Probing new physics through Standard Model Higgs to diphoton signature in pp collisions at $\sqrt{s} = 13$ TeV

By

PRAFULLA SAHA

Enrolment No

PHYS11201704020

National Institute of Science Education and Research, Bhubaneswar

A thesis submitted to the

Board of Studies in Physical Sciences

(as applicable)

In partial fulfillment of requirements

for the Degree of

DOCTOR OF PHILOSOPHY

of

HOMI BHABHA NATIONAL INSTITUTE

June, 2023

Homi Bhaba National Institute

Recommendations of the Viva Voce Committee

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by **Prafulla Saha** entitled **Probing new physics through Standard Model Higgs to diphoton signature in pp collisions at $\sqrt{s}$ = 13 TeV** and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman - Prof. Bedangadas Mohanty	*Bomohanty*	6/10/2023
Guide / Convener - Dr. Prolay Kumar Mal		October 6, 2023
Examiner - Prof. Ritesh Kumar Singh		6/10/2023
Member 1 - Prof. Satyaki Bhattacharya		06/10/2023
Member 2 - Prof. Sanjay Kumar Swain		6/10/23
Member 3 - Prof. Yogesh Kumar Srivastava	*Yogesh*	6/10/2023

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I hereby certify that I have read this thesis prepared under my direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date : October 6, 2023

Place : NISER

Signature

Guide

i

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

Prafulla Saha

DECLARATION

I hereby declare that I am the sole author of this thesis in partial fulfillment of the requirements for Degree of Doctor of Philosophy from National Institute of Science Education and Research (NISER). I authorize NISER to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Prafulla Saha

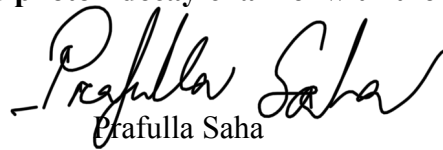<div align="center">**List of Publications arising from the thesis**</div>

Journal

1. CMS Collaboration, "Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV", JHEP 07 (2021) 027,331 doi:10.1007/JHEP07(2021)027, arXiv:2103.06956.

2. CMS Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery", Nature 607, 60(2022) doi:10.1038/s41586-022-04892-x, arXiv:2207.00043

3. CMS Collaboration, "Search for a vector-like quark $T' \to tH$ in the diphoton decay mode of the Higgs boson in proton-proton collisions at $\sqrt{s} = 13$TeV", 10.1007/JHEP09(2023)057, arXiv:2302.12802

Chapters in books and lectures notes

1. Proceedings of the XXIV DAE-BRNS High Energy Physics Symposium, Jatni, India Chapter 33: Recent Measurements of Higgs Boson Properties in the Diphoton Decay Channel with the CMS Detector https://doi.org/10.1007/978-981-19-2354-8_33

Conferences

1. XXV DAE-BRNS Symposium on HEP, IISER Mohali, India **Poster title**: Search for a vector-like quark in the diphoton decay mode of the Higgs boson in proton-proton collisions at $\sqrt{s}$=13 TeV

2. Lepton Photon-2021, University of Manchester, UK **Talk title**: Measurements of Higgs boson couplings and differential cross-sections

3. NUCLEUS-2021, Petersburg, Russia **Talk title: CMS Higgs results**

4. XXIV DAE-BRNS Symposium on HEP, NISER, India **Talk title: Recent measurements of Higgs boson properties in the diphoton decay channel with the CMS detector**

Prafulla Saha

DEDICATIONS

This thesis is dedicated with heartfelt gratitude and love to the memory of my late father, whose unwavering support and guidance continue to inspire me. I also dedicate this work to my cherished mother and brother, whose enduring encouragement and belief in my abilities have been a constant source of strength throughout my academic journey.

Prafulla Sala

ACKNOWLEDGEMENTS

constant source of strength and motivation during the ups and downs of this academic pursuit.

I am deeply grateful to my seniors and labmates whose collaboration and shared experiences have played a crucial role in shaping this journey. I would like to acknowledge the contributions of Chandi, Kuldeep, Amit, Gouranga, Shuvankar, Deepak, Shivam, Srinivas, Ayushkant, Lipsa, Priyanka, Sneh, Rahul, Tanay, Sarjita and among others, for their unwavering support, the exchange of ideas, and their readiness to help each other. I am deeply grateful to all who have played a part, no matter how significant or minor, in completing this thesis.

ABSTRACT

The discovery of the Higgs boson during the Run 1 of LHC has provided the answer to the long-standing puzzle of electroweak symmetry breaking at least within the Standard Model. Recent ATLAS and CMS results strongly indicate the Higgs boson couples with the SM particles as predicted by the SM. Presently, various production mechanisms (ggF, VBF, VH and ttH) and decay modes of the Higgs boson (at the LHC) are being explored to achieve further precision on the existing measurements, while looking for any deviation from the SM predictions. $H \rightarrow \gamma\gamma$ decay is an interesting final state signature due to the involvement of a virtual top quark loop and thus any beyond the Standard Model (BSM) physics involving Higgs boson production would lead to an excess in the diphoton invariant mass spectrum (from Higgs). Furthermore, during the Run 2 of LHC, an additional Higgs boson production mode associated with a single top quark (tH) is explored to improve the sensitivity further for the $H \rightarrow \gamma\gamma$ analysis and also to explore associated possibilities beyond the SM.

After the discovery of the Higgs boson, the SM is complete as a low-energy effective theory describing all known fundamental particles and their interactions. However several problems still remain unanswered including why the quantum correction of the Higgs boson mass diverges to the Plank scale. In many models and extensions of the standard model, inclusion of a new type of fourth generation particles, called vector-like quarks, T' and B', provides a feasible solution. A dedicated analysis is performed to search for production of the T' vector-like quark in proton-proton collisions at a center-of-mass energy of 13 TeV using data corresponding to an integrated luminosity of 137 fb^{-1} collected by the CMS detector during the 2016-2018 LHC run. The search is focused on a T' quark decaying to a SM top quark and a Higgs boson ($T' \rightarrow tH$), with the Higgs boson further decaying to two photons ($H \rightarrow \gamma\gamma$). This analysis presents the asymptotic upper limit of the T' production signal strength ($\mu = \sigma_{\text{obs}}/\sigma_{\text{SM}}$) over the T' mass range of 600 GeV to 1200 GeV. This search is the first T' search to exploit the decay of the Higgs boson in the diphoton channel. The excellent diphoton invariant mass resolution of 1–2% results in an increased sensitivity compared to previous searches, for narrow T' states with masses up to 1.1 TeV.

Contents

Summary

One part of the thesis present a detailed investigation into the Higgs boson production associated with the singe top quark, focusing on the decay of the Higgs boson into two photons and the subsequent leptonic decay of the top quark using the data collected by the CMS detector during period of 2016-2018 LHC run corresponding to an integrated luminosity of 137 fb^{-1}. The signal strength modifiers for the Higgs boson production associated with the single top quark mode is determined through a simultaneous fit with all other Higgs production modes is found to be, $\mu_{tH} = 6.24^{+0.62}_{-0.54}(\text{Syst})^{+3.67}_{-3.33}(\text{Stat})$. It is important to note that the results are currently limited by statistical uncertainties, and further data is required to reduce these uncertainties and draw conclusive interpretations. The cross section of this SM Higgs production mode, convoluted with H $\rightarrow \gamma\gamma$ branching fraction ($\sigma_{tH} \times \mathcal{B}_{H\rightarrow\gamma\gamma}$), is observed to be $1.3^{+0.7}_{-0.7}$ fb. After combining with all other Higgs production mechanisms, the Higgs boson production signal strength in H $\rightarrow \gamma\gamma$ decay mode is observed to be $1.12^{+0.09}_{-0.09}$.

The other part of the thesis presents a dedicated analysis performed to search the Vector Like Quark T$'$ production in proton-proton collision at a center-of-mass energy of 13 TeV using the data corresponding to an integrated luminosity of 137 fb^{-1} collected by the CMS detector during period of 2016-2018 LHC run. Search is focused on T$'$ quark which decays to a SM top quark and the Higgs boson (T$' \rightarrow$ tH), then Higgs is further decayed to the two photon(H $\rightarrow \gamma\gamma$) and top quark decays both hadronically ($t \rightarrow bq\bar{q}$) and leptonically ($t \rightarrow bl\bar{\nu}$). This analysis presents the asymptotic upper limit of the T$'$ production signal strength ($\mu = \sigma_{obs}/\sigma_{SM}$) of T$'$ mass range of 600 GeV to 1200 GeV. This search is the first T$'$ search to exploit the decay of the Higgs boson in the diphoton channel. The excellent diphoton invariant mass resolution of 1–2% results in an increased sensitivity compared to previous searches, for narrow T$'$ states with masses up to 1.1 TeV. The electroweak production of a T$'$ quark with mass up to 960 GeV is excluded at 95% confidence level, assuming a coupling strength $\kappa_T = 0.25$ and a relative decay width $\Gamma/M_{T'} < 5\%$

List of Figures

3

14

List of Tables

21

Chapter 1

Introduction

The matter around us is composed of fundamental elementary particles, a concept dating from the 6th century. Ancient Indian philosophers proposed the existence of elementary particles between the 9th and 5th century BC. They believed that the universe was made up of countless invisible particles known as "permanu". The philosophical doctrine of atomism and the nature of elementary particles was studied by philosophers in ancient Greece and India, as well as early modern European physicists like Pierre Gassendi, Robert Boyle, and Isaac Newton. However, these ideas were purely abstract and philosophical, lacking empirical observation and experimental evidence, and were just one of several lines of thought. In the 19th century, the chemist, physicist, and meteorologist John Dalton introduced an atomic theory in the field of chemistry which concluded that each element was composed of a single element called the "atom", derived from the Greek word atomos, meaning "indivisible" or "uncut". However, by the end of the 19th century, physicists discovered that atoms are not the fundamental particles of nature. The first fundamental particle discovered was the electron, by the British physicist J. J. Thomson in 1897. He demonstrated that cathode rays are composed of negatively charged particles, known as electrons. Throughout the 20th century, the advancement of science and engineering, along with imaginative theorists and nuclear/particle physicists, brought us from studying the scattering of light to studying particle collisions in giant particle accelerators [16]. During the 1950s, 1960s, and 1970s, guided by a series of experimental discoveries and the corresponding inferences, theorists developed a theory of symmetries in nature known as the Standard Model (SM) of particle physics [17, 18]. The SM is a mathematical framework that was developed within

the context of Quantum Field Theory (QFT) to explain the fundamental forces, particles, and their interactions. This model unifies three of the four fundamental forces of nature – the electromagnetic force, the strong nuclear force, and the weak nuclear force – within a QFT framework that includes both force carriers (bosons) and matter particles (fermions). The model is capable of encapsulating numerous experimental observations and predicting a wide range of phenomena. Various aspects of the SM relevant for this dissertation are detailed further in Sec. 1.1.

The discovery of the Higgs boson in 2012 by the ATLAS and CMS experiments [19, 20] at the Large Hadron Collider (LHC) marked the completion of the SM. The accuracy of the SM's predictions has been extensively tested through a variety of experiments in recent decades, and it has proven to be highly successful in explaining the results of these experiments, without any apparent inconsistencies. However there are quite a few fundamentals of our universe that remain unexplained within the SM, such as the dark matter and dark energy, the matter anti-matter asymmetry, the non-zero masses of neutrinos, etc.

In addition, there are several theoretical and conceptual issues in the SM that suggest the necessity for new physics beyond the SM. For example, the hierarchy problem, the naturalness problem [21], and the strong CP problem all point to the existence of new physics at energy scales beyond those accessible by current experiments. Therefore, experimental efforts are underway to discover new particles that could provide evidence for such physics beyond the SM. These efforts include collider experiments such as at the LHC at CERN, as well as experiments studying neutrinos, cosmic rays, and dark matter. In this dissertation, the studies focus on investigating new physics beyond the SM through high-energy proton-proton collisions at the LHC.

This dissertation is structured as follows: the remaining part of this chapter presents a detailed introduction to the SM, the Higgs boson, and the broader context of beyond standard model (BSM) physics. Chapter 2 provides an overview of the LHC and the CMS

detectors. Chapter 3 discusses the search for the production of a SM Higgs boson associated with a single top quark (tH), while Chapter 4 focuses on the search for a vector-like quark, T′ → tH, with the Higgs boson decaying to two photons (H → $\gamma\gamma$).

1.1 Standard Model

The Standard Model of particle physics [17] is currently the most advanced mathematical formalism to explain nature at its fundamental level and accurately describes experimental observations regarding three of the four fundamental forces of nature: electromagnetic, strong, and weak. This theory, which is based on a gauge theory with the special unitary group SU(N), describes the behavior of elementary particles and their interactions. The SM is the non-abelian gauge theory with the gauge group SU(3) × SU(2) × U(1), and it comprises twelve gauge bosons: the photon, three weak bosons, and eight gluons.

The SM particles can be classified into three groups based on their spin-angular momentum: fermions, gauge bosons, and the Higgs boson. Fermions are particles with a spin of half-integer, and they are categorized into generations based on their mass. They can also be further divided into quarks and leptons, based on their interaction under SU(3). The SM contains a total of 24 fermions, including their antiparticles. The gauge bosons mediate the fundamental forces of nature: the photon (γ) for electromagnetism, the gluon for the strong force and the W$^\pm$ and Z bosons for the weak force. The Higgs boson is the particle associated with the Higgs field and is responsible for giving mass to other particles in the SM. A list of all the SM fermions and bosons with their experimentally measured properties [14] is provided in Tab. 1.1 and 1.2.

The presence of gauge symmetries in the model implies that the Lagrangian of the theory can be can be expressed as a sum of multiple parts, with internal symmetries manifesting themselves in the gauge transformations of multiplicative sections. Therefore the

Generation	Fermions				
	Name	Symbol	Spin ($\overline{h}$)	Charge (e)	Mass (MeV)
I	Up	u	½	+2/3	$2.16^{+0.49}_{-0.26}$
	Down	d	½	-1/3	$4.67^{+0.48}_{-0.17}$
II	Charm	c	½	+2/3	$1.27^{+0.02}_{-0.02} \times 10^3$
	Strange	s	½	-1/3	$93.4^{+8.60}_{-3.40}$
III	Top	t	½	+2/3	$172.69^{+0.30}_{-0.30} \times 10^3$
	Bottom	b	½	-1/3	$4.18^{+0.03}_{-0.02} \times 10^3$
I	Electron	e	½	-1	$0.511 \pm (0.15 \times 10^{-9})$
	Electron neutrino	ν_e	½	0	$< 1.1 \times 10^{-6}$
II	Muon	μ	½	-1	$105.66 \pm (0.23 \times 10^{-5})$
	Muon neutrino	ν_μ	½	0	< 0.17
III	Tau	τ	½	-1	1776.86 ± 0.12
	Tau neutrino	ν_τ	½	0	< 18.2

Table 1.1: All the SM particles, quarks and leptons, are listed with their experimentally measured properties [14].

Name	Symbol	Spin	Charge	Mass (MeV)
Photon	γ	1	0	0
Gluon	g	1	0	0
W boson	$W^{\pm}$	1	± 1	80379 ± 12
Z boson	Z	1	0	91187 ± 21
Higgs boson	H	0	0	125250 ± 170

Table 1.2: A list of all SM bosons and their experimentally measured properties [14].

Lagrangian of the SM can be formulated as,

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Fermions}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}} \tag{1.1}$$

The theory of particle physics encompasses a wide range of phenomena, and various segments of the theory explain the interactions among these particles. Quantum Electrodynamics (QED) is a mathematical formulation that explains the interaction between light and matter, including the exchange of photons in the interaction between charged particles. Quantum Chromodynamics (QCD), on the other hand, explains the strong interaction between quarks and gluons. In the following sections, we provide a concise introduction to the mathematical formalisms used in these two theories.

1.2 Quantum Electrodynamics

Quantum electrodynamics is the relativistic quantum field theory of the electromagnetic interaction developed from the $U(1)$ gauge group. This was the first quantum field theory to reach a mature form, describing the interactions of electrons and photons. In classical mechanics, the dynamics of a physical system can be summarized with a function known as the Lagrangian (L). For a discrete system, it is written as the system's kinetic energy (T) minus the system's potential energy (V). In field theory, the Lagrangian in generalised coordinates can be replaced with the Lagrangian density ($\mathcal{L}$) integrate over volume. The independent variables are replaced by a point in spacetime (x, y, z, t) and the dependent variables are replaced by the value of a field at that point in spacetime φ (x, y, z, t). The Lagrangian in three-dimensional space can be expressed as,

$$L = \int \mathcal{L} \, d^3 x \tag{1.2}$$

26

The Lagrangian of quantum electrodynamics represents the interaction between matter, photons, and their mutual interactions. It can be formulated as:

$$\mathcal{L}_{\text{QED}} = \mathcal{L}_{\text{Dirac}} + \mathcal{L}_{\text{Maxwell}} + \mathcal{L}_{\text{interaction}} \tag{1.3}$$

The first term, $\mathcal{L}_{\text{Dirac}}$, corresponds to the Dirac Lagrangian, which describes the motion of a spin-½ particle in the absence of external fields. The second term, $\mathcal{L}_{\text{Maxwell}}$, governs the propagation of the electromagnetic field. The third term, $\mathcal{L}_{\text{interaction}}$, accounts for the interaction between particles and the electromagnetic field.

An significant property of the Dirac Lagrangian is that it is invariant under a $U(1)$ global phase transformation of the form

$$\psi(x) \rightarrow \psi'(x) = e^{i\alpha}\psi(x), \tag{1.4}$$

where, α is a real constant and ψ is the Dirac spinor [22]. On the other hand, it's important to note that the Dirac Lagrangian does not possess invariance under the local U(1) phase transformation of the form

$$\psi(x) \rightarrow \psi'(x) = e^{i\alpha(x)}\psi(x) \tag{1.5}$$

where, $\alpha(x)$ is the space-time field.

The Lagrangian density for free Dirac spinor fields corresponding to spin-½ particles is

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi, \tag{1.6}$$

where $\bar{\psi}$ is the Dirac adjoint of the spinor field, γ^{μ} are the Dirac gamma matrices, and m is the mass of the particle. After applying the local U(1) phase transformation in Eqn. 1.6, the Lagrangian becomes,

$$\mathcal{L}'_{\text{Dirac}} = \mathcal{L}_{\text{Dirac}} - \bar{\psi}\gamma^{\mu}(\partial_{\mu}\alpha(x))\psi \tag{1.7}$$

Unlike the global gauge transformation, the derivative in kinetic term of $\mathcal{L}_{\text{Dirac}}$ is not invariant, $\partial\psi(x) \neq \partial\psi'(x)$, under local gauge transformation. This motivates to introduce a *covariant derivative* (D_μ) with a gauge field in addition of ∂_μ to satisfy the invariance of the Lagrangian, which is of the form

$$D_\mu = \partial_\mu + ieA_\mu(x) \tag{1.8}$$

Where, e is the charge of the particle and A_μ is a vector field that transforms under $U(1)$ transformation as,

$$A_\mu \rightarrow A' = A_\mu + \frac{1}{e}\partial_\mu\alpha(x). \tag{1.9}$$

As a result of incorporating D_μ, the Lagrangian becomes invariant under local gauge transformation. The transformation creates a coupling term of the vector field with the Dirac particle of charge $-e$. This vector field is known as the photon field of QED. In order to represent a new real field, a kinetic energy and the mass term must be added in the Lagrangian and the resulting Lagrangian becomes,

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}m_A^2 A_\mu A^\mu \tag{1.10}$$

Where, $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the electromagnetic field tensor and m_A is the mass of the associated particle of the photon field. When checking for local gauge invariance, it is evident that the first and second term of Eqn. 1.10 exhibit invariance, whereas the third term does not. In order to ensure that the Lagrangian is invariant under U(1) local transformation, the condition $m_A = 0$ must be met. This explains the observed massless nature of photons. The final QED Lagrangian is therefore,

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + e\bar{\psi}\gamma^\mu\psi A_\mu \tag{1.11}$$

The tree level propagator and the interaction terms from the QED Lagrangian are shown in Fig. 1.1
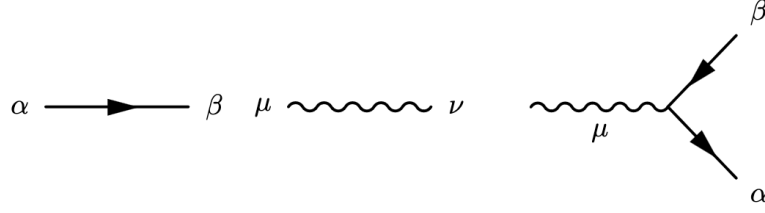
Figure 1.1: The tree level propagator and the interaction vertices of the QED Lagrangian.

1.3 Quantum Chromodynamics

Quantum chromodynamics is a relativistic quantum field theory that deals with the strong interaction between quarks and gluons. It is built upon the SU(3) gauge group and is composed of almost the same components as the QED Lagrangian in Equation 1.11. The only difference is the more complicated $SU(3)$ symmetries in place of $U(1)$, because of the three color charges in place of single charge of QED. Similarly to Eqn. 1.6, the Lagrangian of the quark in absence of any field is written as,

$$\mathcal{L}_{\text{free}} = \bar{q}_n(i\gamma^\mu\partial_\mu - m)q_n \tag{1.12}$$

Where, q_n corresponds to the quark spinor field. The Lagrangian is required to be invariant under the following transformation in order to preserve the $SU(3)$ symmetries.

$$q(x) \to q'(x) = e^{i\alpha_k(x)T_k}q(x) \tag{1.13}$$

Where, T_k refers to a set of linearly independent, traceless, 3×3 matrices known as Gell-Mann matrices and $\alpha_k(x)$ are the transformation parameters.

$$T_a = \lambda_a/2 \tag{1.14}$$

where, λ_a are Gell-Mann matrices and T_a are generator of SU(3). They follow the commutation relation

$$[T_a, T_b] = 2if_{abc}T_c \tag{1.15}$$

29

Where, f_{abc} is the structure constant of the group. Introducing the local SU(3) gauge symmetry breaks the Lagrangian's invariance. Therefore, to preserve the invariance, a covariant derivative of the following form is introduced:

$$D_\mu = \partial_\mu + ig T_k G^k_\mu \tag{1.16}$$

where, G^k_μ is the gluon field and g is the coupling strength to the gluon field. Under the local gauge transformation, G^k_μ transforms as

$$G^k_\mu \to G'^k_\mu = G^k_\mu - \partial_\mu \alpha_k - g f_{ijk} \alpha_i G^j_\mu \tag{1.17}$$

The gauge invariant field strength tensor $G^{\mu\nu}_i$ is analogous to the electromegnetic field strength tensor, $F_{\mu\nu}$, in Equation **??**. $G^{\mu\nu}_i$ is defined as:

$$G^{\mu\nu}_i = \partial^\mu G^\nu_i - \partial^\nu G^\mu_i - g f_{ijk} G^\mu_j G^\nu_k \tag{1.18}$$

Indices i, j and k in Eqn. 1.18 run from 1 to 8. The third term in Equation 1.18 ensures the gauge invariance and implies the self-interaction of gluons which is also observed in experiments. Adding the kinetic terms of the field strength tensor, the Lagrangian is not invariant under the local gauge transformation. This requires the gluon mass to be zero, which agrees well with the experimental observation. The gauge invariant QCD Lagrangian density takes the form:

$$\mathcal{L}_{\text{QCD}} = \bar{q}(i\gamma^\mu D_\mu - m)q - \frac{1}{4} G^i_{\mu\nu} G^{\mu\nu}_i \tag{1.19}$$

$$\mathcal{L}_{\text{QCD}} = \bar{q}(i\gamma^\mu \partial_\mu - m)q - g(\bar{q}\gamma^\mu T_i q)G^i_\mu - \frac{1}{4} G^i_{\mu\nu} G^{\mu\nu}_i \tag{1.20}$$

The tree level propagator and the interaction terms from the QCD Lagrangian is shown in Fig. 1.2.

1.4 Weak Interaction

The formalism of weak interaction is constructed based on the non-abelian $\text{SU}(2)_L$ gauge group, where L denotes that the associated vector fields couple only with the left-handed
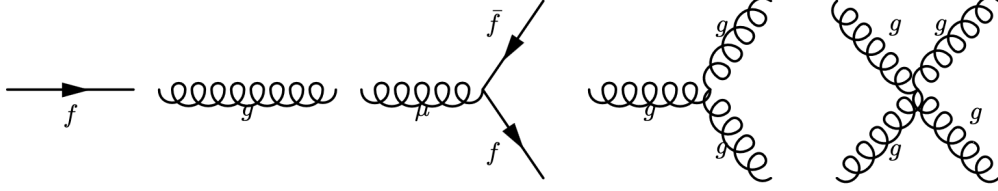
Figure 1.2: The tree level propagator and the interaction vertices of the QCD Lagrangian.

chiral fermions and right-handed chiral anti-fermions.

The vector minus axial vector coupling structure of the weak theory is introduced by the right-handed and left-handed spinors:

$$\psi_R = P_R\psi = \frac{1}{2}(1+\gamma^5)\psi \qquad \psi_L = P_L\psi = \frac{1}{2}(1+\gamma^5)\psi \qquad (1.21)$$

where $P_{L,R}$ are the left and right handed parity operators, respectively, and γ^5 is the product of the four Dirac matrices, $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, which anticommutes with γ^μ and means that objects such as $\bar{\psi}\gamma^5\psi$ are odd under parity. Hence, in the SM, the fermions appear as families with left-handed doublets of quarks Q_L^i and leptons L_L^i and right-handed singlets of quarks $u_R^i(d_R^i)$ and leptons $e_R^i(\nu_R^i)$:

$$
\begin{aligned}
Q_L^i &= \begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, & L_L^i &= \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \\
q_R^i &= u_R, c_R, t_R, & q_R^i &= d_R, s_R, b_R; \\
\ell_R^i &= e_R, \mu_R, \tau_R, & \nu_R^i &= \nu_R^e, \nu_R^\nu, \nu_R^\tau
\end{aligned}
\qquad (1.22)
$$

Following a similar procedure as described in Sec. 1.2 and 1.3, the local gauge invariant Lagrangian for the weak interaction can be written as

$$\mathcal{L}_{\text{weak}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - g_W(\bar{\psi}\tau_i\psi)W_\mu^i - \frac{1}{4}W_{\mu\nu}^i W_i^{\mu\nu} \qquad (1.23)$$

where g_W is a constant, τ_i are 2×2 linearly independent, traceless matrices, and $W_i^{\mu\nu}$ is defined as

$$W_i^{\mu\nu} = \partial^\mu W_i^\nu - \partial^\nu W_i^\mu - g\epsilon_{ijk}W_j^\mu W_k^\nu \qquad (1.24)$$

As before, to obtain the local gauge invariance of the Lagrangian, three gauge fields W_μ^1, W_μ^2 and W_μ^3 had to be introduced with the covariant derivative. However, the inclusion of mass terms corresponding to the field again breaks the invariance of the Lagrangian. The experimental observations of massive bosons associated with weak interaction imply that the mass of the associated particles cannot be zero in this scenario. The resolution to this issue is discussed in the following section, which involves spontaneous symmetry breaking, a mechanism by which massless particles can acquire mass.

1.4.1 Spontaneous Symmetry Breaking and the Higgs Mechanism

The perplexity surrounding the masses of weak bosons can be resolved through a mechanism known as spontaneous symmetry breaking, which enables massless particles to acquire mass. In the simplest idealized relativistic model, the spontaneously broken symmetry can be summarized through an illustrative scalar field theory. To demonstrate this, let us consider a real scalar field ϕ, with a Lagrangian given by:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V \tag{1.25}$$

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{1}{24}\lambda\phi^4 \tag{1.26}$$

$$V(\phi) = \frac{1}{2}m^2\phi^2 + \frac{1}{24}\lambda\phi^4 \tag{1.27}$$

Where, m^2 and λ are the free parameters. Calculating the minimum of the potential:

$$\text{at the minimum, } V'(\phi_0) = 0 \Rightarrow \phi_0 = 0 \tag{1.28}$$

$$\phi_0 = \pm\sqrt{\frac{-6m^2}{\lambda}} \tag{1.29}$$

The coupling parameter λ must be positive to keep the potential bounded from below. On the other hand, the m^2 parameter may have either sign, so it is essential to compare the theory in the $m^2 > 0$ and $m^2 < 0$ regimes. Hence if $m^2 > 0$, the theory has a unique ground state at $\phi = 0$ and is invariant under the phase symmetry. The shape of the potential

is shown in the Fig. 1.3. Now, when $m^2 < 0$, the minimum of the potential as shown in
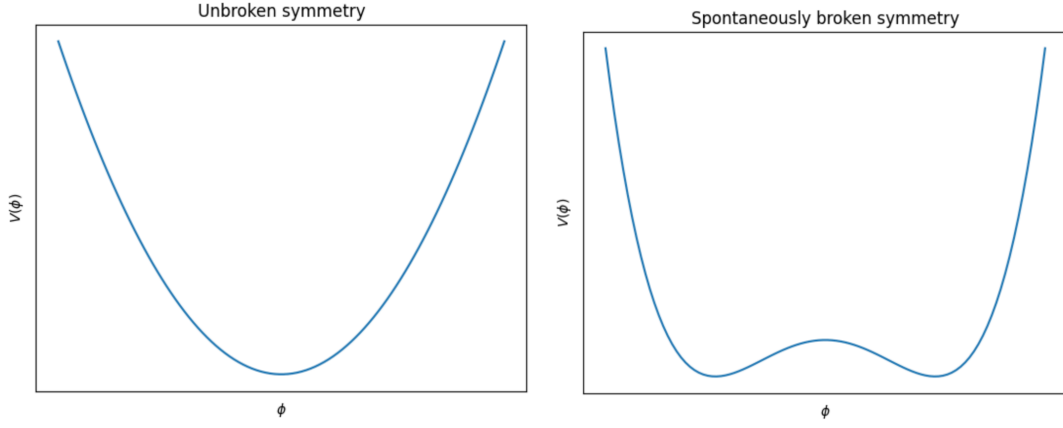
Figure 1.3: The potential $V(\phi)$ of the scalar field ϕ, as defined in equation 1.25, for the two cases: $m^2 > 0$ (left) and $m^2 < 0$ (right).

Fig. 1.3 is no longer at one point. It has two possible vacuum states at $v = \pm\sqrt{\frac{-6m^2}{\lambda}}$, known as the "vacuum expectation value" making none of the minima invariant under the phase symmetry. This phenomenon is known as the spontaneous symmetry breaking. Expanding the classical theory around any of the minima, choosing $\phi = +v$ and a shifting by $\rho(x)$.

$$\phi(x) = v + \rho(x) \tag{1.30}$$

The Lagrangian of Equation 1.25 becomes,

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2}\partial_\mu\rho\partial^\mu\rho - \frac{1}{2}m^2(v+\rho)^2 - \frac{1}{24}\lambda(v+\rho)^4 \\
&= \frac{1}{2}\partial_\mu\rho\partial^\mu\rho - (\frac{1}{2}m^2v^2 + m^2v\rho + \frac{1}{2}m^2\rho^2 + \frac{1}{24}\lambda v^4 + \frac{1}{6}\lambda v^3\rho + \frac{1}{4}\lambda v^2\rho^2 + \frac{1}{6}\lambda v\rho^3 + \frac{1}{24}\lambda\rho^4) \\
&= \frac{1}{2}\partial_\mu\rho\partial^\mu\rho + m^2\rho^2 - \frac{1}{6}\lambda v\rho^3 - \frac{1}{24}\lambda\rho^4 + \text{Const.}
\end{aligned}
$$
$$\tag{1.31}$$

The second term accounts for the mass of the boson associated with this field and the third and fourth terms, its self interaction. Thus a massive scalar field has been generated by the process of spontaneous symmetry breaking.

Now repeating this procedure with a complex scalar field of the form

$$\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2). \tag{1.32}$$

In the interacting complex scalar field theory, the Lagrangian is written as,

$$\mathcal{L} = \partial_\mu \phi \partial^\mu \bar{\phi} - V \tag{1.33}$$

$$\mathcal{L} = \partial_\mu \phi \partial^\mu \bar{\phi} - m^2 \bar{\phi}\phi - \frac{1}{4}\lambda(\phi\bar{\phi})^2 \tag{1.34}$$

$$V(\phi\bar{\phi}) = m^2 \phi\bar{\phi} + \frac{1}{4}\lambda(\phi\bar{\phi})^2 \tag{1.35}$$

The fields ϕ and $\bar{\phi}$ are complex conjugate. $U(1)$ is still a symmetry of the Lagrangian given by $\phi = re^{i\theta}$ and $\phi(x,t) = r(x,t)e^{i\theta(t,x)}$. From the same procedure as above, vacua are obtained by finding the minima of the potential V(r),

$$V(r) = m^2 r^2 + \frac{1}{4}\lambda r^4$$
$$V'(r) = 2m^2 r + \lambda r^3 \tag{1.36}$$

At the minimum, $V'(r) = 0 \Rightarrow r = 0$

$$r = \pm\sqrt{\frac{-2m^2}{\lambda}} \tag{1.37}$$

If $m^2 > 0$, the theory has a unique ground state with a circle of radius r = 0, forming a paraboloid. In the case of $m^2 < 0$, the minima is a radius of circle r, forming a "Mexican hat shape" with a whole circle of an infinite number of vacua. For both cases, the shape of the potential is displayed in Fig. 1.4.

To expand the classical theory around any of the minima, choosing $\phi = r = \frac{-2m^2}{\lambda}$, a shift $\eta(x) + i\xi(x)$ is introduced below.

$$\phi(x) = r(x) + \frac{1}{\sqrt{2}}(\eta(x) + i\xi(x))$$
$$= \frac{1}{\sqrt{2}}(v + \eta(x) + i\xi(x)) \tag{1.38}$$
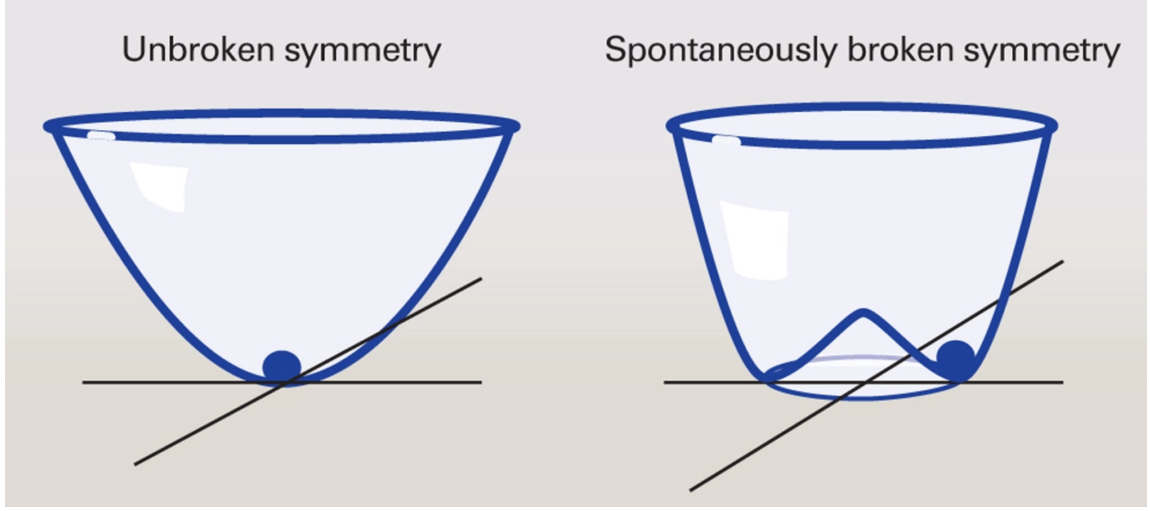
Figure 1.4: A schematic representation of the potential $V(\phi_1, \phi_2)$ of the complex scalar field ϕ, as defined in equation 1.35, is depicted for two cases: $m^2 > 0$ (left) and $m^2 < 0$ (right). The figure is taken from Ref. [1]

Substituting this into the Lagrangian of Equation 1.33 yields

$$\mathcal{L} = \frac{1}{2}\partial_\mu \eta \partial^\mu \eta + \frac{1}{2}\partial_\mu \xi \partial^\mu \xi + m^2\left(\frac{(v+\eta)^2}{2} + \frac{\xi^2}{2}\right) - \frac{\lambda}{4}\left(\frac{(v+\eta)^2}{2} + \frac{\xi^2}{2}\right) \qquad (1.39)$$

$$= \frac{1}{2}\partial_\mu \eta \partial^\mu \eta + \frac{1}{2}\partial_\mu \xi \partial^\mu \xi - m^2 \eta^2 - V_{int}(\eta, \xi) \qquad (1.40)$$

Where, $V_{int}(\eta, \xi)$ contains all the interaction terms. This spontaneously broken quantum theory is a theory of two real scalar fields and two scalar particles associated with these two fields. From Equation 1.39, the particle corresponding to the field η is massive and the particle associated with the field ξ is massless. However, the resulting Lagrangian of the spontaneously broken quantum theory is not invariant under the local gauge transformation. As earlier, a *covariant derivative*, D_μ, is introduced in place of ∂_μ to retain the invariance, defined as

$$D_\mu = \partial_\mu - ieA_\mu. \qquad (1.41)$$

Where the gauge field A_μ transforms as Equation 1.9. Introducing the covariant derivative in Equation 1.39 and inserting the kinetic term corresponding to the field A_μ, the Lagrangian

takes the form

$$\mathcal{L} = \frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \lambda v^2\eta^2 + \frac{1}{2}\partial_\mu\xi\partial^\mu\xi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu A^\mu - V_{int}(\eta,\xi) - evA_\mu\partial^\mu\xi.$$
(1.42)

Therefore, by adding a scalar field with spontaneously broken symmetry, the boson corresponding to the gauge field can have mass without breaking the local gauge symmetry. In doing so, a massive η boson and a massless ξ boson, known as a *Goldstone boson,* have been created. The last term appeared in the above equation describing the interaction between the gauge field and the Goldstone boson is unphysical and can be removed by making the gauge transformation of the form,

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{ev}\partial_\mu\xi$$
(1.43)

The final Lagrangian become,

$$\mathcal{L} = \frac{1}{2}\partial_\mu\eta\partial^\mu\eta - \lambda v^2\eta^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}e^2v^2A_\mu A^\mu - e^2vA_\mu A^\mu\eta$$
$$+ \frac{1}{2}e^2A_\mu A^\mu\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4.$$
(1.44)

The above equation is independent of the field $\xi(x)$, equivalent to taking $\phi = \frac{1}{\sqrt{2}}(v + \eta(x))$ in Equation 1.38. Thus, introducing a complex scalar field with the spontaneously broken symmetry that couples with the locally gauge invariant field A_μ, the mass of the gauge boson is created along with a massive boson for the field, $\eta(x)$. The mass of the gauge boson is ev and that of the scalar boson, η, is $v\sqrt{2\lambda}$. This is known as the Higgs mechanism. The scalar field, η, is called the Higgs field and the corresponding quantum of the field is called the Higgs boson.

1.4.2 Electroweak Theory and Symmetry Breaking

The electroweak theory is a very successful formulation to unify the electromagnetic and weak interactions, constructed based on the product of $SU(2)$ and $U(1)$ gauge groups. Starting from the $SU(2)$ group, three fields, $W_\mu^1, W_\mu^2, W_\mu^3$, are defined in the covariant derivative

36

to make the local gauge invariance of the weak interaction Lagrangian in Equation 1.23 and the corresponding charge is known as weak isospin (I_W). In $U(1)$ symmetry of QED, replacing the electric charge Q with the weak hypercharge Y defines a corresponding field B_μ, which is analogous to the A_μ field of QED. The prime insight of electroweak (EW) unification is that the three bosons that appear from the $SU(2)$ group can be combined with a $U(1)$ boson to form the four physically observed bosons. The relation between electric charge, weak hypercharge, and the third component of weak isospin is defined as

$$Q = I_3 + \frac{1}{2}Y \tag{1.45}$$

In the interacting complex scalar field theory, including the symmetry-breaking potential $V(\phi\bar{\phi}) = m^2\phi\bar{\phi} + \lambda(\phi\bar{\phi})^2$, the Lagrangian is written as

$$\mathcal{L} = \partial_\mu\phi\partial^\mu\bar{\phi} - m^2\bar{\phi}\phi - \lambda(\phi\bar{\phi})^2 \tag{1.46}$$

The ϕ is expanded near the vacuum expectation value, v, with a real scalar field, say $H(x)$, $\phi = v + H(x)$. Introducing the shift and to attain the $U(1)_Y \times SU(2)_L$ symmetry of the Lagrangian, a covariant derivative is introduced of the form

$$D_\mu = \partial_\mu + ig_w T_i W_\mu^i + ig_1 \frac{Y}{2} B_\mu \tag{1.47}$$

Where, T_k are the generators of $SU(2)$, where g_w is the coupling strength and Y is the generator of $U(1)$ with coupling strength g_1. Replacing the D_μ in Equation 1.46, the kinetic term for the Higgs field becomes:

$$D_\mu\phi D^\mu\bar{\phi} = \frac{1}{2}(\partial_\mu H)^2 + \frac{1}{8}g_W^2(v + H)^2|W_\mu^1 + iW_\mu^2|^2 + \frac{1}{8}(v + H)^2|g_W W_\mu^3 - g_1 B_\mu|^2 \tag{1.48}$$

From the above equation, the mass terms are the mixture of B_μ and W_μ^i. Therefore, four new fields are introduced as combination of the B_μ and W_μ^i : two complex and two real

Two complex fields are: $W^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$, such that, $W_\mu^+ = (W_\mu^-)^* \tag{1.49}$

Also, two real fields are:
$$Z_\mu = \frac{1}{\sqrt{g_1^2 + g_W^2}}(g_W W_\mu^3 - g_1 B_\mu);$$
$$A_\mu = \frac{1}{\sqrt{g_1^2 + g_W^2}}(g_1 W_\mu^3 + g_W B_\mu) \tag{1.50}$$

After expansion of Equation 1.48, the fields $W_\mu^\pm$ and Z_μ give rise to the massive bosons, known as the $W^\pm$ and Z bosons, respectively, and the field A_μ is the massless photon field of QED. The masses of the particles are,

$$m_W^\pm = \frac{1}{2}v g_W,$$
$$m_Z = \frac{1}{2}v\sqrt{g_1^2 + g_W^2}, \tag{1.51}$$
$$m_A = 0$$

The mass of the gauge bosons are known once the v and g_w are known. However for the case of Higgs boson, the mass of the Higgs boson ($v\sqrt{2\lambda}$) depends on the free parameter λ, where λ represents the Higgs self coupling.

The combination of the W^3 and B fields produces the Z and γ bosons,

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B \\ W^3 \end{pmatrix} \tag{1.52}$$

Where θ_W is the weak mixing angle which also gives the relation between the masses of the W and Z bosons as

$$\cos\theta_W = \frac{m_W}{m_Z}, \qquad \sin\theta_W = \frac{g_1}{\sqrt{g_W^2 + g_1^2}} \tag{1.53}$$

Therefore, the electroweak spontaneous symmetry breaking (EWSSB) of the SM is described by the four fundamental parameters: the Higgs potential parameters m^2 and λ and the coupling strength parameters g_W and g_1. In summary, the Higgs mechanism applied to the $SU(2_L)$ symmetry generates the masses of the W and Z bosons, as well as agrees with zero mass photons. This mechanism also generates the existence of a scalar boson, known as the Higgs boson (H), which was discovered by CMS and ATLAS experiments.

1.5   Yukawa Coupling and Fermion Masses

The SM treats left-handed and right-handed fermions differently, as shown in Equation 1.22. Since the $\mathrm{SU}(2)_L$ group acts only on the left components of the fermion fields, this poses a fundamental problem for the masses of these particles. The mass term for the Dirac fermions couples to both left-handed and right-handed components and is written as

$$m\psi\bar{\psi} = -m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R). \tag{1.54}$$

This Lagrangian is not invariant under the local phase transformation. Nevertheless, in the theory with spontaneous symmetry breaking, there is a way of giving mass to fermions via a Yukawa coupling to a scalar field. The SM uses the Yukawa interaction to describe the coupling between the Higgs field and massless quark and lepton fields through coupling constants g.

Considering the case of electron, the coupling between the electron doublet, $L_L^e = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$, the Higgs boson doublet ϕ, and the right-handed component of the electron field e_R, the effective Yukawa Lagrangian is:

$$\mathcal{L}_{Yukawa}^e = -g_e(\bar{L}_L^e\varphi e_R + \bar{e}_R\varphi^\dagger L_L^e) \tag{1.55}$$

This Lagrangian is invariant under $SU(2)_L$ symmetry. Now, applying the spontaneous symmetry breaking and expanding the field near the ground state in similar way as in Equation 1.38, the above Lagrangian becomes:

$$\mathcal{L}_{Yukawa}^e = -\frac{g_e v}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{g_e}{\sqrt{2}}(\bar{e}_L e_R + \bar{e}_R e_L)H = -m_e(\bar{e}e) - g_e(e\bar{e})H. \tag{1.56}$$

The first term in Equation 1.56 defines the electron mass term ($m_e = g_e v/\sqrt{2}$), while the second term describes the interaction between the Higgs field and the electron-positron pair. The same mechanism is considered for other charged leptons, although the mass term for neutrinos is currently missing from the SM.

By following a similar approach, the Higgs mechanism can also generate masses for the fermions within the SM through a renormalizable Yukawa interaction between the Higgs field and for all the fermions. A direct consequence of the SSB mechanism is that the strength of the coupling g_f of the Higgs field to a massive fermion f is proportional to its mass $g_f = \sqrt{2}m_f/v$.

1.6 SM Higgs boson production and decays

The SM Higgs boson coupling to the all gauge bosons V, fermions f, as well as its self-coupling are explained in Sec. 1.4 and 1.5. In a hadronic collider experiment, such as the LHC, the production of Higgs bosons involves interactions between quarks or gluons. The SM Higgs boson can be produced at hadron colliders via five primary mechanisms:

- **Gluon-gluon fusion (ggH)**: The Feynman diagram for this process is shown in Fig. 1.5. This is the dominant Higgs production process in the LHC. It involves the fusion of two gluons, each of which carries a fraction of the proton's momentum, to form a Higgs boson. Due to the fact that gluons are massless, there is no direct coupling between the Higgs boson and gluons at the tree-level. As a result, the interaction between gluons and Higgs bosons occurs through a fermionic loop, with the top quark loop contributing the most significantly.

- **Vector boson fusion (VBF)**: At the tree level, this process occurs when two incoming quarks each radiate a Z or W boson and they "fuse" through a t-channel exchange to produce a Higgs boson (with proper charge conservation). This production mode can be used to probe the couplings of Higgs bosons to vector bosons. This process has a relatively low cross section, but can be identified by the presence of two forward jets in the detector. The Feynman diagram for this production process is shown in Fig. 1.6
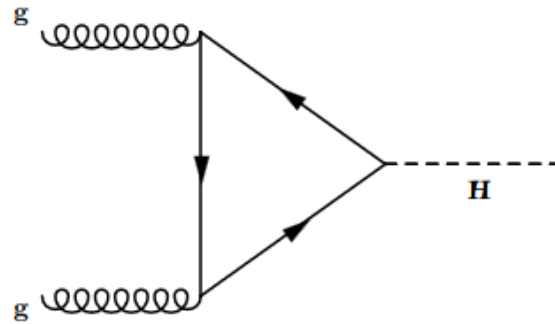
Figure 1.5: The Feynman diagram shown here represent Higgs production via gluon-gluon fusion in a hadron collider.
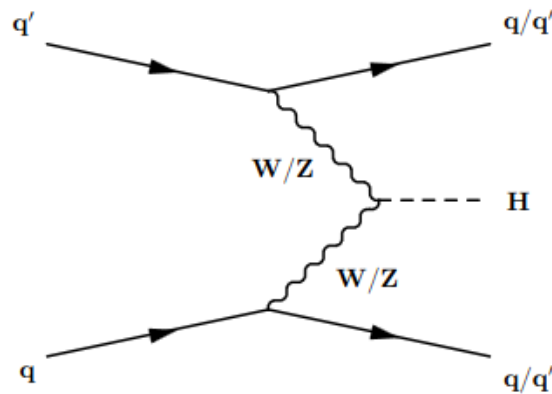
Figure 1.6: The Feynman diagram shown here represents the vector boson fusion process for Higgs production in a hadron collider.

- **Associated Higgs production with vector bosons/Higgsstrahlung (VH)**: In this production mode, the Higgs boson is generally radiated from a vector boson via its coupling to the Higgs field. This process can occur through three different Feynman diagrams, as shown in Fig. 1.7.
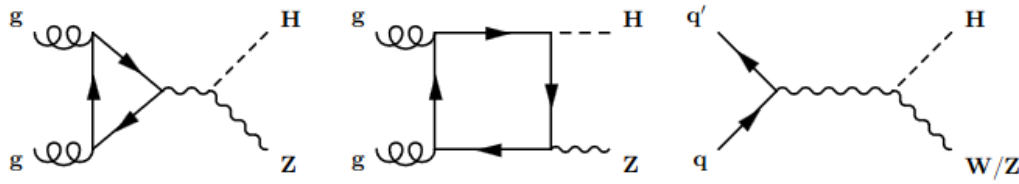
Figure 1.7: The associated production of the Higgs boson with vector bosons in a hadron collider.

- **Associated Higgs production with two top quarks ($t\bar{t}$H)**: Higgs bosons are also produced in association with two top quarks, as shown in Fig. 1.8. This process allows studying the Yukawa coupling of the top quark with the Higgs boson. This process has a relatively low cross section but can be identified by the presence of additional leptons or jets in the detector.
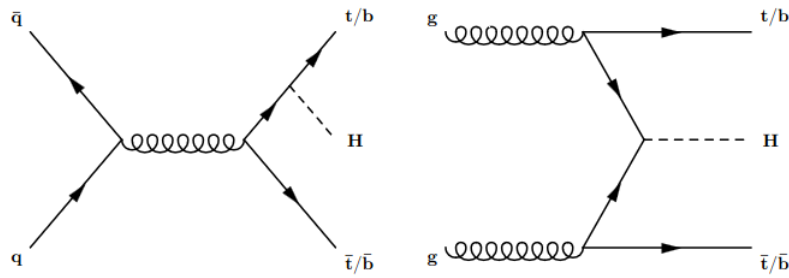
Figure 1.8: Higgs production associated with the two top quarks in hadron collider.

- **Associated Higgs production with single top quarks (tHq/tHW):** This is the rarest production mode considered, shown in Fig. 1.9. This process also allows us to measure the top Yukawa coupling and is also sensitive to the sign of the Yukawa coupling.

In this thesis, we focus on studying this Higgs production mechanism and measuring its production cross section.
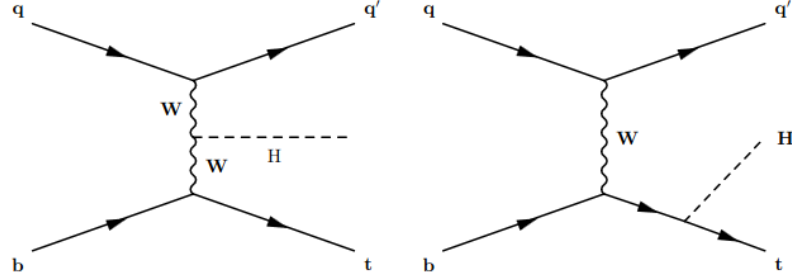
Figure 1.9: The associated production of the Higgs boson with one top quark in a hadron collider, where the Higgs boson is emitted from the heavier legs of top quark and the W boson line.

Figure 1.10a and 1.10b display the theoretical cross sections for these Higgs production processes as a function of the Higgs boson mass (M_H) at a fixed center-of-mass (CoM) energy of $\sqrt{s} = 13$ TeV and as a function of the CoM energy, respectively. As the CoM energy increases, the cross section for Higgs production processes also increases, as shown in Fig. 1.10b. The studies carried out in this thesis are based on a p–p collision CoM energy of 13 TeV. From July 2022, the LHC has increased the CoM energy to 13.6 TeV.

As discussed in Sec. 1.4 and 1.5, all the SM particles gain masses from the spontaneous breaking of the gauge symmetry, through gauge couplings to the Higgs field in the case of vector bosons, and Yukawa couplings in the case of fermions. The SM Higgs boson couples to vector bosons, with an amplitude proportional to the gauge boson mass squared, m_V^2, and to fermions with an amplitude proportional to the fermion mass, m_f. Hence, the Higgs boson can decay to all massive SM particles at tree level; the coupling is stronger for the third generation of quarks and leptons than for those in the 2nd generation. Although the photon is massless, the Higgs boson can decay two photon via a virtual loop of fermions or charged bosons as shown in Fig. 1.11.
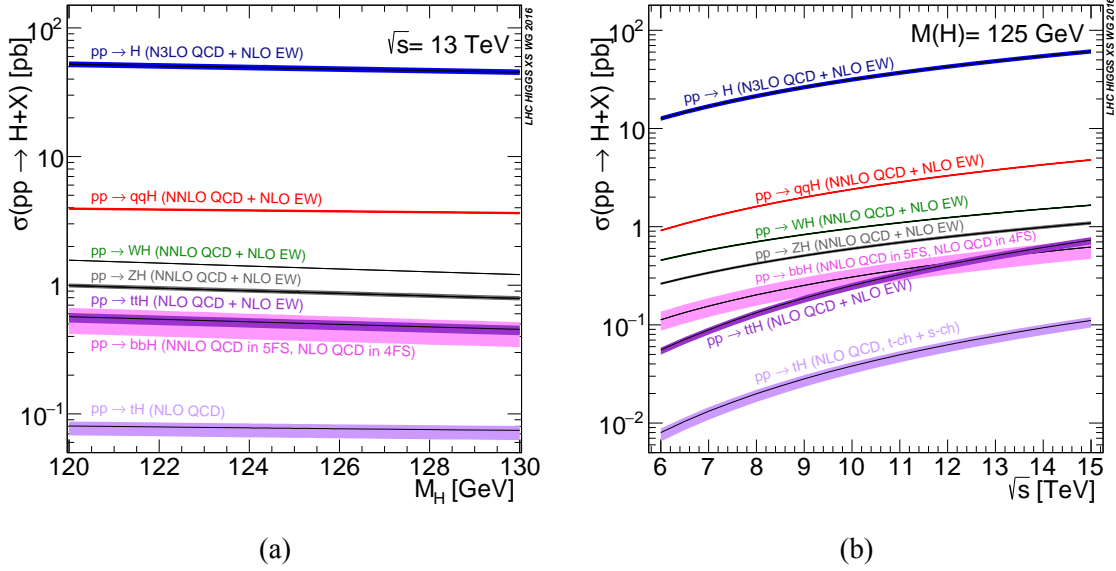
Figure 1.10: Theoretical prediction of the cross section for different Higgs boson production modes with respect to the Higgs boson mass (left) and center-of-mass energy of the collision (right). The plots are taken from the Ref. [2].
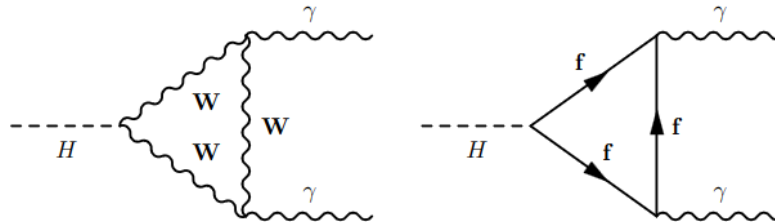
Figure 1.11: The Higgs boson decays to two photon via a virtual loop of top quark or W boson.

The decay modes of the Higgs boson [23], which are prominently probed in experiment, are

$$H \longrightarrow b\bar{b}$$

$$H \longrightarrow WW \longrightarrow l\nu l\nu$$

$$H \longrightarrow \tau\bar{\tau}$$

$$H \longrightarrow \gamma\gamma$$

$$H \longrightarrow ZZ \longrightarrow 4l$$

The Higgs boson, with observed mass of 125.38 ± 0.14GeV [24], is not a stable particle.
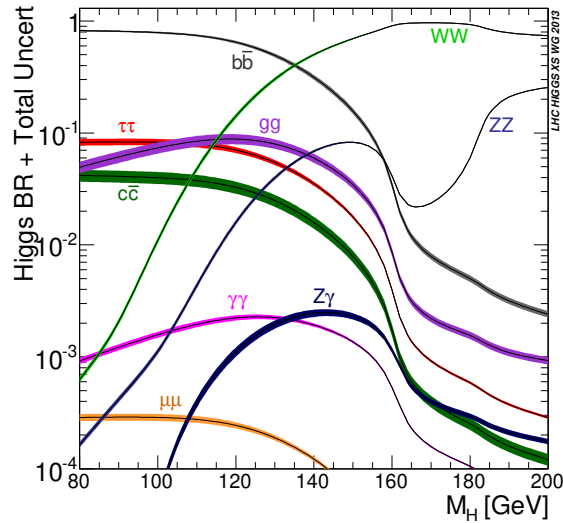
Figure 1.12: Theoretical prediction of the branching fraction for different Higgs boson decay modes with respect to the Higgs boson mass. The plot is taken from Ref. [3].

