

A SEARCH FOR THE STANDARD MODEL HIGGS 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

2017

© 2017 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	6
LIST OF FIGURES	7
ABSTRACT	12
 CHAPTER	
1 INTRODUCTION	14
2 THE LARGE HADRON COLLIDER AND THE CMS EXPERIMENT	16
2.1 The Large Hadron Collider	16
2.2 The Compact Muon Solenoid	18
2.2.1 Silicon Tracker	19
2.2.2 Calorimeters	20
2.2.3 Muon System	22
2.2.3.1 Drift Tubes	23
2.2.3.2 Cathode Strip Chambers	24
2.2.3.3 Resistive Plate Chambers	25
2.2.4 Trigger System	25
2.2.4.1 Level-1 Trigger	25
3 THE STANDARD MODEL	27
3.1 Quantum Field Theory	29
3.1.1 Particles, Symmetries, and Labels	30
3.1.2 The Lagrangian Formalism	30
3.1.3 QFT From Symmetry	35
3.1.3.1 Rotations	35
3.1.3.2 The Lorentz group	39
3.1.3.3 The Poincare group and particle labels	46
3.1.3.4 Building the Lagrangian for a free scalar particle	47
3.1.3.5 Building the Lagrangian for a free spin $\frac{1}{2}$ particle	49
3.1.3.6 Building the Lagrangian for a free spin 1 particle	54
3.1.4 Lagrangians in Quantum Mechanics and QFT	55
3.1.5 Perturbation Theory and Feynman Rules	57
3.1.6 Interactions	64
3.1.7 The Higgs Mechanism	67
3.2 The Standard Model Higgs and the LHC	72

4	THE SEARCH FOR $H \rightarrow \mu^+ \mu^-$	77
4.1	Discovery, Limits, and Measurement, and Sensitivity	77
4.2	The Analysis Strategy	83
4.3	Improving the Level-1 Muon Trigger	85
4.3.1	Metrics of Success	86
4.3.2	The EMTF Regression Project	87
4.3.3	Putting the BDTs into a Look-up Table	89
4.3.4	Results and Conclusions	89
4.4	Data and Monte Carlo Samples	91
4.5	Muon Momentum Calibration	95
4.5.1	Muon Corrections in Data	96
4.5.2	Data-MC agreement in scale, resolution	99
4.5.3	Derivation of systematic uncertainties	105
4.6	Event Selection, Object Reconstruction, and Further Corrections	105
4.6.1	Muons	106
4.6.2	Jets and MET	106
4.6.3	Pile-Up Reweighting	107
4.6.4	Muon Efficiency Scale Factors	107
4.6.5	Jet, MET, and B-Tagging Corrections	108
4.6.6	Event Selection	109
4.7	Maximizing the Sensitivity	109
4.7.1	Training Boosted Decision Trees	110
4.7.2	Event Categorization	112
4.7.3	The Autocategorizer	113
4.7.4	Final Categories	115
4.7.5	Validating The Categories	117
4.7.6	Signal and Background	125
4.8	Modeling the Signal	133
4.9	Modeling the Background	137
4.10	Systematic Uncertainties	139
4.10.1	Signal	139
4.10.1.1	Shape uncertainties	140
4.10.1.2	Category migration uncertainties	140
4.10.1.3	Rate uncertainties	141
4.10.1.4	Impact plots	142
4.10.2	Background	143
4.11	Results	148
4.11.1	Results for 13 TeV Data	148
4.11.2	Results Combining 7, 8, and 13 TeV Data	154
4.11.3	Future Projections	156
	REFERENCES	157
	BIOGRAPHICAL SKETCH	160

LIST OF TABLES

<u>Table</u>		<u>page</u>
4-1 The feature discretization scheme for each mode.	90	
4-2 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.	92	
4-3 The Higgs signal MC samples were generated with POWHEG while the parton shower and hadronization processes are modeled by the PYTHIA8 generator with TuneCUETP8M1.	93	
4-4 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.	94	
4-5 The importance is given by the percentage of the signal and background PDFs along that feature that do not overlap.	111	
4-6 Full Width Half Max (FWHM) of the mass resolution of the signal peak, expected signal and background yields within the FWHM, purity, sensitivity, and the expected limits from data. The expected background yields are estimated by fitting the data in the sidebands.	117	
4-7 The total signal in each category broken down by process.	125	
4-8 Maximum observed bias per category for least biased fit function. The maximum bias is over $m_H = 120, 125, 130$ GeV.	146	

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-1 The CERN Accelerator Complex [Mobs (2016)]	17
2-2 The CMS detector [CMS Collaboration (2011)]	18
2-3 The sagitta measurement	20
2-4 A slice of the CMS detector [CMS Collaboration (2011)]	21
2-5 A Look at the Muon System [CMS collaboration (2013)]	23
2-6 A Drift Tube [CMS Collaboration (1997)]	24
2-7 A Cathode Strip Chamber [CMS Collaboration (1997)]	24
2-8 The L1 Trigger Architecture [CMS Collaboration (2008)]	26
3-1 The Standard Model Particles	28
3-2 Feynman diagrams for the Numerator $N(\lambda^1) = Z_0A + Z_0B + Z_0C$	62
3-3 The Feynman diagrams for $\langle 0 \Phi(x_4)\Phi(x_3)\Phi(x_2)\Phi(x_1) 0 \rangle_{\lambda^1}$ representing the matrix element for $2 \rightarrow 2$ scattering up to first order.	63
3-4 The highest production mode cross sections for the SM Higgs at 14 TeV [Group (2015)]	72
3-5 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 [Group (2015)].	74
3-6 The SM production modes with the highest cross sections. a) Gluon Gluon Fusion (GGF) b) Vector Boson Fusion (VBF) c) Associated Production with a Vector Boson (VH) d) $t\bar{t}H$	75
3-7 The quark-gluon background looks very similar to the GGF production channel when the Higgs decays to two jets. Protons are made of quarks and gluons so this process is extremely common in proton colliders like the LHC.	75
4-1 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 CMS Collaboration (2017).	90
4-2 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 $t\bar{t}H$	91

4-3	The dimuon mass distribution is fit around the Z peak for the positively charged muon with ϕ between 0 and 0.53.	95
4-4	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.	97
4-5	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.	98
4-6	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.	99
4-7	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.	100
4-8	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.	101
4-9	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.	102
4-10	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 Rochester corrections, data and MC have similar resolutions for the Z peak.	103
4-11	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.	104
4-12	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.	105
4-13	The background-rejection vs. signal efficiency ROC curve for the final BDT model. .	112

4-14 The BDT score on the signal and background Monte Carlo (left), where a higher score marks the 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).	113
4-15 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.	114
4-16 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.	115
4-17 The categories produced by the autocategorizer. $ \eta $ refers to the maximum $ \eta $ value of the two muons in the dimuon pair.	115
4-18 The final categorization	116
4-19 Validation of BDT input variables in inclusive events.	118
4-20 Data/MC agreement for the inclusive events.	119
4-21 Data/MC agreement for the most sensitive category.	120
4-22 Data/MC agreement for the most sensitive category.	121
4-23 The BDT distribution in data and MC. Right plot matches the DY p_t spectrum to the one in data.	122
4-24 The BDT distribution in quantile in data and MC. The systematic uncertainty is given by the JES and PU.	123
4-25 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.	124
4-26 Kinematics and object counts for the signal in the inclusive set of events.	126
4-27 More kinematic and object count histograms for the signal in the inclusive set of events.	127
4-28 More kinematic and object count histograms for the signal in the inclusive set of events.	128
4-29 Kinematic and object count histograms for the signal in category c14. Note the high dimuon p_t targeting the signal in general.	129

4-30 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.	130
4-31 More kinematic and object count histograms for the signal in category c14. Note the low MET and 0 btags which exclude $t\bar{t}$	131
4-32 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.	132
4-33 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).	134
4-34 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.	135
4-35 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.	135
4-36 The efficiency times acceptance as function of m_H for the total signal and for each signal process.	136
4-37 Example background-only fits to the dimuon mass spectrum in data for various background models and mass ranges.	138
4-38 Examples of the F-Test for the Bernstein Polynomials (left) and the Sum of Exponentials (right).	139
4-39 Pseudorapidity of the leading jet, and number of associated jets. The uncertainty bands show jet energy scale uncertainty.	142
4-40 Impact plots derived from the Asimov data-set.	143
4-41 Bias results for categories 0-8 with each category's bias indepent 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.	145
4-42 Bias results for categories 9-14 with each category's bias indepent 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.	146
4-43 The plot shows the bias on $\hat{\mu}$ using the combined likelihood for all categories and the background fits of Figure 4-8. In this estimate, the same generating function is used for every category.	147
4-44 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.	149

4-45 The combined upper limit on the signal strength at 95% confidence is presented for 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.	149
4-46 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.	150
4-47 The negative log likelihood is plotted near the minimum for the signal strength μ , and the best fit, $\hat{\mu}$, is given.	150
4-48 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.	151
4-49 The final signal + background fits for the 13 TeV data, part one.	152
4-50 The final Signal + Background fits for the 13 TeV data, part two.	153
4-51 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.	154
4-52 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.	155
4-53 The negative log likelihood is plotted near the minimum for the signal strength μ , and the best fit, $\hat{\mu}$, is given.	155
4-54 The expected p-value on the background only hypothesis assuming a SM Higgs with $m_H = 125$ GeV is plotted versus the amount of 13 TeV data in fb^{-1}	156

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 DECAYING TO TWO MUONS AT THE
CMS EXPERIMENT

By

Andrew Carnes

December 2017

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, 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 at a mass of 125 GeV are 2.64 and $2.08 \times$ respectively. The significance is reported in the same range, and the observed and expected significance at $m_H = 125 \text{ GeV}$ are 0.98σ and 0.74σ 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 [[Sean Carroll \(2007\)](#)]. 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 [[Higgs \(1964b,a, 1966\)](#); [Lancaster and Blundell \(2014\)](#)]. 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 [[ATLAS Collaboration \(2012\)](#); [CMS Collaboration \(2012, 2013\)](#)]. 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 [[CMS Collaboration \(2014a,c,b\)](#); [ATLAS Collaboration \(2015\)](#)]. 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 [[Plehn and Rainwater \(2001\)](#); [Han and McElrath \(2002\)](#)], 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 the 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 $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ [[Evans and Bryant \(2008\)](#)]. 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

CERN's Accelerator Complex

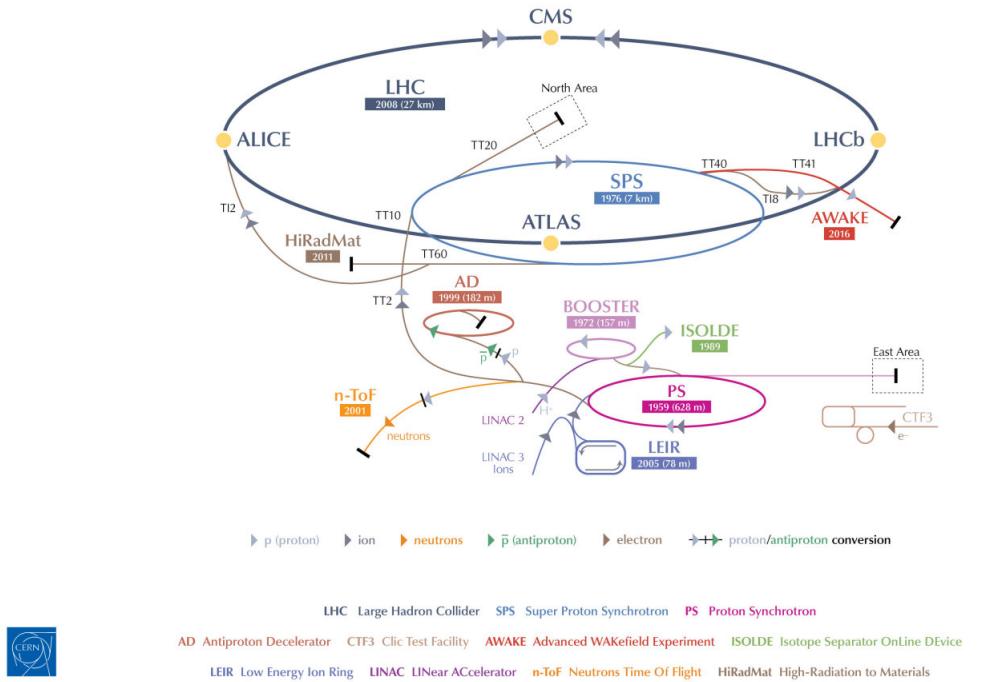


Figure 2-1. The CERN Accelerator Complex [Mobs (2016)]

and electrons. The protons are then sent to a radio frequency quadrupole which focuses the 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 [[CMS Collaboration \(2015\)](#)]. The greatest achievement of the collaboration to date is the discovery of a Higgs like particle in 2012, a feat shared with ATLAS.

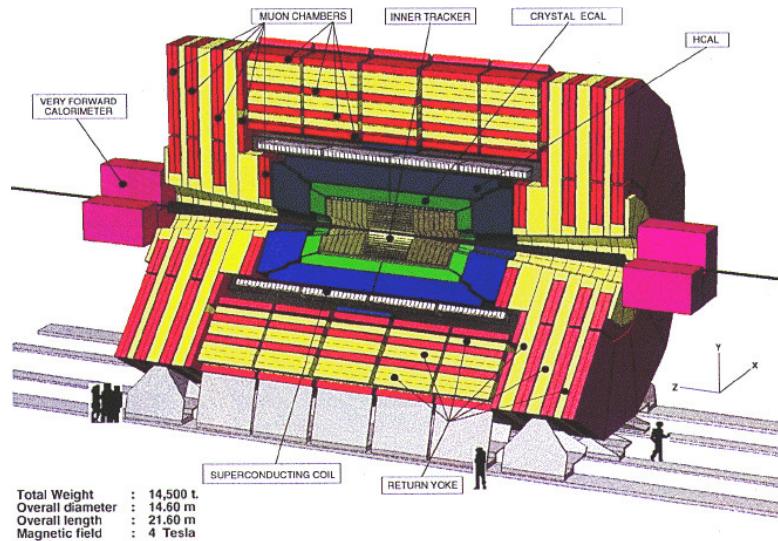


Figure 2-2. The CMS detector [[CMS Collaboration \(2011\)](#)]

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 [CMS Collaboration (2006)]. 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. Jets ¹ 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 [Particle Data Group (2012)].

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

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.

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

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.

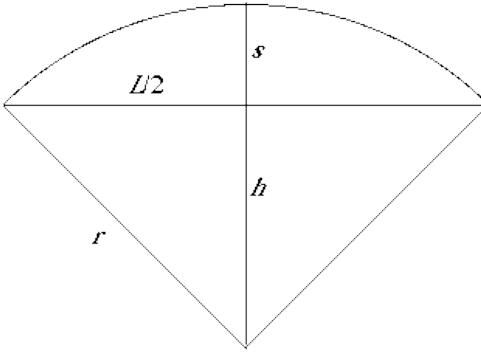


Figure 2-3. The sagitta measurement

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 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

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.

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 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.

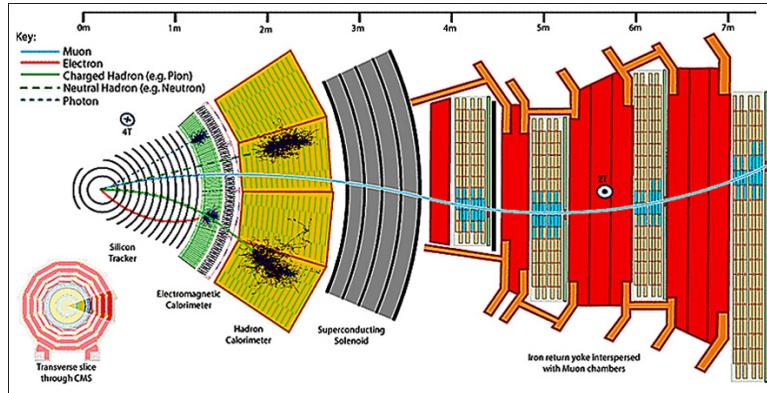


Figure 2-4. A slice of the CMS detector [[CMS Collaboration \(2011\)](#)]

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 [[Particle Data Group \(2012\)](#)].

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

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) [CMS Collaboration (2006)].

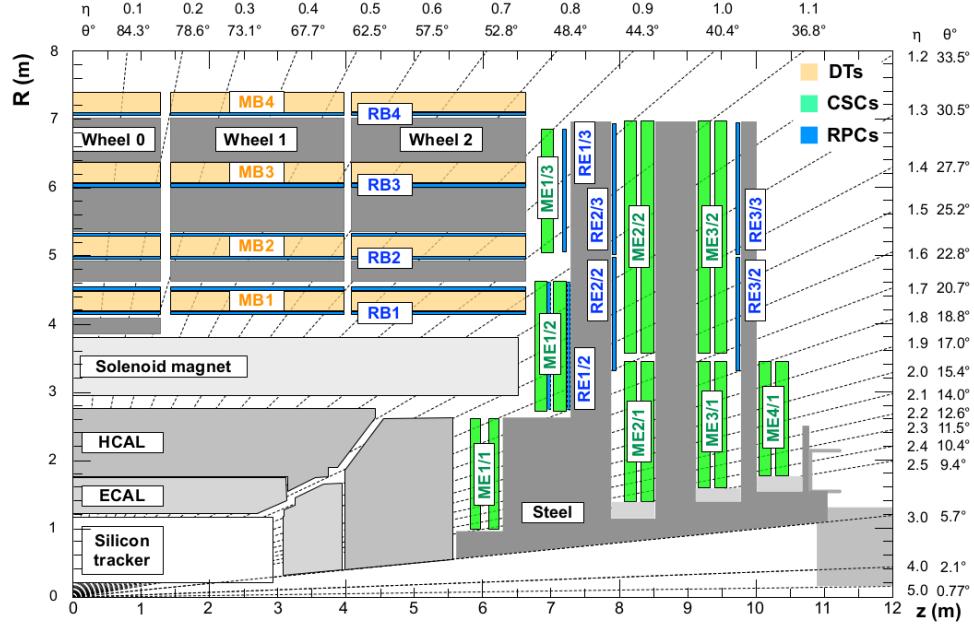


Figure 2-5. A Look at the Muon System [CMS collaboration (2013)]

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

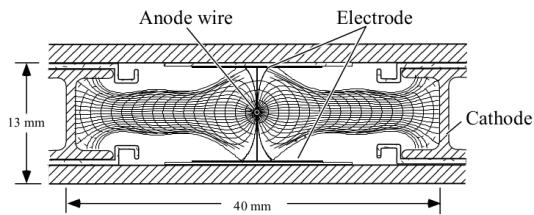


Figure 2-6. A Drift Tube [CMS Collaboration (1997)]

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.

2.2.3.2 Cathode Strip Chambers

The CSCs are located in the endcaps of the detector which range in $|\eta|$ from 0.8 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.

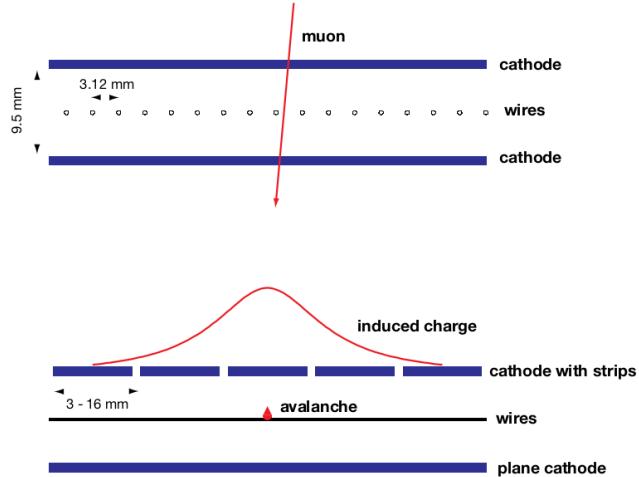


Figure 2-7. A Cathode Strip Chamber [CMS Collaboration (1997)]

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 [CMS Collaboration (2008)].

In this way, 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 10 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 [[CMS Collaboration \(2008\)](#)]. 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.

2.2.4.1 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.

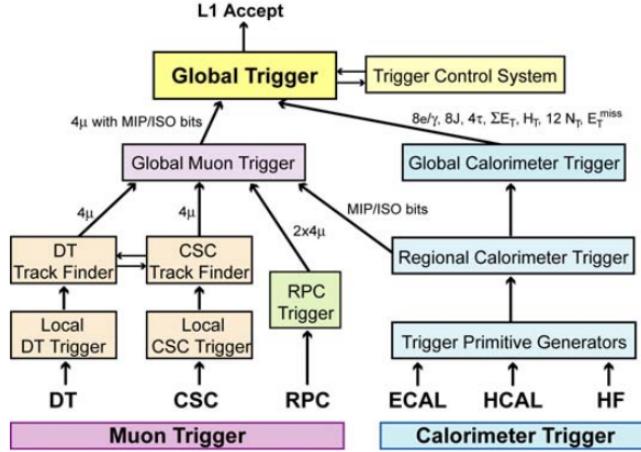


Figure 2-8. The L1 Trigger Architecture [CMS Collaboration (2008)]

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 [Mann (2010)]. Well nearly all, quantum gravity and the non-zero neutrino mass are still a mystery [Mann (2010); Sean Carroll (2007); Zee (2003)]. 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.

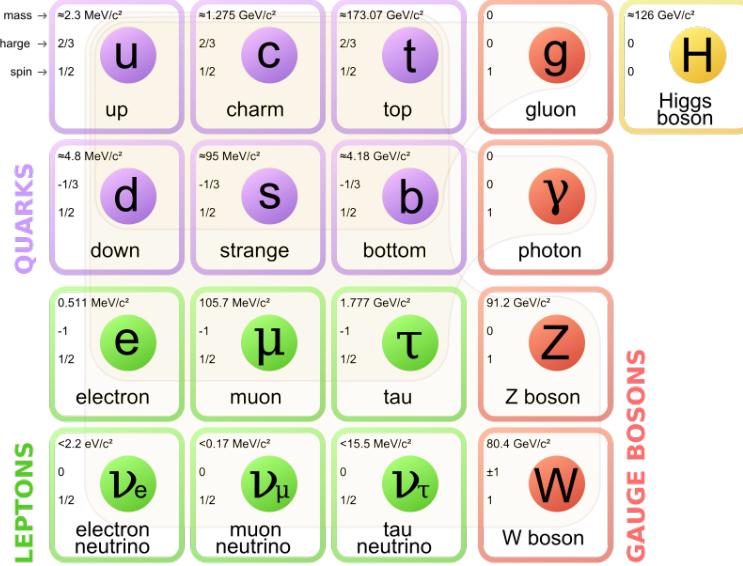


Figure 3-1. The Standard Model Particles

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. With such a large role in the SM, finding this particle or a BSM Higgs has been a huge priority for the CMS collaboration [CMS Collaboration (2006)]. In 2012, a Higgs particle with a mass of 125 GeV was found and to date remains consistent with the Standard Model [ATLAS Collaboration (2012); CMS Collaboration (2012, 2013)]. However, the properties need to be investigated further before declaring the discovered Higgs the Higgs of the Standard Model.

In order to lay out the Standard Model, the next section develops Quantum Field Theory (QFT) and covers 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

The mathematical framework used to describe the physics of the SM as well as other Beyond Standard Model (BSM) field theories is called Quantum Field Theory (QFT). QFT enables the predictions of different measurable probabilities [Peskin and Schroeder (1995)]. One of the most important is the probability for an initial set of particles to collide and produce another set of particles. Another important one is the probability for a single particle to decay into a certain set of particles. These probabilities are encompassed in the cross sections and branching fractions. For example, the theory of the SM predicts the cross section for two protons to collide and create a Higgs. As another example, the SM also predicts the branching fraction for a Z boson to decay into two muons. These probabilities can be measured simply by colliding particles and counting the outcomes which in turn means that different field theories can be tested. Quantum Field Theory is written down in terms of a

Lagrangian, and the math for the various predictions follows from there, but first a brief aside about particles.

3.1.1 Particles, Symmetries, and Labels

Since QFT makes quantitative predictions in terms of particle collisions and particle decays, it's interesting to contemplate what a particle really is. Consider an observer in a frame x with particle p and an observer in another frame x' . If the observer in x' can't identify particle p as well then it doesn't make sense to call p a particle. More concretely, consider a world where in frame x an observer sees a neatly stacked deck of cards, but in x' the observer sees the cards scattered all over the place. Calling the deck of cards a particle doesn't really make sense. On the other hand, both parties can still agree on the individual cards which kept the same suit and value. The suit and value are conserved quantities identifiable between frames. If the two observers get together later and compare notes they can see what happened to each card upon transforming from x to x' and work out a set of rules. The king of hearts may do one thing and the 10 of clubs another. They can then add the different forces into play repeat the process and compare again. Figuring this all out allows the two to determine the laws of physics for the fundamental pieces called particles.

This idea leads to Wigner's view [[Wigner \(1939\)](#)]: a particle is an object with conserved quantities that observers can agree on between frames. In our universe some of these labels are the mass, charge, spin, color, isospin, and weak hypercharge with each label attached to a specific symmetry or group of symmetries. And because different observers can agree on these quantities they can compare notes and work out the laws of physics for the different types of particles. The deep connection between conserved quantities, symmetries, and labels for particles will be seen later. For now the focus is to work out laws that people in different frames can confirm, and this is where the Lagrangian formalism comes into play.

3.1.2 The Lagrangian Formalism

The Lagrangian formalism is a mathematical device that allows physicists to describe the evolution of a physical system over time, and it's within this formalism that QFT can be built

[[Peskin and Schroeder \(1995\)](#); [Lancaster and Blundell \(2014\)](#)]. But before building the full mathematics of QFT, a simple example describing the free Newtonian particle is covered. The Newtonian example serves to illuminate the fact that the laws of physics are indistinguishable in different inertial frames. Moving to the full relativistic theory reveals that the symmetry goes deeper. Investigating the symmetries of the relativistic theory end up leading the way to QFT, a relativistic description of quantum mechanics.

Getting into the framework, the goal of physics is to describe how a physical system evolves over time, and this evolution is usually given by some differential equation describing the state the system will take in the next interval of time given the current time. Moving from state to state, from one interval of time to the next, the system traces out a path in some abstract space of possible states. As a concrete example, the differential equations describing the motion of a Newtonian particle determine the position and velocity of the particle at the next instant of time based upon the values in the previous instant of time. So how does one get the appropriate differential equation? At classical scales, nature tries to minimize the difference between the energy spent³ and the energy available to spend and it minimizes the action S^4 . At the more fundamental level of quantum mechanics, all possible paths contribute with different phases, but those close to the minimum add constructively, contributing the most. As the action gets larger, the paths away from the minimum count less and less and the quantum rule agrees with the classical rule. The classical case is analyzed first in order to begin investigating the symmetries that lead to QFT and an appropriate description of nature.

Equation 3–1 presents the action S , where L is the Lagrangian, T is the kinetic energy and U is the potential energy.

$$S = \int L dt = \int T - U dt \quad (3-1)$$

³ Spent here just means used as kinetic energy.

⁴ Technically, nature looks for stationary action, which could be a minimum or a maximum, but the minimum is the usual case

At an extremum of S , $\delta S = 0$ if S is smooth. S must decrease, go through a slope of zero, and then increase (or vice versa). For a true extremum, δS must be 0 in all directions. So to get the equations of motion vary the parameters of L and solve the system of equations for the values that yield zero change in the action.

$$\begin{aligned}\frac{\delta S}{\delta z_1} &= \int L(z_1 + dz_1, z_2, \dots) dt - \int L(z_1, z_2, \dots) dt = 0 \\ \frac{\delta S}{\delta z_2} &= \int L(z_1, z_2 + dz_2, \dots) dt - \int L(z_1, z_2, \dots) dt = 0 \\ &\dots\end{aligned}\quad (3-2)$$

Following this process provides the Euler-Lagrange equations, describing how the parameters z evolve over time. The z 's may be the position and velocity, or the quantum fields, or the temperature and volume or some other set of parameters that describe the system. The Lagrangian for a Newtonian free particle in one dimension is pretty simple and gets the point across.

$$S = \frac{1}{2} \int m \dot{x}^2 dt \quad (3-3)$$

If the action is at an extremum, perturbing the path $x(t)$ by adding the infinitesimal $\epsilon(t)$ leaves the action unchanged.

$$S' = \frac{1}{2} \int m(\dot{x} + \dot{\epsilon})^2 dt = \frac{1}{2} \int m(\dot{x}^2 + 2\dot{x}\dot{\epsilon} + \dot{\epsilon}^2) dt = \frac{1}{2} \int m(\dot{x}^2 + 2\dot{x}d\dot{x}) dt \quad (3-4)$$

$$\delta S = S' - S = 0 = \frac{1}{2} \int m(\dot{x}^2 + 2\dot{x}\dot{\epsilon}) dt - \frac{1}{2} \int m \dot{x}^2 dt = \int m \dot{x} \dot{\epsilon} dt \quad (3-5)$$

With $x(t)$ fixed at the boundaries of the integral, ϵ must be zero at t_o and t_f , so integrating by parts yields the following equation

$$\delta S = 0 = \epsilon(t_f)\dot{x}(t_f) - \epsilon(t_o)\dot{x}(t_o) + \int m \ddot{x} \epsilon dt = 0\dot{x}(t_f) - 0\dot{x}(t_o) + \int m \ddot{x} \epsilon dt = \int m \ddot{x} \epsilon dt. \quad (3-6)$$

And this equation must be zero for any infinitesimal deviation ϵ ,

$$\delta S = 0 \rightarrow m \ddot{x} = 0, \quad (3-7)$$

implying that a free particle keeps the same velocity over time. Note that a Newtonian boost by constant velocity $v \rightarrow v' = v + u$ ⁵ leaves the equations of motion consistent. In the unprimed frame, the particle has velocity v with 0 acceleration. In the primed frame, the particle has velocity $v + u$ with 0 acceleration. Both observers see the particle act as if there are zero forces in play,

$$S = \frac{1}{2} \int m(v + u)^2 dt = \frac{1}{2} \int m(v')^2 dt \rightarrow \delta S = 0 \rightarrow m \frac{d}{dt}(v + u) = m \frac{d}{dt}(v') = 0. \quad (3-8)$$

If u is not constant but a function of time, $u(t)$, then the equations of motion do not describe the same time evolution,

$$m \frac{d}{dt}(v + u) = m\dot{v} + m\dot{u} = 0 \rightarrow \dot{v} = -\dot{u}. \quad (3-9)$$

In the case where $u(t)$ depends upon time, the difference between the primed and unprimed frames' equations of motion is then δF .

$$\delta F = m \frac{d}{dt}(v + u) - m \frac{dv}{dt} = m\dot{v} + m\dot{u} - m\dot{v} = m\dot{u} \quad (3-10)$$

In the unprimed frame, the particle identified by the mass moves with constant velocity, $\dot{v} = 0$. The observer in the primed frame looks at the particle with the same mass and sees it change velocity given by the equation $\dot{v} = -\dot{u}$. As an example, set v and u , and \dot{u} to zero for all times before $t=0$, and let u , \dot{u} turn on after time 0. Both observers will agree that the particle is stationary up until time 0. After which, the observer in the primed frame will see the particle accelerate in strange ways. Meanwhile, the unprimed frame will continue to observe a stationary particle.

In general, every inertial frame finds $\delta F = 0$ and every accelerating frame finds an extra force δF unique to its acceleration. In this way, no observer in an inertial frame can perform an experiment and determine which inertial frame he or she is in. On the other hand, each

⁵ Renaming \dot{x} as v .

accelerating frame is identified by its δF . Put another way, in every inertial frame, a ball released at rest remains at rest. In an accelerating frame, the ball will accelerate according to the motion of the frame δF and this change in the laws of physics identifies the frame in a unique way. Conversely, the laws of physics remain the same boosting between inertial frames, and this invariance is a symmetry of physics. Of course this example is Newtonian and the correct way to boost is given by the Lorentz transformation from Special Relativity, but this gets the point across.

Delving further along the path of symmetry, the fundamental forces depend only on the distance from the charge and not the direction implying that rotations are also a symmetry. This can be seen by looking at the Lagrangian,

$$L = \frac{1}{2}m\dot{\vec{x}}^2 - U[(\vec{x} - \vec{x}')^2]. \quad (3-11)$$

Rotations leave dot products and consequently the magnitude of vectors unchanged so the Lagrangian is invariant under this transformation. Naturally if the Lagrangian is invariant the equations of motion will be as well,

$$m\frac{d\vec{v}}{dt} = \vec{\nabla}U. \quad (3-12)$$

In the equations of motion of 3-12, both sides are vectors and vectors transform the same way under rotations so the equations of motion are invariant. Note that in the case of rotations both the Lagrangian and the equations of motion are invariant, while for Newtonian boosts, only the equations of motion were invariant. This is due to the fact that Newtonian mechanics is the low velocity limit of relativistic mechanics. In the theory of Special Relativity, the action for a massive free particle is written like so,

$$S = \int \frac{m}{2}u^\mu u_\mu d\tau. \quad (3-13)$$

Just as rotations preserve the dot product, Lorentz transformations (boosts and rotations) preserve the four vector product. Building a Lagrangian out of four vector products then guarantees that both the Lagrangian and the equations of motion will remain invariant under

Lorentz transformations. However, four vectors aren't the most fundamental objects with invariant products. By studying the properties of the Lorentz group, it's possible to find even more fundamental building blocks called spinors. Spinors, vectors, and scalars are all necessary to describe the different types of observed particles and create a Lagrangian that can describe the real world.

3.1.3 QFT From Symmetry

As just explored, the laws of physics are invariant under boosts and rotations, and the Lagrangian provides a mathematical framework for physical predictions. These facts together imply that there's a good shot at building a proper QFT by creating the appropriate invariant Lagrangian. Four vector products remain invariant under Lorentz transformations so they are a natural ingredient, but, as aforementioned, there are other mathematical objects that could be used as well. In this vein, the symmetries under rotations and boosts are investigated in order to look for other building blocks. The goal is to find three different representations of the Lorentz group and use one representation for scalar spin 0 particles, another for the spin $\frac{1}{2}$ fermions, and another for the spin 1 bosons.

3.1.3.1 Rotations

Rotations in three dimensions are described by the SO(3) group. Rotations preserve the lengths of vectors and the angles between them, which means that dot products between vectors remain invariant as well. In three dimensions, one can rotate about any of the three axes. The rotations about the x, y, and z axes may be characterized by the matrices below.

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{pmatrix} \quad (3-14)$$

$$R_y = \begin{pmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{pmatrix} \quad (3-15)$$

$$R_z = \begin{pmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3-16)$$

These rotations may be built up from repeated rotations by an infinitesimally small angle $d\theta$.

The matrices characterizing an infinitesimal rotation are produced by taking the limit as θ goes to zero.

$$dR_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -d\theta_x \\ 0 & d\theta_x & 1 \end{pmatrix} = 1 - id\theta_x \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} = 1 - id\theta_x J_x \quad (3-17)$$

$$dR_y = \begin{pmatrix} 1 & 0 & d\theta_y \\ 0 & 1 & 0 \\ -d\theta_y & 0 & 1 \end{pmatrix} = 1 - id\theta_y \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} = 1 - id\theta_y J_y \quad (3-18)$$

$$dR_z = \begin{pmatrix} 1 & -d\theta_z & 0 \\ d\theta_z & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 1 - id\theta_z \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 1 - id\theta_z J_z \quad (3-19)$$

Repeating an infinitesimal rotation many times builds the finite rotation, and in this way, the J matrices generate rotations along their respective axes. As such, they are aptly referred to as the generators of the group. Consider any of the J matrices applied many times,

$$R = (1 - i\frac{\theta}{N}J)^N = 1 + (-id\theta J) + \frac{1}{2!}(-id\theta J)^2 + \frac{1}{3!}(-id\theta J)^3 + \dots = e^{-i\theta J}. \quad (3-20)$$

Notice that even powers of J yield J^2 and that odd powers of J return J ,

$$= 1 - J^2 + J^2(1 + \frac{i^2}{2!}d\theta^2 + \frac{i^4}{4!}d\theta^4 + \dots) - iJ(d\theta + \frac{i^2}{3!}d\theta^3 + \frac{i^4}{5!}d\theta^5 + \dots) = (1 - J^2) + J^2 \cos \theta - iJ \sin \theta. \quad (3-21)$$

Plugging in J_z reveals that this process does in fact rebuild the rotation matrix R_z ,

$$\begin{aligned}
R_z &= \left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right) + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos \theta_z + \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \sin \theta_z \\
&= \begin{pmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\end{aligned} \tag{3-22}$$

Similarly, the other generators rebuild their respective rotation matrices. The generators of the group are actually more fundamental than the rotation matrices. The multiplication table for the generators describes the algebra of the group, which describes the behavior of rotations at a local level. In fact, the $SO(3)$ rotation matrices are just one of the groups with this local algebra. A specific group obeying the local algebra is analogous to a specific solution of a differential equation: each solution has a different global behavior yet each obeys the same physics at the differential scale. The multiplication table for the algebra can be specified without declaring any particular representation for the generators.

$$\begin{aligned}
J_x * J_y &= iJ_z + J_y * J_x \\
J_y * J_z &= iJ_x + J_z * J_y \\
J_z * J_x &= iJ_y + J_x * J_z
\end{aligned} \tag{3-23}$$

The multiplication table can be specified in a more compact notation using the commutator⁶, $[a, b] = ab - ba$, and the antisymmetric tensor ϵ ,

$$[J_k, J_l] = i\epsilon_{klm}J_m. \tag{3-24}$$

⁶ The Lie Bracket defines the multiplication for a Lie Algebra and this reduces to the commutator for Lie groups of matrices like $SO(3)$. Using the commutator below returns another member of the group and thus the group is closed under commutation.

