

Search for New Physics in the Two Higgs to Four b-quark Final State

by

Alice Cocoros

A dissertation submitted to The Johns Hopkins University
in conformity with the requirements for the degree of
Doctor of Philosophy

Baltimore, Maryland

June, 2017

© Alice Cocoros 2017

All rights reserved

Abstract

The Standard Model of particle physics represents the current understanding of elementary particles and their interactions. However, there are still many phenomena unaccounted for, which can be described by a number of different theories. In some new physics models, heavy particles that couple to the Higgs boson exist. The Higgs boson itself is also newly discovered, and its parameters are still being studied. At the Large Hadron Collider at CERN, the CMS detector records the results of proton proton collisions, in hopes of finding these new particles and studying particles such as the Higgs. Using 35.9^{-1} fb of data taken in 2016 by the CMS detector, a search for these new particles that also probes Higgs properties is performed. Limits are placed on several models of new physics at 95% confidence level.

Primary Reader: Petar Maksimovic

Secondary Reader: Morris Swartz

Acknowledgments

Thanks!

Table of Contents

Abstract	ii
Acknowledgments	iii
1 Introduction	1
1.1 The Standard Model	2
1.1.1 Leptons	5
1.1.2 Quarks	6
1.1.3 Representation of Interactions	7
1.1.4 The Electromagnetic Force:	8
1.1.5 The Strong Force:	9
1.1.6 The Weak Force:	11
1.1.7 Mass and the Higgs Boson	14
1.2 Rounding Up the Standard Model Particles	16
1.3 Beyond the Standard Model	17
2 Theoretical overview	18
2.1 Resonant Production	19
2.2 Nonresonant Production	22

3	Searching for New Particles	28
3.1	The Large Hadron Collider	29
3.1.1	A proton-proton Collision	30
3.2	The Compact Muon Solenoid	32
3.2.1	Coordinates	33
3.2.2	The Tracker	35
3.2.3	The Calorimeters	36
3.2.4	The Solenoid	37
3.2.5	The Muon Chambers	37
3.2.6	What's Missing?	38
3.2.7	The Particle Flow Algorithm	39
3.2.8	Jets	40
3.2.9	Triggers	41
3.2.10	Pileup	42
4	Looking for New Physics	44
4.1	Signal Signature	44
4.2	Distinguishing Signal from Background	47
4.2.1	The Soft-Drop Mass Algorithm	48
4.2.2	The N-subjetiness Algorithm	49
4.2.3	B-Tagging	51
4.2.3.1	Deep CSV Algorithm	52
4.2.3.2	Double-b Algorithm	54
5	Semi-resolved Analysis	56
5.1	Data samples	59

5.2	MC simulation	60
5.3	Event selection	60
5.3.1	Jet kinematics selection	60
5.3.2	H mass selection	64
5.3.3	N-subjettiness selection	65
5.3.4	B-tagging	66
5.3.5	Variable for tt Reduction	67
5.3.6	Invariant mass definition and “reduced mass”	68
5.3.7	Lepton Veto	68
5.3.8	Trigger Choice	69
5.3.9	Summary of the Semi-Resolved Section	74
5.3.10	Selection of Other Analyses	76
5.3.10.1	Boosted Analysis	76
5.3.10.2	Resolved Analysis	77
6	Background Estimate	81
6.1	Alphabet Background Estimate	81
6.2	Closure Test in Data	84
6.3	Closure Test in MC	87
6.4	Signal Region in Data	87
7	Statistical Interpretation	92
7.1	Systematic uncertainties	92

List of Tables

2.1	Parameter values of the final benchmarks selected with number of clusters $N_{clus} = 12$	25
5.1	List of primary datasets for the pp collisions at $\sqrt{s} = 13$ TeV and their corresponding integrated luminosities.	59
5.2	List of signal MC samples used.	79
5.3	List of background Monte Carlo samples used.	80
5.4	The HLT paths used and the corresponding L1 seeds.	80

List of Figures

1.1	Snapshot of a cloud chamber, circa 1933: the vertical line represents a positron traveling through the cloud chamber, curving due to the magnetic field applied to the chamber.	1
1.2	The Standard Model Lagrangian.	3
1.3	Table of Standard Model particles. Image from CERN [2]. . . .	4
1.4	Visual representation of the interaction between two electrons (straight lines) and a photon (wavy line).	7
1.5	A Feynman Diagram showing the interaction of two electrons. Time increases to the right on the x-axis, and the y-axis represents the distance between the two electrons.	8
1.6	A Feynman Diagram showing the annihilation of a positron and an electron.	9
1.7	QCD interactions can be described by three vertices. The curly line represents gluons, while the straight lines represent quarks. Since gluons are colored, they can interact with themselves. . . .	10

1.8 Weak interactions can be described by the following vertices. The wavy lines represent bosons, while the straight lines represent any weak-interacting particle where the appropriate quantities are conserved.	12
1.9 Approximation of the Higgs potential. While the shape is symmetric about the y-axis, the function is not symmetric about the minima.	14
1.10 The Higgs Vertex. The dashed line is the Higgs Boson and the straight lines are any massive particle.	15
1.11 Left: Two gluons interact with a quark, while each gluon interacts with an additional different quark, and these additional different quarks interact with a shared fourth quark and each pair of quarks interacts to form a Higgs boson. Right: Two gluons interact with a quark, while each gluon interacts with an additional quark, and these additional different quarks interact with each other to form a Higgs boson, which interacts with two Higgs bosons. The result of both of these processes is a final state with two Higgs.	16
2.1 A depiction of the fifth extra dimension.	20
2.2 SM Feynman diagrams that contribute to Higgs boson pair production by gluon-gluon fusion at leading order.	22

3.3	The Compact Muon Solenoid detector.	32
3.4	CMS detector with a person for size (top). Slice of the CMS detector, depicting the different components and the signatures of some of the different particles that pass through the detector after a proton-proton collision occurs (bottom).	34
3.5	Coordinates of the CMS detector.	35
3.6	Diagram of particles inside the detector pre-collision (top) and post-collision (bottom), where post-collision shows where the protons entered, which particles were recorded (blue), and the direction of the missing energy (orange).	39
3.7	Typical CMS event with large number of hadrons (tracks represented by green lines, HCAL energy in blue, ECAL energy in red). These hadrons are collected into <i>jets</i> (shown by yellow triangles). The PF algorithm combines information from the trackers and calorimeters to properly measure the total momentum and energy of the jet.	40
4.1	Drawing of decay products with varying degrees of Lorentz boost.	45
4.2	Signature for resolved two Higgs to four b-quarks.	46
4.3	Signature for boosted two Higgs to four b-quarks.	46
4.4	Signature for semi-resolved two Higgs to four b-quarks.	47
4.5	Softdrop mass (right) does a better job of discriminating between QCD (blue) and signal (W boson, in red), than ungroomed mass does (left).	49
4.6	N-subjetiness variable τ_2/τ_{21} for W jets and background jets. . .	51

4.7	Light quark (u, d, c, s) in a detector (left) vs. b-quark in a detector (right).	52
4.8	Deep CSV discriminator for b jets (red), c jets (green), and light jets (blue).	53
4.9	Double b-tagger for signal (red) and various QCD processes (varying shades of blue).	55
5.1	Signature for semi-resolved two Higgs to four b-quarks.	57
5.2	Left. The p_T of the AK8 jet. Right. The η of the AK8 jet.	63
5.3	Left. The p_T of the highest p_T selected AK4 jet. Right. The η of the highest p_T selected AK4 jet.	63
5.4	Left. The p_T of the other selected AK4 jet. Right. The η of the other selected AK4 jet.	63
5.5	The absolute value of $\Delta\eta$ between the AK8 jet and the combined 4 vector of the AK4 jets. We require $ \Delta\eta < 2.0$	64
5.6	The soft-drop mass of the AK8 jet. We require it to be between 105 and 135 GeV.	65
5.7	The dijet mass of the AK4 jet. We require it to be between 90 and 140 GeV.	66
5.8	τ_{21} distribution for the AK8 jet. We require $\tau_{21} < 0.55$	66
5.9	Double-b tagger discriminant for the AK8 jet. We require double b tagger > 0.8	67
5.10	Deep CSV P(b) + P(bb) tagger discriminant for the highest p_T selected AK4 jet (left) and the other selected AK4 jet (right). We require deepCSV > 0.6324	68

5.11 TriAk4jet mass: the invariant mass of the two selected AK4 jets and the nearest unselected AK4 jet. We require triAK4jet mass > 200 GeV.	69
5.12 Reduced mass of the two selected AK4 jets and the AK8 jet.	70
5.13 The trigger efficiency, as a function of reduced mass. Efficiency is defined for different $\Delta\eta$ regions: 0.0–1.0 (left), 1.0–2.0 (right).	73
5.14 The trigger efficiency in QCD MC for the baseline trigger HLT_PFT260, for different $\Delta\eta$ regions: 0.0–1.0 (left) and 1.0–2.0 (right). The percentage difference between one and these turn-on curves are taken as an uncertainty on the trigger efficiency scale factor.	73
5.15 The trigger efficiency scale factors, as a function of M_{jjj}^{red} . The error bars are the combined statistical and systematic errors.	74
6.1 Schematic representation of the regions used to perform our estimate.	83
6.2 First control region (AK4 dijet mass < 70 GeV and remove the triAK4jet mass requirement) for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the first control regions, compared with the true background (black markers). Top is $ \Delta\eta = 0$ –1.0 and bottom is $ \Delta\eta = 1$.0–2.0.	85

6.3	Second control region (AK4 jet 2 deep CSV < 0.6324 and remove the triAK4jet mass requirement) for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the control regions, compared with the true background (black markers). Top is $ \Delta\eta = 0\text{-}1.0$ and bottom is $ \Delta\eta = 1.0\text{-}2.0$.	86
6.4	Signal region in MC for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the first control regions, compared with the true background (black markers). Top is $ \Delta\eta = 0\text{-}1.0$ and bottom is $ \Delta\eta = 1.0\text{-}2.0$.	88
6.5	Signal region in data for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the unblinded signal region. Top is $ \Delta\eta = 0\text{-}1.0$ and bottom is $ \Delta\eta = 1.0\text{-}2.0$.	89
6.6	Signal region in data for a selection rejecting boosted and resolved events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the unblinded signal region. Top is $ \Delta\eta = 0\text{-}1.0$ and bottom is $ \Delta\eta = 1.0\text{-}2.0$.	90

6.7 Signal region in data for a selection retaining boosted and resolved events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the unblinded signal region. Top is $|\Delta\eta| = 0\text{-}1.0$ and bottom is $|\Delta\eta| = 1.0\text{-}2.0$. . . 91

Chapter 1

Introduction

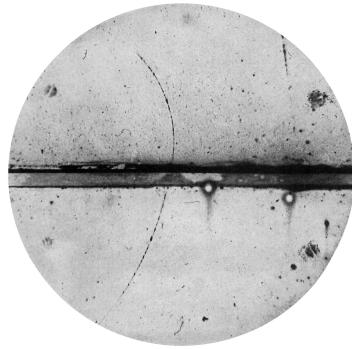


Figure 1.1: Snapshot of a cloud chamber, circa 1933: the vertical line represents a positron traveling through the cloud chamber, curving due to the magnetic field applied to the chamber.

In 1869, Johann Wilhelm Hittorf discovered cathode rays[1], suggesting that there may be subatomic particles. By 1911, these were identified as electrons, observed in many different ways, including cloud chambers, as can be seen in Figure 1.1. Soon after, protons were discovered in 1917, followed by neutrons in 1935.

As physicists advanced the study of quantum mechanics, many new particles were discovered. These particles can be broken down into several elementary particles, whose interactions are governed by *The Standard Model* (SM).

1.1 The Standard Model

The SM describes the interaction of all elementary particles with one formula, shown in Figure 1.2. This lengthy Lagrangian can be summarized as a list of particles and mediating forces, represented in Figure 1.3.

$$\begin{aligned}
& -\frac{1}{2}\partial_\nu g_\mu^a \partial_\nu g_\mu^a - g_s f^{abc} \partial_\mu g_\nu^a g_\mu^b g_\nu^c - \frac{1}{4}g_s^2 f^{abc} f^{ade} g_\mu^b g_\nu^c g_\mu^d g_\nu^e + \\
& \frac{1}{2}ig_s^2 (\bar{q}_i^\sigma \gamma^\mu q_j^\sigma) g_\mu^a + \bar{G}^a \partial^2 G^a + g_s f^{abc} \partial_\mu \bar{G}^a G^b g_\mu^c - \partial_\nu W_\mu^+ \partial_\nu W_\mu^- - \\
& M^2 W_\mu^+ W_\mu^- - \frac{1}{2} \partial_\nu Z_\mu^0 \partial_\nu Z_\mu^0 - \frac{1}{2c_w^2} M^2 Z_\mu^0 Z_\mu^0 - \frac{1}{2} \partial_\mu A_\nu \partial_\mu A_\nu - \frac{1}{2} \partial_\mu H \partial_\mu H - \\
& \frac{1}{2} m_h^2 H^2 - \partial_\mu \phi^+ \partial_\mu \phi^- - M^2 \phi^+ \phi^- - \frac{1}{2} \partial_\mu \phi^0 \partial_\mu \phi^0 - \frac{1}{2c_w^2} M \phi^0 \phi^0 - \beta_h [\frac{2M^2}{g^2} + \\
& \frac{2M}{g} H + \frac{1}{2}(H^2 + \phi^0 \phi^0 + 2\phi^+ \phi^-)] + \frac{2M^4}{g^2} \alpha_h - ig c_w [\partial_\nu Z_\mu^0 (W_\mu^+ W_\nu^- - \\
& W_\nu^+ W_\mu^-) - Z_\mu^0 (W_\mu^+ \partial_\nu W_\mu^- - W_\mu^- \partial_\nu W_\mu^+) + Z_\mu^0 (W_\nu^+ \partial_\nu W_\mu^- - \\
& W_\nu^- \partial_\nu W_\mu^+)] - ig s_w [\partial_\nu A_\mu (W_\mu^+ W_\nu^- - W_\nu^+ W_\mu^-) - A_\nu (W_\mu^+ \partial_\nu W_\mu^- - \\
& W_\mu^- \partial_\nu W_\mu^+) + A_\mu (W_\nu^+ \partial_\nu W_\mu^- - W_\nu^- \partial_\nu W_\mu^+)] - \frac{1}{2} g^2 W_\mu^+ W_\nu^- W_\mu^+ W_\nu^- + \\
& \frac{1}{2} g^2 W_\mu^+ W_\nu^- W_\mu^+ W_\nu^- + g^2 c_w^2 (Z_\mu^0 W_\mu^+ Z_\nu^0 W_\nu^- - Z_\mu^0 Z_\mu^0 W_\nu^+ W_\nu^-) + \\
& g^2 s_w^2 (A_\mu W_\mu^+ A_\nu W_\nu^- - A_\mu A_\nu W_\mu^+ W_\nu^-) + g^2 s_w c_w [A_\mu Z_\nu^0 (W_\mu^+ W_\nu^- - \\
& W_\nu^+ W_\mu^-) - 2A_\mu Z_\mu^0 W_\nu^+ W_\nu^-] - g\alpha [H^3 + H\phi^0 \phi^0 + 2H\phi^+ \phi^-] - \\
& \frac{1}{8} g^2 \alpha_h [H^4 + (\phi^0)^4 + 4(\phi^+ \phi^-)^2 + 4(\phi^0)^2 \phi^+ \phi^- + 4H^2 \phi^+ \phi^- + 2(\phi^0)^2 H^2] - \\
& g M W_\mu^+ W_\mu^- H - \frac{1}{2} g \frac{M}{c_w^2} Z_\mu^0 H - \frac{1}{2} g [W_\mu^+ (\phi^0 \partial_\mu \phi^- - \phi^- \partial_\mu \phi^0) - \\
& W_\mu^- (\phi^0 \partial_\mu \phi^+ - \phi^+ \partial_\mu \phi^0)] + \frac{1}{2} g [W_\mu^+ (H \partial_\mu \phi^- - \phi^- \partial_\mu H) - W_\mu^- (H \partial_\mu \phi^+ - \\
& \phi^+ \partial_\mu H)] + \frac{1}{2} g \frac{1}{c_w} (Z_\mu^0 (H \partial_\mu \phi^0 - \phi^0 \partial_\mu H) - ig \frac{s_w^2}{c_w} M Z_\mu^0 (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \\
& ig s_w M A_\mu (W_\mu^+ \phi^- - W_\mu^- \phi^+) - ig \frac{1-2c_w^2}{2c_w} Z_\mu^0 (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) + \\
& ig s_w A_\mu (\phi^+ \partial_\mu \phi^- - \phi^- \partial_\mu \phi^+) - \frac{1}{4} g^2 W_\mu^+ W_\mu^- [H^2 + (\phi^0)^2 + 2\phi^+ \phi^-] - \\
& \frac{1}{4} g^2 \frac{1}{c_w} Z_\mu^0 Z_\mu^0 [H^2 + (\phi^0)^2 + 2(2s_w^2 - 1)^2 \phi^+ \phi^-] - \frac{1}{2} g^2 \frac{s_w^2}{c_w} Z_\mu^0 \phi^0 (W_\mu^+ \phi^- + \\
& W_\mu^- \phi^+) - \frac{1}{2} ig^2 \frac{s_w^2}{c_w} Z_\mu^0 H (W_\mu^+ \phi^- - W_\mu^- \phi^+) + \frac{1}{2} g^2 s_w A_\mu \phi^0 (W_\mu^+ \phi^- + \\
& W_\mu^- \phi^+) + \frac{1}{2} ig^2 s_w A_\mu H (W_\mu^+ \phi^- - W_\mu^- \phi^+) - g^2 \frac{s_w}{c_w} (2c_w^2 - 1) Z_\mu^0 A_\mu \phi^+ \phi^- - \\
& g^1 s_w^2 A_\mu A_\mu \phi^+ \phi^- - \bar{e}^\lambda (\gamma \partial + m_e^\lambda) e^\lambda - \bar{\nu}^\lambda \gamma \partial u^\lambda - \bar{u}_j^\lambda (\gamma \partial + m_u^\lambda) u_j^\lambda - \\
& d_j^\lambda (\gamma \partial + m_d^\lambda) d_j^\lambda + ig s_w A_\mu [-(\bar{e}^\lambda \gamma^\mu e^\lambda) + \frac{2}{3} (\bar{u}_j^\lambda \gamma^\mu u_j^\lambda) - \frac{1}{3} (\bar{d}_j^\lambda \gamma^\mu d_j^\lambda)] + \\
& \frac{ig}{4c_w} Z_\mu^0 [(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (\bar{e}^\lambda \gamma^\mu (4s_w^2 - 1 - \gamma^5) e^\lambda) + (\bar{u}_j^\lambda \gamma^\mu (\frac{4}{3} s_w^2 - \\
& 1 - \gamma^5) u_j^\lambda) + (\bar{d}_j^\lambda \gamma^\mu (1 - \frac{8}{3} s_w^2 - \gamma^5) d_j^\lambda)] + \frac{ig}{2\sqrt{2}} W_\mu^+ [(\bar{\nu}^\lambda \gamma^\mu (1 + \gamma^5) e^\lambda) + \\
& (\bar{u}_j^\lambda \gamma^\mu (1 + \gamma^5) C_{\lambda\kappa} d_j^\kappa)] + \frac{ig}{2\sqrt{2}} W_\mu^- [(\bar{e}^\lambda \gamma^\mu (1 + \gamma^5) \nu^\lambda) + (d_j^\kappa C_{\lambda\kappa}^\dagger \gamma^\mu (1 + \\
& \gamma^5) u_j^\lambda)] + \frac{ig}{2\sqrt{2}} \frac{m_e^\lambda}{M} [-\phi^+ (\bar{\nu}^\lambda (1 - \gamma^5) e^\lambda) + \phi^- (\bar{e}^\lambda (1 + \gamma^5) \nu^\lambda)] - \\
& \frac{g}{2} \frac{m_e^\lambda}{M} [H(\bar{e}^\lambda e^\lambda) + i\phi^0 (\bar{e}^\lambda \gamma^5 e^\lambda)] + \frac{-ig}{2M\sqrt{2}} \phi^+ [-m_d^\kappa (\bar{u}_j^\lambda C_{\lambda\kappa} (1 - \gamma^5) d_j^\kappa)] + \\
& m_u^\lambda (\bar{u}_j^\lambda C_{\lambda\kappa} (1 + \gamma^5) d_j^\kappa) + \frac{ig}{2M\sqrt{2}} \phi^- [m_d^\lambda (\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger (1 + \gamma^5) u_j^\kappa) - m_u^\kappa (\bar{d}_j^\lambda C_{\lambda\kappa}^\dagger (1 - \\
& \gamma^5) u_j^\kappa)] - \frac{g}{2} \frac{m_u^\lambda}{M} H (\bar{u}_j^\lambda u_j^\lambda) - \frac{g}{2} \frac{m_d^\lambda}{M} H (\bar{d}_j^\lambda d_j^\lambda) + \frac{ig}{2} \frac{m_u^\lambda}{M} \phi^0 (\bar{u}_j^\lambda \gamma^5 u_j^\lambda) - \\
& \frac{ig}{2} \frac{m_d^\lambda}{M} \phi^0 (\bar{d}_j^\lambda \gamma^5 d_j^\lambda) + \bar{X}^+ (\partial^2 - M^2) X^+ + \bar{X}^- (\partial^2 - M^2) X^- + \bar{X}^0 (\partial^2 - \\
& M^2) X^0 + \bar{Y} \partial^2 Y + ig c_w W_\mu^+ (\partial_\mu \bar{X}^0 X^- - \partial_\mu \bar{X}^+ X^0) + ig s_w W_\mu^+ (\partial_\mu \bar{Y} X^- - \\
& \partial_\mu \bar{X}^+ Y) + ig c_w W_\mu^- (\partial_\mu \bar{X}^- X^0 - \partial_\mu \bar{X}^0 X^+) + ig s_w W_\mu^- (\partial_\mu \bar{X}^- Y - \\
& \partial_\mu \bar{Y} X^+) + ig c_w Z_\mu^0 (\partial_\mu \bar{X}^+ X^+ - \partial_\mu \bar{X}^- X^-) + ig s_w A_\mu (\partial_\mu \bar{X}^+ X^+ - \\
& \partial_\mu \bar{X}^- X^-) - \frac{1}{2} g M [\bar{X}^+ X^+ H + \bar{X}^- X^- H + \frac{1}{c_w^2} \bar{X}^0 X^0 H] + \\
& \frac{1-2c_w^2}{2c_w} ig M [\bar{X}^+ X^0 \phi^+ - \bar{X}^- X^0 \phi^-] + \frac{1}{2c_w^2} ig M [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \\
& ig M s_w [\bar{X}^0 X^- \phi^+ - \bar{X}^0 X^+ \phi^-] + \frac{1}{2} ig M [\bar{X}^+ X^+ \phi^0 - \bar{X}^- X^- \phi^0]
\end{aligned}$$

Figure 1.2: The Standard Model Lagrangian.

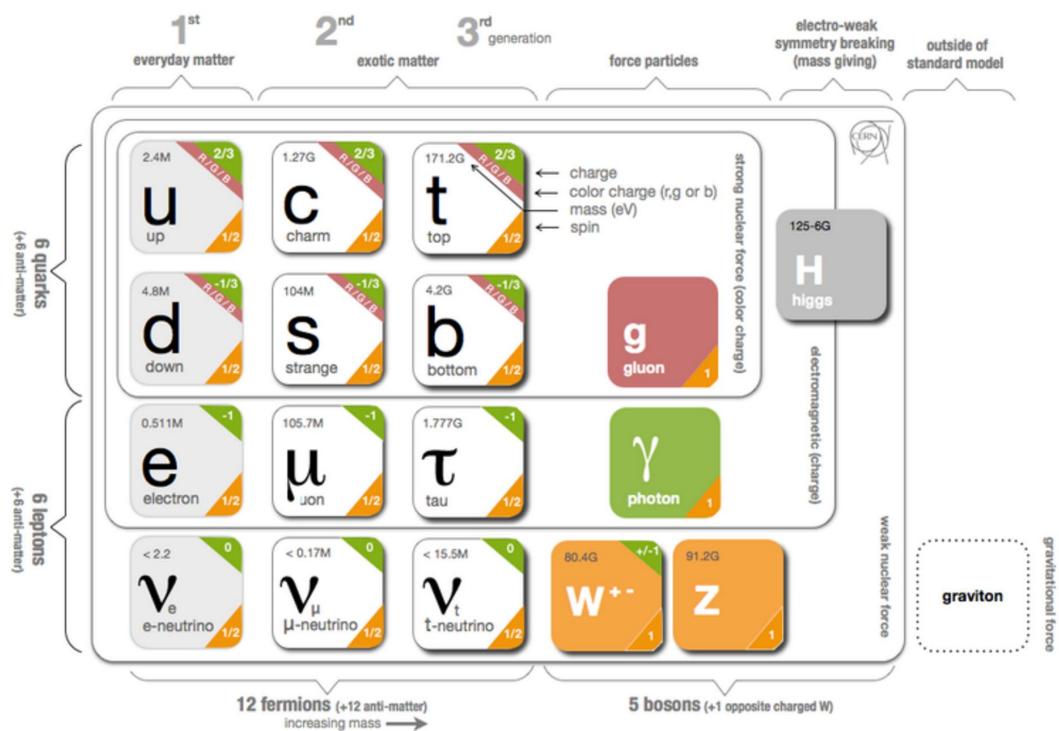


Figure 1.3: Table of Standard Model particles. Image from CERN [2].