With a lifetime of 1.6×10^{-22}s, it decays immediately after it produced. It has a tendency to decay into the heaviest kinematically accessible particles. The most massive fermions have the largest couplings with the Higgs boson. The primary decay mode is into $b\bar{b}$ pairs with a branching fraction ($\mathcal{B}$) of $57.5 \pm 1.9\%$ [14]. However, at the LHC, many energetic

QCD events can mimic this signature, making it challenging to detect. The $H \rightarrow W^+W^-$ mode is the second largest decay, with a $\mathcal{B}$ of 21.4%. The $H \rightarrow \tau\bar{\tau}$ is the next largest decay channel into fermions. Due to the presence of neutrinos (from leptonic decays of the τ-lepton and W bosons) and also constraints in the reconstruction of the particles in the detector (from hadronic decays), it is challenging experimentally to probe all the decays. Each decay mode of the Higgs has its advantages and disadvantages experimentally. The $\mathcal{B}$ of the decay of the Higgs boson to two Z bosons is 2.6%; The subsequent decay of the two Z bosons into four leptons (electrons or muons) has a very clean signal in the detector, but once the $\mathcal{B}$ of the Z boson decay to leptons is considered, this is a very rare decay. Similarly, the Higgs boson decay to two photons also has very clean signature in the detector, but this is also a rare decay with a $\mathcal{B}$ of 0.227%. The studies presented in this thesis specifically focus on the Higgs boson decay channel into two photons.

Review in Higgs studies:

The Higgs boson discovery in 2012 combined the experimental signature of the $H \rightarrow ZZ$ and $H \rightarrow \gamma\gamma$ decay modes to get an observed (obs.) statistical significance of 5 standard deviations (std. dev.) with an expected (exp.) significance of 5.8 std. dev. Individually, the most sensitive channels, $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4\ell$, yielded 4.1 std. dev. obs. (2.8 std. dev. exp.), and 3.2 std. dev. obs. (3.8 std. dev. exp.) respectively.

Using all data collected during Run 1 of the LHC (2010–2012) by the CMS detector (discussed in detail in next section), it was possible to observe separately the bosonic decay channels with significances of 6.5 std. dev. for $H \rightarrow ZZ \rightarrow 4\ell$, 5.6 std. dev. for $H \rightarrow \gamma\gamma$, 4.7 std. dev. for $H \rightarrow WW$, and 3.8 std. dev. for the fermionic decay channel $H \rightarrow \tau\tau$. With the increase of data over years, the properties of the Higgs boson were also studied. The mass was measured to be 125.02 GeV with a precession of 0.2% [25]. Using the angular distributions of the leptons in the bosonic decay channels, the spin (J) and parity (P, a parity transformation that effectively turns a phenomenon into its mirror image) were

also found to be compatible with the SM prediction ($J^P = 0^+$) with a large number of alternative spin-parity hypotheses ruled out at >99.9% confidence level (CL). At this point the individual production modes VBF, VH, and ttH were observed at a level of 3 std. dev, respectively.

Using data from Run 2 of the LHC (2015–2018), it become possible to observe $H \to \tau\tau$, $H \to b\bar{b}$, and $H \to \mu\bar{\mu}$ decays with significance of 5.9, 5.6 and 3 std. dev. Furthermore, the production mode ttH was observed with a significance of 5.2 std. div. One chapter of this thesis presents a measurement of the production cross section for the Higgs boson associated with single top quark using Run 2 data corresponding to an integrated luminosity of 137 fb^{-1}. In certain instances, the integrated luminosity is reported as 138 fb^{-1}; luminosity across various data versions is a result of improvements, which can be attributed to the enhancements in calibration and data processing.

1.7 Probing beyond the Standard Model

The SM of particle physics is a highly successful theory that describes the behavior of fundamental particles and their interactions. However, there are still open questions that the SM cannot explain, such as dark matter [26], neutrino oscillations [27], matter-antimatter asymmetry [28], the Higgs mass hierarchy problem, and the strong CP problem [29]. These reasons imply that the SM is not complete and that there is physics beyond it. Probing beyond the SM involves looking for evidence of new particles, interactions, and phenomena that are not accounted for in the SM.

There are several approaches to probe BSM, including:

- Particle colliders: High-energy particle colliders such as the LHC at CERN can create particles with energies much higher than those accessible in previous experiments. By analyzing the collision products and looking for deviations from the SM predic-

tions, scientists can search for evidence of new physics. Examples of beyond the SM physics that could be probed at the LHC include supersymmetric particles, extra dimensions, and dark matter candidates.

- Precision measurements: The SM makes very precise predictions for the behaviour of particles, and any deviations from these predictions could indicate the presence of new physics. Precision measurements of quantities such as the magnetic moment of the muon, the electric dipole moment of the electron, and the masses of the W and Z bosons can provide constraints on the parameter space of BSM theories.

- Astrophysical observations: Astrophysical phenomena such as dark matter, neutrino oscillations, and cosmic rays can provide clues about new physics beyond the SM. For example, the detection of high-energy neutrinos from distant sources could indicate the presence of new neutrino interactions or the decay of exotic particles.

- Direct detection experiments: Some BSM particles, such as dark matter, may interact only weakly with ordinary matter and therefore be difficult to detect in colliders or astrophysical observations. Direct detection experiments aim to detect these particles by looking for their interactions with matter in specialized detectors.

Overall, probing BSM is a multifaceted effort that involves a combination of theoretical, experimental, and observational approaches. The detection of physics beyond the SM would have notable consequences for our comprehension of the universe. This thesis concentrates on examining BSM scenarios by analyzing data obtained from a particle collider, which aims to resolve the Higgs mass hierarchy problem.

After the discovery of the Higgs boson, the SM is in principle complete as a low-energy effective theory describing all known fundamental particles and their interactions. However, the stability of the Higgs boson mass at the electroweak scale lacks explanation. The

main challenges in understanding the Higgs boson mass is the fact that it is subject to quantum loop corrections from other SM particles in the theory. In particular, the top quark, which is the heaviest elementary particle in the SM, has a large impact on the Higgs boson mass. As the energy scale increases, the contribution from the loop corrections to the Higgs boson self-energy diverges quadratically [30], leading to a large correction pushing the Higgs mass to the Planck scale, which is much larger than the electroweak scale. However, the observed mass of the Higgs boson ($\approx 125.38 \pm 0.14 \text{GeV}$) is much lighter than the Planck mass. Therefore, the question is why the observed Higgs boson mass is so much lighter than the Planck mass. This is popularly known as the "hierarchy problem". One proposed solution to explain the observed Higgs boson mass in the electroweak scale is called "naturalness," there should exist some underlying mechanism or symmetry in the theory that counteracts these large quantum corrections to the Higgs boson mass. Such proposed solution to prevent this divergence from the loop corrections is the introduction of new physics beyond the standard model that predicts additional particles that can cancel out these quantum corrections to the Higgs boson mass. One example of such a particle is a vector-like quark (VLQ), which is a hypothetical particle that is similar to the quarks in the SM but has different properties. VLQs are predicted in many theories beyond the SM, including composite Higgs models [31, 32, 33, 34, 35], little Higgs models [36, 37, 38], and models with a warped extra dimension [39]. VLQs could have a significant impact on the Higgs boson mass because they interact with the Higgs boson. If VLQs exist, they could cancel out the quadratic divergence from the top quark, leading to a more stable Higgs boson mass.

The sections below provide a brief overview of the models and searches considered in this thesis coupled with SM phenomena.

1.7.1 Search of Higgs production associated with a single top quark

A remarkable property of the Higgs boson is its large Yukawa coupling to top quark y_t. The measurements of y_t are particularly crucial for understanding electroweak symmetry breaking and allow for testing theories beyond the standard model. The value of y_t is indirectly tested by measurements sensitive to gluon gluon fusion ggH, the dominant Higgs boson production process at LHC, which receives a large contribution from loop diagrams involving the top quarks. In addition, y_t is probed in the decay of Higgs bosons to two photons, H $\rightarrow \gamma\gamma$, as the decay width also involves loop diagram with a top quark. However, y_t can be directly measured in the production of top - anti-top quark pairs, $t\bar{t}$, associated with a Higgs boson. Interestingly, the tH production is sensitive to the sign of y_t. In the SM the relative sign of the Higgs boson coupling to bosons and fermions is assumed to be positive. Therefore it leads to destructive interference between t-channel diagrams (Fig. 3.1), i.e., the Higgs bosons emitted from the top quark and the W boson lines, resulting in a very small production cross section of about 71 fb. The existing measurements using data collected at the LHC so far are not yet sensitive to this production process. However, in the new physics scenario, an opposite sign in the Yukawa coupling compared to the SM would translate to a significant enhancement ($\sim 10\times$) [40, 41] in the tH production cross section.

We design an analysis to measure the cross section of the various Higgs production modes in proton-proton (pp) collisions at $\sqrt{s}$ = 13 TeV using the data collected during Run 2. and to also facilitate a final combination to measure the Higgs boson signal strength in the H $\rightarrow \gamma\gamma$ channel. As part of this study, I work on measuring the cross section of tHq production in the channel where the Higgs boson decays into two photons and the top quark decays leptonically. The detailed analysis is discussed in Chapter 3

1.7.2 Search of a Vector-Like Quark, $T' \to tH$

Vector-like quarks are hypothetical spin-½ colored particles, labeled T' and B', with electric charges of $+2e/3$ and $-1e/3$, respectively. Their left-handed and right-handed components transform in the same way under the SM gauge group. Therefore, unlike the chiral quarks in the SM, their masses are not generated by a Yukawa coupling to the Higgs boson and have less impact on the production cross section of the Higgs boson. A simplified extension of the SM has the SM field content extended by a single species of vector-like quarks T' of mass $M_{T'}$. After imposing an $SU(3)_c \times U(1)_Q$ gauge symmetry, the Lagrangian of the considered simplified model can be written as

$$\mathcal{L} = \mathcal{L}_{SM} + i\bar{T}'D\,T' + M_{T'}\bar{T}'T' + \Big[h\bar{T}'(\hat{\kappa}_L P_L + \hat{\kappa}_R P_R)u_q + \frac{g}{2c_W}\bar{T}'\,\slashed{Z}\,(\tilde{\kappa}_L P_L + \tilde{\kappa}_R P_R u_q)u_q$$
$$+ \frac{g}{\sqrt{2}}\bar{T}'\,\slashed{W}\,(\kappa_L P_L + \kappa_R P_R)d_q + h.c.\Big],$$

$$(1.57)$$

where, c_W denotes the cosine of the EW mixing, g denotes the weak coupling constant, and $\kappa, \hat{\kappa},$ and $\tilde{\kappa}$ represent the EW coupling of the vector-like quark T. Moreover, u_q and d_q denote the SM up-type and down-type quark fields, Z_μ, W_μ, and h stand for the weak and Higgs boson fields, and P_L and P_R are the usual left-handed and right-handed chirality projectors. The last three terms in the above Lagrangian open the door to single vector-like quark production at hadron colliders. In general, the κ parameters are taken to be small so that the vector-like quark stays narrow.

At the CERN LHC, the vector-like top quark T' can be produced in two production modes. One is pair production through the strong interaction while the other is single production mode through the electroweak interaction. The T' quark can couple to SM quarks and charged or neutral bosons, resulting in the decays into bW, tZ, and tH channels. For the isospin singlet T' VLQ, the T' branching fractions are assumed to be 50, 25, and 25%, respectively, for bW, tH, and tZ decays. We studied the electroweak production of vector-
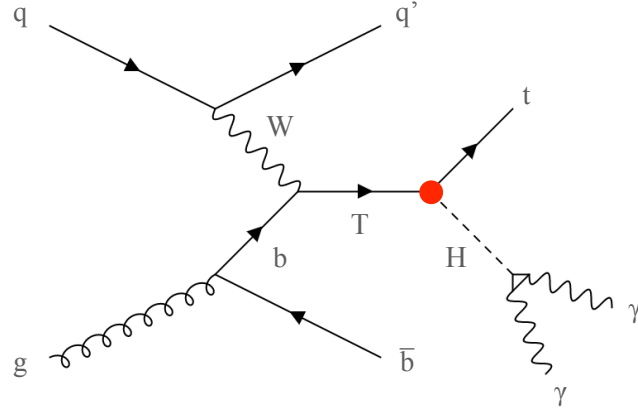
Figure 1.13: Leading-order Feynman diagram for single T′ production in Wb fusion and subsequent decay into tH($\gamma\gamma$).

like top quark partner T′ produced in proton-proton (pp) collisions at $\sqrt{s}$ = 13 TeV, pp → T′ qb, with a decay involving the Higgs boson, T′ → tH, in Higgs to two photon channel (H → $\gamma\gamma$). A leading-order (LO) Feynman diagram is presented in Fig. 4.2. The search is designed to be sensitive to both leptonic and hadronic decays of the top quark together with a Higgs boson decaying to two photons. Chapter 4 provides a detailed analysis of this search.

Chapter 2

Large Hadron Collider & CMS detector

The primary goal for building the Large Hadron Collider was to describe the Higgs mechanism and confirm its role in electroweak symmetry breaking. The experimental study of the Higgs mechanism is also crucial in understanding the mathematical consistency of the SM at energy scales above about 1 TeV. However, numerous other theoretical motivations propose additional symmetries like supersymmetry or introduce new forces and constituents such as strongly-broken electroweak symmetry, technicolor, and more, which can also be potentially verified. The selection of beam energy (7 TeV) and design luminosity ($L = 10^{34} cm^{-2} s^{-1}$) for the Large Hadron Collider was motivated by the aim to investigate physics phenomena at the TeV energy scale. Hence, this substantial increase in energy and luminosity over previous collider experiments opens up opportunities for exploring a broad spectrum of physics inquiries. The potential physics search in increased energy and luminosity required a meticulous design of the detectors.

2.1   Large Hadron Collider (LHC)

LHC is the world's most powerful particle accelerator. It lies in a tunnel 27 km in circumference and as deep as 175 m beneath the France–Switzerland border near Geneva. The LHC is made up of a 27km-long ring of superconducting magnets with a number of acceleration mechanisms to boost the energy of the particles. The accelerator ring consists of two tubes to accelerate two beams in opposite directions which are kept in ultra vacuum with the pressure inside the pipe of the order of 10^{-10} to 10^{-11} mbar. Inside the accelerator, two high-energy proton beams travel nearly at the speed of light before they are made

to collide. A strong magnetic field maintained by superconducting electromagnets is applied to guide the accelerating particle inside the accelerator ring. The magnet include 1232 dipole magnets, 15 metres in length, which bend the beams, and 392 quadrupole magnets, each 5–7m long, which focus the beams. Another kind of magnet is employed to "squeeze" the particles closer together just before the collision to boost the likelihood of collisions. The electromagnets in the LHC are constructed using coils made of unique electric cable capable of operating in a superconducting state. This allows for the efficient conduction of electricity with minimal resistance or energy loss. Achieving this superconducting state requires cooling the electric cables to a temperature as low as $-271.3°C$. A liquid helium distribution system is used to achieve and maintain this temperature. The cryogenic system for the LHC needs 120 tonnes of helium, 40 MW of electricity (10 times more than is required to operate a locomotive), and 40,000 leak-proof pipe seals to keep the magnets at $-271.3°C$. The LHC machine also contains radiofrequency cavities that deliver a "kicks" to the protons, leading to an incremental energy increase of 0.5 MeV per revolution. The luminosity is written as:

$$L = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n\beta^*}F \qquad (2.1)$$

where, γ is the Lorentz factor, f is the revolution frequency, k_B is the number of bunches (groups of protons), N_p is the number of protons per bunch, ϵ_n is the normalized transverse emittance (spread the protons perpendicular to the direction of the beam) with a design value of $3.75\mu m$, β^* is the betatron function at the interaction point (IP), and F is the reduction factor resulting from the crossing angle. β is a measure of the focusing strength or the rate at which the beam of particles converges or diverges in the transverse plane. The nominal energy of each proton beam is 7 TeV. The design luminosity of $L = 10^{34}cm^{-2}s^{-1}$ leads to around 1 billion proton-proton interactions per second. The specific configuration of the beam is determined based on the injection scheme and characteristics of the dump system.

The formation of the bunches occurs within the Proton Synchrotron (PS) with a spacing of 25 ns. The important machine parameters relevant for proton-proton collisions at the LHC are listed in Tab. 2.1. The protons are prepared by a series of systems that gradually

Beam parameters

		pp	Unit
Energy per proton	E	7	TeV
Dipole field at 7 TeV	B	8.33	T
Design Luminosity	L	10^{34}	$cm^{-2}s^{-1}$
Bunch separation		25	ns
No. of bunches	k_B	2808	
No. of protons per bunch	N_p	1.15×10^{11}	
Collisions parameters			
β-value at IP	β	0.55	m
RMS beam radius at IP	σ	16.7	μm
Luminosity lifetime	τ_L	15	hr
No. of collisions/crossing	n_c	≈ 20	

Table 2.1: The design values of the important machine parameters relevant for proton-proton collisions at the LHC

boost their energy before being introduced into the primary accelerator. A schematic view of the CERN accelerator complex comprised of various components used to generate the proton beam for the LHC is displayed in Fig. 2.1. The first system is the linear particle accelerator LINAC4 generating 160 MeV negative hydrogen ions (H$^-$) [42], which feeds the Proton Synchrotron Booster (PSB). There, both electrons are stripped from the hydrogen ions leaving only the nucleus containing one proton. Protons are then accelerated to 2 GeV and injected into the Proton Synchrotron, where they are accelerated to 26 GeV. Afterwards, the beam undergoes further acceleration to reach an energy level of 450 GeV within the Super Proton Synchrotron (SPS) before being transferred to the LHC. The LHC beams are made to collide at the four points in the accelerator ring, which correspond to the locations of the four particle detectors ATLAS (Point 1), CMS (Point 5), ALICE (Point 3) and LHCb (Point 8).
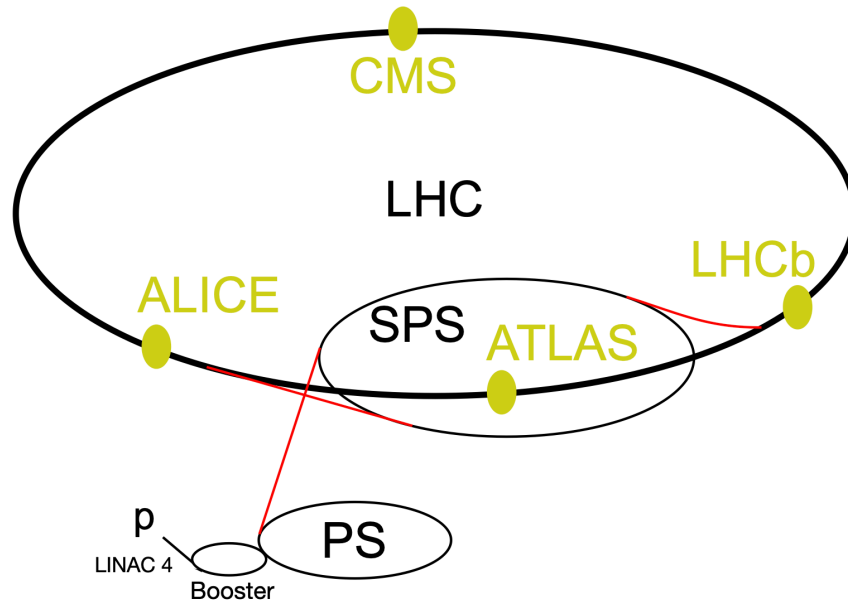
Figure 2.1: The CERN accelerator complex comprises of various components, including the Large Hadron Collider, Super Proton Synchrotron, Proton Synchrotron, LINear ACcelerator (LINAC).

- **A Toroidal LHC Apparatus (ATLAS)**: ATLAS is one of the two general-purpose detectors at the LHC. It looks into a variety of physics topics, including the study of the Higgs boson, and the search for new phenomena such as extra dimensions and particles that might be responsible for dark matter.

- **Compact Muon Solenoid (CMS)**: Another general-purpose detector at the LHC, engages in a comprehensive physics program as ATLAS. Although it uses distinct technical approaches and a different magnet-system design from the ATLAS experiment, it shares the same scientific objectives.

- **LHC-beauty (LHCb)** The "beauty quark," or "b quark," is a particular kind of particle that the LHCb experiment studies to learn more about the minute variations between matter and antimatter.

- **A Large Ion Collider Experiment**: ALICE is a detector at the Large Hadron Collider devoted to heavy-ion physics. It is intended to investigate the physics of strongly interacting matter at extreme energy densities when a state of matter known as quark-gluon plasma forms.

The data used in this thesis were obtained from the CMS detector. Section 2.2 provides a detailed description of the CMS detector.

The general-purpose detectors would observe an event rate of approximately 10^9 inelastic (maximum momentum transfer) events per second at the designed luminosity. This presents a variety of complex experimental challenges. For storage and physics analysis, the online event selection procedure (the "trigger") must reduce the roughly 1 billion interactions per second to no more than about 100 events per second. The design of readout and trigger systems must consider the short 25 ns interval between bunch crossings. A mean of around 20 inelastic collisions will be superimposed on the event of interest at the design luminosity. This means that every 25 ns, around 1000 charged particles will emerge from the interaction region. It is possible to mistake the results of particular interest of interaction with those from other interactions in the same bunch crossing. This problem clearly gets worse when the response time of a detector element and its electronic signal is longer than 25 ns. Therefore, using high-granularity detectors with good time resolution and low occupancy can reduce the impact of this "pile-up". This requires a large number of detector channels. The timing of these millions of detector electronic channels needs to be very accurately synchronized. High radiation levels are caused by the huge flux of particles leaving the interaction region, demanding the use of radiation-hard detectors and front-end electronics.

2.2   CMS detector

The CMS detector is a multi-purpose apparatus operating at the Large Hadron Collider (LHC) at CERN. One crucial factor that drives the design and arrangement of the detector is the selection of the magnetic field configuration for the measurement of the momentum of muons. CMS is 21 m long, 15 m in diameter, and weighs about 14,000 tonnes. Over 4,000 people, representing 206 scientific institutes and 47 countries, form the CMS collaboration who built and now operate the detector. Figure 2.2 illustrates the overall design of CMS. At the center of CMS is a 13 m long, 6 m inner diameter, 3.8 T superconducting solenoid that offers a significant amount of bending power (12 Tm) before the muon system measurement of the muon bending angle. Four muon stations can be integrated to assure robustness and complete geometric coverage, and the return field is large enough to completely saturate 1.5 m of iron. The muon station is made up of many layers of resistive plate chambers (RPC), together with cathode strip chambers (CSC) and aluminium drift tubes (DT) in the endcap and barrel regions, respectively. The bore of the magnet coil is spacious to accommodate both the inner tracker and the calorimetry components. The tracking volume is a cylindrical with a length of 5.8 m and a diameter of 2.6 m. CMS uses ten layers of silicon microstrip detectors to effectively handle the challenges posed by high track multiplicities. The detectors are designed with the necessary granularity and precision for accurate measurements. In addition, four layers of silicon pixel detectors are positioned close to the interaction region to enhance the precision in impact parameter measurements of charged-particle track and the secondary vertex positions. For the electromagnetic calorimeter (ECAL), lead tungstate ($PbWO_4$) crystals are used, covering a pseudorapidity range up to $|\eta| < 3.0$. The scintillation light in the crystal is detected by silicon avalanche photodiodes (APDs) in the barrel region, while in the endcap region, vacuum phototriodes (VPTs) are employed. Additionally, a preshower system is installed in front of the endcap

ECAL to reject π^0 particles. Surrounding the ECAL is a brass/scintillator sampling hadron calorimeter, which covers the pseudorapidity range up to $|\eta| < 3.0$. The scintillation light is converted by wavelength-shifting fibers embedded in the scintillator tiles and guided to photodetectors through clear fibers. This light is detected by photodetectors known as hybrid photodiodes (HPDs), which can operate in high axial magnetic fields and provide gain.
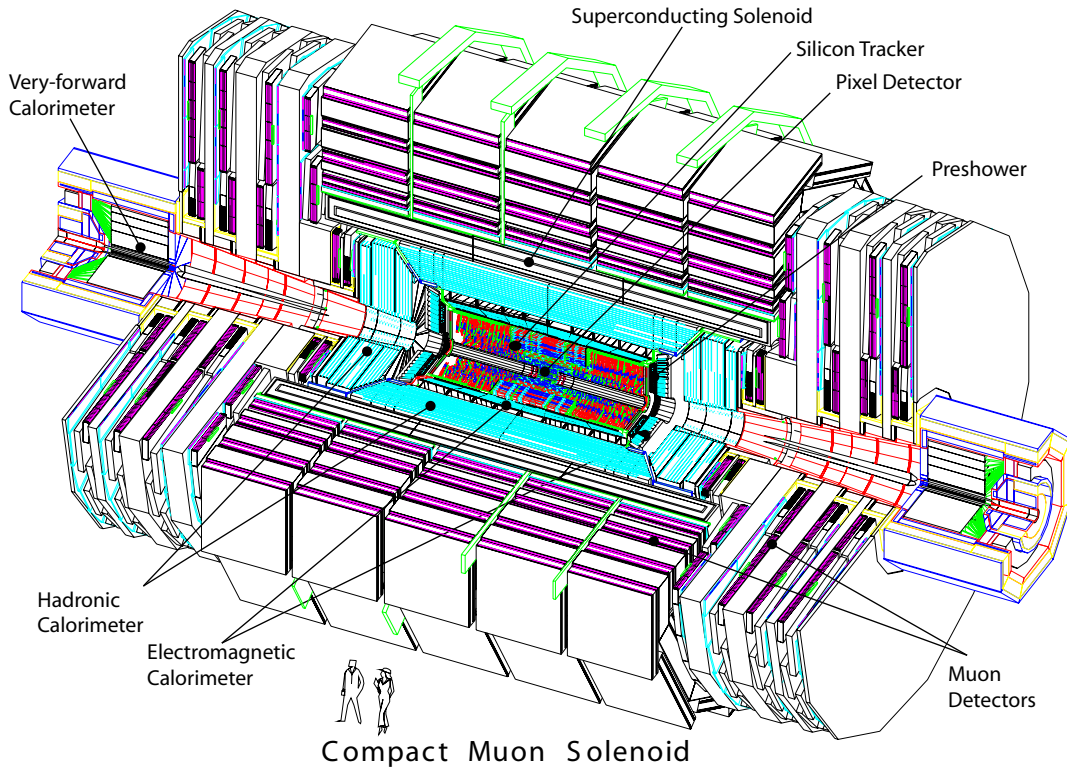
Figure 2.2: A schematic view of the CMS detector. The picture is taken from Ref. [4].

2.2.1 LHC coordinate system

CMS adopts the following coordinate system conventions: the x-axis is directed radially inward toward the center of the LHC, the y-axis points vertically upward, and the origin is situated at the nominal collision point of the experiment. Consequently, the z-axis aligns

with the direction of the beam. The azimuthal angle ϕ is measured from the x-axis within the x-y plane, while the polar angle θ is determined relative to the z-axis. In the experiment it is more useful to have a quantity with better properties under boosts along the beam axis rather than using the polar angle. Such a quantity is rapidity, and is defined as,

$$y = \frac{1}{2} \ln \left(\frac{E + p_z c}{E - p_z c} \right) \qquad (2.2)$$

The only issue with the rapidity is that it requires determining both the energy and the momentum of a particle, which is challenging experimentally. For very energetic particles, there is a definition for a quantity that is nearly identical to the rapidity but considerably simpler to measure than y. This introduces the idea of pseudorapidity and is defined as $\eta = -\ln \tan \frac{\theta}{2}$. In the analysis, the x and y components of momentum are utilized to calculate the transverse momentum (p_T). The transverse plane refers to the plane perpendicular to the beam direction. The imbalance of energy measured after applying energy conservation in the transverse plane is denoted by E_T^{miss}.

2.2.2 Inner tracking system

The total length of the CMS tracker is approximately 540 cm and its outer radius extends to nearly 110 cm. The excellent position resolution of the silicon detectors is attained by splitting a large area of a p-n diode into many small pixels or parallel strips, where each pixel or strip division acts like an independent electrode. Pixel detectors are placed closer to the interaction vertex where the particle flux is the highest ($\approx 10^7/s$ at r $\approx$ 10 cm). Low particle flux in the intermediate region ($20 < r < 55$ cm) enables using silicon microstrip detectors with a minimum cell size of 10 cm $\times 80$ μm, leading to an occupancy of $\approx$ 2–3%/LHC crossing. The further decrease in particle flux in the outermost region ($r > 55$ cm) of the inner tracker permits the use of larger-pitch silicon microstrips with a maximum cell size of 25 cm $\times 180$ μm giving an occupancy of about 1%.

When an incident ionizing particle interacts with the detector medium (depleted region), it produces electron-hole pairs. Due to the electric field in the depletion region, the holes move in the direction of the p$^+$ (negative electrode) and the electrons move in the direction of the n$^+$ (positive electrode). An external reverse bias voltage must be applied to create a deep depletion zone, void of free charge carriers. The collected charges by the electrode produce a current pulse, allowing for the determination of the position of the ionizing particle. Figure 2.3 depicts a strip silicon detector. Currently, the vast majority of silicon
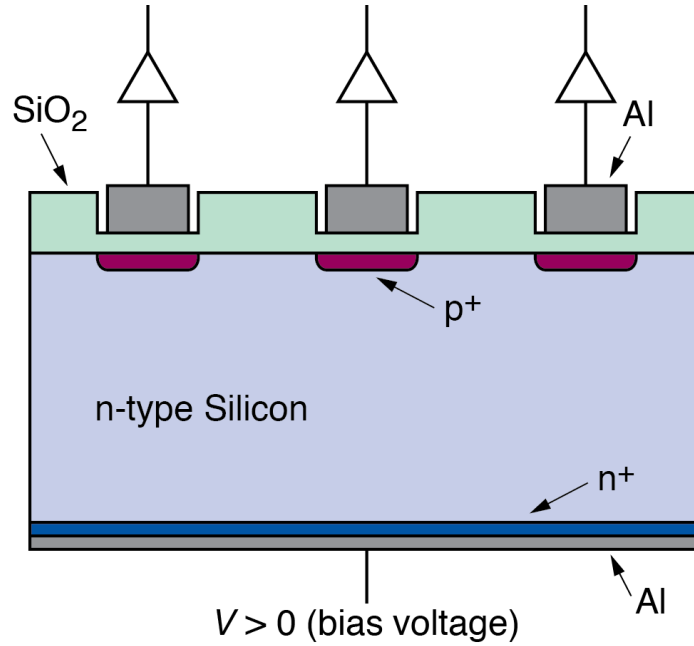
Figure 2.3: A schematic view of the silicon sensor. The picture is taken from Ref. [5].

detectors are implemented as so-called planar structures.

Strip tracker

The total area of the strip detectors is $\approx 200 \; m^2$ providing a coverage up to $|\eta| < 2.4$ and includes 9.6 million silicon strips. The detectors are operated at a temperature of around $-20°C$. There are two sections in the tracker barrel region: a TIB (Tracker Inner Barrel) and

a TOB (Tracker Outer Barrel). The TIB is made of 4 layers of silicon sensors that cover up to $|z| < 65$ cm with sensors of thickness 320 μm and the pitch of the strip that ranges from 80 to 120 μm. The TOB consists of six layers, each with a half-length of $|z| < 110\ cm$. This region experiences lower radiation levels, allowing for the utilization of thicker silicon sensors (500 μm). The choice of longer strip lengths and wider pitch also retains a favorable signal-to-noise ratio. The range of the strip pitch is 120 to 180 μm.

The endcaps of the CMS tracker are subdivided into two sections: the Tracker End Cap (TEC) and the Tracker Inner Disks (TID). Each TEC is made up of nine disks that spread into the region $120\text{cm} < |z| < 280$ cm, and each TID is made up of three small disks that fill the gap between the TIB and the TEC. For TID and three innermost rings of the TEC, the thickness of the sensors is 320 μm, and for the rest of the TEC, the thickness is 500 μm. The modules of the TEC and TID are organized in concentric rings centered on the beam line. These modules consist of strips that are oriented towards the beam line.

Pixel tracker

The pixel detector is comprised of three layers in the barrel region, with two endcap disks on each side. The three barrel layers are positioned at average radii of 4.4 cm, 7.3 cm, and 10.2 cm. Each barrel layer has a length of 53 cm. The two end disks are positioned at $|z| = 34.5$ cm and 46.5 cm on either side, with radii ranging from 6 to 15 cm. The pixel detector has a surface area of around $\approx 1\ m^2$. There are 66 million pixels in the inner tracker. The size of a pixel is $\approx 100 \times 150\ \mu\text{m}^2$, resulting an occupancy of around 10^{-4} per pixel per LHC crossing. A total of 16000 readout chips are bump-bonded to the detector modules to read out the signals of particles passing through the detector.
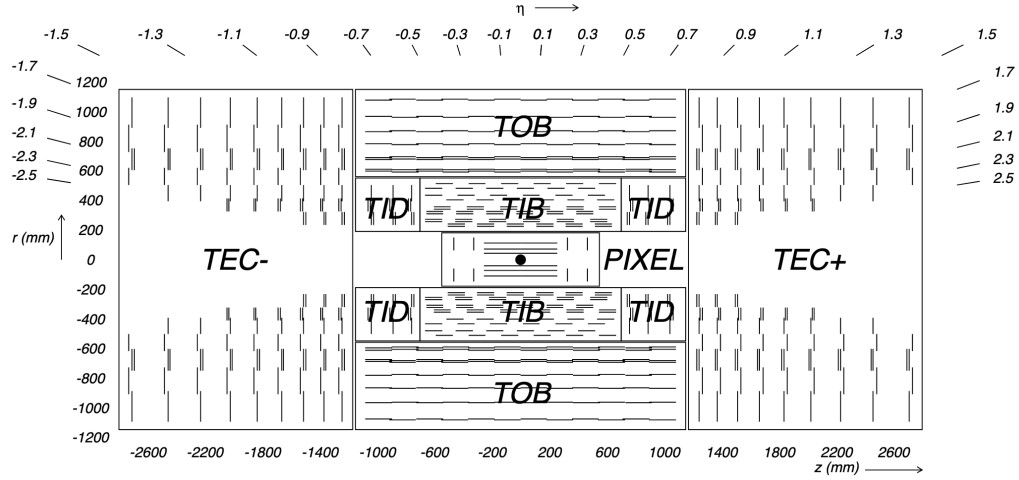
Figure 2.4: The layout of the CMS tracker detector is labeled with different parts of the detector. The picture is taken from [6]

2.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter [9] played an important role in studying the physics of electroweak symmetry breaking, particularly through the discovery of the Higgs boson. The ECAL is used to measure the energy of the electrons and photons generated during the collision. The ECAL is a hermetic, homogeneous calorimeter made up of $PbWO_4$ crystal as the active medium. This choice was based on the facts that $PbWO_4$ is a rapid scintillator, has a short radiation length and a small Molière radius, and is reasonably simple to make from easily accessible raw materials. Radiation length is average distance over which a high-energy electron or positron (beta particle) loses approximately 1/e (about 63.2%) of its energy through bremsstrahlung radiation and e^- - e^+ pair production in a given material. Molière radius is lateral or angular spread of high-energy charged particles (typically e^-/e^+) as they traverse a medium due to multiple Coulomb scattering interactions with atomic nuclei. The ECAL crystals have a radiation length (X_0) of 0.89 cm and and Molière radius of 2.2 cm. The crystal emits 80% of the light within 25 ns and has radiation hardness up to 10 Mrad. In Fig. 2.5, an image of the crystal is displayed. The ECAL is divided into

63
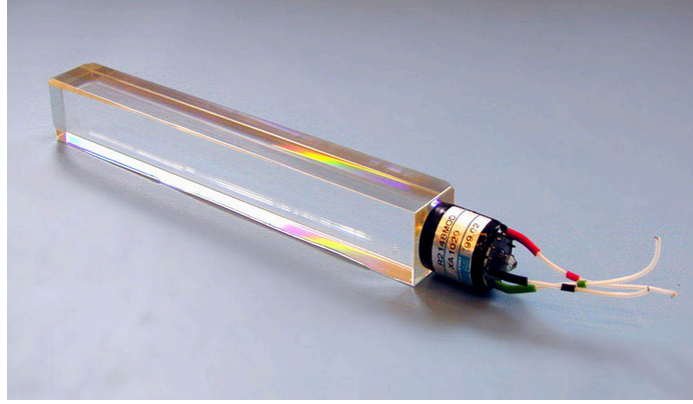
Figure 2.5: Lead tungstate crystal (PbWO$_4$) for the CMS ECAL with attached a photodetector. The figure is taken from [7]

two parts, barrel (EB) and endcap (EE). The EB area is covered with 61,200 crystals total, with 7,324 crystals in each EE. The inner radius of the barrel section is 129 cm. The barrel crystal axes are slanted at 3° with respect to the line from the nominal vertex, in the η and ϕ projections. The crystals are 23 cm in length or 25.8X$_0$, with the front face cross section of 22×22 mm^2 ($0.0174 \times 0.0174°$ in $\eta - \phi$ plane) and the rear face cross section of 26×26 mm^2. The endcaps are at a distance of 314 cm from the vertex and provide a pseudorapidity coverage of $1.479 < |\eta| < 3.0$. Like barrel crystals, the endcap crystals are organized in an x-y grid and off-point from the nominal vertex location. They are all similar and have a length of 22 cm, or 24.7 X$_0$, and a front face cross section of 28.6×28.6 mm^2 (24.7X$_0$). In front of the crystal calorimeter, most of the endcap pseudorapidity range is covered with a sampling calorimeter called a preshower device. It is intended to distinguish π^0 particle decay to two photons, aiming to achieve effective discrimination between π^0 and γ from other process. This discrimination is crucial for minimizing background in the search for the Higgs boson in the H $\rightarrow \gamma\gamma$ channel. The preshower detector is composed of two planes of 1.9 mm-pitch silicon strip detectors positioned behind the lead absorber disks of thickness 2X$_0$ and 1X$_0$, respectively. Different parts of the ECAL detector and the arrangement of crystals can be visualized in Fig. 2.6. The detail calibration of the ECAL can be seen in the Ref [43] Two
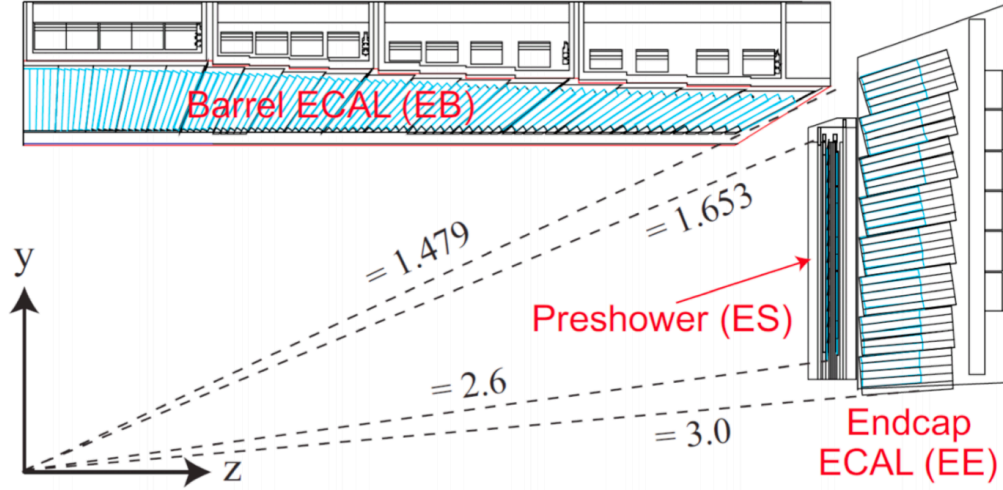
Figure 2.6: Different parts of the ECAL detector and the arrangement of crystals in pseudorapidity. The figure is taken from [8]

different kinds of photodetectors are used to read out the signal from the crystal, depending on the condition of magnetic field and radiation: the EB uses APDs, and the EE uses VPTs. The crystals and the APDs are very sensitive to temperature and require very stable temperature maintenance with the variation of the order of 0.1°C [44]. A multi-gain preamplifier amplifies the signal which then is sampled and digitized at 40 MHz in 1 of 3 chosen 12-bit ADCs for each channel. The energy resolution, measured by fitting a Gaussian function to the reconstructed energy distributions, has been parameterized as a function of energy:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.3}$$

Where, S is the stochastic term, N the noise term and C the constant term. The energy resolution of the PbWO$_4$ calorimeter is the result of many contributions as shown in Fig. 2.7. The stochastic component includes variations in shower confinement and incorporates photostatistics effects, represented as "photon" in the figure. The noise component considers electronic noise and the influence of pileup energy. The "intrinsic" curve in the figure shows the combined shower confinement and a constant term of 0.55%.

65
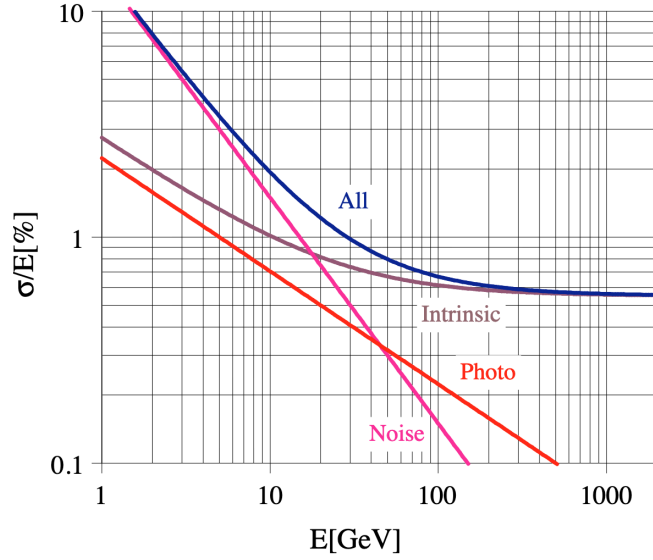
Figure 2.7: The ECAL energy resolution, σ_E/E, as a function of energy as measured from a test beam. The figure is taken from [9]

2.2.4 Hadron calorimeter

The Hadronic Calorimeter (HCAL) measures the hadrons produced in the collision, such as protons, neutrons, pions, and kaons. It also enables indirect measurements of the presence of non-interacting neutral particles such as neutrinos. The choice of magnet parameters strongly influences the design of the HCAL since most of the CMS calorimetry is located inside the magnet coil. The HCAL is a sampling calorimeter made of repeating layers of dense brass absorber plates and tiles of plastic scintillator that maximizes material inside the magnet to increase the total interaction lengths (λ_I). The interaction length is the distance over which a hadron has a 1/e (approximately 37%) probability of passing through without undergoing a nucleus interaction. The set of scintillators is grouped with alternate layers of the absorber, referred to as a "tower". In order to determine a particle's position, energy, and arrival time, the HCAL uses fluorescent "scintillator" materials. When a particle passes through the scintillator, it generates a quick blue-violet light pulse. This light

66

is gathered by special optical fibers and sent into readout boxes to photodetectors, which enhance the signal. The hadron barrel (HB) part of the HCAL has 14 brass layers surrounded by a layer of steel at the front and back yielding an interaction length between 5.8 to 10.6. There are 17 active plastic scintillator tiles interspersed between the stainless steel and brass absorber plates. The initial active layer is positioned right behind the ECAL and features a scintillator thickness roughly twice that of the other layers (9 mm as opposed to 3.7 mm). This greater thickness allows it to actively sample low-energy showering particles originating from the support material situated between the ECAL and HCAL. The final active layer also has a scintillator thickness of 9 mm. It consists of 32 towers covering the pseudorapidity region $|\eta| < 1.4$, resulting in 2304 towers with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The space limitation imposed by the magnet coil prevents the calorimeters from totally containing hadronic showers. So beyond the solenoid, there is an extra outer hadron calorimeter (HO). The HO employs the same scintillator technology as the HB and the solenoid coil as an absorber. The HO detector is made of scintillators that are 10 mm thick. These scintillators are positioned along the exterior of the outer cryostat wall of the coil and cover the region with $|\eta| < 1.26$. As for the HB, it consists of 14 brass layers surrounded by a layer of steel at the front and back. The total thickness of the HB ranges from 5.8 to $10.6\lambda_I$, depending on η^2. For the test beam of 300 GeV pions, energy measured with tracker, EB, and HB, it is observed that nearly 3% of pion events result in missing energy equivalent to 100 GeV. When the HO is taken into consideration, the interaction length becomes $11.8\lambda_I$, reducing energy loss significantly. Energy resolution for pions as a function of beam energy measured with EB + HB and with EB + HB + HO can be seen in the Ref. [45]. Each hadron endcap (HE) of HCAL consists of 14 η towers with $5°$ ϕ segmentation, covering the pseudorapidity region $1.3 < |\eta| < 3.0$. Finally, in the extreme front area, a forward calorimeter is designed to maintain a good performance of the jet reconstruction in the forward region. The requirement to tolerate a very high radiation exposure

was the main feature of its design. The Hadron Forward (HF) calorimeter consists of steel and quartz fibers covering the pseudorapidities between 3.0 and 5.0. The performance of
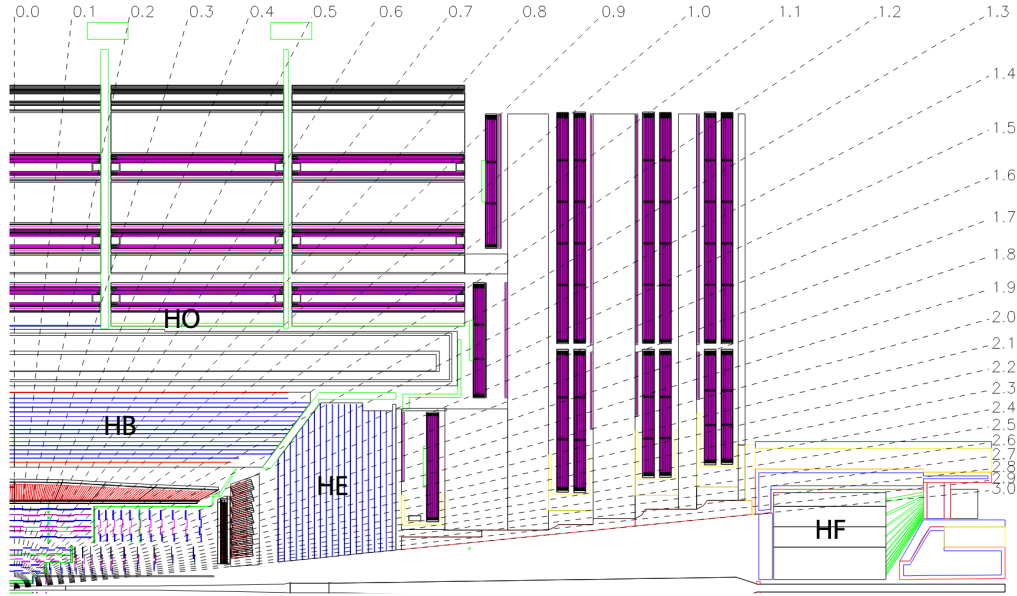
Figure 2.8: The layout of a quarter of the CMS HCAL detector is labeled with different parts (HB, HE, HO and HF) of the detector. The picture is taken from [6]

the HCAL can be checked from the jet energy resolution and the missing transverse energy resolution. The jet transverse energy resolution as a function of the simulated jet transverse energy in all three parts of the HCAL, barrel, endcap and forward is shown in the Fig. 2.9. The normalization in granularity of different parts of the HCAL is considered to compare the plots in different η ranges in the figure. The missing transverse energy (E_T^{miss}) resolution is given by $\sigma(E_T^{miss}) \approx 1.25\sqrt{\sum E_T}$, if energy clustering corrections are not made.

2.2.5   Superconducting magnet

One of the essential aspects of the CMS detector design is the configuration of the magnetic field to measure the momenta of muons. The magnetic field is solenoidal. It is parallel. For
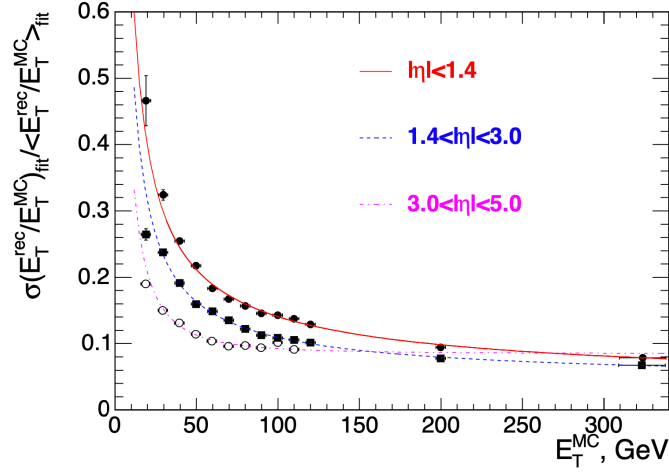
Figure 2.9: The jet transverse energy resolution as a function of simulated jet transverse energy in all three parts of HCAL, barrel, endcap and forward. The figure is taken from [10]

the following reasons, a solenoidal field is preferred: the field is parallel to the beams, the muon track bends in the transverse plane, and momentum measurement can begin at $r = 0$ with a solenoidal field, whereas in the case of a toroidal field, it begins after the absorber material. The required long superconducting solenoid (L = 13 m) has a 5.9 m inner diameter and a 3.8 T uniform magnetic field. A 3.8 T magnetic field has significant advantages not only for muon tracking and inner tracking but also benefits electromagnetic calorimetry. Due to the presence of this intense magnetic field, the low-energy electrons within the scintillator spiral away from each other and the light yield increases.Two critical features of the CMS solenoid are the implementation of an indirect cooling system by thermosyphon cooling system, which cools down to a temperature of 4 Kelvin, and the use of a stabilized conductor made of high-purity aluminium. In particular, a novel conductor with a bigger cross section that can withstand an outward pressure (hoop stress) of 64 atmospheres has been used in a four-layer winding. The conductor has a complex structure and can carry 20 kA of current. Pure aluminum is co-extruded with the Rutherford-type cable [46] to serve as a thermal stabilizer. The conductor's entire cross section measures 64×22 mm^2.

The conductor is produced in continuous lengths, each measuring 2.65 km. Such twenty continuous lengths were produced. Each of the five coil modules is constructed using four of these lengths.

2.2.6 Muon system

Given the total luminosity predicted, muon detection is a potential technique for identifying the fingerprints of intriguing events above the extremely high background rate. For example, it has been said that the Standard Model Higgs boson decay into ZZ or ZZ*, which further decay into 4 leptons, is "gold plated" if all the leptons are muons. The momentum of the generated muons in the collision can be measured in three parts of the CMS detector: in the inner tracker, after the coil, and in the return flux. The momentum measurement simply using the muon system is essentially determined based on the muon bending angle at the exit of the 3.8 T coil with respect to the interaction point. The muon system incorporates three
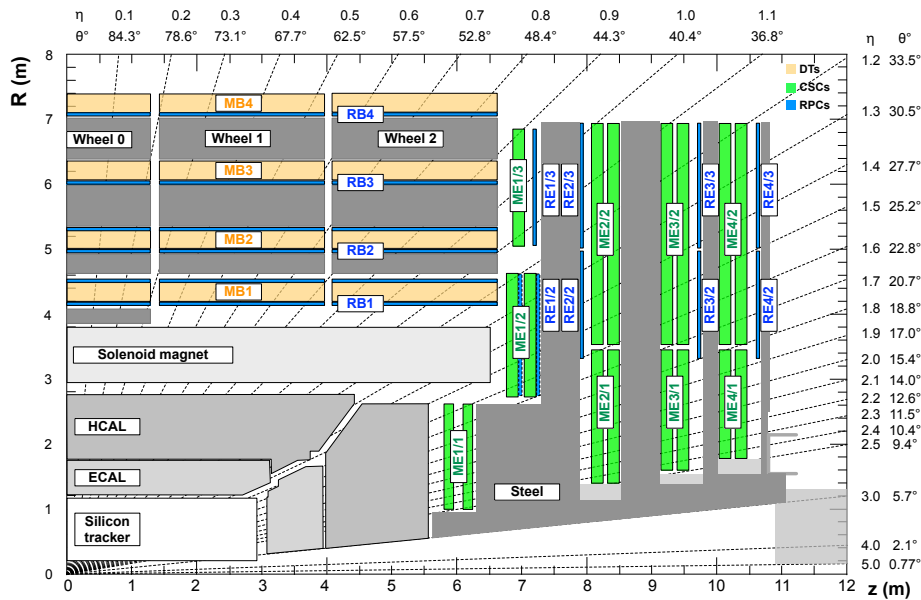
Figure 2.10: Layout of one quarter of the CMS detector with muon detectors in colors. The figure is taken from [11]

types of gaseous detectors: drift tube chambers, cathode strip chambers, and resistive plate

chambers. DT chambers are used in the barrel region ($|\eta| < 1.2$), where the background induced by neutrons is minimal, the muon rate is low, and the residual magnetic field within the chambers is also low. In the endcap regions, CSC detectors are used where both the muon rate and the neutron-induced background rate are high, and the magnetic field is also strong. These CSC detectors cover the region up to $|\eta| < 2.4$. Additionally, RPC detectors are used in both the barrel and endcap regions. Layout of one quarter of the CMS detector with muon detectors in colors is shown in the Fig. 2.11 Various methods are employed to align the muon system, including photogrammetry measurements, optical alignment, and alignment based on particle tracks [47]. In DT drift cell, the transverse distance between the wire and the point at which the muon trajectory crosses the plane containing the wires in the layer is found by hit reconstruction. The anode collects the electrons produced through the gas ionized by a muon crossing the cell. The arrival time, T_{TDC} is recorded by a time-to-digital converter (TDC). In order to reconstruct the position of the DT hit, this time is then adjusted by a time pedestal, T_{ped}, and multiplied by the electron drift velocity, v. The DT drift cell was designed to provide a uniform electric field allowing the drift to be mostly constant for tracks impinging on the cell perpendicularly to the plane of wires.

In the endcaps, the solenoidal field is first parallel to the z direction but then diverges radially. Therefore a muon is first deflected in one azimuthal direction and then deflected in the opposite direction. The deflection is maximum in the first station. The position of the traversing muon in the CSC layer is determined via hit reconstruction by merging data from the cathode strips and anode wires. The strips can precisely measure the ϕ angle since they are radial and individually subtend an angle of around 3 mrad (various chamber types have varying angular strip widths that range from 2.2 to 4.7 mrad). The measured time for a CSC reconstructed hit is additionally calibrated so that muon hits produced promptly in the triggering bunch crossing have a time distribution centered around zero.

A charged particle that enters an RPC causes an electron avalanche in the space between

two plates. With a few nanoseconds of accuracy, this charge creates a signal on an external strip reading plane to identify muons from collision events. The alignment of the strips is along the η direction, with a strip pitch of up to 2 cm. This results in a spatial resolution of a few centimeters in the ϕ coordinate. In an RPC, adjacent strips are grouped to reconstruct one hit since more than one strip might share the ionization charge from a muon.

Compared to DTs or CSCs, RPCs offer a quick reaction with strong time resolution but a cruder position resolution. As a result, RPCs are able to clearly detect the bunch crossing. The DTs or CSCs and the RPCs operate as two separate and complementary sources of information inside the first level trigger system. The complete system results in a reliable, accurate, and adaptable trigger device. Comparison of muon momentum resolution reconstructed from muon system only, inner tracker only, or both is shown in the Fig. 2.10.
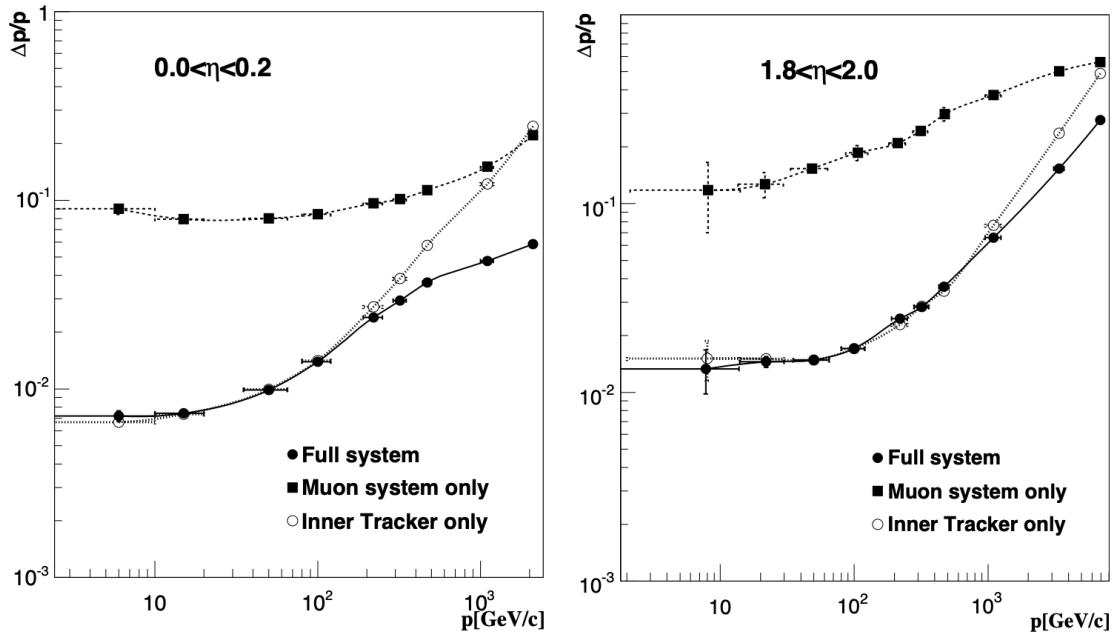
Figure 2.11: Comparison of muon momentum resolution reconstructed from the muon system only, the inner tracker only, or both ("full system") with respect to different muon momentum. The left plot is for barrel $|\eta| < 0.2$ and the right side plot is for endcap, $1.8 < |\eta| < 2.0$. The figure is taken from [10]

72

2.2.7   Trigger and data acquisition system (DAQ)

At design luminosity, the LHC bunch crossing rate of 40 MHz results in $\approx 10^9$ interactions per second. Since archival media can only store data from roughly 10^2 crossings per second, the trigger mechanism must reach a rejection of around 10^7 interactions. The reduction process involves several components, including the detector electronics, level-1 trigger processors (specifically for the calorimeter, muon, and combining different subdetector systems), the readout network, and an online event filter system referred to as the processor farm. The processor farm executes the software for the high-level triggers (HLT).

