Rare Higgs Processes at CMS and Precision Timing Detector Studies for HL-LHC CMS Upgrade

Thesis by
Irene Dutta

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended July 18, 2022

Irene Dutta
ORCID: 0000-0003-0953-4503

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Maria Spiropulu, without whom this thesis wouldn't have been possible. She is a brilliant scientist and innovator. She has always had my back, in both professional and personal situations, and her immensely supportive nature has been the reason for my overall growth as a physicist over the last 5 years.

The initial learning phase of my Ph.D. was nothing short of a vertical slope, and I couldn't have done it without the help and encouragement of Dr. Nan Lu. Nan has been both a mentor and elder sister to me. She is an amazing physicist and has out of the box solutions to every problem. She has taught me how to think, ask questions, and work as a scientist. Thank you Nan for being so patient with me, and I hope to see both you and your baby girl sometime soon at CERN!

I also have to thank Dr. Si Xie, who has been a very inspiring mentor to me, and I have learned a lot by having physics discussions with him. I'm always in awe of his multi-tasking capabilities.

I would also like to thank Dr. Adi Bornheim, Prof. Andrea Rizzi, Dr. Anthony (Tony) LaTorre, Dr. Cristian Pena, Dr. Dustin Anderson, Prof. Harvey Newman, Dr. Javier Duarte, Dr. Jean-Roch Vlimant, Dr. Jennifer Ngadiuba, Dr. Joosep Pata, Lautaro Narvaez, and Dr. Maurizio Pierini, who have all shared their wisdom and helped me during various stages of my PhD journey. A big thank you to Justas Balcas and Andres Moya, who have patiently and promptly solved all the T2 issues I have faced over the years.

My office mates Christina Wang, Jiajing Mao, Olmo Cerri, Dr. Thong Nguyen and Dr. Zhicai Zhang have turned the long work hours more bearable; I'll miss all the office banter and barbeque parties with you guys.

Prof. Markus Klute has been mentoring me through the US-CMS mentorship program for the last 2 years, and I have really enjoyed talking to him about both physics and life. I wish to meet you soon in person, Markus!

To my Caltech friends, Anchal, Annika, Anushri, Cai, Jag, Lena, Nachiket, Prithvi, and Widi: I will miss all the fun parties, cook outs, game nights, restaurant/bar hopping and wonderful trips that I have taken with you guys. I hope our paths keep crossing in the future.

To my IISER Pune friends, Anirban, Deeksha, Kunal, Prachi, Prashali, Shipra, Vaibhav, and Vished: our long video calls, fun trips, and unlimited gossip has helped me get through so much in the past 2-3 years, especially the bleak COVID isolation days. We have grown a lot closer after leaving IISER, and I regret not hanging out enough with you guys during our undergraduate days.

To the two people who have been my friends for the longest time (~17 years): Divya and Shreetoma, thank you for always being there for me.

To Shubham, we went from being mere acquaintances and grew into the definition of *partners*. You make my life peaceful, balanced, and fun. You already know all the reasons for which I'm thankful to you.

And finally to my parents and brother: I could not have been here without your constant love and support. I'm lucky to have such an understanding family. I don't say this enough, but I love you guys.

ABSTRACT

This thesis describes the search for two rare Higgs processes. The first analysis describes the CMS Run 2 search for $H \to \mu\mu$ decays, with 137.3 fb^{-1} of data at $\sqrt{s} = 13$ TeV. The analysis targeted four different Higgs production modes: the gluon fusion (ggH), the vector boson fusion (VBF), the Higgs-strahlung process (VH), and the production in association with a pair of top quarks (ttH). Each category used a dedicated machine learning based classifier to separate the signal from the background processes. A combined fit from all these categories saw a slight excess in the data corresponding to 3.0 standard deviations at $M_H = 125.38$ GeV, and gave the first evidence for the Higgs boson decay to second-generation fermions. The best-fit signal strength and the corresponding 68% CL interval was found to be $\hat{\mu} = 1.19\ ^{+0.41}_{-0.39}\,(\text{stat})^{+0.17}_{-0.16}\,(\text{syst})$ at $M_H = 125.38$ GeV.

The second analysis describes the CMS Run 2 search for $HH \to b\bar{b}b\bar{b}$ with highly boosted Higgs bosons. This analysis used a dedicated jet identification algorithm based on graph neural networks (ParticleNet) to identify boosted H→ bb jets. This search targeted the gluon fusion and the vector boson fusion HH production modes, and put constraints on the allowed values of the various Higgs couplings as: $\kappa_\lambda \in [-9.9, 16.9]$ when $\kappa_V = 1$, $\kappa_{2V} = 1$; $\kappa_V \in [-1.17, -0.79] \cup [0.81, 1.18]$ when $\kappa_\lambda = 1$, $\kappa_{2V} = 1$; $\kappa_{2V} \in [0.62, 1.41]$ when $\kappa_\lambda = 1$, $\kappa_V = 1$. A scenario with $\kappa_{2V} = 0$ was excluded with a significance of 6.3 standard deviations for the first time, when other H couplings are fixed to their SM values. The combined observed (expected) 95% upper limit on the HH production cross section was found to be 9.9 (5.1) × SM.

Finally, this thesis also discusses the planned MIP Timing Detector (MTD) upgrade for CMS at the HL-LHC. The MTD will be a time-of-flight (TOF) detector, designed to provide a precision timing information for charged particles using SiPMs + LYSO scintillating crystals, with a time resolution of ~30 ps. This thesis describes several R&D tests that have been performed for characterizing the sensor properties (time resolution, light yield, etc.) and optimizing the sensor design geometry. This thesis also contains a description of mock test setups for cooling the sensors, since it is known to be an effective way of mitigating the increased dark current rates in the sensors due to radiation damage.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks. Technical report, CERN, Geneva, May 2022. URL http://cds.cern.ch/record/2809466. Accepted by Physical Review Letters.
I.D. was the lead student analyzer of the gluon fusion (ggF) category in this search, and worked on the background estimation, derived the data-driven corrections to signal and background simulations, studied the systematic uncertainties, and performed the final statistical analysis for the ggF category.

[2] A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, Jul 2022. doi: 10.1038/s41586-022-04892-x. URL https://www.nature.com/articles/s41586-022-04892-x.
I.D. worked on two separate analyses [1, 5] used in this Higgs combination paper from CMS.

[3] The CMS MTD collaboration. Test beam characterization of sensor prototypes for the CMS barrel MIP timing detector. *Journal of Instrumentation*, 16(07): P07023, Jul 2021. doi: 10.1088/1748-0221/16/07/p07023. URL https://doi.org/10.1088/1748-0221/16/07/p07023.
I.D. helped develop the test stands (hardware and data acquisition system) at the Caltech Lab, installed them at the Fermilab Test Beam Facility, and took part in data-taking shifts. I.D. also helped with the preliminary analysis of the test beam data.

[4] Irene Dutta, Joosep Pata, Nan Lu, Jean-roch Vlimant, Harvey Newman, Maria Spiropulu, Christina Reissel, and Daniele Ruini. Data analysis with GPU-accelerated kernels. In *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)*, volume 390, page 908, 2021. doi: 10.22323/1.390.0908.
I.D. helped develop and validate the library on the CERN Open Data and also used it to perform the CMS Run 2 $H \to \mu\mu$ analysis.

[5] Albert M Sirunyan et al. Evidence for Higgs boson decay to a pair of muons. *JHEP*, 2101:148. 68 p, Sep 2020. URL https://cds.cern.ch/record/2730058.
I.D. demonstrated the enhanced signal sensitivity of a deep neural network (DNN) for the vector boson fusion (VBF) category over a boosted decision tree (BDT); I.D. helped develop and validate the VBF DNN in control regions with data and simulation comparisons, studied the systematic uncertainties, and performed the statistical analysis and sanity checks to deliver the final sensitivity of the VBF channel.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

For centuries, humanity has pondered over the origins of the universe and its components: *How did the universe form? What are its fundamental constituents? What are the physical laws that govern the universe?* The concept that matter is made up of more elementary particles originated sometime around the 6^{th} century BC [1]. However, those ideas were based more on philosophical reasoning rather than any scientific experimentation. There was a change in temperament around the 1550s, and scientific theories began to be proposed with rigorous testing and evidence. In the 1600-1700s, Kepler's work on planetary motion laid the foundation for Isaac Newton to propose the laws of classical motion of objects and the first understanding of *gravity*, which is a fundamental force in our universe. In the late 1700–1800s, Thomas Young proposed the wave theory of light, Michael Faraday understood that electricity and magnetism are related, and James Clerk Maxwell laid down the fundamental theory of electromagnetism, that could describe light wave propagation in vacuum. In 1898, J.J. Thompson discovered the first known elementary particle, an electron, and put forth a "plum-pudding" model that described an atom as a positive sphere with electrons embedded in it. This was essentially the beginning of a *domino-effect* of new discoveries and advances in knowledge in the 20th century, in a branch of physics that is now known as *particle-physics*.

By 1919, Ernest Rutherford had discovered the atomic nucleus and inferred the existence of protons [2]. Theoretical work done by De Broglie, Schroedinger, Born, Pauli, Dirac, and Heisenberg in the 1920s laid down a framework for quantum mechanics at the atomic level. In 1930, J. Cockroft and E. Walton built the world's first particle accelerator [3]. The 1930's witnessed the discovery of the neutron, positron, and muon (although it was originally assumed to be the pion) [4–6]. Most of my grandparents were born around this time. The 1950's brought around a *zoo* of new particles, although it was still unclear if there was any intrinsic pattern to these particles. By the late 1950s–early 1960s, the works of Schwinger, Bludman, and Glashow suggested the existence of a $weak$ force mediated by heavy charged bosons ($W^{\pm}$) [7, 8]. In 1961, Gell-Mann and Ne'eman proposed the "eight-fold way" to classify hadrons, which became the $SU(3)$ group describing the strong-force in nature and predicted the existence of quarks (which were discovered at SLAC shortly thereafter) [9–11]. All this was happening around the same time my parents were born. Finally, there was a way to classify the plethora of particles that were being discovered!

And thus began the modern theory of elementary particles, now known as the *Standard Model*. In 1967, Steven Weinberg and Abdus Salam proposed the unification of the weak and electromagnetic forces, and predicted the existence of the Z boson [12, 13]. Around the same time, the existence of a Higgs-like particle was proposed by Peter Higgs [14], to explain the electro-weak symmetry breaking and assign particle masses. By the 1990s, all the predicted SM particles were discovered, except the Higgs boson. I was barely a year old when the top quark was discovered (1995) [15, 16].

In 2010, the Large Hadron Collider (LHC) [17], the most powerful particle accelerator in the world, started operations at CERN, with a goal of hunting down the Higgs boson. On 4th July 2012, the ATLAS and CMS experiments from CERN announced that they had independently discovered the Higgs boson [18, 19], thus resolving the last missing piece of the SM. When this announcement happened, I had just graduated from high school, and I was trying to decide where to pursue my undergraduate studies. I saw "God-particle" printed in bold on the front page of local newspapers, and that was the first time I heard about CERN and the LHC. I remember googling CERN, reading the novel "Angels and Demons" by Dan Brown, and I was simply fascinated. I had always liked physics in high school, but I think it was in that moment that I knew that I wanted to major in physics. It has been 10 years since the Higgs discovery, and now I am at the final milestone of my PhD journey.

Even though we discovered the Higgs, we have not measured and validated all of its properties; for example, the first and second generation Yukawa couplings and the Higgs self-coupling are yet to be observed. Additionally, we know that the SM is not a complete theory of our universe and fails to explain gravity, dark matter, neutrino masses, the hierarchy problem, etc. There are several proposed Beyond Standard Model (BSM) theories such as SUSY [20], Two-Higgs-doublet models [21], extra dimensions [22], sterile neutrinos [23], etc., that predict new fundamental particles which could potentially solve one or more of these issues, and/or alter the predictions of the SM. Performing precision tests of the SM and exploring uncharted BSM phase spaces is the future of particle physics, at least for the next several decades.

This thesis is focused on measuring the properties of the Higgs boson, in particular certain rare Higgs processes that are yet to be observed in nature. A fraction of this thesis is also dedicated to hardware R&D for a time-of-flight detector for the upgraded LHC, also known as the High Luminosity LHC.

Chapter 2 will begin with a description of the Standard Model of particle physics and describe the limitations of the theory. It will also discuss the various Higgs production mechanisms and decay channels at the LHC, and give an overview of recent Higgs measurements from the CMS Experiment.

Chapter 3 describes the LHC machine and the CMS Experiment, including a description of the various sub-detectors inside CMS, particle identification and reconstruction techniques, the real-time triggering system, and the global data processing and storage methods employed by CMS. A related note can be found in Appendix A, which will discuss the feasibility of a muon MIP based calibration technique for the CMS HCAL Endcaps during Run 3 data-taking period.

Chapter 4 will discuss the High Luminosity LHC project, and the planned MIP Timing Detector (MTD) upgrade for CMS for the HL-LHC operations. The HL-LHC is an upgraded version of the LHC, with a higher instantaneous luminosity (by a factor of ~3-4), and plans to collect 20 times the data collected by LHC during Run 2. The increased levels of radiation means the existing sub-detectors of CMS will need to be upgraded to maintain the current LHC particle reconstruction efficiency. The MTD will be a time-of-flight (TOF) detector, designed to provide precision timing information for charged particles, with a resolution of ~30 ps. The MTD is divided into a barrel and a endcap timing layer. Chapter 4 will mainly focus on the Barrel Timing Layer (BTL), and the R&D tests that have been performed on the BTL sensors, including various sensor characterization studies (time resolution, light yield, etc.) and optimization of the BTL design geometry. The BTL sensors will be exposed to large amounts of radiation over the 10 year HL-LHC operating period, resulting in a gradual degradation of their time resolution capabilities. Cooling the sensors to sub-zero temperatures is known to be an effective way of mitigating the increased dark current rates in the sensors due to radiation damage. The chapter will include a discussion on the ongoing efforts to understand the thermal properties of the BTL sensors, and minimize the thermal gradients between the cooling system and the sensors. Chapter 4 also has a related appendix note, Appendix B, that describes the commissioning tests performed on the various versions of the BTL prototype readout electronics.

Chapters 5 and 6 describe the search for two separate rare Higgs processes at CMS. Chapter 5 describes the search for $H \to \mu\mu$ decays performed using the full CMS Run-2 dataset, at a center of mass energy of 13 TeV, and corresponding to a total integrated luminosity of 137.3 fb^{-1} [24]. This search resulted in the first evidence for a Higgs boson coupling to muons: a 3σ excess was observed in the invariant mass spectrum of the two muons at 125.38 GeV. This is also the first indication of the existence of second generation Yukawa couplings, that are predicted by the SM. $H \to \mu\mu$ is an incredibly rare process – only 1 in every 5000 Higgs bosons decay to a pair of muons, and the first evidence results were originally not expected until the end of LHC Run 3 data taking (2022-2025). The Run 2 CMS $H \to \mu\mu$ search used four separate analysis categories, each targeting a different Higgs production mode. Each of the four categories used sophisticated machine learning techniques like boosted decision trees and deep neural networks, for enhanced signal sensitivity. The improved analysis techniques accelerated the achievement of the 3σ excess in $H \to \mu\mu$. Chapter 5 also briefly discusses the projections for the $H \to \mu\mu$ search at the HL-LHC, and has two related appendices, Appendix C and Appendix D. Appendix C describes an alternative, mass agnostic machine learning based strategy for one of the four search categories, namely the VBF category. Appendix D describes a fast, kernel based library, known as "hepaccelerate" [25], designed for analyzing billions of events from a typical collider experiment with a small turnaround time. The *hepaccelerate* library relies on a parallel computing architecture and can process data on GPUs or CPUs. This library was used for the data analysis in the CMS Run 2 $H \to \mu\mu$ search, and has also been validated on CERN Open Data.

Chapter 6 switches gears and moves from single Higgs production to double Higgs production (HH). The HH process is another rare but extremely important phenomenon, which will allow us to measure the Higgs self-interaction strength and probe the structure of the Higgs potential. During Run 2, it is estimated that the LHC generated about 7.5 million single Higgs bosons, but only produced about 4500 Higgs boson pairs, which means that the single Higgs production is a 1000 times stronger than the HH process. Chapter 6 will cover the production of $HH \to b\overline{b}b\overline{b}$ with highly boosted Higgs bosons with 138 fb^{-1} of Run 2 CMS data collected at a center of mass energy of 13 TeV [26]. The four-bottom-quark final state has the largest branching ratio (33.9%) amongst all HH decays, but was considered inaccessible until a few years ago, as it is dominated by large backgrounds (QCD and top quark production) and a poor decay channel resolution. To enhance the signal sensitivity in the 4b final state, a dedicated jet identification algorithm based on graph neural networks was developed to identify boosted H→ bb jets. The $HH \rightarrow b\overline{b}b\overline{b}$ search is also very sensitive to anomalous quartic VVHH couplings (c_{2V}) and, for the first time, a scenario with κ_{2V} (c_{2V}/c_{2V}^{SM}) = 0 is excluded at > 5σ, when other Higgs couplings are at their SM values. Chapter 6 will cover the analysis methods used in this search and will discuss the upper limits placed on the inclusive HH production cross section and the final sensitivity to different HH couplings (κ_{λ}, κ_{2V} and κ_{V}). The chapter will conclude with an overview of the current status of HH production at CMS, and projected results for HL-LHC.

Finally, Chapter 7 summarizes the results from all the chapters and has a discussion on the possible future directions for new Higgs measurements at the LHC and future colliders. I sincerely hope that this thesis has made a tiny dent (even if it is one in a trillionth) in expanding the current understanding of the universe that we live in.

Chapter 2

THE STANDARD MODEL OF PARTICLE PHYSICS

2.1 Introduction

The Standard Model (SM) of particle physics is a renormalizable field theory describing the known elementary particles and their interactions. It is described by the $U(1)_Y \times SU(2)_L \times SU(3)_c$ gauge symmetry group, where the subscripts indicate the conserved quantum numbers : Y stands for hypercharge, L stands for left-handedness, and c stands for color charge. It describes three out of the four fundamental forces in nature: the strong interaction ($SU(3)$), electromagnetism ($U(1)$) and the weak force ($SU(2)$). The SM does not explain the fourth fundamental force, gravity. There are two classes of elementary particles in the SM : (1) fermions with spin 1/2, and (2) bosons with integer spins.

The fermions can be further divided into two classes: leptons and quarks, and they constitute the basic building blocks of matter. There are six quarks: up (u), down (d), strange (s), charm (c), bottom (b,) and top (t), and six leptons: electron (e), electron neutrino (ν_e), muon (μ), muon neutrino (ν_μ), tau (τ), and tau neutrino (ν_τ). Fermions obey Fermi-Dirac statistics, and exhibit Pauli's Exclusion Principle. Pairs of fermions that exhibit similar properties are grouped together, forming a generation (or flavor), and thus there are three generations of both leptons and quarks, as can be seen in Fig. 2.1. Quarks carry an electric charge: 2/3 e for the u, c, and t quarks and -1/3 e for d, s, and b quarks, where e is the magnitude of the electron's electric charge. Quarks also carry color charge. There are three types of color charges in nature: red, blue and green. The color confinement principle requires free particles in nature to have a net zero color charge, which can be achieved with all three colors mixed together, or any one of the three colors and its complement. This results in two or more quarks binding together into composite particles known as hadrons. Leptons in the SM can only interact via electromagnetism (except neutrinos which have zero electric charge) and the weak force. Each fermion has its own anti-particle, which has the same mass, but opposite quantum numbers.

The gauge bosons in the SM act as force carriers and are spin 1 particles. The photon (γ) mediates the electromagnetic force between electrically charged particles and is massless. The $W^{\pm}$ and Z bosons mediate the weak force and are massive. There are eight different types of gluons in nature that are massless and mediate the strong force.

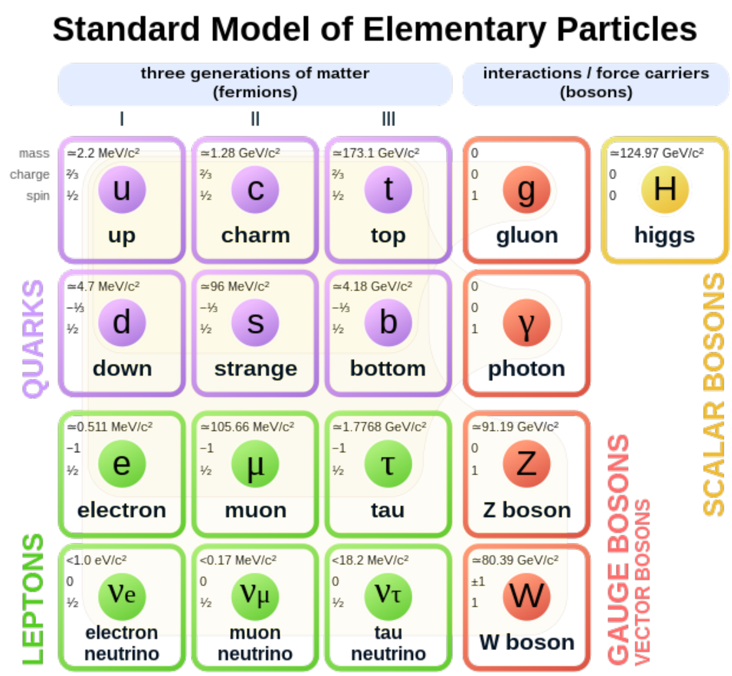


Figure 2.1: The elementary particles in the SM [27].

Finally, the Higgs boson (H) is the only massive spin-0 scalar boson observed in nature. Under the gauge invariant $U(1)_Y \times SU(2)_L \times SU(3)_c$ theory, all particles must be massless in nature, which is inconsistent with observations from various experiments. To solve this problem, the Higgs mechanism was proposed by Peter Higgs in 1964 [28], that predicted the existence of a neutral scalar boson that was responsible for breaking the electro-weak symmetry and giving rise to particle masses. The Higgs boson was later discovered by both the ATLAS and CMS experiments at the LHC, with a mass of about 125 GeV [18, 19].

2.2 The Lagrangian of the Standard Model

The Lagrangian of the SM can be broken down into four different components as

$$L_{\text{SM}} = L_{\text{fermion}} + L_{\text{gauge}} + L_{\text{Higgs}} + L_{\text{Yukawa}} \tag{2.1}$$

L_{fermion} describes the fermion kinematic term, and can be expressed as

$$L_{\text{fermion}} = \bar{\psi}^i (i\gamma^\mu \mathcal{D}_\mu)_{ij} \psi^j \tag{2.2}$$

where ψ is the fermionic field (a Dirac spinor), γ^μ ($\mu = 0, 1, 2, 3$) are the Dirac matrices, $\bar{\psi}$ is $\psi^+\gamma^0$ (ψ^+ is the h.c. of ψ) , and $\mathcal{D}_\mu$ is the covariant derivative (Eqn. 2.9).

In the SM, the fermions form left-handed weak-isospin doublets and right-handed weak-isospin singlets. The weak isospin is denoted by another quantum number, T_3, which is $\pm 1/2$ for left-handed doublets and zero for right-handed singlets. As an example, the first generation of leptons can be written as (no right handed neutrinos observed in nature):

$$L_e = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \tag{2.3}$$

$$R_e = \begin{pmatrix} e_R \end{pmatrix}. \tag{2.4}$$

Here, the left and right handed states are defined by :

$$\nu_L = \frac{1-\gamma^5}{2}\nu_e \tag{2.5}$$

$$e_L = \frac{1-\gamma^5}{2}e \tag{2.6}$$

$$e_R = \frac{1+\gamma^5}{2}e \tag{2.7}$$

where e and ν_e represent the electron and the electron neutrino, and $\gamma^5 = \gamma^0\gamma^1\gamma^2\gamma^3$. Similarly, the first generation of quarks can be written as:

$$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} u_R \end{pmatrix}, \begin{pmatrix} d_R \end{pmatrix}. \tag{2.8}$$

This definition can be extended to second and third generation of leptons and quarks. To have an invariant theory under gauge transformations, the covariant derivative, $\mathcal{D}_\mu$, is introduced as :

$$\mathcal{D}_\mu = \delta_\mu - ig_1 B_\mu Y - ig_2 W^a_\mu T^a - ig_3 G^a_\mu t^a \tag{2.9}$$

where Y, T^a and t^a are the generators for the $U(1)$, $SU(2)$, and $SU(3)$ gauge symmetry groups , respectively. Y imposes the conservation of the weak hypercharge in the combined $SU(2)_L \times U(1)_Y$ symmetry, and is represented as $Y = 2(Q - T_3)$, where Q is the electric charge. The other operators are given by: $T^a = \sigma^a/2$, where

$\sigma^a(a = 1,2,3)$ are the Pauli matrices, and $t^a = \lambda^a/2$, where $\lambda^a(a = 1,2,..8)$ are the Gell-Mann matrices. $B_\mu, W_\mu^a(a = 1,2,3)$ and $G_\mu^a(a = 1,2,...8)$ are the gauge fields for the $U(1)$, $SU(2)$, and $SU(3)$ gauge symmetry groups, respectively. Finally, g_1, g_2 and g_3 are the coupling constants between the fermions and the gauge fields.

We can now write down the kinematic term describing the gauge fields in the SM, L_{gauge}, as :

$$L_{gauge} = -\frac{1}{4}G_{a\mu\nu}G_a^{\mu\nu} - \frac{1}{4}W_{a\mu\nu}W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{2.10}$$

where the field strength tensors are defined as:

$$W_a^{\mu\nu} = \delta^\mu W_\nu^a - \delta^\nu W_\mu^a + g_2\epsilon_{abc}W_\mu^b W_\mu^c \tag{2.11}$$

$$G_a^{\mu\nu} = \delta^\mu G_\nu^a - \delta^\nu G_\mu^a + g_3 f_{abc}G_\mu^b G_\mu^c \tag{2.12}$$

$$B^{\mu\nu} = \delta^\mu B^\nu - \delta^\nu B^\mu \tag{2.13}$$

where $f_{abc}(a,b,c = 1,2,..8)$ and $\epsilon_{abc}(a,b,c = 1,2,3)$ are structure constants for the $SU(3)$ and $SU(2)$ groups, respectively, and are related to the generators as:

$$[T^a, T^b] = i\epsilon^{abc}T^c; \quad [t^a, t^b] = if^{abc}t^c. \tag{2.14}$$

We will now define the photon (A_μ), $W^\pm$ and Z bosons in the SM as linear combinations of B_μ and W_μ^a :

$$\begin{aligned} W_\mu^+ &= \frac{-W_\mu^1 + W_\mu^2}{\sqrt{2}} \\ W_\mu^- &= \frac{-W_\mu^1 - W_\mu^2}{\sqrt{2}} \end{aligned} \tag{2.15}$$

$$\begin{aligned} A_\mu &= \frac{g_2 B_\mu + g_1 W_\mu^0}{\sqrt{g_1^2 + g_2^2}} \\ Z_\mu &= \frac{-g_1 B_\mu + g_2 W_\mu^0}{\sqrt{g_1^2 + g_2^2}}. \end{aligned} \tag{2.16}$$

We can also re-define the coupling constants as

$$
\begin{aligned}
e &= \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}} \\
\sin\theta_w &= \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \\
\cos\theta_w &= \frac{g_2}{\sqrt{g_1^2 + g_2^2}}
\end{aligned}
\tag{2.17}
$$

where e is the electric charge and θ_w is known as the Weinberg angle. θ_w is also referred to as the *weak-mixing angle*, which can be understood from re-writing Eqn 2.16 as

$$
\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_w & \sin\theta_w \\ -\sin\theta_w & \cos\theta_w \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^0 \end{pmatrix}. \tag{2.18}
$$

Finally, we want to introduce a mass term for the fermions and bosons in our theory. To do this, we must first understand how fields change under local gauge transformations. Under the local U(1) symmetry, the fermion and gauge fields transform as

$$
\begin{aligned}
\psi_L &\longrightarrow e^{-i\alpha(x)}\psi_L \\
\psi_R &\longrightarrow e^{-2i\alpha(x)}\psi_R \\
B_\mu &\longrightarrow B_\mu + \frac{2}{g_1}\delta_\mu\alpha(x)
\end{aligned}
\tag{2.19}
$$

and under local SU(2) transformations, the fields change as

$$
\begin{aligned}
\psi_L &\longrightarrow e^{-i\alpha\vec{(x)}.\vec{\sigma}/2}\psi_L \\
W_\mu^a &\longrightarrow W_\mu^a + \frac{2}{g_2}\delta_\mu\alpha^a(x) + \epsilon_{abc}\alpha^b(x)W_\mu^c
\end{aligned}
\tag{2.20}
$$

and finally, under SU(3) transformations, the fields change as

$$
G_\mu^a \longrightarrow G_\mu^a + \frac{2}{g_3}\delta_\mu\alpha^a(x) + f_{abc}\alpha^b(x)G_\mu^c \tag{2.21}
$$

where $\vec{\alpha}$ specifies the rotation angles under the particular symmetry group and $\vec{\sigma}$ are the Pauli matrices. The L_{fermion} and L_{gauge} terms are invariant under these transformations. For gauge bosons, adding a mass term of the form $m_B^2 B_\mu B^\mu$, $m_W^2 W_\mu W^\mu$, or

$m_G^2 G_\mu G^\mu$ would break gauge invariance (although we know gluons are massless). A mass term for a fermion can be written as:

$$-m_f \bar{\psi}\psi = -m_f \bar{\psi}_L \psi_R + h.c. \tag{2.22}$$

The above equation also breaks gauge invariance, since ψ_L and ψ_R have different weak-hypercharges. Thus, it is not possible to add a mass term in the SM by hand. However, we can solve this issue by invoking *Spontaneous Symmetry Breaking*, as is explained in the following section.

2.2.1 Spontaneous Symmetry Breaking (SSB)

A scalar SU(2) doublet field (known as the Higgs field) is introduced in the SM to spontaneously break the electro-weak symmetry in vacuum, known as Electro-Weak Symmetry Breaking (EWSB). The fermions and gauge bosons can interact with this Higgs field to acquire masses. This process is known as the Brout-Englert-Higgs mechanism or simply, the Higgs mechanism [7, 13, 14, 20, 28–32]. The scalar field ϕ is defined as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_2 + i\phi_4 \end{pmatrix}. \tag{2.23}$$

The Lagrangian of the Higgs, L_{Higgs} can be written as

$$L_{\text{Higgs}} = T(\phi) - V(\phi) = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \tag{2.24}$$

The potential $V(\phi)$ is shaped like a Mexican hat, and is also known as the "Mexican-hat" potential, as shown in Fig. 2.2. The minimum of the potential $V(\phi)$ is at:

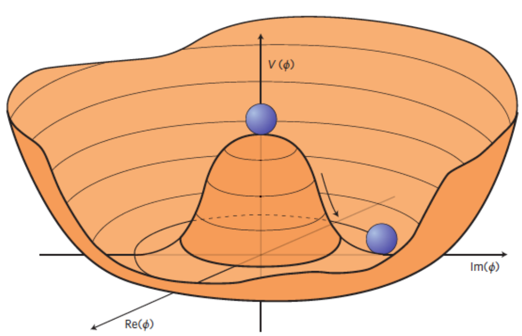


Figure 2.2: The Mexican hat shaped Higgs potential.

$$\phi^\dagger \phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}; \mu^2 < 0 \tag{2.25}$$

where the real component of $\phi^\dagger\phi$ is given by

$$\mathbb{R}(\phi^\dagger\phi) = \frac{\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2}{2}. \tag{2.26}$$

Given the form of Eqn. 2.26, there exist many solutions to Eqn. 2.25. One can choose any direction in this SU(2) phase space to define a vacuum configuration, but we will choose the minimum, ϕ_0, as

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.27}$$

where $v = \sqrt{\frac{-\mu^2}{\lambda}}$ is known as the Higgs vacuum expectation value or v.e.v., and $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$. We can now study tiny perturbations $H(x)$ around this minimum as

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{2.28}$$

Using the above equation, we can re-write the Higgs potential $V(\phi)$ as

$$V(\phi) = \frac{-\mu^4}{4\lambda} - \mu^2 H^2 + \lambda v H^3 + \frac{\lambda}{4}H^4. \tag{2.29}$$

From the above equation, $H(x)$ is a real scalar field that represents the scalar Higgs boson, and the second term in the equation can be re-written as

$$-\mu^2 H^2 = \frac{1}{2}m_H^2 H^2 \Rightarrow m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda}v \tag{2.30}$$

where m_H represents the mass of the Higgs boson. The cubic and quartic terms in Eqn. 2.29 represent the Higgs self-interaction terms. The parameter λ is not specified in the SM and represents the tri-linear Higgs self coupling. Although HH production is yet to be observed experimentally, it is one of the only ways in nature to measure λ and study the shape of the Higgs potential.

Furthermore, substituting Eqn. 2.28 in the kinetic term of L_{Higgs}, $T(\phi)$, gives :

$$\begin{aligned} T(\phi) = (D_\mu\phi)^\dagger(D^\mu\phi) = \frac{1}{2}(\delta_\mu H)^2 + \frac{g_1^2}{8}(-W_\mu^1 - iW_\mu^2)(-W^{1,\mu} - iW^{2,\mu})(v+H)^2 + \\ \frac{1}{8}(g_1 W_\mu^3 - g_2 B_\mu)(g_1 W^{3,\mu} - g_2 B^\mu)(v+H)^2 \end{aligned} \tag{2.31}$$

and then re-defining the above equation in terms of the physical gauge bosons (Eqn. 2.15 and 2.16) gives

$$\begin{aligned} T(\phi) &= (D_\mu\phi)^\dagger(D^\mu\phi) \\ &= \frac{1}{2}(\delta_\mu H)^2 + \frac{g_1^2}{4}W_\mu^+W^{-,\mu}(v+H)^2 + \frac{1}{8}(g_1^2+g_2^2)Z_\mu Z^\mu(v+H)^2 \\ &= \frac{1}{2}(\delta_\mu H)^2 + \frac{1}{2}(vg_2)^2W_\mu^+W^{-,\mu} + \frac{1}{2}vg_1^2W_\mu^+W^{-,\mu}H + \frac{1}{4}g_1^2W_\mu^+W^{-,\mu}H^2 + \\ &\frac{1}{2}(\frac{1}{2}v\sqrt{g_1^2+g_2^2})^2Z_\mu Z^\mu + \frac{1}{4}v(g_1^2+g_2^2)Z_\mu Z^\mu H + \frac{1}{8}(g_1^2+g_2^2)Z_\mu Z^\mu H^2. \end{aligned} \tag{2.32}$$

From the terms, $\frac{1}{2}(vg_2)^2W_\mu^+W^{-,\mu}$ and $\frac{1}{2}(\frac{1}{2}v\sqrt{g_1^2+g_2^2})^2Z_\mu Z^\mu$, we can conclude

$$m_W = \frac{1}{2}(vg_2); \quad m_Z = \frac{1}{2}v\sqrt{g_1^2+g_2^2} \tag{2.33}$$

where m_W and m_Z are the masses of the W and Z boson, respectively. The photon remains massless in this equation. We can re-write $T(\phi)$ as

$$\begin{aligned} T(\phi) = \frac{1}{2}(\delta_\mu H)^2 + m_W^2W_\mu^+W^{-,\mu} + \frac{2m_W^2}{v}W_\mu^+W^{-,\mu}H + \frac{m_W^2}{v^2}W_\mu^+W^{-,\mu}H^2 + \\ \frac{1}{2}m_Z^2Z_\mu Z^\mu + \frac{m_Z^2}{v}Z_\mu Z^\mu H + \frac{m_Z^2}{2v^2}Z_\mu Z^\mu H^2. \end{aligned} \tag{2.34}$$

The above equation describes the kinematics of the Higgs boson, the masses of the $W^\pm$ and Z bosons, and the triple and quartic couplings of the $W^\pm$ and Z bosons with the Higgs. In the interaction terms, the coupling is proportional to the squared mass of the gauge boson, and are known as *gauge couplings* of the Higgs boson.

Finally, the interaction of fermions with the Higgs boson and fermion masses can be explained by adding another L_{Yukawa} term to the SM Lagrangian (taking first generation leptons from Eqns. 2.3 and 2.4 as an example), as

$$\begin{aligned} L^e_{\text{Yukawa}} &= g_e(\bar{L_e}\phi e_R + \phi^\dagger\bar{e_R}L_e) \\ &= \frac{g_e v}{\sqrt{2}}(\bar{e_L}e_R + \bar{e_R}e_L) + \frac{g_e}{\sqrt{2}}(\bar{e_L}e_R + \bar{e_R}e_L)H^{\cdot} \end{aligned} \tag{2.35}$$

L^e_{Yukawa} is invariant under $SU(2)_L \times U(1)_Y$ transformations. Note that there is no mass term for neutrinos in the SM. From Eqn. 2.36, the mass term for the electron is given by $m_e = \frac{g_e v}{\sqrt{2}}$ and the second term gives the interaction strength between the electron and the Higgs boson, which is proportional to the m_e and is known as a *Yukawa coupling*.

Similarly, one can write the L_{Yukawa} for the first generation quarks as :

$$\begin{aligned} L^{Q}_{Yukawa} &= g_d \bar{Q}_L \phi d_R + g_u \bar{Q}_L \phi u_R + h.c. \\ &= m_d d\bar{d} + m_u u\bar{u} + \frac{m_d}{v} d\bar{d}H + \frac{m_u}{v} u\bar{u}H \end{aligned} \tag{2.36}$$

This prescription for adding mass terms for the first generation leptons and quarks can be extended to the second and third generation of fermions.

2.3 Limitations of the SM

The discovery of the Higgs boson completed the last missing piece in the SM. Although the SM is very accurate in its predictions, it is not a complete theory of our universe. It does not account for several observations, some of which are:

- Baryogenesis [33]: the observed universe is mostly made up of matter, although both matter and anti-matter should have been produced in equal amounts in the early universe. This is also known as the "matter anti-matter asymmetry" problem.
- Gravity : The SM does not explain one of the most well understood force.
- Dark Matter : Astronomical and cosmological observations have indicated the existence of a form of matter [34] in our universe that does not interact electromagnetically, and occupies about 27% of the known universe. The SM provides no candidates for dark matter particles.
- Neutrino masses : Neutrino oscillations have been observed in nature [35–37] and indicate that neutrinos must have tiny masses. However, there is no formulation for adding neutrino mass terms in the SM.
- Hierarchy problem : If one considers the SM to be an effective field theory (EFT) at low energies, the Yukawa interaction between the Higgs and fermion fields introduces quadratic corrections to the Higgs mass that go up to the ultraviolet cutoff scale Λ_{UV}:

$$\begin{aligned} m_h^2 &= m_0^2 + \delta m_h^2 \\ \delta m_h^2 &\supset -\frac{|y_f|^2}{8\pi^2}\Lambda_{UV}^2, \end{aligned} \tag{2.37}$$

where m_0 is the bare Higgs boson mass and δm_h^2 are the loop corrections, which are mainly dominated by the top quark loops. The Λ_{UV} cutoff scale is expected to be near the Planck scale, $M_P \sim O(10^{19})$ GeV, while the mass of the Higgs has been experimentally measured to be ~125 GeV. This would mean Eqn. 2.37 is carefully fine tuned, and the subtraction of two extremely large numbers results in a quantity (m_h^2) that is 34 orders of magnitude smaller. This fine-tuning conflicts with the idea of naturalness [38], and is known as the hierarchy problem.

To address these issues, various Beyond Standard Model (BSM) theories such as Supersymmetry [20], Two-Higgs-doublet models [21], extra dimensions [22], sterile neutrinos [23], etc. have been proposed. They all predict new fundamental particles which can solve one or more of the problems listed above, and/or alter the predictions of the SM. Several experiments are underway to test the validity of such BSM theories, either by direct/indirect detection methods or by finding deviations from the SM predictions in precision measurements.

2.4 Higgs boson physics at the LHC

This section will provide an overview of the main Higgs Boson production modes at the LHC, the various Higgs decay modes and a brief compilation of the various measurements of the Higgs properties from the CMS experiment.

2.4.1 Single Higgs production and decay modes at the LHC

At the LHC, the main Higgs production mechanism is the gluon fusion (ggF) process, due to the large gluon density within the colliding protons. The next major production mode is the vector boson fusion (VBF) process, which is characterized by the presence of two additional quarks in the final state, which show up in the detectors as energetic jets produced in the forward regions with a large pseudorapidity separation. The third major production mode is the associated production of Higgs with a vector boson (VH), and one can use the decays of the vector boson to tag events produced via this mechanism. Finally, the Higgs boson can also be produced in association with a top or bottom quark pair. The ttH production mode can be tagged using the experimental signature of top quark decays. The bbH process is soft and similar in kinematics to the ggF process, which makes it very difficult to detect it at the LHC. The tree-level Feynman diagrams for these dominant Higgs boson production processes are shown in Fig. 2.3. The Higgs boson production cross sections for different production mechanisms are summarized in Table 2.1, for a Higgs Boson mass of 125.38 ± 0.14 GeV (the most precise Higgs mass measurement: Ref [39]).

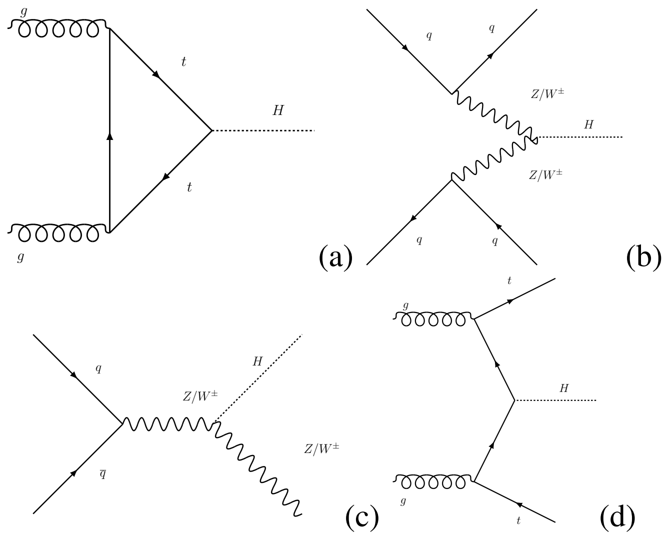


Figure 2.3: Different production modes for the Higgs boson at the LHC : (a) ggH, (b) VBF, (c) VH, and (d) ttH.

Table 2.1: Higgs boson production cross sections for various modes at $\sqrt{s}$ = 13 TeV [40]. The uncertainties on the cross sections are the quadratic sum of the uncertainties from variations of QCD scales, parton distribution functions and the strong interaction coupling strength, α_s.

Process	Cross section (pb)	Perturbative Order
ggH	48.30 ± 2.42	N3LO (QCD) + NLO (EW)
VBF	3.77 ± 0.08	NNLO (QCD) + NLO (EW)
WH	1.358 ± 0.03	NNLO (QCD) + NLO (EW)
ZH	0.8767 ± 0.04	NNLO (QCD) + NLO (EW)
ttH	0.5033 ± 0.05	NLO (QCD) + NLO (EW)
bbH	0.4822 ± 0.12	NLO (QCD)
Total	55.37 ± 2.4	

The main Higgs decay channels with their expected branching ratios at m_H = 125.38 ± 0.14 GeV are summarized in Table 2.2. The Higgs total decay width (Γ_H) is known to be 4.14 ± 0.05 MeV [40].

2.4.2 Double Higgs (HH) production and decay modes at the LHC

At the LHC, the dominant HH production happens via the gluon fusion process, shown in Fig. 2.4, and the second most common production mode is the vector

Table 2.2: Higgs boson decay branching ratios for various modes at $\sqrt{s}$ = 13 TeV [40].

Decay channel	Branching ratio (%)
$H \to b\bar{b}$	57.60
$H \to WW^*$	22.03
$H \to gg$	8.15
$H \to \tau\bar{\tau}$	6.21
$H \to c\bar{c}$	2.86
$H \to ZZ^*$	2.72
$H \to \gamma\gamma$	0.227
$H \to Z\gamma$	0.157
$H \to \mu\bar{\mu}$	2.18×10^{-2}

boson fusion, as shown in Fig. 2.5. The production cross section at 13 TeV for m_H = 125 GeV is σ_{ggHH} = 31.05 fb [41–47] for the gluon fusion production mode and σ_{qqHH} = 1.726 fb [44, 48–51] for the vector boson fusion production mode. The two leading-order diagrams of the gluon fusion production (Fig. 2.4), are known as the box (left) and triangle (right) diagrams. The branching ratios of HH decay to various channels are summarized in Fig. 2.6. HH production is yet to be observed experimentally in nature. As mentioned in Sec. 2.2.1, the HH production is one of the only ways in nature to measure the tri-linear Higgs self coupling, λ_{HHH}, and understand the shape of the Higgs potential (Eqn. 2.29). Measuring λ_{HHH} will also help us to understand the stability of the electroweak vacuum, i.e., it will tell us whether the electroweak vacuum is a "false" or a "true" vacuum state [52].

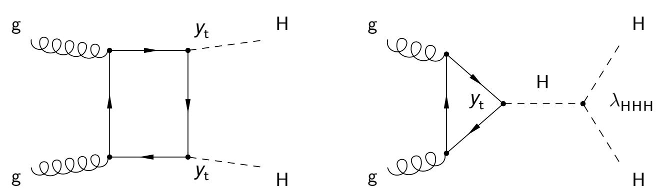


Figure 2.4: Feynman diagrams that contribute to $ggHH$ at leading order. The left and right diagrams correspond to SM processes, referred to as the box and triangle diagrams, respectively.

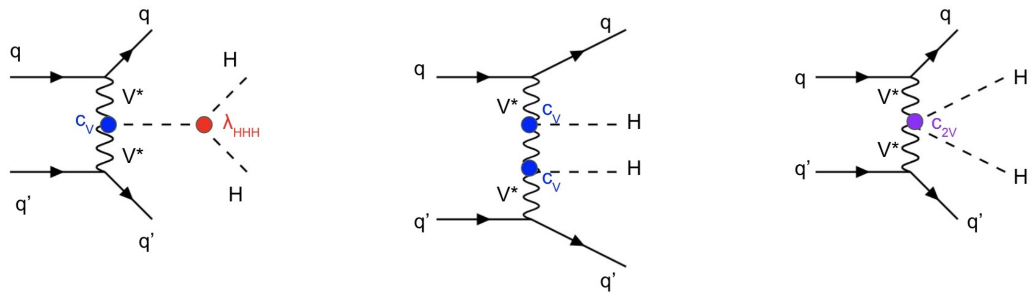


Figure 2.5: Feynman diagrams that contribute to qqHH at leading order. The qqHH process is sensitive to c_V and the c_{2V} couplings.

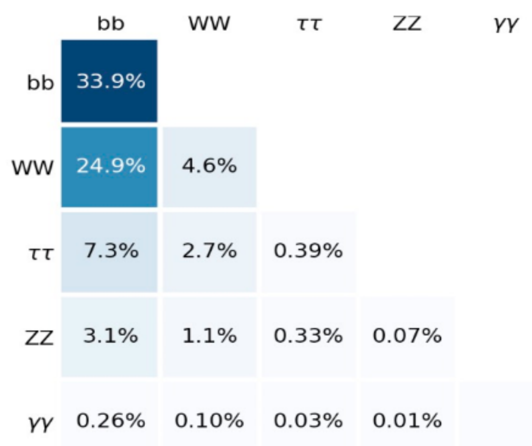


Figure 2.6: The branching ratios of various HH decay modes.

2.4.3 Summary of Higgs boson measurements from the CMS experiment at the LHC

The most precise measurement of the Higgs mass to-date is $m_H = 125.38 \pm 0.14$ [39] and is shown in Fig. 2.7 (top left). Under the assumption of a coupling structure similar to that in the standard model, Ref. [53] constrains the observed Higgs boson width to $3.2^{+2.8}_{-2.2}$ MeV while the expected constraint based on simulation is $4.1^{+5.0}_{-4.0}$ MeV, and is shown in Fig. 2.7 (top right). Fig. 2.7 (bottom) from Ref. [24] shows the best-fit values for the various coupling strengths of different SM particles with the Higgs Boson. All these measurements are consistent with the SM predictions.

Fig. 2.8 shows a summary of the results on upper limits on the inclusive HH production cross section ($\mu_{HH} = \sigma_{HH}/\sigma_{HH}^{SM}$), the VBF HH production cross section in a BSM scenario with $\kappa_{2V} = 0$ ($\mu_{HH} = \sigma_{HH}/\sigma_{HH}^{\kappa_{2V}=0}$), κ_λ ($= \lambda/\lambda^{\mathrm{SM}}$) and κ_{2V} ($=c_{2V}/c_{2V}^{\mathrm{SM}}$) [54] from various CMS physics analyses (Ref. [26, 55–59]). All these measurements are consistent with the predictions of the SM. There are ongoing efforts within CMS to produce a combination of these measurements, which will

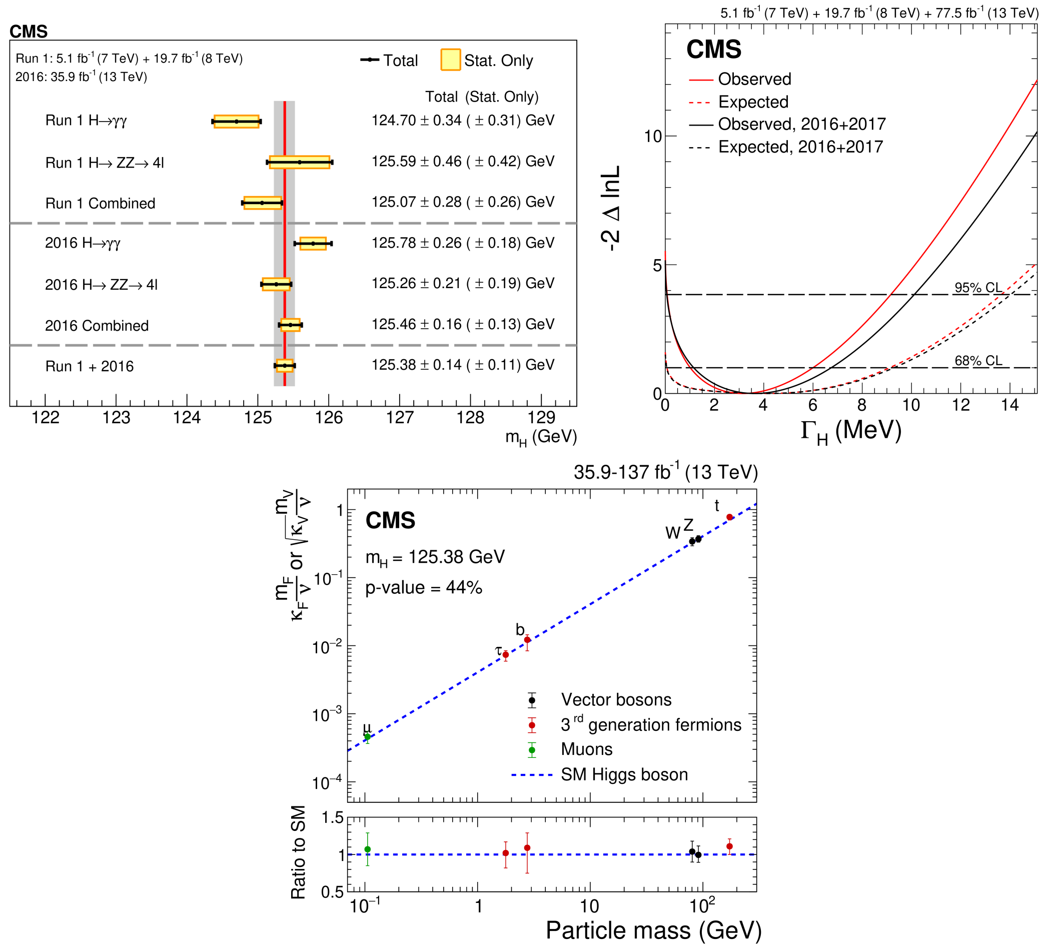


Figure 2.7: Various Higgs measurements from CMS: (Top left) Summary of the measured Higgs boson mass in the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ decay channels, and their combination [39]. (Top right) Likelihood scan of the Higgs decay width [53]. (Bottom) Coupling strengths of various SM particles to the Higgs boson as a function of the particle mass. The bottom panel compares the observed values with the prediction of SM [24].

put very stringent constraints on the allowed values of these parameters. A more detailed discussion on Fig. 2.8 can be found later in Sec. 6.7.

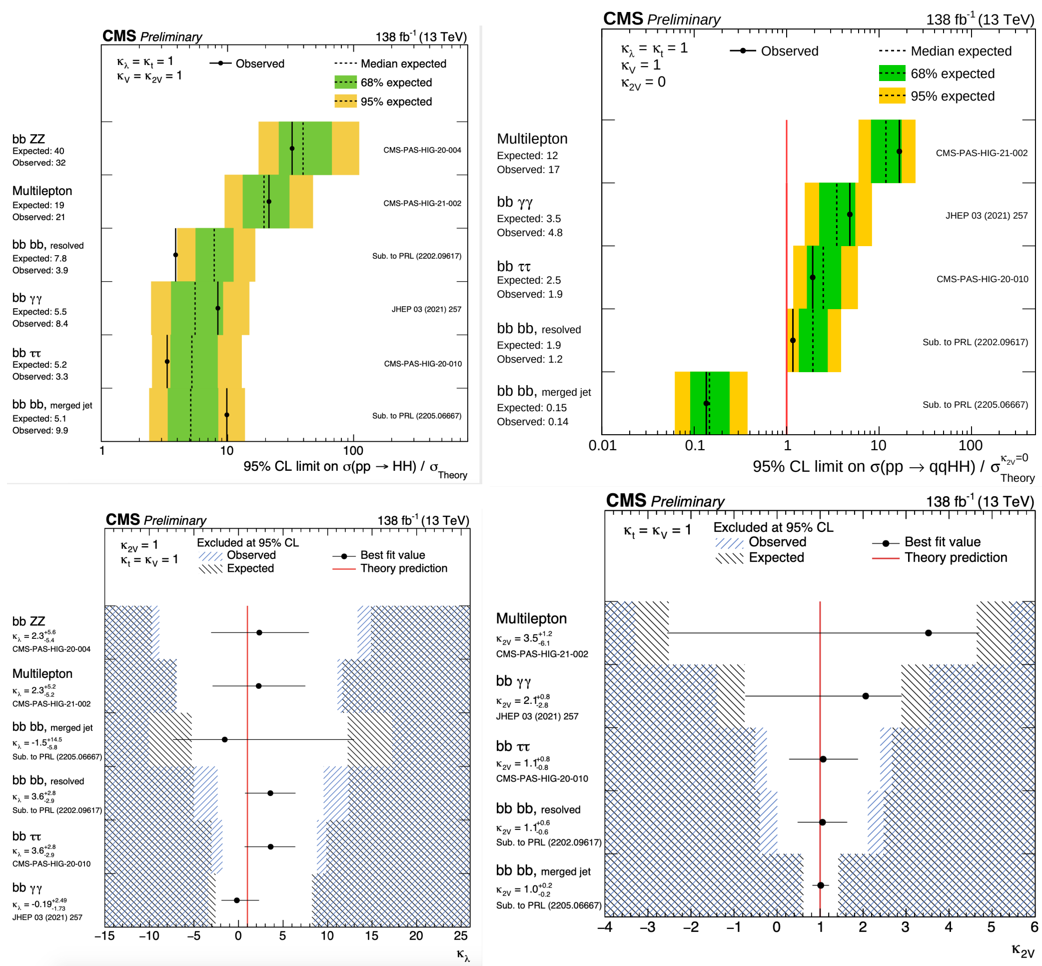


Figure 2.8: Summary of various HH measurements from CMS: (Top) Upper limits at 95% confidence level on $\mu = \sigma_{HH}/\sigma_{HH}^{SM}$ (left) and $\mu = \sigma_{qqHH}/\sigma_{qqHH}^{\kappa_{2V}=0}$ (right). The bbbb resolved and bbbb boosted results are derived using a phase space with a minor overlap of signal events. (Bottom) 95% confidence intervals on κ_λ (left) and κ_{2V} (right). The blue (black) hashed band indicates the observed (expected) excluded regions, respectively. The band around the best fit value corresponds to the one sigma interval. These plots are taken from [54].

Chapter 3

THE COMPACT MUON SOLENOID (CMS) EXPERIMENT AT THE LHC

3.1 The Large Hadron Collider (LHC)

The Large Hadron Collider or the LHC [17] is a 27 km long accelerator ring, located ~100 m underground at the border of Switzerland and France. It is operated by European Organization for Nuclear Research (CERN) and is the world's largest and most powerful particle accelerator. The LHC accelerates, circulates, and focuses two counter-rotating beams of bunched hadrons. The LHC operates in four collision modes: (a) proton-proton (pp) collision, (b) proton-lead (p-Pb) collision, (c) Lead-Lead (Pb-Pb) and (d) Xenon-Xenon (Xe-Xe) collision. There are four major detectors located at four different interaction points (IP), namely ALICE [60], ATLAS [61], CMS [62], and LHCb [63]. Figure 3.1 shows the geographical location of the LHC ring, along with the four major experiments. ATLAS and CMS are general purpose detectors designed to cover a broad physics program, whereas ALICE and LHCb are dedicated for heavy-ion physics and b-physics, respectively. Due to the scope of this thesis, going forward, we will only discuss the pp collisions. During Run 2 of the LHC (2015-2018), pp collisions occurred at $\sqrt{s} = 13$ TeV.

The accelerator complex at CERN is a succession of machines, where each machine boosts the energy of a beam of particles before injecting it into the next machine in the sequence (Fig. 3.2). Proton beams are produced by ripping off electrons from Hydrogen atoms. The protons are passed through radiofrequency quadrupole (RFQ) magnets, where they are focused and bunched together. Each bunch contains about 1.15×10^{11} protons. Bunches of protons are then injected into Linear accelerator 2 or Linac2, and accelerated upto energies of 50 MeV[1]. The ions are then injected into the Proton Synchrotron Booster (PSB), where they are accelerated to 1.4 GeV. They are then subsequently injected into the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS), which accelerate the beam up to 26 GeV and 450

[1] Linac2 was decommissioned at the end of 2018 and is now replaced by Linear accelerator 4 (Linac4), which will accelerate negative hydrogen ions (H^-, consisting of a hydrogen atom with an additional electron) to 160 MeV. The ions will be stripped of their two electrons during injection from Linac4 into the PSB, leaving only protons. These protons will be accelerated to 2 GeV inside the PSB. The remaining chain of acceleration remains the same.

Figure 3.1: An aerial view of the LHC ring at the border of France and Switzerland, along with the four major experiments at the different IPs. Picture Courtesy : CERN

GeV, respectively. The protons are then transferred to the two beam pipes of the LHC, and the two separate beams counter-rotate. It takes roughly 20 minutes for the protons to reach their maximum energy of 6.5 TeV per beam. The two beams collide at the four different IPs once every 25 ns. At a single moment in time, the full LHC tunnel is filled with 2808 proton bunches. Due to the limited space in

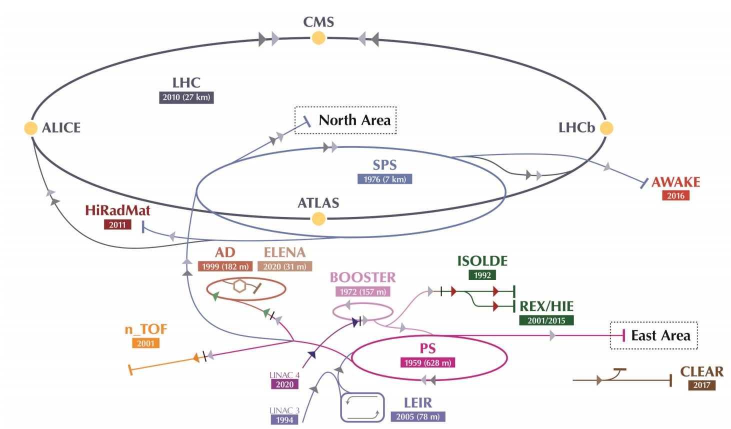


Figure 3.2: The accelerator complex at CERN [64].

the LHC tunnel (diameter of ~3.7 m), a twin-bore magnet design [65] was adopted (instead of separate dipole magnets per beam ring), and the two beam rings share the same cryostat and mechanical structure. The superconducting magnets are made of Nb-Ti cables, cooled by superfluid helium at a temperature of 1.9K, which provides a nominal magnetic field of 8.33 T. There are two main types of magnets in the LHC: 1232 dipole magnets for bending the beams and 392 quadrupole magnets for focusing the beams. There are additional magnets within the dipole, namely the sextupole, octupole and decapole magnets, which correct for small imperfections in the magnetic field.

The event rate of a particular process depends on its cross section and the LHC's instantaneous luminosity ($\mathcal{L}$), given by:

$$\frac{dN}{dt} = \mathcal{L}\sigma \tag{3.1}$$

where $\frac{dN}{dt}$ is the event rate, and σ is the cross section of the process. $\mathcal{L}$ can be further expressed as a function of accelerator parameters as

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

where N_b is the number of particles in a bunch (1.15×10^{11}), n_b is the number of bunches circulating in the ring (2808), f_{rev} is the frequency of the bunch revolution ($f_{rev} = 1/T = \frac{299792\ km/s}{26.659\ km} = 11245\ Hz$ (rings/second)), γ_r is the Lorentz boost factor (= 6929 for 6.5 TeV protons), ϵ_n is the normalized (w.r.t the beam energy) transverse beam emittence (3.75 μm), β^* is the beta function at the collision point (0.55 m), and F is the geometric luminosity reduction factor. The factor $\epsilon_n\beta^*$ defines the beam spot size at the IP and the factor F is defined as

$$F = \frac{1}{\sqrt{1 + (\frac{\theta_c \sigma_z}{2\sigma_{xy}})^2}} \tag{3.3}$$

where θ_c is the crossing angle (285 μrad), σ_z is the root-mean-square (rms) size of the bunch in the longitudinal direction (7.55 cm), and σ_{xy} is the rms size of the bunch in the transverse direction (16.7 μm) . The factor F essentially characterizes the reduction in instantaneous luminosity due to a non-zero bunch crossing angle. Plugging in all the numbers, the value of $\mathcal{L}$ comes to $\sim\ 1.0 \times 10^{34}\ cm^{-2}\ s^{-1}$, which is the design luminosity of the LHC. However, due to improvements in the various operational parameters (like increased maximum bunch intensity or smaller emittance), the peak LHC luminosity during Run 2 reached $\sim 2.0 \times 10^{34}\ cm^{-2}\ s^{-1}$ (for the years 2017 and 2018).

Multiple pp collisions can occur per bunch crossing at the LHC. Most of these interactions produce soft particles with low momentum transfer, and are known as pileup interactions. Pileup interactions pose a challenge for the detectors: pileup particles reduce the efficiency of particle reconstruction from actual hard interactions of interest. With the total p-p inelastic cross-section $\sigma_{inel}^{pp}(80\ mb)$, instantaneous luminosity $\mathcal{L}(\sim 2.0{\times}10^{34}\ cm^{-2}\ s^{-1})$, and time between consecutive bunch crossings ($\Delta T = 25ns$), one can calculate the average pileup ($<\mu>$) as :

$$<\mu>= \mathcal{L} \times \sigma_{inel}^{pp} \times \Delta T. \tag{3.4}$$

Plugging in the values, we obtain $<\mu> = 40$. This value is very close to the actual number of pileup interactions observed in the LHC during 2018. Fig. 3.3 (top) shows the peak instantaneous luminosity for different years of LHC operation since 2010. Fig. 3.3 (bottom) shows pileup event distribution for the same years [66]. The average pileup increases with increasing instantaneous luminosity.

3.2 The CMS Experiment

The Compact Muon Solenoid (CMS) is one of the two large general purpose detectors built at the LHC. It was designed to study many aspects of the SM (including the Higgs boson discovery, precision electroweak tests, heavy flavor physics, etc) and also test various BSM theories (like SUSY, 2HDM, etc). The detector was designed with an aim to probe the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ channels, and thus has an excellent resolution to reconstruct photons and electrons (using the electromagnetic calorimeter or ECAL) and also muons (using the tracker and muon systems). It is capable of accurately tagging hadronic decays of τ leptons and b-quarks. With the nearly hermetic coverage of the detector, it also provides a good measurement of the missing momentum in the transverse plane ($p_{\mathrm{T}}^{\mathrm{miss}}$ or MET), which is essential for several BSM analyses.

The CMS experiment is shown in Fig. 3.4. The detector is 28.7 m long, with a radius of 7.5 m, and weighs 14,000 tonnes. The various subsystems of the detector are built in the shape of concentric cylinders placed around the main interaction point. The innermost sub-detector is the Silicon Tracker that helps to detect charged particle tracks. The ECAL is placed just outside the tracker and is used for reconstructing

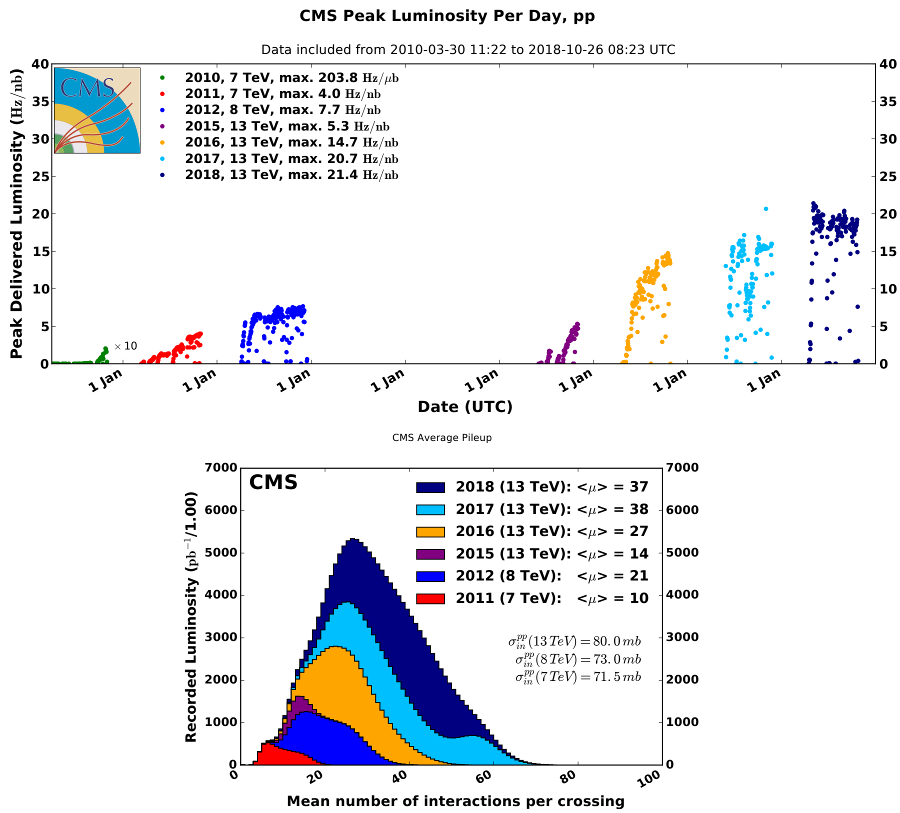


Figure 3.3: Recorded instantaneous luminosity and pileup at CMS during Run 1 and Run 2. (Top) The peak instantaneous luminosity during different years of operation (2010-2012 and 2015-2018). (Bottom) The corresponding average pileup distribution recorded by CMS [66].

electromagnetic showers. The hadron calorimeter (HCAL) is placed outside the ECAL and can measure energy deposits from charged and neutral hadrons. A superconducting solenoid magnet coil is placed outside the HCAL, which provides a magnetic field of 3.8 T, and helps to bend charged particle tracks. Gas filled muon stations are placed outside the solenoid magnet and interleaved in the steel return yokes (that provide a return magnetic field of 2T), and help in the identification of muons.

CMS uses a right-handed cartesian coordinate system where the origin is taken as the nominal interaction point, the x-axis points towards the center of the LHC ring, the y-axis points upwards, i.e., perpendicular to the plane of the LHC ring, and the z-axis is aligned with the beamline. The x-y plane is also known as the transverse plane. In spherical coordinates, the polar angle θ is measured from the positive z-axis and the azimuthal angle ϕ is measured in the plane transverse to the beam axis (Fig. 3.5). The pseudorapidity η is defined as $-\ln\tan\theta/2$. The momentum of

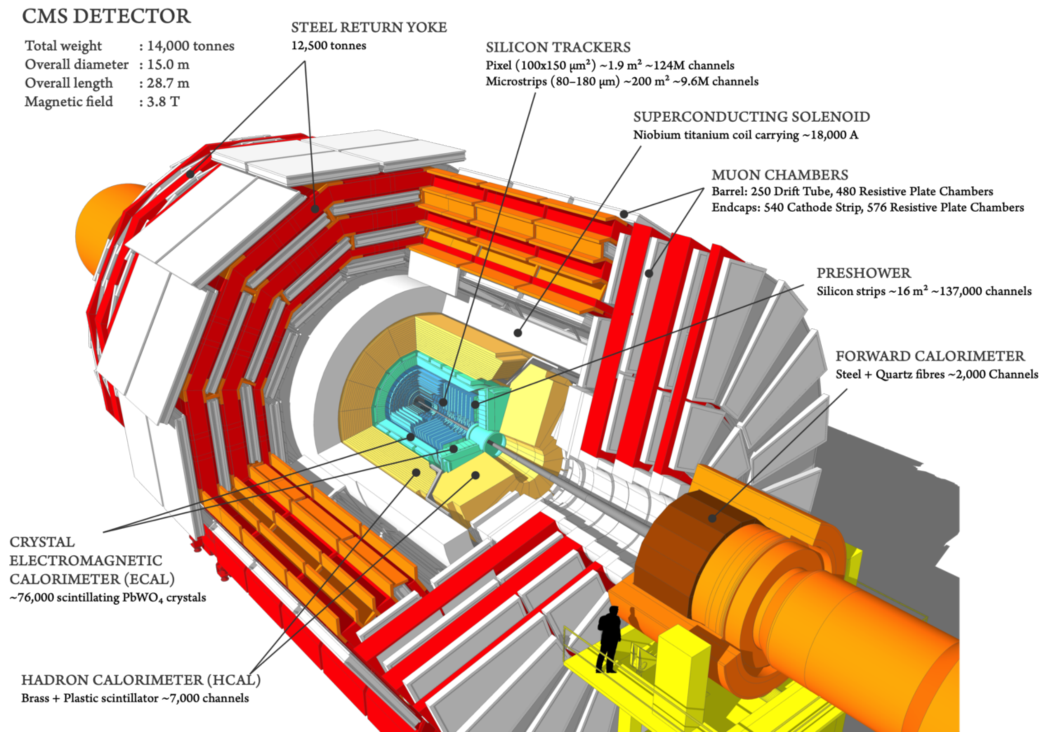


Figure 3.4: A cutaway diagram of the CMS experiment, indicating its various subsystems [67].

outgoing particles is measured in the transverse plane and is defined as $p_{\mathrm{T}} = p \sin\theta$, where p is the momentum in the 3D plane.

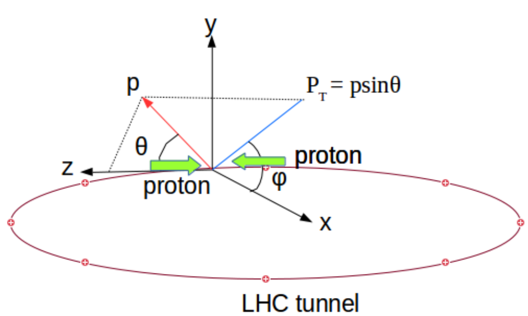


Figure 3.5: The right handed coordinate system used by CMS [68].

The next few sections will describe the various sub-systems of the CMS Experiment in detail.

3.2.1 Tracker

The silicon tracker system [69] is used to detect charged particle hits, thereby helping to reconstruct their momenta as well as find their point of origin (primary or secondary vertex). The innermost layers of the tracker are situated 4.4 cm away from the beamline, and are exposed to heavy radiation dosages, leading to radiation hard silicon as the choice of material. Fig. 3.6 shows a layout of the tracker. The outer radius of the tracker extends to ~ 110 cm and the full length of the tracker is ~ 540 cm, giving a pseudorapidity coverage of $|\eta| < 2.5$ - the central barrel region covers $|\eta| < 1.2$ and the two endcaps disks on either side of the barrel cover $1.2 < |\eta| < 2.5$. The tracker has four pixel layers in the barrel region (at radii of 3.0, 6.8, 10.9, and 16.0 cm), and has six pixel endcap disks (at z of ±29.1, ±39.6, and ±51.6 cm), with a total of 79 (45) million pixels in the barrel (endcaps) [2]. The area of each cell in the pixel layer is about 100×150 μm^2 with 300 μm thickness and provides a position resolution of ~10 μm.

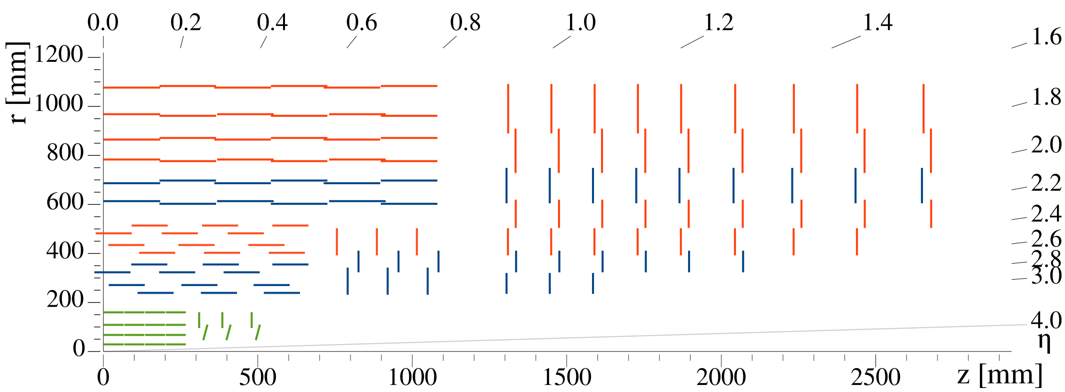


Figure 3.6: A schematic view of one quadrant of the Phase-1 CMS tracking system in the r-z plane. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively [71].

The outer strip tracker consists of 9.3 million strip sensors, with the barrel region divided into a Tracker Inner Barrel (TIB) and a Tracker Outer Barrel (TOB), while the endcap region is divided into the Tracker End Cap (TEC) and Tracker Inner Disks (TID). The TIB consists of 4 layers of 320 μm-thick silicon sensors and a strip pitch (distance between two strips) of 80 to 120 μm. The TOB includes 6 layers of 500 μm-thick sensors and a strip pitch of 120 to 180 μm. The TOB is farther away from the IP, and can afford to use thicker sensors and wider pitch due to smaller particle flux as compared to the TIB. The TEC is made of 9 disks and the TID is made of 3 small disks of 320 μm thick silicon sensors.

[2] The current pixel detector was installed during February-March 2017 (Phase I upgrade) [70]. Before this upgrade, there were 48 million pixels in the barrel region in three layers at radii of 4.3, 7.3, and 10.4 cm, covering z from -27 cm to 27 cm, and 18 million pixels in the endcap populated in four disks at z of ±35.5 and ±46.5 cm.

More details of the particle track and momentum reconstruction can be found in Sec. 3.2.7.

3.2.2 ECAL

The ECAL [72] is a homogeneous calorimeter made of $PbWO_4$ scintillator crystals, and is placed outside the tracker. The ECAL is divided into two regions - a barrel region providing coverage of $|\eta| < 1.479$ and and two endcap disks with $1.479 < |\eta| < 3.0$. There are a total of 61200 (7324) $PbWO_4$ crystals used in the ECAL barrel (each endcap disk). The physical dimension of each crystal is $2.2 \times 2.2 \times 23\ cm^3$ in the barrel regions and $2.86{\times}2.86{\times}22$ cm^3 in the endcaps. The lateral area of the crystals corresponds to $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$. The lateral dimension is equivalent to one Moliére radius (R_M = 2.19 cm) and the longitudinal length corresponds to 24.7 radiation lengths (X_0 = 0.85 cm). This ensures that ~98% of an electromagnetic shower is contained within a crystal. The ECAL's thickness corresponds to about 1 interaction length (λ_I), and the probability of a hadron depositing all its energy in the ECAL is small. This aids in the separation of individual hadrons and jets from electrons and photons. Crystals are also positioned with a small angle relative to the IP (~3°), to avoid particles travelling in the inter-crystal gaps.

Additionally, a thin *preshower* detector is placed in front of each ECAL endcap disk. The preshower is made up of two planes of lead (Pb) absorbers of ~ 2 X_0 and 1 X_0, respectively, interleaved with silicon micro-strip layers. The preshower is highly granular, with silicon sensors that are 1.9 mm wide. The fine transverse granularity of the preshower helps differentiate between prompt photons (i.e., produced at the collision vertex) and photons originating from a neutral pion decay ($\pi^0 \to 2\gamma$), where the angle between the two photons is small. The layout of the ECAL is shown in Fig. 3.7 [73].

The scintillation light produced in the $PbWO_4$ crystals has a spectrum that peaks around 420-430 nm. The photodetectors in the ECAL use avalanche photodiodes (APDs) in the barrel and vacuum phototriodes (VPTs) in the endcaps. There are two APDs glued to the back of each crystal in the barrel, each with an active area of 5 x 5 mm^2, a rise time of < 2 ns, an operating voltage between 340-430 V and a typical dark current of ~3 nA. The VPTs are 25 mm in diameter and are glued to the back of each endcap crystal. VPTs have a lower quantum efficiency than APDs, but are very radiation tolerant and are more suitable for the higher radiation doses in the endcaps.

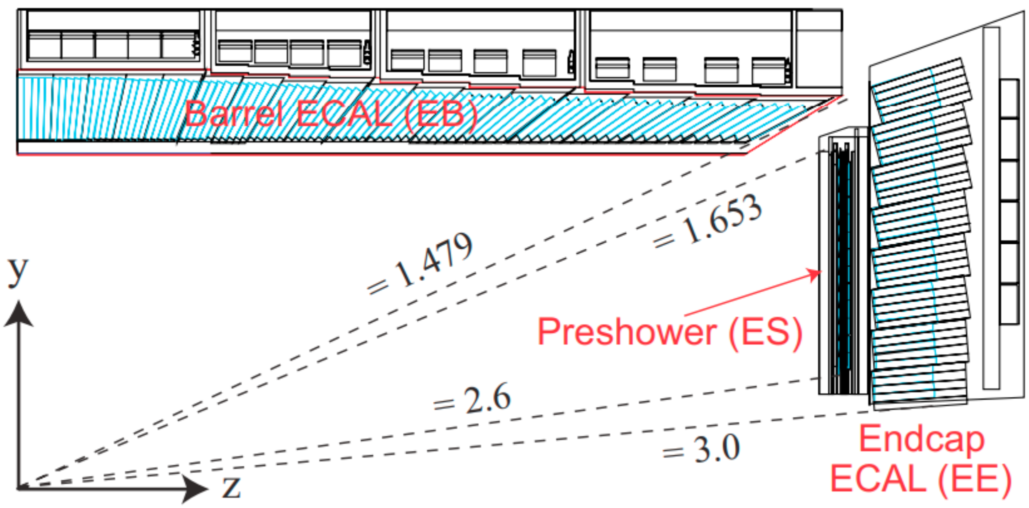


Figure 3.7: A geometric view of one quadrant of the ECAL [73].

The energy resolution of ECAL for an electromagnetic shower is given by:

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{3.5}$$

where the $1/\sqrt{E}$ is the stochastic term, related to statistic fluctuations in the signal, $1/E$ term is related to the electronics noise and pileup, and the constant term is related to the uncertainties associated with channel-by-channel calibration, leakage, dead material, etc.

3.2.3 HCAL

The CMS HCAL [74] is a sampling calorimeter placed right outside the ECAL, and is divided into 4 subsystems: barrel (HB) for $|\eta| < 1.4$, endcap (HE) for $1.3 < |\eta| < 3.0$, forward (HF) for $3.0 < |\eta| < 5.2$, and an additional subsystem in the barrel but outside the solenoid cryostat called HO which covers $|\eta| < 1.2$. The hermetic coverage of HCAL is crucial in jet reconstruction and measurement of missing transverse energy ($p_{\mathrm{T}}^{\mathrm{miss}}$) in an event.

Both HB and HE use plastic scintillators as the active layers interspersed between brass absorbers. The HO also uses plastic scintillators as the active layers, but uses the iron and the solenoidal coil as the absorber. The HF is installed 11m downstream of the IP and uses scintillating quartz fibres as the active material and steel as the absorber. The HF quartz fibers are of two types: (a) Long fibers that cover the full longitudinal range of HF, and (b) Short fibers, that start from 22 cm downstream of the front face of HF. The two different fiber types help to efficiently reconstruct the electromagnetic and hadronic components of hadronic showers. Due to space constraints (the HCAL had to fit in between the ECAL and the solenoid magnet), the longitudinal length of HCAL was limited to ~6 to 10 interaction lengths (λ_I) depending on $|\eta|$. The ECAL + HCAL together reach ~7 λ_I at $\eta = 0$. The total HB+HO depth in the central region is ~11.8 λ_I, which helps to improve energy resolution for high energy pions.

In HB, HE, and HO, the scintillation light (blue-violet, λ = 410-425 nm) is carried by wavelength-shifting (WLS) fibers (green, λ = 490 nm) to hybrid photo-diodes (HPDs) for detection. As part of the Phase I upgrade, HPDs were replaced by silicon photomultipliers (SiPMs) in HE (2017) and HB (2019) as they offer 2.5 times higher photon detection efficiency and 400 times higher response while being insensitive to magnetic fields [75–77]. Each segment of the calorimeter tower in HB, HE ($|\eta|$ < 1.6), and HO has dimensions $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. For $|\eta| > 1.6$, each segment in HE has lateral dimensions $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$. Fig. 3.8 shows the HCAL layout, before (left) and after (right) the Phase I upgrade. For both the HB and HE, the light yield from the different layers with the same color are optically summed up, resulting in a loss of longitudinal depth segmentation information. The HB (HE) is read out as four (six/seven) longitudinal segments.

The scintillating light in HF is carried by WLS fibers and read out by photomultiplier tubes (PMTs). Each segment of the HF has lateral dimensions of about $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$.

