Journal of Physics: Conference Series **2934** (2025) 012037 doi:10.1088/1742-6596/2934/1/012037

Search for additional Higgs bosons in fermionic final states with the CMS experiment

Punnawich Chokeprasert and Chayanit Asawatangtrakuldee*

High Energy Physics Research Unit, Department of Physics, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

E-mail: punnawich.chokeprasert@cern.ch, chayanit@cern.ch

Abstract. A search for additional Higgs boson in gluon-fusion production ($gg\phi$) and association production with at least one additional b quark ($b\phi/b\bar{b}\phi$), in fermionic final states ($b\bar{b}$, $\mu^+\mu^-$, $\tau^+\tau^-$) is performed using proton-proton collisions at a center-of-mass energy of 13 TeV. The data were collected during 2016 – 2018 by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at CERN. The new scalar bosons are predicted by various beyond standard model (BSM) theories, such as the two Higgs doublet models (2HDMs) and the minimal supersymmetric extension of the standard model (MSSM). These models are the simplest extension of the standard model by one additional scalar field, resulting to five physical states of the Higgs boson, which are three neutral Higgs bosons (ϕ represents h, H, A), and two charged Higgs bosons $H^{\pm}$. The production and decay modes allow the CMS experiment to search for additional neutral boson in the mass range of 60 GeV and up to 3500 GeV. The background processes considered include Drell–Yan (DY) process, heavy-flavor QCD multijet process, and single and pair production of t quarks, depending on the final state. No significant excess above the standard model (SM) processes is observed. Finally, the report provides model-independent upper limits at 95% confidence level on the $pp \to \phi + X$ and $pp \to b\phi/b\bar{b}\phi + X$ cross-sections in three fermionic final states, $b\bar{b}$, $\mu^+\mu^-$, $\tau^+\tau^-$, separately.

1 Theoretical Basis

While highly successful, the standard model (SM) of particle physics needs to be completed to explain certain phenomena. This has led to exploring various beyond the standard model (BSM) theories, often predicting an expanded Higgs sector. One such extension is the two Higgs doublet models (2HDMs) [1], which introduces an additional scalar doublet to the SM Higgs sector. This results in five physical Higgs bosons: two charged ($H^{\pm}$) and three neutral Higgs bosons (ϕ represents h, H, A). In the 2HDM framework, the coupling between the two Higgs doublets and quarks or leptons can vary, leading to four distinct types of 2HDMs. Our research focuses on the Type-II and Flipped 2HDMs. In the Type-II model, which mirrors the Higgs sector structure of the Minimal Supersymmetric Standard Model (MSSM) [2], the second doublet pairs to down and up-type quarks separately. The Flipped 2HDM is characterized by its coupling structure where leptons are disfavored compared to quarks. Both these models predict an enhanced coupling to b quarks. This study particularly aims to search for the three neutral Higgs bosons, ϕ, predicted by these models in $b\bar{b}$ final state.

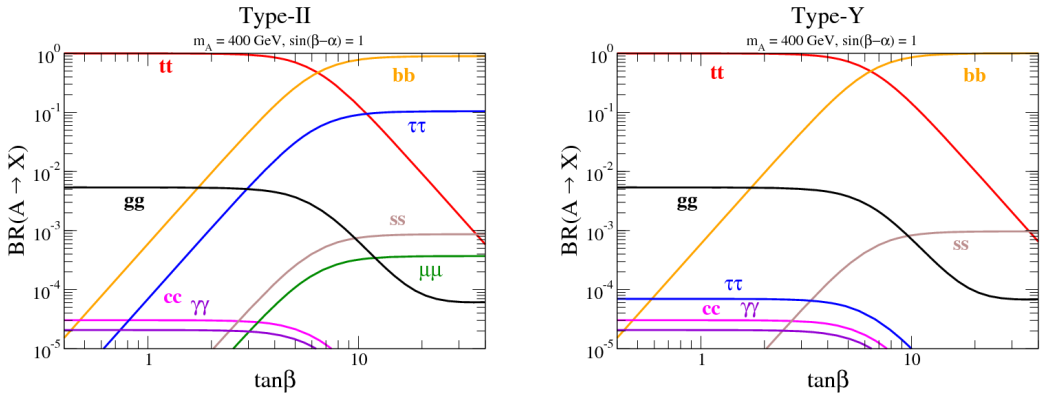

Figure 1: Cross-sections enhanced in bb channel of Type-II (Left) and Flipped or Type-Y (Right). [3]

