

SEARCH FOR SIGNATURES OF LARGE EXTRA DIMENSIONS IN HIGH-MASS
DIPHOTON EVENTS FROM PROTON-PROTON COLLISIONS

AT $\sqrt{s} = 13$ TeV WITH CMS

by

ANDREW THOMAS BUCCILLI

CONOR HENDERSON, COMMITTEE CHAIR

NOBUCHIKA OKADA

TOYOKO ORIMOTO

ANDREAS PIEPKE

PAOLO RUMERIO

DAWN R. WILLIAMS

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Physics and Astronomy
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2019



Abstract

A search is performed for a nonresonant excess of high-mass diphoton events over the Standard Model background prediction. This type of excess could be a signature of new physics, such as from the existence of extra spatial dimensions in the universe. The presence of these extra dimensions could modify the fundamental scale of gravity in such a way as to solve the Standard Model hierarchy problem. The sample of diphoton events were produced from proton-proton collisions using the Large Hadron Collider at a center-of-mass energy of 13 TeV. The data correspond to an integrated luminosity of 35.9 fb^{-1} and were recorded in 2016 with the CMS detector. The dominant, irreducible diphoton background comes from Standard Model diphoton production and is determined using a next-to-next-to-leading order Monte Carlo calculation. The subdominant, reducible photon+jet and dijet backgrounds arise when a jet fakes a photon signature in the detector and one or two become misidentified as photons. A technique using control samples in data is used to estimate this contribution. The data are consistent with the background-only hypothesis and the results are interpreted in the context of the large extra-dimensional model of Arkani-Hamed, Dimopoulos, and Dvali. At 95% confidence level, lower exclusion limits are set on the string mass scale M_S ranging from 5.6 to 9.7 TeV, depending on the model parameters. This result provides the current most stringent limits on searches for large extra dimensions in the diphoton channel.

Dedication

To my father

Acknowledgments

Writing this dissertation has been a time of reflection. During the past six years, graduate school has been an enormous period of growth. This dissertation would not have happened without the help, support, guidance, and encouragement of so many people. Friends, family, teachers, mentors, officemates, roommates, and collaborators have all had an impact.

Having the right advisor is essential. I could not have asked for a better advisor with Conor Henderson. He is always available for his students and constantly devotes an immense amount of time for them, caring deeply about their success. I have come a long way under his guidance and am grateful for his patience, clear explanations, superb mentoring, and camaraderie. As his first graduate student, it has been a pleasure helping him build the Alabama CMS group.

I thank each of the Alabama CMS group members Paolo Rumerio, Otman Charaf, Chris West, and Seth Cooper, all constantly offering support in various ways, especially near the beginning. My friends, officemates, fellow graduate students, professors, and colleagues at Alabama have made my time here enjoyable. I also thank my other dissertation committee members Nobuchika Okada, Toyoko Orimoto, Andreas Piepke, and Dawn Williams for their willingness to serve and accommodate, as well as for their assistance during the past three years, providing constructive feedback along the way.

Spending time at CERN was crucial for my growth as a Ph.D. student and was, arguably, the most important thing I did during this time for both my personal development and dissertation research. I thank my CERN officemates John Hakala, Martin Kwok, and David Yu for their friendship, computing help, and physics discussions. The list of friends I made

while at CERN is vast. Jared Sturdy and Brian Dorney are among them. The XESMP and YALT crews taught me to take a break regardless of how demanding things were.

Josh Kunkle was a tremendous help and mentor on everything HCAL. While I was an HCAL DOC, his assistance was vital, and necessary, for smooth operation, especially during the frequent 4 a.m. phone calls. There are many people within HCAL Operations, DPG, and PFG who have assisted along the way. I thank Pawel de Barbaro, Andrea Benaglia, Mario Galanti, Jae Hyeok Yoo, Lucien Lo, Aram Apyan, Rohan Bhandari, Danny Noonan, Viktor Khristenko, Ted Laird, Federico De Guio, Halil Saka, Dick Kellogg, and others.

There are countless people within CMS who have made this research possible. I thank the Rutgers CMS members John Paul Chou, Antonios Agapitos, and Steve Kaplan for their close collaboration with the diphoton analysis during the last three years. Working on the analysis combination with Simone Pigazzini was smooth thanks to his eagerness to advance it. I thank Pasquale Musella for his help in the early stages of the Run 2 diphoton analysis, especially during our time commissioning and tuning the photon ID. I also thank David Sheffield and Saptaparna Bhattacharya for their prolonged assistance with our Monte Carlo samples. The numerous EXO and EXO Non-Hadronic conveners provided constructive and timely feedback on the analysis throughout its duration. Our ARC was invaluable in ushering along the analysis in its latter stages. Our ARC Chair, Si Xie, made the joint coordination process easy during my time as Analysis Contact. Several internal reviewers and institutes helped polish the analysis as it became publication ready.

Finally, I thank my family. Dad and Mom have made me who I am today, and my brothers Chris, David, and Danny have always been there along the way. I thank my parents-in-law Zhoushun and Wenying for their support. My deepest thanks go to my wife Lijuan. Without her sacrifice, love, and encouragement, this would not have been possible.

This work was supported in part by DOE grant #DE-SC0012447.

Contents

Abstract	ii
Dedication	iii
Acknowledgments	iv
List of Tables	ix
List of Figures	xii
Chapter 1 Introduction	1
1.1 Standard Model	1
1.2 Beyond the Standard Model	11
1.2.1 Hierarchy Problem	11
1.2.2 Large Extra Dimensions	12
1.3 Organization of Dissertation	19
Chapter 2 Experimental Apparatus	20
2.1 Large Hadron Collider	20
2.1.1 The CERN Accelerator Complex	20
2.1.2 The LHC Components	23
2.1.3 The LHC Schedule	25
2.2 The CMS Detector	28
2.2.1 Coordinate System	29
2.2.2 Inner Tracking System	31
2.2.3 Electromagnetic Calorimeter	32
2.2.4 Hadron Calorimeter	35

2.2.5	Muon System	37
2.2.6	Trigger System	39
Chapter 3 Event Reconstruction and Selection		42
3.1	Event Reconstruction	42
3.1.1	Tracks and Vertices	43
3.1.2	Photons and Electrons	43
3.1.3	Muons and Tau Leptons	45
3.1.4	Jets and Missing Transverse Energy	46
3.2	Event Selection	47
3.2.1	Trigger Selection	47
3.2.2	Event Preselection	49
3.2.3	Photon Identification	51
3.2.4	Selection Efficiency	54
3.2.5	Datasets	55
Chapter 4 Background Determination		57
4.1	Background from SM Diphoton Production	57
4.1.1	Monte Carlo Simulation	58
4.1.2	The K Factor	64
4.2	Background from Diphoton Misidentifications	68
4.2.1	Fake Rate Method	69
4.2.2	Fake Rate Calculation	71
4.2.3	Fake Rate Studies	79
4.2.4	Applying the Fake Rate Method	83
4.2.5	Closure Test to the Fake Rate Method	84
4.2.6	Closure Test Studies	97
Chapter 5 Signal Simulation		106
5.1	Signal Generation	106

5.2	The ADD Convention Relations	108
5.3	Signal Samples	113
Chapter 6 Sources of Systematic Uncertainty		119
6.1	Real Background Systematic Uncertainties	119
6.2	Fake Background Systematic Uncertainties	124
6.3	Signal Systematic Uncertainties	124
6.4	Summary of the Systematic Uncertainties	124
Chapter 7 Results and Limit Setting		126
7.1	Results of the High-Mass Diphoton Search	126
7.2	Limits on the ADD Model	134
Chapter 8 Conclusion		142
References		144
Appendix A Example MC Run Cards		154
A.1	Higher Order SM Background Run Card	154
A.2	An ADD Signal Run Card	157
A.3	Leading Order SM Background Run Card	159
Appendix B Additional Fake Rate Material		161
B.1	The MC Fake Rate Fits	161
B.2	The MC Fake Templates	164

List of Tables

1.1	The representations of a single generation of fermions in the SM gauge group. The ordered triplets correspond to their representations under the groups ordered as $(\text{SU}(3)_c, \text{SU}(2)_L, \text{U}(1)_Y)$. Here, the generic symbols e and ν are used to denote the charged and neutral leptons, respectively, and u and d for the up- and down-type quarks, respectively. The left- and right-handed fermions are separated using the respective subscripts L and R	3
1.2	A list of the SM particles, each with their corresponding mass, spin, and charge. The masses are approximated from Ref. [16]. The electric charge is given in units of e , the elementary charge. Due to the large mixing in the neutrino sector, the mass differences between the different neutrino flavors are small and it's difficult to refer to individual neutrino mass eigenstates. Since the mass measurement of ν_e is the most stringent among the different neutrino flavors, the masses of ν_μ and ν_τ have been constrained by it.	5
1.3	The branching ratios for a massive graviton G_{KK} decaying to SM particles with $q = u, d, c, s, b$ and $l = e, \mu$ [44]. The remaining 24 modes are split evenly among the 3 flavored neutrino-antineutrino pair and $\tau^+\tau^-$ decay modes. . . .	17
1.4	A summary of the searches for large extra dimensions in the high-mass diphoton channel performed by the CMS Collaboration's Exotica (EXO) Physics Analysis Group prior to this dissertation. The 95% confidence level observed limits on M_S are given, with the range depending on the model convention. .	19
2.1	A summary of the current planned longterm LHC schedule. A tentative schedule extends past LS3.	26
3.1	Parameters used in the definition of $\overline{\text{Iso}}_\gamma$ based on the η coordinate of the supercluster (η_{SC}).	54
3.2	The high- p_T photon ID. The cut values are listed for the different identification variables. For $\sigma_{i\eta i\eta}$, the values in parenthesis correspond to saturated photons in the ECAL.	54
3.3	The analysis datasets, each listed with their CMS path name and corresponding integrated luminosity.	56

4.1	MC background samples used in this analysis with the number of events and corresponding cross sections. The full, internal CMS path name is specified by appending the dataset path with /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6/MINIAODSIM.	63
4.2	Parameters of the K factor function for the EBEB and EBEE categories. . .	68
4.3	Control datasets used for the photon fake rate calculation.	72
4.4	The MC samples and corresponding cross sections used for deriving the real templates in the fake rate calculation. The full, internal CMS path name is specified by appending the dataset path with /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6/MINIAODSIM. . .	76
4.5	The cut-based definitions of the numerator and denominator objects for the fake rate calculation compared against the high- p_T photon ID. The numerator is determined after numerator candidates are fit with real and fake templates (definitions shown) using the template variable $\sigma_{i\eta i\eta}$. The cuts listed are for both the EB and EE regions unless separately specified.	76
4.6	Fit parameters for the fake rate function.	79
4.7	The Monte Carlo samples and corresponding cross sections used for the closure test to the fake rate method. The full, internal CMS path name is specified by appending the dataset path with /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6/MINIAODSIM.	85
5.1	Cross sections (in units of fb) for the combined ADD+SM samples used in the analysis. For $M_S = 3\text{-}4$ TeV, the last $m_{\gamma\gamma}$ bin is from $2\text{-}M_S$ TeV, i.e., there is no bin splitting into $2\text{-}X$ TeV and $X\text{-}M_S$ TeV. For $M_S = 4.5$ or 5 TeV, the bin edge is at $X = 3$ TeV. For $M_S > 5$ TeV, $X = 4$ TeV.	115
5.2	The LO SM background MC samples. The full, internal CMS path name is specified by appending the dataset path with /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6/MINIAODSIM . .	116
7.1	The three highest $m_{\gamma\gamma}$ events observed in data for the EBEB and EBEE categories. The photon p_T is given in units of GeV. The dataset paths begin with /DoubleEG and end with /MINIAOD, and include $X = 03\text{Feb}2017$. Here, run, LS, and event correspond to the run number, luminosity section, and event number, respectively. Each event was recorded in 2016, where pp collisions occurred (the majority of the time) from May-October.	129
7.2	Event yields for the total pre- and post-fit background predictions compared to the data for different ranges of $m_{\gamma\gamma}$. The total systematic uncertainties are included on the post-fit yields.	132

7.3	Exclusion limits on the mass scale M_S (in units of TeV) for various conventions used to calculate the ADD large extra-dimensional scenario. The total asymmetric uncertainties are shown on the expected limits.	141
-----	---	-----

List of Figures

1.1	Feynman diagrams representing the interactions between the fermions and vector gauge bosons in the SM [24]. The charges of the W bosons are determined by conserving charge with the fermions they interact with.	8
1.2	Feynman diagrams representing the Higgs boson interactions with the fermions (f) and gauge bosons in the SM [25].	9
1.3	Summary of the theory and experimental production cross section measurements of SM processes, measured by the CMS Collaboration [28].	10
1.4	One-loop Feynman diagrams for the quantum corrections to the Higgs boson squared mass due to fermions (left) and bosons (right) [29].	12
1.5	Virtual graviton exchange in the diphoton ($\gamma\gamma$) channel through quark annihilation (left) and gluon fusion (right).	15
2.1	An overview of the location of LHC and its four interaction points [73]. . . .	21
2.2	The CERN accelerator complex [74].	22
2.3	A schematic of the eight LHC insertion points [75].	23
2.4	A cross-section of an LHC dipole magnet [80].	24
2.5	A snapshot of the LHC status display, taken in June 2017 during stable beam pp physics collisions, showing the decay in instantaneous luminosity over time as measured by the LHC experiments.	26
2.6	A timeline showing the cumulative (left) and detailed 2016 (right) integrated luminosity delivered by the LHC [88]. The amount recorded by CMS in 2016 is also shown.	28
2.7	An overview of the CMS detector and its major components [89].	29
2.8	A transverse slice of the CMS detector highlighting how different particles interact with its various detector technologies [90].	30

2.9	An illustration of the coordinate system used by the CMS detector [91]. . . .	31
2.10	A schematic of the CMS inner tracking system [92].	31
2.11	A cutaway view of the ECAL depicting its main features [78].	34
2.12	The ECAL crystal response measured by the ECAL laser monitoring system over the LHC data taking periods [94].	35
2.13	A projection of the HCAL, highlighting its subdetectors [78].	36
2.14	A schematic of the CMS muon system depicting its major components and their detector coverage [98].	38
2.15	The average amount of pileup recorded by the CMS detector during each of its data taking years [88]. The histograms from each year are stacked on the previous year. Also printed are the measured pp interaction cross sections for the different center-of-mass energies used during each data taking period. . .	39
2.16	A flowchart showing how the global, regional, and local components of the L1 trigger system operate to produce an L1A [78].	40
3.1	The relative efficiency of the <code>HLT_DoublePhoton60</code> trigger (measured with reference to the <code>HLT_DoublePhoton40</code> trigger) as a function of the p_T of the subleading photon.	48
3.2	The absolute efficiency of the <code>HLT_DoublePhoton60</code> trigger (measured using a simulated diphoton sample) as a function of $m_X = m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories. The inner boxes show a zoomed-in region of the y -axis.	49
3.3	A plot of $\sigma_{\text{in}}^{\text{in}}$ in various photon p_T bins for saturated (left) and unsaturated photons (right) in the EB. The vertical line is at the value of 0.0105.	52
3.4	Photon selection efficiency (top) in the EB (left) and EE (right) regions and the associated data-over-simulation scale factors (bottom). The statistical and systematic errors are shown.	55
3.5	The product of the event selection efficiency (ε) and the detector accep- tance (A) is shown as a function of signal resonance mass $m_X = m_{\gamma\gamma}$ for the $\Gamma_X/m_X = 1.4 \times 10^{-4}$ signal width hypothesis. The total (black), EBEB (red), and EBEE (blue) curves are shown for the spin (J) hypotheses $J = 0$ (solid line) and $J = 2$ (dashed line).	56
4.1	Representative Feynman diagrams for the LO SM diphoton Born (left) and box (right) processes.	58

4.2	Diagrammatic representation of a proton-proton collision with their partons undergoing the hard scattering process [121]. The remainder of the two protons after collision is known as the beam remnant.	60
4.3	Example PDFs at energy scales $Q = 2, 3.16, 8,$ and 85 GeV , from the CT10 set of PDFs [120].	60
4.4	Representative Feynman diagrams for NLO diphoton production through the born process [122]. In (a) $q\bar{q}$ and (b) gq initiated processes are shown.	61
4.5	Representative Feynman diagrams for NLO diphoton production through the fragmentation of a quark or gluon [122].	62
4.6	Representative Feynman diagrams for the diphoton Born process at NNLO, considered by MCFM [125].	64
4.7	Representative Feynman diagrams for the diphoton box process at LO and NLO, considered by MCFM [125].	64
4.8	The NNLO K factor as a function of $m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories. The renormalization and factorization scales have been set to $m_{\gamma\gamma}$ (middle) and simultaneously varied by factors of 2.0 (upper) and 0.5 (lower).	66
4.9	Ratio of the diphoton cross section calculated by MCFM to that obtained from SHERPA. The curves indicate different orders of the MCFM calculation, and the dashed lines show the effect of the scale variations.	67
4.10	The correlation between $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ for all objects in the jet-triggered control dataset (used for the fake rate calculation) passing the high- p_T photon ID (with the $\sigma_{i\eta i\eta}$ cut relaxed) in the range $450 < p_T < 500 \text{ GeV}$. Beam halo is observed for $\sigma_{i\phi i\phi} < 0.009$	73
4.11	Fake templates scaled to 1 fb^{-1} in various p_T bins in the EB (left) and EE (right) regions.	75
4.12	Real templates scaled to 1 fb^{-1} in various photon p_T bins in the EB (left) and EE (right) regions.	77
4.13	A template fit in the EB with $130 < p_T < 150 \text{ GeV}$, shown in linear (left) and log (right) scale.	78
4.14	The fake rate in different photon p_T bins for the EB (left) and EE (right) categories. The fake rate function (red) is the best fit to the points.	78

4.15	Separate fake rates for the jet- (red) and muon-triggered (green) datasets are shown in each photon p_T bin for the EB (left) and EE (right) categories. The average of the two (black) is the final fake rate.	80
4.16	The number of PVs for events containing at least one photon in the jet-triggered dataset used for the fake rate calculation. The distributions are shown before (red) and after (blue) further filtering of the PVs is imposed. .	81
4.17	The fake rate based on the number of PVs in the events using the jet-triggered dataset in the EB (left) and EE (right).	81
4.18	Comparison of the fake rates using $\sigma_{i\eta i\eta}$ (green) and Iso _{Ch} (red) template variables in the EB (left) and EE (right).	82
4.19	The fake rate using a Iso _{Ch} template with variations in the $\sigma_{i\eta i\eta}$ sideband for the EB (left) and EE (right).	83
4.20	Fake templates extracted from the MC sample used in the fake rate closure test compared against the MC truth fake templates for $50 < p_T < 70$ GeV in the EB (left) and EE (right). The template variable is $\sigma_{i\eta i\eta}$	87
4.21	Fake templates used in the MC based closure test showing their p_T evolution using the nominal Iso _{Ch} sideband definition of 5-10 GeV in the EB (left) and EE (right) regions.	88
4.22	Real templates using $\sigma_{i\eta i\eta}$ showing the evolution with photon p_T in the EB (left) and EE (right).	89
4.23	Template fits using $\sigma_{i\eta i\eta}$ for the MC closure test with $90 < p_T < 130$ GeV in the EB (top) and EE (bottom).	89
4.24	Fake rates for the MC closure test as a function of p_T using $\sigma_{i\eta i\eta}$ templates in the EB (left) and EE (right). Three different choices for the Iso _{Ch} sidebands used in deriving the fake rates are compared against the fake rate using the MC truth information.	90
4.25	Comparison of the ratio of the fake yields obtained from the MC fake prediction using $\sigma_{i\eta i\eta}$ templates over the yields identified by the MC truth information in the EB (left) and EE (right) regions.	91
4.26	Template fits using $\sigma_{i\eta i\eta}$ with fake templates derived using the MC truth information in the simulated sample with the nominal sideband definition for $70 < p_T < 90$ GeV in the EB (top) and EE (bottom) regions.	92
4.27	Fake rate derived using fake templates constructed using the MC truth information in the simulated sample as a function of p_T with $\sigma_{i\eta i\eta}$ templates in the EB (left) and EE (right) regions.	92

4.28	MC fake rate as a function of p_T in the EB (left) and EE (right) using the nominal Iso _{Ch} sideband of 5-10 GeV.	93
4.29	Closure test of the photon variables p_T (top), η (middle), and ϕ (bottom) in the EB (left) and EE (right) regions. The MC denominator objects (black) are reweighted by the MC fake rate (blue) and compared against the distributions of known fakes identified using the MC truth (red).	94
4.30	Closure test of the diphoton $m_{\gamma\gamma}$ distribution of γj objects in the EBEB (left) and EBEE (right) categories. The MC denominator objects (black) are reweighted by the MC fake rate (blue) and compared against the distributions of known fakes identified using the MC truth (red).	95
4.31	Closure test of the diphoton $m_{\gamma\gamma}$ distribution of γj objects ordered as TF (top 4 plots) and FT (bottom 4 plots) in the EBEB, EBEE, EEEB, and EEEE categories (specified in the plot titles). The MC denominator objects (black) are reweighted by the MC fake rate (blue) and compared against the distributions of known fakes identified using the MC truth (red).	96
4.32	Real templates in Iso _{Ch} derived from simulation in the EB (left) and EE (right) categories for different photon p_T bins.	97
4.33	Fake templates in Iso _{Ch} using different $\sigma_{i\eta i\eta}$ sidebands from simulation for $90 < p_T < 130$ GeV in the EB (left) and EE (right) categories, compared against the MC truth prediction.	98
4.34	Fake templates from simulation using the nominal $\sigma_{i\eta i\eta}$ sidebands in the EB (left) and EE (right) categories for different p_T bins.	98
4.35	An example template fit using the Iso _{Ch} variable and nominal $\sigma_{i\eta i\eta}$ sidebands for the fake templates in the EB (top) and EE (bottom) categories with $90 < p_T < 130$ GeV.	99
4.36	The fake rate as a function of p_T using Iso _{Ch} templates from simulation in the EB (left) and EE (right) categories with different $\sigma_{i\eta i\eta}$ sidebands, compared to the MC truth fake rate.	100
4.37	Closure test of the event yields from the fake prediction in simulation using Iso _{Ch} template fits compared to the MC truth yields in the EB (left) and EE (right) categories.	100
4.38	Comparison of the predicted MC fake rates as a function of p_T using $\sigma_{i\eta i\eta}$ (red) and Iso _{Ch} (blue) templates in the EB (left) and EE (right) categories compared to the MC truth fake rate (gold).	101

4.39	Closure test of the event yields from the fake prediction in the simulated sample using $\sigma_{i\eta i\eta}$ (blue) and Iso _{Ch} (red) templates in the EB (left) and EE (right) regions.	101
4.40	Fake rates derived using the MC truth information to construct fake templates in $\sigma_{i\eta i\eta}$ (red) and Iso _{Ch} (blue) as a function of p_T in the EB (left) and EE (right) regions. These are compared against the MC fake rate derived from the fake yields (gold).	102
4.41	The separation in ΔR among two final state generated particles near a reconstructed fake photon in the EB (left) and EE (right) regions. The separation is shown separately for the fake photons that pass (black) and fail (red) the $\sigma_{i\eta i\eta}$ photon ID cut.	103
4.42	Composition of objects in the 5-10 GeV sideband in the simulated sample used for the closure test shown in the EB (left) and EE (right) categories. The fake (black) and fake-like (blue) objects dominate over the real (red) ones.	104
4.43	The fraction of real photons among all denominator objects as a function of p_T , derived from the simulated sample used by the closure test in the EB (left) and EE (right) regions.	105
5.1	The $m_{\gamma\gamma}$ distributions for the ADD signal using the Hewett+ convention with $M_S = 4000$ GeV (blue) compared to the SM background (red).	109
5.2	Combined ADD+SM signal distributions in photon η compared again the corresponding SM background distribution.	110
5.3	Comparison of the $m_{\gamma\gamma}$ spectra between the Hewett+ (blue) and Hewett- (red) conventions using $M_S = 4000$ GeV.	111
5.4	The $m_{\gamma\gamma}$ spectra for different signals from the HLZ convention using $M_S = 2500$ GeV for different values of n_{ED}	111
5.5	The $m_{\gamma\gamma}$ distributions demonstrating that the HLZ convention using $n_{ED} = 4$ (blue) is equivalent to the GRW convention (red), both for $M_S = 4000$ GeV (right), and those from the GRW convention using $M_S = 4000$ GeV for $n_{ED} = 2$ (blue) and 4 (red), demonstrating the equivalence of different n_{ED} values (left).	112
5.6	The $m_{\gamma\gamma}$ distributions for the Hewett+ convention (red) using $M_S = 4000$ GeV and the GRW convention (blue) using $M_S = 4478$ GeV (left). The $m_{\gamma\gamma}$ distributions for the HLZ convention (blue) using $n_{ED} = 6$ with $M_S = 4000$ GeV and for the HLZ convention (red) using $n_{ED} = 3$ with $M_S = 5657$ GeV (right). In each plot, the two conventions have the same value of η_G and their spectra are shown to be equivalent.	113

5.7	The $m_{\gamma\gamma}$ distribution in signal events for the generated values of M_S before (left) and after (right) subtraction of the SM background. The KK conventions used in the plots are HLZ with $n_{\text{ED}} = 2$ (top), HLZ assuming $n_{\text{ED}} = 4$ (middle), and Hewett- (bottom).	117
5.8	The $m_{\gamma\gamma}$ distributions for the ADD signal and interference before (left) and after (right) subtraction of the SM background in the EBEB (top) and EBEE (bottom) categories. The spectra correspond to the KK conventions using HLZ with $n_{\text{ED}} = 2$ and $M_S = 8$ TeV (black) compared against HLZ with $n_{\text{ED}} = 4$ using $M_S = 6, 7, 8, 9$, and 10 TeV.	118
6.1	The K factor as a function of $m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories. The K factor is defined as the ratio of the predicted MCFM NNLO $pp \rightarrow \gamma\gamma$ cross section to that of SHERPA. The renormalization and factorization scales have been set to $m_{\gamma\gamma}$ (black) and varied simultaneously by factors of 0.5 (blue) and 2.0 (red).	120
6.2	The ± 1 standard deviation variations of each of the 26 PDF eigenvectors in the EBEB category. Each distribution is the ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$	122
6.3	The ± 1 standard deviation variations of each of the 26 PDF eigenvectors in the EBEE category. Each distribution is the ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$	123
6.4	The dominant pre-fit systematic uncertainties and total post-fit uncertainty as a function of $m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories.	125
7.1	A comparison of various kinematic distributions between the SM background prediction, fake background measurement, and data. The photons are restricted to the EB and diphotons to the EBEB.	127
7.2	A comparison of various kinematic distributions between the SM background prediction, fake background measurement, and data. The photons are restricted to the EE and diphotons to the EBEE.	128
7.3	An event display of the highest invariant mass diphoton event recorded in 2016 at 1840 GeV in the EBEB category. The two photons are represented by the two large, red ECAL energy deposits.	129
7.4	The (η, ϕ) spatial distribution of the leading (blue) and subleading (red) photons in the diphoton events with $m_{\gamma\gamma} > 1.5$ TeV for the EBEB (left) and EBEE (right) categories.	130

7.5	The $\sigma_{i\eta i\eta}$ (left) and R_9 (right) distributions of the leading (blue) and subleading (red) photons in the diphoton events with $m_{\gamma\gamma} > 1.5$ TeV compared to the photons in events with $m_{\gamma\gamma} < 1.5$ TeV. The y -axis is labeled using arbitrary units.	131
7.6	The diphoton invariant mass distributions in the EBEB (left) and EBEE (right) categories for the pre-fit SM diphoton background prediction and the fake background measurement compared to the data. The last bin includes the overflow. The error bars on the points indicate the statistical uncertainty, and the hatched bands indicate the total pre-fit systematic uncertainties. The bottom panels show the pull distributions, indicating the difference between the data and background prediction, divided by the uncertainty in the background, with error bars representing the statistical uncertainty and shaded bands showing the one standard deviation systematic uncertainty, normalized by the statistical uncertainty.	132
7.7	The diphoton invariant mass distributions in the EBEB (left) and EBEE (right) categories for the post-fit SM diphoton background prediction and the fake background measurement compared to the data. The last bin includes the overflow. The error bars on the points indicate the statistical uncertainty, and the hatched bands indicate the total post-fit systematic uncertainties. Invariant mass distributions from one signal scenario is superimposed. The bottom panels show the pull distributions, indicating the difference between the data and background prediction, divided by the uncertainty in the background, with error bars representing the statistical uncertainty and shaded bands showing the one standard deviation systematic uncertainty, normalized by the statistical uncertainty.	133
7.8	Posterior Gaussian distributions (using a mean of 0 and standard deviation of 1) of the nuisance parameters assigned to PDF eigenvectors 1-13 (first and third columns). Their ± 1 standard deviation variations are shown as a ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$ in the EBEB (solid) and EBEE (dashed) categories (second and fourth columns).	136
7.9	Posterior Gaussian distributions (using a mean of 0 and standard deviation of 1) of the nuisance parameters assigned to PDF eigenvectors 14-26 (first and third columns). Their ± 1 standard deviation variations are shown as a ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$ in the EBEB (solid) and EBEE (dashed) categories (second and fourth columns).	137

7.10	Posterior distributions for all non-PDF nuisance parameters associated with the real and fake backgrounds. Some of their associated ± 1 standard deviation variations are shown as a ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$ in the EBEB and EBEE categories. These are shown for the normalization of the real and fake background in the EBEB and EBEE categories (first row), the fake templates (second row), the NLO uncertainties (third row), and the factorization and renormalization scale variations (fourth row).	138
7.11	The correlation matrix for all 35 independent nuisance parameters used in marginalization of the real and fake background (top), the 26 PDF nuisance parameters (middle), and the remaining 9 non-PDF nuisance parameters (bottom).	139
7.12	The 95% CL exclusion limits on the ADD model signal strength for the GRW (left), Hewett- (middle), and HLZ assuming $n_{\text{ED}} = 2$ (right) conventions. A 100 GeV binning is used.	141
B.1	Template fits for the MC closure test using the $\sigma_{i\eta i\eta}$ template variable, shown in linear and log scales, for the EB (left) and EE (right) regions. The different p_{T} bins are shown in the rows.	162
B.2	Template fits for the MC closure test using the Iso_{Ch} template variable, shown in linear and log scales, for the EB (left) and EE (right) regions. The different p_{T} bins are shown in the rows.	163
B.3	Fake templates in $\sigma_{i\eta i\eta}$ from simulation in the EB (left) and EE (right) categories for various p_{T} bins, shown in the rows.	165
B.4	Fake templates in Iso_{Ch} from simulation in the EB (left) and EE (right) categories for various p_{T} bins, shown in the rows.	166

Chapter 1

Introduction

Since its development in the 1970s, the Standard Model of particle physics has been an enormously successful description of observed phenomena, withstanding many precision tests. However, despite this success, it is widely believed to be incomplete. Among its shortcomings is the hierarchy problem, which arises due to the large difference between two fundamental scales in the theory. The existence of extra spatial dimensions in the universe has been proposed as one possible solution to this problem. This dissertation aims to address this issue by searching for evidence of large extra dimensions at the Large Hadron Collider using proton-proton collisions at the energy frontier.

1.1 Standard Model

The Standard Model (SM) of particle physics is a gauge theory describing the dynamics of the known fundamental particles and their interactions through the strong, weak, and electromagnetic forces (with the notable exception of gravity). A gauge theory is a quantum field theory in which the Lagrangians describing the evolution and interactions of the particles remain invariant under local transformations of a symmetry group. Particles are described as excitations of quantum fields in (3+1)-dimensional spacetime whose properties are defined by their representations under different symmetry groups. There are two types of particles: bosons, which have integer spin with symmetric wave functions under the exchange of

two identical particles, and fermions, which have half-integer spin with antisymmetric wave functions under identical particle exchange. Fermions are the matter particles and their interactions are mediated by bosons, the force carriers.

Gauge theories based on the special unitary group of degree N , $SU(N)$, are known as Yang-Mills theories [1]. The SM is the Yang-Mills theory of the gauge group

$$SU(3)_c \times SU(2)_L \times U(1)_Y$$

The group $SU(3)_c$ corresponds to the strong force and $SU(2)_L \times U(1)_Y$ to the unification of the electromagnetic and weak forces, known as the electroweak force. The gauge transformation of $SU(3)_c$ yields the fields G_μ^α , for $\alpha = 1, 2, \dots, 8$; $SU(2)_L$ yields W_μ^a , for $a = 1, 2, 3$; and $U(1)_Y$ yields B_μ , where μ is the component in (3+1)-dimensions. These gauge fields give rise to gauge bosons, which are also the generators of the group. These gauge bosons are the 8 gluons (g) for $SU(3)_c$ and the $W^\pm$ and Z bosons and photon (γ) for $SU(2)_L \times U(1)_Y$. These are all vector (spin-1) gauge bosons. The gauge bosons $W^\pm$, Z , and γ are obtained after electroweak symmetry breaking, as discussed below. The gluons mediate the strong force, the $W^\pm$ and Z bosons mediate the weak force, and the photon mediates the electromagnetic force.

Fermion fields are associated with every observed matter particle. These are spin-1/2 fermions that exist in two varieties: quarks, which participate in the strong interaction, and leptons, which do not. For each charged fermion there is also a corresponding antifermion of the same mass but opposite electric charge. There are three generations of the fermions, which are identical except for the associated particle masses between generations. Each generation contains an up-type quark, a down-type quark, a charged lepton, and a neutral lepton, called a neutrino. The up-type quarks have an electric charge of $+\frac{2}{3}$ and are the up (u), charm (c), and top (t) quarks. The down-type quarks have an electric charge of $-\frac{1}{3}$ and are the down (d), strange (s), and bottom (b) quarks. The charged leptons have an

electric charge of -1 and are the electron (e^-), muon (μ^-), and tau (τ^-). The neutrinos are electrically neutral and are the electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino (ν_τ). Each fundamental particle is assigned to a particular representation of the SM gauge group.

Each flavor of quark transforms under the fundamental representation **(3)** of $SU(3)_c$, forming “color” triplets. Each quark has a color charge (as denoted by the subscript c) of red, green, or blue. (Antiquarks have a color charge of antired, antigreen, or antiblue.) By analogy with the three primary colors of light, the quark triplets contain a red, green, and blue charged quark state. The leptons are assigned to the trivial representation **(1)** of $SU(3)_c$, forming color singlets. Particles represented as singlets under any symmetry group remain unchanged by the associated transformation and, hence, are not influenced by the corresponding interaction. The left-handed (L) chiral components of the fermions are grouped into $SU(2)_L$ weak doublets, transforming under its fundamental representation **(2)**. The right-handed fermions are $SU(2)_L$ weak singlets with the trivial representation **(1)**. There are no right-handed neutrinos in the simplest form of the theory. The hypercharge (Y) describes the fermion coupling to $U(1)_Y$, chosen to reproduce the electrical charges of the particles. The electrical charge $Q_{EM} = T^3 + \frac{Y}{2}$, where T^3 is the third component of the weak isospin. These representations are summarized in Table 1.1.

Table 1.1: The representations of a single generation of fermions in the SM gauge group. The ordered triplets correspond to their representations under the groups ordered as $(SU(3)_c, SU(2)_L, U(1)_Y)$. Here, the generic symbols e and ν are used to denote the charged and neutral leptons, respectively, and u and d for the up- and down-type quarks, respectively. The left- and right-handed fermions are separated using the respective subscripts L and R .

SM fermion	$q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	u_R	d_R	$\ell_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	e_R
$(SU(3)_c, SU(2)_L, U(1)_Y)$ representation	(3, 2, $\frac{1}{6}$)	(3, 1, $\frac{2}{3}$)	(3, 1, $-\frac{1}{3}$)	(1, 2, $-\frac{1}{2}$)	(1, 1, -1)

The gluons belong to the adjoint representation **(8)** of $SU(3)_c$, forming color octets.

Gluons carry color-anticolor pairs with their eight states given as combinations of these pairs. The gauge boson fields G_μ^α , W_μ^a , and B_μ have the representations $(\mathbf{8}, \mathbf{1}, 0)$, $(\mathbf{1}, \mathbf{3}, 0)$, and $(\mathbf{1}, \mathbf{1}, 0)$, respectively, corresponding the groups ordered as $(\text{SU}(3)_c, \text{SU}(2)_L, \text{U}(1)_Y)$. Gauge invariance prohibits explicit mass terms in the Lagrangians for the gauge bosons. Indeed, the gluons and photon are massless; however, the $W^\pm$ and Z bosons have masses of approximately 80 and 91 GeV¹. This problem is circumvented through the realization that the ground state, or vacuum, of the theory does not need to exhibit the same symmetries as the Lagrangian. This is known as spontaneous symmetry breaking and underlies the Higgs mechanism [2–7].

The spontaneous breaking of a continuous global symmetry implies the existence of massless scalar (spin-0) bosons called Goldstone bosons [8–10]. However, when the symmetry is local, the spontaneous breaking causes the Goldstone bosons to disappear and the Goldstone modes result in the gauge bosons associated with the broken symmetry to become massive through the Higgs mechanism. The Higgs mechanism introduces a Higgs field, which is a complex scalar field that transforms as a doublet under $\text{SU}(2)_L$. The Higgs field has the representation $(\mathbf{1}, \mathbf{2}, \frac{1}{2})$ under the groups ordered as $(\text{SU}(3)_c, \text{SU}(2)_L, \text{U}(1)_Y)$. The spontaneous symmetry breaking of the electroweak gauge group $\text{SU}(2)_L \times \text{U}(1)_Y$ to the quantum electrodynamic (QED) gauge group $\text{U}(1)_{\text{EM}}$ results in the $W^\pm$ and Z bosons acquiring mass. The vacuum is invariant under $\text{U}(1)_{\text{EM}}$ transformations, allowing the photon to remain massless. This mechanism also predicts a massive, electrically neutral scalar boson, called the Higgs (H) boson. This boson was recently discovered by the ATLAS and CMS Collaborations, having mass near 125 GeV with properties consistent with the SM prediction [11–15].

The Higgs mechanism can also explain the presence of fermion masses. Since the weak interaction violates parity, it couples differently to left- and right-handed fermions. Fermion mass terms in the Lagrangian would arise by coupling these separate fields, but this would break $\text{SU}(2)_L$ gauge invariance. Thus, fermions should appear massless in nature, a contra-

¹The convention in particle physics is to use natural units with $\hbar = c = 1$. Hence, energy, momentum, and mass, which have the respective units of eV, eV/ c , and eV/ c^2 , are all given in units of eV.

diction with observation. In contrast to the mass generation for the massive gauge bosons, the fermions acquire mass through Yukawa interactions with the Higgs field. The strength of the Yukawa coupling for a fermion to the Higgs boson is directly proportional to its mass. Despite the absence of right-handed neutrinos in the simplest form of the theory, neutrino oscillation experiments have demonstrated that neutrino masses are nonzero and mix among their three flavor eigenstates, as discussed in Section 63 or Ref. [16]. The mechanism for generating neutrino masses has not been established. A summary of the SM particles, including their masses, is given in Table 1.2.

Table 1.2: A list of the SM particles, each with their corresponding mass, spin, and charge. The masses are approximated from Ref. [16]. The electric charge is given in units of e , the elementary charge. Due to the large mixing in the neutrino sector, the mass differences between the different neutrino flavors are small and it's difficult to refer to individual neutrino mass eigenstates. Since the mass measurement of ν_e is the most stringent among the different neutrino flavors, the masses of ν_μ and ν_τ have been constrained by it.

