Searches for Higgs boson pair production in $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$ final states at Compact Muon Solenoid detector

A THESIS

SUBMITTED FOR THE DEGREE OF

Doctor of Philosophy

IN THE FACULTY OF SCIENCE

by

Lata Panwar

under the supervision of

Dr. Jyothsna Rani Komaragiri

Centre for High Energy Physics

Indian Institute of Science

BANGALORE – 560 012

JULY 2021

DECLARATION

I hereby declare that the thesis entitled **"Searches for Higgs boson pair production in $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$ final states at Compact Muon Solenoid detector"** submitted by me for the degree of *Doctor of Philosophy* to the Indian Institute of Science is a record of bonafide work carried out by me under the supervision of **Dr. Jyothsna Rani Komaragiri**, at the Centre for High Energy Physics, Indian Institute of Science, Bangalore, Karnataka, India.

I further declare that the work reported in this thesis has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university. All the sources are duly acknowledged and cited within this thesis which helped during my research work.

Place: Bangalore

Date: 28th July 2021

Lata Panwar

Certified by:

Dr. Jyothsna Rani Komaragiri

Centre for High Energy Physics,

Indian Institute of Science,

Bangalore, Karnataka, 560012

India

Dr. Jyothsna Komaragiri
Centre for High Energy Physics
IISc Bengaluru - 560 012

To

My Grandparents (Ma and Dada)

and

My Teachers

Acknowledgements

This thesis work is an outcome of support and guidance from many people. From the bottom of my heart, I am grateful to each one of them.

First and foremost, I wish to express my sincere gratitude to my supervisor Dr. Jyothsna Rani Komaragiri; without her motivation and continuous encouragement, this research work was not possible. There are many things to add, but, in short, I would add that five years did not seem like a long journey because it was an exciting time in every way. Also, thanks for improving my communication skills and developing a confident personality in me.

Along with my supervisor, I am also very grateful to Dr. Devdatta Majumder for his guidance and support at the early stage of my research career. Thanks for your patience and bearing with my poor communication skills. I owe a great debt of gratitude to Dr. Maxime Gouzevitch for helping me to understand the physics concepts I needed to carry out this research work. During the time working with them, they both were outstanding mentors. They groomed me well to be a researcher, which resulted in this thesis. Also, thanks to three of them for always believing in me and correcting me on my mistakes. This research is a result of collaboration with the CMS in association with CERN. I want to thank all of the engineers, scientists, and students for their hard work and pro-activeness in solving any technical difficulties we face during this research. I would also like to express my gratitude to the CMS collaboration for providing opportunities to participate in various physics collaborative research activities to advance my career as an experimental high-energy physicist. I want to thank Alexandra Carvalho, Paolo Azzuri, Luca Cadamuro, Nadezda Chernyavskaya, Pallabi Das, Arnab Purohit, Soumya Mukherjee, Ivan Ovtin, Tatyana Dimova, Simone Piggazini, Francesco Micheli, Rajdeep Mohan Chatterjee, Ed Scott, Prof. Kajari Mazumdar, Stepan Zakharov, Jonathon Langford, Geumbong Yu, Sandhya Jain and Spandan Mondal in particular, for being supportive during the studies of my research work. Thanks to Rostyslav Shevchenko, Antoine Lesauvage, Gouranga Kole, Andrej Saibel, Alexandros Attikis, Mauro, Chayanit, Pietro, Heiner Tholen, and Kenichi Hatekayama for all their help within CMS EPR tasks.

I am also grateful to the Indo-French Centre for the Promotion of Advanced Science (CEFIPRA) for funding my international visits, which provided me a platform for engaging with great scientists and students to develop various skills that aided me in my research projects. Thanks to SERC and AEPSHEP

Vita

EDUCATION

1. Doctor of Philosophy in Experimental High Energy Physics at Centre for High Energy Physics, Indian Institute of Science, 2016-Present, Thesis title: **"Searches for Higgs boson pair production in $b\bar{b}b\bar{b}$ and $b\bar{b}\gamma\gamma$ final states at Compact Muon Solenoid detector"**

2. Master of Science in Physics, Department of Physics, University of Rajasthan, 2014-2016

3. Bachelor of Science in Physics Honors, University Maharani College, University of Rajasthan, 2011-2014

ACADEMIC EMPLOYMENT

1. Teaching Assistant, Physical Science Building, August-December 2018, Responsibilities include assisting professors with the preparation of undergraduate courses, grading, and tutoring.

PRESENTATIONS AT PROFESSIONAL MEETINGS

1. India-CMS Collaboration Meeting, Indian Institute of Technology, Madras, India, Contributed talk with the title **"CMS Phase-2: New Physics Searches in $4b$ final states"** (August 2017)

2. XXIII DAE-BRNS HEP Symposium 2018, IITM, Chennai, India, Contributed talk with the title **"Search for vector boson fusion production of a massive resonance decaying to a pair of Higgs bosons in the four b quark final state at the HL-LHC using the CMS Phase-2 detector"** (10-14 December 2018)

3. HIGGS-2020 conference, CERN, Contributed talk with the title **"Searches for new resonances decaying to HH at CMS"**

4. XXIV DAE-BRNS HEP Symposium 2020, NISER, Bhubaneswar, India, Contributed talk with the title **"Searches for non-resonant and resonant Higgs boson pair-production, decaying into $b\bar{b}\gamma\gamma$ final state, in the pp-collisions at $\sqrt{s} = 13$ TeV, using CMS Run2 data"** (14-18 December 2020)

5. Pheno2021 Symposium, University of Pittsburgh, Contributed talk with the title **"Recent HH results (resonant and non-resonant) in CMS"**

6. **"Prospects for physics studies at the High-Luminosity LHC with CMS"**, poster at Posters@LHCC, 28 Feb 2018, CERN

7. **"Projections for di-Higgs production (resonant and non-resonant) in $4b$ final state at High-Luminosity LHC with the CMS phase-2 detector"**, poster at Posters@LHCC, 27 Feb 2019, CERN

AWARDS

1. Inspire Scholarship awarded by DST India 2011

2. Gold Medalist in B.Sc.(Honors) 2014

List of Publications

1. **Prospects for HH measurements at the HL-LHC**, *CMS-PAS-FTR-18-019*, *CERN Yellow Report*, *CERN-LPCC-2018-04*, arXiv:1902.00134, Input to European Particle Physics Strategy, 2018.

2. **Search for vector boson fusion production of a massive resonance decaying to a pair of Higgs bosons in the four b quark final state at the HL-LHC using the CMS Phase 2 detector**, *CMS-PAS-FTR-18-003*, *CERN Yellow Report*, *CERN-LPCC-2018-05*, XXIII DAE High Energy Physics Symposium, *Springer Proc.Phys. 261 (2021) 221-225*, arXiv:1812.07831, Input to European Particle Physics Strategy, 2018.

3. **Search for non-resonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV**, *JHEP 03 (2021) 257*, *CMS-HIG-19-018*, *CERN-EP-2020-222*, arXiv:2011.12373, *CERN Courier*, CMS Physics Briefing, 2021.

4. **Other CMS papers with co-authorship** InspireHEP

Synopsis

Since the discovery of the Higgs boson in 2012 at CERN, the Standard Model (SM) has become a validated theory to understand the fundamental physics of the universe. All of the SM parameters are now established as a result of this discovery. Despite this success, it does not explain several observed phenomena such as dark matter, hierarchy problems, and neutrino masses. This motivates to explore new theories beyond the SM (BSM), which could explain these phenomena while preserving the achievements of SM. Apart from this, Higgs boson self-coupling is yet to be determined precisely as Higgs boson pair production cross section is too low to measure it with current experimental potential. Nevertheless, its range constrained by current measurements allows the possibility of Higgs self-coupling interaction via BSM physics

The Large Hadron Collider (LHC) at CERN is the largest hadron collider, where hadron-hadron collisions occur at high energies. It plays a vital role in the search for new physics and SM precision measurements, including the measurements of Higgs trilinear self-coupling. For the thesis work, we use LHC data collected by Compact Muon Solenoid (CMS) experiment. It is a general-purpose detector at LHC, collecting hadron-hadron collision data to study a wide range of physics phenomena. In the central part of the CMS experiment, a superconducting solenoid of 6 m internal diameter provides a magnetic field of 3.8 T. A silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter are found within the solenoid volume, each consisting of a barrel and two endcap parts. Forward calorimeters expand the barrel and endcap pseudorapidity coverage of the detector. Outside the solenoid, muons are detected in the muon gas-ionization chambers embedded in the steel flux-return yoke. During the collision, a two-tiered trigger scheme is used to select events of physics interest. The first level (L1), made up of custom hardware processors, selects events at a rate of about 100 kHz within a time interval of less than four

microseconds using data from calorimeters and muon detectors. The second stage, named the high-level trigger (HLT), consists of a farm of processors that run a speed-optimized version of the complete event reconstruction process and reduces the event rate to about 1 kHz before data storage. Using the information from the various elements of the CMS detector, the object reconstruction and identification algorithms recreate and classify each particle in an event. The identified particles are the final objects to carry forward any analysis study.

As the last discovered piece of the SM, we use the Higgs boson as a handle to explore BSM predictions. The thesis focuses on the di-Higgs production searches, where at LHC, a pair of Higgs bosons gets produced in the proton-proton collision. We study both resonant and non-resonant di-Higgs production modes.

The non-resonant Higgs boson pair production is the only possible process within SM to study the shape of Higgs potential. Besides, the effective field theory (EFT) approach to study low energy signatures of physics existing at a high energy scale is also suitable for such a production mechanism. We explore BSM resonances for resonant di-Higgs production mode, predicted by the warped extra dimension model and the next-to-minimal supersymmetric model that directly couple with the SM-like Higgs boson. Therefore, it is more accessible to perform direct resonant searches, which could improve the overall sensitivity of di-Higgs searches at LHC.

In the first part of the research work, we perform di-Higgs searches at the high luminosity LHC (HL-LHC) in the final state of four b quarks at 14 TeV center-of-mass energy. With a vast amount of data up to 3 ab^{-1} of the total integrated luminosity, the HL-LHC would allow investigation of known SM mechanisms with high accuracy and rare new particles. The study is made using simulations of phase-2 CMS detector assuming multiple proton-proton collisions (up to 200) within each bunch-crossing. We start with the analysis for the vector boson fusion (VBF) resonant di-Higgs production in a boosted regime. The resonance is a massive spin-2 bulk KK graviton particle predicted by the warped extra dimension model. Previous s-channel searches made by the ATLAS and CMS collaborations for these resonances observed no deviation from the SM predictions, which indicates that the resonances might not directly couple with SM quarks and gluons. This inspires us to study the massive resonance production in the VBF production mode as quark-quark fusion starts contributing equally at high energies, thus enhancing the VBF production cross section. The analysis is the first one that explores the VBF resonant production mechanism. The Higgs bosons coming from the decay of massive

resonance must be sufficiently Lorentz-boosted in order to reconstruct them as a large-area jet. The signal also contains two energetic VBF jets in the forward pseudorapidity regions of the detector. Despite having the largest branching fraction with four final state b quarks among various di-Higgs modes, the channel suffers from high contamination of SM multijet processes. The unique topology of the production and decay would benefit from the upgraded phase-2 CMS detector, which features an extended tracker coverage to identify b quark originated jet and increase signal acceptance. On the other hand, a high granularity calorimeter in the forward pseudorapidity region will help to identify energetic VBF jets in the signal from the background jets. The SM multijet processes are considered as the main background. We optimize event selection for signal topology and the algorithm to identify b quark-originated jets inside the large-area jet from the Higgs boson. The systematic uncertainties are taken following the CMS recommendations for HL-LHC conditions. Expected signal significance for observation of a bulk KK graviton, having a mass between 1.5-3.0 TeV and up to 5% narrow-width, is projected, assuming 1 fb signal cross section and considering di-Higgs invariant mass as an observable. Following a similar boosted analysis strategy, non-resonant di-Higgs production for the SM and effective field theory (EFT) motivated shape benchmarks are also studied in the same final state. A 95% confidence level (CL) upper limit on the product of Higgs boson pair production cross section and branching fraction is presented for the benchmarks. This boosted strategy has not been proven optimal for the SM benchmark. However, the results project significant sensitivity for EFT motivated non-resonant di-Higgs production benchmarks at the HL-LHC.

In the second part, we use the 2016, 2017, and 2018 LHC Run-2 period data collected by the CMS detector at 13 TeV center-of-mass energy with 137 fb^{-1} total integrated luminosity and present the study for resonant di-Higgs production via gluon-gluon fusion in the final state of two photons and two bottom quarks ($b\bar{b}\gamma\gamma$) in a resolved regime. The physics is motivated by the warped extra dimension model where spin-0/2 resonance decays into two Higgs bosons and the next-to-minimal supersymmetric model where spin-0 resonance decays into a Higgs boson and another spin-0 particle different from the discovered Higgs boson. This is the first analysis in CMS collaboration, which explores an NMSSM motivated scenario in $b\bar{b}\gamma\gamma$ state. The analysis benefits from the excellent energy resolution of the CMS electromagnetic calorimeter and good trigger efficiency, which improves the invariant mass of the diphoton system resulting from the decay of the Higgs boson. This channel is the most sensitive among the di-Higgs decay

modes, especially for low resonance masses. The data-driven diphoton QCD background and simulated single Higgs production processes are used as the main backgrounds. The analysis uses machine learning methods and validates them to reject these background contamination. It increases the analysis sensitivity despite having a low di-Higgs branching fraction channel. For signal extraction, the fit is performed in a two-dimensional mass plane of diphoton and dijet invariant mass observables. The impact of systematic uncertainties on final results is found to be around 1-2%. With the narrow-width approximation, a model-independent 95% CL upper limit on the product of resonant di-Higgs production cross section and branching fraction is set for resonance mass up to 1 TeV. The results are also compared with appropriate BSM predictions to exclude allowed resonance mass ranges.

Contents

List of Tables

List of Figures

Keywords

List of Abbreviations

BSM   Beyond Standard Model

CMS   Compact Muon Solenoid

EFT   Effective Field Theory

EWSB   Electroweak Symmetry Breaking

GUT   Grand Unified Theory

LHC   Large Hadron Collider

LO   Leading Order

NLO   Next-to-Leading Order

NMSSM   Next-to-Minimal Supersymmetric Standard Model

QCD   Quantum Chromodynamics

SM   Standard Model

SPS   Super Proton Synchroton

SUSY   Supersymmetry

VEV   Vacuum Expectation Value

WED   Warped Extra Dimension

List of Notations

Λ	Energy scale
λ_{HHH}	Higgs trilinear self coupling
$\mathcal{L}$	Lagrangian
υ	Higgs VEV
G_F	Fermi coupling constant
G_{SM}	SM Gauge Group
H	Higgs boson
$h.c.$	Hermitian Conjugate
m_H	Higgs mass
M_{pl}	Plank mass
S	Action
X	BSM resonance
Y	NMSSM Higgs boson
Y	Hypercharge

Chapter 1

The Higgs Boson

1.1 Introduction

The universe has always been like a mystery box for us. The deeper we dig into it, our curiosity gets increased. For the last many centuries, human beings have been trying to understand this mystery. The foremost curiosity is the science behind the origin of the universe that could explain all the observed physics phenomena.

After studying the results from many experiments, scientists have constructed a particle theory named the standard model (SM) that consists of elementary constituents of the universe and their interactions. There is no doubt that SM is the most successful theory so far after the discovery of Higgs-like particle by the two largest collaborations ATLAS [46] and CMS [47, 48], at CERN.

However, SM only explains about 5% of the universe; the remaining 95% indicates that SM might be an effective theory that shows a low energy signature of physics existing at a high energy scale [49]. It does not explain many physics observations, e.g., gravitational force does not fit within it. However, this is not the only drawback. The SM also does not provide any explanation for dark matter and dark energy [50–53], baryongenesis [54], hierarchy problems [55], neutrino masses [56], amongst others.

The above drawbacks prove that even after being on the right track to unravel the universe, we are still far from the fundamental science behind it. This is why physicists
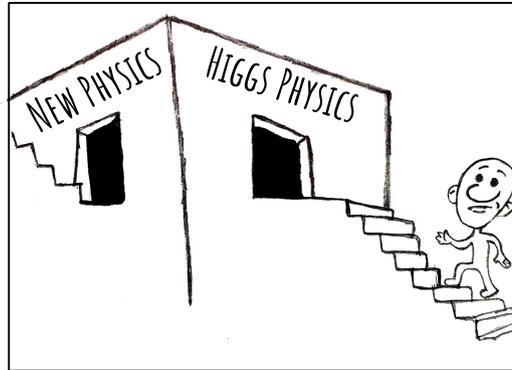
Figure 1.1: Higgs Physics: Door to new physics?

are developing new theories that could reproduce the SM observations with providing an explanation of the above mysterious phenomena. These are known as beyond the standard model (BSM) theories. It comprises supersymmetry (SUSY) [57], warped extra dimension (WED) [58] and many other models. However, the question is where to look for new physics? One possibility is that we use the Higgs boson as a stairway to reach the door of new physics.

After discovering Higgs boson at the large hadron collider (LHC) [59], CERN, all the SM parameters have been measured. The couplings of Higgs boson with the SM gauge bosons and fermions (known as Yukawa couplings) have also been measured within a certain uncertainty by studying various production and decay modes. Considering these uncertainties, we still have experimentally allowed phase space supporting the existence of the BSM physics within the SM itself. Hence SM precision measurements are essential in this prospect.

Following direct and indirect search techniques, colliders can play a crucial role in the search for the new physics (might appear at a very high energy scale) and SM precision measurements (any observed deviation from the SM predictions will be an indication for the new physics). Thus, the goals of LHC include new physics searches at high energy scale and SM precision measurements, including studies related to the self-interactions of

the Higgs boson as the SM predicts that Higgs boson interacts with itself via trilinear self-coupling. However, this trilinear coupling is yet to be determined. Experimentally, the trilinear coupling can be directly measured using the Higgs boson pair production mode $pp \rightarrow HH$, also known as **Di-Higgs production**. Within SM, non-resonant production is the only process for di-Higgs production, while the resonant di-Higgs production has its own importance to look for the new BSM particles. Briefly, we can understand the importance of di-Higgs production in two ways:

1) Non-resonant di-Higgs production: It is a direct probe for the SM Higgs trilinear self-coupling. This approach is also suitable for BSM effective field theories (EFT) searches where resonance might appear at the large TeV scale, but we look for its low energy signatures.

2) Resonant di-Higgs production: There are BSM theories that provide the solutions to the SM inconsistencies like hierarchy problems, dark matter, etc. The predicted resonances by these BSM models with an enhanced cross section directly couple to Higgs boson, which might be easier to probe using the direct search methods.

We have explored these production modes with the two exciting decay channels, 1) both Higgs boson decay into pair of bottom quark resulting in $b\bar{b}b\bar{b}$ final state, 2) one Higgs boson decays into a pair of the bottom quark, and another one decays into a pair of photon resulting in $b\bar{b}\gamma\gamma$ final state. Since the $H \rightarrow b\bar{b}$ has the large branching fraction among all Higgs boson decay modes, therefore $HH \rightarrow b\bar{b}b\bar{b}$ also has the large branching fraction (33%) but, the SM multijet backgrounds make it a challenging final state to perform any study. On the other side, despite of low $H \rightarrow \gamma\gamma$ branching ratio, $HH \rightarrow b\bar{b}\gamma\gamma$ keeps benefit of high purity and selection efficiency due to low background contamination for $H \rightarrow \gamma\gamma$ handle.

This thesis work focuses on the searches using both the di-Higgs production. We can divide this research work into two parts,

1) resonant and non-resonant di-Higgs HL-LHC projection studies in the $b\bar{b}b\bar{b}$ final state using simulations.

2) resonant di-Higgs study in $b\bar{b}\gamma\gamma$ final state with proton-proton collision data.

Within projection studies, a search for a massive resonance produced by vector boson fusion (VBF) at the high luminosity large hadron collider (HL-LHC) in proton-proton collisions at $\sqrt{s} =$ 14 TeV is explored. The resonance decays into a pair of Higgs bosons, each decaying to a $b\bar{b}$ pair. The motivation for this search also comes along with detector configurations for the phase-2 upgrade. The phase-2 CMS detector has a high-granularity calorimeter in the forward pseudorapidity region to reject the background misidentified jets and select signal VBF jets (additional jets along with Higgs boson pair) and larger tracker coverage to identity b quark originated jets (b jets) using the information of the tracks. Therefore, it might help to detect a clean signal for this topology. We analyze heavy resonance searches within this analysis, which decay into highly Lorentz-boosted Higgs bosons (i.e., Higgs bosons are in high transverse momentum (p_T) regime). The Higgs bosons are reconstructed using a single large-area jet. The analysis selections are optimized for the boosted regime. It includes a Higgs boson jet tagging (H tagging) [60] and subjet (jet within a large reconstructed jet is called subjet) b jet identification algorithms [61] to enhance analysis sensitivity. Events are categorized on the basis of the number of subjets identified as b jet. The signal sensitivity is studied for a narrow bulk KK graviton in extra-dimensional scenarios [19, 58] using a simulation of the upgraded phase-2 CMS detector, assuming multiple proton-proton collisions (pileup) in the same bunch crossing (up to 200), for data corresponding to an integrated luminosity of 3 ab^{-1}. The expected significance for different assumed masses of the bulk graviton is calculated using di-Higgs invariant mass m_{HH} as an observable and 1 fb signal cross section. The result projects a possible observation of 3 TeV bulk kk graviton using this analysis strategy. This is the first CMS analysis that explores VBF topology for resonant production mode.

For non-resonant di-Higgs studies in $b\bar{b}b\bar{b}$ final state, a similar VBF resonant analysis strategy has been used to project analysis sensitivity in the boosted regime. The boosted analysis is not optimal for the SM scenario. However, the BSM effective field theory (EFT) motivated non-resonant di-Higgs production [62] shows good analysis sensitivity due to strong contact interactions and high m_{HH} event fractions. We set 95% CL

exclusion limits on the product of non-resonant di-Higgs production cross section and branching fraction.

In the second part of research work, we search for BSM motivated resonant di-Higgs production mode in $b\bar{b}\gamma\gamma$ final state using Run-2 data of compact muon solenoid (CMS) experiment with the total integrated luminosity L=138 fb^{-1}. For these searches, two BSM models are explored.

The first one is the warped extra dimension (WED) model that predicts the existence of small and compactified extra dimensions [19, 58]. The searches are performed in the RS_bulk scenario, where the coupling of resonances with top quark and SM bosons is enhanced. The model provides an initial solution to the SM hierarchy problems. It predicts bulk radion (spin-0) and bulk KK graviton (spin-2) resonances (X) with significantly larger branching fractions to decay into the pair of the SM Higgs boson (H).

Another one is the next-to-minimal supersymmetric standard model (NMSSM) that is the most straightforward supersymmetric extension to the SM where the electroweak scale originates from the SUSY-breaking scale [63]. The model extends the Higgs sector by seven Higgs bosons allowing the heavy Higgs boson to decay into lighter Higgs boson. We study gluon-gluon fusion production of X (spin-0) resonance, i.e., heavy Higgs boson, decaying to an SM Higgs boson H and a Y NMSSM Higgs boson with the largest singlet component [63].

The analysis strategy is motivated from the CMS di-Higgs searches [1,64] with the same final state. Most of the developments and techniques have been adapted from them after a dedicated study for the analysis. Events are selected using similar selections. The analysis strategy mainly focuses on rejecting background events and extracting signal events by fitting reconstructed mass $m_{\gamma\gamma}$ and m_{jj} observables in a two-dimensional (2D) mass plane within an optimized reduced 4-body mass window. For background rejections, a multiclass BDT (boosted decision tree) classifier is trained on signal and background simulations. Based on BDT output and signal significance, the events are categorized into three categories. The impact of systematic uncertainties on the final results is around

1-2% and this analysis is statistically limited because of the purity of $b\bar{b}\gamma\gamma$ channel, therefore, statistical uncertainty dominates. The final results are presented as 95% CL exclusion limits on the product of the resonant production cross section and branching fraction for different signal mass hypotheses. The main highlight of this work is to analyze a new strategy that improves the previously published 2016 CMS data results from 6-25% [64] and NMSSM topology, which we are exploring for the first time in CMS with $b\bar{b}\gamma\gamma$ final state.

The thesis is structured in the following way: Chapter 1, motivates for the collider searches discussed within Chapter 4 and 5. Chapter 2 and 3 convey the details of the collider and experiment that provide us data used for the research work. Chapter 4 is the first part of our research work where we perform projection studies for HL-LHC by studying resonant and non-resonant di-Higgs production modes with the $b\bar{b}b\bar{b}$ final state in the boosted regime. In the second part, Chapter 5 details the collider searches of the resonant di-Higgs production mode in the $b\bar{b}\gamma\gamma$ final state using CMS Run-2 data. In Chapter 6, we conclude with a detailed summary of the main results from both the analyses performed in Chapter 4 and 5. In the appendix A, B and E, additional studies are added on which we have worked, and they are directly or indirectly adapted for our research projects. So let us start reading about my research work with all the above chapters.

1.2   The Standard Model (SM)

The SM [65–67] describes the fundamental non-gravitational forces and fundamental constituents of the universe. The fundamental constituents list contains fermions, including six quarks, three leptons and three neutrinos and gauge bosons, including eight gluons, a Z boson, two W bosons, and a photon as given in Fig. 1.2.

Within the SM, particles are considered quantum fields of a gauge theory invariant under local gauge transformations. Interactions between the gauge fields are mediated via interaction specific gauge boson known as a propagator. The SM gauge theory is
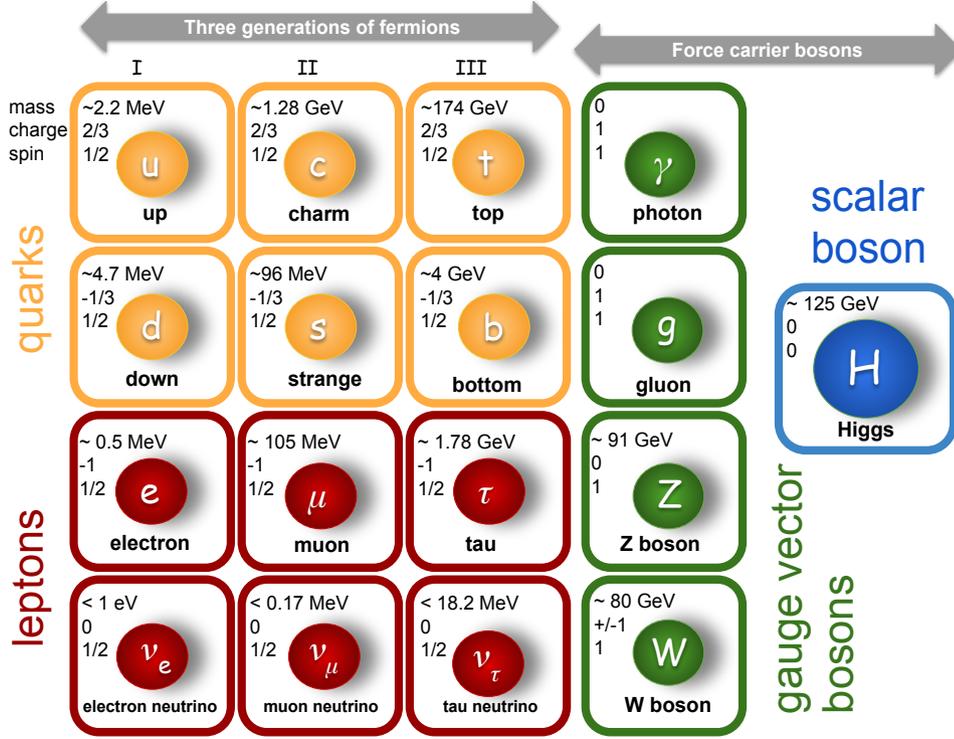
Figure 1.2: Particle spectrum of the Standard Model (image reproduced from Ref. [9])

given by gauge group [68],

$$G_{\mathrm{SM}} = SU(3)_{\mathrm{c}} \times SU(2)_{\mathrm{L}} \times U(1)_{\mathrm{Y}} \tag{1.1}$$

where the subscript c stands for colour, L stands for the left-handed chiral group, whereas Y is the hypercharge. The strong, weak, and electromagnetic forces arise from the SM gauge symmetry, and gauge bosons are quanta of the fundamental forces (i.e., gluons for strong force, W bosons and Z boson for weak force and photon for electromagnetic force). The particle spectrum and their gauge transformation properties under the G_{SM} are given as,

$$Q_i \equiv \begin{pmatrix} u_{Li} \\ d_{Li} \end{pmatrix} \sim \left(3, 2, \frac{1}{6}\right) \quad U_i \equiv u_{Ri} \sim \left(\overline{3}, 1, \frac{2}{3}\right)$$

$$D_i \equiv d_{Ri} \sim \left(\overline{3}, 1, -\frac{1}{3}\right) \tag{1.2}$$

$$L_i \equiv \begin{pmatrix} \nu_{Li} \\ e_{Li} \end{pmatrix} \sim \left(1, 2, -\frac{1}{2}\right) \quad E_i \equiv e_{Ri} \sim (1, 1, -1)$$

In the above expression, the index i runs over the three generations of SM; Q_i represents the left-handed up and down quark doublets; L_i stands for left-handed lepton doublets; U_i and D_i refer to right-handed up and down quarks, and E_i refer to right-handed leptons. The numbers given in parenthesis represents the transformation properties with respect to $SU(3)_c$, $SU(2)_L$ and $U(1)_Y$ groups. For example, for Q_i quark doublets, the equation tells that they transform as a triplet under $SU(3)_c$ of strong interactions, as a doublet under $SU(2)_L$ of weak interaction and carry hypercharge (Y/2) of 1/6.

To preserve the SM gauge symmetry under the local transformations, fermions and gauge boson are required to be massless as the mass terms are not invariant under these transformations; but, the discovery of W and Z bosons at super proton synchroton (SPS), CERN [69], in 1983 with a mass around 80 GeV and 91 GeV respectively, questioned the validation of the SM since the large mass values of W and Z cannot be considered as small fluctuations to keep going on with the SM predictions. That indicates the spontaneous breaking of the $SU(3)_c \times SU(2)_L \times U(1)_Y$ gauge symmetry. Thus, to solve this, the BEH mechanism (Brout-Englert-Higgs or **Higgs mechanism**) [70–72] was proposed by adding a new scalar particle to the SM particle content named Higgs boson.

1.2.1   Higgs mechanism

The Higgs field is a complex $SU(2)$ doublet as shown in Eq. (1.3). Its electric charge is zero; its weak isospin is 1/2, and the third component of weak isospin is -1/2; its weak hypercharge (Y) is 1.

$$\Phi \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} H_1 + iH_2 \\ H + iH_0 \end{pmatrix} \sim \left(1, 2, \frac{1}{2}\right) \tag{1.3}$$

The Higgs mechanism, referred as the spontaneous symmetry breaking of the $SU(2)_L \times U(1)_Y$ gauge symmetry associated with the electroweak force, generates masses for the particles and separates the electromagnetic and weak forces. In other words, at energies much greater than $\mathcal{O}(100)$ GeV, all these particles behave in a similar manner, but, at lower energies, $SU(3)_c \times SU(2)_L \times U(1)_Y$ symmetry is broken so that the photon and the massive W and Z bosons emerge and we end up with $SU(3)_c \times U(1)_{EM}$ gauge symmetry. The most general expression for SM Lagrangian contains three terms:

$$\mathcal{L}_{SM} = \mathcal{L}_{kin} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \tag{1.4}$$

First is the kinetic term originating from Dirac fermion spinors and vector bosons, given as follows:

$$\mathcal{L}_{kin} = i\bar{\Psi}\gamma^\mu D_\mu \Psi - \frac{1}{4}(G^{\mu\nu A}G^A_{\mu\nu} + W^{\mu\nu I}W^I_{\mu\nu} + B^{\mu\nu}B_{\mu\nu}) \tag{1.5}$$

with

$$G^A_{\mu\nu} = \partial_\mu G^A_\nu - \partial_\nu G^A_\mu + g_s f_{ABC}G^B_\mu G^C_\nu$$
$$W^I_{\mu\nu} = \partial_\mu W^I_\nu - \partial_\nu W^I_\mu + g f_{IJK}W^J_\mu W^K_\nu$$
$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$$

where $f_{ABC(IJK)}$ represent the structure constants of the $SU(3)$ $(SU(2))$ non-abelian group and D_μ represents the covariant derivative of the field given as,

$$D_\mu = \partial_\mu - ig_s G^A_\mu \lambda^A - i\frac{g}{2}W^I_\mu \tau^I - ig' B_\mu Y$$

Here Ψ represents fermion field, $A = 1, .., 8$ with G^A_μ representing the $SU(3)_c$ gauge bosons and $I = 1, 2, 3$ with W^I_μ representing the $SU(2)_L$ gauge bosons. The $U(1)_Y$ gauge field is represented by B_μ. g_s, g and g' are the coupling constants corresponding to strong,

weak and electromagnetic interactions.

The second term of Eq. (1.4) is given by

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger D_\mu \Phi - V(\Phi) \tag{1.6}$$

where Higgs potential is given as,

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi \right)^2$$

In the above equation, μ parameter belongs to Higgs boson mass and λ parameter represents Higgs trilinear self-coupling. This is the most generalized renormalizable and $SU(2)$ invariant potential. For $\mu^2 > 0$, the minimum potential energy state is at zero but, for $\mu^2 < 0$, it is not. Thus, the scalar field develops a non-zero VEV (vacuum expectation value). The direction of the minima in $SU(2)_{\text{L}}$ space is not determined because the potential depends only on $\Phi^\dagger \Phi (= H_1^2 + H_2^2 + H^2 + H_0^2)$. Thus we choose VEV (Eq. (1.7)) such that it breaks the $SU(2)_{\text{L}} \times U(1)_{\text{Y}}$ into a $U(1)_{EM}$ which is the desired symmetry.

$$< \Phi > \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.7}$$

Since Higgs is scalar particle, to preserve the electric charge only neutral component ϕ^0 of Higgs doublet in Eq. (1.3) can acquire the VEV and if we work in unitary gauge[1], goldstone bosons H_1, H_2 and H_0 are eaten away to make gauge bosons massive after EWSB (electroweak symmetry breaking); thus, we end up with,

$$\Phi \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ (H + v) \end{pmatrix} \tag{1.8}$$

---

[1]The final results of the Higgs mechanism are independent of the choice of gauge. If we do the same calculation in a gauge other than the unitary gauge, fields are parameterized accordingly. Here unitary gauge has been chosen to make the calculation easier to understand the concept.
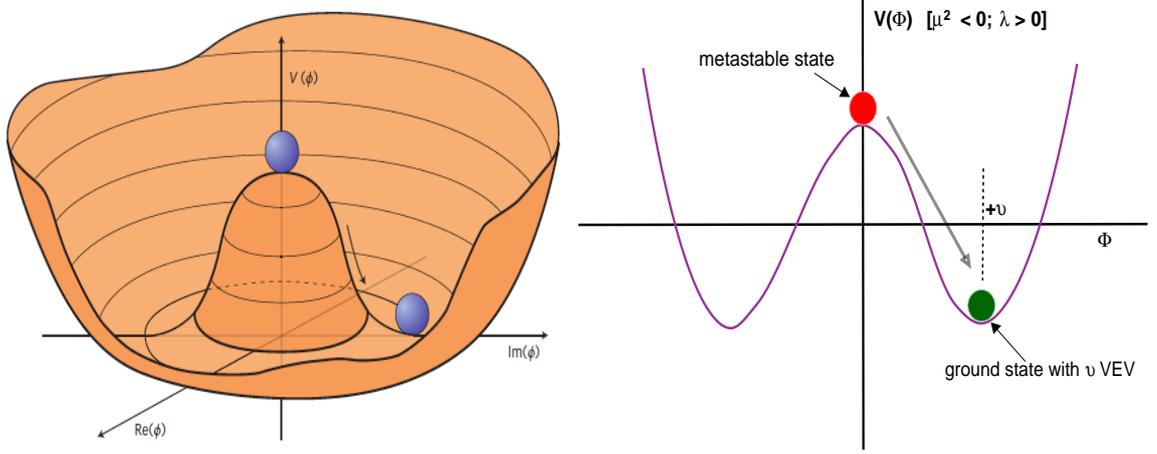
Figure 1.3: Higgs potential [10]

Replacing Φ from Eq. (1.8) within kinetic energy term of Eq. (1.6), we obtain the physical gauge field mass from Higgs mechanism:

$$(D_\mu\Phi)^\dagger D_\mu\Phi = \frac{\upsilon^2}{8}[g^2 W_\mu^\dagger W^\mu + (gW_\mu^3 - g'B_\mu)^2] \tag{1.9}$$

from this equation the mass of W boson is obtained as $m_W = g\upsilon/2$. The second term corresponds to neutral gauge bosons Z and A,
$Z_\mu = (gW_\mu^3 - g'B_\mu)/\sqrt{g^2 + g'^2}$ with mass $m_Z = \upsilon(g^2 + g'^2)/2$, and
$A_\mu = (gW_\mu^3 + g'B_\mu)/\sqrt{g^2 + g'^2}$ with mass $m_\gamma = 0$

Additionally, the fermions also interact with the Higgs boson through the dimensionless Yukawa couplings given by

$$\mathcal{L}_{\text{Yukawa}} = y_{ij}^u \bar{Q}_i U_j \tilde{\Phi} + y_{ij}^d \bar{Q}_i D_j \Phi + y_{ij}^e \bar{L}_i E_j \Phi + h.c. \tag{1.10}$$

where $\tilde{\Phi} = i\sigma^2\Phi$ (σ=Pauli matrices). These couplings make the fermions to attain masses once the gauge symmetry is broken from $G_{\text{SM}} \to SU(3)_c \times U(1)_{EM}$. Exceptionally, neutrinos do not get any mass due to the absence of right-handed neutrinos in the particle spectrum (This makes fermion mass term $m_\Psi^2 \bar{\Psi}\Psi = 0$ for neutrinos).

The parameter υ can be found from the charged current for μ decay ($\mu \to e\bar{\nu}_e\nu_\mu$) using Fermi coupling constant [12], which is measured very accurately to be $G_F =$

1.16638 × 10^{-5} GeV^{-2}. Since the momentum carried by the W boson is of order m_μ, it can be neglected in comparison with m_W, and we make the following identification

$$\frac{G_F}{\sqrt{2}} = \frac{1}{2v^2} \Rightarrow v = (\sqrt{2}G_F)^{\frac{-1}{2}} = 246 \text{ GeV} \tag{1.11}$$

One of the essential points about the Higgs mechanism is that coupling of the Higgs boson to fermions and gauge bosons are entirely determined by coupling constants and fermion or gauge boson masses. Figure 1.4 shows the Feynman diagram of Higgs boson coupling with gauge bosons (a) and fermions (b). The Higgs potential has two free parameters, μ and λ, which can be traded for,

$$v^2 = \frac{\mu^2}{2\lambda}$$
$$m_H^2 = 2v^2\lambda \tag{1.12}$$

Using the Eqs. (1.8) and (1.12) for Higgs potential (given at Eq. (1.6)), we get,

$$V = \frac{m_H^2}{2}H^2 + \frac{m_H^2}{2v^2}vH^3 + \frac{1}{4}\frac{m_H^2}{2v^2}H^4 \tag{1.13}$$

The Higgs boson couplings represent new type of interaction; very weak for light particles, such as up and down quarks and electrons, but strong for heavy particles such as the W and Z bosons and the top quark. More precisely, the SM Higgs boson couplings to fundamental fermions are linearly proportional to the fermion masses, whereas the couplings to bosons are proportional to the square of the boson masses. The SM Higgs boson couplings to gauge bosons and fermions, as well as the Higgs boson self-coupling, are shown in the Fig. 1.4. Because of charge-neutral and color-singlet property, the Higgs boson does not couple at tree level to the massless photons and gluons. Its coupling to gluons is induced at leading order by a one-loop process in which it couples to a virtual $t\bar{t}$ pair (with minor contributions from the other lighter quarks). Likewise, the Higgs boson coupling to photons is also generated via loops. In this case, the one-loop graph with a virtual WW pair provides the dominant contribution. This interferes with the smaller

contribution involving a virtual $t\bar{t}$ pair (as such, the Higgs boson coupling to photons is sensitive to the relative phase of the interactions between bosons and fermions).
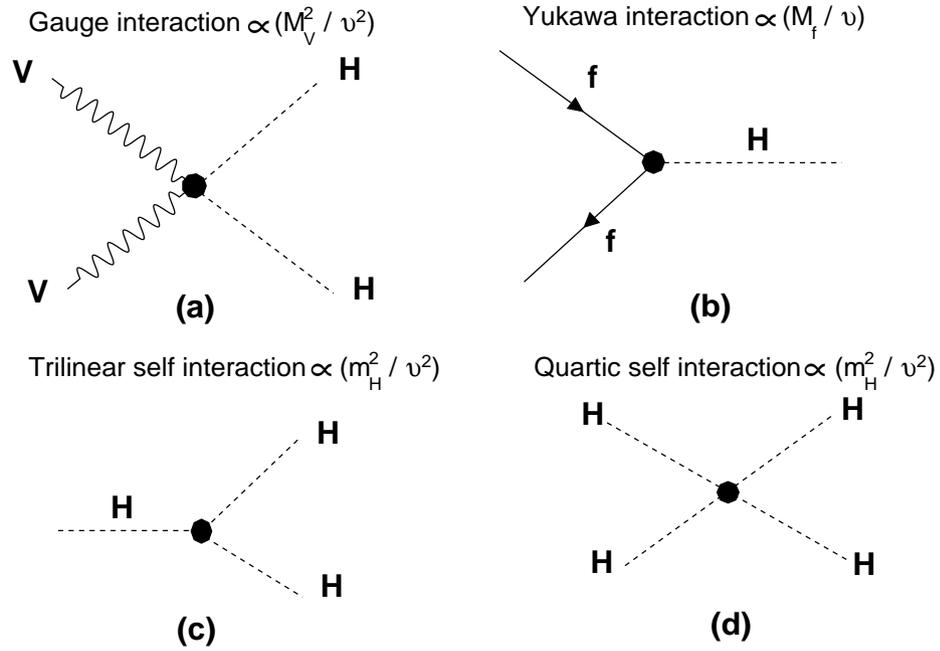
Figure 1.4: Higgs interactions (image motivated from Ref. [11])

The self-interaction of the Higgs boson is determined in terms of the Higgs boson mass. Within Eq. (1.13), the first term is Higgs boson mass, the second one corresponds to Higgs trilinear self-coupling (λ_{HHH} or λ), and the last term is for Higgs quartic coupling (λ_{HHHH}). The corresponding Feynman graphs are shown in Fig. 1.4 ((c) and (d) respectively).

1.2.2 Higgs boson profile at LHC

At LHC, after discovering the Higgs boson, extensive efforts have been made to measure the properties of this new particle. So far, experimental data show results consistent with SM predictions with Higgs boson mass (m_H) around 125 GeV [73–75] as given in Fig. 1.5. Apart from mass determination, measurements for the Higgs boson decay width [76], spin-parity [77] and couplings with other SM particles [78] have also been

made.
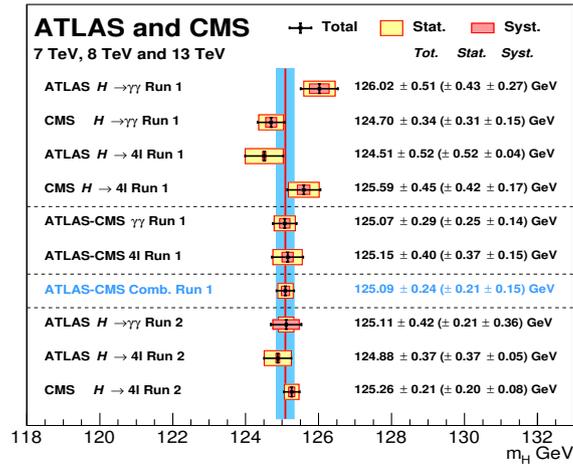
Figure 1.5: SM Higgs boson mass measurements by ATLAS and CMS with Run-1 and Run-2 data [12]

Various production processes have contributed to the precision measurement study of Higgs boson properties. The process list includes gluon-gluon fusion (ggH), vector boson fusion (VBF H), associated production with top (bottom) quarks ttH ($b\bar{b}H$) and vector bosons (VH). The respective leading order Feynman diagrams are shown in Fig. 1.6. These production mechanisms are very crucial for Higgs coupling measurement with other SM particles.

Higgs boson is not a stable particle. Its lifetime is of the order of 10^{-22} sec. As it gets produced during the collision, it decays into the SM particles, which are further identified as the final objects to carry forward any Higgs measurement study. Therefore, we collect information of the final state objects at the detector level and further analyze them to reconstruct the Higgs boson. How reconstruction is done, we will see in the upcoming chapters. However, before that, we first need to understand the probability of Higgs boson decay into any stable particles, which could be photons, jets (originating from the hadronization of quarks coming from Higgs boson decay) or leptons. Here massless neutrinos are not added as Higgs coupling to neutrino is suppressed within SM.

Figure 1.7 represents the branching ratio of Higgs boson decay channels for low m_H
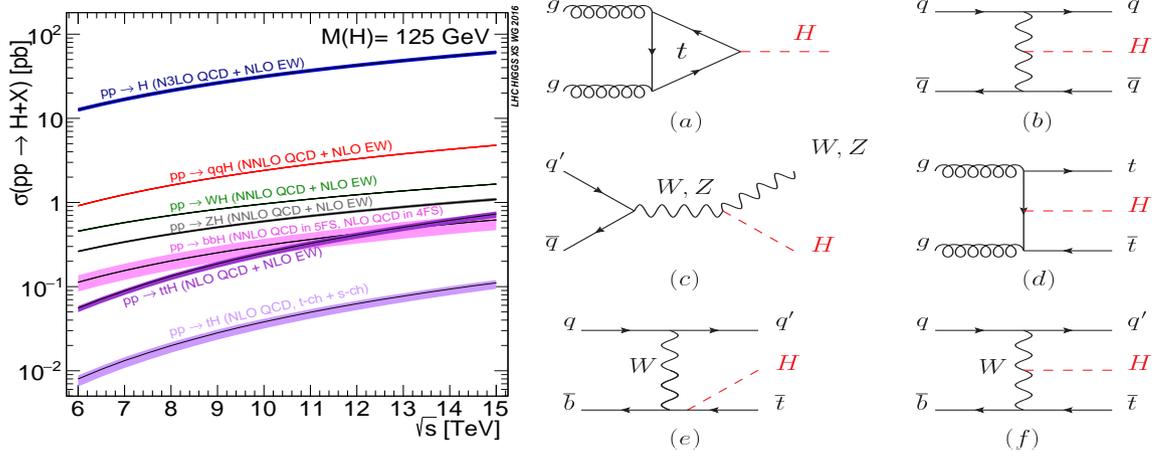
Figure 1.6: Higgs production cross section as function of the centre-of-mass energy [13–15] and corresponding contributing processes at LHC (a) gluon fusion, (b) Vector boson fusion, (c) Higgs-strahlung (or associated production with a gauge boson at tree level from a quark-quark interaction), (d) associated production with a gauge boson (at loop level from a gluon-gluon interaction), (e) associated production with a pair of top quarks (there is a similar diagram for the associated production with a pair of bottom quarks), (f-g) production in association with a top quark [12]

region. From this plot, we conclude that for the measured Higgs mass $m_H = 125$ GeV, Higgs boson decays into pair of b quark, 58% of the times. It should be a golden decay channel for Higgs physics, but, unfortunately, QCD and other SM multijet processes also can produce a similar final state. This makes it a challenging study since there will be less probability to identify the correct pair of b quark for Higgs physics measurements. This example shows how important it is to pick an appropriate decay channel for a precise measurement of any physics parameter with a huge amount of data.

1.2.3 Di-Higgs production

According to the SM predictions, di-Higgs boson production helps to understand the shape of Higgs potential and is a direct probe of Higgs self-coupling. The precise measurement of Higgs boson mass provides a precise value of the Higgs self-coupling. Figure 1.8a and Fig. 1.8b illustrate di-Higgs boson production in the gluon-gluon fusion mode via top-quark box diagram loop (depends on y_t Yukawa coupling) and the top quark triangle diagram loop (depends on Higgs trilinear self-coupling λ_{HHH}). There are
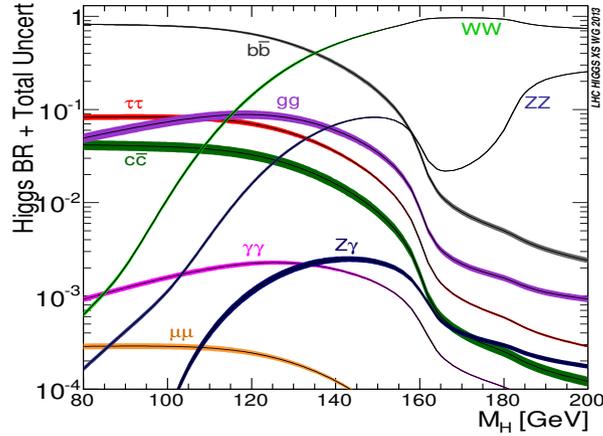
Figure 1.7: Variation in Higgs branching ratio for low m_H region [13–15]

other possible production modes like VBF $HHjj$, VHH and $t\bar{t}HH$ but gluon-gluon fusion mode accounts for 90% of the total di-Higgs production cross section.
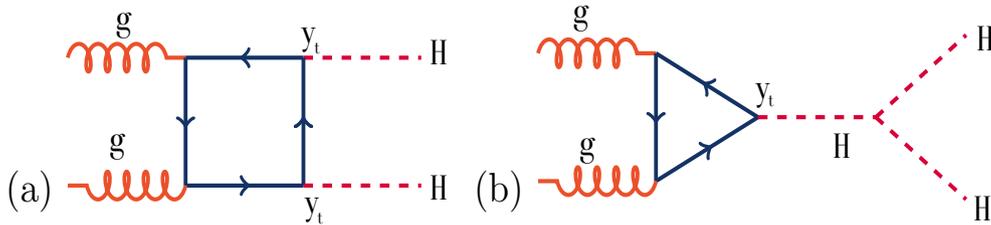
Figure 1.8: Feynman diagrams for the SM gluon-gluon fusion di-Higgs production (a) box diagram (b) triangle diagram [16]

From trilinear coupling term of Eq. (1.13), we have $\lambda_{HHH} = m_H^2/2v^2$. The experimentally measured Higgs boson mass, $m_H = 125.10$ GeV [12], implies that $\lambda_{HHH,SM} = 0.13$. The di-Higgs boson production cross section depends on the trilinear self-coupling and it is roughly 3 orders of magnitude smaller than Higgs boson production cross section. The reason behind a smaller cross section value is not only the small λ_{HHH} value, but also the destructive interference between box and triangle diagrams.

In the search for di-Higgs boson production, the crucial step is to choose a decay channel. This choice depends on the purity, selection efficiency and branching ratio of the channel. Figure 1.10 shows the possible decay channel, which are explored and
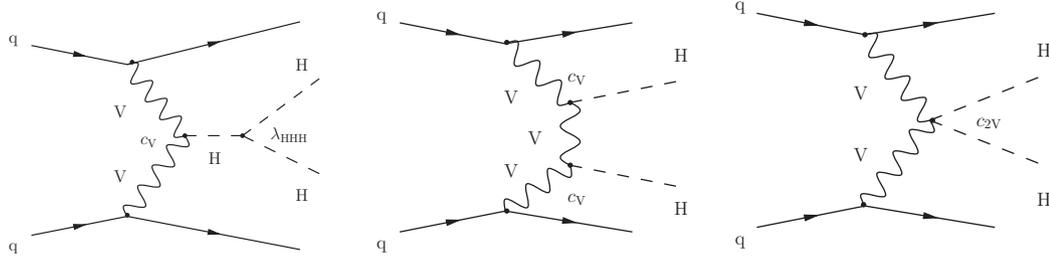
Figure 1.9: Higgs boson pair production via VBF at LO. On the left the diagram involving the HHH vertex, in the middle the diagram with two HVV vertices, and on the right the diagram with the $HHVV$ vertex [1]

analyzed by various experiments to understand di-Higgs physics.
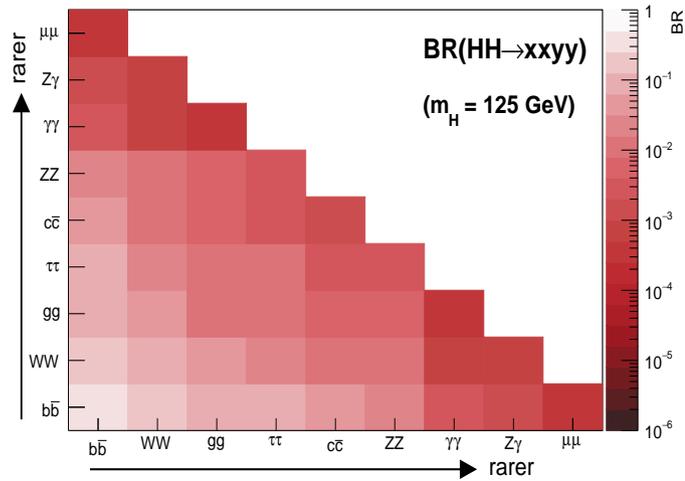
Figure 1.10: SM di-Higgs decay branching ratio (BR) for $m_H =$125 GeV

1.2.4 Why we need new physics theories?

So far the Higgs boson measurement studies, performed using the vast amount of proton-proton collision data delivered by LHC, indicate that the properties of the observed Higgs-like resonance is consistent with the SM Higgs boson. However, uncertainties in the current measurements do not exclude the Higgs boson to have non-standard model coupling as well. This points towards the existence of extended Higgs sector leading to BSM scenario. Additionally, the BSM theories are also required to explain the following

inconsistencies of the SM:

- **Quanta of gravity?:** Within SM, there is a quanta/propagator of each non-gravitational fundamental force. But spin-2 structure of the graviton makes the SM a non-renormalizable theory. Therefore, gravity does not fit within the SM.

- **Dark matter and dark energy:** From the cosmological observations and galaxy rotational curves [79, 80], it has been proven that the universe mostly consists of dark matter (27%) and dark energy (68%). Many experiments have been performed to understand the interaction of dark matter with SM particles, but a direct observation is yet to be seen. Dark energy is an unknown form of energy that affects the universe on the largest scales. It is the most accepted premise to explain the accelerated expansion of the universe. However, the SM does not provide any dark matter candidate or explain the dark energy to understand our observations.

- **Baryon asymmetry:** Following the SM baryon number and charge conservation law, it is assumed that during Big-Bang, there was an equal amount of matter and antimatter. However, with current observations, we find everything mostly made of matter. So it creates curiosity for what happened with antimatter. This asymmetry is known as baryon asymmetry or baryongenesis. No SM mechanism supports the baryongenesis. Therefore theorists has given different ideas to explain it. Sakharov suggested following set of conditions [81] to produce such asymmetry:
  a) B violation
  b) Loss of thermal equilibrium
  c) C, CP violation
  Currently there is no experimental evidence supporting any of these conditions and the amount of CP-violation, that exist in SM, is not enough to produce matter-antimatter symmetry. Thus, baryongenesis is counted as one of the great mysteries of physics.

- **Hierarchy problem:** The hierarchy problem of SM is related to the difference

in Higgs boson mass at the electroweak scale and the Planck scale (Λ_{Pl}) [82]. The calculation of the Higgs boson mass shows that mass term includes radiative corrections. These radiative corrections are quadratic in Λ, the energy scale beyond which SM is no more valid, leading to divergent Higgs boson mass beyond Λ. If $\Lambda \sim M_{Pl}$ Planck mass scale, the Higgs boson would become very massive. However, we observe the Higgs boson to be around 125 GeV after EWSB. SM does not provide any explanation of it.

- **Neutrino masses:** By studying neutrino oscillations, experiments have proven that the neutrinos are massive particle [83–85], a similar concept as kaon-system oscillations where transition or mixing occurs between flavor and mass eigenstate of neutrinos. On the other side, within the SM, neutrinos are considered to be massless particles, and it contradicts experimental observations.

To solve the above mysteries and understand the experimental observation, many BSM models have been constructed by simply extending the SM hypothesis to preserve its achievements within the BSM scenario. In the next section, we will explore the basics of few well-motivated BSM approaches and their advantages, which encourage us to search for new physics in these sectors.

1.3 Let's explore Beyond the Standard Model

As we studied within the last section the shortcomings of the SM and the necessity of the BSM physics, we will briefly go through some interesting BSM scenarios which are related to thesis work. We start with effective field theories, a very nice approach to study the low energy behaviour of physics existing at high energy scale. Apart from this, we study the extra dimensional and supersymmetric models which provide solution to SM's open problems like hierarchy problem, gravity, dark matter etc. So let's start exploring the BSM physics and its advantages as a motivation for this research work.

1.3.1   Effective Field Theory (EFT)

Before we look into EFTs approach used for the new physics searches with di-Higgs boson production mechanism; it is worth mentioning its past achievements to keep ourselves motivated with the EFT studies.

The Fermi theory of weak interactions [86,87] is an EFT for weak interactions at energies below the W and Z boson masses. It is a low-energy EFT constructed from the SM and was used for weak decay calculations even when the scales m_W and m_Z were not known. The global Fermi coupling constant G_F was measured via muon decay process, considering it as a point-like EFT interaction. The momentum transfer in this process is equivalent to muon mass, hence can be ignored, which results in $G_F/\sqrt{2} = g^2/(8M_W^2)$ (well-known equation from the calculation of muon decay using Fermi's theory). Using $G_F \sim 1/\Lambda^2$ gives $\Lambda \sim 300$ GeV. This indicates that we have an EFT with a scale of GeV order. Later, with the discovery of W and Z bosons, the EFT scale had to be replaced by the full SM using dynamical gauge fields.

Now that we have good reasons to explore the EFT scenarios, we return back to our di-Higgs boson searches. The Higgs boson couplings to gauge bosons and top quarks are currently consistent with the SM, but $\mathcal{O}(10\%)$ deviations are still possible [88,89]. Even larger deviations are not ruled out for Yukawa couplings with the first and second generation fermions and the trilinear Higgs coupling. New physics in electroweak gauge interactions usually are constrained to the $\mathcal{O}(1\%)$ level by electroweak precision measurements [90], but anomalous Higgs couplings have the potential to produce much larger effects. It then appears natural to focus firstly on the couplings of the Higgs particle. According to EFT, higher-dimensional operators may alter couplings of Higgs boson and other SM field, resulting in anomalous couplings after electroweak symmetry breaking.

Following this type of argument, one might propose a generalization of the SM in which the gauge interactions remain unchanged (at leading order), but general anomalous couplings for the physical Higgs boson are added. In theory, all of these couplings will deviate from their (dimensionless) SM values by corrections of $\mathcal{O}(1\%)$. As long

as the anomalous couplings in the Higgs sector dominate over other corrections from physics beyond the SM, the Lagrangian in Eq. (1.14) continues to explain the new physics effects for smaller deviations. Thus, di-Higgs boson production in gluon-gluon fusion can be defined as a leading approximation with the Lagrangian (in the context of BSM theories) [62, 91]:

$$
\begin{aligned}
\mathcal{L}_{\mathrm{H_{EFT}}} =\ & \frac{1}{2}\partial_\mu H \partial^\mu H - \frac{1}{2}m_H^2 H^2 - \kappa\lambda\upsilon H^3 - \frac{m_t}{\upsilon}(\upsilon + \kappa_t H + \frac{c_2}{\upsilon}HH) \\
& + \frac{1}{4}\frac{\alpha_s}{3\pi\upsilon}(c_g H - \frac{c_{2g}}{2\upsilon}HH)G_{\mu\nu}G^{\mu\nu}
\end{aligned}
\tag{1.14}
$$

This Lagrangian gives three additional process (in addition to LO SM process shown in Fig. 1.8) for non-resonant di-Higgs boson production via EFT couplings as given in Fig. 1.11 and includes five parameters; the deviation of the Higgs boson trilinear coupling λ_{HHH} (top Yukawa coupling y_t) from its SM value $\lambda_{HHH,SM}$ ($y_{t,SM} = \sqrt{2}m_t/\upsilon$) quantified by $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH,SM}$ ($\kappa_t = y_t/y_{t,SM}$), and the coefficients of three pure BSM operators which describe the contact interaction between two Higgs bosons and two top quarks (c_2), and the gluons contact interaction with one (c_g) and two (c_{2g}) Higgs boson.
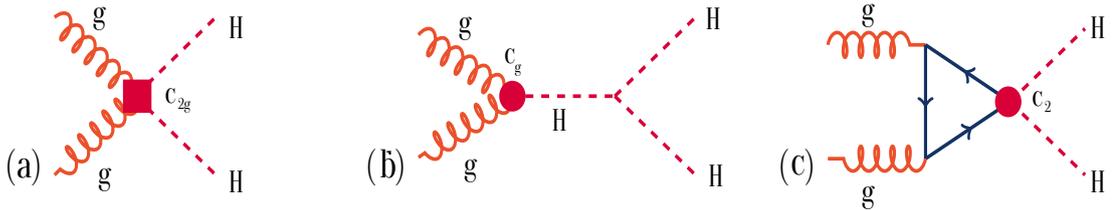
Figure 1.11: Feynman diagrams illustrating non-resonant di-Higgs production in the EFT framework [16]

Hence this five-dimensional coupling parameter space can be used to search for any deviation from SM predictions of the di-Higgs boson production mechanism. How it is done at the analysis level, we will study it in Chapter 4. Briefly, we can say that by choosing various allowed set of these five parameters, we can study which parameter set gives more consistent results with observed data. And, if it is not SM one ($\kappa_\lambda = 1, \kappa_t =$

$1, c_2 = 0, c_g = 0, c_{2g} = 0$), then it might be a hint for new physics.

1.3.2   Warped Extra Dimension (WED) model

Let us start with Newtonian Gravity, which uses an EFT approach to explain the existence of gravitational force. The gravitational potential between two masses, M_1 and M_2, is calculated as follows:

$$V(r, M_1, M_2) = \frac{1}{M_{pl}} \frac{M_1 M_2}{r} \qquad (1.15)$$

where M_{pl} is a constant Planck mass scale (10^{19} GeV) that accounts for weak strength of gravity. However, the electroweak scale exists around $\mathcal{O}(100$ GeV). This inconsistency among strengths of fundamental forces is known as the hierarchy problem.

Theorists have made various efforts to make a consistent model to explain the hierarchy problem. Supersymmetry (SUSY) and grand unified theories (GUT) introduce additional energy scales between the Planck scale and electroweak scale by extending the particle spectrum. On the other side, WED model brings down the Planck scale to a lower scale (TeV) which could be probed at the experiment.

The idea of extra dimension was introduced in 1920 by Kaluza and Klein to unify gravity with electromagnetism by assuming that gravity particles propagate through extra dimensions; hence, in 4D-space, it is not easy to probe gravitational forces. The extra dimension theories assume that the SM fields and interactions are localized on a weak brane, and gravity particles propagate in bulk between weak brane (low energy) and gravity brane (high energy). As it is known, there is no extra dimension evidence within our experimental reach; while introducing such theories, the length of extra dimension is made finite and compactified. It does not contradict our observed four space-time dimensions, which are very large (or infinite). The next question is how we can probe extra dimensions. As we do not see them in the experiment, with the assumption that the size of the extra dimensions is small, we need extremely large energies to be able to see the consequences of the extra dimensions. Thus, making the extra dimensions very

small can effectively hide these dimensions without conflicting the observations.

Two well-known BSM theories involve the concept of extra dimensions to resolve SM's discrepancy with nature. These are described in Sections 1.3.2.1 and 1.3.2.2.

1.3.2.1   Kaluza-Klein (KK) theory

It was postulated by Kaluza and Klein. To understand this concept, let us consider one extra dimension compactified on a circle with radius R [92,93] as given in Fig. 1.12. The coordinates are denoted as $x^M = (x^\mu, y)$, where M runs up to 5, μ up to 4 and y is a 5th component of x (x^5) presenting coordinate of compactified extra dimension. The compactification means that the points y and $y + 2\pi R$ are identified. So the action for a free scalar massless field ϕ can be written as,

$$S = \int d^5 x \frac{1}{2} \partial_M \phi(x^\mu, y) \partial^M \phi(x^\mu, y) \tag{1.16}$$

Because of compactification condition, $\phi(x^\mu, y) = \phi(x^\mu, y + 2\pi R)$ and we can decompose scalar field into its 4-dimension and extra dimension component using Fourier decomposition:

$$\phi(x^\mu, y) = \frac{1}{2\pi R} \sum_{n=-\infty}^{\infty} \phi^n(x^\mu) e^{i\frac{n}{R}y}, n = 0, \pm 1, \pm 2, ... \tag{1.17}$$

Plugging the Fourier series expansion into the action of Eq. (1.16),

$$S = \int d^4 x \frac{1}{2} (\partial_\mu \phi^{(0)} \partial^\mu \phi^{(0)} + \sum_{n=1}^{\infty} [\partial_\mu \phi^{(n)\dagger} \partial^\mu \phi^{(n)} - \frac{n^2}{R^2} \phi^{(n)\dagger} \phi^{(n)}]) \tag{1.18}$$

Eq. (1.18) describes a infinite series of particles with masses $m_{(n)} = n/R$ from 4D point of view. Now we implement the same with a 5D gauge field A. Following the similar approach, we can perform its Fourier decomposition,

$$A(x^\mu, y) = \frac{1}{2\pi R} \sum_{n=-\infty}^{\infty} A^{(n)}(x^\mu) e^{i\frac{n}{R}y}, n = 0, \pm 1, \pm 2, ... \tag{1.19}$$

Using Eq. (1.19) for a gauge field action, we obtain a vector field with a real scalar in

zero-mode, and non-zero modes appear as a longitudinal component of the massive 4D vector field. The interactions become strong at energy scale $E \sim 1/g_5^2 = 1/2\pi R g_4^2$ (g_4 and g_5 are 4D and 5D gauge couplings) and can only be treated as a low energy effective theory with a cutoff $\Lambda \sim 4\pi/g_5^2$. From the 4D point of view, the strong interaction comes from enhancing the number of KK modes accessible at the E energy scale. More generally, if we start with a (4+n) extra dimensional gauge theory with n dimensions compactified on a torus, the zero modes will contain a 4D gauge field together with adjoint scalars, and each non-zero KK level will have a 4D massive vector field and $(n-1)$ massive adjoint scalars.
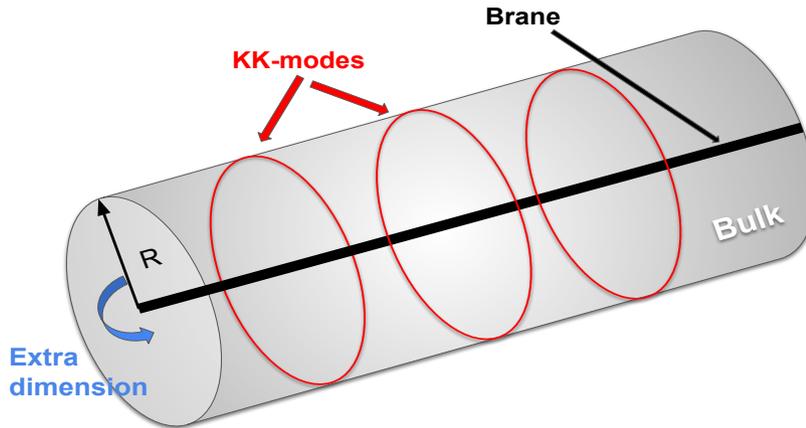
Figure 1.12: An illustration of compactified cylindrical extra dimension with Radius R and KK-modes of gravity particle traveling through bulk (image reproduced from Ref. [17])

Following the same for gravitational field, given by a symmetric metric tensor in (4+n) dimensions, the zero modes contain one 4D graviton, n massless vectors, and $n(n+1)/2$ scalars (including radion). For nonzero modes, each KK level has one massive spin-2 tensor, $(n-1)$ massive vectors, and $n(n-1)/2$ massive scalars. In the particle spectrum from this decomposition, only the 4D graviton, the radion and their KK modes interact with matter. Other fields do not couple to the matter fields on the brane directly.

1.3.2.2 Randall-Sundrum (RS) model

Following the idea of the compactified extra dimension, in 1999, L. Randall and R. Sundrum gave a theory to solve the hierarchy problem [19, 58]. The authors predicted a universe scenario with one compact extra dimension. The compactification scheme allows to describe the extra dimension as a line segment between two 4D branes, known as Planck and TeV brane, as given in Fig. 1.13.
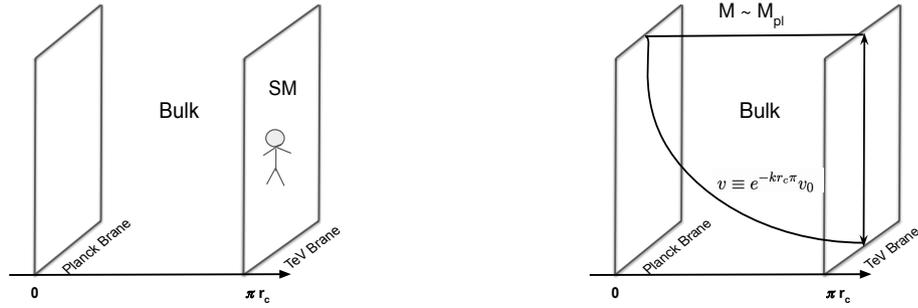
Figure 1.13: Randall-Sundrum approach of WED; SM interactions are localized to TeV brane in RS1 scenario while in RS_bulk SM particle can travel in bulk; warp factor generates the hierarchy between Planck and TeV scale given as $e^{-\kappa r_c \pi}$ (image reproduced from Ref. [18])

The model requires the existence of one extra dimension compactified on a circle of radius r_c with identity between upper and lower half (i.e. $(x^\mu, \phi) = (x^\mu, -\phi)$, ϕ is a angular coordinate for fifth dimension). In this context the most general solution to solve the classical Einstein motion equations maintaining 4D Poincare invariance is:

$$ds^2 = e^{-2\sigma(\phi)}\eta_{\mu\nu}dx^\mu dx^\nu + r_c^2 d\phi^2 \tag{1.20}$$

where μ and ν represent the coordinates of 4D-space, $\eta_{\mu\nu}$ is 4D Minkowski metric and $e^{-2\sigma(\phi)}$ is the warp factor. The full action can be written as,

$$S = S_{gravity} + S_{TeV} + S_{Planck} + S_{matter} \tag{1.21}$$

here S_{matter} represents the action of matter fields and other actions are given as,

$$S_{TeV/Planck} = -\int d^4x\sqrt{g^{(4)}}\Lambda_{TeV/Planck}$$

$$S_{gravity} = -\int d^4x \int_{-\pi}^{\pi} d\phi\sqrt{g^{(5)}}(-\Lambda_{bulk} + 2M_5^3 R) \tag{1.22}$$

where $\Lambda's$ are vacuum energy densities on the branes, R is Ricci tensor, M_5 is 5D Planck mass, $g^{(4)}/g^{(5)}$ refers to 4D/5D space-time metric. When $\Lambda_{bulk} = -\Lambda_{TeV} = \Lambda_{Planck} = \Lambda$, we get warp factor as,

$$\sigma(\phi) = r_c|\phi|\sqrt{\frac{-\Lambda}{24M_5^3}} = r_c\phi\kappa \tag{1.23}$$

where κ is a curvature factor, and it depends on negative cosmological constant and brane tensions. If in Eq. (1.22), the extra dimension is integrated out we get Planck mass, $M_{pl}^2 = M_5^3(1 - e^{2\pi\kappa r_c})/\kappa$. This also explains how an exponential hierarchy between the weak and the gravity scales arise naturally from the theory.

The Higgs mechanism can be combined with the theory assuming Higgs boson mass as M_5. The action for a fundamental Higgs field can be expressed as

$$S_H = \int d^4x e^{-4\pi\kappa r_c}[e^{2\pi\kappa r_c}\eta^{\mu\nu}D_\mu H^\dagger D_\nu H - \lambda(|H|^2 - \upsilon^2)^2] \tag{1.24}$$

where $e^{2\pi\kappa r_c}\eta^{\mu\nu}$ is metric tensor localized at TeV brane. Normalizing the wavefunction, $H \to He^{\pi\kappa r_c}$ gives,

$$S_H = \int [\eta^{\mu\nu}D_\mu H^\dagger D_\nu H - \lambda(|H|^2 - e^{-2\pi\kappa r_c}\upsilon^2)^2] \tag{1.25}$$

If Higgs field is localized to only TeV brane, the vacuum expectation value is suppressed and the physical mass scales are set by a symmetry-breaking scale given by,

$$\upsilon = e^{-\pi\kappa r_c}\upsilon_0 \tag{1.26}$$

and any mass parameter on TeV scale will correspond to physical mass,

$$m = e^{-\pi \kappa r_c} m_0 \tag{1.27}$$

Conclusively, we can say that the electroweak scale is suppressed along the extra dimension and the Planck scale is not that much affected by extra dimension for a large value of κr_c. If 5D Higgs VEV is taken of the order of Planck mass, the separation between Planck mass and electroweak (EW) scales is produced by metric with $\kappa r_c = 11$. This Planck-EW hierarchy reduction is the most successful feature of warped extra dimensional scenarios. Although κ and r_c are the only model parameters but not to deal with very large/small numbers, the free parameters are scaled as $\tilde{\kappa} = \kappa/M_{pl}$ and κr_c.
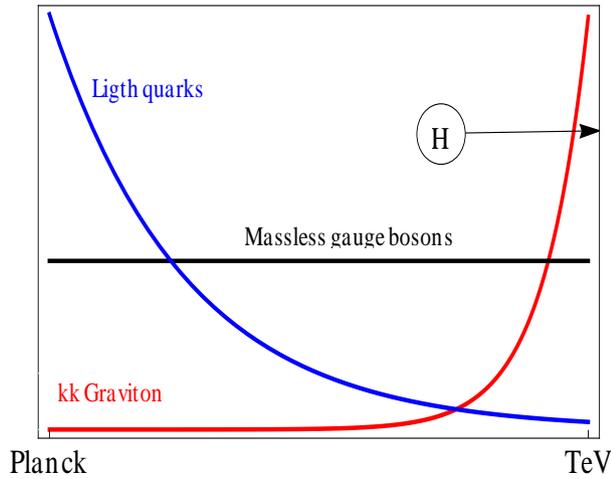
Figure 1.14: RS_bulk scenario [19]

This approach is known as the RS1 model in which only gravity particles are allowed to propagate along the extra dimension. In the updated version, matter particles are also allowed to travel along the extra dimension. This is known as the RS_bulk model. This new approach explains the mass hierarchy (why there is a big difference in the measured masses of SM particles) of the SM particles very well. The model says that heavy SM particles are localized on the TeV scale; therefore, Yukawa interaction with Higgs boson is strong. Lighter particles are localized towards the Planck scale, making the Yukawa

interaction with Higgs boson weak (as Higgs boson is only localized to TeV brane). The massless particles' profile (field component along extra dimension) remain flat along the extra dimension. This can be understood using the Fig. 1.14 which shows profiles of SM particles along extra dimension and between Planck and TeV scales.

In WED model, quantum fluctuations are considered as particles. The fluctuation modes decompose into a 4D tensor, 4D vector and scalar components, using the same Kaluza-Klein theory approach. The tensor fluctuations represent graviton (spin-2) modes, while scalar ones are for radion (spin-0) modes. The KK expansion of these modes provides particle-spectrum where zero modes of graviton and radion are massless and localized to Planck brane. The higher KK-modes of these can be massive and interact with matter particles. They are called bulk KK graviton and bulk radion particles.

In our research work, we only work with the RS_bulk scenario. Detailed phenomenology of such WED theories predicts enhanced di-Higgs boson production cross section. If the mass of bulk KK graviton and bulk radion is sufficiently large, they can decay to the di-Higgs boson final state according to branching ratio given in Fig. 1.16. The dominating resonant production mode is gluon-gluon fusion as given in Fig. 1.17a. The vector boson fusion (VBF) mode, in Fig. 1.17b, also becomes significant as we tend towards masses greater than 1 TeV since quark-quark fusion starts contributing equally which enhance the VBF rate as shown in Fig. 1.15.
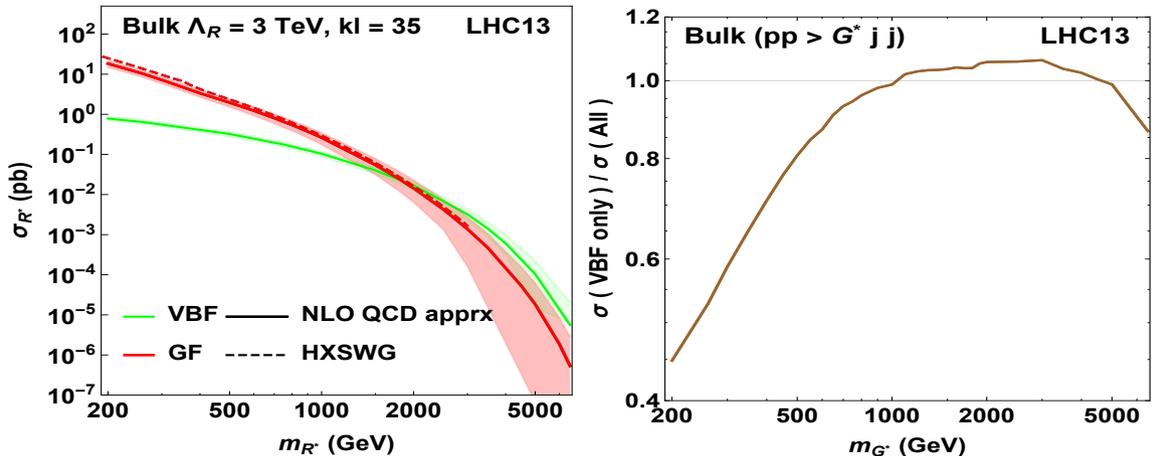
Figure 1.15: Gluon-gluon fusion and vector boson fusion (VBF) production cross section [19]; for heavy masses VBF production dominates
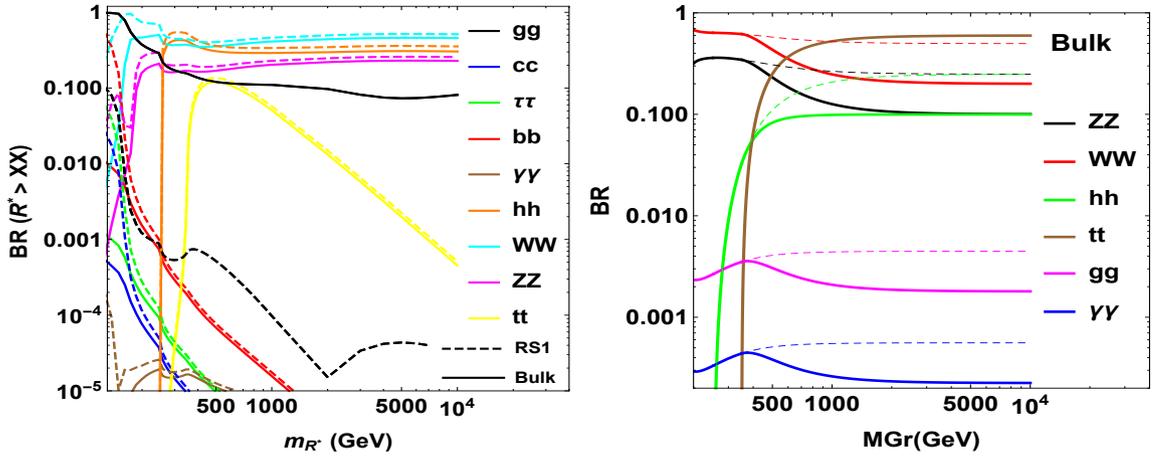
Figure 1.16: Branching fraction for bulk radion and bulk KK graviton decay [19]
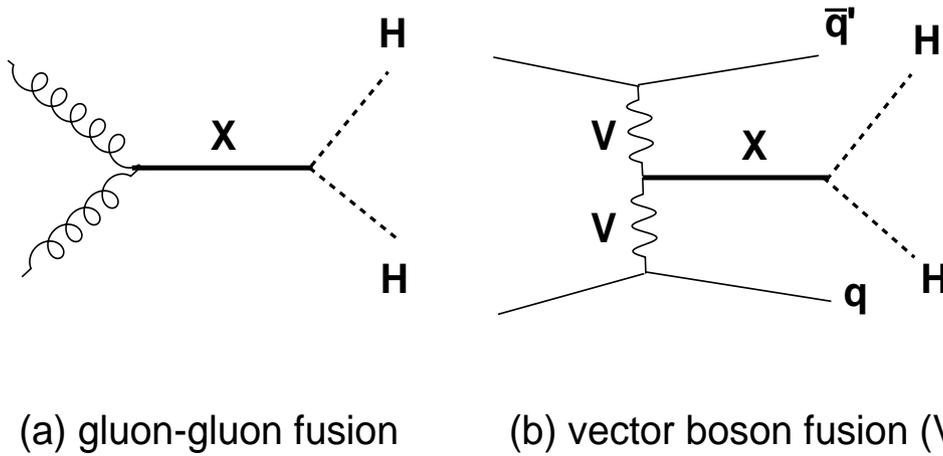
(a) gluon-gluon fusion   (b) vector boson fusion (V=W/Z)

Figure 1.17: WED (RS_Bulk) resonance production modes at LHC with di-Higgs final state; $X=$ bulk radion or bulk KK graviton

1.3.3 Supersymmetric Standard Model (SUSY)

SUSY is a generalization of quantum field theory's space-time symmetries that transform fermions into bosons, and vice versa [94]. It is an extension to the SM to understand physics at high energy scale, i.e., in low energy limits SUSY tends to be SM theory. In its particle spectrum, each SM particle has its superpartner, the spin of which differs by a half-integer, i.e., SM fermions have bosonic superpartners, and SM bosons have fermionic superpartners. The SM-SUSY particle spectrum is shown in Fig. 1.18 where SUSY
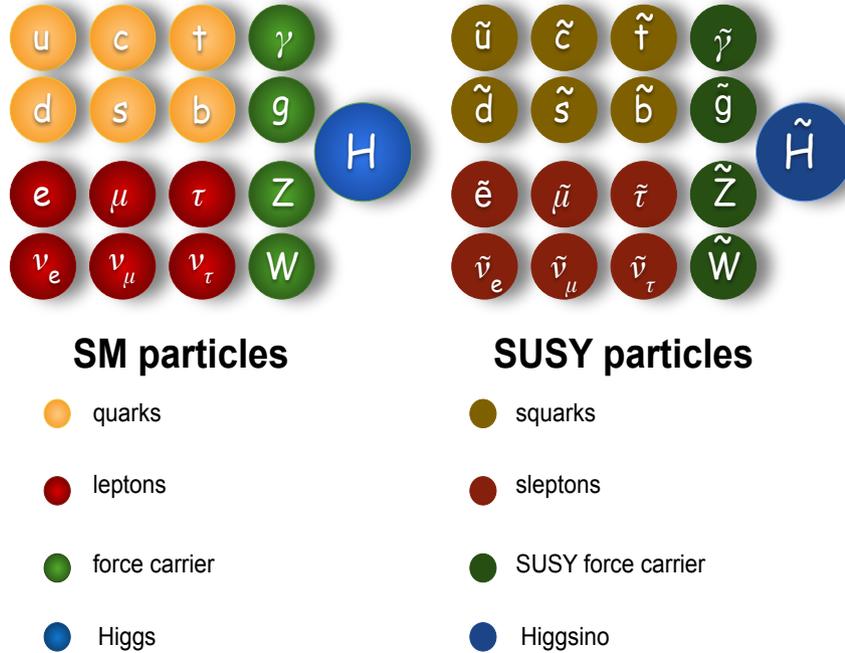
Figure 1.18: Particle spectrum of the Standard Model with their SUSY partners (image reproduced from Ref. [20])

particles are labeled with a "$\sim$" over them. The superpartners of SM quarks are squarks (stop $\tilde{t}$, sbottom $\tilde{b}$, charm squark $\tilde{c}$, strange squark $\tilde{s}$, up squark $\tilde{u}$ and down squark $\tilde{d}$), SM leptons are sleptons (selectron $\tilde{e}$, smuon $\tilde{\mu}$, stau $\tilde{\tau}$ and three sneutrino $\tilde{\nu}^{e/\mu/\tau}$), SM gauge bosons are gaugions (three Wino, Bino, eight gluinos) and SM Higgs boson is Higgsino $\tilde{H}$. The composite state of neutral SUSY particles Bino, Wino, Higgsino is known as neutralino. There are four such neutralinos labeled with $\chi_1^0, \chi_2^0, \chi_3^0, \chi_4^0$. Similarly, the composite state of charged Wino and Higgsino is called chargino and labeled with $\chi_1^\pm, \chi_2^\pm$.