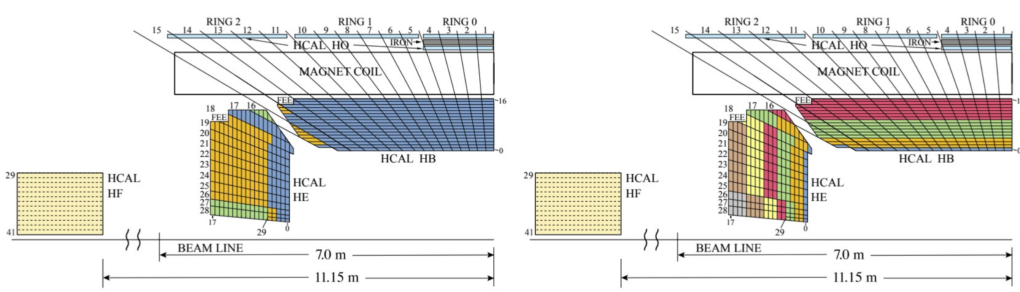


Figure 3.8: The depth segmentation in the previous (left) and upgraded (right) HCAL detector. A depth is a collection of scintillator layers (columns) depicted with the same color, per $\Delta\eta$ division. Light from the different layers in the same depth and $\Delta\eta$ division are optically added together before reaching the photosensors [75].

In general, the hadronic components (h) of a hadronic shower produce a smaller response than the electromagnetic components (e) in the detector, due to an "invisible energy" that results from atomic nuclei breaking up in the calorimeter material. This invisible energy does not add to the calorimetric response to a hadronic shower. This uneven detector response to the various shower components is called non-compensation. The CMS HCAL is a highly non compensating detector, i.e., e/h ∼1.4. This causes intrinsic sensitivity to fluctuations in the electromagnetic component of the shower, and non-linearity. The non-linearity can be corrected, but the fluctuations depend from shower to shower and cause a poor resolution. The combined energy resolution of the CMS HCAL+ECAL for pions is given by:

$$\begin{aligned} \frac{\sigma}{E} &= \frac{110.7\%}{\sqrt{E}} \oplus 7.3\% \text{ (before non-linearity correction)} \\ &= \frac{84.7\%}{\sqrt{E}} \oplus 7.4\% \text{ (after non-linearity correction).} \end{aligned} \tag{3.6}$$

Appendix A discusses the feasibility of a muon MIP based inter-depth energy calibration technique for the HE detector during Run 3 data-taking period.

3.2.4 Superconducting solenoid magnet

The CMS solenoid magnet is a large superconducting coil, with a 6m diameter and a 12.5 m length, and provides a very homogeneous magnetic field of 3.8 T. The solenoid is an alloy of Niobium and Titanium (NbTi) that can conduct currents of ∼ 19.14 kA and can store energies of up to 2.6 Giga-Joules. The magnetic field can technically be extended up to 4 T, but this is avoided to prevent the magnetic field lines from reaching the CMS support structure and causing power losses through Eddy current heating. The solenoid is enclosed in 10,000 tonnes of steel flux return yoke. The return yoke consists of three separate layers. The map of the magnetic field in the entire CMS detector volume was measured with the aid of cosmic muon events, and can be seen in Fig. 3.9.

3.2.5 Muon chambers

Muons with momentum in the range 1-100 GeV mostly lose energy by ionization when passing through a medium. For a muon MIP (a minimum ionizing particle whose mean energy loss rate through a particular material is close to the minimum), the energy lost while traveling through iron (ρ = 7.87 g/cm^3) is ∼11 MeV/cm, as can be inferred from Fig. 3.10. Thus, high energy muons can cross the CMS solenoid magnet without losing most of their energy, and therefore the positioning of the muon chambers outside the solenoid provides a very clean environment for muon detection. The muon chambers consist of 3 different technologies: drift tube chambers (DTs) in $|\eta| < 1.2$, cathode strip chambers (CSCs) in $0.9 < |\eta| < 2.4$,

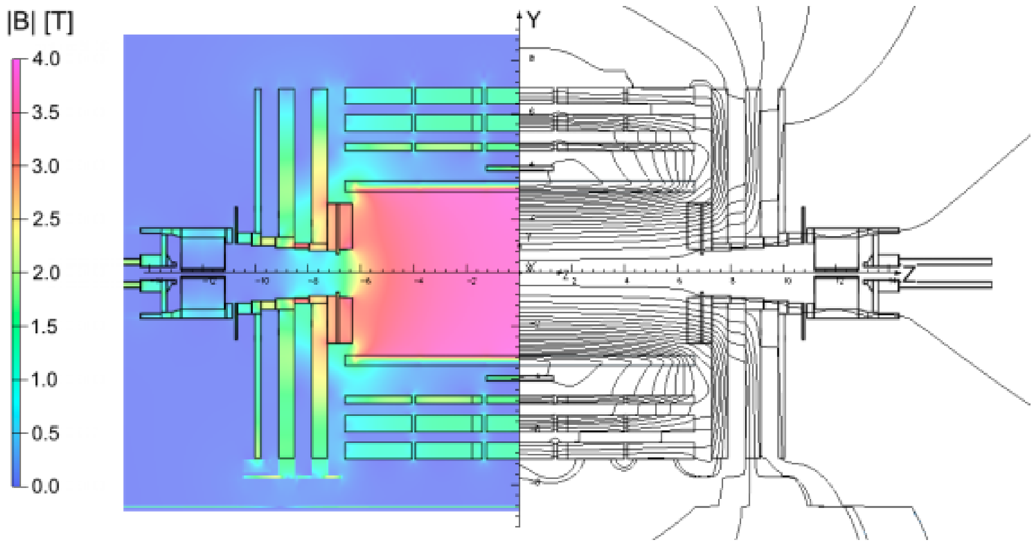


Figure 3.9: The magnetic field and field lines inside the CMS detector volume [78].

and resistive plate chambers (RPCs) in $|\eta| < 1.9$. The DTs are made up of drift cells filled with 85% Ar + 15% CO_2 gas mixture. The drift cells are operated at a high voltage, such that when a muon passes through a tube, it ionizes the gas and generates electrons that travel from the cathode to the anode. The drift time of the electrons can be converted to a position information for the muon track with a resolution of 260 μm [80].

The CSCs are filled with a 40% Ar + 50% CO_2 + 10% CF_4 gas mixture, and consist of arrays of positively-charged "anode" wires crossed with negatively-charged copper "cathode" strips. The anode wires are placed perpendicular to the cathode strips. They are operated at a very high voltage, such that when a muon passes through, it creates an ionization avalanche. Electrons move towards the anode wires and positive ions move towards the copper cathode. Since the strips and the wires are orthogonal to each other, we get a r and ϕ position for each passing particle. The closely spaced wires also make the CSCs fast detectors suitable for triggering.

Resistive plate chambers (RPC) consist of two parallel plates (one anode and one cathode), each made from a high resistivity plastic material and separated by a gas volume (95.2% Freon + 4.5% isobutane + 0.3% SF_6). The RPCs can provide a good time resolution ($\sim$ 1 ns) and a coarse position measurement on the muon hit, and can be used for muon triggering. The schematic layout of the muon chamber system is shown in Fig. 3.11.

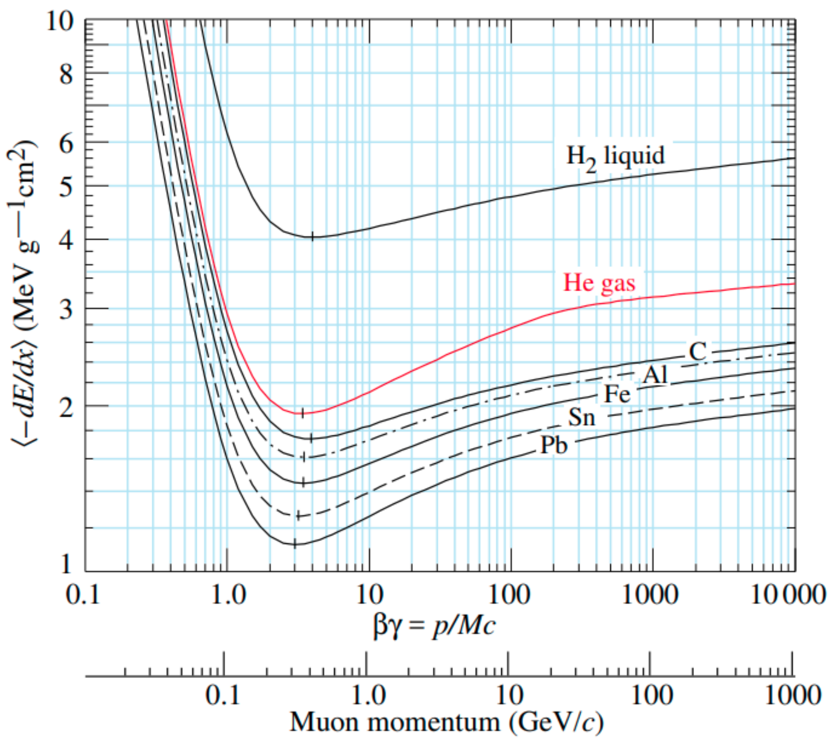


Figure 3.10: Mean energy loss of a muon traversing through different materials as a function of the muon p_{T} [79].

The muon reconstruction in CMS relies on the combined information from the tracker and the muon chambers. For muons with $p_{\mathrm{T}} < 200$ GeV, the tracker resolution dominates the muon p_{T} resolution and for muons in the energy range of 1 TeV, the resolution from the muon chambers dominates. The inner tracker + muon chambers provides a momentum resolution of $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 1\%$ (10%) at 50 GeV (1 TeV) [81].

3.2.6 Trigger and data acquisition

The collision rate of protons at the LHC is 40 MHz, and processing and keeping all those events is not possible due to bandwidth and offline storage constraints. The decision to keep or discard a event is made using a low-latency real time trigger system. The CMS trigger system has two levels: the Level-1 (L1) trigger and the High Level Trigger (HLT). The L1 trigger [82, 83] is a hardware based trigger, running on customized ASICs and FPGAs. It looks for coarsely segmented information from the calorimeters and muon stations, performs a basic physics object reconstruction (like high p_{T} or $p_{\mathrm{T}}^{\mathrm{miss}}$) and makes a positive or negative decision to keep the event. Due to limitations from the current electronics technology, the trigger decision needs to be made in < 4 μs, during which time the event data are

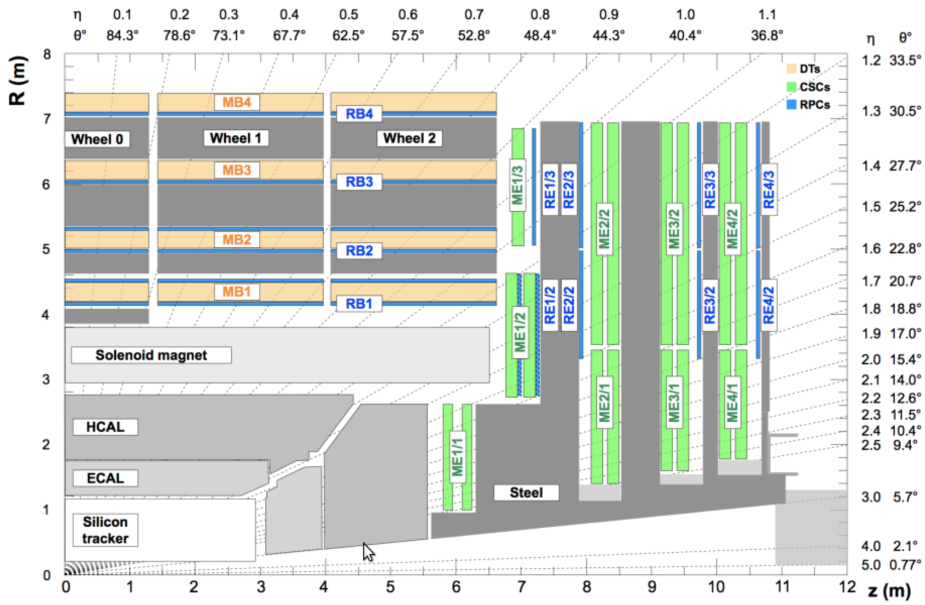


Figure 3.11: A schematic view of one quadrant of the CMS muon chamber. DTs are shown in orange and labeled as MB1/2/3/4; CSCs are shown in green and labeled as MEn/m, where n is the index in the z direction and m is the index in the *r* direction; RPCs are shown in blue and labeled as RB1/2/3/4 for barrel RPCs and REn/m for the endcaps. [81].

kept in a buffer. If a negative trigger decision is made, the event is lost forever. The L1 trigger reduces the input data rate from 40 MHz to 100 kHz.

Once an event is selected with the L1 trigger, it is sent to a CPU farm on the surface. The HLT [84] has access to information from all sub-detectors and does a fast and finer grained physics object reconstruction, including track reconstruction which is the most time consuming. There are several types of HLT algorithms, each one targeting a specific physics process and imposing requirements on multiple physics objects. The HLT reduces the data rate from 100 kHz to 500 Hz. The HLT paths are sometimes *prescaled* with a prescale factor N, meaning only one in every N events passing the HLT algorithm gets saved. This is done for HLT paths targeting physics processes with high event rates or for producing certain dedicated datasets that are only used for detector calibration studies. Additionally, there is a *data scouting* stream [85] for events that can not be stored due to trigger constraints. For such events, only a few reconstructed event quantities at the HLT level get stored permanently (no further offline reconstruction), amounting to a few kBs per event. The scouting stream is useful for CMS to pursue a larger physics phase space, for e.g., the study of B-meson decays.

Data selected by the HLT CPU farm are sent to the Tier-0 (T0) CERN computing center. An express reconstruction on about 10 % of the events is performed to monitor the offline data quality and derive detector alignment and calibration conditions. Once the *prompt* calibration conditions are derived, a full reconstruction on RAW data is performed (*prompt reconstruction*). The *prompt reconstruction* of a dataset needs to happen within 48 hours of data-taking, to avoid overflowing intermediate data storage buffers. The data can be produced and saved in different formats with different event sizes (based on the level of information per event), and is summarized in Table 3.1. From T0, the data are distributed to seven Tier-1 (T1) computing centers: Fermilab (United States), IN2P3 (France), PIC (Spain), ASGC (Taiwan), CLRC (United Kingdom), GridKa (Germany), and INFN (Italy). The T1 sites save a second copy of the RAW data and act as data links to Tier-2 (T2) sites in local regions. T2 sites are local computing centers at various universities, with substantial CPU resources for accessing data and performing physics analyses.

Table 3.1: CMS data formats for physics analyses [86].

Data format	Size (kB/event)
RECO	3000
AOD	400
MINIAOD	50
NANOAOD	1

3.2.7 Event reconstruction

To reconstruct particles in an event, CMS uses the concept of *particle-flow* (or PF) [87], in which information from different sub-detectors are combined and used to assign a trajectory and energy per particle. Fig. 3.12 shows a transverse slice of the CMS detector and the signature left by different types of particles in different sub-detectors.

The PF algorithm begins by reconstructing tracks in an event from all available hits from the tracker. The track reconstruction is performed in an iterative manner: the best quality tracks are first identified from hits in consecutive layers of the tracker and are removed from subsequent iterations. This process is repeated until all the hits are exhausted, and this method is known as the Kalman Filter algorithm [88]. The tracks reconstructed by the PF algorithm are used for primary vertex identification [89]. The vertex reconstruction algorithm clusters the tracks together based on their distance of closest approach to the beam spot, with a spatial resolution of ~ 20 (30)

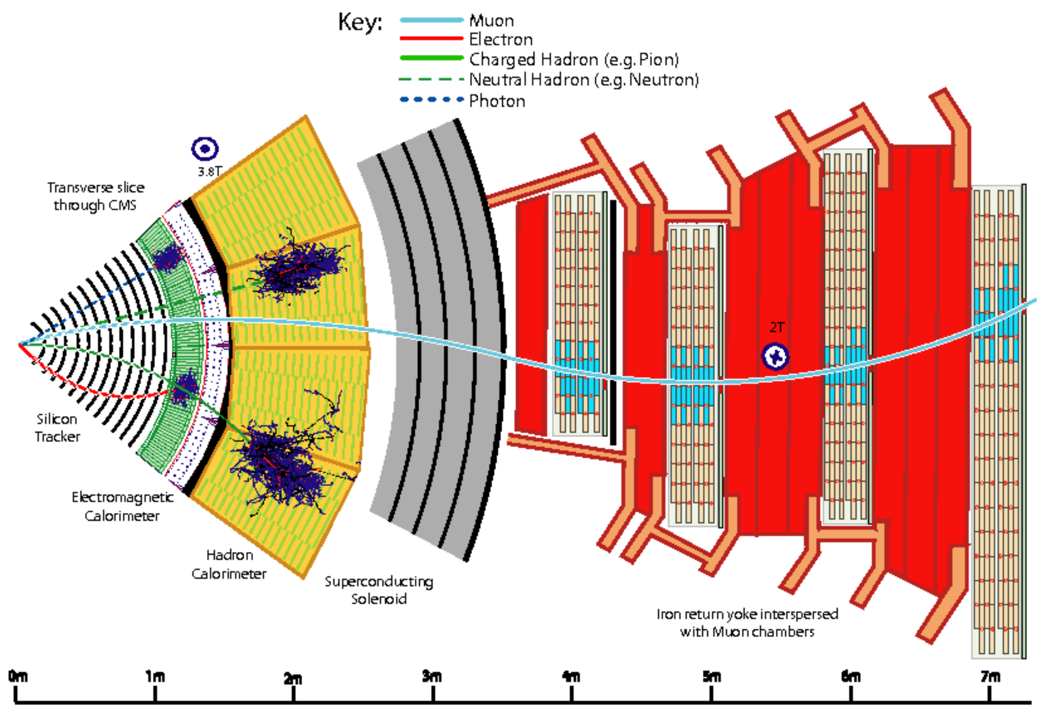


Figure 3.12: A transverse slice of the CMS detector indicating the different signatures left by various particles in the sub-systems [87].

μm in the x-y (z) plane. The precise identification of the primary vertex is crucial for pileup mitigation, as one can reject particles originating from secondary vertices in an event. Additionally, identification of displaced secondary vertices is crucial for reconstructing b-jets. Jets originating from b-quarks are used extensively in Chapter 6.

Electrons are reconstructed using reconstructed tracks from the tracker, and associating them to an energy cluster in the ECAL (based on interpolations of the $\eta - \phi$ of the track). The energy of the electron is determined based on the track momentum and ECAL cluster deposit, and the electric charge is determined by the bending direction of the track in the magnetic field. Similarly, charged hadrons are reconstructed by associating tracks together with ECAL (if available) and HCAL clusters. The neutral hadrons and photons do not leave any hits in the tracker and remain unaffected by the magnetic field. Therefore, a photon is reconstructed using an isolated ECAL cluster with no associated HCAL cluster. Neutral hadrons are reconstructed from HCAL clusters with or without an associated ECAL cluster.

Muons are reconstructed using hits from the tracker and associating them with hits in the muon stations. A combined track fit is performed to all the hits, and the muon

p_{T} is estimated from the global track curvature. Muons are used in this thesis in Chapter 5, and are discussed in greater detail in Sec. 5.3.2.

Quarks and gluons produced in an event undergo hadronization, and produce a cascade of particles known as jets. Jets in CMS are reconstructed using a clustering algorithm that combines information of several neutral and charged particle candidates in a conical structure, known as the anti-k$_T$ algorithm [90]. Jets are used in this thesis in Chapters 5 and 6.

Chapter 4

A MIP TIMING DETECTOR (MTD) FOR CMS AT HL-LHC

4.1 Introduction

The current LHC will undergo major design upgrades during 2026-2028, so that it can increase its instantaneous luminosity by a factor of ~3-4, beyond the LHC's design value. This will launch a new physics program known as the High-Luminosity LHC or HL-LHC, and will be operational starting in 2029. By the end of its first few years of operation at 13 TeV (at the end of 2018), the LHC has collected about ~150 fb^{-1} of data. The HL-LHC aims to collect about 3000 fb^{-1} of data by the end of its 10 year operation period. Figure 4.1 shows the schedule of the LHC program.

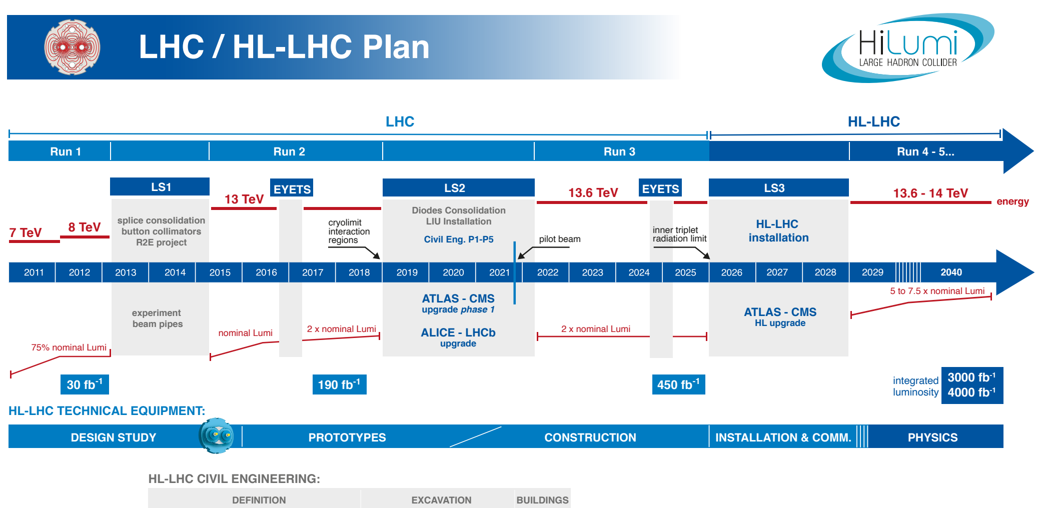


Figure 4.1: The LHC/HL-LHC physics program schedule [91].

The luminosity of the HL-LHC is projected to reach 5 to 7.5 $\times$ 10 ^{34}cm^{-2}s^{-1}, with a corresponding pileup of 140-200 interactions per bunch crossing. To operate in such a high radiation environment, all LHC experiments will undergo major upgrades (known as Phase-2 Upgrades). The CMS detector will also upgrade many of its existing sub-detectors, and install a new time-of-flight (TOF) sub-detector, known as the MIP Timing Detector (MTD) [92].

4.2 MIP Timing Detector (MTD)

The MTD will provide a track-time information for charged particles and is projected to have a timing resolution of 30-40 ps at the beginning of the HL-LHC, which will degrade to 50-60 ps by the end of the HL-LHC operations, because of radiation damage. The time resolution of the MTD is physics motivated. At the HL-LHC, there will be 140-200 pileup interactions with an RMS spread of 180-200 ps per bunch crossing. If one slices a bunch crossing in the temporal phase space with a 30-40 ps resolution, each snapshot in time has a reduced pileup of about ~50 vertices, which is similar to the current LHC levels. Thus, the MTD will help maintain the current levels of particle reconstruction efficiency at the Phase-2 CMS detector.

Figure 4.2 shows a simplified implementation in `GEANT4` [93] of the proposed layout of the MTD in the CMS detector. The MTD is made of a barrel and an endcap region, with different technologies based on different requirements, for e.g., radiation dose, schedule constraints, cost effectiveness, etc. The barrel region of the MTD (described in more detail in Sec. 4.2.1), will be placed in between the tracker and the ECAL and occupy the region where $|\eta| < 1.5$. Since there is less time for R&D of the barrel timing layer (BTL), one needed a mature and established technology, and Silicon Photomultipliers (SiPMs) with crystal scintillators [94–96] were chosen as the photosensors. At an integrated luminosity of 3 ab^{-1} (full lifetime of HL-LHC), the BTL will have received radiation doses of up to 1.9×10^{14} n_{eq}/cm^2 (1 MeV neutron equivalent fluence) or 30 kGy and the endcap timing layer (ETL) will have received doses of up to 1.6×10^{15} n_{eq}/cm^2 or 450 kGy. Due to the higher radiation rates, SiPMS are not a viable technology for the ETL ($1.6 < |\eta| < 3.0$), and the best performance is achieved with low gain avalanche diodes or LGADs [97–99], which are silicon sensors with internal gain of about 10–30. The ETL will be placed as a disk shaped sub-detector, before the calorimeter endcaps, at a distance of 3m from the interaction point (along the beam axis). This thesis will however only focus on the BTL and more details on the ETL can found be in Ref. [92].

Extensive studies have been performed that have looked at the impact of TOF information from MTD, including the reconstruction of physics objects (like b-jets, missing transverse momentum, etc.) as well as the overall effect on the sensitivity of various physics analyses (Ref. [92]). One special physics case is the production of a pair of Higgs bosons decaying to a 4-bottom quark ($HH \to b\bar{b}b\bar{b}$) final state. The CMS Run 2 search for this final state when both Higgs bosons are highly Lorentz

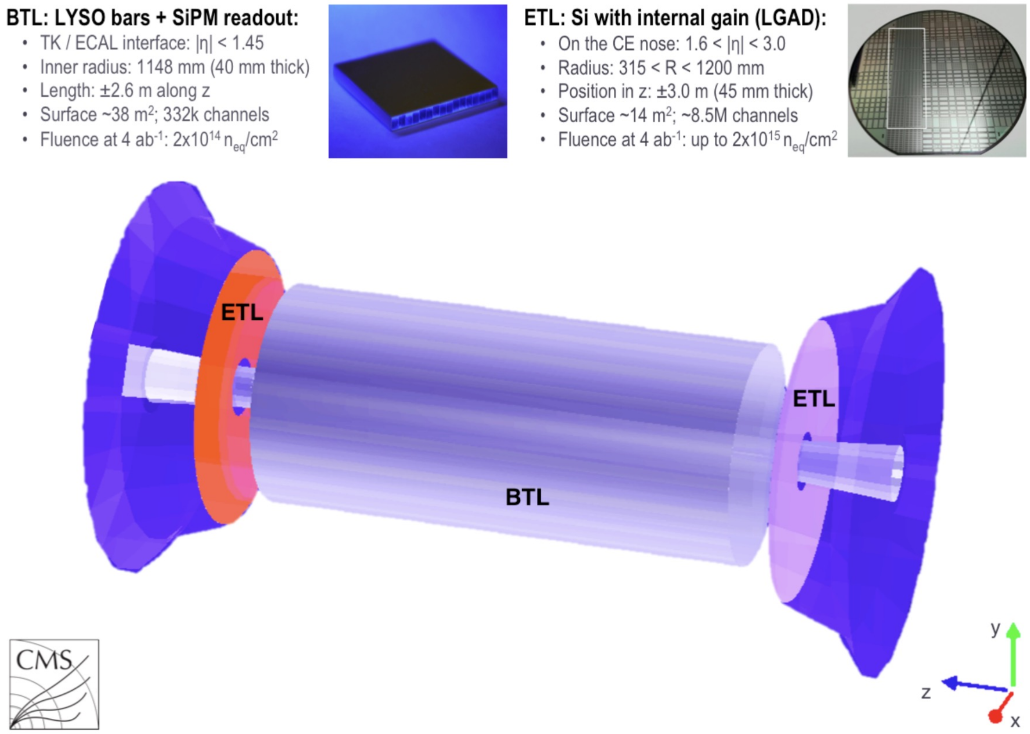


Figure 4.2: A schematic view of the GEANT geometry of the MTD, comprising a barrel layer (grey cylinder), at the interface between the tracker and the ECAL, and two silicon endcap (orange and light violet discs) timing layers in front of the endcap calorimeter [92].

boosted will be discussed extensively in Chapter 6. Studying the HH production and measuring the Higgs trilinear self-coupling (λ_{HHH}) is one of the main physics goals of the HL-LHC. It is projected that for $HH \rightarrow b\bar{b}b\bar{b}$, the signal yield increases by 14% due to the BTL alone, and 18% from the combined power of BTL and ETL, at a constant background rate, as shown in Fig. 4.3.

4.2.1 Barrel Timing Layer (BTL)

The barrel timing layer will cover the pseudorapidity region up to $|\eta| = 1.48$ with a total active surface of about 38 m^2. The sensor unit consists of Lutetium Yttrium Orthosilicate crystals doped with Cerium ($(\mathrm{Lu}_{1-x}\mathrm{Y}_x)_2\mathrm{SiO}_5$:Ce), abbreviated as LYSO:Ce, and read out with SiPMs. LYSO:Ce crystals were chosen because of their high density (7.1 g/cm^3), high light yield (~40000 photons/MeV), fast scintillation rise time (< 100 ps), short decay time (~ 43 ns), very high radiation tolerance (less than few percent loss in transparency over the 10 year HL-LHC operations) and also, LYSO:Ce crystals produce scintillation light at 420 nm wavelength which matches the optical range of SiPMs. SiPMs are widely used in particle physics

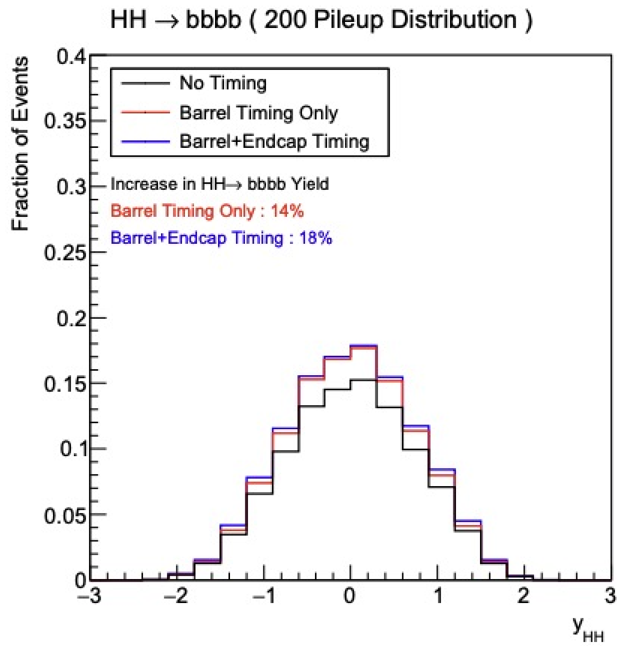


Figure 4.3: Projections for yield enhancement in $HH \to b\bar{b}b\bar{b}$ as a function of the Higgs boson rapidity, for a 200 pile-up scenario at the HL-LHC. The distributions are normalized to the no-timing case [92].

based experiments, are compact, not affected by magnetic fields, very robust (can be exposed to room light without damage), operate at relatively low voltages (30-77 V), and have a low power consumption. They also have a high photo-detection efficiency, PDE, of 20-40% in devices with small cell size (15 μm square pixels). In the BTL, the SiPMs will operate above their breakdown voltage in Geiger mode with a gain of the order of 10^5. The LYSO:Ce crystals have a bar-like geometry with 57 mm length and 3.12 mm width. The radial thickness is varied along η to ensure the slant depth crossed by particles coming from the interaction point remains constant, irrespective of where it hits the BTL. The slant depth varies as 3.7 mm for $|\eta| < 0.7$, 3.0 mm for $0.7 \leq |\eta| \leq 1.1$ and 2.4 mm for $|\eta| > 1.1$. Each bar will be coupled to two SiPMs (at the two ends), whose dimensions will be 3 mm along ϕ and a variable thickness, equal to the bar's radial thickness in each $|\eta|$ interval.

The time resolution per track, measured from two SiPMs at the two ends of a crystal, depends on the following uncertainties added in quadrature:

- CMS clock distribution: 15 ps
- Digitization: 7 ps
- Electronics: 8 ps

- Photo-statistics: 25–30 ps
- Noise (SiPM dark counts): negligible at startup, 50 ps after 3000 fb^{-1}

The electronics and digitization jitter have negligible impact on the overall time resolution, and the key factors affecting the time resolution are the photostatistics and the Dark Count Rate (DCR). The photostatistics term is described by

$$\sigma_t^{phot} = \sqrt{\frac{\tau_r \tau_d}{N_{phe}}} \propto \sqrt{\frac{\tau_r \tau_d}{E_{\text{dep}} * LY * LCE * PDE}} \tag{4.1}$$

where τ_r (100 ps) and τ_d (43 ns) are the rise time and decay time of the scintillation pulse in the LYSO:Ce crystal, respectively. The energy deposited by a MIP in a thin LYSO:Ce crystal, E_{dep}, follows a Landau distribution with the most probable value (MPV) of 0.86 MeV/mm. The number of photoelectrons, N_{phe}, scales linearly with E_{dep}, the crystal light yield (LY) as determined by the crystal thickness and scintillation properties, the light collection efficiency (LCE), i.e., the probability that a photon reaches the SiPM without escaping from lateral faces of the crystal or being absorbed within the material, and with the PDE of the SiPM. These parameters have driven the optimization of the sensor layout (crystal and SiPM configuration), as discussed later in Sec. 4.3. The noise contribution from the DCR in the SiPM scales proportionally to $\sqrt{DCR/N_{phe}}$. The magnitude of the DCR increases with integrated luminosity due to radiation damage creating defects in the silicon, increasing the probability of generating thermal electrons [100].

The BTL will be placed in the Tracker Support Tube (TST) and will share CO_2 cooling with the tracker. Figure 4.4 shows the layout of the BTL inside the TST. 72 trays consisting of LYSO:Ce+SiPM modules will be inserted into the TST, with 2 trays per ϕ interval (10°). Each tray will consist of 6 Readout Units (RU). Each RU (see left of Fig. 4.6) has 12 detector modules, which are made of a copper housing and two BTL modules. Each BTL module (see right of Fig. 4.6) consists of an array of 16 LYSO:Ce bars with 32 SiPMs. The crystals will be wrapped with a reflective material, Enhanced Specular Reflector (ESR) Vikuiti by 3M, between adjacent channels to provide optical isolation. This reflector is chosen because it has a reflectivity for 420 nm light higher than 98.5%, and is sufficiently radiation tolerant. The BTL modules will be cooled to –35 °C with CO_2 cooling and an additional –10 °C with thermo-electric coolers (TECs), to limit SiPM self-heating and DCR. Additionally, annealing the irradiated SiPMs at room temperature is known to mitigate the DCR [100]. Annealing of the BTL modules at room temperature will be scheduled during the yearly shutdowns, to help the SiPMs recover from the radiation induced defects. Fig. 4.5 shows the DCR rate as a function of time (increasing integrated luminosity), in different annealing scenarios [92]. Exploiting the full shutdown period ($\sim$ 4 months) for recovery at room temperature provides a 30% reduction of DCR with respect to a two weeks only annealing scenario.

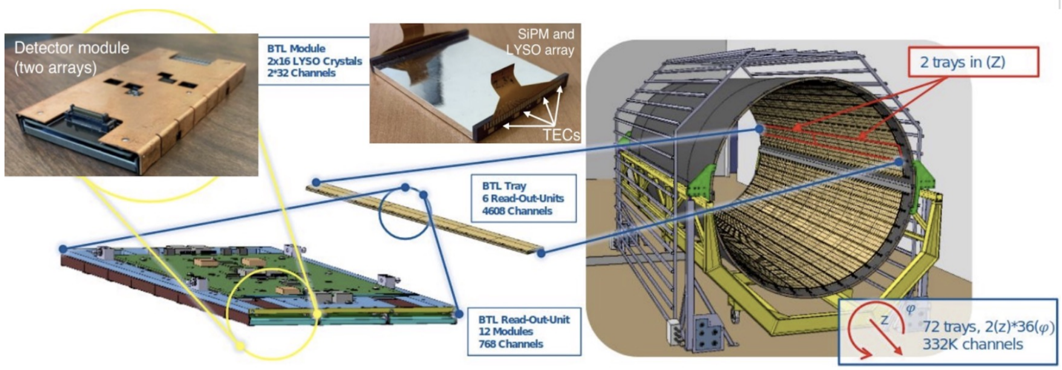


Figure 4.4: A schematic view of the structure and design of BTL [101].

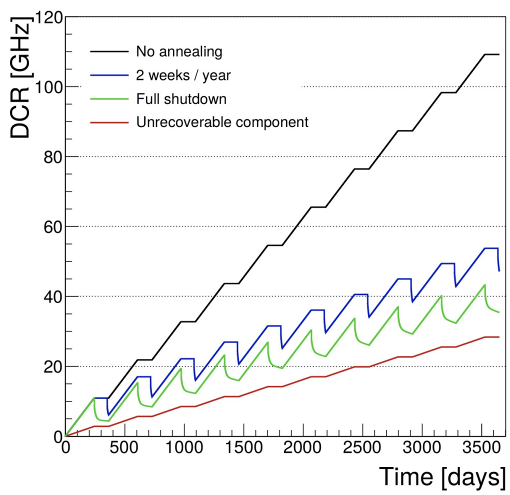


Figure 4.5: Expected growth of DCR for various annealing scenarios at fixed OV of 1.5 V during the detector lifetime [92].

Each RU will consist of 4 front-end (FE) boards. Each FE board will consist of 6 ASICs, which are connected to their SiPMs via flex cables. The 4 FE boards plug into a Concentrator Card (CC), that provides low voltage power, bias voltage, and three low power Gigabit Transceivers (lpGBTs) that carry data and control signals to and from the ASIC.

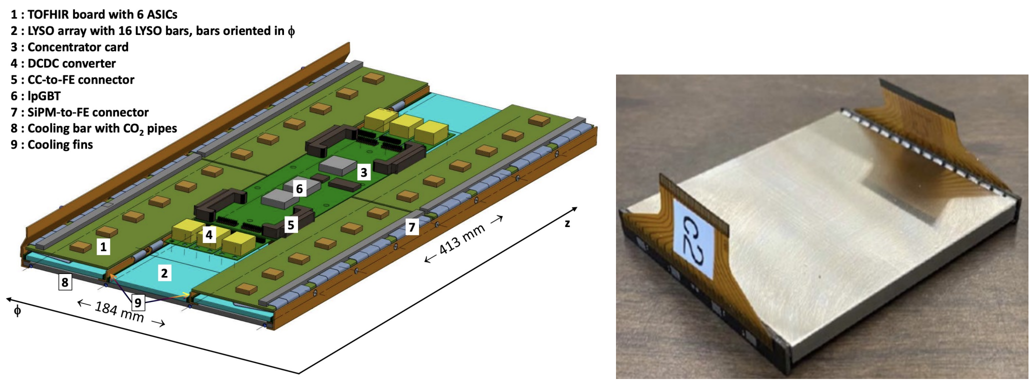


Figure 4.6: Various BTL components. (Left) A BTL Readout unit, consisting of 24 BTL modules of LYSO:Ce bars + SiPMs. Each RU has a total of 768 SiPMs. (Right) A BTL module consisting of 16 LYSO:Ce bars + 32 SiPMs. The module is connected to the FE boards using flex cables, as can be seen in the figure.

The modules will be read out with an ASIC, known as the TOFHIR (Time-of-flight, High Input Rate) chip (Ref. [102]). The TOFHIR chip is being designed to deliver precision timing information at the required rates for the BTL, and will be radiation tolerant, will satisfy the power requirements and will be able to operate at the low BTL temperatures (-35 °C). It will also have a dedicated noise cancelling circuit to mitigate the effect of DCR noise in the SiPMs. Appendix B discusses the commissioning tests performed on the two preliminary versions of the BTL prototype readout electronics, i.e., the TOFPET and the TOFHIR2A ASIC.

4.3 Studies to optimize BTL design parameters

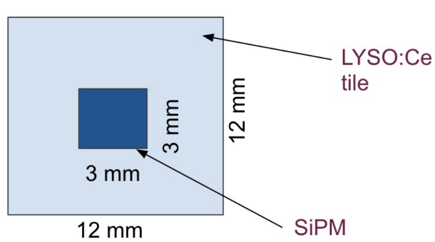
The next few sections will discuss several studies that were performed to optimize different aspects of the BTL design and performance. These tests were carried out at the Caltech Precision Timing (CPT) Lab or at testbeams that were held at CERN and Fermilab.

4.3.1 Characterization of the sensor properties and optimizing the design geometry

The following sections will discuss the considerations made for the sensor geometry optimization, effect of wrapping the crystals with a reflector material and the light output per MIP in the SiPMs. These tests were carried out at the CPT Lab at Caltech.

4.3.1.1 Impact point studies with laser

Before the bar shaped geometry was selected for the LYSO crystals, there was some interest in exploring a square tile shaped geometry for the crystals (Ref. [103]). To understand the properties of such a geometry, we used 3×3 mm^2 SiPMs from Hamamatsu and coupled them to 12×12×4 mm^3 LYSO square tiles using optical grease (see left of Fig. 4.7). We used a 373 nm UV laser to shine light directly into the the sensor configuration (with the light entering the LYSO crystal first, before the SiPM) , and recorded the output using a DRS4 digitizer. The laser pointer was moved slowly from one edge to the other edge of the tile configuration. Fig. 4.7 (Right) shows the amplitude dependence on the MIP impact point in this setup. It can be seen that in such a geometry, the centre of the tile (at the SiPM) records a 30% higher amplitude as compared to the edges, which can be understood from the smaller distance travelled by the scintillation light produced at the tile center, before reaching the SiPM.



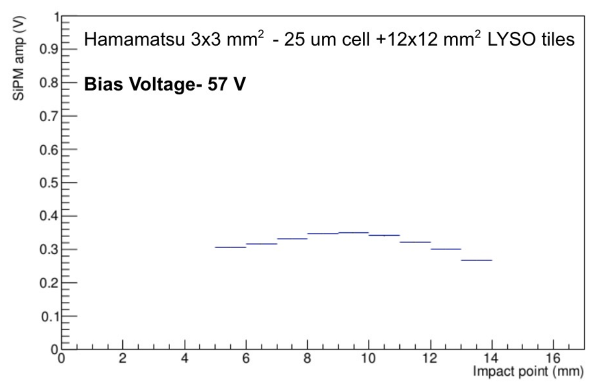


Figure 4.7: Impact point studies with a laser. (Left) A schematic of a 3x3 mm^2 SiPM coupled with a 12x12x4 mm^3 LYSO tile using optical grease. (Right) Amplitude recorded as a function of the MIP impact position on the SiPM +LYSO tile configuration.

This non-uniformity in amplitude also directly impacts the measured time resolution. The time resolution measured at the centre of the tile is slightly better than the time resolution recorded at the edges, as can be seen in Fig. 4.8. To avoid this impact point dependence, the bar geometry was chosen for the LYSO:Ce crystals, with two SiPMs at the two edges.

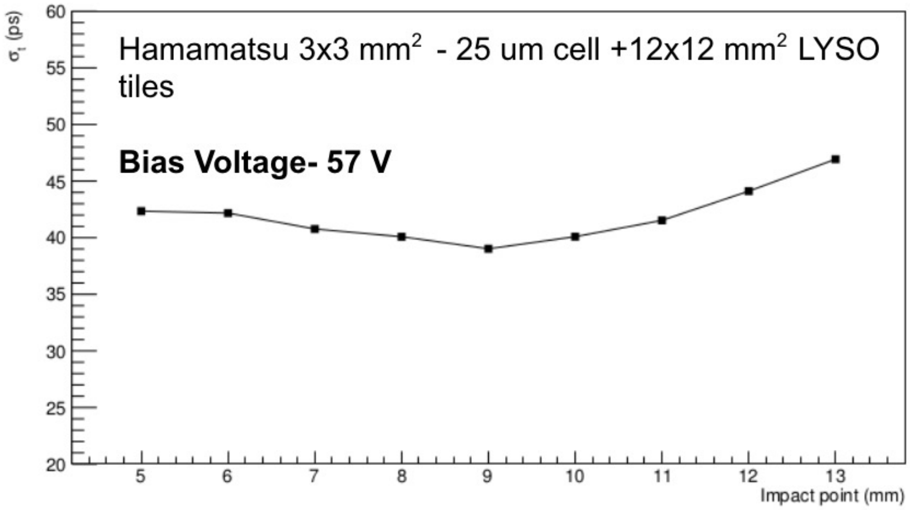


Figure 4.8: The measured time resolution as a function of the MIP impact position on the SiPM +LYSO tile configuration. The time resolution measurement here includes a time walk correction.

4.3.1.2 Effect of wrapping the crystal on light collection efficiency

Before ESR was chosen as the reflector material for wrapping the crystals, we performed tests of the light collection efficiency by studying the difference between wrapped and unwrapped LYSO square tiles (not bars). We used 3×3 mm^2 SiPMs from Hamamatsu and coupled them to 12×12×4 mm^3 LYSO square tiles using optical grease for this study. We used a Keysight S-series scope for data acquisition, with a 6 GHz bandwidth and sampling rate of 20 Gsamples/second. Some specifications provided by the manufacturer for the LYSO tile are quoted below.

- Gain: 7×10^5
- Light yield: 33200 photons/MeV
- Photon detection efficiency (PDE): 25 %

Most photons hitting the LYSO will never reach the SiPM because of the low photon detection efficiency and also because the SiPM active area is much smaller than the LYSO tile area. The idea behind wrapping the crystals is to increase the probability of the photons reaching the SiPM, such that the scintillation light produced in the LYSO bounces back and forth several times before it exits the tile. To demonstrate the power of wrapping on the light collection efficiency (LCE), we placed a ^{22}Na

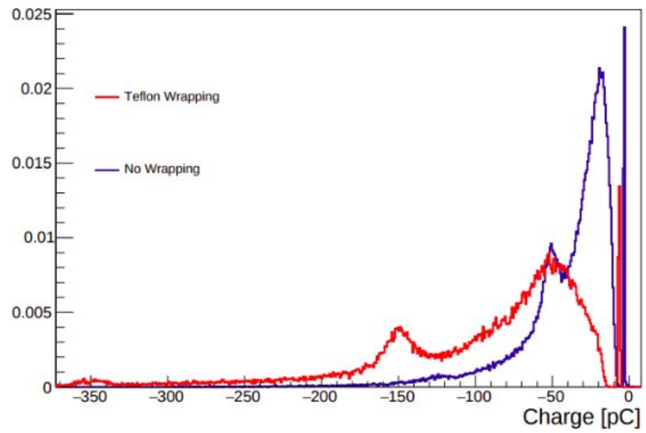


Figure 4.9: ^{22}Na spectrum as recorded from the charge deposited in the SiPM in two different cases: bare configuration and wrapped configuration.

source in front of the SiPM+LYSO configuration and recorded the source spectrum in the following two cases:

- no wrapping
- SiPM + LYSO wrapped in a reflective coating of Teflon tape .

The oscilloscope can be used to measure the ^{22}Na spectrum in a Volts vs time graph. Using the known values of the circuit board amplifier (10) and the overall circuit resistance (50 Ohms), we can take an integral of the recorded pulse and divide it by the circuit impedance to obtain a charge (in pico-Coulombs) vs counts distribution. The charge deposited by a MIP can be written as

$$\text{Charge} = \text{Gain} \times \text{LYSO light yield} \times \text{PDE} \times \text{LCE} \times e \tag{4.2}$$

where e is the charge of an electron. In an ideal case with LCE = 1, the charge deposited would be 475 pC for the 511 keV transition. From Fig. 4.9, the charge deposited corresponding to the 511 keV peak is ≈149 pC with the teflon wrapping and ≈50 pC with no wrapping. From this we can conclude that the LCE is about 3 times bigger in teflon wrapped LYSO+SiPM when compared to the unwrapped case. On using Eqn. 4.2, we can caluclate 31% LCE for the teflon wrapped configuration and only 10.1% LCE for the bare configuration.

The final decision to use ESR instead of Teflon was made based on the poor radiation tolerance of Teflon.

4.3.1.3 Calibration of the number of photoelectrons deposited by a MIP

To measure the number of photoelectrons generated in the sensors per MIP, we need to first calibrate the energy per single photoelectron. To study this, we used a 3×3 mm^2 bare SiPM (no LYSO tile), wrapped it in Teflon (with a tiny gap in the wrapping at the center) and mounted it on a circuit board. We used a Keysight S-series scope for data acquisition, with a 6 GHz bandwidth and sampling rate of 20 Gsamples/second. We recorded the integrated charge spectrum by shining a laser light directly into the SiPM (see Fig. 4.10). For this, we placed an optical filter in front of the SiPM with ND = 2, where ND is defined as

$$ND = (\frac{I}{I_0})^{-x}, \tag{4.3}$$

where x is the value of the ND filter (in this case ND = 2). I_0 and I correspond to the intensities of the light before and after passing the filter. The laser light was also manually tuned to the lowest value where we could record a clean spectrum.

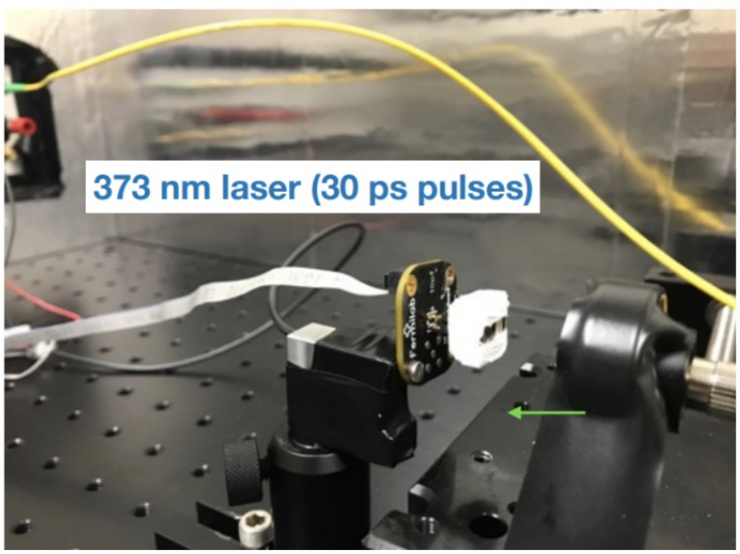


Figure 4.10: The SiPM is wrapped with Teflon and mounted on circuit board. A 373 nm UV laser is directly pointed at the center of the SiPM.

The charge spectrum is shown on the left in Fig. 4.11. The first peak from the right is coming from the noise and trigger turn on. The second, third and fourth peak from the right correspond to one, two and three photo-electrons respectively. The peaks were then fitted with a Gaussian distribution and the integral is plotted against the number of photo-electrons, $N_{\text{p.e.}}$ (with appropriate normalization), as seen on the right in Fig 4.11. We obtain a Poisson distribution with mean 1 and a p-value of 0.67.

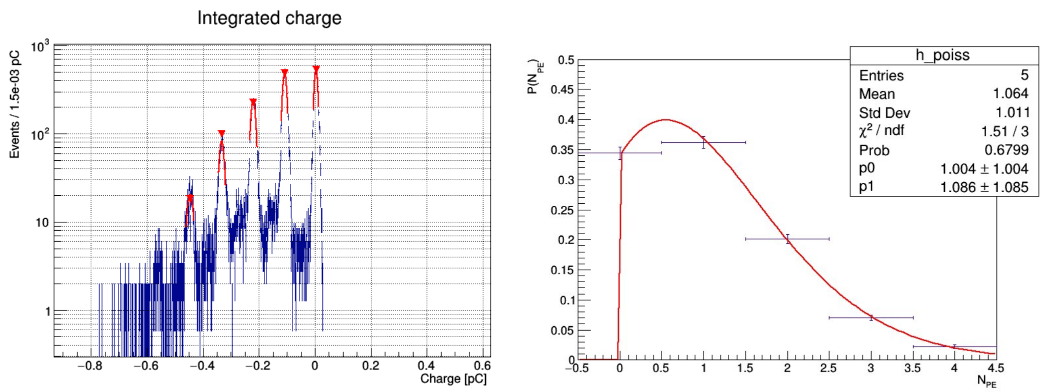


Figure 4.11: Single photoelectron measurements with a UV laser. (Left) The recorded charge spectrum. The different peaks correspond to different values of $N_{\mathrm{p.e.}}$. (Right) Poisson fit to the number of photo-electrons.

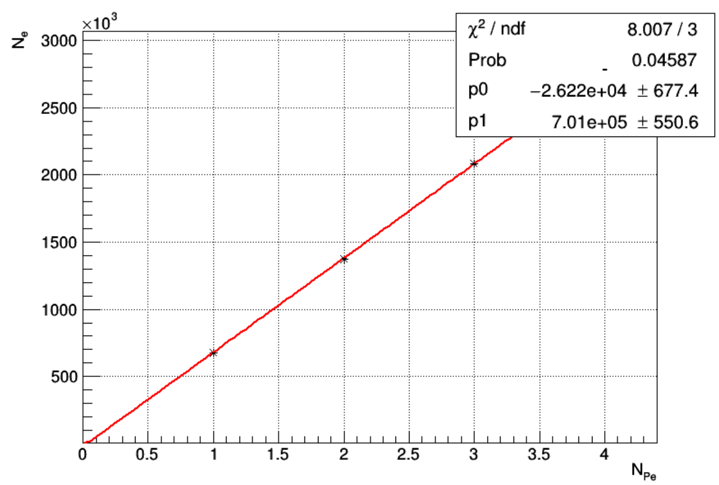


Figure 4.12: The number of electrons plotted against the number of photoelectrons.

We then plotted the the number of electrons (calculated from the integrated charge or area under the curve) from each peak against the number of photo-electrons (peak number), as seen in Fig. 4.12. The error on the number of photo-electrons is 0 and the error on the number of electrons is extracted from the error on the mean parameter from the guassian likelihood fit. We measure a gain of $7 \mathrm{x} 10^5 \pm 550.6$, which is the number quoted by the manufacturer (see Sec. 4.3.1.2).

We also recorded charge histograms by shining laser light into the SiPM through various optical filters. The optical filters used were : No Filter (ND = 0.0), ND = 0.1, ND = 0.2, ND = 0.3, ND = 0.4, and ND = 0.5. These charge histograms were fitted with a Gaussian to give the mean charge deposit. These were then plotted against the ND value (or normalized intensity) as shown in Fig 4.13. The error on the mean

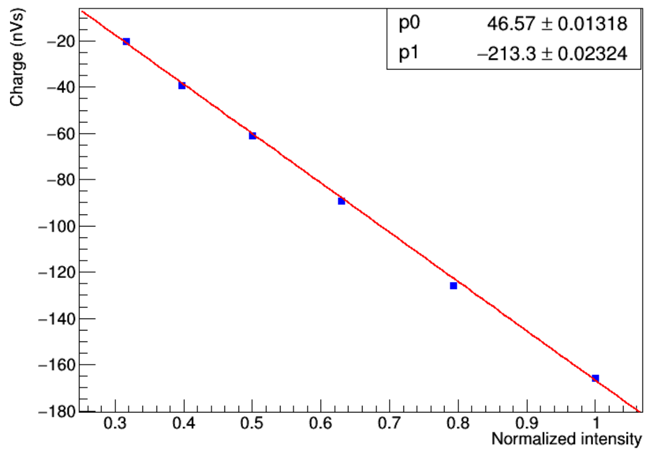


Figure 4.13: The charge deposited in a SiPM scales linearly with the intensity of the laser light.

charge is extracted from the mean parameter from the gaussian likelihood fit. Since the ND value (or normalized intensity) is proportional to the energy deposited, we can confirm that the energy deposited in the SiPM scales linearly with the charge (until the SiPM is saturated).

Once we know the relationship between the number of photo-electrons and the number of electrons (Fig. 4.12) and also the fact that the energy deposit scales linearly with integrated charge in a SiPM, we can determine the number of photo-electrons in a MIP energy deposit. We coupled the 3x3 mm^2 SiPM with a 12x12x4 mm^3 LYSO tile using optical grease, wrapped the configuration with teflon and mounted it on a circuit board. The SiPM was biased at 57 V. We then recorded the spectrum of ^{22}Na. In Fig. 4.14, the two beta decay transitions of the ^{22}Na source (511 keV and 1.27 MeV) are visible (2nd and 3rd peaks from the right, respectively). The 1st peak is interpreted as a convolution of the lower compton edge, noise and trigger turn on. On fitting the peaks with a gaussian, the following results are obtained:

- For the 511 keV peak, the deposited charge is 85.5 pC which corresponds to 535M electrons. Thus the number of photoelectrons is estimated to be ≈ 763 (from Fig. 4.12).
- For the 1.27 MeV peak, the deposited charge is observed to be 187.7 pC which is roughly 1173M electrons. This corresponds to 1674 photoelectrons (from Fig. 4.12).

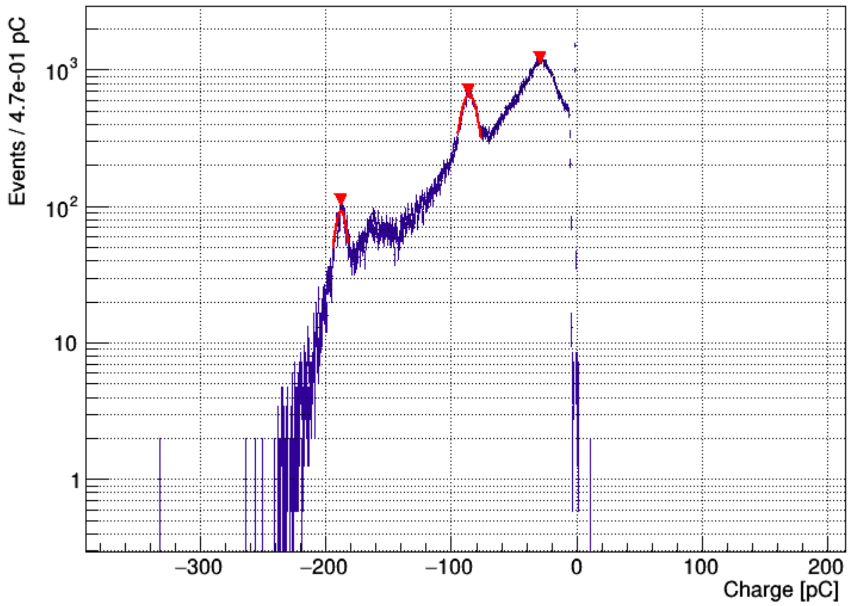


Figure 4.14: The ^{22}Na spectrum recorded by the 3x3 mm^2 SiPM coupled with a 12x12x4 mm^3 LYSO tile using optical grease and wrapped in teflon.

The ratio of the deposited charge in the two peaks, 85.5/187.7 = 0.45 is within 10% of the expected energy ratio of the two peaks, i.e., 0.51/1.27 = 0.40 (systematics not estimated). If we take the Compton edge of the 1.27 MeV photon to be around 160 pC from Fig. 4.14, the ratio of the charge in the compton edge with the 1.27 MeV peak is 160/190 = 0.842, which is less than 1% away from the expected value of 1068/1270 = 0.841. As we already know about the linear relationship between the energy deposited in the SiPM and the integrated charge (Fig. 4.13), we can construct a simple relationship from the energy of the two peaks and the 1.068 MeV compton edge, given by

$$N_{p.e.} = 1200.3 \times E[MeV] + 149.3. \tag{4.4}$$

This procedure was repeated several times, unwrapping and wrapping the tile again (leaving everything else untouched) in order to study the systematic effects produced by the stability of our wrapping procedure. The results obtained did not show any qualitative difference, with the variation in the charge deposited being less than 2%. Thus, we can conclude that for a 3 MeV MIP, the number of photoelectrons generated is $\approx$ 3750. For a 4 MeV MIP, the number of photoelectrons generated is $\approx$ 5000.

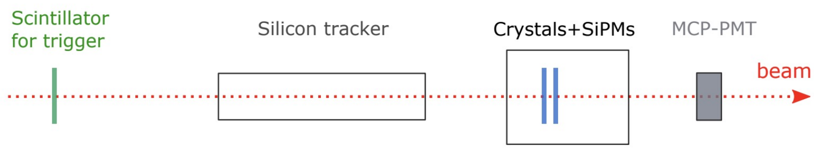


Figure 4.15: The test beam setup at Fermilab. From left to right, the scintillator is used for trigger, the silicon tracker is defines the MIP impact position in the x-y plane, and the MCP-PMT is used to provide a reference time. The LYSO+SiPMs test setups, one for the 1-bar and the other for the 3-bar array, are positioned along the beamline. The beam direction is the z-axis.

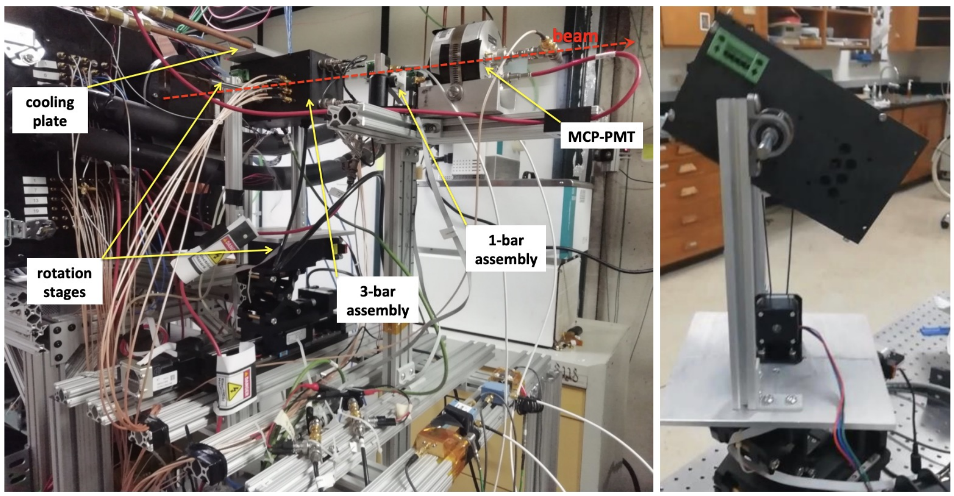


Figure 4.16: Fermilab April 2019 test beam. (Left) The experimental setup at FTBF. (Right) The box and mechanical structure capable of rotating the 3-bar assembly.

4.3.2 Time resolution measurements at the Fermilab Test Beam

The studies in this section (see Ref. [104]) were performed at the Fermilab Test Beam Facility (FTBF), using 120 GeV protons, irradiating single LYSO:Ce bars read out by SiPMs. The experimental setup is shown in Figure 4.15. The trigger was based on a 10 cm^2 scintillation counter located a few meters upstream with respect to the test setup. A silicon tracker telescope, consisting of twelve strip modules with 60 μm pitch in alternating orientation, was located in front of the crystals and SiPMs under test, to determine the impact point of the beam particles in the x-y plane with a precision of about 0.2 mm. The crystals and SiPMs were located inside a temperature controlled (25±1°C) dark box. The box was mounted on a support structure capable of rotating the sensors with respect to the beam direction (Fig. 4.16). A Photek 240 Micro Channel Plate Photomultiplier Tube (MCP-PMT), with a time resolution of 12 ps, was placed downstream of the box and was used to provide a reference time measurement.

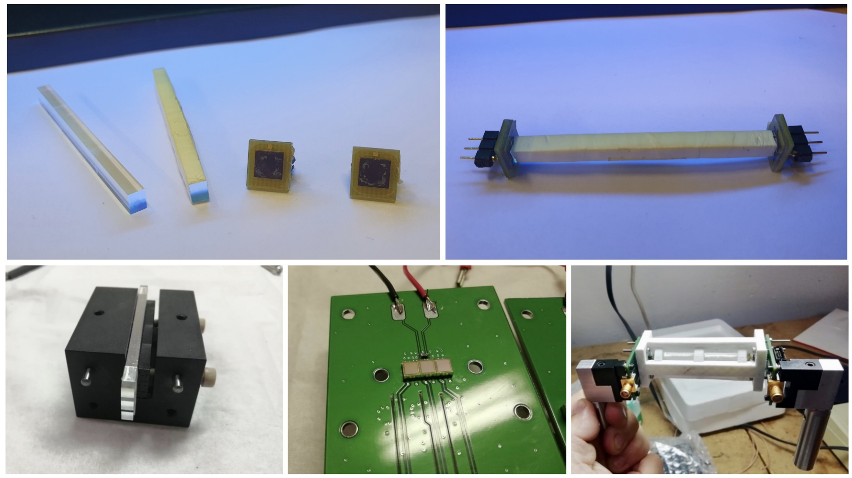


Figure 4.17: Sensors used at the Fermilab April 2019 test beam. (Top) Naked and wrapped individual crystal bars with the two FBK SiPMs (left) and single wrapped bar glued to SiPMs (right). (Bottom) Three-bar assembly in a crystal holder with the screws used to adjust the alignment of the crystals to the SiPMs (left); HPK SiPMs soldered onto a readout board (middle); single bar assembly (right).

The crystal bars of LYSO:Ce used in this test were manufactured by Crystal Photonics Inc. (CPI) in three different geometries of 3×t×57 mm^3, where the thickness, t, varies between 2, 3 and 4 mm. Two different SiPM types were tested, the first type belongs to a set of S12572-015 SiPMs from Hamamatsu (HPK) with an active area of 3×3 mm^2, while the second set was provided by Fondazione Bruno Kessler (FBK), with an active area of 5×5 mm^2 based on the NUV-HD-ThinEpi technology. The HPK and FBK SiPMs provide gains of 1.8×10^5 and 2.5×10^5, respectively, and the PDE weighted by the emission spectra of LYSO:Ce is similar for the two SiPMs and reaches about 36% for 6V over-voltage. The crystal bars were wrapped in Teflon ($\sim$100 μm thick) to improve light collection. Crystal bars coupled to HPK SiPMs were assembled in a system with a three-bar holder, with only one thickness t = 3mm, where the bars are placed parallel to each other and can be rotated with respect to the beam direction. Bars read out by FBK SiPMs were set up with a one bar holder, with the bar thickness varying as t = 2, 3 and 4 mm. The crystals and SiPMs can be seen in Fig. 4.17.

Customized electronic boards were used to apply the bias and perform the readout of the SiPMs, and the full pulse shapes were digitized with a sampling frequency of about 5.12 GSample/s. In the 3-bar setup, each SiPM signal was amplified with a Gali74+ low noise amplifier, filtered with a 500 MHz low-pass filter and then split into two paths. One of the two signals was further amplified by about 44 dB using two Gali52+ amplifiers in cascade to provide a saturated waveform with a steep rising edge, used for extracting the time of arrival using low threshold leading edge discrimination, while the other unsaturated signal was used to measure the deposited energy. For the single bar readout, each SiPM signal was amplified with Gali74+ amplifier and filtered by a 500 MHz low-pass filter. The signal was split and one output was read out directly to measure the signal amplitude while the second output was further amplified by a second stage Hamamatsu C5594 amplifier with a gain of 36 dB and used to measure the MIP arrival time. A CAEN V1742 digitizer [105] hosted in a VME crate was used for the readout of all the waveforms: two readout channels for each SiPM under test and one for the MCP-PMT used as time reference. The digitizer was triggered by TTL-level signals originating from the trigger counter.

The time resolution of the setup can then be defined in a few different ways. We define the following quantities based on the the times of arrival measured at the two SiPMs at the ends (using the high gain channel), t_{left} and t_{right} , and the time measured by the MCP-PMT, t_{MCP}:

- $\Delta t_{\text{bar}} = t_{\text{average}} - t_{\text{MCP}} = (t_{\text{left}} + t_{\text{right}})/2 - t_{\text{MCP}}$
- $t_{\text{diff}} = t_{\text{left}} - t_{\text{right}}$
- $\Delta t_{\text{left}} = t_{\text{left}} - t_{\text{MCP}}$
- $\Delta t_{\text{right}} = t_{\text{right}} - t_{\text{MCP}}$.

Since we use a fixed threshold at the leading edge (of the high gain channel) to measure the time of arrival, the t_{left} (t_{right}) depends on the amplitude of the signal. This variation ($\sim$ few hundreds of ps) in the time of arrival needs to be corrected to achieve the optimal time resolution, and the correction is derived by studying the dependence of $t_{\text{SiPM}} - t_{\text{MCP}}$ on the amplitude of the SiPM signal, and is known as "amplitude-walk correction". Position based corrections are also derived by measuring the dependence of $t_{left} + t_{right}$ on the impinging position of the MIP on the bar, to achieve a uniformity in the time resolution across the bar irrespective of the MIP impact position. A Gaussian fit is then performed to the distributions of Δt_{bar}, Δt_{diff}, Δt_{left} and Δt_{right}, and the time resolution of the bar is then obtained as:

- $\sigma_{t_{\text{average}}} = \sqrt{\sigma^2_{\Delta t_{\text{bar}}} - \sigma^2_{t_{\text{MCP}}}}$

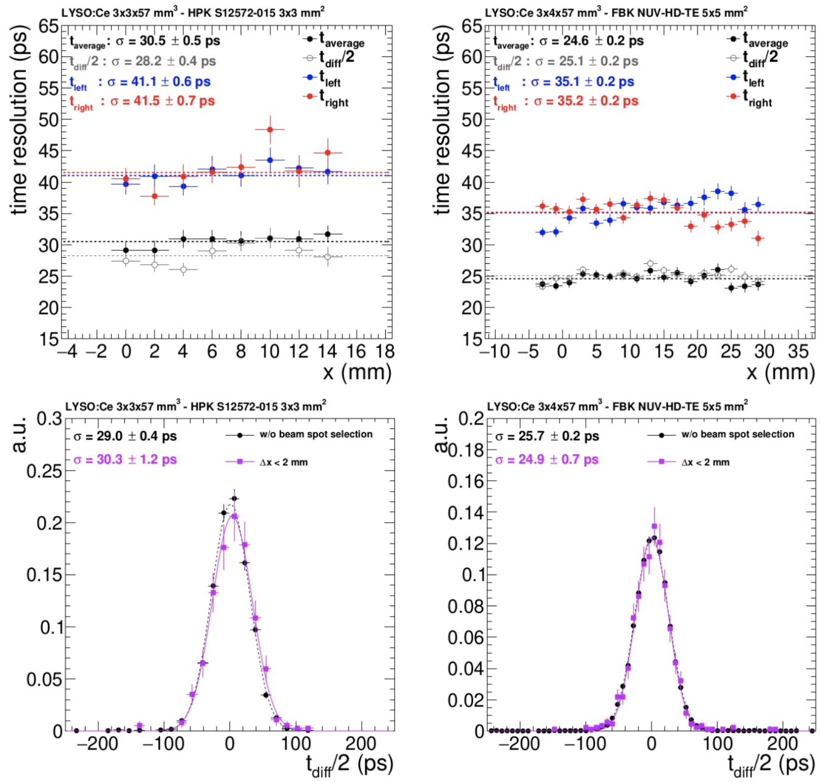


Figure 4.18: Time resolution measurements from the Fermilab April 2019 test beam. (Top) Time resolution of the left and right SiPMs, their average, and half of the time difference as a function of the MIP impact point for a 3×3×57 mm^3 LYSO:Ce bar coupled to HPK SiPMs (left) and for a 3×4×57 mm^3 LYSO:Ce bar coupled to FBK SiPMs (right). (Bottom) Global and local time resolution for a 3×3×57 mm^3 LYSO:Ce bar coupled to HPK SiPMs (left) and for a 3×4×57 mm^3 LYSO:Ce bar coupled to FBK SiPMs (right).

- $\sigma_{t_{\text{average}}} = \frac{1}{2}\sigma_{t_{\text{diff}}} = \frac{1}{2}\sqrt{\sigma^2_{t_{\text{left}}} + \sigma^2_{t_{\text{right}}}}$
- $\sigma_{t_{\text{left}}} = \sqrt{\sigma^2_{\Delta t_{\text{left}}} - \sigma^2_{t_{\text{MCP}}}}$
- $\sigma_{t_{\text{right}}} = \sqrt{\sigma^2_{\Delta t_{\text{right}}} - \sigma^2_{t_{\text{MCP}}}}$.

The final results are shown in Figure 4.18. A very spatially uniform time resolution can be achieved throughout the bar for both types of SiPM, as well for different crystal dimensions. In addition, a global and local resolution is computed using the beamspot information, as shown in Figure 4.18 (bottom). In all the measurements the target time resolution of 30 ps was achieved, demonstrating the feasibility of the BTL design.

4.3.3 Thermal properties of BTL modules

The major factors affecting the time resolution of the BTL were discussed in Sec. 4.2.1, and are summarized in Fig. 4.19 (left). The DCR is the main cause of worsening of the time resolution with increasing integrated luminosity. Studies have shown that the performance degradation caused by an increase of the 1 MeV neutron equivalent fluence of a factor 1.5 can be offset by lowering the operating temperature by 5 °C, as shown in Fig. 4.19 (right).

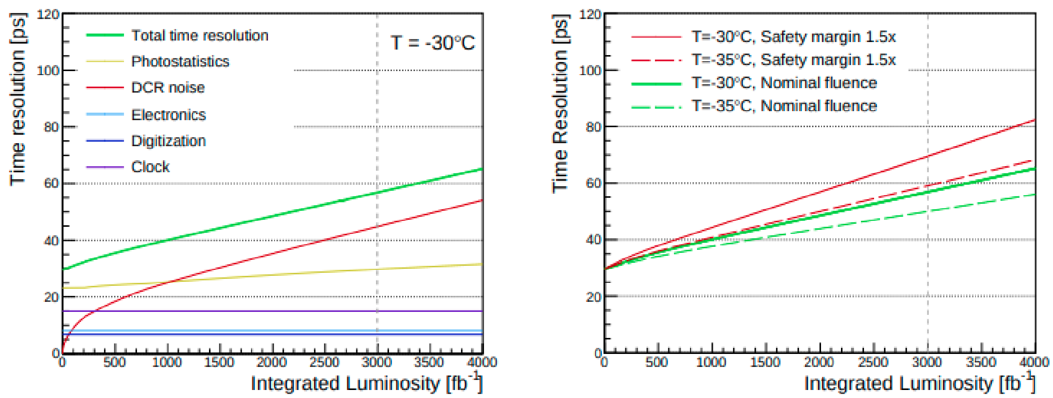


Figure 4.19: Time resolution of BTL sensors as a function of integrated luminosity. (Left) Different factors affecting the time resolution. DCR is the major reason for worsening time resolution as a function of integrated luminosity at the BTL. (Right) The effect of increased radiation fluence by a factor 1.5 (solid red) can be offset by lowering the operating temperature by 5 °C (dashed red) at the BTL. Both plots were taken from [92].

As mentioned before in Sec. 4.2.1, the BTL will be cooled to –35 °C with CO_2 cooling shared with the tracker and the modules will cooled to an additional –10 °C with thermo-electric coolers (TECs) to limit the DCR. For this to be effective, the thermal conductivity between all contact surfaces must be optimized to minimize thermal gradients across the BTL surface. Ideally, the temperature gradient between the cooling pipes and the module should be less than 2 °C.

To understand the thermal gradients that can be expected at the actual BTL detector, we setup a mock cooling experiment in the CPT Lab. We used a BTL module with 32 SiPMs + 16 LYSO:Ce crystals connected with flex cables to a tester readout board. The SiPMs were forward biased and the power was set to 65 mW per SiPM (to emulate a DCR of 50 mW and 15 mW of expected dynamic power consumption per SiPM). The module was placed inside a Copper housing, with a thermal paste (thermal conductivity - 10 W/mK) applied between the SiPMs and the Cu housing. The Cu housing was screwed to a mini-plate, which was in turn screwed to a cold plate. We also applied the thermal paste at the contact surfaces between the Cu housing and the mini-plate, and the mini-plate and the cold plate. The BTL module had two NTC thermistors, one on each edge of the module (with the SiPMs). We placed additional NTC thermistors along the setup at different points. This entire setup was placed inside an Aluminium box, that was sealed and flushed with Nitrogen to bring the humidity inside the box down to < 5%. The temperature and humidity (TH) inside the box was measured using a ChipCap 2-SIP sensor [106]. The TH sensor and all NTC thermistors were readout using an AD5593R [107] analog to digital converter (ADC), controlled by an I^2C bus connection to an external Raspberry Pi. The cold plate was then cooled down to 15°C using cooling pipes that connected it to a Fisher Scientific Isotemp ethylene glycol cooler kept outside the box. Figure 4.20 shows the mock cooling setup.

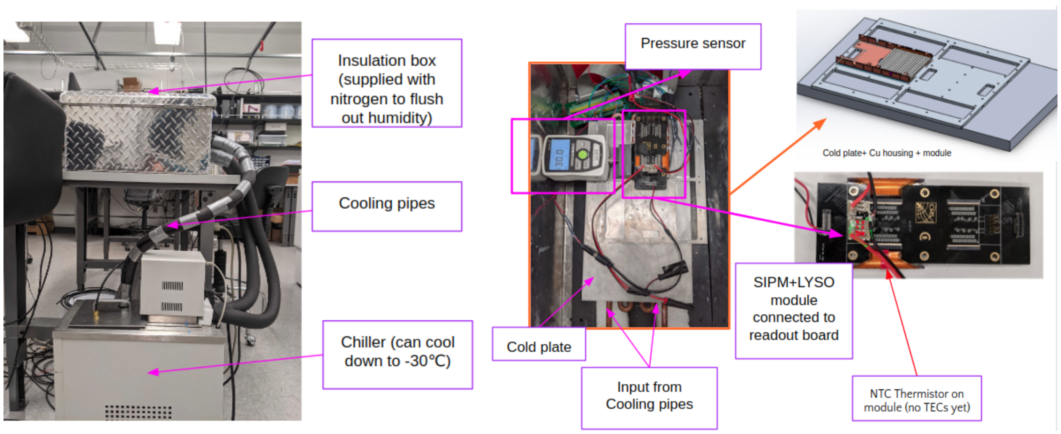


Figure 4.20: Mock cooling setup at the CPT Lab. (Left) The Aluminium box connected to the Isotemp cooler using insulated pipes. (Right) The inside of the Aluminium box consists of a BTL module connected to a tester readout board and placed inside a Copper housing.

Thermistors 2 and 4 measure the temperature at different location on the mini-plate, Thermistors 0 and 1 measure the temperature at the two corners of the same edge of the module, Thermistor 7 (on the module itself) is placed at the center of one edges of the module, Thermistors 5 measures the temperature of the Cu housing and Thermistor 3 measures the temperature of the cold plate (see SOLIDWORKS simulation of the setup in Fig. 4.21). Measurements from the various thermistors were recorded and are shown in Fig. 4.22. The highest temperature gradient recorded between the center of the module and the cold plate is 0.4°C, which is already less than the 2 °C goal.

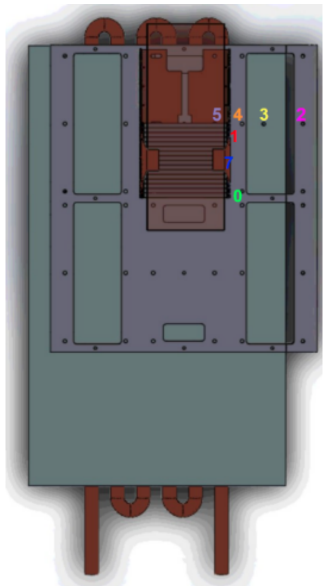


Figure 4.21: A SOLIDWORKS simulation of the cold plate setup. This simulation was performed by Caltech SURF student Esme Knabe in Summer 2021 [108]. Colored numbers indicate the position of the thermistors.

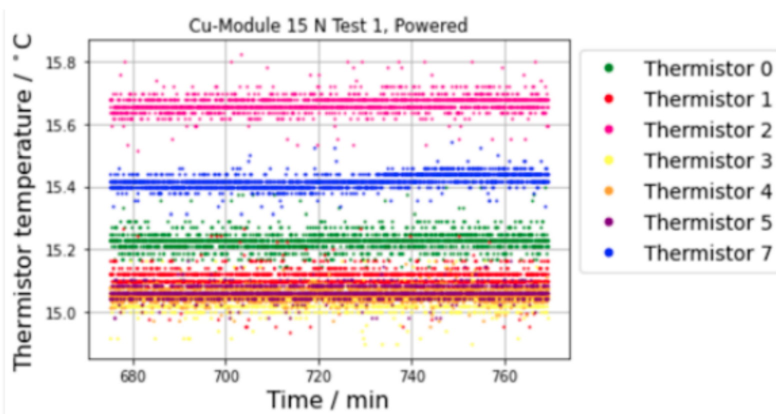


Figure 4.22: Measurements from the various thermistors on the setup when cooled to 15 °C for the powered SIPMs. The highest temperature gradient recorded between the center of the module (Thermistor 7) and the cold plate (Thermistor 3) is 0.4°C. Thermistor 2 records a slightly higher temperature of 15.6 °C, as it is placed farther away from the cooling pipes.

Currently, there are ongoing efforts in the CPT lab to repeat these measurements at –40°C, which is closer to the actual BTL operating temperature. Tests in –40 °C are more challenging due to several factors like cables freezing in the setup, or the Nitrogen air temperature being higher than –40°C, leading to air convection effects playing a significant role in the thermal gradient measurement. There is also an ongoing survey to identify the thermal couplant best suited for BTL (based on ease-of-use nature, radiation hardness, thermal conductivity, etc.). In the near future, we will also have the first prototype versions of modules with TECs on them, and more realistic thermal gradient studies can be performed with those modules.

Chapter 5

FIRST EVIDENCE OF A HIGGS BOSON DECAY TO A PAIR OF MUONS

5.1 Introduction

Various measurements of the Higgs boson interactions with the standard model (SM) particles, including the electroweak gauge bosons and third generation charged fermions, have indicated that the properties of the Higgs are consistent with SM predictions [109–117]. The Higgs-Yukawa couplings with the first and second generation fermions are however yet to be established experimentally. The $\mathcal{B}(H \to f\overline{f})$ is expected to be small for fermions belonging to the first and second generations as the Yukawa couplings are propotional to the mass of the fermion.

Amidst the second generation fermions, the $H \to cc$ branching ratio is the highest, but this particular channel at the LHC gets overwhelmed by multi-jet backgrounds with higher rates. $H \to \mu\mu$ is the cleanest probe at the LHC for establishing second generation Yukawa couplings. The $H \to \mu\mu$ branching ratio for the 125.38 GeV Higgs boson is predicted by the SM to be 2.18×10^{-4} [40].

Previously, a CMS search for $H \to \mu\mu$ decays [118] (performed using a combination of proton-proton collision data collected at 7, 8 and 13 TeV, corresponding to integrated luminosities of 5.0, 19.8 and 35.9 fb^{-1}, respectively) , found the observed (expected) significance to be 0.9 (1.0) standard deviation w.r.t. the background and the observed (expected) upper limit on the production cross section times $\mathcal{B}(H \to \mu\mu)$ to be 0.9 (2.2) times the SM prediction. The most recent search for $H \to \mu\mu$ decays from the ATLAS Collaboration corresponding to the full Run 2 dataset [119], saw an observed (expected) significance over the background-only hypothesis for a Higgs boson with a mass of 125.09 GeV at 2.0 σ (1.7 σ).

This chapter describes the search for $H \to \mu\mu$ decays performed using the full Run-2 dataset collected at the CMS detector, at a center of mass energy of 13 TeV, and corresponding to a total integrated luminosity of 137.3 fb^{-1}. The analysis described here targets the Higgs boson production via gluon fusion (ggH), vector boson fusion (VBF), and in association with a vector boson (VH, V = Z, W$^{\pm}$) or with a top-antitop pair (ttH). The final states of interest are a pair of two prompt, isolated and oppositely charged muons produced in association with 0, 1, or 2 hadronic jets or in association with one or more additional leptons (μ, e). The events are divided into several categories in order to target the various Higgs production modes and increase the signal to background ratio.