Particle	Mass	Spin	Charge
u	2.2 MeV	1/2	+2/3
d	4.7 MeV	1/2	-1/3
c	1.275 GeV	1/2	+2/3
s	95 MeV	1/2	-1/3
t	173 GeV	1/2	+2/3
b	4.18 GeV	1/2	-1/3
e^-	0.511 MeV	1/2	-1
ν_e	< 2 eV	1/2	0
μ^-	106 MeV	1/2	-1
ν_μ	< 2 eV	1/2	0
τ^-	1.78 GeV	1/2	-1
ν_τ	< 2 eV	1/2	0
γ	0	1	0
$W^\pm$	80.4 GeV	1	± 1
Z	91.1876 GeV	1	0
g	0	1	0
H	125.18 GeV	0	0

The theory of quantum chromodynamics (QCD) describes the interactions of the quarks under the strong force, based on the non-Abelian gauge group $SU(3)_c$. At low energies, free quarks are not observed and, instead, become confined to hadrons, which are composite

particles made up of quarks. Hadrons made up of three quarks are called baryons and those made up of a quark-antiquark pair are called mesons. There are also hadrons containing more than three valence quarks. Experimental measurements have shown that the energy scale at which QCD becomes confining is $\Lambda_{\text{QCD}} \approx 300 \text{ MeV}$ [16, 17]. At high energies, however, QCD has a perturbative regime as the strength of the strong interaction becomes small, allowing quarks and gluons to behave as nearly free particles. This behavior is called asymptotic freedom and was demonstrated by Politzer, Gross, and Wilczek [18, 19].

Asymptotic freedom is a particular case of the more general phenomenon of running coupling constants, where the values of the coupling constants change depending on the energy scale of the interaction (Q) due to the participation of virtual particles in the interaction. These virtual particles can lead to infinities through loops in the Feynman diagrams during the calculation. Reparameterizing the bare parameters into physical parameters absorbs the divergences and leads to finite and measurable effects. This process is known as renormalization [20] and introduces the (unphysical) renormalization scale. Once the value of the coupling constant is known at this scale, its dependence on Q can be determined. At $Q \approx 100 \text{ GeV}$, the strong coupling constant $\alpha_s \approx 0.1$. This is in contrast to that of the QED coupling constant, α_{EM} , which increases as Q increases due to charge screening from vacuum polarization. In QCD, its non-Abelian nature results in gluons carrying color, which causes the effective color charge to be spread out in the vacuum. The asymptotic value of $\alpha_{\text{EM}} \approx 1/137$.

The gauge structure of the electroweak theory with gauge group $\text{SU}(2)_L \times \text{U}(1)_Y$ was proposed by Glashow, Weinberg, and Salam [21–23]. The combination of the electroweak and QCD theories became known as the SM. All of the ideas presented here about the SM can be summarized in Lagrangian form. The SM Lagrangian density (see, e.g., Ref. [17]), given in terms of the particle fields and their couplings, is

$$\mathcal{L} = -\frac{1}{4}G_{\mu\nu}^\alpha G^{\alpha\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}$$

$$\begin{aligned}
& + \bar{q}_L i \not{D} q_L + \bar{u}_R i \not{D} u_R + \bar{d}_R i \not{D} d_R \\
& + \bar{\ell}_L i \not{D} \ell_L + \bar{e}_R i \not{D} e_R \\
& + (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \\
& - y_u \bar{q}_L \tilde{\phi} u_R - y_d \bar{q}_L \phi d_R - y_e \bar{\ell}_L \phi e_R
\end{aligned} \tag{1.1}$$

where $G_{\mu\nu}^\alpha$, $W_{\mu\nu}^a$, and $B_{\mu\nu}$ are the gauge field-strength tensors, which are defined in terms of the gauge fields, G_μ^α , W_μ^a , and B_μ , respectively; $\not{D} = D_\mu \gamma^\mu$, where γ^μ are the Dirac matrices and D_μ is the gauge-covariant derivative; ϕ is the Higgs field; $V(\phi)$ is the Higgs potential; $\tilde{\phi} = i\sigma_2 \phi^*$, where σ_a are the 2×2 Pauli matrices; and y_u , y_d , and y_e are the Yukawa couplings of the associated up-type quark, down-type quark, and charged lepton. Each Yukawa coupling is directly related to the corresponding fermion mass. The hermitian conjugate of the three terms with Yukawa couplings are also considered in this equation, but not explicitly listed. The gauge-covariant derivative $D_\mu = \partial_\mu - \frac{i}{2}g_s \tau_\alpha G_\mu^\alpha - \frac{i}{2}g \sigma_a W_\mu^a - \frac{i}{2}g' Y B_\mu$, where τ_α are the 3×3 Gell-Mann matrices and g_s , g , and g' are the gauge coupling constants associated with $SU(3)_c$, $SU(2)_L$, $U(1)_Y$, respectively. If any particle (fermion or Higgs) field is a singlet under one of the SM gauge groups, then the term in D_μ involving the corresponding gauge boson field of that gauge group is omitted. The coupling g_s is directly proportional to α_s and both g and g' can be related to α_{EM} . The gauge couplings g and g' are also related to the masses of the $W^\pm$ and Z bosons and, among others, including the Yukawa couplings, are free parameters in the theory.

Gauge invariance restricts the form of the terms in the SM Lagrangian, which determine the possible particle interactions by constraining their couplings. The first line in Eq. (1.1) describes the kinetic energies and self-interactions of the gauge bosons; the second and third lines describe the kinetic energies and electroweak interactions of the fermions, as well as the strong interactions between the quarks and gluons; the fourth line describes the electroweak and Higgs boson masses and couplings; and the fifth line describes the fermion masses and their couplings to the Higgs boson. Since the mechanism for generating the neutrino

masses has not been established, no term has been added to Eq. (1.1) to explain this. The interactions between the fermions and the vector gauge bosons are depicted in the Feynman diagrams shown in Fig. 1.1. The electroweak bosons are allowed to have cubic and quartic couplings of the form $WW\gamma$, WWZ , $WWWW$, $WW\gamma\gamma$, $WWZZ$, and $WWZ\gamma$. Since gluons carry color and, hence, can self-interact, the cubic and quartic couplings of the form ggg and $gggg$ are possible. The Higgs boson couples to massive SM bosons or fermions and can also self-interact. The Higgs boson cubic and quartic couplings among the SM bosons are of the form HWW , HZZ , HHH , $HHWW$, $HHZZ$, and $HHHH$. Feynman diagrams depicting the Higgs boson interactions among the SM gauge bosons and fermions are shown in Fig. 1.2.

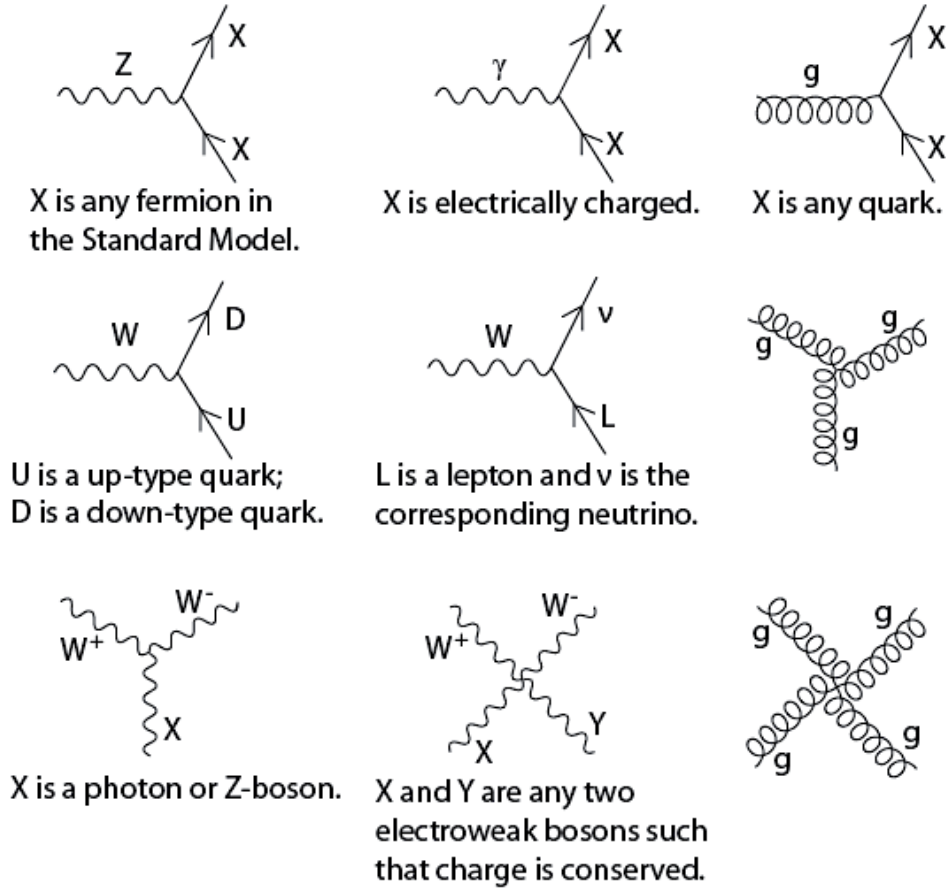


Figure 1.1: Feynman diagrams representing the interactions between the fermions and vector gauge bosons in the SM [24]. The charges of the W bosons are determined by conserving charge with the fermions they interact with.

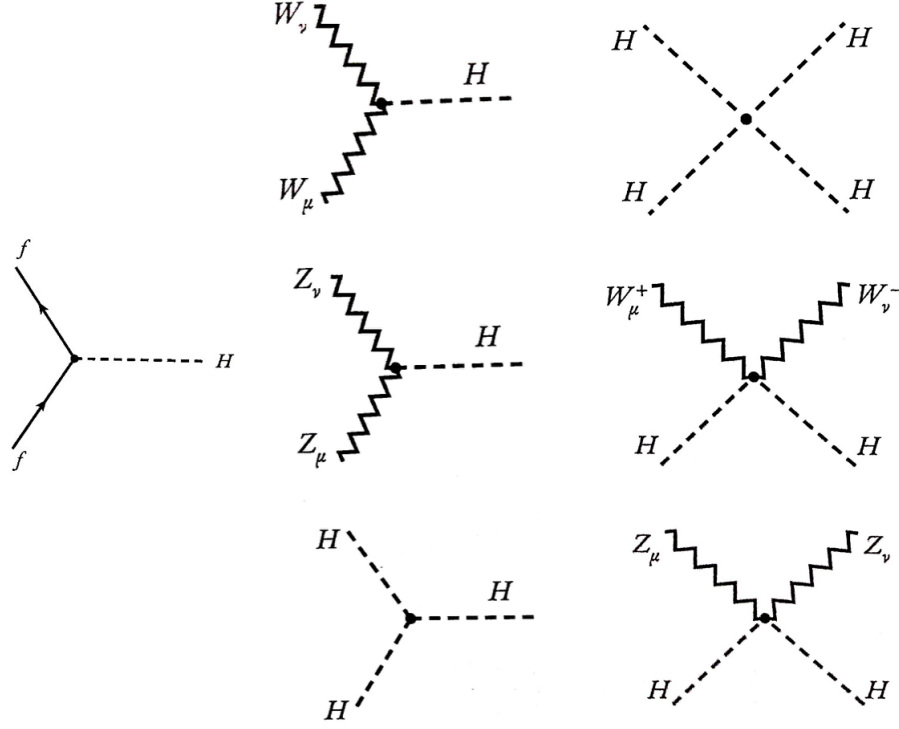


Figure 1.2: Feynman diagrams representing the Higgs boson interactions with the fermions (f) and gauge bosons in the SM [25].

One mechanism for studying the SM predictions and measuring its free parameters is through particle collisions. At the energy frontier, the SM is currently being probed from proton-proton collisions at the Large Hadron Collider (LHC). Protons are baryons made up of three valence quarks (two u quarks and one d quark), gluons, and sea quarks. Any single proton-proton collision actually involves the interactions among its partons, with the largest momentum transfer occurring in the hard parton-parton collision. The distribution among the partons within the protons is determined from experiment and described by parton distribution functions, which will be used and further discussed in Chapter 4.

The success of the SM can be seen in Fig. 1.3, which shows excellent agreement between a wide range of theory and experimental cross section measurements, spanning several orders of magnitude, for the production of various SM processes, as measured by the CMS Collaboration at the LHC. Despite its success, there are several observations not described by the

1.2 Beyond the Standard Model

The shortcomings of the Standard Model suggest that there exists new physics beyond the Standard Model (BSM), offering solutions to its issues. The hierarchy problem, in particular, could be resolved by the existence of extra spatial dimensions in the universe. Among the various extra-dimensional models, the model of large extra dimensions is the focus of this dissertation.

1.2.1 Hierarchy Problem

The hierarchy problem [26, 27] arises from the large difference between the electroweak scale, $M_{\text{EW}} \sim 10^2 \text{ GeV}$, and the Planck scale, $M_{\text{Pl}} = (\hbar c/G_{\text{N}})^{1/2} \sim 10^{19} \text{ GeV}$, where G_{N} is Newton's gravitational constant, and has consequences related to the Higgs boson mass m_H . Since the Higgs boson is a scalar, coupling to mass, it receives large quantum corrections proportional to Λ^2 for an ultraviolet cutoff scale Λ up to where the theory is valid. This scale arises during the integration of the undetermined momenta of the virtual particles in the loop diagrams associated with the quantum corrections and is used to regulate the loop integrals. Since the Higgs boson is the only scalar in the SM, it is the only SM particle not protected from these corrections. The corrections to the physical Higgs boson mass are of the form

$$m_H^2 = m_0^2 + \Delta m_H^2 \tag{1.2}$$

where m_0 is the bare Higgs boson mass and Δm_H^2 are the mass contributions from the radiative corrections. The one-loop corrections due to a fermion f with Yukawa coupling y_f are given by

$$m_H^2 \approx m_0^2 - \frac{|y_f|^2}{8\pi^2} \Lambda^2 \tag{1.3}$$

The contribution from the t quark is largest with coupling $y_t \sim 1$. The dominant one-loop corrections come from the massive gauge bosons, Higgs boson self-coupling, and the t quark

and contribute [25]

$$\Delta m_H^2 = \frac{G_F \Lambda^2}{4\pi^2 \sqrt{2}} (6m_W^2 + 3m_Z^2 + m_H^2 - 12m_t^2) \quad (1.4)$$

where m_W , m_Z , and m_t are the masses of the W boson, Z boson, and t quark, respectively; and G_F is the Fermi constant, which can be related to each particle's coupling constant. The Feynman diagrams depicting these one-loop corrections are shown in Fig. 1.4.



Figure 1.4: One-loop Feynman diagrams for the quantum corrections to the Higgs boson squared mass due to fermions (left) and bosons (right) [29].

The only candidate reference scale for Λ is the Planck scale, at which quantum gravity is expected to emerge. Using this scale, the bare mass and the quantum corrections would have to cancel over 30 orders of magnitude in order to arrive at the experimental value of $m_H \simeq 125 \text{ GeV}$. This extreme degree of cancellation is referred to as fine tuning and seems unnatural. This suggests the possibility of new physics BSM that acts to preserve m_H , eliminating the fine tuning requirement. Many models have been proposed. We consider the addition of extra spatial dimensions in the universe as a possible solution, realized through their modification of the fundamental Planck scale.

1.2.2 Large Extra Dimensions

There are several extra-dimensional models that exist since the original idea by Kaluza and Klein [30–32]. Among these are the model by Arkani-Hamed, Dimopoulos, and Dvali (ADD) [33–35], which proposes flat, large extra dimensions, and the model by Randall and Sundrum [36, 37], which proposes a single, warped extra dimension. A general review of extra-dimensional models and their experimental searches can be found in Section 106 of

Ref. [16] or in the summary by Hewett and Spiropulu [38]. In general, these models modify the fundamental Planck scale, M_D , such that it can be much lower than the observed Planck scale, M_{Pl} . In fact, M_D can be on the order of M_{EW} , which would solve the hierarchy problem. Not only could these models solve the hierarchy problem, but establishing the existence of extra dimensions would fundamentally change our picture of the universe.

In the ADD model, the scale of gravity is fundamentally strong, but the gravitational force effectively becomes diluted by the extra spatial dimensions and appears weak to (3+1)-dimensional observers. This is achieved by confining all SM particles and gauge interactions to a (3+1)-dimensional brane embedded in a larger $(4+n_{\text{ED}})$ -dimensional bulk space with n_{ED} flat extra dimensions assumed to be compactified in a sphere or torus of average radius R . Only gravity is free to propagate through the entire bulk space. Following the ideas in Ref. [33], Gauss's law in this $(4+n_{\text{ED}})$ -dimensional space gives the gravitational potential energy between two test masses m_1 and m_2 separated by a distance $r \ll R$ as

$$U(r) \sim \frac{m_1 m_2}{M_D^{n_{\text{ED}}+2}} \frac{1}{r^{n_{\text{ED}}+1}}, \text{ for } r \ll R \quad (1.5)$$

If these two masses are placed at distances much larger than R , then the usual $1/r$ potential is obtained:

$$U(r) \sim \frac{m_1 m_2}{M_D^{n_{\text{ED}}+2} R^{n_{\text{ED}}}} \frac{1}{r}, \text{ for } r \gg R \quad (1.6)$$

The effective Planck scale M_{Pl} is thus

$$M_{\text{Pl}}^2 \sim M_D^{n_{\text{ED}}+2} R^{n_{\text{ED}}} \quad (1.7)$$

To solve the hierarchy problem, we need $M_D \sim M_{\text{EW}} \sim 1 \text{ TeV}$. Requiring this in Eq. (1.7) yields

$$R \sim 10^{\frac{30}{n_{\text{ED}}}-19} \text{ m} \quad (1.8)$$

Considering values of $n_{\text{ED}} = 1, 2, \dots, 7$, the corresponding sizes of the extra dimensions for

each value are as follows:

$$n_{\text{ED}} = 1, \quad R \sim 10^{11} \text{ m};$$

$$n_{\text{ED}} = 2, \quad R \sim 0.1 \text{ mm};$$

$$n_{\text{ED}} = 3, \quad R \sim 1 \text{ nm};$$

$$n_{\text{ED}} = 4, \quad R \sim 1 \text{ pm};$$

$$n_{\text{ED}} = 5, \quad R \sim 100 \text{ fm};$$

$$n_{\text{ED}} = 6, \quad R \sim 10 \text{ fm};$$

$$n_{\text{ED}} = 7, \quad R \sim 1 \text{ fm}.$$

These distances are large compared to the characteristic distance scales probed by particle physics interactions, even for $n_{\text{ED}} = 6$ or 7 , corresponding to 10- or 11-dimensions, respectively, which is the number of dimensions suggested by different versions of string theory. The case for $n_{\text{ED}} = 1$ would present measurable deviations from Newton's law of gravity, in particular, across solar system distance scales or larger, and can be ruled out. This is not the case for $n_{\text{ED}} \geq 2$. However, tests of Newton's law of gravity at sub-mm distances have started to limit the distance scale for $n_{\text{ED}} = 2$ and astrophysical and cosmological constraints offer small bounds for $n_{\text{ED}} = 2$ and 3 , as summarized in Section 106 of Ref. [16].

Phenomenology

Excitations of the $(4+n_{\text{ED}})$ -dimensional metric correspond to gravitons. The compactification of the n_{ED} extra dimensions, which have periodic boundary conditions, quantizes the graviton's momentum into different eigenmodes as it propagates along the extra dimensions. From the point of view of a $(3+1)$ -dimensional observer, the eigenmodes appear as a tower of graviton excitations called Kaluza-Klein (KK) modes. The momentum eigenmodes in $(4+n_{\text{ED}})$ -dimensions appear as contributions to the mass of each KK mode in $(3+1)$ -dimensions. The massless graviton propagating in $(4+n_{\text{ED}})$ -dimensions is equivalent to

a massive graviton propagating in (3+1)-dimensions. The mass of excitation n is given by

$$m_n^2 = m_0^2 + \left(\frac{n}{R}\right)^2 \quad (1.9)$$

where m_0 is the mass of the ground state KK mode, corresponding to the fundamental graviton mass in (3+1)-dimensions, which is zero.

Each KK mode of the graviton (G_{KK}) couples to SM particles through the stress-energy tensor with a coupling proportional to $1/M_{\text{pl}}^2$. The couplings are enhanced by the large number of modes that are excited at high energy. These G_{KK} modes can decay to SM particles, where they act as propagators in the virtual graviton exchange process. At the LHC, which collides protons, the graviton-mediated processes are initiated through quark annihilation and gluon fusion. The Feynman diagrams depicting these processes with G_{KK} decaying to two photons, i.e., $G_{KK} \rightarrow \gamma\gamma$, are shown in Fig. 1.5.

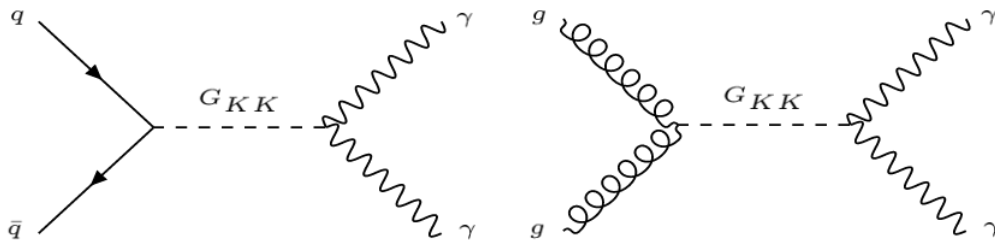


Figure 1.5: Virtual graviton exchange in the diphoton ($\gamma\gamma$) channel through quark annihilation (left) and gluon fusion (right).

The amplitude for any process involving virtual graviton exchange involves a sum over the KK tower of G_{KK} mass states, which can become divergent at high energies, so a cutoff scale needs to be introduced in order to regularize the cross section. Hence, the ADD model is an effective field theory valid only to some cutoff scale Λ_G . This ultraviolet cutoff scale is chosen to be the string scale M_S [39]. This procedure results in a higher dimensional operator with coefficients suppressed by some mass scale [40], which can be parametrized by $\eta_G = \mathcal{F}/M_S^4$, where $\mathcal{F}$ is a dimensionless parameter of order unity for which several conventions exist in the literature. We consider the conventions by Giudice, Rattazzi, and

Wells (GRW) [41]; Han, Lykken, and Zhang (HLZ) [39]; and Hewett [42] expressed as:

$$\mathcal{F} = \begin{cases} 1 & \text{(GRW),} \\ \log\left(\frac{M_S^2}{\hat{s}}\right), & \text{if } n_{\text{ED}} = 2 \\ \frac{2}{n_{\text{ED}}-2}, & \text{if } n_{\text{ED}} > 2 \\ \pm \frac{2}{\pi} & \text{(Hewett),} \end{cases} \quad \text{(HLZ),} \quad (1.10)$$

where $\sqrt{\hat{s}}$, in the context of proton-proton collisions at the LHC, is the center-of-mass energy of the hard interaction from the colliding partons. In the HLZ convention only, $\mathcal{F}$ depends explicitly on n_{ED} and, when $n_{\text{ED}} = 2$, there is an additional dependence on $\hat{s}$.

Considering the interference between the ADD signal and the SM background, the total cross section depends on η_G according to

$$\sigma_{\text{total}} = \sigma_{\text{SM}} + \eta_G \sigma_{\text{int}} + \eta_G^2 \sigma_{\text{ADD}} \quad (1.11)$$

where σ_{SM} is the cross section of the SM background of the desired processes, σ_{ADD} is the cross section of the pure ADD signal, and σ_{int} is the cross section arising from the interference between the two. The HLZ and GRW conventions interfere constructively, while the Hewett convention can interfere either constructively ($\mathcal{F} = +2/\pi$) or destructively ($\mathcal{F} = -2/\pi$).

For KK gravitons in the TeV range, the branching ratios are nearly constant with graviton mass m_G and the total decay width [43, 44] can be expressed as

$$\Gamma_G \simeq \frac{293 m_G^3}{960 \pi \Lambda_G^2} \quad (1.12)$$

The main branching ratios for different SM particles are listed in Table 1.3. These are dominated by dijet (gg and $q\bar{q}$) production, occurring about 2/3 of the time. Because the $\gamma\gamma$ and ZZ modes contain identical particles, they acquire an additional factor of 1/2, as compared to the distinguishable W^+W^- particles. The transverse mode of the Z boson

provides a contribution to the ZZ branching ratio of the same amount as the $\gamma\gamma$ one; however, its longitudinal mode, acquired through the Higgs mechanism, yields one additional contribution to the ZZ fraction not present for $\gamma\gamma$. The ratio of the gg to $\gamma\gamma$ branching ratios is 8, since there are 8 gluons but only a single γ .

Table 1.3: The branching ratios for a massive graviton G_{KK} decaying to SM particles with $q = u, d, c, s, b$ and $l = e, \mu$ [44]. The remaining 24 modes are split evenly among the 3 flavored neutrino-antineutrino pair and $\tau^+\tau^-$ decay modes.

G_{KK} decay mode	HH	gg	$\gamma\gamma$	W^+W^-	ZZ	$t\bar{t}$	$q\bar{q}$	l^+l^-
Branching ratio	2/293	96/293	12/293	26/293	13/293	18/293	90/293	12/293

Searching for this ADD signal is favorable in the high-mass diphoton ($\gamma\gamma$) channel. Even though the KK graviton branching ratio for dijets is dominant among all its decay channels, it suffers from a large background at the LHC. Further, the mass resolution for diphotons is superior to dijets, as provided by the CMS detector at the LHC and discussed in Chapter 2. Compared to the individual dilepton (e^+e^- or $\mu^+\mu^-$) channels, the branching ratio to diphotons is twice as much. The graviton is a spin-2 particle, so its decay to leptons (spin-1/2) results in a restricted dilepton phase space, while photons are spin-1. Therefore, fermions, unlike photons, cannot be produced in s -wave scattering.

In the ADD model, the G_{KK} modes are very finely spaced. Since the energy spacing between adjacent modes goes as $1/R$, they range from about 1 meV to about 100 MeV between $n_{\text{ED}} = 2-7$ [39]. These individual decay modes are too fine to distinguish experimentally, limited by detector resolution. This leads to an effective nonresonant enhancement of the diphoton spectrum at high diphoton invariant mass ($m_{\gamma\gamma}$) resulting in a continuum spectrum of diphotons, rather than discrete resonances. Hence, the experimental signature is the nonresonant production of high-mass diphoton objects above the SM diphoton background.

This class of signatures described for the ADD model is by virtual graviton exchange. Since gravity couples to the stress-energy tensor, a graviton can be added to any fundamental SM vertex, e.g., the s -channel $q\bar{q}\gamma$ production vertex, allowing a search for direct graviton

emission, where a graviton is emitted in the final state, e.g., in $q\bar{q} \rightarrow \gamma G_{\text{KK}}$. These searches depend directly on M_{D} , and are complementary to the searches by virtual graviton exchange, depending on M_{S} , which is expected to be on the order of M_{D} but may be different from it. An example search for direct graviton emission, as described by the ADD model, in the γG_{KK} final state was recently performed by the CMS Collaboration [45]. Interestingly, this model can also be probed by the possibility of microscopic black hole production when the collision energy exceeds M_{D} [46–48]. An example search for microscopic black hole production, as described by the ADD model, by the CMS Collaboration is presented in Ref. [49]. This dissertation searches for signatures of ADD large extra dimensions by virtual graviton exchange in the high-mass diphoton channel using the CMS detector at the LHC.

Limits on ADD Model Parameters

No evidence has yet been seen for the existence of extra dimensions, but they have not been ruled out. Previous searches for extra-dimensional signals in the high-mass diphoton channel at collider experiments have been performed by the ATLAS [50–55], CMS [56–64], CDF [65], and D0 [66] Collaborations. The searches by the ATLAS and CMS experiments used proton-proton collisions at the LHC with center-of-mass energies $\sqrt{s} = 7, 8$, and 13 TeV, while those by the CDF and D0 experiments used proton-antiproton collisions at the Tevatron with $\sqrt{s} = 1.96$ TeV. A summary of the extra-dimensional searches by the CMS experiment using $\sqrt{s} = 7$ and 8 TeV is presented by Landsberg in Ref. [67].

Searches for large extra dimensions arising from collider signals from virtual graviton effects at the LHC have been performed by the ATLAS and CMS Collaborations in the diphoton [55, 64], dilepton [68, 69], ditau [70], and dijet [71] channels. A summary of the corresponding searches in the high-mass diphoton channel by the CMS Collaboration is given in Table 1.4.

The current best limits on the ADD model of large extra dimensions arising from collider signals from virtual graviton effects are due to the CMS dijet angular search [71]. This search

Table 1.4: A summary of the searches for large extra dimensions in the high-mass diphoton channel performed by the CMS Collaboration’s Exotica (EXO) Physics Analysis Group prior to this dissertation. The 95% confidence level observed limits on M_S are given, with the range depending on the model convention.

Search	$\sqrt{s}$	Data	Year	Limits on M_S	Ref.
CMS-EXO-10-026	7 TeV	36 pb ⁻¹	2010	1.31-2.23 TeV	[57]
CMS-EXO-11-038	7 TeV	2.2 fb ⁻¹	2011	2.28-3.50 TeV	[58]

used data corresponding to 35.9 fb⁻¹ with $\sqrt{s} = 13$ TeV. The 95% confidence level limits on M_S range between 8.5-12.0 TeV, depending on the model convention. The use of the dijet angular variable provides superior signal sensitivity than simply considering only the dijet invariant mass. In the diphoton channel, ATLAS sets 95% confidence level limits on M_S between 5.7-8.1 TeV using 37 fb⁻¹ with $\sqrt{s} = 13$ TeV [55]. This dissertation searches in the CMS high-mass diphoton channel using 35.9 fb⁻¹ with $\sqrt{s} = 13$ TeV [64] and aims to extend these results.

1.3 Organization of Dissertation

The remainder of this dissertation is organized as follows. The experimental apparatus used in this large-extra dimensional search is the combination of the LHC and CMS detector, as described in Chapter 2. Event reconstruction and selection within the CMS detector, with an emphasis on photon reconstruction and diphoton selection, are detailed in Chapter 3. Chapter 4 presents the background determination for this nonresonant diphoton search. The ADD large extra-dimensional signal is simulated according to the methods of Chapter 5. Sources of systematic uncertainty are discussed in Chapter 6, and the results are presented in Chapter 7.

Chapter 2

Experimental Apparatus

2.1 Large Hadron Collider

The Large Hadron Collider (LHC) [72] is a superconducting synchrotron that accelerates two counter-rotating, circular beams of hadrons and collides them at four interaction points (IPs) where experiments are placed, as depicted in Fig. 2.1. It is operated by the European Organization for Nuclear Research (CERN), which is headquartered in Geneva, Switzerland, and is their flagship experiment. The LHC is installed in the 26.7 km circumference tunnel that was used by the CERN Large Electron-Positron collider and lies beneath the France-Switzerland border at depths as low as 175 m below the surface. It accelerates separate beams of protons (p) or lead ions (Pb) yielding pp, pPb, or PbPb collisions. The LHC primarily operates using pp collisions. Operation for production physics began in 2010. In October 2017, xenon (Xe) beams were successfully tested and XeXe collisions were achieved.

2.1.1 The CERN Accelerator Complex

The LHC is designed to achieve pp collisions at a center-of-mass energy $\sqrt{s} = 14$ TeV. Protons are delivered to the LHC after going through a series of stages in the CERN accelerator complex, as shown in Fig. 2.2. During each stage, the proton's energy is steadily increased using older accelerators from the previous generations of CERN experiments.

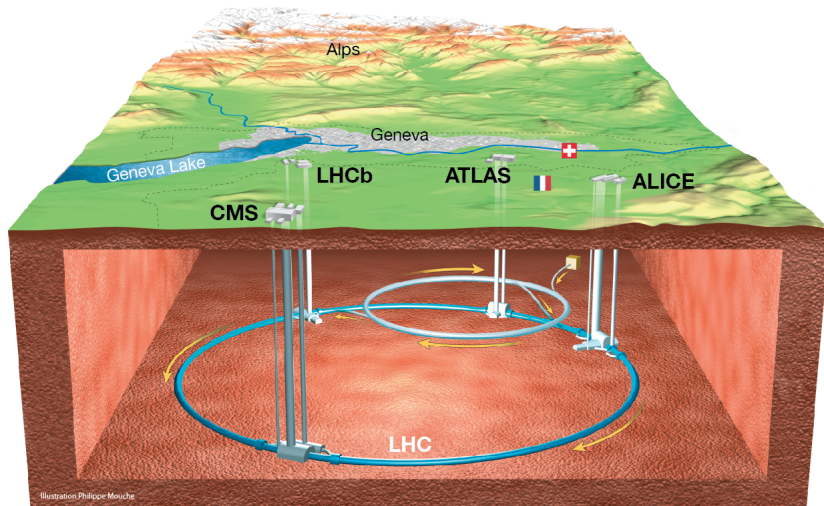


Figure 2.1: An overview of the location of LHC and its four interaction points [73].

Protons are sourced from a bottle of hydrogen gas which is ionized using electric fields. After ionization, the electrons are separated and the protons that are left behind are initially accelerated to 50 MeV using the Linac 2 linear accelerator. The beam is then sent to the circular Proton Synchrotron Booster where it reaches an energy of 1.4 GeV before being injected into the circular Proton Synchrotron and further accelerated to 25 GeV. Next, the beam is transferred to the nearly 7 km circumference Super Proton Synchrotron which accelerates the beam to an energy of 450 GeV. The protons are finally injected into the LHC where they reach their final center-of-mass energy of up to 14 TeV. After acceleration, the two counter-rotating beams are brought together and the protons are collided at the four IPs where detectors are placed.

The LHC has insertions at its Points 1, 2, 5, and 8 housing the IPs where the respective ATLAS [76], ALICE [77], CMS [78], and LHCb [79] experiments are located. ATLAS and CMS are general purpose collider detectors designed to search for and study the Higgs boson, perform precision measurements of SM parameters, study strongly interacting matter, and search for evidence of physics beyond the SM. LHCb is designed to study heavy flavor physics, primarily searching for evidence of new physics in CP violation from beauty and

The CERN accelerator complex *Complexe des accélérateurs du CERN*

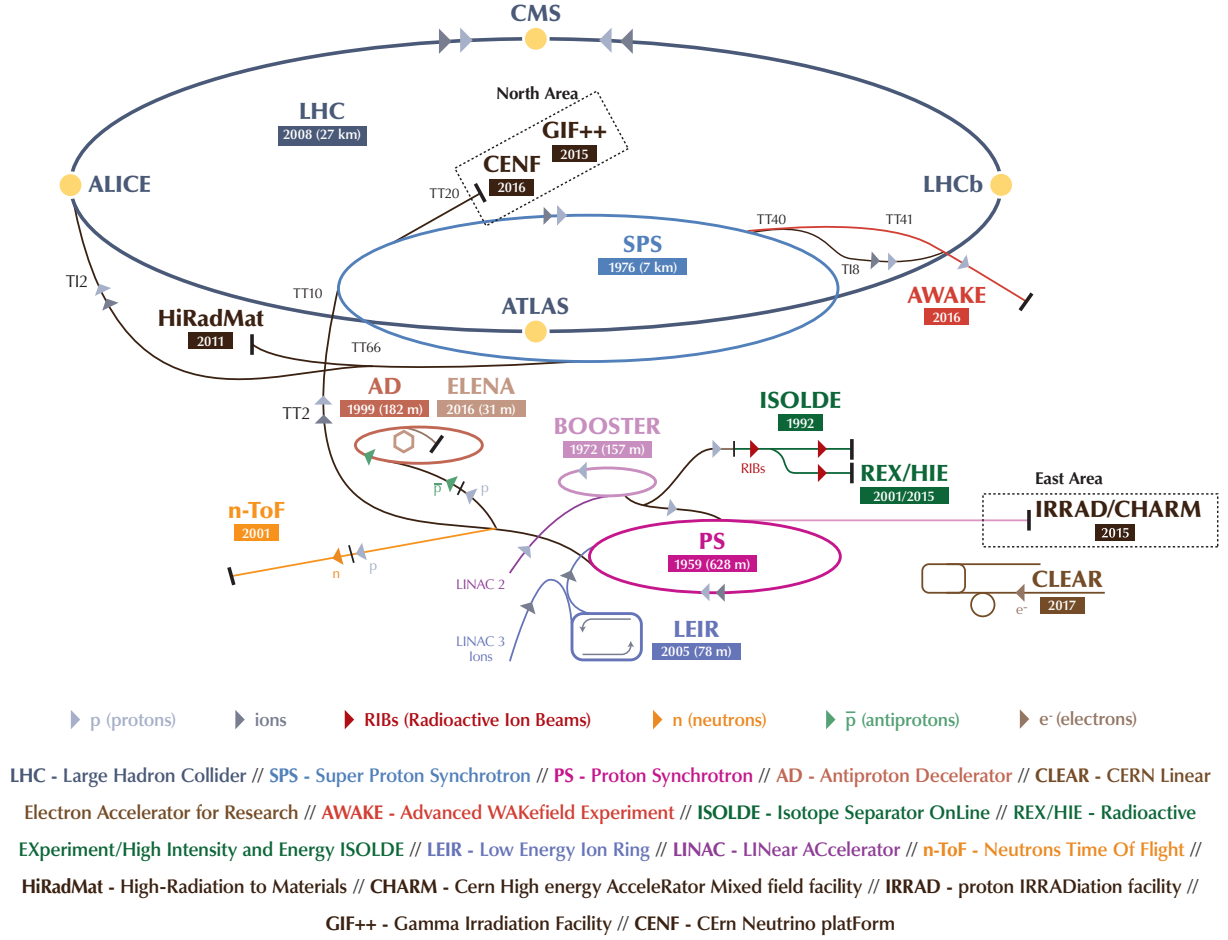


Figure 2.2: The CERN accelerator complex [74].

charm hadron decays. ALICE is a general purpose heavy ion detector designed to study the physics of strongly interacting matter and the quark-gluon plasma. In addition to these four major LHC experiments, four more insertion points are used for LHC operations. Point 4 contains radio frequency (RF) cavities inserted to accelerate the particle beams. The insertions in Points 3 and 7 host collimators used for beam cleaning, and Point 6 contains the beam dump insertion. Fig. 2.3 shows a schematic of the eight LHC insertion points, which includes the four IPs.

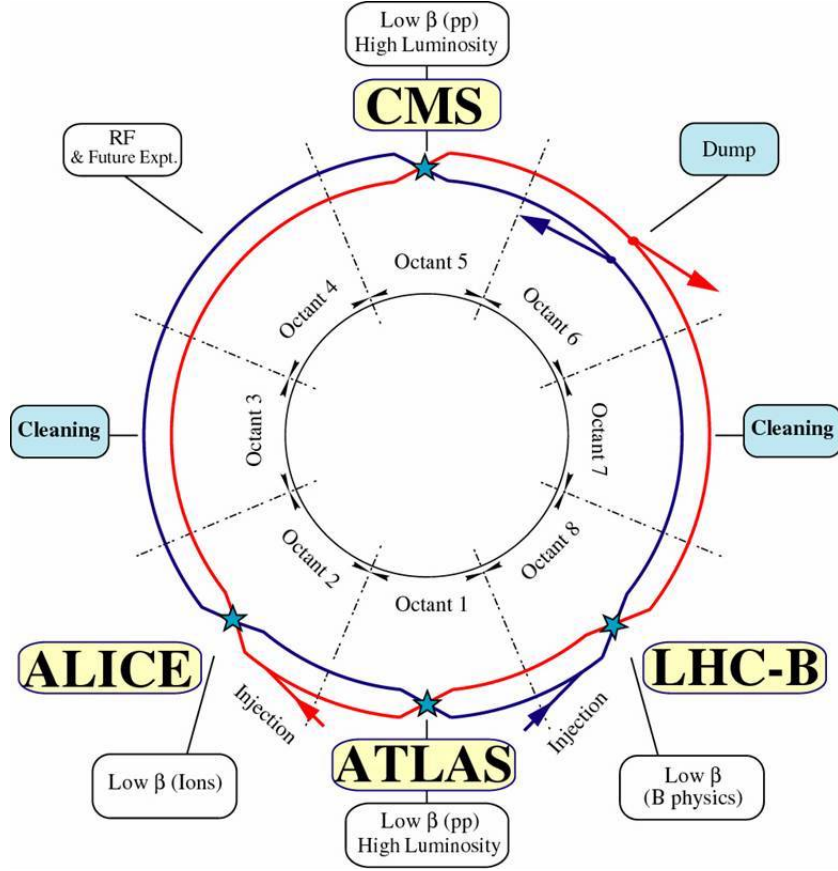


Figure 2.3: A schematic of the eight LHC insertion points [75].

