

Abstract of “A Search for Exotic Particles that Decay into a Higgs Boson and a Photon with the CMS Detector at the CERN LHC” by John Hakala, Ph.D., Brown University, February 2019.

This dissertation presents a search for the signature of the production of previously unobserved particles being produced in the proton-proton collisions of the CERN Large Hadron Collider. The particular signature that this search focuses on is a resonant production of a Higgs boson in conjunction with a photon, a signal that various models that go beyond the standard model of particle physics predict, including theories with a Z' boson. A detailed analysis of the data collected by the Compact Muon Solenoid detector corresponding to 35.9 inverse femtobarns of data is performed to hunt for any “bumps” arising in the invariant mass spectrum of events that contained a photon and a jet consistent with the hadronic decay of a Higgs boson. In order to maximize the sensitivity of this search, new and innovative techniques are employed, including the application of new pile-up mitigation algorithms, advanced jet grooming algorithms, and machine-learning trained discriminating algorithms for identifying the distinctive hadronic jets that develop from the decay of a Higgs boson into a pair of bottom quarks. An overview of LHC physics and the CMS detector is given to contextualize this work.

A Search for Exotic Particles that Decay into a Higgs Boson and a Photon with the CMS Detector
at the CERN LHC

by

John Hakala

B. A., Washington University in St. Louis, 2011

Sc. M., Brown University, 2013

A dissertation submitted in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy
in the Department of Physics at Brown University

Providence, Rhode Island

February 2019

© Copyright 2019 by John Hakala

This dissertation by John Hakala is accepted in its present form by
the Department of Physics as satisfying the dissertation requirement
for the degree of Doctor of Philosophy.

Date _____

Greg Landsberg, Advisor

Recommended to the Graduate Council

Date _____

JiJi Fan, Reader

Date _____

David Cutts, Reader

Approved by the Graduate Council

Date _____

Andrew G. Campbell
Dean of the Graduate School

Preface

I would first like thank my advisor, Greg Landsberg, for his mentorship over the course of my doctoral studies. I owe the many wonderful opportunities I've had the great fortune to have gotten to Greg, and the work presented in this dissertation would not have been possible for me to accomplish without his insights and assistance.

This work and other aspects of my doctoral research would also not have been possible without the help and valuable lessons from my colleagues from Brown: David Yu, Ted Laird, Edmund Berry, Julie Hogan, and Martin Kwok. Thank you all for being excellent mentors and collaborators.

I give many thanks to Lesya Shchutska, Caterina Vernieri, and Lindsey Gray for sharing with me their expertise in the craft and art of particle physics analysis. Thank you also to Yuri Gershtein, Dinko Ferencek, Conor Henderson, Francesco Pandolfi, Halil Saka, and Dave Cutts for their very constructive reviews of this analysis.

I am very grateful to Seth Cooper, Chris West, Danny Noonan, Joe Pastika, Ted Kolberg, Yurii Maravin, Jim Hirschauer, and Dick Kellogg for all that they've taught me and all their help during my work CERN and Fermilab. Thank you all.

Thank you, Mom and Dad, for supporting me throughout my life and making it possible for me to pursue my life's goals.

Finally, thank you, Stephanie, for being a wonderful partner for so many years.

Contents

List of Tables	viii
List of Figures	x
1 Introduction	1
2 Background	3
2.1 The standard model	3
2.1.1 Quantum chromodynamics	4
2.1.2 Electroweak theory	6
2.2 Unsolved mysteries	9
2.2.1 Neutrinos	9
2.2.2 Dark matter	11
2.2.3 Gravity	11
2.2.4 The heirarchy problem	12
2.3 Beyond the standard model	13
2.3.1 Supersymmetry	13
2.3.2 Extra dimensional models	13
2.4 Searching for new physics	15
3 Particle Physics at the Large Hadron Collider	18
3.1 The Large Hadron Collider	18
3.1.1 LHC accelerator complex	19
3.1.2 Machine specifications	20
3.1.3 LHC collider physics	22
3.2 CMS Detector	26
3.2.1 Detector components	27
3.2.2 Trigger, timing, and data acquisition	37
3.2.3 Event reconstruction	41

4	Search for $H\gamma$ resonances	57
4.1	Introduction	57
4.2	Samples	58
4.3	Triggers	59
4.4	Object selection	62
4.4.1	Photons	62
4.4.2	Jets	62
4.4.3	MC background validation	64
4.5	Optimization	66
4.6	Search strategy	77
4.7	Background shape modeling	80
4.8	Signal shape modeling	85
4.9	Bias studies	92
4.10	Sidebands	95
4.11	Double-b tagging scale factors	99
4.12	800 GeV minimum photon+jet invariant mass	103
4.13	Limit calculation exercises	103
4.14	Lower bound of search region	107
4.15	Systematic uncertainties	109
4.16	Signal+background fits	113
4.17	Results	115
4.18	Conclusions	117
A	HCAL data acquisition	118
A.1	Overview	119
A.2	Local runs	123
A.2.1	TCDS	124
A.3	Configuration parameters	125
A.4	Masking	126
	Bibliography	129

List of Tables

2.1	A summary of diboson resonances physics. In the top line of each entry, CMS analyses performed in that channel are listed. Underneath, examples of models which may give rise to such resonances are listed. To denote the models, H denotes additional neutral scalar or pseudoscalar Higgs bosons, ρ denotes neutral or charged techni-rhos, and $\mathcal{Q}$ denotes quirks. [1]	17
3.1	History of integrated luminosity delivered by the LHC. Note that the 2017 pp collisions at 5.02 TeV were performed as a “reference run”. [2]	24
3.2	Nominal LHC operating specifications and beam characteristics [3] [4]. Note that the normalized transverse emittance was reduced from 3.4 μm to 2.0 μm during the 2016 run of the LHC due to the introduction of the “Batch Compression Merging and Splitting” technique in the Proton Synchrotron. Note also that the bunch crossing angle was reduced substantially from 80 μrad to 40 μrad at the beginning of the 2016 run.	25
4.1	Data samples used in this analysis. Listed here are the run numbers, which refer to individual runs of the CMS DAQ system, and the dates of the “re-reco” or subsequent offline rerunning of the CMS reconstruction algorithms.	58
4.2	SM MC background samples used. In the table, the symbol H_T denotes the scalar sum of transverse momenta of all jets in the events.	59
4.3	Final selection requirements in the analysis.	68
4.4	Cut flow, data events	69
4.5	Goodness-of-fit test statistics from the saturated models test, for various background shape fit models, antibtag category.	81
4.6	Goodness-of-fit test statistics from the Kolmogorov-Smirnov test, for various background shape fit models, antibtag category.	81
4.7	Goodness-of-fit test statistics from the saturated models test, for various background shape fit models, btag category.	81
4.8	Goodness-of-fit test statistics from the Kolmogorov-Smirnov test, for various background shape fit models, btag category.	82
4.9	χ^2 calculated for various background shape fit models at their optimal orders.	82

4.10	Goodness-of-fit test statistics for signal fits using the Kolmogorov-Smirnov test statistic: antibtag category.	85
4.11	Goodness-of-fit test statistics for signal fits using the Kolmogorov-Smirnov test statistic: btag category.	85
4.12	Systematic uncertainties	111

List of Figures

2.1	A diagram of the particle content of the Standard Model. The top portion of the diagram shows the SM with unbroken electroweak symmetry, whereas the bottom portion shows the SM with the Higgs at its vacuum expectation value, indicated by the position of the Higgs boson in its “Mexican hat” potential. Taken from [5]. . . .	10
2.2	Gluon fusion production of a Higgs boson. The protons emit gluons, which combine via a top loop to a Higgs boson, and the Higgs boson then decays via a top loop to two photons. This channel was the first channel studied by CMS and ATLAS that provided evidence for the Higgs.	15
2.3	A diagram for the signal model considered in this analysis: $Z' \rightarrow H\gamma$. The leading order diagram corresponds to the one-loop contributions of a fermion F to the $Z' \rightarrow H\gamma$ amplitude. Here, F can be either a top quark or an anomalon. From [6]. Though this particular model is the focus in this analysis note, it is expected that generic narrow $H\gamma$ resonances will have signatures very similar to this particular model.	17
3.1	The LHC inside the LEP/LHC tunnel, during assembly. In this photograph, two segments of the LHC machine structure are separated (they have not yet been connected), showing the interconnection between two segments of the beam pipe [7]. . .	19
3.2	A schematic drawing of the CERN accelerator complex, as well as several of the collider and fixed-target experiments [8].	21
3.3	Coordinate system used. Taken from [9].	26
3.4	A diagram showing the CMS detector with a roughly one-sixth cutaway made to show the interior components of the detector [10].	29
3.5	A diagram showing a cross section through the center of one quarter of the CMS detector. The DT, CSC, and RPC detectors are indicated. Taken from [11]	34
3.6	A sketch of various particle interactions in a wedge of the CMS detector cut through the transverse plane. The muon and pion are positively charged, while the electron is negatively charged. From [12].	45
3.7	A cartoon of the various structures in different types of jets. Notice the two-prong structure in the hadronic decays of boosted bosons. Borrowed from [13].	54

4.1	Trigger efficiency in MC and data. For the data, a logical OR of all SingleMuon dataset triggers is used as the reference trigger. For trigger efficiency vs. p_T , the check is done at preselection level, showing good efficiency but with some residual effects for the lower p_T bins. For trigger efficiency vs. $m_{J\gamma}$, we see some residual turn-on in the invariant mass spectrum up to around 750 GeV.	61
4.2	Efficiencies for various photon identification criteria, as a function of photon p_T , in the ECAL barrel region ($ \eta < 1.4442$). The left plot and right plot are the same except for a different range and binning. Using the $H\gamma$ signal samples, the efficiencies shown are calculated as a ratio: the numerator of the ratio is the number of events with reconstructed photons passing the ID criteria, requiring that the reconstructed photon have $ \eta < 1.4442$ and also be matched with a generator-level photon by requiring $\Delta R < 0.3$ relative to a generated photon. The denominator of the ratio is the number of events with a generated photon in the EB region.	62
4.3	A comparison of the pruned jet mass and PUPPI softdrop jet mass for several signal samples. It is observed that as the signal mass increases, the pruned jet mass corresponding to the Higgs peak drifts downward, while the PUPPI softdrop mass is more stable. This study is performed at the full selection level but without the b-tagging categories applied.	64
4.4	A comparison of signal efficiencies as a function of signal mass for the PUPPI softdrop jets and for CHS pruned jets. The spectra shown are made using the simulated signal samples, with the full selection applied in both categories. The PUPPI softdrop algorithm generally yields better signal efficiency.	65
4.5	A comparison of the shapes of the jet mass spectra near the signal region. For both jet grooming algorithms, no unusual structure is observed near the signal region. . .	65
4.6	Preselection distributions of kinematic variables	70
4.7	Pre-selection MC and data spectra. The spectra reproduce the shapes of the backgrounds reasonably well in all variables used in this analysis; however, cuts are derived using data-driven techniques.	71

4.8	Pre-b-tagging optimization of cuts on the maximum $ \eta $ for selected photons and jets. The MC background and signal samples are plotted with the requirement that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data points plotted have the requirement that the jet mass to be in the 100 GeV to 110 GeV sideband. Aside from the special cut on the jet's invariant mass, the cuts applied in the top plot of the left figure are all the cuts in Table 4.3, except for the cuts on $ \eta_\gamma $, the double-b tagger discriminator, and $p_T^\gamma/m_{J\gamma}$. In the bottom of the left figure, the cuts applied are the same but with a varying cut on $ \eta_\gamma $ applied. Likewise, in the right figure, for the top plot, all cuts in Table 3 are applied except for the cuts on $ \eta_J $, the double-b tagger discriminator, and $p_T^\gamma/m_{J\gamma}$, and for the top plot, a varying cut on $ \eta_J $ is also applied. The varying cuts for both bottom graphs require that $ \eta $ be less than the cut value indicated.	72
4.9	Optimization of the b-tagging selection. The value that separates the untagged and b-tagged categories is chosen to maximize the sensitivity in the b-tagging category. The MC background and signal samples are plotted with the requirement that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data points plotted have the requirement that the jet mass to be in the 100 GeV to 110 GeV sideband. In addition to the special cuts for the jet invariant mass, the cuts applied in the top plot are all the cuts in Table 4.3, except for the cuts on the double-b tagger discriminator and $p_T^\gamma/m_{J\gamma}$. In the bottom, the cuts applied are the same but with a varying cut on the double-b tagger discriminator value. The varying b-tagging selection value requires that the double-b tagger discriminant be greater than the value indicated.	73
4.10	Comparison of the double-b tagger with double subject b-tagging. The double-b tagger outperforms double-subjet b-tagging also at high resonance masses. The MC background and signal samples require that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data require that the event's jet mass be in the 100 GeV to 110 GeV sideband. The cuts applied for making the top plot are the same as in the bottom plot of Figure 4.9, but in the rest of the figures, instead of applying a cut on the double-b tagger discriminator value, different cuts on the leading and subleading subjects' CSVv2 tagger discriminator are applied corresponding to different choices of working points for the leading and subleading subjects.	74

4.11	Additional kinematical cut: the p_T^γ cut at 0.35 is chosen over the similar variable $\cos(\theta^*)$. The MC background and signal samples are plotted with the requirement that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data points plotted have the requirement that the jet mass to be in the 100 GeV to 110 GeV sideband. Along with the special cuts for the jet invariant mass, the cuts applied for the stack plots in the top two figures are all the cuts in Table 4.3, except for the cut on $p_T^\gamma/m_{J\gamma}$. For the graph in the bottom of the top left figure, as well as the green lines of bottom figure, the cuts applied are the same, except that a varying cut on $p_T^\gamma/m_{J\gamma}$ is applied. Likewise, for the graph in the bottom of the top right figure and the red lines of bottom figure, the cuts are again all the cuts in Table 4.3, except no cut on $p_T^\gamma/m_{J\gamma}$ is applied, and instead a varying cut on $\cos(\theta^*)$ is applied. The varying cut on $p_T^\gamma/m_{J\gamma}$ requires that it be greater than the cut value indicated, whereas the varying cut on $\cos(\theta^*)$ requires it to be less than the cut value indicated.	75
4.12	Optimization of sensitivity-maximizing cuts using 50-70 GeV PUPPI softdrop jet mass sideband	76
4.13	No cut on τ_{21} is applied in this analysis, as the modest gains in signal sensitivity for low mass signals would introduce a systematic uncertainty that would not increase the sensitivity of this analysis.	76
4.14	An illustration showing the basic concept of the “bump hunt” strategy employed in this analysis. In the illustration, m_x denotes the mass of the resonance, the smoothly monotonically falling black line is the background fit, and the blue peak is the bump that would arise due to the presence of a narrow resonance. Borrowed from [14]. . .	77
4.15	$m_{J\gamma}$ spectrum before application of the weighting applied to mitigate the trigger turn-on effects. The black points show the data in the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband, while the Monte Carlo backgrounds show the signal region, where the turn-on is more pronounced. To avoid pathological fits due to this turn-on effect, this analysis requires the photon-jet invariant mass to be above 720 GeV, and applies a weight to the accepted events to correct for residual turn-on effects. The cuts applied for this plot are the full analysis cuts except for the $m_{J\gamma}$ minimum, and with no b-tagging categorization applied.	78
4.16	The effects of the reweighting applied to the data. The fractional difference plotted is calculated as $(N_{\text{weighted}} - N_{\text{unweighted}})/N_{\text{weighted}}$. This plot is made using the data in the signal region.	79
4.17	Several fits for the signal region spectra. The function chosen is the dijet2 function, based on studies of the sideband region, where its performance in the KS test is within uncertainty or better than a higher-order dijet function, its relative good performance versus all other functions in both categories, and its χ^2 within uncertainty of other competitive functions.	83

4.18	Shown in the top panels of subfigures (a) and (b) are the final background fits of the $m_{J\gamma}$ spectra in both categories, along with the fit uncertainties calculated using the sampling method described above. Shown in the bottom panels of both subfigures are data's pulls with respect to the fit function, given both statistical and fit uncertainties.	84
4.19	Simulated signal sample fits for the untagged category. The crystal ball fit is shown in red.	86
4.20	Simulated signal sample fits for the untagged category, continued. The crystal ball fit is shown in red.	87
4.21	Simulated signal sample fits for the b-tagged category. The crystal ball fit is shown in red.	88
4.22	Simulated signal sample fits for the b-tagged category, continued. The crystal ball fit is shown in red.	89
4.23	Simulated signal efficiency for the b-tagged and untagged categories fitted using three-parameter functions.	90
4.24	Results of signal shape interpolation. A signal shape is interpolated at every 10 GeV interval between points where a signal was simulated using fullsim. Shown here are only a subset of the interpolated signals.	90
4.25	Interpolation closure test: untagged category.	91
4.26	Interpolation closure test: b-tagged category.	91
4.27	Results of bias studies in the b-tagged and untagged categories carried out using the signal region data. The profile plots characterize the distribution of the medians and means of the pull (i.e. signal significance divided by its uncertainty) . In the b-tagged category, the pull means deviate from 0 at the point where the expected number of events gets close to zero; this is an effect of the best signal+background fit achieving its best fit when the magnitude of the signal is large and negative. (Note that the combine tool will set any negative values for the signal+background fit to 0, and thus a large negative signal component simply forces the signal+background fit to zero at masses at and above the signal mass; therefore in cases when there are empty bins near the tail of the spectrum, a large negative signal artificially attains the best fit.) However, the medians of the pull distributions are stable since the median is not as sensitive to the large-negative signal outliers.	94
4.28	Comparisons of the spectra of several kinematic variables for the data in two different sidebands, plotted along with the spectrum of MC background samples in the signal region. All cuts are applied except for the b-tagging categorization cut.	96
4.29	Comparisons of the spectra of several kinematic variables for the data in two different sidebands, plotted along with the spectrum of MC background samples in the signal region. All cuts are applied except for the b-tagging categorization cut.	97

4.30	Comparisons of the spectra of several kinematic variables for the data in two different sidebands, plotted along with the spectrum of MC background samples in the signal region. All cuts are applied except for the b-tagging categorization cut.	98
4.31	A comparison of the photon+jet invariant mass spectra in the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ and $50 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 70 \text{ GeV}$ sideband regions. The shapes generally agree within uncertainties. This leads to the expectation that the shape of the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband spectrum gives a good prediction of the shape of the signal region.	99
4.32	b-tagged category: Comparisons of the spectra of several kinematic variables for the Monte Carlo background and signal samples with and without the b-tagging scale factors applied.	100
4.33	Untagged category: Comparisons of the spectra of several kinematic variables for the Monte Carlo background and signal samples with and without the b-tagging scale factors applied.	101
4.34	Untagged category: Comparisons of the spectra of several kinematic variables for the Monte Carlo background and signal samples with and without the b-tagging scale factors applied.	102
4.35	Fits of the $m_{J\gamma}$ spectra for events with $m_{J\gamma} > 800 \text{ GeV}$, with no reweighting applied to correct for residual trigger inefficiency. These fits are performed on the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband. It is observed that the fits in the untagged category are nearly identical to the corresponding fits shown in Figure 4.17, however, in the b-tagged category, the fits demonstrate unstable behavior. This instability in the fits results because the region $720 \text{ GeV} < m_{J\gamma} < 800 \text{ GeV}$ is the most densely populated region of the spectrum, and removing those events causes the fits to be underconstrained.	103
4.36	An exercise in calculating limits in untagged, b-tagged, and combined channels given as 95% confidence level limits on cross-section times branching fraction, where instead of performing the analysis strategy on the signal region data spectrum, the signal+background fits are performed on the sideband (i.e. the PUPPI softdrop jet mass is required to be between 100 GeV and 110 GeV) invariant mass spectrum. The background fit is derived from this sideband spectrum, while the signal shapes are derived from MC using the signal window (jet mass between 110 GeV and 140 GeV).	104
4.37	An exercise in calculating limits in untagged, b-tagged, and combined channels given as 95% confidence level limits on cross-section times branching fraction, where instead of performing the analysis strategy on the signal region data spectrum, the signal+background fits are performed on the MC background samples' invariant mass spectrum in the signal region.	105

4.38	An exercise in calculating limits in untagged, b-tagged, and combined channels given as 95% confidence level limits on cross-section times branching fraction, where instead of performing the analysis strategy on the signal region data spectrum, the signal+background fits are performed on the MC background samples' invariant mass spectrum in the signal region.	106
4.39	A demonstration of the kinematic turn-on effect observed, shown in the MC background samples, with the full analysis cuts except for the b-tagging categorization. In this plot, one can observe that as the $m_j^{\text{PUPPI+SD}}$ jet mass window is slid toward higher jet masses, the peak in the $m_{j\gamma}$ moves toward higher masses. The lower bound of the analysis' search region is chosen such that the bump-hunt analysis strategy is carried out only in the region beyond the turn on corresponding to the nominal $m_j^{\text{PUPPI+SD}}$ mass window.	107
4.40	An example of a fit of the turn-on region using the $\exp * \text{erf}$ function. The full $\exp * \text{erf}$ fit is shown in red, and the $m_{j\gamma}$ distribution from the MC background samples is shown in blue. Also shown are the independent $\exp$ and erf components of the fit function: the dark dotted gray line is the pure $\exp$ component of the fit function, while the light dotted gray line is the erf component of the fit function, normalized so that its maximum is equal to the maximum of the MC $m_{j\gamma}$ spectrum. The cuts applied for this plot are the full analysis cuts, except for the b-tagging categorization.	108
4.41	A plot of the exponential term's coefficient b resulting from the $\exp * \text{erf}$ fits as a function of the $m_j^{\text{PUPPI+SD}}$ mass window's center. This fit is then used to provide an estimate for the value of b in the second step of the method, where the fit is repeated but with the value of b fixed. This fit is performed in the MC background samples, using the full analysis cuts except for the b-tagging categorization.	109
4.42	A plot of the calculation of the point at which the $\exp * \text{erf}$ function reaches the "99% turn-on point," defined as the point at which the error function component reaches .99. The vertical line indicates the point at which the error function is equal to .99. This fit is performed with a fixed value of the exponential term's coefficient b , which is done to ensure that the determination of the 99% point is not influenced by the "bending" of the exponential component of the fit function. This plot is performed using the MC background samples, using the full analysis cuts except for the b-tagging categorization.	110
4.43	The final determination of the lower bound of the analysis' search region. The 99% turn-on points as determined using the method shown in Figure 4.42 are plotted as a function of the center of the $m_j^{\text{PUPPI+SD}}$ mass window. A fit is made of this plot, and the lower bound of the search region is taken as the 99% turn-on point in the analysis' $110 \text{ GeV} < m_j^{\text{PUPPI+SD}} < 140 \text{ GeV}$ mass window. This study is done on MC background samples with the full analysis cuts except for the b-tagging categorization.	110

4.44	Changes in signal efficiencies due to varying the double-b tagging scale factors up and down by their uncertainty. Plotted is the difference of efficiency with the scale factor varied upwards by one uncertainty minus the efficiency with the scale factor varied downwards by one uncertainty, divided by the nominal efficiency. As the fractional difference is less than 8%, the systematic uncertainty is conservatively taken to be 4%.	112
4.45	Examples of signal+background fits obtained. Note that these $S + B$ fits are purely for is a demonstration purposes, since the actual limit setting procedure proceeds involves fitting many spectra that have had nusiances applied to account for systematic uncertainties. The signal masses used in the $S + B$ fits shown in these figures are chosen because they represent where the observed signal strength was the largest.	114
4.46	95% confidence level upper limits on cross section times branching fraction for a narrow resonance decaying to $h\gamma$, calculated in the in untagged, b-tagged, and combined channels.	116
A.1	A screen capture of the interface used by shifters to control CMS during normal data-taking. This is the graphical user interface for the level-0 function manager, which is the highest-level function manager of the hierarchy of function managers.	120
A.2	The RCMS state machine. This defines the states that the level-0 FM and all level-1 FMs may have. The solid ovals represent steady states, which the FM may remain in indefinitely; the dashed ovals represent transitional states, which are temporarily passed through to move to the next state, and the underlined words are the commands which initiate the state transition [15]. Some states in the diagram are not used by the HCALFM, but would be allowed if a purpose arose for them, including Fixing-SoftError and RunningSoftErrorDetected (which are used primarily by the tracker and ECAL function managers). The main steady states used by HCAL are Initial, Halted, Configured, and Running, and the RunningDegraded state is used for alerting shifters to serious problems with HCAL.	121
A.3	A schematic of the level-0 FM containing a level-1 HCALFM, which in turn controls level-2 HCALFMs that control xdaq applications.	123
A.4	A screen capture of the HCALFM graphical user interface showing the ability for users to select a local run key.	126
A.5	A screen capture of the HCALFM graphical user interface showing the multipartition masking panel.	127

Chapter 1

Introduction

One takes up fundamental science out of a sense of pure excitement, out of joy at enhancing human culture, out of awe at the heritage handed down by generations of masters.

Leon Max Lederman [16]

In the Presidential Address to the Royal Society of 1928, Ernest Rutherford, acclaimed for his many contributions to nuclear physics, gave a compelling case for why it would be extremely beneficial to nuclear research “to have a supply of electrons and atoms of matter in general, of which the individual energy of motion is greater even than that of the alpha particle,” since, at that time, alpha particles emitted by the decay of radium had the highest energy of any particles that could be produced in the laboratory [17]. Ernest O. Lawrence, then a freshly minted Professor at the University of California, took great inspiration from Rutherford’s charge and began his efforts to mount a “successful attack on the nucleus” by means of artificially accelerating particles to then unheard-of energies. With very little knowledge of the nascent field of accelerator physics (the only published work in the field of accelerator physics that he was aware of was a paper by the Norwegian physicist Rolf Widerøe, written in German, a language Lawrence did not understand!) Lawrence worked out the physics of what is known known as a cyclotron, and developed a design that he, along with his graduate student, Niels Edelfsen, constructed in 1930 [18]. Made of two semicircular electrodes with an alternating voltage, and held between the poles of an electromagnet, the world’s first cyclotron was no larger than a hockey puck, with a diameter of only four inches. This first device was quickly followed up by a series of ever-larger cyclotrons, and by 1939, Lawrence’s laboratory reported the first artificial production of radioactive isotopes. In the meantime, John Cockcroft and Ernest Walton at Cavendish Laboratory had designed an electrostatic accelerator they used to cause the first artificial disintegration of a lithium nuclei into two helium nuclei by bombarding it with

protons.

Despite these early achievements related to nuclear physics, accelerators did not immediately become especially relevant to the physics of fundamental particles. Major breakthroughs in fundamental particle physics continued to come from cosmic ray physics studies, for example the discoveries of the positron (1932), the muon (1936), the pion (1947), the K meson (1947), and the Λ^0 baryon (1950). However, Lawrence continued to push for bigger and better accelerators, and by the late 1940's, innovations by Edward McMillan led to the creation of a new type of accelerator called a synchrotron, which varied the magnetic field used to confine the accelerated particles with their energy in order to keep them in a circular path, including, importantly, the Bevatron at Lawrence Berkeley National Laboratory. Within a year of the beginning of the operation of the Bevatron, in 1955 the antiproton was discovered, which led to the Nobel Prize in Physics being awarded to Emilio Segrè and Owen Chamberlain. From this point onward, most of the major experimental achievements in fundamental particle physics came from experiments that utilized accelerators:

- The 1959 discovery of the neutral Ξ baryon at the Bevatron
- The 1962 discovery of the muon neutrino using the Alternating Gradient Synchrotron at Brookhaven National Laboratory (BNL)
- Late 1960's deep inelastic scattering experiments at the Stanford Linear Accelerator Center (SLAC) that revealed the parton structure of baryons and mesons
- The 1974 discovery of the J/ψ independently by groups at SLAC and BNL
- The 1975 discovery of the τ lepton at SLAC that implied the existence of the charm quark
- The 1977 discovery of the Υ meson at the Fermilab Proton Center Laboratory that implied the existence of the bottom quark
- 1979 experiments at DESY that gave evidence for the existence of gluons
- The 1983 discovery of the W and Z bosons at the Sp $\bar{p}$ S at CERN
- The 1995 discovery of the top quark at the Tevatron, and the 2000 observation of the τ neutrino at the Tevatron.

All of these experimental achievements moved hand-in-hand with developments on the theoretical side, with new experimental discoveries leading to new theories, and new theories guiding the way for experiment. And this progress continues to the present day, with the notable discovery in 2012 of the Higgs boson at the LHC at CERN. The discovery of the Higgs boson represented a very important final piece in the theory that had been built up along the ride – the Standard Model of Particle physics. And there is still much more work to be done looking for any hints of exciting new developments at the LHC! In this spirit, this dissertation presents a work done to continue the long and bountiful development and refinement of the field of particle physics.

Chapter 2

Background

2.1 The standard model

The Standard Model (SM) of particle physics is a quantum field theory that provides a unified description of the elementary particles of nature, the fundamental forces that act between them, and their dynamics. The building blocks of the SM are two families of fermions, quarks and leptons, which can be thought of as the matter particles, plus the bosons, which can be thought of as the force carriers between the fermions. Quarks come in six types, up (u), down (d), charm (c), strange (s), top (t), and bottom (b), and likewise, there are six types of leptons, three of which carry electric charge – the electron (e), muon (μ), and tau (τ) – and three neutral partners of the charged leptons, called neutrinos – the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). Each quark and lepton is described as a Weyl fermion field with spin $\frac{1}{2}$, which represents the lepton and its partner antiparticle, and all of the Weyl fermion fields with the exception of the neutrinos come in both left-handed and right-handed chiralities, and thus can be combined into a Dirac fermion description. Neutrinos (antineutrinos), on the other hand, are only modeled using left-handed (right-handed) Weyl fermion fields, as right-handed (left-handed) neutrinos (antineutrinos) have never been observed in experiment and are thus not a part of the SM.

The interactions between the SM’s matter particles are modeled by spin 1 vector fields that describe two “sectors” of the standard model, the quantum chromodynamics sector, which describes the strong nuclear force, mediated by gluons (g) carrying the so-called color charge, and the electroweak sector, which encompasses both the electromagnetic and weak nuclear forces. In addition to the vector boson and fermion fields of the SM, one final field, a spin 0 scalar field, known as the Higgs field, comprises the “Higgs sector” of the SM, which corresponds to the Higgs boson (H), which couples to the electroweak fields and the quarks, but not the gluons. Unlike the other fields, which may not couple to a potential energy function because of the requirement that the potential be gauge invariant, the Higgs field couples to the so-called Higgs potential, which leads to the Higgs

field taking on a nonzero vacuum expectation value. The Higgs field taking on a nonzero expectation value in the vacuum state leads to an electroweak sector that can be expressed as having one massless Goldstone boson, which can be identified with the photon, while the other modes of the (hitherto massless) electroweak sector gain a term in the SM Lagrangian that is equivalent to the fields acquiring a mass. Similarly, the broken symmetry of the Higgs field leads to the quarks and leptons gaining similar mass terms.

After such a breaking of the electroweak symmetry via the Higgs mechanism, the model can be interpreted in terms of the familiar three fundamental forces of particle physics: the strong, weak, and electromagnetic forces, which act between the massive quarks and leptons. With the exception of the neutrinos, direct measurements have confirmed that all the quarks and leptons have a nonzero mass, and though all direct measurements of the mass of neutrinos have been unable to detect a nonzero mass of the neutrino, recent observations of the phenomenon of neutrino oscillations indicate that neutrinos must have a small but nonzero mass, following from the Higgs mechanism.

This section gives an overview of the important elements of the SM, with a particular focus on the Higgs boson, photons, and the quantum chromodynamics phenomena that lead to jets in high-energy hadron collisions.

2.1.1 Quantum chromodynamics

Quantum chromodynamics is the theory of quarks, which combine to form baryons (such as protons and neutrons, made of three quarks), mesons (such as pions and kaons, made of two quarks), and other, more exotic hadrons (such as tetraquarks), and their interactions via the exchange of gluons, which is referred to as the strong nuclear interaction.

2.1.1.1 Lagrangian formalism

Quarks carry a conserved quantum number called the color charge, which comes in three varieties, red (r), green (g), and blue (b), which obey a local $SU(3)$ gauge symmetry. The gauge bosons that mediate the color interaction, referred to as gluons, therefore correspond to the eight generators of the $SU(3)$ gauge group, and are represented with eight spin-1 gluon fields A_μ^a , where μ is an index running over the spacetime dimensions 0-3 (and likewise for ν), and the index a runs from 1 to 8 for each of the gluons. The Lagrangian of the theory, following Yang-Mills theory, takes the form:

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{\psi}_{q,i} \left(i (\gamma^\mu D_\mu)_{ij} - m_q \delta_{ij} \right) \psi_{q,j} - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \quad (2.1)$$

where ψ_i are quark fields, with the flavor indices i and j running over each of the six flavors (u, d, c, s, t, b) of quarks, $G^{\mu\nu}$ is the field strength tensor of the gluon fields, defined as:

$$G_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g_s f^{abc} A_\mu^b A_\nu^c \quad (2.2)$$

with δ_{ij} denoting the Kronecker delta, γ^μ are the Dirac matrices, and f_{abc} are given by the commutation relationship for the Gell-Mann matrices:

$$[\lambda_a, \lambda_b] = if_{abc}\lambda^c \quad (2.3)$$

and finally, D_μ represents the gauge covariant derivative:

$$D^\mu = \partial^\mu + i\frac{g}{2}\lambda_a A_a^\mu \quad (2.4)$$

where g_s is the dimensionless QCD gauge coupling constant, and m_q are the quark masses, which depend on the quark flavor. The quark masses and the strong coupling constant represent fundamental constants of the theory of QCD and must be determined from experiment. [19] [20]

2.1.1.2 Renormalization

The gauge coupling constant g_s , which is alternatively denoted as $\alpha_s = \frac{g_s^2}{4\pi}$, and the quark masses m_q are observed to take on different values depending on the relevant energy or momentum scales, denoted by the symbol μ_R^2 , in the experiments being modeled. This phenomenon, also familiar in electroweak theory, is known as the “running of the coupling”. Using the Feynman diagrams and Feynman rules of QCD, a perturbative approximation reveals how the coupling runs as a function of μ_R^2 leading to the general relationship:

$$\alpha_s(\mu_s^2) = \frac{\alpha_s(\widehat{\mu_s^2})}{1 + \frac{\alpha_s(\widehat{\mu_s^2})}{12\pi} (11n - 2f) \ln\left(\frac{\mu_s^2}{\widehat{\mu_s^2}}\right)} \quad (2.5)$$

where $\widehat{\mu_s^2}$ represents an experimental energy where the value of the coupling α_s is known, f is the number of quark flavors, and n is the number of colors in the theory, which are $f = 6$ and $n = 3$ in the SM. An important implication of this is that the strong coupling becomes weaker as the experimental energy increases, which is a phenomenon known as “asymptotic freedom.” This can also be interpreted as implying that quark interactions at very short length scales become negligible. [19]

2.1.1.3 Confinement and hadronization

Conversely, the fact that the strength of the strong interaction grows as the length scales between quarks increases suggests the mechanism for why quarks are always observed in nature to live in color-neutral bound states with other quarks. This mechanism can be understood by considering, for example, the quarks that make up a proton, which is composed of two up quarks and a down quark. If one of the quarks in the proton is imparted some very large momentum by, say, a collision with some other particle and travels away from the proton’s other quarks, then, as its distance

to the other quarks increases, the attractive gluon field builds an increasingly large energy density as the outgoing quark increases its distance. As the gluon field gains more and more energy, the probability of generating quark-antiquark pairs from the vacuum grows, and inevitably before the outgoing quark can free itself from the proton, a particle-antiparticle pair is formed, with one quark of the newly generated pair returning to the proton, and the other binding with the outgoing quark to form a meson. [20]

This phenomenon, called “hadronization” explains the abundance of jets (i.e. roughly collinear streams of mesons and their decay products) produced in high energy proton-proton collisions, the fact that unbound quarks are never observed in experiments, and the observed “accidental” (believed to only be approximate) symmetry of baryon number conservation in nature. [21]

2.1.2 Electroweak theory

Electroweak theory is a theory that unites the theory of leptons and photons, which encompasses most elementary quantum mechanics and lays the foundations for chemistry and materials science, with the theory of the weak nuclear force, which mediates the radioactive decay of various atomic nuclei and the interaction of neutrinos with atomic matter. This unification relies crucially on the existence of the Higgs field, and the phenomenon of spontaneously broken symmetry that is induced by the dynamics of the Higgs field under the influence of the Higgs potential.

2.1.2.1 Lagrangian formalism

The electroweak sector of the standard model postulates conservation of two quantum numbers known as weak isospin (T_3) and weak hypercharge (Y), which give rise to the gauge symmetries $SU(2)$ and $U(1)$ respectively. The $U(1)$ symmetry corresponding to weak hypercharge is *not* the same as the $U(1)$ symmetry corresponding to ordinary electromagnetic charge conservation in the earlier theory of Quantum Electrodynamics (QED); in fact $Y = 2(Q - T_3)$ where Q is the ordinary electromagnetic charge. However, the electroweak symmetry breaking does give rise to the $U(1)$ symmetry of QED. The $SU(2)$ group, which corresponds to the conservation of weak isospin, was studied in the context of spin by Pauli and others in the early history of Quantum Mechanics, but the similarities to spin are purely mathematical; isospin is not related to quantum-mechanical spin. Despite the model’s conceptual differences with QED, the gauge theory based on $SU(2) \otimes U(1)$ was exceptionally successful at describing both the electromagnetic force and the weak nuclear force.

The electroweak model embodies the experimental fact that key particles in electroweak processes, neutrinos, all have left-handed chirality in that it models the left handed fermions of the standard model as $SU(2)$ doublets, whereas the right-handed fermions are singlets under $SU(2)$. Before symmetry breaking, all fermions are modeled as being exactly massless. With unbroken symmetry, the electroweak sector’s Lagrangian is composed as the sum of four parts:

$$\mathcal{L} = \mathcal{L}_g + \mathcal{L}_f + \mathcal{L}_y + \mathcal{L}_h \tag{2.6}$$

The first term $\mathcal{L}_g$ describes the kinetic energies of the spin 1 boson gauge fields of weak isospin and weak hypercharge, whose field strength tensors are denoted by $W^{a\mu\nu}$ and $B_{\mu\nu}$:

$$\begin{aligned} W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{abc}W_\mu^b W_\nu^c \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \quad (2.7)$$

where ϵ^{abc} is the Levi-Civita symbol.

The index a (as well as b and c) runs from 1 to 3, and in this picture with unbroken electroweak symmetry, this can be interpreted as meaning that the W field gives rise to three massless bosons, and likewise the B field gives rise to one massless boson:

$$\mathcal{L}_g = -\frac{1}{4} (W_a^{\mu\nu} W_{\mu\nu}^a + B^{\mu\nu} B_{\mu\nu}) \quad (2.8)$$

The second term $\mathcal{L}_f$ corresponds to the kinetic energies of the fermions of the standard model:

$$\mathcal{L}_f = \bar{Q}_i i\gamma^\mu D_\mu Q_i + \bar{u}_i i\gamma^\mu D_\mu u_i + \bar{d}_i i\gamma^\mu D_\mu d_i + \bar{L}_i i\gamma^\mu D_\mu L_i + \bar{e}_i i\gamma^\mu D_\mu e_i \quad (2.9)$$

where the doublet fermion fields Q_i and L_i refer to the doublet left-handed quark and lepton fields, while u_i , d_i , and e_i are the right-handed singlet quark fields and the charged lepton fields, where the index i runs from 1 to 3 indicating the generation of quarks or leptons. The gauge covariant derivative of electroweak gauge theory is given by:

$$D_\mu = \partial_\mu + ig \frac{\sigma^a}{2} W_\mu^a + ig' \frac{Y}{2} B_\mu \quad (2.10)$$

where σ^a are the Pauli matrices, while g and g' are the $SU(2)$ and $U(1)$ gauge coupling constants, respectively, which are related by the electroweak mixing angle θ_W :

$$g' \cos(\theta_W) = g \sin(\theta_W) \quad (2.11)$$

The weak mixing angle is a parameter of the SM and must be measured experimentally.

The final two terms of the electroweak sector's Lagrangian are $\mathcal{L}_h$, which describes the complex scalar doublet Higgs field's kinetic and potential energies, and a possible additional term $\mathcal{L}_y$, which describes the Yukawa coupling of the spin 0 Higgs boson h to the quarks and leptons.

$$\mathcal{L}_h = |D_\mu h|^2 + V(h) \quad (2.12)$$

In this equation, $V(h)$ denotes the Higgs potential, the form of which will be discussed below.

The term that introduces a Yukawa coupling between the Higgs field and the SM fermions is:

$$\mathcal{L}_y = -y_{u,ij} h \bar{Q}_i u_j - y_{d,ij} h \bar{Q}_i d_j - y_{e,ij} h \bar{L}_i e_j + h.c. \quad (2.13)$$

where y_u , y_d , and y_e refer to the Yukawa couplings of the Higgs boson to the quarks, which are free parameters in the SM.

2.1.2.2 The Higgs mechanism

The Lagrangian presented above can be demonstrated to describe an entirely massless theory; in fact, without the incorporation of the Higgs potential, the above Lagrangian is required to have no massive particles by the gauge symmetry. However, if the Higgs potential takes the form of the “Mexican hat” potential:

$$V(h) = \mu^2 h^\dagger h + \lambda (h^\dagger h)^2 \quad (2.14)$$

then, writing the Higgs doublet in the form:

$$h = \begin{pmatrix} h^+ \\ h^0 \end{pmatrix} \quad (2.15)$$

and assuming a negative μ^2 , the Higgs field is most stable with a vacuum expectation value:

$$\begin{pmatrix} h^+ \\ h^0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{\mu}{\sqrt{2\lambda}} \end{pmatrix} \quad (2.16)$$

The value $\frac{\mu}{\sqrt{2\lambda}}$ of the Higgs field’s vacuum expectation value, which we denote as v , can then be used to redefine the nonzero component of the Higgs field in terms of its excitations around its vacuum state:

$$h^0 = \frac{\mu}{\sqrt{2\lambda}} + H + i\eta \quad (2.17)$$

where H and η are real-valued fields with weak hypercharge $Y = 1$, which can be identified with the massive Higgs boson and a massless Goldstone boson. A gauge transformation of the form $h^0 \rightarrow e^{i\theta(x_\mu)} h^0$ is allowed by the local $U(1)$ gauge symmetry of this field, which allows the imaginary component $i\eta$ of h^0 to be eliminated, and therefore the kinetic term of Equation 2.12 may be expanded around the Higgs field’s vacuum expectation value:

$$\begin{aligned} |D_\mu h|^2 &= \left(\partial_\mu + ig \frac{\sigma^a}{2} W_\mu^a + i \frac{g'}{2} B_\mu \right) \begin{pmatrix} 0 \\ H + \frac{\mu}{\sqrt{2\lambda}} \end{pmatrix} \\ &= \frac{1}{2} (\partial_\mu H)^2 + \frac{1}{8} g^2 \left(H + \frac{\mu}{\sqrt{\lambda}} \right)^2 |W_\mu^1 - iW_\mu^2|^2 + \frac{1}{8} \left(H + \frac{\mu}{\sqrt{\lambda}} \right)^2 |gW_\mu^3 - g'B_\mu|^2 \end{aligned} \quad (2.18)$$

In this expression, we can see that the W_μ and B_μ fields have mixed together, so we can write these mixtures of fields in a more natural manner:

$$\begin{aligned} W^\pm &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp iW_\mu^2) \\ Z_\mu &= \cos(\theta_w) W_\mu^3 - \sin(\theta_w) B_\mu \\ A_\mu &= \sin(\theta_w) W_\mu^3 + \cos(\theta_w) B_\mu \end{aligned} \quad (2.19)$$

which trivially implies:

$$gW_\mu^3 - g'B_\mu = \frac{g}{\cos(\theta_w)} Z_\mu \quad (2.20)$$

Keeping in mind that μ^2 was assumed to be negative, we have found that the Higgs coupling to the originally massless vector bosons via the gauge covariant derivative has led to terms in the Lagrangian of the form $-\frac{m^2}{2} |\mathcal{F}_\mu|^2$ where $\mathcal{F}_\mu$ generically denotes the W and Z fields – a mass term – while leaving no such term for the A field. A popular name for this is “dynamical mass generation.” The mass term we have found for the W field is $m_W = \frac{g\mu}{2\sqrt{\lambda}}$, whereas for the Z we have found $m_Z = \frac{g\mu}{2\sqrt{\lambda}\cos(\theta_w)}$.

Similarly, we can investigate whether the field H gains a mass via broken symmetry. Expanding the Higgs potential to second order:

$$V(H) \sim V_0 + \frac{\mu^2}{2} (2vH + H^2) + \frac{\lambda}{4} (4v^3H + 6v^2H^2) = V_0 + \lambda v^2 H^2 \quad (2.21)$$

implying a Higgs mass of $\sqrt{2\lambda}v = \sqrt{2}|\mu|$

If we consider the Yukawa coupling term between the Higgs and fermions $\mathcal{L}_y$, mass terms arise as well, for example for the electron:

$$\mathcal{L}_y = y_e h \bar{L}_i e_j + h.c. = \frac{y_{d,ij}}{\sqrt{2}} (\bar{e}\nu_e) \begin{pmatrix} 0 \\ \frac{\mu}{\sqrt{2\lambda}} + H \end{pmatrix} = y_e \left(\frac{\mu^2}{\sqrt{2\lambda}} + H \right) \bar{e}_L e_R \quad (2.22)$$

giving the electron a mass term corresponding to $m_e = y_e \frac{\mu^2}{\sqrt{2\lambda}}$.

A diagram that gives an overview of the particle content of the SM and various particle properties is shown in Figure 2.1.

2.2 Unsolved mysteries

The SM has been a remarkably successful and fruitful theory of particle physics throughout its history. However, there are many reasons, some quite subtle and others quite glaring, why it is clear that the SM is clearly an incomplete description of nature. Indications that the SM cannot be the end of the story come from many different avenues, a few of which come from particle physics experiments and observational astroparticle physics that have found anomalous results, others – in particular two very significant puzzles in physics – that come from large-scale astronomical observations, yet others that come from the physical interpretation of the meaning of renormalization, and finally many which are based on hints in the mathematical structure of the theory that point to the idea of a more fundamental theory. This section is an overview of the mysteries of particle physics that have yet to be understood.

2.2.1 Neutrinos

In the version of the SM presented above, neutrinos notably do not gain any mass via the Higgs mechanism, and are modeled as massless particles. In addition, neutrinos are modeled as having only one chirality, which implies that they must be massless, since chirality is in general not a Lorentz-invariant property for massive particles. The fact that only left-handed neutrinos are observed in

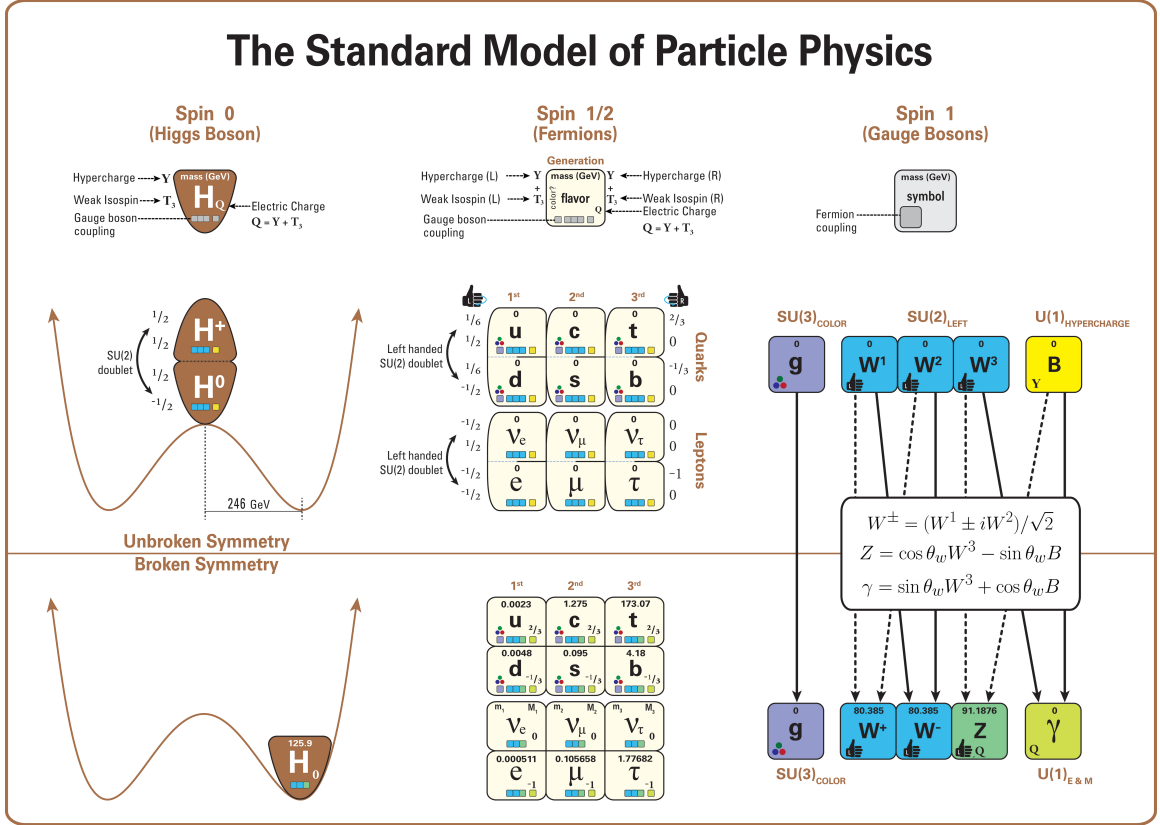


Figure 2.1: A diagram of the particle content of the Standard Model. The top portion of the diagram shows the SM with unbroken electroweak symmetry, whereas the bottom portion shows the SM with the Higgs at its vacuum expectation value, indicated by the position of the Higgs boson in its “Mexican hat” potential. Taken from [5].

nature is what motivates this construction, but one might wonder on purely theoretical grounds why this is so – in a mathematical sense, one could easily include massless neutrinos with right-handed chirality in the theory.