1.1.1 Leptons

The electron is one of the most well known elementary particles. There exist two heavier “versions” of the electron: the muon (μ) which is ~ 200 times as massive as the electron, and the tau (τ) which is ~ 4000 times as massive as the electron. Each of these particles is negatively charged, where the charge is equivalent to -1 which corresponds to -1.602×10^{-19} Coulombs.

For each of these particles, there exists a corresponding neutrino: the electron ν_e , the muon neutrino ν_μ , and the tau neutrino ν_τ . The neutrinos as defined by the SM are chargeless and massless, although experiments have proved that at least two out of three neutrinos observed in nature have mass. The electron and its corresponding neutrino are first generation fermions, while the muon and its neutrino are second generation and the tau and its neutrino are third generation.

A corresponding anti-particle exists for each of these six particles: \bar{e} (e^+), $\bar{\mu}$ (μ^+), $\bar{\tau}$ (τ^+), $\bar{\nu}_e$, $\bar{\nu}_\mu$, and $\bar{\nu}_\tau$. Note that, while neutrinos are chargeless, the anti-particle for the electron, muon, and tau, all have positive (or opposite of their corresponding particle) charge. However, each anti-particle is identical in mass to its corresponding particle.

In addition to charge and mass, particles also have spin, or the intrinsic angular momentum of a particle. All of the above mentioned particles, known as leptons, have spin $\frac{1}{2}$ (or $-\frac{1}{2}$), which means they are fermions. The spin of a particle allows us to define helicity, also known as handedness. This is the sign of the projection of the particle’s spin vector onto its momentum vector.

A particle is said to be right-handed if the spin and momentum align, and left-handed if the spin and momentum point opposite each other. In the Standard Model, there can be difference between left-handed and right-handed particles. For example, we have observed that all neutrinos are left-handed, while all anti-neutrinos are right-handed, which is currently still inexplicable.

1.1.2 Quarks

Many of the particles discovered in the twentieth century are composite particles called hadrons, made up of quarks. There are six quarks: *up* (*u*) and *down* (*d*) (first generation), *strange* (*s*) and *charm* (*c*) (second generation), and *bottom* (*b*) and *top* (*t*) (third generation).

The *u*, *c* and *t* quarks have charge $+\frac{2}{3}$ and the *d*, *s* and *b* quarks have charge $-\frac{1}{3}$. Quarks also have each a different mass. For every quark there is an anti-quark (\bar{u} , \bar{d} , \bar{s}, \bar{c} , \bar{b} and \bar{t}) with an opposite charge. Quarks are also fermions, and therefore have a spin of $\frac{1}{2}$ or $-\frac{1}{2}$.

In addition to charge (by which we mean electric charge), mass, and spin, quarks also have a second kind of charge, known as color. While electric charge allows particles to interact through the electromagnetic force, color charge allows particles to interact through a different type of force, called the strong force. Quarks can be red, green, or blue colored, and anti-quarks can be anti-red, anti-green, or anti-blue colored. In nature, we only see color neutral particles, which is why quarks are only observed as combinations forming composite particles. These composite particles made up of quarks are color neutral because they either contain one quark of each color, or a quark of a particular color and a quark of the corresponding anti-color. For example, the proton is made up of

one u and two d quarks, so one of these must be red, one must be green, and one must be blue.

1.1.3 Representation of Interactions

The Standard Model describes how matter particles, described above, interact with each other through the exchange of force mediating particles. Each term in the Lagrangian in Figure 1.2 represents different types of interactions between particles. Let us consider a simple scenario of electron electron interaction, which happens through the exchange of the electromagnetic mediating particle known as the photon. Instead of considering the mathematical term associated with this process, it is easier to draw the actual interaction using a vertex, as can be seen in Figure 1.4. This vertex can be arranged and combined with

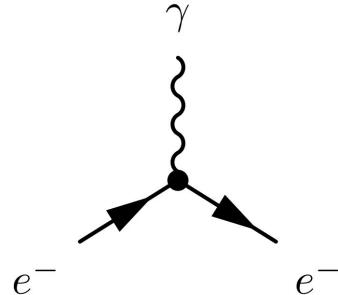


Figure 1.4: Visual representation of the interaction between two electrons (e^-) and a photon (γ).

other vertices in a variety of ways to represent different interactions between electrons, positrons, and photons. For example, we can consider the case where two electrons are propelled towards each other. Using two of the vertex pictured in 1.4, we can see that this must result in two electrons, as can be seen in Figure 1.5, where time propagates to the right on the x-axis, and distance is represented

on the y-axis. Alternatively, if we were to consider the interaction between an

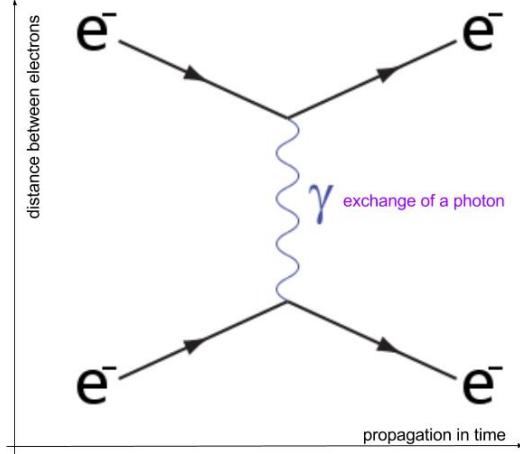


Figure 1.5: A Feynman Diagram showing the interaction of two electrons. Time increases to the right on the x-axis, and the y-axis represents the distance between the two electrons.

electron and a positron, which can be represented as an electron with an arrow pointing backwards in time, we can see that the resulting process is described by Figure 1.6, where an electron and a positron annihilate to a photon, which produces an electron positron pair. In both of these diagrams, we notice that charge is conserved on either side of the diagram. Many more diagrams can be made with this vertex, combining it with other vertices that involve a photon or electrons as well. In understanding diagrams like these, we can explain three of the four fundamental forces at a qualitative level, below, while also introducing new theories in later chapters that have motivated the search for new particles.

1.1.4 The Electromagnetic Force:

Quantum electrodynamics (QED) describes how the electromagnetic force behaves at the quantum level. The force carrier of QED is the photon, γ , which is

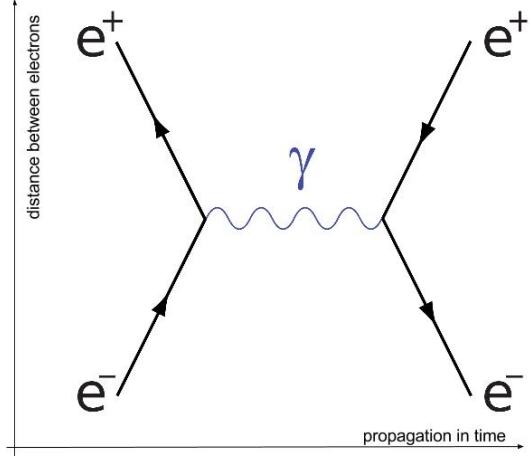


Figure 1.6: A Feynman Diagram showing the annihilation of a positron and an electron.

massless and moves at the speed of light. This theory can be described by diagrams like Figure 1.4, where any fermion with an electric charge can replace the electron. It should be noted that since the photon has neutral electric charge, it cannot couple to itself, and these diagrams with two electrically charged fermions and one photon represent the theory in full. Many experiments have performed precision measurements of QED properties, verifying that they match SM predictions extremely well.

1.1.5 The Strong Force:

Quantum Chromodynamics (QCD) describes quantum interactions related to the strong force. The force carrier of QCD is the gluon, g , which is massless and moves at the speed of light. The gluon does have color charge, just as the quarks do, and therefore there are a number of vertices in QCD, as shown in Figure 1.7. As mentioned before, each quark has a color charge, and each anti-quark has an anti-color charge. Since color must be conserved, we can see

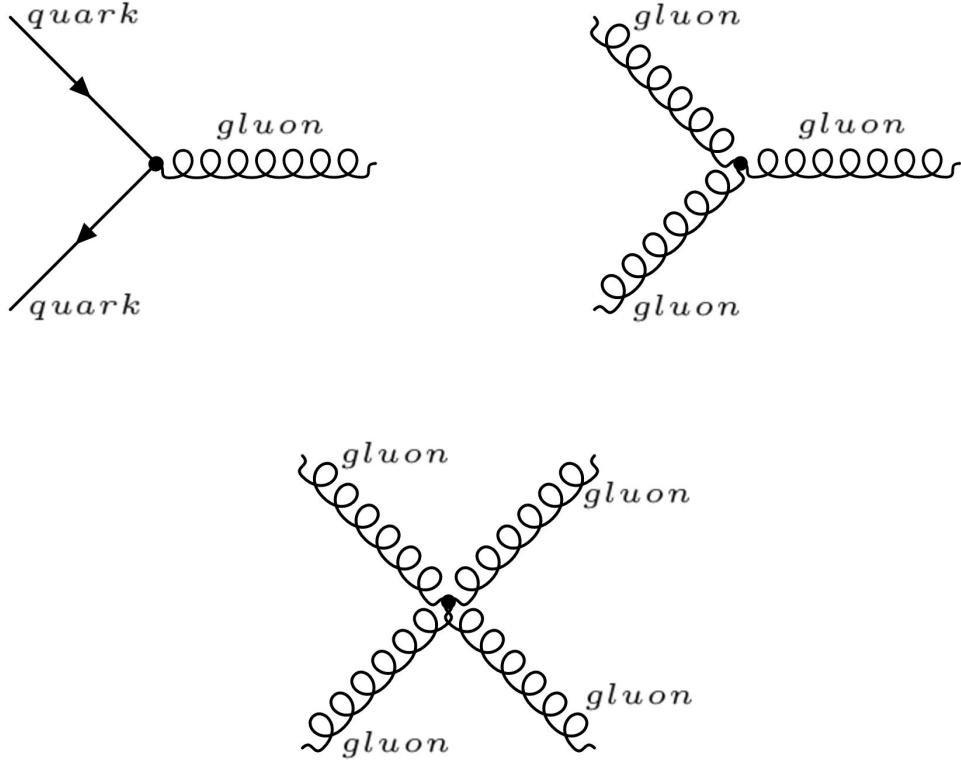


Figure 1.7: QCD interactions can be described by three vertices. The curly line represents gluons, while the straight lines represent quarks. Since gluons are colored, they can interact with themselves.

that the gluon must carry a color and an anti-color for the vertices in Figure 1.7 to be true. Naively, since there are three colors, one would think there are nine gluons, or nine possible combination of three colors and three anticolors. These are summarized below:

$$\begin{array}{cccc}
 \frac{r\bar{r}+b\bar{b}+g\bar{g}}{\sqrt{3}} & & & \\
 \frac{r\bar{b}+b\bar{r}}{\sqrt{2}} & \frac{r\bar{g}+g\bar{r}}{\sqrt{2}} & \frac{b\bar{g}+g\bar{b}}{\sqrt{2}} & \frac{r\bar{r}-b\bar{b}}{\sqrt{2}} \\
 \frac{-i(r\bar{b}-b\bar{r})}{\sqrt{2}} & \frac{-i(r\bar{g}-g\bar{r})}{\sqrt{2}} & \frac{-i(b\bar{g}-g\bar{b})}{\sqrt{2}} & \frac{r\bar{r}+b\bar{b}-2g\bar{g}}{\sqrt{6}}
 \end{array}$$

Both mesons (two-quark bound states) and baryons (three-quark bound states) must be color neutral. We know that a particle can be neutral if it

contains either quarks with color-anti-color, or all three color quarks. Then since the combination of red, green, and blue is neutral, the first combination listed, $\frac{r\bar{r}+b\bar{b}+g\bar{g}}{\sqrt{3}}$, would be color neutral. However, color neutral particles must be non-interacting, otherwise colorless baryons would emit these gluons and interact with one another through the strong force, which we do not observe in nature. Therefore, this first combination is not possible, and there are only eight gluons, called the color octet.

The strong force is responsible for hadronization, or the production of many quarks and gluons when quarks are smashed apart due to an event such as a proton-proton collision. As quarks drift apart after the collision, a color tube of self-interacting gluons is created between the quarks. These tubes are stretched as the quarks drift further apart, increasing energy in the tube due to the constant force exerted from stretching. At distances of roughly 10^{-15} m, it becomes more energetically favorable for two new quarks to be created from the vacuum, through the process shown in the top left vertex in Figure 1.7. If these new quarks are still too energetic to be contained in a particle, this process will repeat until the energy has been sufficiently decreased for all new quark pairs to stay bound. This property is known as confinement, and is critical to the understanding of hadronic activity within a particle detector.

1.1.6 The Weak Force:

When the universe had first begun and was still very hot, the weak force and the electromagnetic force were combined to form the electroweak force. At this time, the electroweak interaction was mediated by four massless bosons: W_1 , W_2 , W_3 and B . As the universe cooled, the bosons eventually began interacting

with the Higgs field (addressed in the following section), and soon, the bosons that interacted with fermions were no longer these four, but a superposition of them. We can write this superpositions as

$$\begin{pmatrix} \gamma \\ Z \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \quad (1.1)$$

$$W^\pm = \frac{(W_1 \mp iW_2)}{\sqrt{2}} \quad (1.2)$$

where θ_W , the mixing angle, is a parameter of the SM. Therefore, while γ is the force carrier for the electromagnetic force, the Z , W^+ , and W^- are the force carriers for the weak force. The weak interaction is interesting because it only acts on left-handed particles and right-handed anti-particles.

The vertexes describing the weak force interactions can be found in Figure 1.8. While the photon is massless, the three weak force carriers are not. The

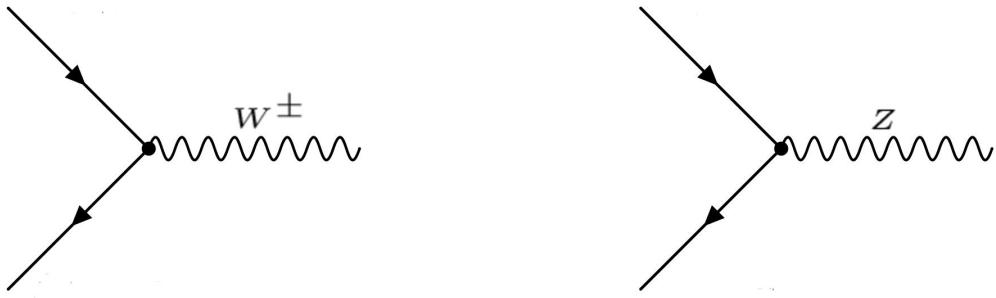


Figure 1.8: Weak interactions can be described by the following vertexes. The wavy lines represent bosons, while the straight lines represent any weak-interacting particle where the appropriate quantities are conserved.

Z boson has a mass of $91.2 \text{ GeV}/c^2$ (where protons have a mass of $\sim 1 \text{ GeV}/c^2$ for reference). The Z boson allows particles to interact with their anti-particles

through the weak force. For example, the process $Z \rightarrow e^-e^+$ is permitted, but the process $Z \rightarrow \mu^+e^-$ is not permitted. The Z boson can couple to both electrically-charged and -neutral particles, as well as particles with and without color.

The W^\pm boson, which has a mass of 80.4 GeV/c², also couples pairs of fermions. Rather than linking particles and anti-particles like the Z boson, it allows the flavors within a generation to interact. For example, a W^+ boson can decay to two quarks ($W^+ \rightarrow u\bar{d}$), and a W^- can decay to a lepton and its neutrino ($W^- \rightarrow \mu\nu_\mu$). Since the W^\pm is electrically charged, electric charge must be conserved at each vertex, and also $\gamma \rightarrow W^+W^-$ is allowed. The decay $Z \rightarrow W^+W^-$ is also a potential vertex.

The probability of a W decaying to different generations is dictated by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, where each of these values has been experimentally measured:

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \quad (1.3)$$

where $|V_{ij}|^2$ is the probability that a quark i decays to a quark j through the emission of a W.

Since all other particles decay through a W to a lighter particle, this explains why our world is built with the lightest two quarks (up and down quarks) and electrons, the building blocks of atoms.

1.1.7 Mass and the Higgs Boson

The Higgs Boson is the most recent confirmed Standard Model particle. It is related to the mass of other SM particles. To understand how this is so, we will look at the Higgs potential, which we can write as $V(\phi) \sim (\phi^2 - \eta^2)^2$ for the purpose of explanation. This is plotted in Figure 1.9. The potential

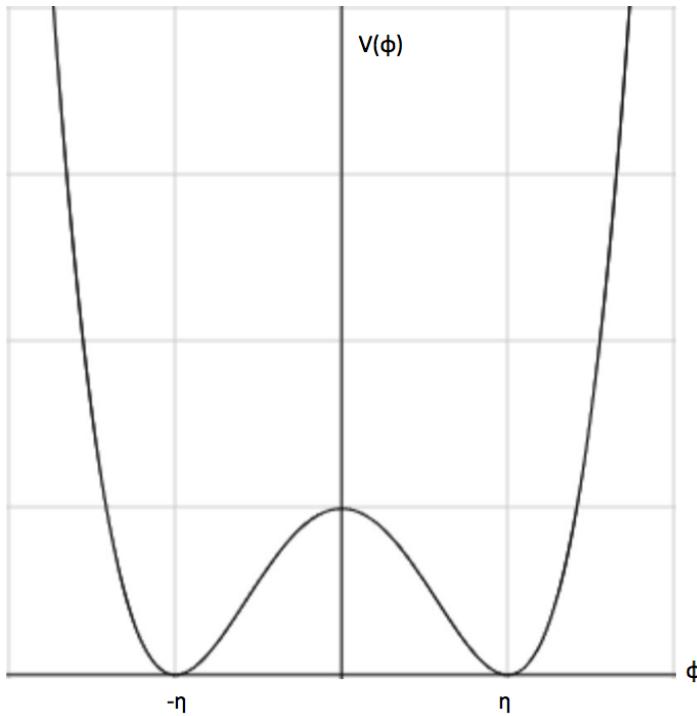


Figure 1.9: Approximation of the Higgs potential. While the shape is symmetric about the y-axis, the function is not symmetric about the minima.

is symmetric about the y-axis, or the axis with units of energy. In the early universe while everything was still very hot, the energies of particles were so high that the bumps at the bottom of the potential had no consequence. As the universe cooled, however, these minima became important, and one or the other minima was chosen. This is known as spontaneous symmetry breaking.

Once the universe had cooled sufficiently, the Higgs field took on a vacuum expectation value (VEV), or an average value in empty space, which was non-zero. This VEV couples to electroweak interactions, as referenced in the previous section, and the photon and weak force bosons mixed to form the states we know today. The spontaneously broken Higgs field also couples to itself, quarks, electrons, muons, taus, Z bosons, and W^\pm bosons. The magnitude of the coupling between the Higgs and a given particle determines the mass of the particle; the larger the coupling, the more massive the particle. However, it does not couple to neutrinos, so the explanation for their mass must lie elsewhere.

While we have been referring to the Higgs field, the Higgs boson is the particle that was recently discovered at the LHC in 2012. This particle is an excitation of the Higgs field in the same way that a photon is an excitation of the electromagnetic field. The SM Higgs has no charge, no spin, and a mass of 126 GeV. It couples to anything with mass (aside from the neutrinos), so it can also couple to itself. These vertexes take the form shown in Figure 1.10.

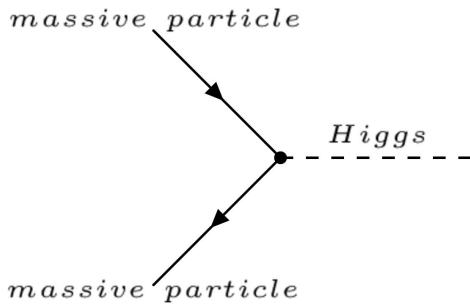


Figure 1.10: The Higgs Vertex. The dashed line is the Higgs Boson and the straight lines are any massive particle.

1.2 Rounding Up the Standard Model Particles

In total, including particles of different colors and all anti-particles, there exist 61 particles in the SM. We can describe their interactions easily using the vertices presented in each section. For example, consider the diagrams in Figure 1.11. These are important physics processes by which two gluons interact with

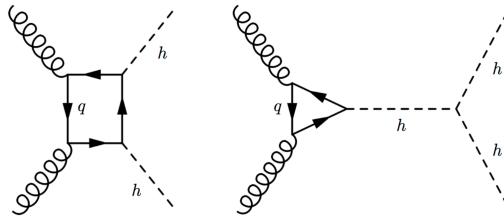


Figure 1.11: Left: Two gluons interact with a quark, while each gluon interacts with an additional different quark, and these additional different quarks interact with a shared fourth quark and each pair of quarks interacts to form a Higgs boson. Right: Two gluons interact with a quark, while each gluon interacts with an additional quark, and these additional different quarks interact with each other to form a Higgs boson, which interacts with two Higgs bosons. The result of both of these processes is a final state with two Higgs.

quarks to produce two Higgs, which is central to the topic of this thesis. This shows an example of the different kind of processes possible using Feynman diagrams: two protons collide and gluons inside of them interact with different quarks to produce two Higgs bosons (left) or one Higgs boson that then produces two Higgs bosons (right).

While this description has been qualitative, and has not drawn distinction between what has been experimentally determined and what has been predicted in the SM, the SM is quite self-consistent. There are only 19 parameters which are experimentally determined, most of which are masses.

1.3 Beyond the Standard Model

While the SM is an impressive feat of mathematics and experiments, it leaves many questions unanswered. For example, the SM only accounts for baryonic matter, which is estimated to be 4.6% of the universe. A remaining 24% is accounted for by dark matter, which can only be gravitationally detected, while the rest of the universe is said to be made up of dark energy.

The SM is also missing one of the four fundamental forces, gravity. While we have convincing understanding of quantum mechanics and general relativity, a unification for these two theories has yet to be discovered. Neutrino mass is unaccounted for by the SM. The Higgs mass is much lower than one might expect from the SM. The universe is dominated by matter, rather than anti-matter.

Many questions remain that require answers beyond the SM, and so we search for new particles and deviations in SM parameters at the LHC in hopes of providing new insight into these difficult problems.

Chapter 2

Theoretical overview

The discovery of a boson with a mass of approximately 125 GeV, and with properties close to those expected for the Higgs boson (H) of the SM [3, 4], has stimulated interest in the exploration of the Higgs potential, described in 1. The production of a pair of Higgs bosons within the SM is a rare process that is sensitive to the structure of this potential through the self-coupling mechanism of the Higgs boson, as discussed in the previous chapter. An effective way to look for new physics is to examine the production cross section of two Higgs at the LHC.

A cross section tells us the probability of a particular final state of events. This depends on the initial conditions of the collision, or the energy going into the collision and what you are colliding. The unit used for cross section is barn, with $1\text{b} = 100\text{fm}^2 = 10^{-28}\text{m}^2$. In the SM, the cross section for the production of two Higgs bosons in pp collisions at 13 TeV is $33.5 \pm 2.5/2.8 \text{ fb}$ for the gluon-gluon fusion process [5, 6, 7], which lies beyond the reach of analyses based on the first run of the CERN LHC. An increase in the cross section beyond SM expectation would be a smoking gun for new physics. This can happen in

one of two ways: there are new particles which decay to HH that contribute to the production of HH, or there are new processes, or additional vertices, that contribute to the production of HH. The first is called "resonant" production, since the increase in cross section is from a resonance (new particle), and the second is called "non-resonant", since it is a new Feynman diagram but not a new particle causing an increase in the cross section.

2.1 Resonant Production

Many theories beyond the SM (BSM) suggest different ways in which the cross section for the production of two Higgs would increase, based on the existence of heavy particles that can couple to a pair of Higgs bosons. Models with a warped extra dimension (WED), as proposed by Randall and Sundrum [8], postulate the existence of one spatial extra dimension compactified between two fixed points, commonly called branes. This would mean that in addition to our three spatial dimension and one temporal dimension, there exists a fifth dimension that's extremely small, such that it would be hard to observe this dimension. This fifth dimensional region between these two points, or branes, is often called the bulk. We define ϕ as the coordinate of this dimension, with the size parametrized by r_c , as can be seen in Figure 2.1. Then the metric for the full five-dimensional spacetime, which as it turns out solves Einstein's equations, can be written as

$$ds^2 = e^{-2kr_c\phi} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 d\phi^2 \quad (2.1)$$

This means that four-dimensional mass scales (the masses we measure) are

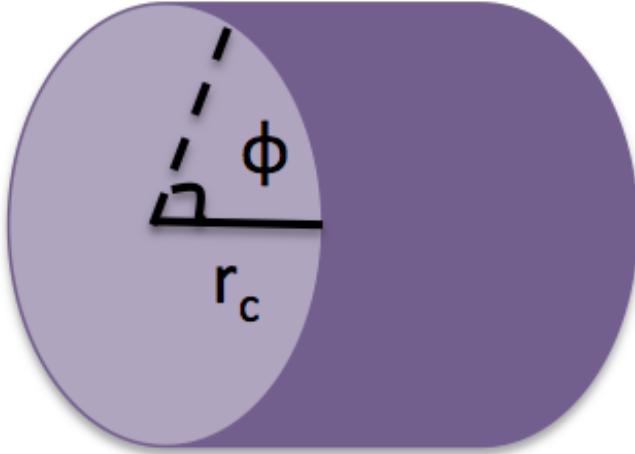


Figure 2.1: A depiction of the fifth extra dimension.

related to five-dimensional mass parameters (the masses predicted by the full five dimensional theory) by the warp factor, $e^{-2kr_c\phi}$. Therefore, this provides a good explanation as to why the Higgs boson mass is predicted to be on the order of the Planck scale ($M_{Pl} \sim 10^{18}$ GeV) but is observed to have a mass on the electroweak scale (125 GeV). This warp factor explains this relationship, without introducing a new hierarchy into the theory, since this large difference between the predicted and observed Higgs mass can be explained with a relatively small r_c . In this framework, we expect gravity to be much stronger in the bulk than in our four-dimensional world, which explains why gravity is observed to be so much weaker than the other three fundamental forces.

