

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON DECAYING TO TWO MUONS
AT THE CMS EXPERIMENT

By

ANDREW CARNES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2018

© 2018 Andrew Carnes

ACKNOWLEDGMENTS

I would like to thank my Mom and Dad for just being all around great parents. Without all of their support, the completion of this Ph.D would never have been possible. I would also like to thank Professors Paul Avery and Darin Acosta for their help. In addition, I would like to thank Andrew Brinkerhoff, Pierluigi Bortignon, and Andrea Marini for all of their support and hard work on the project. I'd also like to thank my friends Stany, Leo, Sean, Sophie, and David for making Geneva a great place to live.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	3
LIST OF TABLES	7
LIST OF FIGURES	8
ABSTRACT	11
 CHAPTER	
1 INTRODUCTION	13
2 THE LARGE HADRON COLLIDER AND THE CMS EXPERIMENT	15
2.1 The Large Hadron Collider	15
2.2 The Compact Muon Solenoid	17
2.2.1 Silicon Tracker	18
2.2.2 Calorimeters	19
2.2.3 Muon System	21
2.2.3.1 Drift Tubes	22
2.2.3.2 Cathode Strip Chambers	23
2.2.3.3 Resistive Plate Chambers	24
2.2.4 Trigger System	24
2.2.5 Level-1 Trigger	25
3 THE STANDARD MODEL	27
3.1 Quantum Field Theory, Symmetries, And The Higgs Boson	29
3.2 The Lagrangian For A Free Scalar Particle	30
3.3 The Lagrangian For A Free Spin $\frac{1}{2}$ Particle	31
3.4 The Lagrangian For A Free Spin 1 Particle	35
3.5 Interactions	36
3.6 The Higgs Mechanism	40
3.7 The Standard Model Higgs Boson And The LHC	45
4 DISCOVERY, LIMITS, MEASUREMENT, AND SENSITIVITY	50
4.1 Discovery	50
4.2 Limits	52
4.3 Measurement and Uncertainty	52
4.4 Sensitivity	54
4.5 Constraints And Nuisance Parameters	54
4.6 A Particle Physics Search	56
4.6.1 Sensitivity	57
4.6.2 Test Statistic	58

4.6.3	Constraints	59
4.7	Using The Test Statistic	59
5	THE $H \rightarrow \mu^+ \mu^-$ SEARCH STRATEGY	61
6	SAMPLES, CORRECTIONS, AND SELECTIONS FOR THE $H \rightarrow \mu^+ \mu^-$ SEARCH	63
6.1	Data And Monte Carlo Samples	63
6.2	Muon Momentum Calibration	67
6.2.1	Muon Corrections In Data	68
6.2.2	Data-MC Agreement In Scale, Resolution	71
6.2.3	Derivation Of Systematic Uncertainties	77
6.3	Event Selection, Object Reconstruction, And Further Corrections	77
6.3.1	Muons	78
6.3.2	Jets And MET	78
6.3.3	Pile-up Reweighting	79
6.3.4	Muon Efficiency Scale Factors	79
6.3.5	Jet, MET, And B-tagging Corrections	80
6.3.6	Event Selection	81
7	OPTIMIZING THE SENSITIVITY OF THE $H \rightarrow \mu^+ \mu^-$ SEARCH	82
7.1	Improving The Level-1 Muon Trigger	82
7.1.1	Metrics Of Success	83
7.1.2	The EMTF Regression Project	84
7.1.3	Putting The BDTs Into A Look-up Table	86
7.1.4	Results And Conclusions	86
7.2	Classifying Events	88
7.2.1	Training Boosted Decision Trees	88
7.2.2	Event Categorization	90
7.2.3	The Autocategorizer	91
7.2.4	Final Categories	93
7.2.5	Validating The Categories	94
7.2.6	Signal And Background	102
8	SIGNAL AND BACKGROUND MODELS FOR THE $H \rightarrow \mu^+ \mu^-$ SEARCH	110
8.1	Modeling The Signal	110
8.2	Modeling The Background	115
8.3	Systematic Uncertainties	117
8.3.1	Signal	117
8.3.1.1	Shape uncertainties	118
8.3.1.2	Category migration uncertainties	118
8.3.1.3	Rate uncertainties	119
8.3.1.4	Impact plots	120
8.3.2	Background	121

9	RESULTS FOR THE $H \rightarrow \mu^+ \mu^-$ SEARCH	126
9.1	Results For 13 TeV Data	126
9.2	Results Combining 7, 8, And 13 TeV Data	132
10	CONCLUSIONS AND OUTLOOK	134
	REFERENCES	136
	BIOGRAPHICAL SKETCH	139

LIST OF TABLES

<u>Table</u>		<u>page</u>
6-1 Overview of the single muon data stream collected during the proton-proton collisions at $\sqrt{s} = 13$ TeV by CMS at the LHC in 2016.	64	
6-2 The Higgs signal MC samples were generated with POWHEG while the parton shower and hadronization processes are modeled by the PYTHIA8 generator with TuneCUETP8M1.	65	
6-3 The background MC samples were generated with amc@NLO. POWHEG and MADGRAPH. Spin effects in multi-boson processes are simulated using MADSPIN. The parton shower and hadronization processes are modeled by the PYTHIA8 generator with TuneCUETP8M1.	66	
7-1 The feature discretization scheme for each mode.	87	
7-2 The importance is given by the percentage of the signal and background PDFs along that feature that do not overlap.	89	
7-3 Full Width Half Max (FWHM) of the signal peak, expected signal and background yields within the FWHM, purity, sensitivity, and the expected upper limits at 95% confidence. The background yields and the expected limits use background models fit to the data in the sidebands.	94	
7-4 The total signal in each category broken down by process.	102	
8-1 Maximum observed bias per category for least biased fit function. The maximum bias is over $m_H = 120, 125, 130$ and GeV.	124	

LIST OF FIGURES

<u>Figure</u>		<u>page</u>
2-1 The CERN Accelerator Complex.	16
2-2 The CMS detector.	17
2-3 The sagitta measurement.	19
2-4 A slice of the CMS detector.	20
2-5 A look at the muon system.	22
2-6 A drift tube.	23
2-7 A cathode strip chamber.	23
2-8 The L1 trigger architecture.	25
3-1 The Standard Model particles.	28
3-2 Higgs production cross sections.	45
3-3 Higgs boson branching fractions.	47
3-4 Feynman diagrams for the Higgs boson production modes.	48
3-5 Quark-gluon scattering.	48
4-1 An illustrated example of a p-value.	51
4-2 Defining the confidence interval.	53
4-3 An example of the signal and background used for the background only and S+B hypotheses.	56
6-1 Standard Model Higgs boson production modes.	63
6-2 An example fit used for the dimuon mass calibrations.	67
6-3 Rochester and Kalman Filter muon corrections on the Z peak mean in data.	69
6-4 Rochester and Kalman filter muon corrections on the Z peak resolution in data.	70
6-5 The uncalibrated Z peak mean and its alignment in data and MC.	71
6-6 The Z peak mean and its alignment in data and MC after Rochester corrections.	72
6-7 The Z peak mean and its alignment in data and MC after Kalman Filter corrections.	73
6-8 The uncalibrated Z peak resolution in data and MC.	74
6-9 The Z peak resolution's alignment in data and MC after Rochester corrections.	75

6-10	The Z peak resolution's alignment in data and MC after Kalman Filter corrections.	76
6-11	Kalman filter statistical uncertainty on the dimuon mass.	77
7-1	EMTF rate reduction and efficiency.	87
7-2	The BDT ROC score for signal vs background classification.	90
7-3	BDT score distributions for signal MC, background MC, and data.	91
7-4	An example of the signal and background histograms for the Autocategorizer.	92
7-5	An example of the autocategorization process.	93
7-6	The final categorization.	94
7-7	Validation plots of the BDT input variables.	95
7-8	Validation plots of the BDT input variables and the dimuon mass.	96
7-9	Data/MC agreement for the most sensitivity category.	97
7-10	Additional Data/MC agreement plots for the most sensitive category.	98
7-11	The BDT score distribution in data and MC.	99
7-12	The BDT distribution in data and MC by quantile.	100
7-13	BDT mass independence plots.	101
7-14	Kinematic and object count distributions for the inclusive signal MC.	103
7-15	More kinematic and object count distributions for the inclusive signal MC.	104
7-16	Even more kinematic and object count distributions for the inclusive signal MC.	105
7-17	Kinematic and object count distributions for the signal MC in the most sensitive category.	106
7-18	More kinematic and object count distributions for the signal MC in the most sensitive category.	107
7-19	Even more kinematic and object count distributions for the signal MC in the most sensitive category.	108
7-20	Background shapes in data in the different categories.	109
8-1	An example of the signal model.	111
8-2	The signal model compared to MC predictions.	112
8-3	The signal composition, resolution, and purity in the different categories.	113

8-4	The efficiency times acceptance versus the Higgs mass.	114
8-5	Some background-only fits.	116
8-6	An example of the F-Test for the ordered background fits.	117
8-7	Uncertainty on some jet kinematics and counts.	120
8-8	Impact plots using the Asimov dataset.	121
8-9	Background bias results for the categories 0 to 8.	123
8-10	Background bias results for the categories 9 to 14.	124
8-11	The net bias of the combined background model.	125
9-1	The upper limit on the signal strength for the individual categories on 13 TeV data alone.	127
9-2	The net upper limit on the signal strength for 13 TeV data alone.	127
9-3	The p-value on the background-only hypothesis using 13 TeV data alone.	128
9-4	A plot of the negative log likelihood near the minimum for the signal strength.	128
9-5	The weighted average of the signal + background fits for 13 TeV data.	129
9-6	Signal + background fits for the individual categories on 13 TeV data.	130
9-7	More signal + background fits for the individual categories on 13 TeV data.	131
9-8	The upper limit on the signal strength combining 7, 8, and 13 TeV data.	132
9-9	The p-value on the background only hypothesis combining 7, 8, and 13 TeV data.	133
9-10	The negative log likelihood for the combined 7, 8, and 13 TeV data.	133
10-1	The expected p-value for a 125 GeV SM Higgs boson versus the amount of 13 TeV data.	135

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON DECAYING TO TWO MUONS
AT THE CMS EXPERIMENT

By

Andrew Carnes

May 2018

Chair: Paul Avery

Major: Physics

In 2012 two collaborations at the Large Hadron Collider announced the discovery of a new particle with properties similar to the Standard Model Higgs Boson. In order to determine whether the boson discovered with a mass of 125 GeV is actually the Standard Model Higgs boson, all of the different ways the particle can decay need to be investigated. If the probabilities for the different decays do not match the predictions of the Standard Model then this would imply new physics.

This dissertation presents the search for the Standard Model Higgs boson decaying to $\mu^+\mu^-$. The search uses the $35.9 \pm 0.9 \text{ fb}^{-1}$ of $\sqrt{s} = 13 \text{ TeV}$ proton-proton collision data recorded by the CMS detector in 2016. The signal strength ($\mu = (\sigma\mathcal{B})/(\sigma\mathcal{B})_{\text{SM}}$) is measured at $0.7^{+1.1}_{-1.0}$ for $m_H = 125 \text{ GeV}$, where σ is the Higgs production cross section and \mathcal{B} is the branching fraction to muons. The observed and expected upper limits on the signal strength at a 95 % confidence level are presented for Higgs masses in the range 120 to 130 GeV. The observed and expected upper limits on the signal strength at a mass of 125 GeV are 2.64 and 2.08 respectively. The significance is reported in the same range, and the observed and expected significance at $m_H = 125 \text{ GeV}$ are 0.74σ and 0.98σ respectively.

Combined results for 5.0 fb^{-1} of 7 TeV, 19.8 fb^{-1} of 8 TeV, and 35.9 fb^{-1} of 13 TeV data are also presented. For $m_H = 125 \text{ GeV}$, the combination yields a measured signal strength of $0.9^{+1.0}_{-0.9}$, observed (expected) upper limits at 95% confidence of 2.64 (1.89), and an observed (expected) significance of 0.98 (1.09) σ . The results correspond to an upper limit on the

$H \rightarrow \mu^+ \mu^-$ branching fraction of 5.7×10^{-4} . These results provide the best results to date on the Higgs coupling to second generation fermions. No deviations from the Standard Model are observed.

CHAPTER 1

INTRODUCTION

The Standard Model (SM) of particle physics is an extremely successful theory shown to correctly predict the behavior of the particles and forces which make up the most basic constituents of the universe. In fact, it correctly describes all of the forces known except for gravity [1]. In particular, the SM predicts that the massive particles of the theory acquire their mass by interacting with a scalar particle called the Higgs boson [2–5]. On July 4, 2012 two collaborations at the Large Hadron Collider (LHC), the A Toroidal LHC Apparatus (ATLAS) and Compact Muon Solenoid (CMS), announced the discovery of a new boson at 125 GeV with properties similar to the Standard Model Higgs [6–8]. This discovery was fueled by the investigation into the Higgs decays to the vector bosons ZZ and $\gamma\gamma$. Soon after, evidence for the Higgs coupling to matter was found through the $\tau^+\tau^-$ and $b\bar{b}$ decays [9–12]. Whether the newly discovered boson is indeed the expected Standard Model Higgs remains to be determined. Insofar, all of the different decay modes will be investigated to search for deviations from the Standard Model predictions.

This leads to the study of the Higgs decay to $\mu^+\mu^-$. Although this decay is the smallest branching fraction expected to be detected [13, 14], the dimuon decay offers high efficiency and excellent momentum resolution, which should lead to a narrow peak over the falling background, mostly Drell Yan events. The tiny branching fraction enables greater sensitivity to small deviations from the predicted decay rate and in this respect offers an advantage over other channels where a minuscule deviation could be drowned out. Furthermore, the Higgs coupling to second generation fermions remains to be determined.

This dissertation presents the search for the Standard Model Higgs Boson decaying to $\mu^+\mu^-$ using the proton-proton collision data recorded by the CMS experiment in 2016. In order to maximize the data available for the search, the first machine learning in the L1 Trigger system at the LHC is developed and deployed for 2016 data collection. To further maximize the sensitivity of the search, an additional machine learning technique is invented to categorize

events based upon the detector resolution and the event kinematics. The search looks for a Higgs boson with a mass between 120 and 130 GeV and presents the expected and observed upper limits on the rate of $H \rightarrow \mu^+ \mu^-$ production in this range as well as the p-values on the background-only hypothesis. The best fit for the rate of production is also presented. The 13 TeV results are then combined with the 2012 $H \rightarrow \mu^+ \mu^-$ results on 7 and 8 TeV data [?].

The dissertation first covers the LHC which accelerates and collides the protons. The dissertation then presents the CMS detector which measures the paths, momentum, and energy of the emerging particles. Next, the dissertation explains the theory underlying the Standard Model and its predictions of the Higgs particle. After, the machine learning implementation in the L1 trigger that reduced the number of fakes in the data by a factor of three is detailed. Finally, the search for H to $\mu^+ \mu^-$ is presented.

CHAPTER 2

THE LARGE HADRON COLLIDER AND THE CMS EXPERIMENT

2.1 The Large Hadron Collider

The Large Hadron Collider is a particle collider near Geneva, Switzerland run by the European Organization for Nuclear Research (CERN). The LHC is the largest and most powerful particle collider ever built, designed to collide protons with a center of mass energy of 14 TeV and a luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [15]. The luminosity is given by

$$L = \frac{n_b f N_p^2 \gamma}{4\pi \epsilon_n \beta^*} \quad (2-1)$$

where n_b is the number of bunches in each ring, f is the frequency for a bunch to circle the ring, N_p gives the number of protons in a bunch, and γ is the Lorentz factor. ϵ_n is the normalized transverse emittance, a measure of the spread of the beam in momentum and position space. β^* measures the focus of the beam at the interaction point, and $\epsilon_n \beta^*$ represents the transverse area at the point of interaction. The large luminosity at the LHC is characterized by a high frequency of bunch crossings (every 25 ns) with about 10^{11} protons in each bunch packed as densely as possible, resulting in a high rate of collisions. With many collisions at high energy, the detectors can collect enough events from yet unexplored energy regimes to discover new physics, to verify old physics, or to discard certain theories of physics.

The collider itself is 26.7 km in circumference and 45-170 m underground. 8.3 T supercooled, superconducting magnets operating at 2 K steer the high energy proton beams. In order to save money, the LHC not only reuses the tunnels of a previous collider, the Large Electron Positron Collider (LEP), but also reuses older accelerators which were state of the art at their time. These older accelerators ramp up the energy of the protons and inject them into the LHC. All of this together makes up the CERN accelerator complex.

First, the protons are created from a source of Hydrogen gas. The hydrogen atoms of the gas are placed into a large electric field that separates the atoms into unbound protons and electrons. The protons are then sent to a radio frequency quadrupole which focuses the

CERN's Accelerator Complex

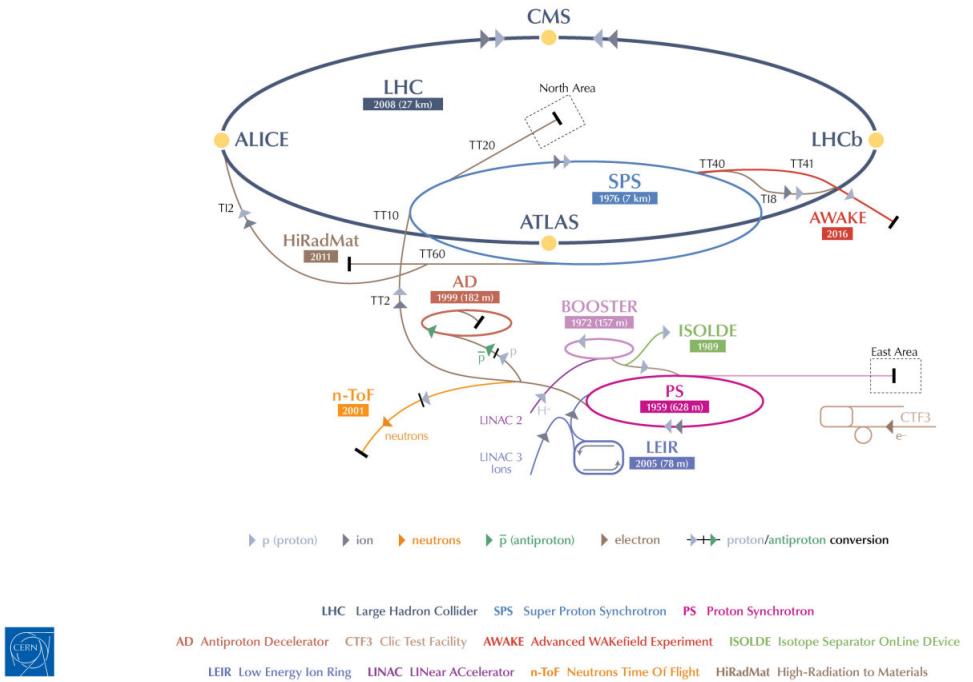


Figure 2-1. The CERN Accelerator Complex [16].

protons and accelerates them. The radio frequency field is stronger for the protons in the back than in the front and consequently squeezes them into a tighter bunch. The protons then proceed to a linear accelerator, LINAC2, where they are accelerated to 50 MeV or 5% of the speed of light (c). The protons then enter a series of synchrotrons. A synchrotron is a device that accelerates particles by guiding them around a fixed circular path with a magnetic field while boosting their speed with an electric field as they pass a certain point. Since a faster particle bends less in the same magnetic field, the magnetic field strength is synchronized with the speed of the accelerating particles to keep them in the fixed circular path.

After LINAC2, the protons enter the first of the synchrotrons, the Proton Synchrotron Booster (PSB) accelerating the protons to 1.4 GeV (0.81c). From here the protons are injected into the Proton Synchrotron (PS) and accelerate to 25 GeV (0.999c). The PS then injects the protons into the Super Proton Synchrotron (SPS) further accelerating them to 450 GeV

(0.99999c). Finally the protons are injected into the LHC where they accelerate up to 6.5 TeV (0.9999999c). Once accelerated to the appropriate collision energy, the proton beams are made to collide in the different detectors located around the ring. By colliding enough protons at large enough energies, it is possible to probe corners of physics that have never been seen before. The two general purpose detectors at the LHC, ALTAS and CMS, are used to look for signs of new physics like the Higgs boson, dark matter, and extra dimensions by measuring the energy, the momentum, and the paths of the particles coming out of the collisions.

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS), located in Cessy, France, is 21.6 m long, 15 m in diameter, and weighs more than the Eiffel Tower. Not only is the CMS detector a massive and complex device it's also run by a huge collaboration involving approximately 3,800 people from 200 institutes spanning 43 different countries [17]. The greatest achievement of the collaboration to date is the discovery of a Higgs like particle in 2012, a feat shared with ATLAS.

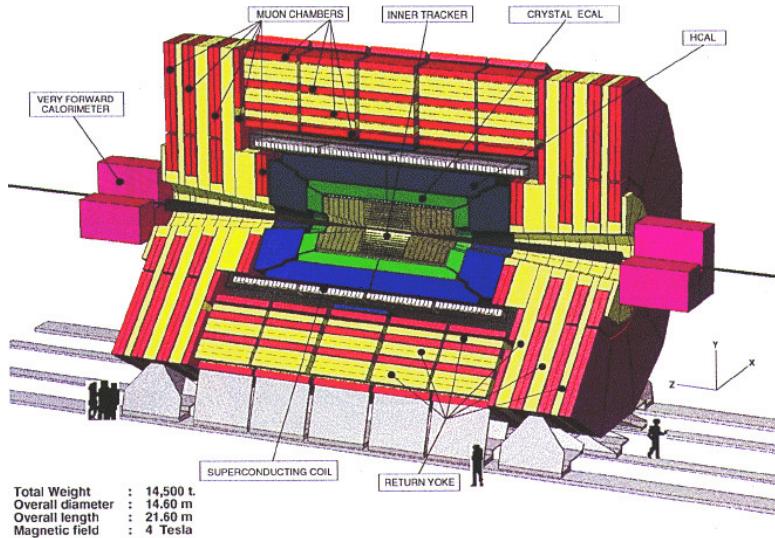


Figure 2-2. The CMS detector [18]

CMS was built primarily to look for the Standard Model Higgs and signs of Beyond Standard Model (BSM) physics like Supersymmetry, extra dimensions, or new heavy weak bosons [19]. Because BSM and Higgs decays to muons and electrons often have the highest

signal to background ratio, CMS is designed to identify and measure these particles with a high accuracy. A high signal to background ratio means that the events of interest have fewer look-alikes. Jets¹ and photons are measured to a high degree of accuracy as well. In order to measure the energy, momentum, and location of the different types of particles CMS deploys a variety of subdetectors working in concert. The defining feature of the detector is an extremely powerful solenoid which enables the accurate measurement of momentum for charged particles. The tracker and calorimeters fit snugly within the 6 m diameter solenoid. The muon detectors reside outside the magnet but within the return yoke.

2.2.1 Silicon Tracker

The 3.8T magnetic field inside the solenoid enables the tracker to measure the transverse momentum of charged particles based upon the curvature of the track. Charged particles with lower transverse momentum (p_t) bend more in a magnetic field than high p_t particles. As such, a measurement of the deviation of a curved track from a straight line, the sagitta, can be used to measure the curvature and determine the momentum [20].

$$p_t \cong \frac{L^2 q B}{8s} \quad (2-2)$$

Here L is the length of the straight line between the first and last position measurements, q is the charge of the particle, B is the magnetic field, and s is the sagitta. The equation for the error in the momentum measurement shows that a higher magnetic field enables better p_t resolution, illuminating the design choice for a powerful magnet.

$$\frac{\delta p_t}{p_t} \propto \frac{p_t}{L^2 B} \quad (2-3)$$

The silicon tracker is made of tiny reverse biased bipolar diodes. When a charged particle travels through one of these diodes, the particle liberates electron hole pairs beyond the

¹ When a quark or gluon is created, it pulls other quarks from the vacuum in order to maintain colorlessness. The result is a cone of colorless particles called a jet.



Figure 2-3. The sagitta measurement.

electrostatic equilibrium, and a current flows. The tracker needs to be small enough such that the particles traveling through it don't deposit much energy. Energy deposition in the tracker would throw off energy measurements in the calorimeters. This means that the tracker needs to be smaller than a few radiation lengths². At CMS, the thickest part of the tracker is one radiation length. The tracker is placed nearest the collision point in order to identify primary and secondary vertices and to measure the momentum of particles before they are tainted by interactions with other detectors.³ Being so near the collision point, the silicon tracker is bombarded by a constant flux of high intensity radiation. As such, the tracker is carefully designed to be robust to this radiation rich environment.

2.2.2 Calorimeters

The Electromagnetic Calorimeter (ECAL) is right outside the tracker and its main goal is to measure the energy of electrons and photons. It's designed to contain entire electromagnetic showers for these particles and is consequently many radiation lengths thick. The ECAL is made of lead tungstate scintillating crystals which release an amount of light proportional to the energy deposition. The light is collected and the total energy is calculated. The separation

² the length scale over which an electron deposits a substantial amount of energy into the material

³ A vertex is an interaction point from which a set of particles emanate.

into individual crystals allows some spatial resolution as well. Particles with larger mass deposit less energy per unit distance into a solid. Many of the hadronic particles make it through the ECAL for this reason.

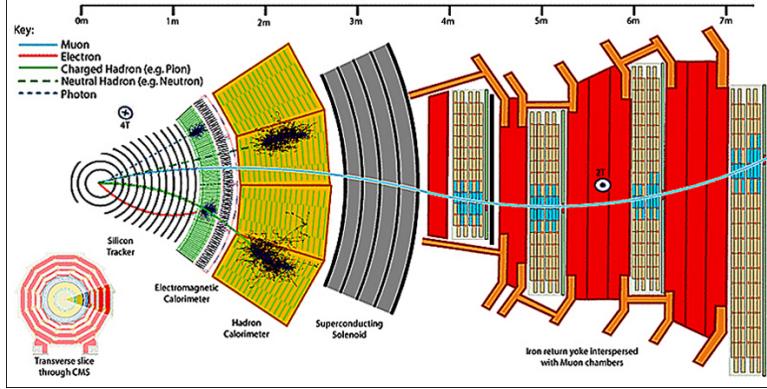


Figure 2-4. A slice of the CMS detector [18].

The HCAL is placed outside the ECAL to collect the energy of the particles that survived the other subsystems. The surviving particles that interact strongly with the HCAL are mostly hadronic particles from jets. The HCAL works in a similar manner to the ECAL except that layers of plastic scintillating material are interspersed with layers of a dense passive absorber like brass or steel. The density of the passive absorber increases the chance of interaction and shower production thus reducing the total length of the hadronic shower and enabling the measurement of the total energy for most of the cascades. Again, the scintillation light is collected to determine the energy. If the ECAL and HCAL were placed outside the magnet the particles would interact with the solenoid material before entering the calorimeters, throwing off their measurements. The showers in the calorimeters are a consequence of the electromagnetic and strong forces, which means that particles without these interactions pass through the materials undetected, e.g. neutrinos or BSM weakly interacting particles. Since the momentum in an interaction is conserved, any imbalance means that some particles escaped the detector. If there is an excess of missing momentum beyond the amount expected due to neutrinos this may indicate the existence of dark matter or some other BSM particle. In order to measure the missing energy correctly, it's important that the HCAL is built without

any gaps and that it is dense enough to collect the energy of the strongly and electrically interacting particles.

While the momentum resolution in the tracker is proportional to the momentum, the energy resolution in the calorimeters decreases with increasing energy [20].

$$\frac{\delta E}{E} = \sqrt{\left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C} \quad (2-4)$$

The first term in the square root describes statistical fluctuations. The energy measured is proportional to the number of photons captured which has poissonian fluctuations and the error (δE) for this term is $\propto \sqrt{E}$. The second term describes noise in the electronics whose error is energy independent, and the last term describes the errors in energy calibration which are proportional to energy.

2.2.3 Muon System

Neutrinos aren't the only Standard Model particles that make it through the tracker, ECAL, and HCAL. Muons have a relatively long lifetime $\sim 10^{-6}$ s with $c\tau \sim 100$ m. The large gamma factor associated with the large energies of the muons at the LHC, in combination with their long lifetime, enables them to travel hundreds of kilometers on average, well through the entire CMS detector before decaying. Muons are charged so their tracks show up in the tracker and some energy is deposited in the calorimeters but, muons are so much more massive than electrons that the energy deposition in the ECAL is minimal. Making it through the ECAL, the muons enter the HCAL. The HCAL is designed to stop strongly interacting hadronic particles and collect their energy. But muons don't interact with the strong force, and they make it through the HCAL as well. This enables the muon system to be placed outside the magnet.

The muon system consists of a few different types of detectors which all involve the same basic principle. The charged muon ionizes some gas and the ionized particles are attracted to charged surfaces initiating a current in the surfaces. With a large enough voltage differential between the charged surfaces the the ionized particles may gain enough kinetic energy to further ionize other atoms in the gas initiating an avalanche effect which reduces the need for

signal amplification later. The muon system uses this strategy in the different detectors. The types of detectors in the muon system are the Cathode Strip Chambers (CSC), the Drift Tubes (DT), and the Resistive Plate Chambers (RPC) [19].

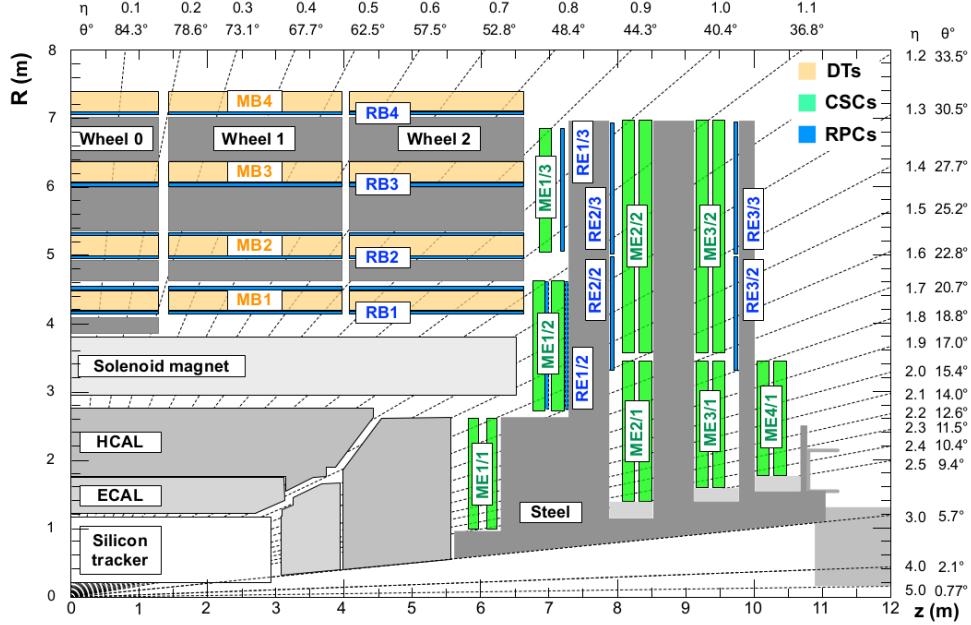


Figure 2-5. A look at the muon system [21].

2.2.3.1 Drift Tubes

The drift tubes are located in the barrel portion of CMS. Throughout the majority of the barrel the magnetic field is basically uniform. The drift tubes have aluminum plates on the top and bottom separated by aluminum I-beams as shown in Figure 2-6. A wire acts as the anode and the I-beams are the cathodes. The tubes are designed to provide a constant drift velocity throughout each tube.

When a charged particle flies through the tube, it ionizes the gas inside. The electrons drift at constant velocity to the anode. The distance from the anode is deduced from the drift time, utilizing the fact that the ionized electrons drift with a constant velocity. This calculation does however require a reference time. In each chamber the drift tubes are placed in layers and the average crossing time in the chamber is used as the reference time.

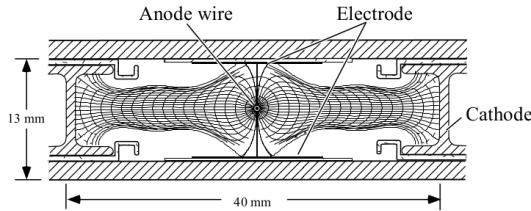


Figure 2-6. A drift tube [22].

2.2.3.2 Cathode Strip Chambers

The CSCs are located in the endcaps of the detector which range in $|\eta|$ from 0.9 to 2.4. The nonuniform magnetic field in the endcaps would adversely affect the drift times in the DTs, so the endcaps use CSCs instead. In this system, there are oppositely charged strips and wires running roughly perpendicular to each other.

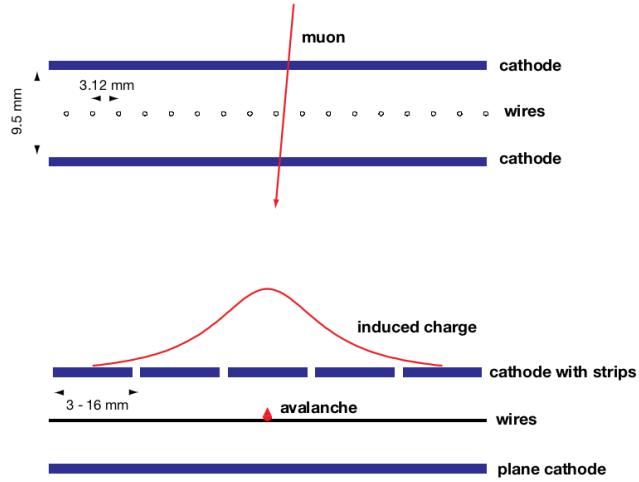


Figure 2-7. A cathode strip chamber [22].

When a muon flies through a CSC, it induces charge on the wires and the strips and ionizes gas in the chamber. The ionized particles in the gas float to the charged strips and wires initiating a current in those components. The induced charge from the muon itself also contributes to the currents. The most intense currents are those associated with the location of the muon, and the position resolution in the ϕ direction is roughly $100 \mu\text{m}$.

2.2.3.3 Resistive Plate Chambers

The RPCs are located both in the barrel and in the endcaps. The RPCs have excellent timing resolution on the order of 1 ns. The RPCs use their excellent timing resolution to determine each particle's bunch crossing of origin. The accurate and rapid timing information helps with the online selection of muons, a huge priority for CMS considering that many interesting collisions produce muons. For this reason, the RPCs focus on efficient online selection of muons instead of accurate offline reconstruction [23].

With their accurate timing measurements, the RPCs complement the DTs and CSCs. The RPCs consist of two high resistance parallel plates surrounding a volume of gas. The outsides of the plates are painted with graphite paint forming the electrodes. A large voltage differential is kept between the electrodes. When a charged particle crosses the plates it induces an electrical discharge in the plates which remains localized in time and space due to the large resistivity.

2.2.4 Trigger System

Collision events come at a rate of 40 MHz with each event taking up roughly a MB of information. If the detector had to store all of the information from each event this would amount to pushing terabytes of information into a storage system every second, which is remarkably infeasible. To deal with this issue, CMS utilizes a trigger system which selects only interesting events and cuts the rate down from 40 MHz to 1 KHz [23]. Since bunch crossings happen every 25 ns, the trigger needs to operate at an incredibly high rate.

CMS tackled this issue by dividing the trigger into different tiers. The Level-1 Trigger is the first stage of the trigger system made from custom hardware which can operate at fantastic speed. The Level-1 Trigger reduces the rate from 40 MHz to 100 KHz. The events passing the L1 Trigger move onto the High Level Trigger (HLT) which further reduces the rate to 1 KHz. Due to the lower input rate the HLT can operate in software.

2.2.5 Level-1 Trigger

The Level-1 (L1) Trigger is made of up different subsystems that work together to decide whether to keep the data from a beam crossing for further processing. The University of Florida works with the Level-1 muon trigger system, the Endcap Muon Track Finder (EMTF) in particular. The muon system needs to determine the transverse momenta of muons and their location and choose the best candidates. Each of the different muon detectors have their own local triggers which send their best muon tracks to the Global Muon Trigger (GMT). The GMT chooses the best muon candidates from that set and passes these on to the Global Trigger (GT). The GT combines the information from the calorimeter triggering system and uses the combined information to check whether the bunch crossing should be sent to the HLT or discarded. The L1 Trigger has many different trigger criteria defining separate triggers. These pass/fail information from the separate triggers define the trigger bits. If an event passes any of the triggers then it is forwarded for further processing.

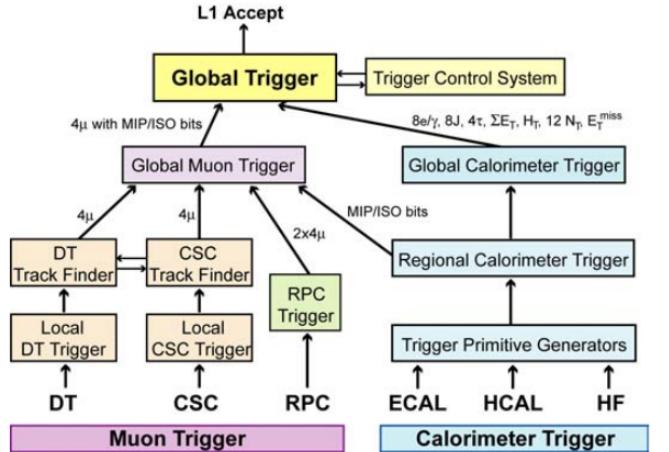


Figure 2-8. The L1 trigger architecture [23].

The Track Finders (TF) play an important role in the L1 Trigger system. The EMTF combines the location and direction information from the different CSC stations into muon tracks and calculates the transverse momenta for the different tracks. The EMTF chooses the best candidates (highest momentum and highest quality) to send to the GMT. The Drift Tube Track Finder (DTTF) performs a similar process for muons in the DT system. The RPC

system calculates the location and direction and forms tracks in the same stage. In the process the RPC trigger system assigns transverse momenta and quality, and like the others chooses the best tracks to send to the GMT.

CHAPTER 3

THE STANDARD MODEL

The Standard Model (SM) of particle physics is an incredibly successful theory that correctly describes the physics of all known particles and forces that make up the universe [24]. Well nearly all, quantum gravity and the non-zero neutrino mass are still a mystery [1, 24, 25]. The particles of the SM come in two types: fermions and bosons. Fermions are the spin $\frac{1}{2}$ particles that make up the different types of matter, and bosons are the integer spin particles responsible for the different forces. Electrons are the most familiar type of fermion, but there are more exotic kinds like the quarks, neutrinos, muons, and the taus. The up and down quarks make up protons and neutrons, which combine to make nuclei, and nuclei combine with electrons to create the atoms that account for nearly all of the matter in our day to day experience. The up and down quarks and the electron are the first of three generations of quarks and leptons¹ with each generation heavier than the next. The up and down quarks are the first generation of quarks, charmed and strange are the next, and top and bottom are the third generation. For the leptons, the electron and electron neutrino are the first generation, the muon and muon neutrino are the second, and the tau and the tau neutrino the third. Each fermion also has a corresponding antiparticle. As an example, the positron is the antiparticle for the electron.

The force carrying particles that allow matter to interact and form more complex objects like atoms, molecules, and even people are the spin 1 bosons. These force carriers are the gluons, photons, and the W and Z particles. Gluons mediate the strong force, photons the electromagnetic force, and the W and Z bosons mediate the weak force. Every force has an associated charge: particles with electric charge can interact through the electromagnetic force, those with color charge may interact via the strong force, and those with isospin or weak

¹ Leptons are fermions like the electron that aren't quarks. The quarks interact with the strong force that binds nuclei together and the leptons do not.



Figure 3-1. The Standard Model particles.

hypercharge may interact through the weak force. The fundamental forces bind the fermions to make the familiar composite objects that surround us. The strong force binds quarks into protons and neutrons and the protons and neutrons into nuclei, while the electromagnetic force binds the electrons and nuclei together to make atoms. The size of the composite objects gives an idea of the relative strength of the forces. A proton is 10^{-15} meters in size while an atom is 10^{-10} meters and a solar system is 10^{12} meters. The more tightly bound the stronger the force. But that isn't quite exact, in fact, the ratio of the strength of the forces is like so $1:10^{-3}:10^{-16}:10^{-41}$, strong : electromagnetic : weak : gravitational ².

Of all the particles predicted by the SM, there is only one spin 0 particle, the Higgs boson, and it plays a special role in the theory. As the universe cooled from the Big Bang, the Higgs field went through a phase transition and settled into a nonzero ground state forming a condensate. The electron, muon, tau and the W and Z particles of the SM interact with the Higgs condensate and acquire mass. The massive particles of the SM are massively only

² Gravity is just included for perspective. The Standard Model does not describe this force and reconciling gravity with quantum mechanics is an open problem.

because of the Higgs boson. With such a large role in the SM, finding this particle or a BSM Higgs boson has been a huge priority for the CMS collaboration [19]. In 2012, a Higgs particle with a mass of 125 GeV was found and to date remains consistent with the Standard Model [6–8]. However, the properties need to be investigated further before declaring the discovered Higgs boson the Higgs particle of the Standard Model.

The next sections build and couple the different Lagrangians to lay out the Standard Model and the Higgs mechanism. After putting the theory together, the more experimental details of the search for $H \rightarrow \mu^+ \mu^-$ are described. In the following sections, \hbar and the speed of light, c , are set to 1. Moreover, 0,1,2,3 and t,x,y,z are used interchangeably to label the components of a four vector. When relevant, 0 represents time and 1,2,3 represent x,y,z respectively. Einstein summation notation is used indicating that repeated indices are summed over, so $x_i x_i y_j y_j$ is shorthand for $\sum_i \sum_j x_i x_i y_j y_j$. Repeated Greek indices assume sums over all four space and time components, while repeated Roman indices assume sums over only the spatial components.