One implication of special relativity is that massless particles cannot spontaneously decay or otherwise oscillate into different mass eigenstates of the SM. However, starting in the observations of solar neutrinos [22] and continuing with further astroparticle experiments [23], as well as reactor-based [24] and accelerator-based experiments [25], strong evidence confirmed that neutrinos oscillate into neutrinos of different flavors.

Various models have been suggested as modifications of the SM in order to reconcile the theory with the observation of neutrino oscillations and the observed smallness of neutrino masses. Currently, the most prevalent theory of neutrino oscillations postulates an extension to the SM that predicts very heavy right-handed neutrinos, resolving the apparent SM disagreement via the “see-saw mechanism.” This idea involves introducing a term into the SM Lagrangian which couples the

normal left-handed to very massive right-handed neutrinos that implies that the mass eigenstates of left-handed neutrinos have masses proportional to the inverse of the very heavy states, which nicely predicts both neutrinos oscillations and explains their extremely small but nonzero mass. [26]

2.2.2 Dark matter

Astronomical observations of the angular velocity of stars around the centers of galaxies have led to the conclusion that the angular velocities of stars near the edges of galaxies greatly exceed that which would be predicted by ordinary gravitational theory, if one models the galaxies as consisting only of the visible stars and other ordinary matter that makes up galaxies. In addition, observations of gravitationally-lensed light from distant galaxies around closer astronomical bodies indicate that the mass of galaxies is much larger and more diffuse than what can be seen in the form of ordinary matter, and that the distribution of the mass is more spherical than the typical accretion-disk shapes of most galaxies. These compelling observations have led astrophysicists to suggest that the mass of ordinary matter in galaxies constitutes only about 20% of the total mass of galaxies.

Many models suggest that the remainder of the mass of galaxies is explained by the existence of particles that do not interact electromagnetically, known as “dark matter.” One well-studied class of models of this sort postulates that the particles that constitute dark matter interact only via the weak nuclear force, which would explain why they cannot be seen using conventional astronomical observatories, and why dark matter does not accrete like ordinary matter. This kind of hypothetical particle is called a “WIMP” or weakly interacting massive particle. Many experiments have been performed in attempts to detect WIMPs, including direct detection experiments that search for rare interactions of astronomical WIMPs with SM particles [27], searches for WIMP production in high-energy colliders [28], and experiments using indirect detection techniques [29].

2.2.3 Gravity

One very obvious problem with the standard model is that it conspicuously omits one very basic piece of human-scale phenomena – a piece with extreme historical significance in the development of physics and the modern understanding of the universe: the force of gravity. The monumental achievements of Sir Isaac Newton, which represented the beginning of what is today considered physics, had a deep focus on gravitation both on the surface of the Earth and in the solar system. The development of General Relativity by Albert Einstein cast a new light on the understanding of gravity as intimately connected with the structure of spacetime, which lead to the modern methods in physics for understanding the cosmological structure and origins of the universe. That the SM does not include any description of the force of gravity is as clear an indication of the incompleteness of the SM as any.

One reason for difficulties of adding a terms like the Einstein-Hilbert action into the SM Lagrangian has to do with issues of renormalizability. The Einstein-Hilbert action would imply including a term of the form:

$$\frac{R}{2\kappa}\sqrt{-g} \quad (2.23)$$

where R is the dimensionless Ricci curvature scalar, g is the determinant of the metric tensor, and κ is Einstein's constant, which has the units of meters per kilogram. In general, quantum field theories that include a dimensionful coupling constant in the Lagrangian cannot be renormalized, which has a historical analogy: Fermi's theory of beta decay, which was based on a four-fermion contact interaction directly between neutrons, protons, electrons, and neutrinos. Because Fermi's theory required a term in the Lagrangian that had a coupling constant with the units energy⁻², it was not realizable as a quantum field theory – no perturbative calculations could be regularized to give non-infinite results – and the phenomenon of beta decay was only solved in the electroweak theory, where the weak bosons mediate beta decay. The presence of intermediate bosons that mediate the weak nuclear force was not known at Fermi's time, due to the fact that they were beyond experimentally accessible energies, but his theory is in retrospect a valid low-energy approximation to electroweak theory, the nature of which could only be probed once energies on the order of the electroweak symmetry breaking scale became accessible. Similarly, one might speculate that if a workable quantum field theory of gravity is to arise, the energies required to probe the quantum nature of gravity would be on the order of where the symmetry between gravitational and other forces is broken, and it is believed that this scale is on the order of the Planck energy, which is approximately one quadrillion times higher than the energy of the LHC. Nonetheless, speculative mathematical frameworks, notably the study of substituting the concept of particles with the concept one-dimensional objects called strings, that attempt to unite gravity with the other known fundamental forces have been studied intensely for half a century, but have not yet become experimentally relevant.

2.2.4 The heirarchy problem

Another problem related to renormalization and the energy scales at which the standard model manifests itself as an effective low-energy theory has to do with the mass of the Higgs boson. To one-loop level in the SM, the Higgs boson receives corrections to its mass from top quarks and other SM particles that go as:

$$\delta m_H^2 = \frac{\Lambda^2}{32\pi^2} \left[6\lambda + \frac{1}{4} (9g^2 + 3g'^2) - y_t^2 \right] \quad (2.24)$$

where Λ is the renormalization scale, or cutoff scale, of the theory, which can be understood as the energy below which the SM becomes a valid low-energy effective field theory of some theory. Above energies on the order of Λ , new physical phenomena occur. In other words, this equation might suggest that the Higgs mass should be on the order of some higher mass scale represented by Λ , such as, perhaps, the Planck scale. Equivalently, one might look at this situation and ask “Why is the energy scale of electroweak theory so much smaller than the Planck energy?” or, in simpler terms, “Why is the force of gravity so much weaker than the electroweak force?” In light of this,

it is natural to expect some physics Beyond the Standard Model, which swoops in at energies near Λ and cancels out the large corrections to the Higgs mass. One might propose, however, that the “bare” coupling constants in this equation, g and g' , have values that offset the terms that diverge with Λ (keeping in mind that bare coupling constants cannot be directly measured). However, this proposal is usually regarded as unlikely and unappealing, because the bare coupling constants would have to exactly offset these large corrections, which is referred to as “fine tuning.”

2.3 Beyond the standard model

In light of the outstanding problems in the SM discussed in the last section, and the suggestive hints that seem to point toward some new physics waiting to be discovered, many theories of physics that go Beyond the Standard Model (BSM) have been proposed.

2.3.1 Supersymmetry

One very well-studied idea to address the problem of the stability of the Higgs mass introduces a fundamental symmetry between fermions and bosons known as “supersymmetry” (SUSY). This hypothetical symmetry introduces one partner particle for every standard model particle with a spin that differs by $\frac{1}{2}$, as well as four additional spin-0 bosons along with their accompanying fermionic partners. Supersymmetry resolves the problem of the stability of the Higgs mass by having partner contributions at all levels of perturbation theory that exactly cancel all the terms that diverge with Λ . Supersymmetry also has the tantalizing feature that, unlike in the standard model, the renormalization flows of the electromagnetic, weak, and strong coupling constants all converge to the same value at an energy known as the “grand unified theory” energy.

Unfortunately, extensive searches for supersymmetry have been undertaken by LEP, the Tevatron, and the LHC, with no evidence found so far. [19] The searches indicate that the phase space of SUSY models originally believed to alleviate the Higgs mass instability have been ruled out, and therefore the SUSY models that are still viable in light of these stringent limits must themselves require a considerable amount of fine tuning; SUSY models with this sort of fine tuning are referred to as “split supersymmetry.”

2.3.2 Extra dimensional models

In an attempt to address the hierarchy problem, several theories have arisen that build on the ideas pioneered and developed by Kaluza and Klein in the 1920's. Kaluza and Klein envisioned that the macroscopic world as we observe, which contains the three spatial dimensions of ordinary physics (and the SM), also contains an additional, unobserved spatial dimension. Their work, which was aimed at creating a unified theory of gravitation and electromagnetism, postulated that this additional dimension has gone unobserved because it is “compactified,” i.e. that the dimension has

a very tiny finite spatial extent in the shape of a circle. The Kaluza-Klein models never enjoyed success, however.

These ideas have been revived in recent times in at least two notable classes of models: the Arkani-Hamed–Dimopoulos–Dvali models (ADD) models, and the Randall-Sundrum (RS) models. Both classes of models postulate that there exist extra dimensions, but that these extra dimensions are not small but are instead macroscopically large. These macroscopic extra dimensions, however, are unobserved because the SM particles that we study in particle physics are confined to a four-dimensional surface within the larger dimensional space, and that only the quanta of gravitation (gravitons) and other exotic particles are able to travel in the bulk space containing the surface (“brane”).

ADD models propose that these large extra dimensions are described by a roughly flat space-time metric, and have radii no bigger than 1 mm. There may be some arbitrarily large number of such extra dimensions. These models help to alleviate or even resolve the hierarchy problem because the gravitational coupling we observe in our 4-dimensional space is not equal to the fundamental gravitational constant – the fundamental gravitational force could be much stronger in the higher-dimensional space, decreasing the apparently huge gap in the scales between gravity and the electroweak interaction.

RS models generically propose only one or two extra dimensions to space, and within this extra-dimensional space, two four-dimensional surfaces exist: one confines the SM particles of the universe, and the other confines gravitons. The gravitational force can again be raised to be dramatically stronger on the gravity brane, and its apparent weakness on our brane is caused by a quantum tunneling effect through an extremely warped anti-de Sitter bulk space separating the two branes, which causes the apparent weakness of gravity on the brane containing SM particles.

As in Kaluza-Klein models, ADD models predict the existence of Z' bosons, which represent the excited modes of a neutral gauge symmetry in the compactified dimensions.

2.4 Searching for new physics

Did I expect that we'd find the fourth quark? No. But was I looking for the structure of the hadrons? Yes.

Burton Richter, on the discovery of the J/ψ meson. [30]

When Samuel Ting and Burton Richter discovered the J/ψ boson in the autumn of 1974, neither Ting's group at Brookhaven National Laboratory nor Richter's group at the Stanford Linear Accelerator Center were specifically motivated to search explicitly for the charm quark. To the contrary, both groups were probing the energy frontier in an effort to explore and characterize the phenomena of previously unexplored energies. The discovery quickly led to an uprooting of the “old” standard model of physics, and the subsequent “November Revolution” in particle physics set the stage for the development of the current standard model.

Presently, the Large Hadron Collider (LHC) at CERN represents the energy frontier of particle physics, and very important and exciting results have come already. Within two years after the first collisions delivered by the LHC, the discovery of a new boson with a mass of 125 GeV [31] [32], now known to be the Higgs boson, was announced on July 4th, 2012 by the ATLAS and CMS collaborations. The Higgs discovery, unlike the discovery of the J/ψ , was widely expected by most of the particle physics community prior to the announcement, and many of the subsequent studies on the newly discovered boson confirmed its properties to be consistent with the SM Higgs boson. The signature of the Higgs seen by CMS and ATLAS that amounted to the first incontrovertible evidence for the Higgs boson was its decay into two photons, shown in Figure 2.2.

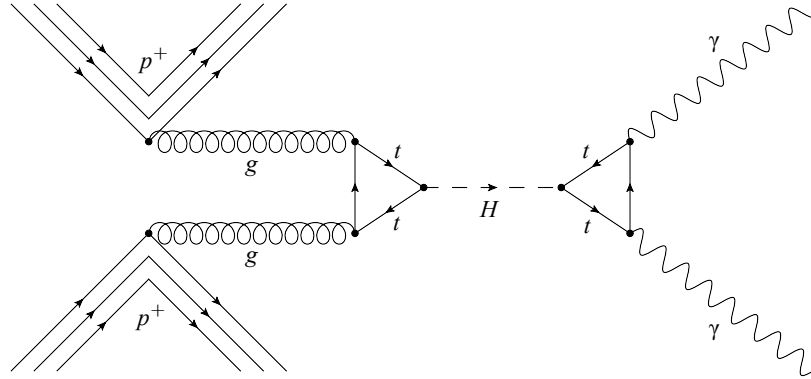


Figure 2.2: Gluon fusion production of a Higgs boson. The protons emit gluons, which combine via a top loop to a Higgs boson, and the Higgs boson then decays via a top loop to two photons. This channel was the first channel studied by CMS and ATLAS that provided evidence for the Higgs.

This channel is a very sensitive channel for detecting the production of the Higgs boson for

a few main reasons. First, even though the branching fraction for the Higgs boson to decay to two photons is quite small (about .0023), this channel has very low backgrounds, thanks to high-performance tracking and high-granularity electromagnetic calorimetry that allows the rejection of most backgrounds. Second, a two-body decay is a very “clean” signature of the presence of a new particle, in that a pronounced peak in the invariant mass of a two body system is a very unambiguous indication of the presence of a resonance, and gives a straightforward method to estimate the backgrounds from the sidebands or by an analytical fit of the spectrum. The use of sidebands or fits to the spectrum of a two-body invariant mass spectrum is very advantageous in terms of providing discovery potential, because it permits a straightforward interpretation of the data without the need for full models of the signal and background processes.

Considering these compelling experimental strengths and track record of discovery potential, and in the spirit of probing and characterizing all the phenomena at the energy frontier, the CMS collaboration has embarked on a program to look for signs of new physics in many (and hopefully, eventually all) diboson channels, especially since the new arrival of the Higgs boson into the family of SM bosons. There are many models that predict new resonances that decay to a pair of SM bosons, ranging from two Higgs doublet models [33], to models with low-scale gravity, e.g., the Randall–Sundrum model [34,35], the ADD model, and models with strong dynamics, such as technicolor [36, 37]. One important fact about searches that focus on new, heavy resonances decaying into two SM bosons is that for sufficiently massive resonances, the SM bosons are created with a significant boost, and for boosted $W^\pm$, Z , and Higgs bosons that decay hadronically, these particles will show up in the detector as large radius jets. Furthermore, because of the specifics of the W , Z , and H decays, provided the individual particles of the jets can be individually reconstructed, a telltale substructure for each kind of particle will develop. Thanks to the excellent performance of the CMS tracking and calorimetry, it is possible to resolve and reconstruct all the particles constituting these large-radius jets, and “tagging” techniques for distinguishing jets from each of the bosons have been developed, which allows the diboson searches to have powerful background rejection that greatly increases the sensitivity of these channels to hints of new physics.

One very exciting diboson channel is the topic of this dissertation: the Higgs + photon channel. Photons are very well studied theoretically, and very well measured objects in the CMS detector, while the Higgs boson is our newest and least well-characterized of the SM bosons, and many recent results from analyses of Higgs-related physics coming from the LHC detectors represent the first of their kind. A summary of the analyses that CMS and ATLAS have published regarding diboson signatures of new physics, and models that predict resonances in those channels, is given in Table 2.1. The

A recent paper by Dobrescu *et. al.* has elaborated on a Z' model, i.e. a model that postulates the existence of a new neutral spin-1 color singlet boson with $U(1)$ gauge symmetry, that decays to $H\gamma$, and proposes that the $H\gamma$ channel has the potential to be the discovery for this kind of Z' . A Z' boson cannot decay to $H\gamma$ at tree level, but that decay may proceed through higher dimensional operators which arise from loops, such as those shown in Figure 2.3. These processes occur due to

Table 2.1: A summary of diboson resonances physics. In the top line of each entry, CMS analyses performed in that channel are listed. Underneath, examples of models which may give rise to such resonances are listed. To denote the models, H denotes additional neutral scalar or pseudoscalar Higgs bosons, ρ denotes neutral or charged techni-rhos, and $\mathcal{Q}$ denotes quirks. [1]

	γ	W	Z	H
γ	[38, 39] $H, G_{KK}, \mathcal{Q}$	underway W_{KK}	[40–42] $H, \mathcal{Q}$	this work Z', Z_{KK}
W		[43–45] H, G_{KK}, ρ	[43, 44, 46] $W', \mathcal{Q}$	[45, 47, 48] $H^\pm, \mathcal{Q}, \rho$
Z			[43, 45, 46, 49] H, G_{KK}, ρ	[45, 47, 50] A, ρ
H				[51, 52] H, G_{KK}

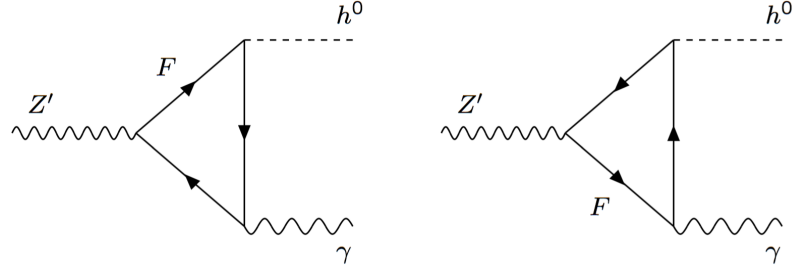


Figure 2.3: A diagram for the signal model considered in this analysis: $Z' \rightarrow H\gamma$. The leading order diagram corresponds to the one-loop contributions of a fermion F to the $Z' \rightarrow H\gamma$ amplitude. Here, F can be either a top quark or an anomalon. From [6]. Though this particular model is the focus in this analysis note, it is expected that generic narrow $H\gamma$ resonances will have signatures very similar to this particular model.

an operator of the form:

$$\frac{c_\gamma e v}{(4\pi)^2 m_0^2} h Z'_{\mu\nu} F^{\mu\nu} \quad (2.25)$$

where $Z'_{\mu\nu}$ is defined similarly to the gauge fields in the SM defined earlier, m_0 is the mass of some particle running in the loop, and $(4\pi)^2$ is related to the loop itegral. Here, c_γ is a model-dependent dimensionless parameter, and is typically of order one or smaller.

Due to this process proceeding to loops, the width of this resonance will very narrow, many orders of magnitude smaller than the Z' mass, and it is possible to show that any higher dimension operators contributing to on-shell decay of $Z' \rightarrow H\gamma$ can be related to that operator. The branching ratio of the Z' to leptons instead of $H\gamma$ is model-dependent, and it can be the case that $h\gamma$ has a higher branching fraction than the lepton channel. In addition, if some new fermion runs in the loop, such as an anomalon (a vector-like fermion), then the branching ratios of $H\gamma$ can be enhanced.

Chapter 3

Particle Physics at the Large Hadron Collider

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest and most powerful particle accelerator in the world as of 2018. A roughly circular hadron collider with a circumference of 26.7 kilometers, the LHC’s main purpose is to collide protons into protons (pp collisions), designed to achieve a center-of-mass energy of 14 teraelectronvolts (TeV). The LHC is also used to collide of various other kinds of particles, such as colliding lead nuclei together (Pb-Pb collisions), or to collide protons with lead nuclei (p-Pb collisions). A summary of the operating periods of the LHC with various types of collisions is given in Table 3.1.

The LHC is located in an underground tunnel (between 45 m and 175 m below the surface) originally constructed for the Large Electron Positron collider, in an area straddling the border between France and Switzerland outside the city of Geneva. The LHC accelerates and circulates two separate, counterrotating particle beams, and collides the particles in the beams by focusing and crossing the “bunches” of particles comprising the beams. [53] [54]

There are four positions around the LHC tunnel at which the beams can be crossed, and at each of these four “interaction points,” various particle detectors have been constructed for probing the physics manifested in these high-energy particle collisions; these include the CMS (Compact Muon Solenoid) detector and the ATLAS (A Toroidal Large Hadron Collider Apparatus), which are both general-purpose hermitic particle detectors, as well as the ALICE (A Large Ion Collider Experiment) detector, which is optimized to study heavy ion (Pb-Pb) collisions, and the LHCb (Large Hadron Collider beauty), which is designed for precision tests of b-quark physics (“b physics”). The LHC has enabled these experiments and others to yield thousands of results, including many precision measurements of the standard model (SM) of particle physics, most of the most stringent limits on models of particle physics that go beyond the SM (BSM), and several discoveries, including the

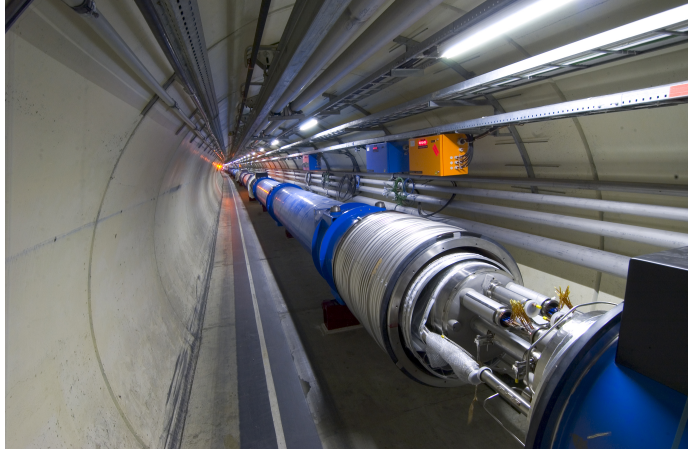


Figure 3.1: The LHC inside the LEP/LHC tunnel, during assembly. In this photograph, two segments of the LHC machine structure are separated (they have not yet been connected), showing the interconnection between two segments of the beam pipe [7].

discovery of the SM Higgs Boson [31] [32] [55], tetraquark candidates [56] [57], and pentaquarks [58]. The LHC is planned to continue operating at least until 2037, with increases in energy and instantaneous luminosity planned, and by that time a total integrated luminosity of 4000 fb^{-1} is expected [59].

3.1.1 LHC accelerator complex

In order to “fill” or supply particles into the LHC’s two beams, a series of different accelerators are used to create beams with the appropriate injection energy, beam profile, and bunch characteristics for optimal performance. Depending on whether a proton or heavy ion beam is being injected into the LHC, different injection schemes are used.

In the case of injecting protons into the LHC, the protons begin as the nuclei of H_2 atoms in a compressed gas cylinder attached to the RFQ2 (Radio Frequency Quadrupole 2) assembly. Within the RFQ2 assembly, a device known as a duoplasmatron is employed as a proton source; the duoplasmatron applies a high electric field to the gas, ionizing the hydrogen atoms into free protons and electrons, and propelling them toward an aperture. Protons leaving the duoplasmatron’s aperture are first accelerated and then focused by four radio frequency (RF) quadrupole magnets in the RFQ2 and bunched (i.e. the protons in the beam are separated into separated groups of protons) in two bunching cavities [60] [61]. The Linac2, a drift tube linear accelerator of standard Alvarez design (consisting of a series of cylindrical conductors with RF voltages applied, arranged such that the protons are attracted into the cylinder by a negative applied voltage, and pushed out of the cylinder by a positive applied voltage), accelerates these bunches to an energy of 50 MeV [62] and are injected into a low energy accelerator called the Proton Synchrotron Booster (PSB), which has four rings with a circumference of 157 m each [63]. Modulation of the field in the PSB’s main

bending magnets as well as adjustments in the position of the beam from the Linac 2 entering the PSB are used to optimize the beam characteristics, in particular, minimizing the beam emittance (i.e. a measure of the size of the beam and the transverse momentum of protons in the beam). After filling all the RF buckets of the PSB with 10^{13} protons per ring, the beams are accelerated to 1.4 GeV and injected into the Proton Synchrotron (PS) accelerator [64].

In the case of injecting heavy ions such as lead nuclei into the LHC, lead atoms are placed in a crucible in the ECR4 (Electron Cyclotron Resonance Ion Source), and oxygen gas is injected into the chamber containing the crucible. Microwave radiation is used to excite the oxygen into a plasma phase, which is contained using a hexapole RF electromagnet, which evaporates and ionizes a portion of the lead atoms. Similar to the preparation of proton beams, these lead ions are then accelerated, focused, and bunched using an RFQ assembly before entering the Linac3, which then accelerates the lead ions to an energy of 4.2 MeV per nucleon. Lead ion bunches are then sent into the LEIR (Low Energy Ion Ring) accelerator, which accelerates and compresses the bunches, which are then ready to be injected into the Proton Synchrotron [65].

For both protons and heavy ions, the bunches are first passed into the Proton Synchrotron (PS) accelerator and then into the Super Proton Synchrotron (SPS) accelerator before injection into the LHC. The PS is a single-ring accelerator with a circumference of 628 m, which accelerates the incoming proton (heavy ion) beam to an energy of 25 GeV (72 MeV). Upon reaching this “flat top” energy (the maximum energy of the PS), the beams are manipulated using various RF harmonics (a process called “RF gymnastics”) such that the bunches, which are initially spaced by 300 ns, are split into the nominal bunch spacing of 25 ns of the LHC [66] [67].

The beams, now with their nominal bunch structure, are then sent into the Super Proton Synchrotron (SPS), a circular collider with a circumference of 6.9 km. The SPS accelerates the bunches to an energy of 450 GeV, and serves to fill the LHC in a pattern of “bunch trains”, i.e. the RF buckets of the LHC are filled in specific patterns such that there are many RF buckets in a row occupied by bunches, followed by a series of empty RF buckets (called “gaps”), which are necessary to allow for the rise and fall times of the “kicker” magnets, i.e. the magnets which direct the beam during; the rise times of the injection kickers is as long as $1\ \mu\text{s}$, and the rise time of the beam dump kicker is $3\ \mu\text{s}$ [68] [69].

A diagram of the CERN accelerator complex, including the accelerators that make up the LHC injector chain, is shown in Figure 3.2.

3.1.2 Machine specifications

The LHC tunnel is roughly circular, but not perfectly; in addition to eight arc sections, the LHC has eight straight sections each with a length of approximately 528 m, which are used as locations for situating (called “insertions”) for experiments, like the main particle detectors, as the location for injection from the SPS transfer tunnels, or as the location of the beam dump. Each of these points, located at the center of each octant of the LHC, is called a “Point”; as an example, the two high-luminosity proton-proton experiments, ATLAS and CMS, are located at diametrically opposite

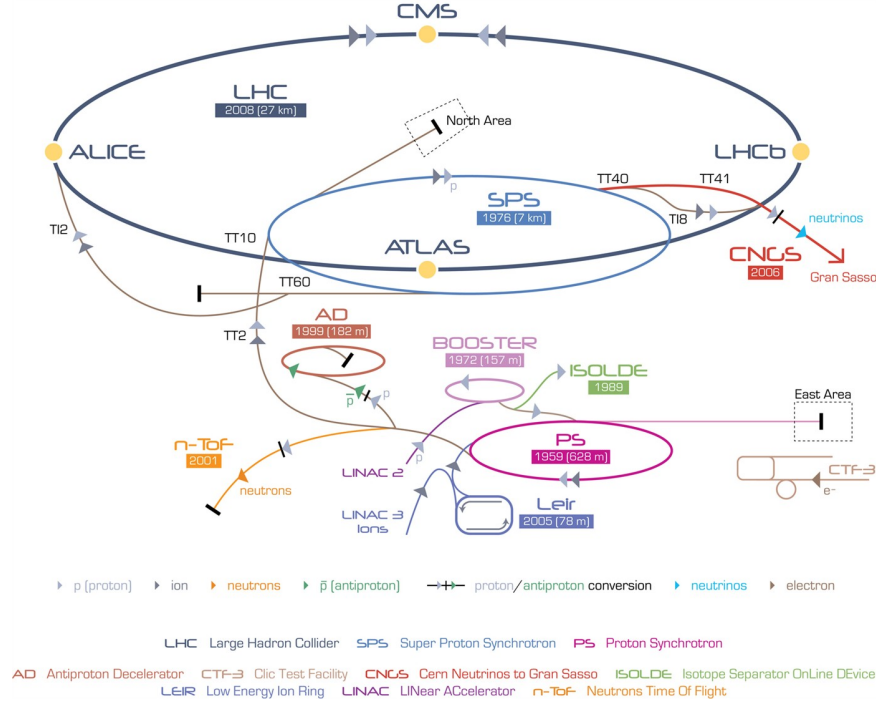


Figure 3.2: A schematic drawing of the CERN accelerator complex, as well as several of the collider and fixed-target experiments [8].

positions around the LHC: Point 1, where ATLAS is located, is in Meyrin, Switzerland, on the main CERN campus, while CMS is located exactly at Point 5, which is on the opposite side of the LHC in Cessy, France. [53] [54]

The LHC contains two counter-rotating beams, which are crossed at four points around the detector, at Points 1 and 5 for the ATLAS and CMS detectors, respectively, as well as at points 2 and 8, where the ALICE and LHCb detectors are located, respectively. The two beams are kept in their circular trajectory using a so-called “two-in-one” or “twin-bore” design superconducting magnets, which serve to bend the beams into the roughly circular path through the LHC. These superconducting magnets are based on NbTi Rutherford cables, but wound in such a way to accommodate both beam channels in a common cold mass and cryostat, with the magnetic flux circulating in opposite directions through the two beam channels. Comprising the bulk of the structure of the LHC accelerator, 1,232 of these NbTi superconducting dipole magnets are cooled to 1.9 K using superfluid helium, which enables that attainment of magnetic fields above 8 T inside the beam pipes, which each have a diameter of 56 mm, and are separated by 194 mm. [53] The dipole magnets and the requisite cryogenics, collectively termed the “Dipole Cryostat”, are housed in an evacuated steel tube of diameter 914 mm.

In each octant of the LHC, a set of corrector magnets are used to focus the beams and maintain

their bunch structure. Housed in so-called “Short Straight Sections”, these corrector magnets include tuning quadrupole magnets, skew quadrupoles, chromaticity sextupoles, skew sextupoles, and octupoles. [53]

In order to accelerating the beam from the injection energy (450 GeV for protons) to the nominal beam energy for physics studies (6.5 TeV, currently, for proton-proton physics), the LHC employs a system of eight single-cell niobium-on-copper superconducting RF cavities. The RF system, which is housed in the underground caverns at Point 4, is cryogenically kept at 4.5 K, and operates at 400.8 MHz, delivering 32 kW of power per beam to accelerate the beams (during “ramping” or the period of increasing the beams’ energy), and 3.7 kW per beam while maintaining stable beams (i.e. once the maximum energy has been attained.) [69] [70]

3.1.3 LHC collider physics

In analogy with classical scattering theory or elementary quantum mechanical scattering theory, the quantity used in particle physics (and generally in quantum field theory) to characterize the probability of a physical process occurring is called the *cross section*, denoted σ_{proc} , which has the units of area. Commonly, the unit of area used for expressing cross sections is called the *barn*, which is equal to $1 \times 10^{-28} \text{ m}^2$, approximately equal to the cross-sectional area of a uranium nucleus.

In general, the cross section of particle physics processes depends on the center-of-mass energy $\sqrt{s}$ of the colliding particles, but in collisions performed at the LHC, the energy of the protons in the bunches is very tightly controlled and is kept constant over the duration of operation (called a “fill”); therefore the cross section σ_{proc} is taken to be a constant. Therefore, in particle collider experiments, the basic relation that is used to calculate the expected number of events N_{exp} in which a particular physical process with cross section σ_{proc} occurs is:

$$N_{\text{exp}} = \sigma_{\text{proc}} \cdot \int \mathcal{L}(t) dt \quad (3.1)$$

where $\mathcal{L}(t)$ is the instantaneous luminosity as a function of time t , which has the units of cross sectional area per unit time. The instantaneous luminosity is a quantity that depends on the experimental conditions, and represents the rate at which particles are exposed to interactions with one another [19]. In the LHC and all modern colliders, the particles in the beams are clustered into bunches, and the shape of each bunch varies as they travels in the collider, due to the constituent particles of a bunch having a nonzero momentum in the plane perpendicular to the direction of beam propagation (called the “transverse” plane) – this spread in the transverse momenta of the particles within bunches is characterized by a quantity called the *normalized emittance*, denoted ϵ_n . The resulting “defocusing” of the beam, due to the nonzero emittance, is corrected using quadrupole magnets, or “focusing magnets”, and the effect of these focusing magnets can be well approximated as a harmonic restoring force. The area of the beam as a function of its position in the collider s is expressed using the “beta function” $\beta(s)$, which, in terms of the RMS transverse beam area $\sigma(s)$ is given by:

$$\sigma(s) = \sqrt{\frac{\beta(s)\epsilon_N}{\gamma}} \quad (3.2)$$

where γ is the relativistic gamma factor.

The variable β^* is defined as the value of the beta function at the point, called the *interaction point*, where the bunches pass through one another (a *bunch crossing*). Assuming that the two bunches are crossed exactly head-on, the following equation relates the bunch shape parameters to the luminosity of one bunch crossing (“BX”):

$$\mathcal{L}_{\text{BX}} = \frac{n_{\rightarrow}n_{\leftarrow}\gamma}{4\pi\epsilon_n\beta^*} \quad (3.3)$$

where $n_{\rightarrow}$ is the number of particles in one bunch, and $n_{\leftarrow}$ is the number of particles in the other bunch.

For a circular collider, there will be many bunches per beam, and a single bunch will circulate through the collider with a frequency of revolution f_{rev} , so the instantaneous luminosity for a circular collider can be written as:

$$\mathcal{L} = \frac{N_b f_{\text{rev}} n_{\rightarrow} n_{\leftarrow} \gamma}{4\pi\epsilon_n\beta^*} \quad (3.4)$$

where N_b is the number of bunches per beam.

Since the LHC mostly performs pp collisions or Pb-Pb collisions, rather than colliding particles with their antiparticles (as opposed to e.g. the LEP collider collided electrons with positrons, and the Sp̄pS collided protons with antiprotons), the beams cannot be collided perfectly head-on, since the beam pipe of protons rotating in one sense must be spatially separated from the beam pipe of the protons rotating in the other sense. Therefore, the beams are crossed so that the bunches pass through one another at a crossing angle θ_c , which requires an additional factor to be inserted into Equation 3.3:

$$\mathcal{L} = \frac{N_b f_{\text{rev}} n_{\rightarrow} n_{\leftarrow} \gamma}{4\pi\epsilon_n\beta^*} \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*} \right)^2 \right) \quad (3.5)$$

where σ_z is the RMS of the longitudinal bunch length, and σ^* is the transverse RMS beam size at the interaction point. This additional factor is called the “geometric luminosity reduction factor”. [53]

The number of bunches per beam and the shape of the bunches may change over the course of a fill, and thus the instantaneous luminosity may vary over time. In the case of the LHC, luminosity has small variations that arise due to adjustments to the accelerator’s focusing and bending magnets, but large changes in luminosity occur over the course of a fill due to the number of particles per bunch decreasing over time due to random losses from the beams, as well as particles being lost during bunch crossings due to collisions. Therefore, during nominal operation, an LHC fill will last for under 20 hours, and the highest instantaneous luminosity occurs during the start of collisions, and at this peak luminosity, the LHC’s operating parameters and beam characteristics are shown in Table 3.2.

With these parameters, Equation 3.5 can be used to give an estimate of the peak instantaneous luminosity of the LHC, yielding $\mathcal{L} = 1.4 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, or 13.98 inverse nanobarns per second, which agrees with the measured peak luminosity in 2016.

Table 3.1: History of integrated luminosity delivered by the LHC. Note that the 2017 pp collisions at 5.02 TeV were performed as a “reference run”. [2]

$\sqrt{s}$	Year	Integrated luminosity
pp collisions		
7 TeV	2010	44.96 pb ⁻¹
	2011	6.10 fb ⁻¹
8 TeV	2012	23.30 fb ⁻¹
	2014	
13 TeV	2015	4.22 fb ⁻¹
	2016	40.82 fb ⁻¹
	2017	49.79 fb ⁻¹
	2018	(ongoing)
pp collisions reference runs		
2.76 TeV	2017	5.61 pb ⁻¹
5.02 TeV	2017	334.3 pb ⁻¹
Pb-Pb collisions		
2.76 TeV/nucleon	2011	184.1 μ b ⁻¹
5.02 TeV/nucleon	2015	0.60 nb ⁻¹
p-Pb collisions		
5.02 TeV/nucleon	2013	36,14 nb ⁻¹
5.02 TeV/nucleon	2016	529.9 pb ⁻¹
8.16 TeV/nucleon	2016	188.3 nb ⁻¹

Table 3.2: Nominal LHC operating specifications and beam characteristics [3] [4]. Note that the normalized transverse emittance was reduced from $3.4 \mu\text{m}$ to $2.0 \mu\text{m}$ during the 2016 run of the LHC due to the introduction of the “Batch Compression Merging and Splitting” technique in the Proton Synchrotron. Note also that the bunch crossing angle was reduced substantially from $80 \mu\text{rad}$ to $40 \mu\text{rad}$ at the beginning of the 2016 run.

	symbol	value
Proton energy	E	6.5 TeV
Lorentz factor	γ	6930
Orbit frequency	f_{rev}	11.2455 kHz
Bunches per beam	N_b	2220
Protons per bunch	$n_{\rightarrow} = n_{\leftarrow}$	10^{11}
Normalized transverse emittance	ϵ_N	$2.0 \mu\text{m}$
Crossing angle	θ_c	$140 \mu\text{rad}$
Value of β function at IP	β^*	40 cm
Longitudinal bunch length	σ_z	7.8 cm

3.2 CMS Detector

The CMS detector is a general-purpose, hermetic particle detector, which takes its name from its relatively compact design considering its comprehensive muon detection systems, as well as from its superconducting solenoid magnet, which has an interior diameter of 6 meters and a length of 13 meters, providing a highly uniform 3.8 T magnetic field in the core of the detector. In addition to the solenoid and muon systems, the CMS detector is comprised of several other “subsystems” which act in concert to provide richly detailed descriptions of particle activity in each bunch crossing event; with a bunch crossing frequency of 40.08 MHz, the CMS detector and its triggering and readout systems handle huge amounts (petabytes per second) of incoming data, which along with its size gives it the label of “Big Science”. The CMS detector, located in the SX55 experimental cavern below LHC Point 5 in Cessy, France, is one of the most advanced particle physics instruments ever constructed, and provides a lens through which to study the highest-energy physics phenomena ever produced in the laboratory.

Situated around the point at which the LHC focuses and crosses its beams called the interaction point constructed in a roughly cylindrical shape, the detector, as well as the physics processes it observes, are typically described using a coordinate system defined relative to the interaction point. The coordinate system, shown in Figure 3.3, is commonly used, and will be the coordinate system used throughout this dissertation.

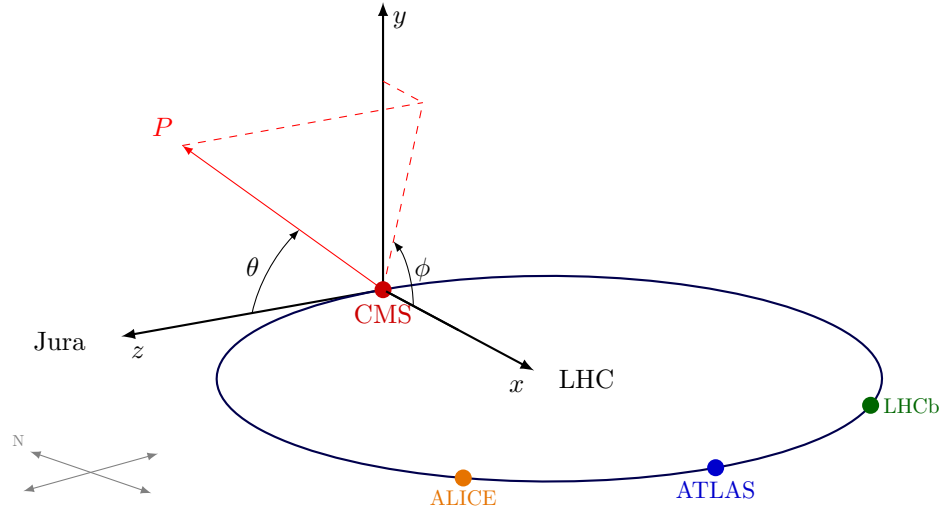


Figure 3.3: Coordinate system used. Taken from [9].

The coordinates x and y point horizontally and vertically in the plane perpendicular to the z axis, which points along the direction of the beams in the counter-clockwise circulating beam, as viewed from above. Typical spherical coordinates are also defined; θ is the angle of a vector in the (x, y, z) space relative to the z axis, ϕ is the angle relative to the x axis to the vector’s projection

into the x, y plane, and r is the magnitude of the vector, which in the context of CMS is commonly associated with the “depth” a particle travels from the interaction point into the mass of the detector. In the context of describing the detector, cylindrical coordinates are also used, where z and ϕ are used in conjunction with ρ , which is the distance from a particle to the z axis. In addition, the commonly-used *pseudorapidity* variable, denoted η , is defined as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (3.6)$$

3.2.1 Detector components

As the results presented in this dissertation were obtained using data taken by the CMS detector during the 2016 pp collisions run of the LHC, the descriptions of the CMS detector given throughout correspond to the configuration of the CMS detector in 2016. Upgrades to CMS that have been done to the detector, in particular a significant upgrade to the tracker done at the end of 2016 as well as significant upgrades to the hadron calorimeter (HCAL) done during the end-of-year stops in 2016 and 2017, are not discussed in this dissertation. A diagram giving a general idea of the layout of the CMS detector is shown in Figure [reffig:sketchUp](#).

3.2.1.1 Solenoid and return yoke

The central feature of the CMS detector is the superconducting solenoid magnet constructed of four concentric coils of 32-strand NbTi superconducting wires, stabilized in a housing of high-purity aluminium. Housed in a cryostat which maintains the superconducting coils at a temperature of 4.65 K, the tube-shaped magnet has an interior diameter of 5.94 m, an external diameter of 7.6 m, and a mass of 225 tons. Operating at a nominal current of 19.14 kA, the CMS magnet generates a magnetic field of 3.8 T inside the magnet’s volume. The magnet is surrounded by five annular dodecagon steel structures called “wheels” (the center-most one supports the magnet cryostat) each of which stands 15 meters tall on two “feet” with heavy-duty air pads that facilitate moving each segment when “opening” the detector. On each end (along the z direction) of the barrel yoke assembly is one 15 m tall endcap yoke, which is comprised of three layers of steel between which the endcap muon detectors can be inserted, and two 15 m tall steel endcap plates, which serves to enclose the detector and provide light leaks, as well as further contain the magnetic field one of each per side. Each wheel is composed of four concentric dodecagonal rings with spokes connecting the concentric layers such that muon detectors can be inserted between the layers of the wheels of the yoke; however the spokes are positioned so that they are staggered in the ϕ coordinate such that a radially or nearly-radially travelling muon cannot avoid passing through any of the inter-layer gaps where muon detectors are located. Made of paramagnetic steel, the yoke has the effect of partially confining the magnetic field (lines) inside the yoke, which allows the solenoid and magnetized steel to maintain a magnetic field of over 2 T throughout the volume of the detector outside the solenoid, where the muon spectrometry instrumentation is installed. The solenoid’s field

stores a total energy of over 2.5 GJ, and is highly symmetrical and homogeneous throughout the detector volume. [71] [72] [73]

3.2.1.2 Silicon trackers

The CMS tracker is innermost part of the CMS detector, closest to the points at which the LHC beams are focused and crossed, called the “interaction point.” Comprised of two main components of the tracker are a silicon pixel detector, closest to the interaction point, contained within a silicon strip tracker outside the pixel detector, comprising a total of about 200 m² of active silicon area.

Pixel tracker

The pixel tracker is the innermost portion of the tracker, named for its use of pixel sensors, which are *n-on-n* sensors with n^+ implants arranged in a grid on n bulk silicon. These silicon pixel sensors are arrayed around three cylindrical layers in the barrel region (BPix), divided into over 700 rectangular modules, comprising a total of 48 million pixels, flanked by two endcap sections (FPix), each consisting of two rotor-shaped layers, comprised of a total of 18 million pixels. The cylindrical BPix layers are located at mean radii of 4.4, 7.3, and 10.2 cm, and are each 53 cm long. On each side of the BPix, at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm, are the rotor-shaped layers of FPix; each layer is divided into 24 rhombus-shaped “blades” arrayed radially outward from the z axis with an inner diameter of 6.1 cm and an outer diameter of 15.0 cm. Covering a region that spans from $\eta = -2.5$ to $\eta = 2.5$, the total area of the silicon pixels sensors in BPix and FPix is 1.06 m². [74]

The passage of a charged particle through the pixel detector, called a “hit”, will cause the production of particle-hole pairs in a depleted region of the bulk silicon. In the 3.8 T magnetic field of the CMS solenoid, the particles traveling in the bulk silicon will experience a Lorentz drift, such that a single hit will lead to charge accumulation on several of the rectangular n^+ implants. This sharing of the charge generated by the hit between different implant regions enables an interpolation method that can determine the position where the hit was incident on a pixel sensor to 10 μm in both directions. Arrays of 52×80 pixels are each read out by a read out chip, or ROC, and groups of 8 to 16 ROCs are read out by Token Bit Manager, or TBM, chips, which transmit the pixel detector’s data to the CMS Data Acquisition, or DAQ, system when a trigger is received by the TBM from the CMS Timing/Trigger Control and Distribution System, called TCDS. [75]

Strip tracker

Surrounding the pixel detector is the Silicon Strip Tracker (SST), named for its use of silicon strip detectors, which use linear p^+ implants on n -type bulk silicon. The strip detectors are assembled into “Tracker Modules”, which contain one or two sensors per module, arrayed in two cylindrical barrel sections, the Tracker Inner Barrel (TIB) and the Tracker Outer Barrel (TOB), along with two endcap regions per side: the Tracker Inner Disks (TID) and the Tracker Endcap (TEC). TIB sits closest to the pixel detector and is composed of four shell structures that span from $\rho = 20$ cm to $\rho = 55$ cm, and extend from $z = -70$ cm to $z = 70$ cm. TID sits next to TIB along the

z direction, consisting of three disk layers in the plane perpendicular to the z axis, extending to $z = \pm 110$ cm. TOB consists of six cylindrical barrel layers extending outward in the ρ direction from the outer layers of TIB and TID to $\rho = 110$ cm. TEC, which begins at $z = \pm 120$ cm, is composed of 9 disk-shaped layers extending to $z = 280$ cm that extend to $\rho = 1100$ cm and have an inner radius that varies along z , following the line $\eta = 2.5$. The overall diameter is 2.4 m and the total length is 5.6 m, and in total is composed of 15,148 tracker modules, which have one strip sensor per module in TIB and TID, with rectangular and trapezoidal shapes, respectively, while there are two strip sensors per module in TOB and TEC, again, respectively, with rectangular or trapezoidal shapes. [76]

As with the pixel sensors, an incident hit will generate particle-hole pairs, and this will lead to charge accumulation on the p^+ implant strips [75], which are parallel to the z axis in the barrel regions, and parallel to the ρ direction in TID and TEC [77]. The charge generated on the strip implants are read out by four or six APV25 readout chips also housed in the tracker module assembly, which receive trigger and timing signals from the TCDS system, and transmit their data to the CMS DAQ system.