For the ggH and VBF enriched channels, the largest backgrounds consist of Drell-Yan events in which an off-shell Z boson decays to a pair of muons and the electro-weak production of a Z boson which decays to a pair of muons (see Fig 5.1). The VBF signal is smaller in cross section compared to the ggF signal by an order of magnitude, but has a distinctive signature: the two muons from the Higgs boson decay are produced in association with two very forward jets, that have a large jet-jet pseudorapidity gap ($\Delta\eta_{\mathrm{jj}}$), and a large dijet invariant mass (m_{jj}). This feature helps to suppress the large Drell-Yan background, making the vector boson fusion channel a very sensitive probe of the $H \rightarrow \mu\mu$ signal. For the VH and ttH channels, the main backgrounds are reprensented by diboson (WZ, ZZ) and $\mathrm{t\bar{t}}$-based ($\mathrm{t\bar{t}}$, $\mathrm{t\bar{t}Z}$) processes, respectively.

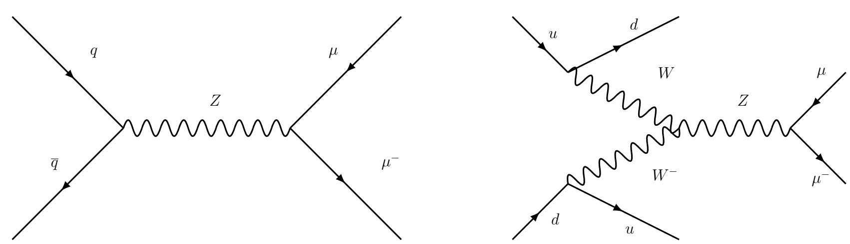


Figure 5.1: The Drell-Yan (left) and electroweak production (right) of a Z boson decaying to a pair of muons.

In relation to the work done for this chapter, Appendix D describes a fast, kernel based library, known as "hepaccelerate" [25], which was designed for analyzing billions of events from a typical collider experiment with a small turnaround time. The *hepaccelerate* library relies on a parallel computing architecture and can process data on GPUs or CPUs. This library was used for the data analysis of the CMS Run 2 $H \rightarrow \mu\mu$ search, and has also been validated on the CERN Open Data.

5.2 Data sets and simulation

5.2.1 Data sets and triggers

This search uses the pp collision data collected by the CMS detector during the data taking years 2016, 2017, and 2018, corresponding to a total integrated luminosity of 137 fb^{-1}.

Signal events in this analysis contain two prompt, isolated, and high p_{T} muons. When the Higgs boson is produced at rest in the transverse plane ($p_{\mathrm{T}}^{\mathrm{H}} \approx 0$), the two muons from its decay are emitted back-to-back with a transverse momentum of about $m_{\mathrm{H}}/2 \approx 60\,\mathrm{GeV}$. At the generator level, the p_{T} distribution for the leading muon produced in both ggH and VBF signal events turns on at about 40 GeV and peaks around 60 GeV. Therefore, signal events for this search were selected in data using online single muon triggers that impose a loose isolation requirement on each muon candidate, and a p_{T} threshold of 27 (24) GeV in 2017 (2016, 2018).

5.2.2 Simulation overview

The processes considered in this analysis have been simulated using either the MadGraph5_amc@nlo (v2.2.2) [120] or the powheg (v2) [121] generators. The matrix element level Monte Carlo (MC) events from these generators are then interfaced with pythia (v8.2 or greater) [122] in order to simulate the fragmentation and hadronization of partons in the initial and final states along with the underlying event. This is done using the CUETP8M1 tune [123] for simulations corresponding to the 2016 data taking era, and using the CP5 tune [124] for the 2017/18 data taking eras.

In the case of the processes simulated with the MadGraph5_amc@nlo generator at leading order (next-to-leading order), jets from the matrix element calculations are matched to the parton shower produced by pythia following the MLM (FxFx) prescription [125, 126]. The 2016 (2017/18) era simulations use the NNPDF 3.0 (3.1) parton distribution functions [127, 128]. The interactions of all final state particles with the CMS detector are simulated using GEANT4 [129]. Lastly, the simulated events include the effects of pileup, with the multiplicity of reconstructed primary vertices matching that in data.

5.2.2.1 Signal simulation

The MC samples for ggH and VBF productions are simulated at next-to-leading order accuracy in QCD using both MadGraph5_amc@nlo and powheg generators. Since the powheg simulation only contains events with positive weights, powheg samples have been used in the training of the BDT multivariate discriminants, which cannot correctly handle negative weights. On the other hand, the MadGraph5_amc@nlo samples are used in the final signal extraction. In contrast, simulated events for VH and ttH processes are produced only via the powheg generator. Additional signal samples, obtained by varying the tune parameters for the underlying event simulation, are also produced and used to estimate the corresponding systematic uncertainty.

It was observed that there is a significant difference in the parton showering done by PYTHIA and HERWIG ++ (v3.0) [130] generators specifically in the case of the VBF process, that has a distinct feature of two jets with a large $\Delta\eta_{\mathrm{jj}}$ separation and no color connection in the rapidity gap. Therefore, VBF signal samples showered using both PYTHIA and HERWIG ++ have been generated and compared to each other. Parton showering with HERWIG ++ is done using the UE-EE-5C tune [123]. For all the other signal production modes (ggH, VH, and ttH), only PYTHIA was considered for the parton showering.

Table 5.1 provides the cross sections for each of the five main Higgs boson production modes at the LHC for a 125.0 GeV SM Higgs boson, along with the respective theoretical uncertainties, as recommended by the LHC Higgs Cross Section Working Group [40]. In addition, simulated events have also been produced for Higgs boson masses of 120 and 130 GeV, allowing to interpolate signal acceptance and lineshape parameters over a 10 GeV mass range.

Table 5.1: Higgs boson production cross sections for various modes at $\sqrt{s}$ = 13 TeV.

Process	Cross section (pb)	Perturbative Order	+QCD scale unc. (%)	-QCD scale unc. (%)	+(PDF+α_s) unc. (%)	-(PDF+α_s) unc. (%)
ggH	48.58	N3LO(QCD) NLO (EWK)	+4.6	-6.7	+3.2	-3.2
VBF	3.782	NNLO (QCD) NLO (EWK)	+0.4	-0.3	+2.1	-2.1
WH	1.373	NNLO (QCD) NLO (EWK)	+0.5	-0.7	+1.9	-1.9
qq $\rightarrow$ ZH	0.884	NNLO (QCD) NLO (EWK)	+3.8	-3.1	+1.6	-1.6
0gg $\rightarrow$ ZH	0.123	NLO (QCD) O(α_s^3)	+25.1	-18.9	+2.4	-2.4
ttH	0.507	NLO (QCD) NLO (EWK)	+5.8	-9.2	+3.6	-3.6

5.2.2.2 Background simulation

The largest contribution to the background in this search comes from Drell-Yan events. We are particularly interested in the off-shell production of the Z boson in the mass range of 110 to 150 GeV (closer to the Higgs peak). Therefore, a Drell-Yan MC sample was generated via MADGRAPH5_aMC@NLO, with NLO precision in QCD and with up to 2 jets in the final state, applying a dimuon mass cut at the matrix element level in the range between 105 and 160 GeV. In addition, in order to gain statistics in a VBF-like phase space, a NLO QCD DY+2-jet sample was produced via MadGraph5_amc@nlo requiring two jets with $m_{jj} > 350$ GeV at the generator level. Dimuon events under the Z peak, however, are also of interest to study the kinematic properties of the Drell-Yan background, to compute certain data-based corrections, and, in general, to assess the reliability of the simulation. For such studies, a Drell-Yan MC sample with a dimuon mass cut of 50 GeV at the matrix element level, produced again using MadGraph5_amc@nlo at NLO in QCD with up to 2 jets in the final state, is used. When dimuon events with VBF-like jets are considered, the contribution from the electroweak production of the Z boson becomes significant. This process was simulated at LO using MadGraph5_amc@nlo and showered via herwig ++. Herwig parton shower is adopted instead of pythia because it is known to better model purely electroweak process without color connection [131]. Two alternative samples for Z-EWK production are available: one in which the invariant mass of the two truth level muons is required to be larger than 50 GeV, another in which $m_{\mu\mu}$ is between 105 and 160 GeV. The former sample is used when data and simulation are compared under the Z-peak, while the latter is used for the signal region.

The next most significant background contribution comes from $t\bar{t}$ events in which both the top quarks decay leptonically. This process was generated at NLO using the powheg generator. Furthermore, there are minor contributions from other top quark processes such as single top (tW, t, and s-channel) production, ttZ, ttW, and tZq, as well as from semi-leptonic $t\bar{t}$ decays. These contributions are also taken into account via dedicated NLO simulation produced with MadGraph5_amc@nlo or powheg. Remaining background can be attributed to the diboson processes (ZZ, WZ, WW) with some very small contributions from triboson production (WWW, WWZ, WZZ, ZZZ) that have also been taken into account. The diboson processes have been simulated using either MadGraph5_amc@nlo or powheg at NLO in QCD, whereas the triboson processes have been simulated using MadGraph5_amc@nlo with NLO precision in QCD corrections.

The cross sections used for normalizing the background expected yields are obtained from the best available theoretical predictions. In particular, the Drell-Yan cross section was obtained from the FEWZ [132–134] generator at NNLO accuracy in QCD, and NLO accuracy in electroweak corrections. Similarly, the cross section for the $t\bar{t}$ was computed at NNLO accuracy in QCD [135].

5.3 Physics objects

Physics objects are reconstructed based on the particle-flow (PF) algorithm [87]. This analysis mainly uses physics objects like muons and jets in order to target the final state signature of $H \to \mu\mu$ decays. We also rely on other physics objects such as additional leptons (muons or electrons) and b-tagged jets in order to tag the different Higgs boson production modes and suppress certain backgrounds like $\mathrm{t\bar{t}}$ and VV production.

5.3.1 Primary vertex

The candidate vertex with the largest value of summed physics-object p_{T}^2 is taken to be the *primary interaction vertex*.

5.3.2 Muons

The muon reconstruction in CMS combines information from both the tracker and muon chambers. There are three types of reconstructed muons [81]:

- **Standalone muon**: Standalone muon tracks are reconstructed using hits in the muon chambers. Hits in the DT or CSC are used as seeds (or starting points), and track segments are built using DT, CSC, and RPC hits, using the iterative Kalman Filter algorithm [88].

- **Tracker muon**: Tracks reconstructed from tracker hits (see Sec. 3.2.7) with $p_{\mathrm{T}} > 0.5$ GeV and a total momentum p > 2.5 GeV are extrapolated to the muon systems. If at least one matching DT or CSC muon segment is found for the extrapolated track, the tracker track is considered as a tracker muon. Low p_{T} muons might only leave hits in the tracker + the innermost muon station, and only get reconstructed as tracker muons. On the other hand, remnants of a hadronic shower can sometimes reach the innermost muon station, known as a *punch-through*, and can look like a (fake) tracker muon.

- **Global muon**: A global muon is reconstructed by matching a standalone muon to a tracker track. A combined fit is performed using information from both the tracker track hits and the standalone muon hits to determine the final parameters of the global muon track. Since the tracker hits have better spatial resolution than the muon chamber hits, the contribution from muon detectors to the muon p_{T} measurement is marginal for muons with $p_{\mathrm{T}} < 200$ GeV. However, for muons with $p_{\mathrm{T}} > 200$ GeV, the tracker track is essentially a straight line, and information from the muon chambers improves the p_{T} measurement significantly, because they provide essential information about the track curvature.

One can define additional muon identification (ID) criteria based on several kinematic variables: the number of hits and the fit quality of the track; the compatibility between the tracker hits and the segments in the muon stations; and the compatibility between the muon track and the primary vertex. The global track fit χ^2 and a kink-finder χ^2 are used as indicators of the fit quality of the global muon track. The compatibility between the tracker track and muon station segments is evaluated with a position match χ^2, and a variable called the segment compatibility. The compatibility between the track and the primary vertex is evaluated with the track impact parameters (closest distance of approach in the x-y plane or z axis). Based on these variables, muons can be further classified into three different types as follows:

- **Loose muon ID** : Muons that are either a tracker or a global muon, with no further requirements.
- **Medium muon ID** : A medium muon satisfies the loose ID criterion, and possesses valid hits on more than 80% of the number of tracker layers that it traverses. Additionally, the muon must be classified either as a *good global muon* or should pass a tight segment compatibility requirement (> 0.451). A *good global muon* has a global track-fit χ^2/ndof < 3, the kink-finder χ^2 < 20, the position match χ^2 < 12, and a segment compatibility > 0.303.
- **Tight muon ID** : A tight muon is a global muon with a global track-fit χ^2/ndof < 10, has at least one hit in the muon chamber and at least six hits in the inner tracker (with > 1 pixel hits). It also satisfies certain track impact parameter (w.r.t. the primary vertex) requirements, which are $d_{xy} < 0.2$ cm and $d_z < 0.5$ cm.

Events in the $H \to \mu\mu$ analysis are required to have at least two oppositely charged Medium ID muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$. Furthermore, each of the muons is also required to pass a relative isolation requirement in which the sum of the p_{T} of charged and neutral hadrons, and photons, lying within a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ around the muon, is less than 25% of the p_{T} of the muon. If a photon in the isolation cone is associated with the final state radiation (FSR) of the muon (see Section 5.3.2.2), then it is not included in the isolation sum. A $\Delta\beta$-correction is also applied to the isolation sum which subtracts half of the p_{T} sum of charged hadron candidates lying within the isolation cone but originating from pile-up vertices. This correction helps to remove the contamination from neutral particles produced by pile-up interactions that spuriously enter into the vicinity of the candidate muon.

The sensitivity of this search depends critically on the muon p_{T} resolution and also on the resolution of the $M_{\mu\mu}$ mass peak. The p_{T} resolution of muons worsens with increasing muon $|\eta|$, the resolution being around 1–2% in the central barrel region of the detector ($|\eta| < 0.9$) , and degrading to 2 to 3.5% for muons passing through the endcaps of the muon system ($|\eta| > 1.2$), as shown in Fig. 5.2. The next few subsections will discuss ways to correct for any mis-reconstruction of the muon p_{T} originating from software bias, uncertainties in the magnetic field or detector misalignment.

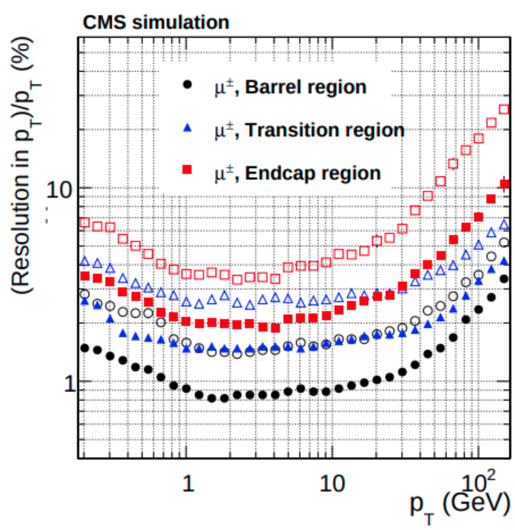


Figure 5.2: Resolution, as a function of p_{T}, for single, isolated muons in the barrel ($|\eta| < 0.9$), transition ($0.9 < |\eta| < 1.4$), and endcap regions ($1.4 < |\eta| < 2.5$). For each bin, the solid (open) symbols correspond to the half-width for 68% (90%) intervals of the residuals distribution [89].

5.3.2.1 Rochester Corrections

The muon momentum scale and resolution are calibrated in bins of muon p_{T}, η and charge, using $Z \to \mu\mu$ decays as a standard candle, following the method described in Ref. [136]. The correction method is briefly summarized as follows

- Negative and positive muons are divided separately into different η and ϕ bins, for both data and simulation. In each bin, the $1/p_{\mathrm{T}}$ distributions for both data and simulation are corrected, so that the mean value ($<1/p_{\mathrm{T}}>$) becomes the same as that in the $Z \rightarrow \mu\mu$ simulation (which is assumed to be very precisely modelled).

- A smearing is applied to the resolution of the $1/p_{\mathrm{T}}$ distribution in the simulation, such that it matches the resolution in data.

- After all the above steps, there may still be some residual offset in each $m_{\mu\mu}$ bin, when compared to the $Z \rightarrow \mu\mu$ simulation. The last step is to apply the ratio of this offset and the nominal Z mass, as a correction factor to the muon p_{T}. The correction is applied in an iterative manner, until the offset is minimized.

5.3.2.2 Final state radiation (FSR) recovery

In a small fraction of signal events (9%), a muon in the final state may radiate a photon, thereby losing some of its momentum. This causes a slight degradation in the resolution of the signal $M_{\mu\mu}$ peak. To recover this loss in the resolution, a procedure was developed to look for FSR photons within the isolation cone of the muon and can be summarized as follows

- Consider all reconstructed muons with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$ as candidates for FSR recovery.

- For a given muon, consider all photons (γ) with $\Delta R(\mu, \gamma) < 0.5$, $p_{\mathrm{T}} > 2\,\mathrm{GeV}$, $0.0 < |\eta| < 1.4442$ or $1.566 < |\eta| < 2.5$ as possible FSR candidates.

- Ignore photons that are associated with the bremsstrahlung of a reconstructed electron.

- In order to strongly suppress the contamination from $\mathrm{H} \rightarrow \mathrm{Z}\gamma \rightarrow \mu\mu\gamma$ decays, FSR photon candidates are required to have $p_{\mathrm{T}}^{\gamma}/p_{\mathrm{T}}^{\mu} < 0.4$.

- Impose a loose isolation requirement on the photon :
  $I_{\gamma}/p_{\mathrm{T}}(\gamma) = (\Sigma_i^{PF} p_{\mathrm{T}}^i(\Delta R(\gamma, i) < 0.3))/p_{\mathrm{T}}(\gamma) < 1.8$, where $p_{\mathrm{T}}(\gamma)$ is the p_{T} of the FSR photon candidate and the index i refers to the PF candidates other than the muon within a cone of $R = 0.3$ around the photon.

- Require photon to be collinear with the muon : $\Delta R(\mu, \gamma)/p_{\mathrm{T}}(\gamma)^2 < 0.012$
- In case of multiple FSR photon candidates, only the one with the smallest value of $\Delta R(\mu, \gamma)/p_{\mathrm{T}}(\gamma)^2$ is considered.

If an FSR photon is associated with a muon, its momentum is added to that of the original muon. This procedure increases the signal efficiency by about 3% and improves the $M_{\mu\mu}$ resolution by around 2%.

5.3.2.3 GeoFit Corrections

In CMS, the muon p_{T} values are primarily computed using the measured radius of curvature (R) of the reconstructed muon track from hits in the inner tracker. This reconstruction has inherent uncertainties, which affect both the track trajectory and measured p_{T}. As prompt muons originate directly from the collision vertex, the measured point of closest approach between the muon track and the collision vertex in the $x - y$ plane (known as the track impact parameter: d_0) should be exactly zero for muons coming from W/Z/H decays. However, if the muon track radius estimation is incorrect, the d_0 value will be non-zero, and is related to the mis-measurement in radius of curvature (ΔR) as follows (see [137] for derivation):

$$|d_0| \sim \frac{\Delta R}{R^2}. \tag{5.1}$$

Since in homogeneous magnetic fields, $p_{\mathrm{T}} \sim R$, we can re-write the above equation as

$$|d_0| \sim \frac{|\delta p_{\mathrm{T}}|}{p_{\mathrm{T}}^2}. \tag{5.2}$$

Therefore, the precision of the muon p_{T} measurement can be improved by including the interaction point position as an additional hit of the muon track. The corresponding adjustment in the p_{T}, δp_{T}, is given by Eqn. 5.2. The resulting improvement in the expected $m_{\mu\mu}$ resolution in signal events ranges from 5% to 10%, depending on muon p_{T}, η, and the data-taking period (see Fig. 5.3).

5.3.3 Jets

Jets used in the analysis are reconstructed by clustering the PF candidates in the event using the anti-k_{t} algorithm [90] with a distance parameter of R=0.4. While clustering the jets, only the charged PF candidates that are associated with the primary interaction vertex are considered (to remove any additional tracks/ calorimetric

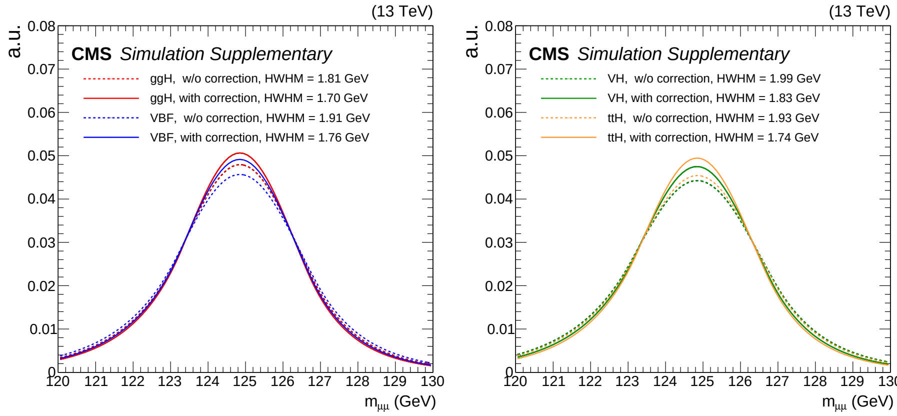


Figure 5.3: The expected $m_{\mu\mu}$ distributions in simulated signal events with m_H= 125 GeV, passing the event selection requirements described in Sec. 5.5, obtained with (solid) and without (dashed) including the interaction point as an additional constraint for the muon track. Left: ggH and VBF processes. Right: VH and ttH. The signal peak is modelled using a double-sided Crystal Ball parametric function. The inclusive improvement in the mass resolution, estimated by comparing the HWHM of the corresponding signal peaks, ranges between 5-10% depending on the production mode.

energy deposits from particles that originate from pile-up vertices). Jet momentum is determined as the vector sum of all particle momenta inside the jet, and is found from simulation to vary, on average, between 5 and 10% of the true momentum over the whole p_{T} spectrum and detector acceptance. An offset correction (L1 correction) is applied to the jet energies to take into account the contribution from neutral particles originating from pile-up interactions within the same (in-time) or adjacent bunch crossings (out-of-time) [138]. To bring the measured energy response of the reconstructed jets from the detector at par with the true energy of the original particle, additional energy corrections are derived from simulation and applied as a function of jet η and p_{T}. Finally, residual difference in the data and simulation are corrected (L2L3 corrections) using *in situ* measurements of the momentum balance in dijet, multijet, γ+jets, and leptonically decaying Z+jets events [139].

Selected jets in this search are required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 4.7$. Certain quality requirements are imposed on the jets to suppress the contamination from pileup jets and mis-reconstructed jets due to detector noise in the fiducial region beyond the coverage of the tracker. Jets are also required to have a geometrical separation of $\Delta R(j, \mu) > 0.4$ from the two muon candidates used to reconstruct the Higgs boson.

5.3.4 Missing transverse momentum

The missing transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) is computed by taking the magnitude of the negative vector sum of the transverse momenta of all the reconstructed PF candidates in a given event. Corrections to the energy scale of jets in the event are propagated to the $p_{\mathrm{T}}^{\mathrm{miss}}$. To reject spurious high $p_{\mathrm{T}}^{\mathrm{miss}}$ events (resulting from a variety of reconstruction failures, detector noise, or non-collision backgrounds), special event filters are used that have an accuracy of 85–90% with a mis-tagging rate smaller than 0.1% [140].

5.3.5 Electrons

Electron candidates used in the analysis are required to have $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$ and are reconstructed using an algorithm that matches fitted tracks in the tracker with ECAL energy deposits [141]. Electron candidates in the transition region between the ECAL barrel and endcaps, $1.44 < |\eta| < 1.57$, are discarded due to non-optimal reconstruction performance. Certain quality criteria are also imposed on these candidates [142], based on the shower shape of the energy deposit, the matching of the electron track to the ECAL energy cluster, the relative amount of energy deposited in the HCAL detector, and the consistency of the electron track with the primary vertex. Electron candidates identified as coming from photon conversions in the detector are rejected. Electrons are also required to be isolated to a certain degree from hadronic activity within a cone of radius $R = 0.3$ around the electron track.

5.3.6 B-tagged jets

Jets arising from the hadronization of bottom quarks are identified with the `DeepCSV` b-tagging algorithm, a neural network that takes tracks displaced from the primary interaction vertex, identified secondary vertices, jet kinematic variables, and information related to the presence of soft leptons in the jet [143], as input variables. Working points (WP) that yield either a 1% (medium WP) or a 10% (loose WP) probability of misidentifying a light-flavour jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ as a b quark jet are used. The corresponding average efficiencies for the identification of the hadronization products of a b quark as a b quark jet are about 70% and 85%, respectively. Simulated events are corrected in order to take into account differences between data and MC measured in the b-tagging efficiency for both bottom quark and light flavor jets.

5.3.7 Track jets (additional *soft* hadronic activity)

Track jets [144, 145] are only comprised of charged tracks that originate from the primary interaction of the event. They are used for measuring low energy jets (down to few GeV) and are less susceptible to pile-up effects, as they do not use energy information from calorimetric deposits. Track jets are often used to estimate the additional hadronic activity in an event (as described later in Section. 5.6.1). Tracks for track jets are required to satisfy certain track quality conditions, have $p_{\mathrm{T}} > 300$ MeV, and have a small impact parameter ($|d_z| < 2$ mm) with respect to the primary interaction vertex. Track jets are then built by clustering the selected tracks with the anti-k_{t} algorithm with a distance parameter of R = 0.4.

5.4 Corrections to data and simulation

In addition to certain corrections already mentioned in Section 5.3, we apply additional corrections to several physical observables in both data and simulation, to account for certain biases in detector response and any mis-modeling of physical processes in our simulation. This is essential for the VBF category (as described later in Section 5.6), where the background modeling fully relies on the prediction from simulation. Additionally, we entirely depend on the simulation to accurately describe the signal processes. The following sub-sections will describe these corrections.

5.4.1 Pileup re-weighting

Our simulations are produced with a pileup profile that is somewhat similar to what is measured in data, but does not match exactly. Therefore, a re-weighting is performed to ensure that the profile of the pileup interactions in simulation matches the estimated pileup distribution integrated over the data taking period. To get the pile-up profile from data, we multiply the estimated luminosity per data-taking period with the minimum bias events cross section of 69.2 mb ± 5%. For simulation, the true number of in-time (from the same proton bunch-crossing) pileup interactions are known, and are used to obtain the pileup profile. The ratio of the data and simulation profiles gives the weight to be applied to a simulated event as a function of the true number of in-time interactions. These pileup weights are derived separately for the 2016, 2017, and 2018 eras.

5.4.2 L1 EGamma pre-firing corrections

The L1 trigger system of CMS has been designed to forbid triggering events in consecutive bunch crossings. During Run 2, the ECAL endcap time alignment of the detector slowly drifted with respect to the central clock, becoming particularly worse towards the end of the 2017 data-taking period. This resulted in an inefficiency in the L1 trigger decision ; the trigger primitives generated by the ECAL deposits, belonging to the $(t-1)^{th}$ bunch crossing, led to inaccurate triggering of the t^{th} bunch-crossing.

The main source of pre-firing are events with forward jets, which particularly affect our VBF events. The ggH, ttH, and VH enriched categories see a small efficiency loss due to pre-firing, whereas the VBF category has inefficiencies on the order of 2–5% in 2016, and 4–10% in 2017 (with no significant loss observed in 2018). This inefficiency has been measured centrally by CMS as a function of both photon and jet p_{T} in the forward η regions of the endcap ($2.5 < |\eta| < 3.0$). Figure 5.4 shows the central maps of pre-firing probabilities ($\mathcal{P}_{\text{prefire}}(j)$) as function of the pseudo-rapidity and the electromagnetic component of the jet transverse momentum.

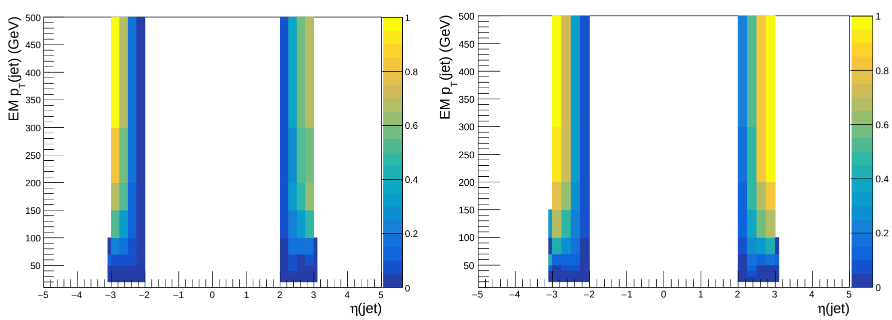


Figure 5.4: Centralized pre-firing probability provided per-jet as function of its electromagnetic transverse momentum and η for 2016 (left) and 2017 (right).

The event efficiency factor ($\varepsilon_j = 1 - \mathcal{P}$) is applied to the MC simulation and it is accounted using all jets in the event, as

$$\varepsilon = \prod_{j \in \text{jets}} 1 - \mathcal{P}_{\text{prefire}}(j).$$

5.4.3 Muon efficiency and trigger scale factors

The efficiencies of the muon identification and isolation requirements used in this analysis have been measured centrally by CMS, using a tag-and-probe technique on Z+jets events in data and simulation. These efficiencies are derived as a function of the muon p_T and η and are used to compute data-to-simulation scale factors that are then applied per muon, in simulated events. Therefore, the identification and isolation scale factors are applied as a product of all the selected muons.

We also apply efficiencies for a muon passing the baseline identification and isolation requirements, to fire the single muon trigger. These are also computed as a function of the muon p_T and η. In a dimuon event, if the per muon trigger efficiencies are ϵ_1 and ϵ_2, then the per-event efficiency to fire the trigger is given by $\epsilon = 1 - (1 - \epsilon_1)(1 - \epsilon_2)$. These efficiencies can be computed separately for data and for simulation, and their ratio gives the scale factor that is then applied as a weight to a simulated dimuon event.

5.4.4 Correction to Higgs boson transverse momentum

The Higgs boson signal produced through gluon fusion is simulated by Monte Carlo generators at NLO in QCD and then matched to the parton shower. The p_T of the Higgs boson is corrected to the prediction from the NNLOPS generator [146, 147], which is the highest order parton shower matched ggH simulation available that includes the finite top quark mass effects. The reweighting is performed in exclusive jet bins, $N_{jets} = 0, 1, 2, 3$ or more jets, where jets are defined at the generator level and required to have $p_T > 30\,\text{GeV}$. More details can be found in Section 4.1 of Ref. [112].

5.5 Analysis strategy

We start by selecting events accepted by the single muon triggers (described in Section 5.2.1), in which there are exactly two opposite charged muons passing the requirements described in Section 5.3.2. These pre-selected events are further divided into exclusive categories which aim to separate the five different Higgs boson production modes: ttH, WH, ZH, VBF, and ggH.

We start by classifying events containing either one medium WP b-tagged jet or two loose WP b-tagged jets into the ttH-category. This category is further divided into two sub-classes: events containing electrons or additional muons, passing the conditions described in Sections 5.3.2 and 5.3.5, respectively, belong to the ttH leptonic category, while events without additional leptons are grouped in the ttH hadronic category.

Events with no b-tagged jets but containing at least one additional lepton (muon or an electron) are classified into the VH-category. The VH category is further divided into two sub-classes: events containing only one additional electron or muon belong to the WH leptonic category, while events containing two additional electrons or muons are grouped into the ZH category.

Events that do not satisfy the requirement of the ttH or VH categories but contain at least two hadronic jets with $p_T > 25\,\text{GeV}$ ($> 35\,\text{GeV}$ for the leading jet), dijet invariant mass $m_{\text{jj}} > 400\,\text{GeV}$, and the jet-jet angular separation $|\delta\eta_{jj}| > 2.5$, are classified into the VBF-enriched category.

Finally, the residual phase space made by events with no additional jets or with just one jet represents the ggH enriched category. Figure 5.5 gives a simple graphical sketch of the event categorization described above.

In the next few sections, I will explain the VBF category in more detail. I will conclude by briefly describing how the search was done in other categories and present the final combined results for this analysis, along with future search projections for $H \to \mu\mu$.

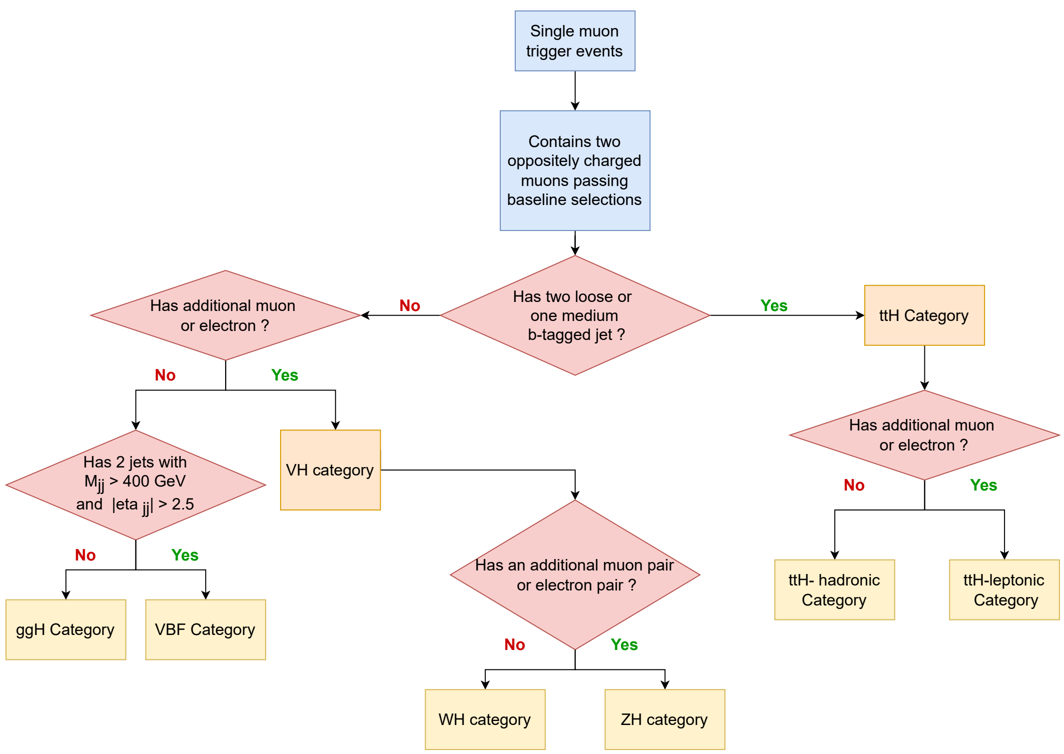


Figure 5.5: Graphical summary of the logical definition of each event category from the baseline dimuon selection. Exclusive channels are defined to target the following production modes: ttH (leptonic or hadronic $t\bar{t}$ final states), leptonic WH and ZH, VBF, and ggH.

5.6 VBF category

The VBF $H \to \mu\mu$ signal has a very distinctive feature: the two muons from the Higgs boson decay are produced in association with two hard jets in the forward regions of the detector. The two jets have a large pseudorapidity gap between them ($\Delta\eta_{\text{jj}}$), and also form a large dijet invariant mass (m_{jj}). This feature helps to suppress the backgrounds in this category, making the VBF channel a very sensitive probe of the $H \to \mu\mu$ signal.

As described in Sec. 5.3 and Sec. 5.5, events must pass the following selections to be classified into the VBF category:

- two muons as described in Sec.5.3.2
- At most 1 loose b-tagged jet and no medium b-tagged jets
- Leading jet $p_{\text{T}} > 35$ GeV and sub-leading jet $p_{\text{T}} > 25$ GeV
- Invariant mass of the di-jet pair: $m_{jj} > 400$ GeV
- Angular separation between the two jets: $|\Delta\eta_{jj}| > 2.5$

Three orthogonal analysis regions are defined, depending on the invariant mass ($m_{\mu\mu}$) of the muon pair:

- Higgs Signal Region: $m_{\mu\mu} \in [115, 135]$ GeV
- Higgs Side-Band Region: $m_{\mu\mu} \in [110, 115] \cup [135, 150]$ GeV
- Z Control Region: $m_{\mu\mu} \in [76, 106]$ GeV

The Z control region and Higgs Side-Band regions are used to control the background shapes and normalization. The Higgs signal region is used for extracting the final signal yields and is blinded in the initial phase of the analysis.

5.6.1 VBF specific kinematic variables

To build a multi-variate discriminant for the VBF signal, several kinematic variables were considered to separate the VBF signature from the Drell-Yan / Electro-weak Z background. They are discussed in details in this section.

5.6.1.1 Di-muon kinematics

We consider the following kinematic variables from the dimuon system:

- The invariant mass of the two selected muons $m(\mu\mu)$: this tends to peak near the Higgs mass for the signal, whereas background processes have a more smoothly falling $m(\mu\mu)$ spectrum.
- The transverse momentum of the dimuon system $p_T(\mu\mu)$.
- The pseudorapidity of the dimuon system $\eta(\mu\mu)$: The dimuon system from the VBF signal process is mostly produced centrally in the detector, unlike other background processes.
- The dimuon mass resolution $\delta M(\mu\mu)$ or the relative mass resolution $\delta M(\mu\mu)/M(\mu\mu)$.

These variables are illustrated in Fig. 5.6 for both the signal and major background processes.

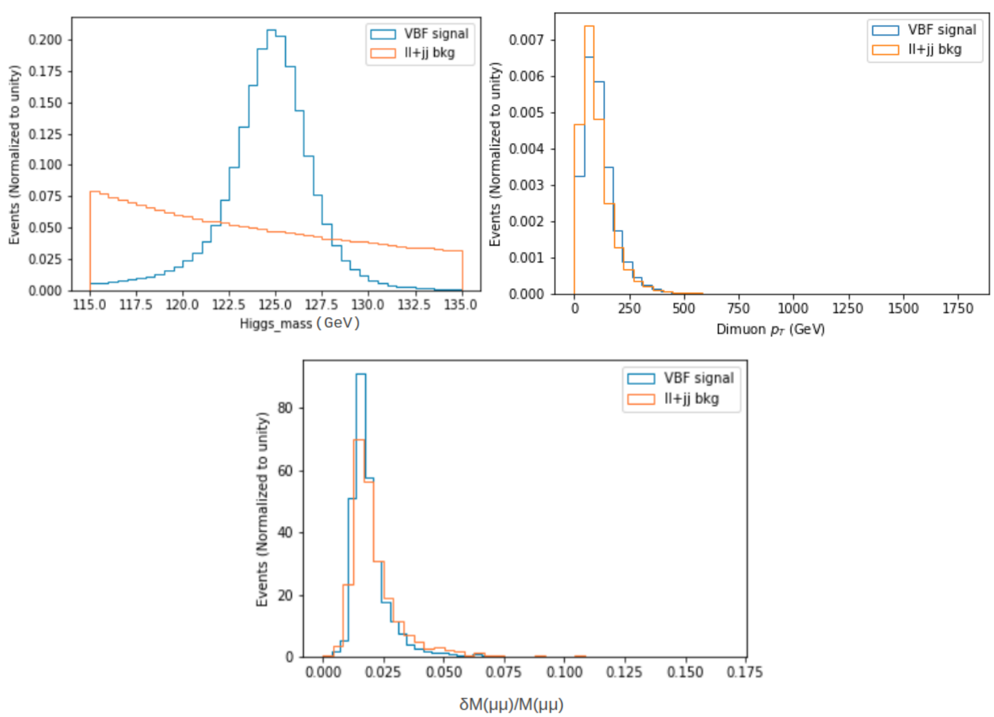


Figure 5.6: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange). (Top Left) $m(\mu\mu)$, (Top Right) $p_T(\mu\mu)$ and (Bottom) $\delta M(\mu\mu)/M(\mu\mu)$.

5.6.1.2 Leading and sub-leading jet kinematics

We consider the p_T, η and ϕ from the two leading candidate VBF-like jets. Some of these variables are illustrated in Fig. 5.7 for both the signal and major background processes.

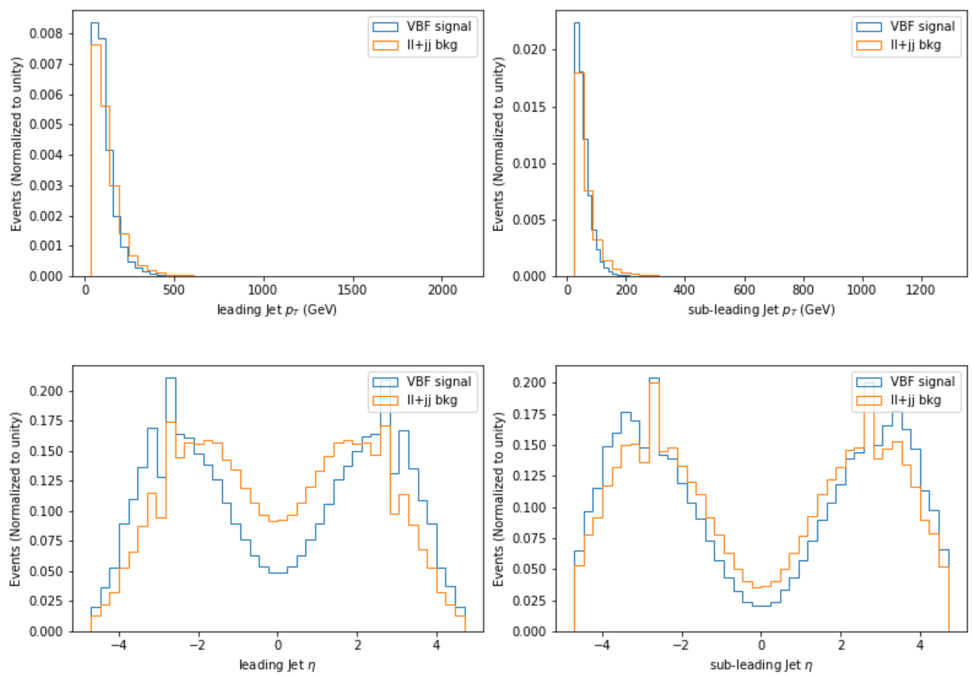


Figure 5.7: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange). (Top Left) $p_T(j_1)$, (Top Right) $p_T(j_2)$, (Bottom Left) $\eta(j_1)$ and (Bottom Right) $\eta(j_2)$.

5.6.1.3 Di-jet kinematics

We consider the following kinematic variables from the dijet system:

- The invariant mass of the two selected jets $m(jj)$: tends to have higher values for the signal process as compared to other background processes.
- The pseudorapidity difference between the two selected jets $|\Delta\eta(jj)| = |\eta(j_1) - \eta(j_2)|$: This difference is higher for VBF-like jets that are produced back-to-back in the forward regions of the detector.

  These variables are illustrated in Fig. 5.8 for both the signal and major background processes.

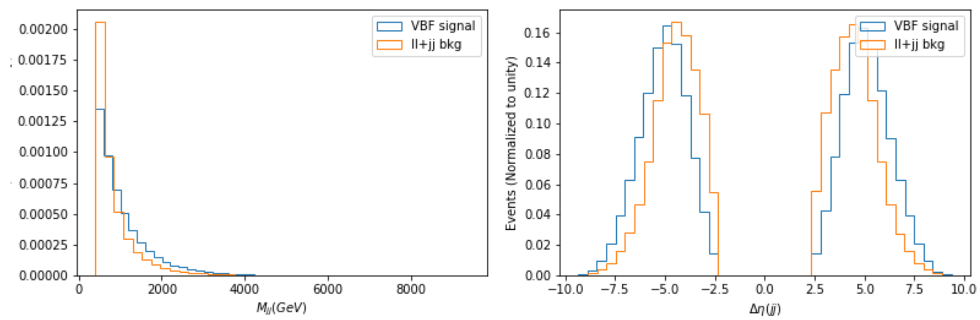


Figure 5.8: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange). (Left) $m(jj)$ and (Right) $|\Delta\eta(jj)|$.

5.6.1.4 Kinematics of the dimuon + dijet system

We consider the following kinematic variables from the dimuon + dijet system:

- The mass of the system composed by the two jets and the two muons is given by $M_{\mu\mu jj}$.

- The Zeppenfeld variable defined by

$$z^{\star} = \frac{y^{\star}}{\mid \eta(j_1) - \eta(j_2) \mid}$$

  where

$$y^{\star} = \eta(\mu\mu) - \frac{\eta(j_1) + \eta(j_2)}{2}.$$

  This variable checks the "centrality" of the dimuon system w.r.t. the two VBF jets, with a value of $|z^*| < 0.5$ when the dimuon pseudorapidity falls in between the two jet pseudorapidities, and $|z^*| > 0.5$ otherwise.

- The transverse momentum balance defined by

$$R(p_{\mathrm{T}}) = \frac{(\vec{p}(j_1) + \vec{p}(j_2) + \vec{p}(\mu_1) + \vec{p}(\mu_2))_T}{p_{\mathrm{T}}(j_1) + p_{\mathrm{T}}(j_2) + p_{\mathrm{T}}(\mu\mu)}.$$

  Background processes tend to have longer tails in this variable than VBF events, as background events may contain additional decay products apart from the selected muons and jets.

- The minimum of the pseudorapidity difference between the selected jets (individually) and the dimuon system

$$\min(|\eta(j_1) - \eta(\mu\mu)|, |\eta(j_2) - \eta(\mu\mu)|).$$

These variables are illustrated in Fig. 5.9 for both the signal and major background processes.

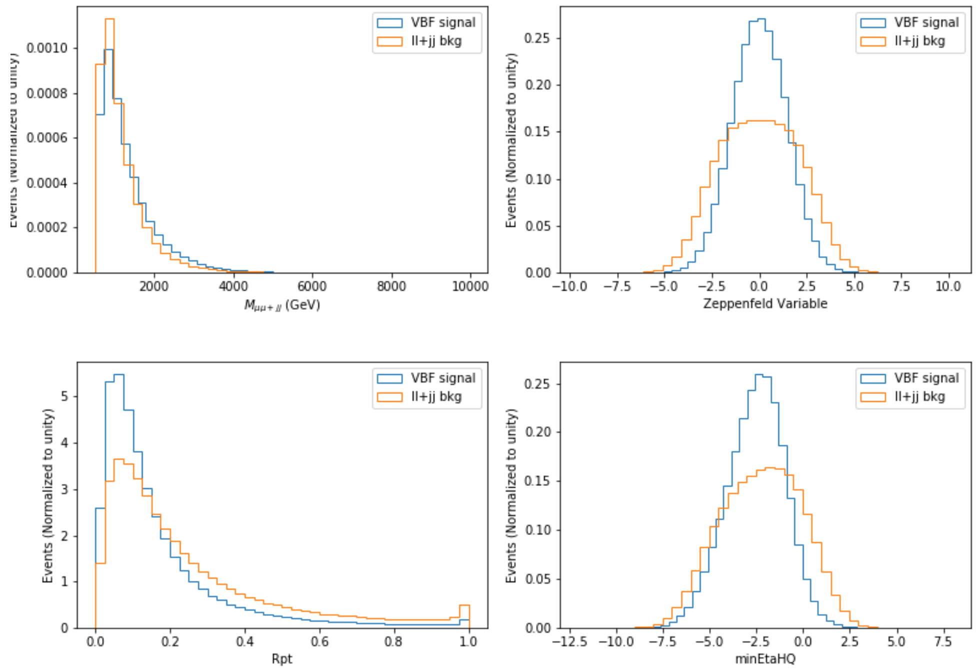


Figure 5.9: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange). (Top left) $p_T(\mu\mu jj)$, (Top right) $z^\star$, (Bottom left) $R(p_\mathrm{T})$ and (Bottom right) $\min(|\eta(j_1) - \eta(\mu\mu)|, |\eta(j_2) - \eta(\mu\mu)|)$.

5.6.1.5 Quark-Gluon Likelihood discriminator for jets

Jets in VBF events are expected to originate solely from quarks whereas events produced via other mechanisms can have jets that originate from both gluons or quarks. The Quark-Gluon likelihood (QGL) [148] is a probability tagger that discriminates gluon-jets from quark-jets using jet constituent information. VBF events usually have high QGL values and the QGL distributions peak at 1 for both leading and subleading jets. The background shapes can peak both at 0 and 1. However, a gluon jet most likely originates from an initial or final state radiation and tends to have a softer p_T spectrum. Therefore, sub-leading jets in background processes often come from gluon hadronization and thus, the QGL output of the subleading jet is more discriminating than the leading one for our scenario. Data-driven polynomial corrections to the QGL shape are also applied to the event weights (derived using Z+jets and dijet events in [148]).

These variables are illustrated in Fig. 5.10 for both the signal and major background processes.

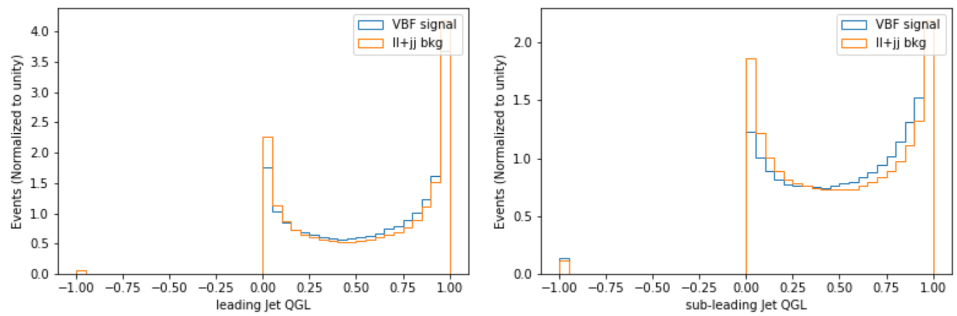


Figure 5.10: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange). (Left) QGL(j_1) and (Right) QGL(j_2). Some jets have spurious QGL values of -1, which indicates the jet had very few constituents for a QGL value to be calculated effectively.

5.6.1.6 Soft activity variables

We consider the following soft-jet kinematic variables in the event:

- N_5^{soft} or the number of track jets (see Section 5.3.7) with $p_T > 5\,\text{GeV}$: These jets are only comprised of charged tracks that do not belong to the selected muons or VBF-jets. Such a selection helps to distinguish the signal process from other backgrounds, since the signal VBF process has very limited extra hadronic activity.

- $H_{T\,x}^{\text{soft}}$ or the scalar sum of the transverse momenta of all the track jets with p_{T} > x GeV, where $x \in [2, 5, 10]$: this also checks for additional soft activity in the event.

  These variables are illustrated in Fig. 5.11 for both the signal and major background processes.

5.6.1.7 Collins-Soper frame variables

As discussed previously, the Drell-Yan process is an irreducible background for $H \to \mu\mu$. However, as the Z/γ boson is a spin-1 particle and the Higgs is a spin-0 particle, the angular distributions of the two muons produced are different in the two processes. In particular, if we look at the frame where the original boson is

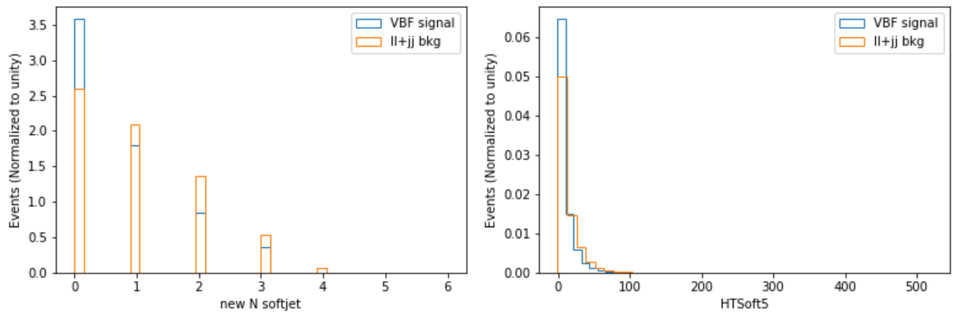


Figure 5.11: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange). (Left) N_5^{soft} and (Right) $H_{T\,5}^{\text{soft}}$.

produced at rest (and the two muons fly back-to-back), we could reconstruct certain angular variables that could help us distinguish between the two processes. This frame is known as the Collin-Soper frame [149] and is illustrated in Fig. 5.12. The hadron plane is defined as the plane containing the two colliding partons. The y-axis is chosen as the normal vector to the hadron plane, and the z-axis is chosen such that it bisects the angle 2β between the vectors of the two colliding partons. The sign of the z-axis is taken as the sign of the z-component of the dimuon momentum in the laboratory frame. The x-axis is chosen accordingly (for a right-handed cartesian coordinate system). The decay products of a spin-0 particle have no preferential direction and this result in a constant angular distribution for muons originating from a Higgs decay. The dimuon system from the Drell-Yan process is however affected by the spin-1 nature the Z/γ and have an angular distribution roughly described by

$$\frac{d\sigma}{d\Omega} \propto 1 + \cos^2 \theta_{CS}$$

where θ_{CS} is shown in Fig. 5.12.

For our purposes here, we consider the following two angles:

- $\cos(\theta_{CS})$: The cosine of the angle between the collinear muons in the dimuon rest frame (green plane) and the z-axis
- $\cos(\phi_{CS})$: The cosine of the angle between the lepton-plane and the hadron-plane.

The $\cos(\theta_{CS})$ variable is illustrated in Fig. 5.13 for both the signal and major background processes.

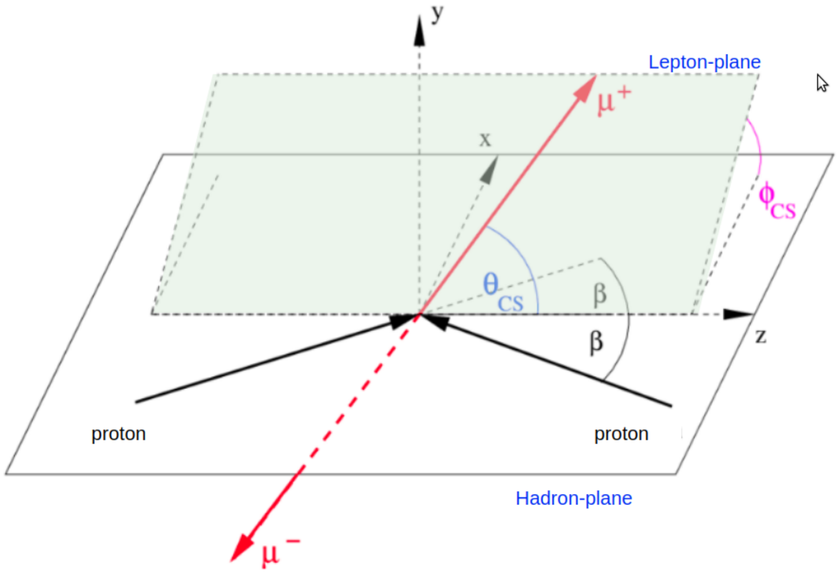


Figure 5.12: An illustration of the Collins-Soper frame.

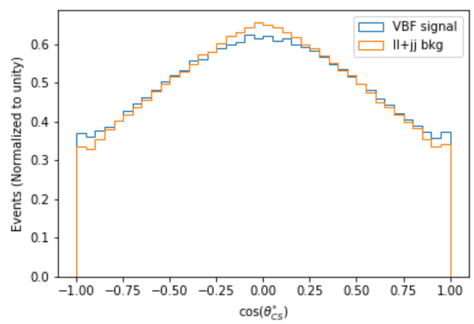


Figure 5.13: Variable comparisons between VBF Higgs signal (blue) and Drell-Yan + Electro-weak Z backgrounds (orange): $\cos(\theta_{CS})$.

Each variable described in this section has been studied in detail in the Z control region for the three data taking periods and any observed differences between the data and the MC were corrected for according to Section 5.3 and 5.4 (see Fig. 5.14, 5.15, and 5.16). Some residual mis-modelling of the simulation compared to data are observed, but are known to be covered by the uncertainties due to jet energy scale and resolution.

5.6.2 Picking a discriminator – BDT vs DNN

Two preliminary Machine Learning algorithms: a Boosted Decision Tree (BDT) and a Deep Neural Network (DNN), were trained to estimate sensitivities, maximize the signal to background separation and compare performances. These preliminary

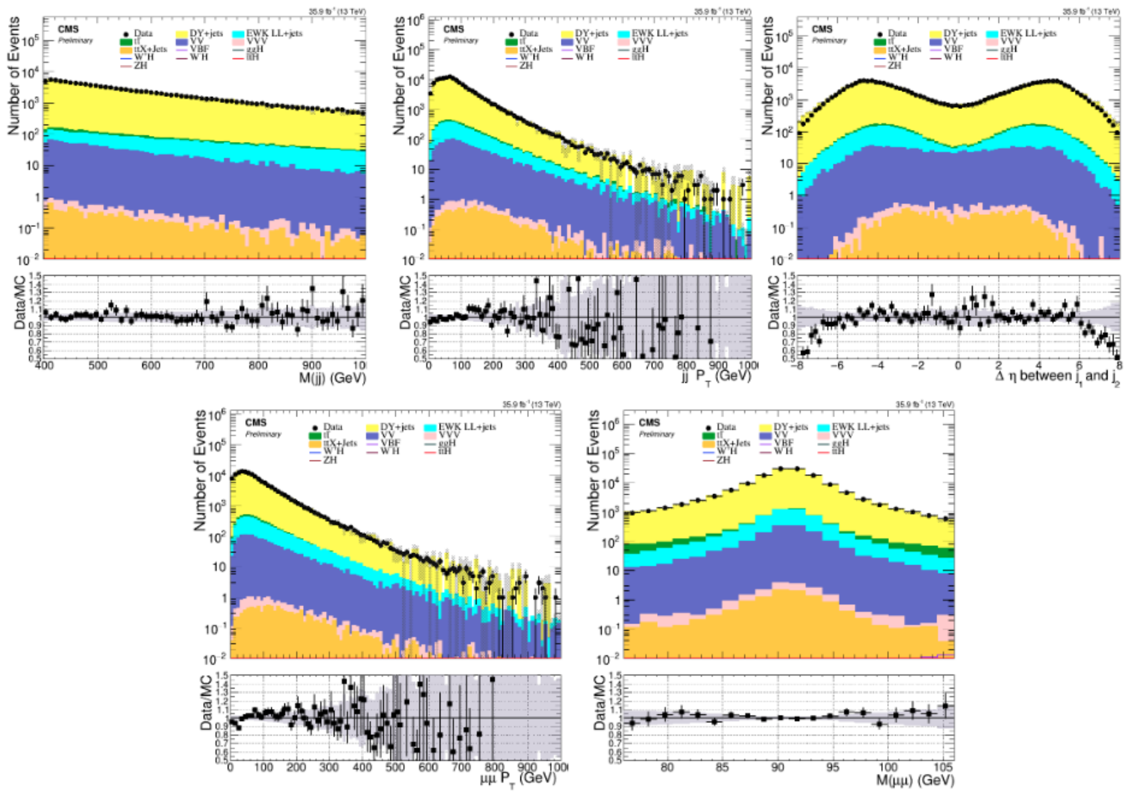


Figure 5.14: Data/MC comparisons for some of the VBF discriminating variables in 2016. (Top) $m(jj)$ (left), p_{T} (jj) (center) and $\eta(jj)$ (right). (Bottom) $p_T(\mu\mu)$ (left) and $M(\mu\mu)$ (right).

discriminators are described in this section. Based on the performance comparison of these discriminators, the final choice to use a DNN for the VBF category was made. The final version of the DNN discriminator used in this analysis is described later in Section 5.6.3.

5.6.2.1 BDT architecture

A preliminary BDT was trained using simulated samples for the three years all mixed together. The simulated samples used in the training are:

- Signal: VBF $H \to \mu\mu$ and ggH $H \to \mu\mu$ with m_H=125.0 GeV
- Background: Drell-Yan Z and Electro-weak Z

The following input variables were used to train the model (for description, see Section 5.6.1):

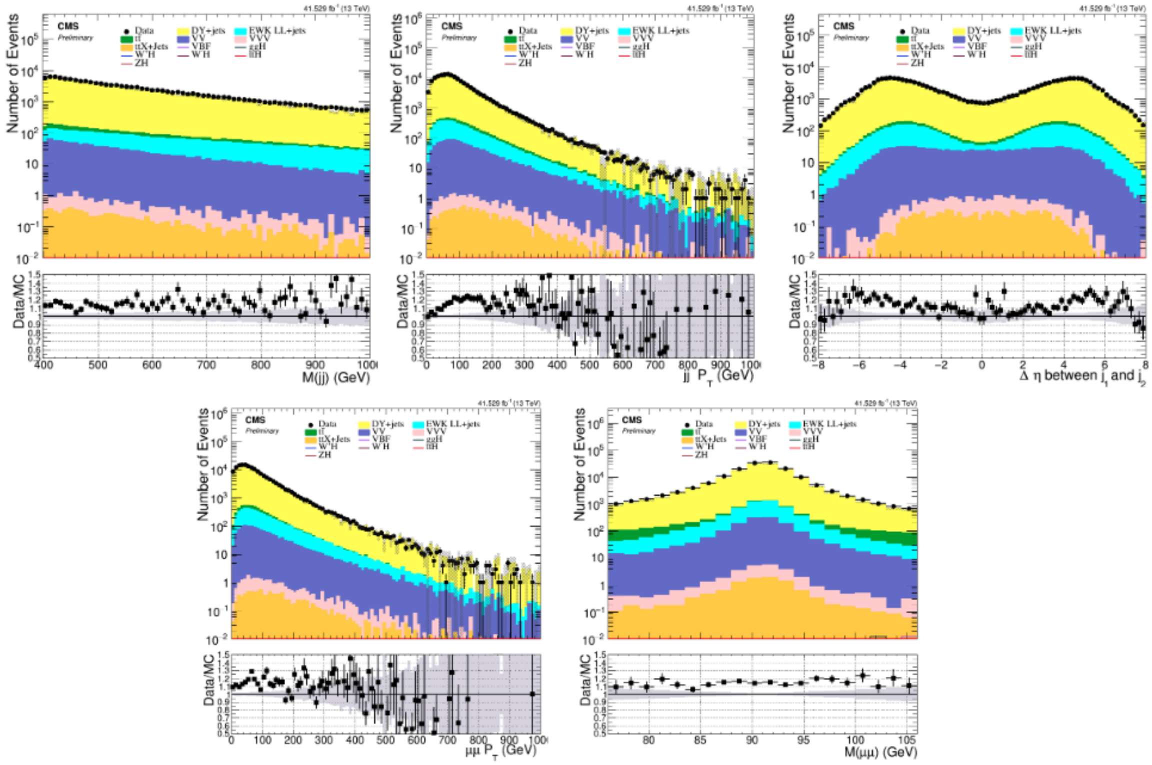


Figure 5.15: Data/MC comparisons for some of the VBF discriminating variables in 2017. (Top) $m(jj)$ (left), p_{T} (jj) (center) and $\eta(jj)$ (right). (Bottom) $p_T(\mu\mu)$ (left) and $M(\mu\mu)$ (right).

- $M_{\mu\mu}$, $p_T(\mu\mu)$ and rapidity $y_{\mu\mu}$ of the dimuon pair
- $m(jj)$
- $R_{p_{\mathrm{T}}}$
- p_{T} centrality $= \frac{p_{\mathrm{T}}^{\mu\mu} - |\vec{p_{\mathrm{T}}}^{j_1} + \vec{p_{\mathrm{T}}}^{j_2}|/2}{|\vec{p_{\mathrm{T}}}^{j_1} - \vec{p_{\mathrm{T}}}^{j_2}|}$
- Single muon variables: $p_{\mathrm{T}}^{\mu_1}/m_{\mu\mu}, p_{\mathrm{T}}^{\mu_2}/m_{\mu\mu}, \eta^{\mu_1}, \eta^{\mu_2}$
- $\Delta\eta(jj), \Delta\phi(jj)$
- z^*
- p_T, η for the two leading jets
- $\cos(\theta_{CS})$, $\cos(\phi_{CS})$
- $\min(|\eta(j_1) - \eta(\mu\mu)|, |\eta(j_2) - \eta(\mu\mu)|), \min(|\phi(j_1) - \phi(\mu\mu)|, |\phi(j_2) - \phi(\mu\mu)|)$
- Jet multiplicity: N_{jets}

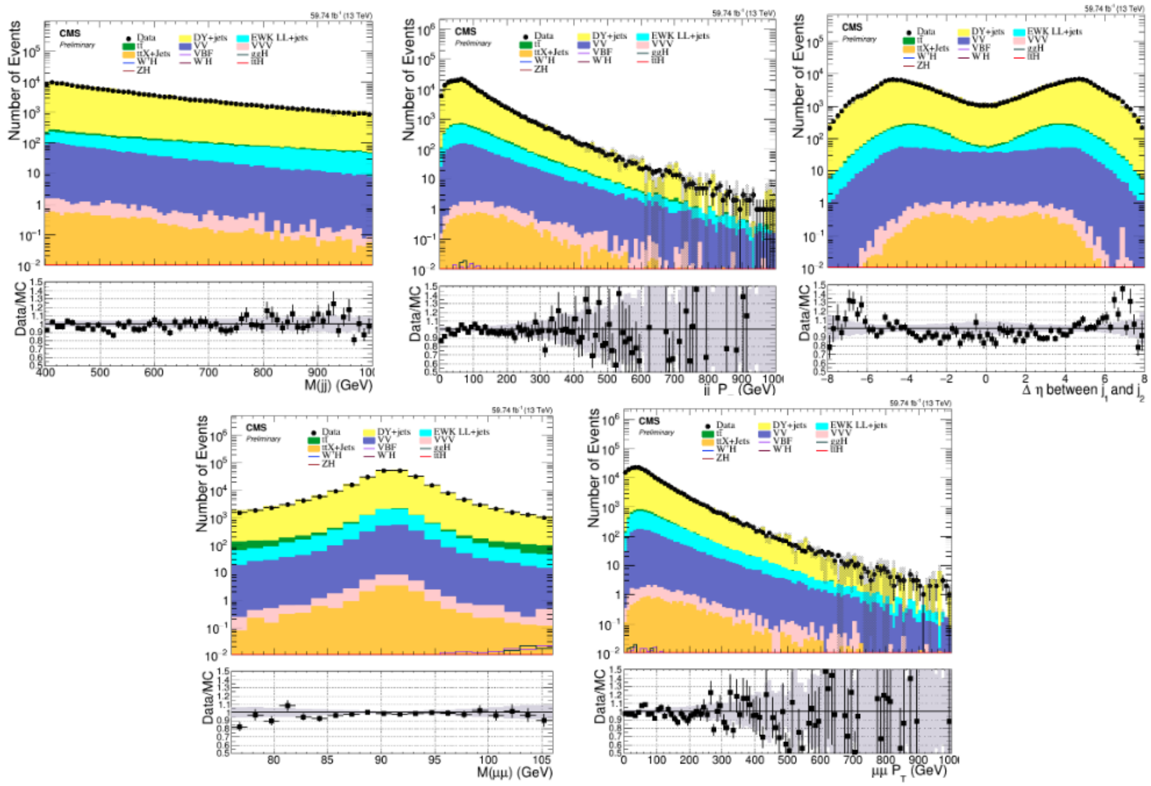


Figure 5.16: Data/MC comparisons for some of the VBF discriminating variables in 2018. (Top) $m(jj)$ (left), p_{T} (jj) (center) and $\eta(jj)$ (right). (Bottom) $p_T(\mu\mu)$ (left) and $M(\mu\mu)$ (right).

- $H_{T\ 2}^{\text{soft}}$
- N_2^{soft}

The BDT training is made aware of the dimuon mass resolution by weighting the signal events proportionally to $\frac{1}{\sigma_{\mu\mu}}$, where $\sigma_{\mu\mu}$ is the calibrated per-event dimuon mass resolution. The $\sigma_{\mu\mu}$ value is not used as an input to the MVA, but only as a weighting factor in the training. The weight is not applied in the evaluation of the MVA score. The BDT is trained using the *Gradient Boost* method. The training is done on 50% of the available simulated events, and the remaining is used for testing purposes. The parameters used for the BDT training are as follows:

- Number of trees = 1000
- Minimum node size = 3%
- Shrinkage = 0.10

- Bagged sample fraction = 0.5
- Number of cuts = 30
- Maximum depth = 4
- Transformation for inputs: (I, N)
- Separation type: Cross-Entropy

5.6.2.2 DNN architecture

A preliminary DNN was trained using simulated samples for the three years all mixed together. The simulated samples used in the training are:

- Signal: VBF $H \to \mu\mu$ and ggH $H \to \mu\mu$ with m_H=125.0 GeV
- Background: Drell-Yan Z and Electro-weak Z

The following 26 input variables were used to train the model (for description, see Section 5.6.1):

- $\eta(\mu\mu)$, $M_{\mu\mu}$, $\delta M(\mu\mu)$,$\delta M(\mu\mu)/M(\mu\mu)$, $p_T(\mu\mu)$ of the dimuon pair
- $m(jj)$, η, ϕ and p_T of the leading dijet pair.
- $\Delta\eta(jj)$
- Mass, η and ϕ of the leading dijet+dimuon pair.
- z^*
- p_T, η and QGL for the two leading jets
- cos(θ_{CS})
- $\Delta\eta(\mu\mu, j_1)$, $\Delta\eta(\mu\mu, j_2)$, $\Delta\phi(\mu\mu, j_1)$ and $\Delta\phi(\mu\mu, j_2)$
- $H_{T\,5}^{\text{soft}}$.

The network was trained with 3 hidden layers and 100 nodes per layer. The `Adam` optimizer was used with a learning rate of 10^{-5}. The `tanh` activation function was used for the inner layers and the `sigmoid` activation was used for the final

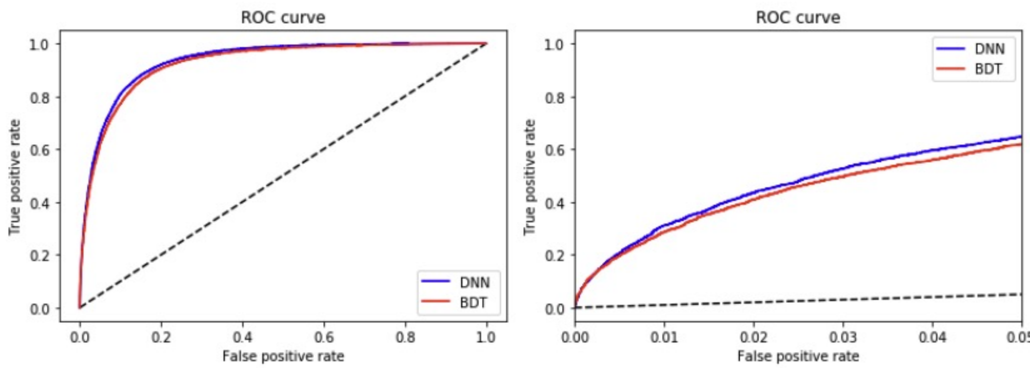


Figure 5.17: BDT vs DNN performance. (Left) A ROC curve comparison of the performances of the preliminary BDT and DNN. (Right) A zoomed in version of the ROC curve. The DNN performs slightly better than the BDT.

output. The loss function used was the `binary cross-entropy`. A 20% drop-out rate and batch-normalization were also used to regularize the training. 60% of the events were used for training and validation, and the remaining 40% were used for testing. The events were weighted during training according to their cross-section. The weights for background events were modified and brought to $O(1)$, in a way such that the relative weight of any process in the background class was preserved (i.e., Electro-weak Z is still rarer than Drell-Yan). Similarly, the signal weights were modified such that the sum of weights of signal events was the sum total of weights of all background processes. Additionally, all input features to the DNN were standardized to have a mean of 0 and standard deviation of 1. Finally, the training was done on events with $m_{\mu\mu} \in [115, 135]$ GeV.

5.6.2.3 Performance comparison

The input variables for the preliminary BDT (Sec. 5.6.2.1) and DNN (Sec. 5.6.2.2) were slightly different, however, the major inputs that contribute to most of the discrimination power were the same ($M(\mu\mu)$, $p_T(\mu\mu)$, $\Delta\eta(jj)$, etc). A comparison of the ROC curves of the preliminary BDT and DNN showed that the DNN performance was better by $\sim$ 5% at 0.1 background efficiency, and $\sim$ 9% at 0.01 background efficiency (see Fig. 5.17 and Table 5.2). Thus, it was concluded that the DNN can separate the signal better from the background and was chosen as the final discriminator. While it is difficult to quantify exactly why the DNN outperforms the BDT, a simple explanation is that the BDT is too "shallow" to capture all the features in the high dimensional VBF phase-space.

Bkg efficiency	Bkg Yield (DY + VBF-Z)	Signal Yield (VBF + ggH)	
		BDT	DNN
0.1	339.36 ± 5.98	8.21 ± 0.05	8.59 ± 0.05
0.01	33.96 ± 1.65	3.03 ± 0.03	3.31 ± 0.03

Table 5.2: The expected signal and background yields for different background efficiency working points for the preliminary BDT and DNN.

5.6.3 The final Deep Neural Network for the VBF category

The final DNN for the VBF category is designed to be more robust and powerful than the one described in Sec. 5.6.2.2, and has a more complex architecture. The training is performed using simulated samples for the three years all mixed together. The motivation behind this is to exploit the higher statistics of the combined samples, compared to performing a dedicated training per year. To account for possible discrepancies in the simulation of different years (due to different detector conditions), the variable "year" is added as an input to the training and serves as a flag.

The simulated samples used in the training are:

- Signal: VBF $H \to \mu\mu$ with m_H=125.0 GeV
- Background: Drell-Yan Z, Electro-weak Z and $t\bar{t}$.

In order to preserve the statistical power of the sample in the final evaluation, a 4-fold procedure is used during training. For every fold, half of the sample size is used for training, a quarter is used for the validation and the remaining quarter is used for testing, i.e., for the final evaluation and extraction of the significance. This procedure is performed for all the samples used in the training. Thus, we end up having 4 slightly different DNN models (with the same inputs; but different training sub-sets). This allows to partially regularize the final result, as the fluctuations of individual subsets of events are reduced by a factor 2. The samples are divided into the four blocks as "`Event number % 4`". The same procedure is also used for samples not used in training (e.g., data and other MC samples), in order to decide which out of the four different models will be used for evaluating the DNN output score for a particular event. A schematic representation of the k-fold procedure with 4 folds is shown in Fig. 5.18.

The events are weighted during training according to their cross-section. The weights for background events are then modified and brought to $O(1)$, in a way

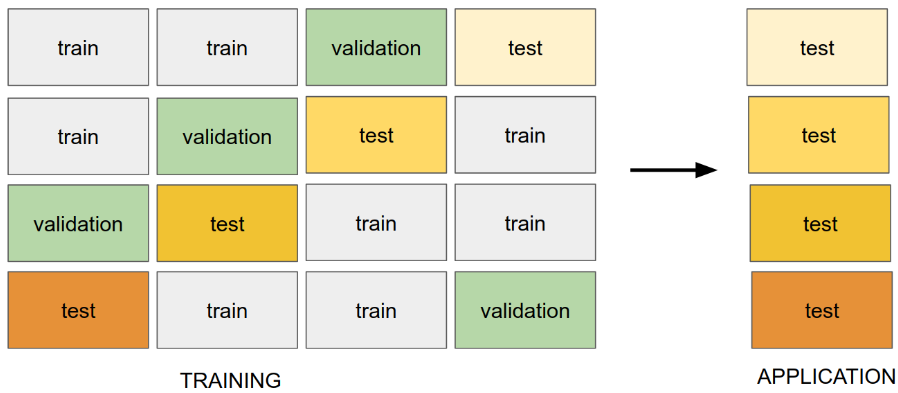


Figure 5.18: Scheme of the 4-fold training, validation and evaluation procedure.

such that the relative weight of any process in the background class is preserved (i.e., $t\bar{t}$ is still rarer than Drell-Yan). Similarly, the signal weights are modified such that the sum of weights of signal events is the sum total of weights of all background processes. This re-weighting is done in order to make the network roughly understand the "strength" or cross-section of each process, without giving too much importance to the background over the signal (since the signal weighted to cross section is a ~100 times smaller than the background). Additionally, all input features to the DNN are standardized. The sample mean (calculated from the training set) is subtracted from each value (per feature) and the result is divided by the sample standard deviation (per feature; also calculated from the training set). As a result, the new input variable distributions (for train/validation/test subsets) have a mean of 0 and standard deviation of 1.

The input variables used for the DNN are (for description, see Section 5.6.1):

- $m(\mu\mu)$, $\Delta m(\mu\mu)$, $\frac{\Delta m(\mu\mu)}{m(\mu\mu)}$
- $m(jj)$, $\log m_{jj}$
- $R(p_T)$
- z^*
- $\cos(\theta_{CS})$, $\cos(\phi_{CS})$
- $\Delta\eta(jj)$
- N_5^{soft}

- $H_{T\,2}^{\text{soft}}$
- $\min(|\eta(j_1) - \eta(\mu\mu)|, |\eta(j_2) - \eta(\mu\mu)|)$
- $p_T(\mu\mu)$, $\log p_T(\mu\mu)$, $\eta(\mu\mu)$
- $p_T(j_1)$, $p_T(j_2)$, $\eta(j_1)$, $\eta(j_2)$, $\phi(j_1)$, $\phi(j_2)$
- $qgl(j_1)$, $qgl(j_2)$
- year: the data-taking period.

The signal region distributions for these variables are shown in Fig. 5.19, 5.20, 5.21, 5.22, 5.23, 5.24, 5.25, and 5.26.

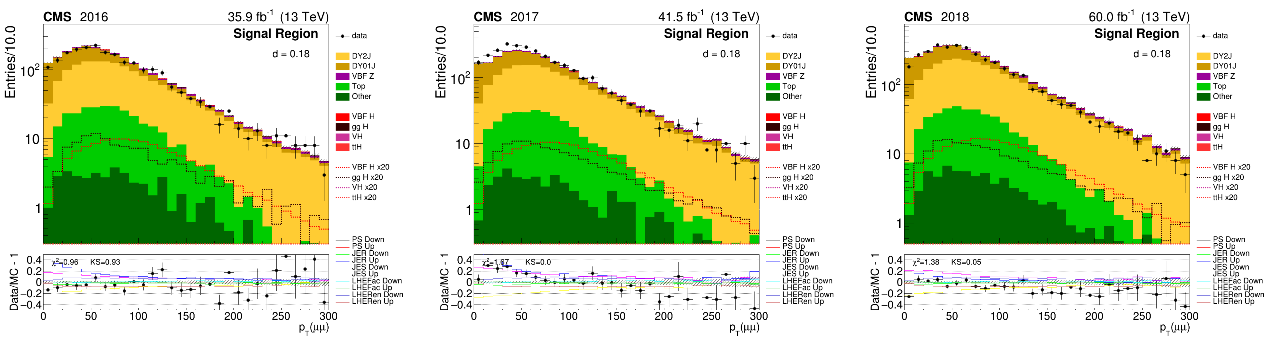


Figure 5.19: Transverse momentum distribution of the dimuon system after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

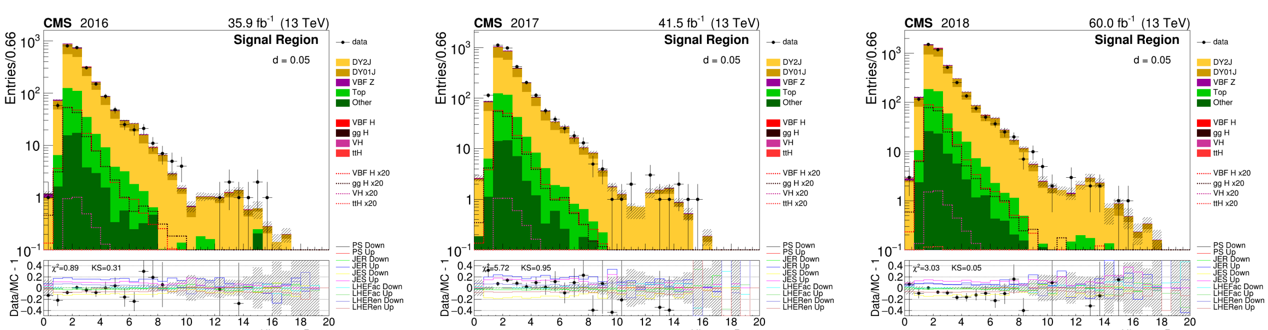


Figure 5.20: Dimuon mass uncertainty $\Delta m_{\mu\mu}$ after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

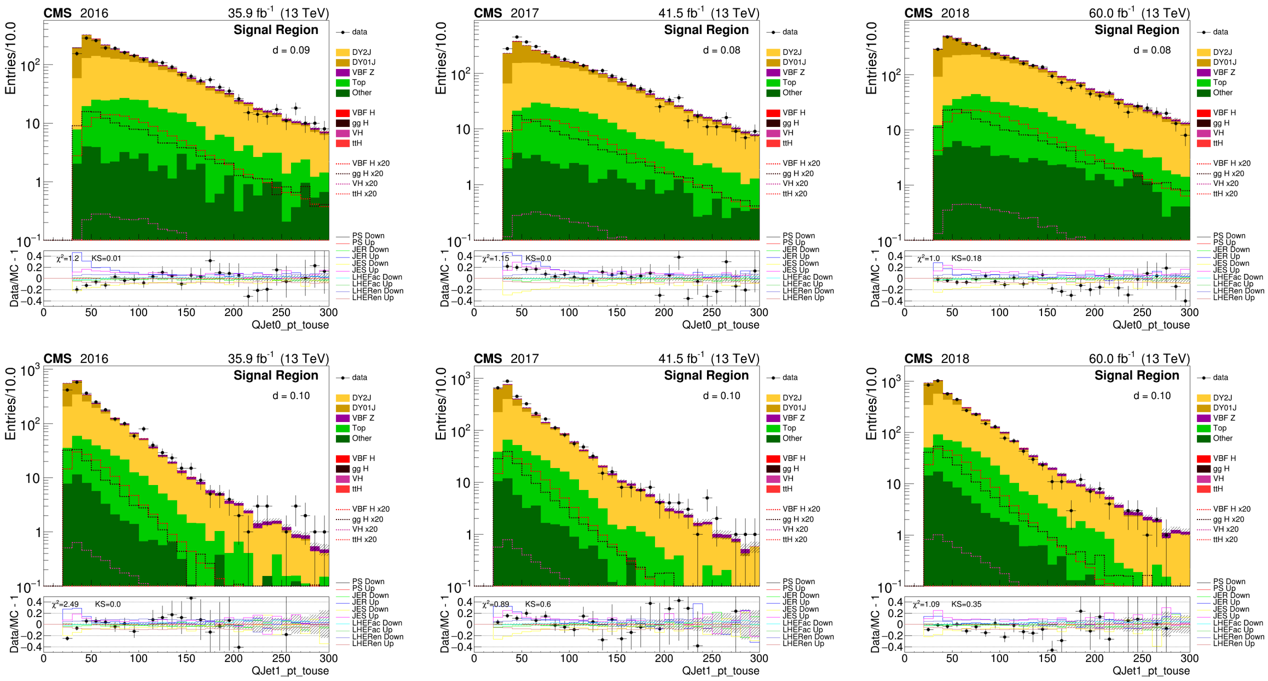


Figure 5.21: Transverse momentum distributions for leading (top) and subleading (bottom) jets after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

The neural network training is performed in multiple steps. Four networks are first optimized independently with different inputs and for different tasks. These independent networks target different backgrounds or event topology, and are described by the following goals:

1. signal vs Electro-weak Z (VBF Z) : The VBF Z process is the most signal-like background and a dedicated network with 3 hidden layers is trained in order to separate signal from this background.