3.1 Quantum Field Theory, Symmetries, And The Higgs Boson

The laws of physics are invariant under boosts and rotations. Any proper Quantum Field Theory (QFT) is described by an invariant Lagrangian respecting these symmetries. Boosts and rotations may be represented by different NxN matrices. The 4x4 matrices act on four vectors, the 2x2 matrices act on spinors, and the 1x1 matrices act on scalars. Scalars, four vector products, and certain spinor products remain invariant under Lorentz transformations so they are a natural ingredients for different QFTs. Scalar fields represent spin 0 particles, spinor fields represent spin 1/2 particles, and four vectors fields represent spin 1 particles. Coupling the spin 1 and spin 1/2 Lagrangians produces forces, and coupling the spin 1/2 and spin 1 fields to a spin 0 field produces mass. The spin 0 particle is the Higgs boson. In order to explain the Higgs coupling, the different Lagrangians are built and coupled.

3.2 The Lagrangian For A Free Scalar Particle

A scalar field remains the same under Lorentz transformations. This representation is used for spin 0 particles. The equations of motion must relate the change in the scalar field Φ at one moment in space and time to the value of Φ at the next moment in space and time, so the Lagrangian must include both the field itself and the four vector derivative, ∂_μ . Note that both Newton's equations of motion, $F = m\ddot{x}$, and the Schrodinger equation for the free particle, $i\partial_t\psi = \frac{1}{2m}(-i\partial_x)(-i\partial_x)\psi$ ³ have at most second order derivatives. The same holds for Maxwell's equations of electromagnetism. Hence, as an assumption, the derivative term in the Lagrangian will be the lowest order possible. As a Lorentz invariant scalar, any power of Φ can be included. On the other hand, ∂_μ is a four vector, and must be paired with a ∂^μ to form the invariant four vector product. Cross terms between these invariant pieces like $\Phi\partial_\mu\partial^\mu\Phi$ or $\partial_\mu\Phi\partial^\mu\Phi$ are also invariant, but lead to feedback between the derivatives and the value of the function. This means that the derivatives $i\partial_t = E$ and $-i\partial_i = P_i$ will change over time, but E and \vec{P} should remain constant for a free particle. The cross terms are thrown out to prevent this. The Φ terms with an order different than Φ^2 cause the same problem, and these are thrown out too. All of these choices lead to the following action,

$$S = \int d^4x (c_0 + c_1\Phi^2 + c_2\partial_\mu\Phi\partial^\mu\Phi + c_3\partial_\mu\partial^\mu\Phi). \quad (3-1)$$

Note that the c_3 term is a total derivative and by the divergence theorem depends only on the values at the boundary, which are fixed. This implies that the contribution to the action from the c_3 term is the same regardless of how Φ changes in the volume. Because the Euler-Lagrange equations depend only on the variation in the volume, this term cannot contribute to δS and c_3 may be set to zero. The c_0 term is a more obvious constant and does

³ Here the one dimensional case is presented for simplicity

not affect δS either, so it may be set to zero as well, leaving,

$$S = \int d^4x (c_1\Phi^2 + c_2\partial_\mu\Phi\partial^\mu\Phi). \quad (3-2)$$

Finding Φ such that $\delta S = 0$ amounts to applying the Euler-Lagrange equations $\partial_\mu \frac{\partial L}{\partial(\partial_\mu\Phi)} = \frac{\partial L}{\partial\Phi}$.

These yield

$$\begin{aligned} 2c_2\partial_\mu\partial^\mu\Phi &= -2c_1\Phi \\ \rightarrow (-E^2 + \vec{P}^2)\Phi &= \frac{c_1}{c_2}\Phi = -m^2\Phi \\ \rightarrow \frac{c_1}{c_2} &= -m^2. \end{aligned} \quad (3-3)$$

In order to get the correct dispersion relation for a relativistic particle, c_1 is set to $-\frac{1}{2}m^2$ and c_2 is set to $\frac{1}{2}$. Notice that including any Φ^n with $n \neq 2$ in the Lagrangian would have contributed to the differential equation via $\frac{\partial L}{\partial\Phi}$ and that E, \vec{P} wouldn't be constant. Thus the equation wouldn't work for a free particle. The resulting equation of motion for the scalar particle, $\partial_\mu\partial^\mu\Phi = -m^2\Phi$, is called the Klein-Gordon equation, and provides the correct description for spin 0 particles. This was all derived using symmetry, a reasonable assumption about the order of the derivatives, and the fact that the Energy shouldn't change over time for a free particle. The final Lagrangian for the scalar particle is

$$S = \int d^4x \frac{1}{2} (\partial_\mu\Phi\partial^\mu\Phi - m^2\Phi^2). \quad (3-4)$$

3.3 The Lagrangian For A Free Spin $\frac{1}{2}$ Particle

With the action for the free scalar particle in hand, the free spin $\frac{1}{2}$ particle is up next. The spin $\frac{1}{2}$ action must combine the \mathcal{L} (left-chiral) and \mathcal{R} (right-chiral) spinors of the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ representations and the four vector, ∂_μ , in a Lorentz invariant way [26]. The \mathcal{L} , \mathcal{R} , and ∂_μ transformations are now analyzed to find the lowest order invariant combinations. A Lorentz transformation for the left-chiral spinor is

$$\Lambda^{(L)} = e^{\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i}, \quad (3-5)$$

and a Lorentz transformation for the right-chiral spinor is

$$\Lambda^{(R)} = e^{\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i}. \quad (3-6)$$

Taking the Hermitian conjugate of the left-chiral transformation gives,

$$(\Lambda^{(L)})^\dagger = e^{-\frac{i}{2}\theta_i\sigma_i^\dagger + \frac{1}{2}\phi_i\sigma_i^\dagger} = e^{-\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i}. \quad (3-7)$$

See Physics from Symmetry [26] for more details. This reveals that $(\Lambda^{(L)})^\dagger$ is the inverse of $\Lambda^{(R)}$. Similarly, $(\Lambda^{(R)})^\dagger$ is the inverse of $\Lambda^{(L)}$. Thus, $\mathcal{L}^\dagger\mathcal{R}$ and $\mathcal{R}^\dagger\mathcal{L}$ are Lorentz invariants, which can be seen below,

$$(\mathcal{L}^\dagger\mathcal{R})' = \mathcal{L}^\dagger(\Lambda^{(L)})^\dagger\Lambda^{(R)}\mathcal{R} = \mathcal{L}^\dagger(\Lambda^{(R)})^{-1}\Lambda^{(R)}\mathcal{R} = \mathcal{L}^\dagger\mathcal{R} \quad (3-8)$$

$$(\mathcal{R}^\dagger\mathcal{L})' = \mathcal{R}^\dagger(\Lambda^{(R)})^\dagger\Lambda^{(L)}\mathcal{L} = \mathcal{R}^\dagger(\Lambda^{(L)})^{-1}\Lambda^{(L)}\mathcal{L} = \mathcal{R}^\dagger\mathcal{L}. \quad (3-9)$$

These are the lowest order invariant pieces involving the field alone. In order to couple the derivative to the field, the \mathcal{L} and \mathcal{R} spinors must attach to ∂_μ in an invariant way. A four vector may be expressed as an outer product of left and right spinors [26],

$$\alpha_{L\tilde{R}} = \mathcal{L}_\alpha \tilde{\mathcal{R}}_\alpha^T. \quad (3-10)$$

The following also works,

$$\alpha_{R\tilde{L}} = \mathcal{R}_\alpha \tilde{\mathcal{L}}_\alpha^T. \quad (3-11)$$

Combining left or right-chiral spinors with a \sim provides a representation of a four vector that preserves the hermitivity and observability of a four vector operator. Unfortunately, the \sim spinors add a bit of bookkeeping.

The \sim transformation sends σ_i to $-\sigma_i*$. So the transformations for the \sim transposed spinors are,

$$(\Lambda^{(\tilde{L})})^T = e^{\frac{i}{2}\theta_i(-\sigma_i*)^T + \frac{1}{2}\phi_i(-\sigma_i*)^T} = e^{-\frac{i}{2}\theta_i\sigma_i^\dagger - \frac{1}{2}\phi_i\sigma_i^\dagger} = e^{-\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i} \quad (3-12)$$

and

$$(\Lambda^{(\tilde{R})})^T = e^{\frac{i}{2}\theta_i(-\sigma_i*)^T - \frac{1}{2}\phi_i(-\sigma_i*)^T} = e^{-\frac{i}{2}\theta_i\sigma_i^\dagger + \frac{1}{2}\phi_i\sigma_i^\dagger} = e^{-\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i}. \quad (3-13)$$

The equations above show that the $(\Lambda^{(\tilde{L})})^T$ transformation is the inverse of $\Lambda^{(L)}$, and that the $(\Lambda^{(\tilde{R})})^T$ transformation is the inverse of $\Lambda^{(R)}$. So the invariant pieces coupling the spinor to the four vector are,

$$\mathcal{R}^\dagger \alpha_{L\tilde{R}} \mathcal{R} = \mathcal{R}^\dagger \mathcal{L}_\alpha \tilde{\mathcal{R}}_\alpha^T \mathcal{R} \quad (3-14)$$

and

$$\mathcal{L}^\dagger \alpha_{R\tilde{L}} \mathcal{L} = \mathcal{L}^\dagger \mathcal{R}_\alpha \tilde{\mathcal{L}}_\alpha^T \mathcal{L}. \quad (3-15)$$

The two types of four vectors have slightly different transformations

$$\alpha'_{L\tilde{R}} = e^{\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i} \alpha_{L\tilde{R}} e^{-\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i} \quad (3-16)$$

$$\alpha'_{R\tilde{L}} = e^{\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i} \alpha_{R\tilde{L}} e^{-\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i}. \quad (3-17)$$

Both transform with positive θ under rotations, but the boosts are a different story. The $L\tilde{R}$ four vector transforms with positive ϕ while the $R\tilde{L}$ four vector transforms with negative ϕ . The $R\tilde{L}$ rank 2 spinor is the parity flipped version of the $L\tilde{R}$ rank 2 spinor with negative spatial components:

$$\begin{aligned} \alpha_{L\tilde{R}} &= \alpha_0\sigma_0 + \alpha_1\sigma_1 + \alpha_2\sigma_2 + \alpha_3\sigma_3 \\ \alpha_{L\tilde{R}} &= \alpha_\mu \sigma^\mu \\ \alpha_{L\tilde{R}} &= \begin{pmatrix} \alpha_0 + \alpha_3 & \alpha_1 - i\alpha_2 \\ \alpha_1 + i\alpha_2 & \alpha_0 - \alpha_3 \end{pmatrix} \end{aligned} \quad (3-18)$$

versus

$$\begin{aligned} \alpha_{R\tilde{L}} &= \alpha_0\sigma_0 - \alpha_1\sigma_1 - \alpha_2\sigma_2 - \alpha_3\sigma_3 \\ \alpha_{R\tilde{L}} &= \alpha_\mu \bar{\sigma}^\mu \\ \alpha_{R\tilde{L}} &= \begin{pmatrix} \alpha_0 - \alpha_3 & \alpha_1 + i\alpha_2 \\ \alpha_1 - i\alpha_2 & \alpha_0 + \alpha_3 \end{pmatrix}. \end{aligned} \quad (3-19)$$

The negative ϕ guarantees that the negative spatial components for a R \tilde{L} rank 2 spinor add up with time the correct way in a boost.

The action for the spin $\frac{1}{2}$ particle is built from the lowest order invariant combinations of the spinors themselves and the lowest order invariant term coupling ∂_μ to the spinors. These ingredients are $\mathcal{L}^\dagger \mathcal{R}$, $\mathcal{R}^\dagger \mathcal{L}$, $\mathcal{R}^\dagger \sigma^\mu \partial_\mu \mathcal{R}$, and $\mathcal{L}^\dagger \bar{\sigma}^\mu \partial_\mu \mathcal{L}$. In the scalar particle case, m coupled the field to itself as potential energy with the same power as ∂_μ ($m^2 \Phi^2 \leftrightarrow \partial_\mu \Phi \partial^\mu \Phi$)⁴. Following that example and throwing an i into the derivative terms⁵ to make them Hermitian gives the following Lagrangian,

$$S = \int d^4x (\mathcal{R}^\dagger \sigma^\mu i\partial_\mu \mathcal{R} + i\mathcal{L}^\dagger \bar{\sigma}^\mu i\partial_\mu \mathcal{L} - m\mathcal{L}^\dagger \mathcal{R} - m\mathcal{R}^\dagger \mathcal{L}). \quad (3-20)$$

The Euler-Lagrange equations reveal how the fields change over time,

$$\begin{aligned} \sigma^\mu i\partial_\mu \mathcal{R} &= m\mathcal{L} \\ \bar{\sigma}^\mu i\partial_\mu \mathcal{L} &= m\mathcal{R}. \\ \sigma^\mu i\partial_\mu \mathcal{R}^\dagger &= -m\mathcal{L}^\dagger \\ \bar{\sigma}^\mu i\partial_\mu \mathcal{L}^\dagger &= -m\mathcal{R}^\dagger. \end{aligned} \quad (3-21)$$

Notice that the mass term couples the left and right-chiral spinors. At rest, P=0 and the equations reduce to $i\partial_0 \mathcal{R} = m\mathcal{L}$ and $i\partial_0 \mathcal{L} = m\mathcal{R}$ showing that $\partial_0^2 \mathcal{R} = -m^2 \mathcal{R}$ and $\partial_0^2 \mathcal{L} = -m^2 \mathcal{L}$. The mass is actually a frequency determining how quickly a particle oscillates between its right-chiral and left-chiral states. Another interesting point is that swapping the sign of m here swaps the roles of the fields and the conjugate fields.

⁴ E, P, and m must have the same power to get $E^2 = m^2 + \vec{P}^2$ right, and $E \sim \partial_0$, $P_i \sim \partial_i$ so m must have the same power as ∂_μ .

⁵ The i gives the appropriate factor for the definition of P_μ , $P_\mu = i\partial_\mu$.

Plugging $\mathcal{R} = \frac{1}{m}\bar{\sigma}^\nu i\partial_\nu \mathcal{L}$ into $\sigma^\mu i\partial_\mu \mathcal{R} = m\mathcal{L}$ shows that the constants chosen for the terms in the Lagrangian provide the correct dispersion relation,

$$\begin{aligned} \sigma^\mu i\partial_\mu \frac{1}{m}\bar{\sigma}^\nu i\partial_\nu \mathcal{L} &= m\mathcal{L} \\ \rightarrow \sigma^\mu \bar{\sigma}^\nu P_\mu P_\nu &= m^2 \\ \rightarrow \eta_\nu^\mu P_\mu P_\nu &= m^2 \\ \rightarrow P^\mu P_\mu &= m^2 \\ \rightarrow E^2 - \vec{P}^2 &= m^2 \end{aligned} \tag{3-22}$$

The action for the spin $\frac{1}{2}$ particle can be rewritten into its more compact, canonical form,

$$S = \int d^4x \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi, \tag{3-23}$$

after defining

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \quad \psi = \begin{pmatrix} \mathcal{L} \\ \mathcal{R} \end{pmatrix}, \text{ and } \bar{\psi} = \psi^\dagger \gamma_0. \tag{3-24}$$

The spin $\frac{1}{2}$ Lagrangian is called the Dirac Lagrangian and the resulting equations of motion are the Dirac equation.

3.4 The Lagrangian For A Free Spin 1 Particle

Last but not least is the Lagrangian for a spin 1 force carrying particle. The main ingredients are the A_μ and ∂_μ four vectors. The lowest order possible invariants are $\partial^\mu A^\nu \partial_\mu A_\nu$, $\partial^\mu A^\nu \partial_\nu A_\mu$, $A^\mu A_\mu$, and $\partial^\mu A_\mu$. However the last term is a total derivative and won't be included. The resulting action and equations of motion are

$$S = \int d^4x (c_0 i\partial^\mu A^\nu i\partial_\mu A_\nu + c_1 i\partial^\mu A^\nu i\partial_\nu A_\mu + c_2 A^\mu A_\mu). \tag{3-25}$$

$$c_2 A^\nu = -\partial_\mu (c_0 \partial^\mu A^\nu + c_1 \partial^\nu A^\mu). \tag{3-26}$$

This is the same form as the equation for the four vector potential in electromagnetism, which has $c_0 = \frac{1}{2}$ and $c_1 = -\frac{1}{2}$. The remaining term, c_2 looks like a mass term, so c_2 is set to

$\frac{1}{2}m^2$ ⁶. The action and equations of motion with the appropriate constants become,

$$S = \int d^4x \left(\frac{-1}{2} \partial^\mu A^\nu \partial_\mu A_\nu + \frac{1}{2} \partial^\mu A^\nu \partial_\nu A_\mu + \frac{m^2}{2} A^\mu A_\mu \right) \quad (3-27)$$

$$- m^2 A^\nu = \partial_\mu (\partial^\mu A^\nu - \partial^\nu A^\mu). \quad (3-28)$$

Taking another derivative reveals that the equations reduce further and take on the correct dispersion relation, confirming the choice of constants,

$$\begin{aligned} - m^2 \partial_\nu A^\nu &= \partial_\nu \partial_\mu \partial^\mu A^\nu - \partial_\nu \partial_\mu \partial^\nu A^\mu = 0 \\ \rightarrow \partial_\nu A^\nu &= 0 \\ \rightarrow -m^2 A^\nu &= \partial_\mu \partial^\mu A^\nu \\ \rightarrow m^2 A^\nu &= i \partial_\mu i \partial^\mu A^\nu \\ \rightarrow m^2 &= E^2 - \vec{P}^2. \end{aligned} \quad (3-29)$$

The action for the spin 1 particle is called the Proca action. It's normally written in an equivalent form in terms of the tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$

$$S = \int d^4x \left(\frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m^2}{2} A^\mu A_\mu \right). \quad (3-30)$$

3.5 Interactions

In order to describe theories where fermions and bosons interact, the spin 1 and spin 1/2 are coupled. The simplest Lorentz invariant term that couples a vector to the Dirac spinors is $\bar{\psi} \gamma^\mu A_\mu \psi$, providing an interacting theory that looks like

$$\begin{aligned} S &= \int d^4x \frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m_A^2}{2} A^\mu A_\mu + \bar{\psi} (i \gamma^\mu \partial_\mu - m) \psi + q \bar{\psi} \gamma^\mu A_\mu \psi \\ &= \int d^4x \frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m_A^2}{2} A^\mu A_\mu + \bar{\psi} [i \gamma^\mu (\partial_\mu - iq A_\mu) - m] \psi. \end{aligned} \quad (3-31)$$

⁶ The order of the mass should match the order of the derivatives.

To describe electromagnetism, the photon field should be massless and the theory should be gauge invariant. Setting m_A to zero and checking whether $A_\mu \rightarrow A_\mu(x) + \partial_\mu\alpha(x)$ is a symmetry, provides

$$S = \int d^4x \frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} [i\gamma^\mu (\partial_\mu - iqA_\mu - iq\partial_\mu\alpha) - m] \psi, \quad (3-32)$$

which isn't gauge invariant. In effect, the gauge transformation on A_μ has shifted the derivative, $\partial_\mu \rightarrow \partial_\mu - iq\partial_\mu\alpha$. In order to retain gauge invariance, a simultaneous transformation that sends $\partial_\mu \rightarrow \partial_\mu + iq\partial_\mu\alpha$ is needed to cancel out the extra term. The correct transformation is $\bar{\psi}\partial_\mu\psi \rightarrow \bar{\psi}e^{-iq\alpha}\partial_\mu e^{iq\alpha}\psi$. So, if

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\alpha(x) \quad (3-33)$$

and

$$\psi(x) \rightarrow e^{iq\alpha(x)}\psi(x) \quad (3-34)$$

the Lagrangian is gauge invariant. The second equation reveals that the field is invariant under local transformations of the phase.

Requiring gauge invariance requires the fermion field to maintain U(1) invariance. By running the logic in reverse and demanding more complicated unitary invariances on the fermion field(s), Lagrangians may be produced that describe new forces. In the case of electromagnetism above, demanding that the fermion field is invariant under a local U(1) transformation, sends $\bar{\psi}\partial_\mu\psi \rightarrow \bar{\psi}e^{-iq\alpha}\partial_\mu e^{iq\alpha}\psi$ effectively shifting the derivative, $\partial_\mu \rightarrow \partial_\mu + iq\partial_\mu\alpha$. Now A_μ needs to cancel the shift requiring

$$\partial_\mu + iq\partial_\mu\alpha + c(A_\mu + a_\mu) = \partial_\mu + cA_\mu. \quad (3-35)$$

Therefore $c=-iq$ and $a_\mu = \partial_\mu\alpha$. Furthermore, $F^{\mu\nu}F_{\mu\nu}$ must remain invariant under $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\alpha(x)$. These conditions plus Lorentz invariance force the Lagrangian to be of the

form,

$$S = \int d^4x \frac{-1}{4} (\partial_\mu A_\nu - \partial_\nu A_\mu) (\partial^\mu A^\nu - \partial^\nu A^\mu) + \bar{\psi} [i\gamma^\mu (\partial_\mu - iqA_\mu) - m] \psi, \quad (3-36)$$

which correctly describes Quantum Electrodynamics (QED) [25].

The QED U(1) symmetry has a single generator and, as a consequence, bypasses any commutation issues. The SU(2) transformation $\psi \rightarrow U\psi$ with $U = e^{\frac{i}{2}q\alpha_i(x)\sigma_i}$ is a bit more complicated. For any unitary transformation, if

$$\bar{\psi} U^\dagger i\gamma^\mu (\partial_\mu + cA'_\mu) U\psi = \bar{\psi} i\gamma^\mu (\partial_\mu + cA_\mu) \psi \quad (3-37)$$

is true when $A_\mu \rightarrow A'_\mu$, then by definition gauge invariance is retained. This is true when $(\partial_\mu + cA'_\mu) U\psi = U(\partial_\mu + cA_\mu)\psi$. Breaking down the transformation on A_μ into $A_\mu \rightarrow UA_\mu U^\dagger + \Delta A_\mu$ helps simplify the conditions for gauge invariance,

$$\begin{aligned} (\partial_\mu + cA'_\mu) U\psi &= \partial_\mu(U\psi) + cA'_\mu U\psi \\ &= (\partial_\mu U)\psi + U(\partial_\mu \psi) + cA'_\mu U\psi \\ &= (\partial_\mu U)\psi + U(\partial_\mu \psi) + cUA_\mu U^\dagger U\psi + c\Delta A_\mu U\psi \\ &= U(\partial_\mu + cA_\mu)\psi + (\partial_\mu U)\psi + c\Delta A_\mu U\psi. \end{aligned} \quad (3-38)$$

To get $U(\partial_\mu + cA_\mu)\psi$ and retain invariance, $(\partial_\mu U)\psi + c\Delta A_\mu U\psi$ must be zero. This requires

$$c\Delta A_\mu U\psi = -(\partial_\mu U)\psi = -(\partial_\mu U)U^\dagger U\psi = U(\partial_\mu U^\dagger)U\psi. \quad (3-39)$$

The last line uses the fact that for a unitary matrix U , $UU^\dagger = 1$ and $\partial_\mu(UU^\dagger) = 0$. Therefore, to preserve gauge invariance under the unitary transformation $\psi \rightarrow U\psi$, cA_μ must transform as

$$cA_\mu(x) \rightarrow cUA_\mu(x)U^\dagger + U\partial_\mu U^\dagger. \quad (3-40)$$

For invariance under SU(2) the derivative can be calculated explicitly

$$\begin{aligned} cA_\mu(x) &\rightarrow cUA_\mu(x)U^\dagger + c\Delta A_\mu \\ &= cUA_\mu(x)U^\dagger - iq\frac{\sigma_i}{2}\partial_\mu\alpha_i(x). \end{aligned} \quad (3-41)$$

Choosing $c=-iq$ and $\Delta A_\mu = \frac{\sigma_i}{2} \partial_\mu \alpha_i(x)$ determines the coupling constant and the transformation for the SU(2) invariant Lagrangian.

A_μ is a matrix which may be expanded in terms of the SU(2) generators, $A_\mu = A_\mu^1 \frac{\sigma_1}{2} + A_\mu^2 \frac{\sigma_2}{2} + A_\mu^3 \frac{\sigma_3}{2}$. Note that U(1) only requires a single particle while SU(2) requires three.

For the SU(N) transformations, there will be a force carrying particle for each generator, and the field may be expanded using the generators as a basis, $A_\mu = T^c A_\mu^c$. In this case, the transformation for each component of A_μ may be written explicitly,

$$A_\mu^c \rightarrow A_\mu^c - f^{abc} \alpha^a A_\mu^b + \partial_\mu \alpha^c, \quad (3-42)$$

where $[T^a, T^b] = if^{abc}T^c$ defines the Lie algebra of the group. For SU(2) in particular, $A_\mu = \frac{\sigma_c}{2} A^c$ and $f^{abc} = \epsilon^{abc}$. This defines half of the SU(2) Lagrangian, but $F_{\mu\nu}$ still needs to be defined so that it remains invariant under equation 3-42. Defining

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - iq[A_\mu, A_\nu] \quad (3-43)$$

makes $F_{\mu\nu}$ covariant, $F_{\mu\nu} \rightarrow UF_{\mu\nu}U^\dagger$, [25] and to make it invariant, the trace is taken. So finally, the full gauge invariant Lagrangian with massless fermions is given by

$$S = \int d^4x \mathcal{N} \text{tr}(F_{\mu\nu} F^{\mu\nu}) + \bar{\psi} i \gamma^\mu D_\mu \psi, \quad (3-44)$$

where $D_\mu = \partial_\mu - iqA_\mu$ is the covariant derivative. The Lagrangian holds for the U(1) and SU(N) gauge transformations. U(1) provides electromagnetism, adding on SU(2) provides the electroweak interactions, and adding SU(3) provides the strong force. All three together, describe the Standard Model. Writing out the trace in terms of the generators, the Lagrangian is given by,

$$S = \int d^4x \frac{-1}{2} \text{tr}(F_{\mu\nu} F^{\mu\nu}) + \bar{\psi} i \gamma^\mu D_\mu \psi = \int d^4x \frac{-1}{4} F_{\mu\nu}^a F_a^{\mu\nu} + \bar{\psi} [i \gamma^\mu (\partial_\mu - iqT^i A_\mu^i)] \psi. \quad (3-45)$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + qf^{abc} A_\mu^b A_\nu^c. \quad (3-46)$$

\mathcal{N} is set to $-\frac{1}{2}$ in order to get the appropriate $-\frac{1}{4}$ in front of the $F_{\mu\nu}^a$ tensors⁷.

The SU(2) invariant Lagrangian of equation 3–45 provides the weak interaction.

Unfortunately, the theory has massless fermions and massless weak force particles, while in real life these particles have mass. Adding mass terms directly ruins the SU(2) invariance, so another mechanism is needed, and this is where the Higgs mechanism comes into play.

3.6 The Higgs Mechanism

The W and Z bosons observed in nature are massive, but directly adding a mass term for a force carrying particle ruins the gauge symmetry,

$$\frac{1}{2}m_A^2 A_\mu A^\mu \rightarrow \frac{1}{2}m_A^2 (UA_\mu U^\dagger + \Delta A_\mu)(UA^\mu U^\dagger + \Delta A^\mu). \quad (3-47)$$

Both the UAU^\dagger and the ΔA are a problem. Meanwhile for the matter particles, the gauge transformation acts on a column of fermions which leaves the $\bar{\psi}m\psi$ term invariant,

$$\begin{pmatrix} \bar{\psi}_1 \\ \dots \\ \bar{\psi}_n \end{pmatrix} m \begin{pmatrix} \psi_1 & \dots & \psi_n \end{pmatrix} \rightarrow \begin{pmatrix} \bar{\psi}_1 \\ \dots \\ \bar{\psi}_n \end{pmatrix} U^\dagger m U \begin{pmatrix} \psi_1 & \dots & \psi_n \end{pmatrix}, \quad (3-48)$$

but this term restricts the fermions to the same mass. In order to describe the massive fermions and the massive W and Z particles seen in nature the Higgs mechanism is needed. Relativistically, mass is energy, so the idea is to produce $m_A^2 A_\mu A^\mu$ and $\bar{\psi}m\psi$ terms via some interaction energy involving a new field. Because mass is a scalar, the interactions require a scalar field, and because the mass is derived from a nonzero interaction energy, the groundstate of the scalar field should be nonzero. For all of these reasons, the Higgs mechanism adds a scalar field with a ϕ^4 potential term to the Lagrangian,

$$\mathcal{L}_\phi^{\text{toy}} = (\partial^\mu \phi - iqA^\mu \phi)^\dagger (\partial_\mu \phi - iqA_\mu \phi) + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2. \quad (3-49)$$

⁷ The calculation uses the fact that the SU(N) the generator matrices form an orthogonal basis, $\text{tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$.

As ϕ goes into the groundstate, $\phi \rightarrow \phi_0$, the force carrying field(s) acquire mass through the $A^2\phi_0^2$ terms. Similarly, interaction terms coupling the left and right spinors provide the fermions with mass as $\phi \rightarrow \phi_0$

$$\mathcal{L}_I^{\text{toy}} = -\beta(L^\dagger \phi R + R^\dagger \phi L). \quad (3-50)$$

The toy examples of equations 3–49 and 3–50 cover the basic ideas behind the Higgs mechanism, but correctly describing the electroweak interaction requires a more complex and intricate theory, the $U(1) \times SU(2)$ Weinberg-Salam Lagrangian. The Weinberg-Salam Lagrangian includes a $U(1)$ gauge field and an $SU(2)$ gauge field. In the theory, the $U(1)$ field and the third component of the $SU(2)$ field mix up, with one orthogonal piece providing the massive Z boson and the other the massless photon. The remaining first and second components of the $SU(2)$ field mix up to provide the massive W^+ and W^- bosons.

The Weinberg-Salam Lagrangian for the electron (e) and electron neutrino (ν_e) of equation 3–51 is written in terms of left and right handed Dirac spinors, e_L , e_R , and ν_e . Left and right handed Dirac spinors for a particle are defined in terms of the left and right handed spinors (s) for the particle as follows, $\psi_L = \begin{pmatrix} s_L \\ 0 \end{pmatrix}$ and $\psi_R = \begin{pmatrix} 0 \\ s_R \end{pmatrix}$. In nature, the weak force treats left and right handed particles differently. The W^+ and W^- particles interact only with left handed particles and interact with those of the same generation symmetrically. To respect this symmetry, the Lagrangian is written to respect the interchange of the left handed, same generation particles. The left swapping symmetry is an $SU(2)$ transformation acting on the column vector of left handed Dirac spinors, $L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}$. The remaining right handed particle is denoted by $R = e_R$. In addition, only left handed neutrinos have been observed in nature, so ν_e

is left handed. The electroweak Lagrangian is then,

$$\begin{aligned}
\mathcal{L} = & (\partial^\mu \phi - iq_{w\phi} \frac{\sigma_i}{2} W_i^\mu \phi - iq_{b\phi} B^\mu \phi)^\dagger (\partial_\mu \phi - iq_{w\phi} \frac{\sigma_i}{2} W_\mu^i \phi - iq_{b\phi} B^\mu \phi) \\
& + \bar{L} i \gamma^\mu (\partial_\mu - iq_{wl} \frac{\sigma_i}{2} W_i^\mu - iq_{bl} B_\mu) L + \bar{R} i \gamma^\mu (\partial_\mu - iq_{br} B_\mu) R \\
& - \beta (\bar{\nu}_e \phi_+ e_R + \bar{e}_R \phi_+^* \nu_e + \bar{e}_L \phi_- e_R + \bar{e}_R \phi_-^* e_L) - \frac{1}{4} G_i^{\mu\nu} G_{\mu\nu}^i - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \\
& + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2.
\end{aligned} \tag{3-51}$$

The SU(2) field, W_μ , in its simplest nontrivial representation is a complex 2x2 matrix that operates on 2x1 complex column vectors. Therefore ϕ is written

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_- \end{pmatrix}. \tag{3-52}$$

To isolate the mass terms, the scalar field is expanded about its minimum, $|\phi_{\min}| = \phi_0 = \sqrt{\frac{m_h^2}{\lambda}}$. The minimum is degenerate, only requiring that $|\phi_+|^2 + |\phi_-|^2 = \phi_0^2$. The particular solution $\phi = \begin{pmatrix} 0 \\ \phi_0 + \frac{1}{\sqrt{2}} h(x) \end{pmatrix}$ fixes the fermion mass terms such that the electron acquires mass in the correct way. The choice also eliminates the strange mass-like coupling between the electron and neutrino,

$$\beta(\bar{L} \phi R + \bar{R} \phi^\dagger L) = \beta(\bar{\nu}_e \phi_+ e_R + \bar{e}_R \phi_+^* \nu_e + \bar{e}_L \phi_- e_R + \bar{e}_R \phi_-^* e_L) \rightarrow \beta(\bar{e}_L \phi_0 e_R + \bar{e}_R \phi_0 e_L). \tag{3-53}$$

Fixing $\phi^+ = 0$ and $\text{Im}\phi^- = 0$ comes at a cost, breaking the SU(2) symmetry of the Lagrangian. The coupling is no longer in the invariant form $\bar{L} \phi R + \bar{R} \phi^\dagger L \rightarrow \bar{L} U^\dagger U \phi R + \bar{R} \phi^\dagger U^\dagger U L$.

As in the toy example, the $(D^\mu\phi)^\dagger(D_\mu\phi)$ term bestows mass onto the force carriers,

$$\begin{aligned}
D_\mu\phi &= (\partial_\mu - iq_{w\phi}\frac{\sigma_i}{2}W_\mu^i - iq_{b\phi}B_\mu)\phi \\
&= \left[\begin{pmatrix} \partial_\mu - iq_{b\phi}B_\mu & 0 \\ 0 & \partial_\mu - iq_{b\phi}B_\mu \end{pmatrix} - \frac{i}{2}q_{w\phi} \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} \right] \begin{pmatrix} 0 \\ \phi_0 + \frac{1}{\sqrt{2}}h \end{pmatrix} \\
&= -\frac{i}{2} \begin{pmatrix} q_{w\phi}\phi_0(W_\mu^1 - iW_\mu^2) + \frac{1}{\sqrt{2}}q_{w\phi}h(W_\mu^1 - iW_\mu^2) \\ i\sqrt{2}\partial_\mu h + \phi_0(2q_{b\phi}B_\mu - q_{w\phi}W_\mu^3) + \frac{1}{\sqrt{2}}h(2q_{b\phi}B_\mu - q_{w\phi}W_\mu^3) \end{pmatrix}.
\end{aligned} \tag{3-54}$$

The ϕ_0^2 terms determine the masses,

$$(D^\mu\phi)^\dagger(D_\mu\phi) = q_{w\phi}^2\frac{\phi_0^2}{4}(W_\mu^1)^2 + q_{w\phi}^2\frac{\phi_0^2}{4}(W_\mu^2)^2 + \frac{\phi_0^2}{4}(q_{w\phi}W_\mu^3 - 2q_{b\phi}B_\mu)^2 + \text{other terms.} \tag{3-55}$$

The orthogonal term $q_{w\phi}W_\mu^3 + 2q_{b\phi}B_\mu$ is missing from the covariant derivative and remains massless [5], providing the photon field. This leaves

$$m_w = \frac{q_{w\phi}\phi_0}{\sqrt{2}}, \quad m_z = \frac{m_w}{q_{w\phi}}, \quad \text{and} \quad m_\gamma = 0. \tag{3-56}$$

The photon, A_μ , is a linear combination of W_μ^3 and B_μ , which implies that the U(1) symmetry corresponding to electromagnetism is also a linear combination,

$$\begin{aligned}
U_A &= e^{i(Q_{wi}T^3+Q_{bi})\alpha(x)}, \\
W_\mu^3 &\rightarrow W_\mu^3 + \frac{1}{g_w}\partial_\mu\alpha, \\
B_\mu &\rightarrow B_\mu + \frac{1}{g_b}\partial_\mu\alpha.
\end{aligned} \tag{3-57}$$

T^3 is the third SU(2) generator for the given representation, and the Qs are the normalized charges defined by,

$$q_{wi} = Q_{wi}g_w \quad \text{and} \quad q_{bi} = Q_{bi}g_b. \tag{3-58}$$

The U(1) gauge symmetry leads to conservation of electromagnetic charge, $Q = Q_wI_3 + Q_b$, implying that the electromagnetic charge for a given particle is $Q_i = Q_{wi}I_{3i} + Q_{bi}$. I_{3i} represents

the eigenvalue of the T^3 generator denoting the particle eigenstate. For example, the operator $T^3 = \frac{\sigma^3}{2}$ acting on L has two eigenstates, ν_e and e_L , corresponding to eigenvalues $I_{3\nu_e} = \frac{1}{2}$ and $I_{3e_L} = -\frac{1}{2}$ respectively. Similarly, the eigenstates $\phi^+ = 0$ and $\phi^- = \phi_0 + h$ correspond to eigenvalues $I_{3\phi^+} = \frac{1}{2}$ and $I_{3\phi^-} = -\frac{1}{2}$.

The symmetry of the W particles towards the left handed particles implies that the left handed particles have the same charge, $Q_{wi} = 1$. The lack of interaction between the W particles and the right handed particles implies that the right handed particles have $Q_{wi} = 0$. The Higgs boson is assumed to interact with the W particles the same way as the left handed ones with a $Q_{wi} = 1$. The electromagnetic charge Q_i and the isospin Q_{wi} values fix the remaining electroweak Q_{bi} values,

$$\begin{aligned} Q_{e_L} &= Q_{we_L} I_{3e_L} + Q_{be_L} = \frac{-1}{2} + Q_{be_L} \\ Q_{e_R} &= Q_{we_R} I_{3e_R} + Q_{be_R} = 0 + Q_{be_R} \\ Q_{\nu_e} &= Q_{w\nu_e} I_{3\nu_e} + Q_{b\nu_e} = \frac{1}{2} + Q_{b\nu_e} \\ Q_{\phi^-} &= Q_{w\phi^-} I_{3\phi^-} + Q_{b\phi^-} = \frac{-1}{2} + Q_{b\phi^-}, \end{aligned} \tag{3-59}$$

providing,

$$Q_{be_L} = \frac{-1}{2}, Q_{be_R} = -1, Q_{b\nu_e} = \frac{-1}{2}, \text{ and } Q_{b\phi^-} = \frac{+1}{2}. \tag{3-60}$$

The electroweak Lagrangian reduces to,

$$\begin{aligned} \mathcal{L} &= (\partial^\mu \phi - ig_w \frac{\sigma_i}{2} W_i^\mu \phi - \frac{i}{2} g_b B^\mu \phi)^\dagger (\partial_\mu \phi - ig_w \frac{\sigma_i}{2} W_i^\mu \phi - \frac{i}{2} g_b B^\mu \phi) \\ &\quad + \bar{L} i \gamma^\mu (\partial_\mu - ig_w \frac{\sigma_i}{2} W_i^\mu + \frac{i}{2} g_b B_\mu) L + \bar{R} i \gamma^\mu (\partial_\mu + ig_b B_\mu) R \\ &\quad - \beta (\bar{L} \phi R + \bar{R} \phi^\dagger L) - \frac{1}{4} G_i^{\mu\nu} G_{i\mu\nu}^i - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2, \end{aligned} \tag{3-61}$$

with $\phi = \begin{pmatrix} 0 \\ \phi_0 + \frac{1}{\sqrt{2}} h(x) \end{pmatrix}$ and $L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}$. In equation 3-61, the W^+ particle is $W_\mu^1 + iW_\mu^2$, and the W^- particle is $W_\mu^1 - iW_\mu^2$. The photon is $g_w W_\mu^3 + g_b B_\mu$, the Z boson is $g_w W_\mu^3 - g_b B_\mu$, and the Higgs boson is $h(x)$. Adding the next two generations of leptons, the three generations

of quarks, and the SU(3) interactions to the electroweak Lagrangian defines the entire Standard Model.

3.7 The Standard Model Higgs Boson And The LHC

The SM Higgs boson interacts with the massive particles of the SM and even with the massless gluons and photons through second order processes. As such, it can be produced by colliding certain combinations of these particles, and it can decay into them as well. At a particle collider like the LHC, the number of particles expected for a certain process is given by the cross section times the integrated luminosity, $N = \sigma_i * L$. The cross section is proportional to the probability for a production process and consequently, describes how likely a collision attempt is to produce some particle(s) of interest. The luminosity roughly describes the density and the frequency of the incoming particles. Some of the Higgs cross sections for 14 TeV proton-proton collisions are shown in Figure 3-2.