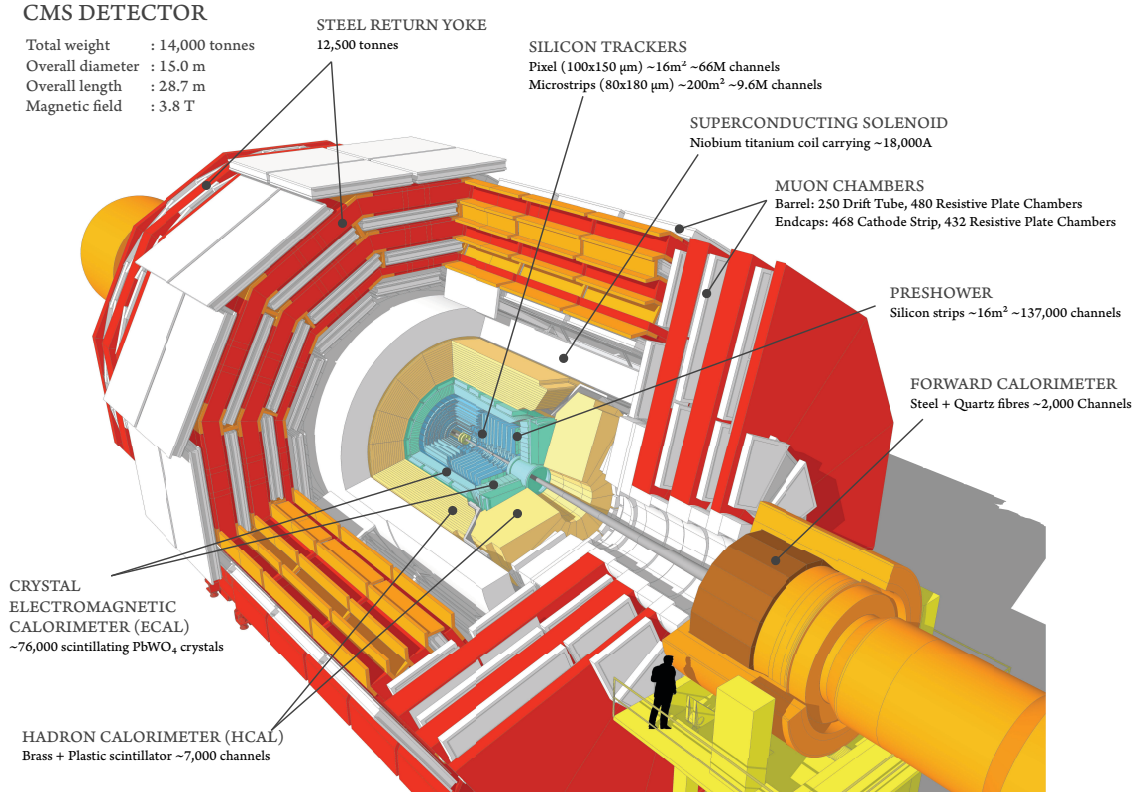


Figure 3.4: A diagram showing the CMS detector with a roughly one-sixth cutaway made to show the interior components of the detector [10].

3.2.1.3 Calorimeters

CMS has extensive calorimetry instrumentation that enables the detection of neutral particles like photons and neutral hadrons, which leave no charge deposits in the silicon trackers, and moreover do not travel a curved trajectory in the magnetic field, and thus must be measured using either absorption or sampling calorimetry. In addition to giving the detector sensitivity to neutral particles, information from the calorimeters is used to supplement measurements from the tracker systems even for charged particles. The two main subsystems of the CMS's calorimetry system are the Electromagnetic Calorimeter (ECAL), which is designed to provide sensitive measurements of photons, electrons, and positrons; and the Hadronic Calorimeter (HCAL) subsystem, which is designed to provide robust detection of neutral and charged hadrons, as well as providing calorimetry of both hadrons, photons, electrons, and positrons in the forward region.

Both calorimeters provide information to the Level-1 trigger system, described later in this dissertation, along with the muon detection systems; however, the tracker system does not. Therefore, the calorimeters also fulfill the crucial role of supplying the measurements used in the process of selecting events to record to long-term storage for physics analysis.

Electromagnetic calorimeter

The CMS ECAL is a total absorption calorimeter that uses lead tungstate (PbWO_4) as the active medium that absorbs photons and electrons emit blue-green scintillation light at wavelengths 420-430 nm which is read out by photodiodes that detect the amount of scintillation light produced, yielding a measurement of the energy of the absorbed particles. Located directly outside the tracker systems, the ECAL is divided as usual into the cylindrical ECAL Barrel (EB) region, which extends to $|\eta| = 1.48$, and the two disk-shaped ECAL Endcap (EE) regions, which extend the ECAL's coverage to $|\eta| = 3.0$. [78] [79]

Lead tungstate is chosen as the scintillating material for its short radiation length, fast response (i.e. the scintillation light is generated quickly; in the case of the EB crystals, 85% of the light is produced in 25 ns), and small Molière radius, which measures the transverse width of the electromagnetic shower. The small Molière radius of lead tungstate, 2.19 cm, motivates the choice of the crystal size; the crystal size is chosen to be as small as possible, leading to increased spatial resolution, while not being much smaller than the natural width of the electromagnetic shower. The dimensions of EB crystals are 23 cm in the r direction, and $2.2 \times 2.2 \text{ cm}^2$ in the perpendicular directions. In EE, the crystals are 22 cm deep, and their front face is $2.86 \times 2.86 \text{ cm}^2$, set back from the interaction point at $z = \pm 3.14 \text{ m}$. In both the EE and EB, the crystals are rotated slightly from being direction directly at the interaction point to ensure that photons emanating from the interaction point do not travel directly along the surface between neighboring crystals. EB has a total of 61,200 crystals, while EE consists of 7,324 crystals. The EB has an inner radius of 1.29 m and outer radius of 1.75 m, and a length of 3 m. [78]

Scintillation light produced in the lead tungstate scintillators is measured by avalanche photodiodes in the EB, and with vacuum phototriodes in the EE. The outputs from the photodetectors

is digitized using ASICs on the Very Front End (VFE) board, which also calculates a compressed form of the data called a “trigger primitive” (TP) to be used by the Level-1 Trigger system. The data and trigger primitives transmitted to off-detector VME electronics, which include a Trigger Concentrator Card (TCC), a Data Concentrator Card (DCC), and a Clock and Control System (CCS) board. The TCC receives the trigger primitive information from the VFE every BX, and process the trigger primitives to elaborate on the TP information. The extended TP information is then propagated to the Level 1 Trigger system. In the event that the Level 1 trigger accepts the event for further usage, a “level 1 accept” (L1A) signal is transmitted back to the ECAL’s CCS, which initiates the readout of the VFE’s data by the DCC board, which the propagates the data to the CMS DAQ system. [80]

Preshower Calorimeter

In addition to the ECAL barrel and endcap portions, the electromagnetic calorimetry is assisted by two lead and silicon sampling calorimeters mounted on the inner face of the EE, these calorimeters are together referred to as the ECAL Preshower (ES) subdetector. The purpose of this detector is to distinguish single photons from closely-spaced photons, which arise commonly in the forward region of the detector from the decay of neutral pions. This measurement requires a higher spatial resolution than the granularity of the ECAL’s lead tungstate crystals, so the ES uses two layers of silicon strip sensors (the two layers have the strips oriented orthogonally), with each layer having a lead absorber layer on its inner surface (10 mm on the inner layer’s inner surface, and 4 mm on the outer layer’s inner surface), located between $298.5 \text{ cm} < |z| < 316.5 \text{ cm}$ in a disk shape with inner radius $\rho = 45 \text{ cm}$ ($\eta = 2.6$) and outer radius $\rho = 120 \text{ cm}$ ($\eta = 1.653$). Incident photons cause electromagnetic showers to form in the lead absorber, and 4,288 silicon strip sensors with 32 strip implants each, otherwise similar to those in the Silicon Strip Tracker, are used to measure the positions of the showers [81]. This enables a spatial resolution of approximately 2 mm to be achieved. Readout of the ES detector uses the Preshower Analog CMS Electronics (PACE3) and auxiliary electronics in a “micromodule”, which is read out using VME off-detector electronics similar to ECAL. [82]

Hadronic calorimeter

The CMS HCAL is a sampling calorimeter that measures light produced by electromagnetic showers in plastic scintillators sandwiched in layers with brass absorber material, with steel structural elements also acting as absorber material. The HCAL is divided into three main regions: the cylindrical HCAL Barrel region (HB), two disk-shaped HCAL Endcap regions (HE), and the HCAL Outer region (HO). HB has an inner radius of 1.811 m and an outer radius of 2.900 m, and is composed of 17 brass/scintillator layers, and joins with the HE in a region between $\eta = \pm 1.2$ to $\eta = \pm 1.48$, forming a hermetic enclosure around the ECAL. HE, composed of 18 brass/scintillator layers, extends the eta coverage to $\eta = 3.0$. Both HB and HE are completely contained within the volume of the CMS solenoid. In total, there are about 70 thousand scintillator tiles in the HCAL. The brass absorber material, commonly known as C26000 or “cartridge brass”, is composed of 70% copper and 30% zinc, and has a radiation length of 1.49 cm and a nuclear interaction length of 16.42

cm. [83]. Just beyond the solenoid lies the HO, which makes use of the volume of the solenoid as absorber material, composed of 2 layers of scintillator separated by a 19.5 mm thick piece of iron in the central region $|\eta| < 0.348$, and one layer of scintillator in the region extending to $|\eta| < 1.24$, which extends the total depth of the calorimeter system to 11.8 interaction lengths, except for a small gap between the barrel and endcap regions, serving as a “tail catcher” for pions that pass through the calorimeters. [84]

The active plastic scintillators are made of SCSN-81, a polystyrene based scintillator in all sections of the HCAL except for in innermost layer of HB and HE, where 9 mm thick BC-408, a polyvinyl toluene-based scintillator, is used. The outermost layer of scintillator in the HB has a thickness of 9 mm, while all the other layers have a thickness of 3.7 mm. These scintillator tiles are instrumented with wavelength-shifting fibers which collect scintillation light inside the calorimeter tiles and transmit the light via normal clear fibers to on-detector electronics assembled into units called Readout BoXes (RBX). Housed within the HB and HE RBXes are photosensors called Hybrid Photodiodes (HPD) which measure the scintillation light collected by the wavelength shifting fibers [as of 2016; subsequent upgrades to HE have replaced HPDs with silicon photomultipliers, however this description of HCAL refers to the configuration in 2016]. Each HPD is exposed to the wavelength shifted light via an “optical decoder unit” [85], which channels light from all scintillator tiles with a given η and ϕ coverage so that they activate the same cell of the HPD; i.e. the readout is combined longitudinally. A similar technique is used to longitudinally combine different scintillator layers in HE; however in the forward regions of HE, the readout of layers is combined so that two or three independent “depths”, or groups of scintillator tiles, are combined longitudinally [83]. In HO, scintillation light collected in the wavelength shifting fibers is measured using Silicon Photomultipliers (SiPM), installed in 2013, which require the use of a Peltier heat pump for temperature regulation [86].

Within the RBXes are further electronics which have the functions of digitizing the analog data from the photodectors, receiving clock signals, receiving and propagating fast control signals and slow control commands, and injecting light into the optical sensing system for calibration purposes, as well as other calibration functions. 16 Charge Integrating Electronics (QIE) boards per RBX are responsible for digitizing the analog signals from the HPDs and SiPMs, which use the QIE8 chip, which does an analog-to-digital conversion of the HPD or SiPM signals [87]. In order to propagate clock signals, or timing and synchronization signals, to the QIE boards, a Clock, Control, and Monitoring (CCM) module in the RBX propagates timing signals from the TCDS subsystem of CMS to the HCAL front-end electronics, via serial connections to an Acromag electronics crate. The CCM module also propagate fast control signals, such as binary control signals called “BGos”, to the rest of the RBX, which are used to send commands, e.g. a QIE resynchronization signal called the “QIE reset”, to the readout electronics [88]. In addition to the QIE boards and CCM, RBXes also contain a module called the “Calibration Unit” (CU), which is mainly used for injecting light-emitting diode light into the photosensor, used for calibrating and monitoring the health of the photosensors. [87]

The HCAL QIE8 boards send their data collected during each bunch crossing to the back-end VME electronics – each QIE8 board has a dedicated link to a back-end VME board called the HCAL Trigger and Readout (HTR) board. The 16 HTR boards in a VME crate propagate their data to two Data Concentrator Card (DCC) boards in the same crate, and each HTR calculates trigger primitives derived from the received data which it sends over a dedicated link to the Level-1 Trigger system. The DCC board buffers the data from the HTR, and receives L1A signals from the TCDS subsystem via a dedicated link. If an L1A signal is received for a given bunch crossing, the DCC propagates the data corresponding to the event passing the Level 1’s trigger requirements, as well as the data from 4 prior and 5 following bunch crossings, to the DAQ subsystem. [83]

Forward calorimeter

In addition to the HB, HE, and HO, a forward calorimeter known as the Hadronic Forward Calorimeter (HF) supplements the calorimetry of CMS by providing calorimetric measurements for both hadrons as well as electrons and photons in the η region beyond the coverage of HCAL, ECAL, and the silicon tracker. The two cylindrical halves of HF lie outside the interior of the solenoid at a distance ± 18 m along the z axis from the interaction point and, with an inner radius of 12.5 cm and outer radius of 130 cm, cover the region $3 < |\eta| < 5$, with a depth of 165 cm totaling 10 interaction lengths. Each half of HF is constructed in 20° wedges machined with a grid of grooves running parallel to the z axis, spaced in the x and y directions by 5 mm. These grooves alternate along the x and y directions between spanning the entire volume of HF, or only spanning from the far end of HF to 22cm beyond the inner face of HF. These grooves are instrumented with $600\mu\text{m}$ quartz fibers, where fibers inserted into the longer grooves are called “long fibers”, and fibers in the shorter grooves are called “short fibers”. High-energy particles traveling through the quartz fibers at a velocity greater than the speed of light in the quartz fibers form a shock wave of electromagnetic radiation, known as Cherenkov radiation, and the quartz fibers totally internally reflect a fraction of the Cherenkov radiation down the fibers. Fibers are grouped into 24 separate regions per wedge corresponding to 13 different layers in η , with two regions in ϕ per layer, except for the two innermost layers. The long and short fibers are read out separately for each of the 24 regions by photomultiplier tubes mounted near the back of each HF half. [89]

Off the detector, but still in the experimental cavern, the analog electrical outputs of the photomultiplier tubes are channeled to Read Out BOXes (ROBOX), which are crates of electronics that hold QIE boards, CCMs, and CUs similar in function to the on-detector electronics in HCAL RBXes, but with substantially different design to accomodate living in an off-detector crate rather than in on-detector enclosures. The data from the HF QIE boards is also handled in a similar way to the data from HCAL; however (since 2016) the back-end electronics underwent an upgrade from the legacy VME HTR and DCC to smaller, more powerful uTCA electronics. The functions of the HTR – reading the data from the ROBOXes and forming TPs for the Level-1 trigger – were replaced using the uHTR uTCA board. The functions of the (two) DCCs – receiving and sending clock and trigger signals from TCDS to the rest of the crate, buffering the data, and transferring it to the

DAQ system when L1As are received – were replaced by the (single) AMC13 board. [90]

3.2.1.4 Muon Detectors

CMS is extensively instrumented with muon detectors outside the solenoid and endcap calorimeters. Three main technologies are used for muon spectrometry: in the barrel, four layers each comprised of both a gas-based Drift Tube (DT) and a Resistive Plate Chamber (RPC) inhabit gaps between the concentric dodecagonal layers of the wheel segments of the magnetic return yoke, which (as described earlier) are arranged such that no regions in η, ϕ space are present for a muon traveling outward from the interaction point. In the endcap regions, Cathode Strip Chambers (CSC) and Resistive Plate Chambers (RPC) are mounted within layers of the two endcap yokes. [91] A diagram of the layout of the various muon detectors can be seen in Figure 3.5.

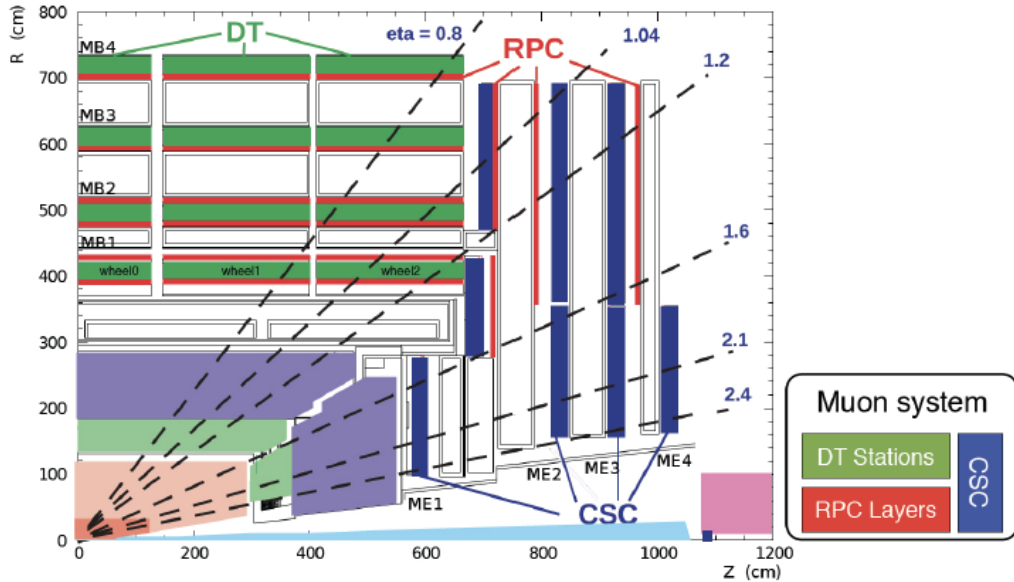


Figure 3.5: A diagram showing a cross section through the center of one quarter of the CMS detector. The DT, CSC, and RPC detectors are indicated. Taken from [11]

Drift tubes

In the barrel region, the muon detector consists of 250 muon “stations”, which contains one DT and either one or two RPCs (discussed later). There is one station located in each of the three interlayer gaps in each of the 12 wedges of each barrel yoke wheel, as well as one station located behind each wedge outside the magnetic return yoke, totaling 250 total stations, each of which consists of one DT and one or two RPCs. The muon barrel detectors inner face is 4 m away from the beam line in the ρ direction, and extends 6.5 m in the z direction, covering a pseudorapidity range of $|\eta| < 1.3$. [91]

Each DT station contains up to 60 individual “drift cells”, which consist of an anode wire stretched

inside a sealed cavity containing a medium of 85% argon and 15% carbon dioxide gas 13 mm thick in the ρ direction and 42 mm wide, with a total length of 2 to 3 m (larger length DT cells are used in the larger muon stations in the outer region of the barrel). High voltages are applied to strip-shaped cathodes that run parallel to each cell's anode wire, which shapes the electric field in the gas medium such that free electrons in the gas are channeled to the anode wire. When muons pass through the gas mixture inside a drift cell, electrons are liberated via ionization of the gas, and their accumulation on the anode wire creates an electrical signal that can be read out using on-detector electronics. Within a single DT station, these many drift cells are arranged in three main groups, called superlayers, which are each composed of where two of the superlayers per DT station have their cells' anode wires running parallel to the beam line, which yields a measurement of the muon's ϕ coordinate, while a third layer located in between the other two has its cells arranged roughly parallel to the ϕ direction, which yields a measurement of the η coordinate. The drift cells inside each superlayer are arranged in four planar layers, where layers are offset by one-half a cell width between each layer. In total, the DTs have about 172 thousand sensitive wires. [92] Signals carried by the anode wires in each DT station are digitized by a set of electronics called a "minirate" attached to every DT station, of which the main component are known as the "Read Out Boards" (ROB). Depending on the number of channels in a DT station, a minirate may have between 3 and 7 ROB, each of which have 4 highly programmable ASICs known as "High Performance Time To Digital Converters" (HPTDC). HPTDCs read out the electrical signals coming from the DT anode wires in 16 independent channels, and create digitized timing measurements of hits in the DTs, with a time resolution of 265 ps, which suffices to obtain a resolution of 300 μm in position measurements. An Altera CPLD manages merging the data from all the HPTDCs on an ROB and propagates the data to the off-detector "Read-Out Server" board, which resides in a VME crate in the experimental cavern. The "Trigger Sector Collector" (TSC) boards in the same crate collect data from the DT chambers and calculate trigger primitives which are propagated to the Level 1 trigger system. When an L1A is received, the ROS boards then transfer their data to VME boards in the service cavern, called the "Detector-Dependent Unit" (DDU), which propagates the DT data to the DAQ system. [93]

Cathode strip chambers

The endcap muon spectrometer system consists of CSCs and RPCs instrumented throughout four annular rings in each of the two endcap regions of the detector; each side consists of one plane of CSCs on the inner face of the endcap yoke, two planes within the endcap yoke, and one plane between the outside of the endcap yoke and within the endcap plate. The endcap muon detectors' innermost face is 6 m from the interaction point and extend 7 m outward in the ρ direction, covering a pseudorapidity range from $\eta = 0.9$ to $\eta = 2.4$. Each plane is divided into concentric inner and outer (larger and smaller $|\eta|$, respectively) rings, except for the right closest to the interaction point, which is divided into three concentric rings. Each of these rings consists of two overlapping layers (except the outermost ring of the plane closest to the interaction point) of wedge-shaped CSC detectors

that span 20° or 30° in ϕ , which are arranged so that they overlap one another. The wedge shaped CSC detectors house a gas-based detection system which is similar to a proportional wire chamber augmented with orthogonal cathode strips. CSCs are filled with a mixture of 50% carbon dioxide, 40% argon, and 10% tetrafluoromethane gasses, within which wires are stretched such that they form six planes of parallel wires. Interleaved within the planes formed by the wires are panels etched with conducting strips that run in the perpendicular direction to the wires. The trapezoidal sections of the CSC detector are instrumented such that the wires go outward in the ρ direction, while the strips run at lines of roughly constant η . During operation, the wires are given a large positive voltage while the strips are given a large negative voltage; a muon passing through the gas creates electrons and positive ions which accelerate in the electric field within the chamber, which in turn causes more ions and electrons to be formed in the gas, called “avalanche multiplication.” This charge accumulates on the anode wires, creating electrical signals which can be read out by on-detector electronics, and the combination of readings from the wires and the strips gives a measurement of the muon’s η and ϕ coordinates. Altogether, the CSC system is comprised of 468 chambers with 6000 m^2 sensitive area with 217,728 cathode channels and 183,168 anode channels. [94]

Each chamber has 480 cathode strip channels and 384 anode wire channels, and the electrical signals they carry are read out using on-detector electronics called “Anode Front-End Boards” (AFEB), which can number between 18 and 42 per CSC, and “Cathode Front-End Boards” (CFEB), of which there are typically 5 per CSC. The AFEB and CFEB amplify the anode and cathode signals they respectively receive and propagate fast trigger information to another on-detector board called the “Anode Local Charged Track Board”. These on-detector electronics propagate the trigger information to off-detector electronics located in the experimental cavern called “Trigger Motherboards” (TMB) which forms TPs by checking for coincidences between cathode and anode hit patterns and propagates them to both the Level 1 Trigger system as well as a second piece of off-detector electronics called the “Data acquisition MotherBoard” (DMB) which buffers the TP information. Upon receipt of an L1A from the trigger system, the DMB collects all the data from the AFEB and DFEB, and propagates the data and the cached TP information to a DDU, similar to the DT readout. [95]

Resistive plate chambers

The RPC detectors are instrumented in both the barrel and endcap regions to compliment the DTs and CSCs, respectively, by serving as a dedicated muon detector for generating trigger primitives for the Level 1 trigger. RPCs are installed in the barrel and endcaps in all DT stations and CSC chambers up to $\eta = 1.8$, as shown in Figure 3.5. The RPCs are constructed in “double-gap” modules, referring to the two gas cavities in each RPC module. These two gas cavities have a gap width of 2 mm and are surrounded by bakelite electrodes, a high-resistivity ($10^{10} \text{ } \Omega\text{m}$) plastic, coated on one side with graphite paint, allowing them to be charged up with a high positive voltage. The gaps are filled with a mixture of 95.2% $\text{C}_2\text{H}_2\text{F}_4$, 4.5% C_4H_{10} , and 0.3% SF_6 , and the high-voltage resistive electrodes causes the gas to be ionized by a passing muon; the electrons liberated from a passing muon then proceed to form an avalanche of other electrons, which are deposited on the

high-resistance electrode. Because of the high resistance, the charge deposited on the electrode remains localized long enough to induce a charge on a nearby readout strip. The readout strip is located between the two gaps, such that the charge read out is the sum of the induced charges from avalanche deposits on the electrodes of both chambers. The RPC strips are read out by up to 18 on-detector “Front-End Boards” (FEB) per RPC, which are home to up to four custom ASICs known as FEB chips, which digitize the incoming electrical signals. Up to 12 FEB chips are read out by an “LBox”, which contain a “Link Board”, which propagate their signals to the L1 Trigger electronics in the service cavern.

3.2.2 Trigger, timing, and data acquisition

With a bunch-crossing frequency of 40.08 MHz, the LHC is a very precisely timed instrument, and thus the CMS detector must be both precisely synchronized with the LHC, and also precisely timed internally; i.e. between the various subdetectors of CMS. Furthermore, since the total size of the data from a single event recorded by CMS is approximately 1 MB, this naïvely would imply that the CMS requires a data throughput of nearly 50 TB per second to be written to long term (i.e. disk or tape) storage. Such a high data throughput would be prohibitively expensive and extremely challenging in terms of computer systems engineering. These challenges of synchronization, data collection, and data storage are handled by several systems that make up the CMS Data Acquisition systems. Four main systems—the Timing and Control Distribution System (TCDS), the Level 1 trigger, the “Data Acquisition” (DAQ) subsystem, and the High-Level trigger—are designed to ensure proper synchronization with the LHC and within the CMS detector, the analysis of trigger primitives from the various subsystems to calculate whether an event is interesting enough to be further propagated toward long-term storage (an L1A), the propagation of L1As to the subsystems, the propagation of “backpressure” from subsystems back to the TCDS system if the demand for data corresponding to L1As cannot be met, the readout of the data from the subdetectors for each L1A, and the final filtering of the data before finally being written to long-term storage. This system overall lowers the rate of events stored into long-term storage by the CMS detector down from 40.08 MHz to the manageable rate of 100 Hz, or 100 MB/s. [96]

3.2.2.1 Clock and synchronization

One of the main tasks of the CMS TCDS subsystem is to receive a synchronization or “clock” signal from the LHC and propagate this clock signal to all the on-detector and off-detector electronics of the CMS detector subsystems as well as to the other elements of the CMS data acquisition system. Upstream from the CMS TCDS system, clock signals must be propagated between the various accelerators in the LHC injection chain; e.g. the SPS and PS must have very accurate synchronization for filling the SPS. Similarly, during LHC fills, the LHC receives its clock signals from the SPS. However, after the LHC has been filled, the LHC’s own clock (i.e. the one which stays in synchronization with the RF at P4 and thus in synchronization with the beams) defines the timing of bunch crossings within the various LHC detectors. CMS receives its clock signals

from the LHC’s clock distribution system via a “Radio Frequency Receiver” (RFRx) board, which is converted into the CMS standard “Timing, Trigger, and Control” format for beam-synchronous signals by the “Radio Frequency To TTC” (RF2TTC) board [97]. This “TTC stream” is then split and then propagated to a TCDS electronics crate, which are designed for the uTCA crate electronics standard, and feed into an AMC13 board specifically designated to act as the “Central Partition Manager” (CPM) which propagates clock to the rest of the TCDS system. [98] The uTCA crate holding the CPM distributes its clock signal over the uTCA crate’s backplane to up to 12 uTCA boards known as “Local Partition Managers” (LPM), each of which can propagate clock to an entire crate of 12 “Partition Interface” boards, which connect directly to the subsystems’ electronics for delivering the clock signals to the detector front-ends. Each PI board corresponds to an independent “partition” of the clock “tree” and can be operated independently from other partitions. Careful measurements of latencies at each stage in the clock distribution chain are made and corrected for to ensure synchronization between all electronics within and between partitions. [98]

3.2.2.2 Level-1 trigger

The Level-1 trigger system is designed to perform fast analysis and low-level reconstruction on the trigger primitives from the various CMS subdetectors for determining whether the event meets certain criteria, e.g. whether some interesting process has occurred, and to issue an L1A (Level 1 accept) to propagate to the subdetectors’ readout systems to initiate a full readout of the data from that event. The Level-1 Trigger is instrumented with custom uTCA electronics, and receives input from the calorimeters as well as the muon detectors (note that the tracker systems do not participate trigger primitives to the Level-1 trigger.) The trigger primitives that are sent as inputs to the Level-1 trigger must be processed and the L1A must be issued within in $3.8 \mu\text{s}$, as this restraint is imposed by the finite capacity of the various subdetectors’ readout electronics to buffer incoming data. Two main parts of the Level-1 trigger handle these two main inputs: the calorimeter trigger and the muon trigger. [99] The electronics of the calorimeter trigger implement various algorithms to quickly reconstruct electron and photon candidates, tau lepton candidates, and jet candidates, and to compute various energy sums of these candidates, such as vector sums for calculating missing transverse energy (MET) and scalar sums such as HT, the scalar sum of jet candidates’ transverse energy. The trigger primitives from various regions of the calorimeters are organized according to the granularity of the HCAL and forward calorimeter, e.g. in the central region of the detector, a trigger tower consists of all channels in a 0.087×0.087 region of η, ϕ space. In the calorimeter trigger, the grouping and distribution of data is performed by Calorimeter Trigger Processor uTCA cards (CTP7), called “Layer-1”, while the algorithms to reconstruct the various physics object candidates are implemented in different uTCA cards known as Master Processor Virtex 7 (MP7) cards, referred to as “Layer-2”. [99]

Similarly, the muon trigger instrumentation is designed to reconstruct muon candidates from the three main muon detectors, and is divided into three main parts that correspond to the η regions covered by the muon detectors whose TPs are being received by the respective parts: the Barrel

Muon Track Finder (BMTF) receives TPs from the DT and RPC detectors in the $|\eta| < 0.83$ region, the Endcap Muon Track Finder (EMTF) receives TPs from the CSC and RPC detectors in the region $|\eta| > 1.24$, and the Overlap Muon Track Finder receives TPs from DTs, CSCs, and RPCs in the region where they may overlap, which is $0.83 < |\eta| < 1.24$. Muons are reconstructed by combining hits from each of the muon detectors and running fast fitting algorithms to reconstruct tracks and to determine the quality of the track. From these tracks, the muons' energies are calculated. The BMTF is instrumented using MP7 cards, whereas EMTF and OMTF are implemented using a different uTCA board called the Modular Track Finder (MTF7), which provides a greater memory capacity for storing look-up-tables that expedite the fitting algorithms. [99]

The output from the calorimeter and muon triggers are then propagated to the Global Trigger (μ GT), which implements a set of up to 128 physics-objectbased algorithms to calculate the final L1A decision for each event. The algorithms performed by the μ GT are implementations designed to identify processes of interest [100]; for example, a hypothetical algorithm the μ GT may run could be an algorithm that checks for a muon candidate reconstructed by the muon trigger, plus a jet reconstructed by the calorimeter trigger, plus some missing transverse energy which the μ GT determines by running a vector sum over all candidates reconstructed by the muon and calorimeter triggers, which could serve as a Level-1 trigger algorithm to identify events that may contain a top quark.

After running all the algorithms that correspond to checking for events that might be of interest, the μ GT calculates a final decision, the L1A, as the logical *OR* of the pass/fail decisions from all the different algorithms. In all, algorithms implemented in the Level-1 trigger are tuned to achieve a reduction of the overall event rate from 40.08 MHz down to approximately 100 kHz or less; depending on the instantaneous luminosity of the LHC, various thresholds and parameters of the algorithm may be tuned on the fly to achieve a 100 kHz event rate; e.g., over the course of an LHC fill, the instantaneous luminosity decreases over time, and the Level-1 trigger changes, among other things, the prescales applied to various μ GT algorithms (i.e. a prescale factor of f will only issue an L1A for one out of every f events passing the μ GT algorithm with that prescale.) If an L1A is issued, the Level-1 trigger system propagates the L1A TTC signal to the TCDS system. [99]

3.2.2.3 TTC and TTS

The TCDS system has various functionalities in addition to propagating clocks to the detector's read out electronics, including propagating the L1A TTC signal from the Level-1 trigger to the various readout electronics that buffer the full data from the detector. TCDS also enforces various "trigger rules"; e.g. TCDS will not permit two L1As to be issued in two consecutive bunch crossings, and will not permit more than 2 L1As to be issued per every 25 bunch crossings. The TCDS system also embellishes the TTC stream with various "fast commands"—commands that must be issued synchronously with particular bunch crossings—to the subdetector systems, referred to as "BGos", such as commands that indicate the first bunch crossing in an orbit, or e.g. commands that fire the ECAL's calibration laser, or a command to have the HCAL fire its LEDs into the photodetectors,

which are both actions that those subdetectors take for calibration that must happen at particular segments within the LHC orbit (in this case, these actions occur during the “calibration gap,” a group of empty LHC RF buckets where no collisions occur.) The final function of the TCDS system is the “Trigger Throttling System” (TTS), is a system whereby the subdetectors of CMS may send a signal to the TCDS system to prevent L1As from being issued. The readout electronics from each subdetector are equipped with either a fiber optic or electrical wire link to a compact PCI (cPCI) board known as “Fast Merging Modules” which compute an overall TTS state for a subset of the subdetector electronics, and a chain of FMMs combine these to form an overall TTS state for the entire detector that is propagated to the TCDS system, which then throttles the issuing of L1As to the entire detector. This system is implemented to allow the subdetectors to give feedback to the trigger/TCDS system when triggers are arriving in such a way that the subdetector would not be able to provide good data; e.g. if too many L1As are issued, the readout electronics may not be able to transmit the data quickly enough, or may not be able to buffer the data correctly. Thus, the TTS system is an important aspect of the TCDS system that helps to ensure the integrity of the data. [98]

3.2.2.4 Data acquisition

The role of the CMS DAQ system is to collect the data from the subsystems’ readout electronics upon the issuing of an L1A, to compile all the data corresponding to a given L1A into a single “event” containing all the subdetectors’ data from that particular BX, and to buffer that data for processing by the High-Level Trigger (HLT). With an L1A rate of 100 kHz from the Level-1 trigger and an event size of approximately 1 MB per event, the CMS DAQ system must handle a throughput of 100 GBps, making it the highest throughput DAQ system in high-energy physics to date. The electronics belonging to each subdetector which buffer the data and transfer it to the DAQ system are collectively referred to as “Front-End Drivers” (FED), and these FEDs are connected via fiber optic data transfer links to the DAQ system’s “Front-End Readout Optical Links” (FEROL), which are cPCI-based electronics. The FEROL then acts as a 10 GBps ethernet NIC to send the data via TCP/IP to the next element of the DAQ system, the Readout Unit (RU), via commercial 10/40 GBps ethernet switches. [101]

The RUs are 84 Dell PowerEdge R620 computers with high-speed ethernet and infiniband NICs, which merge the data received over ethernet TCP/IP from several (typically 12) FEROLs. An infiniband network connects the RUs to a system of 64 Builder Unit (BU) PCs, which serve to assemble the data compiled by the various RUs and write them to disk storage in the form of files, where the “Filter Unit” (FU) machines, on which the HLT system runs. [101]

3.2.2.5 High-level trigger

The CMS HLT is implemented in software that is part of the CMS Software (CMSSW) framework (but not the full-fledged reconstruction) and runs on a cluster of commodity computers called the

Event Filter Farm (EVF). The RU machines which buffer the incoming data deliver the completely-assembled data from an event to an FU machine on which the HLT algorithms are run, which then unpacks the data and performs event reconstruction and trigger filtering. The EVF comprises upwards of 13,000 CPU cores, with a single event typically taking about 90 ms on one core. The HLT performs its reconstruction in several discrete steps, for instance, selections that rely only on the calorimeters and muon detectors are applied such that CPU-intensive reconstruction of tracks using tracker hits can be avoided for a large subset of events. This step-by-step approach is referred to as a “trigger path”. The HLT filters the event rate from 100 kHz down to about 100 Hz, resulting in an rate of data output of 100 MBps. After all algorithms for all trigger paths have completed, the data from events passing selection are transferred to Storage Manager for transfer to long-term storage, with subsequent transfers delivering the data to the CMS Tier-0 computing center, where the full-fledged data reconstruction is performed using the same CMSSW framework. The data handled by the Storage Manager is organized on an event-by-event basis into different datasets corresponding to the HLT triggers that have been fired for that event, for example, events firing HLT triggers that correspond to the presence of one photon will be added into the SinglePhoton “primary dataset” (PD). [100]

3.2.3 Event reconstruction

Heuristically, the overall design of the CMS detector can be thought of as having the following main features:

- A strong and extensive magnetic field (a large field integral) throughout the detector, providing clean, easily modeled, well isolated tracks
- Extensive tracking systems in the region $|\eta| < 2.4$ with a high-granularity inner pixel detector and outer tracking strip sensors which allow a precise track measurement
- A very precise electromagnetic calorimeter, which yields excellent energy measurements for isolated photons, with high granularity
- A hermetic hadron calorimeter that can reconstruct the energy of neutral hadrons
- A muon system that enables highly efficient identification and precise spectrometry of muons
- Extensive forward calorimetry that ensures hermeticity of the detector, enabling reliable measurements of missing transverse momentum (MET)

These distinctive features of the CMS detector have enabled a new technique in event reconstruction to be made possible (having only been utilized successfully in the past for detectors at electron/positron collider experiments such as the ALEPH experiment at LEP): the Particle Flow (PF) approach to event reconstruction. Before the advent of particle flow in CMS, the techniques employed to reconstruct physics objects like muons, electrons, photons, and jets typically only used a subset of the detector’s information; for example, in the “Energy Flow” paradigm, jets were only

reconstructed using information from calorimeter systems. The PF algorithm, on the other hand, takes a holistic view to reconstructing the physics objects, and attempts to reconstruct all the particles observed in each event by incorporating all available information from all subdetectors of CMS. This approach greatly improves measurements of jet energies, reconstruction of tau leptons, tagging of b-jets, and the reconstruction of many other physics objects. [102]

This section discusses those elements of the particle flow algorithm that are relevant to the reconstruction and identification of photons and jets, with additional discussion of tagging of jets that are believed to originate from b quarks. For a detailed description of the particle flow algorithm as it pertains to electrons, muons, missing transverse momentum, and tagging of jets originating from other heavy particles such as tau leptons, top quarks, and W bosons, see [12], [103], [104].

3.2.3.1 Tracking

The first step in CMS event reconstruction is the identification of charged particle tracks from hits measured in the pixel and strip tracker subdetectors using an iterative-tracking algorithm [105] known as the Combinatorial Track Finder (CTF), based upon the Kalman filter method [106] [107]. This iterative procedure begins by forming a “seed” of two hits in consecutive layers of the pixel detector consistent with the trajectory of a charged particle leaving the interaction region, which gives a rough estimate of the track’s parameters such as the η, ϕ direction of the charged particle candidate and the vertex from which it originated. This initial step in seeding the tracks is performed with very stringent criteria on the quality of the pixel hits, ensuring a very low probability of tracks being seeded from hits caused by separate particles in the event or detector noise—thus the initial seeds are intended to minimize the “fake rate” of reconstructed tracks while preserving moderate track reconstruction efficiency. Using the initial estimate of the seed’s trajectory, the tracking reconstruction algorithm extrapolates the trajectory to determine the next tracker layer where the charged candidate is expected to leave a hit, according to the equations of motion of a charged particle in a constant magnetic field taking into account possible scattering of the charged particle off layers of the tracker material as well as energy lost during traversal of the tracking material or support structure. If a hit in the next layer can be unambiguously identified with the seed’s trajectory, this hit can be incorporated into the track’s fit to update the track parameters for better precision, according to the Kalman filter formalism. These unambiguous hits are then removed from subsequent iterations of the CTF algorithm, and the algorithm proceeds iteratively with progressively looser track seeding criteria, which increases the efficiency for reconstructing tracks for more difficult classes of tracks, such as low p_T charged particle tracks or highly displaced tracks.

This process proceeds iteratively, incorporating an additional layer of the tracker at each iteration. When several hits in a tracker layer are compatible with the predicted trajectory, one further trajectory candidate is created per hit, and these track candidates are extrapolated likewise in an iterative procedure – when a trajectory is found to have no hits in layers where the track is predicted to leave a hit, that track candidate can be ruled out as an “invalid hit”. [107] [105] [12]

3.2.3.2 Calorimetry

In addition to the detailed measurements of charged particle candidates from the CTF algorithm’s reconstruction of tracker hits, the PF algorithm also relies heavily on ECAL and HCAL measurements in four main roles:

- Detection and energy measurement of stable neutral particles such as photons and neutral hadrons
- Distinguishing neutral particles’ energy deposits from charged hadron energy deposits
- Reconstruction and identification of electrons
- Improvement of energy measurements for charged hadrons in cases where the tracking measurement alone cannot yield a reliable estimate, such as is the case for low-quality tracks or tracks corresponding to charged candidates with very high p_T

The first step in the reconstruction of photons and neutral hadrons is a set of algorithms that form clusters of energy deposits in the calorimeters; a clustering of energy deposits is performed independently for each of ECAL, HCAL, and ES (while in HF no specific clustering is used as each channel is regarded as its own cluster). PF calorimeter clusters begin with the identification of “cluster seeds” which are individual calorimeter clusters with energy exceeding a given threshold (240 MeV in EB, 600 MeV in EE, 800 MeV in HB, 1.1 GeV in HE, and 0.12 MeV in ES.) and exceeding the energy of all neighboring channels. The cluster is then extended by forming a “topological cluster”, whereby any channels with energy deposits greater than twice the channel’s noise level are included into the cluster. An expectation-maximization algorithm based on a Gaussian-mixture model is then employed to reconstruct clusters from the topological clusters, whereby each seed is modeled with a corresponding Gaussian distribution, and a maximum likelihood fit of the sum of Gaussians is performed on the spatial distribution of hits in the calorimeter under consideration. The η and ϕ coordinates as well as the amplitudes of each Gaussian term of the pdf are allowed to float in this fit, while the width of the Gaussian are fixed according to the subdetector under consideration (1.5 cm in ECAL, 10 cm in HCAL, and 0.2 cm in ES), and the best-fit results for these Gaussian components are then used as the final values of the η, ϕ coordinates and energies of each cluster in the calorimeters. [12]

3.2.3.3 Links

After the CTF and calorimeter deposit clustering algorithms have calculated the track and cluster objects, the PF algorithm proceeds with a “link” algorithm which establishes connections between tracks and calorimeter deposits as the next step in reconstructing the event’s candidate particles, called “PF candidates”. The link algorithm considers nearest neighbor tracks and calorimeter clusters in a pairwise process and calculates a measure of the quality of the link, called the “distance”, which is used as a rough probability that the tracks or clusters resulted from the same particle in the event. Tracks and clusters that are directly linked or linked through a chain of intermediary

links are then associated into a “PF block” which is then passed to the PF candidate identification algorithms.

A link between a track and a calorimeter cluster is formed by extrapolating the charged particle’s track to the η, ϕ plane – in the case of the ECAL at a depth corresponding to the expected maximum of a typical longitudinal electron shower profile in the ECAL, or in the case of the HCAL at a depth of one interaction length – and a link is established between the track and the cluster if the extrapolated trajectory is within the area of the cluster extended in all directions by one additional calorimeter channel. The distance parameter is then defined as the distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between the extrapolated track’s position and the cluster’s calculated η, ϕ coordinates. If more than one HCAL cluster is linked to a track, or if more than one track is linked to an ECAL cluster, all links besides the one with the smallest distance are discarded. [12]

Links between ES and ECAL clusters as well as links between ECAL and HCAL clusters are based on whether the cluster of the calorimeter with finer granularity lies within the envelope of the cluster envelope of the coarser grained calorimeter; in particular, an ES cluster is linked to an ECAL cluster when the coordinates of the cluster in the x, y plane are within the envelope of the ECAL cluster, and similarly, an ECAL cluster is linked to an HCAL cluster if the η, ϕ coordinates of the ECAL cluster are within the cluster envelope of an HCAL cluster, and in both cases the distance parameter is defined in a similar manner to the definition of ΔR above. If multiple HCAL clusters are linked to the same ECAL cluster, or if multiple ECAL clusters are linked to the same ES cluster, all links besides the one with the smallest distance parameter are discarded. [12]

Tracks of charged particles are also assigned links that correspond to the tracks sharing a secondary vertex. If at least three tracks are linked, corresponding to one incoming charged particle and at least two outgoing charged particles, the secondary vertex where they meet is saved; however if fewer than three tracks are linked, then the reconstructed secondary vertex is discarded.

A diagram that gives a basic idea of how information from various different subdetectors is combined to give an overall description of particles in an event is shown in Figure 3.6.

3.2.3.4 Photons

Photons emanating from the primary vertex (prompt photons) that travel directly into the ECAL would in principle be easy to reconstruct as they are fully absorbed in the ECAL, like electrons, but do not leave hits in the silicon tracker, and thus one might assume that photons are easily reconstructed by identifying an ECAL deposit with no collinear HCAL clusters or muon detector hits as well as no consistent tracks in the tracker. However, only 30% of photons reach the ECAL without first converting into pairs or single electrons or positrons, due to the considerable probability of electrons interacting with the intervening material of the tracker. Reconstruction of the 70% of prompt photons that undergo conversion into electrons and/or positrons requires the reconstruction of the $e^\pm$ candidates in the event, and in turn, since $e^\pm$ s in the event emit bremsstrahlung photons while traversing the tracker volume, this requires a reconstruction of the nonprompt photons present in the event. [12]

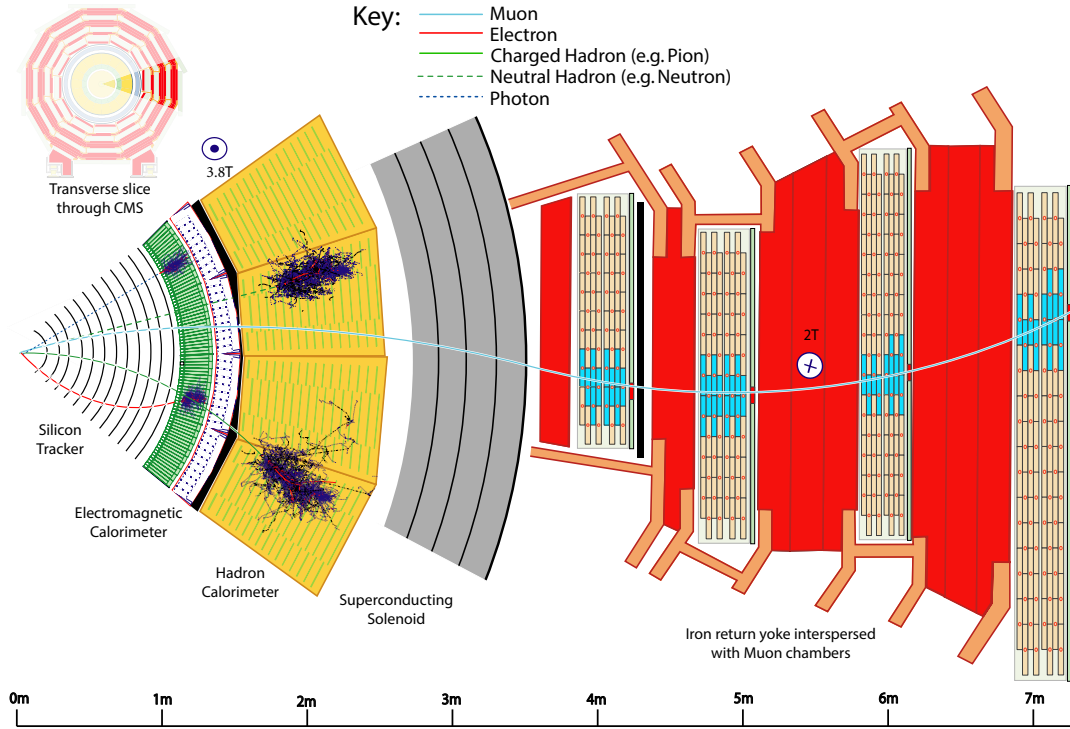


Figure 3.6: A sketch of various particle interactions in a wedge of the CMS detector cut through the transverse plane. The muon and pion are positively charged, while the electron is negatively charged. From [12].