Level-1 trigger

The Level-1 trigger system relies heavily on specialized hardware components for its logic implementation. This includes custom application apecific integrated circuits (ASICs), semi-custom and gate-array ASICs, field programmable gate arrays (FPGAs), programmable logic devices (PLDs), as well as discrete logic elements like random access memories (RAMs) used for memory look-up tables (LUTs). To minimize the transit time between the signals from the detector front-end electronics to the Level-1 trigger logic and return back to the front-end electronics, the Level-1 trigger logic farm is kept in the service cavern of the CMS experiment. A total of 3.2 μs is allotted for the transit and for deciding whether to preserve or delete data from a specific bunch crossing. The detector data must be held in a buffer during that time while trigger data is gathered from the front-end electronics and judgments are made to dismiss a large portion of events keeping the small portion of interactions of interest. The time allocation for Level-1 trigger decision is less than 1 μs of the overall latency.

The Level-1 trigger involves the information from calorimetry and muon systems, along with inter-system correlated information. The decision-making process is based on the detection of "trigger primitive" particles, including photons, electrons, muons, and jets,

which pass a specified threshold for E_T and p_T. Additionally, global sums of E_T and E_T^{miss} are taken into account. Trigger objects are formed using reduced-granularity and reduced-resolution data. During the Level-1 decision-making period, all the high-resolution data is held in pipelined memories. After the Level-1 trigger pass, further decisions are made using more sophisticated algorithms that leverage detailed information from all detectors.

High-Level triggers

For the HLT decision, a farm of high-performing computers with approximately 13,000 CPU cores runs software implementing the desired criteria. Upon receipt of a Level-1 trigger, the data from the pipelines are transferred to front-end readout buffers. Each event has an approximate size of 1.5 MB (pp interactions) and is contained in several hundred front-end readout buffers. Data for a given event are sent to a processor via the event building "switch." Each processor executes the same HLT software code, aiming to reduce the Level-1 output rate from 100 kHz to 100 Hz for storage purposes. The HLT code is developed using a variety of ways. When it is feasible, just the objects and detector areas that are really required are reconstructed, as opposed to all possible objects in an event.

2.3   Particle reconstruction

The reconstruction of particles in the CMS experiment is a complex process that involves a combination of experimental measurements and theoretical calculations. It requires sophisticated detectors and algorithms, as well as a large team of physicists and engineers working together to analyze the data.

Particles produced from collisions first meet the all-silicon tracker, which leaves their footmark as charged particles traverse through the detector, allowing us to measure their momentum. The tracker is surrounded by the ECAL in which electrons, positrons, and photons deposit their energy. The HCAL is designed to detect "hadrons". The tracker,

74

ECAL, and HCAL are inside the solenoidal magnetic coil with a magnetic field of 3.8 T. As the name indicates, CMS is also designed to measure muons precisely. Muons are detected and measured using the muon chamber. Neutrinos escape from CMS undetected and can be estimated indirectly from the "missing transverse energy" in the event. This simplified view is graphically summarized in Fig. 2.12, which displays a sketch of a transverse slice of the CMS detector. The data used for the studies in the thesis includes information from all parts of the detector.
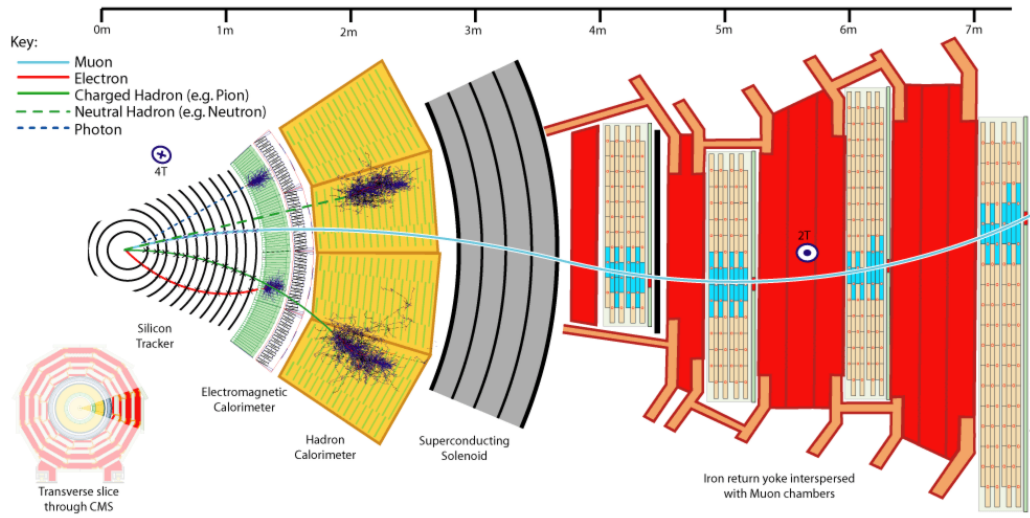
Figure 2.12: A slice of the CMS detector with a cross-sectional view showing different particles depositing their energies in the different detector parts. The figure is taken from [12]

2.3.1 Particle flow algorithm

The particle flow (PF) algorithm is a key component of the particle reconstruction process in the CMS experiment. CMS is well-suited to use the PF approach thanks to a combination of features including a robust magnetic field, finely-segmented subdetectors, and high hermeticity across the tracker, calorimeters, and muon system. This allows for effective identification of distinct types of objects and enables energy measurement to be driven by

the subdetectors with the most precise resolution. The goal of the PF algorithm is to reconstruct all the particles produced in a collision event by combining optimally the information from the different subdetectors in the detector. The PF algorithm requires a number of key ingredients to be efficient. These include:

- High-quality tracking: The tracking detectors must be able to accurately measure the momentum and trajectory of charged particles. This requires a combination of high spatial resolution, good track-finding algorithms, and excellent detector alignment.

- Robust clustering algorithms: The PF algorithm must be able to group together the energy deposits in the calorimeters that belong to the same particle. This requires sophisticated clustering algorithms that can handle a wide range of particle energies and angles.

- Accurate energy calibration: The energy measurements in the calorimeters must be calibrated to a high precision, in order to accurately reconstruct the energy of the particles.

- Robust particle identification algorithms: The PF algorithm must be able to distinguish between different types of particles, based on their energy deposits, track properties, and other characteristics. This requires sophisticated particle identification algorithms that can handle a wide range of particle energies and angles.

In simpler terms, the algorithm can be summarized as follows, with a more comprehensive explanation in Ref. [48]. First, the algorithm extends the tracks into the calorimeters, and if these tracks intersect with one or more clusters, the clusters are linked to the track. This combination of a track and its associated cluster(s) represents a charged hadron, and the individual components are no longer considered within the subsequent steps of the algorithm. Muons are identified in advance to ensure that their tracks do not contribute to the creation of

a charged hadron. The electrons are more difficult to deal with. This is primarily due to the frequent emission of Bremsstrahlung photons. As a result, a specialized track reconstruction method [49] becomes necessary, along with a dedicated approach to correctly attach photon clusters to the corresponding electron and prevent any double counting of energy. After processing all the tracks, any remaining clusters are identified as either photons (in the case of the ECAL) or neutral hadrons (in the HCAL). Once all the particle's energy deposits are associated, its identity can be determined, and information from the sub-detectors is combined to optimally calculate its four-momentum. In situations where the calibrated calorimeter energy of the clusters, which is essentially a linear combination of the ECAL and HCAL energy deposits associated with a track, exceeds the track momentum by more than one standard deviation, the excess energy is attributed to an overlapping neutral particle (either a photon or hadron). This overlapping particle is assigned an energy equivalent to the difference between the two measurements. The resulting particle list, consisting of charged hadrons, photons, neutral hadrons, electrons, and muons, serves multiple purposes. It is used for reconstructing jets, calculating the transverse energy imbalance (E_{miss}^{T}), identifying τ particles based on their decay products, and measuring the isolation levels of the particles. The following sections provide a detailed explanation of how individual particles are reconstructed in the subdetectors prior to their input into the PF algorithm.

2.3.2 Muons

Muons and other charged particles traverse the gaseous medium of a detector and ionize the gas creating an electrical signal in the electrode of the detector. The connected electronics then read out the electrical signal that along with the position of the signal is referred to as a "hits," which will be connected with the hits from other detectors to reconstruct the tracks of the particles. Different algorithms are used depending on different detector technologies to measure the precise location of each hit reconstructed from the electronic signals.

Track reconstruction

In the standard CMS reconstruction [11, 50, 51] procedure for pp collisions, the tracks are first separately reconstructed in the inner tracker (tracker track) and in the muon system (standalone muon track) and are then utilized as input for muon track reconstruction.

- *Standalone-muon tracks* are created by using data from muon subdetectors to gather all CSC, DT, and RPC signals along a muon trajectory using a Kalman-filter technique [52]

- *Tracker muon tracks* are built by propagating tracker tracks to the muon system with loose matching criteria to DT or CSC segments. The technique is called the "inside-out" approach.

- *Global muon tracks* are built "outside-in", i.e by propagating standalone-muon tracks to tracker tracks. The matching is done by comparing parameters of the two tracks propagated onto a common surface. A combined fit is performed using information from both the tracker track and standalone-muon track, using the Kalman filter technique.

About 99% of muons that are produced within the geometrical acceptance of the muon system are reconstructed either as a global muon track or as a tracker muon track. These reconstructed muons are further fed into the PF algorithm [53]. Muon candidates reconstructed using the standalone, global, or tracker muon tracks algorithms are subjected to a set of selection criteria by the PF algorithm.

Identification

In order to enable each analysis to fine-tune the ideal balance between efficiency and purity of signal, a set of variables is analyzed and selection criteria are specified. The selection

criteria are applied to the variables associated with the muon reconstruction, such as the number of hits per track (either in the inner tracker or in the muon system, or both), track fit χ^2 and the degree of matching between tracker tracks and standalone-muon tracks (for global muons). The method gives the values in the range between 0 and 1, with 1 denoting the maximum level of compatibility. The primary muon identification types employed in CMS physics analyses using these characteristics are:

- *Loose muon identification (ID)*: the goal is to identify prompt muons that originate at the primary vertex, muons from light and heavy flavor decays, and also maintain a low rate of being mistaken for charged hadrons as muons.

- *Medium muon ID*: is optimized for muons from heavy flavor decay and also for prompt muons. A medium muon is a loose muon with a tracker track that uses hits from more than 80% of the inner tracker layers it traverses.

- *Tight muon ID*: aims to prevent muons from hadronic shower and from in-flight decay. A tight muon is a loose muon with a tracker track using hits from at least six layers of the inner tracker, including at least one pixel hit. The muon must be reconstructed as both a global muon and a tracker muon. Additionally, it needs to be consistent with the primary vertex, with a transverse impact parameter $|dXY|$ less than 0.2 cm and a longitudinal impact parameter $|dz|$ less than 0.5 cm.

- *Soft muon ID*: Tuned for B-physics and quarkonia studies with low-p_T muons. A soft muon is a tracker muon that uses hits from at least six levels of the inner tracker, including at least one pixel hit, and meets the high purity flag of tracker track [51].

- *High momentum muon ID*: is designed for muons with p_T more than 200 GeV. A high momentum muon must be both a tracker muon and a global muon.

The standard method employed by CMS to calculate the muon momentum is the Tune-P algorithm [50]. Specifically, the Tune-P algorithm makes individual decisions for each muon, selecting among three fitting methods: the tracker-only fit, the Tracker-Plus-First-Muon-Station (TPFMS) fit, and Picky fits [50], which initiate from the hit list of the global-muon track. Initially, the algorithm uses the Picky fit and subsequently switches to the tracker-only fit if it provides a significantly better fit quality. Next, it compares the fit quality of the chosen track with that of TPFMS, and TPFMS is favored if it yields superior results. In the case of high-pT muons, Tune-P predominantly opts for the TPFMS and Picky algorithms, with roughly equal usage rates, while the tracker-only fit is employed in only a small percentage of events. This algorithm is validated using muons from pp collisions, cosmic ray muons, and Monte Carlo simulations generated using various misalignment scenarios. The comparison plots with different algorithms can be seen in the Ref. [50]. Furthermore, to discriminate between prompt muons and those from weak decays within jets, the isolation of a muon is determined relative to its p_T by adding the energy in the geometric cone, $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, around the muon.

2.3.3   Jets

In collider experiments, a quark or gluon hadronizes into a narrow cone of hadrons and other particles known as a "jet". Due to quantum chromodynamics (QCD) confinement, which only permits colorless states, particles with a color charge, such as quarks, cannot exist in free form. Each fragment that breaks off from a color-charged item takes part of the color charge with it. These fragments create further colored objects around them to form colorless objects. Since all of the particles have a tendency to go in the same direction, producing a narrow jet of particles, the group of these objects is known as a jet. Jets are observed and studied in particle detectors to determine the characteristics of the original quarks. The tracker accurately measures the energy of charged hadrons, which account

for around 60% of the average jet energy, while the ECAL properly measures the energy of photons, which account for another 30%. The HCAL measures the remaining 10% of neutral hadron energy. In the HCAL, readout cells are placed in a tower pattern in η, ϕ space, projective to the nominal interaction point. The segmentation of the cells in the barrel region is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ and they get gradually bigger in the endcap and forward regions. Since the ECAL granularity is much better than the HCAL, ECAL plus HCAL towers are formed by addition of ECAL signals in η, ϕ bins corresponding to individual HCAL cells. Calorimeter noise contributions can have a considerable influence on the reconstruction of low-E_T jets; careful selection of these inputs is necessary to provide the best performance of higher-level objects reconstructed from calorimeter towers. Many studies has been performed on strategies for reducing noise and pile-up contributions to jet energy; more details can be found in reference [54]. Jets in CMS are identified using the anti-kT clustering technique [55] with a distance value of R = 0.4. The algorithm groups these potential clusters of energy deposits starting with the two closest objects and defines a distance metric based on the p_T and angular parameters of the candidate clusters of energy deposits. The process ends when one object is closest to the beampipe, at which point the group of objects is known as a jet. The calibration of jet energy aims to establish a consistent relationship between the energy measurement obtained for a detector jet and the energy of the corresponding true particle jet. This true particle jet is formed by clustering, using the same algorithm employed for detector jets, all stable particles originating from the parton fragmentation process, as well as particles from the underlying event activity. A correction in the form of a multiplicative factor is applied to each component of the initial detector jet four-momentum vector [56].

2.3.4 Missing transverse energy

Neutrinos and other neutral weakly interacting particles typically escape from collider detectors without leaving any direct signature in the detector components. An imbalance in total momentum must be used to infer the existence of such particles. Missing transverse momentum, here referred as $\vec{E}_T^{\mathrm{miss}}$, is the vector momentum imbalance in the plane perpendicular to the direction of the beam that is particularly useful in hadron colliders. It is crucial observable in hadron collider experiments because it helps to account for undetected particles, search for new physics beyond the Standard Model, study neutrinos, calibrate detectors, and make precise measurements of known processes. $\vec{E}_T^{\mathrm{miss}}$ is typically the negative of the vector sum of all final state particles' transverse momenta that have been reconstructed in the detector. Three distinct algorithms have been developed by CMS to reconstruct $\vec{E}_T^{\mathrm{miss}}$: (a) PF E_T^{miss}, which uses a complete particle-flow method [48]; (b) Calo E_T^{miss}, which relies on calorimeter energies and the geometry of the calorimeter towers and does not include contributions from muons; and (c) TC E_T^{miss}, which corrects Calo E_T^{miss} by adding tracks that were reconstructed in the inner tracker after adjusting for the expected energy depositions of the tracks in the calorimeter. From the reconstructed PF particles, PF E_T^{miss} is determined. Different instrumental sources and techniques for determining E_T^{miss} occasionally can cause abnormal E_T^{miss} readings. Sometimes, anomalous signals related to particles striking the transducers are captured by the CMS ECAL and HCAL. Rare random discharges of the readout detectors can also result in anomalous signals in the HCAL. Some of these impacts have previously been noted when collecting cosmic and test beam data.

2.3.5 Electron and Photon

Nearly all of the energy that electrons and photons carry are deposited in the ECAL, and the shower spreads over several crystals. Around 94% and 97% of the incoming energy

of a single electron or photon is contained in a matrix of 3×3 and 5×5 crystals, respectively. The electrons also produce hits in the tracker layer as it traverses through. The signal pulse in the ECAL crystals is fitted with multiple template functions to subtract the contribution from out-of-time pileup. Throughout the entire LHC Run 2 data collection period, this method was employed for both the HLT and offline event reconstruction. The tracker detector material in front of the ECAL may interact with the incoming electron or photon as it travels through it, causing an electron to release bremsstrahlung photons or a photon to change into an electron-positron pair. This means that by the time the electron or photon reaches the ECAL, it could not just be one particle, but rather a shower of several electrons and photons. In order to obtain the energy of the primary electron or photon, a specialized algorithm is used to merge the clusters from individual particles into a single entity. In addition, the electron loses momentum as it emits bremsstrahlung radiation, altering its trajectory in the magnetic field. To estimate the track parameters for electrons, a specialized tracking technique built on the Gaussian sum filter (GSF) is used [49]. Electron and photon reconstruction in CMS is fully integrated into the PF framework. The phases of reconstruction are briefly outlined below.

- The first step of the energy reconstruction procedure is the formation of ECAL clusters by assembling crystals with energies over a certain threshold.

- To include photon conversions and bremsstrahlung losses, superclusters (SC) are formed by combining ECAL clusters within a predetermined geometric region (the "window") surrounding the seed cluster. "Superclustering" is the term used to describe this process.

- The GSF, is used for electrons to estimate the track parameters. The GSF tracking step is seeded using trajectory seeds in the pixel detector that are consistent with the SC location and the trajectory of an electron.

- A special algorithm [57] is used to find the generic tracks that are most likely to result from photons changing into e^+e^- pairs.

- The PF algorithm that connects the components together into blocks of particles imports information connected to ECAL clusters, SCs, GSF tracks, and generic tracks connected to electrons, as well as conversion tracks and related clusters.

- Starting from either a GSF track or a SC, these blocks are resolved into electron and photon (e and γ) objects, respectively. There is no distinction between electron and photon candidates at this time.

- Based on loose selection criteria [58], electron or photon objects are constructed from refined SCs. If an accompanying GSF track exists, the objects that passes the selection is designated as electron; if not, they are designated as photons.

- Finally, in the PF framework, a stricter selection is made to these e/γ objects to determine if they should be regarded as electrons or isolated photons.

2.3.6 Vertex

Vertex is the specific spatial point where the two proton initially collide (primary vertex) or where they interact or decay after the collision (secondary vertex). Vertex finding and vertex fitting are often the first two steps in vertex reconstruction. Tracks are grouped into vertex candidates during vertex finding. The choice of vertex finding algorithm can vary significantly depending on the specific physics scenario: identifying primary or secondary vertices, reconstructing exclusive decays, and more. Vertex fitting plays a crucial role in determining the most accurate estimation of vertex parameters such as vertex position, covariance matrix, and track parameters for a given set of tracks. Additionally, indicators of the fit quality, such as total χ^2, the number of degrees of freedom, or track weights, are taken

into consideration. The well-known Kalman filter [52] is the vertex fitting method that is most frequently utilized. It is theoretically equal to a global least-squares minimization, which is the best estimator when the measurements are Gaussian and the fitted parameters depend linearly on those measurements. The vertex and the generated track-to-track co-variance matrices can both be used as constraints in the filter in order to obtain an improved estimate of the track momenta. To get a better approximation of the track momenta, the filter might be constrained by the vertex and the derived track-to-track covariance matrices. The trimmed Kalman fitter is a standard, reliable version of the Kalman vertex fitter that starts with the least compatible track and removes each track from the vertex one at a time until the vertex is fit [59].

Chapter 3

Search for a Higgs production associated with a single top quark

The large Yukawa coupling of the Higgs boson in the standard model to the top quark y_t^{SM} is an important property. In order to understand electroweak symmetry breaking and to evaluate hypotheses that go beyond the standard model (BSM), measurements of y_t is particularly crucial. The value of y_t are indirectly tested by measurements sensitive to ggH, the dominant Higgs boson production process at LHC, in which the production process heavily relies on loop diagrams involving top quarks. Additionally, the y_t can also be probed through the decay of Higgs bosons to two photons, H$\rightarrow \gamma\gamma$, as the decay width also involves loop diagram with a top quark. However, y_t can be directly measured in the production of top - anti-top quark pairs, $t\bar{t}$, in associated with a Higgs boson. The production of a Higgs boson in association with a single top quark is also sensitive to y_t. Within the SM, it is assumed that the relative sign of the Higgs coupling to bosons and fermions is positive. As a result, a destructive interference occurs between the t-channel diagram, where Higgs bosons are emitted from the top quark, and the W boson lines. The Feynman diagram of the production process is displayed in the Fig 3.1. The single top quark plus Higgs production cross section would significantly increase if the Higgs boson were to be negatively coupled to fermions as opposed to bosons. The final state of this production process consists of a quark, a top quark and a Higgs boson. The top quark decays to a bottom quark and W boson with the W boson further decaying to either hadrons or leptons. In this analysis we focus on the final state where Higgs boson decay to a pair of photons and the W bosons decay to a lepton and a neutrino. This study analyzes the data collected by the CMS experiment
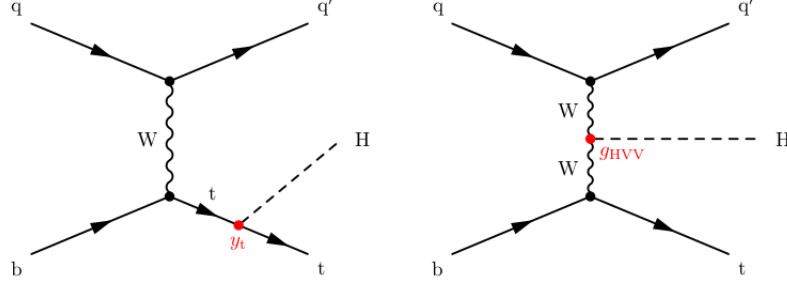
Figure 3.1: Dominant Feynman diagrams for the production of a Higgs boson associated with a single top quark, where the Higgs boson is emitted from the heavier legs of top quark and the W boson lines.

between the years 2016 and 2018. Within the SM framework, the decay process $H \to \gamma\gamma$ decay exhibits a relatively low branching fraction of approximately 0.23% when the Higgs boson mass (m_H) is around 125 GeV. However, its unique final-state topology involving two accurately reconstructed photons allows for a narrow invariant mass ($m_{\gamma\gamma}$) peak, providing effective differentiation from background processes. Therefore, the $H \to \gamma\gamma$ is the one of the key channels for precise measurements of Higgs boson properties. Moreover, it is among the few decay channels that possess sensitivity to all major Higgs boson production modes. The analysis is structured to facilitate measurements within the simplified template cross section (STXS) framework [60]. This is a consistent approach to conducting precise measurements of the Higgs boson by combining all the decay channels of Higgs bosons. The primary goals of this framework are to reduce the direct effects of standard model predictions on the results, minimize the theory dependence of Higgs boson measurements, and enable access to kinematic regions likely to be affected by BSM physics. Additionally, this approach enables the employment of sophisticated analytic methods to maximize sensitivity. The SM simulation that was used to simulate the experimental acceptance of the signal processes is still a factor in the results in the STXS framework, and BSM scenarios might change it. It is preferable to reduce theory dependence since it makes measurements easier to reinterpret and less affected by changes in theoretical predictions, extending their useful-

ness. This makes it possible to measure the production of the Higgs boson in all of its major production modes: ggH, VBF, VH, ttH, and tH. However, this thesis focuses only on the production of a Higgs boson associated with a single top quark where the final state consists of electrons/muons, quarks, photons, and neutrinos, referred to as the "leptonic tHq" final state. They all (except for the neutrino) leave their signature in different parts of the CMS detector. The reconstruction of all the particles in the CMS detector is briefly discussed in the previous chapter.

3.1   Analysis strategy

This study focuses on searching for the production of a Higgs boson in association with a single top quark, where the Higgs boson subsequently decays into two photons. The accuracy of photon reconstruction and their energy determination plays a crucial role in this analysis. The analysis started with the selection of photon (Sec. 3.3) using a diphoton event trigger (Sec. 3.4). A photon identification MVA (Sec. 3.5) is specifically trained to distinguish genuine ("prompt") photons from jets that mimic a photon signature. Additionally, the precise determination of the primary vertex, from which the two photons originate, is crucial to the analysis. The central discriminating factor in this analysis is the Higgs invariant mass (or diphoton invariant mass, $m_{\gamma\gamma}$). Two mass regions are defined accordingly:

- Signal window: $m_{\gamma\gamma} \in [115, 135]$ GeV

- Sideband region: $m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV,

where the signal window is blinded throughout the development of the analysis. The analysis preselection focuses on the leptonic decays of the top quark and the following steps are performed:

1. Train binary multivariate (MVA) classification algorithms using simulated tHq samples as the signal and simulation of related standard model processes as the back-

ground. and a more comprehensive explanation can be found in Sec. 3.7.1 The training specifications are as follows:

- A single MVA is trained, including only ttH as the background.

- One MVA is trained to suppress SM non-resonant backgrounds (NRB).

2. Define a signal region by applying rectangular cuts on the MVA scores. These cuts are determined to maximize the significance, as discussed in Sec. 3.8.1

3. Parametric signal and background models are constructed by fitting the $m_{\gamma\gamma}$ distribution in each category.

- The tHq signal processes and the SM Higgs background processes are modeled with a sum of Gaussian functions.

- The non-resonant background is modeled using data from the $m_{\gamma\gamma}$ sideband, using various functional forms. The choice of the functional form is treated as a discrete nuisance parameter, following the discrete profiling method [61].

Specifically, the fitting procedure is carried out separately for each year and channel to account for the year dependent $m_{\gamma\gamma}$ resolution. The signal strength modifier for the Higgs boson production associated with the single top quark mode is measured from simultaneous fit with all other Higgs production modes. Finally, the best fit cross sections along with their corresponding 68% confidence intervals are presented.

3.2 Data samples and simulated events

The analysis is performed using the data from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV, recorded during the periods of 2016, 2017, and 2018. The integrated luminosities corresponding to these years are 35.9, 41.5, and 59.4 fb^{-1}, respectively.

During the online selection of events from 2016 (2017 and 2018), a diphoton high-level trigger is used, which applies asymmetric thresholds on the transverse momenta (p_T) of the photons: 30 (30) and 18 (22) GeV with the addition of calorimetric selection factors such as the electromagnetic shower-shape, the isolation of the photon candidate, and the distribution of its hadronic and electromagnetic energy deposits. Tab. **??** contains a list of the data samples collected between the year 2016 and 2018.

The simulated samples of each Higgs boson production mechanism are generated at next-to-leading order accuracy [62] in perturbative QCD using the MG5_aMC@NLO (version 2.4.2) Monte Carlo generator. These events are generated with Higgs boson masses (m_H) set at 120, 125, and 130 GeV. The total cross sections and branching fractions, as recommended by the LHC Higgs boson cross section working group [23], have been adopted for the SM Hiigs boson production processes. For parton showering and hadronization, all parton-level samples are interfaced with Pythia version 8.226 (8.230) [63], with the CUETP8M [64] (CP5 [65]) tune being used for the simulation of the 2016 (2017 and 2018) data. The functions that describes the probability distribution of finding a particular type of parton (quark or gluon) within a hadron (Parton distribution functions) are taken from the NNPDF 3.0 [66] (3.1 [67]) set, when simulating 2016 (2017 and 2018) data. The Sherpa (version 2.2.4) [68] generator is used to generate the diphoton backgrounds. The γ+jets and jet+jet backgrounds are simulated at leading order with Pythia 8.205 [63], after applying a filter at generator level to enrich the production of jets with a high electromagnetic activity. The background processes, $t + X$, $t\bar{t} + X$, $W\gamma$, and $Z\gamma$, are simulated with MG5_aMC@NLO , whereas diboson events are produced at the leading order with Pythia. The nonresonant diphoton samples are simulated with Sherpa 2.2.4 [68] which includes tree-level processes with up to three additional partons, as well as box diagrams. The GEANT4 package [69] is used to simulate the response of the CMS detector. The multiple proton-proton interactions occurring during each bunch crossing is also included in the simulation. These interactions

90

can happen either during the actual bunch crossing (in-time pileup) or with neighbouring bunches before or after (out-of-time pileup), which is also taken into account during the simulation. Only out-of-time pileup within a range of $[-12, +3]$ bunch crossings around the nominal bunch crossing is simulated, in which the effects on the observables reconstructed in the detector are most relevant. The number of interaction vertices observed in the actual data is replicated in the simulated events by applying weights. The 2016 dataset had an average of 23 interactions per bunch crossing, while the 2017 and 2018 datasets had an average of 32 interactions.

3.3 Photon preselection

The sensitivity of this analysis depends greatly on how efficiently photons can be reconstructed with a precise energies determination. In this section, the procedures used for reconstructing photon energy and the photon preselection criteria are elaborated in detail. The photon reconstruction in the CMS detector is briefly covered in the section 2.3.5. Here is a brief overview of standard photon reconstruction variables used for selecting photons/electrons and also used as input features in the photon identification training within the analysis:

Shower-shape variables:

- $E_{2\times2}/E_{5\times5}$: The energy ratio is determined by comparing the energy in a 2×2 crystal matrix, which includes the first and second most energetic crystals, to the energy in a 5×5 crystal matrix centered around the seed crystal of the supercluster. Here, the supercluster is defined as a group of adjacent crystals that have detected energy deposits from a single particle.

- $cov_{i\eta i\phi}$: The covariance of the η and ϕ of the individual crystal is expressed in terms

91

of crystal cells within the 5×5 crystals centered on the supercluster seed crystal.

- $\sigma_{\eta\eta}$: The measurement of the lateral spread of the electromagnetic shower in terms of crystal cells.

- R_9: $E_{3\times3}/E_{SC}$, where $E_{3\times3}$ is the energy sum of the 3 by 3 crystals surrounding the supercluster seed crystal and E_{SC} is the energy sum of the supercluster.

- σ_η: The logarithmic energy weighted standard deviation of single crystal η within the supercluster.

- σ_ϕ: The logarithmic energy weighted standard deviation of single crystal ϕ within the supercluster.

- *Preshower* σ_{RR}: The standard deviation of the spread of the electromagnetic shower in the x and y planes of the preshower, which is only defined in the endcap region.

Isolation variables:

- *PF Photon iso (I_{ph})*: defined as the transverse energy sum associated with all particles identified as photons by the PF algorithm, around the photon candidate, falling inside a cone size R=0.3.

- *PF Charged iso*: defined as the transverse energy sum associated with all particles identified as charged hadrons by the particle-flow algorithm falling inside a cone size R=0.3 around the photon candidate direction. The measurement is performed with respect to the selected vertex.

- *Charge hadron iso*: the energy in HCAL from charged particles inside a cone (R = 0.3) around the photon candidate.

- *H/E*: the energy collected by the HCAL towers within a cone of R=0.15 centered on the supercluster position, divided by the supercluster energy.

- *Tracker isolation in a hollow cone (I_{tk})*: The variable represents the sum of transverse momenta from all tracks falling within a cone of radius R=0.3 around the direction of the photon candidate. Tracks falling in an inner cone of size R=0.04 are not included in the p_T sum.

The energy of a photon is calculated by adding the energy depositions in all the crystals in the supercluster, calibrated, and compensated for response-time differences [58]. The photon energy is corrected for the imperfect containment of the electromagnetic shower and the energy losses from converted photons. The correction is computed with a multivariate regression technique trained on simulated photons, which simultaneously estimates the energy of the photon and its uncertainty.

After applying the correction derived from simulation, there is still some disagreement between the data and the simulation. To address this, a series of additional corrections are implemented to enhance the agreement between the data and simulation. This is achieved using $Z \rightarrow ee$ events where the electrons are reconstructed as photons ignoring the tracks. The first correction involves compensating for any residual drift in the energy scale over time in bins corresponding approximately to the duration of one LHC fill. The second step focuses on adjusting the energy scale in the data and the energy resolution in the simulation. Correction factors are determined to align the mean of the dielectron mass spectrum in the data with the expected value from the simulation. Additionally, the resolution in the simulation is smeared to match the observed resolution in the data. These corrections are simultaneously derived in bins of $|\eta|$ and R_9. For more detailed information regarding this procedure, please refer to Ref. [24].

Figure 3.2 presents the comparisons between data and simulation once all the necessary

corrections have been applied. It includes two scenarios: one where both electrons are reconstructed in the ECAL barrel and another where they are reconstructed in the ECAL endcaps. In both cases, the dielectron invariant mass spectra for the data and simulation are found to be compatible within the uncertainties. After the photon energy correction is
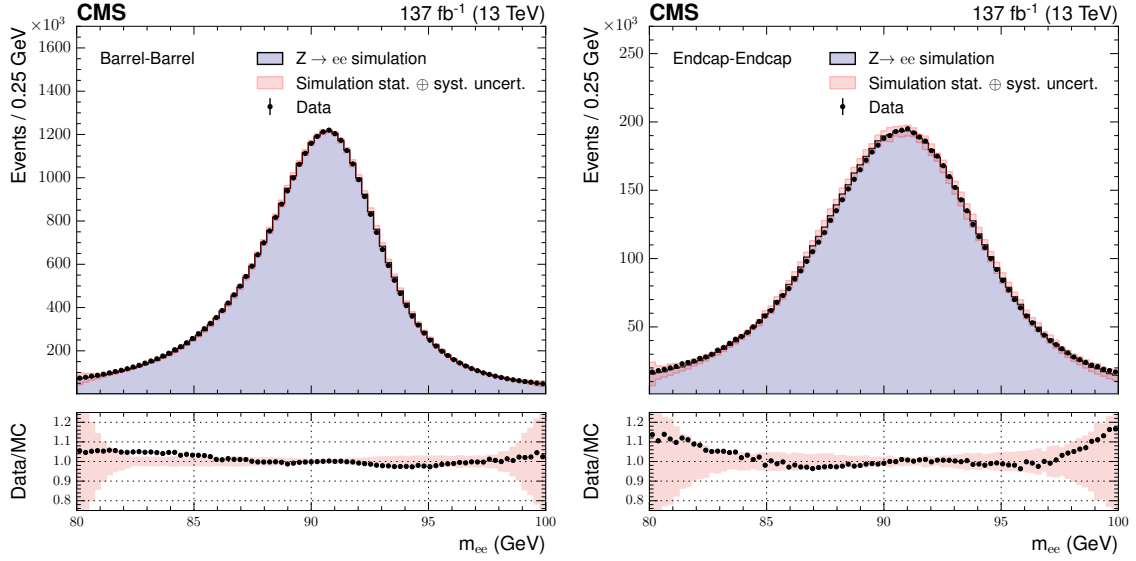
Figure 3.2: The dielectron invariant mass spectra for $Z \rightarrow ee$ events, where electrons are reconstructed as photons, are compared between data (black points) and simulation (blue histogram) after applying energy scale corrections to data and energy smearing to the simulation. The pink band illustrates the statistical and systematic uncertainty on the simulation. The comparison is presented separately for events where both electrons are reconstructed in the ECAL barrel (left) and in the ECAL endcaps (right). The lower panels display the ratio of the data to the MC simulation in black points, with the uncertainty on the ratio demonstrated by the pink band.

applied, the photon candidates have to pass a preselection criterion before being used to form diphoton candidates. This preselection criterion involves imposing requirements on the photon kinematic, shower-shape, and isolation variables. These requirements are set to be at least as stringent as those applied in the trigger. The selection criteria on standard photon reconstruction variables are tabulated in Tab. 3.1:

Table 3.1: Selection criterion on standard photon reconstruction variables.

	R_9	H/E	$\sigma_{\eta\eta}$	I_{ph} (GeV)	I_{tk} (GeV)
Barrel	[0.50, 0.85]	<0.08	<0.015	< 4.0	< 6.0
	>0.85	<0.08	—	—	—
Endcap	[0.80, 0.90]	<0.08	<0.035	< 4.0	< 6.0
	>0.90	<0.08	—	—	—

3.4 Diphoton triggers

Every LHC proton-proton bunch crossing with a potential to produce high-energy collisions in the CMS experiment is evaluated by the trigger system. The events entering an analysis must first pass a hardware level trigger followed by a software level trigger decision. Events are selected using a diphoton high-level trigger with asymmetric photon p_T thresholds of 30 (30) and 18 (22) GeV in 2016 (2017 and 2018) data. A calorimetric selection is applied at trigger level, based on the shape of the electromagnetic shower, the isolation of the photon candidate, and H/E of the shower. The HLT path used in this analysis is seeded by at least one hardware level 1 (L1) electromagnetic candidate. Given an L1 seed, the ECAL clustering algorithm is performed by the HLT from the detector readout units overlapping a rectangle centered on the L1 candidate and the HLT processing time is kept short by clustering only around the L1 seed in a first step. The requirements of the HLT diphoton path are then applied. Since any photon candidate in the event can produce an L1 seed, the overall efficiency for a diphoton path seeded by a single L1 seed is higher than requiring two L1 seeds. Due to bandwidth limitations at the L1, the lowest transverse energies of unprescaled single electromagnetic L1 candidates is 40 GeV; however if the EM L1 candidate is found to be isolated the crossponding threshold is 30-32GeV depending on the data taking period. This creates an inefficiency at the lowest transverse energy of the analysis selection. It is possible that events which would otherwise pass the analysis preselection

will fail the diphoton trigger decision. In order to estimate this effect, the trigger efficiency is computed and defined as the fraction of events passing the analysis preselection that also pass the trigger decision. To study the efficiency of the trigger selection, the tag-and-probe technique is applied on a sample of $Z \rightarrow ee$ events, where one electron passing a tight identification (tag) and one passing a loose identification (probe). The efficiency is computed separately for the pixel detector seeded and unseeded electrons. These two efficiencies are then multiplied to obtain a per-event (diphoton) correction, which is applied to the simulation. This trigger efficiency is measured using data collected at the LHC by CMS and it is applied to simulation samples.

The HLT selection is the same in all three years (apart from the E_T cut on the unseeded leg, at 18 GeV in 2016 and 22 GeV in 2017 and 2018) and can be split into general, isolation plus calorimeter identification (CaloId+Iso), and high R_9 requirements. The general criteria are applied to all objects, while either the CaloId+Iso or the high R_9 selection is required. The variables used are:

- General variables: H/E, R_9, $m_{\gamma\gamma}$, and E_T of diphoton.

- Variables used in CaloId + Iso filters: $\sigma_{i\eta i\eta}$, ρ-corrected ECAL particle flow cluster isolation (I_{ph}), tracker isolation in a hollow cone (I_{tk}).

Every L1-seeded leg of the HLT is required to have $E_T > 30$ GeV. The cluster must be within the region $|\eta| < 2.5$ and have an $R_9 > 0.5$ (0.8) in EB (EE). Additionally, a requirement of H/E < 0.12 (0.1) EB (EE) is applied. The clusters are then required to have $R_9 > 0.85$ (0.9) in the EB (EE). Those passing the R_9 selection continue to the unseeded step and are not required to pass the CaloId + Iso filters. Those failing the R_9 selection must pass the CaloId + Iso requirements in order to continue to the unseeded step. These consist of $\sigma_{i\eta i\eta} < 0.015(0.035)$ EB(EE) and ECAL isolation $I_{ph}^{HLT} < 6.0 + 0.012 E_T$. Finally, the mass of the diphoton object is required to be above 90 GeV, in order to reduce the trigger

rate and respect the limited bandwidth.

3.5   Photon identification

After applying the preselection criteria, a BDT is used to make the final photon selection This photon identification BDT (ID BDT) is specifically trained to distinguish genuine ("prompt") photons from jets that mimic a photon signature. The BDT is trained using the simulated sample of $\gamma +$ jets events, where prompt photons are treated as the signal and jets are considered as the background. Various input variables, including shower-shape variables, isolation variables, photon energy and η, as well as global event variables sensitive to pileup effects such as the median energy density per unit area ρ [70], are used in the ID BDT. The inputs of BDT, specifically shower-shape and isolation variables, are adjusted to match the data by exploring a chained quantile regression (CQR) method [71]. This method is used to enhance the agreement in the output of the ID BDT between data and simulation, thereby reducing the associated systematic uncertainty. The corrections are determined by using an unbiased sample of electrons obtained by applying a tag-and-probe technique in $Z \rightarrow ee$ events. The CQR method consists of a series of BDTs that estimate the cumulative distribution function (CDF) for a specific input variable. These predictions are based on three electron kinematic variables (p_T, η, ϕ) and ρ. By extracting CDFs from both data and simulated events using the CQR, correction factors can be derived and applied to individual simulated electrons. These correction factors adjust the simulated shower-shape and isolation variable's CDF to align with the observed distribution in data. A systematic uncertainty is incorporates in the analysis related to this corrections. To estimate this uncertainty, the corrections are recalculated using equally sized subsets of the $Z \rightarrow ee$ events that are used for training. The magnitude of the uncertainty is determined by the standard deviation of the differences in the corrected ID BDT output scores obtained from the two

training subsets on an event-by-event basis.

Figure 3.3 (left) displays the photon ID BDT distribution for the least scoring photon in signal events and various background components. Additionally, a comparison between data and simulation is presented using $Z \to ee$ events, where electrons are reconstructed as photons. The $Z \to ee$ events serve as a suitable validation tool due to the similarity in detector signatures and reconstruction procedures for electrons and photons. In this context, electrons being reconstructed as photons implies that track information is disregarded, and the energy determination follows the algorithm and corrections specific to photons rather than electrons. Additionally, the distribution of the photon ID BDT is validated using photons in $Z \to \mu\mu\gamma$ events as shown in Fig 3.3 (right), where data and simulation are found to agree within uncertainties.

3.6 Vertex Identification

The accurate determination of the primary vertex, from which the two photons originate, plays a crucial role in achieving a high-resolution $m_{\gamma\gamma}$ measurement. When the position along the beam axis (z) of the interaction producing the diphoton is known with a precision of approximately 1 cm, the $m_{\gamma\gamma}$ resolution is primarily determined by the photon energy resolution. In the data collected from 2016 to 2018, the root mean square (RMS) of the distribution in z for the reconstructed vertices ranges from 3.4 to 3.6 cm. To ensure consistency between simulation and data, the distribution of the primary vertex in each year's simulation is adjusted through reweighting to match the distribution observed in the data. The assignment of the diphoton vertex is accomplished using a vertex identification BDT. This BDT takes into account various track-related observables associated with the recoiling tracks against the diphoton system [70]. It is trained specifically on simulated ggH events and is capable of identifying a single vertex in each event. The performance of the vertex
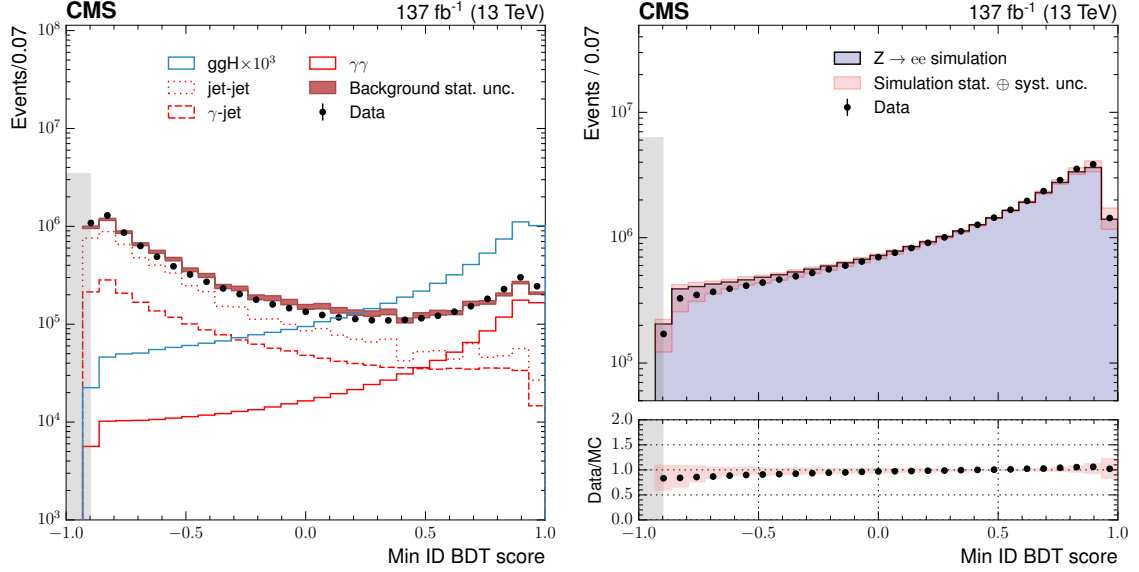
Figure 3.3: The distribution of the photon identification BDT score of the lowest scoring photon is shown for diphoton pairs with $100 < m_{\gamma\gamma} < 180$ GeV (left). The black points represent data events that pass the preselection criteria, while the red band illustrates the simulated background events. Additionally, histograms for various components of the simulated background are displayed. The blue histogram corresponds to simulated Higgs boson signal events. On the right plot, the same distribution is presented for $Z \rightarrow$ ee events in both data and simulation, where the electrons are reconstructed as photons. The statistical and systematic uncertainty in simulation is demonstrated by the pink band. The gray shaded area represents the range of identification BDT scores from -1 to -0.9, and photons with scores below -0.9 are eliminated from the analysis right at the initial stages.

identification BDT is assessed by validating it with $Z \rightarrow \mu^+\mu^-$ events. In this validation process, to simulate a diphoton system, the vertices are refitted while excluding the muon tracks. The efficiency of correctly assigning the vertex is evaluated as a function of the dimuon p_T and is displayed in Fig. 3.4 (left plot). The efficiency of correctly assigning the diphoton vertex within a 1 cm range of the true vertex in simulated $H \rightarrow \gamma\gamma$ events is approximately 79%.

A secondary multivariate discriminant, known as the vertex probability BDT, is used to estimate the likelihood of the vertex chosen by the vertex identification BDT being within a 1 cm range of the originating diphoton vertex. The vertex probability BDT is trained on sim-

ulated H $\rightarrow \gamma\gamma$ events using various input variables associated with the event vertices, their corresponding vertex identification BDT scores, the number of photons with associated conversion tracks, and the p_T of the diphoton system. Figure 3.4 (right plot) demonstrates the agreement between the average vertex probability and the vertex efficiency observed in the simulation.
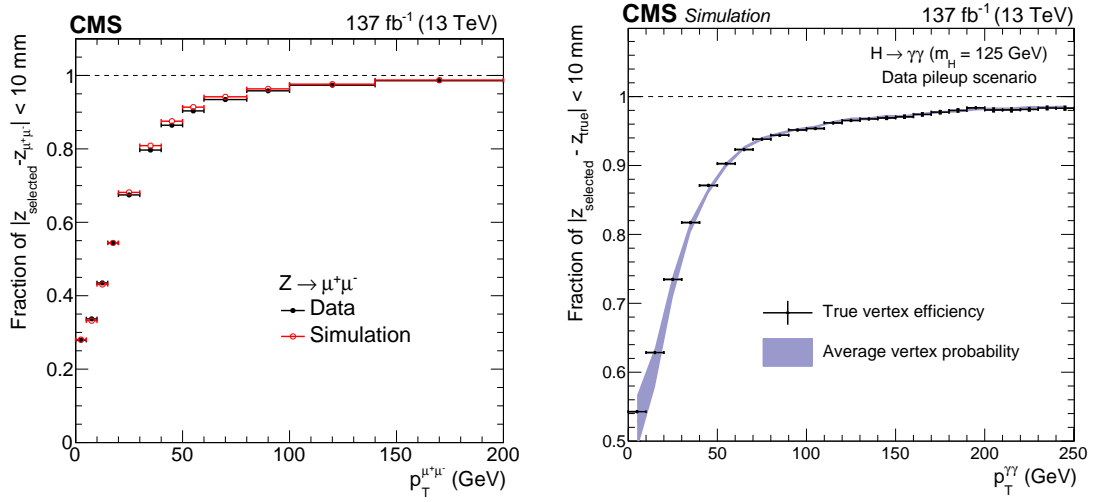
Figure 3.4: The left plot illustrates the validation process of the H $\rightarrow \gamma\gamma$ vertex identification algorithm on $Z \rightarrow \mu^+\mu^-$ events. In this validation process, the muon tracks are excluded while performing the event reconstruction. This enables a comparison between the fraction of events with accurately assigned vertices estimated using simulation and the data, as a function of the transverse momentum of the dimuon system, serving as a validation of the vertex identification BDT. The simulated events are appropriately weighted to match the pileup and vertex distribution along the beam axis observed in data. The right plot demonstrates that the true vertex efficiency in simulated events agrees with the average vertex probability to be within 1 cm of the true vertex.

3.7 Preselection for leptonic tHq

The event selection for this analysis requires that the two leading preselected photon candidates to have p_T (leading photon)$> m_{\gamma\gamma}/3$ and p_T (subleading photon)$> m_{\gamma\gamma}/4$, respectively, with an invariant mass in the range $100 < m_{\gamma\gamma} < 180$ GeV. Both photons are

required to have a photon identification BDT score of at least -0.7. Other criteria applied to the associated objects (electron, muon, jets) in the events are mentioned below. This analysis requires the presence of at least one electron or muon satisfying with the following criteria. Tau leptons are not considered in this analysis.

Muons: The identification of muon follows the standard CMS recommendations as briefly discussed in Sec. 2.3.2. Muons are required to pass the tight selection, to have a p_T of at least 5 GeV, and $|\eta| < 2.4$. Moreover, the isolation of each muon is required to be no greater than 0.25 (loose working point for isolation). For each muon, the angular distances (ΔR) between it and any photon or jet must be greater than 0.4.

Electrons: The identification of electron is based on the standard CMS recommendations as briefly discussed in Sec. 2.3.5. Electrons are required to pass the loose selections [58], including isolation to have p_T of at least 5 GeV, and $|\eta| < 2.4$. Candidates in the gap between the ECAL barrel and endcap, $1.4442 < |\eta| < 1.566$, are removed. In addition, for each electron, the angular distance (ΔR) between it and any photon or jet must be greater than 0.4. The electrons are also required to not have an invariant mass with either photon that would be within 5 GeV of Z boson mass ($|m_{e\gamma} - m_Z| > 5$ GeV).

Jets: We follow the standard recommendations for jets, following the reconstruction procedure mentioned briefly in Sec. 2.3.3. The reconstruction of jets is based on the anti-kT clustering algorithm [55] with a distance parameter of 0.4 (AK4). The jets originate from PF candidates; however the charged hadrons that originate from non-primary vertices are subtracted from the jet. The jet energy corrections and energy smearing are applied according to the recipes provided by the Ref. [72]. In addition, tight pileup jet identification [72] is used for 2017 and 2018, while loose identification is used for 2016, in order to maintain a consistent efficiency for reconstructed jets for all years. Each jet in the analysis satisfies the following criteria:

- $|\eta| < 4.5$

- $p_T > 25\,\mathrm{GeV}$

- $\Delta R(\text{jet, leading photon}) > 0.4$

- $\Delta R(\text{jet, subleading photon}) > 0.4$

- $\Delta R(\text{jet, lepton}) > 0.4$

- At least one jet in an event

Among the jets in the event, the one with the highest $|\eta|$ value is designated as the forward jet.

b-jets: The b-jets selection starts from the set of jets selected with the mentioned jet selection. Jets originating from the hadronization of b-quarks are tagged using the DeepCSV algorithm [73]. The loose working point [74] of the DeepCSV algorithm is used to identify b-jets. Additionally, the b-jets are required to be in the region ($|\eta| < 2.5$). To eliminate any potential overlap between the b-jet originating from the top quark decay and the forward jet within the event, the highest eta jet satisfying the b-jet criterion is removed from the collection of b-jets. Ultimately, the event is required to fulfill the below b-jet criterion.

- At least one such tagged jet is required

Missing E_T: The missing transverse momentum vector $\vec{p_T}^{\mathrm{miss}}$ is computed as the negative vector sum of the p_T of all the PF candidates in the event. The magnitude of $\vec{p_T}^{\mathrm{miss}}$ is denoted as MET. The $\vec{p_T}^{\mathrm{miss}}$ is modified to account for corrections to the p_T scale of the reconstructed objects in the event.

The event and the object selections mentioned aboove are summarized in the table 3.6

Given that the Higgs boson is radiated from one of the heavier legs in the t-channel process, it is anticipated that the light quark will have high rapidity. This characteristic

Objects		Object selection	Event selection
Diphoton		$p_T\ (\gamma_1)> m_{\gamma\gamma}/3$ and $p_T\ (\gamma_2)> m_{\gamma\gamma}/4$ γ_1 and γ_2 ID MVA > -0.7 $100\ \text{GeV} < \text{m}_{\gamma\gamma} < 180\ \text{GeV}$	≥ 0
Leptons (e/μ)		$|M(ee/\mu\mu) - M(Z)_{\text{PDG}}|> 5\ \text{GeV},$ Loose(Tight) cut based $e(\mu)$ ID [58]([11]) $p_T > 10\ \text{GeV}$ $\Delta R(e/\mu$ and $\gamma_{1/2}) > 0.4$	≥ 0
	e	$|\eta| < 2.4$ excluding of $[1.4442, 1.566]$	
	μ	$|\eta| < 2.4$ Relative muon PF isolation < 0.25	
Jets		$p_T > 25\ \text{GeV}$ and $|\eta| < 4.5$ $\Delta R(\text{Jet}$ and $\gamma_1/\gamma_2/\text{leptons}) > 0.4$	≥ 1
	b-jet	loose working point of b-jet identification training [74] $|\eta| < 2.5$	≥ 1

Table 3.2: Summary of associated objets and event selection.

is evident in the distribution of the forward jet η, as shown in Fig 3.5. Hence, specific observables such as the forward jet η, lepton charge asymmetry (due to the pp collision), and angular separation between the forward jet and different final state objects are analyzed to capture these distinctive features.

3.7.1   Background rejection

After the preselection stage, significant backgrounds remain in the signal region.. Therefore, a classifier is used to differentiate between the background and the signal. For this purpose, we employed a multivariate analysis technique [75]. A deep neural network (top-DNN) is trained to distinguish tHq from the ttH backgrounds and outputs a score for each event between 0 to 1. Events with scores closer to 1 are more likely to be ttH events, while those with scores closer to 0 are more likely to be tHq events. A boosted decision tree, BDT-bkg, is trained to distinguish tHq events from non-Higgs backgrounds (

Figure 3.5: The eta distribution of the jet with the largest $|\eta|$ (left) and the charge distribution of the lepton with the largest p_T (right). The events are selected from the $m_{\gamma\gamma}$ sidebands, fulfilling $m_{\gamma\gamma} \in [100, 120] \cup [130, 180]$ GeV. All simulated processes, tHq and backgrounds, are normalized according to the cross section and luminosity. Nevertheless, the tHq process is magnified 200 times for a shape comparison.

$\gamma\gamma, \gamma + \mathrm{jet}, tt + \gamma\gamma, tt + \gamma + \mathrm{jet}, t + \gamma + \mathrm{jet}$). BDT-bkg assigns a score to each event between -0.5 to 0.5, with events having scores closer to 0.5 being more likely to be tHq events. The kinematic variables of all the final state particles are used as inputs for both top-DNN and BDT-bkg training. We use the same top-DNN that was trained to differentiate tHq from the ttH category. In order to prevent a possible correlation between $m_{\gamma\gamma}$ and the training output scores, the ratios $p_T(\gamma_1)/m_{\gamma\gamma}$, $p_T(\gamma_2)/m_{\gamma\gamma}$, and $p_T(\gamma\gamma)/m_{\gamma\gamma}$ are provided as input to the BDT training, rather than using $m_{\gamma\gamma}$ directly. The linear correlation between the BDT score and $m_{\gamma\gamma}$ is found to be less than 5% in all samples. The inputs for the DNN are similar to those used in the ttH background discriminant described in Ref. [76]. The following is a list of variables used for the BDT:

- leading (subleading) photon $p_T/m_{\gamma\gamma}$

- leading (subleading) photon photon ID BDT

- leading (subleading) pixel detector seed

- leading lepton charge, p_T, η.