Lastly, this class of models predicts the existence of new particles. One of these particles would be the massless spin-2 graviton, which would give us insight into the inner workings of quantum gravity. There are also other new particles, such as the spin-0 radion [9, 10, 11], and the spin-2 bulk graviton [12, 13, 14].

The radion is a particle that helps stabilize the size of the extra dimension. We consider the case where no mixing between the radion and Higgs boson (in other words, the two have separate mass states unrelated to each other). The couplings of SM particles to the bulk graviton depend on where SM particles can be located. In this analysis, we consider a scenario where SM particles are allowed in the bulk [15].

Supersymmetry is a class of theories that predict a supersymmetric new particle for every SM particle that currently exists. Some supersymmetric models also predict one spin-0 resonance that, when sufficiently massive, decays to a pair of SM Higgs bosons. Those would be additional Higgs bosons [16, 17]. The signal modeling for a spin-0 particle is identical if it is a radion or an additional Higgs boson.

In searching for the bulk graviton and radion, we will compare the cross section of di-Higgs production we observe in data to the cross sections predicted by these theories. If no new particles are found, then depending on the sensitivity of the analysis and the value of the cross section for bulk graviton and radion decay to di-Higgs, we may be able to rule out these particles below a certain mass. For the bulk graviton and radion signals that we consider, the tools used to calculate the cross sections for the production of KK graviton in the bulk and RS1 models are described in Ref. [18, 19]. The implementation of the calculations is described in Ref. [20].

Searches for narrow particles decaying to two Higgs bosons have already been performed by the ATLAS [21, 22, 23] and CMS [24, 25, 26] collaborations in pp collisions at $\sqrt{s} = 7$ and 8 TeV. Until now their reach was limited to $M_X = 1.5$ TeV. Moreover, some of the models that predict the coupling of the new

resonance to HH also expect it to couple to W^+W^- or Z^0Z^0 [27]. Searches for these final states were performed by ATLAS and CMS [28, 29, 30, 31, 32].

2.2 Nonresonant Production

In the SM, non resonant pair production occurs primarily through gluon-gluon fusion via an internal fermion loop, which is dominated by the top quark, as can be seen in Figure 2.2. Assuming there are no new light states that we have

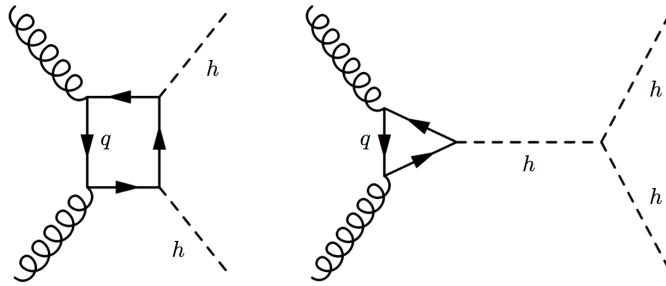


Figure 2.2: SM Feynman diagrams that contribute to Higgs boson pair production by gluon-gluon fusion at leading order.

not discovered yet, the gluon fusion Higgs boson pair production at the LHC can be described by five parameters, to leading order: κ_λ , κ_t , c_g , c_{2g} , and c_2 . The Higgs boson tri-linear coupling ($\lambda_{SM} = m_h^2/2v^2$, where v is the vacuum-expectation value of the Higgs field), and the top Yukawa interaction exist in the SM Lagrangian. Any deviations from the SM value for these two quantities is parametrized by κ_λ and κ_t , respectively. However, the interaction of Higgs and gluons, as well as two Higgs and two gluons or $t\bar{t}$, are not predicted by the SM. These are instead parametrized by the absolute couplings c_g , c_{2g} , and c_2 .

We can then write out the Lagrangian describing all of these SM interactions and potential BSM interactions, assuming no other light states besides for SM particles, as an effective field theory:

$$\begin{aligned} \mathcal{L}_h = & \frac{1}{2}\partial_\mu h\partial^\mu h - \frac{1}{2}m_h^2 h^2 - \kappa_\lambda \lambda_{SM} v h^3 \\ & - \frac{m_t}{v}(v + \kappa_t h + \frac{c_2}{v} h h)(\bar{t}_L t_R + h.c.) + \frac{1}{4} \frac{\alpha_s}{3\pi v} (c_g h - \frac{c_{2g}}{2v} h h) G^{\mu\nu} G_{\mu\nu}, \end{aligned} \quad (2.2)$$

where the first two terms are the kinetic and mass term respectively for the Higgs, the third term is related to SM Higgs self-interactions (parametrized by κ_λ), the fourth term is related to both SM Higgs-top-antitop interactions (parametrized by κ_t) and BSM Higgs-Higgs-top-antitop interactions (parametrized by c_2), and the last term is related to BSM Higgs-gluon-gluon interactions (parameterized by c_g) and BSM Higgs-Higgs-gluon-gluon interactions (parameterized by c_{2g}). In the SM, $\kappa_\lambda=\kappa_t=1.0$ and the other three parameters are set to 0. The Feynman diagrams contributing to the di-Higgs signal at leading order can be found in Figure 2.3.

This shows that the phase space for the Higgs boson couplings in the BSM scenario has 5 parameters, where constraints come from measurements of single Higgs boson production and other theoretical considerations. For example, assuming electroweak symmetry is linearly realized, it turns out $c_{2g} = -c_g$ [33, 34]. While this phase space is large, the kinematics of di-Higgs production also depend on these five parameters. In particular, the distribution of the di-Higgs invariant mass and the modulus of the cosine of the polar angle of one Higgs boson with respect to the beam axis (modulus of the direction the protons come from), are similar across many different parameter combinations. A statistical

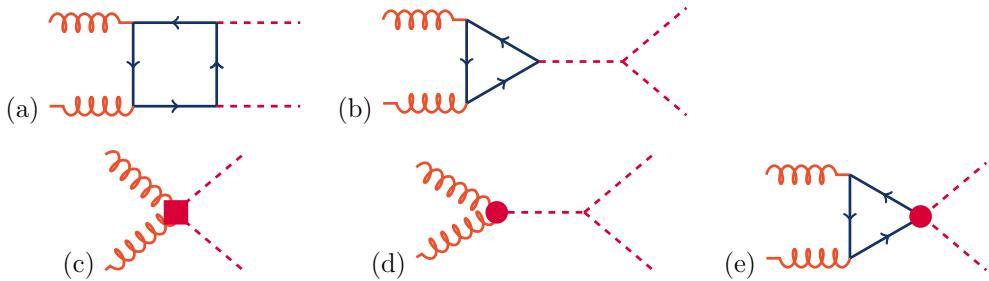


Figure 2.3: SM Feynman diagrams that contribute to Higgs boson pair production by gluon-gluon fusion at leading order. Diagrams (a) and (b) correspond to SM-like processes, while diagrams (c), (d), and (e) correspond to pure BSM effects: (c) and (d) describe contact interactions between the Higgs boson and gluons, and (e) exploits the contact interaction of two Higgs bosons with top quarks.

approach was developed to identify twelve clusters of models with similar distribution in both kinematic variables, described in full in Reference [35]. Twelve benchmarks, one from each cluster, were chosen to as the model that best represents each cluster. These are described in Table 2.1, along with the SM values of the parameters. The invariant mass distributions of the twelve BSM scenarios can be found in Figure 2.4, while the modulus of the cosine of the polar angle of one Higgs boson with respect to the beam axis ($|\cos\theta^*|$) can be found in Figure 2.5.

There is still much to be learned about the Higgs boson, the potential of the Higgs field, and other properties related to the Higgs. In particular, probing the Higgs potential by examining non-resonant production of di-Higgs is one of the most important tasks that the next generation LHC will tackle. While we don't currently have enough sensitivity to effectively probe the Higgs potential, we

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

Table 2.1: Parameter values of the final benchmarks selected with number of clusters $N_{clus} = 12$.

are able to test out strategies to find the best way to probe the Higgs potential in the future.

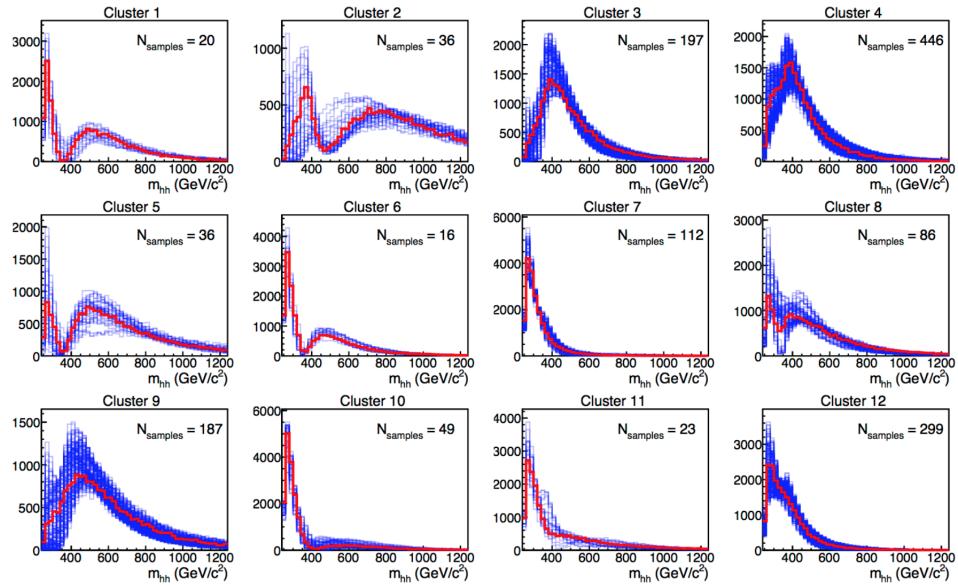


Figure 2.4: The invariant mass of the two Higgs for Monte Carlo simulation of different parameter combinations, clustered into 12 classes of shapes. The red distributions correspond to the benchmark model, while the blue distributions are the other parameter combinations similar to the benchmark that are part of that cluster.

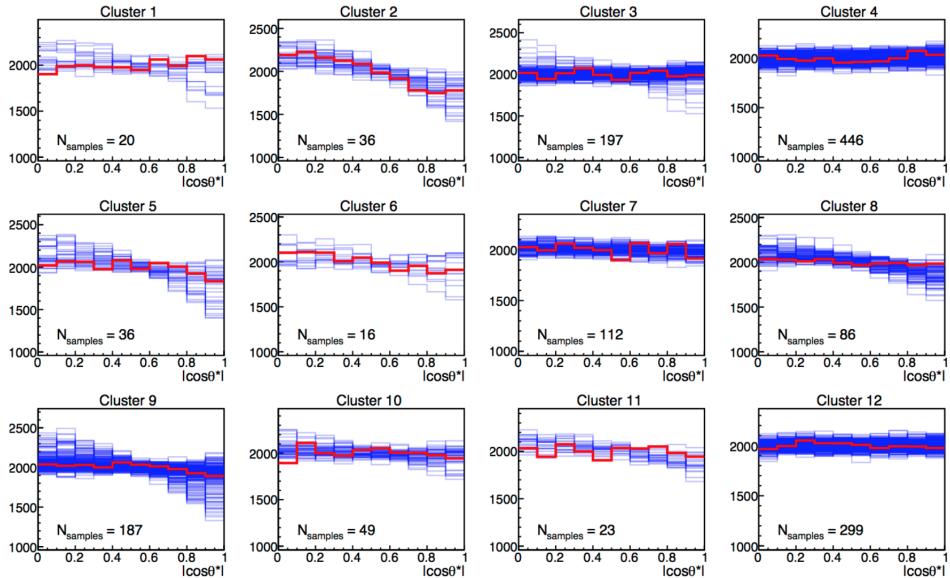


Figure 2.5: The modulus of the cosine of the polar angle of one Higgs boson with respect to the beam axis for Monte Carlo simulation of different parameter combinations, clustered into 12 classes of shapes. The red distributions correspond to the benchmark model, while the blue distributions are the other parameter combinations similar to the benchmark that are part of that cluster.

Chapter 3

Searching for New Particles

One of the most effective ways to probe the SM further and have a chance at discovering new BSM particles is to smash SM particles together at high energies in hopes of discovering something new. One of the easiest colliders to imagine is an electron-positron collider. Such colliders have existed over the past century and the experiments performed there account for many precision SM measurements. High energy collisions allow for a greater chance of something new and massive appearing, and allow for more precise measurements. However, it is hard to produce high energy collisions with electrons and positrons. Charged particles radiate energy when they are traveling through an electromagnetic field, where the power radiated is given by

$$P = \frac{e^4}{6\pi m^4 c^5} E^2 B^2 \quad (3.1)$$

Due to the low mass of electrons and positrons, they radiate much more energy than a heavier particle would. In order to reach high energies, then, colliding protons provides the most effective solution to date, and this is exactly what the Large Hadron Collider at CERN was built to do.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [36] is the largest collider ever built, colliding protons at the highest energy produced inside a collider, and producing more data than all other experiments combined. The LHC is located in a tunnel 100m underground on the border of Switzerland and France at CERN (European Center for Nuclear Research), near Geneva, Switzerland. It has been running since 2007, colliding protons at a center of mass energy 7 TeV from 2007-2011, 8 TeV from 2011-2015, and now at 13 TeV.

In order to create the proton beams used for collision in the LHC, the energy of the protons used for collisions is ramped up slowly. Hydrogen is stripped of electrons and fed into a linear accelerator, and then a succession of increasingly large circular synchrotrons. A schematic of the LHC and the smaller accelerators that feed into the main ring can be found in Figure 3.1, along with an aerial view of CERN. Once the two beams of protons enter the LHC, they are accelerated to

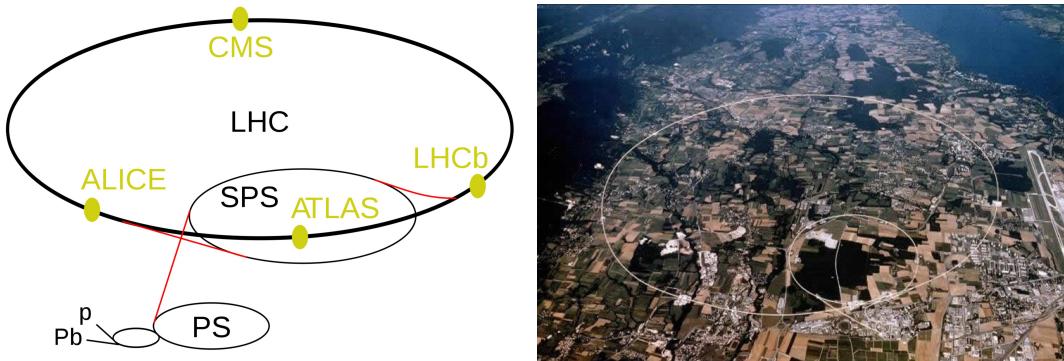


Figure 3.1: (left) Schematic of the LHC and the smaller accelerators which feed into the LHC. (right) Aerial view of Geneva. The larger circle marks the location of the LHC, safely underneath the ground.

collision energies (currently 13 TeV, or 6.5 TeV per beam). These beams travel

in opposite directions around the ring, which contains 1,232 dipole magnets to direct and accelerate the beams and 392 quadrupole magnets to focus the beams. The superconducting magnets are cooled to less than 2K and produce a 7T field. The protons travel in bunches of roughly 115 billion each, with a spacing of 25 nanoseconds between crossings, barreling along at 99.9999999% the speed of light to collide within four detectors: ALICE, ATLAS, CMS, and LHCb.

3.1.1 A proton-proton Collision

Because protons are composite particles, their collisions are not simple. While we speak of protons as being comprised of two up quarks and one down quark, these are only the valence quarks of the proton. The inner-workings of a proton are actually a complicated, ongoing interaction between these three valence quarks, where this interaction involves gluons. These gluons also interact with each other, spontaneously producing different quarks which interact to become gluons again. Therefore, when protons are collided at the energy scales of a machine such as the LHC, we are actually colliding quarks and gluons, rather than protons. We call these quarks and gluons partons.

This makes it much more difficult to know what is actually interacting in each collision. We rely on Parton Distribution Functions [37, 38] for a statistical understanding of what may be happening. Figure 3.2 shows two different simulated PDF sets. The x-axis represents the fraction of energy of the proton that belongs to a particular parton, while the y-axis tells us the probability of a particular parton having this particular energy. The various curves represent

different partons, where the subscript "v" indicates a valence quark. The left plot shows a low energy scale, while the right plot is for the LHC energy scale. For example, at the LHC scale, there is a $\sim 50\%$ probability that an up valence quark is carrying 10% of the proton's energy. It is of interest to note that at

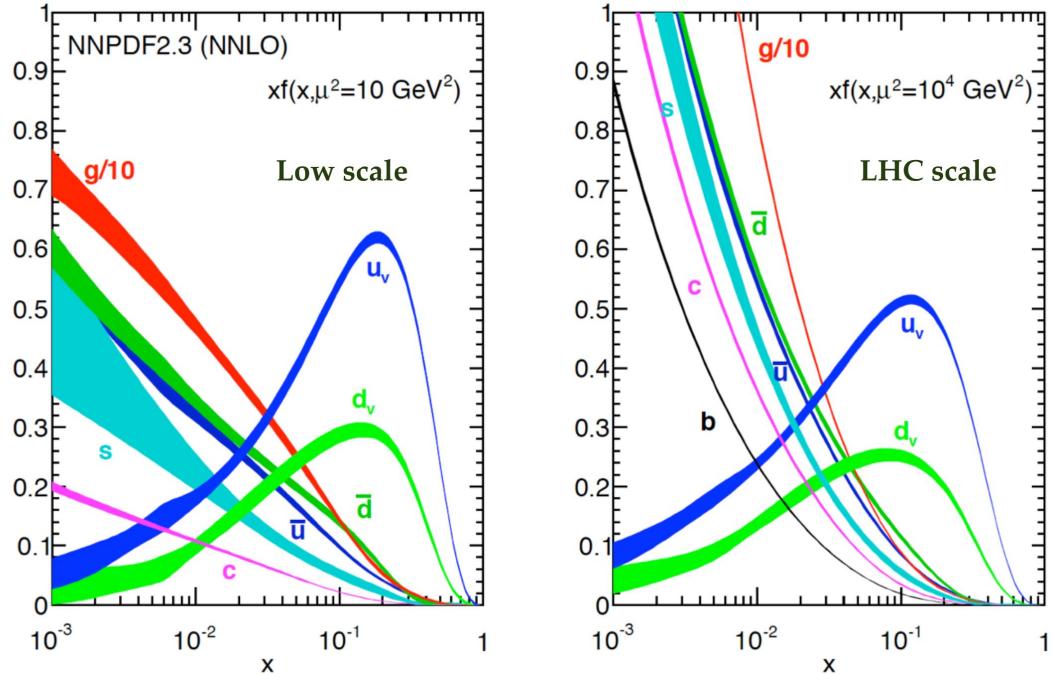


Figure 3.2: PDF for 10 GeV (left) and 10 TeV (right) collisions. Each curve represents a parton in a proton, and tells us the probability (y-axis) of finding that parton carrying momentum fraction x of the total momentum of the proton.

higher energies, gluon-gluon or quark-gluon interactions occur more often than they do at lower energies. It is important to have correct PDFs to predict the amount of new physics one might see at the LHC. For example, the non-resonant production considered in this thesis is only through gluon-gluon interactions, so this must be taken into account when producing simulated signal events. Unfortunately, it is hard to compute PDFs, so in order to have the best possible idea of what PDFs look like, partial models are combined with many measurements

performed at fixed target and collider experiments to get more accurate values. There is a systematic uncertainty associated with this process which will be discussed in more detail in later chapters.

3.2 The Compact Muon Solenoid

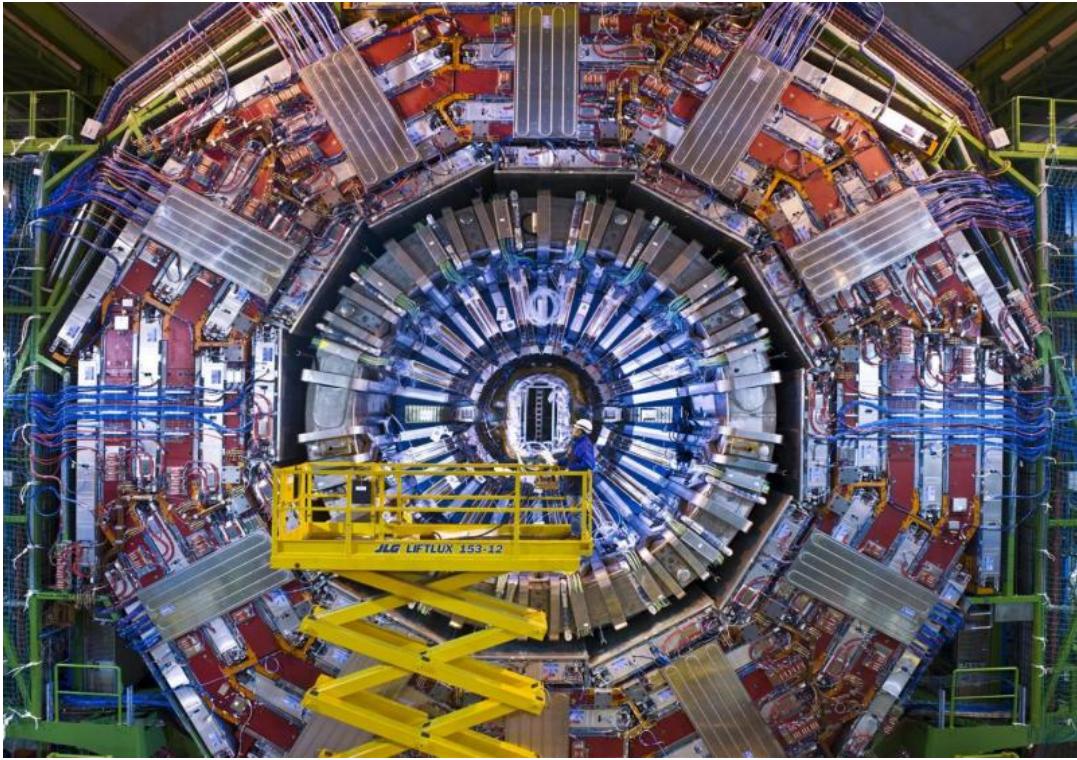


Figure 3.3: The Compact Muon Solenoid detector.

In order to make use of these proton-proton collisions, a carefully designed detector is necessary to capture the data. There are two general purpose detectors on the LHC, which serve to look for new physics and test current SM predictions. The analysis presented in this thesis was performed on data taken with one of these two detectors: the Compact Muon Solenoid (CMS)[39], as

can be seen in Figure 3.3. Armed with a team of roughly 4,000 scientists, this five-story, 14,000 ton piece of hardware has been the subject of many previous theses and papers. A brief overview is presented here so as to give context to the data analysis performed in this thesis. A slice of the detector is sketched in Figure 3.4; each component will be described in the following sections to build an understanding of the detector as a whole.

3.2.1 Coordinates

The CMS detector geometry approximates that of a large cylinder, with detectors in the barrel of the cylinder and at either end (endcaps). Coordinates in the detector are defined as shown in Figure 3.5. The z-axis is parallel to the beam line; protons come in from the +z-axis and -z-axis. The y-axis points directly up and the x-axis points directly sideways. The azimuthal angle ϕ is in reference to the x-axis, where $\phi = 0$ is along the x-axis. The polar angle θ is parametrized in terms of pseudorapidity η , where

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (3.2)$$

We use η instead of θ because pseudorapidity is a Lorentz-invariant quantity, meaning that it is the same in any Lorentz frame. This is critical for any quantity measured with respect to the beam line, since partons may not be symmetric in the laboratory rest frame because they may carry different fractions of energy of their respective protons.