In test beam studies, 120 GeV electrons impinging directly on the center of an ECAL crystal deposit about 97% of their energy into a 5×5 crystal array. However, in CMS, bremsstrahlung in the tracker on average cause between one third and four fifths of electron's energy to be radiated as photons, depending on the amount of intervening tracker material, which is smallest at $\eta = 0$ and largest at $|\eta| = 1.4$. The bremsstrahlung photons mainly spread the electron's energy deposits in the ECAL along the ϕ direction, due to the electron's bending in the solenoid's magnetic field. ECAL “seed clusters” are PF calorimeter clusters above 240 MeV in EB and 600 MeV in EB are extended into ECAL “superclusters”, which are formed using the dedicated “moustache algorithm”. The moustache algorithm groups together PF ECAL clusters that fall within a region of η, ϕ space within an envelope whose shape is dependent on the seed cluster's energy and η coordinate. This shape typically is 30 ECAL crystals long in the ϕ direction, with a width in the η direction that increases as the distance in ϕ from the seed cluster, with smaller widths for higher energy seed clusters; clusters falling within this envelope are not required to pass any lower threshold, but clusters that are too energetic with respect to their deflection from the seed cluster are rejected.

Tracks linked with superclusters are fit using the “Gaussian Sum Filter” (GSF) method [108],

which reconstructs the $e^\pm$'s track before and after all bremsstrahlung emissions inferred from secondary clusters in the supercluster. The reconstructed GSF tracks are required to pass basic quality requirements on the goodness of the track fit and the number of tracker hits. ECAL clusters inside the supercluster not matched to Kalman filter tracks are assumed to be bremsstrahlung photons emitted by the GSF electrons. After reconstructing the GSF tracks of electrons and their associated bremsstrahlung photons, the “converted bremsstrahlung finder” algorithm is used to identify $e^\pm$ tracks that were created by conversion of a bremsstrahlung photon. Tangents to the GSF tracks are calculated along the GSF track's trajectory at tracker layer, representing the trajectory of a potential bremsstrahlung photon, and Kalman Filter tracks that are consistent with originating from that tangent line are identified and linked to the PF block. ECAL clusters that are consistent with the converted bremsstrahlung tracks are then grouped with the supercluster in a “refined supercluster” which is linked to the PF block. Any tracks or clusters linked to the PF block that link to an HCAL cluster are discarded from the PF block. [109]

To distinguish electrons-positron pairs created by conversion of photons from those created in massive meson decays, nuclear interactions, or vertices from misreconstructed tracks, the fact that the photon is massless, and therefore conversion electrons are created with parallel momenta. Therefore, a cut on the angular variable $\Delta \cot(\theta) < 0.1$ is applied, and the point of closest approach of the two tracks is required to be positive (i.e. non-intersecting.) In addition, the point at which the two tracks are tangent is required to be within the tracker volume. Electron-positron pairs consistent with being produced by conversion of a prompt photon are stored as PF candidate photons, along with ECAL superclusters not linked with any tracker tracks. Prompt photons that are not converted and are absorbed in the ECAL typically deposit over 97% of their energy in a single 5×5 matrix of crystals. [109]

Final photon identification criteria must be applied before final use in analysis, but these criteria can vary between various CMS analyses. These final identification criteria are meant to reduce backgrounds, the dominant of which is photon candidates that arise from neutral mesons such as π^0 mesons, which decay into a pair of photons. A π^0 with $p^T > 15$ GeV produced in CMS that decays to $\gamma\gamma$ will be separated by a distance roughly the same as the width of the face of an ECAL crystal. Background such as these rely on variables that discriminate based on shower shape and the distribution of energy within the supercluster. Other variables are meant to correct for detector effects such as anomalous noise from the APDs. Variables used in confirming a good quality photon include:

R_9 This variable is the sum of the energies of the 3×3 crystals centered on the most energetic crystal in the supercluster divided by the energy of the supercluster. This variable is sharply peaked near .95 for photons that do not convert before reaching the ECAL.

$\sigma_{\eta\eta}$ This variable corresponds to the lateral extension of the photon shower, given by the energy weighted spread of the shower within the 5×5 crystal matrix centered on the crystal with the largest energy of all crystals in the supercluster. This variable is generally small compared to backgrounds.

- $q_{n\phi}$ This variable is the diagonal component of the covariance matrix centered on the same 5×5 matrix.
- s_4 This is the “swiss cross” variable which is defined as the energy in the highest energy crystal divided by the four sharing a side.
- I Various photon isolation variables are calculated by summing (separately) the momenta of other reconstructed objects in the event within a radius of δR of the reconstructed photon: charged hadrons, photons, neutral hadrons

ID criteria used by CMS include cuts on these variables, and also on an ID scheme based on a Multivariate Analysis (MVA) which is implemented as a Boosted Decision Tree (BDT) trained on MC samples.

3.2.3.5 Hadrons

After the reconstruction of electrons, photons, and muons has been performed, these are all removed from further consideration in the PF algorithm, and the remaining particles in the event to be identified are charged hadrons, including $\pi^\pm$, $K^\pm$, and protons, as well as neutral hadrons, mainly K_L^0 . Of the remaining PF blocks, those that contain tracks linked with HCAL and ECAL clusters are identified as charged hadrons, and their energy and momenta are initially determined via fits of the tracks, using the assumption that the mass of the charged hadron is the mass of a $\pi^\pm$. The sum of these initial track momenta is then compared with the sum of the calorimetric energies from the linked ECAL and HCAL blocks. If the energy estimates from the calorimeters agree within the calorimeters’ energy resolution for hadrons, no neutral particles are inferred to have been present within those calorimeter clusters; however, if the energy from the calorimeters is larger than the estimate from the sum of the tracks, the algorithm infers the presence of neutral particles leading to the calorimeter deposits. For an energy measurement from the calorimeters that is greater than the sum of the track momenta by 500 MeV and smaller than the total ECAL energy, the presence of a photon is inferred, with an energy corresponding to the excess. However, if the excess of calorimetric energy over the sum of track momenta is greater than 1 GeV and larger than the total ECAL energy measurement (indicating a large contribution to the excess by energy measured in the HCAL), a neutral hadron is identified in the event, with an energy equal to any remaining excess after the identification of any photons in that PF block. [12]

3.2.3.6 Jets

Jet clustering

Jets are reconstructed from PF candidates by clustering them with the anti- k_T [110], k_T [111], Cambridge-Aachen [112, 113], or cone [114] algorithms. All jet clustering algorithms are based around the idea of grouping particles together based on their initial momenta in η, ϕ space in order

to group together the particles that originated from the hadronization of a quark or a gluon into a jet, so as to give an estimate of the original momentum of the initial quark or gluon. Currently the most widely-used jet clustering algorithm, the anti- k_T algorithm can be understood as seeding jets from their most energetic constituents, and in a pairwise fashion, calculating a decision on whether nearby constituents should be clustered into the jet based on their separation ΔR and their energy.

The anti- k_T , k_T , and Cambridge-Aachen algorithms all define a d_{ij} for each pair of particles, labeled with the indices i, j for each particle, given by:

$$d_{ij} = \min(p_{T,i}^{2\alpha}, p_{T,j}^{2\alpha}) \frac{\Delta_{ij}^2}{R^2} \quad (3.7)$$

where $\Delta(i, j)$ is equal to the standard ΔR between two particles in η, ϕ space, R is a parameter which can be thought of as giving a scale to the width of the jet, and p is a parameter introduced to govern the relative power of the energy versus geometrical (Δ_{ij}) scales. In the k_T algorithm, $p = 1$, and the Cambridge-Aachen algorithm, $p = 0$, and $p = -1$ for the anti- k_T method. Subjects, sometimes also referred to as pseudojets, are then formed by finding the minimum d_{ij} of all particles in the event and combining their four-vectors to form a subjet. A subjet is only formed in this manner if d_{ij} is also less than $p_{T,i}^{2\alpha}$, otherwise the particle is not clustered into any jet and removed from the set of particles being clustered. If a subjet is formed, however, the two original constituents are replaced by the subjet formed from the sum of the two subjets momenta, and the algorithm proceeds to repeat the same calculation considering all particles and subjets. If a subjet is found to have a $p_{T,i}^{2\alpha}$ greater than all d_{ij} , then that subjet is declared to be a jet. Because a jet is formed in this pairwise procedure – by combining two particles, two subjets, or one particle and one subjet – every jet is always fundamentally comprised of two constituents. In this picture, even if one (or both) of the final two objects being clustered is a particle rather than a true subjets, that particle (or both particles) are nonetheless referred to as the subjets of the jet.

Compared to the other algorithms, the anti- k_T algorithm has several major advantages. One major advantage is known as “IRC safety”, which refers to the property that jets formed using the anti- k_T jet are robust against soft radiation, in that the final jets the algorithm produces are not drastically different from the final jets that would be produced from the same set of particles but with some additional low- p_T particles added. The algorithm is also robust against drastic changes to the final jets if a particle in the initial event is exchanged for two nearby collinear particles each sharing half the p_T or the original – this is known as “collinear safety.” [115] Jets clustered with a radius parameter R are commonly referred to by $R \times 10$ in their name; for example, an anti- k_T jet formed using $R = 0.8$ is referred to as an “AK8 jet.”

Jet energy corrections

In the CMS data reconstruction paradigm, the PF objects are first clustered as-is, without the application of more finely-tuned pileup mitigation or jet grooming algorithms; these are sometimes referred to as “raw jets”. This level of jets constitutes the starting point for more sophisticated

types of jets, but at the level of raw jets, there are various detector effects which require a first-pass calibration of the energy of these jets. This first-pass of calibrating the jet energy measurements is referred to as “Jet energy corrections” or JECs, which result in jets that are robust enough for many analyses that do not depend very sensitively on exact jet mass measurements, and which also form the starting point for more sophisticated algorithms used for precision measurements of quantities such as the invariant mass of a jet.

The first correction applied to raw jets is referred to as the “L1 Residual Offset Correction” which is intended to roughly correct the jets for the fact that the clustering algorithms inevitably cluster in pileup particles. Neutral hadrons which enter into jets cannot be unambiguously identified with the primary vertex, and in addition have a relatively poor energy resolution compared to e.g. charged hadrons and photons. Since the HCAL is a sampling calorimeter, it is inevitable that collision data must be used for an overall energy scale (referred to as the HCAL response corrections) and these corrections are derived by studying low-pileup events that feature a jet recoiling against an isolated charged track. This method gives an overall scale for the HCAL energy response, but does not mitigate the effect of neutral hadron pileup in events. In addition, the HCAL’s response to both neutral and charged hadrons is intrinsically nonlinear; i.e. the light deposited in the scintillators and read out by the photoelectronics does not scale linearly with the energy of the incident hadrons, and thus the HCAL response corrections can only calibrate the absolute response of the detector at a given reference energy, which is usually taken to be 50 GeV (for a charged pion). Therefore, the L1 JEC additionally fills the role of linearizing the energy reading from the HCAL such that they accurately provide an estimate of the neutral hadron energies. Therefore, to unfold the effects of the HCAL’s nonlinear response and to mitigate neutral hadron pileup in raw jets, the L1 jet energy correction is derived from studies that compare MC simulation of detector response to events in the data, taking into account the effects of neutral pileup, and corrects for these two effects.

The second level of JECs is referred to as the “L2 Residual Relative η Correction”, which is designed to ensure uniformity of the jet energy measurements as a function of jets’ η . Because of the nonuniformity of the detector, especially in the transition region between the barrel and the endcaps, difficulties in fully measuring and reconstructing jet energies arise, for example because of particles travelling through the small gap region between the ECAL barrel and endcaps. In addition to MC studies to help mitigate this effect, a data-based “tag-and-probe” method has been designed to correct mismeasurement of jets due to these detector nonuniformities. This method utilizes events that are characterized mainly by the presence of two jets plus another object in the event which leads to some asymmetry in the dijet system, such that the two jets lie in separate $|\eta|$ regions of the detector. The “tag” jet is required to have an $|\eta| < 1.3$, the region of the detector with the best absolute jet energy scale (thanks to robust tracking in this region), and the “probe” jet is left unconstrained. Two quantities are then studied in order to ensure the uniformity of the jet energy

scale as a function of η :

$$DB = \frac{p_{T,\text{probe}}}{p_{T,\text{tag}}}$$

$$MPF = 1 + \frac{\vec{p}_T^{\text{miss}} \cdot \vec{p}_{T,\text{tag}}}{p_{T,\text{probe}}^2} \quad (3.8)$$

where $\vec{p}_T^{\text{miss}}$ denotes the negative of transverse component of the vector sum of all PF candidates in the event, corresponding to missing transverse momentum that might arise from particles leaving the detector undetected, like e.g. a neutrino. *DB* stands for ‘dijet balance’ and *MPF* stands for ‘missing momentum fraction.’ *DB* corresponds to the degree to which the dijet system displays a mismatch between the tag and probe jets, which may indicate miscalibration of the jet energy measurements in the probe region. *MPF* is analogous, but takes into account the presence of undetected particles in the event, a larger *MPF* indicates that the missing transverse momentum in the event is not balanced against the jet activity in the event, which can again indicate miscalibration of jet energies in the event. These two quantities are studied in QCD multijet events where little neutrino activity is expected, and the two quantities are studied, and typically a minimization of some combination of the two is used to derive the L2 JECs. Note that in the denominators of Equation 3.8, in the case when it is possible that the tag jet is not ensured to be in a well-calibrated region, is often advantageous to replace $p_{T,\text{tag}}$ with another reference quantity, sometimes taken as the average of tag and probe p_T ’s.

The last commonly used level of JECs applied to raw jets is the “L3 Residual Absolute p_T Correction” which refines the linearization of absolute jet response with dedicated studies of several processes that also provide a tag-and-probe of jet energy response. The events used for these studies use final states which include, among others, $Z(e^+e^-) + j$, $Z(\mu^+\mu^-) + j$, $\gamma + j$, and multijet events. For example, in events that have a jet recoiling against a Z which decays according to $Z \rightarrow \mu^+\mu^-$, a precise measurement of the momenta of the two muons is possible thanks to the excellent resolution of the trackers and muon chambers. A p_T and η dependent correction to the measured jet responses then can be found by allowing a free jet correction parameter to be applied to the jet’s 4-vector and then minimizing the imbalance of the transverse momenta of the muons recoiling against the jet in a global fit of all such events and jet energy corrections. The various channels apply with varying degrees of efficacy in different energy and η regions, which means that the various channels act complementarily to cover a wide range of energies and in different regions of the detector, but share enough overlap to ensure the consistency of the derived JECs from each method.

Pileup mitigation

Due to the fact that many proton-proton collisions occur in each LHC bunch crossing, known as “in-time pile up” (where pile up is abbreviated as PU), it is inevitable, due to the nature of the jet clustering algorithms, that particles originating from many different collisions in the event can be clustered into a particular jet, and if one is interested in the physics of just one particular proton-proton collision in the event (which is certainly the case for most applications), this is certainly an

unwanted contamination of the jet. Various strategies have been developed to mitigate the effects of pileup and remove the unwanted jet constituents from jet objects used for physics analysis. The most advanced pileup mitigation strategy developed by the CMS collaboration as of the present is the “PileUp Per-Particle Identification” (PUPPI) algorithm. [116]

The PUPPI algorithm operates on PF candidates before any jet clustering algorithm is applied. The method applies a weight to the 4-vector of each particle according to an estimate of the likelihood that the particle originated from a PU vertex, where ideally a weight of 0 would be applied to particles originating from PU and a weight of 1 would be applied to particles originating from the primary vertex (PV). For each particle, a shape α_i is defined:

$$\alpha_i = \log \sum_{j \in \text{event}} \xi_{ij} \times \Theta(\Delta R_{ij} - R_{\min}) \times \Theta(R_0 - \Delta R_{ij}) \quad \text{where } \xi_{ij} = \frac{p_{T,j}}{\Delta R_{ij}} \quad (3.9)$$

where Θ represents the Heaviside step function, and R_{ij} is the common ΔR distance between particles labelled by indices i and j . The step functions define an annular region around each particle, such that only particles within a distance R_0 but beyond $R_{\min}$ enter the calculation of α , where typically $R_0 = 0.3$ and $R_{\min} = 0.02$. α is thus determined by only nearby particles, but does not consider very close particles, to ensure collinear safety of the algorithm. In general, α is observed to be smaller for particles from PU, for both charged and neutral PU. A weight w_i is then calculated for each neutral particle in the event according to the formula:

$$w_i = F_{\chi^2, \text{NDF}=1} \left(\Theta(\alpha_i - \bar{\alpha}_{\text{PU}}) \times \frac{(\alpha_i - \bar{\alpha}_{\text{PU}})^2}{\sigma_{\text{PU}}^2} \right) \quad (3.10)$$

where $\bar{\alpha}_{\text{PU}}$ and σ_{PU}^2 denote the mean and standard deviation, respectively, of the α distributions observed in each event for charged PU particles, and F_{χ^2} is the cumulative distribution function of the χ^2 distribution. Charged PU particles are then assigned a weight of 0, and neutral particles in the event with $w > w_{\min}$ assigned the weight given by the above equation. Particles with $w < w_{\min}$ are assigned a weight of 0. This algorithm thus uses event-by-event information to characterize the pileup in that event from measurements of the charged PU and uses that to reweight neutrals which may have originated in PU.

Jet grooming

In order to achieve more precise estimates of a jet’s invariant mass, a jet grooming algorithm known as “soft drop” (SD) [117] has recently been adopted by several CMS Collaboration analyses. The SD algorithm, which is an extension of a modified mass drop tagger [118], reclusters the jet, starting from its original constituents using the Cambridge-Aachen algorithm, and removes soft, wide-angle components of the jet at each step in the clustering process. The following SD grooming parameters are used: $z_{\text{cut}} = 0.1$ and $\beta = 0$, where the z_{cut} specifies trailing subjet p_T relative to the whole jet p_T at which the jet declustering into subjet pairs is stopped. The parameter β adds additional angular requirements on the jet declustering; for $\beta = 0$ these requirements are neglected. The groomed jet mass (M_J^{pruned}) is computed from the sum of the four-momenta of the remaining constituents.

An important feature for identifying W , Z , or Higgs bosons is by cutting (or otherwise discriminating) the candidate jets based on their invariant mass. In order to ensure that the PU mitigated and groomed jet masses are reliable and unbiased values, a comprehensive set of checks have been made to ensure proper calibration of the reconstructed jet masses. This is achieved by deriving an η and p_T dependent correction of the jet mass that adjusts for the systematic downward bias that these algorithms introduce – this downward biasing is due to the fact that much of the algorithms’ efficacy comes from their removing pileup particles, noise, or poorly reconstructed neutral hadrons from the event, but a certain degree of over-removal occurs.

The derivation of these mass corrections proceeds in two steps. First, using MC samples of hadronically decaying W bosons, a correction for PUPPI softdrop jets is derived by observing the effect of the application of the PUPPI and softdrop algorithms to generator-level particles. In other words, without having run any detector simulation, the clustering, PUPPI, and SD algorithms are run over the generator’s simulated objects, taking as a given the final state particles momenta directly from the PYTHIA hadronizer’s final state particles. A generator-level calculation of the W jets’ invariant masses m_{gen} is then found. A correction for m_{gen} is then derived as a function of the jets’ generator-level p_T , which can be thought of as restoring the jets’ generator-level masses to match the true W mass. The numerical values of this p_T -dependent correction are found by first binning the spectrum of true mass ($m_{\text{PDG}} = 80.4$ GeV) divided by the PUPPI softdrop invariant mass $\frac{m_{\text{PDG}}}{m_{\text{gen}}}$ in bins of the jets’ p_T . Then, a power-law fit is made to this spectrum, and the fit function is taken to give the generator-level correction c_{gen} as a function of p_T . This correction is found to be between 2% and 3%.

Once the generator-level correction is found, it is then applied to the reconstructed W jets of the same sample – i.e. the factor found from the fit of $\frac{m_{\text{PDG}}}{m_{\text{gen}}}$ is used as a weight to multiply the 4-momenta of the reconstructed jets that are found in a full simulation of event occurring in the detector. A second level correction is then found to restore the reconstructed-level W jets mass m_{reco} such that they more accurately match the true W jet. This reconstructed-level correction is found as a function of jet p_T separated into two regions: $|\eta| < 1.3$ and $|\eta| > 1.3$. To derive the values of this

correction for each region in η , a polynomial fit is performed to the spectrum of $(m_{\text{reco}} - m_{\text{gen}})/m_{\text{reco}}$ binned as a function of the jet p_T . These two fits then give rise to p_T -dependent corrections c_{reco} for both regions in η .

Finally, the corrected PUPPI-softdrop masses $M^{\text{PUPPI+SD}}$ used for both MC and data for reconstructed jets is given by:

$$M^{\text{PUPPI+SD}} = M_{\text{uncorr}}^{\text{PUPPI+SD}} \times c_{\text{gen}}(p_T) \times c_{\text{reco}}(p_T, \eta) \quad (3.11)$$

where $M_{\text{uncorr}}^{\text{PUPPI+SD}}$ is the uncorrected PUPPI softdrop jet's invariant mass. This correction is found to greatly improve the downward biasing of W , Z , and H jets' invariant masses, and reduces an issue in uncorrected jet masses whereby higher p_T boson jets' invariant masses tend to peak at values lower than the true mass of the bosons. [119]

Structure of jets

As discussed in Chapter 2, the W , Z , Higgs, and other interesting particles that come from the decay of a high-mass resonance – one of the holy grails of experimental particle physics – would be produced with a significant boost. Therefore, their decay products will be roughly collimated, and in the case of a hadronic decay of one of a boson, the decay products will end up as being consistent with one large-radius jet. These decays are quite dominant for these bosons: W bosons have a 30% branching fraction to decay to $c\bar{s}$, Z bosons similarly have greater than 30% probability to decay to $c\bar{c}$ or $b\bar{b}$, and Higgs bosons have a 58% chance to decay to $b\bar{b}$. These decay modes all lead to the constituent particles inside these large-radius jets being distributed in ways which can be used to distinguish these jets from ordinary QCD jets, which constitute the main background to detecting these bosons, and from one another. Tagging these bosons builds off the considerable contributions that have been made toward identifying the heavy-flavor jets on their own, and the techniques of tagging jets from boosted bosons are a natural extension to work that has happened in the worlds of charm tagging and b-tagging. The interesting difference in the substructure relative to heavy flavor jets is that there is a definite two-pronged structure to the distribution of particles in these jets, so the study of the structure of subjets inside these jets, along with the related idea of the τ axes (discussed later), is a burgeoning topic. An illustration of the basic structures of these fat jets is shown in Figure 3.7. The distinguishing features that can be seen in these different types of jets can be characterized in many ways, either through simple intuitive variables, through more advanced but traditional algorithms such as subjet reclustering and iterative approaches to identifying structures, or through advanced modern techniques such as boosted decision trees or the application of deep neural networks to combine various possible variables and for automated pattern recognition.

b-tagging

b quarks typically hadronize into B mesons, which have a considerable lifetime relative to various other heavy mesons. Therefore, the use of identifying secondary vertices within a jet is of crucial utility in identifying jets from b-quarks. The “Combined Secondary Vertex” (CSVv2) algorithm [120]

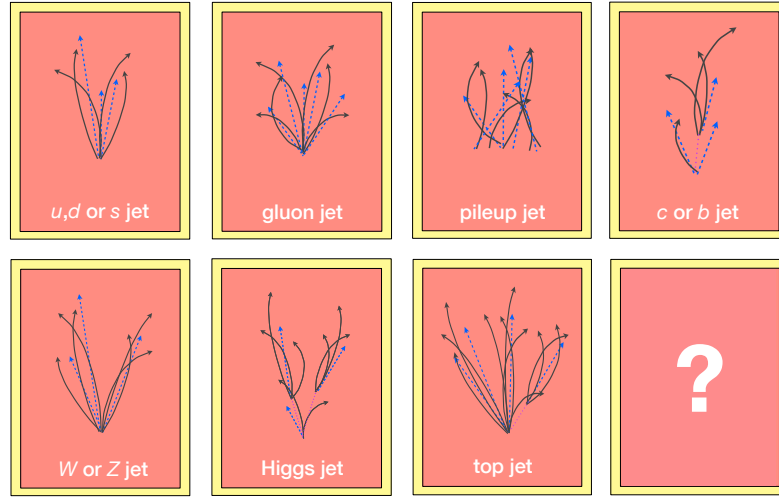


Figure 3.7: A cartoon of the various structures in different types of jets. Notice the two-prong structure in the hadronic decays of boosted bosons. Borrowed from [13].

and its successors, including the Deep CSV and Deep Flavor tagging algorithms, use a set of variables that characterize the presence of secondary vertices in jets.

In the CSVv2 algorithm, secondary vertices (SV) inside jets are reconstructed along the tracks consistent with originating at the primary vertex that pass several track quality requirements in an iterative manner that uses the adaptive vertex fit algorithm. [121] The algorithm then identifies each of the jet's tracks compatible with originating from that secondary vertex, removing those tracks from further consideration, and then iterates the procedure to find further secondary vertices. The CSV algorithm then calculates the 3D impact parameter (IP) of each track with respect to the PV, defined as the vector from the PV to the point of closest approach to the track extrapolated inward toward the PV. The significance of the impact parameter, ($S_{IP,i}$) is then calculated for each track's IP, defined as the magnitude of IP divided by the uncertainty on the magnitude of the IP, multiplied by a sign that corresponds to whether the IP vector's projection along the jet's momentum vector is has positive or negative magnitude. The IP values of jet tracks, the S_{IP} values, and several other variables of the jet including the vertex mass, the distance of flight to the SV in the transverse plane, the number of tracks in the jet, are then used to construct two likelihood ratios that are then combined to give a final discriminator value that can be used for b-tagging.

Characterizing jet substructure

In order to characterize whether a jet has a structure that that is consistent with it being the result of the hadronization of two main subcomponents, various discriminating variables have been proposed. The best well known of these is known as n-subjettiness [122]. This algorithm starts by first identifying N candidate subjets, which ideally would run some complete minimization over all

particles in a jet, but typically some clustering algorithm is rerun on the jet's particles. Typically, the algorithm used by CMS is the exclusive k_T clustering algorithm [123] with a requirement placed so that it returns exactly N jets.

Once the N subjets are identified, a variable τ_N is defined, which is designed to give a characterization of how likely the jet is composed of the distinct subjets identified:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min \{ \Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k} \} \quad (3.12)$$

where k is an index that runs over the constituent particles in the jet each with momentum $p_{T,k}$, $\Delta R_{J,k}$ is the usual distance ΔR in η, ϕ space between the k^{th} particle and the J^{th} identified subjet, and d_0 is a normalization factor:

$$d_0 = \sum_k p_{T,k} R_0 \quad (3.13)$$

where R_0 is the distance parameter of the original jet, e.g. 0.8 for an AK8 jet. Intuitively, this variable tends towards its minimum value, 0, when most of the particles in the event are aligned with the N subjet axes, and a τ_N that is larger indicates that a better fit for the jet would be that it is composed of a larger number of subjets (e.g. a large τ_N indicates that the jet is more likely composed of two or more subjets.) Comparing between jets that arise from gluons or light-flavor quarks (QCD jets) and jets that originate from vector or higgs boson hadronization, a general observation is that QCD jets, with their wide-angle, fairly diffusely spread radiation, can have a significant τ_1 , simply because of the unpredictable structure of these jets forming. However, it is observed that the QCD jets with high τ_1 are correlated with also having a high τ_2 . On the other hand, jets with a genuine two-prong structure tend to display a large τ_1 but do not gain a significant τ_2 , so often in the use of n -subjettiness to tag W s, Z s, or Higgses, one focuses on the variable $\frac{\tau_1}{\tau_2}$. The τ_N variables are generically useful for characterizing jets even beyond tagging bosons; for example, a boosted top quark can give rise to a jet that has a three-pronged structure. In this case, a useful variable is $\frac{\tau_3}{\tau_2}$. The entire set of τ_N variables can potentially be used in MVA or boosted decision tree-based discriminators.

Double-b tagging

For Higgs bosons decaying to $b\bar{b}$ that are produced with a boost corresponding to a $p_T > 250$ GeV, the b -quarks will be produced with momenta close enough in η, ϕ such that a single AK8 jet will cluster the decay products of both b -jets into one “fat jet.” This type of jet, with two characteristic subjets each a characteristic secondary vertex from a B meson, has been studied extensively and a dedicated “Double b -tagger” (DBT) has been developed by the CMS collaboration for its identification.

Many of the same variables as those used in the CSV algorithm are used, but several other important variables are also considered which are more specific to the two-subjet structure of a boosted Higgs jet. The DBT uses the Inclusive Vertex Finder algorithm to identify SVs, and the

SVs are each associated with the SV's track's closest τ axis, which is given by the axes of the last two subjets clustered into the AK8 jet by the anti- k_T algorithm. For each τ axis, the secondary vertex masses are all calculated, and the tracks that comprise all associated secondary vertices are summed to derive the subjet's p_T . Variables that are specifically used for the DBT are the S_{IPs} of each track but calculated with respect to its associated τ axis, rather than to the primary vertex, the ratios of the energies of the tracks associated with each τ axis to the total AK8 jet energy, the 2D flight distance for the SV with smallest uncertainty corresponding to each of the two τ axes, the total invariant mass of all tracks associated with each of the two τ axes, a variable that characterizes the two-secondary vertex system taking into account the kinematic properties that help distinguish $b\bar{b}$ background from gluon splitting, and the z variable, given by:

$$z = \Delta R(SV_0, SV_1) \frac{p_T(SV_1)}{m(SV_0, SV_1)} \quad (3.14)$$

where SV_0 and SV_1 are the secondary vertices with the smallest 3D flight distance uncertainty associated with the two τ axes, and $m(SV_0, SV_1)$ is the invariant mass corresponding to the summed four-momenta of the two secondary vertices.

This algorithm uses machine learning techniques, and the discriminator variable from the DBT is calculated using a neural network trained and validated using MC samples and b-enriched samples from the data. This new algorithm is implemented and optimized using the TMVA package inside the ROOT framework. [124]

Chapter 4

Search for $H\gamma$ resonances

4.1 Introduction

As discussed in Section 2.4, this analysis probes the data collected by the CMS detector for evidence of massive, narrow $H\gamma$ resonances, where the Higgs boson is produced with a large boost. As the Higgs decays to $b\bar{b}$ with a branching ratio of 58%, the telltale final state of such a resonance will be a photon plus a large-radius jet (“fat jet”), and that jet will display the characteristic features described in 3.2.3.6. Production of a narrow-width resonance of this type will lead to a peak in the invariant mass spectrum of the photon plus jet system ($J\gamma$), or a “bump”. In order to scan the spectrum for the characteristic bump, a shape-based approach is used, whereby a smooth fit to the spectrum is made to estimate the expected background, and it is compared with a fit that represents the background plus a narrow bump from the signal, and the signal significance is extracted from this signal plus background ($S+B$) fit. The invariant mass of the fat jet is required to be consistent with the mass of the Higgs boson (it is required to be in a “window” of masses close to the Higgs mass), and in order to obtain the best possible invariant mass measurement of the jet, pileup mitigation and grooming algorithms are used. In order to reduce nonresonant backgrounds that result in a $J\gamma$ final state, b-tagging methods are applied. Because of the use of b-tagging, the sensitivity of this analysis, as compared with analyses that have been reinterpreted in the context of $H\gamma$, is expected to be at least one order of magnitude greater.

While many aspects of the analysis follow strategies similar to that of the other diboson analyses, significant differences in the $H\gamma$ final state, such as a lower background level and the higher branching ratio of the Higgs boson to $b\bar{b}$, necessitate a few innovations relative to previous work. Comparing to, e.g., the $Z(q\bar{q})\gamma$ analysis [42], some important factors lead to the need for modifications of the search strategy: the invariant mass window for jets that are consistent with a Z decay must obviously be changed to that for a Higgs decay, and it is found that the newer double b-tagging algorithm (designed specifically for jets arising from nearby pairs of b quarks) is more effective for identifying the merged $b\bar{b}$ jets from the Higgs decay compared to the b-tagging algorithm used in

the $Z(qq)\gamma$ analysis. Furthermore, it is observed that the 50-70 GeV jet mass sideband employed for optimizing the event selection criteria in the $Z\gamma$ analysis does not translate well for $H\gamma$, as it does not reproduce the shapes of many kinematic variables in the Higgs jet mass window. However, thanks to the up-to-then unprecedented integrated luminosity collected by the CMS detector during the 2016 datataking campaign, a 100-110 GeV jet mass sideband can be used. The $H\gamma$ analysis therefore uses this sideband for optimizing the choice of variables and cuts, which is validated with cross-checks against Monte Carlo (MC) background samples.

4.2 Samples

The data used for this analysis was collected by the CMS detector during the 2016B-2016H proton-proton datataking periods. The CMS collaboration’s Data Quality Monitoring and Data Certification teams thoroughly validate the performance of the detector as well as the integrity and quality of the data both in real-time during the collection of the data, as well as in offline checks of the collected data. The checks on the data are performed with a granularity of “lumi sections”, which correspond to ~ 30 s periods of data collection, and all data in each lumi section must be of good quality in order to be certified. [125] All of the data certified in this manner is used in the analysis presented in this dissertation; in all, the sample corresponds to 35.9 fb^{-1} of p-p collisions data with the solenoid delivering a magnetic field of 3.8T and the LHC colliding with 25 ns bunch-crossing spacing.

Table 4.1: Data samples used in this analysis. Listed here are the run numbers, which refer to individual runs of the CMS DAQ system, and the dates of the “re-reco” or subsequent offline rerunning of the CMS reconstruction algorithms.

PD	runII era	campaign	run range
SinglePhoton	B	Feb. 3 re-reco v2	272007-275376
	C	Feb. 3 re-reco	275657-276283
	D	Feb. 3 re-reco	276315-276811
	E	Feb. 3 re-reco	276831-277420
	F	Feb. 3 re-reco	277772-278808
	G	Feb. 3 re-reco	278820-280385
	H	Feb. 3 re-reco v2, v3	280919-284044

As will be explained later in this dissertation, simulated standard model background Monte Carlo (MC) samples were used for various cross-checks of the analysis methods; in particular, since the background estimate is derived from data, MC background samples were used as a cross-check of the background estimation methods. These background samples were produced in the Summer 16 campaign of the CMS MC sample generation, and the cross sections corresponding to these SM processes were validated by the Exotica (EXO) physics analysis group. The NNPDF3.0 [126] parton distribution functions (PDFs) are used for all simulated samples. The fragmentation and hadronization are described with PYTHIA version 8.205 [127] using the underlying event tune CUETP8M1 [128].

The CMS detector response is simulated using GEANT4 [129]. All simulated samples include effects of pileup by superimposing on hard scattering events simulated minimum bias collisions, with their multiplicity matching that observed in data.

Table 4.2: SM MC background samples used. In the table, the symbol H_T denotes the scalar sum of transverse momenta of all jets in the events.

sample description	N_{events}	$\sigma \times k^{NLO/LO}$ (pb)
γ +jets, $40 \text{ GeV} < H_T < 100 \text{ GeV}$	4816232	20790 ± 9
γ +jets, $100 \text{ GeV} < H_T < 200 \text{ GeV}$	5015403	9238 ± 4
γ +jets, $200 \text{ GeV} < H_T < 400 \text{ GeV}$	10424189	2305 ± 0.7
γ +jets, $400 \text{ GeV} < H_T < 600 \text{ GeV}$	2476770	274.4 ± 0.2
γ +jets, $600 \text{ GeV} < H_T < \infty \text{ GeV}$	2550765	93.46 ± 0.06
QCD multijets, $100 \text{ GeV} < H_T < 200 \text{ GeV}$	81637494	27990000 ± 4073
QCD multijets, $200 \text{ GeV} < H_T < 300 \text{ GeV}$	18718905	1712000 ± 376.3
QCD multijets, $300 \text{ GeV} < H_T < 500 \text{ GeV}$	19826197	347700 ± 74.81
QCD multijets, $500 \text{ GeV} < H_T < 700 \text{ GeV}$	19664159	32100 ± 7
QCD multijets, $700 \text{ GeV} < H_T < 1000 \text{ GeV}$	15356448	6831 ± 1.7
QCD multijets, $1000 \text{ GeV} < H_T < 1500 \text{ GeV}$	4963895	1207 ± 0.5
QCD multijets, $1500 \text{ GeV} < H_T < 2000 \text{ GeV}$	3868886	119.9 ± 0.06
QCD multijets, $2000 \text{ GeV} < H_T < \infty \text{ GeV}$	1912529	25.24 ± 0.02
W(qq)+jets, $600 \text{ GeV} < H_T < \infty \text{ GeV}$	1008034	$95.14 \pm 3.57 \times 1.21$
W(qq)+jets, $H_T = 180 \text{ GeV}$		2788×1.21
Z(qq)+jets, $600 \text{ GeV} < H_T < \infty \text{ GeV}$	855865	5.67×1.23
Drell-Yan + Jets $H_T=180$	12813158	1187×1.23

Signal samples were also produced by the CMS EXO MC group, using LHE files adapted from originals created by the authors of Ref. [6] using FeynRules 2.0 [130] and MADGRAPH aMC@NLO, followed by processing with the same PDFs, hadronizer, and detector simulation as those used for the SM background samples listed above. The $Z' \rightarrow H\gamma$ resonance masses generated with this full simulation chain were 650, 750, 850, 1000, 1150, 1300, 1450, 1600, 1750, 1900, 2050, 2450, 2850, and 3250 GeV.

4.3 Triggers

Events used in this analysis were required to fire either of following HLT triggers:

- HLT_Photon175_v*: requires there to be at least one HLT-level photon with $p_T > 175 \text{ GeV}$.
- HLT_Photon165_HE10_v*: requires there to be at least one HLT-level photon with $p_T > 165 \text{ GeV}$ with the additional requirement that the photon's supercluster energy in the ECAL be greater than ten times any linked HCAL cluster energy.

where the v^* is a version number depending on the version of the trigger deployed during a given data-taking era.

As demonstrated in Reference [131], MC simulation accurately describes trigger behavior starting at $p_T^\gamma > 180$ GeV, so in studies using MC samples, no scale factor is applied to correct the trigger behavior in the MC.

To check the trigger efficiency, the efficiency of this logical combination of triggers was studied both in the data and in the MC signal samples. In each check, we calculate the efficiency as the number of events with a reconstructed photon that fire the trigger divided by the number of events with a reconstructed photons. This efficiency is calculated for events binned in photon p_T and in the photon+jet invariant mass $m_{J\gamma}$. The jet and photon objects used for these calculation are the same as those described in Section 4.4. For the calculation of efficiency versus photon+jet invariant mass, the cuts applied on the events considered in this calculation are the full selection cuts except that no b-tagging categorization is performed. For the calculation of efficiency versus photon p_T , the preselection-level cuts are applied, except that the cut on photon p_T is relaxed to $p_T^\gamma > 150$ GeV, so as to demonstrate the turn-on. This calculation is performed using a combination of all the MC signal samples, and also using a 2016 data sample, where the data chosen is required to fire any of the triggers of the SingleMuon primary dataset. The results are shown in Figure 4.1.

The preselection cuts are:

- one reconstructed jet and one reconstructed photon are required to be present in the event
- $p_T^\gamma > 200$ GeV, $|\eta_\gamma| < 2.4$ (excluding ECAL gap)
- photon MVA ID > 0.2 (0.9 signal efficiency WP), $e^\pm$ veto
- $p_T^J > 250$ GeV, $|\eta_J| < 2.6$
- tight jet ID + tight lepton veto
- $m_J^{\text{PUPPI+SD}} > 30\text{GeV}$
- $\Delta R(J, \gamma) > 1.1$

The full selection cuts add further requirements beyond the preselection cuts:

- $|\eta_\gamma| < 1.4442$
- $|\eta_J| < 2.2$, $110 < m_J^{\text{PUPPI+SD}} < 140\text{GeV}$
- $p_T^\gamma/m_{J\gamma} > 0.35$

These cuts are discussed thoroughly in Section 4.5 and summarized in Table 4.3.

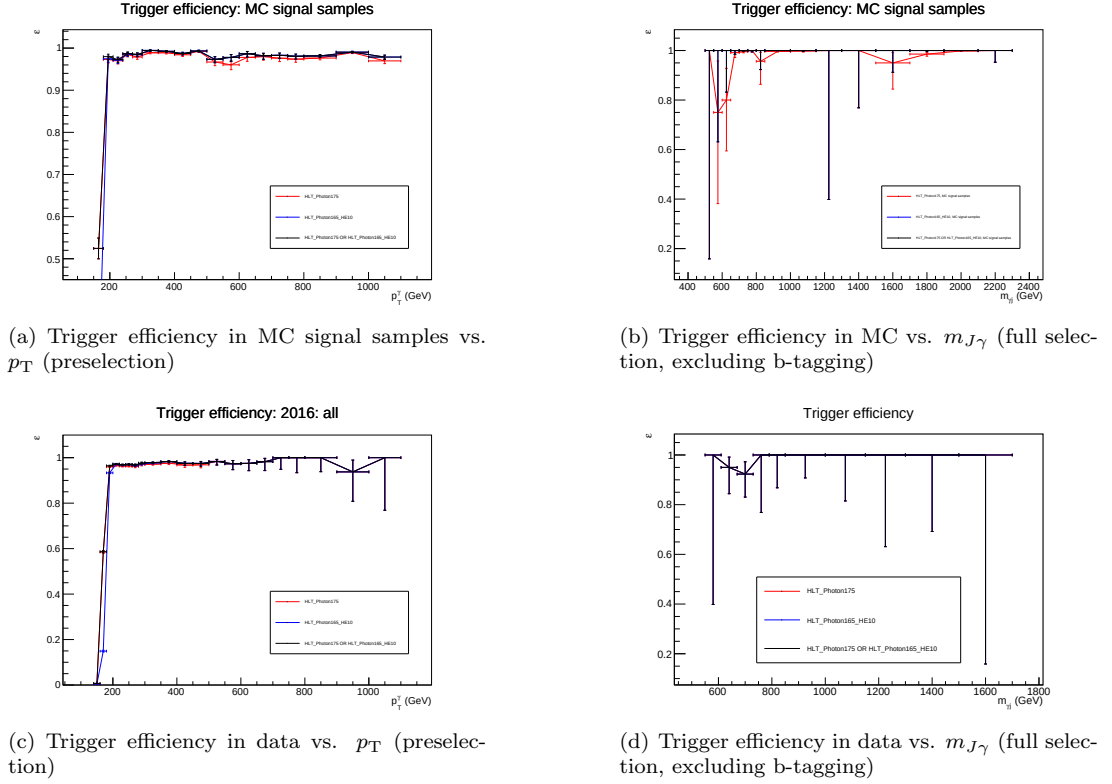


Figure 4.1: Trigger efficiency in MC and data. For the data, a logical OR of all SingleMuon dataset triggers is used as the reference trigger. For trigger efficiency vs. p_T , the check is done at preselection level, showing good efficiency but with some residual effects for the lower p_T bins. For trigger efficiency vs. $m_{J\gamma}$, we see some residual turn-on in the invariant mass spectrum up to around 750 GeV.

4.4 Object selection

4.4.1 Photons

Several alternative photon identification criteria were considered, including the loose, medium, and tight cut-based photon IDs as recommended by the Egamma POG [132], as well as the MVA photon ID at the loose and tight working points. As the dominant background in this analysis is from standard model photon+jet events, where a real photon is present, there will be a negligible effect from events that have a reconstructed photon from a jet or other non-photon object in the event. Therefore, the photon ID criteria are chosen to maximize the photon efficiency. Based on the efficiencies of the considered ID criteria, the chosen photon ID was the MVA ID at the 90% efficiency working point [133].

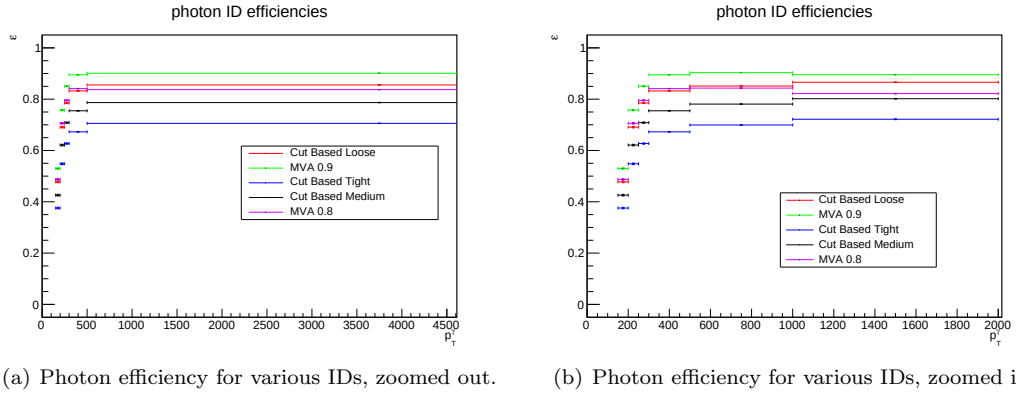


Figure 4.2: Efficiencies for various photon identification criteria, as a function of photon p_T , in the ECAL barrel region ($|\eta| < 1.4442$). The left plot and right plot are the same except for a different range and binning. Using the $H\gamma$ signal samples, the efficiencies shown are calculated as a ratio: the numerator of the ratio is the number of events with reconstructed photons passing the ID criteria, requiring that the reconstructed photon have $|\eta| < 1.4442$ and also be matched with a generator-level photon by requiring $\Delta R < 0.3$ relative to a generated photon. The denominator of the ratio is the number of events with a generated photon in the EB region.

4.4.2 Jets

Jets are clustered using the anti- k_T algorithm with a radius parameter of 0.8, and the PUPPI pileup mitigation algorithm [116] is applied to mitigate the effects of pileup in the event. Further, the soft drop jet grooming algorithm [117] is applied to sharpen the mass peak for jets originating from boosted higgs bosons, and to reduce the QCD background in the higgs mass region. The following jet identification criteria, corresponding to the tight Jet ID [134] with the tight lepton veto requirement, are applied:

- constituent multiplicity > 1

- neutral hadron energy fraction < 0.9
- charged hadron energy fraction > 0
- charged hadron multiplicity > 0
- neutral electromagnetic energy fraction < 0.9
- charged electromagnetic energy fraction < 0.9
- muon energy fraction < 0.8

To correct for detector effects, the effects of pileup, and the residual dependence on the p_T and η of jets, jet energy corrections are applied to each jet in the event [135]. The CMS collaboration has a dedicated “subgroup” of researchers, called the “Jet Energy and Resolution” subgroup, who study the CMS jet reconstruction algorithm, and they provide jet energy corrections officially for use by the entire CMS collaboration that are updated every few months in order to provide the optimal corrections corresponding to the varying detector properties (e.g. scintillator aging in the calorimeters), which are given version names for each update. The following jet energy corrections were applied to the data, following the recommendation of the JERC subgroup:

- Summer16_23Sep2016BCDV4_DATA
- Summer16_23Sep2016EFV4_DATA
- Summer16_23Sep2016GV4_DATA
- Summer16_23Sep2016HV4_DATA

For all Monte Carlo samples, the 80X_mcRun2_asymptotic_2016_TracheIV_v8 jet energy corrections were applied. In addition, photon p_T -dependent k-factors were applied for the photon+jets MC background samples; these k-factors are 1.8 for $100 \text{ GeV} < p_T^\gamma < 400 \text{ GeV}$, 1.4 for $400 \text{ GeV} < p_T^\gamma < 600 \text{ GeV}$, and 1.0 for $p_T^\gamma > 600 \text{ GeV}$ [136].