2.1.2 The LHC Components

The LHC uses RF cavities to accelerate the particle beams. Superconducting dipole electromagnets are used to bend the beams and quadrupole magnets are used for focusing. There are 1232 dipole and 392 quadrupole magnets around the LHC ring, each 15 m and 3 m long, respectively. The dipole magnets are constructed using NbTi superconducting coils capable of generating an 8.3 T magnetic field. A cross-section of a dipole magnet is depicted in Fig. 2.4.

The proton beams are divided into bunches containing on the order of 10^{11} protons/bunch. At designed operation, bunches are spaced so that they occur 25 ns apart. Given the LHC orbit length of $89.1 \mu\text{s}$, a maximum of 3564 bunches are possible in each beam. Bunches are injected in the LHC in bunch trains separated by empty bunches which allow for beam

LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

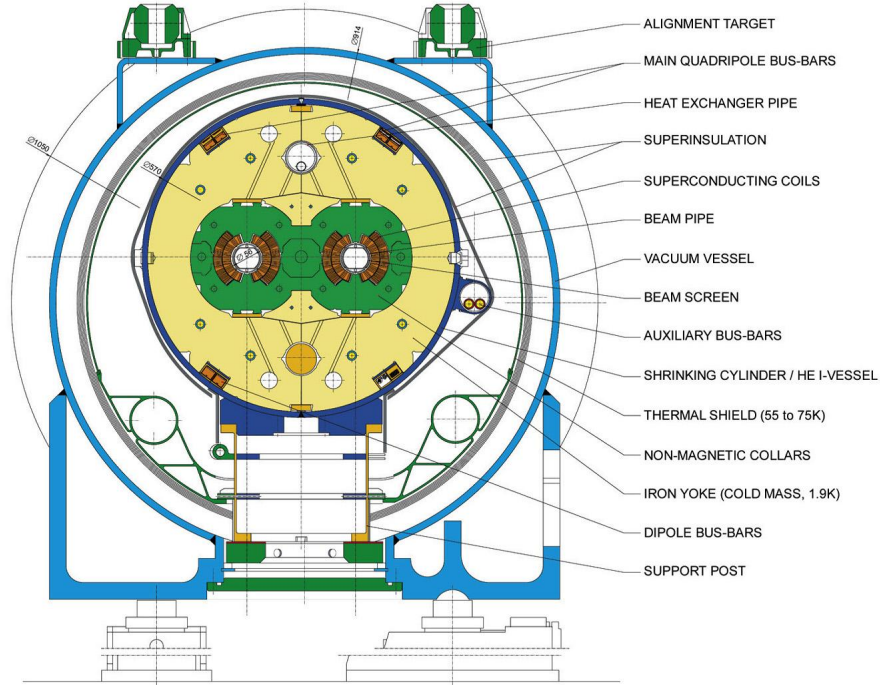


Figure 2.4: A cross-section of an LHC dipole magnet [80].

monitoring, calibration, and control. At nominal operation, 2808 bunches are filled in each beam. The total number of events N produced in the collisions are related to the integrated luminosity L of the LHC fill by

$$L = N\sigma \quad (2.1)$$

where σ is the proton-proton interaction cross section. The total proton-proton cross section at $\sqrt{s} = 14$ TeV is expected to be roughly 100 mb [81]. At $\sqrt{s} = 13$ TeV, CMS measured the inelastic portion of this cross section to be approximately 70 mb in a restricted phase space of the detector [82]. The integrated luminosity is calculated from the instantaneous luminosity $\mathcal{L}$ by the anticipated relation $L = \int \mathcal{L} dt$. The instantaneous luminosity is determined from the beam parameters by

$$\mathcal{L} = \frac{N_b^2 f n_b}{4\pi\sigma_x^* \sigma_y^*} F \quad (2.2)$$

where N_b is the number of protons/bunch, or beam intensity; n_b is the number of bunches;

$f = 11.2455$ kHz is the LHC revolution frequency; $\sigma_{x,y}^*$ are the transverse RMS beam widths at the IP; and F is a geometric loss factor used to correct for the beam crossing angle, shape, and longitudinal beam size. For a Gaussian beam at nominal LHC operation, both σ_x^* and σ_y^* equal $16.6 \mu\text{m}$. The loss factor is about 0.83 [81]. The instantaneous luminosity is not constant throughout the period of duration during stable beam collisions. This is primarily due to beam intensity degradation and emittance growth. The luminosity decays with a lifetime τ according to

$$\mathcal{L} = \mathcal{L}_0 e^{-t/\tau} \quad (2.3)$$

where $\mathcal{L}_0$ is the peak instantaneous luminosity, which, at nominal pp LHC operation, is on the order of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. A snapshot of the LHC status display, taken during pp collisions, is given in Fig. 2.5, showing the decay in instantaneous luminosity over time as measured by the LHC experiments. The lifetime during stable beam collisions is about 20 hours. However, during typical operation, stable beams can be lost to electrical trips or contamination in the beam pipe.

2.1.3 The LHC Schedule

The LHC operates during production cycles known as Runs, which typically last for 3-4 years. During any year within an LHC running period, stable proton-proton collisions used for production physics are only performed for part of that year. For the other part of that year, the LHC is shutdown for about 3-4 months to accommodate upgrades and repairs and additional running time is allocated for beam commissioning, special requests by the experiments for beam use, and heavy ion collisions. Runs are followed by an extended period of accelerator inactivity without beams known as a Long Shutdown (LS). During this time, major upgrades and repairs are performed on the accelerator and each of its experiments. The first pp collisions were achieved in November 2009 and physics production began in March 2010, marking the start of the Run 1 LHC physics program. The current planned

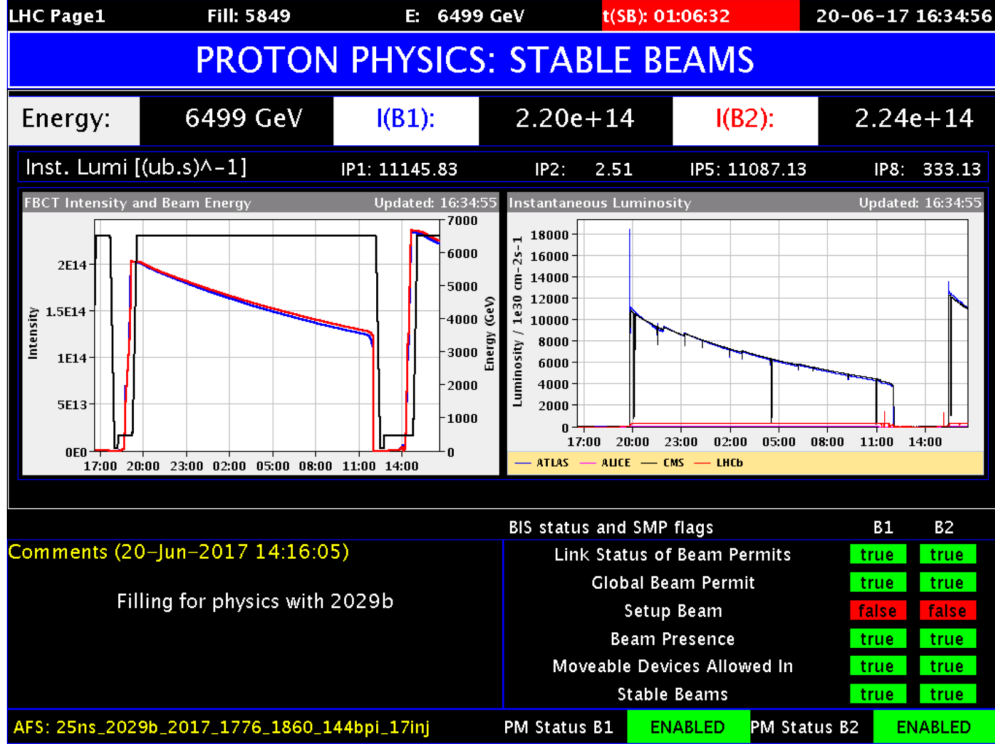


Figure 2.5: A snapshot of the LHC status display, taken in June 2017 during stable beam pp physics collisions, showing the decay in instantaneous luminosity over time as measured by the LHC experiments.

longterm LHC schedule is summarized in Table 2.1.

Table 2.1: A summary of the current planned longterm LHC schedule. A tentative schedule extends past LS3.

LHC period	Duration
Run 1	2010-2012
LS1	2013-2014
Run 2	2015-2018
LS2	2019-2020
Run 3	2021-2023
LS3	2024-

During the first part of Run 1 from 2010-2011, the LHC performed pp collisions at $\sqrt{s} = 7$ TeV. In 2010, the LHC delivered an integrated luminosity of 44.9 pb^{-1} , of which 41.5 pb^{-1} were recorded by CMS. In 2011, this increased to 6.1 fb^{-1} being delivered and 5.6 fb^{-1} recorded. The collision energy for pp collisions was increased to $\sqrt{s} = 8$ TeV during

the second part of Run 1 beginning in 2012. This resulted in the LHC providing an integrated luminosity of 23.3 fb^{-1} , of which CMS recorded 21.8 fb^{-1} and certified 19.7 fb^{-1} for physics analysis use. During Run 1, a 50 ns bunch spacing was used.

Starting with Run 2, the collision energy was further increased to $\sqrt{s} = 13 \text{ TeV}$. During the first two months of physics data taking, a 50 ns bunch spacing was used, but, in August 2015, this spacing was reduced to the nominal LHC value of 25 ns, where it remains. In 2015, the LHC delivered 4.2 fb^{-1} with 3.8 fb^{-1} of it recorded by CMS. This increased as the LHC delivered an integrated luminosity of 40.8 fb^{-1} to CMS in 2016 and 49.8 fb^{-1} in 2017, of which 37.8 fb^{-1} and 45 fb^{-1} was recorded, respectively. In 2018, 68.2 fb^{-1} were delivered by the LHC and 64 fb^{-1} were recorded by the CMS detector, using the current best offline data reconstruction. Future improvements in the offline reconstruction could change the final numbers. This marks the end of Run 2. In total, 163 fb^{-1} were delivered by the LHC and 150.5 fb^{-1} were recorded by CMS with 139.7 fb^{-1} being certified for physics analysis use in Run 2. During Runs 1 and 2, a combined integrated luminosity of approximately 192.5 fb^{-1} were delivered by the LHC to the CMS detector. A summary of this during each year of Runs 1 and 2 is shown in Fig. 2.6. Also shown is a detailed breakdown for 2016, which is the year the dataset used by the analysis in this dissertation was recorded. For more information on the CMS luminosity measurement in 2016, see Ref. [83].

A major goal during LS2 is to prepare the LHC for reaching its nominal collision energy of $\sqrt{s} = 14 \text{ TeV}$, planned for Run 3 operation. By the end of Run 3, it is expected that many of the LHC components will suffer severe radiation damage and will need to be replaced. Therefore, at the start of LS3, the LHC is planned to be upgraded to the High-Luminosity Large Hadron Collider (HL-LHC) [84]. The HL-LHC is being designed to integrate ten times more luminosity than the LHC. This high luminosity environment poses significant challenges to the LHC experiments and will require them to undergo major changes, especially in the forward region of the detectors. To accommodate this, the CMS Collaboration, in particular, is preparing for its Phase II upgrade [85, 86], which includes a proposal to replace its existing

endcap calorimeters with a high granularity calorimeter [87].

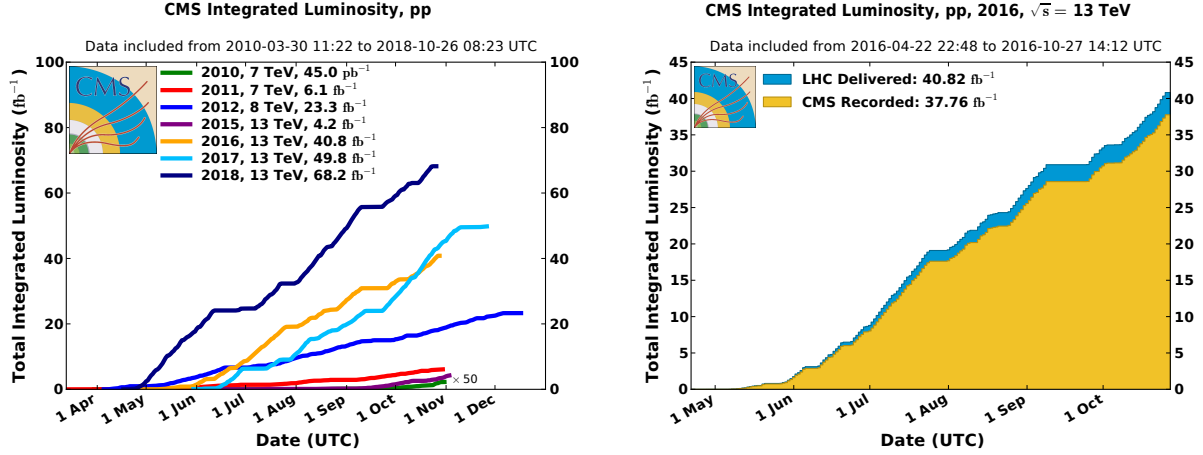


Figure 2.6: A timeline showing the cumulative (left) and detailed 2016 (right) integrated luminosity delivered by the LHC [88]. The amount recorded by CMS in 2016 is also shown.

2.2 The CMS Detector

The Compact Muon Solenoid (CMS) detector [78] is a general purpose particle physics experiment designed for high performance in the search for the Higgs boson and for BSM physics at the TeV scale. It operates as one of the LHC experiments located underground at Point 5 of the LHC ring located in Cessy, France. The overall detector has a cylindrical form approximately 15 m in diameter and 28.7 m in length, weighing about 14000 metric tons. A defining feature of CMS is its 6 m internal diameter superconducting solenoid, providing a uniform magnetic field of 3.8 T inside the majority of its detector volume. CMS comprises several subdetectors arranged in a barrel and endcap configuration, centered on the beam axis, consisting of an inner silicon pixel and strip tracking system; a crystal electromagnetic calorimeter (ECAL); a brass and scintillator hadron calorimeter (HCAL), all of which are mostly contained inside of the solenoid; and gas ionization muon chambers embedded in the steel return yoke of the solenoid. Fig. 2.7 depicts a 3D cutaway model of the CMS detector, highlighting its major components in this configuration.

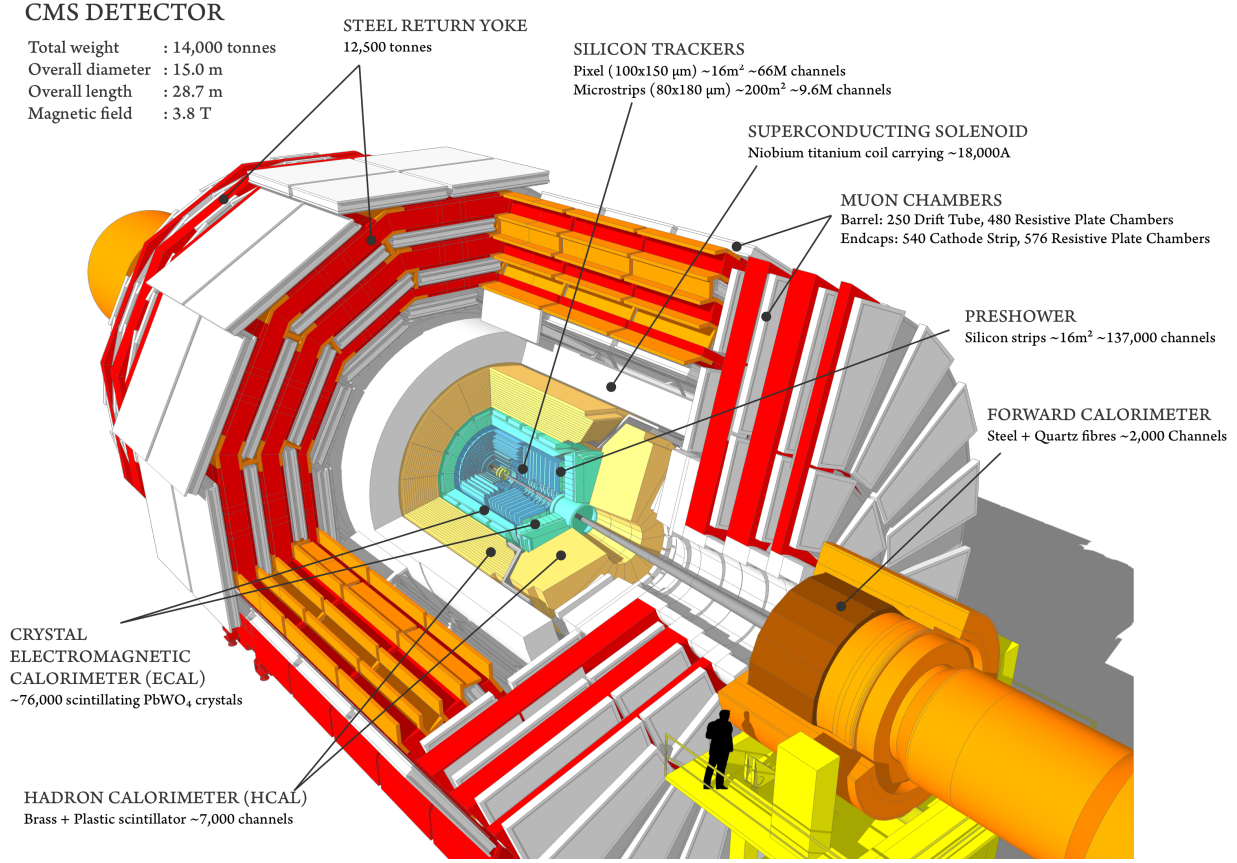


Figure 2.7: An overview of the CMS detector and its major components [89].

Particle detection in CMS relies on information from each subdetector, starting with the inner most detectors. Different particles interact differently with the various detector technologies, possibly leaving behind tracks or energy deposits in different regions, as depicted in Fig. 2.8. This information is read out using a sophisticated trigger and data acquisition system, consisting of the detector electronics, Level-1 trigger, readout network, and high-level trigger. After describing the CMS coordinate system, we will discuss each of its major components.

2.2.1 Coordinate System

CMS uses a right-handed coordinate system with its origin located at the nominal LHC interaction point inside of the detector. The x -axis points inward towards the center of

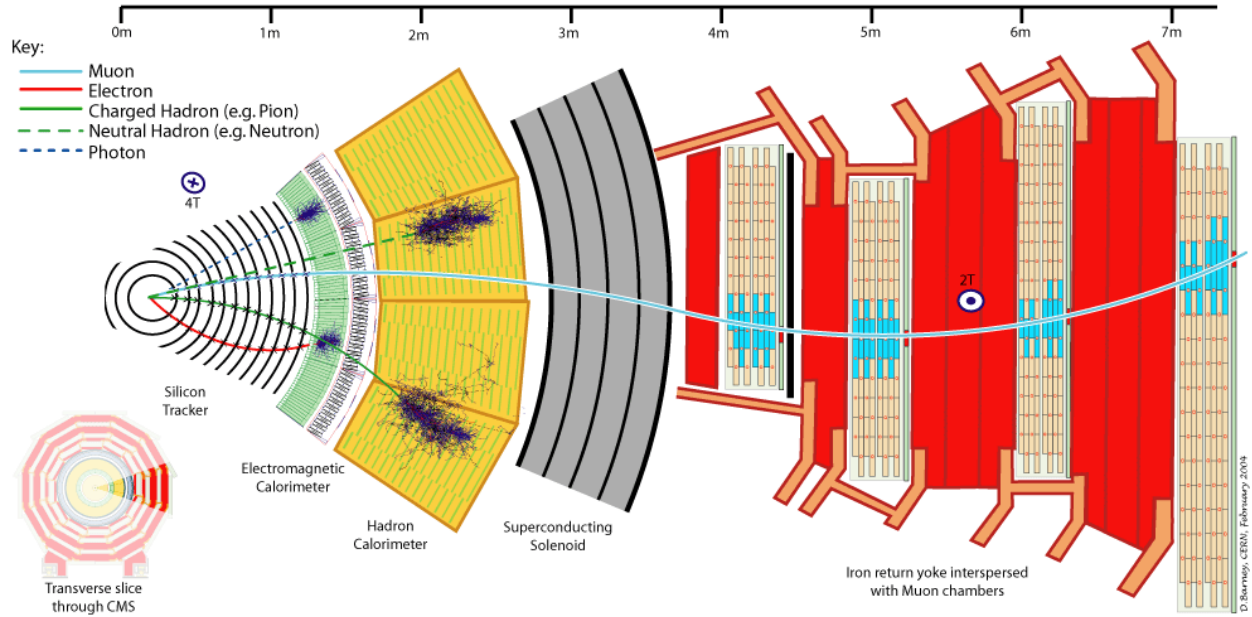


Figure 2.8: A transverse slice of the CMS detector highlighting how different particles interact with its various detector technologies [90].

the LHC, the y -axis points vertically upward perpendicular to the LHC plane, and the z -axis points along the direction of the counterclockwise rotating beam. Due to its geometry, cylindrical coordinates of the form (z, θ, ϕ) are used. The azimuthal angle ϕ is measured from the x -axis in the transverse x - y plane, and the polar angle θ is measured from the z -axis. A diagram illustrating both the Cartesian and cylindrical coordinate systems is shown in Fig. 2.9.

Measurements of particle energy E and momentum p in the transverse plane are denoted $E_T = E \sin \theta$ and $p_T = p \sin \theta$, respectively. Typically, the pseudorapidity, defined as $\eta = -\ln[\tan(\theta/2)]$, is preferred to θ . Particle coordinates are often expressed as (p_T, η, ϕ) , with distances in the η - ϕ plane given by $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$.

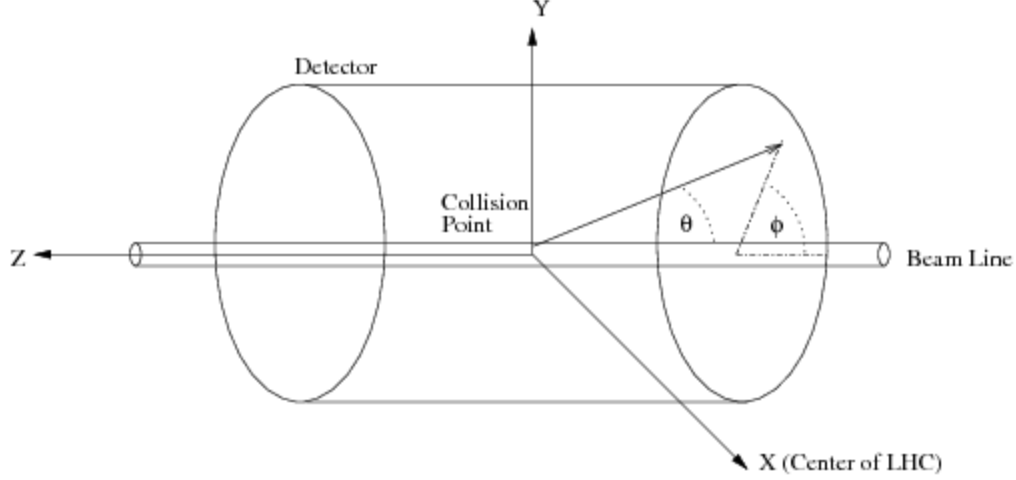


Figure 2.9: An illustration of the coordinate system used by the CMS detector [91].

2.2.2 Inner Tracking System

The inner tracking system of CMS is designed to measure the trajectories of charged particles emerging from the collisions, as well as perform the reconstruction of secondary vertices, such as from b quark and τ lepton decays. This system consists of the silicon pixel and strip tracker subdetectors, laid out according to Fig. 2.10.

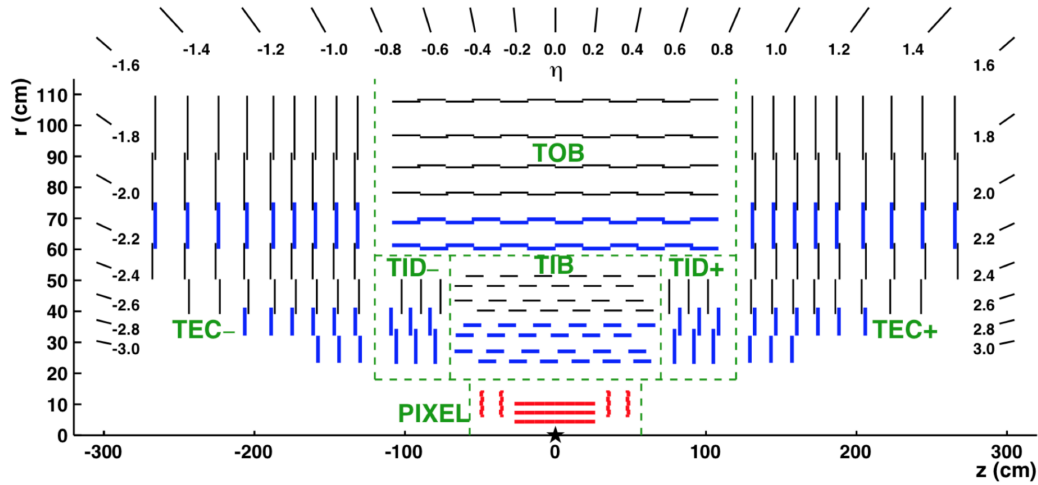


Figure 2.10: A schematic of the CMS inner tracking system [92].

The silicon pixel detector consists of 3 barrel layers (BPix) and 2 endcap disks (FPix), which surround the interaction point, providing coverage in the range $-2.5 < \eta < 2.5$. The

BPix layers are 53 cm long and located at radii of 4.4, 7.3, and 10.2 cm. The FPix are placed at $z = \pm 34.5$ and ± 46.5 cm with radii ranging from 6-15 cm. A single silicon pixel cell has an area of $100 \times 150 \mu\text{m}$. In total, the BPix (FPix) consists of 48 (18) million pixels covering an area of 0.78 (0.28) m^2 .

The silicon strip tracker consists of four subsystems: the tracker inner barrel (TIB), tracker inner disks (TIDs), tracker outer barrel (TOB), and tracker endcaps (TECs). Both the TIB and TIDs are located out to $r = 55$ cm and comprise $320 \mu\text{m}$ thick silicon microstrip sensors arranged in 4 barrel layers and 3 disks, respectively. This inner barrel and disk system is surrounded by the TOB, which has an outer radius of 116 cm and extends between $z = \pm 118$ cm. The TOB consists of 6 barrel layers of $500 \mu\text{m}$ thick silicon microstrip sensors. The TECs cover the range $124 < |z| < 282$ cm, surrounding the rest of the system. Each TEC comprises 9 disks with radii ranging between 22.5-113.5 cm, using silicon strips either 320 or $500 \mu\text{m}$ thick. The entire strip tracker has a total of 9.3 million silicon strips covering a total area of 198 m^2 .

The performance of the inner tracking system is presented in Ref. [93], and the reconstruction of tracks and vertices is discussed in Section 3.1.

2.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is designed to detect and measure the energy of electrons and photons. The ECAL is a hermetic, homogeneous calorimeter comprising approximately 75000 lead tungstate (PbWO_4) scintillating crystals separated into a barrel (EB) and two endcap (EE) sections, which give coverage up to $|\eta| < 3.0$. These crystals were chosen for their high granularity, short radiation length, high radiation tolerance, and fast response time. Avalanche photodiodes are used to detect the scintillation light in the EB and vacuum phototriodes are used in the EEs. The scintillation light is caused by high-energy electrons or photons initiating an electromagnetic shower in the crystals. This occurs for electrons when they undergo bremsstrahlung until their energy loss becomes dominated

by ionization and for photons when they undergo pair production until their energies fall below the pair production threshold. The electromagnetic shower ionizes the atoms in the crystals and light is emitted when they de-excite. The amount of light produced is proportional to the energy of the incident particles. A general description of the passage of particles through matter and, in particular, their detection using calorimeters is given in Sections 33 and 34 of Ref. [16].

The EB is located at $y = 1.29$ m and covers the range $|\eta| < 1.479$. It is made up of 61200 crystals constructed into 36 identical supermodules, each covering half of the barrel length. Each supermodule is made up of 4 modules, and these modules are made up of various submodules each containing 2 rows of 5 crystals. The gap between adjacent crystals is $320 \mu\text{m}$. A single crystal in the EB covers an area of about $22 \times 22 \text{ mm}^2$ in η - ϕ , giving a granularity of $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$. Each crystal is of length 230 mm, corresponding to 25.8 radiation lengths (X_0). The crystals are tilted at 3° with respect to the origin in both the η and ϕ directions, allowing for the detection of particles that would otherwise travel between the gaps of the crystals.

Each EE is located at $z = \pm 3.14$ m and covers the range $1.479 < |\eta| < 3.0$. They are made up of a total of 14648 crystals. Both are divided into two half-disk arrangements constructed with crystals tapered in an x - y grid, grouped into 5×5 structures called supercrystals. A single crystal in an EE covers an area of about $28.6 \times 28.6 \text{ mm}^2$ in x - y and is of length 220 mm, corresponding to $24.7 X_0$. The gap between adjacent crystals is $320 \mu\text{m}$ and between different 5×5 units is 2 mm.

A preshower detector (ES) is placed in front of each EE in the region $1.653 < |\eta| < 2.6$. These are primarily used to help distinguish between photons and π^0 decays which mimic a single photon signature. The ES is a sampling calorimeter made of 2 disks of lead and 2 planes of silicon strip detectors, providing about $3 X_0$ of material. A cutaway view of the ECAL highlighting its main features is shown in Fig. 2.11.

The ECAL uses an active laser monitoring system to measure the relative response of

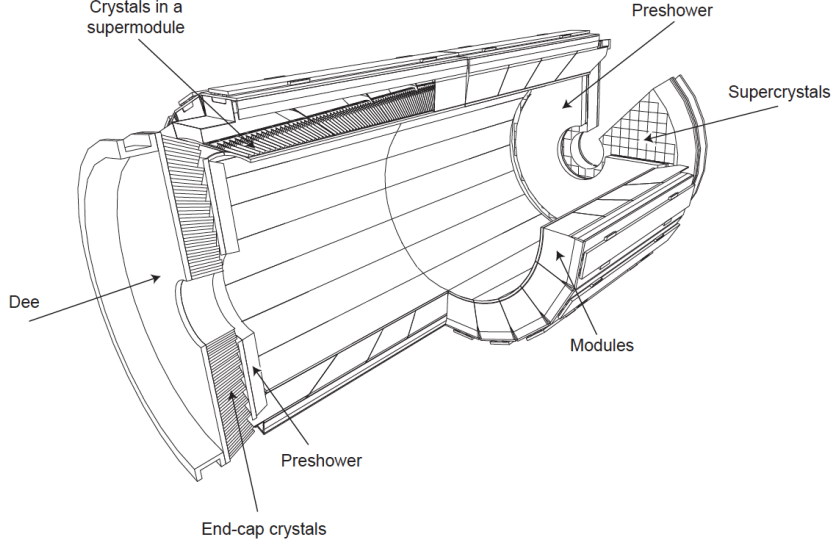


Figure 2.11: A cutaway view of the ECAL depicting its main features [78].

laser light injected into the crystals. The response degrades over time as high radiation doses are received, with limited recovery during periods without collisions, as shown in Fig. 2.12. These measurements are performed every 40 minutes and are used to correct the physics data.

The energy resolution of the ECAL as a function of energy E is characterized by separate terms added in quadrature of the form

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (2.4)$$

for a stochastic term a , a noise term b , and a constant term c . The stochastic term arises from the statistical nature of the shower containment and is primarily caused by intrinsic shower fluctuations and photoelectron statistics. The noise term is primarily due to electronic noise, and the constant term is largely dependent on the intercalibration of the readout channels among different crystals. At high energies, the constant term dominates, but is corrected for and reduced by the laser monitoring system. All of these parameters were measured using an electron test beam and were determined to be $a = 2.8\%$, $b = 12\%$, and $c = 0.3\%$ [95]. The ECAL provides an energy resolution of about 2% in the EB and about 2-

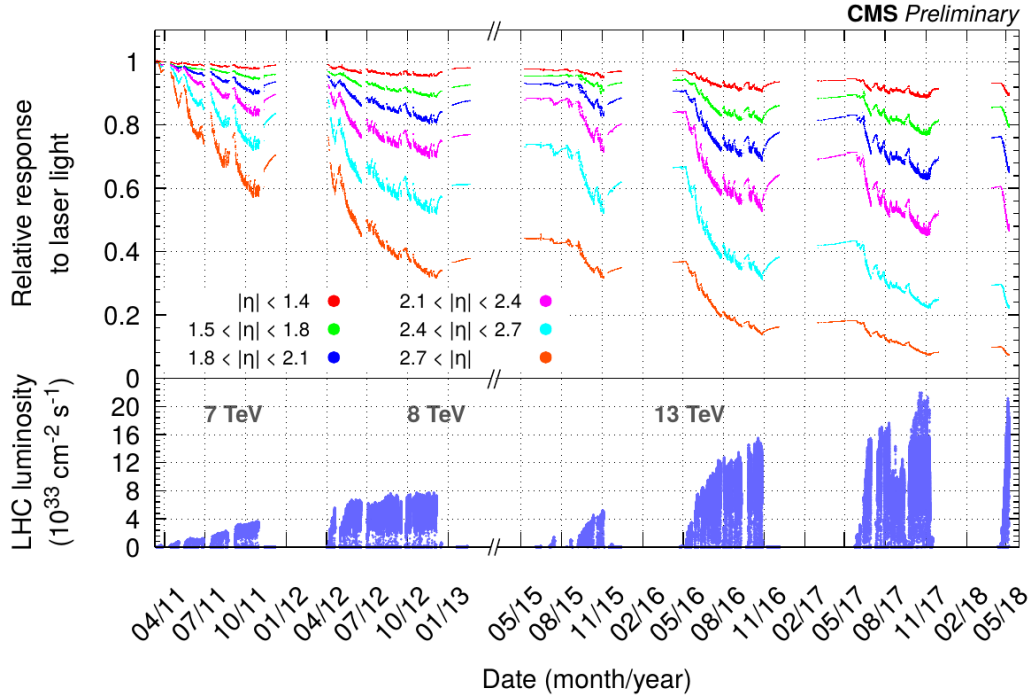


Figure 2.12: The ECAL crystal response measured by the ECAL laser monitoring system over the LHC data taking periods [94].

5% in the EE [95], as measured using $Z \rightarrow e^+e^-$ decays. A detailed description of the energy resolution measurement can be found in Ref. [96] for photons and Ref. [97] for electrons. The reconstruction of electrons and photons is discussed in Section 3.1.

2.2.4 Hadron Calorimeter

The hadron calorimeter (HCAL) is a hermetically constructed sampling calorimeter used to detect and measure the energy of hadronic particles. This is achieved by using alternating layers of absorber and scintillator materials. In most of the HCAL, brass is used as the absorber and plastic scintillator tiles are used as the sampling material. As a hadron travels through the HCAL, it interacts with the absorber, causing a hadron shower. This shower produces scintillation light in the tiles, which is carried through wavelength-shifting fibers to photodetectors. The photodetectors collect and convert the light pulses into electrical signals. The sum of the light across the entire shower is proportional to the energy of the

incident hadrons.

The HCAL comprises four subdetectors: the hadron barrel (HB), hadron endcap (HE), hadron outer (HO), and hadron forward (HF) detectors. Fig. 2.13 shows a 2D projection of a quarter of the CMS detector, highlighting the η coverage of each HCAL subdetector.

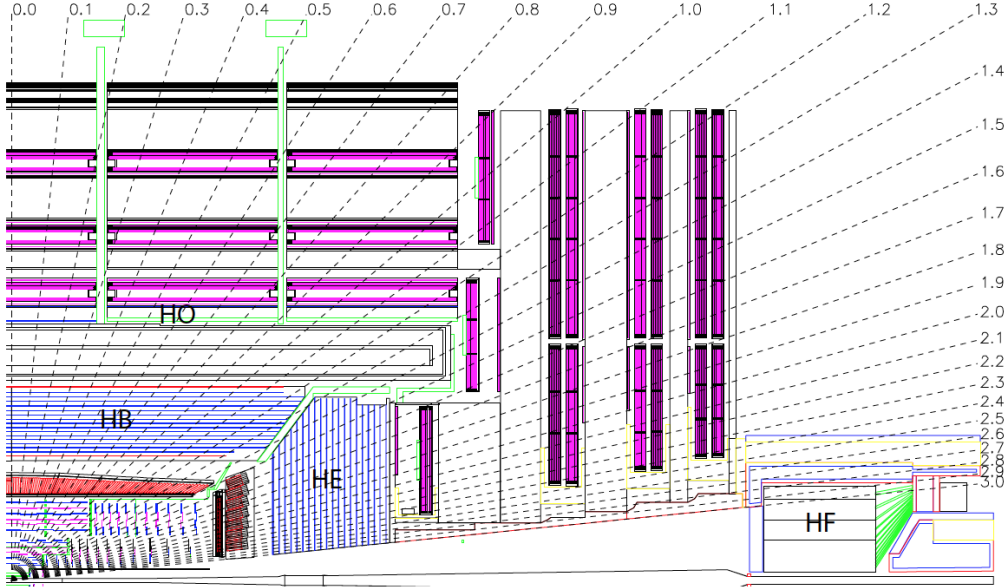


Figure 2.13: A projection of the HCAL, highlighting its subdetectors [78].

The HB comprises 36 identical wedges forming two cylindrical half-barrels, each 4.3 m long. This subdetector extends radially between $1.77 < r < 2.95$ m and is located between the EB and solenoid, covering the region $|\eta| < 1.3$. Although the HB provides excellent shower containment in this region, the HO is added outside of the solenoid, which it uses as an absorber, to provide additional detection for shower tails that may extend outside of the HB. The combination of the HB and HO provide an effective thickness of a minimum of 11.8 nuclear interaction lengths (λ_I). Each HE is located between $3.9 < |z| < 5.7$ m and covers the range $1.3 < |\eta| < 3.0$. The granularity provided by the HB and HEs is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ for $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.174 \times 0.174$ for $|\eta| \geq 1.6$. The scintillation light from the tiles across the depth of a single $\Delta\eta \times \Delta\phi$ region in the HB or an HE is optically added, forming a tower.

The coverage provided by the HEs is extended by the HFs to $|\eta| = 5.2$. Each HF is located at $z = 11.2$ m and uses quartz fibers embedded in a steel absorber to deal with some of the highest particle fluxes in the detector, minimizing radiation damage. Particle showers in the fibers emit Cherenkov radiation, which is detected by photomultiplier tubes. Each HF uses both long and short fibers interlaced throughout its detector volume. Both fibers run the full HF depth of 165 cm ($10 \lambda_I$), but the short fibers start 22 cm from the front end. Both fibers are read out separately to allow for the discrimination between electromagnetic and hadron showers. The former deposit most of their energy within the first 22 cm, while the latter produce nearly equal signals in both calorimeter segments.