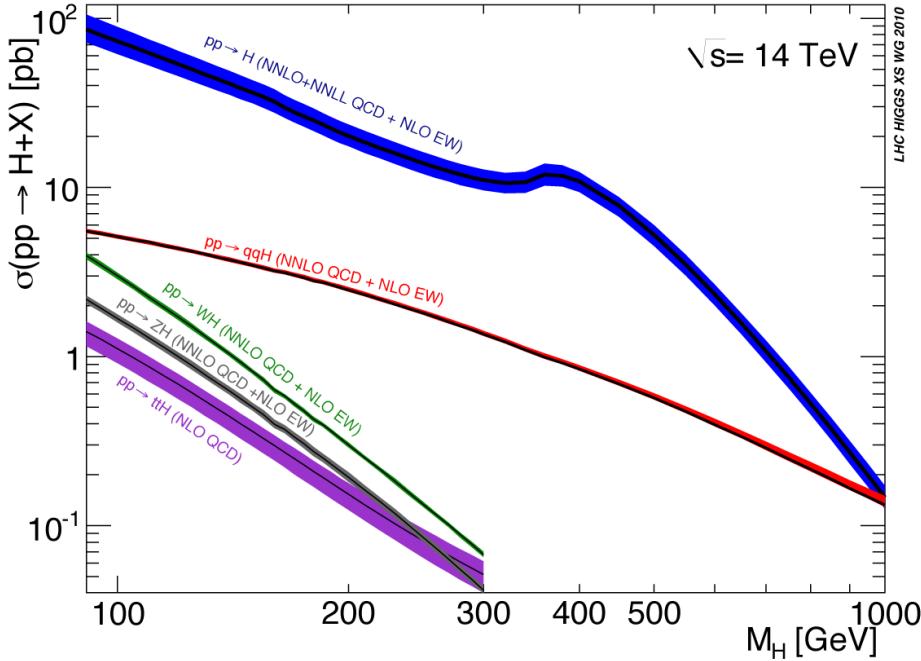


Figure 3-2. The highest production mode cross sections for the SM Higgs boson at 14 TeV [27].

The Higgs cross sections are functions of the mass of the Higgs boson as well as the energy of the collisions. For a given collision energy, as in Figure 3-2, the cross sections

decrease as the Higgs mass increases. When a larger portion of the collision energy was used to create the mass of the particle, there is less energy to distribute among the kinematic degrees of freedom and therefore fewer possibilities for distribution. On the other hand, for a specific Higgs mass, the cross section grows with collision energy at the LHC. This contrasts with cross sections involving collisions of fundamental particles, e.g. electron antielectron collisions, due to the fact that the LHC collides protons together.

Protons behave like a quantum superposition of an infinite number of quark-antiquarks, an infinite number of gluons, and the usual uud. As a consequence, the total momentum of a proton in a collision is divided up amongst these constituents called partons. This experimentally verified phenomena is modeled by the parton distribution function, which describes the number of partons with a given fraction of the total momentum. In general, there are many partons with very little of the momentum, and this behavior implies that the cross section should increase with increasing proton momentum. The minimum energy required to create a particular particle is a constant, and at larger proton momentum, this constant is a smaller fraction of the total proton momentum. With more partons at this smaller fraction, there are effectively more partons colliding with the necessary energy. This effective increase in the density of energetic partons results in a growth of the cross section with collision energy.

If a SM Higgs boson is produced, it's predicted to decay in about 10^{-22} seconds⁸, which means that the particle itself can't be directly detected at the LHC, only the decay products can. The SM decay probabilities for the different products are listed in Figure 3-3. These probabilities are determined by the coupling, which for the Higgs boson is the mass of the particle. With a stronger coupling to more massive particles, the decays to the more massive particles are more probable.

⁸ Assuming a 125 GeV SM Higgs boson

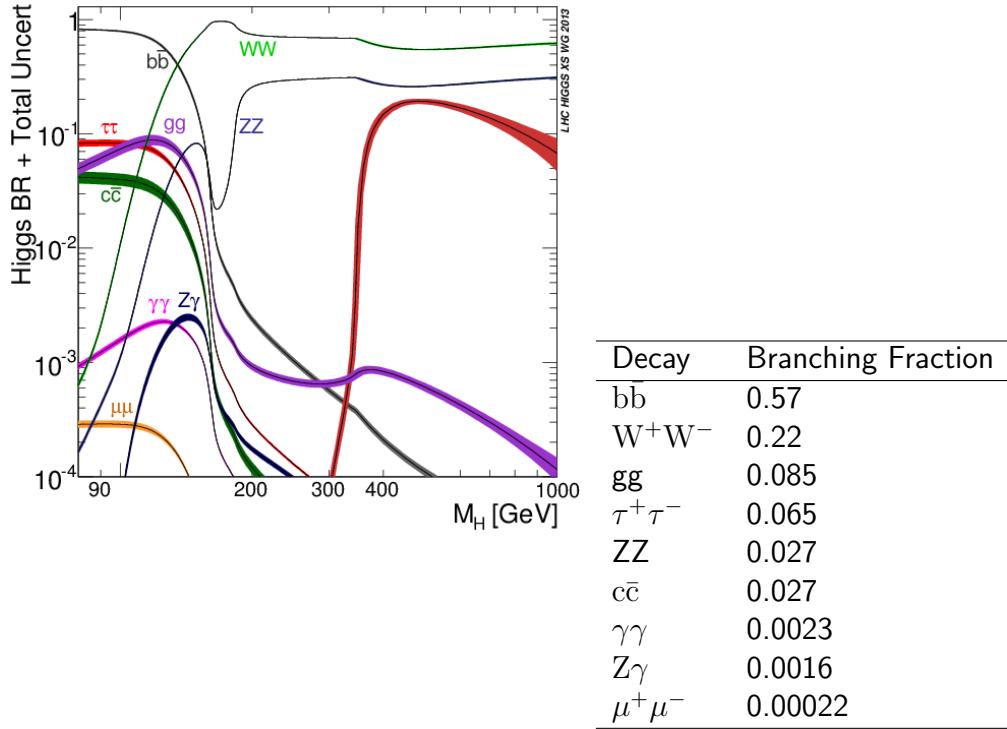


Figure 3-3. The graphic on the top left presents the SM Higgs branching fractions as functions of mass while the table on the bottom right displays the branching fractions for a 125 GeV SM Higgs boson [27].

The muon has the lowest mass⁹ of the particles in Figure 3-3 and consequently $H \rightarrow \mu^+\mu^-$ has the lowest branching fraction in the set.¹⁰ The gluons and photons are massless and do not couple to the Higgs boson at leading order, but through second order processes. Gluons interact with the Higgs boson through a loop of top quarks, as seen in Figure 3-4a. The extremely heavy mass of the top quark, about 173 GeV, balances the fact that the loop production is a higher order mechanism. The photons interact with the Higgs boson through either a loop of W bosons or a loop of top quarks. Figure 3-4 shows the highest probability production mechanisms at the LHC. At $M_h = 125$ GeV, $\sqrt{s} = 13$ TeV, the

⁹ excluding the photon and gluon

¹⁰ The Higgs boson also couples to the electron and the first generation quarks but the masses are so light that CMS does not expect to see the SM Higgs boson in those modes.

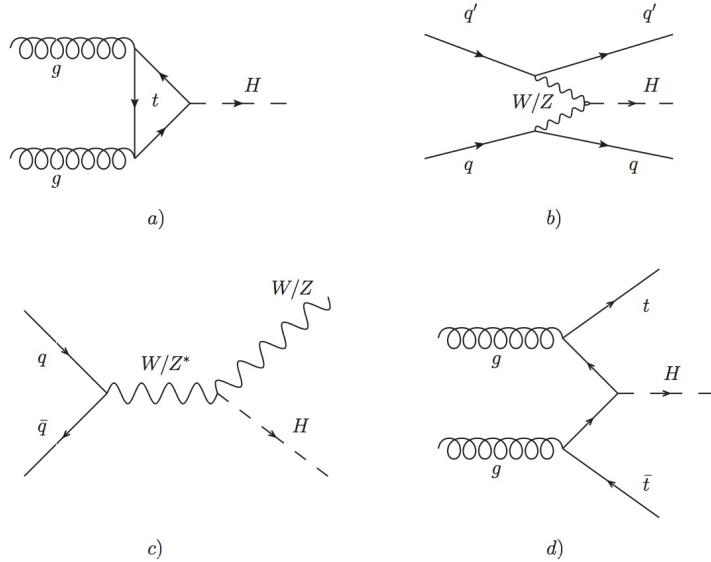


Figure 3-4. The SM production modes with the highest cross sections. a) Gluon Gluon Fusion (GGF/ggH) b) Vector Boson Fusion (VBF/qqH) c) Associated Production with a Vector Boson (VH) d) $t\bar{t}H$.

GGF channel comprises 87% of the total Higgs production cross section, VBF 7%, VH 4%, and $t\bar{t}H$ 1% [27]. Besides $t\bar{t}H$, the process $q + \bar{q} \rightarrow H$ isn't considered due to its low cross section. The low masses of these other quarks suppress the process.

Colliding protons full of quarks and gluons results in many quark-gluon (qg) scattering events like the one in Figure 3-5. Quarks and gluons are detected at CMS as collimated jets of energy deposition and not single particles with well defined tracks. Because of this, the different quarks and gluons are difficult to differentiate from one another, and it's difficult to differentiate the quark/gluon Higgs decays from this large background. The Higgs decay

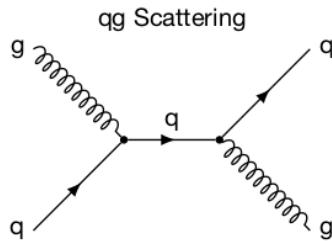


Figure 3-5. The quark-gluon background looks very similar to the GGF production channel when the Higgs boson decays to two jets. Protons are made of quarks and gluons so this process is extremely common in proton colliders like the LHC.

to b-quarks is an exception as the b-quark upon production forms a reasonably long lived hadron, which travels from the initial collision point and then decays. In this way, jets that come from displaced vertices are probably b-jets and these events can be collected for study without collecting the majority of the enormous qg background. The relative clarity over the qg background makes $H \rightarrow b\bar{b}$ an important and viable process that allows scientists at the LHC to study the Higgs coupling to fermions and to third generation quarks in particular. The other viable decays to study are those with isolated lepton or photon final states that distinguish them from the overwhelming jet background.

CHAPTER 4

DISCOVERY, LIMITS, MEASUREMENT, AND SENSITIVITY

This chapter briefly covers the statistical methods needed for the $H \rightarrow \mu^+ \mu^-$ search.

The test of a coin for bias is used as a simple example to explain the concepts of discovery, limits, measurement and sensitivity. These concepts are then extended to a binned counting experiment typical in particle physics. The chapter starts with discovery.

4.1 Discovery

Many scientific endeavors attempt to disprove some null hypothesis in order to claim the observation of a new effect. For example, to test whether a coin is biased, the experimenter assumes the null hypothesis that the coin is unbiased and then performs an experiment, tossing the coin many times. If all of the tosses are heads, then it's very unlikely that the coin is unbiased, and it's therefore reasonable to rule out the null and to declare the discovery of a biased coin.

In order to quantify how rarely an unbiased coin would yield an experiment with N_{heads} , a model for the probability density function (PDF) is needed. The binomial distribution with $x = N_{\text{heads}}$, $\rho = p_{\text{heads}} = 0.5$, and $N = N_{\text{tosses}}$, is the appropriate PDF,

$$p(x, N; \rho) = \frac{N!}{x!(N-x)!} \rho^x (1-\rho)^{N-x}. \quad (4-1)$$

With the PDF, the probability for the null to produce different experiments can be compared in terms of p-values. The p-value assuming the null, $P(x \succ Y | \text{null})$, is the probability to observe something at least as extreme, \succ , as the outcome Y given the null. Declaring the cutoff p-value for the coin flipping experiment sets the minimum threshold of heads, h_{cutoff} , needed for discovery. Any observation of $x = N_{\text{heads}}$ with a p-value rarer than the cutoff, i.e. $P(x \geq h_{\text{cutoff}} | \text{null}) < p_{\text{cutoff}}$, will rule out the null hypothesis.

Traditionally, different fields require different p-values, and in high energy physics 3σ leads to an "observation" and 5σ leads to a "discovery". These correspond to p-values of 0.3% and 0.00006% respectively. As an example, observing 65 heads or greater in 100 tosses occurs just

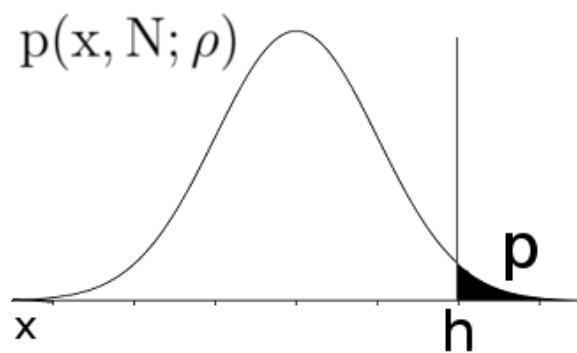


Figure 4-1. The shaded region represents the p-value for an observation of $N_{\text{heads}} = h$. If the p-value for h is lower than the cutoff threshold, then the null hypothesis can be ruled out and a discovery may be declared.

less than 0.3% of the time for an unbiased coin. Therefore, any experiment with $N_{\text{heads}} \geq 65$ would invalidate the null at 3σ and lead to an observation.

4.2 Limits

Besides discovery, setting a limit is also important. Upon tossing a coin 100 times and finding $N_{\text{heads}} = 40$, the experimenter may ask which values of p_{heads} are too high to yield such a low observation. In this case, all values of p_{heads} that predict experiments with 40 heads or fewer at too rare a probability may be ruled out at some confidence. For 95% confidence, $p_{\text{heads}} \geq 0.488$ may be ruled out as $p_{\text{heads}} \geq 0.488$ yields 40 heads or fewer less than 5% of the time, while smaller values for p_{heads} do not. Therefore, observing 40 heads in 100 tosses places an upper limit of $p_{\text{heads}} = 0.488$ on the bias of the coin at 95% confidence.

4.3 Measurement and Uncertainty

Finally, measuring values is also important. In the case of the coin, the experimenter flips the coin 100 times and attempts to measure how biased the coin is. The value stated as the measured value is usually the best fit, and the best fit value is the one that maximizes the (log) likelihood of seeing the data observed. In practice, minimizing the *negative log likelihood* (N_{LL}) is more convenient,

$$-\frac{\partial}{\partial \rho} \ln(p(x, N; \rho)) = 0 \rightarrow \hat{\rho} = \frac{x}{N}. \quad (4-2)$$

When performing many independent experiments the PDFs for each experiment multiply and the negative log likelihood is,

$$-\ln(p) = -\ln \left(\prod_i p_i \right) = -\sum_i \ln(p_i). \quad (4-3)$$

Over many coin flipping experiments, the best fit value for ρ is

$$-\frac{\partial}{\partial \rho} \ln(p) = 0 \rightarrow \hat{\rho} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n N_i}. \quad (4-4)$$

When N_i is the same in each experiment the best fit is just the mean over all experiments,

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{N} = \frac{1}{n} \sum_{i=1}^n \hat{\rho}_i. \quad (4-5)$$

Note that the different experiments all contribute to determine the best fit, and that the results of other experiments constrain the influence of a particular experiment to declare the best fit value.

In order to quantify the uncertainty of the measured value, limiting values of ρ are computed. The upper limit ρ_{hi} and the lower limit ρ_{lo} define the confidence interval $[\rho_{\text{lo}}, \rho_{\text{hi}}]$, which quantifies the uncertainty on $\hat{\rho}$. Conceptually, the interval is a range of values for which the observed data (summarized by $\hat{\rho}$) is not too extreme. The confidence interval stated usually corresponds to 1σ or 68%, constructed from upper and lower limits that exclude 16% of their respective PDFs. The construction is illustrated in Figure 4-2, and the construction guarantees that in many experiments the true value will be contained in the confidence interval 68% of the time.

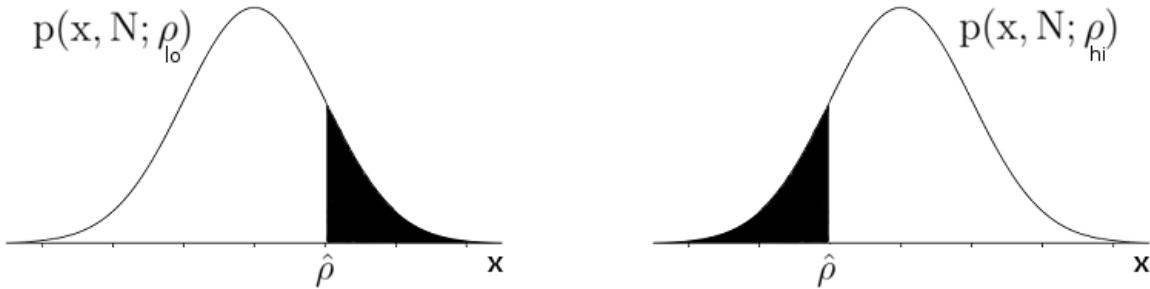


Figure 4-2. The 1σ uncertainty on $\hat{\rho}$ is determined by finding the appropriate ρ_{lo} and ρ_{hi} . ρ_{lo} is the value of ρ just low enough that an observation corresponding to $\hat{\rho}$ or greater occurs only 16% of the time. ρ_{hi} is the value of ρ just high enough that an observation of $\hat{\rho}$ or less occurs only 16% of the time. The shaded region represents 16% of the respective PDF.

Observing 50 heads in 100 tosses leads to $\hat{\rho} = 0.5$ and a corresponding 1σ confidence interval of $[0.44, 0.56]$. The measurement is then reported as $\hat{\rho} = 0.5^{+0.6}_{-0.6}$. An experiment with more data generally has a lower uncertainty on the measured value. For example, observing 450 heads in 900 tosses leads to $\hat{\rho} = 0.5^{+0.2}_{-0.2}$. Note that the uncertainty shrinks with $\sqrt{N_{\text{tosses}}}$, $\frac{0.6}{0.2} = 3 = \sqrt{\frac{900}{100}}$.

4.4 Sensitivity

An analysis is often designed to maximize the chance of discovery by minimizing the expected p-value. The lower the expected p-value given the null, the higher the sensitivity. This section uses a limiting case of the binomial distribution to point out the factors that contribute to a sensitive analysis. Consider an experiment flipping a coin N times,

$$p(x, N; \rho) = \frac{N!}{x!(N-x)!} \rho^x (1-\rho)^{N-x}. \quad (4-6)$$

The experimenter may ask what p-value is expected if the coin has a true value $\rho = \rho_i$. For a coin with ρ_i , the observed N_{heads} is most frequently $N\rho_i$. Therefore, given the null with $\rho = \rho_{\text{null}}$ one expects

$$p(x = N\rho_i, N; \rho = \rho_{\text{null}}) = \frac{N!}{(N\rho_i)!(N - N\rho_i)!} \rho_{\text{null}}^{N\rho_i} (1 - \rho_{\text{null}})^{N - N\rho_i}. \quad (4-7)$$

When $N\rho$ is far enough from zero, the binomial distribution may be approximated by a Gaussian with mean $\mu = N\rho$, and standard deviation $\sigma = \sqrt{N\rho(1-\rho)}$. In this limit, the relationship between the expected p-value and the hypotheses are easy to see, because the p-values for a Gaussian are determined by the number of standard deviations away from the mean, Z . The sensitivity is given by $Z = \frac{(x-\mu)}{\sigma}$, and for a coin with ρ_i , the expected sensitivity is

$$Z = \frac{(x - \mu)}{\sigma} = \frac{N(\rho_i - \rho_{\text{null}})}{\sqrt{N\rho_{\text{null}}(1 - \rho_{\text{null}})}} = \frac{\sqrt{N}(\rho_i - \rho_{\text{null}})}{\sqrt{\rho_{\text{null}}(1 - \rho_{\text{null}})}}. \quad (4-8)$$

The larger Z is, the smaller the p-value, and the more sensitive the experiment. Note that the sensitivity scales as the \sqrt{N} , which shows that collecting more data improves the sensitivity to discovery.

4.5 Constraints And Nuisance Parameters

As alluded to in section 4.3, additional measurements provide constraints on the parameters of a PDF. Consider two experiments tossing the same coin with y heads observed in the first experiment and x in the second. In this case, the likelihoods of the individual

experiments are multiplied to determine the net likelihood,

$$p(x, y; \rho) = p(x; \rho)p(y; \rho), \quad (4-9)$$

Upon minimizing the negative log-likelihood, the observations of the first experiment fight with the second experiment to determine the best fit for ρ .

Sometimes it is convenient to replace the actual likelihood of the previous experiment with a suitable representative. Often, a Gaussian is used along with the best fit and the uncertainty of the previous experiment, $\hat{\rho}_y$ and σ_y . The net likelihood then becomes,

$$p(x; \rho) = p(x; \rho)\mathcal{N}(\hat{\rho}_y; \mu = \rho, \sigma = \sigma_y). \quad (4-10)$$

The Gaussian (\mathcal{N}) fights to keep $\hat{\rho}$ near $\hat{\rho}_y$ with a strength dependent on σ_y . Other constraints like the log-normal or Poisson distributions are also used. Constraint terms are especially useful when they replace a complicated likelihood with many data points.

Imagine a case where the distribution of dirt on a coin affects the net bias ρ . The fraction of heads expected then depends on the bias of the coin itself and the additional bias from the dirt, $\rho = \rho_{\text{coin}} + \rho_{\text{dirt}}$. If a complicated measurement is made to determine the distribution of dirt on the coin, thus estimating ρ_{dirt} , then the bias of the coin itself ρ_{coin} can be extracted.

The net likelihood includes the likelihood for the set of observations y

$$p(x, y; \rho_{\text{coin}}, \rho_{\text{dirt}}) = p(x; \rho_{\text{coin}}, \rho_{\text{dirt}})p(y; \rho_{\text{dirt}}). \quad (4-11)$$

The observations y determine the best fit $\hat{\rho}_{y-\text{dirt}}$ and uncertainty $\sigma_{y-\text{dirt}}$ from that experiment, and the likelihood becomes,

$$p(x, \hat{\rho}_{y-\text{dirt}}; \rho_{\text{coin}}, \rho_{\text{dirt}}) = p(x; \rho_{\text{coin}}, \rho_{\text{dirt}})\mathcal{N}(\hat{\rho}_{y-\text{dirt}}; \mu = \rho_{\text{dirt}}, \sigma = \sigma_{y-\text{dirt}}), \quad (4-12)$$

where a Gaussian constraint replaces the complicated likelihood of the ρ_{dirt} measurement involving the data points y . The best fit for ρ_{coin} can be calculated using the maximum likelihood estimates for ρ_{coin} and ρ_{dirt} given x and $\hat{\rho}_{y-\text{dirt}}$. Limits and uncertainty on ρ_{coin} can

be calculated by *profiling* the likelihood,

$$p(x, \hat{\rho}_{y-\text{dirt}}; \rho_{\text{coin}}, \hat{\rho}_{\text{dirt}}) = p(x; \rho_{\text{coin}}, \hat{\rho}_{\text{dirt}}) \mathcal{N}(\hat{\rho}_{y-\text{dirt}}; \mu = \hat{\rho}_{\text{dirt}}, \sigma = \sigma_{y-\text{dirt}}). \quad (4-13)$$

The double hat designates a maximum likelihood estimate of the parameter for fixed ρ_{coin} . The parameters other than the parameter of interest are called *nuisance parameters*, and profiling sets the nuisance parameters to their best fit estimates for some fixed value of the parameter of interest.

The distribution for $p(x, \hat{\rho}_{y-\text{dirt}}; \rho_{\text{coin}}, \hat{\rho}_{\text{dirt}})$ can be simulated using Monte Carlo methods, where $\hat{\rho}_{y-\text{dirt}}$ is a random variable drawn from $\mathcal{N}(\hat{\rho}_{y-\text{dirt}}; \mu = \hat{\rho}_{\text{dirt}}, \sigma = \sigma_{y-\text{dirt}})$. The randomness of $\hat{\rho}_{y-\text{dirt}}$ smears the distribution for $\hat{\rho}_{\text{coin}}$ and widens its uncertainty. The uncertainty of $\hat{\rho}_{y-\text{dirt}}$ also leads to looser the limits on ρ_{coin} .

4.6 A Particle Physics Search

This section extends the previous concepts to the search for a new particle in high energy particle physics. Such a search makes use of two hypotheses: the background only (B) and the signal plus background (S+B). The background only predicts the number of events if the particle does not exist, and the signal plus background (S+B) predicts the number of events if the particle does exist. The signal (S) represents the number of additional events expected if the new particle exists.

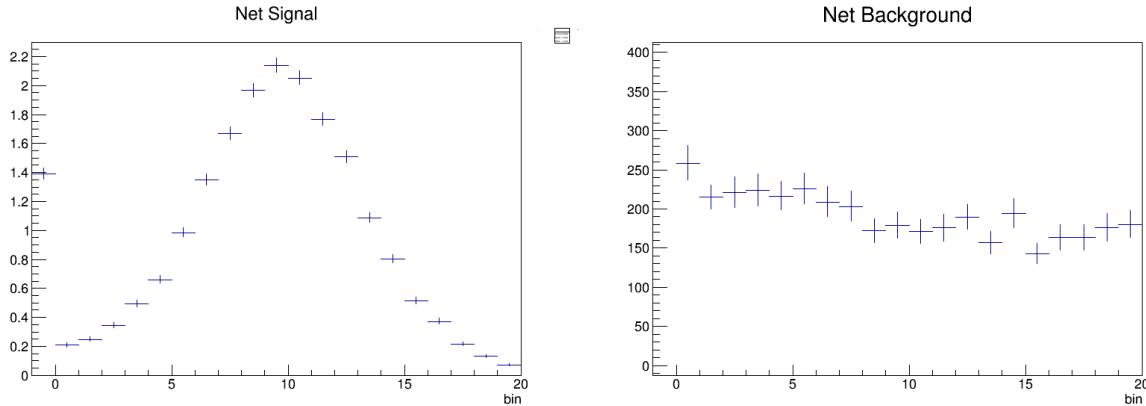


Figure 4-3. An example of the binned signal and background used for the background only and S+B hypotheses.

The S+B and B hypotheses are usually one dimensional distributions binned along some variable. If the observed distribution of events is too different from the background only, the background only hypothesis is ruled out and the experiment proclaims the discovery of a new particle. The distribution describing the amount of signal (S) is often reported in terms of the signal strength (ν) ¹ and the SM prediction (s) allowing the signal hypothesis to be written $S = \nu s$. A signal strength of 1 means that the signal model S is simply the Standard Model prediction, and a signal strength of 2 means that S has twice the signal as the SM prediction in each bin.

The probability to observe a particular count in a bin is described by the Poisson distribution

$$p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad (4-14)$$

where λ is the expected number of events and x is the observed number of events. A valid hypothesis provides the expected number of events in each bin and the data provides the observed number in each bin. The expected number of events in a bin is given by λ and the standard deviation is $\sigma = \sqrt{\lambda}$. When λ is far enough from zero the Poisson distribution may be described by a Gaussian with mean $\mu = \lambda$ and standard deviation $\sigma = \sqrt{\lambda}$.

4.6.1 Sensitivity

The sensitivity becomes much more interesting with binned distributions. First consider the sensitivity for a single bin. If the Standard Model or some other theory predicts the expected number of events in a bin to be $\lambda_i = S_i + B_i$ and the null predicts $\lambda_i = B_i$, then the expected sensitivity for discovery is,

$$Z = \frac{(x_i - \mu_i)}{\sigma_i} = \frac{(S_i + B_i - B_i)}{\sqrt{B_i}} = \frac{S_i}{\sqrt{B_i}} = \frac{\sqrt{N} \rho_{si}}{\sqrt{\rho_{bi}}}. \quad (4-15)$$

¹ Usually μ denotes the signal strength, but μ was already used in this section to designate the Gaussian mean.

As before, this scales with the \sqrt{N} . So, again, one way to ensure a sensitive experiment is to collect a lot of data, but that's not the only way. With many bins, the PDFs for each bin multiply and the N_{LL} in the Gaussian limit is,

$$-\ln \left(\frac{p}{C} \right) = -\ln \left(\prod_i p_i \right) + \ln(C) = -\sum_i \ln(p_i) + \ln(C) = \sum_i \frac{(x_i - \mu_i)^2}{2\sigma_i^2}, \quad (4-16)$$

where C is the sum of the normalizations for the Gaussians. After normalizing by C and the factor of 2, the N_{LL} is a sum of χ^2 variables and therefore itself a χ^2 variable,

$$\chi_{nll}^2 = -2\ln \left(\frac{p}{C} \right) = \sum_i \frac{(x_i - \mu_i)^2}{\sigma_i^2}. \quad (4-17)$$

This shows that the expected sensitivity with many Poissonian bins may be estimated by

$$Z^2 = \frac{(x_{nll} - \mu_{nll})^2}{\sigma_{nll}^2} = \sum_i \frac{S_i^2}{B_i}. \quad (4-18)$$

By concentrating the fixed amount of signal into a few bins with low background the sensitivity may improve regardless of the data available. This is indicative of the idea that the null may be invalidated when the data observed is many times the expected fluctuations.

4.6.2 Test Statistic

The negative log likelihood can be used to form the test statistic t , a random variable that summarizes the total discrepancy between the observed data and a given theory. The statistic is nice because it reduces the multitude of discrepancies in the many bins to a single number.

The test statistic t is given by,

$$t = -2\ln \left(\frac{p(x_i, \theta)}{p(x_i, \hat{\theta})} \right). \quad (4-19)$$

The normalization, $p(x_i, \hat{\theta})$, is the PDF with the parameters set to the best fit values, and this term as in Equation 4-16 sets the range of t from zero to infinity. When there is only one parameter of interest ν the likelihood may be reduced to one dimension by profiling the nuisance parameters θ ,

$$t = -2\ln \left(\frac{p(x_i, \nu, \hat{\theta})}{p(x_i, \hat{\theta})} \right). \quad (4-20)$$

As before, the $\hat{\theta}$ parameters are the best fit values for a fixed parameter of interest ν . The profiled test statistic may be used to set limits on the parameter of interest. Asymptotically, the test statistic is a χ^2 variable. See the analysis of section 4.6.1. Constraints may be included as necessary.

4.6.3 Constraints

When previous fits have been performed to determine certain parameters and their uncertainty, the likelihoods can be included to encode the information. The measurements on the nuisance parameters θ are included like so,

$$p(x_i; \nu, \theta) = \prod_i \text{Poisson}(x_i, \lambda_i(\nu, \theta)) \prod_j p(y_j, \theta_j), \quad (4-21)$$

where y_j represents the previously observed data and ν represents the signal strength. As in section 4.5, the observations y_j are summarized by the best fits ($\bar{\theta}_j$) rather than keeping track of the actual data. The net likelihood becomes,

$$p(x_i; \nu, \theta) = \prod_i \text{Poisson}(x_i; \lambda_i(\nu, \theta)) \prod_j C(\bar{\theta}_j, \theta_j, \sigma_j), \quad (4-22)$$

where C designates some constraint PDF. Constraints implement the various systematic and theoretical uncertainties involved in the analysis.

4.7 Using The Test Statistic

With a model for the expected yields in each bin, the distribution for t may be approximated by Monte Carlo methods, or by assuming that with enough data t reduces to a χ^2 distribution. The expected p-value against the background only may be calculated using the expected yields for S+B as the expected hypothesis and the background only yields as the null. Similarly, the expected upper limit on S may be calculated using t with S+B as the null and the background as the expected hypothesis. The expected upper limit at 95% confidence is calculated by finding the value of S high enough that observing the expected background-only has a p-value of 5%. A higher expected sensitivity and lower expected upper limit are important goals in the design of a physics analysis.

Upon collecting data, the test statistic t may be used with the data observed and the background only as the null to check for discovery. The observed upper limit on S at 95% confidence is computed by finding a high enough value of S such that observing the data has a p-value of 5%. These can be compared to the expected p-value and the expected upper limit.

In high energy physics the CLs method is often used to set the upper limits. The CLs method reports an upper limit using an adjusted p-value, $p_{\text{CLs}} = \frac{p_\mu}{p_{\text{bkg-only}}}$, where p_μ is the probability for a theory with a true parameter μ to observe x or less data and $p_{\text{bkg-only}}$ is the probability for the background only to observe x or less data. Using the CLs method at 95% confidence, p_μ must be $0.05p_{\text{bkg-only}}$ or less. When the data observed is much greater than the median expected background, the integrated probability less than the observed is most of the distribution, $p_{\text{bkg-only}}$ goes to 1, and the CLs upper limit approaches the standard upper limit. But in general, $p_{\text{bkg-only}}$ is less than one yielding a p_{CLs} greater than p_μ and a CLs limit that is more conservative than the standard upper limit. The CLs construction guarantees that hypotheses with low p_μ must also differ substantially from the background to be ruled out.

CHAPTER 5

THE $H \rightarrow \mu^+\mu^-$ SEARCH STRATEGY

The goal of the $H \rightarrow \mu^+\mu^-$ search is to discover the decay, and barring that to set an upper limit on the rate of production. The best fit for the rate of production is also reported. In order to do all of this, the analysis needs the data and the expected yields for the signal and background. The expected yields determine the likelihood, the PDF for the test statistic t , and the limits/p-values. The analysis fits Monte Carlo samples along $m_{\mu^+\mu^-}$ for the signal, and fits data along $m_{\mu^+\mu^-}$ for the background. The data at the collected luminosity has far better statistics than the Monte Carlo available for the background, and the larger statistics provide lower uncertainty in the bins of the PDF. The data and MC samples used in the analysis are covered in Chapter 6, and the models for the signal and background are covered in Chapter 8.

There is too much data to process to use all of it in the statistical analysis, and that would be wasteful anyways, so basic selections are made to look only at data likely to contain $H \rightarrow \mu^+\mu^-$ signal and unlikely to miss it. With a reasonable amount of data to look at, the Monte Carlo is validated against the data to make sure it agrees with observation. Some corrections like efficiency scale factors, trigger scale factors, and momentum corrections are applied to the Monte Carlo to ensure better agreement with the data. The background Monte Carlo is used to help with the data Monte Carlo validation, and to optimize the sensitivity, but not for the limit setting, discovery p-values, or measurement. The event selection and corrections are also covered in Chapter 6.

The analysis attempts to maximize the odds of discovering $H \rightarrow \mu^+\mu^-$ by maximizing the sensitivity. Because the sensitivity increases with the amount of data, improvements to the Level-1 Muon Trigger are made to increase the $\mu^+\mu^-$ data saved for the analysis. Moreover, the sensitivity improves with increased signal and worsens with increased background – recall that in the limit of large statistics, $Z^2 = \sum_i \frac{S_i^2}{B_i}$. As such, the ideal situation is one where the signal is concentrated into a single bin with little or no background. This isn't possible in

practice, but regions of feature space may be found where the signal is more concentrated or where there is little background or both.

The analysis corrects the muon momentum to narrow the $m_{\mu^+\mu^-}$ peak, improving its resolution, concentrating the signal, and improving the data/MC agreement. To further improve the sensitivity, Boosted Decision Trees (BDTs) [28] are trained to separate signal and background. The training produces a discriminating feature with a higher concentration of signal and lower concentration of background at large values. A novel algorithm then optimally and automatically categorizes the events by resolution and BDT score to maximize the expected sensitivity of the analysis. When the S+B fit is along a single dimension like $m_{\mu\mu}$, low sensitivity regions merge with high sensitivity regions of feature space polluting the chance for discovery. The categorization extracts the high sensitivity regions to regain the sensitivity.

The muon momentum corrections are covered in Chapter 6. The work to optimize the sensitivity is covered in Chapter 7. The limits, p-values, and signal strength measurement are presented in Chapter 9.

CHAPTER 6

SAMPLES, CORRECTIONS, AND SELECTIONS FOR THE $H \rightarrow \mu^+\mu^-$ SEARCH

This chapter presents the data and MC samples used in the analysis, the corrections to improve the data/MC agreement, and the selections defining the events used.

6.1 Data And Monte Carlo Samples

The proton-proton collision data used for the $H \rightarrow \mu^+\mu^-$ search is listed in Table 6.1.

The signal processes in 6-1 are represented by the MC in Table 6.1. The signal PDF is formed by fitting the GGF, VBF, W^+H , W^-H , ZH , and $t\bar{t}H$ MC then adding the fits together. The background MC used to optimize the categorization and to validate the simulation is listed in Table 6.1. While the background MC is used for validation and optimization, the background PDF used in the limit setting is formed by fitting the data in signal free control regions.

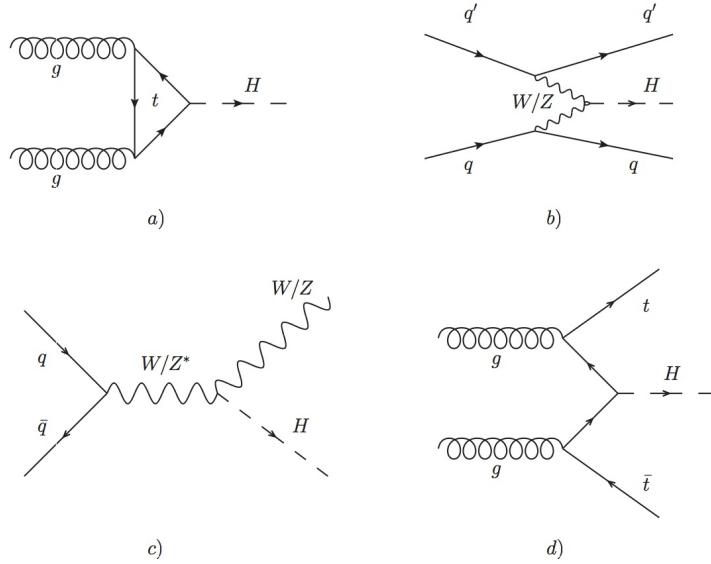


Figure 6-1. The SM production modes considered in the analysis. a) Gluon Gluon Fusion (GGF) b) Vector Boson Fusion (VBF) c) Associated Production with a Vector Boson (VH) d) Associated production with top quarks $t\bar{t}H$

Table 6-1. Overview of the single muon data stream collected during the proton-proton collisions at $\sqrt{s} = 13$ TeV by CMS at the LHC in 2016.

Dataset	Run Range	Integrated Luminosity [fb $^{-1}$]
/SingleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD	272007-275376	5.788
/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD	275657-276283	2.573
/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD	276315-276811	4.248
/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD	276831-277420	4.009
/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD	277772-278808	3.102
/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD	278820-280385	7.540
/SingleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD	280919-284044	8.606
/SingleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD		
/SingleMuon/Run2016*-03Feb2017-v*/MINIAOD	272007-280385	35.866

Luminosity mask: [Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt](#)

Table 6-2. The Higgs signal MC samples were generated with POWHEG while the parton shower and hadronization processes are modeled by the PYTHIA8 generator with TuneCUETP8M1.

Higgs signal MC samples	Events	Cross section [fb]	$\times \text{sec} \times \text{BR}$ [fb]
/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8/RunlISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6-v1/MINIAODSIM	250000	48.58	10.571
/VBF_HToMuMu_M125_13TeV_powheg_pythia8/RunlISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6-v1/MINIAODSIM	249200	3.7817	0.8229
/WPlusH_HToMuMu_M125_13TeV_powheg_pythia8/RunlISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6-v1/MINIAODSIM	124547	0.09426	0.02051
/WWplusH_HToMuMu_M125_13TeV_powheg_pythia8/RunlISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6-v1/MINIAODSIM	125000	0.0593	0.013019
/ZH_HToMuMu_M125_13TeV_powheg_pythia8/RunlISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6-v1/MINIAODSIM	249748	0.17762	0.03865

Table 6-3. The background MC samples were generated with amc@NLO, POWHEG and MADGRAPH. Spin effects in multi-boson processes are simulated using MADSPIN. The parton shower and hadronization processes are modeled by the PYTHIA8 generator with TuneCUETP8M1.

Background MC	Events	Cross Section [pb]
Drell-Yan		
/DY_Tol_L_01_1.3TeV-amcatnloFXFX-pythia8/RunlISummer16MiniAOv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext2_v1_MINIAODSIM	122055388	5765
/DY_Tol_L_11_1.3TeV-amcatnloFXFX-pythia8/RunlISummer16MiniAOv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	49579613	4754
/DY_Tol_L_13_1.3TeV-amcatnloFXFX-pythia8/RunlISummer16MiniAOv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	49902571	888.9
/DY_Tol_L_21_1.3TeV-amcatnloFXFX-pythia8/RunlISummer16MiniAOv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v2_MINIAODSIM	42324802	348.8
/DY_Tol_L_23_1.3TeV-amcatnloFXFX-pythia8/RunlISummer16MiniAOv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	47974554	348.8
Single Top		
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	5372991	35.85
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	3256650	35.85
/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	5425134	35.85
/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	3256407	35.85
Top Pair		
/TT_Jets_Dilept_TuneCUETP8M1_1.3TeV-madgraphMLM_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	6094476	85.66
/TT_Jets_Dilept_TuneCUETP8M1_1.3TeV-madgraphMLM_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	24350202	85.66
/TT_Jets_Dilept_TuneCUETP8M2T4_1.3TeV-amcatnloFXFX-pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	14529280	85.66
DiBoson		
/WWTo2L2Nu_1.3TeV_powheg/RunlISummer16MiniAOv2_FUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	1999000	12.46
/WZTo3LNu_TuneCUETP8M1_1.3TeV_amcatnloFXFX-pythia8/RunlISummer16MiniAOv2_FUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	11887464	2.113
/WZTo2L2Q_1.3TeV_amcatnloFXFX-pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	26517272	4.409
/ZZTo2l_2Nu_1.3TeV_powheg_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	8842475	0.564
/ZZTo2l_2Q_1.3TeV_amcatnloFXFX_madspin_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	15348572	3.22
/ZZTo4l_13TeV_amcatnloFXFX_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	10709784	1.212
TriBoson		
/WWW_4F_TuneCUETP8M1_1.3TeV_amcatnlo_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	240000	0.2086
/WWZ_TuneCUETP8M1_1.3TeV_amcatnlo_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	250000	0.1651
/WZZ_TuneCUETP8M1_1.3TeV_amcatnlo_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	246800	0.08565
/ZZZ_TuneCUETP8M1_1.3TeV_amcatnlo_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_v1_MINIAODSIM	249237	0.01398
SingleTop++X		
/tZq_ll_4f_13TeV_amcatnlo_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	14509520	0.0758
Top Pairs		
/TTWJets_Tol_Nu_TuneCUETP8M1_1.3TeV_amcatnloFXFX-madspin_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v3_MINIAODSIM	2160168	0.2043
/TTWJets_Tol_Nu_TuneCUETP8M1_1.3TeV_amcatnloFXFX-madspin_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext2_v1_MINIAODSIM	3120397	0.2043
/TTZToLLNuNu_M_10_TuneCUETP8M1_1.3TeV_amcatnlo_pythia8/RunlISummer16MiniAOv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1_MINIAODSIM	1992438	0.2529

6.2 Muon Momentum Calibration

The search for $H \rightarrow \mu^+ \mu^-$ looks for the SM signal peak in the $m_{\mu\mu}$ spectrum. This peak has a theoretical width of around 4 MeV, but the detector resolution on the order of a GeV dominates the observed width, diluting the sensitivity. Improving the dimuon mass resolution in data is critical to the analysis. Moreover, the mean and the resolution of the dimuon mass peaks in Monte Carlo, must match in data in order to set limits accurately. CMSSW provides two packages to address these issues, the Rochester Muon Corrections and the Kalman Filter Muon Corrections.