The commutator closes the group.

Finding a 3x3 representations of the generators and then repeating the infinitesimal transformations builds the SO(3) rotation group. The group acts on real 3x1 objects called vectors, and these 3x1 vectors are a suitable candidate for a Newtonian Lagrangian. Finding another representation obeying this algebra will provide a more fundamental ingredient for the Lagrangian and allow the construction of a proper QFT. Similar to the way real numbers are built from the squares of imaginary numbers, vectors are built from spinors.

Looking for the lowest order nontrivial $n \times n$ matrices satisfying the algebra gives the 2x2 Pauli matrices. The 1x1 matrices are the trivial solution⁷. The solution for the 2x2 case, the Pauli matrices, are given by

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3-25)$$

However, plugging these into the commutator reveals a factor of two difference,

$$[\sigma_k, \sigma_l] = 2i\epsilon_{klm}\sigma_m. \quad (3-26)$$

Defining $J_k = \frac{1}{2}\sigma_k$ fixes this. These matrices act on an array of 2x1 complex numbers called spinors, and this new rotation group is called SU(2). By starting with vectors and analyzing the SO(3) rotation group along with its underlying algebra, new mathematical objects have been discovered. The 1x1 matrices satisfying the rotation algebra make up the spin 0 representation of SU(2), the complex 2x2 matrices acting on complex 2x1 objects satisfying the algebra make up the spin $\frac{1}{2}$ representation, and the complex 3x3 matrices acting on complex 3x1 objects satisfying the algebra make up the spin 1 representation. The pattern continues on. There are in fact many representations of SU(2). It's now possible to use these

⁷ 1x1 matrices are simply scalar complex numbers, which commute. The only way to satisfy the algebra is to set the J matrices to 0. The objects these 1x1 operators act on are 1x1 numbers called scalars which remain invariant under rotations and correspond to spin 0.

representations to build rotationally invariant Lagrangians, using the 2x1 spinors of SU(2) to model fermions and the 3x1 representation to model bosons. While rotationally invariant Lagrangians are important for nonrelativistic theories, the real goal is to break down four vectors in the same way to find the most fundamental ingredients for relativistic Lagrangians.

3.1.3.2 The Lorentz group

Four vector products are invariant with regards to rotations and boosts. This statement is defined by the mathematical equation below, where the Λ matrices represent the rotation/boost matrices of the Lorentz group⁸ and the η matrix is the Minkowski metric $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$.

$$x'_\mu x'^\mu = x_\mu x^\mu \rightarrow \eta_{\sigma\rho} \Lambda_\mu^\sigma \Lambda_\nu^\rho x^\mu x^\nu = \eta_{\mu\nu} x^\mu x^\nu \rightarrow \eta_{\sigma\rho} \Lambda_\mu^\sigma \Lambda_\nu^\rho = \eta_{\mu\nu} \quad (3-27)$$

In this $3 + 1$ dimensional space the rotations and their corresponding generators are now given by the following R and J matrices,

$$R_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos \theta_x & -\sin \theta_x \\ 0 & 0 & \sin \theta_x & \cos \theta_x \end{pmatrix}, J_x = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} \quad (3-28)$$

$$R_y = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta_y & 0 & \sin \theta_y \\ 0 & 0 & 1 & 0 \\ 0 & -\sin \theta_y & 0 & \cos \theta_y \end{pmatrix}, J_y = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \end{pmatrix} \quad (3-29)$$

⁸ The Lorentz group dealt with here is the proper orthochronous Lorentz group SO(1,3).

$$R_z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta_z & -\sin \theta_z & 0 \\ 0 & \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, J_z = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (3-30)$$

where the R matrices satisfy equation 3 – – 27. The boosts are given by the B matrices,

$$B_x = \begin{pmatrix} \cosh \omega_x & \sinh \omega_x & 0 & 0 \\ \sinh \omega_x & \cosh \omega_x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3-31)$$

$$B_y = \begin{pmatrix} \cosh \omega_y & 0 & \sinh \omega_y & 0 \\ 0 & 1 & 0 & 0 \\ \sinh \omega_y & 0 & \cosh \omega_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3-32)$$

$$B_z = \begin{pmatrix} \cosh \omega_z & 0 & 0 & \sinh \omega_z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh \omega_z & 0 & 0 & \cosh \omega_z \end{pmatrix}. \quad (3-33)$$

These also leave the four vector product invariant. Looking at the differential boosts yields the generators K.

$$dB_x = \begin{pmatrix} 1 & d\omega_x & 0 & 0 \\ d\omega_x & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 1 + d\omega_x \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 1 + d\omega_x K_x \quad (3-34)$$

$$dB_y = \begin{pmatrix} 1 & 0 & d\omega_y & 0 \\ 0 & 1 & 0 & 0 \\ d\omega_y & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 1 + d\omega_y \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 1 + d\omega_y K_y \quad (3-35)$$

$$dB_z = \begin{pmatrix} 1 & 0 & 0 & d\omega_z \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ d\omega_z & 0 & 0 & 1 \end{pmatrix} = 1 + d\omega_z \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = 1 + d\omega_z K_z \quad (3-36)$$

As before, with the algebra for rotations in three dimensions, the Lorentz algebra is defined by its multiplication table. The multiplication table is given by the commutation relations below, which can be confirmed by brute force computation.

$$[J_i, J_j] = i\epsilon_{ijk} J_k \quad (3-37)$$

$$[J_i, K_j] = i\epsilon_{ijk} K_k \quad (3-38)$$

$$[K_i, K_j] = -i\epsilon_{ijk} J_k \quad (3-39)$$

Notice that the commutator between J matrices returns another J matrix, but that the commutator between K matrices returns a J matrix. This means that the J operators form their own subgroup, but the K operators don't. On the other hand, mixing the J s and K s up by defining the Y^\pm operators allows the Lorentz algebra to be represented by two independent subgroups.

$$Y^\pm = \frac{1}{2}(J_i \pm iK_i) \quad (3-40)$$

$$[Y_i^\pm, Y_j^\pm] = i\epsilon_{ijk} Y_k^\pm \quad (3-41)$$

$$[Y_i^\pm, Y_j^\mp] = 0 \quad (3-42)$$

Both of the Y groups have the same commutation relations as $SU(2)$. In this way, the Lorentz algebra can be viewed as if two orthogonal $SU(2)$ rotation algebras have been glued

together. This is similar to the way in which orthogonal basis vectors are stuck together to create a larger dimensional space. Now, in order to figure out how to use this space to build the appropriate Lagrangians, the individual subspaces must be investigated. Looking at Y^+ alone is akin to looking along the Y^+ axis by setting Y^- to zero. Using (y_+, y_-) to label the representation, the simplest nontrivial case along the Y^+ axis is spin $\frac{1}{2} \times$ spin 0, given by $(\frac{1}{2}, 0)$.

Using $Y_i^- = \frac{1}{2}(J_i - iK_i) = 0$ implies that $J_i = iK_i$, and since Y_i^+ is the 2x2 representation obeying the SU(2) algebra, $Y_i^+ = \frac{\sigma_i}{2}$. Putting this together,

$$\begin{aligned} J_i &= iK_i \\ \rightarrow Y_i^+ &= \frac{1}{2}(J_i + iK_i) = \frac{\sigma_i}{2} = \frac{1}{2}(J_i + J_i) = J_i \\ \rightarrow J_i &= \frac{1}{2}\sigma_i \\ \rightarrow K_i &= \frac{-i\sigma_i}{2}. \end{aligned} \tag{3-43}$$

Finally, the finite Lorentz transformations for the $(\frac{1}{2}, 0)$ representation are given by

$$R^{(L)} = e^{i\theta_i J_i} = e^{i\theta_i \frac{\sigma_i}{2}} \tag{3-44}$$

for rotations, and

$$B^{(L)} = e^{i\phi_i K_i} = e^{\phi_i \frac{\sigma_i}{2}} \tag{3-45}$$

for boosts. These act on 2x1 objects called left-chiral spinors, \mathcal{L} . A general Lorentz transformation on a left-chiral spinor can be written

$$\Lambda^{(L)} = e^{\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i}. \tag{3-46}$$

Replacing K_i with $-K_i$ takes Y_i^+ to Y_i^- , and gives the finite Lorentz transformations for the $(0, \frac{1}{2})$ representation

$$R^{(R)} = e^{i\theta_i J_i} = e^{\frac{i}{2}\theta_i \sigma_i} \tag{3-47}$$

for rotations, and

$$B^{(R)} = e^{i\phi_i K_i} = e^{-\phi_i \frac{\sigma_i}{2}} \quad (3-48)$$

for boosts. These act on 2×1 objects called right-chiral spinors, \mathcal{R} . The general Lorentz transformation on a right-chiral spinor can be written

$$\Lambda^{(R)} = e^{\frac{i}{2}\theta_i \sigma_i - \frac{1}{2}\phi_i \sigma_i}. \quad (3-49)$$

Last but not least, rank 2 spinors are given by the $(\frac{1}{2}, \frac{1}{2})$ representation. This representation is a tensor combining two spinors via outer product, and this is the representation that describes four vectors.

$$\alpha = \mathcal{L}\mathcal{R}^T \quad (3-50)$$

In order to transform α , both \mathcal{L} and \mathcal{R} must be transformed.

$$\alpha' = \Lambda^{(L)} \mathcal{L} \mathcal{R}^T \Lambda^{(R)T} = e^{\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i} \mathcal{L} \tilde{\mathcal{R}}^T e^{\frac{i}{2}\theta_i \sigma_i^T - \frac{1}{2}\phi_i \sigma_i^T} \quad (3-51)$$

However, in order to preserve the hermiticity of the operator α – an observable operator should remain observable under a Lorentz transformation – it's important that the transformation term on the right side is the Hermitian conjugate of the transformation on the left side. This would be the case if σ_i^T was $-\sigma_i^\dagger = -(\sigma_i^*)^T$, and this requires a transformation that turns σ into $-\sigma^*$. So \mathcal{R} is rearranged such that $\mathcal{R} \rightarrow \tilde{\mathcal{R}} = t\mathcal{R}$ where $t\sigma_i t^{-1} = -\sigma_i^*$. The matrix $t = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ satisfies the requirements. This change of basis for \mathcal{R} redefines α ,

$$\alpha = \mathcal{L} \tilde{\mathcal{R}}^T. \quad (3-52)$$

Defined in this manner, the $(\frac{1}{2}, \frac{1}{2})$ representation now transforms like so,

$$\alpha' = e^{\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i} \mathcal{L} \tilde{\mathcal{R}}^T e^{-\frac{i}{2}\theta_i \sigma_i^\dagger + \frac{1}{2}\phi_i \sigma_i^\dagger}. \quad (3-53)$$

Considering the fact that $\sigma_i^\dagger = \sigma_i$, the transformation reduces further,

$$\alpha' = e^{\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i} \alpha e^{-\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i}. \quad (3-54)$$

A transformation of the form $M' = HMH^\dagger$ where H is a Hermitian matrix, preserves the Hermitivity of the matrix M . Namely, M' will be Hermitian if M is Hermitian,

$$M'^\dagger = (HMH^\dagger)^\dagger = HM^\dagger H^\dagger = HMH^\dagger = M'. \quad (3-55)$$

On the other hand, M' will be anti-Hermitian if M is anti-Hermitian,

$$M'^\dagger = (HMH^\dagger)^\dagger = HM^\dagger H^\dagger = H(-M)H^\dagger = -HMH^\dagger = -M'. \quad (3-56)$$

The transformation of the $(\frac{1}{2}, \frac{1}{2})$ object α is purposely defined this way to preserve hermitivity and observability. Note that a general complex matrix can be broken up into a Hermitian piece and an anti-Hermitian piece, and that these pieces remain independent under the $(\frac{1}{2}, \frac{1}{2})$ Lorentz transformations constructed here. Then note that a general complex 2×2 matrix has 8 free parameters, 4 from the Hermitian part and 4 from the anti-Hermitian part. Meanwhile, α has only 4, two from each spinor⁹. This means that a general α may be represented by the Hermitian space alone. An appropriate basis for this space is the collection of Pauli matrices plus the identity,

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \text{ and } \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3-57)$$

Thus any $(\frac{1}{2}, \frac{1}{2})$ object α may be written in terms of its four independent parameters like so,

$$\begin{aligned} \alpha &= \alpha_0 \sigma_0 + \alpha_1 \sigma_1 + \alpha_2 \sigma_2 + \alpha_3 \sigma_3 \\ \alpha &= \begin{pmatrix} \alpha_0 + \alpha_3 & \alpha_1 - i\alpha_2 \\ \alpha_1 + i\alpha_2 & \alpha_0 - \alpha_3 \end{pmatrix}. \end{aligned} \quad (3-58)$$

⁹ A spinor has 2 complex components yielding 4 parameters. Two constraints reduce the number of free parameters to 2. One constraint requires a magnitude of 1, and the other requires that the overall phase doesn't matter.

Boosting α along the z direction hints that the four components transform like a four vector.

$$\alpha' = e^{\frac{1}{2}\phi_3\sigma_3} \begin{pmatrix} \alpha_0 + \alpha_3 & \alpha_1 - i\alpha_2 \\ \alpha_1 + i\alpha_2 & \alpha_0 - \alpha_3 \end{pmatrix} e^{\frac{1}{2}\phi_3\sigma_3} \quad (3-59)$$

Exponentiating the σ_3 matrix and multiplying everything yields,

$$\begin{pmatrix} \alpha'_0 + \alpha'_3 & \alpha'_1 - i\alpha'_2 \\ \alpha'_1 + i\alpha'_2 & \alpha'_0 - \alpha'_3 \end{pmatrix} = \begin{pmatrix} e^{\frac{1}{2}\phi_3} & 0 \\ 0 & e^{-\frac{1}{2}\phi_3} \end{pmatrix} \begin{pmatrix} \alpha_0 + \alpha_3 & \alpha_1 - i\alpha_2 \\ \alpha_1 + i\alpha_2 & \alpha_0 - \alpha_3 \end{pmatrix} \begin{pmatrix} e^{\frac{1}{2}\phi_3} & 0 \\ 0 & e^{-\frac{1}{2}\phi_3} \end{pmatrix} \quad (3-60) \\ = \begin{pmatrix} e^{\phi_3}(\alpha_0 + \alpha_3) & (\alpha_1 - i\alpha_2) \\ (\alpha_1 + i\alpha_2) & e^{-\phi_3}(\alpha_0 - \alpha_3) \end{pmatrix}.$$

Then solving the systems of equations makes the transformation clearer,

$$\begin{aligned} \alpha'_1 &= \alpha_1 \\ \alpha'_2 &= \alpha_2 \\ \alpha'_0 &= (\cosh \phi_3)\alpha_0 + (\sinh \phi_3)\alpha_3 \\ \alpha'_3 &= (\sinh \phi_3)\alpha_0 + (\cosh \phi_3)\alpha_3. \end{aligned} \quad (3-61)$$

Finally the transformation can be written as a 4x4 matrix acting on the 4x1 four vector,

$$\alpha' = \begin{pmatrix} \cosh \phi_3 & 0 & 0 & \sinh \phi_3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh \phi_3 & 0 & 0 & \cosh \phi_3 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}. \quad (3-62)$$

The other transformations reproduce the usual 4x4 Lorentz transformations as well. In this way, four vectors are just rearranged versions of the $(\frac{1}{2}, \frac{1}{2})$ rank 2 spinors. In practice it's easier to work with 4 vectors as 4x1 column vectors, though one could use the $(\frac{1}{2}, \frac{1}{2})$ rank 2 spinor representation. With four vectors, spinors, and scalars in hand, the invariant Lagrangians describing the different types of particles can now be built. Before doing this, however, there is a brief aside about group representations, particles, and labels.

3.1.3.3 The Poincare group and particle labels

The Poincare group is the Lorentz group plus translations. Physics experiments should be the same with a rotated apparatus, an apparatus moving at a different constant velocity, and also at another location. Adding translations to the group amounts to adding another the set of translation generators P_i . In the "Particles, Symmetries, and Labels" section the argument was made that particles are things that observers agree on between frames. Distinguishable particles have labels that remain invariant. A spin $\frac{1}{2}$ particle with mass m remains a spin $\frac{1}{2}$ particle with mass m for all observers. As it turns out these labels relate to different subspaces in the representation of the group.

When the generators of the group are represented a certain way, say as some $N \times N$ matrices, there is a likelihood that there are some invariant subspaces. As an example, a representation of a group may act on a 5 dimensional space spanned by vectors e_1, e_2, e_3, e_4, e_5 . The group may always transform vectors in the e_1, e_2, e_3 subspace into one another, and those in e_4, e_5 into one another, but never mix up e_1, e_2, e_3 with e_4, e_5 . In this example the 5×5 operators could be decomposed into a 3×3 operator acting only on the e_1, e_2, e_3 subspace and a 2×2 operator acting only on the e_4, e_5 subspace. Each 5×5 member of the group could be written as a block diagonal matrix with a 3×3 block and a 2×2 block. The same goes for the generators. Since these subspaces retain their identity under the transformations of the group¹⁰, they may be considered different particles. The question is whether there are labels for the subspaces.

Labels in physics need to be measurable, and by the current understanding of quantum mechanics these must be eigenvalues of Hermitian operators. The operators that label these subspaces are called the Casimir operators and must be built from the generators of the group [Weinberg (2005); Wigner (1939)]. The operators should give the same values

¹⁰ This assumes that the 3×3 and 2×2 pieces have no invariant subspaces besides themselves and 0.

for e_1, e_2 , and e_3 since they transform into one another and represent the same particle, which implies that the Casimir labeling operator must be proportional to the identity in that subspace. The same goes for the operator in the e_4, e_5 subspace. The Casimir operator for $SU(2)$ is the J^2 operator which labels the spin of the particle. The Poincare group has two Casimir operators corresponding to two labels the mass, m , and the spin, j . Looking at the irreducible representations¹¹ of the Poincare group, the mass and spin arise naturally as labels for particles.

3.1.3.4 Building the Lagrangian for a free scalar particle

Analyzing the symmetries of the $SO(3)$ rotation group revealed the $SU(2)$ group, the 2×2 complex Pauli matrices, and the complex 2×1 spinors. Analyzing the symmetries of the Lorentz group revealed the $(\frac{1}{2}, 0)$ left-chiral spinors, the $(0, \frac{1}{2})$ right-chiral spinors, and the four vectors encoded in the $(\frac{1}{2}, \frac{1}{2})$ representation. Now, these pieces are put together to form relativistically invariant Lagrangians for free particles.

The $(0, 0)$ representation of the Lorentz group is a scalar, which means that it 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, $i\partial_t\psi = \frac{1}{2m}(-i\partial_x)(-i\partial_x)\psi$ ¹² 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

¹¹ This is the group theory term for the block diagonal pieces and the corresponding invariant subspaces. The irreducible representations, irreps, are those that can't be broken down in terms of smaller invariant subspaces and smaller block diagonal matrices. An arbitrary representation of the group is built from these irreps.

¹² Here the one dimensional case is presented for simplicity

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,

$$S = \int d^4x (c_0 + c_1\Phi^2 + c_2\partial_\mu\Phi\partial^\mu\Phi + c_3\partial_\mu\partial^\mu\Phi). \quad (3-63)$$

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

$$S = \int d^4x (c_1\Phi^2 + c_2\partial_\mu\Phi\partial^\mu\Phi). \quad (3-64)$$

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

$$\begin{aligned} 2c_2\partial_\mu\partial^\mu\Phi &= -2c_1\Phi \\ \rightarrow (-E^2 + \vec{P}^2)\Phi &= \frac{c_1}{c_2}\Phi = -m^2\Phi \\ \rightarrow \frac{c_1}{c_2} &= -m^2. \end{aligned} \quad (3-65)$$

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

$$S = \int d^4x \frac{1}{2} (\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2). \quad (3-66)$$

3.1.3.5 Building 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} and \mathcal{R} spinors of the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ representations and the four vector, ∂_μ , in a Lorentz invariant way. Moving forward in this regard, the \mathcal{L} , \mathcal{R} , and ∂_μ transformations are now analyzed to find the lowest order invariant combinations. The transformation for the left-chiral spinor is

$$\Lambda^{(L)} = e^{\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i}, \quad (3-67)$$

and the transformation for the right-chiral spinor is,

$$\Lambda^{(R)} = e^{\frac{i}{2}\theta_i \sigma_i - \frac{1}{2}\phi_i \sigma_i}. \quad (3-68)$$

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

$$(\Lambda^{(L)})^\dagger = e^{-\frac{i}{2}\theta_i \sigma_i^\dagger + \frac{1}{2}\phi_i \sigma_i^\dagger} = e^{-\frac{i}{2}\theta_i \sigma_i + \frac{1}{2}\phi_i \sigma_i}. \quad (3-69)$$

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,

$$(\mathcal{L}^\dagger \mathcal{R})' = \mathcal{L}^\dagger (\Lambda^{(L)})^\dagger \Lambda^{(R)} \mathcal{R} = \mathcal{L}^\dagger (\Lambda^{(R)})^{-1} \Lambda^{(R)} \mathcal{R} = \mathcal{L}^\dagger \mathcal{R} \quad (3-70)$$

$$(\mathcal{R}^\dagger \mathcal{L})' = \mathcal{R}^\dagger (\Lambda^{(R)})^\dagger \Lambda^{(L)} \mathcal{L} = \mathcal{R}^\dagger (\Lambda^{(L)})^{-1} \Lambda^{(L)} \mathcal{L} = \mathcal{R}^\dagger \mathcal{L}. \quad (3-71)$$

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. Recall that

a four vector may be expressed as an outer product of left and right spinors,

$$\alpha_{L\tilde{R}} = \mathcal{L}_\alpha \tilde{\mathcal{R}}_\alpha^T. \quad (3-72)$$

The following also works,

$$\alpha_{R\tilde{L}} = \mathcal{R}_\alpha \tilde{\mathcal{L}}_\alpha^T. \quad (3-73)$$

Recall that the \sim transformation sent σ_i to $-\sigma_i^*$. So the transformations for the \sim transposed spinors are,

$$(\Lambda^{(\tilde{L})})^T = e^{\frac{i}{2}\theta_i(-\sigma_i^*)^T + \frac{1}{2}\phi_i(-\sigma_i^*)^T} = e^{-\frac{i}{2}\theta_i\sigma_i^\dagger - \frac{1}{2}\phi_i\sigma_i^\dagger} = e^{-\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i} \quad (3-74)$$

and

$$(\Lambda^{(\tilde{R})})^T = e^{\frac{i}{2}\theta_i(-\sigma_i^*)^T - \frac{1}{2}\phi_i(-\sigma_i^*)^T} = e^{-\frac{i}{2}\theta_i\sigma_i^\dagger + \frac{1}{2}\phi_i\sigma_i^\dagger} = e^{-\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i}. \quad (3-75)$$

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,

$$\mathcal{R}^\dagger \alpha_{L\tilde{R}} \mathcal{R} = \mathcal{R}^\dagger \mathcal{L}_\alpha \tilde{\mathcal{R}}_\alpha^T \mathcal{R} \quad (3-76)$$

and

$$\mathcal{L}^\dagger \alpha_{R\tilde{L}} \mathcal{L} = \mathcal{L}^\dagger \mathcal{R}_\alpha \tilde{\mathcal{L}}_\alpha^T \mathcal{L}. \quad (3-77)$$

The two types of four vectors have slightly different transformations

$$\alpha'_{L\tilde{R}} = e^{\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i} \alpha_{L\tilde{R}} e^{-\frac{i}{2}\theta_i\sigma_i + \frac{1}{2}\phi_i\sigma_i} \quad (3-78)$$

$$\alpha'_{R\tilde{L}} = e^{\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i} \alpha_{R\tilde{L}} e^{-\frac{i}{2}\theta_i\sigma_i - \frac{1}{2}\phi_i\sigma_i}. \quad (3-79)$$

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 . Behaving the same way under rotations implies that the x,y,z components mix up the same way in both types of four vector. In this respect, rotating x,y,z or $-x,-y,-z$ are both valid.

Exemplifying this, an infinitesimal rotation around z by $d\theta_z$ gives

$$\begin{aligned} x' &= x - yd\theta_z \\ y' &= y + xd\theta_z \end{aligned} \tag{3-80}$$

for positive x,y,z and

$$\begin{aligned} -x' &= -x - (-y)d\theta_z \\ -y' &= -y' + (-x)d\theta_z \end{aligned} \tag{3-81}$$

for $-x,-y,-z$. These transformations are equivalent. Moving onto boosts, the opposite ϕ sign naively implies that the x,y,z terms mix up with t in the opposite way for the two types of rank 2 spinors. However, if both $L\tilde{R}$ and $R\tilde{L}$ represent a four vector, then the boost must yield the same transformation on the components. Therefore, to get from $L\tilde{R}$ to $R\tilde{L}$ take t,x,y,z to $t,-x,-y,-z$. Illuminating this, an infinitesimal boost along x by $d\phi_x$ gives

$$\begin{aligned} t' &= t + xd\phi_x \\ x' &= x + td\phi_x \end{aligned} \tag{3-82}$$

for the $L\tilde{R}$ representation and

$$\begin{aligned} t' &= t + (-x)(-d\phi_x) \\ -x' &= -x + t(-d\phi_x) \end{aligned} \tag{3-83}$$

for $R\tilde{L}$. The negative ϕ and the negative spatial components counteract to transform the four vector correctly. The $R\tilde{L}$ rank 2 spinor encodes a four vector in the same way as a $L\tilde{R}$ rank 2 spinor, except that the spatial components are coded into the rank 2 spinor with the opposite

sign. Writing the representations out in terms of the four vector components yields

$$\begin{aligned}\alpha_{L\tilde{R}} &= \alpha_0\sigma_0 + \alpha_1\sigma_1 + \alpha_2\sigma_2 + \alpha_3\sigma_3 \\ \alpha_{L\tilde{R}} &= \alpha_\mu\sigma^\mu \\ \alpha_{L\tilde{R}} &= \begin{pmatrix} \alpha_0 + \alpha_3 & \alpha_1 - i\alpha_2 \\ \alpha_1 + i\alpha_2 & \alpha_0 - \alpha_3 \end{pmatrix}\end{aligned}\tag{3-84}$$

and

$$\begin{aligned}\alpha_{R\tilde{L}} &= \alpha_0\sigma_0 - \alpha_1\sigma_1 - \alpha_2\sigma_2 - \alpha_3\sigma_3 \\ \alpha_{R\tilde{L}} &= \alpha_\mu\bar{\sigma}^\mu \\ \alpha_{R\tilde{L}} &= \begin{pmatrix} \alpha_0 - \alpha_3 & \alpha_1 + i\alpha_2 \\ \alpha_1 - i\alpha_2 & \alpha_0 + \alpha_3 \end{pmatrix}.\end{aligned}\tag{3-85}$$

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 ($m^2\Phi^2 \leftrightarrow \partial_\mu\Phi\partial^\mu\Phi$)¹³. Following that example and throwing an i into the derivative terms¹⁴ to make them Hermitian gives the following Lagrangian,

$$S = \int d^4x (\mathcal{R}^\dagger\sigma^\mu i\partial_\mu\mathcal{R} + i\mathcal{L}^\dagger\bar{\sigma}^\mu i\partial_\mu\mathcal{L} - m\mathcal{L}^\dagger\mathcal{R} - m\mathcal{R}^\dagger\mathcal{L}).\tag{3-86}$$

¹³ 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 .

¹⁴ The i gives the appropriate factor for the definition of P_μ , $P_\mu = i\partial_\mu$.

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

$$\begin{aligned}\sigma^\mu i\partial_\mu \mathcal{R} &= m\mathcal{L} \\ \bar{\sigma}^\mu i\partial_\mu \mathcal{L} &= m\mathcal{R}. \\ \sigma^\mu i\partial_\mu \mathcal{R}^\dagger &= -m\mathcal{L}^\dagger \\ \bar{\sigma}^\mu i\partial_\mu \mathcal{L}^\dagger &= -m\mathcal{R}^\dagger.\end{aligned}\tag{3-87}$$

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.

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,

$$\begin{aligned}\sigma^\mu i\partial_\mu \frac{1}{m} \bar{\sigma}^\nu i\partial_\nu \mathcal{L} &= m\mathcal{L} \\ \rightarrow \sigma^\mu \bar{\sigma}^\nu P_\mu P_\nu &= m^2 \\ \rightarrow \eta_\nu^\mu P_\mu P_\nu &= m^2 \\ \rightarrow P^\mu P_\mu &= m^2 \\ \rightarrow E^2 - \vec{P}^2 &= m^2\end{aligned}\tag{3-88}$$

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

$$S = \int d^4x \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi,\tag{3-89}$$

after defining

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \bar{\sigma}^\mu & 0 \end{pmatrix}, \quad \psi = \begin{pmatrix} \mathcal{L} \\ \mathcal{R} \end{pmatrix}, \text{ and } \bar{\psi} = \psi^\dagger \gamma_0.\tag{3-90}$$

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

3.1.3.6 Building 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

$$S = \int d^4x (c_0 i \partial^\mu A^\nu i \partial_\mu A_\nu + c_1 i \partial^\mu A^\nu i \partial_\nu A_\mu + c_2 A^\mu A_\mu). \quad (3-91)$$

$$c_2 A^\nu = -\partial_\mu (c_0 \partial^\mu A^\nu + c_1 \partial^\nu A^\mu). \quad (3-92)$$

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 $\frac{1}{2}m^2$ ¹⁵. The action and equations of motion with the appropriate constants become,

$$S = \int d^4x \left(\frac{-1}{2} \partial^\mu A^\nu \partial_\mu A_\nu + \frac{1}{2} \partial^\mu A^\nu \partial_\nu A_\mu + \frac{m^2}{2} A^\mu A_\mu \right) \quad (3-93)$$

$$-m^2 A^\nu = \partial_\mu (\partial^\mu A^\nu - \partial^\nu A^\mu). \quad (3-94)$$

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

$$\begin{aligned} -m^2 \partial_\nu A^\nu &= \partial_\nu \partial_\mu \partial^\mu A^\nu - \partial_\nu \partial_\mu \partial^\nu A^\mu = 0 \\ \rightarrow \partial_\nu A^\nu &= 0 \\ \rightarrow -m^2 A^\nu &= \partial_\mu \partial^\mu A^\nu \\ \rightarrow m^2 A^\nu &= i \partial_\mu i \partial^\mu A^\nu \\ \rightarrow m^2 &= E^2 - \vec{P}^2. \end{aligned} \quad (3-95)$$

¹⁵ The order of the mass should match the order of the derivatives.

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 $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$

$$S = \int d^4x \left(\frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m^2}{2} A^\mu A_\mu \right). \quad (3-96)$$

3.1.4 Lagrangians in Quantum Mechanics and QFT

The Lagrangians for the different types of free particles have been derived. Now it's important to know how to use them to predict results that can be verified by experiment. Quantum behavior is different than the classical behavior covered earlier. Classically, each particle takes a single well defined path, but in the quantum regime particles behave as if they take all possible paths from some initial point to some final point. Analogously, the fields from which particles arise behave as if they take all possible configurations between states. Each possibility contributes to the net probability as a unit complex number with a phase given by the action [Feynman and Hibbs (1965)]. These contributions are called probability amplitudes,

$$\mathcal{A}_j = e^{iS_j}. \quad (3-97)$$

Adding the amplitudes for every possibility gives the net amplitude, and squaring the net amplitude gives the probability,

$$P = \left| \sum_j \mathcal{A}_j \right|^2. \quad (3-98)$$

Consider a free non-relativistic quantum particle in one dimension with the Lagrangian $\mathcal{L} = \frac{1}{2}m\dot{x}^2$. The amplitude for the particle to start at x_0, t_0 and end at x_f, t_f can be calculated by discretizing time, $t \in t_0, t_1, \dots, t_f$. A path is then determined by specifying x at each point in time, $x(t) \in x(t_0), x(t_1), \dots, x(t_f)$. In this way, the action for a path x , $S(x)$, can be written

$$S(x) = \int dt L = \frac{m}{2} \int_{t_0}^{t_f} dt \dot{x}^2 = \frac{m}{2} \sum_i \frac{(x(t_i) - x(t_{i-1}))^2}{(t_i - t_{i-1})^2} (t_i - t_{i-1}) = \frac{m}{2} \sum_i \frac{(x(t_i) - x(t_{i-1}))^2}{\Delta t}. \quad (3-99)$$

The amplitude for the path is then given by,

$$\mathcal{A}(x) = e^{i \frac{m}{2} \sum_i \frac{(x(t_i) - x(t_{i-1}))^2}{(t_i - t_{i-1})^2}} = e^{i \frac{m}{2} \frac{(x(t_f) - x(t_{f-1}))^2}{\Delta t}} e^{i \frac{m}{2} \frac{(x(t_{f-1}) - x(t_{f-2}))^2}{\Delta t}} \dots e^{i \frac{m}{2} \frac{(x(t_1) - x(t_0))^2}{\Delta t}}. \quad (3-100)$$

Summing over all paths, $x(t)$, with fixed endpoints, $x(t_0) = x_0$ and $x(t_f) = x_f$, provides the net amplitude

$$\begin{aligned} \mathcal{A} &= \sum_x \mathcal{A}(x) = \sum_{x(t_{f-1})} \sum_{x(t_{f-2})} \dots \sum_{x(t_1)} e^{i \frac{m}{2} \frac{(x(t_f) - x(t_{f-1}))^2}{\Delta t}} e^{i \frac{m}{2} \frac{(x(t_{f-1}) - x(t_{f-2}))^2}{\Delta t}} \dots e^{i \frac{m}{2} \frac{(x(t_1) - x(t_0))^2}{\Delta t}} \\ &= C \int dx_{t_{f-1}} dx_{t_{f-2}} \dots dx_{t_1} e^{i \frac{m}{2} \frac{(x(t_f) - x(t_{f-1}))^2}{\Delta t}} e^{i \frac{m}{2} \frac{(x(t_{f-1}) - x(t_{f-2}))^2}{\Delta t}} \dots e^{i \frac{m}{2} \frac{(x(t_1) - x(t_0))^2}{\Delta t}} \\ &= \int \mathcal{D}[x(t)] e^{i \frac{m}{2} \int_{t_0}^{t_f} dt \partial_t x \partial_t x}. \end{aligned} \quad (3-101)$$

Integrating by parts and using the average velocity $v = \frac{x_f - x_0}{t_f - t_0}$ for the boundary velocities yields,

$$\begin{aligned} \mathcal{A} &= \int \mathcal{D}[x(t)] e^{i \frac{m}{2} x \partial_t x|_{t_0}^{t_f} + i \frac{m}{2} \int_{t_0}^{t_f} dt (-x \partial_t^2 x)} \\ &= \mathcal{N} e^{i \frac{m}{2} (x(t_f) - x(t_0))v} \int \mathcal{D}[x(t)] e^{i \frac{m}{2} \int_{t_0}^{t_f} dt (-x \partial_t^2 x)}. \end{aligned} \quad (3-102)$$

The final solution for the propagator is obtained by repeatedly integrating by parts using the average values, $\frac{d^n a}{dt^n} = 0$, for subsequent boundary terms,

$$\begin{aligned} \mathcal{A} &= \mathcal{N} e^{i \frac{m}{2} (x(t_f) - x(t_0))v} \left[e^{i \frac{m}{2} [(x(t_f) - x(t_0))a + (x(t_f) - x(t_0))\dot{a} + \dots]} \right] \\ &= \mathcal{N} e^{\frac{-im(x_f - x_0)^2}{2(t_f - t_0)}} \\ &= \sqrt{\frac{m}{2i\pi(t_f - t_0)}} e^{\frac{-im(x_f - x_0)^2}{2(t_f - t_0)}}. \end{aligned} \quad (3-103)$$

The normalization \mathcal{N} is fixed by requiring that the probability for the particle to be somewhere at t_f is one. In other words, integrating over all x_f for constant x_0, t_0, t_f should give one. A more rigorous derivation of this propagator can be found in [Feynman and Hibbs \(1965\)](#) for example.