2.2.5 Muon System

The muon system is designed to provide robust muon identification with precise measurements of muon momenta. This subdetector is arranged in a muon barrel (MB) and muon endcap (ME) configuration, consisting of three types of gas ionization chambers: drift tube chambers (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). The DTs are segmented into drift cells that measure the positions of muons by determining a drift time through an electric field to an anode wire of a cell. The CSCs are multi-wire proportional counters using a highly segmented cathode strip readout capable of providing an accurate measurement of the positions of the muons in the r - ϕ magnetic bending plane. The RPCs are double-gap chambers that operate in avalanche mode with a fast response and primarily provide timing information for the muon trigger. These chambers are embedded in the steel flux-return yoke outside of the solenoid. Fig. 2.14 shows a schematic highlighting these different features of the muon system.

The MB covers the region with $|\eta| < 1.2$ using DTs, chosen for the suitable detector conditions in that region: a uniform magnetic field, low neutron induced background, and small muon rates. The DTs are arranged into 4 stations, containing 8 chambers capable of measuring the muon position in the r - ϕ plane. The first 3 stations have 4 additional chambers

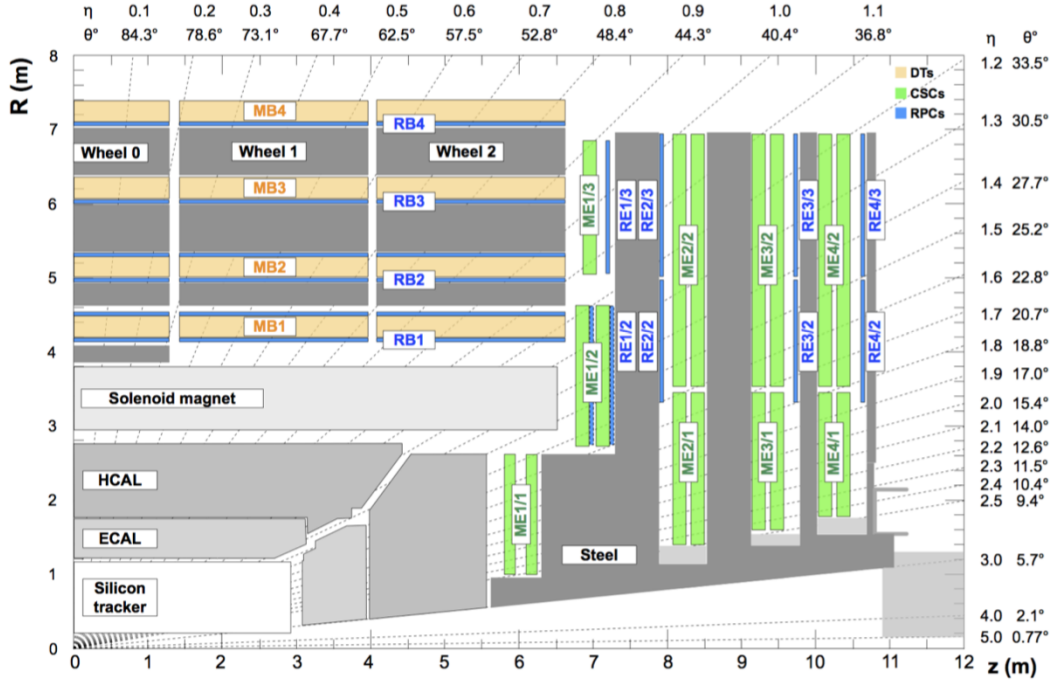


Figure 2.14: A schematic of the CMS muon system depicting its major components and their detector coverage [98].

allowing for position measurements along the z -axis. The MEs cover the region between $0.9 < |\eta| < 2.4$ using CSCs, which can handle the non-uniform magnetic field, high background levels, and large muon rates. Each ME has 4 stations of CSCs comprised of chambers positioned perpendicular to the beam line, providing the full position measurements and beam crossing times for the muons. To complement the DTs and CSCs, the RPCs are added to both the MB and MEs, covering $|\eta| < 1.6$. Their fast response provides good time resolution, but, compared to the DTs and CSCs, a coarser position resolution. There are 6 layers of RPCs embedded in the MB and 3 layers in each ME. Altogether, this provides a muon system capable of triggering on muon p_T with high efficiency ($>90\%$ over the entire η range) and excellent background rejection, while providing a muon p_T resolution of less than 7% in the MB for muon $p_T < 1$ TeV and a muon timing resolution of approximately 1.4 ns [98]. The performance of the muon system is presented in Ref. [98], and further discussion on muon reconstruction is given in Section 3.1.

2.2.6 Trigger System

The LHC provides proton-proton collisions at high rates, occurring every 25 ns, corresponding to a frequency of 40 MHz. This challenging environment is further complicated by the presence of pileup, which is multiple pp interactions occurring within the same event from a single proton-proton bunch crossing. This type of pileup is in-time with the bunch crossing, but pileup can also occur out-of-time from bunch crossings just before and after the collision of interest. When a subdetector's electronics are integrated for more than 25 ns, its response becomes sensitive to out-of-time pileup. The amount of pileup recorded by the CMS detector during each of its data taking years is plotted in Fig. 2.15. An average of 37 simultaneous pp interactions occurred during the 2018 LHC data taking period using $\sqrt{s} = 13$ TeV. At these high rates with this large amount of pileup, it is not feasible to fully process, reconstruct, and store all of the data for each event, so a significant reduction needs to be made. This is accomplished by the trigger system, which functions in two stages: the Level-1 (L1) trigger and high-level trigger (HLT) stages.

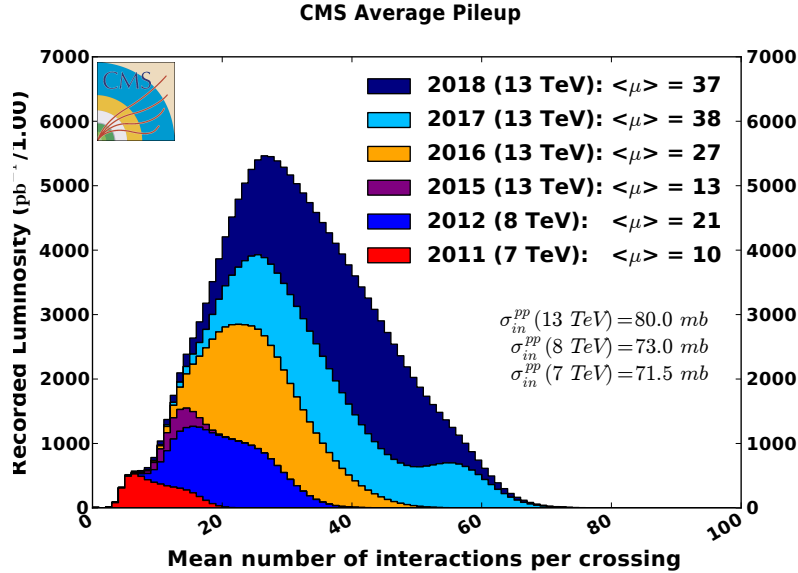


Figure 2.15: The average amount of pileup recorded by the CMS detector during each of its data taking years [88]. The histograms from each year are stacked on the previous year. Also printed are the measured pp interaction cross sections for the different center-of-mass energies used during each data taking period.

The L1 trigger is made up of custom designed and programmable hardware. It analyzes every bunch crossing and is capable of selecting events in order to determine whether or not they should be further processed. This is achieved at a rate of around 100 kHz, with a latency of $3.2 \mu\text{s}$. The L1 trigger uses information from the ECAL, HCAL, and muon system. The inner tracking system takes too long for readout at this stage and is only read out after the event of interest passes the L1 trigger. The calorimeters provide the L1 trigger with trigger primitives based on the presence of large energy deposits. This information is passed to regional and global components of the trigger, where event-level quantities are determined, such as jets, photon/electron candidates, traverse energy, and missing transverse energy. The muon system works to provide track information and muon candidates. The L1 trigger takes the final information provided by these subsystems and generates an L1 Accept (L1A) if trigger criteria are met, marking the event to be fully read out. Upon generation of an L1A, the detector information is passed to the HLT, which reconstructs the event and decides whether or not it should be saved. These steps are described in the flowchart in Fig. 2.16. A more detailed description can be found in Ref. [99].

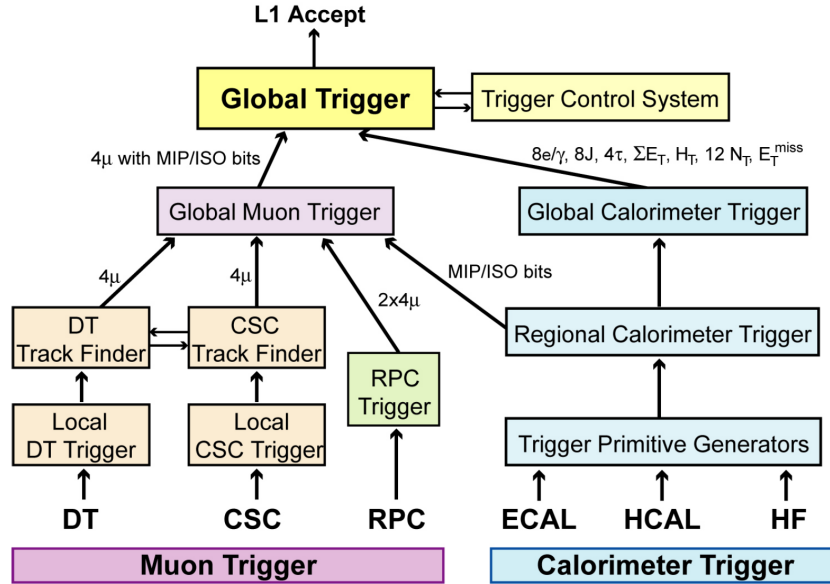


Figure 2.16: A flowchart showing how the global, regional, and local components of the L1 trigger system operate to produce an L1A [78].

The HLT consists of a processor farm running the full CMS online event reconstruction software, which can be continuously updated to meet experimental needs. A sophisticated data acquisition (DAQ) system works with the trigger system to collect and analyze the event information from each LHC bunch crossing. At designed luminosity, the LHC delivers about 1 MB/event of zero-suppressed data containing the full CMS detector information. The DAQ is designed to sustain a maximum input rate of 100 kHz and a data flow rate of approximately 100 GB/s. The HLT reduces the event rate to about 1 kHz. Events are uniquely specified by their run number, luminosity section of the LHC fill, and event number, all specific to the CMS DAQ system.

Chapter 3

Event Reconstruction and Selection

Events are recorded and reconstructed in the CMS detector, then further selection criteria are imposed in order to determine analysis-level physics objects. General event reconstruction will be discussed, followed by the specific diphoton event selection used in this analysis.

3.1 Event Reconstruction

After events are recorded by the CMS detector, they are reconstructed through a variety of algorithms into their constituent components. All electrically charged particles produce signals in the pixel and strip tracking system. Muons are detected within the muon system before, generally, leaving the detector. Electrons and photons deposit most of their energies in the ECAL, while charged and neutral hadrons mostly deposit in the HCAL. The electrical signals produced by each subdetector are read out and converted into physics object candidates. Each subdetector can individually generate different particle candidates; however, CMS uses particle-flow (PF) global event reconstruction [100] to identify and reconstruct individual particles (namely, electrons, muons, photons, charged hadrons, and neutral hadrons) using information from all subdetectors. Jets, missing transverse energy, and other composite objects are built using the PF algorithm.

3.1.1 Tracks and Vertices

The trajectories of charged particles form tracks as they propagate as helices between the tracker layers. Tracks are fundamental for event reconstruction, contributing towards the subsequent reconstruction of other physics objects in the event. CMS uses the Combinatorial Track Finder (CTF) algorithm, which is based on a Kalman filter, for track reconstruction, as described in Ref. [93]. The CTF fits tracker hits, while taking into account the uncertainty of the hit positions and the effects of multiple Coulomb scattering and energy loss. The fit returns track information, including charge, initial momentum, and its impact parameter with respect to the primary vertex.

Primary vertices are those compatible with the hard scattering from the pp interactions along the beam line. Secondary vertices in the event can result from particle decays after the hard interaction. The primary vertex is reconstructed by selecting and clustering tracks compatible with the LHC beam spot, which is the 3D luminous region where the LHC beams collide. The track clustering is done using a deterministic annealing algorithm, and the position of each vertex is fit for using its associated tracks, as described in Ref. [93]. Due to pileup, multiple primary vertices can be reconstructed in an event. However, the main primary vertex is chosen to be the reconstructed vertex with the largest value of summed charged-particle track p_T^2 , with additional constraints described in Section 9.4.1 of Ref. [85]. This selected vertex is simply referred to as the primary vertex. The position resolution for primary vertices is between 10-12 μm for each of its three spatial coordinates [93].

3.1.2 Photons and Electrons

Photon candidates are reconstructed from energy deposits in the ECAL. Before entering the ECAL, a photon may convert in the tracker into an electron-positron pair, whose paths bend under the influence of the magnetic field, causing their energy deposits to be extended along the ϕ direction. Individual energy deposits are grouped into superclusters to allow

maximal recovery, using different procedures in the EB and EE regions. ECAL crystals are aligned in an (η, ϕ) grid in the EB and an (x, y) grid in either EE. First, clusters are built starting with a seed crystal, which is located by selecting the crystal with the highest E_T above a certain threshold among its neighboring crystals. Next, in the EB region, a 5×1 bar of crystals centered on a seed is formed along the η direction. Connected bars within ± 17 crystals along ϕ from the seed crystal are added if their total energy is above some threshold, forming a cluster. Other clusters, built in this way, having seeds aligned in η with this cluster's seed are added, based on a further threshold requirement, to form superclusters. In the EE regions, a cluster of a 5×5 array of crystals centered on a seed is first built. The crystals surrounding the 5×5 boundary are allowed to seed new 5×5 clusters using crystals not included in any cluster yet. The clusters are allowed to partially overlap, and a supercluster is formed from the connected clusters. In either the EB or an EE, the superclusters formed for unconverted photons are typically 5×5 matrices. Further details are described in Ref. [96].

The raw energy for photons is obtained by summing the energy deposits in the supercluster, corrected for detector effects, as described in Section 2.2.3. Energy deposited in an ECAL preshower detector is added to the photon energy measurement from the adjacent EE. This raw energy needs to be corrected, mainly to account for photon shower loss and contamination from pileup. The shower may be lost between gaps in the crystals or through gaps in the ECAL modules, resulting in parts of the shower not being clustered into the supercluster. Data-driven scale factors for the photon energy scale and resolution are extracted from $Z \rightarrow e^+e^-$ decays, with electrons reconstructed as photons, using an energy regression based on a boosted decision tree (BDT) [96]. Because of the limited dynamic range of the ECAL readout electronics, they saturate when the energy deposited in a single crystal is greater than an average value of about 1.7 (2.8) TeV in the EB (EEs), with a spread of about 10 (25)% among different channels. Energy corrections to account for this are derived using a simulated sample of photon candidates with energies spanning a large range. The

photon energy scale and resolution are corrected by comparing the difference between data and simulation from $Z \rightarrow e^+e^-$ events, calibrated at the Z boson peak. The difference in energy scale is used to correct the data. The energy resolution in simulation needs additional Gaussian smearing to match that of data, which is considered as a correction. In the EB, photons in the range of tens of GeV have an energy resolution of about 1%, while photons above this range have a resolution of about 1.3% up to $|\eta| = 1$, rising to about 2.5% at $|\eta| = 1.4$. In each EE, the energy resolution for photons in the range of tens of GeV is about 2.5%, while photons above this range have a resolution between 3 and 4% [96].

Electrons are reconstructed by matching a track to an ECAL supercluster, determined using the same method as for photons. The electron track reconstruction uses a Gaussian sum filter (GSF) to fit the tracker hits. The emitted bremsstrahlung photons, which can carry away a significant fraction of the initial electron energy, can convert into electron-positron pairs, yielding secondary tracks. This is taken into account by the GSF. The details are described in Ref. [97].

3.1.3 Muons and Tau Leptons

Muon reconstruction uses information from both the muon system and inner tracker. Muon tracks are first reconstructed in the muon system to form track segments generated by requiring a certain number of hits in the layers within a single DT or CSC chamber. These segments are then fit, with information from the RPCs included, to form standalone muons. These standalone candidates are then matched to tracks from the inner tracking system by projecting the standalone muon trajectories to the outer tracker surface and using a method similar to the CTF algorithm (see Ref. [98] for details). Muon performance was discussed in Chapter 2.

The τ lepton signature is more complex than for electrons or muons. They are the heaviest of the charged leptons, with a short lifetime, and will typically decay before reaching the active material of the detector. About 2/3 of the time, τ leptons decay hadronically and can

be reconstructed indirectly through their PF candidate decay products. The performance of τ lepton reconstruction and identification is presented in Refs. [101, 102].

3.1.4 Jets and Missing Transverse Energy

A typical proton-proton collision will produce quarks or gluons either through the hard interaction directly or through QCD radiation from the initial or final state partons. Due to QCD confinement, colored partons are not freely observed. Instead, they transform into a collection of color-neutral hadrons, observed as a cluster of particles called jets. Because the number of color states for gluons is larger than for quarks, gluon-initiated jets, in general, have higher particle multiplicity, a softer fragmentation function, and are less collimated than quark-initiated jets [103, 104]. So, quark jets tend to have lower multiplicity, harder constituents, and are narrower. In addition, quarks have electrical charge, while gluons do not. Jets can be further classified according to their quark flavor. Heavy-flavor jets originate from b or c quarks, while light-flavor jets are those originating from u , d , or s quarks. In some cases, jet identification techniques are able to discriminate between these different types of jets [105, 106]. Top quarks decay before they hadronize and are identified through their decay products (see, e.g., Ref. [107]).

Jets are built from PF candidates, primarily detected as charged hadrons (e.g., $\pi^\pm$, $K^\pm$, and protons), neutral hadrons (e.g., K_L^0 and neutrons), or nonisolated photons (e.g., from π^0 decays). The energy deposits are calibrated to account for the nonlinear response of the calorimeters, and reconstructed tracks are compared to the clusters. The energy deposits are clustered into jets using the infrared and collinear safe anti- k_T algorithm [108, 109], with a distance parameter of $R = 0.4$. Infrared (collinear) safety means the jet algorithm remains invariant under soft emission (co-moving / collinear momentum splitting) of the clustered particles. Heavy-flavor jets will typically produce a secondary vertex. For example, a B meson such as a B^+ will travel an average distance of approximately 500 μm before decaying. This is well within the vertex resolution of the tracking system. Jets originating from b quarks

are identified with the combined secondary vertex (CSVv2) b tagging algorithm [106]. Since the lifetime of a c quark is shorter than that of a b quark, a similar, but modified, procedure is used for c jet identification, discussed in Ref. [106].

Neutrinos interact only through the weak interaction and escape through the detector. This is also true for any (stable) electrically neutral, colorless particle. Nonetheless, their presence can be inferred through a momentum imbalance in the collision using conservation of momentum, given that the initial transverse momentum of the colliding system is negligible. In the transverse plane, this would show up as missing E_T , referred to as MET, equal to $|\sum \vec{p}_T|$. Along the longitudinal axis, the missing E_z is not useful since the initial longitudinal momenta of the colliding partons within the protons are unknown.

3.2 Event Selection

The signature in this analysis consists of two high-energy, prompt photons. Events containing this signature are reconstructed and required to pass a trigger selection. These photons form diphoton candidates, which are required to pass a preselection and further identification criteria, imposed to obtain high event purity with high search sensitivity.

3.2.1 Trigger Selection

Events are selected which pass the CMS HLT trigger with path name `HLT_DoublePhoton60`. These events contain at least two photon candidates, each with $p_T > 60$ GeV, across the range $|\eta| < 3.0$. The primary selection at this level requires each photon candidate to have $H/E < 0.15$. The longitudinal shower shape variable H/E is defined as the ratio of the hadronic to electromagnetic energy fraction of the photon candidate. The electromagnetic energy fraction is defined as the corrected supercluster energy. At this trigger level, the hadronic energy fraction is defined as the total HCAL energy within a cone of $\Delta R < 0.15$ from the supercluster position, as used in the variable referred to as Hadronic/E in Section 4.2.

Later, at the analysis level, this fraction will mean the amount of HCAL energy in the tower directly behind the photon seed crystal. This trigger is seeded by the logical OR of a suite of `SingleEG`, `DoubleEG`, `SingleJet`, and `SingleTau` CMS L1 seeds, with varying thresholds. These thresholds are primarily sensitive to one or two large electromagnetic energy deposits, a large hadronic energy deposit, or a signature in the muon and calorimeter systems consistent with the presence of a single τ lepton decay, respectively.

The trigger efficiency measured relative to the `HLT_DoublePhoton40` trigger, which requires the two photon candidates to each have $p_T > 40$ GeV, is shown in Fig. 3.1. The photons are ordered by p_T in each event, with the leading photon having the highest p_T and the subleading photon having the second highest. The efficiency shown here is measured as a function of the subleading photon p_T . This selection is fully efficient above photon $p_T > 75$ GeV.

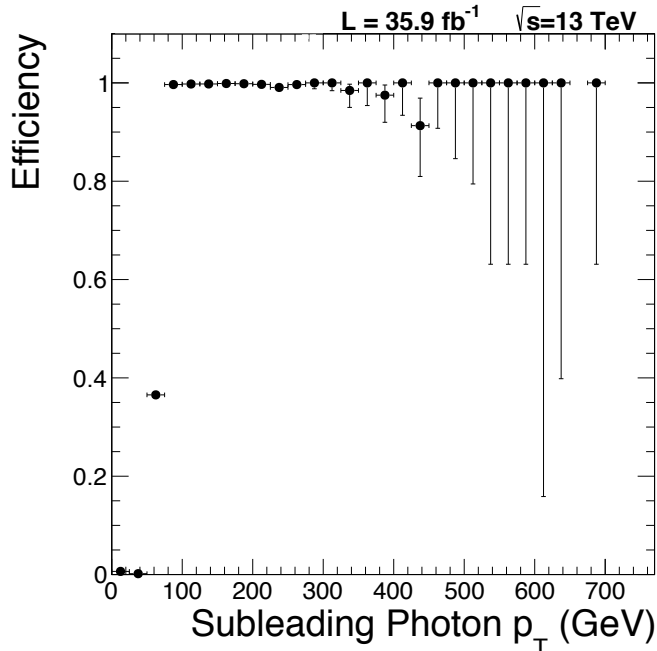


Figure 3.1: The relative efficiency of the `HLT_DoublePhoton60` trigger (measured with reference to the `HLT_DoublePhoton40` trigger) as a function of the p_T of the subleading photon.

The absolute trigger efficiency of `HLT_DoublePhoton60` was measured using a simulated sample of diphoton events, similar to those described in Chapter 5. In this simulated sample,

the diphoton invariant mass $m_{\gamma\gamma}$ is known and can be compared before and after imposing the trigger selection. Fig. 3.2 shows the trigger efficiency as a function of $m_{\gamma\gamma}$, which is fully efficient for events with $m_{\gamma\gamma} > 500$ GeV. The two plots shown correspond to the different analysis event categories, which will be explained in the following section.

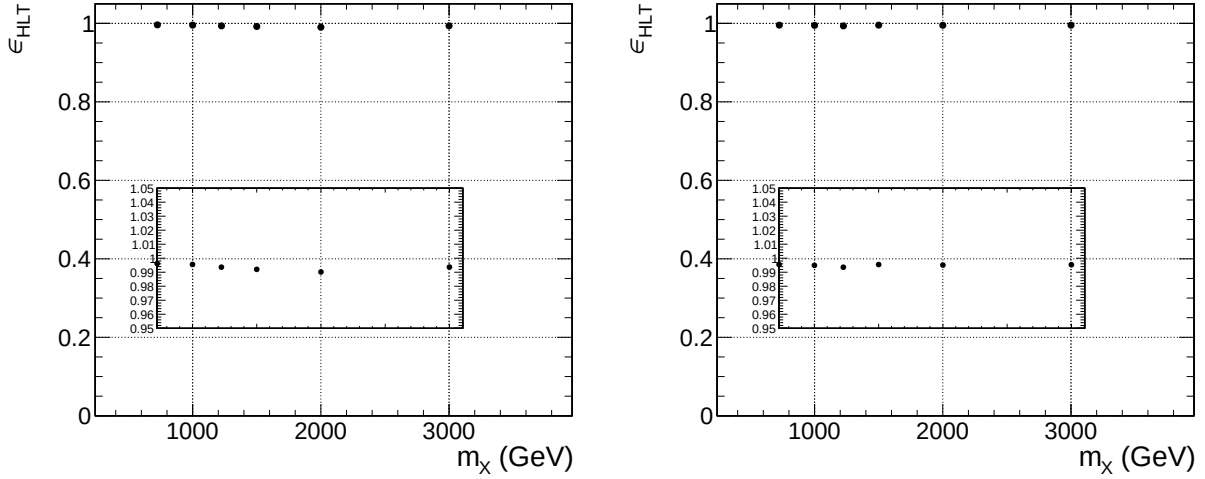


Figure 3.2: The absolute efficiency of the HLT_DoublePhoton60 trigger (measured using a simulated diphoton sample) as a function of $m_X = m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories. The inner boxes show a zoomed-in region of the y -axis.

To protect against potentially missing high-mass diphoton candidates at the trigger level, the backup trigger HLT_ECALHT800 is applied in addition to HLT_DoublePhoton60. This trigger solely on the presence of large energy deposits in the ECAL without imposing additional criteria, such as a selection in HLT_DoublePhoton60 that could be responsible for causing the diphoton candidates to be rejected. The contribution from HLT_ECALHT800 to the analysis dataset is minimal (yielding ~ 10 events).

3.2.2 Event Preselection

After trigger selection, candidate diphoton events are required to satisfy the following kinematic selection:

- each photon candidate is required to have $p_T > 75$ GeV;

- one photon is required to be in the EB with $|\eta| < 1.4442$, and another in either the EB or in an EE, where it must have $1.566 < |\eta| < 2.5$;
- the photon pair must satisfy $\Delta R > 0.45$; and
- $m_{\gamma\gamma} > 500$ GeV.

The trigger efficiency measurements are one of the motivations for the photon $p_T > 75$ GeV and $m_{\gamma\gamma} > 500$ GeV requirements. In this analysis, there are two signal regions considered: one with both photons in the EB, denoted EBEB, and the other with one photon in the EB and the other in either EE, denoted EBEE. Events with both photon candidates in the EEs are dominated by SM production and have negligible sensitivity to the target BSM signal, and thus are omitted from this analysis. Additional categories were considered, but were also excluded for this reason. We use the fiducial region of the ECAL, defined with a restricted η range, as specified above, which rejects events outside of the tracker coverage and in the gap between the EB and EEs. The η coordinate specifying this region corresponds to that of the supercluster position with respect to the fixed detector coordinate system, i.e., it is not relative to the shifting position of the primary vertex of each event. The ΔR requirement has a minimal effect on the overall event selection, but is motivated by a consistency requirement with the background simulation, as explained in Section 4.1. If more than one diphoton candidate satisfies this set of selection criteria, only the photon pair with the largest scalar sum of photon p_T is retained.

The CMS standard primary vertex algorithm, as described in Section 3.1.1, is suboptimal when used for events based on neutral particles because these particles do not contribute to the selection variable $\sum p_T^2$ of charged particle tracks. In this case, improvements to the vertex finding can be made by analyzing the correlation between observables related to tracks recoiling against the diphoton system. The algorithm used by the $H \rightarrow \gamma\gamma$ analysis [110, 111] is incorporated in this analysis and is based on using this information as input to a multivariate classifier (a BDT). This algorithm is found to converge to the standard primary

vertex algorithm at high $m_{\gamma\gamma}$. For diphoton events above $m_{\gamma\gamma} > 500$ GeV, the interaction vertex is correctly assigned 90% of the time, as measured in simulation and compared against data. If the position of the vertex along the z -axis is known to better than about 10 mm, then the diphoton mass resolution is dominated by the photon energy resolution. This is achieved by this algorithm at high $m_{\gamma\gamma}$.

3.2.3 Photon Identification

A dedicated set of photon identification criteria were developed for this analysis, tuned for high-mass photons. These criteria are based on photon shower shape and isolation variables, which help suppress misidentifications from jets and electrons (described further in Section 4.2) while providing a high selection efficiency. One or two jets can fragment in such a way as to mimic a photon signature in the ECAL, causing the jet to be misreconstructed as a photon. For example, the decay of a neutral pion to two photons, $\pi^0 \rightarrow \gamma\gamma$, can fake a photon signature when both photon showers overlap in the ECAL. Electron showers are similar to photon showers in the ECAL, and if tracks are not properly assigned to an electron, they can fake a photon signature as well. In order to improve photon purity, we use identification criteria for photon candidates that are sensitive to the presence of more than one electromagnetic shower in the ECAL and to extra activity around the shower. Individual shower shape and isolation variables are used to achieve this.

The lateral shower shape variable $\sigma_{i\eta i\eta}$ is the spatial second moment of the supercluster about its average η position. It can be thought of as the log energy-weighted RMS of the photon shower in units of the crystals, calculated as

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_{i \in 5 \times 5} (\eta_i - \bar{\eta})^2 w_i}{\sum_{i \in 5 \times 5} w_i}} \quad (3.1)$$

with $w_i = \max(0.0, 4.7 + \ln E_i/E_{5 \times 5})$, where $\bar{\eta}$ is the average supercluster position in η , η_i is the η coordinate of crystal i in the 5×5 array, E_i is the energy of crystal i , and $E_{5 \times 5}$

is the energy of the full 5×5 array centered on the seed crystal. A larger value of $\sigma_{i\eta i\eta}$ indicates a larger photon shower, essentially measuring the width of the photon shower in the η direction. Prompt photons are well isolated in $\sigma_{i\eta i\eta}$. In fact, $\sigma_{i\eta i\eta}$ is part of a covariance matrix describing the spatial extension of the photon shower in (η, ϕ) space:

$$\begin{pmatrix} \sigma_{i\eta i\eta} & \sigma_{i\eta i\phi} \\ \sigma_{i\eta i\phi} & \sigma_{i\phi i\phi} \end{pmatrix}$$

where $\sigma_{i\phi i\phi}$ and $\sigma_{i\eta i\phi}$ have analogous definitions as for $\sigma_{i\eta i\eta}$. The variable $\sigma_{i\phi i\phi}$ is used in the background calculation, as described in Section 4.2. High-energy photons causing saturation in the ECAL will have an effective wider shower in $\sigma_{i\eta i\eta}$, as shown in Fig. 3.3. The shower only appears wider since the central peak in the $\sigma_{i\eta i\eta}$ distribution is suppressed, due to the readout suppression above a certain threshold. In Fig. 3.3, the tall edge on the left of the $\sigma_{i\eta i\eta}$ distributions are understood to be caused by prompt photons depositing their energy into the cracks between the ECAL crystals along the η direction.

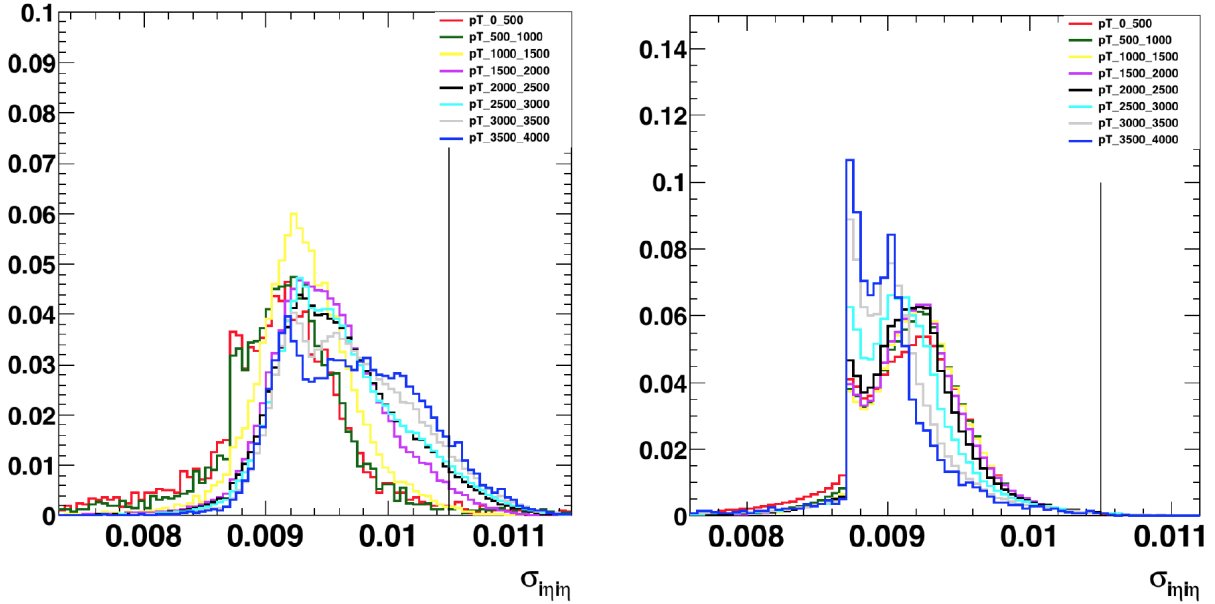


Figure 3.3: A plot of $\sigma_{i\eta i\eta}$ in various photon p_T bins for saturated (left) and unsaturated photons (right) in the EB. The vertical line is at the value of 0.0105.

The variable H/E , as defined above in Section 3.2.1, is the ratio of the hadronic to

electromagnetic energy fraction of the photon candidate. This variable is close to zero for prompt photons, as the photon shower is well contained in the ECAL with little leakage into the HCAL. The R_9 variable is defined as the energy sum of the 3×3 array of crystals centered on the seed crystal divided by the energy of the supercluster. R_9 is sensitive to photon conversions, with unconverted photons having high values.

Isolation variables are based on the total scalar p_T sum of PF candidates assigned to the selected vertex with coordinates (η, ϕ) reconstructed within a cone of size $\Delta R = 0.3$ around the photon candidate with coordinates $(\eta_\gamma, \phi_\gamma)$, where $\Delta R = \sqrt{(\eta - \eta_\gamma)^2 + (\phi - \phi_\gamma)^2}$. In cases where the PF candidates share part of their energy with the photon candidate, they are excluded from the isolation sums. Separate variables are constructed for PF charged hadrons (Iso_{Ch}) and PF photons (Iso_γ). Prompt photons are well isolated for both variables. These PF candidates are calculated with respect to the diphoton vertex chosen using the $H \rightarrow \gamma\gamma$ BDT. The Iso_γ variable was found to be sensitive to both pileup and photon p_T . The corrected photon isolation has been pileup subtracted with p_T dependence removed according to:

$$\overline{\text{Iso}}_\gamma = \alpha + \text{Iso}_\gamma - \rho A - \kappa p_T \quad (3.2)$$

where κ is a coefficient governing the p_T dependence; ρ is the average pileup p_T flow density per unit area in the event, calculated using the jet area method [112, 113]; A is an effective area, which is the geometrical area of the isolation cone times an η -dependent correction factor; and α is an empirical constant chosen to make the corrected isolation peak near zero. General pileup removal algorithms utilized by the CMS Collaboration are found in Ref. [114]. The values for α , A , and κ are listed in Table 3.1 based on the η coordinate of the supercluster (η_{SC}). When $\overline{\text{Iso}}_\gamma$ is considered in simulated data, an additional stochastic correction is applied to account for the mismodeling of the pileup dependence of Iso_γ in simulation. This correction assumes that the effect of pileup on Iso_γ is to add a discrete number of photons in its isolation cone.

Electrons and photons leave similar energy deposits in the ECAL, so an electron veto is

Table 3.1: Parameters used in the definition of $\overline{\text{Iso}}_\gamma$ based on the η coordinate of the super-cluster (η_{SC}).

Detector region	α (GeV)	A	κ
$ \eta_{\text{SC}} < 0.9$	0.99	0.15	0.0016
$0.9 < \eta_{\text{SC}} < 1.4442$	0.99	0.13	0.0016
$1.566 < \eta_{\text{SC}} < 2.0$	0.77	0.093	0.00075
$2.0 < \eta_{\text{SC}} < 2.2$	0.77	0.15	0.00075
$2.2 < \eta_{\text{SC}} < 2.5$	0.77	0.21	0.00075

applied to reject electrons in the selection. Electrons are vetoed based on hits in the silicon pixel and strip trackers with a further check to ensure photon candidates associated with electron tracks are incompatible with those resulting from photon conversions. This is known as the conversion-safe electron veto (CSEV).

The cut values applied to each of these identification variables are listed in Table 3.2, separately for photons in the EB and EEs. This collection of photon identification criteria is known as the high- p_{T} photon ID. This ID was developed and tuned specifically for this analysis, with significant contributions from the author of this dissertation.

Table 3.2: The high- p_{T} photon ID. The cut values are listed for the different identification variables. For $\sigma_{i\eta i\eta}$, the values in parenthesis correspond to saturated photons in the ECAL.

Photon category	Iso_{Ch} (GeV)	$\overline{\text{Iso}}_\gamma$ (GeV)	H/E	$\sigma_{i\eta i\eta}$	R_9	CSEV
EB	5	2.75	0.05	0.0105 (0.0112)	0.8	applied
EE	5	2.0	0.05	0.028 (0.030)	0.8	applied

The background which remains from misidentifications not suppressed by the high- p_{T} photon ID is discussed in Section 4.2.

3.2.4 Selection Efficiency

The efficiency of the high- p_{T} photon ID was measured using the tag-and-probe technique, similar to the procedure done in Ref. [115]. This technique is used in both data, using a sample of electron-triggered events, and in simulation, using a $Z \rightarrow e^+e^-$ sample. The tag

is an electron required to be in the event, with $p_T > 30$ GeV, passing the standard CMS electron ID. The probe is a photon with $p_T > 20$ GeV passing the high- p_T photon ID with an inverted CSEV. The invariant mass of the tag-and-probe system is required to be between 70-110 GeV. Figure 3.4 shows the photon selection in the EB and EE regions, including data-over-simulation scale factors. The overall efficiency is about 90 (87)% for single photons in the EB (EE) region. The scale factors are compatible with unity, but are still applied to the simulation described in Chapters 4 and 5.

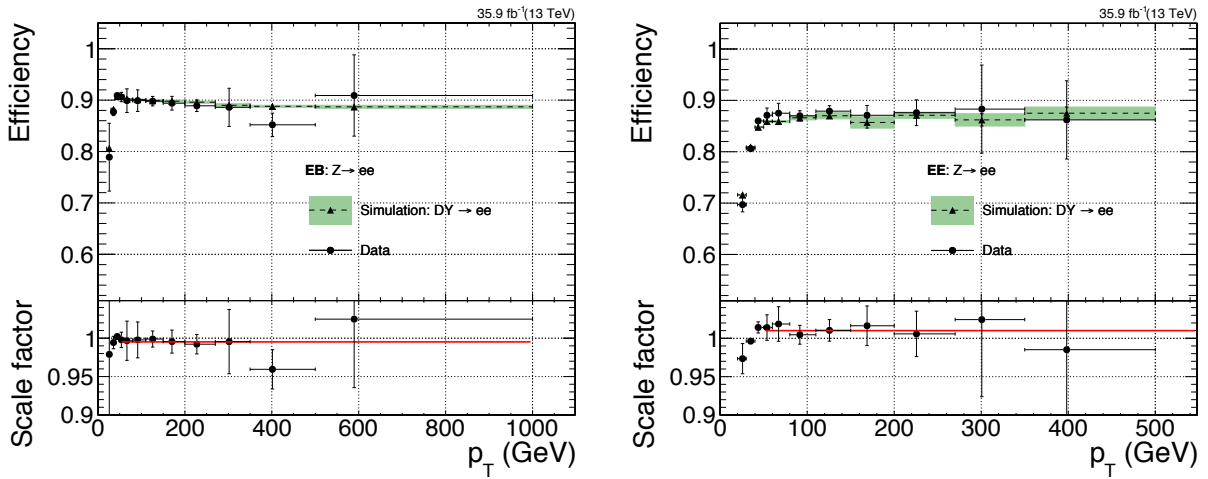


Figure 3.4: Photon selection efficiency (top) in the EB (left) and EE (right) regions and the associated data-over-simulation scale factors (bottom). The statistical and systematic errors are shown.