To assess the performance of the different corrections, dimuon mass histograms encompassing the Z peak are plotted for various windows of muon kinematic variables. The histograms for each kinematic window are fit with a Voigtian (the convolution of a Breit Wigner and a Gaussian), where the intrinsic width is set to the theoretical width of the Z peak. The mean and resolution of the fit are extracted and then plotted against the kinematic variable. One of these fits is shown in Figure 6-2 as an example.

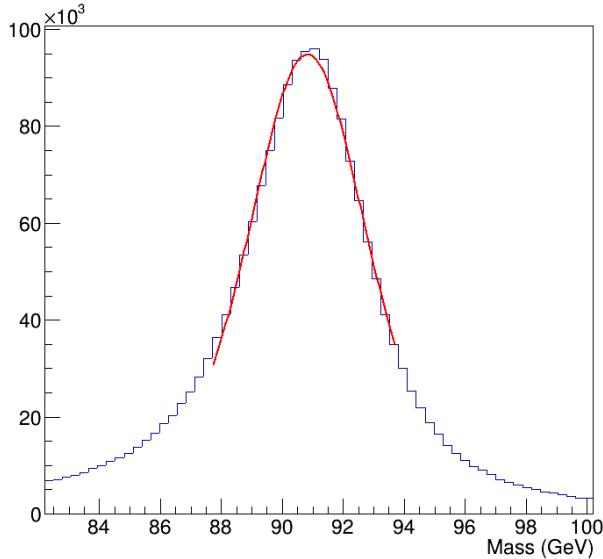


Figure 6-2. The dimuon mass distribution is fit around the Z peak for the positively charged muon with ϕ between 0 and 0.53.

As seen in the following sections the Rochester and Kalman filter corrections perform similarly. The search chooses to err on the side of robustness and use the Kalman filter corrections. The Rochester corrections are derived on the Z and should naturally perform a bit better. The Kalman filter corrections are derived on the J/Ψ and Z and are therefore expected to extrapolate to the Higgs peak more reliably. Moreover, the Kalman filter corrections are used in the measurement of $W \rightarrow \mu\nu$, which requires far greater precision on the mass.

6.2.1 Muon Corrections In Data

The expected behavior of the corrections on the Higgs peak in data is studied by examining the effects on the Z peak in data. Figure 6-3 shows the mean of the fitted Z peak plotted separately vs. ϕ for the positively and negatively charged muon for the Rochester, Kalman Filter, and uncorrected Particle Flow measurements. Also shown is the mean plotted vs. η and p_t^μ for the positively and negative charged muon. Finally, the mean is plotted vs. dimuon p_t as well. The corrections should provide consistent measurements of the Z peak in the different ϕ and η regions of the detector. By aligning the observed mass in ϕ and η , the sets of peaks that make up the inclusive set will align and the net resolution should improve. Note that after corrections, the mean is shifted closer to the theoretical value of 91.2 GeV. Furthermore, the Rochester corrections reduce the variation in phi from 0.1 GeV to 0.0025 GeV, an improvement of 98%. Figure 6-4 shows the Z resolution plotted against the same variables as in Fig. 6-3. The Rochester corrections improve the Z resolution by 1.6%, which translates to about 20 MeV, roughly 4 times the theoretical Higgs width. The Kalman filter corrections improve the resolution in data by 0.076%, which is about 10 MeV or 2 theoretical Higgs widths.

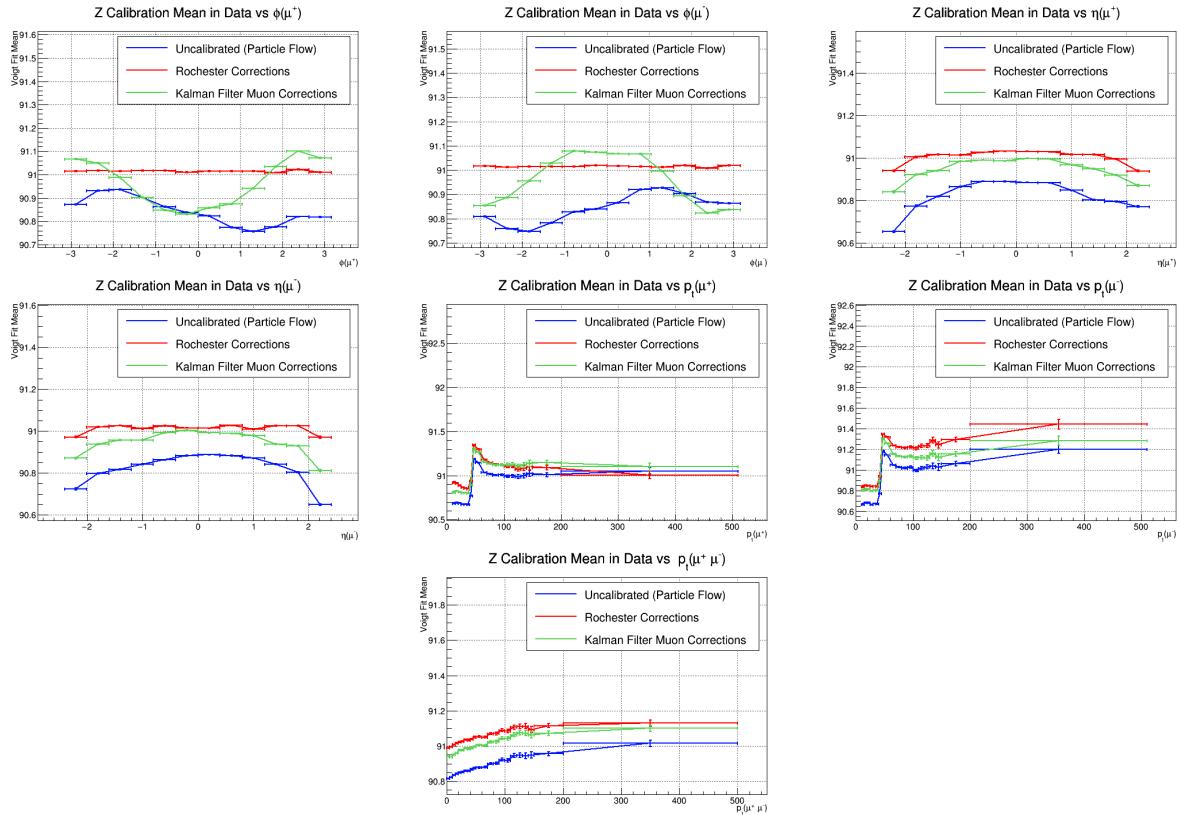


Figure 6-3. The Rochester (Roch) and Kalman Filter (KaMu) muon corrections are applied to data and compared to the uncalibrated Particle Flow (PF) predictions. The fitted mean of the Z peak is plotted vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for dimuon p_t as well.

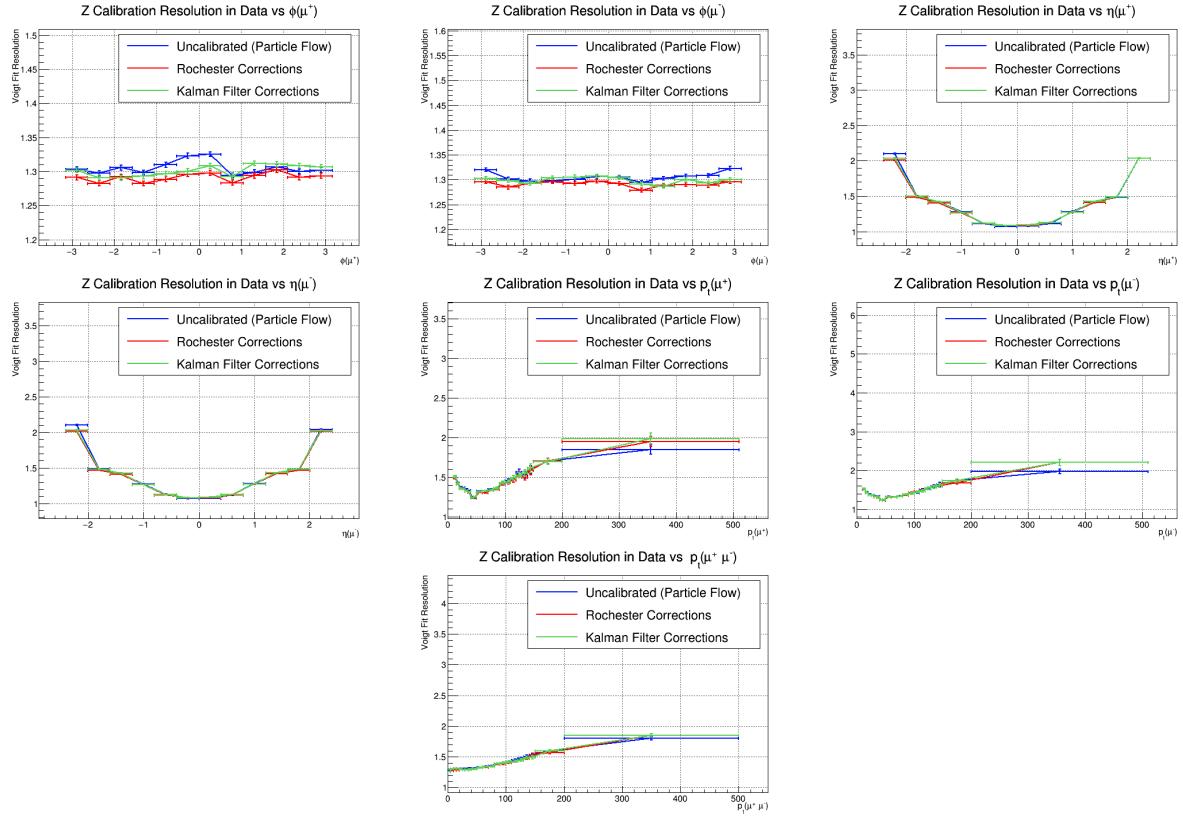


Figure 6-4. The Rochester (Roch) and Kalman Filter (KaMu) muon corrections are applied to data and compared to the uncalibrated Particle Flow (PF) measurement. The fitted Z peak resolution is plotted vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well.

6.2.2 Data-MC Agreement In Scale, Resolution

As aforementioned, the muon corrections should not only improve the resolution of the dimuon mass peaks in data, but align the scale and resolution between data and Monte Carlo as well. With the scale and resolution aligned between MC and data the signal peak (if the hypothesis is true) will appear in data with the width and mass predicted by the MC. When the MC simulates reality accurately, the limits will be set correctly for the different values of the Higgs mass. Figure 6-5 shows the Data/MC agreement on the Z peak mean before any corrections, while Figure 6-6 shows the agreement after Rochester corrections and 6-7 shows the agreement after the Kalman filter muon corrections. Both the Rochester and Kalman muon corrections successfully align Data/MC in terms of the Z peak mean.

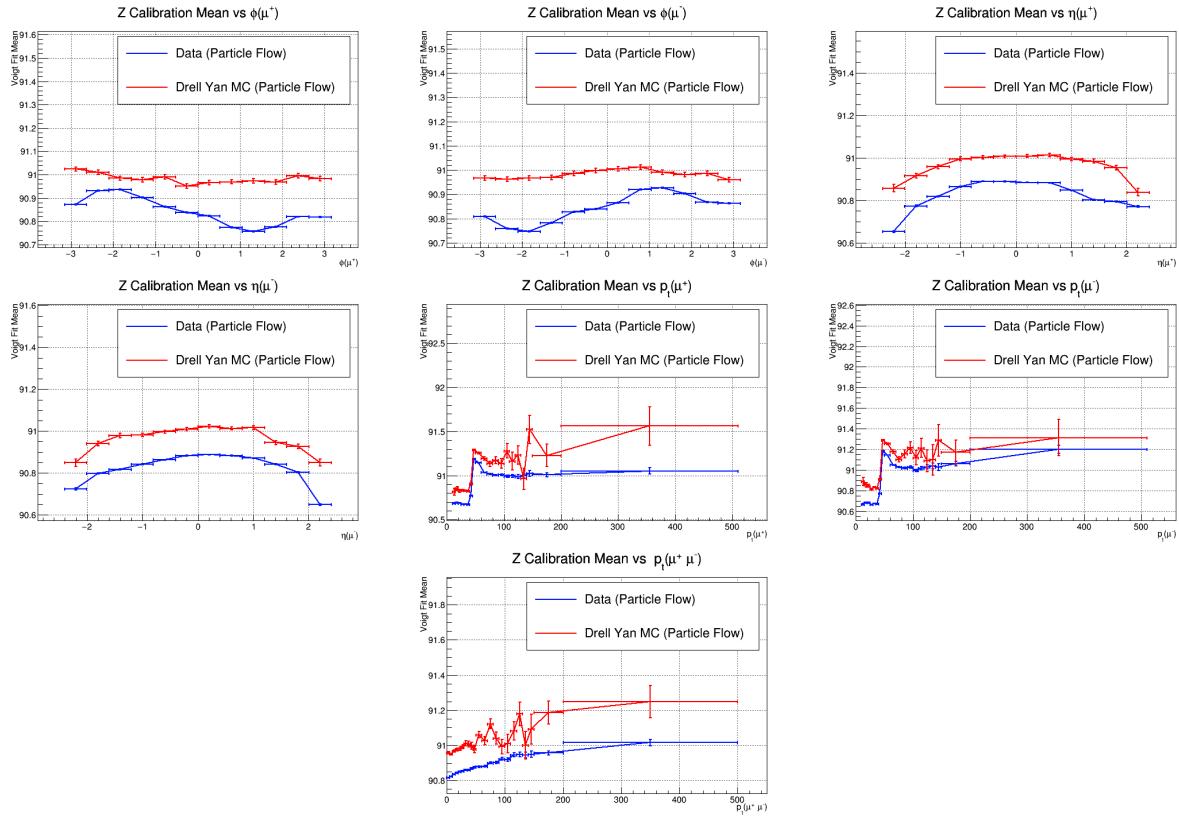


Figure 6-5. The fitted mean of the Z peak without any corrections is plotted in data and MC. The mean is plotted vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC are not aligned in terms of the Z peak mean before corrections.

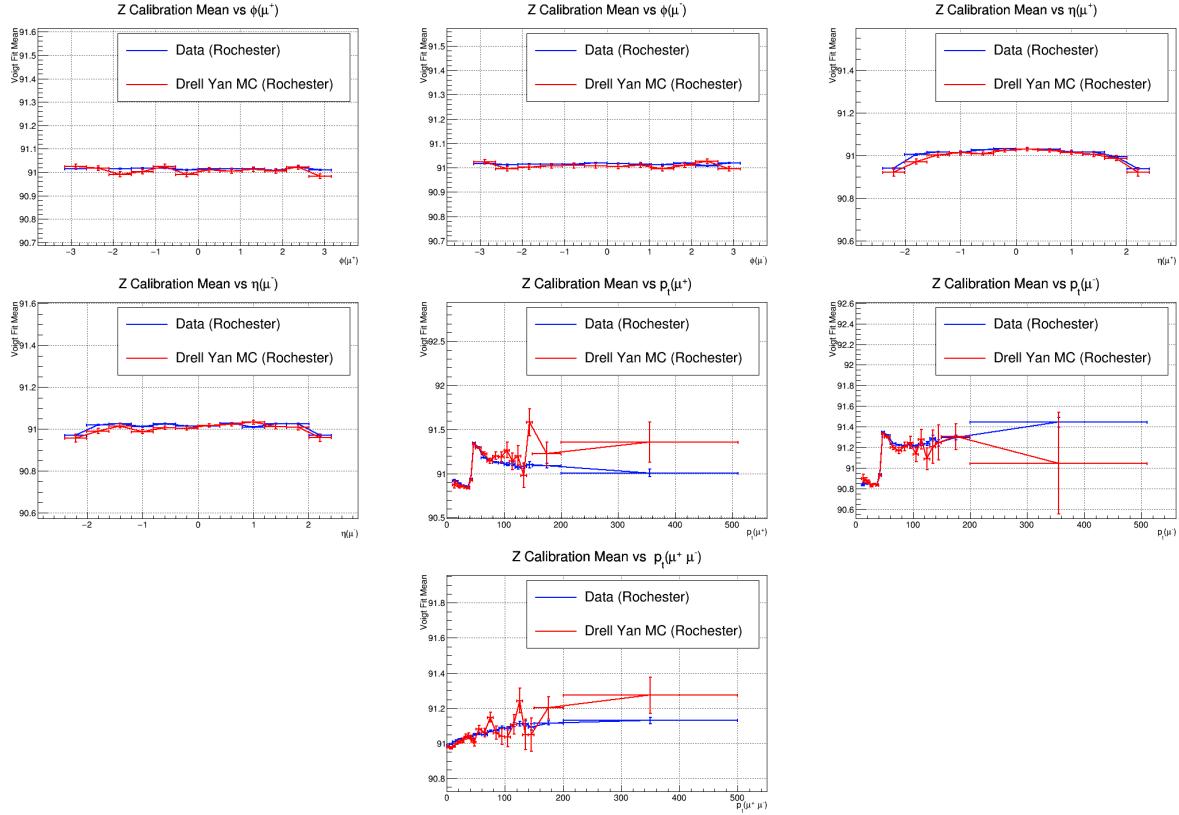


Figure 6-6. The fitted mean of the Z peak is plotted for the Rochester corrections. The mean is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC align very well in terms of the Z peak mean after applying the Rochester muon corrections.

Both corrections also succeed in aligning the Z peak resolution. The mismatch before corrections is seen in Figure 6-8 and the agreement after corrections is seen in Figure 6-9 and Figure 6-10.

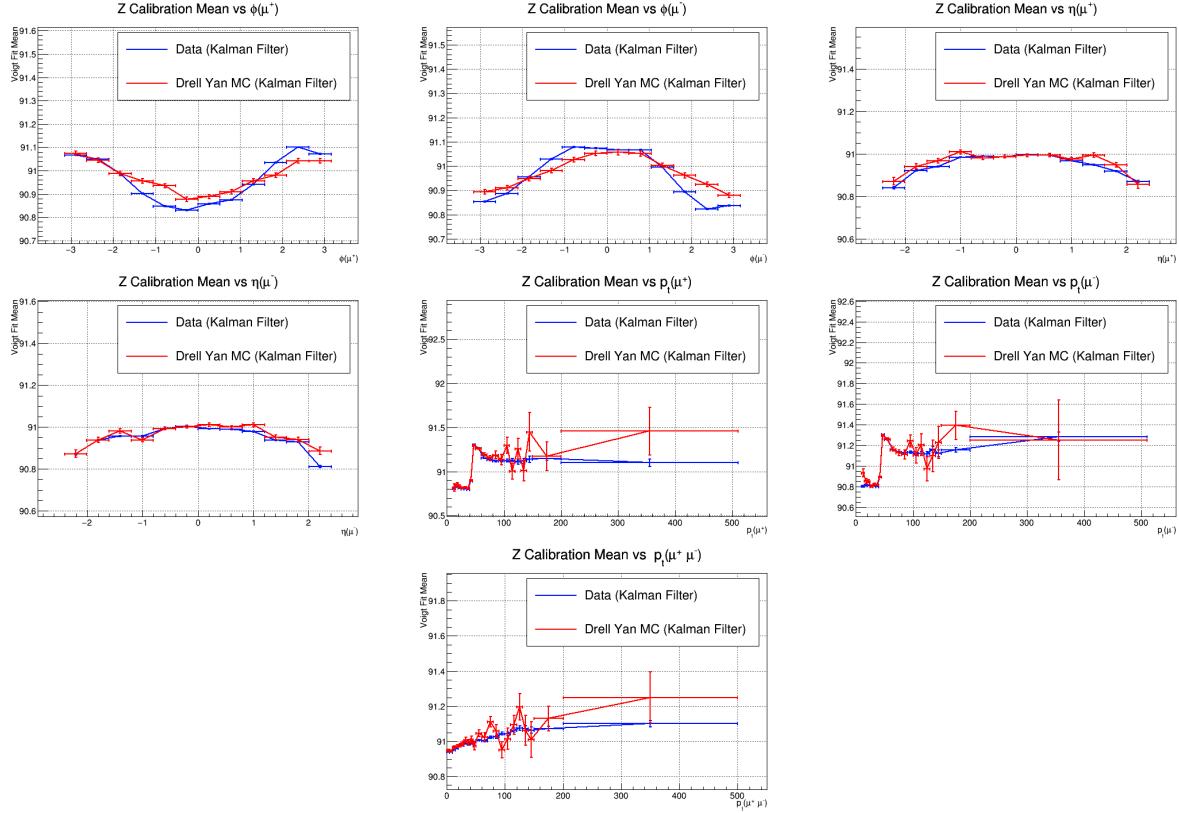


Figure 6-7. The fitted mean of the Z peak is plotted for the Kalman Filter corrections. The mean is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC align very well in terms of the Z peak mean after applying the Kalman filter muon corrections.

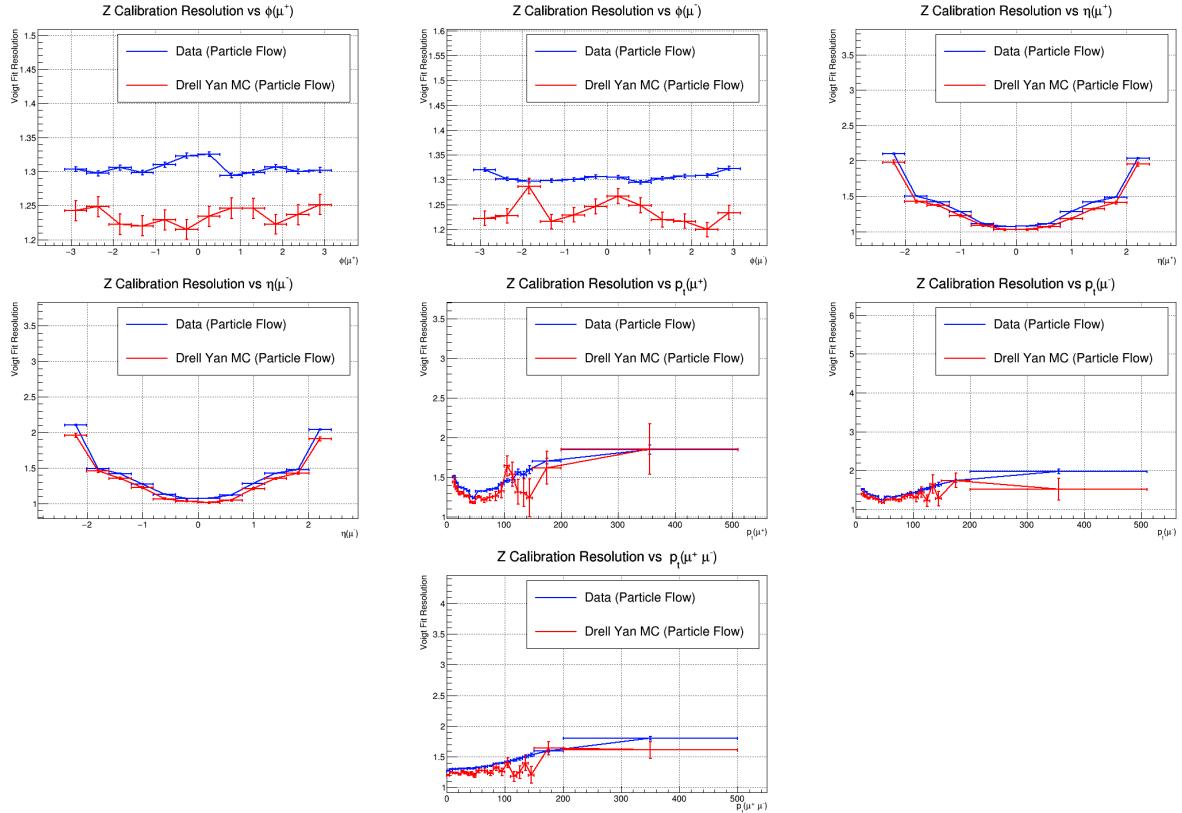


Figure 6-8. The fitted resolution of the Z peak without corrections is plotted. The resolution is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC do not show similar resolutions for the Z peak before corrections.

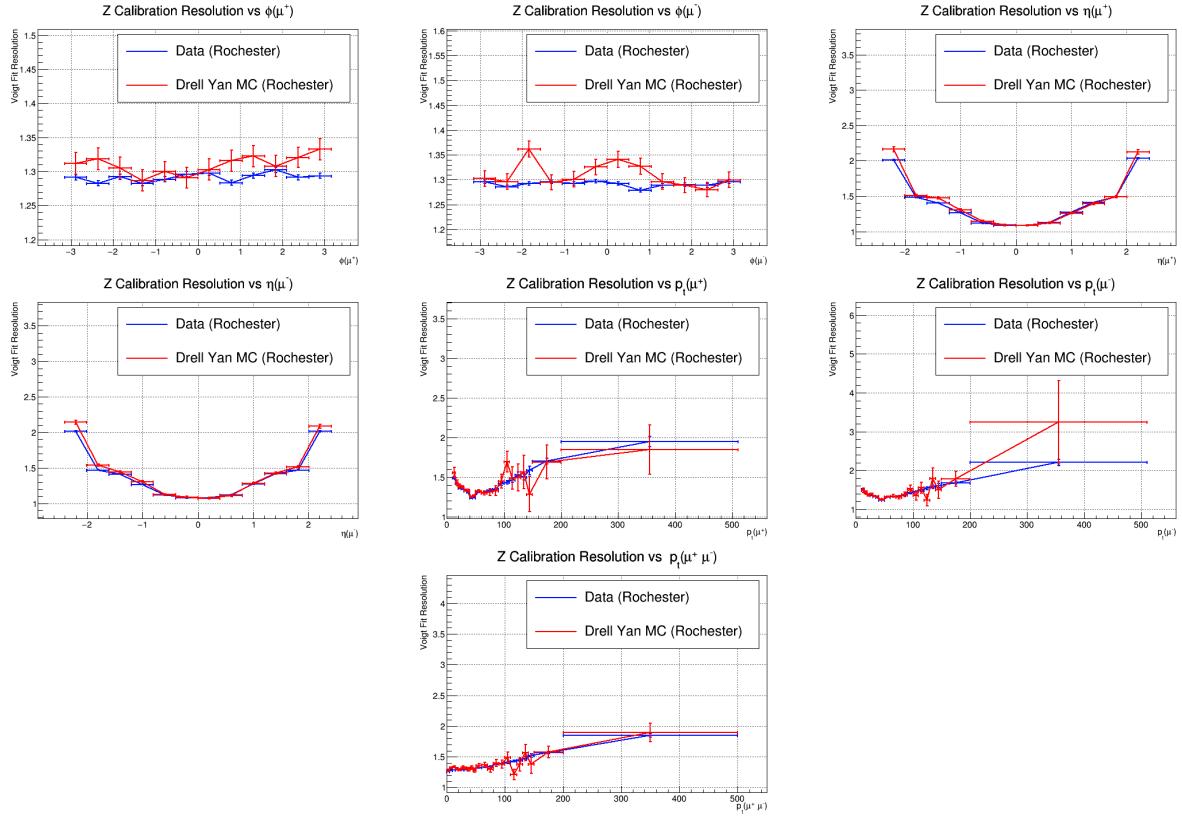


Figure 6-9. The fitted resolution of the Z peak is plotted for the Rochester corrections. The resolution is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. After Rochester corrections, data and MC have similar resolutions for the Z peak.

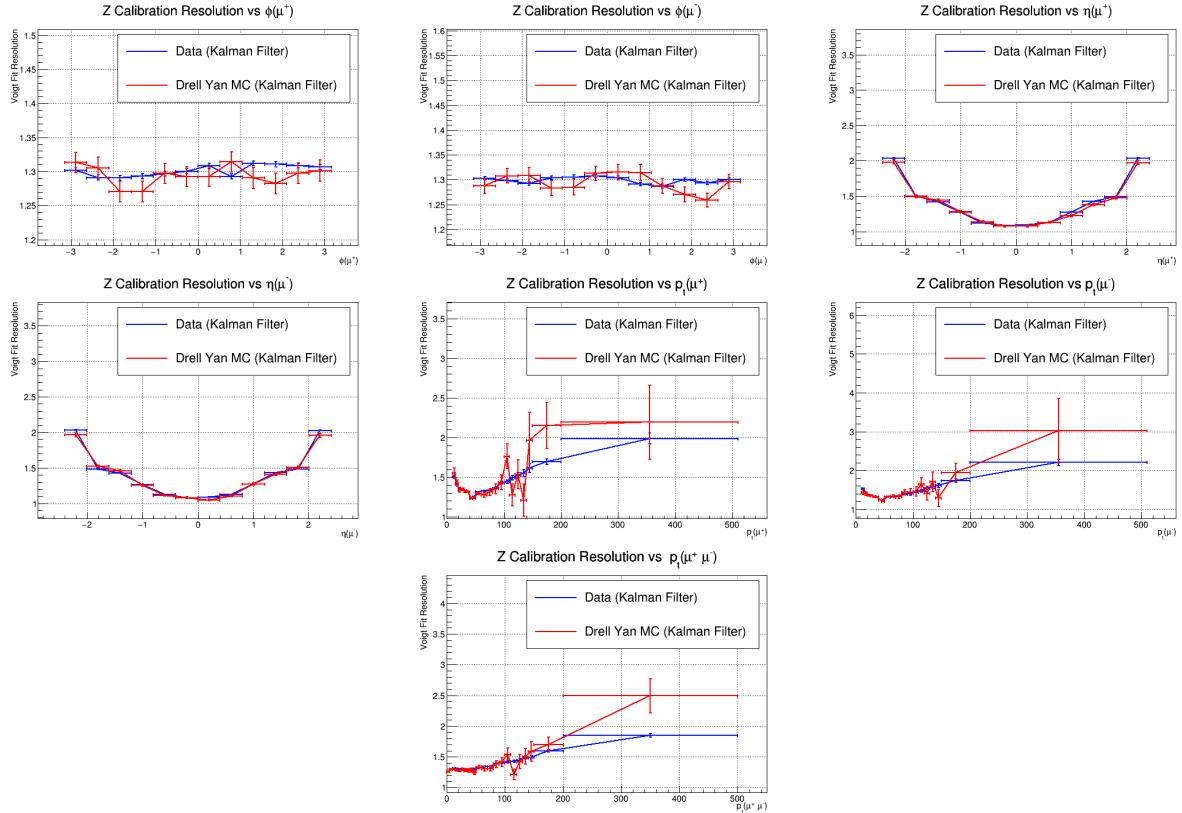


Figure 6-10. The fitted resolution of the Z peak is plotted for the Rochester corrections. The resolution is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. After Kalman filter muon corrections, data and MC have similar resolutions for the Z peak.

6.2.3 Derivation Of Systematic Uncertainties

Uncertainties are provided both for Kalman filter muon corrections and Rochester corrections. For the Kalman filter corrections, uncertainties include the non-closure on data. Histograms of the Kalman filter's most probable true mass value for the signal MC events are shown in Figure 6-11 for the least sensitive and most sensitive categories. No particular structure is observed as a function of category or production process. A conservative number of 0.05% is used for the scale uncertainty. Resolution corrections derived with the Kalman filter corrections have an uncertainty of 10%.

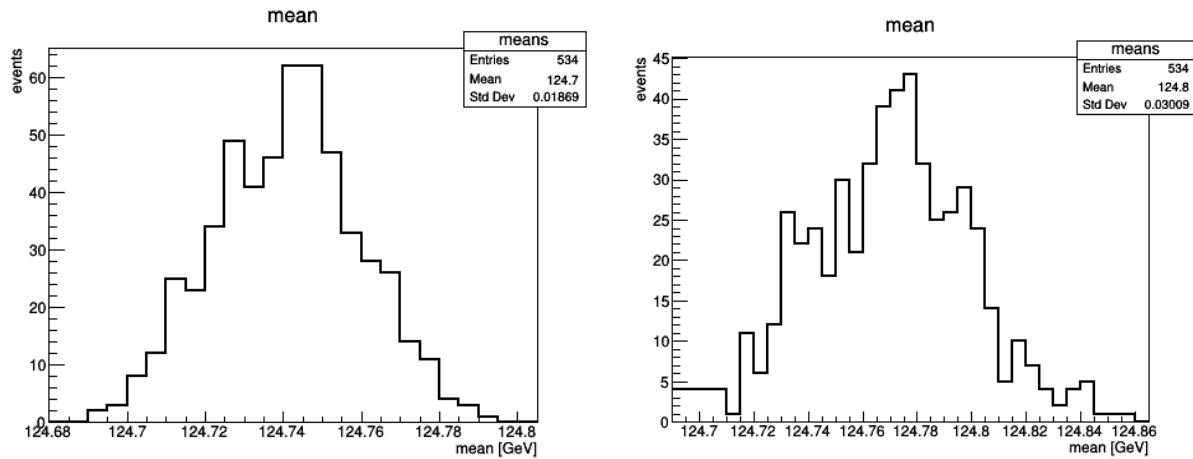


Figure 6-11. The Kalman filter's most probable true mass values are plotted for the signal MC. The least sensitive category is shown on the left and the most sensitive category is shown on the right.

T

6.3 Event Selection, Object Reconstruction, And Further Corrections

The Particle Flow (PF) algorithm reconstructs the detector level measurements at CMS into higher level objects like electrons, muons, photons, and jets. The $H \rightarrow \mu^+ \mu^-$ search uses the PF muons, jets, and missing transverse energy (MET) with four vector information like E , p_t , η , and ϕ to perform the analysis. Certain selections are made to reduce the size of the data while keeping the signal events. In addition, the η coverage of the subdetectors limits the available data for different objects. The analysis follows the Physics Object Group (POG) recommendations for 2016.

6.3.1 Muons

The PF algorithm forms muon objects by matching hits in the silicon tracker with hits in the muon chambers. Because the muon chambers end at $|\eta| = 2.4$, the analysis is limited to muons with $|\eta| \leq 2.4$. Moreover, the $H \rightarrow \mu^+ \mu^-$ signal is expected to occur around 125 GeV, which on average has 62 GeV muons. As such, the analysis requires muons with $p_t \geq 20$ GeV. Some muons originate from hadronic decays, surrounded by activity, while the signal produces isolated muons directly from the Higgs decay. To eliminate the hadronic muon background, only energetically isolated muons are considered. The isolation is determined by considering the energetic activity in a cone $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$ relative to the muon's p_T ,

$$I_{\text{rel}}^{\text{PF}} = (p_T^{\text{ch}} + \max(0, E_T^\gamma + E_T^{\text{nh}} - 0.5 * p_T^{\text{chPU}})) / p_T^\mu. \quad (6-1)$$

The term p_T^{ch} is the sum of the charged hadron p_T excluding hadrons from pile-up vertices, E_T^γ is the sum of the photon E_T , and E_T^{nh} is the sum of the neutral hadron transverse energy. The $0.5p_T^{\text{chPU}}$ term estimates the energy contribution from neutral pile-up as half the p_t of the charged pile-up. Muons with $I_{\text{rel}}^{\text{PF}} < 0.25$ in the cone are kept. In order to remove cosmic ray muons, muons from mid-flight decays, hadronic punch through, and muons with a poor p_t assignment, only muons that pass the Medium ID are considered.

6.3.2 Jets And MET

Jets are used to distinguish the VBF channel from the overwhelming Drell-Yan background. To avoid fake jets from hot calorimeter cells and readout effects, the jets must pass the Loose ID requirements. Moreover, to distinguish hard scattering jets from pile-up, the jets must also pass the Loose Pile-Up ID. Forward jets are especially indicative of VBF events, and the forward hadronic calorimeter goes all the way up to $|\eta| = 4.7$, so all jets with $|\eta| \leq 4.7$ are considered. High p_t jets are better measured and better distinguish VBF from Drell-Yan, so in order to keep these jets without being inundated, jets must have a $p_T \geq 30$ GeV. Some objects PF declares to be jets may actually be muons, so any jet within a $\Delta R = 0.4$ a muon is not

considered. All of the jets used in the analysis are anti- k_t PF jets with a distance parameter of 0.4.

Jets tagged as bottom quarks are used to reduce $t\bar{t}$ background. To this end, the analysis uses the centrally provided combined secondary vertex b-tagging algorithm (CSVv2) to identify likely b-jets. The CSVv2 medium working point is used, providing a 60% efficiency (true-positive) for b-jets, and a misidentification (false-positive) rate of about 1% for u, d, s, and gluon jets. The silicon tracker is needed to identify the secondary vertices corresponding to b-jets, and this subdetector ends at $|\eta| = 2.4$. Therefore the b-tagging algorithm is only applied to jets with $|\eta| \leq 2.4$.

The $H \rightarrow \mu^+ \mu^-$ search also uses the MET to reduce the $t\bar{t}$ background. By conservation of momentum, the \vec{p}_t for all of the particles in an event should add up to zero. A non-zero \vec{p}_t indicates that detector missed the p_t from some undetected neutral particles. The $|\vec{p}_t|$ associated with the "invisible" particles like neutrinos is called the MET¹

6.3.3 Pile-up Reweighting

Pile-up (PU) refers to particles emerging from collision vertices other than the vertex of interest. With high pile-up there are many more hits in the detector and the chance to misassign different quantities increases. As a consequence, discrepancies in the PU distributions between data and MC propagate to create discrepancies in the variables used in the analysis. To correct the PU and the concomitant effects, the MC is reweighted to match the data in the PU spectrum.

6.3.4 Muon Efficiency Scale Factors

In data, only a certain percentage of events that should pass the muon trigger, isolation, and ID requirements actually succeed. This probability is called the efficiency, and labelled ε .

¹ MET usually corresponds to neutrinos which have a tiny mass and a p_t basically equal to the E_t .

To correct the efficiency in the MC samples, scale factors (SFs) are applied to the MC events by p_t and η to scale the efficiency to the efficiency observed in data.

When there are multiple muons in an event, only one has to pass the trigger criteria, and the probability that at least one passes is given by $\varepsilon_{\text{trg}} = 1 - \prod_{i=1}^N (1 - \varepsilon_{\mu_i})$, where N is the total number of trigger matched muons with $p_t \geq 26$ GeV. The trigger SF is then,

$$SF_{\text{trg}} = \frac{\varepsilon_{\text{trg}}^{\text{data}}}{\varepsilon_{\text{trg}}^{\text{MC}}}. \quad (6-2)$$

On the other hand, every muon in the event must pass the isolation and ID requirements.

These efficiencies are given by $\varepsilon_{\text{id/iso}} = \prod_{i=1}^N \varepsilon_{\mu,i}$, where N is the number of muons, and the SFs are,

$$SF_{\text{id/iso}} = \frac{\varepsilon_{\text{id/iso}}^{\text{data}}}{\varepsilon_{\text{id/iso}}^{\text{MC}}}. \quad (6-3)$$

The efficiencies in data are different for data taking runs B through F compared to G and H, so the final scale factors are given by the weighted average,

$$w_{\text{BF}} = \frac{\mathcal{L}_{\text{BF}}}{\mathcal{L}_{\text{tot}}} \quad (6-4)$$

$$w_{\text{GH}} = \frac{\mathcal{L}_{\text{GH}}}{\mathcal{L}_{\text{tot}}} \quad (6-5)$$

$$SF = w_{\text{BF}} \cdot (SF_{\text{trg}}^{\text{BF}} SF_{\text{id}}^{\text{BF}} SF_{\text{iso}}^{\text{BF}}) + w_{\text{GH}} \cdot (SF_{\text{trg}}^{\text{GH}} SF_{\text{id}}^{\text{GH}} SF_{\text{iso}}^{\text{GH}}). \quad (6-6)$$

The scale factors are centrally provided by the Muon POG, and the weights are determined by the integrated luminosity.

6.3.5 Jet, MET, And B-tagging Corrections

Centrally provided corrections from the JetMET POG correct the jet energy to remove residual pile-up contributions, to reinstate known energy loss from reconstruction effects, and to align data and MC. The corrections are propagated to the Missing Transverse Energy (MET). While the MET is often physical, large values may arise due to reconstruction issues like cosmic ray muons, noise, or beam-halo particles. To remove these contributions to the reported MET value, the POG recommended MET filters are applied.