In general the amplitude to go from $\phi(t_0) = \phi_0$ to $\phi(t_f) = \phi_f$ is determined by the path integral,

$$\mathcal{A}(\phi_f, t_f; \phi_0, t_0) = \langle \phi_f, t_f | \phi_0, t_0 \rangle = \int \mathcal{D}[\cdot] e^{i \int_{t_0}^{t_f} dt L}. \quad (3-104)$$

Furthermore, the expectation value for an operation performed at time t_1 when the system starts in the state $\phi(t_0) = \phi_0$ and ends in $\phi(t_f) = \phi_f$ is

$$|\langle \phi_f, t_f | \mathcal{O}(t_1) |\phi_0, t_0 \rangle|^2 = \left| \int \mathcal{D}[\cdot] \mathcal{O}(t_1) e^{i \int_{t_0}^{t_f} dt L} \right|^2. \quad (3-105)$$

Of particular importance for QFT are scattering events. When two particles are measured at (t_1, \vec{x}_1) and (t_2, \vec{x}_2) some time in the past, what is the amplitude to measure two particles at (t_3, \vec{x}_3) and (t_4, \vec{x}_4) some time in the future? Considering that particles are in fact disturbances in the field, an equivalent question may be asked. How is the field amplitude correlated between spacetime points $(t_4, \vec{x}_4), (t_3, \vec{x}_3), (t_2, \vec{x}_2)$, and (t_1, \vec{x}_1) [Zee (2003)]. So for two to two scattering the amplitude is given by

$$\langle 0 | \psi_i(t_4, \vec{x}_4) \psi_j(t_3, \vec{x}_3) \psi_k(t_2, \vec{x}_2) \psi_l(t_1, \vec{x}_1) | 0 \rangle = \frac{\int \mathcal{D}[\cdot] \psi_i(t_4, \vec{x}_4) \psi_j(t_3, \vec{x}_3) \psi_k(t_2, \vec{x}_2) \psi_l(t_1, \vec{x}_1) e^{i \int_{-\infty}^{\infty} dt L}}{\int \mathcal{D}[\cdot] e^{i \int_{-\infty}^{\infty} dt L}}, \quad (3-106)$$

where ψ is a field, 0 denotes the vacuum state, and i, j, k, l specify the types of particles observed [Peskin and Schroeder (1995)]. Calculating this path integral for the full Lagrangian determines the amplitude for the scattering event. Usually the correlation functions can only be calculated for the free fields where the Lagrangian density is quadratic in the fields. Interactions generally spoil this simplicity, but even if the Lagrangian is more complicated, the path integral for the full theory can usually be expanded in terms of the correlation functions from the free field. However, this only works in practice if the additional terms in the Lagrangian are small compared to the free part of the Lagrangian. This method of expansion is called perturbation theory.

3.1.5 Perturbation Theory and Feynman Rules

Perturbation theory solves an intractable path integral with interactions by expanding it as a series of the solvable free solutions. When the magnitude of the interacting part is small compared to the free part of the Lagrangian, only the first few terms of the series expansion are needed. The best way to showcase perturbation theory is to run over the process for the scalar particle. The first step is to figure out the free solutions. The two point correlation

function for the free case is given by

$$\langle 0 | \Phi(t_2, \vec{x}_2) \Phi(t_1, \vec{x}_1) | 0 \rangle = \frac{\int \mathcal{D}[\Phi] \Phi(t_2, \vec{x}_2) \Phi(t_1, \vec{x}_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} (\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} (\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)}}, \quad (3-107)$$

and it describes a particle going from (t_1, \vec{x}_1) to (t_2, \vec{x}_2) . Integrating by parts and throwing away the boundary term turns the path integrals into Gaussians

$$\langle 0 | \Phi(t_2, \vec{x}_2) \Phi(t_1, \vec{x}_1) | 0 \rangle = \frac{\int \mathcal{D}[\Phi] \Phi(t_2, \vec{x}_2) \Phi(t_1, \vec{x}_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi (-\partial_\mu \partial^\mu - m^2) \Phi}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi (-\partial_\mu \partial^\mu - m^2) \Phi}}. \quad (3-108)$$

Focusing on the integral in the denominator and looking at the integral as a matrix multiplication simplifies things

$$\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi (-\partial_\mu \partial^\mu - m^2) \Phi} = \int d\Phi_1 d\Phi_2 \dots d\Phi_N e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j}. \quad (3-109)$$

Transforming to the eigenbasis yields a multidimensional Gaussian integral

$$\begin{aligned} \int d\Phi_1 d\Phi_2 \dots d\Phi_N e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j} &= \int d\tilde{\Phi}_1 d\tilde{\Phi}_2 \dots d\tilde{\Phi}_N e^{i \frac{1}{2} \tilde{\Phi}_i \tilde{K}_{ij} \tilde{\Phi}_j} = \int d\tilde{\Phi}_1 d\tilde{\Phi}_2 \dots d\tilde{\Phi}_N e^{\frac{-1}{2i} \tilde{k}_i \tilde{\Phi}_i^2} \\ &= \prod_{i=1}^N \left(\frac{2i\pi}{\tilde{k}_i} \right)^{\frac{1}{2}} = \frac{(2i\pi)^{N/2}}{(\det K)^{\frac{1}{2}}} \equiv Z_0. \end{aligned} \quad (3-110)$$

The solution uses the fact that the determinant is the product of eigenvalues. The integral from the numerator of Equation 3-108 can be dealt with by taking derivatives of the moment generating function

$$\int d\Phi_1 d\Phi_2 \dots d\Phi_N \Phi_k \Phi_l e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j} = \frac{1}{i} \frac{\delta}{\delta J_k} \frac{1}{i} \frac{\delta}{\delta J_l} \Big|_{J=0} \int d\Phi_1 d\Phi_2 \dots d\Phi_N e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j + i J_i \Phi_i}. \quad (3-111)$$

And the integral on the right of Equation 3-111 can be solved using the substitution $\Phi' = \Phi + K^{-1}J$ to complete the square,

$$\begin{aligned} \int d\Phi_1 d\Phi_2 \dots d\Phi_N e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j + i J_i \Phi_i} &= \int d\Phi'_1 d\Phi'_2 \dots d\Phi'_N e^{i \frac{1}{2} \Phi'_i K_{ij} \Phi'_j - \frac{i}{2} J_i K_{ij}^{-1} J_j} \\ &= \frac{(2i\pi)^{N/2}}{(\det K)^{\frac{1}{2}}} e^{\frac{-i}{2} J_i K_{ij}^{-1} J_j} \\ &= Z_0 e^{\frac{-i}{2} \int d^4x d^4y J(y) K^{-1}(y, x) J(x)}, \end{aligned} \quad (3-112)$$

where the last line reverts back to continuous functions and operators. Applying the derivatives provides the numerator of the two point correlation function

$$\begin{aligned} \int d\Phi_1 d\Phi_2 \dots d\Phi_N \Phi_k \Phi_l e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j} &= \frac{1}{i} \frac{\delta}{\delta J_k} \frac{1}{i} \frac{\delta}{\delta J_l} \Big|_{J=0} \int d\Phi_1 d\Phi_2 \dots d\Phi_N e^{i \frac{1}{2} \Phi_i K_{ij} \Phi_j + i J_i \Phi_i} \\ &= \frac{1}{i} \frac{\delta}{\delta J_k} \frac{1}{i} \frac{\delta}{\delta J_l} \Big|_{J=0} Z_0 e^{\frac{-i}{2} J_i K_{ij}^{-1} J_j} \\ &= Z_0 i K_{ij}^{-1}. \end{aligned} \quad (3-113)$$

Then plugging in the solutions for the numerator and denominator of Equation 3-108 shows that the two point correlation function is given by the Feynman propagator, \mathcal{D}_F ,

$$\begin{aligned} \langle 0 | \Phi(x_2) \Phi(x_1) | 0 \rangle &= \frac{\int \mathcal{D}[\Phi] \Phi(x_2) \Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}} \\ &= \frac{Z_0 i K^{-1}(x_2, x_1)}{Z_0} \\ &= i K^{-1}(x_2, x_1) = i(-\partial_\mu \partial^\mu - m^2)^{-1} = \mathcal{D}_F(x_2, x_1). \end{aligned} \quad (3-114)$$

In momentum space K is diagonal and $\mathcal{D}_F = \frac{i}{p^2 - m^2 + i\epsilon}$. Converting to the spacetime basis yields $\mathcal{D}_F(x_2, x_1) = \int \frac{d^4k}{\sqrt{2\pi}^4} \frac{ie^{-ik \cdot (x_2 - x_1)}}{p^2 - m^2 + i\epsilon}$, which represents a free particle propagating from one place to another.

The Feynman propagators are the building blocks for the higher order correlation functions. Take the four point function as an example,

$$\begin{aligned} \langle 0 | \Phi(x_4) \Phi(x_3) \Phi(x_2) \Phi(x_1) | 0 \rangle &= \frac{\int \mathcal{D}[\Phi] \Phi(x_4) \Phi(x_3) \Phi(x_2) \Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}} \\ &= \frac{\frac{1}{i} \frac{\delta}{\delta J(x_4)} \frac{1}{i} \frac{\delta}{\delta J(x_3)} \frac{1}{i} \frac{\delta}{\delta J(x_2)} \frac{1}{i} \frac{\delta}{\delta J(x_1)} \Big|_{J=0} \int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi + i J \Phi}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}} \\ &= \frac{\delta}{\delta J(x_4)} \frac{\delta}{\delta J(x_3)} \frac{\delta}{\delta J(x_2)} \frac{\delta}{\delta J(x_1)} \Big|_{J=0} e^{-\frac{1}{2} \int d^4x d^4y J(y) \mathcal{D}_F(y, x) J(x)} \\ &= \frac{\delta}{\delta J(x_4)} \frac{\delta}{\delta J(x_3)} \frac{\delta}{\delta J(x_2)} \frac{\delta}{\delta J(x_1)} \Big|_{J=0} \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{1}{2} \int d^4x d^4y J(y) \mathcal{D}_F(y, x) J(x) \right)^n. \end{aligned} \quad (3-115)$$

When $J=0$ only the term with four Js and four derivatives survives. The four derivatives hit the Js and leave terms like $\mathcal{D}_F(d, c) \mathcal{D}_F(b, a)$. In this case, there are $4!=24$ ways to rearrange

a,b,c, and d. However, the order of the spacetime points within the propagator doesn't matter nor does the order of the propagators, so this reduces the total to $\frac{4!}{2!2^2} = 3$ ways. The $\frac{1}{n!}$ and the $\frac{1}{2^n}$ from the expansion naturally take care of the degeneracy. The solution to the four point function is

$$\begin{aligned} \langle 0 | \Phi(x_4)\Phi(x_3)\Phi(x_2)\Phi(x_1) | 0 \rangle &= \frac{\delta}{\delta J(x_4)} \frac{\delta}{\delta J(x_3)} \frac{\delta}{\delta J(x_2)} \frac{\delta}{\delta J(x_1)} \Big|_{J=0} \frac{1}{8} \int d^4y_j d^4x_j d^4y_i d^4x_i J(y_j)J(x_j)J(y_i)J(x_i) \mathcal{D}_F(y_j, x_j) \mathcal{D}_F(y_i, x_i) \\ &= \mathcal{D}_F(x_4, x_3)\mathcal{D}_F(x_2, x_1) + \mathcal{D}_F(x_4, x_2)\mathcal{D}_F(x_3, x_1) + \mathcal{D}_F(x_3, x_2)\mathcal{D}_F(x_4, x_1). \end{aligned} \quad (3-116)$$

Any higher order correlation functions are built from the propagators in the same way: add up all possible combinations of the propagators for the appropriate order and account for the degeneracies.

This solves the free theory, which is nice, but it isn't that useful. Particles interact and the interacting solutions are the ones needed to describe scattering and decays. Fortunately, the interacting theories can be expanded in terms of the free correlations. For example, the two point correlation function for an interacting theory is given by,

$$\begin{aligned} \langle 0 | \Phi(x_2)\Phi(x_1) | 0 \rangle &= \frac{\int \mathcal{D}[\Phi] \Phi(x_2)\Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi + \mathcal{L}_{\text{int}}}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi + \mathcal{L}_{\text{int}}}} \\ &= \frac{\int \mathcal{D}[\Phi] \Phi(x_2)\Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} e^{i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{\text{int}}}}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} e^{i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{\text{int}}}} \\ &= \frac{\int \mathcal{D}[\Phi] \Phi(x_2)\Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} \sum_{n=0}^{\infty} \frac{1}{n!} \left(i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{\text{int}} \right)^n}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} \sum_{n=0}^{\infty} \frac{1}{n!} \left(i \int_{-\infty}^{\infty} d^4x \mathcal{L}_{\text{int}} \right)^n}, \end{aligned} \quad (3-117)$$

where \mathcal{L}_{int} may be written as a powerseries in Φ . Expanding this way, the numerator becomes a series of integrals of the form

$$\int d^4x_{i_1} d^4x_{i_2} \dots d^4x_{i_k} \int \mathcal{D}[\Phi] \Phi(x_{i_1}) \dots \Phi(x_{i_k}) \Phi(x_2) \Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}, \quad (3-118)$$

which are just correlation functions of the free field. Likewise, the denominator is also an expansion of the free correlations. In this way, any solution of the full theory may be written in terms of the free field solutions. Furthermore, when the interaction is small compared to the free part of the Lagrangian, the series expansions converge, and only a few terms are needed.

The simplest interacting theory is $\mathcal{L}_{\text{int}} = -\frac{\lambda}{4!}\Phi^4(x)$, and the two point correlation function describing the propagation of a particle is given by

$$\langle 0 | \Phi(x_2) \Phi(x_1) | 0 \rangle = \frac{\int \mathcal{D}[\Phi] \Phi(x_2) \Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\int_{-\infty}^{\infty} d^4z \frac{-i\lambda}{4!} \Phi(z)^4 \right)^n}{\int \mathcal{D}[\Phi] e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\int_{-\infty}^{\infty} d^4z \frac{-i\lambda}{4!} \Phi(z)^4 \right)^n}. \quad (3-119)$$

Consider the numerator up to first order in λ

$$\begin{aligned} N(\lambda^1) &= \int \mathcal{D}[\Phi] \Phi(x_2) \Phi(x_1) e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi} \left(1 + \int_{-\infty}^{\infty} d^4z \frac{-i\lambda}{4!} \Phi(z)^4 \right) \\ &= Z_0 \mathcal{D}_F(x_2, x_1) + \int_{-\infty}^{\infty} d^4z \frac{-i\lambda}{4!} \int \mathcal{D}[\Phi] \Phi(x_2) \Phi(x_1) \Phi(z)^4 e^{i \int_{-\infty}^{\infty} d^4x \frac{1}{2} \Phi(-\partial_\mu \partial^\mu - m^2) \Phi}. \end{aligned} \quad (3-120)$$

The first term is just the Feynman propagator, but the order λ term contains a correlation function with six Φ s. Solving the six point integral requires adding up all combinations of terms of the form $\mathcal{D}_F(a, b)\mathcal{D}_F(c, d)\mathcal{D}_F(e, f)$ while accounting for the appropriate degeneracies. There are $\frac{6!}{3!2^3} = 15$ possible propagator triples: 12 ways to put x and y in separate propagators and 3 ways to put them in the same one. With this information the perturbation series for the numerator is finally in hand,

$$\begin{aligned} N(\lambda^1) &= Z_0 \underbrace{\mathcal{D}_F(x_2, x_1)}_A + Z_0 \frac{12}{4!} (-i\lambda) \underbrace{\int_{-\infty}^{\infty} d^4z \mathcal{D}_F(x_2, z) \mathcal{D}_F(x_1, z) \mathcal{D}_F(z, z)}_B \\ &\quad + Z_0 \frac{3}{4!} (-i\lambda) \underbrace{\mathcal{D}_F(x_2, x_1) \int_{-\infty}^{\infty} d^4z \mathcal{D}_F(z, z) \mathcal{D}_F(z, z)}_C. \end{aligned} \quad (3-121)$$

The perturbation series for any n -point correlation function can be represented in terms of Feynman diagrams, which assign symbols to the different pieces of math. In this simple

scalar interacting theory, a line represents a factor of $\mathcal{D}_F(a, b)$ ($\mathcal{D}_F(z, z)$ if it's a loop) and an internal vertex representing an interaction, provides a factor of $-i\lambda \int d^4z$. The Feynman diagrams for $N(\lambda^1)$ are shown in Figure 3-2. In those diagrams, A and B are fully connected while C is disconnected. The separation reveals the fact that the pieces may be calculated independently and then multiplied. The fully connected pieces are the fundamental expressions from which any correlation function may be derived. There are two types of connected diagrams, those with external points and those with no external points. Those with zero external points represent vacuum fluctuations. The denominator is a normalization that serves

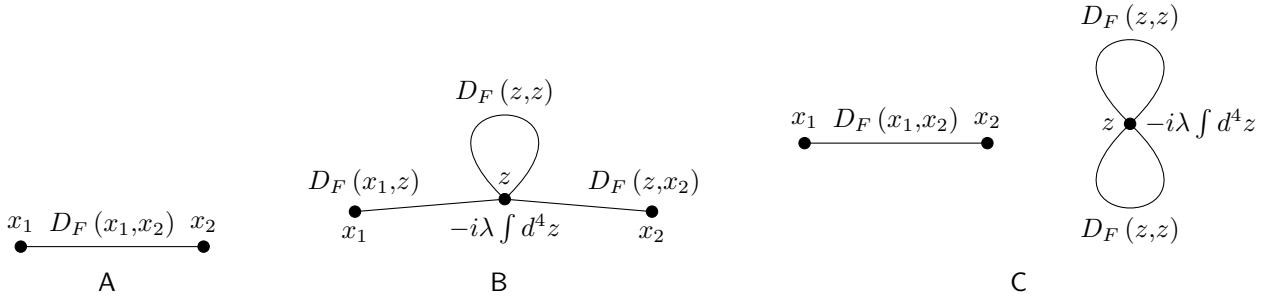


Figure 3-2. Feynman diagrams for $N(\lambda^1) = Z_0A + Z_0 \frac{12}{4!}B + Z_0 \frac{3}{4!}C$.

to eliminate Z_0 and the disconnected diagrams with vacuum fluctuations like C. It removes the contributions where the vacuum fluctuates independently from the process of interest. Finally, the propagator for the fully interacting theory up to first order in λ is

$$\langle 0 | \Phi(x_2)\Phi(x_1) | 0 \rangle_{\lambda^1} = A + \frac{12}{4!}B = \mathcal{D}_F(x_2, x_1) + \frac{12}{4!}(-i\lambda) \int_{-\infty}^{\infty} d^4z \mathcal{D}_F(x_2, z) \mathcal{D}_F(x_1, z) \mathcal{D}_F(z, z). \quad (3-122)$$

The factors out front are called symmetry factors and provide a weight for each diagram. In this theory the symmetry factor for a diagram is the number of nondegenerate ways to place pairs of points into the propagators with a factor of by $\frac{1}{4!}$ for every interaction vertex.

The Feynman rules for a theory can be used to build the perturbation series for any process, and this is why they are so ubiquitous. Any n-point correlation function is the

weighted¹⁶ sum of all possible diagrams with n external points – excluding those with disconnected vacuum fluctuations. For $2 \rightarrow 2$ scattering, the correlation function to first order requires the six diagrams shown in Figure 3-3. The first three diagrams are of the form

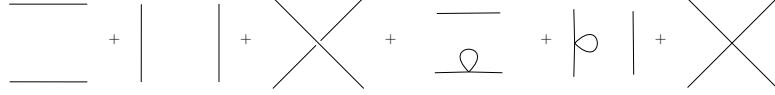


Figure 3-3. The Feynman diagrams for $\langle 0 | \Phi(x_4)\Phi(x_3)\Phi(x_2)\Phi(x_1) | 0 \rangle_{\lambda^1}$ representing the matrix element for $2 \rightarrow 2$ scattering up to first order.

A^*A , and these represent the three ways to put four points into $\mathcal{D}_F\mathcal{D}_F$. The next two terms are of the form A^*B , and the last one is a new diagram where the particles actually scatter instead of propagating separately. And the last term is built from the following factors: the interaction vertex provides a factor of $-i\lambda \int d^4z$ and the four propagators provide a factor of $\mathcal{D}_F(x_1, z)\mathcal{D}_F(x_2, z)\mathcal{D}_F(x_3, z)\mathcal{D}_F(x_4, z)$. The symmetry factor is $1 = \frac{4!}{4!}$ with the $\frac{1}{4!}$ from the internal vertex and the $4!$ for the ways to place the z s among x_1, x_2, x_3, x_4 . All of this results in a mathematical expression for the last diagram,

$$D = -i\lambda \int d^4z \mathcal{D}_F(x_1, z)\mathcal{D}_F(x_2, z)\mathcal{D}_F(x_3, z)\mathcal{D}_F(x_4, z). \quad (3-123)$$

Summing the six diagrams yields the full four point correlation function to first order,

$$\begin{aligned} \langle 0 | \Phi(x_4)\Phi(x_3)\Phi(x_2)\Phi(x_1) | 0 \rangle_{\lambda^1} &= \mathcal{D}_F(x_1, x_3)\mathcal{D}_F(x_2, x_4) + \mathcal{D}_F(x_1, x_2)\mathcal{D}_F(x_3, x_4) + \mathcal{D}_F(x_1, x_4)\mathcal{D}_F(x_2, x_3) \\ &\quad + \mathcal{D}_F(x_1, x_3)\frac{12}{4!}(-i\lambda) \int_{-\infty}^{\infty} d^4z \mathcal{D}_F(x_2, z)\mathcal{D}_F(x_4, z)\mathcal{D}_F(z, z) \\ &\quad + \mathcal{D}_F(x_2, x_4)\frac{12}{4!}(-i\lambda) \int_{-\infty}^{\infty} d^4z \mathcal{D}_F(x_1, z)\mathcal{D}_F(x_3, z)\mathcal{D}_F(z, z) \\ &\quad + -i\lambda \int d^4z \mathcal{D}_F(x_1, z)\mathcal{D}_F(x_2, z)\mathcal{D}_F(x_3, z)\mathcal{D}_F(x_4, z). \end{aligned} \quad (3-124)$$

¹⁶ The weights are the symmetry factors.

The Feynman rules for the actual interactions of the Standard Model are derived from the interacting Lagrangians in a similar way.

3.1.6 Interactions

The spin 1 and spin $\frac{1}{2}$ Lagrangians were obtained after a lengthy adventure through the Lorentz group symmetries. Now the goal is to couple them to produce theories where fermions and bosons interact as observed in nature. 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

$$\begin{aligned} S &= \int d^4x \frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m_A^2}{2} A^\mu A_\mu + \bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi + q\bar{\psi}\gamma^\mu A_\mu\psi \\ &= \int d^4x \frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{m_A^2}{2} A^\mu A_\mu + \bar{\psi} [i\gamma^\mu (\partial_\mu - iqA_\mu) - m] \psi \end{aligned} \quad (3-125)$$

To describe electromagnetism, the photon field should be massless and the interacting 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

$$S = \int d^4x \frac{-1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi} [i\gamma^\mu (\partial_\mu - iqA_\mu - iq\partial_\mu\alpha) - m] \psi, \quad (3-126)$$

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

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu\alpha(x) \quad (3-127)$$

and

$$\psi(x) \rightarrow e^{iq\alpha(x)}\psi(x) \quad (3-128)$$

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

$$\partial_\mu + iq\partial_\mu\alpha + c(A_\mu + a_\mu) = \partial_\mu + cA_\mu. \quad (3-129)$$

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,

$$S = \int d^4x \frac{-1}{4}(\partial_\mu A_\nu - \partial_\nu A_\mu)(\partial^\mu A^\nu - \partial^\nu A^\mu) + \bar{\psi}[i\gamma^\mu(\partial_\mu - iqA_\mu) - m]\psi, \quad (3-130)$$

effectively deriving the Lagrangian for Quantum Electrodynamics (QED).

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. If

$$\bar{\psi}U^\dagger i\gamma^\mu(\partial_\mu + cA'_\mu)U\psi = \bar{\psi}i\gamma^\mu(\partial_\mu + cA_\mu)\psi \quad (3-131)$$

is true when $A_\mu \rightarrow A'_\mu$ then the derivative term will be gauge invariant under SU(2). This is true when $(\partial_\mu + cA'_\mu)U\psi = U(\partial_\mu + cA_\mu)\psi$. The transformation $A_\mu \rightarrow UA_\mu U^\dagger + \Delta A_\mu$ helps simplify the conditions for gauge invariance,

$$\begin{aligned} (\partial_\mu + cA'_\mu)U\psi &= \partial_\mu(U\psi) + cA'_\mu U\psi \\ &= (\partial_\mu U)\psi + U(\partial_\mu\psi) + cA'_\mu U\psi \\ &= (\partial_\mu U)\psi + U(\partial_\mu\psi) + cUA_\mu U^\dagger U\psi + c\Delta A_\mu \\ &= U(\partial_\mu + cA_\mu)\psi + (\partial_\mu U)\psi + c\Delta A_\mu \\ &\rightarrow c\Delta A_\mu = -(\partial_\mu U)\psi = -(\partial_\mu U)U^\dagger U\psi = U\partial_\mu U^\dagger\psi \end{aligned} \quad (3-132)$$

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

$$cA_\mu(x) \rightarrow cUA_\mu(x)U^\dagger + U\partial_\mu U^\dagger. \quad (3-133)$$

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

$$\begin{aligned} cA_\mu(x) &\rightarrow cUA_\mu(x)U^\dagger + c\Delta A_\mu \\ &= cUA_\mu(x)U^\dagger - iq\frac{\sigma_i}{2}\partial_\mu\alpha_i(x). \end{aligned} \quad (3-134)$$

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 a general unitary transformation, there will be a force carrying particle for each generator, and an expansion $A_\mu = T^c A_\mu^c$. By representing A_μ in terms of the generators, the transformation for each component, A_μ^c , may be calculated for a general unitary transformation,

$$A_\mu^c \rightarrow A_\mu^c - f^{abc}\alpha^a A_\mu^b + \partial_\mu\alpha^c, \quad (3-135)$$

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 the D_μ part of the $SU(2)$ Lagrangian, but $F_{\mu\nu}$ still needs to be defined so that it remains invariant under equation 3-135. Defining

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - iq[A_\mu, A_\nu] \quad (3-136)$$

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

$$S = \int d^4x \mathcal{N} \text{tr}(F_{\mu\nu}F^{\mu\nu}) + \bar{\psi}i\gamma^\mu D_\mu\psi, \quad (3-137)$$

where $D_\mu = \partial_\mu - iqA_\mu$ is the covariant derivative. The Lagrangian holds for any unitary gauge transformation, $U(1)$ 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. For $SU(N)$ invariant theories the Lagrangian is given by,

$$S = \int d^4x \frac{-1}{2} \text{tr}(F_{\mu\nu}F^{\mu\nu}) + \bar{\psi}i\gamma^\mu D_\mu \psi = \int d^4x \frac{-1}{4} F_{\mu\nu}^a F_a^{\mu\nu} + \bar{\psi} [i\gamma^\mu (\partial_\mu - iqT^i A_\mu^i)] \psi. \quad (3-138)$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + qf^{abc} A_\mu^b A_\nu^c. \quad (3-139)$$

\mathcal{N} is set to $-\frac{1}{2}$ in order to get the appropriate $-\frac{1}{4}$ in front of the $F_{\mu\nu}^a$ tensors¹⁷.

The $SU(2)$ invariant Lagrangian of equation 3-138 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 comes into play.

3.1.7 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,

$$\frac{1}{2}m_A^2 A_\mu A^\mu \rightarrow \frac{1}{2}m_A^2 (UA_\mu U^\dagger + \Delta A_\mu)(UA^\mu U^\dagger + \Delta A^\mu). \quad (3-140)$$

Meanwhile for the matter particles, the gauge transformation acts on a column of fermions which leaves the $\bar{\psi}m\psi$ term invariant,

$$\begin{pmatrix} \bar{\psi}_1 \\ \dots \\ \bar{\psi}_n \end{pmatrix} m \begin{pmatrix} \psi_1 & \dots & \psi_n \end{pmatrix} \rightarrow \begin{pmatrix} \bar{\psi}_1 \\ \dots \\ \bar{\psi}_n \end{pmatrix} U^\dagger m U \begin{pmatrix} \psi_1 & \dots & \psi_n \end{pmatrix}, \quad (3-141)$$

¹⁷ The calculation uses the fact that the $SU(N)$ the generator matrices form an orthogonal basis, $\text{tr}(T^a T^b) = \frac{1}{2}\delta^{ab}$.

but this term restricts the fermions to the same mass. In order to describe the massive fermions and massive W and Z particles as 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,

$$\mathcal{L}_\phi^{\text{toy}} = (\partial^\mu \phi - iqA^\mu \phi)^\dagger (\partial_\mu \phi - iqA_\mu \phi) + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2. \quad (3-142)$$

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 $\phi \rightarrow \phi_0$

$$\mathcal{L}_I^{\text{toy}} = -\beta(L^\dagger \phi R + R^\dagger \phi L). \quad (3-143)$$

The toy examples of equations 3-142 and 3-143 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 3-144 is written in terms of left and right handed Dirac spinors. Left and right handed Dirac spinors are defined in terms of the left and right handed spinors as follows,
 $\psi_L = \begin{pmatrix} L \\ 0 \end{pmatrix}$ and $\psi_R = \begin{pmatrix} 0 \\ R \end{pmatrix}$. 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. In this regard, the Lagrangian is written to

respect the interchange of the left handed particles with the SU(2) transformation acting on $L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}$. The remaining right handed particles are 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,

$$\begin{aligned} \mathcal{L} = & (\partial^\mu \phi - iq_{w\phi} \frac{\sigma_i}{2} W_i^\mu \phi - iq_{b\phi} B^\mu \phi)^\dagger (\partial_\mu \phi - iq_{w\phi} \frac{\sigma_i}{2} W_\mu^i \phi - iq_{b\phi} B^\mu \phi) \\ & + \bar{L} i \gamma^\mu (\partial_\mu - iq_{wl} \frac{\sigma_i}{2} W_i^\mu - iq_{bl} B_\mu) L + \bar{R} i \gamma^\mu (\partial_\mu - iq_{br} B_\mu) R \\ & - \beta (\bar{\nu}_e \phi_+ e_R + \bar{e}_R \phi_+^* \nu_e + \bar{e}_L \phi_- e_R + \bar{e}_R \phi_-^* e_L) - \frac{1}{4} G_i^{\mu\nu} G_{\mu\nu}^i - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \\ & + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2. \end{aligned} \quad (3-144)$$

The SU(2) field, W_μ , is a complex 2x2 matrix expanded in terms of the Pauli matrices which operates on 2x1 complex column vectors. Therefore ϕ is written

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_- \end{pmatrix}. \quad (3-145)$$

To isolate the mass terms, the scalar field is expanded about its minimum, $|\phi| = \phi_0 = \sqrt{\frac{m_h^2}{\lambda}}$.

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 and that the strange mass-like coupling between the electron and neutrino disappears,

$$\beta(\bar{L} \phi R + \bar{R} \phi^\dagger L) = \beta(\bar{\nu}_e \phi_+ e_R + \bar{e}_R \phi_+^* \nu_e + \bar{e}_L \phi_- e_R + \bar{e}_R \phi_-^* e_L) \rightarrow \beta(\bar{e}_L \phi_0 e_R + \bar{e}_R \phi_0 e_L). \quad (3-146)$$

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,

$$\begin{aligned}
 D_\mu\phi &= (\partial_\mu - iq_{w\phi}\frac{\sigma_i}{2}W_\mu^i - iq_{b\phi}B_\mu)\phi \\
 &= \left[\begin{pmatrix} \partial_\mu - iq_{b\phi}B_\mu & 0 \\ 0 & \partial_\mu - iq_{b\phi}B_\mu \end{pmatrix} - \frac{i}{2}q_{w\phi} \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} \right] \begin{pmatrix} 0 \\ \phi_0 + \frac{1}{\sqrt{2}}h \end{pmatrix} \\
 &= -\frac{i}{2} \begin{pmatrix} q_{w\phi}\phi_0(W_\mu^1 - iW_\mu^2) + \frac{1}{\sqrt{2}}q_{w\phi}h(W_\mu^1 - iW_\mu^2) \\ i\sqrt{2}\partial_\mu h + \phi_0(2q_{b\phi}B_\mu - q_{w\phi}W_\mu^3) + \frac{1}{\sqrt{2}}h(2q_{b\phi}B_\mu - q_{w\phi}W_\mu^3) \end{pmatrix}.
 \end{aligned} \tag{3-147}$$

The ϕ_0^2 terms determine the masses,

$$(D^\mu\phi)^\dagger(D_\mu\phi) = q_{w\phi}^2\frac{\phi_0^2}{4}(W_\mu^1)^2 + q_{w\phi}^2\frac{\phi_0^2}{4}(W_\mu^2)^2 + \frac{\phi_0^2}{4}(q_{w\phi}W_\mu^3 - 2q_{b\phi}B_\mu)^2 + \text{other terms.} \tag{3-148}$$

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

$$m_w = \frac{q_{w\phi}\phi_0}{\sqrt{2}}, \quad m_z = \frac{m_w}{q_{w\phi}}, \quad \text{and} \quad m_\gamma = 0. \tag{3-149}$$

The photon, A_μ , is a linear combination of W_μ^3 and B_μ . The U(1) symmetry corresponding to electromagnetism is then

$$\begin{aligned}
 U_A &= e^{i(Q_{wi}T^3+Q_{bi})\alpha(x)} \\
 W_\mu^3 &\rightarrow W_\mu^3 + \frac{1}{g_w}\partial_\mu\alpha \\
 B_\mu &\rightarrow B_\mu + \frac{1}{g_b}\partial_\mu\alpha
 \end{aligned} \tag{3-150}$$

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

$$q_{wi} = Q_{wi}g_w \quad \text{and} \quad q_{bi} = Q_{bi}g_b. \tag{3-151}$$

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_3 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 $\frac{1}{2}$ and $-\frac{1}{2}$ respectively. Similarly, the eigenstates $\phi^+ = 0$ and $\phi^- = \phi_0 + h$ correspond to eigenvalues $\frac{1}{2}$ and $-\frac{1}{2}$.

In order to agree with experiment, the left handed particles and the scalar field are assigned $Q_{wi} = 1$, and the right handed particles are assigned $Q_{wi} = 0$. The electromagnetic charge Q_i and the isospin Q_{wi} values fix the remaining electroweak Q_{bi} values,

$$\begin{aligned} Q_{e_L} &= Q_{we_L} I_3 + Q_{be_L} = \frac{-1}{2} + Q_{be_L} \\ Q_{e_R} &= Q_{we_R} I_3 + Q_{be_R} = 0 + Q_{be_R} \\ Q_{\nu_e} &= Q_{w\nu_e} I_3 + Q_{b\nu_e} = \frac{1}{2} + Q_{b\nu_e} \\ Q_\phi &= Q_{w\phi} I_3 + Q_{b\phi} = \frac{-1}{2} + Q_{b\phi}, \end{aligned} \tag{3-152}$$

providing,

$$Q_{be_L} = \frac{-1}{2}, Q_{be_R} = -1, Q_{b\nu_e} = \frac{-1}{2}, \text{ and } Q_{b\phi} = \frac{+1}{2}. \tag{3-153}$$

The electroweak Lagrangian reduces to,

$$\begin{aligned} \mathcal{L} &= (\partial^\mu \phi - ig_w \frac{\sigma_i}{2} W_i^\mu \phi - \frac{i}{2} g_b B^\mu \phi)^\dagger (\partial_\mu \phi - ig_w \frac{\sigma_i}{2} W_i^\mu \phi - \frac{i}{2} g_b B^\mu \phi) \\ &\quad + \bar{L} i \gamma^\mu (\partial_\mu - ig_w \frac{\sigma_i}{2} W_i^\mu + \frac{i}{2} g_b B_\mu) L + \bar{R} i \gamma^\mu (\partial_\mu + ig_b B_\mu) R \\ &\quad - \beta (\bar{L} \phi R + \bar{R} \phi^\dagger L) - \frac{1}{4} G_i^{\mu\nu} G_{i\mu\nu}^i - \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + \frac{m_h^2}{2} \phi^\dagger \phi - \frac{\lambda}{4} (\phi^\dagger \phi)^2 \end{aligned} \tag{3-154}$$

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 3-154, the W^+ particle is $W_\mu^1 + iW_\mu^2$, and the W^- particle is $W_\mu^1 - iW_\mu^2$. As for the neutral bosons, 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.