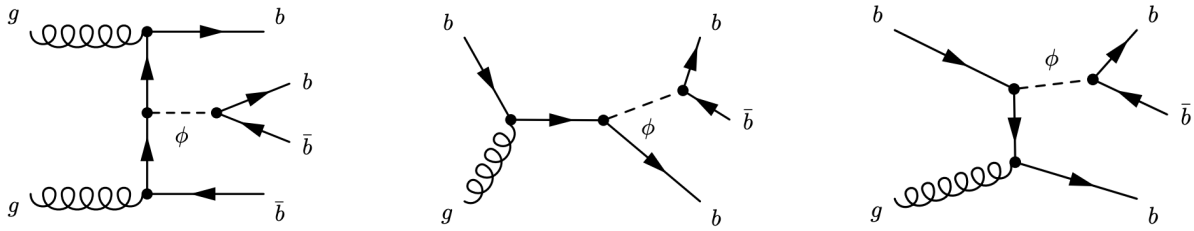

Figure 2: Feynman diagrams for the signal processes. [4]

2 Analysis Strategy

We focus on the decay of the Higgs boson into a pair of b quarks ($b\bar{b}$). The approach leverages the b-associated production mechanism as shown in Fig. 2, effectively suppressing the significant background arising from Quantum Chromodynamics (QCD) multijet process. Dedicated triggers that identify b jets during data collection enhance significantly reduce the high rate of multijet background events while maintaining sensitivity to the Higgs boson decay signals.

The measurements were done with the CMS detector using data obtained in $\sqrt{s} = 13$ TeV pp collisions during the 2017 run at the LHC. The analysis has two channels: fully hadronic (FH) and semi-leptonic (SL), each targeting different Higgs mass ranges. The FH channel probes Higgs boson masses from 300 GeV to 1800 GeV and selects events where the Higgs boson decays into a pair of b jets involving multiple high-energy jets. On the other hand, the SL channel, which focuses on Higgs masses ranging from 125 GeV to 700 GeV, includes events where one of the leading jets contains a muon, reflecting the semi-leptonic decay mode of b hadrons.

The analysis is divided into three specific regions to effectively isolate potential Higgs boson events, thereby validating our search for the Higgs.

- **Signal region (SR)** This is the primary focus of our search for the Higgs boson. Events enter the SR when the third jet achieves the specified criteria for b-tagging efficiency.
- **Control region (CR)** The CR is used to model the QCD background. It includes events in which the third jet does not meet the basic b-tagging efficiency criteria.
- **Validation region (VR)** Situated between the SR and CR, the VR validates our background model and analytical methods. Events categorized into this region contribute to the robustness and reliability of our final results.

3 Event Selection

Event selection strategy involves both trigger and offline selection criteria. The FH channel employs triggers that select events with multiple high-energy jets, requiring at least two jets to be b-tagged online.

For the SL channel, triggers are set to select events with a muon within one of the leading jets. Offline, we refine our selection further: in the FH channel, we require events to have at least three b-tagged jets identified using the DeepJet algorithm [5, 6], with the leading jets having transverse momentum (p_T) $>$ 110 GeV and 100 GeV, respectively. For the SL channel, the leading jets must have $p_T > 60$ GeV and 50 GeV, respectively, and the third jet exceeding 40 GeV in both channels. To maintain a clear dataset focus in the FH channel, we exclude events with a muon within a jet, thus differentiating these events from those in the SL channel. We impose additional geometric constraints. Jets must be within a pseudorapidity range of $|\eta| < 2.2$. The difference in pseudorapidity ($\Delta\eta_{1,2}$) between the two leading jets must be less than 1.5. The angular separation ($\Delta R_{i,j}$) between any two jets (where $i, j = 1, 2, 3$) must be greater than 1. This separation helps ensure that the jets are sufficiently distinct from one another.

4 Signal Modeling

Monte Carlo (MC) are simulated at next-to-leading order (NLO) using the POWHEG BOX-v2 [7, 8], interfaced with PYTHIA8 [9, 10] for parton showering and hadronization. This approach provides a realistic representation of Higgs production and decay processes. It covers 14 Higgs mass points per channel ranging from 125 GeV to 1800 GeV to cover the broad spectrum of potential BSM Higgs masses.

The signal distributions are modeled using a double-sided crystal ball function, which captures the core Gaussian shape of the signal and accommodates the non-Gaussian tails due to detector resolution effects and final state radiation (FSR). The width of the signal peaks is influenced by the experimental resolution and unrecovered FSR, ensuring that even small deviations from the background can be identified. The analysis covers various mass points and explores different coupling scenarios within the MSSM and 2HDMs frameworks, assessing how these affect the production and decay rates of the Higgs bosons.