2. signal vs DY : A network with 3 hidden layers is dedicated to separate signal from DY events.

3. mass independent signal -vs- background : A network with 3 hidden layers is trained with the 22 input variables uncorrelated with the dimuon mass. This network is trained against all three backgrounds.

4. mass + mass resolution signal vs background : A network with two hidden layers is trained using only three input variables: the dimuon mass and its

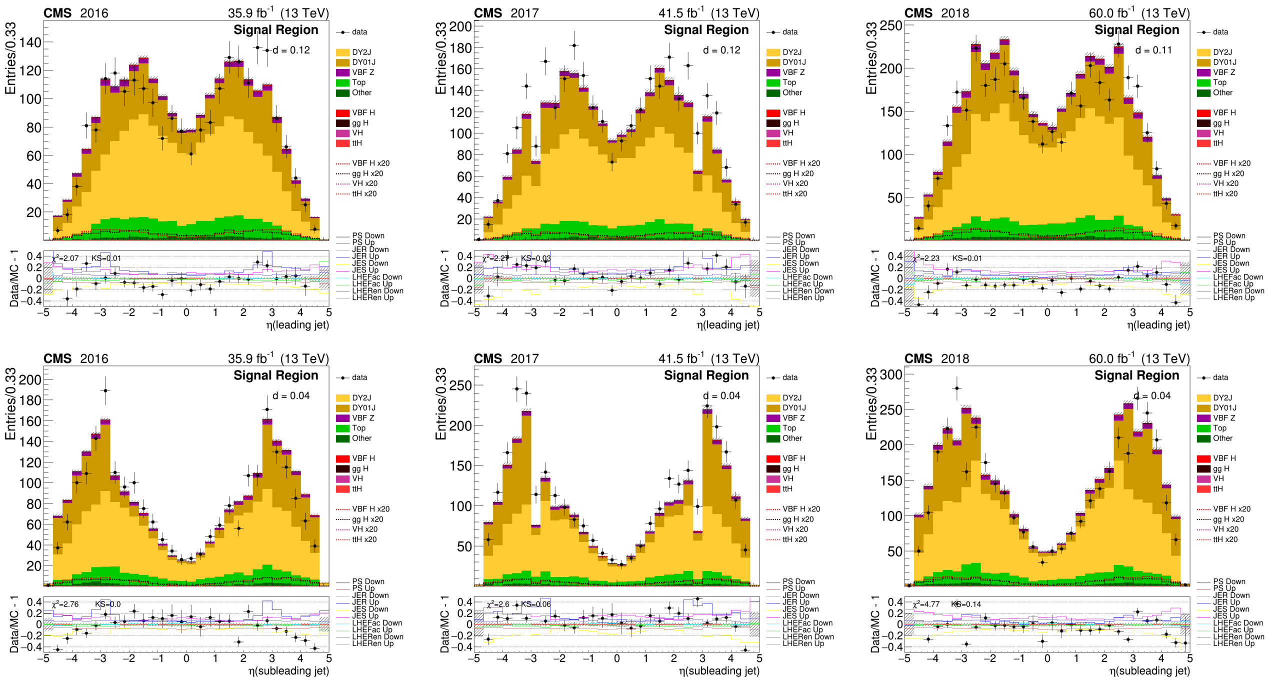


Figure 5.22: Pseudorapidity distributions for leading (top) and subleading (bottom) jets after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

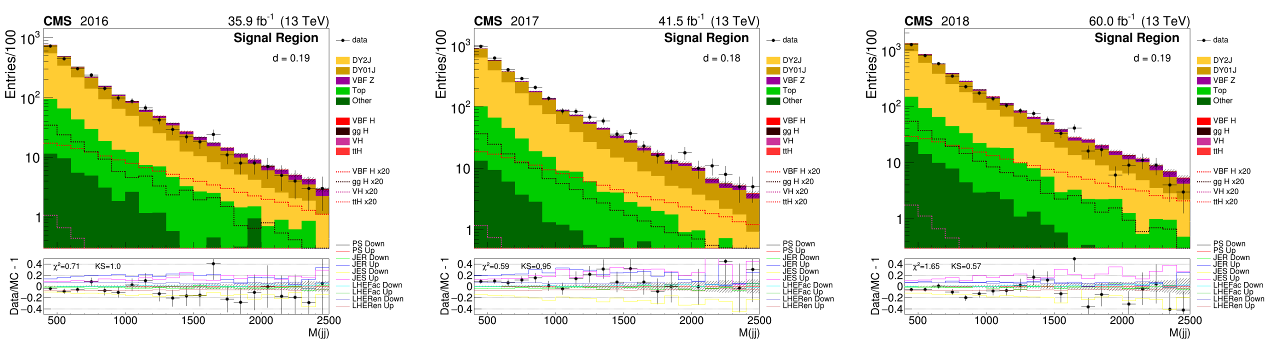


Figure 5.23: Dijet invariant mass distributions after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

absolute and relative resolutions. This network is trained against all three backgrounds.

The last hidden layer of all the four networks are then merged and combined to build the final classifier. A final, fifth network with 2 hidden layers is fine-tuned as

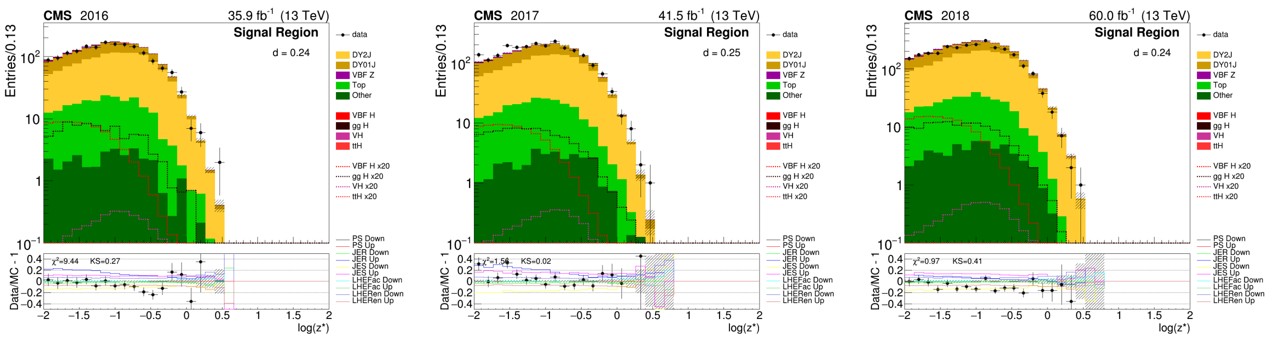


Figure 5.24: Distributions of the *Zeppenfeld* variable z^* after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

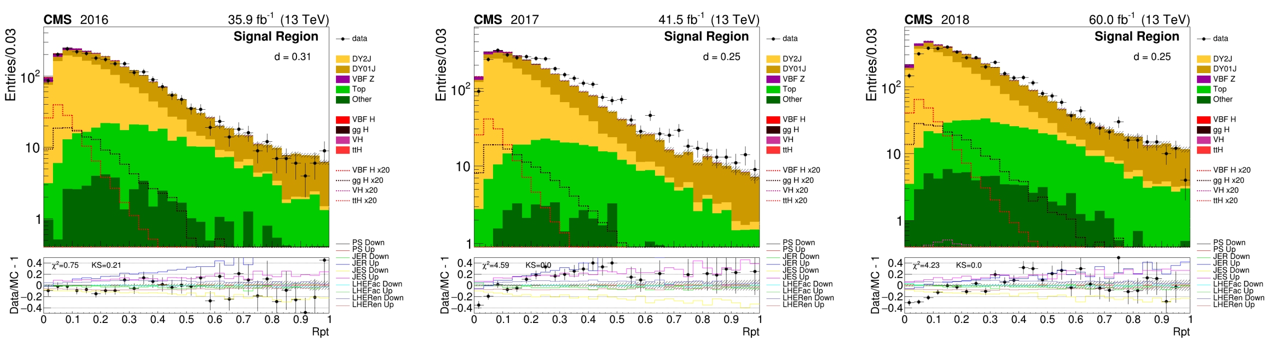


Figure 5.25: Distributions of the transverse momentum balance $R(p_{\mathrm{T}})$ after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

- Step 1: the node weights of the networks 1, 2, and 3 are fixed to values obtained by the previous training while the network (4) weights are free to float.

- Step 2: the node weights of the network 3 are also free to float.

This architecture is visually represented in Fig. 5.27. The orange blocks denote the input features, the grey blocks indicate the DNNs optimized for their specific tasks, with their outputs in blue. The last hidden layer from the 4 networks are merged into a single vector and is used as an input for the combination (5th) network, whose output is shown in red.

Each network has a few dozen nodes in each hidden layer and employs a 20% dropout rate for regularizing the model. A batch size of 1024 events is used, except for the Step 2 in the final network, where 10240 events are used. Based on the validation

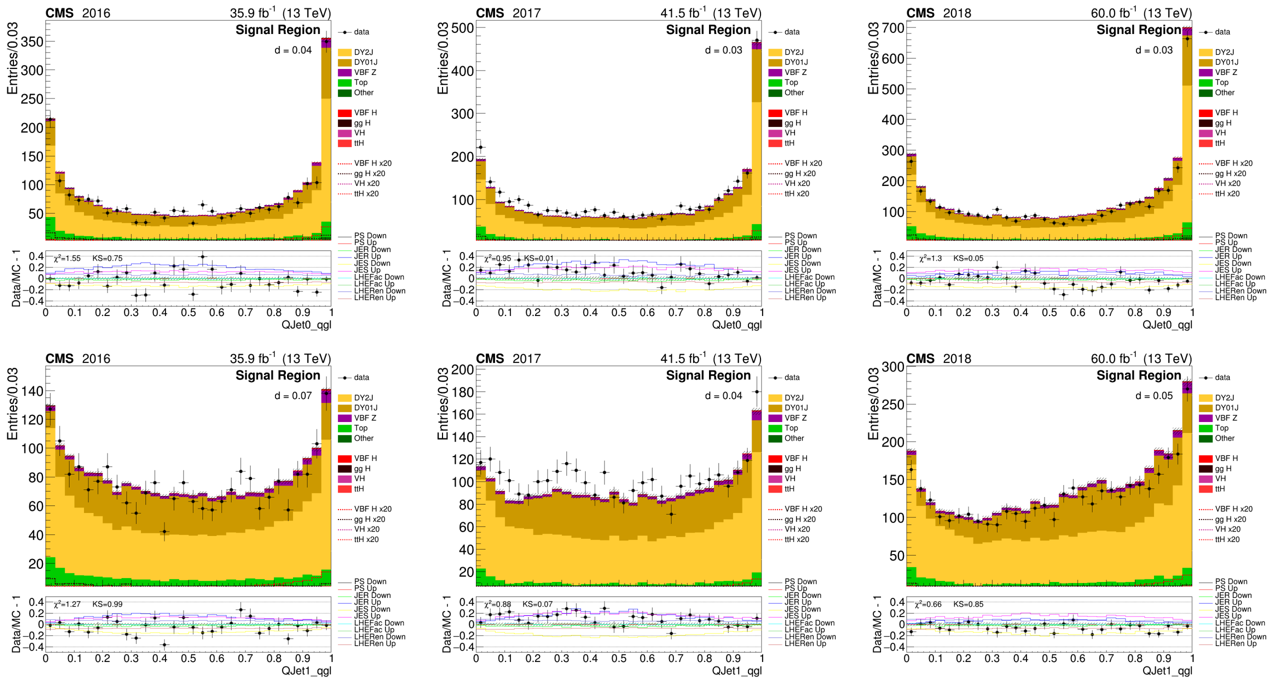


Figure 5.26: QGL output distributions for leading (top) and subleading (bottom) jets after the event selection in the Signal Region for 2016 (left), 2017 (center), and 2018 (right) [150].

loss, the learning rate is also gradually decreased over training epochs from a start value of 0.05. The loss function used at each step is the binary cross-entropy or $log - loss$, which is the maximum likelihood estimator for binary classification problems.

The best trained model is chosen using the estimated Asimov significance [151], which is computed both for the training and validation folds. The minimum significance is used as to pick the best model. The Asimov significance is given by

$$\text{Asimov significance} = \sqrt{\sum_{i=1}^{N} 2[(S_i + B_i)\log(1 - \frac{S_i}{B_i}) - S_i]} \tag{5.3}$$

where $S_i$and B_i are the expected of signal and background events in the i-th bin. The binning of the DNN (for significance calculation and final template fitting) is constructed as follows (starting from high output scores and moving towards lower scores):

- the bin has to contain at least 0.5 expected signal events,

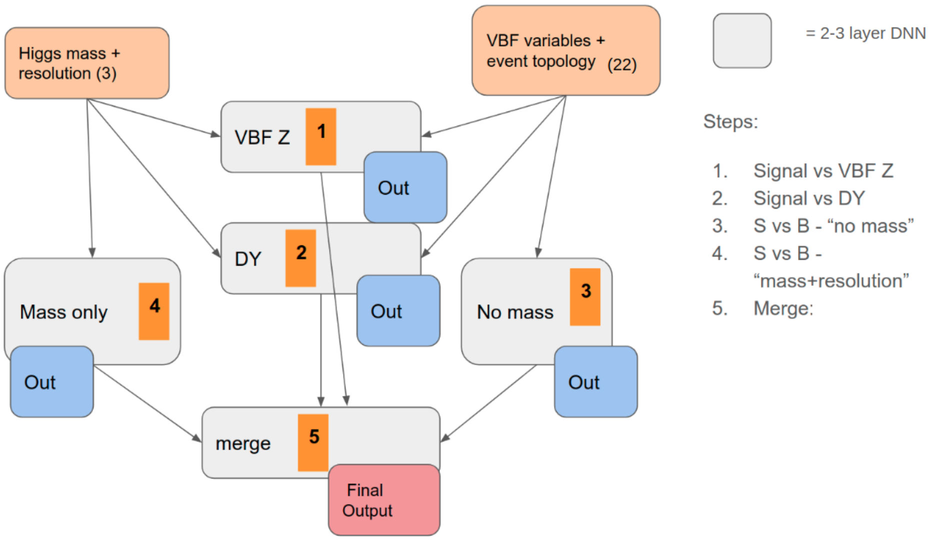


Figure 5.27: Schematic representation of the DNN architecture: the training procedure involves optimizing for individual tasks, combining the network outputs and fine-tuning the final model by unfreezing upstream weights.

- the bin yield of the background must have a relative error smaller than 30%,
- use the smallest bin width that satisfies the two previous conditions.

The signal and background DNN output distributions are shown in figure 5.28.

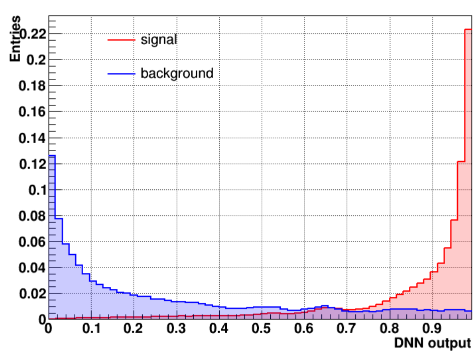


Figure 5.28: Plot of the signal and background normalized distributions for the DNN output score. The simulated samples of 2016, 2017, and 2018 are used all together.

In an alternative approach, Appendix C describes a mass agnostic machine learning based strategy for the VBF category. To perform a data-driven fit to the $m_{\mu\mu}$ distribution (unlike the template based fit to the DNN score output chosen for the VBF category), one must develop a $m_{\mu\mu}$ independent discriminator, to avoid sculpting the distribution of the backgrounds. However, for a simple DNN classifier, it is easy to learn the dimuon mass with the input kinematic variables, even if the mass is not explicitly given as an input to the training. Appendix C describes an adversarial training technique based on Ref. [152], which was used to develop a mass agnostic neural network.

5.6.4 Systematic uncertainties

In this section, we will discuss the systematic uncertainties considered for the VBF category, which are divided into experimental and theoretical sources. They can affect both the overall signal acceptance and the shape of the DNN output. Experimental uncertainties account for differences in true values of observables and their actual measurements. Theoretical uncertainties are related to the mis-modelling of the MC sample generation (due to missing higher order corrections, etc).

5.6.4.1 Experimental uncertainties

This section describes experimental uncertainties considered in this analysis.

- **Luminosity (normalization only)**: The luminosity affects the signal and background normalization. The uncertainty on the luminosity measurement is about 2.5% in each era [153–155], and is modelled via a log-Normal nuisance parameter whose effect is correlated across all DNN categories, processes, and eras.

- **Single muon trigger and muon selection efficiencies (normalization and shape)**: Uncertainties on the data driven corrections for the trigger, and offline muon selections amount to a total of 2-3%. Their effect is correlated across DNN bins, analysis regions, and eras.

- **Muon scale and resolution (normalization and shape):** The muon momentum scale and resolution correction uncertainties are computed by the Rochester method (see Sec. 5.3.2.1 and [136]). Their effect is correlated across DNN bins, analysis regions, and eras.

- **Pileup uncertainty (shape):** An uncertainty on the number of primary vertices is derived by varying the minimum bias cross section used in the pileup re-weighting applied to MC samples. The corresponding effect is correlated across DNN bins, analysis regions, and eras.

- **Jet energy scale (normalization and shape):** This uncertainty is obtained by varying the transverse momentum of each jet up and down by one standard deviation for each source of uncertainty, as recommended centrally for all CMS data analyses [139]. The full chain of event-selection to DNN evaluation is performed again, taking into account the shifted energies. The corresponding variations in each DNN bin yield for each process are used to estimate the uncertainty. This results in a set of uncertainties affecting both signal and background acceptance as well as the shape of the DNN output.

- **Jet energy resolution (normalization and shape):** This uncertainty is obtained by smearing the transverse momentum of each jet by the smearing factors provided centrally for all CMS data analyses [139]. Jets are divided in six exclusive categories:

  - central jets with $|\eta| < 1.93$.
  - central jets with $1.93 < |\eta| < 2.5$.
  - forward jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| < 3.139$.
  - forward jets with $p_{\mathrm{T}} > 50$ GeV and $|\eta| < 3.139$.
  - forward jets with $p_{\mathrm{T}} < 50$ GeV and $|\eta| > 3.139$.
  - forward jets with $p_{\mathrm{T}} > 50$ GeV and $|\eta| > 3.139$.

  The nuisance parameters across categories are set to be uncorrelated.

- **Quark-gluon likelihood (shape):** Uncertainties on the QGL discriminator are evaluated using data-driven polynominal corrections derived from Z+jets and dijet events, following [148]. The variations to the nominal distributions are computed by applying those corrections twice for up variation and not applying the corrections for the down variation. The corresponding effect is correlated across DNN bins, analysis regions, and eras.

- **Prefiring (normalization and shape):** The uncertainty due to the L1 ECAL pre-firing condition (Section. 5.4) affects only the 2016 and 2017 eras. This uncertainty varies from 0.3–1.5% (0.7–2%) as a function of the DNN score in the 2016 (2017) era, and its effect is correlated across DNN bins and regions.

- **MC Simulation size:** The per-bin statistical uncertainty arising from the limited size of the simulated samples is taken into account for all signal and background processes. This mostly affects the high score region of the DNN. This statistical uncertainty is modelled using the Barlow-Beeston method [156, 157].

- **b-tagging (normalization and shape):** To account for differences in the b-tagging selection efficiency in the data and MC, centrally derived scale factors are added to the event weights. These scale factors are varied up and down according to the source of uncertainty to derive their final effect.

- **Drell-Yan contribution from pileup and noise (normalization):** A significant fraction (about 30–40%) of the DY background populating the low score DNN bins comprises of events where the leading or the subleading jet falls in the forward region of the detector ($|\eta| > 3.0$), but is not matched with a jet at the generator level. These jets originate either from soft emission produced by the parton shower or from pileup interactions, and are promoted above the p_{T} thresholds used in the analysis due to inefficiencies of the detector response. The remainder of the DY events contain two jets matched to generator-level jets primarily arising from the quarks at matrix element level.

  To account for this, we treat the DY process as two different backgrounds in the final fit, called DY+2jets and DY+pu/noise. The normalization of the DY+pu/noise is left floating in the fit and is directly constrained by the observed data. The normalization of the DY+2jets is taken from the simulation and constrained by the data within the described uncertainties. Due to significant variation of the detector response over the data taking period, this normalization is treated uncorrelated across years.

5.6.4.2 Theory uncertainties

- **Signal inclusive cross section (normalization):** uncertainties in the production cross section for gg$H$, VBF, VH, and tt$H$ processes from QCD scale and PDF variations are taken from Ref. [40] and are listed in Tab. 5.1.

- **ggH Simplified Template Cross-Sections (STXS) (normalization and shape):** These uncertainties are evaluated following the recommendations of the LHC Higgs Cross-Section working group (LHCHXSWG) [158]. This recipe provides a set of independent sources of uncertainty, modelled via log-Normal nuisance parameters correlated across categories and eras. These sources account for variations in the estimate of the gg$H$ acceptance in bins of Higgs boson p_{T} and $\mathrm{N_{jets}}$. The size of this uncertainty is around ~15-25% for the ggH process in the VBF category.

- **VBF simplified Template Cross-Sections (STXS) (normalization and shape):** These uncertainties are evaluated following the recommendations of the LHC Higgs Cross-Section working group (LHCHXSWG) [158]. This recipe provides a set of independent sources of uncertainty, modelled via log-Normal nuisance parameters correlated across categories and eras. They account for variation of the VBF signal acceptance as a function of Higgs boson p_{T}, $\mathrm{N_{jets}}$, and m_{jj}. The size of this uncertainty is around ~2-4% for the VBF process in the VBF category.

- **Perturbative QCD scale variation(normalization and shape):** The perturbative QCD renormalization and factorization uncertainties are estimated by changing the scales μ_R and μ_F up and down by a factor of 2 from their default values used in the matrix element calculation. They are treated as correlated between regions and eras but uncorrelated between processes.

- **Parton distribution functions or PDFs (normalization and shape):** The uncertainty due to PDFs is evaluated by taking the RMS of the predictions from the 100 replicas provided by NNPDF3.0 [127] in the 2016 samples, or taking the sum-in-quadrature of the variations provided by the 33 Hessian components of NNPDF3.1 [128]. For each process and PDF replica, the ratios between the replicas and the nominal DNN histogram is computed. These ratios are then fitted with two different functions: $y = mx + b$ and $y = ax^2 + b$, where x is the value of the DNN. In order to build an up variation of the nominal histogram, the bin contents of the nominal distributions are multiplied by either b', $m'x$ or $a'x^2$, where x is the value of the bin center and m', a', b' are the RMS of the parameters m, a, b computed by the previous fits. Down variations are obtained by flipping the sign of m', a', and b'. Normalization effects are removed for the two non constant variations. The three corresponding nuisance parameters are named *PDFX 0,1,2* and their effects are correlated across DNN bins, regions, and eras but uncorrelated between processes.

- **Parton shower acceptance uncertainty (normalization and shape) for**

VBF-H: The parton shower uncertainty for the signal accounts for acceptance (5-10%) and/or shape differences that are observed in the predictions obtained with different combinations of matrix element generators. We use the prediction from POWHEG +PYTHIA with dipole recoil as our nominal prediction and treat the full difference with the POWHEG +HERWIG prediction as an up variation of the uncertainty. The down variation is calculated by flipping the sign of the difference between the two predictions.

- **Parton shower acceptance uncertainty (normalization and shape) for VBF-Z:** The PS uncertainty for the VBF-Z sample is calculated using observed differences in predictions from different generators. The PYTHIA shower in global recoil mode (standard p_{T}-ordered shower) is known to mis-model the additional hadronic activity in VBF-Z. Events showered with POWHEG + PYTHIA with dipole recoil mode are expected to show a better agreement with data, however such a sample was unavailable at the time of this study. Predictions from a PYTHIA + HERWIG sample are in better agreement with the observed data and is therefore used to derive the central prediction for the VBF-Z process. 20% of the difference between the predictions from the two PS programs is considered as an uncertainty, which varies between 2–8%. The 20% fraction is chosen because it accounts for 2× of the parton shower uncertainty provided by the PYTHIA generator with the PS-weight mechanism.

5.6.5 Fitting strategy

The signal strength is extracted by performing a binned maximum likelihood fit of the DNN distribution in the signal region. The fit is simultaneously performed in the Signal Region and Side-Band Region to constrain nuisance parameters. Since the value of $m_{\mu\mu}$ is important for the DNN evaluation and events that are farther away from the Higgs peak have a smaller output DNN score, we replace the $m_{\mu\mu}$ value in the Side-Band regions with a fixed value of 125 GeV. This is known as the mass-decorrelated DNN. This helps to reproduce the main features of the DNN distribution in the Side-Band and is helpful for constraining the systematic uncertainties, especially in the high score DNN bins.

The signal strength extraction from the fit is heavily dependent on the binning choice. The binning choice needs to be balanced in a manner such that we achieve optimal sensitivity while minimizing the total number of bins. For each year, we define thirteen bins based on the VBF Higgs simulation yield in the signal region.

The highest score bin is defined such that it has a fixed signal yield of 0.6 expected events. All the other bins are consecutively defined such that the signal yields changes linearly down to the lowest score bin. This procedure leads to different bin boundaries for each year. The binning in the Side-Band and Signal Region are chosen to be the same per year. To obtain a visual focus on the high score bins of the DNN, both the score and the binning are transformed as $DNN' = atanh(DNN)$.

Limits on the signal strength are computed with the Asymptotic CLs method [151]. Systematic uncertainties sources described in Sec. 5.6.4 are taken into account with additional nuisance parameters that modify the likelihood function in a frequentist manner [159]. The likelihood function for extracting the signal strength modifier μ (ratio of the measured cross-section and the Standard Model cross-section) is defined as:

$$\mathcal{L}(data|\mu, \boldsymbol{\theta}) = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i} \prod_k p(\theta_k|\tilde{\theta_k}) \tag{5.4}$$

where $\frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-\mu s_i - b_i}$ is defined as the Poisson probability to observe n_i events in bin i. s_i and b_i represent the mean number of entries in the i-th bin. Each uncertainty is parametrized by a nuisance parameter θ_k and is described by a probability density function (or p.d.f.), $p(\theta_k|\tilde{\theta_k})$, where $\tilde{\theta_k}$ is the initial estimate of the parameter.

To test the compatibility of the data with the background-only and signal+background hypotheses, we construct the test statistic as a profile likelihood ratio $\tilde{q_\mu}$ described by:

$$\tilde{q_\mu} = -2\ln\left(\frac{\mathcal{L}(data|\mu, \hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}})}{\mathcal{L}(data|\hat{\mu}, \hat{\boldsymbol{\theta}})}\right), \qquad 0 \leq \hat{\mu} \leq \mu \tag{5.5}$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}}$ refers to the conditional maximum likelihood estimators of $\hat{\boldsymbol{\theta}}$, given the signal strength parameter μ. The parameters $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ correspond to the global maximum of the likelihood. We then find the experimentally observed value of the test statistic $\tilde{q}_\mu^{obs}$, for a given signal strength modifier µ. We also find the parameters, $\hat{\boldsymbol{\theta}}_{\mathbf{0}}$ and $\hat{\boldsymbol{\theta}}_\mu$ that best describe the observed data, with a background only and signal+background hypothesis, respectively. We then generate two sets of Monte Carlo pseudo-data with p.d.f.s of the likelihood ratios given by $f(\tilde{q_\mu}|\mu = 0, \hat{\boldsymbol{\theta}}_{\mathbf{0}}^{\boldsymbol{obs}})$ and $f(\tilde{q_\mu}|\mu, \hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}}^{\boldsymbol{obs}})$, corresponding to the two hypotheses.

We then define two probabilities given by:

$$p_\mu = P(\tilde{q}_\mu \geq \tilde{q}_\mu^{obs}|signal + background) = \int_{\tilde{q}_\mu^{obs}}^{\infty} f(q|\mu, \hat{\boldsymbol{\theta}}_{\boldsymbol{\mu}}) \tag{5.6}$$

$$p_b = P(\tilde{q}_\mu < \tilde{q}_\mu^{obs} | background\ only) = \int_0^{\tilde{q}_\mu^{obs}} f(q|\mu, \hat{\boldsymbol{\theta}}_\mu) \quad (5.7)$$

where p_μ and p_b correspond to the the signal + background and background-only hypothesis, respectively. The ratio of the two probabilities

$$CL_s = \frac{p_\mu}{1 - p_b} \quad (5.8)$$

is the exclusion limit for signal strength modifier μ. The upper limit on the signal strength at 95% confidence level is given by that value of μ where CL_s= 0.05.

The p-value is defined for $\mu = 1$ as $\frac{p_1}{1-p_b}$ and denotes the probability that the observed fluctuation in data can be described by the background-only null hypothesis. The p-value can be converted into an equivalent significance Z as $Z = \Phi^{-1}(1 - p)$, where Φ^{-1} is the quantile (inverse of the cumulative distribution) under one-tail of a Gaussian. A value of $Z = 3$ (3σ evidence of new physics) and $Z = 5$ (5σ observation of new physics) corresponds to a null hypothesis p-value of 1.3×10^{-3} and 2.87×10^{-7}, respectively.

5.6.6 Results

We perform independent fits to the signal strength modifier μ in all three years, and take in to account the signal region and the side-band simultaneously per fit. These same fits are also used to derive the signal significance. A combined fit is then performed to the global μ defined as $(\sigma \times \mathcal{B}(H \to \mu\mu))_{obs}/(\sigma \times \mathcal{B}(H \to \mu\mu))_{SM}$.

The expected and observed significance for each of the fits are shown in table 5.3. The data and simulation distributions for the fitted DNN is shown in Figures 5.29, 5.30, 5.31 and 5.32. The best fit signal strength for m_H 125.38 GeV is $\mu = 1.36^{+0.69}_{-0.61}$ corresponding to an observed significance of 2.40 σ.

Year	Significance	
	Observed	Expected
2016	2.01	1.04
2017	0.00	0.88
2018	2.21	1.21
Combined	2.40	1.77

Table 5.3: The observed and expected significance in the VBF category for excluding the background-only null hypothesis for m_H=125.38 GeV, for each year and for the combined data-taking period.

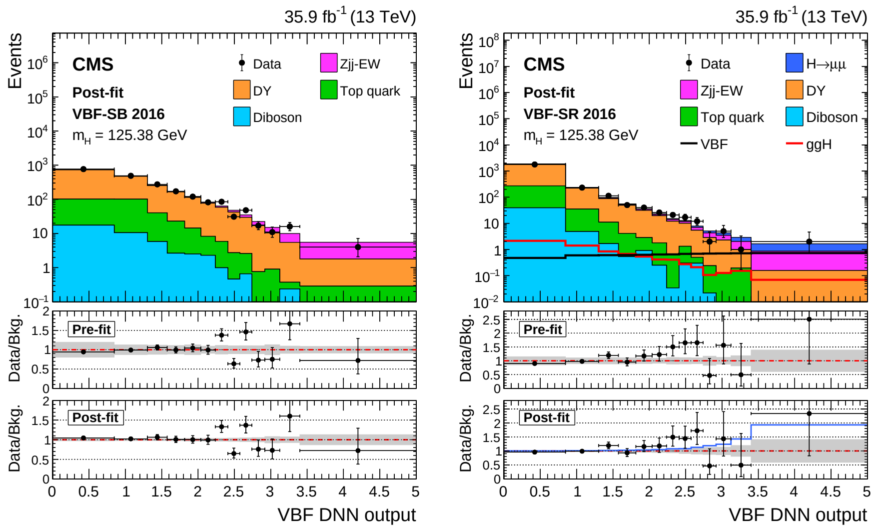


Figure 5.29: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the 2016 data-taking period. In the second panel, the ratio between data and the pre-fit background prediction is shown. The gray band indicates the total pre-fit uncertainty obtained from the systematic sources previously described. The third panel reports the ratio between data and the post-fit background prediction from the signal-plus-background fit. The gray band indicates the total background uncertainty after performing the fit, while the blue histogram refers to the total signal extracted from the fit.

An unbiased mass distribution representative of the fit result in the VBF category is obtained by weighting both simulated and data events from the VBF-SR and VBF-SB regions by the per-event $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ ratio, computed as a function of the mass-decorrelated DNN output, defined in Sec. 5.6.5, for events within $m_{\mu\mu}$ = 125.38 GeV± HWHM. The best-fit estimates for the nuisance parameters and signal strength are propagated to the $m_{\mu\mu}$ distribution. Figure 5.33 shows the observed and predicted weighted $m_{\mu\mu}$ distributions for events in the VBF-SB and VBF-SR regions, combining 2016, 2017, and 2018 data. The lower panel shows the residuals between the data and the post-fit background prediction, along with the post-fit uncertainty obtained from the background-only fit. The best-fit signal contribution with $m_H = 125.38$ GeV is indicated by the blue line. An excess is observed in the weighted data distribution that is consistent with the expected resonant mass distribution for the signal with m_H near 125 GeV and compatible with the excess observed at high DNN score in Fig. 5.32.

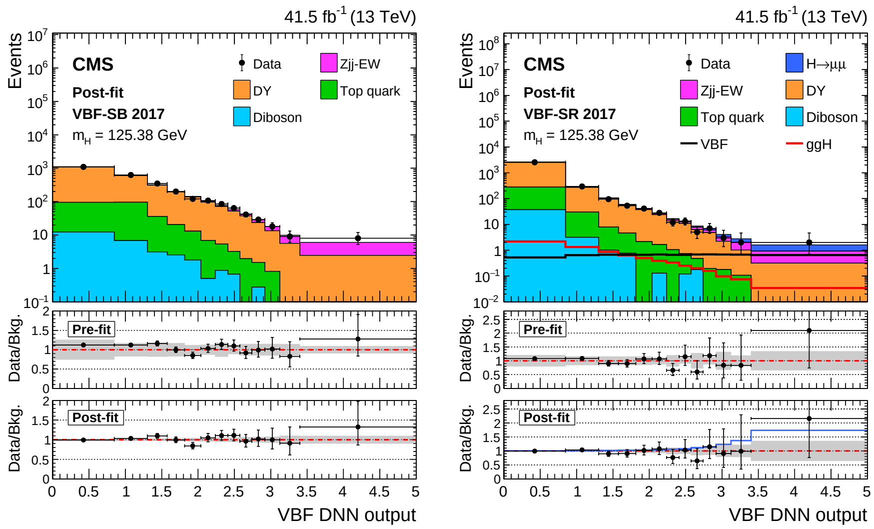


Figure 5.30: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the 2017 data-taking period. The description of ratio panels is the same as in Fig. 5.29

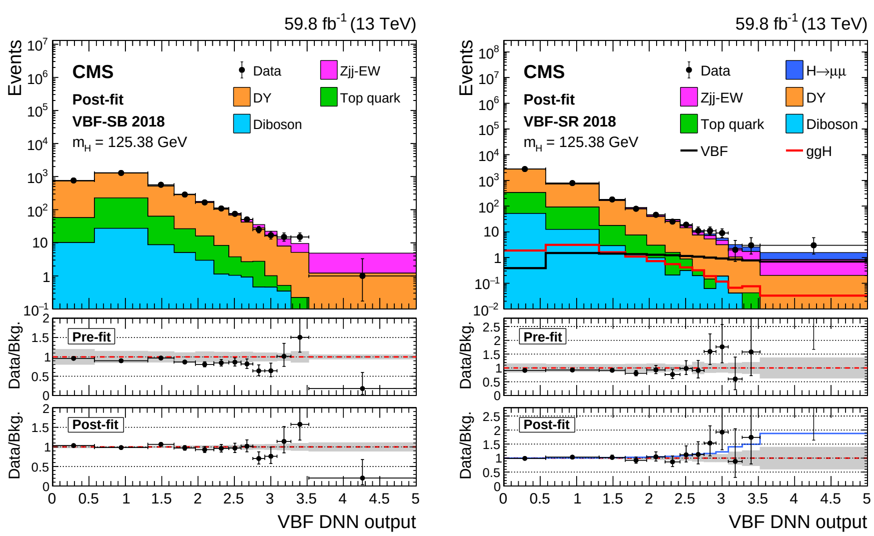


Figure 5.31: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the 2018 data-taking period. The description of ratio panels is the same as in Fig. 5.29.

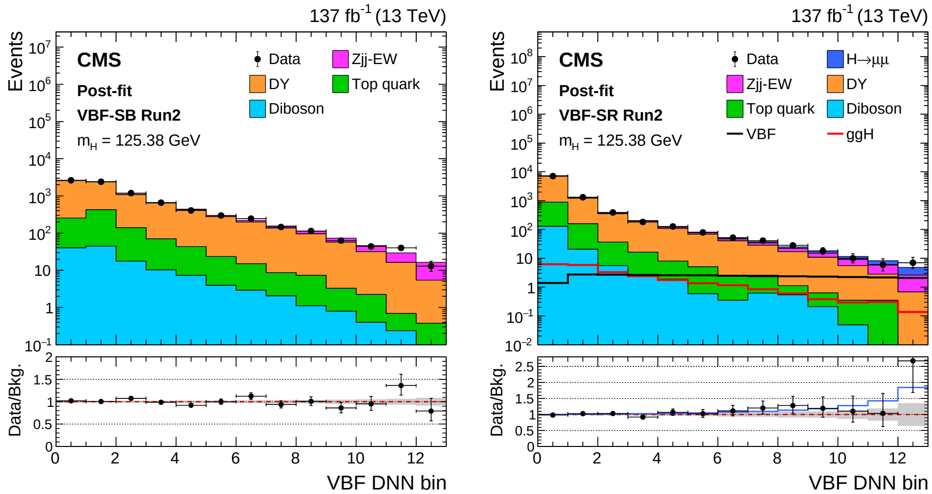


Figure 5.32: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the full Run-2 data-taking period. The lower panel shows the ratio between data and the post-fit background prediction from the S+B fit. The best fit $H \to \mu\mu$ signal contribution for m_H= 125.38 GeV is indicated by the blue solid line, while the grey band indicates the total background uncertainty.

5.7 The gg$H$, VH, and tt$H$ categories

In this section, I will briefly describe the search strategy for the remaining categories of this analysis. The basic analysis strategy for these categories is as follows: events are first divided into independent subcategories based on the output score of a multivariate discriminator specifically trained to discriminate between the signal and the major backgrounds in that particular category. The signal strength is then extracted from each subcategory by searching for a narrow peak over a smoothly falling background in the $m_{\mu\mu}$ distribution. The total background is estimated directly from a fit to the data in the mass sidebands.

5.7.1 ggH production

Events in the ggH category are required to contain two muons passing the baseline selections detailed in Sec. 5.3 and Sec. 5.5. Events with additional electrons or muons are rejected. Events that contain zero or more jets that are spatially separated ($\Delta R > 0.4$) from either of the two muons are considered in this category. In order to maintain orthogonality with the VBF category, events containing two or more jets with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ are only considered if the leading jet has $p_{\mathrm{T}} < 35\,\mathrm{GeV}$, or $m_{\mathrm{jj}} <$

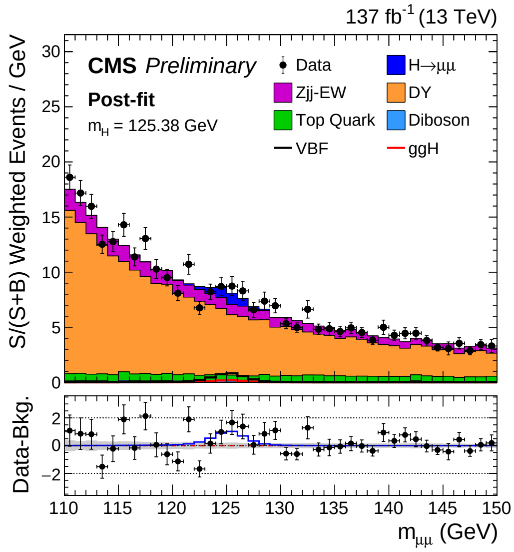


Figure 5.33: The $m_{\mu\mu}$ distribution for the weighted combination of VBF-SB and VBF-SR events. Each event is weighted proportionally to the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ ratio, calculated as a function of the mass-decorrelated DNN output. The lower panel shows the residuals after subtracting the background prediction from the signal-plus-background fit. The best-fit $H \to \mu\mu$ signal contribution is indicated by the blue line, and the grey band indicates the total background uncertainty from the background-only fit.

400 GeV, or $|\Delta\eta_{\mathrm{jj}}| < 2.5$. Lastly, events containing at least two loose b-tagged jets or at least one medium b-tagged jet are also ignored to maintain orthogonailty with the ttH category.

5.7.1.1 Multivariate discriminator

A BDT is trained in this category to discriminate between signal and background events. To account for the evolution in the detector response during data-taking periods, the BDT discriminant is trained separately for the 2016, 2017, and 2018 years using the TMVA package [160], resulting in three independent BDT outputs. The input variables are chosen such that the BDT discriminators are effectively uncorrelated with $m_{\mu\mu}$. The input variables for the BDT are :

- $p_{\mathrm{T}\ \mu\mu}$, $y_{\mu\mu}$,
- ϕ_{CS}, and $\cos\theta_{\mathrm{CS}}$,
- the η_{μ} and the $p_{\mathrm{T}\ \mu}/m_{\mu\mu}$ for each of the two muons,
- the p_{T} and η of the leading jet (if present), and $\Delta\eta(\mu\mu, \mathrm{j}_1)$ and $\Delta\phi(\mu\mu, \mathrm{j}_1)$,

- $m_{\rm jj}$, $\Delta\eta_{\rm jj}$, and $\Delta\phi_{\rm jj}$ of the two highest-$p_{\rm T}$ jets (if present),
- the Zeppenfeld (z^*) variable, $\min[\Delta\eta(\mu\mu, {\rm j}_1), \Delta\eta(\mu\mu, {\rm j}_2)]$, and $\min[\Delta\phi(\mu\mu, {\rm j}_1), \Delta\phi(\mu\mu, {\rm j}_2)]$,
- the total number of jets (N_{jets}) in the event with $p_{\rm T} > 25$ GeV and $|\eta| < 4.7$

The signal simulation considered in the training of the BDT are ggH, VBF, VH, and $t\bar{t}H$ processes. The ggH sample used in the training is generated via POWHEG since it provides positively weighted events at NLO in QCD. For the final signal extraction fit, the MADGRAPH5_AMC@NLO sample is used instead, since it provides a more accurate description of gluon fusion events accompanied by more than one jets. The background simulation consists of DY, $t\bar{t}$, single top, diboson, and electroweak Z processes. Only events with $m_{\mu\mu}$ in the range 115–135 GeV are included in the training. Signal events in the BDT training are assigned a weight proportional to the expected mass resolution ($1/(\sigma_{\mu\mu}/m_{\mu\mu})$), derived from the uncertainties in the $p_{\rm T}$ measurement from the individual muon tracks. This weighting provides increased importance to the high-resolution signal events.

The $p_{\rm T}(\mu\mu)$ is one of the most discriminating observables in the BDT for the ggH category. Discrepancies between data and simulation in the $p_{\rm T}(\mu\mu)$ spectrum similar to those reported in Ref. [161] are also observed in this search. In order to correctly model the $p_{\rm T}(\mu\mu)$ spectrum of the DY background during the training of the BDT discriminant, corrections are derived for each data-taking period by re-weighting the $p_{\rm T}(\mu\mu)$ distribution of the DY simulation to reproduce the observation in data for dimuon events with $70 < m_{\mu\mu} < 110$ GeV. These corrections are obtained separately for events containing zero, one, and two or more jets with $p_{\rm T} > 25$ GeV and $|\eta| < 4.7$.

Figure 5.34 (left) shows the summed BDT score distribution from all three years, for events with $110 < m_{\mu\mu} < 150$ GeV. Five event categories are defined based on the summed BDT score output. The category boundaries are determined via an iterative process that aims to maximize the expected sensitivity, which is estimated from signal-plus-background fits to the $m_{\mu\mu}$ distribution in simulated events with $110 < m_{\mu\mu} < 150$ GeV. In these fits, the Higgs boson signal is modelled using the double-sided Crystal Ball function (DCB) as

$$\mathrm{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-\hat{m})^2/2\sigma^2} & -\alpha_\mathrm{L} < \frac{m_{\mu\mu}-\hat{m}}{\sigma} < \alpha_\mathrm{R} \\ \left(\frac{n_\mathrm{L}}{|\alpha_\mathrm{L}|}\right)^{n_\mathrm{L}} e^{-\alpha_\mathrm{L}^2/2}\left(\frac{n_\mathrm{L}}{|\alpha_\mathrm{L}|} - |\alpha_\mathrm{L}| - \frac{m_{\mu\mu}-\hat{m}}{\sigma}\right)^{-n_\mathrm{L}} & \frac{m_{\mu\mu}-\hat{m}}{\sigma} \leq -\alpha_\mathrm{L} \\ \left(\frac{n_\mathrm{R}}{|\alpha_\mathrm{R}|}\right)^{n_\mathrm{R}} e^{-\alpha_\mathrm{R}^2/2}\left(\frac{n_\mathrm{R}}{|\alpha_\mathrm{R}|} - |\alpha_\mathrm{R}| - \frac{m_{\mu\mu}-\hat{m}}{\sigma}\right)^{-n_\mathrm{R}} & \frac{m_{\mu\mu}-\hat{m}}{\sigma} \geq \alpha_\mathrm{R} \end{cases} . \tag{5.9}$$

The core of the DCB function consists of a Gaussian distribution of mean $\hat{m}$ and standard deviation σ, while the tails on either side are modelled by a power-law function with parameters α_L and n_L (low-mass tail), and α_R and n_R (high-mass tail). The total expected background is modelled with a modified form of the Breit–Wigner function (mBW),

$$\mathrm{mBW}(m_{\mu\mu}; m_Z, \Gamma_Z, a_1, a_2, a_3) = \frac{e^{a_2 m_{\mu\mu} + a_3 m_{\mu\mu}^2}}{(m_{\mu\mu} - m_\mathrm{Z})^{a_1} + (\Gamma_\mathrm{Z}/2)^{a_1}}, \tag{5.10}$$

where the parameters m_Z and Γ_Z refer to the measured Z boson mass of 91.19 GeV and width 2.49 GeV [162], and the parameters a_1, a_2, and a_3 have flat priors.

The summed BDT output score from all three years is scanned in quantiles of signal efficiency (1% to 99%), and for every quantile x, two sets of events (x%, 100-x%) are considered. The total expected significance is calculated for both sets, and the first BDT boundary is set at the location where the expected significance is maximized. This process is repeated recursively to define additional category boundaries until the further gain in the expected significance is less than 1%. The optimized event categories are labelled as ggH-cat1, ggH-cat2, ggH-cat3, ggH-cat4, and ggH-cat5 corresponding to signal efficiency intervals of 0–30%, 30–60%, 60–80%, 80–95%, and >95%, respectively. The grey vertical boxes in Figure 5.34 (left) indicate the various BDT boundaries for the optimized event categories.

5.7.1.2 Signal extraction

A simultaneous binned maximum-likelihood fit to the observed $m_{\mu\mu}$ distributions in each of the five categories is performed over the mass range 110–150 GeV. A bin size of 50 MeV is chosen for the $m_{\mu\mu}$ distribution. The signal distributions from the different production modes (ggH, VBF, WH, ZH, and $t\bar{t}H$) are modelled independently with DCB functions (Eqn. 5.9, and the best-fit values of the DCB tail parameters are treated as constants in the final fit to the data. The $\hat{m}$ and σ parameters of the DCB function are allowed to vary, within Gaussian constraints, with widths corresponding to the muon momentum scale (up to 0.2%) and resolution uncertainties (up to 10%) in each event category. Figure 5.34 (right) shows the total signal model for $m_H = 125\,\text{GeV}$ obtained by summing the contributions from the different production modes in ggH-cat4 and ggH-cat1.

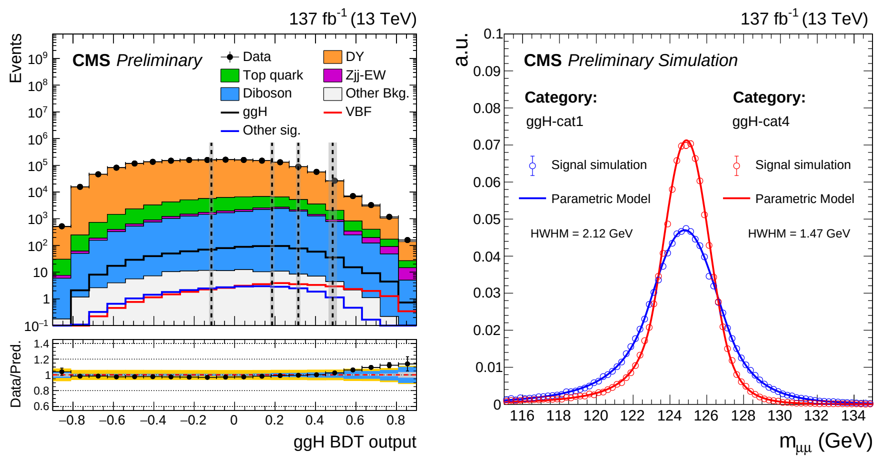


Figure 5.34: The ggH category. (Left) the observed BDT output distribution in the ggH category for events with $m_{\mu\mu}$ between 110–150 GeV. The gray vertical boxes indicate the BDT boundaries for the optimized event categories defined in the text. In the lower panel, the ratio between data and the expected background is shown. The grey band indicates the uncertainty due to the limited size of the simulated samples. The azure band corresponds to the sum in quadrature between the statistical and experimental systematic uncertainties, while the orange band additionally includes the theoretical uncertainties affecting the background prediction. (Right) the signal shape model for the simulated $H \to \mu\mu$ sample with $m_H = 125\,\text{GeV}$ for ggH-cat4 (red) and ggH-cat1 (blue) categories.

The theoretical and experimental sources of systematic uncertainty affecting the expected signal rate in each event category are similar to those described in the VBF analysis. Experimental uncertainties in the measurement of the muon selection efficiencies (0.5–1% per category), jet energy scale (1–4% per category) and resolution (1–6% per category), the modelling of the pileup conditions (0.3–0.8% per category), the integrated luminosity (about 2.5% per year), and the efficiency of vetoing b quark jets (0.1–0.5% per category) are considered. Theoretical uncertainties in the prediction of the Higgs boson production cross section, decay rate, and acceptance are also included, corresponding to a total uncertainty in the ggH process yield ranging from 6–12% depending on category. Rate uncertainties are modelled in the signal extraction as nuisance parameters acting on the relative signal yield with log-normal priors.

The background contribution in each category is modelled with analytical functions and is constrained directly by the observed data in the signal-plus-background fit. The background shape in $m_{\mu\mu}$ is similar in all categories and arises mainly from the DY + jets process. The function describing the background in each event category is defined as the product of a core shape that is common among all event categories, with parameters correlated across categories, and a polynomial term (shape modifier) specific to each event category that modulates the core shape. This background modelling approach is referred to as the core-pdf method. The core background shape is obtained from an envelope of three distinct functions: the modified Breit–Wigner (mBW) defined in Eqn. 5.10, a sum of two exponential functions, and the product of a non-analytical shape derived from the FEWZ v3.1 generator [132] and a third-order Bernstein polynomial. Each of these functions contains three freely floating shape parameters. The non-analytical shape derived from the FEWZ generator is obtained by simulating DY events at NNLO precision in QCD corrections and NLO accuracy in EW theory and smoothing out the resulting $m_{\mu\mu}$ distribution using a spline function [133, 134]. In a given category, each of the three core functions is modulated by either a third- (ggH-cat1 and ggH-cat2) or a second-order polynomial, with parameters uncorrelated across categories. A discrete profiling method [163] is employed, which treats the choice of the core function used to model the background as a discrete nuisance parameter in the signal extraction.

The choice of parametric function for the background model can induce a bias in the measured signal strength. To estimate the size of this uncertainty, in each event category, background-only fits to the data are performed using different types of analytical functions: the modified Breit–Wigner (mBW), a sum of two exponentials, a sum of two power laws, a Bernstein polynomial, the product between the FEWZ spline and a Bernstein polynomial, the product between a Bernstein polynomial and a BWZ function (Eqn. 5.11) , and the BWZGamma function (Eqn. 5.12).

$$\mathrm{BWZ}(m_{\mu\mu}; a, m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}) = \frac{\Gamma_Z e^{a m_{\mu\mu}}}{(m_{\mu\mu} - m_{\mathrm{Z}})^2 + (\Gamma_{\mathrm{Z}}/2)^2}, \tag{5.11}$$

$$\mathrm{BWZGamma}(m_{\mu\mu}; a, f, m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}) = f \times \mathrm{BWZ}(m_{\mu\mu}; a, m_{\mathrm{Z}}, \Gamma_{\mathrm{Z}}) + (1 - f) \times \frac{e^{a m_{\mu\mu}}}{m_{\mu\mu}^2}. \tag{5.12}$$

The BWZ function is a Breit–Wigner distribution with an exponential tail. The BWZGamma function is the sum of a BWZ function and a $1/m_{\mu\mu}^2$ term, used to model the Z boson and the photon contributions to the $m_{\mu\mu}$ spectrum in DY events, respectively. For the functions including Bernstein polynomials, a Fisher test [11] is used to determine the maximum order of the polynomials to be considered in the fit. The chosen functional forms are able to fit the data with a χ^2 p-value > 0.05 in all categories. Pseudo-data sets are generated across all event categories from the post-fit background shapes obtained for each type of function in each category, and injecting a given number of signal events. Signal-plus-background fits are performed on the pseudo-data sets using the core-pdf method. The median difference between the measured and injected signal yields, relative to the post-fit uncertainty on the signal yields, gives an estimate of the bias due to the choice of the background model. The measured bias changes the overall uncertainty by less than 1% in the signal rate, and is therefore neglected.

Figure 5.35 shows the $m_{\mu\mu}$ distributions after performing a binned maximum-likelihood fit in each of the 5 ggH categories. Table 5.4 reports the signal composition in each ggH category as well as the HWHM of the expected signal shape. In addition, the estimated number of background events, the observation in data, the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$, and the $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios computed within the HWHM range around the signal peak are also listed.

Category	Sig.	ggH (%)	VBF (%)	$VH + t\bar{t}H$ (%)	HWHM (GeV)	Bkg. in HWHM	$\mathrm{S}/(\mathrm{S}+\mathrm{B})$ (%) in HWHM	$\mathrm{S}/\sqrt{\mathrm{B}}$ in HWHM	Data in HWHM
gg$H$-cat1	267.6	93.7	2.9	3.4	2.12	86359	0.20	0.60	86632
gg$H$-cat2	311.5	93.5	3.4	3.1	1.75	46347	0.46	0.98	46393
gg$H$-cat3	131.4	93.2	4.0	2.8	1.60	12655	0.70	0.80	12738
gg$H$-cat4	125.6	91.5	5.5	3.0	1.47	8259	1.03	0.96	8377
gg$H$-cat5	53.8	83.5	14.3	2.2	1.50	1678	2.16	0.91	1711

Table 5.4: The total expected number of signal events with $M_H = 125.38\,\mathrm{GeV}$, the HWHM of the signal peak, the estimated number of background events and the observation in data within ± HWHM, and the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and the $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios within ± HWHM, for each of the optimized ggH event categories.

5.7.2 VH production

Events considered in the VH category contain at least two muons passing the selection requirements listed in Sec. 5.3 and Sec. 5.5. The VH category is required to be orthogonal to all other categories in the analysis. Events are also required to have at least one additional lepton (electron or muon), which is expected from the leptonic decay of the W or Z boson. Electrons and muons are required to pass

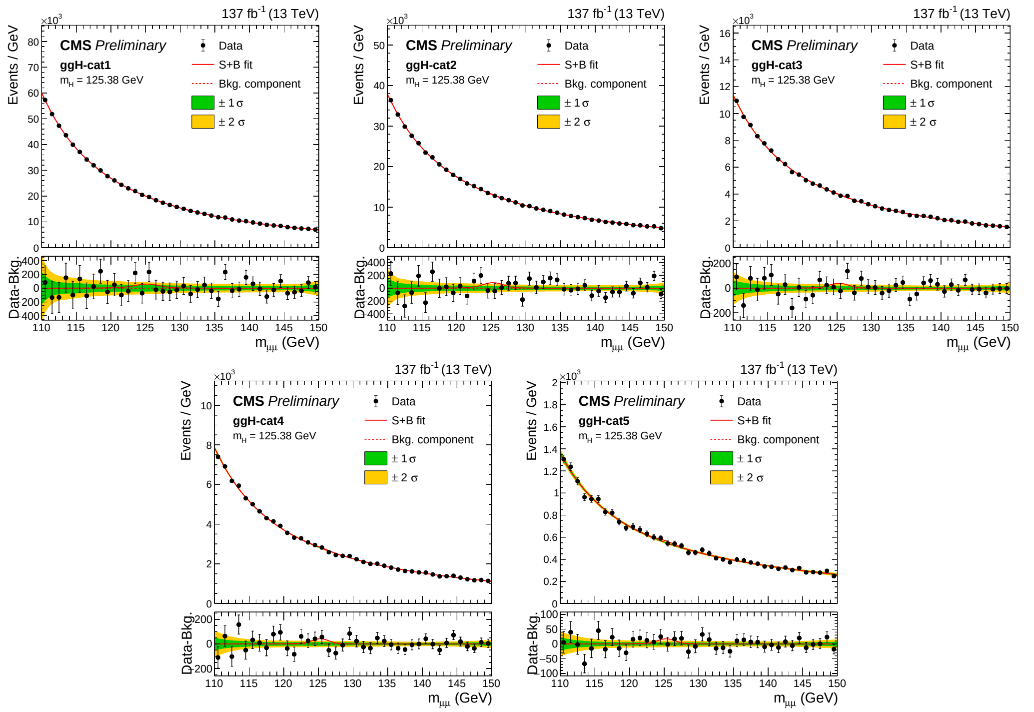


Figure 5.35: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across the ggH categories. First row, from left to right: ggH-cat1, ggH-cat2, and ggH-cat3. Second row, from left to right: ggH-cat4 and ggH-cat5. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after background subtraction and the red line indicates the signal with M_H = 125.38 GeV extracted from the fit.

the medium WP of a multivariate discriminant developed to identify and suppress non-prompt leptons [164], with a selection efficiency of about 90 (85)% per prompt muon (electron).

Events containing exactly one additional lepton belong to the WH category. If the additional lepton is a muon, the two pairs of oppositely charged muons are required to have $m_{\mu\mu} > 12$ GeV to suppress background events from quarkonium decays. Moreover, neither of the two oppositely charged muon pairs can have an invariant mass consistent with m_Z within 10 GeV. Finally, at least one of these two muon pairs must have $m_{\mu\mu}$ in the range 110–150 GeV. If both $m_{\mu\mu}$ pairs satisfy this criterion, the highest-p_T pair is considered as the Higgs boson candidate. If the additional lepton is an electron, the only requirement imposed is that $110 < m_{\mu\mu} < 150$ GeV.

The ZH category targets signal events where the Higgs boson is produced in association with a Z boson that decays to a pair of electrons or muons. Events in the

ZH category are therefore required to contain four leptons, with a combined lepton number and electric charge of zero. The invariant mass of each pair of same-flavour opposite-charge leptons is required to be greater than 12 GeV. An event is rejected if it does not contain exactly one pair of same-flavour and oppositely charged leptons with invariant mass compatible with the Z boson within 10 (20) GeV for muon (electron) pairs. In addition, each event must contain one oppositely charged muon pair satisfying $110 < m_{\mu\mu} < 150$ GeV. For events with four muons, the muon pair with $m_{\mu\mu}$ closer to m_Z is chosen as the Z boson candidate, while the other muon pair is selected as the Higgs boson candidate. A summary of the selection criteria applied in the WH and ZH production categories is reported in Table 5.5.

Selection	WH leptonic		ZH leptonic	
	$\mu\mu\mu$	$\mu\mu$e	4μ	2μ2e
Number of loose (medium) b-tagged jets	≤ 1 (0)	≤ 1 (0)	≤ 1 (0)	≤ 1 (0)
N(μ) passing id.+iso.	3	2	4	2
N(e) passing id.+iso.	0	1	0	2
Lepton charge	$\sum q(\ell) = \pm 1$		$\sum q(\ell) = 0$	
Low mass resonance veto	$m_{\ell\ell} > 12$ GeV			
N($\mu^+\mu^-$) pairs with $110 < m_{\mu\mu} < 150$ GeV	≥ 1	1	≥ 1	1
N($\mu^+\mu^-$) pairs with $\lvert m_{\mu\mu} - m_Z \rvert < 10$ GeV$\lvert$	0	0	1	0
N(e$^+$e$^-$) pairs with $\lvert m_{ee} - m_Z \rvert < 20$ GeV$\lvert$	0	0	0	1

Table 5.5: Summary of the kinematic selection used to define the WH and ZH production categories.

The main backgrounds of the WH category are the WZ (off-shell Z boson decay), ZZ (one lepton is not reconstructed) and the DY process (with associated lepton production). The main backgrounds of the ZH category are the ZZ and ggZZ processes.

5.7.2.1 Multivariate discriminator

Two separate BDT discriminants are trained to discriminate between signal and background events in the WH and ZH categories. The BDT input variables are not significantly correlated with the $m_{\mu\mu}$ of the Higgs boson candidate. During the BDT training, weights are applied to the signal events that are inversely proportional to the per-event uncertainty on the measured $m_{\mu\mu}$, as described in Section 5.7.1.

The input variables to the WH category BDT are :

- $p_{\mathrm{T}\ \mu\mu}$ of the Higgs boson candidate, the η_μ of the two muons, and the angular separation $\Delta R_{\mu\mu}$ between them.

- the flavour and the p_{T} of the additional lepton ℓ_{W}.

- $\Delta\eta(\mu\mu, \ell_{\mathrm{W}})$, $\Delta\phi(\mu\mu, \ell_{\mathrm{W}})$, $\Delta\eta(\mu_1, \ell_{\mathrm{W}})$ and $\Delta\phi(\mu_2, \ell_{\mathrm{W}})$

- The $\vec{H}_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative vector sum of the p_{T} of all jets in the event with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 4.7$. The transverse mass and angular distances in ϕ and η of the combined ℓ_{W} and $\vec{H}_{\mathrm{T}}^{\mathrm{miss}}$ system are also considered.

The input variables to the ZH category BDT are :

- the mass m_{ll}, $p_{\mathrm{T}}(ll)$, the η_{ll} and the $\Delta R(ll)$ of the two leptons from the Z boson candidate.

- $\Delta\eta(\mu\mu, ll)$ and $\cos\theta_{CS}(\mu\mu, ll)$

- the flavour of the lepton pair associated to the Z boson decay

Figure 5.36 shows the output of the BDT classifiers in the WH (left) and ZH (right) categories. Based on these outputs, events in the WH category are further divided into three subcategories termed WH-cat1, WH-cat2, and WH-cat3. Similarly, events in the ZH category are divided into two subcategories, labelled ZH-cat1 and ZH-cat2. The boundaries of these categories, defined in terms of the BDT discriminant and indicated in Fig. 5.36 by black dashed vertical lines, are chosen via an optimization strategy analogous to that described in Section 5.7.1 for the ggH category. In this category, the BWZ function (Eqn. 5.11) is used to estimate the total background instead of the mBW (Eqn. 5.10).

5.7.2.2 Signal extraction

The systematic uncertainties considered in this analysis are similar to the ggH category (Section 5.7.1). Figure 5.37 show the $m_{\mu\mu}$ distributions in the WH (first row) and ZH (second row) event categories. The signal is extracted via a binned maximum-likelihood fit in each event category, where the signal is modelled with a DCB function and the background is modelled with the BWZGamma function in WH-cat1, as defined in Eqn.(5.12) and the BWZ function in the remaining categories, as defined in Eqn.(5.11). A bias test on the choice of the background

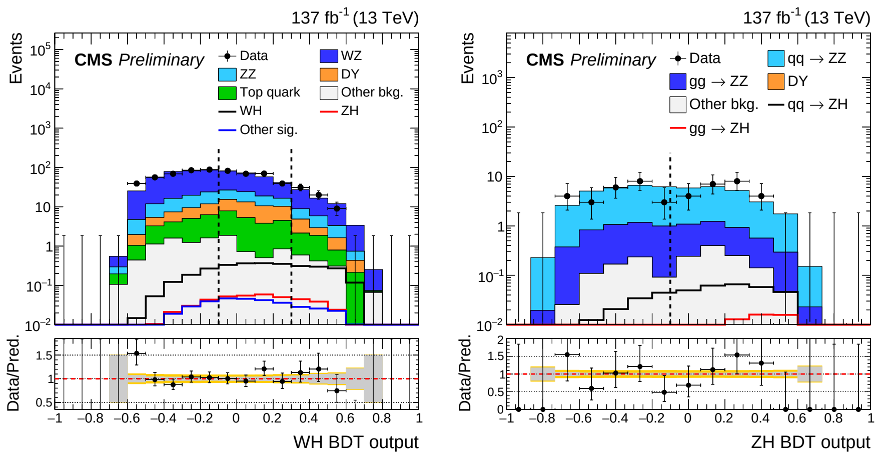


Figure 5.36: The observed BDT output distribution in the WH (left) and ZH (right) categories compared to the prediction from the simulation of various SM background processes. Signal distributions expected from different production modes of the 125 GeV Higgs boson are overlaid. The description of the ratio panel is the same as in Fig. 5.34. The dashed vertical lines indicate the boundaries of the optimized event categories.

modelling function is performed similar to Section 5.7.1.2 and is observed to be small and is therefore neglected in the signal extraction. Finally, Table 5.6 reports the signal composition in the WH and ZH categories, along with the HWHM of the expected signal shape. In addition, the estimated number of background events, the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios, and the observation in data within the HWHM of the signal peak are also listed.

Category	Sig.	WH (%)	$qqZH$ (%)	$ggZH$ (%)	$t\bar{t}H+\mathrm{t}H$ (%)	HWHM (GeV)	Bkg. in HWHM	$\mathrm{S}/(\mathrm{S}+\mathrm{B})$ (%) in HWHM	$\mathrm{S}/\sqrt{\mathrm{B}}$ in HWHM	Data in HWHM
WH-cat1	0.82	76.2	9.6	1.6	12.6	2.00	32.0	1.54	0.09	34
WH-cat2	1.72	80.1	9.1	1.5	9.3	1.80	23.1	4.50	0.23	27
WH-cat3	1.14	85.7	6.7	1.8	4.8	1.90	5.48	12.6	0.35	4
ZH-cat1	0.11	—	82.8	17.2	—	2.07	2.05	3.29	0.05	4
ZH-cat2	0.31	—	79.6	20.4	—	1.80	2.19	8.98	0.14	4

Table 5.6: The total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observed number of events within ± HWHM, and the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and the $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios computed within the HWHM of the signal peak for each of the optimized event categories defined along the WH and ZH BDT outputs.

5.7.3 ttH production

The $t\bar{t}H$ process has the smallest cross section among the main Higgs boson production modes at the LHC. However, this production mode benefits from the clean signature provided by the presence of a pair of top quarks. Top quarks decay pre-

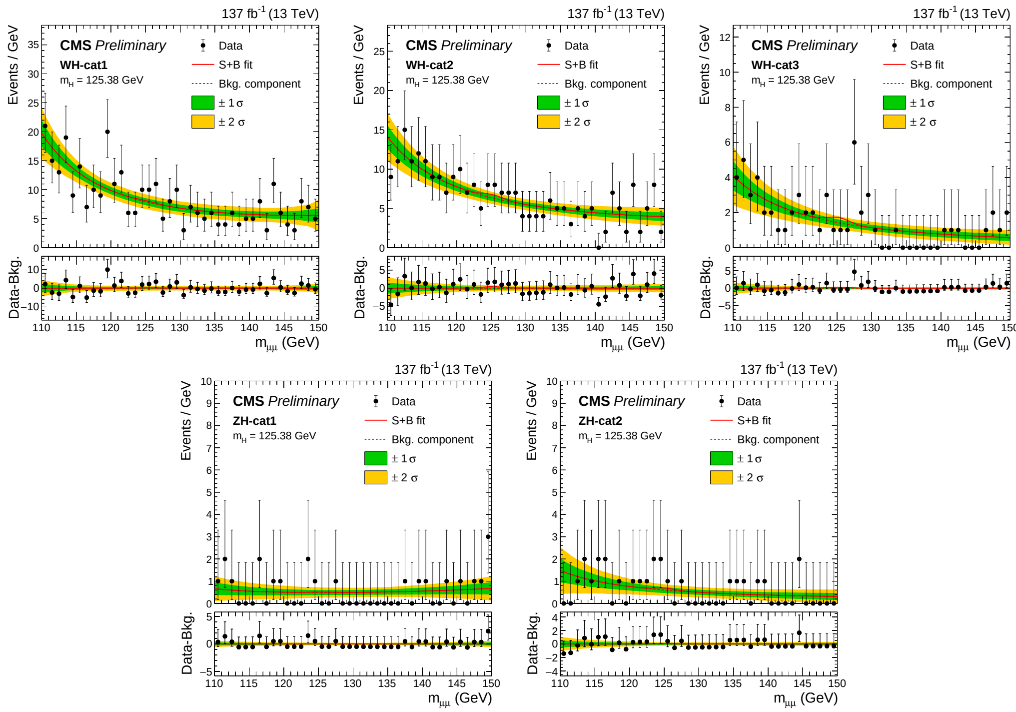


Figure 5.37: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across WH (first row) and ZH (second row) event categories. First row, from left to right: WH-cat1, WH-cat2, and WH-cat3. Second row, from left to right: ZH-cat1 and ZH-cat2. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after the background subtraction, where the red line indicates the signal with $m_H = 125.38\,\text{GeV}$ extracted from the fit.

dominantly into a b quark and a W boson, which then decays either to a lepton and a neutrino ($\mathcal{B}(\text{W} \to \ell\nu) \approx 0.33$), or into two quarks ($\mathcal{B}(\text{W} \to qq') \approx 0.66$). Therefore, events in the $t\bar{t}H$ category are required to contain at least two jets passing the loose WP of the DeepCSV algorithm, or at least one jet passing the medium WP. This requirement suppresses background processes that do not contain b quarks in their final state, and the dominant backgrounds in this channel are the $t\bar{t}$, DY and the $t\bar{t}Z$ processes.

To target the leptonic decays of the top quark in the $t\bar{t}H$ signal events and increase the signal selection efficiency, the muon isolation requirement is relaxed to be less than 40% of the muon p_{T}, as compared to the baseline event selection detailed in Section 5.3.2. In addition, the isolation cone size decreases dynamically with the muon p_{T} ($R = 0.2$ for $p_{\text{T}} < 50\,\text{GeV}$, $R = 10/p_{\text{T}}$ for $50 < p_{\text{T}} < 200\,\text{GeV}$, and $R = 0.05$ for $p_{\text{T}} > 200\,\text{GeV}$), following the same approach used in Ref. [165]. The isolation requirements for an electron (Sec. 5.3.5) is also modified following the same strategy as for muons, and the magnitude of the transverse and longitudinal electron track impact parameters must be smaller than 0.05 and 0.1 cm, respectively. To suppress backgrounds containing non-prompt leptons produced in the decay of heavy quarks, electrons and muons are rejected when found within $\Delta R < 0.4$ of a medium b-tagged jet with $p_{\mathrm{T}} > 15\,\mathrm{GeV}$. Furthermore, just like the VH category, additonal electrons and muons are required to pass the medium WP of a multivariate lepton identification discriminant specifically designed to reject nonprompt leptons [164], resulting in a selection efficiency of about 90 (85)% per prompt muon (electron).

Events are further divided into two separate categories: $t\bar{t}H$ hadronic and $t\bar{t}H$ leptonic. Events with exactly two oppositely charged muons with $110 < m_{\mu\mu} < 150\,\mathrm{GeV}$ and at least three jets in the final state with invariant mass (m_{jjj}) between 100 and 300 GeV belong to the $t\bar{t}H$ hadronic category. Each jet must have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 4.7$. Events with one or two additional leptons in the final state are grouped in the $t\bar{t}H$ leptonic category, in which at least one of the two top quarks decays leptonically. An event in the $t\bar{t}H$ leptonic category containing one (two) additional leptons is further required to have the net sum of the lepton electric charges equal to one (zero). In the case of events with more than one pair of oppositely charged muons with $110 < m_{\mu\mu} < 150\,\mathrm{GeV}$, the pair with the largest dimuon p_{T} is chosen as the Higgs boson candidate. The invariant mass of each pair of same-flavour opposite-charge leptons is required to be greater than 12 GeV. An event is vetoed if it contains a pair of oppositely charged electrons or muons with an invariant mass in the range 81–101 GeV, consistent with the decay of an on-shell Z boson. A summary of the selection criteria used to define the $t\bar{t}H$ hadronic and leptonic categories is reported in Table 5.7.

Selection	$t\bar{t}H$ hadronic	$t\bar{t}H$ leptonic
Number of b quark jets	> 0 medium or > 1 loose b-tagged jets	
Number of leptons	2	3 or 4
Lepton charge	$\sum q(\ell) = 0$	$\mathrm{N}(\ell) = 3\ (4) \rightarrow \sum q(\ell) = \pm 1\ (0)$
Jet multiplicity ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$, $\lvert\eta\rvert < 4.7$)	≥ 3	≥ 2
Leading jet p_{T}	> 50 GeV	> 35 GeV
Jet triplet mass	$100 < m_{jjj} < 300\,\mathrm{GeV}$	—
Z mass veto	—	$\lvert m_{\ell\ell} - m_Z\rvert > 10\,\mathrm{GeV}$
Low mass resonance veto	—	$m_{\ell\ell} > 12\,\mathrm{GeV}$

Table 5.7: Summary of the kinematic selections used to define the $t\bar{t}H$ hadronic and leptonic production categories.

5.7.3.1 Multivariate discriminator

Two separate BDT-based multivariate discriminants are trained for the $t\bar{t}H$ hadronic and leptonic categories. The input variables are uncorrelated with $m_{\mu\mu}$. During the BDT training, weights are applied to the signal events that are inversely proportional to the per-event uncertainty on the measured $m_{\mu\mu}$, as described in Section 5.7.1. A common set of observables are used as input to the two BDT discriminators which include:

- $p_{\mathrm{T}}(\mu\mu)$, $y^{\mu\mu}$, ϕ_{CS}, and $\cos\theta_{\mathrm{CS}}$.
- the η and $p_{\mathrm{T}}^{\mu}/m_{\mu\mu}$ of the two muons
- p_{T} and η of the three leading jets, the maximum DeepCSV value of jets (spatially separated from leptons) , the number of jets, and the scalar (vectorial) p_{T} sum (H_{T} ($H_{\mathrm{T}}^{\mathrm{miss}}$)) of all identified leptons and jets with $|\eta| < 2.5$.
- $p_{\mathrm{T}}^{\mathrm{miss}}$
- the $\Delta\zeta$ variable, which is defined as the projection of the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ on the bisector of the dimuon system in the transverse plane.

In the $t\bar{t}H$ leptonic category, several additional input variables are used to target the leptonic top quark decay. These include the ϕ separation between the Higgs boson candidate and the highest-p_{T} additional lepton, the invariant mass formed by the leading additional lepton and the jet with the highest DeepCSV score, and the transverse mass formed by the additional lepton and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ in the event. In the $t\bar{t}H$ hadronic category, the resolved hadronic top tagger (RHTT), which combines a kinematic fit and a BDT-based multivariate discriminant, is used to identify top quark decays to three resolved jets. The jet triplet with the highest RHTT score is selected as a hadronic top quark candidate. The corresponding RHTT score is used as input to the BDT discriminant. Furthermore, the p_{T} of the top quark candidate and the p_{T} balance of the top quark and the muon pair are also considered.

Figure 5.38 shows the output of the BDT discriminant in the $t\bar{t}H$ hadronic (left) and leptonic (right) categories. The high BDT score region of the $t\bar{t}H$ hadronic category is enriched in events with large jet multiplicity, where the $t\bar{t}$ and DY background simulations are known to be mis-modelled. The signal prediction, however, relies largely on jets from the ME. Since the background prediction is extracted from data, the observed differences between data and background simulation do not affect the fit result. Based on the BDT output, events in the $t\bar{t}H$ leptonic category are further divided into two subcategories, termed $t\bar{t}H$ lep-cat1 and $t\bar{t}H$ lep-cat2. Similarly, events in the $t\bar{t}H$ hadronic category are divided into three subcategories labelled $t\bar{t}H$ had-cat1, $t\bar{t}H$ had-cat2, and $t\bar{t}H$ had-cat3. The BDT score boundaries of these event categories, indicated in Fig. 5.38 by black dashed vertical lines, are optimized following the same strategy described in Section 5.7.1 for the ggH category. Bernstein polynomials are chosen for the analytical function used to model the background in the $t\bar{t}H$ had-cat1 and $t\bar{t}H$ had-cat2, while a sum of two exponentials and a single exponential functions are used in the $t\bar{t}H$ had-cat3 and $t\bar{t}H$ leptonic categories, respectively.

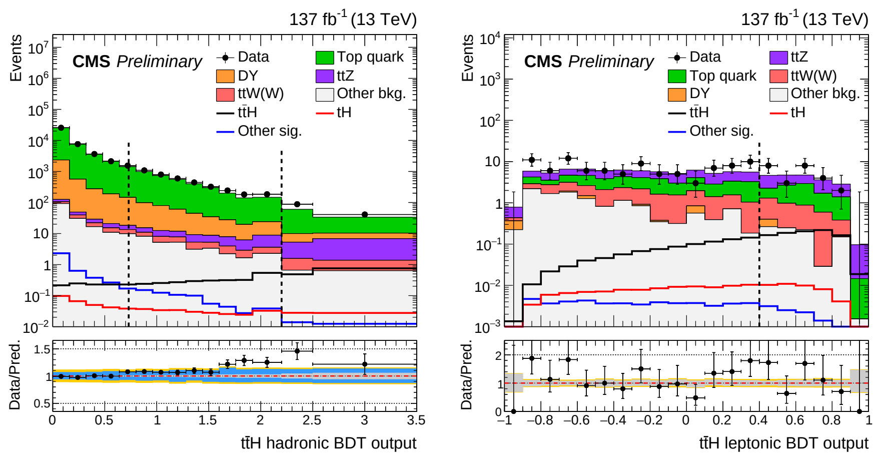


Figure 5.38: The observed BDT output distribution in the $t\bar{t}H$ hadronic (left) and leptonic (right) categories compared to the prediction from the simulation of various SM background processes. The dashed vertical lines indicate the boundaries of the optimized event categories. The description of the ratio panels is the same as in Fig. 5.34.

5.7.3.2 Signal extraction

The systematic uncertainties considered here are similar to the ggH and VH categories (Sec. 5.7.1 and Sec. 5.7.2). Figure 5.39 shows the $m_{\mu\mu}$ distributions in the $t\bar{t}H$ hadronic (first row) and leptonic (second row) event categories. The signal is extracted by performing a binned maximum-likelihood fit to these $m_{\mu\mu}$ distributions, where the signal is modelled using the DCB function and the background is modelled using a second-order Bernstein polynomial or a sum of two exponentials (single exponential) in the $t\bar{t}H$ hadronic (leptonic) categories. The potential bias due to the choice of the parametric function used to model the background is estimated in a manner similar to the ggH analysis (Section 5.7.1.2), and is found to be small. Table 5.8 reports the signal composition of each $t\bar{t}H$ category, along with the HWHM of the expected signal shape. In addition, the estimated number of background events, the observation in data, and the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios within the HWHM of the signal shape are shown.

Category	Sig.	$t\bar{t}H$ (%)	ggH (%)	VH (%)	tH (%)	VBF+$b\bar{b}H$ (%)	HWHM (GeV)	Bkg. in HWHM	$\mathrm{S}/(\mathrm{S}+\mathrm{B})$ (%) in HWHM	$\mathrm{S}/\sqrt{\mathrm{B}}$ in HWHM	Data in HWHM
$t\bar{t}H$ had-cat1	6.87	32.3	40.3	17.2	6.2	4.0	1.85	4298	1.07	0.07	4251
$t\bar{t}H$ had-cat2	1.62	84.3	3.8	5.6	6.2	—	1.81	82.0	1.32	0.12	89
$t\bar{t}H$ had-cat3	1.33	94.0	0.3	1.3	4.2	0.2	1.80	12.3	6.87	0.26	12
$t\bar{t}H$ lep-cat1	1.06	85.8	—	4.7	9.5	—	1.92	9.00	7.09	0.22	13
$t\bar{t}H$ lep-cat2	0.99	94.7	—	1.0	4.3	—	1.75	2.08	24.5	0.47	4

Table 5.8: The total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observed number of events within $\pm$ HWHM, and the $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ and $\mathrm{S}/\sqrt{\mathrm{B}}$ ratios computed within the HWHM of the signal peak, for each of the optimized event categories defined along the $t\bar{t}H$ hadronic and leptonic BDT outputs.

5.8 Results

A simultaneous fit is performed across all the event categories, with a single overall signal strength modifier (μ) with a flat prior. Confidence intervals on the signal strength are estimated using a profile likelihood ratio test statistic, as described previously in Sec. 5.6.5. Theoretical uncertainties affecting the signal prediction are correlated among the event categories. Similarly, experimental uncertainties in the measurement of the integrated luminosity in each year, jet energy scale and resolution, modelling of the pileup conditions, and selection efficiencies of muons and electrons are also correlated across categories. Uncertainties in the b quark jet identification are uncorrelated. Because of the different analysis strategy employed in the VBF category, the acceptance uncertainties from the muon energy scale and resolution are correlated only among the ggH, WH, ZH, and $t\bar{t}H$ categories. Furthermore, their effect on the position and width of the signal peak are assumed to be uncorrelated across event categories.