The JetMET POG also provides MC scale factors accounting for data/MC discrepancies in b-tagging efficiency and misidentification. A single SF for each MC event accounts for all of the jets in the event and their flavors (u,d,s,c,g) using the underlying MC truth information.

6.3.6 Event Selection

In order to consider a proton-proton event in the analysis, the event must have two oppositely charged muons from the primary vertex, with one muon matching either the HLT-IsoMu-24 or the HLT-IsoTkMu-24 trigger. The trigger matched muon must have a $p_T > 26$ GeV. The triggers chosen are the lowest p_T , unprescaled triggers available for the entire data taking period. In addition, the event must have at least one valid Primary Vertex (PV). The PV for the event is the one with the largest scalar sum of the p_t^2 and at least four associated tracks.

CHAPTER 7

OPTIMIZING THE SENSITIVITY OF THE $H \rightarrow \mu^+\mu^-$ SEARCH

This chapter covers the work to optimize the sensitivity of the analysis. The work on the Level-1 Muon Trigger is covered first, and the classification and categorization of events is covered after. The work on the trigger improves the sensitivity by increasing the amount of data available to the $H \rightarrow \mu^+\mu^-$ search, and the classification and categorization of events groups the events to maximize the chance of discovery.

7.1 Improving The Level-1 Muon Trigger

Saving as much relevant data as possible is important as the chance of discovery improves with a larger dataset. With this in mind, improvements are made to the L1 Muon Trigger to improve its capability to save events with high momentum muons and discard events without them. Towards this end, Boosted Decision Trees (BDTs) are trained to distinguish high momentum muons from low momentum muons. The BDTs are loaded onto Field Programmable Gate Arrays (FPGAs) in the L1 Muon Trigger and run online. The improvements increase the amount of data with high momentum muons available for the $H \rightarrow \mu^+\mu^-$ search.

The LHC collides bunches of protons every 25 ns at a center of mass energy of 13 TeV. The CMS experiment detects the resulting particles and measures their kinematics using a variety of subdetectors working in concert. With 40 million proton bunch crossings per second amounting to roughly 40 TB of data each second, saving the information from every event is not feasible. As such, the CMS trigger system chooses the interesting events to save to disk, operating in two stages [29]. The Level-1 (L1) trigger runs in hardware online reducing the throughput of data from 40 MHz to 100 KHz. From there, the High Level Trigger (HLT) operates in software online reducing the rate from 100 KHz to 1 KHz. In the end, about 1 GB/s is saved to disk.

With 40 MHz of input, the L1 Trigger has only 4 μ s to decide whether to keep the information for an event. The Endcap Muon Track Finder (EMTF) – part of the L1 Trigger

dedicated to muons – has only about 500 ns to determine the location, tracks, and momentum of the muons passing through the Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC) in the endcaps of CMS [30]. High momentum muons are an important object for many physics analyses at CMS and the most important object for $H \rightarrow \mu^+ \mu^-$. As such, an accurate momentum assignment distinguishing low momentum muons (background) from high momentum muons (signal) is key to the L1 Trigger. In order to meet the timing requirements, the EMTF’s logic is implemented in Field Programmable Gate Arrays (FPGAs), a type of reprogrammable hardware that allows vast parallelization and speeds much greater than even the best CPUs.

To improve the transverse momentum (p_t) assignment for muons in the endcaps at Level-1, the EMTF team trained Boosted Decision Trees (BDTs) offline using TMVA [31], and stored the prediction scheme into a 1.2 GB Look-Up Table (LUT). The FPGAs then use the LUT online to assign the p_t in a single operation. Using the LUT to turn the BDT p_t assignment into a simple look-up enables the EMTF to utilize the power of a robust machine learning algorithm for its momentum predictions while still operating at the required time scale. Putting a parallelized version of the BDTs directly into the FPGAs, while hypothetically possible, would require more than the available number of logic gates. Such an implementation would still be slower than the LUT method, and changing the p_t assignment would require reprogramming the FPGA logic each time. The LUT method provides a simple way to run any machine learning evaluation at high speed by turning the evaluation into a single operation.

7.1.1 Metrics Of Success

Two metrics are used to measure the success of the EMTF: the rate and the efficiency. The rate at X GeV is defined as the number of muons with a predicted p_t greater than X GeV. In other words, the rate consists of both true and false positives above the p_t threshold. The efficiency at X GeV is defined as the number of muons with both predicted p_t and true p_t greater than X GeV divided by the number of muons with true p_t above X GeV. Put another way, the efficiency measures the percentage of muons with true p_t above X GeV

correctly predicted above X GeV. A good trigger will minimize the data saved without losing the interesting high p_t events where unexplored physics lies, i.e. it will minimize rate while maximizing the efficiency.

7.1.2 The EMTF Regression Project

A muon traveling through the endcap detectors has a chance to leave hits in four sequential stations labeled 1, 2, 3, and 4. The specific combination of hits like 1,3,4 is called the mode. Each station records the ϕ and θ location of a hit, among other information. The CSCs have better spatial resolution, so the ϕ and θ information is taken from the CSCs by default, but the RPC measurements for the station are used if the CSCs missed the hit in the same station. The charged muons travel through a magnetic field following curved paths due to the Lorentz force. The force causes the high p_t muons in a magnetic field to bend less and the low p_t muons to bend more. The difference in ϕ and θ between stations i and j , $\Delta\phi_{ij}$ and $\Delta\theta_{ij}$, quantify the curvature of the track. With most of the curvature accounted for by the $\Delta\phi$ variables, the $\Delta\phi$ s provide the majority of the p_t discrimination.

A major difficulty in minimizing the rate is the steeply falling p_t distribution. A typical interesting event has p_t greater than 25 GeV, and there are about one thousand 5 GeV muons for every 25 GeV muon. With so many more low p_t events, predicting the low momentum muons poorly will drastically increase the rate. Moreover, in addition to the large number of low p_t muons, there are other notable difficulties: the muons travel through a non-uniform magnetic field, some scatter between detector stations, and those with high p_t often shower charged particles upon interacting with the detector material. Moreover, low p_t muons may spiral completely before getting to the next station. The scattering, showering, and spiraling add noise to the underlying true behavior, while the number of low p_t muons requires that the regression focus on the low momentum regime to prevent an explosion in the rate.

In order to assign p_t in a robust way and deal with the aforementioned difficulties, a BDT is trained for each possible mode using the discretized values for the features of Table 7-1. The loss function and weights are chosen to focus on the low p_t events and minimize the rate while

maintaining acceptable efficiency. Features are chosen for each mode to give the BDT the information needed to predict the p_t while dealing with the non-uniform magnetic field and the problematic scattering and showering effects.

The $\Delta\phi$ variables available for each mode are used as features to determine the curvature and get most of the p_t discrimination. However, the power of these variables depends largely on the track position in θ . The magnetic field varies as a function of θ affecting the magnitude of the curvature for a given p_t , thus correlating $\Delta\phi$, p_t , and θ . The link between these three makes θ the next most important training feature.

Variables modeling the mean and RMS of the available $\Delta\phi$ s for the mode are also used as features in order to identify scattering and showering effects. If a muon were to scatter or shower between stations the recorded hit in a station may not be the true hit of the muon. Any $\Delta\phi$ involving this station will be an outlier. To determine the severity of the deviation and the likelihood of scattering/showering, the idea is to identify the outlier station and to compare the mean and RMS $\Delta\phi$ with and without the outlier station. The greater the difference the greater the severity. The nominal mean and RMS $\Delta\phi$ features are calculated using all available $\Delta\phi$ s for the mode. The exclusive mean and RMS are calculated using all available $\Delta\phi_{ij}$ for the mode with i or $j \neq S_{out}$, where S_{out} is the outlier station. S_{out} is the excluded station such that leaving it out of the sum minimizes the mean and RMS. The outlier station, S_{out} , is also used as a feature. Including the nominal mean and RMS of $\Delta\phi$, the exclusive RMS and mean of $\Delta\phi$, and S_{out} as features helps the BDT differentiate scattering, showering, and normal events.

The features described above are the most important features, but not the whole collection. The front-rear (FR) bit designates whether the muon hit a front or rear CSC chamber in the station, and it is also included. The $\Delta\theta$ s provide additional curvature information, and these are included as well. The B feature for each station is included as well, and it flags whether the ϕ , θ information for the station came from the CSCs or the RPCs. If there are bits available for the B_i feature it also includes information about the single

station $\Delta\phi$ bend angle within a CSC chamber. Lastly, the $+/-$ feature stores the signs of the later $\Delta\phi$ s relative to the first $\Delta\phi_{ij}$ for the mode.

7.1.3 Putting The BDTs Into A Look-up Table

After training BDTs for each mode, the mode and the fundamental features from which the others can be derived are discretized and fit into a 30 bit word. The discretization scheme is different for each mode, detailed in Table 7-1. With the feature space compressed into 30 bits, there are 2^{30} possibilities that need to be assigned a p_t . A LUT is created by looping over all 2^{30} possible bit words, decoding each word into the fundamental features, deriving the secondary features, and sending the values to the BDT to assign the p_t prediction. Using 9 bits for the p_t , this amounts to a 1.2 GB LUT where each bit word value is an address and the p_t is the value in memory. Discretizing the feature space and creating a LUT turns the p_t assignment into a single operation. The LUT is then used by the FPGA logic online to assign p_t to muon tracks in the EMTF. The LUT method is a simple way to run any machine learning method quickly, but compressing the features into 30 or so bits may not always be feasible for the application.

7.1.4 Results And Conclusions

The LUT scheme utilizing the BDT predictions has been implemented in the EMTF for 2016 and 2017 data taking. As seen in Figure 7-1, the upgraded system – compared to the legacy system – reduces the rate at 25 GeV by a factor of three with no loss in efficiency. The legacy system was used in the endcaps until 2015. The improvements in the EMTF, in essence, save 3x the amount of data in the endcaps for $H \rightarrow \mu^+\mu^-$, which is a huge gain for the analysis.

Table 7-1. The feature discretization scheme for each mode.

Four Station Modes

Mode	Feature	$\Delta\phi_{12}$	$\Delta\phi_{23}$	$\Delta\phi_{34}$	+/-	$\Delta\theta_{14}$	B ₁	B ₂	B ₃	B ₄	FR ₁	θ	Mode
1-2-3-4	Bits	7	5	4	2	2	2	1	1	1	1	3	1

Three Station Modes

Mode	Feature	$\Delta\phi_{12}$	$\Delta\phi_{23}$	+/-	$\Delta\theta_{13}$	B ₁	B ₂	B ₃	FR ₁	FR ₂	θ	Mode
1-2-3	Bits	7	5	1	3	2	1	1	1	1	5	3

Mode	Feature	$\Delta\phi_{12}$	$\Delta\phi_{24}$	+/-	$\Delta\theta_{14}$	B ₁	B ₂	B ₄	FR ₁	FR ₂	θ	Mode
1-2-4	Bits	7	5	1	3	2	1	1	1	1	5	3

Mode	Feature	$\Delta\phi_{13}$	$\Delta\phi_{34}$	+/-	$\Delta\theta_{14}$	B ₁	B ₃	B ₄	FR ₁	FR ₃	θ	Mode
1-3-4	Bits	7	5	1	3	2	1	1	1	1	5	3

Mode	Feature	$\Delta\phi_{23}$	$\Delta\phi_{34}$	+/-	$\Delta\theta_{24}$	B ₂	B ₃	B ₄	FR ₂	-	θ	Mode
2-3-4	Bits	7	5	1	3	2	1	1	1	-	5	4

Two Station Modes

Mode	Feature	$\Delta\phi_{XY}$	$\Delta\theta_{XY}$	B _X	B _Y	FR _X	FR _Y	θ	Mode
X-Y	Bits	7	3	3	3	1	1	5	7

X-Y runs through the possible two station combinations: 1-2, 1-3, 1-4, 2-3, 2-4, 3-4.

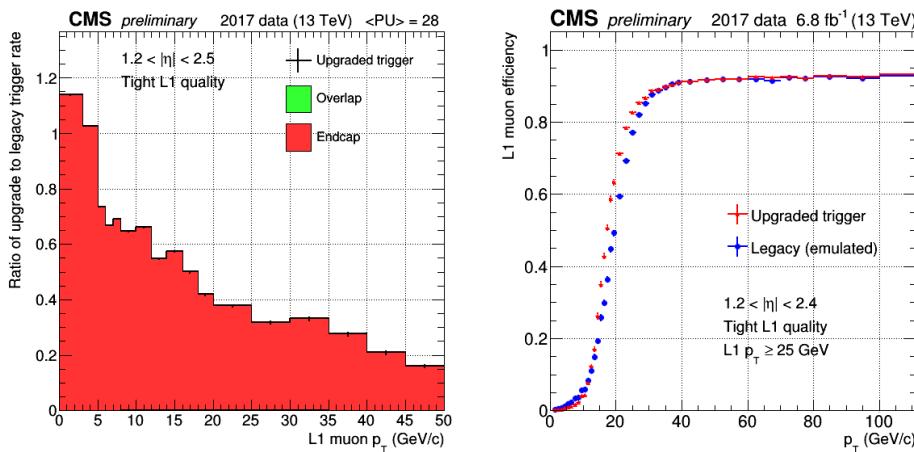


Figure 7-1. On the left, the upgraded EMTF rate divided by the legacy rate is shown for a variety of p_T thresholds. On the right, the upgraded and legacy efficiencies are presented for a 25 GeV threshold. The upgraded EMTF has a 3x lower rate than the legacy system at 25 GeV with virtually no difference in plateau efficiency for the same threshold. Plots are taken from [32].

7.2 Classifying Events

Practically, it is difficult to perform a statistical analysis with a PDF in many dimensions. In a binned analysis, the density of the data points in each bin decays rapidly as the number of dimensions increase. To mitigate this problem, the $H \rightarrow \mu^+\mu^-$ analysis uses a one dimensional PDF along the dimuon mass spectrum. Unfortunately, collapsing the muon and jet information mixes areas of high sensitivity with areas of low sensitivity and undercuts the power for discovery. To recover the lost sensitivity, the analysis uses the muon and jet information in the MC to train Boosted Decision Trees (BDTs) to distinguish signal from background. The BDTs map the events from the many kinematic dimensions onto the BDT score, whose spectrum places a large density of signal and low density of background at large values.

The high BDT score regions will have $m_{\mu\mu}$ distributions with more signal in the peak over less background, recovering the sensitivity. The sensitivity can be further improved by concentrating the signal peak into a narrower region. To this end, the events are extracted and placed into categories based upon both the BDT score and the mass resolution.

7.2.1 Training Boosted Decision Trees

BDTs are trained using ROOT's TMVA to distinguish signal from the major backgrounds, Drell-Yan (DY) and $t\bar{t}$. VBF is easiest to distinguish from Drell-Yan as the VBF channel has two jets at leading order, while Drell-Yan has none. VBF events characteristically have two forward jets with large dijet mass. A top quark decays most often to a W and b, and to be seen in $\mu^+\mu^-$ data the W decays to a neutrino and a muon. As a result, the $t\bar{t}$ events may be distinguished by the presence of b-jets and above average MET. GGF and DY are very similar. In GGF, two gluons fuse and make a Higgs, which then decays to two muons. In DY, two quarks fuse and make a Z/photon which then decays to two muons. In both cases, the final state at leading order is just the two muons. However, gluons are a bit more likely to radiate jets than quarks, so the average dimuon p_t is a bit higher for GGF. The dimuon p_t provides some discrimination power as do the jet variables, but the GGF and DY events are still very similar on average.

The features used to train the BDTs are chosen to maximize the discrimination capabilities without changing the $m_{\mu\mu}$ distribution. The features are

- the p_t and η of the dimuon system,
- the $|\Delta\eta|$ and $|\Delta\phi|$ between the muons,
- the η values of the two highest- p_t jets,
- the masses of the two highest-mass dijet pairs,
- the $|\Delta\eta|$ between the jets in the two highest-mass pairs,
- the number of jets with $|\eta| < 2.4$ and $|\eta| > 2.4$,
- the number of jets passing the CSVv2 medium b-tag working point,
- and the MET.

The p_t for the individual muons is left out to prevent any reshaping of $m_{\mu\mu}$. If any peaks or troughs are arbitrarily created in the $m_{\mu\mu}$ spectrum, this could lead to a false discovery. The feature importance at first order is listed in Table 7-2. The ranking plots histograms of the signal and background along the feature and finds the percentage of the total area that does not overlap.

Table 7-2. The importance is given by the percentage of the signal and background PDFs along that feature that do not overlap.

Feature	Importance
Dimuon p_t	5.342e-02
Dimuon η	3.896e-02
$ \delta\phi(\mu\mu) $	3.500e-02
Number of medium b-tags	2.842e-02
$\eta(\text{jet1})$	1.724e-02
MET	1.706e-02
Number of forward jets	1.218e-02
$ \delta\eta(jj_1) $	8.310e-03
Number of central jets	8.036e-03
$\eta(\text{jet2})$	7.460e-03
$ \delta\eta(\mu\mu) $	6.793e-03
$M(jj_1)$	6.546e-03
$ \delta\eta(jj_2) $	3.304e-03
$M(jj_2)$	2.199e-03

The BDT training and optimization uses 50% of the available DY and $t\bar{t}$ MC for training and 50% to evaluate the effectiveness of the training. On the other hand, 25% of the major signal MC (GGF, VBF, and VH) is used for training and 25% for testing, and 50% is excluded

from training/evaluation. The excluded 50% of the signal MC is used at a later stage for an unbiased report of the limits. The background PDF used for the limits is estimated from the sidebands in the data, so it's not necessary to withhold any of the background MC. The number of trees, nodes, and learning rate were optimized based upon the area under the Receiver Operating Characteristic (ROC) curve. The final BDT settings use 500 trees, a depth of 5, and a learning rate of 0.1. The gradient boosting optimizes the binary cross entropy loss function.

Histograms over the features are shown in Figures 7-7 and 7-8 for the events passing the initial event selection. The separation of the signal and background along the BDT score is plotted in Figure 7-3. The Receiver Operating Characteristic (ROC) curve for the final BDT model is shown in Figure 7-2. With the BDT score the events may now be categorized.

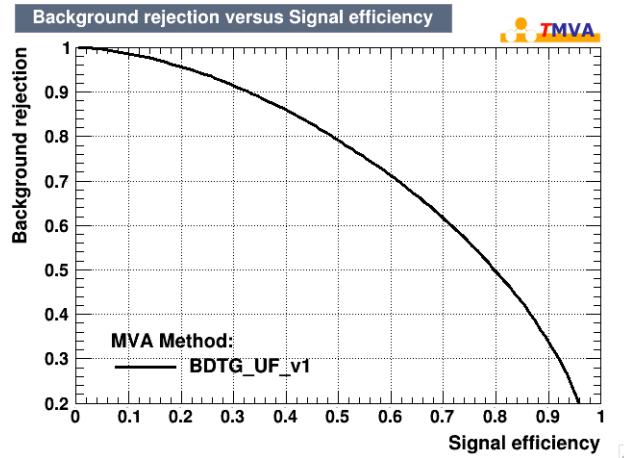


Figure 7-2. The background-rejection vs. signal efficiency ROC curve for the final BDT model.

7.2.2 Event Categorization

The analysis optimizes the sensitivity of the $H \rightarrow \mu^+ \mu^-$ search, by categorizing the events based on two criteria: the BDT output score and the muon $|\eta|$ information. The output score of the BDT is used to separate signal from background events, where the score ranges from -1 (background-like) to +1 (signal-like). The maximum $|\eta|$ value of the two candidate muons is used to select events with better mass resolution. In order to optimize the choice of BDT and $|\eta|$ cuts, the analysis designed and implemented a decision tree autocategorizer to create categories by greedily optimizing a metric corresponding to the sensitivity. The BDT training

and the autocategorization are performed on 50% of the signal MC events and 100% of the background MC events. The autocategorizer uses the same background and signal MC as the BDT training. After training the autocategorizer on the signal and background MC to determine the categories, the expected limits are calculated using data and the remaining 50% of simulated signal events. Separating the signal MC into exclusive sets for optimization and limit evaluation, and excluding the data from the training phase, ensure an unbiased estimate for the limits.

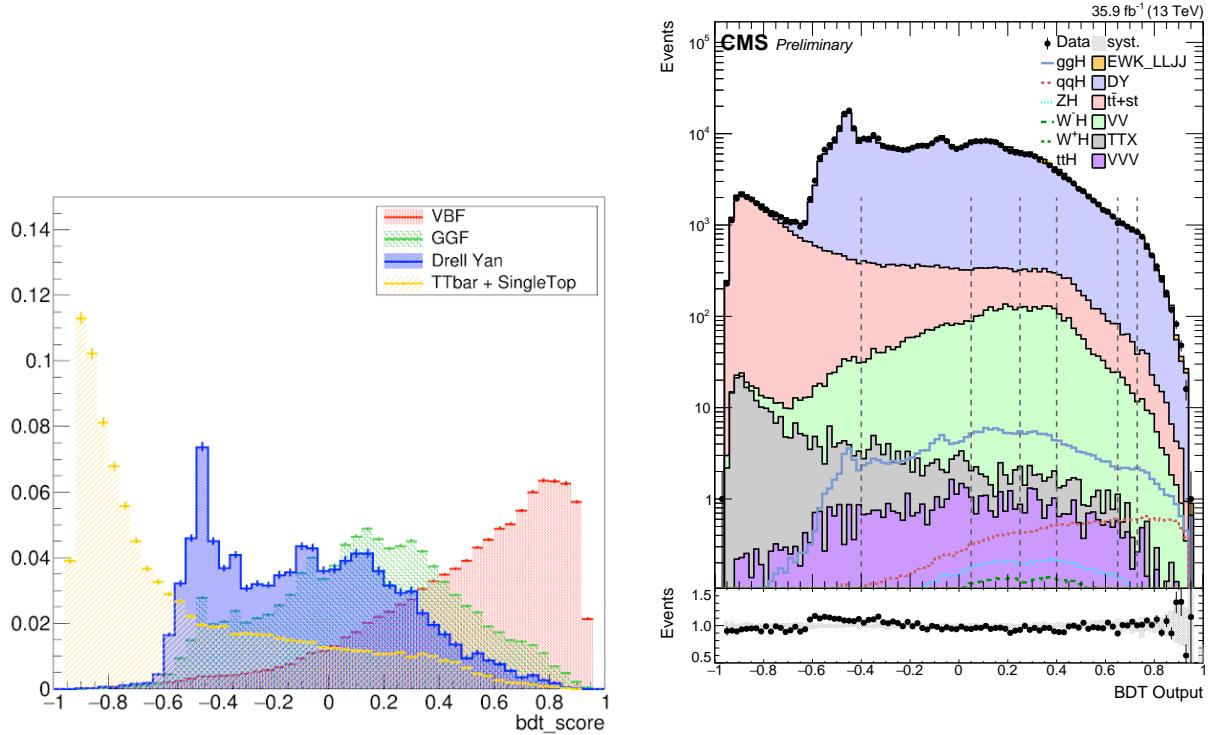


Figure 7-3. The BDT score on the signal and background Monte Carlo (left). A higher score indicates signal-like events. Note $t\bar{t}$ on the left as the most distinguishable background and VBF on the right as the most distinguishable signal. The BDT score in data is modeled well by the amc@NLO Drell–Yan MC (right).

7.2.3 The Autocategorizer

A novel algorithm has been designed to categorize the events and maximize the sensitivity of the search. The categorization takes into account both the signal/background discrimination and the resolution of the mass peak. In order to account for the resolution, the signal and background events are binned in half-GeV bins in the signal region of the dimuon mass

spectrum, 120 to 130 GeV. The significance in each bin is then given by S/\sqrt{B} . The

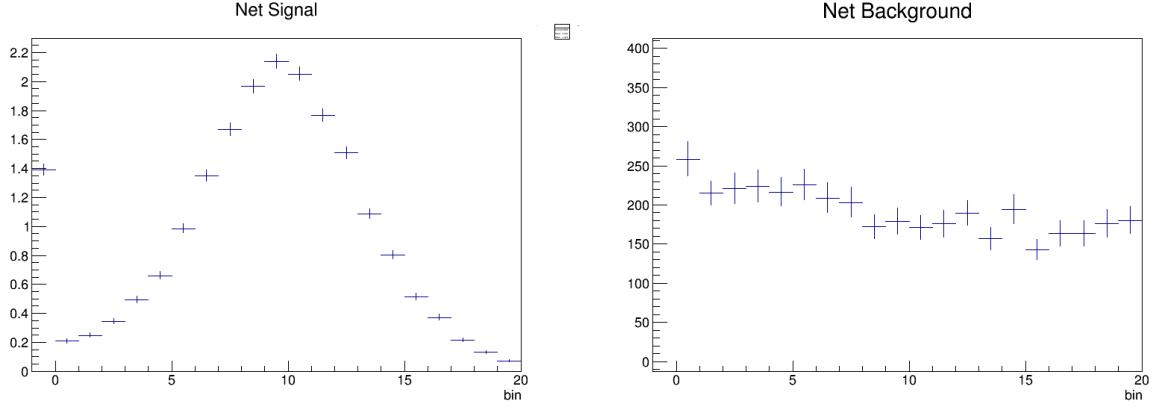


Figure 7-4. An example of the binned signal and background necessary for the significance metric. The binning keeps the most sensitive bins from being drowned out by the others, providing a measure that accounts for both signal–background discrimination and resolution.

significance for independent bins adds in quadrature, so the net significance is given by

$$\text{Net Significance} = \sum_{c,i} S_{c,i}^2 / B_{c,i} \quad (7-1)$$

where i labels the bin in the signal region, and c denotes the category. $S_{c,i}$ is the number of expected signal events in the i^{th} bin in that category, and $B_{c,i}$ is the number of expected background events in the i^{th} bin. By binning finely enough, the central mass bins contribute the most and the resolution is accounted for.

On the first iteration, the autocategorizer calculates the net significance for the set of all events, the inclusive set. The algorithm then searches over the inclusive set, checking all possible split values of the maximum muon $|\eta|$. Events with maximum $|\eta|$ values less than the split go in one candidate category, and those with max $|\eta|$ values greater than or equal to the split value go into the other candidate category. For every split candidate, the algorithm calculates the net significance in the two categories delineated by the split value. The maximum $|\eta|$ cut value that provides the largest gain in significance over the inclusive set is stored. The gain is defined in the equation below where $c1$ and $c2$ are the prospective

categories created from c by splitting on the feature.

$$\text{Gain} = \sum_i S_{c1,i}^2/B_{c1,i} + \sum_i S_{c2,i}^2/B_{c1,i} - \sum_i S_{c,i}^2/B_{c,i} \quad (7-2)$$

The autocategorizer then searches over the BDT score values and stores the BDT score that provides the largest gain in significance. The algorithm chooses to split on either the maximum muon $|\eta|$ or the BDT score (whichever has the higher gain), creating two categories from the inclusive set of events. At the next iteration, the autocategorizer repeats the procedure for the two new categories and chooses to split the category that provides the most gain. This process continues, each time greedily choosing to split the category with the most gain, until the number of categories desired is reached.

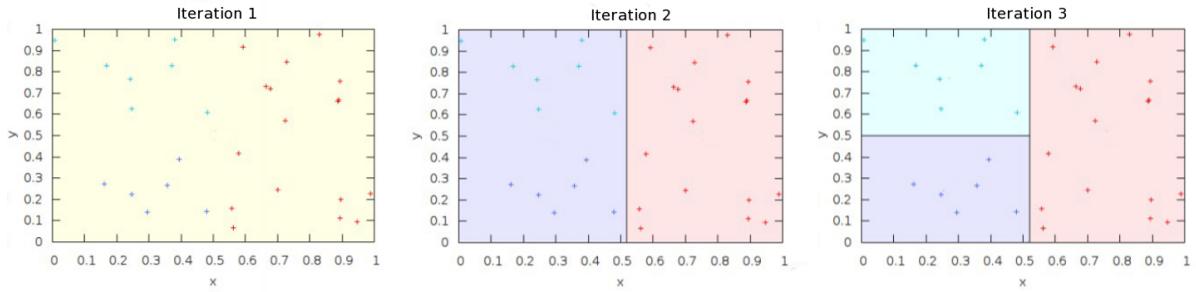


Figure 7-5. An example of the categorization process for some toy features x and y and three categories. The colored crosses represent the events that should be grouped together for optimum sensitivity. After three iterations, the categorizer correctly groups those events. The autocategorizer chooses $x=0.52$ for the first split and $y=0.50$ for the second.

7.2.4 Final Categories

After training the BDT and autocategorizing based upon the maximum muon $|\eta|$ and the BDT score, the resulting categorization is formed by rounding some of the cuts. Care is taken so that the simplification does not worsen the expected limit. The 15 BDT categories produced by the Autocategorizer in Figure 7-6 show a 23% improvement compared to the 15 Run I $H \rightarrow \mu^+\mu^-$ categories. For the comparison, the expected upper limits are evaluated on the same 36 fb^{-1} of 2016 data and signal Monte Carlo in the same way for an even comparison. The categories are ordered by BDT Score and $|\eta|$.

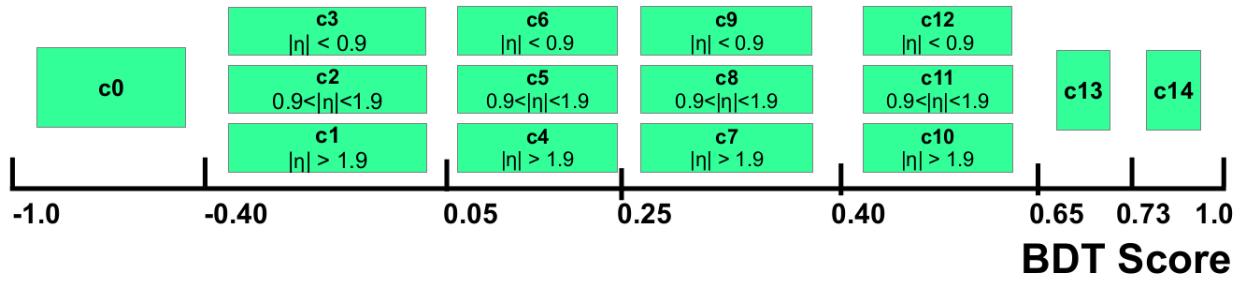


Figure 7-6. The final categorization.

The FWHM of the mass resolution, the expected signal and background yields, the sensitivity, and the expected limit for each of the final 15 categories are shown in Table 7-3.

Table 7-3. Full Width Half Max (FWHM) of the signal peak, expected signal and background yields within the FWHM, purity, sensitivity, and the expected upper limits at 95% confidence. The background yields and the expected limits use background models fit to the data in the sidebands.

Category	Signal FWHM (GeV)	Signal	Background	S/B (%)	S/\sqrt{B}	Limit
c0	4.4	13.3	14400	0.000927	0.111	19
c1	6.0	13.9	7990	0.00174	0.156	16
c2	4.3	27.7	10000	0.00276	0.277	7.9
c3	3.1	8.70	2600	0.00334	0.170	12
c4	6.3	7.66	2750	0.00279	0.146	17
c5	4.3	19.6	4320	0.00453	0.298	7.1
c6	3.1	9.86	1540	0.00638	0.250	8.4
c7	6.2	3.44	928	0.00370	0.113	22
c8	4.3	14.0	2310	0.00607	0.292	7.9
c9	3.2	8.85	1050	0.00838	0.272	7.3
c10	6.2	3.62	641	0.00564	0.142	16
c11	4.5	14.4	1830	0.00788	0.338	6.1
c12	3.4	10.1	855	0.0118	0.348	6.1
c13	4.2	6.15	422	0.0145	0.299	7.7
c14	4.5	9.84	387	0.0254	0.500	4.6

7.2.5 Validating The Categories

It is important that the signal is correctly modeled in each category in order to report an accurate expected upper limit. Figures 7-7 and 7-8 show the data/MC agreement for the BDT features as well as the dimuon mass for the inclusive set of events with dimuon mass greater than 60 GeV.

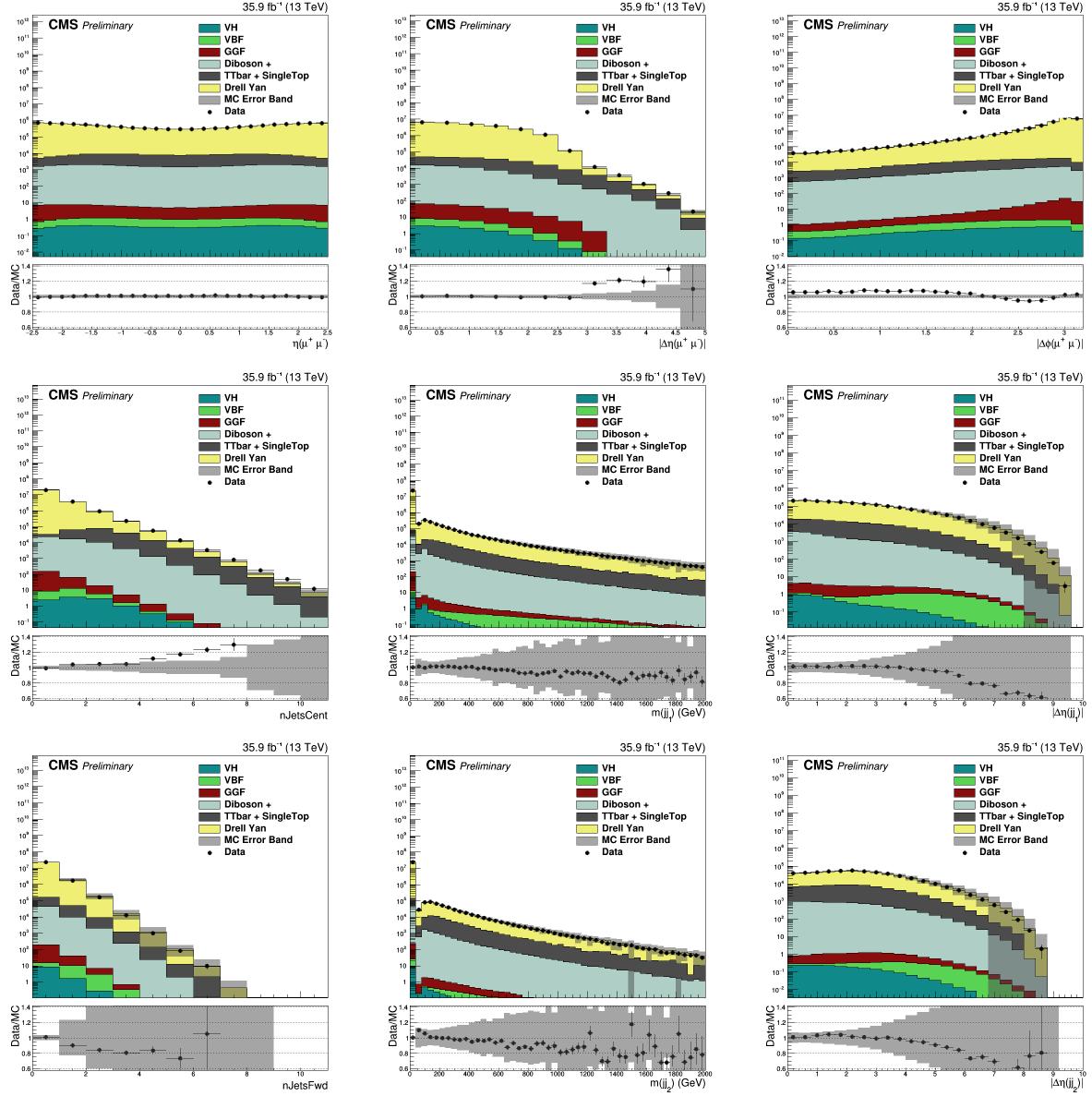


Figure 7-7. Validation of BDT input variables in inclusive events.

The Data and MC agree well. The most important variables, the BDT score, muon η , and the dimuon mass both show agreement within 10%. Figures 7-9 and 7-10 show the data/MC agreement for c14, the most sensitive, most VBF-like category. The BDT distribution in quantile is presented in Fig. 7-12 with the systematic uncertainties given by the JES and PU. The discrepancy from -0.6 to -0.4 in the BDT score is due to the dimuon p_t spectrum in Drell Yan. The data/MC comparison of the bdt with (left) and without (right) reweighting DY

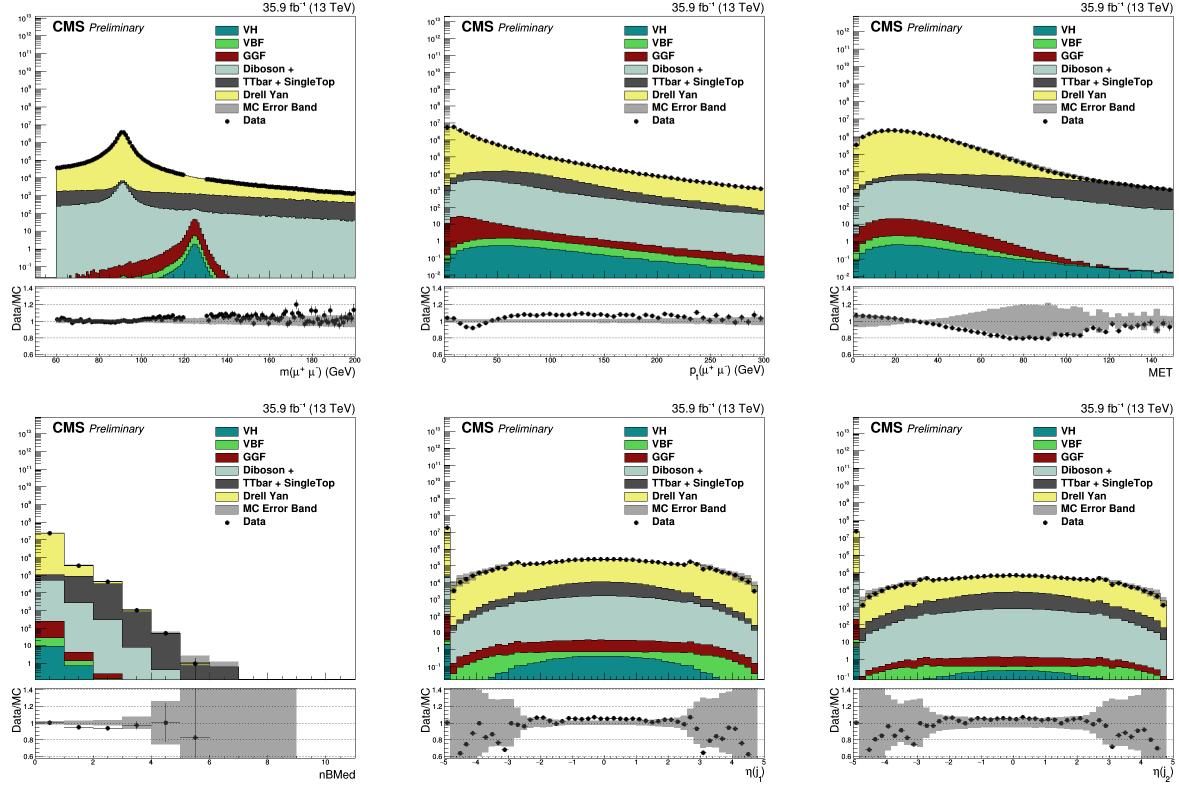


Figure 7-8. Data/MC agreement for the inclusive events.

dimuon p_t to data are shown in Fig. 7-11. Note also the systematic 10% disagreement in BDT score from -1 to -0.6 in Figure 7-11 where $t\bar{t}$ -like events dominate. This region of the BDT score is defined by nBMed greater than 0 and MET greater than 40. The 15% disagreement in MET around 50 GeV is the largest contributor. The 10% uncertainty on the $t\bar{t}$ cross section may play a role as well. Note that the disagreement takes place where the BDT score is less than -0.4 defining the least sensitive category. Insofar, this Data/MC discrepancy does not affect the expected limits. Beyond the mismatch in $t\bar{t}$ -like events, there is a notable Data/MC mismatch with jet $|\eta| > 2.4$, but this is expected as these are calorimeter only jets. The disagreement in the forward jets affects high BDT score, but the discrepancy is covered by the jet energy scale and resolution uncertainties as shown in Fig. 7-12. These uncertainties also cover the discrepancy in MET.