The main motivations behind the ideas of SUSY can be understood by the following points:

- **Extended Higgs sector:** With additional Higgs doublet, the SUSY models extends the Higgs sector by predicting more than one Higgs bosons and enables cascade decays of Higgs. It motivates to search for di-Higgs sector using SUSY scenarios with enhanced di-Higgs production cross section.

- **Solution to hierarchy problem:** By guaranteeing the quadratic divergence of all orders cancels out in perturbation theory, SUSY could solve major hierarchy problems in gauge theory (like divergent Higgs mass radiative corrections).

- **Candidate for dark matter:** SUSY models predict the weakly interacting massive particle (WIMP), like neutralino, that does not interact via strong or electromagnetic forces. Thus if it gets to produce, it can escape from the detector without any signal. The WIMP may be a good candidate for dark matter searches.

- **Gauge coupling unification:** Within SUSY, the SM particles and their superpartners interact with same coupling strength. It predicts that SM interactions have the same strength at very high energy, a basis for GUT showing that there was only one fundamental interaction/force in the early universe after the Big Bang.

If we consider SUSY as an ideal symmetry, i.e. it does not break, SUSY sparticles will share the same mass and internal quantum numbers (besides spin) as their SM-partners. However, since we have not observed any SUSY spectrum so far, it points towards SUSY-breaking. This breaking mechanism gives higher masses to SUSY sparticles compared to their SM counterpart and keeps SUSY a valid theory within allowed ranges of masses and couplings of sparticles. In other words, we can label SUSY as a valid model only from a certain energy scale ($\mathcal{O}(\text{TeV})$) with extended particle spectrum, which can solve major problems of SM.

Now, we start going through two supersymmetric models with Sections 1.3.3.1 and 1.3.3.2 and try to understand their importance from thesis point of view. Firstly, the minimal supersymmetric extension of SM, known as the minimal supersymmetric standard model (MSSM) is described. After studying its features, its shortcomings are explained which leads towards NMSSM motivated di-Higgs boson searches.

1.3.3.1   MSSM

Pierre Fayet proposed the MSSM in 1977 as the first practical SUSY version of the standard model. It is based on the same symmetry group as SM and extends the minimal particle content [12, 68, 94, 95]. Within it, each SM particle is pair-wise combined with its superpartner into a superfield. The SUSY models are constructed using the concept of superspace that includes space-time coordinates and fermionic anti-commuting coordinates. The superfield is a function of the superspace coordinates. We deal with two types of superfields: chiral and vector superfields. The MSSM particle content is built by replacing the SM vector field with a vector superfield and the SM matter field with a chiral superfield. In simple words, the gauge boson fields and their gaugino partners are included in vector superfields, and the spin-0 Higgs supermultiplets and spin-1/2 matter fields are added in the chiral superfields. This doubles the SM particle content as we see in Fig. 1.18. The MSSM particle content and its transformation under G_{SM} is given below:

$$Q_i \equiv \begin{pmatrix} u_{Li} & \tilde{u}_{Li} \\ d_{Li} & \tilde{d}_{Li} \end{pmatrix} \sim \left(3, 2, \frac{1}{6} \right) \quad U_i^c \equiv \begin{pmatrix} u_i^c & \tilde{u}_i^c \end{pmatrix} \sim \left(\overline{3}, 1, \frac{2}{3} \right)$$

$$D_i \equiv \begin{pmatrix} d_i^c & \tilde{d}_i^c \end{pmatrix} \sim \left(\overline{3}, 1, -\frac{1}{3} \right) \tag{1.28}$$

$$L_i \equiv \begin{pmatrix} \nu_{Li} & \tilde{\nu}_{Li} \\ e_{Li} & \tilde{e}_{Li} \end{pmatrix} \sim \left(1, 2, -\frac{1}{2} \right) \quad E_i \equiv \begin{pmatrix} e_i^c & \tilde{e}_i^c \end{pmatrix} \sim (1, 1, -1)$$

The notations in Eq. (1.28) are followed from Eq. (1.2) and i stands for generation index. The main feature of MSSM is two Higgs doublets which is required for anomaly cancellation. The MSSM consists the same spectrum of chiral spin-1/2 particles (quarks and leptons) so their anomalies get canceled. The corresponding superpartners are scalars which do not contribute to anomaly. Apart from it, the superpartners of gauge bosons are left-right symmetric Majorana fermions. Hence, they contribute zero to anomalies. The only new fermionic particles are Higgsinos which may contribute to anomaly loop.

Therefore, by adding two Higgs doublet with opposite hypercharge, the anomalies caused due to one Higgs doublet get canceled by another Higgs doublet. Therefore, MSSM consists two special features; a) it uses conjugate of the right handed particles instead of right handed particles itself, b) it has two Higgs superfields- one gives mass to up-type quarks and other gives mass to down-type quarks and charged leptons (also known as TypeII-2HDM model). The two Higgs superfields and its transformation under G_{SM} are given as,

$$
\begin{aligned}
H_u &\equiv \begin{pmatrix} H_1^0 & \tilde{H}_1^0 \\ H_1^- & \tilde{H}_1^- \end{pmatrix} \sim \left(1, 2, \frac{-1}{2}\right) \\
H_d &\equiv \begin{pmatrix} H_2^+ & \tilde{H}_2^+ \\ H_2^0 & \tilde{H}_2^0 \end{pmatrix} \sim \left(1, 2, \frac{1}{2}\right)
\end{aligned}
\tag{1.29}
$$

In the above equation, $\tilde{H}$ are the Higgsinos and they are fermionic superpartners of Higgs.

After constructing the MSSM particle spectrum, it should also be tested if MSSM Lagrangian is renormalizable including gauge invariance under SM gauge symmetry. The conservation of the baryon and lepton number arise naturally within the SM construction. But this is not the case with MSSM. Therefore, in MSSM, R-parity symmetry ($R = (-1)^{L+3B+2S}$) is imposed to evade the baryon and lepton number violating terms. The value of the R quantum number is 1 and -1 for SM fields and their superpartners, respectively that makes it a conserved symmetry for every superfield. Also, the R-parity condition claims existence of a weakly interacting stable neutral SUSY particle which might be a good dark matter candidate, e.g., neutralino.

Another important point is the cancellation of divergent Higgs boson mass corrections. In SM, Higgs radiative mass corrections are quadratic due to the fermionic loop, and corresponding superpartner scalar squarks contribute to Higgs mass correction with an opposite sign as shown in Fig. 1.19. Thus, both get to cancel out and theory change logarithmically with the new physic scale. However, for a considerable large fine-tuning, the SUSY particle masses should not exceed the TeV scale. With this constraint,

the MSSM can be considered a renormalizable theory.

Figure 1.19: Cancellation of the Higgs boson quadratic mass corrections between fermionic quark loop and scalar squark loop Feynman diagrams in MSSM (image reproduced from Ref. [21])

The MSSM Lagrangian is constructed by including all possible interactions that follow SM gauge invariance and conservation laws. The SM gauge couplings, Higgsino mass parameter and Yukawa couplings of one generation of left and right-handed leptons and quarks, and their superpartners to the Higgs bosons and Higgsinos, all are SUSY-conserving parameters. Within MSSM Lagrangian, these parameters enters via R-parity conserving superpotential,

$$W_{MSSM} = Y_u \hat{Q} \hat{H}_u \hat{U} - Y_d \hat{Q} \hat{H}_d \hat{D} - Y_e \hat{L} \hat{H}_d \hat{E} + \mu \hat{H}_u \hat{H}_d; \ \mu = \text{Higgsino mass parameter} \tag{1.30}$$

After this, soft SUSY breaking terms are added, which are consistent with the gauge symmetries. The tree level MSSM Higgs potential is given in Eq. (1.31) which has SUSY breaking two real squared Higgs mass parameters m_1^2, m_2^2 and one complex squared-mass parameter m_{12}^2.

$$V_{MSSM}^{soft} = (m_1^2 + |\mu|^2) H_d^\dagger H_d + (m_2^2 + |\mu|^2) H_u^\dagger H_u + (m_{12}^2 H_u H_d + h.c.) +$$
$$\frac{1}{8}(g^2 + g'^2)(H_d^\dagger H_d - H_u^\dagger H_u)^2 + \frac{1}{2} g^2 |H_d^\dagger H_u|^2 \tag{1.31}$$

where $m_{u,d}^2 = \mu^2 + m_{1,2}$ and other terms correspond to Higgs quartic couplings given in terms of gauge couplings as a consequence of SUSY.

The MSSM two Higgs doublets holds eight degree of freedom (dof). After EWSB, both Higgs doublets get VEV v_d and v_u and five Higgs bosons get generated consisting two CP-even Higgs (H and H'), a CP-odd Higgs (A) and two charged Higgs ($H^\pm$) bosons. Other three dof are Goldstone bosons which makes SM vector bosons massive particles. We can choose MSSM Higgs potential minima such that v_d and v_u are real and positive with constraint $v = \sqrt{v_d^2 + v_u^2} \sim 246$ GeV. The ratio of both the VEVs is defined as parameter $tan\beta = v_u/v_d$; $(0 \le \beta \le \pi/2)$. For MSSM, there are only two free parameters at tree level, $tan\beta$ and m_A (mass of CP-odd Higgs A). Other Higgs masses are given in terms of these two parameters. The tree-level light Higgs mass takes maximum value for $m_A >> m_Z$ with condition $m_H \le m_Z cos2\beta$ [96].

The discovered SM-like Higgs boson, if interpreted as the lightest MSSM Higgs boson H with a mass of around 125 GeV, at tree-level couplings of two CP-even MSSM Higgs bosons to vector boson $V(=W/Z)$ will be given as,

$$g_{HVV} = g_V m_V sin(\beta - \alpha),$$
$$g_{H'VV} = g_V m_V cos(\beta - \alpha) \; ; \; g_V = \frac{2m_V}{v} \tag{1.32}$$

Where α with $(-\pi/2 \le \alpha \le 0)$ represents the mixing between Higgs doublets to form CP-even Higgs bosons. From above equation we observe that in the limit $cos(\beta - \alpha) \to 0$, the lightest CP-even Higgs will behave as SM Higgs boson (H). A and $H^\pm$ do not couple to gauge bosons at tree level. The relation between Yukawa couplings and fermions masses is given by

$$y_{b,\tau} = \frac{\sqrt{2}m_{b,\tau}}{vcos\beta}, \; y_t = \frac{\sqrt{2}m_t}{vsin\beta} \tag{1.33}$$

and the coupling to all three Higgs with fermions $(f\bar{f})$ are given as,

$$
\begin{aligned}
Hb\bar{b}/h\tau^{+}\tau^{-} &\propto \frac{-sin\alpha}{cos\beta}, \ \ Ht\bar{t} \propto \frac{cos\alpha}{sin\beta} \\
H'b\bar{b}/H\tau^{+}\tau^{-} &\propto \frac{cos\alpha}{cos\beta}, \ \ H't\bar{t} \propto \frac{sin\alpha}{sin\beta} \\
Ab\bar{b}/A\tau^{+}\tau^{-} &\propto \gamma_5 tan\beta, \ At\bar{t} \propto \gamma_5 cot\beta
\end{aligned}
\tag{1.34}
$$

MSSM Higgs Phenomenology: Presence of MSSM signatures may modify SM Higgs boson couplings and branching ratio [96]. If the MSSM lightest CP-even Higgs is SM Higgs boson, it can also decay into the lightest SUSY particle. Apart from this, coupling of light stops, sbottoms with photon and/or gluon can also sizeably deviate Higgs loop-induced interaction to gg and $\gamma\gamma$. At hadron colliders, gluon fusion, mediated by loops comprising heavy top and bottom quarks and the corresponding SUSY partners, is the dominant neutral Higgs boson production mechanism at moderate values of $tan\beta$. In the larger $tan\beta$ region, neutral Higgs boson coupling with bottom-quarks gets enhanced. The SM di-Higgs boson production box and triangle diagrams, in low-energy supersymmetric models, receive additional contributions from loops mediated by the third generation of squarks. Furthermore, the triangle diagrams are linked to the trilinear Higgs coupling, which is crucial for reconstructing the Higgs potential. As a result, new physics contributions are especially sensitive to the HH production searches. ATLAS and CMS have studied various channel considering all possibilities for MSSM signal. The precision measurements of Higgs physics in the future, combined with data from heavy Higgs boson searches, might provide a depth of information about the SUSY parameter space.

Why MSSM fails?: The Higgs mass parameter μ given in Eq. (1.34) is a SUSY preserving parameter up to Planck mass scale, but for EWSB, it is required to be at electroweak scale. This is known as **μ-problem** [97]. A Type-II 2HDM model with an additional singlet (NMSSM) provides a solution to the μ-problem. This is explained in the next paragraph.

1.3.3.2 NMSSM

The NMSSM is the most straightforward supersymmetric extension of the SM where the electroweak scale originates from the SUSY-breaking scale only. It accommodates not only the SM Higgs-like boson but also preserves all the MSSM advantages. It adds a complex scalar singlet field $\hat{S}$ to MSSM Lagrangian, which couples only with Higgs doublet fields of MSSM in the superpotential by redefining MSSM μ parameter with VEV of $\hat{S}$ field [12, 94, 95] as $\mu_{eff} = \lambda_s v_s$. λ_s is coupling between Higgs singlet and doublet fields. For Higgs potential minimization when $\hat{S}$ also acquires a VEV v_s, all three VEVs should be the order of M_{SUSY} effective SUSY-breaking scale. This brings down the μ_{eff} from M_{pl} to electroweak scale. Hence μ-problem gets solved. With this solution, the NMSSM enriches the Higgs sector by extending the MSSM Higgs sector with two more Higgs bosons. Thus, we have seven Higgs bosons; three CP-even scalars, 2 CP-odd pseudoscalars, and two charged Higgs bosons. The NMSSM superpotential and soft-breaking terms in the Higgs-sector are given by Eqs. (1.35, 1.36). In these equations, λ_s and κ are dimensionless coupling coefficients and A_{λ_s} and A_κ are the soft-breaking parameters. Other notations are as same as MSSM superpotential and soft-breaking Higgs potential in Eqs. (1.30, 1.31). At the tree level, the Higgs sector of the NMSSM can be represented by six independent parameters, λ_s, κ, $tan\beta$, μ, m_A, A_κ with $m_A^2 = 2\mu(A_\mu + \kappa v_s)/sin2\beta$.

$$W_{NMSSM} = Y_u \hat{Q}\hat{H}_u\hat{U} - Y_d\hat{Q}\hat{H}_d\hat{D} - Y_e\hat{L}\hat{H}_d\hat{E} + \lambda_s\hat{H}_u\hat{H}_d\hat{S} + \frac{1}{2}\kappa\hat{S}^3 \tag{1.35}$$

$$V_{NMSSM}^{soft} = m_u^2|H_u|^2 + m_d^2|H_d|^2 + m_s^2|S|^2 + (A_{\lambda_s}\lambda_s\hat{S}\hat{H}_u\hat{H}_d + \frac{A_\kappa}{3}\kappa^3 + h.c.) \tag{1.36}$$

The $\lambda_s\hat{H}_u\hat{H}_d\hat{S}$ term in superpotential changes the tree-level Higgs mass terms as $m_H^2 = m_Z^2 cos^2 2\beta + \lambda_s v^2 sin^2 2\beta$. It enhances the SM Higgs mass at the tree level. The mixing between Higgs singlet and doublet fields can further increase the SM Higgs mass up to 125 GeV. Such phenomena make large radiative corrections from top-squark loops unnecessary to predict a Higgs boson at 125 GeV. Due to singlet-doublet Higgs fields mixing, the three CP-even fields mix to form H_1, H_2, H_3 and 2 CP-odd fields mix to form

A_1, A_2. Therefore, physics of the Higgs bosons, i.e., masses, couplings and branching ratios, can differ significantly from the MSSM. In many regions of the allowed parameter space, one of the lighter two CP-even Higgs bosons will have a large singlet component. Similarly, the A_1 can easily have a large singlet component.

NMSSM Higgs phenomenology: In the NMSSM scenario, gluon-gluon fusion mode of the Higgs pair-production dominates at the LHC as given in Fig. 1.20. Exploring the NMSSM phenomenology in light of the 125 GeV Higgs boson and identifying benchmark scenarios with new topologies that include Higgs boson decay chains has sparked much interest. It also allows to have a fundamental theory valid up to the Planck scale and yielding a rich and interesting Higgs boson phenomenology at the LHC. The extended CP-even and CP-odd Higgs sectors allow for the possibility of additional Higgs-to-Higgs decays compared to the MSSM.

In this thesis work, the gluon-gluon fusion production of NMSSM neutral heavy Higgs boson X has been analyzed. X further decays to NMSSM neutral light Higgs boson Y, and SM CP-even Higgs H (X and Y can be a CP-even or CP-odd Higgs bosons) [63] with a mass of $m_Y < m_X - m_H$. The Y Higgs boson could have a dominant admixture of singlet, leading to the suppression of its couplings to SM particles and thus of its direct production at the LHC. The production of an X boson and subsequent decay into YH might become the dominant Y production mechanism. In spite of the small couplings to SM particles, the branching fraction of Y is assumed to be as same as H. This happens in NMSSM where Y is closer to SM Higgs boson mass since mixing angle of Y and H are largest. Therefore couplings of Y with SM particles are proportional to H following the same branching fractions for Y and H. As Y tends to be heavy particle, this is no longer true.

Although considering MSSM possibilities, various searches have been performed at LHC and LEP. However, this is the first time searches for the NMSSM Higgs pair production process using CMS data at LHC are performed. A wide range of allowed X and Y mass parameters are explored within this search, and exclusion limits are set
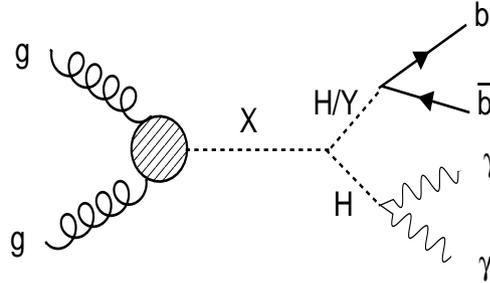
Figure 1.20: Feynman diagram showing the gluon-gluon production of NMSSM X resonance that further decays to Y and H particle with a $b\bar{b}\gamma\gamma$ final state

on resonance production cross section.

1.4   Current status of di-Higgs boson searches at LHC

This section presents an overview of the results of previous searches similar to signal process we are studying in this thesis work. Searches for di-Higgs boson production processes, in which the Higgs boson is used as a platform to look for BSM phenomena, have been carried out in a number of different ways, depending on the subsequent decays of the Higgs boson. At LHC, both resonant and non-resonant di-Higgs boson production have already been performed by ATLAS and CMS collaborations using various decay channels as $b\bar{b}b\bar{b}$, $b\bar{b}VV$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$, $WW\gamma\gamma$, $WWWW$ using Run-1 and Run-2 data.

1.4.1   Non-resonant di-Higgs boson searches

ATLAS and CMS have performed searches for non-resonant di-Higgs boson production for Higgs trilinear self-coupling and cross section or signal strength ($\mu = \sigma_{HH}/\sigma_{HH}^{SM}$) measurements, using 2016 Run-2 data at $\sqrt{s} = 13$ TeV. The SM predicts an inclusive 31.05 fb (NNLO) gluon-gluon fusion di-Higgs boson production cross section [98] which is small. Therefore, to obtain measurements with accuracy various decay channels are combined for the final results. The combined results from both the collaborations for 2016 Run-2 data at 13 TeV, are added in Figs. 1.21 and 1.22. In Table 1.1, the 95% CL exclusion limits on di-Higgs boson production cross section and constraints on "κ_λ"

parameter are provided. For BSM EFT shape benchmarks (defined in Ref. [62]), CMS has also provided upper limits on cross section as given in Fig. 1.21c. The results are consistent with the SM model predictions within 2σ deviation.

Searches, using 2016+2017+2018 Run-2 data, are also being carried out by both the collaborations. CMS has made results public for $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$ and $bbZZ$ channels while ATLAS has public results for $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$ and $b\bar{b}l\nu l\nu$ channels with gluon-gluon fusion and vector boson fusion di-Higgs boson production modes. The results are summarized in Table 1.2.

Table 1.1: Combined observed (Expected) 95% CL exclusion limits for non-resonant SM di-Higgs boson searches by ATLAS and CMS using 2016 Run-2 data at 13 TeV

95% CL limits on	ATLAS	CMS
$\sigma(pp \to HH)$ [fb]	6.9 (10) × SM	22.2 (12.2) × SM
κ_λ	-5.0 (-5.8) < κ_λ < 12.0 (12.0)	-11.8 (-7.1) < κ_λ < 18.8 (13.6)

Table 1.2: Observed (Expected) 95% CL exclusion limits for non-resonant SM di-Higgs boson searches by ATLAS and CMS using 2016+2017+2018 Run-2 data at $\sqrt{s} = 13$ TeV [1–6]

limits	$b\bar{b}\gamma\gamma$	$b\bar{b}b\bar{b}$	$b\bar{b}l\nu l\nu$
(ATLAS) μ	(ggF+VBF) 130(180)	(VBF) 840(550)	(ggF) 40(29)
limits	$b\bar{b}\gamma\gamma$	$b\bar{b}b\bar{b}$	$bbZZ$
(CMS) σ×BR [fb]	(VBF+ggF) 7.7(5.2)×SM	(VBF+ggF) 3.6(7.3)×SM	(ggF) 30(37)×SM

1.4.2   Resonant di-Higgs boson searches

For resonant searches, model-independent limits are set as function of resonance mass by both collaborations. The spin-0 and spin-2 narrow-width signal hypotheses with wide range of mass points are studied in resolved and boosted scenarios. For low masses $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$ channel are more sensitive while as we tend to higher masses $b\bar{b}b\bar{b}$ channel becomes sensitive. The 2016 Run-2 combination results are consistent with the SM background prediction; no significant excess is found across the entire mass range for either a spin-0 or a spin-2 resonance, as shown in Figs. 1.23 and 1.24.

Apart from spin-0 and spin-2 searches, 95% CL exclusion limits are also set for EWK-singlet and hMSSM BSM scenarios and large part of parameter space is excluded. For EWK-singlet model, Fig. 1.25a excludes the heavy scalar from 260 GeV to 700 GeV mass range at $\tan\beta = 1.0$ depending upon $\sin\alpha$. Fig. 1.25b excludes heavy pseudoscalar particle mass from 190 GeV to 550 GeV depending upon $\tan\beta$ within hMSSM scenario.

Using 2016+2017+2018 Run-2 data also, CMS and ATLAS have made some results public for different BSM resonant scenarios. Using same strategy as non-resonant vector boson fusion, ATLAS has performed spin-0 searches for narrow-width and broad-width approximations in $4b$ channel [3]. For $b\bar{b}\tau\tau$ channel, studies have been performed for boosted scenarios for mass > 1 TeV using a di-τ tagging technique considering it as a merged single final state object [24]. For $b\bar{b}\gamma\gamma$ channel, ATLAS recently made results public which set 0.9-0.1 fb upper limit on product of resonant production cross-section and branching fraction for resonance mass up to 1 TeV [6].

On the other side, CMS has extended search criteria up to NMSSM predictions. It has made NMSSM motivated model-independent results public using narrow-width approximation, in $b\bar{b}\tau\tau$ final state [25]. In Fig. 1.27, the maximum allowed ranges for $\sigma(pp \to H \to h_S(b\bar{b})h(\tau\tau))$ (H=heavy NMSSM Higgs boson, h=SM Higgs boson and h_S=light NMSSM Higgs boson) within the NMSSM can be constrained for m_H masses between 400 and 600 GeV. The strongest constraint is achieved for m_H=450 GeV and $m_{h_S} = [60, 80]$ GeV, where the maximum allowed value for $\sigma(pp \to H \to h_S(b\bar{b})h(\tau\tau))$ is reduced by a factor of five by this search.

(a) 95% CL limits on signal strength μ

(b) 95% CL limits on $\sigma(pp \to HH)$ as function of κ_λ
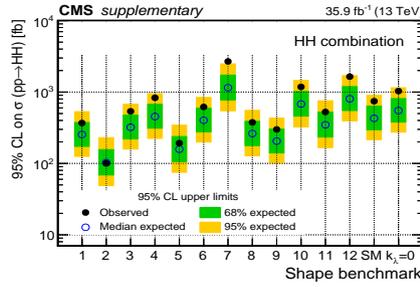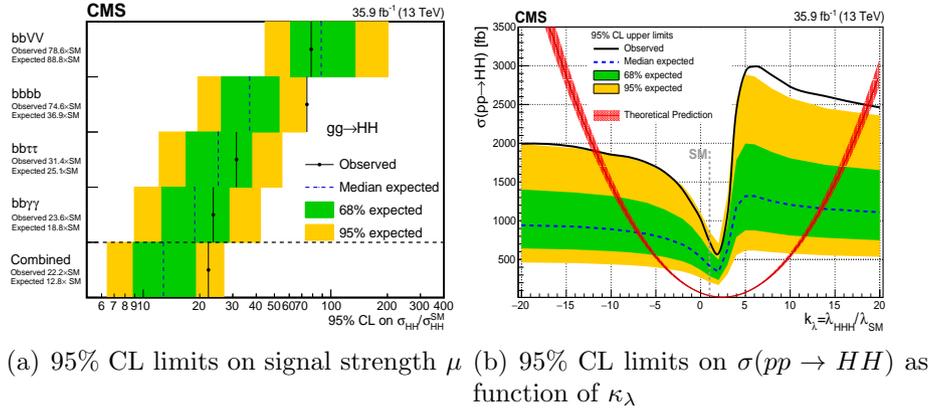
(c) The 95% CL exclusion limits on the non-resonant di-Higgs boson production cross section for different EFT benchmark topologies, for the SM production and for $\kappa_\lambda = 0$ production.

Figure 1.21: Non-resonant di-Higgs boson searches by CMS collaboration at $\sqrt{s} = 13$ TeV with 2016 Run-2 data [22]
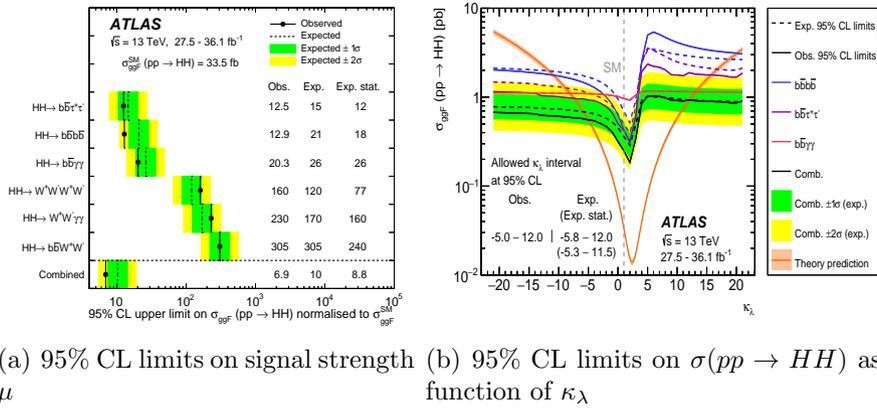
(a) 95% CL limits on signal strength μ

(b) 95% CL limits on $\sigma(pp \to HH)$ as function of κ_λ

Figure 1.22: Non-resonant di-Higgs boson searches by ATLAS collaboration at $\sqrt{s} = 13$ TeV with 2016 Run-2 data [23]
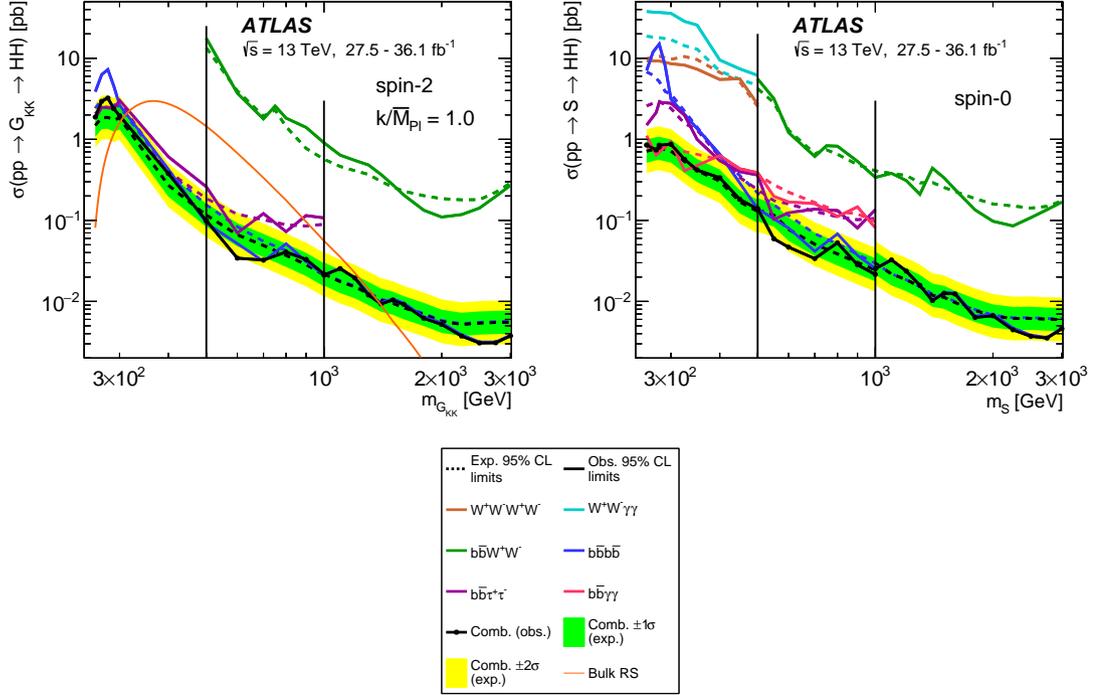
Figure 1.23: Resonant di-Higgs boson searches by ATLAS collaboration using 13 TeV 2016 Run-2 data; 95% CL exclusion limits as a function of mass for narrow-width spin-0 and spin-2 signal hypotheses; no deviation from SM expectations [23] has been observed.
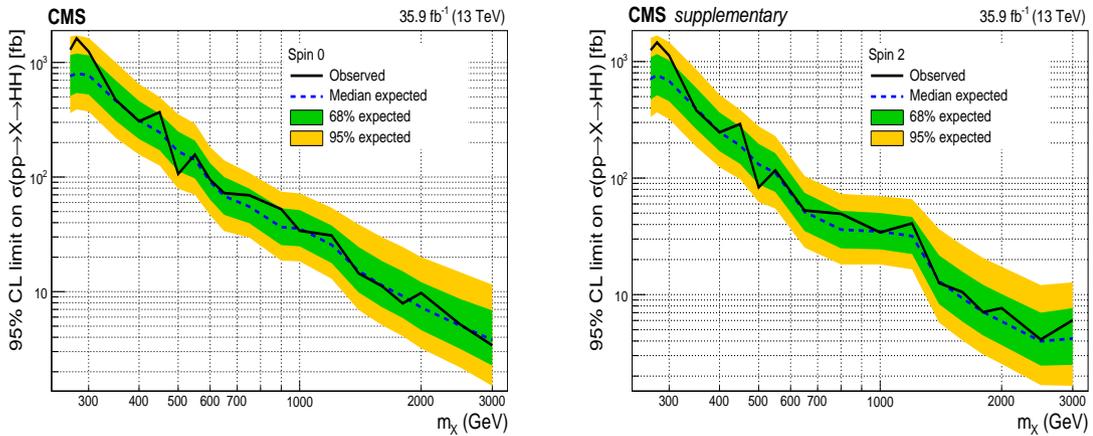
Figure 1.24: Resonant di-Higgs searches by CMS collaboration using 13 TeV 2016 Run-2 data; 95% CL exclusion limits as function of mass for narrow-width spin-0 and spin-2 signal hypotheses; no deviation from SM expectations [22]
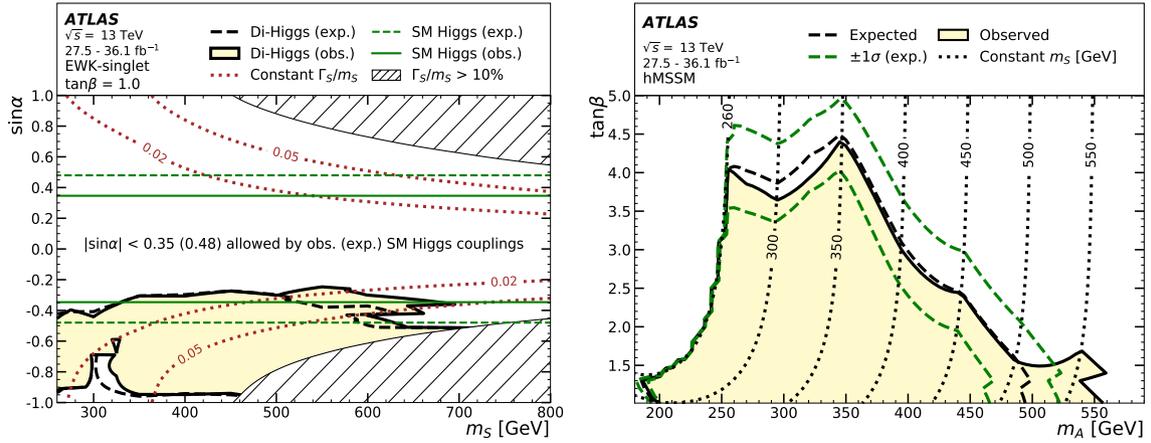
(a) Excluded regions for the EWK-singlet spin-0 resonance searches; the indirect constraints from SM Higgs boson coupling measurements are shown with vertical lines.

(b) Excluded regions for the hMSSM model using the experimental upper limits obtained in the spin-0 resonance searches
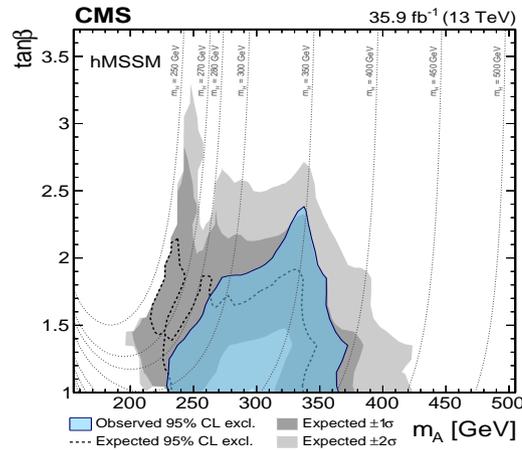
(c) exclusion limit, in the context of the hMSSM model, parametrized as a function of the $\tan\beta$ and m_A parameters; the CP-even lighter scalar (h) is assumed to be the observed 125 GeV Higgs boson, while H and A are the CP-even heavier scalar and CP-odd scalar respectively. The dotted lines indicate trajectories in the plane corresponding to equal values of m_H

Figure 1.25: Resonant di-Higgs boson interpretations for other BSM models [23, 24]
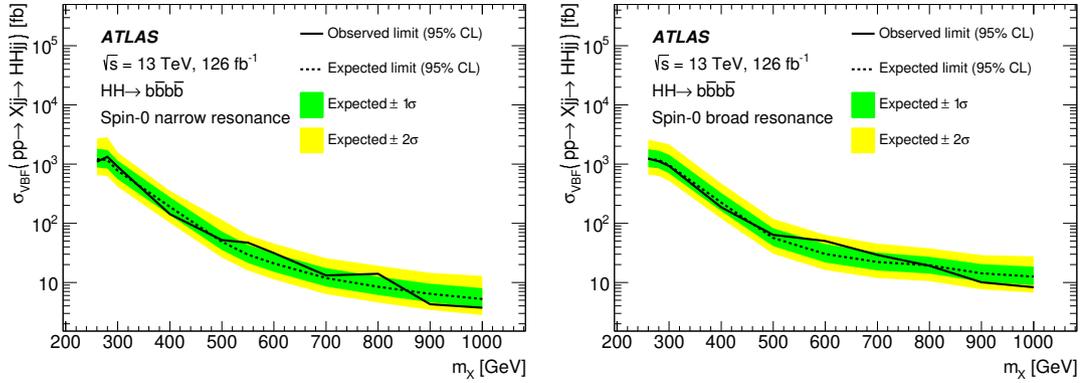
(a) VBF $X \to HH \to 4b$ narrow-width

(b) VBF $X \to HH \to 4b$ broad-width
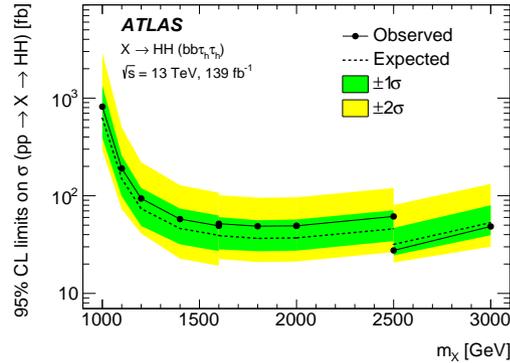
(c) ggF $X \to HH \to b\bar{b}\tau^{+}\tau^{-}$ narrow-width (boosted)

Figure 1.26: Resonant di-Higgs searches by ATLAS using 2016+2017+2018 Run-2 data at 13 TeV [3, 24]; Observed and expected 95% CL upper limits on the production cross section for resonant HH production.
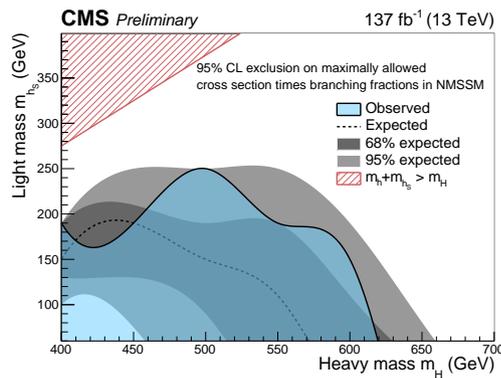
Figure 1.27: Resonant di-Higgs boson searches by CMS using 2016+2017+2018 Run-2 data at 13 TeV [25]; 95% CL exclusion limits on $\sigma(pp \to H \to h_S(b\bar{b})h(\tau\tau))$ for mass-range in m_H and m_{h_S}.

Chapter 2

The Large Hadron Collider (LHC)

2.1 What is LHC?

The LHC (Large Hadron Collider) machine is a particle accelerator at CERN that accelerates protons or ions close to the speed of light (c). In its name, "Large" refers to its size, "Hadron" refers to protons or ions getting accelerated within the ring, and "Collider" stands for a circular machine where collisions occur between accelerated hadronic particles. It is the most powerful collider in the world. It is a 27 km ring of superconducting magnets connected by a series of accelerating structures that raise the energy of the particles as they move along the ring. It consists of two concentric beamlines made up of quadrupole and superconducting dipole magnets. In the beam tube, it collides high-energy hadrons circulating along the concentric beamlines in opposite directions. LHC program started in 2009, and at a high instantaneous luminosity, proton-proton collisions have occurred at the center-of-mass energies of 7 TeV (2011), 8 TeV (2012), and 13 TeV (2015-2018).

2.1.1 Dipole magnet

As the LHC collides beams of particles with the same charge, the two beams must have opposite magnetic dipole fields to make them collide. Because of the 3.7 m internal diameter of the LHC tunnel, installing two completely separate proton rings is extremely

47

difficult, which led to the use of a twin-bore magnet configuration with two sets of coils. To bend the beam around a circular ring with a design center of mass energy of 14 TeV for proton-proton collisions, 1232 dipole magnets with a peak dipole field of 8.33 T are required. The superconducting dipole magnets of LHC use superfluid helium at 1.9 K to cool the magnet's Nb-Ti alloy coils to a superconducting state, allowing them to generate the required magnetic field. Figure 2.1 shows the cross-sectional view of LHC dipole magnet.

LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

ALIGNMENT TARGET

MAIN QUADRIPOLE BUS-BARS

HEAT EXCHANGER PIPE

SUPERINSULATION

SUPERCONDUCTING COILS

BEAM PIPE

VACUUM VESSEL

BEAM SCREEN

AUXILIARY BUS-BARS

SHRINKING CYLINDER / HE I-VESSEL

THERMAL SHIELD (55 to 75K)

NON-MAGNETIC COLLARS

IRON YOKE (COLD MASS, 1.9K)

DIPOLE BUS-BARS

SUPPORT POST
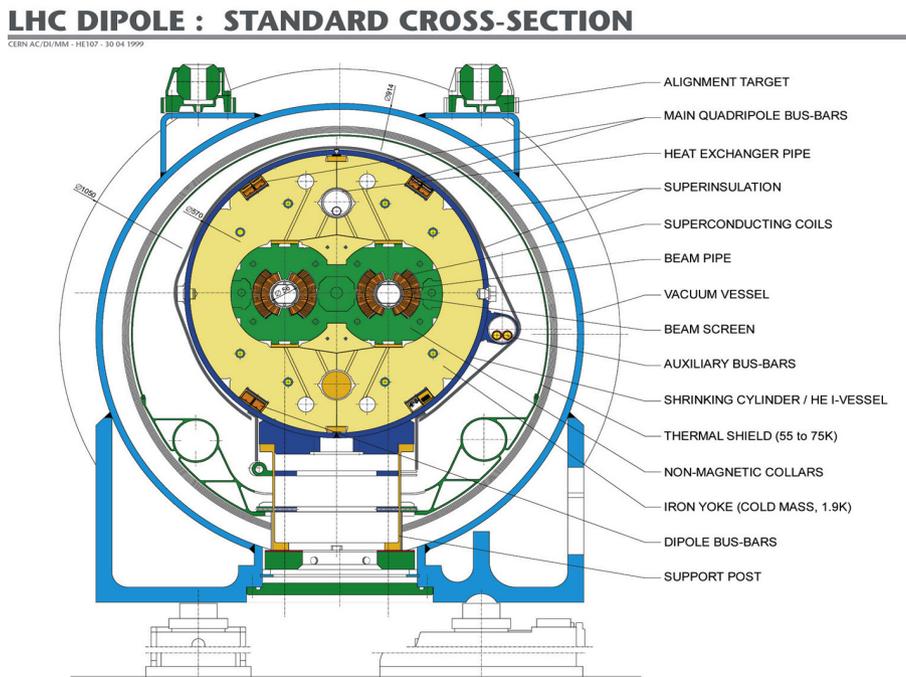
Figure 2.1: Cross-sectional view of LHC dipole magnet [26]

2.2   How does LHC work?

The CERN accelerator complex contains a chain of machines, as given in Fig. 2.2, that sequentially increase the energy of injected particle beam. Before the beam gets injected into the next machine, each machine accelerates the beam up to a certain energy limit. The next unit increases the energy of the beam to a higher level than the previous one, and so on. The LHC is the target element of this chain, where the beams are split

into two different directions. The beams collide at their highest energies (beam velocity becomes close to c) at LHC to produce interesting physics events.

The proton beam acceleration begins with a duoplasmatron source [99], which produces protons by ionizing the hydrogen atoms. Protons with 90 keV energies from the duoplasmatron source are sent to a radio frequency quadrupole (RFQ) cavity [100], where they are accelerated to 750 keV. The protons are then accelerated to 50 MeV using a linear accelerator (LINAC2) [101], after which they are injected into a booster ring. It boosts their energy to 1.4 GeV. They are then transported to the proton synchrotron (PS) [102], where their energy increases to 26 GeV. In the next step, beams are injected into the super proton synchrotron (SPS). Here, they are accelerated to 450 GeV before being injected into the LHC ring. In LHC, they are accelerated to their maximum energy (up to 7 TeV per beam). Eight RFQ cavities around the LHC ring oscillate at 400 MHz and use an electric field to accelerate each beam with a voltage of 2 MV per cavity.

There are two LHC run happened so far. They are known as Run-1 (7-8 TeV) and Run-2 (13 TeV). For the Run-2 period, the important figures of the LHC are tabulated in Table 2.1. With given 25 ns bunch spacing and 6.5 TeV beam energy the instantaneous luminosity ($\mathcal{L}$) of Run-2 is 10^{34} cm^{-2}s^{-1}. The event rate for a physical process with cross section σ is given by,

$$\frac{dN}{dt} = \mathcal{L}.\sigma \tag{2.1}$$

As a result, the number of events per unit time corresponding to that physical process for a given data collection duration is proportional to the "integrated luminosity" $L = \int dt \mathcal{L}$. Figure 2.3a shows the integrated luminosity delivered by LHC for Run-1 (2011-2012) and Run-2 (2015-2018) data-taking years. With a higher integrated luminosity, the probability of the rare physics events increases; therefore, it helps to study the rare physics signatures with more accuracy. However, during this bunch crossing, more than one proton-proton collisions occur, so whatever collision is picked as an interesting event, the contribution from other collisions is taken as a pileup effect. The pileup (PU) profile for Run-1 and Run-2 is given in Fig. 2.3b, which shows how PU (labeled with $< \mu >$) varies with luminosity. PU suppression is one of the main challenges for any analysis.

Various algorithms and techniques are used to mitigate PU. We will study this within the Chapter 3.
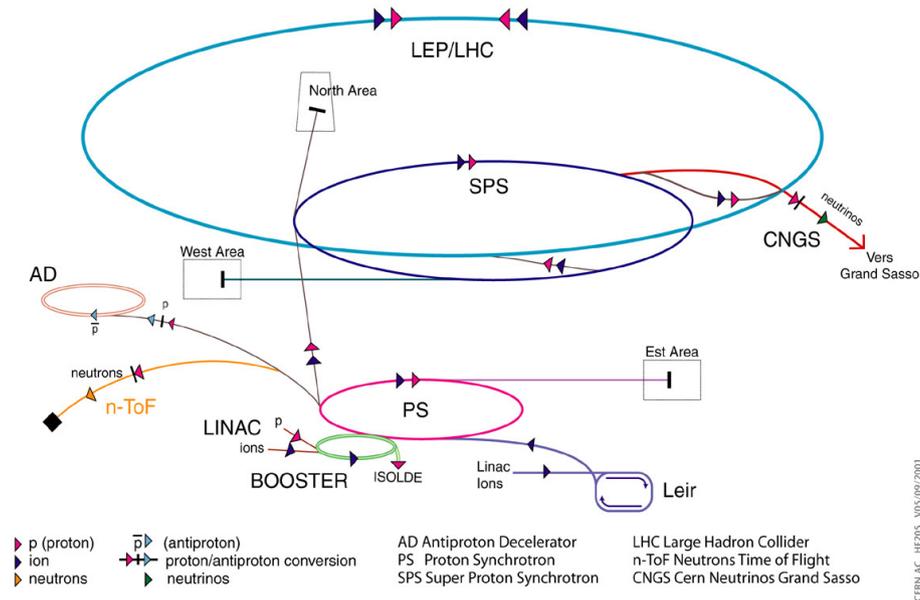
Figure 2.2: Accelerator chain at CERN [27]

2.3   Goals of LHC

The discovery of the Higgs boson at CERN was the first step of LHC towards achieving its goals. Still, there are many open challenges, as we studied in Chapter 1. Run-2 data at higher energy (almost double of Run-1 energy) and luminosity might help to study rare signals that could solve open questions of SM. The searches are performed in many sectors, e.g., Higgs physics, SM precision measurement, BSM signatures, B-physics, quark-gluon plasma study, to name a few. At CERN, various detectors are used to collect data for these searches. The construction of each detector is specific to its physics goal. We will have a quick overview of all the detectors in the next section.

(a) Integrated luminosity

(b) PU profile
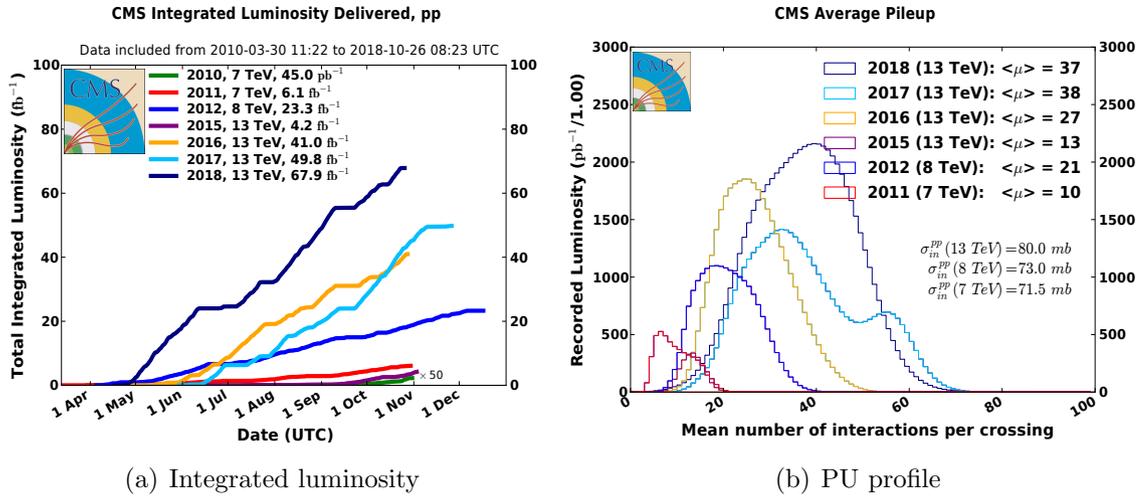
Figure 2.3: Integrated luminosity and Pile-up profile for Run-1 and Run-2 [28]

2.4   Experiments at LHC

At LHC, linear and circular accelerators make proton beams collide with high energy to produce a possible rare signal at a significant rate. At every collision point, a detector works like a camera to capture the collisions with high resolution. This high-resolution image information is obtained in different discrete pieces. It needs to join them using various algorithms and techniques since the detector is not ideally efficient to catch the signal. It combines background events with signal events. This makes it very tedious work to separate an interesting event from the background events. At LHC, we have four main experiments, known as ATLAS, CMS, ALICE, and LHCb. ATLAS and CMS are the general-purpose detectors where SM measurements and BSM-specific studies are performed using proton-proton collision data. ALICE detector uses proton-proton and heavy-ion collision data to study quark-gluon plasma, an early universe state just after the Big Bang. On the other side, LHCb concentrates on B-physics and CP-violation-specific studies. Figure 2.4 shows the position of these experiments within LHC ring. Our research work has used data collected by CMS detector. We will thoroughly study it in the Chapter 3.

Quantity	Number
Circumference	26,659 m
Dipole operating temperature	1.9 K (-271.3C)
Number of magnets	9593
Number of main dipoles	1232
Number of main quadrupoles	392
Number of RF cavities	8 per beam
Nominal energy, protons	6.5 TeV
Nominal energy, ions	2.56 TeV/u (energy per nucleon)
Nominal energy, protons collisions	13 TeV
No. of bunches per proton beam	2808
No. of protons per bunch (at start)	1.2×10^{11}
Number of turns per second	11245
Number of collisions per second	1 billion

Table 2.1: Important figures of LHC Run-2 program [7]
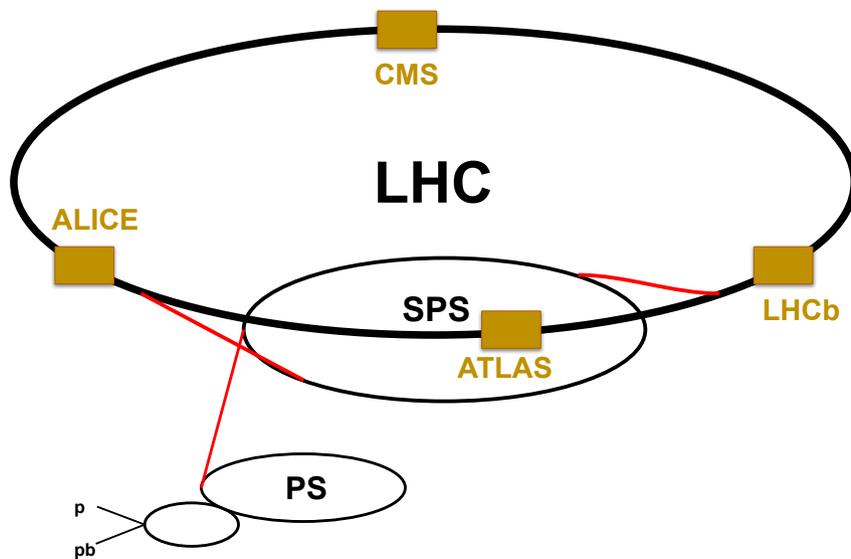
Figure 2.4: Major experiments at LHC (image reproduced from Ref. [29])

Chapter 3

Compact Muon Solenoid (CMS) Detector

One of the four major experiments at the LHC is the CMS detector [103]. It is a general-purpose detector exploring a wide range of physics searches. A number of collaborating institutes around the world have contributed for assembling the detector. During the data-taking, the LHC instantaneous luminosity is very challenging task for detector. With an instantaneous luminosity of about 2×10^{34} cm^{-2} s^{-1}, a peak pileup (PU) of almost 50 extra pp-interactions is expected during collisions. This additional activity raises detector occupancy while lowering detector performance.

In achieving the wide range physics goals, the main challenges for CMS are good electron and muon detection, high trigger-efficiency and offline tagging of leptons and jets associated with b quarks, good electromagnetic energy resolution and good dijet, diphoton and dielectron mass resolution, and identification of missing transverse energy. Apart from efficient readout electronics performance, by construction the detector is prone to radiation damage, particularly in the forward regions.

The CMS coordinate system is right-handed as given with Fig. 3.1. The origin of the CMS coordinate system is located within the detector, with the y-axis pointing vertically upward, the x-axis pointing radially inward to the middle of the LHC ring, and the z-axis pointing along the beam direction. The detector design can be understood using the

spherical coordinate system where r is the distance from the z-axis, ϕ is the azimuthal angle (measured from the x-axis in the x-y plane), and θ is the polar angle (measure from the y-axis in y-z plane). The rapidity is defined, for a particle having momentum 4-vector (E, px, py, pz), as ρ and in high energy limit rapidity converges to pseudorapidity (η) i.e. $(\rho \to \eta \equiv -ln(|tan(\theta/2)|) = 1/2\ ln((p+p_z)/(p-p_z)))$. The transverse momentum is defined as momentum projections on x-y plane and given as $p_T = \sqrt{p_x^2 + p_y^2}$. The usage of η and p_T at hadron colliders is motivated by the facts that the difference between pseudorapidity is a Lorentz-invariant quantity and beams enter along $\pm$z-axis within the detector; therefore, the transverse component for colliding particles is zero. So after the collision, all outgoing particles should have the sum of transverse momenta equal to zero by following the momentum conservation law.

The schematic diagrams of the CMS detector and its workflow is given in Fig. 3.3. The description of each CMS subdetector part is given in the next section.
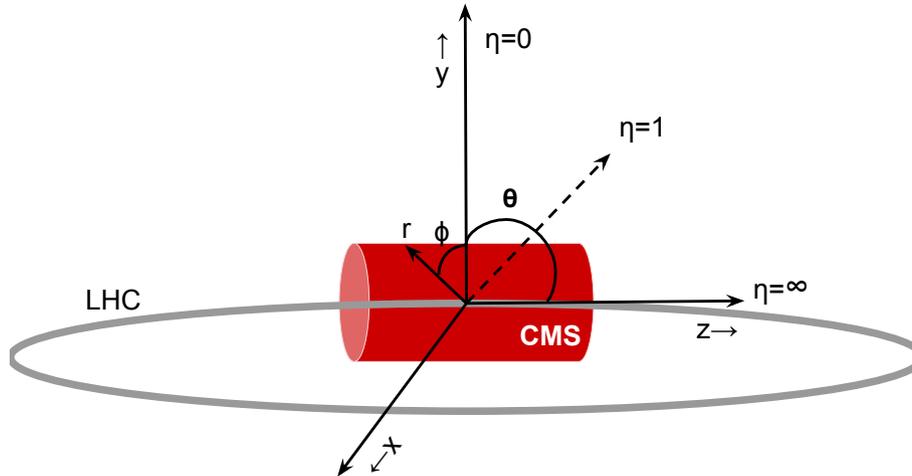
Figure 3.1: An illustration of the CMS detector with spherical co-ordinate system (image reproduced from Ref. [17])

3.1 Components of the CMS detector

The CMS detector collects data information using its subdetector components. The output information from each subdetector is further used to identify the final state. The overview of the CMS subdetectors (Fig. 3.2) is given by following subsections.

3.1.1 Superconducting magnet

A complex arrangement of niobium-titanium (Nb-Ti) coils, capable of carrying a current of 19.5 kA and cooled by liquid helium, works as a superconducting solenoid magnet and generates a 3.8 T magnetic field. This magnet has an inner diameter of 6 m and a length of 12.5 m. The tracker and calorimeters are entirely contained inside it. It is the main feature of the CMS detector, which bends the path of charged particles while passing through the magnetic field. The curvature of the path within the tracker depends on the energy and mass of the charged particle. It helps in particle identification and provides good momentum resolution. A 14 m iron return yoke surrounds the magnet coils and returns the magnetic flux before the muon chamber.

3.1.2 Tracker (TRK)

The CMS tracker [104] is mainly designed to precisely determine the trajectory of charged particles within the magnetic field and distinguish the primary and secondary vertices of an event. The tracker (used up to 2016 data-taking period with CMS phase-0 detector) can be divided into two subsystems; a pixel detector that ranges radially from 4.4 cm to 10.2 cm away from the beamline and a silicon strip detector covering the radii between 20 cm and 116 cm. The silicon modules cover a pseudorapidity range of $|\eta| < 2.5$ and have a total active area of more than 200 m^2.

The pixel detector is the closest part of the tracker to the collision point and divided into 66 M n+ pixels with a scale of 100 μm by 150 μm that are implanted into an n-type bulk with a thickness of 285 μm and a p-type backside. The barrel region of the detector has three layers with radii of 4.3 cm, 7.2 cm, and 11 cm, and two discs on either side of

the barrel (endcap regions) with radii of 34.5 cm and 46.5 cm from the interaction point. Each of the 15840 readout chips (ROC) in the pixel detector reads an array of 52 by 80 pixels. The ROCs are organized into modules, with 1312 readout links transmitting data. In 2017, pixel layers were increased up to four [105] during phase-1 upgrade of CMS detector. It increased tracker acceptance from pseudorapidity 2.4 to 2.5. The upgrade results in better momentum resolution and helps in all the track-based object identification algorithms explained in Section 3.5. The general design of pixel detector is similar to phase-0 pixel detector. The additional layer in phase-1 pixel detector uses chips based on column-drain readout architecture. The chips can handle hit rates up to 600 MHz/cm^2.

The silicon strip detector surrounds the pixel detector completing the tracking system. It is segmented into 9.6 M p+ strips implanted into n-type bulk with 320 μm (500 μm) thickness in the inner (outer) layers and n-type backside. The pitch of strips ranges 80–205 μm. The detector has ten tracking layers (4 inner and 6 outer layers) in the barrel area, covering radii of 25 cm to 110 cm and a z-axis of up to 120 cm. It also contains 12 disks (3 inner and 9 outer) in the endcap region with radii up to 110 cm and in z-axis up to 280 cm. Stereo modules are mounted in four layers in the barrel and multiple layers in the endcap regions, allowing for 2D measurement. Two silicon sensors are installed back-to-back in these modules, with their strips aligned at a 100 mrad relative angle.

Both the tracker subdetectors are read using a set of analogue electronic and optical links that can send absolute pulse height and pixel coordinates. All data processing for the strips takes place in off-detector electronics.

Given the LHC collision conditions, the main challenges for the tracker system are granularity, response time, and radiation hardness. It helps to detect the charged particles by constructing their trajectory and measuring their momentum.

3.1.3   Electromagnetic calorimeter (ECAL)

The CMS has a homogeneous electromagnetic calorimeter (ECAL) [30, 106] made of fine-grained 75,848 lead tungstate (PbWO$_4$) crystals, just outside of the tracker system.

The homogeneous medium minimizes the sampling fluctuations and provides a better energy resolution for photons and electrons. It is highly transparent and scintillates as electrons and photons pass through it. In other words, it generates light in proportion to the particle's energy. The crystals emit 80% of their light in less than 25 nanoseconds, which is the nominal time between successive bunch crossings at the LHC; this satisfies the requirement for quick detector response under LHC conditions. Since the light yield of $PbWO_4$ is temperature-dependent, a cooling system is needed to keep the crystals at ~18 degree Celsius. The photo-detectors are used to detect this scintillated light and covert it into an amplified electric signal.

The crystals are arranged in barrel region (EB), covering pseudorapidity up to $|\eta| =$ 1.48, and in two endcap regions (EE), covering $|\eta| =$3.0. EB has a crystal length of 230 mm (220 mm in EE) with 26 (25) radiation lengths. The crystals on the front face have a transverse dimension of 2.2×2.2 cm^2 in EB (2.86×2.86 cm^2 in EE). The total volume of the crystal is 11 m^3, and its weight is 92 t. The barrel calorimeter is divided into 36 supermodules, each with 1,700 crystals. The endcaps are divided into two dees, each with 3,662 crystals.

The photon separation is improved by a preshower detector (ES) based on lead absorber and silicon strips sensors (4,288 sensors, 137,216 strips, 1.9061 mm^2 with x-y view) mounted in front of the endcaps at 1.65< $|\eta|$ <2.6. The ES has a cumulative thickness of around three radiation lengths. It resolves the highly collimated photon-pairs coming from the light and short-lived π^0-meson decay, which is not possible to resolve using ECAL.

The ECAL energy reconstructions are crucial for the rare physics searches with final states having charged leptons and photons, e.g., $H \rightarrow \gamma\gamma$, $H \rightarrow ZZ \rightarrow 4l$ or many others.

3.1.4 Hadron calorimeter (HCAL)

Hadron calorimeter [30, 107] completes the CMS calorimetric systems. It is a sampling calorimeter which uses alternating layers of absorber and fluorescent scintillator

materials. HCAL determines a particle's location, energy, and arrival time as the particle passes through calorimeter. The scintillating light is collected by special optical fibres and fed into readout boxes, where photo-detectors amplify the signal. The total amount of light in a given area, known as tower, is a measure of a particle's energy which is summed up over several layers of scintillator tiles in depth. HCAL also consists of two parts: the barrel region (HB) and the endcap region (HE).

The sampling calorimeter is made up of the active material (4 mm thick plastic scintillator tiles) placed between copper absorber plates. The absorber plates are 5 cm thick in the HB region and 8 cm thick in the HE region. The active elements are read out using wavelength-shifting (WLS) plastic fibres. The depth of the barrel HB is around 79 cm or 5.15 nuclear interaction length.

The CMS also uses a separate forward calorimeter (HF) 6 m downstream of the HE endcaps. It extends the hermeticity of the central HCAL system to a pseudorapidity of 5.0 (as needed for an excellent missing transverse energy measurement). Quartz fibres are used as the active medium, and they are contained in a copper absorber matrix. It is specifically sensitive to Cerenkov light from neutral pions due to the quartz fibre active element. As a result, it has the unique and attractive property of providing a highly localized response to hadronic showers.

Along with measuring the energy of hadrons, HCAL also allows the detection of non-interacting and uncharged particles as missing transverse energy (MET). Measuring these particles is crucial because the measurement can reveal whether new particles have formed, such as the supersymmetric particles (much heavier versions of the standard particles). The decay products of these new particles leave no trace of their existence in any part of the CMS detector. To detect them, the HCAL must be hermetic, which means it must catch any particle that emerges from the collisions to the greatest extent possible. We can deduce the existence of the invisible particles if we see particles fly out one side of the detector but not on the other side, with an imbalance in momentum and energy. More about MET reconstruction has been explained in Section 3.4.2.

3.1.5   Muon chamber

As the name of the detector "Compact Muon Solenoid" indicates, muon physics is a vital task for CMS. Muons are the charged leptons similar to electrons but 200 times heavier than electron mass. Despite being a charged particle, it can penetrate the detector for several meters as it does not interact and deposit energy within calorimetric systems. Therefore, the outer part of the detector is entirely covered by muon chambers [30, 108] to detect muons, almost the only surviving particles till the range of the muon chambers. There are four muon stations outside that solenoid and interleaved with iron return yoke plates. They are used to reconstruct the hits made by muons while passing through them. Muon also leaves hits within the tracker. The strong solenoidal magnet field bends the muon track which helps in measuring muon's momentum. The hits within the tracker are combined with hits within the muon champers for high energetic muons.

The muon system contains gas ionization chambers. There are 1400 chambers in total. The 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) monitor the particles' positions and provide a trigger, and 610 resistive plate chambers (RPCs) form a redundant trigger network that quickly determines whether or not to hold the muon data acquired. All these components are robust and capable of suppressing background noise. The muon barrel (MB) region contains RPCs and DTs, while the endcap contains RPCs and CSCs. The arrangements depend on the muon rate in MB (low rate) and ME (high rate) region.

The DTs cover the pseudorapidity region $|\eta|$ <1.2. They can reconstruct the muon track from its hits within the stations with excellent time resolution and efficiency. A gas mixture of 85% Ar+15% CO_2 is surrounded by a gold-plated stainless-steel anode wire in each cell, resulting in a drift time of 380 ns.

The CSCs cover a pseudorapidity region 0.9< $|\eta|$ <2.4. In each endcap, the 468 trapezoidal CSCs are arranged into four stations. Six anode planes are interleaved among seven cathode panels in each chamber, with wires running azimuthally. The ME chambers uses a gas admixture of 50% CO_2 + 40% Ar and 10% CF_4.

The RPCs are interspersed in both the MB and ME covering $|\eta|$ <1.9 region. The

RPCs are made of two resistive Bakelite plates separated by a gas volume. They provide an independent triggering system and a fast response with good time resolution (less than 25 ns) for muons.
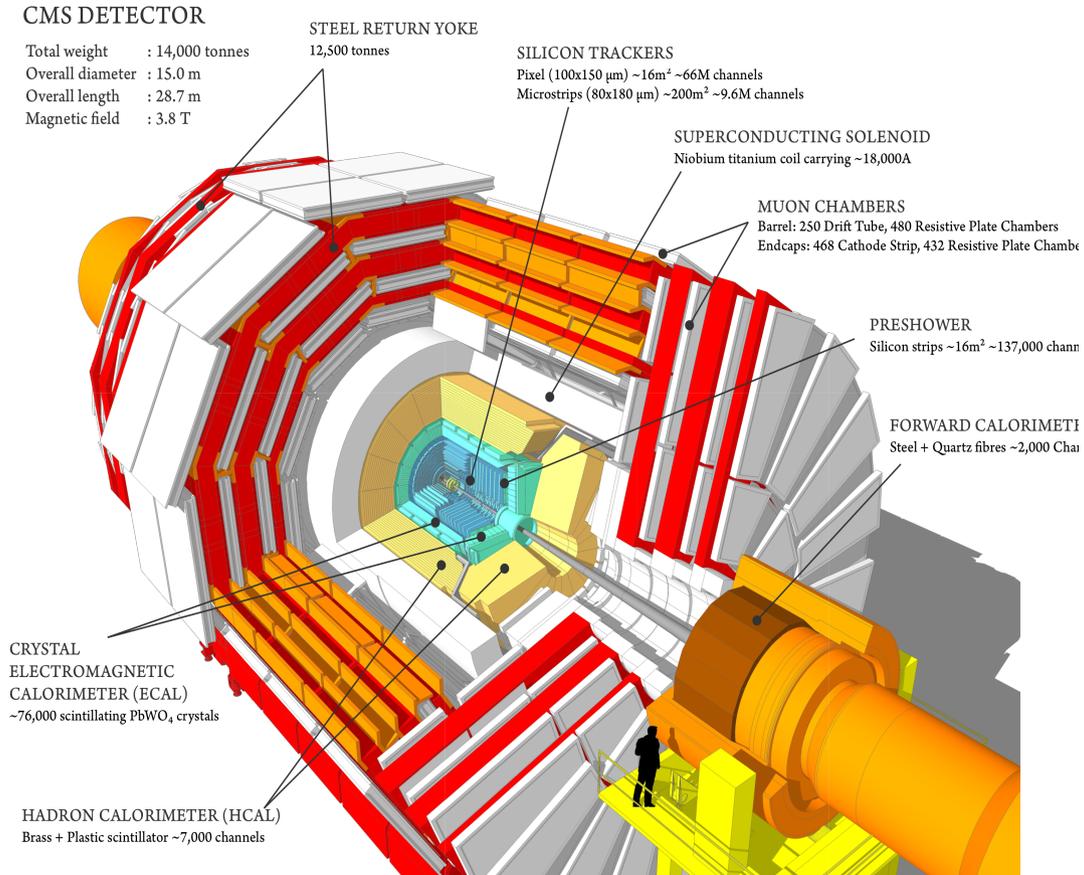
Figure 3.2: Components of the CMS detector [30]

3.2 Collecting and collating the data

The data collection is a very important step of the CMS detector. The amount of the collision data is too huge to store everything. That is why identifying only interesting events and storing them is very crucial. In this section, the workflow of the CMS detector is explained. After following it, the data are used for any analysis.

3.2.1 Pattern recognition
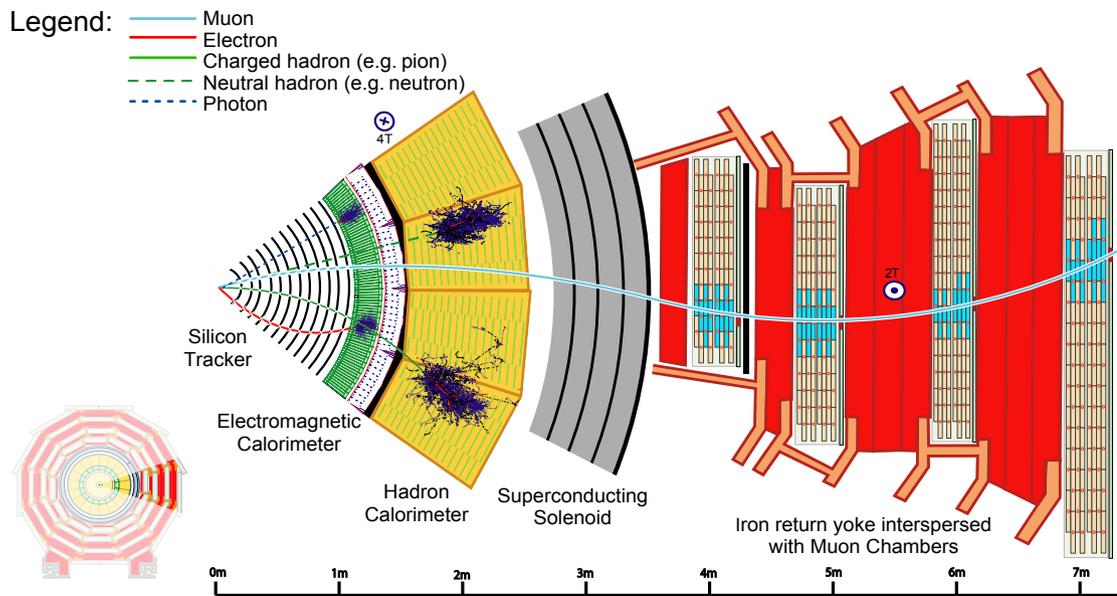
Figure 3.3: Pattern recognition within the CMS detector [30]

The main virtue of any new physics signal event is that its final decay products carry different kinematics properties. These properties are used to identify the event and to decide if it has something interesting to analyze. However, there is always a probability for background events to mimic the signal. This is something we have to deal with while doing any analysis.

Every different particle leaves a different pattern within the subdetector depending upon the interactions with its components. Based on this pattern, the objects are

recognized and used at the trigger level to decide if they should be discarded or not. Also, to avoid mixing of particles from two events, detectors must have a high time resolution, and signals from millions of electronic channels must be synchronized so that they can all be detected as coming from the same event.

The main final state objects are photons (γ), electrons (e), muons (μ) and hadrons. In Fig. 3.3, the pattern of interactions of each object is shown. All the charged particles leave (e, μ, charged hadrons) the curved tracks within the tracker while passing through the strong magnetic field. They further deposit the energy within the calorimetric systems, while neutral objects (γ, neutral hadrons) only deposit energy within the calorimetric systems. Within the ECAL, the electrons and photons deposit the energy, but hadrons hardly deposit any energy. Hadrons only interact within HCAL and lose all their energy.

3.2.2   Trigger system

At LHC, the proton-proton collision occurs at very high luminosity, which leads to producing a rare physics signal at a good rate. However, most of the collisions are soft (low energy), so they do not produce any interesting physics event. Also, The size of each event is around 1 MB, and the frequency of collision is 40 MHz, i.e., 40 TB data per second get generated during the collision. Considering the fact that in this huge data collection, only a few events are of physics

Figure 3.4: Flowchart of CMS trigger system (image reproduced from Ref. [31])

interest, a trigger system is used to select potentially interesting events. Only this

fraction of data is stored on a computer disk for subsequent analysis. The full trigger system decreases the rate of interesting events to 1k per second. A series of trigger levels are used to achieve this. The detector stores all of the data from each crossing in buffers. A small amount of key data is used to perform a fast, approximate calculation to identify features of interest such as high-energy jets, muons, or missing energy. The levels are known as "L1-trigger" or level-1 trigger and "HLT" or high level trigger as given in Fig. 3.4.

3.2.2.1 L1-trigger

The L1-trigger is based on hardware. It uses a rapid and completely automated method that scans the basic signs of interesting physics, such as particles with high energy or rare combinations. From the billion events, 100k events are selected at this level with a latency of few microseconds using a simplified readout of the calorimeters and muon subdetectors.

A simple schematic for CMS L1-trigger is given in Fig. 3.5. The trigger primitives (TP) from ECAL and HCAL as well as muon detectors (drift tubes (DT), cathode strip chambers (CSC), and resistive-plate chambers (RPC)) are processed in several steps until the combined event information is evaluated in the global trigger (GT). After this, a decision is made to accept the event. A regional calorimeter trigger (RCT) and a global calorimeter trigger (GCT) make up the L1 global calorimeter trigger (GCT). The RCT receives the transverse energies. The RCT processes this data in parallel and sends objects and their energy information as outputs. The GCT sorts the objects using their energy information and classified them as isolated, non-isolated, central, forward jets, and several global quantities (E_T^{miss}).

To ensure good coverage and redundancy, each of the three muon detector systems participates in the L1 muon trigger. The front-end trigger electronics of DTs and CSCs identify tracks (hits) and transmit them in regional track finders. They further identify muons based on pattern recognition algorithms and measure their energy. In the overlap region of the DT track finder and CSC track finder, the information is shared for efficient

coverage. For RPC hits, the information is sent to pattern comparator trigger logic boards via front-end electronics that identify muon candidates. The three regional track finders sort the muon candidates that have been detected and send them to the global muon trigger (GMT) with their p_T and position information. The GMT then combines muon candidates identified by multiple systems to exclude candidates that pass multiple muon triggers. The GMT also conducts a consistency assignment so that candidates can be rejected at the final trigger stage if their quality is poor and they can only be reconstructed by one muon track finder.

The global trigger (GT) completes the CMS L1 trigger scheme by implementing a menu of triggers, a set of up to 128 selection requirements applied to the final list of objects required by the HLT algorithms to satisfy the physics data-taking objectives.

3.2.2.2   High-level trigger (HLT)

The HLT is made of software farm that runs on scientific Linux systems. The event selection process at the HLT is as similar as it is done in offline processing. It reduces the event rate down to 100 Hz. Objects such as electrons, muons, and jets are reconstructed for each event, and identification criteria are used to select only interesting events for
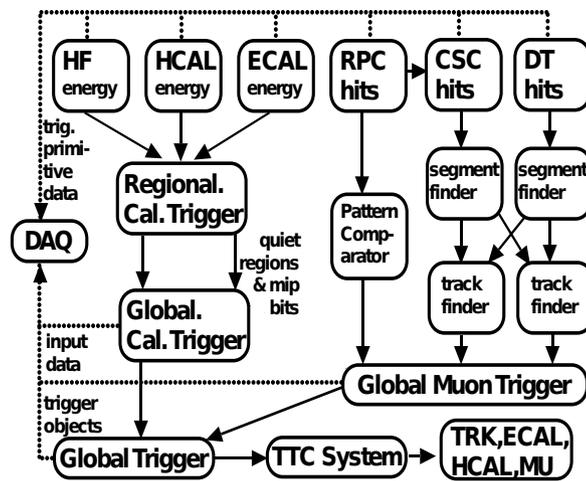
Figure 3.5: Overview of the CMS L1 trigger system. The information from the GCT and GMT is combined in a global trigger (GT), which makes the final trigger decision. This decision is sent to the tracker (TRK), ECAL, HCAL or muon systems (MU) via the trigger, timing and control (TTC) system. The data acquisition system (DAQ) reads data from various subsystems for offline storage after HLT decision. MIP stands for minimum-ionizing particle. [32]

data analysis. After the HLT
level selection, the data is stored
on tape for the analyses.

In HLT operation, the data from the readout buffers are sent to a processor farm
with 16000 CPUs. This trigger level is made up of a series of increasingly complex filters.
The filtering process uses the complete detector information from all the subdetectors,
starting from reconstruction to selection. In simple words, the HLT considers full data
events to decide if the event should be kept or not. In order to create datasets with
different physics signatures, the final stage of HLT processing involves the reconstruction
and event filtering. The time duration it takes to process an event varies depending on
the algorithms used. The average time between events is about 60 ms, but some events
can take up to a second.

3.2.3 Data analysis

Data events that pass the triggering stages and have been taped are duplicated using the
Grid to additional storage sites worldwide for easy access and redundancy. Even after
trigger selection, CMS produces huge amount of data that is handled by a distributed
grid computing and data storage infrastructure. The computing centres available to CMS
around the world are distributed and configured in a tiered architecture that functions
as a single coherent system. Each of the three tier levels provides different resources
and services. The T0-tier accepts RAW data from the CMS Trigger System and repack
them to tape and send their copy to all tier T1. Prompt calibration and reconstruction
is also performed at T0 and saved in RECO format. RECO is also copied to T1. From
RECO to AOD/miniAOD format, only skimmed information of final state objects is
saved. From T1, only miniAOD format is copied to T2 which is accessible to perform
any analysis. Physicists may then use the Grid to access the data from T2 and run their
analyses.

