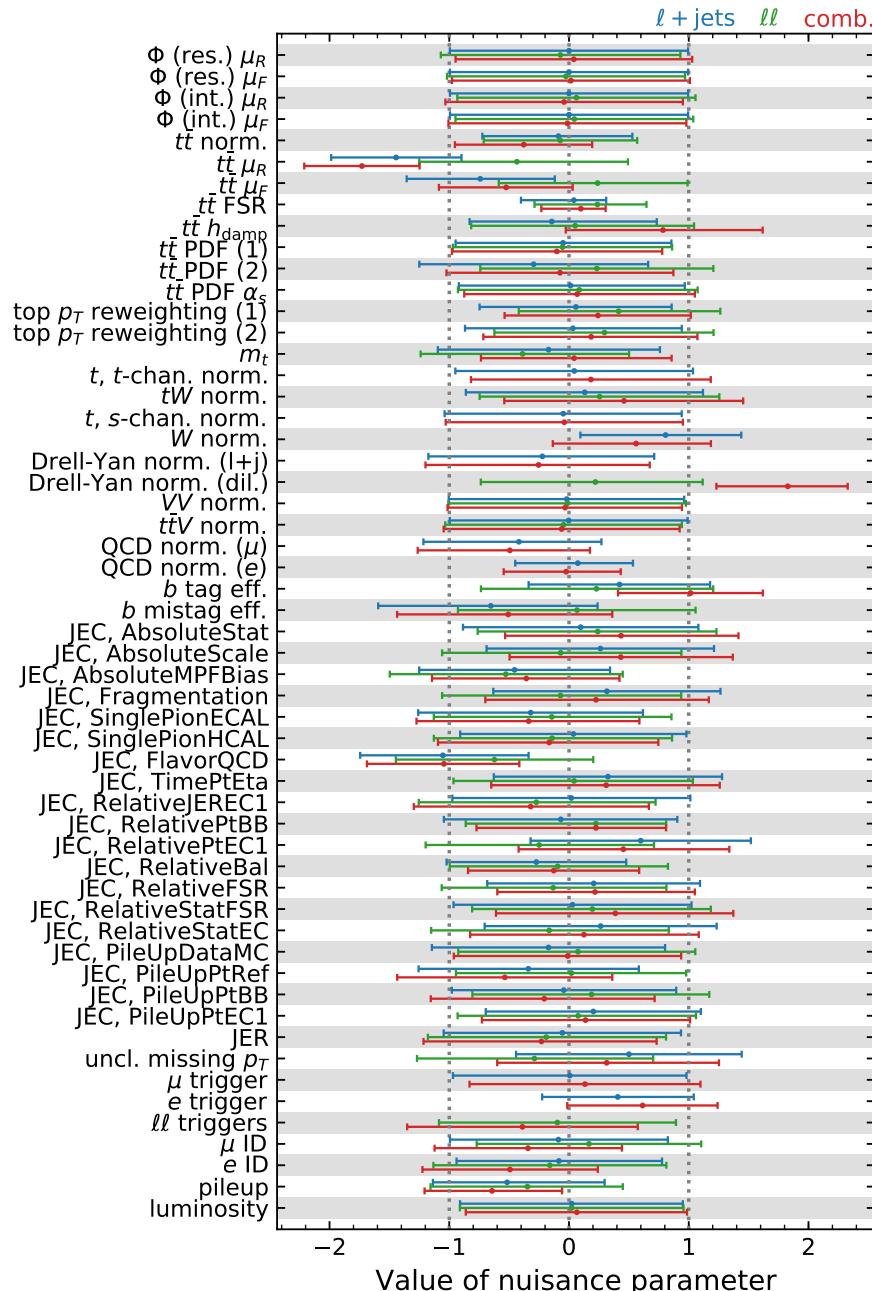


Figure D.18: Constraints on nuisance parameters from a fit to data, shown separately for the lepton-plus-jets (green) and the dilepton channel (red), and their combination (black).

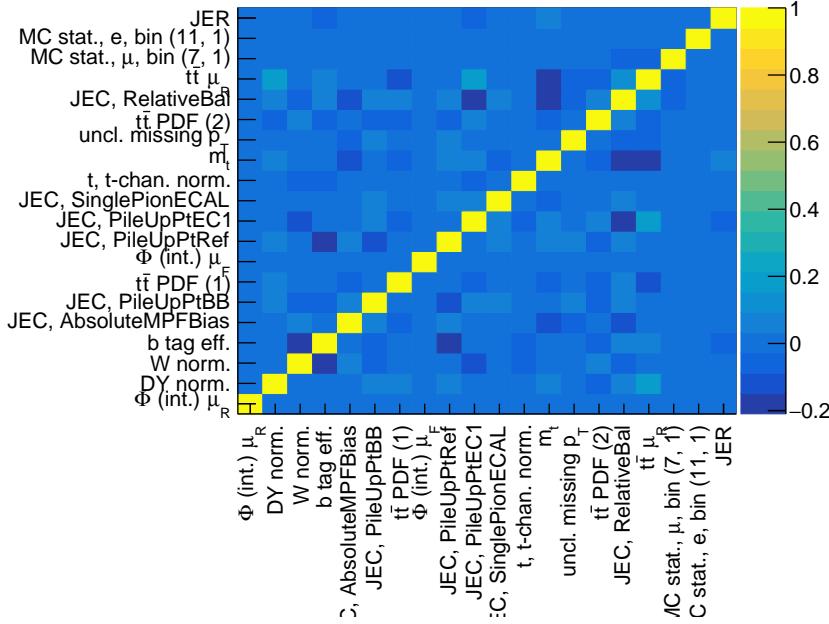


Figure D.19: Correlations of 20 nuisance parameters with largest impact from a fit to the Asimov dataset. Both the correlations and the set of nuisance parameters with largest impacts are obtained using the benchmark point of $m_A = 500 \text{ GeV}$ and a relative width of 5%.

Figure D.19 shows the correlations of the 20 nuisance parameters with largest impact, obtained from a fit using $m_A = 500 \text{ GeV}$ and a relative width of 5%, and the signal strength corresponding to the central expected limit.