Several studies were performed to compare the performance of tagging jets in signal events originating from the boosted Higgs boson, using either PUPPI softdrop (SD) jets or jets groomed with an alternative set of algorithms: charged hadron subtracted (CHS) pruned jets [137]. Charged hadron-subtracted pruned jets led to a lower signal efficiency, shown in Figure 4.4, and the peak in the jet invariant mass spectrum displayed a downward shift away from the nominal Higgs mass as the signal mass grew for pruned jets, as can be seen in Figure 4.3. Because this analysis estimates the background using a fit of the signal region, it is important that the grooming algorithm chosen does not introduce any unwanted structure in the Higgs mass window; e.g., a grooming algorithm that artificially leads to a bump in the jet mass spectrum in the signal region would reduce the analysis’s sensitivity due to weaker background rejection. Figure 4.5 shows that the shape of the jet mass spectrum is smooth for both grooming algorithms. Considering its better signal efficiency and the Higgs mass peak’s stability as a function of signal mass, and noting that the jet mass spectrum

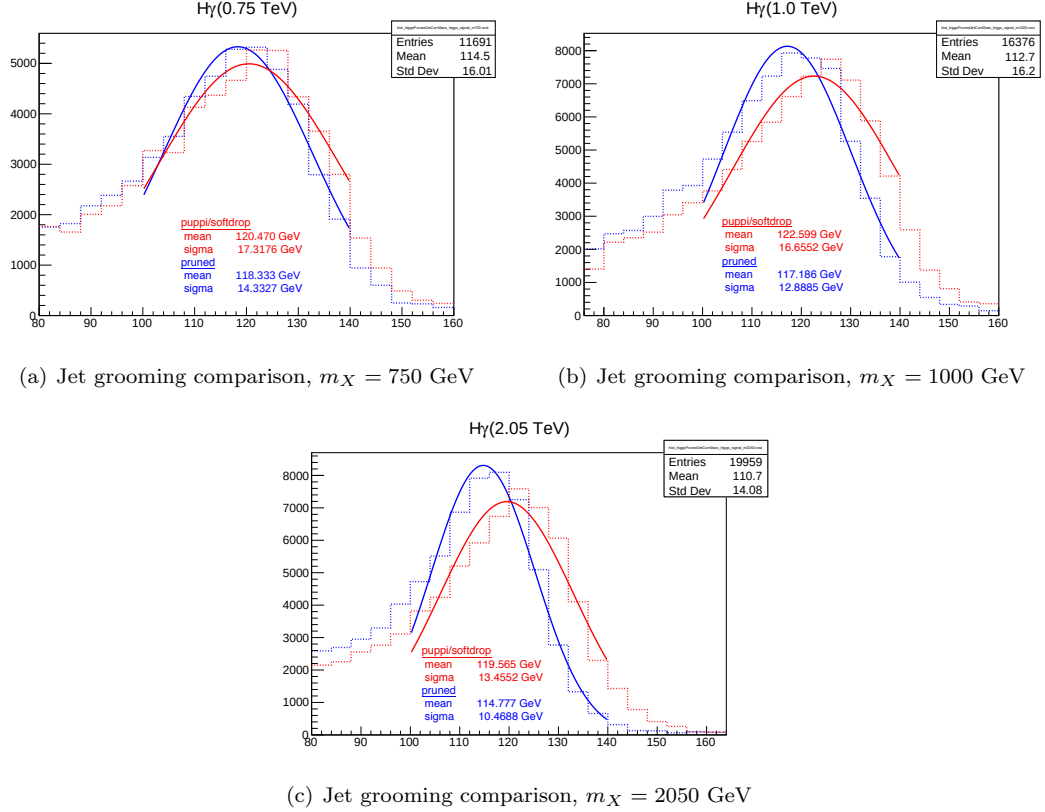


Figure 4.3: A comparison of the pruned jet mass and PUPPI softdrop jet mass for several signal samples. It is observed that as the signal mass increases, the pruned jet mass corresponding to the Higgs peak drifts downward, while the PUPPI softdrop mass is more stable. This study is performed at the full selection level but without the b-tagging categories applied.

for MC backgrounds is smooth for both algorithms, PUPPI softdrop jets were chosen in favor of charged hadron-subtracted pruned jets .

4.4.3 MC background validation

As explained in the next section, in order to maximize sensitivity to an $H\gamma$ resonance, cuts are applied on several variables, where the cut values are determined in a data-driven manner by using the 100 GeV-110 GeV PUPPI softdrop jet mass sideband to give an estimate of the background distribution, and the MC signal samples in the 110-140 GeV PUPPI softdrop mass signal region to give an estimate of the signal distributions. To check that the sideband region reproduces the expected background in the signal region reasonably well, it is useful to compare to the expected backgrounds predicted by background Monte Carlo samples. Therefore, it is important that the MC background samples describe the data well even at the preselection level of cuts. Figures 4.6 and 4.7, which use preselection cuts, demonstrate that the spectra of the main variables of interest

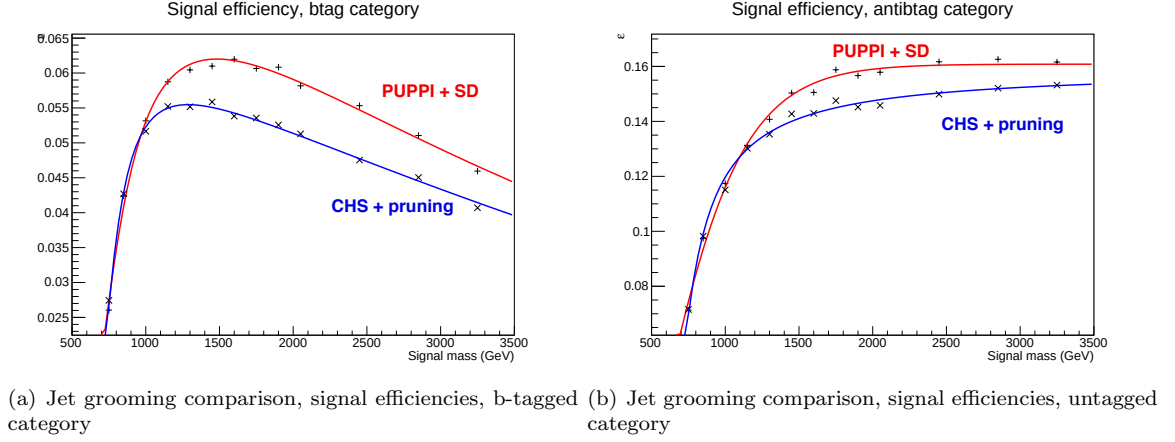
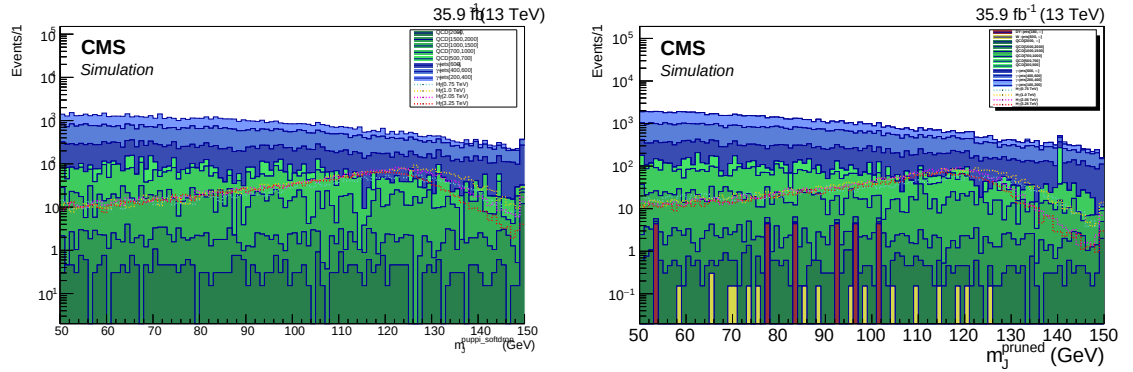


Figure 4.4: A comparison of signal efficiencies as a function of signal mass for the PUPPI softdrop jets and for CHS pruned jets. The spectra shown are made using the simulated signal samples, with the full selection applied in both categories. The PUPPI softdrop algorithm generally yields better signal efficiency.



(a) Jet mass spectrum near the signal region, PUPPI softdrop jets (b) Jet mass spectrum near the signal region, CHS+pruned jets

Figure 4.5: A comparison of the shapes of the jet mass spectra near the signal region. For both jet grooming algorithms, no unusual structure is observed near the signal region.

used in the analysis are well reproduced by the MC background samples.

The cuts at the preselection level and at the level of full selection are shown in Table 4.3.

4.5 Optimization

Cuts on relevant variables were optimized individually and iteratively by releasing one variable's cut at a time, and then finding its cut value yielding the highest sensitivity with all other cuts held constant. Distributions with these “N-1” cuts are shown along with signal sensitivity test statistics corresponding to a sliding value for a possible cut choice. Cuts were chosen to optimize the test statistics shown. As the signal samples are generated without a theoretical model for their cross section, and thus the test statistics are shown with arbitrary units (as they depend on the arbitrary normalization chosen on the number of signal events). Blinding is applied from this point onward.

A cut on $\Delta R(\gamma, J)$ was chosen to ensure isolation between the photon and jet, where the jet typically has a radius of $R = 0.8$ while the photon isolation cone is $R = 0.3$; therefore, to avoid any ambiguity or misreconstruction that can occur if the photon and jet are overlapping, the cut chosen was $\Delta R(\gamma, J) > 1.1$. Before applying a cut on a b-tagging discriminant, photon and jet eta cuts were chosen by observing the test statistic $S/\sqrt{B}$ as a function of their cut values, as shown in Figure 4.8.

To verify that the cuts applied increase the sensitivity of the analysis and are optimized to yield the best sensitivity, a $S/\sqrt{B}$ test statistic is used to determine the optimal cut value for various different signal masses. For given signal mass, S or B is calculated as the number of signal or background events, respectively, passing a given set of cuts, with the added requirement that the photon-jet invariant mass of an event be consistent with that expected for a signal event. In order to impose this requirement, we place a cut on events that enter into the calculation of S or B such that for a given signal mass, the photon-jet invariant mass must be within a window centered on the generator-level invariant mass of that signal. The width of this photon-jet invariant mass window is chosen to extend approximately 2σ above and below the signal sample's generator-level mass, where σ is the RMS of the photon-jet invariant mass distribution for that signal sample.

S is calculated by placing the “signal-region” cut on the PUPPI softdrop jet mass (110-140 GeV, corresponding to the Higgs mass window), and is equal to the observed number of signal MC events passing all cuts, with the further requirement that the photon-jet invariant mass be within the 2σ window described in the last paragraph. B is calculated in the same way with one exception: the PUPPI softdrop jet mass cut is chosen to be outside the signal region in a nearby sideband ($100 \text{ GeV} < m_{J\gamma} < 110 \text{ GeV}$). This permits a data-driven approach to estimate the background: the background estimate B is determined from the data, using this sideband.

This 100-110 GeV PUPPI softdrop jet mass sideband was chosen because the spectra of various kinematical variables from this sideband are quite similar in shape to the spectra from MC background samples in the signal region. In addition, use of the 100-110 GeV jet mass sideband avoids effects due to the presence of Z or W boson resonances. Other sidebands more distant from

the signal region, where signal contamination is not expected and contamination from the W and Z boson resonances is not expected, were investigated, and the optimization procedure was carried out on these alternate sidebands to cross-check the results from the 100-110 GeV sideband region-based optimization procedure. These alternate sidebands yield similar conclusions to that of the 100-110 primary sideband. A study of the 100-110 GeV and 50-70 GeV sidebands, justification for the primary use of the 100-110 GeV jet PUPPI softdrop mass sideband, is shown in Section 4.10.

Loosening the selection requirements on photon and jet pseudorapidity by extending into the forward region was observed not to increase the analysis’s sensitivity to signal: as demonstrated in Figure 4.8, sensitivity was level beyond cut values of $|\eta_\gamma| < 1.4442$, the edge of the ECAL barrel region, and $|\eta_J| < 2.2$, avoiding jet leakage beyond the coverage of the HCAL endcap region.

As the boosted Higgs decays to $b\bar{b}$ with a branching ratio of 0.57, b-tagging algorithms were investigated for maximization of background rejection and signal sensitivity. The b-tagging algorithms investigated were the double b-tagging algorithm [138] (DBT), and the double subjet CSV b-tagging algorithm [120]. The “loose” working point of the double-b tagger corresponds to a tagging efficiency of 79% and a mistagging probability of 10%, the “medium” working point has a tagging efficiency 68% and a mistag probability of 5%, and the “tight” working point has a 36% tagging efficiency and a mistag probability 1%. The CSV b-tagging algorithm uses similar techniques for identifying B hadrons inside subjets, but the algorithm considers only one (sub)jet – to tag fat jets originating from a $H \rightarrow b\bar{b}$ decay, one must run the CSV algorithm on every subjet within the fat jet, and a (possibly different) working point is applied on each of the two candidate b subjets.

Scale factors were applied for these b-tagging events to account for residual discrepancies between the b-tagging efficiency in data and in simulation, where the scale factor for the DBT was derived in Ref. [139] using a sample of high p_T jets enriched in $b\bar{b}$ from gluon splitting, and the double-subjet tagging follows the method of Ref. [131]. The scale factors applied are derived centrally by the b-tagging POG and can be found in Ref. [140].

As signal sensitivity in the low mass regions, which have higher backgrounds, (up to about $m_X \approx 1$ TeV) is expected to come from the b-tagged category, the double-b tagging working point was chosen by observing the $S/\sqrt{B}$ test statistic, due to the sufficiently large backgrounds in this regime (see Figure 4.9.) At higher resonance masses, where backgrounds are lower and b-tagging efficiency drops, b-tagging working points are optimized by maximizing the CL95 expected limit in a mass window of $\pm 5\%$ in the region around the resonance mass.

The double-b tagger at its tight working point, with a working point for the discriminator value of 0.9, showed the best performance, and comparisons with double-subjet working points showed better performance versus any combination of the tight, medium, loose, and untagged working points of the CSVv2 tagger, as shown in Figure 4.10.

In addition to applying a selection on the double-b tag discriminator, cuts on additional kinematic variables were investigated, in particular cuts on $\cos(\theta^*)$ as well as on the related variable $p_T^\gamma/m_{J\gamma}$. As these variables are mathematically isomorphic to one another, the cut is applied on only one of these two variables. The results of this study are shown in Figure 4.11.

To maximize sensitivity, a cut on $p_T^\gamma/m_{J\gamma} > 0.35$ is applied. Table 4.3 summarizes the cuts used in this analysis, and Table 4.4 shows the number of events in the 2016 data passing each successive cut.

Table 4.3: Final selection requirements in the analysis.

Preselection: at least one γ and one AK8 Jet	
Trigger	HLT_Pho175_* OR HLT_Pho165_HE10_*
Photon	$p_T^\gamma > 200$ GeV, $ \eta_\gamma < 2.4$ (excluding ECAL gap) MVA ID > 0.2 (0.9 signal efficiency WP), $e^\pm$ veto
Jet	$p_T^J > 250$ GeV, $ \eta_J < 2.6$ tight jet ID + tight lepton veto PUPPI softdrop groomed jet mass $m_J^{\text{PUPPI+SD}} > 30\text{GeV}$
$\Delta R(J, \gamma) > 1.1$	
Full selection	
Photon	$ \eta < 1.4442$
Jet	$ \eta < 2.2$, $110 < m_J^{\text{PUPPI+SD}} < 140\text{GeV}$
$p_T^\gamma/m_{J\gamma} > 0.35$	
b-tagged category: double-b tagger discriminant > 0.9	
untagged category: double-b tagger discriminant < 0.9	
Search region	$m_{J\gamma} > 720$ GeV

Because the 100-110 GeV jet PUPPI softdrop mass sideband used for optimization studies is very close to the 110-140 GeV jet mass signal region, it is possible that signal contamination is present in this sideband, as can be observed in Figure 4.3. Such signal contamination in the sideband could bias the choice of the cuts that were chosen on the basis of maximizing signal sensitivity. The two cuts that were chosen on the basis of maximizing sensitivity, i.e., the cut on the double b-tagging working point and the cut on $p_T^\gamma/m_{J\gamma}$, were therefore double checked using the 50-70 GeV jet PUPPI softdrop mass window, which is free from signal contamination. The results of this study are shown in Figure 4.12 – it is observed that the optimal cut choices are unchanged if we use this distant sideband where signal contamination is minimal.

Table 4.4: Cut flow, data events

Preselection	337836
Higgs window	30558
Search region	19462
η_J	19048
η_γ	12910
$p_T/m_{J\gamma}$	8878
b-tagged category	186
untagged category	8692

The n-subjettiness variables τ_1 and τ_2 , and, in particular, their ratio τ_{21} are variables commonly employed for tagging jets deriving from W bosons (and in a few analyses, Z bosons). This motivates investigating their use in this analysis as an additional categorization, to obtain the highest signal sensitivity. Shown in Figure 4.13 is a “N minus one” plot created in the same way as those shown above, with all cuts applied except for the b-tagging categorization cuts. A modest improvement in the $S/\sqrt{B}$ test statistic is apparent only at the lowest signal mass points, where the improvement in signal sensitivity would be roughly ten percent if no consideration is made of any additional systematic uncertainties. However, systematic uncertainties associated with cuts on the τ_{21} variable at a value of .35 are between 12 and 18 percent [141], with increasing uncertainties as the cut value is made smaller. Therefore, applying a cut on τ_{21} is not expected to improve the sensitivity of this analysis. Thus no cut on these n-subjettiness variables is applied.

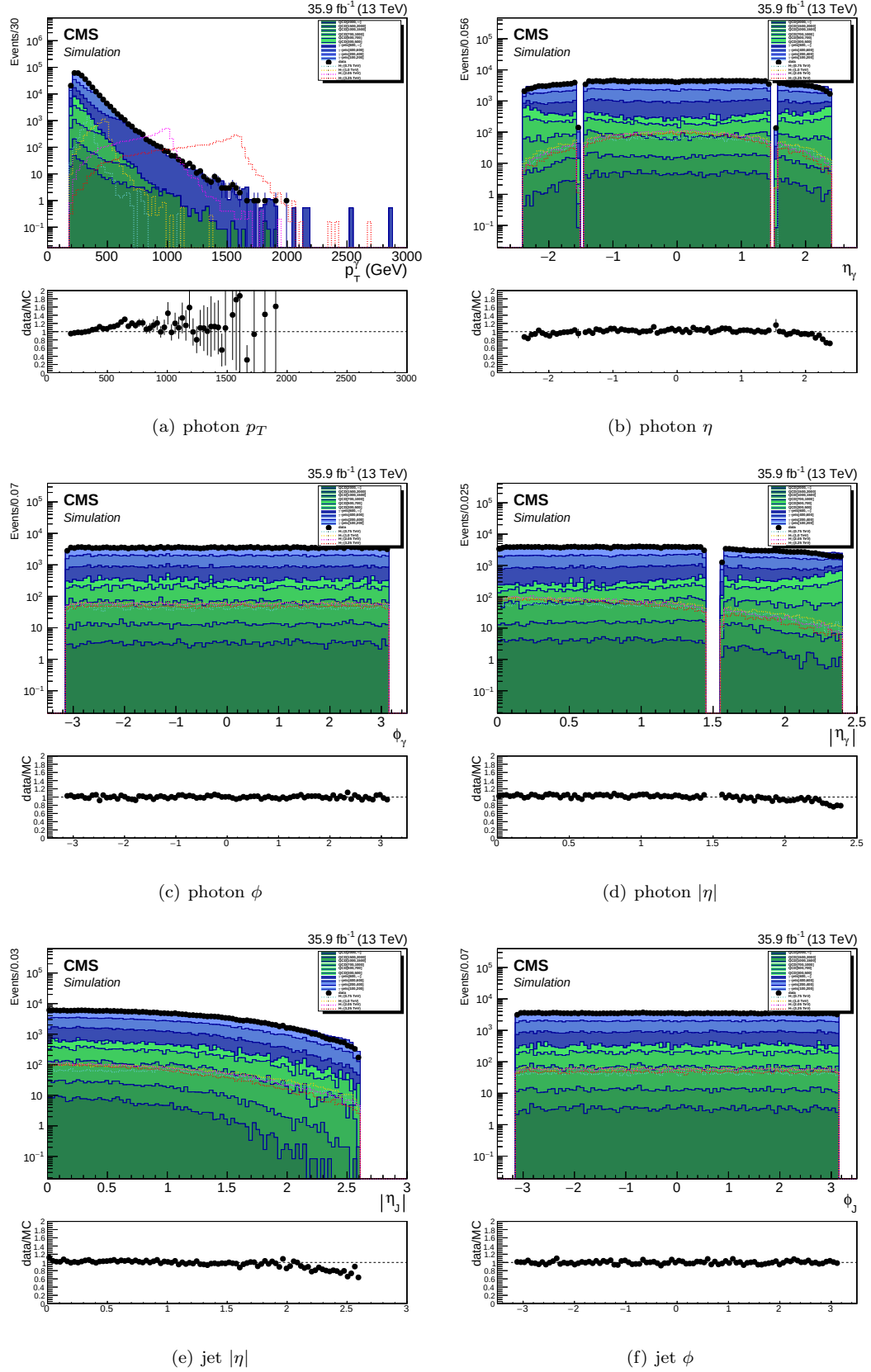


Figure 4.6: Preselection distributions of kinematic variables

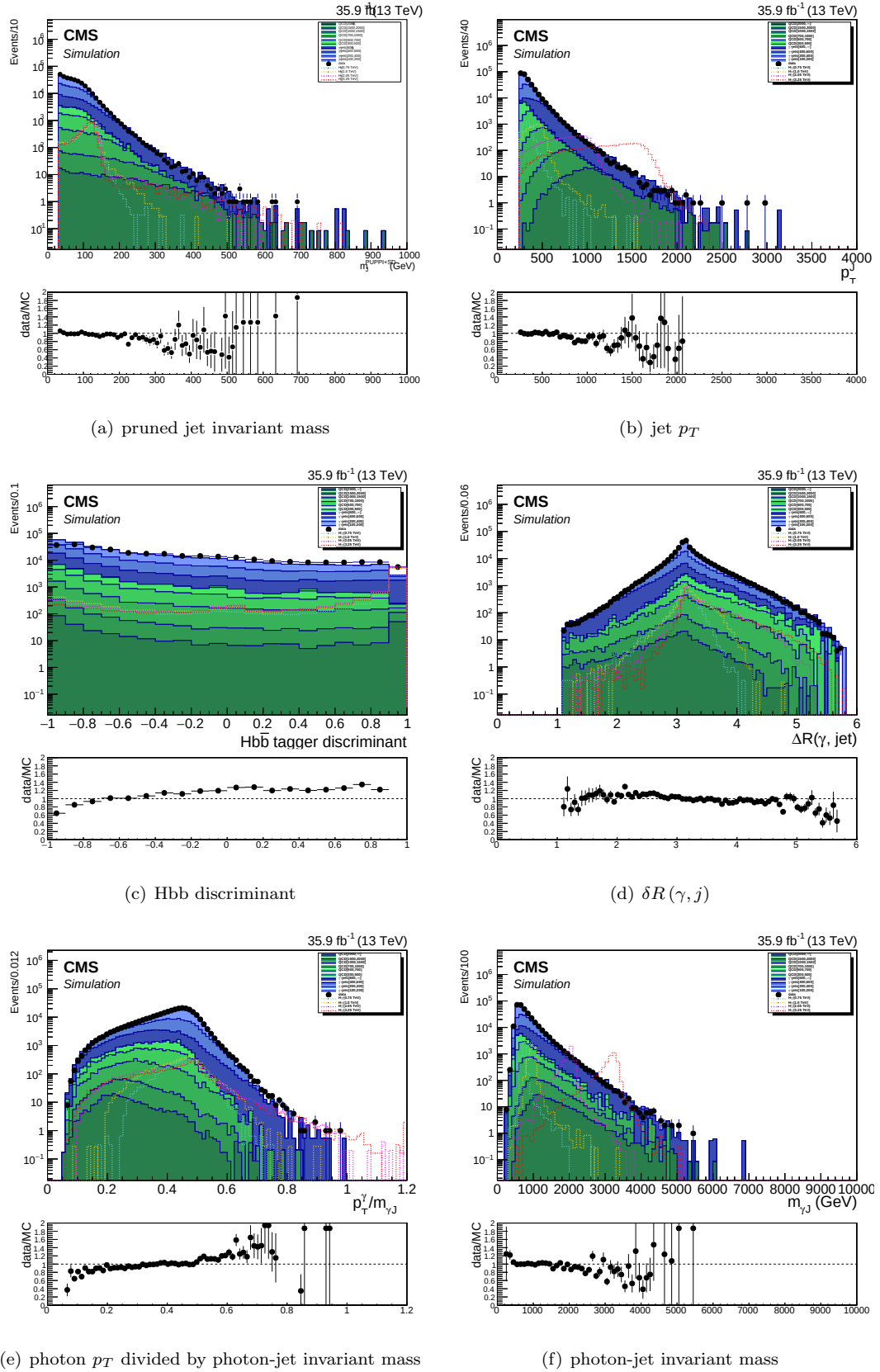


Figure 4.7: Pre-selection MC and data spectra. The spectra reproduce the shapes of the backgrounds reasonably well in all variables used in this analysis; however, cuts are derived using data-driven techniques.

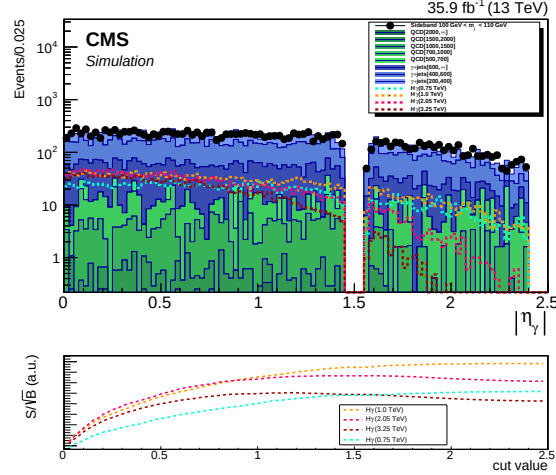
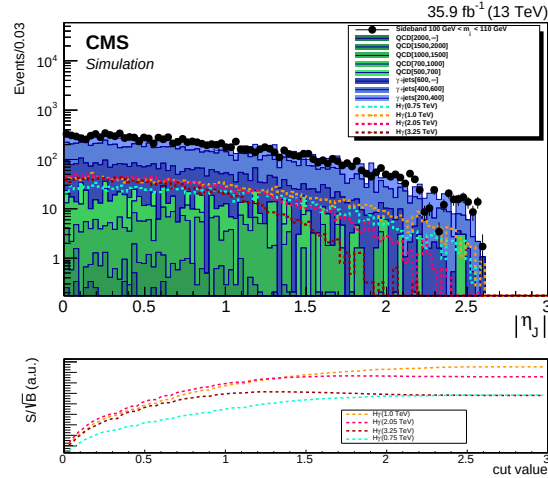
(a) Photon $|\eta|$ optimization(b) Jet $|\eta|$ optimization

Figure 4.8: Pre-b-tagging optimization of cuts on the maximum $|\eta|$ for selected photons and jets. The MC background and signal samples are plotted with the requirement that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data points plotted have the requirement that the jet mass to be in the 100 GeV to 110 GeV sideband. Aside from the special cut on the jet's invariant mass, the cuts applied in the top plot of the left figure are all the cuts in Table 4.3, except for the cuts on $|\eta_\gamma|$, the double-b tagger discriminator, and $p_T^\gamma/m_{J\gamma}$. In the bottom of the left figure, the cuts applied are the same but with a varying cut on $|\eta_\gamma|$ applied. Likewise, in the right figure, for the top plot, all cuts in Table 3 are applied except for the cuts on $|\eta_J|$, the double-b tagger discriminator, and $p_T^\gamma/m_{J\gamma}$, and for the top plot, a varying cut on $|\eta_J|$ is also applied. The varying cuts for both bottom graphs require that $|\eta|$ be less than the cut value indicated.

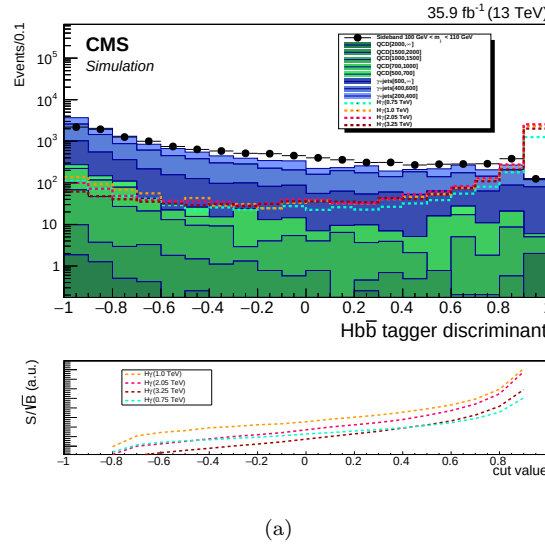
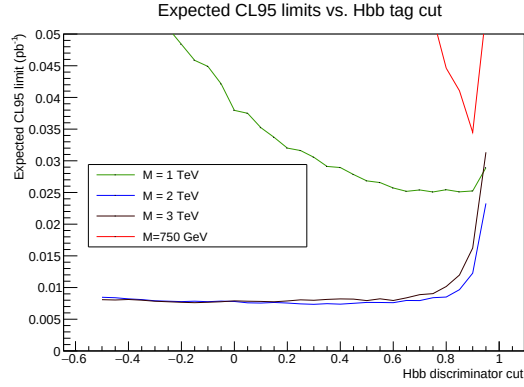


Figure 4.9: Optimization of the b-tagging selection. The value that separates the untagged and b-tagged categories is chosen to maximize the sensitivity in the b-tagging category. The MC background and signal samples are plotted with the requirement that the event’s jet have an invariant mass between 110 GeV and 140 GeV, whereas the data points plotted have the requirement that the jet mass to be in the 100 GeV to 110 GeV sideband. In addition to the special cuts for the jet invariant mass, the cuts applied in the top plot are all the cuts in Table 4.3, except for the cuts on the double-b tagger discriminator and $p_T^{\gamma}/m_{J\gamma}$. In the bottom, the cuts applied are the same but with a varying cut on the double-b tagger discriminator value. The varying b-tagging selection value requires that the double-b tagger discriminant be greater than the value indicated.



(a) CL95 expected limits for DBT working points

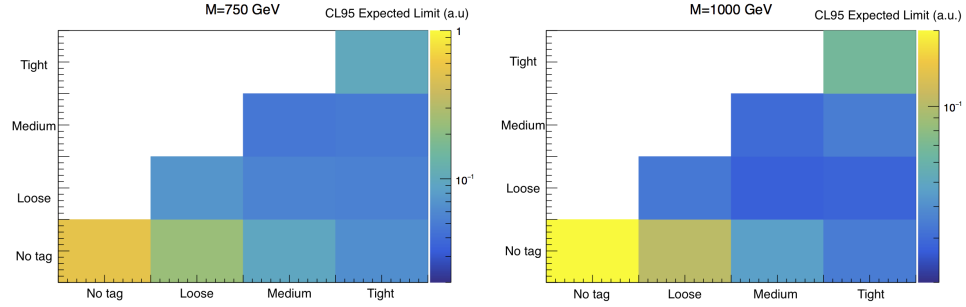
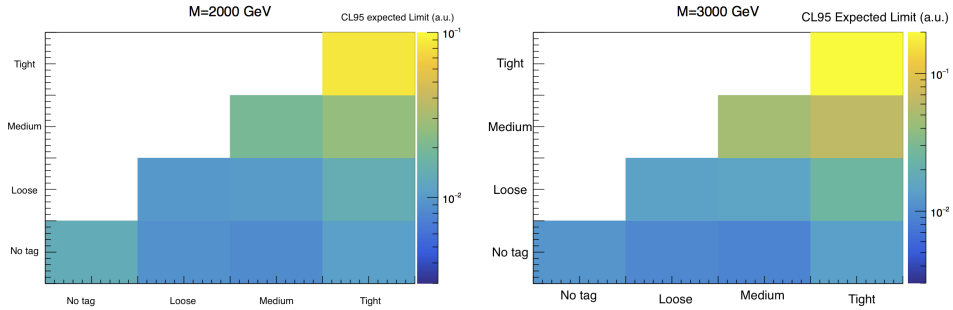
(b) double-subject CSV: expected CL95 limits at various sets of working points on subjects, $m_X = 750$ GeV(c) double-subject CSV: $m_X = 1000$ GeV(d) double-subject CSV: $m_X = 2000$ GeV(e) double-subject CSV: $m_X = 3000$ GeV

Figure 4.10: Comparison of the double-b tagger with double subject b-tagging. The double-b tagger outperforms double-subject b-tagging also at high resonance masses. The MC background and signal samples require that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data require that the event's jet mass be in the 100 GeV to 110 GeV sideband. The cuts applied for making the top plot are the same as in the bottom plot of Figure 4.9, but in the rest of the figures, instead of applying a cut on the double-b tagger discriminator value, different cuts on the leading and subleading subjects' CSVv2 tagger discriminator are applied corresponding to different choices of working points for the leading and subleading subjects.

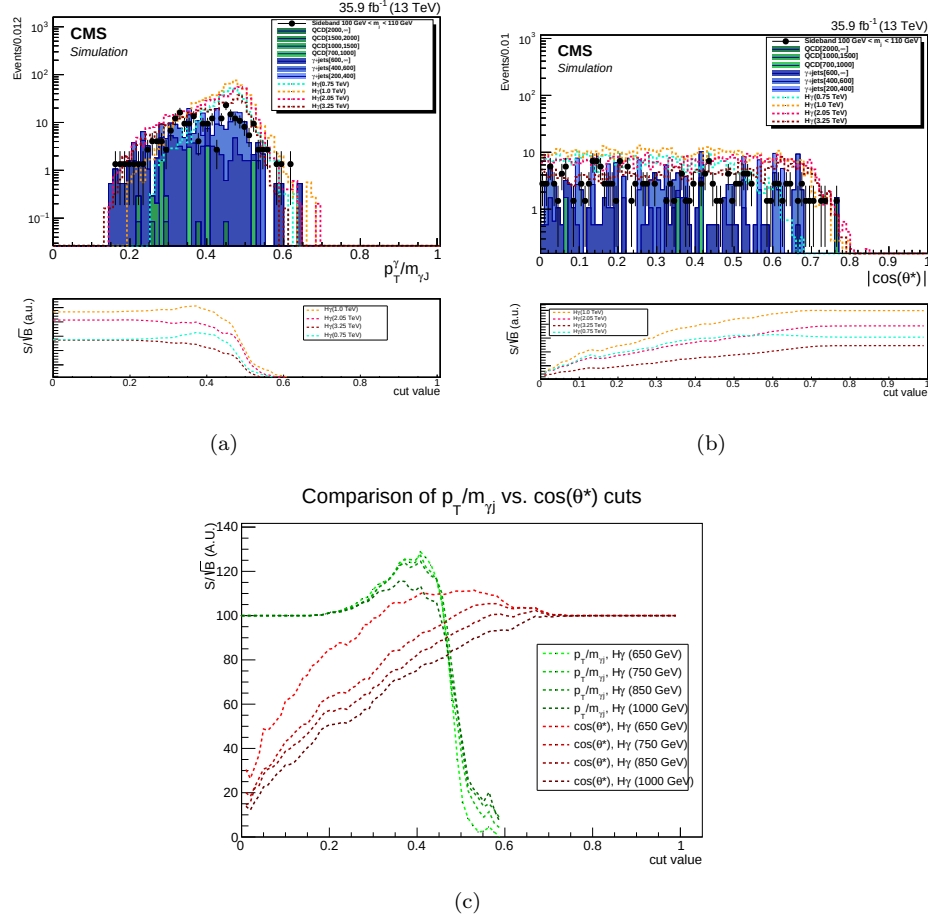


Figure 4.11: Additional kinematical cut: the p_T^{γ} cut at 0.35 is chosen over the similar variable $\cos(\theta^*)$. The MC background and signal samples are plotted with the requirement that the event's jet have an invariant mass between 110 GeV and 140 GeV, whereas the data points plotted have the requirement that the jet mass to be in the 100 GeV to 110 GeV sideband. Along with the special cuts for the jet invariant mass, the cuts applied for the stack plots in the top two figures are all the cuts in Table 4.3, except for the cut on $p_T^{\gamma}/m_{J\gamma}$. For the graph in the bottom of the top left figure, as well as the green lines of bottom figure, the cuts applied are the same, except that a varying cut on $p_T^{\gamma}/m_{J\gamma}$ is applied. Likewise, for the graph in the bottom of the top right figure and the red lines of bottom figure, the cuts are again all the cuts in Table 4.3, except no cut on $p_T^{\gamma}/m_{J\gamma}$ is applied, and instead a varying cut on $\cos(\theta^*)$ is applied. The varying cut on $p_T^{\gamma}/m_{J\gamma}$ requires that it be greater than the cut value indicated, whereas the varying cut on $\cos(\theta^*)$ requires it to be less than the cut value indicated.

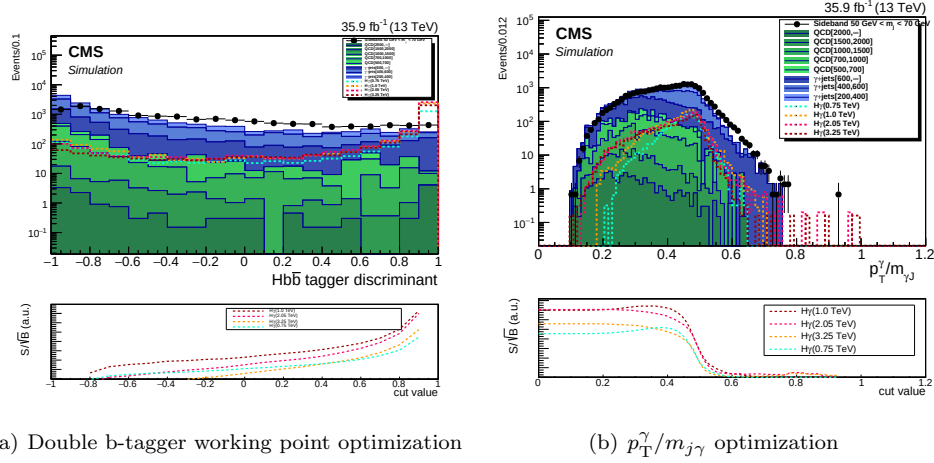


Figure 4.12: Optimization of sensitivity-maximizing cuts using 50-70 GeV PUPPI softdrop jet mass sideband

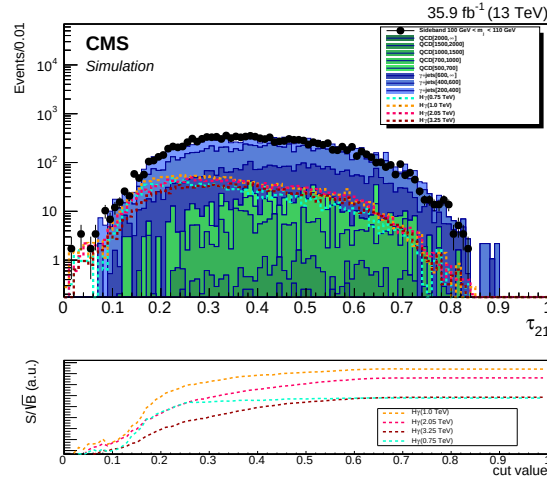


Figure 4.13: No cut on τ_{21} is applied in this analysis, as the modest gains in signal sensitivity for low mass signals would introduce a systematic uncertainty that would not increase the sensitivity of this analysis.

4.6 Search strategy

After all other cuts have been applied, the events passing the Hbb tagger discriminator are considered as one category (the “b-tagged” category), while the remainder of the events passing all other cuts are considered as a second category (the “untagged” category). In both categories, limits are set using the following procedure, and the limits are then combined for a final limit.

A smooth background function is fitted to the signal region’s $J\gamma$ invariant mass spectrum to estimate the background; this function is chosen by conducting an investigation of several alternative background fit functions. The function giving the best goodness-of-fit in these studies is used to make a background prediction. To model the signals’ $J\gamma$ invariant mass spectrum, a crystal ball function is fitted to create a signal template shape for each of the signal samples. Using the smoothly falling background model and the signal shape template, the sum of the background plus signal shapes is fitted to the signal region’s $m_{J\gamma}$ spectrum, and the signal significance is extracted from the resultant fit. This procedure is carried out for signal resonant masses from 720 GeV to 3250 GeV. A cartoon illustrating the basic concept is given in Figure 4.14

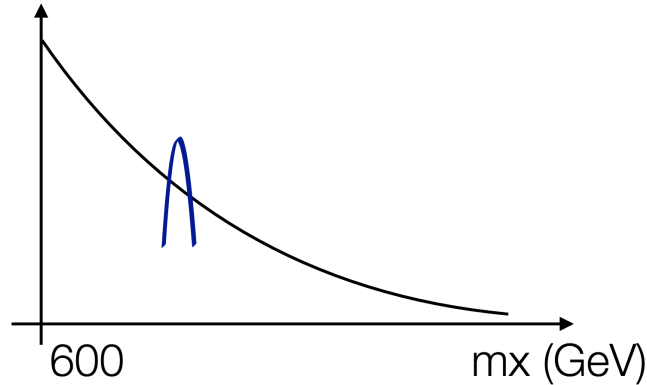
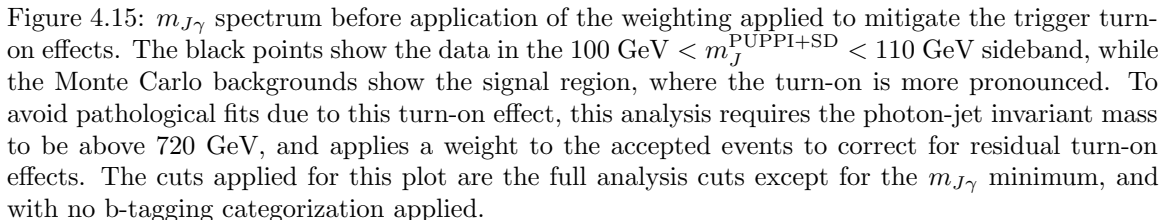


Figure 4.14: An illustration showing the basic concept of the “bump hunt” strategy employed in this analysis. In the illustration, m_x denotes the mass of the resonance, the smoothly monotonically falling black line is the background fit, and the blue peak is the bump that would arise due to the presence of a narrow resonance. Borrowed from [14].

One additional refinement to the method is implemented to account for the observed trigger-turn inefficiency shown in Figure 4.1. To correct any remaining turn-on effects due to the trigger above the 720 GeV cut on the invariant mass, each event in the data is weighted by the inverse of the observed trigger efficiency according to the transverse momentum of the selected photon in that event. The efficiencies used in the calculation are the same as those shown in Figure 4.1, where we use the values from the study of efficiency versus photon p_T performed in the 2016 SingleMuon data. Figure 4.15 shows the spectrum before this weighting, with all cuts applied except for the



The simulated narrow-width resonance invariant mass spectra are fit using a function characteristic of narrow-mass resonances, and are then templated to interpolate a signal shape at every 10 GeV throughout the signal region. The interpolation procedure is explained in detail in Section 4.8. A subset of the interpolated signal shapes is shown in Figure 4.24. For each of these signal templates, the data is fit using a sum of the smooth background function plus the signal template, with one free parameter corresponding to the cross-section times branching ratio.

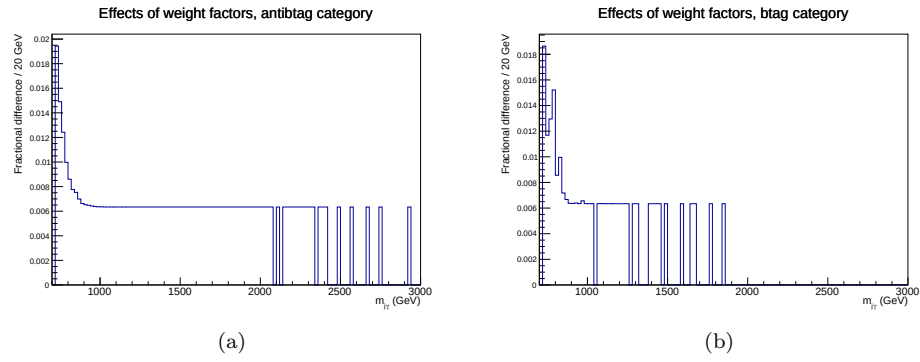


Figure 4.16: The effects of the reweighting applied to the data. The fractional difference plotted is calculated as $(N_{\text{weighted}} - N_{\text{unweighted}}) / N_{\text{weighted}}$. This plot is made using the data in the signal region.

4.7 Background shape modeling

The background fitting functions considered for the for this analysis were chosen in two steps. First, using the sideband spectrum, several categories of functions were investigated, where each family shares a similar functional form, but the order (i.e. the number of free parameters) varies within the family. The order of the best-fit function from within each family is chosen by picking the lowest order such that an increase in the function’s order does not yield a test statistic from the Kolmogorov-Smirnov test beyond the uncertainty of the test statistic of one lower order. Then, the nominal background fit function was chosen as the function with the smallest χ^2 test statistic in the sideband studies. This led to the choice for the “dijet” function of order 2 (the lowest usable order within the dijet family functions.). For all fits of the invariant mass spectrum, the fitting range chosen was 720-4720 GeV, with the lower edge corresponding to the end of the turn-on region, as discussed in Section 4.14, and the upper edge extending beyond the highest invariant mass event observed.

Shown below are the results of running the same KS and saturated models tests using the signal region, shown in Table 4.8 for the untagged category, and Table 4.6 for the b-tagged category. Also shown are the results by using the goodness-of-fit test using the saturated models test [142], in Tables 4.5 and 4.7, showing similar results.

$$\frac{dN}{dm} = P_0 \times (m/\sqrt{s})^{P_1 + P_2 \times \log(m/\sqrt{s})} \quad (4.1)$$

We refer to this as the “dijet” function of order 2.

- **expow**: exponential $\times$ power law $m^{P_1} e^{\sum_{i=2}^{N+1} P_i m^i}$
- **pow**: power law $\sum_i f_i m^{\alpha_i}$
- **vvdijet**: “VV” dijet $\frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2 + \sum_i P_i \times \log(m/\sqrt{s})}}$
- **atlas**: ATLAS function $\frac{P_0(1 - m/\sqrt{s})^{P_1/3}}{(m/\sqrt{s})^{P_2 + \sum_i P_i \times \log^{P_i}(m/\sqrt{s})}}$

The fits of all functions are shown in Figure 4.17.

Shown in shown in Table 4.9 are the χ^2 values for the various fits of the signal region. In the signal region, the dijet2 function did not achieve the smallest χ^2 ; however, its χ^2 was within uncertainty of the other best-performing functions, and to avoid biasing the choice of background function based on observations in the signal region, the dijet2 function was kept as the nominal fit function based on the studies of the sideband region.

Shown in Figure 4.18 are the final fit functions that were chosen based on the results of the goodness-of-fit tests described above, along with the pulls of the data relative to these fits. The error bars shown on the data points are given by the Garwood confidence intervals [143]. The function chosen for the background fits in both categories is the dijet2 function, and it is shown

Table 4.5: Goodness-of-fit test statistics from the saturated models test, for various background shape fit models, antibtag category.

Model	Order	Lower Bound	Test Statistic	Upper Bound
atlas	1	1261.59	1314.36	1365.6
vvdijet	1	1396.44	1451.6	1508.13
expow	1	1408.35	1457.93	1509.3
exp	1	1409.29	1459.78	1510.97
dijet	2	1431.79	1486.31	1539.77
expow	2	1435.55	1491.36	1543.52
vvdijet	2	1447.08	1500.01	1548.78
dijet	3	1445.27	1503.22	1559.05

Table 4.6: Goodness-of-fit test statistics from the Kolmogorov-Smirnov test, for various background shape fit models, antibtag category.

Model	Order	Lower Bound	Test Statistic	Upper Bound
vvdijet	2	0.00468405	0.00585454	0.0070481
dijet	3	0.00469837	0.00591273	0.0071553
expow	2	0.00470134	0.00600357	0.00728229
dijet	2	0.00479846	0.00610737	0.00746914
expow	1	0.00465967	0.0061239	0.00771571
vvdijet	1	0.00473874	0.00621016	0.00754898
exp	1	0.00531809	0.00718211	0.009116
atlas	1	0.00619424	0.00880725	0.0113205

Table 4.7: Goodness-of-fit test statistics from the saturated models test, for various background shape fit models, btag category.

Model	Order	Lower Bound	Test Statistic	Upper Bound
atlas	1	507.865	537.021	570.193
vvdijet	2	510.951	541.026	571.092
expow	2	508.441	541.48	574.359
dijet	2	509.574	541.855	573.87
vvdijet	1	512.085	542.864	574.148
expow	1	511.826	543.347	574.516
exp	1	513.787	546.411	577.571
pow	1	530.247	565.753	600.008
atlas	3	1124.31	1182.65	1244.92
dijet	3	1124.31	1182.65	1244.92

with its one- and two-sigma uncertainty bands. The one-sigma (two-sigma) uncertainty band on the background fits are found using a sampling method which generates a number of curves with

Table 4.8: Goodness-of-fit test statistics from the Kolmogorov-Smirnov test, for various background shape fit models, btag category.