3.2 The Standard Model Higgs and the LHC

The SM Higgs 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 per second, for a certain process, is given by the cross section times the luminosity, $R = \sigma_i * L$. 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 3-4.

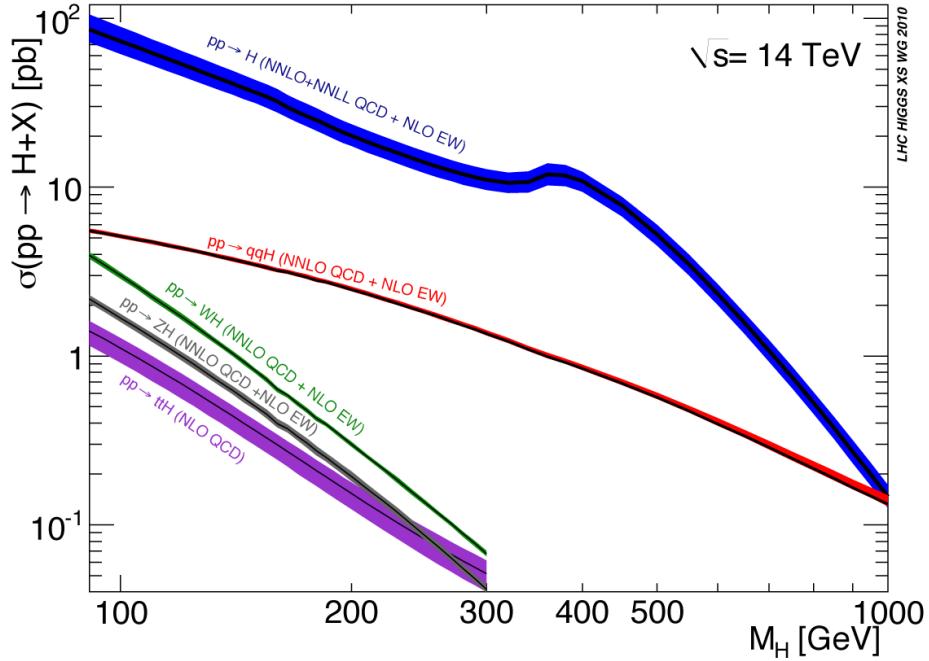


Figure 3-4. The highest production mode cross sections for the SM Higgs at 14 TeV [Group (2015)]

The Higgs cross sections are functions of the mass of the Higgs as well as the energy of the collisions. For a given collision energy, as in Figure 3-4, 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 modelled 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. With a larger proton momentum, the minimum creation energy for the process is a smaller fraction than before, and with more partons at this smaller fraction, it's as if a larger number with the necessary energy are now colliding. This effective increase in the density of energetic partons results in a growth of the cross section with collision energy.

If a SM Higgs is produced, it's predicted to decay in about 10^{-22} seconds¹⁸, which means that the particle itself can't be directly detected at the LHC, only the the decay products can. The SM decay probabilities for the different products are listed in Figure 3-5. These probabilities are determined by the coupling, which for the Higgs is the mass of the particle. With a stronger coupling to more massive particles, the decays to the more massive particles are more probable.

¹⁸ Assuming a 125 GeV SM Higgs boson

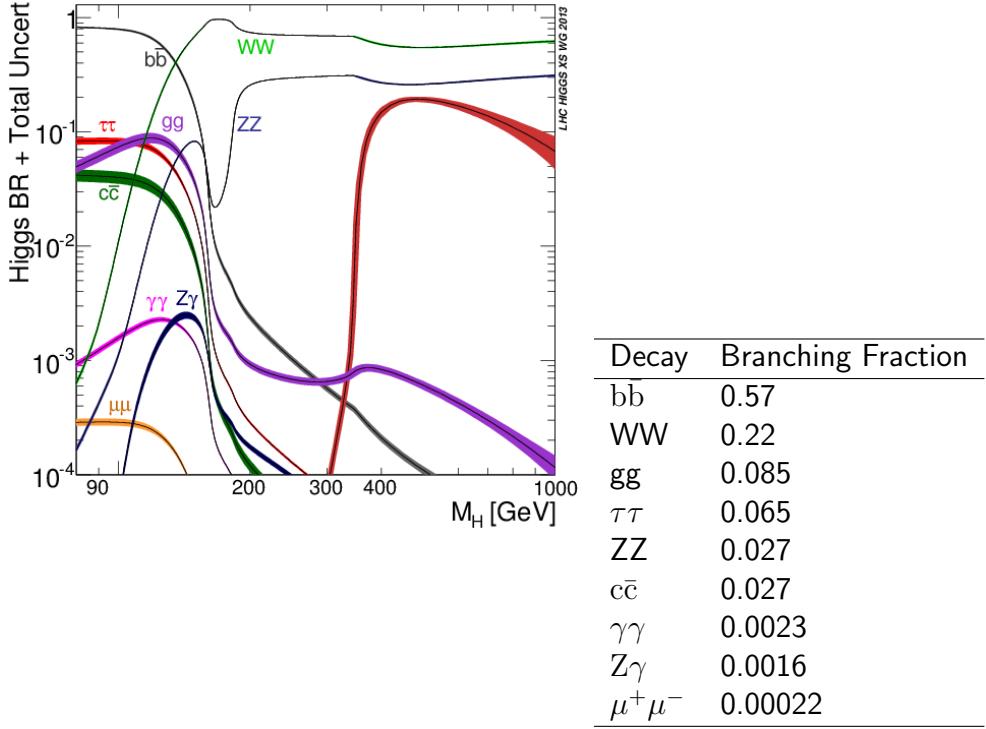


Figure 3-5. The graphic on the top left presents the SM Higgs branching fractions as functions of mass while the table on the bottom right displays the branching fractions for a 125 GeV SM Higgs [Group (2015)].

The muon has the lowest mass¹⁹ of the particles in Figure 3-5 and consequently $H \rightarrow \mu^+\mu^-$ has the lowest branching fraction in the set.²⁰ The gluons and photons are massless and do not couple to the Higgs at leading order, but through second order processes. Gluons interact with the Higgs through a loop of top quarks, as seen in Figure 3-6a. 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 through either a loop of W bosons or a loop of top quarks. Figure 3-6 shows the highest probability production mechanisms at the LHC. At $M_h = 125$ GeV, $\sqrt{s} = 13$ TeV, the GGF channel

¹⁹ excluding the photon and gluon

²⁰ The Higgs 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 in those modes.

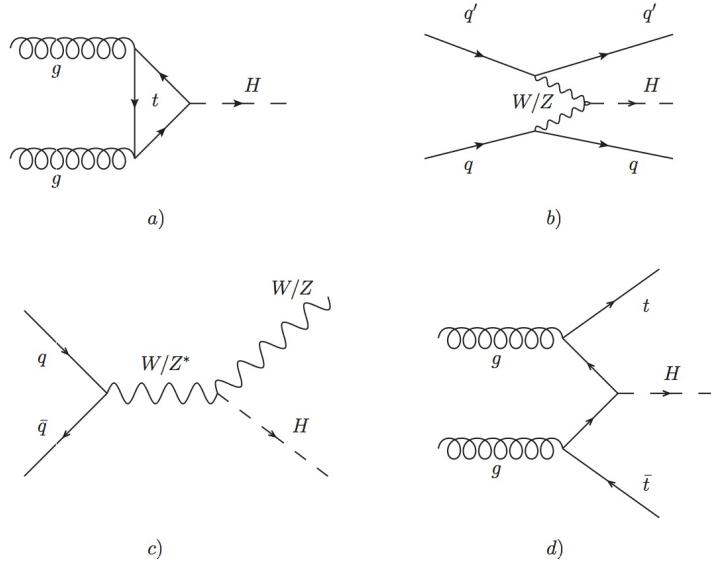


Figure 3-6. The SM production modes with the highest cross sections. a) Gluon Gluon Fusion (GGF) b) Vector Boson Fusion (VBF) c) Associated Production with a Vector Boson (VH) d) $t\bar{t}H$

comprises 87% of the total Higgs production cross section, VBF 7%, VH 4%, and $t\bar{t}H$ 1% [Group (2015)]. 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 3.2. 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

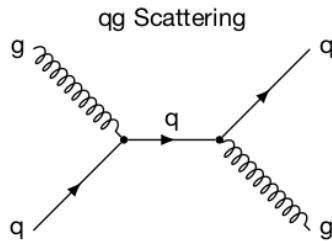


Figure 3-7. The quark-gluon background looks very similar to the GGF production channel when the Higgs decays to two jets. Protons are made of quarks and gluons so this process is extremely common in proton colliders like the LHC.

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. In this regard, $H \rightarrow b\bar{b}$ is 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

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

If the Higgs decays to $\mu^+ \mu^-$, the signal should be highly concentrated in the di muon mass spectrum. The mass remains invariant,

$$m_{\text{mother}} = m_{\text{sys}} = \sqrt{E_{\text{sys}}^2 - \vec{p}_{\text{sys}}^2}, \quad (4-1)$$

so, the mass of the decay products is concentrated at the mass of the mother particle.

However, the mass nature assigns to temporary particles is stochastic, and the mass spectrum centered at m_{mother} appears spread out in a Gaussian-like shape. Incorrect momentum measurements can make the spread worse. As a SM Higgs-like particle with a mass of 125 GeV has been observed in various other decays, the analysis searches for a peak in the dimuon mass spectrum from 120 to 130 GeV. The details of the analysis strategy are covered in Section 4.2, but first some of the necessary statistical methods are outlined in Section 4.1.

4.1 Discovery, Limits, and Measurement, and Sensitivity

This section briefly covers some of the statistical methods needed and provides context for the goals of 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. After covering these important topics, the analysis strategy for $H \rightarrow \mu^+ \mu^-$ is covered. The section starts with 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 experimentalist assumes 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,

$$p(x, N; \rho) = \frac{N!}{x!(N-x)!} \rho^x (1-\rho)^{N-x}. \quad (4-2)$$

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 , needed for discovery. If observing N_{heads} is rarer than the cutoff, h , i.e. $P(N_{\text{heads}} \geq h|\text{null}) < p_{\text{cutoff}}$, then the null hypothesis is ruled out. 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, when tossing an unbiased coin 100 times, observing 64 heads or greater should occur just more than 0.3% of the time, so observing 64 heads would be just too likely to invalidate the null. On the other hand, an experiment with an observation of 65 heads or greater should occur less than 0.3% of the time and anything greater than this is even more rare. Therefore, any experiment with $N_{\text{heads}} \geq 65$ would invalidate the null at 3σ and lead to an "observation".

Besides discovery, setting a limit is also important. Upon tossing a coin 100 times and finding $N_{\text{heads}} = 40$, the experimentalist may ask which values of p_{heads} are too high to yield an experiment with a value so low. In this case, all p_{heads} that predict experiments with 40 heads or fewer at too rare a probability may be ruled out at the $x\%$ level. For 3σ , $p_{\text{heads}} \geq 0.542$ may be ruled out as $p_{\text{heads}} \geq 0.542$ yields 40 heads or fewer less than 0.3% of the time, while smaller values for p_{heads} do not. Observing 40 heads in 100 tosses places an upper limit of $p_{\text{heads}} = 0.542$ on the bias of the coin at 3σ .

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 (NLL) is more convenient,

$$-\frac{\partial}{\partial p_{\text{heads}}} \ln[p(x, N; \rho)] = 0 \rightarrow \hat{\rho} = \frac{x}{N}. \quad (4-3)$$

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

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

Over many coin flipping experiments, the best fit value for $\hat{\rho}$ is the mean of the individual best fit values,

$$-\frac{\partial}{\partial p_{\text{heads}}} \ln[p] = 0 \rightarrow \hat{\rho} = \frac{1}{n} \sum_i \frac{x_i}{N_i} \quad (4-5)$$

The Poisson and Gaussian distributions are limits of the binomial with $\lambda = N\rho$ and $\mu = N\rho$, so the best fit parameters for λ and μ are also means of the individual experiments. This is why averaging the measurements of many experiments provides a better estimate of the true value.

In order to quantify the uncertainty of the measured value, limiting values of μ are computed. The idea is to find $\mu = \mu_{\text{lo}}$ such that observing anything higher than $\hat{\mu}$ occurs less than $x\%$ of the time, and to find a value μ_{hi} such that observing anything lower than $\hat{\mu}$ occurs less than $x\%$ of the time. This defines a confidence interval of $100\% - 2x\%$. Generally, a confidence interval of $1\sigma = 68\%$ is taken leaving $x = 16\%$. This construction guarantees that in many experiments the true value will be contained in the interval 68% of the time.

When the likelihood is Gaussian the interval is simply $\hat{\mu} \pm \sigma$. In the limit of large statistics, the likelihood becomes Gaussian, and the negative log-likelihood, simplifies to

$$\text{NLL} = -\ln[p] = -\ln \left[A e^{\frac{(x-\mu)^2}{2\sigma^2}} \right] = -\ln[A] + \frac{(x-\mu)^2}{2\sigma^2}. \quad (4-6)$$

Expanding about the minimum, $\hat{\mu}$, provides an estimate of σ and hence the confidence interval,

$$\text{NLL}(\hat{\mu}) + 0(x - \hat{\mu}) + \frac{1}{2} \text{NLL}''(\hat{\mu})(x - \hat{\mu})^2 = -\ln[A] + \frac{(x - \hat{\mu})^2}{2\sigma^2} \rightarrow \sigma^2 = \frac{1}{\text{NLL}''(\hat{\mu})}. \quad (4-7)$$

Note that σ may be determined by moving x away from the minimum until $\Delta\text{NLL} = 1$. The $\Delta\text{NLL} = 1$ method is sometimes used as an estimate of the uncertainty even when the likelihood is not Gaussian. In some cases the PDF may be multidimensional, and in those cases, σ^2 is the covariance matrix with $\partial_{\theta_i}\partial_{\theta_j}\text{NLL}(\hat{\theta}) = (\sigma^2)_{ij}^{-1}$. Multidimensional or otherwise, σ can be used to estimate the uncertainty on the best fit values.

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. Consider an experiment flipping a coin N times,

$$p(x, N; \rho) = \frac{N!}{x!(N-x)!} \rho^x (1-\rho)^{N-x}. \quad (4-8)$$

The experimenter may ask what p-value, given the null, 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

$$p(x = N\rho_i, N; \rho = \rho_{\text{null}}) = \frac{N!}{(N\rho_i)!(N - N\rho_i)!} \rho_{\text{null}}^{N\rho_i} (1 - \rho_{\text{null}})^{N - N\rho_i}. \quad (4-9)$$

With enough statistics, the binomial distribution may be approximated by a Gaussian with $\mu = N\rho$, and $\sigma = \sqrt{N\rho(1-\rho)}$. The p-values for a Gaussian are determined by the number of standard deviations away from the mean, Z . The larger Z is the smaller the p-value and the more sensitive the experiment. The sensitivity is given by $Z = \frac{(x-\mu)}{\sigma}$, and for a coin with ρ_i , the expected sensitivity is

$$Z = \frac{(x - \mu)}{\sigma} = \frac{\sqrt{N}(\rho_i - \rho_{\text{null}})}{\sqrt{\rho_{\text{null}}(1 - \rho_{\text{null}})}}. \quad (4-10)$$

Note that the sensitivity scales as the \sqrt{N} . In the limit of large statistics, the Poisson distribution may also be approximated as a Gaussian with $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$. In particle physics, one often tries to rule out the background only hypothesis by observing enough signal. If the Standard Model or some other theory predicts $\mu = S + B$ and the null predicts $\mu = B$.

Then the expected sensitivity for discovery is,

$$Z = \frac{(x - \mu)}{\sigma} = \frac{(S + B - B)}{\sqrt{B}} = \frac{S}{\sqrt{B}} = \frac{\sqrt{N}\rho_s}{\sqrt{\rho_b}}. \quad (4-11)$$

This also scales with the \sqrt{N} . So 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 NLL in the high statistics limit is,

$$-\ln\left[\frac{p}{C}\right] = -\ln\left[\prod_i p_i\right] = -\sum_i \ln[p_i] = -C + \sum \frac{(x_i - \mu_i)^2}{2\sigma_i^2}, \quad (4-12)$$

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

$$\chi_{\text{nll}}^2 = -2\ln\left[\frac{p}{C}\right] = \sum \frac{(x_i - \mu_i)^2}{\sigma_i^2}. \quad (4-13)$$

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

$$Z^2 = \frac{(x_{\text{nll}} - \mu_{\text{nll}})^2}{\sigma_{\text{nll}}^2} = \sum_i \frac{S_i^2}{B_i}. \quad (4-14)$$

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.

The test statistic t is a random variable with its own distribution,

$$t = -2\ln\left[\frac{p(x_i, \theta)}{p(x_i, \hat{\theta})}\right]. \quad (4-15)$$

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 4-12 sets the range of t from zero to infinity. When there is only one parameter of interest, μ , the likelihood may be reduced to one dimension by using the maximum likelihood estimates for the uninteresting parameters, θ . The method is called

"profiling" and provides the best fit for the one dimensional PDF along μ ,

$$t = -2\ln\left[\frac{p(x_i, \mu, \hat{\theta})}{p(x_i, \hat{\theta})}\right]. \quad (4-16)$$

The $\hat{\theta}$ parameters are the best fit values with a fixed μ .

The $H \rightarrow \mu^+ \mu^+$ search uses histograms to compare the expected number of events to the observed number of events, and the PDF for a given bin is described by the Poisson distribution. The net likelihood is therefore,

$$p(x_i, \mu, \theta) = \prod_i \text{Poisson}(x_i, \nu_i(\mu, \theta)), \quad (4-17)$$

where $\nu_i = \mu S_i + B_i$ represents the expected number of events in a bin. S_i and B_i are the expected SM signal and the background yields determined by some models with parameters θ . The factor μ is the signal strength, representing a theory with μ times the expected SM signal yield.

Sometimes previous experiments have been performed to determine the uncertainty on the parameters. In this case, the likelihood for the previous measurements on the different θ s can be included in the net likelihood,

$$p(x_i, \mu, \theta) = \prod_i \text{Poisson}(x_i, \nu_i(\mu, \theta)) \prod_j p(y_j, \theta_j), \quad (4-18)$$

where y_j represents the previously observed data. Usually y_j is summarized by the best fit $\bar{\theta}_j$ rather than keeping track of the actual data. In this case, the likelihood for the previous measurement is reduced to a function of $\theta_j, \bar{\theta}_j$ and σ_j and usually referred to as a "constraint". The net likelihood becomes,

$$p(x_i, \theta) = \prod_i \text{Poisson}(x_i, \nu_i(\theta)) \prod_j p(\bar{\theta}_j, \theta_j, \sigma_j). \quad (4-19)$$

Upon profiling, the previous measurement fights to keep the best fit for each θ_j near $\bar{\theta}_j$ with a strength depending on σ_j .

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 for the Standard Model against the background only null may be calculated using the expected yields for the S+B and the background only yields. Similarly, the background only expected upper limit on S may be calculated using t with S+B as the null and the background as the expected hypothesis. The background only expected upper limit at 95% confidence is calculated by finding the value of S high enough that observing the expected background has a p-value of 5%. A higher expected sensitivity and lower background only 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 for the SM against the null and to the expected upper limit on S when assuming the background-only as truth.

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, $p_{\text{bkg-only}}$ goes to 1 and the CLs upper limit approaches the standard upper limit. However, in general, $p_{\text{bkg-only}}$ is less than one, p_{CLs} is larger than p_μ , and the CLs method sets more conservative upper limits. In this way, only hypotheses with low p_μ that differ substantially from the background only will be ruled out.

4.2 The Analysis 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/pvalues. 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.

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 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, and concentrating the signal. To further improve the sensitivity, Boosted Decision Trees (BDTs) ? 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.

4.3 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 [Khachatryan et al. \(2017\)](#). 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 [Tapper and Acosta \(2013\)](#). 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 [Hoecker et al. \(2007\)](#), 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.

4.3.1 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.

4.3.2 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 noteable 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 4-1. 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.

4.3.3 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 4-1. 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.

4.3.4 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 4-1, 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 4-1. 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.

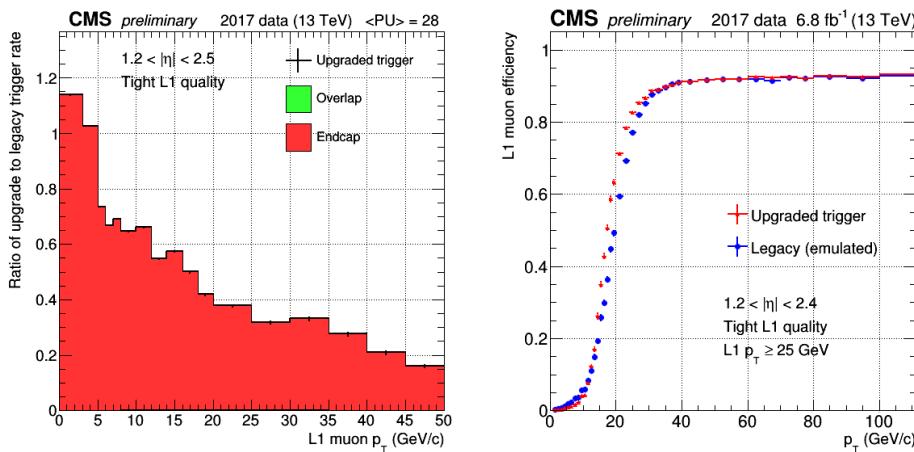


Figure 4-1. On the left, the upgraded EMTF rate divided by the legacy rate is shown for a variety of p_T thresholds. On the right, the upgraded and legacy efficiencies are presented for a 25 GeV threshold. The upgraded EMTF has a 3x lower rate than the legacy system at 25 GeV with virtually no difference in plateau efficiency for the same threshold. Plots are taken from [CMS Collaboration \(2017\)](#).

4.4 Data and Monte Carlo Samples

The proton-proton collision data used for the $H \rightarrow \mu^+ \mu^-$ search is listed in Table 4.4.

The signal processes in 4-2 are represented by the MC in Table 4-3. 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 4-4. 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.

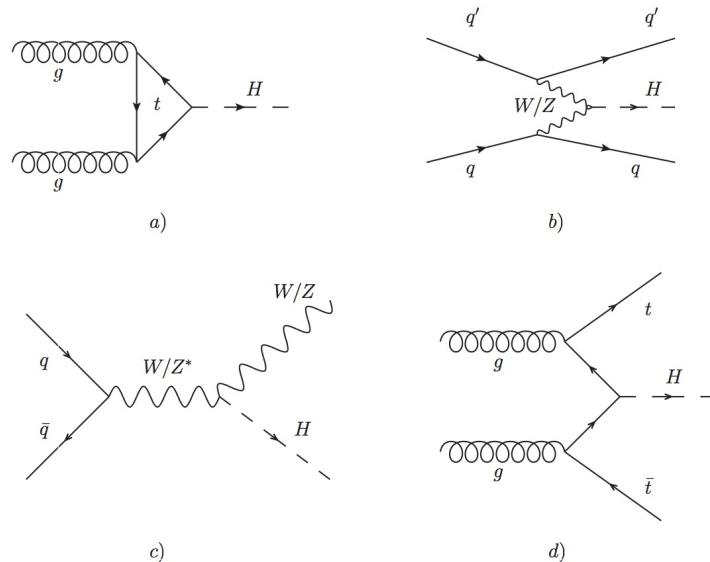


Figure 4-2. The SM production modes considered in the analysis. a) Gluon Gluon Fusion (GGF) b) Vector Boson Fusion (VBF) c) Associated Production with a Vector Boson (VH) d) Associated production with top quarks $t\bar{t}H$

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 4-2. 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.

Higgs signal MC samples	Events	Cross section [pb]	$\times \text{sec} \times \text{BR} [\text{fb}]$
/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	250000	48.58	10.571
/VBF_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	249200	3.7817	0.8229
/WPlusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	124547	0.09426	0.02051
/WWminusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	125000	0.05983	0.013019
/ZH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	249748	0.17762	0.03865

Table 4-3. The Higgs signal MC samples were generated with POWHEG while the parton shower and hadronization processes are modeled by the PYTHIA8 generator with TuneCUETP8M1.

Background MC		Events	Cross Section [pb]
Drell-Yan			
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext2_v1/MINIAODSIM	122055388	5765	
/DYToLL_0_1_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	49579613	4754	
/DYToLL_1_1_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	883.9		
/DYToLL_2_1_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	42324802	343.8	
/DYToLL_2_1_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	47974534	343.8	
SingleTop			
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	5372991	35.85	
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	3256650	35.85	
/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	5425134	35.85	
/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV_powheg_TuneCUETP8M1_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	3256407	35.85	
Top Pair			
/TT_Jets_Dilept_TuneCUETP8M1_13TeV_madgraphMLM_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	6094476	85.556	
/TT_Jets_Dilept_TuneCUETP8M1_13TeV_madgraphMLM_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	2435022	85.556	
/TT_Jets_Dilept_TuneCUETP8M2_14_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	1452920	85.556	
DiBoson			
/WWTo2L2Nu_13TeV_powheg_RunIISummer16MiniAODv2_FUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	1999000	12.46	
/WWTo3LNu_TuneCUETP8M1_13TeV_amcatnloFXFX_pythia8_RunIISummer16MiniAODv2_FUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	11887464	2.113	
/WZTo2L2Q_13TeV_amcatnloFXFX_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	2617272	4.409	
/ZZTo2L2Nu_13TeV_powheg_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	8842475	0.564	
/ZZTo2L2Q_13TeV_amcatnloFXFX_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	15345572	3.22	
/ZZTo4L_13TeV_amcatnloFXFX_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	10709784	1.212	
TriBoson			
/WWWW_4F_TuneCUETP8M1_13TeV_amcatnlo_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	240000	0.2086	
/WWZZ_TuneCUETP8M1_13TeV_amcatnlo_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	250000	0.1651	
/WZZ_TuneCUETP8M1_13TeV_amcatnlo_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	246800	0.05565	
/ZZZ_TuneCUETP8M1_13TeV_amcatnlo_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	249237	0.01398	
Single Top + X			
/tZq_1_14f_13TeV_amcatnlo_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	14509520	0.0758	
Top pairs			
/TTWJetsToLNu_TuneCUETP8M1_13TeV_amcatnloFXFX_madspin_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	2160168	0.2043	
/TTWJetsToLNu_TuneCUETP8M1_13TeV_amcatnloFXFX_madspin_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext2_v1/MINIAODSIM	3120397	0.2043	
/TTZToLNuNu_M-10_TuneCUETP8M1_13TeV_amcatnlo_pythia8_RunIISummer16MiniAODv2_PUMoriond17_80X_mcRun2_asymptotic_2016_TranchelV_v6_ext1_v1/MINIAODSIM	1992438	0.2529	

Table 4-4. 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.

4.5 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. In this regard, 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 4-3 as an example.

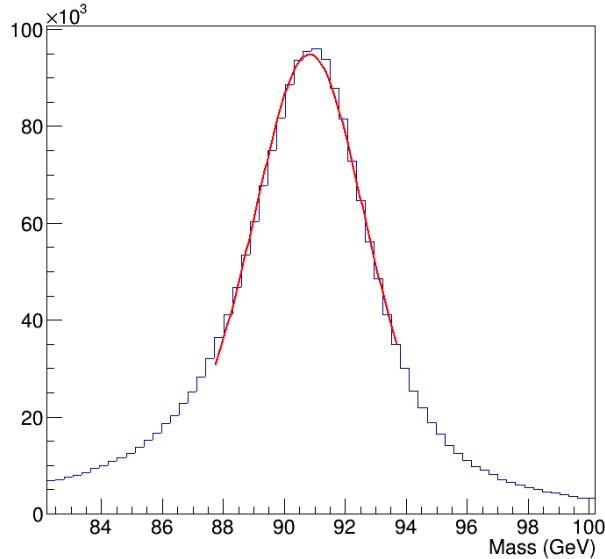


Figure 4-3. The dimuon mass distribution is fit around the Z peak for the positively charged muon with ϕ between 0 and 0.53.

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.

4.5.1 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 4-4 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 4-5 shows the Z resolution plotted against the same variables as in Fig. 4-4. 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.

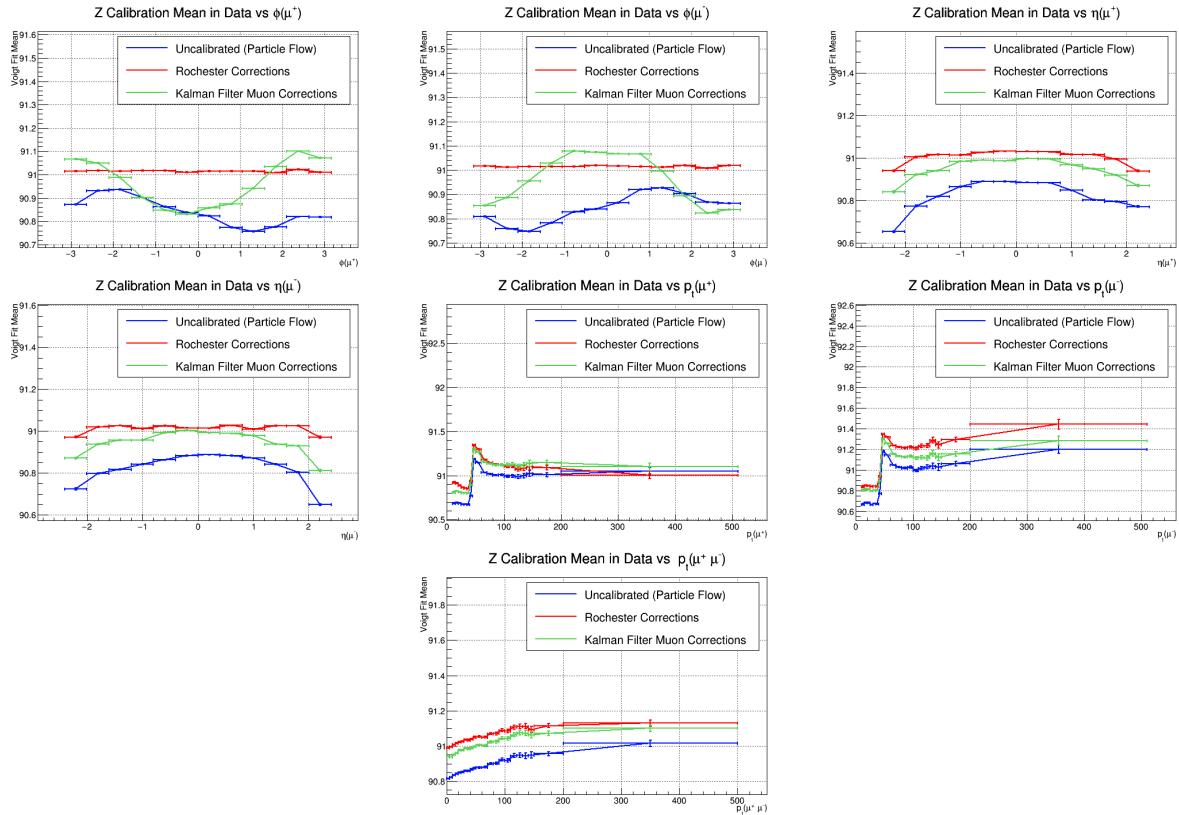


Figure 4-4. The Rochester (Roch) and Kalman Filter (KaMu) muon corrections are applied to data and compared to the uncalibrated Particle Flow (PF) predictions. The fitted mean of the Z peak is plotted vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for dimuon p_t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

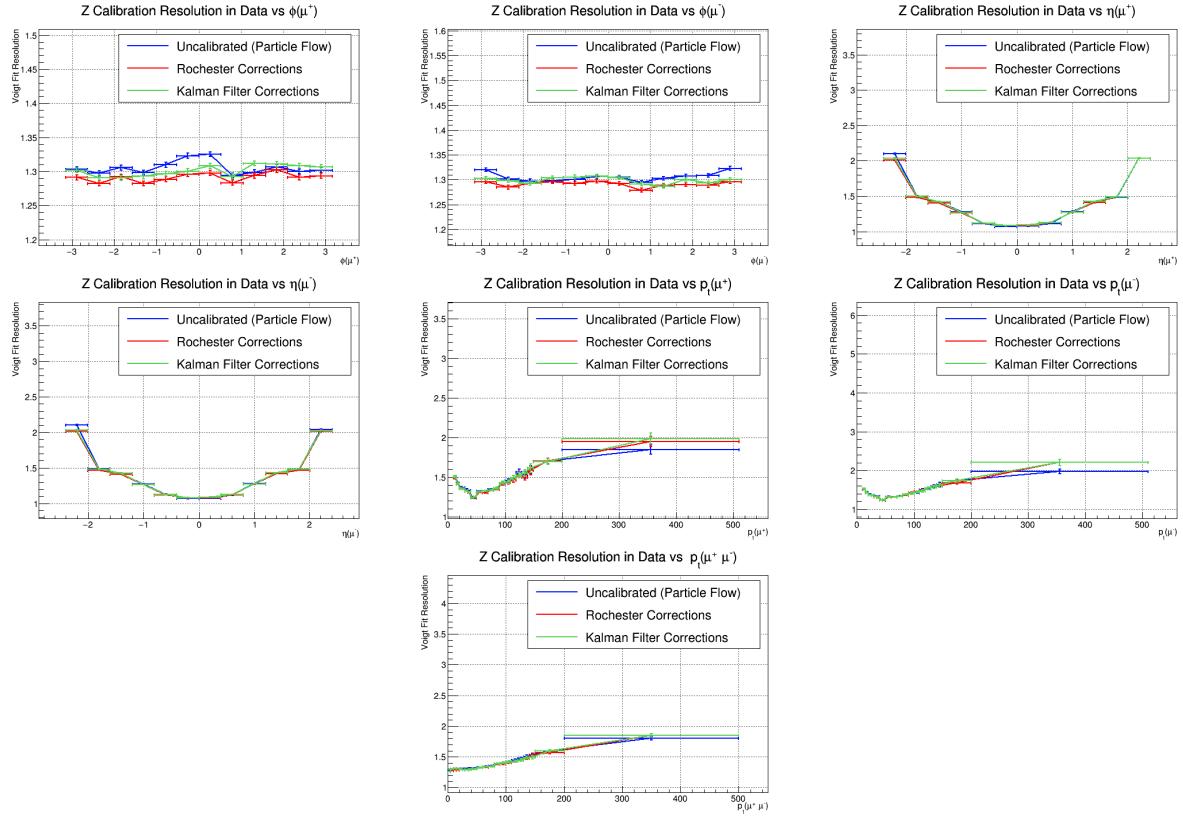


Figure 4-5. The Rochester (Roch) and Kalman Filter (KaMu) muon corrections are applied to data and compared to the uncalibrated Particle Flow (PF) measurement. The fitted Z peak resolution is plotted vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well.

4.5.2 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 at with the width and mass predicted by the MC. In this way, the limits will be set correctly for the different values of the Higgs mass. Figure 4-6 shows the Data/MC agreement on the Z peak mean before any corrections, while figure 4-7 shows the agreement after Rochester corrections and 4-8 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.