3.3   CMS simulation tools

The simulation plays a fundamental role in data analysis to perform any measurement or extract any relevant physics parameter. It consists of complete information on the physics process used for event generation and corresponding particle content. For event generation, Monte-Carlo (MC) based event generators are used. They use numerical MC based techniques to produce collisions at the high energy as they occur in detectors. The MC event generators provide a complete picture of the collision process from initial to final stages, including the strongly inelastic interaction, the radiation process, parton-hadronization, and the underlying event description. After the physical events are produced using information from theoretical models incorporated into the generator, the effects that a detector introduces into the basic theory must be considered. The choice of event generator depends on the physics process.

In simulation each event has a weight that corresponds to its differential cross section. Having equal weights for every event in simulation produces natural distributions as observed in collision. While generating events at NLO order, the infra-red (IR) divergences in the real-emission corrections and virtual corrections are taken into account using IR subtraction method. The contribution of the soft singularities makes the matrix element of real-emission finite. Therefore, the method adds a subtraction term to make virtual corrections finite. The events are generated separately for Born and real-emission phase space. For the events simulating real-emission, weights become negative if simulation over-estimates real-emission matrix element. The negative weights contributes as the negative differential cross section term for real-emission. This results in reduction of effective simulation statistics. Generally the fraction of events with negative weights is small, but still problematic while doing higher order calculation that requires more computational sources per event. Therefore, the event generators that take care of the negative weights issue are used for NLO order event generation [109, 110]. CMS physics simulations are mostly based on the following event generators:

MadGraph: MADGRAPH [111] generates the matrix element for the multi-particle final state process. The updated version allows for matrix element calculations at NLO

accuracy (controls theoretical uncertainty) and provides a technique for parton shower matching. Depending upon the Lagrangian of any renormalizable or effective theory, it can generate events for any physics process predicted by that theory.

PowHeg: PowHeg [112, 113] also produces NLO-accurate calculations of the hard scattering subprocess. PowHeg uses a condition in cross section calculation that do not consider events with negative weights in calculation. The Refs. [110, 114, 115] have detailed description of PowHeg NLO calculations.

Pythia: The Pythia [116] is a standard tool for the generation of events in high-energy collisions, which works for various SM and BSM processes. It contains the models of hard processes and initial- and final-state parton showers, matching and merging methods between hard processes and parton showers, multiparton interactions, beam remnants, string fragmentation, and particle decays.

The last step of the simulation process is to pass the generated events through the detector response simulation. In CMS, it is done using Geant4 [117] toolkit. The detector simulation includes the detector geometry, particle interactions with detector's materials, magnetic field effects, real conditions during detector operation, and electronic readout. The additional pp-interactions (pileup) during collision are also super imposed to the event at hit level using pileup mixing module. The module deals with a sequence of bunch crossings to properly simulate the contributions affecting the in-time bunch crossing. The next step in the event is modeling the response of the detector readout electronics, which is also known as Digitization. The digitized signal is further reconstructed and use for physics analysis after skimming steps.

Most of the aspects of simulations are integrated with standard analysis software known as CMSSW [118]. The MC events are reconstructed using the same methods used for real data, allowing for a consistent comparison between the data and the simulation. We will briefly study these reconstruction algorithms in the next section.

3.4   Object   reconstruction   and   identification algorithms

The event reconstruction within CMS is performed using two steps. The first step is to reconstruct each object with a particle flow algorithm, whether they are leptons, jets or missing energy.

After the collision, particles enter the tracker after leaving the beam interaction point, where signals (hits) in the sensitive tracker layers are used to reconstruct charged-particle trajectories (tracks) and origins (vertices). The tracker is embedded in a magnetic field, which bends the trajectories and allows for measuring charged particle electric charge and momentum. ECAL absorbs electrons and photons as they pass through it. The corresponding electromagnetic showers are observed as energy clusters in neighbouring cells, which are used to calculate the particle's energy and direction. Hadrons, both charged and neutral, may induce a hadronic shower in the ECAL, which is then fully absorbed in the hadron calorimeter (HCAL). The energies and directions are calculated using corresponding clusters. Muons and neutrinos pass through the calorimeters with very few interactions. Muons generate hits in the muon detectors, located outside the calorimeters, while neutrinos escape undetected.

All the above information from subdetectors are basis of CMS reconstruction algorithms.

3.4.1   Tracks and Primary Vertices (PV) reconstruction

For any data analysis, it is essential to understand how the tracks and origin of every physics process (also known as primary vertex) can be identified dealing with the large PU interactions. A detailed procedure of CMS tracks and vertex reconstruction is given in Ref. [119].

Tracks are essential for determining the production vertex of charged particles and measuring their momenta. For their reconstruction, firstly, the hits within the pixel and strip detectors are determined. They estimate the momentum and position parameters

(longitudinal and transverse parameters) of the particles. For this estimation, tracking software called combinatorial track finder (CTF) is used.

The CTF reconstruction method computes six iterations in total. Tracks with p_T >0.8 GeV and three pixel hits originating near the nominal interaction point (beam spot) are chosen in the first iteration. Tracks with just two-pixel hits are chosen in the next iteration. The subsequent iterations are used to find tracks that originate far away from the beam location and new tracks that were not discovered previously. It should be stated that at the start of each iteration, the previous iteration's hits are no longer taken into account.

After these iterations, track fitting is performed to determine trajectories of the particles and smoothen the track path. After fitting, the final track is selected based on the quality of the fit by checking χ^2 of the track fit, the number of hits and missing hits in the track.

Primary vertex reconstruction depends on the selected tracks where it finds a common meeting point (vertex) among a set of tracks. It aims to determine the position and associated uncertainty of all proton-proton interaction vertices, including the signal vertex and any vertices from pileup collisions. It consists of three steps:

(1) Tracks selection

(2) Clustering of tracks originated from the same interaction vertex

(3) Track fitting for each vertex's location using its corresponding tracks

Among the fitted vertices, one with the largest square-sum of the transverse momenta of all tracks is considered the hardest scattering interactions or the "primary vertex". Other vertices are treated as PU vertices. The vertex-finding algorithms differ depending on the physics study.

3.4.2   Particle Flow (PF) algorithm

The PF algorithm aims to classify and reconstruct all of the particles from a collision by integrating the information from the various subdetectors in the most optimal way. For each collision, the set of the identified and reconstructed particles (PF candidates)

by the algorithm provides a global event description that leads to phenomenal CMS performance for jet and τ hadronic decay reconstruction, MET determination, and e and μ identification. This method also identifies particles from PU interactions; therefore, it is used to build efficient PU mitigation techniques.

Any physics study is conducted purely based on final decay products, including the features of the physics process. In this part, we study how the PF algorithm works for final observable states that leave signals in the detector.

- **Electron (e) and Photons (γ):** Nearly all of the energy of electrons and photons is deposited in the ECAL, where electrons also create hits in the tracker layers. Extrapolation from the last measured hit in the tracker to any cluster in the ECAL is used to link them. The signals in the ECAL crystals are reconstructed by subtracting the PU contributions. This technique has been used for both the HLT and offline event reconstruction during the entire LHC Run-2 data-taking period.

  While travelling through the tracker material in front of the ECAL, an electron can emit bremsstrahlung photons, and a photon can convert to an electron-positron pair, producing the shower in the tracker. A dedicated algorithm is used to combine the clusters from the individual particles into a single object that recovers the energy of the primary electron or photon. For energy reconstruction, first, energy is clustered around the group of ECAL crystal having one of the highest energy deposited in any specific region, with a minimum transverse energy of 1 GeV. After clustering, to include bremsstrahlung and photon conversions, losses are included and called superclusters. The curvature of the tracker is also impacted by the trajectory of an electron losing momentum by emitting bremsstrahlung photons. All reconstructed tracks are tested for compatibility with an electron trajectory. A dedicated algorithm is used to select generic tracks originated from the photon conversion process in the tracker. All these ECAL clusters, superclusters, electrons tracks, tracks from photon conversion are used as input to link this element into a block of particles. Starting from electron tracks or superclusters, respectively, the blocks are divided into electron and photon. At this point the supercluster is

called a refined supercluster. A further track selection criteria are applied to these object to reconstruct "PF electron". Without passing this selection, the particle is labelled as "PF photons".

- **Muons (μ):** Muons are reconstructed using information from the tracker and the muon systems in CMS [120]. The reconstruction is performed with three following methods:

  - The local reconstruction is the first step in the muon reconstruction chain. First, digitized electronic signals are used to recreate hits in DTs, CSCs, and RPCs. Hits are then matched within each DT and CSC chamber to form segments (track stubs) using the Kalman filter method. The reconstructed muons are labelled as "standalone muons".

  - A search is performed for tracks that fit each standalone muon track among those reconstructed in the inner tracking system, with the best-matching tracker track being chosen. The track fitting, using all hits in both tracks, is performed for each tracker track-standalone muon pair, again using the Kalman filter technique. The result is a collection of objects referred to as "global muons".

  - Tracker muon tracks are built from the inner tracker trajectory reconstruction. The tracker-muon algorithm is beneficial for identifying low-p_T muons that may not leave enough hits in the muon stations to be reconstructed as a standalone muon. Tracker muons should not be used without additional specifications because the default conditions for tagging a tracker track as "tracker muon" are pretty loose.

The resulting muon candidate collections are used as input for PF muon identification. The PF muon reconstruction has been fine-tuned to distinguish muons within jets with high accuracy, resulting in a low rate of false positives due to misidentified charged hadrons.

- **Jets and Missing Transverse Energy (MET):** As the collisions occur at the high energy the process may end up with having partons in the final state carrying colour charge. Since they can not exist in a free state, they hadronize to produce stable colourless hadrons as a result of QCD confinement. The produced hadrons appear to move in the same direction as its origin parton, creating collimated bunches of particles known as jets. After the identification of muons, electrons, and isolated photons and their extraction from the PF blocks, the remaining particles to be detected are hadrons originating from jet fragmentation and hadronization. The jets are reconstructed by placing their components together using various algorithm that follows collinear and infrared safety principles [121]. A final state parton from hard-scattering and hadrons can have multiple collinear splitting or soft emissions which create infrared collinear (IRC) divergences. In theory, these divergences get cancel out with one order loop correction. But this cancellation does not happen within jet algorithm. If algorithm is not IRC safe, it will result in unrealistic infinite cross-section. Also, jet defined by any algorithm should be invariant of choice of algorithm.

  The anti-k_T algorithm and Cambridge/Aachen (C/A) are the two of them [122]. The C/A algorithm is based on sequential recombination. It combines the closest particles with a condition on the distance of two particles. The clustering process uses distances d_{ij} between particles i and j and d_{iB} between particle i and the beam (B). The clustering proceeds by identifying the smallest of the distances and if it is a d_{ij} recombining particles i and j, while if it is d_{iB} calling i a jet and removing it from the list of particles. The distances are recalculated and the procedure repeated until no particles are left. The general definition of the distance measures is given as:

  $d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2}$ and $d_{iB} = k_{ti}^{2p}$

  where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. The k_{ti}, y_i and ϕ_i represent the transverse momentum, rapidity and azimuthal angle of the particle i, respectively. The R represent the jet clustering distance parameter. The C/A algorithm uses $p = 0$ in

d_{ij} and d_{iB} for jet clustering.

On the other hand, anti-k_T can be considered the generalized version of the C/A algorithm. It uses $p = -1$ for distance measure. Therefore, distance measure is taken as $d_{ij} = \min(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2})\frac{\Delta_{ij}^2}{R^2}$ for jet clustering. The algorithm allows clustering of the hard particles. First, it identifies the highest energetic particle and grows a jet around that particle. Since the algorithm involves a combination of energy and angle in its distance measure, hence it becomes IRC-safe growth. The achievement of anti-k_T is that it gives circular-shaped jets without using a cone-based jet algorithm. The anti-k_T algorithm is widely used to define a set of collimated particles as a PF jet. The spread area of the jet is given in terms of jet distance parameter R. The value of R depends on the collision energy. For the boosted scenario where hadrons get produced at very high energy, it is hard to separate the particles from two different hadronic processes. Therefore, the bunches of particles from the hadrons are merged and reconstructed using a large R in the algorithm. They are called fat or large jet.

At 13 TeV, CMS generally uses AK4 PF jets, i.e., anti-k_T jets clustered with distance parameter $R = 0.4$. For large area jets, the value of R is taken 0.8, and they are called AK8 PF jets. Despite the excellent functionality of the jet clustering algorithm, we get discrepancies between the energy of partons and the reconstructed energy of jets due to the efficiency and acceptance of the detector. That is why jet energy calibration (JEC) and jet energy resolution (JER) are applied on the reconstructed jets including uncertainty associated with them [123].

The MET is identified as momentum imbalance in the transverse direction and defined as the negative vector/scalar sum of the transverse momentum of the reconstructed PF objects in the event [124, 125]. It originates from weakly interacting neutrinos or any BSM particle that hardly leaves any sign within the detector. Since W bosons, top quarks, and tau leptons may decay into neutrinos, CMS uses MET to reconstruct them. Furthermore, several BSM physics models, such as dark matter models, supersymmetric models, and models with warped

extra dimensions, predict the presence of particles that might be invisible and can carry momentum. Accurate MET reconstruction is complex because it requires the precise reconstruction of all visible particles in an event. The CMS detector meets these requirements with its highly granular electromagnetic calorimeters, hermetic hadronic calorimeters, redundant muon systems, and silicon trackers in a strong magnetic field.

There are two types of PF Jets used in CMS analyses. For PF CHS (Charged Hadron Subtraction) jets, charged particles from non-primary vertices (pileup) are removed before clustering. Another one is PF PUPPI (Pile-Up Per Particle Identification) jets which use the PUPPI [126] algorithm. It is explained in detail within Section 3.5.6. Apart from PF jets and PF MET, Calo jets/MET and TRK jets/met is also used for various studies. Calo objects are reconstructed using the energies in calorimeter towers, and their direction and TRK objects are reconstructed using hits information from the inner tracker.

3.5   Other object specific algorithms

Once we have objects reconstructed and identified with the PF algorithm, specific algorithms are applied depending upon the analysis final state to discriminate the object contribution coming from similar background final states. These algorithms are based on the energy of the physics process, type of final state and how stable decay product behave within the detector.

For example, a jet can be originated from a light quark or a heavy quark. In both conditions, it leaves different signs in the detector. Similarly, a new physics originated photon or lepton will have some features to build a discriminator to remove the fake object contamination.

We have worked in the boosted scenarios for HL-LHC projection studies. The boosted Higgs jet tagging (H tagging using n-subjettiness) with subjet (jet within a large area AK8 jet) b tagging for $H \rightarrow b\bar{b}$ handle is used. PUPPI and jet grooming algorithms

remove the soft and pileup contributions.

For the project with real data in Chapter 5, we have b jets and photons in the final state. To identify and discriminate them from background objects, neural network-based b tagging discriminant and multivariate method (MVA) based photonID have been used. The following subsections explain only those algorithms which are specifically used within the thesis Chapters 4 and 5:

3.5.1   Jet b tagging

As the name reflects, b tagging aims to identify a heavy quark jet from a light quark jet. Identification of the origin of jets is critical for studying and characterizing different channels, such as top quark/Higgs boson events and a variety of new physics scenarios. The long lifetime, high mass, and large momentum fraction of b hadrons and the existence of soft leptons from semileptonic b hadron decays are used to develop CMS b tagging algorithms.
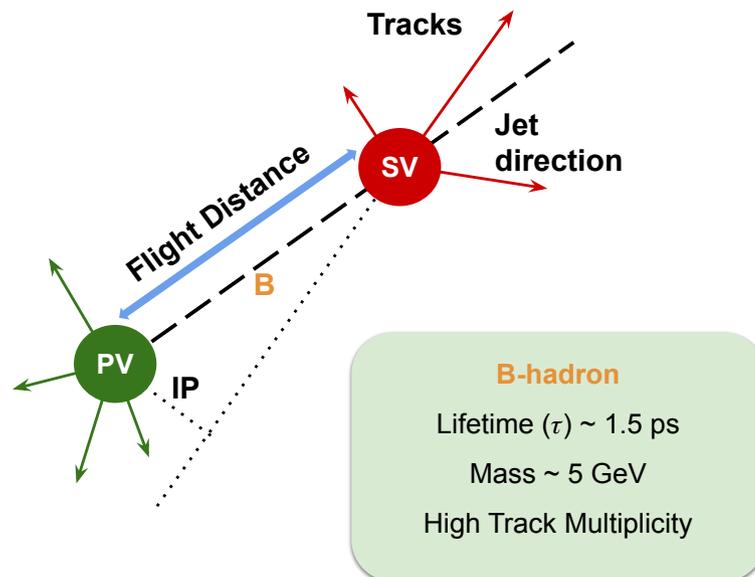
Figure 3.6: An illustration of B-hadron decay and corresponding Impact parameter (IP) (image motivated from Ref. [33])

Semileptonic decays of b hadrons give rise to b jets that contain a muon. A cascade decay of $b \rightarrow c \rightarrow l$ also gives muon in the final state. Since the CMS muon systems identify the origin of muons with high efficiency and resolution, this information helps to tag the b jets.

Due to long lifetime, b hadron travels within tracker, and the point, it decays, is called the secondary vertex (SV) shown in Fig. 3.6. The distance between PV and SV is called flight distance. With the high resolution of the CMS tracking system, it is possible to reconstruct the SV. In the SV vertex finding process, the tracks associated with PV are not considered.

The distance of the closest approach of a track to the primary vertex is known as impact parameter (IP) given in Fig. 3.6. IP's sign is positive/negative if the track passes closest to its associated jet direction down/upstream of the PV. It is calculated in three-dimension, which benefit from the good x-y-z resolution from the pixel detector. IP is a Lorentz invariant parameter. Due to the long lifetime, the IP from b jets is mostly positive, while for light jets, the impact parameter remains symmetric around zero. A tight selection on impact parameter helps to reject the tracks associated with the background process.

The important variables for b tagging algorithms are IP significance of the tracks, the position of the secondary vertex and transverse momentum of muon with respect to jet direction. CMS used jet-probability (JP) (uses impact parameter significance of the tracks) and combined secondary vertex (CSVv2) (combines the information of displaced tracks with the information of secondary vertices associated with the jet using a multivariate technique) taggers during Run-1 [127]. For Run-2, the new versions of the b tagging algorithms known as DeepCSV and DeepJet have been developed using deep neural network (DNN) [128] training.

DeepCSV and DeepJet: DeepCSV uses input tracks and secondary vertex information similar to JP and CSVv2 taggers including track-variables up to six tracks, p_T, η information of all flavour jets to learn the correlation between jet kinematics and other input variables. The DNN training is performed using jets with p_T up to 1 TeV and

within the tracker acceptance. A mixture of $t\bar{t}$ and multijet events is used so that training could learn about heavy flavour and light flavour jets and discriminate them. The neural network uses four hidden layers that are fully connected. A normalized exponential function is used to activate the nodes in the last layer so that the output value can be interpreted as a likelihood/probability P(f) for a specific jet flavour category. There are five such categories known as jet having one b hadron P(b), at least two b hadrons P(bb), one c hadron without any b hadron P(c), at least two c hadrons without any b-hadron P(cc) and not any b/c hadron P(udsg). A b discriminator P(b)+P(bb) is combined to tag a b jet in physics analyses. The performance of DEEPCSV improves with the 2017 and 2018 data-taking years.

Apart from DEEPCSV, CMS has developed one more DNN based b tagger DEEPJET [34] which uses PF candidates as input, which results in similar and better performance for 2016, 2017 and 2018 data-taking years. Unlike DEEPCSV, it has seven output categories. Four are similar to DEEPCSV. The rest of the three (P(lepb), P(g) and P(uds)) discriminate against b jet having lepton or jet originating from gluon or light quark. The difference between both DNN based b taggers is explained in Fig. 3.7.

	DeepJet	DeepCSV
Input	Particle Flow candidates	Same observables as CSVv2 b-tagger
Additional feature	Use soft lepton information	Use more charge particle tracks
	Discriminate uds jets with gluon jets	----
	Recovers degradation at high momentum	----
	Similar performance for phase0 and phase1 pixel configuration	Performance degrades for phase0 (2016)
Category	7 P (b/bb/c/cc/uds/g/lepb)	5 P (b/bb/c/cc/udsg)
b-discriminator	P (b + bb + lepb)	P (b + bb)

(a) Difference between DeepCSV and DeepJet
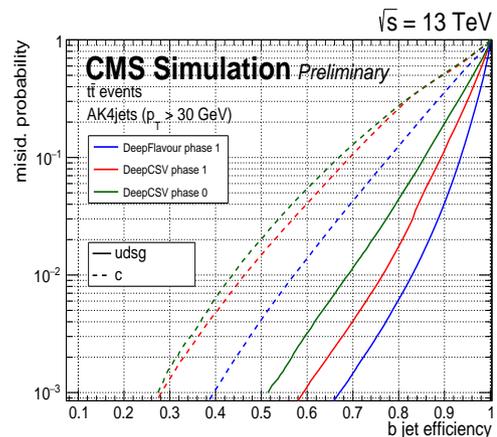
(b) DeepJet and DeepCSV performance curve for AK4 jets

Figure 3.7: DEEPJET and DEEPCSV b taggers [34]

Within an analysis, we can use either a selection on b discriminant or full shape of b discriminant. The selection is applied depending upon the Loose, Medium or Tight

working point corresponding to 10%, 1% or 0.1% misidentification probability or mistag rate. While using b tagging algorithms, we need to consider how much differently the algorithms behave for jets reconstructed in simulations and data. For this purpose, CMS provides the scale factors (SFs) to

(a) correct separately for the final yield of jets tagged as heavy or light flavour (working point-based calibration), or

(b) correct for the whole shape of the discriminator (shape calibration or reshaping), if the analysis needs it (for example, as an input to MVA training).

The SFs of a jet depends on its flavour, p_T and η (and jet discriminator value for reshaping). Depending upon the properties of the jet, SFs are calculated and applied to simulations to minimize discrepancy with data.

The approach of tagging a b jet in a boosted regime is different from general AK4 b jet tagging. We will study it in Section 3.5.5 after going through the Higgs jet tagging methods for the boosted $H \rightarrow b\bar{b}$ process.

3.5.2 b jet energy regression

The b jet composition contains leptons and neutrinos. Due to the weakly interacting nature of neutrinos, they can escape from the detector easily. That is why the energy of jets originating from b quarks cannot be fully reconstructed, making the jet energy resolution worse. An analysis-independent b jet energy regression, developed by CMS collaboration, is used to improve it. It is based on the neural network and trained on jet properties and jet composition of truth b jet. With this technique, a correction factor is calculated and applied for p_T and energy of b jet. The regression results in around 20% improvement in the signal dijet mass (m_{jj}) resolution within the analysis phase space without altering the reconstructed dijet background mass shapes much. Additional details on this regression can be found in the paper HIG-18-027 [35].
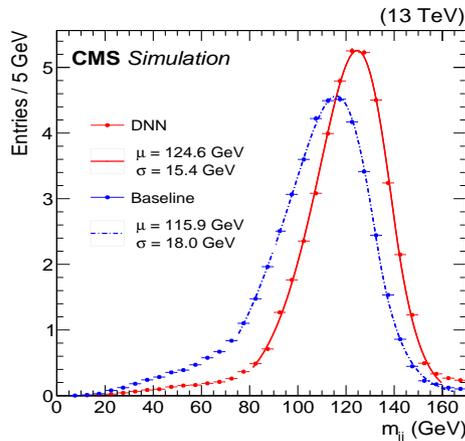
Figure 3.8: Dijet invariant mass distribution before and after b jet energy regression for $H \to b\bar{b}$ events [35]

3.5.3 Photon energy regression, scales and smearing methods

Due to shower leakages, gaps, and dead crystals, the ECAL is not suitable for collecting all of the energy deposited by photons. They cause systematic variations in measured ECAL energy and reduce photon energy resolution. A multivariate technique-based correction is used to minimize the impact of these losses [36]. The target of this regression is the ratio between generator level and reconstructed level photon energy[1], and its output correction factors are applied to reconstructed energy of data and simulated events to obtain the best estimate of the true energy.

There remains a slight variation in energy scales and resolutions for data and simulation after applying the energy regression, which is fixed by scaling and smearing correction factors. The studies are performed using simulated $Z \to ee$ events. For photons, only ECAL reconstructed information of $Z \to ee$ events is used [36].

The photon energy scales are adjusted by changing the data scale to match the scale observed in simulated events. The results of fitting the invariant mass m_{ee} distributions in different eta regions, obtained from data and simulated events separately, are compared to derive a scale offset. Spreading method extracts corrections to both the energy

---

[1]The methods are applicable for both electrons and photons, but I am explaining them considering photon as the object of analysis interest.
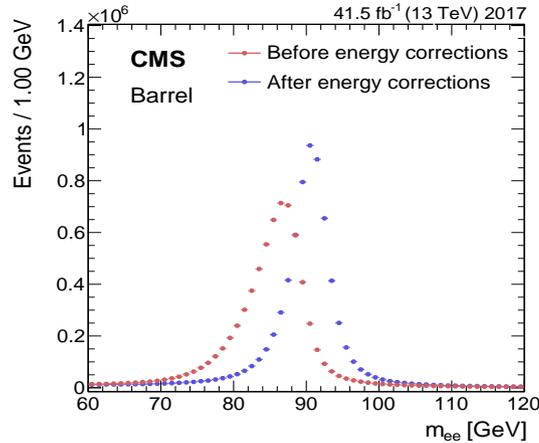
Figure 3.9: Dielectron invariant mass distribution before and after all the energy corrections (regression and scale corrections) for barrel for $Z \to ee$ events [36]

resolution in the simulation and the scale for the data in bins of $|\eta|$ and R_9 in the second stage. It fixes the residual discrepancy between data and simulation in m_{ee} distributions by applying an energy Gaussian spreading function to simulated events.

Depending on the pseudorapidity region and energy loss in the detector material, the ultimate energy resolution after all corrections (regression and scale corrections) ranges from 2 to 5%.

3.5.4 Photon Identification

There are two different methods to identify the photons in the final state of any physics process at LHC. One is the cut-based method, where some selections are applied on object-specific identification variables. The second one is the multivariate technique, where using the same identification variables, BDT is trained to discriminate signal and background photons. For precision measurements, the latter one is more optimal. This section will briefly describe the identification variables and multivariate photon identification technique [36].

3.5.4.1 Identification Variables:

A photon from new physics is identified as a prompt (originated from PV) and isolated photon. These features are used as identification variables to reject background photons, e.g., jets reconstructed from light neutral mesons π^0 or η, emerging from fragmentation, can give two photons which is the most critical background for prompt photons. To develop photon identification methods, the following identification variables are used:

- **Isolation variables:** This is constructed by the sum of the transverse momenta of all hadrons and photons within a $\Delta R = 0.3$ isolation cone in $\eta - \phi$ plane around photon object. The isolation thresholds depend on the energy of the photon objects.

- **Shower shape variables:** Another strategy for rejecting high-electromagnetic-content jets take advantage of the ECAL electromagnetic shower shape. The energetic jets with photons from hadronic decay make a wider shower within ECAL in comparison to an isolated single photon. The following are two of the most relevant variables used for photon identification depending upon the geometric shower shape from prompt and background photons:

  H/E ratio: The ratio of energy stored in the HCAL in a cone of radius $\Delta R =0.15$ around the supercluster direction compared to the energy of the photon candidates is known as the H/E ratio. For low energy photons, HCAL contribution comes due to HCAL noise and pileup, while for high energy photons, it is due to leakage of photons through the inter-module gaps.

  $\boldsymbol{\sigma_{i\eta i\eta}}$: This variable is designed to filter out ECAL noise by ensuring that crystals with energy deposits of at least 0.9% of $E_{5\times5}$ (the energy deposited in a 5×5 crystal matrix around the most energetic crystal) contribute to $\sigma_{i\eta i\eta}$ defined as:

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5\times5} \omega_i (\eta_i - \bar{\eta}_{5\times5})^2}{\sum_i^{5\times5} \omega_i}} \tag{3.1}$$

where η_i is the pseudorapidity of the ith crystal, $\bar{\eta}_{5\times5}$ is mean pseudorapidity of crystal matrix, and ω_i is weight factor defined as $\max(0, 4.7 + ln(E_i/E_{5\times5}))$, and

is nonzero if $ln(E_i/E_{5\times5}) > -4.7$ that implies E_i 0.9% of $E_{5\times5}$.

For single photon or electron shower, $\sigma_{i\eta i\eta}$ the distribution should be narrow, but for two-photon showers resulting from neutral meson decays, it should be wider.

- **R_9 variable:** It is the ratio between the energy deposited in a 3×3 ECAL crystal matrix around the most energetic crystal in the supercluster and the supercluster energy. Photons that convert before reaching ECAL have a lower R_9 value than unconverted photons.

Using the above criterion, the photon identification methods are build and used for physics analyses which are given in Sections 3.5.4.2 and 3.5.4.4.

3.5.4.2 Cut-based photon identification

In this approach, selections are applied to photon identification variables. According to different working points, different thresholds are optimized on these variables.

3.5.4.3 Electron Veto

This selection is used to reject electron from photon identification. There are two methods for vetoing the electrons, 1) conversion safe electron veto 2) pixel seed electron veto. The first one requires the absence of charged-particle tracks, as well as a hit in the pixel detector's innermost layer that does not match with reconstructed conversion vertex pointing to the ECAL photon cluster. It is appropriate for analyses where electrons do not constitute a major background. The latter rejects the photons having at least two-pixel hits in the pixel detector points to the ECAL around photon supercluster. It is used when electrons are major background.

3.5.4.4 Multivariate technique based photon identification

It is commonly used photon identification based on BDT training. A BDT is trained using very loose selections on photon identification variables with the median energy per unit area (ρ), η and uncorrected energy of photon supercluster as input. The BDT

output discriminates prompt photons from background jets which can be misidentified as photons.

A comparison of the performance between cut-based identification and BDT identification for photons is shown in Fig. 3.10 which clearly shows for a fix background misidentification rate, MVA based photon identification performs better.
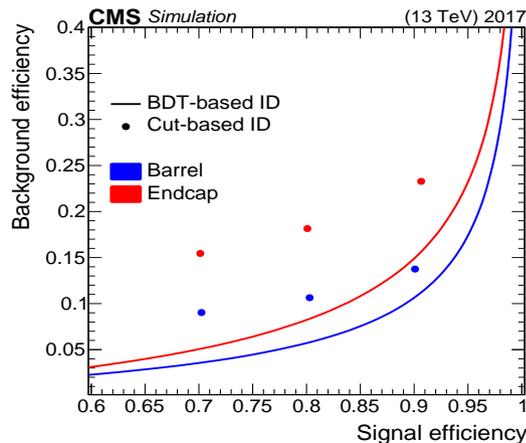
Figure 3.10: Performance of the photon BDT and cut-based identification algorithms. Cut-based identification is shown for three different working points, loose, medium and tight [36]

3.5.4.5 Diphoton vertex identification method

For $H \rightarrow \gamma\gamma$, the $m_{\gamma\gamma}$ mass resolution depends on the photon energy and angle between two photons. By choosing a diphoton vertex within 10 mm distance from PV, removes mass resolution dependency on angle variable. The methods can be studied in Ref. [129]. The kinematic features of the diphoton system and their correlations with the kinematic properties of the recoiling tracks are used to identify the diphoton vertex indirectly. For each reconstructed PV, three discriminating variables are used as BDT training input: 1) $\vec{p}_T^2$, 2) $-\sum(\vec{p}_T \cdot (\vec{p}_T^{\gamma\gamma}/|\vec{p}_T^{\gamma\gamma}|))$ and 3) $(|\sum \vec{p}_T| - |\vec{p}_T^{\gamma\gamma}|)/(|\sum \vec{p}_T| + |\vec{p}_T^{\gamma\gamma}|)$ (the sum runs over the transverse momentum vectors of the charged tracks, $\vec{p}_T$ and $\vec{p}_T^{\gamma\gamma}$ represents the transverse momentum vector of the diphoton system). Additionally, to deal the case where photon conversion leads to leave tracks within tracker, a pull variable $g_{conv} = |z_e - z_{vtx}|/\sigma$ is also used where z_e and z_{vtx} are the longitudinal position of estimated PV and reconstructed

vertex. The method is validated to $Z \to \mu\mu$ simulation events by removing muon tracks to mimic it like a diphoton event and checked how much efficient the training is to find a vertex within 10 mm of true vertex. The use of tracks from converted photons for vertex finding is validated on γ+jets simulated events. The ratio of the measured efficiency in data and simulations is within one percent of unity. It, as function of Z boson p_T, is further used for vertex finding in simulated Higgs boson events which results 80-90% efficient. Apart from this, a second vertex identification is also used to select event by event a diphoton vertex within 10 mm of diphoton interaction point. For this method, BDT is trained with following inputs from each event:

- vertex identification BDT output for the three most likely vertices from first method

- the total number of reconstructed vertices

- the transverse momentum of the diphoton system

- the distances between the chosen vertex and the second- and third-best vertices from first method

- the number of photons with an associated conversion track(s).

This BDT training is validated using similar approach as first one. The vertex probability BDT score holds a linear relationship with the probability of identifying a close-enough vertex (vertex probability).

3.5.5   Higgs jet tagging

Within HL-LHC projection studies in Chapter 4, BSM di-Higgs boson searches are explored in the boosted regime. Therefore understanding of the boosted Higgs jet identification is essential. As the SM predicts the highest branching ratio for $b\bar{b}$ as the final state for Higgs boson decay, boosted $H \to b\bar{b}$ tagging might be crucial for new physics searches. In boosted regime, jet mass is the first jets' property that distinguishes signal jet from background one. The jet mass of the signal should peak around resonance

mass (Higgs boson in our case). However, this happens only in an ideal picture. In reality, the jet mass distribution gets worse, and its peak gets shifted due to soft emissions, pileup and underlying events. For $H \to b\bar{b}$ decay, QCD multijet processes (mainly gluon splitting in $b\bar{b}$) are dominant backgrounds. The background jets acquire mass through showering, which grows as a function of transverse momentum. Boosted hadronic objects keep a different energy pattern than QCD jets of comparable invariant mass. It motivates to go beyond jet mass and exploit jet substructure [130, 131]. It removes soft radiation contamination and identifies and selects features related to hard decay.

The main idea of H tagging is to reconstruct a large area jet with jet substructure (grooming and tagging) algorithms that remove soft contribution and try to understand the subjet structure and jet-shape of the jet, which discriminate signal jets from background jets. At this stage, selected jet is called the Higgs jet. In the end, using boosted b tagging algorithms, subjets are identified as b jets, and the Higgs jet is identified as a jet with two b-hadrons. Let's understand all these steps one by one.

- **Jet grooming:** When a boosted large jet originate from Higgs boson, its mass should peak near Higgs boson mass. Jet grooming removes background contamination and related component from the jet mass calculation. Soft drop declustering is one of the grooming technique. It suppresses wide-angle soft radiation from a jet in order to mitigate the effects of contamination from initial state radiation (ISR), underlying event (UE), and multiple hadron scattering (pileup).

  Starting from AK8 reclustered jet with radius R_0, it does declustering of the last step and drop the soft contributions unless

  $$soft\ drop\ condition = \frac{min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut}(\frac{\Delta R_{12}}{R_0})^{\beta} \qquad (3.2)$$

  where p_{Ti} is transverse momenta of constituents, ΔR_{12} is distance in $\eta - \phi$ plane, z_{cut} is the soft drop threshold, and β is angular exponent. For wide-angle soft radiation, soft drop condition does not fulfil. The efficiency and performance of

grooming depends on z_{cut} and β ($\beta \to \infty$ for ungroomed jet). For $\beta < 0$, it works as tagger vetoing jets that do not have two well-separated hard prongs. While for $\beta > 0$, it works as groomer changing the constituents of a jet. The steps to understand the workflow are (starting from AK8 jet):

1) recluster jet with Cambridge-Aachen (C/A) algorithm for pair wise clustering depending upon the distance between jet constituents.

2) break the jets by undoing the last step and form two subjets

3) check soft drop condition for two subjets. If pass then jet is final soft-drop jet. Otherwise take highest p_T subjet as jet and repeat all the steps.

4) if j can not be declustered further, one can remove j from selection (tagging mode) or consider as final soft drop jet (grooming mode).

This technique is IRC safe and removes all soft contributions. CMS uses this technique as a standard choice for jet grooming.

- **Jet shapes:** Jet shape methods effectively tag boosted objects with jet-based observables that take advantage of the different energy flow in signal and background jets' decay patterns. Within thesis we use "N-subjettiness (τ_N)" [132] which effectively count number of subjets by looking at the energy lobes distributed in a given boosted jet and help to reject background jets by cutting on one-dimensional function $f(\tau_1, ..., \tau_N)$. It also provide a inclusive jet shape and is independent of any other jet substructure algorithm. For a boosted jet with n subjets τ_N is defined as,

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \; min(\Delta R_{1,k}, ..., \Delta R_{N,k}) \tag{3.3}$$

where k runs over all constituents particle with $p_{T,k}$ transverse momenta in a given jet of radius R_0 and $\Delta R_{N,k} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is distance between Nth subjet and kth constituent particle. The normalization factor d_0 is given as $d_0 = \sum_k p_{T,k} R_0$. Jets with $\tau_N \sim 0$ will have N subjets. For higher τ_N jets will have at least N+1 subjets. An appropriate τ_N depends on correct choice of all subjet directions for

which it acquires a minima. In that case, it becomes decreasing function of N, and follow $0 < \tau_N/\tau_{N-1} < 1$.

For $H \to b\bar{b}$ decay, we expect maximum 2 subjets thus, in Eq. (3.3), N might be equal 1 or 2. This leads to take τ_2/τ_1 as discriminator for Higgs jet tagging.

- **boosted b tagging:** For boosted jets, b tagging can be applied either on the AK8 jet or its subjets. For boosted training, jet mass is used to identify b jets from Higgs boson decay. These jets are first corrected with jet substructure techniques to remove soft-radiation and identify subjets. For thesis work, we use subjet b tagging using DEEPCSV b tagger where both subjets from $H \to b\bar{b}$ decay are required to be tagged. For high p_T regime, subjet b tagging gives the best performance to reject background light jets as given in Ref. [61].

3.5.6   Pile-Up Per Particle Identification (PUPPI)

For HL-LHC projection studies, pileup is a big challenge that might increase up to 200. PUPPI [126] is one of the new idea proposed for pileup mitigation. For projection studies, we have used PUPPI weighted jets. The algorithm uses global information (like PV) of an event and local information (like tracks) at particle level to identify pileup. Within this algorithm, a weight is calculated, using this global and local information, for each particle in the event. It is calculated by a shape parameter α, which distinguishes parton shower-like radiation from pileup-like radiation for every particle in the event. Apart from the shape parameter, p_T also helps for pileup mitigation as p_T distribution for pileup falls much faster. The tracking information also helps to differentiate between charged tracks that originate from the primary vertex (PV) and the charged tracks that originate from the pileup. The PF algorithm can be used to relate these tracks to particles. Using PF, particles can be sorted into three class: charged hadrons from PV, charged hadrons from pileup and neutral hadrons from both pileup and PV. For all neutral particles, the algorithm assumes them to be originated from PV and assign weights depending upon their p_T. Therefore, for soft contributions from pileup, it assigns low weights, which

reduce pileup neutral contribution from the event. Hence PUPPI even works for the region where tracking is not available.

The PUPPI weights are further used to rescale the four momenta of the particles. Ideally, this weight is one for particles from hard scattering, and zero for pileup originated particles. However, in the real picture, the weights can be in fractional form, depending upon the particles' properties. Particles with a very small weight are discarded. Using PUPPI weighted particles, one can perform jet clustering without any other treatment for pileup. This algorithm has proven very efficient to correct jet p_T, jet mass and missing transverse energy in a high pileup collision conditions.

Chapter 4

Searches for Higgs pair production in $b\bar{b}b\bar{b}$ final state at HL-LHC

4.1 What is High Luminosity LHC (HL-LHC)?

The HL-LHC project aims to enhance the LHC's performance in order to boost its discovery potential after 2027 [133]. It consists of more powerful beam bending magnets with upgraded accelerator chains (PS, SPS, PSB, LINAC). The aim is to increase luminosity by a factor of ten compared to the LHC's design value. The HL-LHC will generate about 250 inverse femtobarns of data per year. For both ATLAS and CMS experiments, the HL-LHC project will deliver proton-proton collisions at 14 TeV with an integrated luminosity of 3 ab^{-1}.

4.1.1 Phase-2 CMS experiment upgrades for HL-LHC

To handle the high pileup collisions with large amount of radiations, experiments at LHC need an upgrade to work efficiently in HL-LHC collision conditions. Therefore a long shutdown (LS) period has been scheduled in HL-LHC plans. During second and third LS, CMS will upgrade its inner tracking system, trigger system, calorimeter, and muon detection systems. This is called CMS Phase-2 detector shown in Fig. 4.1.

Trigger/HLT/DAQ
- Track Information in Trigger (Hardware)
- Trigger latency 12.5 μs - output rate 750kHz
- HLT output 7.5 kHz

Barrel EM calorimeter
- New FE/BE electronics with improved time resolution
- Lower operating temperature

Muon Systems
- New DT & CSC FE/BE electronics
- Complete RPC coverage 1.6 <|η| < 2.4
- Muon tagging 2.4 < |η| < 3

New endcap calorimeter
- Radiation tolerant
- High granularity: increased transverse and longitudinal segmentation
- Precise timing capabilities

New Tracker
- Rad. tolerant, increased granularity, lighter
- 40 MHz selective readout in outer tracker for trigger
- Extended coverage to |η| ~ 3.8

Figure 4.1: Phase-2 upgrades for the CMS detector [37–45]

4.2 Goals of HL-LHC projection studies

The number of collisions within a particular amount of time is related to luminosity, an essential metric of an accelerator's performance. Thus, the more data experiments can collect, the more rare processes they will observe.

The HL-LHC would allow physicists to investigate known mechanisms in more depth, such as the Higgs boson, and observe rare new occurrences. Precise SM measurements, searches for BSM physics, flavor physics of heavy quarks and leptons, investigations of the characteristics of the Higgs boson, and investigations of QCD matter at high density and temperature are the five primary goals of the HL-LHC. For example, the HL-LHC will produce at least 15 million Higgs bosons every year, compared to roughly three million from the LHC in 2017 for precise Higgs boson measurements. For the thesis work, SM and BSM Higgs boson physics sensitivities in the HL-LHC scenario are projected. The work reveals future prospects for SM precision measurements and new particle searches in the Higgs sector with the boosted regime only. Now we will go through the HL-LHC projection studies with Sections 4.3 and 4.4.

4.3 VBF resonant di-Higgs production

The search for new physics resonances decaying to a pair of Higgs bosons (H) is motivated by several BSM scenarios. Such models include warped extra dimensions (WED) [58] having particles such as the spin-0 radion [134–136] and the spin-2 first Kaluza-Klein (KK) excitation of the graviton [137–139]. Others, such as the two-Higgs doublet models [140] (particularly, the minimal supersymmetric model [141]) and the Georgi-Machacek model [142] also contain spin-0 resonances. These resonances may have a sizeable branching fraction to decay into a HH pair.

Searches for a new particle X in the HH decay channel have been performed by the ATLAS [143–145] and CMS [146–150] Collaborations in proton-proton (pp) collisions at $\sqrt{s} = 7$ and 8 TeV. The ATLAS Collaboration published limits on the production of a KK bulk graviton decaying to HH in the final state with a pair of b quark and

antiquark ($b\bar{b}b\bar{b}$), using pp collision data at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 3.2 fb^{-1} [151] and 13.3 fb^{-1} [152]. More recently, the CMS Collaboration has published limits on the production of a KK bulk graviton and a radion, decaying to HH, in the $b\bar{b}b\bar{b}$ final state, with pp collision data at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 35.9 fb^{-1} [153]. Overall, the searches from ATLAS and CMS set a limit on the production cross sections and the branching fractions $\sigma(pp \to X)B(X \to HH \to b\bar{b}b\bar{b})$ for masses of X (m_X) up to 3 TeV.

The above searches looked at the s-channel production of a narrow resonance X produced from the SM quark-antiquark or gluon-gluon interactions. The WED models are used in the interpretations of the results. In these models, the extra spatial dimension is compactified between two branes (called the bulk) via an exponential metric κl (the warp factor) and l the coordinate of the extra spatial dimension [154]. The fundamental scale is the reduced Planck scale ($\overline{M_{pl}} \equiv M_{pl}/8\pi$, M_{pl} being the Planck scale) and the ultraviolet cutoff of the theory $\Lambda_{\mathrm{R}} \equiv \sqrt{6}e^{-\kappa l}\overline{M_{pl}}$ [134]. Assuming $\Lambda_{\mathrm{R}} = 3$ TeV, a spin-0 radion of mass below 1.4 TeV is excluded at a 95% confidence level [153].

The absence of evidence of such a signal may point to highly suppressed couplings of X with the SM quarks and gluons. In such cases, alternative production modes may become the most dominant ones, such as through vector boson fusion (VBF). Moreover, the particles may have a substantial width compared to their masses, owing to which the sensitivity of dedicated searches for narrow resonances may be vastly decreased. In either case, a much higher amount of data may be needed to detect these particles, owing to a smaller VBF production cross section and the difficulty of locating a wide resonance over the backgrounds.

In this analysis, the prospects of a search for a massive resonance produced through VBF and decaying to HH (Fig. 4.2) using pp collisions at a center-of-mass energy of 14 TeV at the HL-LHC, assuming a data set corresponding to an integrated luminosity of 3 ab^{-1} collected by the upgraded Phase-2 CMS detector, is explored. This is the first CMS analysis which studies the resonant production in VBF mode. The produced Higgs bosons are highly Lorentz-boosted for a high mass resonance, and each reconstructed as
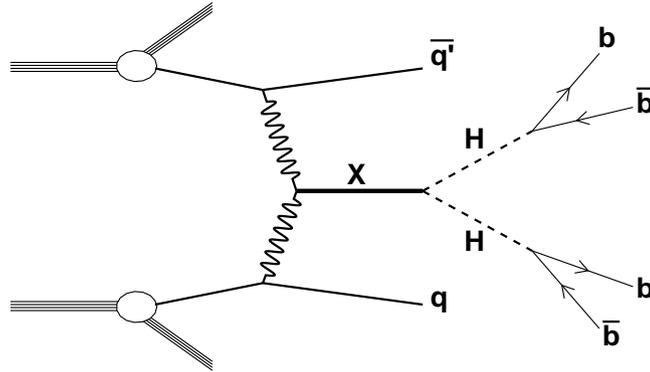
Figure 4.2: The vector boson fusion mode of production of a resonance X decaying to a pair of Higgs bosons H, with both Higgs bosons decaying to $b\bar{b}$ pairs.

a large area jet (Higgs jet). In addition, a signal event is characterized by two energetic jets at large pseudorapidity η, θ being the polar angle of the jet measured in the CMS detector coordinate system), arising from the VBF production mode. The background consists mainly of the standard model multijet processes.

The unique topology of the production and decay would benefit from an upgraded CMS detector as shown in Fig. 4.1. First, the production is accompanied by two energetic forward jets. A high granularity calorimeter (HGCAL) would enable identification of the VBF jets with an improved rejection of jets arising from additional pp collisions in the same or adjacent LHC bunch crossings (pileup) accompanying the main event. Second, the Higgs jets would benefit from the increased coverage of the tracker and the calorimeters. Higgs jet tagging includes jet flavor tagging, which would improve with the extended tracker coverage. The proposed CMS upgrades to calorimetry, the HGCAL, would also enable a more precise measurement of the jet energies, thus improving the resolution of the di-Higgs jet resonance. We study the search sensitivity of a bulk graviton of mass between 1.5–3.0 TeV and a width narrow to up to 5% of the resonance mass.

This study has been organized in the following way: Section 4.3.1 provides the details of simulations used for the studies. All the event selections are described in Section 4.3.2. The basic kinematic distributions are added in Section 4.3.3. Background estimation method is explained within Section 4.3.4. In the end, results are added in Section 4.3.5.

4.3.1 Simulation samples

Simulated bulk graviton signals of different mass and width configurations are given in Table 4.1. The cross sections are calculated for the narrow width approximation (NWA).

Table 4.1: All the generated signal mass points with respective widths and cross section, using the narrow width approximation (NWA) [8].

Mass of Graviton (TeV)	Widths	Cross section using NWA [fb]
1.5	1% 5%	1713
2.0	1%	416.2
2.5	1%	121.6
3.0	1%	39.20

The main backgrounds in this search are multijet events from quantum chromodynamics (QCD) interaction. Trace amounts of $t\bar{t}$+jets events may be present, estimated to be $< 1\%$ in the CMS analysis using 2016 data [153, 155] and is neglected in the present analysis. The names of the simulated signal and background samples are given in Table 4.2.

Table 4.2: Simulations used for modeling the VBF bulk graviton production and the backgrounds. The signal cross sections are what are used to normalize the figures.

Dataset	σ [pb]	Events
Signals		
/VBF_RS_bulk_M1500_W01pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-noPU_93X_upgrade2023_realistic_v2-v1	1×10^{-3}	80200
/VBF_RS_bulk_M1500_W01pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-PU200_93X_upgrade2023_realistic_v2-v1	1×10^{-3}	27720
/VBF_RS_bulk_M2000_W01pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-noPU_93X_upgrade2023_realistic_v2-v2	1×10^{-3}	100000
/VBF_RS_bulk_M2000_W01pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-PU200_93X_upgrade2023_realistic_v2-v2	1×10^{-3}	81850
/VBF_RS_bulk_M3000_W01pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-noPU_93X_upgrade2023_realistic_v2-v1	1×10^{-3}	91351
/VBF_RS_bulk_M3000_W01pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-PU200_93X_upgrade2023_realistic_v2-v2	1×10^{-3}	87507
/VBF_RS_bulk_M3000_W05pc_14TeV-madgraph/PhaseIITDRFall17MiniAOD-PU200_93X_upgrade2023_realistic_v2-v1	1×10^{-3}	42552
Backgrounds		
/QCD_Mdijet-1000toInf_TuneCUETP8M1_14TeV-pythia8/PhaseIITDRFall17MiniAOD-noPU_93X_upgrade2023_realistic_v2-v1	99.1990	4098542
/QCD_Mdijet-1000toInf_TuneCUETP8M1_14TeV-pythia8/PhaseIITDRFall17MiniAOD-PU200_93X_upgrade2023_realistic_v2-v2	99.1990	3802314
/TT_TuneCUETP8M2T4_14TeV-powheg-pythia8/PhaseIITDRFall17MiniAOD-noPU_93X_upgrade2023_realistic_v2-v1	864.5	4979816
/TT_TuneCUETP8M2T4_14TeV-powheg-pythia8/PhaseIITDRFall17MiniAOD-PU200_93X_upgrade2023_realistic_v2-v3	864.5	2874776

4.3.2 Event selection

The first task is to pick optimized event selection depending upon the signal topology. The characteristics of the signal events is shown in Fig. 4.3 (AK8 jet p_T), Fig. 4.4 (AK8 jet soft-drop mass), Fig. 4.5 (AK8 jet N-subjettiness). The VBF characteristic of the event is given by the separation $\Delta\eta(j_1, j_2)$ between two AK4 jets in the forward region, as shown in Fig. 4.6.
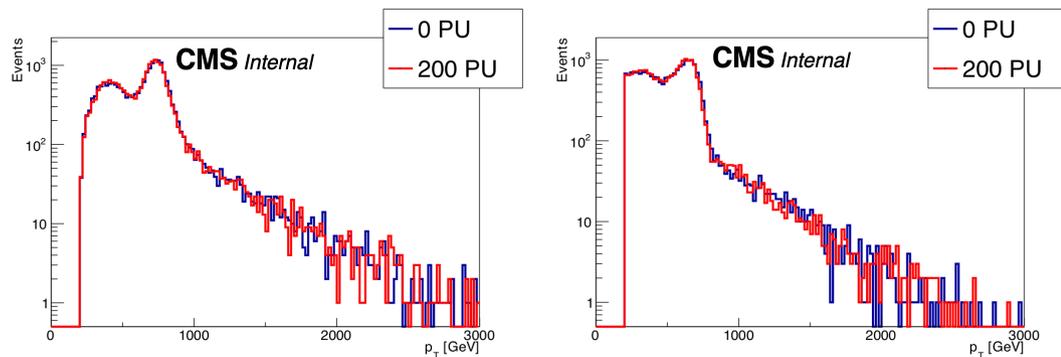
Figure 4.3: Leading (left) and subleading (right) AK8 jet p_T for $M_{BG} = 1.5$ TeV
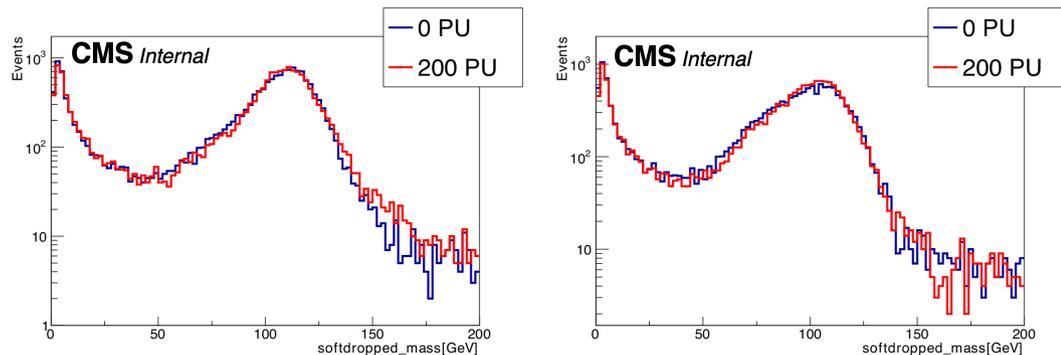
Figure 4.4: Soft drop mass of leading (left) and subleading (right) AK8 jet p_T for $M_{BG} = 1.5$ TeV

The AK8 jets are first groomed to remove soft and wide-angle radiation using the soft-drop algorithm explained in Section 3.5.5. The soft-drop algorithm gives two subjets each, for J_1 and J_2, by undoing the last stage of the jet clustering. The
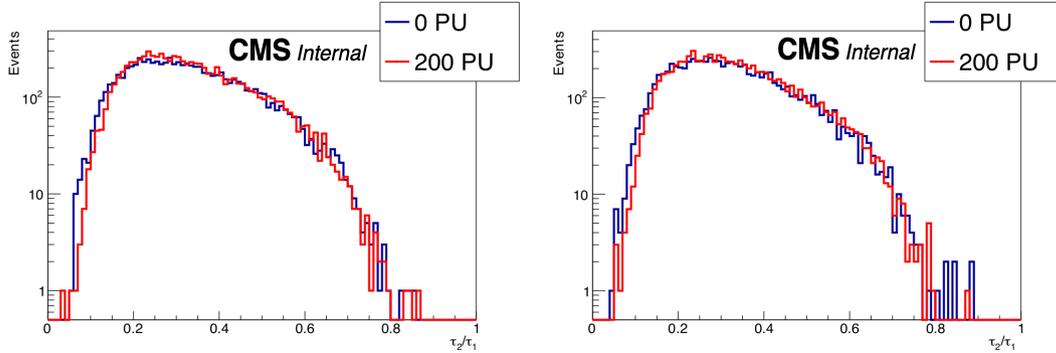
Figure 4.5: N-subjettiness variable τ_2/τ_1 of leading (left) and subleading (right) AK8 jet p_T for $M_{BG} = 1.5$ TeV
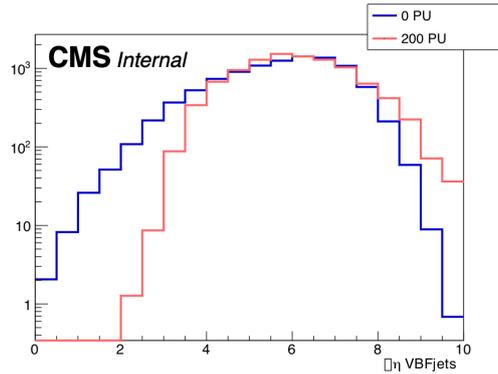
Figure 4.6: $\Delta\eta$ between the VBF jets for $M_{BG} = 1.5$ TeV

invariant mass of the two subjets is the soft-drop mass of each AK8 jet, which has a distribution with a peak near the Higgs boson mass $m_H = 125$ GeV. The leading two AK8 jets (J_1 and J_2) in the event are required to pass the Higgs jet criteria listed in Table 4.3. It comprises selections on the jet p_T and η, the soft-drop mass, and subjet b tagging. In Appendix A.1.1, studies to select optimal b jet tagger between CSVv2 and DEEPCSV have been discussed. The subjet b tagging uses the DEEPCSV algorithm with a choice between the DeepCSVL (discriminator > 0.1522) loose threshold with 10% misidentification rate or DeepCSVM (discriminator > 0.4941) medium threshold with 1% misidentification rate. We performed a detailed optimization of the soft-drop

mass window and the subjet b tagging operating points (OP) based on expected signal significance as given in Sections 4.3.2.2 and 4.3.2.1. If the event does not contain two Higgs jet candidate, the event is discarded.

Table 4.3: Higgs jet selection based on AK8 jets.

Variable	Applied selection criteria
Boosted jets	$p_T > 300$ GeV
Central jets	$|\eta| < 3$
Soft drop mass	90-140 GeV
N-subjettines	$\tau_2/\tau_1 < 0.6$
Subjet b tagging	≥ 1 pass DeepCSV medium

The AK4 jets must be separated from the H-tagged AK8 jets by $\Delta R > 1.2$. An event is required to have at least two AK4 jets (j_1 and j_2) passing the VBF jet requirements in Table 4.4 to be selected as a signal event. The variation in signal and background efficiencies with selection cuts is shown in Fig. 4.7[1].
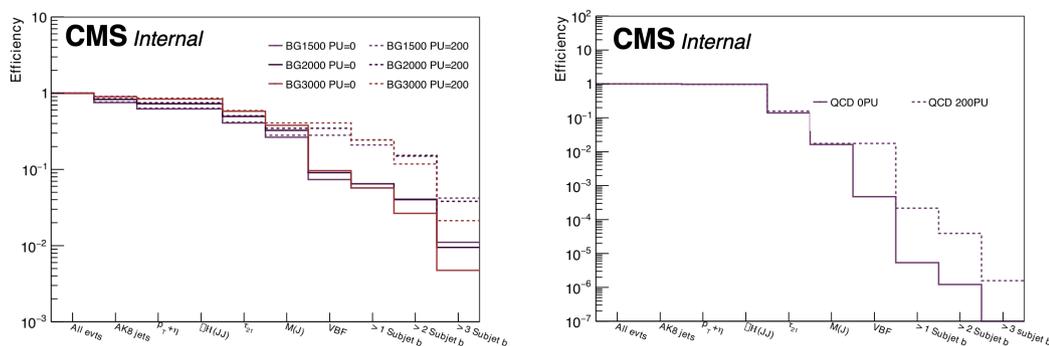
Figure 4.7: Variation in selection efficiencies with selection criteria for bulk gravitons (BG) of masses 1.5, 2.0 and 3.0 TeV with 1% widths (left) and QCD background(right), 200 pileups. The plots are shown using the DeepCSVM operating points for the subjet b tagging.

---

[1]Although it should be mentioned that additional requirements on selecting VBF jets do not bring any improvement in the analysis with current algorithms and techniques as shown in Fig. 4.7. Because of the high pileup, we are much more likely to select the pileup jets as VBF jets. We expect to have a better VBF jet tagging efficiency with future developments and techniques.

Table 4.4: VBF jet selection based on AK4 jets

Variable	Applied selection criteria
Separation from H jets	$\Delta R(\text{AK4, Higgs jets}) > 1.2$
Minimum AK4 jet p_T	$p_T > 50$ GeV
Forward jets	$\|\eta\| < 5$
Large η separation	$\Delta \eta(j_1, j_2) > 5$
Jets in opposite η	$\eta_{j_1} * \eta_{j_2} < 0$
Invariant mass	$m_{JJ} > 300$ GeV

After passing the Higgs jets and VBF selections, events are categorized according to the number of b tagged subjets-3b or 4b-among the four subjets from the two Higgs jets. Finally, the bulk graviton invariant mass m_{JJ} is reconstructed from the 4-momenta of two Higgs jets in the event belonging to the 3b or 4b categories separately.

4.3.2.1   Choice of subjet b tagging in the event selection

We looked at the optimization of event selection based on the choice of the operating point (Loose or Medium) for DEEPCSV subjet b tagger, with events categorized into those having three of four subjets b tagged, the 3b or the 4b categories, respectively.

Table 4.5: Subjet b tagging operating point optimization for bulk graviton (3 Subjets with DEEPCSV operating point loose (0.1522))

Mass[TeV](Width)	yields(PU=0)	yields(PU=200)	$S/\sqrt{B}(PU = 0)$	$S/\sqrt{B}(PU = 200)$
1.5 (1%)	197.58	736.80	3.26	1.85
2.0 (1%)	237.09	891.39	3.90	2.23
3.0 (1%)	230.97	971.75	3.80	2.43
3.0 (5%)	225.14	965.38	3.70	2.41
QCD	3683.89	159181.12		

Table 4.6: Subjet b tagging operating point optimization for bulk graviton (4 Subjets with DEEPCSV working point loose (0.1522))

Mass[TeV](Width)	yields(PU=0)	yields(PU=200)	$S/\sqrt{B}(PU = 0)$	$S/\sqrt{B}(PU = 200)$
1.5 (1%)	117.12	437.23	5.77	3.32
2.0 (1%)	129.84	489.64	6.40	3.71
3.0 (1%)	104.73	453.80	5.16	3.44
3.0 (5%)	104.77	466.02	5.17	3.54
QCD	411.21	17373.37		

Table 4.7: Subjet b tagging operating point optimization for bulk graviton (3 Subjets with DEEPCSV operating point medium (0.4941))

Mass[TeV](Width)	yields(PU=0)	yields(PU=200)	$S/\sqrt{B}(PU=0)$	$S/\sqrt{B}(PU=200)$
1.5 (1%)	120.22	447.19	5.48	3.38
2.0 (1%)	121.80	457.61	5.55	3.46
3.0 (1%)	79.54	353.90	3.62	2.68
3.0 (5%)	77.80	358.85	3.54	2.71
QCD	481.93	17461.61		

Table 4.5 shows the signal and multijets yields and the metric $S/\sqrt{B}$, for PU=0 and 200, for the event selection using DeepCSVL subjet b tagging and for the 3b category. Table 4.6 shows the same for the 4b category. Tables 4.7 and 4.8 show the signal and multijets yields and the metric $S/\sqrt{B}$, for PU=0 and 200, for the event selection using DeepCSVM for the 3b and 4b categories, respectively.

Table 4.8: Subjet b tagging operating point optimization for bulk graviton (4 Subjets with DEEPCSV operating point medium (0.4941))

Mass[TeV](Width)	yields(PU=0)	yields(PU=200)	$S/\sqrt{B}(PU=0)$	$S/\sqrt{B}(PU=200)$
1.5 (1%)	33.22	126.19	4.92	3.16
2.0 (1%)	28.47	114.43	4.21	2.86
3.0 (1%)	14.21	63.87	2.10	1.60
3.0 (5%)	14.27	70.29	2.11	1.76
QCD	45.63	1594.91		

At PU=200, using DeepCSVL, the 4b category performs better than the 3b category. However, using DeepCSVM at PU=200, the 3b category performs better than the 4b category. To have a most optimal b tagging choice, we combine both categories, and estimate the signal significance (including systematic uncertainties) as described in Appendix A.2, Fig. A.22. Based on the significance estimates, the DeepCSVM using the 3b or the 4b categories is our final choice for event selection[2].

---

[2]This study to adapt optimal b tagger working point has been performed with first choice of soft-drop mass window. The final choice of optimal working point is used with final choice of optimized mass window selection.

4.3.2.2   Choice of the soft-drop mass window

The soft-drop mass of the AK8 jets after the "N-1" selections is given in Fig. 4.8 for the $3b$ and the $4b$ event categories, using DEEPCSV medium. The distribution for the $4b$ category is a bit sparse, but there are enough events in the range of interest, i.e., around the Higgs boson mass. Using these distributions, summing up them for the two event categories, and using a figure of merit of $S/\sqrt{B}$, the soft-drop mass window was optimized. The figures of merit are given in Table 4.9. Initially, we used different soft-drop mass windows for the leading and 2nd-leading p_T AK8 jets. The dedicated study and plots are added in Appendix A.1.1. However, the ranking in p_T is subject to jet energy correction uncertainties, and hence, in the tuning of the soft-drop mass, we stick with the same mass window for both the AK8 jets. We used a soft-drop mass of **90–140 GeV** for the two AK8 jets based on optimization studies. It gives an overall improvement of 20-30% on the expected significance in comparison to the first choice.
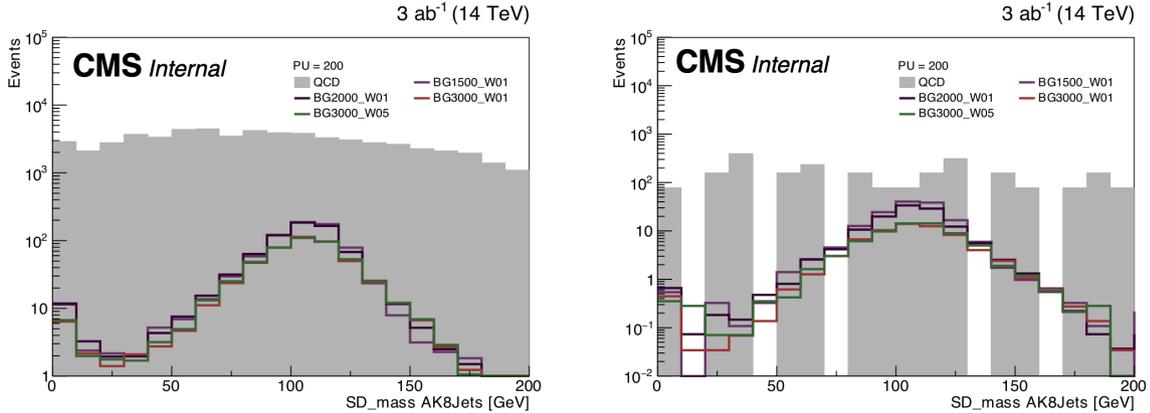
Figure 4.8: The soft-drop masses of the leading+2nd-leading p_T subjets for the $3b$ (left) and $4b$ (right) categories with pileup = 200. The DEEPCSV medium threshold is used.

4.3.3   The N-1 kinematic distributions

The N-1 distributions for a few kinematic variables are shown here (PU=0 is not a realistic scenario. Therefore, plots are for PU=200 only.):

Table 4.9: The figure of merit, $S/\sqrt{B}$, for the event selections reported in Table 4.3, with the exception of the soft-drop mass selection which is tuned, and Table 4.4.

Soft-drop mass [GeV]		$S/\sqrt{B}$	Comments
Leading-p_T AK8 jet	2nd-p_T AK8 jet		
80–160	60–140	2.06	first
80–160	80–160	3.12	New
60–140	60–140	2.73	New
90–140	90–140	3.34	New
90–130	90–130	3.25	New

- Fig 4.9: $3b$ category and PU=200;

- Fig 4.10: $4b$ category and PU=200.

Since the N-1 plots hold the complete event selections, except for the concerned variable, very few background events survive, especially in the $4b$ category. Hence, in the next section, a method of estimating the background distributions after the entire event selection is developed, which allows the prediction of the background shape with accuracy, without the limitations of a small sample of events after the final selection.
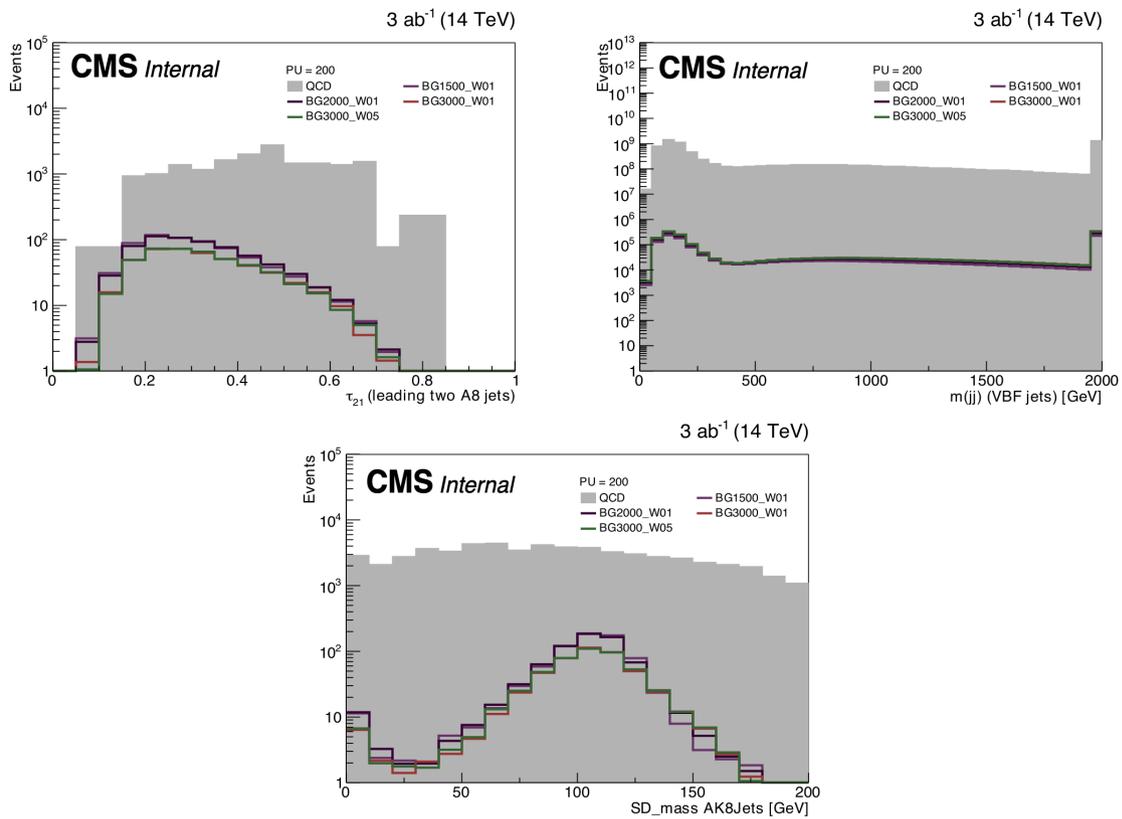
Figure 4.9: The N-1 distributions for the 3b category and PU=200 for τ_{21} (upper left), VBF dijet invariant mass (upper right), and the AK8 soft-drop mass (lower).
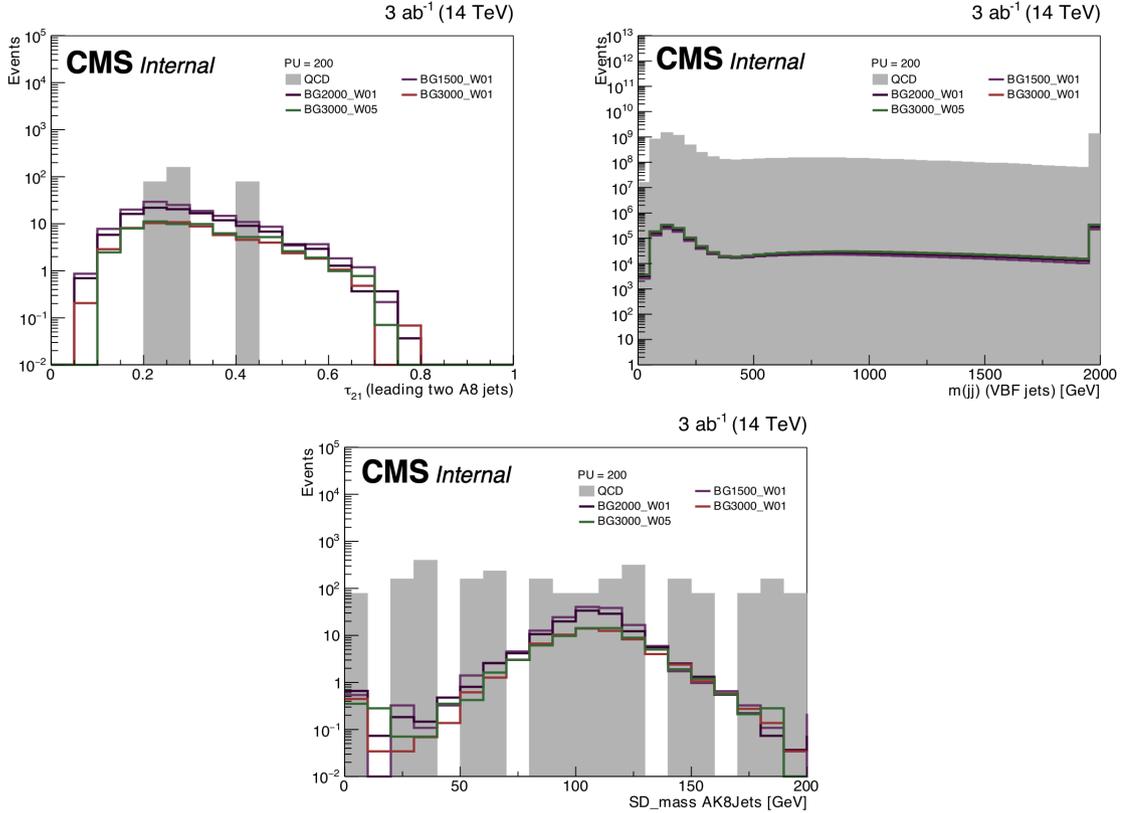
Figure 4.10: The N-1 distributions for the 4*b* category and PU=200 for τ_{21} (upper left), VBF dijet invariant mass (upper right), and the AK8 soft-drop mass (lower).

4.3.4 Background estimation

The simulated multijet background events have a sample size of ~ 4 million, and no events survive after the entire selections. Therefore, for background estimation, the subjet *b* tagging efficiency is determined using a loose set of selections which require events to have J_1 and J_2 passing only the soft-drop mass and τ_{21} requirements. The *b* tagging efficiency is obtained for the different subjet flavors and as a function of p_T and η. These efficiencies are shown in Fig. 4.11 using the DeepCSVM operating point. The efficiencies are estimated in events with an average pileup of 200. The *b* tag efficiency maps for zero and 200 pileup for both DeepCSV loose and medium are given in Appendix A.3 in

Figs. A.24 and A.23. The final results of this analysis focus on the 200 pileup scenario, for which the results are presented here.
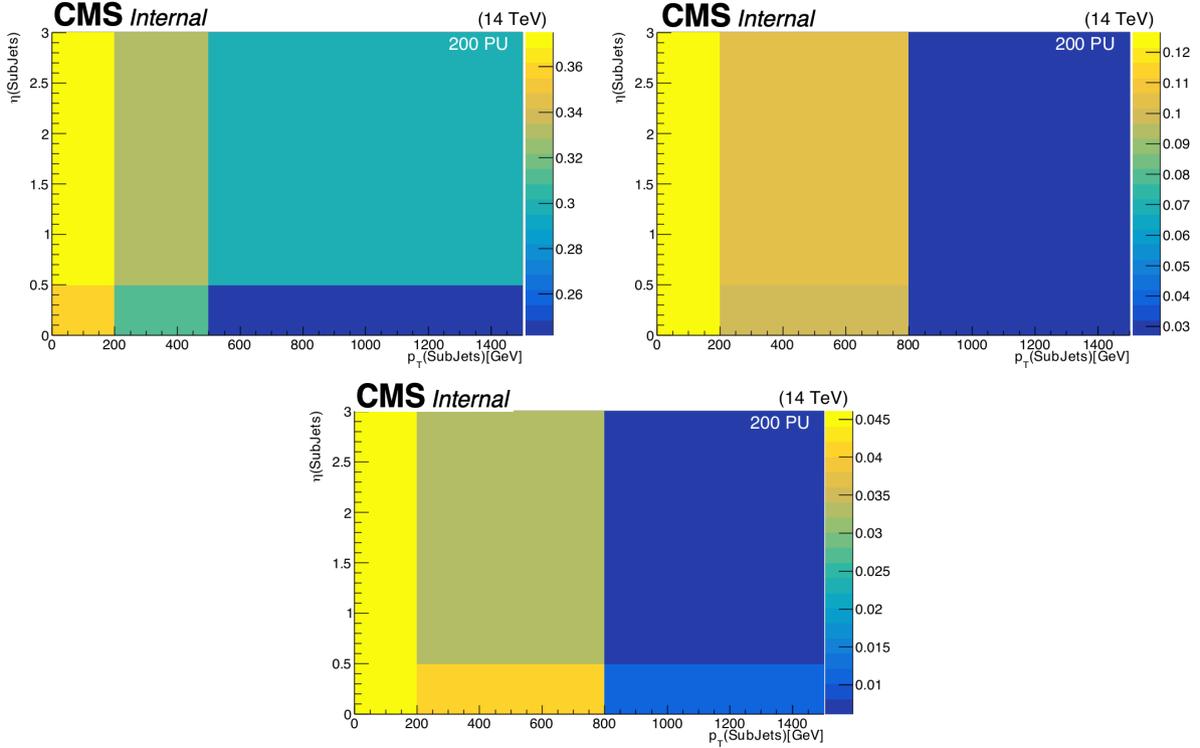
Figure 4.11: The b tagging efficiency for subjets from a b quark (upper left), c quark (upper right), and light quark or a gluon (lower) in events with an average pileup of 200. These figures are using the DeepCSVM operating point.

This b tagging rate is then applied to the multijet events passing all selections except the subjet b tagging one. Based on the subjet flavors, the events are assigned a probability of passing the $3b$ or the $4b$ categories. The m_{JJ} distributions of the estimated backgrounds and the signals are shown in Fig. 4.12.

2016 data analysis [153] shows the ratio of the QCD multijet yields in the data to that in the simulations is about 0.7, i.e., the simulations overestimate the multijet backgrounds by a factor of 1.43. Hence, in the background estimation, we have scaled the QCD multijet backgrounds by 0.7 to attain a more realistic background yield. The event yields after full selection are shown in Table 4.10.
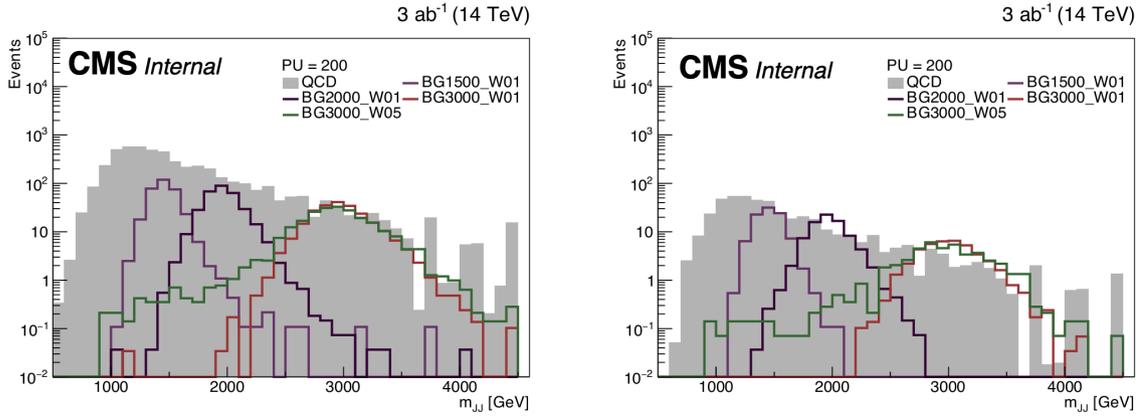
Figure 4.12: The estimated multijets background and the signal m_{JJ} distributions for the $3b$ (left) and $4b$ (right) subjet b tagging categories for bulk gravitons (BG) of masses 1.5, 2.0 and 3.0 TeV with 1% (and 5% for 3.0 TeV) widths, assuming a signal cross section of 1 fb. These figures are using the DeepCSVM operating point and PU=200.

Table 4.10: Event yields and efficiencies for the signal and multijet background. The product of the cross sections and branching fractions of the signals $\sigma(pp \to X \to HH \to b\bar{b}b\bar{b})$ is assumed to be 1 fb.

	PU = 200			
Process	Category $3b$		Category $4b$	
	Events	Efficiency (%)	Events	Efficiency (%)
Multijets	4755	1.6×10^{-3}	438	1.5×10^{-4}
BG (M=1.5 TeV, width = 1%)	326	11	95.2	3.2
BG (M=2.0 TeV, width = 1%)	316	11	81.2	2.7
BG (M=3.0 TeV, width = 1%)	231	7.7	41.4	1.4
BG (M=3.0 TeV, width = 5%)	233	8	45.8	1.5

4.3.5   Results

The uncertainty on the product of the signal selection efficiency and fiducial acceptances are as follows: H jet selection efficiency because of the $H \to b\bar{b}$ parton shower modeling (10%), selection on the N-subjettiness τ_2/τ_1 (12.5%), H jet mass scale (2%), luminosity measurement (2.5%), and mismatch in the pileup profile in the simulation with respect to the real beam conditions (3%). All systematic uncertainties are projections based on the analysis performed using 2016 data. The currently implemented systematic uncertainties are given in Table 4.11.

Table 4.11: List of systematic uncertainties. All systematics are applicable to the signal samples. The multjets background is "data-driven" and hence do not have these uncertainties. The uncertainties obtained from Run-2 4b analysis are scaled by half.