The result of considering the event selection efficiency (ε) and detector acceptance (A) together is shown in Fig. 3.5. This was measured using simulation, similar to the samples described in Chapter 5, using events of known $m_{\gamma\gamma}$ and determining which accepted events pass the high- p_T ID requirements. The combination of each across both analysis categories is approximately 60%.

3.2.5 Datasets

The datasets used in this analysis were recorded in 2016 and correspond to an integrated luminosity of 35.9 fb⁻¹. They were produced using LHC proton-proton collisions at $\sqrt{s} =$

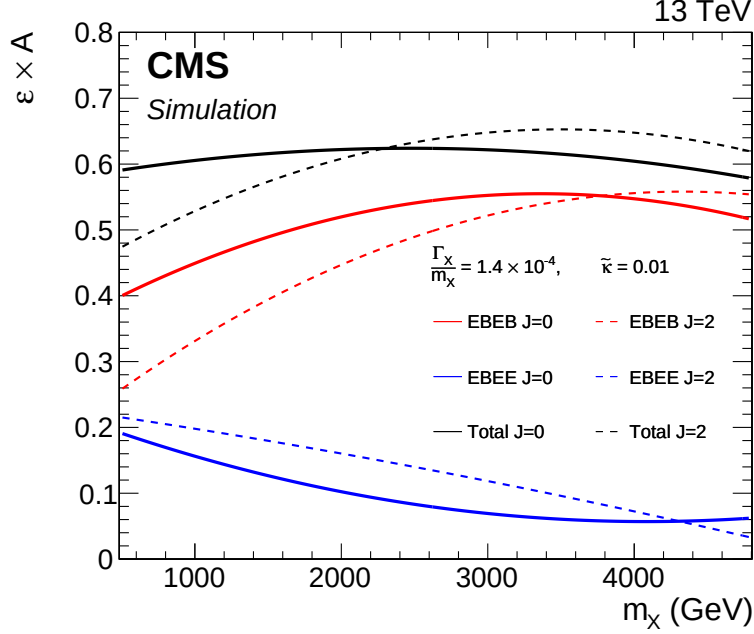


Figure 3.5: The product of the event selection efficiency (ε) and the detector acceptance (A) is shown as a function of signal resonance mass $m_X = m_{\gamma\gamma}$ for the $\Gamma_X/m_X = 1.4 \times 10^{-4}$ signal width hypothesis. The total (black), EBEB (red), and EBEE (blue) curves are shown for the spin (J) hypotheses $J = 0$ (solid line) and $J = 2$ (dashed line).

13 TeV and recorded by the CMS detector. Each dataset is listed in Table 3.3 according to their internal CMS path and corresponding integrated luminosity. These DoubleEG triggered primary datasets contain at least two electron or photon candidates.

Table 3.3: The analysis datasets, each listed with their CMS path name and corresponding integrated luminosity.

Dataset path	Int. lumi. (fb^{-1})
/DoubleEG/Run2016B-03Feb2017_ver2-v2	5.788
/DoubleEG/Run2016C-03Feb2017-v1/MINIAOD	2.573
/DoubleEG/Run2016D-03Feb2017-v1/MINIAOD	4.248
/DoubleEG/Run2016E-03Feb2017-v1/MINIAOD	4.009
/DoubleEG/Run2016F-03Feb2017-v1/MINIAOD	3.102
/DoubleEG/Run2016G-03Feb2017-v1/MINIAOD	7.540
/DoubleEG/Run2016H-03Feb2017_ver2-v1/MINIAOD	8.391
/DoubleEG/Run2016H-03Feb2017_ver3-v1/MINIAOD	0.215

Chapter 4

Background Determination

This search is for signatures of large extra-dimensional signals, as described by the ADD model, producing a nonresonant excess of high-mass diphoton events above the SM diphoton background. A defining feature of this search is the approach used to determine its background—a technique yielding a full background prediction. The background arises from two components: the real background, which comes from prompt SM diphoton production; and the fake background, which occurs when jets are misidentified as photons in the detector. The real background is dominant. This irreducible background is calculated using a next-to-next-to-leading order Monte Carlo calculation. The subdominant fake background is reducible and is estimated from control samples in data.

4.1 Background from SM Diphoton Production

This analysis is searching for high-mass photon pairs in the region $m_{\gamma\gamma} > 500 \text{ GeV}$. At the LHC, the dominant background in this search region arises from prompt SM diphoton production occurring through quark annihilation (the Born process) and gluon fusion (known as the “box” process). The Born process dominates over the box process. The leading order (LO) Feynman diagrams for these processes are shown in Fig. 4.1. A qg -initiated process is also possible, as illustrated in Fig. 4.4, but is formally higher order than the Born process and is suppressed compared to the box process due to the large gluon parton distribution

function (PDF), as discussed in the next section. The $\gamma\gamma$ final state of these processes is the same as the signal and, therefore, this real component of the background is irreducible. A next-to-next-to-leading order (NNLO) Monte Carlo calculation is used to determine the real background.

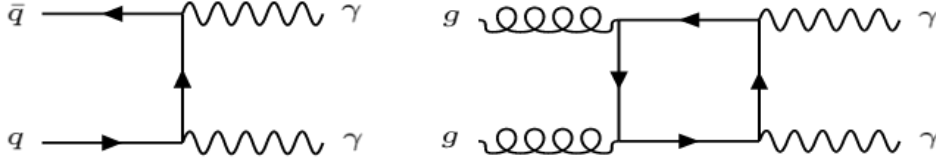


Figure 4.1: Representative Feynman diagrams for the LO SM diphoton Born (left) and box (right) processes.

4.1.1 Monte Carlo Simulation

Monte Carlo (MC) event generation in high-energy physics (a detailed review can be found in Section 4.1 of Ref. [16]) uses a divide and conquer philosophy. The phase space is typically split into distinct domains involving a matrix element (ME) and a parton shower (PS) calculation. The ME corresponds to the calculation of the hard interaction for specific particle processes. Cross section information is provided at this stage. The evolution of these parton-level processes through radiation and fragmentation is then handled in the PS calculation. A full event history of each particle is obtained by merging the two spaces. As quarks and gluons emerge from the collisions, they radiate gluons, which split into quark-antiquark pairs. This splitting process is known as fragmentation and is dominated by the emission of soft and collinear radiation until the scale Λ_{QCD} is reached and the partons become confined into hadrons, a process known as hadronization. Hadronization is a non-perturbation process, described by different phenomenological models. For the real background calculation, we use SHERPA v.2.1.1 [116], a full Monte Carlo (MC) event generator capable of performing these steps. SHERPA incorporates the cluster model [117, 118] to describe hadronization. To characterize the LHC proton-proton collisions, SHERPA is embedded with the CT10 set of

PDFs [119, 120].

The PDF allows us to assign a fraction x , called the the Bjorken x variable, of the total proton momentum to the colliding partons within each proton. The PDF $f_{a/A}(x_a, \mu_f^2)$ describes the probability that a parton a within some hadron A carries the fraction x_a of the total hadron momentum, where μ_f is the factorization scale, which, roughly, separates the long- and short-distance processes. Recall, that at short distances, or equivalently high energies, QCD is perturbative, allowing these processes to be calculated, while at long distances the regime is non-perturbative. The factorization scale is usually set at the energy scale of the hard scattering interaction, $\mu_f \sim Q^2$, where Q^2 is the momentum transfer in the collision. The interaction cross section is found by summing over the partons and integrating the PDFs according to:

$$\sigma_{AB \rightarrow X} = \sum_{a,b} \int dx_a dx_b f_{a/A}(x_a, \mu_f^2) f_{b/B}(x_b, \mu_f^2) \hat{\sigma}_{ab \rightarrow X} \quad (4.1)$$

for partons a and b , within hadrons A and B , carrying momentum fractions x_a and x_b , respectively, where $\hat{\sigma}_{ab \rightarrow X}$ is the parton hard scattering cross section. See Fig. 4.2 for a diagram representing this hard scattering process. In addition to μ_f , this calculation also requires the choice of the renormalization scale μ_r , the scale for the QCD running coupling. Some example PDFs from the CT10 PDF set are shown in Fig. 4.3.

SHERPA can be configured to produced events from user specified processes generated from the proton collisions. Both the Born and box processes are included in the SHERPA ME calculation and treated separately. The box process is calculated at the LO 1-loop level using a dedicated loop-induced ME generator. Even though the box process is formally NNLO, its contribution to the cross section is comparable to the LO terms due to the large gluon PDFs. Up to three additional final state partons are included in the Born process. Both quarks and gluons are allowed in the initial state of the Born process yielding higher order terms involving the additional final state partons through $jj \rightarrow \gamma\gamma + nj$, for a jet j and

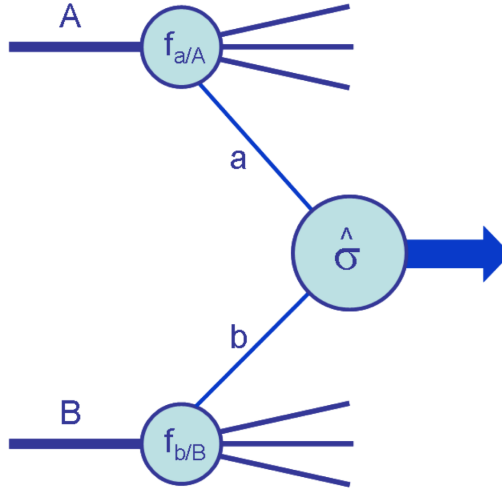


Figure 4.2: Diagrammatic representation of a proton-proton collision with their partons undergoing the hard scattering process [121]. The remainder of the two protons after collision is known as the beam remnant.

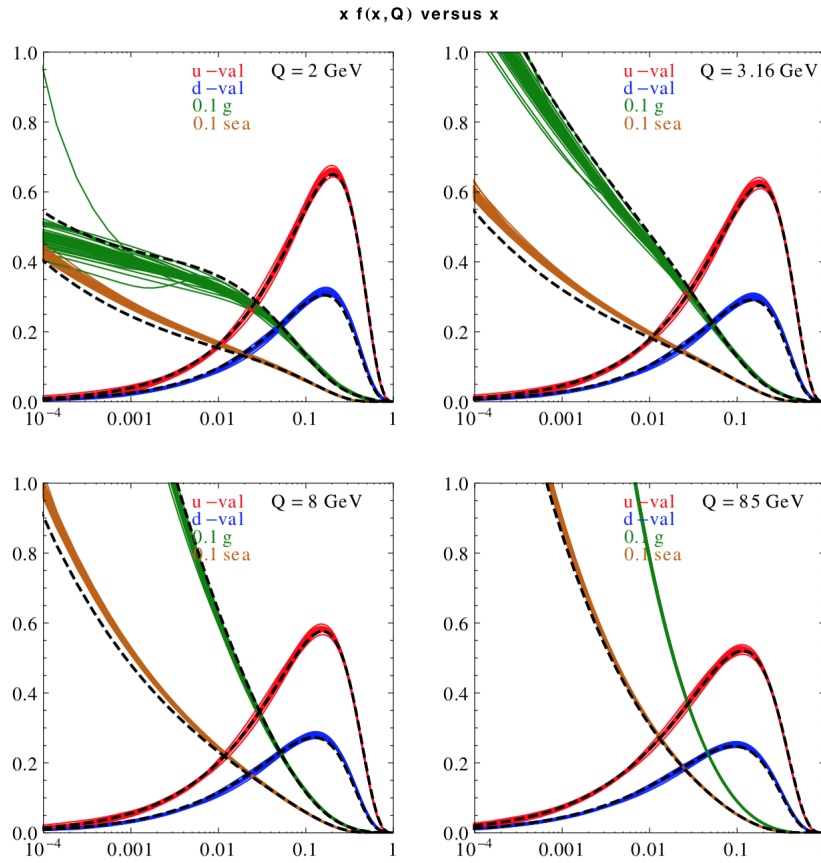


Figure 4.3: Example PDFs at energy scales $Q = 2, 3.16, 8,$ and 85 GeV, from the CT10 set of PDFs [120].

integer $n = 0, 1, 2$, or 3 . However, SHERPA only incorporates the real radiation component of these higher order processes and does not include virtual corrections. Fig. 4.4 shows some representative Feynman diagrams for NLO diphoton production in the Born process.

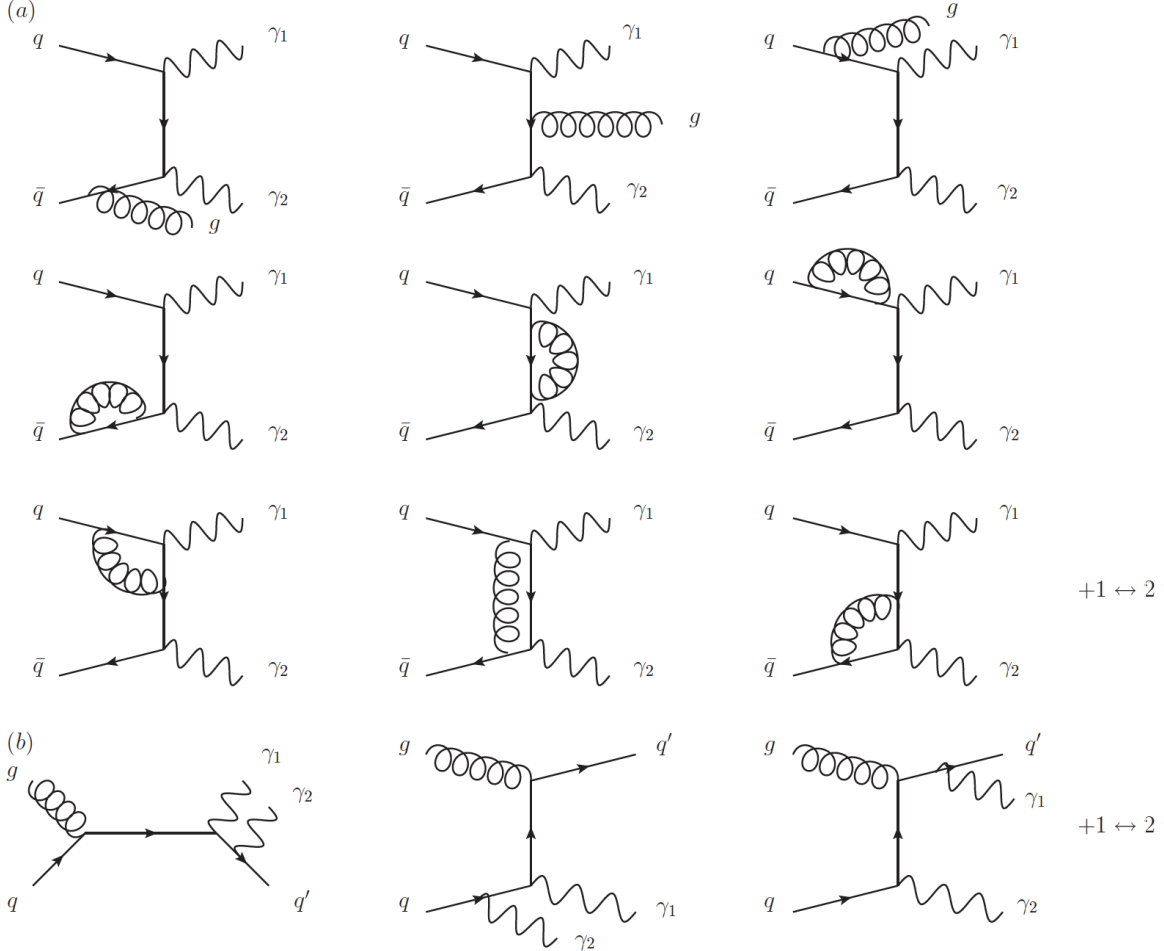


Figure 4.4: Representative Feynman diagrams for NLO diphoton production through the born process [122]. In (a) $q\bar{q}$ and (b) gq initiated processes are shown.

The inclusion of additional final state jets is important for sampling the diphoton phase space as demonstrated in Ref. [61]. Here it was determined using Run 1 data that the variable $\Delta\phi_{\gamma\gamma}$, which is the difference in the ϕ coordinate between the two photons forming the diphoton object in the event, is sensitive to the presence of these additional jets and using a LO simulation was insufficient at accurately calculating the process. For computational convenience, the probability for producing these higher order processes is altered

within SHERPA by applying weights to the additional final state partons. This is corrected for in the final analysis by extracting the weights and undoing them when incorporating this background calculation. In addition to these specific processes, SHERPA also interlaces photons into the QCD+QED PS [123]. This provides prompt photon contributions from the fragmentation of quarks and gluons and is taken into account when estimating the fake background. Representative Feynman diagrams for this fragmentation process are shown in Fig. 4.5.

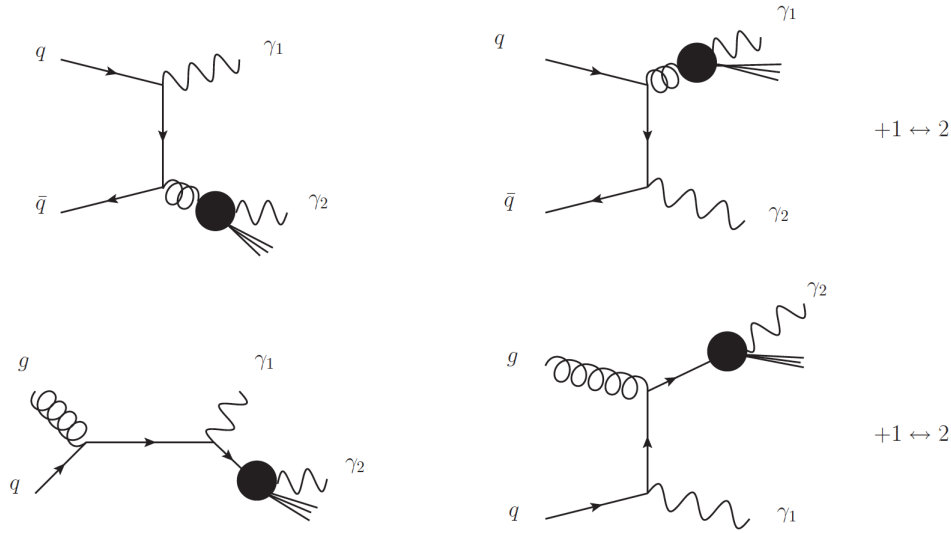


Figure 4.5: Representative Feynman diagrams for NLO diphoton production through the fragmentation of a quark or gluon [122].

Within SHERPA a kinematic selection of $p_T > 50 \text{ GeV}$ and $|\eta| < 2.8$ is imposed on the photons during event generation. To improve statistics the event generation is broken into eight samples corresponding to the diphoton invariant mass $m_{\gamma\gamma}$ bins of 60-200, 200-500, 500-1000, 1000-2000, 2000-4000, 4000-6000, 6000-8000, and 8000-13000 GeV. These parameters are specified and configured in SHERPA using run cards. An example run card used to generate one of these samples is shown in Appendix A.1. Among all samples, a total of 6,150,000 events were generated. Table 4.1 summarizes the samples and their corresponding cross sections. For each sample, the interactions of the generated particles are fully simulated through the CMS detector material using GEANT4 [124] and their response

to the detector’s electronic readout is then reconstructed. The simulation and reconstruction use CMS detector conditions that match those used during the 2016 data-taking period and include the effects of pileup from the proton-proton collisions.

Table 4.1: MC background samples used in this analysis with the number of events and corresponding cross sections. The full, internal CMS path name is specified by appending the dataset path with `/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/MINIAODSIM`.

Dataset path	Number of events	Cross section (pb)
/GGJets_M-60To200_Pt-50_13TeV-sherpa	1,000,000	5.785
/GGJets_M-200To500_Pt-50_13TeV-sherpa	2,500,000	2.244
/GGJets_M-500To1000_Pt-50_13TeV-sherpa	1,000,000	1.510
/GGJets_M-1000To2000_Pt-50_13TeV-sherpa	600,000	1.084e-02
/GGJets_M-2000To4000_Pt-50_13TeV-sherpa	300,000	3.690e-04
/GGJets_M-4000To6000_Pt-50_13TeV-sherpa	300,000	2.451e-06
/GGJets_M-6000To8000_Pt-50_13TeV-sherpa	250,000	1.753e-08
/GGJets_M-8000To13000_Pt-50_13TeV-sherpa	200,000	7.053e-11

These SHERPA samples contain individual events which have been fully simulated and reconstructed through the CMS detector, but higher order terms are absent. Unfortunately, at present, the MC programs capable of calculating the SM diphoton background at higher orders are not interfaced with GEANT4 and, hence, cannot provide simulated events. However, these higher order MC programs are capable of calculating (currently, at full NNLO) the diphoton cross section at generator-level in a specified phase space, yielding distributions. SHERPA also provides these same distributions at generator-level, allowing for a comparison. The ratio of these generator-level background predictions is called the K factor. We will follow this general approach to produce a K factor which will be subsequently applied to the simulation-level SHERPA samples. This procedure allows us to reweight the simulated events, providing a full NNLO diphoton background prediction, accounting for the missing higher order terms in the original samples.

4.1.2 The K Factor

The MC program MCFM 8.0 [125] is capable of calculating a full NNLO diphoton cross section at generator-level and will be used to form the K factor. MCFM includes the 1-loop virtual corrections to the Born process, as shown in Fig. 4.4, as well as for the box process. In addition, MCFM includes the NNLO diagrams for the Born process and NLO diagrams for the box process in its calculation, as depicted in Figures 4.6 and 4.7, respectively. MCFM allows light fermions in the box process to circulate in the loop, but further considers the finite mass of the t quark by also allowing it to circulate. The effects of this become important at high $m_{\gamma\gamma}$ [125].

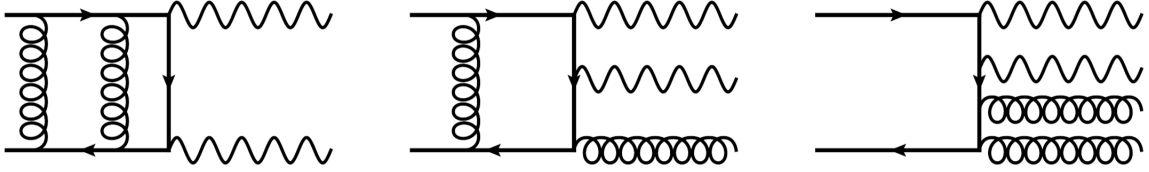


Figure 4.6: Representative Feynman diagrams for the diphoton Born process at NNLO, considered by MCFM [125].

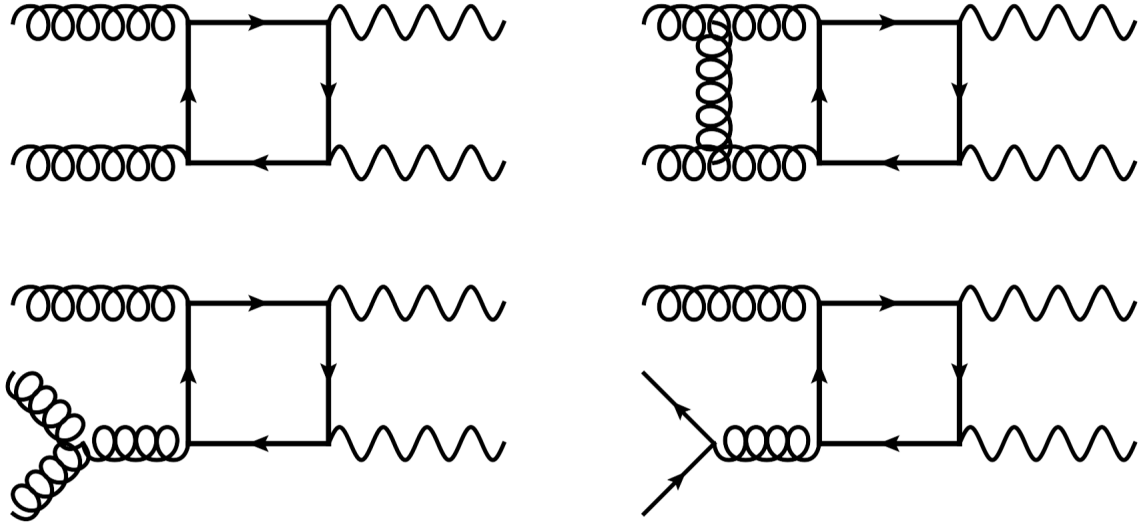


Figure 4.7: Representative Feynman diagrams for the diphoton box process at LO and NLO, considered by MCFM [125].

For the calculation, we will interface MCFM with the CT10 set of PDFs, as was done for the SHERPA calculation. MCFM can only provide parton-level cross section information for different diphoton kinematic variables, instead of individual, simulation-level events, since it is not interfaced with GEANT4. The NNLO contributions to the SM diphoton production processes are calculated as functions of $m_{\gamma\gamma}$. Kinematic cuts are placed on the generator-level photons during the parton-level calculation. These are the same kinematic cuts used offline in the analysis, requiring photons to have $p_T > 75$ GeV and restricting them to reside within the EB or an EE using the appropriate values of photon η . Photon isolation is calculated using the method by Frixione [126], which imposes an isolation requirement that depends on the ΔR from the location of the photon. The isolation at all distances must be within 5 GeV. Diphoton objects are then selected and categorized into the EBEB and EBEE regions. The four $m_{\gamma\gamma}$ bins of 500-750, 750-1000, 1000-1500, and 1500-4000 GeV are used during the calculation to improve the statistics of the overall calculation and reduce any potential statistical bias when a large range of $m_{\gamma\gamma}$ is separately considered. Finally, we reject events with very small $\Delta R_{\gamma\gamma}$ (< 0.45) to avoid an infrared divergence from collinear photon emissions. SHERPA is configured in an analogous way, providing similar diphoton cross section distributions as a function of $m_{\gamma\gamma}$, allowing for the formation of the K factor.

The K factor is defined as the ratio of the MCFM NNLO generator-level prediction to the SHERPA generator-level prediction, both produced using equivalent settings. Fig. 4.8 (middle) shows the MCFM LO (black), NLO (blue), and NNLO (red) calculations as a function of $m_{\gamma\gamma}$ compared against the SHERPA generator-level prediction (magenta). The ratio plots show the K factor distributions for each order with the final NNLO K factor shown in red. The K factor ranges from 1.4 to 2.2 (2.5) in the EBEB (EBEE) region for $m_{\gamma\gamma} < 3$ TeV.

Fig. 4.8 also shows the effects of simultaneously varying the renormalization μ_r and factorization μ_f scales by a factor of two up (upper) and down (lower) around the nominal prediction (center). The standard convention is to set these scales equal to each other and simultaneously vary them according to:

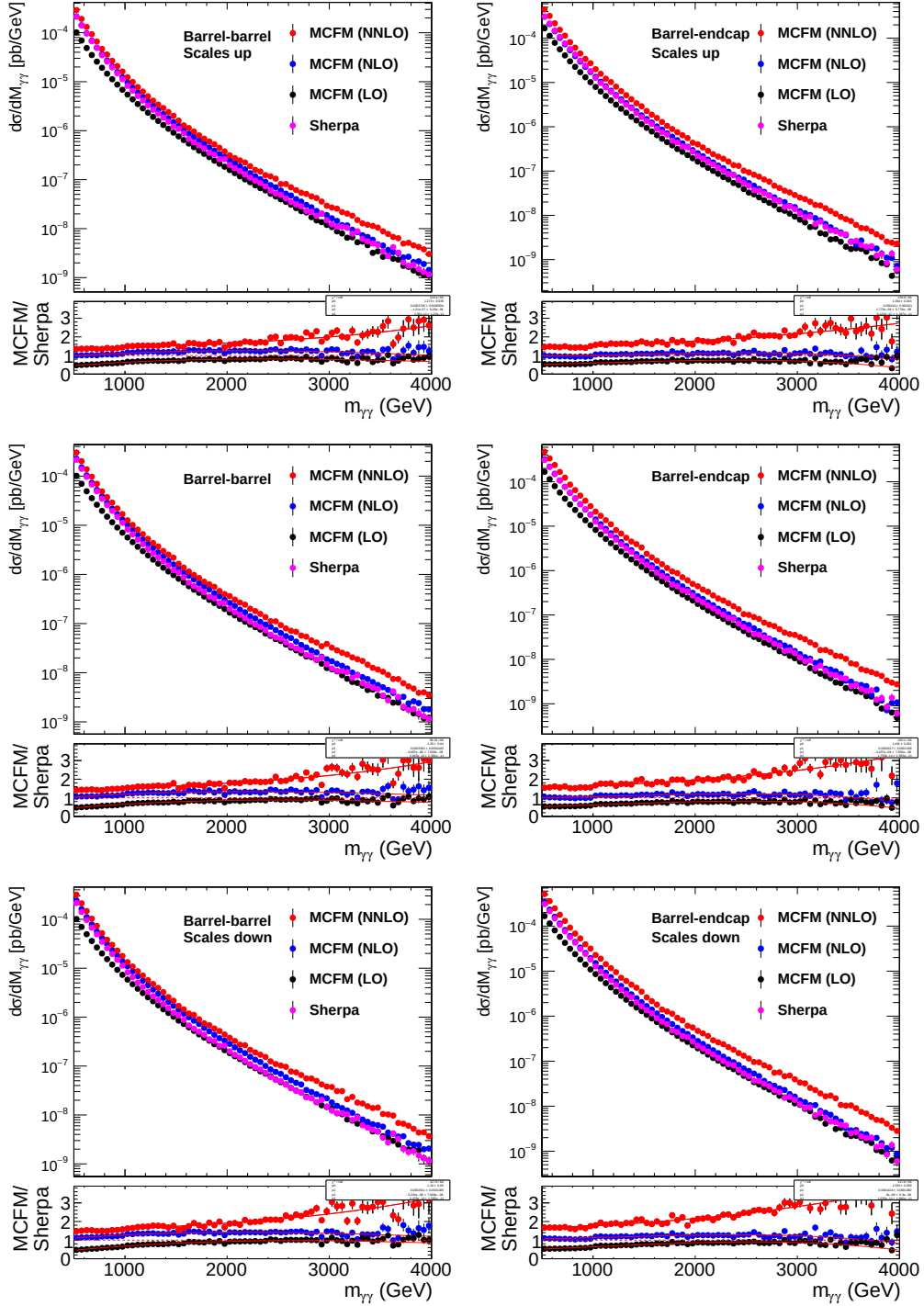


Figure 4.8: The NNLO K factor as a function of $m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories. The renormalization and factorization scales have been set to $m_{\gamma\gamma}$ (middle) and simultaneously varied by factors of 2.0 (upper) and 0.5 (lower).

- $\mu_r = \mu_f = 2.0m_{\gamma\gamma}$;
- $\mu_r = \mu_f = 1.0m_{\gamma\gamma}$; and
- $\mu_r = \mu_f = 0.5m_{\gamma\gamma}$.

These separate variations in the EBEB region are summarized for each order in Fig. 4.9, which shows an envelope around the nominal K factor prediction with these scales set to $1.0m_{\gamma\gamma}$. This effect is treated as a systematic uncertainty on the analysis and is discussed in Chapter 6. Note, by default in MCFM, the fragmentation scale is set to the renormalization scale. The fragmentation scale is the scale of the fragmentation functions, which describe the probability for a parton to fragment into a hadron. The variation of the renormalization scale indirectly varies this scale as well; however, its contribution is negligible when using Frixione isolation.

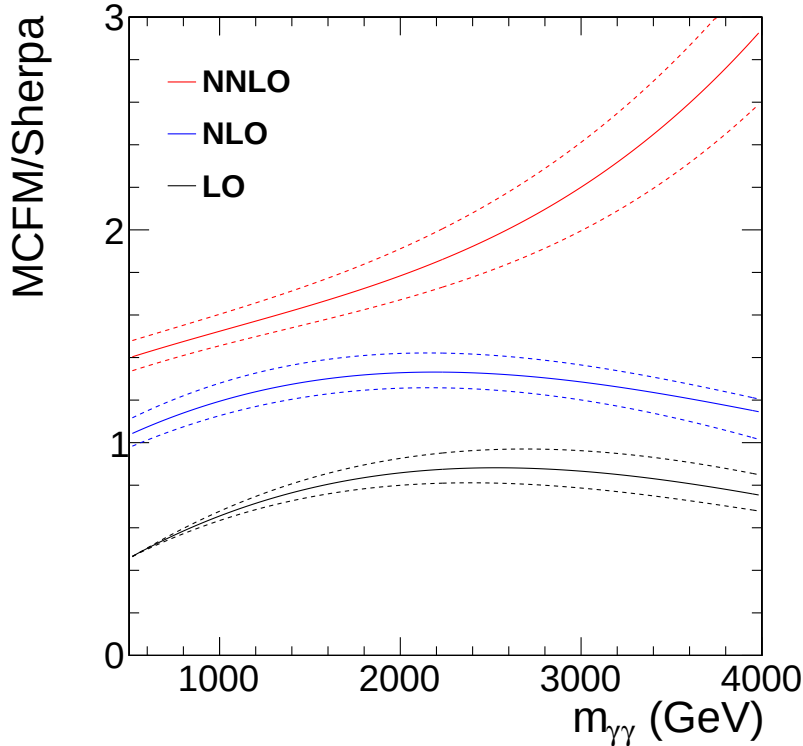


Figure 4.9: Ratio of the diphoton cross section calculated by MCFM to that obtained from SHERPA. The curves indicate different orders of the MCFM calculation, and the dashed lines show the effect of the scale variations.

The final prediction for the real SM diphoton background is determined by applying the K factor to the fully simulated SHERPA samples to reweight its events. The K factor distributions are parameterized as a function of $m_{\gamma\gamma}$, separately for the EBEB and EBEE categories, using a third degree polynomial function f of the form

$$f(m_{\gamma\gamma}) = p_0 + p_1 m_{\gamma\gamma} + p_2 m_{\gamma\gamma}^2 + p_3 m_{\gamma\gamma}^3 \quad (4.2)$$

where p_i , for $i = 0, 1, 2$, and 3 , are free parameters determined by a best fit to each distribution. The values of p_i are shown in Table 4.2, separately for each region. These functions are applied to the simulation-level SHERPA events, yielding a full NNLO prediction of the real SM diphoton background.

Table 4.2: Parameters of the K factor function for the EBEB and EBEE categories.

	EBEB	EBEE
p_0	1.25 ± 0.04	1.43 ± 0.05
p_1	$3.36\text{e-}04 \pm 1.05\text{e-}04$	$2.02\text{e-}04 \pm 1.19\text{e-}04$
p_2	$-9.09\text{e-}08 \pm 7.01\text{e-}08$	$-4.44\text{e-}09 \pm 7.99\text{e-}08$
p_3	$2.82\text{e-}11 \pm 1.38\text{e-}11$	$1.76\text{e-}11 \pm 1.59\text{e-}11$

4.2 Background from Diphoton Misidentifications

A subdominant contribution to the total background in this search occurs when particles are misidentified as photons in the CMS detector. These particles fake a photon signature and are misreconstructed as photons. The background arising from this is referred to as the fake background. In this analysis, we consider fake photons arising from the misreconstruction of both electrons and jets. The rate for electrons faking photons is almost negligible and is readily measured from an MC calculation. To measure the rate for jets faking photons, a sophisticated technique using control samples in data is utilized.

The subdominant contribution to the fake background occurs when one or two electrons

are misidentified as a photon in the detector. This mainly occurs when the electron track is misreconstructed or an electron converts late in the tracker leaving behind missing hits. This is measured using the MC simulation of Drell-Yan dielectron, $W\gamma$, $Z\gamma$, W , and $t\bar{t}\gamma$ events. The measurement is done by inverting the electron veto in the high- p_T photon ID and counting how many events pass this modified ID. The contribution of these fakes to the total background is calculated to be less than 1%.

The dominant contribution to the fake background occurs when a jet with large electromagnetic activity is misidentified as a photon in the CMS detector, allowing it to pass the high- p_T photon ID criteria and, hence, fake a photon signature. The typical example of this occurs when a jet fragments to a boosted neutral meson, such as a π^0 or η , which then decays at high momentum to two photons. Both of these photons can then deposit their energy in the ECAL in such a way that their deposits overlap and resemble the signature of a single photon candidate. The fact that two photon energy deposits can become indistinguishable from one is due to the finite granularity and energy resolution of the ECAL. In a single diphoton event, this can occur from either one or two jets resulting in a photon+jet (γj) or dijet (jj) background event, respectively. This type of jet fragmentation process is rare, so instead of relying on the MC simulation to accurately calculate it, a data-driven approach is preferred, which uses control samples in data to estimate to estimate the fake background contribution. The approach utilized is referred to as the fake rate method.

4.2.1 Fake Rate Method

The fake rate method involves determining a fake rate f from a jet-enriched data control sample and applying it to the analysis dataset containing diphotons passing the final analysis selection. The fake rate is defined as the ratio of the number of jets passing two categories of jet fragmentation processes. The first category, the numerator of the ratio, is the number of jets passing the high- p_T photon ID. The second category, the denominator of the ratio, is the number of jets passing a looser, photon-like ID. The denominator ID is chosen such that

the two categories are orthogonal. In equation form, the fake rate is

$$f = \frac{(\text{numerator})}{(\text{denominator})} = \frac{\# \{\text{jets passing the photon ID}\}}{\# \{\text{jets passing a photon-like ID}\}} \quad (4.3)$$

where $\#$ is used to denote the number of objects in that category. This fake rate f is simply the ratio of its defined categories and is not the probability that a generic jet will fake a photon.

A complication arises when determining the numerator. Although the control data sample is jet enriched, there will inevitably be a contribution to the numerator from genuine photons in it passing the photon ID, which must be removed. This is done through a fitting procedure where template models of real and fake photons are fit to the data in the control sample. These templates are constructed using a photon ID variable that is sensitive to the differences between real and fake photons, such as a shower shape or isolation variable. After the fit is performed, the contribution to the numerator from real photons will be subtracted from the total. In general, the choice of template variable will evolve with photon p_T and η . Therefore, the fake rate f will be measured as a function of photon p_T and categorized in the EB and EE regions.

After the final fake rate function $f(p_T)$ is determined separately in the EB and EE categories (from the data control sample), it will then be applied to the final, analysis-level diphoton dataset. This is done by selecting objects in the diphoton dataset that pass the denominator, photon-like ID and reweighting them using $f(p_T)$, separately in the EBEB and EBEE categories. The assumption in doing this final step is that taking denominator objects and reweighting them by the fake rate does indeed produce an accurate representation of the kinematics of known fake events. Physically, we expect this to be true because jet fragmentation is largely independent of the rest of the event. Thus, events where a jet fragments to look like a denominator, photon-like jet should be representative of other events where the jet fragmented differently and ended up faking the photon ID, since we want to

measure the number of jets passing the photon ID by using jets that pass the denominator ID. Verification of the ability of the fake rate method to achieve this is demonstrated in simulation and is described in Section 4.2.5.

To validate the fake rate method, a closure test is performed on a sample of simulated data produced from an MC calculation. The idea is to apply the fake rate method to the simulated sample in the identical way as it is done in the jet-enriched control dataset. This yields a fake prediction in the simulated sample. Now, because we are using simulation, we have access to the MC truth information for the particles, i.e., we know the properties and full event history of all the particles in the simulated sample. In particular, we know the identities of all of the particles in it passing the high- p_T photon ID. Therefore, using the truth information, we can tag objects passing the photon ID as having arisen from jets or from real, prompt photons. Finally, we can compare these fake yields to the fake prediction in the simulated sample. In addition, we also perform further studies to validate the method, as will be explored in Section 4.2.5.