- Jet, b-jet and central jet($|\eta| < 1$) multiplicity

- $p_T, |\eta|$, and b-tag score for leading three jets

- $p_T, |\eta|$, and b-tag score for leading three b-jets

- $p_T, |\eta|$, and b-tag score of the forward jet

- Angular separation (ΔR) between each photon, the leading lepton, the leading b-jet, and the forward jet

The data and simulation comparison of each input distribution of the BDT training are checked. Figure 3.7,3.8, 3.9 demonstrate the good agreement observed between the data and simulation

105

The distributions of each variable used as input for the BDT training are compared between data and simulation in Figs. 3.7, 3.8 and 3.9 in the $m_{\gamma\gamma}$ sideband region, defined by $m_{\gamma\gamma} < 115$ or $m_{\gamma\gamma} > 135$ GeV. Good agreement is seen. The samples from 2016, 2017, and 2018 are trained separately. Finally, the data and simulation comparison for each of the BDT-bkg and top-DNN output distributions of the leptonic categories are checked in the $m_{\gamma\gamma}$ sideband region as shown in Fig. 3.6. All comparison plots shows an overall statistical agreement between data and simulation with a χ^2 p-value of 5%. The fluctuations observed in some of the distributions are associated with a limited simulated sample size.
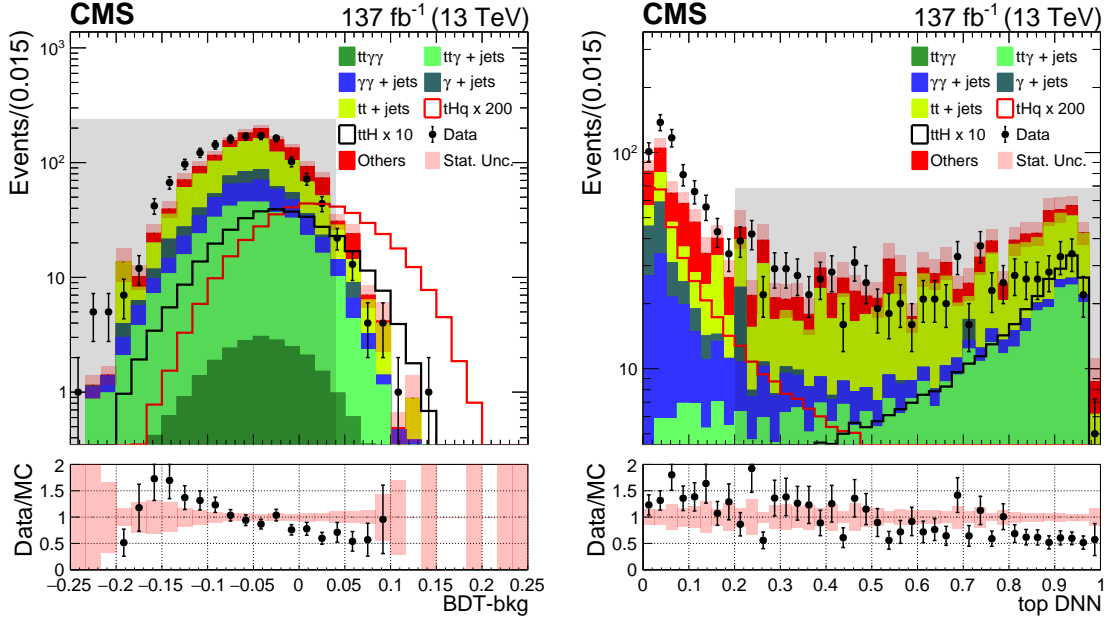
Figure 3.6: The distributions of the BDT-bkg (left) and top-DNN (right) data(black dot), training scores for the tHq (black), ttH (red), and data (black dots), as well as background events without Higgs bosons. The rejected events are represented by the shaded region in the figure. All simulated processes, including tHq and backgrounds, are scaled to the appropriate cross section and luminosity. Additionally, the tHq and ttH processes are scaled 200 and 10 times, respectively, for shape comparison.
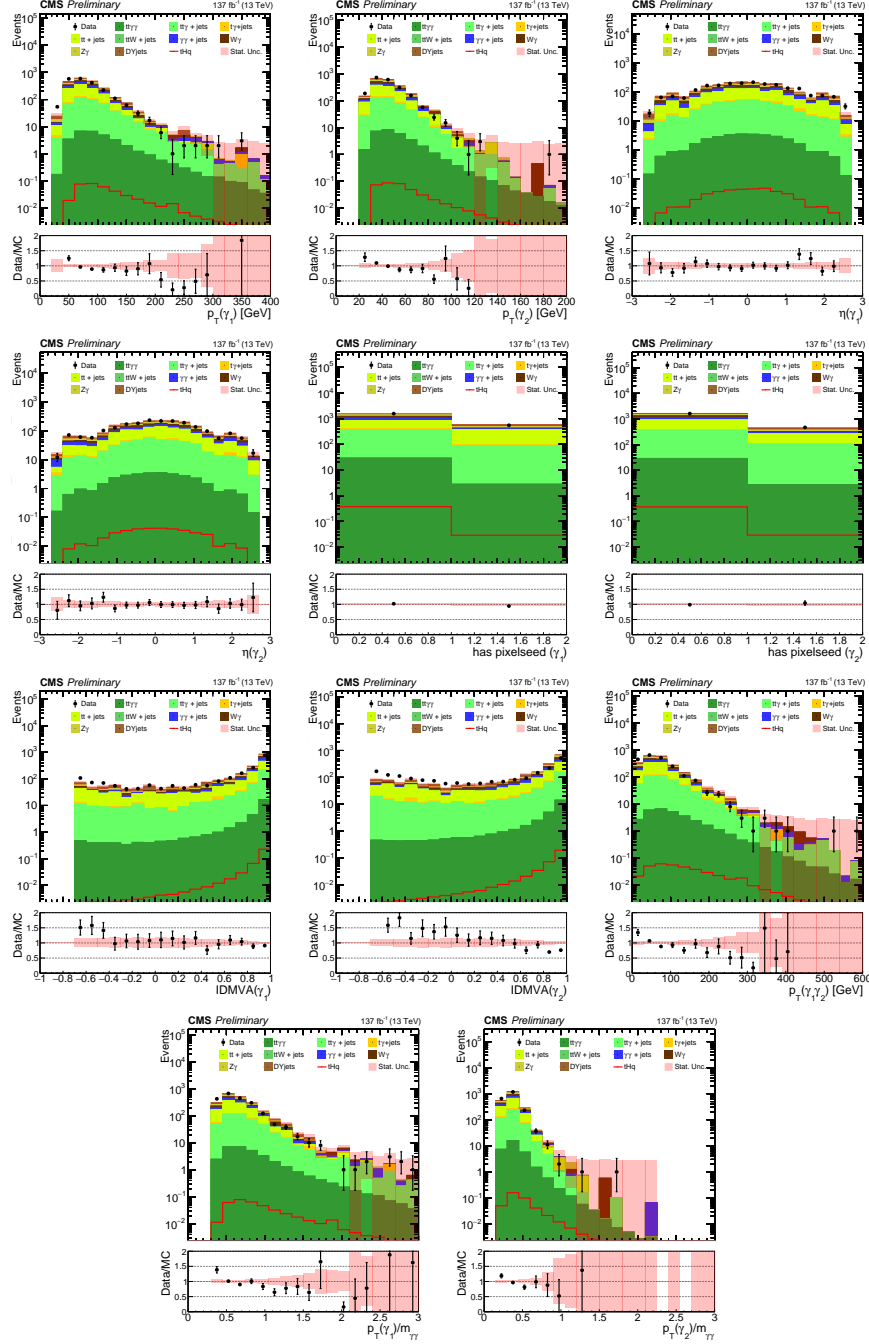
Figure 3.7: Comparison between data and MC for the distributions of photon kinematic variables used as input for BDT-bkg. All simulated processes, tHq and backgrounds, are normalized according to the cross section and luminosity.
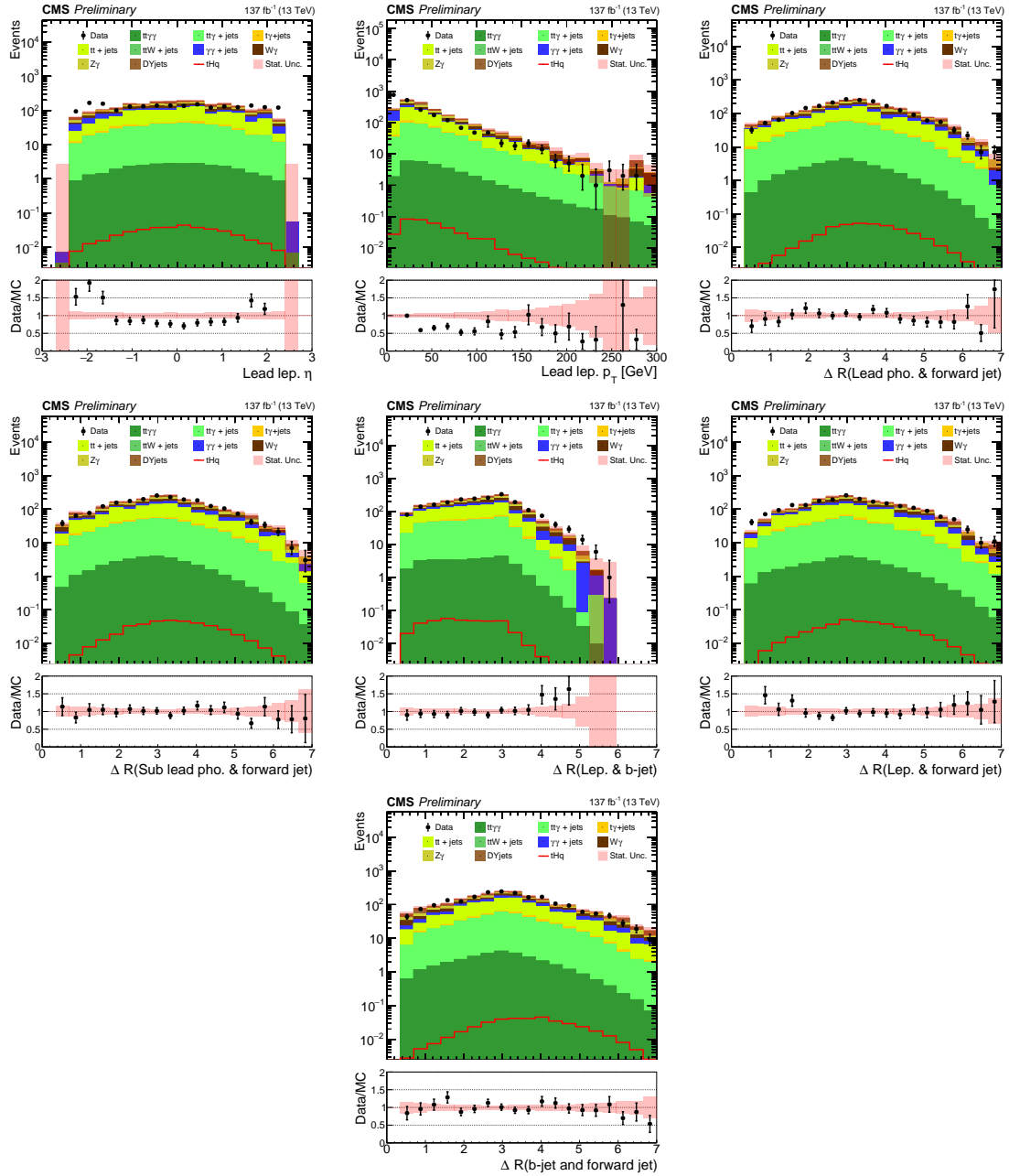
Figure 3.8: The data and MC comparison of the lepton kinematic variables and ΔR of different objects as input features in the tHq Vs. Non-Higgs BDT training. All simulated processes, tHq and backgrounds, are normalized appropriately according to the cross section and luminosity.

Figure 3.9: The data and MC comparison of the kinematic variables of jets as input features in the tHq Vs. Non-Higgs BDT training. All simulated processes, tHq and backgrounds, are normalized according to the cross section and luminosity.

3.8 BDT and DNN training configuration

The BDTs are trained using the TMVA framework [75]. For classification, the binomial log-likelihood loss is used.

$$L(F, y) = ln\left(1 + e^{-2F(x)y}\right)$$

The function F(x) represents a weighted sum of parametrized base functions, where y represents the true values. Each BDT is trained with a maximum depth of 2–4 layers. The learning rate, or shrinkage parameter, is fixed at 0.10. To improve stability and prevent overfitting, a bagging-like resampling procedure is used, where only 50% of random subsamples from the training events are used in each iteration for growing trees. This procedure helps maintain the robustness of the machine learning algorithms. When constructing a decision tree, each node is optimized by iteratively exploring 20 cut values within the variable range. Once a certain minimum number of events is reached, the node stops splitting. The training and testing samples are created by splitting the input events using an alternate split mode, allowing for independent evaluation of the trained BDTs using the test samples.

The DNN training is implemented using KERAS [77] with a TENSORFLOW [78] backend. For optimization, the DNN uses the Adam optimizer [79] with a learning rate set to 10^{-3} and uses a binary cross-entropy loss function. During training, an early-stopping procedure is applied, where the batch size is progressively increased [80]. Initially, it begins with a batch size of 1024 and continues training until the improvement in 1-AUC, as computed on the validation set, after each epoch falls below 1%. At this point, the batch size is quadrupled. This iterative process is repeated until the batch size reaches 50,000, after which it is capped. Training concludes when the validation AUC no longer exhibits improvements. In addition, dropout [81] is applied between the fully-connected layers in order to reduce overfitting and improve performance. The hyperparameters used for BDT and DNN training are summarized in Tab 3.3 and 3.4, respectively.

Hyper parameter	Value(s)
BoostType	Grad
NTrees	900-1000
MinNodeSize	1.0%
Shrinkage	0.10
Bagged sample fraction	0.50
nCuts	20
MaxDepth	2-4

Table 3.3: Hyperparameters used for the BDT training, specifically NTrees and MaxDepths parameters can vary with in the mention range for different cases.

Hyper parameter	Value(s)
Number of nodes (fully connected layers)	300, 200, 200, 200, 200
Number of nodes (LSTM layers)	150, 150, 150
L2-normalization constraint ("maxnorm")	3
Dropout rate	0.1
Learning rate	10^{-3}
Batch momentum	0.99
Activation function (LSTM)	hyperbolic tangent
Activation function (fully-connected layers)	exponential linear unit
Activation function (output layer)	sigmoid

Table 3.4: Hyperparameters used for the DNN training.

3.8.1   MVA threshold optimization

The signal significance $\frac{s}{\sqrt{s+b}}$ is used to optimize the thresholds on the training output scores. Here, s represents the number of tHq events that pass the threshold, and b represents the total number of background events that pass the threshold, ttH for the case of top-DNN and the total number of non-Higgs for the case of BDT-bkg. The efficiency for tHq, ttH and, non-Higgs boson events with respect to different MVA scores for the 2016, 2017 and 2018 samples, separately, are shown in the Fig. 3.11 and 4.13. The efficiency profile of the BDT-bkg is checked on the events that pass the preselection criteria mentioned in Sec. 3.7. The shaded regions in the figures correspond to the regions rejected for the leptonic category.
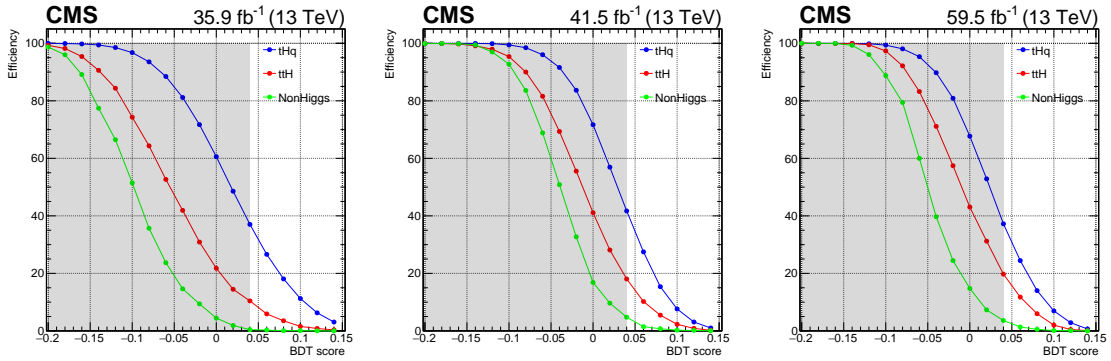
Figure 3.10: The efficiency of tHq (blue), ttH (red) and non-Higgs (green) events w.r.t. the BDT scores trained for 2016 (left), 2017 (middle), and 2018 (right) samples. The efficiency profile is checked on the events that pass the preselection criteria, mentioned in Sec. 3.7. The shaded area in the plot represents the rejected region of BDT score for the leptonic tHq category. The threshold is chosen based on the calculated significance and the efficiency of tHq events.
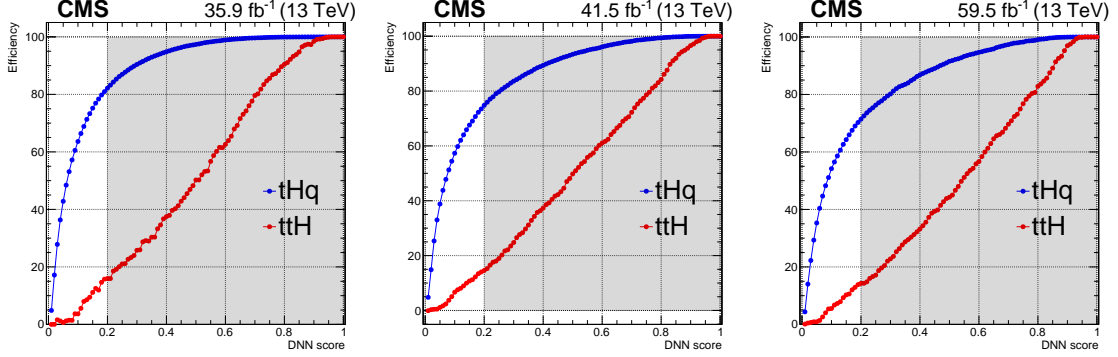
Figure 3.11: The efficiency profiles of tHq (blue) and ttH (red) events with respect to DNN scores for 2016 (left), 2017 (middle), and 2018 (right) samples. The DNN is trained specifically to distinguish between ttH and tHq events. The efficiency is evaluated on events that satisfy the preselection criteria mentioned in Sec. 3.7 and the BDT selection criterion described above. The shaded area in the plot corresponds to the rejected region of DNN score in the leptonic tHq category, which is selected based on the calculated significance and the efficiency of tHq events.
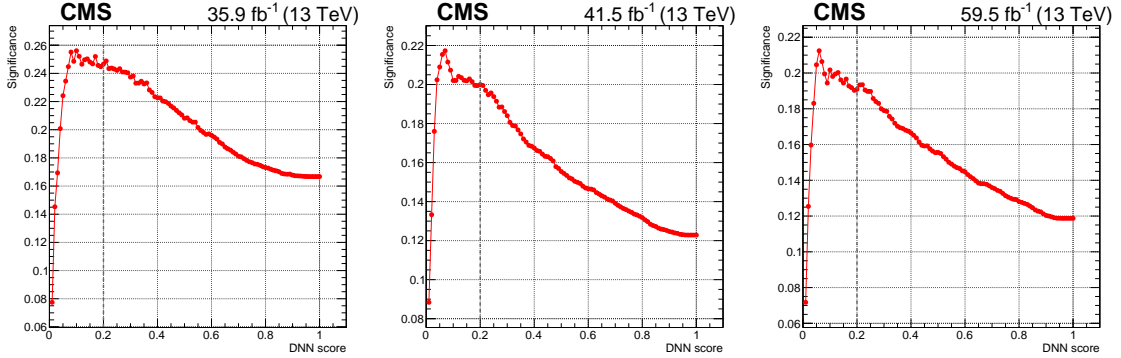
Figure 3.12: The significance that was calculated in order to determine the appropriate threshold for the DNN training in 2016 (left), 2017 (middle), and 2018 (right) samples. The threshold is chosen to be slightly below the point of maximum significance, as choosing that point corresponded to a sharp decrease in the efficiency of tHq events. The chosen threshold value is indicated by the black dotted line.

Table 3.5: Thresholds for the two MVAs in leptonic tHq tagger

Year	top-DNN threshold	BDT-bkg threshold
2016	0.2	0.04
2017	0.2	0.04
2018	0.2	0.04

Table 3.6: Yield of tHq and ttH events after both MVA selections, combined over all years.

Process	No. of events after preselection	No. of events after BDT-bkg cut	No. of events DNN-top cut
tHq	1.065	0.411(38.5%)	0.311(29.2%)
ttH	17.516	2.977(17%)	0.436(2.5%)

3.8.2 Validation of MVAs in tZq ($Z \rightarrow ee$) and ttZ ($Z \rightarrow ee$) control regions

The effectiveness of the MVAs, which are designed to distinguish tHq from non-Higgs and ttH backgrounds, is validated by analyzing ttZ and tZq events in control regions. The kinematics of tZq and ttZ events with the Z boson decaying to two electrons are highly similar to those of tHq and ttH, respectively, providing an excellent means of testing the agreement between the data and MC in the MVA distributions. A Z$\rightarrow ee$ control region is defined by reversing the electron veto selection of the H $\rightarrow \gamma\gamma$ analysis. Electron veto excludes the photon candidate if its supercluster in the ECAL is close to the projected trajectory of a track that matches an electron. Tracks associated with a reconstructed photon conversion vertex are exempt from this rejection process. Additionally, other supplementary selections implemented are,

- $|m_Z - m_{ee}| < 10$ GeV

- At least one jet with $p_T > 25$ GeV and $|\eta| < 4.5$.

- At least one b-jet with loose working point of b-jet identification training (PFdeepCSV)

Figure. 3.13 illustrates the agreement between the data and MC in the tZq and ttZ control regions. In Fig. 3.13, tZq events are found to accumulate towards the lower top-DNN score side, while ttZ events tend to accumulate towards the higher score side. Similarly, tZq events tend to cluster towards the higher BDT score side.
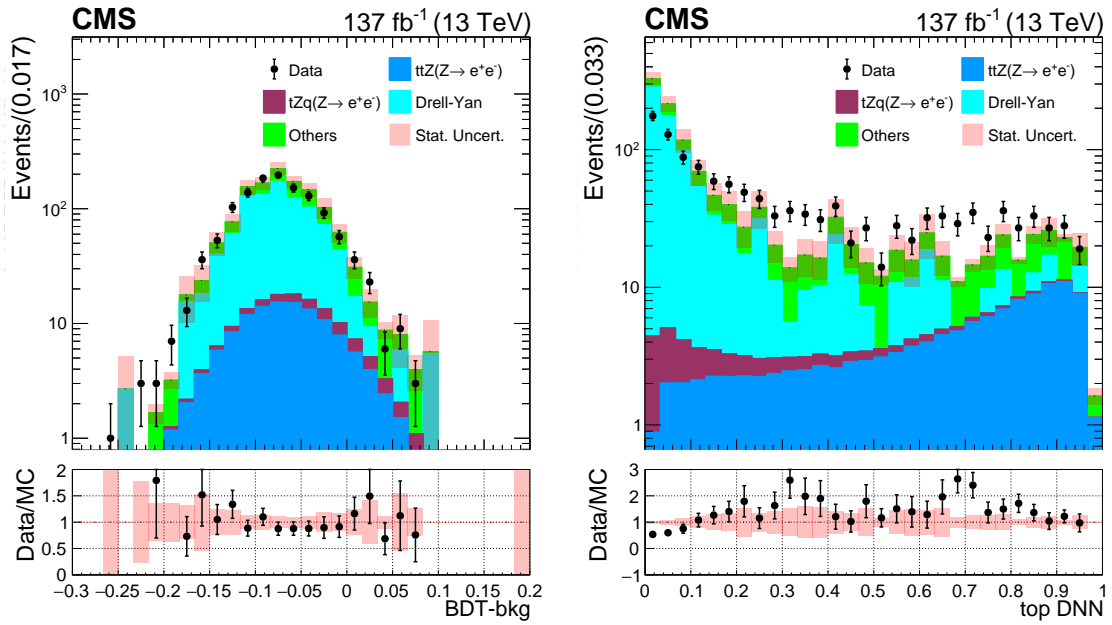
Figure 3.13: The distribution of the BDT-bkg (left) and top-DNN (right) MVA outputs for data compared to simulation in the tZq($Z \rightarrow ee$), ttZ($Z \rightarrow ee$) control region. The tZq events are accumulate towards the higher BDT score side, while we observe that the tZq events tend to cluster towards the lower DNN score side and ttZ events tend to cluster towards the higher DNN score side.

3.9   Statistical interpretation

The statistical methodology employed in this analysis follows the procedure outlined in Ref. [25], which was developed collaboratively by the ATLAS and CMS Collaborations. Binned maximum likelihood fits are performed simultaneously on the $m_{\gamma\gamma}$ distributions within the range of $100 < m_{\gamma\gamma} < 180$ GeV. For the analysis category, a likelihood function is defined, from analytical models to characterize the $m_{\gamma\gamma}$ distributions of both signal and background events. Nuisance parameters are introduced to account for the uncertainties arising from experimental and theoretical systematic effects.

The analytic signal model used in this analysis is extracted from simulation, with a distinct model constructed for each particle. Both the shape and normalization of the model are parameterized as functions of the Higgs boson mass (m_H). On the other hand, the background model is directly determined from the observed $m_{\gamma\gamma}$ distribution in the collected data. The background model can adopt various functional forms, all of which exhibit a smoothly decreasing spectrum.

3.9.1   Signal model

The shape of the signal in the $m_{\gamma\gamma}$ distribution within the analysis category, assuming a nominal value for m_H, is generated using simulation data. In order to achieve precise measurements, signal models are created separately for each data-taking year to account for variations in detector performance. This approach helps to ensure that the signal model accurately reflects the detector performance for each year and allows for year-specific systematic uncertainties in the signal estimate to be incorporated into the final fit.

For building the signal model, the correct vertex ($\Delta z < 1$ cm) and wrong vertex ($\Delta z > 1$ cm) scenarios are taken into account separately since the distribution of $m_{\gamma\gamma}$ depends on whether the vertex associated with the candidate diphoton was properly recog-

nized within 1 cm. A separate function is constructed for each vertex scenario by fitting the $m_{\gamma\gamma}$ distribution using a sum of at most five Gaussian functions. The number of Gaussian function combinations is determined by the form of the $m_{\gamma\gamma}$ distribution and is chosen to minimize the χ^2 per degrees of freedom, where the number of degrees of freedom is equal to the number of $m_{\gamma\gamma}$ bins (each having at least one event) minus the number of shape parameters in the fitted function. Figure 3.14 displays fits using varying numbers of Gaussians for simulated tHq events. Figure 3.15 shows the signal models. The diphoton mass
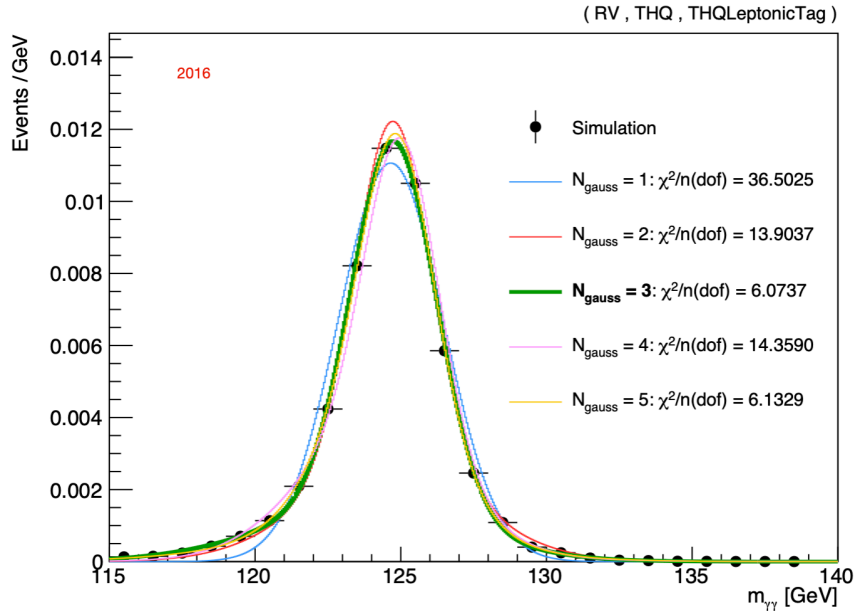
Figure 3.14: The optimal number of Gaussian functions required to fit the signal peak is determined in the leptonic tHq category for the 2016 simulated tHq events with the Higgs boson mass $m_H = 125.0$ GeV. The events in the correct vertex scenarios are considered here. It is concluded from the χ^2/ndof value that the most suitable choice for this scenario is three Gaussian functions.

resolution is quantified using the σ_{eff}, which is defined as half of the smallest interval containing 68.3% of the invariant mass distribution. In the plot, the models for the individual year are shown together with the combined three years, and the corresponding σ_{eff} values are displayed.
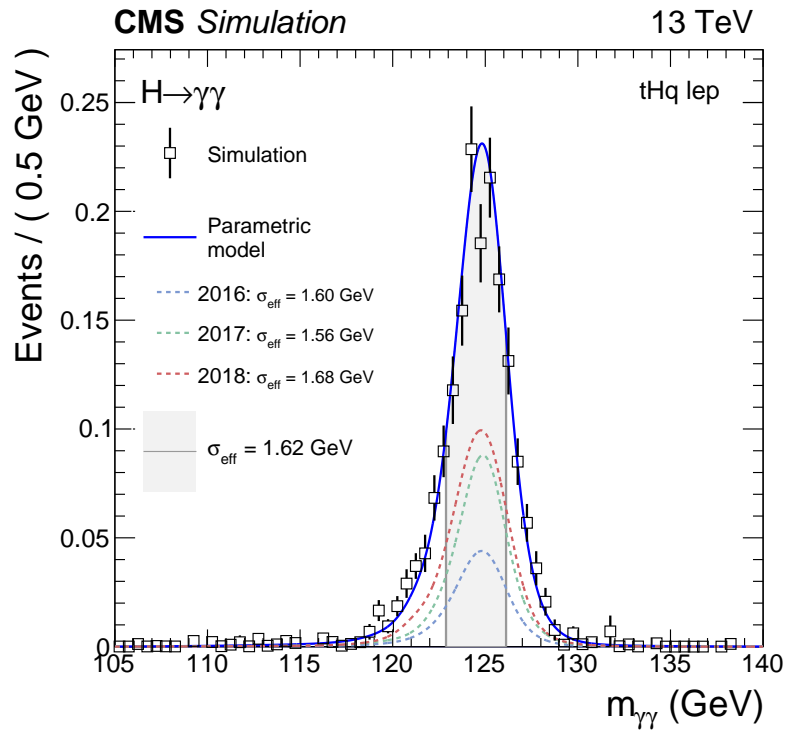
Figure 3.15: The shape of the parametric signal model for each year of simulated tHq events as well as for all three years combined. The dotted lines represent the contribution of the signal model from each individual year of simulation.

3.9.2 Background model

After the final selection, a considerable number of events enter the analysis category that does not originate from Higgs boson production and form a smoothly falling $m_{\gamma\gamma}$ distribution. Therefore the signal peak is expected to reside on top of the falling $m_{\gamma\gamma}$ distribution. The model used to describe these nonresonant backgrounds is extracted from fitting data with different analytic falling functions. A large set of candidate function families is considered, including exponential functions, Bernstein polynomials, Laurent series, and power law functions [82]. The considered functions can be expressed as follows:

1. Exponential:

$$f_N(x) = \sum_{i=0}^{N} a_i \exp\left(-b_i x\right), \tag{3.1}$$

2. Power Law:

$$f_N(x) = \sum_{i=0}^{N} a_i x^{-b_i}, \tag{3.2}$$

3. Bernstein polynomial:

$$f_N(x) = \sum_{i=0}^{N} a_i \binom{N}{i} x^i (1-x)^{N-i}, \tag{3.3}$$

4. Laurent series:

$$f_N(x) = \sum_{i=0}^{N} a_i x^{-4+\sum_{j=0}^{i}(-1)^j j}, \tag{3.4}$$

The systematic uncertainties associated with selecting a certain analytic functions to fit the background $m_{\gamma\gamma}$ distribution is estimated using the discrete profiling method [83, 82]. This technique uses the considered functions as a discrete nuisance parameter in the likelihood fit to the data. The maximum order of parameters to be used for each family of functions is decided using an F-test, and the minimum order is established by imposing a requirement on the goodness-of-fit to the data with a loose requirement on the goodness-of-fit. Finally, in the fit of extracting the parameter of interest, the discrete parameters that characterize the

selection of background functions, as well as the parameters associated with the functions, are both considered and treating as nuisance parameters. Figures 3.16 displays the different functions fitted to the data distributions. The fitted background shapes are normalized with the total number of events in the range of $m_{\gamma\gamma} \in 100\text{--}180$ GeV. The associated normalization uncertainty is also propagated in the final fit and is found to be one of the dominant uncertainties in the analysis.
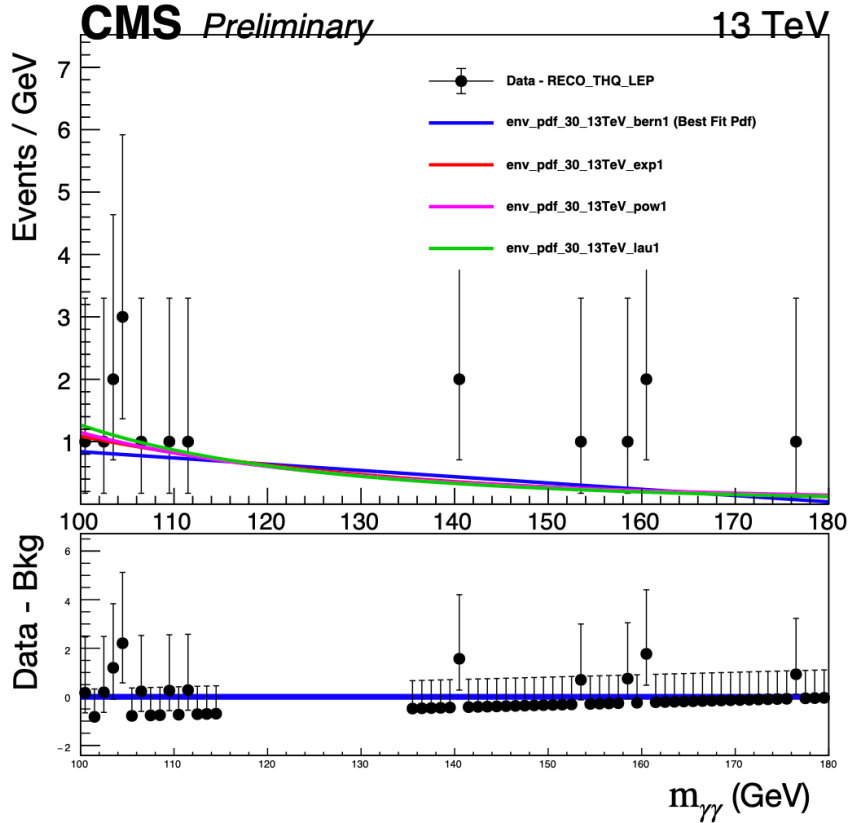
Figure 3.16: Model for nonresonant background which passes the final analysis selections. The residuals after the background component subtraction are shown in the bottom panel. The data points are fitted with Eqn. 3.1, 3.2, 3.3, and 3.4 indicated with exo1, paw1, bern1, and lau1, respectively, all with the order of parameters of 1 (determined by the F-test).

3.10 Systematic uncertainties

The potential systematic uncertainties taken into account in the analysis are listed in this section. As mentioned before, the systematic uncertainties associated with the data-driven background estimation are accounted for by fitting the data $m_{\gamma\gamma}$ distribution with different functions and treat each functions as a discrete nuisance parameter. The uncertainties that affect $m_{\gamma\gamma}$ distribution can alter both the mean and width of the Gaussian distributions so they are taken as nuisance parameters in the signal model. Again there exist some uncertainties that do not affect the shape of the $m_{\gamma\gamma}$ distribution but do affect the signal normalization. These uncertainties are handled by representing them as a log-normal variation in the event yield. The details of the sources of the theoretical and experimental uncertainties [84] are listed bellow.

3.10.1 Theoretical uncertainties

The theoretical uncertainty that have been taken into account in this analysis are as follows:

- *QCD scale uncertainty*: In this analysis, the computation of the expected cross section and event kinematics of the simulated events involved accounting for uncertainty resulting from changes in the normalization and factorization scales, which reflect the absence of higher order terms in perturbative calculations. These adjustments were made in accordance with the recommendations outlined in Ref. [2]. To estimate the overall normalization uncertainty, the renormalization and factorization scales were varied by two and changed simultaneously in the same direction. The magnitude of this uncertainty varies depending on the extent to which QCD is involved in a production process, ranging from approximately 0.05% for VH production (which is primarily driven by electroweak interaction) to 16% for tHq production. However, the effect of these uncertainties on the analysis is negligible, with an impact of less

than 1%.

- *PDF (parton density functions) uncertainties*: This uncertainty arises from the limited understanding of the composition of the proton and its impact on the selection of the most probable partons for initiating high-energy events. The determination of overall normalization uncertainties follows the prescription provided by PDF4LHC [85, 3], whereas the uncertainties in event kinematics are calculated using the MC2hessian method [86] with the NNPDF3.0 PDF set [66]. The overall normalization uncertainties for all processes range from 1.5% to 4%.

- *Uncertainty in the $H_{\gamma\gamma}$ branching fraction*: The calculation of the SM expected cross section depends on the branching fraction of Higgs decaying to two photons. However, the branching fraction is currently known with an estimated 2% uncertainty according to the latest findings [2]. This uncertainty has an effect on the overall normalization of the signal processes, which, in turn, affects the measurement of signal strength. The impact of this uncertainty on the signal strength measurement is 1.5%.

3.10.2 Experimental uncertainties

The uncertainties which affect the shape of the signal $m_{\gamma\gamma}$ distribution are listed below.

- *Photon energy scale and resolution*: To determine the uncertainties associated with the correction applied to the resolution in simulation and the photon energy scale in the data, simulated $Z \to ee$ events are utilized. The calculation involves varying the electron selection criteria, the distribution of R_9, and the regression training scheme. The estimated uncertainty for most photons is between 0.05% to 0.015% in terms of energy scale. However, for high p_T photons, the resulting effect may be between 0.5% to 3%.

- *Shower-shape corrections*: An uncertainty in the shower-shape correction is taken into consideration in this analysis, resulting from the imperfect modeling of the correction in the simulation. The uncertainty is determined by comparing the energy scale before and after applying the correction to the shower-shape variables, as very briefly mentioned in Sec. 3.5. The degree of uncertainty in the energy scale is dependent on the photon's $|\eta|$ and R_9 values, with a magnitude ranging from 0.01% to 0.15%.

- *Modelling of material in front of the ECAL*: As the objects generated by the collision pass through a certain amount of material before reaching the ECAL, the behavior of the electromagnetic shower is significantly affected by this material. Due to imperfect modeling of this in the simulation, an uncertainty in the photon energy scale is estimated by examining dedicated samples and varying the upstream material. The resulting uncertainty ranges from 0.02% to 0.05% for the most central and forward photons.

The uncertainties which only modify the event yield include:

- *Integrated luminosity*: the CMS luminosity monitoring has determined uncertainties of 1.2%, 2.3%, and 2.5% for the 2016, 2017, and 2018 data sets, respectively. These are partially correlated across the different data sets to account for common sources of uncertainty in the luminosity measurement schemes.

- *Photon identification BDT score*: The uncertainty arising from the photon identification BDT is evaluated by changing the set of events used to train the quantile regression corrections. This uncertainty is then carried through the complete event selection process to estimate the associated uncertainty in the signal yield. The impact of this uncertainty on the signal strength measurement in the analysis categories is found to be less than 1%.

- *Jet energy scale and smearing corrections*: The uncertainty from the jet energy scale is several percent and is dependent on p_T and η. This scale is evaluated using the p_T balance of jets with Z bosons and photons in $Z \rightarrow ee$, $Z \rightarrow \mu\mu$ and γ+jets events, in addition to the p_T balance between jets in dijet and multijet events [87]. To determine the effect of jet energy scale uncertainties on the signal yield, the jet energy corrections are varied within their respective uncertainties, and the variation is propagated to the result. The analysis is found to have an impact of around 1% due to the scale uncertainties, while the resolution affected by less than 1%.

- *Per-photon energy resolution estimate*: The uncertainty in the per-photon resolution is modelled by a $\pm 5\%$ rescaling from its nominal value. This uncertainty is intended to account for any discrepancies between the data and simulation in the distribution, which is obtained from the energy regression. The resulting variation in the yield in the analysis category is typically up to 5%.

- *Trigger efficiency*: The trigger selection efficiency is evaluated through the tag-and-probe method using $Z \rightarrow ee$ events. The corresponding uncertainty is estimated to be less than 1%.

- *Photon preselection*: The uncertainty in the preselection efficiency from cuts on R9, H/E, p_T, $\sigma_{i\eta i\eta}$ and isolations (details of cuts are mentioned in Tab. 3.1) is determined by computing the ratio of the efficiency measured in data to that in simulation. The magnitude of this uncertainty is less than 1%.

- *Missing transverse momentum*: The computation of this uncertainty involves shifting the reconstructed transverse momentum of the particles that enter the missing transverse momentum calculation. The amount of shift applied to each reconstructed object is determined based on its corresponding momentum scale uncertainty, as ex-

plained in Ref. [87].

- *Lepton isolation and identification*: The uncertainty associated with the selection of leptons in the analysis is estimated by varying the ratio of efficiencies obtained from measurements in both data and simulation using the tag and probe method in $Z \to \ell\ell$ events. The resulting effect on the selection of leptons in the category is less than 0.2%.

- *Jet* b-*tagging*: To calculate the uncertainties in the b tagging efficiency, the b tag discriminator in both data and simulations is compared. These uncertainties also account for the statistical component of the estimate of the fraction of heavy and light flavor jets in both data and simulation.

- *Effect of HEM 15/16 failure*: The HEM15/16 issue refers to two HCAL modules whose power supply died in the middle of the data taking. To mitigate the impact of the issue, data recorded during the affected period were carefully analyzed, and appropriate corrections were applied. The effect is expected to be marginal in the analysis as in the similar study of the all hadronic analysis [88] the effect for given jets was below 1%. In a conservative scenario, we quotes a 1% systematic uncertainty for this issue.

- *Prefiring of L1 trigger*: During the years 2016 and 2017, the ECAL in the CMS detector experienced a gradual timing shift. This timing shift was not correctly accounted for in the propagation of level-1 trigger primitives (TP). The level-1 TP are the initial set of criteria used at the first level of the CMS trigger system to quickly decide whether an event should be kept for further analysis or discarded. Consequently, a notable portion of high η TP was incorrectly associated to the previous bunch crossing. The influence on signal and SM Higgs boson production modes due to the L1

prefiring is evaluated and the event weight accounting for the issue is calculated as the product over non-prefiring probabilities of all jets and photons in the event. The corresponding systematic uncertainty in the efficiency is taken as either 20% or summation of per-object uncertainties in quadrature (whichever is larger).

3.10.3   Profile Likelihood method

In high energy physics, the maximum likelihood method is a widely used statistical technique for parameter estimation and hypothesis testing. It plays a crucial role in data analysis and model fitting. The maximum likelihood method aims to find the values of the model parameters that maximize the likelihood function given the observed data. The likelihood function represents the probability of obtaining the observed data as a function of the model parameters. By maximizing the likelihood function, we can determine the most probable values for the parameters.

In addition to parameter estimation, the maximum likelihood method is also used for hypothesis testing in high-energy physics. The likelihood ratio test is commonly used to assess the consistency of theoretical hypotheses or models by comparing their goodness-of-fit. The likelihood ratio is the ratio of the maximum likelihood values obtained under different hypotheses. By comparing the likelihood ratios with appropriate statistical distributions, we can determine the statistical significance of a particular hypothesis or the compatibility of different models with the data.

The probability of observing a tHq event in a single pp collision out of the total number of collisions can be expressed as:

$$p(\text{tHq event}) = \frac{L\sigma(pp \to \text{tHq})A\epsilon_{ff}}{L\sigma(pp)} \tag{3.5}$$

The probability to observe $n_{\text{tHq}}^{\text{obs}}$ tHq events out of N p-p collisions:

$$P(n_{\text{tHq}}^{\text{obs}}) = \frac{N!}{n_{\text{tHq}}^{\text{obs}}!(N - n_{\text{tHq}}^{\text{obs}})!} p^{n_{\text{tHq}}^{\text{obs}}}(1-p)^{N-n_{\text{tHq}}^{\text{obs}}} \tag{3.6}$$

126

For $N \rightarrow \infty$

$$P(n_{\text{tHq}}^{\text{obs}}) = Poiss(n_{\text{tHq}}^{\text{obs}}, \lambda) = \frac{e^{-\lambda}\lambda^{n_{\text{tHq}}^{\text{obs}}}}{n_{\text{tHq}}^{\text{obs}}!} \tag{3.7}$$

where, λ is the average number of events per unit of time or space. Considering that the events are independent of each other and that the probability of a single collision producing a tHq event is constant throughout the data taking period, λ can be written as

$$\lambda = Np = L\sigma(pp).\frac{L\sigma(pp \rightarrow tHq)A\epsilon_{ff}}{L\sigma(pp)} = n_{\text{tHq}}^{\text{exp}} \tag{3.8}$$

For a counting experiment $n_{\text{tHq}}^{\text{exp}} = \mu s(\text{tHq}) + \text{b}$, where $s(\text{tHq}) = L\sigma_{\text{SM}}A\epsilon_{ff}$ and

$$\mu = \frac{L\sigma_{\text{obs}}(\text{tHq})}{L\sigma_{\text{SM}}(\text{tHq})} = \frac{\sigma_{\text{obs}}(\text{tHq})}{\sigma_{\text{SM}}(\text{tHq})} \tag{3.9}$$

The method of maximum likelihood is a technique for estimating the values of the parameters given a finite sample of data. If we consider a random variable x distributed according to a probability distribution function $f(x; \theta)$, where θ is the unknown parameter, the probability to observe x in the interval $[x_1, x_1 + dx_1]$ is $p_1 = f(x_1; \theta)dx_1$. Therefore the probability that x_i is in $[x_i, x_i + dx_i]$ for all i is $\Pi_{i=1}^{n}P_i$. In the usual likelihood method we determine the parameters relevant to the shapes of distribution. In the extended maximum likelihood [89] we take the absolute normalization of the Passion probability of the observed number of events.

In that case the probability to observe x_1 in the interval $[x_1, x_1 + dx_1]$ is

$$p_1 = f(x_1; \theta)dx_1 \times \frac{e^{-\lambda_1(\mu)}\lambda_1(\mu)^{n^{obs}}}{n^{obs}!}$$

The likelihood is defined as,

$$L(\mu, \theta) = \Pi_{i=1}^{n}p_i \tag{3.10}$$

127

The best fit values and confidence intervals for the parameters of interest are estimated using a profile likelihood test statistic [90],

$$q(\mu) = -2\ln\left(\frac{L(\mu, \hat{\hat{\theta}})}{L(\hat{\mu}, \hat{\theta}_\mu)}\right) \tag{3.11}$$

The likelihood functions in the numerator and denominator of Eqn. 3.11 are constructed by taking the product over the likelihood functions defined for each analysis category. The quantities $\hat{\mu}$ and $\hat{\theta}$ represent the unconditional maximum likelihood estimates for the parameters of interest and the nuisance parameters, respectively. Furthermore, $\hat{\hat{\theta}}$ corresponds to the conditional maximum likelihood estimate obtained for fixed values of the parameters of interest, $\hat{\mu}$. In this analysis, the parameters of interest can vary and may include signal strengths, cross sections, or coupling modifiers, depending on the specific fit being performed. In all fitting procedures, the value of m_H is kept fixed at its most accurately measured value of 125.38 GeV [24]. This choice ensures that all measurementsare reported consistently.

3.11 Results

The signal strength modifier, denoted as μ, is a commonly used parameter in high energy physics experiment that compares the observed cross section to the expected value predicted by some model. Specifically, here, μ is defined as the ratio of the Higgs boson cross section and diphoton branching fraction to the expected value predicted by the SM. During fitting the model, the Higgs boson mass, m_H, is typically held at a fixed value of 125.38 GeV, which is the most precisely measured value. However, if m_H is profiled without any constraint, the impact on the best fit signal strength values is found to be relatively small, resulting in changes of 0.7–1.8%. The signal strength modifier for the Higgs boson production associated with the single top quark mode measured from simultaneous fit with

all other Higgs production modes is found to be

$$\mu_{\text{tH}} = 6.24^{+0.62}_{-0.54}(\text{Syst})^{+3.67}_{-3.33}(\text{Stat}) \tag{3.12}$$

The profiled likelihood scan of the expected signal strength of tHq with fixed ttH, ggH, VBF and WH signal strength to 1 is displayed in Fig. 3.18. Signal strength modifiers and the best fit cross sections along with their corresponding 68% confidence intervals for all Higgs boson production categories are shown in Fig. 3.19. The VH hadronic and VH leptonic processes are categorized as μ_{VH} scale, whereas the VBF production mode falls under μ_{VBF}. Similarly, μ_{top} equally scales the ttH, tHq, and tHW production modes, while μ_{ggH} scales both ggH and bbH production modes. The Fig. 3.20 presents the values of signal strength modifiers of individual production modes and their uncertainties. The overall signal strength modifier in the H $\rightarrow \gamma\gamma$ channel is determined to be $\mu = 1.12^{+0.09}_{-0.09} = 1.12^{+0.06}_{-0.06}(\text{theo})^{+0.03}_{-0.03}(\text{syst})^{+0.07}_{-0.06}(\text{stat})$ by combining all the production modes. The best-fit signal-plus-background model with data points, keeping signal strength modifiers of all production modes in their best-fit value, is shown in Fig. 3.17. Figure 3.21 shows an event display for one data event passing the selection for Higgs boson production in association with a single top quark, with a diphoton invariant mass of 125.52 GeV. Over the past ten years since the discovery of the Higgs boson, the CMS experiment has observed it in various fermionic and bosonic decay channels, established its spin-parity quantum numbers, determined its mass, and measured its production cross sections in various modes. An effort has been made for a combined measurement of all Higgs boson production modes and Higgs boson decay channels, as shown in Fig. 3.22. Combining the measurement of the H $\rightarrow \gamma\gamma$ decay channel with other decay channels, the signal strength modifier for Higgs boson production associated with a single top quark is observed to be

$$\mu_{\text{tH}} = 6.05^{+1.69}_{-1.38}(\text{Syst})^{+2.06}_{-1.99}(\text{Stat}) \tag{3.13}$$