Model	Order	Lower Bound	Test Statistic	Upper Bound
vvdijet	2	0.0323142	0.0410799	0.0501311
expow	2	0.0323616	0.0415958	0.0507536
dijet	2	0.0329181	0.0423721	0.0527441
expow	1	0.0330378	0.04275	0.0537952
vvdijet	1	0.0365541	0.0484789	0.0612254
pow	1	0.0364707	0.0497403	0.0641687
exp	1	0.0378725	0.0505348	0.06327
dijet	3	0.0371042	0.0535085	0.0689778
atlas	1	0.0426085	0.0610218	0.0778366
atlas	2	0.0426085	0.0610218	0.0778366

Table 4.9: χ^2 calculated for various background shape fit models at their optimal orders.

antibtag			btag	
model	χ^2		model	χ^2
expow3	2392.4		atlas3	104.23
dijet3	2408.0		pow1	110.41
atlas2	2408.5		dijet2	110.61
vvdijet2	2408.5		dijet3	110.66
dijet2	2412.5		atlas2	110.70
expow2	2421.0		expow1	110.72
expow1	2430.8		vvdijet1	110.73
exp1	2433.1		vvdijet2	110.81
vvdijet1	2434.8		expow2	110.84
atlas1	2502.6		exp1	110.90
pow1	3032.7		atlas1	111.08

parameters calculated by applying variations of the nominal parameter values drawn from a multi-variate Gaussian probability distribution function constructed from the fit results of the covariance matrix. Using this family of curves with variations of the parameters, the one-sigma (two-sigma) uncertainty band at each value of $m_{J\gamma}$ is determined by finding a central interval which covers 68% (95%) of the values of the functions evaluated at $m_{J\gamma}$. The total number of varied functions sampled is selected so that at least 30 functions evaluate outside the central interval at each value of $m_{J\gamma}$, which is minimally 100 (659) samples for the one-sigma (two-sigma) uncertainty band calculation.

The pulls shown in the lower panels of Figure 4.18 are calculated according to the formula $\text{pull} = (N_{\text{obs}} - N_{\text{exp}}) / \text{unc.}$, where the expected number of events $N_{\text{exp.}}$ is given by the integral of

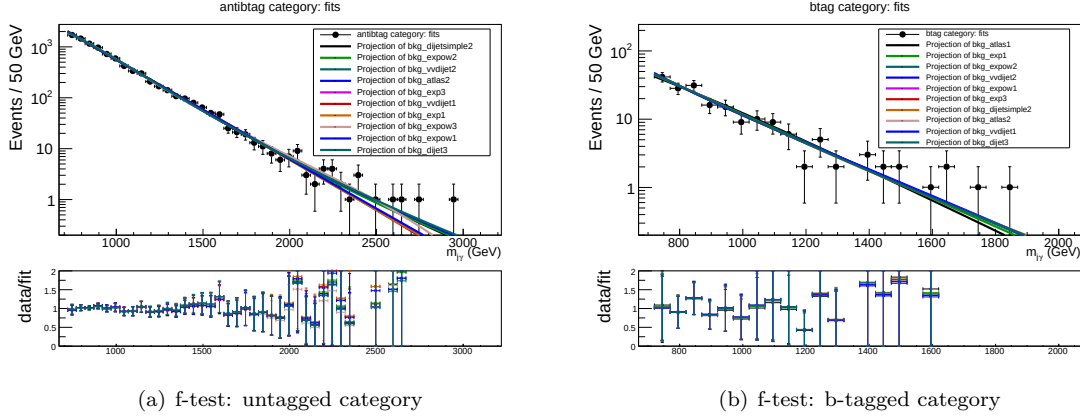


Figure 4.17: Several fits for the signal region spectra. The function chosen is the dijet2 function, based on studies of the sideband region, where its performance in the KS test is within uncertainty or better than a higher-order dijet function, its relative good performance versus all other functions in both categories, and its χ^2 within uncertainty of other competitive functions.

the background fit function over that bin, and $N_{\text{obs.}}$ is given by the number of data events in that bin. The uncertainty in the denominator of this definition is equal to the statistical uncertainty on the data in a bin added in quadrature with the fit function's uncertainty at the center of that bin.

Also shown in Figure 4.18 are the $m_{J\gamma}$ spectra for various signal samples. These samples are drawn with arbitrary normalization, since, as described in Section 2.4, the cross sections in the model under consideration are not known *a priori*, so the normalization of the signal samples is chosen for illustrative purposes.

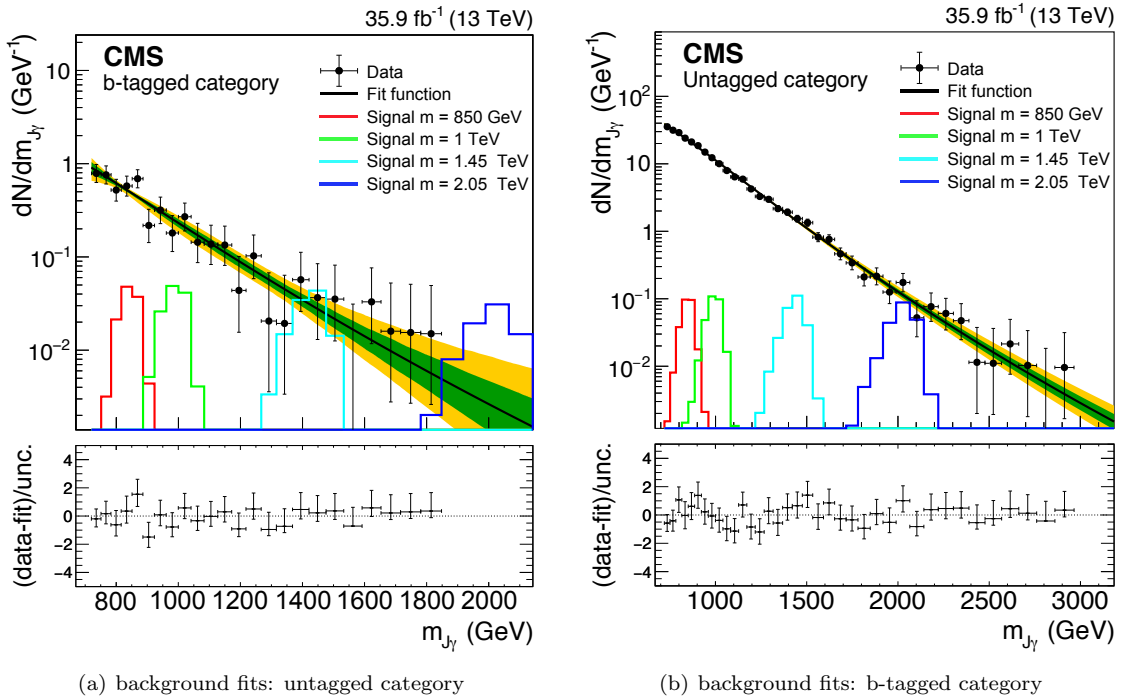


Figure 4.18: Shown in the top panels of subfigures (a) and (b) are the final background fits of the $m_{J\gamma}$ spectra in both categories, along with the fit uncertainties calculated using the sampling method described above. Shown in the bottom panels of both subfigures are data's pulls with respect to the fit function, given both statistical and fit uncertainties.

4.8 Signal shape modeling

The signal shape was modeled using a crystal ball function. Goodness-of-fit test statistics for the crystal ball fits of the signal samples' $m_{J\gamma}$ spectra are shown in Tables 4.10 and 4.11, where the KS test statistic shown is the Kolmogorov-Smirnov probability.

Table 4.10: Goodness-of-fit test statistics for signal fits using the Kolmogorov-Smirnov test statistic: antibtag category.

signal mass (GeV)	KS probability
750	0.999
850	0.963
1000	0.999
1150	0.892
1300	0.999
1450	0.943
1600	0.999
1750	0.954
1900	0.801
2050	0.999
2450	0.783
2850	0.999
3250	0.762

Table 4.11: Goodness-of-fit test statistics for signal fits using the Kolmogorov-Smirnov test statistic: btag category.

signal mass (GeV)	KS probability
750	0.999
850	0.824
1000	0.928
1150	0.978
1300	0.895
1450	0.742
1600	0.409
1750	0.865
1900	0.856
2050	0.999
2450	0.999
2850	0.994
3250	0.999

The signal shape was fit at all the resonance mass points for which MC was generated with powheg+Pythia 8, along with pythia hadronization and fullsim detector simulation. This shape was then interpolated between the mass points according to the linear interpolation method discussed in Reference [144], as implemented in the RooFit class RooIntegralMorph. For each pair of

fullsim-simulated signal samples, this method takes as inputs the crystal ball probability distributions determined by performing a fit on the fullsim signal samples. Then, a linear interpolation of the parameters is performed for each intermediate point. By construction, the interpolated shapes approach the original fit shapes, with their difference in p.d.f. parameters is proportional to the difference in simulated signal mass. These templated shapes were constructed as probability distribution functions, and to model the magnitude of the signal, a normalization based on the signal efficiency was applied (see Figure 4.23). The results of the interpolation are shown in Figure 4.24.

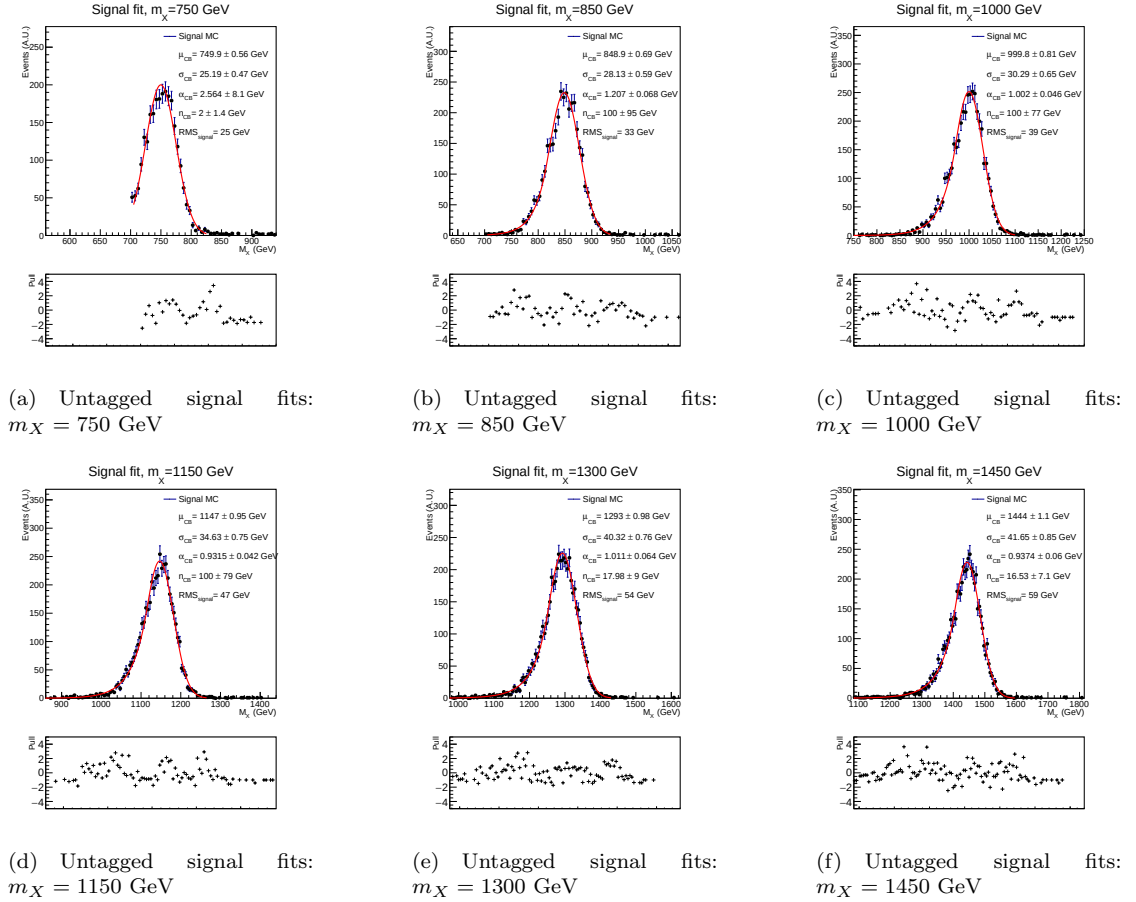
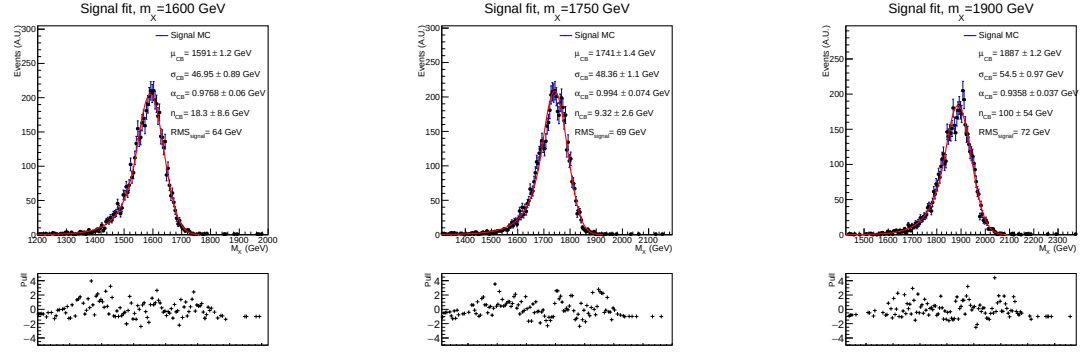


Figure 4.19: Simulated signal sample fits for the untagged category. The crystal ball fit is shown in red.

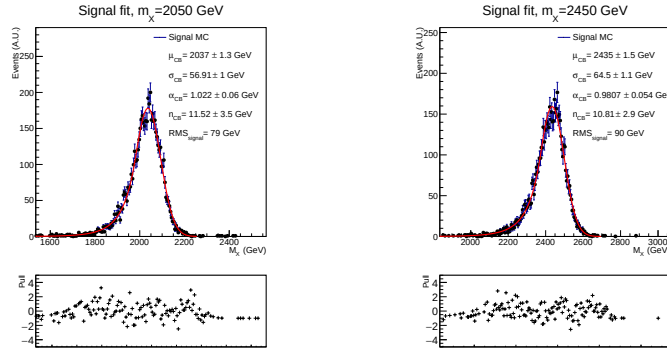
A closure test was performed to verify that the interpolated signal shapes give a reasonable approximation of the signal shapes that come directly from fitting a fullsim simulation at that point. To perform the closure test, a fullsim sample is chosen, and it is removed as an input into the interpolation algorithm. Then, the interpolation algorithm is run so that it creates an interpolated signal shape for the mass of the omitted fullsim sample, thus allowing a comparison between an interpolated signal shape for a given mass with the signal spectrum from a fullsim simulation at



(a) Untagged signal fits:
 $m_X = 1600$ GeV

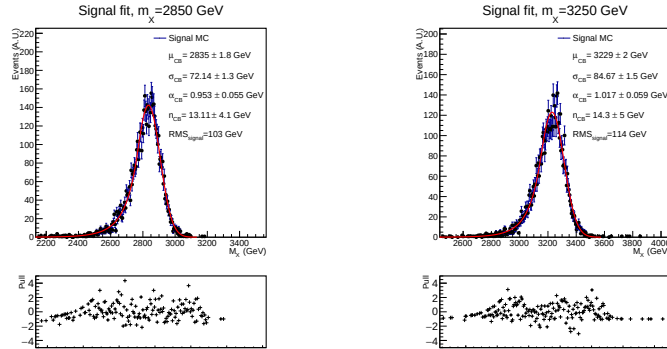
(b) Untagged signal fits:
 $m_X = 1750$ GeV

(c) Untagged signal fits:
 $m_X = 1900$ GeV



(d) Untagged signal fits:
 $m_X = 2050$ GeV

(e) Untagged signal fits:
 $m_X = 2450$ GeV

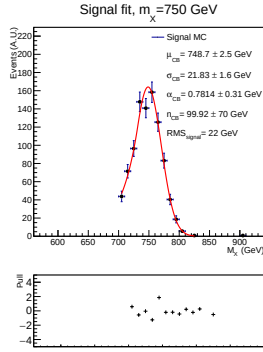


(f) Untagged signal fits:
 $m_X = 2850$ GeV

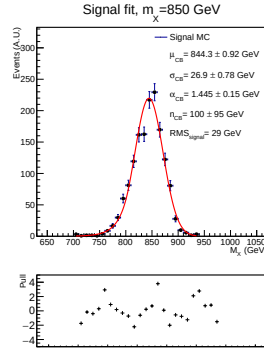
(g) Untagged signal fits:
 $m_X = 3250$ GeV

Figure 4.20: Simulated signal sample fits for the untagged category, continued. The crystal ball fit is shown in red.

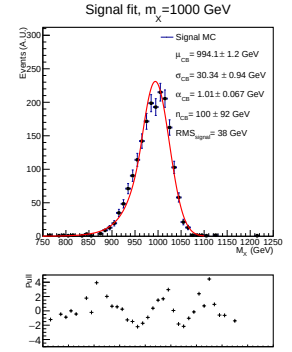
that mass. This was performed at several signal mass points in each category, shown in Figure 4.25 and Figure 4.26.



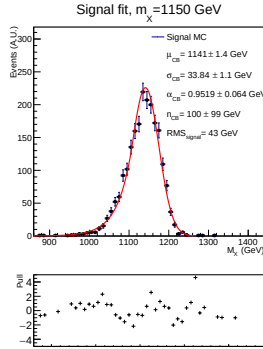
(a) b-tagged signal fits:
 $m_X = 750$ GeV



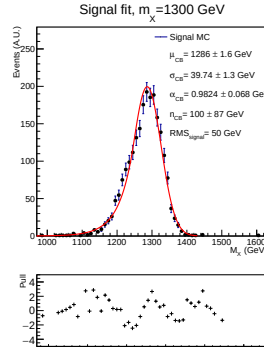
(b) b-tagged signal fits:
 $m_X = 850$ GeV



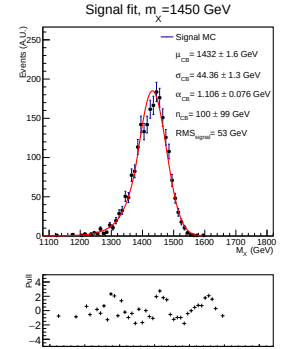
(c) b-tagged signal fits:
 $m_X = 1000$ GeV



(d) b-tagged signal fits:
 $m_X = 1150$ GeV

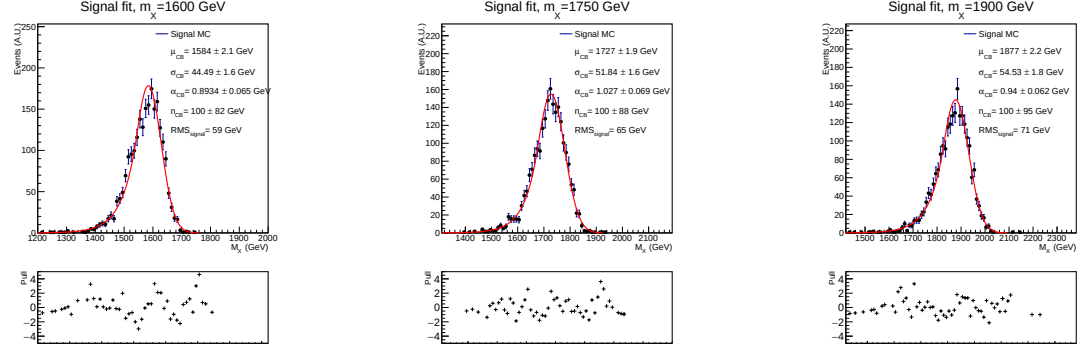


(e) b-tagged signal fits:
 $m_X = 1300$ GeV



(f) b-tagged signal fits:
 $m_X = 1450$ GeV

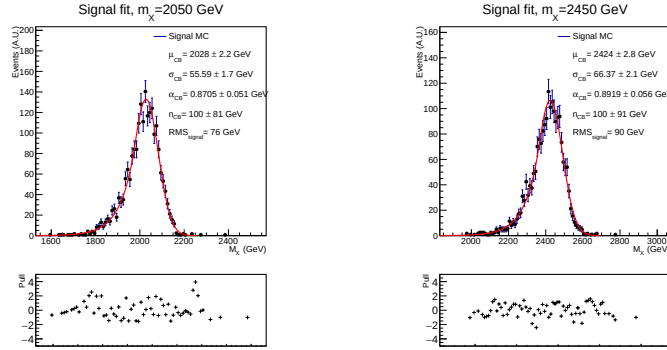
Figure 4.21: Simulated signal sample fits for the b-tagged category. The crystal ball fit is shown in red.



(a) b-tagged signal fits:
 $m_X = 1600$ GeV

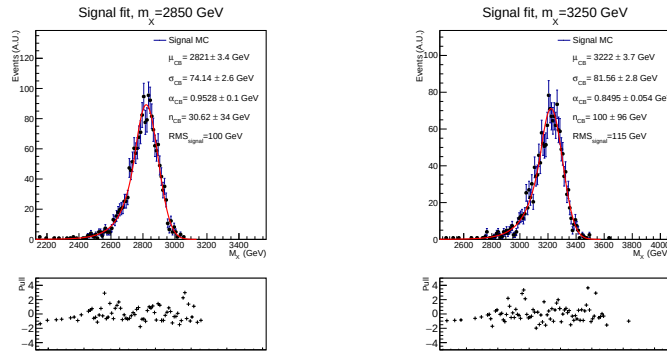
(b) b-tagged signal fits:
 $m_X = 1750$ GeV

(c) b-tagged signal fits:
 $m_X = 1900$ GeV



(d) b-tagged signal fits:
 $m_X = 2050$ GeV

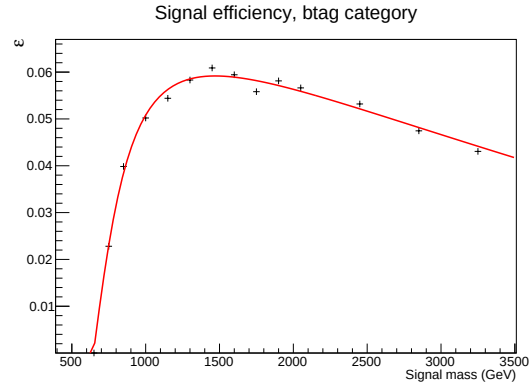
(e) b-tagged signal fits:
 $m_X = 2450$ GeV



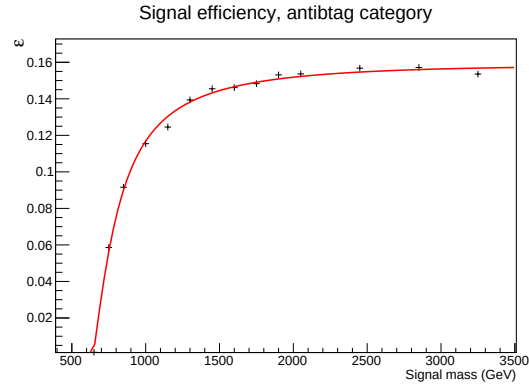
(f) b-tagged signal fits:
 $m_X = 2850$ GeV

(g) b-tagged signal fits:
 $m_X = 3250$ GeV

Figure 4.22: Simulated signal sample fits for the b-tagged category, continued. The crystal ball fit is shown in red.

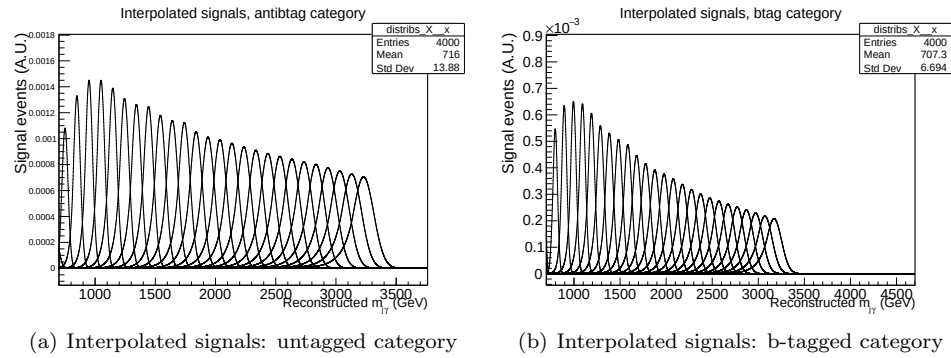


(a) Efficiency fit: b-tagged category



(b) Efficiency fig: untagged category

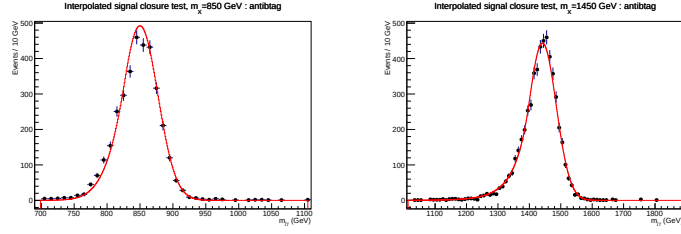
Figure 4.23: Simulated signal efficiency for the b-tagged and untagged categories fitted using three-parameter functions.



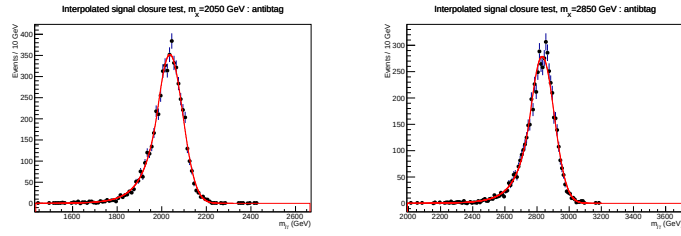
(a) Interpolated signals: untagged category

(b) Interpolated signals: b-tagged category

Figure 4.24: Results of signal shape interpolation. A signal shape is interpolated at every 10 GeV interval between points where a signal was simulated using fullsim. Shown here are only a subset of the interpolated signals.

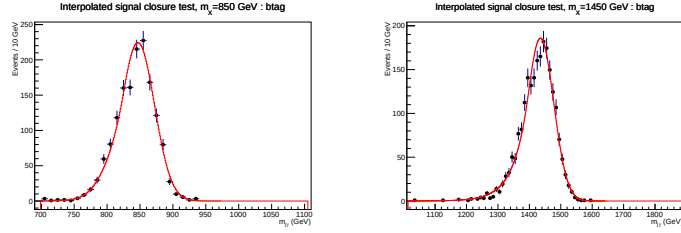


(a) Untagged signal interpolation closure test: $m_X = 850$ GeV (b) Untagged signal interpolation closure test: $m_X = 1450$ GeV

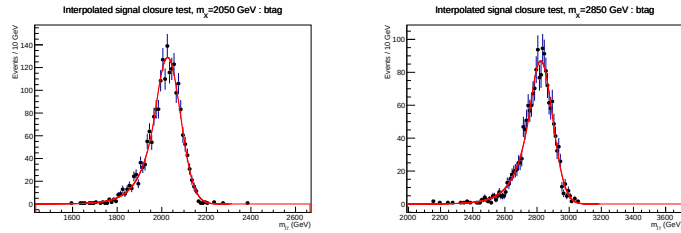


(c) Untagged signal interpolation closure test: $m_X = 2050$ GeV (d) Untagged signal interpolation closure test: $m_X = 2850$ GeV

Figure 4.25: Interpolation closure test: untagged category.



(a) b-tagged signal interpolation closure test: $m_X = 850$ GeV (b) b-tagged signal interpolation closure test: $m_X = 1450$ GeV



(c) b-tagged signal interpolation closure test: $m_X = 2050$ GeV (d) b-tagged signal interpolation closure test: $m_X = 2850$ GeV

Figure 4.26: Interpolation closure test: b-tagged category.

4.9 Bias studies

Because the estimate of the background in this analysis comes from a fit of the signal region, it is important to verify that the background fit function chosen does not introduce some systematic bias that would lead to the signal significance being artificially inflated or diminished relative to other possible background fit functions. In order to test whether the nominal background fit function (the dijet2 function, which was chosen according to the procedure in Section 4.7) causes any pathological effects on the results, a dedicated bias study was performed.

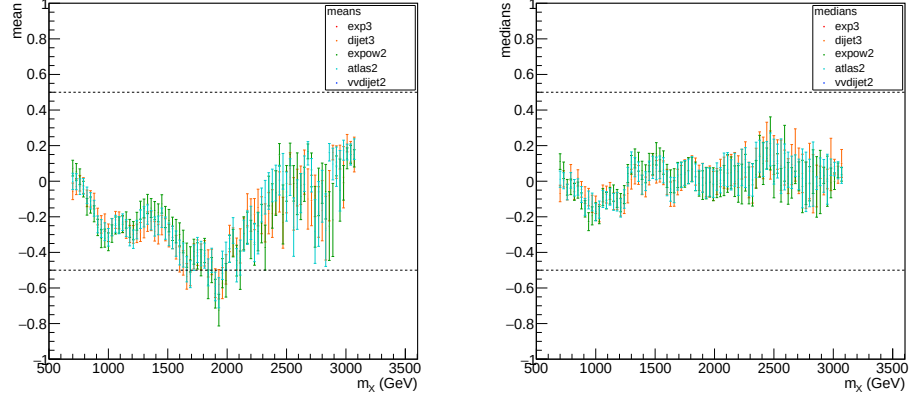
The method performed for the bias studies is also described in the $Z(q\bar{q})\gamma$ analysis. [131] This procedure involves picking an alternative background function as a template with which a simulated background distribution is generated. This simulated background distribution, or “toy” distribution, is created by treating the background function as a probability distribution function, and then for a given luminosity, a pseudo-data spectrum is generated according to that probability distribution function. Using the toy spectrum, as well as a particular signal template, the signal significance (in units of cross section), is found by fitting a sum of the nominal background fit and a signal fit to the toy spectrum, and an associated error on signal significance is found. Thus, aside from the use of use of the data to determine the nominal and alternative background fit functions, the bias study procedure is free from data.

This procedure is performed with 500 different toy distributions generated from the alternative background fit function. The pull for each toy is then calculated by dividing the signal significance by the uncertainty associated with this signal significance. With these 500 different values of pull, the distribution of the pulls is characterized in two different ways. The first characterization involves fitting a Gaussian to the spectrum of pulls, and then from this fit, the “pull mean” is defined as the mean of the Gaussian, and the error on the pull mean is defined as the standard deviation of the Gaussian fit. The second characterization of the pull spectrum involves finding the median, lower quartile, and upper quartile of the distribution of the pulls. The “pull median” is defined as the median of the pulls distribution, and the error on the “pull median” is given as one-half the difference between the upper quartile and lower quartile of the pulls distribution. Therefore we find a pull median and a pull mean, along with their errors, for a given signal template.

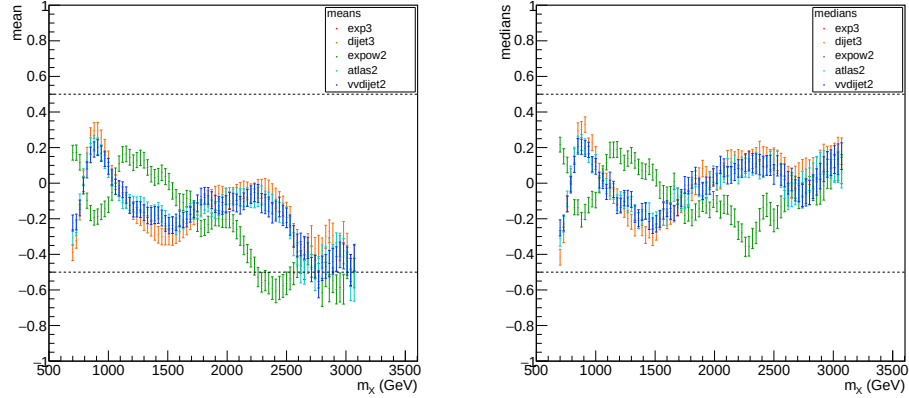
This procedure is done for all of the signal templates in both the b-tagged and untagged categories, and it is repeated using several different choices of alternative background fit function. For each of the b-tagged and untagged categories, the alternative background functions used in the bias study were the three functions besides the nominal fit function with the best χ^2 test statistic. The results of running this procedure using the dijet2 nominal background fit of the $100 \text{ GeV} < m_{j\gamma} < 110 \text{ GeV}$ sideband data are shown below in Figure 4.27.

For two background fit functions that achieve a good fit of the spectrum, it is expected that the toy spectrum from a particular pseudo-experiment may lead to signal+background fit that indicates non-negligible signal significance due to statistical fluctuations, but that for many pseudo-experiments, on average the signal significance should be small. If the magnitude of the pull medians

exceed 0.5, the systematic uncertainty due to the bias would be roughly 10%, and as the overall systematic uncertainties in the analysis are about 10%, a pull median exceeding 0.5 would suggest introducing an additional systematic uncertainty (called the “bias term”) to account for the effects of the background fit function biasing the results. However, it is observed that the pull medians do not exceed this threshold.



(a) Bias studies: pull means, b-tagged category (b) Bias studies: pull medians, b-tagged category



(c) Bias studies: pull means, untagged category (d) Bias studies: pull medians, untagged category

Figure 4.27: Results of bias studies in the b-tagged and untagged categories carried out using the signal region data. The profile plots characterize the distribution of the medians and means of the pull (i.e. signal significance divided by its uncertainty). In the b-tagged category, the pull means deviate from 0 at the point where the expected number of events gets close to zero; this is an effect of the best signal+background fit achieving its best fit when the magnitude of the signal is large and negative. (Note that the combine tool will set any negative values for the signal+background fit to 0, and thus a large negative signal component simply forces the signal+background fit to zero at masses at and above the signal mass; therefore in cases when there are empty bins near the tail of the spectrum, a large negative signal artificially attains the best fit.) However, the medians of the pull distributions are stable since the median is not as sensitive to the large-negative signal outliers.

4.10 Sidebands

Shown in Figure 4.28 are plots that compare the two sidebands employed in this analysis: $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ and $50 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 70 \text{ GeV}$. Shown are the spectra of data in these sidebands overlaid with the Monte Carlo distribution in the signal region, where the sideband data spectra are normalized so that the total number of events in the Monte Carlo signal region. Aside from the cut on the PUPPI softdrop jet mass, all other cuts of the analysis are applied here except for the b-tagging categorization cut. This study is meant to motivate the choice of sideband; we observe that the best agreement between the sideband spectra and MC signal region spectra is obtained using the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband. Therefore, this sideband window is the primary sideband used for the investigations into optimizing this analysis's choice of cuts, except in cases where signal contamination into the sideband might affect those choices. These plots also demonstrate that in the two cases where a particular value for a cut on a variable (in particular, the cuts on double-b tagger working point and $p_T^\gamma/m_{J\gamma}$), these variables do not suffer from great disagreement between the two possible sidebands, and that the lower sideband, where signal contamination is not expected, is valid for cross-checking the choice of cut value.

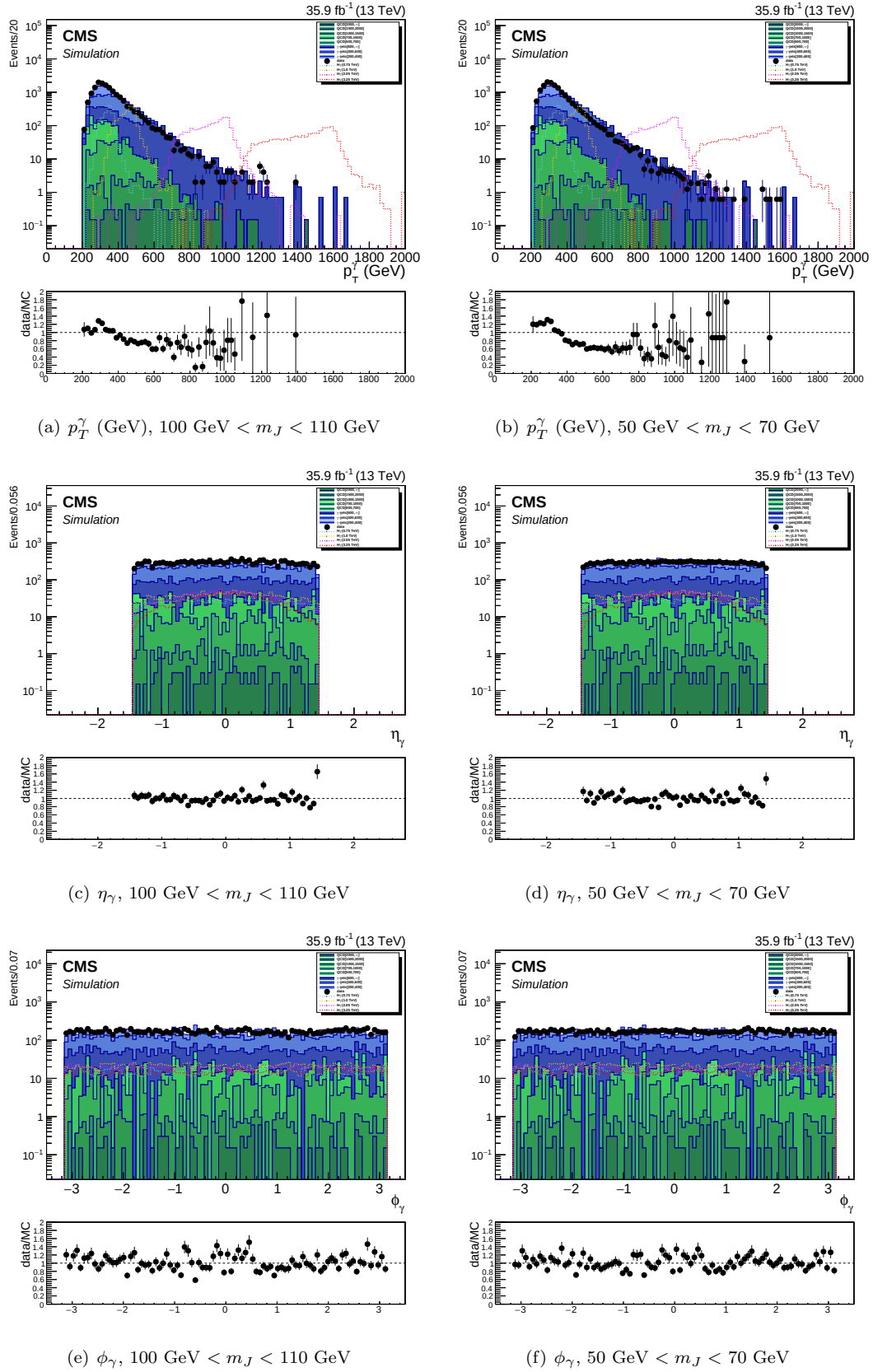


Figure 4.28: Comparisons of the spectra of several kinematic variables for the data in two different sidebands, plotted along with the spectrum of MC background samples in the signal region. All cuts are applied except for the b-tagging categorization cut.

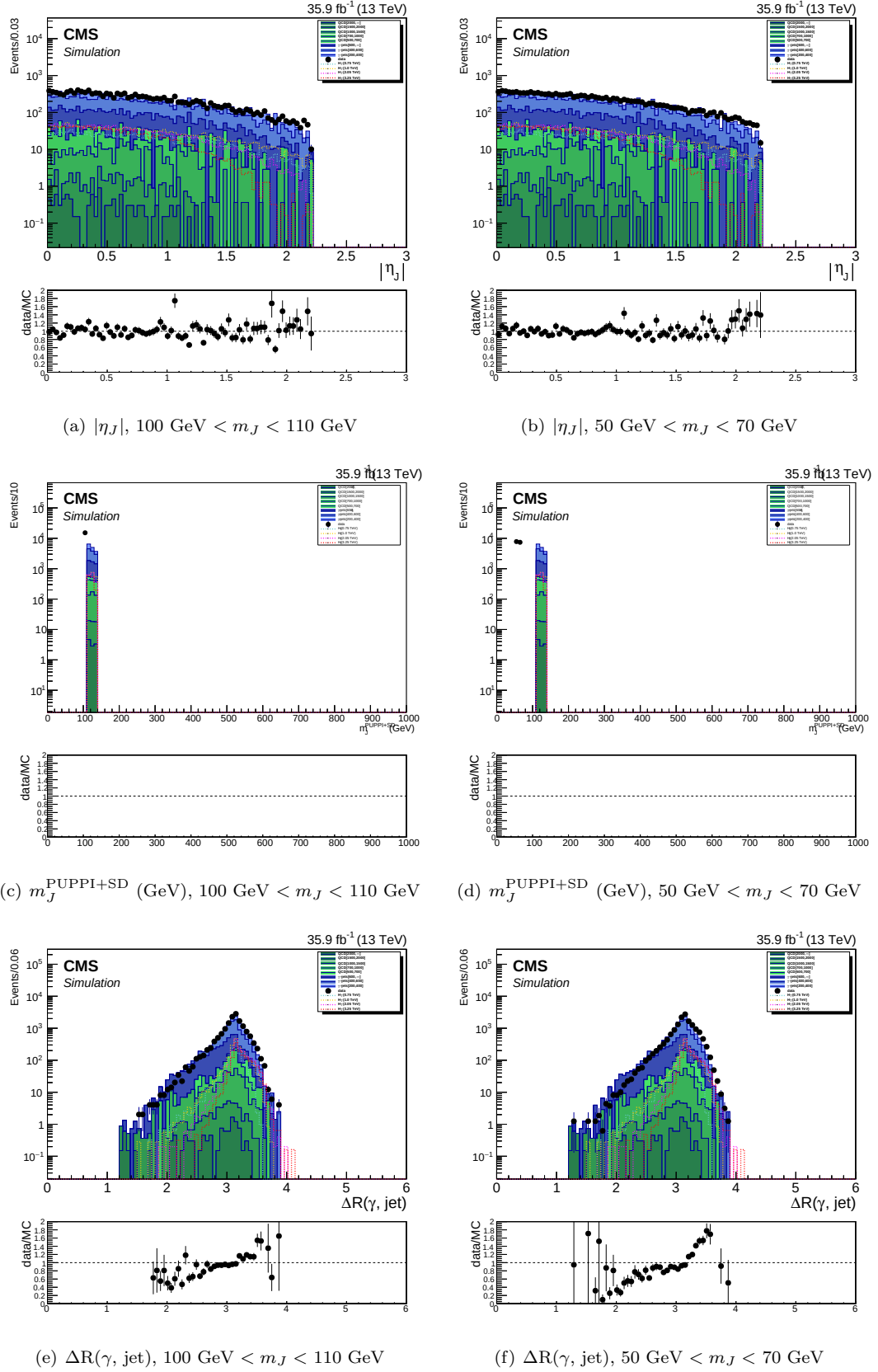


Figure 4.29: Comparisons of the spectra of several kinematic variables for the data in two different sidebands, plotted along with the spectrum of MC background samples in the signal region. All cuts are applied except for the b-tagging categorization cut.

As a check of whether the shape of the photon+jet invariant mass spectrum of the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ reliably gives an approximation of the expected shape of the photon+jet invariant mass spectrum in the signal region, a study was performed to check whether or not a significant difference in shape is observed between the photon+jet invariant mass spectra of the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband region versus that of the $50 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 70 \text{ GeV}$ sideband region. Shown in Figure 4.31 are the shapes of the photon+jet invariant mass spectra in both sidebands, for both b-tagged and untagged categories. In these plots, the 50-70 GeV sideband spectrum is scaled so that its normalization matches that of the 100-110 GeV sideband.

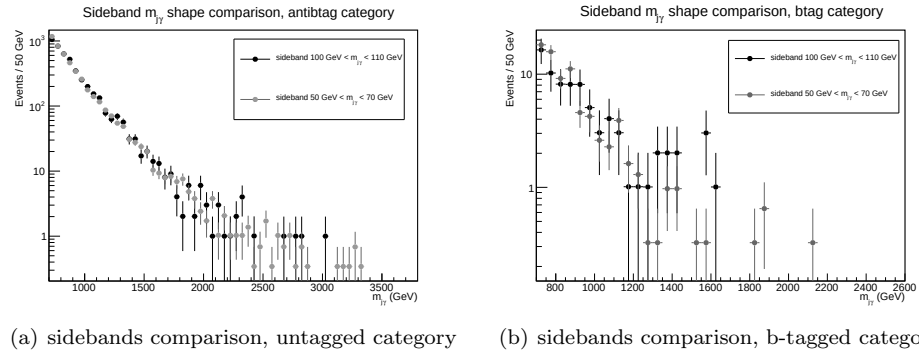
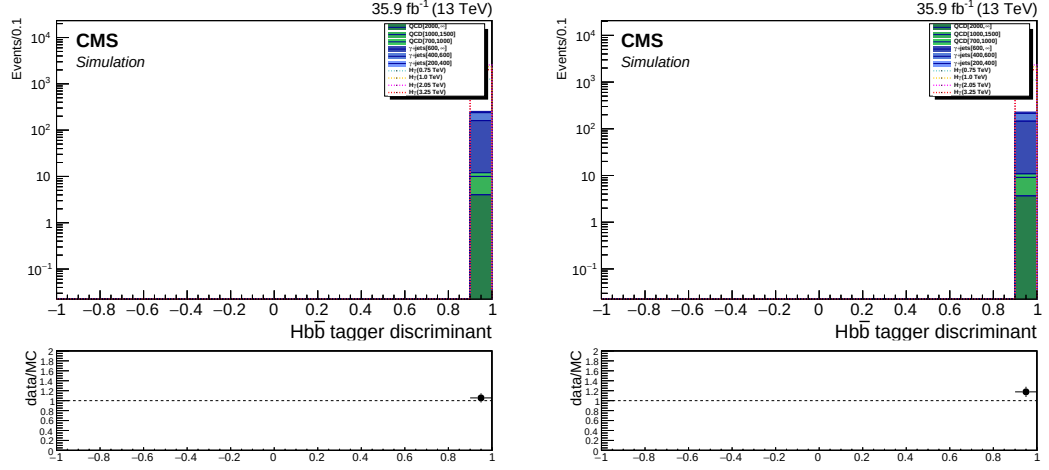


Figure 4.31: A comparison of the photon+jet invariant mass spectra in the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ and $50 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 70 \text{ GeV}$ sideband regions. The shapes generally agree within uncertainties. This leads to the expectation that the shape of the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband spectrum gives a good prediction of the shape of the signal region.

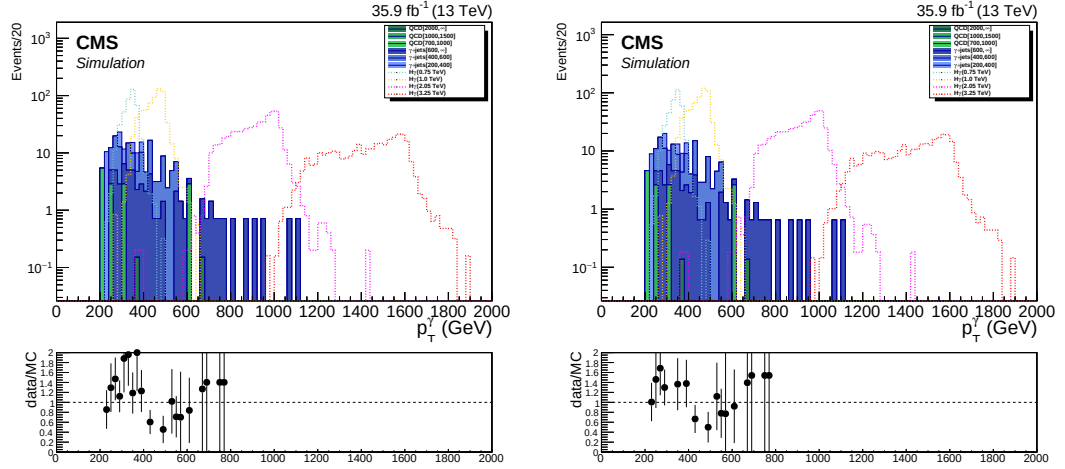
It is observed that the shapes of the two sideband spectra agree well within uncertainties. Therefore it is expected that the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ photon+jet invariant mass spectrum gives a good estimation for the expected shape of the invariant mass spectrum in the signal region, thus supporting the reliability of the studies performed using the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband.

4.11 Double-b tagging scale factors

Shown in Figure 4.32 - 4.34 are comparisons of several relevant variables' spectra before and after the application of the b-tagging scale factors. The comparisons are presented for both the b-tagged and untagged categories. The Monte Carlo samples are plotted after the full set of cuts used in this analysis for both categories. It is observed that no large changes arise in the shapes of the spectra of the relevant variables.