4.2.2 Fake Rate Calculation

The data control sample used for the photon fake rate calculation comes from a combination of jet- and muon-triggered data samples. The list of samples using the full, internal CMS dataset path names are tabulated in Table 4.3. This combination of datasets has been constructed to help eliminate any potential bias arising from the differences in the hadronization between quarks and gluons. In general, gluon jets contain a higher number of particles, as discussed in Section 3.1.4, and are therefore less likely to pass the isolation variables in the photon ID, so they should have a lower probability of faking a photon than a quark jet. It is difficult to distinguish between quark and gluon jets in data, but these datasets result from different processes and will help better sample jet flavor in the control sample.

As the proton beams circulate, they may interact with the LHC collimators or residual gas in the beam pipe due to the non-perfect vacuum, creating “beam halo.” When the

Table 4.3: Control datasets used for the photon fake rate calculation.

Dataset path
/JetHT/Run2016B-23Sep2016-v3/MINIAOD
/JetHT/Run2016C-23Sep2016-v1/MINIAOD
/JetHT/Run2016D-23Sep2016-v1/MINIAOD
/JetHT/Run2016E-23Sep2016-v1/MINIAOD
/JetHT/Run2016F-23Sep2016-v1/MINIAOD
/JetHT/Run2016G-23Sep2016-v1/MINIAOD
/JetHT/Run2016H-PromptReco-v2/MINIAOD
/JetHT/Run2016H-PromptReco-v3/MINIAOD
/DoubleMuon/Run2016B-23Sep2016-v3/MINIAOD
/DoubleMuon/Run2016C-23Sep2016-v1/MINIAOD
/DoubleMuon/Run2016D-23Sep2016-v1/MINIAOD
/DoubleMuon/Run2016E-23Sep2016-v1/MINIAOD
/DoubleMuon/Run2016F-23Sep2016-v1/MINIAOD
/DoubleMuon/Run2016G-23Sep2016-v1/MINIAOD
/DoubleMuon/Run2016H-PromptReco-v2/MINIAOD
/DoubleMuon/Run2016H-PromptReco-v3/MINIAOD

protons interact with the particles associated with these sources, jets will be formed and can subsequently fragment to other high-energy particles, which are not produced from the collision point of the bunch crossing. This will produce charged pions ($\pi^\pm$), which will travel an average distance of about 10 m before decaying, primarily to muons ($\mu^\pm$). These muons will travel parallel to the beam pipe and can traverse the whole detector, producing high-energy bremsstrahlung photons in the ECAL. Beam halo could fake a photon signature and must be eliminated in the fake rate measurement. This source can be identified and rejected by using the photon shower shape variables $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$. Beam halo will be well contained in the ϕ direction but spread out in the η direction as it interacts with the CMS detector. This is in contrast to prompt photons, which are well contained in the η direction. Fig. 4.10 shows the correlation between $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ for all objects in the jet-triggered dataset passing the high- p_T photon ID (with the $\sigma_{i\eta i\eta}$ cut relaxed) in the range $450 < p_T < 500$ GeV. In this high-energy range, cutting at $\sigma_{i\phi i\phi} > 0.009$ will safely reject the beam halo.

The denominator objects are chosen to pass a less strict version of the high- p_T photon ID. These photon-like jets should have high electromagnetic energy content but are required to

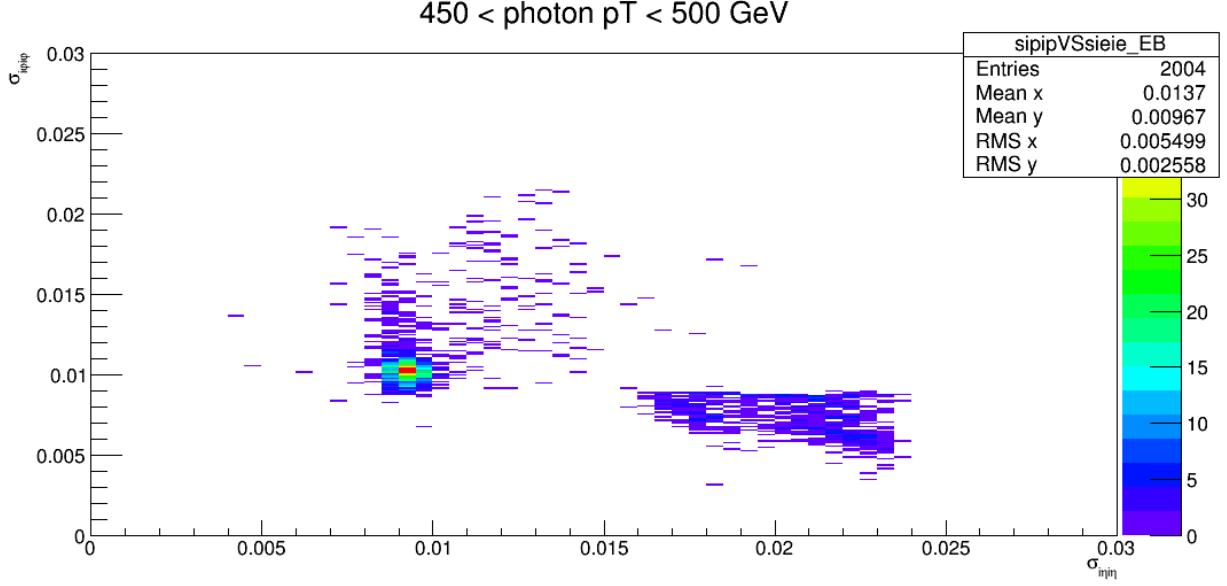


Figure 4.10: The correlation between $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ for all objects in the jet-triggered control dataset (used for the fake rate calculation) passing the high- p_T photon ID (with the $\sigma_{i\eta i\eta}$ cut relaxed) in the range $450 < p_T < 500$ GeV. Beam halo is observed for $\sigma_{i\phi i\phi} < 0.009$.

fail at least one of the photon ID variables, ensuring the denominator category is orthogonal to the numerator category, as well as eliminating any contamination from real photons. We want these objects to be representative of the kinematics of the reweighted events in the diphoton data sample. They should be reasonably isolated since not all of the energy of the particles in a cone of ΔR around the object is added to it, while, on the other hand, we need to ensure there is enough statistics after loosening the isolation criteria. Different cuts are imposed in the EB and EE regions to optimize the selection in each category. Denominator objects in the EB are required to fail at least one of the variables H/E , $\sigma_{i\eta i\eta}$, Iso_{Ch} , or $\overline{\text{Iso}}_\gamma$. In addition, they are required to pass both $\text{Iso}_{\text{Ch}} < 0.2 p_T$ and $\overline{\text{Iso}}_\gamma < 0.2 p_T$. Because of pileup mismodeling in the Iso_γ variable during the 2016 data-taking period, denominator objects in the EE category are required to pass the photon ID cut of $\overline{\text{Iso}}_\gamma < 2.0$ GeV, instead of the looser selection of $\overline{\text{Iso}}_\gamma < 0.2 p_T$. In the EE regions, the denominator objects are also required to pass $\text{Iso}_{\text{Ch}} < 0.2 p_T$ and fail at least one of the variables H/E , $\sigma_{i\eta i\eta}$, or Iso_{Ch} . These cuts will be summarized in Table 4.5 below after some further discussion.

An additional cut for denominator objects on the ratio the hadronic over electromagnetic energies using Hadronic/E is imposed in both the EB and EE regions. As discussed in Section 3.2.1, the variable H/E using a tower-based isolation to sum the hadronic energy deposits while Hadronic/E uses a cone-based calculation. The trigger selection used in the analysis (HLT_DoublePhoton60) requires Hadronic/E at the HLT level and is not fully efficient for denominator objects, even for those passing $H/E < 0.05$, while it is for objects passing the high- p_T photon ID. This is remedied by applying an additional cut of $\text{Hadronic}/E < 0.1$ to the denominator objects.

The numerator objects result from the selection of objects in the jet-enriched dataset passing the high- p_T photon ID and adjusted so that any contamination from real photons is removed through a template fitting procedure. The template variable used is the shower shape variable $\sigma_{i\eta i\eta}$. The data to be fit to are the numerator candidates in the jet-enriched dataset consisting of all objects passing the high- p_T photon ID with $\sigma_{i\eta i\eta}$ relaxed, to be summarized in Table 4.5 below. Since $\sigma_{i\eta i\eta}$ evolves with photon p_T , the data are divided into the nine photon p_T bins of 50-70, 70-90, 90-110, 110-130, 130-150, 150-200, 200-250, 250-300, and 300-600 GeV. (The denominator objects are also split into these bins resulting in a fake rate for each bin.) The numerator candidates in each p_T bin are split into their real and fake components using templates derived from respective models of real and fake photons.

The fake templates are derived from the jet-enriched dataset. Since real photons are well isolated after pileup subtraction, in particular, the high- p_T photon ID requires $\text{Iso}_{\text{Ch}} < 5$ GeV, fake photons can be modeled by inverting this isolation requirement. Objects in the fake templates must pass the high- p_T photon ID with $\sigma_{i\eta i\eta}$ relaxed but are taken from the sideband of $5 < \text{Iso}_{\text{Ch}} < 10$ GeV. Notice, we are trying to describe fake photons with $\text{Iso}_{\text{Ch}} < 5$ GeV by modeling them from the selection using $5 < \text{Iso}_{\text{Ch}} < 10$ GeV. The evolution of the fake templates with p_T from the jet-triggered dataset is shown in Fig. 4.11. These templates are scaled to 1 fb^{-1} and shown separately in the EB and EE regions.

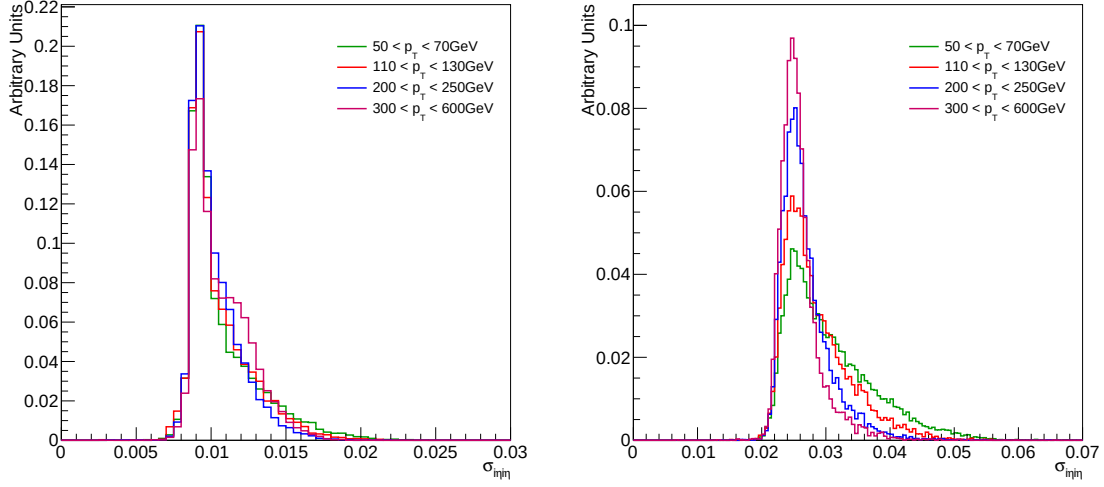


Figure 4.11: Fake templates scaled to 1 fb^{-1} in various p_T bins in the EB (left) and EE (right) regions.

The real templates are derived from a photon-rich simulated sample produced using an MC calculation. The $\gamma\gamma$ MC background samples, listed in Table 4.1, are used for this. In addition, the γj MC samples produced centrally within the CMS Collaboration are added to allow for a source of prompt photons coming from processes using the γj final state. The full list of samples is given in Table 4.4. Within the photon-rich MC sample, all objects passing the high- p_T photon ID with relaxed $\sigma_{i\eta i\eta}$ are selected, but only those coming from known prompt photons are allowed in the real templates. The definitions for both the real and fake templates are summarized in Table 4.5. It's possible to identify these real photons since the MC truth information is provided for all generated particles.

The procedure used to ensure only real photons are selected is to first find the best match in ΔR among all the final state generated particles to the reconstructed photon object passing the photon ID. Next, the best match is required to be a generated photon within $\Delta R < 0.1$. If the matched generated photon can be traced back to the hard interaction, then the reconstructed photon object is included in the real template. Fig. 4.12 shows the evolution of these real templates (each scaled to 1 fb^{-1}) with photon p_T in both the EB and EE categories.

Table 4.4: The MC samples and corresponding cross sections used for deriving the real templates in the fake rate calculation. The full, internal CMS path name is specified by appending the dataset path with /RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/MINIAODSIM.

Dataset path	Cross section (pb)
/GGJets_M-60To200_Pt-50_13TeV-sherpa	5.785
/GGJets_M-200To500_Pt-50_13TeV-sherpa	2.244
/GGJets_M-500To1000_Pt-50_13TeV-sherpa	1.510
/GGJets_M-1000To2000_Pt-50_13TeV-sherpa	1.084e-02
/GGJets_M-2000To4000_Pt-50_13TeV-sherpa	3.690e-04
/GGJets_M-4000To6000_Pt-50_13TeV-sherpa	2.451e-06
/GGJets_M-6000To8000_Pt-50_13TeV-sherpa	1.753e-08
/GGJets_M-8000To13000_Pt-50_13TeV-sherpa	7.053e-11
/GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.121e+04
/GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.863e+03
/GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.298e+03
/GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.816e+02
/GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.465e+01

Table 4.5: The cut-based definitions of the numerator and denominator objects for the fake rate calculation compared against the high- p_T photon ID. The numerator is determined after numerator candidates are fit with real and fake templates (definitions shown) using the template variable $\sigma_{i\eta i\eta}$. The cuts listed are for both the EB and EE regions unless separately specified.

Cut [EB {sat.} (EE {sat.})]	Photon ID	Num.	Real template	Fake template	Denom. (EB)	Denom. (EE)
Photon ID cuts						
H/E < 0.05	pass	pass	pass	pass		
$\sigma_{i\eta i\eta} < 0.0105 \{0.0112\} (0.028 \{0.030\})$	pass	template	template	template		
$\text{Iso}_{\text{Ch}} < 5 \text{ GeV}$	pass	pass	pass	not applied	fail at least one	fail at least one
$\text{Iso}_{\gamma} < 2.75 (2.00) \text{ GeV}$	pass	pass	pass	pass		pass
$R_9 > 0.8$	pass	pass	pass	pass	pass	pass
Electron veto						
CSEV	pass	pass	pass	pass	pass	pass
Additional fake rate cuts						
$\sigma_{i\phi i\phi} > 0.009$	not applied	pass	pass	pass	pass	pass
$5 < \text{Iso}_{\text{Ch}} < 10 \text{ GeV}$	not applied	not applied	not applied	pass	not applied	not applied
Hadronic/E < 0.1	not applied	not applied	not applied	not applied	pass	pass
$\text{Iso}_{\text{Ch}} < 0.2 p_T$	not applied	not applied	not applied	not applied	pass	pass
$\text{Iso}_{\gamma} < 0.2 p_T$	not applied	not applied	not applied	not applied	pass	not applied

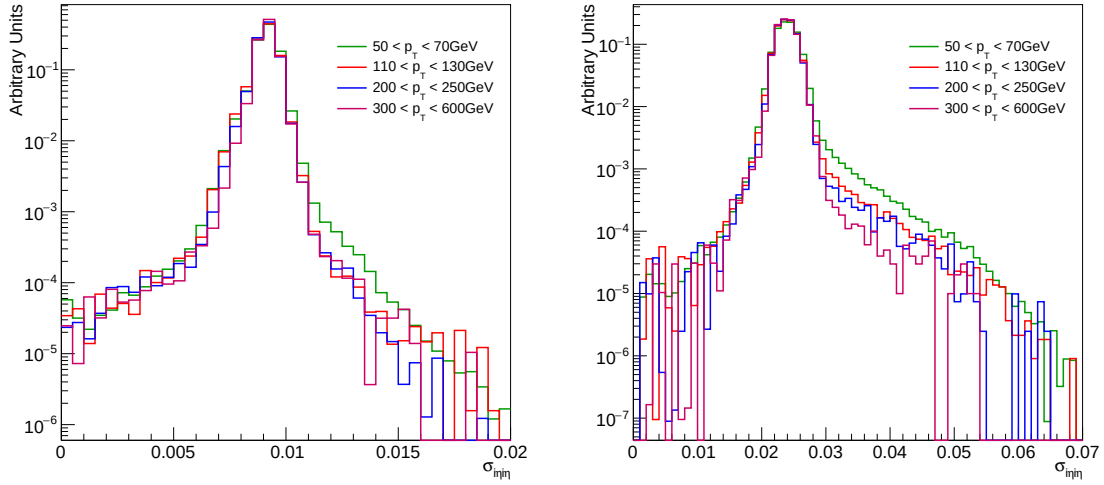


Figure 4.12: Real templates scaled to 1 fb^{-1} in various photon p_T bins in the EB (left) and EE (right) regions.

With the construction of the fake and real templates, they can both be fit to the numerator candidates in the control data sample. The shapes of the templates are fully determined from their construction, but their normalization is determined from the fit to the data. First the templates are normalized to 1 fb^{-1} , then a simultaneous fit using a maximum likelihood estimation is performed. A negative log-likelihood is used in a procedure similar to what is employed in Chapter 7. An example fit in the EB region for the 130-150 GeV p_T bin is shown in Fig. 4.13. We see that the tail region with $\sigma_{i\eta\eta} > 0.011$ is dominated by fakes, which largely fixes the overall fake normalization. The general trend showing that the real template dominates the overall normalization in the peak and the fake template in the tail holds true for the other p_T bins and also in the EE category. This final fit shows the contribution of fakes within the data. To extract the numerator, the fake template is integrated from 0 to $\sigma_{i\eta\eta}$ cut value, which differs between the EB and EE categories. This process is repeated for all p_T bins in both categories.

Dividing the numerator yields by the denominator gives the fake rates in the nine p_T bins for both the EB and EE categories, as shown in Fig. 4.14. These are represented by the black data points with horizontal error bars showing the bin widths. The statistical uncertainties

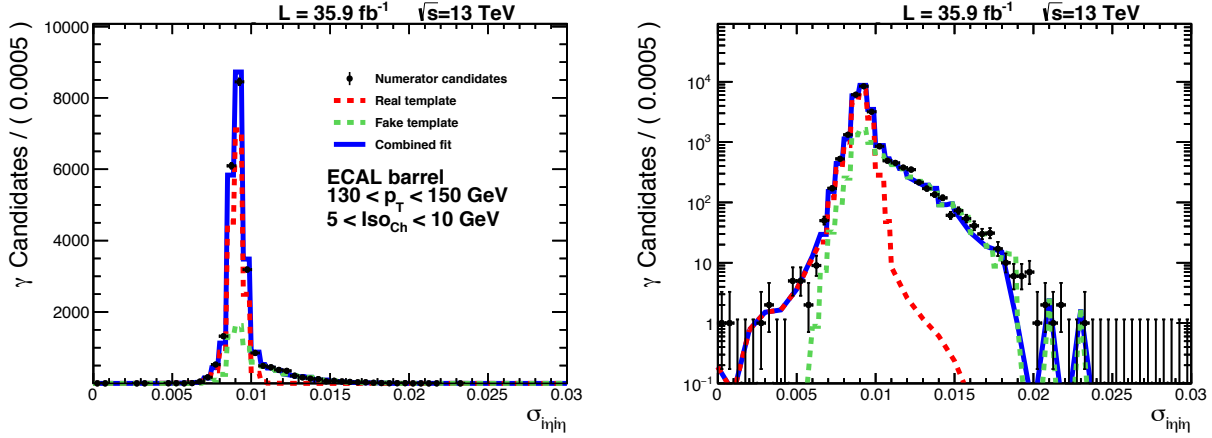


Figure 4.13: A template fit in the EB with $130 < p_T < 150$ GeV, shown in linear (left) and log (right) scale.

from the denominator distributions and the template fits are represented by the vertical, black error bars. The vertical, blue error bars represent the change in the fake rates under a variation of the Iso_{Ch} sideband, to be discussed in more detail below in Section 4.2.3. The larger fake rate in the EE category is caused by imposing the $\overline{\text{Iso}}_\gamma < 2.0$ GeV cut in the denominator selection, which is different for the EB category. The best fit to these fake rate points is shown in red, giving the final fake rate as a function of photon p_T .

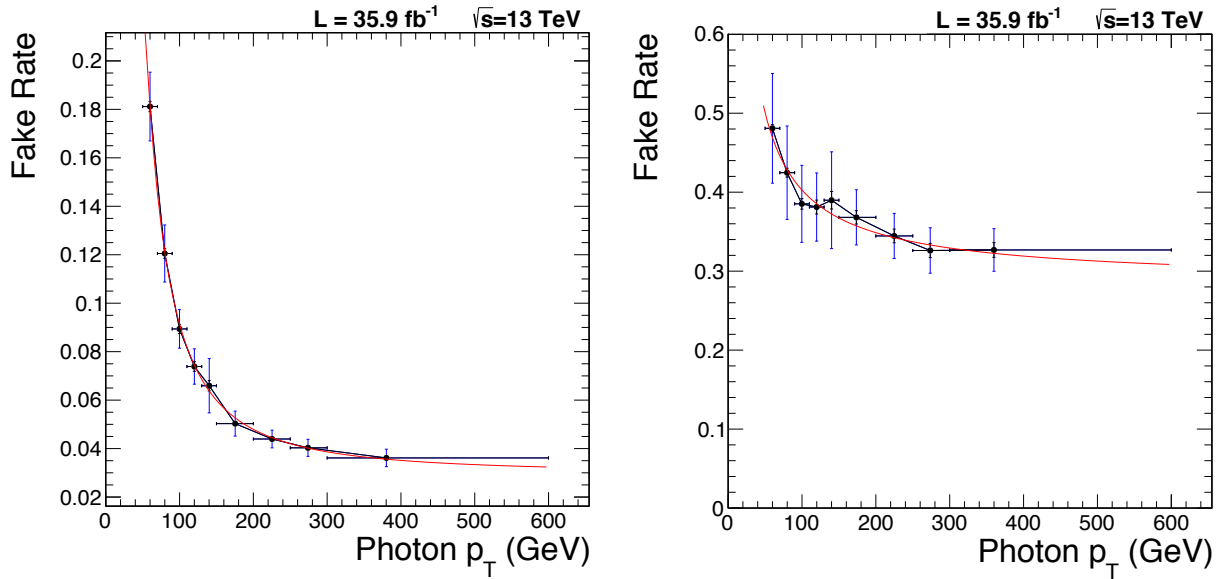


Figure 4.14: The fake rate in different photon p_T bins for the EB (left) and EE (right) categories. The fake rate function (red) is the best fit to the points.

The fake rate is parameterized by an inverse function with three free parameters p_0 , p_1 , and p_2 of the form

$$f(p_T) = p_0 + \frac{p_1}{p_T^{p_2}} \quad (4.4)$$

These free parameters are determined by the best fit to the fake rate points taking into account the black statistical error bars shown in Fig. 4.14. The fit parameters are tabulated in Table 4.6, separately for the EB and EE categories.

Table 4.6: Fit parameters for the fake rate function.

	EB	EE
p_0	0.030 ± 0.006	0.28 ± 0.05
p_1	205.9 ± 401.4	6.36 ± 9.77
p_2	1.76 ± 0.44	0.86 ± 0.41

The fake rate points shown in Fig. 4.14 are actually the result of deriving separate fake rates from the jet- and muon-triggered datasets and taking their average. To help reduced trigger bias in the jet-triggered dataset, the objects passing the high- p_T photon ID are required to be matched within $\Delta R < 0.6$ to the leading jet within the dataset. The separate fake rates from these two datasets are shown in Fig. 4.15. Their average is also shown, which is the same as the fake rate in Fig. 4.14.

4.2.3 Fake Rate Studies

The effect of jet flavor on the fake rate has been considered in the above calculation. Here we check how robust the fake rate is among other considerations. We check its variation with pileup, change of template variable, and change in sideband definition. These effects are covered with a systematic uncertainty on the overall analysis and will be discussed in Chapter 6.

The fake rate is sensitive to the amount of pileup in an event. One measure for this is the number of primary vertices (PVs) within an event. In addition to the selection criteria

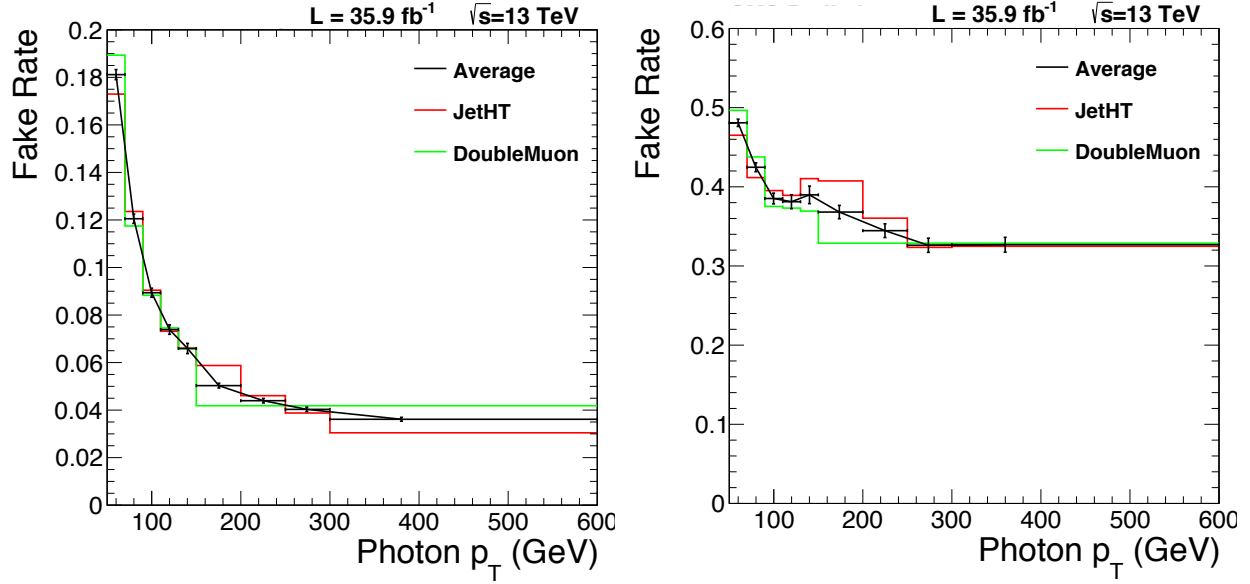


Figure 4.15: Separate fake rates for the jet- (red) and muon-triggered (green) datasets are shown in each photon p_T bin for the EB (left) and EE (right) categories. The average of the two (black) is the final fake rate.

imposed for primary vertex reconstruction, described in Section 3.1, we further require PVs to have

- $n_{\text{dof}} \geq 2$;
- $|z| < 24$ cm; and
- $d_0 \leq 2$ cm;

where n_{dof} is the number of degrees of freedom, z is the z -coordinate, and d_0 is the impact parameter of the vertex. The n_{dof} is a criteria used in the vertex fitting algorithm based on weights assigned to the reconstructed tracks, indicating track compatibility with the vertex. Fig. 4.16 shows the number of PVs for events in the jet-triggered dataset containing at least one photon passing the high- p_T photon ID. The distribution is shown both before and after these PV quality cuts are imposed.

Based on these distributions of PVs, the fake rate calculation was performed using the jet-triggered dataset in bins containing events with 0-13, 14-17, 18-20, 21-24, and 25 or more

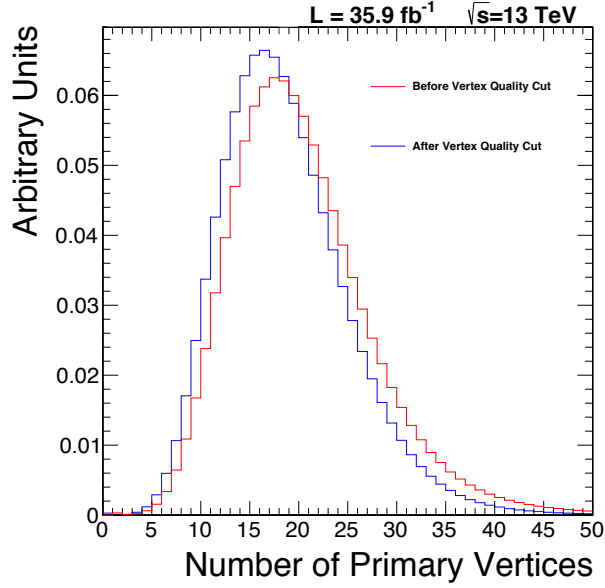


Figure 4.16: The number of PVs for events containing at least one photon in the jet-triggered dataset used for the fake rate calculation. The distributions are shown before (red) and after (blue) further filtering of the PVs is imposed.

PVs. These bins were chosen to span the range and offer an approximately uniform amount of statistics within each bin. The results are shown in Fig. 4.17 for the EB and EE regions. In general, the fake rate increases with the number of PVs.

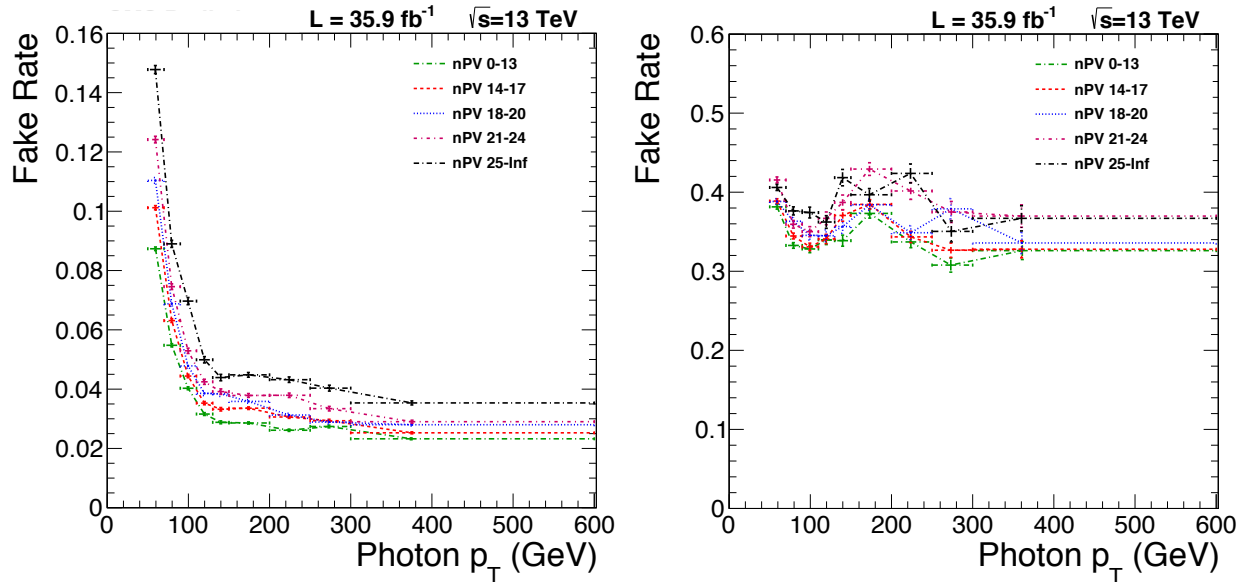


Figure 4.17: The fake rate based on the number of PVs in the events using the jet-triggered dataset in the EB (left) and EE (right).

Another assessment of the fake rate results in its change from using an alternate choice of template variable. The isolation variable Iso_{Ch} , like $\sigma_{i\eta i\eta}$, is sensitive to the differences between real and fake photons and is the best alternate choice among the photon ID variables. The sideband variable must simultaneously be changed and $\sigma_{i\eta i\eta}$ is the natural choice. Note, the Iso_{Ch} and $\sigma_{i\eta i\eta}$ variables are correlated for fake photons due to the underlying jet fragmentation process, but are largely uncorrelated for real photons. The effect of this correlation is studied using simulation in Section 4.2.5. The sideband cuts start at their corresponding cut in the photon ID and are $0.0105 < \sigma_{i\eta i\eta} < 0.0150$ in the EB and $0.0280 < \sigma_{i\eta i\eta} < 0.0400$ in the EE. Fig. 4.18 shows a comparison of the fakes rates produced considering Iso_{Ch} and $\sigma_{i\eta i\eta}$ as separate template variable with the appropriate sideband variable. The fake rates are found to be in close agreement in both the EB and EE categories. Further investigations of this study, along with the templates produced using the Iso_{Ch} variable in a simulated sample, are presented in the closure test studies detailed in Section 4.2.6.

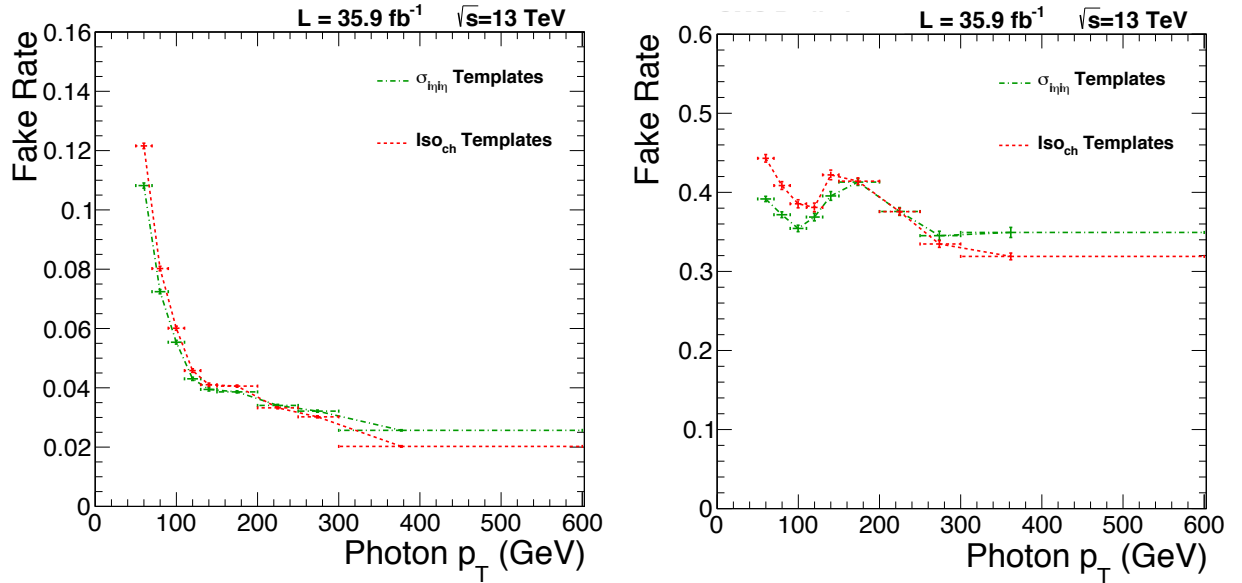


Figure 4.18: Comparison of the fake rates using $\sigma_{i\eta i\eta}$ (green) and Iso_{Ch} (red) template variables in the EB (left) and EE (right).

The choice of cuts for the sideband also affects the fake rate. When using Iso_{Ch} as the template variable, three different $\sigma_{i\eta i\eta}$ sideband choices are considered in each ECAL region:

0.0105-0.0150, 0.0105-0.2000 and 0.0105-1.0000 in the EB and 0.0280-0.0400, 0.0400-0.0600, and 0.028-1.000 in the EE. Fig. 4.19 shows the variation of the fake rate using these sideband choices in each category. Differences are observed between the orthogonal sideband choices. Further investigation of this effect will be considered during the closure test study below, including the variation using the Iso_{Ch} sidebands.

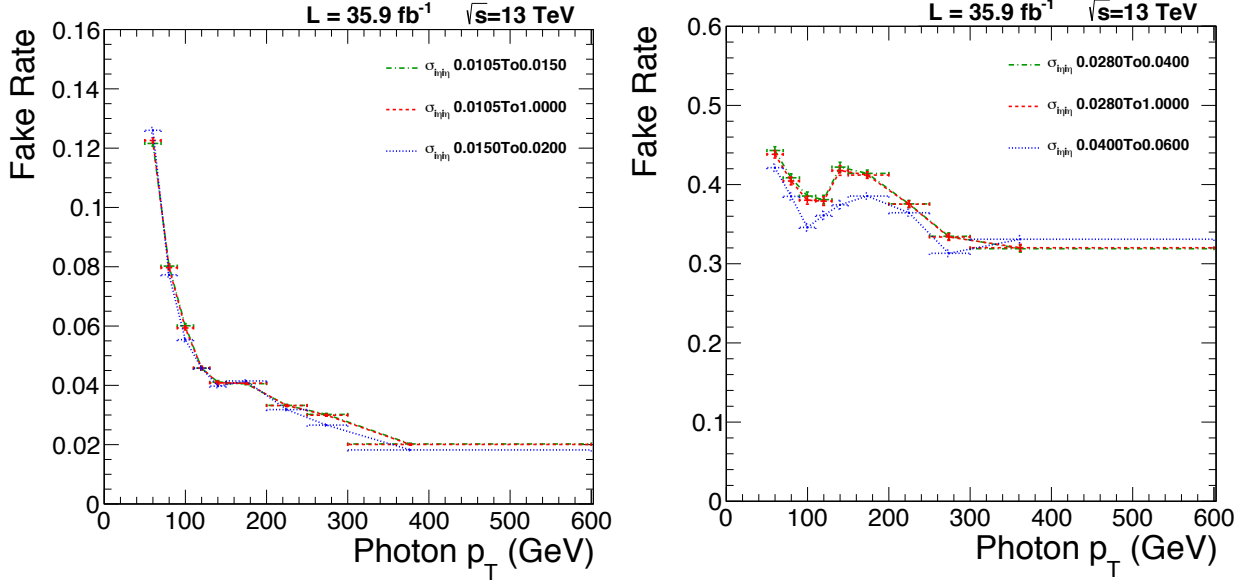


Figure 4.19: The fake rate using a Iso_{Ch} template with variations in the $\sigma_{i\eta i\eta}$ sideband for the EB (left) and EE (right).

4.2.4 Applying the Fake Rate Method

The fake rate function f in Eq. (4.4) can now be applied to events in the diphoton data sample. Events are selected with two objects passing either the high- p_T photon ID, the denominator definition, or a combination of both. Let T denote an object passing the photon ID to signify a true photon passing this tight ID, and let F denote an object passing the denominator definition to signify a potential fake photon failing the photon ID / passing this looser ID. If we label the leading and subleading objects using the index $i = 1$ or 2, respectively, then we have 4 categories of events: T_1T_2 , T_1F_2 , F_1T_2 , and F_1F_2 . Events containing an F object get reweighted by f to be used in the background estimation.

The jj background is estimated by reweighting each of the two F objects for all events in the data sample:

$$\text{jj} = (f_1 F_1)(f_2 F_2) \quad (4.5)$$

where f_i is the fake rate function f evaluated at the p_T of photon i , for $i = 1, 2$. The γj background estimation takes more care because the T objects are comprised of real photons (R) and not real photons (N), i.e., $T = R + N$, where $N = fF$. The γj background is estimated over all data events according to:

$$\gamma\text{j} = (R_1 N_2) + (N_1 R_2) \quad (4.6)$$

$$= (T_1 - f_1 F_1)(f_2 F_2) + (f_1 F_1)(T_2 - f_2 F_2) \quad (4.7)$$

$$= T_1(f_2 F_2) + (f_1 F_1)T_2 - 2(f_1 F_1)(f_2 F_2) \quad (4.8)$$

Therefore, the γj background is estimated by reweighting the TF and FT samples minus twice the FF sample.

4.2.5 Closure Test to the Fake Rate Method

An MC based approach is used to validate the fake rate method performed for determining the jet-faking-photon rate. The general idea is that we perform the fake rate method on a simulated sample using the same procedure as was used on the data sample. This yields a fake prediction in the simulated sample. Since the simulated sample was produced using an MC calculation, we have access to the MC truth information for the generated particles, which provides information about the event history and properties of these particles. From this information, we can identify which types of particles pass the high- p_T photon ID. In particular, we can determine which particles faked a photon signature in the simulated sample and compare this yield to the fake prediction. The QCD jj and γj MC samples listed in Table 4.7 are produced for collaboration-wide use by the CMS Collaboration and contain a sufficient number of fake photons to be used as the simulated sample for this closure test.