5 Background Modeling

The dominant background mainly from QCD multijet production, which is characterized by multiple high-energy jets. This poses significant challenges for accurate simulation, especially after applying the stringent triple b-tag selection criteria. Given the complexity and statistical limitations of Monte Carlo (MC) simulations in this context, we adopt a data-driven approach to model the background in the SR.

Our method relies on describing the background in the SR using data from the CR combined with a transfer factor. The CR provides a robust dataset to model the background as it mirrors the SR but is less likely to include signal events. The shape of the invariant mass distribution of the two leading b jets (M_{12}) is very similar in both the SR and CR, as demonstrated by our simulations. This similarity allows us to use the M_{12} distribution from the CR to predict the background in the SR.

A transfer factor is applied to account for any subtle differences in the M_{12} distributions between the SR and CR. This factor, which corrects for minor shape discrepancies, is derived from the ratio of SR to CR in QCD MC simulations and is fitted to data. The transfer factor is typically a more straightforward function inspired by this ratio. It helps to ensure that the background model in the SR accurately reflects the observed data in the CR. For the precise modeling of the CR data, we employ a modified Gaussian function with exponential tails to fit the M_{12} distribution. The transfer factor is also parameterized using a second-degree Chebychev polynomial, which effectively captures the shape differences between the SR and CR.

6 Systematic Uncertainties

The systematic uncertainties are included into the analysis to ensure the accuracy and stability of the results. These uncertainties arise from both experimental sources and theoretical modeling.

- **Luminosity** Uncertainty in the measurement of the integrated luminosity impacts the normalization of both signal and background estimates. This is typically 2.3% for the 2017 CMS dataset.
- **Jet Energy Scale (JES) and Resolution (JER)** Variations in the jet energy calibration affect the selection efficiency and mass resolution. These are accounted for by varying the JES and JER within their uncertainties.
- **b-tagging Efficiency** The efficiency of b-jet identification and the rate of non-b jets being misidentified as b-jets introduce significant uncertainties. These are evaluated for both online (trigger) and offline (reconstruction) stages.
- **Pileup Reweighting** Uncertainties in the pileup modeling, due to variations in the number of overlapping interactions per bunch crossing, are quantified and incorporated.

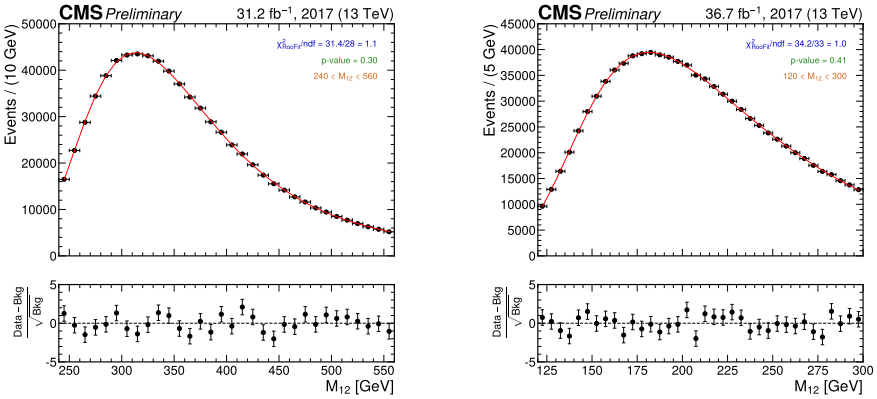

Figure 3: Background of FH (Left) and SL (Right) processes.

- **Parton Distribution Functions (PDFs)** Variations in PDFs affect the momentum distribution of partons inside protons and are included in the uncertainty evaluation.
- **Function Choice** The choice of functional form for the background fit, such as the use of Chebychev polynomials, introduces uncertainty. This is evaluated by comparing fits with different functions.
- **QCD Factorization and Renormalization Scales** Uncertainties in the choice of these scales affect the predicted cross-sections for both signal and background processes. These are evaluated by varying the scales within a factor of two.

All systematic uncertainties are integrated into the limit-setting procedure using tools like the CMS Combine framework [11], which handles both statistical and systematic uncertainties comprehensively.

7 Results

The invariant mass distributions, M_{12}, for the SR are derived from the two distinct datasets. The fits demonstrate a good agreement with the expected background distributions, indicating that no significant excess of signal events is observed above the SM background predictions.