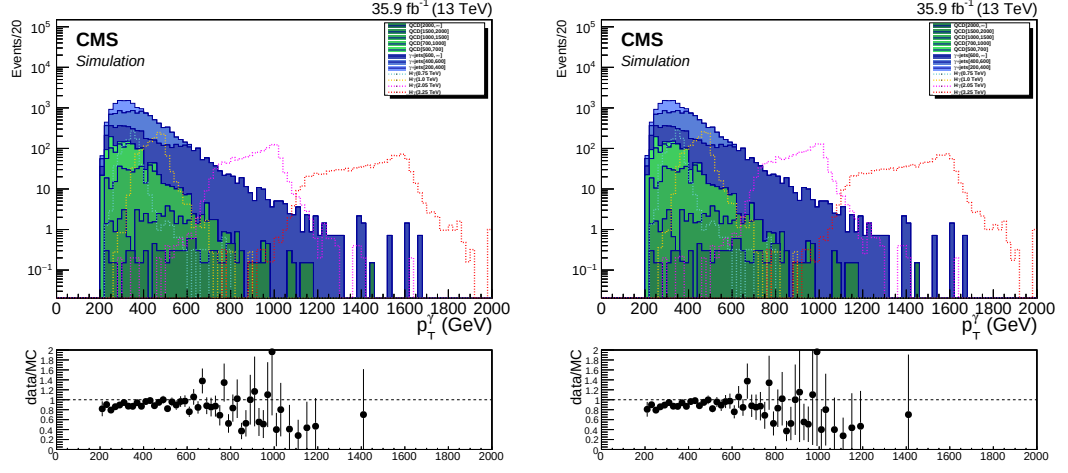


(a) b-tagged category: $Hb\bar{b}$ tagger discriminant, no scale factors (b) b-tagged category: $Hb\bar{b}$ tagger discriminant, with scale factors

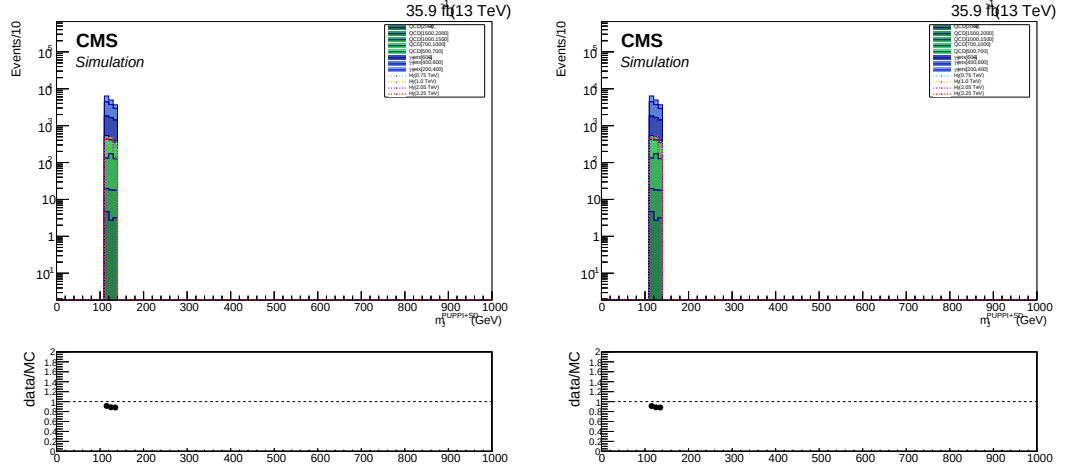


(c) b-tagged category: p_T^γ (GeV), no scale factors (d) b-tagged category: p_T^γ (GeV), with scale factors

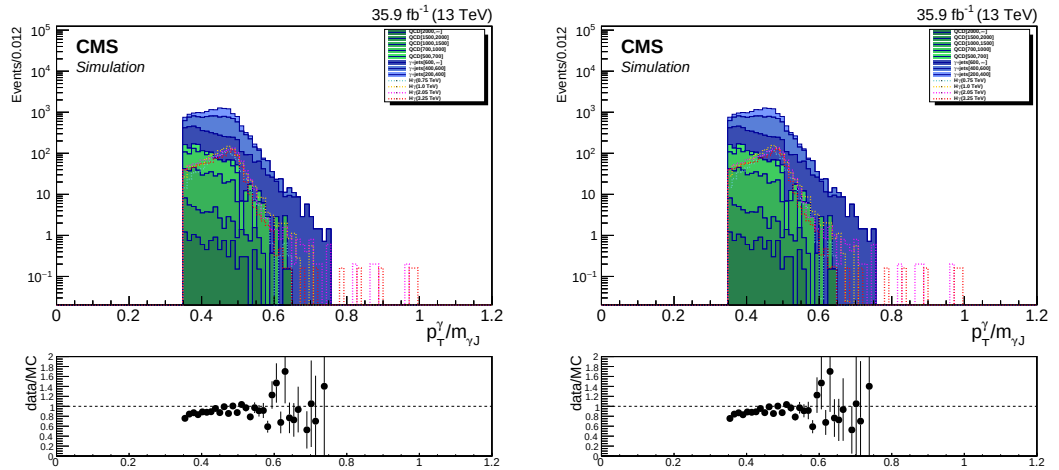
Figure 4.32: b-tagged category: Comparisons of the spectra of several kinematic variables for the Monte Carlo background and signal samples with and without the b-tagging scale factors applied.



(a) Untagged category: p_T^{γ} (GeV), no scale factors (b) Untagged category: p_T^{γ} (GeV), with scale factors



(c) Untagged category: $m_J^{\text{PUPPI+SD}}$ (GeV), no scale factors (d) Untagged category: $m_J^{\text{PUPPI+SD}}$ (GeV), with scale factors



(e) Untagged category: $p_T^{\gamma}/m_{j\gamma}$, no scale factors (f) Untagged category: $p_T^{\gamma}/m_{j\gamma}$, with scale factors

Figure 4.34: Untagged category: Comparisons of the spectra of several kinematic variables for the Monte Carlo background and signal samples with and without the b-tagging scale factors applied.

4.12 800 GeV minimum photon+jet invariant mass

A study was performed on the possibility of omitting the trigger efficiency reweighting procedure discussed in Section 4.6, and in order to avoid the residual observed turn-on effects visible in Figure 4.1, we instead increase the minimum required photon+jet invariant mass to be above 800 GeV rather than 720 GeV. The background shape fits that result from considering only events with a photon+jet invariant mass greater than 800 GeV, performed using the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband in the data, are shown in Figure 4.35.

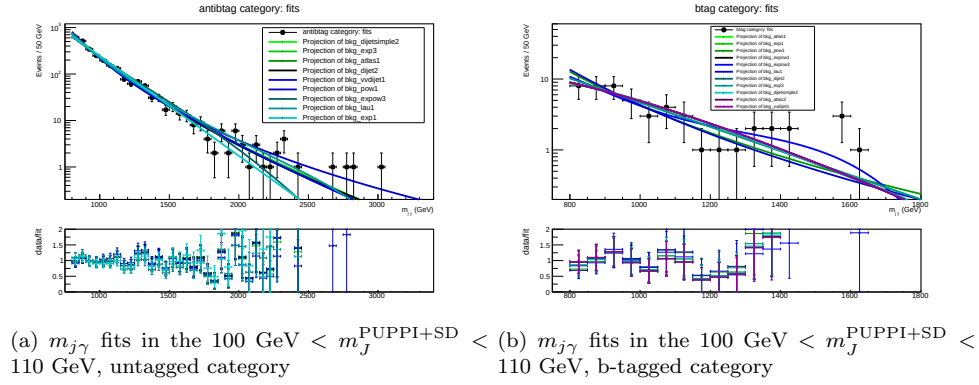


Figure 4.35: Fits of the $m_{J\gamma}$ spectra for events with $m_{J\gamma} > 800 \text{ GeV}$, with no reweighting applied to correct for residual trigger inefficiency. These fits are performed on the $100 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 110 \text{ GeV}$ sideband. It is observed that the fits in the untagged category are nearly identical to the corresponding fits shown in Figure 4.17, however, in the b-tagged category, the fits demonstrate unstable behavior. This instability in the fits results because the region $720 \text{ GeV} < m_{J\gamma} < 800 \text{ GeV}$ is the most densely populated region of the spectrum, and removing those events causes the fits to be underconstrained.

It is found that discarding the events from the $720 \text{ GeV} < m_{J\gamma} < 800 \text{ GeV}$ region leads to instability of the fits in the b-tagged category, due to the fact that this region is the most densely populated part of the spectrum, and thus provides the needed statistics for constraining the fits.

4.13 Limit calculation exercises

As an exercise to validate the analysis strategy without unblinding the data in the signal region, expected limits were calculated for the b-tagged, untagged, and combined categories using a few proxies for the data's signal region: the two sideband spectra in the data as well as the signal region spectrum in the MC background samples. The results of these exercises are shown in Figure 4.36 and serve as validation for the analysis strategy.

Shown in Figure 4.37 is a similar exercise in the limit-setting procedure, where limits are set by running the limit setting procedures on the MC background samples, using the $110 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 140 \text{ GeV}$ signal region.

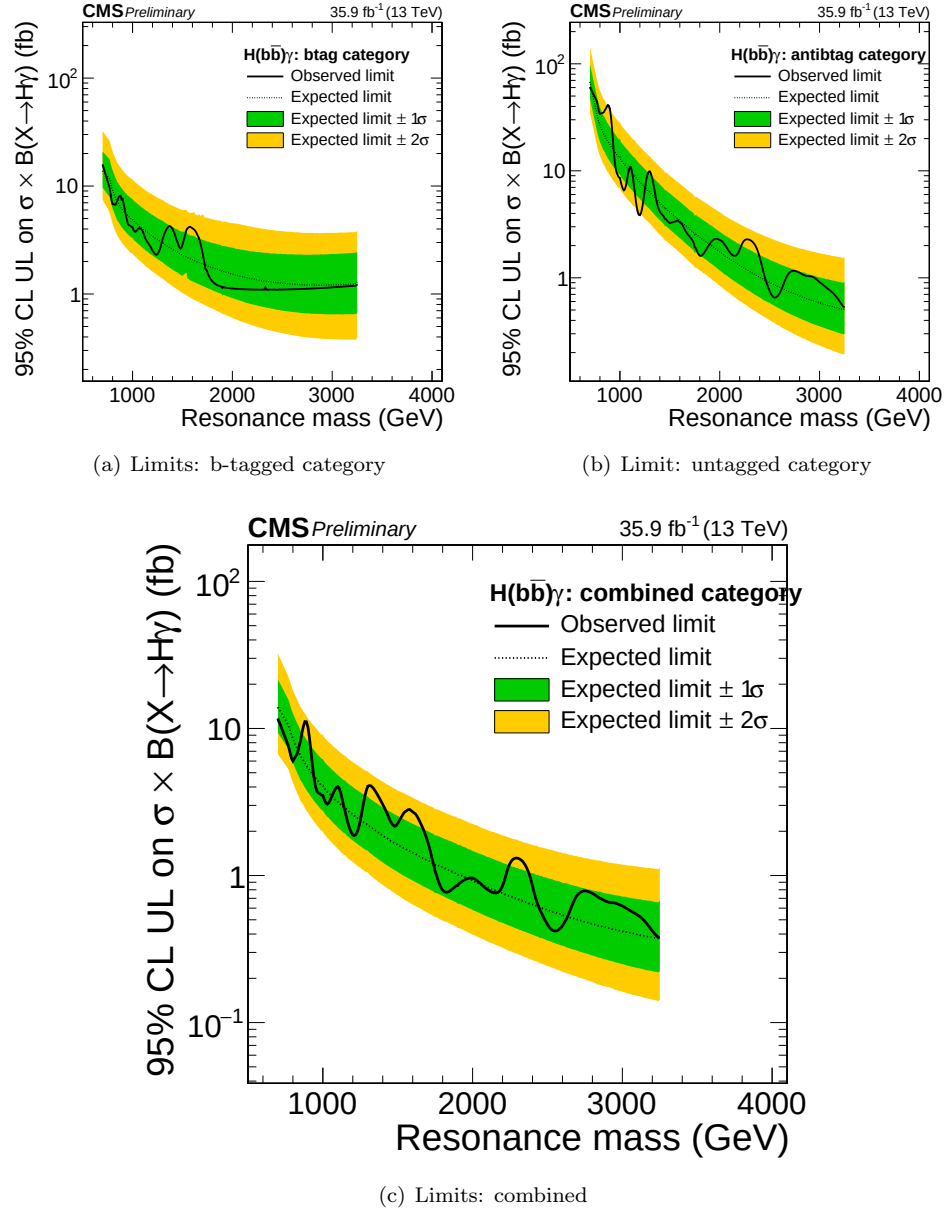
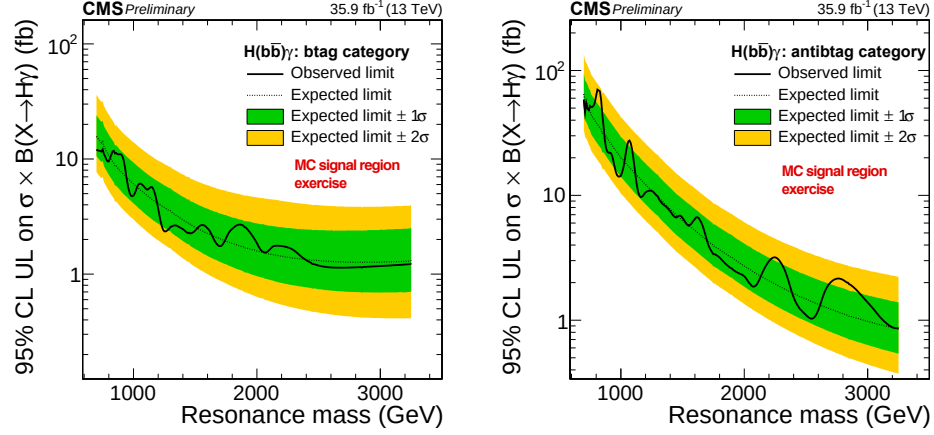
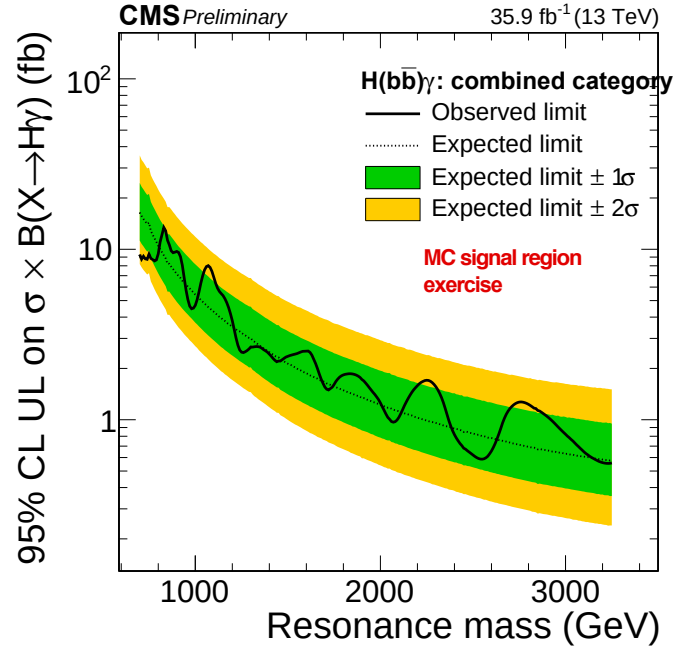


Figure 4.36: An exercise in calculating limits in untagged, b-tagged, and combined channels given as 95% confidence level limits on cross-section times branching fraction, where instead of performing the analysis strategy on the signal region data spectrum, the signal+background fits are performed on the sideband (i.e. the PUPPI softdrop jet mass is required to be between 100 GeV and 110 GeV) invariant mass spectrum. The background fit is derived from this sideband spectrum, while the signal shapes are derived from MC using the signal window (jet mass between 110 GeV and 140 GeV).

Shown in Figure 4.38 is a similar exercise in the limit-setting procedure, where limits are set by running the limit setting procedures on the $50 \text{ GeV} < m_J^{\text{PUPPI+SD}} < 70 \text{ GeV}$ sideband.

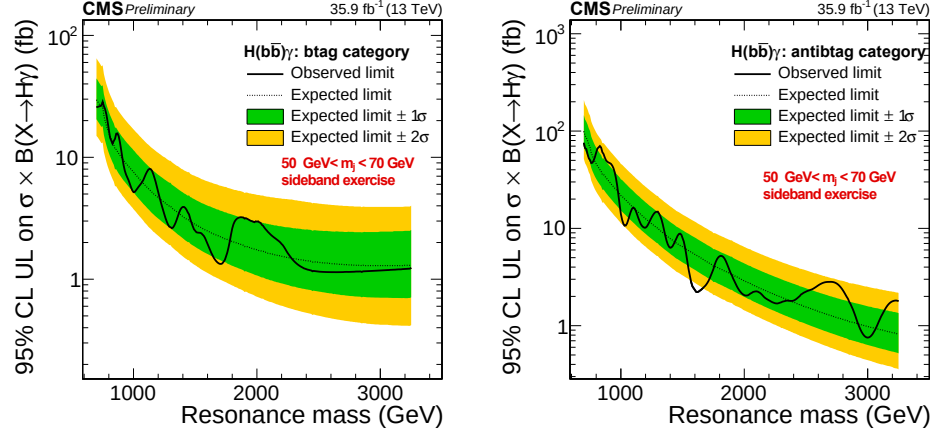


(a) Limits: b-tagged category, using MC back-grounds (b) Limits: untagged category, using MC back-grounds

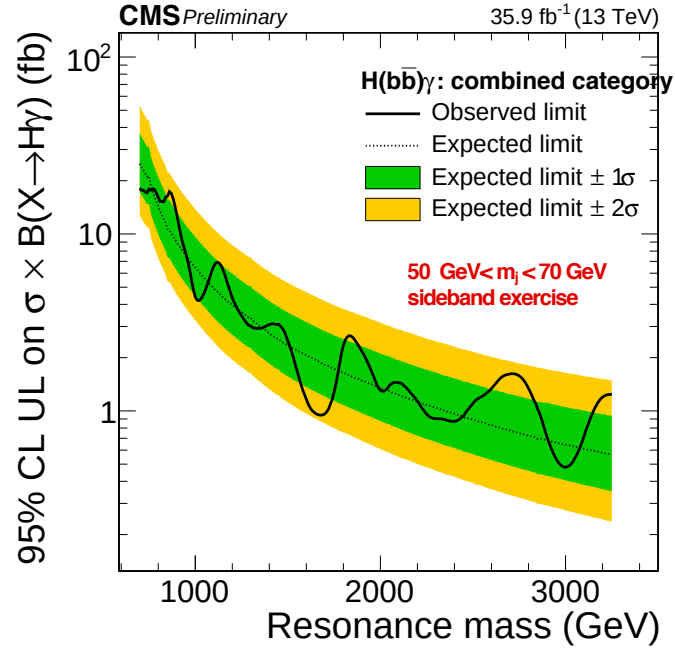


(c) Limits: combined, using MC backgrounds

Figure 4.37: An exercise in calculating limits in untagged, b-tagged, and combined channels given as 95% confidence level limits on cross-section times branching fraction, where instead of performing the analysis strategy on the signal region data spectrum, the signal+background fits are performed on the MC background samples' invariant mass spectrum in the signal region.



(a) Limits: b-tagged category, using the 50-70 GeV jet mass sideband (b) Limits: untagged category, using the 50-70 GeV jet mass sideband



(c) Limits: combined, using the 50-70 GeV jet mass sideband

Figure 4.38: An exercise in calculating limits in untagged, b-tagged, and combined channels given as 95% confidence level limits on cross-section times branching fraction, where instead of performing the analysis strategy on the signal region data spectrum, the signal+background fits are performed on the MC background samples' invariant mass spectrum in the signal region.

4.14 Lower bound of search region

In order to carry out the bump-hunt strategy used in the search for $h\gamma$ resonant production, a vital assumption is that the background can be estimated in the $m_{j\gamma}$ spectrum using a smooth, monotonically falling background fit function. However, it is observed that the $m_{j\gamma}$ spectrum manifests a turn-on effect in which the choice of $m_j^{\text{PUPPI+SD}}$ jet mass window affects the shape of the turn on. Because of the correlation between the jet mass and the jet's transverse momentum, a mass window requiring more massive jets will cause the $m_{j\gamma}$ spectrum to peak at a higher mass. Because the analysis strategy applies only in the region where the background is monotonically falling, the analysis should be carried out at points in the $m_{j\gamma}$ spectrum where we are beyond the effects of this turn-on. The turn-on effect can be observed in the MC background samples, as shown in Figure 4.39.

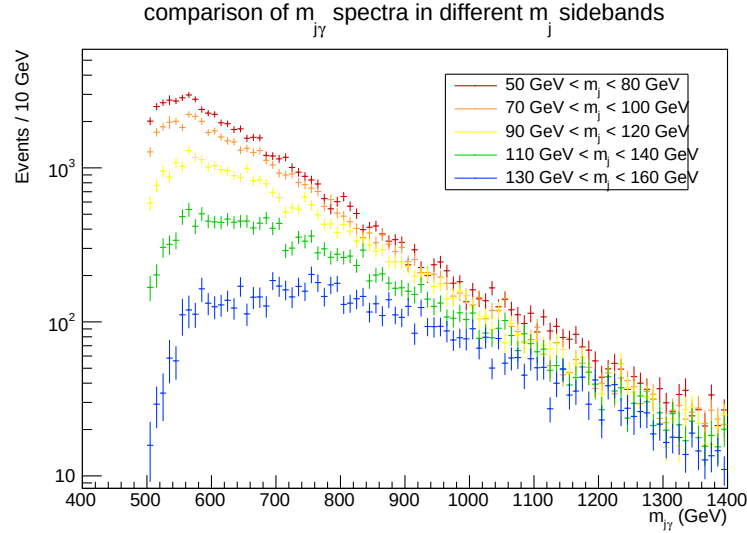


Figure 4.39: A demonstration of the kinematic turn-on effect observed, shown in the MC background samples, with the full analysis cuts except for the b-tagging categorization. In this plot, one can observe that as the $m_j^{\text{PUPPI+SD}}$ jet mass window is slid toward higher jet masses, the peak in the $m_{j\gamma}$ moves toward higher masses. The lower bound of the analysis' search region is chosen such that the bump-hunt analysis strategy is carried out only in the region beyond the turn on corresponding to the nominal $m_j^{\text{PUPPI+SD}}$ mass window.

Therefore, the lower bound of the analysis' search region was chosen using a method which quantifies the extent of the turn-on in the $m_{j\gamma}$ spectrum, and identifies a point at which the turn-on effect is negligible, given the $110 \text{ GeV} < m_j^{\text{PUPPI+SD}} < 140 \text{ GeV}$ jet mass window of the signal region. This method is carried out in two steps, and uses the MC background samples throughout. In both steps, we consider a function which models the combined effects of the kinematic turn on and the smoothly falling behavior of the background.

To model the turn-on, we assume that the turn-on shape is due to a large number of (bounded) random variables, and therefore we assume that the turn-on has the shape of an error function $\text{erf}(x)$,

where we impose the boundary conditions that this function range from 0 (completely “turned off”) at low invariant mass to 1 (completely turned on) at high invariant mass. To model the smoothly falling background function, we use an exponential function $\exp(x) = ae^{m/b}$ for constants a and b . Note that this function is used instead of the nominal background function of the analysis, the dijet2 function, since that function $m^{c+d \log(m)}$ is not monotonically decreasing e.g. for parameters $c = 25$, $b = -2.5$. The full behavior is thus modeled as the convolution of these two functions: $\text{erf} * \exp$.

The lower bound of the search region is then determined in two steps. First, the $\text{erf} * \exp$ function is then used to perform a fit of the $m_{j\gamma}$ spectrum in 30 GeV wide m_j windows: (30 GeV - 60 GeV), (31 GeV - 61 GeV), (32 GeV - 62 GeV), ..., (149 GeV - 179 GeV), and (150 GeV - 180 GeV). An example of a fit is shown in Figure 4.40

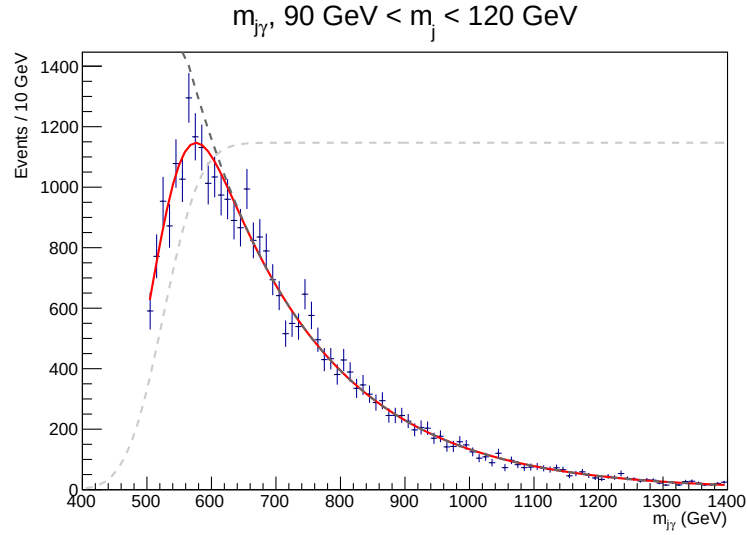


Figure 4.40: An example of a fit of the turn-on region using the $\exp * \text{erf}$ function. The full $\exp * \text{erf}$ fit is shown in red, and the $m_{j\gamma}$ distribution from the MC background samples is shown in blue. Also shown are the independent $\exp$ and erf components of the fit function: the dark dotted gray line is the pure $\exp$ component of the fit function, while the light dotted gray line is the erf component of the fit function, normalized so that its maximum is equal to the maximum of the MC $m_{j\gamma}$ spectrum. The cuts applied for this plot are the full analysis cuts, except for the b-tagging categorization.

After performing the $\exp * \text{erf}$ fits on each of the spectra of differing m_j windows, the fitted value for the coefficient b in the exponential term is extracted, and we plot the resulting values of the exponential coefficient b . To estimate the value of b for a given mass window, we perform a fit of this plot. The plot of the exponential coefficient b along with its fit is shown in Figure 4.41.

In the second step of this method, the $\exp * \text{erf}$ fits are repeated, but instead of allowing the value of the term b in the exponential component of the fit to float, b is fixed to be the value predicted from the fit shown in Figure 4.41. Then, using the resulting fit, we calculate the point at which the erf component of the fit is equal to .99, indicating that the turn-on behavior is within 1% of being fully turned-on at that point. An illustration of the determination of the 99% point is shown

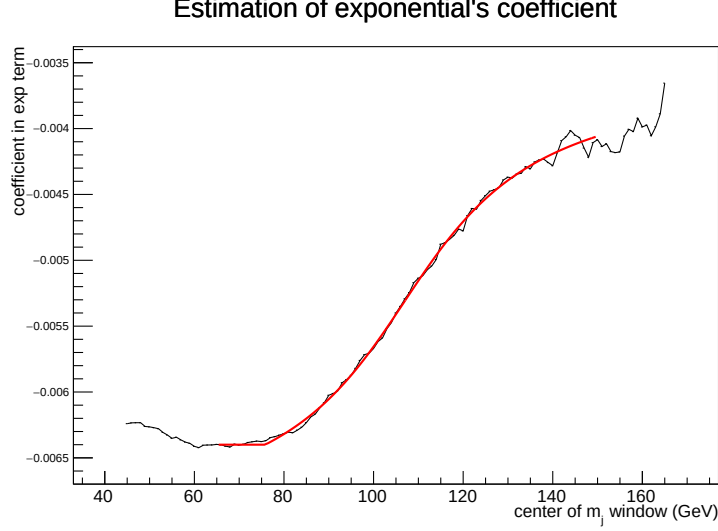


Figure 4.41: A plot of the exponential term's coefficient b resulting from the exp*erf fits as a function of the $m_j^{\text{PUPPI}+\text{SD}}$ mass window's center. This fit is then used to provide an estimate for the value of b in the second step of the method, where the fit is repeated but with the value of b fixed. This fit is performed in the MC background samples, using the full analysis cuts except for the b-tagging categorization.

in Figure 4.42.

Having calculated the 99% turn-on points for each of the 30 GeV wide $m_j^{\text{PUPPI}+\text{SD}}$ mass windows, we then plot the 99% turn-on points as a function of the center of the m_j mass window. A fit is performed on this plot. Finally, the 99% turn on point for the $110 \text{ GeV} < m_j^{\text{PUPPI}+\text{SD}} < 140 \text{ GeV}$ is extracted from the fit. The value found from this fit, i.e. the expected point at which the turn-on has reached 99% in the analysis' search region, is 717 GeV. This result is shown in Figure 4.43. Rounding to the nearest 10 GeV, we take the lower bound of the search region for this analysis to be 720 GeV.

4.15 Systematic uncertainties

Various systematic uncertainties in this analysis were taken into account in the limit setting procedure. These uncertainties stem from effects which may lead to a imperfect estimation of the signal rate and shape, including experimental uncertainties in the integrated luminosity [145], jet energy scale and resolution [146, 147], photon energy scale and resolution which varies depending on the $J\gamma$ mass [42], pileup, groomed jet mass scale, and various identification efficiencies, which are dominated by the double b-tagger efficiency uncertainty [124]. Also accounted for is an uncertainty in signal acceptance due to the PDF choice, which is based on the PDF4LHC recommendations [148] using the NNPDF3.0 replicas [126]. Since the correction for the trigger inefficiency in data never

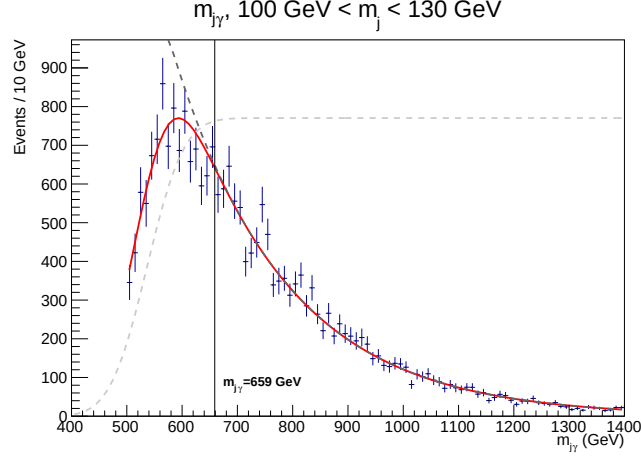


Figure 4.42: A plot of the calculation of the point at which the $\exp * \text{erf}$ function reaches the “99% turn-on point,” defined as the point at which the error function component reaches .99. The vertical line indicates the point at which the error function is equal to .99. This fit is performed with a fixed value of the exponential term’s coefficient b , which is done to ensure that the determination of the 99% point is not influenced by the “bending” of the exponential component of the fit function. This plot is performed using the MC background samples, using the full analysis cuts except for the b-tagging categorization.

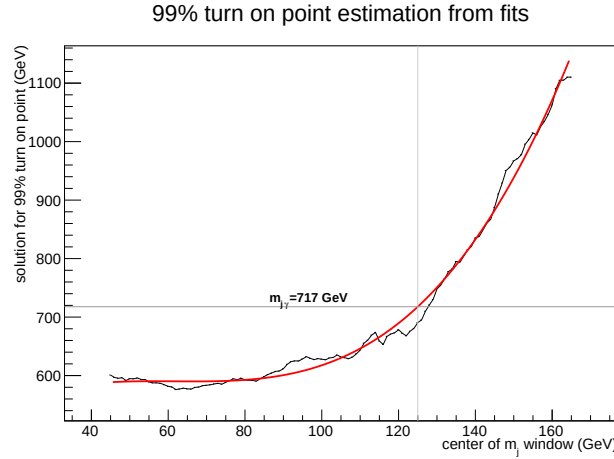


Figure 4.43: The final determination of the lower bound of the analysis’ search region. The 99% turn-on points as determined using the method shown in Figure 4.42 are plotted as a function of the center of the $m_j^{\text{PUPPI+SD}}$ mass window. A fit is made of this plot, and the lower bound of the search region is taken as the 99% turn-on point in the analysis’ $110 \text{ GeV} < m_j^{\text{PUPPI+SD}} < 140 \text{ GeV}$ mass window. This study is done on MC background samples with the full analysis cuts except for the b-tagging categorization.

exceeds 2%, the uncertainty due to this correction is always much smaller than the statistical uncertainty of data and therefore has been ignored. These systematic uncertainties are summarized in

Table 4.12. Because the background is estimated from a fit of the signal region, the only uncertainties in the background estimate stem from possible bias in the background function choice, but as shown in Figure 4.27, this effect is small compared to the statistical uncertainty on the background normalization.

Luminosity	2.5%
Jet energy scale+resolution	2.0%
Photon energy scale+resolution	0.1–2.3%
Pileup	1.0%
PUPPI+SD jet mass scale	5.0%
Double-b tagger efficiency	4.0%
Parton distribution function	2.0%

Table 4.12: Systematic uncertainties

The systematic related to the uncertainty on the double-b tagger scale factors was derived as follows. Using the double-b tagger scale factor uncertainties supplied by the b-tagging POG, the signal efficiency was calculated at each of the fullsim-generated MC signal samples with the b-tagging scale factor changed to its nominal value plus its uncertainty, as well as with the b-tagging scale factor changed to its nominal value minus its uncertainty. The signal efficiency with the scale factor varied upward is denoted ε_{up} , and likewise, the signal efficiency with the scale factor varied downward is denoted ε_{down} . The fractional change in the efficiency was then calculated as $(\varepsilon_{up} - \varepsilon_{down}) / \varepsilon_{nom}$, where ε_{nom} denotes the signal efficiency using the nominal double-b tagger scale factor. Shown in Figure 4.44 are the fractional changes in the signal efficiency for each of the fullsim generated points. The observed fractional difference between signal efficiencies in the b-tagged category with upward- and downward-varied b-tagging scale factors was observed to be between 7.4% and 7.8%, which implies a systematic uncertainty between 3.7% and 3.9%. Therefore, the systematic was conservatively taken to be 4% for all signal masses in both categories.

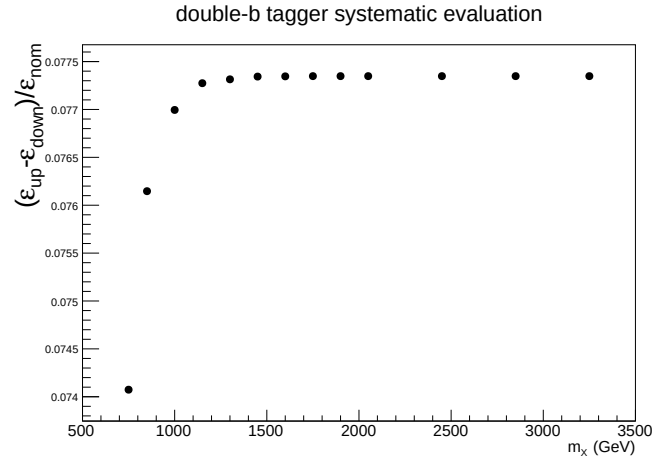


Figure 4.44: Changes in signal efficiencies due to varying the double-b tagging scale factors up and down by their uncertainty. Plotted is the difference of efficiency with the scale factor varied upwards by one uncertainty minus the efficiency with the scale factor varied downwards by one uncertainty, divided by the nominal efficiency. As the fractional difference is less than 8%, the systematic uncertainty is conservatively taken to be 4%.

4.16 Signal+background fits

Shown in Figure 4.45 are the results of a $S + B$ fit to the $H\gamma$ invariant mass spectrum, shown alongside the nominal background-only fit and the one- and two-sigma uncertainty bands derived via the sampling method described in Section 4.7. The $S + B$ fits shown in Figure 4.45 are performed without incorporating any nuisance parameters to incorporate systematic uncertainties, and are thus shown as a guide for interpreting the limits in Section 4.17. It can be seen that the $S + B$ fit does find regions of both positive and negative signal significance in both categories where bump-like features are present; however, it is also apparent that the deviations from the background fit are in general consistent with the statistical and systematic uncertainties associated with the background-only hypothesis.

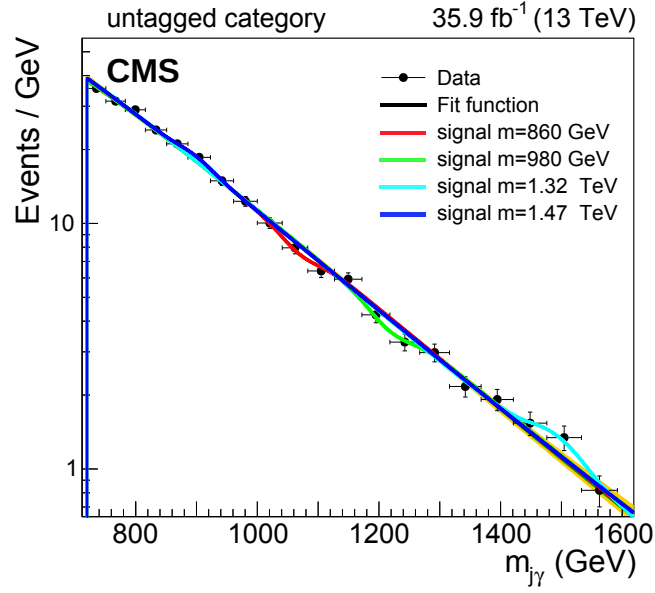
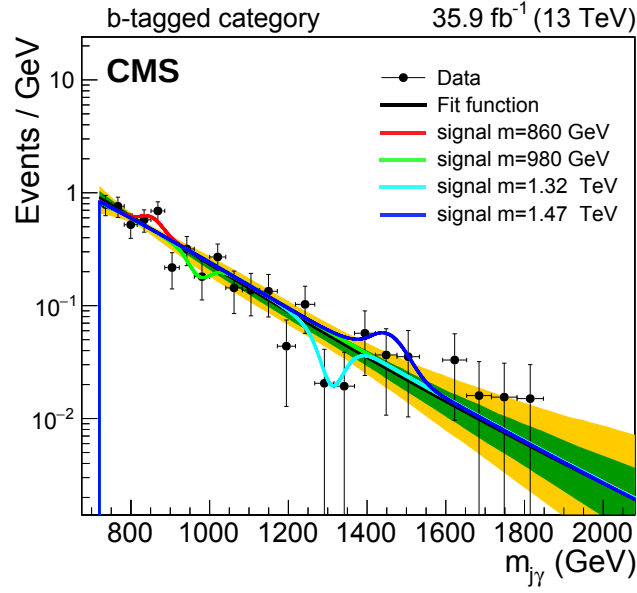
(a) $S + B$ example fit: untagged category(b) $S + B$ example fit: b-tagged category

Figure 4.45: Examples of signal+background fits obtained. Note that these $S + B$ fits are purely for demonstration purposes, since the actual limit setting procedure proceeds involves fitting many spectra that have had nuisances applied to account for systematic uncertainties. The signal masses used in the $S + B$ fits shown in these figures are chosen because they represent where the observed signal strength was the largest.

4.17 Results

Upper limits are set on the production cross section of narrow spin-1 resonances using the modified frequentist CL_S criterion [149–151], with a likelihood ratio used as a test statistic, and uncertainties incorporated as nuisance parameters with log-normal priors, as implemented in the Combine tool. Results are shown in Figure 4.46. In the case of setting limits on the combined category, the Combine tool’s limit setting procedure considers both categories in a simultaneous fit for both categories; for details on this procedure see Refs. [151] and [152]

From these results, the observed limit is compatible with the background-only hypothesis within two sigma at all points in the mass range.

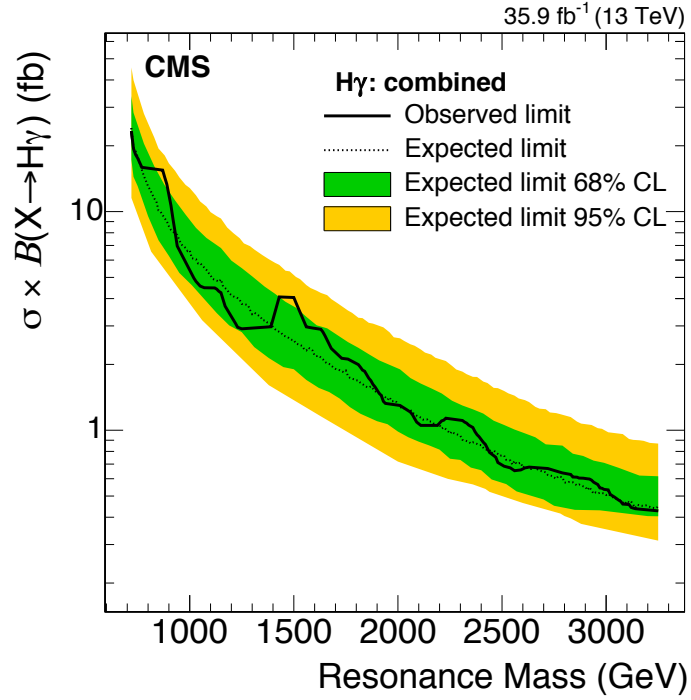
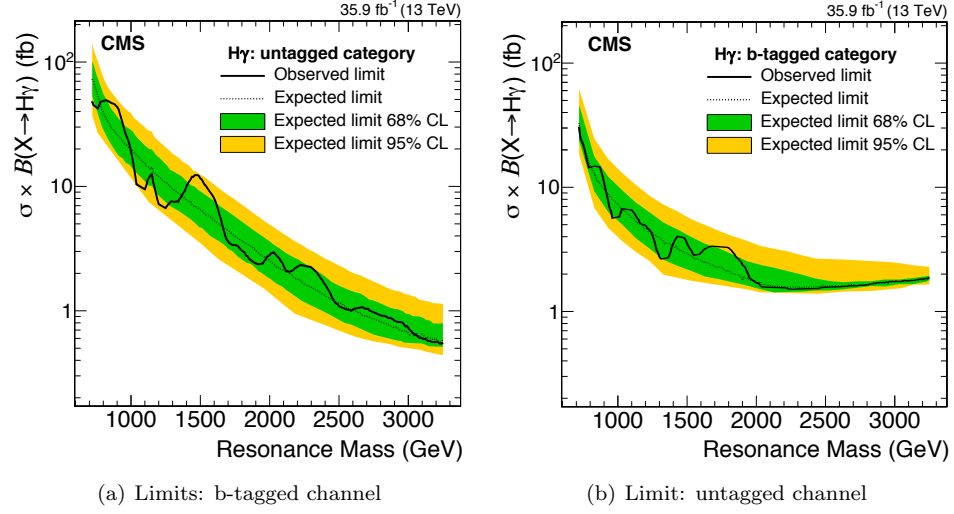


Figure 4.46: 95% confidence level upper limits on cross section times branching fraction for a narrow resonance decaying to $h\gamma$, calculated in the untagged, b-tagged, and combined channels.

4.18 Conclusions

Figure 4.46 shows the results of the $H\gamma$ analysis, the first analysis of this channel by the CMS collaboration. Stringent upper limits on production cross section times branching fraction to $H\gamma$ for a narrow spin-1 resonance are set at 95% confidence level between 20 and 0.3 fb for a resonance mass between 0.72 and 3.25 TeV. As the SM predicts no resonances that decay to $H\gamma$ in this analysis' search region, this result can be interpreted as validating the SM and showing no significant evidence of BSM phenomena. Comparing the limits derived in the untagged and b-tagged categories, it is clear that the dramatic background suppression that the application of b-tagging benefits the sensitivity of the analysis in the low mass region, up to about 2 TeV. Beyond this region, where the background in both categories becomes negligible, the untagged category lends the most sensitivity in this analysis, and the novel approach of using both b-tagged and untagged categories combines the strengths of both categories.

Comparing with recent results from the ATLAS search for $H\gamma$ resonances [153], which only considered a b-tagged category in their search, these limits significantly exceed the limits set throughout the search region considered by the ATLAS analysis. The analysis reported here also considers a significantly extended search region – below 1 TeV and above 2 TeV these limits are the only ones available.

Appendix A

HCAL data acquisition

As described in Section 3.2.1, CMS’s subdetectors, as well its triggering and data acquisition systems, comprise a hugely complex and very heterogenously structured system which must be operated in a very coordinated and synchronized way in order to ensure high-performance data taking. This requires a control system that is not only able to reliably get the detector operating whenever the LHC begins to collide its beams, which can often follow a very chaotic schedule, but also able to quickly and reliably handle any problem occurring during datataking and recover CMS back to a functioning state. On the other hand, while the LHC is not providing collisions, the detector does not simply sit idle waiting for the next collisions; CMS is constantly collecting data for other important studies. For example, this includes taking various kinds runs for monitoring aging effects in the detector so that appropriate corrections to calibrations can be derived, collecting events where cosmic muons pass through the detector (which are an important tool for studying the alignment of the detector and validation of particle-flow reconstruction), taking runs to commission new or upgraded components, *etc.* In addition to running all the subdetectors of CMS in the normal holistic “global” manner, the subdetectors are often operated in a stand-alone or “local” manner, meaning runs that are taken with only one subdetector. Local runs are an essential tool for the upgraders and maintainers of the subdetectors to do their work without impacting the workflow of the other subdetectors; as an example, the task of synchronizing the readout of the HCAL (i.e. making sure every channel is reporting data from the correct bunch crossing, and that the bunch crossing boundaries are timed to frame the actual collisions appropriately) involves finding a timing calibration for each of the many thousands of channels, and this requires taking many short runs, adjusting the timing, inspecting the data, and repeating.

Timing adjustments are just one example of the very large number of “configuration parameters” that are sent to the readout electronics; there are many other parameters relevant to the operation of the detector, some of which are not just calibration adjustments but instead are used to put the electronics into different modes of functioning; for example, the HCAL’s RBXes can be put into a mode that changes the RBX from reading out the ADC values of its QIE chips to reading out a test pattern serves as that channel’s “finger print”, which is used for validating the fibering of the

connections between the RBXes or ROBOXes and the back-end uTCA hardware (called FiberID mode). Overall, the detector can be run in countless different ways, corresponding to different subsets of detector components in the run as well as different configuration parameters. Therefore, operating CMS requires a considerable amount of “online software” with the power to configure the huge amount of hardware and firmware with the right configuration parameters, with the reliability to handle the unpredictable schedule of LHC beams and unforeseen problems, and with the usability that enables its operators to flexibly run the detector and its individual sub-components in the myriad of different possible ways.

Because of the fact that the subdetectors are to a large degree independently operable, and because they have different subdetector-specific modes of running for specific studies, the framework of CMS online software is designed in a modular way reflecting the detector’s modularity. Each subdetector has a significant amount of custom online software to implement the desired modular functionality, and each is a valuable case study in the design of detector control systems. This Appendix describes the how the HCAL subdetector is controlled and monitored in the global and local modes by means of the “Run Control and Monitoring System” (RCMS), xdaq, and other kinds of online software, and describes work done over the last several years to enable more effective and flexible datataking with the HCAL, and serves to demonstrate the kinds of practical solutions that serve to handle the challenges of operating large-scale detectors.

A.1 Overview

The general way CMS is operated, from the “top-down” perspective (as opposed to the “bottom-up” approach in Section 3.2.1) is that during nominal data taking (during global running), the detector is operated by a rotating crew of people called “shifters” who typically work around the clock above-ground in the CMS control room, which neighbors the Point 5 assembly hall above the experimental cavern. The shifters are CMS collaborators, but typically have no particular expertise in the control software or detector operations aside from rudimentary training and a $\sim$ week of shadowing other shifters, and in order to operate the detector, they use a web-based interface (accessible only to those with permission to log into the CMS private network) that presents them with the high-level functionalities of running the detector – for example, starting the control software, configuring the detector, running the detector, stopping the detector, and recovering from potential issues – as well as providing real-time information about the status of the detector. The interface is designed to be intuitive enough for non-experts to use effectively but also to provide the flexibility requisite for the many modes of operating the detector. This high-level interface is accessed in a web browser like any normal website; i.e. it is implemented in HTML, JSP, JavaScript, and CSS, and it serves as the “front-end” (in the web development sense) for interacting with the “back-end” of the web application. The back-end web application is a CMS-developed framework written in Java called the “Run Control and Monitoring System” (RCMS). An image of the RCMS web interface that shifters use during global running is shown in Figure A.1.