Lastly, it is important that the BDT score is mass independent. One way to show this is to demonstrate that the classifier score remains the same for signal MC with different

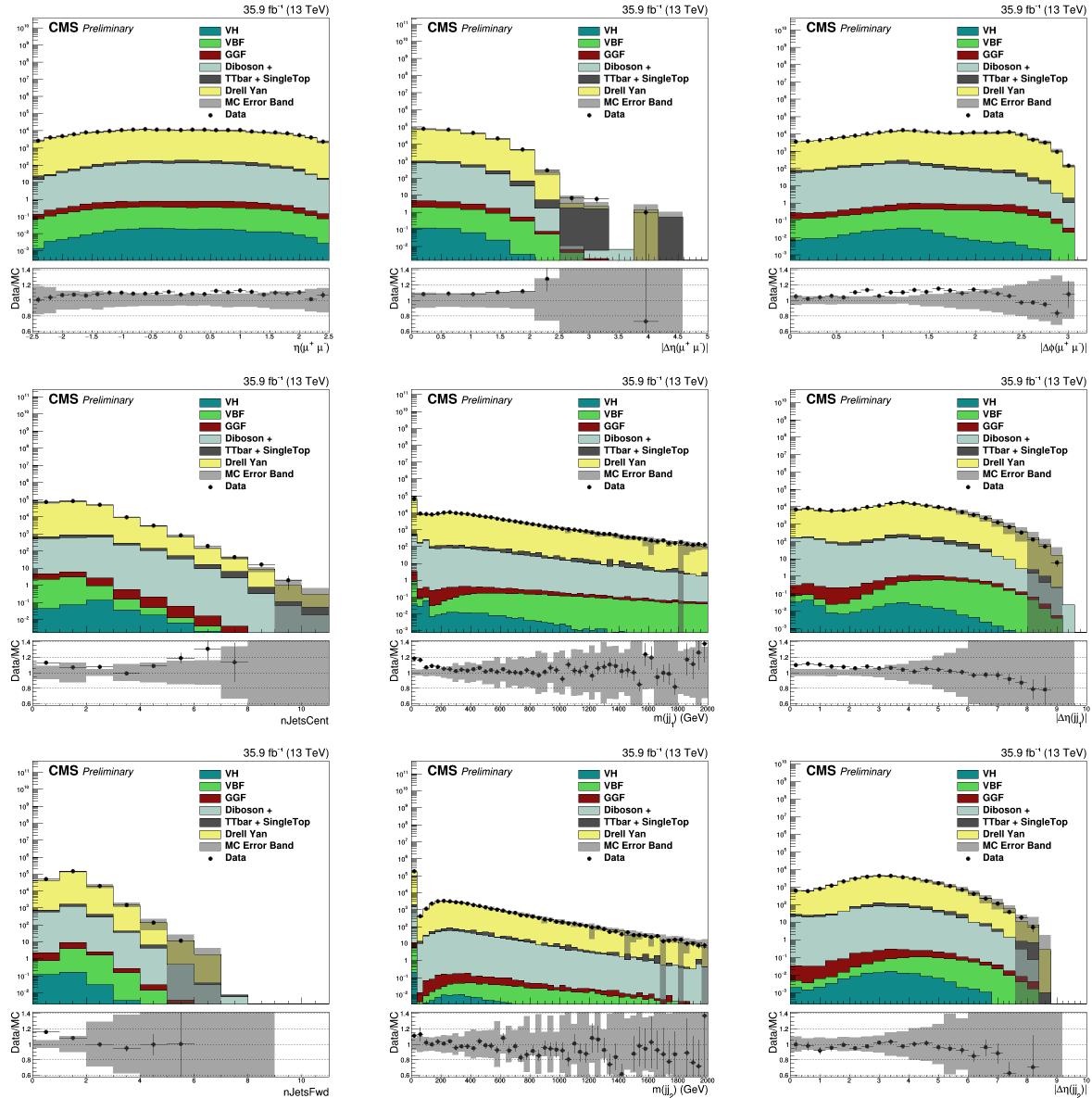


Figure 7-9. Data/MC agreement for the most sensitive category.

values of m_H . If the BDT score does not change as the mass changes, then it is clearly mass independent. As shown in Figure 7-13, the BDT score does not change significantly with the dimuon mass.

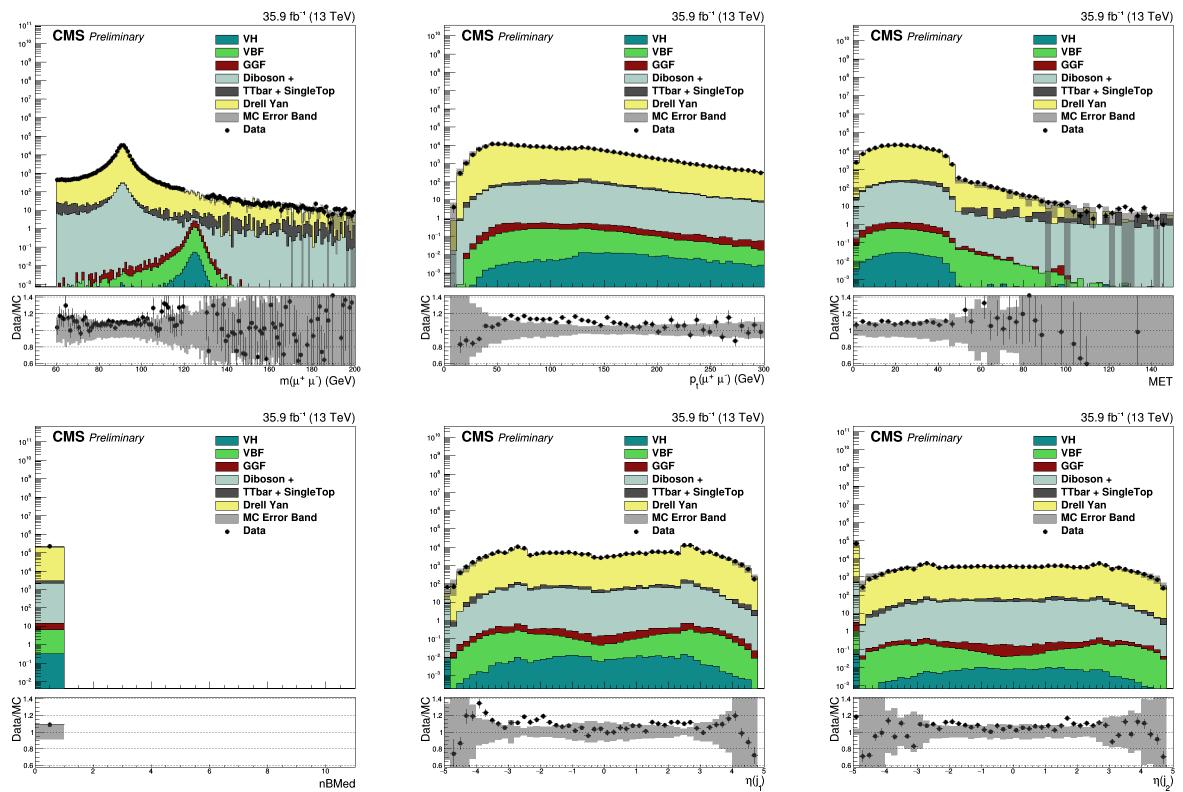


Figure 7-10. Data/MC agreement for the most sensitive category.

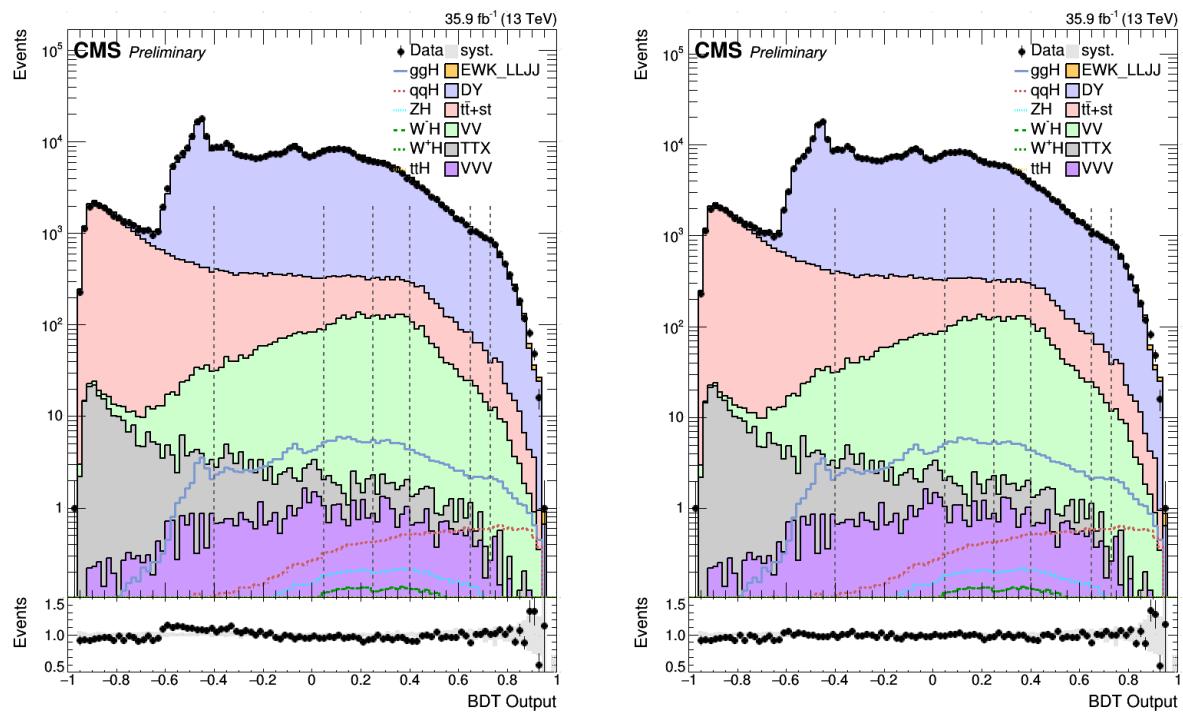


Figure 7-11. The BDT distribution in data and MC. Right plot has DY MC with p_t corrections to improve Data/MC agreement.

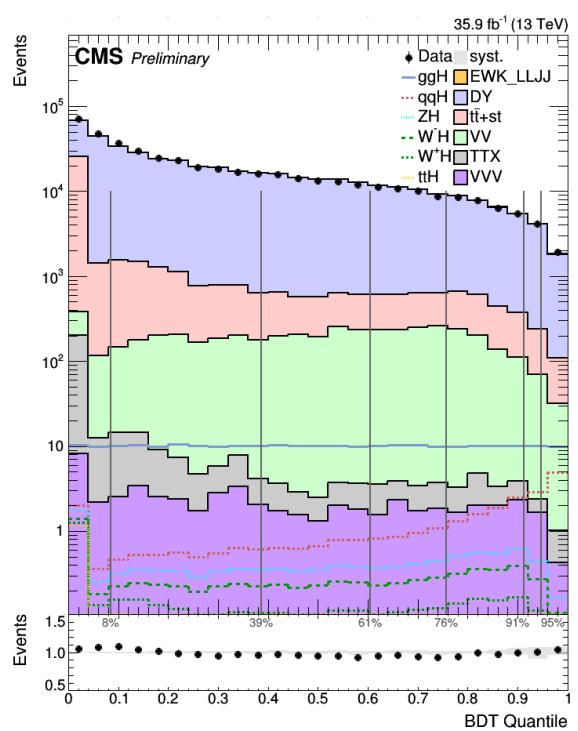


Figure 7-12. The BDT distribution in quantile in data and MC. The systematic uncertainty is given by the JES and PU.

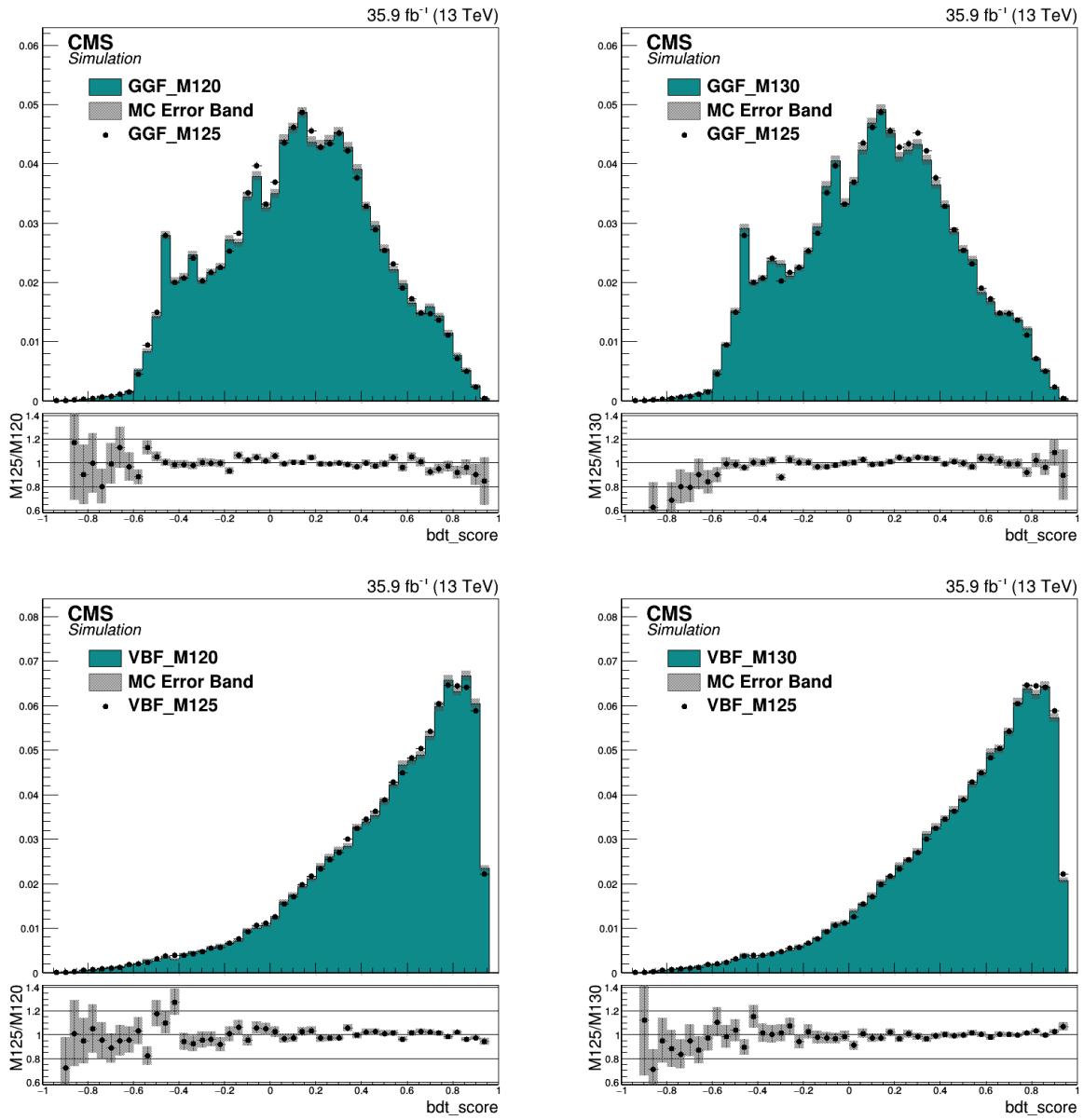


Figure 7-13. Plots showing the agreement between $m_H = 125 \text{ GeV}$ vs. 120 GeV and $m_H = 125 \text{ GeV}$ vs. 130 GeV for ggH and VBF.

7.2.6 Signal And Background

This section presents some plots and tables of interest for the signal and background.

Table 7-4 breaks down the signal yields by category and process. Then there are some plots showing the kinematic distributions and object counts for the signal. The distributions are presented for the inclusive set of events starting with Figure 7-14 and then for the most sensitive category (c14) starting with Figure 7-17. After looking at the signal, Fig. 7-20 presents the background shapes in data for each category and compares each to the inclusive shape.

Table 7-4. The total signal in each category broken down by process.

Category	All Signal	GGF	VBF	VH	t <bar>H</bar>
Inclusive	253	224	18.0	9.213	1.94
c0	21.2	18.7	0.397	0.790	1.26
c1	22.3	21.1	0.504	0.652	0.050
c2	41.1	38.9	0.829	1.16	0.206
c3	12.7	12.0	0.226	0.356	0.138
c4	11.8	10.8	0.510	0.474	0.012
c5	29.2	27.1	0.961	1.03	0.051
c6	14.5	13.7	0.337	0.436	0.035
c7	5.2	4.40	0.443	0.313	0.006
c8	20.3	18.4	1.07	0.790	0.027
c9	13.1	12.2	0.464	0.405	0.025
c10	5.8	4.49	0.906	0.378	0.008
c11	20.3	16.6	2.56	1.08	0.043
c12	13.7	11.8	1.20	0.667	0.035
c13	8.6	6.27	1.90	0.435	0.019
c14	13.7	7.66	5.72	0.252	0.027

Comparing the signal kinematics in Figures 7-14 through 7-16 and 7-17 through 7-19 notice that a high BDT score, exemplified by c14, removes tt}-like events by requiring 0 btags and low MET. Aside from removing tt}, high BDT score also targets VBF-like events by requiring large dijet mass and large separation in η between the jets. Also note that a high BDT score picks out high dimuon p_t since the signal processes generally have higher dimuon p_t than the background.

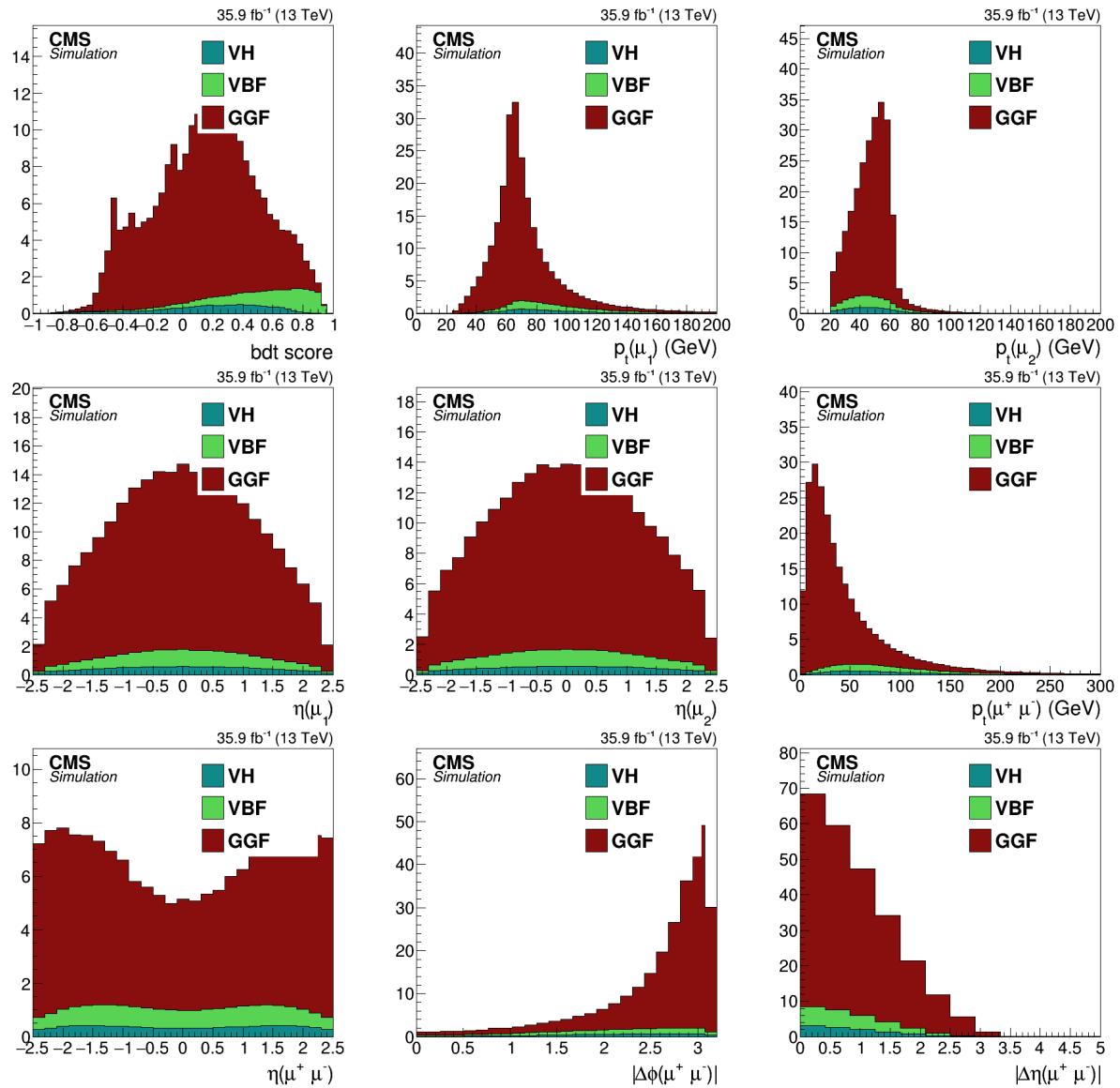


Figure 7-14. Kinematics and object counts for the signal in the inclusive set of events.

The factors influencing the shape of the background are the mass resolution (determined by muon η) and the BDT score. High resolution (low muon η) categories will have a narrower Z peak with more events proportionally on the high side of the Z peak—lower mass on the plots shown in Fig. 7-20. See c9 as an example. A high BDT score has a similar effect: since there are fewer $t\bar{t}$ like events, there are proportionally more events at lower mass than in the tail. See c14 for this effect. In Figure 7-20, the background shape for each category

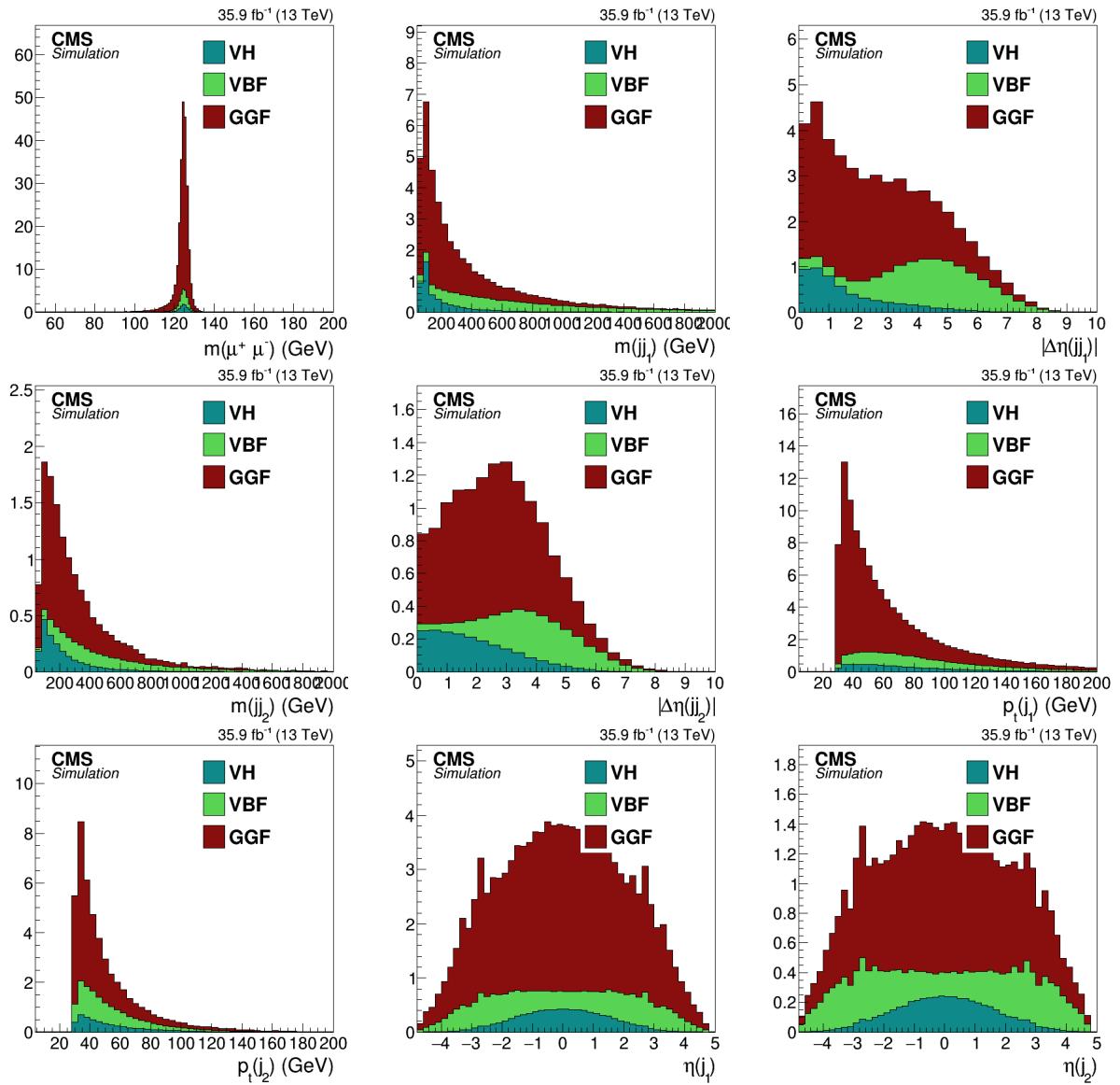


Figure 7-15. More kinematic and object count histograms for the signal in the inclusive set of events.

is normalized and compared to the inclusive set of events so that it's easy to compare the variation in shape between categories.

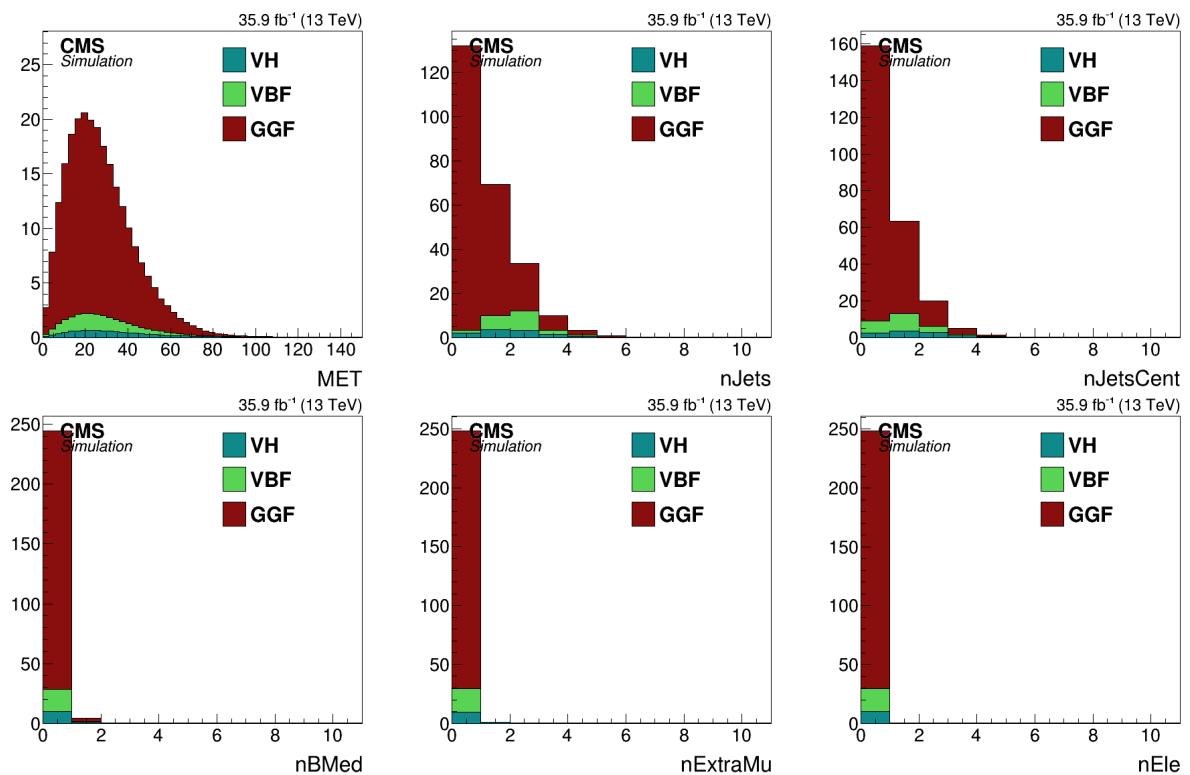


Figure 7-16. More kinematic and object count histograms for the signal in the inclusive set of events.

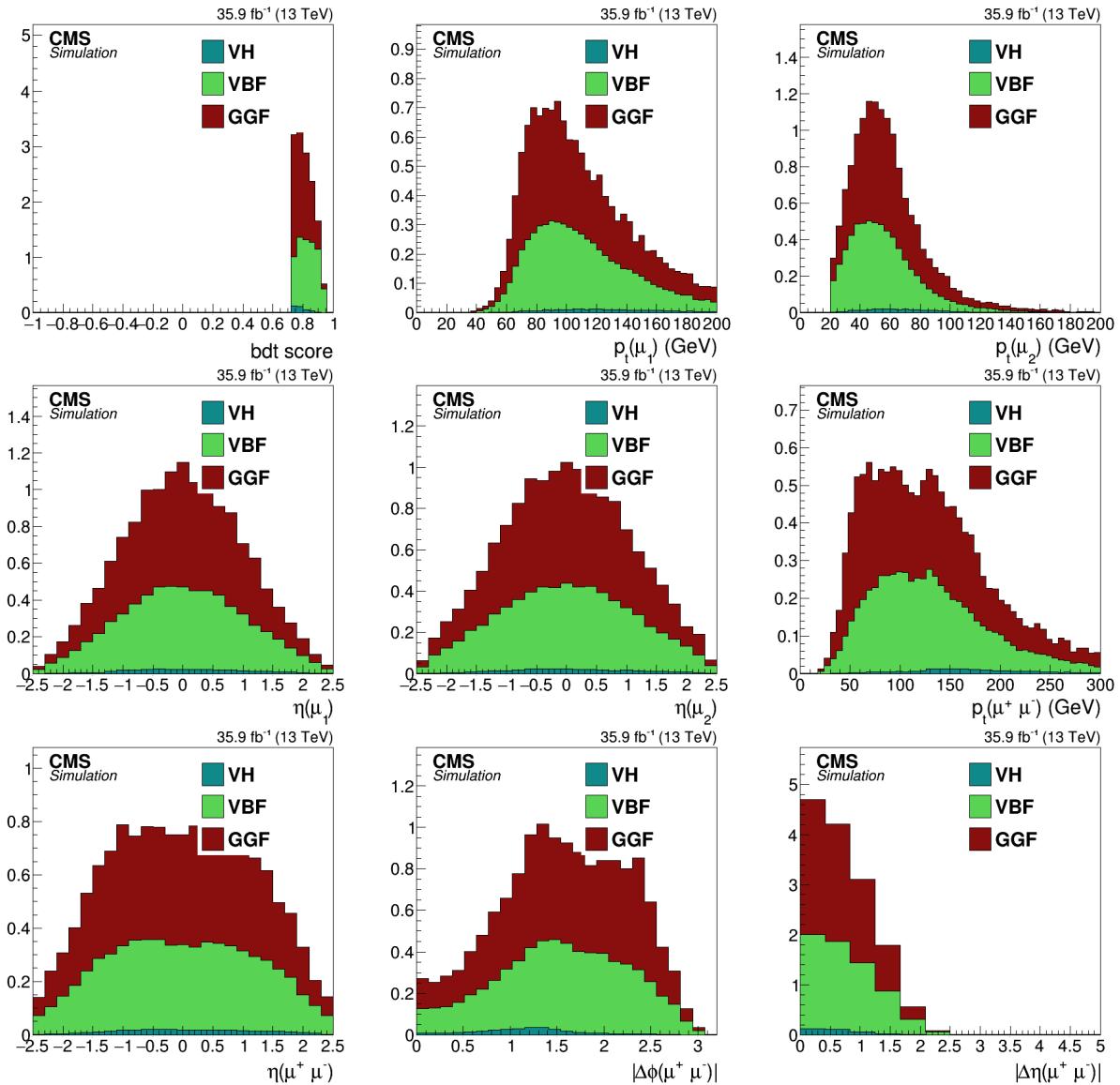


Figure 7-17. Kinematic and object count histograms for the signal in category c14. Note the high dimuon p_t targeting the signal in general.

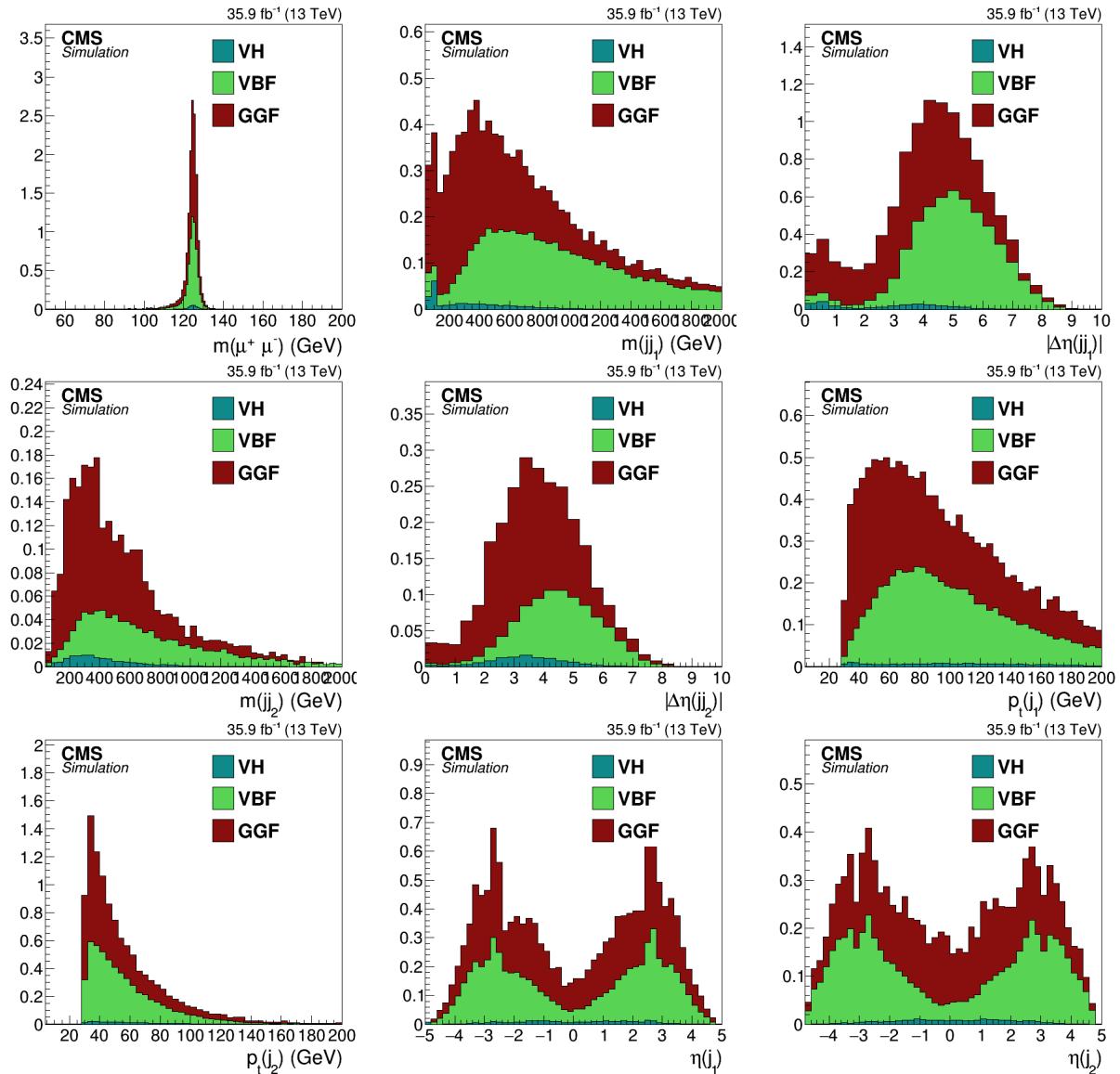


Figure 7-18. More kinematic and object count histograms for the signal in category c14. Note large dijet mass and the large separation in eta for the jets targeting VBF.

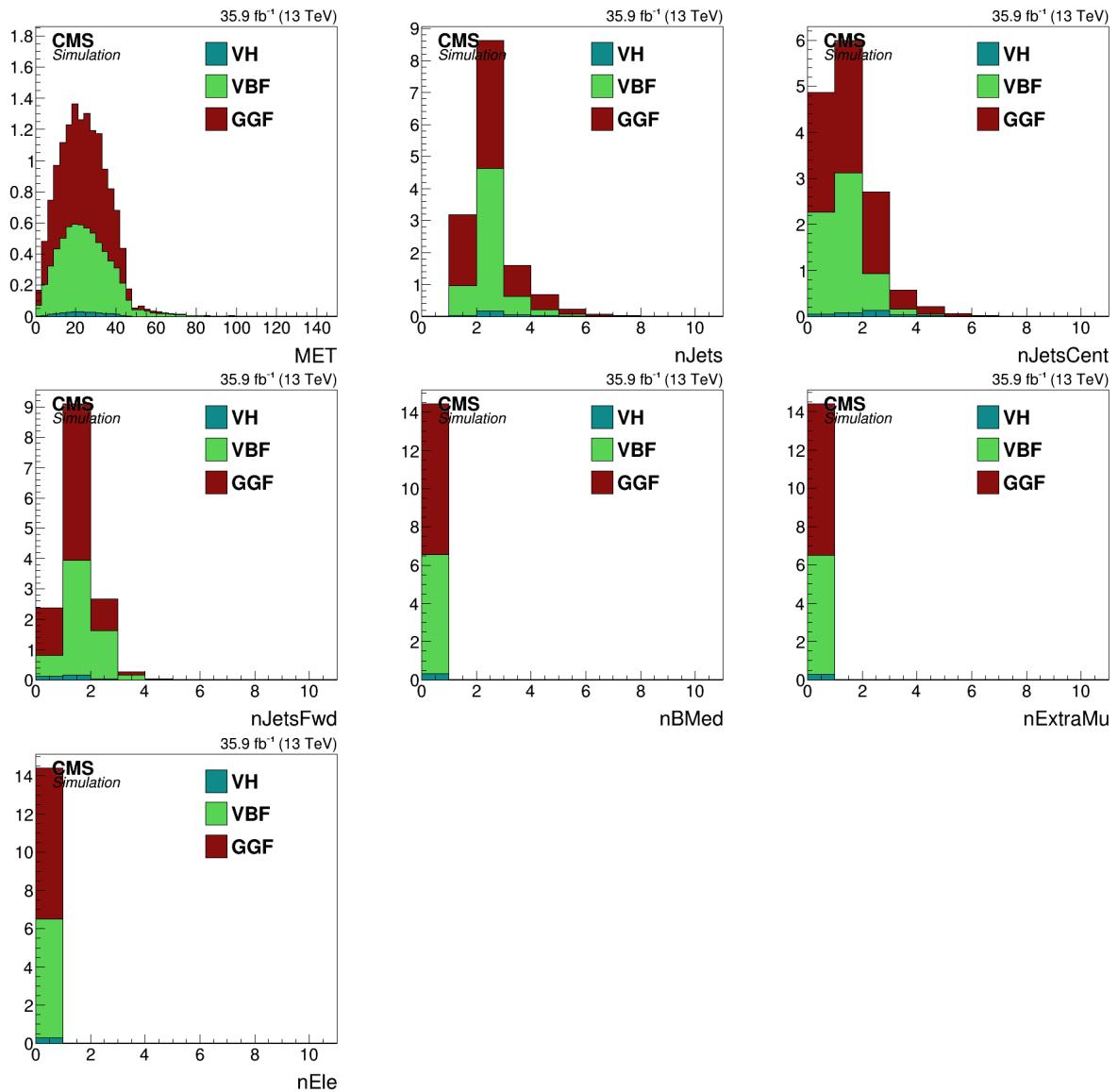


Figure 7-19. More kinematic and object count histograms for the signal in category c14. Note the low MET and 0 btags which exclude $t\bar{t}$.

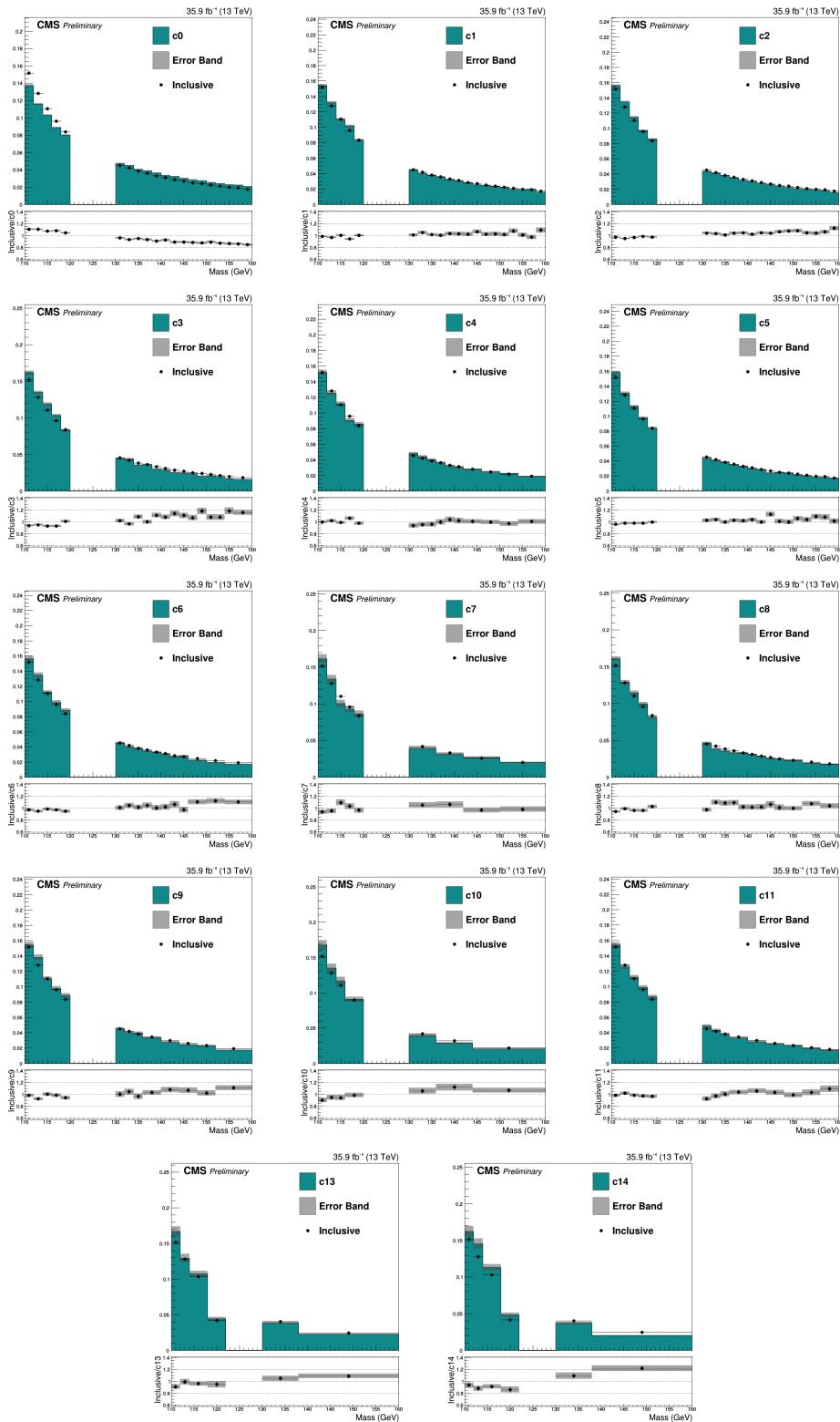


Figure 7-20. The difference in background shapes between the categories in data. The plots show the data with the signal region blinded. Each category's background shape is compared to the inclusive shape.