Source	Value	Obtained from
H jet mass scale and resolution	1%	Run-2, scaled by 0.5
H jet τ_{21} selection	13%	Run-2, scaled by 0.5
H-tagging correction factor	3.5%	Run-2, scaled by 0.5
Pileup modelling	1%	Run-2, scaled by 0.5
PDF and scales	1%	Run-2, scaled by 0.5
Luminosity	1.5%	CMS recommended
Jet energy scale	1%	CMS recommended
b tagging	1%	CMS recommended

The expected signal significance of a bulk graviton of mass between 1500 and 3000 GeV, produced through vector-boson fusion, and decaying into a pair of Higgs bosons, each of which decays to a $b\bar{b}$ pair, is given in Fig. 4.13 for an integrated luminosity of 3 ab^{-1}. With this analysis strategy, the results project a possible observation of 3 TeV bulk KK graviton assuming signal cross section 1 fb for all mass hypotheses. Results are made public for 1% width only.
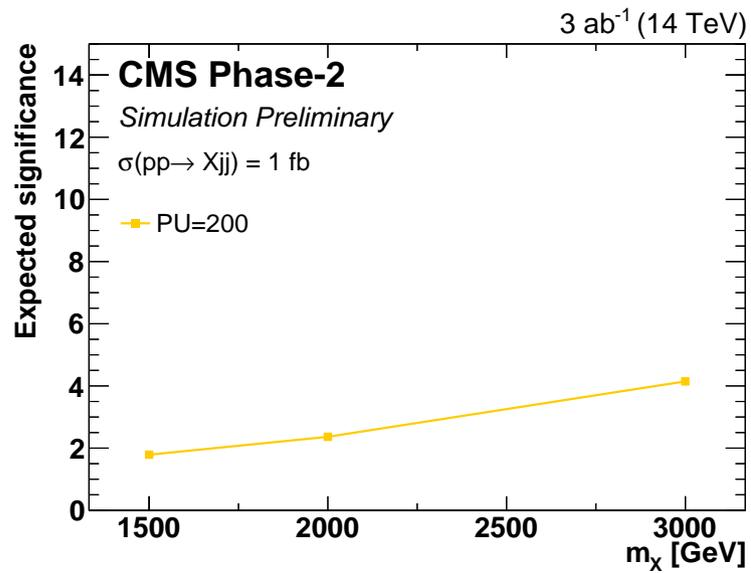
Figure 4.13: The expected signal significance for three mass points of the bulk graviton with 1% width assuming a production cross section of 1 fb and for PU=200.

4.4   Non-resonant di-Higgs production

This work is an extension to the previous Section 4.3 to study the non-resonant di-Higgs production in the boosted regime. The study of non-resonant Higgs boson pair production (HH) is one of the main goals of the scientific program of the CMS experiment at the HL-LHC. It gives direct experimental access to the Higgs boson self-coupling λ_{HHH} and hence to the structure of the Higgs scalar potential itself and the origin of the mass of the Higgs boson.

The experimental program of the CMS experiment involves several decay channels of the HH system. One of the experimentally most interesting decay channels is $HH \rightarrow b\bar{b}b\bar{b}$, which holds the most significant branching fraction among all the decay channels (33%). Searches made by the ATLAS and CMS collaborations with the data collected at $\sqrt{s} = 8$ and 13 TeV have proven that the $b\bar{b}b\bar{b}$ channel is indeed one of the most sensitive and a vital part of the program of study of HH production. The HL-LHC not only constitutes an unprecedented opportunity to study HH production in this decay channel but also represents a formidable challenge because of the high instantaneous luminosity and number of simultaneous pp interactions, or pileup, that will range between 140–200 or beyond. An essential program of the CMS detector upgrades is planned to mitigate the effects of the pileup and improve the performance in the trigger and reconstruction of $b\bar{b}b\bar{b}$ events.

This work aims to estimate the sensitivity to non-resonant HH production in the $b\bar{b}b\bar{b}$ final state in the luminosity and pileup conditions of the HL-LHC within a boosted regime. Boosted object reconstruction has been proven to be highly important in exploring the scalar sector of the standard model and physics beyond it. While the non-resonant HH production in the SM is typically associated with low p_T Higgs bosons and hence well separated b jets, BSM physics may manifest in the existence of new contact interactions, such as gHH (c_g), $ggHH$ (c_{2g}), and $ttHH$ (c_2), that mostly enhance the cross section in the high m_{HH} region, above 1 TeV. The $b\bar{b}b\bar{b}$ final state is one of the most sensitive in this exploration because of the reduced background contamination at high m_{HH}.

A study is carried out using the methods and tools developed for the prospect search for VBF production of resonances that decay to $b\bar{b}b\bar{b}$ in Section 4.3. The selection of events is identical to the resonant VBF analysis apart from the VBF jet selections. The analysis is run on the same set of backgrounds, while the signal is from a fast parametric simulation with the DELPHES software [156]. The behavior of the upgraded detector has been carefully studied, and the performance in the jet energy determination and b jet identification and discrimination from gluon and light flavor quark jets has been parametrized to generate samples with a DELPHES fast simulation.

The signal models are considered with the SM coupling of the Higgs boson and twelve other sets of values for the coupling parameters known as EFT shape benchmarks as we studied in Section 1.3.1, given in Table 4.12. These parameters are chosen based on representative kinematic features in the HH production process [16]. Signals corresponding to the SM and shape benchmark number 2 were generated. The other shape benchmarks, i.e., 1 and 3–12, were obtained by reweighting the events of shape benchmark 2 based on the values of m_{HH} and $\cos\theta^*_{HH}$ [16,62]. 95% upper limits of HH production cross section are set for SM and BSM shape benchmarks.

Table 4.12: Parameter values of the couplings corresponding to the twelve shape benchmarks and the SM prediction.

Shape benchmark	κ_λ	κ_t	c_2	c_g	c_{2g}
1	7.5	1.0	-1.0	0.0	0.0
2	1.0	1.0	0.5	-0.8	0.6
3	1.0	1.0	-1.5	0.0	-0.8
4	-3.5	1.5	-3.0	0.0	0.0
5	1.0	1.0	0.0	0.8	-1.0
6	2.4	1.0	0.0	0.2	-0.2
7	5.0	1.0	0.0	0.2	-0.2
8	15.0	1.0	0.0	-1.0	1.0
9	1.0	1.0	1.0	-0.6	0.6
10	10.0	1.5	-1.0	0.0	0.0
11	2.4	1.0	0.0	1.0	-1.0
12	15.0	1.0	1.0	0.0	0.0
SM	1.0	1.0	0.0	0.0	0.0

This section has three subsections. Section 4.4.1 explains MC reweighting method

for the BSM shape benchmarks. Section 4.4.2 explains the analysis strategy which is motivated from Section 4.3. The results are discussed in Section 4.4.3.

4.4.1   Signal MC reweighting

The analytic formula used to calculate cross sections for any point of the parameter space (κ_λ, κ_t, c_2, c_g, c_{2g}) is reported in Ref. [62]. A script implementing this calculation can be found here [157]. In this section we explain how the reweighting has been implemented to reweight the simulated MC signal samples to any point in (κ_λ, κ_t, c_2, c_g, c_{2g}) representation. We are considering a LO $2 \rightarrow 2$ process. The two Higgs bosons are produced with identical transverse momenta, and they are back-to-back before parton shower. The final state can then be entirely defined by three kinematic variables (ignoring the irrelevant azimuthal angle of emission of the bosons). Furthermore, one of these three variables can be used to isolate all the information related to the PDF of the colliding partons, which is also irrelevant to the physics of the production process once one focuses on a specific initial state (the gluon-gluon fusion process). The magnitude of the boost of the center-of-mass frame is taken as the variable factorizing out the PDF modeling.

The two remaining variables, which provide direct information on the physics of ggHH production, can be the invariant mass of the HH system (m_{HH}) and the modulus of the cosine of the polar angle of one Higgs boson with respect to the beam axis ($\cos\theta^*_{HH}$) in the rest from of di-Higgs system. Since we are using parton-level information, this last variable is equivalent to the polar angle in the Collins-Soper frame [158], which is commonly used in experimental analysis. The Collins-Soper frame has the z-axis bisecting the beam and target momentum vectors. The x-z plane is formed by four vectors of two interacting protons. The two Higgs boson remain back to back in this frame. The variables m_{HH} and $\cos\theta^*_{HH}$ can thus be used to characterize the final state kinematics produced entirely by different choices of the value of anomalous Higgs boson (self-) coupling parameters.

By construction, the simulated signal samples listed in Table 4.13 are good representatives of the kinematic phase space. Therefore using the generator level m_{HH}

and $\cos\theta_{HH}^*$ we can reweight the generated MC samples to any other point in (κ_λ, κ_t, c_2, c_g, c_{2g}) parameter space. Using this and the analytic formula parametrizing the cross section of any point of the parameter space (κ_λ, κ_t, c_2, c_g, c_{2g}) from Ref. [62] we reweighted the mixture of all generated signal MC described above to match the clustered benchmark points described in Ref. [16] and to the SM samples as well. Fig. 4.14 shows the distribution of the Higgs boson pair mass at the generator level for 12 clustered benchmarks. One can see that the distributions at the generator level obtained with the reweighting procedure correspond to the distributions from Ref. [16].
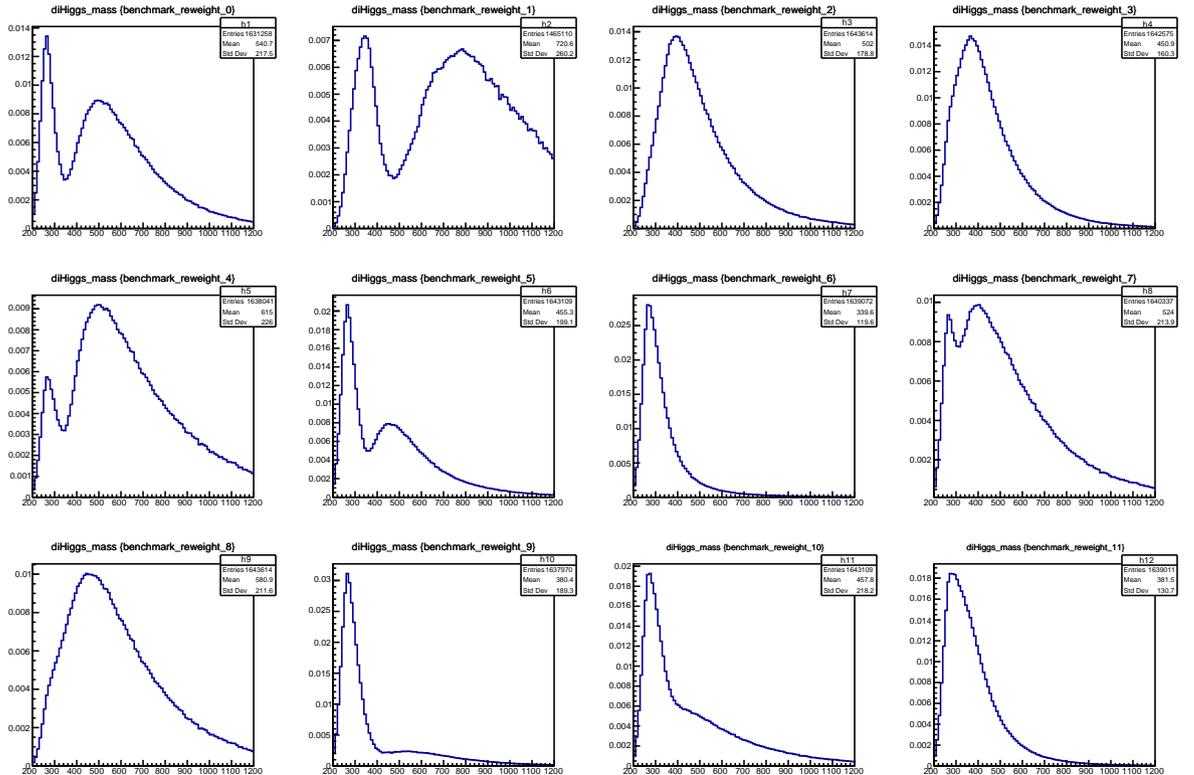
Figure 4.14: Distributions of the di-Higgs invariant mass m_{HH} at the generator level for 12 clustered BSM benchmarks points.

4.4.2 Analysis strategy

The analysis looks for two large-area jets and reconstructs the m_{HH} spectrum from their invariant mass, m_{JJ}. Substructure techniques are used to identify subjets originating

from b quarks and discriminate from jets from other flavors, i.e., the charm, light quarks, and gluon induced jets. The b tagging and misidentification rates for large area jets are parametrized using the fully simulated samples from resonant VBF analysis and applied in the DELPHES simulation. The selection criteria for the boosted $H \to b\bar{b}$ jets is given in Table 4.3. Events with two boosted $H \to b\bar{b}$ are selected. No additional selection criteria from resonant VBF analysis are imposed to keep the selection as general as possible for different BSM effects on the final state.

Two event categories defined respectively as containing 3 or 4 b tagged subjets are used to look for the presence of a signal in the m_{JJ} distribution, as shown in Fig. 4.15. The longer tail in the m_{JJ} distributions of the shape benchmark number 2 is clearly visible, wherein lies the sensitivity of this analysis over the fully-resolved regime containing four separate b quark jets.

The resonant VBF analysis was developed for $m_{JJ} \geq 1$ TeV and hence used a dedicated QCD sample with a selection on the invariant mass of the two leading-p_T partons: $m_{qq} \geq 1$ TeV. Thus its contribution is lacking in the lower m_{JJ} region. To complete the background description for lower m_{JJ}, we use the QCD HT-binned[3] background samples listed in Table 4.13. Events from these samples are required to satisfy the requirement $m_{qq} < 1$ TeV. This selection makes the phase space of the HT-binned samples orthogonal to that of the high m_{qq} QCD samples. The HT-binned samples are b-enriched at the matrix element level and lack significant contribution from light flavor jets. However, it was checked that the contribution from the light jets background was less than 4%, that is why they are not included.

4.4.3   Results

After the entire event selections, the selection efficiency for the SM signal is 0.1%, while it is 1.8% for the shape benchmark number 2. The values of the $S/\sqrt{B}$ is 0.025 for the SM signal and 0.52 for the shape benchmark number 2, assuming the same production cross section (for both signals) with a new background modeling strategy. For the BSM

---

[3]HT-binning refers to different range of scalar sum of transverse energy of decay products.

Table 4.13: Simulated signal and background samples

Physics process	GEN dataset	N. evts.
Signal		
$gg \to HH \to b\bar{b}b\bar{b}$ (SM)	GluGluToHHTo4B_node_SM_14TeV-madgraph	500000
$gg \to HH \to b\bar{b}b\bar{b}$ (BSM node 2)	GluGluToHHTo4B_node_2_14TeV-madgraph	500000
Backgrounds		
QCD b enr., $200 < H_T < 300$ GeV	QCD_bEnriched_HT200to300_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	10704113
QCD b enr., $300 < H_T < 500$ GeV	QCD_bEnriched_HT300to500_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	9726551
QCD b enr., $500 < H_T < 700$ GeV	QCD_bEnriched_HT500to700_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	9819938
QCD b enr., $700 < H_T < 1000$ GeV	QCD_bEnriched_HT700to1000_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	7719024
QCD b enr., $1000 < H_T < 1500$ GeV	QCD_bEnriched_HT1000to1500_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	6860134
QCD b enr., $1500 < H_T < 2000$ GeV	QCD_bEnriched_HT1500to2000_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	9537208
QCD b enr., $H_T > 2000$ GeV	QCD_bEnriched_HT2000toInf_TuneCUETP8M1_14TeV-madgraphMLM-pythia8	9932246

shape benchmark 2, the associated upper limits on cross section varies depending on the specific combination of EFT couplings. However, the analysis is sensitive to parts of this phase space and can exclude some values of anomalous couplings. The systematic uncertainties are similar to resonant VBF analysis given in Table 4.11.

The event yields for the signals and the background are given in Table 4.14. With the event yields and distributions shown in Fig. 4.15 we calculate the 95% confidence level (CL) upper limits on the non-resonant HH productions in the SM, and for other BSM couplings (nodes 1–12), as shown in Fig. 4.16. For this, the events from $3b$ and $4b$ categories are combined. The plot shows the highest expected sensitivity for shape benchmark 2.

(a) 3 b tag

(b) 4 b tag
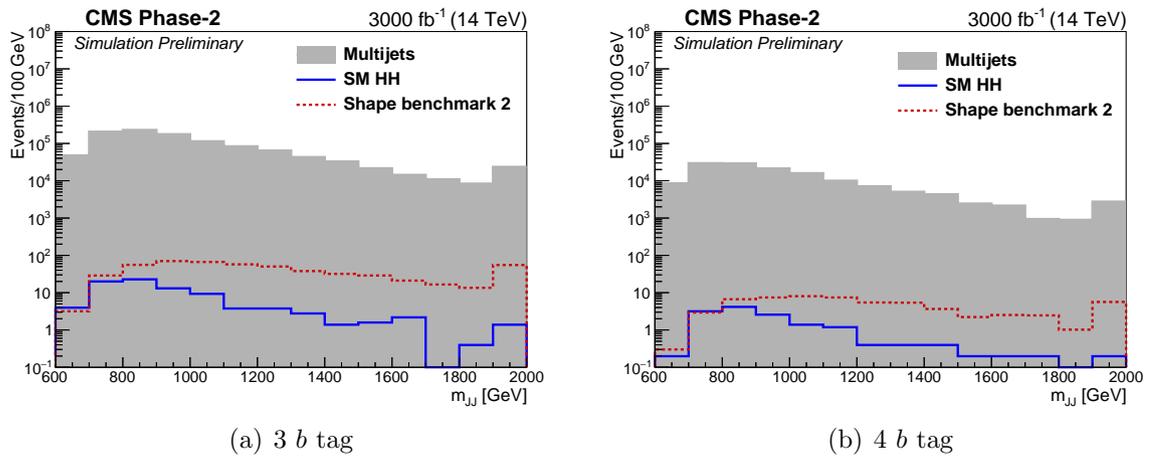
Figure 4.15: Distribution of the two large radius jet invariant mass, m_{JJ}, in the 3 (a) and 4 (b) b tag regions.

Table 4.14: Yields and significance for all 13 shape benchmarks with $3b$ and $4b$ categories as well as for the background. The cross section is assumed to be 10 fb for all the nodes, including the SM.

Shape benchmark	yields(3b)	yields(4b)	$S/\sqrt{B}(3b)$	$S/\sqrt{B}(4b)$
SM	26	4	0.03	0.01
1	136	17	0.13	0.05
2	537	61	0.52	0.16
3	85	11	0.08	0.03
4	44	4	0.04	0.01
5	261	31	0.25	0.08
6	80	10	0.08	0.03
7	14	2	0.01	0.01
8	139	17	0.13	0.05
9	191	23	0.18	0.06
10	50	6	0.05	0.02
11	108	13	0.10	0.03
12	16	2	0.02	0.01
QCD	1079041	140526		
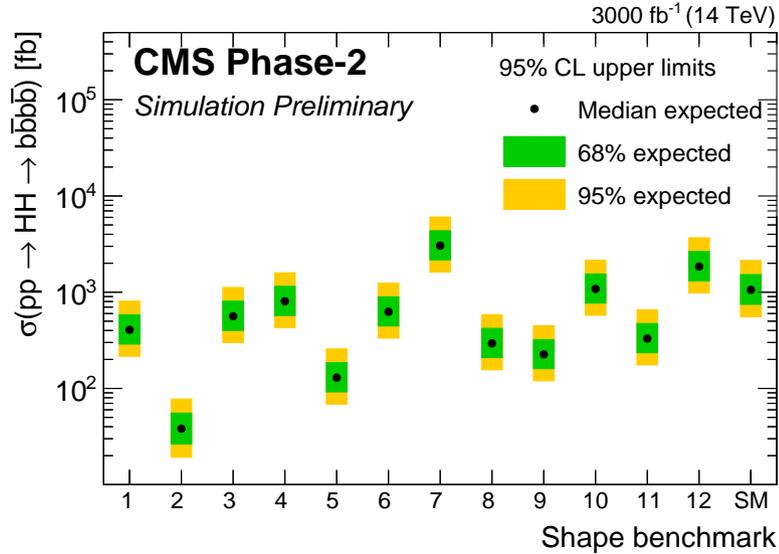
Figure 4.16: The expected upper limits for non-resonant HH production in the standard model and other shape benchmarks (1–12). The inner (green) and the outer (yellow) bands indicate the regions containing the 68 and 95% deviation, respectively, of the distribution of limits expected under the background-only hypothesis.

Chapter 5

Searches for resonant Higgs pair production in $b\bar{b}\gamma\gamma$ final state

This analysis presents searches for resonant Higgs boson pair production in the final state with two bottom quarks and two photons ($b\bar{b}\gamma\gamma$) in proton-proton collisions at 13 TeV center-of-mass energy using LHC Run-2 data collected by the CMS detector with 138 fb^{-1} total integrated luminosity. A model-independent approach with narrow-width approximation in a warped extra dimension (WED) and next-to-minimal supersymmetric standard model (NMSSM) motivated scenarios is explored. It performs a direct search for $pp \to X \to YH \to b\bar{b}\gamma\gamma$ process where X can be a spin-0 or spin-2 particle, H is the SM Higgs boson, and Y may be a H or another spin-0 boson.

As described in Section 1.3.2, the WED model predicts the existence of small and compactified extra dimensions along which gravity particles propagate. The searches are performed in the RS_bulk scenario, where matter particles are also allowed to propagate along the extra dimensions. The model provides an initial solution to the SM hierarchy problem. In the RS_bulk scenario of the model, the coupling of resonances with the top quark and SM bosons is enhanced. It predicts bulk radion (spin-0) and bulk KK graviton (spin-2) resonances (X) with significantly larger branching fraction ($\mathcal{O}(10\%)$) to decay into the pair of the SM Higgs bosons (H).

The other BSM scenario is motivated by SUSY as studied in Section 1.3.3. The

NMSSM is a further extension to MSSM, which generates an electroweak scale from the SUSY-breaking scale providing a solution to well-known MSSM μ-problem. NMSSM also enriches the Higgs sector by extending it up to seven Higgs bosons (three CP-even, two CP-odd, and two charged Higgs bosons).

We concentrate on gluon-gluon fusion production of X (spin-0) resonance, i.e., a heavy Higgs decaying to an SM-like Higgs boson H and a Y spin-0 Higgs boson with the largest singlet component. Because of the dominant singlet nature, couplings of Y with matter fields are suppressed. In such a case, the $X \to YH$ process becomes essential for Y production. However, the branching fraction of Y is considered as same as H. This happens due to mixing of singlet and doublet fields which make couplings of Y to quarks, leptons and gauge fields proportional to H giving the same branching fraction for Y. This statement does not hold when we tend to heavy Y masses.

Figure 5.1 presents the Feynman diagram for the physics process. With this channel, the analysis benefits from 1) $H \to \gamma\gamma$ handle, having high purity and selection efficiency because of excellent resolution of CMS electromagnetic calorimeter, and 2) $H \to b\bar{b}$ handle, having the large branching fraction among Higgs decay modes. The $HH \to b\bar{b}\gamma\gamma$ process yields 0.26% of the total branching fraction of HH-decay. For the NMSSM process, by choice, it is set for H to decay into $\gamma\gamma$ final state to preserve the benefits of $H \to \gamma\gamma$ handle.

The ATLAS and CMS have published results in the same final state for $X \to HH$ searches using 2016 Run-2 LHC data [64, 159]. ATLAS explored mass range $m_X =$[260, 1000] GeV and set 0.4-2.4 fb observed upper limits on product of resonant production cross section and branching ratio. On the other side, CMS studied mass range $m_X =$[260, 900] GeV and set 0.3-4 fb observed upper limits on product of resonant production cross section and branching ratio. Both the results are comparable. In this work, we have used newly developed object identification algorithms and machine learning tools that efficiently reject background contamination. **Using this analysis method, we produce 2016 only expected results and get 6-25% improvement compared to the previously published CMS expected results. For resonance**

mass up to 1 TeV and with 2016, 2017 and 2018 data, ATLAS has made results public for $X \rightarrow HH \rightarrow b\bar{b}\gamma\gamma$ searches which set 1.6–0.12 (0.95–0.11) fb observed (expected) upper limits on cross section [6] while with our analysis we set 0.82–0.07 (0.74–0.08) fb observed (expected) upper limits on cross section. Our analysis gives $\sim$30–40% better results. More details about this comparison study are added in Appendix B.5. Another main highlight of this analysis is $X \rightarrow YH \rightarrow b\bar{b}\gamma\gamma$ searches explored in the CMS for the first time using LHC data.
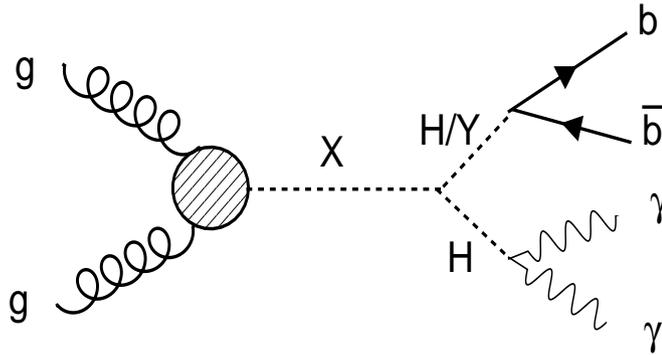
Figure 5.1: Feynman diagram for BSM X resonance production at LHC via gluon-gluon fusion; X further decays to two scalars (both SM Higgs (H) or a combination of SM Higgs (H) and NMSSM Higgs (Y)) with a $b\bar{b}\gamma\gamma$ final state

This chapter is organized in the following way: The details related to data and simulations have been added in Section 5.1. Section 5.2 briefly explains all the analysis steps described in upcoming sections. Event reconstruction and selection with background rejection methods have been thoroughly discussed in Section 5.3. Signal and background modeling have been described in Section 5.4 and 5.5, respectively. Section 5.6 details the systematic uncertainties. In the end, results are presented in Section 5.7.

5.1   Data and simulation samples

The data used for the analysis were collected by the CMS detector, during the Run-2 period comprising 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}), and 2018 (59.8 fb^{-1}) data-taking years with total integrated luminosity 138 fb^{-1} and 13 TeV center-of-mass energy. For data collection, the E/gamma triggers requiring two isolated energy deposits in the electromagnetic calorimeter are used. The trigger details are briefly explained in Section 5.3.1.

For the signal process, we study the resonant production from gluon-gluon fusion. The resonance further decays into two scalars, either a pair of the SM Higgs boson or a combination of the SM-like Higgs boson and NMSSM Y boson. Considering both the BSM models and corresponding BSM couplings, samples have been generated using MADGRAPH5_AMC@NLO, version 2.2.2 (2016)/2.4.2 (2017, 2018) (for WED) and 2.6.5 (for NMSSM) [160] for resonance (X) mass range up to 1000 GeV.

For WED, the complete list of the signal samples is available in Table 5.1. For NMSSM, various Y mass points have been considered depending upon the allowed phase space with the mass of X resonance. Figure 5.2 shows the (X, Y) mass grids, which are studied for the NMSSM scenario. Table 5.2 contains the list of NMSSM signal samples. We start with 90 GeV mass for Y and 300 GeV mass for X to get rid of the turn-on issue in m_{jj} (around 70 GeV) and $\tilde{M}_X$ (=$m_{jj\gamma\gamma}$- $m_{\gamma\gamma}$- m_{jj} + m_H + m_Y) distributions (Fig. 5.37).

For background simulations, two types of background processes are considered: resonant (single Higgs production) and non-resonant (diphoton QCD processes), where resonant and non-resonant refer to the distribution of the diphoton invariant mass of the background processes.

Non-resonant backgrounds are estimated in a data-driven way, while we rely on MC simulations for single Higgs backgrounds. MC simulation for the non-resonant background is used only in the multivariate discriminant, optimization, and validation studies of the analysis. The main non-resonant backgrounds for the $b\bar{b}\gamma\gamma$ final state come from the SM multijet processes, with well-isolated photons coming from the hard scatter

Table 5.1: Signal samples for WED resonances bulk radion and bulk KK graviton

Signal	Dataset
Radion 2016	GluGluToRadionToHHTo2B2G_M-{mass}_narrow_13TeV-madgraph, mass = [260, 900] GeV
	GluGluToRadionToHHTo2B2G_M-1000_narrow_TuneCUETP8M1_13TeV-madgraph-pythia8
2017	GluGluToRadionToHHTo2B2G_M-{mass}_narrow_13TeV-madgraph_correctedcfg, mass = [260, 1000] GeV
2018	GluGluToRadionToHHTo2B2G_M-{mass}_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8, mass = [260, 1000] GeV
Graviton 2016	GluGluToBulkGravitonToHHTo2B2G_M-{mass}_narrow_13TeV-madgraph, mass = [260, 1000] GeV
2017	GluGluToBulkGravitonToHHTo2B2G_M-{mass}_narrow_13TeV-madgraph_correctedcfg, mass = [260, 1000] GeV
	GluGluToBulkGravitonToHHTo2B2G_M-320_narrow_TuneCP5_13TeV-madgraph-pythia8_correctedcfg
2018	GluGluToBulkGravitonToHHTo2B2G_M-{mass}_narrow_TuneCP5_PSWeights_13TeV-madgraph-pythia8, mass = [260, 1000] GeV

Table 5.2: Signal samples for NMSSM X resonance

Signal	Dataset
2016/17/18 NMSSM	NMSSM_XToYHTo2b2g_MX-{massX}_Tunetune_13TeV-madgraph-pythia8, massX = [300, 1000] GeV, massY = [90,800] GeV, tune={2016: CUETP8M1, 2017/18: CP5}

(prompt photons) or jets misidentified as photons (fake photons). The prompt-prompt background events are simulated with the SHERPA generator [161]; it includes the Born processes with up to 3 additional jets at LO accuracy and the box processes at LO. We studied other non-resonant background contributions and found them less than 1% after event selections. Therefore, we neglect them.

For resonant background, we consider all possible Higgs production mechanisms: gluon-gluon fusion (ggH), vector-boson fusion (VBF H), associated production with a vector boson (VH), top quarks (ttH), and b quarks ($b\bar{b}H$) where Higgs decays into two photons. The corresponding cross sections and decay branching ratios are taken from Ref. [162].

The background samples are listed in Table 5.3 and 5.4.

Table 5.3: Non-resonant background samples

Background (non-resonant)	Dataset
2016/17/18	DiPhotonJetsBox_MGG-80toInf_13TeV-Sherpa GJet_Pt-20to40_DoubleEMEnriched_MGG-80toInf_Tune{tune}_13TeV_Pythia8 GJet_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_Tune{tune}_13TeV_Pythia8 tune={2016: CUETP8M1, 2017/18: CP5}
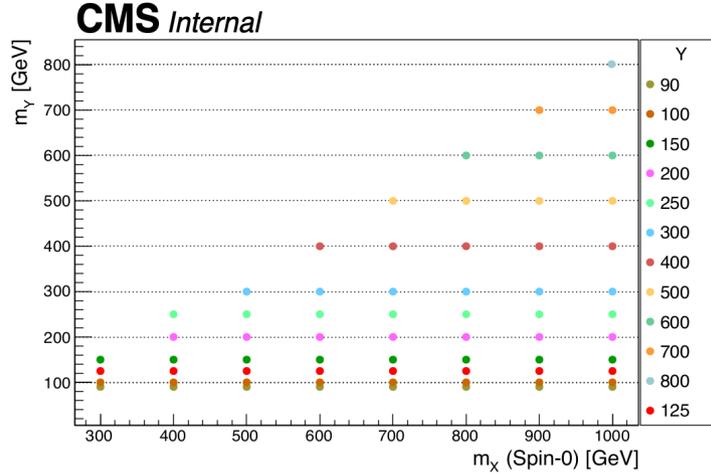
Figure 5.2: $X - Y$ mass point grid for NMSSM simulations

Table 5.4: List of the resonant background samples

Resonant background	Dataset
2016	GluGluHToGG_M-125_13TeV_powheg_pythia8
	VHToGG_M125_13TeV_amcatnloFXFX_madspin_pythia8
	VBFHToGG_M-125_13TeV_powheg_pythia8
	ttHToGG_M125_13TeV_powheg_pythia8_v2
	bbHToGG_M-125_4FS_yb2_13TeV_amcatnlo
	bbHToGG_M-125_4FS_ybyt_13TeV_amcatnlo
2017	GluGluHToGG_M-125_13TeV_powheg_pythia8
	VHToGG_M125_13TeV_amcatnloFXFX_madspin_pythia8
	VBFHToGG_M-125_13TeV_powheg_pythia8
	ttHToGG_M125_13TeV_powheg_pythia8
	bbHToGG_M-125_4FS_yb2_13TeV_amcatnlo
	bbHToGG_M-125_4FS_ybyt_TuneCP5_13TeV-amcatnlo-pythia8
2018	GluGluHToGG_M-125_13TeV_powheg_pythia8
	VHToGG_M125_13TeV_amcatnloFXFX_madspin_pythia8
	VBFHToGG_M-125_13TeV_powheg_pythia8
	ttHToGG_M125_TuneCP5_PSWeights_13TeV_powheg_pythia8
	bbHToGG_M-125_4FS_ybyt_TuneCP5_13TeV-amcatnlo-pythia8
	bbHToGG_M-125_4FS_yb2_TuneCP5-13TeV-amcatnlo-pythia8

All simulated samples are interfaced with PYTHIA8 for parton showering and fragmentation with the standard p_T-ordered parton shower (PS) scheme. The underlying event is modeled with PYTHIA8, using the CUETP8M1 (CP5) tune for 2016 (2017, 2018) [163,164]. PDFs are taken from the NNPDF3.0 NLO (2016) and NNPDF3.1 NLO (2017, 2018) set [165–170]. The detector response is modeled using GEANT4 [117] toolkit. The simulated events include similar additional proton-proton interactions (pileup) as observed in the data.

5.2   Analysis strategy

A basic flowchart for analysis strategy is given in Fig. 5.3. The upcoming sections explain each step of this flowchart. Our study follows the approach of the di-Higgs searches in the same final state, made public by CMS [1, 64] with Run-2 data[1]. Most of the developments and techniques have been directly used from these analyses after a dedicated study. Events are selected using similar online and offline selections. The final state physics objects, the photons and b jets, are initially reconstructed and selected using the loose object preselections. The diphoton and dijet systems are reconstructed using event preselections from these photons and b jets. The analysis strategy then focuses on rejecting backgrounds and extracting signal by fitting $m_{\gamma\gamma}$ and m_{jj} observables in a 2D mass plane within an optimized $\tilde{M}_X$ mass window.

For non-resonant background rejections, a multiclass BDT classifier is trained on signal and background simulations which is explained in Section 5.3.5. The training inputs are the same as given in Ref. [1]. We also studied additional inputs to improve the training performance as described in Appendix B.2, but overall improvement was less than 5%. The event categorization is performed within MVA categories optimized using the simulations. An NN-based ttHkiller discriminant (developed by CMS collaboration) is used further to reject dominating ttH single Higgs contribution (Section 5.3.6).

The signal is extracted from a simultaneous two-dimensional (2D) fit to the invariant mass of diphoton system $m_{\gamma\gamma}$ and dijet system m_{jj} in all categories. This method benefits from different shapes of $m_{\gamma\gamma}$ and m_{jj} distributions for signal and data. The signal keeps peaking $m_{\gamma\gamma}$ and m_{jj} distributions while background events from data show falling distributions of these observables. Also both observables have negligible correlation. This fit is performed within the optimized $\tilde{M}_X$ window, which also suppresses resonant background contribution. The impact of systematic uncertainties on final results is around 1–2% on top of the dominated statistical uncertainty. The results are presented as 95% CL exclusion limit on the product of resonance production cross section and

---

[1]This analysis uses the standard tools and frameworks from $H \rightarrow \gamma\gamma$ searches. We contributed for non-resonant $HH \rightarrow b\bar{b}\gamma\gamma$ analysis [1] and further extended study to resonant searches for thesis work.

Online:

Event passing

diphoton triggers

Non-resonant

backgrounds:

MVA Classifier

MVA categorization

Signal:

$m_{\gamma\gamma}$:Sum of Gaussians

m_{jj}: Double Sided Crystal ball /

Crystal ball + Gaussian

Event selections

Analysis optimization methods

Statistical Analysis

Signal Extraction ($m_{\gamma\gamma}$:m_{jj}) fit

Offline:

select photon-pair

and b jet pair

Resonant

backgrounds:

Selections on NN-based

discriminant and $\tilde{M}_X$

window

Data driven Backgrounds:

2D envelope method

Resonant MC backgrounds:

$m_{\gamma\gamma}$:Sum of Gaussians,

m_{jj}: varies with production process

Figure 5.3: Analysis flowchart

branching fraction. The limits are also compared with their appropriate BSM predictions to exclude allowed mass parameters.

5.3   Event reconstruction and selection

The analysis objects are initially reconstructed for trigger level selections. After that, analysis-specific offline selections are applied to select events with two photons and two b jets. We further make the event selection on machine learning-based discriminant and resonance mass windows to suppress non-resonant and resonant background contributions. The events which pass all the selection requirements are used for the final results. Let us go through all these studies one by one with upcoming subsections.

5.3.1   Online event selections

Online events are selected using standard $H \to \gamma\gamma$ double photon triggers in which both photons are required to be isolated with some minimum p_T threshold with more than 90 GeV diphoton invariant mass. The trigger paths for three data-taking years are as

follows:

2016: $HLT\_Diphoton30\_18\_R9Id\_OR\_IsoCaloId\_AND\_HE\_R9Id\_Mass90$

2017: $HLT\_Diphoton30\_22\_R9Id\_OR\_IsoCaloId\_AND\_HE\_R9Id\_Mass90$

2018: $HLT\_Diphoton30\_22\_R9Id\_OR\_IsoCaloId\_AND\_HE\_R9Id\_Mass90$

The p_T threshold was changed from 18 GeV (2016) to 22 GeV (2017/18) to keep sustainable HLT bandwidth for subleading photons. Selections on photon isolation variables or R_9 variable maintains the high trigger efficiency. More details related to the L1-trigger and HLT algorithm can be found in Section 3.2.2. Trigger weights and uncertainty associated with them are also considered and explained in Section 5.6. The trigger selections are thoroughly described in Ref. [129] for $H \to \gamma\gamma$ analysis.

5.3.2   Photon reconstruction

Photons are reconstructed with some preselection requirements and using ECAL energy deposits. The preselections are taken from $H \to \gamma\gamma$ analysis [129] and applied on the globally reconstructed particle-flow (PF) algorithm-based photon candidates. The ECAL energy scale is corrected in data using $Z \to ee$ simulation and multivariate regression, and photon energy is smeared within simulated events to match the resolution with data as discussed in Section 3.5.3. The photon multivariate identification method has been discussed in Section 3.5.4. Within this, the loose selections are applied on shower shapes (R_9), photon identification variables (H/E), kinematics, and isolation variables (CHI) which are slightly tighter than the trigger ones to improve Data/MC discrepancies. This multivariate identification method, trained on photon shower-shape and kinematic variables, discriminates the signal photons from background photons. A very loose cut on photonID (output variable from this multivariate method) with 99% efficiency is also applied. This is further used as input for the BDT classifier discussed in Section 5.3.5. Only events that have at least one diphoton candidate passing the preselections requirements are considered. Table 5.5 summarizes the photon preselections keeping more than 99% trigger efficiency for selected events. Scale factors are used to mitigate Data/MC discrepancies. The corresponding systematics are included in the

final list of systematics.

Table 5.5: Preselections for photon candidates

Requirements	Leading Photon	Subleading Photon
E_T	30 GeV	20 GeV
$|\eta|$	< 2.5 and outside 1.442 < $|\eta|$ < 1.566	
Shower shape and Isolation	$R_9 > 0.8$ or CHI < 20 GeV or CHI/E_T < 0.3	
Identification	H/E < 0.08 and MVAID> -0.9	

5.3.3   Jet reconstruction

Jets are reconstructed using anti-k_T algorithm with a distance parameter R=0.4 as discussed in Section 3.4.2; thus, we use AK4 jets. The preselections are directly used from non-resonant $HH \to b\bar{b}\gamma\gamma$ analysis [1]. Events, passing jet identification [171] criteria to reject jets from calorimeter noise, are required to have jets with $p_T > 25$ GeV and $|\eta| <$ 2.4 (2016) / 2.5 (2017, 2018) thresholds[2]. The $|\eta|$ selection is based on selecting jets within tracker coverage to benefit from track information to identify the jets originated from b quark hadronization.

5.3.4   Offline event selections

To identify events with diphoton object within ECAL and tracker coverage $|\eta| < 2.5$, they should pass selection on $p_T(\gamma)/m_{\gamma\gamma}$ ($p_T^{\gamma 1}/m_{\gamma\gamma} > 1/3$ and $p_T^{\gamma 2}/m_{\gamma\gamma} > 1/4$) and 100 $< m_{\gamma\gamma}$ < 180 GeV, where p_T^{γ} and $m_{\gamma\gamma}$ are transverse momenta and invariant mass of selected photon-pair, respectively. The asymmetric p_T^{γ} selection avoids the turn-on present in data $m_{\gamma\gamma}$ distribution. It also rejects the misidentified background photons. An electron veto (from Section 3.5.4.3) requirement is also applied not to select an electron as a photon. It suppresses dominant $Z \to ee$ background contribution. The primary pp vertex is identified using a diphoton vertex identification multivariate technique as given in Section 3.5.4.5. It has been proven 99.9% efficient for correct vertex assignment for

---

[2]As $H \to \gamma\gamma$ handle is crucial for analysis sensitivity, we do not compromise with $H \to \gamma\gamma$ selections. For a completeness we tried with $H \to b\bar{b}$ p_T selection of 20 GeV to gain in signal efficiency, especially for low masses, but it made the BDT training performance worse

$b\bar{b}\gamma\gamma$ final state because of the additional jet selection requirements explained within the next paragraph.

The reconstructed jets are corrected with p_T based b jet energy regression as described in Section 3.5.2. It improves the b jet energy resolution, resulting in a better m_{jj} resolution with 20% improvement. A $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.4$ selection is also applied to have selected photons and jets isolated from each other. For tagging the jets as b jets, deep neural network and secondary vertex algorithm-based tagger DEEPJET is used (discussed in Section 3.5.1). We do not apply any selection on the b discriminator. It is used as one of the inputs to train a multiclass BDT classifier. The corresponding reshaping scale factors and uncertainties are taken into account which are included in Section 5.6. Additionally, a wider $70 < m_{jj} < 190\ (1200)$ GeV selection is also applied. Within the m_{jj} mass window, a dijet system is reconstructed using two jets with the highest sum of DEEPJET discriminator score in an event. The m_{jj} window is kept wider for the NMSSM signals to scan over the allowed Y mass points. All the offline event selections are tabulated in Table 5.6. To use optimal b tagger in the analysis,

Table 5.6: Event selection criteria

Photons		Jets	
Variable	Selection	Variable	Selection
photon p_T	$> m_{\gamma\gamma}/3,\ m_{\gamma\gamma}/4$	p_T	$> 25.$ [GeV]
electron veto		$|\eta|$	$< 2.4(2016)/2.5(2017/18)$
$m_{\gamma\gamma}$	$[100, 180]$ [GeV]	$\Delta R_{\gamma j}$	> 0.4
		njets	> 1
		m_{jj}	$[70, 190/1200]$ [GeV] (WED/NMSSM)
		DEEPJET	> 0,with highest sum of b -Tag score

we integrated the DEEPJET in the analysis framework and studied the performance with non-resonant $HH \rightarrow b\bar{b}\gamma\gamma$ analysis specific signal and background samples listed in Table 5.7. From a comparison study between DEEPJET and DEEPCSV in Fig. 5.4, it can be seen that the usage of DEEPJET b tagger gives up to 10% improvement in b-efficiency for 1% mistag-rate.

The variation in signal efficiency with offline selections, for different mass hypotheses,

Table 5.7: Samples used for DEEPJET performance study

signal	GluGluToHHTo2B2G_node_SM_13TeV-madgraph
background	QCD_Pt-40toInf_DoubleEMEnriched_MGG-80toInf_TuneCUETP8M1_13TeV_Pythia8
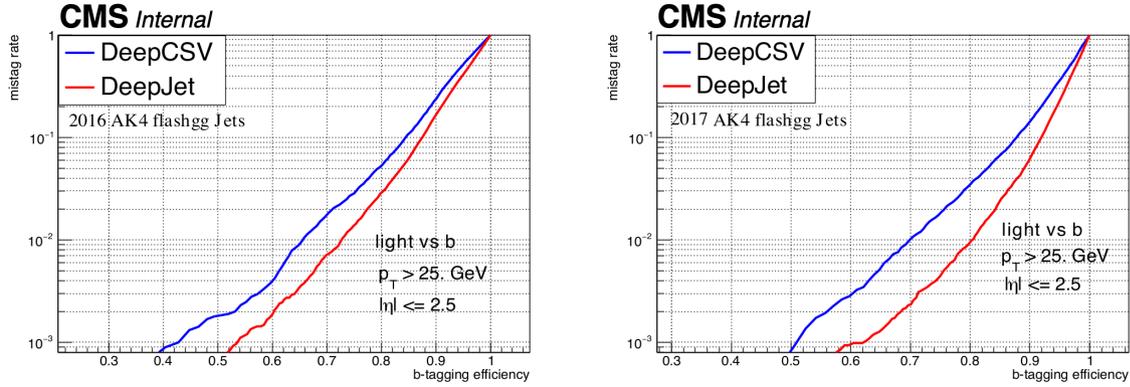
Figure 5.4: Performance comparison between DEEPJET and DEEPCSV for 2016 (left) and 2017 (right) data-taking years

is given in Fig. 5.5. It shows that signal efficiency increases along the x-axis. It happens because more events pass selections as $H \rightarrow \gamma\gamma$ and $Y \rightarrow b\bar{b}$ handles start entering in a high p_T regime with a higher X mass. Differing from this trend, signal efficiency along the y-axis first increases because of m_{jj} selection, which is a bit tighter for the $Y < 125$ GeV mass region. Hence, it lowers the efficiency for $Y < 125$ GeV. As it starts entering the higher Y mass region, the efficiency decreases because the $H \rightarrow \gamma\gamma$ handle becomes less energetic in this region. Therefore, photon selections dominate and make the efficiency low. This can also be understood using Figs. 5.7 and 5.6 where the cut by cut effect on signal efficiency, for each Y, is shown.
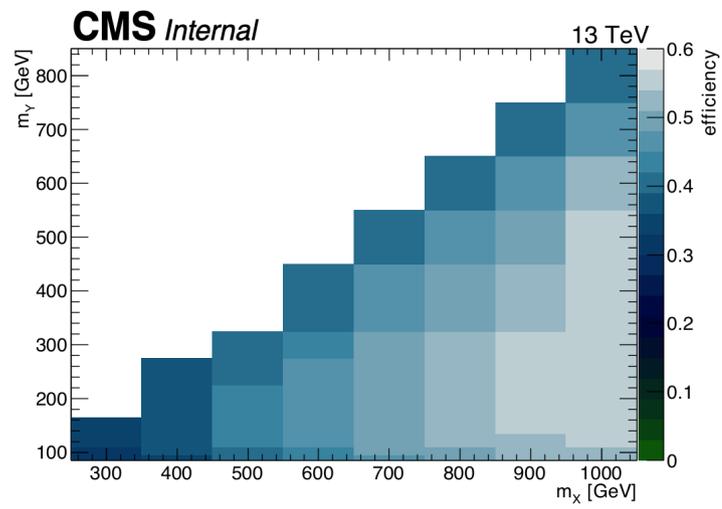
Figure 5.5: Signal efficiency for each X-Y mass point after offline selections
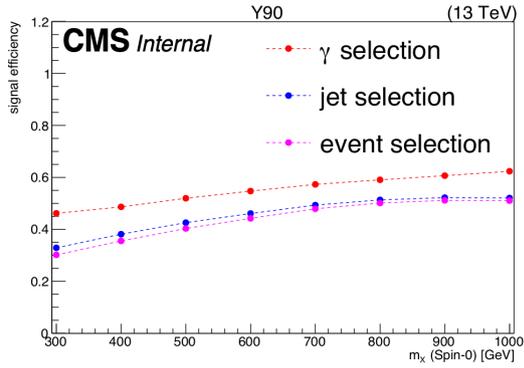
(a) Y=90 GeV
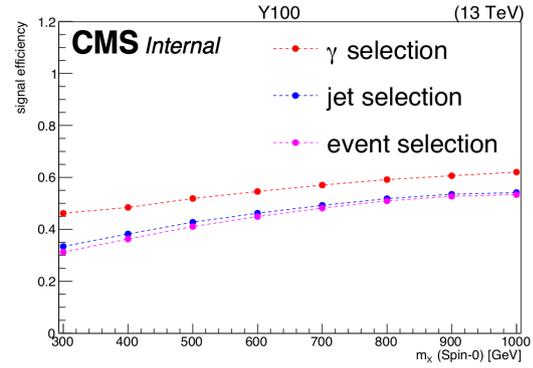
(b) Y=100 GeV
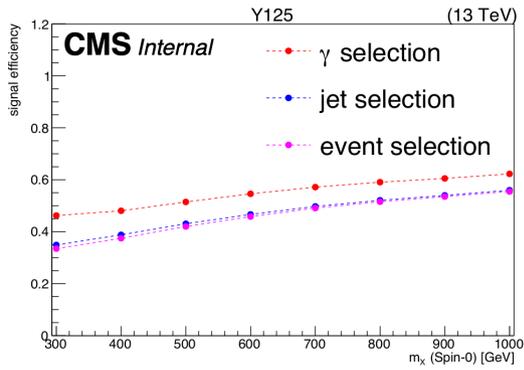
(c) Y=125 GeV
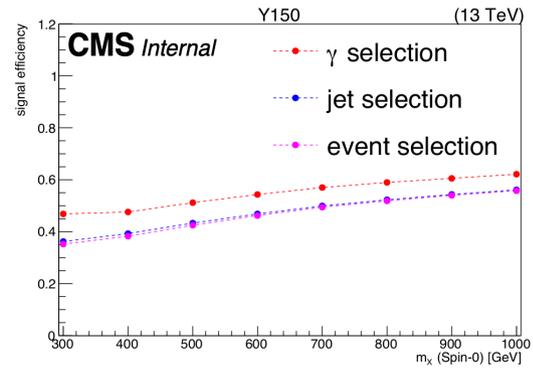
(d) Y=150 GeV
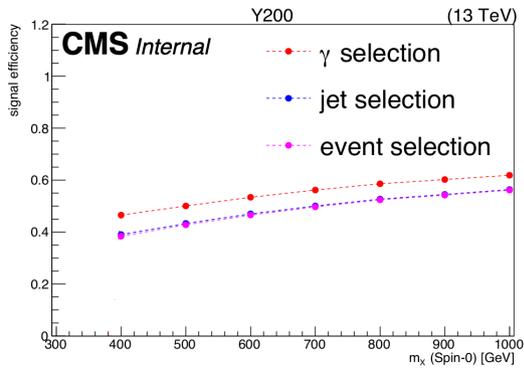
(e) Y=200 GeV
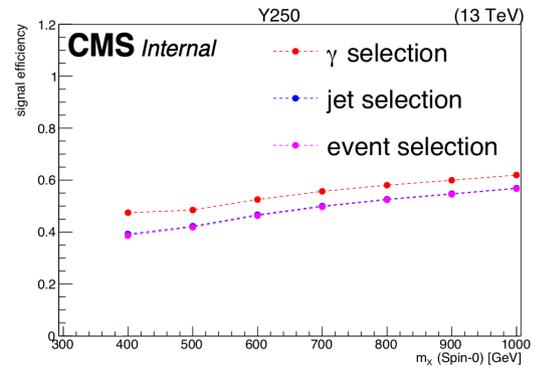
(f) Y=250 GeV

Figure 5.6: Signal efficiency cutflow graphs for each Y mass point
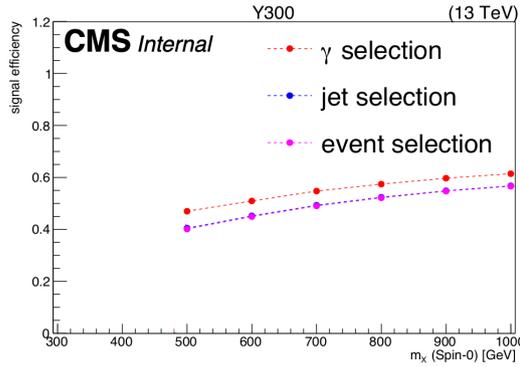
(a) Y=300 GeV
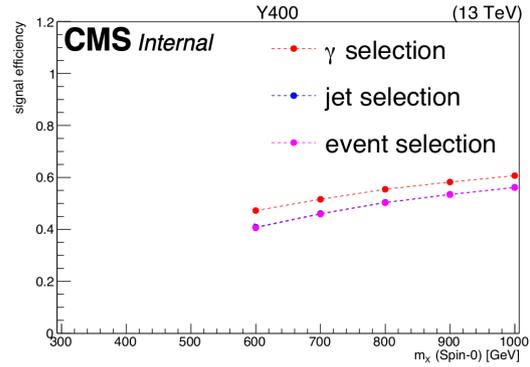
(b) Y=400 GeV
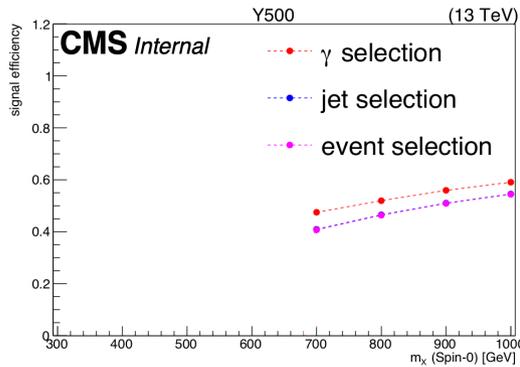
(c) Y=500 GeV
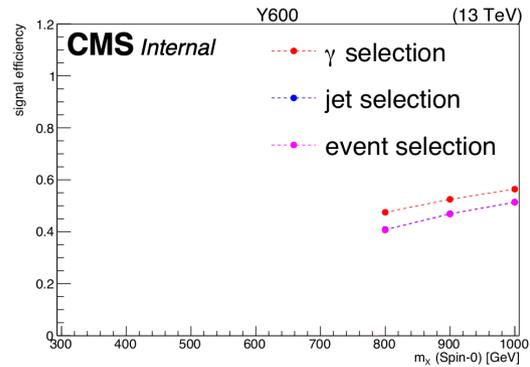
(d) Y=600 GeV
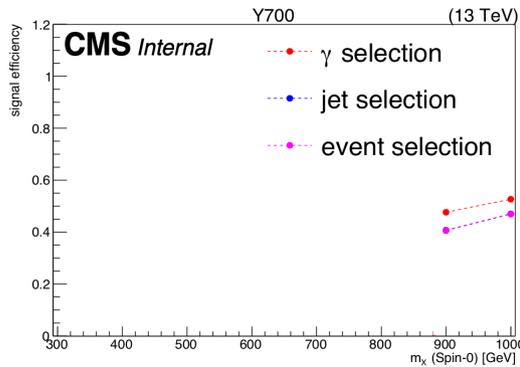
(e) Y=700 GeV
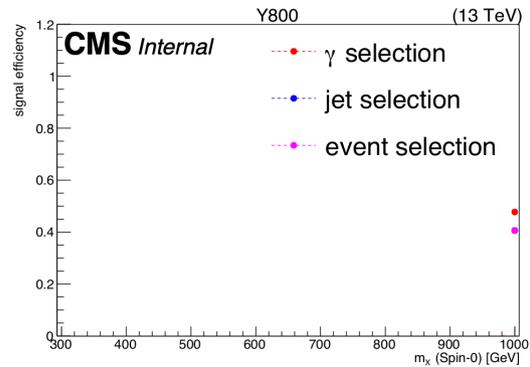
(f) Y=800 GeV

Figure 5.7: Signal efficiency cutflow graphs for each Y mass point

The bar graphs for the signal efficiency of bulk radion and bulk KK graviton after passing the offline selections are in Fig. 5.8. The numbers in the bar graph for bulk radion are consistent with numbers in the 2D efficiency plot for $Y = 125$ GeV case as the kinematic distributions do not differ much for both the signal hypotheses as shown in Fig. 5.9. For bulk KK graviton efficiency is slightly higher because of the spin-2 structure explained in Section 5.3.5.6. In this case, the angular distribution of decay products remain in more central region of detector; therefore events are in high p_T regime.
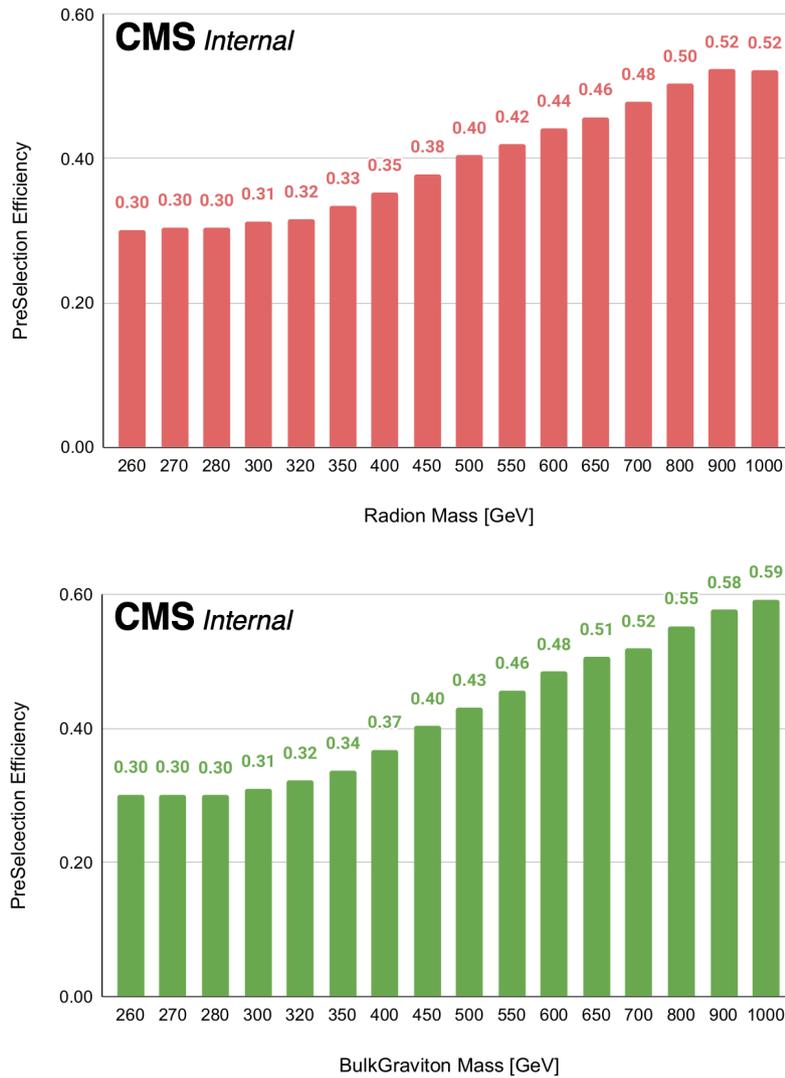
Figure 5.8: Signal efficiency for bulk radion and bulk KK graviton

(a) leading jet p_T

(b) subleading jet p_T

(c) leading photon p_T

(d) subleading photon p_T

(e) leading jet η

(f) leading photon η
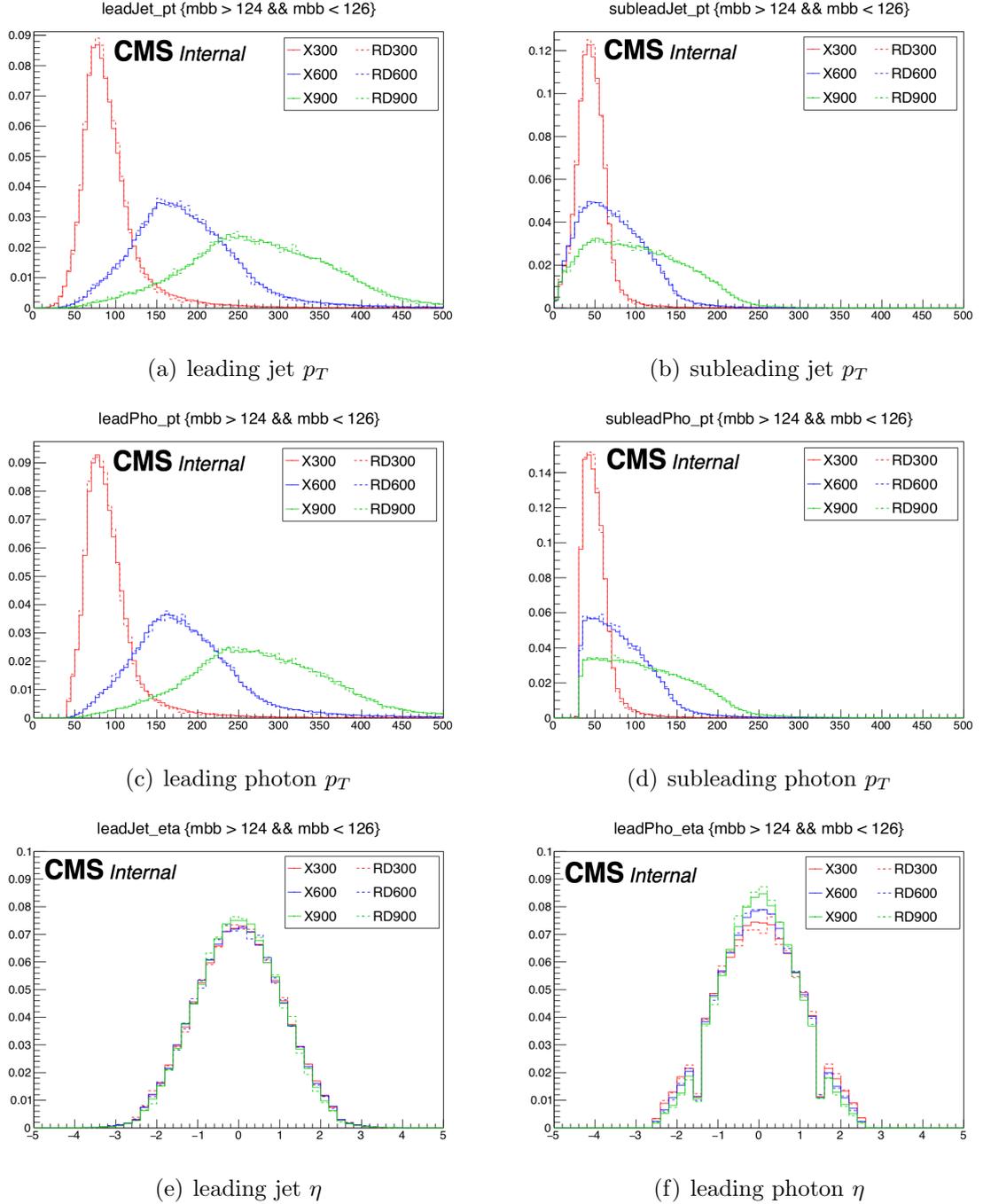
Figure 5.9: Generator level kinematic comparisons between NMSSM (Y=125 GeV) and WED spin-0 hypotheses (legend RD=bulk radion and X=NMSSM X resonance)

5.3.5   Non-resonant background rejection

A machine learning tool, XGBOOST [172] with SCIKIT-LEARN [173], is used to have a multiclass BDT classifier to discriminate signal and dominating non-resonant background processes. To train the BDT, simulated samples are used. The samples are put in three different classes (one for signal and two for background processes) to use as the input for the classifier.

Signal = Signal simulations (NMSSM)

Background-1 = $\gamma\gamma$ + jets and **Background-2** = $j\gamma/jj$+jets

Since we deal with various resonance X masses, which vary from 260 to 1000 GeV (for both BSM models) and various Y masses (for NMSSM case only), kinematics differ significantly in specific X-Y mass ranges. Therefore, BDT is trained separately for six different mass ranges to achieve efficient training performance for every signal. In this division, three ranges are in m_X (lowX: $m_X < 500$ GeV, midX: $m_X = 500$–700 GeV and highX: $m_X > 700$ GeV) and the rest of three in m_Y (lowY: $m_Y < 300$ GeV, midY: $m_Y = 300$–500 GeV and highY: $m_Y > 500$ GeV) as labeled in Fig. 5.10. We call these six mass regions lowX_lowY, midX_midY, highX_highY, midX_lowY, highX_midY and highX_lowY. The basis of this division is Lorentz boost factor [174] defined with $m_X/(m_H + m_Y)$, where m stands the pole mass. Fig. 5.10 also contains the boost factors for each $X - Y$ pair. The $X - Y$ mass regions have the following properties:

- **lowX_lowY, midX_midY and highX_highY:** boost-factor $\lesssim 2$; where jets are distributed in different hemispheres, and objects are soft and can be treated as resolved.

- **midX_lowY and highX_midY:** $1.5 \lesssim$ boost-factor $\lesssim 3$; where jets are in the same hemisphere and objects can be treated as resolved.

- **highX_lowY**: $2.5 \lesssim$ boost-factor $\lesssim 5$; where jets and photons are tend to be close as they are in high p_T regime.

For three m_Y ranges, training considers events in three different and overlapping wide m_{jj} intervals (or bins). While deciding these intervals, we ensure that we have

the appropriate m_{jj} observable shape of every signal belonging to the corresponding m_Y range without compromising signal efficiency much.

lowY = [90, 250] GeV, m_{jj}-bin = [70, 400] GeV

midY = [300, 500] GeV, m_{jj}-bin = [150, 560] GeV

highY = [600,800] GeV, m_{jj}-bin = [300, 1000] GeV

This will make each training more efficient to suppress background within a specific m_{jj} interval where we model m_{jj} in data as an observable for signal extraction.
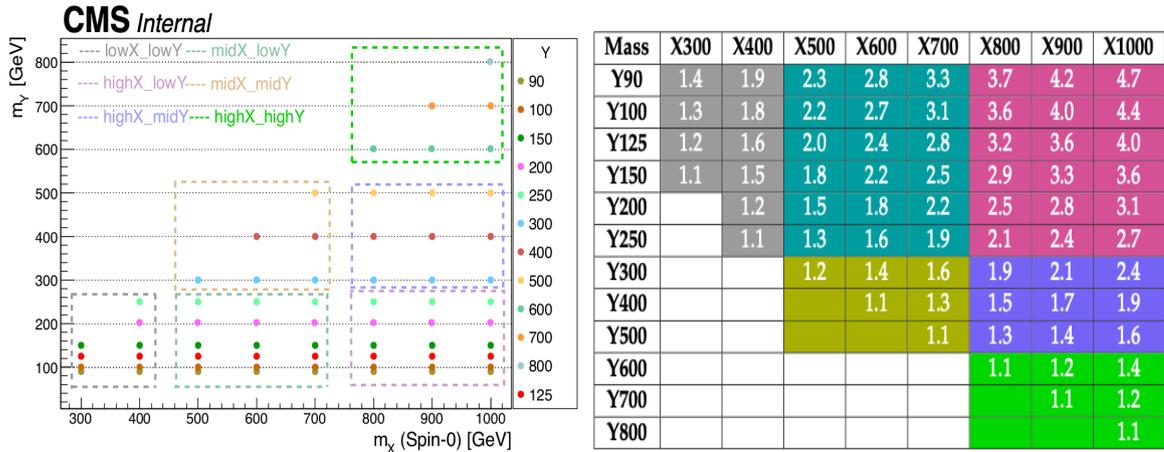
Figure 5.10: $X - Y$ mass bins for MVA training using boost factors

5.3.5.1 Run-2 training strategy

We build a training workflow for our analysis to make the BDT training procedure easier for full Run-2 training and further validated it. We perform a merged training for all three years by adding a year label as training input and separate training for each data-taking year, using the same hyperparameters[3] and MC samples. We compare training outputs of merged training and year-wise separate training. The year-label for the merged case does not bring any improvement but only helps to pick training output

---

[3]In machine learning, hyperparameter is a configuration that is external to the model and whose value cannot be estimated from data. It helps to discover the parameters of the model that result in the meaningful predictions of data.

year-wise from the merged training file. Figure 5.11 shows that from both the training, the output ROC performance is same for both the background classes. Thus without making any compromise in training performance, it makes the training workflow faster and easier to deal with BDT training for full Run-2[4].

---

[4]We have made this study only for bulk radion signal process and adapted same workflow for NMSSM default training.

(a) 2016

(b) 2016

(c) 2017

(d) 2017

(e) 2018

(f) 2018
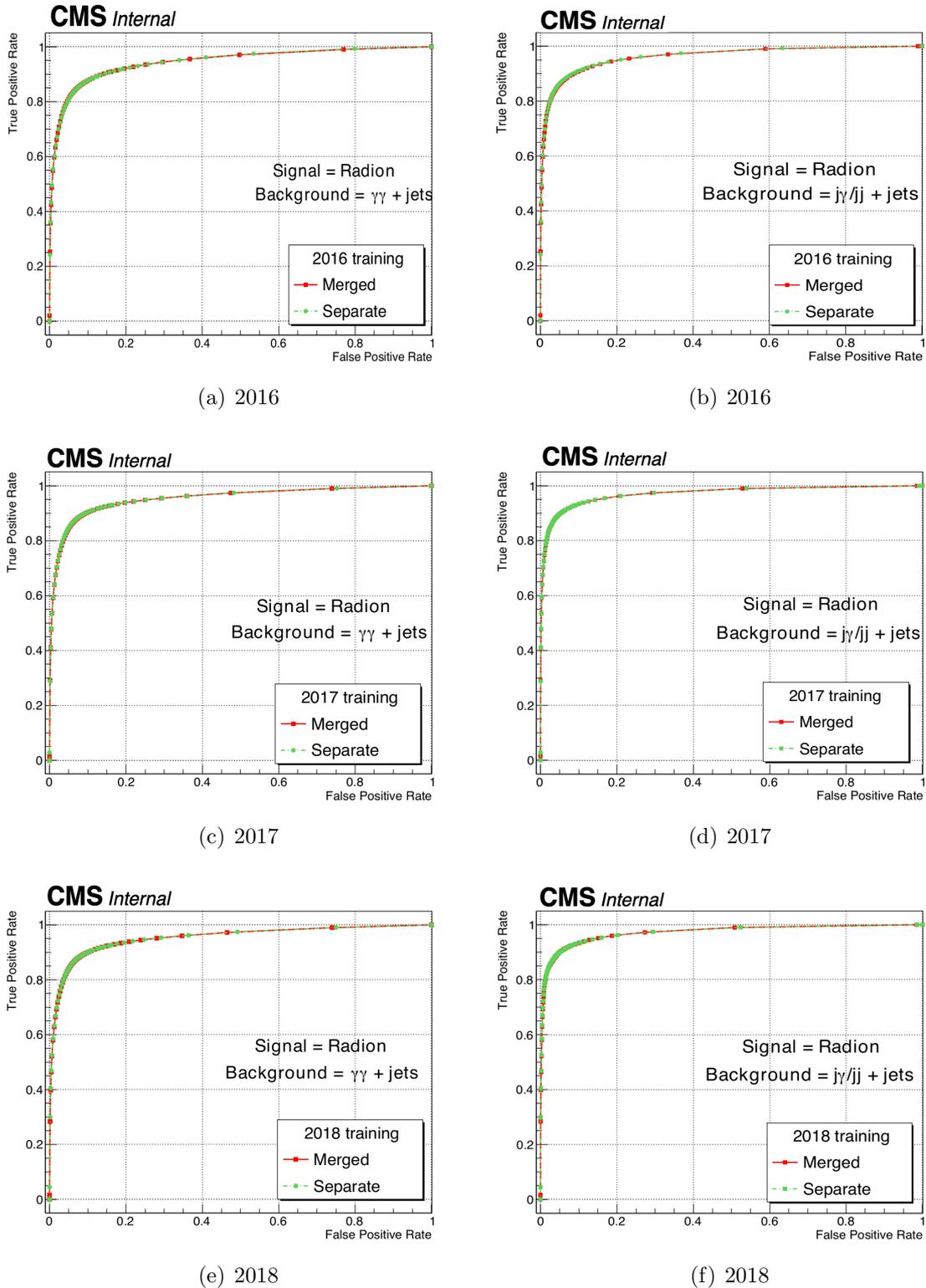
Figure 5.11: ROC curve performance comparison between merged and year wise separate training

5.3.5.2   Training Input Variables:

For BDT training, three groups of input variables are used:

1) kinematic variables that have discriminating distribution for signal and background processes.

- Helicity angles ($|\cos\theta_{HY}^{\mathrm{CS}}|$, $|\cos\theta_{bb}^{\mathrm{CS}}|$, $|\cos\theta_{\gamma\gamma}^{\mathrm{CS}}|$), here CS refers to the Collins-Soper frame. $|cos(\theta_{xx}^{CS})|$ is defined as the angle between the particle x and the direction defined by the $H \to xx$ candidate, where $x = \gamma$ or b

- First two minimum angular distance between photons and jets ($\Delta\mathrm{R}(\gamma,\mathrm{jets})$)

- $p_T(jj)/m_{jj\gamma\gamma}$ and $p_T(\gamma\gamma)/m_{jj\gamma\gamma}$

- Leading and subleading photons $p_T(\gamma)/m_{\gamma\gamma}$ and jets $p_T(j)/m_{jj}$

2) object identification variables that could reject background misidentified objects

- Leading and subleading photonID to reject misidentified photon contribution

- Leading and subleading jets b tagging score from DEEPJET algorithm to reject light jets

3) resolution variables

- Leading and subleading photons resolution variables:
  a) energy resolution (σ_E/E)
  b) mass resolution ($\sigma_{m_{\gamma\gamma}}/m_{\gamma\gamma}$)

- Leading and subleading jets resolution variables:
  a) energy resolution ($\sigma_{p_T}(j)/p_T(j)$)
  b) mass resolution ($\sigma_{m_{jj}}/m_{jj}$)

In the first set, the Helicity angles help to reject diphoton QCD backgrounds. For resonant signal, Helicity distributions are uniform due to the spin-0 nature of resonance, while for the background, it will be peaking around high values because of collinear

emissions. On the other side, transverse momentum is also a good discriminating variable. For the background, it has a steep fall starting from the low p_T region. Scaling transverse momentum with mass, decorrelate the classifier from $m_{\gamma\gamma}$ and m_{jj}. $p_T(\gamma)/m_{\gamma\gamma}$ and $p_T(j)/m_{jj}$ variables mainly improve the sensitivity in the low $\tilde{M}_X$ region, where analysis is more sensitive. In this region, the Higgs bosons are produced nearly at rest. Hence, the photons and jet momenta have values of around $p_T > m_{H,Y}/2$. Also, the jet and photon pairs are produced isotropically due to the scalar nature of the Higgs and Y bosons. These dimensionless variables will therefore improve the performance of the BDT. Although a low $\tilde{M}_X$ region motivated these variables, they are used for training across the entire resonance mass region.

In the second set, the output scores of photonID and DEEPJET b discriminant are used. These variables reject the background photons and jets, which can mimic the signal objects after passing the offline selections.

The last set corresponds to resolution variables. Due to the excellent ECAL response of the CMS detector, photon yields better resolution, which is fed into training. The resolution is directly taken from photon energy regression training. Similarly, for b jets, the resolution variable from the b jet energy regression training is used.

Shape comparison for the signal and background distributions corresponding to the above variables are added in Figs. 5.12, 5.13, 5.14 and 5.15 for three X mass ranges in low m_{jj}-bin. Mid and high m_{jj} bin plots are added in Appendix B.1. The Data/MC comparison for these variables had been studied for non-resonant $HH \to b\bar{b}\gamma\gamma$ analysis [1] as the BDT inputs remain the same for both the analyses.

Apart from all these variables, median energy density ρ is also added to include pileup information in training, which differs among 2016, 2017, and 2018 data-taking years. It helps the BDT to learn the pileup profile present in real data. For this analysis, we perform a merged training combining three data-taking years as explained in Section 5.3.5.1 in every training mass range.

5.3.5.3   Training and hyperparameter optimization:

For signal, the MC samples are mixed according to six mass range definitions. It is checked that in every mass range, the training is efficient enough to reject background events. Signal events for each resonance mass keep the same weightage during the BDT training. Signal and backgrounds are normalized to unity before using as training input, where for normalization of backgrounds, the relative fraction of different background processes is preserved. This normalization keeps the signal and background on equal footing to have a shape-based training depending upon the given inputs. The samples are divided into train and test samples after randomization of events, where BDT uses the training sample to learn the model for the classifier, and a test sample is used to check the performance of that classifier.

The machine learning (ML) software used for BDT training is based on python libraries and used for fast solutions of classification and regression problems. Multiclass features of these ML tools help to train for one signal hypothesis with respect to more than one background hypothesis and further check the training performances for individual background hypotheses.

For this analysis, fine-tuning of input hyperparameters is done using **RandomizeCVSearch** function of SCIKIT-LEARN [175] with **5-fold cross-validation**. This optimization helps to improve the training performance and stops over-training. A predefined range is used for the optimization of these parameters. This function takes the best possible combinations of hyperparameters and returns the best estimator after comparing performance among them. More details about these hyperparameters are available in XGBOOST Documentation at [176].

Best Estimator:

XGBClassifier(base_score=0.5,   booster='gbtree',   colsample_bylevel=1, colsample_bynode=1,   colsample_bytree=1,   gamma=0,   learning_rate=0.1,   max_delta_step=0, max_depth=5, min_child_weight=1e-04, missing=None, n_estimators=2500, n_jobs=1,   nthread=4,   objective='binary:logistic',   random_state=0,   reg_alpha=0.01,   reg_lambda=0.3, scale_pos_weight=1, seed=0,silent=True, subsample=1, verbosity=1)

An early-stopping function is also used to control the over-training. A loss-function evaluates if the model is learning the given data well. The early-stopping function keeps track of the loss function with each iteration and takes decisions accordingly to stop over-training if the loss function does not improve during the training.

Here we use the XGBOOST recommended cross-entropy loss function, a more generic form of logarithmic loss which serves multiclass classification problems. The mathematical expression for loss function $= -(1/N) \sum_{i=1}^{N} \sum_{j=1}^{M} y_{ij} log\ p_{ij}$, where N refers to number of events and M stands for number of BDT outputs. y_{ij} stands for the BDT output for ith event with jth BDT output and p_{ij} is the probability associated with y_{ij}.