Even though in Figs 3.19, 3.20, and 3.22, μ_{tH} result is presented with the all other Higgs boson production mode measurements, this thesis only contributed to the measurements of the Higgs boson production mode associated with the single top quark (tH).
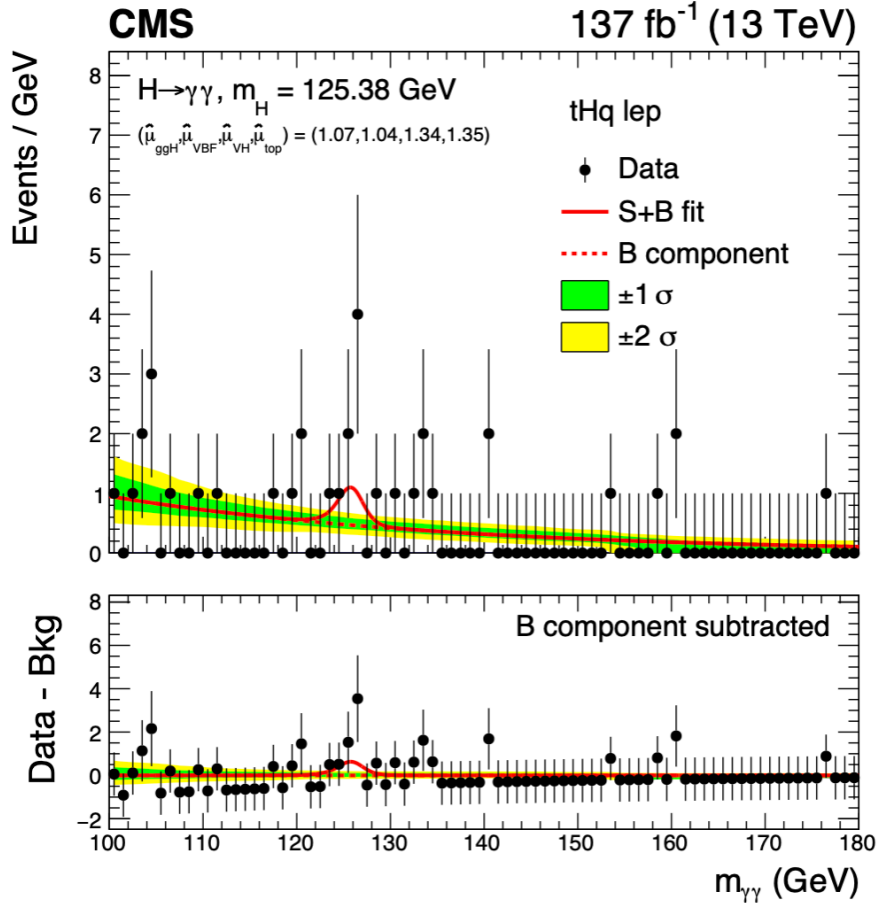
Figure 3.17: The best-fit signal-plus-background model with data points (in black) in the leptonic tHq category is presented while maintaining the signal strength modifiers of all production modes at their best-fit value. The uncertainties in the background model are demonstrated by the green (1 std. dev.) and yellow (2 std. dev.) bands. The total signal-plus-background contribution is presented by the solid red line, while the background component alone is represented by the dashed red line. The residuals after the background component subtraction are shown in the bottom panel.
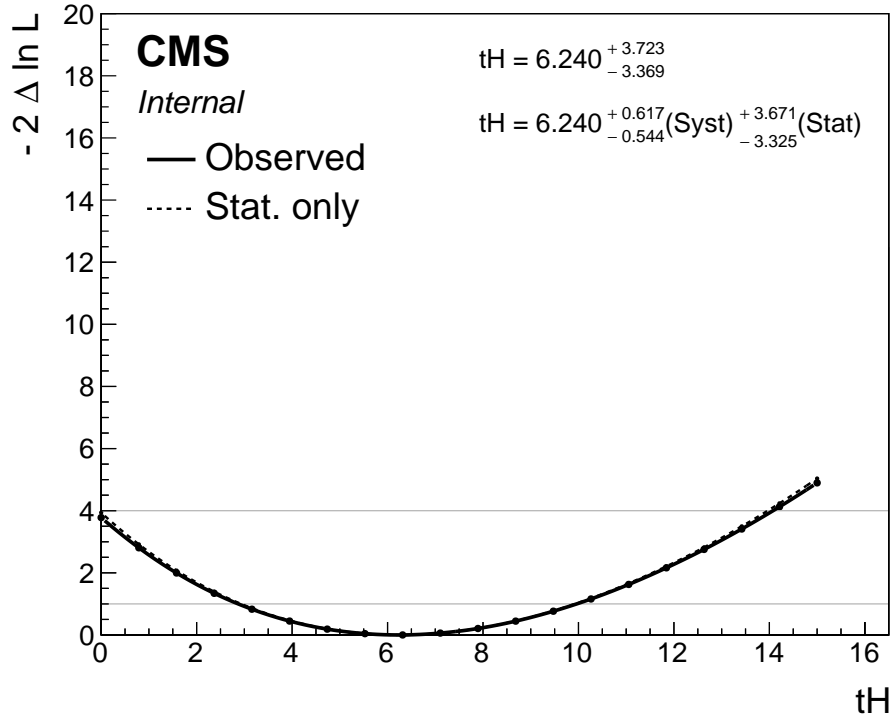
Figure 3.18: The profiled likelihood scan of the expected signal strength of tHq with fixed ttH, ggH, VBF and WH signal strength to 1.
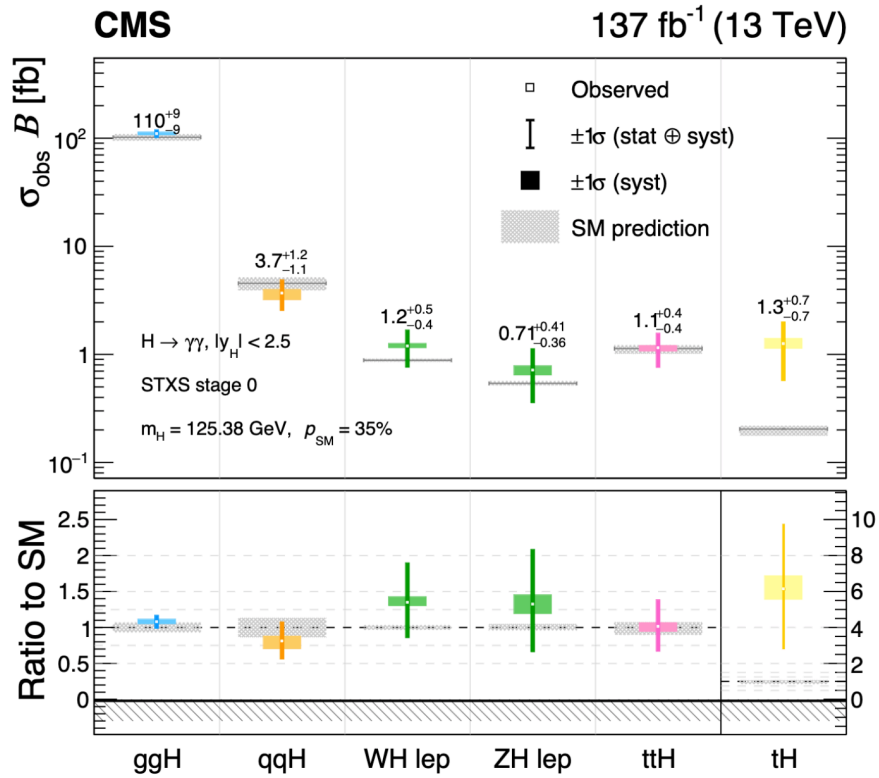
Figure 3.19: Observed results in the stage 0 STXS bins, displaying the best fit cross sections along with their corresponding 68% confidence intervals. The colored boxes illustrate the systematic components of the uncertainty in each parameter, while the hatched grey boxes depict the theoretical uncertainties in the SM predictions. The bottom panel displays the ratio of the fitted values to the SM predictions. The p-value for the compatibility of this fit with the SM prediction is approximately 66%.
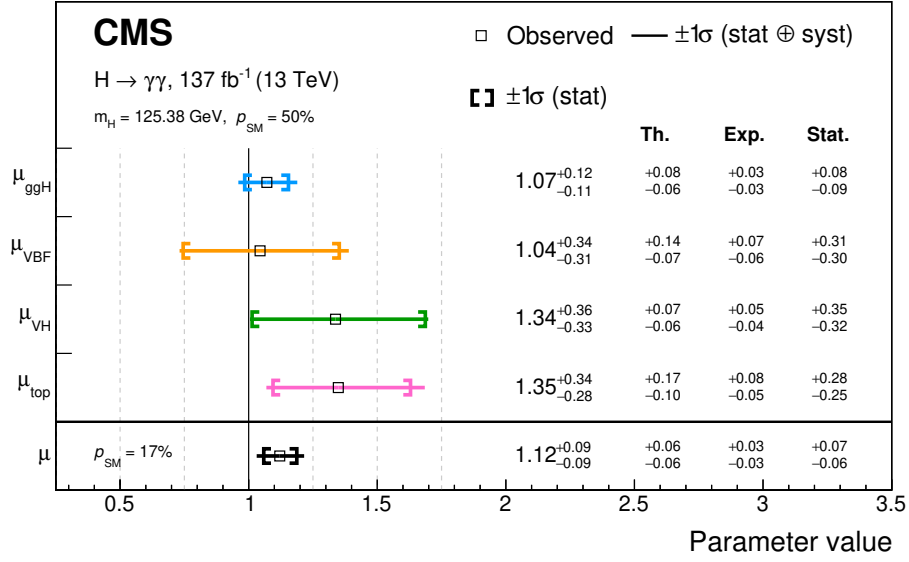
Figure 3.20: The fit results for the signal strength modifiers of the four main production modes. The individual contributions of theoretical systematic, experimental systematic, and statistical components to the total uncertainty in each parameter are displayed. The compatibility of this fit with the SM prediction is evaluated using a p-value, which is approximately 50%. Additionally, the fit results for the inclusive signal strength modifier are presented in black, and the corresponding p-value is 17%.
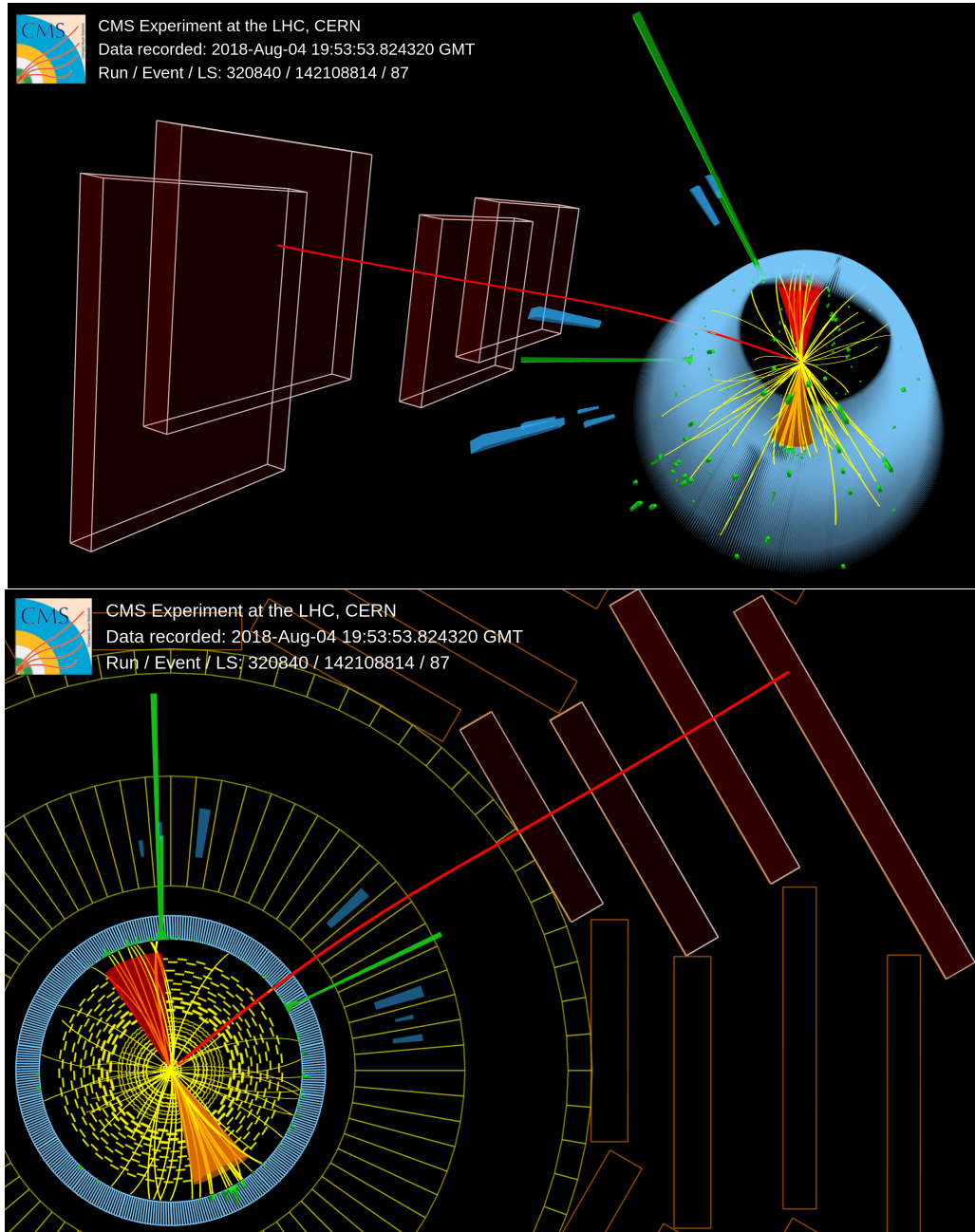
Figure 3.21: Visualisations of a candidate single top-associated production event in data. The event is selected in the leptonic tHq analysis category and is characterised by two photon candidates with a diphoton invariant mass of 125.52 GeV, shown by the green rectangles. The top quark decays into a W boson and a b quark. The long red line depicts the muon from the decay of the W boson and the red cone depicts the b-tagged jet originating from the b quark. The jet produced from the additional quark is shown as the orange cone.
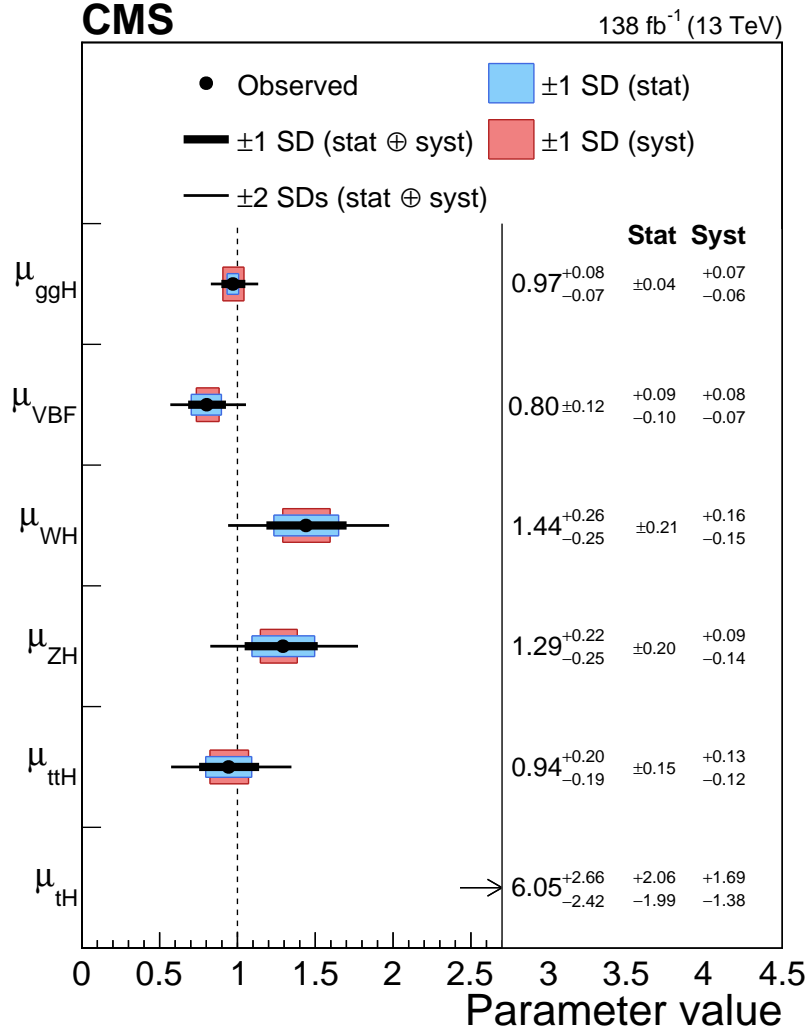
Figure 3.22: The signal strength parameters for different production modes μ_i obtained assuming $\mathcal{B} = (\mathcal{B})_{SM}$. The 1 (2) std. dev. confidence intervals are represented by thick(thin) black lines. The systematic and statistical components of the 1 std. dev. interval are shown by the red and blue bands respectively. The value of μ_i in the SM is represented by the vertical dashed line at unity. The p-value indicating the probability of deviation from the SM prediction (i.e., the likelihood of the observed result deviating as much or more from the SM prediction) is 3.1%.

3.12 Summary

The study presents measurements of the cross section of Higgs boson production associated with a single top quark (tH), with the Higgs boson decaying into a pair of photons. This

rare Higgs boson production mode is explored for the first time in the H $\to \gamma\gamma$ channel and included in the legacy measurements of Higgs production cross section. The analysis involves the selection of events with two photons from proton-proton collisions at a center-of-mass energy of $\sqrt{s}$ = 13 TeV recorded with the CMS detector at the LHC from 2016 to 2018, corresponding to an integrated luminosity of 137 fb^{-1}. This study focues only on tH production, where the top quark decays only leptonically. The novel multivariate analysis techniques are used to separate likely signal events from background processes, including the standard model Higgs boson production associated with the top and anti-top quark.

The measurement of the signal strength modifiers for Higgs boson production in association with the single top quark mode was obtained through a simultaneous fit with all other Higgs production modes and is observed to be $\mu_{\text{tH}} = 6.24^{+0.62}_{-0.54}(\text{Syst})^{+3.67}_{-3.33}(\text{Stat})$. Additionally, by combining the measurement from the H $\to \gamma\gamma$ decay channel with other decay channels, the signal strength modifier for this production mode is observed to be $\mu_{\text{tH}} = 6.05^{+1.69}_{-1.38}(\text{Syst})^{+2.06}_{-1.99}(\text{Stat})$

Chapter 4

Search for a Vector-Like Quark, $\mathrm{T}' \to \mathrm{tH}$ $(\mathrm{H} \to \gamma\gamma)$

Vector-like quarks (VLQs) are hypothetical spin-1/2 colored particles. In contrast to the SM chiral quarks, their left-handed and right-handed components transform in the same way under the standard model electroweak gauge group $\mathrm{SU(2)_L \times U(1)_Y}$. Therefore, unlike the chiral quarks in the SM, their masses are not generated by a Yukawa coupling to the Higgs boson and have less impact on the production cross section of the Higgs boson. A variety of new physics models introduce vector-like quarks. Such models include Composite Higgs model [31, 32, 33, 34, 35], Little Higgs models [36, 37, 38], and models with extra dimensions [39], etc. In theory, the existence of such VLQs might explain the experimentally measured mass of the Higgs boson by canceling out the leading quantum loop corrections coming from top quarks [13, 91]. In minimal models, VLQs may only exist as electroweak singlets and doublets, carrying respective electric charges of $+2/3$e and $-1/3$e, denoted as T' and B', respectively. Additionally, doublets and triplets further incorporate two additional particles X and Y with exotic charges of $+5/3$ and $-4/3$, respectively.

As discussed earlier, a vector-like top quark partner T' has two production modes. One is pair production through the strong interaction while the other is the single production mode through the electroweak interaction. For pair production through the strong interaction, the available parton level center-of-mass energy is shared between the two heavy particles. In contrast, in the EW production of a single T', a larger kinematic phase space is accessible and heavier masses can be probed [13]. The maximum single heavy quark production cross section at the LHC with center-of-mass energy of 13 TeV, for selected quark multiplets is

shown in Fig. 4.1 and the crossection is compared with the pair production of VLQ through strong interaction. The T′ quark could couple to bW, tZ, or tH, resulting in the corresponding T′ quark decays. For EW production of an isospin singlet T′ VLQ, considered in this search, the T′ branching fractions are assumed to be 50, 25, and 25%, respectively, for bW, tZ, and tH decays [13].
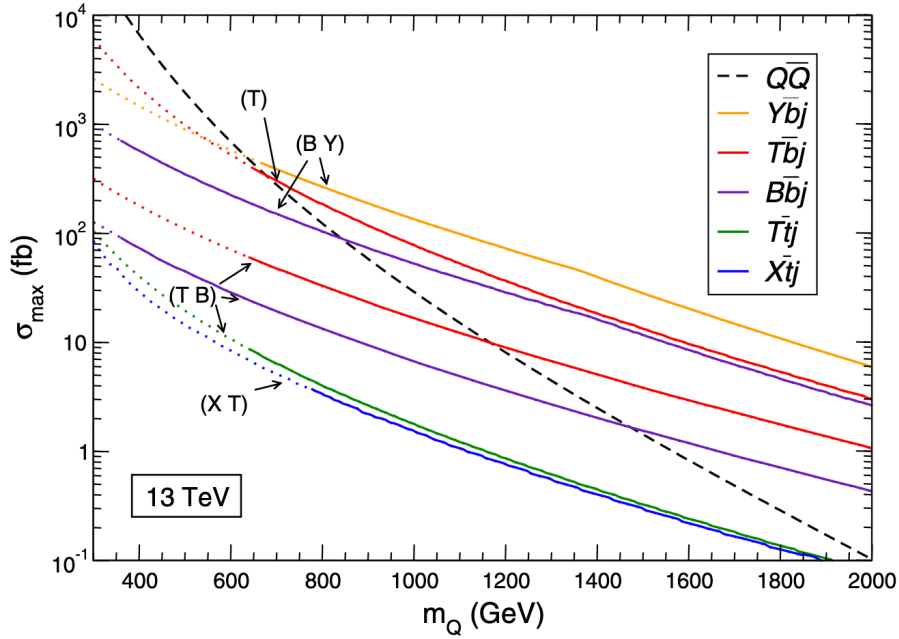
Figure 4.1: The single heavy quark production cross section at the LHC with centre-of-mass energy of 13 TeV, for selected quark multiplets. The dotted part of the lines indicates the range of masses already excluded by direct searches. The figure is taken from Ref. [13]. The "T" and "B" in the figure corresponds to the "T′ " and "B′" mentioned in this thesis. Furthermore, the symbol "Q" in the figure represents a broader category of heavy quark T', B', X, and Y.

The total cross section for EW production of T′ and subsequent decay of a T′ quark of mass $m_{T'}$ can be written as follows [92]:

$$\sigma\left(C_1, C_2, m_{T'}, \Gamma_{T'}\right) = C_1^2 \, C_2^2 \, \tilde{\sigma}_{FW}\left(m_{T'}, \Gamma_{T'}\right) \tag{4.1}$$

where C_1 and C_2 are couplings corresponding to production of a T′ quark and its subsequent decay, respectively, and $\Gamma_{T'} = \Gamma(C_i, \, m_{T'}, \, m_{\text{decays}})$ is the total width of the T′ quark. When

considering narrow width approximation (NWA), in which the Γ_X/m_X ratio approaches zero, the production and decay can be factorised out and the cross section can be expressed as

$$\sigma\left(C_1, C_2, m_{T'}, \Gamma_{T'}\right) = \sigma_P\left(C_1, m_{T'}\right) BR_{T' \rightarrow \text{decay channel}} = C_1^2\, \hat{\sigma}_{NWA}(m_{T'})\, BR_{T' \rightarrow \text{decay channel}}$$

(4.2)

where C_1 is the coupling corresponding to the production of the T' quark and C_2 and $\Gamma_{T'}$ are absorbed in the branching ratio, $BR_{T' \rightarrow \text{decay channel}}$. To be more specific, the equation can be written as

$$\sigma_{T' \rightarrow tH}\left(C_1, C_2, m_{T'}, \Gamma_{T'}\right) = C_1^2\, \hat{\sigma}_{NWA}(m_{T'})\, BR_{T' \rightarrow tH}$$

(4.3)

In a simple extension of the SM with just one VLQ representation, the coupling of a T' quark produced via the electroweak interaction involving a t-channel W boson and a b quark can be parametrized as

$$C_1 = \frac{e}{\sqrt{2}s_w}\kappa_{T'}$$

(4.4)

where s_w is the sine of the Weinberg angle θ_w, $\kappa_{T'}$ is a coupling strength, and e is the unit of electric charge. Combining Eqn. 4.3 and Eqn. 4.4, the resulting formulas are shown as follows:

$$\sigma_{T' \rightarrow tH} = \left(0.458486 \times \kappa_{T'}\right)^2\, \hat{\sigma}_{\text{NWA}}(m_{T'})\, BR_{T' \rightarrow tH}$$

(4.5)

$$\kappa_{T'} = 2.18109 \times \sqrt{\frac{\sigma_{T' \rightarrow tH}}{\hat{\sigma}_{\text{NWA}}(m_{T'})\, BR_{T' \rightarrow tH}}}$$

(4.6)

This study focuses on the single electroweak production of a vector-like top quark partner T' in proton-proton (pp) collisions at $\sqrt{s}$ = 13 TeV, pp $\rightarrow T'$ qb, with a decay involving the Higgs boson, $T' \rightarrow tH$, and the Higgs boson decaying to two photons (H $\rightarrow \gamma\gamma$). A leading-order (LO) Feynman diagram along with the H $\rightarrow \gamma\gamma$ decay is displayed in figure 4.2. The EW production cross section explicitly depends on the couplings of the VLQ

to third-generation quarks [13, 93]. Here, the coupling κ_T can significantly change based on the choice of the VLQ mass and width. In this study, the VLQ production and decay are parametrized using the narrow width approximation (NWA) [94], where the T' natural width (Γ) is approximately 1% relative to its mass ($\text{M}_{\text{T}'}$). The NWA is valid up to $\Gamma/\text{M}_{\text{T}'} \approx 10\text{--}15\%$, beyond which the large width of T' and its interference with the SM background become important [93, 95]. The leptonic and hadronic decay modes of the top quark are treated separately to achieve the best possible search sensitivity while being the first T' search by the LHC experiments in $\text{H} \rightarrow \gamma\gamma$ channel.
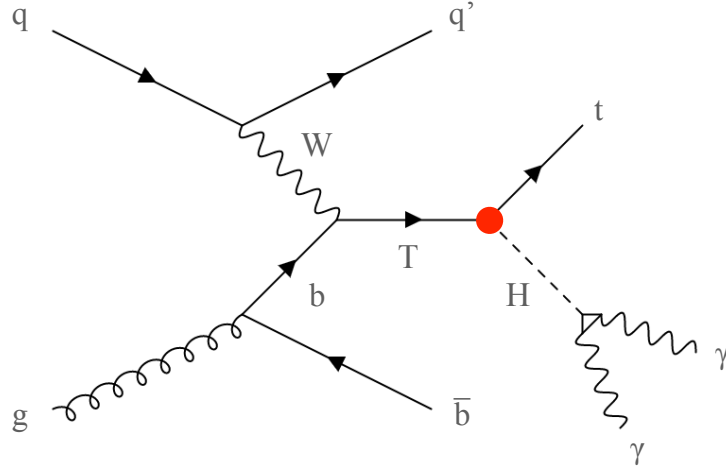
Figure 4.2: Leading-order Feynman diagram for single T' production in Wb fusion and subsequent decay into $\text{tH}(\gamma\gamma)$.

Previous searches at the LHC for vector-like quarks have been performed by both the ATLAS and CMS collaborations. Some of the recent publications at $\sqrt{s} = 13$ TeV are

- ATLAS 13 TeV, 36.1 fb^{-1}, pair production, $\text{T}' \rightarrow \text{Zt/Wb/Ht}$, $\text{B}' \rightarrow \text{Zb/Wt/Hb}$ [96]

- ATLAS 13 TeV, 139 fb^{-1}, single production, $\text{T}' \rightarrow \text{tH}(b\bar{b})$ [97]

- CMS 13 TeV, 35.9 fb^{-1}, pair production, fully hadronic final states [98]

- CMS 13 TeV, 35.9 fb^{-1}, electroweak production, fully hadronic final states [99]

These previous searches used the reconstructed T' invariant mass or transverse mass as the main observable. The present analysis exploits the excellent resolution of 1–2% for the reconstructed H mass in the diphoton decay channel to search for a signal characterized by a peak at the H mass above the falling diphoton mass ($m_{\gamma\gamma}$) continuum. The statistical methodologies and the non-Higgs background determination techniques are identical to the ones used in the SM H $\to \gamma\gamma$ measurements.

The chapter is structured as follows: This portion begins with the motivation behind searches for vector-like quarks and highlights the latest publications on the subject. In Sec. 4.1, the overall strategy for the analysis is outlined. Section **??** and 4.3 provides details on the samples and objects used in the study, as well as analysis specifics for both the leptonic and hadronic channels. Techniques used to reject backgrounds and optimize signal regions are discussed in Sec. 3.8 and 4.4, respectively. In Sec. 4.5, the analysis delves into the modeling of signal and background. Section 4.6 study to evaluate any potential bias in the model for the parameter of interest. Section 4.7 explores various sources of systematic uncertainties and their impact on the limits. Section 4.8 provides a means to assess the degree of compatibility between the observed data and the probability density function of the model. Subsequently, Sec. 4.9 presents the physics results, which are then summarized in Sec. 4.10.

4.1 Analysis strategy

This analysis is a search for the production of a vector-like quark, T', which subsequently decays into a top quark and a Higgs boson that further decays into two photons. Similarly to the previous studies, the analysis greatly depends on how well photons can be efficiently reconstructed with precise energy determination. The primary discriminant in this analysis

is the Higgs invariant mass (or diphoton invariant mass, $m_{\gamma\gamma}$), which serves as the main observable for identifying T' events. The reconstruction of photons and diphoton candidates employs the same methodologies as used in the $\mathrm{H} \to \gamma\gamma$ measurements discussed in the previous chapter. Two mass regions are defined accordingly:

- Signal window: $m_{\gamma\gamma} \in [115, 135]$ GeV

- Sideband region: $m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV,

where the signal window is blinded throughout the development of the analysis. Two loose preselection criteria are implemented to maintain a high level of efficiency in identifying signal events. One preselection focuses on the leptonic decays of the top quark, while the other is designed for the hadronic decays. In each channel, the following steps are performed:

1. Train binary MVA classification algorithms using simulated T' samples as the signal and simulation of related standard model processes as the background. The training specifications are as follows:

   - In the leptonic channel, a single MVA is trained, including only standard model Higgs production modes as the background.

   - In the hadronic channel, two MVAs are trained due to the large background: One MVA is trained to suppress SM Higgs boson (SMH) processes and another MVA is trained to reject SM non-resonant backgrounds (NRB).

2. Define a signal region by applying rectangular cuts on the MVA scores and reconstructed T' mass ($\mathrm{M}_{\mathrm{T}'}$). These cuts are determined to maximize the significance, as discussed in Sec. 4.4

3. Parametric signal and background models are constructed by fitting the $m_{\gamma\gamma}$ distribution in each category.

- The T′ signal processes and the SM Higgs background processes are modeled with a sum of Gaussian functions.

- The non-resonant background is modeled using data from the $m_{\gamma\gamma}$ sideband, using various functional forms. The choice of the functional form is treated as a discrete nuisance parameter, following the discrete profiling method [61].

Specifically, the fitting procedure is carried out separately for each year and channel to account for the year-dependent $m_{\gamma\gamma}$ resolution. Subsequently, a simultaneous fit of the $m_{\gamma\gamma}$ spectra is conducted for both the leptonic and hadronic channels, aiming to derive a cross section limit for each mass point. The analysis workflow, illustrating these steps, is depicted in Fig. 4.3. In order to investigate the possibility of a potential low T′ event
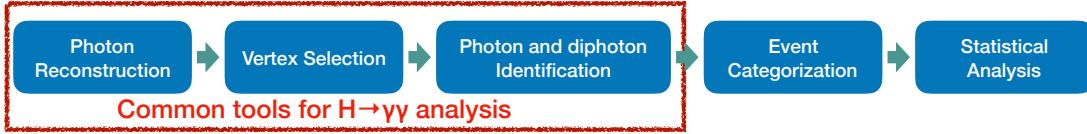
Figure 4.3: Diagram of analysis workflow.

excess, it has been observed that using a single training approach for the entire mass range of 600 to 1200 GeV is not optimal. To address this, a strategy for the analysis is developed by defining three distinct T′ mass categories. These considerations include the similarity of kinematics among signal processes when T′ masses are not significantly different and also to avoid the excessive complexity in training a multiple sets of MVAs. As a result, the following three mass ranges are defined to ensure proper grouping of signal samples for training:

- $M_{T'} \in [600, 700]$ GeV

- $M_{T'} \in [800, 1000]$ GeV

- $M_{T'} \in [1100, 1200]$ GeV.

The analysis procedure described above is followed for each mass category, yielding the final results. Additionally, it is confirmed that the search for T' masses falling between the defined mass categories, specifically in the ranges of [700, 800] GeV and [1000, 1100] GeV, can also be effectively described using this strategy.

4.2   Collision data and simulated events

The analysis is conducted using proton-proton collision data at a center-of-mass energy of $\sqrt{s} = 13$ TeV, collected by the CMS experiment during the 2016, 2017, and 2018 data-taking periods. These data correspond to integrated luminosities of 35.9, 41.5, and 59.4 fb^{-1}, respectively. The event selection in the data is performed using a diphoton high-level trigger with different photon transverse momentum thresholds: 30 GeV for the leading photon and 18 GeV (22 GeV) for the subleading photon in the 2016 (2017 and 2018) dataset. Additionally, at the trigger level, a calorimetric selection is applied based on criteria such as the electromagnetic shower-shape, photon candidate isolation, and the ratio of its hadronic and electromagnetic energy deposits. The data samples used in the analysis are the same as the previous study in Chap. 3; details of the data samples can be found in Tab. **??**.

The signal process, pp $\to T'$ ($\to$ tH) bq, is generated at leading order using the Monte Carlo (MC) event generator MadGraph5_aMC@NLO 2.3.3 [100, 101]. Ten different mass points of T' ranging from 600 to 1200 GeV are considered in the generation process. The decay of the Higgs boson (H $\to \gamma\gamma$) is simulated using Pythia 8.2 [102]. The mass of the Higgs boson is set to 125.0 GeV, and the mass of the top quark is set to 172.5 GeV. The NNPDF3.0 parton distribution function is employed [66]. The signal samples are generated assuming left-handed chirality of the T' quark. The details of the T' samples, along with their corresponding cross sections, can be found in Tab. 4.1. For all MC samples generated with MadGraph5_aMC@NLO, the parton showering and hadronization have been implemented

via Pythia, with the underlying event tune CUETP8M1 [103] and CP5 [104] respectively for the 2016 and the 2017–2018 datasets.

T$'$ Mass (GeV)	NLO $\sigma_{T'bq} \times \mathcal{B}_{T' \to tH}$ (fb)
600	176.40
625	148.85
650	121.30
675	104.95
700	88.60
800	45.90
900	25.10
1000	14.50
1100	8.67
1200	5.36

Table 4.1: List of Monte Carlo signal samples used in the analysis and their corresponding NLO cross section [15]

The SM background processes contributing to the $m_{\gamma\gamma}$ spectrum ($\in [100, 180]$ GeV) can be classified into two categories: SM Higgs boson background (SMH) and SM non-resonant background (NRB). The SM Higgs boson production processes, including gluon fusion [105], vector-boson fusion [106], production in association with top quarks [107], or with a vector boson [108], are simulated using the MG5_aMC@NLO generator at next-to-leading order in QCD. The total cross sections and branching fractions have been adjusted according to the recommendations of the LHC Higgs boson cross section working group [23].

The nonresonent background processes t+X, t$\bar{\text{t}}$+X, Wγ, and Zγ are simulated using the MG5_aMC@NLO generator. Diboson events are generated at leading order using Pythia 8.205 [102]. The nonresonant diphoton samples are simulated with Sherpa 2.2.4 [68], which includes tree-level processes with up to three additional jets, as well as box diagrams. In all Monte Carlo (MC) samples, parton showering and hadronization are implemented using Pythia, using the underlying event tune CUETP8M1 [103] for the 2016 dataset and

CP5 [104] for the 2017-2018 datasets. The SM Higgs processes and the nonresonant simulated samples used in the analysis are the same as the previously mentioned in Sec. 3.2.

After generation, events are propagated through the simulation of the CMS detector using the Geant4 detector simulator [109]. The effects of multiple proton-proton interactions are also modeled in the package by adding simulated minimum-bias interactions to the simulated samples. The number of pileup interactions agrees with the one observed in data after reweighting the simulated pileup distribution.

4.3   Event reconstruction

The final state of the EW T$'$ production process consists of a Higgs boson, a top quark and another quark. The top quark decays to a bottom quark and W boson, with the W boson then decaying in either hadronic or leptonic modes. As mentioned before, in order to maximize the search sensitivity, the leptonic and hadronic decay modes of the top quark are handled as independent channels. Therefore, similarly as in Chapter. 3, the final state of the process may contain leptons (electrons, muons, taus), quarks, photons and neutrinos. All of them (apart from the neutrino) leave their mark on various subdetectors within the CMS detector. A simplified selection criteria for each individual channel is outlined here. More detailed information on object reconstruction, particularly for photons, utilized from the H → $\gamma\gamma$ measurements can be found in Sec. 3.3, 3.5 and 3.6 of Chap. 3. The HLT triggers used in this analysis are the same as the trigger used for the H → $\gamma\gamma$ analysis. Details of the triggers, along with their performance and validation, are discussed in Sec. 3.4

4.3.1   Leptonic channel

The channel is designed to tag the T$'$ → tH (H → $\gamma\gamma$) events with the top quark decaying leptonically. The event selection in this analysis category primarily requires the two photon candidates and at least one electron or muon present in the event. The additional criteria

applied to the associated objects in the events are mentioned in Tab. 3.6

Reconstruction of top quark

In the leptonic decay of the top quark, the presence of a neutrino introduces a missing component in the event reconstruction. Since there is only one neutrino in the final state of the T$'$ signal, the missing transverse momentum of the event is assumed to be equal to the transverse momentum of the neutrino. To determine the longitudinal momentum of the neutrino, an analytical solution is used, which can be derived as follows:

1. Applying energy momentum conservation at $W \to l\nu$ vertex,

$$p^W = p^l + p^\nu$$

$$p^{W2} = (p^l + p^\nu)^2$$

$$m^{W2} = m^{l2} + m^{\nu2} - 2p_x^l p_x^\nu - 2p_y^l p_y^\nu - 2p_z^l p_z^\nu + 2E^l\sqrt{(p_T^{\nu2} + p_z^{\nu2}) - m^{\nu2}}$$

2. By neglecting the masses of the lepton and neutrino in comparison to their energies, we can express the following equation:

$$m^{W2} = 2E^l\sqrt{p_T^{\nu2} + p_z^{\nu2}} - 2p_x^l p_x^\nu - 2p_y^l p_y^\nu - 2p_z^l p_z^\nu$$

$$E^l\sqrt{p_T^{\nu2} + p_z^{\nu2}} = \frac{m^{W2}}{2} + p_x^l p_x^\nu + p_y^l p_y^\nu + p_z^l p_z^\nu$$

$$= \Lambda + p_z^l p_z^\nu$$

where

$$\Lambda = \frac{m^{W2}}{2} + p_x^l p_x^\nu + p_y^l p_y^\nu = \frac{m^{W2}}{2} + \vec{p_T^l} \cdot \vec{p_T^\nu}$$

3. By solving the quadratic equation mentioned above, we can obtain two solutions for p_z.

$$p_z^\nu = \frac{\Lambda p_z^l}{p_T^{l\,2}} \pm \frac{1}{p_T^{l\,2}} \sqrt{\Lambda^2 p_z^{l\,2} - p_T^{l\,2}(E_l^2 p_T^{\nu\,2} - \Lambda^2)} \qquad (4.7)$$

The resolution of this equation can yield three potential outcomes depending on the terms inside the square root:

- Case 1: In this case, the equation produces real values for p_z. This occurs when the term inside the square root is positive, indicating that there are two distinct real roots. For example, for $M_{T'} = 600$ GeV case, there are around 73% of events corresponding to this case.

- Case 2: When the term inside the square root is negative, resulting in a complex value, the longitudinal momentum component p_z is also complex. This occurs under the assumption of the pole mass of the W boson (80.44 GeV). To address this, we employ a random generation process for the W boson mass, using the full Breit-Wigner function as the probability distribution. We then attempt to find a real solution for p_z by conducting trials. Remarkably, we succeed in finding a real solution for p_z in approximately 27% of the events after less than 1,000 trials.

- Case 3: If, even after 1,000 trials, the longitudinal momentum component p_z remains complex, it is assumed to be equal to the real part of the equation mentioned earlier, resulting in $p_z^\nu = \frac{\Lambda.p_z^l}{p_T^{l\,2}}$. Here, the mass of the W boson is fixed at 80.44 GeV. Events falling into this category constitute approximately 0.13% of the total.

A majority of events (case 1 and case 2) have two solutions for p_Z from the quadratic equation. The solution with lower absolute value is taken as an estimate of the longitudinal momentum of the neutrino. Combining the Lorentz vectors of neutrino, bottom quark, and the lepton we can calculate the invariant mass of the top quark. The T′ invariant mass can be reconstructed by combining the Lorentz vector of the top quark and the diphoton. Figure 4.4 illustrates the comparison between the p_z component of the neutrino in the reconstructed

level and the p_z component of the neutrinos at the generated level. The reconstructed top mass and the reconstructed mass of T′, $m_{T'}$ using this derived solution for p_z is shown in the figure 4.5, and 4.6. The distributions of every reconstructed T′ mass point and their mean and resolution are also summarized in Fig. 4.7 and Tab. 4.2, respectively.
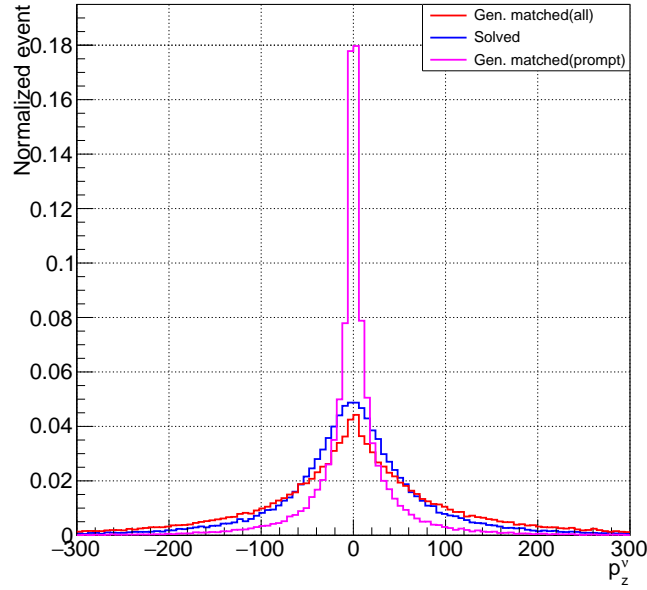
Figure 4.4: Comparison between the reconstructed p_z component of the neutrino, the p_z component of the neutrino from top quark decay at generator level, and the p_z sum over all final-state neutrinos at generator level.
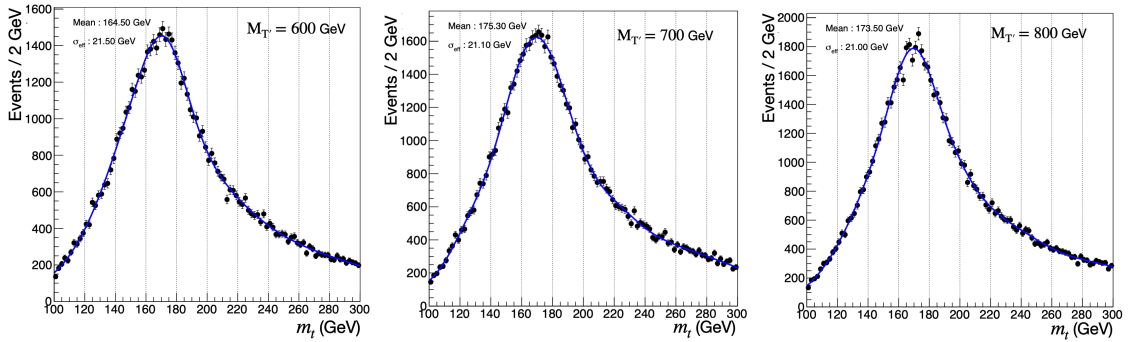
Figure 4.5: Reconstructed top quark mass using calculated neutrino p_z in simulated T′ quark samples with $M_{T'}$ = 600 (left), 700 (center) and 800 GeV (right).
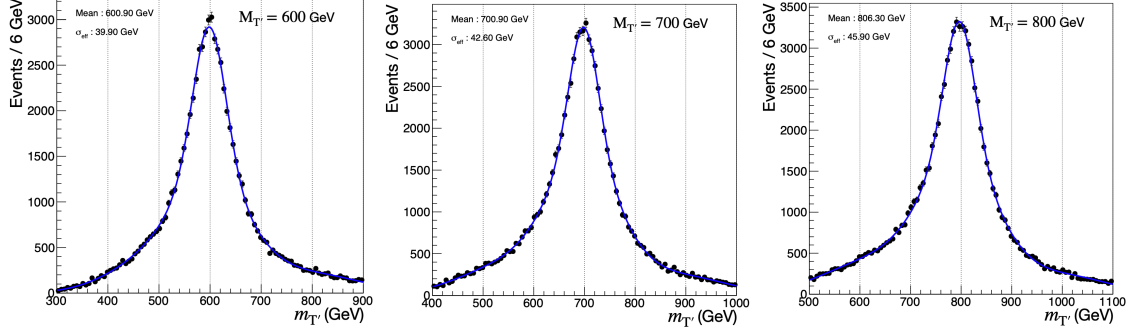
Figure 4.6: Reconstructed T′ mass in simulated T′ quark samples with $M_{T'}$ = 600 (left), 700 (center) and 800 GeV (right).
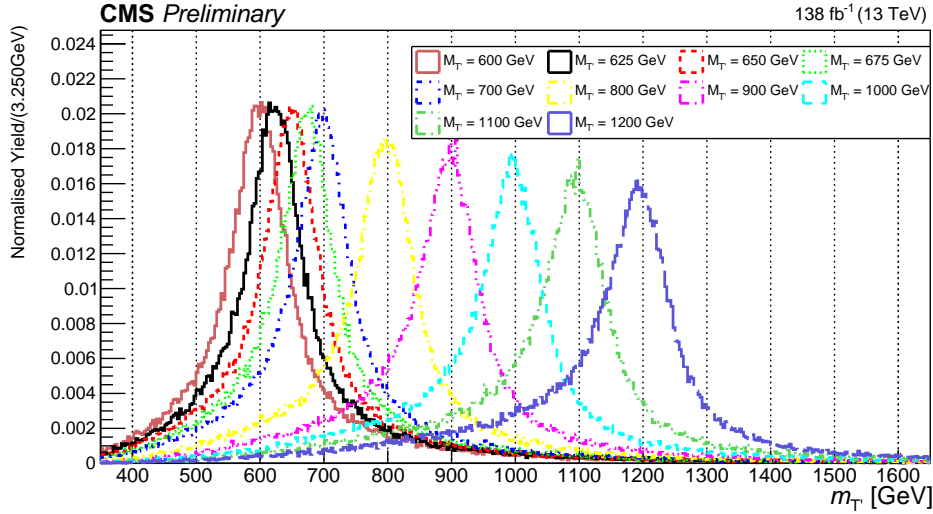
Figure 4.7: Reconstructed T′ mass of all simulated T′ mass points in the range [600, 1200] GeV. All the histograms are normalised to unity.

$M_{\mathrm{T}'}$(GeV)	Reconstructed	
	Mean (GeV)	Width (GeV)
600	600.9 ± 0.40	39.9 ± 0.38
625	614.5 ± 0.40	40.8 ± 0.38
650	646.7 ± 0.38	41.4 ± 0.36
675	675.9 ± 0.38	42.3 ± 0.36
700	700.9 ± 0.40	42.6 ± 0.38
800	806.3 ± 0.44	45.9 ± 0.44
900	897.9 ± 0.46	46.2 ± 0.45
1000	984.7 ± 0.53	47.4 ± 0.53
1100	1091.3 ± 0.65	48.6 ± 0.65
1200	1181.1 ± 0.60	49.2 ± 0.60

Table 4.2: Mean and resolution of reconstructed T$'$ mass. The mean and sigma are parameters from a Gaussian fit to M$_{\mathrm{T}'}$ distributions. "Reconstructed" represents T$'$ mass reconstructed from the mentioned analytical solution.

Background Rejection

After preselection, a significant amount of background is still present in the search region ($115 < m_{\gamma\gamma} < 135$ GeV). Hence a classifier is required to distinguish the signal from the backgrounds. Table 4.3 is a list of all the main standard model backgrounds and their respective yields. One dedicated boosted decision tree is trained to separate T$'$ from the SM Higgs process; it is also found to reject a sufficient amount of non-resonant background events (see figure 4.12 and figure 4.13). In training, the simulated T$'$ are treated as signals while the simulated Higgs boson production events from the standard model are treated as background. The configuration of the BDT training are explained in Sec. 3.8. To ensure an even training distribution across various T$'$ mass points in the relevant mass category, signal events are assigned a single flat weight, while background events are scaled by their respective cross sections. The event weights are subsequently renormalized to ensure that the total signal weights and the total background weights have the same value. The hyperpa-

rameters of the BDT training procedure are tuned for best performance and are summarized in Tab. 3.3.

The input features for the training include the kinematic variables of all the final state particles. Since the kinematic features largely depend on T$'$ mass, using one training for the whole mass range has been found sub-optimal in order to check a potential low T$'$ mass excess. Therefore training is performed for three different sets of T$'$ mass points [600, 625, 650, 675, 700], [800, 900, 1000], and [1100, 1200]. The list of variables used as the input features in the training and corresponding ranking of their importance is provided in Tab. 4.4. For each data-MC comparison plot below in Fig. 4.8, 4.9, 4.10, events in data and non-resonant background MC are blinded in the region $m_{\gamma\gamma} \in$ [115, 135] GeV. The non-resonant background MC distribution are normalized to the number of observed data events. All comparison plots shows an overall statistical agreement between data and simulation with a χ^2 p-value of 5%. The fluctuations observed in some of the distributions are associated with a limited simulated sample size.

Performance of BDTs

The discriminating power of the trained BDTs can be characterized by using the receiver operating characteristic (ROC) and area under the curve (AUC) as figures of merit. They are displayed in figure 4.11. The signal efficiencies at set background efficiency (1%, 10%, 30%) for the BDTs trained in each T$'$ mass category are shown in Tab. 4.5 to further quantify the performance. Figure 4.12 presents the BDT responses for various T$'$, SMH, and nonresonant background events. They show a good agreement between data and simulation. The efficiency of T$'$, Higgs boson, and non-resonant background events in various MVA score of three sets of training are shown in figure 4.13. The efficiency is calculated based on the number of events passed the preselection. The training is also effective in separating the background processes that are not Higgs-related ($\gamma\gamma$, γ+jet, $tt+\gamma\gamma$, $tt+\gamma$+jet, $t+\gamma$+jet),

Process	Yield	Frac. of total bkg.(%)
$\gamma\gamma$ + Jets	232.47	15.43
γ + Jets	136.14	9.04
$t\bar{t} + \gamma\gamma$	28.30	1.88
$t\bar{t} + \gamma$ + Jets	307.40	20.40
$t\bar{t}$ + Jets	419.93	27.87
$V + \gamma$	358.73	23.81
All non-resonant bkg.	1482.97	98.42
$t\bar{t}H$	19.04	1.26
ggH	0.71	0.05
VBF	0.16	0.01
VH	2.52	0.17
tHq	1.34	0.09
All SM Higgs bkg.	23.77	1.58
T' (600)	3.91	−
T' (625)	3.36	−
T' (650)	2.80	−
T' (675)	2.47	−
T' (700)	2.11	−
T' (800)	1.15	−
T' (900)	0.65	−
T' (1000)	0.38	−
T' (1100)	0.23	−
T' (1200)	0.14	−

Table 4.3: The expected yield of signal and background, as well as the corresponding fraction of total background by process for the leptonic channel preselections using MC only as background. The yields of non-resonant background, VLQ signal and SM Higgs processes are integrated in the region $m_{\gamma\gamma} \in [100, 180]$ GeV. The yields are corresponding to the integrated luminosity of 138 fb^{-1}

Rank	BDT-SMH [600,700]		BDT-SMH [800,1000]		BDT-SMH [1100,1200]	
	Variable	Importance	Variable	Importance	Variable	Importance
1	$p_T(\gamma_1)/m_{\gamma\gamma}$	7.024e-02	$p_T(\gamma_1)/m_{\gamma\gamma}$	6.879e-02	lead lepton η	6.146e-02
2	Scalar T′ p_T/All p_T	5.007e-02	$\eta(\gamma_1)$	5.484e-02	dipho_leadEta	5.758e-02
3	$p_T(\gamma_2)/m_{\gamma\gamma}$	4.631e-02	lead lepton η	5.204e-02	$p_T(\gamma_1)/m_{\gamma\gamma}$	5.546e-02
4	jet2 p_T	4.421e-02	$p_T(\gamma_2)/m_{\gamma\gamma}$	5.012e-02	ΔR (lepton, b-jet)	4.968e-02
5	ΔR(b-jet, forward jet)	4.201e-02	jet1 p_T	4.610e-02	jet1 p_T	4.515e-02
6	$\eta(\gamma_1)$	4.168e-02	Scalar T′ p_T/All p_T	4.210e-02	$p_T(\gamma_2)/m_{\gamma\gamma}$	4.217e-02
7	b-jet1 b-tagging score	4.023e-02	jet2 p_T	3.920e-02	b-jet1 η	4.065e-02
8	jet2 η	3.836e-02	ΔR(b-jet, forward jet)	3.915e-02	ΔR (lepton, forward jet)	3.794e-02
9	no. of central jets ($|\eta| < 1$)	3.739e-02	no. of central jets ($|\eta| < 1$)	3.663e-02	jet2 p_T	3.766e-02
10	ΔR (γ_1, forward jet)	3.713e-02	lead lepton p_T	3.621e-02	jet1 η	3.601e-02
11	lead lepton η	3.663e-02	b-jet1 η	3.513e-02	missing p_T	3.513e-02
12	jet1 η	3.657e-02	ΔR (lepton, forward jet)	3.479e-02	$\eta(\gamma_1)$	3.504e-02
13	ΔR (lepton, forward jet)	3.627e-02	ΔR (γ_1, forward jet)	3.448e-02	ΔR(b-jet, forward jet)	3.464e-02
14	jet1 p_T	3.405e-02	$\eta(\gamma_1)$	3.398e-02	lead lepton p_T	3.283e-02
15	lead lepton charge	3.333e-02	b-jet1 b-tagging score	3.233e-02	ΔR (γ_1, forward jet)	3.248e-02
16	forward jet p_T	3.187e-02	forward jet p_T	3.187e-02	forward jet p_T	3.116e-02
17	ΔR (lepton, b-jet)	3.118e-02	jet2 η	3.159e-02	b-jet1 b-tagging score	2.996e-02
18	IDMVA (γ_2)	3.074e-02	ΔR (lepton, b-jet)	2.943e-02	Scalar T′ p_T/All p_T	2.945e-02
19	no. of jets	3.060e-02	lead lepton charge	2.876e-02	ΔR (γ_2, forward jet)	2.731e-02
20	IDMVA(γ_1)	2.958e-02	jet1 η	2.832e-02	IDMVA(γ_2)	2.687e-02
21	ΔR (γ_2, forward jet)	2.931e-02	ΔR (γ_2, forward jet)	2.801e-02	jet2 η	2.674e-02
22	b-jet1 η	2.842e-02	missing p_T	2.718e-02	b-jet1 p_T	2.617e-02
23	missing p_T	2.819e-02	no. of jets	2.699e-02	no. of jets	2.587e-02
24	lead lepton p_T	2.788e-02	IDMVA(γ_1)	2.603e-02	IDMVA(γ_1)	2.556e-02
25	$\eta(\gamma_1)$	2.763e-02	IDMVA(γ_2)	2.474e-02	no. of central jets ($|\eta| < 1$)	2.481e-02
26	forward jet b-tagging score	2.426e-02	b-jet1 p_T	2.179e-02	ΔR (tH, forward jet)	2.340e-02
27	no. of b-jets	2.006e-02	no. of b-jets	2.119e-02	forward jet b-tagging score	2.242e-02
28	b-jet1 p_T	1.921e-02	forward jet b-tagging score	2.031e-02	lead lepton charge	1.868e-02
29	ΔR (tH, forward jet)	1.659e-02	ΔR (tH, forward jet)	1.788e-02	no. of b-jets	1.840e-02
30	haspixelseed(γ_1)	0.000e+00	haspixelseed(γ_1)	0.000e+00	haspixelseed(γ_1)	6.229e-03
31	haspixelseed(γ_2)	0.000e+00	haspixelseed(γ_2)	0.000e+00	haspixelseed(γ_2)	3.100e-03

Table 4.4: Leptonic Channel: Rank of feature importance for the trained BDT-SMH in each of the three mass categories. The value of importance is proportional to the frequency at which the variables are used to split decision tree nodes.
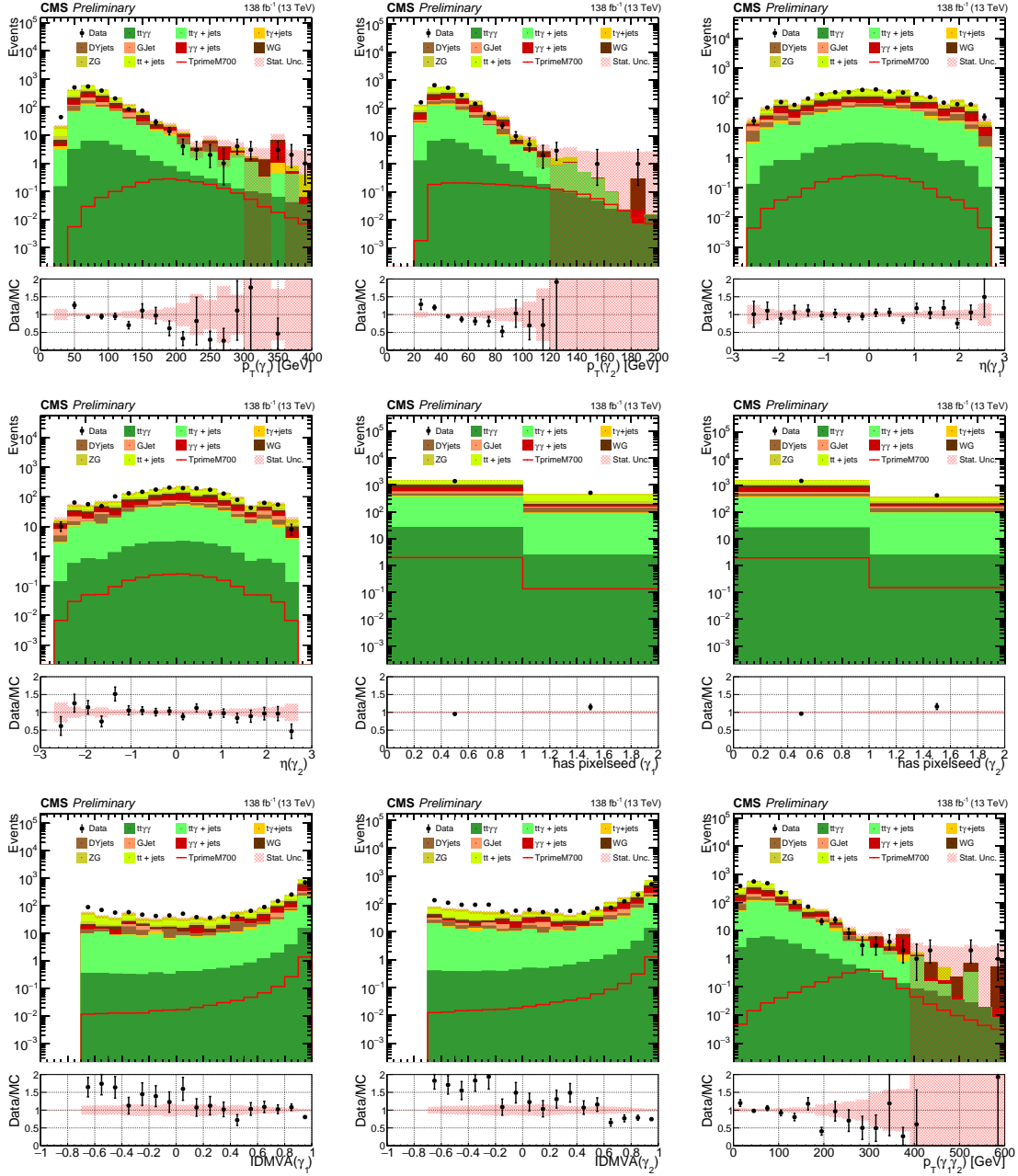
Figure 4.8: Comparison of data and simulation of input features to the BDTs in the VLQ leptonic channel. All simulated processes, T′ and backgrounds, are normalized according to the cross section and luminosity.
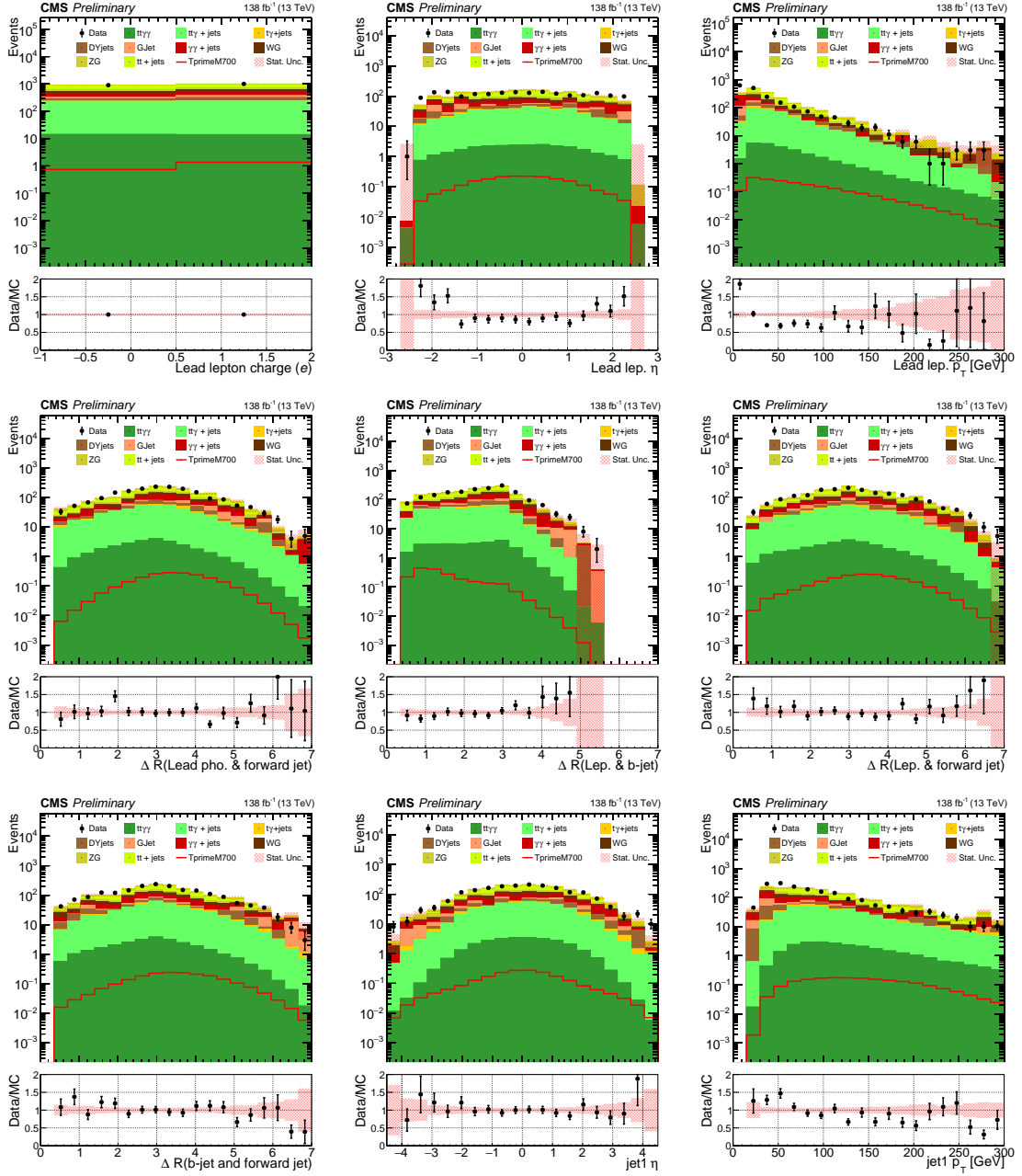
Figure 4.9: Comparison of data and simulation of input features to the BDTs in the VLQ leptonic channel. All simulated processes, T′ and backgrounds, are normalized according to the cross section and luminosity.
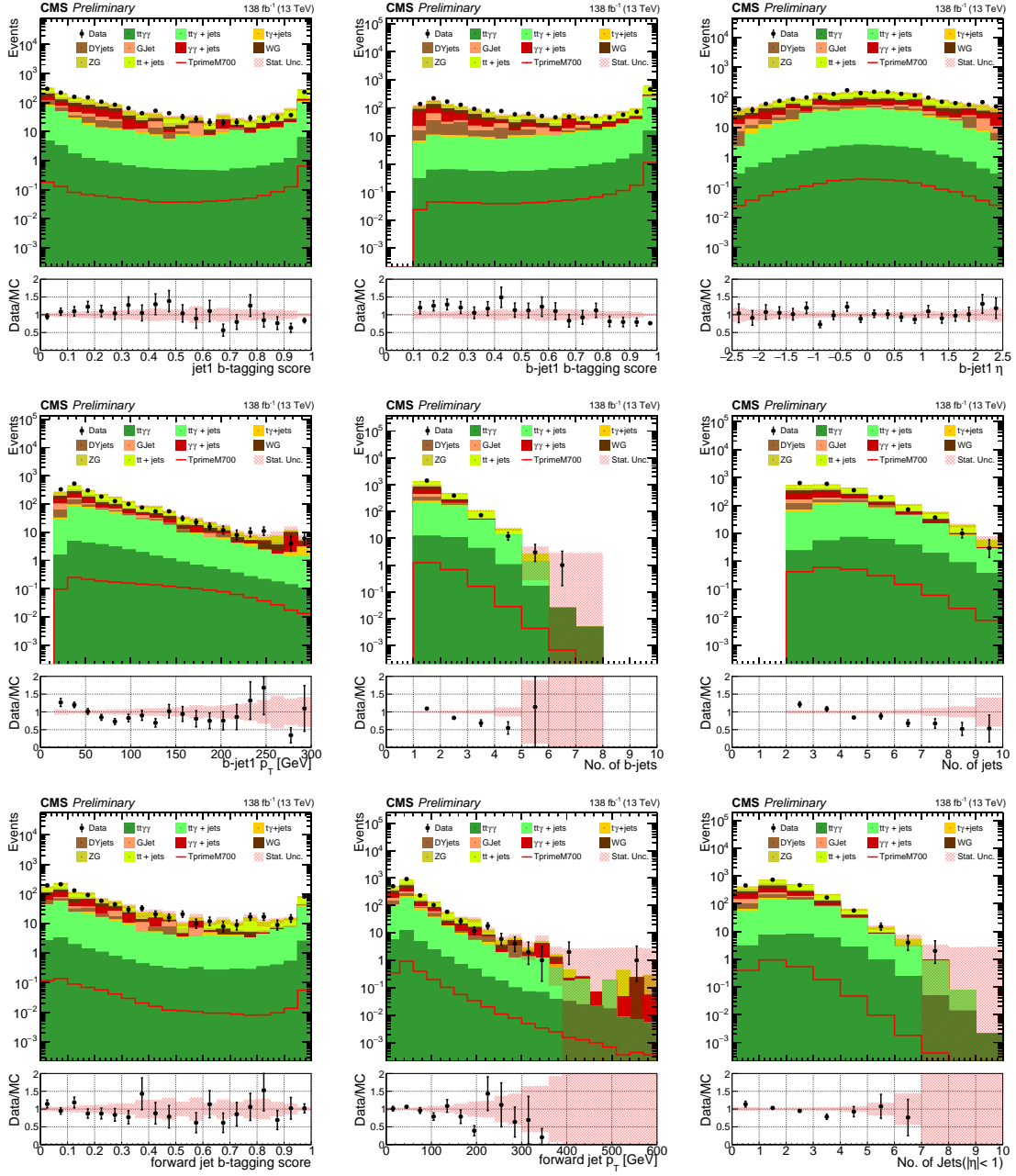
Figure 4.10: Comparison of data and simulation of input features to the BDTs in the VLQ leptonic channel. All simulated processes, T′ and backgrounds, are normalized according to the cross section and luminosity.

as seen in Fig. 4.12 and 4.13. Finally, the category is optimised based on the BDT score and the T' mass window to maximize the signal significance; the procedure is explained in Sec. 4.4.