The signal and background distributions that were obtained for the VBF category in Figure 5.33 are interpolated with a spline function in order to obtain a continuous spectrum that can be summed with the analytical fit results in the ggH, WH, ZH, and $t\bar{t}H$ categories (Figs. 5.35, 5.37 and 5.39). Figure 5.40 shows the $m_{\mu\mu}$ distribution for the weighted combination of all event categories. The ggH, VH, and $t\bar{t}H$ categories are weighted proportionally to the corresponding $\mathrm{S}/(\mathrm{S}+\mathrm{B})$ ratio, where

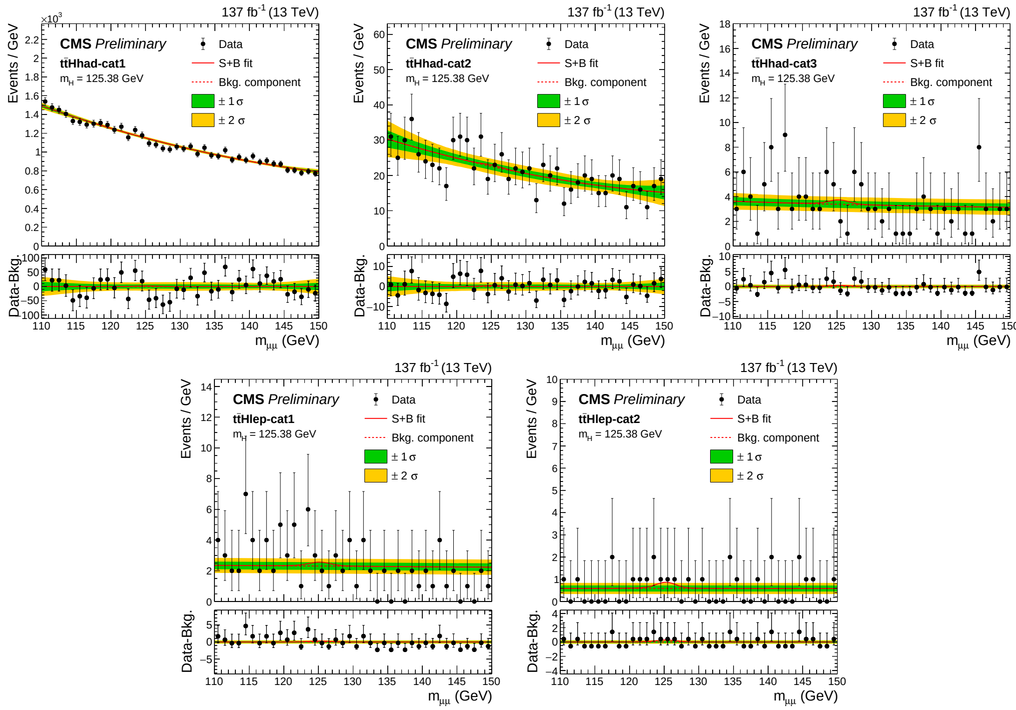


Figure 5.39: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across $t\bar{t}H$ hadronic (first row) and leptonic (second row) event categories. First row, from left to right: $t\bar{t}H$ had-cat1, $t\bar{t}H$ had-cat2, and $t\bar{t}H$ had-cat3. Second row, from left to right: $t\bar{t}H$ lep-cat1 and $t\bar{t}H$ lep-cat2. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after the background subtraction, where the red line indicates the signal with $m_H = 125.38$ GeV extracted from the fit.

S and B are the number of expected signal and background events with mass within ± HWHM of the expected signal peak with $m_H = 125.38$ GeV. The lower panel shows the residuals after background subtraction, with the best-fit SM signal contribution with $m_H = 125.38$ GeV indicated by the red line. An excess of events over the background-only expectation is observed near $m_{\mu\mu} = 125$ GeV.

5.8.1 p-values vs m_H scan

A local p-value (defined in Sec. 5.6.5) vs m_H scan is also performed for each of the four categories. In the ggH, VH, and $t\bar{t}H$ categories, in order to evaluate p-values for masses different than 125 GeV, signal models are derived using additional alternative $H \to \mu\mu$ signal samples generated with m_H fixed to 120 and 130 GeV. Signal shape parameters and the expected rate for each production mode in each event category are then interpolated within $120 < m_H < 130$ GeV, providing a signal model for any

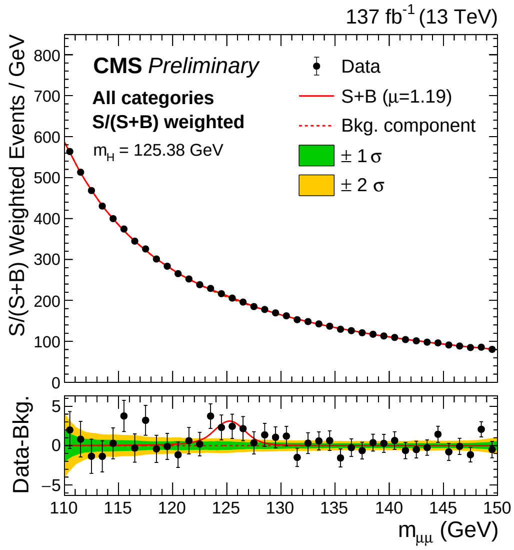


Figure 5.40: The $m_{\mu\mu}$ distribution for the weighted combination of all event categories. The upper panel is dominated mainly by the ggH categories with small $\mathrm{S}/(\mathrm{S}+\mathrm{B})$. The lower panel shows the residuals after background subtraction, with the best-fit SM $H \to \mu\mu$ signal contribution with $m_H = 125.38$ GeV indicated by the red line.

mass value in the $m_H = 125 \pm 5$ GeV range. A different strategy is employed in the VBF category since $m_{\mu\mu}$ is a DNN input variable. By using the mass decorrelated DNN, a potential signal with mass m' different from 125 GeV can be extracted by fitting the data with an alternative set of signal and background templates, obtained by fixing $m_H = m'$ during the DNN evaluation and adjusting the expected signal yields by the corresponding differences in the production cross section and decay rate. This procedure is also applied to the data, yielding for each value of m', a different observed DNN distribution. Figure 5.41 (left) shows the observed local p-value for the combined fit and for each individual production category as a function of m_H in a 5 GeV window around the expected Higgs boson mass. Figure 5.41 (right) shows the expected p-values computed for the combined fit and for each production category on an Asimov data set [151] generated from the background expectation obtained from the signal-plus-background fit injecting a signal at $m_H = 125.38$ GeV. The solid markers indicate the mass points for which the observed p-values are computed. Throughout the explored mass range, $120 < m_H < 130$ GeV, the VBF category has the highest expected sensitivity to $H \to \mu\mu$ decays, followed by the ggH, $t\bar{t}H$, and VH categories, respectively. The observed (expected for $\mu = 1$) significance at $m_H = 125.38$ GeV of the incompatibility with the background-only hypothesis is 3.0 (2.5) σ. Fluctuations in the observed p-value of the VBF category and for the combined fit are due to the nature of the signal extraction fit used in the VBF analysis. When evaluating the DNN for each tested mass point, event migrations in data between neighbouring bins in the high score DNN region produce discrete variations in the observed p-value.

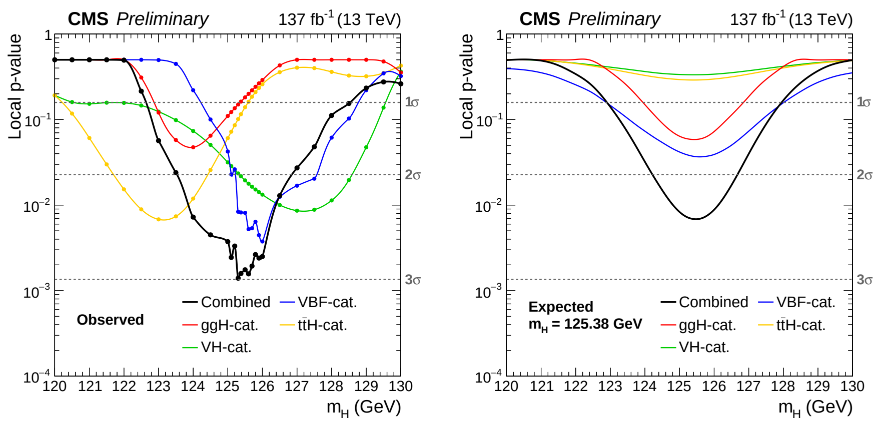


Figure 5.41: p-value vs m_H scan. (Left) observed local p-values as a function of m_H, extracted from the combined fit as well as from each individual production category, are shown. (Right) the expected p-values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$.

5.8.2 Limits on signal strength and $H \to \mu\mu$ branching ratio

The observed (expected for $\mu = 0$) upper limit on μ at 95% CL for $m_H = 125.38$ GeV is 1.9 (0.8). The best-fit signal strength for the Higgs boson with mass of 125.38 GeV, and the corresponding 68% CL interval, is $\hat{\mu} = 1.19\,^{+0.41}_{-0.39}\,(\text{stat})^{+0.17}_{-0.16}\,(\text{syst})$. The statistical component of the post-fit uncertainty is obtained by performing a likelihood scan as a function of μ in which systematic uncertainties are removed. The systematic uncertainty component is then taken as the difference in quadrature between the total and the statistical uncertainties. The individual contributions to the uncertainty in the measured signal strength from experimental uncertainties, the limited size of the simulated samples, and theory uncertainties are also evaluated following a similar procedure. The individual uncertainty components are summarized in Table 5.9, and the dominating uncertainty is the data statistics. Assuming SM production cross sections for the various modes, the $H \to \mu\mu$ branching fraction is constrained at 95% CL to be within $0.8 \times 10^{-4} < \mathcal{B}(H \to \mu\mu) < 4.5 \times 10^{-4}$.

Figure 5.42 (left) reports a summary of the best-fit values for the signal strength and the corresponding 68% CL intervals obtained from a profile likelihood scan in each production category. A likelihood scan is performed in which the four main Higgs boson production mechanisms are associated to either fermion (ggH and $t\bar{t}H$) or vector boson (VBF and VH) couplings. Two signal strength modifiers, denoted as $\mu_{ggH,t\bar{t}H}$ and $\mu_{\mathrm{VBF,V}H}$, are varied independently as unconstrained parameters in the fit. Figure 5.42 (right) shows the 1σ and 2σ contours, computed as variations around the likelihood maximum for $m_H = 125.38\,\mathrm{GeV}$, for the signal strength modifiers $\mu_{ggH,t\bar{t}H}$ and $\mu_{\mathrm{VBF,V}H}$. The best-fit values for these parameters are $\hat{\mu}_{ggH,t\bar{t}H} = 0.66^{+0.67}_{-0.66}$ and $\hat{\mu}_{\mathrm{VBF,V}H} = 1.84^{+0.89}_{-0.77}$, consistent with the SM expectation.

Uncertainty source	$\Delta\mu$	
Total uncertainty	+0.44	−0.42
Statistical uncertainty	+0.41	−0.39
Total systematic uncertainty	+0.17	−0.16
Size of simulated samples	+0.07	−0.06
Total experimental uncertainty	+0.12	−0.10
Total theoretical uncertainty	+0.10	−0.11

Table 5.9: Major sources of uncertainty in the measurement of the signal strength μ and their impact. The total post-fit uncertainty on μ is separated into four components: statistical, size of the simulated samples, experimental, and theoretical.

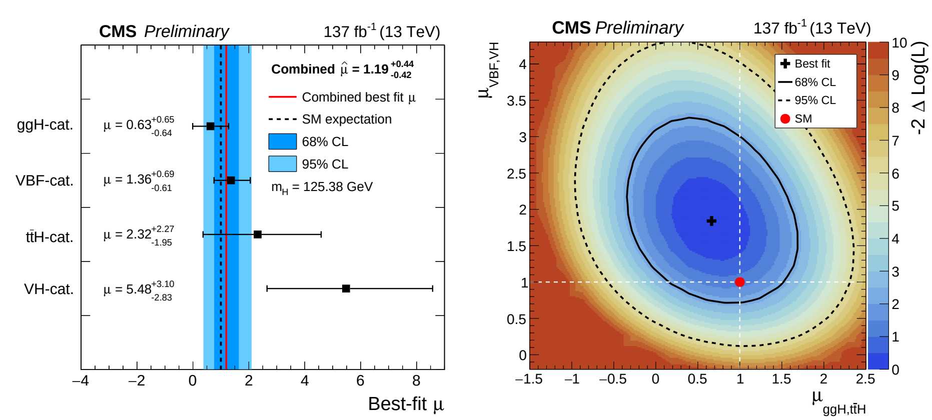


Figure 5.42: Results on the $H \to \mu\mu$ signal strength. (Left) signal strength modifiers measured for $m_H = 125.38\,\mathrm{GeV}$ in each production category (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed grey line). (Right) scan of the profiled likelihood ratio as a function of $\mu_{ggH,t\bar{t}H}$ and $\mu_{\mathrm{VBF,V}H}$ with the corresponding 1σ and 2σ uncertainty contours. The black cross indicates the best-fit values $(\hat{\mu}_{ggH,t\bar{t}H}, \hat{\mu}_{\mathrm{VBF,V}H}) = (0.66, 1.84)$, while the red circle represents the SM expectation.

5.8.3 Combination with CMS Run 1 results

The Run 2 analysis is combined with the 7+8 TeV search from CMS [166], using updated values for the Higgs boson production cross sections and the branching ratios as reported in Ref. [40]. Systematic uncertainties on the inclusive signal production cross sections and $\mathcal{B}(H \to \mu\mu)$ are correlated across the 7, 8, and 13 TeV analyses. Experimental uncertainties affecting the measured properties of the various physics objects (muons, electrons, jets, and b quark jets), the measurement of the integrated luminosity, and the modelling of the pileup conditions are assumed to be uncorrelated between the 7+8 and 13 TeV analyses. The combination improves upon the 13 TeV result by about 1%. Table 5.10 reports the observed and expected significances over the background-only expectation at $m_H = 125.38$ GeV and the 95% CL upper limits on μ in each production category as well as for the 13 TeV and the 7+8+13 TeV combined fits. Figure 5.43 shows the observed (solid black) and the expected (dashed black) local p-values derived from the 7+8+13 TeV combined fit as a function of m_H in a 5 GeV window around the expected Higgs boson mass, using the same strategy as described in Sec. 5.8.1.

Production category	Observed (expected) Significance	Observed (expected) UL on μ
VBF	2.40 (1.77)	2.57 (1.22)
ggH	0.99 (1.56)	1.77 (1.28)
$t\bar{t}H$	1.20 (0.54)	6.48 (4.20)
VH	2.02 (0.42)	10.8 (5.13)
Combined $\sqrt{s}$ = 13 TeV	2.95 (2.46)	1.94 (0.82)
Combined $\sqrt{s}$ = 7, 8, 13 TeV	2.98 (2.48)	1.93 (0.81)

Table 5.10: Observed and expected significances for the incompatibility with the background-only hypothesis for $m_H = 125.38$ GeV and the corresponding 95% CL upper limits on μ (in absence of $H \to \mu\mu$ decays) for each production category as well as for the 13 TeV and the 7+8+13 TeV combined fits.

5.8.4 Higgs couplings to muons

The signal strength measured in the $H \to \mu\mu$ analysis cannot be translated directly into a measurement of the Higgs boson coupling to muons because it is also sensitive to the interactions between the Higgs boson and the top quark/vector bosons. We combined the results of our search with those presented in Ref. [167], to obtain constraints on the Higgs couplings to SM particles.

Under the assumption that there are no BSM particles contributing to the Higgs boson natural width, the Higgs boson production and decay rates in each category are expressed in terms of coupling modifiers within the so-called κ-framework [168,

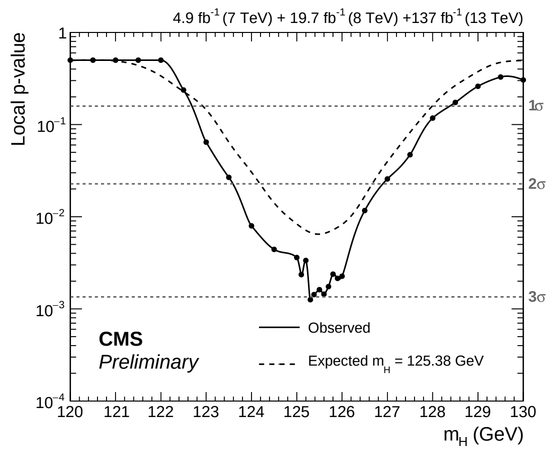


Figure 5.43: Observed (solid black) and expected (dashed black) local p-values as a function of m_H, extracted from the combined fit performed on data recorded at $\sqrt{s} = 7,\ 8,$ and 13 TeV, are shown. The expected p-values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$.

169]. We will define $\kappa_i = y_i/y_i^{SM}$, where y_i is the coupling strength of a SM particle i to the Higgs boson, as observed experimentally and y_i^{SM} is the coupling strength predicted by the SM. Six free coupling parameters are introduced in the likelihood (κ_{W}, κ_Z, κ_{t}, κ_τ, κ_b, and κ_μ) and are extracted from a simultaneous fit across all categories. The coupling modifiers are allowed to float freely and modify both the Higgs boson production cross sections and decay rates, with the constraint of keeping the total Higgs boson width fixed to the SM value.

Figure 5.44 (left) shows the observed profile likelihood ratio as a function of κ_μ for $m_H = 125.38$ GeV. The corresponding 68% and 95% CL intervals for the κ_μ parameter are $0.91 < \kappa_\mu < 1.34$ and $0.65 < \kappa_\mu < 1.53$, respectively. The best-fit values for the various coupling parameters are compatible with the SM prediction, as can be seen in Figure 5.44 (right).

5.9 Future of $H \to \mu\mu$

For HL-LHC, the CMS detector will install a new upgraded tracker system with an extended coverage up to $|\eta| = 2.8$, which will improve the mass resolution of a reconstructed muon pair by a factor of 2, as shown in Fig. 5.45 (left) [170]. The muon spectrometers will also be upgraded [171], and the combined tracker + muon chamber coverage will extend up to $|\eta| = 4.0$. The increased acceptance of muons and improvement in muon p_{T} resolution will be a key factor in improving

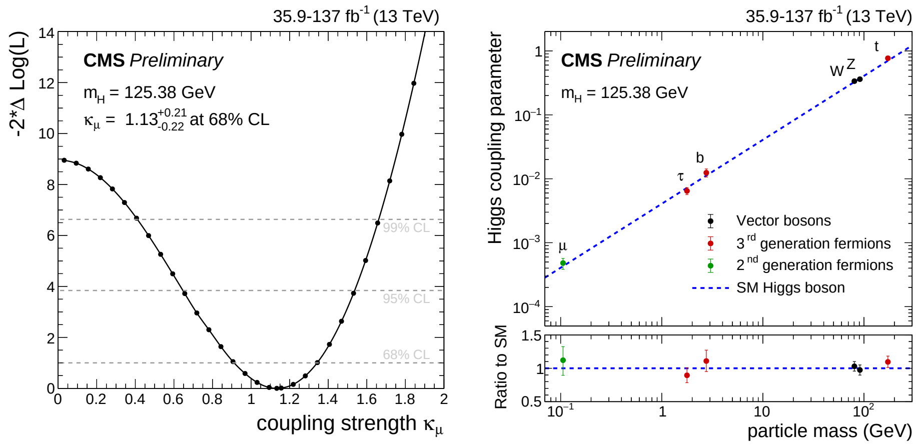


Figure 5.44: Higgs boson coupling modifiers. (Left) observed profile likelihood ratio as a function of κ_μ for $m_H = 125.38\,\text{GeV}$, obtained from a combined fit with Ref. [167] in the κ-framework model. The best-fit value for κ_μ is 1.13 and the corresponding observed 68% CL interval is $0.91 < \kappa_\mu < 1.34$. (Right) the best-fit estimates for the coupling parameters compared to their corresponding prediction from the SM, as function of the particle mass. The error bars represent 68% CL intervals for the measured parameters. The lower panel shows the ratios of the measured coupling modifiers values to their SM predictions.

the $H \to \mu\mu$ analysis sensitivity for HL-LHC. Additionally, the existing ECAL and HCAL detectors will be replaced with a new high granularity calorimeter (HGCAL) [172], which will greatly help with the reconstruction of VBF category jets, due to improved jet-energy-resolution and rejection of pileup jets in the endcap and forward regions of the detector.

A combined ATLAS + CMS projection on the $H \to \mu\mu$ search at the HL-LHC was studied in Ref. [173]. The $H \to \mu\mu$ process is expected to be discovered at the HL-LHC (5 σ excess), and κ_μ is expected to be measured with a total uncertainty of 4%, as shown in Fig. 5.45 (right).

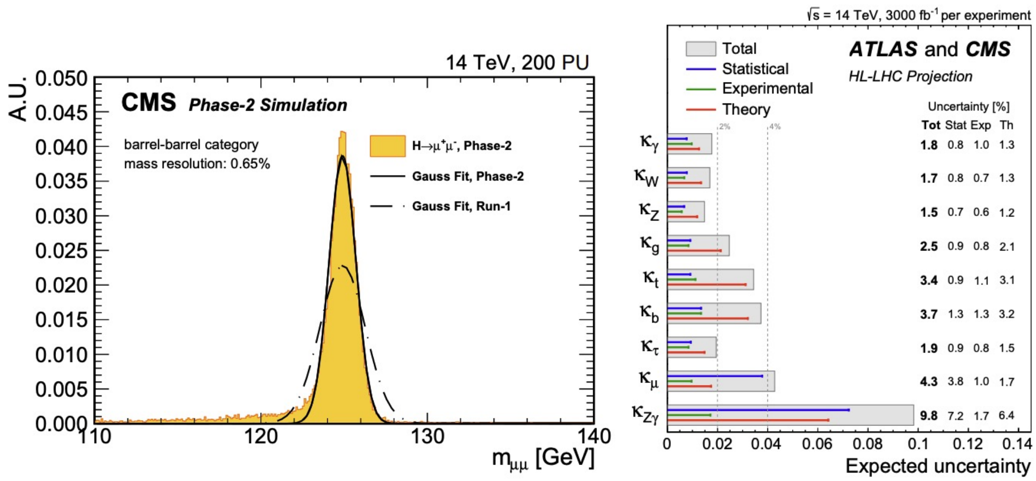


Figure 5.45: HL-LHC projection for $H \to \mu\mu$ channel. (Left) The di-muon invariant mass distribution for $H \to \mu\mu$ decays for muons in the central region, simulated with the Phase-2 CMS detector [170]. (Right) Summary plot showing the total expected $\pm 1\sigma$ uncertainties on the coupling modifier parameters from the combination of ATLAS and CMS measurements at the HL-LHC. For each measurement, the total uncertainty is indicated by a grey box while the statistical, experimental and theory uncertainties are indicated by a blue, green and red line, respectively. κ_μ is expected to be measured with a total uncertainty of 4% [173].

Chapter 6

NONRESONANT PAIR PRODUCTION OF HIGHLY ENERGETIC HIGGS BOSONS DECAYING TO BOTTOM QUARKS

6.1 Introduction

The Higgs boson pair production (HH) in the SM provides unique sensitivity to explore the structure of the Higgs potential. Measurements of its production cross section allow us to directly access the tri-linear Higgs self coupling, λ_{HHH}, and also the quartic coupling between two Higgs bosons and two vector bosons, known as c_{2V}. The value of λ_{HHH} is calculated from the SM as

$$\lambda = \frac{m_H^2}{2v^2} \approx 0.13, \tag{6.1}$$

where v is the Higgs v.e.v. (~246 GeV). As mentioned in Sec. 2.4.2, the dominant HH production happens via the gluon fusion process, and the second most common production mode is the vector boson fusion process. The HH production cross section (~32 fb at 13 TeV) is roughly 3 orders of magnitude smaller than the single Higgs production (~55 pb at 13 TeV), and is therefore a much more difficult process to measure.

The two leading-order diagrams of the gluon fusion production (Fig. 2.4), are known as the box (left) and triangle (right) diagrams. The triangle diagram depends on the λ_{HHH}, and also on the top-quark Yukawa coupling (y_t) through a top-quark loop. The box diagram only depends on y_t, and therefore, when the mass of the Higgs pair (m_{HH}) exceeds two times the top mass threshold ($2m_t \approx 350$ GeV), the HH production probability increases, as can be seen from Fig 6.1. It decreases eventually due to the decreasing probability of finding two high momentum gluons inside the protons. In the SM, the box and triangle diagrams interfere destructively, which makes the $ggHH$ cross section even smaller. The overall m_{HH} distribution peaks near 400 GeV, a very important feature that we will come back to later in this chapter.

To investigate the effect of deviations of the interaction strengths from their SM values, without assuming any particular BSM model, we will express them within

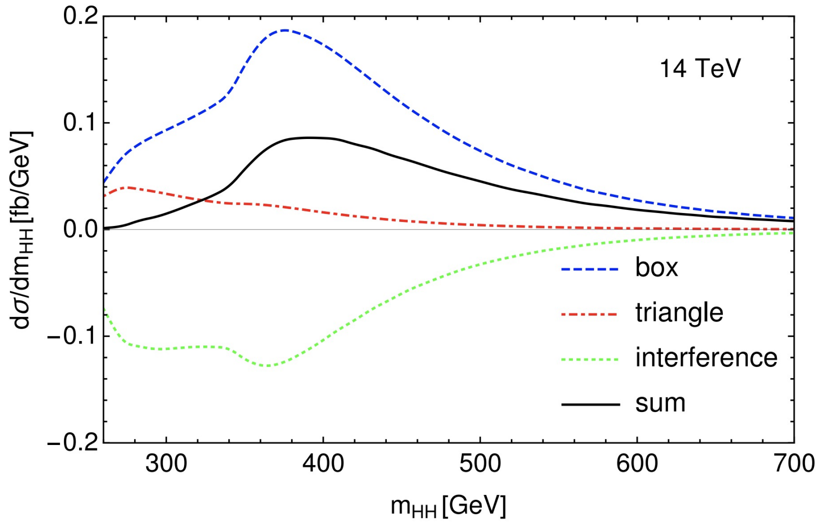


Figure 6.1: In the SM, the box (blue dashed line) and triangle diagram (red dashed line) for the $ggHH$ process interfere destructively. The dependence of the interference term as a function of m_{HH} is shown with the green dashed line. This results in a smaller overall cross-section for the $ggHH$ process, as shown by the black solid line [174].

the κ-framework [168, 169]. In the subsequent sections, we will use $\kappa_\lambda = \lambda/\lambda^{\rm SM}$ and $\kappa_t = y_t/y_t^{\rm SM}$. κ_t has been measured to be consistent with the SM [112, 175]. The effect of having a κ_λ very different from the SM can be understood from Fig. 6.2. For large values of $|\kappa_\lambda|$ (>1), the kinematic peak in the m_{HH} distribution shifts from 400 GeV to 270 GeV. In this region, the triangle diagram dominates over the box diagram, and the Higgs boson in the propagator of $gg \to H^* \to HH$ is off-shell and barely above $2m_H(\approx 250$ GeV). These differences in shape can be used to put strong constraints on the allowed values of κ_λ

The VBF or qqHH production cross section can be parametrized as a function of both the VVH coupling (c_V) and the VVHH coupling (c_{2V}), as shown in Fig. 2.5. In the subsequent sections, we will use $\kappa_{2V} = c_{2V}/c_{2V}^{\rm SM}$ and $\kappa_V = c_V/c_V^{\rm SM}$. For small values of κ_{2V}, the $qqHH$ production cross section increases and leads to a harder m_{HH} spectrum, as shown in Fig. 6.3.

The ATLAS and CMS Collaborations have performed studies of Higgs boson pair production at $\sqrt{s}$ = 7, 8, and 13 TeV in the $b\overline{b}\gamma\gamma$ [58, 178–180], $b\overline{b}\tau^+\tau^-$ [59, 181, 182], $b\overline{b}b\overline{b}$ [57, 183–187], $b\overline{b}VV$ [188–191] channels, as well as combinations of channels [192–194]. For the non-resonant HH production, the current best observed (expected) 95% CL upper limit on the cross section corresponds to 3.1

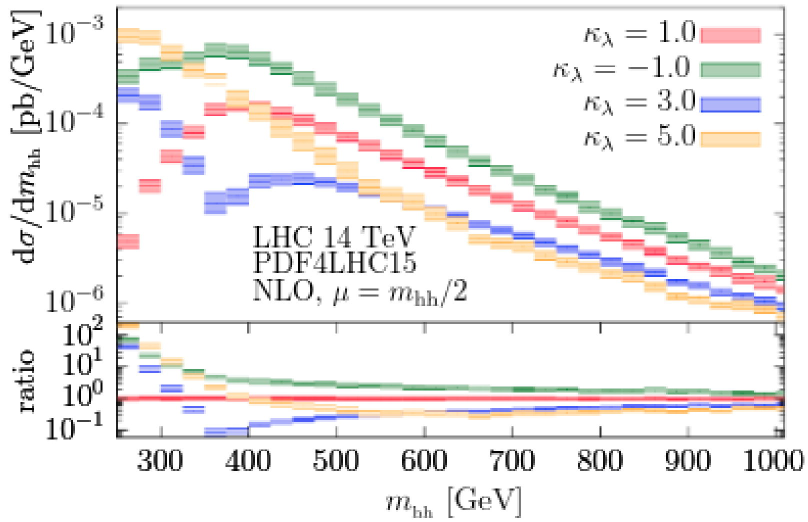


Figure 6.2: The m_{HH} spectra for different values of κ_λ [176].

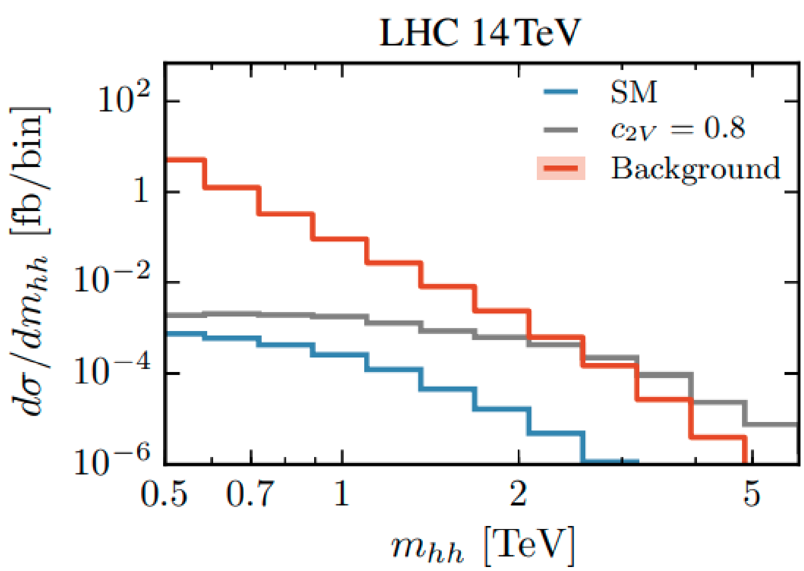


Figure 6.3: The m_{HH} spectra for different values of κ_{2V} [177].

(3.1) × SM [193]. The current best observed (expected) 95% CL constraints on the self-coupling modifier are: $-1.0 < \kappa_\lambda < 6.6$ ($-1.2 < \kappa_\lambda < 7.2$) [193].

This analysis searches for the nonresonant HH production in the $b\bar{b}b\bar{b}$ decay mode where both Higgs bosons decay to two b quarks [26]. Despite having the highest branching ratio amongst all possible HH decays ($\mathcal{B}(HH \to b\bar{b}b\bar{b}) \approx 33.9\%$), this decay channel is traditionally dominated by a large QCD multi-jet background and offers a poor decay channel resolution. We target the phase space where both the Higgs bosons have a high transverse momentum ($p_{\mathrm{T}} > 300$ GeV), otherwise known as the boosted regime. In this region, the two b quarks that decay from each H are sufficiently close together geometrically that they merge into a single large-radius jet. One can then exploit fat-jet sub-structures to obtain a better S/B in this decay channel. Recall that for κ_λ values close to 1, most of the HH signal is populated around m_{HH} = 400 GeV. Therefore, boosted searches like this one have a good sensitivity to SM-like phase space of κ_λ. Additionally, small values of κ_{2V} leads to a harder m_{HH} spectrum, and thus boosted searches also have a good sensitivity to BSM scenarios with small κ_{2V} values. This chapter will be mainly focused on the $ggHH$ analysis, and will give a brief overview of the $qqHH$ analysis. The combination of the $ggHH$ and qqHH analysis is discussed in Sec. 6.6.

To identify these merged Higgs candidates, we use a graph neural network (GNN) based $H \to b\overline{b}$ classifier called ParticleNet [195], explained in more details in Sec. 6.3.4.2. Although the tagger was developed originally for single merged $H \to b\overline{b}$ jet identification, it is significantly more powerful for identifying $HH \to b\overline{b}b\overline{b}$ production, as the requirement for at least one of the Higgs bosons to be produced with a large transverse momentum automatically boosts the transverse momentum of the second Higgs boson. As a result, while the acceptance for the first H to have $p_{\mathrm{T}} > 250\,\mathrm{GeV}$ is only 11%, the acceptance for the second H to have a similarly high p_{T} is large (around 50%). This can be seen from the right plot in Figure 6.4, where we observe that the truth level p_{T} of the second H peaks above 250 GeV, if the p_{T} of the first truth level H is required to be above 250 GeV. Therefore, the requirement for the second H to be successfully tagged as a single large radius $H \to b\overline{b}$ jet achieves about 10^3 background suppression without paying the price of the percent level acceptance.

6.2 Datasets and simulated samples

6.2.1 Datasets

This search uses the pp collision data collected by the CMS detector during the data taking years 2016 (36.33 fb^{-1}), 2017 (41.48 fb^{-1}), and 2018 (59.830 fb^{-1}), corresponding to a total integrated luminosity of 138 fb^{-1}. A combination of several trigger algorithms is used, all requiring the total hadronic transverse energy in the event (H_{T}) or jet p_{T} to be above a given threshold, and/or a combination of b-tagging requirements. In addition, some triggers also have a minimum threshold on the jet mass after removing the remnants of soft radiation [196] (to reduce the H_{T} or p_{T} thresholds). The trigger efficiency varies between 10-95% for jets with

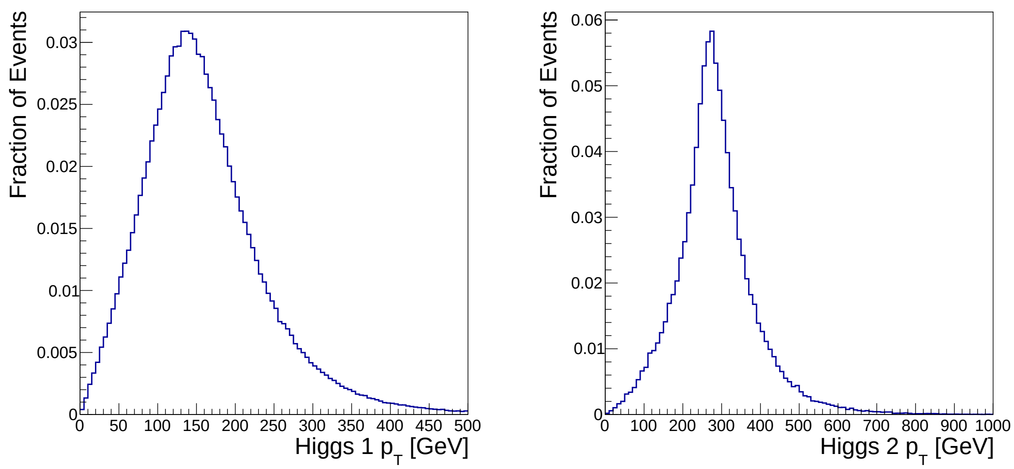


Figure 6.4: Truth level Higgs p_{T} distributions in the HH signal. (Left) The p_{T} distribution of the first Higgs boson in the HH signal. (Right) The p_{T} distribution of the second Higgs boson after requiring $p_{\mathrm{T}} > 250\,\mathrm{GeV}$ for the first Higgs boson.

$300 < p_{\mathrm{T}} < 450\,\mathrm{GeV}$ and is fully efficient for jets with $p_{\mathrm{T}} > 500\,\mathrm{GeV}$. The trigger efficiency measurement is explained in details in 6.4.3.1.

6.2.2 Simulation overview

The processes considered in this analysis have been simulated using either the MadGraph5_amc@nlo 2.6.5 [120] or the powheg 2.0 [197–205] generators. Parton showering, hadronization, and the underlying event are modeled by pythia 8.205 [206] with parameters set by the CUETP8M1 [123] and CP5 tunes [207] used for samples simulating the 2016 and 2017–2018 conditions, respectively. The NNPDF 3.0 [127] and 3.1 [128] parton distribution functions (PDFs) are used in the generation of all simulated samples. The Geant4 [93] package is used to model the response of the CMS detector, and simulated minimum bias interactions are mixed with the hard interactions in simulated events to model additional pp interactions within the same or nearby bunch crossings (pileup). The simulated events are weighted to match the pileup distribution measured in data.

6.2.2.1 Signal simulation

Simulated samples for the ggF process are generated at next-to-leading order (NLO) accuracy using powheg 2.0, and corrected as a function of HH mass (m_{HH}) based on Ref. [208]. The VBF HH samples are generated at leading order (LO) accuracy using MadGraph5_amc@nlo. A range of samples corresponding to different combinations of the κ_V, κ_{2V}, and κ_λ couplings are generated (see Table 6.1). Samples for other coupling combinations are constructed as linear combinations of the original generated samples by applying appropriate event weights as described in Ref. [57].

Process	$\sigma \times \mathcal{B}(HH \to b\bar{b}b\bar{b})$ [fb]
ggHH SM (κ_λ =1)	10.53
ggHH BSM (κ_λ =0)	23.64
ggHH BSM (κ_λ =2.45)	4.46
ggHH BSM (κ_λ =5)	31.10
VBF HH SM (κ_λ =1, κ_V =1, κ_{2V} =1)	0.585
VBF HH BSM (κ_λ =1, κ_V =1, κ_{2V} =0)	9.169
VBF HH BSM (κ_λ =1, κ_V =1, κ_{2V} =2)	4.823
VBF HH BSM (κ_λ =0, κ_V =1, κ_{2V} =1)	1.558
VBF HH BSM (κ_λ =2, κ_V =1, κ_{2V} =1)	0.482
VBF HH BSM (κ_λ =1, κ_V =0.5, κ_{2V} =1)	3.656
VBF HH BSM (κ_λ =1, κ_V =1.5, κ_{2V} =1)	22.412

Table 6.1: Summary of various HH signal simulation samples used. The cross sections reported include the branching fraction of $HH \to b\bar{b}b\bar{b}$.

The ggF samples are normalized to the NNLO cross sections [41–43, 45–47, 209, 210] corresponding to the coupling values considered. The VBF sample with SM couplings is normalized to the N^{3}LO cross section [51], and the same N^{3}LO/LO correction is applied to the VBF samples with modified couplings.

6.2.2.2 Background simulation

Samples for $t\bar{t}$ background process are generated at NLO accuracy using POWHEG v2.0 [197–199, 211, 212] with the FxFx jet matching and merging scheme [126], and normalized to the theoretical cross section calculated at NNLO precision using TOP++ v2.0 [213]. The differential $t\bar{t}$ cross section as a function of the top quark p_T is corrected to the NNLO QCD + NLO electroweak accuracy [214]. Samples for other background processes, such as the QCD multijet process, single top quark, Z and W single-boson and diboson production processes, as well as single Higgs boson production in all relevant production modes, are also generated.

6.3 Physics objects

The main physics objects used in this analysis are two large radius jets and their associated $H \to b\bar{b}$ taggers. We additionally use muons and electrons to derive certain corrections to the top quark background in the semi-leptonic top control regions. Small radius jets are used to identify VBF jets that are used to explicitly exclude events selected by the qqHH analysis.

6.3.1 Muons

Muons used in this analysis are reconstructed using the tight ID (see Sec. 5.3.2) and satisfy certain relative isolation conditions, called "mini-isolation", first suggested in Ref. [215]. When jets are boosted, their decay products are highly collimated, and the chance of accidental overlap from leptons or other jets/pile-up is lower. In a two-body decay of a massive particle with mass M and large p_{T}, the angular separation of the daughter particles scales as $\Delta R \approx 2M/p_{\mathrm{T}}$. One can then define lepton p_{T}-dependent cone sizes to remove overlaps between b-jets and leptons originating from a boosted top quark. We define the cone size $R^{mini-iso}$ as:

$$R^{mini-iso} = \begin{cases} 0.2, & p_{\mathrm{T}}^{l} < 50\,\mathrm{GeV} \\ 10/p_{\mathrm{T}}^{l}, & 50\,\mathrm{GeV} < p_{\mathrm{T}}^{l} < 200\,\mathrm{GeV} \\ 0.05, & p_{\mathrm{T}}^{l} > 200\,\mathrm{GeV}. \end{cases} \tag{6.2}$$

Relative mini-isolation, I_{rel}^{mini}, is then defined as the transverse energy of particles (with a $\delta\beta$ correction, as described in Sec. 5.3.2) in a cone of radius $R^{mini-iso}$ around the lepton, divided by the lepton p_{T}. We require muon candidates to have $I_{rel}^{mini} < 0.2$, $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ and $|\eta| < 2.4$.

6.3.2 Electrons

Electron candidates used in this analysis are reconstructed as described in Sec. 5.3.5. We impose mini-isolation (see Sec. 6.3.1) requirements on the electron as $I_{rel}^{mini} <$ 0.2. Electrons are required to have $p_{\mathrm{T}} > 50\,\mathrm{GeV}$ and $|\eta| < 2.5$.

6.3.3 AK4 jets

AK4 jets are small radius jets with cone size R=0.4, and have been previously described in Sec. 5.3.3. In order to remain orthogonal to the VBF HH analysis, we use the AK4 jets to identify and veto VBF signal events. The selected jets are required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 4.7$.

6.3.4 AK8 jets

The large radius jets used in this analysis are reconstructed by clustering PF candidates with the anti-k_T algorithm [90], using a distance parameter of $R = 0.8$, and are called AK8 jets. The PUPPI algorithm [216] is used to correct for pileup contributions to the jet. Certain quality requirements are imposed on the jets to suppress the contamination from pileup jets and mis-reconstructed jets due to detector noise in the fiducial region beyond the coverage of the tracker. Corrections to the measured jet energies are applied according to the description in Sec. 5.3.3. The jet mass scale and jet mass resolution corrections for this analysis are derived in a boosted semi-leptonic top control region, using a muon as a "tag" object to select suitable events. The AK8 jet in the opposite hemisphere (representing a top/W jet) is then used as a "probe" object to derive correction factors as data-to-simulation ratios.

We select AK8 jets with $p_T > 300\,\text{GeV}$, and $|\eta| < 2.4$. For an AK8 jet resulting from a boosted $H \to b\bar{b}$ decay, we expect the jet to have a two-prong sub-structure. We will exploit this feature and use certain special jet sub-structure identifiers to tag our AK8 jets, as discussed in the next few sub-sections.

6.3.4.1 Soft-drop mass

The soft-drop algorithm [217] is a procedure to remove soft and wide-angled radiation from a jet. This helps to remove contamination from initial state radiation (ISR), underlying event (UE) and pile-up. The procedure employed is as follows: the last step of a jet clustering process is undone, resulting in two separate jet constituents. The softer constituent is discarded from the jet if the following condition fails:

$$\frac{min(p_T^1, p_T^2)}{p_T^1 + p_T^2} = z_{cut}\left(\frac{\Delta R_{12}}{R_o}\right)^{\beta}. \tag{6.3}$$

Here, R_o is the jet radius, p_T^i are the transverse momenta of the two jet constituents, ΔR_{12} is the angular separation between the two constituents in the $\eta - \phi$ plane, and β is an angular exponent. If the softer constituent gets dropped, the remaining jet constituent is again de-clustered into two separate constituents. This process is repeated recursively, until no further de-clustering is possible.

For this analysis, we use the “modified mass drop tagger” [218] approach, with angular exponent $\beta = 0$, soft cutoff threshold $z < 0.1$, and characteristic radius $R_o = 0.8$. This grooming process is known to improve the jet mass resolution, and hence we will use the "soft-drop mass" or m_{SD} of an AK8 jet in our analysis.

6.3.4.2 ParticleNet Xbb tagger

The most crucial feature of the $HH \to b\bar{b}b\bar{b}$ search is the efficient reconstruction of $H \to b\bar{b}$ jet, and achieving good rejection of fake jets originating from light quarks (u, d, s, c) and gluons. b-jet identification relies on the large lifetime of B hadrons ($\sim$ 150 ps), which results in a displaced secondary vertex and displaced tracks inside the jet, and possibly non-isolated electrons and muons (see Fig. 6.5).

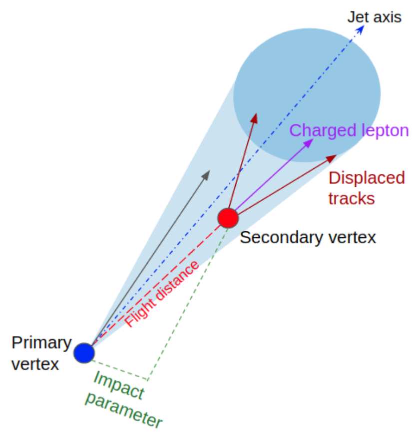


Figure 6.5: The hadronization of a b quark results in a B meson, which travels $O(\sim\ mm)$ before decaying, resulting in a displaced secondary vertex (SV) inside the jet. This feature is exploited for b-jet tagging.

This is one of the first CMS searches that selects AK8 jets originating from a $H \to b\bar{b}$ decay, using the novel ParticleNet jet tagger [195, 219]. The ParticleNet tagger is a dynamic graph convolutional neural network [220] trained to classify jets according to their flavour type. Compared to the previous state-of-the-art jet tagging algorithms, like DeepAK8 tagger [221], ParticleNet relies on a highly sophisticated jet representation and network architecture. The ParticleNet algorithm treats a jet as an unordered set (permutation-invariant) of its constituents ("particle cloud"). This technique provides a natural representation of the jet , unlike other algorithms that treat jets as images or particles ordered by their p_{T}. The graph network architecture allows the algorithm to efficiently explore the correlations between the various jet constituents. Particle-flow candidates and secondary vertices within the jet cone are used as inputs to the algorithm. These input variables are detailed in Table 6.2.

The ParticleNet algorithm has two versions:

- the "nominal" version is designed for maximum performance but may introduce sculpting in the mass spectrum of the background jets,

Table 6.2: Input variables of each jet used by the ParticleNet tagger are the constituent particle-flow candidates and secondary vertices (SV). The track-based variables are only defined for charged particles, and a value of 0 is assigned to neutral particles.

Variable	Definition
	For particle-flow candidates.
$\log p_{\mathrm{T}}$	logarithm of the particle's p_{T}
$\log E$	logarithm of the particle's energy
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the particle and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the particle and the jet axis
$\lvert\eta\rvert$	absolute value of the particle's pseudorapidity
q	electric charge of the particle
`pvAssociationQuality`	flag related to the association of the track to the primary vertices
`lostInnerHits`	quality flag of the track related to missing hits on the pixel layers
χ^2/dof	χ^2 value of the trajectory fit normalized to the number of degrees of freedom
`qualityMask`	quality flag of the track
d_z	longitudinal impact parameter of the track
d_z/σ_{d_z}	significance of the longitudinal impact parameter
d_{xy}	transverse impact parameter of the track
$d_{xy}/\sigma_{d_{xy}}$	significance of the transverse impact parameter
η_{rel}	pseudorapidity of the track relative to the jet axis
$p_{\mathrm{T,rel}}$ ratio	track momentum perpendicular to the jet axis, divided by the magnitude of the track momentum
$p_{\mathrm{par,rel}}$ ratio	track momentum parallel to the jet axis divided by the magnitude of the track momentum
$d_{3\mathrm{D}}$	signed 3D impact parameter of the track
$d_{3\mathrm{D}}/\sigma_{3\mathrm{D}}$	signed 3D impact parameter significance of the track
`trackDistance`	distance between the track and the jet axis at their point of closest approach
	For Secondary Vertices (SVs) within the jet cone.
$\log p_{\mathrm{T}}$	logarithm of the SV's p_{T}
m_{SV}	mass of the SV
$\Delta\eta(\text{jet})$	difference in pseudorapidity between the SV and the jet axis
$\Delta\phi(\text{jet})$	difference in azimuthal angle between the SV and the jet axis
$\lvert\eta\rvert$	absolute value of the SV's pseudorapidity
N_{tracks}	number of tracks associated with the SV
χ^2/dof	χ^2 value of the SV fit normalized to the number of degrees of freedom
$d_{2\mathrm{D}}$	signed 2D impact parameter (i.e., in the transverse plane) of the SV
$d_{2\mathrm{D}}/\sigma_{2\mathrm{D}}$	signed 2D impact parameter significance of the SV
$d_{3\mathrm{D}}$	signed 3D impact parameter of the SV
$d_{3\mathrm{D}}/\sigma_{3\mathrm{D}}$	signed 3D impact parameter significance of the SV

- the "mass decorrelated" (MD) version which is designed to be largely decorrelated with respect to the mass of a jet, at the cost of slight degradation in the discrimination power.

To train the mass decorrelated tagger [219], the same inputs and network architecture as the nominal ParticleNet algorithm are used. Dedicated simulation samples were used as follows:

- Signal : Lorentz-boosted spin-0 particles (X), with a flat mass spectrum between 15 to 250 GeV, and subsequently decaying to a pair of quarks (X $\rightarrow b\overline{b}$, X $\rightarrow c\overline{c}$, X $\rightarrow q\overline{q}$).
- Background : QCD multijet process.

Jets from the signal and background samples were also re-weighted to yield flat distributions in both p_{T} and m_{SD} for the training. We use the mass decorrelated version in this analysis.

The tagger assigns a set of output classifier scores for each jet, corresponding to the probability of the jet originating from a resonance that decays into a pair of quarks ($X \to b\overline{b}$, $X \to c\overline{c}$, $X \to q\overline{q}$, i.e., X decaying to light quarks), or non-resonant quark-and-gluon jet (QCD_{bb}, QCD_b, QCD_{cc}, QCD_c and QCD_{others}).

To focus on the discrimination power between $X \to b\overline{b}$ and QCD jets, the $b\overline{b}$-tagging discriminant T_{Xbb} that we use for this analysis is defined to be:

$$T_{\mathrm{Xbb}} = \frac{P_{\mathrm{Xbb}}}{P_{\mathrm{QCDbb}} + P_{\mathrm{QCDb}} + P_{\mathrm{QCDcc}} + P_{\mathrm{QCDc}} + P_{\mathrm{QCDothers}} + P_{\mathrm{Xbb}}} \\ = \frac{P_{\mathrm{Xbb}}}{1 - P_{\mathrm{Xcc}} - P_{\mathrm{Xqq}}}, \quad (6.4)$$

where the last equality comes from the fact that the algorithm is trained such that the output probabilities sum to one, $P_{\mathrm{QCDbb}} + P_{\mathrm{QCDb}} + P_{\mathrm{QCDcc}} + P_{\mathrm{QCDc}} + P_{\mathrm{QCDothers}} + P_{\mathrm{Xbb}} + P_{\mathrm{Xcc}} + P_{\mathrm{Xqq}} = 1$. Therefore, our $X \to b\overline{b}$ tagger is specifically optimized to enhance $X \to b\overline{b}$ resonances over all of the other non-resonant QCD-type jets and is constrained to be between 0 and 1. The tagger score distribution for signal and background jets is demonstrated in Fig. 6.6.

The performance of the tagger is shown in the ROC curve in Figure 6.7. We achieve a background rejection of about 200 for a working point that yields 50% signal efficiency.

6.3.4.3 ParticleNet Xbb mass regression

We will use the mass distribution of one of the AK8 jets to extract the final signal strength in this analysis (explained later in Sec. 6.4.1), and therefore, the sensitivity of this analysis is driven by the jet mass resolution.

Nominally, we use the mass distribution obtained after applying the soft drop algorithm, as described in Sec. 6.3.4.1. This grooming algorithm removes soft and wide angled constituents from the jets, likely coming from background sources such as the underlying event or pile-up. The resulting m_{SD} distribution peaks at the Higgs boson mass for signal events and pushes the mass of background quark and gluon initiated events towards lower values. However, depending on the level of grooming,

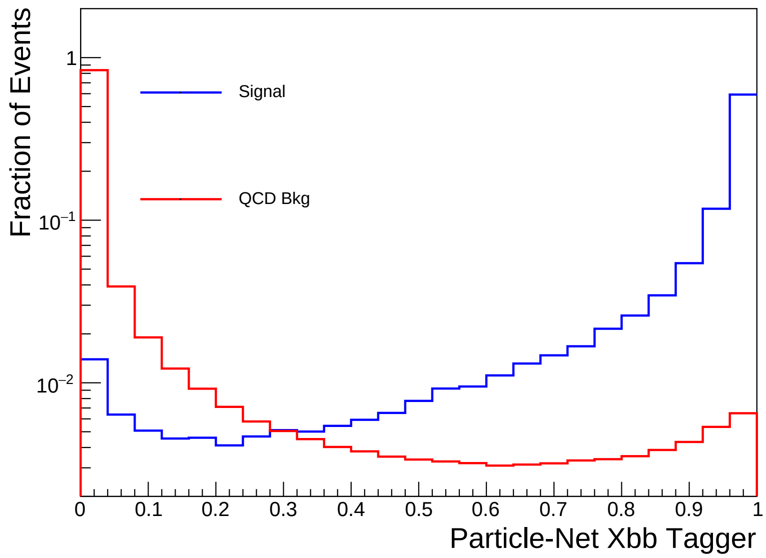


Figure 6.6: The distribution for the ParticleNet $X \to b\bar{b}$ tagger is shown for QCD background and HH signal. The events selected for this figure contain two AK8 jets with $p_{\mathrm{T}} > 300$ GeV and having a soft drop mass > 40 GeV each. Here the Particle-Net Xbb Tagger is the $b\bar{b}$-tagging discriminant, T_{Xbb}, as mentioned in the text.

it can also over-subtract genuine hard jet constituents from the Higgs jet and this results in low values of m_{SD} even for the signal.

To improve the signal efficiency, we will use the mass distribution resulting from the ParticleNet Mass Regression. This algorithm uses the same input variables (Table 6.2) and network architecture as the ParticleNet $X \to b\bar{b}$ tagger, with the exception of the final probabilistic output of the tagger being replaced by a jet mass regression. As opposed to the classification case for the tagger, the regressed value to be learned is a single real number, known as the target mass M_{Target}. The most naive training procedure would be to simply use a soup of boosted Higgs and QCD jets and define M_{Target} as 0 for QCD jets and m_H otherwise. However, this approach would cause the network to learn only two definite values for the mass (and would be equivalent to a binary classification), which would result in substantial sculpting of the QCD jets mass distribution near the Higgs peak. A more sensible choice is to use a continuous target for QCD, such as the soft drop mass computed using hadron-level constituents of the truth level jet or m_{SD}^{gen}. This resolves the problem of the binary classification, however, we are still left with a substantial fraction of QCD jets where $M_{Target} = m_H$. To avoid this scenario, the signal simulation

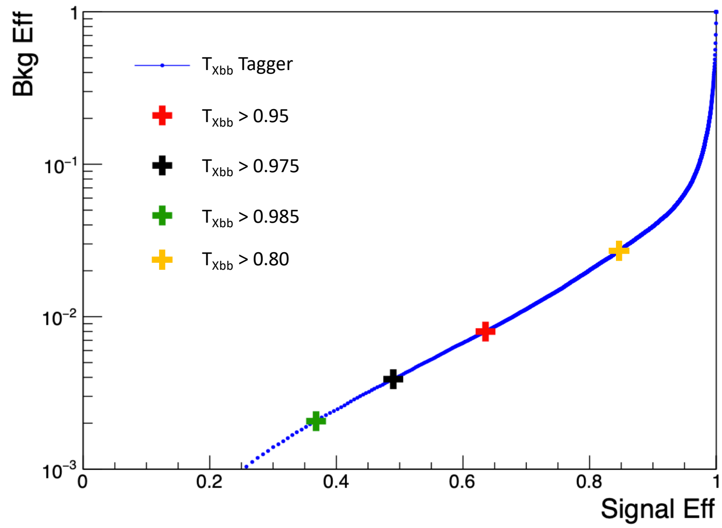


Figure 6.7: The ROC curve for the ParticleNet $X \to b\bar{b}$ tagger is shown in this figure. The signal here corresponds to reconstructed AK8 jets matched to truth level Higgs bosons from the $ggHH$ simulation samples, and the background corresponds to AK8 jets from the QCD multijet sample. The events selected for this figure contain two AK8 jets with $p_{\mathrm{T}} > 300$ GeV and a soft drop mass > 40 GeV each. The signal efficiency is plotted against the background efficiency as one scans the tagger score for the jets. The red, black, green, and orange crosses correspond to various working points for the tagger, relevant for this analysis.

samples used to train the regression were the same samples that were used to train the mass-decorrelated tagger (Sec. 6.3.4.2), where the spin-0 X particle mass ranges as $M_X \in [15, 250]$ GeV. The target mass is defined as

$$M_{Target} = \begin{cases} m_{SD}^{gen} & \text{if jet is QCD;} \\ M_X & \text{otherwise.} \end{cases} \tag{6.5}$$

The training was performed using AK8 jets that satisfy $p_{\mathrm{T}} > 170$ GeV and $m_{SD}^{gen} < 260$ GeV for QCD jets. The jets from both signal and background samples are re-weighted such that their distribution in the 2-dimensional $(\log(p_{\mathrm{T}}), M_{Target})$ space is uniform. The mass regression is trained using a "log-cosh" loss function defined as:

$$L(m_{\mathrm{reg}}, M_{Target}) = \sum_{i \in jets} \log(\cosh(m_{\mathrm{reg}}^{(i)} - M_{Target}^{(i)})). \tag{6.6}$$

where m_{reg} is the regressed mass output from the network.

The performance of the mass regression is summarized in Figure 6.8, where the AK8 jet mass distributions reconstructed with the soft drop and regression algorithms are compared for one of the categories of the $ggHH$ analysis (defined later in Sec. 6.4.1.1). This regression technique improves the jet mass resolution by ~3% and also the mass scale by ~1%, w.r.t the soft drop algorithm.

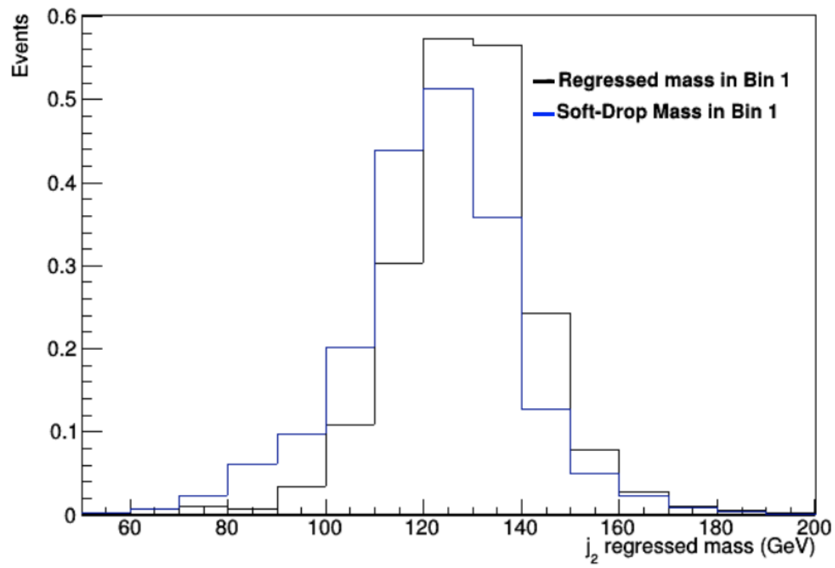


Figure 6.8: Comparison of the mass resolution for one of the $H \to b\bar{b}$ jets from the SM ggHH signal in the most sensitive category of this analysis (category 1, as explained in Sec. 6.4.1.1). The performance of the soft drop mass algorithm is shown in blue and the regressed mass algorithm is shown in black. The regressed jet mass algorithm improves our overall signal sensitivity by 25%.

6.3.4.4 N-subjettiness

QCD jets have a fundamentally different pattern of energy deposits in the detector when compared to a boosted W/Z/top jet. To identify an N sub-jet structure within a jet, we define a variable called the N-subjettiness τ_N as:

$$\tau_N = \frac{1}{d_o} \sum_k p_{\mathrm{T}}^k min\{\Delta R_{1,k}, \Delta R_{2,k}, .., \Delta R_{N,k}\} \tag{6.7}$$

$$d_o = \sum_k p_{\mathrm{T}}^k R_o. \tag{6.8}$$

Here, k runs over the constituent particles in a given jet, p_{T}^k are their transverse momenta, $\Delta R_{J,k}$ is the distance in the $\eta - \phi$ plane between a candidate subjet-J and a constituent particle k and R_o is the jet radius. For $\tau_N \approx 0$, all the radiation is along the candidate subjet, and the jet has N (or fewer) subjets. For $\tau_N \gg 0$, a large fraction of the jet energy is scattered away from the candidate subjet, which implies the jet has at least N+1 subjets. τ_3 is expected to be small for top jets. However, QCD jets can randomly have small values of τ_3 too, and to increase the discrimination power, τ_3/τ_2 or τ_{32} is used as an identifier for top jets. Smaller the value of τ_{32}, more likely it is that the jet is a top jet.

We require $\tau_{32} < 0.46$ (with a mis-tag rate of 0.5%) in certain hadronic top control regions (described later in Sec. 6.4.2.1), for selecting a very pure sample of top jets. Simulated events in the top quark control regions are corrected in order to take into account differences in the top-tagging efficiency between data and MC.

6.3.5 Missing Transverse Energy (MET)

The missing transverse momentum used in this analysis has been previously described in Sec. 5.3.4.

6.4 The $ggHH$ analysis

The next few sub-sections will discuss the $ggHH$ analysis in detail.

6.4.1 Event selection

We select events with at least two AK8 jets with $p_T > 300\,\mathrm{GeV}$. If there are more than two such AK8 jets in an event, then the two jets with the highest T_{Xbb} score are selected as H candidates. The jets are arranged in a descending order of their T_{Xbb} score. We apply additional pre-selection requirements as follows:

- Two or more AK8 jets with $p_T > 300\,\mathrm{GeV}$,
- Jet 1 soft drop mass > 50 GeV,
- Jet 2 regressed mass > 50 GeV,
- Jet 1 $T_{\mathrm{Xbb}} > 0.8$.

We use the mass distribution of the 2nd jet (jet with second highest T_{Xbb} score) to extract the final signal strength and hence impose requirements on the regressed mass for Jet 2. The second jet in VH background events tends to be the W or Z boson instead of the Higgs boson. In the SM, the $W \to b\bar{b}$ decay is prohibited and the $Z \to b\bar{b}$ decay has a branching ratio of 15% only. Hence, the W or Z boson tends to have a lower T_{Xbb} score as compared to the H jet in VH background events and gets classified as the second jet, as shown in Figure 6.9. Thus, the second jet mass has a significantly larger distinguishing power than the first jet mass.

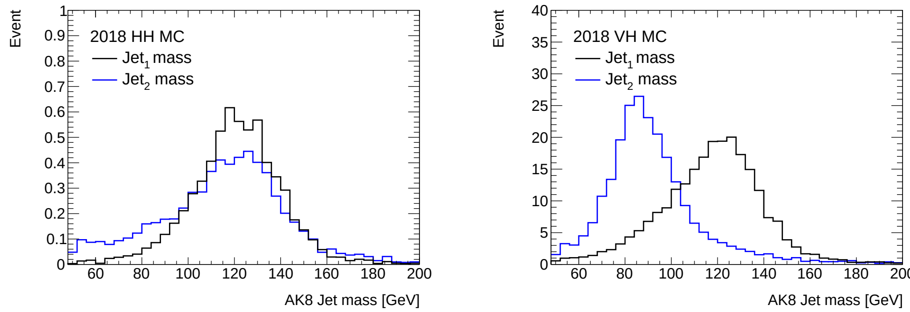


Figure 6.9: After pre-selection requirements, the distributions of the first and second jet mass for HH signal (left) and VH background (right) are shown in this figure. The second jet in VH background events tends to be the W or Z boson instead of the Higgs boson.

The $T_{\mathrm{Xbb}} > 0.8$ requirement helps to reduce the QCD jet contributions, resulting in a total background that is roughly 50% QCD multijet and 50% $t\bar{t}$ background. We also impose orthogonality to the VBF category (see Sec. 6.5.1 for description), by removing events which satisfy all of the following conditions :

- ≥ 2 AK4 jets (j), with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 4.7$,
- the AK4 jets do not overlap with leptons, i.e., $\Delta R(j, \ell) > 0.4$, where ℓ are muons (electrons) with $p_{\mathrm{T}} > 5(7)$ GeV, and,
- the two highest-p_{T} AK4 jets have $m_{jj} > 500$ GeV and $\Delta\eta_{jj} > 4.0$.

Lastly, we designed a boosted decision tree (BDT) discriminator trained to enhance HH signal over QCD multijet and top quark background. We will define several event categories based on a 2D grid of the output score of the BDT discriminator as well as the T_{Xbb} score for the second jet. The BDT is described in the following section.

6.4.1.1 The BDT architecture and performance

To improve on the search sensitivity for the $ggHH$ analysis, we train a boosted decision tree (BDT) to discriminate between the HH signal events, and QCD and top quark background events. The input training variables for the BDT include the

Jet 1 and dijet system kinematics, the p_{T} of Jet 2, the τ_{32} values of both jets and the various ParticleNet scores for Jet 1. As mentioned in Sec. 6.4.1, our final fit categories will be based on a 2D grid of the output score of the BDT discriminator as well as the T_{Xbb} score for the second jet. Hence, we need to have zero correlation between the BDT output score and the Jet 2 T_{Xbb} score and we explicitly do not include the Jet 2 ParticleNet scores for training the BDT. The mass of the dijet system (m_{jj}) is calculated using the soft drop and regressed masses of jet 1 and jet 2, respectively. The full list of training variables is shown in Table 6.3. The data and simulation comparisons of these input variables are shown later in Sec. 6.4.1.2.

Table 6.3: List of input variables used for the BDT training.

Variable	Definition
$p_{\mathrm{T}jj}$	Dijet system p_{T}
η_{jj}	Dijet system η
m_{jj}	Dijet system mass
$p_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse energy
τ_{32}^{j1}	Jet 1 τ_{32}
τ_{32}^{j2}	Jet 2 τ_{32}
m_{SD}^{j1}	Jet 1 soft-drop mass
p_{T}^{j1}	Jet 1 p_{T}
η^{j1}	Jet 1 η
T_{Xbb}^{j1}	Jet 1 T_{Xbb} score
P_{QCDb}^{j1}	Jet 1 QCD_b score
P_{QCDbb}^{j1}	Jet 1 QCD_{bb} score
$P_{\mathrm{QCDothers}}^{j1}$	Jet 1 QCD_{others} score
p_{T}^{j2}	Jet 2 p_{T}
$p_{\mathrm{T}}^{j1}/m_{jj}$	Jet 1 p_{T} over dijet system mass
$p_{\mathrm{T}}^{j2}/m_{jj}$	Jet 2 p_{T} over dijet system mass
$p_{\mathrm{T}}^{j1}/p_{\mathrm{T}}^{j2}$	Jet 2 p_{T} / Jet 1 p_{T}

The BDT is trained using the *XGBoost* method. 60% of the events were used for training and validation, and the remaining 40% were used for testing. The parameters used for the BDT training are as follows:

- Number of trees = 400
- Maximum depth = 3
- Learning rate = 0.1

- L2 regularization strength : 1.0.

We specifically ensured that none of the training variables induced any sculpting of the Jet 2 mass as tighter requirements on the BDT output score are imposed. In Figure 6.10, we show the shape of the Jet 2 regressed mass in the QCD background simulation as the requirement on the BDT score is increased. The plot shows that the background has an exponentially falling shape for different selections cuts on the BDT output score, and no sculpting is observed.

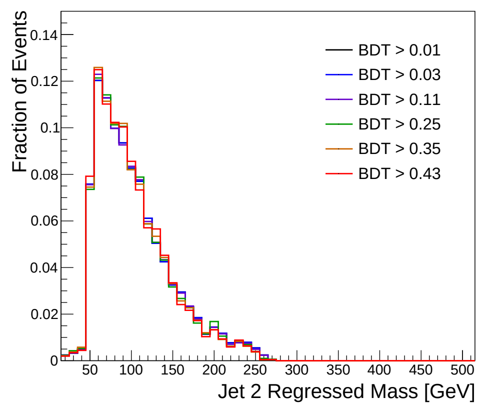


Figure 6.10: The jet 2 mass distribution is shown for QCD multijet background from simulation samples as the requirements on the BDT output score is tightened. No mass sculpting is observed here.

The discriminator output score shape for the $ggHH$ signal and main backgrounds are shown in Figure 6.11. The backgrounds tend to peak at lower BDT score values, and the signal has a higher concentration at the larger BDT score values.

The performance of the BDT is shown in a ROC curve in Figure 6.12, where we plot the signal efficiency against the background efficiency as one scans the requirement on the BDT output score. The background here consists of both the $t\bar{t}$ and QCD. To estimate the power of the BDT, we define a simple alternative cut based selection strategy (not used in this analysis) as follows:

- Jet 1 $p_{\mathrm{T}} > 350\,\mathrm{GeV}$, Jet 2 $p_{\mathrm{T}} > 310\,\mathrm{GeV}$
- Jet 1 soft drop mass $\in [105, 135]\,\mathrm{GeV}$,
- Jet 2 regressed mass $> 50\,\mathrm{GeV}$,
- Jet 1 $T_{\mathrm{Xbb}} > 0.985$.

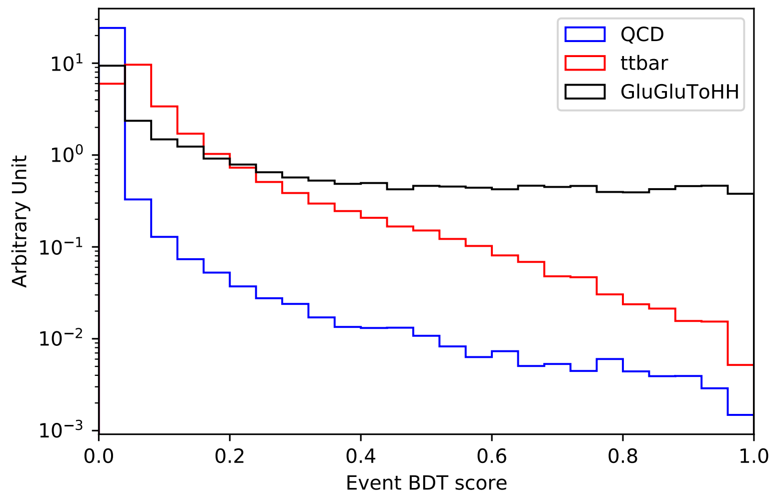


Figure 6.11: The $ggHH$ signal and main background distributions of the BDT output score are shown in this figure. The histograms are normalized to unity. The backgrounds tend to peak at lower BDT score values, and the signal has a higher concentration at the larger BDT score values.

To make a fair comparison against the cut-based analysis and the BDT based analysis, we compare the value of the signal efficiency at Jet 2 $T_{\mathrm{Xbb}} > 0.98$. We observe that at a signal efficiency equal to the cut-based analysis selection, the BDT has twice the background rejection power.

6.4.1.2 Data and MC comparisons of BDT input variables

The most important BDT variables are shown in Fig. 6.13, in the hadronic $t\bar{t}$ control region (defined later in Sec. 6.4.2.1). The data is well described by the simulation for these input variables.

6.4.1.3 BDT event categories

To obtain event categories with an optimized signal sensitivity, we scan across different values of the BDT output score and the T_{Xbb} score of the second jet, and compute the signal and background yields for Jet 2 mass in the range 120–130 GeV. The total background prediction in the range 120–130 GeV is obtained using a simple linear interpolation of the data in the Jet 2 mass sidebands, i.e., $[110, 120] \cup [130, 140]$ GeV. We then choose those working points for the BDT score and the Jet 2 T_{Xbb}, that yield the best expected upper limit on the SM $ggHH$ cross section. We keep repeating this procedure, excluding the events that were

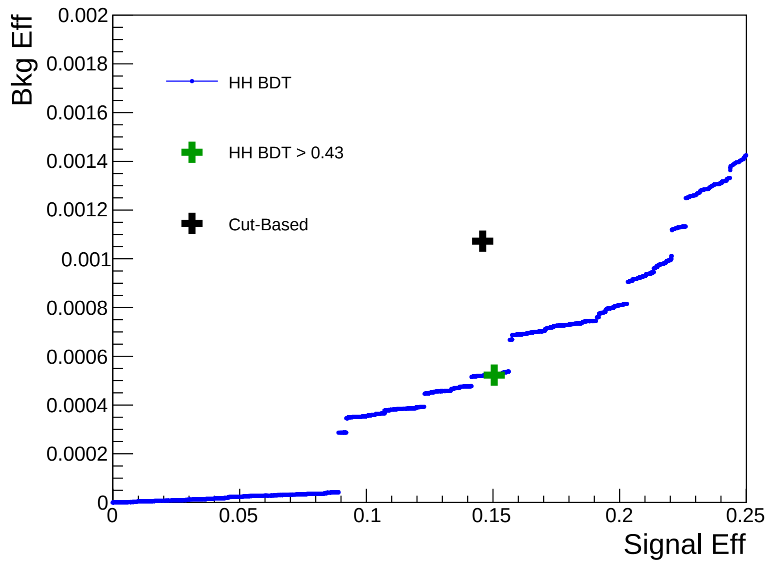


Figure 6.12: The ROC curve for the BDT is shown in this figure. The signal efficiency vs the background efficiency is plotted as one scans the requirement on the BDT output score. The black cross corresponds to the working point for the cut-based analysis selection (defined in text). The green cross corresponds to BDT score > 0.43, which is the requirement in the most optimized analysis category (defined in Sec. 6.4.1.3). The discontinuities observed here arise due to certain large weight events in the QCD simulation samples.

previously selected to form an event category, until the remaining events have negligible contribution to the total sensitivity.

The specific event categories obtained with this optimization procedure are mutually exclusive and summarized as follows:

- Category 1: Jet 2 $T_{\mathrm{Xbb}} > 0.980$ and BDT > 0.43
- Category 2: Jet 2 $T_{\mathrm{Xbb}} > 0.980$ and $0.11 <$ BDT ≤ 0.43, or Jet 2 $T_{\mathrm{Xbb}} < 0.980$ and BDT> 0.43 (excluding category 1)
- Category 3: Jet 2 $T_{\mathrm{Xbb}} \geq 0.95$ and BDT > 0.03 (excluding category 1 and 2).

The event categories are visually represented in Figure 6.14. Event category 1 is the most sensitive, with a signal-to-background ratio of about 1:6, assuming SM $ggHH$ cross section. Event category 2 is the second most sensitive, with a S/B of 1:33 and category 3 has a S/B ratio of about 1:200. The fail region in Fig. 6.14 is used

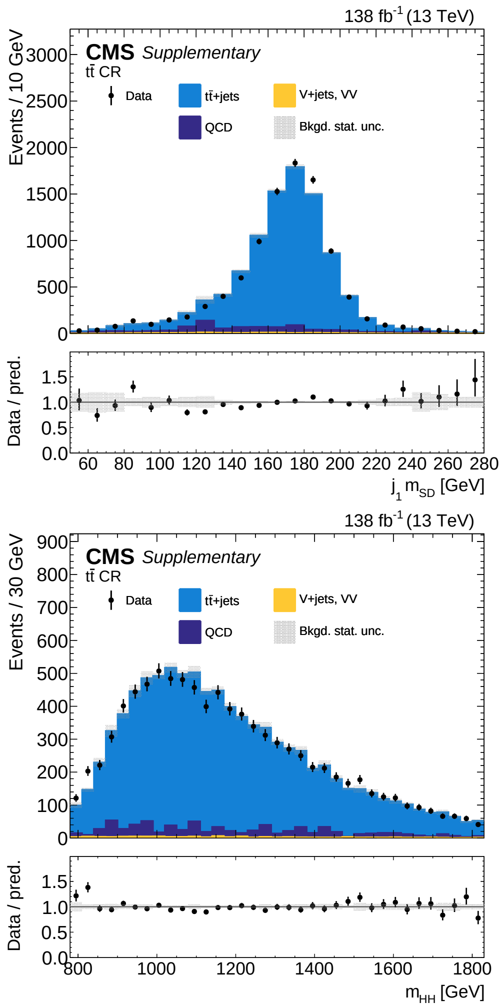


Figure 6.13: The data and expected background distributions from simulation in the $t\bar{t}$ control region (Sec. 6.4.2.1) for some of the most discriminating BDT input variables, including the Jet 1 m_{SD} (upper left) and T_{Xbb} (upper right), m_{HH} (lower left), and $p_{\mathrm{T\ j_1}}/m_{HH}$ (lower right). The lower panel shows the ratio of the data and the total background prediction, with its statistical uncertainty represented by the shaded band. The error bars on the data points represent the statistical uncertainties.

as a control region for QCD background estimation, and will be discussed later in Sec. 6.4.2.2.

6.4.2 Background estimation

The dominant background for this analysis is QCD multijet production and top quark pair production. They account for more than 90% of the total background and roughly contribute equally in the signal region categories. Other background processes include single Higgs production, diboson production and V+Jets production.

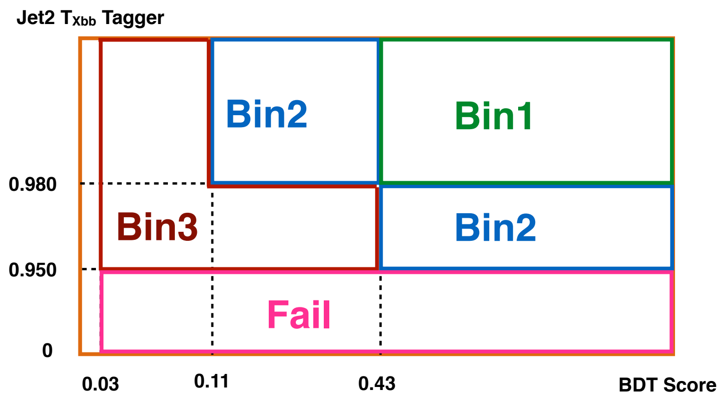


Figure 6.14: The event categories in the $ggHH$ analysis are visually represented in a 2D grid of the BDT score and the jet 2 T_{Xbb} score.

The QCD multijet background will be estimated from the data in control regions using the parametric alphabet fit method [222], and is described in Sec. 6.4.2.2. The QCD background prediction from this method also predicts the contribution from gluon fusion and VBF single Higgs boson production (see Sec. 6.4.2.2).

The top quark background is predicted using simulation, with several data-driven corrections derived from dedicated control regions and is described in Sec. 6.4.2.1.

Finally, the remaining backgrounds like $t\bar{t}H$, VH, VV, and V+jets are predicted using the simulation.

6.4.2.1 $t\bar{t}$ backgrounds

In this subsection, we describe all corrections applied to the simulation prediction of the top quark backgrounds and their associated uncertainties.

T_{Xbb} shape correction

The distribution of the T_{Xbb} score for the $t\bar{t}$ background may not be modeled perfectly in the simulation. To correct for this discrepancy, we measured its shape in a control region dominated by semi-leptonic $t\bar{t}$ events. The events in this control region were selected using online single muon or electron triggers. Additionally, events were required to satisfy the following conditions:

- At least one selected electron (Sec. 6.3.2) or muon (Sec. 6.3.1)
- At least one AK8 jet with $p_{\mathrm{T}} > 300\,\mathrm{GeV}$, $\tau_{32} < 0.46$, $m_{\mathrm{SD}} > 140\,\mathrm{GeV}$
- ΔR (lepton, AK8 jet) > 1.0

- $p_{\mathrm{T}}^{\mathrm{miss}} > 100\,\mathrm{GeV}$.

The requirements on jet τ_{32}, jet SD mass (m_{SD}), and $p_{\mathrm{T}}^{\mathrm{miss}}$ are applied in order to suppress the QCD multijet and W+jets events. The $t\bar{t}$ events make up for 90% of this control region, as can be seen in Figure 6.15.

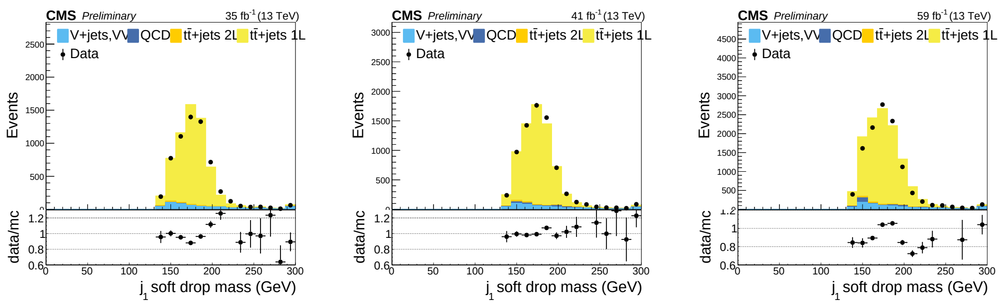


Figure 6.15: The distribution of the AK8 jet SD mass in the semi-leptonic $t\bar{t}$ control region after imposing all requirements. The 2016, 2017, and 2018 datasets are shown in the left, center, and right plots, respectively.

In Figure 6.16, we show the distribution of the ParticleNet T_{Xbb} tagger. The first bin shown here includes an in-situ normalization correction (between 0.0 and 0.8) to offset any discrepancies that might occur due to mismodeling of the lepton trigger/identification/isolation efficiencies or τ_{32} top tagging efficiencies. Figure 6.17 is a zoomed in version of this plot showing the most important bins. The remaining bins (between 0.8 and 1.0) are used to derive the correction to T_{Xbb}. We subtract the simulation prediction from the data for all the non top-quark backgrounds and compare the shape of the residual data to the top quark simulation prediction. We then derive correction factors as residual data to $t\bar{t}$ simulation ratios in different T_{Xbb} bins. The size of these corrections in different T_{Xbb} bins varies between a few percent to ~20%, with uncertainties below 2%.

The impact of these corrections on the distribution of the Jet 2 T_{Xbb} score and the event BDT score for the $t\bar{t}$ background, for the different event categories in the signal region, is shown in Fig. 6.18 and 6.19. These corrections tend to shift the top background towards lower T_{Xbb} and BDT scores, which is desirable.

$t\bar{t}$ recoil correction

The recoil of the $t\bar{t}$ system is mis-modeled by the simulation in the phase space relevant for this analysis. Therefore, we will measure the recoil in-situ using an

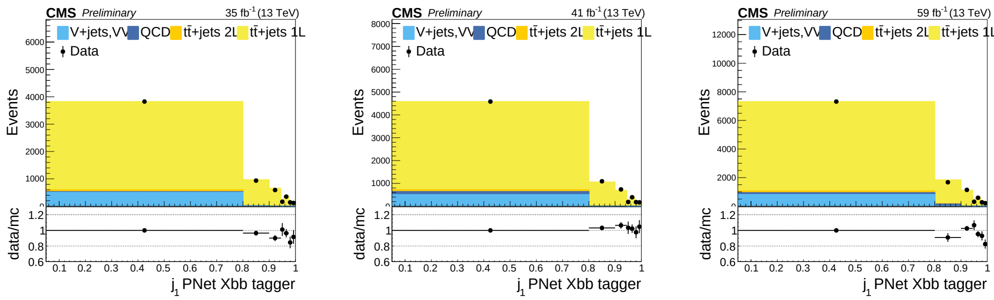


Figure 6.16: The distribution of the jet T_{Xbb} score in the 1ℓ+1j $t\bar{t}$ control region for 2016 (left), 2017 (middle), and 2018 (right); the first bin includes underflow events in each plot.