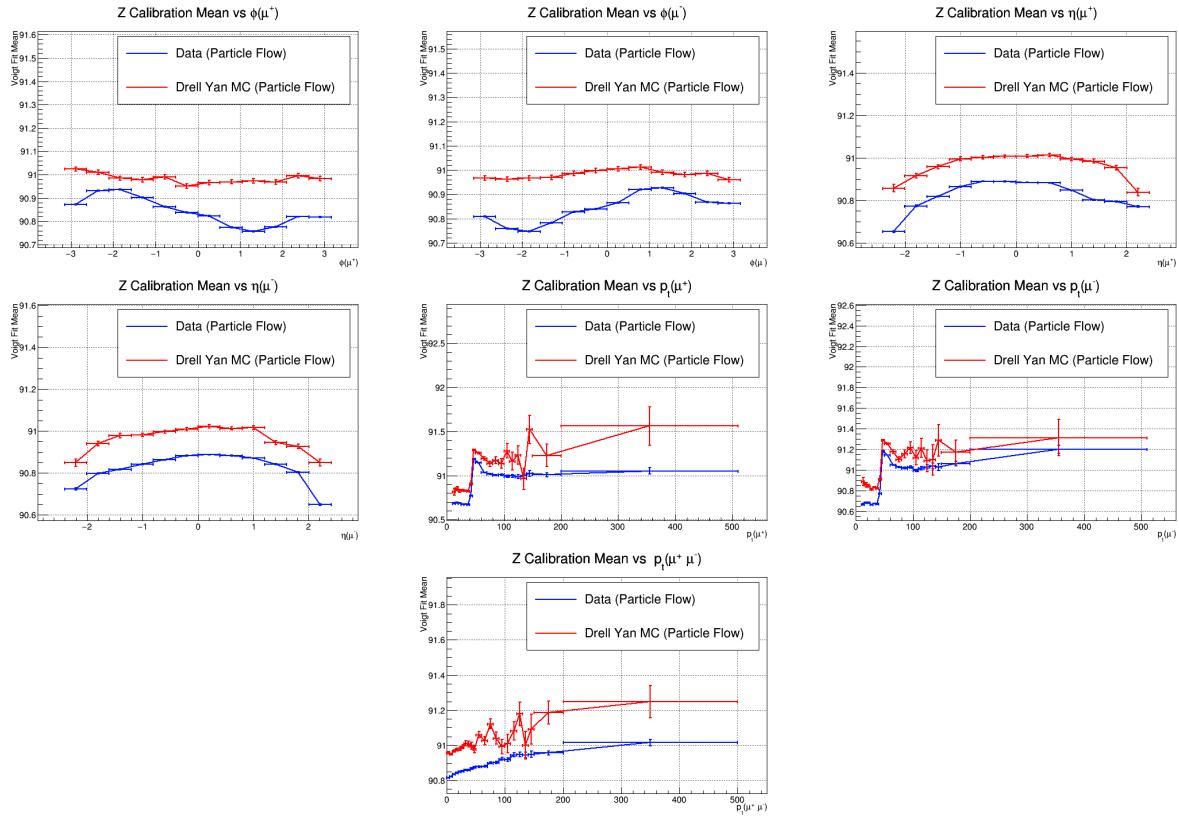


Figure 4-6. The fitted mean of the Z peak without any corrections is plotted in data and MC. The mean is plotted vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC are not aligned in terms of the Z peak mean before corrections.

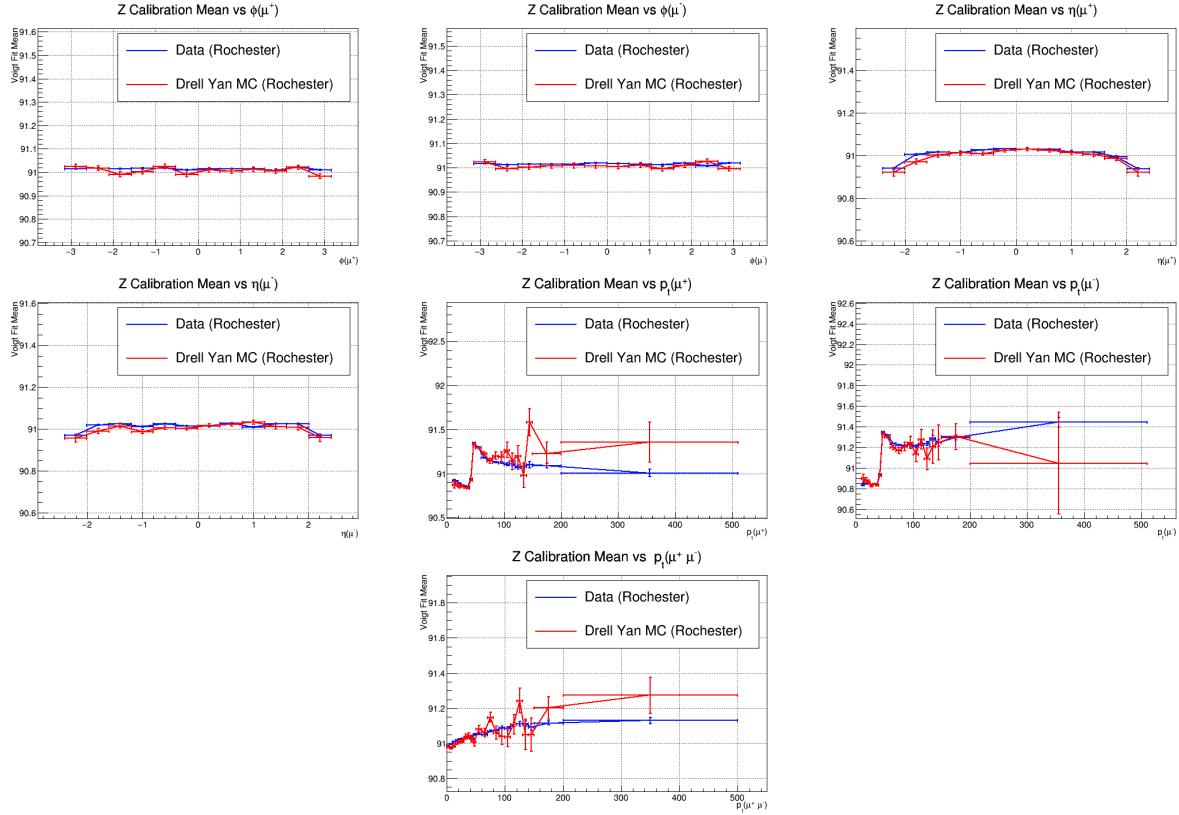


Figure 4-7. The fitted mean of the Z peak is plotted for the Rochester corrections. The mean is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC align very well in terms of the Z peak mean after applying the Rochester muon corrections.

Both corrections also succeed in aligning the Z peak resolution. The mismatch before corrections is seen in Figure 4-9 and the agreement after corrections is seen in Figure 4-10 and Figure 4-11.

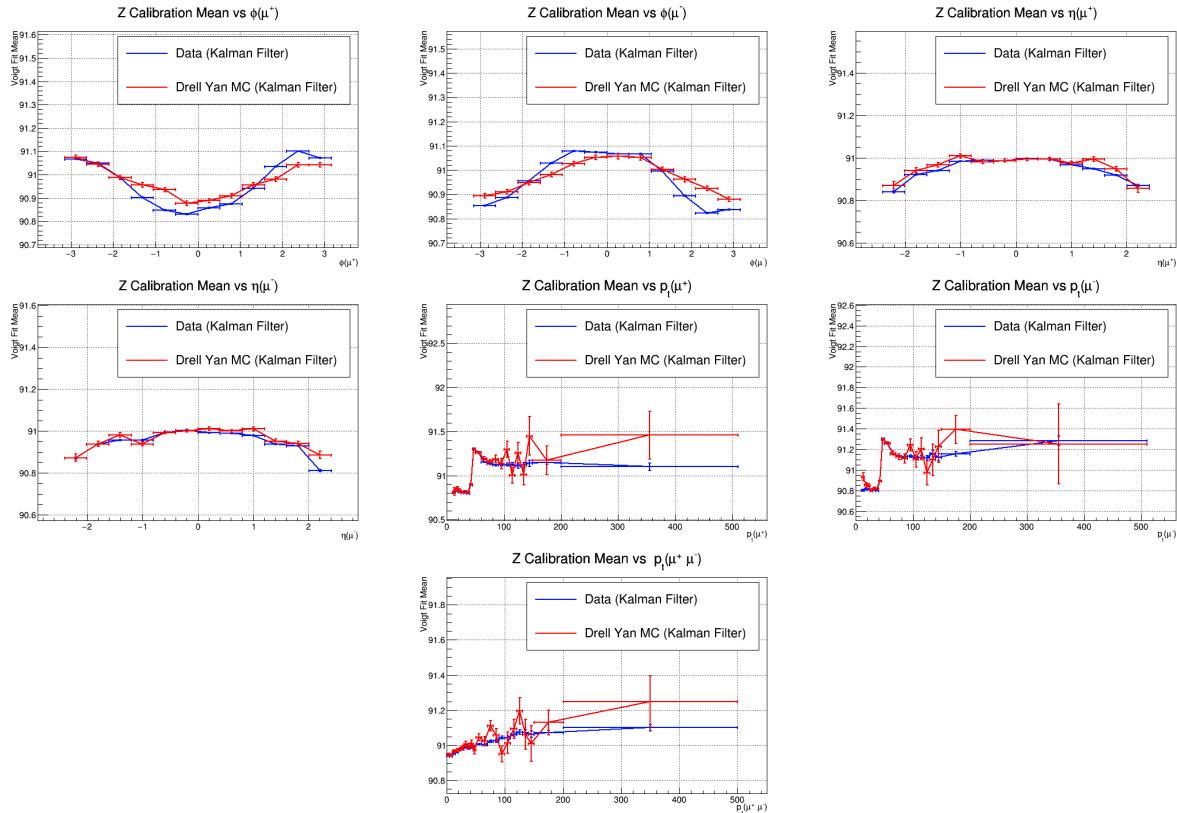


Figure 4-8. The fitted mean of the Z peak is plotted for the Kalman Filter corrections. The mean is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC align very well in terms of the Z peak mean after applying the Kalman filter muon corrections.

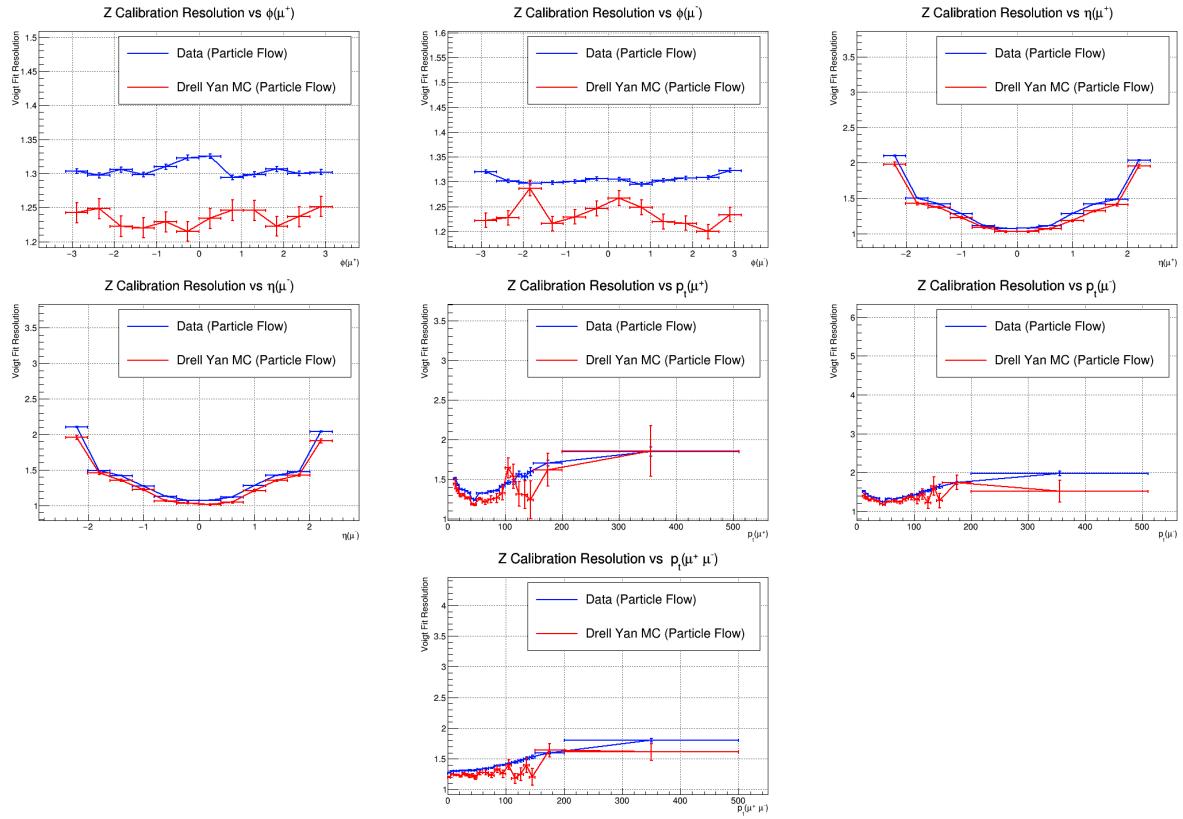


Figure 4-9. The fitted resolution of the Z peak without corrections is plotted. The resolution is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. Data and MC do not show similar resolutions for the Z peak before corrections.

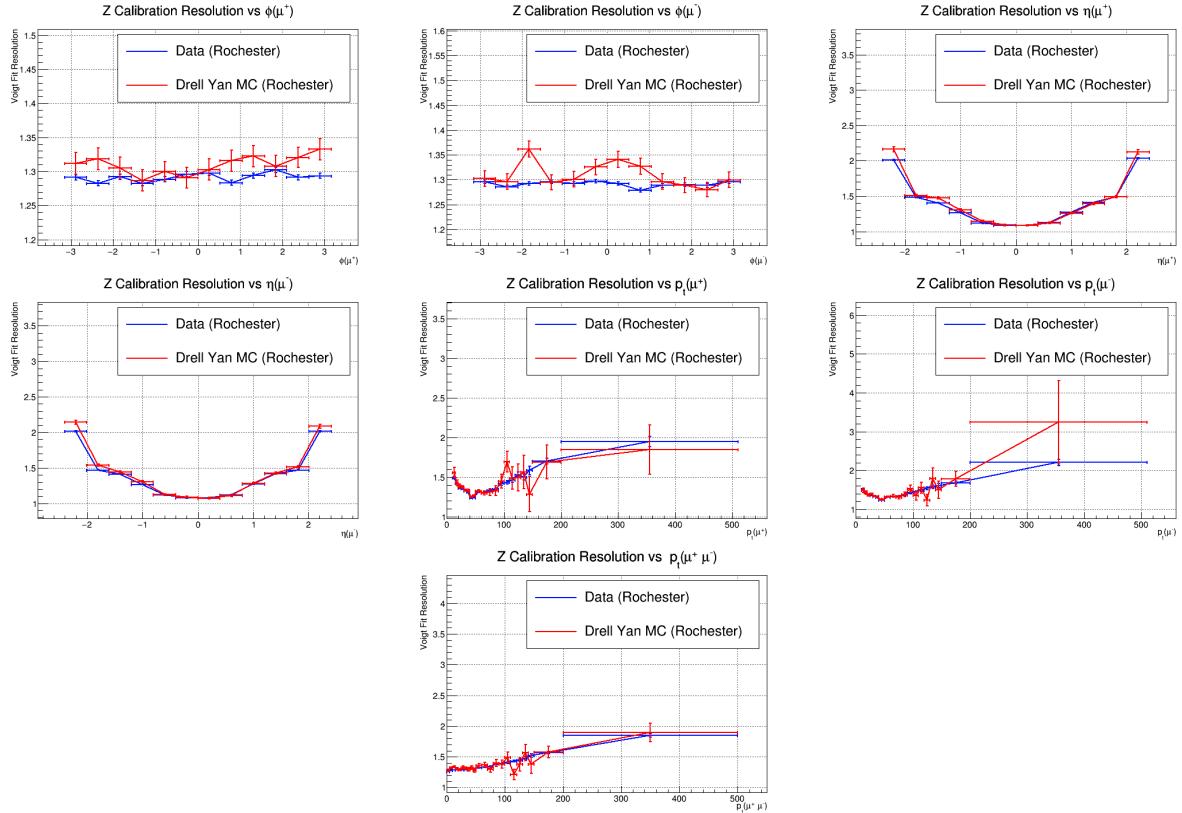


Figure 4-10. The fitted resolution of the Z peak is plotted for the Rochester corrections. The resolution is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. After Rochester corrections, data and MC have similar resolutions for the Z peak.

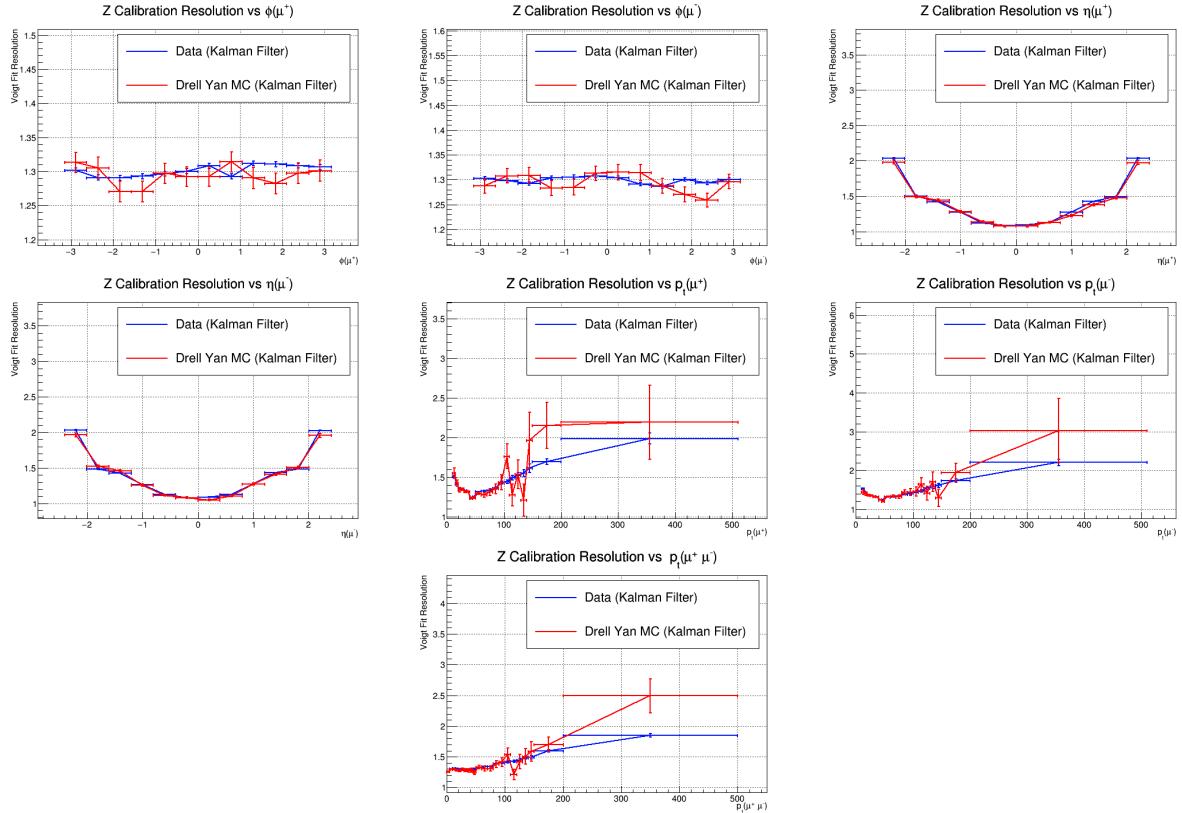


Figure 4-11. The fitted resolution of the Z peak is plotted for the Rochester corrections. The resolution is plotted in data and MC vs. ϕ , η , and p_t^μ for the positively and negatively charged muon separately, and for the dimuon p_t as well. After Kalman filter muon corrections, data and MC have similar resolutions for the Z peak.

4.5.3 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 4-12 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%.

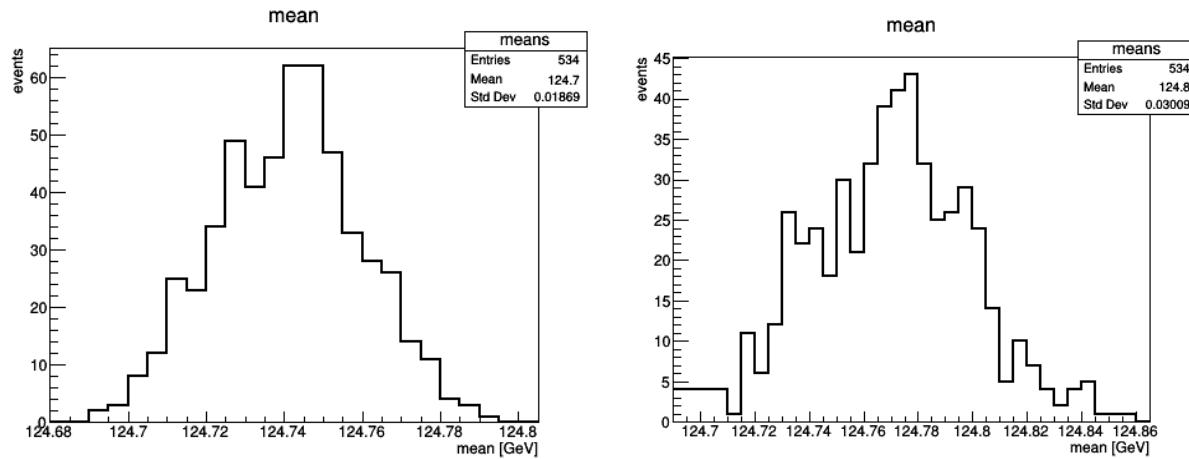


Figure 4-12. The Kalman filter's most probable true mass values are plotted for the signal MC. The least sensitive category is shown on the left and the most sensitive category is shown on the right.

4.6 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.

4.6.1 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 ,

$$I_{\text{rel}}^{\text{PF}} = (p_T^{\text{ch}} + \max(0, E_T^\gamma + E_T^{\text{nh}} - 0.5 * p_T^{\text{chPU}})) / p_T^\mu. \quad (4-20)$$

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.

4.6.2 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

4.6.3 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.

4.6.4 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,

$$SF_{\text{trg}} = \frac{\varepsilon_{\text{trg}}^{\text{data}}}{\varepsilon_{\text{trg}}^{\text{MC}}}. \quad (4-21)$$

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,

$$SF_{\text{id/iso}} = \frac{\varepsilon_{\text{id/iso}}^{\text{data}}}{\varepsilon_{\text{id/iso}}^{\text{MC}}}. \quad (4-22)$$

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,

$$w_{\text{BF}} = \frac{\mathcal{L}_{\text{BF}}}{\mathcal{L}_{\text{tot}}} \quad (4-23)$$

$$w_{\text{GH}} = \frac{\mathcal{L}_{\text{GH}}}{\mathcal{L}_{\text{tot}}} \quad (4-24)$$

$$SF = w_{\text{BF}} \cdot (SF_{\text{trg}}^{\text{BF}} SF_{\text{id}}^{\text{BF}} SF_{\text{iso}}^{\text{BF}}) + w_{\text{GH}} \cdot (SF_{\text{trg}}^{\text{GH}} SF_{\text{id}}^{\text{GH}} SF_{\text{iso}}^{\text{GH}}). \quad (4-25)$$

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

4.6.5 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.

4.6.6 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 $p_T > 26 \text{ GeV}$. 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.

4.7 Maximizing the Sensitivity

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. In this way, the sensitivity is recovered. The sensitivity can be further improved by concentrating the signal peak into a narrower region. In this regard, the events are extracted and placed into categories based upon both the BDT score and the mass resolution.

4.7.1 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 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 4-5. 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 4-5. 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 4-19 and 4-20 for the events passing the initial event selection. The separation of the signal and background along the BDT score is plotted in Figure 4-14. The Receiver Operating Characteristic (ROC) curve for the final BDT model is shown in Figure 4-13. With the BDT score the events may now be categorized.

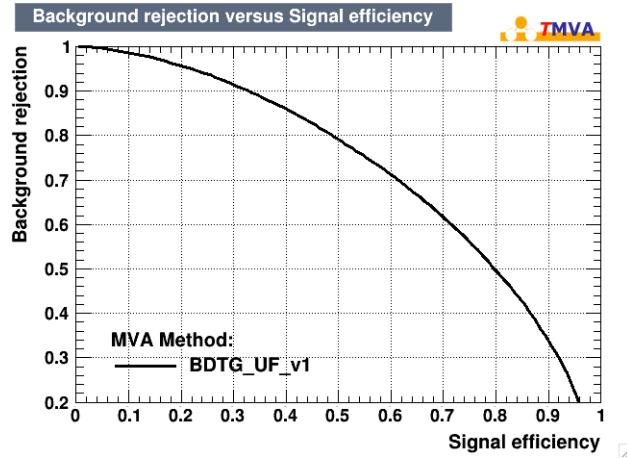


Figure 4-13. The background-rejection vs. signal efficiency ROC curve for the final BDT model.

4.7.2 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.

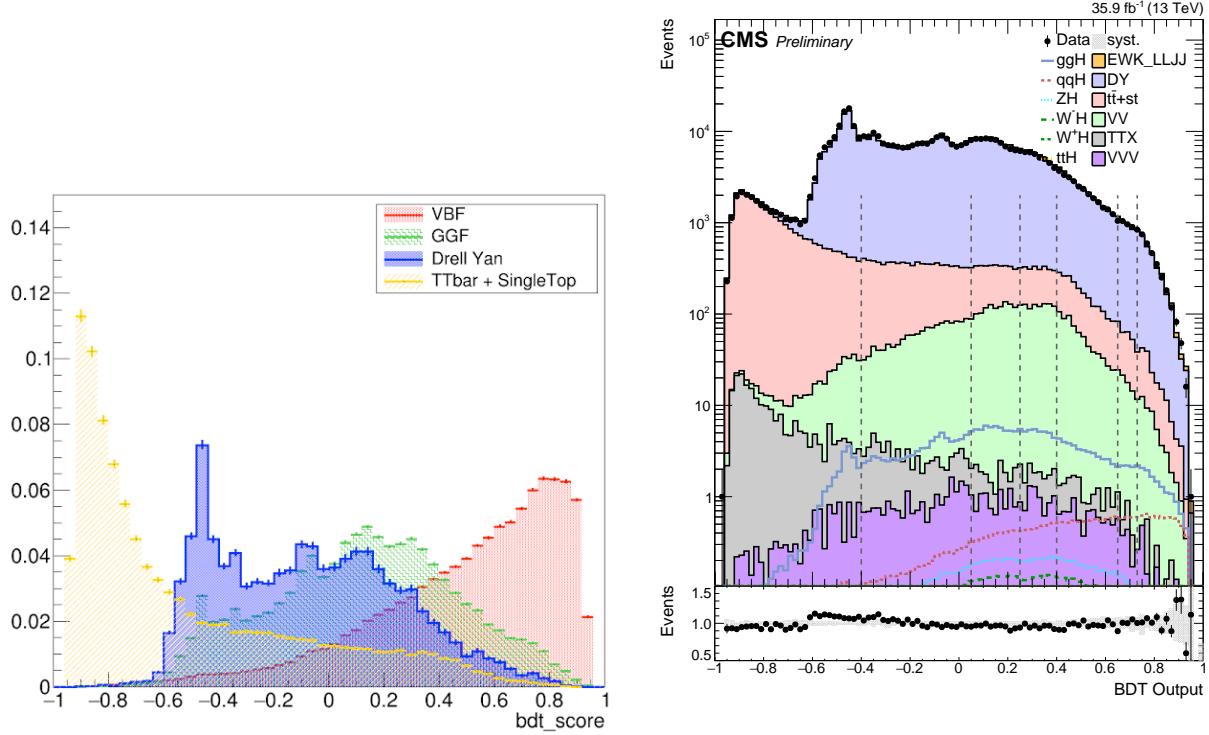


Figure 4-14. The BDT score on the signal and background Monte Carlo (left), where a higher score marks the signal-like events. Note $t\bar{t}$ on the left as the most distinguishable background and VBF on the right as the most distinguishable signal. The BDT score in data is modeled well by the amc@NLO Drell–Yan MC (right).

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

$$\text{Net Significance} = \sum_{c,i} S_{c,i}^2 / B_{c,i} \quad (4-26)$$

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 $B_{i,c}$ is the number of expected

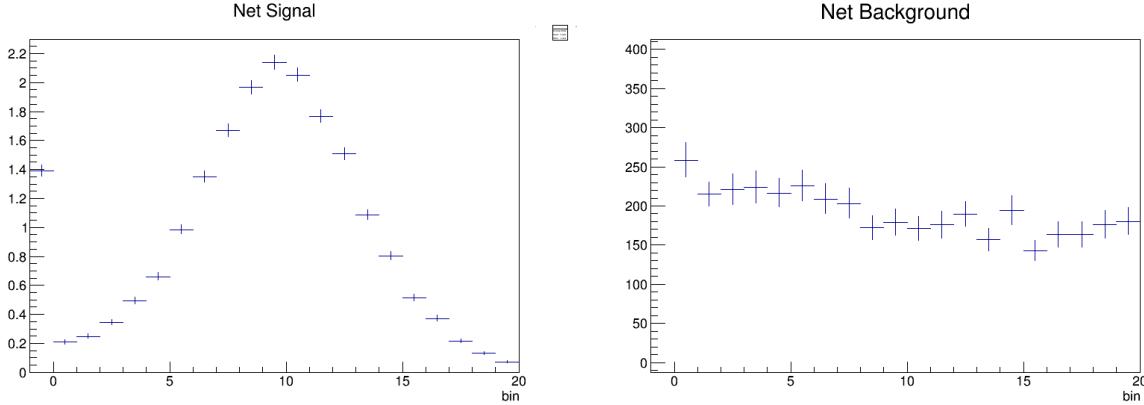


Figure 4-15. An example of the binned signal and background necessary for the significance metric. The binning keeps the most sensitive bins from being drowned out by the others, providing a measure that accounts for both signal–background discrimination and resolution.

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.

$$\text{Gain} = \sum_i S_{c1,i}^2 / B_{c1,i} + \sum_i S_{c2,i}^2 / B_{c1,i} - \sum_i S_{c,i}^2 / B_{c,i} \quad (4-27)$$

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.

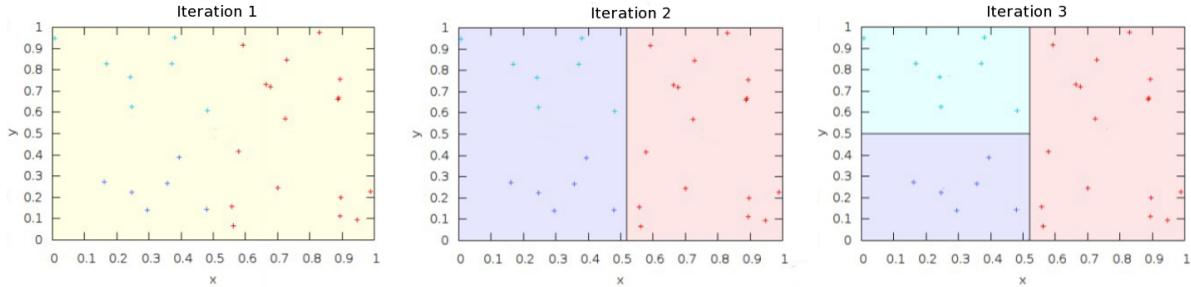


Figure 4-16. An example of the categorization process for some toy features x and y and three categories. The colored crosses represent the events that should be grouped together for optimum sensitivity. After three iterations, the categorizer correctly groups those events. The autocategorizer chooses $x=0.52$ for the first split and $y=0.50$ for the second.

4.7.4 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 13 BDT categories

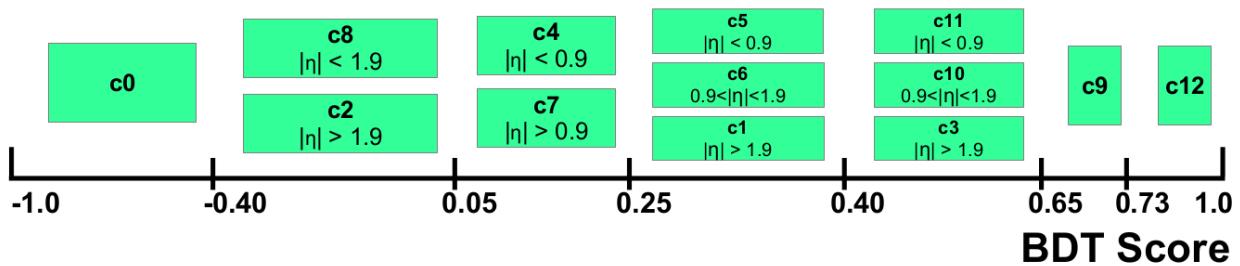


Figure 4-17. The categories produced by the autocategorizer. $|\eta|$ refers to the maximum $|\eta|$ value of the two muons in the dimuon pair.

produced by the Autocategorizer in Figure 4-17 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 BDT categories are named such that the numbers go from low expected sensitivity to high expected sensitivity, e.g. c0 has the lowest expected sensitivity and c12 has the highest expected sensitivity.

Splitting c7 and c8 from Figure 4-17 for symmetry and renaming by BDT Score and $|\eta|$, yields the final 15 category scheme shown in Figure 4-18. After these adjustments, categories c0 to c14 are no longer ordered by expected sensitivity, but c14 remains the most sensitive and c0 the least sensitive.

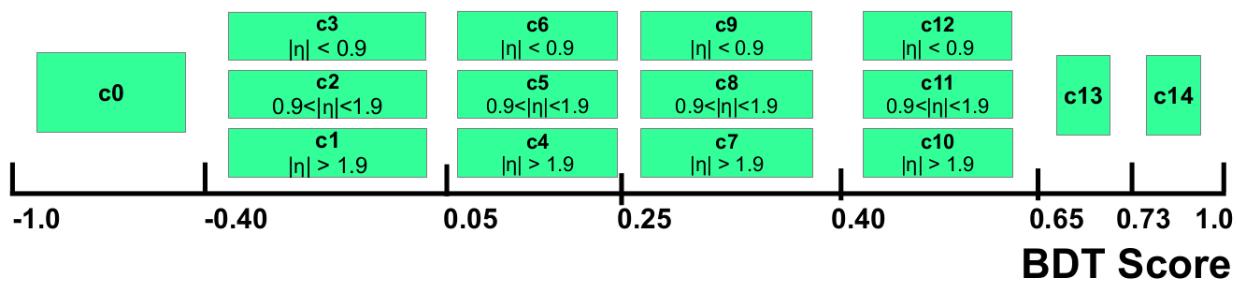


Figure 4-18. The final categorization

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 4-6.

Table 4-6. Full Width Half Max (FWHM) of the mass resolution of the signal peak, expected signal and background yields within the FWHM, purity, sensitivity, and the expected limits from data. The expected background yields are estimated by fitting 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

4.7.5 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 4-19 and 4-20 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.