Performance Metrics

$M_{T'}$ (GeV)	MVAs	AUC	Signal efficiency at fixed background efficiency		
			ϵ_{sig} ($\epsilon_{\text{bkg}} = 1\%$)	ϵ_{sig} ($\epsilon_{\text{bkg}} = 10\%$)	ϵ_{sig} ($\epsilon_{\text{bkg}} = 30\%$)
$[600, 700]$	BDT-SMH	0.923	0.299	0.757	0.945
$[800, 1000]$	BDT-SMH	0.969	0.470	0.926	0.994
$[1100, 1200]$	BDT-SMH	0.989	0.715	0.990	1.000

Table 4.5: Leptonic Channel: Signal efficiencies at various background efficiencies for BDT-SMH

4.3.2 Hadronic channel

Hadronic category considers signal events $T' \to tH$ (H $\to \gamma\gamma$) with the top quark decaying hadronically, $t \to bW(jj)$. It is defined to be orthogonal to the leptonic channel through a lepton veto, which means that events are required to have no lepton. The main backgrounds in this channel are QCD, $\gamma + \text{jets}$, and $\gamma\gamma + \text{jets}$. In addition, the SM Higgs boson production modes, $t\bar{t}H$, tH, *gg*H, VBF, and VH, have non-negligible contributions because they could have a similar final-state topology as the VLQ signal. Therefore, suppression of non-resonant background and differentiation of the VLQ signal from the SM Higgs processes are major tasks in this channel.

Pre-selection for the hadronic category

Object selection in this channel is same as for the leptonic channel mentioned in Tab 3.6. However, the event selection consists of the following criteria:

Background rejection vs Signal selection
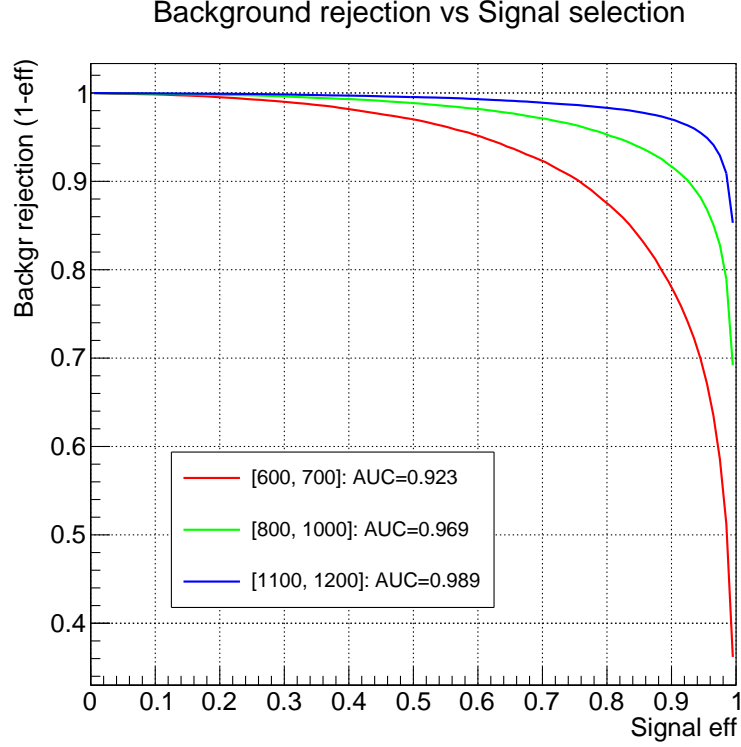
Figure 4.11: Leptonic Channel: Background efficiency vs. signal efficiency (ROC curves) for each of the BDTs used in signal region definition
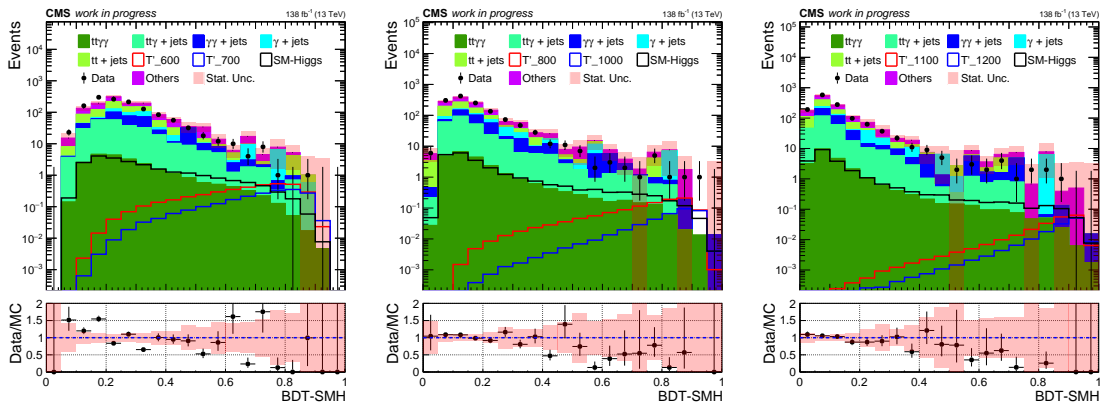
Figure 4.12: Leptonic Channel: The output scores of different events in the three T' mass range training. Events in data and non-resonant background MC are blinded in the region $m_{\gamma\gamma} \in [115, 135]$ GeV. Non-resonant background MC distributions are normalized to the number of observed data events.
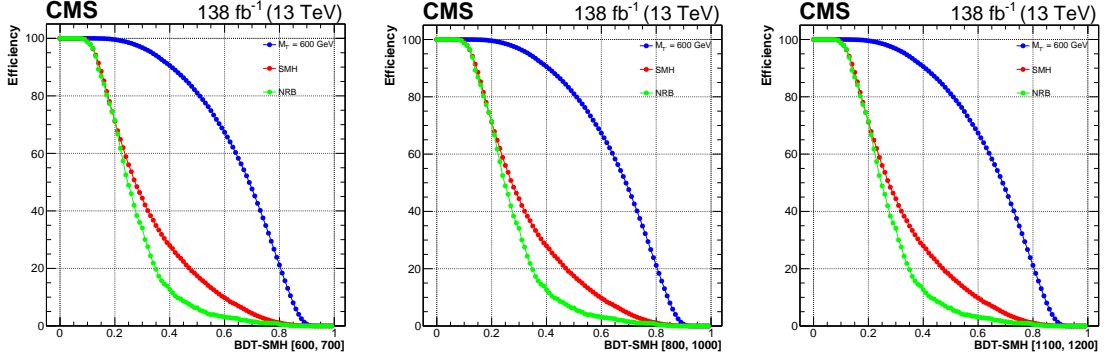
Figure 4.13: Leptonic Channel: The efficiency of T', Higgs and non-Higgs events w.r.t different MVA score of three sets of training is calculated based on the number of events passed the preselection. The top left plot corresponds to the training [600, 700] mass points, top right plot corresponds to the training [800, 1000] mass points and bottom plot corresponds to the training of [1100, 1200] mass points. It can be observed that the non-resonant background has a higher rejection rate than the SM-Higgs background because a wide range of variables help discriminate between VLQ signal and non-resonant backgrounds as well.

- Number of jets $>= 3$

- Number of leptons $= 0$

Background description and data-driven method

In this analysis, the non-resonant SM background is estimated through the data events in the range $m_{\gamma\gamma} \in [100, 180]$ GeV. Given that the non-resonant background is obtained from data in the final fit, a thorough understanding of the background processes and their composition is not important. Nevertheless, Monte Carlo (MC) samples provide the following advantages:

1. knowledge of the dominant background processes and their kinematics

2. training and optimizing classification algorithms

3. optimizing signal region boundaries

Therefore, in order to accurately model the backgrounds, a combination of simulation and data-driven techniques is employed. While some backgrounds are simulated, others such

as γ + jets and multi-jets, which are poorly described by the simulation, are estimated using data-driven methods. It is important to note that these γ + jets, $\gamma\gamma$ + jets and multi-jets account for approximately 95% of the total background. Specifically, the γ + jets and multi-jets backgrounds are estimated by inverting the selection on the photon ID MVA, allowing for the inclusion of events with misidentified photons contributing from γ + jets events. Since the simulation of the $\gamma\gamma$ + jets process is reliable only at the leading order, additional normalization factors are determined through a simultaneous template fit to the photon ID MVAs [76]. These normalization factors are then applied to both the data-driven samples (γ+jets and multi-jets) and the simulated $\gamma\gamma$ + jets samples. Table 4.6 and Tab. 4.7 shows the list of all the main standard model backgrounds and their respective yields in the sideband region, $m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV, before and after applying the data-driven description of γ + jets and multi-jet QCD processes, respectively. Also, the yields in the signal region, $m_{\gamma\gamma} \in [115, 135]$, after applying the data-driven description are shown in Tab. 4.8

Process	Yield	Frac. of total bkg. (%)
$\gamma\gamma$ + jets	34077 ± 412	48.35
QCD	21768 ± 13943	30.89
γ + jets	10579 ± 489	15.01
$t\bar{t} + \gamma\gamma$	503 ± 27	0.71
$t\bar{t} + \gamma$ + jets	1388 ± 59	1.97
$t\bar{t}$ + jets	1006 ± 71	1.43
V + γ	1152 ± 28	1.64
All non-resonant bkg.	70476 ± 13958	99.99
Data	115728 ± 340	164.20

Table 4.6: Hadronic Channel: Yields and fraction of total background by process using MC only as background. Events in data and non-resonant background are evaluated in the sideband region, $m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV. A large disagreement is observed between data and MC.

Process	Yield	Frac. of total bkg. (%)
$\gamma\gamma$ + Jets	34077 ± 412	29.42
(γ) + Jets (Data Sideband)	77697 ± 326	67.08
$t\bar{t} + \gamma\gamma$	503 ± 27	0.43
$t\bar{t} + \gamma$ + Jets	1388 ± 59	1.20
$t\bar{t}$ + Jets	1006 ± 71	0.87
V + γ	1152 ± 28	1.00
All non-resonant bkg.	115825 ± 535	100.00
Data	115728 ± 340	N/A

Table 4.7: Hadronic Channel: Yields and fraction of total background by process after application of the data-driven description of γ + jets and multi-jet QCD processes, which are denoted as (γ) + jets. Events in data and non-resonant background are evaluated in the sideband region, $m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV. The uncertainties reflect only the statistical uncertainty.

Reconstruction of the top and T' quarks

After the preselection in the hadronic channel, the QCD+X backgrounds (QCD, γ + jets, and $\gamma\gamma$ + jets) are the dominant processes. The VLQ signal will differ kinematically from these backgrounds due to the presence of a top quark. To leverage this difference, the top quark is reconstructed using a minimum χ^2 method, enabling the selection of the appropriate combination of jets. The method aims at picking up the jets that are more likely coming from top hadronic decay, $\mathrm{t} \to \mathrm{bW(jj)}$, with the on-shell mass constraints of W boson and top quark, as well as the covariance between each mass term being taken into account. The three jets coming from top decays are determined by minimum χ^2 sorting using a χ^2 formula defined by:

$$\chi^2 = \begin{pmatrix} m_{jj} - m_W \\ m_{bjj} - m_{top} \end{pmatrix}^T \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^{-1} \begin{pmatrix} m_{jj} - m_W \\ m_{bjj} - m_{top} \end{pmatrix}$$

where V is a 2×2 covariance matrix obtained from the simulated signal samples with MC-truth-matched particles. Each covariant term is estimated by the formula:

$$V_{\alpha\beta} = \mathrm{cov}(m_\alpha, m_\beta) = \frac{1}{N} \sum_{i=1}^{N} \left(m_\alpha^i - E(m_\alpha)\right) \left(m_\beta^i - E(m_\beta)\right)$$

Process	Yield	Frac. of total bkg. (%)
$\gamma\gamma$ + Jets	14990 ± 270	29.43
(γ) + Jets (Data Sideband)	33944 ± 216	66.64
$t\bar{t} + \gamma\gamma$	164 ± 18	0.32
$t\bar{t} + \gamma$ + Jets	600 ± 37	1.18
$t\bar{t}$ + Jets	425 ± 45	0.84
$V + \gamma$	477 ± 18	0.94
All non-resonant bkg.	50603 ± 352	99.35
$t\bar{t}H$	48.72 ± 0.31	0.10
ggH	214.82 ± 1.98	0.42
VBF	25.15 ± 0.33	0.05
VH	37.49 ± 0.33	0.07
tHq	6.21 ± 0.01	0.01
All SM Higgs bkg.	332.38 ± 2.06	0.65
T' (600)	18.04 ± 0.03	N/A
T' (625)	15.51 ± 0.03	N/A
T' (650)	12.95 ± 0.02	N/A
T' (675)	11.40 ± 0.02	N/A
T' (700)	9.79 ± 0.02	N/A
T' (800)	5.38 ± 0.01	N/A
T' (900)	3.07 ± 0.00	N/A
T' (1000)	1.84 ± 0.00	N/A
T' (1100)	1.12 ± 0.00	N/A
T' (1200)	0.71 ± 0.00	N/A

Table 4.8: Hadronic Channel: Yields and fraction of total background by process with data-driven description of γ + jets and multi-jet QCD processes, which are denoted as (γ) + jets. The yields of VLQ signal, SM Higgs and non-resonant background processes are integrated in the signal window, $m_{\gamma\gamma} \in [115, 135]$ GeV. The uncertainties reflect only the statistical uncertainty. Data events are blinded in the signal window.

where the notations of above formula are described as following:

- $\alpha, \beta \in \{1, 2\}$

- N is the number of signal events passing hadronic preselection with MC-truth-matching

  criteria applied.

- m_1 denotes m_{jj} with jets being MC-truth-matched particles

- m_2 denotes m_{bjj} with jets being MC-truth-matched particles

- $\text{E}(m_1)$ denotes the mean value of m_{jj}

- $\text{E}(m_2)$ denotes the mean value of m_{bjj}

With the χ^2 formula defined above, for each selected event, all the possible jet triplet combinations, one b-jet and two jets out of the rest reconstructed jets, are considered. The combination with the minimum χ^2 value is selected. After selecting the b-jet and w-jets that result in the minimum χ^2 value, a top quark can be reconstructed by summing up the four-momenta of the jets subsequently. Furthermore, with the derived kinematics of the top quark, we are able to reconstruct the vector-like quark ($T' \rightarrow tH(\gamma\gamma)$) by summing the four momenta of the top quark and the diphoton candidate. A set of kinematic variables can be derived from the reconstruction, including normalized transverse momentum ($p_T(T')/M_{bjj\gamma\gamma}$), pseudo-rapidity ($\eta_{T'}$), and helicity angle of the vector-like quark.

The used covariant matrix and the resulting minimum χ^2 values are shown in figure 4.14. The data and MC comparison for the reconstructed top and T′ masses are shown in figure 4.15. The distribution of each reconstructed T' mass and corresponding width are presented in figure 4.16 and Table 4.9. All comparison plots shows an overall statistical agreement between data and simulation with a χ^2 p-value of 5%. The fluctuations observed in some of the distributions are associated with a limited simulated sample size. The mean and width of every reconstructed T′ mass points are summarized in Tab. 4.9.

Background rejection

After preselection and reconstruction of the top and T′ quarks, a MVA is used to distinguish VLQ signal from background events. Discussions about input features, signal regions defined by BDT output scores, and performance are presented in the following sections sequentially.
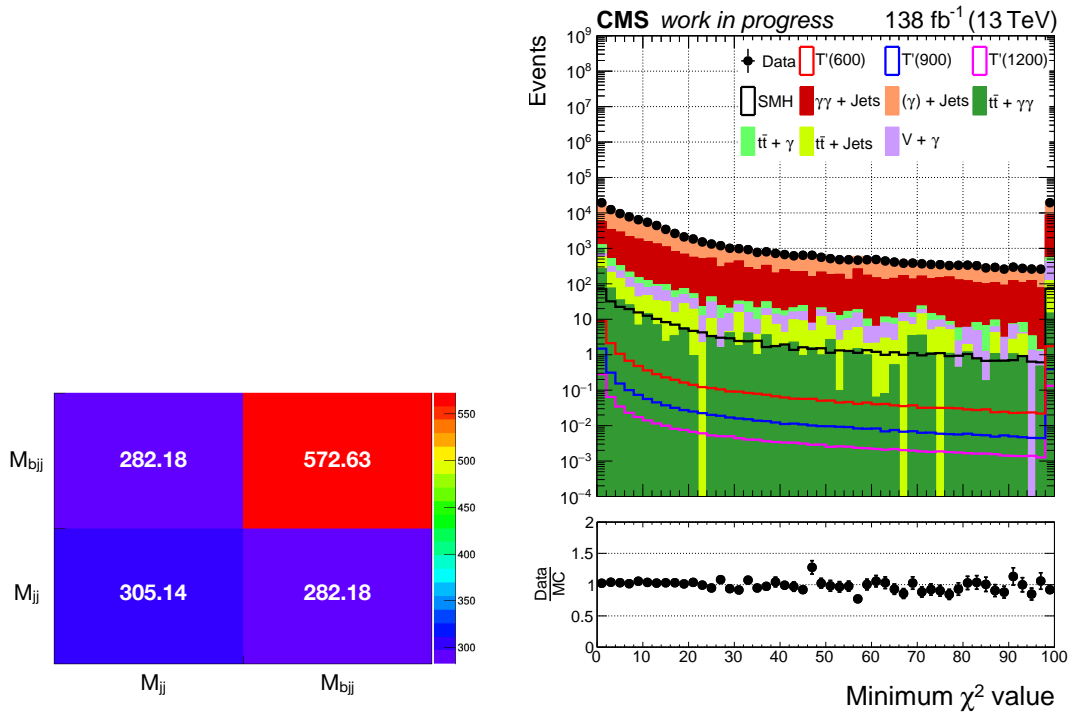
Figure 4.14: The covariance matrix used in the χ^2 formula and the minimum χ^2 distribution for events from data and simulation entering the hadronic channel preselections. Events in data and non-resonant background samples are blinded in the region $m_{\gamma\gamma} \in [115, 135]$ GeV. $\gamma\gamma$+jets and (γ) + jets are applied with additional scale factors from the data-driven study; while the other MC samples are normalized to an integrated luminosity of 138 fb^{-1}.
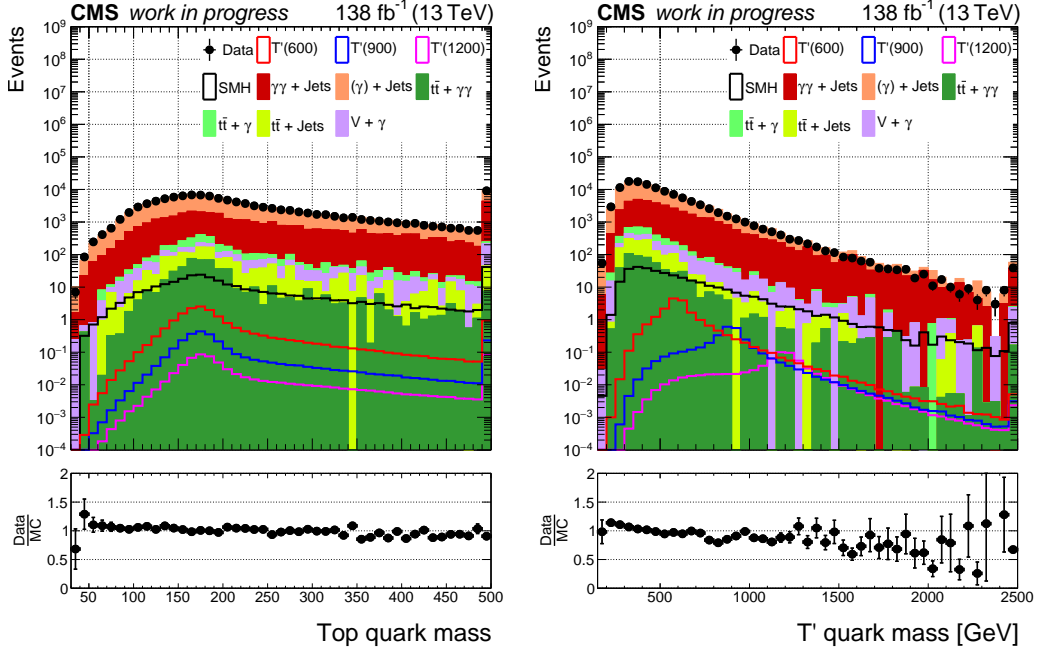
Figure 4.15: Comparison of data and simulation for the reconstructed mass of top quark and T' quark in the VLQ Hadronic channel. Events in data and non-resonant background samples are blinded in the region $m_{\gamma\gamma} \in [115, 135]$ GeV. $\gamma\gamma$+jets and (γ) + jets are applied with additional scale factors from the data-driven study; while the other MC samples are normalized to an integrated luminosity of 138 fb^{-1}.
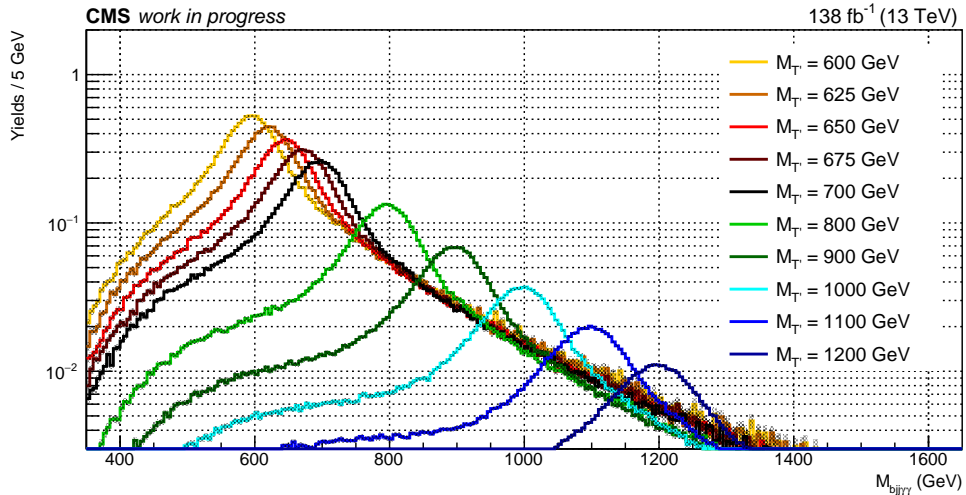
Figure 4.16: Reconstructed T′ mass in the VLQ hadronic channel. The scale of yields reflects magnitude of cross section with respect to T′ mass. The granularity of the distributions shows that the signal samples well cover the T′ mass range of [600, 1200] GeV.

$M_{\mathrm{T}'}$ (GeV)	Reconstructed	
	Mean (GeV)	Width (GeV)
600	594.5 ± 0.2	42.5 ± 0.3
625	619.4 ± 0.2	43.1 ± 0.3
650	644.4 ± 0.2	43.0 ± 0.3
675	670.1 ± 0.2	43.3 ± 0.3
700	694.6 ± 0.2	44.0 ± 0.3
800	794.3 ± 0.2	48.1 ± 0.3
900	894.8 ± 0.2	50.0 ± 0.3
1000	995.4 ± 0.2	53.3 ± 0.4
1100	1094.9 ± 0.2	61.5 ± 0.3
1200	1196.3 ± 0.3	65.0 ± 0.4

Table 4.9: Mean and resolution of T' mass. "Reconstructed" represents T' mass reconstructed from the χ^2 method; The mean and width are parameters from a Gaussian fit to $M_{\mathrm{T}'}$ distributions.

Input features A set of physics variables summarizing the kinematics of each event is used as input to the BDTs trained on the hadronic channel (shown in Fig. 4.17). These variables are expected to show some difference between VLQ signal and backgrounds. In addition, to prevent the BDTs from being sensitive to the diphoton invariant mass ($m_{\gamma\gamma}$) and reconstructed T' mass ($m_{bjj\gamma\gamma}$), the transverse momenta of the leading two photons and reconstructed T' quark are normalized by the corresponding mass terms. In the hadronic channel, multiple binary classification BDTs are trained to differentiate the VLQ signal from some subsets of the background processes. Then, the output scores of the BDTs are used to define a single signal region for the statistical analysis. In practice, this will be done by optimizing a threshold on BDT scores. The details of the training and the weight treatment of the training samples are explained in Sec. 3.8. To ensure an even training distribution across various T' mass, the scaling of the cross sections follow the same prescription mentioned for the leptonic channel.

Like the leptonic channel, three sets of T' mass ranges are defined: [600, 700] GeV, [800,

Objects		Variables			# Variables
PHOTON RELATED	leading photon	pT/mγγ	pixel seed veto	eta	3
	subleading photon	pT/mγγ	pixel seed veto	eta	3
	photon relevant	max IDMVA	min IDMVA	dR(photon, photon)	3
	di-photon	pT/mγγ	cos φ	| cos (helicity angle) |	3
		Rapidity			1
JETS RELATED	jet relevant	n_jets	n_bjets		2
		max b-tag score	min b-tag score	HT	3
	leading jet 1	pT / Mbjj	eta	b-tag score	3
	leading jet 2	pT / Mbjj	eta	b-tag score	3
	leading jet 3	pT / Mbjj	eta	b-tag score	3
	leading jet 4	pT / Mbjj	eta	b-tag score	3
	b-jet	pT / Mbjj	eta	b-tag score	3
	w-jet 1	pT / Mjj	eta	b-tag score	3
	w-jet 2	pT / Mjj	eta	b-tag score	3
	W boson	pT / Mjj	eta	Mjj	3
		dR(W boson, b-jet)	dR(w-jet1, w-jet2)	Minimum χ2 value	3
	top quark	pT / Mbjj	eta	Mbjj	3
OTHERS	T' quark	pT / Mbjjγγ	eta	(pT(bjj) + pT(γγ)) / HT	3
		| cos (helicity angle) |			1
	MET	MET			1
					53

Figure 4.17: The input features used for training BDTs in the hadronic channel. Here, helicity angle is the orientation of a spin of T' relative to its momentum direction and HT is the sum of the p_T of all the jets in the event

1000] GeV, and [1100, 1200] GeV. Given that the non-resonant background is the largest background component after preselection and that SM Higgs processes have non-negligible contributions in the $m_{\gamma\gamma}$ signal window [115, 135] GeV, we train two BDTs to suppress the two different types of background in each T′ mass category. Therefore, there are 6 BDTs in total in the study of the hadronic channel. The two types of BDTs are summarized as the following:

- Non-resonant background BDT: BDT-NRB

  – Signal: single vector-like quark production

  – Background: (γ) + jets, $\gamma\gamma$ + jets, tt+X, and $V + \gamma$

- SM Higgs BDT: BDT-SMH

  – Signal: single vector-like quark production

  – Background: t$\bar{\text{t}}$H, tH, ggH, VBF, and VH

The BDT responses, BDT-NRB and BDT-SMH, are presented in Figs. 4.18 and 4.19. They show a good agreement between data and simulation. The optimization of the signal region is performed based on the BDT-NRB and BDT-SMH variables. Further details regarding the optimization of the signal region can be found in Sec. 4.4.

4.4 Signal region optimization

The selection criteria on the BDT discriminants for the three different $M_{T′}$ ranges have been optimized to maximize signal sensitivity with respect to the background. The optimization is carried out independently for the leptonic and hadronic categories, with each category being optimized by minimizing a χ^2-like estimator.

$$\chi^2 = -\frac{S}{\sqrt{B}} + \text{Pen} \tag{4.8}$$
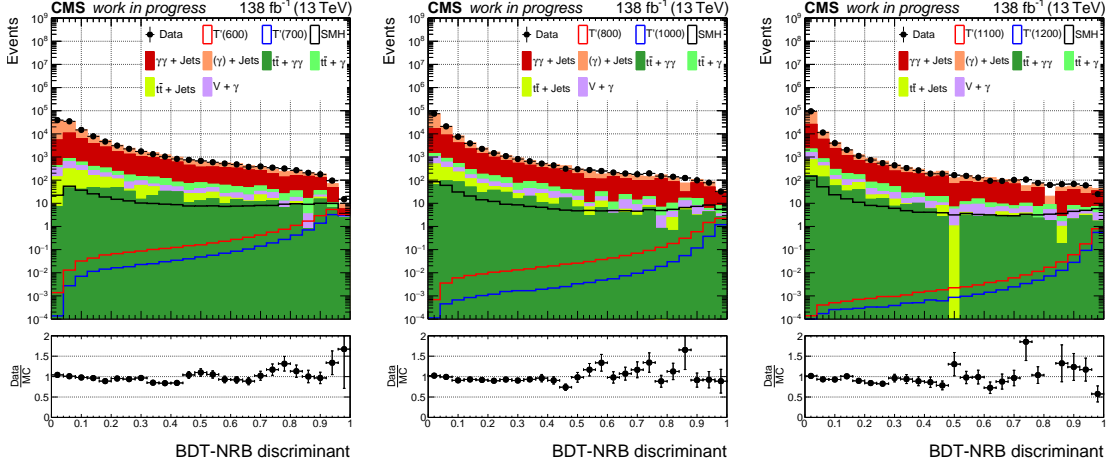
Figure 4.18: Hadronic Channel: BDT-NRB distributions for events from data and simulation entering the hadronic channel preselection. The BDTs are trained with T′ samples in the mass category of [600, 700] (left), [800, 1000] (middle), and [1100, 1200] (right). Events in the data and non-resonant background samples are blinded in the region $m_{\gamma\gamma} \in$ [115, 135] GeV. Additional scale factors from the data-driven study are applied to the $\gamma\gamma$ + jets and γ + jets samples, while the other MC samples are normalized to an integrated luminosity of 138 fb^{-1}.
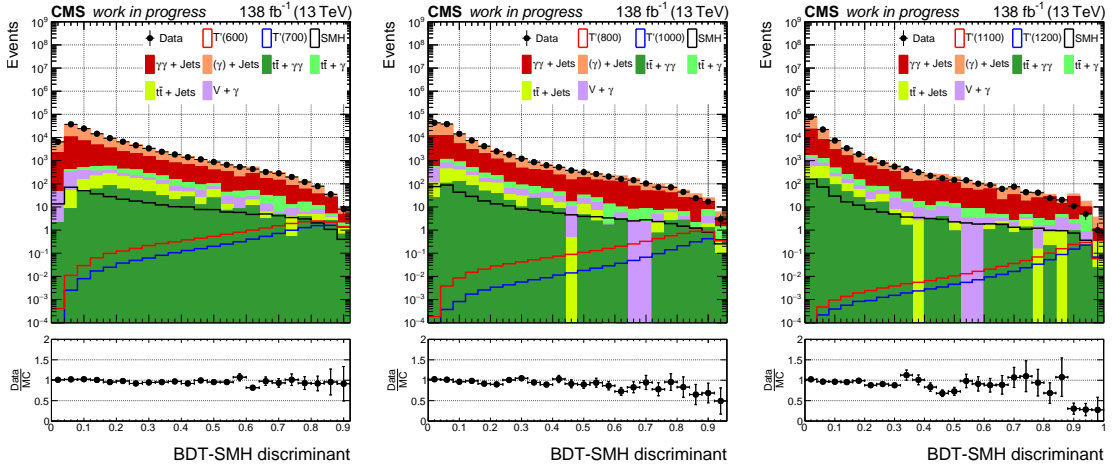
Figure 4.19: Hadronic Channel: BDT-SMH distributions for events from data and simulation entering the hadronic channel preselection. The BDTs are trained with T′ samples in the mass category of [600, 700] (left), [800, 1000] (middle), and [1100, 1200] (right). Events in the data and non-resonant background samples are blinded in the region $m_{\gamma\gamma} \in$ [115, 135] GeV. Additional scale factors from the data-driven study are applied to the $\gamma\gamma$ + jets and γ + jets samples, while the other MC samples are normalized to an integrated luminosity of 138 fb^{-1}.

The signal, denoted as S, is obtained by integrating the signal MC sample over the range of $100 < m_{\gamma\gamma} < 180$ GeV, and is weighted according to the cross section and luminosity. The background, denoted as B, is considered either in the side-band region ($100 < m_{\gamma\gamma} <$ 115 GeV or $135 < m_{\gamma\gamma} < 180$ GeV) when data are used, or in the full range of $100 < m_{\gamma\gamma} < 180$ GeV when MC background is considered. To ensure the statistical reliability of modeling the NRB from data, having a minimum number of events (B_{min}) in the signal sideband regions of $m_{\gamma\gamma}$ is necessary. In our study, we maintained a minimum of eight events in these regions to minimize the risk of fit failure. Therefore, a penalty term, denoted as "Pen", is introduced. The Pen term is proportional to the ratio $S/\sqrt{B}$, multiplied by a stability factor (0.1) chosen for minimization stability.

$$\text{Pen} = 0.1 \times \ \min(0, B - B_{\min}) \times (B - B_{\min}) \times \frac{S}{\sqrt{B}} \tag{4.9}$$

It is essential to determine the optimal combination of cuts in two or three dimensions. The optimization process takes into account a minimal cut on BDT scores and a loose $M_{T'}$ window for both categories. To simplify the structure, the optimization is performed in specific $M_{T'}$ range BDTs, namely BDT I, II, and III. The optimized selection criteria for the analyses are presented in Tab. 4.10 , while the signal efficiency in each category can be found in Fig. 4.20. The expected yields of the VLQ, nonresonant background, and SM Higgs boson background processes within the signal region (SR), $m_{\gamma\gamma} \in [115, 135]$, are provided in Tab. 4.11 along with the corresponding signal windows. The yields for different background contributions in individual channels and each T$'$ masses with $m_{\gamma\gamma} \in [115, 135]$ are presented in Tab. 4.14 and Tab. 4.15.

An event with a signature consistent with a vector-like quark (T$'$) produced and decaying to a top quark and a Higgs boson as displayed in Fig. 4.21. The top quark decays to a b quark and a W boson, producing three jets, one from the b and the other two from the decay of the W boson. The Higgs boson decays to two photons.

BDT	I	II	III
M$_{\mathrm{T}'}$ (GeV)	[600, 700]	[700, 1000]	[1000, 1200]
Hadronic analysis			
BDT-NRB score	> 0.94	>0.96	>0.95
BDT-SMH score	>0.80	>0.80	>0.80
m_{tH} window	[480, 800]	[550, 1150]	[650, 1600]
Leptonic analysis			
BDT score	>0.60	>0.40	>0.40
m_{tH} window	[480, 800]	[550, 1150]	[650, 1600]

Table 4.10: Signal selection criteria for the three BDTs and $m_{\mathrm{T}'}$ windows.

BDT	M$_{\mathrm{T}'}$ (GeV)	Leptonic yield T$'$	Nonres. bkgd.	SM H bkgd.	Obs.	Hadronic yield T$'$	Nonres. bkgd.	SM H bkgd.	Obs.
	600	1.7				3.2			
	625	1.7				3.5			
I	650	1.6	11.0 ± 9.0	1.3 ± 0.1	1	3.6	1.6 ± 0.9	1.8 ± 0.1	4
	675	1.6				3.7			
	700	1.5				3.6			
	800	1.6				2.9			
II	900	1.2	19.0 ± 14.4	2.3 ± 0.1	16	3.0	7.3 ± 4.0	2.0 ± 0.1	6
	1000	0.8				2.5			
III	1100	0.7	14.4 ± 13.7	1.4 ± 0.1	10	2.3	9.0 ± 5.3	2.4 ± 0.2	7
	1200	0.5				1.8			

Table 4.11: The expected yields of different processes in each signal window for events with a T$'$ with mass in the range M$_{\mathrm{T}'} \in [600, 1200]$ GeV, and the observed number of events in the signal region $m_{\gamma\gamma} \in [115, 135]$. Here, the yields for the T$'$ are for κ_{T} fixed at 0.2.
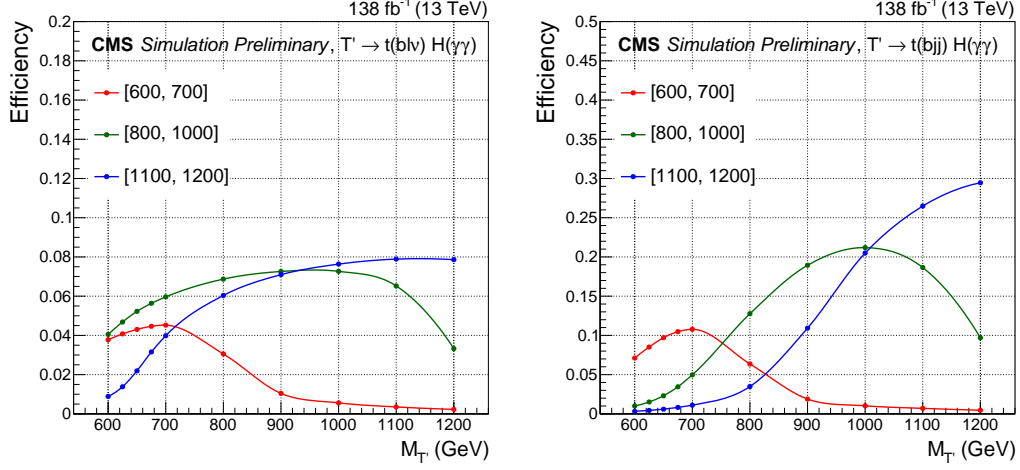
Figure 4.20: Signal efficiency in the optimized signal regions for the leptonic channel (left) and hadronic channel (right). The efficiency is defined as yields in the optimized region divided by expected yields, $\mathcal{L} \times \sigma \times \mathcal{B}(\mathrm{H} \to \gamma\gamma)$.

4.5 Signal & background models

4.5.1 Signal and Resonant background models

Models of signal and the SM Higgs background processes are created by fitting the $m_{\gamma\gamma}$ distribution from MC samples with a sum of at most five Gaussian functions. To capture any year-dependent dependencies of the $m_{\gamma\gamma}$ resolution, a separate fit of each process is performed for each of the three years during Run-II period. The resulting three models for each process are summed together (with proper scale factors accounting for luminosity) to create the final model of the process. As a validation of the procedure, an extensive search for the Higgs boson is done in a control region. The complete study is explained in Sec. 4.8. Signal processes are modelled separately for the leptonic and hadronic event categories. The analysis is performed for the 10 different T′ mass hypotheses. Therefore each mass hypothesis has a separate signal model. The diphoton invariant mass ($m_{\gamma\gamma}$) distributions from 100-180 GeV are fitted using a sum of at most five Gaussian functions. The χ^2 per degrees of freedom is evaluated in each signal model to find the best order of the Gaussian
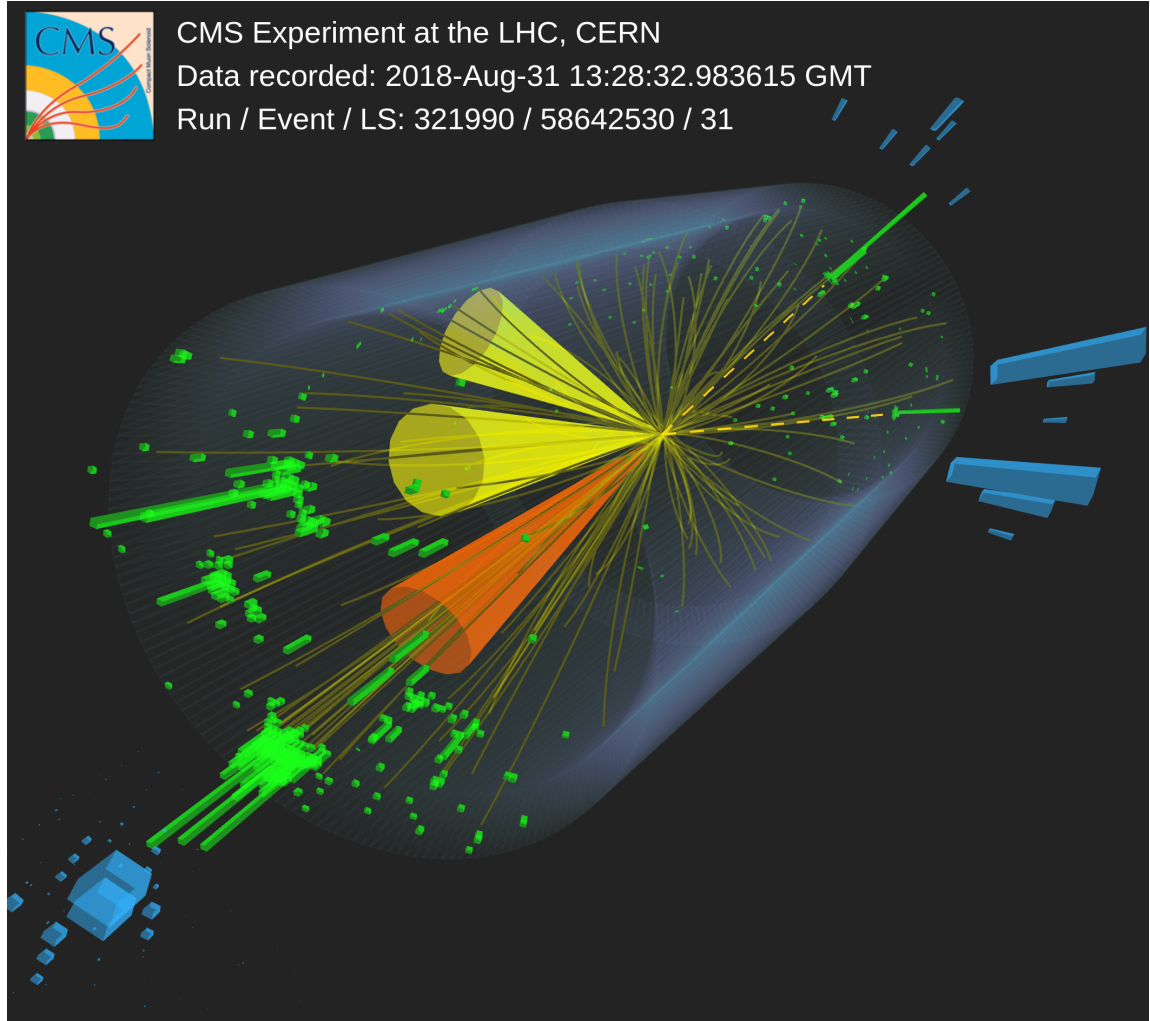
Figure 4.21: An event with a signature consistent with a vector-like quark (T') produced and decaying to a top quark and a Higgs boson. The top quark decays to a b quark and a W boson, producing three jets, one from the b and the other two from the decay of the W boson. The Higgs boson decays to two photons. The three jets are represented by the cones where the orange cone represents the b jet and the yellow cones the jets from the W boson decay. The photons are represented by the dashed lines; their energy deposit in the electromagnetic calorimeter is shown by the green boxes.

fit to simulated signal events.

Systematic uncertainties which modify the $m_{\gamma\gamma}$ distribution, associated with the smearing and scale of the individual photon energies, are incorporated into the signal model as nuisance parameters. The signal models for each T$'$ mass point in [600, 1200] GeV in both leptonic and hadronic channels are displayed in Fig. 4.22–4.23. The model of the ttH and tHq production processes are displayed in Fig 4.24 and 4.25, respectively

4.5.2 Non-resonant background model

The non-resonant background processes are modelled using the $m_{\gamma\gamma}$ spectrum from data, $m_{\gamma\gamma} \in [100, 180]$ GeV. Since the exact shape of the $m_{\gamma\gamma}$ distribution is not known and choosing any particular function would lead to a potential bias, a set of functions is employed to fit the $m_{\gamma\gamma}$ sideband region. The functions used to fit the distribution are mentioned in Sec. 3.9.2. The choice of the background function is then treated as a discrete nuisance parameter in the final fit with the associated systematic uncertainty estimated by the discrete profiling method [61].

In most situations, higher-order functions would give a better fit. To prevent over-fitting, an F-test [110] is performed for each family of functions when modelling non-resonant background processes. Figures 4.26, 4.27, and 4.28 present the resulting fitting curves for both leptonic and hadronic channels in different training ranges of $M_{T'}$ = [600–700], [800–1000] and [1100–1200], respectively.

Figure 4.32–4.35 presents the overlaid signal (μ = 1) and background model for the leptonic and hadronic channel of different mass points. The red line in the plot is the best-fitted background model and the blue line is the best-fitted signal model. The peak in the background component shows the considered irreducible SM Higgs boson (ggH, VBF, VH, ttH and tH) contribution.
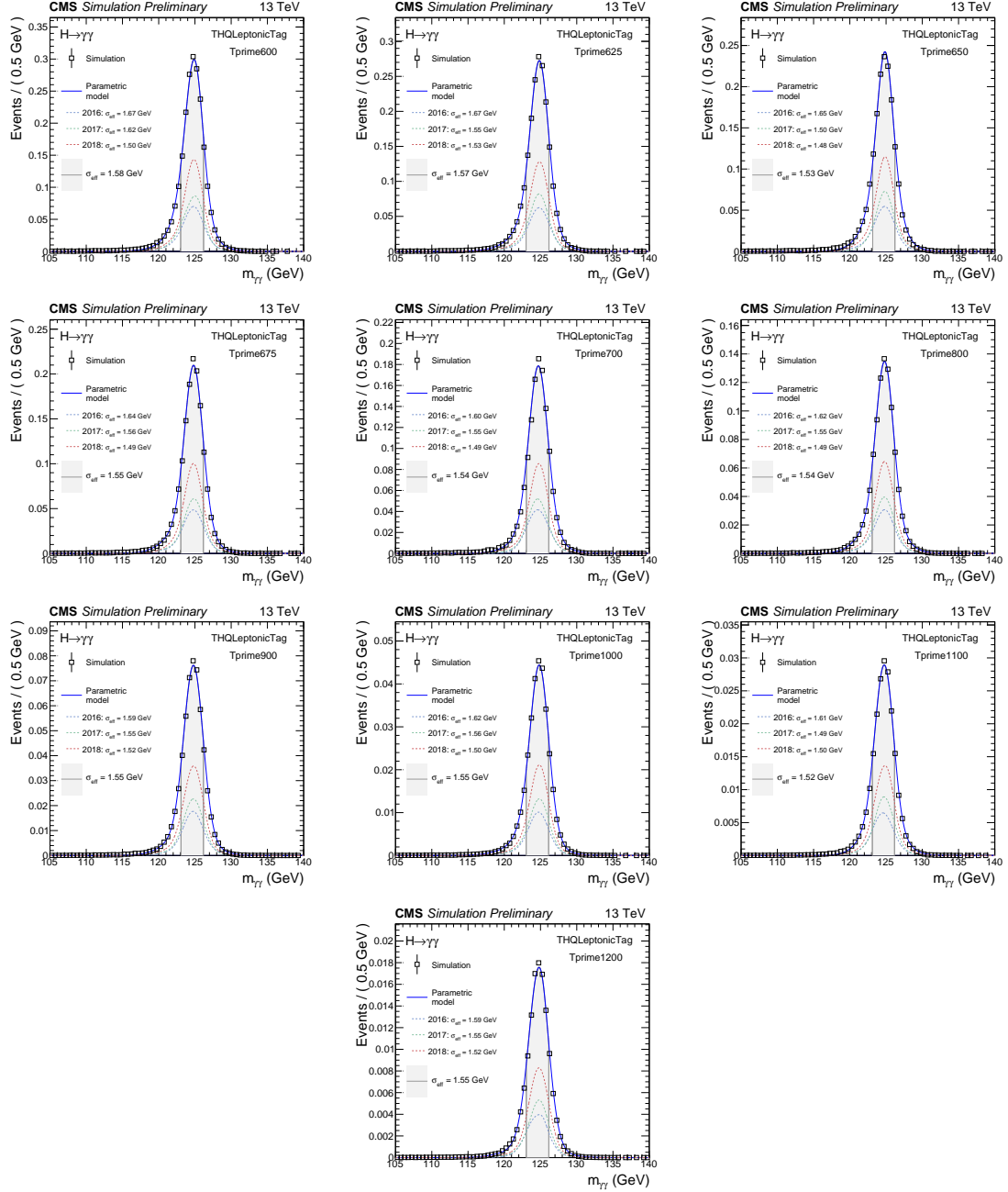
Figure 4.22: Signal model for the events passing the leptonic category selection for the ten $M_{T'}$ mass points from 600 to 1200 GeV. The legend "THQLeptonicTag" in the figure refers to the leptonic channel.
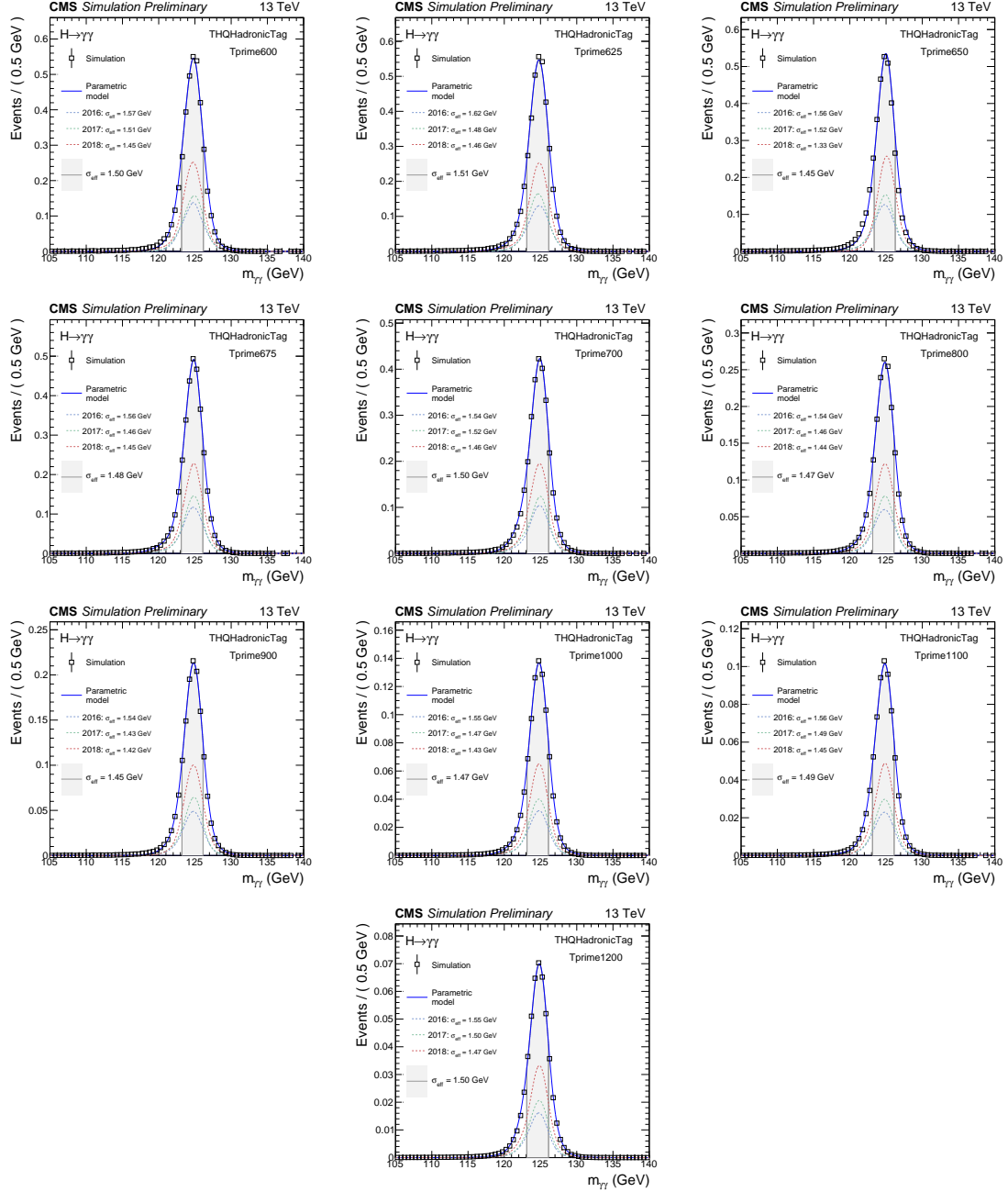
Figure 4.23: Signal model for the event passing the hadronic category selection for the ten $M_{T'}$ mass points from 600 to 1200 GeV. The legend "THQHadronicTag" in the figure refers to the hadronic channel.
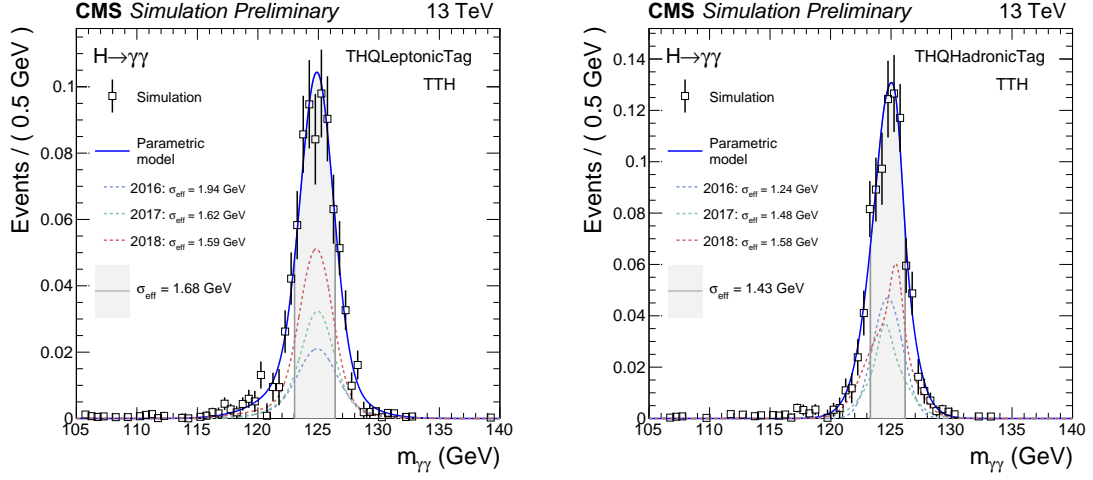
Figure 4.24: Model for the ttH SM Higgs background which passing the leptonic (left) and the hadronic (right) signal regions defined by BDT(s) trained in T′ mass range of [600-700] GeV.
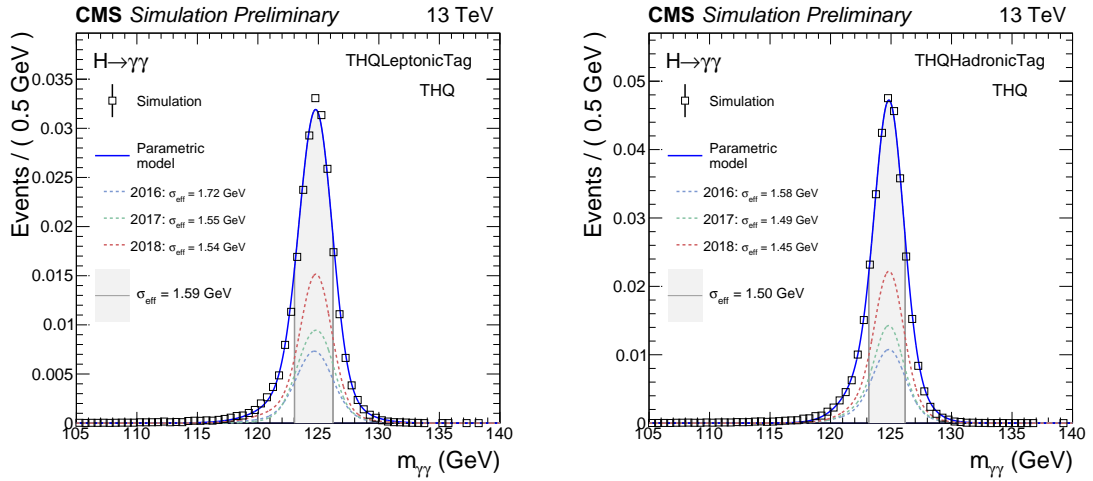
Figure 4.25: Model for the tHq SM Higgs background which passing the leptonic (left) and the hadronic (right) signal regions defined by BDT(s) trained in T′ mass range of [600-700] GeV.
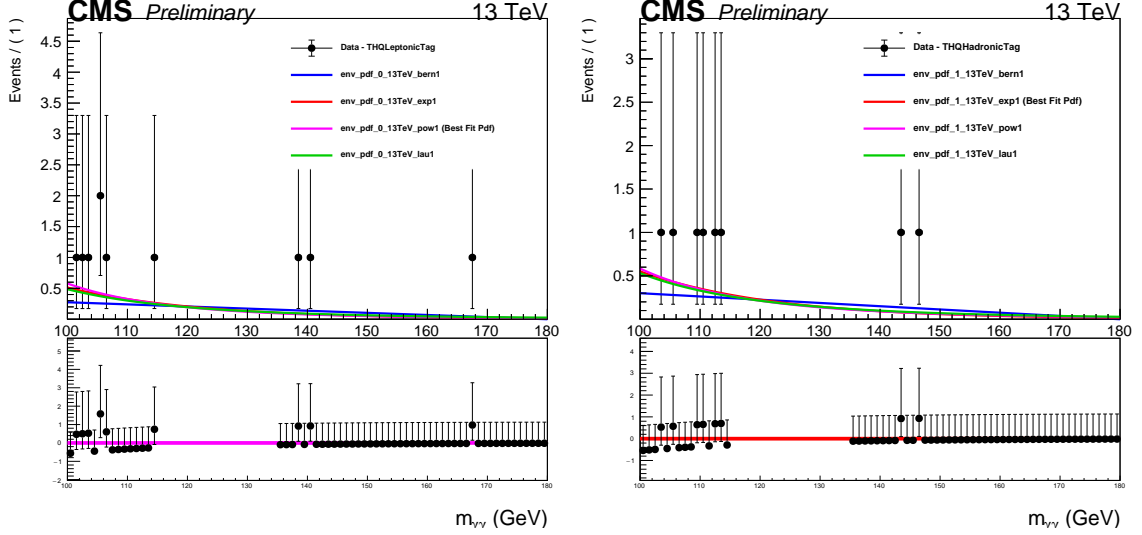
Figure 4.26: Model for non-resonant background which passes the leptonic (left) and the hadronic (right) signal regions defined by BDT(s) trained in T′ mass range of [600-700] GeV. The solid colored lines represent the distinct functions fitted to the dataset. The differences between the observed data and the background component (best-fit probability density function) after subtraction are shown in lower panel.
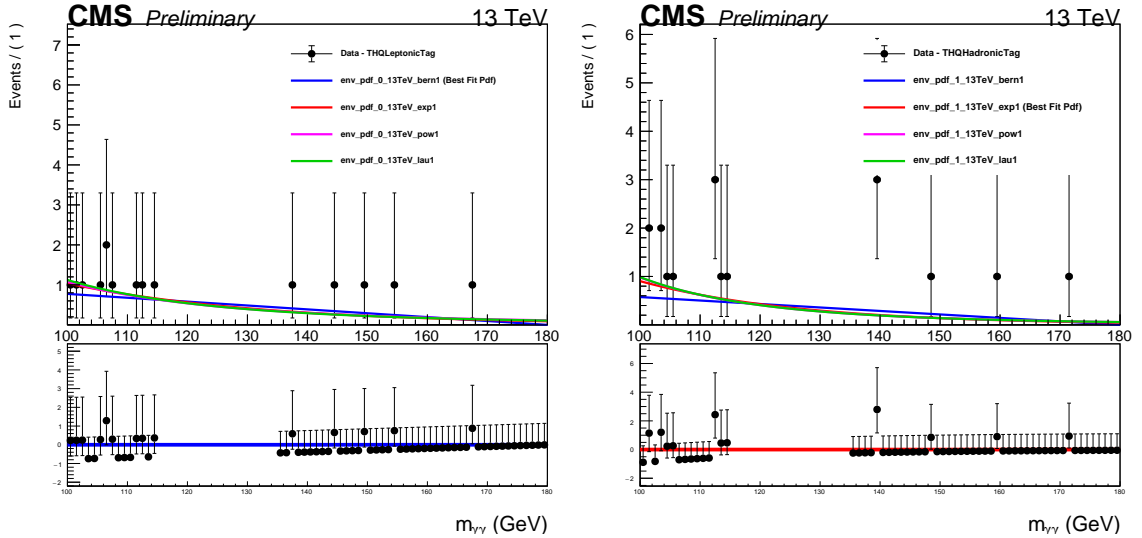
Figure 4.27: Model for non-resonant background which passes the leptonic (left) and the hadronic (right) signal regions defined by BDT(s) trained in T′ mass range of [800-1000] GeV. The solid colored lines represent the distinct functions fitted to the dataset. The differences between the observed data and the background component (best-fit probability density function) after subtraction are shown in lower panel.