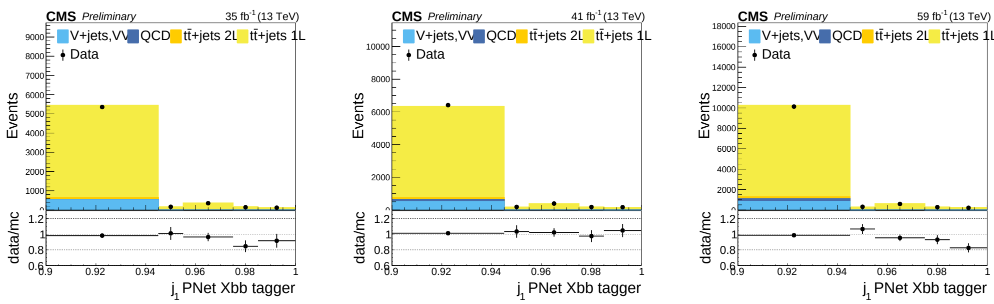


Figure 6.17: The distribution of the jet T_{Xbb} score in the 1ℓ+1j $t\bar{t}$ control region for 2016 (left), 2017 (middle), and 2018 (right) showing the most important bins.

all-hadronic $t\bar{t}$ control region. For this control region, we will require events to have two AK8 jets with each jet satisfying the following conditions:

- Jet $p_{\mathrm{T}} > 450\,\mathrm{GeV}$,
- Jet $m_{\mathrm{SD}} > 50\,\mathrm{GeV}$,
- Jet $T_{\mathrm{Xbb}} > 0.1$ (looser than signal region to increase statistics),
- Jet $\tau_{32} < 0.46$
- Jet contains an AK4 sub-jet that passes the loose WP of the DeepCSV b-tagging algorithm (described in Sec. 5.3.6).

These selections result in a pool of $t\bar{t}$ events with a purity of about 90%. Figure 6.20 shows the p_{T} of the $t\bar{t}$ system in this control region.

We subtract all non-$t\bar{t}$ backgrounds from the data, and then divide the residual data by the $t\bar{t}$ simulation prediction to obtain ratios in bins of $p_{\mathrm{T}}^{\mathrm{jj}}$. These ratios linearly

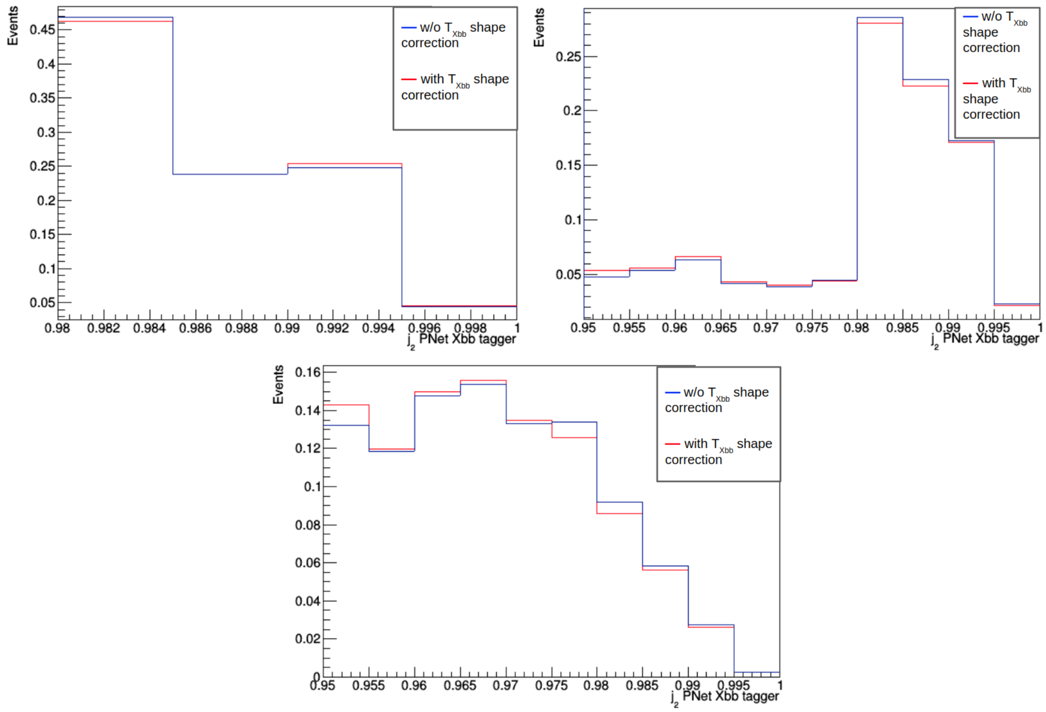


Figure 6.18: The distribution of the second jet T_{Xbb} score in the signal region for the $t\bar{t}$ background, for event category 1 (top left), category 2 (top right), and category 3 (bottom). The histograms are normalized to unit area.

increase until about 300 GeV, and then decrease linearly above 300 GeV. We fit these ratios with a “ramp-model“ defined as follows:

- Fit a linear function with a positive slope $p1$ and constant p_0, for $p_{\mathrm{T}}^{\mathrm{jj}} \in [0,$ $300]$ GeV as: $p_0 + p_1 \times p_{\mathrm{T}}$;
- Fit another linear function with a negative slope $p3$ and constant p_2, for $p_{\mathrm{T}}^{\mathrm{jj}} \in$ $[300, 1000]$ GeV as: $p_2 + p_3 \times p_{\mathrm{T}}$;
- Assume a flat correction for $p_{\mathrm{T}}^{\mathrm{jj}} > 1000$ GeV : $p_2 + p_3 \times 1000$.

The results from this ramp-model fit can be found in Figure 6.21. The ramp-model is used to derive correction factors for top quark simulation in various bins of $p_{\mathrm{T}}^{\mathrm{jj}}$. The fit uncertainties (shown with the dotted blue line) are propagated as systematic uncertainties for this correction.

Figure 6.22 shows the distribution of the p_{T} of the $t\bar{t}$ system after applying the “ramp-model” recoil correction and we observe a good level of agreement between the simulation prediction and data. Furthermore, we also check the modeling

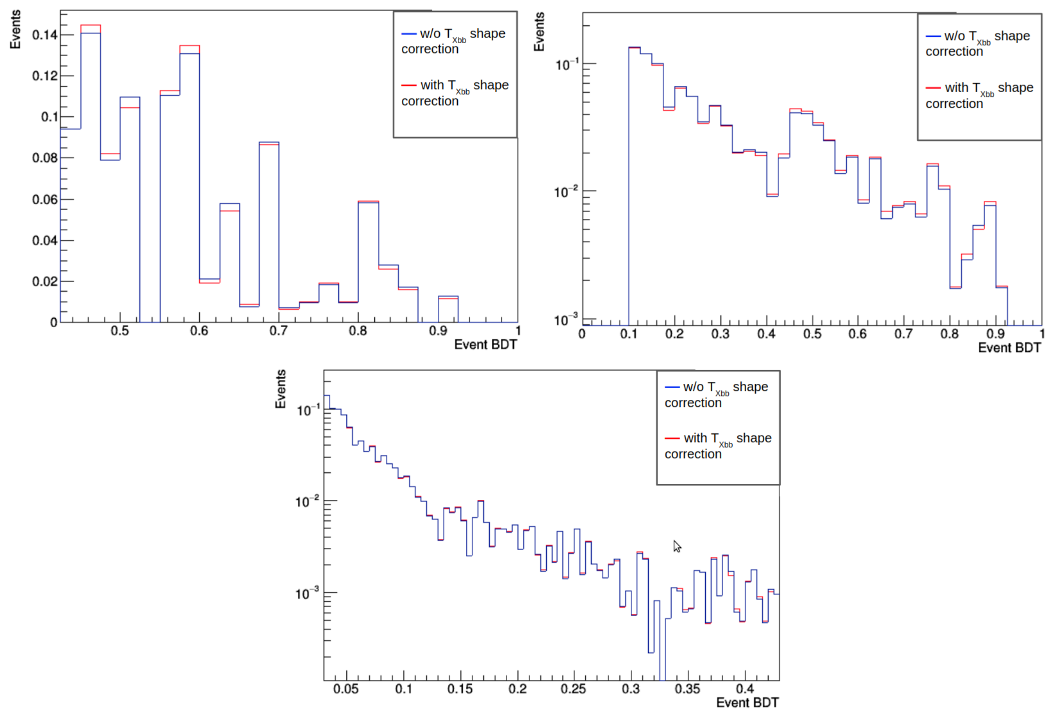


Figure 6.19: The distribution of the event BDT score in the signal region for the $t\bar{t}$ background, for event category 1 (top left), category 2 (top right), and category 3 (bottom). The histograms are normalized to unit area.

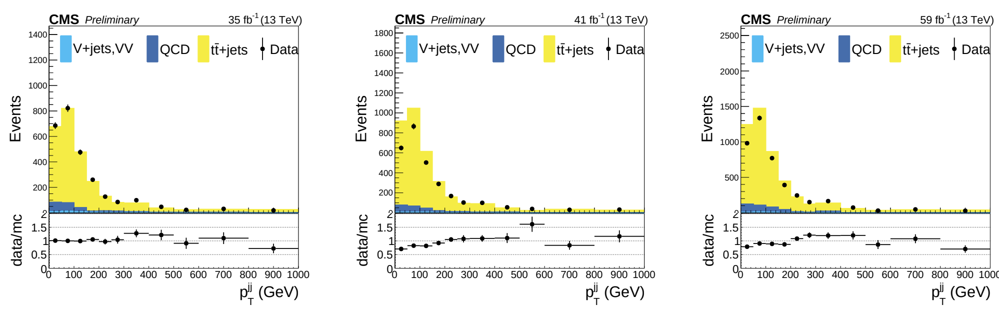


Figure 6.20: The distribution of the p_T of the $t\bar{t}$ system is shown for events in the all hadronic top control region, for 2016 (left), 2017 (middle), and 2018 (right) datasets.

of the p_T of the two AK8 jets individually and the mass of the dijet system in Figures 6.23, 6.24, and 6.25. All distributions show a good level of agreement between the simulation prediction and data.

For the ramp-model correction, we assumed that our QCD simulation is modelled well enough in the hadronic $t\bar{t}$ control region. To support this statement, we performed a check. When looking at the 80–120 GeV region in Jet 1 m_{SD} (Figure 6.26), the relative contribution of QCD and $t\bar{t}$ is similar and the data matches the MC

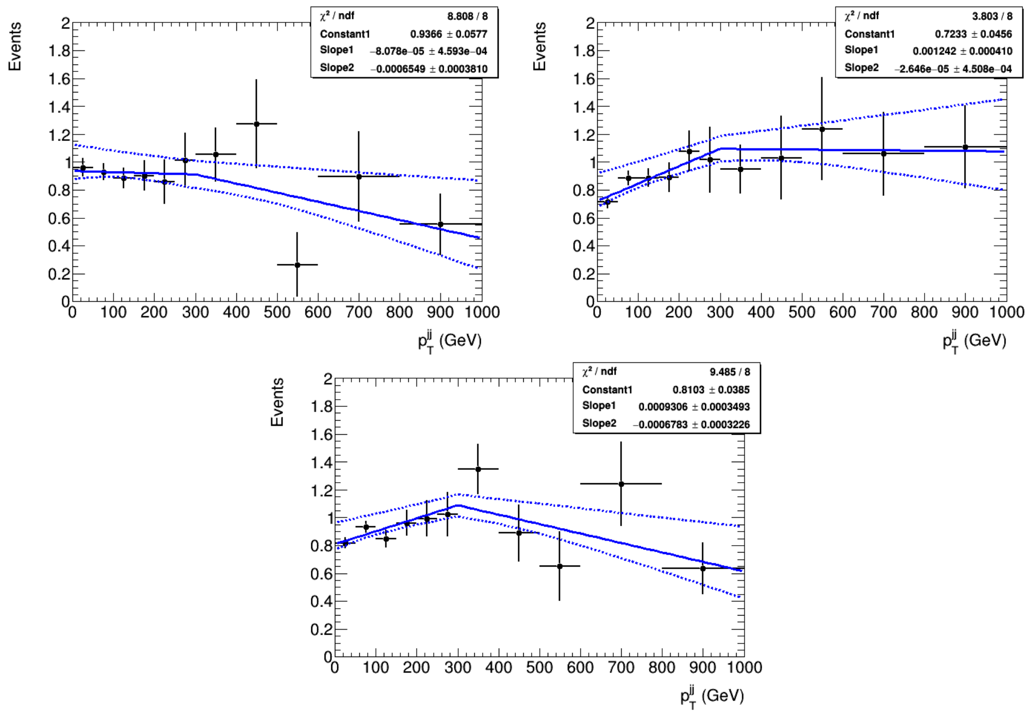


Figure 6.21: This figure shows the results of the "ramp-model" fit (solid blue line) in the hadronic $t\bar{t}$ control region, for 2016 (top left), 2017 (top right), and 2018 (bottom) datasets. The black points here represent the residual data (data subtracted with non-top backgrounds) divided by the $t\bar{t}$ simulation in different bins of $p_{\mathrm{T}}^{\mathrm{jj}}$. The fit uncertainties (shown with the dotted blue line) are propagated as systematic uncertainties for this correction in the final signal extraction fit.

prediction relatively well. Thus, this region shows that an arbitrary increase or decrease (let us say by a factor of 2) to the QCD background yield from simulation would most likely be incompatible with the observed data.

Nevertheless, we tried changing the QCD background yield per bin by a factor of 2 and 0.5 and then re-derived the ramp-model corrections. We observed that the impact of varying the QCD background in this manner is already covered by the total uncertainty on the $t\bar{t}$ recoil correction considered in the nominal fit. Considering these two pieces of evidence together, we can rule that the impact of the QCD background is negligible for this correction.

Top jet mass scale

We would also like to ensure that the top mass scale as modelled by the regressed mass for top jets is well simulated in the hadronic $t\bar{t}$ control region, defined in Section 6.4.2.1. The regressed mass distribution is shown in Figure 6.27 for different bins of the BDT score. The mass distributions are then fitted to a Gaussian, and the

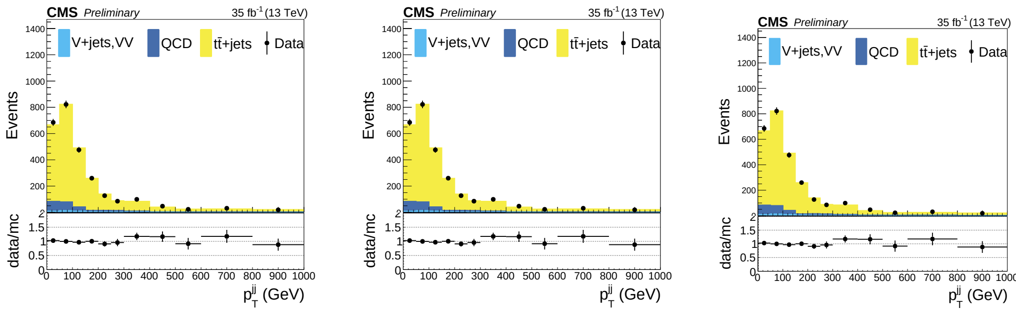


Figure 6.22: The distribution of the p_T of the $t\bar{t}$ system is shown for events in the hadronic $t\bar{t}$ control region after applying the ramp-model recoil correction, for 2016 (left), 2017 (middle), and 2018 (right) datasets separately.

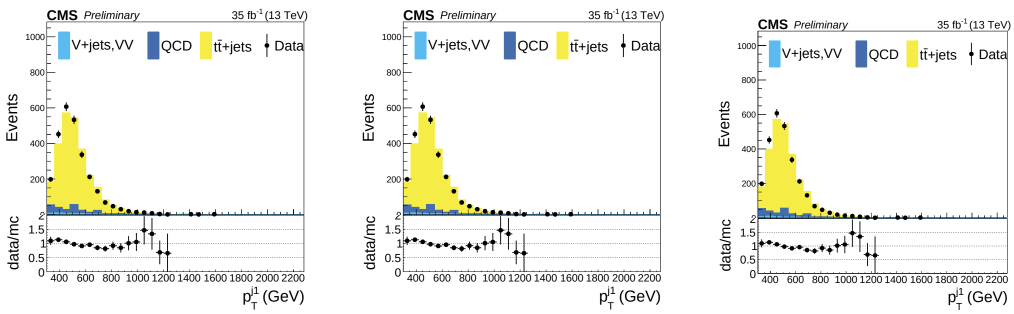


Figure 6.23: The distribution of the p_T of the AK8 Jet 1 (jet with highest T_{Xbb} score) is shown for events in the hadronic $t\bar{t}$ control region after applying the ramp-model recoil correction, for 2016 (left), 2017 (middle), and 2018 (right) datasets separately.

location of the fitted peak is plotted as a function of the event BDT and shown in Figure 6.28. We observe that the simulation agrees with the data, and predicts the mass scale very well in all BDT bins.

Modeling of second jet regressed mass in the $t\bar{t}$ control region

As mentioned previously, the variable used for the final signal extraction is the regressed mass of the second jet. We would like to ensure that this variable is well modelled by the simulation prediction in the hadronic $t\bar{t}$ control region. We apply the following corrections and check the distribution of the Jet 2 regressed mass:

- corrections for the T_{Xbb} score shape,
- corrections for the recoil of the $t\bar{t}$ system, and

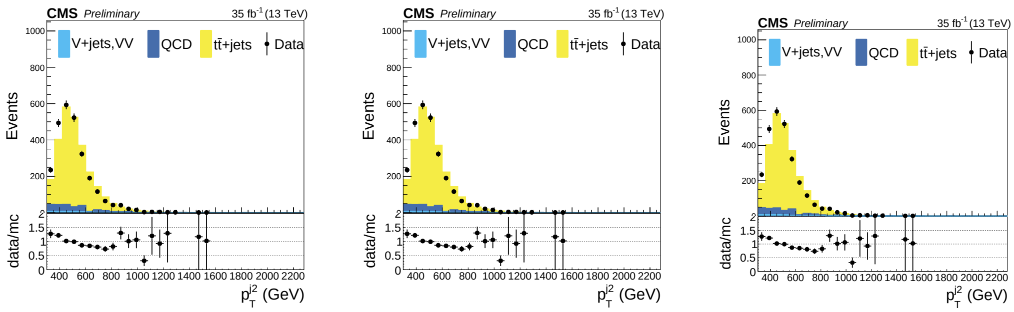


Figure 6.24: The distribution of the p_{T} of the AK8 Jet 2 (jet with second highest T_{Xbb} score) is shown for events in the hadronic $t\bar{t}$ control region after applying the ramp-model recoil correction, for 2016 (left), 2017 (middle), and 2018 (right) datasets separately.

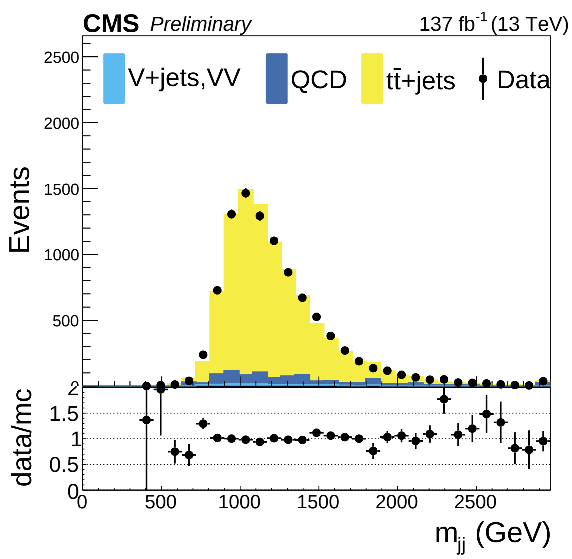


Figure 6.25: The distribution of the dijet mass is shown for events in the top control region using the full Run 2 dataset corresponding to 137 fb^{-1}.

- corrections to the regressed mass scale and mass resolution, as as recommended centrally for all CMS data analyses [139].

We observe good agreement between the pre-fit predicted shape and the data, as shown on the left in Fig. 6.29. Furthermore, we verified that the residual disagreement observed is covered by the jet mass resolution uncertainty, by performing a fit in which we allowed that nuisance parameter to float within its constraints. The post-fit distribution of the jet 2 regressed mass is shown on the right of Figure 6.29, where we observe improved shape agreement with data.

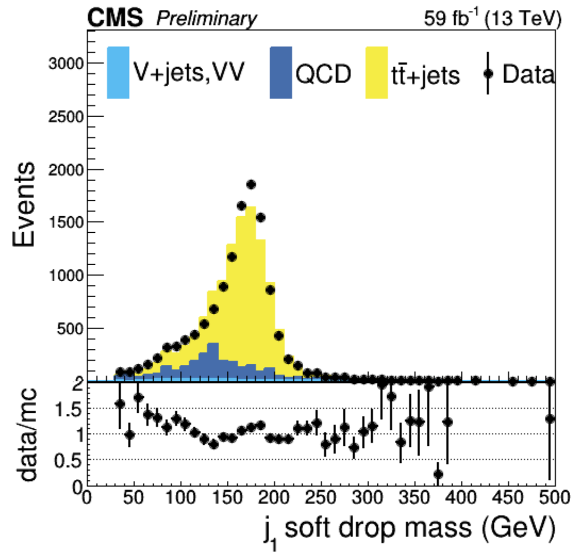


Figure 6.26: The Jet 1 m_{SD} in the $t\bar{t}$ hadronic control region. The 80–120 GeV region shows that a factor of two increase or decrease to the QCD background would be incompatible with the observed data.

BDT shape uncertainty

Lastly, we also check the event BDT score distribution in the $t\bar{t}$ all-hadronic control region. For this, we tighten the selection on the τ_{32} of both AK8 jets to $\tau_{32} < 0.39$, in order to achieve very high $t\bar{t}$ purity in all BDT bins. Figure 6.30 shows the event BDT shape in the control region, after applying the T_{Xbb} shape corrections and the $t\bar{t}$ recoil corrections derived in the previous sections. The left plot shows the BDT distribution in the $t\bar{t}$ all-hadronic control region with the nominal binning. The right plot shows the same BDT distribution, but with the last bin split into three bins, to confirm that variations in bin size do not alter the data to simulation agreement.

From the left plot of Figure 6.30, we subtract all non-$t\bar{t}$ background from the data and measure a small difference between the residual data and the $t\bar{t}$ simulation prediction. This small difference is propagated as a shape systematic uncertainty for the $t\bar{t}$ background prediction in the final signal extraction fit.

6.4.2.2 QCD

Since QCD is a relatively more difficult process to model accurately with simulation, we would like to estimate it using a data-driven approach from a signal-depleted region. For this purpose, we will use the Fail region, mentioned previously in Sec. 6.4.1.3, defined as the phase space of events where Jet 2 $T_{\mathrm{Xbb}} < 0.95$ and BDT

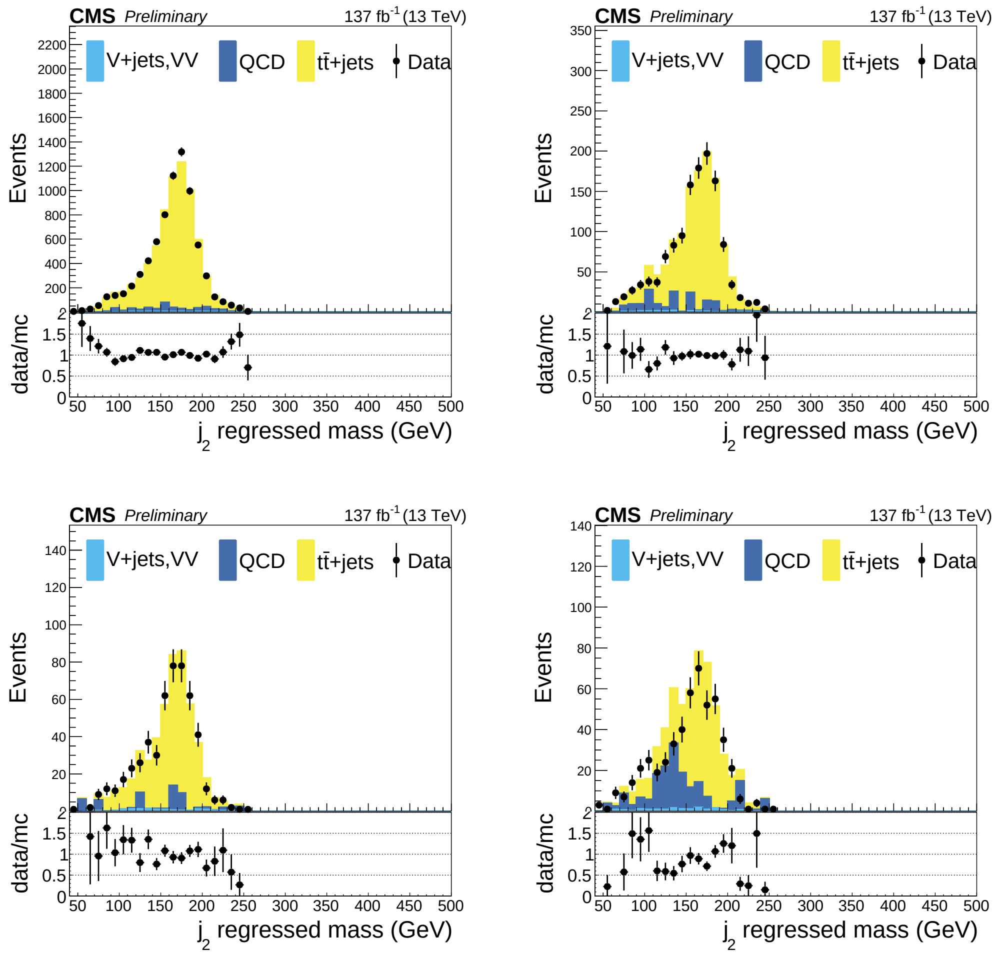


Figure 6.27: The top mass in different BDT bins (BDT range of top left: 0 - 0.00008; top right: 0.00008 - 0.0002; bottom left: 0.0002 - 0.0004; bottom right: 0.0004 - 1.0) for the full Run2 dataset. Here, the top recoil correction and the T_{Xbb} shape correction are already applied.

score > 0.03 (see Fig. 6.14). The fail region is enriched in backgrounds and can be used to extrapolate the shape of the QCD in the signal region as :

$$p^{i\ pass}_{\mathrm{QCD}}(\text{Jet 2 } m_{\mathrm{reg}}) = TF^{i}(\text{Jet 2 } m_{\mathrm{reg}}) * p^{fail}_{\mathrm{QCD}}(\text{Jet 2 } m_{\mathrm{reg}}) \tag{6.9}$$

where $p^{i\ pass}_{QCD}(\text{Jet 2 } m_{\mathrm{reg}})$ and $p^{fail}_{QCD}(\text{Jet 2 } m_{\mathrm{reg}})$ are the QCD regressed mass distributions of Jet 2 in the i-th pass region (signal event category 1, 2 and 3) and fail regions, respectively. $TF^{i}(\text{Jet 2 } m_{\mathrm{reg}})$ is a n-th order polynomial transfer factor for the i-th pass region, as a function of the Jet 2 m_{reg}. This is known as the parametric alphabet fit method. The gluon fusion and VBF single Higgs boson production background exhibits the same shape as the QCD multijet background, as the second jet in those events originate from initial state radiation (ISR), same as the second

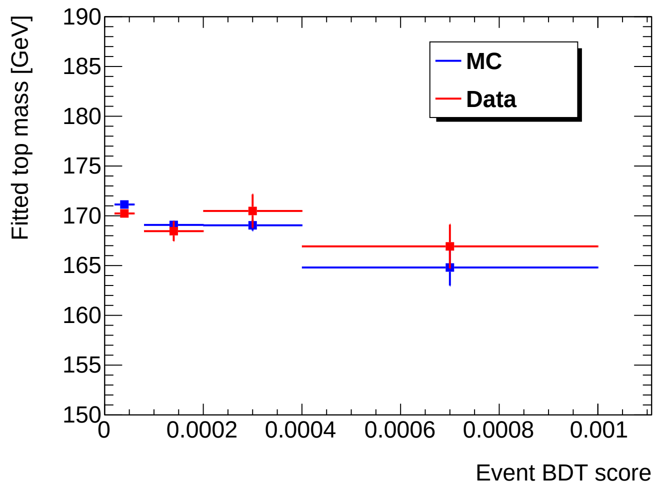


Figure 6.28: The fitted top mass (from a Gaussian fit to Figure 6.27) is shown as a function of the event BDT. Here, the top recoil correction and the T_{Xbb} shape corrections are already applied.

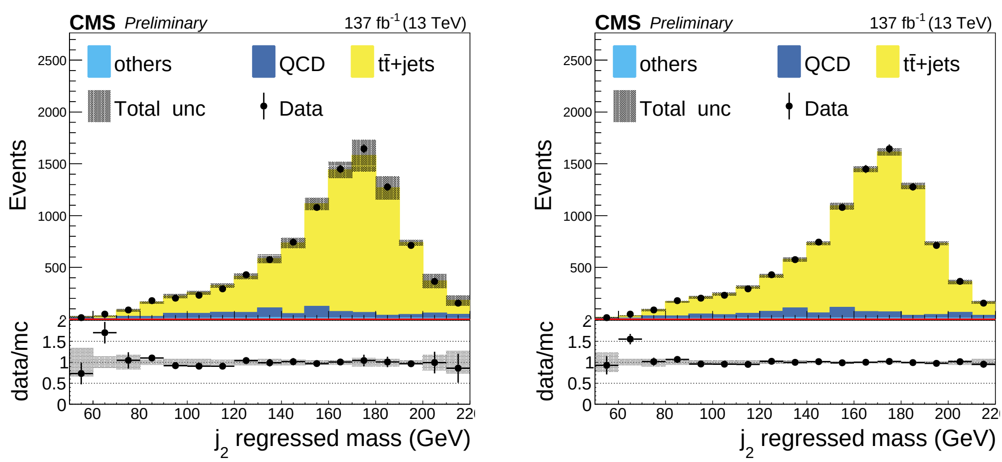


Figure 6.29: The distribution of the jet 2 regressed mass is shown for events in the top all-hadronic control region, for the full Run 2 dataset corresponding to 137 fb^{-1}. The pre-fit result is shown on the left, and on the right we show a post-fit result where the jet mass resolution uncertainty nuisance parameter was allowed to float within its constraints.

jet in QCD multijet production. Therefore, the QCD background prediction from this method also predicts the contribution from gluon fusion and VBF single Higgs Boson background.

The mass shape of the QCD multijet background in the fail region (p_{QCD}^{fail}(Jet 2 m_{reg})) is defined as the data subtracted with all non-QCD backgrounds as predicted by the simulation (the $t\bar{t}$ background prediction used here includes all the corrections

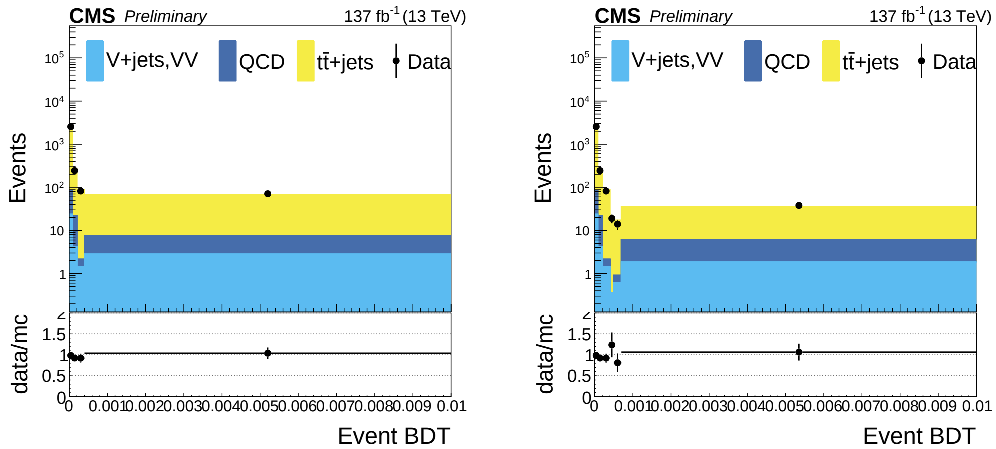


Figure 6.30: The event BDT distribution in the $t\bar{t}$ all-hadronic control region is shown for the data and simulation prediction, for the full Run2 dataset. The binning edges are [0.000-0.00008], [0.00008-0.0002], [0.0002-0.0004], [>0.0004] for the left plot, and [0.000-0.00008], [0.00008-0.0002], [0.0002-0.0004], [0.0004-0.0005], [0.0005-0.0007], [>0.0007] for the right plot. The left plot is used to derive uncertainties for the BDT shape modeling in the $t\bar{t}$ simulation.

described in Section 6.4.2.1). The Jet 2 regressed mass shape of the QCD multijet background in the different event categories ($p^{i\ pass}_{QCD}$(Jet 2 m_{reg})) is then estimated from a simultaneous likelihood fit to the pass and fail regions. The order of the transfer factor for each individual event category is determined by performing a F-test and a goodness-of-fit (GOF) test. We will first describe these tests and then explain the procedure to derive the polynomial order of the transfer factors.

F-Test and goodness-of-fit (GOF) Test

We perform the Fisher test (F-Test) to determine the minimum number of parameters that can describe the transfer factors well. For two fit models, 1 and 2, where model 1 with p_1 parameters is "nested" within model 2 with p_2 parameters (p2 > p1), F-test determines whether model 2 gives a significantly better fit to the data. For example, a polynomial function of zero-th order is nested within a 1st order polynomial. A random variable X will follow the F-distribution with parameters (d_1, d_2)

$$X = \frac{U_1/d_1}{U_2/d_2} \tag{6.10}$$

where U_1 and U_2 are independent variables that follow chi-squared distributions, with d_1 and d_2 degrees of freedom, respectively.

For data that is distributed like a Gaussian, we have the chi-squared distribution defined as

$$\chi^2 = \sum_j \frac{(f_j - d_j)^2}{\sigma_j^2} \tag{6.11}$$

where $d_i \pm \sigma_i$ is the i-th measured data point with rms deviation σ_i, and f_i is the model prediction. For the same data and model, the likelihood is given by:

$$\mathcal{L} = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(d_i - f_i)^2/2\sigma_i^2}. \tag{6.12}$$

The goodness-of-fit (GOF) test is a test of the null hypothesis in the absence of an alternate hypothesis. In a situation where we only have the data and a model to fit it with, one can always come up with an alternative model where $f_i = d_i$ at every measured value. Such a model is called a saturated model. In Eqn. 6.12, if we set $f_i = d_i$, the saturated model [79, 223, 224] becomes:

$$\mathcal{L}_{saturated} = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}}. \tag{6.13}$$

The ratio of the two likelihoods (Eqn. 6.12 and 6.13), λ, is then

$$\lambda = \prod_i e^{-(d_i - f_i)^2/2\sigma_i^2}, \tag{6.14}$$

and more importantly,

$$\chi^2 = -2ln\lambda. \tag{6.15}$$

Thus, the likelihood ratio asymptotically follows a chi-squared distribution and is a suitable candidate for a GOF test. In case of Poissonian distribution, one can write the GOF test statistic as

$$-2ln\lambda = 2\sum_j^{n_{bins}} (f_j - d_j + d_j \times ln(d_j/f_j)). \tag{6.16}$$

Eqn. 6.16 asymptotically follows a chi-squared distribution with $n_{bins} - p_i$ degrees of freedom, where p_i is the polynomial order. Therefore, we can construct the F-statistic (from Eqn. 6.10) as,

$$F = \frac{\frac{-2\log\lambda_1 + 2\log\lambda_2}{p_2 - p_1}}{\frac{-2\log\lambda_2}{n_{\text{bins}} - p_2}}. \tag{6.17}$$

For our case, n_{bins} is the number of Jet 2 regressed mass bins. Under the null hypothesis that model 2 does not provide a significantly better fit than model 1, Eqn. 6.17 will have an F-distribution with $(p_2 - p_1, p_2 - n)$ degrees of freedom. We may reject the null hypothesis if the value of Eqn. 6.17, calculated from data is greater than the critical value of the F-distribution for the desired false rejection probability of $\alpha = 0.05$.

The polynomial order of the transfer factors

The procedure for deriving the polynomial order of the transfer factors using an F-test and GOF test is as follows:

1. Start with the lowest n-th order polynomial (p_1) that can describe the data (usually start with zero-th order). We will call this a "poly-n" model.

2. Fit p_1 to the data and generate pseudo-data from the resulting fit. Calculate GOF test statistic (Eqn. 6.16) for p_1.

3. Define p_2 to be a polynomial of order $n + 1$ (one degree higher than p_1). Fit the data and the pseudo-datasets using the p_2 model.

4. Compute the F test statistic (Eqn. 6.17) for the toy datasets and the data.

5. If the F value for the data falls on the F-distribution of the pseudo-dataset with a p-value > 0.05, and the GOF test for the p_1 model passes with p-value>0.05, then we accept model p_1 as the best model for the QCD background.

6. If model p_1 is not selected, then go back to step (1), using the model p_2 as the initial lower order polynomial model, and repeat all previous steps again. Iterate until the stopping condition is satisfied.

Figures 6.31, 6.32, and 6.33 show the F-test and GOF test procedure for all three event categories.

To illustrate how the polynomial order is selected, we will discuss the second event category, whose F-test and GOF test plots are shown in Figure 6.32. The selection proceeds as follows:

- The F-value observed from the data (shown with a blue arrow) falls on the F-distribution comparing poly-0 vs poly-1, with a p-value = 0.12, which passes

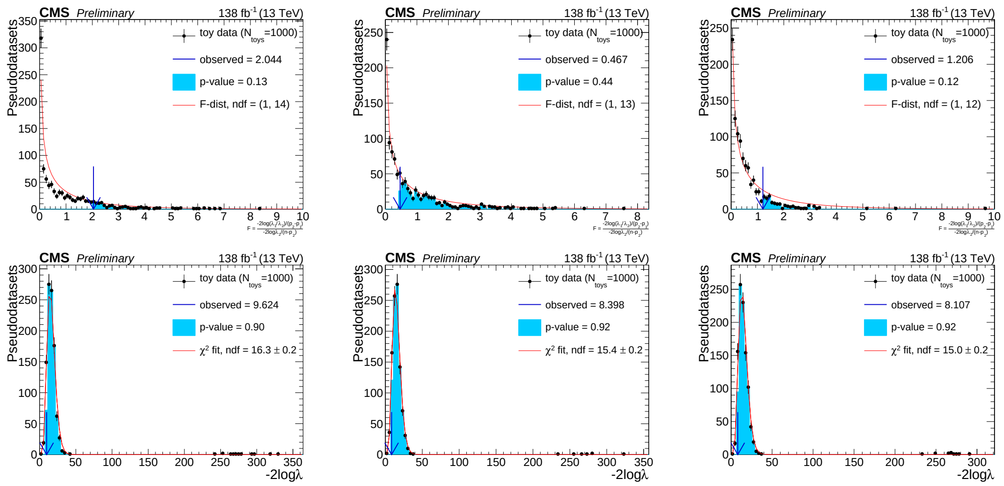


Figure 6.31: Deriving transfer factors for QCD parametric alphabet fit. (Top) F-distribution of poly-0 vs poly-1 (left), poly-1 vs poly-2 (middle), poly-2 vs poly-3 (right); (Bottom) GOF test of poly-0 (left), poly-1 (middle), poly-2 (right). These plots are for event category 1. Poly-0 is the selected model from the F-test and GOF-tests as it passes both tests.

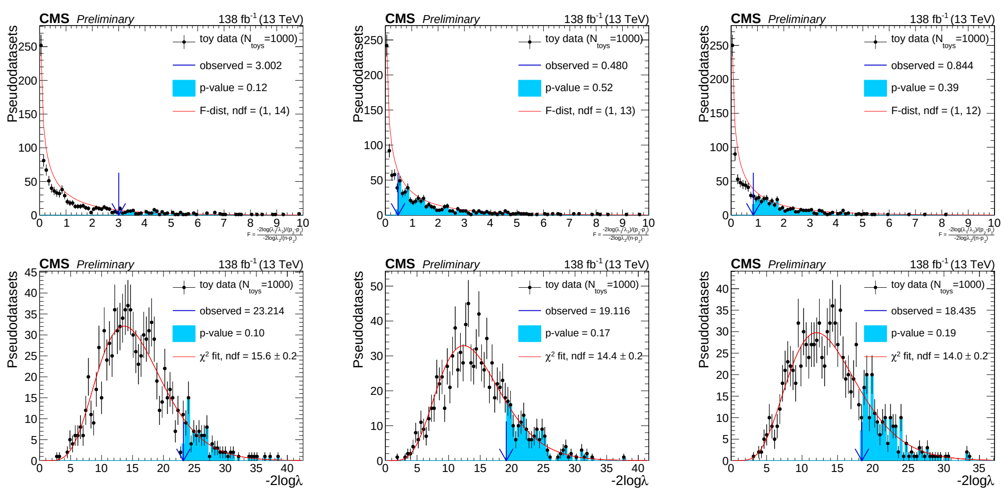


Figure 6.32: Deriving transfer factors for QCD parametric alphabet fit. (Top) F-distribution of poly-0 vs poly-1 (left), poly-1 vs poly-2 (middle), poly-2 vs poly-3 (right); (Bottom) GOF test of poly-0 (left), poly-1 (middle), poly-2 (right). These plots are for event category 2 . Poly-0 is the selected model from the F-test and GOF-tests as it passes both tests.

the p-value > 0.05 requirement. The GOF test on poly-0 also gives a p-value (= 0.10) which is larger than 0.05. So poly-0 can fit the data well and we can stop the iterative procedure.

- We conclude poly-0 is the best transfer factor for event category 2.

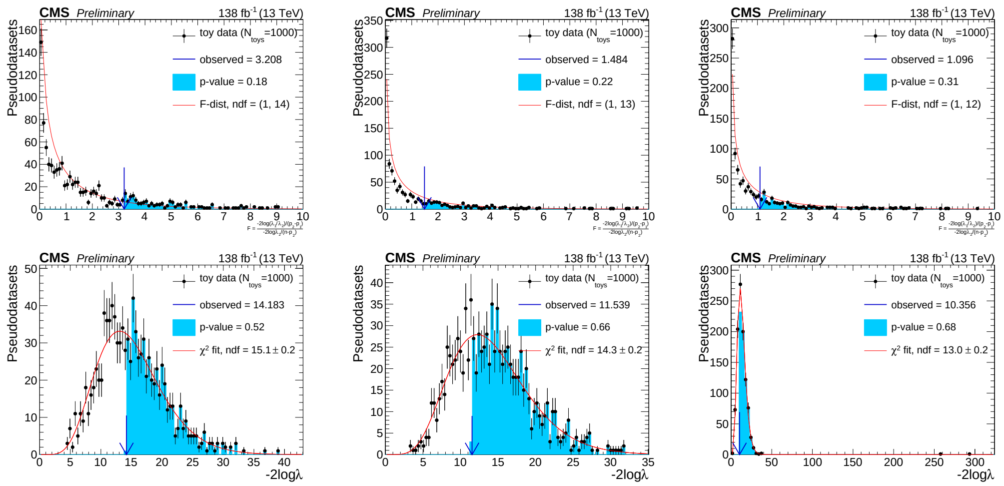


Figure 6.33: Deriving transfer factors for QCD parametric alphabet fit. (Top) F-distribution of poly-0 vs poly-1 (left), poly-1 vs poly-2 (middle), poly-2 vs poly-3 (right); (Bottom) GOF test of poly-0 (left), poly-1 (middle), poly-2 (right). These plots are for event category 3. Poly-0 is the selected model from the F-test and GOF-tests as it passes both tests.

Following the same method, we determined that the best transfer factor model for event categories 1, 2, and 3 are poly-0, poly-0, and poly-0, respectively. This implies that each transfer factor is just a number with an uncertainty determined by the fit, and is the same in all bins of the Jet 2 regressed mass, per event category. The transfer factors determined by the background-only fit to the data in the Jet2 regressed mass are presented in Table 6.4.

Table 6.4: Summary of transfer factors obtained from background-only fit.

event category	transfer factor
Event category 1	$0.60^{+0.06}_{-0.13}$
Event category 2	$1.08^{+0.06}_{-0.06}$
Event category 3	$1.18^{+0.04}_{-0.04}$

6.4.3 Corrections to data and simulation

We have previously described corrections to the jet energies used in this analysis in Sec. 6.3.4 and the corrections to the top background are mentioned in Sec. 6.4.2.1. We apply some additional corrections to the data and simulation in this analysis, which are described in the next few sub-sections.

6.4.3.1 Trigger efficiencies

We measure the trigger efficiency by selecting events using online single muon triggers that impose a loose isolation requirement on the muon candidate and require the muon p_{T} to be above a certain threshold. These are our reference triggers.

In this muon dataset, we measure the efficiency for the events that additionally contain an AK8 jet (defined in section 6.3.4). The offline muon candidate is required to have $p_{\mathrm{T}} > 30$ and $|\eta| < 2.4$. The AK8 jet is required to be well separated from the muon with ΔR(muon, AK8 jet)> 1.5. These selections are known as our reference selections.

The trigger efficiency is measured in events that additionally pass the main analysis triggers (as described in 6.2.1), as a function of three variables: the AK8 jet T_{Xbb} tagger score, p_{T}, and soft drop mass. The total trigger efficiency for each simulated event is calculated assuming that each of the two selected AK8 jets in an event can independently pass the trigger requirements. However, this assumption ignores potential correlations between the trigger efficiency for the two jets. The efficiency per jet, ε_j, and efficiency per event, $\varepsilon_{\mathrm{trig}}$, is defined as:

$$\varepsilon_j = \frac{N(\text{pass analysis signal trigger AND reference selection})}{N(\text{pass reference selection})}(p_{\mathrm{T}}, \text{mass}, T_{\mathrm{Xbb}}) \tag{6.18}$$

$$\varepsilon_{\mathrm{trig}} = 1.0 - (1.0 - \varepsilon_{\mathrm{Jet\ 1}})(1.0 - \varepsilon_{\mathrm{Jet\ 2}}). \tag{6.19}$$

The measured trigger efficiencies are shown in Figures 6.34, 6.35, and 6.36, for the 2016, 2017, and 2018 datasets, respectively. The trigger efficiency increases steadily for $p_{\mathrm{T}} \in [300, 550]$ GeV, and is fully efficient for p_{T} above 550 GeV. This happens because most of our signal triggers have a high H_T or jet p_{T} threshold. We note here that future improvements to the trigger efficiency at smaller values of the jet p_{T}, particularly between 300 GeV and 600 GeV, could yield a substantial improvement to our measured $ggHH$ signal yield.

The statistical uncertainties of the trigger efficiency measurement are below 5% for most bins and can be up to 10–15% for the least populated bins. These uncertainties are propagated to the signal yield predictions as detailed in Sec. 6.4.4.

To quantify the systematic error of our assumption of independent trigger efficiencies for the two jets, we repeat the trigger efficiency measurements using simulated semi-leptonic $t\bar{t}$ samples. We select events that pass the single muon reference triggers,

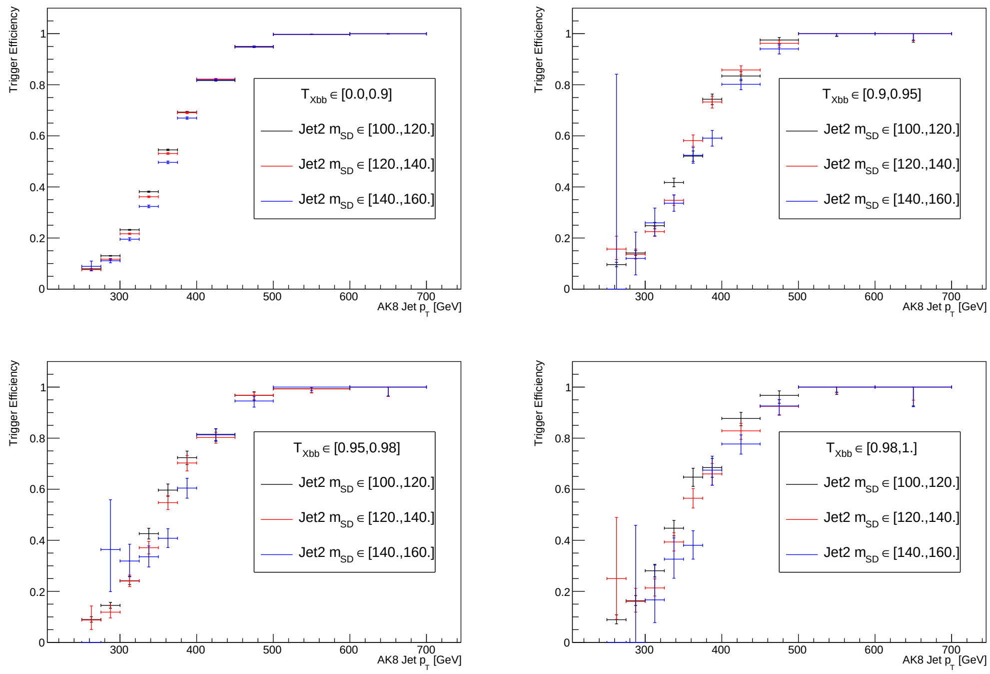


Figure 6.34: The trigger efficiency for the 2016 dataset is measured as a function of the AK8 jet p_{T} in four bins of the T_{Xbb} score: 0.0–0.90 (top left), 0.90–0.95 (top right), 0.95–0.98 (bottom left), 0.98–1.00 (bottom right). The different colors represent different ranges of jet m_{SD}.

contain at least one isolated muon, and an AK8 jet that is well separated from the muon. We then measure the efficiency that these events also pass the signal analysis triggers described in 6.2.1. To properly account for the fact that some of our analysis signal triggers were disabled for some fraction of the data-taking period, we calculate the effective luminosity of each individual trigger in data, and flip the trigger bit to false on a random basis to achieve the correct effective luminosity per trigger in the $t\bar{t}$ simulation samples. We then apply the efficiencies measured in the $t\bar{t}$ simulation sample on the $ggHH$ signal sample, under the same assumption that the trigger efficiencies for each of the two jets are independent, and calculate the expected signal yield for the signal region (described in 6.4.1.3). We compare this prediction for the signal yield with prediction obtained by requiring that the simulated triggers fired in the $ggHH$ sample. The difference between the two predictions is observed to be between $\sim 3-5\%$. The size of this difference is propagated as a systematic uncertainty in the final fit, as a function of the jet p_{T}.

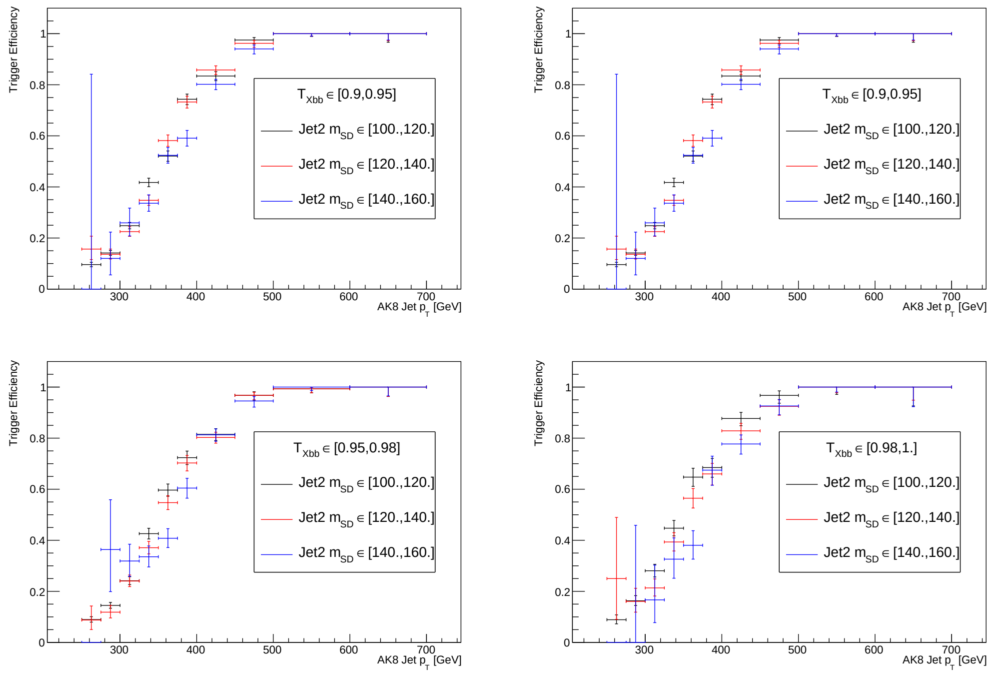


Figure 6.35: The trigger efficiency for the 2017 dataset is measured as a function of the AK8 jet p_{T} in four bins of the T_{Xbb} score: 0.0–0.90 (top left), 0.90–0.95 (top right), 0.95–0.98 (bottom left), 0.98–1.00 (bottom right). The different colors represent different ranges of jet m_{SD}.

6.4.3.2 ParticleNet Tagger Efficiency measurement for signal jets

The ParticleNet T_{Xbb} score is the key discriminator for this analysis, and gives us the main enhancement in signal sensitivity. The T_{Xbb} score for Jet 1 is also used in the analysis BDT, and therefore the T_{Xbb} score shape also has some impact on the shape of the BDT distribution. It is therefore important to accurately predict the T_{Xbb} shape distribution for the signal jets.

To do this, we tried measuring the T_{Xbb} shape in a control region of boosted $Z \to b\overline{b}$ jets. However, due to extremely large backgrounds (mainly QCD), it was very difficult to find a phase space with a high purity sample of $Z \to b\overline{b}$ AK8 jets. Instead, we use the method described in [225] that selects b jets from gluon splitting to a pair of b quarks. A high purity $g \to b\overline{b}$ sample can be selected from a QCD simulation with large statistics. The QCD AK8 jets are classified as b-type, c-type, and light-type jets based on the truth level particle information. The b-type jets are then chosen as the proxy jets to mimic the $H \to b\overline{b}$ jets. The proxy AK8 jets are also required to have at least one secondary vertex (SV) matched to each of the two sub-jets within the AK8 jet.

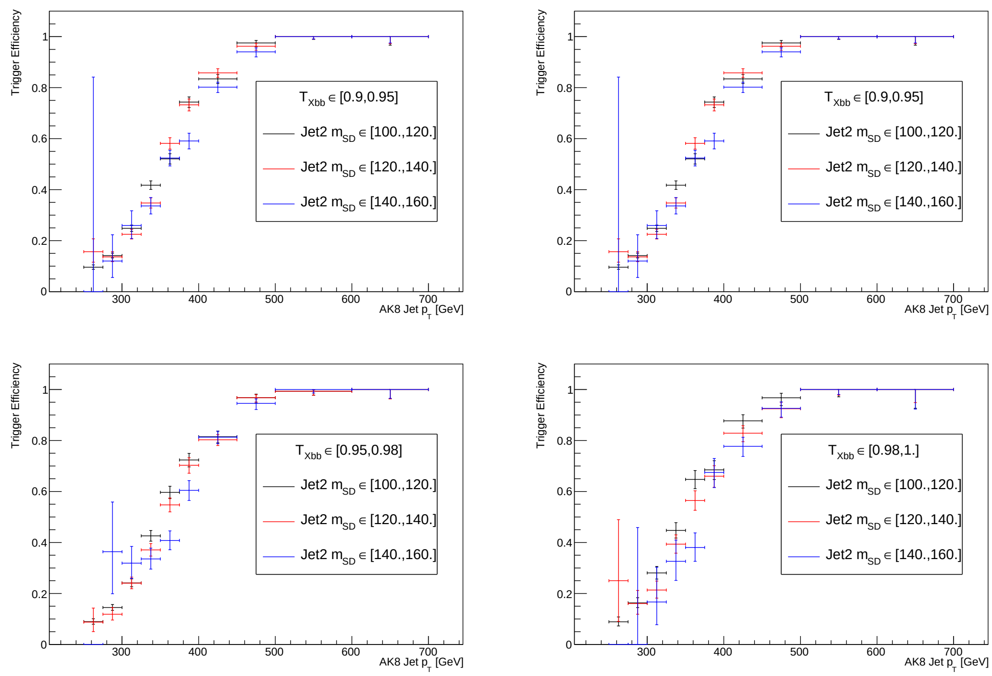


Figure 6.36: The trigger efficiency for the 2018 dataset is measured as a function of the AK8 jet p_T in four bins of the T_Xbb score: 0.0–0.90 (top left), 0.90–0.95 (top right), 0.95–0.98 (bottom left), 0.98–1.00 (bottom right). The different colors represent different ranges of jet m_SD.

A multivariate discriminant ("sfBDT") was developed to enhance the similarity between the proxy jets and the signal jets, and is designed to separate jets with a clean composition of quarks, which are more like $H \to b\bar{b}$ jets, against the ones with large contamination of extra gluons. The gluon contamination rate is defined by a variable κ_g, which is given by

$$\kappa_g = \frac{\sum p_\mathrm{T}(g)}{\sum p_\mathrm{T}(g,q)} \tag{6.20}$$

where $\sum p_\mathrm{T}(g)$ is the scalar p_T sum of all final state gluons and the $\sum p_\mathrm{T}(g,q)$ is the scalar p_T sum of all final-state gluons and quarks. The gluons and quarks are selected from the parton-level truth particles associated with a jet. The proxy signal jets are selected from the QCD $g \to b\bar{b}$ when they satisfy $\kappa_g < 0.15$ and proxy background jets satisfy $\kappa_g > 0.85$. The input variables to the sfBDT involve kinematics of the sub-jets, tracks and secondary vertices associated with the jet. These variables include :

- Jet τ_{21}: 2-prong N-subjettiness or τ_2/τ_1,

- $m(s_{j_1})$: Mass of sub-jet 1,
- $m(s_{j_2})$: Mass of sub-jet 2,
- N($SV_{1,2}$ tracks): Number of tracks associated with SV of either sub-jet,
- p_T (SV_1): p_T of SV (with the highest track impact parameter) for sub-jet 1,
- p_T (SV_2): p_T of SV (with the highest track impact parameter) for sub-jet 2.

We divide the proxy jets into four jet p_T bins: 250–300, 300–400, 400–500, and >500 GeV. In every p_T bin, we require a selection on the sfBDT discriminant such that the fraction of signal-like proxy jets with T_{Xbb} larger than 0.80 is the same as the fraction of actual $H \to b\bar{b}$ signal jets with T_{Xbb} larger than 0.80. This fraction varies between between 0.2 and 0.5. This drives the phase-space of $g \to b\bar{b}$ proxy signal jets to look more like the signal $H \to b\bar{b}$ jets. The actual selection on the sfBDT discriminant varied depending on the p_T bin and the data-taking year.

Finally, the logarithm of the secondary vertex mass (with the highest track impact parameter) is used as a fit variable to distinguish the b-type jets from the c- and light-type jets. This fit variable is used because the mass of the SV of a b-jet tends to peak at the B meson mass (~5 GeV) and peaks at the D meson mass for a c-jet, (~2 GeV). We divide the proxy jets into a 2D grid of (Jet T_{Xbb}, Jet p_T), with five T_{Xbb} bins: 0.8–0.9, 0.9–0.95, 0.95–0.98, 0.98-0.99, 0.99–1.0, and the four jet p_T bins mentioned previously. We measured the correction factors to the T_{Xbb} shape in these bins, for 2016, 2017 and 2018 data-taking years separately. In each of these 2D bins, we perform a *Tag and probe* simultaneous fit of simulation to data in the pass and fail (invert the T_{Xbb} score) regions. We consider three floating parameters in the fit corresponding to overall normalization of the *b*, *c* and *light* template, and vary them in the range [0.5,2] while keeping the total yield in the pass+fail region constant. The final "b" template fitted normalization parameter is used as the correction factor to the T_{Xbb} shape in that particular (Jet T_{Xbb}, Jet p_T) bin. An example fit plot in the pass and fail regions of the efficiency measurement is shown on the left and right of Figure 6.37, respectively. The resulting correction factors usually change the particular (Jet T_{Xbb}, Jet p_T) bin yield up or down by 10–20%.

The total uncertainty on the correction factor measurement is dominated by the systematic uncertainty on the similarity between the proxy $g \to b\bar{b}$ jets and the signal $H \to b\bar{b}$ jets. This systematic uncertainty is obtained by imposing looser

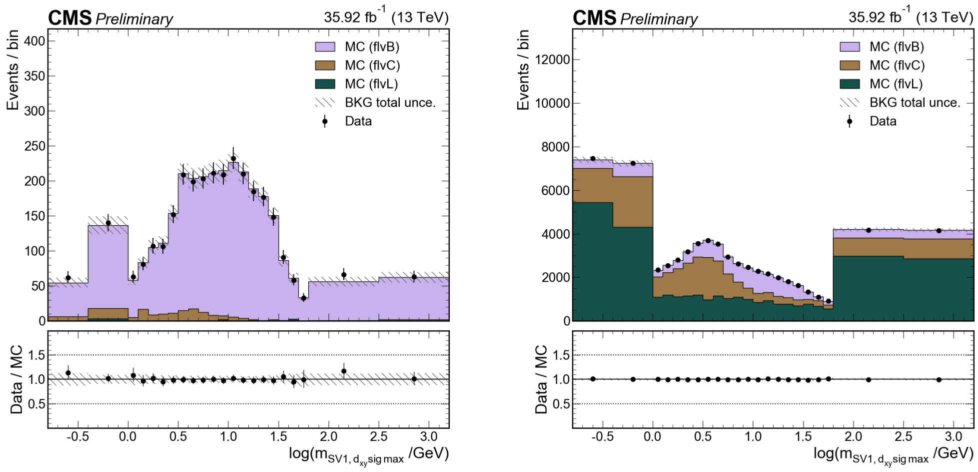


Figure 6.37: Example fit plots for the pass and fail regions of the T_{Xbb} efficiency measurement is shown here. This example corresponds to the bin with jet T_{Xbb} score between 0.99 and 1.00, jet p_{T} between 400 and 500 GeV and the 2016 data-taking period. The fail region here is defined by inverting the T_{Xbb} score, i.e., jet T_{Xbb} score < 0.99. The different colors represent different jet flavours. We used the fitted normalization of the "b" template (purple) as the final correction factor in the particular (Jet T_{Xbb}, Jet p_{T}) bin.

and tighter requirements on the sfBDT selection, which is used to obtain samples of proxy jets that place upper and lower bounds around the behavior of the signal jets, and this uncertainty ranges in 20 – 40%. Upon re-weighting the shape of the secondary vertex mass distribution, one measures a slightly different correction factor and this feature is also propagated as an uncertainty on the fit, and has a size of about 10–15%. The statistical uncertainty from the fit has a size of about 3–9%. All of the above uncertainties are propagated assuming no correlation across different bins. Finally, other systematic uncertainties include pileup uncertainties and the flavor composition uncertainty in the simulation, and contribute to a size of~ 4%, which is propagated as correlated across all bins.

The impact of these correction factors on the shape of the signal T_{Xbb} distribution is shown on the left of Figure 6.38, and their impact on the BDT discriminant shape is shown on the right of Figure 6.38. These corrections tend to shift the $ggHH$ signal to higher Jet 2 T_{Xbb} and BDT score bins, which is desirable.

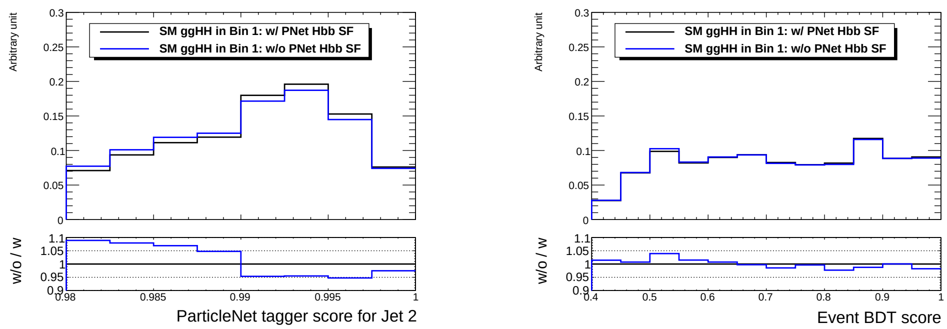


Figure 6.38: The impact of the measured T_{Xbb} correction factors on the ParticleNet tagger score for Jet 2 (left) and BDT discriminant (right) for the SM ggHH signal in event category 1. The histograms are normalized to unit area to show the change in shape.

6.4.3.3 Pileup re-weighting

This is done to match the observed pile-up interaction profile in data to what is predicted by the simulation, and was described previously in Sec. 5.4.1.

6.4.3.4 L1 EGamma pre-firing corrections

This corresponds to the same issue previously described in Sec. 5.4.2, and has been corrected for in this analysis too.

6.4.3.5 m_{HH} theory re-weighting

The $ggHH$ NLO signal simulation sample used in this analysis has large numerical instabilities for values of m_{HH} above 1 TeV. To correct for this, we re-weight the m_{HH} prediction in our samples with NLO Pade approximations for virtual corrections based on the high-energy expansion of the form factors. We take the full size of the correction as a systematic uncertainty, i.e., the up shape variation has no correction factors applied and the down shape variation has twice the correction, as presented in Figure 6.39 [226].

6.4.4 Systematic uncertainties

The systematic uncertainties considered for the $ggHH$ analysis are listed as follows:

- **Trigger uncertainty (normalization and shape)**: The uncertainty due to the triggering efficiency is applied and has been explained previously in Sec. 6.4.3.1.

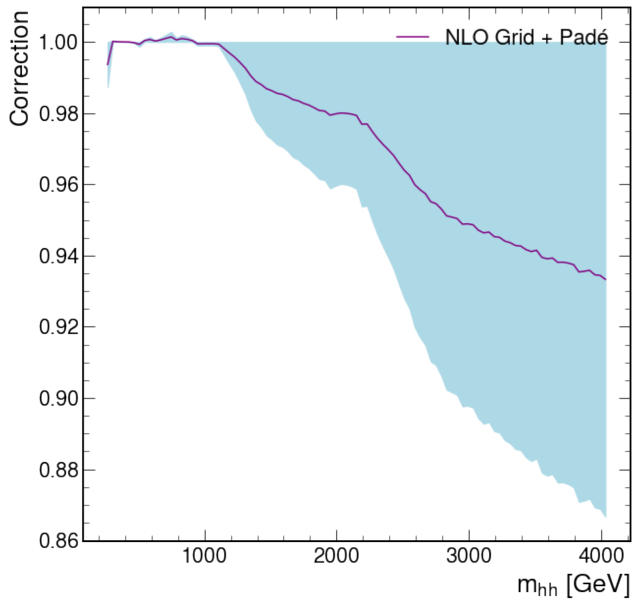


Figure 6.39: Correction factors to re-weight the current NLO calculation of m_{HH} in our signal simulation, to NLO calculations based on Pade approximations for virtual corrections, based on the high-energy expansion of the form factors [226].

- **Luminosity (normalization only)**: The luminosity uncertainty affects the signal and background normalization. The uncertainty on the luminosity measurement is about 2.5% in each era [153–155], and is modelled via a log-Normal nuisance parameter whose effect is correlated across all event categories, processes, and eras.

- **Uncertainty on the QCD multijet background (normalization and shape)**: This a dominant systematic uncertainty for this analysis. There are two components to this uncertainty: (1) bin-by-bin uncertainty in the Jet 2 regressed mass distribution, and (2) overall uncertainty on the transfer factor per event category.

- **Pileup uncertainty (shape):** An uncertainty on the number of primary vertices is derived by varying the minimum bias cross section used in the pileup re-weighting applied to simulation samples. The corresponding effect is correlated across event categories.

- **Jet energy scale (normalization and shape):** This uncertainty is obtained by varying the transverse momentum of each jet up and down by one standard deviation for each source of uncertainty, as recommended centrally for all CMS data analyses [139]. The full analysis chain is performed again, taking into account the shifted energies. The corresponding variations in each Jet 2 regressed mass bin yield for each process are used to estimate the uncertainty. This results in a set of uncertainties affecting both signal and background acceptance as well as the shape of the Jet 2 regressed mass.

- **Jet energy resolution (normalization and shape):** This uncertainty is obtained by smearing the transverse momentum of each jet by the smearing factors provided centrally for all CMS data analyses [139].
- **Jet mass scale (normalization and shape):** This uncertainty is obtained by varying the mass scale of each jet up and down by one standard deviation, as recommended centrally for all CMS data analyses [139].
- **Jet mass resolution (normalization and shape):** This uncertainty is obtained by smearing the Soft-Drop/Regressed mass resolution of each jet by the smearing factors provided centrally for all CMS data analyses [139].
- **T_{Xbb} shape uncertainties (normalization and shape)**: These uncertainties have been explained for the top background in Sec. 6.4.2.1 and for the signal jets in Sec. 6.4.3.2.
- **$t\bar{t}$ recoil uncertainties (normalization and shape)**: These uncertainties have been explained for the top background in Sec. 6.4.2.1.
- **BDT shape uncertainty for top background (shape)**: These uncertainties have been explained for the top background in Sec. 6.4.2.1.
- **Jet 2 regressed mass shape uncertainty for top background (shape)**: Due to the low $t\bar{t}$ simulation statistics in the event category 1, the nominal shape of the Jet 2 regressed mass in this category for the $t\bar{t}$ process is taken from a loosened requirement of $T_{\mathrm{Xbb}} > 0.9$ (instead of $T_{\mathrm{Xbb}} > 0.98$, which is event category 1 definition). This shape is normalized to the $t\bar{t}$ yield predicted in event category 1. This procedure is justified because the T_{Xbb} score and the Jet 2 Regressed mass are uncorrelated. The up and down variations of the nominal shape are taken from a modified selection of $T_{\mathrm{Xbb}} > 0.92$ and $T_{\mathrm{Xbb}} > 0.2$, respectively, as shown in Figure 6.40.
- **$ggHH$ and VBFHH Simplified Template Cross-Sections (STXS) (normalization and shape):** These uncertainties are evaluated following the recommendations of the LHC Higgs Cross-Section working group (LHCHXSWG) [158].

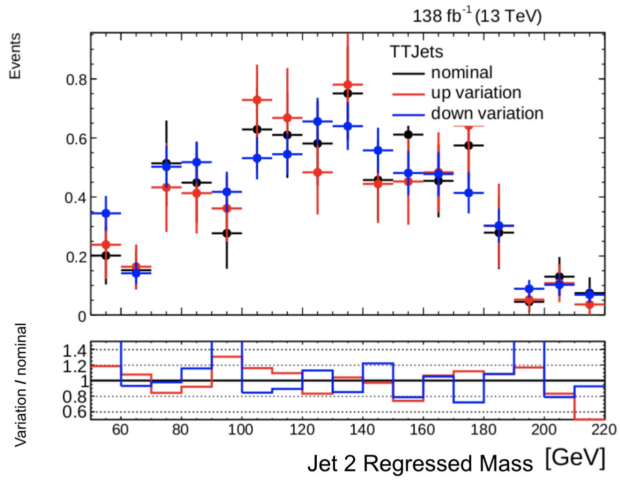


Figure 6.40: Uncertainty on the Jet 2 regressed mass shape in event category 1 for the $t\bar{t}$ process due to different T_{Xbb} score selections. The nominal shape is obtained after imposing a requirement of $T_{\mathrm{Xbb}} > 0.9$, and the up and down variations are obtained by requiring $T_{\mathrm{Xbb}} > 0.92$ and $T_{\mathrm{Xbb}} > 0.2$, respectively.

This recipe provides a set of independent sources of uncertainty, modelled via log-Normal nuisance parameters correlated across categories and eras.

- **Perturbative QCD scale variation (normalization and shape):** The perturbative QCD renormalization and factorization uncertainties are estimated by changing the scales μ_R and μ_F. We consider values of (μ_R, μ_F) = (0.5,0.5), (0.5,1.0), (1.0,0.5), (1.0,1.0), (1.0,2.0), (2.0,1.0), (2.0,2.0)), and obtain an envelope on the shape variation, which is treated as the uncertainty. The different (μ_R, μ_F) variations are shown for the $ggHH$ signal in Figure 6.41. This uncertainty is also considered for the VBF HH, $t\bar{t}H$ and VH processes. These uncertainties are treated as correlated between regions and eras but uncorrelated between processes.

- **Parton distribution functions or PDFs (normalization and shape):** The uncertainty due to PDFs is evaluated by taking the sum-in-quadrature of the variations provided by the 103 replicas of NNPDF3.0. This uncertainty is considered for the $ggHH$ signal, VBF HH, $t\bar{t}H$ and VH processes. This uncertainty is correlated across event categories and eras but uncorrelated between processes.

- **Initial and Final State Radiations (ISR and FSR) (normalization and shape):** The uncertainty in the parton shower calculation due to a initial or a final state radiation is also considered for the $ggHH$ signal, VBF HH, and

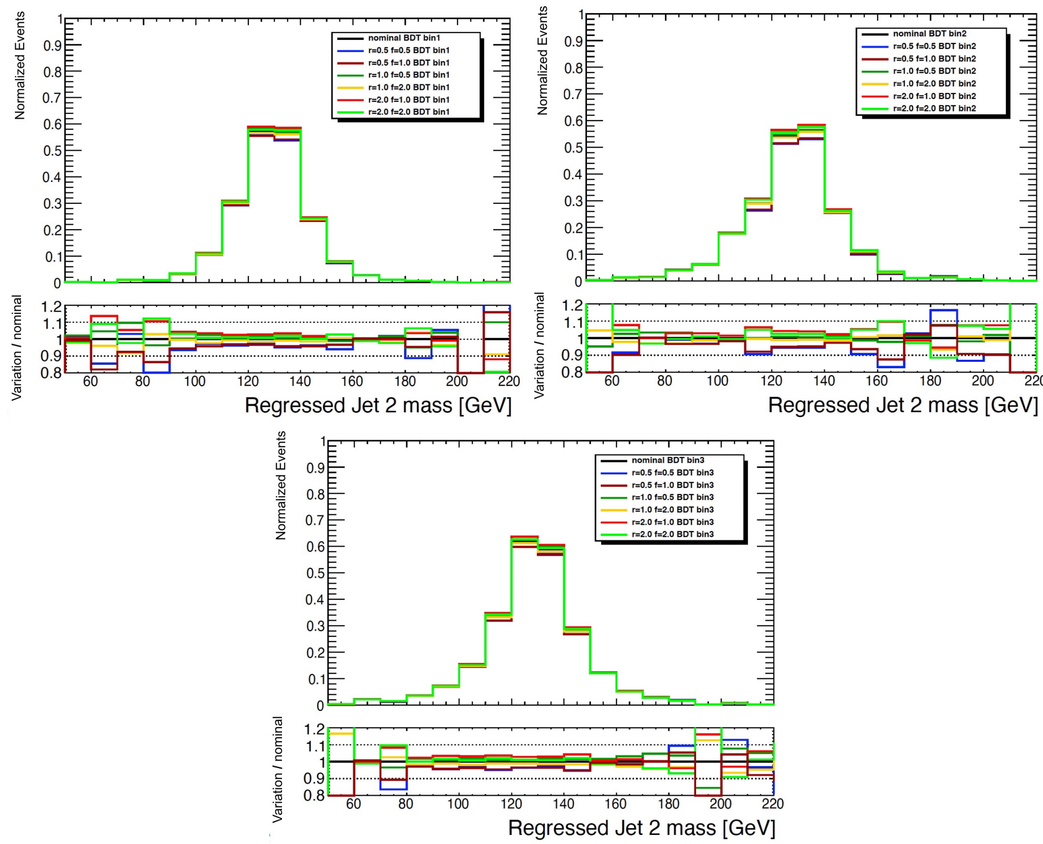


Figure 6.41: The QCD scale uncertainty on the $ggHH$ signal acceptance in event category 1 (top left), 2 (top right) and 3 (bottom) is obtained by taking the envelope of the Jet 2 mass distribution as one varies the (μ_R, μ_F).

V+jets processes. This uncertainty is correlated across event categories and eras but uncorrelated between processes.

- $\mathcal{B}(H \to b\bar{b})$ **(normalization):** The theoretical uncertainty on the branching ratio of Higgs boson decaying to a pair of b-quarks ($\sim 2.5\%$) is also included.

- m_{HH} **theory uncertainty (normalization and shape)**: The uncertainty on the m_{HH} theory prediction is also applied and has been explained previously in Sec. 6.4.3.5.

- **Simulation sample size (normalization and shape):** The per-bin statistical uncertainty arising from the limited size of the simulated samples is taken into account for all signal and background processes. This statistical uncertainty is modelled using the Barlow-Beeston method [156, 157].

The HH production signal strength, is defined as

$$\mu = \sigma_{HH}/\sigma_{HH}^{SM}. \tag{6.21}$$

The major systematic uncertainties are listed in Table 6.5, along with their impact on the measurement of the signal strength. The impact of each uncertainty is calculated by varying the post-fit error up and down by one standard deviation and re-evaluating the signal strength μ by fitting to the data. The impact $\Delta\mu_{HH}$ is defined as :

$$\Delta\mu_{HH} = \sigma_{HH}/\sigma_{HH}^{SM} - \hat{\sigma_{HH}}/\sigma_{HH}^{SM} \tag{6.22}$$

where $\hat{\mu_{HH}} = \hat{\sigma_{HH}}/\sigma_{HH}^{SM}$ is the best fit signal strength obtained by floating all parameters and fitting to the data. $\sigma_{HH}/\sigma_{HH}^{SM}$ is the new signal strength obtained after varying that particular uncertainty whose impact is being measured. The impact pulls of the top-ranked systematic uncertainties on the signal strength measurement are shown in Figure 6.42. No significant pulls are observed.

Uncertainty source	$\Delta\mu$	
Statistical	+2.55	−2.30
Signal extraction	+2.32	−2.06
QCD multijet modeling	+1.12	−1.01
$t\bar{t}$ modeling	+0.28	−0.19
Systematic	+2.09	−0.89
Simulated sample size	+0.55	−0.55
T_{Xbb} selection	+0.72	−0.32
Jet energy and mass scale and resolution	+0.54	−0.39
Trigger selection	+0.26	−0.03
Luminosity measurement	+0.13	−0.04
Pileup modeling	+0.05	−0.06
Other experimental uncertainties	+0.05	−0.03
Theoretical	+0.63	−0.63
Total	+3.30	−2.47

Table 6.5: Major sources of uncertainty in the measurement of the signal strength modifier μ, and their observed impact ($\Delta\mu$) from a fit to the combined data set. Decompositions of the statistical, systematic, and theoretical components of the total uncertainty are specified. The impact of each uncertainty is evaluated by computing the uncertainty excluding that source and subtracting it in quadrature from the total uncertainty. The sum in quadrature for each source does not in general equal the total uncertainty of each component because of correlations in the combined fit between nuisance parameters corresponding to different sources.

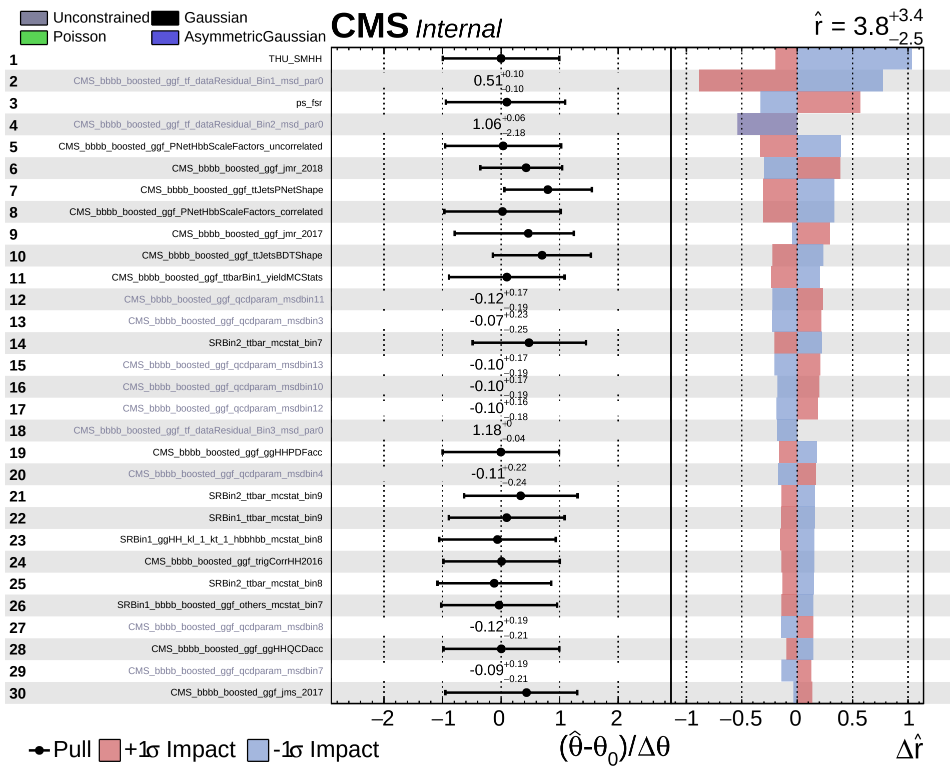


Figure 6.42: The impact pulls of top-ranked systematic uncertainties. Tho dominant sources of uncertainty are the QCD multi-jet modelling, the $t\bar{t}$ background modelling, the $ggHH$ signal modelling and the jet energy and mass scale and resolution.

6.4.5 Results

The signal strength (Eqn. 6.21) is extracted by performing a binned maximum likelihood fit of the Jet 2 regressed mass in the signal region. The fit is simultaneously performed across all event categories and the fail region. The backgrounds for this fit are estimated according to the description in Sec. 6.4.2. The likelihood function for extracting the signal strength was previously described in Sec. 5.6.5. For this analysis, we extract results assuming $m_H = 125\,\text{GeV}$.

6.4.5.1 Expected and observed yields

The expected and observed yields in the various event categories in the signal region are presented in Table 6.6 and Figure 6.52.

6.4.5.2 Upper limits on the SM HH production cross section

We set 95% CL upper limits on $\sigma_{HH} \times \mathcal{B}(HH \to b\bar{b}b\bar{b})$, the product of the $ggHH$ cross section and the $H \to b\bar{b}$ branching fraction, using the Asymptotic CLs

Process	Event category 1	Event category 2	Event category 3
$ggHH$ ($\kappa_\lambda = 1, \mu = 3.8$)	4.94 ± 1.50	4.99± 1.49	5.57± 1.68
VBFHH ($\kappa_\lambda = 1, \mu = 3.8$)	0.04± 0.01	0.04± 0.01	0.05± 0.01
ttH	0.39± 0.04	1.16± 0.09	2.86± 0.25
VH	0.74± 0.09	1.74± 0.17	2.88± 0.24
V+jets,VV	0.55± 0.26	2.61± 0.46	12.79± 0.96
tt+jets	2.15± 0.24	15.59± 1.13	155.27± 5.29
QCD+ggH+VBFH	2.60± 0.27	22.88± 0.68	102.40± 1.53
total signal+background	11.35± 1.57	49.02± 2.06	276.20± 5.60
data	9	71	304

Table 6.6: Summary of post-fit signal and background yields and the observed number of events in each event category in the Jet 2 m_{reg} distribution under the Higgs peak region, i.e., Jet 2 $m_{\text{reg}} \in [110., 140.]$ GeV.

method [151]. The likelihood function calculation uses the modified frequentist approach [159]. The observed and expected upper limits in each of the three event categories and for the combined analysis is presented in Figure 6.44. We place an observed (expected) upper limit of 10 (5) × SM on $\sigma_{HH} \times \mathcal{B}(HH \to b\overline{b}b\overline{b})$.