CHAPTER 8

SIGNAL AND BACKGROUND MODELS FOR THE $H \rightarrow \mu^+\mu^-$ SEARCH

This chapter presents the signal and background modeling for the $H \rightarrow \mu^+\mu^-$ search. The signal MC samples are fit to model the signal, and the data is fit to model the background. The systematic uncertainties and the evaluation of the background for bias are covered in Section 8.3.

8.1 Modeling The Signal

The signal in the $H \rightarrow \mu^+\mu^-$ search is a peak in the $m_{\mu\mu}$ spectrum, and it is modeled by fitting the signal MC after corrections, event selection, and categorization. The PDFs for the GGF, VBF, ZH, W^+H , W^-H , and $t\bar{t}H$ processes are fit separately, and the net PDF for the signal is the sum of the individual PDFs. The signal for each process in each category is fit with the same parametric form, a triple Gaussian¹ ,

$$S_{pc}(x, m_H, \theta_{pc}) = \sum_{g=1}^3 f_{pcg} \mathcal{N}(x, \mu_{pcg}, \sigma_{pcg}). \quad (8-1)$$

In 8–1, x runs along $m_{\mu\mu}$, p labels the signal process, c labels the category, g labels the Gaussian, and $\mathcal{N}(x, \mu, \sigma)$ is a Gaussian with mean μ and width σ . One Gaussian picks up the peak at the center, another wider Gaussian fits the remaining width at the core, and the final Gaussian accounts for the slightly higher probability in the low mass tail. Energy loss effects like Final State Radiation (FSR) and brehmsstrahlung increase the probability to measure a lower than average Higgs mass.

The PDFs for the signal in each category are fit at $m_H = 120, 125$, and 130 GeV, using the signal MC at those masses. To get the signal PDFs for any m_H value in the [120, 130] GeV search window, the fit parameters are interpolated (linearly) between the best fit values at 120, 125, and 130 GeV. Figure 8–1 shows an example of a fit on the left, its interpolation in the center, and the efficiency times acceptance on the right.

¹ Three of the process, category fits have very low stats and use only two Gaussians.

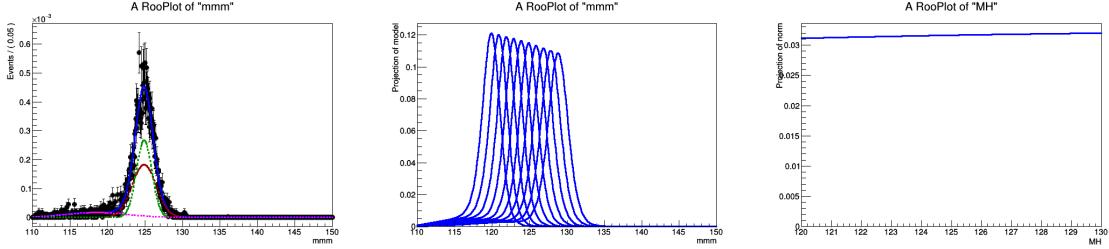


Figure 8-1. An example Higgs signal fit for a specific production process and category at a mass of 125 GeV (left), its interpolation to lower and higher mass (center), and its efficiency times acceptance vs m_H (right).

The PDFs describe the shape of the signal distribution, but not the total number of signal events. The models describing the signal yields are determined by normalizing the PDFs to the expected SM yields.

$$\text{Yield}_{\text{pc}} = \mathcal{L} \sigma(\text{pp} \rightarrow \text{H}) \mathcal{B}(\text{H} \rightarrow \mu^+ \mu^-) \varepsilon_{\text{pc}} A_{\text{pc}} \quad (8-2)$$

$\mathcal{L}\sigma(\text{pp} \rightarrow \text{H})\mathcal{B}(\text{H} \rightarrow \mu^+ \mu^-)$ provides the total number of $\text{H} \rightarrow \mu^+ \mu^-$ events expected by the SM, and the efficiency times acceptance, $\varepsilon_{\text{pc}} A_{\text{pc}}$, provides the fraction expected for the given process in the given category,

$$\varepsilon_{\text{pc}} A_{\text{pc}} = \frac{1}{N} \sum_{i \in p} w_i \cdot r_i \cdot s f_i \cdot I\{i \in c\}. \quad (8-3)$$

The sum runs over all MC events for the signal process. The factors w_i , r_i , and $s f_i$ are the MC weight, the PU reweighting factor, and the scale factor for the event, respectively. The indicator function, $I\{i \in c\}$, is 0 if the event i is not in the category and 1 if it is. The normalization, N , is given by the sum of MC weights over all the signal processes, $\sum_i w_i$. The cross section, $\sigma(\text{pp} \rightarrow \text{H})$, and branching fraction, $\mathcal{B}(\text{H} \rightarrow \mu^+ \mu^-)$, are taken from the Yellow Report 4, and provided centrally by combine.

The left side of Figure 8-2 shows the sum of the signal models over all processes and categories. The right side of the figure shows the net signal for one of the best mass resolution categories. The plots are divided by the number of expected events, $\mathcal{L}\sigma\mathcal{B}$, to present the fraction of signal expected per GeV, rather than the net yield per GeV. The blue line represents

the sum of the fits and the points represent the MC predictions. Figure 8-3 shows the

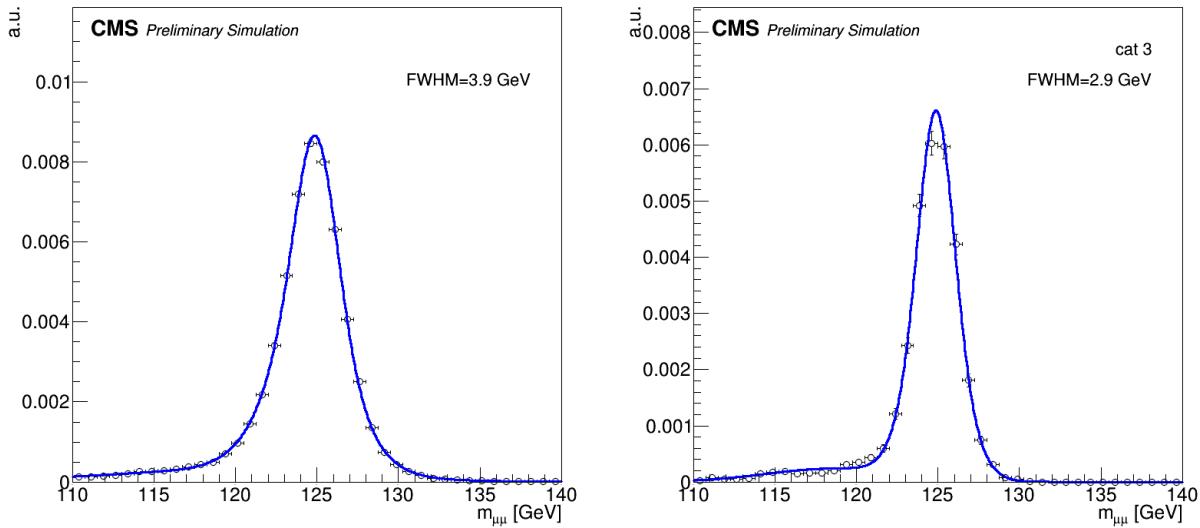


Figure 8-2. Signal model compared to MC predictions, summing up the contribution from all process and all categories (left), and for one of the categories with the best mass resolution.

composition of the signal model, the mass resolution, and the S/(S+B) in each category.

Figure 8-4 shows the efficiency times acceptance of the selection – prior to categorization – for the total signal and for each signal process.

CMS Preliminary Simulation

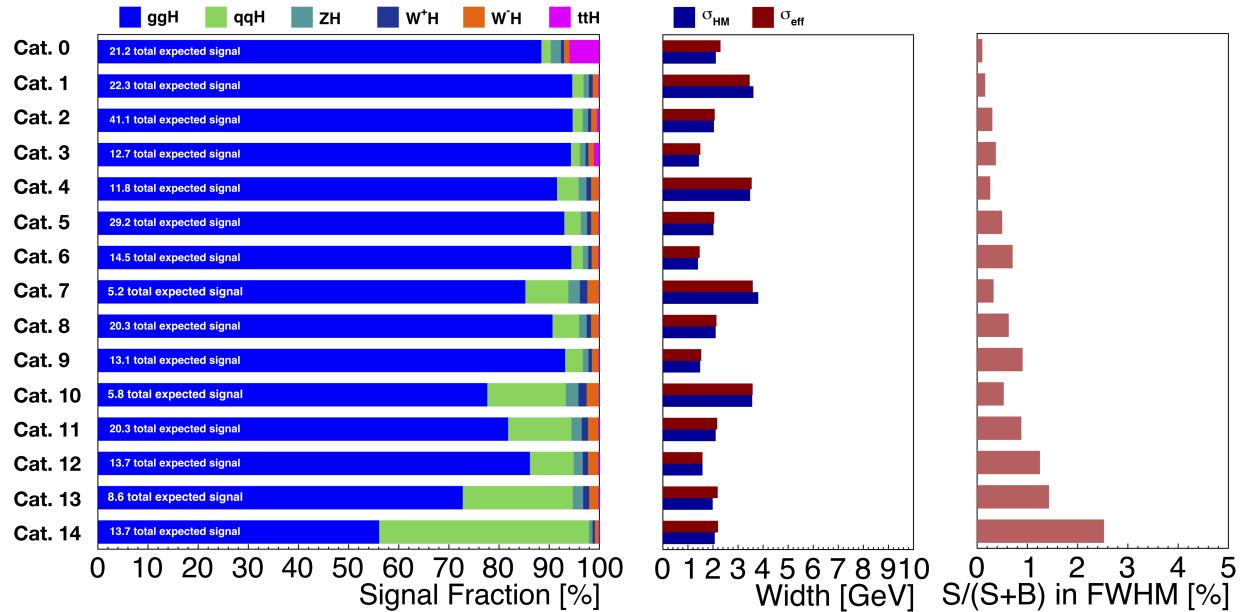


Figure 8-3. The signal composition, the mass resolution in Half Width Half Max (σ_{HM}) and 68% probability (σ_{eff}), and the S/(S+B) of the different categories.

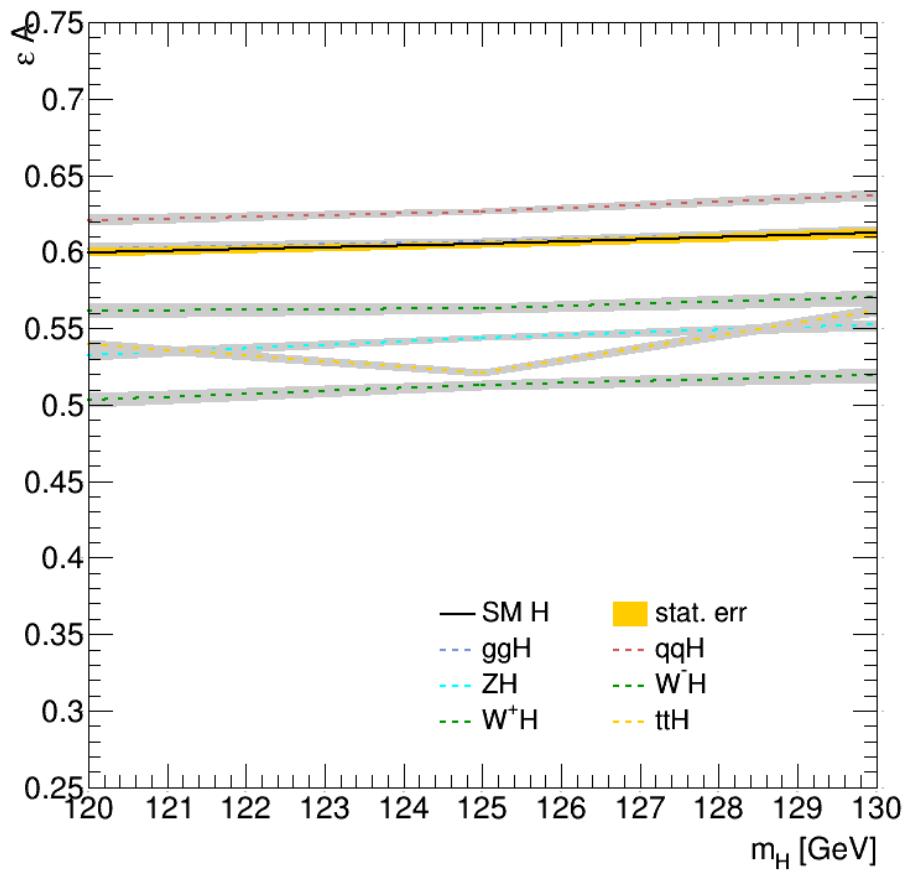


Figure 8-4. The efficiency times acceptance as function of m_H for the total signal and for each signal process.

8.2 Modeling The Background

The background for $H \rightarrow \mu^+ \mu^-$ is a smoothly falling spectrum in $m_{\mu\mu}$ dominated by Drell-Yan and $t\bar{t}$. The background in the signal region, 120 to 130 GeV, is estimated by fitting the data in the sidebands, $[110, 120] \cup (130, 150]$ GeV. While MC samples are available for the background, fitting the data provides a much better estimate. First, the data collected far exceeds the amount of MC available, and the larger statistics in data provide a reduced uncertainty on the estimated background. Second, the MC has many additional uncertainties like those from higher order QCD and EW contributions, resummation effects, and estimates of the parton density function, among others. With a lower the uncertainty on the estimated background, it's easier to tell a signal peak from a background fluctuation.

In order to find an accurate background fit with the lowest uncertainty, many different models are studied. The models are organized into three classes. The first class is physically motivated, using analytic functions based upon the Breit-Wigner for the Z peak. The functions modify the Breit-Wigner to control the decay at large mass where $t\bar{t}$ kicks in.

$$\text{BWZ: } B(x) = \frac{e^{ax}\sigma_z}{(x - \mu_z)^2 + (\frac{\sigma_z}{2})^2} \quad (8-4)$$

$$\text{BWZRedux: } B(x) = \frac{e^{a_2 x + a_3 x^2}}{(x - \mu_z)^{a_1} + (\frac{2.5}{2})^{a_1}} \quad (8-5)$$

$$\text{BWZGamma: } B(x) = f \frac{e^{ax}\sigma_z}{(x - \mu_z)^2 + (\frac{\sigma_z}{2})^2} + (1 - f) \frac{e^{ax}}{x^2} \quad (8-6)$$

These models have been validated by fitting the FEWZ (NNLO QCD) generated mass shapes and by fitting the data in the sidebands. The second class of models are the FEWZ shapes of the DY spectrum for the fully inclusive spectrum (NNLO), DY with one jet (NLO), and DY with two jets (LO). The FEWZ shapes are implemented as splines of the FEWZ histograms modulated by Bernstein polynomials in the same way as Equation 8-9. The polynomials multiplying the theoretical FEWZ shapes account for resolution smearing effects and the contributions from backgrounds besides DY. The third class are the general purpose functions,

which can fit any smoothly falling background with enough terms.

$$\text{Bernstein: } B(x) = \sum_{i=0}^n \alpha_i [\binom{n}{i} x^i (1-x)^{n-i}] \quad (8-7)$$

$$\text{SumExponentials: } B(x) = \sum_{i=1}^n \beta_i e^{\alpha_i x} \quad (8-8)$$

The analysis also uses a modified version of the BWZRedux (BWZR(x), Eq. 8-5) modulated by a Bernstein polynomial of degree n (BERN(x, n), Eq. 8-7, normalized to 1):

$$\text{BWZRedux} \cdot \text{Bern: } B(x) = \text{BWZR}(x) \cdot (1 + \text{BERN}(n, x)) \quad (8-9)$$

$$(8-10)$$

The modulated BWZRedux makes the model more flexible and helps to reduce any bias that may exist towards the unknown underlying true distribution.

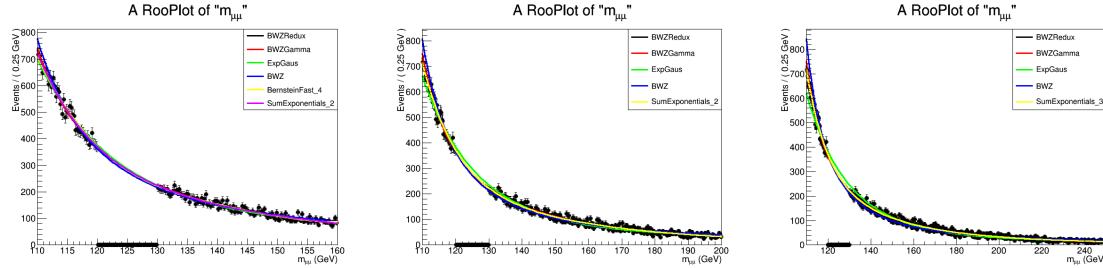


Figure 8-5. Example background-only fits to the dimuon mass spectrum in data for various background models and mass ranges.

For each category, the order of each FEWZ shape and of each general purpose function is selected using the F-Test method at 95% CL. The method is described in detail below.

- Null hypothesis: There is no difference between orders n and $n + 1$.
- If there is no difference, the likelihoods for the two models will show little difference in goodness of fit on the data.
- To reject this hypothesis at 95% confidence, the p-value for $2\Delta N_{LL}$ must be less than 5%, $p - \text{value}(\chi^2, \text{ndf}) < 5\%$.
- Perform the background only fit to the data for orders n and $n + 1$.
- Use $\chi^2 = -2\Delta \log \mathcal{L}$ using the asymptotic approximation of the Likelihood function.

- Compute the difference in number of degrees of freedom n.d.f. = $NDF_{n+1} - NDF_n$.
- The χ^2 variable is distributed as a χ^2 -distribution with n.d.f. degrees of freedom.
- Compute the χ^2 p-value.
- For a p-value less than 5%, reject n and move on to test n+1.
- For a p-value greater than 5%, we stop and select order n for this category, for this functional family.

Figure 8-6 shows an example of the F-Test for the Bernstein Polynomials and for the Sum of Exponentials.

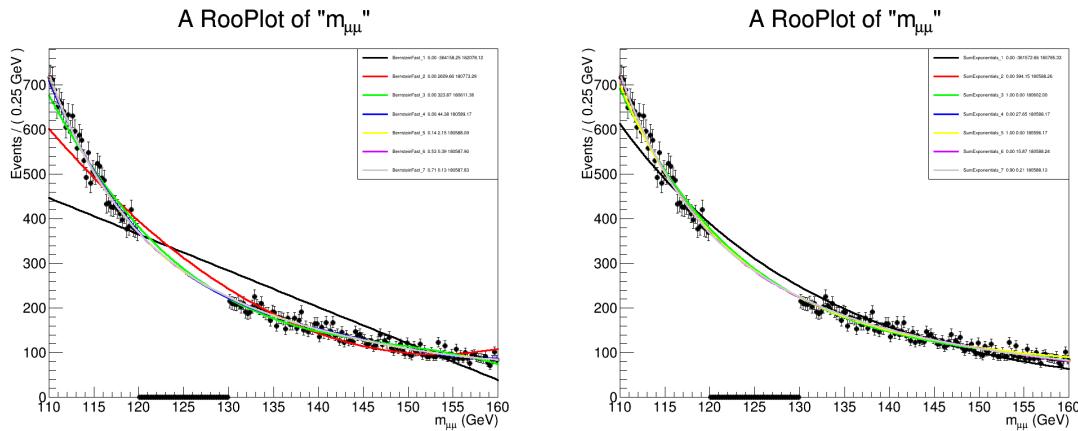


Figure 8-6. Examples of the F-Test for the Bernstein Polynomials (left) and the Sum of Exponentials (right).

8.3 Systematic Uncertainties

In order to report accurate results, it's essential to incorporate all known uncertainties affecting the signal and background models. These so-called "systematics" broaden the likelihood's PDF, worsen the limits, and increase the uncertainty of the measurement. The known contributions for the signal model are covered in the next section, and the background uncertainties are covered after.

8.3.1 Signal

The uncertainties on the signal model are broken down into shape, category migration, and rate uncertainties. Shape uncertainties affect the shape of the signal PDF. Category migration uncertainties quantify the migration of signal events from one category to another,

and rate uncertainties affect the total number of $H \rightarrow \mu^+\mu^-$ events expected. The WG1 uncertainties on GGF affect all three, and these are mentioned in the rate uncertainty section. A quantitative comparison of the different uncertainties is outlined in the impacts plot, Figure 8-8.

8.3.1.1 Shape uncertainties

The shape of the $m_{\mu\mu}$ signal peak is determined by the p_t , η , and ϕ measurements for the muons. The centrally provided Kalman Filter muon momentum corrections provide uncertainties on the muon momenta in terms of the scale and resolution. The muon η and ϕ measurements are precise enough to be neglected in comparison.

- **Muon Scale** The scale of the momentum affects the position of the $m_{\mu\mu}$ peak, and propagating this uncertainty to the mass shifts the signal peak a maximum of 0.05%. A single nuisance is used to quantify the scale uncertainty, correlated across the categories.
- **Muon Resolution** The resolution affects the width of the peak and propagating this uncertainty to the mass affects the width of the signal peak up to 10%. Again, a single nuisance is used correlated across the categories.

8.3.1.2 Category migration uncertainties

Moving the values of certain factors in the analysis may cause a signal event to migrate from one category to another affecting the yields in the categories. The uncertainties in these factors therefore affect the uncertainty of the yields across the categories. The major migration uncertainties are discussed below.

- **Jet Energy Scale** After applying the jet energy corrections *Summer16_23Sep2016A/V4*, the energy scale of the jet is varied as provided by the JetMET group. The variation shifts the jet energy and propagates the shift to the MET for the event. The jet energy variations lead to variations in category yields, especially in the high sensitivity VBF-like categories. The jet energy scale is the dominant experimental uncertainty with up to 6% variation in the yields and the largest impact on the signal strength measurement. The effect of the scale uncertainty can be seen in the error bands in Fig. 8-7, showing coverage for data/MC discrepancies in jet eta and the number of jets.
- **Jet Energy Resolution** The uncertainty in jet energy resolution also causes category migrations. Upon widening the resolution, events migrate along the steeply falling p_t spectrum. More events move from the highly populated low p_t regime into the less

populated high p_t regime, causing events to move between categories. The resolution uncertainty leads to variations in the yields up to 3%.

- **Pile-up Re-weighting** The pile-up re-weighting procedure uses the minimum bias cross section to estimate the amount of pile-up in data. The amount of Pile-up has two major effects. First, it can reduce the efficiency of the muon selection by increasing the nearby hadronic activity, and, second, it may create random clusters of energy that become identified as jets. But thanks to the PU removal techniques in the jet reconstruction and muon isolation, the PU uncertainty only leads to 1% variations in the signal yield.
- **b-jet Efficiency** The uncertainty in the b-tagging efficiency also leads to yield variations, but due to the fact that b-jets are vetoed in the most sensitive categories to suppress the $t\bar{t}$, the uncertainty yields $\simeq 1\%$ variations.
- **b-jet Fake Rate** Same as above but correcting the efficiency of light-flavored jets to fake a b-jet ($\simeq 1\%$).
- **MC R/F Scale** Varying the renormalization and factorization scales up and down by a factor 2, yields up to $\simeq 6\%$ migration. The contribution to the total signal yield is described in the next section on rate uncertainties.
- **MC PDF** Parton distribution functions are varied using the NNPDF3.0 weights available in the production. These variations amount to $\simeq 2/3\%$ signal yield variation. The contribution to the total sample normalization is covered in the next section on rate uncertainties.
- **MC Tune** Tune uncertainty is derived using the centrally provided up and down MC variations of the MC tune for pythia8.
- **Parton Shower** Parton Shower uncertainty is derived using the centrally produced up and down MC variations of the pythia8 showering settings.

8.3.1.3 Rate uncertainties

Rate uncertainties affect the total number of SM $H \rightarrow \mu^+\mu^-$ events expected. The theoretical uncertainties include those on the branching fraction and the signal production cross section. The experimental uncertainties include those on the luminosity and the muon scale factors.

- $H \rightarrow \mu^+\mu^-$ **Branching Fraction** The branching fraction has an uncertainty of 1.7% and applies to all signal production.
- **R/F – Cross Section** The GGH cross section has an uncertainty of 3.9% due to renormalization and factorization (R/F), VBF 0.4%, ZH 3.8%, WH 0.6%, and $t\bar{t}H$ -9.9% to $+6.8\%$.

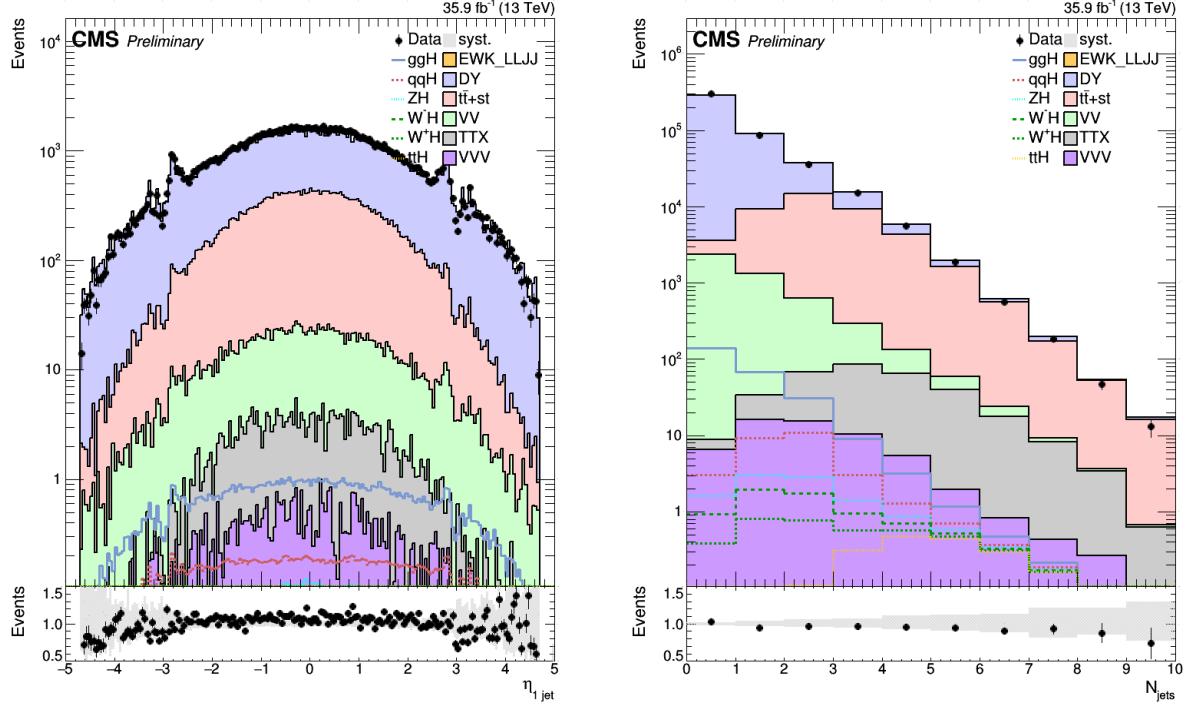


Figure 8-7. Pseudorapidity of the leading jet, and number of associated jets. The uncertainty bands show jet energy scale uncertainty.

- **PDF – Cross Section** The parton distribution function contributes an uncertainty of 3.2% to the GGH cross section, 2.1% to VBF, 1.6% to ZH, 1.9% to WH, and 3.7% to $t\bar{t}H$.
- **WG1 uncertainties** Additional uncertainties on GGF are applied using the WG1 recommendations. These apply uncertainties ontop of the R/F and PDF uncertainties accounting for different effects like the uncertainty in yield over jet p_t , the number of jets, and other factors. The WG1 corrections affect the shape, category migration, and rate. These are marked THU in the impact assessment of Figure 8-8.
- **Luminosity** The luminosity measurement has an uncertainty of 2.5%.
- **Muon Scale Factors** The muon scale factors provided by the POG have an uncertainty of 0.5% for the trigger, 1.0% for the muon ID, and 0.5% for the isolation. Uncertainties on the different runs are taken to be correlated.

8.3.1.4 Impact plots

The influence of the different uncertainties on the signal strength measurement are shown in Fig. 8-8. The uncertainties are sorted in decreasing order of their impact, $\Delta\hat{\mu}$, induced by a $\pm 1\sigma$ variation of the systematic under evaluation.

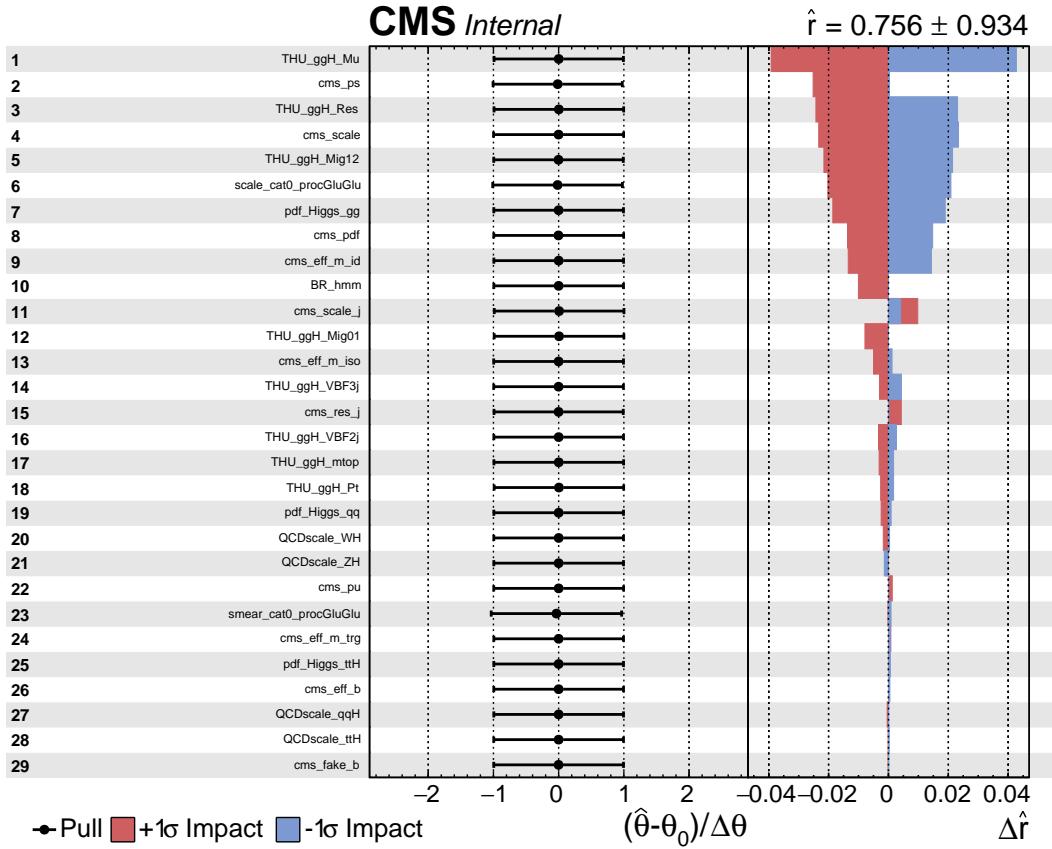


Figure 8-8. Impact plots derived from the Asimov dataset.

8.3.2 Background

The background is modeled by fitting the data, and the likelihood from the fit naturally accounts for the uncertainties in the fit parameters leaving fewer systematics to incorporate as additional constraints. Unfortunately, when using an arbitrary model to estimate the background it is unknown whether the form of the model can accurately fit the underlying true distribution in each of the categories. If the model cannot accurately fit the underlying true distribution, then the signal measured will differ from the truth and the model is "biased". In order to report accurate results it is therefore important to assess the existence of any bias, and account for it appropriately.

To quantify the bias of a particular model, the following process is used. The analysis chooses a set of reasonable models $\beta = B_1, B_2, B_3, \dots, B_N$. Then the bias of model B_i towards

hypothesis B_j is evaluated by assuming that B_j is the underlying true model. B_j is fit to the sidebands in data, and the best fit for B_j is used along with the signal model to form the $S+B_j$ PDF, emulating a world where $S+B_j$ represents the truth. This PDF is used to generate pseudodata at the given integrated luminosity with a signal strength (μ , sometimes called r) of 1 representing the SM. The pseudodata is fit in the sidebands with B_i and the best fit of the signal strength ($\hat{\mu}_{ij}$) is determined. With a true μ of one the expected $\hat{\mu}_{ij}$ should be 1. To get a distribution for $\hat{\mu}_{ij}$, the pseudodata is generated by $S+B_j$ and fit by $S+B_i$ to extract $\hat{\mu}_{ij}$. If the median $\hat{\mu}_{ij}$ is not 1 then B_i is biased towards B_j . As such, the bias is defined to be,

$$b_{ij} = \text{Median} \left[\frac{\hat{\mu}_{ij} - \mu}{\sigma_{\mu,ij}} \right] \quad (8-11)$$

with $\mu = 1$. B_i is unbiased towards B_j if b_{ij} is close to zero. The median is taken over 2500 trials yielding an uncertainty in b_{ij} of about 2%. The bias for all $B_i \in \beta$ is evaluated for all $B_j \in \beta$, providing the NxN matrix b .

If a bias is found, the background model tends to under or over estimate the background and measure too much or too little signal. When this is the case, a "spurious signal" is added to the background model to compensate. The spurious signal has the same shape as the signal model, with a normalization constrained by a Gaussian with mean 0 and $\sigma = b_{ij}\sigma_\mu N$, where N is the expected number of signal events in the category. The spurious signal allows the background more freedom in the signal region, fixing the tendency to mismeasure the signal. On the other hand, this added freedom in the signal region increases the background uncertainty there, reducing the sensitivity.

A bias less than 0.20 has a negligible effect on the upper limits (< 2%). In this case, the spurious signal can usually be neglected. With this information, the analysis attempts to find fits for each category such that the maximum likelihood estimate of μ is unbiased ($b_{ij} < 0.20$) with respect to all $B_j \in \beta$. The bias matrices for $m_H = 125\text{GeV}$ are shown in Figures 8-9 and 8-10. The matrices present the bias towards $\hat{\mu}$ for the category's likelihood alone and zero spurious signal.

The set of reasonable functions used is $\beta = \{\text{BWZRedux} \cdot \text{poly}, \text{BWZGamma}, \text{SumExp}, \text{FEWZ} - \text{full} \cdot \text{poly}, \text{FEWZ} - 1 - \text{jet} \cdot \text{poly}, \text{FEWZ} - 2 - \text{jet} \cdot \text{poly}\}$. Polynomial orders 0-4 are examined for BWZRedux·poly, and the F-Test at 95% confidence is used to select the orders for SumExp and the FEWZ·poly models. The Bernstein polynomials proved biased towards all of the other functions, so they are considered unreasonable and excluded from β .

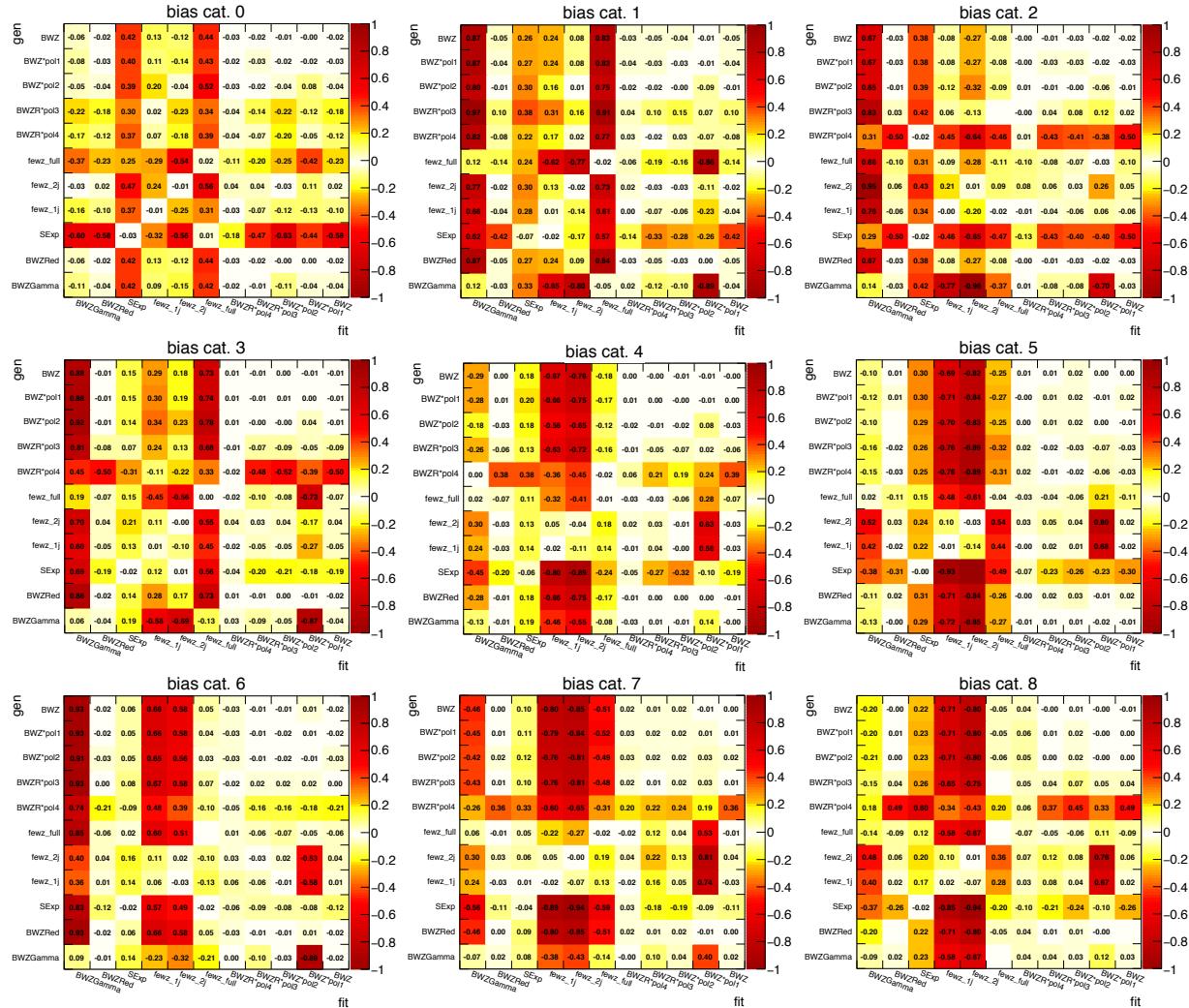


Figure 8-9. Bias results for categories 0-8 with each category's bias independent of the others. The results shown are for $m_H = 125$ GeV. The x-axis shows the fit model and the y-axis shows the pseudodata generating model.