Figure 5.16 shows the distributions of the classifier output for signal (blue) and background (red) events for all six mass ranges, which are good discriminator to separate signals and backgrounds. Figures 5.17 and 5.18 show the ROC curve performances of the classifiers for both the background classes ($\gamma\gamma$+jets and $j\gamma/jj$+jets) from all six training. In these ROCs, the performance to reject $j\gamma/jj$+jets background is higher due to photonID input in the BDT training. This helps to suppress misidentified photon contributions coming from $j\gamma/jj$+jets. As the training is performed on NMSSM signals, a study has been made to check training performance for WED signals. This is added in Section 5.3.5.6.
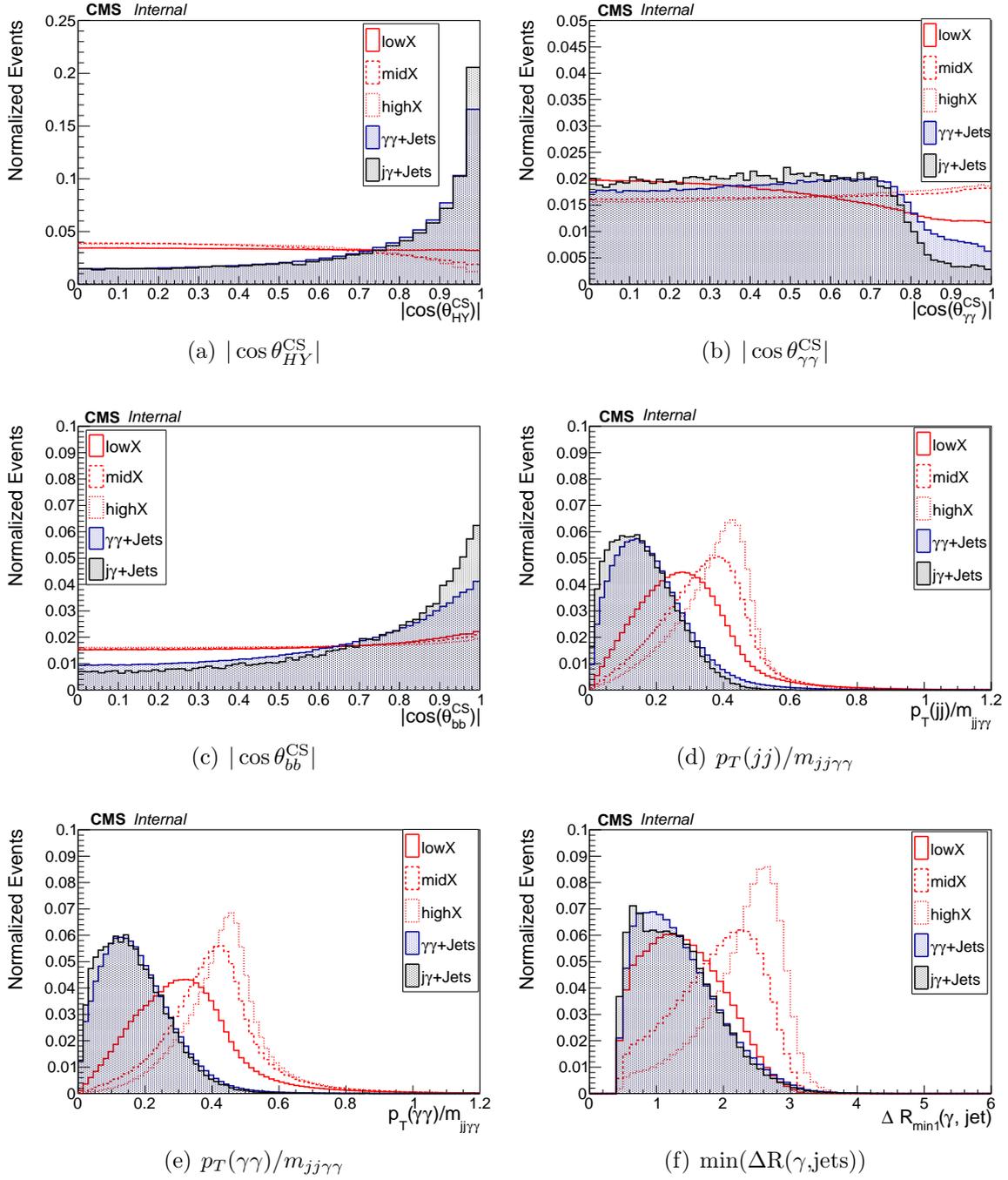
Figure 5.12: Signal and background kinematics for lowY mass region

(a) other min($\Delta R(\gamma,$jets$)$)

(b) $\sigma_{m_{\gamma\gamma}}/m_{\gamma\gamma}$

(c) $\sigma_{m_{jj}}/m_{jj}$

(d) leading jet b Discriminant

(e) leading jet bRegressionNN

(f) $p_T(j1)/m_{jj}$
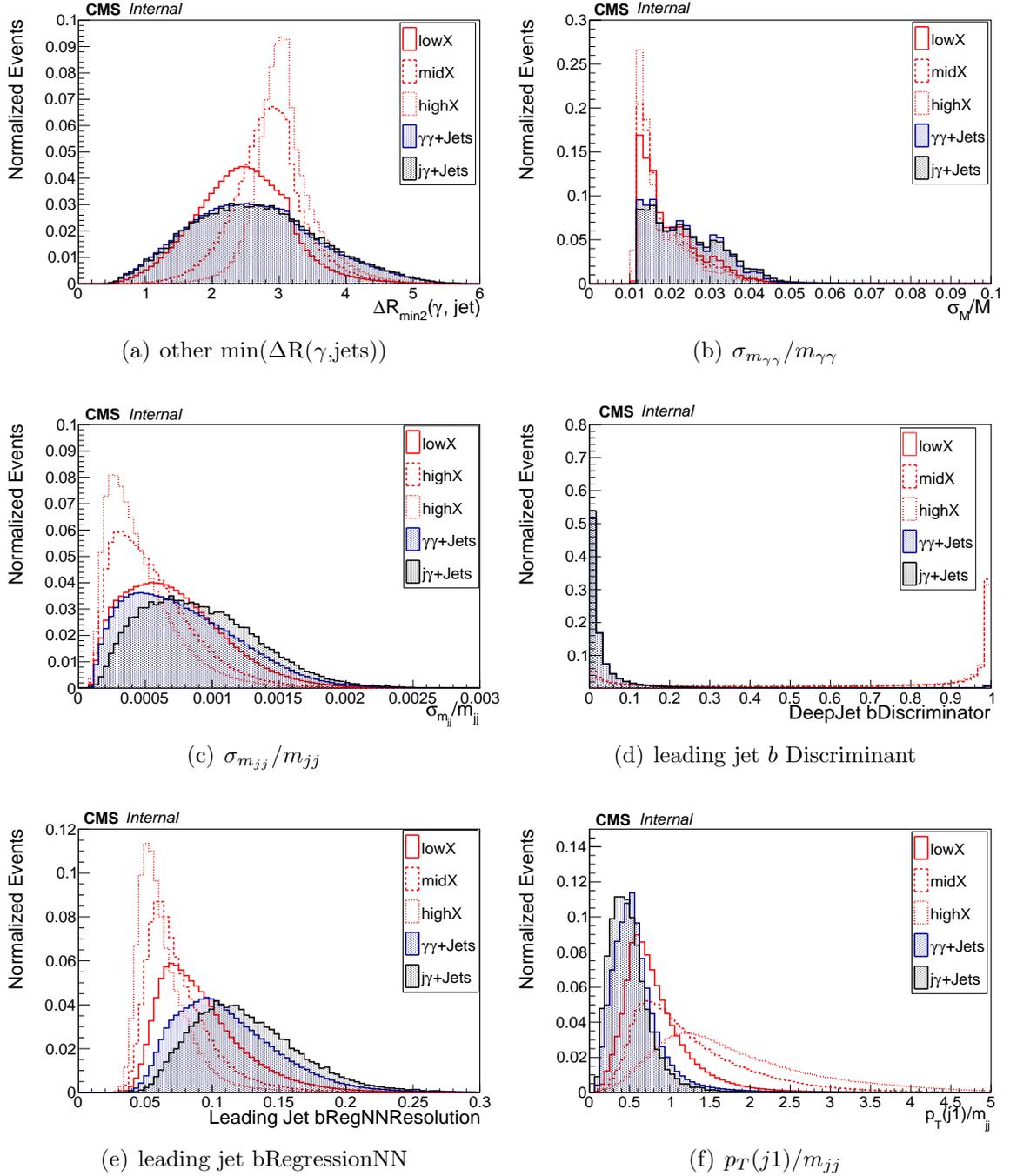
Figure 5.13: Signal and background kinematics for lowY mass region

(a) leading photonID MVA

(b) $p_T(\gamma 1)/m_{\gamma\gamma}$

(c) leading photon σ_E/E

(d) subleading Jet b Discriminant

(e) subleading Jet bRegressionNN

(f) $p_T(j2)/m_{jj}$
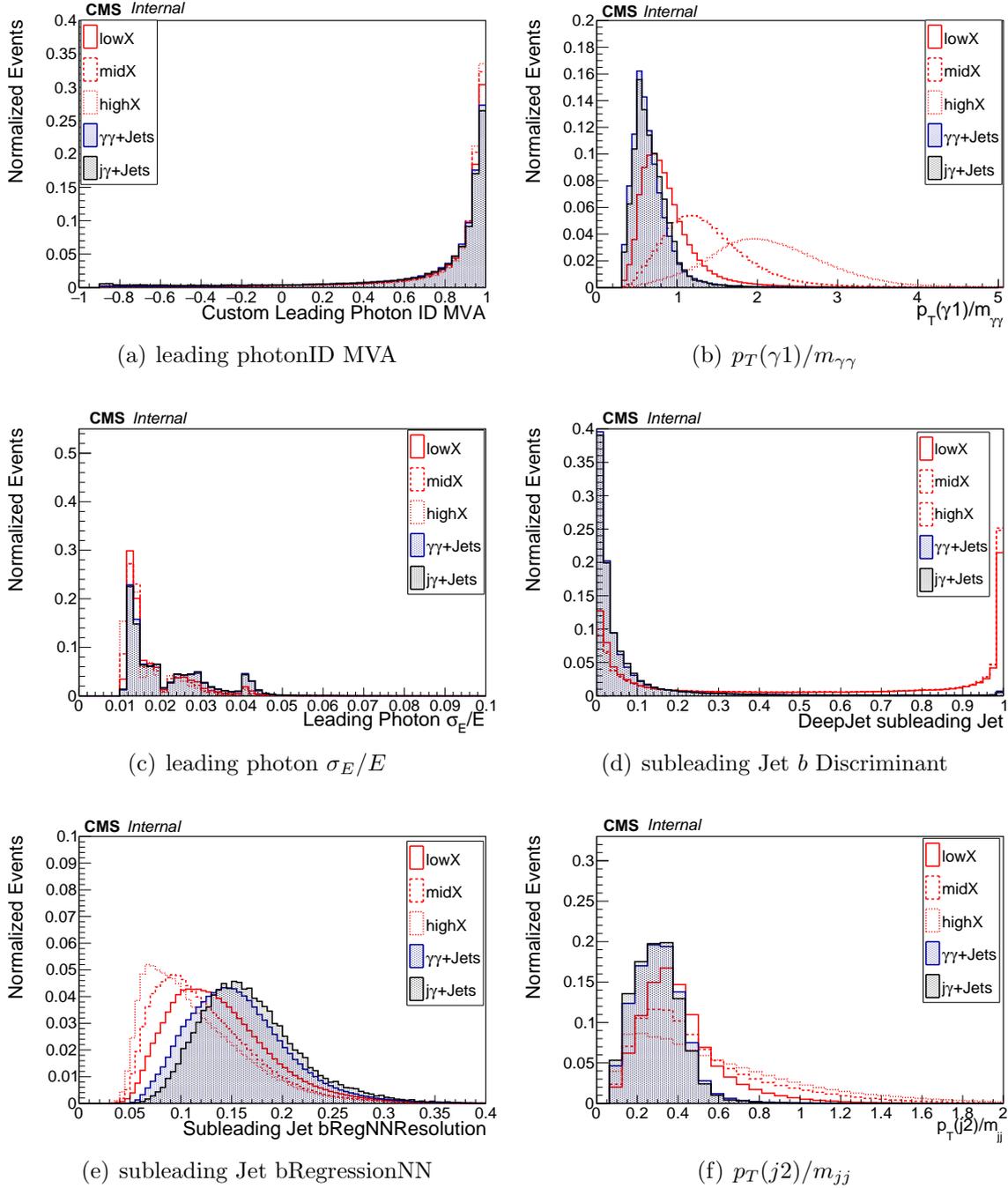
Figure 5.14: Signal and background kinematics for lowY mass region

(a) subleading photonID MVA

(b) $p_T(\gamma 2)/m_{\gamma\gamma}$

(c) subleading photon σ_E/E
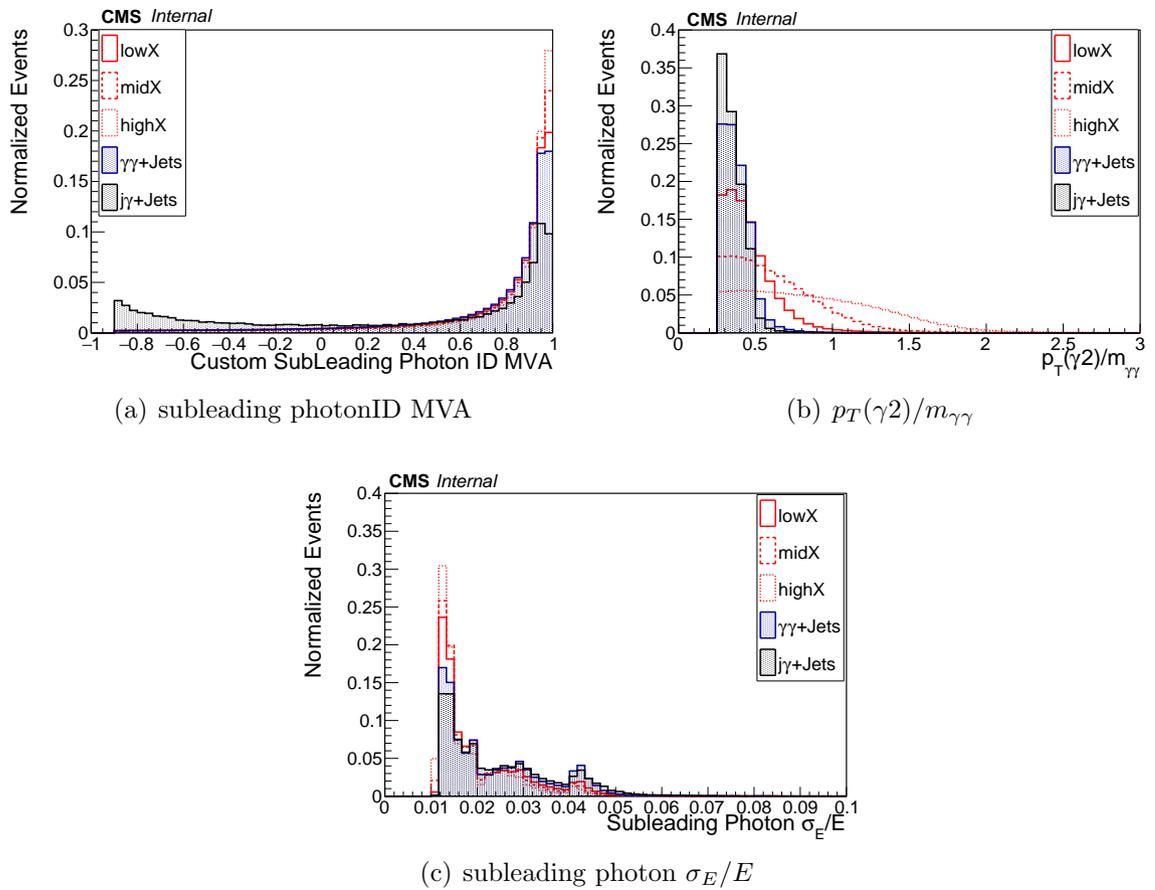
Figure 5.15: Signal and background kinematics for lowY mass region

(a) lowX_lowY

(b) midX_lowY

(c) highX_lowY

(d) midX_midY

(e) highX_midY
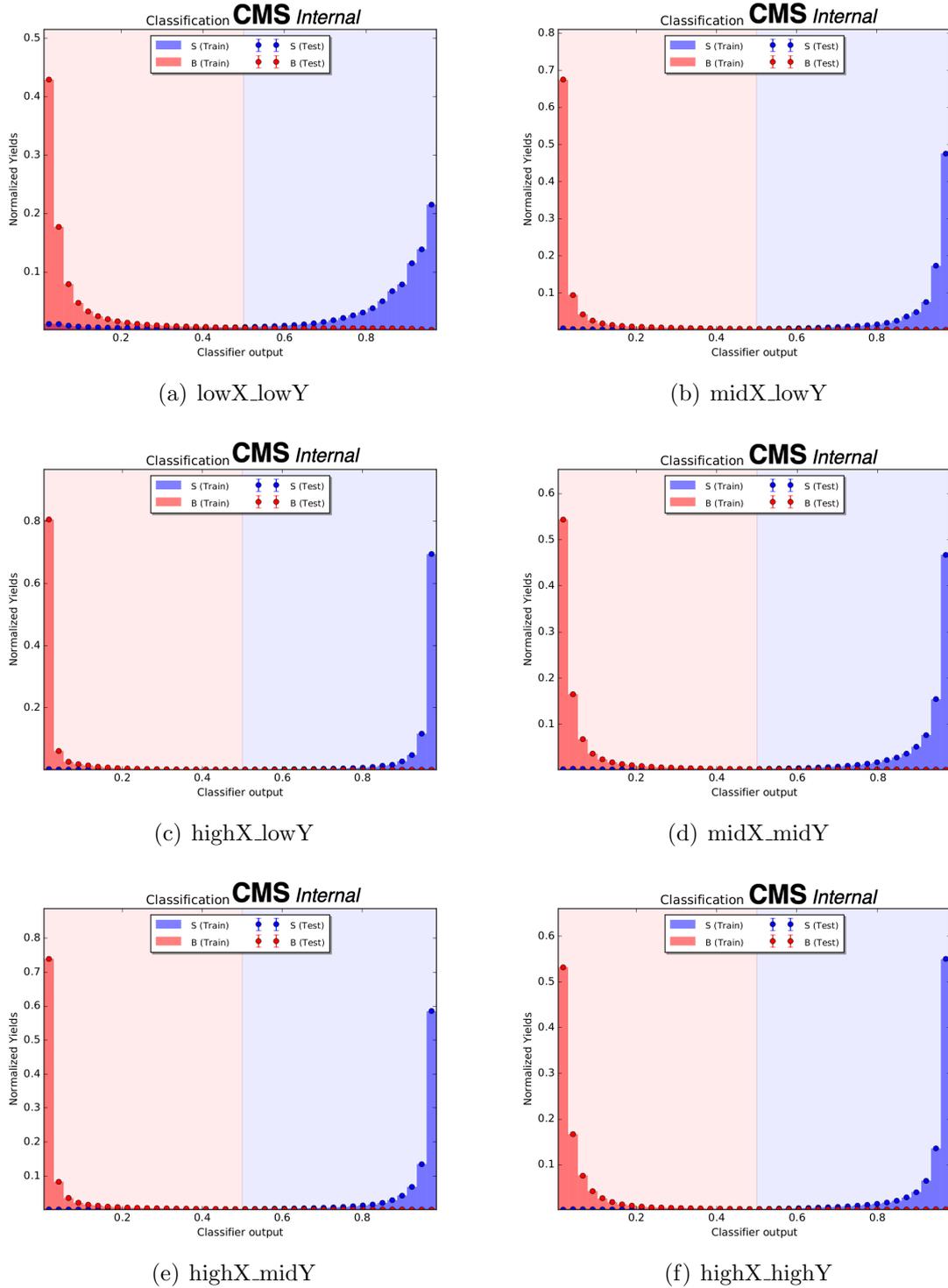
(f) highX_highY

Figure 5.16: BDT classifier output

(a) lowX_lowY

(b) midX_lowY

(c) highX_lowY

(d) midX_midY

(e) highX_midY
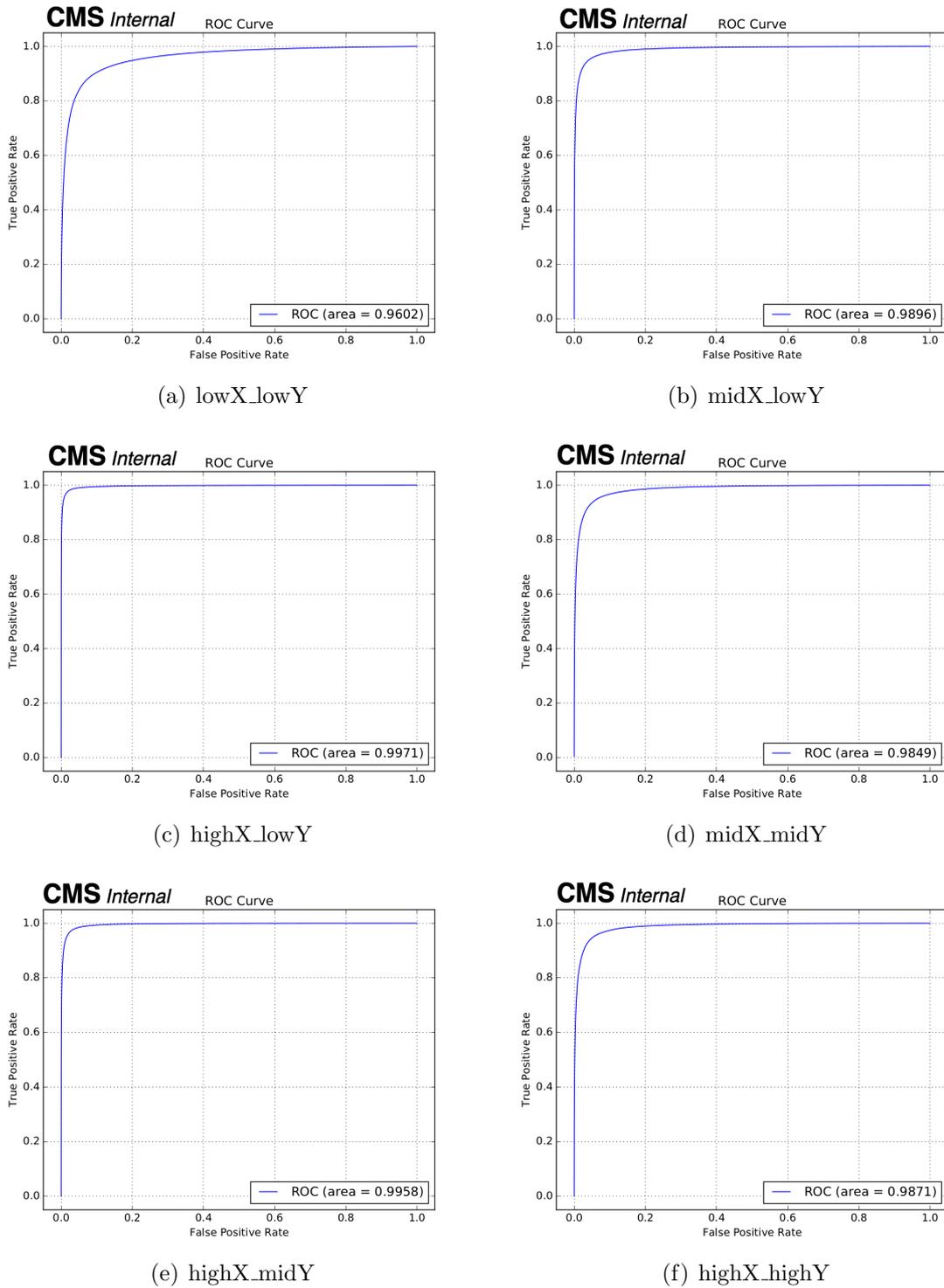
(f) highX_highY

Figure 5.17: ROC performance for $\gamma\gamma$+jets
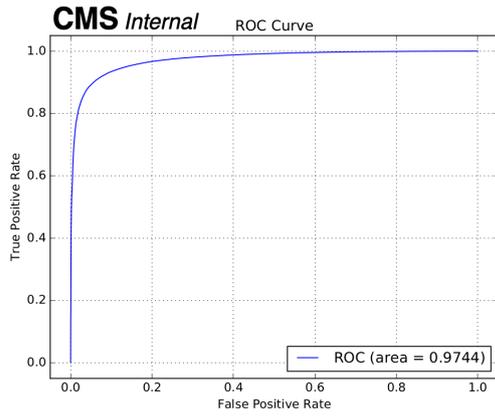
(a) lowX_lowY
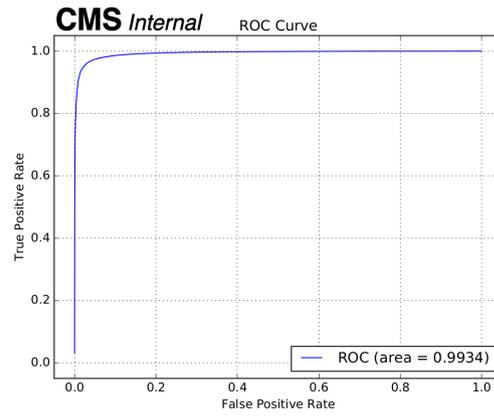
(b) midX_lowY
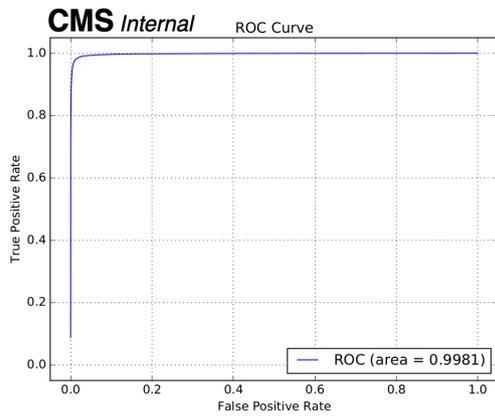
(c) highX_lowY
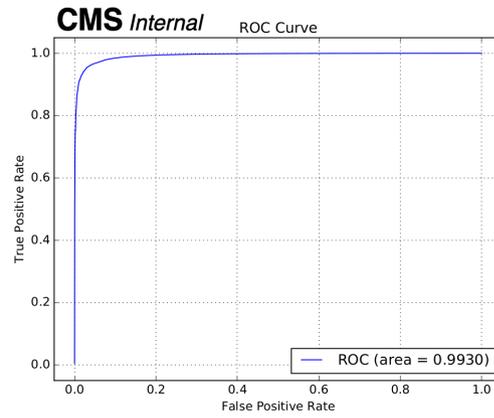
(d) midX_midY
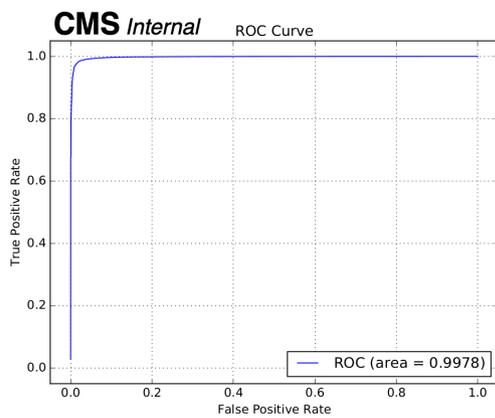
(e) highX_midY
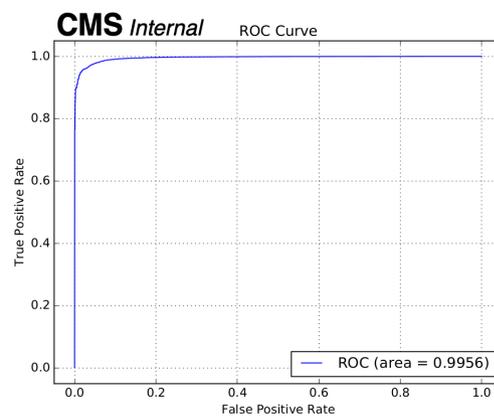
(f) highX_highY

Figure 5.18: ROC performance for $j\gamma/jj$+jets

5.3.5.4 MVA flattening:

This analysis is statistically very limited due to the high selection efficiency and purity of $b\bar{b}\gamma\gamma$ channel; therefore, applying a tight selection on BDT output is not that optimal approach as statistics might lack for a robust background modeling. Event categorization is an excellent approach to enhance analysis sensitivity in such cases. Following this approach, we categorize events into MVA categories. The BDT output at Fig. 5.16 shows that there is a sharp peak for signal events around one. This indicates a higher density of signal events around region one, making categorization hard to optimize final MVA boundaries. To make the categorization easier, a cumulative transformation is applied to both signal and background events with respect to the signal.

This transformation makes signal BDT output uniform without compromising the training performance. The background follows the falling BDT distribution only. Fig. 5.19 has transformed MVA output distributions for signal (yellow line) and background (stacked histograms) events from all six mass ranges where we train the BDT.

5.3.5.5 Data/MC comparison for MVA Output:

In this analysis, non-resonant background simulation samples are only used for BDT training and optimization studies. The data-driven method is used to consider non-resonant background contributions. Therefore, to validate the transformed BDT output, a comparison between BDT output of non-resonant background simulations (normalized to its cross-section times luminosity values) and Run-2 data is made so that it could confirm that the trained BDT is efficient enough to reject non-resonant backgrounds coming from data. In this comparison, signal region $m_{\gamma\gamma} = [115, 135]$ GeV is kept blinded, and a scale factor (SF) is used to minimize data/MC inconsistencies. The SF is a ratio of event yields from non-resonant background MC and data in the control region (side-bands of blinded $m_{\gamma\gamma}$ distribution). This differs among six training output distributions due to statistics and kinematics of mass ranges (varies from 2 to 5). Fig. 5.19 shows that background MC BDT output describes the data well.

(a) lowX_lowY

(b) midX_lowY

(c) highX_lowY

(d) midX_midY

(e) highX_midY

(f) highX_highY
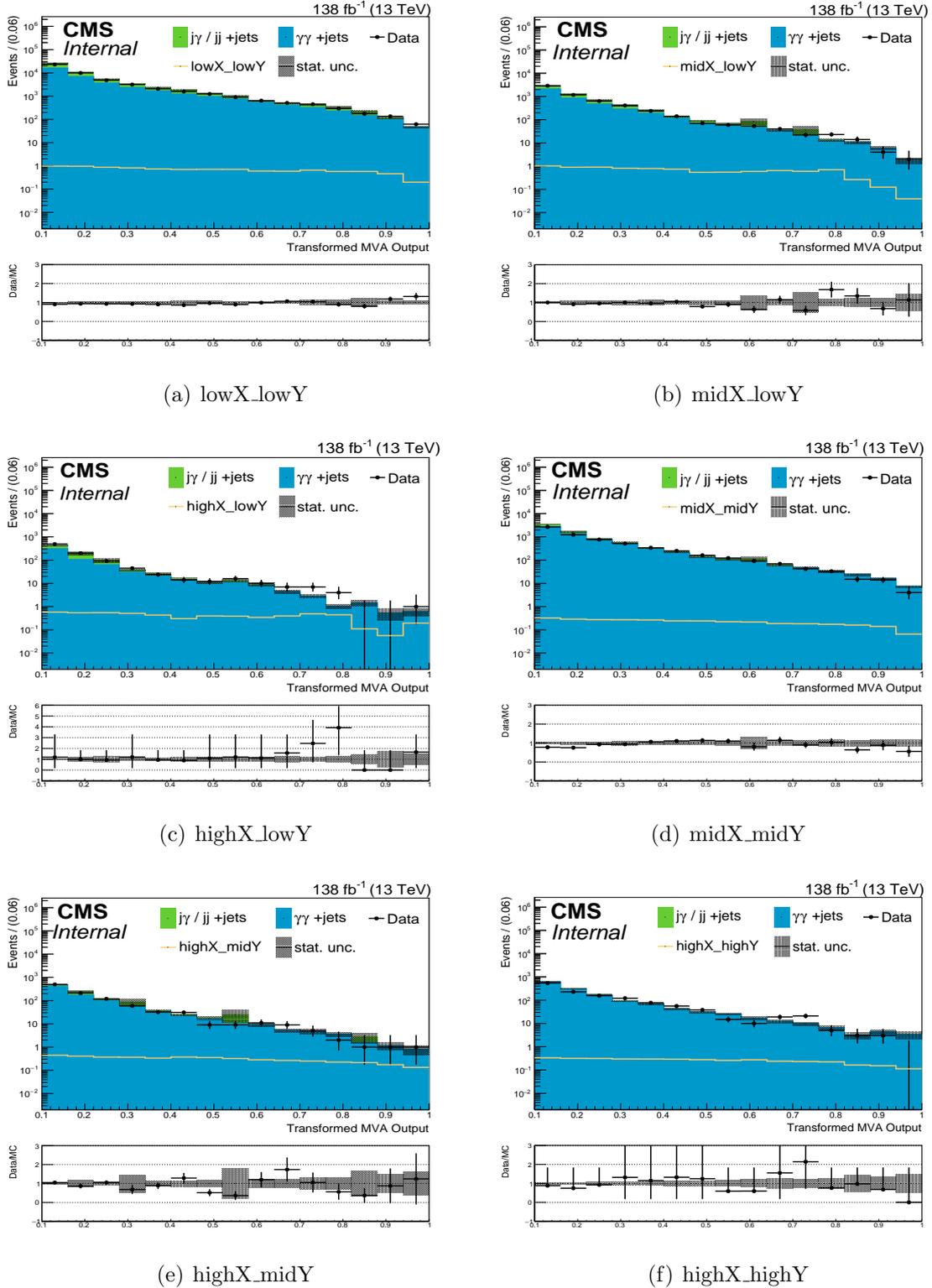
Figure 5.19: Data/MC comparison of transformed MVA Output for all six BDT training; the dominant contribution comes from irreducible $\gamma\gamma$+jets background

5.3.5.6 NMSSM training validation on WED signal

We apply the same BDT training, produced for NMSSM lowY mass ranges to deal with bulk radion and bulk KK graviton searches. We do the following checks to show that NMSSM training is optimal for all three signal hypotheses.

As we observed in Fig. 5.9, kinematics for bulk radion and NMSSM Y=125 are similar. For bulk KK graviton, the $|\cos\theta_{HY}^{\mathrm{CS}}|$ differs due to the spin-2 nature of resonance. Spin-0 and spin-2 kinematic comparisons are added in Fig. 5.21.

Variations in other kinematics are directly correlated with $|\cos\theta_{HY}^{\mathrm{CS}}|$ for spin-2 case; thus, having only one different kinematic shape for spin-2 does not differ the training output much for spin-2 compared to the spin-0 case. We produce separate bulk KK graviton (bulk radion) signal training and compare the training results to validate the above statements. We apply bulk KK graviton (bulk radion) training on bulk KK graviton (bulk radion) sample (dotted line in Fig. 5.20) and NMSSM training on bulk KK graviton (bulk radion) sample (solid line in Fig. 5.20) and produce upper limits on cross section. This shows that NMSSM $X \rightarrow YH$ training can be used for all three signals as the results do not change much.
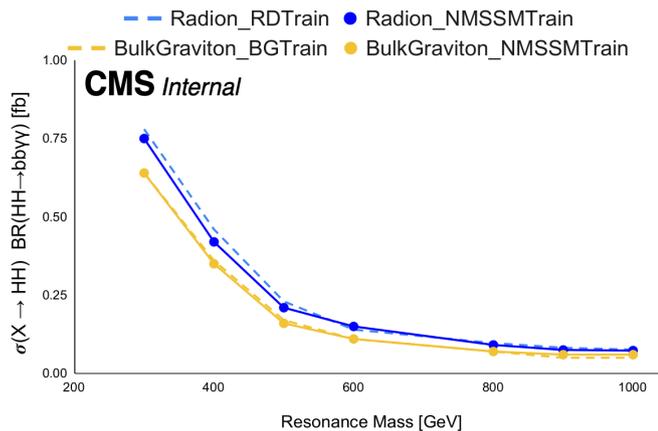
Figure 5.20: Validation of NMSSM training on WED signals; in legend RDTrain = bulk radion training, BGTrain = bulk KK graviton training, NMSSMTrain = NMSSM training; solid and dotted yellow (blue) lines compare the results for bulk KK graviton (bulk radion) from NMSSM training and bulk KK graviton (bulk radion) training
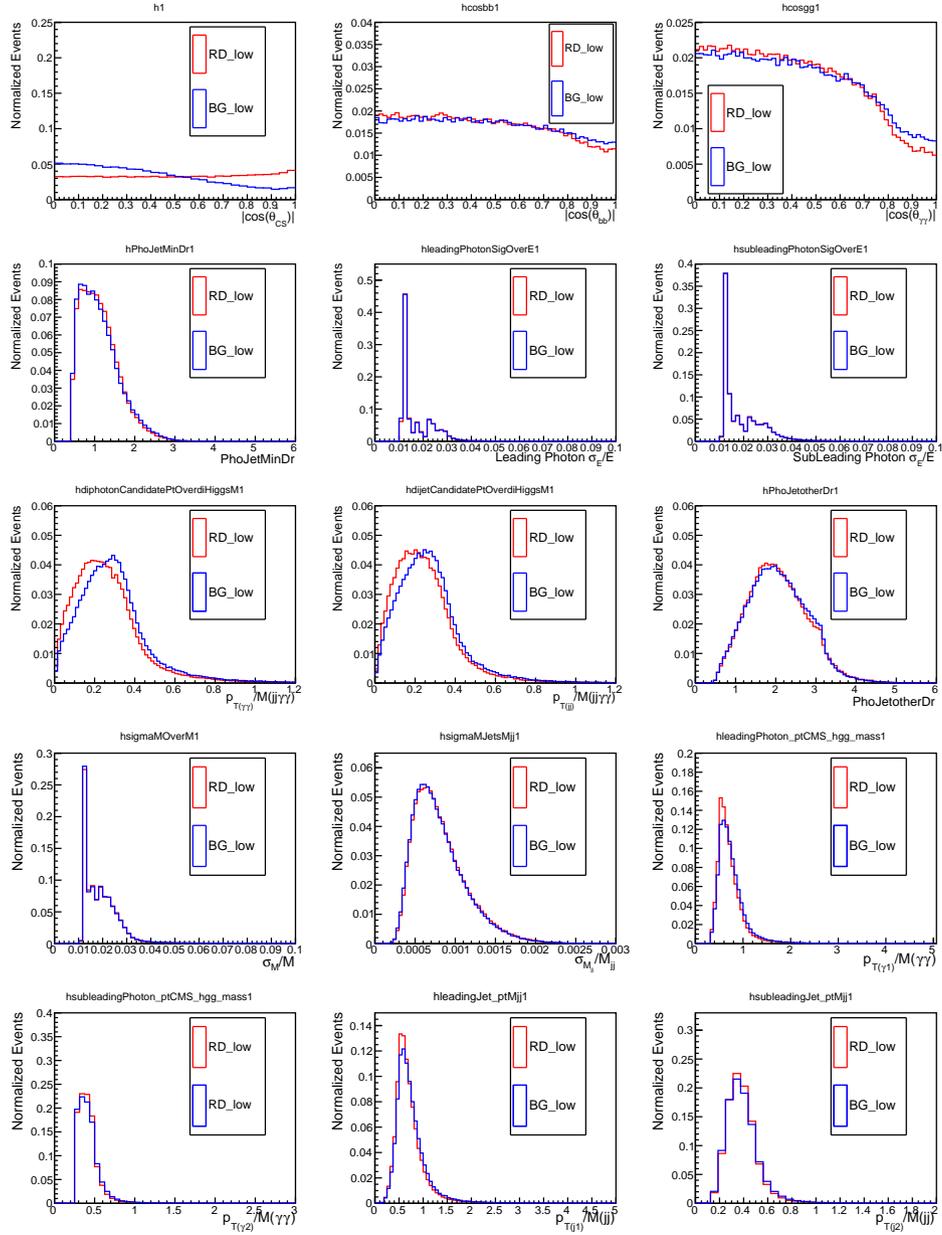
Figure 5.21: Kinematic comparisons between spin-0 bulk radion (RD) and spin-2 bulk KK graviton (BG) for lowX ranges

5.3.5.7 MVA Categorization:

To boost the sensitivity of the analysis, we divide events into different categories based on the transformed MVA output. As explained in Section 5.3.5.1, we train the BDT for all three years together to have one training output. Similarly, we optimize event categories for all three years together. The MVA categories are optimized based on Punzi figure of merit (FOM) [177]. The simulation samples (non-resonant backgrounds and signal) of 2016, 2017, and 2018 data-taking years are used for this which are merged and scaled to corresponding cross section times luminosity value. For signal, 1 fb cross section is used to give the same weightage to all mass points.

A Root package Minuit [178] is used with MIGrad minimizer to decide final MVA boundaries. Minuit is a tool to find the minimum value of a multi-parameter function and analyze the shape of the function around the minimum. The number of iteration and tolerance are set manually. The input function is Punzi FOM $S_{eff}/(1+\sqrt{B})$, where S_{eff} and B are signal efficiency and background yields. The category optimization remains independent of assuming any signal cross section since Punzi FOM only uses signal efficiency. It prevents categorization bias towards any specific signal hypothesis, and category boundaries remain the same for all signals within a m_X-m_Y mass range. To serve the same purpose, $\tilde{M}_X$ selection (Section 5.3.6) is not considered during BDT category optimization and applied only during categorization. The minimizer decides boundaries on BDT output to get the maximum root-square-sum of Punzi FOM from all categories. Four boundaries are provided by this optimization, among which the background-dominated BDT region is always discarded, and the rest of the three are considered. The optimized boundaries are tabulated in Table 5.8. These categories are labeled as CAT 0, CAT 1 and CAT 2 depending upon the value of Punzi FOM. For CAT 0, CAT 1 and CAT 2, Punzi FOM ranges 0.01–0.05, 0.003–0.02 and 0.001–0.006, respectively. We have six sets of category boundaries corresponding to six training m_X-m_Y ranges.

During the optimization, a constraint is used to keep enough background statistics in each category for proper event modeling. It has been also checked that MVA boundary

optimization does not get affected by the scale factor used in Data/MC comparison.

Table 5.8: Boundaries for MVA categorization

mass range & category	lowX_lowY	midX_lowY	highX_lowY	midX_midY	highX_midY	highX_highY
CAT 2	[0.174, 0.329]	[0.213, 0.401]	[0.215, 0.304]	[0.180, 0.352]	[0.177, 0.239]	[0.129, 0.286]
CAT 1	[0.329, 0.627]	[0.401, 0.550]	[0.304, 0.500]	[0.352, 0.600]	[0.239, 0.350]	[0.286, 0.400]
CAT 0	[0.627, 1.000]	[0.550, 1.000]	[0.500, 1.000]	[0.600, 1.000]	[0.350, 1.000]	[0.400, 1.000]

The number of MVA categories was optimized for non-resonant analysis and three MVA categories were found to be enough for a good analysis sensitivity. We also tried with four categories just for a cross-check, but it affected the results <1%.

5.3.5.8   Mass sculpting checks for background:

While dealing with the background within each MVA category, we need to ensure that the categorization does not sculpt the shape of the background; otherwise, it could be problematic while modeling the background distributions. Therefore, we compared the $m_{\gamma\gamma}$ and m_{jj} distributions among all three categories using non-resonant background simulations to check the effect of categorization. All the shape comparison plots are in Figs. 5.22 and 5.23.

We observe that for highX mass ranges, BDT sculpts the background shapes a bit, but background modeling can be handled by the envelope method [179]. For lowX and midX, it seems to work fine. Overall we do not see any turn-on or other issue with the $m_{\gamma\gamma}$ and m_{jj} shapes of the background within each category.

5.3.6   Resonant background rejection

The single Higgs backgrounds are the essential resonant backgrounds for the HH/HY signal process, and among them, the most dominant background is ttH. For these resonant searches, it contributes only for resonance mass < 600 GeV. As we tend to the higher resonance masses, resonant background contamination becomes less than 1%. Thus, it is required to remove ttH contribution to improve results for the lower m_X region where analysis has been proven more sensitive from previously published results.
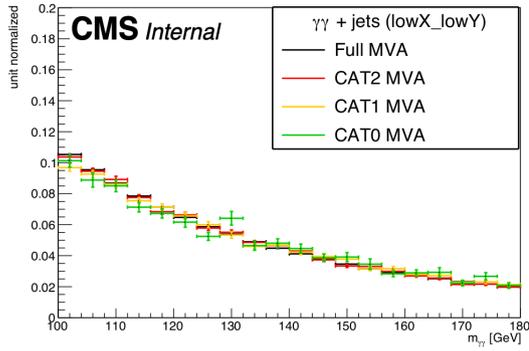
(a) lowX_lowY $m_{\gamma\gamma}$
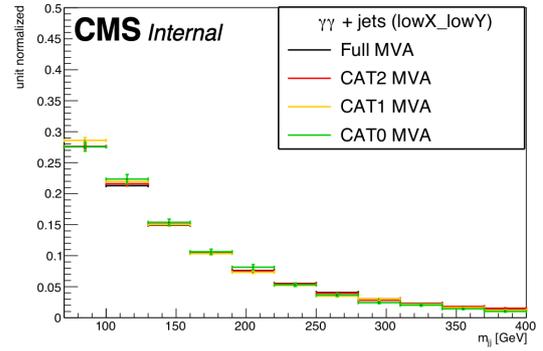
(b) lowX_lowY m_{jj}
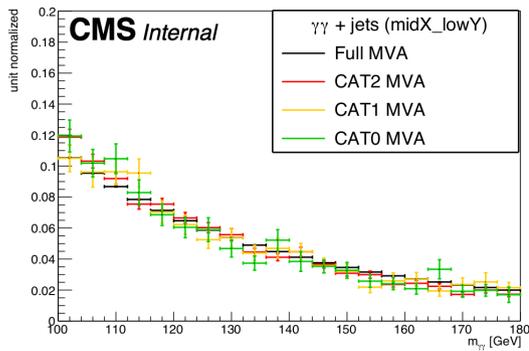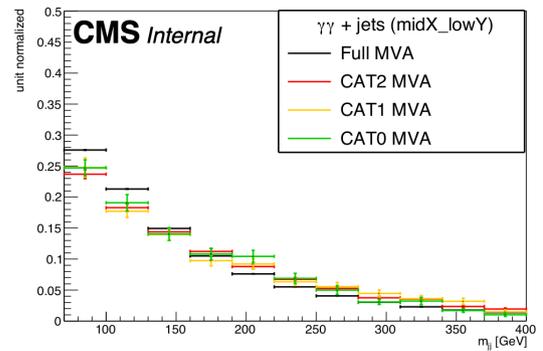
(c) midX_lowY $m_{\gamma\gamma}$

(d) midX_lowY m_{jj}
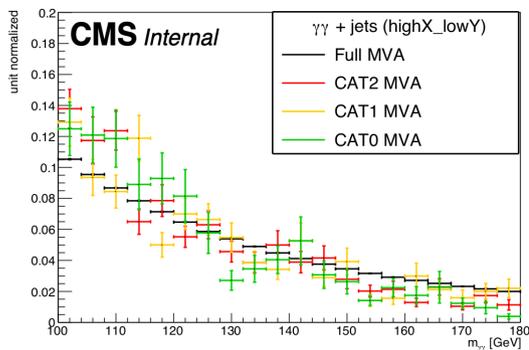
(e) midX_midY $m_{\gamma\gamma}$
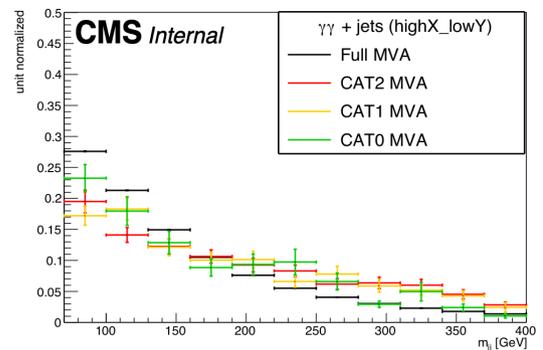
(f) midX_midY m_{jj}

Figure 5.22: Background MC $m_{\gamma\gamma}$ and m_{jj} shape comparisons within each MVA category for all six training.

(a) highX_lowY $m_{\gamma\gamma}$

(b) highX_lowY m_{jj}

(c) highX_midY $m_{\gamma\gamma}$

(d) highX_midY m_{jj}

(e) highX_highY $m_{\gamma\gamma}$

(f) highX_highY m_{jj}
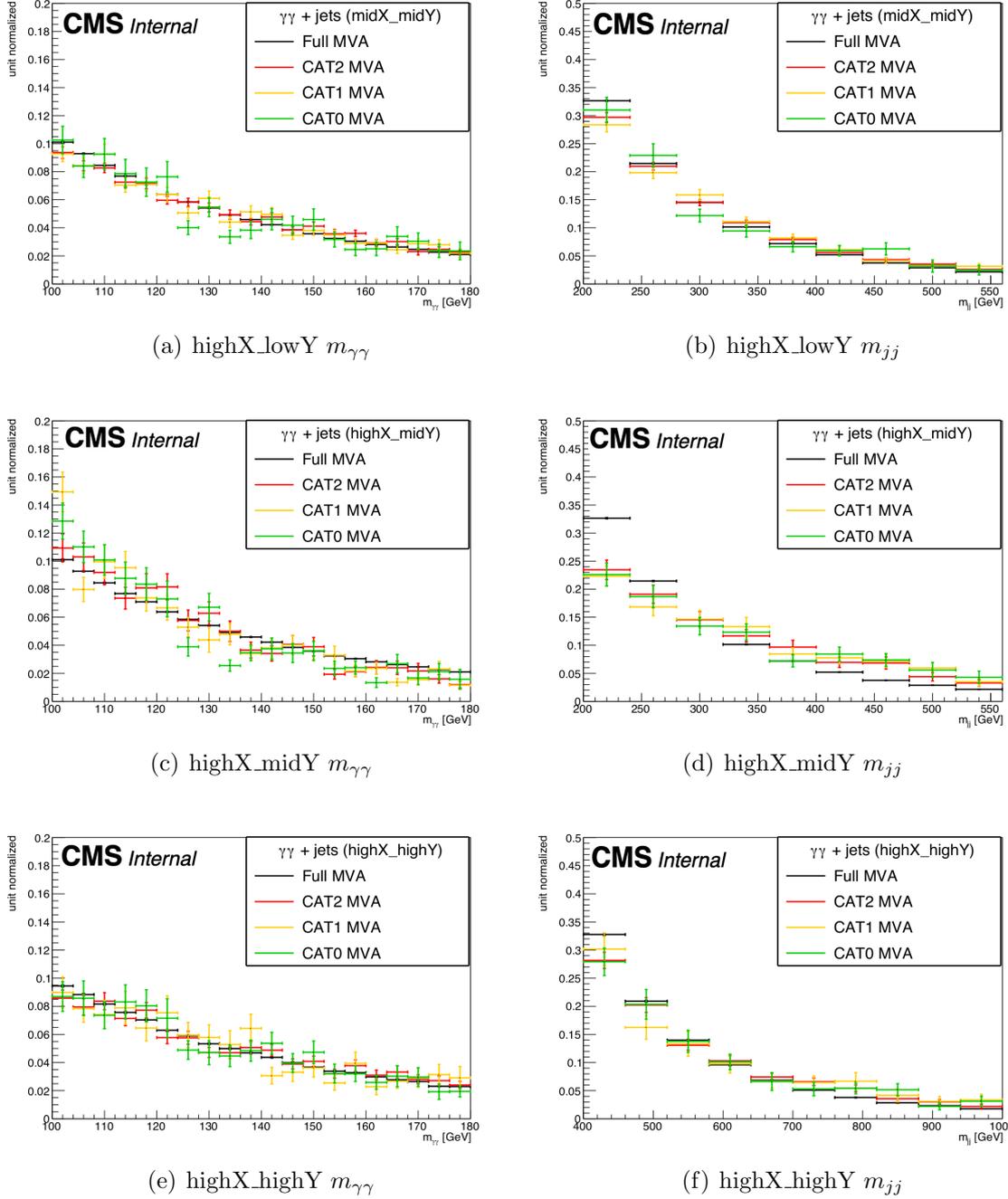
Figure 5.23: Background MC $m_{\gamma\gamma}$ and m_{jj} shape comparisons within each MVA category for all six training.

There are two following selections which help to achieve this motive:

- **ttHkiller discriminant:** A neural network discriminator was developed for non-resonant $HH \to b\bar{b}\gamma\gamma$ analysis, which is named as ttHkiller or ttHScore. For the non-resonant analysis, the optimization study to select a cut on this discriminant was performed using pseudo data to increase the analysis sensitivity. This study has been briefly added in Appendix B.3. The study concluded that a 0.26 cut on ttHScore could make up to 10% improvement in the final results. Therefore, it is directly adapted and applied for resonant analysis (only for resonance masses < 600 GeV). With the inclusion of this cut, 80% ttH background gets rejected, keeping 95% signal efficiency and improvement on low mass resonance results goes up to 10% as shown in Fig. 5.24. After ttHkiller cut (as checked using background simulations), only 1% resonant background contribution of total background in three categories contaminates the signal. This number goes around 6-7% for signal dominated region $m_{\gamma\gamma} = [120, 130]$ GeV.

  In Fig. 5.26, the shapes of $m_{\gamma\gamma}$ and m_{jj} are checked for background MC to observe how it gets modified with various ttHkiller thresholds. No turn-on is observed. It should also be mentioned that final results do not strongly depend on the choice of ttHKiller threshold (Fig. 5.25)[5].

---

[5]For low mass resonances, the effect of ttHkiller cut on MVA categorization is checked. We find the same MVA boundaries before and after the ttHkiller cut.
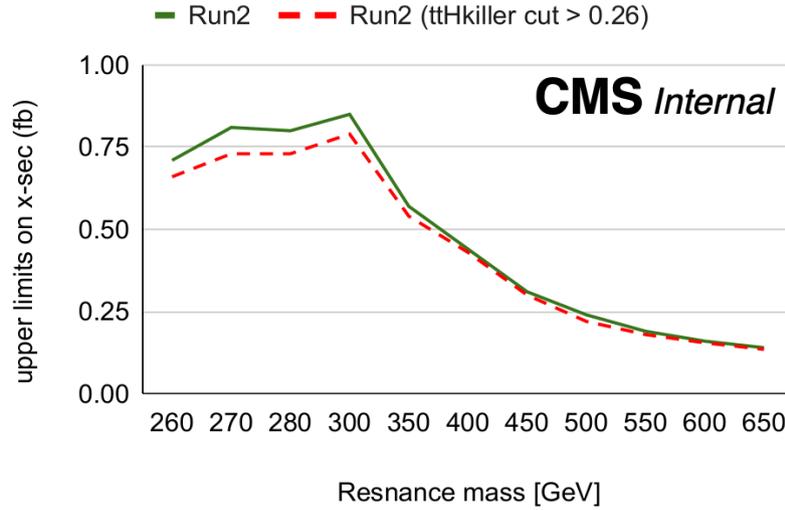
Figure 5.24: Effect of ttHkiller cut on upper limits of low resonance masses
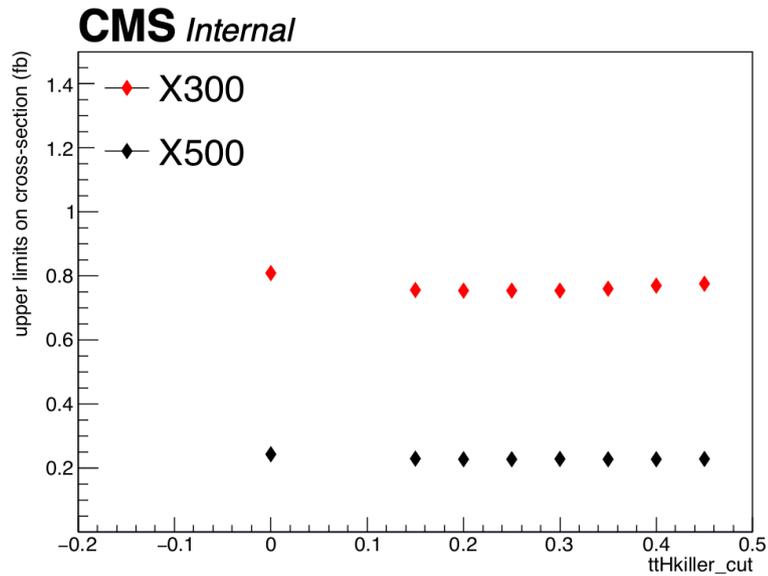
Figure 5.25: Variation in upper limits with ttHkiller thresholds

(a) lowX_lowY $m_{\gamma\gamma}$

(b) lowX_lowY m_{jj}

(c) midX_lowY $m_{\gamma\gamma}$

(d) midX_lowY m_{jj}
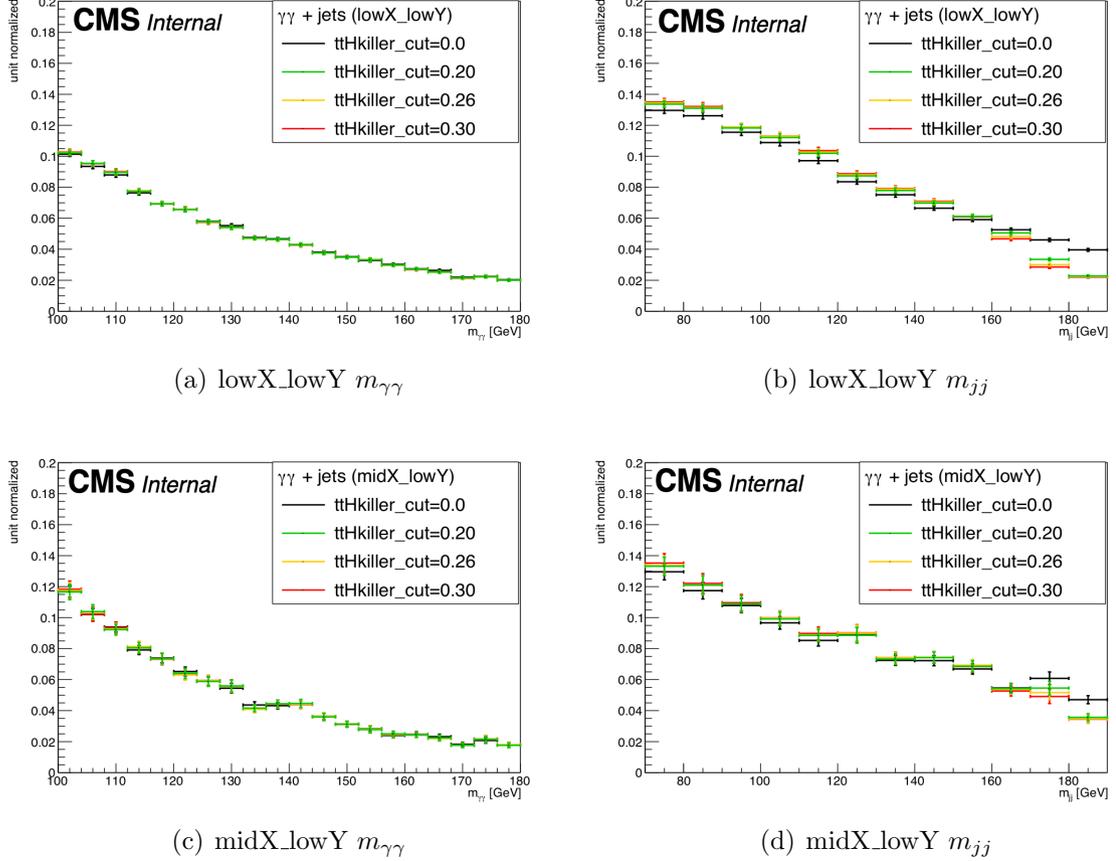
Figure 5.26: Effect of various ttHkiller thresholds on $m_{\gamma\gamma}$ and m_{jj} shapes

- **Mass window optimization:** Apart from the ttHkiller selection, $\tilde{M}_X$ mass window selection (adapted from 2016 published analysis [64]) also helps to suppress resonant background contribution. By construction, $\tilde{M}_X$ is 4-body invariant mass where mass resolutions of $m_{\gamma\gamma}$ and m_{jj} are subtracted from $m_{\gamma\gamma jj}$. It yields better $\tilde{M}_X$ resolution in comparison to $m_{\gamma\gamma jj}$. The improvement in $\tilde{M}_X$ resolution depends on the boost factor of the signal. For the signal, where decay products are in high p_T regime, $m_{\gamma\gamma}$ and m_{jj} resolutions are small; hence, their subtraction improves $\tilde{M}_X$ resolution only around 30%. As we tend to the signal having decay products in low p_T regime, $m_{\gamma\gamma}$ and m_{jj} resolutions get large, which subtraction brings this improvement up to 90%. The $\tilde{M}_X$ distributions are given in Figs. 5.28 and 5.29.

After $\tilde{M}_X$ reconstruction, we optimize a tight mass window selection on it for each

m_X using HH signals. The optimization covers more than 60% of signal events and achieves the best expected results using data from the control region[6]. With good $\tilde{M}_X$ resolution, this mass window selection helps to reject background keeping the high signal efficiency, thus increasing the signal-to-background ratio.

The optimized mass window on HH signals is directly applied on HY signals. For low m_Y, $\tilde{M}_X$ resolution of HY signals remains similar to HH signals. While for high m_Y, $\tilde{M}_X$ keeps up to 90% improvement in the resolution that gives high signal efficiency with the same mass window. Also, it is checked that the optimized mass window does not significantly affect the correlation between $m_{\gamma\gamma}$ and m_{jj} observables. The $\tilde{M}_X$ selection boundaries are shown in Fig. 5.27.

To verify the resonant background suppression, we compare the single Higgs background yields between resonant analysis after ttHkiller and $\tilde{M}_X$ selections and non-resonant analysis where we apply ttHkiller cut but do not have this $\tilde{M}_X$ selection. In Fig. 5.30, the bar graph shows that the single Higgs resonant backgrounds become negligible for resonant di-Higgs analysis after $\tilde{M}_X$ selection. It makes resonant backgrounds contribution less than 1% of total background.
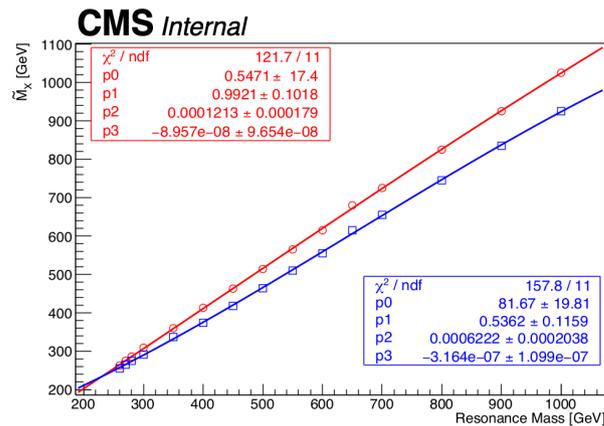
Figure 5.27: Mass window for each resonance mass; Red and Blue graphs show upper and lower boundaries of the $\tilde{M}_X$ selection

---

[6]We used blinded data for $\tilde{M}_X$ window optimization since this is a very tight selection, and it becomes crucial as we tend for higher masses where data events are very less. Thus to have enough data events to perform robust background modeling, it is done in this way.

(a) Y=90 GeV

(b) Y=100 GeV

(c) Y=125 GeV

(d) Y=150 GeV

(e) Y=200 GeV

(f) Y=250 GeV
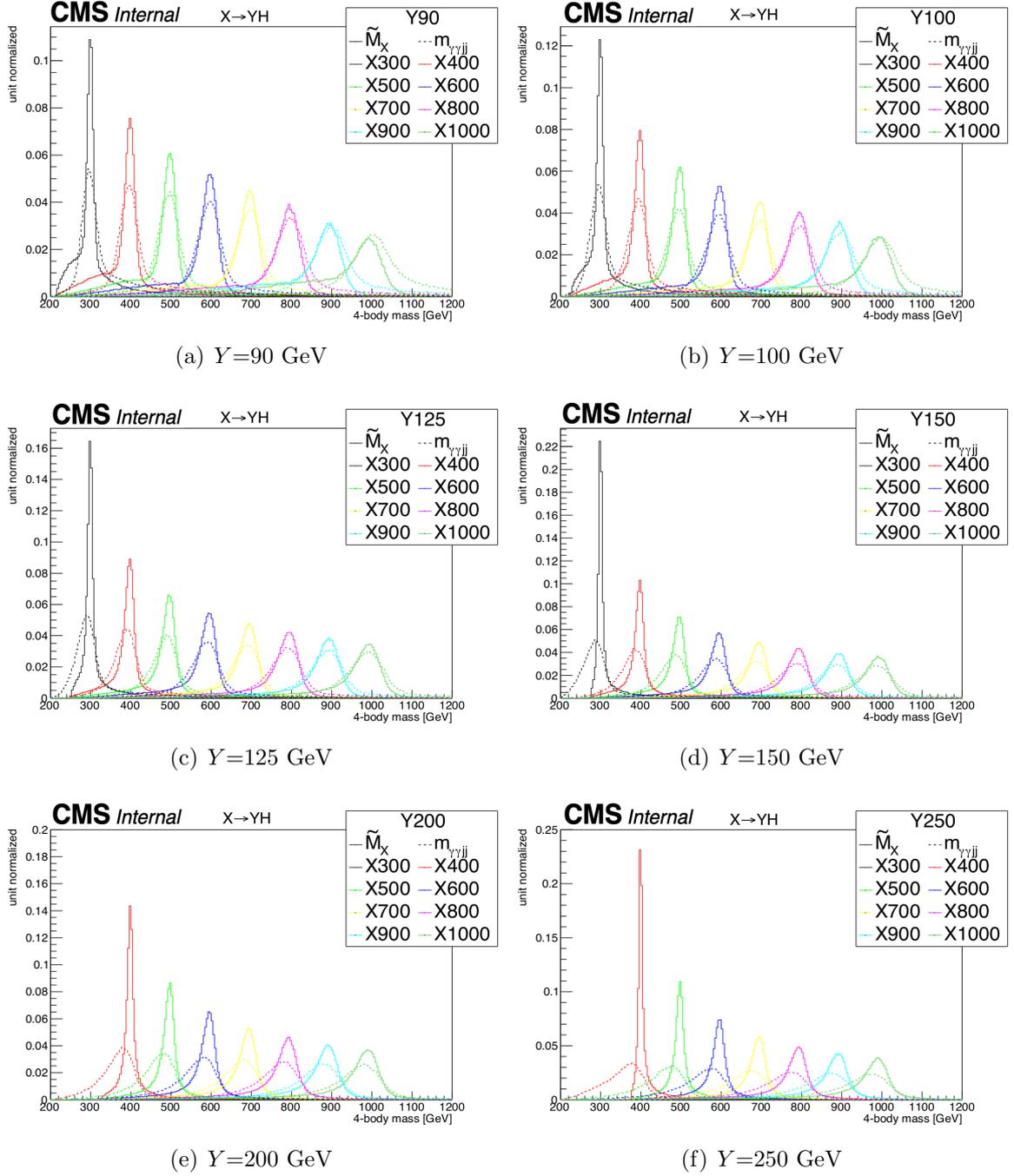
Figure 5.28: $\tilde{M}_X$ and $m_{jj\gamma\gamma}$ distribution comparison among different X resonance mass for each Y mass point; $\tilde{M}_X$ yields better resolution

(a) Y=300 GeV

(b) Y=400 GeV

(c) Y=500 GeV

(d) Y=600 GeV

(e) Y=700 GeV

(f) Y=800 GeV
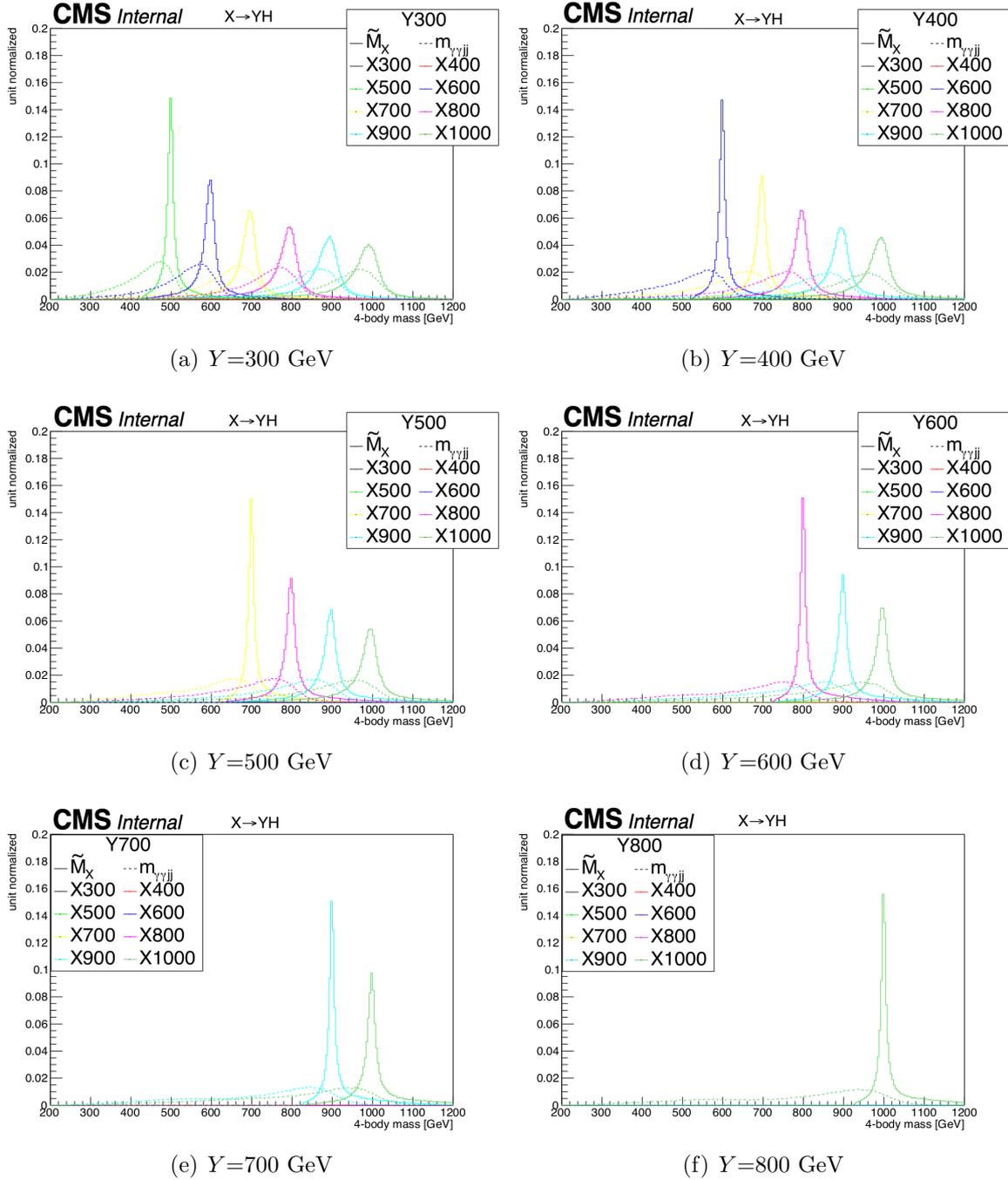
Figure 5.29: $\tilde{M}_X$ and $m_{jj\gamma\gamma}$ distribution comparsion among different X resonance mass for each Y mass point; $\tilde{M}_X$ yields better resolution
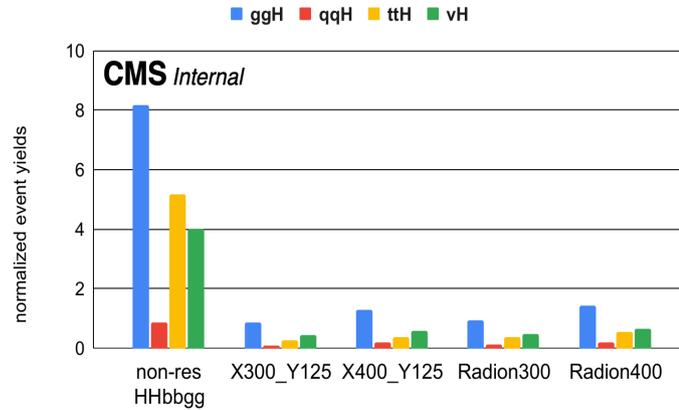
Figure 5.30: Single Higgs yields comparison between resonant and non-resonant analysis showing benefit of $\tilde{M}_X$ selection which reduces resonant background contribution for resonant $b\bar{b}\gamma\gamma$ analysis.

5.4 Signal model

To estimate the expected signal strength, signal simulations are used. After passing the event selections, categorized events are fitted in $m_{\gamma\gamma}$:m_{jj} plane with a product of two parametric signal models: a sum of Gaussian distributions and a double-sided Crystal Ball (CB) or sum of CB and Gaussian functions.

The $m_{\gamma\gamma}$ shape is parameterized using the sum of up to five Gaussian functions, where the number of Gaussian functions is chosen from the F-test [180]. The m_{jj} shape is modeled using a double-sided crystal ball (CB) function or sum of CB and Gaussian function depending upon the χ^2 of fit [181]. This signal model is good enough to model the $m_{\gamma\gamma}$ with higher resolution as well as m_{jj} with lower resolution for each resonance mass of signal hypothesis. The final signal model is a product of $m_{\gamma\gamma}$ and m_{jj} fit functions.

The no-correlation hypothesis is checked by comparing the two-dimensional $m_{\gamma\gamma}$ and m_{jj} distributions from the simulated signal samples (Fig. 5.36). For the average expected number of signal events in this analysis, the impact of such correlations are found to be negligible. This is expected from the fact that Higgs bosons decay independently and are not angularly correlated.

Signal fit for both $m_{\gamma\gamma}$ and m_{jj} are available in Figs. 5.31 and 5.32. These fits are only for the high purity category CAT 0 for each resonance mass.

(a) bulk radion300   (b) X700Y500   (c) X1000Y800
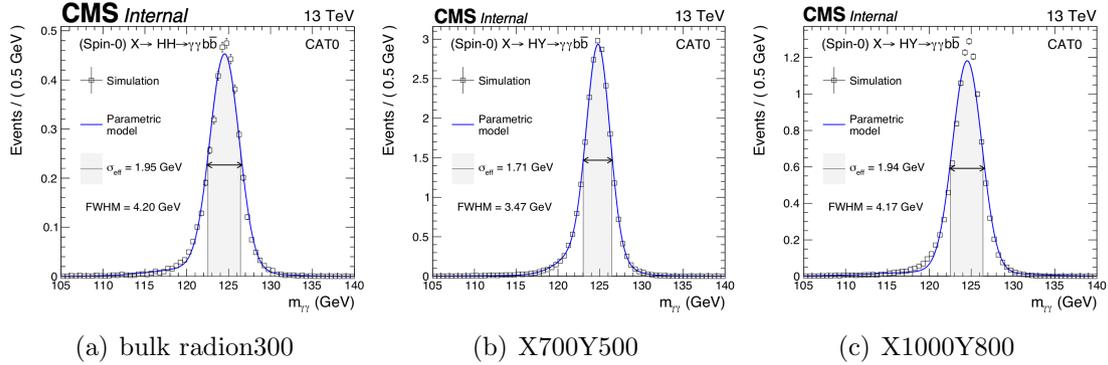
Figure 5.31: $m_{\gamma\gamma}$ model for three different Y mass region

(a) bulk radion300   (b) X700Y500   (c) X1000Y800
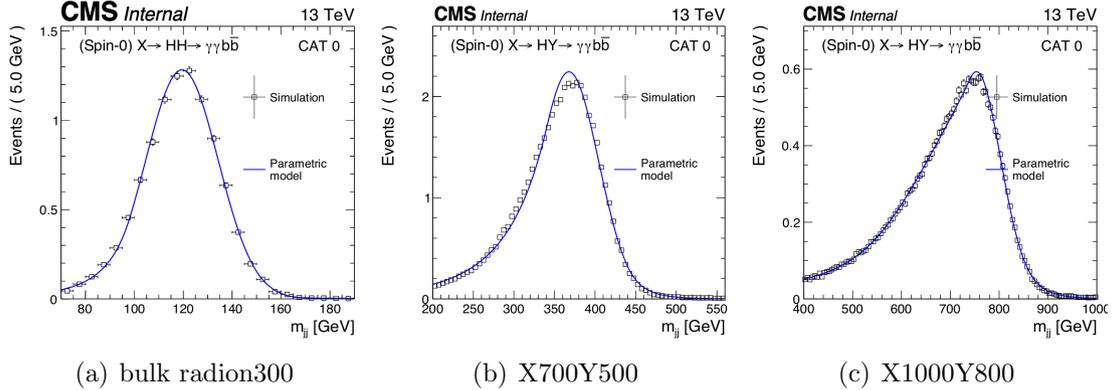
Figure 5.32: m_{jj} model for three different Y mass region

5.5   Background model

The total background model is built as a sum of the non-resonant diphoton QCD background continuum and single Higgs boson production. The probability density for the resonant single Higgs boson production is built using MC simulations and normalized to the corresponding SM production cross section. The non-resonant background is estimated in a data-driven way. The nominal background model is built as a factorized two-dimensional probability density function in the $m_{\gamma\gamma}{:}m_{jj}$ plane. For signal extraction,

we perform a 2D fit in this plane. The fit region for $m_{\gamma\gamma}$ observable is defined as 100-180 GeV. For m_{jj}, the fit region depends upon the mass of Y for the NMSSM scenario. For WED one, it is kept the same as non-resonant $HH \to b\bar{b}\gamma\gamma$ analysis (70–190 GeV). A bias study is also performed to validate the 2D envelop method for background modeling, which is explained within Section 5.5.3.

5.5.1 Single Higgs background model

For the single Higgs mechanisms, the $m_{\gamma\gamma}$ from simulations is modeled using the same approach as signal $m_{\gamma\gamma}$ modeling in Section 5.4, using the sum of Gaussian functions since two photons is the decay product of the SM Higgs boson.

However, their m_{jj} distributions differ from the signal, and an appropriate function is chosen based on the production process. We use Crystal ball (for VH), Gaussian (for ttH) and Bernstein (for ggH, $b\bar{b}H$, VBF H) function depending upon the background process. For VH, the additional two jets from the hadronic decay of vector bosons might mimic like signal b jets. The asymmetric distribution with a peak around vector boson mass 80–90 GeV can be modeled with the CB function. For ttH, the two b jets coming from top quark decay might gain (2*mass of top)/3 fraction of its energy-momentum. Therefore, a Gaussian function with a peak around 115 GeV is used to model this m_{jj} distribution. This is the most dominating single Higgs contribution, which fully mimics the signal with low Y mass. The rest of the three productions (ggH, VBF H, $b\bar{b}H$) is modeled with Bernstein polynomials as additional jets from these processes do not come from any massive particle; hence, they show a falling m_{jj} spectrum in the m_{jj} region of analysis interest. The representative plots are added in Fig. 5.33. Because of limited statistics in the single Higgs simulations after the categorization, it is hard to construct the shapes in each category per year. In $m_{\gamma\gamma}$ distribution, for all single Higgs production modes, a narrow Higgs peak is observed. Therefore, if there are not enough MC statistics in a given category, the $m_{\gamma\gamma}$, shape constructed for ttH is used, which has good MC statistics in all categories. For m_{jj}, the events are combined from simulations of all three years to have robust modeling.

(a) ttH CAT2 $m_{\gamma\gamma}$

(b) ggH CAT2 $m_{\gamma\gamma}$

(c) VH CAT2 $m_{\gamma\gamma}$

(d) qqh CAT2 $m_{\gamma\gamma}$

(e) $b\bar{b}H$ _yb2 CAT2 $m_{\gamma\gamma}$

(f) $b\bar{b}H$ _ybyt CAT2 $m_{\gamma\gamma}$

(g) ttH CAT2 m_{jj}

(h) ggH CAT2 m_{jj}

(i) VH CAT2 m_{jj}

(j) qqh CAT2 m_{jj}

(k) $b\bar{b}H$ _yb2 CAT2 m_{jj}

(l) $b\bar{b}H$ _ybyt CAT2 m_{jj}
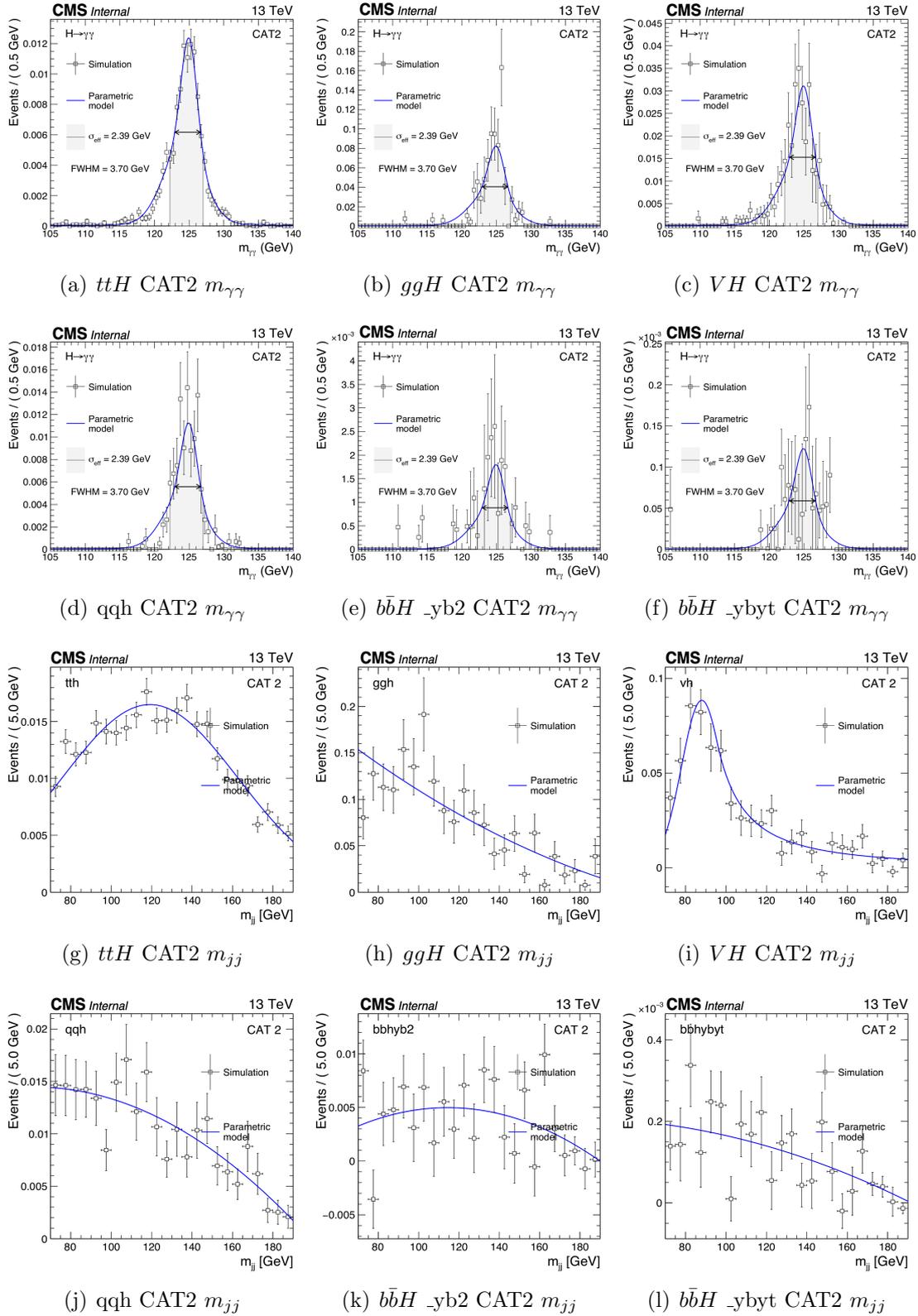
Figure 5.33: Resonant background model

5.5.2   Non-resonant background model

A discrete profiling method [179] is used to model the data-driven diphoton QCD continuum. Whenever we study a rare signal, the choice of background and signal functions change the results; thus, there is always uncertainty associated with the choice of functions. The method considers this choice as a source of systematic uncertainty modeled as a nuisance parameter. For the discrete profiling method, three families of the functions are considered: polynomials in the Bernstein basis, exponential, and power-law functions. For each of the analysis categories, F-test [180] selects the right dof of the functions from each of these families to proceed with discrete profiling by comparing $2\Delta NLL$ between two consecutive fits. This $2\Delta NLL$ is distributed as a $\chi^2(n)$ distribution, where the number of dof n equals the difference in the number of free parameters between two consecutive orders. We continue to increase the order of the function until the next order gives no significant improvement to the fit to the data. It is done for $m_{\gamma\gamma}$ and m_{jj} variables separately, and then, combinations of best fit functions are chosen for $m_{\gamma\gamma}$ and m_{jj} projections. The choice of a function to define the background is regarded as a discrete nuisance parameter in the final $2\Delta NLL$ minimization. Choosing the minimum envelope of selected functions creates a systematic uncertainty on the background modeling choice.