The 1D likelihood scan of the signal strength μ (Eqn. 6.21) is presented in figure 6.45. In this measurement, all Higgs boson couplings are assumed to be at their SM values. The observed (expected) value of the signal strength is found to be is $\mu = 3.8^{+3.4}_{-2.5}$ ($\mu = 1.0^{+2.4}_{-2.0}$) at 68% CL.

6.4.5.3 Constraints on the Higgs boson self-coupling

The 1D likelihood scan of the Higgs boson self-coupling factor, κ_λ, is presented in figure 6.46. In this measurements, all other Higgs boson couplings are assumed to be at their SM values. The observed (expected) value of the Higgs self coupling is found to be $\kappa_\lambda = -4.2^{+20.9}_{-4.2}$ ($\kappa_\lambda = 1.0^{+10.5}_{-4.7}$) at 68% CL.

6.4.5.4 Constraints on the HHVV coupling

The 1D likelihood scan of the HHVV coupling factor κ_{2V}, is presented in figure 6.47. In this measurement, all other Higgs boson couplings are assumed to be at their SM values. In particular, for this measurement, we process the VBF HH signal samples and fix the $ggHH$ process rate at its SM prediction and treat it as a background. Although our event selection does not target VBF HH production, and we veto the VBF category events from the $ggHH$ events selection, we find that our analysis is

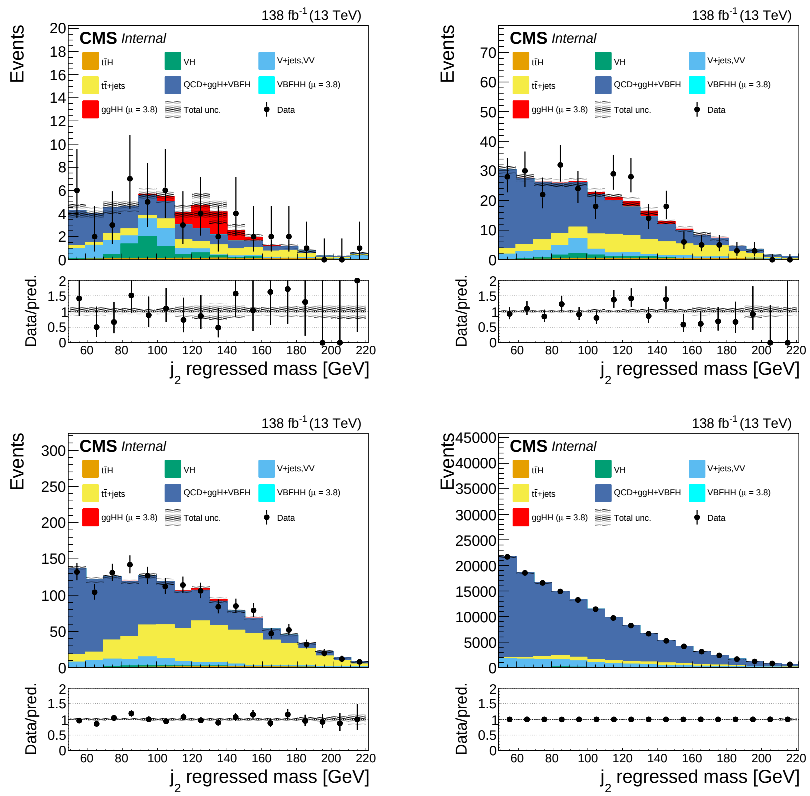


Figure 6.43: This figure shows post-fit distributions for the different event categories corresponding to the best-fit parameters. (Top left) Event category 1, (top right) Event category 2, (bottom left) Event category 3 and (bottom right) QCD fail region.

sensitive to κ_{2V} because of our looser p_{T} requirements on the AK8 jets. The κ_{2V} coupling factor is observed to be $\kappa_{2V} = 0.5^{+0.4}_{-0.2}$ at 68% CL.

6.5 The VBF HH analysis

The next few sub-sections will briefly discuss the VBF HH analysis, which can be found in Ref. [227].

6.5.1 Event selection

For the VBF categories, we require two AK8 jets with $m_{\mathrm{reg}} \in [50, 200]$ GeV. The the p_{T}-leading (p_{T}-subleading) AK8 jet is required to have $p_{\mathrm{T}} > 500$ (400) GeV. The VBF process is characterized by the presence of two forward jets with a large dijet invariant mass and a gap in pseudorapidity. Hence, we require two additional AK4 jets with $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 4.7$, referred to as the VBF jets. The two AK4 jets with the highest p_{T}, separated from both Higgs boson candidate jets by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} > 1.2$, are considered as the VBF jet candidates. We also

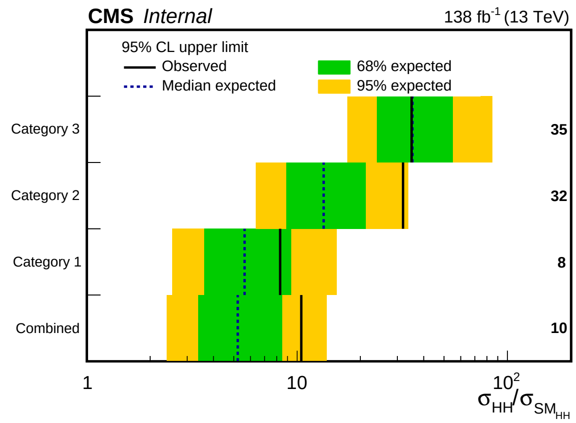


Figure 6.44: Observed (expected) 95% CL upper limits on the $\sigma_{HH} \times \mathcal{B}(HH \rightarrow b\overline{b}b\overline{b})$, in different event categories, and by combining all categories are presented by the solid (dashed) black line. The green and yellow bands represent the expected one and two standard deviation uncertainty bands on the expected limit, respectively.

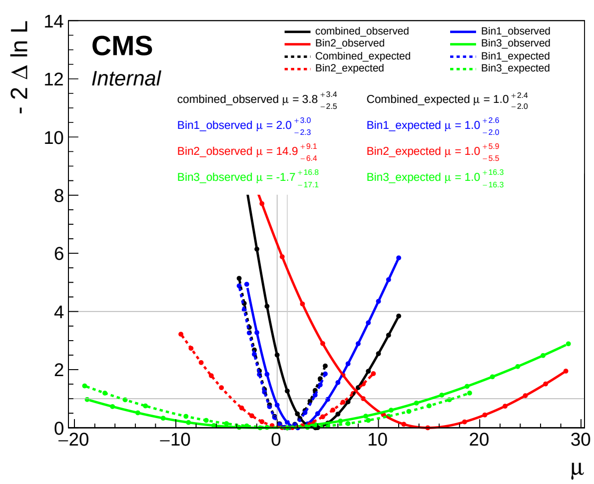


Figure 6.45: The observed and expected 1D likelihood scan of the SM HH production signal strength μ, assuming all Higgs boson couplings to be at their SM values. The different colors represent the different event categories.

require $\Delta\eta_{\mathrm{jj}}^{\mathrm{VBF}} > 4.0$ (absolute pseudorapidity difference between the two VBF jets), and the invariant mass of the two VBF jets ($m_{\mathrm{jj}}^{\mathrm{VBF}}$) must be larger than 500 GeV. Additionally, we require the azimuthal angle between the two AK8 jets, $\Delta\phi_{\mathrm{j_1j_2}} > 2.6$, and the absolute difference in pseudorapidity, $\Delta\eta_{\mathrm{j_1j_2}} > 2.0$. We also veto events with additional electrons or muons.

Three VBF event categories are defined based on the T_{Xbb} scores of the two Higgs boson candidate AK8 jets. Events in the high-purity (HP) category must have both jets with $T_{\mathrm{Xbb}} > 0.98$. Events not in the HP category are categorized as medium

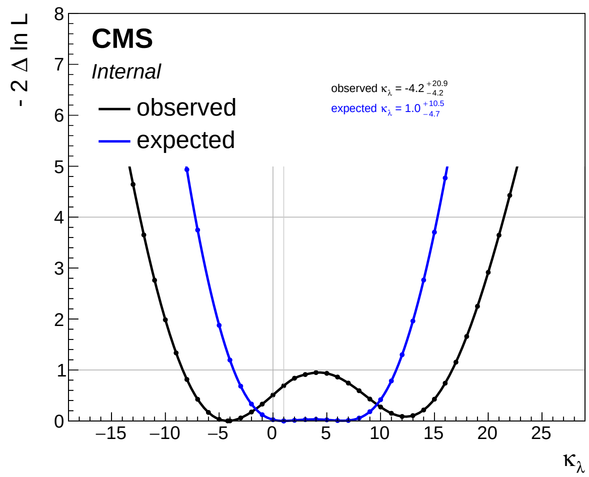


Figure 6.46: The observed and expected 1D likelihood scan of the Higgs boson self-coupling factor, κ_λ, assuming all other Higgs boson couplings to be at their SM values.

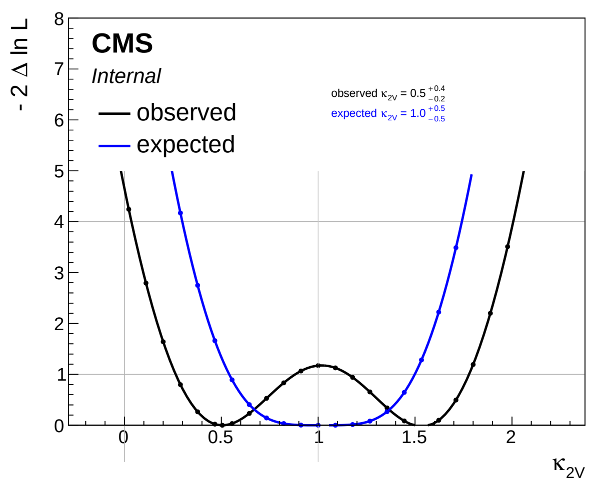


Figure 6.47: The observed and expected 1D likelihood scan of the HHVV coupling factor, κ_{2V}, assuming other Higgs boson couplings to be at their SM values.

purity (MP) if both jets have $T_{\text{Xbb}} > 0.94$, and events not in the HP or MP categories are categorized as low purity (LP) if both jets have $T_{\text{Xbb}} > 0.90$.

Finally, the masses of the two leading AK8 jets are required to be within the Higgs peak window, i.e., Jet 1 $m_{\text{reg}} \in [110, 150]$ GeV and Jet 2 $m_{\text{reg}} \in [100, 145]$ GeV. The final fit variable for signal extraction in the VBF category is m_{HH}, i.e., the invariant mass of the two AK8 jets.

6.5.2 Background estimation

Just like the ggF category, the two major backgrounds in the VBF category are the QCD multi-jet process and the Top quark. The next two sub-sections will briefly discuss the dominant background estimation procedure for the VBF analysis.

6.5.2.1 QCD

Just like the ggF category, the QCD background estimation in the VBF category is done in a data-driven manner, using the **ABCD** method. The procedure for this method is as follows

- Define the signal region **D**.
- Define a control region **C** that is orthogonal to the search region but enriched in the background process to be estimated.
- Modify another set of requirements (uncorrelated with the selections altered to obtain region C), to obtain two additional regions, **A** and **B**.
- Use regions A and B to derive correction factors, known as *transfer factors* (TFs), that quantify the difference in background normalization (and possibly also in the shape) between the regions A and B.
- The transfer factors are then used to normalize the background estimate obtained from region C to the search region D.

In the VBF catgeory, a control region (CR) C enriched in QCD multijet events is selected by requiring the T_{Xbb} score of both Higgs boson candidate AK8 jets to be between 0.1 and 0.9. Since the region C also has a non-negligible $t\bar{t}$ contamination, the QCD multijet and $t\bar{t}$ yields are estimated simultaneously via a binned maximum likelihood fit to the observed data in the m_{HH} distribution. The resulting m_{HH} distribution for QCD is then normalized to the search region D (signal region (SR): HP, MP and LP) with transfer factors w_i, which are derived separately using the side-bands of the sub-leading AK8 jet mass. We define control region A with the same T_{Xbb} selections as region C, and region B is defined by the same T_{Xbb} selections as region D. However, for both regions A and B, the sub-leading jet m_{reg} is required to be either below ($50 < m_{\mathrm{reg}}^{subl} < 100\,\mathrm{GeV}$) or above ($145 < m_{\mathrm{reg}}^{subl} < 200\,\mathrm{GeV}$) the signal mass window. To ensure sufficient statistics for the transfer factor derivation, the selection requirement on the leading jet mass is also relaxed, requiring $50 < m_{\text{reg}}^{lead} < 200\,\text{GeV}$. The TFs are then defined as:

$$w_i = \frac{N^B_{QCD,i}}{N^A_{QCD,i}}. \tag{6.23}$$

A separate TF is derived for each m_{HH} bin, in each search category (HP, MP and LP), and for each year.

The QCD background in the search region D is then defined as :

$$N^D_{QCD,i} = w_i \times N^C_{QCD,i}. \tag{6.24}$$

Binned m_{HH} distributions for all SR and CR event categories are fitted simultaneously, with the QCD yields in CRs unconstrained in the fit, and with the QCD yield in the SR categories determined by the CR yields via the TFs. Fig. 6.48 illustrates how the ABCD method is applied in this analysis.

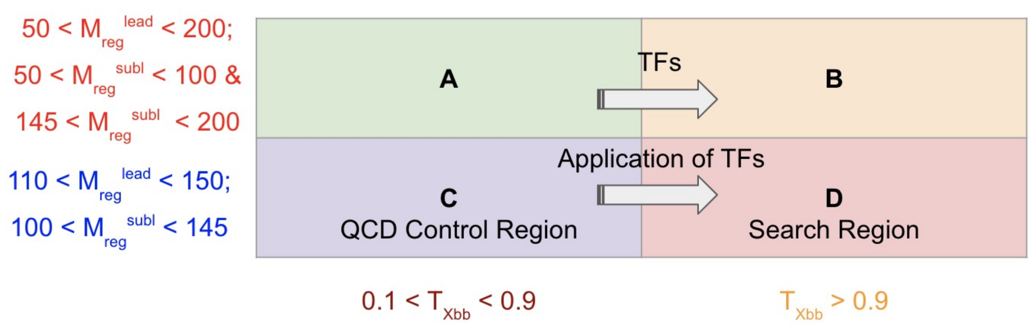


Figure 6.48: The ABCD method used for QCD background estimation in the VBF analysis.

6.5.2.2 Top

For the VBF categories, the top background is estimated from the simulation, with certain corrections derived from data.

A region enriched in semi-leptonic $t\bar{t}$ events is used to extract two sets of corrections, one accounting for the difference in the $H \to b\bar{b}$ mis-identification efficiency in $t\bar{t}$ events between data and simulation, and another to correct the overall normalization of the $t\bar{t}$ process in our signal region. These corrections are derived using a "tag-and-probe" method. A muon, in combination with $\vec{p}_{\text{T}}^{\text{miss}}$ (representing the leptonic W boson decay) is used as the "tag", and an AK8 jet in the opposite hemisphere is considered as the "probe" jet. We define this region with a set of selections that closely follows the definition of the $t\bar{t}$-enriched region in Ref. [221]. The mass distribution of the AK8 jet is then fitted to data (as detailed in Ref. [221]), to derive a set of jet p_{T} dependent scale factors to correct the response of the T_{Xbb} discriminant in each of the three categories (HP, MP and LP), as well as for the overall normalization of the $t\bar{t}$ process.

6.5.3 Corrections to data and simulation

The VBF analysis uses corrections similar to the ggF analysis, as detailed in Sec. 6.4.3.

6.5.4 Systematic uncertainties

The VBF analysis uses systematic uncertainties similar to the ggF analysis, as detailed in Sec. 6.4.4.

6.5.5 Results

A binned maximum-likelihood fit using the m_{HH} templates is performed simultaneously with all SR and CR event categories. The SR distributions of the m_{HH} after a background-only fit to the data are shown in Fig. 6.49. A signal hypothesis corresponding to $\kappa_{2V} = 0$, with other couplings set to the SM values, is overlaid for illustration. No excess is found in data compared to the background-only hypothesis.

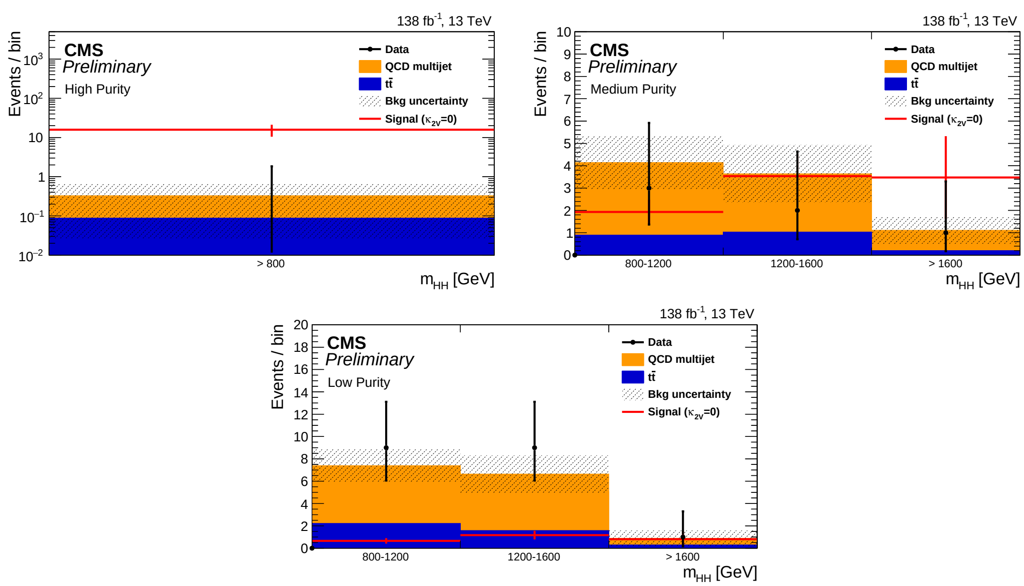


Figure 6.49: The m_{HH} distribution is shown here with a background-only fit to the data, for the HP (upper), MP (middle), and LP (lower) categories.

The 95% CL exclusion limits are set on $\sigma(pp \to qqHH) \times \mathcal{B}(HH \to b\bar{b}b\bar{b})$ for a range of signal hypotheses corresponding to different values of κ_{2V} and κ_V couplings. In Fig. 6.50 (Left), the expected and observed limits are presented as a function of the κ_{2V} coupling assuming that all other Higgs boson couplings are at their SM values. The κ_{2V} is constrained in the range $0.6 < \kappa_{2V} < 1.4$ at 95% CL. A hypothesis of vanishing coupling, namely $\kappa_{2V} = 0$ with other couplings equal to 1, is excluded at a CL higher than 99.99%. Similarly, in Fig. 6.50 (Right), the 95% CL exclusion limits are shown as a function of κ_V, with other couplings fixed to 1. The κ_V coupling is constrained in the range $1.2 < \kappa_V < 0.8$ or $0.8 < \kappa_V < 1.2$.

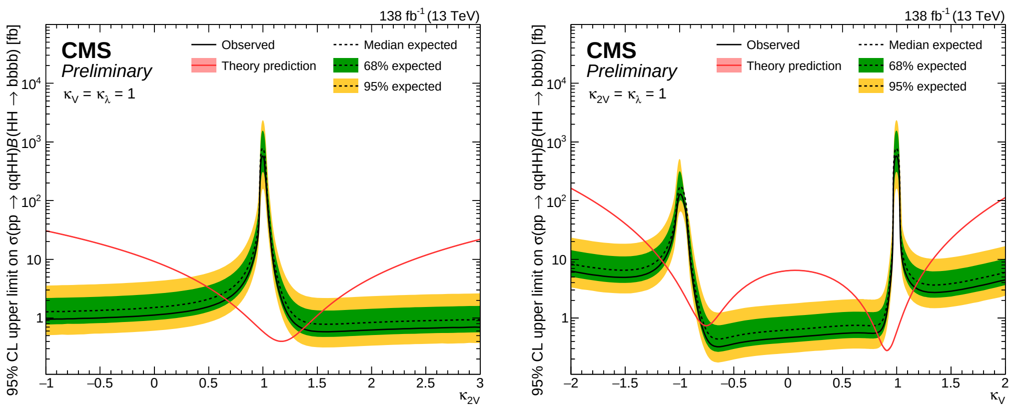


Figure 6.50: (Left): Observed (solid line) and expected (dashed line) 95% CL exclusion limits on $\sigma(pp \to qqHH) \times \mathcal{B}(HH \to b\bar{b}b\bar{b})$, as a function of the κ_{2V} coupling, with other couplings fixed to the SM values. (Right): Observed (solid line) and expected (dashed line) 95% CL exclusion limits on $\sigma(pp \to qqHH) \times \mathcal{B}(HH \to b\bar{b}b\bar{b})$, as a function of the κ_V coupling, with other couplings fixed to the SM values.

6.6 Combination results

This section will describe the results from combining the $ggHH$ (Sec. 6.4) and VBF HH (Sec. 6.5) analyses.

6.6.1 Systematic uncertainty treatment

Most systematic uncertainties between the two analyses are treated as fully correlated, including pileup, luminosity, jet energy scale and resolution, and theory uncertainties, like $\mathcal{B}(H \to b\bar{b})$, PDFs, QCD scale, and parton shower effects. However, for the T_{Xbb} score, the efficiency scale factors depend on the jet p_{T}, and for the ggF category, most of the events are in the 300–500 GeV p_{T} range (87% of

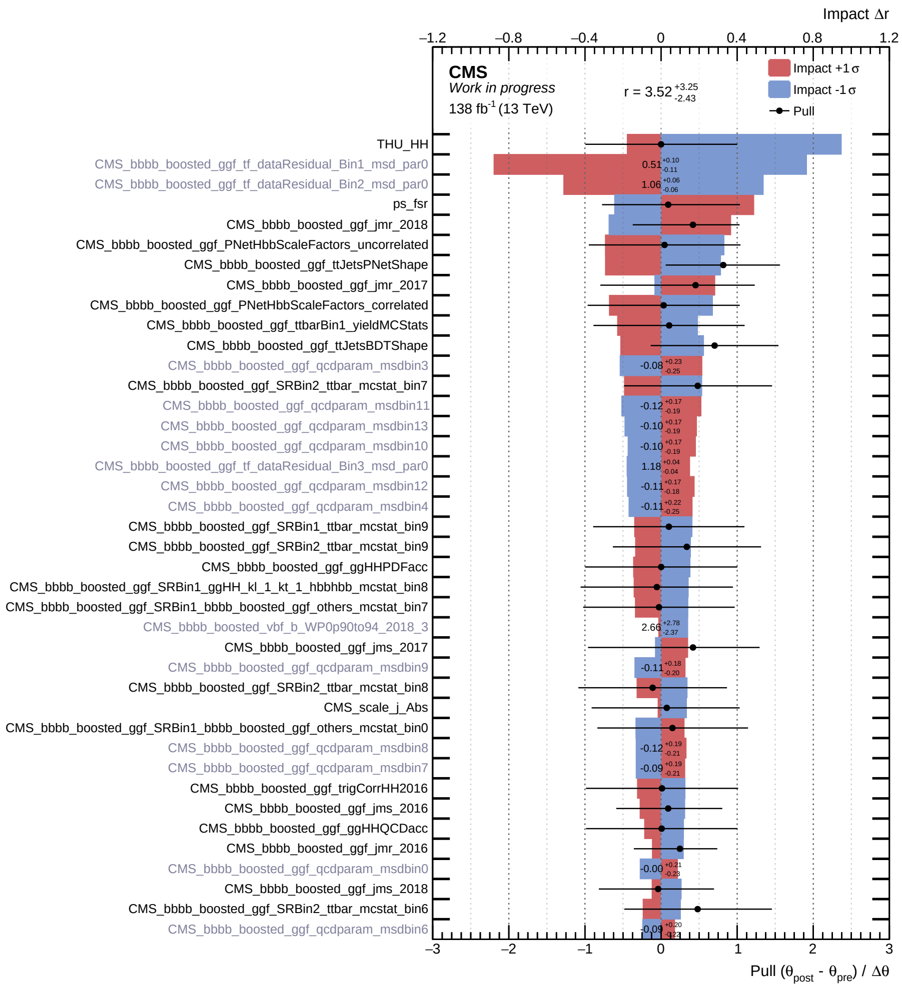


Figure 6.51: Pulls and impacts of the first 40 nuisance parameters ranked by impact.

$ggHH$ signal has at least one AK8 jet p_{T} between 300–500 GeV). On the contrary, for the VBF category, most signal events have jet p_{T} >500 GeV. The jet mass scale and resolution are therefore treated as uncorrelated due to the different phase space. Fig. 6.51 shows the pulls and impacts of the first 40 nuisance parameters ranked by impact. No significant pulls are observed.

6.6.2 The Jet 2 m_{reg} and m_{HH} distributions

A binned maximum likelihood fit to the observed Jet 2 m_{reg} (for ggF categories) and m_{HH} (for VBF SR) is performed using the sum of the signal and background contributions. These distributions in the various event categories are presented in Figure 6.52.

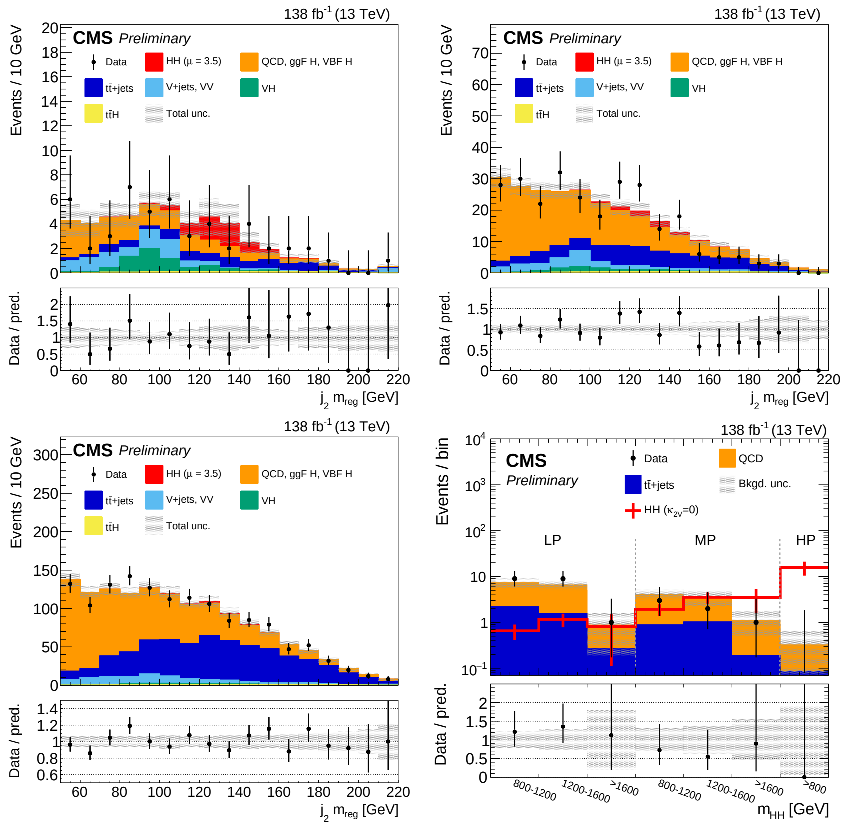


Figure 6.52: This figure shows post-fit distributions for the different event categories in the ggF and VBF analyses, corresponding to the best-fit parameters. (Top left) ggF event category 1, (top right) ggF event category 2, (bottom left) ggF event category 3 and (bottom right) VBF Signal Region.

6.6.3 Upper limit on the inclusive HH production cross section

Fig. 6.53 shows the upper limits on the inclusive HH production cross section. The combined observed (expected) 95% upper limit on the cross section is 9.9 (5.1) × SM, and this result is mainly driven by the ggF channel.

6.6.4 Constraints on the various Higgs boson couplings

Figure 6.54 shows the 1D expected and observed likelihood scans in κ_λ (top), κ_V (bottom left), and κ_{2V} (bottom right) for the ggF and VBF channels. As shown in Fig. 6.54, the sensitivity to κ_V and κ_{2V} is largely driven by the VBF channel, while the sensitivity to κ_λ is largely driven by the ggF channel.

Figure 6.55 show the 1D upper limits on the inclusive HH cross section as a function of κ_λ (top), κ_V (bottom left), and κ_{2V} (bottom right) for the combined ggF and VBF channels. For the combined analysis, at 95% CL, we put constraints on the allowed values of the couplings as: $\kappa_\lambda \in [-9.9, 16.9]$ when $\kappa_V = 1$, $\kappa_{2V} = 1$;

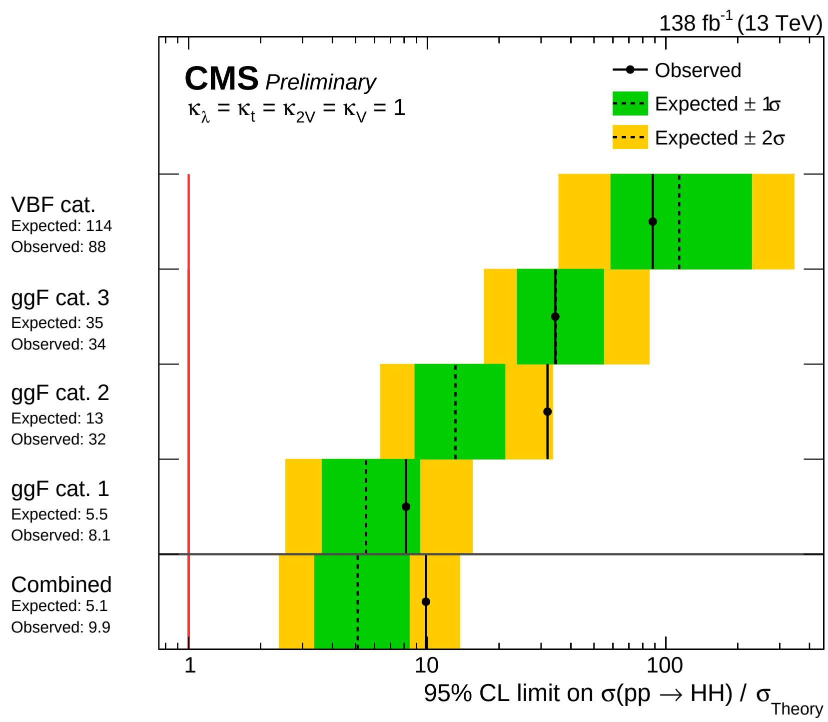


Figure 6.53: Upper limits on the inclusive SM signal strength.

$\kappa_V \in [-1.17, -0.79] \cup [0.81, 1.18]$ when $\kappa_\lambda = 1$, $\kappa_{2V} = 1$; $\kappa_{2V} \in [0.62, 1.41]$ when $\kappa_\lambda = 1$, $\kappa_V = 1$. We exclude $\kappa_{2V} = 0$ with a significance of 6.3 standard deviations for the first time, when other H couplings are fixed to their SM values.

Finally, Fig. 6.56 shows the 2D profile likelihood scan as a function of κ_λ and κ_{2V} (left) and κ_V and κ_{2V} (right) for the combined ggF and VBF channels.

6.7 Current results on HH production from CMS

This section will briefly discuss Fig. 2.8 from Sec. 2.4.3, that summarizes the current results on Higgs pair production from different CMS analyses (Ref. [26, 55–59]). To put things in to perspective, the boosted $HH \to b\bar{b}b\bar{b}$ result described in this chapter produced an upper limit on the inclusive Higgs production cross section ($\mu = \sigma_{HH}/\sigma_{HH}^{SM}$), which is ~30 × better than the previous CMS analysis using just the 2016 dataset [228]. The $b\bar{b}$ ZZ and Multilepton analyses are completely new channels targeting the HH production. The resolved $b\bar{b}b\bar{b}$, $b\bar{b}$ $\gamma\gamma$ and $b\bar{b}$ $\tau\tau$ have all performed ~ 3-5 × better than their previous counterparts that used just the 2016 dataset [229]. It is evident from Fig. 2.8 (top left), that the boosted $HH \to b\bar{b}b\bar{b}$ result described in this chapter has the best sensitivity to the SM HH production cross section, amongst all other CMS analyses.

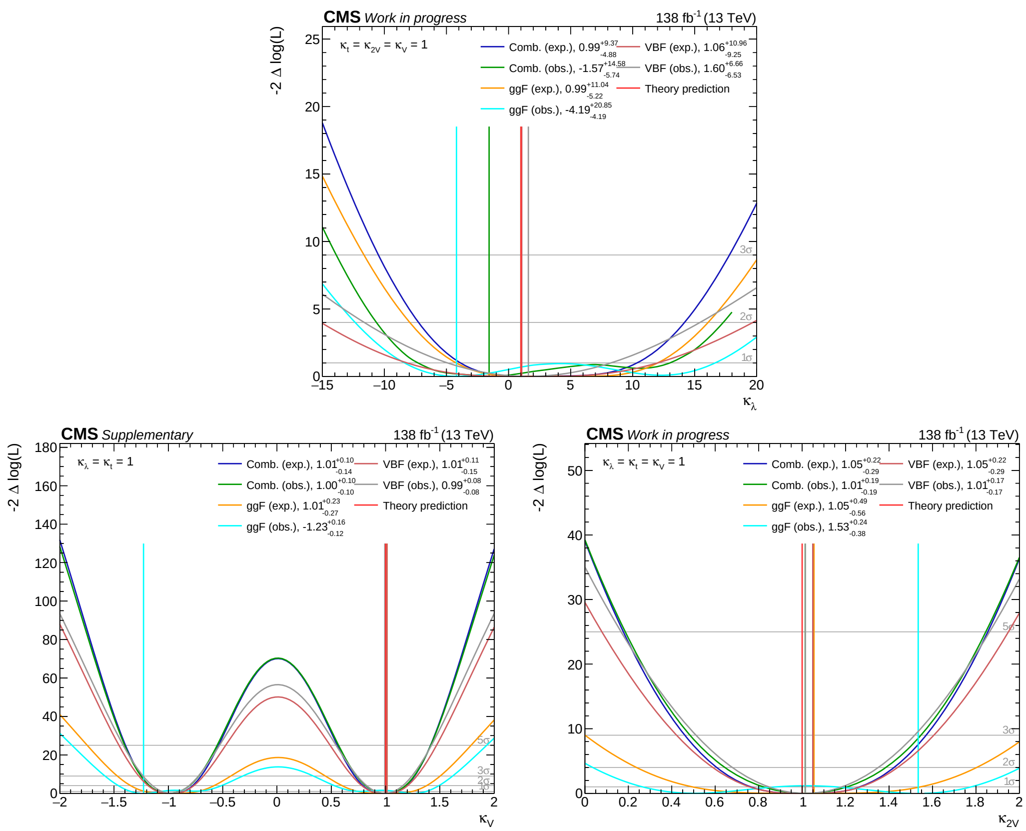


Figure 6.54: Comparison of the 1D expected and observed likelihood scans in κ_λ (top), κ_V (bottom left), and κ_{2V} (bottom right) for the ggF channel, VBF channel, and combined ggF and VBF channels.

Additionally, it is clear from Fig. 2.8 (top right and bottom right), that the boosted $HH \to b\bar{b}b\bar{b}$ result described in this chapter has the best sensitivity to the VBF HH production cross section in a BSM scenario with $\kappa_{2V} = 0$, and also puts the best constraints on the allowed values of κ_{2V}, amongst all other CMS analyses. The $b\bar{b}\ \gamma\gamma$ analysis [58] puts the best constraints on the allowed values of κ_λ amongst all other CMS analyses.

6.8 Future of HH production

A combined ATLAS + CMS projection on the Higgs pair production at the HL-LHC was studied in Ref. [173]. Several ATLAS and CMS results on the 2016 or 2016 + 2017 dataset were combined and projected signal significance and constraints on κ_λ were derived for 3 ab^{-1} of pp collision data. The results are summarized in Table 6.7 and Fig. 6.57. The expected combined ATLAS + CMS significance for the HH process is 4 σ at HL-LHC, and the 68% confidence intervals for κ_λ are $0.52 \leq \kappa_\lambda$

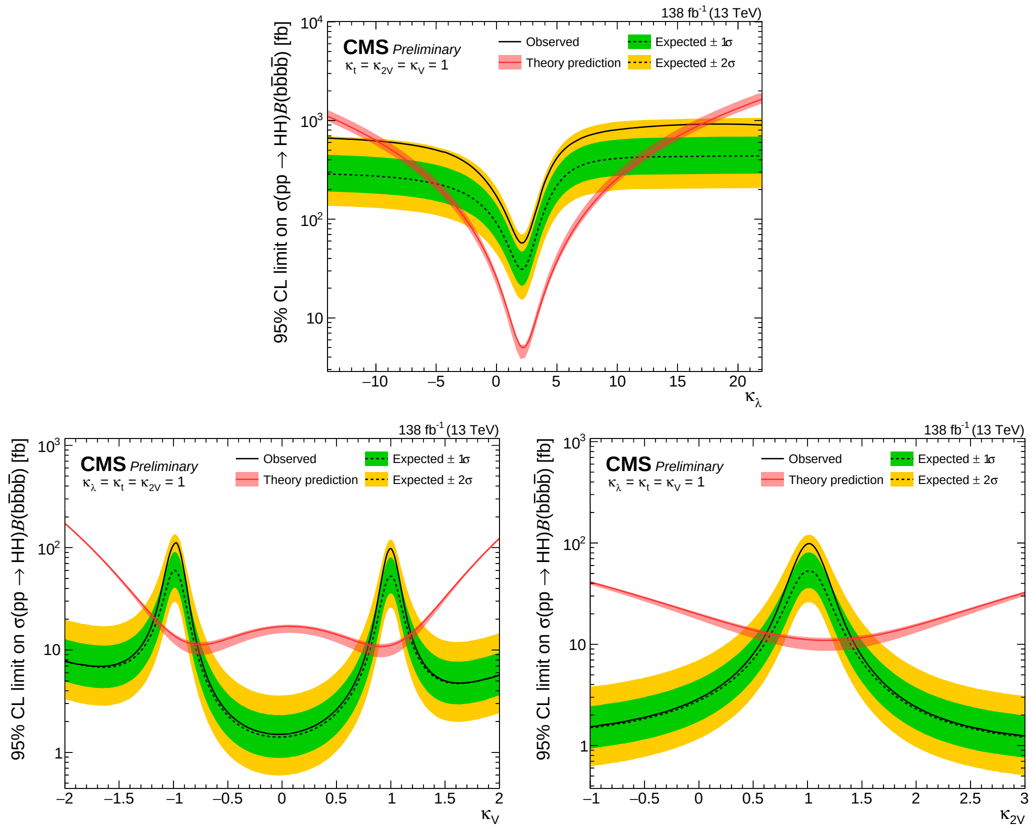


Figure 6.55: 1D upper limits on the inclusive HH cross section as a function of κ_λ (top), κ_V (bottom left), and κ_{2V} (bottom right) for the combined ggF and VBF channels.

≤ 1.5. However, this projection does not come from the full Run 2 dataset and does not consider the enhancements that come from new and improved analyses, like the boosted $HH \to b\bar{b}b\bar{b}$ result described in this chapter. Overall, with advanced physics object reconstruction techniques and improved ways to constrain systematic uncertainties, there is a potential for discovery of the HH process at the HL-LHC.

Table 6.7: Projected significance in standard deviations of the individual HH production channels as well as their combination, at the HL-LHC [173].

Target channel	ATLAS	CMS
$HH \to b\bar{b}b\bar{b}$	0.61	0.91
$HH \to b\bar{b}\ \tau\tau$	2.1	1.4
$HH \to b\bar{b}\ \gamma\gamma$	2.0	1.8
$HH \to b\bar{b}\ VV\ (ll\nu\nu)$	-	0.56
$HH \to b\bar{b}\ ZZ\ (4l)$	-	0.37
combined	3.0	2.6
Combined ATLAS + CMS	4.0	

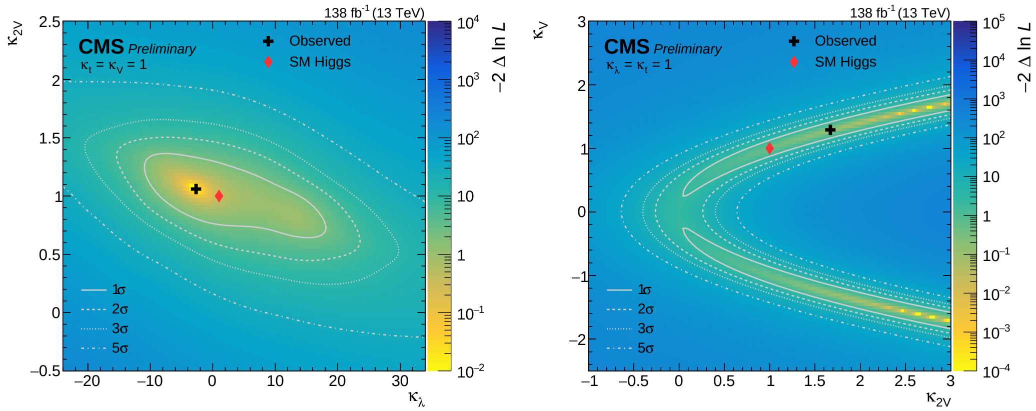


Figure 6.56: 2D profile likelihood scan as a function of κ_λ and κ_{2V} (left) and κ_V and κ_{2V} (right).

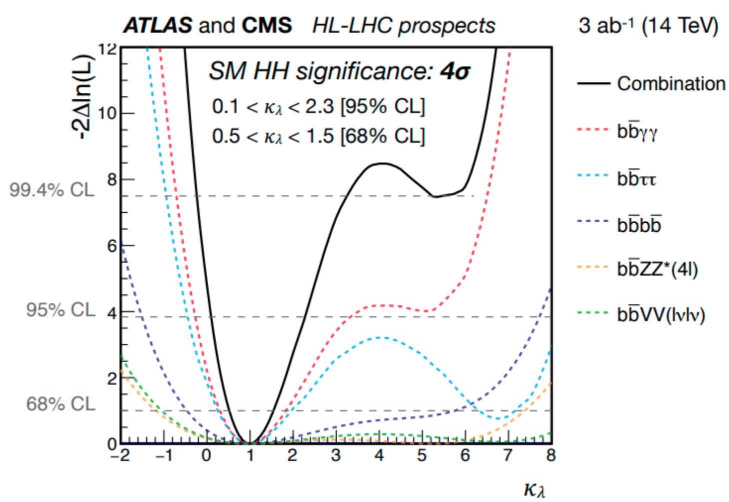


Figure 6.57: A projected likelihood scan of κ_λ at HL-LHC, calculated by performing a conditional signal+background fit to the background and SM signal. The coloured dashed lines correspond to the combined ATLAS and CMS results by channel, and the black line to their combination. Plot is taken from [173].

Chapter 7

SUMMARY AND OUTLOOK

It has been 10 years since the ATLAS and CMS Collaborations at the LHC announced the landmark discovery of a Higgs boson at a mass of around 125 GeV. The Higgs discovery was made within 2 years of the beginning of the LHC operations, and with a total of ~ 10 fb^{-1} of data collected at $\sqrt{s} = 7$ and 8 TeV. Since then, the LHC had ramped up its operation to $\sqrt{s} = 13$ TeV, and collected about 15 times more data during Run 2 (2015-2018). With the increased data, we have been able to test several properties of the Higgs boson and compare them to predictions from the SM. The gauge couplings of the Higgs to SM bosons and Yukawa couplings to third generation fermions have been confirmed, its spin-parity quantum numbers have been determined, and its mass has been established to a precision of 0.1%. However, we are yet to observe certain rare Higgs decay modes in nature. Painting a complete picture of the Higgs boson is essential to understanding the precision of the SM, as any new physics process can alter one or more of the predicted properties of the Higgs. This thesis described the search for two separate rare Higgs processes in nature.

Chapter 5 described the CMS Run 2 search for $H \rightarrow \mu\mu$ decays, which have a tiny branching fraction of 2.18×10^{-4} at $m_H = 125.38$ GeV. It may surprise the reader to learn that the $H \rightarrow ZZ^* \rightarrow 4l$ ($l = e/\mu$) process, which was one of the discovery channels for the Higgs in 2012, has a branching fraction of 1.25×10^{-4}, which is very similar to $H \rightarrow \mu\mu$. The major background for $H \rightarrow \mu\mu$ is the Drell-Yan + jets process, with a $\sigma \times \mathcal{B}$ of 640.6 pb. On the other hand, the major background for $H \rightarrow ZZ^* \rightarrow 4l$ is the SM production of $ZZ^* \rightarrow 4l$ with a $\sigma \times \mathcal{B}$ of 7.34 pb. Thus, the major background for $H \rightarrow \mu\mu$ is ~90 times stronger than the major background for $H \rightarrow ZZ^* \rightarrow 4l$, making $H \rightarrow \mu\mu$ a very difficult process to observe at the LHC. To mitigate this issue, the Run 2 CMS $H \rightarrow \mu\mu$ search used four dedicated analysis strategies, each targeting a different Higgs production mode, and used sophisticated machine learning techniques like boosted decision trees and deep neural networks, to enhance the signal sensitivity. The improved analysis techniques lead to the first evidence (3 σ excess) for $H \rightarrow \mu\mu$ decays. The best-fit signal strength with the corresponding 68% CL interval for $m_H = 125.38$ GeV was found to be $\hat{\mu} = 1.19\,^{+0.41}_{-0.39}\,(\text{stat})^{+0.17}_{-0.16}\,(\text{syst})$. Assuming SM production cross sections for the various modes, the $H \to \mu\mu$ branching fraction is constrained at 95% CL to be within $0.8 \times 10^{-4} < \mathcal{B}(H \to \mu\mu) < 4.5 \times 10^{-4}$. The corresponding 68% and 95% CL intervals for the κ_μ are $0.91 < \kappa_\mu < 1.34$ and $0.65 < \kappa_\mu < 1.53$, respectively. The $H \to \mu\mu$ process is expected to be discovered at the HL-LHC, and the $_\mu$ will be measured with an overall uncertainty of 4%.

Another rare Higgs process is the Higgs pair (HH) production, which is an important probe for studying the Higgs self coupling λ, which will allow us to understand the stability of the electroweak vacuum and possibly the origins of matter anti-matter asymmetry in our universe. However, the HH process has an extremely small cross section of ~32 fb^{-1} at 13 TeV, and is yet to be observed in nature. Chapter 6 described the CMS Run 2 search for $HH \to b\overline{b}b\overline{b}$ with highly boosted Higgs bosons. The four-bottom-quark final state suffers from large backgrounds and a poor decay channel resolution, despite having the largest branching ratio (33.9%) amongst all HH decays. To enhance the signal sensitivity in the 4b final state, a dedicated jet identification algorithm based on graph neural networks was used to identify boosted H→ bb jets. The search was divided into two parts, each targeting a separate HH production mode. The search put constraints on the allowed values of the couplings as: $\kappa_\lambda \in [-9.9, 16.9]$ when $\kappa_V = 1$, $\kappa_{2V} = 1$; $\kappa_V \in [-1.17, -0.79] \cup [0.81, 1.18]$ when $\kappa_\lambda = 1$, $\kappa_{2V} = 1$; $\kappa_{2V} \in [0.62, 1.41]$ when $\kappa_\lambda = 1$, $\kappa_V = 1$. A scenario with $\kappa_{2V} = 0$ was excluded with a significance of 6.3 σ for the first time, when other H couplings are fixed to their SM values. The combined observed (expected) 95% upper limit on the HH production cross section was found to be 9.9 (5.1) × SM. Chapter 6 also has a brief discussion on the status of current HH results from CMS, and the boosted $HH \to b\overline{b}b\overline{b}$ analysis has the best sensitivity to the SM HH production cross section, the best sensitivity to the VBF HH production cross section in a BSM scenario with $\kappa_{2V} = 0$, and also puts the best constraints on the allowed values of κ_{2V}, amongst all other CMS analyses. It is expected that with improved physics object reconstruction techniques and smart ways to constrain systematic uncertainties, the HH process has a potential for discovery at the HL-LHC.

To expand the current physics program, the LHC will be undergoing major upgrades after Run 3, to operate at a higher instantaneous luminosity. The upgraded machine will be known as the High Luminosity LHC, and will collect about 3 ab^{-1} of data by the end of its 10 year operation period. The HL-LHC will be a Higgs factory, and is expected to produce > 150 million Higgs bosons (over 1 million for each of the main production mechanisms, spread over many decay modes). The goal of the HL-LHC is to perform precision measurements in several rare Higgs decay channels and potentially observe the HH process. It will also target other precision tests of the SM (like lepton flavor universality or LFU) and expand the available search phase-space of BSM theories (for example, long lived particles of LLPs). The increased instantaneous luminosity at the HL-LHC will however also mean increased levels of radiation, and the existing sub-detectors of the CMS Experiment will be undergoing major upgrades to maintain the LHC levels of particle reconstruction efficiency. Chapter 4 described the planned MIP Timing Detector (MTD) upgrade for CMS at the HL-LHC. The MTD will be a time-of-flight (TOF) detector, designed to provide a precision timing information for charged particles, with a resolution of ~30 ps. The MTD is divided into a barrel and a endcap timing layer. Chapter 4 mainly focused on the Barrel Timing Layer (BTL), and described several R&D tests that have been performed on the BTL sensors (SiPMs + LYSO crystals), including various sensor characterization studies (time resolution, light yield, etc.) and optimization of the BTL design geometry. The chapter also discussed mock test setups to cool the sensors, since it is known to be an effective way of mitigating the increased dark current rates in the sensors due to radiation damage. In the end, carefully controlling all aspects of the BTL design will help it deliver its promised timing performance. This will help to maintain/improve the reconstruction efficiencies of physics objects like b-jets (precise secondary vertex reconstruction from TOF information) despite the increased levels of pile-up, and in turn, will enhance the sensitivity of various physics analyses, like $HH \to b\bar{b}b\bar{b}$.

This thesis covered topics that targeted the single Higgs and double Higgs production modes at the LHC. Another interesting phenomena that can be explored in the future is the triple Higgs production (HHH), which is sensitive to the Higgs quartic coupling λ_4 (the constant in front of the H^4 term in Eqn. 2.29). According to the SM, the value of the Higgs tri-linear self coupling (λ_3) is equal to the Higgs quartic coupling. However, it is widely accepted that SM is a low-energy approximation (effective field theory or EFT) of a more complex theory valid at higher energy scales, and this could imply $\lambda_3 \neq \lambda_4$. The HHH process is therefore important to understand the shape of the Higgs potential at higher energy scales. The SM production cross section for the HHH process is very tiny, with $\sigma_{HHH}^{NNLO} = 0.103$ fb at 13 TeV, implying that only about ~400 HHH events will be produced even at the HL-LHC. However, in certain Two Higgs- Doublet (2HDM) extensions of the SM, this rate can be further enhanced, through the production of a new heavier

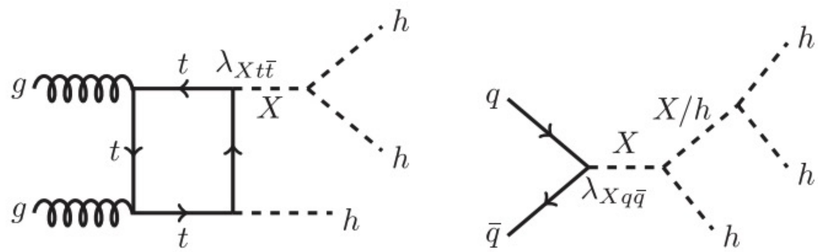


Figure 7.1: Semiresonant (left) and other resonant and nonresonant (right) tri-Higgs production diagrams via heavier scalar particle X [230].

Higgs like scalar particle [230], as shown in Fig. 7.1. One can then study the HHH production rates at Run 3 or HL-LHC as a probe of extended BSM scalar sectors, using various combinations of final state Higgs decays (6b, 4b2τ, 4b2γ, etc).

Run 3 of the LHC is beginning from July 2022, and will collect $\sim$150 fb^{-1} of data by the year 2025, at $\sqrt{s}$ = 13.6 TeV. Many major physics investigations (SM precision tests and BSM searches) will continue to advance at CMS during this period, and also at the HL-LHC in the future.

Appendix A

FEASIBILITY STUDIES FOR A MUON MIP CALIBRATION TECHNIQUE FOR CMS HCAL ENDCAPS

A.1 Introduction

The HCAL detector was discussed in Sec. 3.2.3. After the Phase I upgrade [74], the Hadron Calorimeter (HCAL) is read out with SiPMs in the HB and HE, and has more longitudinal segmentation to improve pileup rejection (see Fig. 3.8). The HB (HE) is read out as four (six/seven) longitudinal segments. This depth segmentation motivates inter-calibration of the different depths using muon MIPs, as the energy deposited by a muon MIP in each HCAL active layer (columns within a depth) is expected to be the same everywhere.

The HCAL is divided into various segments with some lateral $\Delta\eta \times \Delta\phi$ dimensions (see Sec. 3.2.3). Each division in η and ϕ is numbered for convenience, and is called $i\eta$ or $i\phi$, respectively. The HCAL $i\eta - i\phi$ map is shown in Fig. A.1. The HE covers the range 17 <= $|i\eta|$ <= 29. There is full tracker coverage in $|\eta| < 2.4$ (or $|i\eta| < 26$), $|i\eta| = 26$ has partial tracker coverage and $|i\eta| = 27$, 28, and 29 has no tracker coverage. The muon chamber coverage also extends only up to $|\eta| < 2.4$. This makes calibration of the high $|i\eta|$ regions of the HE extremely difficult. A detailed view of the HE depth segmentation is shown in Fig. A.2. In the past, the calibration values derived per depth for HE $|i\eta| = 25$ were also used for $|i\eta| > 25$, with the assumption that the energy deposited by a muon MIP in each HCAL active layer remains the same. However, this method is not necessarily accurate, since $|i\eta| > 25$ has 7 depth segmentations, whereas $|i\eta| = 25$ has only 6 depth segmentations.

We will explore the feasibility of calibrating the outside tracker regions of HE using $Z \to \mu\mu$ events, in the 2018 pp collision data, in a tag and probe fashion. We will require one of the muons from the Z decay to be well reconstructed and isolated (tag muon) inside the tracker region ($21 \leq |i\eta| \leq 24$), such that it enters and exits the same HE $i\eta$ tower. The other muon will be used as a probe, and can be in HE $|i\eta| =$ 27/28/29. The details of this method are described in the next few sections.

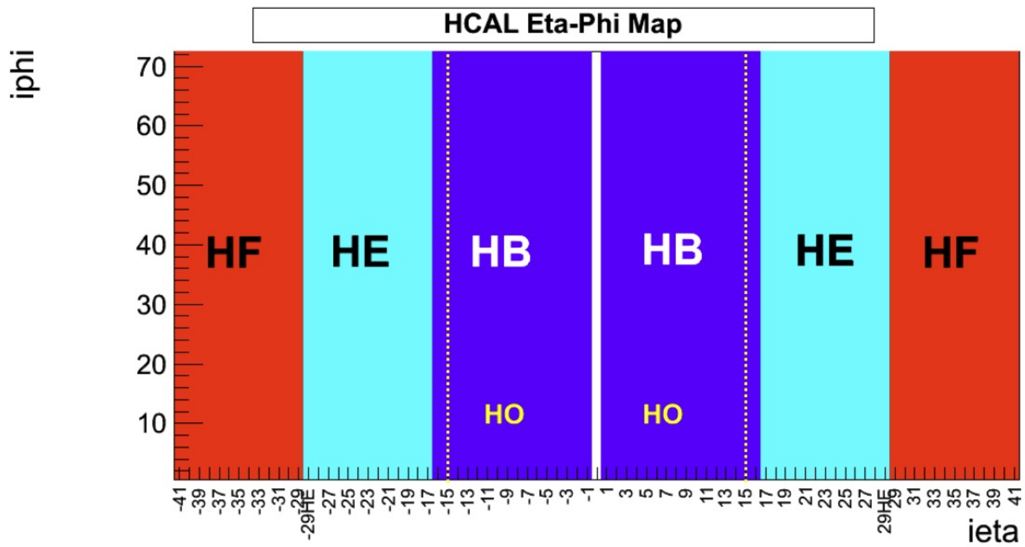


Figure A.1: The HCAL $i\eta - i\phi$ map [231]. HB covers the range |iη|<=16; HE covers the range 17 <= |$i\eta$| <= 29; HF covers the range |iη|>=29; HO covers the range |iη|<=15 (the region within the dashed yellow lines).

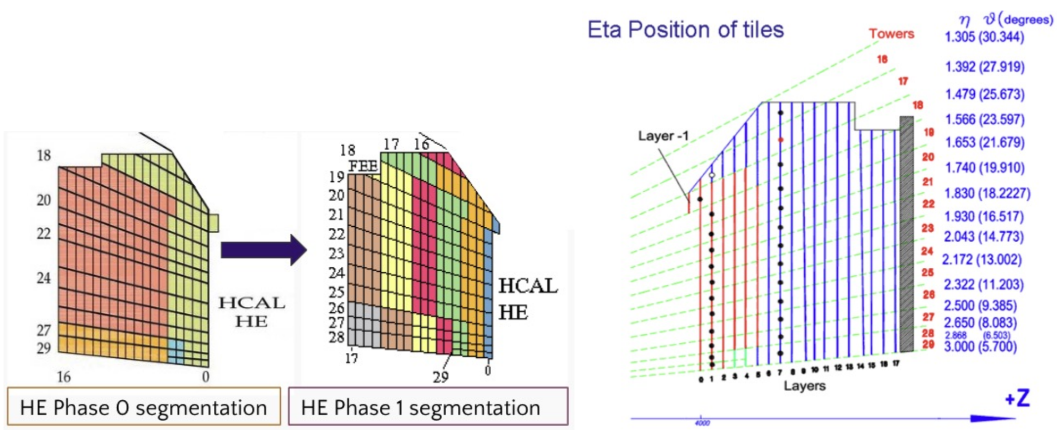


Figure A.2: Longitudinal segmentation of HE before and after the Phase I upgrade and the iη and pseudorapidity mapping of HE.

A.2 Event selection

We will describe a set of selections for choosing the events we used in this study. The event selection proceeds as

- **Tag muon or μ_1** : The tag muon is required to have p_T > 25 GeV and should be inside the tracker coverage area, i.e., $21 \leq |i\eta| \leq 24$. The tag muons used in this analysis are reconstructed with the Tight ID (Sec. 5.3.2) and must be well isolated (surrounding energy summed from other particles is less than 15% of muon p_T). Additionally, the muon is required to enter and exit the same HE iη tower, and this tower is required to have hot cells i.e. this tower should have a larger measured energy than the surrounding HCAL towers.

- **Veto events with fully reconstructed $Z \to \mu\mu$ decays** : Reject events where there is a second well reconstructed and isolated muon in the event, such that it is oppositely charged from the tag muon, has p_T > 20 GeV, $|\eta|$ < 2.5, and satisfies $|M_{\mu_1\mu}$ - 91.1$|$ < 2 GeV.

- **Guess the Lorentz four-momentum of the probe muon or μ_2** : We use the information from $Z \to \mu\mu$ simulation events to make an educated guess about the four-momentum of the probe muon (outside the tracker coverage). We select a truth level muon with $|i\eta|$ = 27 (2.5 < $|\eta|$ < 2.65), originating from a $Z \to \mu\mu$ decay. The most probable p_T value of such a muon is 45.31 GeV, as shown in Fig. A.3 (left), and we use this value as the guess p_T value of a probe muon iin $|i\eta|$ = 27 (similar procedure for $|i\eta|$ = 28/29). We use the average pseudorapidity of 2.575 for $|i\eta|$ = 27, and since both the muons originating from the Z decay are expected to be on the same side of the detector, the sign of the probe muon pseudorapidity is taken to be the same as that of the tag muon. The two muons from the Z decay are usually produced back-to-back, and thus we require the $\phi_{\mu_2} = \phi_{\mu_1} \pm \pi$. For example, if a lorentz-vector is denoted as (p_T, η, ϕ, M) and for μ_1 it is known to be (from reconstruction): (85.3 GeV, -2.3, 1.57, 0.105 GeV), then the guess Lorentz-vector for μ_2 in $|i\eta|$ = 27 is : (45.31 GeV, -2.575, -1.57, 0.105 GeV).

- We also require the energy deposited in the ECAL in the $i\eta$ tower of the probe muon to be < 500 MeV.

- Using the tag muon and the guess probe muon four momenta, we can reconstruct an invariant mass (just a discriminator for background rejection) of

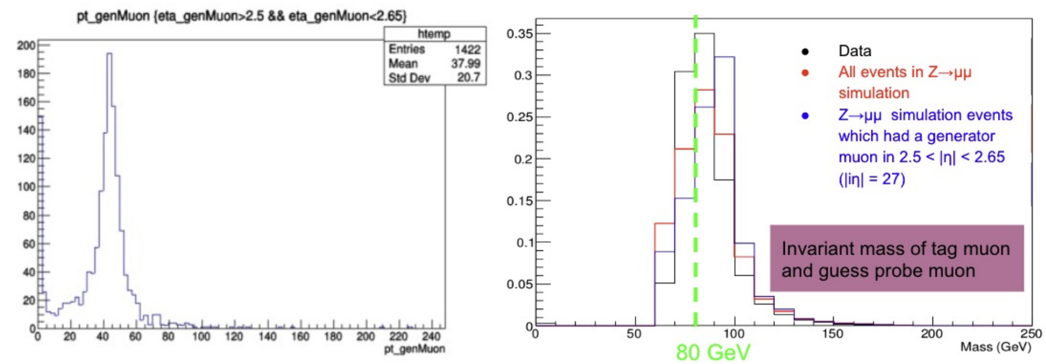


Figure A.3: Reconstructing the probe muon. (Left) The truth level p_{T} for a muon, originating from a $Z \to \mu\mu$ decay, and located in $2.5 < |i\eta| < 2.65$ ($|i\eta| = 27$). The most probable p_{T} value of such a muon is 45.31 GeV. (Right) Distribution of the $M_{\mu_1\mu_2}$ discriminant for events passing the tag and probe muon selections. The data is shown in black, $Z \to \mu\mu$ MC simulation which includes both signal and background events is shown in red, and the $Z \to \mu\mu$ MC simulation events with a truth level muon with $2.5 < |\eta| < 2.65$ ($|i\eta| = 27$) is shown in blue and includes only the signal events. All histograms were normalized to unity to compare the shapes.

the muon pair as $M_{\mu_1\mu_2}$. We require $M_{\mu_1\mu_2} > 80$ GeV, as shown in Fig. A.3 (Right).

- **Energy deposit per depth** : The energy deposited in different depths has the following requirements (based on $Z \to \mu\mu$ simulation events with a probe muon in $|i\eta| = 27/28/29$): Depth 1 , Depth 6 and Depth 7 energy > 0. GeV; Depth 2 energy > 0.05 GeV; Depth 3 energy < 0.1 GeV, Depth 4 and Depth 5 energy < 0.8 GeV.

A.3 Validation checks in $|i\eta| = 25$

We will first check the signal efficiency and background contamination resulting from our event selections in a suitable control region. To do this, we select a control region with the same selections as described in the previous section, except we will now require the probe muon to be in $|i\eta| = 25$, where there is tracker coverage.

A.3.1 Signal efficiency

We look at events where there is a real reconstructed second muon in $|i\eta| = 25$, and compare the energy distributions per depth in both data and $Z \to \mu\mu$ simulation. An example is shown in Fig. A.4, which compares the energy profiles in Depth 6. The signal is modeled using a convolution of Gaussian and Landau distributions (in

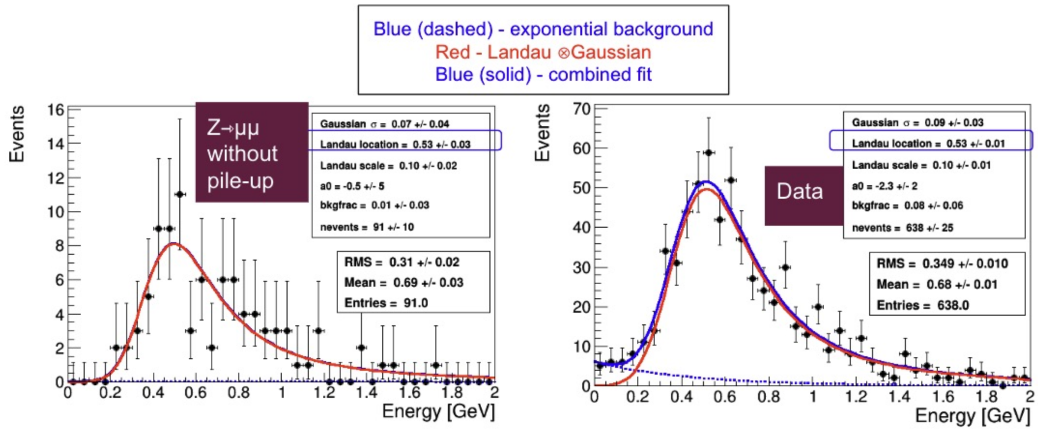


Figure A.4: Comparison of deposited energy profiles in Depth 6 for $|i\eta| = 25$ (for both HE detectors combined), in events with a real reconstructed second muon in that $i\eta$ tower. The left distribution is for $Z \to \mu\mu$ MC (no pileup) simulation events and the right distribution is data. The signal is modeled using a convolution of Gaussian and Landau distributions (in red), while the background is modeled by an exponential function (dashed blue). The combined signal + background fit is shown in solid blue. The most probable value of deposited energy is similar in both the simulation and data.

red), while the background is modeled by an exponential function (dashed blue). The combined signal + background fit is shown in solid blue. The most probable value of deposited energy is similar in both the simulation and data (~0.53 MeV). We can derive an "Energy/layer MPV" (= 0.53/4 = ~0.13 MeV for Fig. A.4) based on these fits, and Table A.1 summarizes these values for different depths in $|i\eta| = 25$.

Table A.1: The average energy/layer MPV value for $|i\eta| = 25$ (for both HE detectors combined), in events with a real reconstructed second muon in that $i\eta$ tower.

Depth	$Z \to \mu\mu$ (no pileup) simulation	Data
4	0.13 ± 0.01	0.13 ± 0.01
5	0.13 ± 0.03	0.15 ± 0.01
6	0.13 ± 0.02	0.14 ± 0.01

The observed difference between data and simulation in the average energy/layer MPV value is small (< 14%).

A.3.2 Background contamination

We also look at events where there are no reconstructed second muons in $|i\eta| = 25$, and compare the energy distributions per depth in both data and single muon

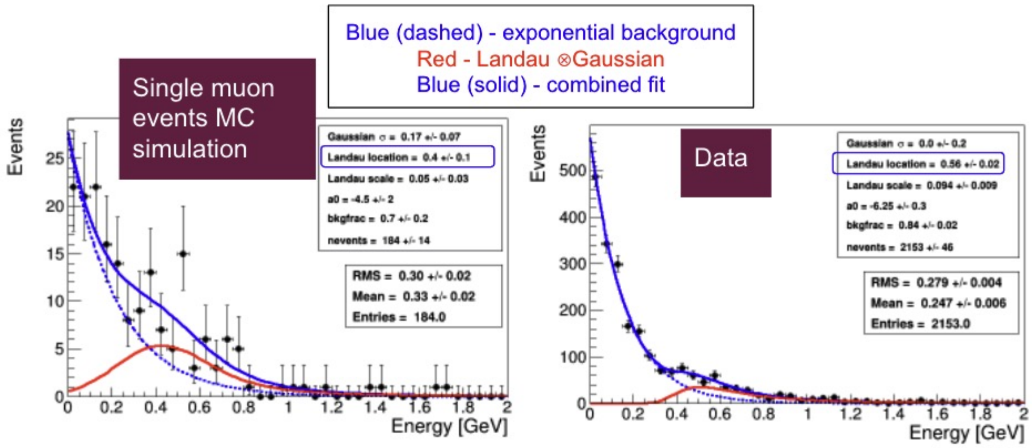


Figure A.5: Comparison of deposited energy profiles in Depth 6 for $|i\eta| = 25$ (for both HE detectors combined), in events with no reconstructed second muon in that $i\eta$ tower. The left distribution is for single muon (background) simulation events and the right distribution is data. The signal is modeled using a convolution of Gaussian and Landau distributions (in red), while the background is modeled by an exponential function (dashed blue). The combined signal + background fit is shown in solid blue. There is a small MIP like peak visible in both the data and the simulation, indicating background sculpting.

MC (background) simulation. An example is shown in Fig. A.5, which compares the energy profiles in Depth 6. The signal is modeled using a convolution of Gaussian and Landau distributions (in red), while the background is modeled by an exponential function (dashed blue). The combined signal + background fit is shown in solid blue. There is a small MIP like peak visible in both the data and the simulation, indicating background sculpting due to our depth energy selections. The most probable value of deposited energy the simulation and data is ~0.4 MeV and ~0.56 MeV, respectively.

A.4 Conclusions

Our current event selections have a signal purity of 60–70% and tend to introduce a visible MIP-like background sculpting in the energy/layer distribution due to our depth energy selections. One possible way to reduce the background contamination is to veto events when they contain b-jets, photons or electrons in the $i\eta$ tower of interest. The invariant mass distribution of the two muons with a real reconstructed second muon in $|i\eta| = 25$, after rejecting events that contain b-jets, photons or electrons in $|i\eta| = 25$, is shown in Fig. A.6. The data agrees very well with the simulation prediction, and the $S/(S + B)$ value is found to be 1/10. This indicates

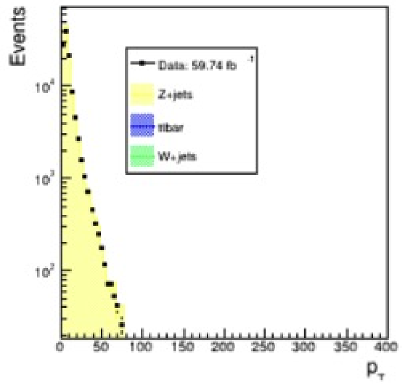


Figure A.6: The invariant mass distribution of the two muons with a real reconstructed second muon in $|i\eta| = 25$, after rejecting events that contain b-jets, photons or electrons in $|i\eta| = 25$. The data agrees very well with the simulation prediction.

that there is promise in moving forward with this strategy. Due to lack of dedicated background simulation samples (b-jets, photons or electrons in the high $i\eta$ regions), we could not finish this study for Run 2 (2018 pp collision data). However, this strategy can be used in Run 3 to calibrate the HCAL endcaps.

Appendix B

COMMISSIONING TESTS FOR PROTOTYPE BTL READOUT ELECTRONICS

This appendix discusses the commissioning tests performed on the two preliminary versions of the BTL prototype readout electronics, i.e., the TOFPET and the TOFHIR2A ASIC.

B.1 TOFPET2

The TOFPET2 (Time-of-Flight Positron Emission Tomography) ASIC [232] was the first version of the prototype readout electronics chip. It is a 64 channel ASIC, with two outputs per channel: the time and energy of a pulse. It operates based on three thresholds: a low threshold `th1` for assigning a time stamp to the input pulse, a high threshold `th2` for rejecting noise and starting the charge integration, and `th3` threshold for event triggering and finishing the charge integration. The three thresholds are shown in Fig. B.1. The TOFPET 2 was designed with UMC 110 nm technology, and can handle the signal rate expected at the BTL, but was primarily designed for PET applications and not HEP. The final readout chip (TOFHIR2), needs to be more radiation tolerant and include a noise cancellation filter, which were missing in the TOFPET2 ASIC.

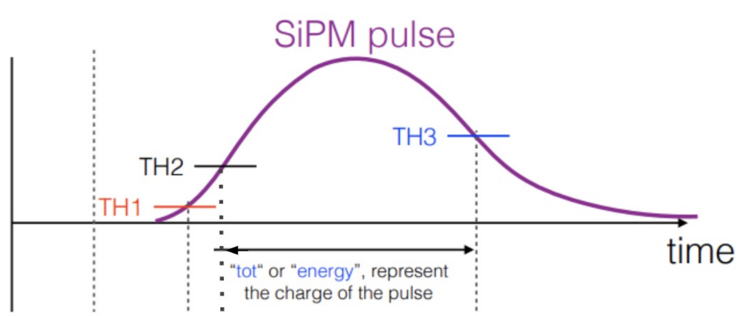


Figure B.1: The discriminator thresholds of the TOFPET 2 ASIC.

B.1.1 CPT Lab tests

To perform commissioning tests, we used an array of 16 LYSO crystal bars of dimensions 3x3x57 mm^3. Each bar had two SiPMS air gap coupled at the two ends.

The SiPM active area was 3 x 3 mm^2. The module was connected to a SiPM adapter board, which was in turn connected to the TOFPET Front-End Board (FEB/D) using samtec connectors (Fig. B.2). The SiPMs were powered at their Geiger break down mode, with an over voltage (OV) of 7V above breakdown voltage (V_{br}).

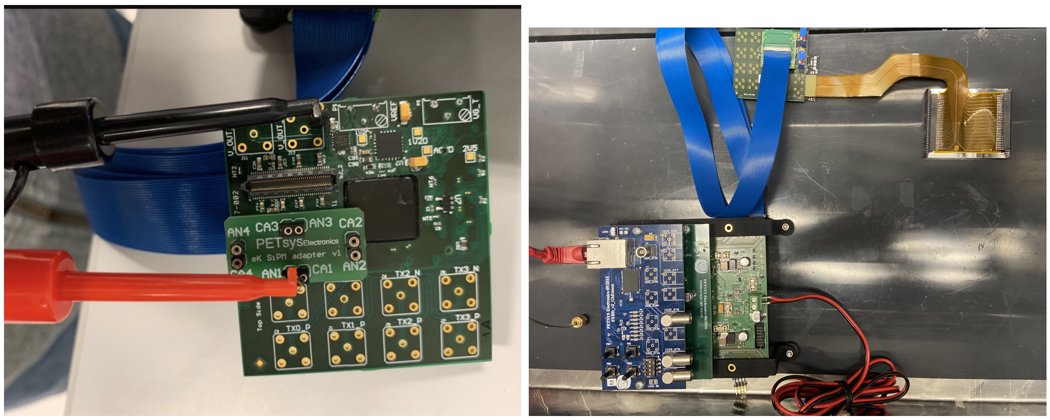


Figure B.2: The readout test setup. (Left) The SiPM adapter board. (Right)The TOFPET FEB/D is connected to the LYSO bar + SiPM array through an adapter board.

We measured the energy spectrum of Na^{22} with this setup. The unit of energy read by the TOFPET is in arbitrary ADC units. The 511 keV peak from Na^{22} for each of the 64 channels was fitted individually using a Gaussian, and the fit for one such channel is shown in Fig. B.3. Fig. B.4 shows the peak position and resolution obtained from the Gaussian fit for every channel of the SiPM module. The fitted position and resolution is consistent across all channels, except for a few channels that were known to be broken.

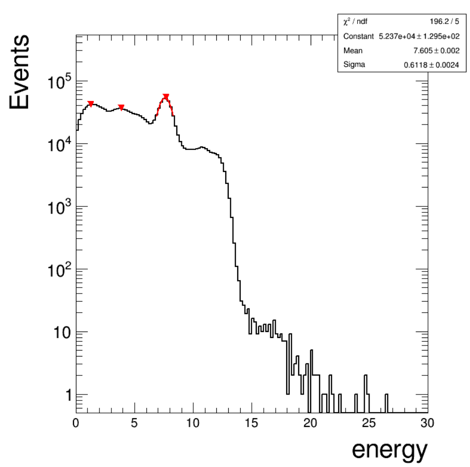


Figure B.3: The energy spectrum of channel 32 and the fit of the 511 keV peak.

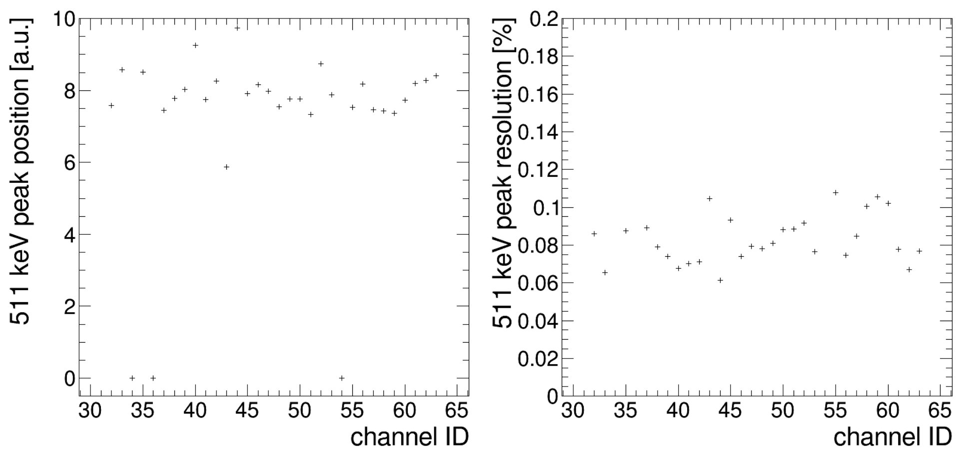


Figure B.4: The fitted position for each channel (Left) and the resolution of each channel (Right).

B.1.2 Fermilab test beam

We also performed commissioning tests on the TOFPET at the Fermilab December 2017 test beam. We used two HPK 3x3 mm^2 SIPMs with 25μm cell pitch coupled to 12x12 mm^2 LYSO tiles using optical grease, and wrapped in Teflon. The two SiPM+LYSO tiles were connected to a SiPM adapter board, which was in turn connected to the TOFPET FEB/D using samtec connectors. A silicon tracker telescope, consisting of twelve strip modules with 60 μm pitch in alternating orientation, was located in front of the crystals and SiPMs under test to determine the impact point of the beam particles in the x-y plane with a precision of about 0.2 mm. The crystals and SiPMs were located inside a dark box placed along the beamline, while the FEB/D was placed outside of the beam line. A Photek 240 Micro Channel Plate-Photomultiplier Tube (MCP-PMT), with a time resolution of 12 ps, was placed downstream of the box and was used to provide a reference time measurement, and also acted as a trigger. Figure B.5 shows the actual test setup.

The MCP signal was fed to the TOFPET and also a NIM readout crate. The NIM readout crate also received input signals from the TOFPET, and additionally provided a feedback MCP trigger to the TOFPET. A schematic view of the Data Acquisition (DAQ) system can be found in Fig. B.6.

The energy spectrum recorded from the two SiPMs showed the MIP peak as well as the intrinsic peaks from the Lutetitum radioactivity (Fig. B.7 left). The time resolution as a function of the SiPM OV was also recorded and is shown in Fig. B.7 (right). Although, the time resolution is of the order of a few 100 ps, it is seen that increasing the OV reduces the measured time resolution, which is expected,

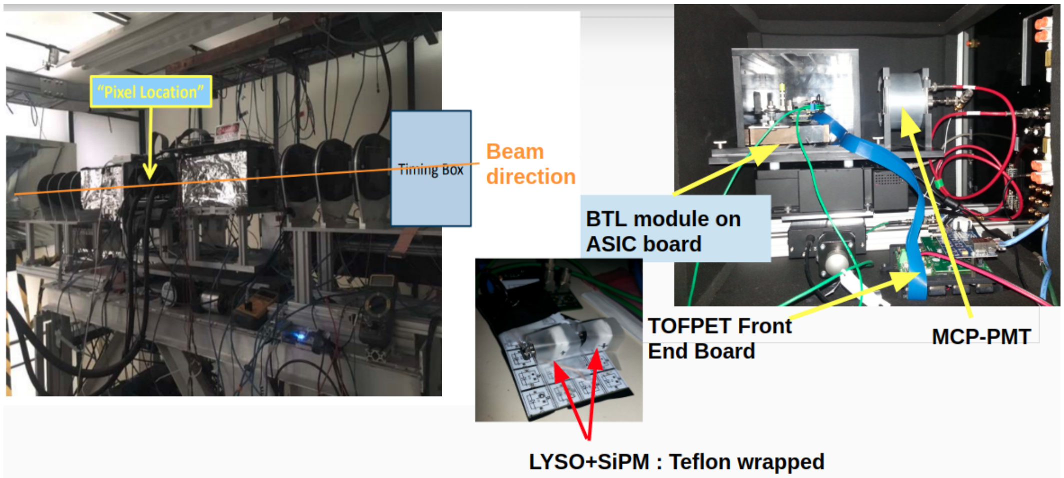


Figure B.5: The experimental setup at the Fermilab December 2017 test beam.

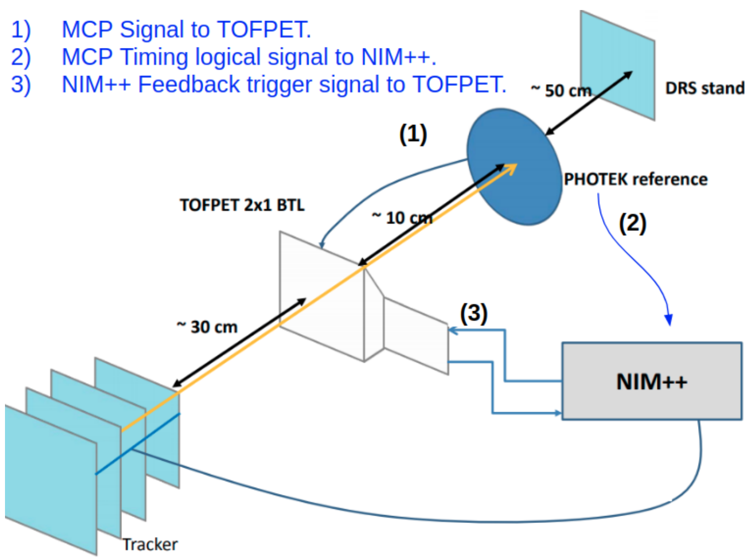


Figure B.6: The DAQ at the Fermilab December 2017 test beam.

since the SiPM PDE increases with increasing over-voltage. The time resolution measured here is dominated by electronics noise of TOFPET, since it was developed for PET applications (~200 ps resolution). This feature will be fixed in the TOFHIR ASIC chip, which will be specifically developed for BTL.