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 4-21 and 4-22 show the data/MC agreement for c14, the most sensitive, most VBF-like category. The BDT distribution in quantile is presented in Fig. 4-24 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 dimuon p_t to data are shown in Fig. 4-23. Note also the systematic 10% disagreement in BDT score from -1 to -0.6 in Figure 4-23 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

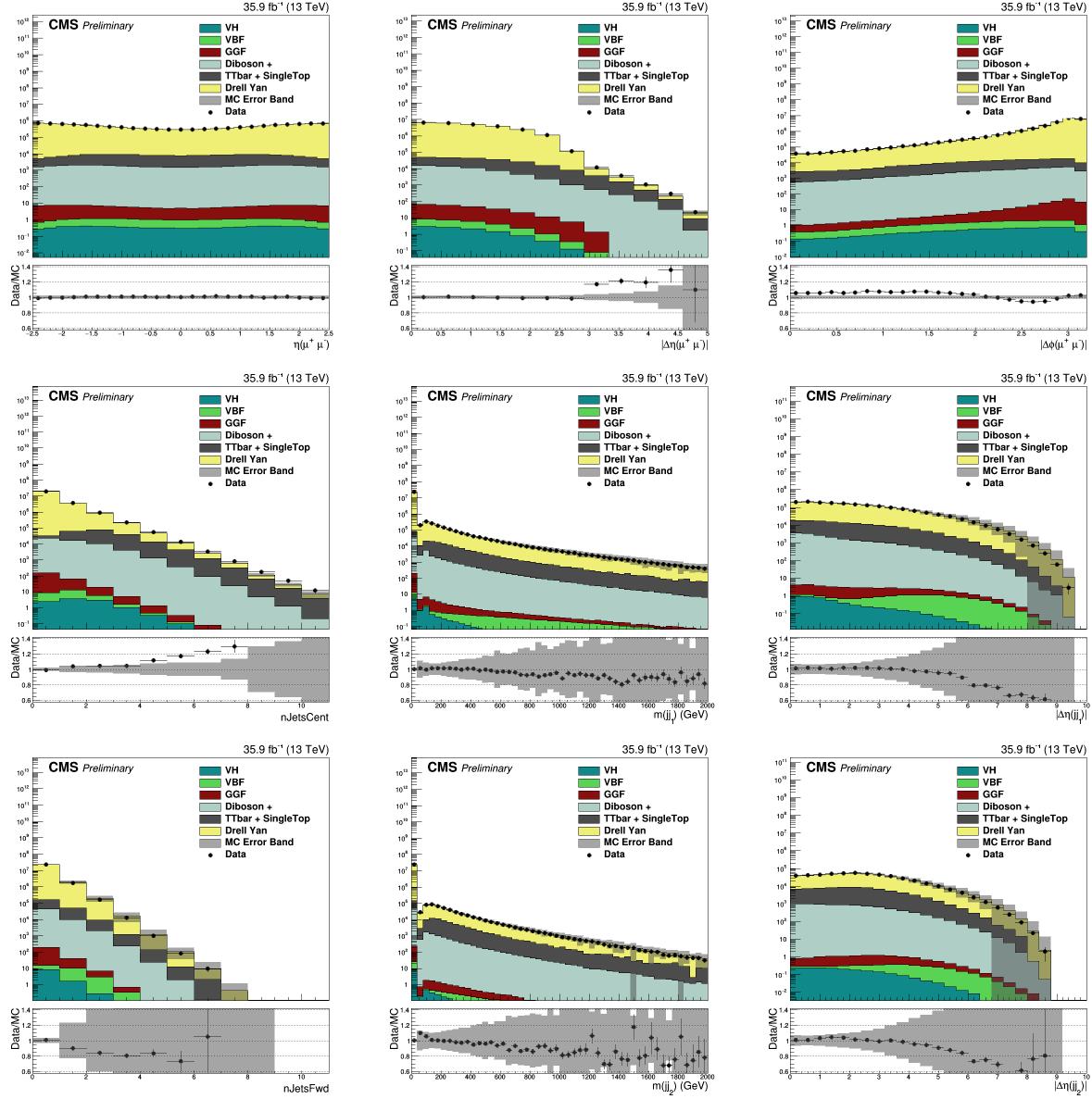


Figure 4-19. Validation of BDT input variables in inclusive events.

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

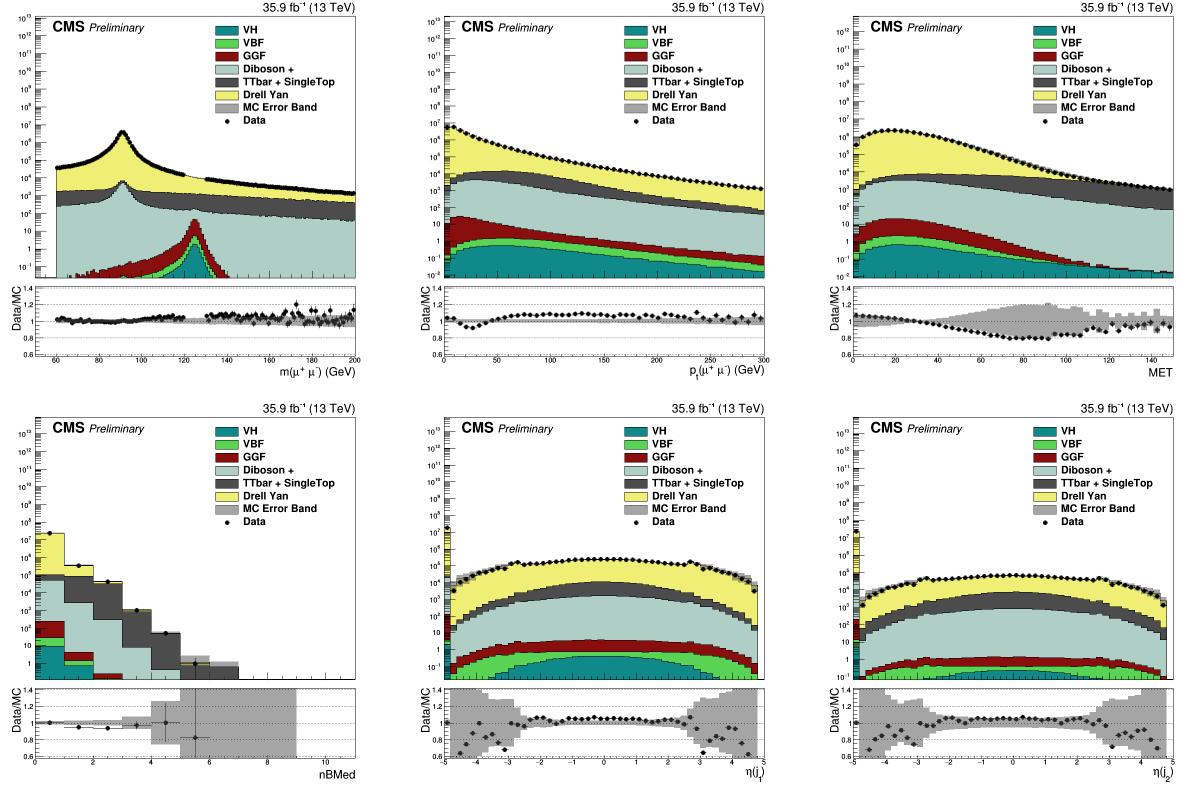


Figure 4-20. Data/MC agreement for the inclusive events.

jet energy scale and resolution uncertainties as shown in Fig. 4-24. 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 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 4-25, the BDT score does not change significantly with the dimuon mass.

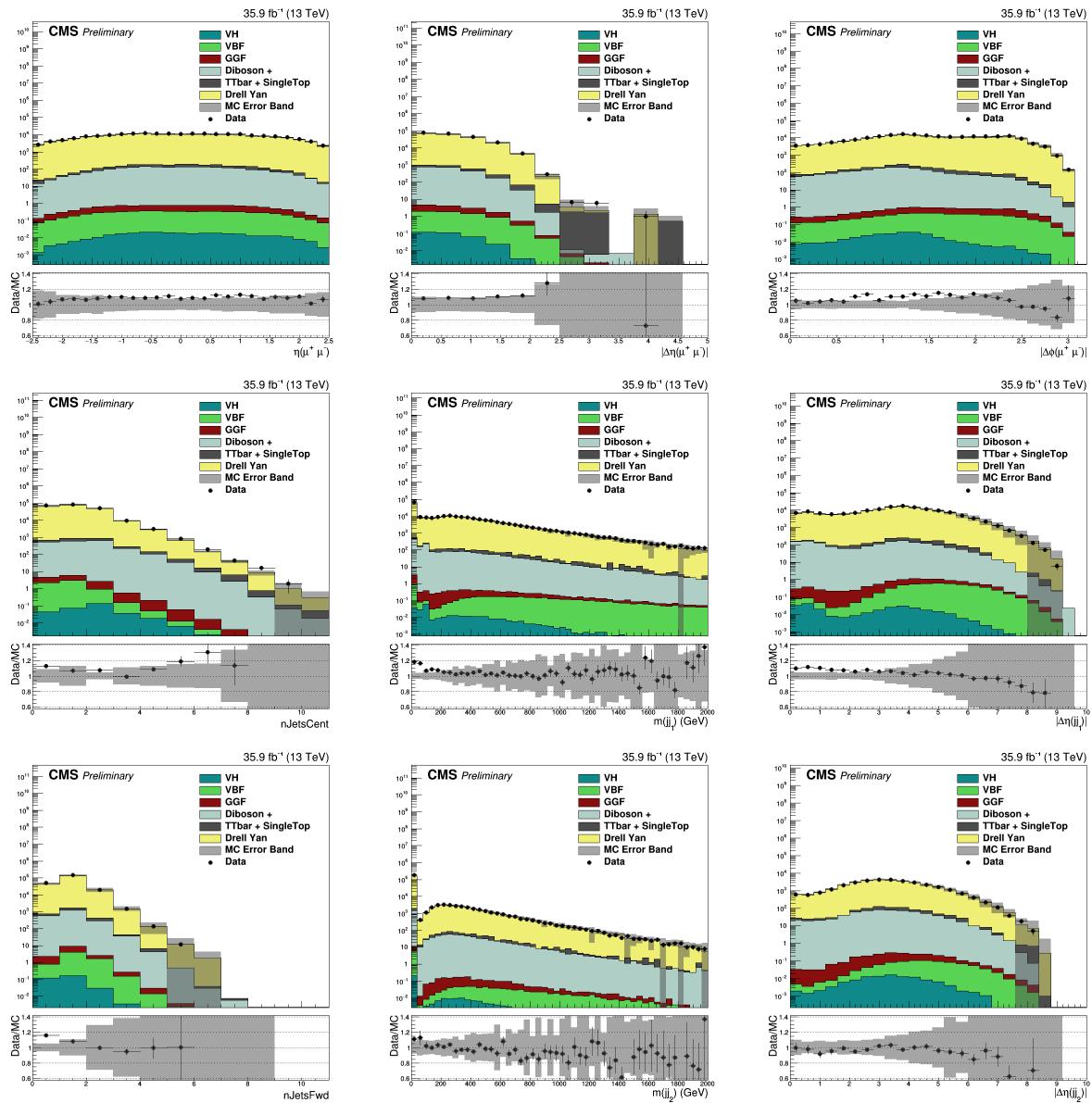


Figure 4-21. Data/MC agreement for the most sensitive category.

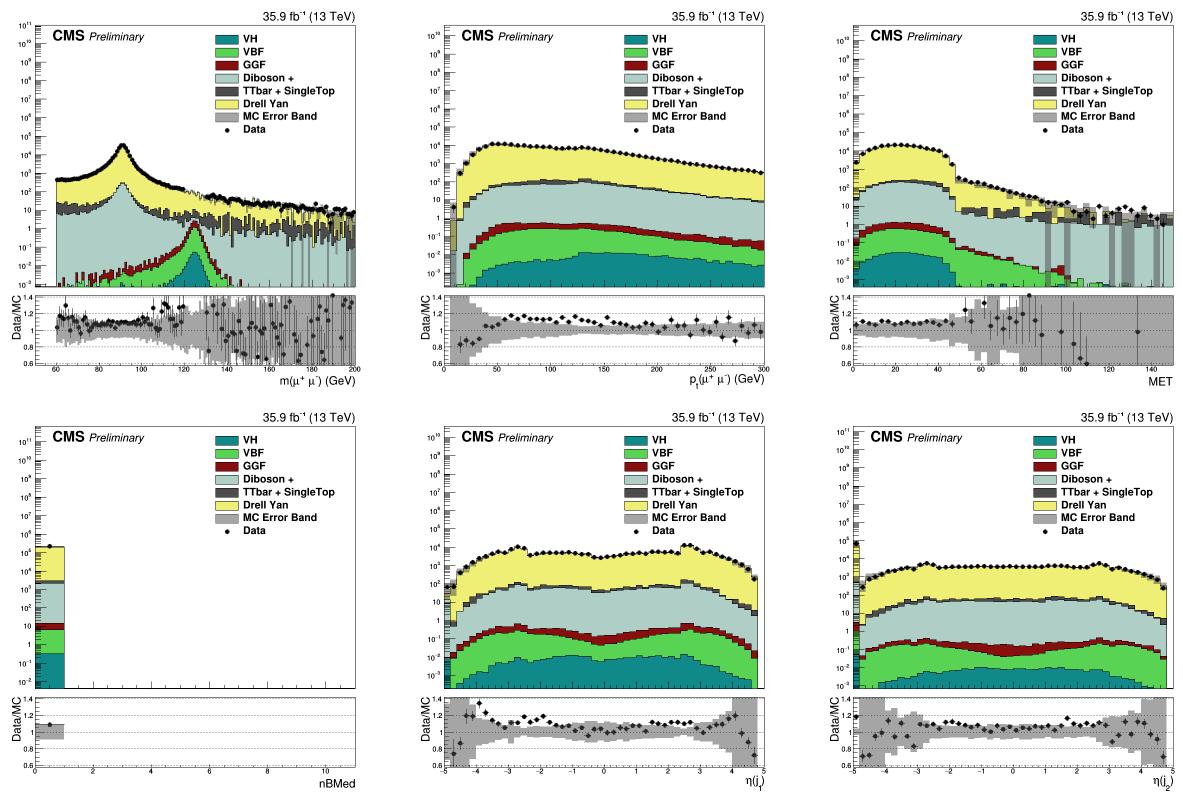


Figure 4-22. Data/MC agreement for the most sensitive category.

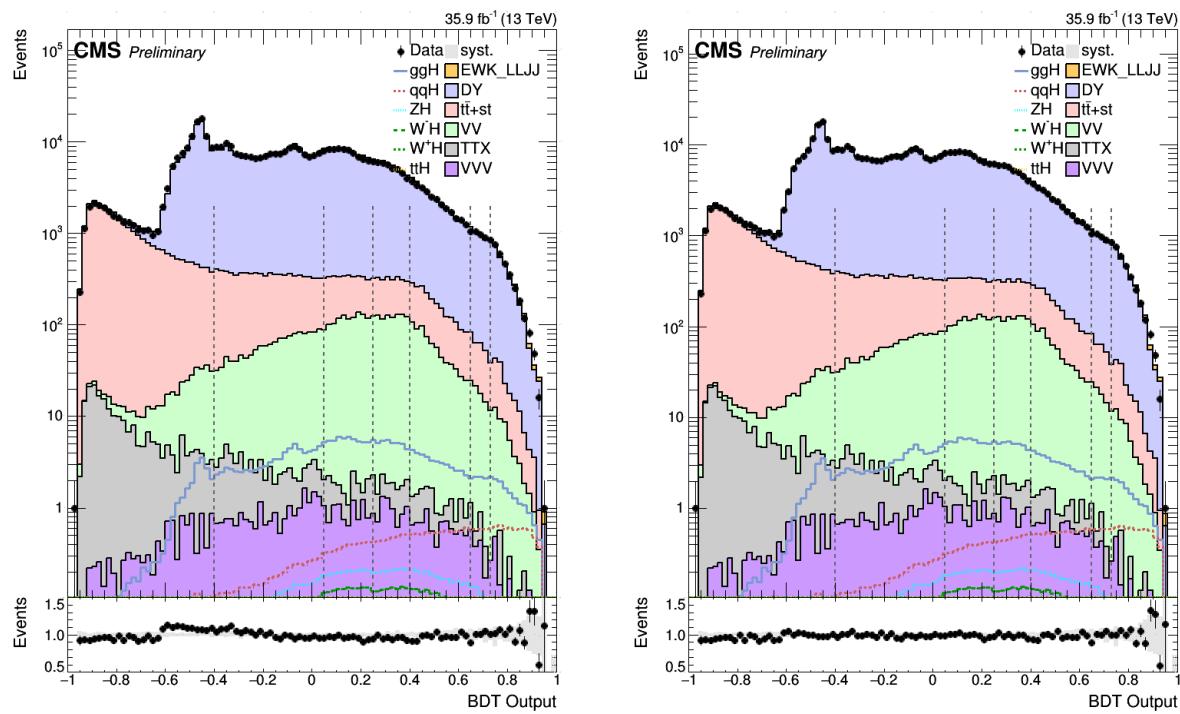


Figure 4-23. The BDT distribution in data and MC. Right plot matches the DY p_t spectrum to the one in data.

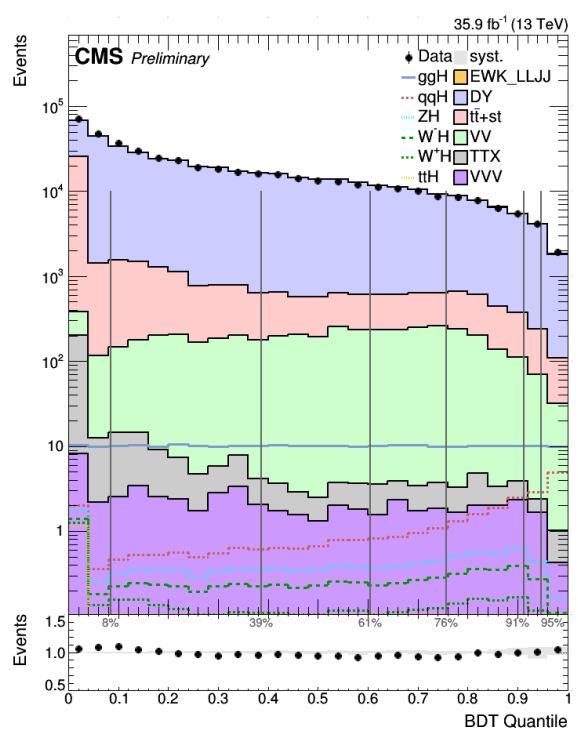


Figure 4-24. The BDT distribution in quantile in data and MC. The systematic uncertainty is given by the JES and PU.

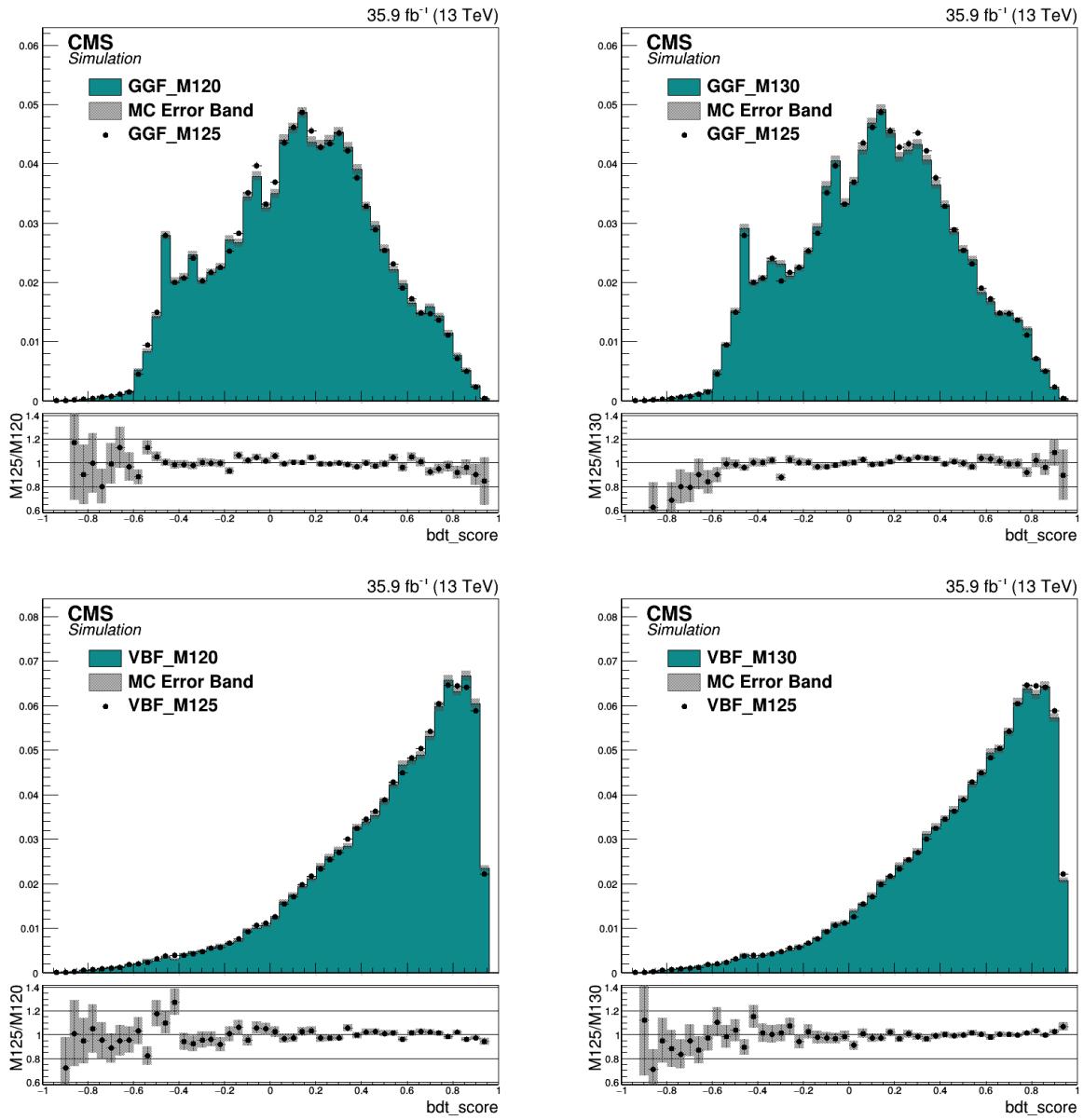


Figure 4-25. Plots showing the agreement between $m_H = 125 \text{ GeV}$ vs. 120 GeV and $m_H = 125 \text{ GeV}$ vs. 130 GeV for ggH and VBF.

4.7.6 Signal and Background

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

Table 4-7 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 4-26 and then for the most sensitive category (c14) starting with Figure 4-29. After looking at the signal, Fig. 4-32 presents the background shapes in data for each category and compares each to the inclusive shape.

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

Category	All Signal	GGF	VBF	VH
Inclusive	243	214	18.9	10.3
c0	18.8	17.6	0.401	0.833
c1	20.7	19.5	0.505	0.693
c2	39.0	36.9	0.862	1.21
c3	12.2	11.6	0.236	0.374
c4	11.1	10.0	0.518	0.522
c5	27.5	25.4	1.01	1.10
c6	13.8	13.0	0.355	0.462
c7	5.00	4.20	0.451	0.352
c8	19.8	17.8	1.12	0.899
c9	12.5	11.6	0.492	0.448
c10	5.52	4.14	0.932	0.447
c11	20.0	16.1	2.67	1.27
c12	13.8	11.8	1.30	0.749
c13	8.76	6.22	2.01	0.521
c14	14.1	7.70	6.07	0.313

Comparing the signal kinematics in Figures 4-26 through 4-28 and 4-29 through 4-31 notice that a high BDT score, exemplified by c14, removes $t\bar{t}$ -like events by requiring 0 btags and low MET. Aside from removing $t\bar{t}$, 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.

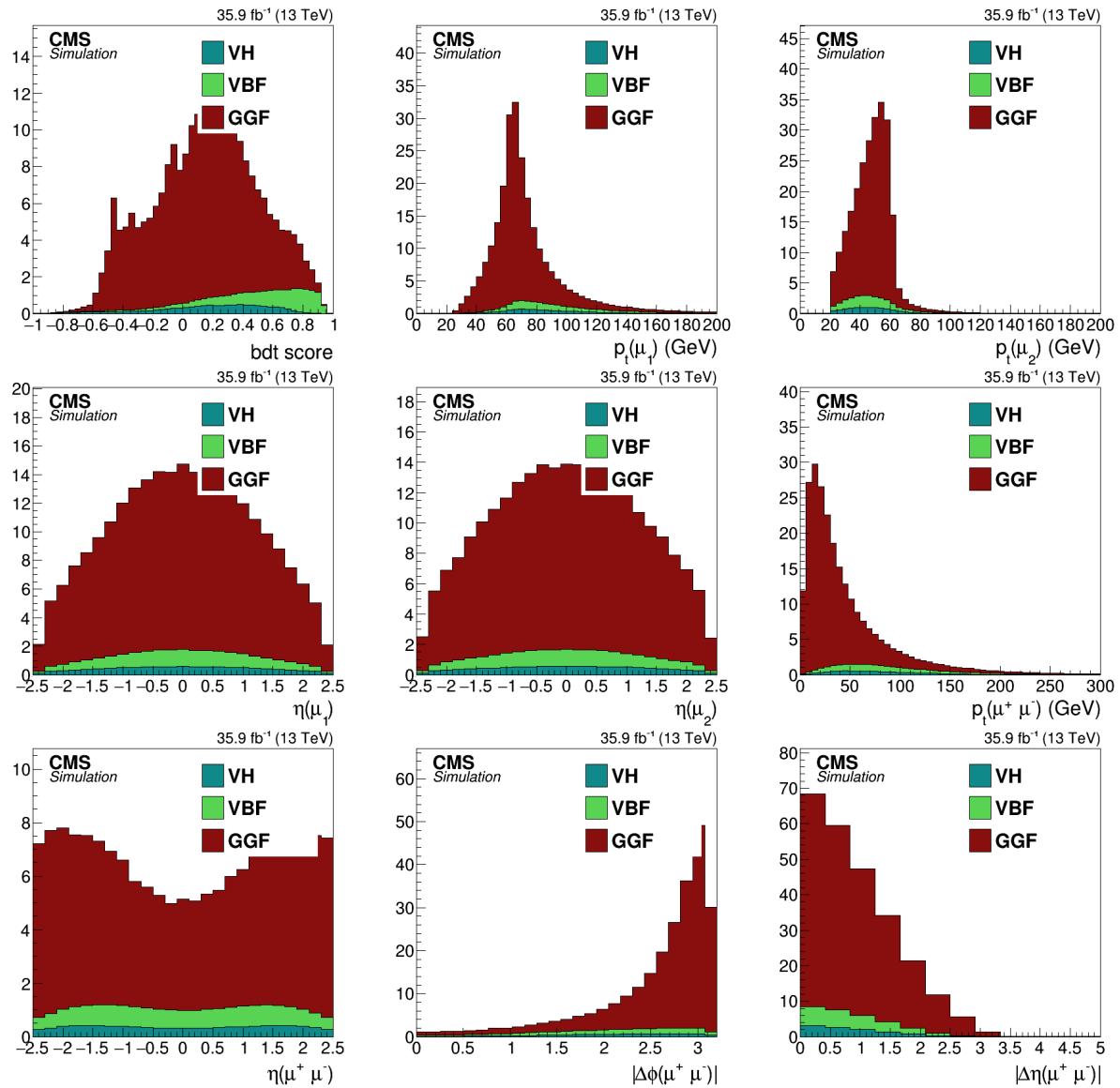


Figure 4-26. Kinematics and object counts for the signal in the inclusive set of events.

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. 4-32. 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 4-32, the background shape for each category

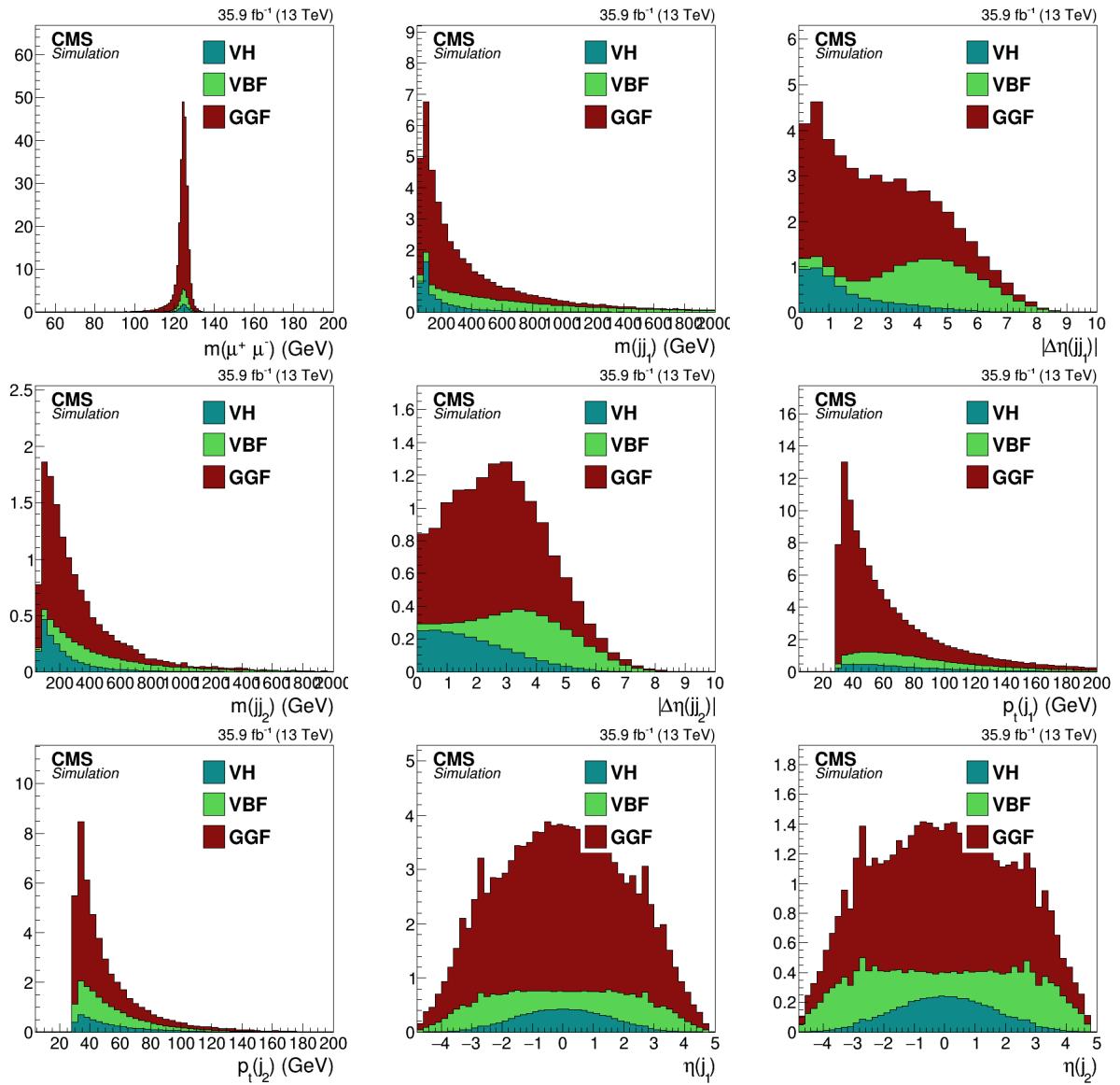


Figure 4-27. More kinematic and object count histograms for the signal in the inclusive set of events.

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

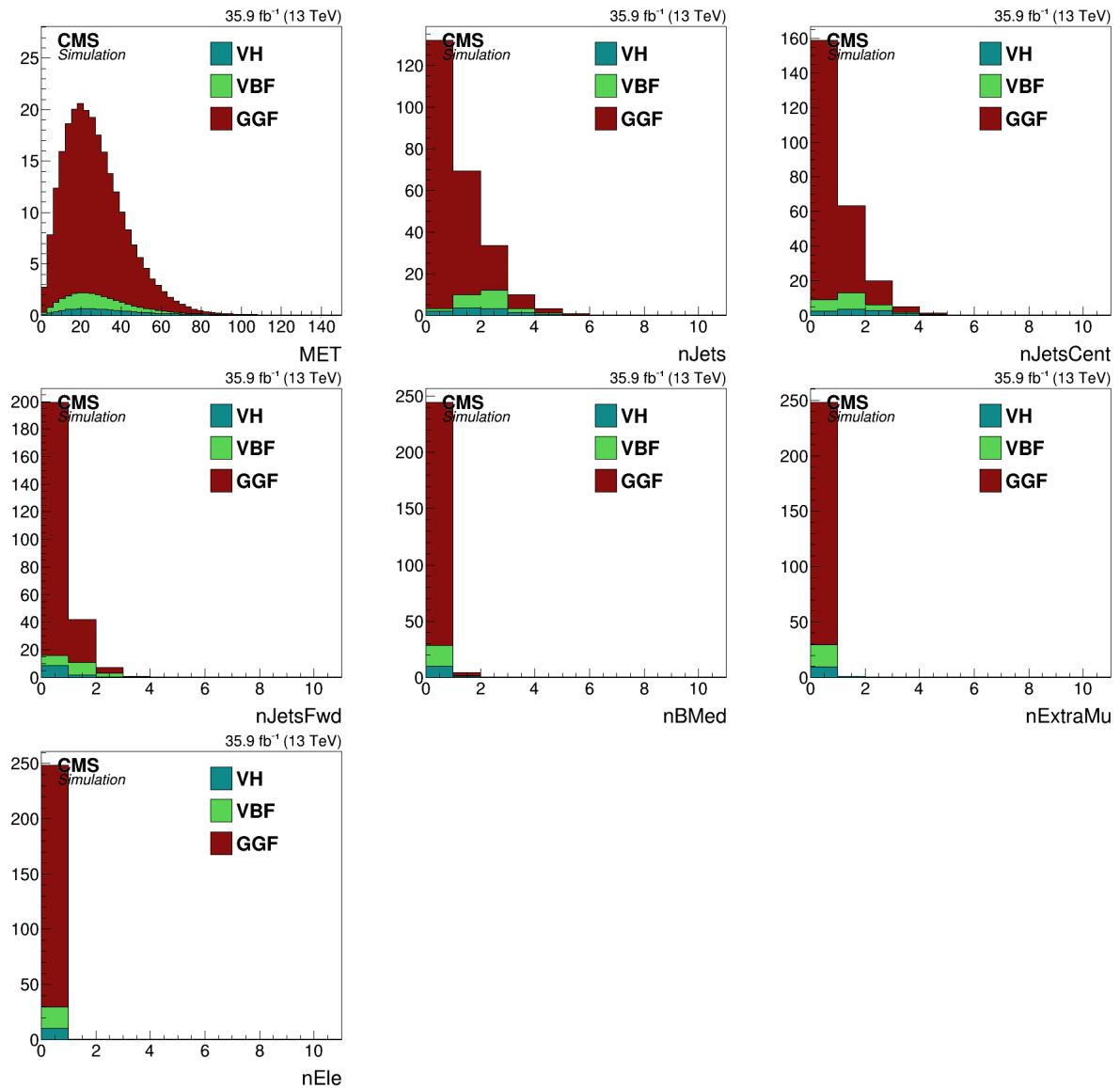


Figure 4-28. More kinematic and object count histograms for the signal in the inclusive set of events.

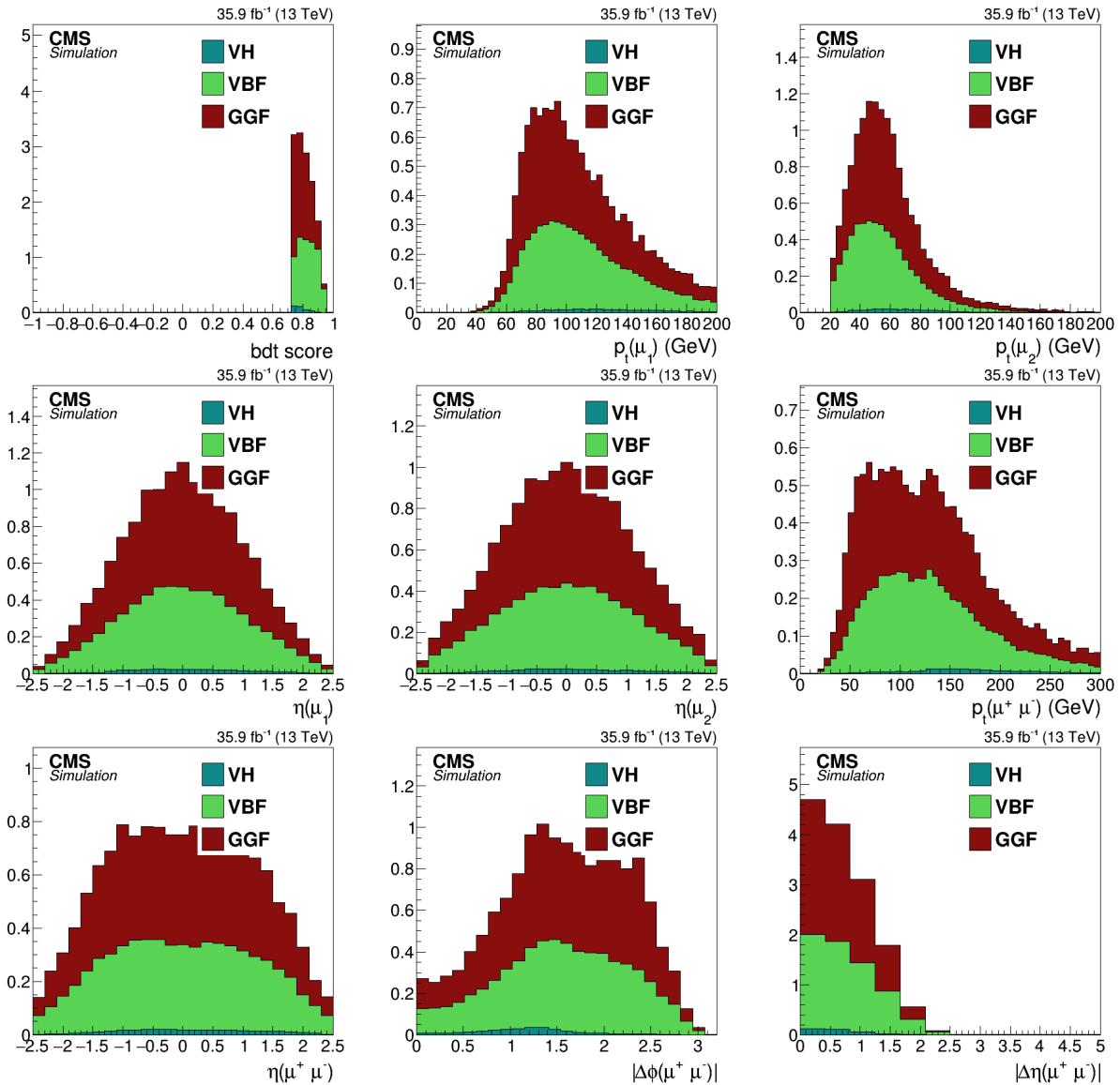


Figure 4-29. Kinematic and object count histograms for the signal in category c14. Note the high dimuon p_t targeting the signal in general.