Three quantities are recorded for each particle passing through the CMS detector: azimuthal angle ϕ , pseudorapidity η , and transverse momentum p_T ,

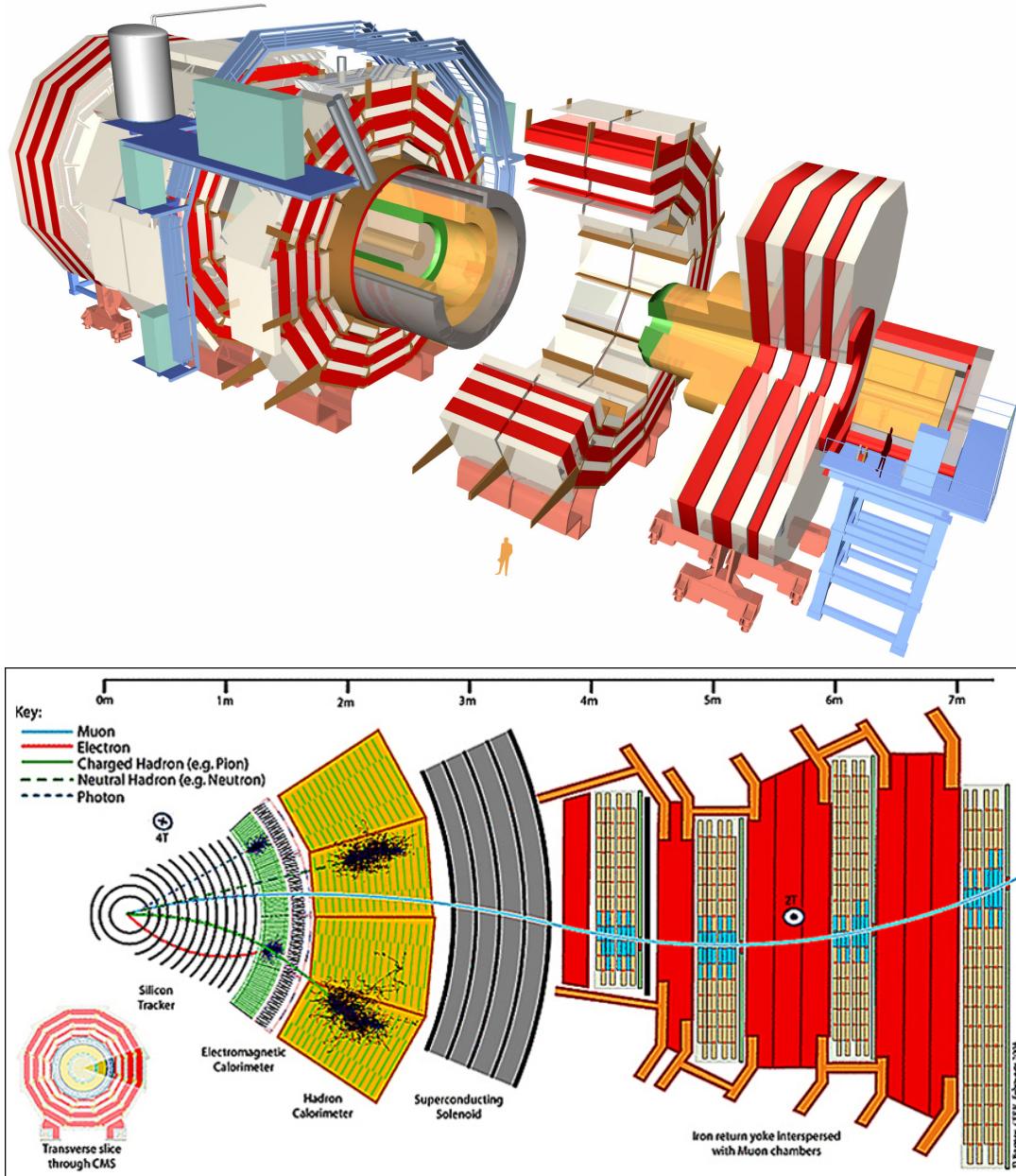


Figure 3.4: CMS detector with a person for size (top). Slice of the CMS detector, depicting the different components and the signatures of some of the different particles that pass through the detector after a proton-proton collision occurs (bottom).

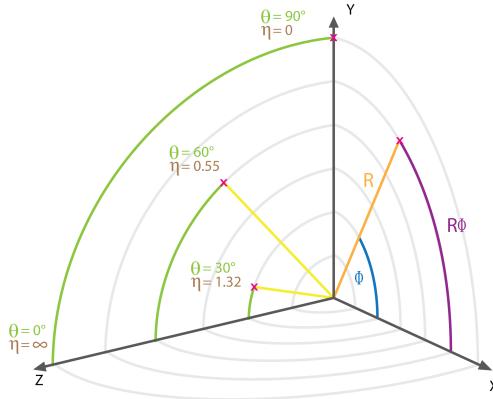


Figure 3.5: Coordinates of the CMS detector.

defined as $p_T = \sqrt{p_X^2 + p_Y^2}$, or the momentum in the plane transverse to the beam line. Transverse momentum is used rather than momentum because the initial state of collisions has $p_T = 0$, while p_Z is hard to determine. Momentum in Cartesian coordinates can be calculated with the following transformations,

$$p_X = p_T \cos \phi, \quad p_Y = p_T \sin \phi, \quad p_Z = p_T \cosh \eta, \quad (3.3)$$

although this is rarely necessary.

3.2.2 The Tracker

The innermost layer of the detector is a tracker made out of silicon. The tracker records the paths of charged particles as they pass through. It has 13 layers in the barrel, and 14 in the endcaps. The first four layers are made up of silicon pixels, while the remaining layers are strips. Each pixel is $100 \times 150 \mu\text{m}$ in area, while each strip is $180 \mu\text{m}$ by 10 cm or 25 cm, depending on where the strip is located. While this adds up to over 200 m^2 of silicon sensors, the detector itself

is not much larger than a shoe box.

Charged particles passing through the detector interact with the silicon to produce "hits" which allows us to determine a particle's location to within $10\mu\text{m}$. Because the tracker is inside a strong magnetic field (described below), the tracks of these charged particles bend. This allows us to make a measurement on the charge and momentum of the particle; the higher momentum the particle, the less curved the track.

3.2.3 The Calorimeters

The electromagnetic calorimeter is next to the tracker, followed by the hadronic calorimeter. While the purpose of a tracker is to allow a particle to pass through the detector in order to record the full track of a particle, the purpose of a calorimeter is to stop a particle so that a full measurement of the particle's energy can be obtained. Unlike the tracker, both charged and neutral particles interact with the calorimeter.

The electromagnetic calorimeter (ECAL) is built to measure the energy deposited by electrons, positrons, and photons. It is made of $\sim 80,000$ lead-tungstate (PbWO_4) crystals. Lead-tungstate crystal emits light, or scintillates, when particles deposit energy in the compound. It was chosen for its short radiation length and short Moliere length, both of which force particles to stop faster, allowing for a compact ECAL. By measuring the light produced by a particle in the ECAL, we can measure the energy of the particle. Particles heavier than an electron or photon continue on through the detector to be measured by the hadronic calorimeter.

The hadronic calorimeter (HCAL) is built to measure the energy deposited by hadrons. It is comprised of layers of an absorber, which causes an electromagnetic shower when a particle deposits energy, alternated with layers of an active medium, which measures the light emitted in the electromagnetic shower. The absorber chosen for the HCAL is brass, made out of Russian naval shells recycled from WW2, and chosen for its short interaction length (once again allowing for a more compact detector) and because brass is non-magnetic. The active medium is a plastic scintillator, which allows us to measure the amount of light caused by both charged and neutral hadron interactions with the brass, and therefore allows us to measure the energy of these particles.

3.2.4 The Solenoid

By choosing the material for the inner three components carefully, the CMS design allows for all three of these detectors to fit within a large solenoid magnet (hence "**Compact Muon Solenoid**", which produces a 3.8 Tesla magnetic field. This field causes the tracks of charged particles to bend within the inner three detectors, allowing us to make momentum measurements from tracks in the tracker.

3.2.5 The Muon Chambers

The last and largest component of the detector is the muon chambers. It is designed to measure the location, momentum, and energy of muons, which are too heavy to be stopped by the ECAL but not heavy enough to be stopped by the HCAL. This detector is comprised of three different types of detectors

(drift tube chambers, cathode strip chambers, resistive plate chambers), all of which operate on the same principle: as the muon travels through the detector, it knocks electrons off of gas atoms, which are then collected to measure the energy and location of the muon. These detectors are alternated with layers of steel, which stop non-muons from passing through. These layers also direct and contain the weak magnetic field outside of the solenoid, which allows for a precise momentum measurement of the muon (hence "Compact **Muon** Solenoid"). This is done by measuring the muon's curvature both in the tracker and in the muon chambers (it curves opposite directions in each of these, as can be seen in Figure 3.4).

3.2.6 What's Missing?

Some particles, such as neutrinos, are so weakly interacting that they pass right through the detector. We would potentially expect the same from some BSM particles, such as dark matter. However, because of the law of conservation of energy, we can infer a direction and magnitude for these missing particle(s). In the x and y direction, or the transverse plane, which is perpendicular to the beam line, the initial state condition requires that total momentum is 0, where all of the momentum is in the z-direction (Figure 3.6, top). Having measured all particles that are registered by the detector carefully after the collision, with detectors that leave very little space uncovered for a particle to slip through, we are then able to reconstruct whether there is any missing energy in the transverse plane based off of energy conservation laws (Figure 3.6, bottom). While we can't say anything definitive about what particle(s) is(are) missing,

we can report the total missing transverse momentum p_T and the azimuthal angle (ϕ) associated with this missing energy.

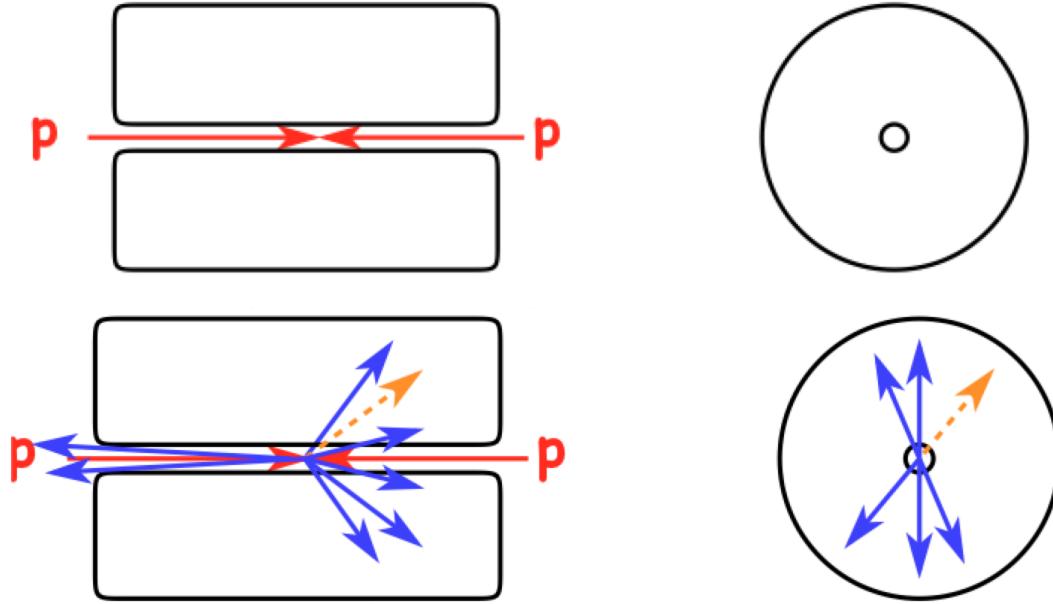


Figure 3.6: Diagram of particles inside the detector pre-collision (top) and post-collision (bottom), where post-collision shows where the protons entered, which particles were recorded (blue), and the direction of the missing energy (orange).

3.2.7 The Particle Flow Algorithm

Information from all of the detectors is collected and analyzed by the Particle Flow Algorithm (PF Algorithm, or PF) [40], which allows for an accurate event by event reconstruction by combining information from detectors, rather than treating the information separately.

3.2.8 Jets

As mentioned in Chapter 1, all particles that exist in nature must be color-neutral. When proton-proton collision result in bare quarks or gluon, the process of hadronization begins immediately, and new quarks are produced from the vacuum until it is no longer energetically favorable. This creates a spray of hadronic activity; wherever one quark or gluon is, many more will exist in a conical structure that points in the direction the original quark or gluon was moving. A depiction of these in a real event at the LHC is shown in Figure 3.7, where the yellow cones represent these hadronic showers, which are called jets. Jets are

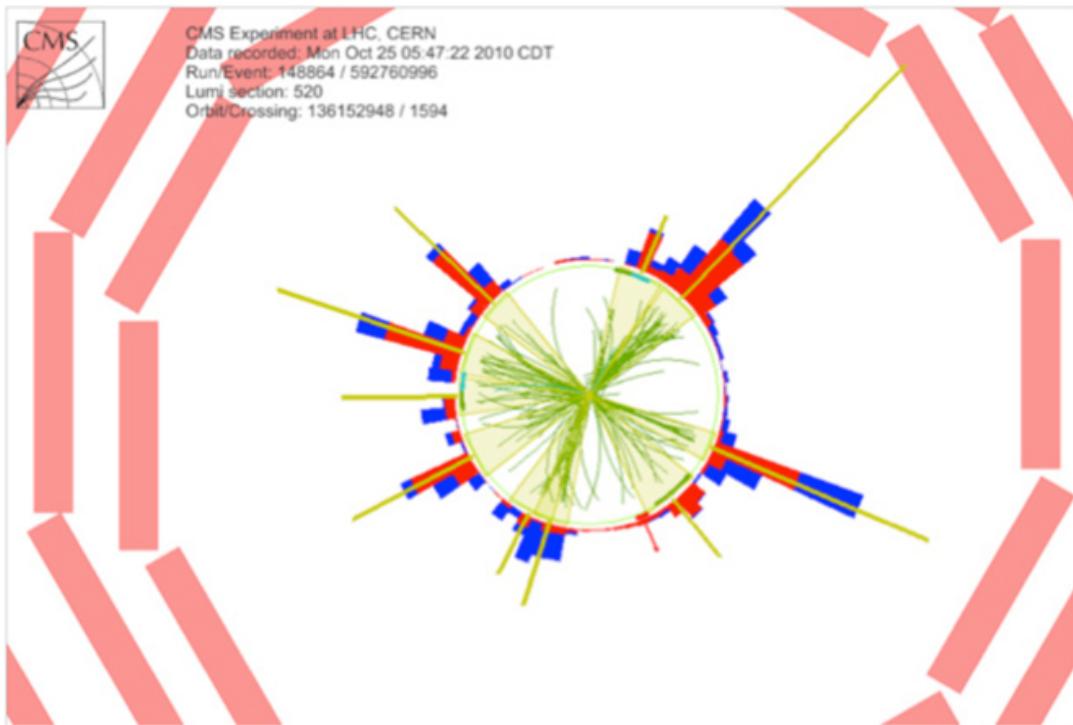


Figure 3.7: Typical CMS event with large number of hadrons (tracks represented by green lines, HCAL energy in blue, ECAL energy in red). These hadrons are collected into *jets* (shown by yellow triangles). The PF algorithm combines information from the trackers and calorimeters to properly measure the total momentum and energy of the jet.

composed of constituents which are defined by combining tracker, calorimeter, and muon system information from the PF Algorithm. These constituents are clustered into jets, making educated guesses as to which constituents belong to which jets. A number of algorithms exist for clustering, but the jets used in this analysis are clustered with the anti- K_T algorithm [41]. Using this method, every PF candidate, or particle reconstructed with the PF Algorithm, is compared with all of the other candidates, measuring a distance-like parameter between each pair. The two closest constituents are paired to become a new constituent. This process continues until the distance-like parameter between the jet and the beam is equal to $1/p_T^2$ of this new conglomerate constituent. Then this constituent is considered a jet and is removed from consideration. This continues until no constituents remain.

The anti- K_T algorithm tends to produce conical jets with smooth, rounded edges. Jets of this type are referred to as AKR jets, where R is the radius of the jet. Two types of jet are used in this analysis: AK8 and AK4, with R=0.8 and R=0.4 respectively, depending on how collimated we expect decay products to be.

3.2.9 Triggers

The detector output from each collision would require roughly a megabyte of space to store. Given that roughly 600 million collisions happen per second, there is no way to store all of the data. In fact, over 99% of data is thrown out, and only the interesting events are kept. The trigger system is responsible for determining which events matter, and which do not.

There are two stages of the trigger system. The Level 1 Trigger is comprised of hardware. Detector output is temporarily stored and analyzed to look for "interesting" physics, such as large amounts of HCAL energy or a high energy muon. This stage removes 99% of collisions, passing the remaining events to the next stage.

The High Level Trigger (HLT), comprised of software, analyzes the output of the Level 1 trigger. Events are reconstructed more fully and sorted into different categories should they pass the HLT criteria. Any remaining events that do not pass selection are discarded. Different HLT paths are associated with different interesting physics. For example, events with a muon with $p_T > 50$ GeV would pass the *HLT_Mu50* trigger while events with total transverse energy, known as H_T , larger than 800 would pass the *HLT_HT800*. If an event passes both criteria, then it passes both triggers. Most triggers retain all events passing the required criteria for that particular trigger, but some triggers, called prescale triggers, pass only a fraction of the events that pass the criteria. These triggers tend to have looser criteria and are designed to be used for measuring the efficiency of unprescaled triggers for a given analysis. The comparison of this efficiency measurement in data and simulation plays an important role in systematics, which will be discussed in later chapters.

3.2.10 Pileup

While we tend to speak of collisions at the LHC as a clean event, in which two protons enter the detector and collide, this is not entirely accurate. The protons come in bunches, and therefore, multiple protons can collide in the same event.

Some events can even contain up to forty individual interactions. Because of this, we define a primary vertex for the event, which is a point along the beam line from which the highest value of the sum of square of the p_T of Particle Flow objects originate.

Pileup, or the rest of the event aside from the interaction originating from this primary vertex, can affect the algorithms for creating jets. We account for the effects of pileup on jet related quantities such as mass in the evaluation of systematics, discussed in later chapters.

Chapter 4

Looking for New Physics

In the previous chapters we learned about the Standard Model, interesting new physics that we can test for at the LHC related to the Higgs boson, and the design of the LHC and CMS. Armed with all of this information, we can now understand how to go about finding this new physics (or proving that it does not exist in a certain phasespace).

4.1 Signal Signature

We are looking for events in CMS data that have two Higgs boson. Inside the detector, the Higgs boson immediately decays, and we observe its decay products rather than the Higgs itself. The most common decay mode for a Higgs boson is into two b quarks, so this is what we choose to look at in our analysis. Since there are two Higgs bosons, we are looking for events with four b-quarks. B-quarks show up in the CMS detector as jets. However, the number of jets we expect to see is dependent upon how much momentum each Higgs boson has. The more momentum a particle has, the more collimated its decay products tend to be. In Figure 4.1, we see that if a particle is produced

with no momentum (left), its decay products will be scattered in the detector. However, as a particle is produced with higher (middle) and even higher (right) momentum, this Lorentz boost causes the decay products to become collimated.

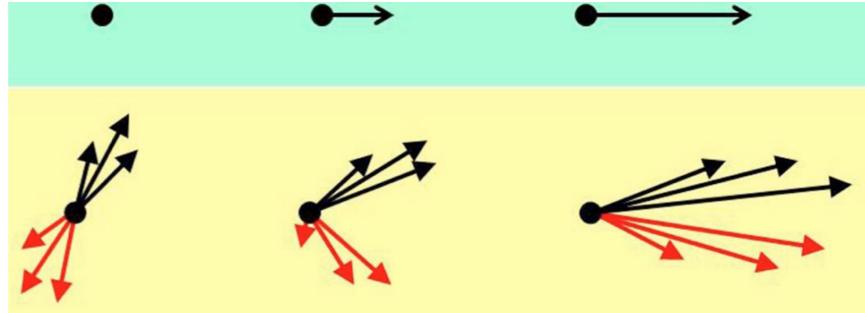


Figure 4.1: Drawing of decay products with varying degrees of Lorentz boost.

Since we are looking for events with two Higgs bosons, we expect that the total energy E from the pp collision must be at minimum $2M_H = 250$ GeV. If the total energy $E \sim 250$ GeV or slightly greater, then we expect each Higgs boson to be produced with little momentum. This means that the decay products of each Higgs will not be collimated. Therefore, we would expect to see an event with four distinct AK4 (small) jets, most likely paired off but not too close to any other jet. We would expect the event signature to look something like that of Figure 4.2. This analysis, called the resolved case, is important, but is being performed by other CMS collaborators, so we will only refer to this analysis when speaking about combining results across the different scenarios.

On the other hand, it is also possible that two protons come in and smash with a combined energy $E \gg 2M_H$. Both Higgs bosons would have considerable momentum in this case, and we would expect the decay products to be collimated. Rather than two distinct AK4 jets per Higgs, we would expect these

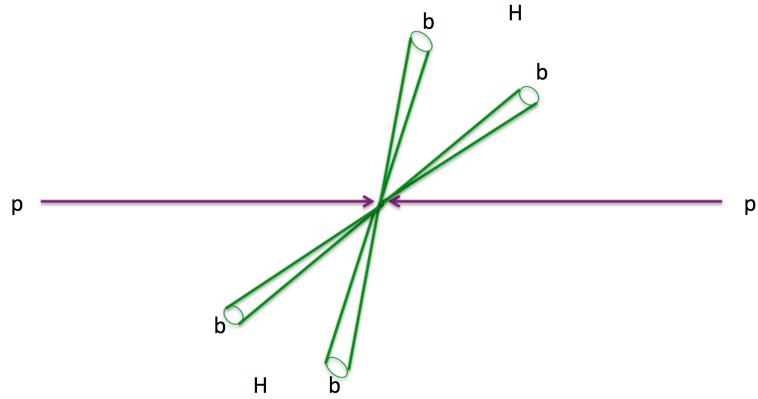


Figure 4.2: Signature for resolved two Higgs to four b-quarks.

AK4 jets to merge into one AK8 (large) jet. This means that these events would have two large jets, each representing one Higgs, similar to Figure 4.3. These events are called boosted events.

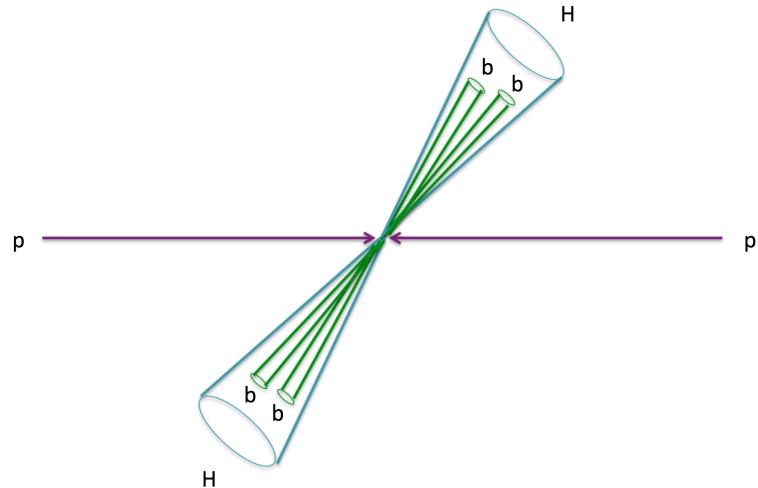


Figure 4.3: Signature for boosted two Higgs to four b-quarks.

Lastly, given these two extremes, we expect some events to have an energy E that is in the middle of these two cases, such that one Higgs boson is produced with enough momentum to collimate two AK4 jets into one big AK8 while the

other Higgs boson does not have enough momentum to have collimated decay products and is reconstructed as two small AK4 jets (Figure 4.4). We then expect one AK8 jet which is far away from two AK4 jets which are close to each other, but not so close as to have merged into one AK8 jet. We call these events semi-resolved since one half of the event looks like the boosted case and the other half looks like the resolved case. In this thesis, we will focus on the

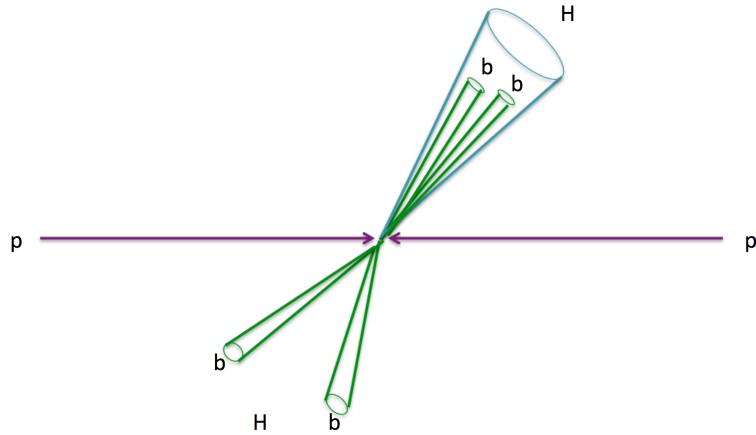


Figure 4.4: Signature for semi-resolved two Higgs to four b-quarks.

semi-resolved case and its combination with the boosted case.

4.2 Distinguishing Signal from Background

The main background for any hadronic analysis is multi-jet QCD events caused by random hadronic activity. This is quite common at the LHC, and comprises the majority of events observed. Signal, or two Higgs decaying to four b quarks ($HH \rightarrow bbbb$), on the other hand, is quite rare. In order to be able to make a discovery or rule out phasespace related to this analysis, we must find ways to

distinguish signal from background. In this section, we will cover several different tools that provide a good way to reduce background events while retaining as many signal events as possible.

4.2.1 The Soft-Drop Mass Algorithm

One of the best discriminating variables is the mass of a given jet or combination of jets. We expect that any AK8 jet representing a Higgs boson should have a mass close to that of the Higgs, or close to 125 GeV. The mass of a particular jet is calculated by summing up the momentum vectors of all jet constituents into a combined object. While one might expect that only top quarks and Higgs are reconstructed jets with a large mass, it turns out that background QCD processes, when combined with pileup inside the detector, can also have fairly large mass.

There are several algorithms that exist to "groom" jets such that any constituents resulting from pileup would be removed, making it less likely that background QCD processes would have a mass similar to that of the Higgs. For this analysis we use the Soft-Drop Mass Algorithm [42]. This algorithm begins by undoing the most recent clustering step in the jet algorithm. Then we have two pseudo-jets: the main one (jet 1) and the last constituent to be added to the main one (jet 2). These two jets are evaluated based on the following equation

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z \quad (4.1)$$

The value of z determines how stringent the grooming algorithm is; in our case, we use $z=0.1$. If the above condition is not met, that is to say the lowest p_T of the two pseudo-jets is not at least 10% of the total p_T of the two constituents, then

the lowest p_T constituent is discarded. This process is repeated, unclustering the jet step by step, until the above condition is met. At that time, the mass of the remaining constituents is defined as the soft drop mass. This is meant to remove, or "drop", low momentum, or "soft", constituents with the assumption that these are more likely to be a result of scattered background jets rather than actual decay products of a signal jet. The difference between ungroomed mass (left) and soft drop mass (right) can be seen comparing a W jet (in this case, signal) and a gluon jet (background), in Figure 4.5 [43]. Soft-drop mass provides a more narrowly peaked distribution for signal, while shifting the background peak lower, improving the discrimination power of mass.