Given the absence of a notable signal, we use the results to set upper limits at the 95% confidence level (CL) on the production cross-section times the branching fraction, $\sigma(pp \to b\phi + X)B(\phi \to b\bar{b})$. These limits are calculated using the modified frequentist CLs method and the asymptotic approximation. Systematic uncertainties are incorporated as nuisance parameters within the statistical inference. We use log-normal priors for uncertainties that affect the signal yield, while Gaussian priors are applied to shape uncertainties. This comprehensive approach ensures a robust and accurate determination of the upper limits on the Higgs boson production cross-section in the $b\bar{b}$ final state.

These upper limits indicate that there are excesses greater than 2σ due to fluctuations in the data. However, when these results are combined, the excesses disappear.

Additionally, recent studies have also focused on the $\tau^+\tau^-$ [12] and $\mu^+\mu^-$ [13] final states. In the $\mu^+\mu^-$ channel, no excess events were observed beyond the expected SM background. The 95% confidence level upper limits on the production of neutral MSSM Higgs bosons were set. In the $\tau^+\tau^-$ channel, upper limits were also set on the product of the branching fraction for decay into τ leptons and the cross-section for resonance production via gluon fusion and b-associated mechanisms. Two excesses were observed in the $\tau^+\tau^-$ channel for $gg\phi$ production at $m_\phi = 100$ GeV and 1.2 TeV, with local p-values equivalent to about 3σ.

8 Summary

The search for additional Higgs bosons in $gg\phi$ production and associated production with at least one additional b quark in fermionic final states ($b\bar{b}$, $\mu^+\mu^-$, $\tau^+\tau^-$) has been presented, with no significant excess above the SM processes observed. The analysis provides model-independent upper limits at 95% CL in the FH and SL channels of the 2017 CMS datasets. These comprehensive analyses in the three channels significantly improve sensitivity in the search for 2HDM and MSSM Higgs bosons, enhancing our understanding of potential BSM physics.

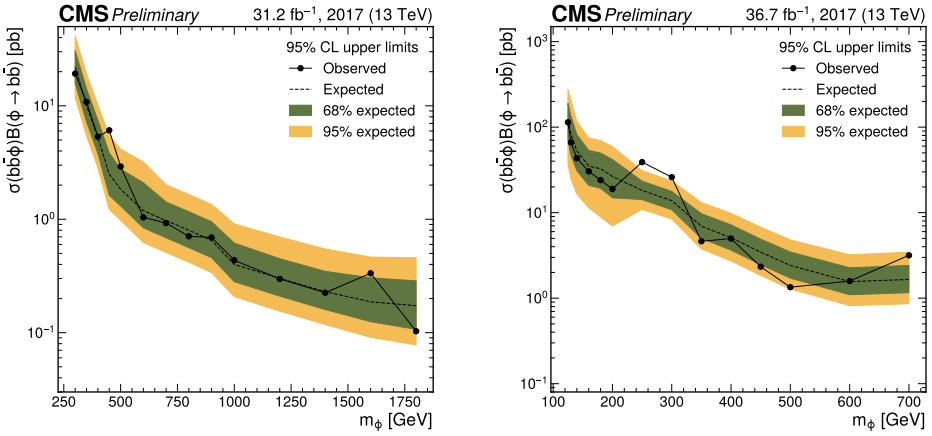

Figure 4: Expected and Observed upper limits for the Higgs b-associated production cross- section times branching fraction of the decay into a b-quark pair at 95% CL of FH (Left) and SL (Right) channels.

9 Acknowledgments

This work has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [**B39G670016**].

References

[1] Branco G C *et al.* 2012 *Phys. Rep.* **516** 1–102

[2] Nilles H P 1984 *Phys. Rep.* **110** 1–162

[3] Asner D *et al.* 2013 *arXiv preprint arXiv:1310.0763*

[4] CMS Collaboration 2018 *J. High Energy Phys.* **08** 113

[5] CMS Collaboration 2018 *J. Instrum.* **13** P05011

[6] Bols E *et al.* 2020 *J. Instrum.* **15** P12012

[7] Nason P 2004 *J. High Energy Phys.* **2004** 040

[8] Alioli S *et al.* 2010 *J. High Energy Phys.* **2010** 1–58

[9] CMS Collaboration 2020 *Eur. Phys. J. C* **80** 4

[10] Sjöstrand T *et al.* 2015 *Comput. Phys. Commun.* **191** 159–177

[11] CMS Collaboration 2024 *arXiv preprint arXiv:2404.06614*

[12] Belforte S *et al.* 2023 *J. High Energy Phys.* **2023** 1–72

[13] CMS Collaboration 2019 *arXiv preprint arXiv:1907.03152*