179
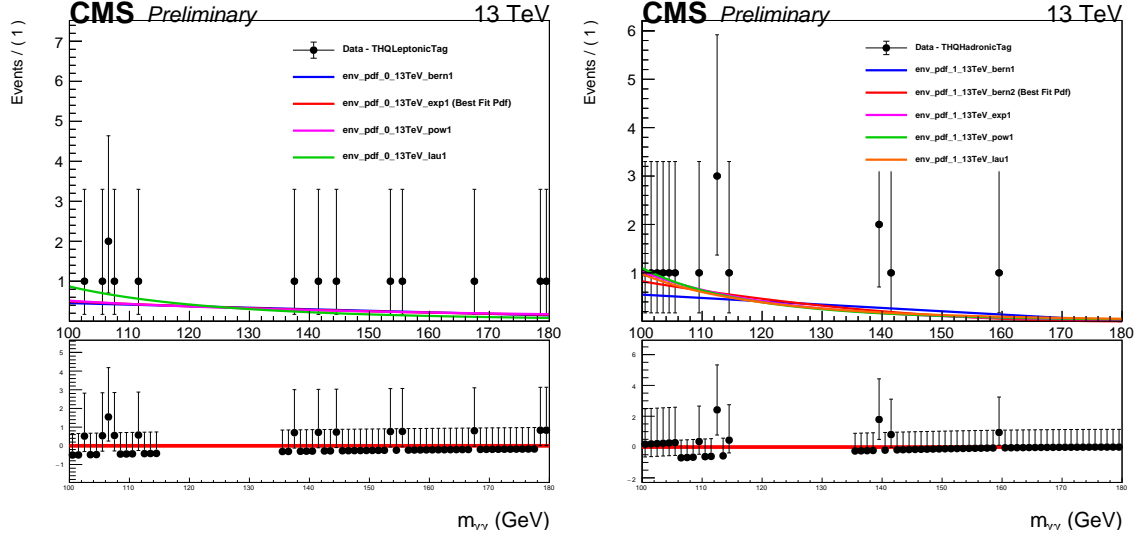
Figure 4.28: Model for non-resonant background which passes the leptonic (left) and the hadronic (right) signal regions defined by BDT(s) trained in T′ mass range of [1100-1200] GeV. The solid colored lines represent the distinct functions fitted to the dataset. The differences between the observed data and the background component (best-fit probability density function) after subtraction are shown in lower panel.

4.6 Bias studies

The likelihood function is created using the combined Poissonian probability and the signal plus background probability distribution function as a function of our parameter of interest (signal strength, μ) as defined in Eqn. 3.10. The bias test attempts to evaluate any potential bias in the model for a certain value of μ. Here, the test is performed for leptonic and hadronic modes combined; however, this test can also be done separately for each channel.

In bias studies, a toy dataset is generated ("thrown") using a selected background function, and then fitted using a different background function. These toy datasets are generated with a predetermined signal strength value (e.g., μ=1), referred to as μ_{truth}. The fitted value of the signal strength is denoted as μ_{fit}, with some associated uncertainty σ_{fit}. To assess the agreement between the fitted and true signal strength, a pull value (P) is computed for each

toy dataset using the following formula:

$$P = \frac{(\mu_{\text{truth}} - \mu_{\text{fit}})}{\sigma_{\text{fit}}}$$

The following steps outline the process:

- The study is conducted using various injected values of the signal strength, denoted as μ, ranging from 1 to 10.

- For each μ value, 1000 pseudoexperiments (toys) are generated.

- The generated pseudoexperiments are then fitted using the signal plus background hypothesis.

- The bias is calculated using the above formula, P.

Pull distributions of μ are fitted with the Gaussian function to extract the mean. The pull distributions, $(\mu_{\text{fit}} - \mu_{\text{truth}})$, are analyzed by fitting them with a Gaussian function to determine the mean. Figure 4.29 illustrates the pull distributions $(\mu_{\text{fit}} - \mu_{\text{truth}})$ for different cases where μ_{truth} takes values of 1, 5, and 10. The similar plots of the pull distributions for all other μ_{truth} points are created, but only a subset of three points is shown in the figure here. The bias is evaluated for each T′ mass point model. The outcomes of the bias test, covering mass points ranging from 600 to 1200 GeV, are presented in Fig. 4.30 and 4.31. It is worth noting that no biases exceeding 20% of σ_{fit} (the fitted uncertainty) are observed in the study.

4.7 Systematic uncertainties

The systematic uncertainty associated with the data-driven background estimation are accounted for by fitting the sideband data with different functions treating the choice of function as a discrete nuisance parameter [61]. The uncertainties which influence the shape of
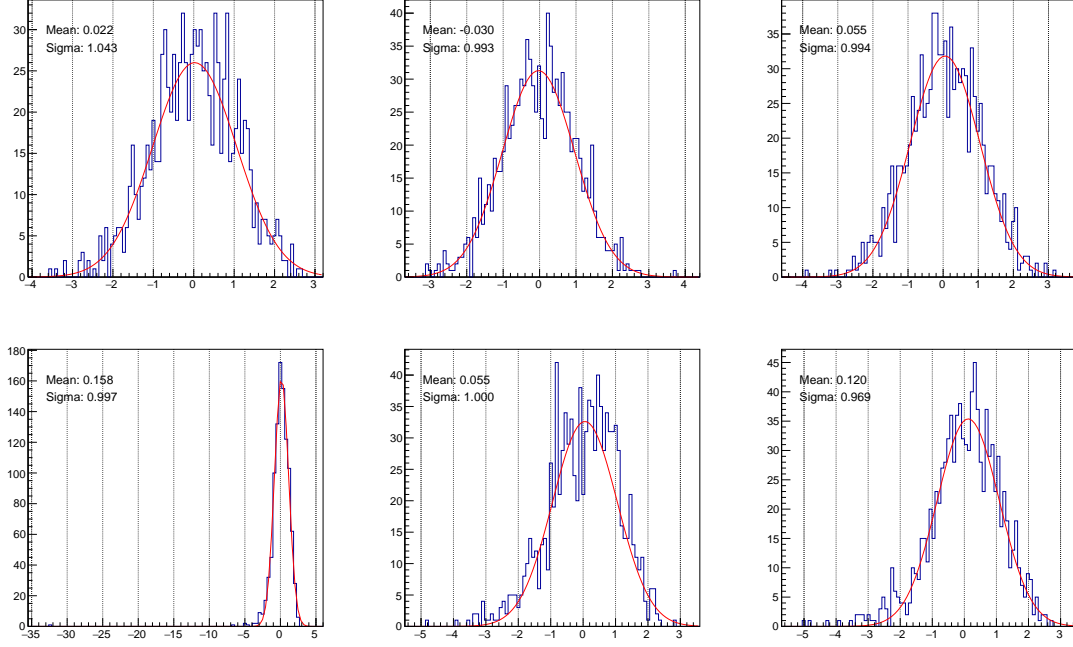
Figure 4.29: The pull distributions for different values of μ_{truth}. The left column corresponds to $\mu_{\text{truth}} = 1$, the middle column represents $\mu_{\text{truth}} = 5$, and the right column shows $\mu_{\text{truth}} = 10$. The upper row of plots displays the pulls for the $M_{T'} = 600$ GeV model, while the lower two plots exhibit the pulls for the $M_{T'} = 1200$ GeV model.
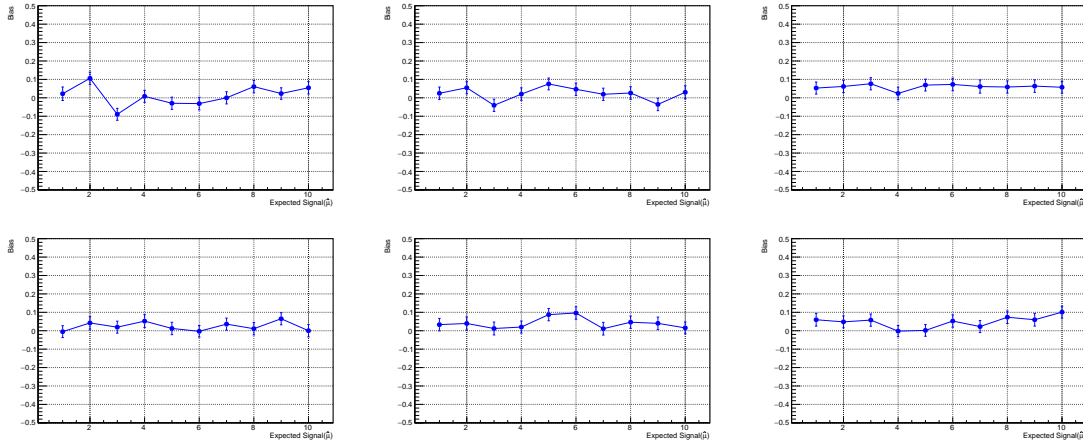
Figure 4.30: Bias test for $M_{T'} = 600, 625, 650$ GeV (top row) and $M_{T'} = 700, 800, 900$ GeV (bottom row) models with $\mu_{\text{truth}} \in [1, 10]$. No bias larger than 14% of σ_{fit} is observed.
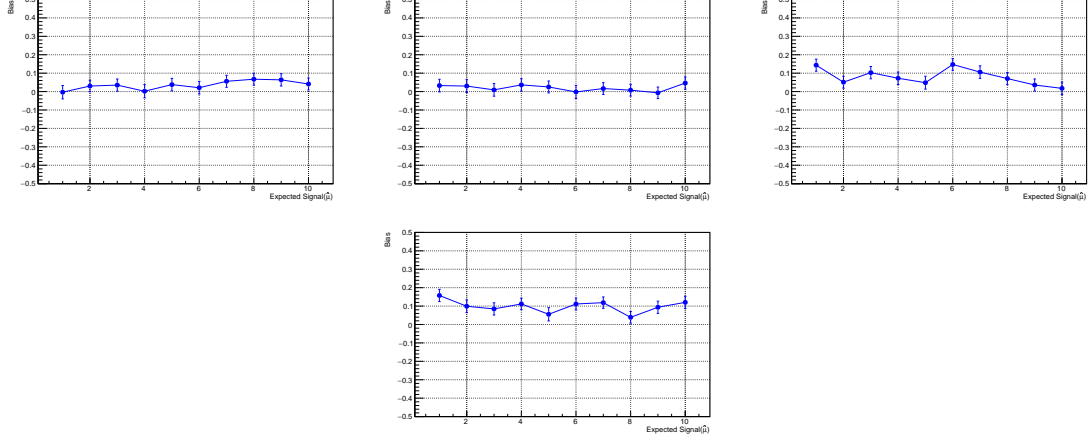
Figure 4.31: Bias test for $M_{T'} = 900, 1000$ GeV (top row) and $M_{T'} = 1100, 1200$ GeV (bottom row) models with $\mu_{\text{truth}} \in [1, 10]$. No bias larger than 20% of σ_{fit} is observed.

the $m_{\gamma\gamma}$ distribution are considered as nuisance parameters in the signal model and can affect the mean and width of the Gaussian distributions. The few uncertainties which won't affect the shape of $m_{\gamma\gamma}$ are treated as the log-normal variation in the event yield.

The details of the sources of the theoretical and experimental uncertainties are explained in Sec. 4.7. All uncertainties are correlated among the leptonic and hadronic categories.

4.7.1 Impacts

The effect of systematic uncertainties on the signal and background expectations is modeled through nuisance parameters (NPs), collectively denoted as θ. These parameters are restricted by Gaussian or log-normal probability density functions. Log-normal distributions are employed specifically for normalization uncertainties to avoid negative normalization factors during the fitting process. The expected numbers of signal and background events in each bin are functions of θ. The parameterization of each NP is carefully selected to ensure that the projected signal and background yields in each bin follow a log-normal distribution when the θ is normally distributed. The prior knowledge of the nuisance parameters is incorporated as a penalty term in the likelihood function (in Eqn. 3.10) which
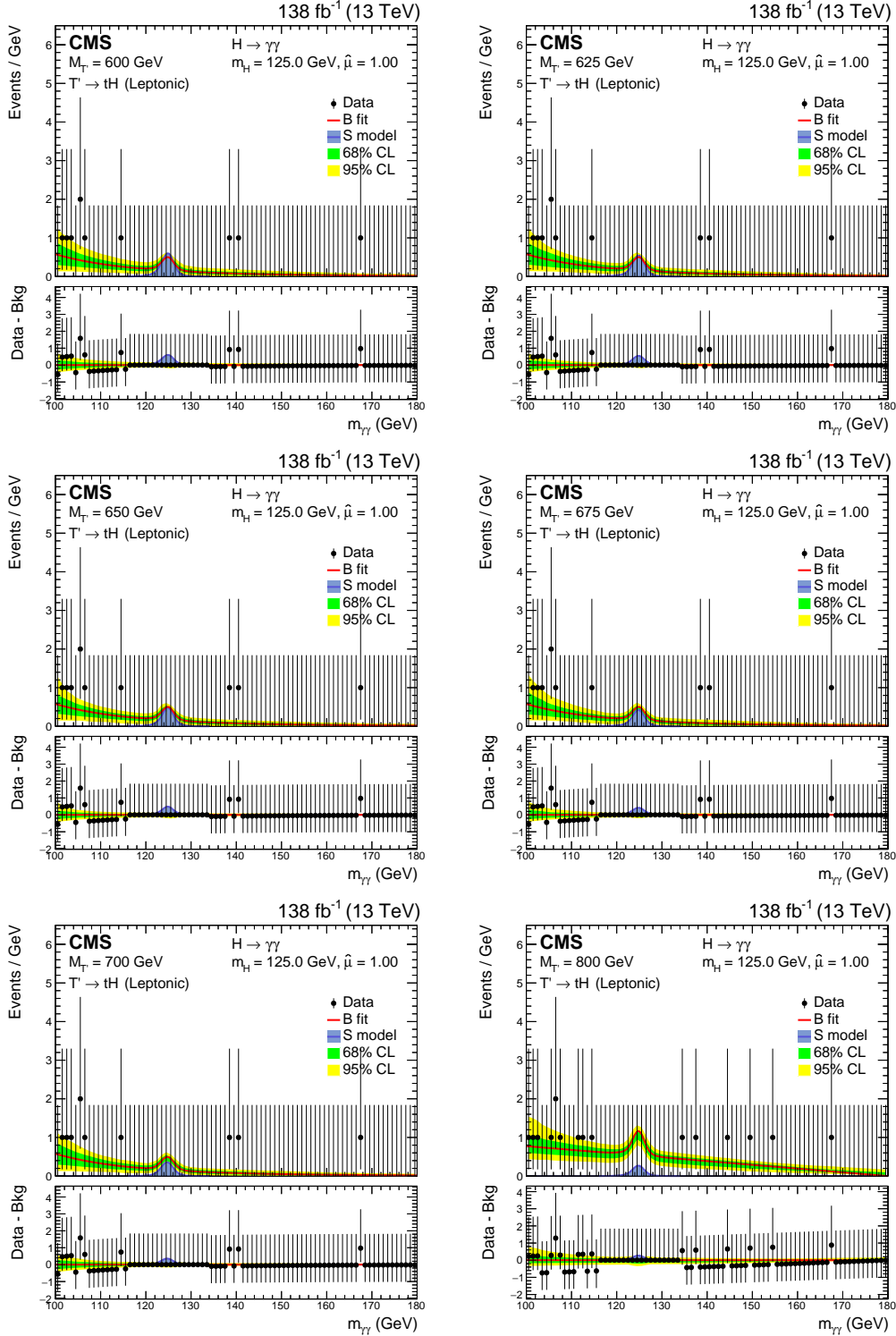
Figure 4.32: Overlaid signal and background model for the leptonic channel of mass points: 600 GeV (top left), 625 GeV (top right), 650 GeV (middle left), 675 GeV (middle right), 700 GeV (bottom left) and 800 GeV (bottom right). The red line in the plot is the best fitted background model and the blue line is the best fitted signal model.

184
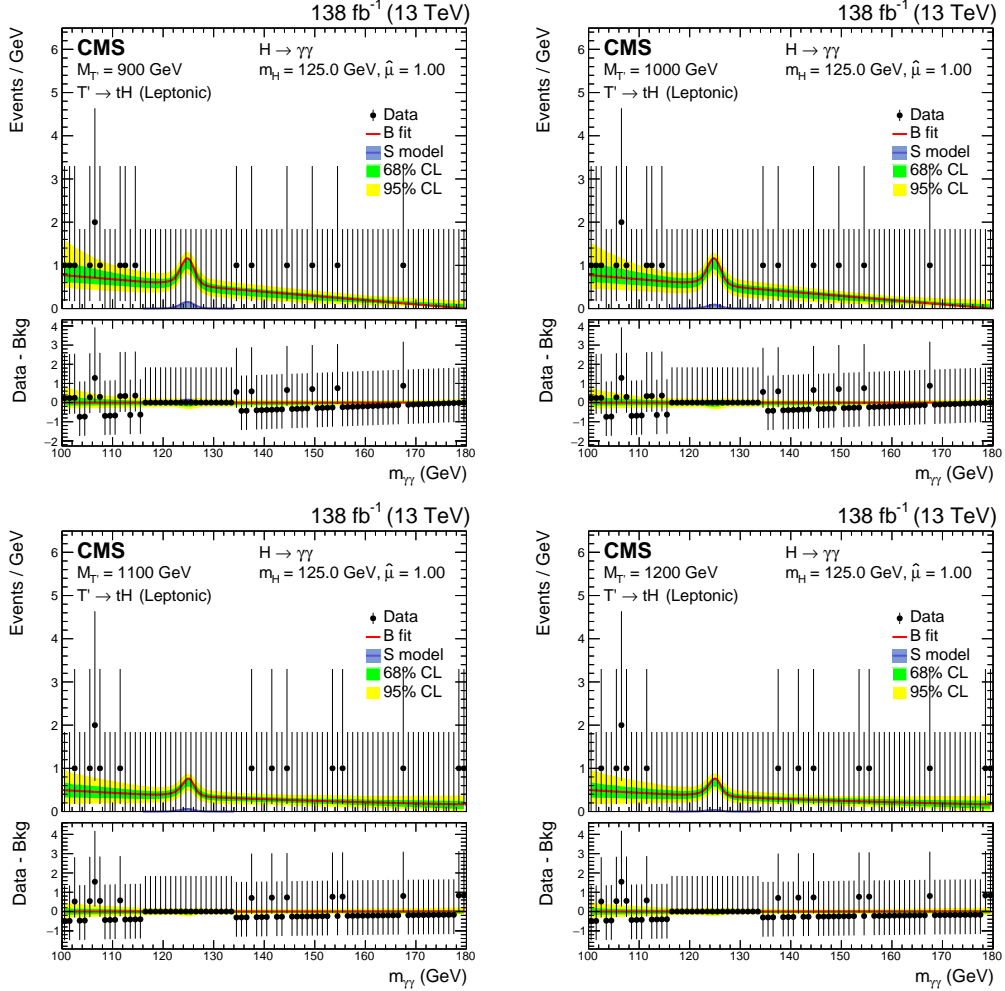
Figure 4.33: Overlaid signal and background model for the leptonic channel of mass points: 900 GeV (top left), 1000 GeV (top right), 1100 GeV (bottom left) and 1200 GeV (bottom right). The red line in the plot is the best fitted background model and the blue line is the best fitted signal model.
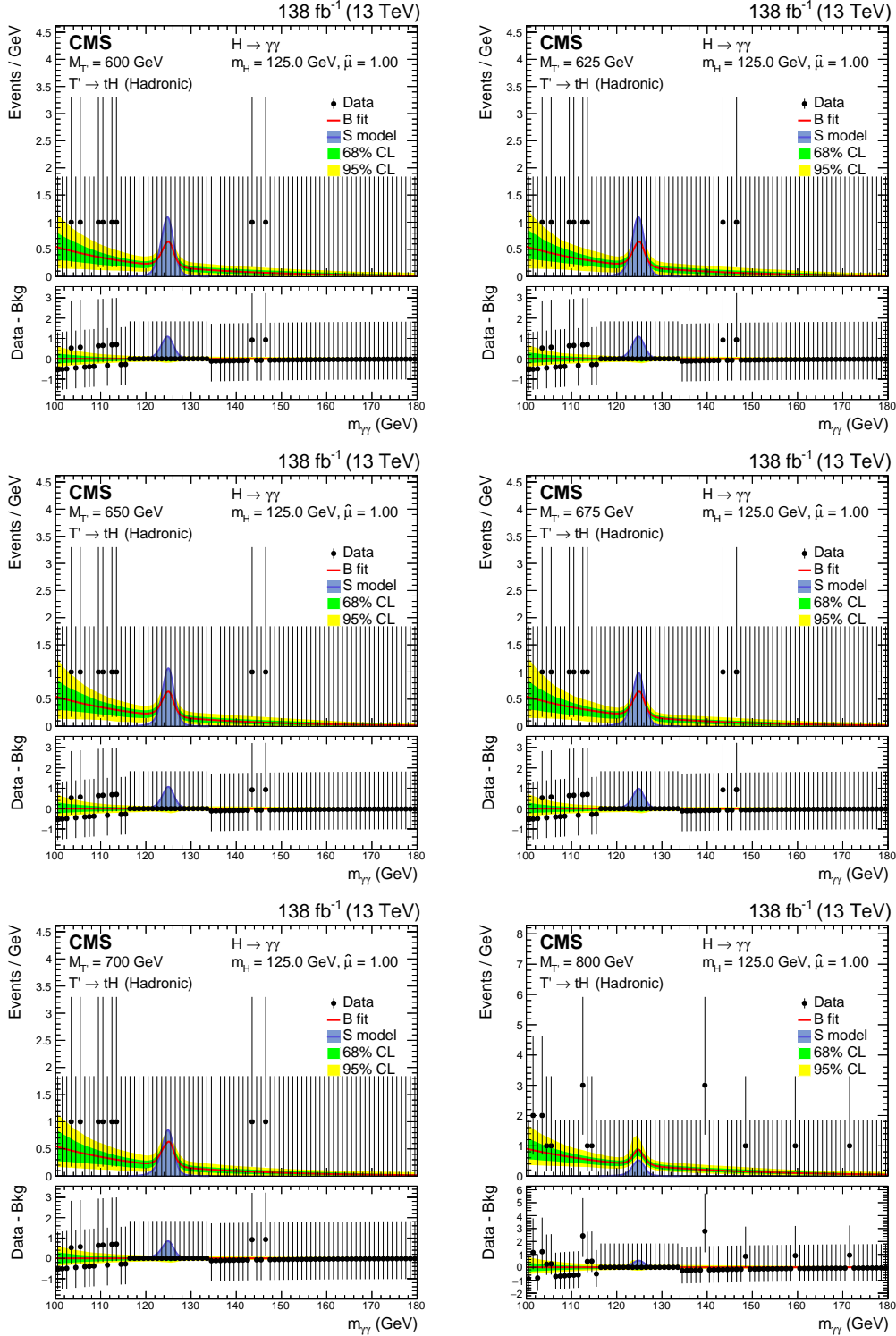
Figure 4.34: Overlaid signal and background model for the hadronic channel of mass points: 600 GeV (top left), 625 GeV (top right), 650 GeV (middle left), 675 GeV (middle right), 700 GeV (bottom left) and 800 GeV (bottom right). The red line in the plot is the best fitted background model and the blue line is the best fitted signal model.
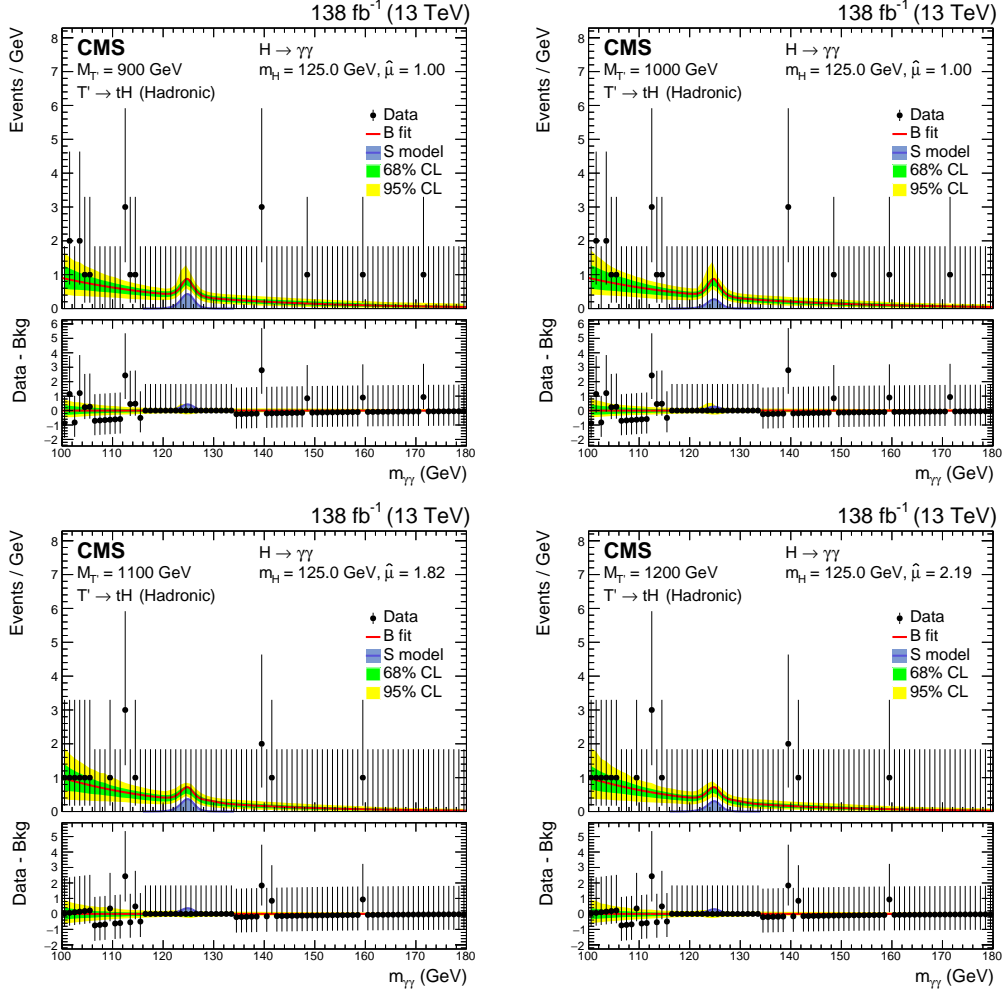
Figure 4.35: Overlaid signal and background model for the hadronic channel of mass points: 900 GeV (top left), 1000 GeV (top right), 1100 GeV (bottom left) and 1200 GeV (bottom right). The red line in the plot is the best fitted background model and the blue line is the best fitted signal model.

reduces the likelihood when θ deviates from its nominal value [111].

Gaining insight into the systematic uncertainties that exert the most significant impact on the parameter of interest (POI) warrants careful consideration. To achieve this, a ranking of the NPs is established. Each systematic uncertainty undergoes an iterative fit process, wherein the fit is performed while keeping the corresponding NP fixed at its fitted value, $\hat{\theta}$. The NP is then systematically shifted 1σ up and down by its fitted uncertainty, while allowing all other parameters to vary. This approach properly accounts for the correlations among different systematic uncertainties. By observing the induced shift ($\Delta\mu$) in the fit, the relative impact of each systematic uncertainty can be assessed and ranked accordingly. This metric serves as an indication of the correlation between the NP and the POI, and proves valuable in identifying NPs that contribute significantly to the uncertainty of the POI. To increase the computational efficiency and facilitate more comprehensive fit analyses, certain NPs that have a minimal impact on the expected uncertainty of the fitted parameter (μ) are disregarded or excluded.

Figure 4.36-4.37 show the impact of the systematic uncertainties on the observed measurement of the signal strength for the VLQ signal with $M_{T'}$ of 800 GeV and 1100 GeV, respectively. In the left panel of the figure, the quantity $(\theta-\theta_0)/\Delta\theta$ is presented, where θ represents the post-fit value of the nuisance parameter, θ_0 represents the pre-fit (initial estimation from global fit) value, and $\Delta\theta$ denotes the pre-fit uncertainty. The error bars displayed in an asymmetric manner illustrate the post-fit uncertainty divided by the pre-fit uncertainty. This indicates that parameters exhibiting error bars narrower than ± 1 are constrained by the fit.
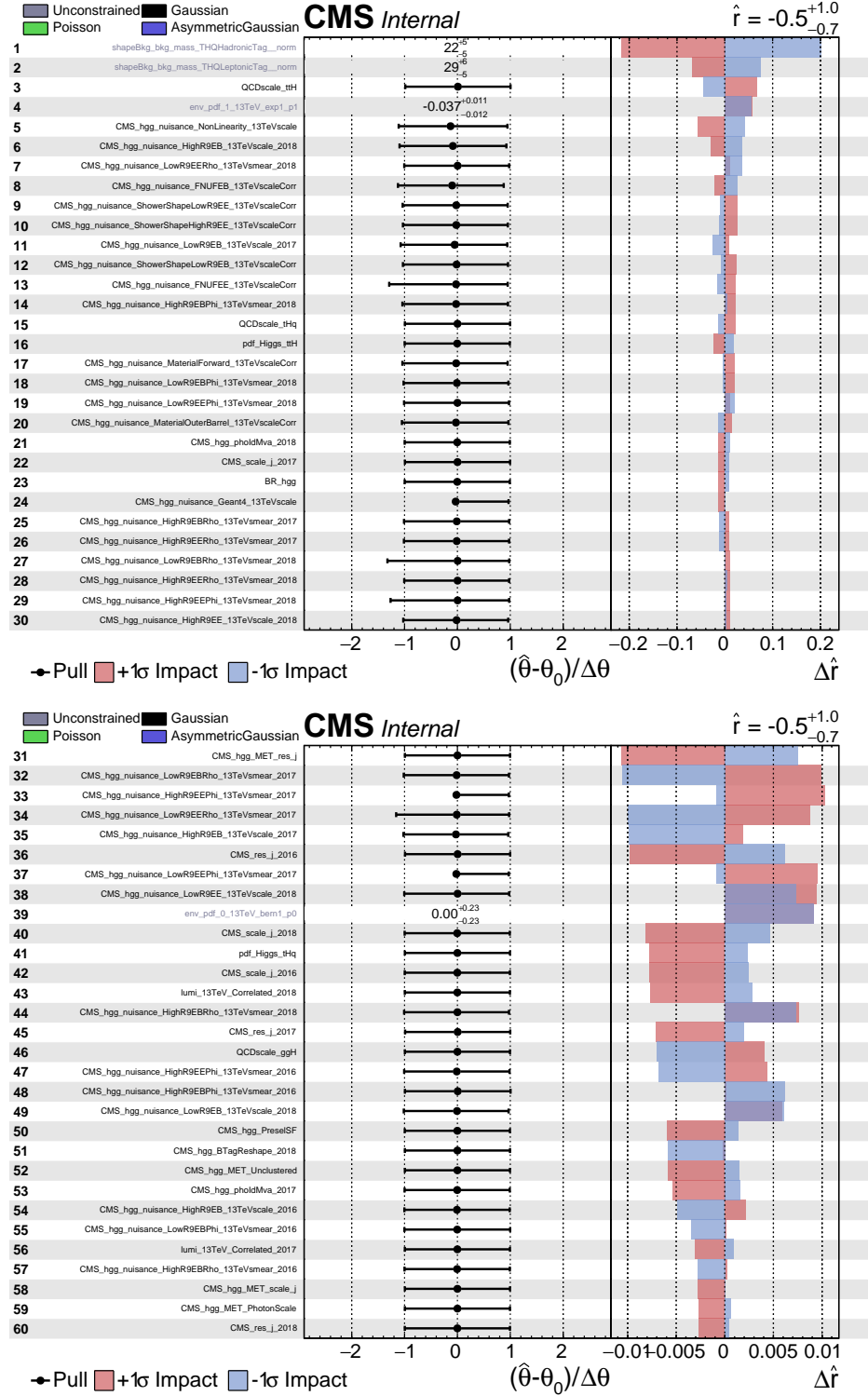
Figure 4.36: Impact of systematic uncertainties on the observed measurement of the signal strength for the VLQ with $M_{T'}$ = 800 GeV. Here, "r" in the figure corresponds to the signal strength, μ, mentioned in this thesis.
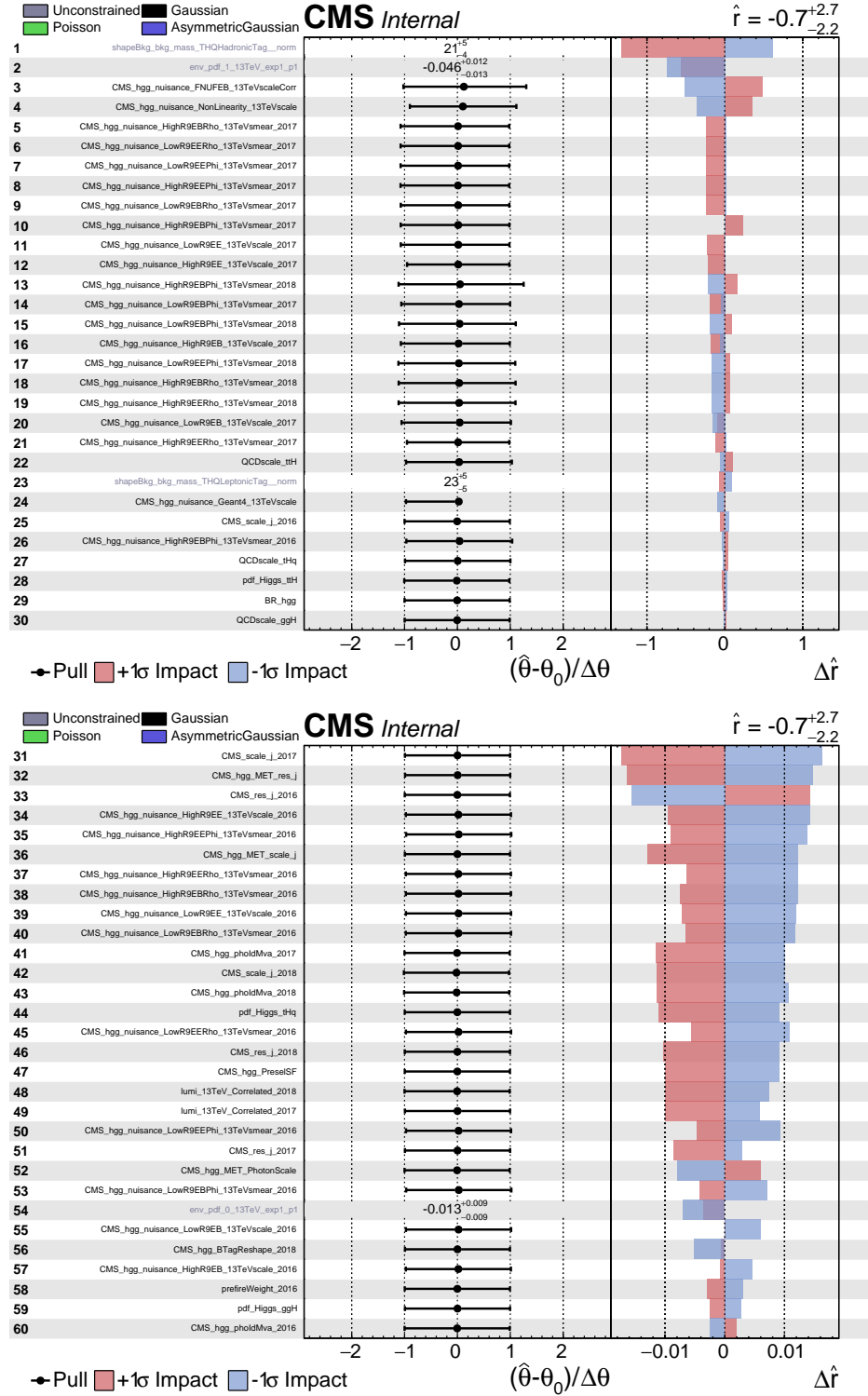
Figure 4.37: Impact of systematic uncertainties on the observed measurement of the signal strength for the VLQ with $M_{T'} = 1100$ GeV. Here, "r" in the figure corresponds to the signal strength, μ, mentioned in this thesis.

4.8   Goodness-of-fit test in control region

The goodness-of-fit (GoF) method provides a means to assess the degree of compatibility between the observed data and the probability density function (PDF) of the model. The purpose of this study is to validate the reliability of the fit in a control region prior to fitting the signal region. To conduct this test, control regions are defined, taking into account the following considerations:

- The signal, vector-like quark, is negligible in comparison to the background.

- A small number of SM Higgs events (approximately 10) are retained to ensure accurate modeling of the signal.

- A limited number of data events in the $m_{\gamma\gamma}$ sideband are preserved for effective modeling of the background.

To separate the control regions from the primary analysis region, a mass cut is applied: $M_{T'} < 480$ GeV. Next, we look for possible regions within the phase space defined by the BDT discriminants, which are trained for the mass category of $[600, 700]$ GeV. By adjusting the cuts on the BDT discriminants, two viable control regions are identified:

- Hadronic channel: $M_{T'} < 480\mathrm{GeV} \ \& \ \mathrm{BDT\text{-}SMH} < 0.80 \ \& \ \mathrm{BDT\text{-}NRB} \in [0.58, 0.78]$

- Leptonic channel: $M_{T'} < 480\mathrm{GeV} \ \& \ \mathrm{BDT\text{-}SMH} \in [0.20, 0.60]$

A schematic representation of the control regions is presented in Fig. 4.38, providing a visual overview. Furthermore, Tab. 4.12 provides a summary of the event yields for the different samples within the control regions. The table highlights the notable occurrence of a few number of SM Higgs events, predominantly contributed by $t\bar{t}H$ events. Within these control regions, the presence of the T' signal remains minimal.
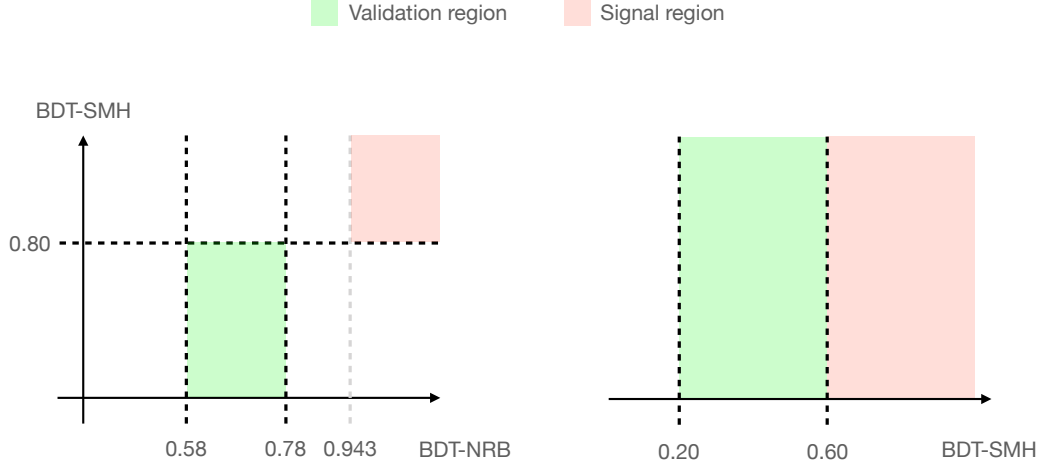
Figure 4.38: A schematic view of the BDT phase space for both the hadronic channel (left) and the leptonic channel (right) is shown. The schematic displays the 2-dimensional BDT phase space, with BDT-NRB and BDT-SMH axes in the hadronic channel, as well as the 1-dimensional BDT phase space with BDT-SMH axis in the leptonic channel. The control regions, marked in green, and the signal regions, marked in red, are clearly indicated within the respective BDT phase spaces.

Sample	Hadronic channel	Leptonic channel
Data in $m_{\gamma\gamma}$ Sideband	458	557
SMH	11.36	7.71
VLQ $M_{T'}$ = 600 GeV	0.35	0.22
VLQ $M_{T'}$ = 625 GeV	0.23	0.14
VLQ $M_{T'}$ = 650 GeV	0.15	0.10
VLQ $M_{T'}$ = 675 GeV	0.11	0.07
VLQ $M_{T'}$ = 700 GeV	0.09	0.05

Table 4.12: Table displaying the event yields within the control regions, emphasizing the presence of a higher number of SM Higgs events. In these control regions, the T′ signal remains minimal, while a small number of SM Higgs events are available for modeling the Higgs peak in the $m_{\gamma\gamma}$ spectrum. Furthermore, the data events within the $m_{\gamma\gamma}$ sideband are used for modeling the distributions of non-resonant background components.

In the defined control regions, the signal and background are modeled from the $m_{\gamma\gamma}$ spectrum. Models of signal are created by fitting the $m_{\gamma\gamma}$ distribution for ggH, VBF, VH, ttH and tHq MC samples with a sum of at most five Gaussian functions. Fig. 4.39 and Fig. 4.40 display the signal models for the hadronic channel and leptonic channel, respectively.

The non-resonant background is modeled using the data $m_{\gamma\gamma}$ spectrum, $m_{\gamma\gamma} \in [100, 180]$ GeV and used similar procedure mentioned in Sec. 3.9.2. Figure 4.41 shows the background models for the both channels while Fig. 4.42 presents the post-fit signal-plus-background model in the $m_{\gamma\gamma}$ region of [100, 180] GeV.
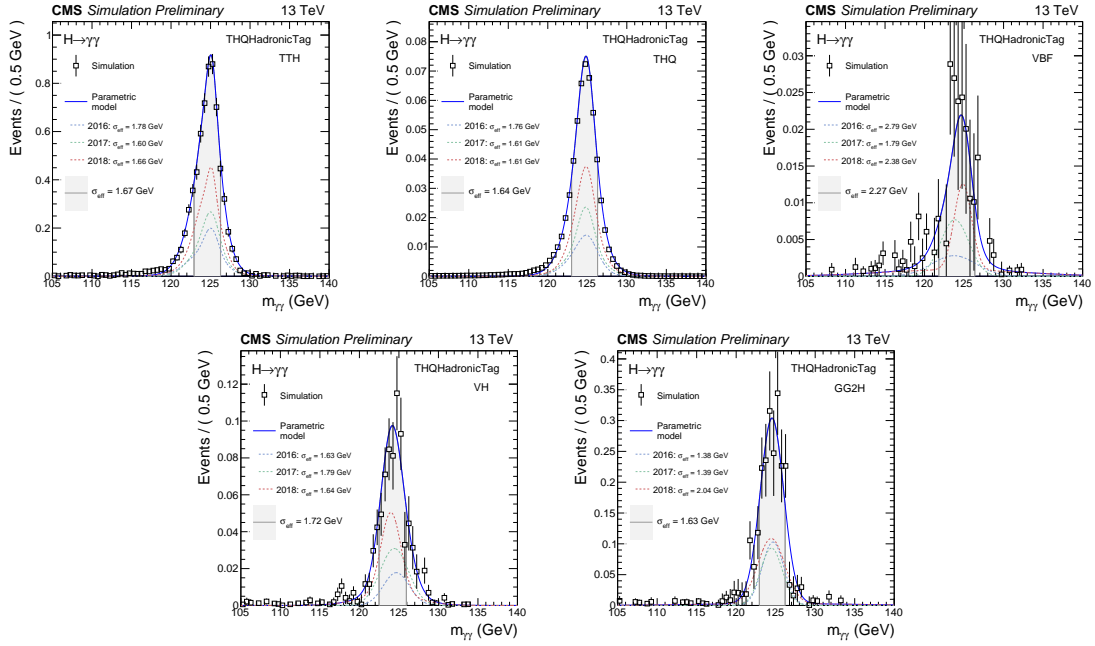
Figure 4.39: Model of SM Higgs boson production entering hadronic control region defined by the BDTs trained in [600-700] mass range.

A likelihood fit is performed for a shared signal strength parameter μ_{H}, which scales all Higgs production modes collectively. The expected and observed signal strengths obtained from the fit are $1.000^{+0.755}_{-0.687}$ and $1.848^{+0.817}_{-0.762}$, respectively, as shown in Fig. 4.43. The observed result is compatible with the latest ttH analysis [76], where the observed best fit signal strength for the ttH signal is $\mu_{\mathrm{ttH}} = 1.34^{+0.36}_{-0.29}$.
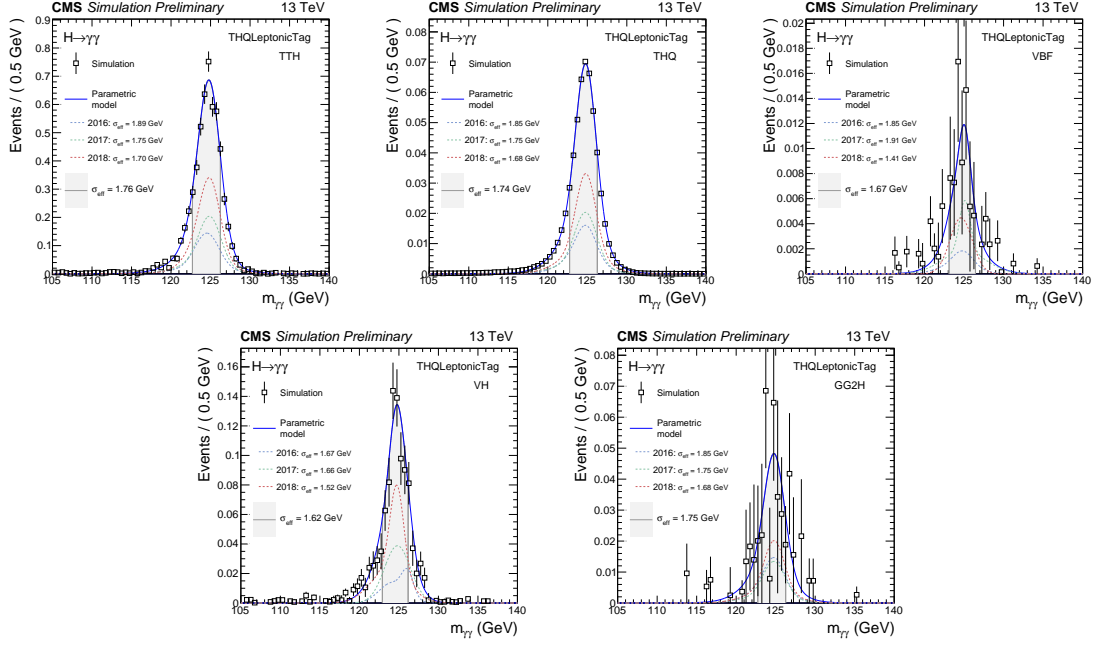
Figure 4.40: Model of SM Higgs boson production entering the leptonic control region defined by the BDT trained in [600-700] mass range.
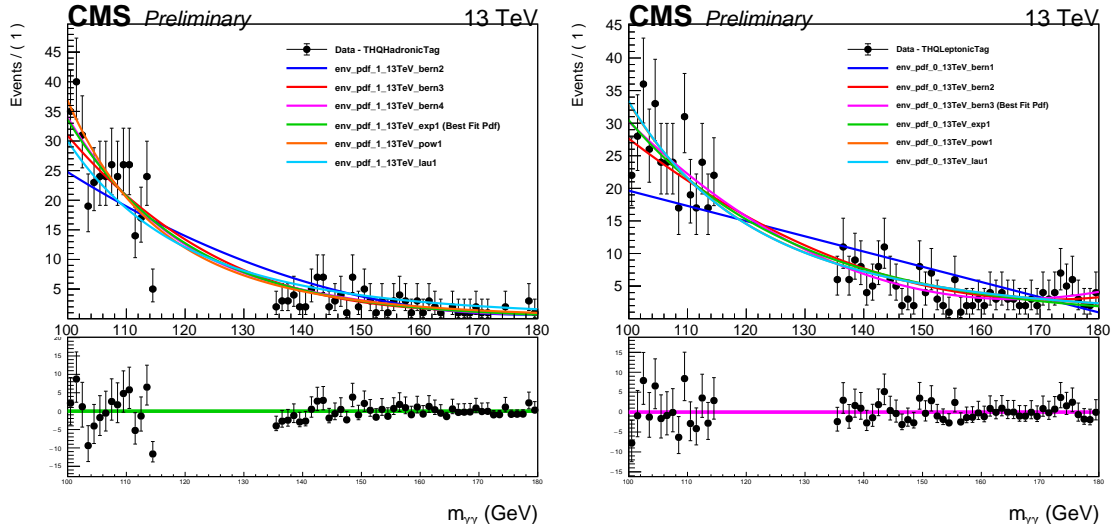
Figure 4.41: Model for non-resonant background which enter the hadronic (left) and leptonic (right) control regions with [600–700] mass range training.
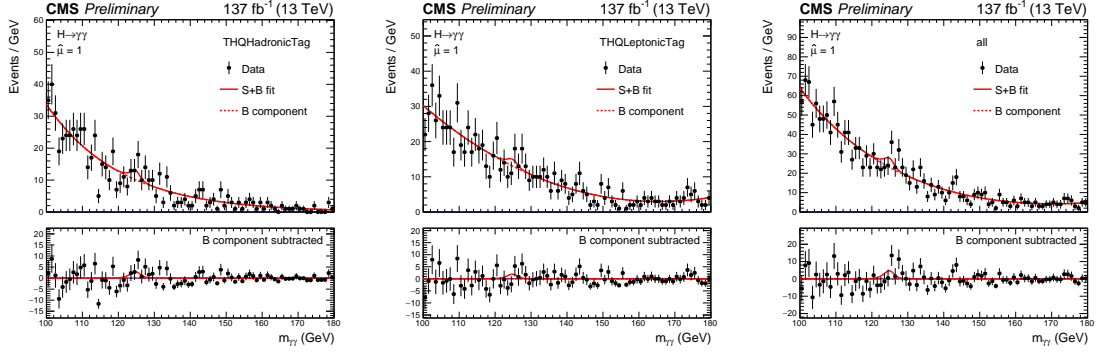
Figure 4.42: The probability density function for the signal plus background is superimposed on the distribution of the data $m_{\gamma\gamma}$ in the range of 100 to 180 GeV. The distributions are shown separately for the hadronic channel (left), leptonic channel (middle), and the combined channels (right).
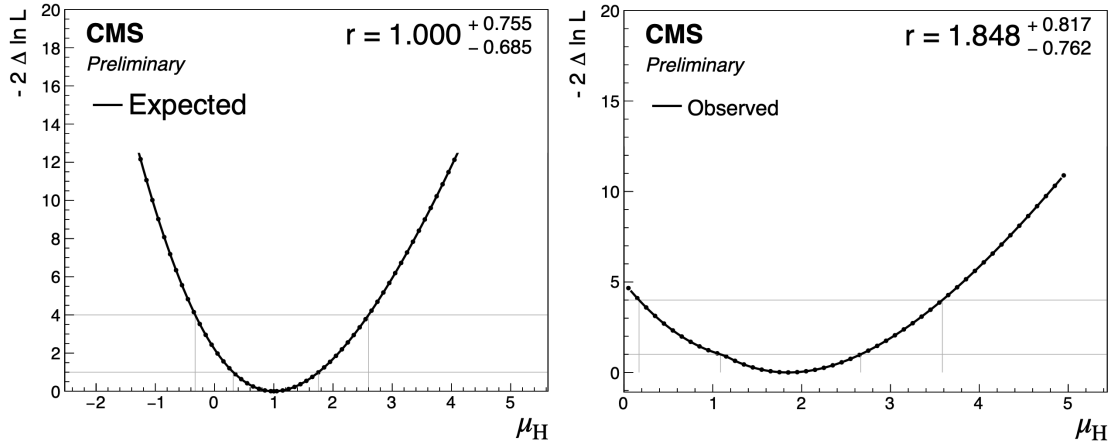
Figure 4.43: Expected signal strength (left) and the observed signal strength (right) for Higgs boson production in the control region. Here, "r" in the figure corresponds to the signal strength, μ, mentioned in this thesis.