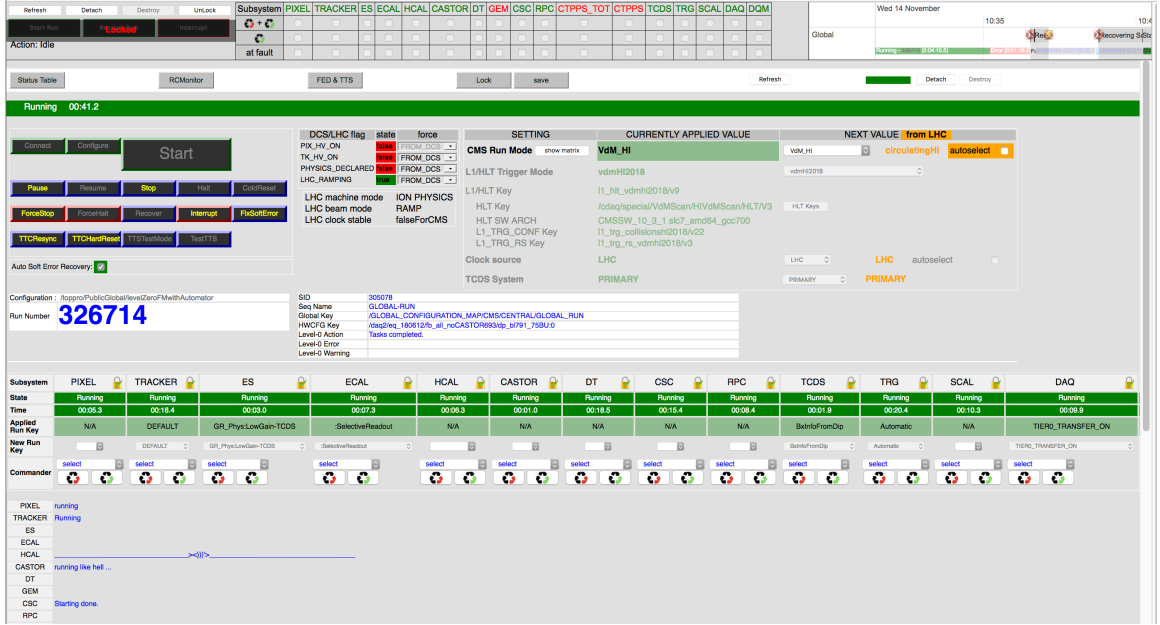


Figure A.1: A screen capture of the interface used by shifters to control CMS during normal data-taking. This is the graphical user interface for the level-0 function manager, which is the highest-level function manager of the hierarchy of function managers.

RCMS is based around Java servlets, which allows stateful representation of (the software operating) the overall detector as well as each of its sub-components. The states include, for instance, “Configuring”, “Configured”, “Starting”, and “Running”. The objects within the RCMS framework that are used to control all subdetectors as well as the overall CMS detector are called “Function Managers” (FM), which implement the RCMS state machine, shown in Figure A.2, and implement the subdetector-specific functionalities needed for controlling and monitoring lower-level software, referred to as “resources”. The resources include, for instance, programs which communicate with the readout electronics, intermediate level programs that manage other programs, and child function managers. As an example, the interface that the shifters interact with during global running to control the entire CMS detector is the interface to the “level-0 function manager”, and the resources it controls, called “child resources”, are all RCMS function managers, each of which controls a “subsystem” corresponding to different subdetectors or the trigger system or the TCDS system, *etc.* Each subsystem is represented in the horizontal grid in the middle of the interface shown in Figure A.1. The function managers that control each subsystem are referred to as “level-1” function managers, and each subdetector implements their function manager with the desired functionality for that subdetector.

Function managers, such as the level-0 FM, are dynamically created and gotten rid of (“destroyed”) by the RCMS Java web application (webapp), which runs in an Apache Tomcat servlet container, and runs as a service on the relevant detector control computers at P5. The information

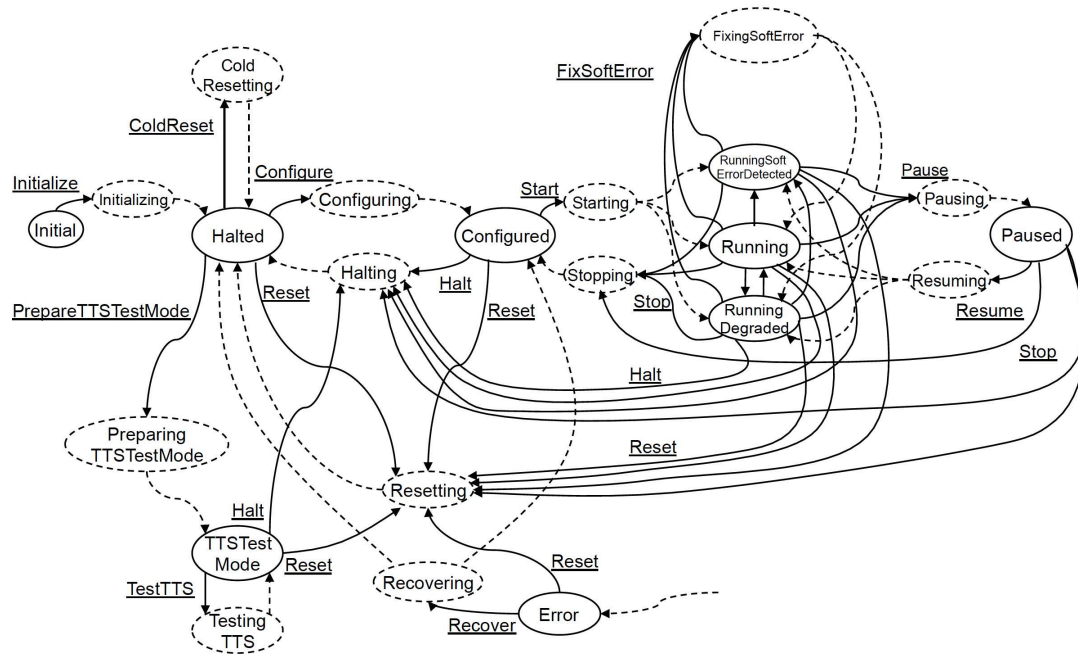


Figure A.2: The RCMS state machine. This defines the states that the level-0 FM and all level-1 FMs may have. The solid ovals represent steady states, which the FM may remain in indefinitely; the dashed ovals represent transitional states, which are temporarily passed through to move to the next state, and the underlined words are the commands which initiate the state transition [15]. Some states in the diagram are not used by the HCALFM, but would be allowed if a purpose arose for them, including FixingSoftError and RunningSoftErrorDetected (which are used primarily by the tracker and ECAL function managers). The main steady states used by HCAL are Initial, Halted, Configured, and Running, and the RunningDegraded state is used for alerting shifters to serious problems with HCAL.

that defines how the RCMS webapp should create one or more function managers – including e.g. the location of the Java archive file (“jar”) that contains the compiled code of the level-1 FM, the structure of its child resources (collectively known as the function manager’s “qualified group”), and initialization parameters for the function managers and child resources – is served by an Oracle database hosted on computers at P5, called the “Resource Service database” (RS database). This full set of information describing a set of FMs, their qualified groups, and all associated parameters is collectively referred to as a “configuration”; an RCMS user, upon logging in, is presented a menu of different available configurations to select amongst, where different configurations could correspond to e.g. different subsets of the detector or different configuration parameters. A separate schema within the oracle database, called the “RunInfo database“, is a database that FMs can write archival information to, such as information about particular hardware configuration parameters and conditions read back from the detector.

A level-1 FM itself may have a hierarchy of controlled resources in its qualified group, which

are typically either other function managers (level-2 function managers) or detector-control software implemented within the xdaq framework. xdaq is a CMS-developed framework implemented in C++, and is intended to be suitable for low-level communication with the detector electronics and the operation of the DAQ subsystem. For instance, The HCAL function manager (HCALFM), which is used for operating the HCAL as well as for operating test setups of HCAL hardware (e.g. at the Building 904 integration facility and H2 test beam area, both at CERN, and at the WH14 HCAL upgrade laboratory at Fermilab) is designed such that every configuration must have one level-1 HCALFM that controls at least one level-2 HCALFM, and the level-2 HCALFM controls resources that are all xdaq applications. A schematic of an HCALFM within a global configuration is shown in Figure A.3. Level-2 function managers spawn and destroy xdaq software through an xdaq service called JobControl. JobControl runs as a service on all the machines involved in HCAL data taking, and is responsible for spawning the Linux processes that contain the xdaq applications, known as xdaq executives. The xdaq processes spawned by JobControl are dictated by XML configurations generated by the HCALFM (based on what is dictated in the configuration plus other forms of user input) called “executive XML”, which is sent to JobControl via a SOAP message. The executive XML dictates what xdaq applications should be run within an executive, various user-defined parameters such as the desired port for the executive to listen to TCP/IP requests on, and, importantly, information about what other xdaq applications are to be “connected” with a given xdaq application.

As an example of these “connections,” one important instance of these connections defined in executive XML are those that appear in the hcalSupervisor application. The hcalSupervisor application is a privileged xdaq application because it serves as the point-of-contact between the HCALFM and the rest of the xdaq applications, except for the steps of the HCALFM initializing and terminating all the xdaq executives via JobControl. There is one hcalSupervisor per level-2 HCALFM, and each application in the configuration has a connection defined in its executive XML to exactly one hcalSupervisor. The supervisor receives configuration parameters dictated by the HCALFM and propagates them to all the applications involved in configuring the hardware, and then monitors them to compute an overall state of the applications it is connected to. The hcalSupervisor’s state is then reported via asynchronous notifications to the HCALFM, with which the HCALFM bases its state transitions through the states of the state machine.

When a subsystem in CMS being controlled by the level-0 FM is (temporarily) not needed for datataking or needs to be removed from the global run, for whatever reason, or if the subsystem is in an irrecoverable error state, the level-0 FM sends it a “Destroy” command, and the level-1 function manager for that subsystem carries out the tasks of destroying all child resources (e.g. level-2 FM’s and their child resources), leaving the system in a state where all processes that should no longer be running are gone, and then the function managers terminate themselves. For HCAL, the level-1 and level-2 HCALFMs destroy all of the HCAL xdaq processes that are relevant for configuring and running the detector, such as xdaq applications that configure uHTRs, AMC13s, the electronics that manage operating the laser for abort-gap operations, etc. This design ensures that the software

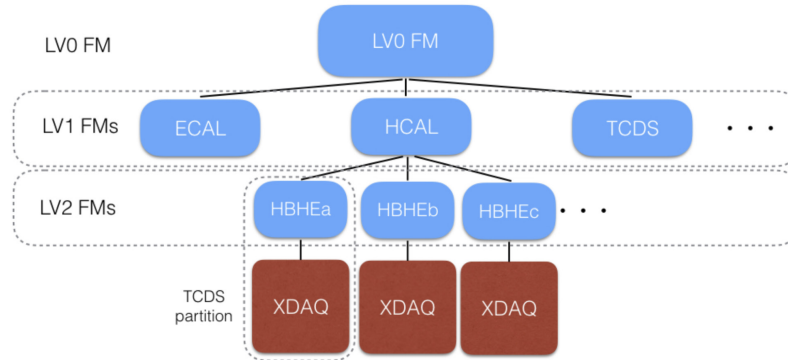


Figure A.3: A schematic of the level-0 FM containing a level-1 HCALFM, which in turn controls level-2 HCALFMs that control xdaq applications.

crucial for operating the detector can be completely stopped and restarted in the event of any error state or software malfunction. There are various other HCAL pieces of software that run as long-lived services, such as HCAL’s Detector Control System (DCS), which handles safety-related aspects of operating the detector (such as controlling powering of the system and automatic shutdowns in the event of smoke or water), as well as applications for monitoring the hardware status and sending alerts to HCAL experts in the event of issues. However, the HCAL online software design principle is that any software that is important for data-taking and does not need to run continuously for supporting the DCS system should be terminated completely at the end of use of one configuration and recreated from a “ground state” at the beginning start of the life-cycle of a new configuration.

A.2 Local runs

If the HCAL has been destroyed in the level-0, then the HCALFM can operate to take data from the HCAL in local runs. This mode of running differs significantly from how datataking works in global mode, because the data is not propagated to the DAQ system’s FEROLs; instead, the data gets read out from the AMC13’s monitor buffer by software, which is propagated over the network to the HCAL-specific EventBuilder application. This mode of running is primarily used when there are not stable beams delivering collisions to CMS for calibration, commissioning, and debugging studies for CMS, and is the primary mode of operation of HCAL test setups at the teststand facilities involved in prototyping, testing, and “burning-in” of upgrades to HCAL electronics.

The xdaq applications which gather event fragments from the monitor buffer are called readout applications, such as the DTCReadout (which gathers data from the AMC13 monitor buffer), the DCCReadout (which gathers data in a similar way from the DCC boards of HO), and the hcalQD-CTDCReadout, which reads out specialized hardware used for triggering at the H2 testbeam facility. The events collected in this manner must be triggered with an L1A as per usual, but in local running, no inputs from the level-1 trigger are received. Instead, a specialized application called the

TriggerAdapter sends a request to the TCDS system (e.g. at P5 or 904) to issue a burst of triggers, typically 100 triggers, to the AMC13s. At other teststands, where a TCDS setup is not available, either other triggering hardware like the TTCci or TTCvi VME boards are used in a similar way, or, at Fermilab, the TriggerAdapter communicates with the DTCManager and the AMC13 is operated in a special manner such that it can generate its own bursts of triggers.

The trigger adapter keeps track of the delivery of the data to its destination – the EventBuilder application – through a system which is similar to the exchange of “trigger credits” which function as semaphores for the status of each event. For each trigger in a given burst, the TriggerAdapter starts off for one trigger credit per trigger. For each trigger it requests to be issued, it expends one trigger credit, and ceases to request more triggers after the trigger credits are expended. After the triggers are issued, the AMC13s will all collect the data from all the uHTRs in their crate and store it into their monitor buffer, and this exposes the data to the DTCReadout application (and analogously for other readout applications). The readout applications then send these data fragments in the form of unsigned TCP packets over TCP/IP to the EventBuilder application. The EventBuilder then waits until all the data from all the triggers in the burst have been collected, at which point it sends another batch of trigger credits to the TriggerAdapter. The data collected by the EventBuilder is then transferred to an application which stands in for the normal HLT filter units, called the TrivialFU. These three applications, the TriggerAdapter, EventBuilder, and TrivialFU, are responsible for triggering and collecting the data, which means there can only be one of them present in the configuration for a given run, are therefore referred to as the “EvMTrig” (event manager and triggering) trio. The EvMTrig trio has a software requirement that it be run in the same xdaq executive as a DTCReadout, and therefore it must be included in an executive that corresponds to some AMC13 present in a given run.

A.2.1 TCDS

One main functionality that is utilized heavily by the HCAL operations team is to operate various portions of the detector independently. There are several constraints on the granularity at which this can be achieved. One very important constraint is imposed by the supply of clock and fast control signals (triggers and BGos) from the TCDS hardware (or, for certain teststands, the older TTCci/vi hardware). TCDS, as a subsystem, has its hardware and software designed around the principle that during global runs, the Central Partition Manager (CPM) supplies clock and fast control signals to all the downstream hardware, such as the Partition Interfaces (PI); however, subsystems that are not in a global run can use the TCDS Local Partition Manager (LPM) to supply an alternative set of clock and fast control signals to multiple Partition Interfaces (PI), or the PI can be operated in standalone mode to supply clock only to its associated hardware. The set of hardware that a single PI drives clock and fast control signals for is known as a TCDS partition; HCAL consists of three partitions for the HCAL barrel and endcap (HBHEa, HBHEb, and HBHEc), as well as one partition each for the HO and HF, plus another partition that serves the laser hardware and various electronics used for noncritical studies. This provides a natural granularity for the detector, and thus

one level-2 HCALFM is used for each partition. Therefore, in principle it is possible to operate each of these partitions independently, or one cluster of partitions running together with other partitions running independently, in addition to running in unison. However, prior to late 2015, for all of these different modes of collecting local run data would have required the manual creation of a new configuration in the RS database to correspond to every possible mode, due to the necessity of having to locate it inside an executive corresponding to an AMC13 present in one of the partitions being used. Development on the HCALFM since that time has improved the situation greatly by providing much more flexible methods for operating the detector, both in terms of operating the detector with arbitrary sets of available partitions, and also in terms of various other aspects, discussed below.

A.3 Configuration parameters

One significant improvement that has been implemented in the HCALFM is the improvement of how users set particular configuration parameters for the HCAL. For example, users may want to use the calibration units' LEDs to check photodetector response, or they may want to not use the LEDs to measure the pedestals from the photodetectors and readout electronics; users may want to update the lookup tables (LUT) used in the uHTRs' calculations of trigger primitives; and users may want to enable the FiberID mode. In the past, it was necessary to generate a new configuration in the RS database for each of these types of runs, since the unix paths to the text files that stored these parameters (called "snippets") were stored in the RS database, and editing of the RS database is a serious impediment to many users. This unfortunate situation, which led to the proliferation of dozens of mostly-identical configurations being created laboriously for any minor change, was improved by implementing in the HCALFM and its graphical user interface a system of "Local Run Keys", that allow a drop-down menu selection of what different options for local run modes, and a simple command-line method to create new local run keys that would automatically populate the user interface without any need for edits to the RS database. An image of the local run key interface is shown in Figure A.4.

The text files that dictate what local run keys are available to users are version-controlled XML files called "grandmaster snippets" that are validated by XML schema declarations (XSD), helping to ensure they are syntactically parsable using Java's in-built DOM and JAX XML parsers. The grandmaster snippets define the available set of local run keys by their name, a second snippet that it references called the "mastersnippet", and various other specifics of that particular local run mode such as whether any resources should be masked and how many events should be taken for that run. The mastersnippets, which correspond to each local run key, contain a list of other snippets containing the relevant configuration parameters for that particular kind of run, plus additional configuration parameters of its own; this mastersnippet gets sent to the hcalSupervisor, which then parses the information contained in the list of snippets as well as the info in the mastersnippet. Local run keys can also contain specific configuration parameters, to enable the reuse of a single

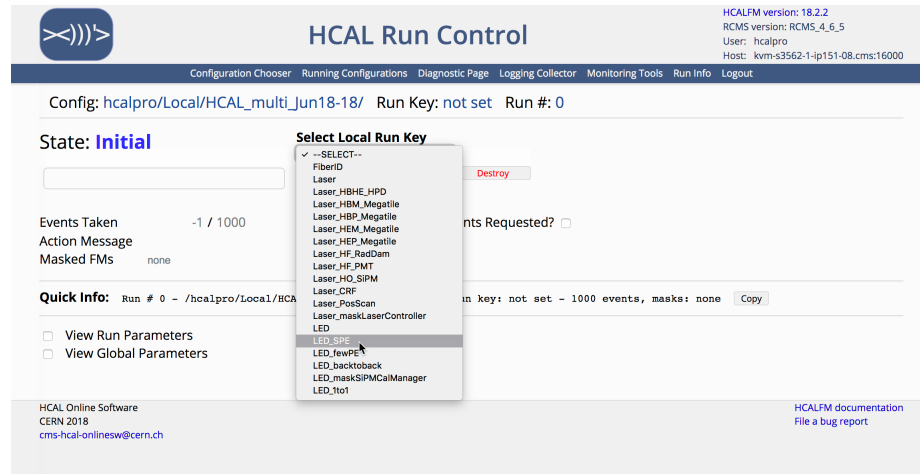


Figure A.4: A screen capture of the HCALFM graphical user interface showing the ability for users to select a local run key.

mastersnippet for many different closely-similar types of runs. This system has proved highly versatile for rapidly generating new kinds of local runs, greatly reducing the amount of time and effort needed to accomplish this over the prior system, and allowing those comfortable with manipulating text files via the command line but not with editing the Oracle database to prepare new run types.

A.4 Masking

As mentioned earlier, operating the HCAL or other systems with multiple TCDS/TTCci partitions (e.g. there are three at the Building 904 integration facility) is possible with any set of TCDS partitions. Furthermore, because the HCALFM has the ability to change the executive XML stored in the RS database, it is possible to do fine-grained edits to what xdaq applications are spawned in any given run. Since early 2016, developments in the HCALFM have enabled users to easily pick arbitrary sets of partitions and to eliminate applications/executives from being spawned for a given run, without needing to create new configurations in the RS database.

The masking of individual xdaq applications and executives is controllable by the user via a parameter in the local run keys stored in the grandmaster snippet. A user can easily edit this text file, and to remove an application from the configuration, one simply adds the application name (this is a unique name consisting of the application’s class name, followed by an underscore, followed by its instance number, which is an integer starting at 0 and labelling each application of that class, e.g. “uHTRManager.4”) to the list named “maskedapps”. The HCALFM parses this list and removes that application’s “context” from the executive XML of the executive in which it runs, and then identifies all other places in all executives’ XML that indicate connections to or from the masked application. This fine-grained approach to masking applications must therefore be performed prior to the executive being initialized by JobControl; which happens at a very early step in the RCMS

state machine, the Initial→Initializing→Halted transition, as can be seen in Figure A.2.

For masking entire partitions, the user is presented with a panel in the HCALFM graphical user interface, which has two panels, the “Multipartition” panel (shown in Figure A.5) and the “Single Partition” panel. The selection panels provide either a checkbox or a radio button for selecting what level-2 HCALFMs (which correspond to partitions) will be present in the run. Depending on the panel the user makes their selection in, a parameter is stored by the HCALFM that informs the server-side code to switch between two alternative sets of snippets with which to configure the TCDS service applications; if the user made a selection in the Single Partition panel, the snippets are chosen to inform the PI to not propagate clock and fast control signals from the LPM, but instead to generate its own. If the user makes their selection from the Multipartition panel, then the LPM will be used in that run and will supply clocks and fast control signals to the PIs to propagate to the detector electronics; because HCAL has (perhaps until recently) only had one LPM available, this implies that only one multipartition run can be taken at a time; however, a multipartition run can be taken in parallel with single partition runs for each of the other partitions.



Figure A.5: A screen capture of the HCALFM graphical user interface showing the multipartition masking panel.

The selection made by the user in the masking panel is then formed into a JSON list of the names of the level-2 function managers which are masked, which is sent to the HCALFM (server side). The HCALFM then analyzes the qualified groups of each of the masked function managers and forms a list of all the child resources and executives of the masked function managers. In addition, the qualified groups of each level-2 HCALFM are checked to see whether that level-2 FM is suitable to be one that contains the EvMTrig trio. In order to provide the ability for a single configuration to

operate any partition or set of partitions, the configurations are now all designed such that every level2 FM has its own EvMTrig trio. Therefore, the level-1 HCALFM picks an unmasked level-2 FM to be the one holding the EvMTrig trio, and further adds all the other level-2 HCALFM EvMTrig trio applications into the list of masked resources. This list is then used in the same fine-grained masking routine as is used for masking individual applications. This strategy enables a single configuration to run the HCAL with any set of partitions and allows different partitions to be run independently in parallel, provided only one multipartition run is being taken.

The main benefits provided by the local run keys and the masking functionalities of the HCALFM are the ability to greatly reduce the amount of work needed for maintaining many different configurations, since the number of configurations has been reduced from over 20 to just one. Duplicates of this one configuration can be used by different users at once taking runs simultaneously with different parts of the HCAL. With on-the-fly changing of what parts of the HCAL are in use, combined with a user-friendly interface for changing configuration parameters, HCAL operations and upgrades teams can operate much more effectively and with much less overhead.

Bibliography

- [1] N. Craig et al., “The unexplored landscape of two-body resonances”, [arXiv:1610.09392](#).
- [2] “Public CMS Luminosity Information ”, August, 2018.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [3] M. Hostettler et al., “Luminosity Scans for Beam Diagnostics”, [arXiv:1804.10099](#).
- [4] D. Nisbet, “LHC Operation in 2016”, January, 2017.
https://indico.cern.ch/event/580313/contributions/2359285/attachments/1396590/2135891/Operation_in_2016_v1_1.pdf.
- [5] L. Boyle, “Standard Model Of Particle Physics, Most Complete Diagram”, February, 2014.
https://commons.wikimedia.org/wiki/File:Standard_Model_Of_Particle_Physics--Most_Complete_Diagram.png.
- [6] B. A. Dobrescu, P. J. Fox, and J. Kearney, “Higgs-photon resonances”, [arXiv:1705.08433](#).
- [7] M. Brice, “View of an open LHC interconnection. Vue d’une interconnection ouverte”, (Oct, 2005).
- [8] “The CERN accelerator complex”, August, 2016.
<https://cds.cern.ch/record/2197559?ln=en>.
- [9] I. Neutelings, “Simple example of 3D axes with spherical coordinates.”, June, 2017.
https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates.
- [10] T. Sakuma and T. McCauley, “Detector and event visualization with sketchup at the cms experiment”, *Journal of Physics: Conference Series* **513** (2014), no. 2, 022032.
- [11] P. Paolucciet al., “CMS Resistive Plate Chamber overview, from the present system to the upgrade phase I”, *JINST* **8** (2013), no. 04, P04005.
- [12] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, [arXiv:1706.04965](#).

- [13] N. Tran, “The evolution of jets at the LHC”, <http://www.pha.jhu.edu/groups/particle-theory/seminars/talks/S15/Tran-slides.pdf>.
- [14] G. Landsberg et al., “Pre-approval of EXO-16-020: Search for resonant decays in $V(qq)\gamma$ and combination with $V(l\bar{l})\gamma$ ”, February, 2016.
https://indico.cern.ch/event/504234/contributions/1182309/attachments/1235581/1814571/preEX016020_LindseyGray_29022016.pdf.
- [15] CMS Trigger/DAQ Group, “Finite State Machine Model for Level 1 Function Managers”, 2009. http://cmsdoc.cern.ch/cms/TRIDAS/RCMS/Docs/Manuals/manuals/level1FMFSM_1_10_1.pdf.
- [16] L. M. Lederman, “The Value of Fundamental Science”, *Scientific American* **251** (November, 1984) 40–47, doi:10.1038/scientificamerican1184-40.
- [17] U. Amaldi and G. Hall, “Particle Accelerators: From Big Bang Physics to Hadron Therapy”. SpringerLink : Bücher. Springer International Publishing, 2014. ISBN 9783319088709.
- [18] Nobel Media AB, “Nobel Lecture – Ernest Lawrence”, 1939.
<https://www.nobelprize.org/prizes/physics/1939/lawrence/lecture/>.
- [19] Particle Data Group Collaboration, “Review of particle physics”, *Phys. Rev. D* **98** (Aug, 2018) 030001, doi:10.1103/PhysRevD.98.030001.
- [20] D. J. Griffiths, “Introduction to Elementary Particles; 2nd Rev. Version”. Physics textbook. Wiley, New York, NY, 2008.
- [21] K. G. Wilson, “Confinement of Quarks”, *Phys. Rev. D* **10** (1974) 2445–2459, doi:10.1103/PhysRevD.10.2445. [,319(1974)].
- [22] R. Davis, “Solar Neutrinos. II. Experimental”, *Phys. Rev. Lett.* **12** (Mar, 1964) 303–305, doi:10.1103/PhysRevLett.12.303.
- [23] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.
- [24] M. Apollonio et al., “Limits on neutrino oscillations from the CHOOZ experiment”, *Phys. Lett. B* **466** (1999), no. 2-4, 415–430.
- [25] A. Aguilar-Arevalo et al., “Improved search for $\nu_\mu \rightarrow \nu_e$ oscillations in the minibooone experiment”, *Phys. Rev. Lett.* **110** (2013), no. 16, 161801.
- [26] O. Miranda and J. Valle, “Neutrino oscillations and the seesaw origin of neutrino mass”, *Nuclear Physics B* **908** (2016) 436 – 455, doi:<https://doi.org/10.1016/j.nuclphysb.2016.03.027>. Neutrino Oscillations: Celebrating the Nobel Prize in Physics 2015.

- [27] LUX Collaboration, “Results from a search for dark matter in the complete LUX exposure”, *Phys. Rev. Lett.* **118** (2017), no. 2, 021303, doi:10.1103/PhysRevLett.118.021303, arXiv:1608.07648.
- [28] CMS Collaboration, “Search for dark matter and large extra dimensions in monojet events in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **09** (2012) 094, doi:10.1007/JHEP09(2012)094, arXiv:1206.5663.
- [29] S. Desai et al., “Search for dark matter WIMPs using upward through-going muons in Super-Kamiokande”, *Phys. Rev. D* **70** (October, 2004) 083523, doi:10.1103/PhysRevD.70.083523, arXiv:hep-ex/0404025.
- [30] Nobel Media AB, “Burton Richter – Interview”, August, 2008.
<https://www.nobelprize.org/prizes/physics/1976/richter/interview/>.
- [31] ATLAS Collaboration, “Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [32] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [33] J. F. Gunion, H. E. Haber, G. L. Kane, and S. Dawson, “The Higgs Hunter’s Guide”, volume 80 of *Frontiers in Physics*. Westview Press, Boulder, Colorado, 2000.
ISBN 073820305X,
- [34] L. Randall and R. Sundrum, “A large mass hierarchy from a small extra dimension”, *Phys. Rev. Lett.* **83** (1999) 3370, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.
- [35] L. Randall and R. Sundrum, “An alternative to compactification”, *Phys. Rev. Lett.* **83** (1999) 4690, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.
- [36] C. T. Hill, “Topcolor assisted technicolor”, *Phys. Lett. B* **345** (1995) 483, doi:10.1016/0370-2693(94)01660-5, arXiv:hep-ph/9411426.
- [37] E. Eichten and K. Lane, “Low-scale technicolor at the Tevatron and LHC”, *Phys. Lett. B* **669** (2008) 235, doi:10.1016/j.physletb.2008.09.047, arXiv:0706.2339.
- [38] CMS Collaboration, “Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Submitted to: Phys. Rev.* (2018) arXiv:1809.00327.
- [39] CMS Collaboration, “Search for high-mass diphoton resonances in proton–proton collisions at 13 TeV and combination with 8 TeV search”, *Phys. Lett. B* **767** (2017) 147–170, doi:10.1016/j.physletb.2017.01.027, arXiv:1609.02507.

- [40] CMS Collaboration, “Search for high-mass $Z\gamma$ resonances in $e^+e^-\gamma$ and $\mu^+\mu^-\gamma$ final states in proton-proton collisions at $\sqrt{s} = 8$ and 13 TeV”, *JHEP* **01** (2017) 076, doi:10.1007/JHEP01(2017)076, arXiv:1610.02960.
- [41] CMS Collaboration, “Search for high-mass $Z\gamma$ resonances in proton-proton collisions at $\sqrt{s} = 8$ and 13 TeV using jet substructure techniques”, *Phys. Lett. B* **772** (2017) 363, doi:10.1016/j.physletb.2017.06.062, arXiv:1612.09516.
- [42] CMS Collaboration, “Search for $Z\gamma$ resonances using leptonic and hadronic final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”, (2017). arXiv:1712.03143. Submitted to JHEP.
- [43] CMS Collaboration, “Search for massive resonances decaying into pairs of boosted bosons in semi-leptonic final states at $\sqrt{s} = 8$ TeV”, *JHEP* **08** (2014) 174, doi:10.1007/JHEP08(2014)174, arXiv:1405.3447.
- [44] CMS Collaboration, “Search for massive resonances decaying into WW, WZ or ZZ bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **03** (2017) 162, doi:10.1007/JHEP03(2017)162, arXiv:1612.09159.
- [45] CMS Collaboration, “Combination of searches for heavy resonances decaying to WW, WZ, ZZ, WH, and ZH boson pairs in proton-proton collisions at $\sqrt{s} = 8$ and 13 TeV”, *Phys. Lett. B* **774** (2017) 533, doi:10.1016/j.physletb.2017.09.083, arXiv:1705.09171.
- [46] CMS Collaboration, “Search for exotic resonances decaying into WZ/ZZ in pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **02** (2013) 036, doi:10.1007/JHEP02(2013)036, arXiv:1211.5779.
- [47] CMS Collaboration, “Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **02** (2016) 145, doi:10.1007/JHEP02(2016)145, arXiv:1506.01443.
- [48] CMS Collaboration, “Search for massive WH resonances decaying into the $\ell\nu b\bar{b}$ final state at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **76** (2016) 237, doi:10.1140/epjc/s10052-016-4067-z, arXiv:1601.06431.
- [49] CMS Collaboration, “Search for a narrow spin-2 resonance decaying to a pair of Z vector bosons in the semileptonic final state”, *Phys. Lett. B* **718** (2013) 1208, doi:10.1016/j.physletb.2012.11.063, arXiv:1209.3807.
- [50] CMS Collaboration, “Search for narrow high-mass resonances in proton-proton collisions at $\sqrt{s} = 8$ TeV decaying to a Z and a Higgs boson”, *Phys. Lett. B* **748** (2015) 255, doi:10.1016/j.physletb.2015.07.011, arXiv:1502.04994.
- [51] CMS Collaboration, “Search for heavy resonances decaying to two Higgs bosons in final states containing four b quarks”, *Eur. Phys. J. C* **76** (2016), no. 7, 371, doi:10.1140/epjc/s10052-016-4206-6, arXiv:1602.08762.

- [52] CMS Collaboration, “Search for resonant pair production of Higgs bosons decaying to two bottom quark–antiquark pairs in proton–proton collisions at 8 TeV”, *Phys. Lett. B* **749** (2015) 560–582, doi:10.1016/j.physletb.2015.08.047, arXiv:1503.04114.
- [53] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008), no. 08, S08001.
- [54] L. Evans, “The Large Hadron Collider”, *Annual Review of Nuclear and Particle Science* **61** (2011), no. 1, 435–466, doi:10.1146/annurev-nucl-102010-130438, arXiv:https://doi.org/10.1146/annurev-nucl-102010-130438.
- [55] CMS Collaboration, “Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **06** (2013) 081, doi:10.1007/JHEP06(2013)081, arXiv:1303.4571.
- [56] LHCb Collaboration, “Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \rightarrow J/\psi\phi K^+$ decays”, *Phys. Rev. Lett.* **118** (2017), no. 2, 022003, doi:10.1103/PhysRevLett.118.022003, arXiv:1606.07895.
- [57] LHCb Collaboration, “Amplitude analysis of $B^+ \rightarrow J/\psi\phi K^+$ decays”, *Phys. Rev.* **D95** (2017), no. 1, 012002, doi:10.1103/PhysRevD.95.012002, arXiv:1606.07898.
- [58] LHCb Collaboration, “Observation of $J/\psi p$ Resonances Consistent with Pentaquark States in $\Lambda_b^0 \rightarrow J/\psi K^- p$ Decays”, *Phys. Rev. Lett.* **115** (2015) 072001, doi:10.1103/PhysRevLett.115.072001, arXiv:1507.03414.
- [59] G. Apollinari, O. Brüning, T. Nakamoto, and L. Rossi, “High Luminosity Large Hadron Collider HL-LHC”, *CERN Yellow Report* (2015), no. 5, 1–19, doi:10.5170/CERN-2015-005.1, arXiv:1705.08830.
- [60] M. Weiss, “The RFQ2 complex: The future injector to CERN Linac 2”, *Conf. Proc. C* **920324** (1992) 539–541. [,539(1992)].
- [61] M. Vretenar, “The radio-frequency quadrupole”, *CERN Yellow Report* (2013) 207–223, doi:10.5170/CERN-2013-001.207, arXiv:1303.6762. CERN-2013-001.
- [62] J.-P. Burnet et al., “Fifty years of the CERN Proton Synchrotron : Volume 2”, *ArXiv e-prints* (September, 2013) doi:10.5170/CERN-2013-005, arXiv:1309.6923. CERN-2013-005.
- [63] S. G. for CPS Improvements, “The Second Stage CPS Improvement Study: 800 MeV Booster Synchrotron”, 1967. <http://cds.cern.ch/record/109292/files/cm-p00051778.pdf>.
- [64] M. Barnes et al., “The PS complex as proton pre-injector for the LHC: design and implementation report”. CERN Yellow Reports: Monographs. CERN, Geneva, 2000.

- [65] CMS Collaboration, “Search for high-mass resonances in Z(qq)gamma final state with 2016 data”,.
- [66] S. Gilardoni et al., “The proposed CERN Proton-Synchrotron upgrade program”, in *IPAC 2011 - 2nd International Particle Accelerator Conference*, pp. 2520–2522. 01, 2011.
- [67] R. Garoby, “Plans for Upgrading the CERN Proton Accelerator Complex”, in *J. Phys. Conf. Ser.*, volume 110. 05, 2008.
- [68] K. Schindl, “The injector chain for the LHC”, in *LEP performance. Proceedings, 9th Workshop, Chamonix, France, January 26-29, 1999*, pp. 47–52. 1999.
- [69] D. Boussard and T. Linnekar, “The LHC Superconducting RF System”, *Adv. Cryog. Eng.* **45A** (2000) 835–844.
- [70] E. Ciapala et al., “Commissioning of the 400 MHz LHC RF System”, Technical Report CERN-LHC-PROJECT-Report-1147, Sep, 2008.
- [71] CMS Collaboration, “CMS, the magnet project: Technical design report”, Technical Report CERN-LHCC-97-10, 1997.
- [72] F. Kircher et al., “Final design of the CMS solenoid cold mass”, in *Applied Superconductivity, IEEE Transactions on*, volume 10, pp. 407 – 410. 04, 2000.
- [73] F. Kircher et al., “A Review of the Magnetic Forces in the CMS Magnet Yoke”, Technical Report CERN-CMS-NOTE-1998-048, 1998.
- [74] G. B. Cerati, “Overview of the CMS pixel detector”, in *2008 IEEE Nuclear Science Symposium Conference Record*, pp. 2196–2199. Oct, 2008.
doi:10.1109/NSSMIC.2008.4774788.
- [75] W. Erdmann, “The CMS pixel detector”, *Int. J. Mod. Phys. A* **25** (2010) 1315–1337,
doi:10.1142/S0217751X10049098. [41(2010)].
- [76] P. Azzurri, “The CMS Silicon Strip Tracker”, *J. Phys. Conf. Ser.* **41** (2006), no. 1, 127.
- [77] G. Pasztor, “The CMS silicon strip tracker”, *AIP Conf. Proc.* **842** (2006) 1067–1069,
doi:10.1063/1.2220455, arXiv:physics/0602117. [1067(2006)].
- [78] C. Biino, “The CMS Electromagnetic Calorimeter: overview, lessons learned during Run 1 and future projections”, *J. Phys. Conf. Ser.* **587** (2015), no. 1, 012001.
- [79] CMS Collaboration, “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV”, *JINST* **8** (2013), no. 09, P09009.
- [80] CMS Collaboration, “The CMS ECAL Trigger and DAQ system: electronics auto-recovery and monitoring”, technical report, Jun, 2018. arXiv:1806.09136.

- [81] P. Bloch et al., “Silicon sensors for the CMS preshower detector”, *Nucl. Instrum. Meth. A* **479** (2002), no. 2, 265 – 277, doi:[https://doi.org/10.1016/S0168-9002\(01\)00900-7](https://doi.org/10.1016/S0168-9002(01)00900-7).
- [82] E. Tournefier, “The preshower detector of CMS at LHC”, *Nucl. Instrum. Meth. A* **461** (2001), no. 1-3, 355–360.
- [83] CMS Collaboration, “Design, performance, and calibration of CMS hadron-barrel calorimeter wedges”, *Eur. Phys. J. C* **55** (May, 2008) 159–171, doi:[10.1140/epjc/s10052-008-0573-y](https://doi.org/10.1140/epjc/s10052-008-0573-y).
- [84] B. S. Acharya et al., “The CMS Outer Hadron Calorimeter”, Technical Report CMS-NOTE-2006-127, CERN, Geneva, Jun, 2006.
- [85] J. Freeman, “Innovations for the CMS HCAL”, *Int. J. Mod. Phys. A* **25** (2010), no. 12, 2421–2436.
- [86] J. Anderson, J. Freeman, S. Los, and J. Whitmore, “Upgrade of the CMS Hadron Outer Calorimeter with SIPMs”, *Physics Procedia* **37** (2012) 72 – 78, doi:<https://doi.org/10.1016/j.phpro.2012.02.358>. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).
- [87] T. M. Shaw et al., “Front end readout electronics for the CMS Hadron Calorimeter”, in *2002 IEEE Nuclear Science Symposium Conference Record*, volume 1, pp. 194–197 vol.1. Nov, 2002. doi:[10.1109/NSSMIC.2002.1239297](https://doi.org/10.1109/NSSMIC.2002.1239297).
- [88] S. Holm, T. M. Shaw, J. Elias, and S. Surgueev, “Design and testing of a radiation tolerant Clock, Control and Monitor (CCM) Module for the CMS HCAL electronics”, in *2002 IEEE Nuclear Science Symposium Conference Record*, volume 1, pp. 229–232 vol.1. Nov, 2002. doi:[10.1109/NSSMIC.2002.1239305](https://doi.org/10.1109/NSSMIC.2002.1239305).
- [89] CMS Collaboration, “Design, performance, and calibration of CMS forward calorimeter wedges”, *Eur. Phys. J. C* **53** (2008), no. 1, 139–166, doi:[10.1140/epjc/s10052-007-0459-4](https://doi.org/10.1140/epjc/s10052-007-0459-4).
- [90] S. I. Cooper, “Phase I Upgrade of the CMS Hadron Calorimeter”, *Nuclear and Particle Physics Proceedings* **273-275** (2016) 1002 – 1007, doi:<https://doi.org/10.1016/j.nuclphysbps.2015.09.157>. 37th International Conference on High Energy Physics (ICHEP).
- [91] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:[10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [92] M. Fouz, “The CMS muon system”, *Nuclear Physics B - Proceedings Supplements* **120** (2003) 264 – 267, doi:[https://doi.org/10.1016/S0920-5632\(03\)01913-3](https://doi.org/10.1016/S0920-5632(03)01913-3). Proceedings of the 8th International Conference on B-Physics at Hadron Machines.

- [93] J. M. Celaet et al., “CMS Drift Tube Chambers Read-Out Electronics”, Technical Report CMS-CR-2008-018, CERN, Geneva, Mar, 2008.
- [94] D. Acosta et al., “Large CMS cathode strip chambers: design and performance”, *Nucl. Instrum. Meth. A* **453** (2000), no. 1, 182 – 187, doi:[https://doi.org/10.1016/S0168-9002\(00\)00627-6](https://doi.org/10.1016/S0168-9002(00)00627-6). Proc. 7th Int. Conf on Instrumentation for colliding Beam Physics.
- [95] J. Liu, M. Matveev, and P. Padley, “Remote access to Xilinx programmable devices in the CSC endcap muon electronic system at CMS”, *JINST* **6** (2011), no. 11, C11001.
- [96] V. Brigljevic et al., “The CMS Event Builder”, *ArXiv Physics e-prints* (June, 2003) arXiv:physics/0306150.
- [97] J. Troska, T. Christiansen, and A. Holzner, “Installation and Commissioning of the CMS Timing, Trigger and Control Distribution System”, 2007. <https://cds.cern.ch/record/1091748>.
- [98] J. Hegeman et al., “The CMS Timing and Control Distribution System”, in *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–3. Oct, 2015. doi:10.1109/NSSMIC.2015.7581984.
- [99] L. Cadamuro, “The CMS Level-1 trigger system for LHC Run II”, *JINST* **12** (2017), no. 03, C03021.
- [100] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017), no. 01, P01020.
- [101] T. Baweji et al., “The New CMS DAQ System for Run-2 of the LHC”, *IEEE Transactions on Nuclear Science* **62** (June, 2015) 1099–1103, doi:10.1109/TNS.2015.2426216.
- [102] CMS Collaboration, F. Beaudette, “The CMS Particle Flow Algorithm”, in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pp. 295–304. 2013. arXiv:1401.8155.
- [103] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.
- [104] CMS Collaboration, “Missing transverse energy performance of the CMS detector”, *JINST* **6** (2011), no. 09, P09001.
- [105] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr, 2009.
- [106] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”, *Nucl. Instrum. Meth. A* **262** (1987) 444–450, doi:10.1016/0168-9002(87)90887-4.

- [107] W. Adam, B. Mangano, T. Speer, and T. Todorov, “Track reconstruction in the CMS tracker”, 2005. <http://cds.cern.ch/record/934067>.
- [108] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc”, *J. Phys. G* **31** (2005), no. 9,.
- [109] “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 08, P08010.
- [110] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [111] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev.* **D48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [112] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms”, *JHEP* **08** (1997) 001, doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.
- [113] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279. 1998. arXiv:hep-ph/9907280.
- [114] G. C. Blazey et al., “Run II jet physics”, in *QCD and weak boson physics in Run II. Proceedings, Batavia, USA, March 4-6, June 3-4, November 4-6, 1999*, pp. 47–77. 2000. arXiv:hep-ex/0005012.
- [115] R. Atkin, “Review of jet reconstruction algorithms”, *J. Phys. Conf. Ser.* **645** (2015), no. 1, 012008.
- [116] D. Bertolini, P. Harris, M. Low, and N. Tran, “Pileup Per Particle Identification”, *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.
- [117] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05** (2014) 146, doi:10.1007/JHEP05(2014)146, arXiv:1402.2657.
- [118] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, “Towards an understanding of jet substructure”, *JHEP* **09** (2013) 029, doi:10.1007/JHEP09(2013)029, arXiv:1307.0007.
- [119] CMS Collaboration, “Identification and calibration of boosted hadronic W/Z bosons at 13 TeV”, 2016. CMS-AN-16-215, http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%20AN-2016/215.
- [120] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

- [121] W. Waltenberger, R. Frühwirth, and P. Vanlaer, “Adaptive vertex fitting”, *J. Phys. G* **34** (2007), no. 12, N343.
- [122] J. Thaler and K. Van Tilburg, “Identifying Boosted Objects with N-subjettiness”, *JHEP* **03** (2011) 015, doi:10.1007/JHEP03(2011)015, arXiv:1011.2268.
- [123] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev.* **D48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [124] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [125] CMS Collaboration, “The CMS data quality monitoring software: experience and future prospects”, *J. Phys. Conf. Ser.* **513** (2014), no. 3, 032024.
- [126] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [127] T. Sjöstrand et al., “An introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.
- [128] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [129] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [130] A. Alloul et al., “FeynRules 2.0—A complete toolbox for tree-level phenomenology”, *Computer Physics Communications* **185** (2014), no. 8, 2250–2300.
- [131] CMS Collaboration, “Search for high-mass resonances in Z(qq)gamma final state”, CMS physics analysis summary CMS EXO-16-020, CERN, 2016.
- [132] S. Kyriacou et al., “Cut Based Photon ID for Run 2”, January, 2016.
<https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2>.
- [133] I. Kravchenko et al., “General Purpose MVA-based Photon Identification”, December, 2015.
<https://twiki.cern.ch/twiki/bin/view/CMS/MultivariatePhotonIdentificationRun2>.
- [134] “Jet Identification for the 13 TeV data Run2016”, December, 2016.
<https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVRun2016>.
- [135] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017), no. 02, P02014.

- [136] R. B. Garg, “G+Jets Xsec”. <https://indico.cern.ch/event/489946/>.
- [137] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, “Techniques for improved heavy particle searches with jet substructure”, *Phys. Rev.* **D80** (2009) 051501, doi:10.1103/PhysRevD.80.051501, arXiv:0903.5081.
- [138] CMS Collaboration, “Performance measurement of the double b-tagger at 13 TeV”, CMS analysis note CMS AN-2016/032, CERN, 2016.
- [139] CMS Collaboration, “Identification of double-b quark jets in Boosted Topologies”, CMS Physics Analysis Summary CMS-PAS-BTV-15-002, CERN, 2015.
- [140] “Usage of b/c Tag Objects for 13 TeV Data in 2016 and 80X MC ”. https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80XReReco#Boosted_event_topologies.
- [141] “W/Z-tagging of Jets ”. <https://twiki.cern.ch/twiki/bin/view/CMS/JetWtagging>.
- [142] CMS Collaboration, “Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms”, note, University of California, Los Angeles, 2010.
- [143] F. Garwood, “Fiducial limits for the Poisson distribution”, *Biometrika* **28** (1936) 437, doi:10.1093/biomet/28.3-4.437.
- [144] A. Read, “Linear Interpolation of Histograms”, *Nucl. Instrum. Meth. A* **425** (1999) 357.
- [145] CMS Collaboration, “CMS luminosity measurement for the 2016 data taking period”, CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.
- [146] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [147] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [148] J. Butterworth et al., “PDF4LHC recommendations for LHC Run II”, *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.
- [149] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [150] A. L. Read, “Presentation of search results: The CLs technique”, *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.

- [151] ATLAS and CMS Collaborations, “Procedure for the LHC Higgs boson search combination in summer 2011”, ATL-PHYS-PUB-2011-011, CMS NOTE-2011/005, CERN, 2011.
- [152] “Documentation of the RooStats-based statistics tools for Higgs PAG”.
<https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit>.
- [153] ATLAS Collaboration, “Search for heavy resonances decaying to a photon and a hadronically decaying $Z/W/H$ boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, (2018).
[arXiv:1805.01908](https://arxiv.org/abs/1805.01908). Submitted to Phys. Rev. D.