The γj samples used here are same samples that were used in the real template construction, listed in Table 4.4.

Table 4.7: The Monte Carlo samples and corresponding cross sections used for the closure test to the fake rate method. The full, internal CMS path name is specified by appending the dataset path with `/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6/MINIAODSIM`.

Dataset path	Cross section (pb)
/QCD_Pt_50to80_TuneCUETP8M1_13TeV_pythia8	19204300
/QCD_Pt_80to120_TuneCUETP8M1_13TeV_pythia8	2762530
/QCD_Pt_120to170_TuneCUETP8M1_13TeV_pythia8	471100
/QCD_Pt_170to300_TuneCUETP8M1_13TeV_pythia8	117276
/QCD_Pt_300to470_TuneCUETP8M1_13TeV_pythia8	7823
/QCD_Pt_470to600_TuneCUETP8M1_13TeV_pythia8	648.2
/QCD_Pt_600to800_TuneCUETP8M1_13TeV_pythia8	186.9
/QCD_Pt_800to1000_TuneCUETP8M1_13TeV_pythia8	32.293
/QCD_Pt_1000to1400_TuneCUETP8M1_13TeV_pythia8	9.4183
/QCD_Pt_1400to1800_TuneCUETP8M1_13TeV_pythia8	0.84265
/QCD_Pt_1800to2400_TuneCUETP8M1_13TeV_pythia8	0.114943
/QCD_Pt_2400to3200_TuneCUETP8M1_13TeV_pythia8	0.00682981
/QCD_Pt_3200toInf_TuneCUETP8M1_13TeV_pythia8	0.000165445
/GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.121e+04
/GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.863e+03
/GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.298e+03
/GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2.816e+02
/GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9.465e+01

The SHERPA $\gamma\gamma$ background prediction described in Section 4.1 takes into account the direct photon contribution due to the fragmentation of quarks and gluons according to Ref. [123]. This suggests that the fakes we need to account for are those coming directly from hadron decays, such as from π^0 or η meson decays. The procedure for identifying these fake photons in the simulated sample is by using the MC truth information associated with the generated particles, as described in the following way.

Reconstructed photons passing the high- p_T photon ID in the simulated sample are matched to the best final state generated particle within $\Delta R \leq 0.1$ around the reconstructed photon. There may not be any match within this cone. If the reconstructed photon is matched directly to a final state generator-level hadron, then it is considered fake. If instead the

reconstructed photon is matched to a final state generator-level photon, then the generated photon’s mothers are used to determine whether or not it is fake. Since in general a generated particle can have multiple mothers within the SHERPA framework, its sole mother is chosen from the best match in ΔR among all of its mothers. If the generated photon’s mother is matched to a generated hadron, then the reconstructed photon is considered fake. This approach isolates the fake photons from hadron decays and, for example, ignores prompt photons directly from the hard interaction.

The validation of the fake rate method is done in three steps. In each case, the method is performed on the simulated sample, treating it as if it were data, which gives a prediction that can be compared to the correct result obtained using the MC truth information on the same sample. First, the accuracy of fake templates produced using the sideband method is tested by comparing the fake templates obtained in simulation using the sideband definition, as done in data, with the fake template that is obtained using the MC truth information. Second, we test the use of the templates to fit the discriminator variable and extract the fake yields. This is done by performing the fitting procedure on the simulation in the same manner as done in data and comparing against the actual fake yields using the MC truth information, both as a function of p_T . Third, we test the validity of using reweighted denominator objects to predict the kinematics of actual known fakes—in particular, the diphoton invariant mass distribution. These three tests are described sequentially below.

Testing the Fake Template Derived using the Sideband Definition

The MC fake templates are derived from the combined jj and γj MC sample using the same selection and Iso_{Ch} sideband definitions as are used when performing the fake rate on data. These fake templates are p_T dependent so the closure test is performed in photon p_T bins of 50-70, 70-90, 90-130, 130-200, and 200-600 GeV. These bins are coarser than the ones used in the actual fake rate method due to the limited statistics in the MC sample. Because we are limited in choosing a sideband outside of the photon ID cut of $\text{Iso}_{\text{Ch}} > 5 \text{ GeV}$, and due

to its correlation with the template variable $\sigma_{i\eta i\eta}$, we know that the choice of sideband is biased. The target fake template is constructed using the MC truth information to form the $\sigma_{i\eta i\eta}$ distribution of the actual fakes in the MC sample. The fake templates derived from this sample using the $\sigma_{i\eta i\eta}$ template variable with different Iso_{Ch} sideband choices are compared against the MC truth fake templates and are shown in Fig. 4.20. The comparison is done using the lowest p_{T} bin of 50-70 GeV separated in the EB and EE regions. The remaining p_{T} bins are shown in Appendix B.2.

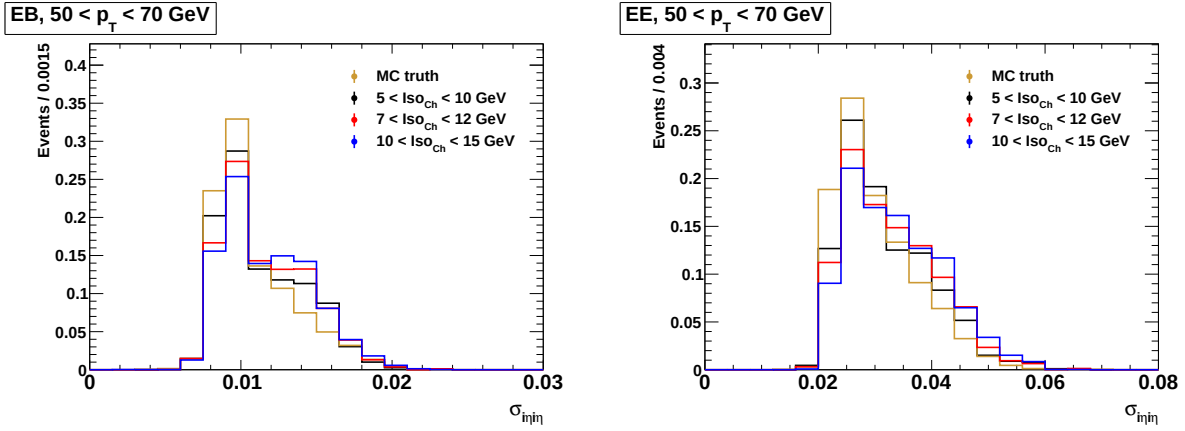


Figure 4.20: Fake templates extracted from the MC sample used in the fake rate closure test compared against the MC truth fake templates for $50 < p_{\text{T}} < 70$ GeV in the EB (left) and EE (right). The template variable is $\sigma_{i\eta i\eta}$.

We see that as the sideband definition is pushed further away from the region within the photon ID cut, the resulting template becomes more biased towards larger $\sigma_{i\eta i\eta}$ compared to the truth information. We choose our nominal sideband to have the selection of $5 < \text{Iso}_{\text{Ch}} < 10$ GeV. Fig. 4.21 shows the p_{T} evolution of these fake templates using the nominal Iso_{Ch} sideband choice.

We see from this study that the sideband-defined template does a reasonable job of replicating the truth template for known fakes. There is clearly some bias, as expected, but, unfortunately, some bias is unavoidable since fakes in a sideband are intrinsically different from fakes that pass the photon ID. This bias is limited but is the main source of systematic uncertainty on the fake rate method.

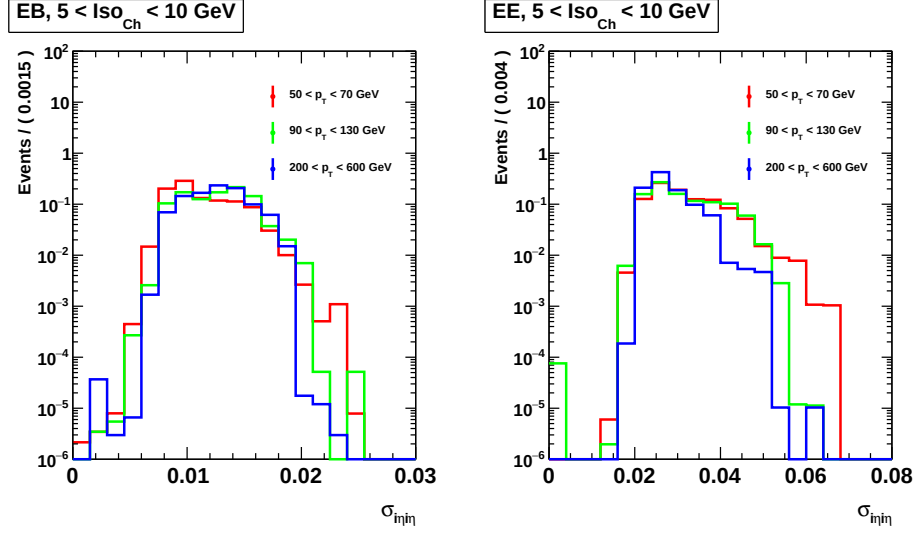


Figure 4.21: Fake templates used in the MC based closure test showing their p_T evolution using the nominal Iso_{Ch} sideband definition of 5-10 GeV in the EB (left) and EE (right) regions.

Testing the Fit Method using Templates to Extract Fake Yields

With the creation of the fake templates validated in the simulated sample, we can next test the fake prediction obtained by using these templates to fit to the numerator $\sigma_{i\eta i\eta}$ distribution. This fit uses the same real templates that are used in the data fake rate method, but using the different p_T binning. Fig. 4.22 shows the real templates normalized to 1 fb^{-1} in the EB and EE categories comparing different photon p_T bins. We see the higher p_T photons are more isolated.

To better handle the lower statistics in the MC sample, the template fit is performed (in $\sigma_{i\eta i\eta}$) using a negative log-likelihood function with errors calculated using the method by Barlow and Beeston [127] incorporating an error term in the likelihood function. The fits are done for each p_T bin in the EB and EE separately. An example fit for photons in each category with $90 < p_T < 130 \text{ GeV}$ is shown in Fig. 4.23. The fits for all p_T bins are given in Appendix B.1.

After the fit, the photon ID cut for $\sigma_{i\eta i\eta}$ is applied to the fake template and it is integrated in order to provide a fake prediction in each separate p_T bin. This is the numerator of the fake

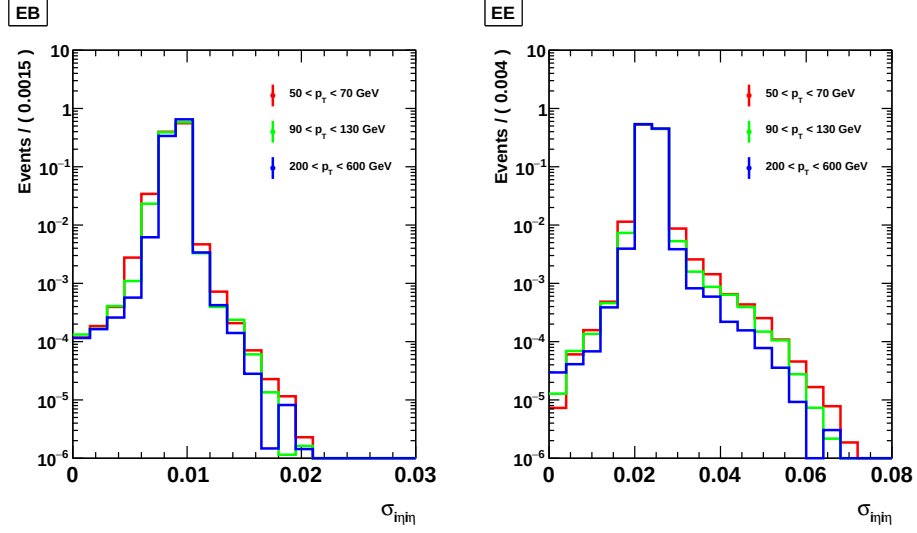


Figure 4.22: Real templates using $\sigma_{\eta\eta}$ showing the evolution with photon p_T in the EB (left) and EE (right).

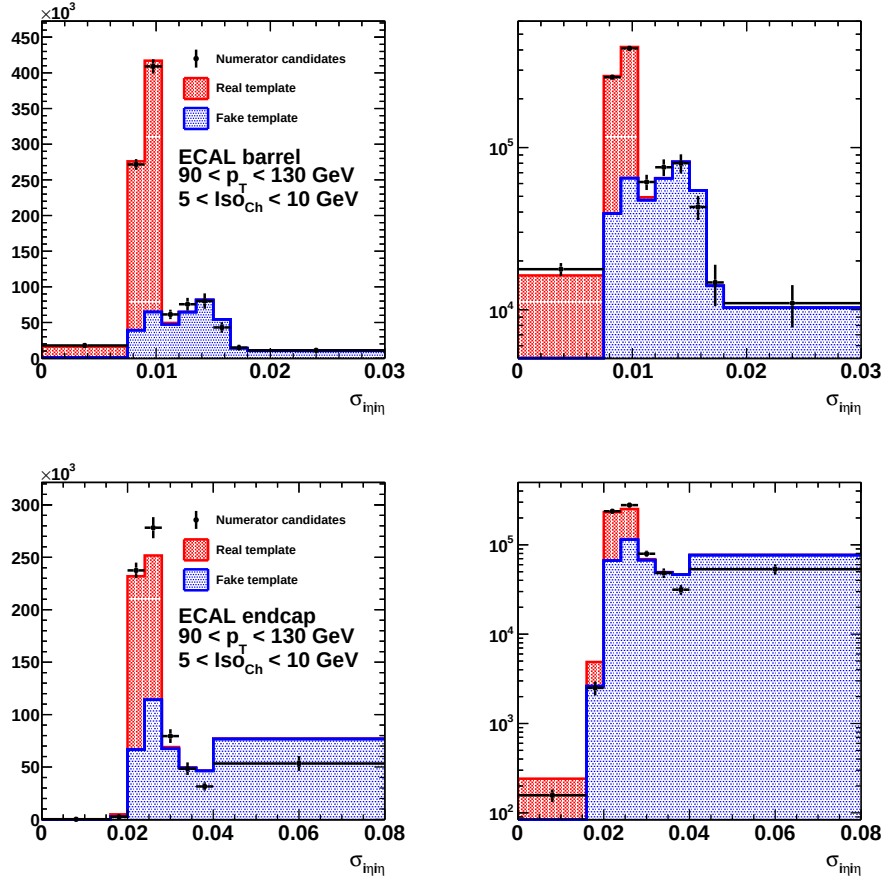


Figure 4.23: Template fits using $\sigma_{\eta\eta}$ for the MC closure test with $90 < p_T < 130$ GeV in the EB (top) and EE (bottom).

rate in the MC sample as a function of p_T . Dividing by the denominator extracted from the MC sample yields MC fake rate as a function of p_T . This procedure was repeated using fake templates constructed from sidebands with $7 < \text{Iso}_{\text{Ch}} < 12 \text{ GeV}$ and $10 < \text{Iso}_{\text{Ch}} < 15 \text{ GeV}$ in the MC sample. The three MC fake rates corresponding to all three different sideband choices are shown in Fig. 4.24. In addition, since we have access to the MC truth information, we can simply count the number of fake photons in each p_T bin and divide by the same denominator to get an MC truth fake rate. This result is also shown in Fig. 4.24.

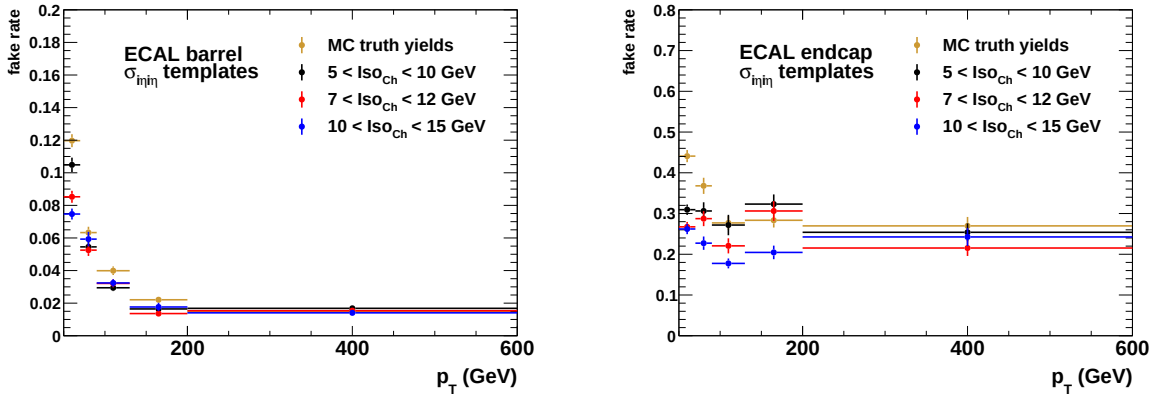


Figure 4.24: Fake rates for the MC closure test as a function of p_T using σ_{inj} templates in the EB (left) and EE (right). Three different choices for the Iso_{Ch} sidebands used in deriving the fake rates are compared against the fake rate using the MC truth information.

We see that, in general, the template fit method somewhat underpredicts the actual number of fakes. This underprediction is attributed to the bias in the fake template constructed from the sideband definition, as seen above. We see that the underprediction is minimized by choosing a sideband in Iso_{Ch} as close as possible to the photon ID region ($\text{Iso}_{\text{Ch}} < 5 \text{ GeV}$), as expected from the above study. We quantify the underprediction by calculating the ratio of the derived fake rate with the known MC truth fake rate as a function of p_T . This ratio is shown in Fig. 4.25 using our nominal Iso_{Ch} sideband for the MC fake rate prediction. In general, we observe an underprediction within an uncertainty of about 30%, a degree of nonclosure consistent with this method based on its use in previous CMS analyses, such as in Ref. [61].

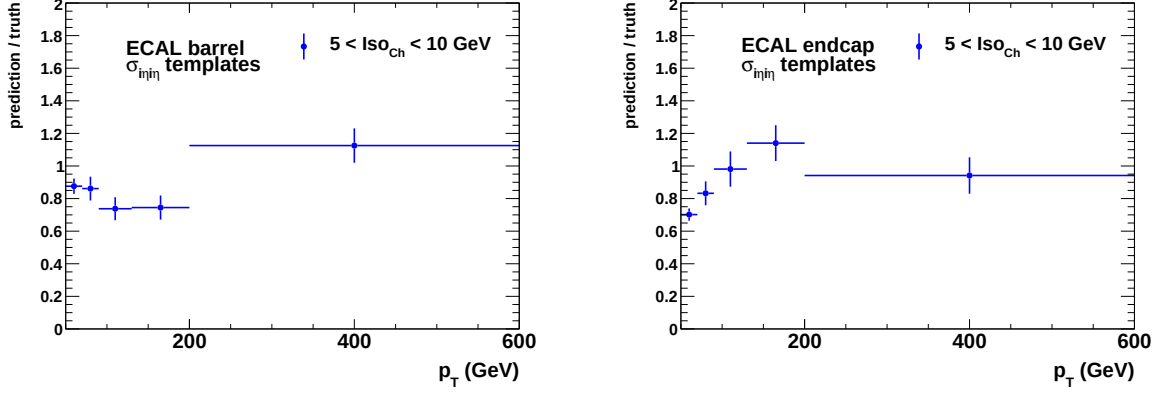


Figure 4.25: Comparison of the ratio of the fake yields obtained from the MC fake prediction using $\sigma_{inj\eta}$ templates over the yields identified by the MC truth information in the EB (left) and EE (right) regions.

An alternative way of obtaining the MC truth fake rate is to use the MC truth information to derive the fake templates and then use them in the fitting procedure to produce the fake rate. This alternative procedure was performed and an example template fit using these fake templates in the bin $70 < p_T < 90$ GeV is shown in Fig. 4.26, separately for the EB and EE regions. The corresponding fake rates are shown in Fig. 4.27. These fake rates are in good agreement with those obtained using the MC truth yields, previously presented in Fig. 4.24, helping to validate the fitting procedure.

Testing the Kinematics of Reweighted Denominator Objects

The final step in the MC closure test is to verify that taking denominator objects (selected to be photon-like jets) and reweighting them by the fake rate does indeed produce an accurate representation of the kinematics of known fake events. This is to validate the final step of the actual fake background estimation technique used in the analysis where denominator objects are selected in the diphoton dataset and are reweighted by the fake rate measured in the jet-enriched dataset.

The fake rate for the MC sample is obtained by fitting a function of the same functional form as in Eq. (4.4) to the MC fake rate derived using the nominal Iso_{Ch} sideband selection,

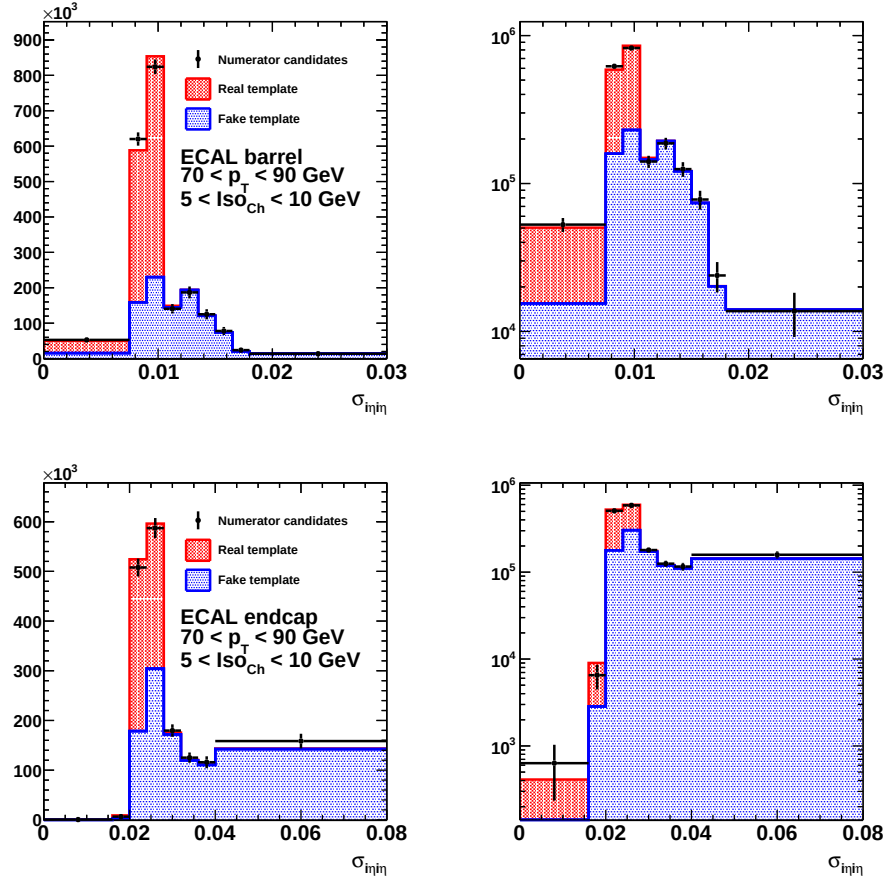


Figure 4.26: Template fits using $\sigma_{i\eta i\eta}$ with fake templates derived using the MC truth information in the simulated sample with the nominal sideband definition for $70 < p_T < 90$ GeV in the EB (top) and EE (bottom) regions.

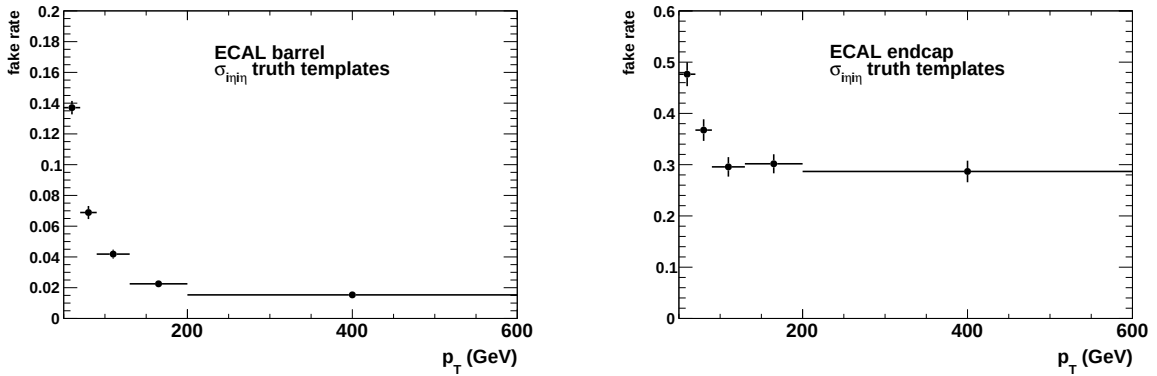


Figure 4.27: Fake rate derived using fake templates constructed using the MC truth information in the simulated sample as a function of p_T with $\sigma_{i\eta i\eta}$ templates in the EB (left) and EE (right) regions.

shown in Fig. 4.24, separately in both the EB and EE. The fitted function and its associated fit parameters p_0, p_1 , and p_2 are shown in Fig. 4.28. This function is used to reweight the denominator objects selected in the MC sample.

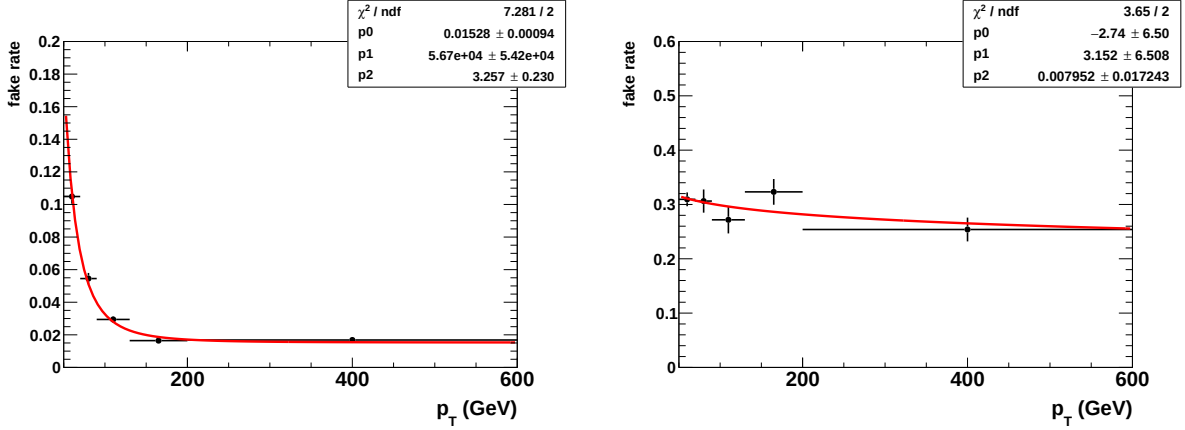


Figure 4.28: MC fake rate as a function of p_T in the EB (left) and EE (right) using the nominal Iso_{Ch} sideband of 5-10 GeV.

Fig. 4.29 shows the single photon kinematic variables p_T , η , and ϕ in the EB and EE regions for the MC denominator objects, the reweighted MC denominator objects, and distributions from the actual fakes in the MC sample determined using the MC truth information. Overall, the agreement is excellent in each region, but with better agreement in the EB than in the EE category. There are noticeable discrepancies at high values of η , but the level of disagreement is within the uncertainty ascribed to this fake rate method. With more statistics, this dependence could be accounted for by further subdividing the η region and producing separate fake rates in each region, instead of just in the EB and EE regions.

We also check the kinematics of the diphoton system. Diphotons are categorized in the EBEB, EBEE, EEEB, or EEEE regions depending on whether the leading or subleading photon is in the EB or EE. The $m_{\gamma\gamma}$ spectra are calculated using two real photons ($\gamma\gamma$), one real and one fake photon (γj), or two fake photons (jj). Because the MC statistics are limited, we only compare the MC truth γj spectra to the MC fake rate weighted γj spectra, omitting the jj comparisons. These spectra are further categorized by whether the leading or subleading photons are real (T) or fake (F).

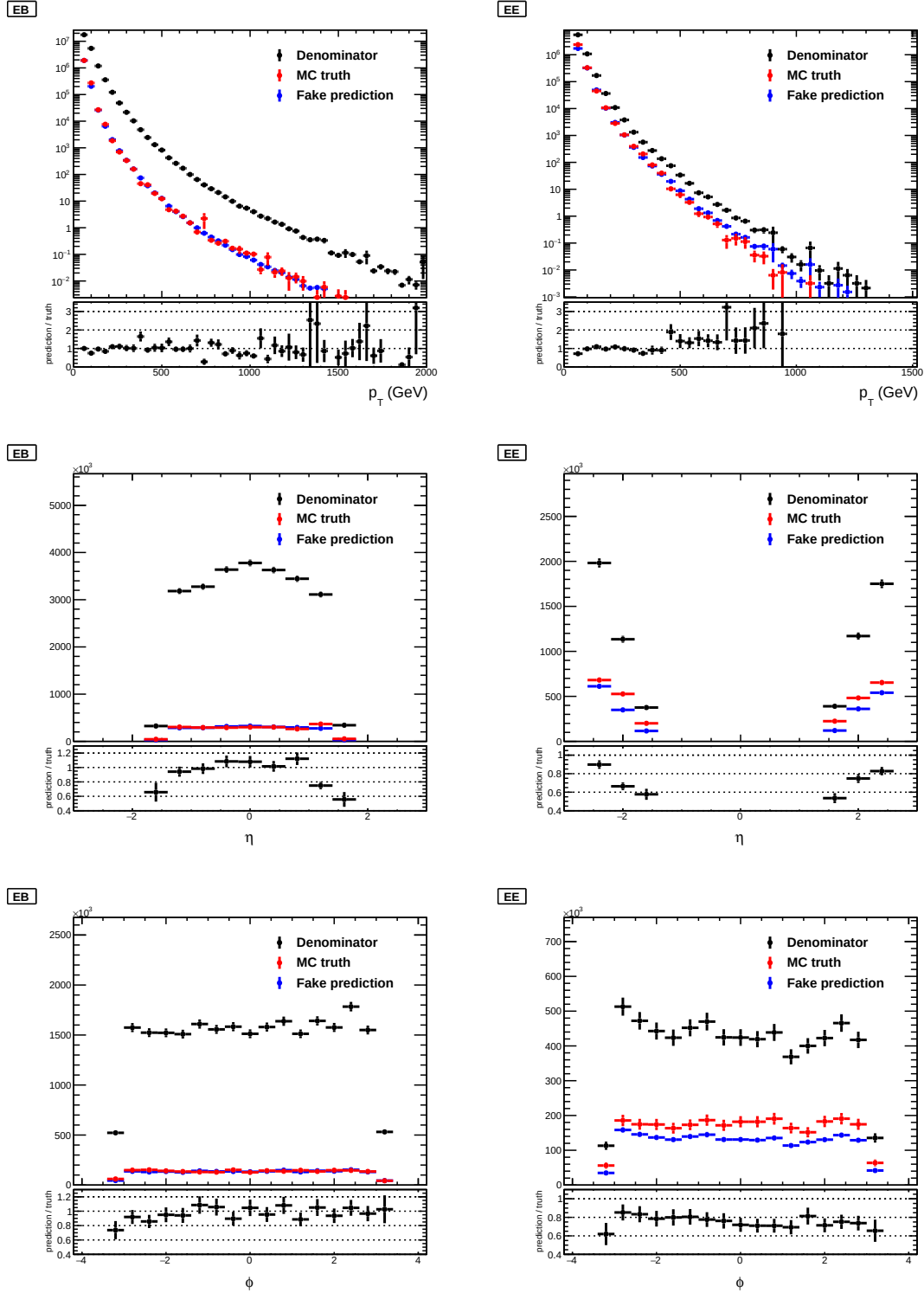


Figure 4.29: Closure test of the photon variables p_T (top), η (middle), and ϕ (bottom) in the EB (left) and EE (right) regions. The MC denominator objects (black) are reweighted by the MC fake rate (blue) and compared against the distributions of known fakes identified using the MC truth (red).

All possible categories using the γj $m_{\gamma\gamma}$ distributions are merged to show the distribution as used in the main analysis, which combines the TF and FT ordering, drops the EEEE region, and merges the distinct EBEE and EEEB regions into the analysis-level EBEE category, as shown in Fig. 4.30. The distributions for the individual categories of this type are shown in Fig. 4.31. We can see that the diphoton invariant mass distribution obtained from events with denominator objects reweighted by the fake rate does indeed reproduce well the distribution that would be obtained for known fakes using the MC truth information.

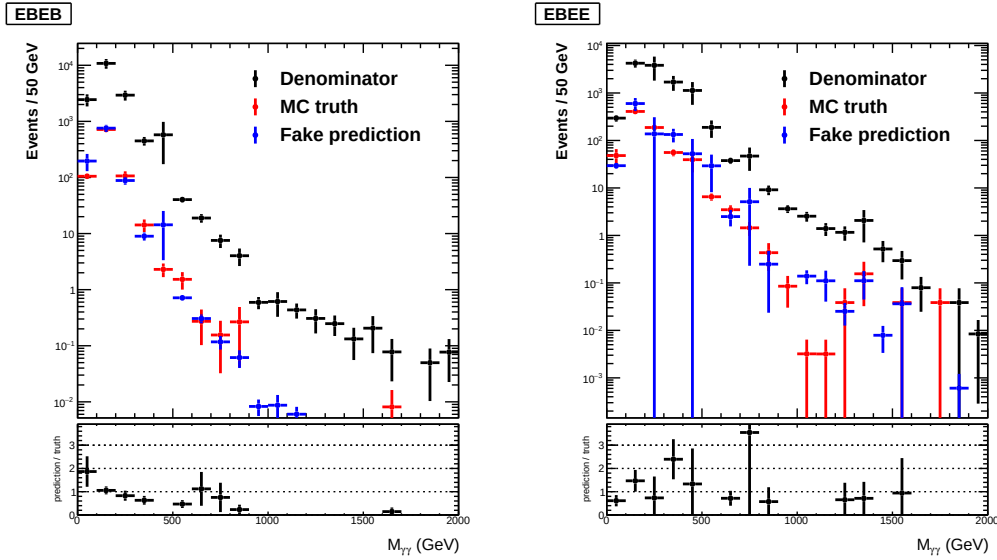


Figure 4.30: Closure test of the diphoton $m_{\gamma\gamma}$ distribution of γj objects in the EBEB (left) and EBEE (right) categories. The MC denominator objects (black) are reweighted by the MC fake rate (blue) and compared against the distributions of known fakes identified using the MC truth (red).

This concludes the MC closure test and demonstrates the validity of the photon fake rate method. Further studies involving different aspects of the closure test are discussed in the next section.

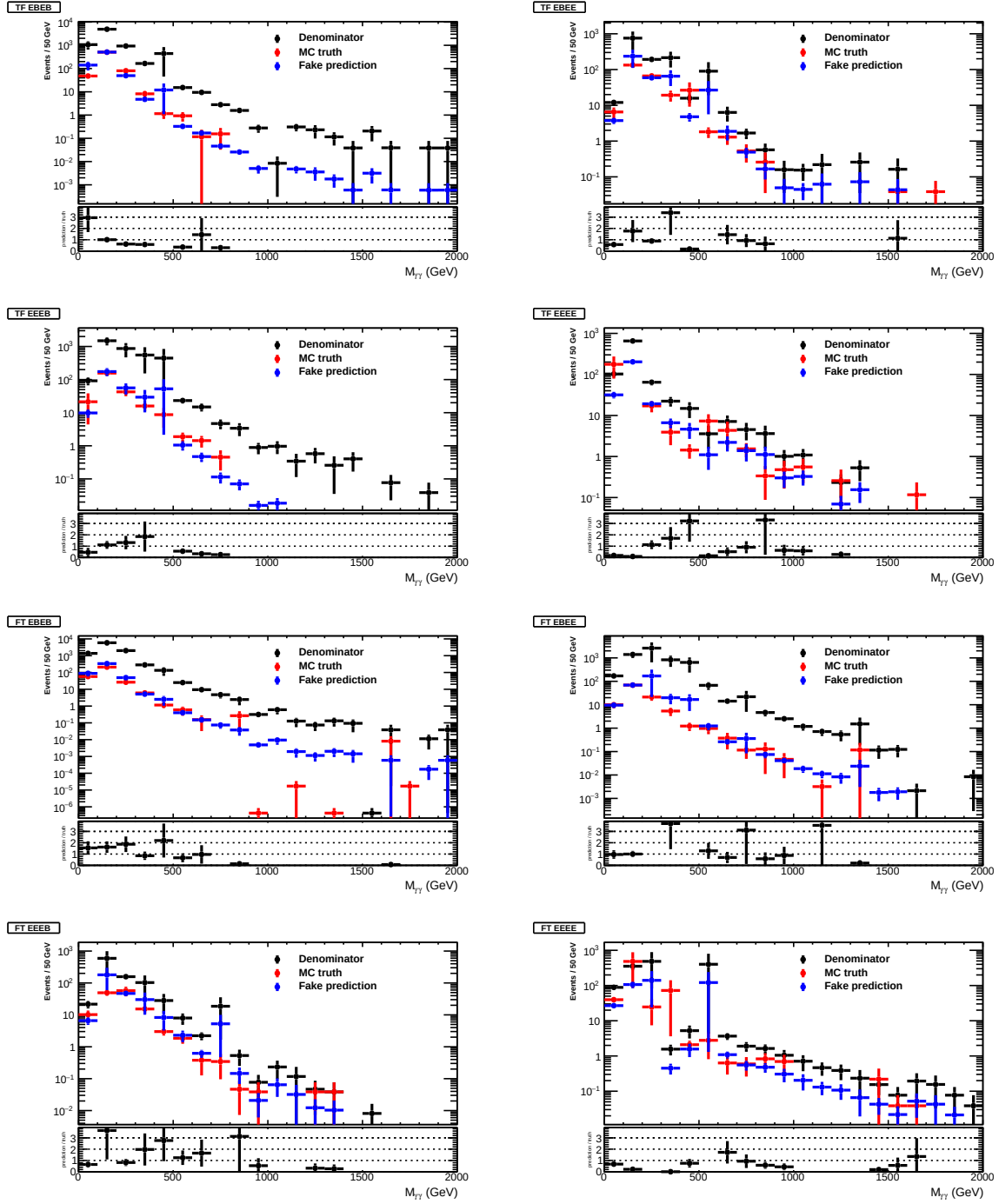


Figure 4.31: Closure test of the diphoton $m_{\gamma\gamma}$ distribution of γj objects ordered as TF (top 4 plots) and FT (bottom 4 plots) in the EBEB, EBEE, EEEB, and EEEE categories (specified in the plot titles). The MC denominator objects (black) are reweighted by the MC fake rate (blue) and compared against the distributions of known fakes identified using the MC truth (red).

4.2.6 Closure Test Studies

Investigating the Change in Template Variable

All aspects of the above closure test were done using $\sigma_{i\eta i\eta}$ as the template variable with fake templates derived using a Iso_{Ch} sideband. We now repeat the closure test using the inverted relationship between these two variables, i.e., using Iso_{Ch} as the template variable and $\sigma_{i\eta i\eta}$ as the sideband variable. The real templates constructed using the Iso_{Ch} variable are formed from the simulated sample using the datasets listed in Table 4.4, and Fig. 4.32 shows their evolution as a function of photon p_{T} . A nice feature of using this variable is that the templates are independent of photon p_{T} .