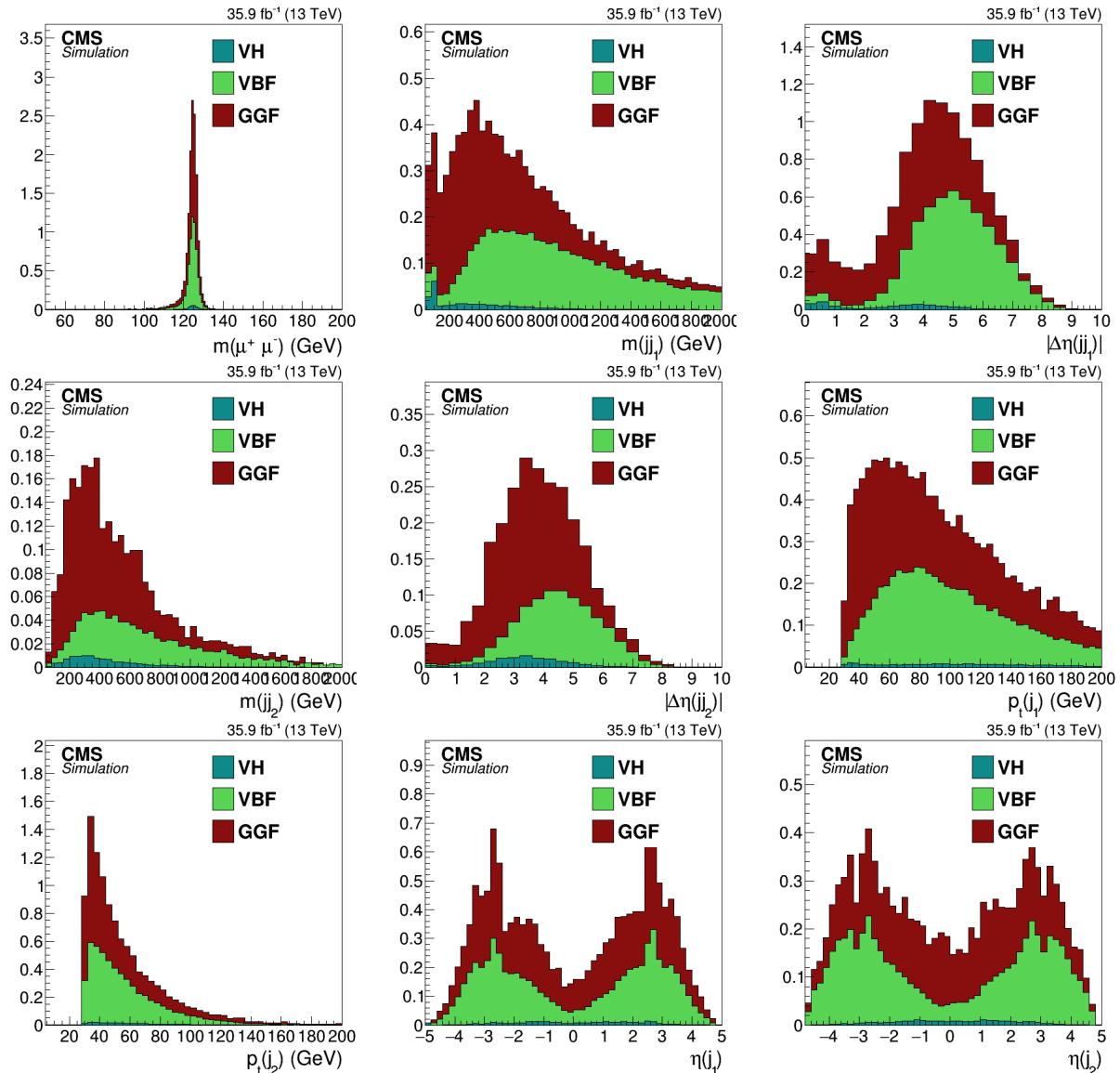


Figure 4-30. More kinematic and object count histograms for the signal in category c14. Note large dijet mass and the large separation in eta for the jets targeting VBF.

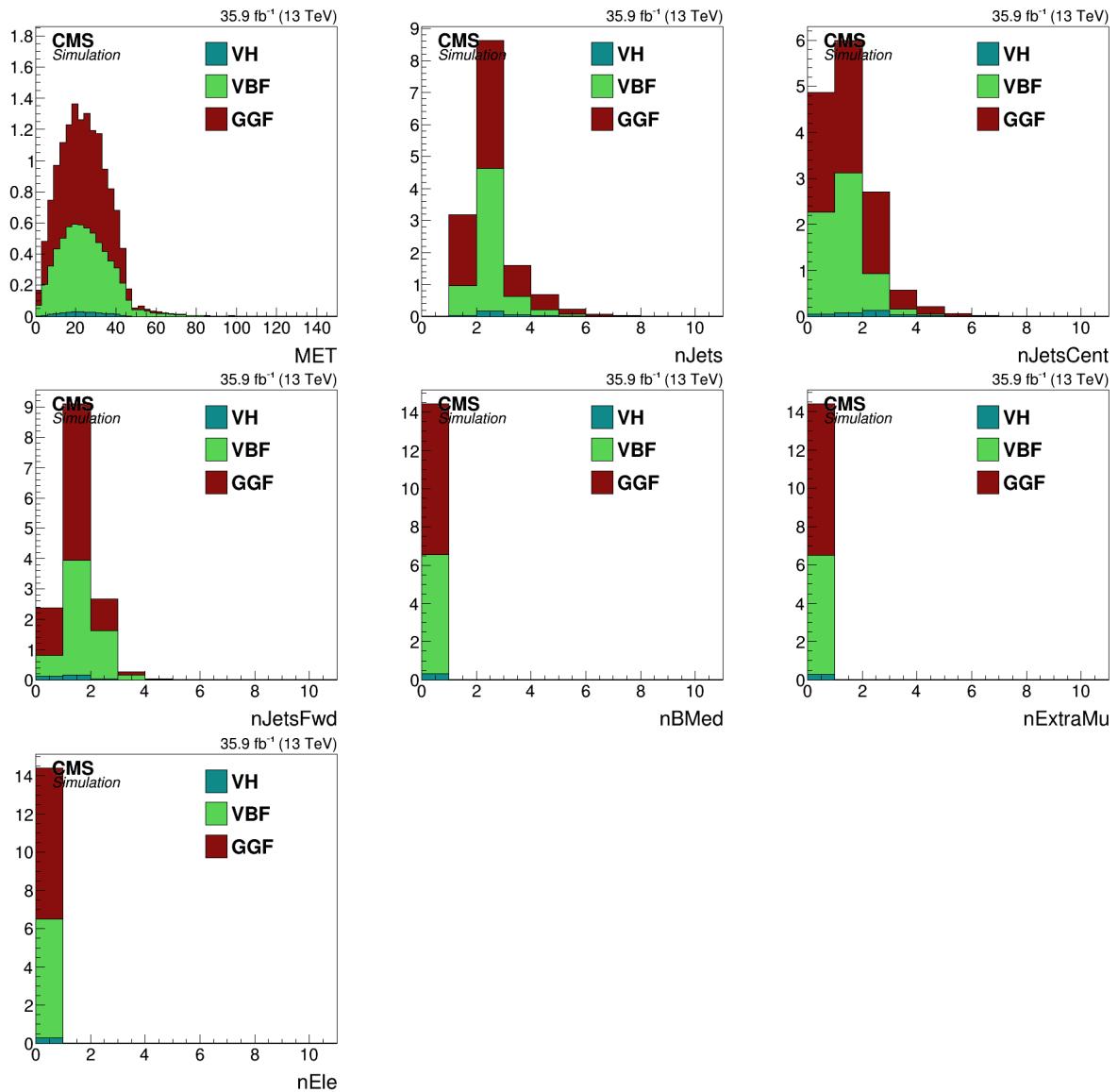


Figure 4-31. More kinematic and object count histograms for the signal in category c14. Note the low MET and 0 btags which exclude $t\bar{t}$.

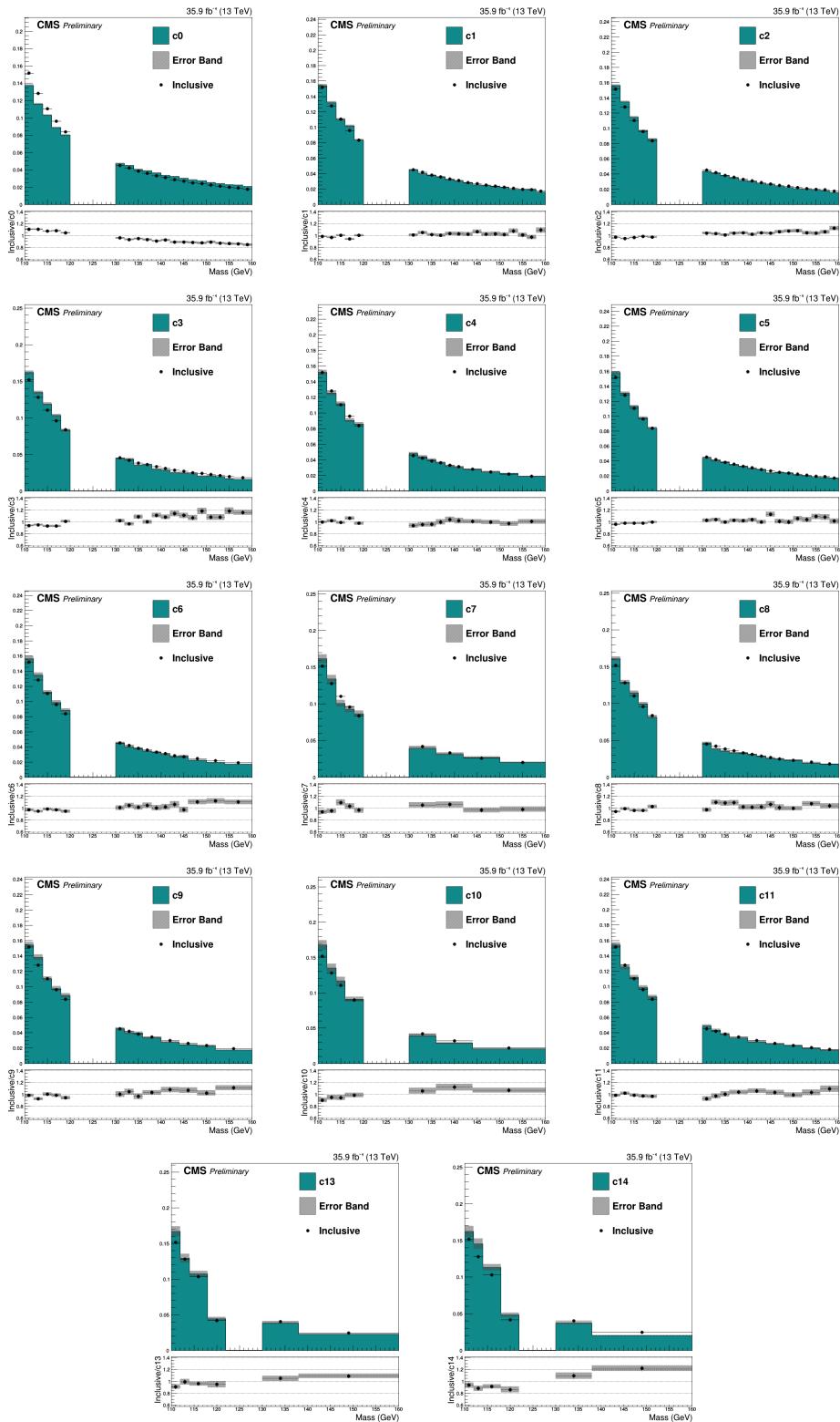


Figure 4-32. The difference in background shapes between the categories in data. The plots show the data with the signal region blinded. Each category's background shape is compared to the inclusive shape.

4.8 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 Gaussian²,

$$S_{pc}(x, m_H, \theta_{pc}) = \sum_{g=1}^3 f_{pcg} \mathcal{N}(x, \mu_{pcg}, \sigma_{pcg}). \quad (4-28)$$

In 4-28, 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 4-33 shows an example of a fit on the left, its interpolation in the center, and the efficiency times acceptance on the 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.

$$\text{Yield}_{pc} = \mathcal{L} \sigma(pp \rightarrow H) \mathcal{B}(H \rightarrow \mu^+ \mu^-) \varepsilon_{pc} A_{pc} \quad (4-29)$$

² Three of the process,category fits have very low stats and use only two Gaussians.

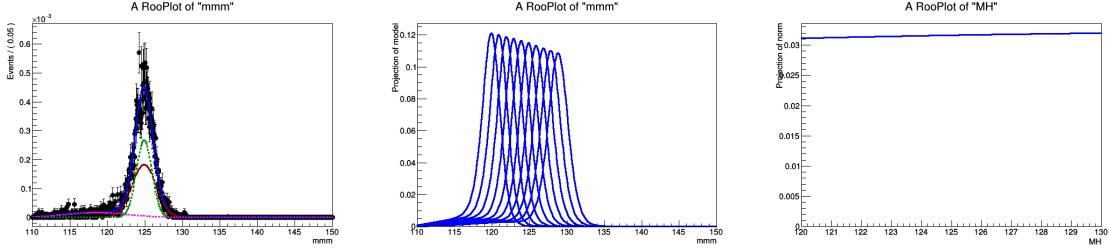


Figure 4-33. An example Higgs signal fit for a specific production process and category at a mass of 125 GeV (left), its interpolation to lower and higher mass (center), and its efficiency times acceptance vs m_H (right).

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

$$\varepsilon_{pc}A_{pc} = \frac{1}{N} \sum_{i \in p} w_i \cdot r_i \cdot sf_i \cdot I\{i \in c\}. \quad (4-30)$$

The sum runs over all MC events for the signal process. The factors w_i , r_i , and sf_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, $\sigma(pp \rightarrow H)$, and branching fraction, $\mathcal{B}(H \rightarrow \mu^+\mu^-)$, are taken from the Yellow Report 4, and provided centrally by combine.

The left side of Figure 4-34 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 4-35 shows the composition of the signal model, the mass resolution, and the $S/(S+B)$ in each category. Figure 4-36 shows the efficiency times acceptance of the selection – prior to categorization – for the total signal and for each signal process.

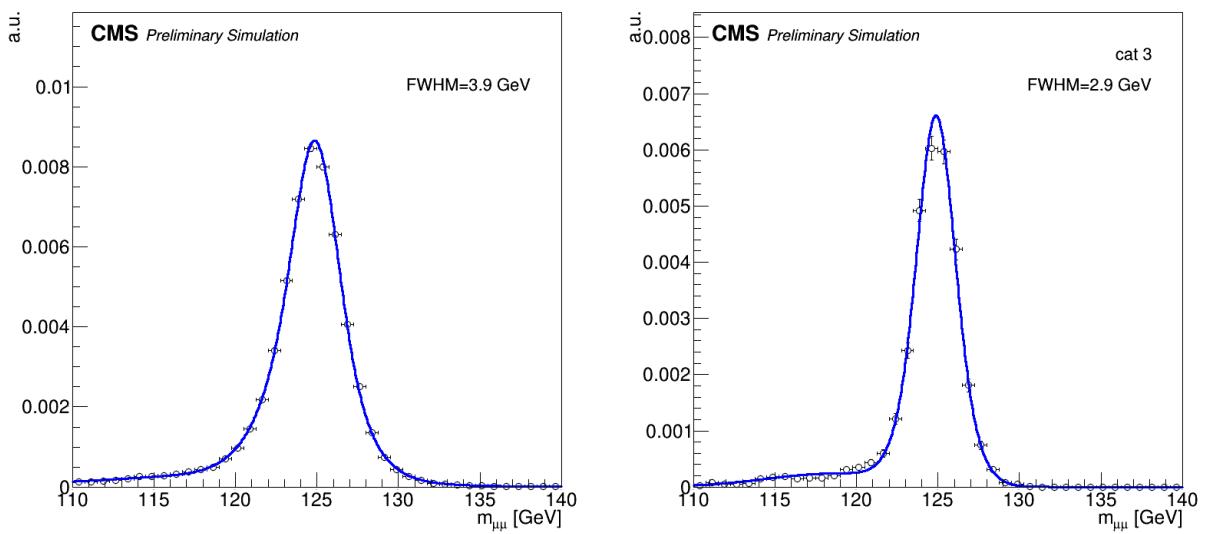


Figure 4-34. Signal model compared to MC predictions, summing up the contribution from all process and all categories (left), and for one of the categories with the best mass resolution.

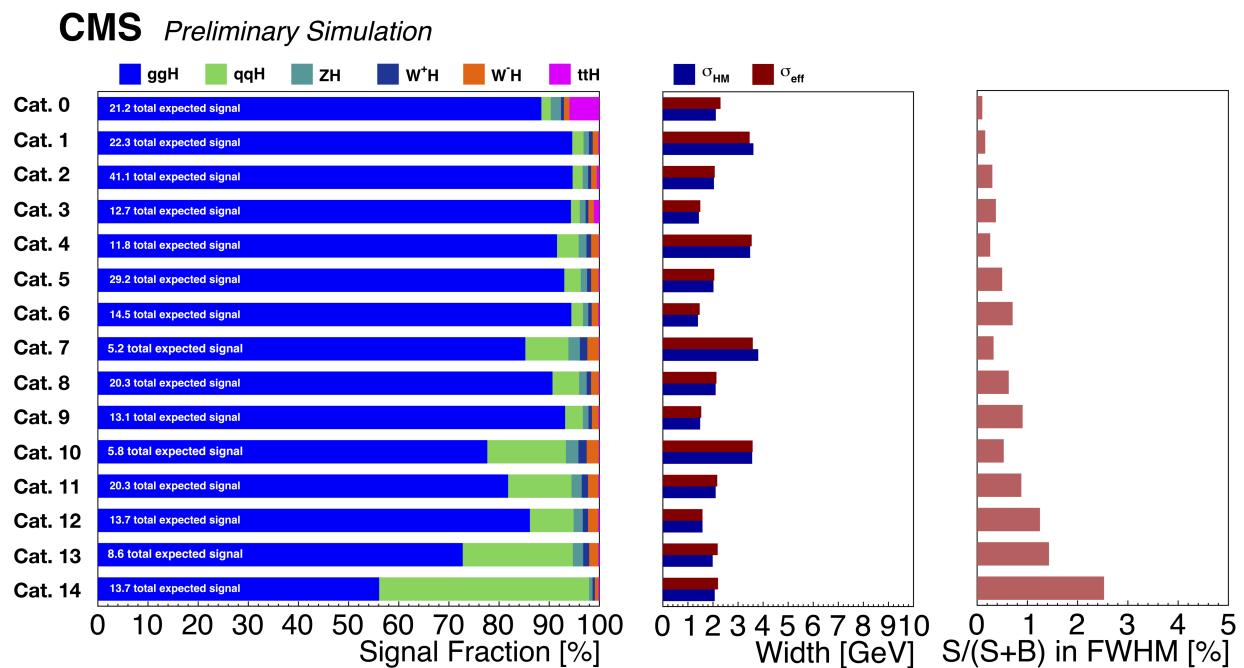


Figure 4-35. The signal composition, the mass resolution in Half Width Half Max (σ_{HM}) and 68% probability (σ_{eff}), and the $S/(S+B)$ of the different categories.

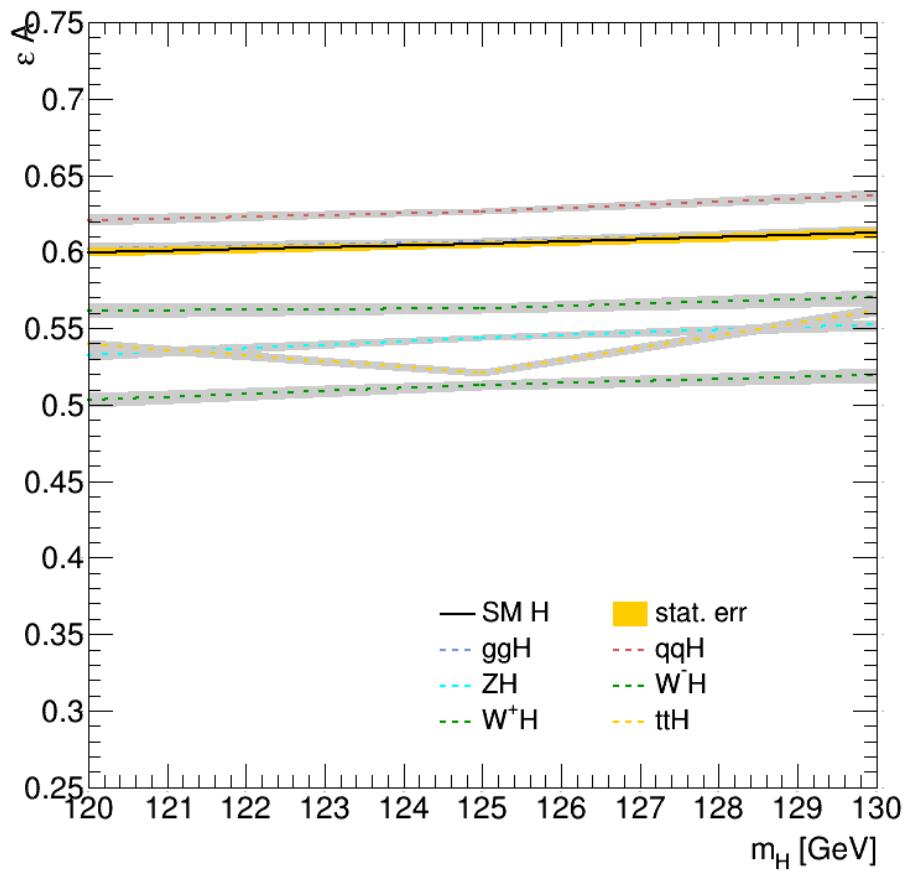


Figure 4-36. The efficiency times acceptance as function of m_H for the total signal and for each signal process.

4.9 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.

$$\text{BWZ: } B(x) = \frac{e^{ax}\sigma_z}{(x - \mu_z)^2 + (\frac{\sigma_z}{2})^2} \quad (4-31)$$

$$\text{BWZRedux: } B(x) = \frac{e^{a_2 x + a_3 x^2}}{(x - \mu_z)^{a_1} + (\frac{2.5}{2})^{a_1}} \quad (4-32)$$

$$\text{BWZGamma: } B(x) = f \frac{e^{ax}\sigma_z}{(x - \mu_z)^2 + (\frac{\sigma_z}{2})^2} + (1 - f) \frac{e^{ax}}{x^2} \quad (4-33)$$

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 4-36. 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.

$$\text{Bernstein:B}(x) = \sum_{i=0}^n \alpha_i \left[\binom{n}{i} x^i (1-x)^{n-i} \right] \quad (4-34)$$

$$\text{SumExponentials:B}(x) = \sum_{i=1}^n \beta_i e^{\alpha_i x} \quad (4-35)$$

The analysis also uses a modified version of the BWZRedux (BWZR(x), Eq. 4-32) modulated by a Bernstein polynomial of degree n (BERN(x, n), Eq. 4-34, normalized to 1):

$$\text{BWZRedux} \cdot \text{Bern: B}(x) = \text{BWZR}(x) \cdot (1 + \text{BERN}(n, x)) \quad (4-36)$$

$$(4-37)$$

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

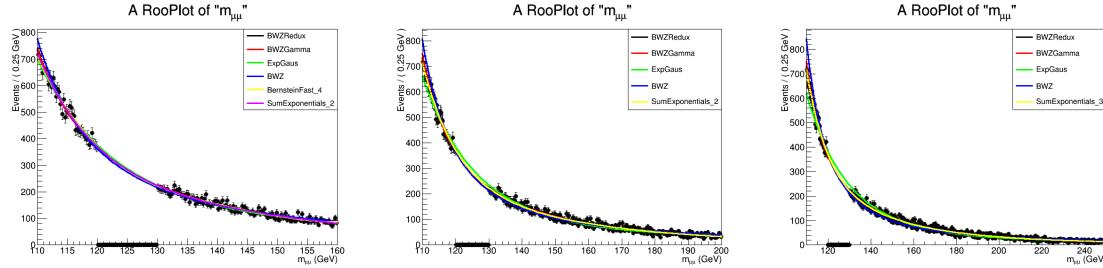


Figure 4-37. Example background-only fits to the dimuon mass spectrum in data for various background models and mass ranges.

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 $2\Delta\text{NLL}$ 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 4-38 shows an example of the F-Test for the Bernstein Polynomials and for the Sum of Exponentials.

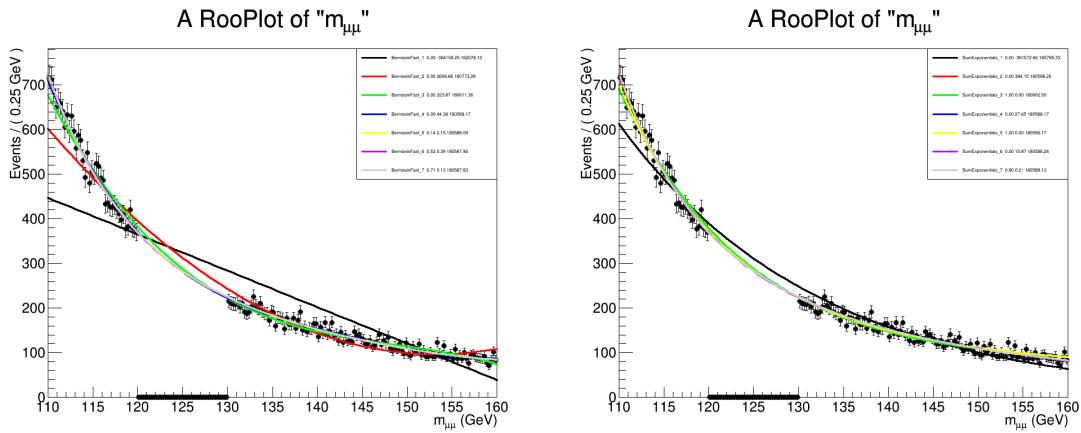


Figure 4-38. Examples of the F-Test for the Bernstein Polynomials (left) and the Sum of Exponentials (right).

4.10 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.

4.10.1 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 4-40.

4.10.1.1 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.

4.10.1.2 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. 4-39, 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.

4.10.1.3 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 $+6.8\%$.

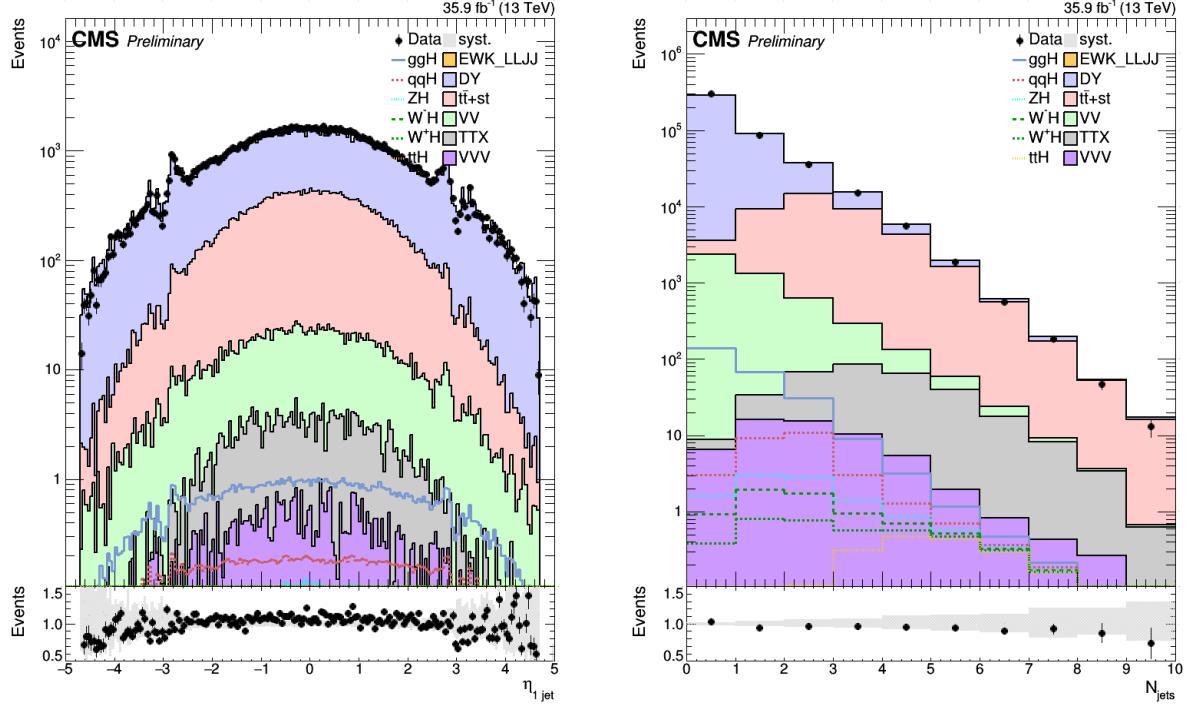


Figure 4-39. Pseudorapidity of the leading jet, and number of associated jets. The uncertainty bands show jet energy scale uncertainty.

- **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 4-40.
- **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.

4.10.1.4 Impact plots

The influence of the different uncertainties on the signal strength measurement are shown in Fig. 4-40. 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.

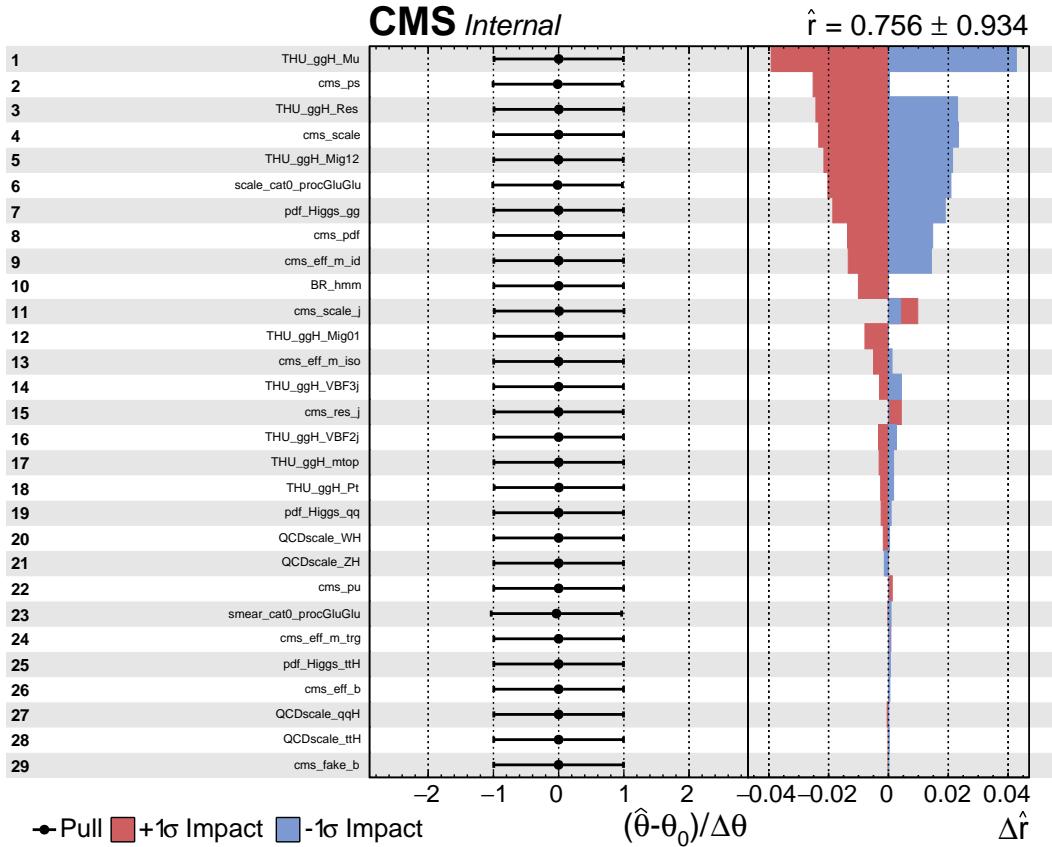


Figure 4-40. Impact plots derived from the Asimov data-set.

4.10.2 Background

The background is modeled by fitting the data, and the likelihood naturally accounts for the uncertainties in the fit parameters. In this regard, fitting the data leaves fewer systematics to account for. 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,

$$b_{ij} = \text{Median} \left[\frac{\hat{\mu}_{ij} - \mu}{\sigma_{\mu,ij}} \right] \quad (4-38)$$

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 (j2%). 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 ($b_{i,j} < 0.20$) with respect to all $B_j \in \beta$. The bias matrices for $m_H = 125\text{GeV}$ are shown in Figures 4-41 and 4-42. 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 $\beta = \text{BWZRedux} \cdot \text{poly}, \text{BWZGamma}, \text{SumExp}, \text{FEWZ} - \text{full} \cdot \text{poly}$. 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 .

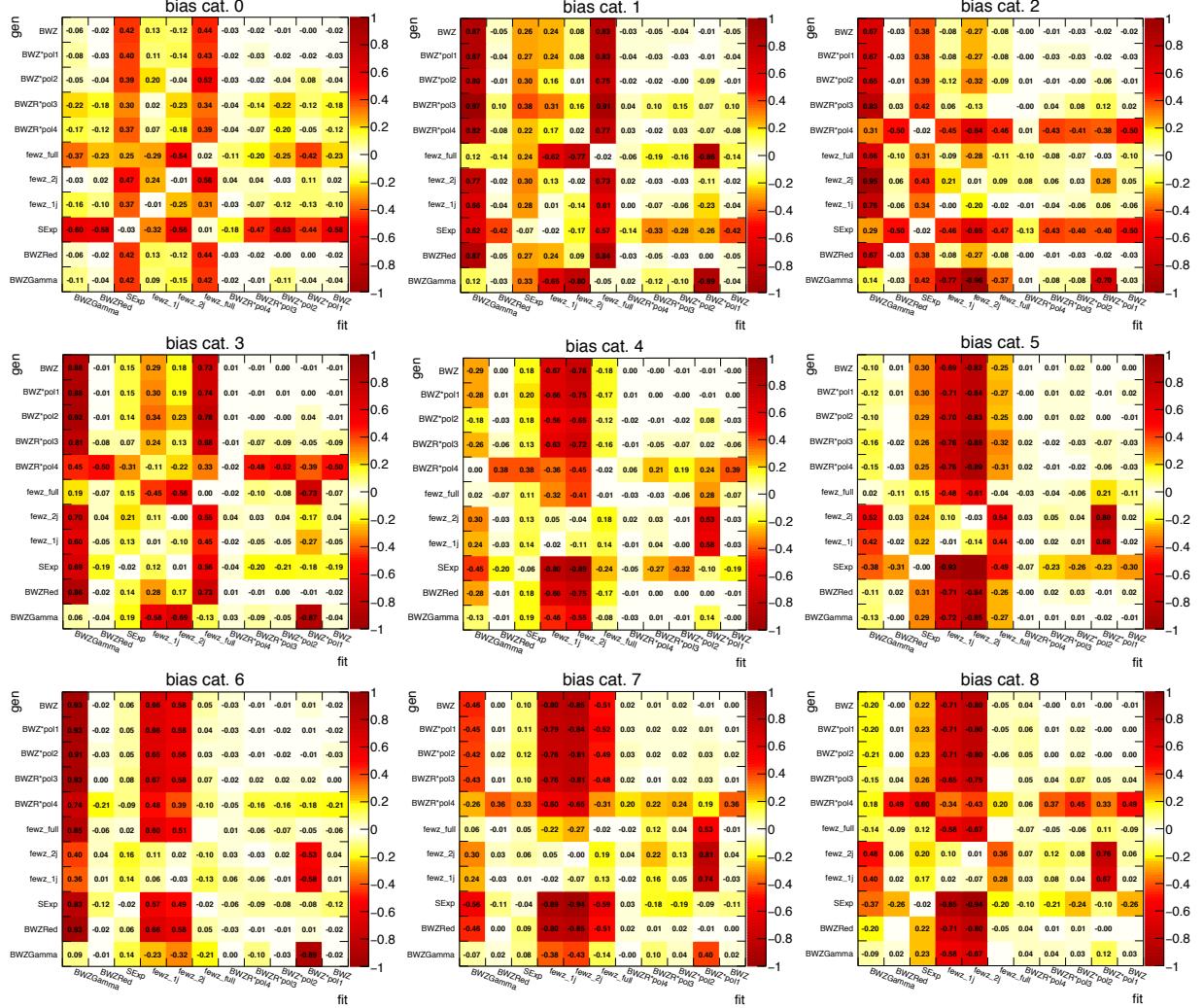


Figure 4-41. Bias results for categories 0-8 with each category's bias independent of the others. The results shown are for $m_H = 125$ GeV. The x-axis shows the fit model and the y-axis shows the pseudodata generating model.