The GoF is evaluated for the combined hadronic and leptonic channels. To calculate the GoF, 500 toy datasets are generated with signal-plus-background hypothesis. The test statistic, as defined in Eqn. 3.11, is computed by fitting the signal-plus-background hypothesis to each toy dataset and comparing the resulting test statistic distribution with the test statistic obtained from fitting the signal-plus-background to the observed data. Figure 4.44 shows the distribution of the test statistic, and the green vertical line shows the test statistic of the observed data fit, indicating a p-value of 82% for the goodness of fit. The results demonstrate that the observed best-fit test statistics align well with the expected values. Collectively, these findings offer validation for the soundness of the signal-plus-
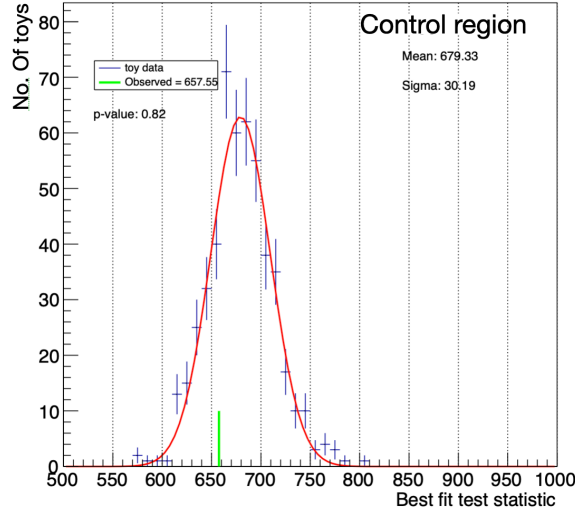
Figure 4.44: Goodness of fit of S+B model on 500 pseudo-experiments for the combination of the hadronic channel and leptonic channel in the control regions.

background hypothesis employed in the experiment, indicating that the observed data aligns consistently with this hypothesis within the control region.

4.9   Results

In our search for a signal process that lies beyond the boundaries of the standard model, it is customary to establish an upper limit on the cross section multiplied by the branch-

ing fraction of the process. This is conducted under the presumption that our dataset does not reveal a substantial breakthrough in new physics. The results of such an investigation involve quantifying the level of agreement between the observed data and a specific hypothesis, denoted as H, by calculating a p-value. This p-value represents the probability, assuming hypothesis H, of encountering data that exhibit equal or greater incompatibility with the predictions of H. The degree of incompatibility can be assessed based on various factors, such as the number of events detected within specific regions of certain distributions or through the likelihood ratio between the signal and background. If the p-value falls below a predetermined threshold, the hypothesis can be considered excluded.

In this search, the expected limits are determined through a maximum likelihood fit of the $m_{\gamma\gamma}$ distributions in the range of 100–180 GeV for both the leptonic and hadronic channels. This fitting process is conducted individually for each T′ mass point within the range of 600 to 1200 GeV. Figure 4.49, 4.50 and 4.51 display the data distributions alongside the corresponding best fitted signal-plus-background model to the $m_{\gamma\gamma}$ distribution for the leptonic, hadronic, and combined (leptonic plus hadronic) channels, respectively, focusing on $M_{T'}$ values of 600, 900, and 1200 GeV. Notably, no statistically significant deviations beyond the expected SM backgrounds are observed across any of the channels or mass ranges. For various $M_{T'}$ values, upper limits on the signal strength modifiers $\mu_{\text{obs}} = (\sigma)_{\text{obs}}/(\sigma)_{\text{th}}$ and $\mu_{\text{exp}} = (\sigma)_{\text{exp}}/(\sigma)_{\text{th}}$, are obtained by performing a maximum likelihood fit of the $m_{\gamma\gamma}$ distributions. Throughout the fitting process, the M_{H} parameter of the model is held constant at 125 GeV. The estimation of expected and observed upper limits is carried out at the 95% confidence level (CL) based on the CLs criterion [112, 113] using the asymptotic approximation [114, 115] for the test statistic. Finally, the upper limits on μ_{obs} and μ_{exp}, as shown in Fig. 4.45, are converted into upper limits on $\sigma_{T'\text{bq}} \times \mathcal{B}_{T'\rightarrow\text{tH}}$. These limits are presented in Fig. 4.46 for the individual leptonic and hadronic channels and in Fig. 4.47 for both the channels combined, along with the theoretical cross sections for singlet T′ pro-

duction. The theoretical cross sections are shown for representative κ_T values of 0.1, 0.15, 0.2, and 0.25, under the condition $\Gamma/M_{T'} < 5\%$. To ensure the validity of the results, they are validated through pseudoexperiments and further scrutinized with the goodness of fit test. The test statistic of the data fit aligns with the median of the test statistic distribution obtained from toy fits as shown in Fig. 4.52. These limits are presented in Fig. 4.46 for the individual leptonic and hadronic channels and in Fig. 4.47 for combining both channels, along with the theoretical cross sections for singlet T' production.
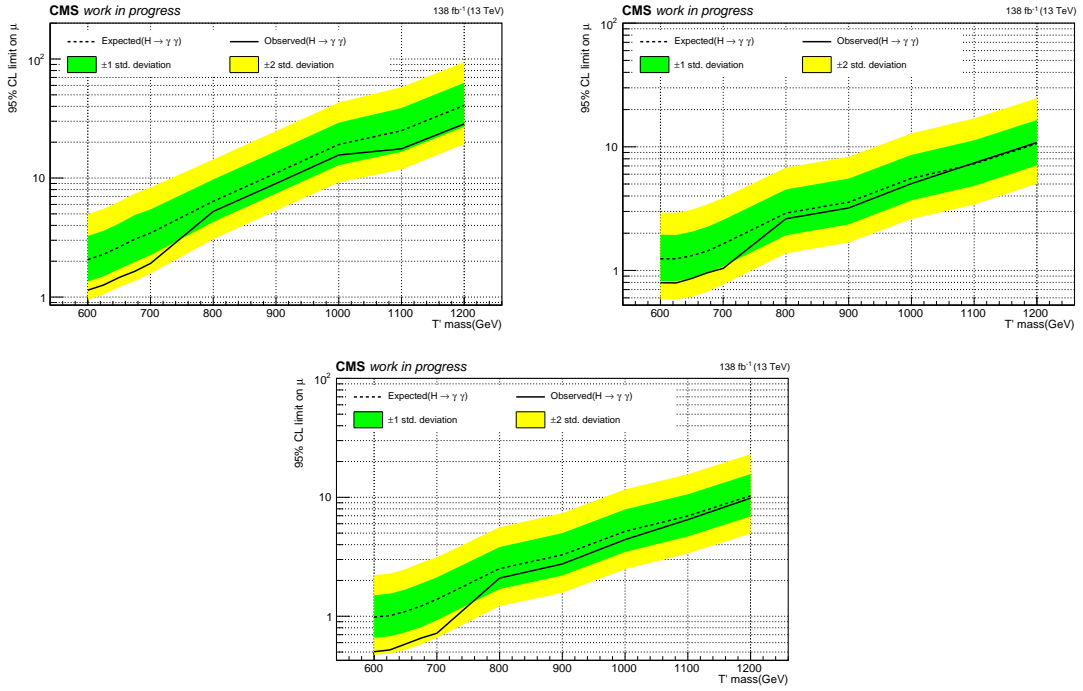
Figure 4.45: The expected (dotted black) and observed (solid black) upper limits at 95% CL on signal strength in the leptonic channel (top left), hadronic channel (top right) and combined (below) with $M_{T'} \in [600, 1200]$ GeV and m_H fixed at 125 GeV. The green (yellow) band represents the 68% (95%) of the limit values expected under the background-only hypothesis.

The limits and yields for individual channels and each T' masses with $m_{\gamma\gamma} \in [115, 135]$ are presented in Tab. 4.14 and Tab. 4.15. Additionally, Tab. 4.13 displays the observed and expected limits on signal strength after combining the leptonic and hadronic channels, with
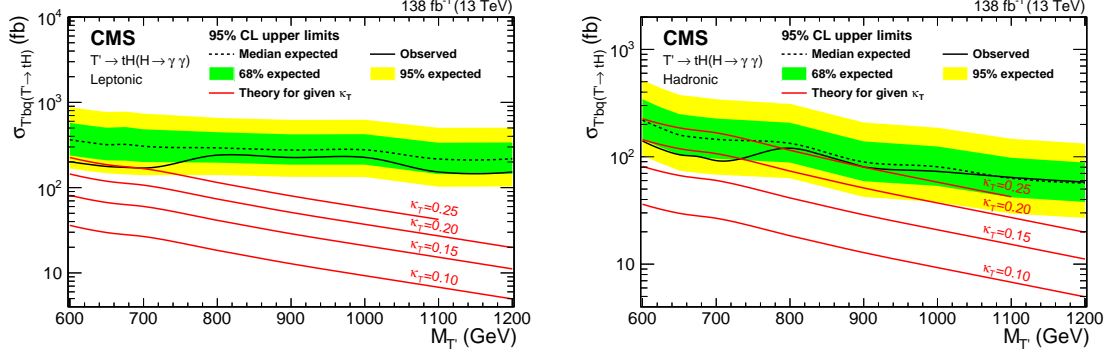
Figure 4.46: The expected (dotted black) and observed (solid black) upper limits at 95% CL on $\sigma_{\text{T}'\text{bq}} \times \mathcal{B}_{\text{T}' \rightarrow \text{tH}}$ are displayed as a function of $\text{M}_{\text{T}'}$ in the leptonic channel (left) and the hadronic channel (right). The green (yellow) band represents the 68% (95%) of the limit values expected under the background-only hypothesis. The theoretical cross sections for the singlet T' production with representative κ_{T}-values fixed at 0.1, 0.15, 0.2 and 0.25 (for $\Gamma/\text{M}_{\text{T}'} < 5\%$) are shown as red lines.
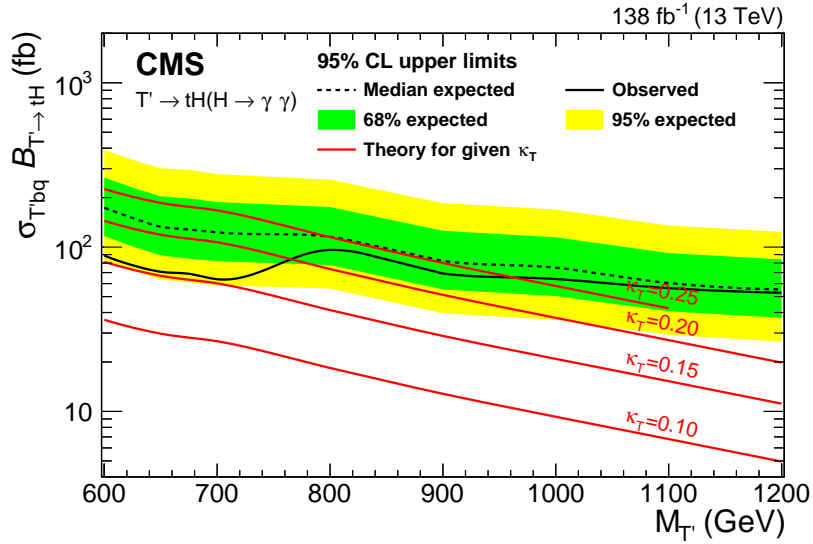
Figure 4.47: The combined expected (dotted black) and observed (solid black) upper limits at 95% CL on $\sigma_{\text{T}'\text{bq}} \times \mathcal{B}_{\text{T}' \rightarrow \text{tH}}$ are displayed as a function of $\text{M}_{\text{T}'}$. The green (yellow) band represents the 68% (95%) of the limit values expected under the background-only hypothesis. The theoretical cross sections for the singlet T' production with representative κ_{T}-values fixed at 0.1, 0.15, 0.2 and 0.25 (for $\Gamma/\text{M}_{\text{T}'} < 5\%$) are shown as red lines.

the Higgs mass fixed at 125 GeV. Numbers of data events in $m_{\gamma\gamma}$ window and sideband are summarized in Tab. 4.16. Furthermore, Fig. 4.48 presents the distribution of data events in the 2D phase space of $m_{\gamma\gamma}$ and $m_{\text{T}'}$.

Figure 4.49 and Fig. 4.50 display the S+B models for each T′ mass in each category while Fig. 4.52 shows the goodness of fit of S+B models in combined channels with representative T′ masses of 600, 900 and 1200 GeV.

The coupling strength can be translated from cross section through Eq. 4.6. Accordingly, we set limits on $\kappa_{\text{T}'}$ as shown in Fig. 4.53.

4.10   Summary

In summary, a search for a vector-like quark decaying to a top quark and a Higgs boson which in turn decays into two photons, T′ → tH(H → γγ), has been performed using proton-proton collision data at $\sqrt{s} = 13\text{TeV}$ recorded with the CMS detector in 2016–2018, and corresponding to an integrated luminosity of 138 fb^{-1}. The search has been carried out based on the model of T′ electroweak production in the narrow width approximation with the ratio of T′ width relative to its mass $(\Gamma/\text{M}_{\text{T}'}) \approx 1\%$. The sensitivity of this analysis extends up to $\Gamma/\text{M}_{\text{T}'} \approx 5\%$, which roughly corresponds to the experimental resolution of $\text{M}_{\text{T}'}$. Both the hadronic and leptonic decay modes of the top quark are considered in the search. A multivariate analysis incorporating three separately optimized boosted decision trees is exploited to separate likely signal events from background processes, including the standard model production of Higgs bosons. No statistically significant excess over the expected background prediction is observed. Assuming a coupling to third generation quarks of $\kappa_{\text{T}} = 0.25$ and a relative decay width of $\Gamma/\text{M}_{\text{T}'} < 5\%$, the singlet T′ masses are excluded up to 960 GeV at 95% confidence level. This electroweak search for the vector-like quark T′ is the most sensitive to date for $\text{M}_{\text{T}'}$ up to 1.1TeV.
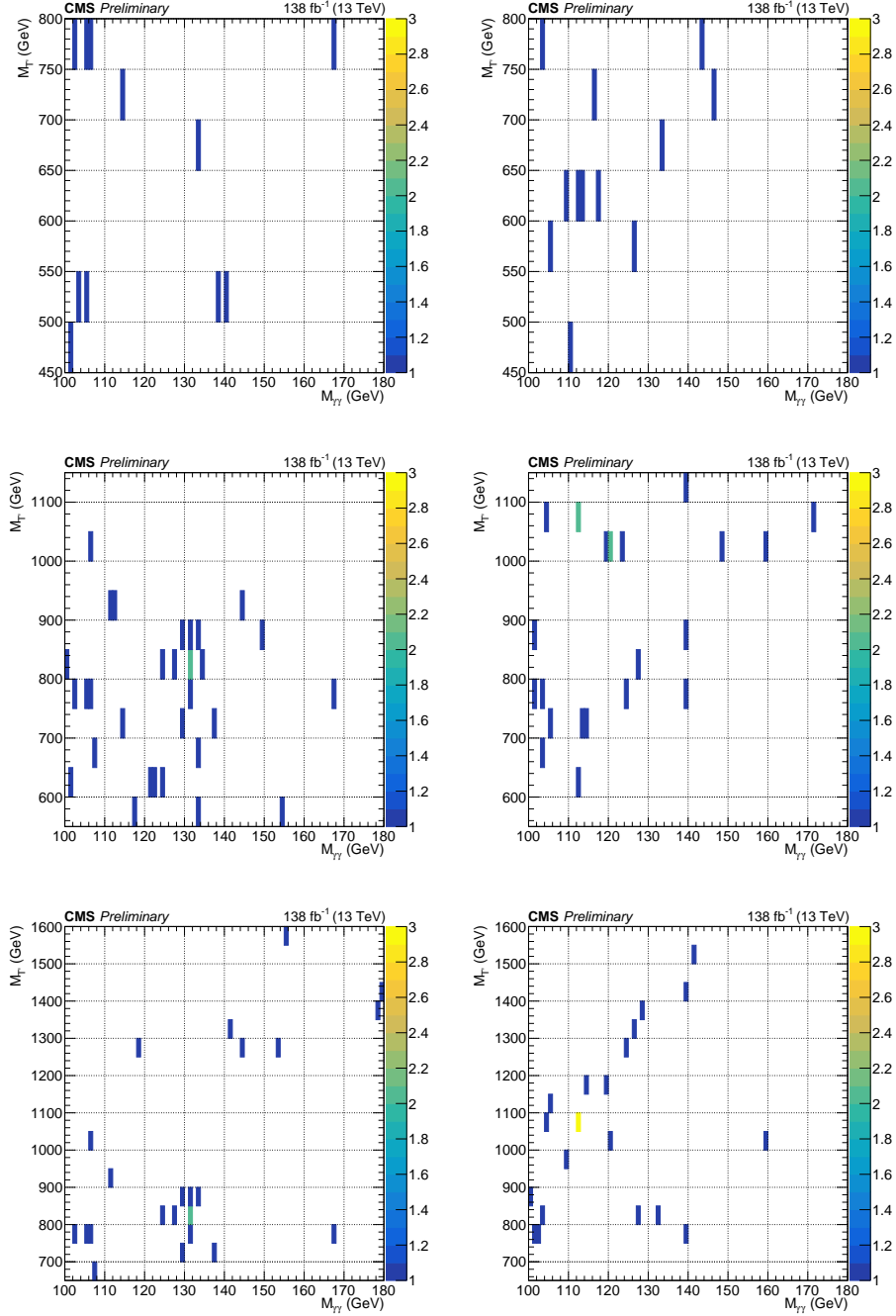
Figure 4.48: The distributions of data events in the 2D plane defined by the diphoton invariant mass and reconstructed T′ mass. The left column corresponds to the leptonic channel, while the right column represents the hadronic channel. Each row corresponds to a specific regions of BDT I, II and III, arranged from top to bottom.
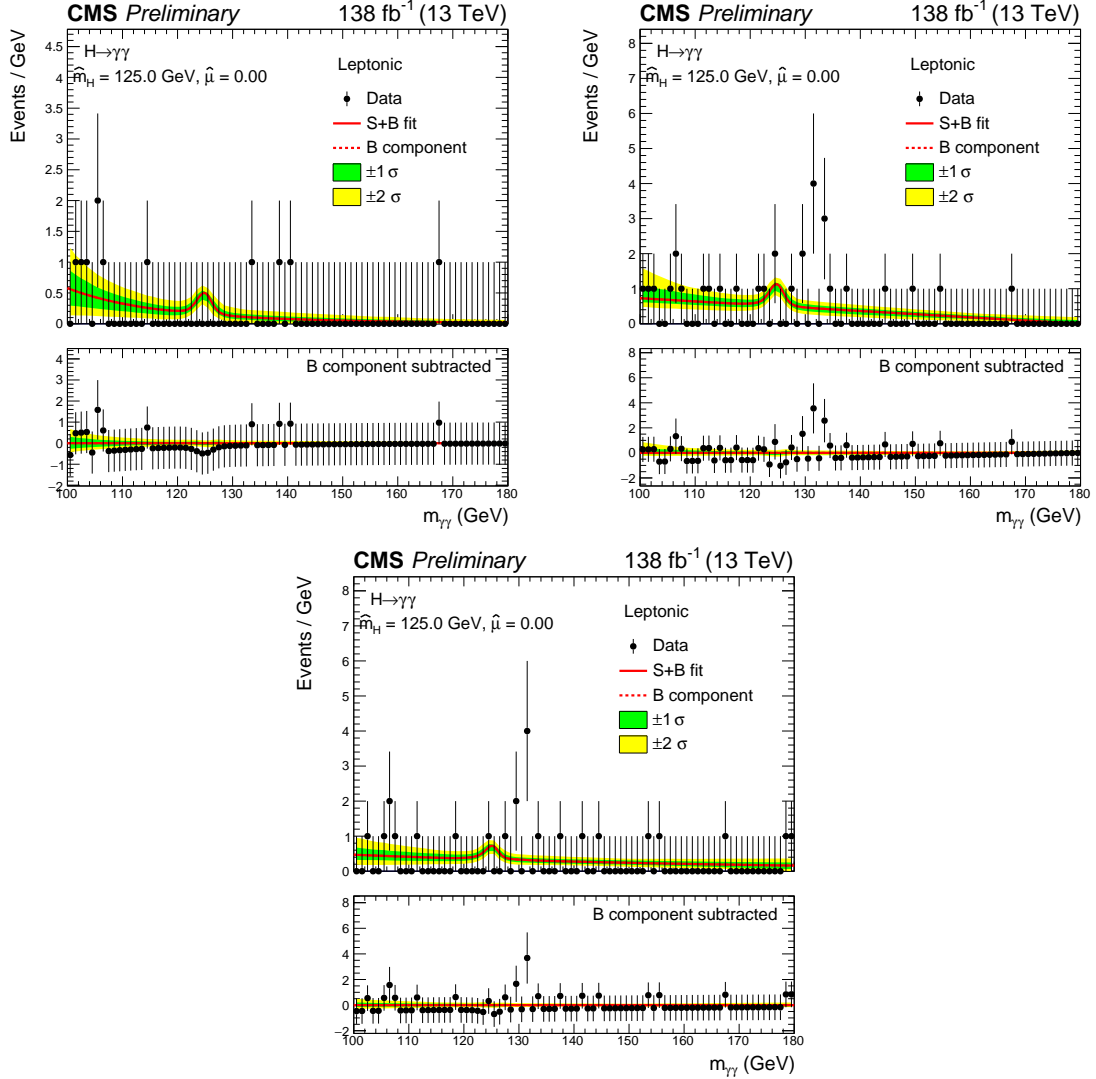
Figure 4.49: Leptonic channel: The distributions for data (black dots) and $m_{\gamma\gamma}$ signal-plus-background model fits (red line) for a VLQ signal with $\mathrm{M_{T'}}$ of 600 (upper left), 900 (upper right), and 1200 GeV (lower). The green (yellow) band represents the 68% (95%) CL in the background component of the fit. The peak in the background component shows the considered irreducible SM Higgs boson (ggH, VBF, VH, $\mathrm{t\bar{t}H}$ and tH) contribution. Here, $\hat{\mu}$ is the best fit value of the signal strength parameter μ, which is zero for the three $\mathrm{M_{T'}}$ values considered. The lower panel shows the residuals after the subtraction of the background component.
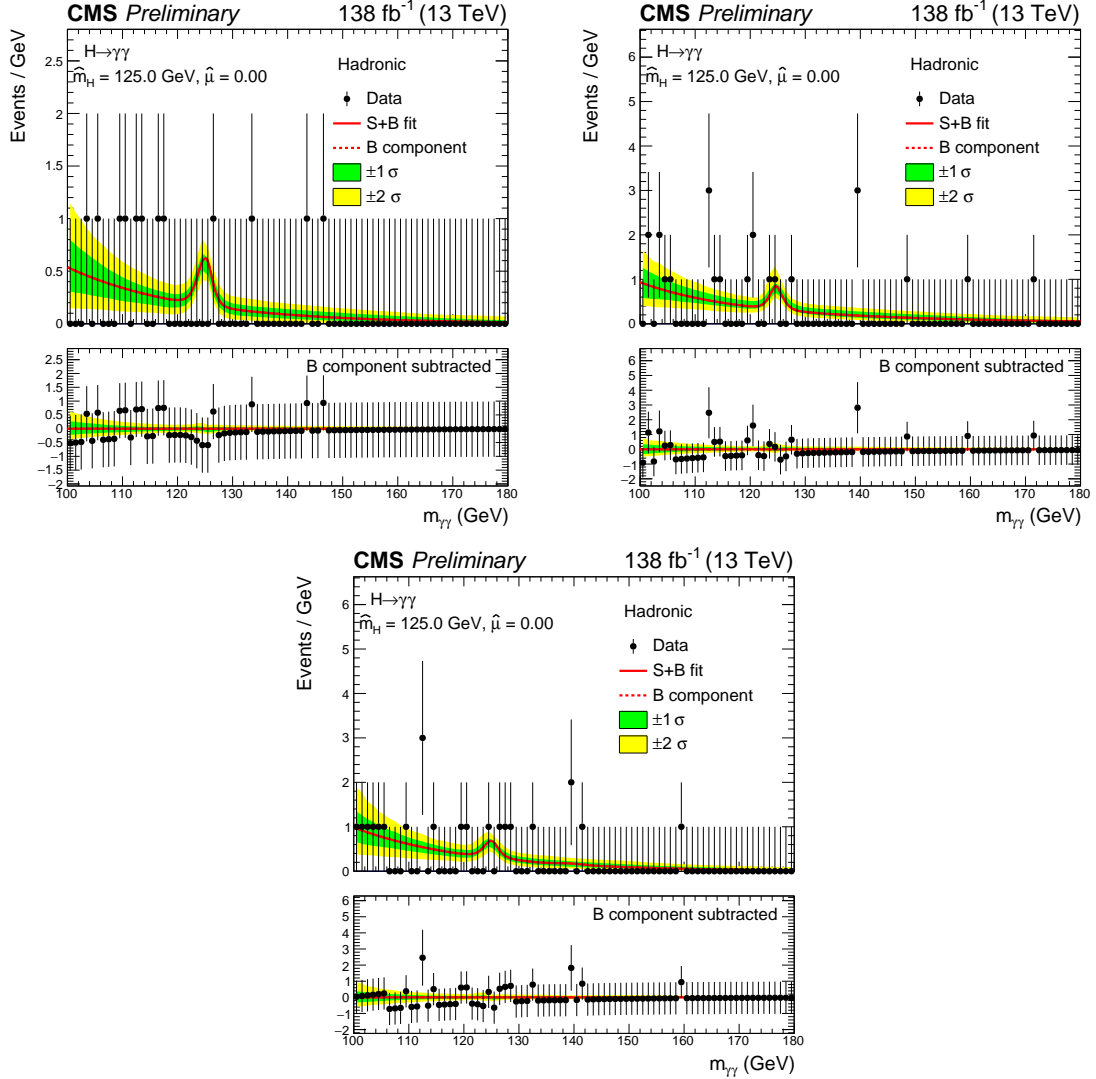
Figure 4.50: Hadronic channel: The distributions for data (black dots) and $m_{\gamma\gamma}$ signal-plus-background model fits (red line) for a VLQ signal with $M_{T'}$ of 600 (upper left), 900 (upper right), and 1200 GeV (lower). The green (yellow) band represents the 68% (95%) CL in the background component of the fit. The peak in the background component shows the considered irreducible SM Higgs boson (ggH, VBF, VH, tt̄H and tH) contribution. Here, $\hat{\mu}$ is the best fit value of the signal strength parameter μ, which is zero for the three $M_{T'}$ values considered. The lower panel shows the residuals after the subtraction of the background component.
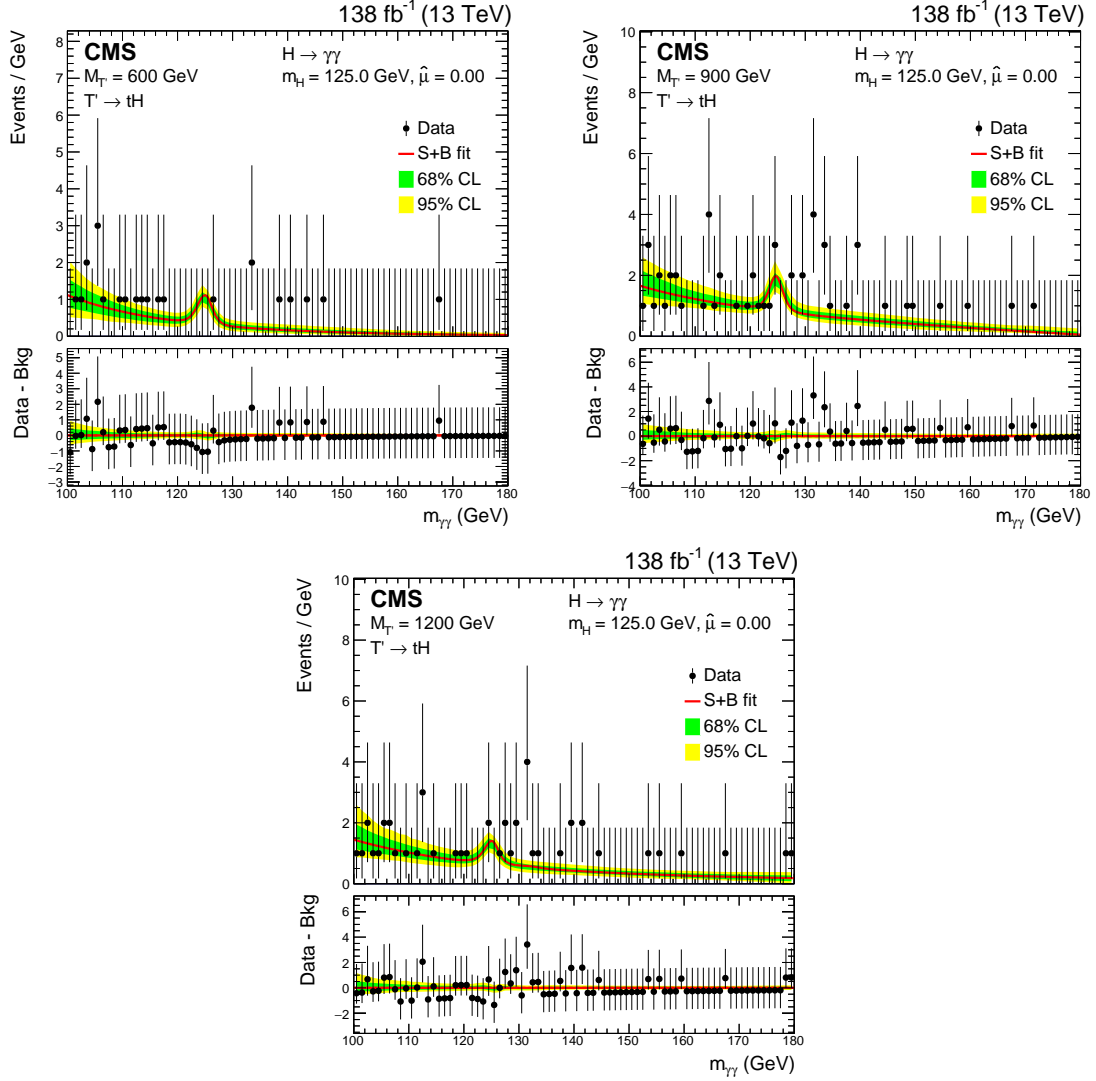
Figure 4.51: The combined, leptonic plus hadronic, distributions for data (black dots) and $m_{\gamma\gamma}$ signal-plus-background model fits (red line) for a VLQ signal with $\text{M}_{\text{T}'}$ of 600 (upper left), 900 (upper right), and 1200 GeV (lower). The green (yellow) band represents the 68% (95%) CL in the background component of the fit. The peak in the background component shows the considered irreducible SM Higgs boson (ggH, VBF, VH, $\text{t}\bar{\text{t}}\text{H}$ and tH) contribution. Here, $\hat{\mu}$ is the best fit value of the signal strength parameter μ, which is zero for the two $\text{M}_{\text{T}'}$ values considered. The lower panel shows the residuals after the subtraction of the background component.
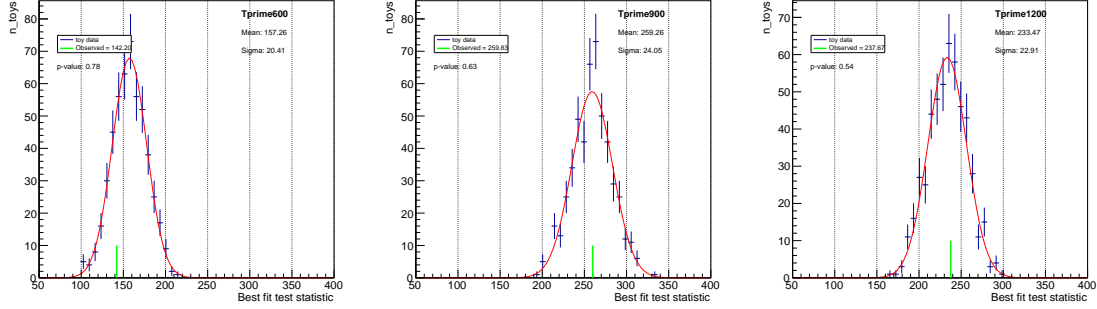
Figure 4.52: The goodness of fit for the signal-plus-background (S+B) model in the combined (leptonic plus hadronic) channel for VLQ signal masses of 600 GeV (left), 900 GeV (middle), and 1200 GeV (right). The test statistic of the data fit aligns with the median of the test statistic distribution obtained from toy fits. The green vertical line represents the test statistic of the observed data fit, and the corresponding p-values are indicated in the figure.
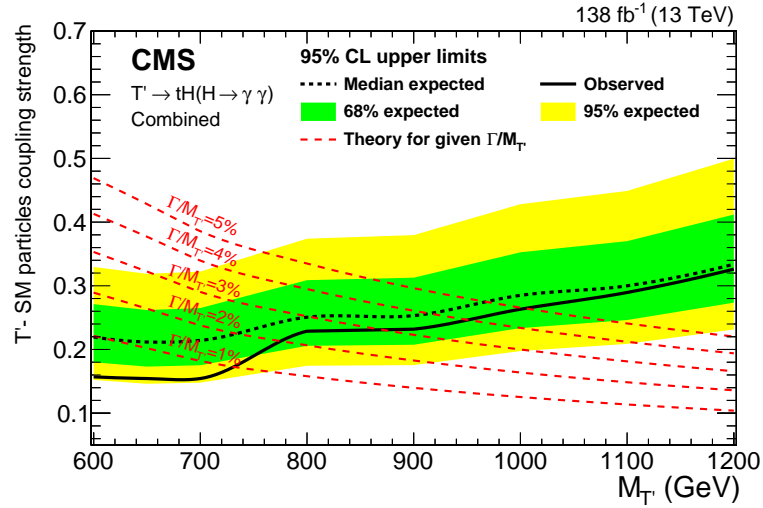
Figure 4.53: The combined expected (dotted black) and observed (solid black) upper limits at 95% CL on the T′ coupling with the SM particles, κ_{T}, under the narrow width approximation displayed as a function of $M_{\mathrm{T'}}$. The green (yellow) band represents the 68% (95%) of the limit values expected under the background-only hypothesis. The theoretical κ_{T} values corresponding to the $\Gamma/M_{\mathrm{T'}}$-values fixed at 1, 2, 3, 4, and 5% are shown as red dashed lines.

$M_{T'}$ (GeV)	Observed limit	Expected limit
600	0.50	0.98
625	0.52	1.01
650	0.58	1.10
675	0.66	1.22
700	0.72	1.39
800	2.09	2.52
900	2.75	3.28
1000	4.41	5.17
1100	6.49	6.97
1200	9.84	10.31

Table 4.13: Result of observed and expected limits on signal strength after combining the leptonic channel and the hadronic channel with the Higgs boson mass fixed at 125 GeV.

				Leptonic							
$M_{T'}$	Limits			Yield	Bkg Composition		SM Higgs Composition (%)				
(GeV)	observed	expected	VLQ	Tot. Bkg	Non-Res. Bkg	SM Higgs	ttH	ggH	tHq	VH	VBF
600	1.14	2.06	2.21								
625	1.26	2.28	2.00								
650	1.46	2.63	1.72	12.23 ± 9.08	10.94 ± 8.99	1.29 ± 0.09	61.2	3.8	18.6	14.7	1.6
675	1.65	3.06	1.53								
700	1.92	3.45	1.31								
800	5.24	6.38	0.97								
900	8.99	11.00	0.56	21.34 ± 14.51	19.04 ± 14.39	2.30 ± 0.12	76.0	0.4	12.17	10.0	1.3
1000	15.61	19.12	0.33								
1100	17.61	25.00	0.19	15.83 ± 13.78	14.40 ± 13.68	1.43 ± 0.10	74.8	0.69	12.58	11.8	0.69
1200	28.48	40.75	0.11								

Table 4.14: Leptonic channel: Expected limits and yields by process for each signal region targeting VLQ events with $M_{T'} \in [600, 1200]$ GeV. Yields are shown for events within the signal window, $115 < m_{\gamma\gamma} < 135$ GeV. The yields coming from MC samples are only used as a cross check; their uncertainties are not involved in limit calculation.

	Limits		Yield		Bkg Composition		SM Higgs Composition (%)				
$M_{T'}$ (GeV)	observed	expected	VLQ	Tot. Bkg	Non-Res. Bkg	SM Higgs	ttH	ggH	tHq	VH	VBF
600	0.80	1.24	3.88								
625	0.79	1.24	3.92								
650	0.86	1.32	3.64	3.35 ± 0.95	1.60 ± 0.94*	1.75 ± 0.10	50.3	24.9	19.1	5.2	0.6
675	0.96	1.44	3.40								
700	1.04	1.64	2.96								
800	2.62	2.91	1.81								
900	3.20	3.56	1.47	9.29 ± 4.00	7.30 ± 4.00	1.99 ± 0.12	50.8	29.7	14.9	3.6	1.0
1000	5.04	5.56	0.95								
1100	7.40	7.28	0.71	11.34 ± 5.31	8.98 ± 5.31	2.36 ± 0.15	31.6	50.6	9.5	4.8	3.5
1200	10.84	10.62	0.49								

Table 4.15: Hadronic channel: Expected limits and yields by process for each signal region targeting VLQ events with $M_{T'} \in [600, 1200]$ GeV. Yields are shown for events within the signal window, $115 < m_{\gamma\gamma} < 135$ GeV. The yields coming from MC samples are only used as a cross check; their uncertainties are not involved in limit calculation.

Signal region	$m_{\gamma\gamma} \in [100, 115] \cup [135, 180]$ GeV	$m_{\gamma\gamma} \in [115, 135]$ GeV
Leptonic, $M_{T'}$ = [600, 700] GeV	10	1
Leptonic, $M_{T'}$ = [800, 1000] GeV	15	16
Leptonic, $M_{T'}$ = [1100, 1200] GeV	14	10
Hadronic, $M_{T'}$ = [600, 700] GeV	8	4
Hadronic, $M_{T'}$ = [800, 1000] GeV	17	6
Hadronic, $M_{T'}$ = [1100, 1200] GeV	15	7

Table 4.16: Number of data events in $m_{\gamma\gamma}$ sideband and window of each optimized signal region.

Chapter 5

Summary and Conclusions

Recent results from the ATLAS and CMS experiments have confirmed the Higgs boson coupling with SM particles, in line with SM predictions. Ongoing research focuses on exploring different Higgs boson production mechanisms (ggH, VBF, VH, and ttH) and decay modes at the LHC to improve precision in measurements and search for deviations from SM predictions.

One significant aspect of the Higgs boson is its strong Yukawa coupling to the top quark, represented by y_t. Accurate measurements of y_t are vital for understanding electroweak symmetry breaking and testing theories beyond the SM. Indirect measurements of y_t rely on processes like ggH and Higgs boson to two photons ($H \rightarrow \gamma\gamma$), which involves top quark loop diagrams, while a direct measurement can be obtained from the production of Higgs boson associated with the top-antitop quark pairs ($t\bar{t}$)

Notably, the production of Higgs boson association with a single top quark (tH) is sensitive to the sign of y_t. In the SM, the relative sign of the Higgs boson's couplings to bosons and fermions is assumed to be positive, leading to destructive interference between the t-channel diagram, where the Higgs bosons are emitted from the top quark and W boson lines. As a result, the production cross section is very small, approximately 71 fb. Current measurements using data collected at the LHC have yet to achieve sensitivity to this production process. In contrast, in scenarios involving new physics, an opposite sign in the Yukawa coupling compared to the SM would result in a substantial enhancement (≈ 10 times) in the tH production cross section, while considering the coupling of Higgs boson to bosons at its SM value. In addition, during LHC Run 2, the exploration of an additional production

mode of the Higgs boson associated with a single top quark aims to enhance the sensitivity for the H $\to \gamma\gamma$ analysis and investigate any potential connections to theories beyond the SM.

As illustrated in Chapter 3, this thesis presented a detailed investigation into the tH production process, focusing on the decay of the Higgs boson into two photons and the subsequent leptonic decay of the top quark using data collected by the CMS detector during the 2016-2018 LHC run corresponding to an integrated luminosity of 137 fb^{-1}. The signal strength modifiers for the Higgs boson production associated with the single top quark mode is determined through a simultaneous fit with all other production modes of the Higgs boson and is found to be $\mu_{\mathrm{tH}} = 6.24^{+0.62}_{-0.54}(\mathrm{Syst})^{+3.67}_{-3.33}(\mathrm{Stat})$. It is important to note that the results are currently limited by statistical uncertainties, and further data is required to reduce these uncertainties and draw conclusive interpretations. The cross section of this SM Higgs production mode, convoluted with H $\to \gamma\gamma$ branching fraction ($\sigma_{\mathrm{tH}} \times \mathcal{B}_{\mathrm{H}\to\gamma\gamma}$), is observed to be $1.3^{+0.7}_{-0.7}$ fb. After combining with all other Higgs production mechanisms, the Higgs boson production signal strength in H $\to \gamma\gamma$ decay mode is observed to be $1.12^{+0.09}_{-0.09}$.

After the discovery of the Higgs boson, the SM is complete as a low-energy effective theory describing all known fundamental particles and their interactions. However, the stability of the Higgs boson mass at the electroweak scale lacks explanation. In SM, the observable Higgs mass is a residue of the bare Higgs mass subtracted by the quantum loop corrections for other SM particles. Here, the loop corrections relating to the top quarks are quite important due to its large mass, and overall the quantum corrections including the self-coupling can be enormously large up to the limit of the Planck scale (10^{19} GeV). However, experimentally the Higgs boson mass is measured to be only $125.25 \pm 0.17 GeV$; such a precarious cancellation is believed to be quite unnatural and referred as the Higgs mass Hierarchy problem. In many models and extensions of the standard model, inclusion of a new type of fourth generation of particles, called vector-like quark, T$'$ and B$'$ provides

a feasible solution.

As presented in Chapter 4, a dedicated analysis is performed to search the vector-like quark T′ production in proton-proton collisions at a center-of-mass energy of 13 TeV using data corresponding to an integrated luminosity of 137 fb^{-1} collected by the CMS detector during the 2016-2018 LHC run. The search is focused on a T′ quark which decays to a SM top quark and a Higgs boson (T′ → tH), with the Higgs boson further decaying to two photons(H → $\gamma\gamma$) and the top quark decaying either hadronically ($t \rightarrow bq\bar{q}$) or leptonically ($t \rightarrow bl\bar{\nu}$). This analysis presents the asymptotic upper limit of the T′ production signal strength ($\mu = \sigma_{obs}/\sigma_{SM}$) over the T′ mass range of 600 GeV to 1200 GeV. This search is the first T′ search to exploit the decay of the Higgs boson in the diphoton channel. The excellent diphoton invariant mass resolution of 1–2% results in an increased sensitivity compared to previous searches, for narrow T′ states with masses up to 1.1 TeV. The electroweak production of a T′ quark with mass up to 960 GeV is excluded at 95% confidence level, assuming a coupling strength $\kappa_T = 0.25$ and a relative decay width $\Gamma/M_{T'} < 5\%$

References

[1] Adrian Cho. Spontaneous symmetry breaking decanted. *Science*, 337(6100):1289–1289, 2012.

[2] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2/2017, 2016.

[3] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 2013.

[4] Detector Drawings. `https://cds.cern.ch/record/1433717`, 2012.

[5] M. Krammer. Silicon detectors in High Energy Physics experiments. *Scholarpedia*, 10(10):32486, 2015. `http://www.scholarpedia.org/article/Silicon_detectors_in_High_Energy_Physics_experiments`.

[6] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[7] https://cms-ecal-ee-rc.web.cern.ch/index.htm.

[8] A. Benaglia. The CMS ECAL performance with examples. *JINST*, 9:C02008, 2014.

[9] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. `https://cds.cern.ch/record/349375`.

[10] G. L. Bayatian et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. 2006. `https://cds.cern.ch/record/922757`.

[11] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018.

[12] David Barney. CMS Detector Slice. `https://cds.cern.ch/record/2120661`, 2016.

[13] J. A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Pérez-Victoria. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev. D*, 88(9):094010, 2013.

[14] R. L. Workman et al. Review of Particle Physics. *PTEP*, 2022:083C01, 2022.

[15] CMS Collaboration. Cross-sections for Top Partners in Single Production. *CMS top partners cross sections on GitHub*, 2018. `https://github.com/CrossSectionsLHC/TopPartners_SingleProduction`.

[16] Giuseppe Iurato. Pathways of history of elementary particle physics, 2014.

[17] Mary K. Gaillard, Paul D. Grannis, and Frank J. Sciulli. The Standard model of particle physics. *Rev. Mod. Phys.*, 71:S96–S111, 1999.

[18] Tom W. B. Kibble. The Standard Model of Particle Physics. 2014.

[19] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. arXiv: 1207.7214.

[20] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. arXiv: 1207.7235.

[21] Robert Harlander and Joshua Rosaler. Higgs naturalness and renormalized parameters. *Foundations of Physics*, 49(9):879–897, 2019.

[22] Michael Edward Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995. Reading, USA: Addison-Wesley (1995) 842 p.

[23] CERN. CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector. 2017.

[24] Albert M Sirunyan et al. A measurement of the Higgs boson mass in the diphoton decay channel. *Phys. Lett. B*, 805:135425, 2020.

[25] Vardan Khachatryan et al. Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV. *Eur. Phys. J. C*, 75(5):212, 2015.

[26] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005.

[27] G. Bellini, L. Ludhova, G. Ranucci, and F. L. Villante. Neutrino oscillations. *Adv. High Energy Phys.*, 2014:191960, 2014.

[28] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and Antimatter in the Universe. *New J. Phys.*, 14:095012, 2012.

[29] Jihn E. Kim and Gianpaolo Carosi. Axions and the Strong CP Problem. *Rev. Mod. Phys.*, 82:557–602, 2010. [Erratum: Rev.Mod.Phys. 91, 049902 (2019)].

[30] Leonard Susskind. Dynamics of spontaneous symmetry breaking in the Weinberg–Salam theory. *Phys. Rev. D*, 20:2619, 1979.

[31] Roberto Contino, Leandro Da Rold, and Alex Pomarol. Light custodians in natural composite Higgs models. *Phys. Rev. D*, 75:055014, 2007.

[32] Roberto Contino, Thomas Kramer, Minho Son, and Raman Sundrum. Warped/Composite Phenomenology Simplified. *JHEP*, 2007(05):074–074, 2007. arXiv: hep-ph/0612180.

[33] David B. Kaplan. Flavor at SSC energies: A new mechanism for dynamically generated fermion masses. *Nucl. Phys. B*, 365(2):259–278, 1991.

[34] Michael J. Dugan, Howard Georgi, and David B. Kaplan. Anatomy of a composite Higgs model. *Nucl. Phys. B*, 254:299–326, 1985.

[35] Simone Blasi and Florian Goertz. Softened Goldstone-Symmetry Breaking. *Phys. Rev. Lett.*, 123(22):221801, 2019. arXiv: 1903.06146.

[36] Maxim Perelstein, Michael E. Peskin, and Aaron Pierce. Top Quarks and Electroweak Symmetry Breaking in Little Higgs Models. *Phys. Rev. D*, 69(7):075002, 2004. arXiv: hep-ph/0310039.

[37] Oleksii Matsedonskyi, Giuliano Panico, and Andrea Wulzer. Light Top Partners for a Light Composite Higgs. *JHEP*, 01:164, 2013.

[38] Martin Schmaltz and David Tucker-Smith. Little Higgs Review. *Annual Review of Nuclear and Particle Science*, 55(1):229–270, 2005. arXiv: hep-ph/0502182.

[39] Lisa Randall and Raman Sundrum. A Large Mass Hierarchy from a Small Extra Dimension. *Phys. Rev. Lett.*, 83(17):3370–3373, 1999. arXiv: hep-ph/9905221.

[40] Marco Farina, Christophe Grojean, Fabio Maltoni, Ennio Salvioni, and Andrea Thamm. Lifting degeneracies in Higgs couplings using single top production in association with a Higgs boson. *JHEP*, 05:022, 2013.

[41] Sanjoy Biswas, Emidio Gabrielli, and Barbara Mele. Single top and Higgs associated production as a probe of the Htt coupling sign at the LHC. *JHEP*, 01:088, 2013.

[42] Fermilab. Concepts rookie book. `https://operations.fnal.gov/rookie_books/concepts.pdf`.

[43] Serguei Chatrchyan et al. Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV. *JINST*, 8:P09009, 2013.

[44] P. Adzic et al. Results of the first performance tests of the CMS electromagnetic calorimeter. *Eur. Phys. J. C*, 44S1:1–10, 2006.

[45] S. Abdullin et al. Design, performance, and calibration of the CMS Hadron-outer calorimeter. *Eur. Phys. J. C*, 57:653–663, 2008.

[46] E Barzi, N Andreev, C Boffo, E Borissov, L Elementi, L Del Frate, R Yamada, and A V Zlobin. Development and study of rutherford-type cables for high-field accelerator magnets at fermilab. *Supercond Sci Technol*, 17(5):S213, 2004.

[47] Pablo Martinez Ruiz del Arbol. The cms muon system alignment. *Journal of Physics: Conference Series*, 219(2):022014, 2010.

[48] Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical report, CERN, Geneva, 2009.

[49] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC. *J Phys G Nucl Part Phys*, 31(9):N9–N20, 2005.

[50] Serguei Chatrchyan et al. Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.

[51] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[52] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987.

[53] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[54] Regina Demina, James Dolen, Christopher Justus, Paul Tipton, Marek Zielinski, Anwar Bhatti, and Robert M Harris. Calorimeter Cell Energy Thresholds for Jet Reconstruction in CMS. Technical report, CERN, Geneva, 2006. `https://cds.cern.ch/record/926541`.

[55] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[56] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6(11):P11002–P11002, 2011.

[57] Vardan Khachatryan et al. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV. *JINST*, 10(08):P08010, 2015.

[58] Albert M Sirunyan et al. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16(05):P05014, 2021.

[59] S. Gennai, F. Moortgat, L. Wendland, A. Nikitenko, S. Wakefield, G. Bagliesi, S. Dutta, A. Kalinowski, M. Konecki, and D. Kotlinski. Tau jet reconstruction and tagging at high level trigger and off-line. *Eur. Phys. J. C*, 46S1:1–21, 2006.

[60] Nicolas Berger et al. Simplified Template Cross Sections - Stage 1.1. 2019. `https://doi.org/10.48550/arXiv.1906.02754`.

[61] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes: the discrete profiling method. 2014.

[62] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[63] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[64] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3):155, 2016.

[65] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020.

[66] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015.

[67] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017.

[68] Enrico Bothmann, Gurpreet Singh Chahal, Stefan Höche, Johannes Krause, Frank Krauss, Silvan Kuttimalai, Sebastian Liebschner, Davide Napoletano, Marek Schönherr, Holger Schulz, and et al. Event generation with sherpa 2.2. *SciPost Physics*, 7(3), 2019.

[69] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[70] A. M. Sirunyan et al. Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:185, 2018.

[71] Eleftherios Spyromitros-Xioufis, Grigorios Tsoumakas, William Groves, and Ioannis Vlahavas. Multi-target regression via input space expansion: treating targets as inputs. *Machine Learning*, 104(1):55–98, 2016.

[72] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12(02):P02014, 2017.

[73] Daniel Guest, Julian Collado, Pierre Baldi, Shih-Chieh Hsu, Gregor Urban, and Daniel Whiteson. Jet flavor classification in high-energy physics with deep neural networks. *Phys. Rev. D*, 94(11), 2016.

[74] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[75] A. Hoecker et al. TMVA - Toolkit for Multivariate Data Analysis. *arXiv:physics/0703039*, 2009. arXiv: physics/0703039.

[76] Albert M Sirunyan et al. Measurements of $t\bar{t}$H Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel. *Phys. Rev. Lett.*, 125(6):061801, 2020.

[77] F. Chollet et al. Keras, 1999. `https://keras.io/`.

[78] M. Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous systems, 1999. `https://www.tensorflow.org/`.

[79] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[80] Samuel L. Smith, Pieter-Jan Kindermans, Chris Ying, and Quoc V. Le. Don't decay the learning rate, increase the batch size, 2018.

[81] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *JMLR*, 15(56):1929–1958, 2014.

[82] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes: the discrete profiling method. *JINST*, 10(04):P04015, 2015.

[83] Vardan Khachatryan et al. Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties. *Eur. Phys. J. C*, 74(10):3076, 2014.

[84] Albert M Sirunyan et al. Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s}$ = 13 TeV. *JHEP*, 07:027, 2021.

[85] Jon Butterworth, Stefano Carrazza, Amanda Cooper-Sarkar, Albert De Roeck, Joel Feltesse, Stefano Forte, Jun Gao, Sasha Glazov, Joey Huston, Zahari Kassabov, Ronan McNulty, Andreas Morsch, Pavel Nadolsky, Voica Radescu, Juan Rojo, and Robert Thorne. PDF4LHC recommendations for LHC Run II. *J Phys G Nucl Part Phys*, 43(2):023001, 2016. arXiv: 1510.03865.

[86] Stefano Carrazza, Stefano Forte, Zahari Kassabov, Jose Ignacio Latorre, and Juan Rojo. An Unbiased Hessian Representation for Monte Carlo PDFs. *Eur. Phys. J. C*, 75(8):369, 2015. arXiv: 1505.06736.

[87] Jet algorithms performance in 13 TeV data, 2017. `https://cds.cern.ch/record/2256875`.

[88] Search for production of single vector-like quarks decaying to tH or tZ in the all-hadronic final state in pp collisions at sqrt(s) = 13 TeV. 2023.

[89] Louis Lyons. *Statistics for Nuclear and Particle Physicists*. Cambridge University Press, 1986.

[90] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[91] Andrea De Simone, Oleksii Matsedonskyi, Riccardo Rattazzi, and Andrea Wulzer. A First Top Partner Hunter's Guide. *JHEP*, 04:004, 2013.

[92] Alexandra Carvalho, Stefano Moretti, Dermot O'Brien, Luca Panizzi, and Hugo Prager. Single production of vectorlike quarks with large width at the Large Hadron Collider. *Phys. Rev. D*, 98(1):015029, 2018.

[93] Aldo Deandrea, Thomas Flacke, Benjamin Fuks, Luca Panizzi, and Hua-Sheng Shao. Single production of vector-like quarks: the effects of large width, interference and NLO corrections. *JHEP*, 08:107, 2021.

[94] Avik Roy, Nikiforos Nikiforou, Nuno Castro, and Timothy Andeen. Novel interpretation strategy for searches of singly produced vectorlike quarks at the LHC. *Phys. Rev. D*, 101:115027, 2020.

[95] Shankha Banerjee, Geneviève Bélanger, Disha Bhatia, Benjamin Fuks, and Sreerup Raychaudhuri. Phenomenological analysis of multi-pseudoscalar mediated dark matter models. *JHEP*, 07:111, 2022.

[96] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeloos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, and et al. Combination of the searches for pair-produced vector-like partners of the third-generation quarks at s=13 tev with the atlas detector. *Phys. Rev. Lett.*, 121(21), 2018.

[97] ATLAS Collaboration. Search for single production of a vectorlike quark decaying into a higgs boson and top quark with fully hadronic final states using the ATLAS detector. *Phys. Rev. D*, 105(9), 2022.

[98] CMS Collaboration. Search for pair production of vector-like quarks in the fully hadronic final state. *Phys. Rev. D*, 100(7):072001, 2019. arXiv: 1906.11903.

[99] CMS Collaboration. Search for electroweak production of a vector-like T quark using fully hadronic final states. *JHEP*, 2020(1):36, 2020. arXiv: 1909.04721.

[100] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[101] Pierre Artoisenet, Rikkert Frederix, Olivier Mattelaer, and Robbert Rietkerk. Automatic spin-entangled decays of heavy resonances in monte carlo simulations. *JHEP*, 2013(3), 2013.

[102] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.

[103] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3), 2016.

[104] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, and et al. Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1), 2020.

[105] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. *JHEP*, 2012, 2012.

[106] Paolo Nason and Carlo Oleari. NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG. *JHEP*, 2010, 2010.

[107] H. B. Hartanto, B. Jäger, L. Reina, and D. Wackeroth. Higgs boson production in association with top quarks in the POWHEG BOX. *Phys. Rev. D*, 91, 2015.

[108] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano. HW/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO. *JHEP*, 2013, 2013.

[109] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3), 2003.

[110] R. Fisher. On the interpretation of χ^2 from contingency tables, and the calculation of p. *J R Stat Soc*, 85(1):87–94, 1922.

[111] Georges Aad et al. Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector. *JHEP*, 01:069, 2015.

[112] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999.

[113] Alexander L. Read. Presentation of search results: the CLs technique. *J. Phys. G*, 28:2693, 2002.

[114] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[115] Gary J. Feldman and Robert D. Cousins. A unified approach to the classical statistical analysis of small signals. *Phys. Rev. D*, 57:3873, 1998.