B.2 TOFHIR2A

The TOFHIR2A ASIC [102] is a prototype BTL readout ASIC, developed with CMOS 130nm technology of the TSMC foundry. This ASIC has an improved behavior under radiation as compared to the previous readout prototype, TOFPET (also known as TOFHIR 1). The TOFHIR2A test board (T2TB) has two ASICs

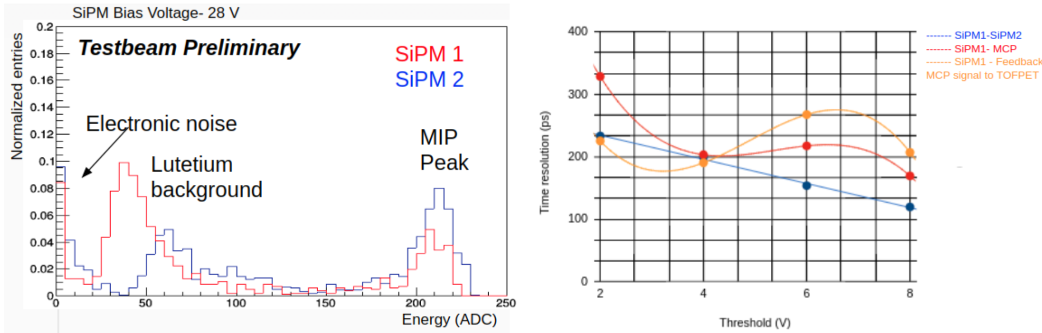


Figure B.7: Results obtained from the TOFPET test setup at the Fermilab December 2017 test beam. (Left) Energy spectrum of the two SiPMs. The intrinsic Lutetium radioactivity from the LYSO tile and the MIP peaks are visible for both the SiPMs. (Right) Time resolution of the SiPMs as a function of the over voltage. The different colors indicate different ways of calculating the time resolution (similar to what was discussed in Sec. 4.3.2.)

(ASIC0 and ASIC1), and each ASIC has 32 channels. The T2TB can be connected to a FEB/D board through an adaptor board. The T2TB should also be connected to the adapter board through an auxiliary cable (2x8 pin IDC cable with 2.54 mm pitch), to provide bias voltage and test pulses to the SiPMs. Figure B.8 shows the TOFHIR2A ASIC.

ASIC0 has an ALDO2 chip which is used to provide bias voltages to the SiPMs, and has flex cable connectors for the BTL module. Half of the channels in ASIC0 are DC coupled and the other half are AC coupled. ASIC1 does not have an ALDO chip, and uses standard samtec connectors to provide bias voltage through the FEB/D. All channels in ASIC 1 are AC coupled. The TOFHIR2A also triggers on SiPM pulses based on three different thresholds, similar to the ones mentioned in Sec. B.1.

To perform commissioning tests, we used an array of 16 LYSO crystal bars of dimensions 3x3x57 mm^3. Each bar had two SiPMS air gap coupled at the two ends. The SiPM active area was 3 x 3 mm^2. We measured the energy spectrum of Na^{22} for various channels of the module with this setup, for different over voltages. The experimental setup is shown in Fig. B.9. The energy spectrum recorded by one of the channels is shown in Fig. B.10. Both the 511 keV and 1.27 MeV peaks are visible in the spectrum indicating that the test setup works well.

The final version of the BTL ASIC, TOFHIR2B, is currently undergoing testing and is in the review process.

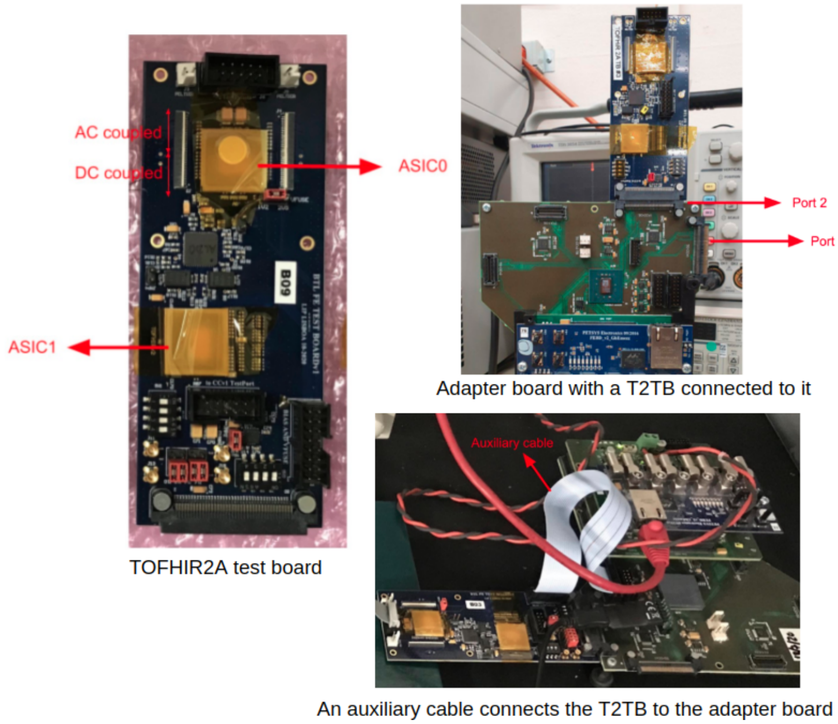


Figure B.8: The TOFHIR2A ASIC.

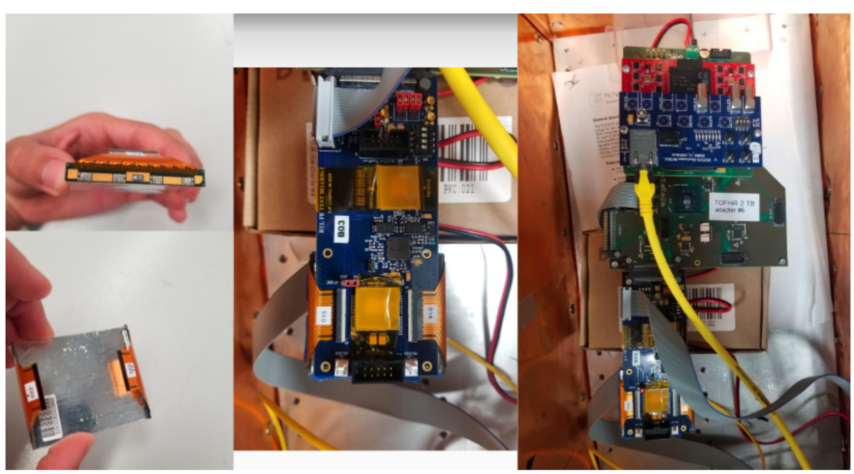


Figure B.9: The experimental setup with the TOFHIR2A.

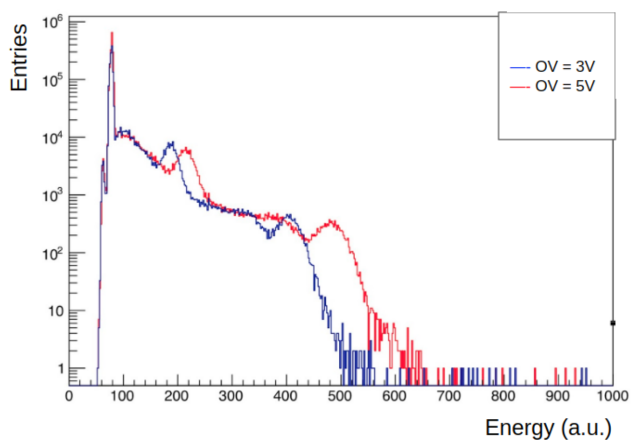


Figure B.10: The energy spectrum recorded in one of the BTL module channels connected to the T2TB tester board. The spectrum shifts to the right when increasing the SiPM over voltage.

Appendix C

A MASS AGNOSTIC NEURAL NETWORK FOR $H \to \mu\mu$

C.1 Introduction

The template-based fit method was chosen for the VBF category (Section. 5.6), unlike the data-driven fit for the ggH, VH and ttH categories, since it produced better expected results. To perform a data-driven fit, however, the most discriminating variable is the $m_{\mu\mu}$, and therefore one must develop a $m_{\mu\mu}$ independent DNN. Before we made the choice to perform a simulation based template fit to the VBF category, we invested efforts in developing a neural network which is agnostic to the dimuon mass in the VBF phase space (to avoid sculpting the $m_{\mu\mu}$ distribution of the background). However, for a simple DNN classifier, it is easy to learn the dimuon mass with the input kinematic variables, even if the mass is not explicitly given as an input to the training. Therefore, we used adversarial training techniques [152] to develop a neural network, where we penalized our classifier loss function based on how well it could learn the dimuon mass. This method is described in more detail in the next few sections.

C.2 DNN architecture

We designed a network where we have two models competing to outperform one another, one is the classifier and the other one is the adversarial. The classifier (C) takes as input several high level features over which it can train and give an output prediction probability of an event being signal or background like. The adversarial regression (A) then takes as input, the prediction of the classifier and tries to estimate the dimuon mass. Based on how close the prediction of the adversarial is to the true dimuon mass, the classifier gets penalized in its loss function ($\mathcal{L}_C$). In other words, if A can accurately predict the dimuon mass from C's output, then C should be heavily penalized, since we want to build a classifier that is completely blind to the dimuon mass.

70% of the events were used for training and validation, and the remaining 30% were used for testing. The signal weights were then modified such that the sum of weights

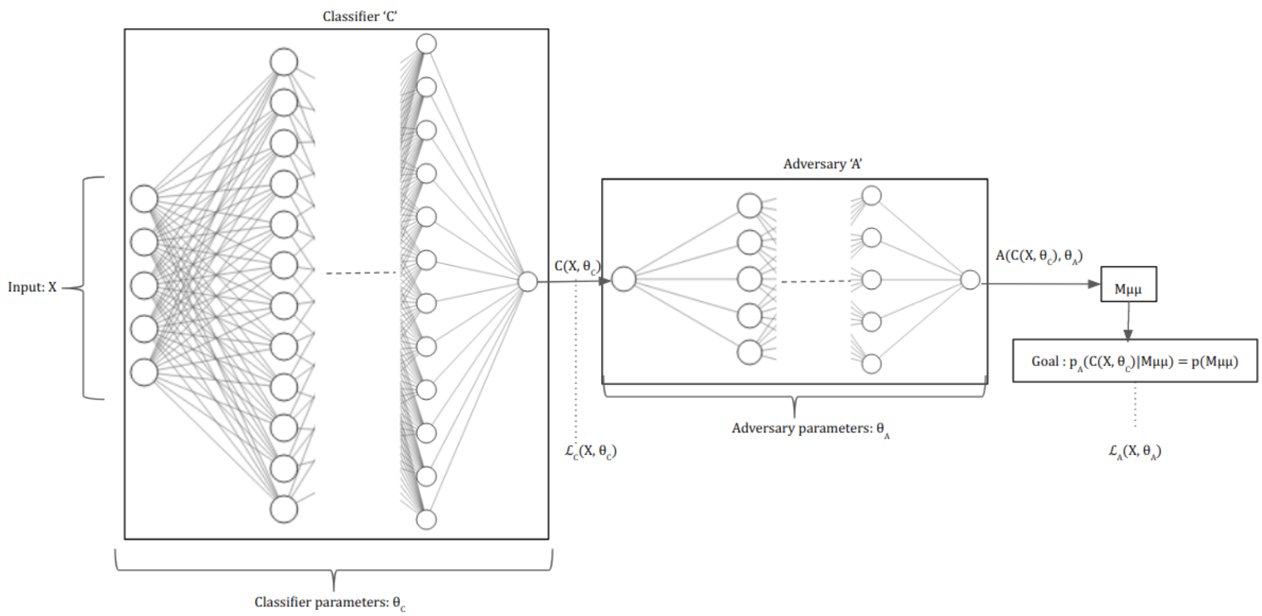


Figure C.1: Adversarial network.

of signal events was equal to the sum of weights of background events. Additionally, all input features to the DNN were standardized to have a mean of 0 and standard deviation of 1. The training was done on events with $m_{\mu\mu} \in [115, 135]$ GeV.

The classifier network was trained with 5 hidden layers and 100 nodes per layer. The Adam optimizer was used with a learning rate of 10^{-5}. The tanh activation function was used for the inner layers and the `sigmoid` activation was used for the final output. The loss function used was the `binary cross-entropy`. A 10% drop-out rate and batch-normalization were also used to regularize the training. The adversarial regression network was trained with 3 hidden layers and 50 nodes per layer. The Adam optimizer was used with a learning rate of 10^{-5} and the loss function ($\mathcal{L}_{\mathcal{A}}$) used was the *mean squared error (MSE)*. A 10% drop-out rate and batch-normalization were also used to regularize the training. The combined loss of function of the network was $\mathcal{L} = \mathcal{L}_C - \alpha\mathcal{L}_{\mathcal{A}}$, where α is a tunable parameter denoting the strength of the adversarial training. We tested several different values of α. Ultimately, one needs to achieve a balance between reducing the background sculpting and finding the optimal classifier performance. We found that $\alpha = 10^{-4}$ gave us the desired performance. A visual representation of this network is given in Fig. C.1

A preliminary adversarial network was trained using simulated samples for the three years all mixed together. The simulated samples used in the training are:

- Signal: VBF $H \to \mu\mu$ with m_H=125.0 GeV
- Background: Drell-Yan Z.

The following 28 input variables were used to train the model (for description, see Section 5.6.1):

- $\eta(\mu\mu)$, $p_T(\mu\mu)$ of the dimuon pair
- ΔR, $\Delta\phi$ and $\Delta\eta$ between the two leading muons
- $m(jj)$, η, ϕ and p_T of the leading dijet pair
- $\Delta\eta(jj)$ and $\Delta\phi(jj)$
- Mass, η, ϕ and p_T of the leading dijet+dimuon pair
- z^*
- Minimum ΔR between a muon and jet
- Minimum ΔR between dimuon and jet
- Maximum ΔR between a muon and jet
- Maximum ΔR between dimuon and jet
- p_T, η and QGL for the two leading jets
- $\cos\theta_{\mathrm{CS}}$
- N_5^{soft}

C.3 Results

We compared two sets of neural networks, a simple classifier (trained with the parameters described for the classifier part of the full classifier + adversarial network in Section. C.2) and the adversarial network. Fig. C.2 shows the output score distributions for the two networks and Fig. C.3 shows the background DY distributions for $m_{\mu\mu}$ after imposing different cuts on the output scores of the two networks. As is evident, the simple classifier can outperform the adversarial network in terms of a better signal and background separation, but ends up sculpting the background dimuon mass by quite a bit.

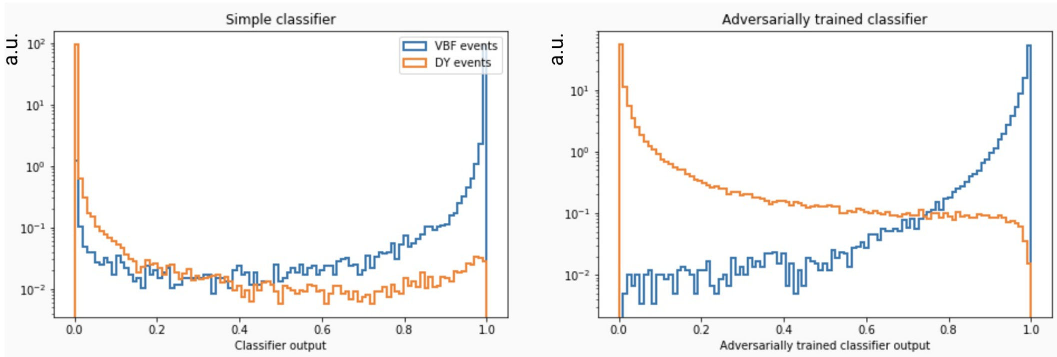


Figure C.2: The output score of the network. (Left) Simple classifier and (Right) Adversarial network. The simple classifier has a better separation power.

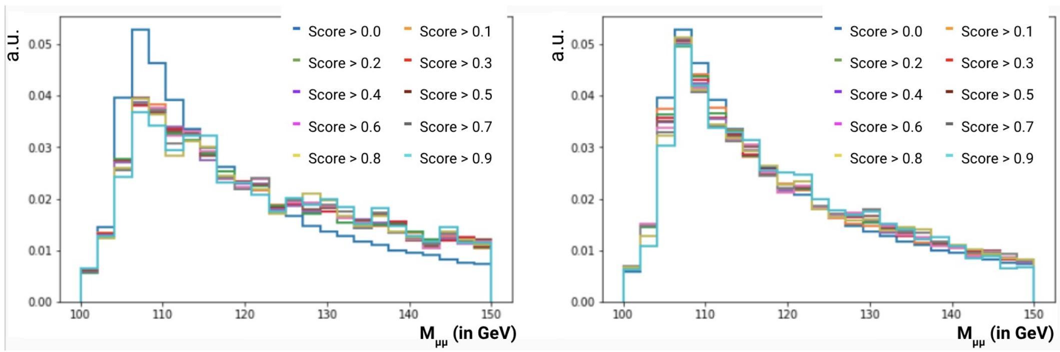


Figure C.3: The $m_{\mu\mu}$ distributions for the DY background for different cuts on the output score of the network. (Left) Simple classifier and (Right) Adversarial network. The simple classifier tends to sculpt the background $m_{\mu\mu}$ mass increasingly with higher selection cuts on the output score of the classifier. The adversarial network does not have this issue since it is specifically trained to be mass agnostic.

C.3.1 Signal extraction

The signal is extracted by fitting the dimuon mass spectrum distributions via analytical functions, in the region where the adversarial network score is greater than 0.5. The signal events are mainly produced via VBF and ggH, and are modelled with a sum of three Gaussian functions (see Fig. C.4). The signal shape parameters are extracted from the simulation and are fixed in the fit to the data.

The background (only Drell-Yan was considered here) is modelled with an exponential modified Breit-Wigner function (BWZ). The BWZ function is defined as:

$$BWZ(m_{\mu\mu}; \Gamma_Z, m_Z, a_1, a_2, a_3) = \frac{e^{a_2 m_{\mu\mu} + a_3 m_{\mu\mu}^2}}{(m_{\mu\mu} - m_Z)^{a_1} + (\Gamma_Z/2)^{a_1}}, \tag{C.1}$$

where m_Z and Γ_Z are the mass and width of the Z boson, and a_i are free-parameters of the fit. The post-fit function is used to generate pseudo-datasets (see Fig. C.5).

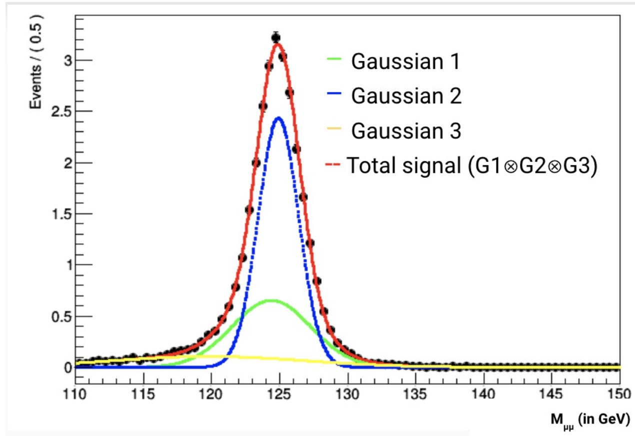


Figure C.4: The ggH + VBF signal is modelled with a sum of 3 Gaussian functions.

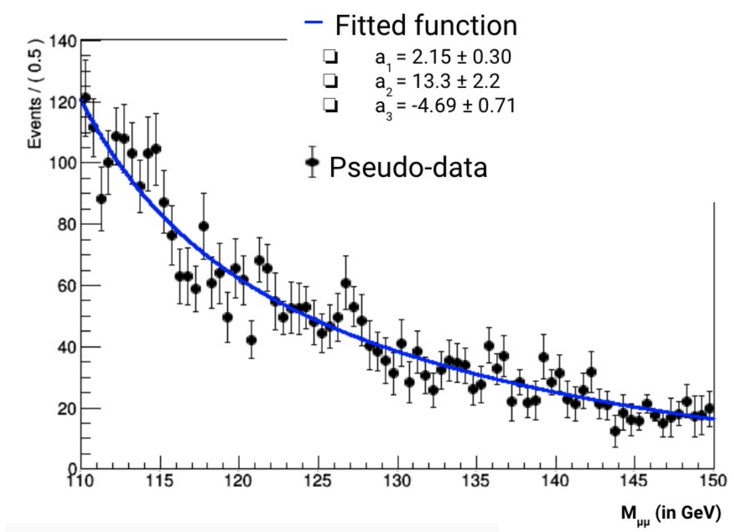


Figure C.5: Pseudo-data (black points) is generated from an analytical fit to the DY background (blue solid line).

The pseudo-dataset is then injected with a signal strength of 1. The expected signal strength is then extracted with a signal plus background fit to the pseudo-dataset to get a measure of the sensitivity of the adversarial network. The expected 95% C.L. upper limit on the signal strength extracted from this fit is 2.3× SM.

Appendix D

HEP DATA PROCESSING WITH GPU-ACCELERATED KERNELS

D.1 Introduction

At the general-purpose experiments of the Large Hadron Collider such as CMS or ATLAS, the data formats typically consist of columns of physics related variables or features for the recorded particles such as electrons, muons, jets and photons for each event in billions of rows. In addition to the columns of purely kinematic information, each particle carries a number of features that describe the reconstruction details and other high-level properties of the reconstructed particles. For example, for muons, we might record the number of tracker layers where an associated hit was found. A typical event for such reduced data formats are on the order of a few kilobytes per event.

A typical physics analysis at the LHC requires tens to hundreds of iterations over such datasets. For each iteration of the analysis, hundreds of batch jobs of custom reduction software is run over these data. By reducing the complexity and increasing the speed of these workflows, HEP experiments could deliver results from large datasets with faster turn-around times.

Recently, heterogeneous and highly parallel computing architectures beyond consumer x86 processors such as GPUs, TPUs and FPGAs have become increasingly prevalent in scientific computing. We investigate the use of array-based computational kernels that are well-suited for parallel architectures for carrying out the final data analysis in HEP. Although we use two simple analyses as an example, the approach based on array computing with accelerated kernels is generic and can be used for other collider analyses. The purpose of this report is to document the efforts of processing terabyte-scale data in HEP fast and efficiently using the example kernels in the hepaccelerate [233] library.

In section D.2, we describe the structure of the data and discuss how data sparsity is handled efficiently. We introduce physics-specific computational kernels in section D.3 and describe the measured performance under a variety of conditions in section D.4. Finally, we conclude with a summary and outlook in section D.5.

D.2 Data structure

We can represent HEP collider data in the form of two-dimensional matrices, where the rows correspond to events and columns correspond to features in the event. Due to the random nature of the underlying physics processes that produce a varying number of particles per event, the standard HEP software framework based on ROOT includes mechanisms for representing dynamically-sized arrays as well as complete C++ classes with arbitrary structure as the feature columns, along with a mechanism for serializing and deserializing these dynamic arrays [234].

Based on the approach first introduced in the `uproot` [235] and `awkward-array` [236] python libraries, HEP data files with a varying number of particles per event can be represented and efficiently loaded as sparse arrays with an underlying one-dimensional array for a single feature. Event boundaries are encoded in an offset array that records the particle count per event. Therefore, the full structure of N events, M particles in total, can be represented by a contiguous offset array of size $N + 1$ and a contiguous data array of size M for each particle feature. We illustrate this sparse or jagged data structure on Fig. D.1.

In practice, analysis-level HEP event data are stored in compressed files of raw columnar features in a so-called "flat analysis ntuple" form, implemented via the ROOT library. Out of hundreds of stored features, a typical analysis might use approximately 50 to 100, discarding the rest and only using them rarely for certain auxiliary calibration purposes. When the same features are accessed multiple times, the cost of decompressing the data can be significant. There is a tradeoff between disk usage, CPU time spent for decompression and CPU time spent for physics analysis. In the following benchmarks, we create an analysis-specific cache of the necessary columns from the full input using LZMA compression, such that the overhead from decompression would be minimal. Despite the reduced compression, the resulting "skimmed" cache is typically smaller than the input, due to keeping only analysis-specific rows and columns. The choice of compression algorithms should be addressed further in optimizing the file formats for cold storage and analysis [237], as well as using efficient, vectorizable compression algorithms [238].

D.3 Computational kernels

In the context of this report, a kernel is a function that is evaluated on the elements of an array to transform the underlying data, for example, compute the square root of all the values in the array. When the individual kernel calls across the data are

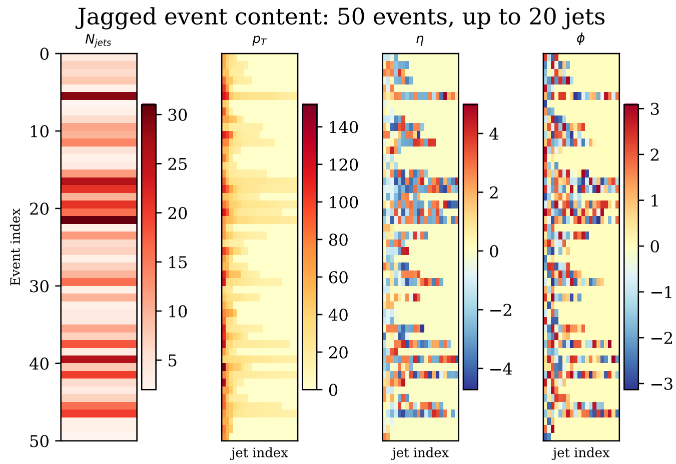


Figure D.1: A visual representation of the jagged data structure of the jet p_T, η and ϕ content in 50 simulated events. On the leftmost figure, we show the number of jets per event, one event per row, derived from the offset array. In the three rightmost columns, we show the jet content in events, visualizing the p_T, η and ϕ of the first 20 jets for each event.

independent of each other, they can be evaluated in parallel over the data using single-instruction, multiple-data (SIMD) processors.

We note that the columnar data analysis approach based on single-threaded kernels is already recognized in HEP [239]. Our contribution is to further extend the computational efficiency and scalability of the kernels to parallel hardware such as multi-threaded CPUs and propose a GPU implementation.

A prototypical HEP-specific kernel would be to find the scalar sum $H_T = \sum_{i \in \text{event}} p_T^i$ of all particles passing some quality criteria in an event. We show the Python implementation for this on Listing D.1. This kernel takes as input the M-element data array of all particle transverse momenta `pt_data` and an $N+1$-element array of the event offsets. In addition, as we wish to include only selected particles in selected events in this analysis, we use an N-element boolean mask for events and M-element boolean mask for the particles that have passed selection. These masks can be propagated to further functions, making it possible to efficiently chain computations without resorting to data copying. Finally, the output is stored in a pre-allocated array of size N.

Other generic kernels that turn out to be useful are related to finding the minimum or maximum value within the offsets (for example, finding the jet with the highest p_T in an event) or retrieving or setting the m-th value of an array within the event offsets. Several such kernels can already be found in [233].

```
1  def sum_ht(
2    pt_data, offsets,
3    mask_rows, mask_content,
4    out):
5
6    N = len(offsets) - 1
7    M = len(pt_data)
8
9    #loop over events in parallel
10   for iev in prange(N):
11     if not mask_rows[iev]:
12       continue
13
14     #indices of the particles in this event
15     i0 = offsets[iev]
16     i1 = offsets[iev + 1]
17
18     #loop over particles in this event
19     for ielem in range(i0, i1):
20       if mask_content[ielem]:
21         out[iev] += pt_data[ielem]
```

Listing D.1: Python code for the kernel computing the scalar sum of selected particle momenta H_T. The inputs are `pt_data`, an M-element array of p_T data for all the particles, the $N+1$-element `offsets` array with the indices between the events in the particle collections, as well masks for events and particles that should be considered. On line 10, the kernel is executed in parallel over the events using the Numba `prange` iterator, which creates multithreaded code across the loop iterations. On line 19, the particles in the event are iterated over sequentially.

These kernels have been implemented in Python and just-in-time compiled to either multithreaded CPU code using the Numba package [240] or using CUDA [241] for GPUs. We have chosen Python and Numba to implement the kernels in the spirit of quickly prototyping this idea, but this approach is not restricted to a particular programming environment.

D.4 Analysis benchmark

CMS Open Data

We benchmark this approach in a prototypical top quark pair analysis using CMS Open Data from 2012 [242]. The datasets are processed from the experiment-specific format to a columnar format using a publicly-available tool [243]. We stress that the output of this tool is not meant to be used for deriving physics results, but

```
1 @cuda.jit
2 def sum_ht_cudakernel(
3   pt_data, offsets,
4   mask_rows, mask_content,
5   out):
6
7     xi = cuda.grid(1)
8     xstride = cuda.gridsize(1)
9     for iev in range(xi, offsets.shape[0]-1, xstride):
10         if mask_rows[iev]:
11             start = np.uint64(offsets[iev])
12             end = np.uint64(offsets[iev + 1])
13
14             #loop over particles in this event
15             for ielem in range(start, end):
16               if mask_content[ielem]:
17                 out[iev] += pt_data[ielem]
```

Listing D.2: Python code for the kernel computing the scalar sum of selected particle momenta H_T. The inputs are `pt_data`, an array of p_T data for all the particles, the `offsets` array with the indices between the events in the particle collections, as well masks for events and particles that should be considered. On line 9, the kernel is executed in a GPU compatible style which is parallel over the events using CUDA. On line 15, the particles in the event are iterated over sequentially.

rather to replicate the computing conditions that are encountered in the data analysis groups at the experiments. The resulting derived datasets, corresponding to about 60GB of simulation and 40GB of data with the single muon trigger, are further processed in our benchmark analysis.

The benchmark analysis implements the following features in a single end-to-end pass:

- event selection based on event variables: trigger bit, missing transverse energy selection
- object selection based on cuts on objects: jet, lepton selection based on p_T, η
- matching of pairs of objects: jet-lepton ΔR matching, jet to generator jet matching based on index
- event weight computation based on histogram lookups: pileup, lepton efficiency corrections

- jet energy corrections based on histogram lookups: *in-situ* jet energy systematics, increasing the computational complexity of the analysis by about 40x
- high-level variable reconstruction: top quark candidate from jet triplet with invariant mass closest to $M = 172$ GeV
- evaluation of $\simeq 40$ typical high-level inputs to a deep neural network (DNN)
- Multilayer, feedforward DNN evaluation using `tensorflow`
- saving all DNN inputs and outputs, along with systematic variations, to histograms ($\simeq 1000$ individual histograms).

We perform two benchmark scenarios of this analysis: one with a partial set of systematic uncertainties to emulate a simpler IO-limited analysis, and one with the full set of systematic uncertainties to test a compute-bound workload. The timing results from the benchmark are reported on figure D.2. In the former case, we observe the GPU-accelerated version performing about 12x faster than a single multi-threaded CPU. For the latter case, where the main workload is repeated around 40x, a single GPU is about 15x faster than 1 multi-threaded CPU.

We have also found that it is beneficial to ensure that the kernels are called on sufficiently large datasets to reduce the overhead of kernel scheduling with respect to the time spent in the computation.

Although the approximately 10-15x performance improvement on the GPU with respect to a single CPU thread is relatively modest, it is promising to see that most of the physics analysis methods can be implemented with relative ease on a GPU, such that with further optimizations, a small number multi-GPU machines might be a viable alternative to a large server farm in the future. Ultimately, the choice of the computing platform will be driven by availability and pricing of resources. We stress that we do not claim a GPU is necessarily faster than a large number of CPU threads, but rather demonstrate that it is possible to implement a portable end-to-end analysis on various backends by using relatively standard tools.

ATLAS Open Data

To demonstrate the capability of this library to run on data formats from different collider experiments, we tried to reproduce the classic $H \longrightarrow ZZ \longrightarrow 4l$ with the 13

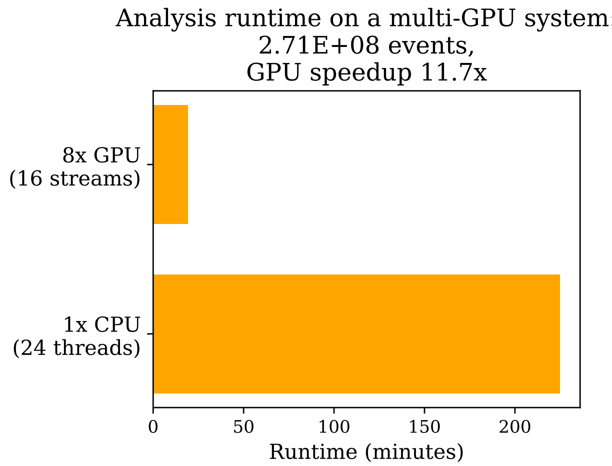


Figure D.2: Benchmarks of the full analysis with 270M events, 100GB of numerical data on a multi-GPU system. We find that by using 8x nVidia GTX 1080 GPUs, 2 compute streams per device, we can reduce the analysis runtime by a factor of 12x, compared to using multiple threads on the CPU alone.

TeV ATLAS Open Data [244]. The benchmark analysis implements the following features in a single end-to-end pass:

- object selection based on lepton variables: p_T, η, charge
- event weight computation based lepton efficiency corrections, trigger weights
- high-level variable reconstruction: 4 lepton invariant mass closest to $M = 125$ GeV.

The invariant 4 lepton mass is shown in Fig. D.3.

D.5 Summary and outlook

We demonstrate that it is possible to carry out prototypical end-to-end high-energy physics data analysis from a relatively simple code in a HPC context. This is achieved with efficient input data preparation, array processing approaches and a small number of specialized kernels for jagged arrays implemented in Python using the Numba package, compiled to native machine code with LLVM. It is also possible to offload parts of these array computations to accelerators such as GPUs, which are highly efficient at parallel processing and alleviate compute-bound analyses such as those with many systematic variations or DNN evaluations. We demonstrate a prototypical top quark pair analysis implementation using these computational

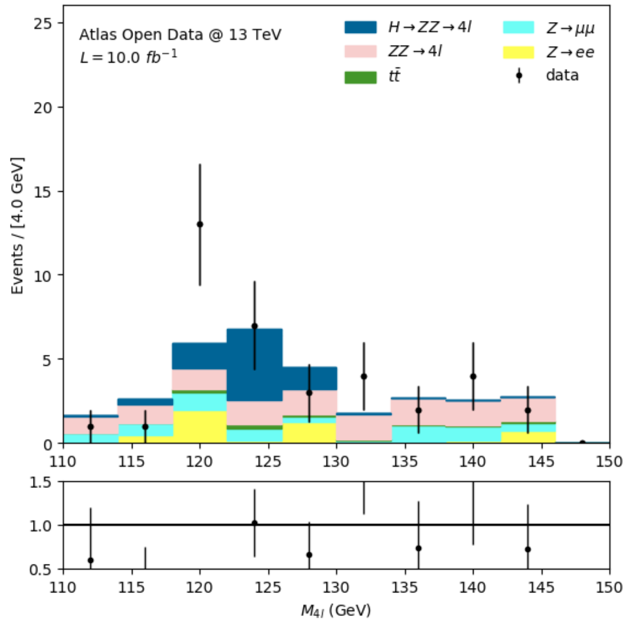


Figure D.3: The 4 lepton invariant mass peaks near 125 GeV.

kernels with optional GPU offloading that can run an order of magnitude faster on a multi-GPU machine, compared to using a multithreaded CPU approach alone.

Several improvements are possible, among them optimizing data access with direct-memory access IO, using accelerators for data decompression, scaling horizontally across multiple machines, scaling vertically in a single server by optimizing the threading performance as well as kernel scheduling and fusion on heterogeneous architectures. Setting up integrated "analysis facilities" at HPC centers, with relevant datasets available on a high-performance filesystem coupled with multithreaded processors and GPUs would allow high-energy physics experiments to iterate through analysis data without the typical network overheads from distributed computing. We hope that this report sparks further discussion and development of efficient and easy-to-use analysis tools which would be useful for scientists involved in HEP data analysis and in other data intensive fields.

Bibliography

[1] History of subatomic physics - Wikipedia. URL https://en.wikipedia.org/wiki/History_of_subatomic_physics. Site last accessed in June 2022.

[2] Physics at the British Association. *Nature*, 106:357–358, 1920. URL https://doi.org/10.1038/106357a0.

[3] Guy Hartcup and T.E. Allibone. *Cockcroft and the atom*. Bristol [Avon]: A. Hilger, 1984. URL https://www.worldcat.org/oclc/12666364.

[4] James Chadwick. The existence of a neutron. *Proc. R. Soc. Lond. A*, 136: 692–708, 1932. URL https://doi.org/10.1098/rspa.1932.0112.

[5] Carl D. Anderson. The positive electron. *Phys. Rev.*, 43:491–494, Mar 1933. doi: 10.1103/PhysRev.43.491. URL https://link.aps.org/doi/10.1103/PhysRev.43.491.

[6] J. C. Street and E. C. Stevenson. New evidence for the existence of a particle of mass intermediate between the proton and electron. *Phys. Rev.*, 52:1003–1004, Nov 1937. doi: 10.1103/PhysRev.52.1003. URL https://link.aps.org/doi/10.1103/PhysRev.52.1003.

[7] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961. ISSN 0029-5582. doi: https://doi.org/10.1016/0029-5582(61)90469-2. URL https://www.sciencedirect.com/science/article/pii/0029558261904692.

[8] Sidney A. Bludman. The First gauge theory of weak interactions and the prediction of weak neutral currents. In *3rd International Symposium on the History of Particle Physics: The Rise of the Standard Model*, pages 403–411, 6 1992.

[9] Murray Gell-Mann. The eightfold way: A theory of strong interaction symmetry. doi: 10.2172/4008239. URL https://www.osti.gov/biblio/4008239. Technical Report.

[10] Yuval Ne'eman. Derivation of strong interactions from a gauge invariance. *Nuclear Physics*, 26(2):222–229, 1961. ISSN 0029-5582. doi: https://doi.org/10.1016/0029-5582(61)90134-1. URL https://www.sciencedirect.com/science/article/pii/0029558261901341.

[11] R. A. Fisher. On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society*, 85:87, 1922. ISSN 09528385. URL http://www.jstor.org/stable/2340521.

[12] Abdus Salam and J.C. Ward. Electromagnetic and weak interactions. *Physics Letters*, 13(2):168–171, 1964. ISSN 0031-9163. doi: https://doi.org/10.

1016/0031-9163(64)90711-5. URL https://www.sciencedirect.com/science/article/pii/0031916364907115.

[13] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. doi: 10.1103/PhysRevLett.19.1264. URL https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[14] Peter W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132–133, 1964. ISSN 0031-9163. doi: https://doi.org/10.1016/0031-9163(64)91136-9. URL https://www.sciencedirect.com/science/article/pii/0031916364911369.

[15] CDF Collaboration. Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr 1995. doi: 10.1103/PhysRevLett.74.2626. URL https://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[16] D0 Collaboration. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, Apr 1995. doi: 10.1103/PhysRevLett.74.2632. URL https://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[17] Lyndon Evans and Philip Bryant. LHC machine. *JINST*, 3(08):S08001–S08001, Aug 2008. doi: 10.1088/1748-0221/3/08/s08001. URL https://doi.org/10.1088/1748-0221/3/08/s08001.

[18] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020.

[19] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021.

[20] Abdus Salam. *Weak and electromagnetic interactions*, pages 244–254. doi: 10.1142/9789812795915_0034. URL https://www.worldscientific.com/doi/abs/10.1142/9789812795915_0034.

[21] Gautam Bhattacharyya and Dipankar Das. Scalar sector of two-Higgs-doublet models: A minireview. *Pramana*, 87(3):40, 2016. doi: 10.1007/s12043-016-1252-4.

[22] Abdel Pérez-Lorenzana. An introduction to extra dimensions. *Journal of Physics: Conference Series*, 18:224–269, jan 2005. doi: 10.1088/1742-6596/18/1/006. URL https://doi.org/10.1088/1742-6596/18/1/006.

[23] Zurab G. Berezhiani and Rabindra N. Mohapatra. Reconciling present neutrino puzzles: Sterile neutrinos as mirror neutrinos. *Phys. Rev. D*, 52:6607–6611, Dec 1995. doi: 10.1103/PhysRevD.52.6607. URL https://link.aps.org/doi/10.1103/PhysRevD.52.6607.

[24] CMS Collaboration. Evidence for Higgs boson decay to a pair of muons. *JHEP*, 2101:148. 68 p, Sep 2020. doi: 10.1007/JHEP01(2021)148. URL https://cds.cern.ch/record/2730058.

[25] Irene Dutta, Joosep Pata, Nan Lu, Jean-roch Vlimant, Harvey Newman, Maria Spiropulu, Christina Reissel, and Daniele Ruini. Data Analysis with GPU-Accelerated Kernels. In *Proceedings of 40th International Conference on High Energy physics — PoS(ICHEP2020)*, volume 390, page 908, 2021. doi: 10.22323/1.390.0908.

[26] CMS Collaboration. Search for nonresonant pair production of highly energetic Higgs bosons decaying to bottom quarks. Technical report, CERN, Geneva, May 2022. URL http://cds.cern.ch/record/2809466. Submitted to Physical Review Letters.

[27] Standard Model - Wikipedia. URL https://en.wikipedia.org/wiki/Standard_Model.

[28] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. doi: 10.1103/PhysRevLett.13.508. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[29] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. doi: 10.1103/PhysRevLett.13.321. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[30] 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. doi: 10.1103/PhysRevLett.13.585. URL https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[31] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966. doi: 10.1103/PhysRev.145.1156. URL https://link.aps.org/doi/10.1103/PhysRev.145.1156.

[32] T. W. B. Kibble. Symmetry breaking in non-abelian gauge theories. *Phys. Rev.*, 155:1554–1561, Mar 1967. doi: 10.1103/PhysRev.155.1554. URL https://link.aps.org/doi/10.1103/PhysRev.155.1554.

[33] James M. Cline. Baryogenesis, 2006. URL https://arxiv.org/abs/hep-ph/0609145.

[34] Matts Roos. Dark matter: The evidence from astronomy, astrophysics and cosmology. URL https://arxiv.org/abs/1001.0316.

[35] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, Aug 1998. doi: 10.1103/PhysRevLett.81.1562. URL https://link.aps.org/doi/10.1103/PhysRevLett.81.1562.

[36] Q. R. Ahmad et al. Direct Evidence for Neutrino Flavor Transformation from Neutral-Current Interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002. doi: 10.1103/PhysRevLett.89.011301. URL https://link.aps.org/doi/10.1103/PhysRevLett.89.011301.

[37] K. Eguchi et al. First Results from KamLAND: Evidence for Reactor Antineutrino Disappearance. *Phys. Rev. Lett.*, 90:021802, Jan 2003. doi: 10.1103/PhysRevLett.90.021802. URL https://link.aps.org/doi/10.1103/PhysRevLett.90.021802.

[38] Porter Williams. Naturalness, the autonomy of scales, and the 125 GeV Higgs. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 51:82–96, 2015. ISSN 1355-2198. doi: https://doi.org/10.1016/j.shpsb.2015.05.003. URL https://www.sciencedirect.com/science/article/pii/S1355219815000416.

[39] CMS Collaboration. A measurement of the Higgs boson mass in the diphoton decay channel. *Physics Letters B*, 805:135425, 2020. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2020.135425. URL https://www.sciencedirect.com/science/article/pii/S037026932030229X.

[40] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. 2016. doi: 10.23731/CYRM-2017-002.

[41] S. Dawson, S. Dittmaier, and M. Spira. Neutral Higgs boson pair production at hadron colliders: QCD corrections. *Phys. Rev. D*, 58:115012, 1998. doi: 10.1103/PhysRevD.58.115012.

[42] S. Borowka, N. Greiner, G. Heinrich, S.P. Jones, M. Kerner, J. Schlenk, U. Schubert, and T. Zirke. Higgs Boson Pair Production in gluon fusion at next-to-leading order with full top-quark mass dependence. *Phys. Rev. Lett.*, 117(1):012001, 2016. doi: 10.1103/PhysRevLett.117.079901. [Erratum: Phys.Rev.Lett. 117, 079901 (2016)].

[43] Daniel de Florian and Javier Mazzitelli. Higgs Boson pair production at next-to-next-to-leading order in QCD. *Phys. Rev. Lett.*, 111:201801, 2013. doi: 10.1103/PhysRevLett.111.201801.

[44] J. Baglio, A. Djouadi, R. Gröber, M.M. Mühlleitner, J. Quevillon, et al. The measurement of the Higgs self-coupling at the LHC: Theoretical status. *JHEP*, 1304:151, 2013. doi: 10.1007/JHEP04(2013)151.

[45] Ding Yu Shao, Chong Sheng Li, Hai Tao Li, and Jian Wang. Threshold resummation effects in Higgs boson pair production at the LHC. *JHEP*, 07:169, 2013. doi: 10.1007/JHEP07(2013)169.

[46] Daniel de Florian and Javier Mazzitelli. Higgs pair production at next-to-next-to-leading logarithmic accuracy at the LHC. *JHEP*, 09:053, 2015. doi: 10.1007/JHEP09(2015)053.

[47] Massimiliano Grazzini, Gudrun Heinrich, Stephen Jones, Stefan Kallweit, Matthias Kerner, Jonas M. Lindert, and Javier Mazzitelli. Higgs boson pair production at NNLO with top quark mass effects. *JHEP*, 05:059, 2018. doi: 10.1007/JHEP05(2018)059.

[48] R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, P. Torrielli, E. Vryonidou, and M. Zaro. Higgs pair production at the LHC with NLO and parton-shower effects. *Phys. Lett. B*, 732:142–149, 2014. doi: 10.1016/j.physletb.2014.03.026.

[49] Liu-Sheng Ling, Ren-You Zhang, Wen-Gan Ma, Lei Guo, Wei-Hua Li, and Xiao-Zhou Li. NNLO QCD corrections to Higgs pair production via vector boson fusion at hadron colliders. *Phys. Rev. D*, 89(7):073001, 2014. doi: 10.1103/PhysRevD.89.073001.

[50] Frédéric A. Dreyer and Alexander Karlberg. Fully differential vector-boson fusion Higgs pair production at next-to-next-to-leading order. *Phys. Rev. D*, 99(7):074028, 2019. doi: 10.1103/PhysRevD.99.074028.

[51] Frédéric A. Dreyer and Alexander Karlberg. Vector-boson fusion Higgs pair production at N^3LO. *Phys. Rev. D*, 98(11):114016, 2018. doi: 10.1103/PhysRevD.98.114016.

[52] Tommi Markkanen, Arttu Rajantie, and Stephen Stopyra. Cosmological aspects of Higgs vacuum metastability. *Frontiers in Astronomy and Space Sciences*, 5, 2018. ISSN 2296-987X. doi: 10.3389/fspas.2018.00040. URL https://www.frontiersin.org/article/10.3389/fspas.2018.00040.

[53] 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. *Phys. Rev. D*, 99:112003, Jun 2019. doi: 10.1103/PhysRevD.99.112003. URL https://link.aps.org/doi/10.1103/PhysRevD.99.112003.

[54] CMS Collaboration. CMS Results on HH production from Run II, . URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/SummaryResultsHIG.

[55] 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.

[56] CMS Collaboration. Search for Higgs boson pairs decaying to WWWW, WW$\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s}$ = 13 TeV. Technical report, CERN, Geneva, 2022. URL https://cds.cern.ch/record/2799151.

[57] CMS Collaboration. Search for Higgs boson pair production in the four b quark final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV, 2022. URL https://arxiv.org/abs/2202.09617.

[58] CMS Collaboration. Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *JHEP*, 03:257, 2021. doi: 10.1007/JHEP03(2021)257.

[59] CMS Collaboration. Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two tau leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV. Technical report, CERN, Geneva, 2022. URL https://cds.cern.ch/record/2803419.

[60] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3(08):S08002–S08002, Aug 2008. doi: 10.1088/1748-0221/3/08/s08002. URL https://doi.org/10.1088/1748-0221/3/08/s08002.

[61] The ATLAS Collaboration. The ATLAS experiment at the CERN large hadron collider. *JINST*, 3(08):S08003–S08003, aug 2008. doi: 10.1088/1748-0221/3/08/s08003. URL https://doi.org/10.1088/1748-0221/3/08/s08003.

[62] The CMS Collaboration. The CMS experiment at the CERN LHC. *JINST*, 3(08):S08004–S08004, aug 2008. doi: 10.1088/1748-0221/3/08/s08004. URL https://doi.org/10.1088/1748-0221/3/08/s08004.

[63] The LHCb Collaboration. The LHCb detector at the LHC. *JINST*, 3(08):S08005–S08005, Aug 2008. doi: 10.1088/1748-0221/3/08/s08005. URL https://doi.org/10.1088/1748-0221/3/08/s08005.

[64] CERN. Immersive tour of the accelerator complex. URL https://home.cern/science/accelerators/accelerator-complex/panoramas. Site last accessed in June 2022.

[65] J. P. Blewett. 200-GeV intersecting storage accelerators. *eConf*, C710920:501, 1971.

[66] CMS Collaboration. CMS luminosity public results, . URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Site last accessed in May 2022.

[67] Tai Sakuma. Cutaway diagrams of CMS detector. May 2019. URL http://cds.cern.ch/record/2665537.

[68] Irene Dutta. A search for pair produced vector-like T quarks in dilepton and multi-jet final states at sqrt (s) = 13 TeV. Master's thesis, 2017. URL https://cds.cern.ch/record/2272680.

[69] V Karimäki, M Mannelli, P Siegrist, H Breuker, A Caner, R Castaldi, K Freudenreich, G Hall, R Horisberger, M Huhtinen, and A Cattai. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. URL https://cds.cern.ch/record/368412.

[70] A Dominguez et al. CMS technical design report for the pixel detector upgrade. Technical report, Sep 2012. URL https://cds.cern.ch/record/1481838.

[71] CMS Collaboration. Sketches of the CMS tracker detector, . URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/DPGResultsTRK?sortcol=0;table=1;up=0#Sketches_of_the_CMS_Tracker_Dete. Site last accessed in June 2022.

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

[73] G L Bayatian et al. *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. There is an error on cover due to a technical problem for some items.

[74] CMS Collaboration. CMS technical design report for the Phase 1 upgrade of the hadron calorimeter. 9 2012. doi: 10.2172/1151651.

[75] Nadja Strobbe. The upgrade of the CMS hadron calorimeter with silicon photomultipliers. *JINST*, 12:C01080. 8 p, 2017. doi: 10.1088/1748-0221/12/01/C01080. URL http://cds.cern.ch/record/2291316.

[76] Federico De Guio. First results from the CMS SiPM-based hadronic endcap calorimeter. *Journal of Physics: Conference Series*, 1162:012009, Jan 2019. doi: 10.1088/1742-6596/1162/1/012009. URL https://doi.org/10.1088/1742-6596/1162/1/012009.

[77] CMS Collaboration. Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data. *JINST*, 5(03):T03012–T03012, Mar 2010. doi: 10.1088/1748-0221/5/03/t03012. URL https://doi.org/10.1088/1748-0221/5/03/t03012.

[78] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *JINST*, 5(03):T03021–T03021, mar 2010. doi: 10.1088/1748-0221/5/03/t03021. URL https://doi.org/10.1088/1748-0221/5/03/t03021.

[79] Particle Data Group, P A Zyla, et al. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2020(8), 08 2020. ISSN 2050-3911. doi: 10.1093/ptep/ptaa104. URL https://doi.org/10.1093/ptep/ptaa104. 083C01.

[80] CMS Collaboration. Performance of the CMS drift tube chambers with cosmic rays. *JINST*, 5(03):T03015–T03015, Mar 2010. doi: 10.1088/1748-0221/5/03/t03015. URL https://doi.org/10.1088/1748-0221/5/03/t03015.

[81] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV. *JINST*, 13(06): P06015–P06015, Jun 2018. doi: 10.1088/1748-0221/13/06/p06015. URL https://doi.org/10.1088/1748-0221/13/06/p06015.

[82] R Frazier, S Fayer, G Hall, C Hunt, G Iles, D Newbold, and A Rose. A demonstration of a time multiplexed trigger for the CMS experiment. *JINST*, 7(01):C01060–C01060, Jan 2012. doi: 10.1088/1748-0221/7/01/c01060. URL https://doi.org/10.1088/1748-0221/7/01/c01060.

[83] CMS Collaboration. Performance of the CMS level-1 trigger in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *JINST*, 15(10):P10017–P10017, oct 2020. doi: 10.1088/1748-0221/15/10/p10017. URL https://doi.org/10.1088/1748-0221/15/10/p10017.

[84] CMS Collaboration. The CMS high level trigger. *The European Physical Journal C*, 46(3):605–667, apr 2006. doi: 10.1140/epjc/s2006-02495-8. URL https://doi.org/10.1140%2Fepjc%2Fs2006-02495-8.

[85] Javier Duarte. Fast reconstruction and data scouting, 2018. URL https://arxiv.org/abs/1808.00902.

[86] Tommaso Boccali. Computing models in high energy physics. *Reviews in Physics*, 4:100034, 2019. ISSN 2405-4283. doi: https://doi.org/10.1016/j.revip.2019.100034. URL https://www.sciencedirect.com/science/article/pii/S2405428319300449.

[87] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017. doi: 10.1088/1748-0221/12/10/P10003.

[88] R. Frühwirth. Application of Kalman filtering to track and vertex fitting. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 262(2):444–450, 1987. ISSN 0168-9002. doi: https://doi.org/10.1016/0168-9002(87)90887-4. URL https://www.sciencedirect.com/science/article/pii/0168900287908874.

[89] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009–P10009, Oct 2014. doi: 10.1088/1748-0221/9/10/p10009. URL https://doi.org/10.1088/1748-0221/9/10/p10009.

[90] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[91] CERN. The High-Luminsoity LHC program and schedule. URL https://home.cern/science/accelerators/high-luminosity-lhc.

[92] 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.

[93] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. doi: 10.1016/S0168-9002(03)01368-8.

[94] S. Gundacker et al. Time of flight positron emission tomography towards 100ps resolution with L(Y)SO: An experimental and theoretical analysis. *JINST*, 8(07):P07014–P07014, Jul 2013. doi: 10.1088/1748-0221/8/07/p07014. URL https://doi.org/10.1088/1748-0221/8/07/p07014.

[95] D. Anderson et al. On timing properties of LYSO-based calorimeters. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 794:7–14, 2015. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2015.04.013. URL https://www.sciencedirect.com/science/article/pii/S0168900215004829.

[96] Dustin Anderson et al. Precision timing calorimeter for high energy physics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 824:670–673, 2016. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2015.11.129. URL https://www.sciencedirect.com/science/article/pii/S0168900215015223. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[97] S. White. R&D for a dedicated fast timing layer in the CMS endcap upgrade. *Acta Physica Polonica B Proceedings Supplement*, 7(4):743, 2014. doi: 10.5506/aphyspolbsupp.7.743. URL https://doi.org/10.5506%2Faphyspolbsupp.7.743.

[98] G. Pellegrini et al. Technology developments and first measurements of Low Gain Avalanche Detectors (LGAD) for high energy physics applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 765:12–16, 2014. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2014.06.008. URL https://www.sciencedirect.com/science/article/pii/S0168900214007128. HSTD-9 2013 - Proceedings of the 9th International "Hiroshima" Symposium on Development and Application of Semiconductor Tracking Detectors.

[99] N. Cartiglia et al. Design optimization of ultra-fast silicon detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 796:141–148, 2015. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2015.04.025. URL https://www.sciencedirect.com/science/article/pii/S0168900215004982. Proceedings of the 10th International Conference on Radiation Effects on Semiconductor Materials Detectors and Devices.

[100] E. Garutti and Y. Musienko. Radiation damage of SiPMs. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 926:69–84, 2019. doi: 10.1016/j.nima.2018.10.191. URL https://doi.org/10.1016%2Fj.nima.2018.10.191.

[101] Badder Marzocchi. CMS MTD Barrel Timing Layer: precision timing at the HL-LHC. In *Proceedings of The European Physical Society Conference on High Energy Physics — PoS(EPS-HEP2021)*, volume 398, page 813, 2022. doi: 10.22323/1.398.0813.

[102] Joao Varela for the TOFHIR team. The TOFHIR2 readout ASIC of the CMS Barrel MIP Timing Detector. Technical report, CERN, Geneva, Jan 2021. URL https://cds.cern.ch/record/2753168.

[103] CMS Collaboration. Technical proposal for a MIP timing detector in the CMS experiment Phase 2 upgrade. Technical report, CERN, Geneva, Dec 2017. URL https://cds.cern.ch/record/2296612.

[104] The CMS MTD collaboration. Test beam characterization of sensor prototypes for the CMS barrel MIP timing detector. *JINST*, 16(07):P07023, Jul 2021. doi: 10.1088/1748-0221/16/07/p07023. URL https://doi.org/10.1088/1748-0221/16/07/p07023.

[105] CAEN V1742: 32+2 Channel 12bit 5 GS/s switched capacitor digitizer. URL https://www.caen.it/products/v1742/. Site last accessed in June 2022.

[106] ChipCap 2-SIP Humidity and Temperature Sensor. URL https://www.digikey.com/catalog/en/partgroup/chipcap-2-sip/44852. Site last accessed in June 2022.

[107] AD5593R, 8-Channel, 12-Bit, Configurable ADC/DAC with On-Chip Reference, I^2C Interface. URL https://www.analog.com/media/en/technical-documentation/data-sheets/ad5593r.pdf. Site last accessed in June 2022.

[108] Esme Knabe. Thermal studies on the BTL readout Module for the CMS Phase II MIP Timing Detector. MTD talk at APS April 2022. URL https://indico.cern.ch/event/1145906. Site last accessed in June 2022.

[109] CMS Collaboration. Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector. *Phys. Lett.*, B779:283–316, 2018. doi: 10.1016/j.physletb.2018.02.004.

[110] CMS Collaboration. Observation of $t\bar{t}H$ production. *Phys. Rev. Lett.*, 120(23):231801, 2018. doi: 10.1103/PhysRevLett.120.231801.

[111] CMS Collaboration. Observation of Higgs boson decay to bottom quarks. *Phys. Rev. Lett.*, 121(12):121801, 2018. doi: 10.1103/PhysRevLett.121.121801.

[112] CMS Collaboration. Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79:421, 2019. doi: 10.1140/epjc/s10052-019-6909-y.

[113] ATLAS Collaboration. Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D99:072001, 2019. doi: 10.1103/PhysRevD.99.072001.

[114] ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett.*, B784: 173–191, 2018. doi: 10.1016/j.physletb.2018.07.035.

[115] ATLAS Collaboration. Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector. *Phys. Lett.*, B786:59–86, 2018. doi: 10.1016/j.physletb.2018.09.013.

[116] ATLAS Collaboration. Measurement of the Higgs boson mass in the $H \to ZZ^* \to 4\ell$ and $H \to \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector. *Phys. Lett.*, B784:345–366, 2018. doi: 10.1016/j.physletb.2018.07.050.

[117] ATLAS Collaboration. Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett.*, B789:508–529, 2019. doi: 10.1016/j.physletb.2018.11.064.

[118] CMS Collaboration. Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122(2):021801, 2019. doi: 10.1103/PhysRevLett.122.021801.

[119] ATLAS Collaboration. A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector. *Physics Letters B*, 812:135980, 2021. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2020.135980. URL https://www.sciencedirect.com/science/article/pii/S0370269320307838.

[120] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.

[121] Paolo Nason. A New method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[122] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A brief introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. doi: 10.1016/j.cpc.2008.01.036.

[123] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J.*, C76(3):155, 2016. doi: 10.1140/epjc/s10052-016-3988-x.

[124] CMS Collaboration. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. 2019. doi: 10.1140/epjc/s10052-019-7499-4. URL https://doi.org/10.1140%2Fepjc%2Fs10052-019-7499-4.

[125] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008. doi: 10.1140/epjc/s10052-007-0490-5.

[126] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012. doi: 10.1007/JHEP12(2012)061.

[127] Richard D. Ball et al. Parton distributions for the LHC Run II. *JHEP*, 04:040, 2015. doi: 10.1007/JHEP04(2015)040.

[128] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J.*, C77(10):663, 2017. doi: 10.1140/epjc/s10052-017-5199-5.

[129] S. Agostinelli et al. Geant4 : a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250 – 303, 2003. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(03)01368-8. URL http://www.sciencedirect.com/science/article/pii/S0168900203013688.

[130] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J.*, C76(4):196, 2016. doi: 10.1140/epjc/s10052-016-4018-8.

[131] CMS Collaboration. Electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78(7):589, 2018. doi: 10.1140/epjc/s10052-018-6049-9.

[132] Ye Li and Frank Petriello. Combining QCD and electroweak corrections to dilepton production in FEWZ. *Phys. Rev.*, D86:094034, 2012. doi: 10.1103/PhysRevD.86.094034.

[133] Dimitri Bourilkov. Photon-induced background for dilepton searches and measurements in pp collisions at 13 TeV. 2016. doi: 10.48550/ARXIV.1606.00523. URL https://arxiv.org/abs/1606.00523.

[134] Dimitri Bourilkov. Exploring the LHC landscape with dileptons. 2016. doi: 10.48550/ARXIV.1609.08994. URL https://arxiv.org/abs/1609.08994.

[135] Michal Czakon, Paul Fiedler, and Alexander Mitov. Total top-quark pair-production cross section at hadron colliders through $O(\alpha\frac{4}{S})$. *Phys. Rev. Lett.*, 110:252004, 2013. doi: 10.1103/PhysRevLett.110.252004.

[136] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov. Extracting muon momentum scale corrections for hadron collider experiments. *Eur. Phys. J. C*, 72:2194, 2012. doi: 10.1140/epjc/s10052-012-2194-8.

[137] X. Zuo. *A search for the Standard Model Higgs boson decaying into two muons in the vector boson associated production mode at the CMS experiment*. PhD thesis, University of Florida, 2021. URL https://ufdc.ufl.edu/UFE0057405/00001.

[138] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119, 2008. doi: 10.1016/j.physletb.2007.09.077.

[139] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12:P02014, 2017. doi: 10.1088/1748-0221/12/02/P02014.

[140] CMS Collaboration. Performance of missing transverse momentum in pp collisions at $\sqrt{s}$=13 TeV using the CMS detector. *JINST 14 P07004*, 2019. doi: 10.1088/1748-0221/14/07/p07004. URL https://doi.org/10.1088%2F1748-0221%2F14%2F07%2Fp07004.

[141] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10: P06005, 2015. doi: 10.1088/1748-0221/10/06/P06005.

[142] Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16(05):P05014, 2021. doi: 10.1088/1748-0221/16/05/p05014. URL https://doi.org/10.1088%2F1748-0221%2F16%2F05%2Fp05014.

[143] CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13:P05011, 2018. doi: 10.1088/1748-0221/13/05/P05011.

[144] CMS Collaboration. Performance of jet reconstruction with charged tracks only. Technical report, CERN, Geneva, Aug 2009. URL http://cds.cern.ch/record/1198681.

[145] CMS Collaboration. Commissioning of trackjets in pp collisions at 7 tev. Technical report, CERN, Geneva, 2010. URL http://cds.cern.ch/record/1275133.

[146] Keith Hamilton, Paolo Nason, Emanuele Re, and Giulia Zanderighi. NNLOPS simulation of Higgs boson production. *JHEP*, 10:222, 2013. doi: 10.1007/JHEP10(2013)222.

[147] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator. *JHEP*, 05:140, 2015. doi: 10.1007/JHEP05(2015)140.

[148] CMS Collaboration. Jet algorithms performance in 13 TeV data. Technical report, CERN, Geneva, 2017. URL https://cds.cern.ch/record/2256875.

[149] John C. Collins and Davison E. Soper. Angular distribution of dileptons in high-energy hadron collisions. *Phys. Rev. D*, 16:2219–2225, Oct 1977. doi: 10.1103/PhysRevD.16.2219. URL https://link.aps.org/doi/10.1103/PhysRevD.16.2219.

[150] Darin Acosta et al. $H \rightarrow \mu\mu$ search with CMS Run 2 datasets - CMS Internal Note AN-19-124. URL http://cms.cern.ch/iCMS/jsp/openfile.jsp?tp=draft&files=AN2019_124_v7.pdf. Site last accessed in June 2022.

[151] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71: 1554, 2011. doi: 10.1140/epjc/s10052-011-1554-0,10.1140/epjc/s10052-013-2501-z. [Erratum: Eur. Phys. J.C73,2501(2013)].

[152] Gilles Louppe, Michael Kagan, and Kyle Cranmer. Learning to pivot with adversarial networks. 2016. doi: 10.48550/ARXIV.1611.01046. URL https://arxiv.org/abs/1611.01046.

[153] CMS Collaboration. Cms luminosity measurements for the 2016 data taking period. Technical report, CERN, Geneva, 2017. URL https://cds.cern.ch/record/2257069.

[154] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV. Technical report, CERN, Geneva, 2018. URL https://cds.cern.ch/record/2621960.

[155] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV. Technical report, CERN, Geneva, 2019. URL https://cds.cern.ch/record/2676164.

[156] J.S. Conway. Incorporating nuisance parameters in likelihoods for multi-source spectra. pages 115–120. 6 p, Mar 2011. doi: 10.5170/CERN-2011-006.115. URL http://cds.cern.ch/record/1333496. Comments: Presented at PHYSTAT 2011, CERN, Geneva, Switzerland, January 2011, to be published in a CERN Yellow Report.

[157] Roger Barlow and Christine Beeston. Fitting using finite Monte Carlo samples. *Computer Physics Communications*, 77(2):219–228, 1993. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(93)90005-W. URL https://www.sciencedirect.com/science/article/pii/001046559390005W.

[158] Marco Delmastro et al. Simplified Template Cross Sections – Stage 1.1. Technical report, CERN, Geneva, Apr 2019. URL https://cds.cern.ch/record/2669925.

[159] ATLAS Collaboration, CMS Collaboration, and LHC Higgs Combination Group. Procedure for the LHC Higgs boson search combination in Summer 2011. Technical report, CERN, Geneva, Aug 2011. URL https://cds.cern.ch/record/1379837.

[160] Andreas Hoecker et al. TMVA - Toolkit for multivariate data analysis, 2007. URL https://arxiv.org/abs/physics/0703039.

[161] CMS Collaboration. Measurements of differential Z boson production cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 12:061, 2019. doi: 10.1007/JHEP12(2019)061.

[162] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, 2018. doi: 10.1103/PhysRevD.98.030001.

[163] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies. Handling uncertainties in background shapes. *JINST*, 10(04):P04015, 2015. doi: 10.1088/1748-0221/10/04/P04015.

[164] CMS Collaboration. Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 122:132003, 2019. doi: 10.1103/PhysRevLett.122.132003.

[165] CMS Collaboration. Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV. *JHEP*, 08:066, 2018. doi: 10.1007/JHEP08(2018)066.

[166] CMS Collaboration. Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC. *Phys. Lett.*, B744:184–207, 2015. doi: 10.1016/j.physletb.2015.03.048.

[167] CMS Collaboration. Combined Higgs boson production and decay measurements with up to 137 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-HIG-19-005, CERN, Geneva, 2020. URL https://cds.cern.ch/record/2706103.

[168] LHC Higgs Cross Section Working Group et al. LHC HXSWG interim recommendations to explore the coupling structure of a Higgs-like particle, 2012. URL https://arxiv.org/abs/1209.0040.

[169] C T Potter et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group, 2013. URL http://cds.cern.ch/record/1559921.

[170] CMS Collaboration. The Phase-2 upgrade of the CMS tracker. Technical report, CERN, Geneva, Jun 2017. URL https://cds.cern.ch/record/2272264.

[171] CMS Collaboration. The Phase-2 upgrade of the CMS muon detectors. Technical report, CERN, Geneva, Sep 2017. URL https://cds.cern.ch/record/2283189. This is the final version, approved by the LHCC.

[172] CMS Collaboration. The Phase-2 upgrade of the CMS endcap calorimeter. Technical report, CERN, Geneva, Nov 2017. URL https://cds.cern.ch/record/2293646.

[173] M. Cepeda et al. Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC. *CERN Yellow Rep. Monogr.*, 7:221–584. 364 p, Dec 2018. doi: 10.23731/CYRM-2019-007.221. URL http://cds.cern.ch/record/2650162.

[174] Biagio Di Micco et al. Higgs boson potential at colliders: status and perspectives. Technical report, CERN, Geneva, Sep 2019. URL https://cds.cern.ch/record/2690841.

[175] ATLAS Collaboration. Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. *Phys. Rev. D*, 101:012002, 2020. doi: 10.1103/PhysRevD.101.012002.

[176] Gudrun Heinrich, S. Jones, M. Kerner, Gionata Luisoni, and Ludovic Scyboz. Trilinear Higgs boson coupling variations for di-Higgs production with full NLO QCD predictions in Powheg. *Journal of Physics: Conference Series*, 1525:012009, 04 2020. doi: 10.1088/1742-6596/1525/1/012009.

[177] F. Bishara, R. Contino, and J. Rojo. Higgs pair production in vector-boson fusion at the LHC and beyond. *Eur. Phys. J. C*, 77(7):481, 2017. doi: 10.1140/epjc/s10052-017-5037-9. URL https://doi.org/10.1140/epjc/s10052-017-5037-9.

[178] 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. Submitted to *Phys. Rev. D*, 2021. URL https://arxiv.org/abs/2112.11876.

[179] CMS Collaboration. Search for Higgs boson pair production in the $\gamma\gamma\mathrm{b\bar{b}}$ final state in pp collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 788:7, 2019. doi: 10.1016/j.physletb.2018.10.056.

[180] ATLAS Collaboration. 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(8):081802, 2015. doi: 10.1103/PhysRevLett.114.081802.

[181] ATLAS Collaboration. Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$

TeV with the ATLAS detector. *Phys. Rev. Lett.*, 121(19):191801, 2018. doi: 10.1103/PhysRevLett.121.191801. [Erratum: Phys.Rev.Lett. 122, 089901 (2019)].

[182] CMS Collaboration. Search for Higgs boson pair production in the $bb\tau\tau$ final state in proton-proton collisions at $\sqrt{s}$ = 8TeV. *Phys. Rev. D*, 96(7):072004, 2017. doi: 10.1103/PhysRevD.96.072004.

[183] ATLAS Collaboration. Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion production using proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *JHEP*, 07:108, 2020. doi: 10.1007/JHEP07(2020)108.

[184] CMS Collaboration. Search for nonresonant Higgs boson pair production in the $\mathrm{b\bar{b}b\bar{b}}$ final state at $\sqrt{s}$ = 13 TeV. *JHEP*, 04:112, 2019. doi: 10.1007/JHEP04(2019)112.

[185] 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. *JHEP*, 01:030, 2019. doi: 10.1007/JHEP01(2019)030.

[186] 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. *Phys. Rev. D*, 94(5):052002, 2016. doi: 10.1103/PhysRevD.94.052002.

[187] ATLAS Collaboration. 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(9):412, 2015. doi: 10.1140/epjc/s10052-015-3628-x.

[188] ATLAS Collaboration. Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s}$ = 13 TeV. *Phys. Lett. B*, 801:135145, 2020. doi: 10.1016/j.physletb.2019.135145.

[189] ATLAS Collaboration. Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s}$ = 13 TeV with the ATLAS detector. *JHEP*, 04:092, 2019. doi: 10.1007/JHEP04(2019)092.

[190] CMS Collaboration. Search for resonant pair production of Higgs bosons in the $bbZZ$ channel in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *Phys. Rev. D*, 102(3):032003, 2020. doi: 10.1103/PhysRevD.102.032003.

[191] CMS Collaboration. Search for resonant and nonresonant Higgs boson pair production in the $\mathrm{b\bar{b}}\ell\nu\ell\nu$ final state in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *JHEP*, 01:054, 2018. doi: 10.1007/JHEP01(2018)054.

[192] CMS Collaboration. Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *Phys. Rev. Lett.*, 122(12):121803, 2019. doi: 10.1103/PhysRevLett.122.121803.

[193] ATLAS Collaboration. Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector. *Phys. Lett. B*, 800: 135103, 2020. doi: 10.1016/j.physletb.2019.135103.

[194] ATLAS Collaboration. Searches for Higgs boson pair production in the $hh \to bb\tau\tau, \gamma\gamma WW^{*}, \gamma\gamma bb, bbbb$ channels with the ATLAS detector. *Phys. Rev. D*, 92:092004, 2015. doi: 10.1103/PhysRevD.92.092004.

[195] Huilin Qu and Loukas Gouskos. ParticleNet: Jet Tagging via Particle Clouds. *Phys. Rev. D*, 101(5):056019, 2020. doi: 10.1103/PhysRevD.101.056019.

[196] D. Krohn, J. Thaler, and L. Wang. Jet trimming. *JHEP*, 02:084, 2010. doi: 10.1007/JHEP02(2010)084.

[197] Paolo Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11:040, 2004. doi: 10.1088/1126-6708/2004/11/040.

[198] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

[199] 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. doi: 10.1007/JHEP06(2010)043.

[200] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini. Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM. *JHEP*, 02:088, 2012. doi: 10.1007/JHEP02(2012)088.

[201] G. Heinrich, Stephen Philip Jones, M. Kerner, G. Luisoni, and E. Vryonidou. NLO predictions for Higgs boson pair production with full top quark mass dependence matched to parton showers. *JHEP*, 08:088, 2017. doi: 10.1007/JHEP08(2017)088.

[202] G. Heinrich, S. P. Jones, M. Kerner, G. Luisoni, and L. Scyboz. Probing the trilinear Higgs boson coupling in di-Higgs production at NLO QCD including parton shower effects. *JHEP*, 06:066, 2019. doi: 10.1007/JHEP06(2019)066.

[203] Stephen Jones and Silvan Kuttimalai. Parton shower and NLO-matching uncertainties in Higgs boson pair production. *JHEP*, 02:176, 2018. doi: 10.1007/JHEP02(2018)176.

[204] Gudrun Heinrich, Stephen Philip Jones, Matthias Kerner, and Ludovic Scyboz. A non-linear EFT description of $gg \to HH$ at NLO interfaced to POWHEG. *JHEP*, 10:021, 2020. doi: 10.1007/JHEP10(2020)021.

[205] G. Buchalla, M. Capozi, A. Celis, G. Heinrich, and L. Scyboz. Higgs boson pair production in non-linear Effective Field Theory with full m_t-dependence at NLO QCD. *JHEP*, 09:057, 2018. doi: 10.1007/JHEP09(2018)057.

[206] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015. doi: 10.1016/j.cpc.2015.01.024.

[207] CMS Collaboration. Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80:4, 2020. doi: 10.1140/epjc/s10052-019-7499-4.

[208] Joshua Davies, Gudrun Heinrich, Stephen P. Jones, Matthias Kerner, Go Mishima, Matthias Steinhauser, and David Wellmann. Double Higgs boson production at NLO: Combining the exact numerical result and high-energy expansion. *JHEP*, 11:024, 2019. doi: 10.1007/JHEP11(2019)024.

[209] Julien Baglio, Francisco Campanario, Seraina Glaus, Margarete Mühlleitner, Michael Spira, and Juraj Streicher. Gluon fusion into Higgs pairs at NLO QCD and the top mass scheme. *Eur. Phys. J. C*, 79:459, 2019. doi: 10.1140/epjc/s10052-019-6973-3.

[210] J. Baglio, F. Campanario, S. Glaus, M. Mühlleitner, J. Ronca, and M. Spira. $gg \to HH$: Combined uncertainties. *Phys. Rev. D*, 103:056002, 2021. doi: 10.1103/PhysRevD.103.056002.

[211] Tomáš Ježo, Jonas M. Lindert, Paolo Nason, Carlo Oleari, and Stefano Pozzorini. An NLO+PS generator for $t\bar{t}$ and Wt production and decay including non-resonant and interference effects. *Eur. Phys. J. C*, 76:691, 2016. doi: 10.1140/epjc/s10052-016-4538-2.

[212] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction. *JHEP*, 09:126, 2007. doi: 10.1088/1126-6708/2007/09/126.

[213] Michal Czakon and Alexander Mitov. Top++: A program for the calculation of the top-pair cross-section at hadron colliders. *Comput. Phys. Commun.*, 185:2930, 2014. doi: 10.1016/j.cpc.2014.06.021.

[214] Michal Czakon, David Heymes, Alexander Mitov, Davide Pagani, Ioannis Tsinikos, and Marco Zaro. Top-pair production at the LHC through NNLO QCD and NLO EW. *JHEP*, 10:186, 2017. doi: 10.1007/JHEP10(2017)186.

[215] Keith Rehermann and Brock Tweedie. Efficient identification of boosted semileptonic top quarks at the LHC. *JHEP*, 2011(3), Mar 2011. doi: 10.1007/jhep03(2011)059. URL https://doi.org/10.1007%2Fjhep03%282011%29059.

[216] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup Per Particle Identification. *JHEP*, 10:059, 2014. doi: 10.1007/JHEP10(2014)059.

[217] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft drop. *JHEP*, 05:146, 2014. doi: 10.1007/JHEP05(2014)146.

[218] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani, and Gavin P. Salam. Towards an understanding of jet substructure. *JHEP*, 09:029, 2013. doi: 10.1007/JHEP09(2013)029.

[219] CMS Collaboration. Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques. 2020. URL https://cds.cern.ch/record/2707946.

[220] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph CNN for learning on point clouds. *ACM Trans. Graph.*, 38, 2019. doi: 10.1145/3326362.

[221] CMS Collaboration. Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques. *JINST*, 15(06):P06005, 2020. doi: 10.1088/1748-0221/15/06/P06005.

[222] CMS Collaboration. Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. *JHEP*, 12:85, 2020. doi: 10.1007/JHEP12(2020)085.

[223] Steve Baker and Robert D. Cousins. Clarification of the use of chi-square and likelihood functions in fits to histograms. *Nuclear Instruments and Methods in Physics Research*, 221(2):437–442, 1984. ISSN 0167-5087. doi: https://doi.org/10.1016/0167-5087(84)90016-4. URL https://www.sciencedirect.com/science/article/pii/0167508784900164.

[224] Robert D. Cousins. Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models , with Application to Histograms. 2013. URL https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[225] CMS Collaboration. Calibration of the mass-decorrelated ParticleNet tagger for boosted $b\bar{b}$ and $c\bar{c}$ jets using LHC Run 2 data. CMS Detector Performance Note CMS-DP-2022-005, 2022. URL https://cds.cern.ch/record/2805611.

[226] Joshua Davies, Gudrun Heinrich, Stephen P. Jones, Matthias Kerner, Go Mishima, Matthias Steinhauser, and David Wellmann. Double Higgs boson production at NLO: combining the exact numerical result and high-energy expansion. *JHEP*, 11:24, 2019. doi: 10.1007/JHEP05(2019)024.

[227] CMS Collaboration. Search for Higgs boson pair production via vector boson fusion with highly Lorentz-boosted Higgs bosons in the four b quark

final state at $\sqrt{s}$ = 13 TeV. Technical report, CERN, 2021. URL https://cds.cern.ch/record/2776802.

[228] CMS Collaboration. Search for production of higgs boson pairs in the four b quark final state using large-area jets in proton-proton collisions at $\sqrt{s}$=13 TeV. *JHEP*, 2019(1), Jan 2019. doi: 10.1007/jhep01(2019)040. URL https://doi.org/10.1007%2Fjhep01%282019%29040.

[229] CMS Collaboration. Search for nonresonant higgs boson pair production in the $b\bar{b}b\bar{b}$ final state at $\sqrt{s}$ = 13 TeV. *JHEP*, 2019(4), apr 2019. doi: 10.1007/jhep04(2019)112. URL https://doi.org/10.1007%2Fjhep04%282019%29112.

[230] Daniel Egana-Ugrinovic, Samuel Homiller, and Patrick Meade. Multi-Higgs boson production probes Higgs sector flavor. *Phys. Rev. D*, 103:115005, Jun 2021. doi: 10.1103/PhysRevD.103.115005. URL https://link.aps.org/doi/10.1103/PhysRevD.103.115005.

[231] HCAL subdetector ieta-iphi mapping. URL https://indico.cern.ch/event/645831/contributions/2622664/attachments/1473654/2281324/HcalDQMShiftOffline.pdf. Site last accessed in June 2022.

[232] PETsys TOFPET2 ASIC. URL https://www.petsyselectronics.com/web/public/products/1. Site last accessed in June 2022.

[233] Joosep Pata. hepaccelerate: v0.2.0. Zenodo, October 2019. URL https://doi.org/10.5281/zenodo.3491423.

[234] I. Antcheva et al. ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization. *Computer Physics Communications*, 180(12):2499–2512, 2009. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2009.08.005. URL https://www.sciencedirect.com/science/article/pii/S0010465509002550.

[235] Jim Pivarski et al. uproot, . URL https://github.com/scikit-hep/uproot. Site last accessed in June 2022.

[236] Jim Pivarski et al. awkward, . URL https://github.com/scikit-hep/awkward-array.

[237] Brian Bockelman, Zhe Zhang, and Jim Pivarski. Optimizing ROOT IO for analysis. *CoRR*, abs/1711.02659, 2017. URL http://arxiv.org/abs/1711.02659.

[238] Ritesh A. Patel, Yao Zhang, Jason Mak, Andrew Davidson, and John D. Owens. Parallel lossless data compression on the gpu. In *2012 Innovative Parallel Computing (InPar)*, pages 1–9, 2012. doi: 10.1109/InPar.2012.6339599.

[239] Lindsey Gray, Nicholas Smith, Mauro Verzetti, Joosep Pata, Anna Woodard, Pratyush Das, Nick Amin, Andreas, areinsvo, and Lukas. coffea v0.6.8. Zenodo, August 2019. URL https://doi.org/10.5281/zenodo.3358981.

[240] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: A llvm-based python jit compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450340052. doi: 10.1145/2833157.2833162. URL https://doi.org/10.1145/2833157.2833162.

[241] NVIDIA CUDA Home page. URL https://developer.nvidia.com/cuda-zone.

[242] CMS Collaboration. Simulated datasets in AODSIM format for 2012 collision data. CERN Open Data Portal, . doi:10.7483/OPENDATA.CMS.42GY.2VJI, doi:10.7483/OPENDATA.CMS.2DSE.HYDF, doi:10.7483/OPENDATA.CMS.7RZ3.0BXP, doi:10.7483/OPENDATA.CMS.ARKO.6NV3, doi:10.7483/OPENDATA.CMS.REHM.JKUH, doi:10.7483/OPENDATA.CMS.DELK.2V7R, doi:10.7483/OPENDATA.CMS.HHCJ.TVXH, doi:10.7483/OPENDATA.CMS.DELK.2V7R, doi:10.7483/OPENDATA.CMS.KAYE.XLAH.

[243] CMS Collaboration. Tool to convert CMS AOD data format to NanoAOD format for the purpose of education and outreach, . URL https://github.com/cms-opendata-analyses/AOD2NanoAODOutreachTool.

[244] ATLAS Collaboration. ATLAS 13 TeV samples collection at least four leptons (electron or muon), for 2020 Open Data release. CERN Open Data Portal, 2020. URL http://opendata.cern.ch/record/15005.