The m_{jj} and $m_{\gamma\gamma}$ distributions in data are shown in Fig. 5.34 together with a set of candidate functions obtained with F-test[7].

---

[7]For background modeling, data events with $m_{\gamma\gamma}$ side-band are considered for m_{jj} modeling. m_{jj} region is not blinded. Also, since we have different Y-masses, signal modeling for m_{jj} is done in Y specific m_{jj} range where it covers full m_{jj} shape corresponding to that Y mass-point.
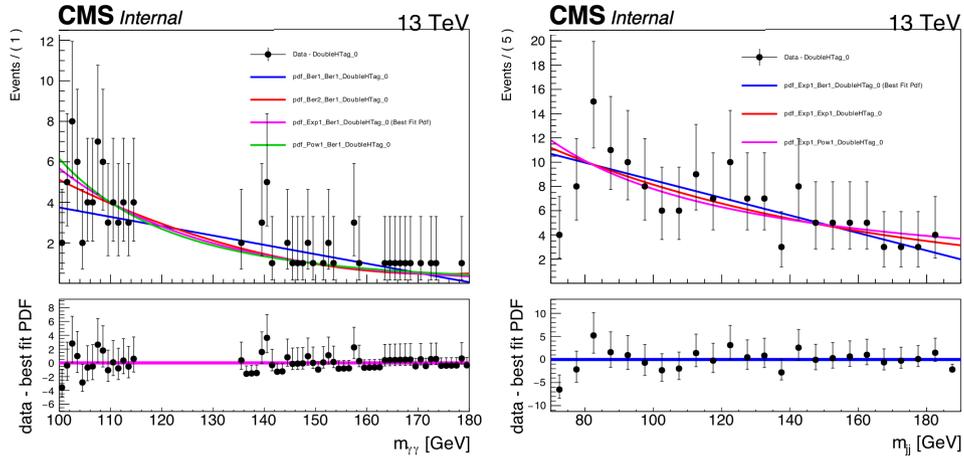
(a) bulk radion300 CAT 0 $m_{\gamma\gamma}$

(b) bulk radion300 CAT 0 m_{jj}
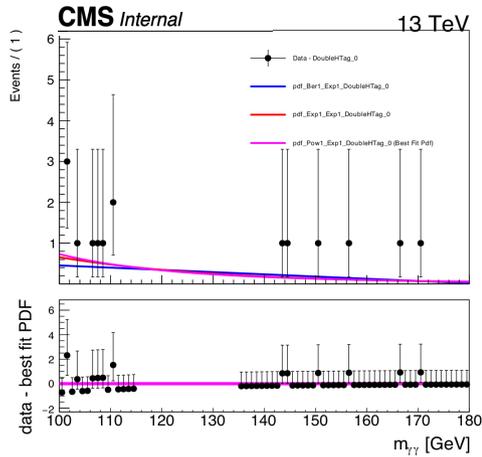
(c) X800Y400 CAT 0 $m_{\gamma\gamma}$

(d) X800Y400 CAT 0 m_{jj}
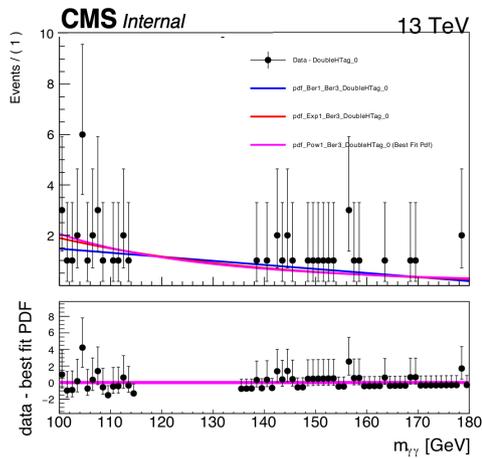
(e) X1000Y800 CAT 0 $m_{\gamma\gamma}$
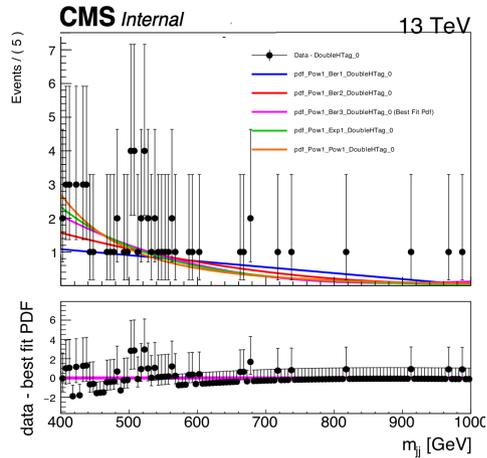
(f) X1000Y800 CAT 0 m_{jj}

Figure 5.34: Background model for CAT 0 in three different Y mass region

5.5.3 Bias study

After deciding the function for signal and background models, it is required to perform the bias study, which indicates how much potential bias can be present given a certain choice of event modeling functions. These studies are conducted using Higgs Combine Tool [182]. The steps to perform the bias study are described below:

- signal injection value $(\mu_{inj}) = 1.0$

- generates 1500-2000 toys using the alternative choice of background modeling functions $(\mu_{inj}.s + b)$ which could also define data side-bands well. We could do this by switching the pdf index (pdf index refers to a different order of the other background modeling function families) for the envelope method.

- Fit the toys using final background model choice from envelope method with fit parameter μ range [-5, 5].

- bias is extracted from fitted toys using ($(\mu_{fit} - \mu_{inj})/\sigma_{fit}$) where σ_{fit} is uncertainty on fitted μ.

The reason behind choosing signal injection $= 1$ is to deal with negative pdf error as in some bins background events are less, which leads to a negative value of $\mu_{inj}.s + b$ if μ_{inj} is set equal to the expected upper limit value.

Also, since the choice of background modeling functions remains the same for WED signals, it is enough to perform this study using any signal as both signals differ in spin structure, but that does not alter the m_{jj}, and $m_{\gamma\gamma}$ shapes much. For the NMSSM scenario, the bias study is performed for the masses, representing all the six mass ranges as we define for training.

All the results can be found in Figs. B.9, B.10 and B.11 for both WED and NMSSM signals. In the results, where the bias is not under the 14-20% set threshold with a few background modeling functions, those are the alternative functions that do not define background side-bands well; thus, corresponding bias can be ignored. These functions are explicitly compared with the best fit pdfs in Table 5.9.

Table 5.9: Pdf functions showing high bias (3rd column) with best fit pdfs (last column)

Signal	CAT	pdfs	Best fit pdf
WED_X260	2	pdf_Pow1_Pow1_DoubleHTag_2	pdf_Exp1_Ber2_DoubleHTag_2
NMSSM X500_Y300	2	pdf_Ber3_Ber1_DoubleHTag_2, pdf_Ber3_Ber2_DoubleHTag_2, pdf_Ber3_Ber3_DoubleHTag_2, pdf_Ber3_Exp1_DoubleHTag_2, pdf_Ber3_Pow1_DoubleHTag_2	pdf_Pow1_Exp1_DoubleHTag_2
NMSSM X600_Y400	2	pdf_Ber3_Ber1_DoubleHTag_2, pdf_Ber3_Ber2_DoubleHTag_2, pdf_Ber3_Ber3_DoubleHTag_2, pdf_Ber3_Exp1_DoubleHTag_2, pdf_Ber3_Pow1_DoubleHTag_2	pdf_Pow1_Exp1_DoubleHTag_2
	1	pdf_Pow1_Ber3_DoubleHTag_1, pdf_Pow1_Pow1_DoubleHTag_1	pdf_Ber2_Exp1_DoubleHTag_2
NMSSM X800_Y600	2	pdf_Pow1_Exp1_DoubleHTag_2,pdf_Pow1_Pow1_DoubleHTag_2	pdf_Ber2_Pow1_DoubleHTag_2
NMSSM X900_Y700	2	pdf_Pow1_Exp1_DoubleHTag_2,pdf_Pow1_Pow1_DoubleHTag_2	pdf_Exp1_Pow1_DoubleHTag_2
NMSSM X1000_Y800	2	pdf_Pow1_Ber1_DoubleHTag_2, pdf_Pow1_Ber2_DoubleHTag_2, pdf_Pow1_Ber3_DoubleHTag_2, pdf_Pow1_Ber4_DoubleHTag_2	pdf_Exp1_Pow1_DoubleHTag_2

5.6   Systematic uncertainties

Several sources of systematic uncertainty are considered in the analysis. The systematic uncertainties affect mainly the signal model and the resonant single Higgs background. The non-resonant background model is constructed using a data-driven method with the uncertainties associated with the choice of background fit function using the discrete profiling method as described in Section 5.5. Systematic uncertainties can have a number of effects on signal and single Higgs background models. As parametric nuisance parameters, systematic uncertainties altering the shapes of the $m_{\gamma\gamma}$ and m_{jj} distributions are introduced into the models. The systematic uncertainties affecting only the yields are treated as log-normal uncertainties.

All the systematics are in common with the standard $H \rightarrow \gamma\gamma$ analysis and implemented in the same way. We give a summary of the dominant sources for this analysis. Since the analysis is statistically dominated, the effect of the systematics is of the order of 1–2%.

The major systematic uncertainties considered in this analysis are the following:

- *Jet energy scale and smearing corrections*: The p_T balance of jets with Z bosons and photons in $Z \rightarrow ee$, $Z \rightarrow \mu\mu$, and γ+jets events, as well as the p_T balance

between jets in dijet and multijet events, are used to determine the energy scale of jets. The jet energy scale has a small amount of uncertainty and is influenced by p_T and η. By varying the jet energy corrections within their uncertainties and propagating the effect to the final result, the impact of jet energy scale and smearing uncertainties on event yields is evaluated. It is found to be a size of 1%. These are considered uncorrelated across three data-taking years since Jet measurements are performed independently for 2016, 2017 and 2018 data.

- *Shape of the b discriminant*: To correct the shape of DeepJet discriminant coming from simulation to match with data, reshaping scale factors (SFs) are used, depending on p_T, η and flavor of the jets. These corrections are determined from multijet events, enriched in heavy-flavor content by requiring the presence of a muon, and in $t\bar{t}$+jets events. The uncertainty associated with this SFs measurement is uncorrelated across three years (since b discriminant training and corresponding SF measurements are independent in data-taking years) and obtained by comparing b discriminator distribution between data and simulation [183]. The impacts from different systematic uncertainty sources are studied and found to be highest where the light flavor jet gets mistagged as b jet. The average size of uncertainty varies from 4–6%. Other sources have size of less than 1%.

- *Background modeling*: The choice of background parameterization is handled using the discrete profiling method. This automatically leads to an uncertainty on the choice of background function as described in Section 5.5.

- *Parton density functions (PDF) uncertainties*: After re-weighting the simulated signal sample events, the impact of the uncertainty from the PDF choice is evaluated by measuring the relative yield variation in each process and category. The re-weighting is performed according to PDF4LHC15 Ref. [168], combined PDF set and NNPDF3.0 using the MC2hessian procedure in Ref. [167]. The overall normalization variation is taken from Ref. [184]. Therefore, it is studied with

correlating the uncertainties across three data-taking years. The average size of pdf uncertainties is around 2-3%.

- *QCD scale uncertainty:* It is related to checking the impact of varying renormalization and factorization scales. For signal, the uncertainty is -5/+2.2%.

- *Integrated luminosity*: These are included as described in CMS Refs. [185–187]. The uncertainty range from 1–2.5% for 2016–2018 data-taking years and has a combined impact of 1.6%. Following CMS recommendation, we consider uncertainties from both correlated and uncorrelated sources of luminosity measurements. Correlation is considered from common sources like Beam current calibration, beam-beam effect (2017–2018 only), x-y correlations of the beam shape, etc.

- *Trigger efficiency*: It is calculated with the tag-and-probe method using $Z \rightarrow ee$ events by considering the no correlation among three data-taking years. Its average size is around 1–2% and the impact on the final result is less than 1%. More details are given in Ref. [188].

- *Photon preselection*: It is studied on the ratio of efficiency measured in data and simulation after passing the preselections considering no correlations across three years. The tag-and-probe method is used for data preselection efficiency on $Z \rightarrow ee$ events [188]. Overall it has an impact of less than 1% on the final results.

- *Photon energy scale and resolution*: Electrons from $Z \rightarrow ee$ simulations are used to study scale and resolution corrections, which are then applied to photons. The different interactions of electrons and photons with the ECAL are the primary source of systematic uncertainty. Changes in the R_9 distribution, the regression training, and the electron selection utilized to create the corrections are used to quantify uncertainty. The additional energy smearing uncertainty is assigned, and the uncertainties on the various $|\eta|$ and R_9 bins are propagated to the Higgs boson signal phase space. In both cases, nuisance parameters are incorporated in the signal model as additional systematic terms and amount to less than a few percent

depending on the category. These uncertainty are not correlated across three data-taking years and directly applied on data for scale corrections and simulation for resolution corrections [189].

Apart from the above-listed ones, the impact of other systematic sources was observed negligible. We calculate the effect of these systematics in all six mass ranges where we train BDT as some systematic uncertainties depend on the kinematics, which differs in every mass range. We find the highest impact on the lowX_lowY mass range. The corresponding prefit impacts are shown in Fig. 5.35. Apart from this, JetHEM (0.2-0.5%) and prefire-probability (varies from 0.2-1%, large for high masses) impacts are also considered.
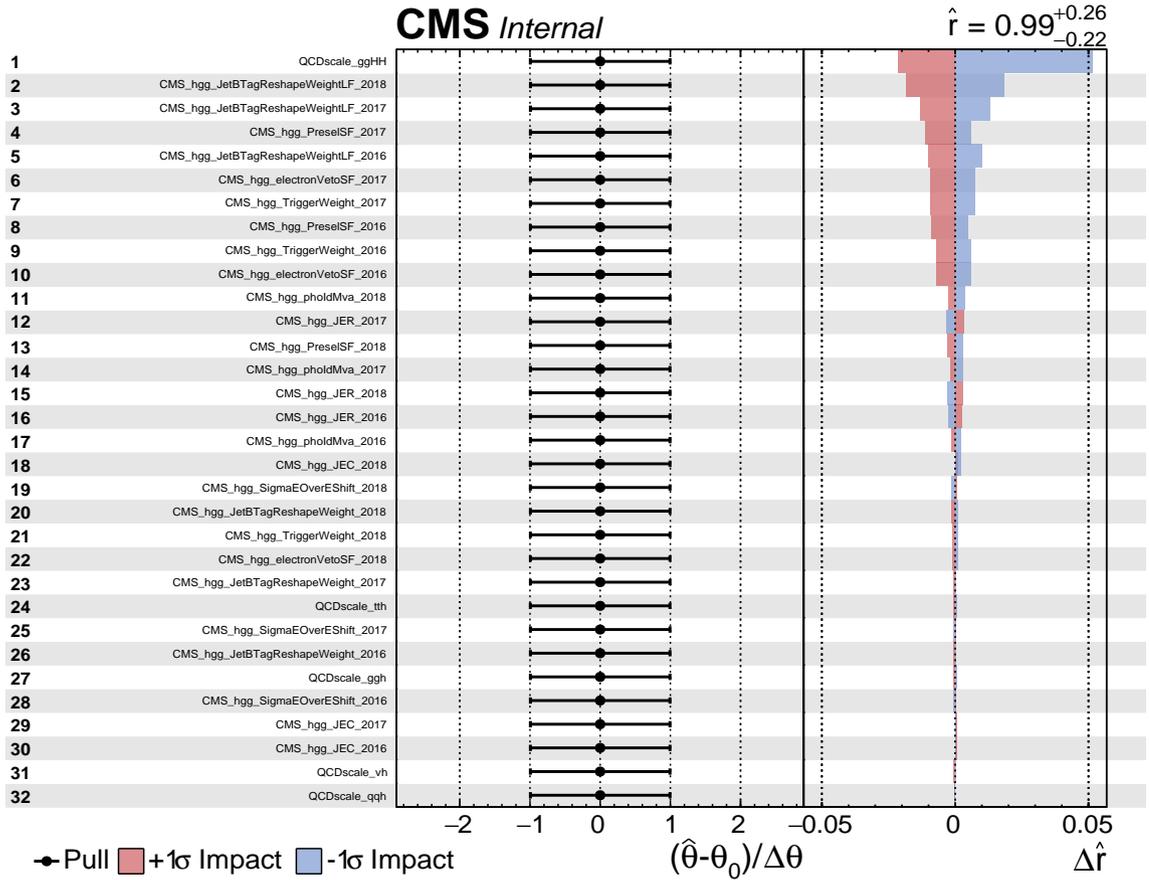
Figure 5.35: Prefit impacts for combined fits of $m_{\gamma\gamma} \times m_{jj}$ distributions for low mass range

5.7   Signal extraction and results

Two possible 2D-fit signal extraction methods could be used within this analysis. One is simultaneous fit in $m_{\gamma\gamma}$:m_{jj} plane, and another one is $m_{\gamma\gamma}$:$\tilde{M}_X$ 2D-fit. We performed a few cross-checks to adapt the final method, which is explained in Section 5.7.1. This study concludes that $m_{\gamma\gamma}$:m_{jj} 2D-fit is the more optimal one.

5.7.1   No $m_{\gamma\gamma} : \tilde{M}_X$ fit; why?

Correlation issue:

We made a comparison between the correlation factors of $m_{\gamma\gamma}$:$\tilde{M}_X$ and $m_{\gamma\gamma}$:m_{jj} variables after the selections in Fig. 5.36. It is observed that for $m_{\gamma\gamma} : m_{jj}$, correlation is always negligible but for $m_{\gamma\gamma} : \tilde{M}_X$, it is getting increased as we tend to higher X masses.
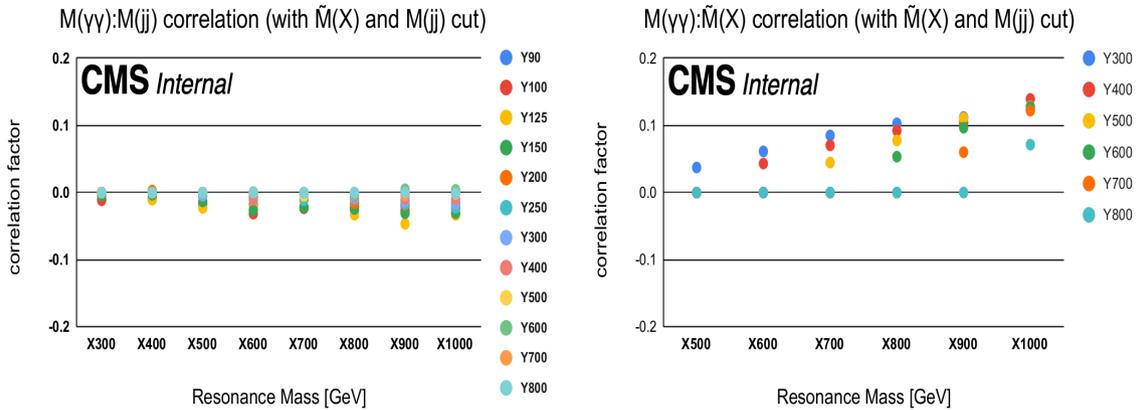
Figure 5.36: Correlation factor for $m_{\gamma\gamma} : m_{jj}$ (left) and $m_{\gamma\gamma} : \tilde{M}_X$ (right)

Turn-on issue

Figure 5.37 has shape comparisons for $\tilde{M}_X$ distributions, coming from the signal and background processes. For low X masses, the turn-on seems to be an issue to model the background for $m_{\gamma\gamma}$:$\tilde{M}_X$ 2D signal extraction method. This also supports not to use $m_{\gamma\gamma}$:$\tilde{M}_X$ 2D-fit method.

Following the above study, signal extraction and limit setting in this analysis is performed with a 2D-fit in the $m_{\gamma\gamma}$:m_{jj} plane. However, for the background, they are
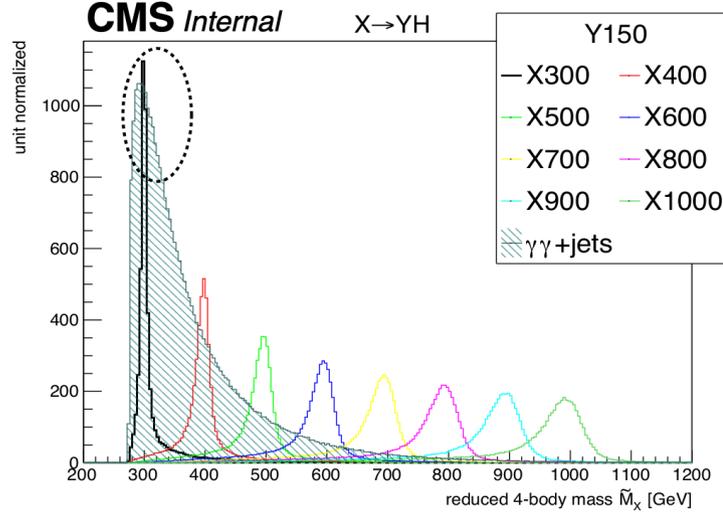
Figure 5.37: Background turn-on around signal low X masses in $\tilde{M}_X$ distributions

expected to be uncorrelated. For non-resonant $HH \rightarrow b\bar{b}\gamma\gamma$ analysis, a detailed study was performed using toy MC and data which showed no correlations between the mass dimensions. With this assumption, we can construct background function models as $f(m_{\gamma\gamma})*f'(m_{jj})$, where $f(f')$ are our functional choices to fit the diphoton (dijet) mass spectrum. This 2D approach constrains the impact of the single Higgs backgrounds since their structure in m_{jj} differs from the signal.

5.7.2 Results

For 2D $m_{\gamma\gamma}$:m_{jj} fit signal extraction method, a likelihood function is defined using signal and background analytic models of $m_{\gamma\gamma}$ and m_{jj} distributions with nuisance parameters related to the uncertainties explained in Section 5.6. A simultaneous unbinned maximum likelihood fit to the $m_{\gamma\gamma}$ and m_{jj} observable distributions is performed within three categories to extract the signal. Figures B.13 and B.14, representing the data and signal-plus-background fit for $m_{\gamma\gamma}$ and m_{jj} observables in CAT 0 after unbinding, are added in Appendix B.4. No deviation is observed from the background-only hypothesis; therefore, we set the upper limit at 95% CL on the product of resonant production cross section and branching ratio to $b\bar{b}\gamma\gamma$ channel. Figure 5.39 shows the observed upper limit at 95% CL 0.82–0.07 fb and 0.78–0.06 fb on production cross section of spin-0 and spin-2

resonances for HH searches, while the expected ranges are 0.74–0.08 fb and 0.65–0.06 fb, respectively. These results exclude masses up to 600 GeV for spin-0 bulk radion signal at $\Lambda_R =$6 TeV and up to 850 GeV for spin-2 bulk KK graviton signal with 0.5 coupling factor. Figure 5.40 puts 95% CL exclusion limit on $pp \to X \to YH \to b\bar{b}\gamma\gamma$ signal with different m_X and m_Y mass hypotheses. The corresponding observed and expected limits are 0.90–0.04 fb and 0.79–0.05 fb, respectively. Tables 5.10 and 5.11 show the observed (expected) upper limits for bulk radion, bulk KK graviton, and $X \to YH$ signals.

For $X \to YH$ signal, width is a pure experimental choice, so NMSSM interpretations are not made. The results show that limits follow a flat distribution for various mass hypotheses. However, they become worse for high Y masses as objects start to get into a boosted regime where the same analysis methods are not optimal. The m_X beyond 1 TeV is not studied because of the significantly better sensitivity of the 4b channel in this mass region.

In addition to this, Fig. 5.41 compares the spin-0 results between bulk radion and $X \to YH$ ($Y=$ 125 GeV) signals. The difference comes from b jet selection in the wider m_{jj} window for the latter case. It has been checked that if it is done with the same window as bulk radion, both signal hypotheses will give similar results.

Also, Y and H are different bosons that decay into different final states; therefore, we only cover half of the phase space for the $X \to YH$, $Y=$125 GeV signal in the $b\bar{b}\gamma\gamma$ analysis (since we set $H \to \gamma\gamma$ decay and $Y \to b\bar{b}$ decay at generator level; so we do not consider efficiencies for $Y \to \gamma\gamma$ and $H \to b\bar{b}$ case).

Table 5.10: Run-2 95% CL observed (expected) limits ($\sigma(X \to HH \to b\bar{b}\gamma\gamma)$) for WED spin-0 and spin-2 signal hypotheses

m_X [GeV]	(Spin-0) [fb]	(Spin-2) [fb]
260	0.82 (0.71)	0.78 (0.65)
270	0.74 (0.70)	0.61 (0.62)
280	0.69 (0.72)	0.63 (0.63)
300	0.47 (0.74)	0.41 (0.65)
320	0.46 (0.68)	0.40 (0.59)
350	0.73 (0.58)	0.63 (0.50)
400	0.39 (0.42)	0.34 (0.35)
450	0.23 (0.29)	0.17 (0.22)
550	0.26 (0.20)	0.20 (0.16)
600	0.16 (0.14)	0.12 (0.11)
650	0.28 (0.14)	0.22 (0.11)
700	0.26 (0.12)	0.21 (0.10)
800	0.07 (0.09)	0.06 (0.07)
900	0.14 (0.08)	0.11 (0.06)
1000	0.07 (0.08)	0.06 (0.06)

Table 5.11: Run-2 95% CL observed (expected) limits ($\sigma(X \to YH \to b\bar{b}\gamma\gamma$ [fb]) for NMSSM spin-0 X and Y signal hypotheses

mass [GeV]	Y90	Y100	Y150	Y200	Y250	Y300	Y400	Y500	Y600	Y700	Y800
X300	0.90 (0.79)	0.67 (0.73)	0.44 (0.64)	-	-	-	-	-	-	-	-
X400	0.52 (0.44)	0.46 (0.42)	0.19 (0.39)	0.37 (0.42)	0.45 (0.60)	-	-	-	-	-	-
X500	0.18 (0.29)	0.26 (0.27)	0.13 (0.23)	0.30 (0.21)	0.35 (0.25)	0.39 (0.27)	-	-	-	-	-
X600	0.15 (0.18)	0.26 (0.18)	0.21 (0.14)	0.13 (0.13)	0.09 (0.13)	0.35 (0.15)	0.27 (0.23)	-	-	-	-
X700	0.25 (0.16)	0.23 (0.15)	0.17 (0.13)	0.10 (0.10)	0.18 (0.11)	0.12 (0.08)	0.19 (0.11)	0.58 (0.25)	-	-	-
X800	0.20 (0.15)	0.18 (0.14)	0.11 (0.10)	0.10 (0.09)	0.06 (0.08)	0.08 (0.08)	0.07 (0.07)	0.15 (0.10)	0.18 (0.21)	-	-
X900	-	-	0.12 (0.10)	0.07 (0.09)	0.06 (0.07)	0.05 (0.06)	0.11 (0.07)	0.05 (0.06)	0.13 (0.10)	0.15 (0.18)	-
X1000	-	-	0.13 (0.10)	0.08 (0.08)	0.08 (0.07)	0.11 (0.06)	0.04 (0.05)	0.06 (0.06)	0.04 (0.06)	0.09 (0.09)	0.12 (0.16)
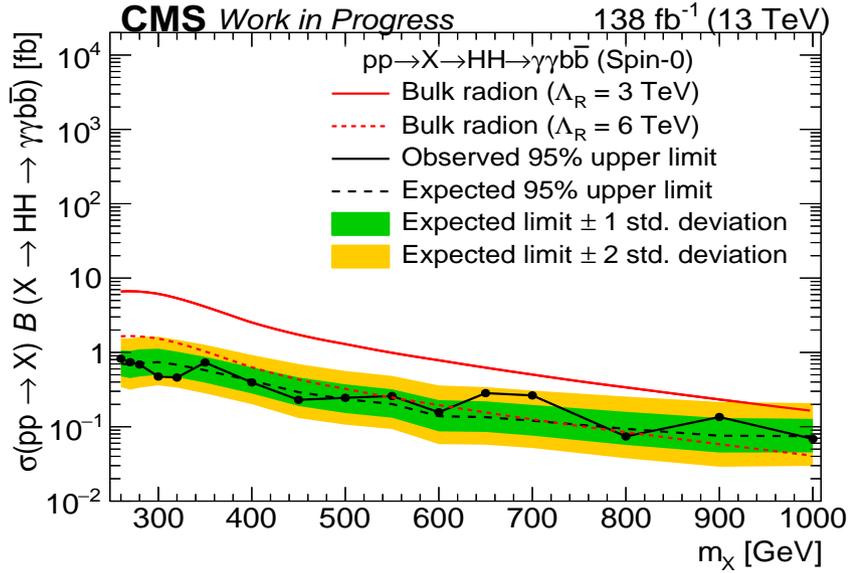
Figure 5.38: 95% CL observed (expected) upper limits on cross section for bulk radion (spin-0). The green and yellow bands represent the 1 and 2 standard deviation for expected limit. The red lines show the theoretical predictions.
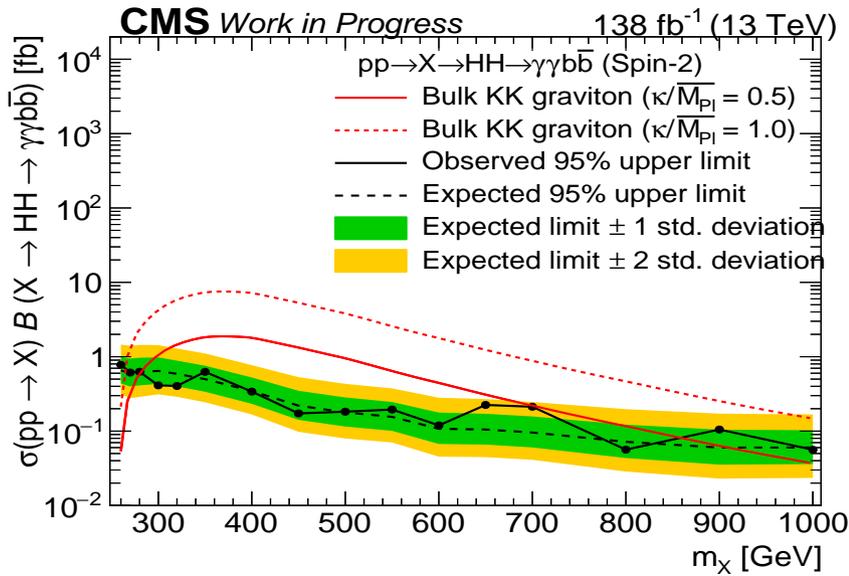
Figure 5.39: 95% CL observed (expected) upper limit on cross section for bulk KK graviton (spin-2). The green and yellow bands represent the 1 and 2 standard deviation for expected limit. The red lines show the theoretical predictions.
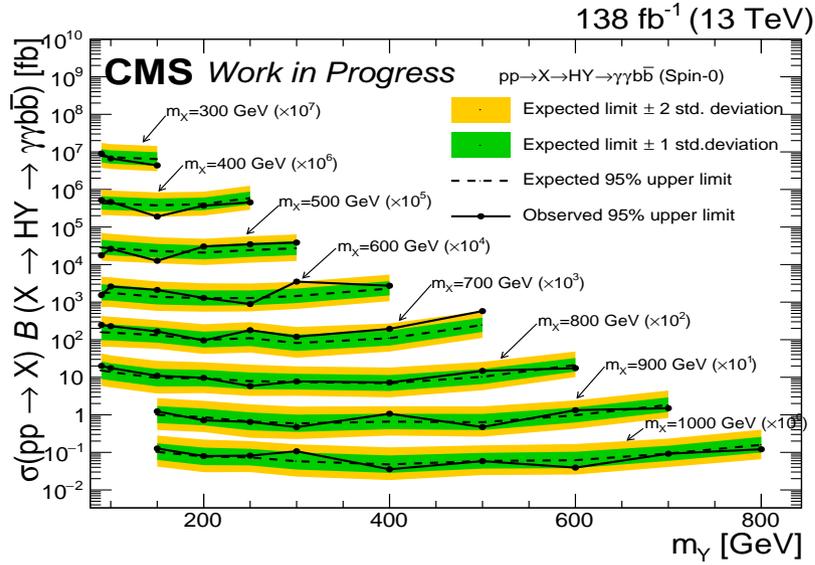
Figure 5.40: 95% CL observed (expected) upper limit on production cross section for NMSSM signal hypotheses. Limits are scaled with order of 10 depending upon the resonance X mass.
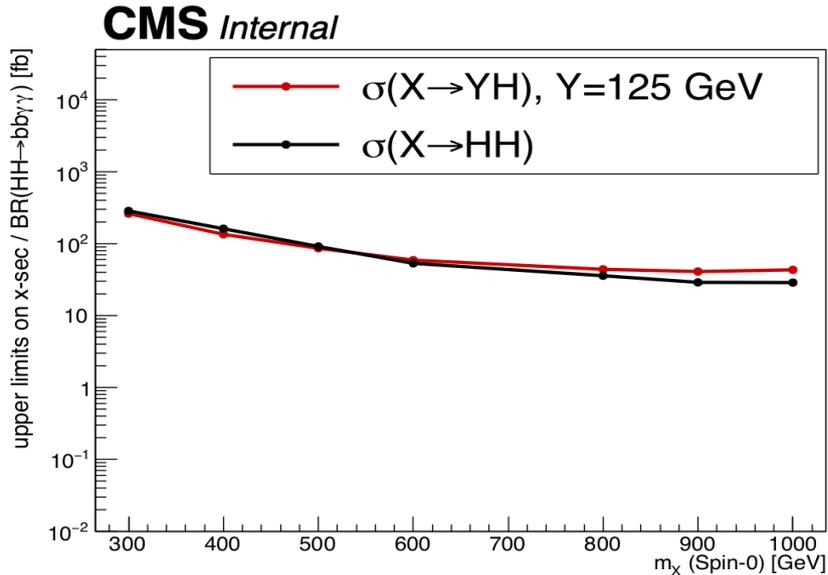
Figure 5.41: Comparison between expected median of upper limits on cross section for spin-0 signal hypotheses of WED and NMSSM models (Scaled with $BR(HH \rightarrow b\bar{b}\gamma\gamma)$, assuming $BR(Y125 \rightarrow b\bar{b}) = BR(H \rightarrow b\bar{b})$ as for Y=125 GeV, $Y \rightarrow b\bar{b}$, $H \rightarrow \gamma\gamma$ is indistinguishable from $H \rightarrow b\bar{b}$, $Y \rightarrow \gamma\gamma$; the difference between limits comes from the order of selection as for NMSSM we select b jet pair within a wider mass window.

Chapter 6

Summary and Conclusions

The thesis concentrates on searches for the Higgs boson pair production modes at large hadron collider (LHC). The four b quark final state is explored using simulations for the HL-LHC projection studies. Furthermore, the final state having two photons with two b quarks is studied using real LHC data collected by the compact muon solenoid (CMS) detector. Both the studies are summarized and concluded within the following sections:

6.1   HL-LHC projection studies

Prospects of the new physics searches using Higgs boson pair production at the high luminosity LHC (HL-LHC) with 14 TeV center-of-mass energy are presented. We search for a massive spin-2 bulk KK graviton, produced via vector boson fusion, decaying to a pair of Higgs bosons. The bulk gravitons are predicted in various new physics scenarios like the warped extra-dimensional models, which aim to explain the so-called hierarchy problems of the SM. The search focuses on the final state where both the Higgs bosons decay to b quark-antiquark pairs that are boosted, thus forming Higgs jets. Apart from the large amount of data at the HL-LHC, this search would benefit from the upgraded phase-2 CMS detector, which would help to improve the signal identification techniques in the high event rate expected at the HL-LHC. The analysis shows that the upgraded CMS experiment would be sensitive to observing a bulk graviton of mass 3.0 TeV, assuming

a signal production cross section of 1 fb, using 3 ab^{-1} proton-proton collision data. For non-resonant production mode, SM and effective field theory (EFT) motivated thirteen shape benchmarks are studied using a similar boosted analysis strategy. These benchmarks are the probe to study the shape of Higgs potential and Higgs-self coupling interaction. A 95% confidence level upper limit on the product of Higgs boson pair production cross section and branching fraction is presented for the benchmarks. The results show significant sensitivity for EFT motivated non-resonant di-Higgs production with strong contact interaction at the HL-LHC.

6.2 Resonant Higgs pair production analysis with Run-2 CMS data

A search for resonant Higgs boson pair production in the final state with two b quarks and two photons is presented. The search uses proton-proton collision data, collected by CMS experiment at LHC, corresponds to 138 fb^{-1} integrated luminosity with 13 TeV center-of-mass energy. The study is motivated by the warped extra dimension (WED) model and the next-to-minimal supersymmetric model (NMSSM). In signal topology, resonance X decays into either a pair of the Higgs boson H or a Higgs boson and another spin-0 particle Y with mass condition $m_Y < m_X - m_H$. Masses up to 1 TeV for X and up to 800 GeV for Y are covered. Optimized event selections, multivariate and neural network methods are used to enhance the analysis sensitivity. The signal is extracted by fitting a two-dimensional mass plane of the dijet and diphoton invariant mass. The data were found to be compatible with the background-only hypothesis. Results are presented as the upper limits at 95% confidence level on the product of the production cross section of X and its branching fraction to the $b\bar{b}\gamma\gamma$ final state, through either HH or HY. The observed limits for a spin-0 resonance decaying to HH ranges from 0.82–0.07 fb, while the expected limits are 0.74–0.08 fb. For a resonance decaying to HY, the observed limits are 0.90–0.04 fb, while the expected limits lie in the range 0.79–0.05 fb. Bulk radions decaying to HH are excluded for masses up to 600 GeV for Λ_R =6 TeV, while

the mass limit on a bulk KK graviton extends to 850 GeV assuming a coupling factor 0.5. The results are sensitive for low resonance masses among all the Higgs pair production searches performed at LHC.

Appendix A

Appendix to Chapter 4
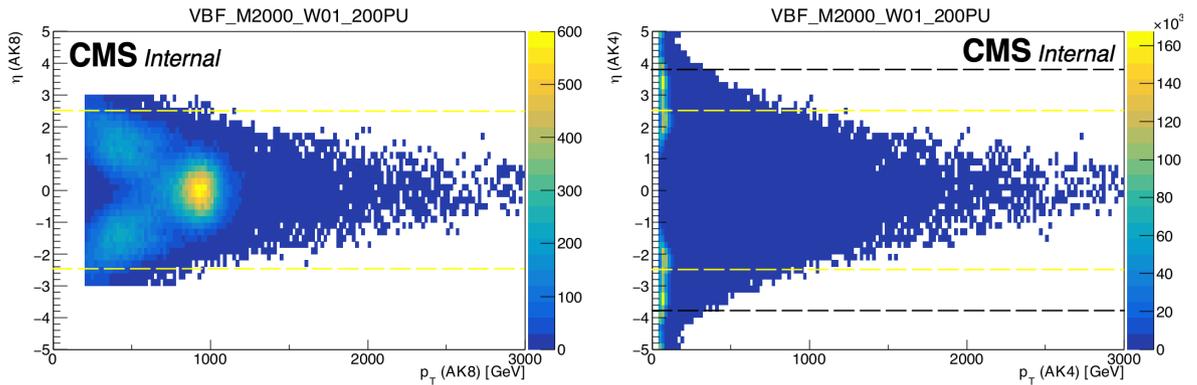
Figure A.1: 2D-map between Higgs AK8 jet p_T and η distribution (left plot); yellow dashed line corresponds to current tracker pseudorapidity coverage showing gain in signal acceptance with extended CMS Phase-2 tracker. (right plot) 2D-map between VBF AK4 jet p_T and η distribution; black dashed line corresponds to CMS Phase-2 tracker pseudorapidity coverage which shows that beyond this tracker coverage HGCAL would be helpful for full signal gain.

A.1   Comparing FullSim and Delphes

Early on in this analysis, we made private Delphes samples using the official CMS geometry cards and compared with the FullSIM samples. The details of the Delphes simulations are:

- Version `delphes3.4.2pre05`;

- CMS card: `CMS_PhaseII_Substructure_PIX4022`.

In this appendix, we show the comparison between the distributions in the FullSim and the Delphes signal samples. In Figs. A.2, A.3, A.4. We compare the FullSim distributions from Section 4.3.2 with their Delphes equivalents.
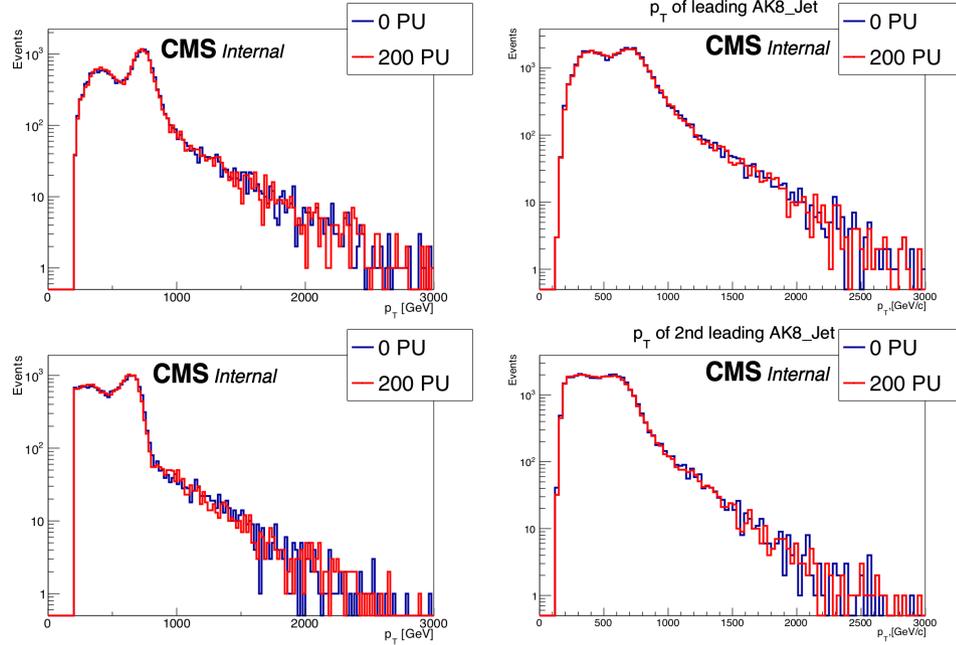
Figure A.2: Leading (upper) and subleading (lower) AK8 jet p_T. Fullsim (left) and Delphes (right) for $M_{BG} = 1.5$ TeV
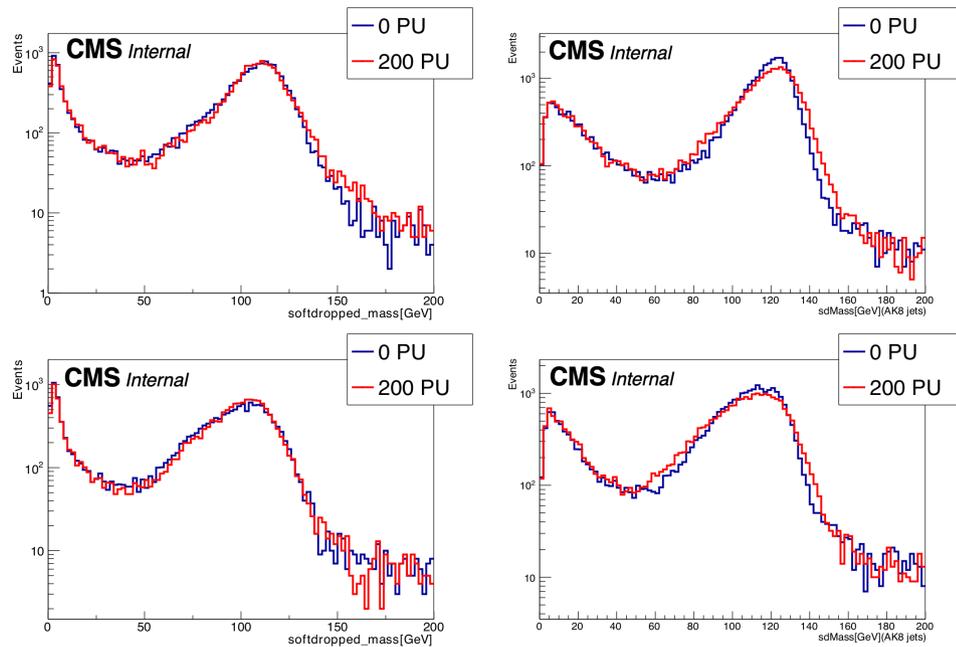
Figure A.3: Soft-drop mass of leading (upper) and subleading (lower) AK8 jet p_T. Fullsim (left) and Delphes (right) for $M_{BG} = 1.5$ TeV
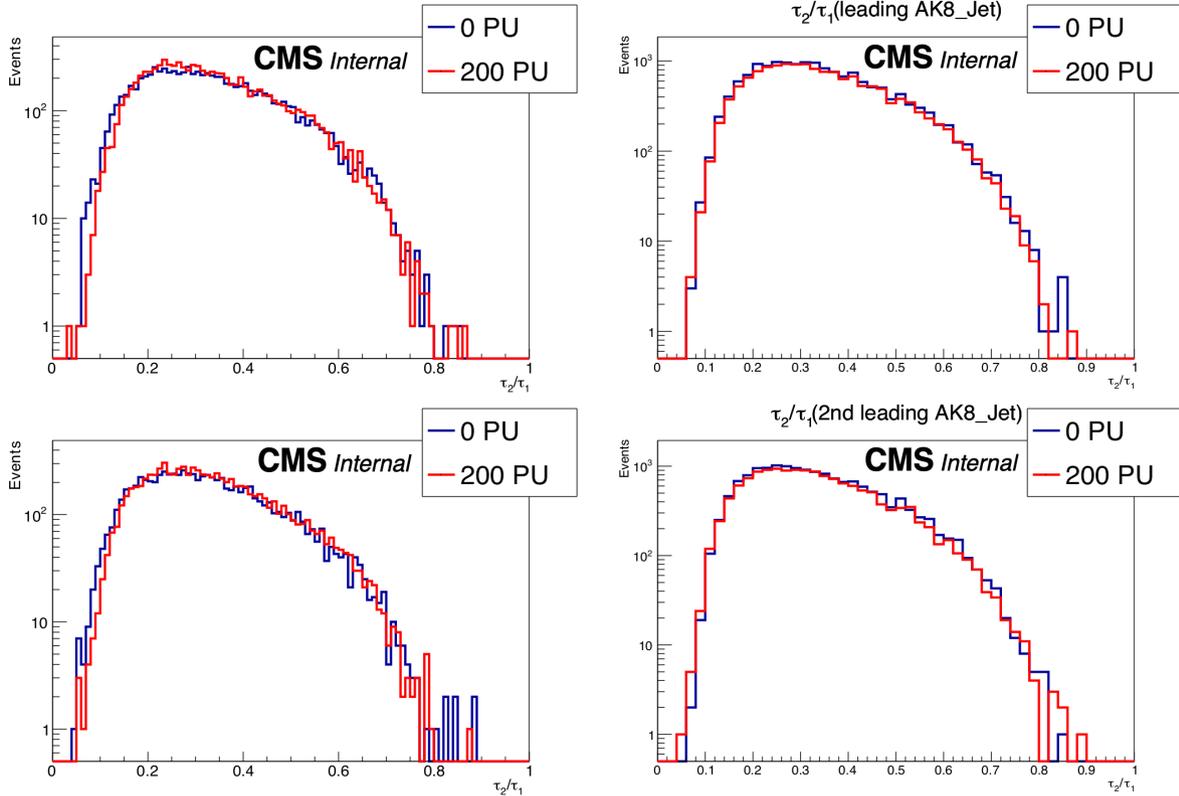
Figure A.4: N-subjettiness variable τ_2/τ_1 of leading (upper) and subleading (lower) AK8 jet p_T. Fullsim (left) and Delphes (right) for $M_{BG} = 1.5$ TeV

A.1.1 b tagging and jet substructure

This section includes the performance of the b tagging and soft-drop algorithms, comparing Run-2 and Phase-2. We have compared b tagging performance for two b taggers CSVv2 and DeepCSV (our choice for this analysis), for a couple of operating points (Loose and Medium) for Run-2 and Phase-2. We also compare the AK8 jet soft-drop mass distributions in Run-2 and Phase-2 to determine an optimal selection on the AK8 jet soft-drop mass for tagging the Higgs boson jets.

The comparison of the subjet b tagging efficiencies and mistag rate from light quarks and gluons between Run-2 and Phase-2 are as a function of the subjet p_T are shown in:

- Fig. A.5 for multijets for CSVv2 loose;

- Fig. A.6 for multijets for CSVv2 medium;

- Fig. A.8 for multijets for DeepCSV loose;

- Fig. A.9 for multijets for DeepCSV medium.

- Fig. A.7 for bulk gravtion signal for CSVv2 loose and medium;

- Fig. A.10 for bulk gravtion signal for DeepCSV loose and medium.

The comparisons in different subjet η regions are given in

- Fig. A.12 for multijets for DeepCSV medium;

- Fig. A.11 for bulk gravtion signal for DeepCSV medium.
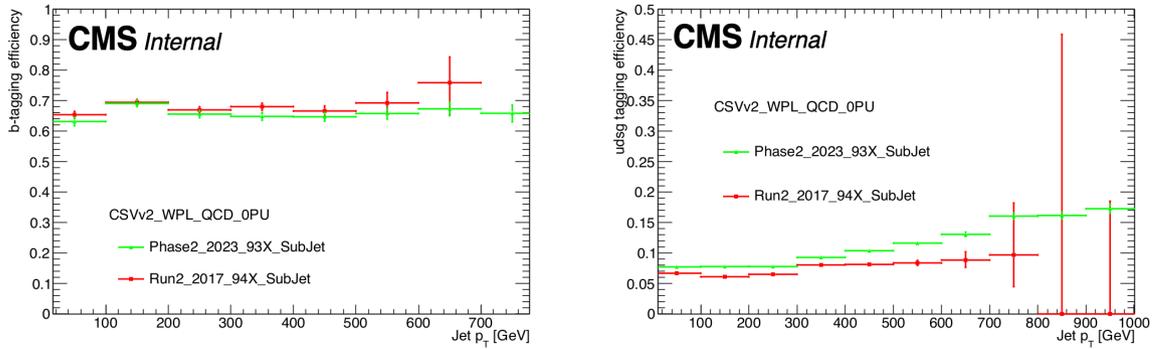
Figure A.5: b tagging efficiency (left) and mistag rate (right) using QCD sample for CSVv2 Loose operating point for $\eta \leq 2.5$ (Run-2) or $\eta \leq 3.5$ (Phase-2).
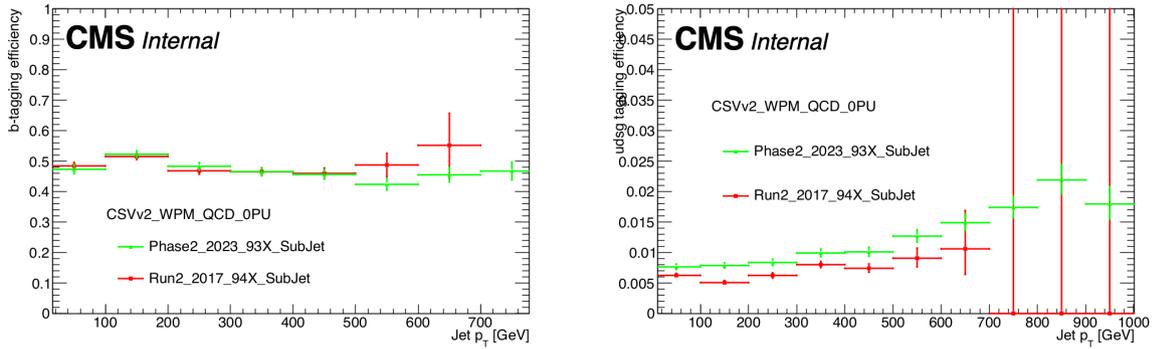
Figure A.6: *b* tagging efficiency (left) and mistag rate (right) using QCD sample for CSVv2 Medium operating point for $\eta \leq 2.5$ (Run-2) or $\eta \leq 3.5$ (Phase-2).
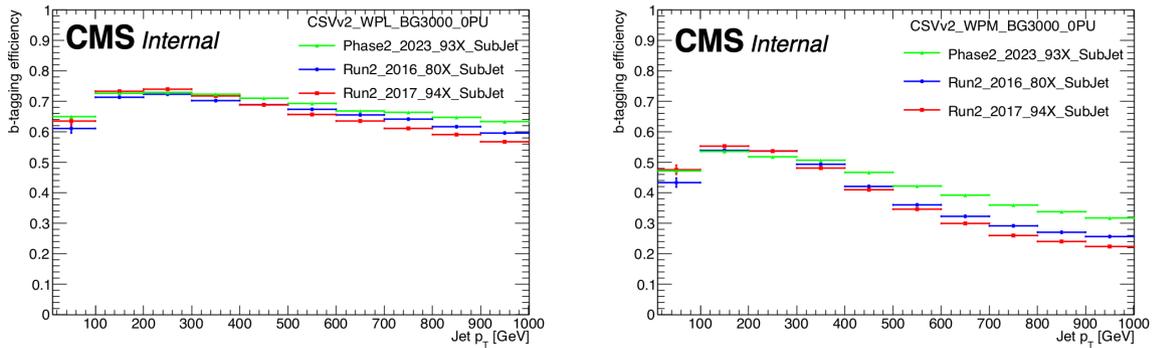
Figure A.7: *b* tagging efficiency using bulk graviton sample for CSVv2 Loose (left) and Medium (right) operating points for $\eta \leq 2.5$ (Run-2) or $\eta \leq 3.5$ (Phase-2).
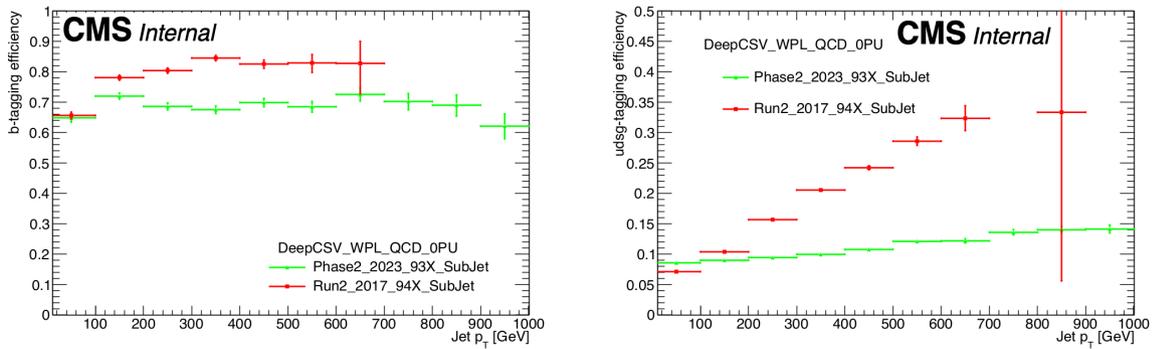
Figure A.8: *b* tagging efficiency (left) and mistag rate (right) using QCD sample for DeepCSV Loose operating point for $\eta \leq 2.5$ (Run-2) or $\eta \leq 3.5$ (Phase-2).
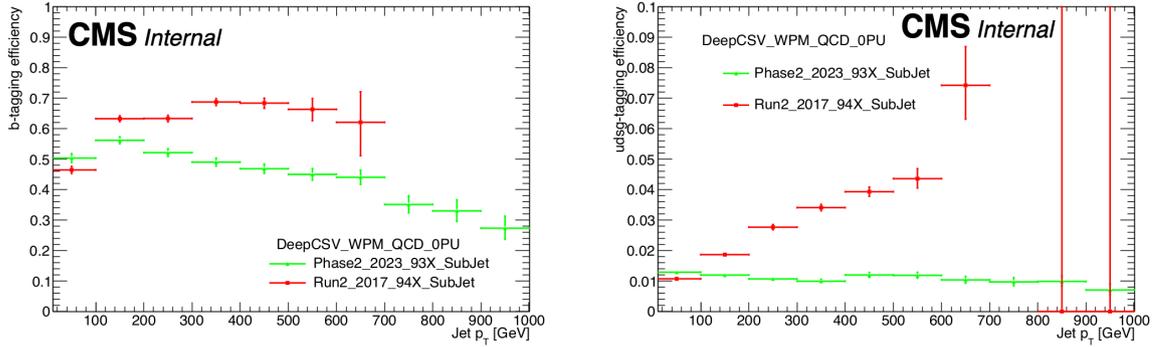
Figure A.9: *b* tagging efficiency (left) and mistag rate (right) using QCD sample for DeepCSV Medium operating point for $\eta \leq 2.5$ (Run-2) or $\eta \leq 3.5$ (Phase-2).
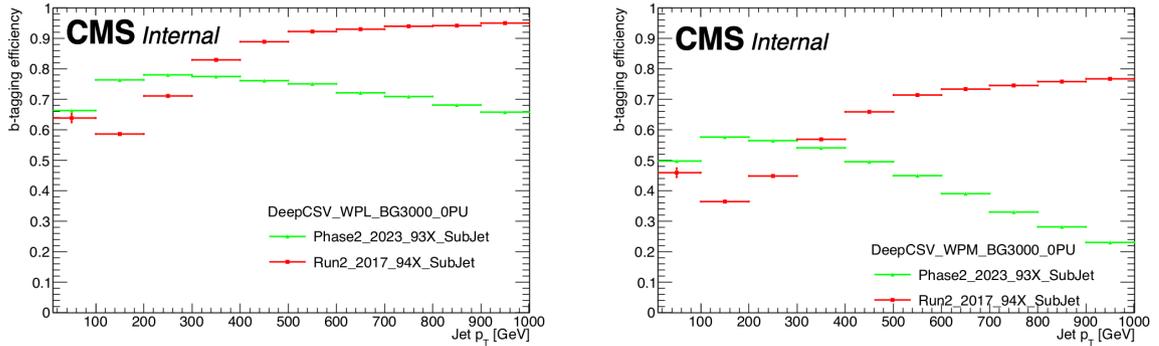
Figure A.10: *b* tagging efficiency using bulk graviton sample for DeepCSV Loose (left) and Medium (right) operating points for $\eta \leq 2.5$ (Run-2) or $\eta \leq 3.5$ (Phase-2).
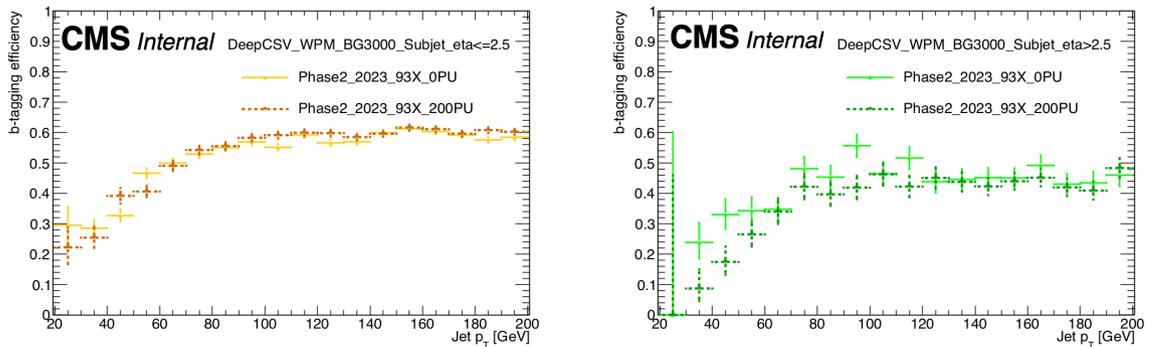
Figure A.11: *b* tagging efficiency using bulk graviton sample for DeepCSV Medium operating point for for $|\eta| \leq 2.5$ (left) and $2.5 < |\eta| \leq 3.5$ (right).
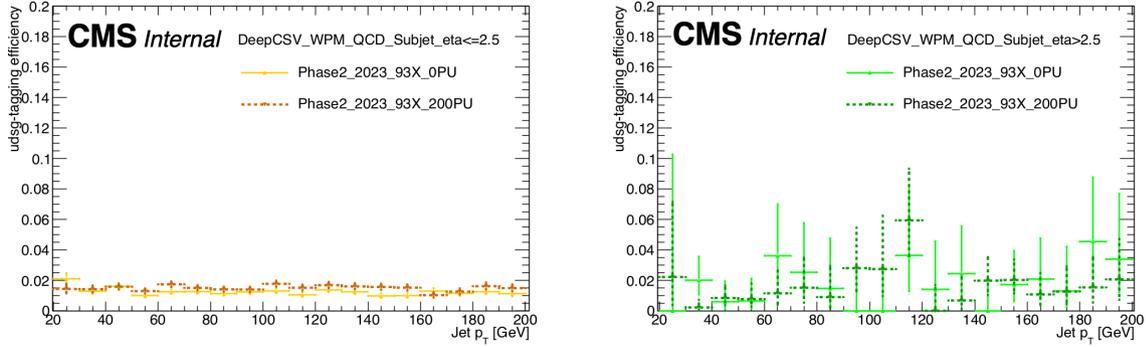
Figure A.12: Mistag rate using QCD sample for DeepCSV Medium operating point for $|\eta| \leq 2.5$ (left) and $2.5 < |\eta| \leq 3.5$ (right).

The following receiver operating characteristics curves are shown:

- CSVv2 for subjets of all AK8 jets comparing Run-2 and Phase-2 (PU=0): Fig A.13. The Run-2 and Phase-2 performances are equivalent for high efficiency (right region) whereas the Phase-2 is better at high purity (left region).

- DeepCSV for subjets comparing Run-2 and Phase-2 (PU=0), with subjet $p_T > 20$ GeV and $|\eta| < 2.5$ (Run-2) or $|\eta| < 3$ (Phase-2): Fig A.14. The figure shows Run-2 to have a slightly better performance. However, it is expected that with a dedicated Phase-2 training, the DeepCSV should perform better than the Run-2.

- DeepCSV for subjets in Phase-2 with pileups of 0 and 200, with subjet $p_T < 200$ GeV and in two η ranges, $|\eta| \leq 2.5$ and $2.5 < |\eta| < 3.5$: Fig A.15. These curves show that the b tagging is resilient against pileup.
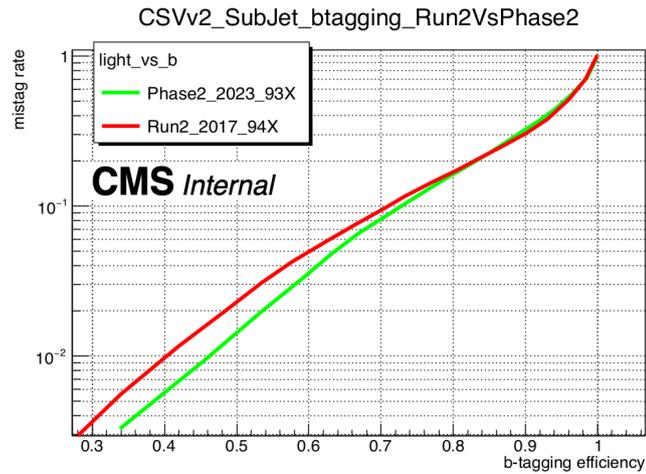
Figure A.13: Comparison of ROC curve for CSVv2 between Run-2 and Phase2 for subjets of all AK8 jets with subjet $p_T > 20$ GeV and $\eta \leq 2.5$ (Run-2) or $\eta \leq 3$ (Phase-2).
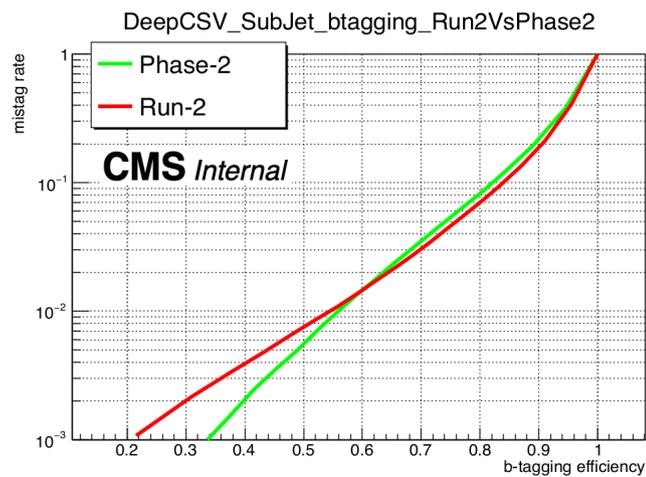
Figure A.14: Comparison of ROC curve for DeepCSV between Run-2 and Phase-2 for subjets of all AK8 jets with subjet $p_T > 20$ GeV and $\eta \leq 2.5$ (Run-2) or $\eta \leq 3$ (Phase-2).
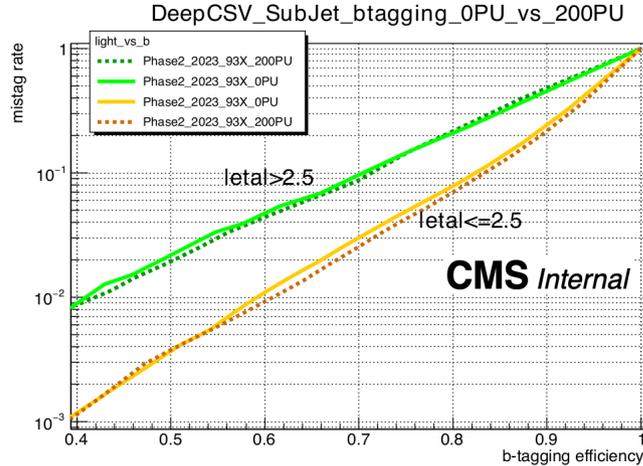
Figure A.15: Comparison of ROC curve for DeepCSV for PU of 0 and 200. Two different η ranges are shown: $|\eta| \leq 2.5$ and $2.5 < |\eta| \leq 3.5$.

A comparison of the soft-drop mass distributions of the leading and subleading-p_T AK8 jets in the signal events between Run-2 and Phase-2 simulation samples is shown in Fig. A.16. The soft-drop mass windows used in Run-2 and Phase-2 to select Higgs-tagged jets are also demarcated. The Phase-2 soft-drop mass resolution has a slightly poorer resolution compared to the Run-2, but this is compensated by widening the Higgs tagged jet mass window, as discussed in Section 4.3.2.

The stability of the soft-drop mass distribution as a function of pileup is studied. As shown in Fig. A.17 the soft-drop mass distribution with respect to the number of reconstructed PVs (a good measure of the number of pileup events) is relatively flat (given by the profile histogram shown as black markers). Hence the chosen mass window is deemed to be optimal for all pileup conditions.
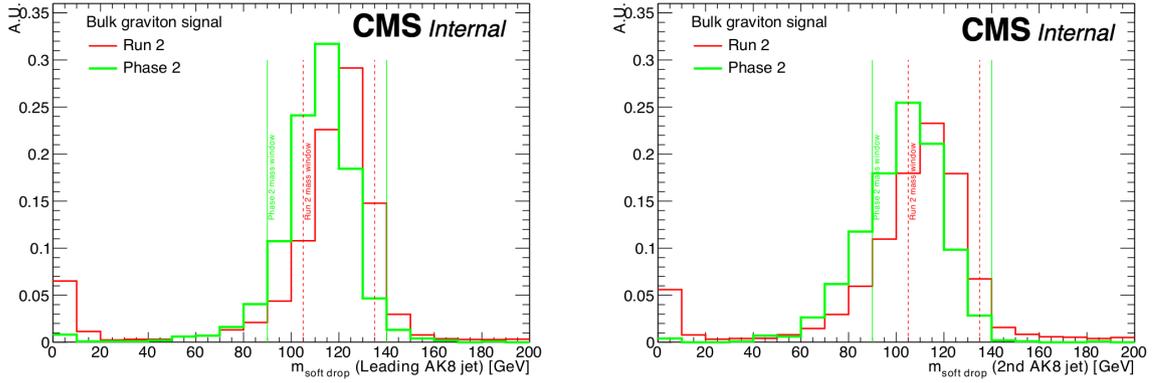
Figure A.16: Comparison of the leading-p_T (left) and second leading-p_T (right) AK8 jet soft-drop mass in bulk graviton signal simulations (mass = 3 TeV) in Run-2 (red histogram) and Phase-2 (green histogram). The Run-2 and the Phase-2 mass windows used to selection the Higgs tagged jets are also shown as red lines and green lines, respectively. The Phase-2 soft-drop mass resolution has a slightly poorer resolution compared to the Run-2, but this is compensated by widening the Higgs tagged jet mass window.
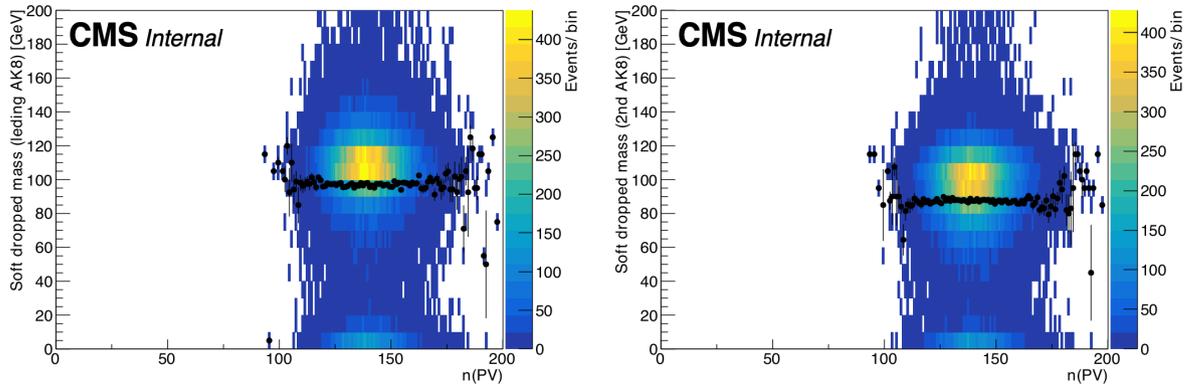
Figure A.17: Comparison of the leading-p_T (left) and second leading-p_T (right) AK8 jet soft-drop mass distributions as a function of the number of primary vertices in bulk graviton signal simulations (mass = 3 TeV) in Phase-2 with PU=200.

Table A.1: Event yields and efficiencies for the signal and QCD multijet background **using the DeepCSVM operating point for subjet b tagging, which is our final choice, based on expected signal significance (Fig. A.22)**. The product of the cross sections and branching fractions of the signals is assumed to be 1 fb.

	PU = 0			
Process	Category $3b$		Category $4b$	
	Events	Efficiency (%)	Events	Efficiency (%)
Multijets	482	5.37×10^{-4}	45.6	1.30×10^{-4}
BG (M=1.5 TeV, width = 1%)	120	5.7	33.2	4.0
BG (M=2.0 TeV, width = 1%)	122	6.5	28.5	4.1
BG (M=3.0 TeV, width = 1%)	79.5	5.7	14.2	2.6
BG (M=3.0 TeV, width = 5%)	77.8	5.5	14.3	2.6
	PU = 200			
Process	Category $3b$		Category $4b$	
	Events	Efficiency (%)	Events	Efficiency (%)
Multijets	17462	2.16×10^{-2}	1595	3.92×10^{-3}
BG (M=1.5 TeV, width = 1%)	447	21	126	14.9
BG (M=2.0 TeV, width = 1%)	458	24	114	15
BG (M=3.0 TeV, width = 1%)	354	24	63.9	12
BG (M=3.0 TeV, width = 5%)	359	24	70.3	12

A.2 The m_{JJ} and significance with preliminary selections

This appendix documents the results using the initial selection on the AK8 jet soft-drop mass: 80–160 GeV for the leading and 60–140 GeV for the subleading-p_T AK8 jets. In addition, both the DeepCSVL (discriminator > 0.152) and DeepCSVM (discriminator > 0.4941) were tested to find the optimal choice. The full analysis chain was run and signal significance computed with systematic uncertainties included. Based on this, the choice of the DeepCSV tagger was found to be DeepCSVM with 3 and 4 b tagged subjet categories.

The m_{JJ} distributions are shown in Figs. A.18 and A.19, for the DeepCSVL and the DeepCSVM subjet b tagging operating points, respectivtely.

The efficiencies for the signal selection for bulk gravitons (BG) of masses 1.5, 2.0 and
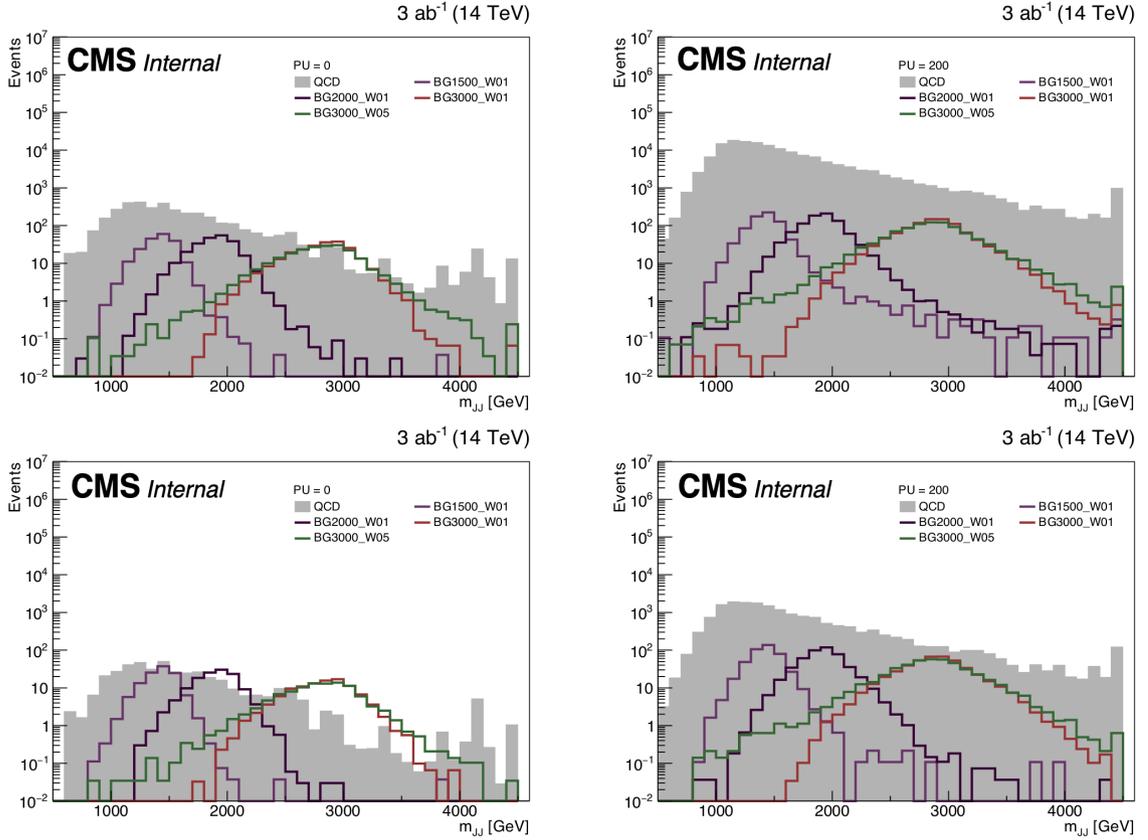
Figure A.18: Upper (lower) plot for the estimated QCD background and the signal m_{JJ} distributions with PU=0 (left) and PU=200 (right) and for the 3*b* (4*b*) subjet *b* tagging categories for bulk gravitons (BG) of masses 1.5, 2.0, and 3.0 TeV with 1% (and 5% for 3.0 TeV) widths, assuming a signal cross section of 1 fb. These figures are using the DeepCSVL operating point.

3.0 TeV, and for 0 and 200 pileups, are shown in Fig. A.20 (left) and efficiencies on QCD sample are shown in Fig. A.20 (right). Two choices for subjet *b* tagging–the DeepCSVL and DeepCSVM–are shown. From Fig. A.20 it can be seen that the efficiency of VBF selection depends strongly on pileup. Moreover, VBF efficiency is much higher for QCD at PU=200 than for signal. This can be seen from the number of VBF jet pairs, shown in Fig A.21. The multijet events do not have true VBF pairs at PU=0. With the addition of pileup, the number of VBF jet pairs in the multijet events drastically increases, and so does its efficiency. The efficiency increase is relatively modest for the signal since these
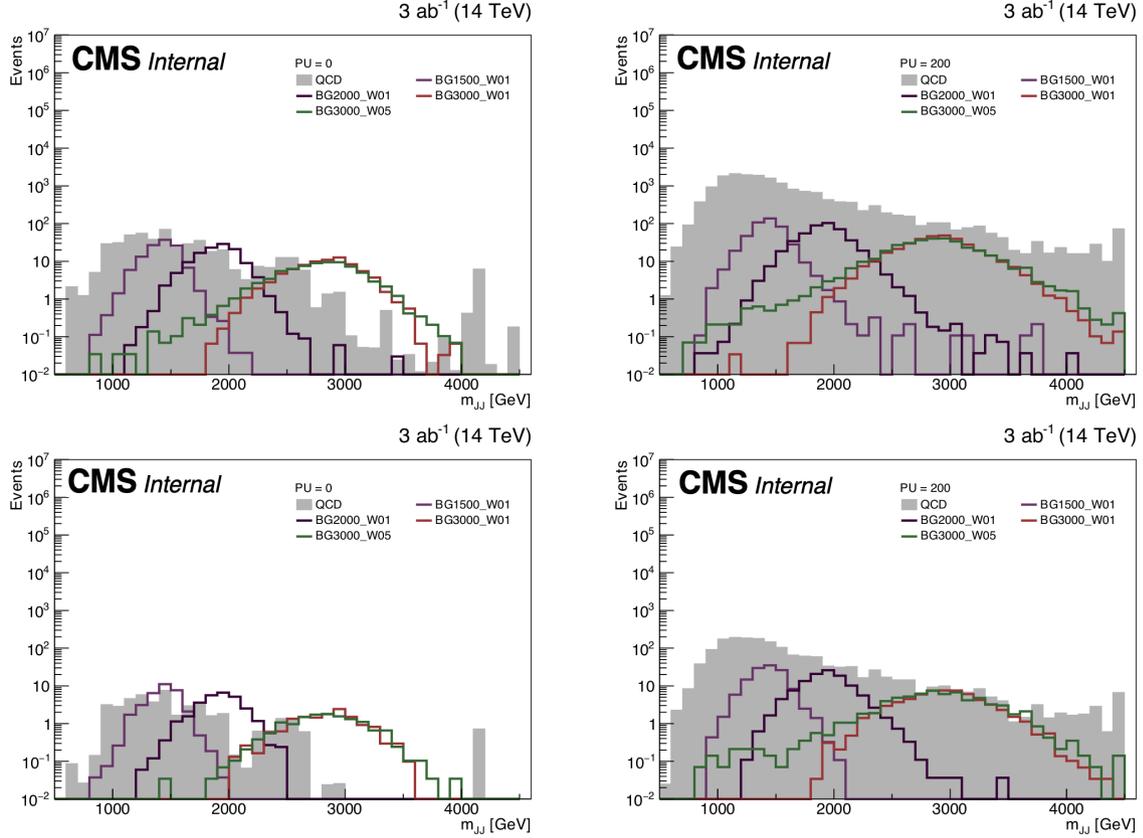
Figure A.19: Upper (lower) plot for the estimated QCD background and the signal m_{JJ} distributions with PU=0 (left) and PU=200 (right) and for the 3b (4b) subjet b tagging categories for bulk gravitons (BG) of masses 1.5, 2.0, and 3.0 TeV with 1% (and 5% for 3.0 TeV) widths, assuming a signal cross section of 1 fb. These figures are using the DeepCSVM operating point.

events already have true VBF pairs even at PU=0.