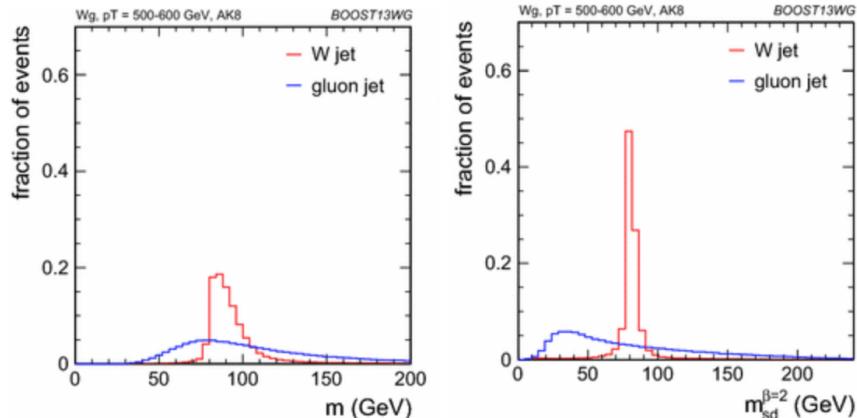


Figure 4.5: Softdrop mass (right) does a better job of discriminating between QCD (blue) and signal (W boson, in red), than ungroomed mass does (left).

4.2.2 The N-subjettiness Algorithm

Signal jets tend to have a certain amount of substructure, while background jets tend to be more chaotic. For a Higgs jet that decays to two b-quarks, we expect that if the Higgs jet is contained within one AK8 jet, there would be

two subjets within that AK8 jet, each representing one of the b-quarks. The N-subjetiness Algorithm [44, 45] defines the likelihood of a given number of subjets in a jet, allowing for an easier discrimination between jets with substructure and jets without. The algorithm compares different jet constituents to subjet axes (direction in which a particular subjet points) to determine how likely it is that a jet has a certain number of subjets. This is done by defining τ_N ,

$$\tau_N = \frac{1}{d_0} \sum_i p_{T,i} \times \min(\Delta R_{1,i}, \Delta R_{2,i}, \dots, \Delta R_{N,i}) \quad (4.2)$$

where N is the number of subjets, ΔR measures the distance between a given subjet axis and a constituent i , and $d_0 = \sum p_{T,i} R_0$ where $p_{T,i}$ is the momentum of a constituent and R_0 is the radius of the jet (0.8 for AK8 jets). This then measures the sum of each constituent's p_T multiplied by the distance to the closest subjet axis, divided by the sum of each constituent's p_T multiplied by the jet radius.

Since we are looking for jets with two subjets, and attempting to reject jets with no substructure (ie one subjet, which is just the jet itself), we compare τ_2/τ_1 , which is indicative of how likely a jet is to have two subjets rather than one subjet. The closer to 0 this value is, the more likely the jet is to have two subjets, whereas the closer to 1, the more likely the jet is to have no substructure, or one subjet. This variable can be seen in Figure 4.6 [46], where W is also expected to have two subjets when it decays hadronically, but QCD is not expected to have much substructure.

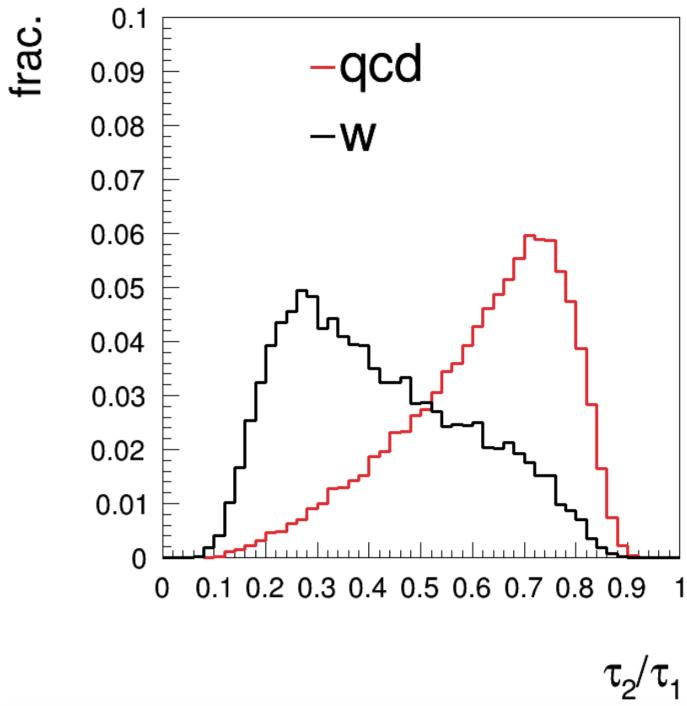


Figure 4.6: N-subjetiness variable τ_2/τ_1 for W jets and background jets.

4.2.3 B-Tagging

We expect to see four b-quarks in each event for the signal $HH \rightarrow bbbb$. For the semi-resolved case, this means that the resolved Higgs should have two AK4 jets which are each from a b-quark, and the boosted Higgs should have one AK8 jet with two b-quark subjets inside. As it turns out, b-quarks leave a signature in the detector that is unique from any other hadronic activity. Bare quarks inside the detector immediately form hadrons. The lightest quarks hadronize immediately, so their jets are formed at the interaction point, or primary vertex (Figure 4.7, left). However, b hadrons have a longer lifetime, so they travel a short distance in the detector before decaying (Figure 4.7, right). This means that most of the jet arises from a secondary vertex, rather than the primary

vertex. Conversely, a top quark has such a short lifetime that it does not have

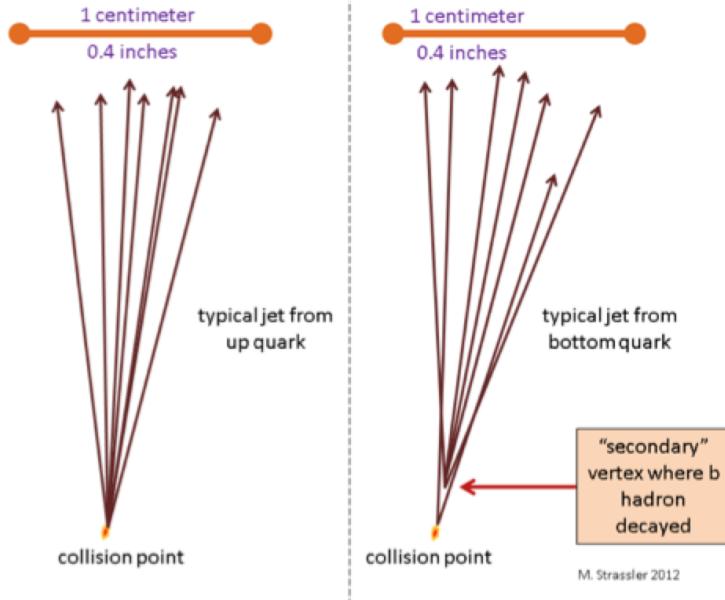


Figure 4.7: Light quark (u, d, c, s) in a detector (left) vs. b-quark in a detector (right).

any time to hadronize and immediately decays. Thus, a b-quark has a unique signature in comparison with the other five quarks and this signature can be used to identify jets coming from b-quarks. We use two different algorithms for identifying b-quarks in this analysis: the Deep CSV algorithm identifies AK4 b jets and the Double-b Algorithm identifies AK8 jets with two b-quarks inside. A full description of b-tagging algorithms used by CMS can be found in Reference [47].

4.2.3.1 Deep CSV Algorithm

In order to identify AK4 jets as b-jets, we use the deep CSV algorithm, a multivariate measurement of jets with a secondary vertex, which takes into

account information from the displaced tracks and from the secondary vertices associated with the jet. The algorithm is trained to recognize b-jets and non b-jets depending on these variables, using a deep neural network to "learn" the difference between signal and background so as to assign a certain likelihood of a jet containing a b-quark based off of these different variables (0 is unlikely, 1 is very likely). The discriminator can be seen in Figure 4.8, and is a combination of the probability of a jet containing one b hadron and the probability of a jet containing two b hadrons. Overall, this algorithm outperforms the other b-tagging algorithms used on CMS data.

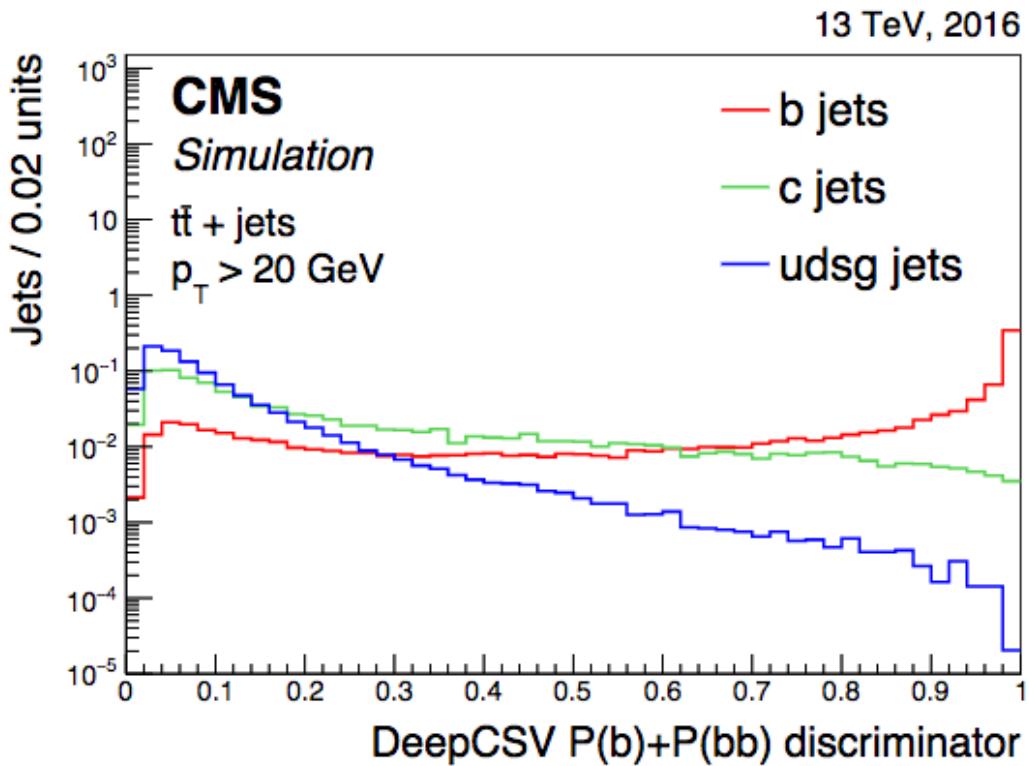


Figure 4.8: Deep CSV discriminator for b jets (red), c jets (green), and light jets (blue).

4.2.3.2 Double-b Algorithm

The double-b algorithm is designed to identify AK8 jets with two b-quarks inside of the jet. It was specifically designed to identify $H \rightarrow bb$ jets. Just like the deep CSV algorithm, the double-b algorithm uses a multivariate approach to identify jets with two b-quarks, taking into account variables related to the τ axes calculated by the N-subjettiness algorithm as well as variables related to the displaced tracks and secondary vertices. In particular, this algorithm is designed to be p_T and mass independent, which allows for a wide range of AK8 jets to be properly tagged. The algorithm is trained to recognize jets with two b-quarks and jets without two b-quarks depending on the aforementioned variables, using a boosted decision tree to "learn" the difference between signal and background. The algorithm assigns a likelihood of a jet containing two b-quarks, ranging from -1 (unlikely) to 1 (very likely). This can be seen in Figure 4.9. The double-b algorithm outperforms any other tagger used by CMS to identify jets with two b-quarks when the background mostly consists of QCD or jets with two b-quarks coming from gluon fusion, as well as when jets have high p_T . Since this is our case, this algorithm was chosen to identify the AK8 jet with two b-quarks that are arising from a Higgs decay.

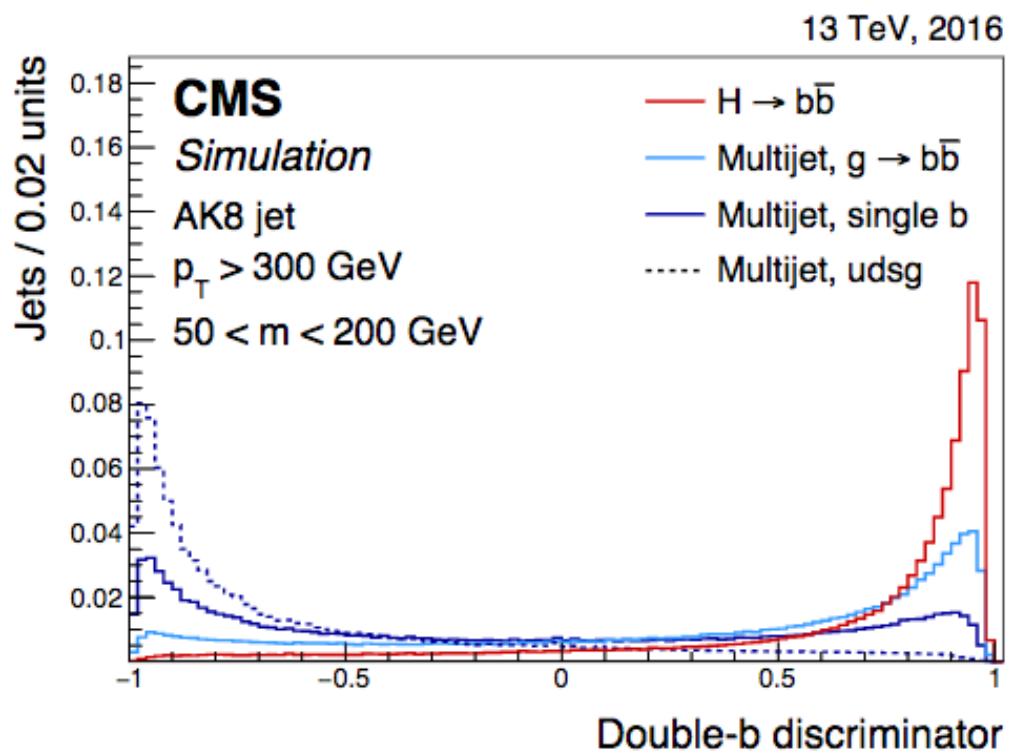


Figure 4.9: Double b-tagger for signal (red) and various QCD processes (varying shades of blue).

Chapter 5

Semi-resolved Analysis

This thesis is focused on events at the LHC with two Higgs bosons that each decay to two b quarks, or $\text{HH} \rightarrow \text{bbbb}$. We are looking for resonant production coming from a particle such as the bulk graviton or radion, and non-resonant production through SM and perhaps other BSM diagrams. However, these processes are much less common than QCD multijet processes at the LHC. In order to effectively probe this new physics, we must reduce as much background as possible, while retaining as much signal as possible.

We begin the process of optimizing the ratio of signal to background by considering important characteristics of our signal events. Recalling the event depiction, shown in Figure 5.1, we can identify a number of variables that may help with the elimination of background. First, we focus on the topology of the event. Signal events should have one large AK8 jet, representing the boosted Higgs. In addition, there should be two AK4 jets that are outside the cone of the AK8 jet, and close to each other. These two AK4 jets representing the resolved Higgs. We expect a high amount of hadronic activity in each event. While the process of hadronization can (and often does) produce leptons, we

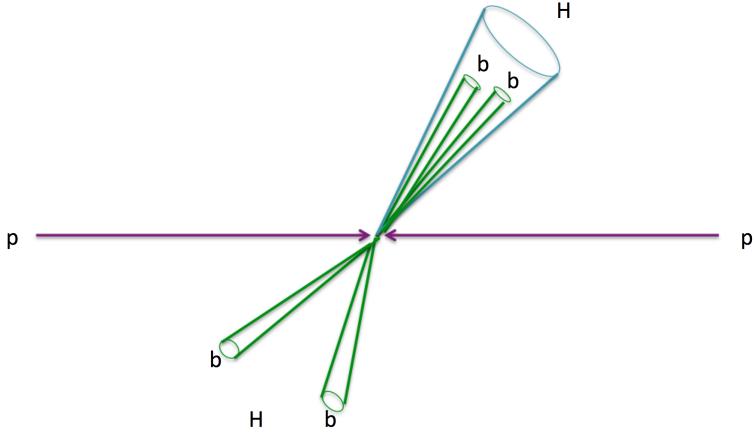


Figure 5.1: Signature for semi-resolved two Higgs to four b-quarks.

expect that these electrons or muons would not be isolated from the hadronic activity, and would be low momentum. We expect that the combination of the two AK4 jets should be roughly opposite that of the AK8 jet, which is to say they should have similar pseudorapidity.

Now, we break it down, side by side. We expect that the AK8 jet will have a high momentum, since it had enough momentum for the Higgs decay products to be collimated into one jet. We expect that this jet will not be too close to the beamline, or the pseudorapidity should be low. We expect that the softdrop mass of this jet should be close to the Higgs mass. The τ_{21} of the jet should be close to 0, since we expect this jet to be more likely to have two subjets than no substructure. This jet should have a double-b tagger value close to 1, indicating that it is likely to have two b-quarks inside the jet.

For the AK4 jet side, we expect that these jets should not have low momentum. Since they are the individual b-jets rather than the combined Higgs jet, and the Higgs did not have enough momentum to collimate these two jets into

one big jet, we do not necessarily expect that these jets would have a momentum as large as the AK8 jet, but the momentum of these two jets should still be higher than that of typical QCD jets. We also expect that these jets will not be too close to the beamline, or the pseudorapidity should be low. If we add up the four-vectors of the two AK4 jets, the mass of this combination should be close to that of the Higgs. Each AK4 jet should have a deepCSV value close to 1, since we expect both jets to be the result of a b-quark.

All of these considerations should help eliminate QCD multijet background, while retaining most of our signal events. However, this is not the only background we must consider. It turns out that top-anti-top ($t\bar{t}$) can also contribute to the background. Top quarks decay into a W boson and a b-quark, so events with two tops contain two W's and two b-jets. While W bosons can decay leptonically, they decay hadronically 2/3 of the time. Because of the high mass of the top quark (even higher than the Higgs at 172 GeV), these events can mimic $H\bar{H} \rightarrow bbbb$ events more easily than QCD multijet processes can. Luckily, the rate at which $t\bar{t}$ events are produced is much lower than that of QCD multijets. Still, we must take into consideration how to eliminate as much $t\bar{t}$ as possible in addition to eliminating QCD multijet background.

Data comes out of the LHC as information from various detectors, and we can only infer what the actual events may have been. In order to better understand the data we see, we use Monte Carlo simulations of signal and background processes to study different properties of signal and background events. Monte Carlo (MC) simulations are generated by random sampling, where statistical information on what could happen in a given event informs what is in each MC event. Once MC events have been produced, the events are

sent through a simulation of the CMS detector in order to mimic data taken by CMS. We then can compare signal and background MC distributions of many different variables to figure out how to prune the data to get the best signal to background ratio possible.

In the following sections, we will cover the data and MC used as well as the event selection of this analysis, detailing each choice and demonstrating its efficacy through looking at these variables in MC simulation of signal and background.

5.1 Data samples

The analysis is performed using pp interactions collected with the CMS detector at $\sqrt{s} = 13$ TeV. The data samples are summarized in Table 5.1. The JetHT dataset was chosen due to the high amount of hadronic activity in these events. Each run corresponds to a different time period in which data was taken, and the total dataset corresponds to an integrated luminosity, or total amount of data, of 35.9 fb^{-1} .

Table 5.1: List of primary datasets for the pp collisions at $\sqrt{s} = 13$ TeV and their corresponding integrated luminosities.

Dataset	Processing	Int. lumi. (fb^{-1})
JetHT/Run2016B	03Feb2017	5.9
JetHT/Run2016C	03Feb2017	2.6
JetHT/Run2016D	03Feb2017	4.4
JetHT/Run2016E	03Feb2017	4.1
JetHT/Run2016F	03Feb2017	3.2
JetHT/Run2016G	03Feb2017	7.7
JetHT/Run2016H	03Feb2017	8.9
Total		35.9 fb^{-1}

5.2 MC simulation

The signal MC samples used for this analysis includes spin-0 bulk graviton and spin-2 radion resonances which both decay to $\text{HH} \rightarrow \text{bbbb}$, as well as non-resonant $\text{HH} \rightarrow \text{bbbb}$, given in Table 5.2. We examine a range of different masses for the bulk graviton and radion resonances, chosen such that we would expect at least some events to present as semi-resolved, rather than fully resolved or fully boosted. We also examine all of the different non-resonant BSM benchmarks, as well as the SM non-resonant production.

We also consider different background processes that are likely to contribute to the analysis. The two largest are QCD multijet and $t\bar{t}$ MC, as given in Table 5.3. The QCD samples are separated into H_T ranges, where H_T is the summed p_T of all jets in the event. Other MC samples related to diboson production were examined, but found to have no appreciable impact.

5.3 Event selection

Events are required to have at least one reconstructed pp collision vertex. Many additional vertices, corresponding to other overlapping pp collisions (pileup), are usually reconstructed in an event using charged particle tracks. We assume that the primary interaction vertex (PV) corresponds to the one that maximizes the sum in p_T^2 and the magnitude of $\sum p_T$ from the associated physics objects.

5.3.1 Jet kinematics selection

Individual particles are reconstructed using the PF algorithm described previously. The five classes of PF candidates are muons, electrons, photons, charged

hadrons, and neutral hadrons.

This analysis used AK8 jets and AK4 jets. In order to mitigate the effect of pileup on jet observables, we take advantage of pileup per particle identification (PUPPI) [48] for AK8 jets. This method uses local shape information, event pileup properties and tracking information together in order to compute a weight describing the degree to which a particle is pileup-like. No additional pileup corrections are applied to AK8 jets clustered from these weighted inputs.

The jet 4-momenta are corrected to account for the difference between the measured and the expected momentum at the particle level, using a standard CMS correction procedure described in Refs. [49, 50]. The `Summer16_23Sep2016V4` jet energy corrections [51] were used. The `Spring16_25nsV10` jet energy resolutions are used. All AK8 jets are further required to pass TightLepVeto jet identification requirements [52], which was chosen over Tight to reject leptons more efficiently. These requirements ensure that the jet is comprised of mostly charged hadrons, with more than one constituent, and limits the amount of neutral hadrons, electrons, and muons in the jet. Figs. 5.2, 5.3, 5.4 show the p_T and η of the three selected jets. A preselection is applied to QCD MC, $t\bar{t}$ MC, and bulk graviton signal points for masses 600, 800, 1000, and 1200 GeV as follows:

- Passes trigger selection (discussed later in this chapter);
- At least one AK8 jet in the event with $p_T > 300GeV$ and $|\eta| < 2.4$;
- At least two AK4 jets with $p_T > 30GeV$ and $|\eta| < 2.4$, with a deep CSV value > 0.2219 ;

- The AK8 jet and AK4 jets that represent the three signal jets are chosen in the following way. First we pick out AK4 jets that have $p_T > 30$, $|\eta| < 2.4$, and $\text{deep CSV} > 0.6324$. Out of this group we find all AK4 jet pairs where each AK4 jet is at least $\Delta R > 0.8$ away from the highest p_T AK8 jet, and within ΔR of 1.5 of each other. If there are more than one pair, we pick the pair with the highest deepCSV values. If there are no pairs, we repeat this second step for the second highest p_T AK8 jet and check if we get a pair of AK4 jets that match the criteria. Once again, if there are more than one pair, we pick the pair with the highest deepCSV values, and if there are none, the event is rejected.
- Lepton veto (discussed later on);
- Fails selection of the fully boosted and the selection of the fully resolved analysis, which is documented in Sec. 5.3.10.

All figures in this section have the same preselection, with QCD MC (yellow) and $t\bar{t}$ (red) weighted by their cross section multiplied by the luminosity of the dataset and divided by the total number of events. The signal samples (different dashed lines) are weighted by 50 multiplied by the cross section multiplied by the luminosity of the dataset and divided by the total number of events, so that they are easy to see on the plots.

The analysis is performed in two different $\Delta\eta$ regions, 0–1.0 and 1.0–2.0, where $\Delta\eta = |\eta_{AK8} - \eta_{1AK4+2AK4}|$ where AK8 is the AK8 jet and the two AK4 jets are added together. This variable can be found in Figure 5.5.

As discussed previously, we expect a high p_T AK8 jet, so we require AK8 jet > 300 GeV. The two AK4 jets are required to have > 30 GeV each, to prevent

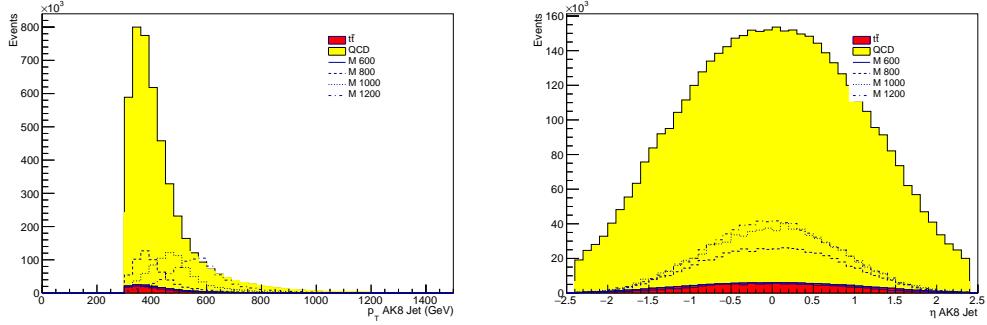


Figure 5.2: Left. The p_T of the AK8 jet. Right. The η of the AK8 jet.