The maximum observed bias for each category's least biased function is tabulated in Table 4-8. The maximum bias is calculated over $m_H = 120, 125, 130$ GeV. These functions are used for the final limit setting. If the statistics in a category are low enough, the uncertainty

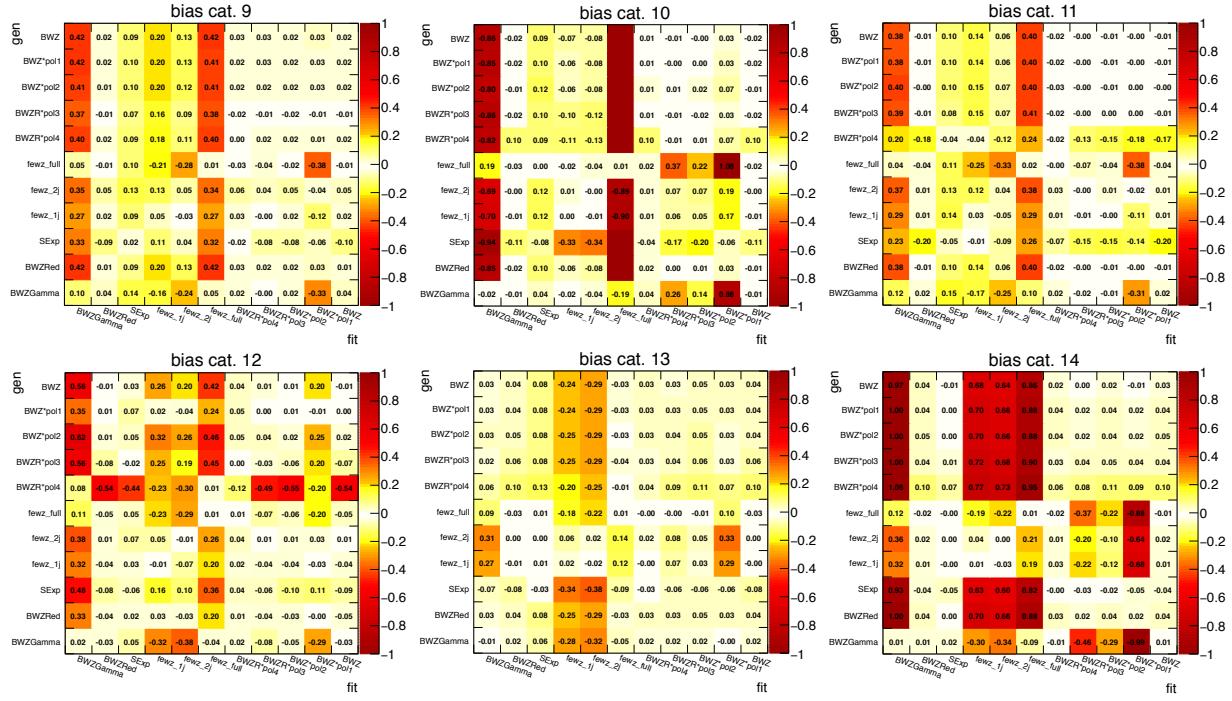


Figure 4-42. Bias results for categories 9-14 with each category's bias independent of the others. The results shown are for $m_H = 125$ GeV. The x-axis shows the fit model and the y-axis shows the pseudodata generating model.

Table 4-8. Maximum observed bias per category for least biased fit function. The maximum bias is over $m_H = 120, 125, 130$ 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 $\sigma_{\hat{\mu}}$ use the full statistics. The evaluations of Figure 4-43 estimate the bias on $\hat{\mu}$ using the combined likelihood over all categories with the fits of Figure 4-8 in each category. The final fits of 4-8 provide a net bias under 20% for

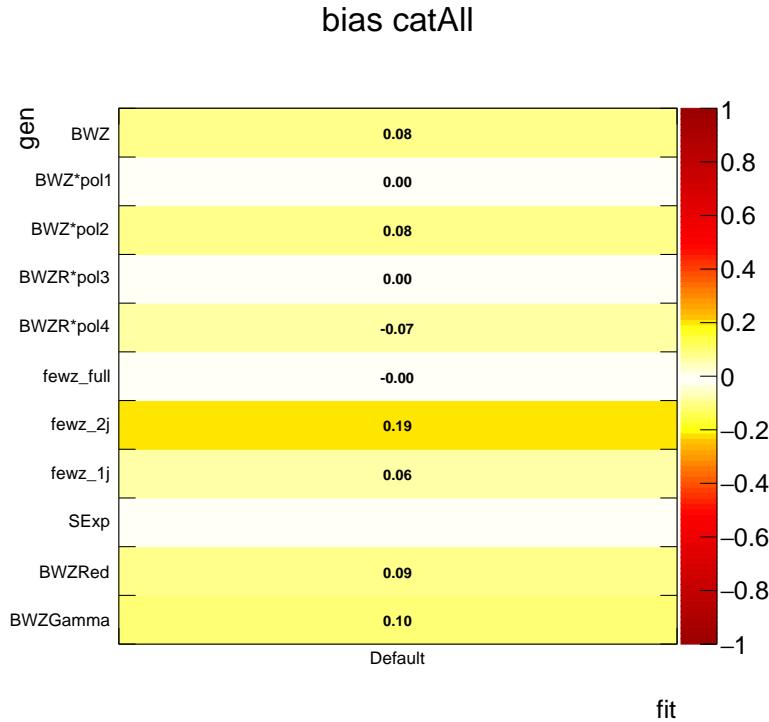


Figure 4-43. The plot shows the bias on $\hat{\mu}$ using the combined likelihood for all categories and the background fits of Figure 4-8. In this estimate, the same generating function is used for every category.

$m_H = 120, 125, 130 \text{ GeV}$. 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.

4.11 Results

With the signal and background models pinned down and the systematic uncertainties accounted for, 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 use the CLs method.

4.11.1 Results for 13 TeV Data

Figures 4-44 to 4-47 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 at 95% confidence is 2.64, while the expected (median) upper limit is 2.08 if the background only hypothesis is true. 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 is 0.98σ if the Higgs is SM. 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 4-48, and the individual fits per category are shown in Figures 4-49 and 4-49.

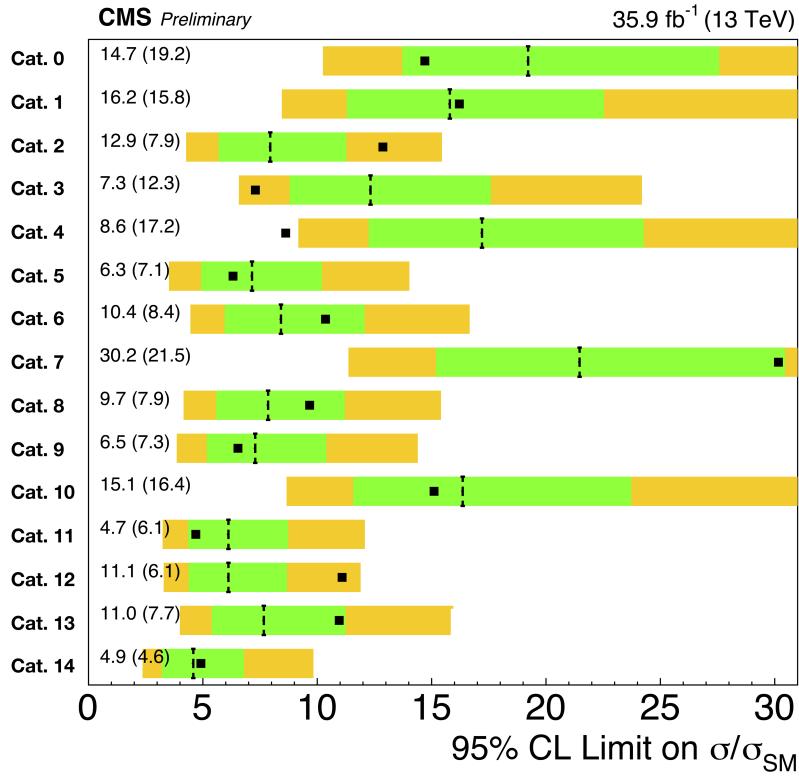


Figure 4-44. The upper limit on the signal strength at 95% confidence is presented for each category independent of the others. The results are for the 13 TeV data only.

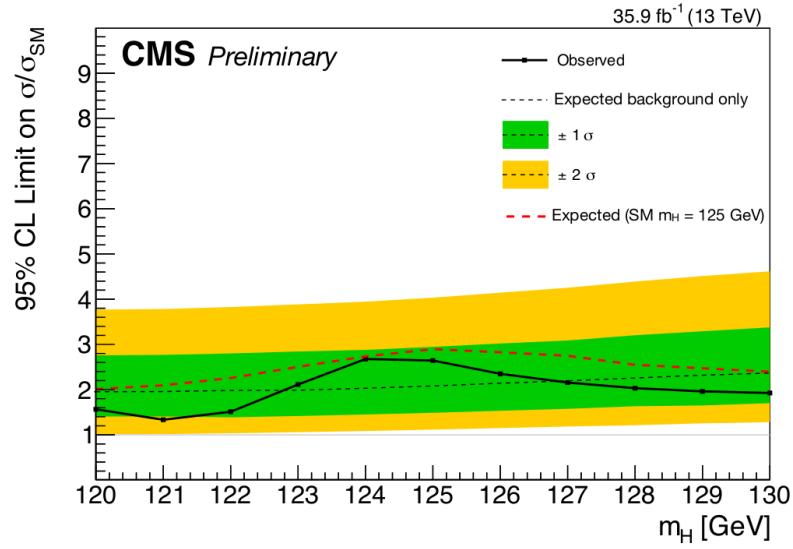


Figure 4-45. The combined upper limit on the signal strength at 95% confidence is presented for 13 TeV data. The red dashed line represents the expected (median) observation if the SM $m_H = 125 \text{ GeV}$ hypothesis is true. The black dashed line represents the expected observation if the background only hypothesis is true.

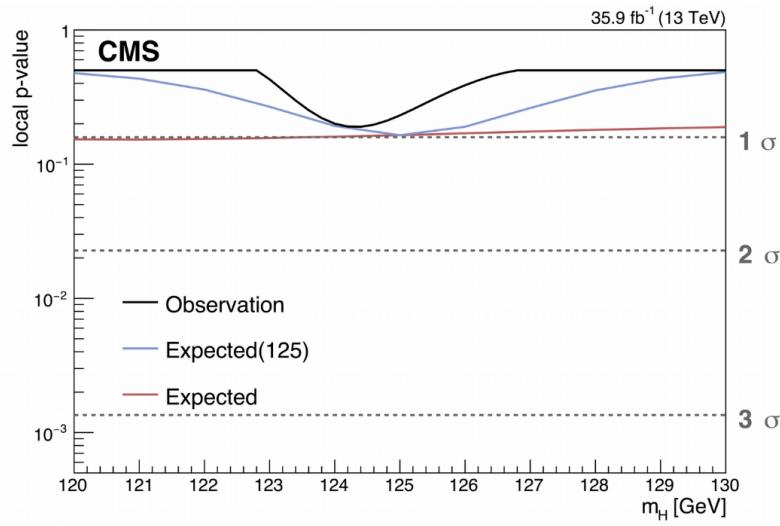


Figure 4-46. The p-value is presented for 13 TeV data. The black line is the observed p-value on the background only hypothesis. The blue line represents the expected (median) p-value on the background only if the SM $m_H = 125$ GeV hypothesis is true. The red line represents the expected p-value at $m_H = x$ GeV if the SM $m_H = x$ GeV hypothesis is true.

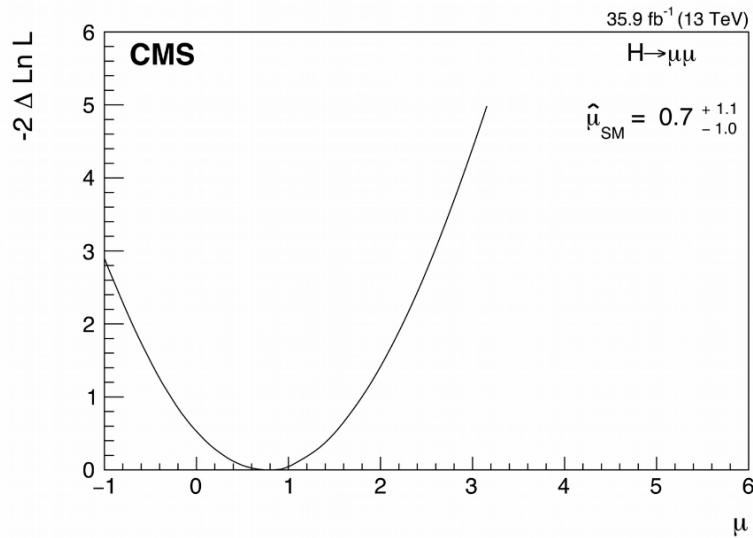


Figure 4-47. The negative log likelihood is plotted near the minimum for the signal strength μ , and the best fit, $\hat{\mu}$, is given.

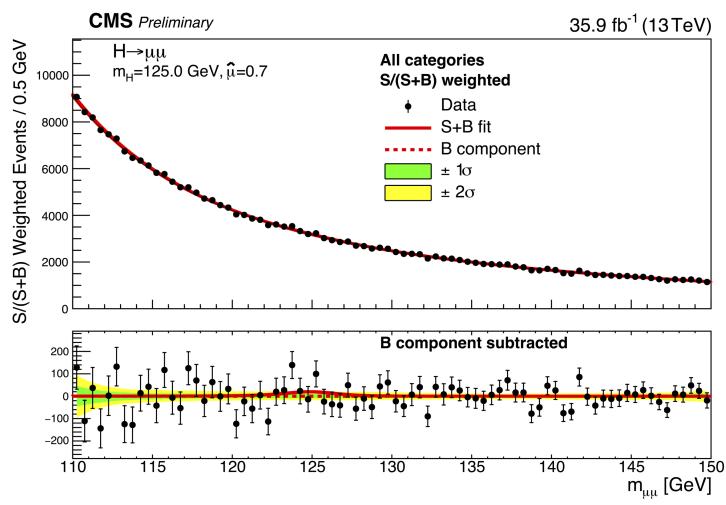


Figure 4-48. The weighted average of the signal + background fits over the categories. The fit from each category is weighted by $S/(S+B)$ to provide an idea of the overall shape while counting those categories with larger signal purity to a greater degree.

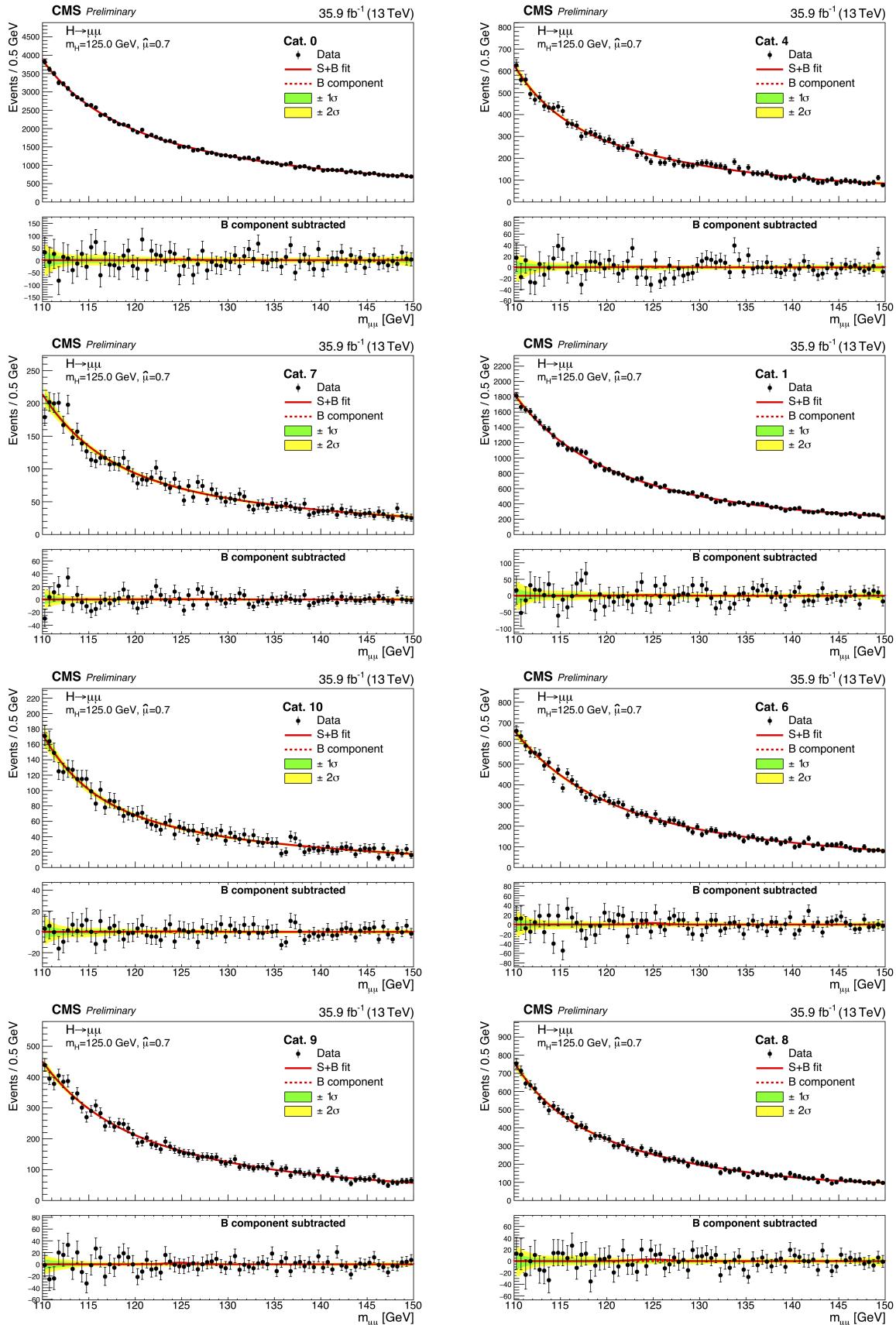


Figure 4-49. The final signal + background fits for the 13 TeV data, part one.

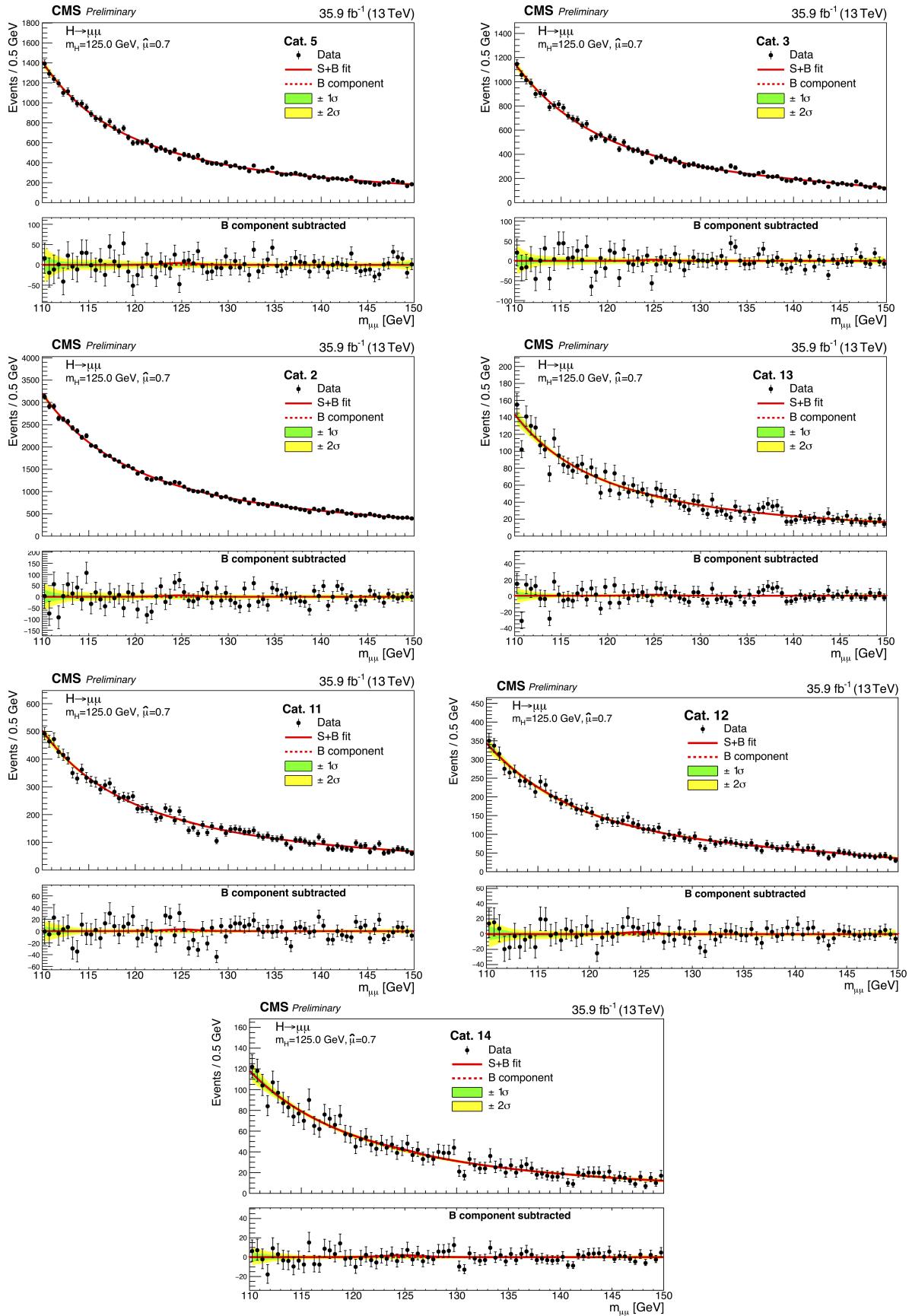


Figure 4-50. The final Signal + Background fits for the 13 TeV data, part two.

4.11.2 Results Combining 7, 8, and 13 TeV Data

Figures 4-51 to 4-53 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 $0.9^{+1.0}_{-0.9}$.

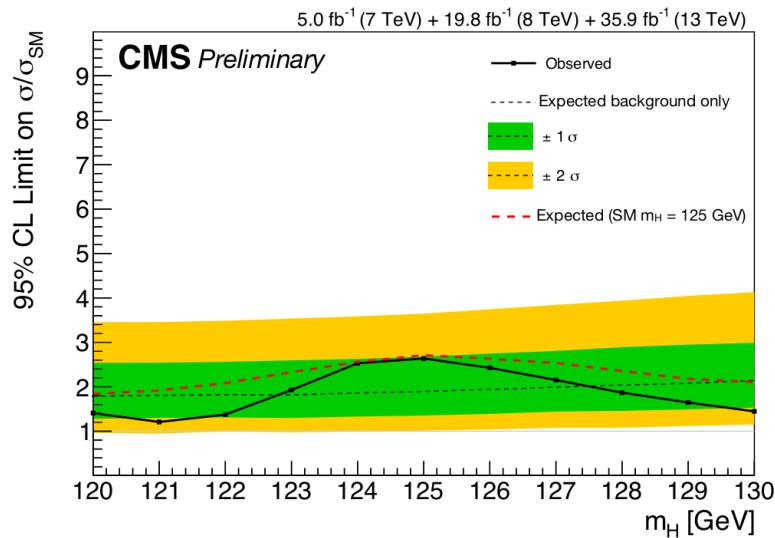


Figure 4-51. The upper limit on the signal strength at 95% confidence is presented for the combination of 7, 8, and 13 TeV data. The red dashed line represents the expected (median) observation if the SM $m_H = 125$ GeV hypothesis is true. The black dashed line represents the expected observation if the background only hypothesis is true.

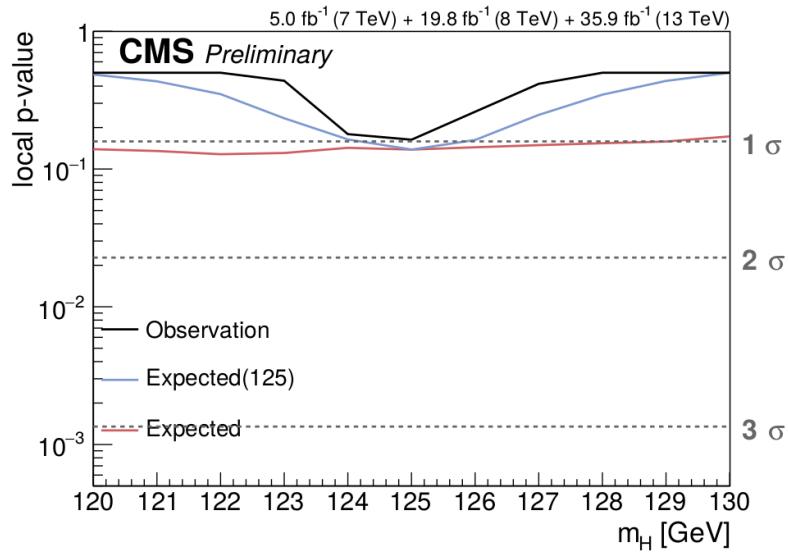


Figure 4-52. The p-value is presented for the combination of 7, 8, and 13 TeV data. The blue line represents the expected (median) p-value on the background only if the SM $m_H = 125$ GeV hypothesis is true. The red line represents the expected p-value at $m_H = x$ GeV if the SM $m_H = x$ GeV hypothesis is true.

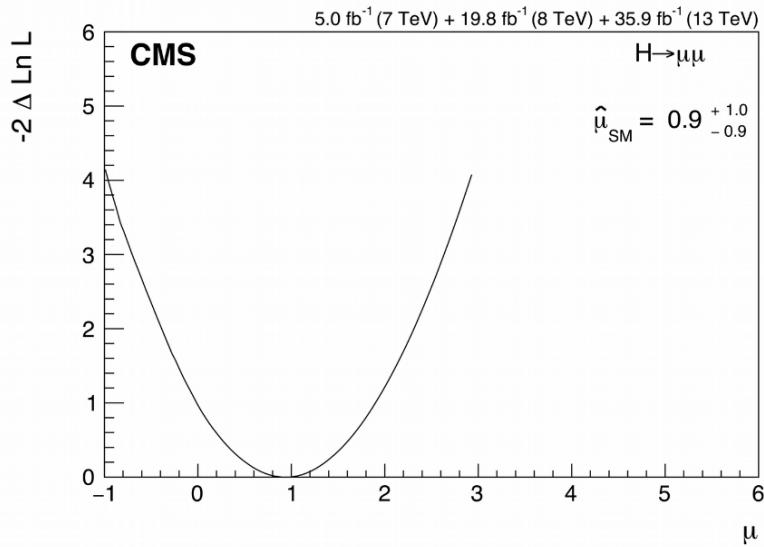


Figure 4-53. The negative log likelihood is plotted near the minimum for the signal strength μ , and the best fit, $\hat{\mu}$, is given.

4.11.3 Future Projections

In order to estimate the amount of data needed to observe a SM Higgs at 125 GeV, the expected background only p-value is plotted as a function of the integrated luminosity. See Figure 4-54. The calculation uses the signal and background PDFs from the 13 TeV analysis and scales them to different integrated luminosities. Note that 400 fb^{-1} of data is required to observe $H \rightarrow \mu^+ \mu^-$ with a p-value of 3σ .

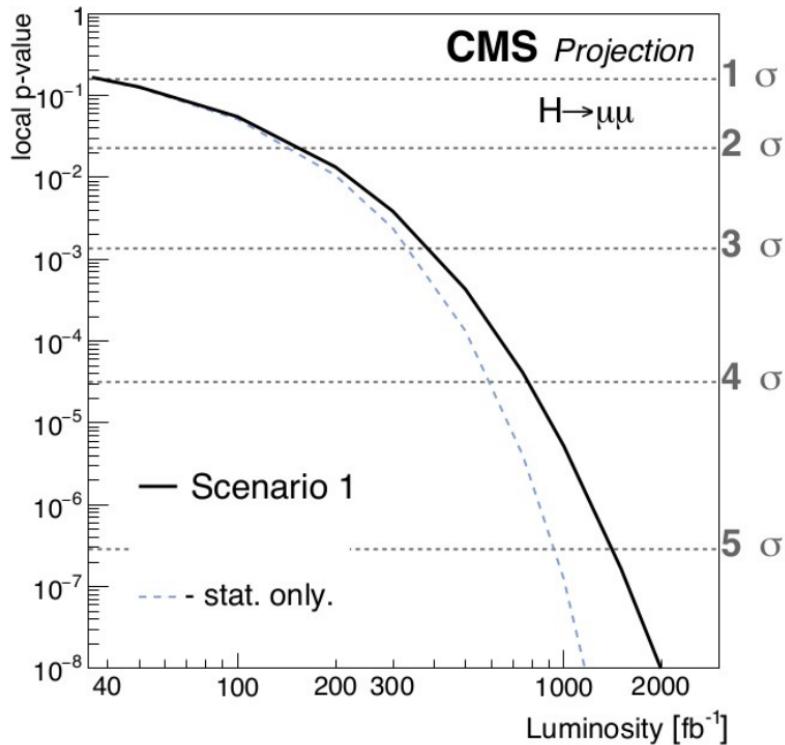


Figure 4-54. The expected p-value on the background only hypothesis assuming a SM Higgs with $m_H = 125 \text{ GeV}$ is plotted versus the amount of 13 TeV data in fb^{-1} .

REFERENCES

- 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 (2012).1: 1 – 29.
URL <http://www.sciencedirect.com/science/article/pii/S037026931200857X>
- _____. "Search for the bb decay of the standard model Higgs boson in associated (W/Z)H production with the ATLAS detector." *Journal of High Energy Physics* 2015 (2015).1: 69.
URL [https://doi.org/10.1007/JHEP01\(2015\)069](https://doi.org/10.1007/JHEP01(2015)069)
- CMS Collaboration. "The CMS muon project: Technical Design Report." Tech. Rep. CERN-LHCC-97-032, CMS-TDR-3, 1997.
- _____. "CMS Technical Design Report, Volume I: Detector Performance and Software." Tech. Rep. CERN-LHCC-2006-001, CMS-TDR-008-1, 2006.
- _____. "The CMS experiment at the CERN LHC." *Journal of Instrumentation* 3 (2008).08: S08004.
URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08004>
- _____. "CMS Detector." *CMS Web* (2011).
URL <http://cms.web.cern.ch/>
- _____. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." *Physics Letters B* 716 (2012).1: 30 – 61.
URL <http://www.sciencedirect.com/science/article/pii/S0370269312008581>
- _____. "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 (2013).6: 81.
URL [https://doi.org/10.1007/JHEP06\(2013\)081](https://doi.org/10.1007/JHEP06(2013)081)
- CMS collaboration. "The performance of the CMS muon detector in proton-proton collisions at s = 7 TeV at the LHC." *Journal of Instrumentation* 8 (2013).11: P11002.
URL <http://stacks.iop.org/1748-0221/8/i=11/a=P11002>
- CMS Collaboration. "Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons." *Journal of High Energy Physics* 2014 (2014a).5: 104.
URL [https://doi.org/10.1007/JHEP05\(2014\)104](https://doi.org/10.1007/JHEP05(2014)104)
- _____. "Evidence for the direct decay of the 125 GeV Higgs boson to fermions." *Nature Physics* 10 (2014b): 1038.
URL <http://dx.doi.org/10.1038/nphys3005>

- _____. “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 (2014c): 012003.
- URL <https://link.aps.org/doi/10.1103/PhysRevD.89.012003>
- _____. “CMS Collaboration.” *CMS Web* (2015).
- URL <http://cms.web.cern.ch/content/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>
- Evans, Lyndon and Bryant, Philip. “LHC Machine.” *Journal of Instrumentation* 3 (2008).08: S08001.
- URL <http://stacks.iop.org/1748-0221/3/i=08/a=S08001>
- Feynman, Richard P. (Richard Phillips) and Hibbs, Albert R. *Quantum mechanics and path integrals*. International series in pure and applied physics. 1965.
- Group, LHC Higgs Cross Section Working. “Higgs cross sections and decay branching ratios.” *CERN Wiki Page* (2015).
- URL <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>
- Han, Tao and McElrath, Bob. “ h^+ via gluon fusion at the LHC.” *Physics Letters B* 528 (2002).1: 81 – 85.
- URL <http://www.sciencedirect.com/science/article/pii/S037026930201208X>
- Higgs, Peter W. “Broken Symmetries and the Masses of Gauge Bosons.” *Phys. Rev. Lett.* 13 (1964a): 508–509.
- URL <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>
- _____. “Broken symmetries, massless particles and gauge fields.” *Phys. Lett.* 12 (1964b): 132–133.
- _____. “Spontaneous Symmetry Breakdown without Massless Bosons.” *Phys. Rev.* 145 (1966): 1156–1163.
- Hoecker, Andreas, Speckmayer, Peter, Stelzer, Joerg, Therhaag, Jan, von Toerne, Eckhard, and Voss, Helge. “TMVA: Toolkit for Multivariate Data Analysis.” *PoS ACAT* (2007): 040.
- Khachatryan, Vardan et al. “The CMS trigger system.” *JINST* 12 (2017).01: P01020.
- Lancaster, Tom and Blundell, Stephen J. *Quantum field theory for the gifted amateur*. Oxford: Oxford University Press, 2014.

URL <https://cds.cern.ch/record/1629337>

Mann, Robert B. *An Introduction to Particle Physics and The Standard Model*. Boca Raton, FL: CRC Press, 2010.

Mobs, Esma. "The CERN accelerator complex. Complexe des accrateurs du CERN." (2016). General Photo.

URL <https://cds.cern.ch/record/2197559>

Particle Data Group. "Review of Particle Physics." *Phys. Rev. D* 86 (2012): 010001.

URL <https://link.aps.org/doi/10.1103/PhysRevD.86.010001>

Peskin, Michael E and Schroeder, Daniel V. *An Introduction to Quantum Field Theory*. Boulder, CO: Westview, 1995.

Plehn, Tilman and Rainwater, David. "Higgs decays to muons in weak boson fusion." *Physics Letters B* 520 (2001).1: 108 – 114.

URL <http://www.sciencedirect.com/science/article/pii/S0370269301011571>

Sean Carroll. *Dark Matter, Dark Energy: The Dark Side of the Universe*. The Teaching Company, 2007.

Tapper, A and Acosta, Darin. "CMS Technical Design Report for the Level-1 Trigger Upgrade." Tech. Rep. CERN-LHCC-2013-011. CMS-TDR-12, 2013. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon 1, didier.claude.contardo@cern.ch.

URL <https://cds.cern.ch/record/1556311>

Weinberg, Steven. *The Quantum Theory of Fields, Volume 1: Foundations*. Cambridge University Press, 2005.

Wigner, E. "On Unitary Representations of the Inhomogeneous Lorentz Group." *Annals of Mathematics* 40 (1939).1: 149–204.

URL <http://www.jstor.org/stable/1968551>

Zee, A. *Quantum Field Theory in a Nutshell*. Nutshell handbook. Princeton, NJ: Princeton Univ. Press, 2003.

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.