The event yields after full selection are shown in Table A.1 **using the DeepCSVM operating point for subjet b tagging**. We estimate the expected significance of the signal using the distributions shown in Figs. A.18 and A.19, for the DeepCSVL and the DeepCSVM subjet b tagging operating points, respectively. The expected significance is compared to make a choice between the two operating points in the final event selection for the PAS. The Higgs combine tool is used. It can be seen that the DeepCSVM performs better than the DeepCSVL over most of the mass range, particularly at PU=200. **Hence,**
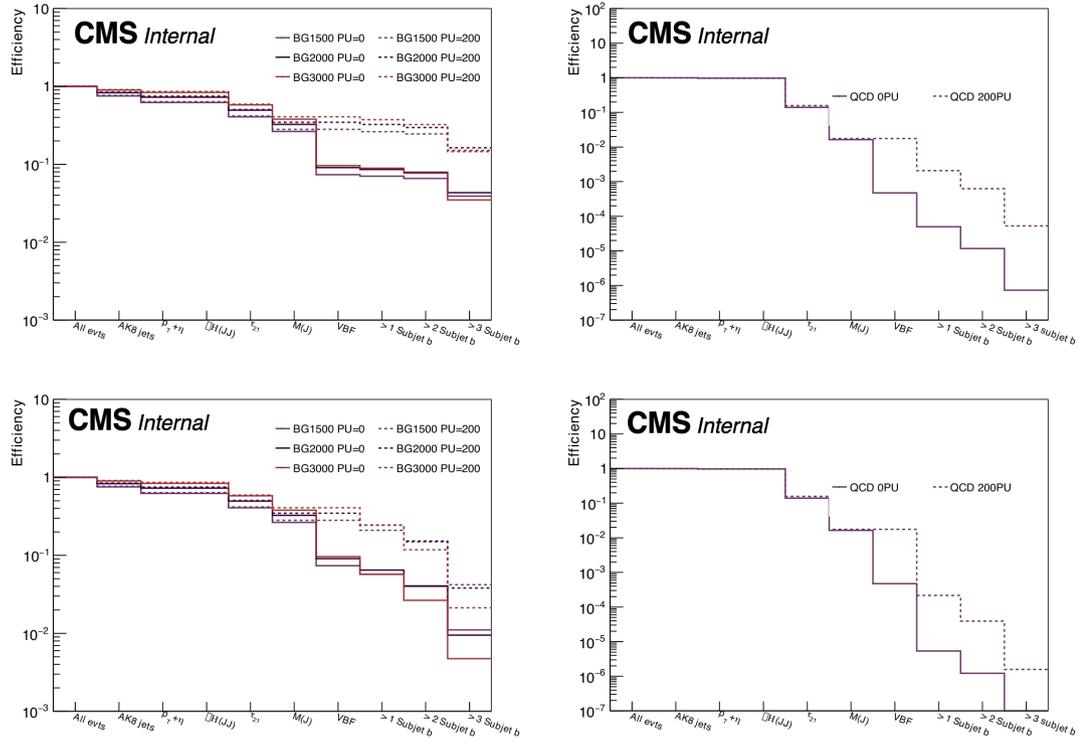
Figure A.20: The signal efficiencies for the various steps in the selection criteria for bulk gravitons (BG) of masses 1.5, 2.0 and 3.0 TeV with 1% (also 5% for 3.0 TeV) widths (left) and QCD background(right), for 0 and 200 pileup. The results are shown using the DeepCSVL (upper) and DeepCSVM (lower) operating points for the subjet b tagging. **The DeepCSVM is our final choice, based on expected signal significance (Fig. A.22).**

the DeepCSVM is our final choice, based on expected signal significance (Fig. A.22).
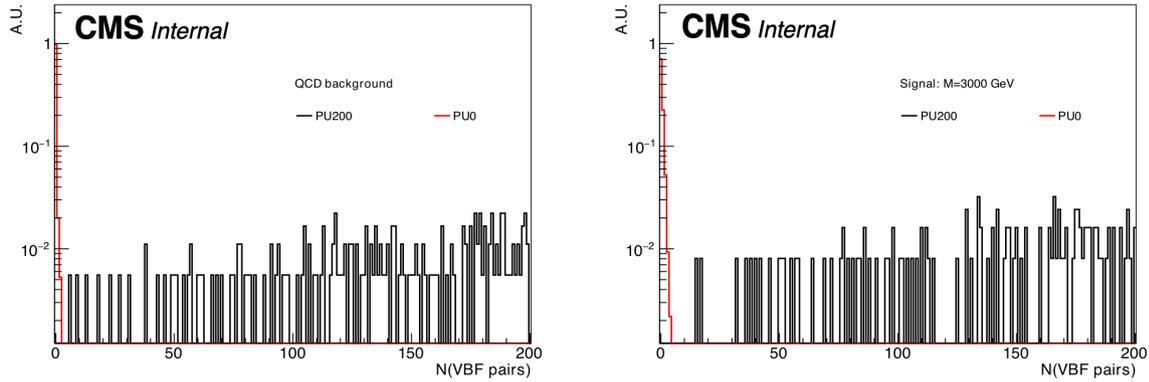
Figure A.21: The total number of VBF jet pairs in the multijet background (left) and signal bulk graviton of mass 3000 GeV (right). Shown are the distributions of PU=0 (red) and PU=200 (black).
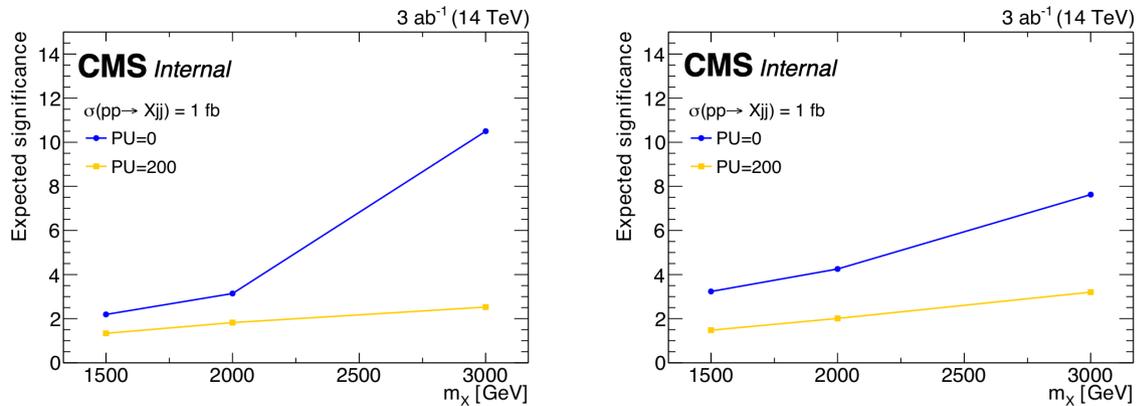
Figure A.22: The expected signal significance for three mass points of the bulk graviton assuming a production cross section of 1 fb. The left (right) figures are for the subjet b tagging using the DeepCSVL (DeepCSVM) operating points. It can be seen that the DeepCSVM performs better than the DeepCSVL over most of the mass range, particularly at PU=200. Hence this operating point is our final choice for the event selection used in this analysis.

A.3 Subjet b tagging efficiency for DeepCSV loose

Fig. A.24 and A.23 shows the subjet b tagging efficiency using the DeepCSVL and DeepCSVM operating points, respectively, for different jet flavours and as a function of the subjet p_T and η. The efficiencies are estimated in events with an average pileup of zero or 200.
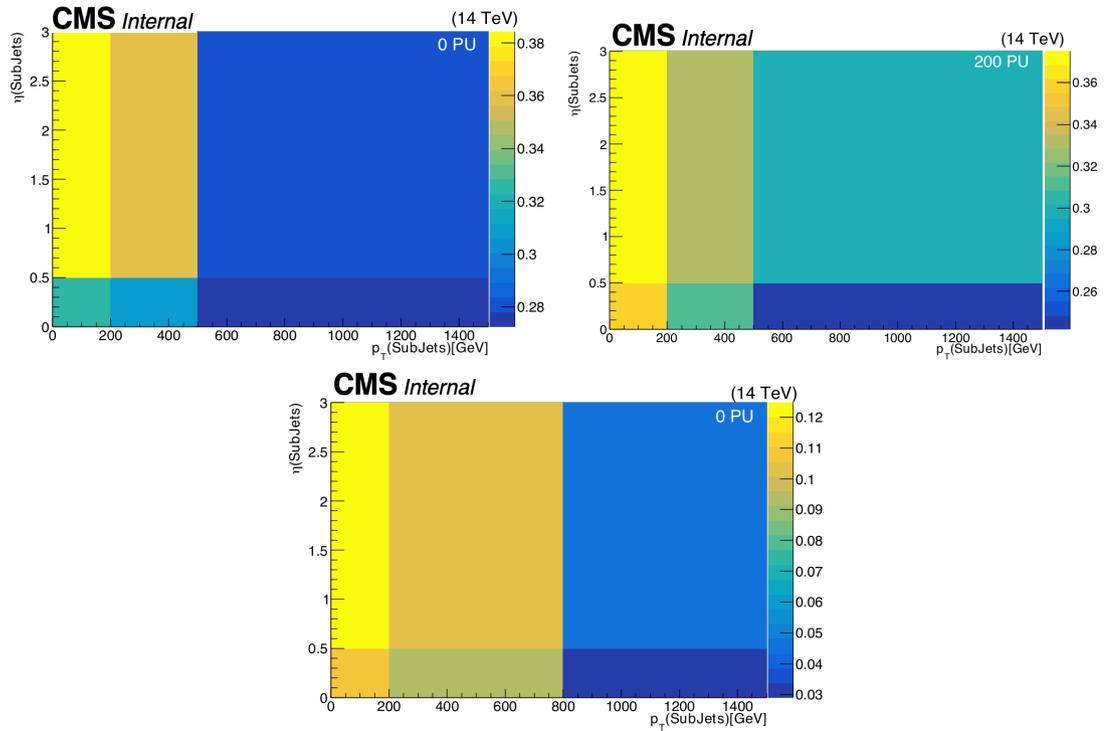
Figure A.23: The b tagging efficiency for subjets from a b quark (upper left), c quark (upper right), and light quark or a gluon (lower) in events with an average pileup of zero. These figures are using the DeepCSVM operating point.
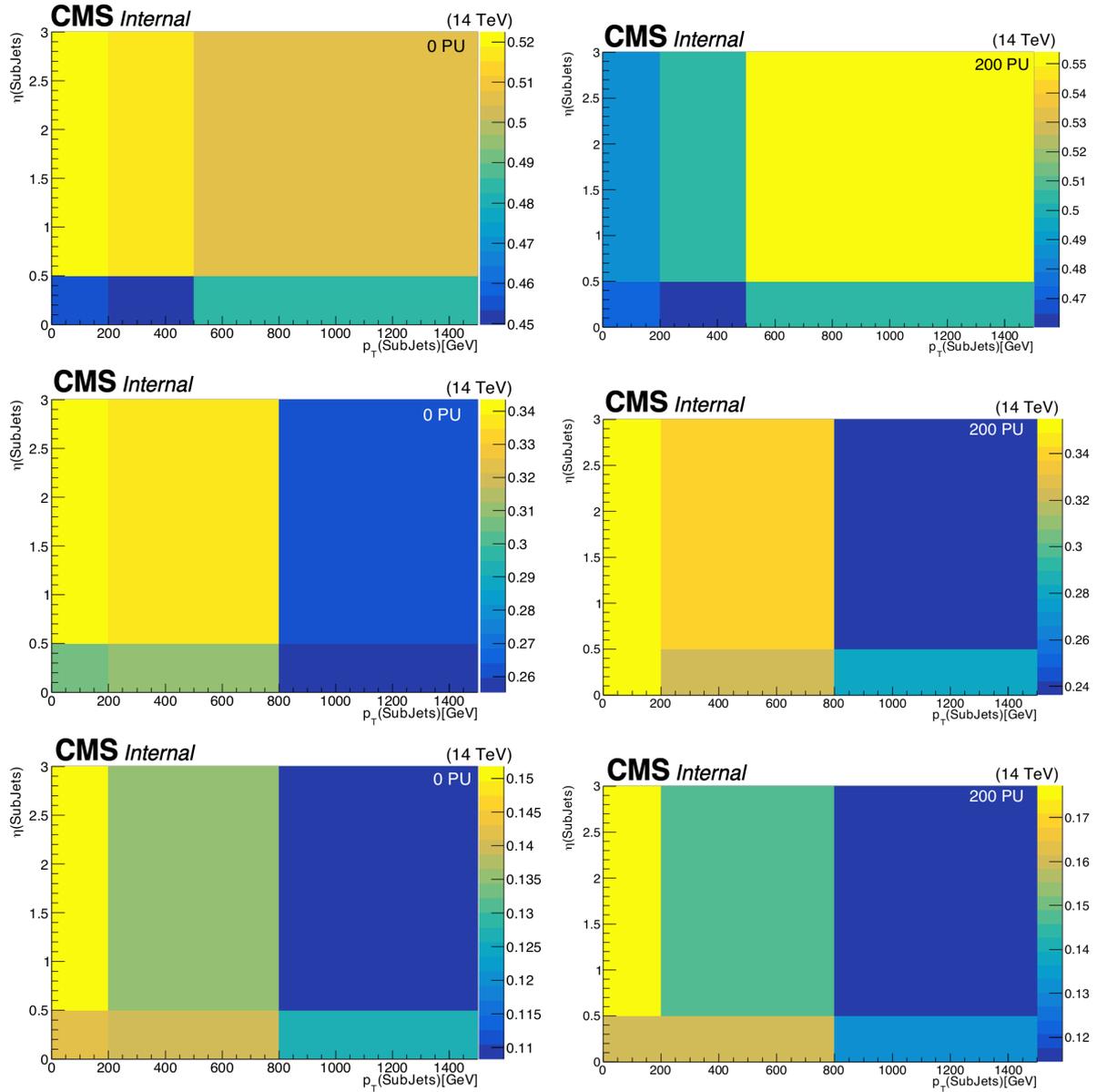
Figure A.24: The b tagging efficiency for subjets from a b quark (upper left), c quark (upper right), and light quark or a gluon (lower) in events with an average pileup of zero (left) and 200 (right). These figures are using the DeepCSVL operating point.

Appendix B

Appendix to Chapter 5

B.1 Other kinematic distributions

In the chapter 5, we add the kinematic distribution which have been used as BDT training input only for lowY mass ranges. We add midY and highY distributions within Fig. B.1, B.2, B.3, B.4, B.5, B.6.

(a) $|\cos\theta_{HY}^{\mathrm{CS}}|$

(b) $|\cos\theta_{\gamma\gamma}^{\mathrm{CS}}|$

(c) $|\cos\theta_{bb}^{\mathrm{CS}}|$

(d) $p_T(\mathrm{jj})/m_{jj\gamma\gamma}$

(e) $p_T(\gamma\gamma)/m_{jj\gamma\gamma}$

(f) $\min(\Delta\mathrm{R}(\gamma,\mathrm{jets}))$

(g) other $\min(\Delta\mathrm{R}(\gamma,\mathrm{jets}))$

(h) $\sigma_{m_{\gamma\gamma}}/m_{\gamma\gamma}$
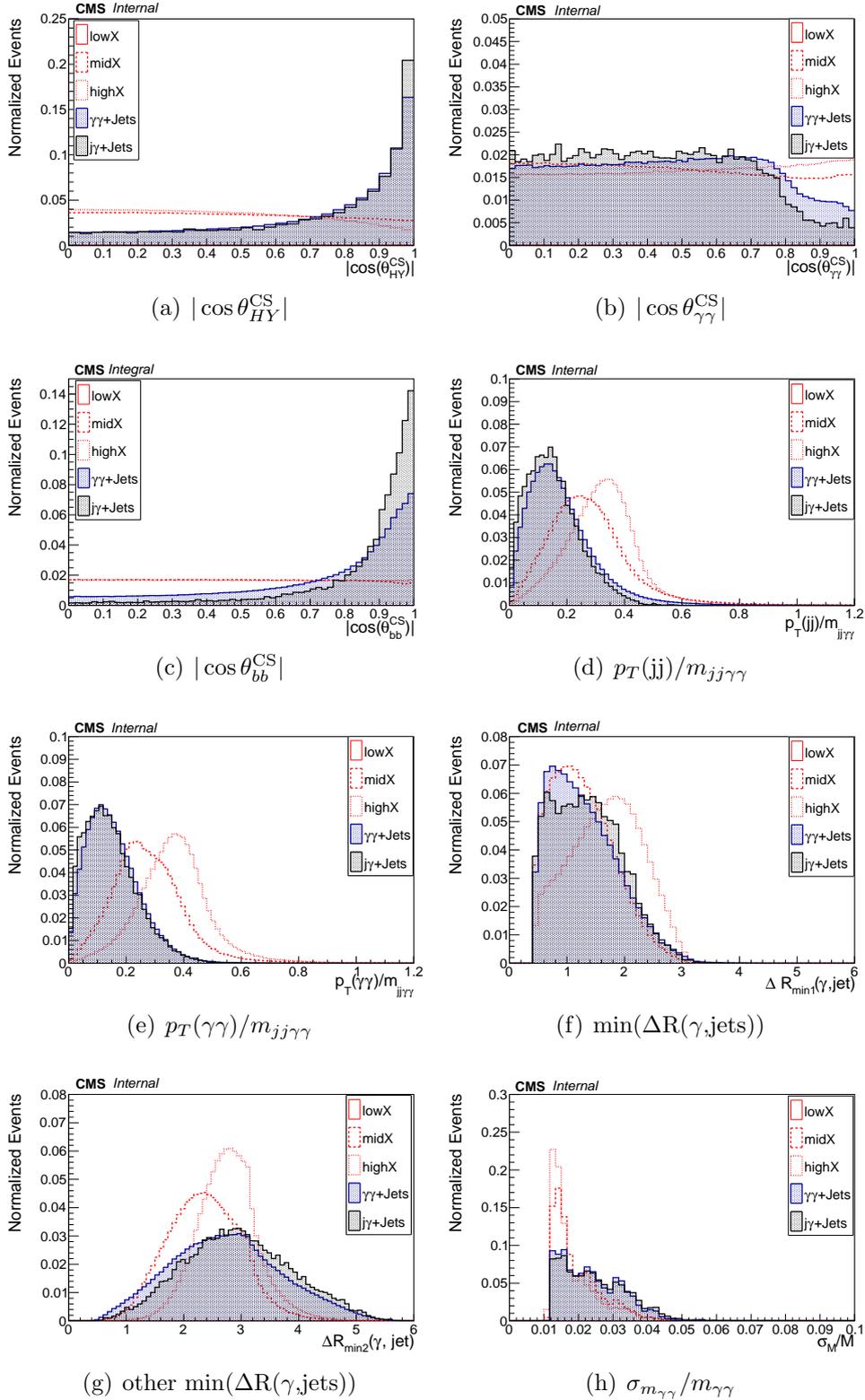
Figure B.1: Signal and background kinematics for midY mass region

(a) $\sigma_{m_{jj}}/m_{jj}$

(b) leading jet b Discriminant

(c) leading jet bRegressionNN

(d) $p_T(j1)/m_{jj}$

(e) leading photonID MVA

(f) $p_T(\gamma 1)/m_{\gamma\gamma}$

(g) leading photon σ_E/E

(h) subleading Jet b Discriminant
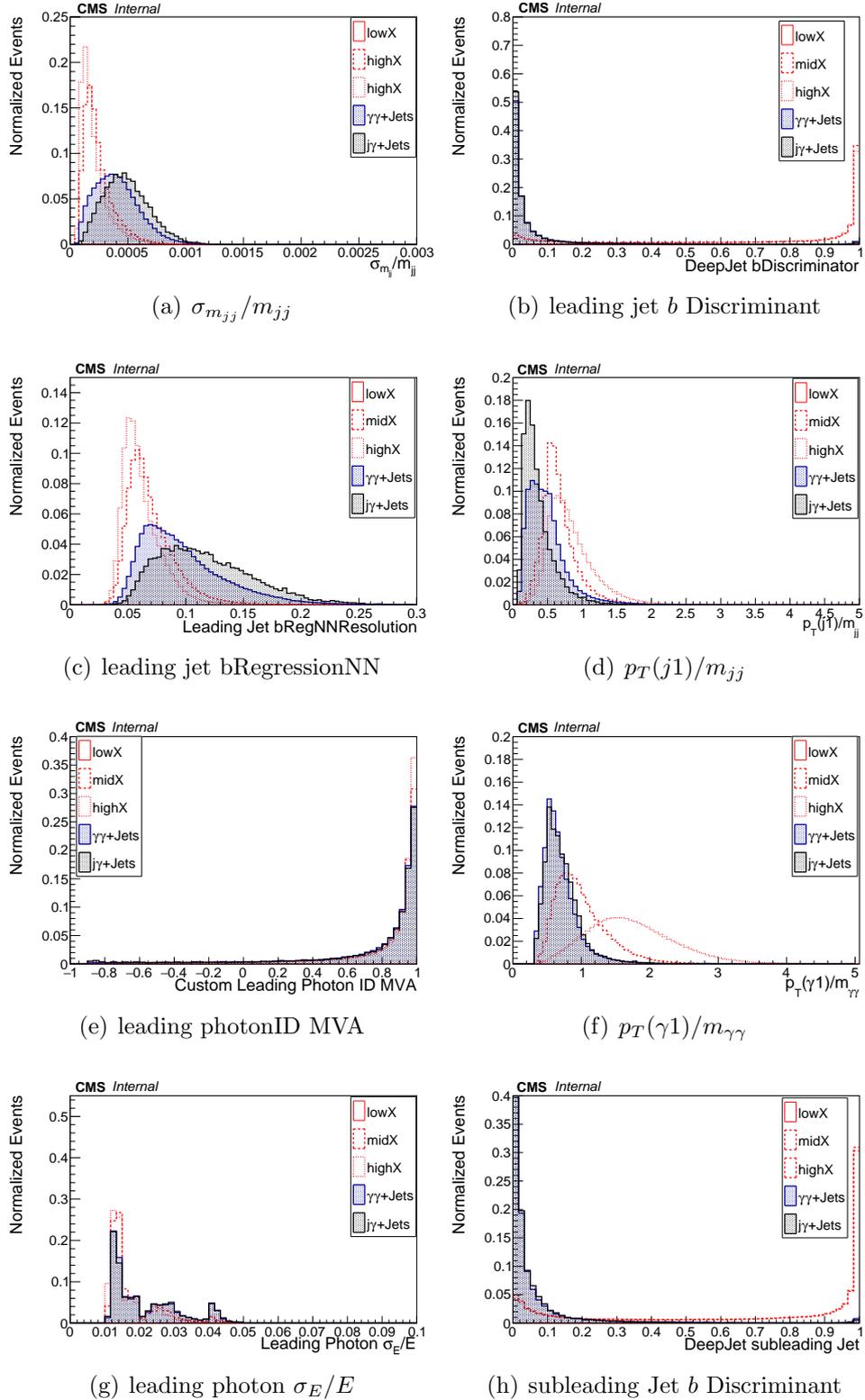
Figure B.2: Signal and background kinematics for midY mass region

(a) subleading Jet bRegressionNN

(b) $p_T(j2)/m_{jj}$

(c) subleading photonID MVA

(d) $p_T(\gamma2)/m_{\gamma\gamma}$

(e) subleading photon σ_E/E
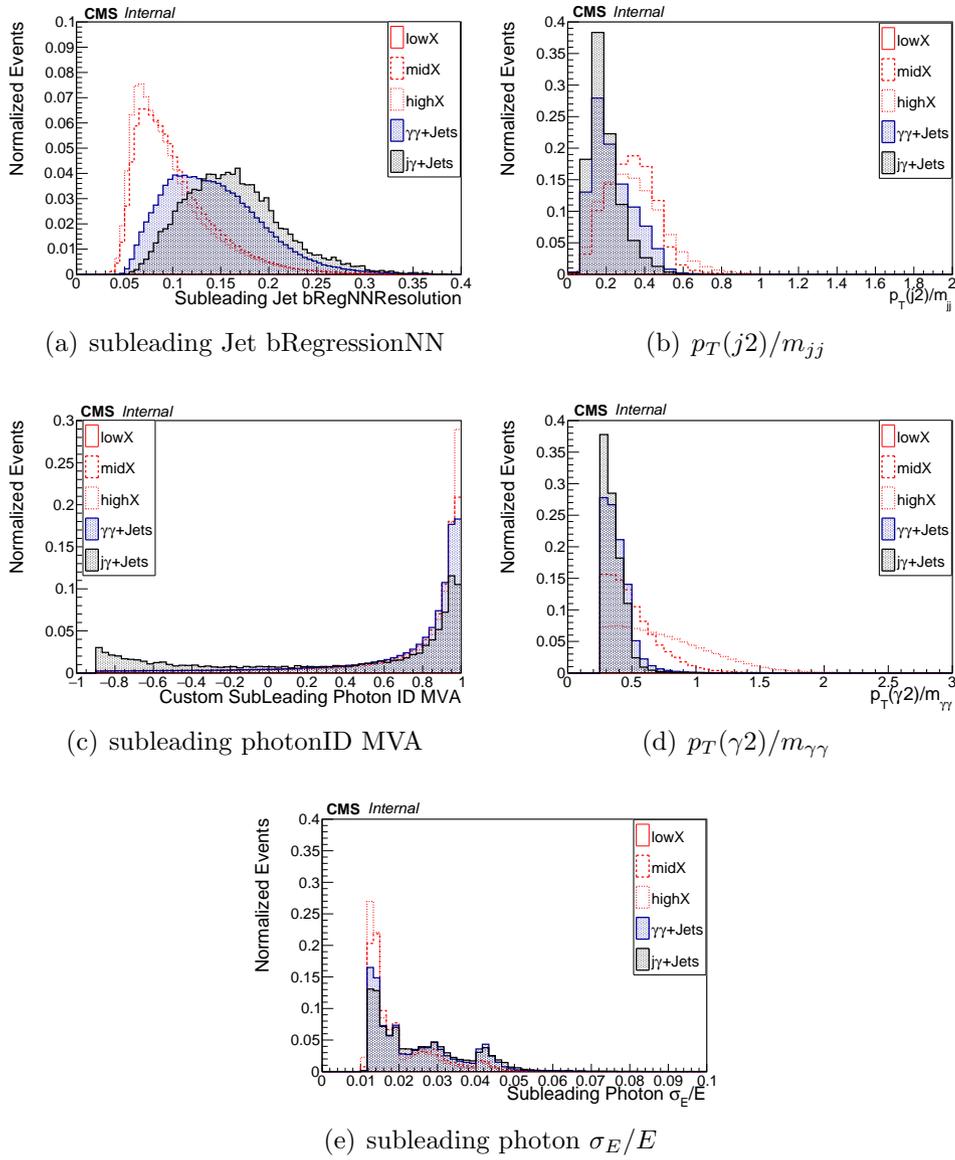
Figure B.3: Signal and background kinematics for midY mass region
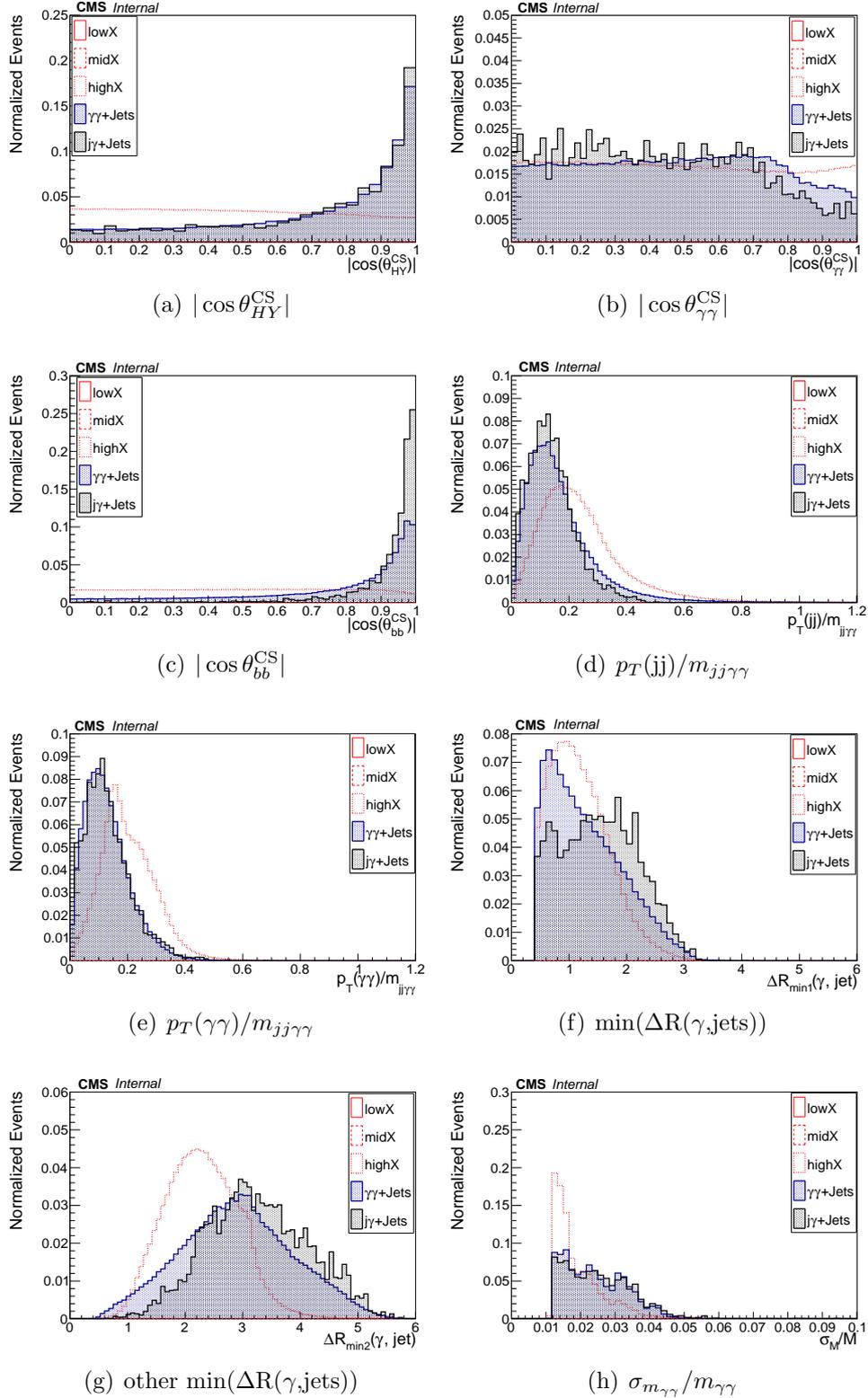
Figure B.4: Signal and background kinematics for highY mass region

(a) $\sigma_{m_{jj}}/m_{jj}$

(b) leading jet b Discriminant

(c) leading jet bRegressionNN

(d) $p_T(j1)/m_{jj}$

(e) leading photonID MVA

(f) $p_T(\gamma1)/m_{\gamma\gamma}$

(g) leading photon σ_E/E

(h) subleading Jet b Discriminant
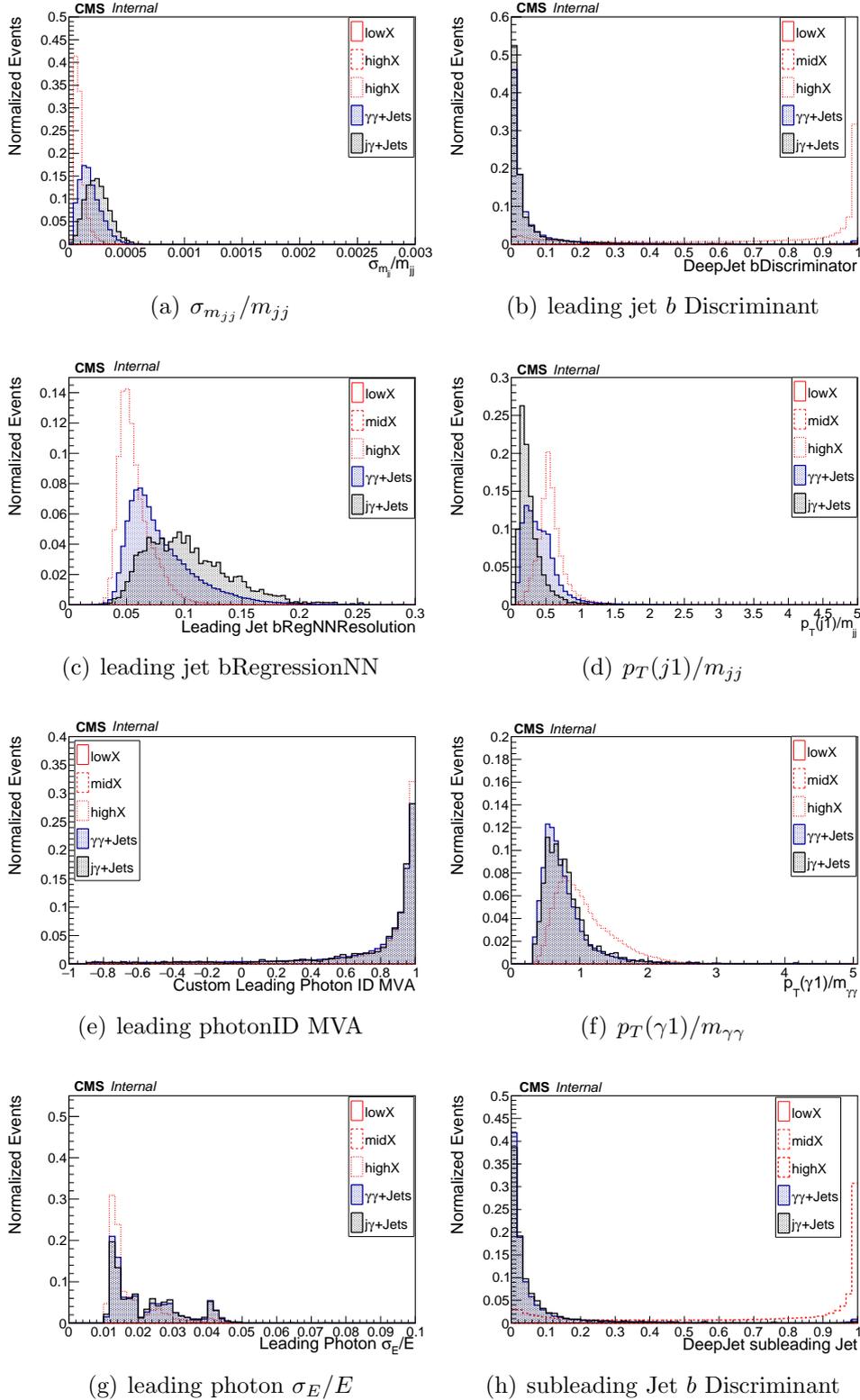
Figure B.5: Signal and background kinematics for highY mass region

(a) subleading Jet bRegressionNN

(b) $p_T(j2)/m_{jj}$

(c) subleading photonID MVA

(d) $p_T(\gamma 2)/m_{\gamma\gamma}$

(e) subleading photon σ_E/E
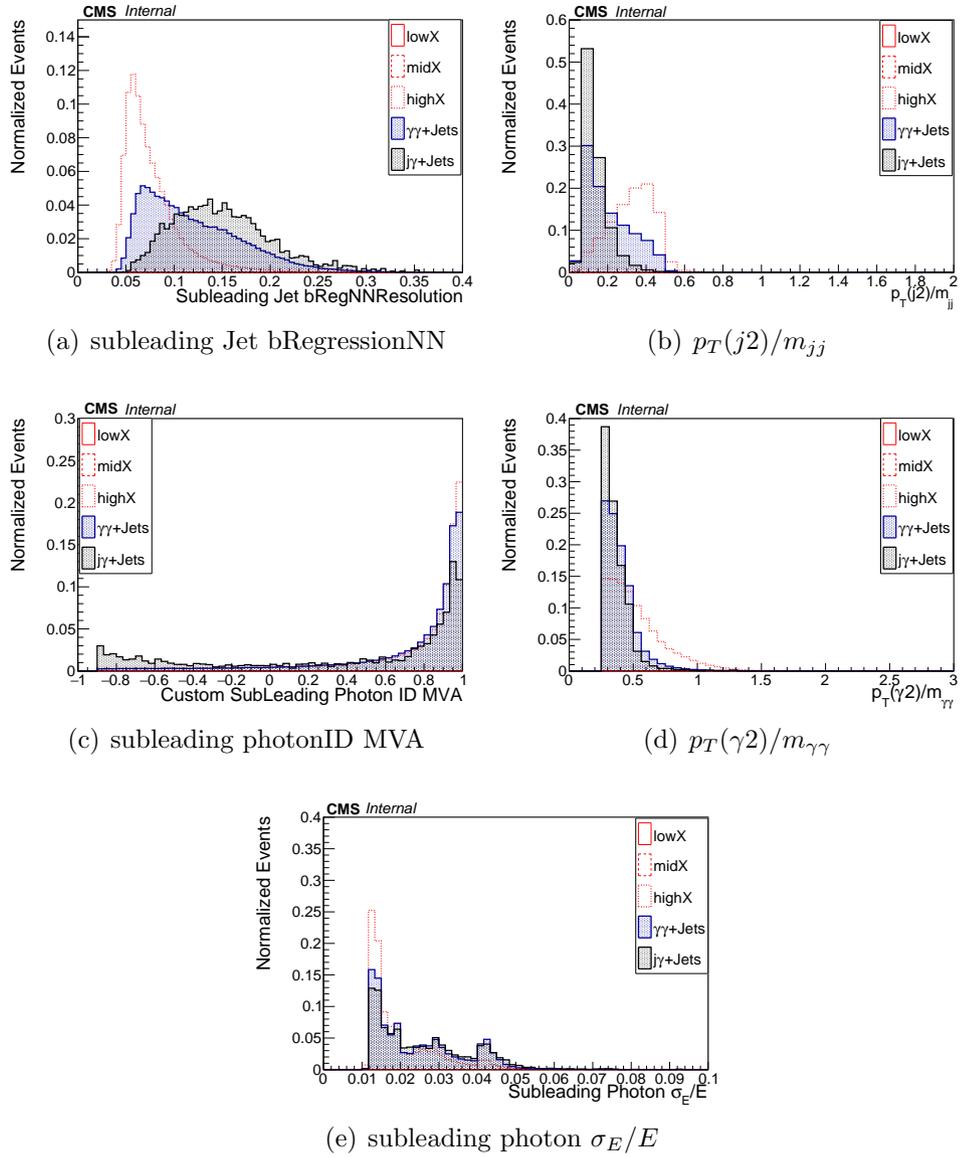
Figure B.6: Signal and background kinematics for highY mass region

B.2   Additional study for BDT training

To make BDT training more efficient, we performed some additional studies by making selection loose and adding more variables to the training.

As we cannot compromise with $H \to \gamma\gamma$ selections, which is the crucial handle for $b\bar{b}\gamma\gamma$ final state, we tried with $H \to b\bar{b}$ p_T selection 20 GeV to gain signal efficiency, especially for low masses. However, it made the training performance worse. Thus we excluded this idea.

Apart from this, we also train by adding $\Delta\eta_{HH}$ to the BDT inputs. However, since we already have $\cos\theta_{HH}$ as input, it does not bring any improvement because both represent the same physics in a different frame. The plots in Fig.B.7 show that there is a negligible improvement with this.
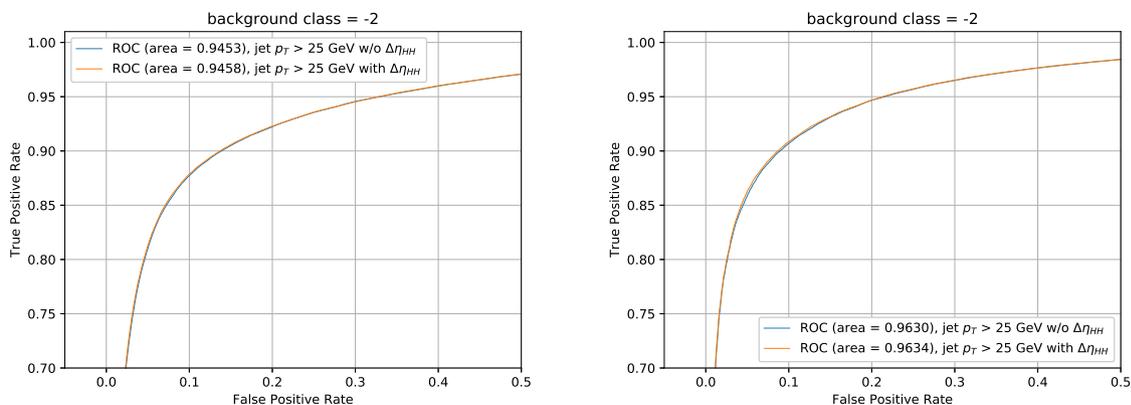
Figure B.7: Training performance with and without $\Delta\eta_{HH}$ variable for $\gamma\gamma$+jets (left) and γj/jj+jets (right) using bulk radion as signal

B.3   ttHkiller cut optimization

This optimization study was performed for non-resonant $HH \to b\bar{b}\gamma\gamma$ analysis to select an optimal cut on ttHkiller discriminator to reject ttH contribution. We used pseudo data (mixture of non-resonant and resonant background simulations) for this study. First, we made a discriminator shape comparison between pseudo data and data from the control

region.

After that, we study the expected improvement on the final limit with different ttHKiller selection thresholds using pseudo data as real data. As shown in Fig. B.8, the optimal selection on the ttHScore lies in the region ttHScore [0.15-0.3]. We optimize the 0.26 cut on the ttHScore, maximizing the expected significance. The improvement with this selection keeps the same order what we observe for resonant analysis.
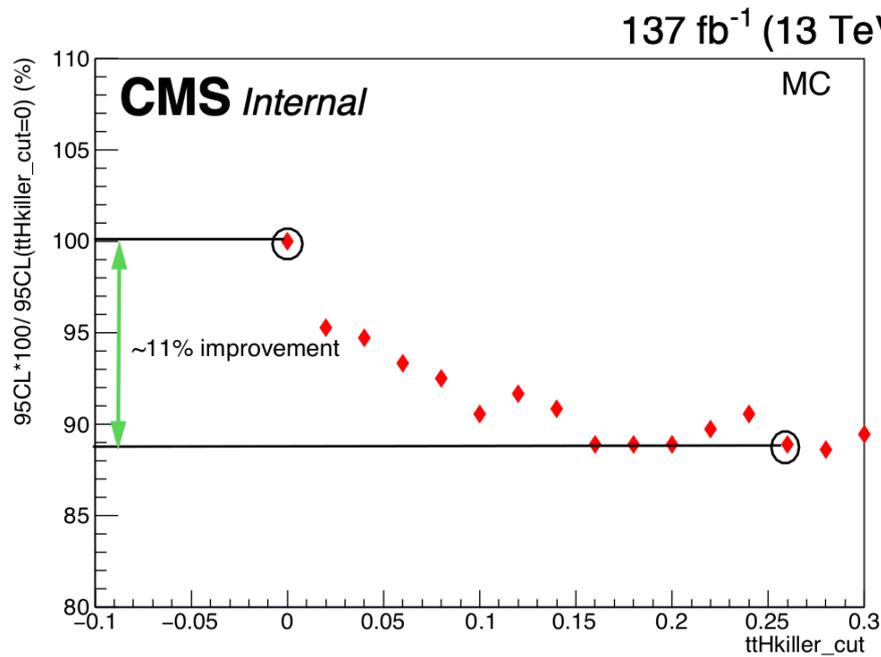
Figure B.8: ttHkiller cut optimization using pseudo data

(a) X=260 GeV

(b) X=270 GeV

(c) X=280 GeV

(d) X=300 GeV

(e) X=320 GeV

(f) X=350 GeV

(g) X=400 GeV

(h) X=450 GeV

(i) X=550 GeV

(j) X=600 GeV

(k) X=650 GeV

(l) X=700 GeV

(m) X=800 GeV

(n) X=900 GeV

(o) X=1000 GeV
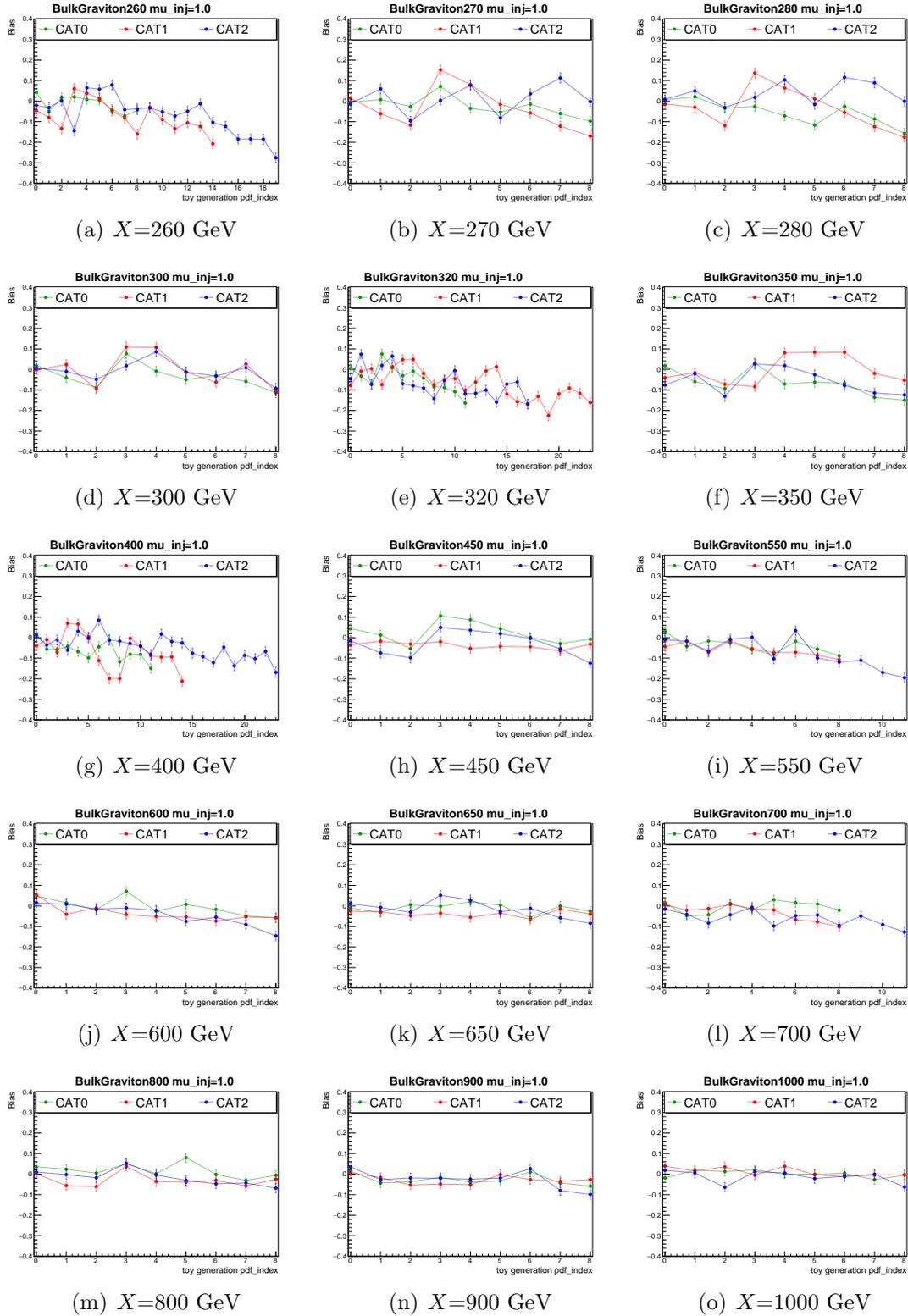
Figure B.9: Bias test results for WED signal and background model

(a) X=300 GeV; Y=90 GeV (b) X=300 GeV; Y=150 GeV (c) X=400 GeV; Y=125 GeV

(d) X=500 GeV; Y=90 GeV (e) X=500 GeV; Y=200 GeV (f) X=500 GeV; Y=300 GeV

(g) X=600 GeV; Y=125 GeV (h) X=600 GeV; Y=250 GeV (i) X=600 GeV; Y=400 GeV

(j) X=800 GeV; Y=125 GeV (k) X=800 GeV; Y=400 GeV (l) X=800 GeV; Y=600 GeV
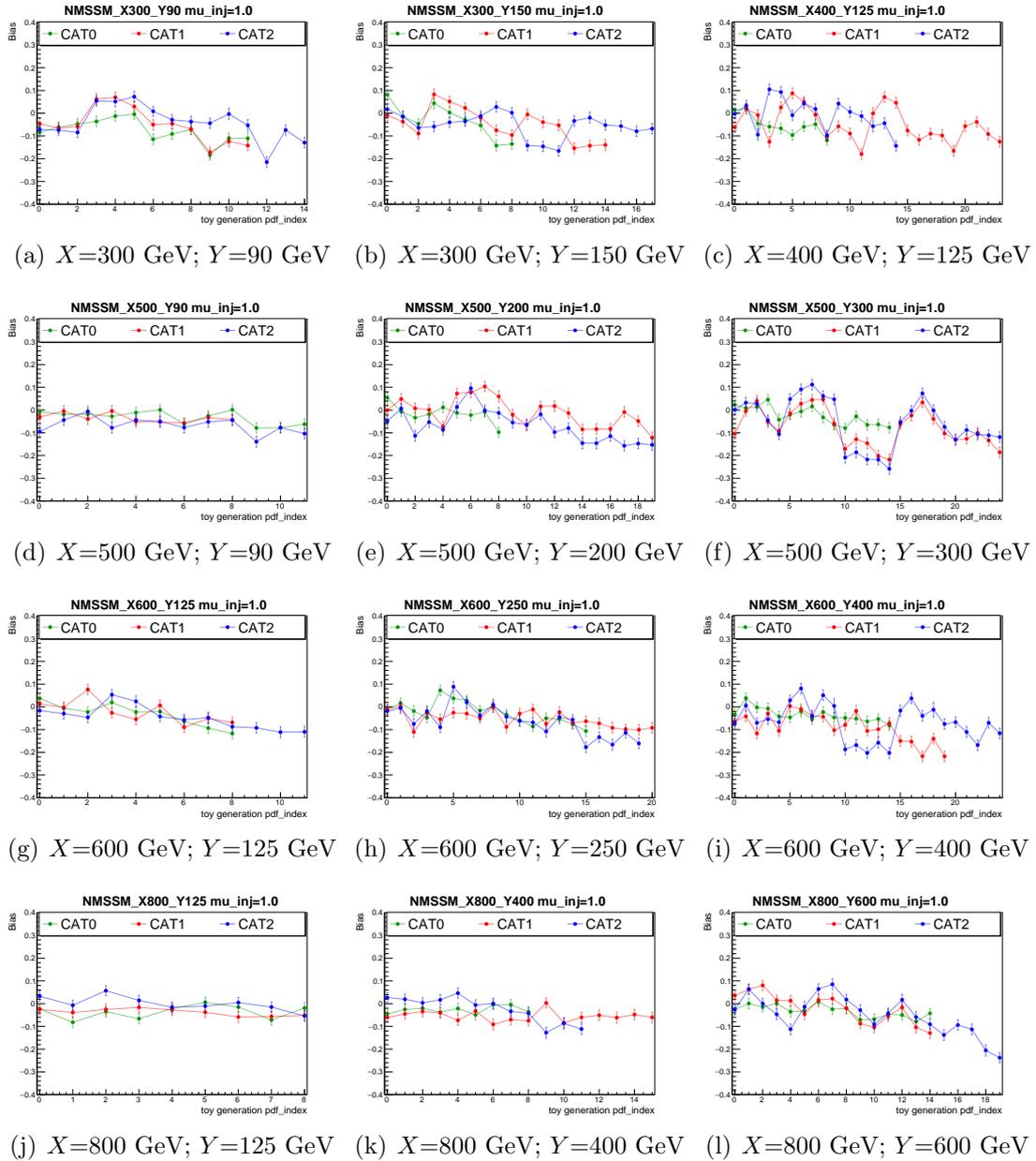
Figure B.10: Bias test results for NMSSM signal and background model

(a) X=900 GeV; Y=200 GeV (b) X=900 GeV; Y=300 GeV (c) X=900 GeV; Y=700 GeV

(d) X=1000 GeV; Y=250 GeV (e) X=1000 GeV; Y=600 GeV (f) X=1000 GeV; Y=800 GeV
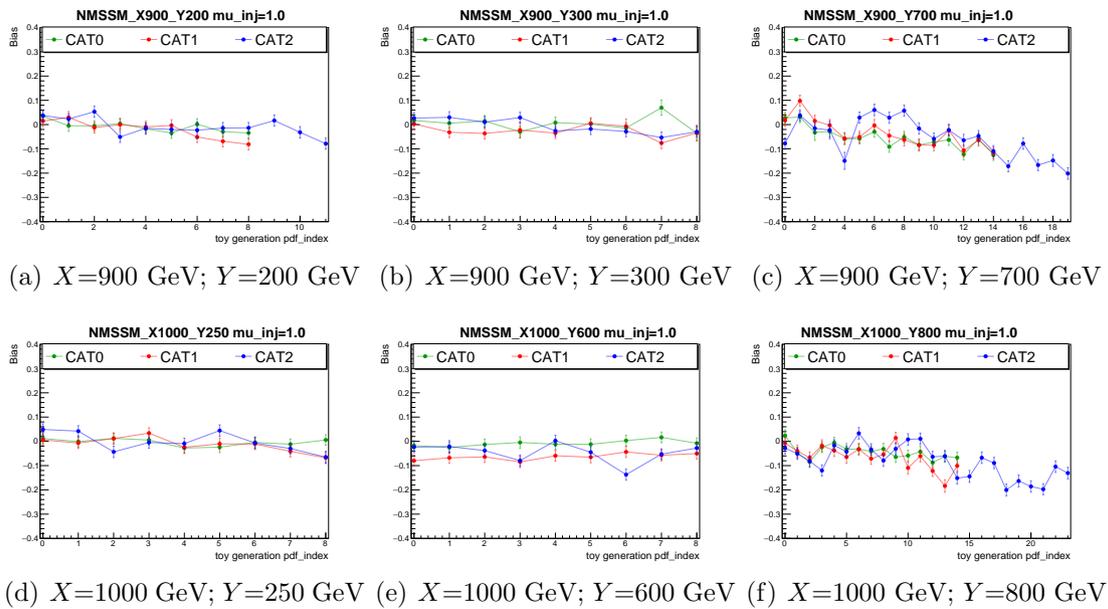
Figure B.11: Bias test results for NMSSM signal and background model
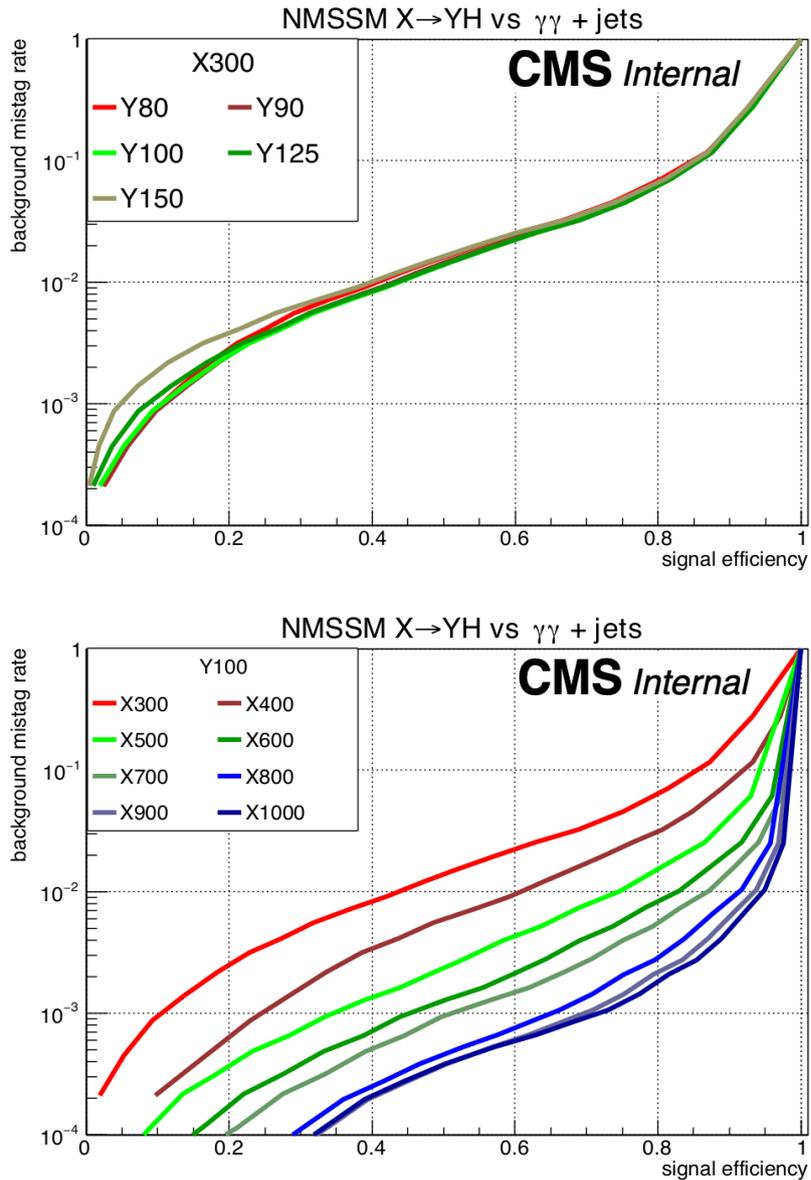
Figure B.12: (Upper) Same training performance for $X=300$ GeV for each Y mass point proves that BDT learns for each signal point; (lower) for constant Y mass training performance improves with X mass as object tends to be in high p_T regime which make kinematics more discriminating from background

B.4   Signal-plus-background fit in Data

The Figs. B.13 and B.14 show signal-plus-background fit of observables in data after unblinding.
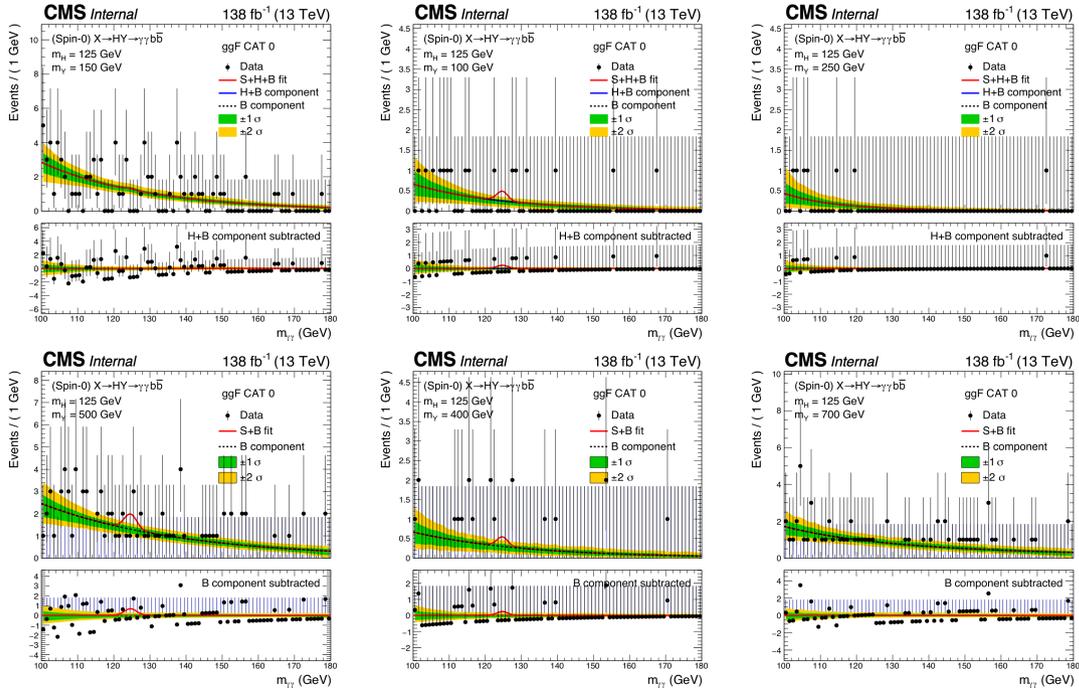
Figure B.13: $m_{\gamma\gamma}$ distributions in data for CAT0 corresponding to six different signal hypotheses
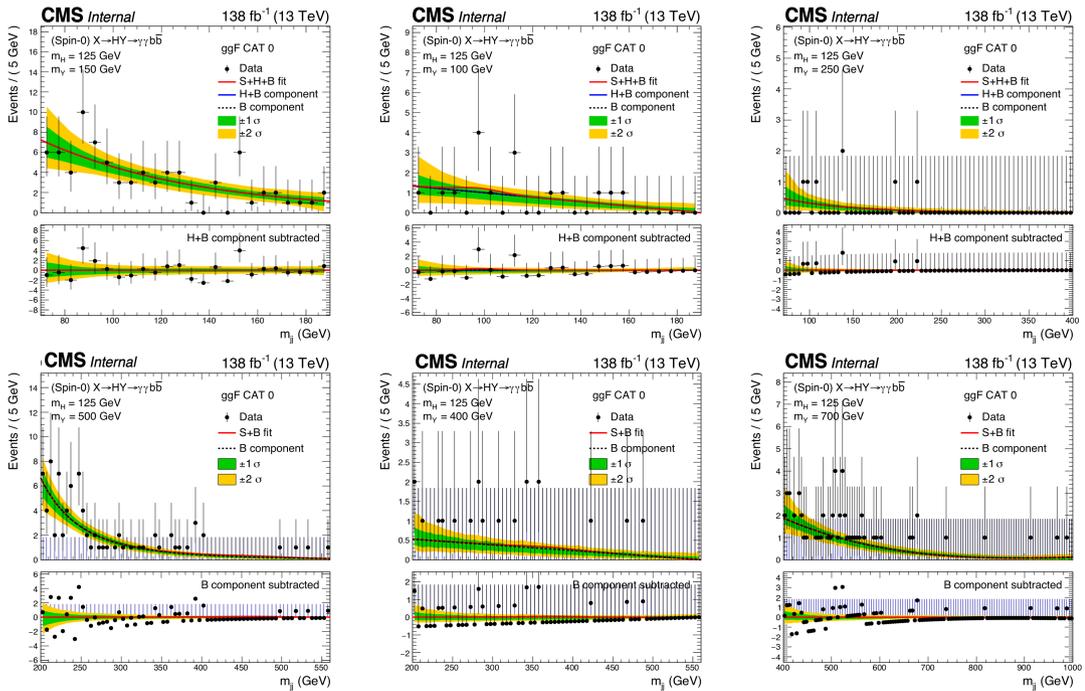
Figure B.14: m_{jj} distributions in data for CAT0 corresponding to six different signal hypotheses

B.5   Comparison of ATLAS and CMS Run-2 results

The Run-2 13 TeV observed and expected results between ATLAS and CMS for $pp \to X \to HH \to bb\gamma\gamma$ process are compared and found up to 30–40% better with CMS analysis strategy even after having the similar performance of object identification methods. Below we mention the points which were helpful to achieve good analysis sensitivity.

- *ttH* **rejection**: The training performed by CMS to minimize contribution of the resonant background *ttH* is more optimal with NN-based method. While ATLAS uses BDT training.

- *HH* **discriminant:** The BDT training performed by ATLAS to build a *HH* discriminant uses all mass points together while we perform training separately for different mass ranges. This helps to make training efficient for each mass point.

- **Event classification:** ATLAS applies a selection of *HH* discriminant and uses events passing that selection while CMS categorizes the events into three MVA categories after a soft selection on *HH* discriminant, which helps to increase the sensitivity.

- **Signal extraction method:** ATLAS uses 1D fit method using $m_{\gamma\gamma}$ observable for fit while we use two-dimensional $m_{\gamma\gamma} : m_{jj}$ fit, which gives up to 10% better sensitivity.

B.6   Approximate   interpretations   of   NMSSM   motivated searches

The results shown in Fig. 5.40 are model-independent; therefore, we are free to compare any BSM model that is close to our assumption of these physics processes. Here we have used decay width of X resonance 1 MeV at generator level. The model which supports the asymmetrical decay to X into two scalars where one is always Higgs boson
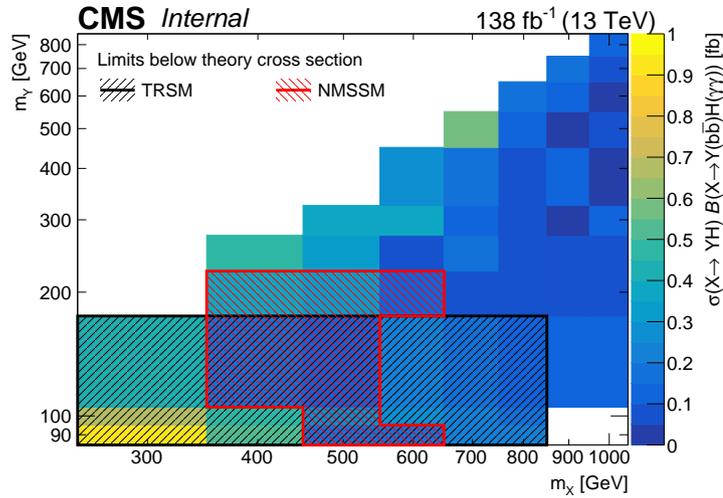
Figure B.15: Approximate interpretation of NMSSM and TRSM models

is NMSSM. However, in terms of width, we are not close to a realistic picture of NMSSM. Therefore, we can make an approximate interpretation.

Similarly, the Two-real-singlet model (TRSM) [190], where two real singlet fields are added to SM Lagrangian, can also be added in interpretation. It gives three neutral CP-even scalar bosons, out of which one is identified with the observed Higgs boson at 125 GeV. Therefore, it supports the decay of one scalar field to SM Higgs boson and another scalar different from H. However, the TRSM model is also very far from our assumption of decay width.

Figure B.15 shows the approximate interpretations of NMSSM and TRSM models where shaded black and red mass regions are ruled out mass parameters for these models. Note that if searches are done for width close to NMSSM and TRSM models, the experimental limits will be worse, and the ruled out mass parameters still can hold. So an exact measurement with BSM model-dependent scenario is required for such interpretations.

Appendix C

Outlook

This chapter focuses on the additional studies which could be made with future research projects to have improvement in the results which look for the same signal. Since thesis research work is divided on HL-LHC and Run-2 data-based searches, I have divided this chapter into two sections corresponding to each search.

C.1   Possible improvements for HL-LHC searches:

- Since analyses is based on object and event selections. Machine learning-based background rejection methods can enhance the projected future sensitivity.

- For VBF topology, the current algorithms and techniques are not efficient to identify VBF jets, so developing a VBF tagger based on machine learning could help in improving the results.

- the subjet b tagging training is based on Run-2 b tagging, which we used in the analysis. Therefore, a Phase-2 specific b tagger training could be important to check how much improvement it could give.

- The VBF analysis only focuses on narrow width searches. We targeted for wide width searches as well but could not add due to the technical issues of MADGRAPH.

It would be interesting to study how much decrements in the significance we get while dealing with wide-width resonances using the same analysis strategy.

C.2 Possible improvements for Run-2 searches:

- It could be worth checking how much improvement we can make with use for DNN based training to build resonant HY-tagger.

- The m_{jj} regression, used for HIG-19-018, is not part of the analysis since it is not trained in the mass range we are exploring. So having that training to improve m_{jj} resolution might provide better sensitivity.

- $b\bar{b}\gamma\gamma$ channel has never explored resonance mass more than 1 TeV. Going for high masses with building boosted strategy would be a completely new analysis in CMS.

- In the future, with more data, we can use more than three event categories with enough events for robust background modeling.

- Currently ttH killer is used from HIG-19-18, training it for resonant searches might improve the results.

- We have only worked for narrow width approximation. Working on wide-width scenarios is more interesting to interpret various physics models like 2HDM, NMSSM and TRSM. With narrow-width approximation, we are far from a real physics scenario to make a realistic interpretation.

- The limits are model-independent, therefore any BSM model with new resonance having significant BR to decay in HH or any two scalars (one close to H) and any theory having two Higgs doublets or two Higgs doublets + one singlet can be interpreted with this. Also, the analysis does not differentiate between scalar and pseudo-scalar, so model interpretation for a pseudo-scalar process can also be added.

Appendix D

Other contributions at CMS

I am the main contributor for the two HL-LHC analyses and one Run-2 analysis, which are added in Chapters 4 and 5. The resonant $b\bar{b}\gamma\gamma$ analysis is led by me as an analysis contact. I have also contributed to various service tasks in physics object groups and physics analysis groups. Moreover, I was part of two more physics analyses related to Higgs physics searches. I am listing other contributions below.

- **Service tasks:**
  MC generator contact for Higgs subgroup: handled the simulation sample requests made by Higgs analysts.
  AlcaDB contact: served as AlCaDB (Alignment and Calibration DataBase) contact for BTV and Particle Flow hadron calibration group; handled database integration of calibrations, performed corresponding validation studies, and followed up with AlcaDB group.

- **microAOD production contact for Higgs subgroup**: handled microAOD production for $H \rightarrow \gamma\gamma$ analyses.

- **Run-2 2016 analysis for VBF $H \rightarrow bb$ searches**: Contributed with the team working on bias study to validate the background modeling and produced expected upper limit results. The study has been added in Appendix E.

- **Run-2 measurement of Higgs trilinear self-coupling with $b\bar{b}\gamma\gamma$ channel**: Served as one of the main contributors of the analysis reported in Ref. [1]. Developed mass regression method to improve dijet mass resolution. Performed validation and optimization studies along with analysis-related service and framework integration tasks.

Appendix E

Bias study for VBF SM Higgs with $b\bar{b}$ final state

We perform this analysis to have a robust background modeling for SM Higgs production signal via VBF mode. We use 2016 data with 35.9 fb^{-1}. The analysis is performed in 9 categories defined according to the selection (SingleB or DoubleB) and the multivariate discriminant (BDT) output. For the SingleB and the DoubleB selection, 5 and 4 categories are defined. The higher categories of each selection have larger significance than the lower ones. All BDT categories are listed in table E.1 together with their purity levels (S/B) and significance values ($S/\sqrt{B}$). The comparison of the actual m_{bb}

Table E.1: Definition of categories in the Higgs search, according to the selection and the discriminating BDT output.

Category	Selection	BDT Range	GF/(GF+VBF)	S/B	$S/\sqrt{B}$ (35.9 fb^{-1})
0	DoubleB	$-1.00 - 0.087$	0.751	0.0002	0.4690
1	DoubleB	$0.087 - 0.578$	0.482	0.0006	0.5605
2	DoubleB	$0.578 - 0.787$	0.298	0.0012	0.5594
3	DoubleB	$0.787 - 1.000$	0.174	0.0025	0.5110
4	SingleB	$-1.00 - 0.077$	0.369	0.000016	0.2649
5	SingleB	$0.077 - 0.578$	0.230	0.0009	0.6203
6	SingleB	$0.578 - 0.777$	0.142	0.0022	0.7849
7	SingleB	$0.777 - 0.877$	0.091	0.0043	0.9138
8	SingleB	$0.877 - 1.000$	0.050	0.0089	1.1054

shapes in data for different MVA output values is shown in Fig. E.1. The histograms are normalized to the unit area, and the Higgs signal region is excluded.
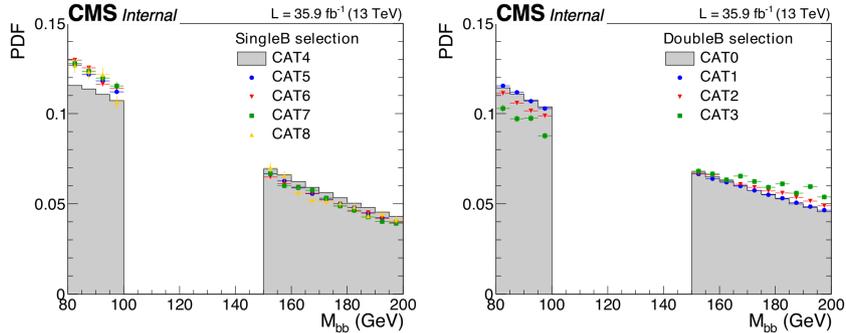
Figure E.1: Shape comparison of the m_{bb} distributions for defined event categories. The excluded region $100 < m_{b\bar{b}} < 150$ GeV corresponds to the $H \to b\bar{b}$ signal region.

The quantitative differences of these shapes are shown in Fig. E.2, where the ratios to the control regions are shown. These transfer functions can be fitted high order polynomial, and the shaded region represents the corresponding uncertainty.

E.1   Signal mass templates

Next to a QCD component and transfer functions, signal templates are required. The stacked VBF and GF signal distributions are fit per mass and per selection with the product of a 3rd order polynomial (combinatorial background) and a crystal-ball function (signal component). The templates are shown in Figs. E.3 and E.4. In Fig. E.5 a summary of the Gaussian core parameters of the templates is shown.

E.2   Background mass templates

The full model is composed of two additional background templates. The top templates (ttbar + single top) are fit with a Gaussian curve. In contrast, the Z templates are fit with the product of a 3rd order polynomial (for the combinatorial background) and a crystal-ball function (the Z peak component). The templates are shown in Figs. E.6-E.9.

E.3   Fits of the mass spectrum

Combining all the parts the full model reads:

$$f_i(m_{bb}) = N_{i,\text{qcd}} \cdot R_i(m_{bb}) \cdot Q(m_{bb}; \vec{p}) + N_{i,\text{top}} \cdot T_i(m_{bb}; k_{\text{JES}}, k_{\text{JER}}) + \cdot N_{i,Z} \cdot Z_i(m_{bb}; k_{\text{JES}}, k_{\text{JER}})$$
$$+ \mu_H \cdot N_{i,\text{H}} \cdot H_i(m_{bb}; k_{\text{JES}}, k_{\text{JER}})$$

In the above equation, the index i runs over the categories, $N_{i,\text{qcd}}$ is the free normalization parameter for the QCD background, and $N_{i,\text{top}}$, $N_{i,Z}$ are the top and Z yields, allowed to be floating within 20% of the expectation. The Higgs normalization is multiplied with the μ_{H} signal strength parameter. The top, Z and Higgs components share the k_{JES} and k_{JER} nuisance parameters. The QCD shape $Q(m_{bb}; \vec{p})$ is a 5th order polynomial for DoubleB selection categories and 7th order polynomial for SingleB selection categories. Linear transfer functions $R_i(m_{bb})$ account for the small differences between the control and other categories.

E.4   Fit bias studies

For the Higgs signal search, the functional form for the QCD background shape is chosen following robustness studies of the fitting procedure, verifying that the signal can be measured in an unbiased way even if the QCD model follows different functional shapes compatible with the data distributions. Again, since we cannot determine the correct parameterization of the QCD template shape from physics arguments, we have to take into account that a variety of functions - giving a good fit to the QCD background - could be the true underlying model. Therefore, we have to look for a functional form that can measure the amount of signal without introducing a bias, meaning we can measure the correct amount of signal independent of the underlying model producing the QCD contribution. From previous studies, it was shown that a polynomial form is an appropriate parameterization to guarantee such a demand. The order of the polynomial

is determined by performing the bias studies.

In the following, we will show that, independent of the smooth functional form that is producing the QCD contribution, we can measure the amount of Higgs boson signal in an unbiased way.

E.4.1   Alternative QCD modeling functions

To parameterize the QCD shape, many different functional forms have been taken into account. The final selection is based on those that are able to fit the m_{bb} data distribution in the lowest categories (CAT0 and CAT4) with a χ^2 probability larger than 5%, and with a reasonable number of parameters. With χ^2 probability, F-probability is also calculated to get the appropriate degree of the function for QCD-fit in data. When F-probability is larger than 5% for an order of any function, then lower order is used to fit the QCD background.

The following functions are finally used:

- $\exp(\mathrm{pol(x)}) = \exp(-ax - bx^2....)$

- $\mathrm{modG(x)} = \exp(-ax - bx^2....) \cdot \mathrm{erfc}[(b - cx)]$

- $\tanh(\mathrm{x}) = [a - b\tanh(cx + d) - e\tanh^2(cx + d)....]$

- $\sin\mathrm{e(x)} = [1 + a\sin(bx + c) + d\sin^2(bx + c)...]$

- $\mathrm{x}^{Pol(x)} = x^{(ax + bx^2...)}$

- $\mathrm{polydijet} = x^{(a + b\log(x))} * (1 + cx + dx^2 + ...)$

- $\mathrm{Invpolydijet} = x^{(a + b\log(x))} * (1 + c(1/x) + d(1/x)^2 + ...)$

Each represents quite different shapes with respect to the polynomial family that is chosen to perform the QCD fit.

The fit of SingleB CAT4 data with the above alternative functions are shown in Fig. E.10, while the fits of CAT4 data with the polynomial family is shown in Fig. E.11.

For alternative functions as we conclude from Tables E.2 - E.8, that in SingleB CAT4 category tanh and sine functions with 1st order, Invpolydijet with 2nd order, polydijet and $x^{Pol(x)}$ with 3rd order, and modG and exp(pol(x)) with 4th order fit well. It can be noted from Fig. E.10 that to describe the data correctly with a polynomial function, a 7th degree would be sufficient.

The fit of DoubleB CAT0 data with the above functions are shown in Fig. E.12, while the fits of CAT0 data with the polynomial family is shown in Fig. E.13. For alternative functions as we conclude from Tables E.2 - E.8, that in DoubleB CAT0 category modG and sine functions with 1st order, Invpolydijet, exp(pol(x)) and tanh with 2nd order, polydijet and $x^{Pol(x)}$ with 3rd order fit well. It can be noted from Fig. E.10 that to describe the data correctly with a polynomial function, a 5th degree would be sufficient.