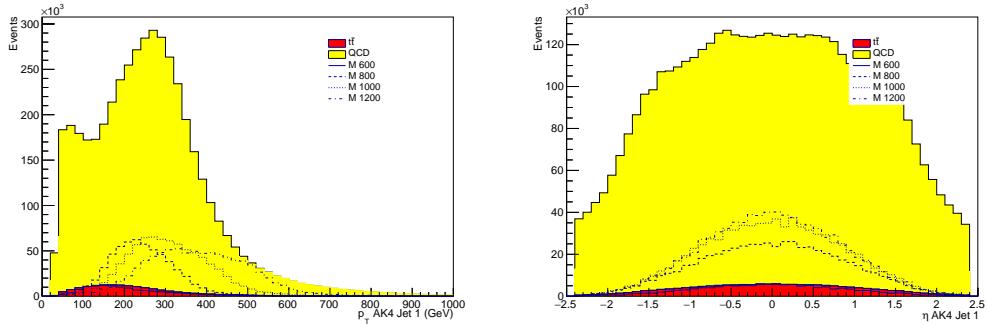


Figure 5.3: Left. The p_T of the highest p_T selected AK4 jet. Right. The η of the highest p_T selected AK4 jet.

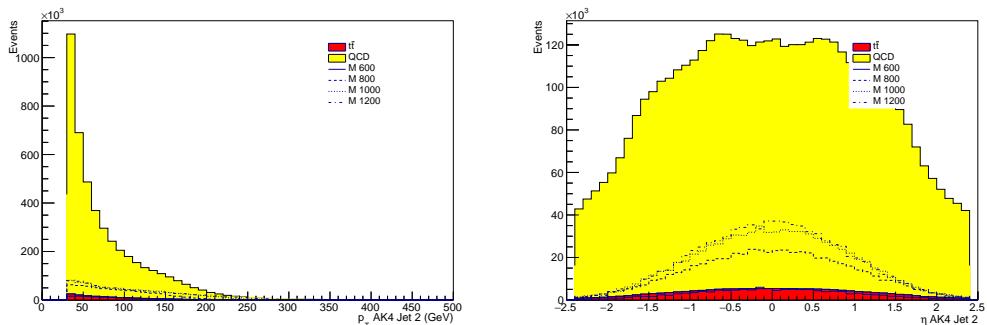


Figure 5.4: Left. The p_T of the other selected AK4 jet. Right. The η of the other selected AK4 jet.

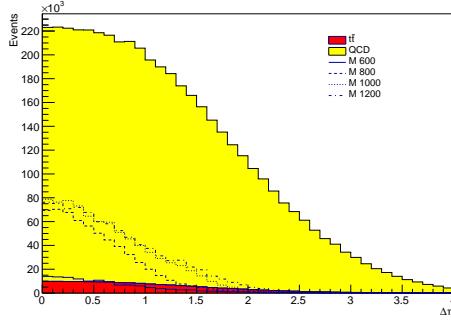


Figure 5.5: The absolute value of $\Delta\eta$ between the AK8 jet and the combined 4 vector of the AK4 jets. We require $|\Delta\eta| < 2.0$.

low p_T QCD jets. All three jets are required to have $|\eta| < 2.4$, ensuring that the jets are not too close to the beamline. Lastly, we require $|\Delta\eta| < 2.0$ so that the AK8 jet and the result of the addition of the two AK4 jets have similar η , to ensure that they are in similar positions in the detector.

5.3.2 H mass selection

The AK8 softdrop mass is also used to limit the amount of background. A dedicated jet energy calibration is applied to the softdrop mass as derived in Ref. [53]. The correction is derived in two steps. First, a weight to account for a p_T dependent softdrop jet mass shift introduced prior to MC being sent through the detector simulation is calculated. Second, to account for any residual p_T and η dependence, an additional weight is calculated based on the difference between the reconstructed (post-detector simulation) and the generated (pre-detector simulation) softdrop mass. The difference in reconstructed and generated softdrop mass is a 5-10% effect. Figure 5.6 shows the AK8 soft-drop corrected jet mass.

The softdrop corrected mass, or $M_{\text{soft drop}}$ signal mass window is restricted to between 105 and 135 GeV, to avoid overlapping with other analyses targeting W and Z resonances.

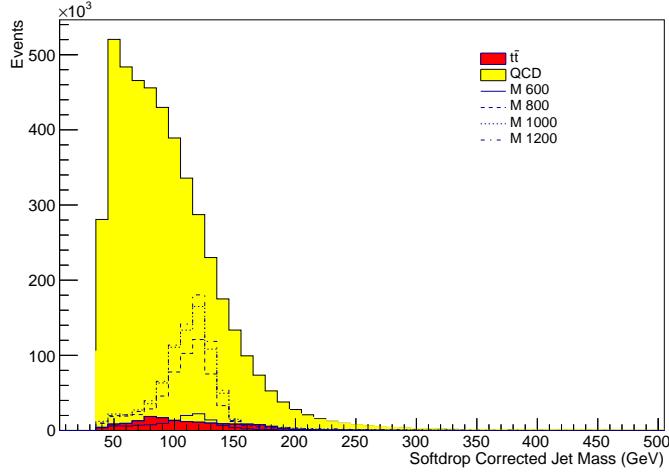


Figure 5.6: The soft-drop mass of the AK8 jet. We require it to be between 105 and 135 GeV.

The combined mass of the two AK4 jets, defined as the mass of the two AK4 jet vectors added together, or the mass of $(\text{AK4jet1} + \text{AK4jet2})$, can also be used to suppress the multijet and $t\bar{t}$ backgrounds as well, shown in Figure 5.7. The AK4 dijet mass window is slightly larger, ranging from 90 to 140 GeV, because there is less worry of an overlap with another analysis.

5.3.3 N-subjettiness selection

The ratio $\tau_{21} = \tau_2/\tau_1$ is calculated for the AK8 jet after PUPPI has been applied. It is required to have $\tau_{21} < 0.55$. The τ_{21} spectra for the AK8 jet is shown in Figure 5.8.

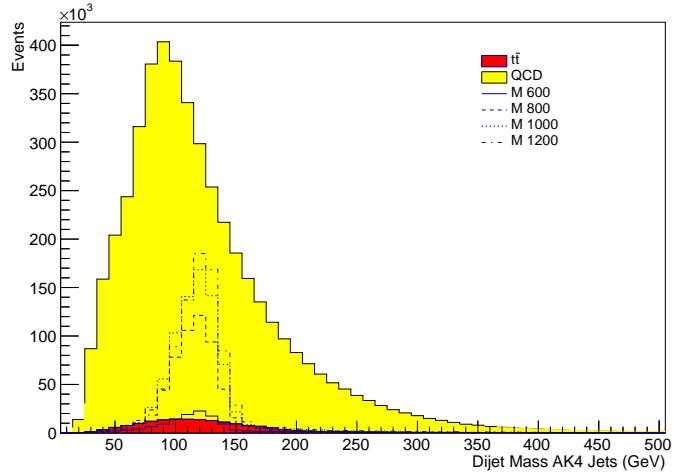


Figure 5.7: The dijet mass of the AK4 jet. We require it to be between 90 and 140 GeV.

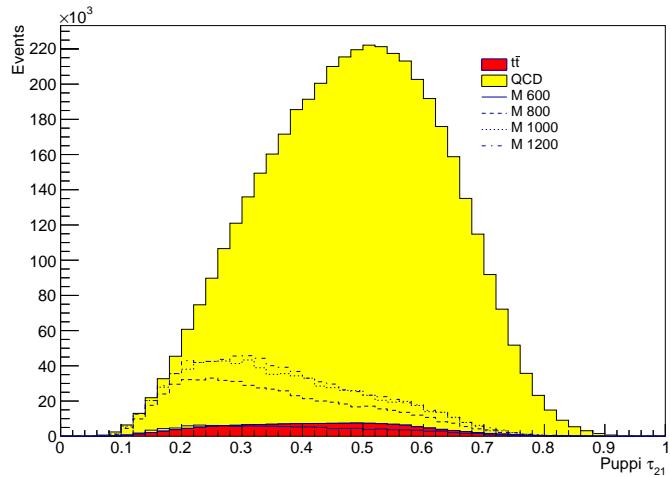


Figure 5.8: τ_{21} distribution for the AK8 jet. We require $\tau_{21} < 0.55$.

5.3.4 B-tagging

In order to identify the AK8 jet most likely to contain two b quarks, we use the double-b tagger discriminant, shown in Figure 5.9. We require double b-tagger

> 0.8 .

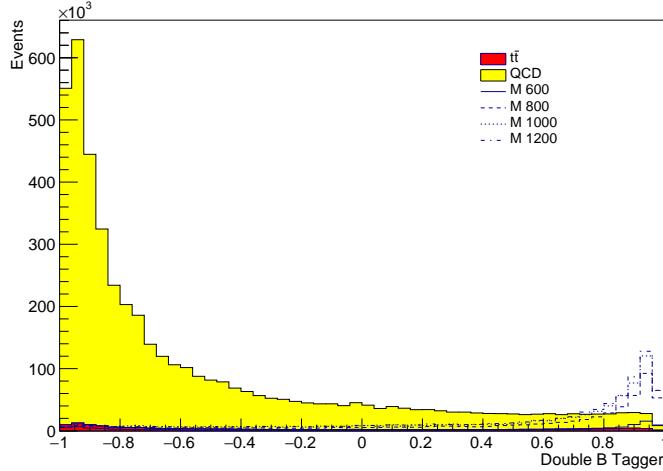


Figure 5.9: Double-b tagger discriminant for the AK8 jet. We require double b tagger > 0.8 .

For the AK4 jets, we use the deep CSV tagger to b-tag the jets. We require deep CSV > 0.6324 , found in Figure 5.10.

5.3.5 Variable for $t\bar{t}$ Reduction

Lastly, we studied many variables related to unselected AK4 jets present in selected events to determine how to remove as much $t\bar{t}$ as possible. The variable that provided the most discriminating power was the invariant mass of the two selected AK4 jets combined with the nearest unselected AK4 jet to one of the selected AK4 jets that is not part of the AK8 jet. We call this combined mass the triAK4jet mass. This variable can be seen in Figure 5.11, where we place a cut requiring the value of the invariant mass to be larger than 200 GeV.

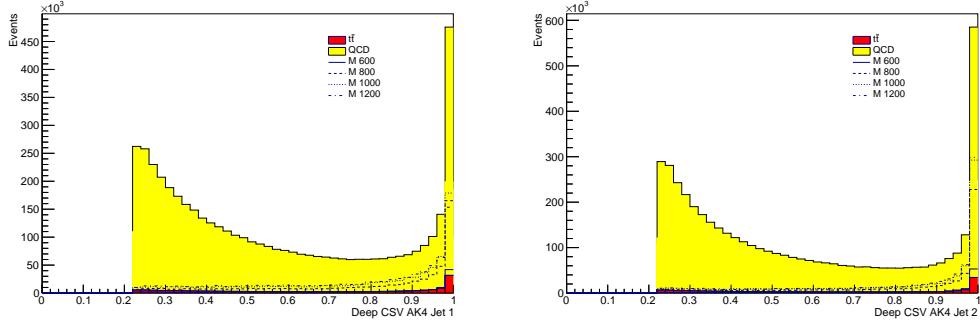


Figure 5.10: Deep CSV $P(b) + P(bb)$ tagger discriminant for the highest p_T selected AK4 jet (left) and the other selected AK4 jet (right). We require $\text{deepCSV} > 0.6324$.

5.3.6 Invariant mass definition and “reduced mass”

The invariant M_{jjj} mass distribution of the AK8 jet and two AK4 jets in the event corresponds to the invariant mass of the resonance searched for. We use instead the reduced mass, defined

$$M_{jjj}^{red} \equiv M_{jjj} - (M_{AK8} - M_H) - (M_{1AK4+2AK4} - M_H) \quad (5.1)$$

since it provides resolution improvement and the mean position of M_{jjj}^{red} remains at $\approx M_{jjj}$. In this equation, M_{AK8} is the softdrop corrected mass of the AK8 jet, $M_{1AK4+2AK4}$ is the dijet mass of the two AK4 jets, and M_H is 125 GeV. The reduced mass can be found in Figure 5.12. We require the reduced mass > 750 GeV because this analysis has little sensitivity below this level.

5.3.7 Lepton Veto

A lepton veto is applied. Events are vetoed if they contain one strictly defined or two loosely defined opposite-sign same-flavour isolated electrons or muons

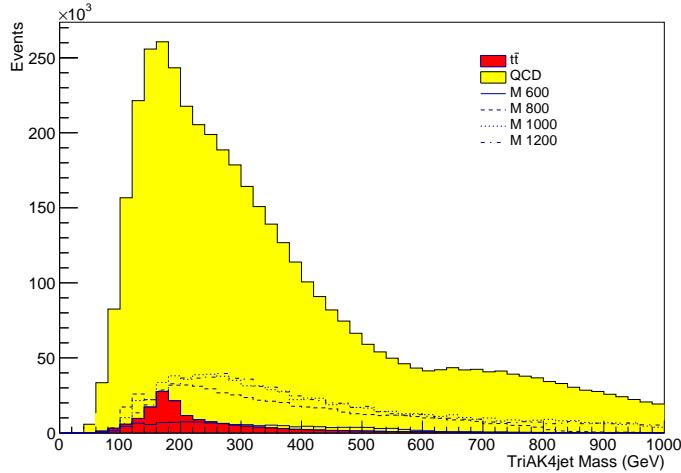


Figure 5.11: TriAk4jet mass: the invariant mass of the two selected AK4 jets and the nearest unselected AK4 jet. We require triAK4jet mass > 200 GeV.

with p_T in excess of 20 GeV. No veto on the presence of isolated photons is applied.

5.3.8 Trigger Choice

As mentioned in the previous chapter, events in a particular dataset are organized by which triggers they pass. In this case, we chose a combination of triggers, and we require that each event pass at least one of these triggers. The triggers chosen place requirements on the scalar sum of jet transverse energy, H_T , jet p_T , the jet groomed mass, and b-tagging. The trigger paths used are listed in Table 5.4. Recall that HLT triggers are sorted by which Level 1 (L1) triggers were passed, which is also listed in the table. The H_T triggers are chosen because we expect signal events to have a high H_T . However, the L1 H_T trigger paths were observed to have an inefficiency, so we also used jet p_T based triggers to recover these events that did not pass the H_T triggers. In the last

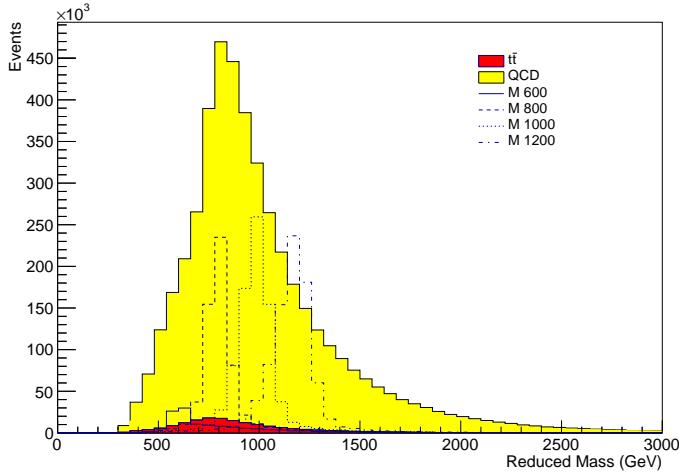


Figure 5.12: Reduced mass of the two selected AK4 jets and the AK8 jet.

data run (Run H), the `HLT_PFHT800` trigger was prescaled (not all events that passed the trigger were kept, as explained in the previous chapter). Therefore, we added the `HLT_PFHT900` trigger to compensate for this. For the L1 triggers, the names of each trigger explain what is required to pass the trigger. "HTT" means that H_T in that event is greater than the number that comes after, for example $H_T > 160$ GeV. "SingleJet" means that there was at least one jet, with p_T greater than the number listed, for example jet 1 has $p_T > 180$ GeV. The first HLT trigger requires $H_T > 650$ GeV, where two jets have a combined mass > 900 GeV and the difference in the pseudorapidity between the two jets is less than 1.5. The second and third HLT trigger require $H_T > 650$ GeV and 700 GeV respectively, with a mass jet > 50 GeV. The fourth and fifth trigger require $H_T > 800$ GeV and 900 GeV respectively. The sixth trigger requires an AK8 jet with $p_T > 360$ GeV, and a mass > 30 GeV, while the last trigger requires two AK8 jets with $p_T > 280$ GeV and 200 GeV respectively, with mass > 30

GeV and b-tag value > 0.20 (using the CSV b-tagging algorithm, the precursor to deepCSV). Most signal events should pass at least one of these triggers.

The trigger requirement is applied to both the data and the MC (signal and tt). In order to compensate for the difference in trigger response between the data and the simulation, trigger efficiency scale factors, defined as the ratio of the trigger efficiency as measured in the data to that in the MC, are applied to the simulated events. Trigger efficiency is the number of events that pass a defined selection and the trigger, divided by the number of events that pass that same defined selection regardless of what triggers they pass. A baseline trigger of HLT_PFJET260, a prescaled trigger which requires at least one jet with $p_T > 260$ GeV, is used to select events for the measurement of the trigger efficiency. This trigger is prescaled, yet provides enough events for measurement of the efficiencies and the scale factors. Events passing the baseline trigger are further required to pass selection criteria close to the signal selection in the actual analysis:

- At least one AK8 jet in the event with $p_T > 300GeV$ and $|\eta| < 2.4$;
- At least two AK4 jets with $p_T > 30GeV$ and $|\eta| < 2.4$, with a deep CSV value > 0.6324 ;
- Out of the AK8 and AK4 jets that satisfy the above requirements, one AK8 jet and two AK4 jets are chosen to be the three signal jets. The highest p_T AK8 jet with $M_{\text{soft drop}} > 40$ GeV that is $\Delta R > 0.8$ away from two AK4 jets that are within $\Delta R < 1.5$ of each other is chosen. If there are more than two AK4 jets that satisfy this criteria, the two AK4 jets with the highest deepCSV value are chosen.

- The soft drop mass of the AK8 jet is $105 < M_{\text{soft drop}} < 135 \text{ GeV}$, with all necessary jet mass corrections applied;
- The combined mass of the two AK4 jet is $90 < M_{\text{dijet}} < 140 \text{ GeV}$;
- Reduced mass, introduced in Eqn. 5.1, $> 750 \text{ GeV}$, since the analysis has little sensitivity below this value;
- $\Delta\eta = |\eta_{\text{AK8}} - \eta_{\text{1AK4+2AK4}}| < 2.0$;
- Lepton veto.

This selection is used in lieu of the full selection because the full selection can only be applied to data when the analysis has been fully verified, so as to avoid biasing results. In addition to the selection, the trigger efficiency is calculated separately for events with $\Delta\eta < 1.0$ and events with $1.0 \leq \Delta\eta < 2.0$. Then the trigger efficiency for data and MC is defined as

$$\frac{\text{N. events passing selection} + \text{HLT_PFJET260} + \text{at least one trigger listed}}{\text{N. events passing selection} + \text{HLT_PFJET260}} \quad (5.2)$$

Trigger efficiency is measured as a function of the “reduced mass” introduced in Eqn. 5.1 of Section 5.3.6. The efficiency is shown in Figure 5.13. The trigger efficiency at low mass in the high $\Delta\eta$ region is higher than expected due to the inefficiency of HLT_PFJET260 at low mass.

For $M_{jjj}^{\text{red}} < 1000 \text{ GeV}$, the trigger efficiencies are higher for smaller $\Delta\eta$, where most of the signal lie, and are smaller at larger values of $\Delta\eta_{jj}$. Furthermore, the data/MC scale factor too varies depending on $\Delta\eta_{jj}$. Since we begin the search from M_{jjj}^{red} well below 1000GeV, one needs to be especially careful of

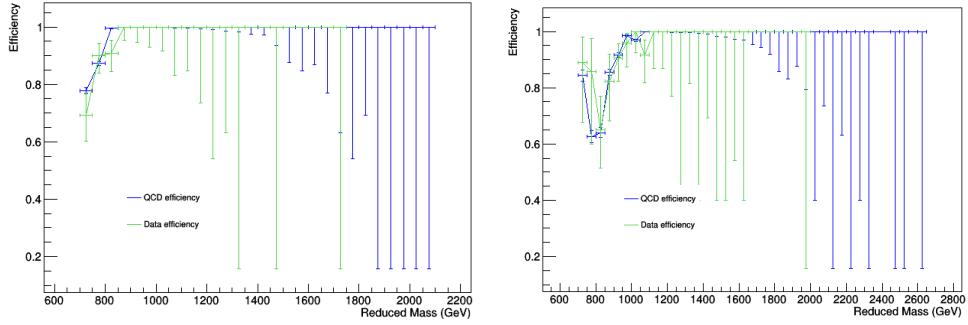


Figure 5.13: The trigger efficiency, as a function of reduced mass. Efficiency is defined for different $\Delta\eta$ regions: 0.0–1.0 (left), 1.0–2.0 (right).

the modelling of the trigger efficiency turn-on curves in the data and the simulations. Since the baseline trigger HLT_PFTjet260 too has some inefficiency for low M_{jjj}^{red} , we measure it in QCD MC and assign an uncertainty to the trigger efficiency scale factor based on this. This efficiency is defined as

$$\frac{\text{N. events passing selection + HLT_PFTjet260}}{\text{N. events passing selection}} \quad (5.3)$$

Figure 5.14 shows the HLT_PFTjet260 trigger turn-on curves for different

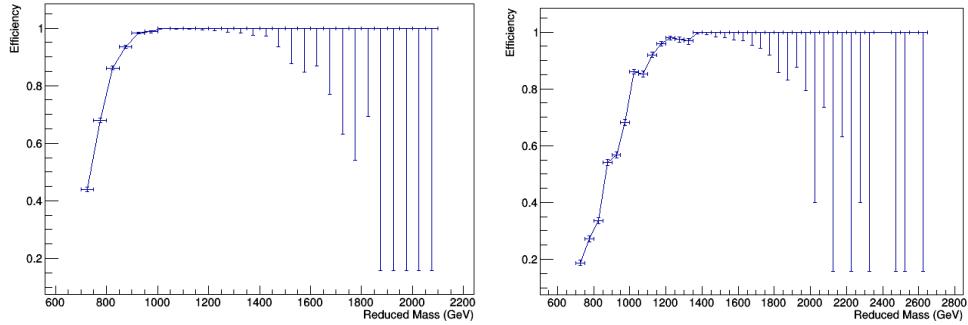


Figure 5.14: The trigger efficiency in QCD MC for the baseline trigger HLT_PFTjet260, for different $\Delta\eta$ regions: 0.0–1.0 (left) and 1.0–2.0 (right). The percentage difference between one and these turn-on curves are taken as an uncertainty on the trigger efficiency scale factor.

$\Delta\eta$ regions in MC. The difference between unity and the trigger efficiency is propagated to the scale factor as a systematic uncertainty. In addition to this systematic uncertainty, we assign a 0.5% uncertainty based on the use of the last trigger listed in Table 5.4, since it has a b-tagging requirement. The trigger efficiency difference between data and MC is calculated comparing QCD and data, but QCD has a lower percentage of b-quarks than signal does, and this trigger efficiency scale factor gets applied to signal, so we had to correct for this difference in composition. The trigger efficiency scale factor is calculated as

$$\frac{\text{efficiency in data}}{\text{efficiency in MC}} \quad (5.4)$$

The scale factor and errors are shown in Figure 5.15.

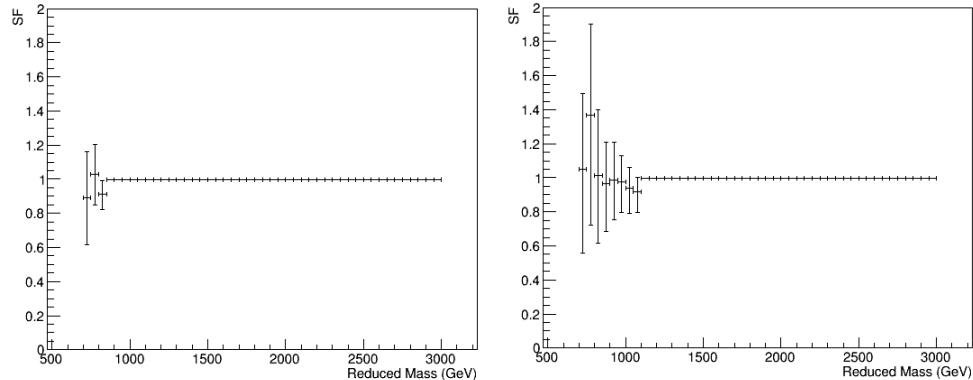


Figure 5.15: The trigger efficiency scale factors, as a function of M_{jjj}^{red} . The error bars are the combined statistical and systematic errors.