The maximum observed bias for each category's least biased function is tabulated in

Table 8-1. The maximum bias is calculated over $m_H = 120, 125, 130$ GeV. These functions are

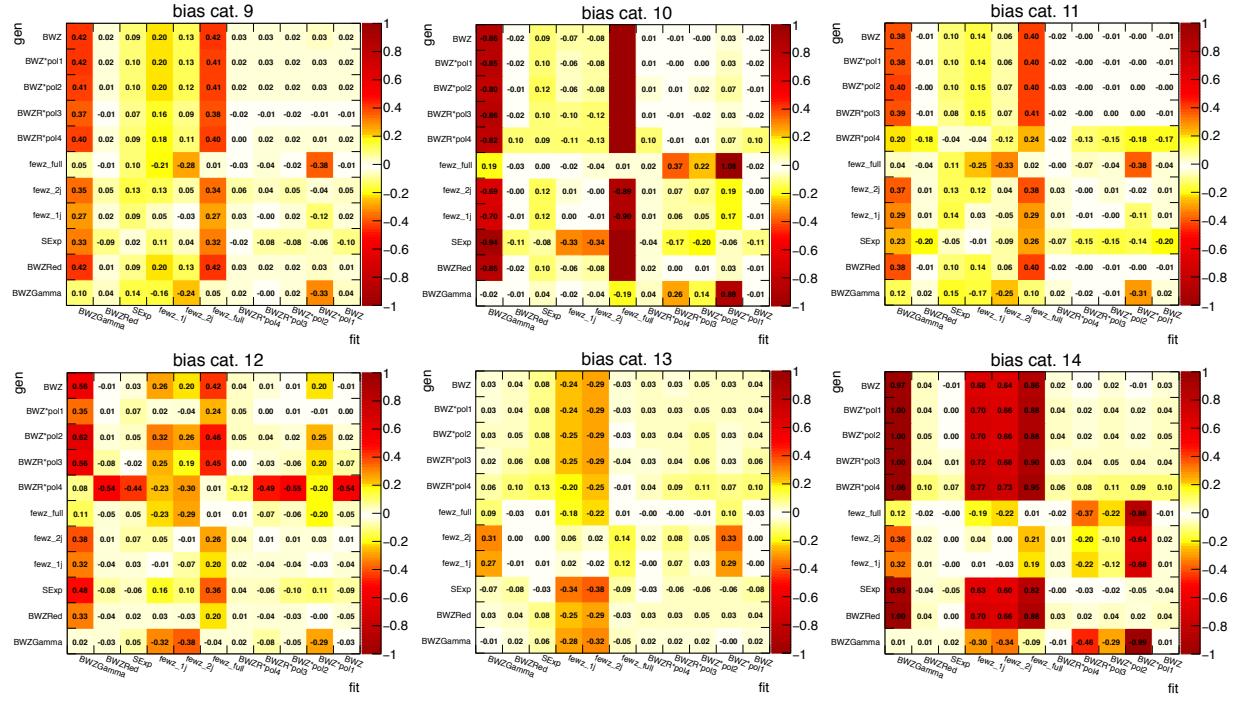


Figure 8-10. Bias results for categories 9-14 with each category's bias independent of the others.
The results shown are for $m_H = 125$ GeV. The x-axis shows the fit model and the y-axis shows the pseudodata generating model.

used for the final limit setting. If the statistics in a category are low enough, the uncertainty

Table 8-1. Maximum observed bias per category for least biased fit function. The maximum bias is over $m_H = 120, 125, 130$ and GeV.

Category	Best Model	Max Bias
cat0	BWZRedux · Bern4	18%
cat1	BWZRedux · Bern4	17%
cat2	BWZRedux · Bern4	21%
cat3	BWZRedux · Bern4	10%
cat4	BWZRedux · Bern4	12%
cat5	BWZRedux · Bern4	20%
cat6	BWZRedux · Bern4	6%
cat7	BWZRedux · Bern4	20%
cat8	BWZRedux · Bern4	7%
cat9	BWZRedux	9%
cat10	SumExp (n=2)	19%
cat11	BWZRedux	22%
cat12	BWZRedux · Bern4	13%
cat13	BWZRedux	10%
cat14	BWZRedux	10%

on the signal strength will be large compared to the signal strength. As a consequence, each b_{ij} ($\frac{\Delta\mu_{ij}}{\sigma_{\mu,ij}}$) can be made artificially low. It is therefore important to calculate the bias using the combined likelihood over all categories where $\hat{\mu}$ and σ_μ use the full statistics. The evaluations of Figure 8-11 estimate the bias on $\hat{\mu}$ using the combined likelihood over all categories with the fits of Figure 8-1 in each category. The final fits of 8-1 provide a net bias under 20% for m_H

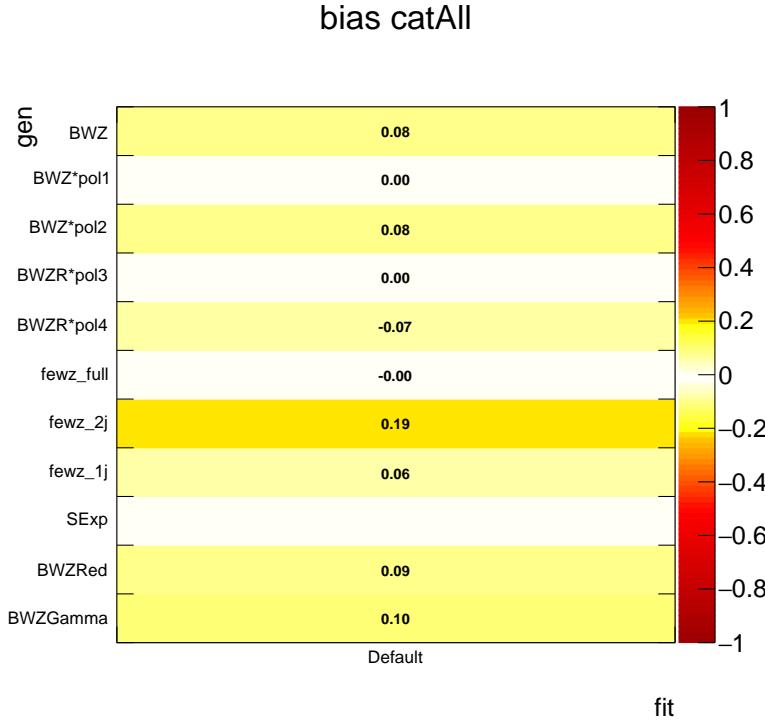


Figure 8-11. The plot shows the bias on $\hat{\mu}$ using the combined likelihood for all categories and the background fits of Figure 8-1. In this estimate, the same generating function is used for every category.

$= 120, 125,$ and 130 GeV . Moreover, including a spurious signal in each category affects the $m_H = 125 \text{ GeV}$ expected upper limit at 95% confidence by only 0.7% . This effect is considered negligible and no spurious signal is used in the results.

CHAPTER 9

RESULTS FOR THE $H \rightarrow \mu^+\mu^-$ SEARCH

With the categories finalized, the signal and background models pinned down, and the systematic uncertainties set, the results are calculated. The upper limit on the signal strength, the background only p-value, and the best fit of the signal strength are calculated for the 13 TeV data. These results are then combined with the 2012 results from 7 and 8 TeV data. The limits presented use the CLs method.

9.1 Results For 13 TeV Data

Figures 9-1 to 9-4 present the upper limits, background only p-value, and the best fit signal strength for 13 TeV data. For $m_H = 125$ GeV, the observed upper limit on the signal strength at 95% confidence is 2.64, while the expected (median) upper limit if the background only hypothesis is true is 2.08. For $m_H = 125$ GeV, the observed p-value on the background only hypothesis is 0.74σ , while the expected p-value on the background only if $H \rightarrow \mu^+\mu^-$ obeys the SM is 0.98σ . The best fit signal strength at $m_H = 125$ GeV is $0.7^{+1.1}_{-1.0}$. A weighted average of the S+B fits in each category is shown in Figure 9-5, and the individual fits per category are shown in Figures 9-6 and 9-6.

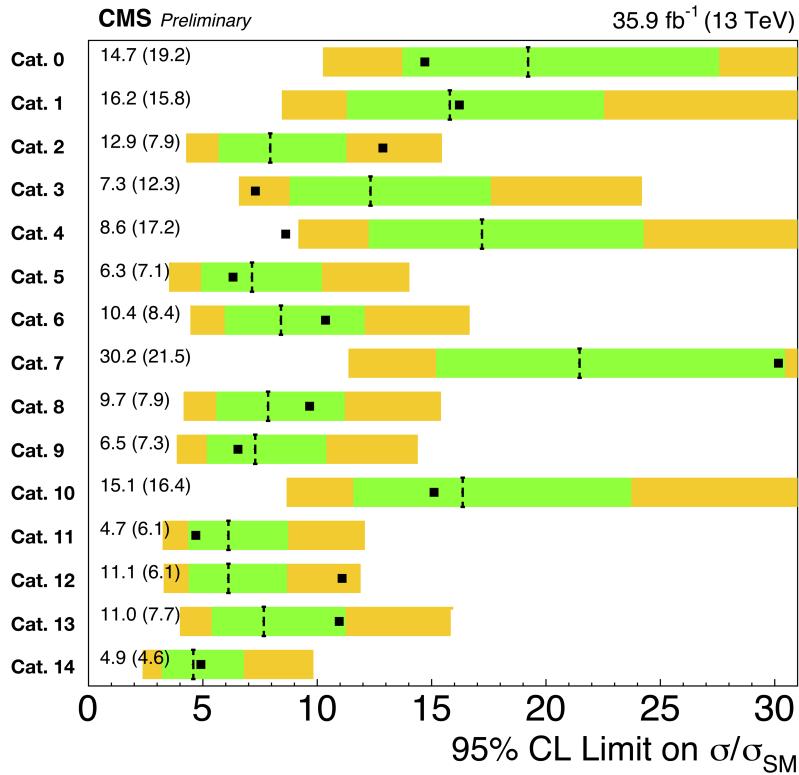


Figure 9-1. The upper limit on the signal strength at 95% confidence is presented for each category independent of the others. The results are for the 13 TeV data only.

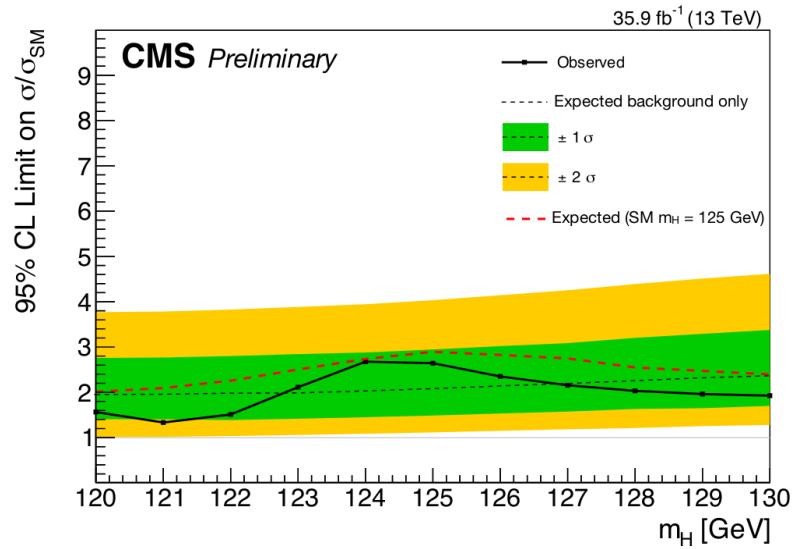


Figure 9-2. The net upper limit on the signal strength at 95% confidence is presented for 13 TeV data. The result uses the combined likelihood of all the categories. The red dashed line represents the expected (median) observation if the SM $m_H = 125 \text{ GeV}$ hypothesis is true. The black dashed line represents the expected observation if the background only hypothesis is true.

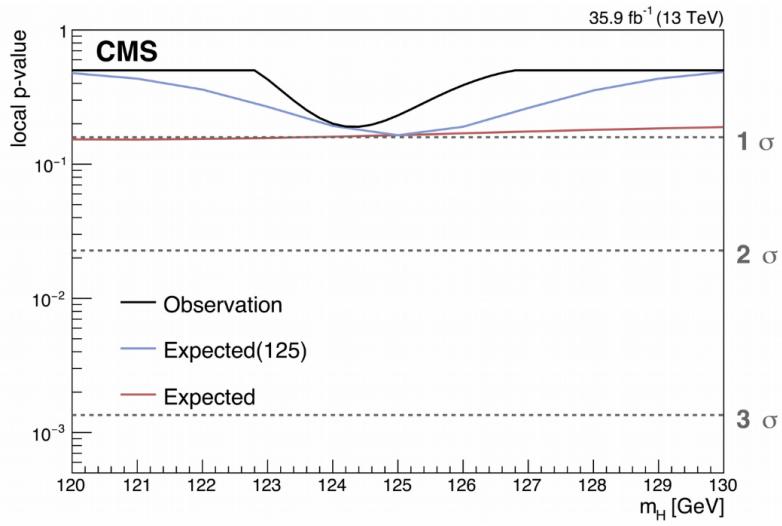


Figure 9-3. The p-value is presented for 13 TeV data. The black line is the observed p-value on the background only hypothesis. The blue line represents the expected (median) p-value on the background only if the SM $m_H = 125$ GeV hypothesis is true. The red line represents the expected p-value at $m_H = x$ GeV if the SM $m_H = x$ GeV hypothesis is true.

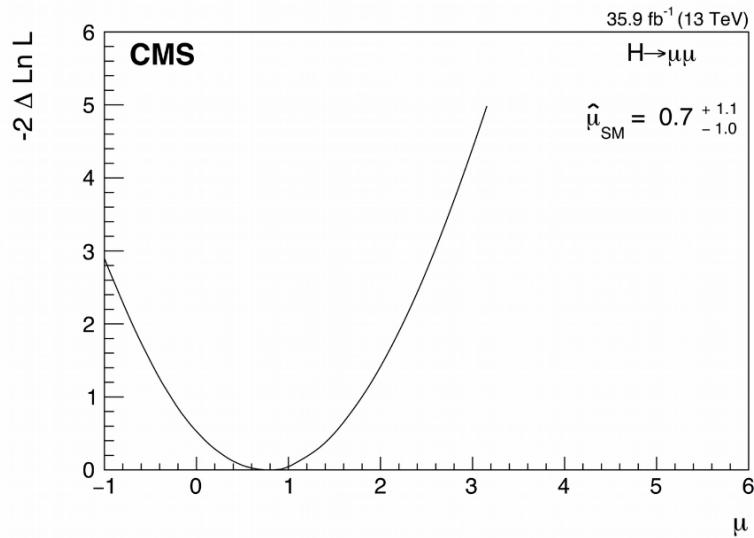


Figure 9-4. The negative log likelihood is plotted near the minimum for the signal strength μ , and the best fit, $\hat{\mu}$, is given.

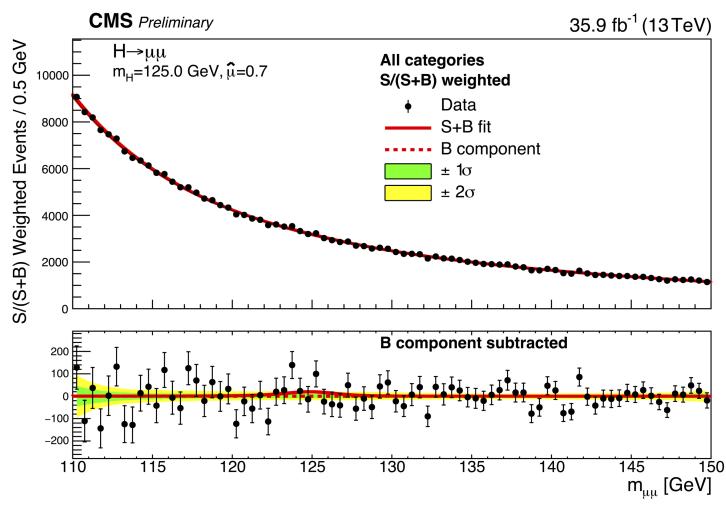


Figure 9-5. The weighted average of the signal + background fits over the categories. The fit from each category is weighted by $S/(S+B)$ to provide an idea of the overall shape while counting those categories with larger signal purity to a greater degree.

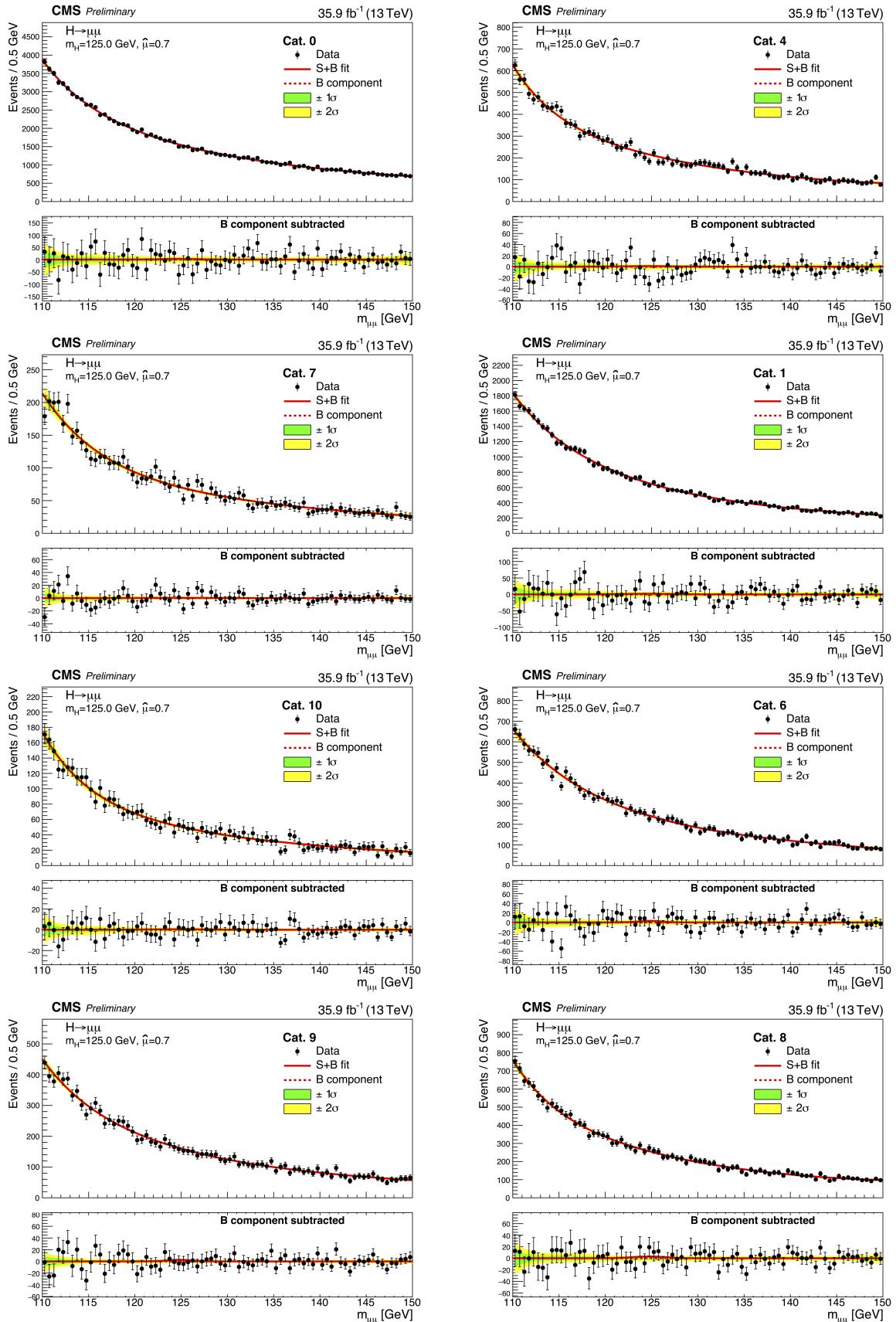


Figure 9-6. The final signal + background fits for the 13 TeV data, part one.

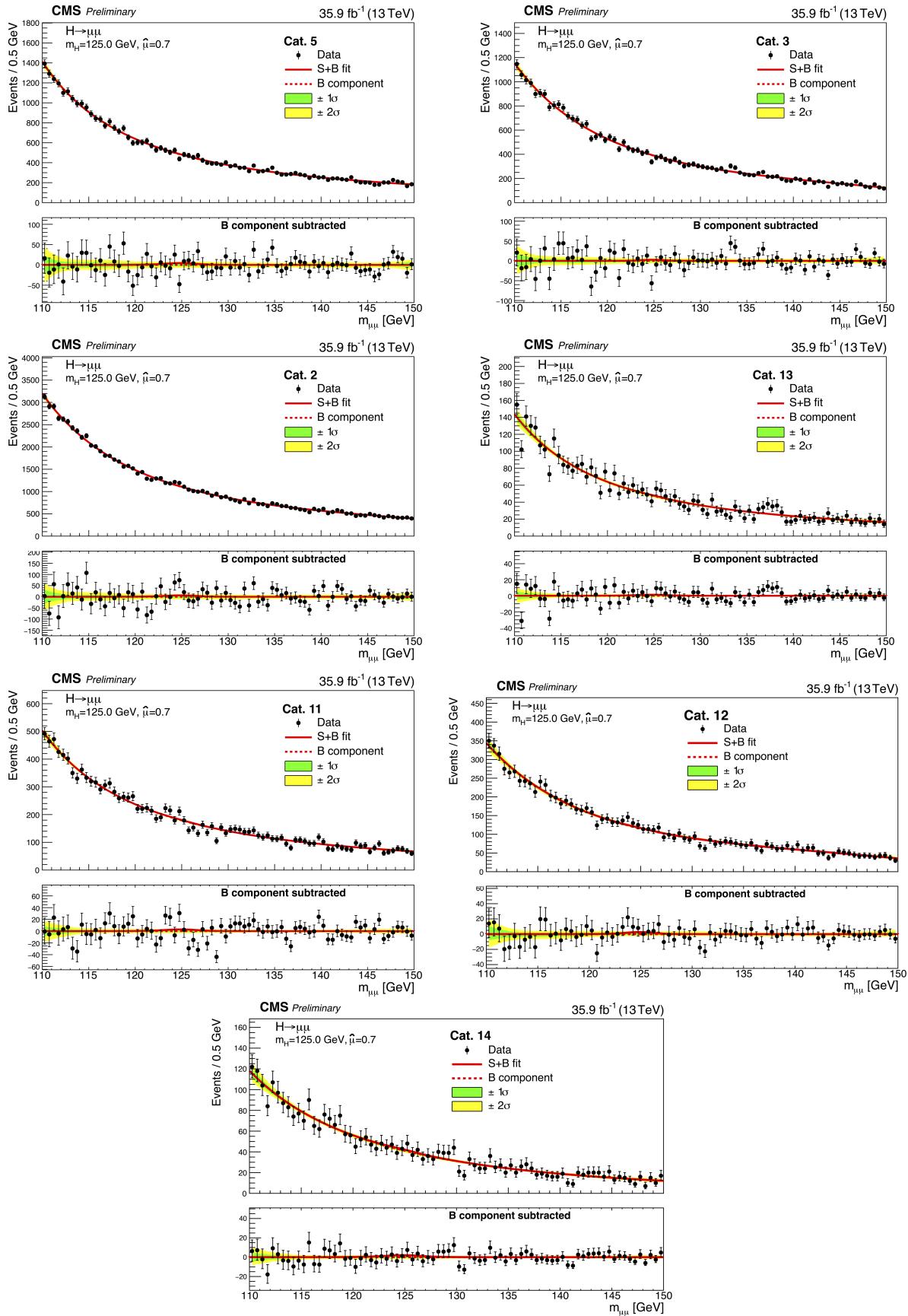


Figure 9-7. The final Signal + Background fits for the 13 TeV data, part two.

9.2 Results Combining 7, 8, And 13 TeV Data

Figures 9-8 to 9-10 present the upper limits, background only p-value, and the best fit signal strength for the combined 7, 8, and 13 TeV data. For $m_H = 125$ GeV, the observed upper limit at 95% confidence is 2.64, while the expected (median) upper limit is 1.89 if the background only hypothesis is true. For $m_H = 125$ GeV, the observed p-value on the background only hypothesis is 0.98σ , and the expected p-value if the Higgs is SM is 1.09σ . The best fit signal strength at $m_H = 125$ GeV is $0.9^{+1.0}_{-0.9}$.

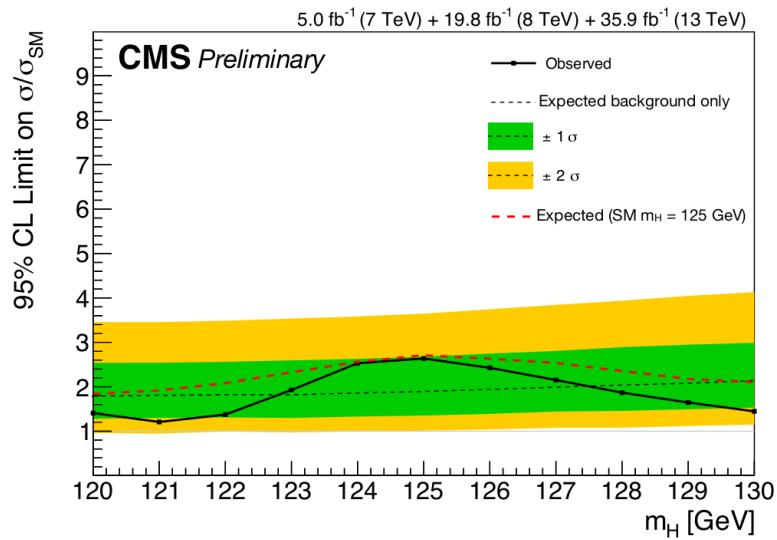


Figure 9-8. The upper limit on the signal strength at 95% confidence is presented for the combination of 7, 8, and 13 TeV data. The red dashed line represents the expected (median) observation if the SM $m_H = 125$ GeV hypothesis is true. The black dashed line represents the expected observation if the background only hypothesis is true.

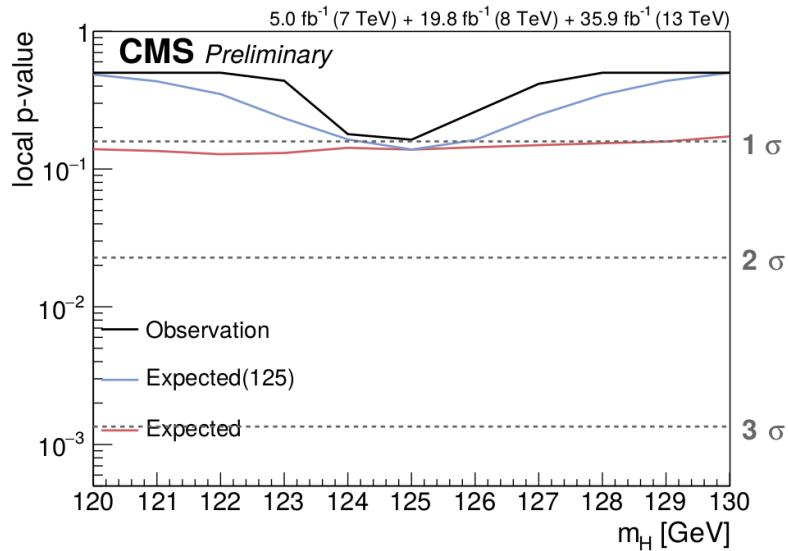


Figure 9-9. The p-value is presented for the combination of 7, 8, and 13 TeV data. The blue line represents the expected (median) p-value on the background only if the SM $m_H = 125$ GeV hypothesis is true. The red line represents the expected p-value at $m_H = x$ GeV if the SM $m_H = x$ GeV hypothesis is true.

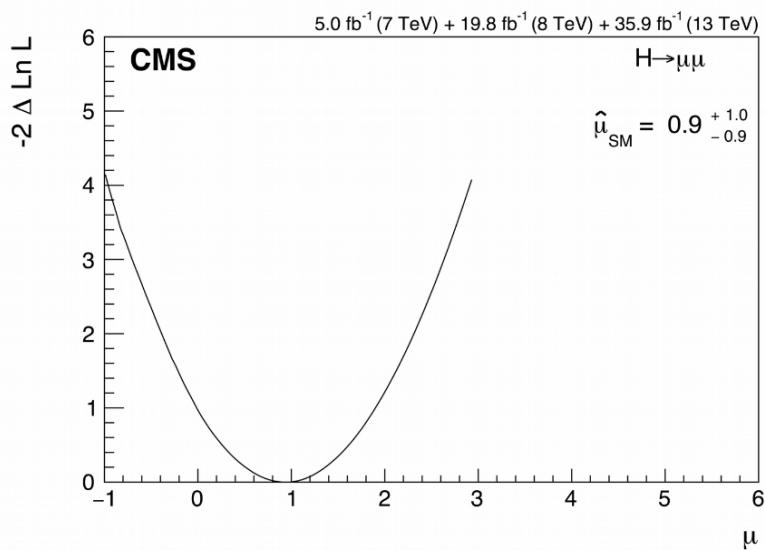


Figure 9-10. The negative log likelihood is plotted near the minimum for the signal strength μ , and the best fit, $\hat{\mu}$, is given. The plot uses the combined 7, 8, and 13 TeV data.

CHAPTER 10

CONCLUSIONS AND OUTLOOK

A Higgs boson with a mass of 125 GeV was discovered in 2012 [6–8], but the exact behavior of the Higgs coupling to fermions remains undetermined. The SM predicts a 2.2×10^{-4} branching fraction for $H \rightarrow \mu^+ \mu^-$, but various BSM theories predict branching fractions greater than the SM value [33–35]. Moreover, the SM also predicts that the branching fractions for the different fermions are proportional to the square of the fermion mass, and specifically that $\mathcal{B}_{\tau\tau}/\mathcal{B}_{\mu\mu} = m_\tau^2/m_\mu^2$. Measuring the decay rates to the different fermions with greater precision reduces the number of acceptable theories, and if any new measurements happen to rule out the Standard Model, then this implies new physics.

ATLAS and CMS used 2011 and 2012 LHC collisions to investigate the branching fractions of a 125 GeV Higgs boson to fermions and to muons in particular. The CMS experiment reported an observed (expected) upper limit on the $H \rightarrow \mu^+ \mu^-$ signal strength of 7.4 (6.5) at 95% confidence [36]. The results used 5.0 fb^{-1} of 7 TeV data and 19.7 fb^{-1} of 8 TeV data. At the same time, ATLAS reported observed (expected) upper limits of 7.1 (7.2) at 95% confidence using 4.5 fb^{-1} of 7 TeV data and 20 fb^{-1} of 8 TeV data [37].

This dissertation presents new results for $H \rightarrow \mu^+ \mu^-$, using 35.9 fb^{-1} of 13 TeV data collected by the CMS experiment in 2016. A new observed (expected) upper limit on the signal strength is reported at 2.64 (2.08) at 95% confidence. An observed (expected) significance of 0.75σ (0.98σ) and a best fit signal strength of $0.7_{-1.0}^{+1.1}$ are also reported. Meanwhile, the latest results from ATLAS report an observed (expected) upper limit of 3.0 (3.1) using 36.1 fb^{-1} of 13 TeV data [38].

Combined results for 5.0 fb^{-1} of 7 TeV, 19.8 fb^{-1} of 8 TeV, and 35.9 fb^{-1} of 13 TeV CMS data are also presented. For $m_H = 125 \text{ GeV}$, the combination yields a measured signal strength of $0.9_{-0.9}^{+1.0}$, observed (expected) upper limits at 95% confidence of 2.64 (1.89), and an observed (expected) significance of 0.98 (1.09σ). The results correspond to an upper limit on the $H \rightarrow \mu^+ \mu^-$ branching fraction of 5.7×10^{-4} . Any theories predicting a larger rate are ruled

out at this confidence level. Upon combining 4.5 fb^{-1} of 7 TeV, 20 fb^{-1} of 8 TeV, and 36.1 fb^{-1} of 13 TeV data, ATLAS reports an observed (expected) upper limit of 2.8 (2.9) at 95% confidence [38].

Scaling the signal and background PDFs from the 13 TeV analysis of this dissertation to different integrated luminosities provides estimates of the expected p-value for a 125 GeV SM Higgs boson versus the integrated luminosity. About 400 fb^{-1} of data is required to observe $H \rightarrow \mu^+ \mu^-$ with a p-value of 3σ . The entire extrapolation can be seen in Figure 10-1.

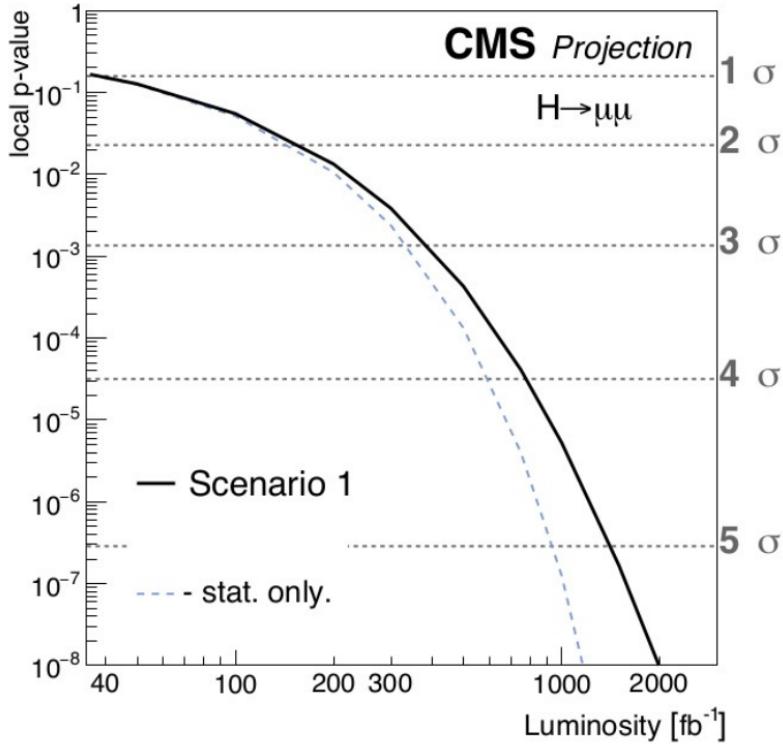


Figure 10-1. The expected p-value on the background only hypothesis assuming a SM Higgs with $m_H = 125 \text{ GeV}$ is plotted versus the amount of 13 TeV data in fb^{-1} .

To date, observations of the Higgs branching fractions to fermions show no significant deviations from the Standard Model predictions [38–41]. In particular, the latest CMS results on $H \rightarrow b\bar{b}$ [40] and $H \rightarrow \tau^+ \tau^-$ [39] report signal strengths of $1.06^{+0.31}_{-0.29}$ and $0.98^{+0.18}_{-0.18}$ for combined 2011, 2012, and 2016 data at $m_H = 125 \text{ GeV}$. The latest measurement of the signal strength for $H \rightarrow \mu^+ \mu^-$ ($0.9^{+1.0}_{-0.9}$) is also consistent with the SM for $m_H = 125 \text{ GeV}$.

REFERENCES

- [1] Sean Carroll. *Dark Matter, Dark Energy: The Dark Side of the Universe* (The Teaching Company, 2007).
- [2] Higgs, P. W. Broken symmetries, massless particles and gauge fields. *Phys. Lett.* **12**, 132–133 (1964).
- [3] Higgs, P. W. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.* **13**, 508–509 (1964). URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [4] Higgs, P. W. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.* **145**, 1156–1163 (1966).
- [5] Lancaster, T. & Blundell, S. J. *Quantum field theory for the gifted amateur* (Oxford University Press, Oxford, 2014). URL <https://cds.cern.ch/record/1629337>.
- [6] ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B* **716**, 1 – 29 (2012). URL <http://www.sciencedirect.com/science/article/pii/S037026931200857X>.
- [7] CMS Collaboration. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B* **716**, 30 – 61 (2012). URL <http://www.sciencedirect.com/science/article/pii/S0370269312008581>.
- [8] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *Journal of High Energy Physics* **2013**, 81 (2013). URL [https://doi.org/10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081).
- [9] CMS Collaboration. Evidence for the 125 gev higgs boson decaying to a pair of τ leptons. *Journal of High Energy Physics* **2014**, 104 (2014). URL [https://doi.org/10.1007/JHEP05\(2014\)104](https://doi.org/10.1007/JHEP05(2014)104).
- [10] CMS Collaboration. Search for the standard model higgs boson produced in association with a w or a z boson and decaying to bottom quarks. *Phys. Rev. D* **89**, 012003 (2014). URL <https://link.aps.org/doi/10.1103/PhysRevD.89.012003>.
- [11] CMS Collaboration. Evidence for the direct decay of the 125 gev higgs boson to fermions. *Nature Physics* **10**, 1038 (2014). URL <http://dx.doi.org/10.1038/nphys3005>.
- [12] ATLAS Collaboration. Search for the $b\bar{b}$ decay of the standard model higgs boson in associated (w/z)h production with the atlas detector. *Journal of High Energy Physics* **2015**, 69 (2015). URL [https://doi.org/10.1007/JHEP01\(2015\)069](https://doi.org/10.1007/JHEP01(2015)069).
- [13] Plehn, T. & Rainwater, D. Higgs decays to muons in weak boson fusion. *Physics Letters B* **520**, 108 – 114 (2001). URL <http://www.sciencedirect.com/science/article/pii/S0370269301011571>.

- [14] Han, T. & McElrath, B. $h \rightarrow \mu^+ \mu^-$ via gluon fusion at the lhc. *Physics Letters B* **528**, 81 – 85 (2002). URL <http://www.sciencedirect.com/science/article/pii/S037026930201208X>.
- [15] Evans, L. & Bryant, P. Lhc machine. *Journal of Instrumentation* **3**, S08001 (2008). URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>.
- [16] Mobs, E. The CERN accelerator complex. (2016). URL <https://cds.cern.ch/record/2197559>. General Photo.
- [17] CMS Collaboration. Cms collaboration. *CMS Web* (2015). URL <http://cms.web.cern.ch/content/cms-collaboration>.
- [18] CMS Collaboration. Cms detector. *CMS Web* (2011). URL <http://cms.web.cern.ch/>.
- [19] CMS Collaboration. CMS Technical Design Report, Volume I: Detector Performance and Software. Tech. Rep. CERN-LHCC-2006-001, CMS-TDR-008-1 (2006).
- [20] Particle Data Group. Review of particle physics. *Phys. Rev. D* **86**, 010001 (2012). URL <https://link.aps.org/doi/10.1103/PhysRevD.86.010001>.
- [21] CMS collaboration. The performance of the cms muon detector in proton-proton collisions at $\sqrt{s} = 7$ tev at the lhc. *Journal of Instrumentation* **8**, P11002 (2013). URL <http://stacks.iop.org/1748-0221/8/i=11/a=P11002>.
- [22] CMS Collaboration. The CMS muon project: Technical Design Report. Tech. Rep. CERN-LHCC-97-032, CMS-TDR-3 (1997).
- [23] CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation* **3**, S08004 (2008). URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08004>.
- [24] Mann, R. B. *An Introduction to Particle Physics and The Standard Model* (CRC Press, Boca Raton, FL, 2010).
- [25] Zee, A. *Quantum Field Theory in a Nutshell*. Nutshell handbook (Princeton Univ. Press, Princeton, NJ, 2003).
- [26] Schwichtenberg, J. *Physics from Symmetry*. Undergraduate Lecture Notes in Physics (Springer International Publishing, Switzerland, 2015).
- [27] Group, L. H. C. S. W. Higgs cross sections and decay branching ratios. *CERN Wiki Page* (2015). URL <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>.
- [28] Friedman, J. H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **29**, 1189–1232 (2001).
- [29] Khachatryan, V. et al. The CMS trigger system. *JINST* **12**, P01020 (2017). [1609.02366](https://arxiv.org/abs/1609.02366).
- [30] Tapper, A. & Acosta, D. CMS Technical Design Report for the Level-1 Trigger Upgrade. Tech. Rep. CERN-LHCC-2013-011, CMS-TDR-12 (2013). URL <https://cds.cern.ch/>

[record/1556311](#). Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch
Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

- [31] Hoecker, A. *et al.* TMVA: Toolkit for Multivariate Data Analysis. *PoS ACAT*, 040 (2007). [physics/0703039](#).
- [32] CMS Collaboration. Level-1 muon trigger performance in 2017 data and comparison with the legacy muon trigger system (2017). URL <https://cds.cern.ch/record/2286327>.
- [33] Dery, A., Efrati, A., Hochberg, Y. & Nir, Y. What if $\text{BR}(h \rightarrow \mu\mu)/\text{BR}(h \rightarrow \tau\tau)$ does not equal m_μ^2/m_τ^2 ? *JHEP* **05**, 039 (2013). [1302.3229](#).
- [34] Vignaroli, N. Searching for a dilaton decaying to muon pairs at the LHC. *Phys. Rev.* **D80**, 095023 (2009). [0906.4078](#).
- [35] Giudice, G. F. & Lebedev, O. Higgs-dependent Yukawa couplings. *Phys. Lett.* **B665**, 79–85 (2008). [0804.1753](#).
- [36] CMS Collaboration. Search for a standard model-like higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the lhc. *Physics Letters B* **744**, 184 – 207 (2015). URL <http://www.sciencedirect.com/science/article/pii/S0370269315002117>.
- [37] ATLAS Collaboration. Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector. *Phys. Lett.* **B738**, 68–86 (2014). [1406.7663](#).
- [38] Collaboration, A. Search for the dimuon decay of the higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the atlas detector. *Phys. Rev. Lett.* **119**, 051802 (2017). URL <https://link.aps.org/doi/10.1103/PhysRevLett.119.051802>.
- [39] CMS Collaboration. Observation of the Higgs boson decay to a pair of tau leptons (2017). [1708.00373](#).
- [40] CMS Collaboration. Evidence for the Higgs boson decay to a bottom quark-antiquark pair (2017). [1709.07497](#).
- [41] CMS Collaboration. Search for the standard model Higgs boson decaying into two muons in pp collisions at $\text{sqrt}(s)=13\text{TeV}$. Tech. Rep. CMS-PAS-HIG-17-019, Geneva (2017). URL <https://cds.cern.ch/record/2292159>.

BIOGRAPHICAL SKETCH

This section is where your biographical sketch is typed in the [bio.tex](#) file. It should be in third person, past tense. Do not put personal details such as your birthday in the file. Again, to make a full paragraph you must write at least three sentences.