Table E.2: Selection of sine function with 1st order for both SingleB CAT4 and DoubleB CAT0, based on 5% thresholds on both χ^2 probabiity and F-probabiity

Selection	Order	χ^2	NDF	χ^2/NDF	Probabiity	F-probabiity
SingleB	n=1	1254.75	1197	1.05	0.12	-
	n=2	1251.69	1194	1.05	0.12	0.08
	n=3	1251.68	1191	1.05	0.11	0.92
	n=4	1252.27	1188	1.05	0.09	0.00
	n=5	1257.22	1185	1.06	0.07	0.00
DoubleB	n=1	1173.32	1197	0.98	0.62	-
	n=2	1171.59	1194	0.98	0.67	0.19
	n=3	1132.88	1191	0.95	0.88	0.00
	n=4	1133.26	1188	0.95	0.87	0.00
	n=5	1134.26	1185	0.95	0.85	0.00

Table E.3: Selection of tanh function with 1st and 2nd order for SingleB CAT4 and DoubleB CAT0,respectively , based on 5% thresholds on both χ^2 probability and F-probability

Selection	Order	χ^2	NDF	χ^2/NDF	Probability	F-probability
	n=1	1253.92	1196	1.05	0.12	-
	n=2	1253.80	1193	1.05	0.11	0.73
SingleB	n=3	1252.51	1190	1.05	0.10	0.26
	n=4	1247.76	1187	1.05	0.11	0.03
	n=5	1242.37	1184	1.05	0.12	0.02
	n=1	1158.85	1196	0.97	0.77	-
	n=2	1147.93	1193	0.96	0.82	0.00
DoubleB	n=3	1147.77	1190	0.96	0.81	0.70
	n=4	1149.03	1187	0.97	0.78	0.00
	n=5	1149.00	1184	0.97	0.76	0.87

Table E.4: Selection of Invpolydijet function with 2nd for both SingleB CAT4 and DoubleB CAT0, based on 5% thresholds on both χ^2 probability and F-probability

Selection	Order	χ^2	NDF	χ^2/NDF	Probability	F-probability
	POL1, n=1	1385.19	1197	1.16	0.00	-
	POL2, n=2	1269.89	1196	1.06	0.07	0.00
SingleB	POL3, n=3	1268.00	1195	1.06	0.07	0.00
	POL4, n=4	1268.07	1194	1.06	0.07	0.00
	POL5, n=5	1269.60	1193	1.06	0.06	0.00
	POL1, n=1	1263.86	1197	1.06	0.09	-
	POL2, n=2	1231.90	1196	1.03	0.23	0.00
DoubleB	POL3, n=3	1231.95	1195	1.03	0.22	0.00
	POL4, n=4	1234.30	1194	1.03	0.20	0.00
	POL5, n=5	1234.98	1193	1.03	0.19	0.00
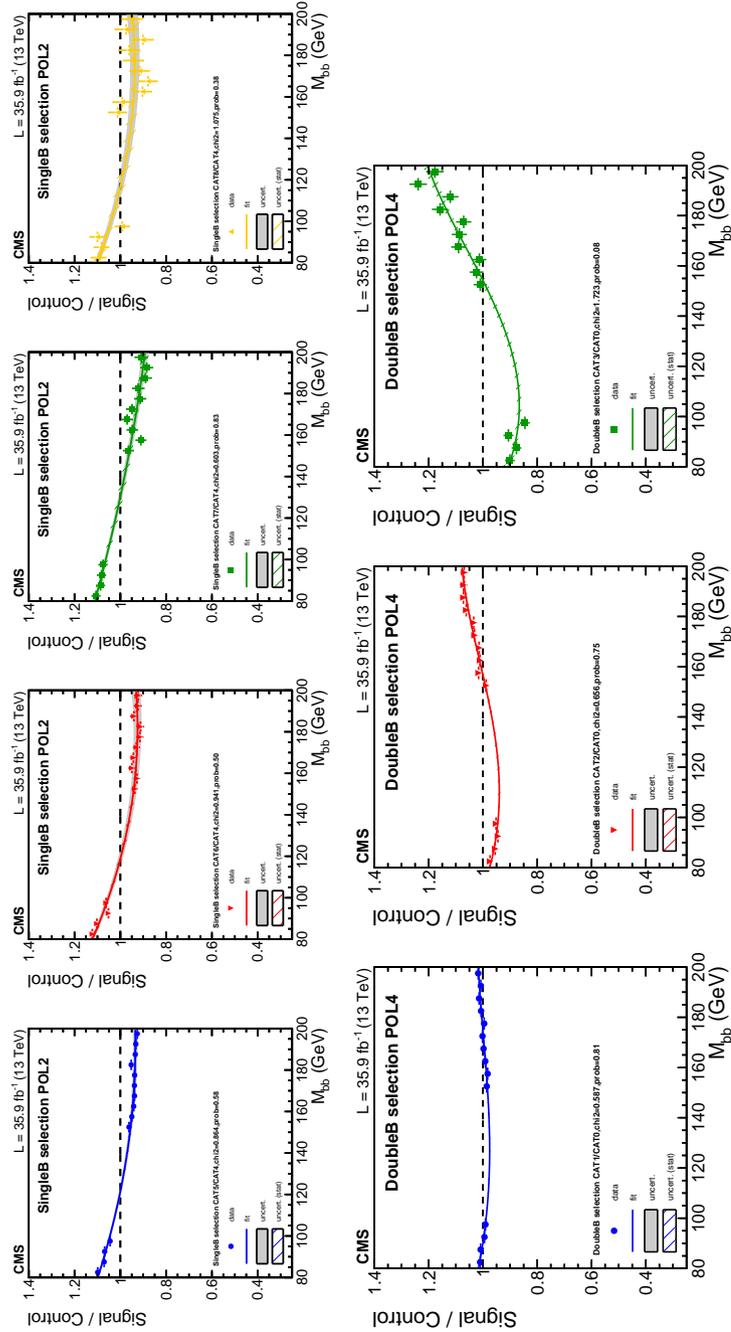
Figure E.2: Pre-fitted QCD shape transfer functions excluding the $100 < m_{bb} < 150$ GeV region. The top panel shows the fits for the SingleB transfer functions, the down panel the fits for the DoubleB transfer functions
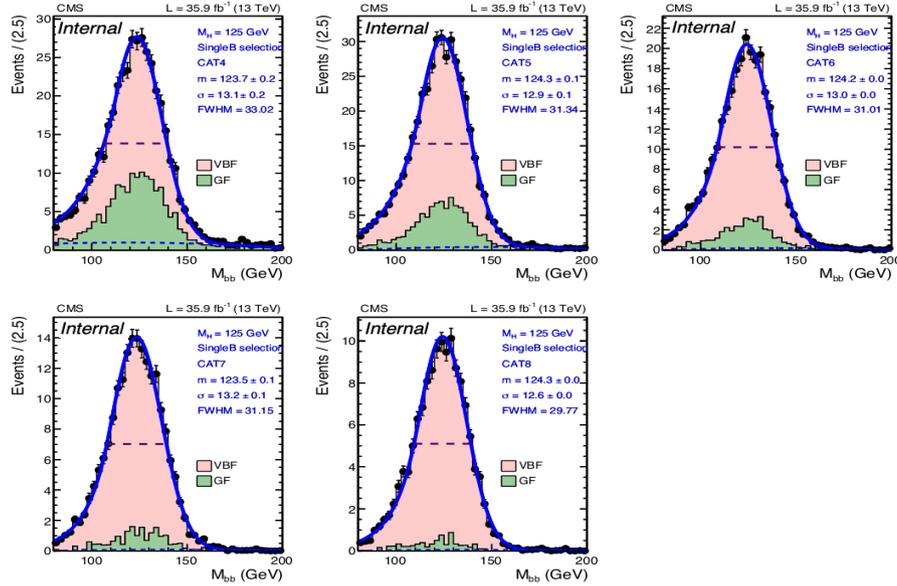
Figure E.3: Templates of m_{bb} for $m_H = 125$ GeV in the SingleB selection. The histograms are scaled to the expected number of events. The VBF and GF contributions are weighted with the respective cross sections and stacked.
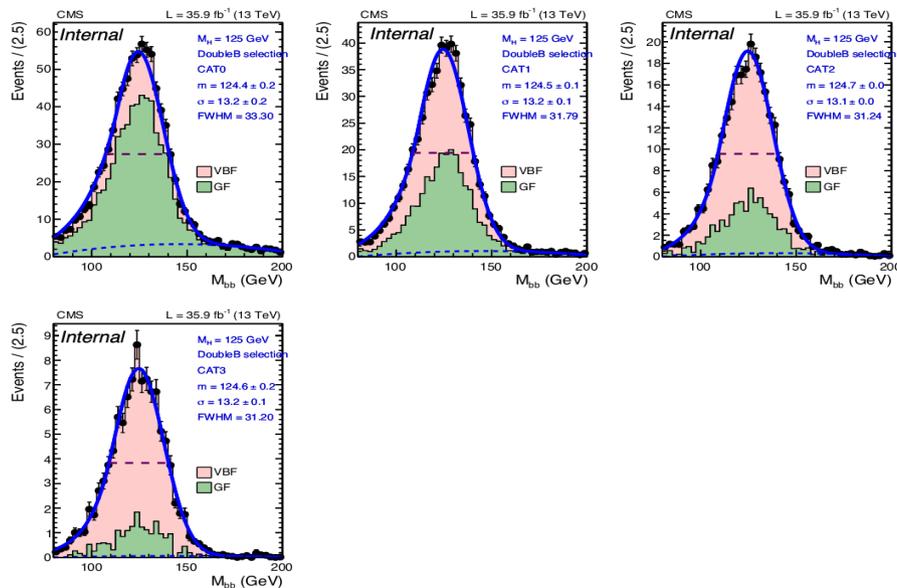
Figure E.4: Templates of m_{bb} for $m_H = 125$ GeV in the DoubleB selection. The histograms are scaled to the expected number of events. The VBF and GF contributions are weighted with the respective cross sections and stacked.
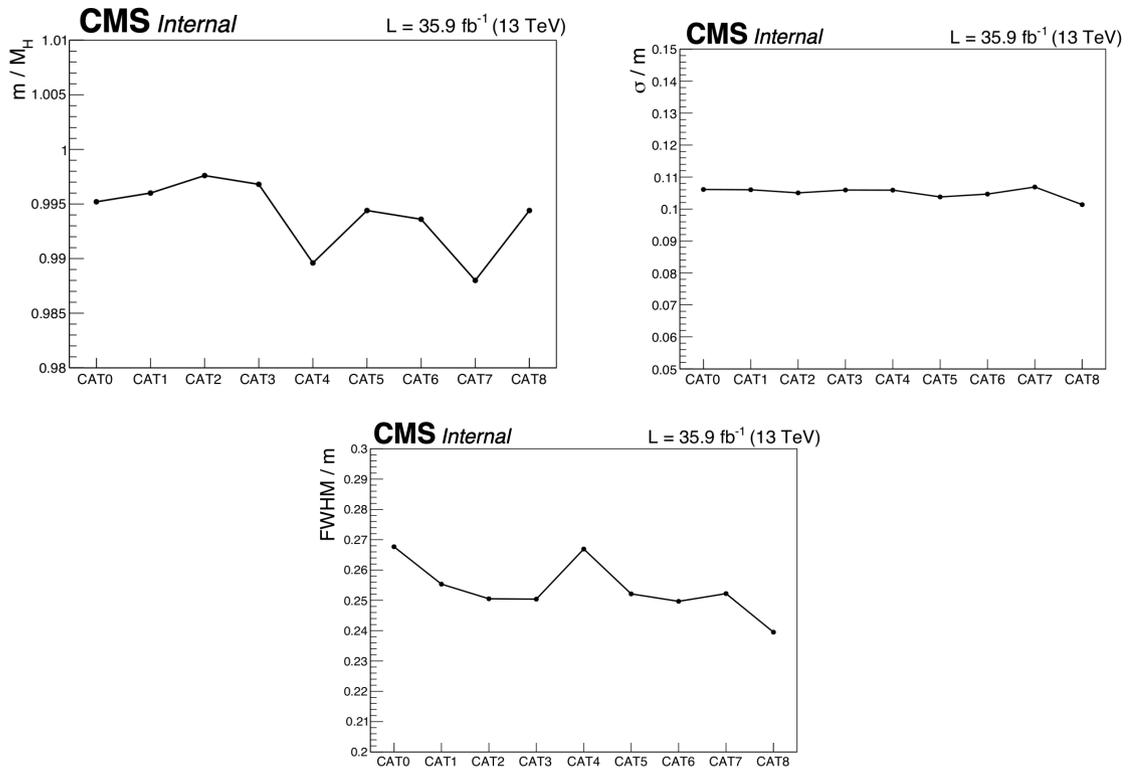
Figure E.5: Summary of the Gaussian core parameters of the Higgs templates. The categories $0 - 3$ refer to the DoubleB selection, and the categories $4 - 8$ refer to the SingleB selection.
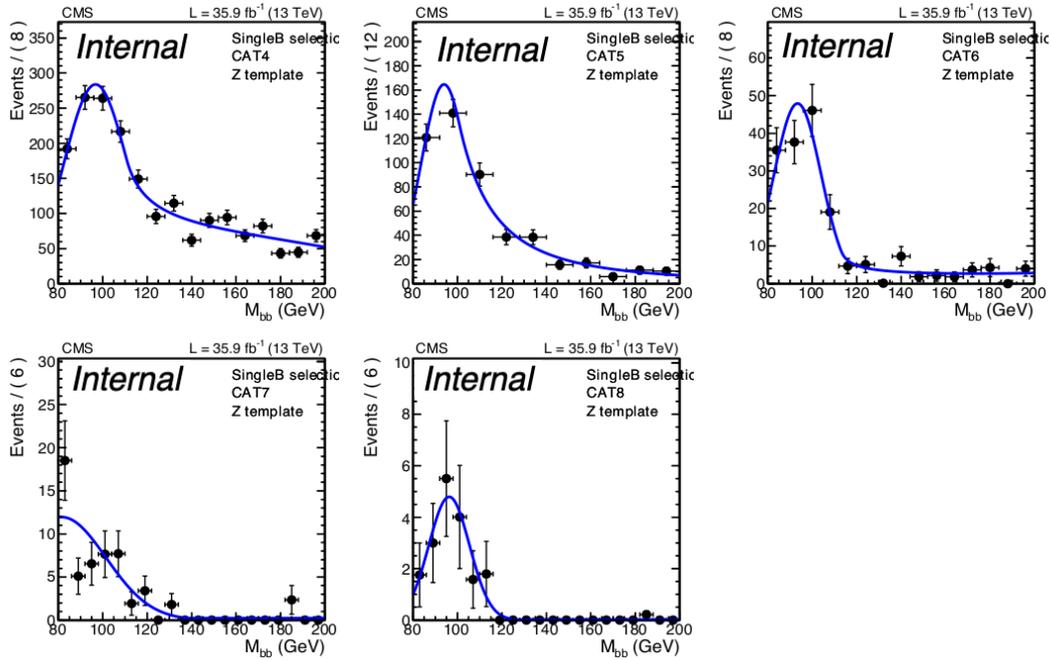
Figure E.6: Templates of m_{bb} for the $Z/W+$jets background in the SingleB selection.
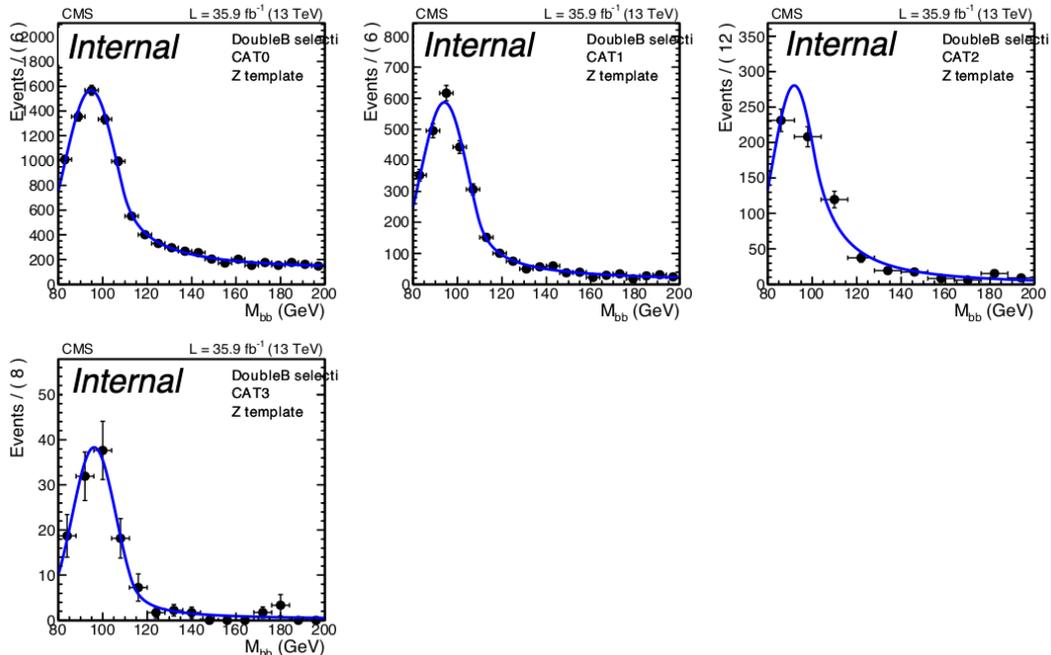
Figure E.7: Templates of m_{bb} for the $Z/W+$jets background in the DoubleB selection.
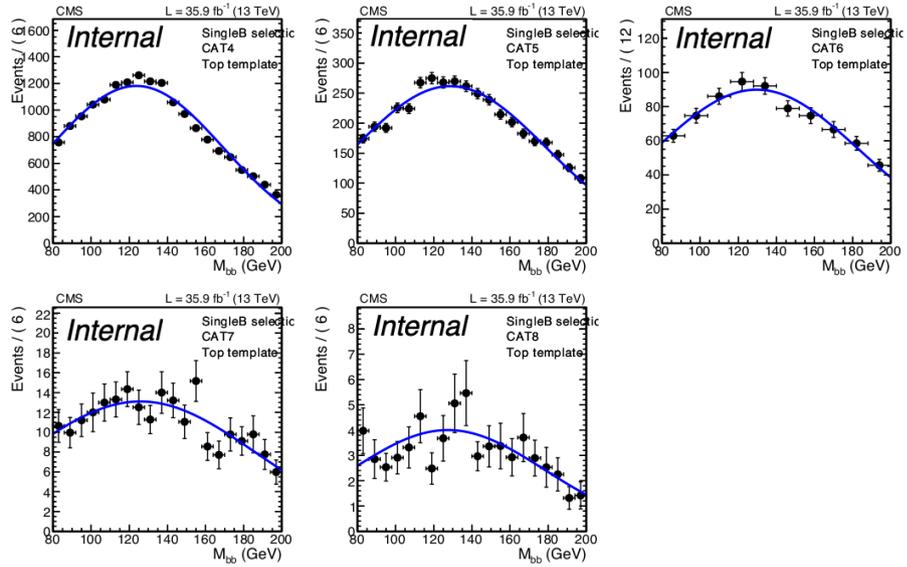
Figure E.8: Templates of m_{bb} for the top (ttbar + single top) background in the SingleB selection.
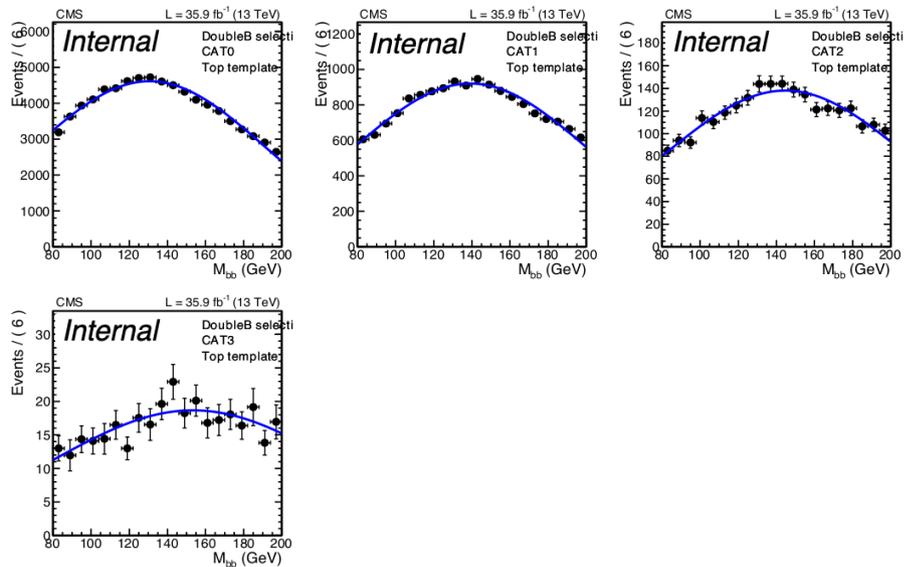
Figure E.9: Templates of m_{bb} for the top (ttbar + single top) background in the DoubleB selection.
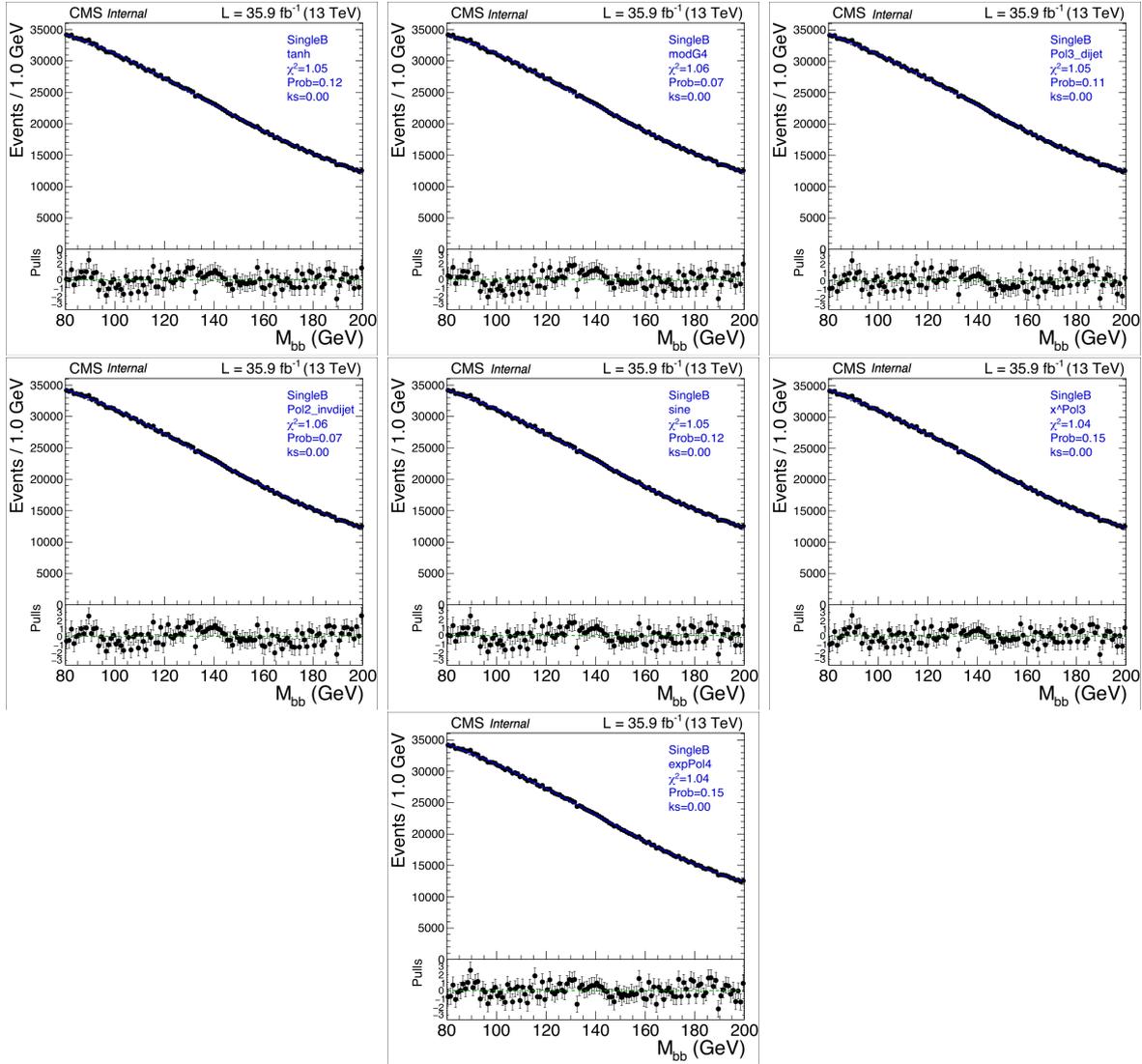
Figure E.10: Fit for the Higgs boson: background model fit in the SingleB CAT4 control region with the alternative QCD shape functions.
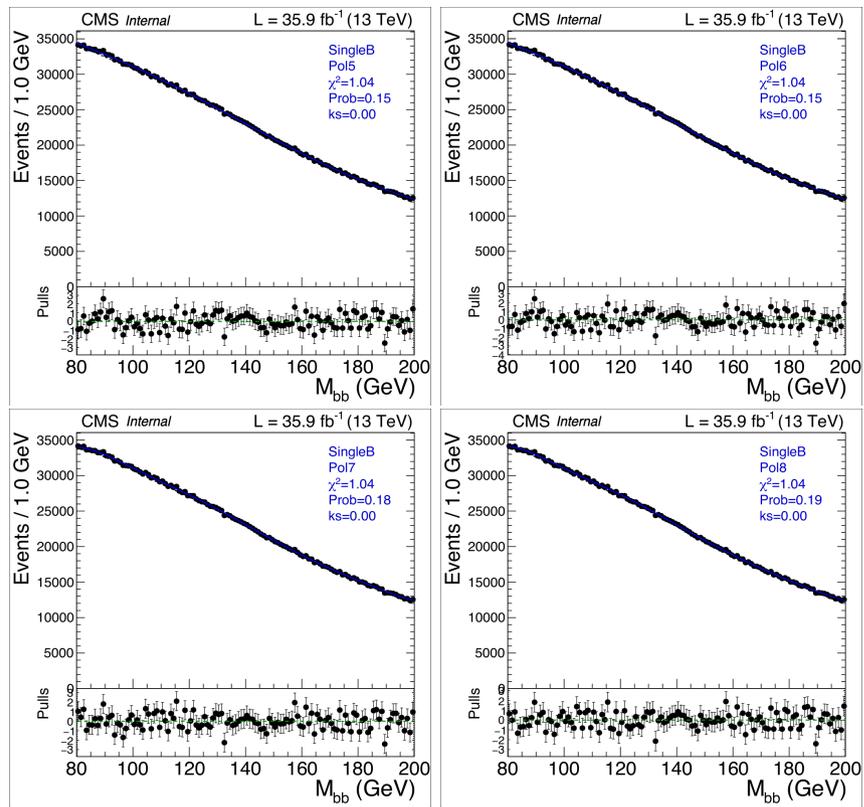
Figure E.11: Fit for the Higgs boson: background model fit in the SingleB CAT4 control region with polynomial QCD shape functions.
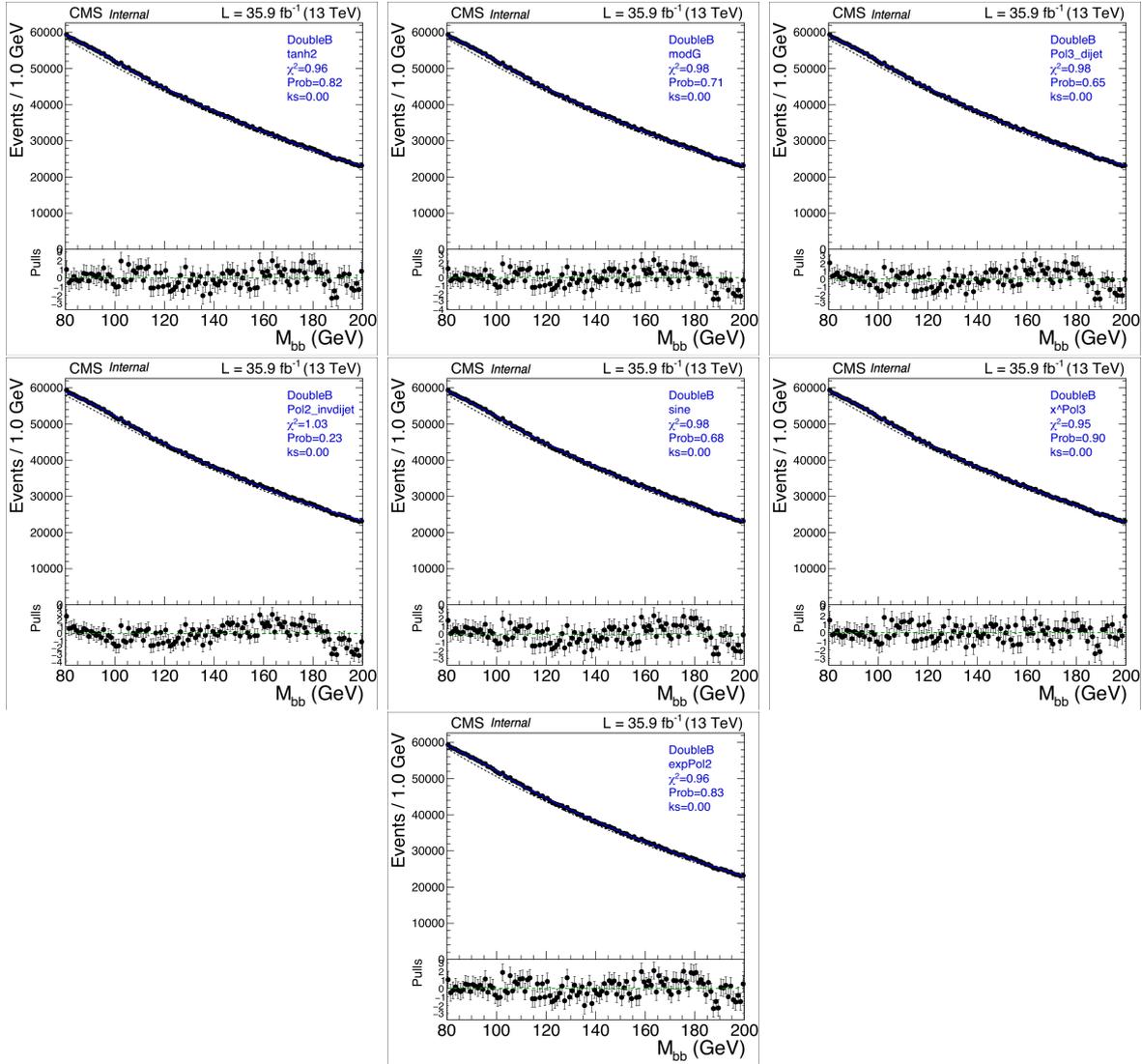
Figure E.12: Fit for the Higgs boson: background model fit in the DoubleB CAT0 control region with the alternative QCD shape functions.
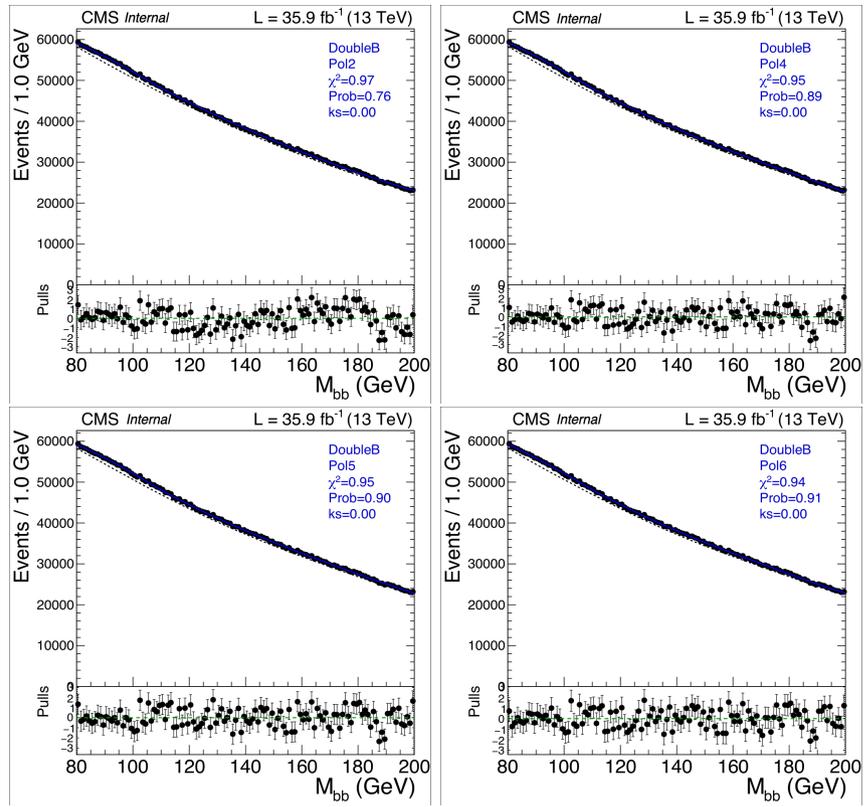
Figure E.13: Fit for the Higgs boson: background model fit in the DoubleB CAT0 control region with polynomial QCD shape functions.

Table E.5: Selection of Polydijet function with 3rd for both SingleB CAT4 and DoubleB CAT0, based on 5% thresholds on both χ^2 probability and F-probability

Selection	Order	χ^2	NDF	χ^2/NDF	probability	F-probabiity
SingleB	POL1, n=1	1313.88	1197	1.10	0.01	-
	POL2, n=2	1274.06	1196	1.06	0.06	0.00
	POL3, n=3	1256.48	1195	1.05	0.10	0.00
	POL4, n=4	1254.04	1194	1.05	0.11	0.12
	POL5, n=5	1251.29	1193	1.05	0.12	0.10
	POL6, n=6	1245.81	1192	1.04	0.14	0.02
	POL7, n=7	1244.99	1191	1.04	0.13	0.37
	POL8, n=8	1245.43	1190	1.05	0.13	0.00
DoubleB	POL1, n=1	1277.70	1197	1.07	0.05	-
	POL2, n=2	1269.89	1196	1.06	0.07	0.00
	POL3, n=3	1175.15	1195	0.98	0.65	0.00
	POL4, n=4	1174.99	1194	0.98	0.65	0.70
	POL5, n=5	1173.43	1193	0.98	0.65	0.20
	POL6, n=6	1495.52	1192	1.25	0.00	0.00
	POL7, n=7	5027.35	1191	4.22	0.00	0.00

Table E.6: Selection of x$^{Pol(x)}$ function with 3rd for both SingleB CAT4 and DoubleB CAT0, based on 5% thresholds on both χ^2 probability and F-probability

Selection	Order	χ^2	NDF	χ^2/NDF	Probability	F-probability
SingleB	POL1, n=1	1163.36	1198	0.97	0.76	-
	POL2, n=2	1147.39	1197	0.96	0.84	0.00
	POL3, n=3	1133.01	1196	0.95	0.90	0.00
	POL4, n=4	1132.75	1195	0.95	0.90	0.60
	POL5, n=5	1133.67	1194	0.95	0.89	0.00
DoubleB	POL1, n=1	1163.36	1198	0.97	0.76	-
	POL2, n=2	1147.39	1197	0.96	0.84	0.00
	POL3, n=3	1133.01	1196	0.95	0.90	0.00
	POL4, n=4	1132.75	1195	0.95	0.90	0.60
	POL5, n=5	1133.67	1194	0.95	0.89	0.00

Table E.7: Selection of modG function with 4th and 1st order for SingleB CAT4 and DoubleB CAT0, respectively, based on 5% thresholds on both χ^2 probability and F-probability

Selection	Order	χ^2	NDF	χ^2/NDF	Probability	F-probability
SingleB	n=1	1282.30	1197	1.07	0.04	-
	n=2	1284.48	1196	1.07	0.04	0.00
	n=3	1277.89	1195	1.06	0.05	0.01
	n=4	1267.21	1194	1.06	0.07	0.00
	n=5	1259.82	1193	1.06	0.09	0.00
	n=6	517785.50	1192	434.38	0.00	0.00
SingleB	n=1	1169.31	1197	0.98	0.71	-
	n=2	1164.39	1196	0.97	0.74	0.03
	n=3	1145.13	1195	0.96	0.85	0.00
	n=4	1148.95	1194	0.96	0.82	0.00
	n=5	1151.17	1193	0.97	0.80	0.00
	n=6	2963.24	1192	2.48	0.00	0.00

Table E.8: Selection of exp(pol(x)) function with 4th and 2nd order for SingleB CAT4 and DoubleB CAT0, respectively, based on 5% thresholds on both χ^2 probability and F-probability

Selection	Order	χ^2	NDF	χ^2/NDF	Probability	F-probability
SingleB	POL1, n=1	4662.98	1199	3.90	0.00	-
	POL2, n=2	1373.08	1198	1.15	0.00	0.00
	POL3, n=3	1277.82	1197	1.07	0.05	0.00
	POL4, n=4	1247.59	1196	1.04	0.15	0.00
	POL5, n=5	1247.59	1195	1.04	0.14	0.98
	POL6, n=6	1247.59	1195	1.04	0.14	0.97
DoubleB	POL1, n=1	1480.83	1199	1.23	0.00	-
	POL2, n=2	1152.05	1198	0.96	0.83	0.00
	POL3, n=3	1144.52	1197	0.96	0.86	0.00
	POL4, n=4	1132.71	1196	0.95	0.90	0.00
	POL5, n=5	1132.70	1195	0.95	0.90	0.96
	POL5, n=5	1132.70	1194	0.95	0.90	0.00

E.4.2   Evaluation of biases

As for the Z-peak fit, the bias evaluations are done in two steps.

1. Generation of pseudo-data using the full model with different QCD parameterizations:

   The full model is the one described by Eq. E.1 for all categories. The Z signal, W, and top shapes and normalization are from a fit to MC in every category. The QCD shape is determined from a fit to data. For each QCD function 1000 pseudo-datasets are generated with a bin width of 0.1 GeV and m_{bb} in [80, 200]. This is repeated for a signal injection of 0, 1, and 2 times the SM expectation.

2. Fitting to the pseudo data :

   A binned maximum likelihood fit is performed with the full model, where a polynomial of varying order describes the QCD shape. The distributions of fit results are finally checked to access the measurement bias for each QCD parameterization.

For both the generation of the pseudo-data and the fit, we make use of the Combine tool.

The signal bias is defined as the difference of the mean measured signal strength and the injected amount of signal divided by the fitted signal uncertainty, as Bias $= (\mu_{\mathrm{measured}} - \mu_{\mathrm{injected}})/\sigma_\mu)$. The results are given a QCD fit model containing a polynomial of 5th and 7th order as QCD component.

Figures E.14 (E.15) show the distributions of fitted signal strengths in pseudo-data, and the corresponding observed biases to the alternative QCD modeling when using respectively a polynomial of degree 7 (5) as the main QCD model to fit the SingleB (DoubleB) categories.

It is clear that bias effects of the polynomial family can be problematic with respect to the alternative functions, and become overall acceptable only when the degree of the main polynomial is raised to higher order. For this reason, the chosen QCD functions, to perform the Higgs fit, are a 5th, and a 7th-degree polynomial for DoubleB and
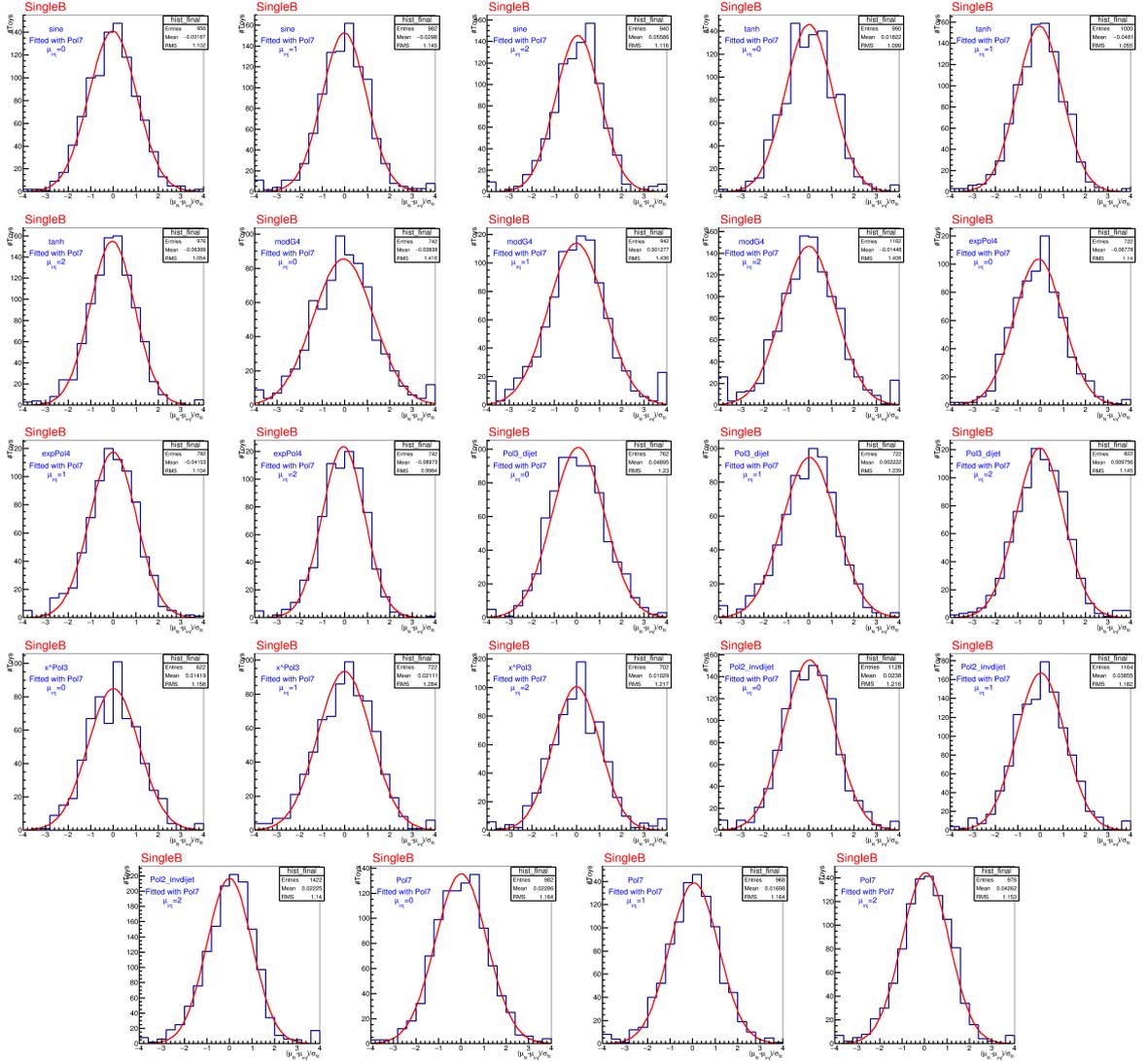
Figure E.14: Fits for the Higgs signal with the SingleB sample. Distributions of pseudo-data fit results using a polynomial of **7th** degree as a QCD model, and alternative QCD shapes to produce the pseudo-data. Results from the closure test are also included where toys are generated and fitted with nominal 7th order polynomial in SingleB selection. Results of fits are shown for pseudo-data with zero, one, and two times the SM injected signal strengths.
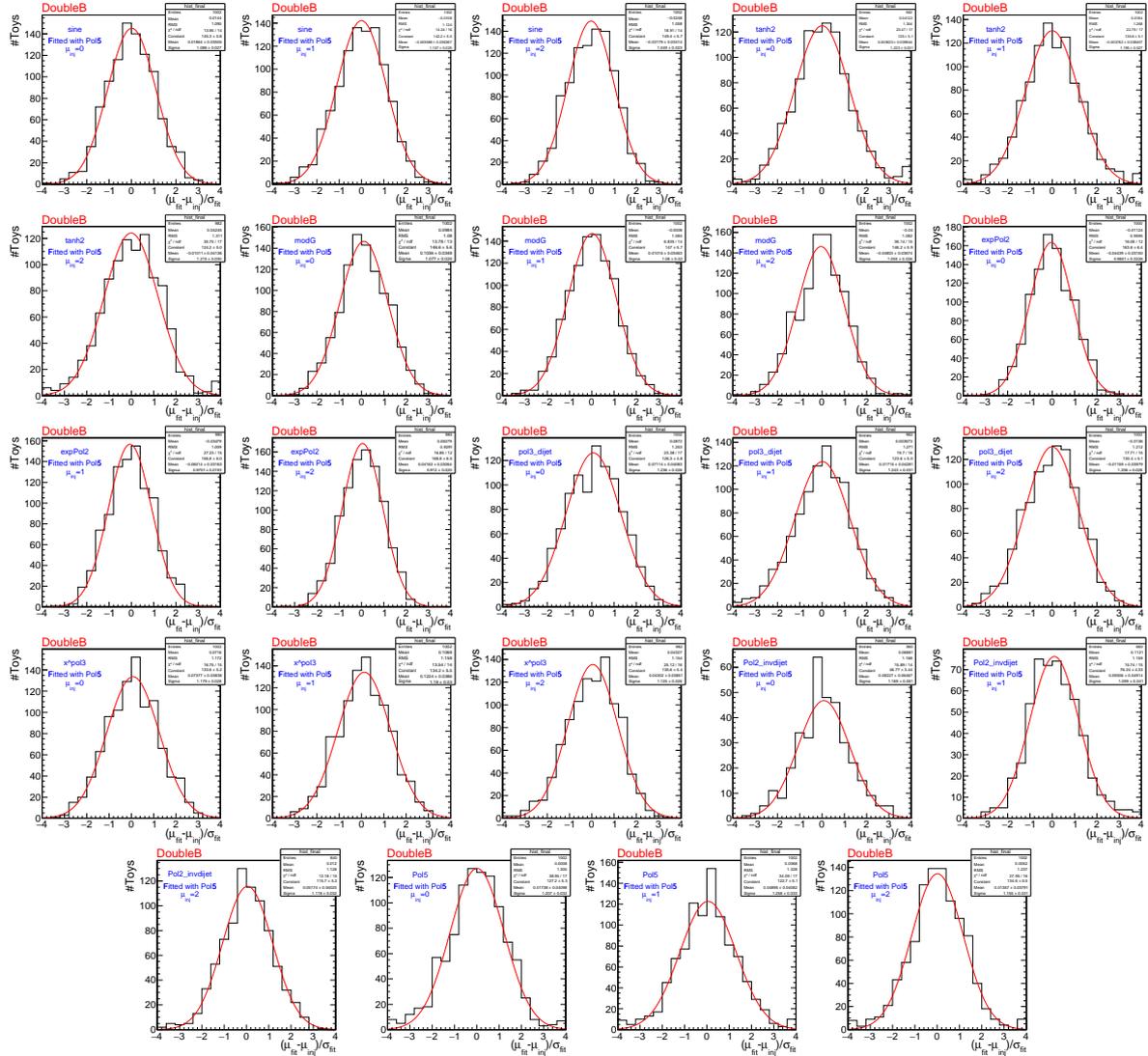
Figure E.15: Fits for the Higgs signal with the DoubleB sample. Distributions of pseudo-data fit results using a polynomial of **5th** degree as a QCD model, and alternative QCD shapes to produce the pseudo-data. Results from closure test are also included where toys are generated and fitted with nominal 5th order polynomial in DoubleB selection. Results of fits are shown for pseudo-data with zero, one and two times the SM injected signal strengths.
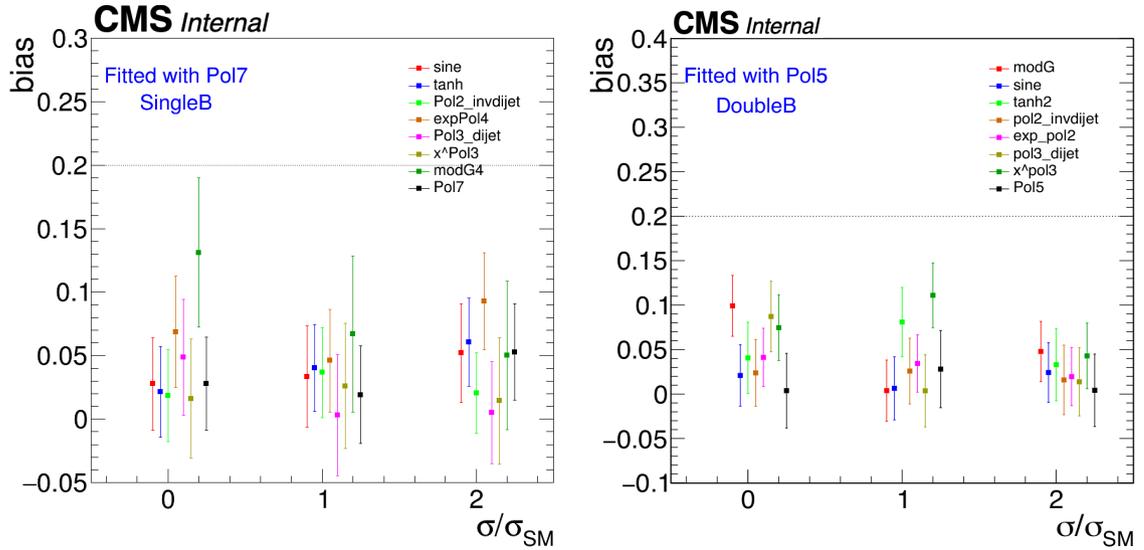
Figure E.16: Summary for observed bias from Figs. E.14 and E.15 for mu injection value 0, 1 and 2.

SingleB selection categories and E.16 shows that observed bias is within 15% threshold, so conclude that 5th and 7th order functions for DoubleB and SingleB categories are good enough to fit QCD background in data.

References

[1] CMS Collaboration. Search for non-resonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2021(3):257, April 2021. `arXiv:2011.12373`, `doi:10.1007/JHEP03(2021)257`.

[2] CMS Collaboration. Search for nonresonant Higgs boson pair production in the 4 leptons plus 2 b jets final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2020. URL: `https://cds.cern.ch/record/2725691`.

[3] ATLAS Collaboration. Search for the HH $\to b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Journal of High Energy Physics*, 2020(7):108, January 2021. `arXiv:2001.05178`, `doi:10.1007/JHEP07(2020)108`.

[4] ATLAS Collaboration. Search for non-resonant Higgs boson pair production in the $bbl\nu l\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 801:135145, February 2020. `arXiv:1908.06765`, `doi:10.1016/j.physletb.2019.135145`.

[5] CMS Collaboration. Search for Higgs boson pair production in the four b quark final state. Technical report, CERN, Geneva, 2021. URL: `https://cds.cern.ch/record/2771912`.

[6] ATLAS Collaboration. Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical report, CERN, Geneva, Mar 2021. All figures including auxiliary figures are available at

https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2021-016. URL: `https://cds.cern.ch/record/2759683`.

[7] CERN. CERN Documentation Server. URL: `https://home.cern/resources/faqs/facts-and-figures-about-lhc`.

[8] Bulk graviton cross sections in WED. `https://github.com/CrossSectionsLHC/WED/blob/master/KKGraviton_Bulk/Xjj_LO_14TeV_ktilda_0p1.txt`. A. Oliveira: June, 2016.

[9] Wikipedia. General description of the Standard Model in Wikipedia. URL: `https://en.wikipedia.org/wiki/Standard_Model`.

[10] John Ellis. Higgs Physics. *arXiv e-prints*, page arXiv:1312.5672, December 2013. `arXiv:1312.5672`.

[11] Suresh Emre. Experimental verification of Higgs boson couplings to elementary particles. URL: `https://sureshemre.wordpress.com/2020/09/12/experimental-verification-of-higgs-boson-couplings-to-elementary-particles/`.

[12] Particle Data Group. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. 083C01. `arXiv:https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf`, `doi:10.1093/ptep/ptaa104`.

[13] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *CERN-2011-002*, CERN, Geneva, 2011. `arXiv:1101.0593`, `doi:10.5170/CERN-2011-002`.

[14] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 2. Differential Distributions. *CERN-2012-002*, CERN, Geneva, 2012. `arXiv:1201.3084`, `doi:10.5170/CERN-2012-002`.

[15] LHC Higgs Cross Section Working Group, S. Heinemeyer, C. Mariotti, G. Passarino, and

R. Tanaka (Eds.). Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN-2013-004*, CERN, Geneva, 2013. `arXiv:1307.1347`, `doi:10.5170/CERN-2013-004`.

[16] Alexandra Carvalho, Martino Dall'Osso, Tommaso Dorigo, Florian Goertz, Carlo A. Gottardo, and Mia Tosi. Higgs Pair Production: Choosing Benchmarks With Cluster Analysis. *JHEP 04*, page 126, 2016. `arXiv:1507.02245`, `doi:10.1007/JHEP04(2016)126`.

[17] Ram Krishna Dewanjee. *Search for Neutral MSSM Higgs bosons and other exotic resonances in their decays involving 2 hadronically decaying τ leptons at CMS*. PhD thesis, CERN, Geneva, 2019. URL: `https://cds.cern.ch/record/2743311`.

[18] Jakob Calvén. Warped Extra Dimensions in the Randall-Sundrum Model and a Simple Implementation in Pythia 8. In *Conference Proceedings*, 2012.

[19] Alexandra Carvalho. Gravity particles from Warped Extra Dimensions, predictions for LHC. *arXiv e-prints*, page arXiv:1404.0102, March 2014. `arXiv:1404.0102`.

[20] LHC CERN. Supersymmetry:Taking a closer look at LHC. URL: `https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.supersymmetry`.

[21] Monika Blanke, David Curtin, and Maxim Perelstein. Supersymmetric Yukawa sum rule and LHC tests. *Phys. Rev. D*, 82, 08 2010. `doi:10.1103/PhysRevD.82.035020`.

[22] The CMS Collaboration. Combination of Searches for Higgs Boson Pair Production in Proton-Proton Collisions at $\sqrt{s}$ = 13 TeV. *Phys. Rev. Lett.*, 122:121803, Mar 2019. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.122.121803`, `doi:10.1103/PhysRevLett.122.121803`.

[23] The ATLAS Collaboration. Combination of searches for Higgs boson pairs in pp collisions at s=13TeV with the ATLAS detector. *Physics Letters B*, 800:135103, 2020. URL: `https://www.sciencedirect.com/science/article/pii/S0370269319308251`, `doi:https://doi.org/10.1016/j.physletb.2019.135103`.

[24] CMS Collaboration. Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton-proton collisions at sqrt(s) = 13 TeV. *arXiv e-prints*, page arXiv:1707.02909, July 2017. `arXiv:1707.02909`.

[25] CMS Collaboration. Search for the decay of a heavy Higgs boson H into two lighter Higgs bosons h and h_S in the $h(\tau\tau)h_S(bb)$ final state at 13 TeV. Technical report, CERN, Geneva, 2021. URL: https://cds.cern.ch/record/2758271.

[26] AC Team. Diagram of an LHC dipole magnet. Schéma d'un aimant dipôle du LHC. 1, Jun 1999. URL: https://cds.cern.ch/record/40524.

[27] CERN. Cern documentation server. URL: http://cdsweb.cern.ch/search?recid=43487.

[28] CERN. CMS Lumi TWiki. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots.

[29] Wikipedia. Large Hadron Collider. URL: https://en.wikipedia.org/wiki/Large_Hadron_Collider.

[30] CMS Collabortion. CMS workflow. URL: https://cms.cern/detector.

[31] Christopher J Seez. The CMS Trigger System. Technical report, CERN, Geneva, May 2003. URL: https://cds.cern.ch/record/687840, doi:10.1140/epjcd/s2004-04-016-8.

[32] V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al. The CMS trigger system. *Journal of Instrumentation*, 12(1):P01020, January 2017. arXiv:1609.02366, doi:10.1088/1748-0221/12/01/P01020.

[33] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, J. Erö, and et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(5):P05011, May 2018. arXiv:1712.07158, doi:10.1088/1748-0221/13/05/P05011.

[34] The CMS Collaboration. Performance of b tagging algorithms in proton-proton collisions at 13 TeV with Phase 1 CMS detector. Technical report, CERN, Jun 2018. URL: https://cds.cern.ch/record/2627468.

[35] CMS Collaboration. A deep neural network for simultaneous estimation of b jet energy and resolution. *arXiv e-prints*, page arXiv:1912.06046, December 2019. `arXiv:1912.06046`.

[36] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *arXiv e-prints*, page arXiv:2012.06888, December 2020. `arXiv:2012.06888`.

[37] CMS Collaboration. The Phase-2 Upgrade of the CMS Beam Radiation Instrumentation and Luminosity Detectors. Technical report, CERN, Geneva, Mar 2021. This is a dummy document, only to get the LHCC reference label. URL: `https://cds.cern.ch/record/2759074`.

[38] CMS Collaboration. The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger. Technical report, CERN, Geneva, Mar 2021. This is a temporary submission, just to get the LHCC reference number, to be used in this and other CMS documents. URL: `https://cds.cern.ch/record/2759072`.

[39] CMS Collaboration. The Phase-2 Upgrade of the CMS Level-1 Trigger. Technical report, CERN, Geneva, Apr 2020. Final version. URL: `https://cds.cern.ch/record/2714892`.

[40] CMS Collaboration. A MIP Timing Detector for the CMS Phase-2 Upgrade. Technical report, CERN, Geneva, Mar 2019. URL: `https://cds.cern.ch/record/2667167`.

[41] CMS Collaboration. The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical report, CERN, Geneva, Nov 2017. URL: `https://cds.cern.ch/record/2293646`.

[42] CMS Collaboration. The Phase-2 Upgrade of the CMS Muon Detectors. Technical report, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC. URL: `https://cds.cern.ch/record/2283189`.

[43] CMS Collaboration. The Phase-2 Upgrade of the CMS Barrel Calorimeters. Technical report, CERN, Geneva, Sep 2017. This is the final version, approved by the LHCC. URL: `https://cds.cern.ch/record/2283187`.

[44] CMS Collaboration. The Phase-2 Upgrade of the CMS Tracker. Technical report, CERN, Geneva, Jun 2017. URL: `https://cds.cern.ch/record/2272264`.

[45] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical design report. CMS. CERN, Geneva, 2006. URL: `http://cds.cern.ch/record/922757`.

[46] 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. URL: `https://www.sciencedirect.com/science/article/pii/S037026931200857X`, `doi:https://doi.org/10.1016/j.physletb.2012.08.020`.

[47] The 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. URL: `https://www.sciencedirect.com/science/article/pii/S0370269312008581`, `doi:https://doi.org/10.1016/j.physletb.2012.08.021`.

[48] Serguei Chatrchyan et al. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 06:081, 2013. `arXiv:1303.4571`, `doi:10.1007/JHEP06(2013)081`.

[49] Ilaria Brivio and Michael Trott. The Standard Model as an Effective Field Theory. *Physics Reports*, 793, June 2019. URL: `https://www.sciencedirect.com/science/article/pii/S0370157318303223`, `arXiv:1706.08945`, `doi:https://doi.org/10.1016/j.physrep.2018.11.002`.

[50] Steven W. Allen, August E. Evrard, and Adam B. Mantz. Cosmological Parameters from Observations of Galaxy Clusters. *Annual Review of Astronomy and Astrophysics*, 49(1):409–470, September 2011. `arXiv:1103.4829`, `doi:10.1146/annurev-astro-081710-102514`.

[51] Edvige Corbelli and Paolo Salucci. The extended rotation curve and the dark matter halo of M33. *Monthly Notices of the Royal Astronomical Society*, 311(2):441–447, January 2000. `arXiv:astro-ph/9909252`, `doi:10.1046/j.1365-8711.2000.03075.x`.

[52] Ze-Long Yi and Tong-Jie Zhang. Constraints on Holographic Dark Energy Models Using the Differential Ages of Passively Evolving Galaxies. *Modern Physics Letters A*, 22(1):41–53, January 2007. `arXiv:astro-ph/0605596`, `doi:10.1142/S0217732307020889`.

[53] L. Chen, R. de Grijs, and J. L. Zhao. Mass Segregation in Very Young Open Clusters: A Case Study of NGC 2244 and NGC 6530. *The Astronomical Journal*, 134(4):1368–1379, October 2007. `arXiv:0706.2723`, `doi:10.1086/521022`.

[54] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and antimatter in the universe. *New Journal of Physics*, 14(9):095012, September 2012. `arXiv:1204.4186`, `doi:10.1088/1367-2630/14/9/095012`.

[55] Fred Jegerlehner. The hierarchy problem of the electroweak Standard Model revisited. *arXiv e-prints*, page arXiv:1305.6652, May 2013. `arXiv:1305.6652`.

[56] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. Search for Neutrinos from the Sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.20.1205`, `doi:10.1103/PhysRevLett.20.1205`.

[57] J. Wess and B. Zumino. Supergauge transformations in four dimensions. *Nuclear Physics B*, 70(1):39–50, 1974. URL: `https://www.sciencedirect.com/science/article/pii/0550321374903551`, `doi:https://doi.org/10.1016/0550-3213(74)90355-1`.

[58] Lisa Randall and Raman Sundrum. Large Mass Hierarchy from a Small Extra Dimension. *prl*, 83(17):3370–3373, October 1999. `arXiv:hep-ph/9905221`, `doi:10.1103/PhysRevLett.83.3370`.

[59] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001–S08001, August 2008. `doi:10.1088/1748-0221/3/08/s08001`.

[60] Jesse Thaler and Ken van Tilburg. Maximizing boosted top identification by minimizing N-subjettiness. *Journal of High Energy Physics*, 2012:93, February 2012. `arXiv:1108.2701`, `doi:10.1007/JHEP02(2012)093`.

[61] CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(5):P05011, May 2018. `arXiv:1712.07158`, `doi:10.1088/1748-0221/13/05/P05011`.

[62] Alexandra Carvalho, Martino Dall'Osso, Pablo de Castro Manzano, Tommaso Dorigo, Florian Goertz, Maxime Gouzevich, and Mia Tosi. Analytical parametrization and shape

classification of anomalous HH production in the EFT approach. *arXiv e-prints*, page arXiv:1608.06578, July 2016. `arXiv:1608.06578`.

[63] Zhaoxia Heng, Xue Gong, and Haijing Zhou. Pair production of Higgs boson in NMSSM at the LHC with the next-to-lightest CP-even Higgs boson being SM-like. *Chinese Physics C*, 42(7):073103, June 2018. `arXiv:1805.01598`, `doi:10.1088/1674-1137/42/7/073103`.

[64] Albert M Sirunyan et al. Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 788:7, 2019. `arXiv:1806.00408`, `doi:10.1016/j.physletb.2018.10.056`.

[65] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961. `doi:10.1016/0029-5582(61)90469-2`.

[66] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970. URL: `https://link.aps.org/doi/10.1103/PhysRevD.2.1285`, `doi:10.1103/PhysRevD.2.1285`.

[67] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.19.1264`, `doi:10.1103/PhysRevLett.19.1264`.

[68] Sudhir K. Vempati. Introduction to MSSM. *arXiv e-prints*, page arXiv:1201.0334, January 2012. `arXiv:1201.0334`.

[69] Di Lella, Luigi and Rubbia, Carlo. The Discovery of the W and Z Particles. *Adv. Ser. Dir. High Energy Phys.*, 23:137–163, 2015. URL: `https://cds.cern.ch/record/2103277`, `doi:10.1142/9789814644150_0006`.

[70] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.321`, `doi:10.1103/PhysRevLett.13.321`.

[71] Peter W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145:1156–1163, May 1966. URL: `https://link.aps.org/doi/10.1103/PhysRev.145.1156`, `doi:10.1103/PhysRev.145.1156`.

[72] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.13.585`, `doi:10.1103/PhysRevLett.13.585`.

[73] The ATLAS and CMS Collaboration. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, May 2015. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.114.191803`, `doi:10.1103/PhysRevLett.114.191803`.

[74] The ATLAS Collaboration. Measurement of the Higgs boson mass in the HZZ4 and H channels with s=13 TeV pp collisions using the ATLAS detector. *Physics Letters B*, 784:345–366, 2018. URL: `https://www.sciencedirect.com/science/article/pii/S0370269318305884`, `doi:https://doi.org/10.1016/j.physletb.2018.07.050`.

[75] Sirunyan, Albert M and others. Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:047, 2017. `arXiv:1706.09936`, `doi:10.1007/JHEP11(2017)047`.

[76] CMS Collaboration. Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state. *Physical Review D*, 99(11):112003, June 2019. `arXiv:1901.00174`, `doi:10.1103/PhysRevD.99.112003`.

[77] CMS Collaboration. Study of the Mass and Spin-Parity of the Higgs Boson Candidate via Its Decays to Z Boson Pairs. *Physical Review Letters*, 110(8):081803, February 2013. `arXiv:1212.6639`, `doi:10.1103/PhysRevLett.110.081803`.

[78] CMS Collaboration. Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79:421, May 2019. `arXiv:1809.10733`, `doi:https://doi.org/10.1140/epjc/s10052-019-6909-y`.

[79] N. Jarosik and et al. Seven-year wilkinson microwave anisotropy probe (wmap) Observations: Sky Maps, systematic errors, and basic results. *The Astrophysical Journal Supplement Series*, 192(2):14, jan 2011. `doi:10.1088/0067-0049/192/2/14`.

[80] Virginia Trimble. Existence and Nature of Dark Matter in the Universe. *Annual Review of Astronomy and Astrophysics*, 25(1):425–472, 1987. `arXiv:https://doi.org/10.1146/annurev.aa.25.090187.002233`, `doi:10.1146/annurev.aa.25.090187.002233`.

[81] Andrei D. Sakharov. SPECIAL ISSUE: Violation of CP in variance, C asymmetry, and baryon asymmetry of the universe. *Soviet Physics Uspekhi*, 34(5):392–393, May 1991. `doi:10.1070/PU1991v034n05ABEH002497`.

[82] Kenneth G. Wilson. Renormalization Group and Strong Interactions. *Phys. Rev. D*, 3:1818–1846, Apr 1971. URL: `https://link.aps.org/doi/10.1103/PhysRevD.3.1818`, `doi:10.1103/PhysRevD.3.1818`.

[83] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. Search for Neutrinos from the Sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.20.1205`, `doi:10.1103/PhysRevLett.20.1205`.

[84] Y. Fukuda and et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998. URL: `https://link.aps.org/doi/10.1103/PhysRevLett.81.1562`, `doi:10.1103/PhysRevLett.81.1562`.

[85] M. C. Gonzalez-Garcia and Yosef Nir. Neutrino masses and mixing: evidence and implications. *Rev. Mod. Phys.*, 75:345–402, Mar 2003. URL: `https://link.aps.org/doi/10.1103/RevModPhys.75.345`, `doi:10.1103/RevModPhys.75.345`.

[86] E. Fermi. An attempt of a theory of beta radiation. 1. *Z. Phys.*, 88:161–177, 1934. `doi:10.1007/BF01351864`.

[87] Aneesh V. Manohar. Introduction to Effective Field Theories. *arXiv e-prints*, page arXiv:1804.05863, April 2018. `arXiv:1804.05863`.

[88] The ATLAS Collaboration. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. Technical Report ATLAS-CONF-2015-044, CERN, Geneva, Sep 2015. URL: `https://cds.cern.ch/record/2052552`.

[89] The CMS Collaboration. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC

pp collision data at $\sqrt{s} = 7$ and 8 TeV. Technical Report CMS-PAS-HIG-15-002. ATLAS-CONF-2015-044, CERN, Geneva, 2015. URL: `https://cds.cern.ch/record/2053103`.

[90] The LEP Collaborations: ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, and the LEP Electroweak Working Group. A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model. *arXiv e-prints*, pages hep–ex/0612034, December 2006. `arXiv:hep-ex/0612034`.

[91] Gerhard Buchalla. Nonlinear EFT. Technical report, CERN, Mar 2016. URL: `https://cds.cern.ch/record/2137956`.

[92] Csaba Csaki. TASI Lectures on Extra Dimensions and Branes. *arXiv e-prints*, pages hep–ph/0404096, April 2004. `arXiv:hep-ph/0404096`.

[93] Hsin-Chia Cheng. 2009 TASI Lecture – Introduction to Extra Dimensions. *arXiv e-prints*, page arXiv:1003.1162, March 2010. `arXiv:1003.1162`.

[94] Stephen P. Martin. *A Supersymmetry Primer*, pages 1–98. arXiv, 1998. `doi:10.1142/9789812839657_0001`.

[95] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev. D*, 98(3):030001, 2018. `doi:10.1103/PhysRevD.98.030001`.

[96] Ian J R Aitchison. Supersymmetry and the MSSM: An Elementary Introduction. *arXiv e-prints*, pages hep–ph/0505105, May 2005. `arXiv:hep-ph/0505105`.

[97] Jihn E. Kim and Hans Peter Nilles. Symmetry Principles Toward Solutions of the μ Problem. *Modern Physics Letters A*, 9(38):3575–3583, January 1994. `arXiv:hep-ph/9406296`, `doi:10.1142/S0217732394003415`.

[98] M. Grazzini, G. Heinrich, S. Jones, S. Kallweit, M. Kerner, J. M. Lindert, and J. Mazzitelli. Higgs boson pair production at NNLO with top quark mass effects. *Journal of High Energy Physics*, 2018(5):59, May 2018. `arXiv:1803.02463`, `doi:10.1007/JHEP05(2018)059`.

[99] M A Hone. The duoplasmatron ion source for the new CERN LINAC preinjector. Technical report, CERN, Geneva, Nov 1979. URL: `http://cds.cern.ch/record/2640736`.

[100] A M Lombardi. The radio frequency quadrupole (RFQ). *CERN Document Server*, 2006. URL: `https://cds.cern.ch/record/1005049`, `doi:10.5170/CERN-2006-012.201`.

[101] CERN. Linear accelerator 2. *CERN Document Server*, Sep 2012. URL: `https://cds.cern.ch/record/1997427`.

[102] Giorgio Brianti. The CERN synchrotrons. *CERN Document Server*, 1997. URL: `https://cds.cern.ch/record/340514`, `doi:10.5170/CERN-1997-004.29`.

[103] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008. `doi:10.1088/1748-0221/3/08/S08004`.

[104] V. Veszpremi. Operation and performance of the CMS tracker. *Journal of Instrumentation*, 9(3):C03005, March 2014. `arXiv:1402.0675`, `doi:10.1088/1748-0221/9/03/C03005`.

[105] N. Parashar. CMS Pixel Detector Upgrade. *arXiv e-prints*, page arXiv:1110.2125, October 2011. `arXiv:1110.2125`.

[106] D. J. A. Cockerill. The CMS Electromagnetic Calorimeter at the LHC. *arXiv e-prints*, page arXiv:0810.0381, October 2008. `arXiv:0810.0381`.

[107] CMS Collabortion. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. URL: `https://cds.cern.ch/record/357153`.

[108] CMS collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV. *Journal of Instrumentation*, 13(6):P06015, June 2018. `arXiv:1804.04528`, `doi:10.1088/1748-0221/13/06/P06015`.

[109] Katharina Danziger, Stefan Höche, and Frank Siegert. Reducing negative weights in Monte Carlo event generation with Sherpa. *arXiv e-prints*, October 2021. `arXiv:2110.15211`.

[110] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. `arXiv:hep-ph/0409146`, `doi:10.1088/1126-6708/2004/11/040`.

[111] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011:128, June 2011. `arXiv:1106.0522`, `doi:10.1007/JHEP06(2011)128`.

[112] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010:43, June 2010. `arXiv:1002.2581`, `doi:10.1007/JHEP06(2010)043`.

[113] C. Oleari. The Powheg Box. *Nuclear Physics B Proceedings Supplements*, 205:36–41, August 2010. `arXiv:1007.3893`, `doi:10.1016/j.nuclphysbps.2010.08.016`.

[114] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 06:043, 2010. `arXiv:1002.2581`, `doi:10.1007/JHEP06(2010)043`.

[115] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. `arXiv:0709.2092`, `doi:10.1088/1126-6708/2007/11/070`.

[116] 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. *Computer Physics Communications*, 191:159–177, June 2015. `arXiv:1410.3012`, `doi:10.1016/j.cpc.2015.01.024`.

[117] S. Agostinelli and et al. Geant4a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. URL: `https://www.sciencedirect.com/science/article/pii/S0168900203013688`, `doi:https://doi.org/10.1016/S0168-9002(03)01368-8`.

[118] CMS Collaboration. CMSSW TWiki. URL: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework`.

[119] The CMS Collaboration. Description and performance of track and primary-vertex

reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009, October 2014. `arXiv:1405.6569`, `doi:10.1088/1748-0221/9/10/P10009`.

[120] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 13(6):P06015, June 2018. `arXiv:1804.04528`, `doi:10.1088/1748-0221/13/06/P06015`.

[121] Gavin P. Salam. Towards jetography. *European Physical Journal C*, 67(3-4):637–686, June 2010. `arXiv:0906.1833`, `doi:10.1140/epjc/s10052-010-1314-6`.

[122] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *Journal of High Energy Physics*, 2008(4):063, April 2008. `arXiv:0802.1189`, `doi:10.1088/1126-6708/2008/04/063`.

[123] The CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(2):P02014, February 2017. `arXiv:1607.03663`, `doi:10.1088/1748-0221/12/02/P02014`.

[124] The CMS Collaboration. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(2):P02006, February 2015. `arXiv:1411.0511`, `doi:10.1088/1748-0221/10/02/P02006`.

[125] CMS Collaboration. Missing transverse energy performance of the CMS detector. *Journal of Instrumentation*, 6(9):9001, September 2011. `arXiv:1106.5048`, `doi:10.1088/1748-0221/6/09/P09001`.

[126] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014:59, October 2014. `arXiv:1407.6013`, `doi:10.1007/JHEP10(2014)059`.

[127] The CMS Collaboration. Identification of b-quark jets with the CMS experiment. *Journal of Instrumentation*, 8(4):P04013, April 2013. `arXiv:1211.4462`, `doi:10.1088/1748-0221/8/04/P04013`.

[128] 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.

Physical Review D, 94(11):112002, December 2016. `arXiv:1607.08633, doi:10.1103/
PhysRevD.94.112002`.

[129] CMS Collaboration. Observation of the diphoton decay of the Higgs boson and measurement of its properties. *The European Physical Journal C*, 74:1434–6052, July 2014. `arXiv:1407.0558, doi:10.1140/epjc/s10052-014-3076-z`.

[130] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet Substructure as a New Higgs-Search Channel at the Large Hadron Collider. *Physical Review Letters*, 100(24):242001, June 2008. `arXiv:0802.2470, doi:10.1103/
PhysRevLett.100.242001`.

[131] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *Journal of High Energy Physics*, 2013:29, September 2013. `arXiv:1307.0007, doi:10.1007/JHEP09(2013)029`.

[132] Jesse Thaler and Ken Van Tilburg. Identifying boosted objects with N-subjettiness. *Journal of High Energy Physics*, 2011:15, March 2011. `arXiv:1011.2268, doi:10.1007/
JHEP03(2011)015`.

[133] I Bjar Alonso, O Brning, P Fessia, L Rossi, L Tavian, and M Zerlauth. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2020. URL: `http://cds.cern.ch/record/2749422, doi:
10.23731/CYRM-2020-0010`.

[134] Walter D. Goldberger and Mark B. Wise. Modulus stabilization with bulk fields. *Phys. Rev. Lett.*, 83:4922, 1999. `arXiv:hep-ph/9907447, doi:10.1103/PhysRevLett.83.
4922`.

[135] Csaki, Csaba and Graesser, Michael and Randall, Lisa and Terning, John. Cosmology of brane models with radion stabilization. *Phys. Rev. D*, 62:045015, 2000. `arXiv:hep-ph/
9911406, doi:10.1103/PhysRevD.62.045015`.

[136] Csaki, Csaba and Graesser, Michael L. and Kribs, Graham D. Radion dynamics and electroweak physics. *Phys. Rev. D*, 63:065002, 2001. `arXiv:hep-th/0008151, doi:10.
1103/PhysRevD.63.065002`.

[137] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo. Phenomenology of the Randall-Sundrum Gauge Hierarchy Model. *Phys. Rev. Lett.*, 84:2080, 2000. `arXiv:hep-ph/9909255`, `doi:10.1103/PhysRevLett.84.2080`.

[138] O. DeWolfe, D. Z. Freedman, S. S. Gubser, and A. Karch. Modeling the fifth dimension with scalars and gravity. *Phys. Rev. D*, 62:046008, 2000. `arXiv:hep-th/9909134`, `doi:10.1103/PhysRevD.62.046008`.

[139] Kaustubh Agashe, Hooman Davoudiasl, Gilad Perez, and Amarjit Soni. Warped gravitons at the LHC and beyond. *Phys. Rev. D*, 76:036006, 2007. `arXiv:hep-ph/0701186`, `doi:10.1103/PhysRevD.76.036006`.

[140] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:01, 2012. `arXiv:1106.0034`, `doi:10.1016/j.physrep.2012.02.002`.

[141] Abdelhak Djouadi. The anatomy of electroweak symmetry breaking. Tome II: the Higgs bosons in the minimal supersymmetric model. *Phys. Rept.*, 459:01, 2008. `arXiv:hep-ph/0503173`, `doi:10.1016/j.physrep.2007.10.005`.

[142] Howard Georgi and Marie Machacek. Doubly charged Higgs bosons. *Nuclear Physics B*, 262(3):463, 1985. `doi:=10.1016/0550-3213(85)90325-6`.

[143] Georges Aad et al. Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector. *Phys. Rev. Lett.*, 114:081802, 2015. `arXiv:1406.5053`, `doi:10.1103/PhysRevLett.114.081802`.

[144] Georges Aad et al. Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 75:412, 2015. `arXiv:1506.00285`, `doi:10.1140/epjc/s10052-015-3628-x`.

[145] Georges Aad et al. Searches for Higgs boson pair production in the $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector. *Phys. Rev. D*, 92:092004, 2015. `arXiv:1509.04670`, `doi:10.1103/PhysRevD.92.092004`.

[146] Vardan Khachatryan et al. Searches for heavy Higgs bosons in two-Higgs-doublet models

and for $t \to ch$ decay using multilepton and diphoton final states in pp collisions at 8 TeV. *Phys. Rev. D*, 90:112013, 2014. `arXiv:1410.2751, doi:10.1103/PhysRevD.90.112013`.

[147] Vardan Khachatryan et al. Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 8 TeV. *Phys. Lett. B*, 749:560, 2015. `arXiv:1503.04114, doi:10.1016/j.physletb.2015.08.047`.

[148] Vardan Khachatryan et al. Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with h $\to \tau\tau$. *Phys. Lett. B*, 755:217, 2016. `arXiv:1510.01181, doi: 10.1016/j.physletb.2016.01.056`.

[149] Vardan Khachatryan et al. Search for two Higgs bosons in final states containing two photons and two bottom quarks in proton-proton collisions at 8 TeV. *Phys. Rev. D*, 94:052012, 2016. `arXiv:1603.06896, doi:10.1103/PhysRevD.94.052012`.

[150] Vardan Khachatryan et al. Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks. *Eur. Phys. J. C*, 76:371, 2016. `arXiv:1602.08762, doi:10.1140/epjc/s10052-016-4206-6`.

[151] Morad Aaboud et al. Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 94:052002, 2016. `arXiv:1606.04782, doi:10.1103/PhysRevD.94.052002`.

[152] ATLAS Collaboration. Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton$-$proton collisions at $\sqrt{s} = 13 \ TeV$ with the ATLAS detector. Technical Report ATLAS-CONF-2016-049, CERN, Geneva, Aug 2016. URL: `https://cds.cern.ch/record/2206131`.

[153] CMS Collaboration. Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 781:244, 2018. `arXiv:1710.04960, doi:10.1016/j.physletb.2018.03.084`.

[154] Gian F. Giudice, Riccardo Rattazzi, and James D. Wells. Graviscalars from higher dimensional metrics and curvature Higgs mixing. *Nucl. Phys. B*, 595:250, 2001. `arXiv:hep-ph/0002178, doi:10.1016/S0550-3213(00)00686-6`.

[155] Vieri Candelise, Ching Wei Chen, Alice Cocoros, Leonardo Giannini, Maxime Gouzevitch, Raman Khurana, Michael Krohn, Gregorio de Leon III, Devdatta Majumder, Petar Maksimovic, Marc Osherson, Andrea Rizzi, Caterina Vernieri, Steve Wagner, Shin-Shan Eiko Yu. Search for heavy resonances decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV. CMS Analysis Note CMS-AN-16-300, CERN, 2015.

[156] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014:57, February 2014. `arXiv:1307.6346`, `doi:10.1007/JHEP02(2014)057`.

[157] `https://github.com/acarvalh/HH_reweighting_cpp`.

[158] John C. Collins and Davison E. Soper. Angular distribution of dileptons in high-energy hadron collisions. *Phys. Rev. D*, 16:2219–2225, Oct 1977. URL: `https://link.aps.org/doi/10.1103/PhysRevD.16.2219`, `doi:10.1103/PhysRevD.16.2219`.

[159] ATLAS Collaboration. Search for Higgs boson pair production in the $\gamma\gamma b\overline{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment. *Journal of High Energy Physics*, 2018(11):40, November 2018. `arXiv:1807.04873`, `doi:10.1007/JHEP11(2018)040`.

[160] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 1407:079, 2014. `arXiv:1405.0301`, `doi:10.1007/JHEP07(2014)079`.

[161] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *Journal of High Energy Physics*, 2009(2):007, February 2009. `arXiv:0811.4622`, `doi:10.1088/1126-6708/2009/02/007`.

[162] D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, R. Tanaka, R. Aggleton, and et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. *arXiv e-prints*, page arXiv:1610.07922, October 2016. `arXiv:1610.07922`.

[163] Vardan Khachatryan et al. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016. `arXiv:1512.00815`, `doi:10.1140/epjc/s10052-016-3988-x`.

[164] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J.*, 2019. `arXiv:1903.12179`.

[165] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. `arXiv:1410.8849`, `doi:10.1007/JHEP04(2015)040`.

[166] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77:663, 2017. `arXiv:1706.00428`, `doi:10.1140/epjc/s10052-017-5199-5`.

[167] Stefano Carrazza, Jose I. Latorre, Juan Rojo, and Graeme Watt. A compression algorithm for the combination of PDF sets. *Eur. Phys. J. C*, 75:474, 2015. `arXiv:1504.06469`, `doi:10.1140/epjc/s10052-015-3703-3`.

[168] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016. `arXiv:1510.03865`, `doi:10.1088/0954-3899/43/2/023001`.

[169] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Marco Guzzi, Joey Huston, Pavel Nadolsky, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. New parton distribution functions from a global analysis of quantum chromodynamics. *Phys. Rev. D*, 93:033006, 2016. `arXiv:1506.07443`, `doi:10.1103/PhysRevD.93.033006`.

[170] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75:204, 2015. `arXiv:1412.3989`, `doi:10.1140/epjc/s10052-015-3397-6`.

[171] Albert M Sirunyan et al. Pileup mitigation at CMS in 13 TeV data. *JINST*, 15:P09018, 2020. `arXiv:2003.00503`, `doi:10.1088/1748-0221/15/09/P09018`.

[172] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, page 785, New York, NY, USA, 2016. ACM. `doi:10.1145/2939672.2939785`.

[173] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[174] Maxime Gouzevitch, Alexandra Oliveira, Juan Rojo, Rogerio Rosenfeld, Gavin P. Salam, and Veronica Sanz. Scale-invariant resonance tagging in multijet events and new physics in Higgs pair production. *Journal of High Energy Physics*, 2013:148, July 2013. `arXiv: 1303.3636`, `doi:10.1007/JHEP07(2013)148`.

[175] scikit learn. scikit-learn page. URL: `https://scikit-learn.org/stable/modules/ generated/sklearn.model_selection.RandomizedSearchCV.html`.

[176] XGBoost. XGBoost Documentation Page. URL: `http://xgboost.readthedocs.io/ en/latest/python/python_api.html`.

[177] Giovanni Punzi. Sensitivity of Searches for New Signals and Its Optimization. In L. Lyons, R. Mount, and R. Reitmeyer, editors, *Statistical Problems in Particle Physics, Astrophysics, and Cosmology*, page 79, January 2003. `arXiv:physics/0308063`.

[178] ROOT::MINUIT. Minuit Documentation. URL: `https://root.cern.ch/download/ minuit.pdf`.

[179] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes: the discrete profiling method. *Journal of Instrumentation*, 10(4):P04015, April 2015. `arXiv:1408.6865`, `doi:10.1088/1748-0221/10/04/P04015`.

[180] R. A. Fisher. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of p. *J. Royal Stat. Soc*, 85:87, 1922. `doi:10.2307/2340521`.

[181] Mark Joseph Oreglia. *A study of the reactions $\psi' \to \gamma\gamma\psi$*. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236. URL: `http://www.slac.stanford.edu/cgi-wrap/ getdoc/slac-r-236.pdf`.

[182] CERN. Combine Documentation. URL: `https://cms-analysis.github.io/ HiggsAnalysis-CombinedLimit/`.

[183] Albert M Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018. `arXiv:1712.07158, doi: 10.1088/1748-0221/13/05/P05011`.

[184] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. *CERN Document Server*, 2016. `arXiv:1610.07922`.

[185] Albert M Sirunyan et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81:800, Apr 2021. `arXiv: 2104.01927, doi:10.1140/epjc/s10052-021-09538-2`.

[186] CMS Collaboration. CMS luminosity measurements for the 2016 data taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, CERN, 2017. URL: `https://cds.cern.ch/record/2257069`.

[187] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-18-002, CERN, 2019. URL: `https://cds.cern.ch/record/2676164`.

[188] Serguei Chatrchyan et al. Measurement of the inclusive w and z production cross sections in pp collisions at $\sqrt{s} = 7$ tev. *JHEP*, 10:132, 2011. `arXiv:1107.4789, doi:10.1007/ JHEP10(2011)132`.

[189] Chatrchyan, Serguei and others. Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10:P08010, 2015. `arXiv:1502.02702, doi:10.1088/1748-0221/10/08/P08010`.

[190] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt. Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios. *European Physical Journal C*, 80(2):151, February 2020. `arXiv:1908.08554, doi:10.1140/epjc/s10052-020-7655-x`.