5.3.9 Summary of the Semi-Resolved Section

Taking into account all of the above information, we can summarize the selection as follows:

- Passes triggers described in Table 5.4;
- One AK8 jets, $p_T > 300$ GeV, $|\eta| < 2.4$, tight jet ID;
- Two AK4 jets, $p_T > 30$ GeV, $|\eta| < 2.4$;
- Out of the AK8 and AK4 jets that satisfy the above requirements, one AK8 jet and two AK4 jets are chosen to be the three signal jets. The highest p_T AK8 jet with $M_{\text{soft drop}} > 40$ GeV that is $\Delta R > 0.8$ away from two AK4 jets that are within $\Delta R < 1.5$ of each other is chosen. If there are more than two AK4 jets that satisfy this criteria, the two AK4 jets with the highest deepCSV value are chosen.
- $|\Delta\eta| < 2.0$ between the AK8 jet and the combined AK4 jets;
- $M_{jjj}^{red} > 750$ GeV;
- triAK4jet mass > 200 GeV;
- Lepton veto;
- The soft drop mass of the AK8 jet is $105 < M_{\text{soft drop}} < 135$ GeV, with all necessary jet mass corrections applied;
- $90 < \text{Dijet mass of the two AK4 jets} < 140$ GeV
- double b-tagger > 0.8
- deep CSV > 0.6324
- $\tau_{21} < 0.55$ for the AK8 jet;

5.3.10 Selection of Other Analyses

5.3.10.1 Boosted Analysis

The boosted analysis selection is documented in Reference [54].

The selection is identical to the boosted side of the semi-resolved analysis by design; this analysis came first and established best practices for this channel, and the semi-resolved analysis was modeled after the work done in this analysis.

The boosted analysis requires:

- Passes triggers described in Table 5.4;
- Two AK8 jets, $p_T > 300$ GeV, $|\eta| < 2.4$, tight jet ID;
- $|\Delta\eta| < 1.3$ between two AK8 jets;
- $M_{jj}^{red} > 750$ GeV;
- Lepton veto;
- The soft drop mass of both AK8 jets is $105 < M_{\text{soft drop}} < 135$ GeV, with all necessary jet mass corrections applied;
- Falls into one of the following categories: both AK8 jets have $\text{double-}b > 0.8$, one AK8 jet has $\text{double-}b > 0.8$ and the other has $0.3 < \text{double-}b < 0.8$, both AK8 jets have $0.3 < \text{double-}b < 0.8$.

It should be noted, however, that this is only the events in the signal region. Because the boosted analysis uses the same background estimate method as this analysis (described in later chapters), we must also account for any event that was used to estimate the background. This way, we prevent any overlap of

events, whether in the data signal region or in a region of data used to estimate the QCD background contribution. This set of events is as follows:

- Passes triggers described in Table 5.4;
- Two AK8 jets, $p_T > 300 \text{ GeV}$, $|\eta| < 2.4$, tight ID;
- $|\Delta\eta| < 1.3$ between two AK8 jets;
- $\tau_{21} < 0.55$ for both AK8 jets;
- $M_{jj} > 750 \text{ GeV}$;
- The soft drop mass of **only the subleading AK8 jet** is required to be between $105 < M_{\text{soft drop}} < 135 \text{ GeV}$, with all necessary jet mass corrections applied;
- The double-b of **only the subleading AK8 jet** is required to satisfy $\text{double-b} > 0.3$.

We note that since the entire mass spectrum and double-b spectrum of the leading AK8 jet is used either in the signal region or as part of the background estimate, we cannot place any restrictions on these values. Then this selection defined directly above is the selection rejected when limits are calculated to be combined with the fully boosted analysis.

5.3.10.2 Resolved Analysis

The fully resolved non-resonant and resonant analysis have different selections. However, in this thesis, we only talk about combining with the fully resolved non-resonant analysis, so this is the only relevant selection.

Selection is as follows:

- Passes the triggers HLT_QquadJet45_TripleBTagCSV_p087 and/or HLT_DoubleJet90_Double30_TripleBTagCSV_p087;
- Four AK4 jets, $p_T > 30$ GeV, $|\eta| < 2.4$, and CMVA > 0.4432 (a b-tagger similar to deepCSV)

This selection is relatively loose because they optimize events differently, using a boosted decision tree (type of machine learning) to decide which events to keep.

Table 5.2: List of signal MC samples used.

Resonant Bulk graviton
GluGluToBulkGravitonToHHTo4B_M-500_narrow_13TeV-madgraph
GluGluToBulkGravitonToHHTo4B_M-550_narrow_13TeV-madgraph
GluGluToBulkGravitonToHHTo4B_M-600_narrow_13TeV-madgraph
GluGluToBulkGravitonToHHTo4B_M-650_narrow_13TeV-madgraph
GluGluToBulkGravitonToHHTo4B_M-750_narrow_13TeV-madgraph
GluGluToBulkGravitonToHHTo4B_M-800_narrow_13TeV-madgraph
GluGluToBulkGravitonToHHTo4B_M-900_narrow_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-1000_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-1200_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-1400_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-1600_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-1800_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-2000_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-2500_13TeV-madgraph
BulkGravTohhTohbhbb_narrow_M-3000_13TeV-madgraph
Radion
GluGluToRadionToHHTo4B_M-600_narrow_13TeV-madgraph
GluGluToRadionToHHTo4B_M-650_narrow_13TeV-madgraph
GluGluToRadionToHHTo4B_M-750_narrow_13TeV-madgraph
GluGluToRadionToHHTo4B_M-800_narrow_13TeV-madgraph
RadionTohhTohbhbb_narrow_M-1000_13TeV-madgraph
RadionTohhTohbhbb_narrow_M-1200_13TeV-madgraph
RadionTohhTohbhbb_narrow_M-1400_13TeV-madgraph
RadionTohhTohbhbb_narrow_M-1600_13TeV-madgraph
Non-resonant Signal
GF_HHTo4B_node10_13TeV-madgraph-pythia8
GF_HHTo4B_node11_13TeV-madgraph-pythia8
GF_HHTo4B_node12_13TeV-madgraph-pythia8
GF_HHTo4B_node1_13TeV-madgraph-pythia8
GF_HHTo4B_node2_13TeV-madgraph-pythia8
GF_HHTo4B_node3_13TeV-madgraph-pythia8
GF_HHTo4B_node4_13TeV-madgraph-pythia8
GF_HHTo4B_node5_13TeV-madgraph-pythia8
GF_HHTo4B_node6_13TeV-madgraph-pythia8
GF_HHTo4B_node7_13TeV-madgraph-pythia8
GF_HHTo4B_node8_13TeV-madgraph-pythia8
GF_HHTo4B_node9_13TeV-madgraph-pythia8
GluGluToHHTo4B_node_SM_13TeV-madgraph

Table 5.3: List of background Monte Carlo samples used.

Background
QCD_HT-100to200
QCD_HT-200to300
QCD_HT-300to500
QCD_HT-500to700
QCD_HT-700to1000
QCD_HT-1000to1500
QCD_HT-1500to2000
QCD_HT-2000toinf
TT_TuneCUETP8M1_13TeV-powheg-pythia8

Table 5.4: The HLT paths used and the corresponding L1 seeds.

HLT path	L1 seeds
PFHT650_WideJetMJJ900DEtaJJ1p5	HTT160-255
AK8PFHT650_TrimR0p1PT0p03Mass50	HTT240-320
AK8PFHT700_TrimR0p1PT0p03Mass50	HTT240-320
PFHT800	HTT160-255
PFHT900	HTT160-255
AK8PFJet360_TrimMass30	SingleJet180/200
AK8DiPFJet280_200_TrimMass30_BTagCSV_p20	SingleJet180/200

Chapter 6

Background Estimate

Once we have made our event selection, we must find a way to determine if new physics has been discovered. In the CMS collaboration, this is done by estimating the background in a data-driven manner and then comparing this estimate to the actual data chosen through the event selection. This is done by finding a sideband of data that is kinematically similar to the signal region, or data events chosen through the event selection. This sideband region must also be statistically independent from the signal region. In this analysis we use the Alphabet Background Estimate to estimate the QCD multijet contribution to the background, which comprises roughly 90% of the background. The remaining 10% is $t\bar{t}$, which is estimated using $t\bar{t}$ MC.

6.1 Alphabet Background Estimate

We measure the QCD background with a data-driven method which exploits a number of sidebands of the signal region. These sidebands are defined with respect to the AK8 jet based on its mass and the value of the double-b tagger. Considering these two variables, we may define several regions, as outlined in

Figure 6.1, where the figure consists of all events which pass all selection requirements except the soft-drop mass requirements and double b-tagger requirements on the AK8 jet. The *signal* region is the subset of those events which pass the mass requirement (softdrop mass 105-135 GeV) and pass the double b-tagger requirement we impose on our signal (double b-tagger > 0.8). The *anti-tag* region is the subset of events which pass the mass requirement (softdrop mass 105-135 GeV) but fail the double b-tagger requirement (double b-tagger < 0.8). All other events in the pre-tag region constitute the mass-sideband (softdrop mass < 105 or > 135 GeV, both passing and failing double b-tagger requirement). For QCD background, we expect the shapes but not normalizations of distributions of events in the signal and anti-tag regions to be similar. If there were no correlation between the jet mass and the double b-tagger variable, we could simply measure the ratio of passing to failing events in the mass sideband and scale the anti-tag region by this ratio to obtain an estimate of the signal region (this method is often referred to as the ABCD method).

However, the tagging variable we consider has a dependence on mass. Thus, to obtain the normalization, we must measure a *conversion rate* or the pass-fail ratio, $R_{p/f}$, which depends on mass. We can fit for such a dependence in the mass-sidebands. The pass-fail ratio $R_{p/f}$ is correlated with the value chosen for the cut on the double b-tagger, in this case, 0.8. Should a higher value have been chosen, the ratio would be lower, and should a lower value have been chosen, the ratio would be higher.

This background estimate provides a good estimate of QCD background because it tends to have a relatively smooth distribution in softdrop mass. However, tt does not have a smooth distribution in softdrop mass. Since it is a small

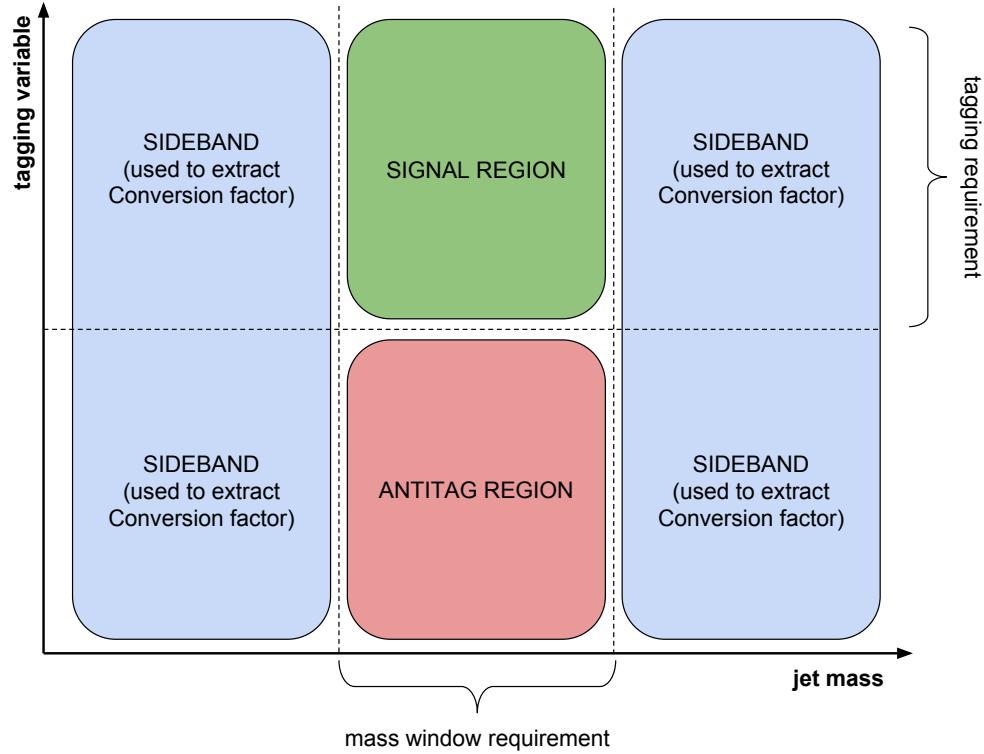


Figure 6.1: Schematic representation of the regions used to perform our estimate.

contribution to the background and $t\bar{t}$ is modeled relatively well by MC, we can use $t\bar{t}$ MC to estimate this background contribution. However, we do not want to double count for this background while performing the QCD estimate, so before the estimate of QCD background is made from data, we subtract the $t\bar{t}$ MC distribution from data. Once the estimate is performed, we add $t\bar{t}$ MC in the signal region to this estimate for a full estimate of the background.

The background estimation is performed in two different control regions in data as well as in the signal region with MC to ensure that it is properly working first, documented in the next two sections. Once this was verified,

the background estimation was performed with the full selection in order to compare this estimate with data in the signal region to identify any differences between the estimation and data that might suggest a discovery of new physics, documented in the last section.

6.2 Closure Test in Data

We look at two control regions in data, one where we require AK4 dijet mass < 70 GeV and remove the triAK4jet mass requirement, and the other where we require AK4 jet 2 deep CSV < 0.6324 and remove the triAK4jet mass requirement. Both control regions reject events that pass the boosted selection, and are statistically independent from the actual selection used to estimate the signal region while still having similar kinematic properties. We use a quadratic fit in the mass sidebands of the conversion rate $R_{p/f}$ to pass the requirement double- $b > 0.8$, after first subtracting tt MC from data. A quadratic function was chosen based off of a Fisher F-test which determined that a linear function was not sufficient, but a cubic function provided no appreciable improvement in the fit.

Figure 6.2 shows the $R_{p/f}$ (left) and the background estimate as a function of M_{jjj}^{red} (right) for the two different categories for the first closure region. The true value of the conversion rate in this mass window is also shown in the figure, as well as the true values for M_{jjj}^{red} . The conversion rate is then applied (as a function of mass) to the anti-tag region to obtain an estimate of the signal region. The estimated and true background in signal region are shown on the right. The background estimate, much like the trigger efficiency calculation, is

separated into two categories: $|\Delta\eta(\text{AK8}, 1\text{AK4}+2\text{AK4})|$ 0-1.0 (top) and 1.0-2.0 (bottom).

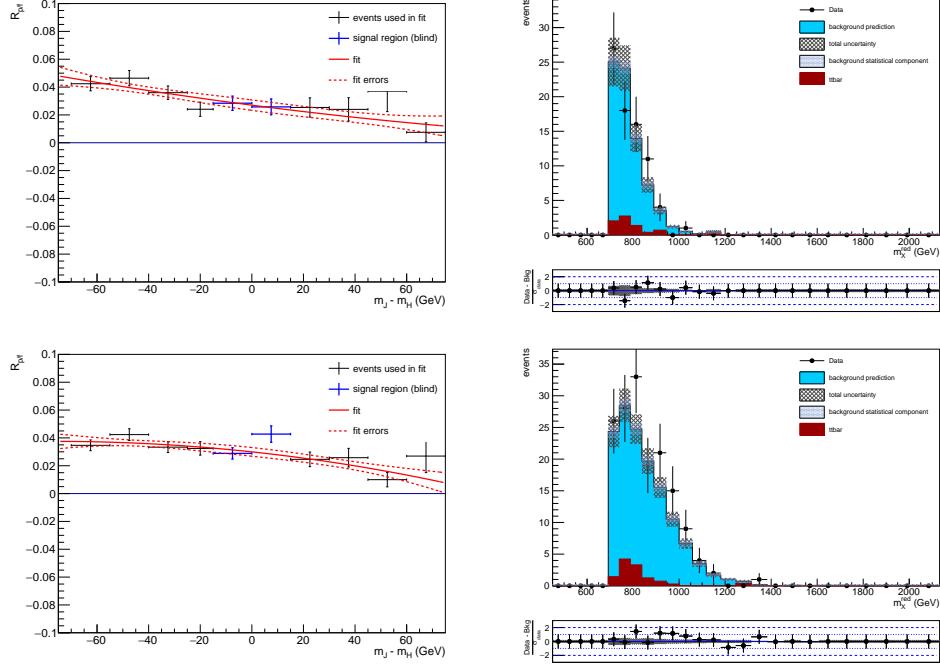


Figure 6.2: First control region (AK4 dijet mass $< 70\text{GeV}$ and remove the triAK4jet mass requirement) for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the first control regions, compared with the true background (black markers). Top is $|\Delta\eta| = 0\text{-}1.0$ and bottom is $|\Delta\eta| = 1.0\text{-}2.0$.

Two systematic uncertainties arise naturally from this estimate. The dominant error is the uncertainty in the fit to the mass sideband regions, which is shown as a dashed line enveloping the fit. This error can be treated as fully correlated between all mass bins when setting limits. The second source of error comes from propagating the statistical uncertainty in the anti-tag region to the signal region. This error is uncorrelated between bins and is smaller than the fitting error. Both errors are shown on the plots on the right as total uncertainty,

and are further discussed in the following chapter.

The $R_{p/f}$ and M_{jjj}^{red} are shown for the second control region in 6.3. We note that the second control region has more statistics.

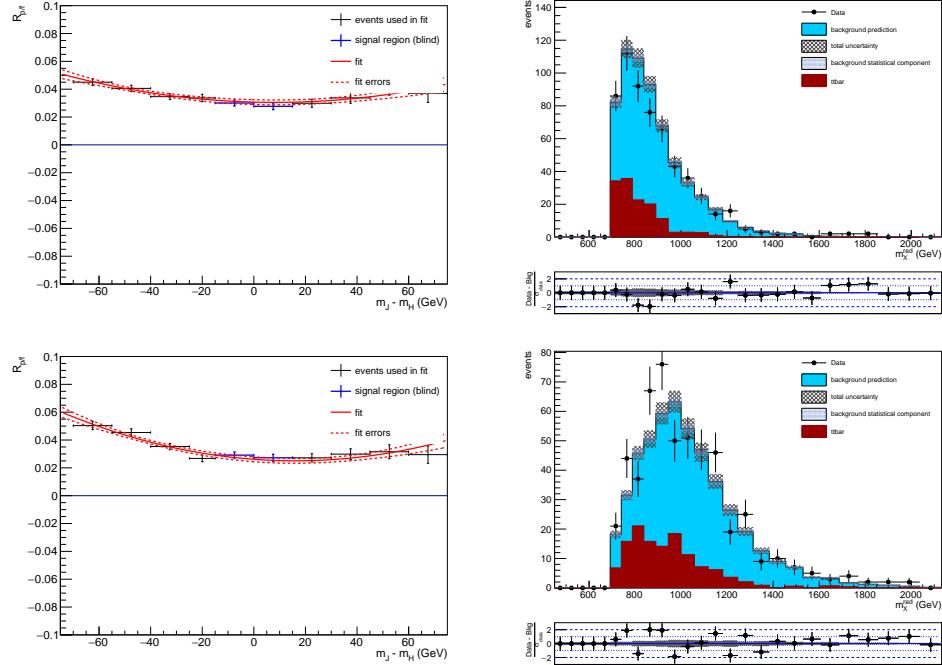


Figure 6.3: Second control region (AK4 jet 2 deep CSV < 0.6324 and remove the triAK4jet mass requirement) for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the control regions, compared with the true background (black markers). Top is $|\Delta\eta| = 0-1.0$ and bottom is $|\Delta\eta| = 1.0-2.0$.

Both control regions show a relatively good agreement between the estimate and the true data, leading us to conclude that the estimation is working well in these control regions.

6.3 Closure Test in MC

We also examine the signal region in MC for a selection rejecting boosted events, as shown in Figure 6.4. This was done by using QCD MC instead of data, and not subtracting $t\bar{t}$. Due to low statistics and reweighting MC, the distributions, in particular the $|\Delta\eta| = 1.0\text{-}2.0$ distribution, are not smooth. This is because each QCD sample has a certain number of events and a certain cross section. The low H_T samples have relatively few events that pass selection and a high cross section, while the high H_T samples have more events that pass selection and a lower cross section. Since the weight of each MC event is equal to the cross section multiplied by the total luminosity of data and divided by the total number of MC events in that sample previous to any cut being applied, this means that the distributions may not be smooth if there are a couple highly weighted events added together with several lower weighted events.

6.4 Signal Region in Data

The unblinded conversion rate $R_{p/f}$ and application of that fit to the anti-tag region to estimate the background in the signal region in data can be found in Figure 6.5 for a selection rejecting boosted events, in Figure 6.6 for a selection rejecting boosted and resolved events, and in Figure 6.7 for a selection retaining both boosted and resolved events. The agreement for a selection rejecting boosted events is good, as well as the agreement for a selection rejecting both boosted and resolved events.

There is a statistical fluctuation that causes the background to underestimate data in the signal region for the selection rejecting no other analyses'

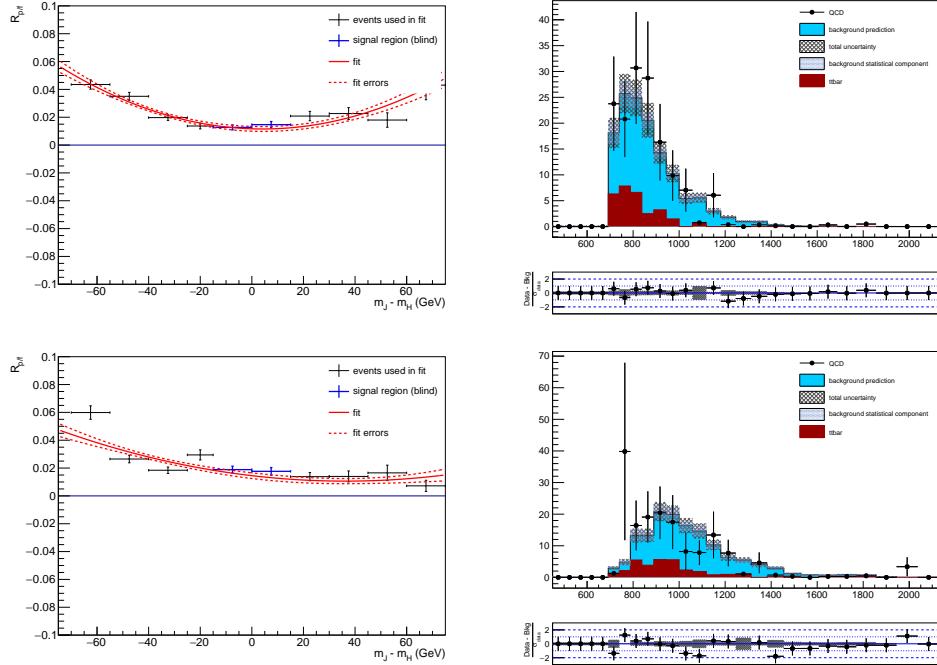


Figure 6.4: Signal region in MC for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the first control regions, compared with the true background (black markers). Top is $|\Delta\eta| = 0\text{-}1.0$ and bottom is $|\Delta\eta| = 1.0\text{-}2.0$.

events in the $|\Delta\eta| = 0\text{-}1.0$ region. Studies were performed to determine if this excess was due to $t\bar{t}$ or a contribution from W or Z boson. By varying the amount of $t\bar{t}$, we found that this did not have an appreciable impact on the estimate, and W and Z boson contributions were found to be negligible. We also studied the $R_{p/f}$ fit with different binning in the sidebands and found that this had an impact on the agreement, further supporting the idea that this is merely a statistical fluctuation.

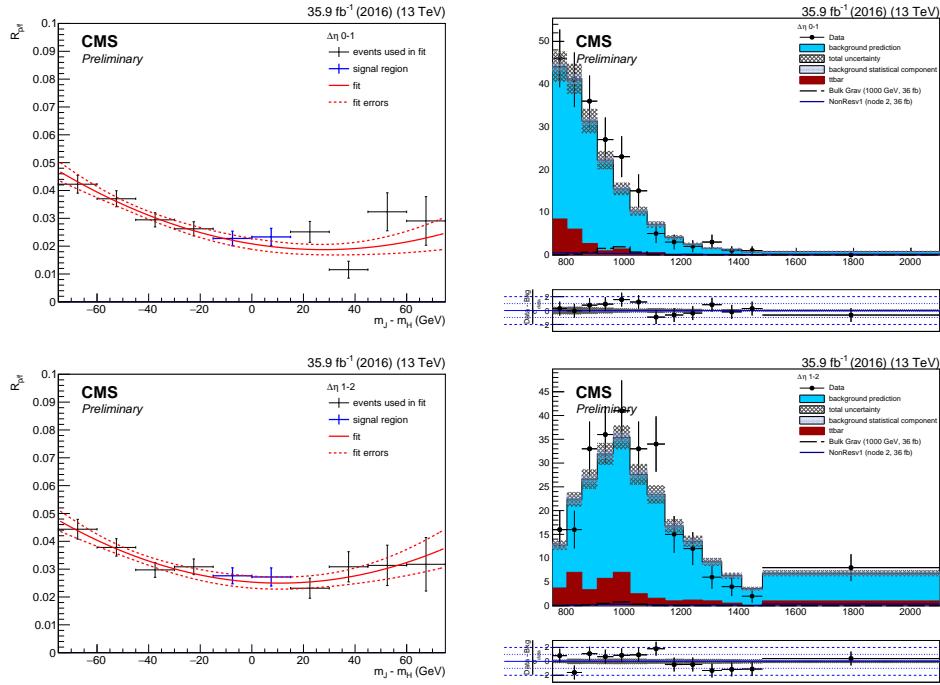


Figure 6.5: Signal region in data for a selection rejecting boosted events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the unblinded signal region. Top is $|\Delta\eta| = 0-1.0$ and bottom is $|\Delta\eta| = 1.0-2.0$.

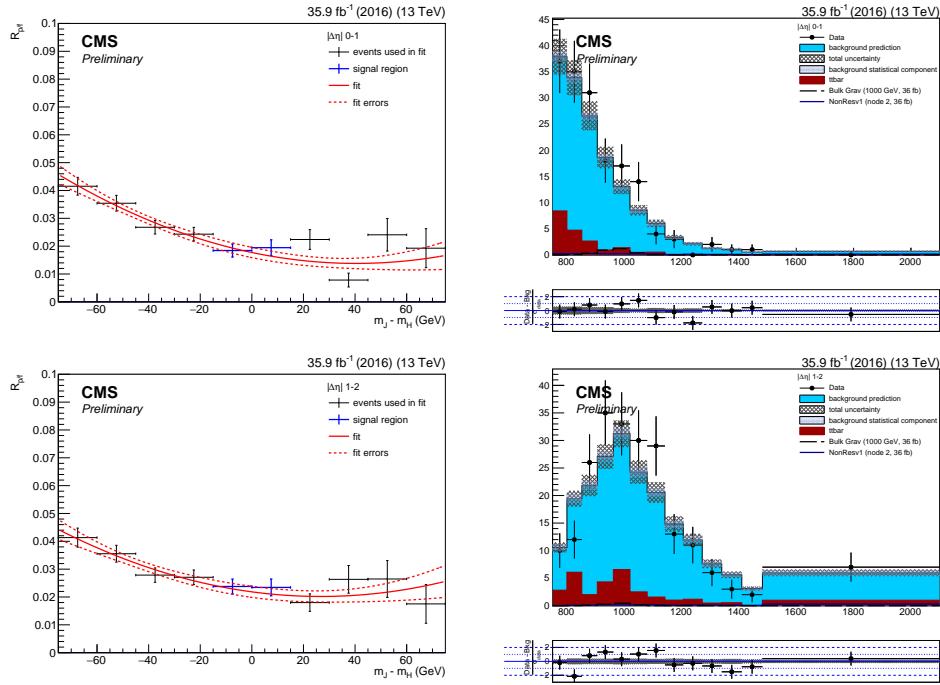


Figure 6.6: Signal region in data for a selection rejecting boosted and resolved events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the unblinded signal region. Top is $|\Delta\eta| = 0-1.0$ and bottom is $|\Delta\eta| = 1.0-2.0$.

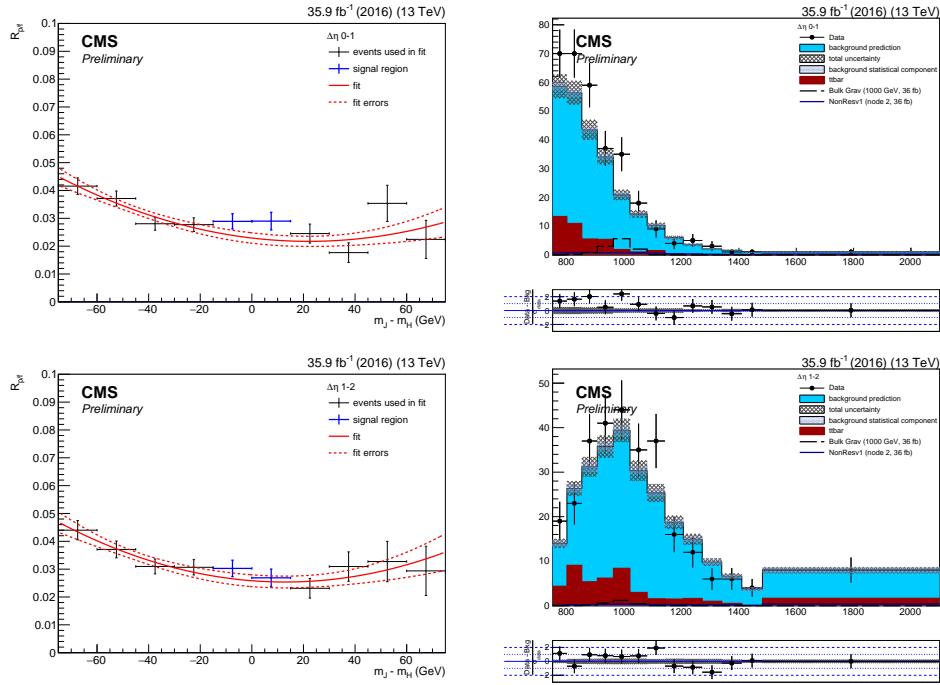


Figure 6.7: Signal region in data for a selection retaining boosted and resolved events. (left) Fits in the mass sideband regions for the conversion rates $R_{p/f}$. (right) Application of those fits to the anti-tag region to estimate the background in the unblinded signal region. Top is $|\Delta\eta| = 0\text{--}1.0$ and bottom is $|\Delta\eta| = 1.0\text{--}2.0$.

Bibliography

- [1] “Preliminary communications on electric discharges in rarefied gases,” Monatsberichte der Kauniglich Preussischen Akademie der Wissenschaften zu Berlin, p. 286, 1876.
- [2] CERN, “Go on a Particle Quest at the First CERN Webfest,” <https://cds.cern.ch/journal/CERNBulletin/2012/35/News%20Articles/1473657>, 2012.
- [3] S. Chatrchyan *et al.*, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,” *Phys. Lett. B*, vol. 716, p. 30, 2012.
- [4] G. Aad *et al.*, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B*, vol. 716, p. 1, 2012.
- [5] D. de Florian *et al.*, “Handbook of LHC Higgs cross sections: 4. deciphering the nature of the Higgs sector,” CERN, CERN Report CERN-2017-002-M, 2016.
- [6] D. de Florian and J. Mazzitelli, “Higgs Boson Pair Production at Next-to-Next-to-Leading Order in QCD,” *Phys. Rev. Lett.*, vol. 111, p. 201801, 2013.

- [7] J. Baglio, A. Djouadi, R. GrÃüber, M. M. MÃijhlleitner, J. Quevillon, and M. Spira, “The measurement of the Higgs self-coupling at the LHC: theoretical status,” *JHEP*, vol. 04, p. 151, 2013.
- [8] L. Randall and R. Sundrum, “A large mass hierarchy from a small extra dimension,” *Phys. Rev. Lett.*, vol. 83, p. 3370, 1999.
- [9] W. D. Goldberger and M. B. Wise, “Modulus stabilization with bulk fields,” *Phys. Rev. Lett.*, vol. 83, p. 4922, 1999.
- [10] O. DeWolfe, D. Z. Freedman, S. S. Gubser, and A. Karch, “Modeling the fifth dimension with scalars and gravity,” *Phys. Rev. D*, vol. 62, p. 046008, 2000.
- [11] C. Csaki, M. Graesser, L. Randall, and J. Terning, “Cosmology of brane models with radion stabilization,” *Phys. Rev. D*, vol. 62, p. 045015, 2000.
- [12] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, “Phenomenology of the Randall-Sundrum Gauge Hierarchy Model,” *Phys. Rev. Lett.*, vol. 84, p. 2080, 2000.
- [13] C. Csaki, M. L. Graesser, and G. D. Kribs, “Radion dynamics and electroweak physics,” *Phys. Rev. D*, vol. 63, p. 065002, 2001.
- [14] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, “Warped Gravitons at the LHC and Beyond,” *Phys. Rev. D*, vol. 76, p. 036006, 2007.
- [15] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, “Searching for the Kaluza-Klein Graviton in Bulk RS Models,” *JHEP*, vol. 09, p. 013, 2007.

- [16] A. Djouadi, “The anatomy of electroweak symmetry breaking. Tome II: the Higgs bosons in the minimal supersymmetric model,” *Phys. Rept.*, vol. 459, p. 1, 2008.
- [17] R. Barbieri, D. Buttazzo, K. Kannike, F. Sala, and A. Tesi, “One or more Higgs bosons?” *Phys. Rev. D*, vol. 88, p. 055011, 2013.
- [18] K. Agashe, O. Antipin *et al.*, “Warped Extra Dimensional Benchmarks for Snowmass 2013,” in *Community Summer Study 2013: Snowmass on the Mississippi (CSS2013) Minneapolis, MN, USA, July 29-August 6, 2013*, 2013. [Online]. Available: <http://cp3-origins.dk/research/units/ed-tools>
- [19] P. de Aquino, K. Hagiwara, Q. Li, and F. Maltoni, “Simulating graviton production at hadron colliders,” *JHEP*, vol. 06, p. 132, 2011. [Online]. Available: <http://feynrules.irmp.ucl.ac.be/wiki/RSmodel>
- [20] A. Oliveira, “Gravity particles from Warped Extra Dimensions, a review. Part I - KK Graviton,” 2014, submitted on arXiv.
- [21] G. Aad *et al.*, “Search For Higgs Boson Pair Production in the $\gamma\gamma b\bar{b}$ Final State using pp Collision Data at $\sqrt{s} = 8$ TeV from the ATLAS Detector,” *Phys. Rev. Lett.*, vol. 114, p. 081802, 2015.
- [22] ———, “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*, vol. 75, p. 412, 2015.

- [23] ——, “Searches for Higgs boson pair production in the $hh \rightarrow bb\tau\tau, \gamma\gamma WW^*, \gamma\gamma bb, bbbb$ channels with the ATLAS detector,” *Phys. Rev. D*, vol. 92, p. 092004, 2015.
- [24] V. Khachatryan *et al.*, “Searches for heavy Higgs bosons in two-Higgs-doublet models and for $t \rightarrow ch$ decay using multilepton and diphoton final states in pp collisions at 8 TeV,” *Phys. Rev. D*, vol. 90, p. 112013, 2014.
- [25] ——, “Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 8 TeV,” *Phys. Lett. B*, vol. 749, p. 560, 2015.
- [26] ——, “Searches for a heavy scalar boson H decaying to a pair of 125 GeV Higgs bosons hh or for a heavy pseudoscalar boson A decaying to Zh, in the final states with $h \rightarrow \tau\tau$,” *Phys. Lett. B*, vol. 755, pp. 217–244, 2016.
- [27] J. Brehmer *et al.*, “The Diboson Excess: Experimental Situation and Classification of Explanations; A Les Houches Pre-Proceeding,” 2015, submitted on arXiv.
- [28] G. Aad *et al.*, “Search for high-mass diboson resonances with boson-tagged jets in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *JHEP*, vol. 12, p. 055, 2015.
- [29] ——, “Search for production of WW/WZ resonances decaying to a lepton, neutrino and jets in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *Eur. Phys. J. C*, vol. 75, p. 209, 2015.

- [30] ——, “Search for resonant diboson production in the $\ell\ell q\bar{q}$ final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector,” *Eur. Phys. J. C*, vol. 75, p. 69, 2015.
- [31] V. Khachatryan *et al.*, “Search for massive resonances in dijet systems containing jets tagged as W or Z boson decays in pp collisions at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 08, p. 173, 2014.
- [32] ——, “Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV,” *JHEP*, vol. 08, p. 174, 2014.
- [33] G. F. Giudice, C. Grojean, A. Pomarol, and R. Rattazzi, “The Strongly-Interacting Light Higgs,” *JHEP*, vol. 06, p. 045, 2007.
- [34] W. BuchmÃjller and D. Wyler, “Effective lagrangian analysis of new interactions and flavour conservation,” *Nuclear Physics B*, vol. 268, no. 3, pp. 621 – 653, 1986. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0550321386902622>
- [35] A. Carvalho, M. Dall’Osso, T. Dorigo, F. Goertz, C. A. Gottardo, and M. Tosi, “Higgs pair production: choosing benchmarks with cluster analysis,” *Journal of High Energy Physics*, vol. 2016, no. 4, p. 126, Apr 2016. [Online]. Available: [https://doi.org/10.1007/JHEP04\(2016\)126](https://doi.org/10.1007/JHEP04(2016)126)
- [36] “Lhc the guide,” <http://cds.cern.ch/record/1092437/files/CERN-Brochure-2008-001-Eng.pdf>.
- [37] D. Bourilkov, R. C. Group, and M. R. Whalley, “LHAPDF: PDF use from the Tevatron to the LHC,” 2006.

- [38] R. D. Ball *et al.*, “Parton distributions for the LHC Run II,” *JHEP*, vol. 04, p. 040, 2015.
- [39] S. Chatrchyan *et al.*, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, ser. Technical Design Report CMS. Geneva: CERN, 2006, there is an error on cover due to a technical problem for some items.
- [40] CMS Collaboration, “Particle-flow event reconstruction in CMS and performance for jets, taus, and missing e_T ,” CMS Physics Analysis Summary CMS-PAS-PFT-09-001, 2009. [Online]. Available: <http://cdsweb.cern.ch/record/1194487>
- [41] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm,” *JHEP*, vol. 04, p. 063, 2008.
- [42] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop,” *JHEP*, vol. 05, p. 146, 2014.
- [43] D. Adams *et al.*, “Towards an Understanding of the Correlations in Jet Substructure,” *Eur. Phys. J.*, vol. C75, no. 9, p. 409, 2015.
- [44] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness,” *JHEP*, vol. 1103, p. 015, 2011.
- [45] ——, “Maximizing Boosted Top Identification by Minimizing N-subjettiness,” *JHEP*, vol. 1202, p. 093, 2012.

- [46] H. Khanpour, S. Khatibi, M. Khatiri Yanehsari, and M. Mohammadi Najafabadi, “Single top quark production as a probe of anomalous $tq\gamma$ and tqZ couplings at the FCC-ee,” *Phys. Lett.*, vol. B775, pp. 25–31, 2017.
- [47] A. M. Sirunyan *et al.*, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV,” *Submitted to: JINST*, 2017.
- [48] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification,” *JHEP*, vol. 10, p. 059, 2014.
- [49] S. Chatrchyan *et al.*, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS,” *JINST*, vol. 6, p. P11002, 2011.
- [50] CMS Collaboration, “Status of the 8 TeV Jet Energy Corrections and Uncertainties based on 11 fb^{-1} of data in CMS,” 2013, cMS detector performance summary. [Online]. Available: <https://cds.cern.ch/record/1545350>
- [51] “Recommended jet energy corrections and uncertainties for data and mc,” 2015, https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#Recommended_for_Data. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#Recommended_for_Data
- [52] “Jet identification recommendations for 13 tev data analysis,” 2015, https://twiki.cern.ch/twiki/bin/view/CMS/JetID#Recommendations_for_13_TeV_data. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMS/JetID#Recommendations_for_13_TeV_data

- [53] CMS Collaboration, “Search for heavy resonances in the W/Z-tagged dijet mass spectrum at 13 TeV,” CERN, CMS Analysis Note CMS-AN-16-235, 2016.
- [54] A. M. Sirunyan *et al.*, “Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV,” 2017.
- [55] CMS Collaboration, “CMS Luminosity Measurement for the 2016 Data Taking Period,” CERN, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. [Online]. Available: Inpreparation
- [56] “Utilities for accessing pileup information for data: 2015 pileup json files,” 2015, https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#2015_Pileup_JSON_Files. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#2015_Pileup_JSON_Files
- [57] CMS Collaboration, “Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC: Boosted event topologies,” CERN, CMS TWiki r14, 2017. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation80XReReco#Boosted_event_topologies
- [58] ——, “Measurement of W-tagging data/MC scale factors using ttbar semi-leptonic events using full 2016 dataset,” CERN, CMS Analysis Note CMS-AN-17-051-v2, 2017.
- [59] ——, “W/Z-tagging of Jets,” CERN, CMS TWiki r37, 2016. [Online]. Available: <https://twiki.cern.ch/twiki/bin/view/CMS/JetWtagging>

- [60] “Search for heavy resonances decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV,” CERN, Geneva, Tech. Rep. CMS-PAS-B2G-16-026, 2017. [Online]. Available: <http://cds.cern.ch/record/2264684>
- [61] J. S. Conway, “Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra,” in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, 2011, pp. 115–120. [Online]. Available: <https://inspirehep.net/record/891252/files/arXiv:1103.0354.pdf>

Alice A. Cocoros

2 Somerville Ct. Apt. G, Parkville, MD 21234, asady1@jhu.edu, 775-781-7336

Education and Honors

- Ph.D., Physics and Astronomy The Johns Hopkins University, expected June 2018
M.A., Physics and Astronomy The Johns Hopkins University, 2014, 4.00
B.A. with Honors, Astrophysics Williams College, 2013, 3.50
- National Science Foundation Graduate Research Fellow

Clare Boothe Luce Fellow

Mellon Mays Undergraduate Fellow

Research Experience

- **CMS Experiment at CERN**, JHU Dept. Physics & Astronomy, Baltimore, MD; June 2013 - present
 - Thesis: Search for New Physics in the Two Higgs to Four b-quark Final State*
 - Designed, developed, and implemented data processing for a search for new particles and processes that decay to two Higgs bosons which each decay to two b quarks, using data taken with the Compact Muon Solenoid (CMS) detector on the Large Hadron Collider (LHC) at CERN in 2016 and Monte Carlo (MC) simulation produced at CERN. Compared signal MC and background MC to optimize signal to background ratio, using numerical methods and data visualization. Performed data driven background estimate to compare with signal region data to look for any significant deviations in the expected Standard Model distribution.

–Wrote simulation software which was added to an existing CMS detector simulation to model the merging of hits in the pixel detector, which will improve the performance of the simulation in high energy environments. Aided in the update of the current pixel detector simulation to make the framework more adaptable to changes, using plugins for each process instead of hard-coding everything.

–Performed calibration testing on the double b-tagger, a machine-learning variable used to identify the likeliness of a jet containing two b-quarks. Evaluated the difference in performance between data and MC for a multivariate analysis variable designed to identify a jet with two b-quarks.

–Designed a novel variable that distinguishes signals with a lepton inside a jet from multijet QCD background. Advisor: Professor Petar Maksimovic (JHU)

- **Exploring a Theory of Dark Matter**, Williams College, Williamstown, MA; June 2012 - June 2013
 - Senior Honors Thesis: Exploring a Theory of Dark Matter*
 - Developed a new theory that could explain the nature of dark matter. Examined an extension to the Standard Model that includes a dark matter candidate, a new particle called Z' that mediates interactions between the Standard Model particles and dark matter particles, and an additional new particle that explains how the Z' has mass. Ran relic abundance simulations, produced Monte Carlo simulations of new physics, studied dark matter direct detection limits and Z' detection limits from current experiments to place constraints on the model. Advisor: Professor David

Tucker-Smith (Williams)

- **ATLAS Experiment at CERN**, Boston University, Geneva, Switzerland; Jan-Aug 2012
 - Measured properties of the Standard Model with the Top Quark Pair Differential Cross Section Group within A Toroidal LHC Apparatus (ATLAS) Collaboration.
 - Searched for new particles using data taken by ATLAS on the LHC, comparing Monte Carlo (MC) simulation of new physics signals with Standard Model background to achieve a more comprehensive understanding of kinematic phasespace for new physics models and determine which variables improve the ratio of signal to background.
 - Research performed as part of a study abroad program, Boston University Geneva Physics Program, Spring 2012. Advisor: Professor Kevin Black (BU)
- **Photoionization Models of Planetary Nebulae in M31**, Williams College, Williamstown, MA; June-Dec 2011
 - Analyzed spectra and modeled 16 planetary nebulae in M31 to learn about the properties of both the nebula and its central star, helping us understand how they are formed and how material originally created in the star is recycled. Advisor: Professor Karen Kwitter (Williams)

Sample Publications and Talks

- –A. Cocoros and the CMS Collaboration. Search for heavy resonances decaying to a pair of Higgs bosons in four b quark final state in proton-proton

collisions at $\sqrt{s}=13$ TeV. Paper. CERN, Geneva, 2017. arxiv:1710.04960.

Submitted to Physics Letters B.

– “*Search for heavy resonances decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*”. Talk.

APS April Meeting 2017, Washington, D.C., 2017.

– A. Cocoros and the CMS Collaboration. Identification of double-b quark jets in boosted event topologies. Paper. CERN, Geneva, 2017. arxiv:1712.07158.

Submitted to Journal of Instrumentation.

– “*Identifying Boosted New Physics With Non-Isolated Leptons*”. Talk.

APS April Meeting 2015, Baltimore, MD, 2015.

– C. Brust, P. Maksimovic, A. Sady (Cocoros) et al. Identifying Boosted New Physics With Non-Isolated Leptons. Paper. Baltimore, MD, 2014. 10.1007/JHEP04(2015)079.

– “*Exploring a Theory of Dark Matter*”. Poster. Joint Fall 2012 Meeting of the American Physical Society New England Section and the American Association of Physics Teachers, Williams College, 2012.

– “Measurement of top quark pair differential cross section with ATLAS in pp collisions at $\sqrt{s} = 7$ TeV.” Tech. Rep. ATL-COM-PHYS-2012-1137, CERN, Geneva, 2012.

– “*Photoionization Models of Planetary Nebulae in M31*”. Talk and publication (along with M. Hosek and A. King). Keck Northeast Astronomy Consortium Student Research Symposium, Wellesley College, 2011.

Teaching Experience

- **Preparing Future Faculty Certificate Recipient**, October 2017
 - Three stage program designed to introduce graduate students and post-docs to pedagogy and effective teaching practices.
 - Workshops through CIRTL (first stage): Creating Effective Learning Communities in Teaching and Research, Reaching and Teaching Diverse Learners, Developing Work-Life Resilience, and Writing an Effective Teaching Philosophy Statement.
 - Johns Hopkins Teaching Institute (second stage): Three-day intensive exploring the benefits of active learning, ongoing assessment, and responsiveness to diversity, examining a variety of teaching practices and principles and participating in peer-evaluated micro-teaching exercises.
 - Taught winter intersession course, “Exploring the Building Blocks of the Universe” (third stage).
- **Electricity and Magnetism I Teaching Assistant**, Johns Hopkins University, Spring 2017, 4 credits
 - Aided in active learning style lectures, answering questions about practice problems, guiding students to the answer in the time allotted.
 - Taught a problem solving section every week, directing small groups of students to solve practice problems, managing group dynamics to ensure all students are learning the material.
 - Held office hours and counseled students in best test-taking strategies.

- **Intersession Instructor**, Johns Hopkins University, January 2016, January 2017, 1 credit
 - Designed and taught “Exploring the Building Blocks of the Universe”, a three week Winter Intersession physics course for non-majors.
 - Wrote and delivered lectures on particle physics, incorporated demonstrations and audio-visual aids, designed and guided labs on how to analyze LHC data, wrote and delivered and graded a final exam.
 - Adjusted Jan. 2017 curriculum based on Jan. 2016 student reviews, adding demos, shortening lecture material, and lengthening the amount of time for labs.
- **Observatory TA**, Williams College, September 2010 - June 2011, 5 hrs/week
 - Directed students in introductory astronomy courses in night sky observing involving constellation, binocular, 70 mm telescope visual, and .6 m telescope CCD camera projects, answered homework questions.
- **Johns Hopkins University Center For Talented Youth Teaching Assistant**, Santa Cruz, CA, June-July 2010, 40 hrs/week
 - Center for Talented Youth Teaching Assistant for Introduction to Astronomy, aided instructor in teaching a basic course on astronomy for gifted students grades 7-12 enrolled in CTY program in Santa Cruz, CA.
 - Designed and taught several lessons in basic physics and astronomy.

Leadership and Outreach

- **Chair JHU Physics/Astronomy Diversity Group:**

- Restarted the Physics and Astronomy Diversity Group, served as chair for four years.
- Started and maintained a mentorship program, aimed at pairing students, postdocs, and junior faculty with volunteer mentors to foster relationships and support outside of traditional academic relationships (such as advisors, teachers), took on several undergraduate mentees through the program, helping them address a range of issues, such as difficulty in classes, securing internships, and fitting into the social network of physicists at JHU.
- Organized and conducted meetings several times a year, ranging from planning meetings of the committee to meetings for underrepresented groups to network.
- Scheduled biannual diversity speakers, on a range of subjects as they pertain to the physics community, including issues related to LGBTQ folks, disabilities, gender, race, and mentorship.

- **Member JHU Physics/Astronomy Graduate Students Outreach:**
 - Taught physics to Baltimore high school students with fellow Outreach members through a partnership with a local school, bringing demonstrations of simple physics concepts.
 - Constructed a portable planetarium for Baltimore elementary school students and brought the planetarium to the elementary school, teaching an all-day unit on space with fellow Outreach members.
- **Member Johns Hopkins Diversity Leadership Council:**
 - Attended regular monthly meetings of the Diversity Leadership Council,

focused on addressing widespread issues of diversity across the institution.

–Participated in two subcommittees, one related to STEM diversity and the other related to faculty diversity. Documented existing STEM pipeline recruitment and retention activities at JHU Krieger School of Arts and Sciences. Aided in efforts to examine faculty recruitment and retention on a university-wide level.

- **Creator/Co-President Society for Physics Students at Williams:**

–Co-started the SPS group at Williams. Created infrastructure for a mentorship program, to help younger students gain access to older students. Designed a website and gathered alumni advice on applying to graduate school, applying to jobs, senior thesis projects, summer research, etc.. Started an outreach program with a local rural elementary school, bringing demonstrations to teach them more about physics.

Skills and Interests

- **Programming Skills:** Python, C++/C, multivariate analysis, Monte Carlo simulation, GitHub, UNIX, Mathematica, Root, L^AT_EX

- **Interests:** fluent in French, 20 years classical piano, 18 years classical ballet, 8 years vocal training, 7 years performing in community musical theatre