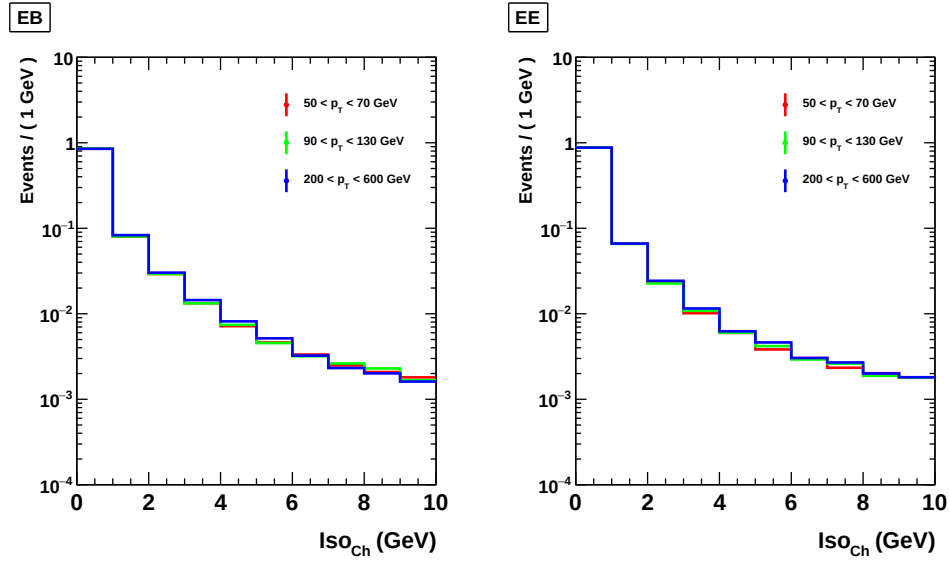


Figure 4.32: Real templates in Iso_{Ch} derived from simulation in the EB (left) and EE (right) categories for different photon p_{T} bins.

Similarly, the fake templates constructed using the Iso_{Ch} variable are formed from the simulated sample using the datasets listed in Table 4.7. The same sideband definitions for the $\sigma_{i\eta i\eta}$ variable that were used in the construction of the fake templates in Section 4.2.3 are also used here. These definitions differ between the EB and EE regions. Fig. 4.33 shows these fake templates compared to the distribution of MC truth fakes for $90 < p_{\text{T}} < 130$ GeV. The

distributions using the other p_T bins considered in the closure test are shown in Appendix B.2. The expected bias in these templates is observed. We choose our nominal $\sigma_{i\eta i\eta}$ sideband to be 0.0105-0.0150 in the EB and 0.0280-0.0400 in the EE category. Using these choices, the p_T dependence of the Iso_{Ch} fake templates is assessed and shown in Fig. 4.34. These templates are also observed to be p_T independent.

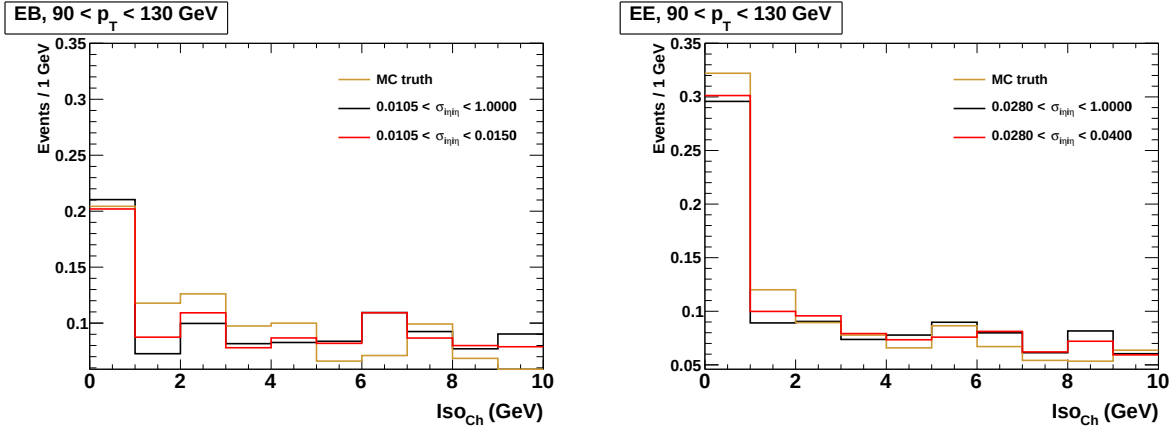


Figure 4.33: Fake templates in Iso_{Ch} using different $\sigma_{i\eta i\eta}$ sidebands from simulation for $90 < p_T < 130$ GeV in the EB (left) and EE (right) categories, compared against the MC truth prediction.

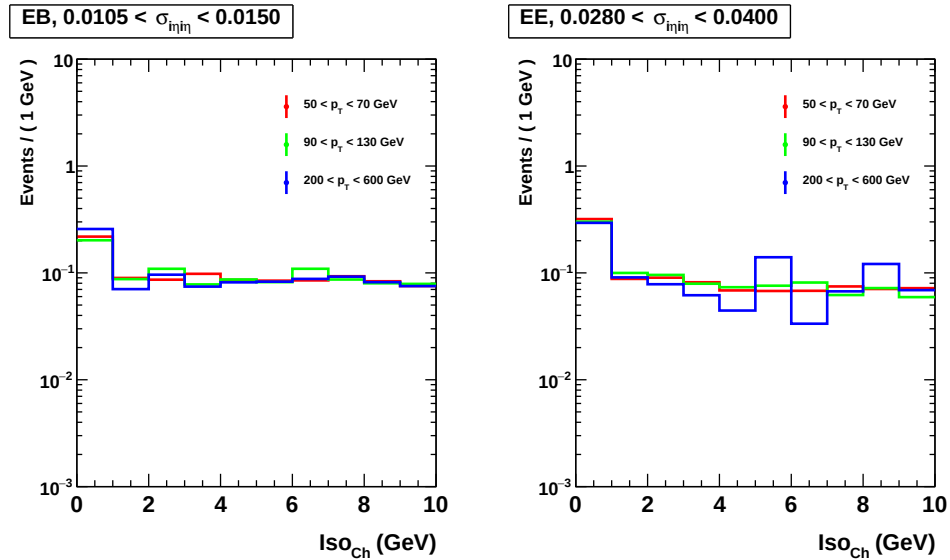


Figure 4.34: Fake templates from simulation using the nominal $\sigma_{i\eta i\eta}$ sidebands in the EB (left) and EE (right) categories for different p_T bins.

Similar to the procedure done in the standard closure test, these templates were fit to numerator candidates using the simulated sample. Fig. 4.35 shows an example fit using the nominal $\sigma_{i\eta i\eta}$ sideband choices in the EB and EE regions for $90 < p_T < 130$ GeV. The distributions using the other p_T bins are provided in Appendix B.1. From the fits, we can extract the fake prediction in the simulated sample and derive a corresponding fake rate. This process was repeated using each of the sideband choices. The fake rates derived from each choice are compared against the yields of the MC truth fakes in Fig. 4.36. There is some disagreement due to the mismodeling of the fakes by the fake templates. The ratio of the fake rate obtained using the nominal sideband choices to the MC truth fake rate is shown in Fig. 4.37. The final closure is about 30% in the EB and 10% in the EE category.

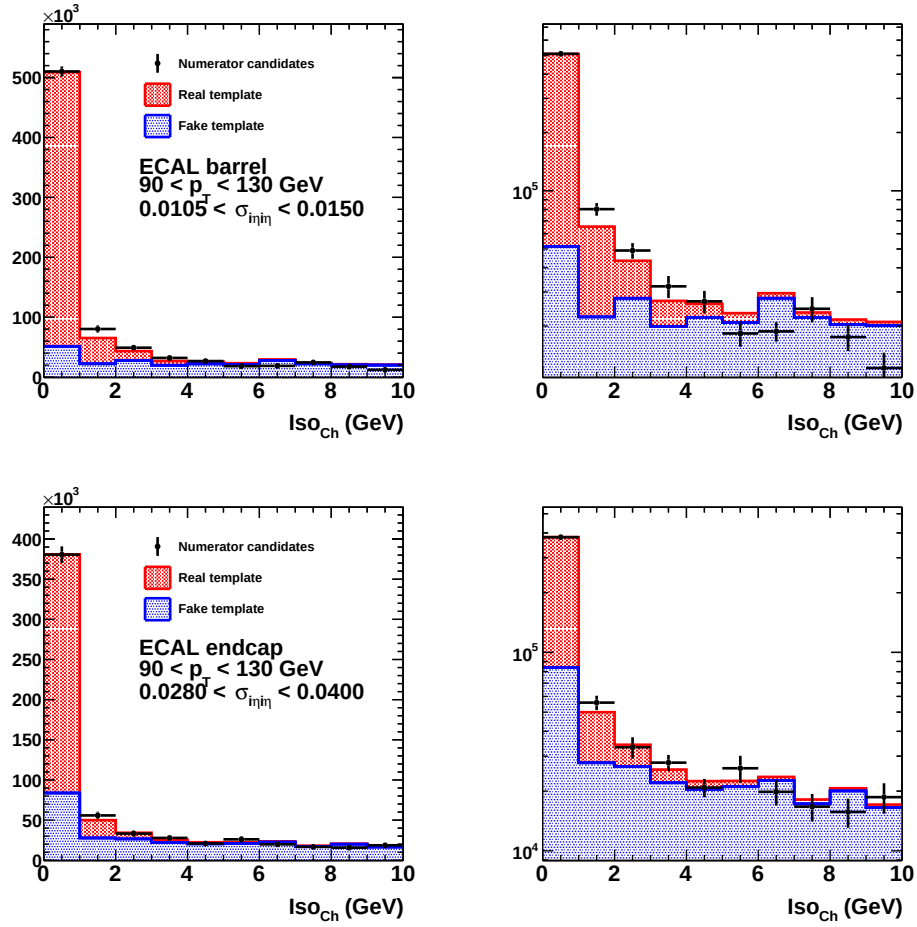


Figure 4.35: An example template fit using the Iso_{Ch} variable and nominal $\sigma_{i\eta i\eta}$ sidebands for the fake templates in the EB (top) and EE (bottom) categories with $90 < p_T < 130$ GeV.

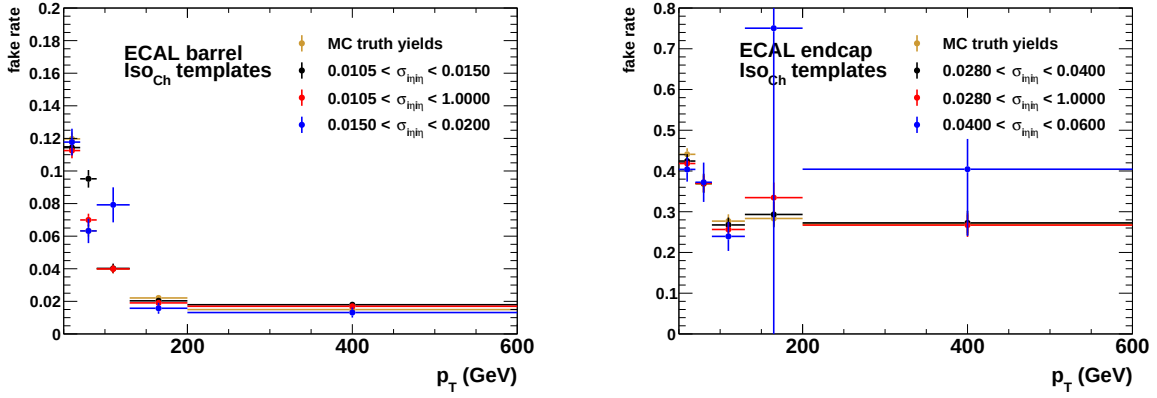


Figure 4.36: The fake rate as a function of p_T using Iso_{Ch} templates from simulation in the EB (left) and EE (right) categories with different $\sigma_{i\eta i\eta}$ sidebands, compared to the MC truth fake rate.

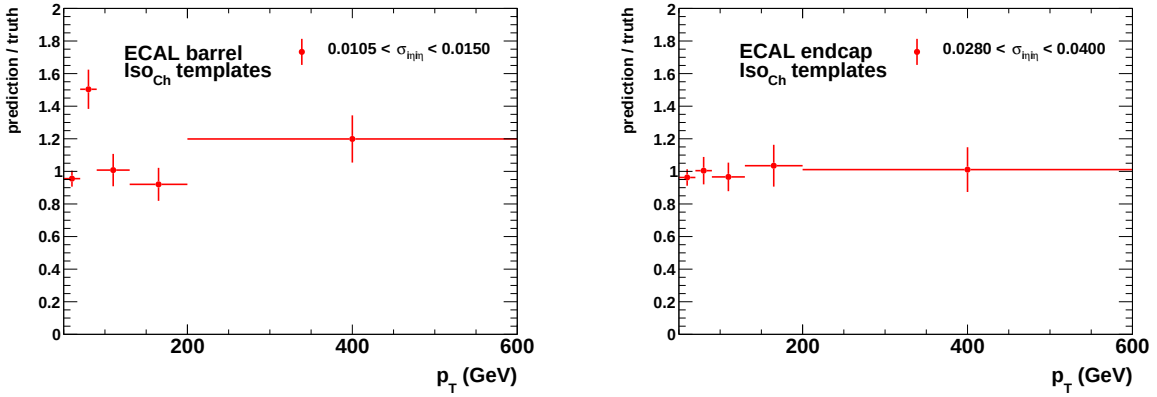


Figure 4.37: Closure test of the event yields from the fake prediction in simulation using Iso_{Ch} template fits compared to the MC truth yields in the EB (left) and EE (right) categories.

Next, we compare the above fake rates derived using the nominal sideband definitions to the ones obtained earlier using the $\sigma_{i\eta i\eta}$ templates. Both fake rates are shown in Fig. 4.38 and compared against the MC truth fake rate, separately in the EB and EE categories. There is good agreement among all three fake rates in both categories. For both template variables, the ratios of the predicted MC fake rates to the MC truth fake rate are given in Fig. 4.39. Both template choices largely agree within statistical uncertainty with some disagreement at low p_T in the EB. We stick with $\sigma_{i\eta i\eta}$ as the nominal template variable used for the fake rate method.

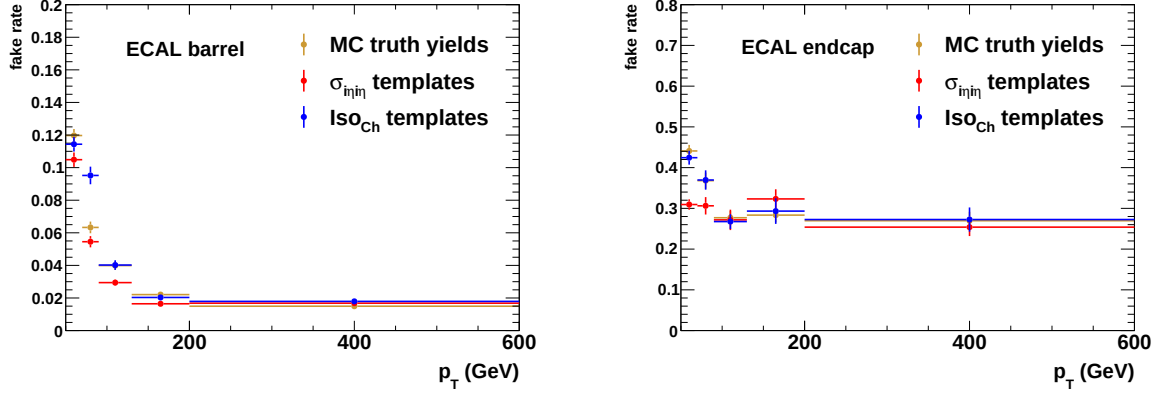


Figure 4.38: Comparison of the predicted MC fake rates as a function of p_T using $\sigma_{i\eta i\eta}$ (red) and Iso_{Ch} (blue) templates in the EB (left) and EE (right) categories compared to the MC truth fake rate (gold).

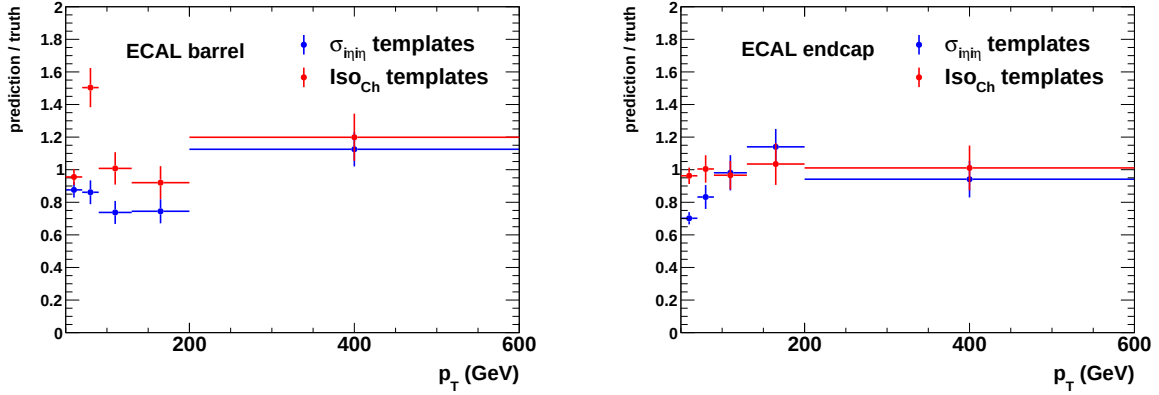


Figure 4.39: Closure test of the event yields from the fake prediction in the simulated sample using $\sigma_{i\eta i\eta}$ (blue) and Iso_{Ch} (red) templates in the EB (left) and EE (right) regions.

A final study between these two template variables is a comparison of their predicted fake rates using the MC truth information to form their fake templates. These fake rates using the $\sigma_{i\eta i\eta}$ template variable were derived previously and shown in Section 4.2.5. The same procedure was repeated using the Iso_{Ch} templates and the resulting fake rates are shown in Fig. 4.40, compared to the MC truth fake rate derived directly from the fake yields. All three of these are in excellent agreement, further validating the fitting procedure.

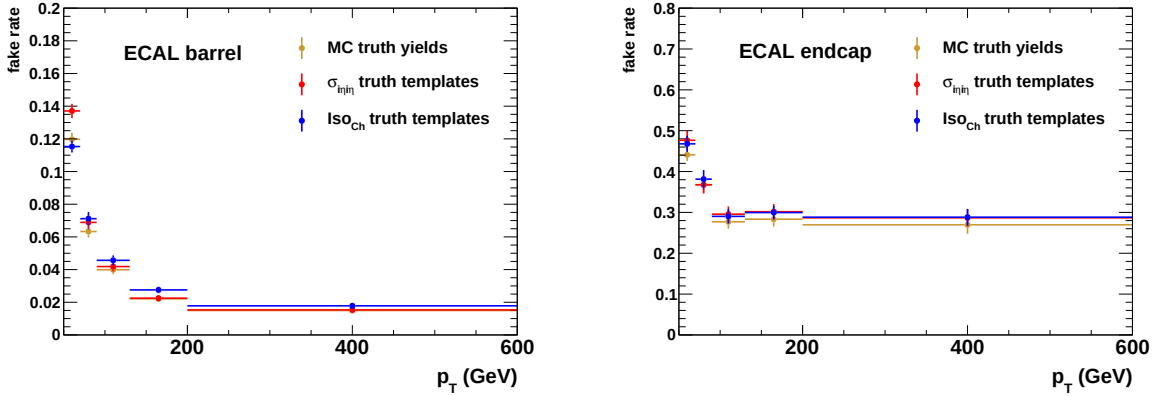


Figure 4.40: Fake rates derived using the MC truth information to construct fake templates in $\sigma_{i\eta i\eta}$ (red) and Iso_{Ch} (blue) as a function of p_T in the EB (left) and EE (right) regions. These are compared against the MC fake rate derived from the fake yields (gold).

Investigation of the Fake Composition from MC Truth

In general, the fake templates using $\sigma_{i\eta i\eta}$ (see, e.g., Fig. 4.20) are peaked near the cut value of $\sigma_{i\eta i\eta}$ used in the high- p_T photon ID. This peak is usually expected for real photons and is more pronounced at low- p_T . We can see how isolated these reconstructed, known fake photons are by using the MC truth information available for the generated particles in the simulated sample used to select them.

Considering a region of $\Delta R < 0.1$ around the fake photons reveals that there are mostly two final state generated particles present, but often there are one, and sometimes, three particles. This does not depend on whether or not the fake photon passes the $\sigma_{i\eta i\eta}$ cut. Further investigation reveals that these final state generated particles mostly come from

a single generated hadron mother, as expected. In the case of a single hadron mother producing a single final state generated particle, this creates a well isolated electromagnetic energy deposit (passing the $\sigma_{i\eta i\eta}$ cut) giving rise to the peak in the templates. When the single hadron mother decays to two final state generated particles, we need to consider their separation.

When there are two final state generated particles within a region of $\Delta R < 0.1$ from the fake reconstructed photon, we calculate their separation in ΔR . This is done separately for the cases when the fake photon passes and fails the $\sigma_{i\eta i\eta}$ cut value of the photon ID. Fig. 4.41 shows this separation. A single crystal in the EB has a segmentation of $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$. We see that most often these two particles are deposited within the same ECAL crystal, giving rise to a well isolated $\sigma_{i\eta i\eta}$ shower shape and contributing to the peak in the fake templates inside of the $\sigma_{i\eta i\eta}$ cut.

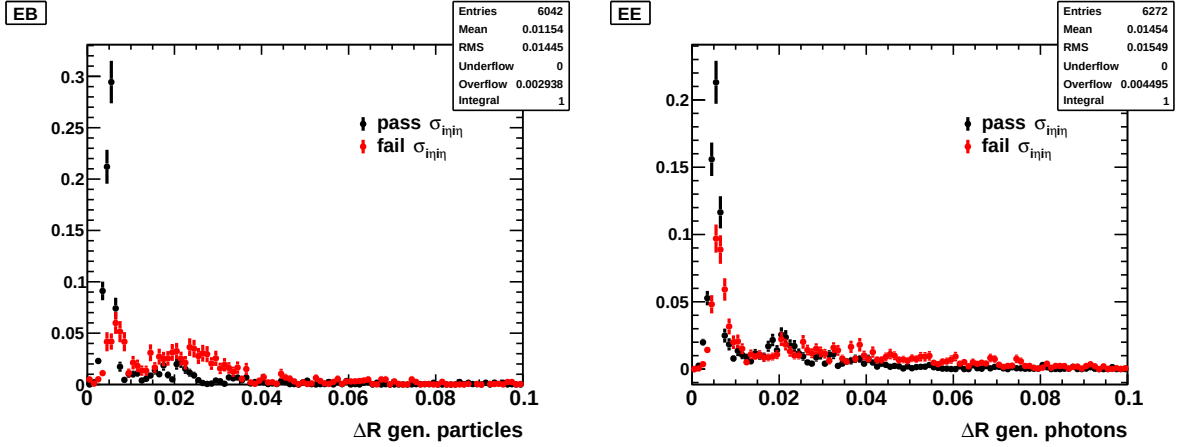


Figure 4.41: The separation in ΔR among two final state generated particles near a reconstructed fake photon in the EB (left) and EE (right) regions. The separation is shown separately for the fake photons that pass (black) and fail (red) the $\sigma_{i\eta i\eta}$ photon ID cut.

Real Photon Contamination

We now check for possible contamination of real photons in the sideband, which is used to model fake photons, and in the denominator, which, unlike the numerator, is not corrected for. We check both cases separately, starting with the former.

Shown in Fig. 4.42 are the $\sigma_{i\eta i\eta}$ distributions in the EB and EE regions of different objects in the 5-10 GeV sideband from the simulated sample used in the closure test. Every reconstructed object in the sideband is matched to a generated particle using the MC truth information. A true source of fakes are those that are matched to a hadron mother. A true source of real photons are those that are matched to a hard scattering photon. There still remains some objects that are not matched and don't fall into either category. This mixture is predominantly fake-like (as can be inferred from their $\sigma_{i\eta i\eta}$ distribution) and are indicated as “other” in Fig. 4.42 below. This plot shows that the fraction of real photons is small in the sideband. This small contamination is accounted for in using the variation of the sideband definition as a systematic uncertainty.

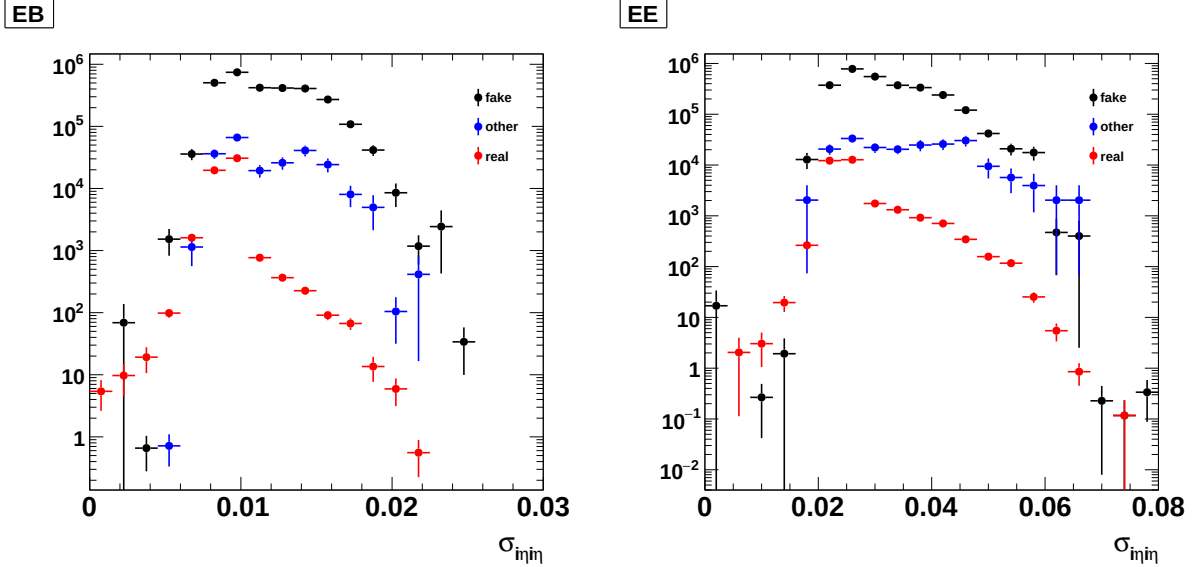


Figure 4.42: Composition of objects in the 5-10 GeV sideband in the simulated sample used for the closure test shown in the EB (left) and EE (right) categories. The fake (black) and fake-like (blue) objects dominate over the real (red) ones.

Although the denominator objects are required to fail at least one of the photon ID variables, ensuring that the two categories are mutually exclusive, small contamination is still possible. To understand the real photon contamination in the denominator, we use the MC truth information available for the simulated sample used by the closure test. We take the reconstructed objects that pass the denominator definition and match them to a

generated particle, as done above. Again, the real photons are those that are matched to a generated photon directly from the hard scattering. The fraction of these real photons over the total denominator objects as a function of p_T in the EB and EE categories is shown in Fig. 4.43. This contribution is less than 10% over the majority of p_T range up to about 600 GeV. Although this fraction does rise with p_T , the effect is accounted for in the overall uncertainty attributed to the fake background.

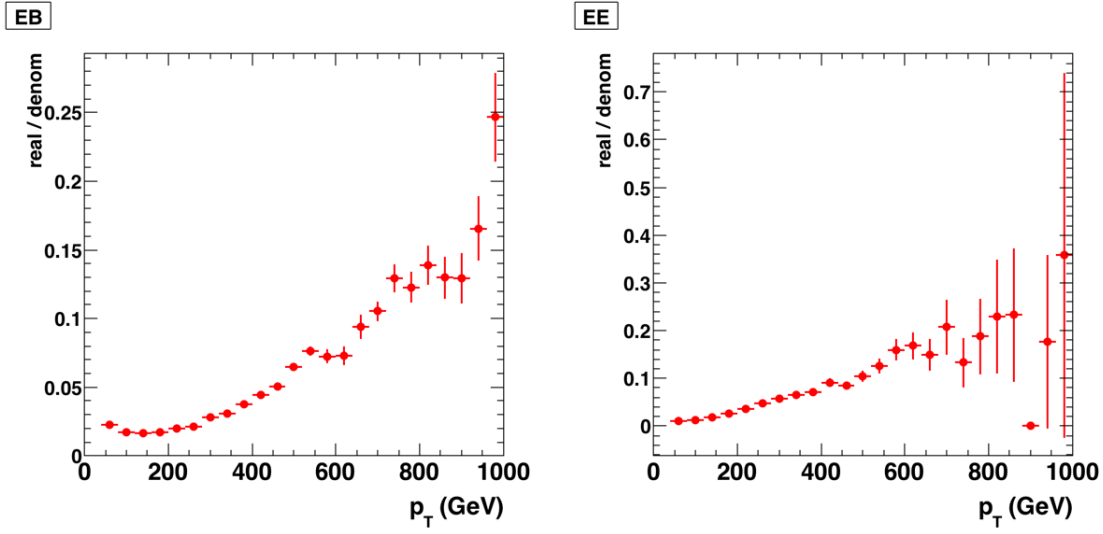


Figure 4.43: The fraction of real photons among all denominator objects as a function of p_T , derived from the simulated sample used by the closure test in the EB (left) and EE (right) regions.

Chapter 5

Signal Simulation

This analysis is searching for large extra-dimensional signals as described by the ADD model. This BSM signal produces a nonresonant excess over the SM diphoton background and is modeled using MC simulation. A large parameter space is considered in the search.

5.1 Signal Generation

For this search, the final state of ADD signal comes from the excited state of the KK graviton decaying to two high-mass photons ($G_{\text{KK}} \rightarrow \gamma\gamma$). This final state is initiated through quark annihilation and gluon fusion at the LHC, as depicted in the Feynman diagrams in Fig. 1.5, and is mediated by virtual graviton exchange. Since proton-proton collisions are used at the LHC, the gluon-initiated processes dominate over the quark-initiated ones, due to the larger gluon contribution to the proton PDFs, as previously shown in Fig. 4.3.

Similar to the real background calculation, the ADD signal is calculated with the MC event generator SHERPA v.2.1.1 using the CT10 set of PDFs. In order to account for the large interference effects between the ADD signal and SM background, discussed in Chapter 1, the ADD signal MC samples are generated in conjunction with the SM diphoton background processes in the high- $m_{\gamma\gamma}$ region considered here. However, it is not computationally feasible to generate the combined signal and background processes together at high orders, so only the LO background processes are considered. Hence, unlike the background calculation, no

additional final state jets are added to the Born diagrams for the signal generation. This yields samples containing the combined ADD signal and SM background processes, which we denote ADD+SM. As explained in Chapter 4, the inclusion of the additional jets primarily affects the sampling of the angular phase space.

Even though the additional final state jets are absent in the ADD+SM samples, what is important is extracting the final ADD signal (which includes the interference effects) in an appropriate manner. This is done by generating separate SM-only background samples that are identical to the ADD+SM samples, but exclude the ADD signal contribution. These SM-only samples allow us to subtract the SM background from the ADD+SM samples yielding the final samples containing only the contributions from the signal and the interference effects, which are needed for the limit setting procedure discussed in Chapter 7. When validating this approach, some samples were able to be produced with the inclusion of one additional jet. The $m_{\gamma\gamma}$ shapes of the final distributions using one additional jet matched those produced using no additional jets.

The implementation of the ADD model within SHERPA [40] is parameterized by the string cutoff scale M_S and the number of extra dimensions n_{ED} . In this implementation, the cutoff scale M_S is related to the fundamental Planck scale M_D by

$$M_S = 2\sqrt{\pi} [\Gamma(n_{\text{ED}}/2)]^{1/(n_{\text{ED}}+2)} M_D \quad (5.1)$$

where Γ is the gamma function. Since the cross section is cutoff at M_S , the $m_{\gamma\gamma}$ distribution is truncated at the chosen value of M_S during signal generation. This conservative estimate sets the cross section to zero above $m_{\gamma\gamma}$. Each of the three cutoff conventions, GRW, HLZ, and Hewett, as discussed in Chapter 1, are separately considered. Note, another common choice for the cutoff scale is the mass scale Λ_T , which can be related to M_S . For example, in the GRW convention, $M_S = \Lambda_T$ for virtual graviton exchange processes, while for real graviton emission processes $M_S = M_D$.

These different conventions are implemented in SHERPA and can be selected by choosing an appropriate value for the internal parameter KK . Using $\text{KK} = 0$, the constant mass parameter, the contributions from the ADD signal during generation are effectively turned off, leaving just the SM background without any interference. Both $\text{KK} = 1$ and $\text{KK} = 2$ correspond to the HLZ convention with $\text{KK} = 1$ being the “simplified” sum and $\text{KK} = 2$ being the “exact” sum of the internal KK propagators. The simplified sum is a LO approximation and is specified in the corresponding term in Eq. (1.10). This form is typically quoted in the literature, including previous CMS searches, and is considered in this analysis. The Hewett convention offers both positive ($\mathcal{F} = +\frac{2}{\pi}$) and negative ($\mathcal{F} = -\frac{2}{\pi}$) interference scenarios. The value $\text{KK} = 3$ is used to select Hewett with positive interference (Hewett+) and $\text{KK} = 4$ is used for Hewett with negative interference (Hewett-). The GRW convention is specified using $\text{KK} = 5$. These are summarized as follows:

- $\text{KK} = 0 \rightarrow$ constant mass;
- $\text{KK} = 1 \rightarrow$ HLZ (simplified);
- $\text{KK} = 2 \rightarrow$ HLZ (exact);
- $\text{KK} = 3 \rightarrow$ Hewett+;
- $\text{KK} = 4 \rightarrow$ Hewett-; and
- $\text{KK} = 5 \rightarrow$ GRW.

5.2 The ADD Convention Relations

As an illustration, the signal was generated for Hewett+ using $M_S = 4000$ GeV and its diphoton invariant mass $m_{\gamma\gamma}$ distribution is shown in Fig. 5.1. This distribution includes the total combined signal, background, and interference components. Notice $m_{\gamma\gamma}$ is cut off at $M_S = 4000$ GeV. This is compared against the invariant mass from only the SM background

(also truncated at $m_{\gamma\gamma} = 4000$ GeV for illustration). Note, to improve statistics, the generation was done in $m_{\gamma\gamma}$ bins of 500-1000, 1000-2000, and 2000-4000 GeV, but some statistical fluctuations can still be observed near the bin edges. In this case, the SM background was generated by specifying $KK = 0$. Both distributions are normalized to have cross sections of 1 fb^{-1} . The nonresonant ADD signal enhancement over SM background is evident at high mass.

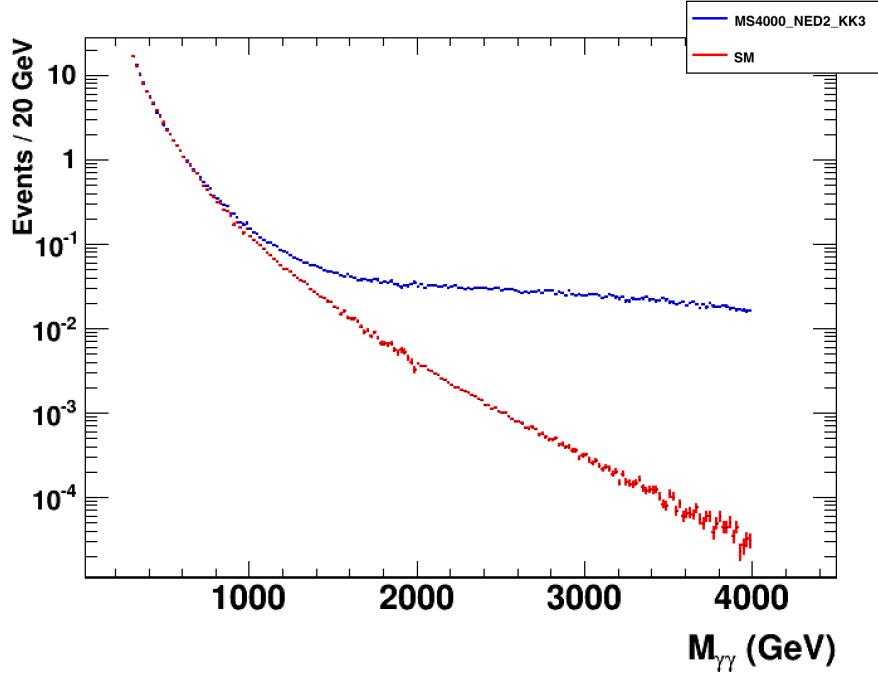


Figure 5.1: The $m_{\gamma\gamma}$ distributions for the ADD signal using the Hewett+ convention with $M_S = 4000$ GeV (blue) compared to the SM background (red).

Fig. 5.2 compares photon η of the combined ADD+SM distributions from the GRW convention at values of $M_S = 2000, 4000$, and 6000 GeV against that of only the SM background. The signal compared to background tends to be peaked in the central region of η . This allows for more sensitivity in the barrel region of the detector over the endcap, in particular, in the EB. We see for larger values of M_S , the signal appears more SM like, as expected since the SM background is recovered as $M_S \rightarrow \infty$.

In general, the diphoton invariant mass distributions from different conventions are expected to differ. For example, Fig. 5.3 compares the distributions between Hewett+ and

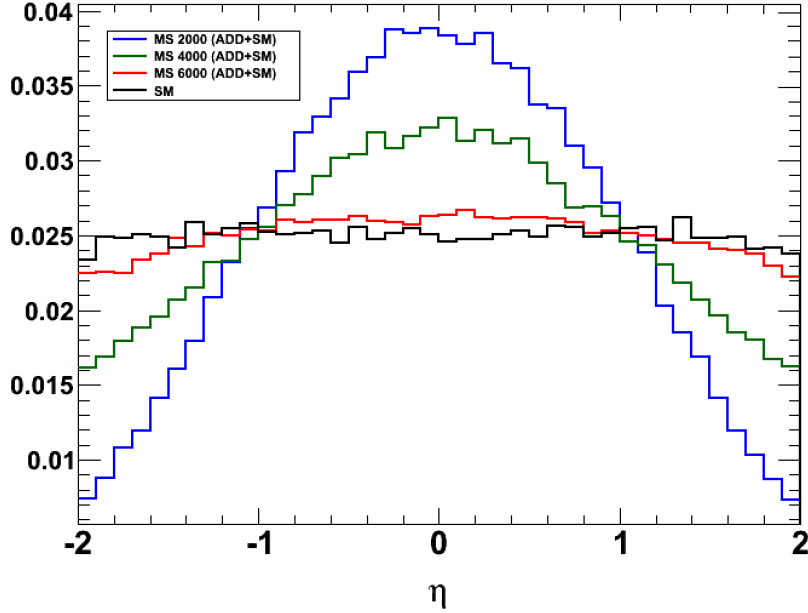


Figure 5.2: Combined ADD+SM signal distributions in photon η compared against the corresponding SM background distribution.

Hewett- using $M_S = 4000$ GeV. The negative interference gives rise to a pronounced valley in the distribution. Fig. 5.4 shows a comparison of the invariant mass distributions within the HLZ convention using $M_S = 2500$ GeV for different values of n_{ED} . The shape of the distribution using $n_{\text{ED}} = 2$ differs from the others due to the different form of $\mathcal{F}$ as shown in Eq. (1.10). The distributions for $n_{\text{ED}} > 2$ are more signal like for lower values of n_{ED} and become more background-like at higher values.

In some cases, the parameter $\mathcal{F}$ is the same between different conventions. This can be observed from Eq. (1.10) and allows for a relation among different conventions. For example, the value of $\mathcal{F}$ using the HLZ convention with $n_{\text{ED}} = 4$ equals that of GRW, so these two conventions should be equivalent. This is demonstrated in Fig. 5.5 (left) by comparing the diphoton invariant mass distributions from each convention. Notice the value of η_G is the same for both signals. In addition, the GRW and Hewett conventions do not depend on n_{ED} . We verify that this is in fact the case in the SHERPA implementation by comparing the $m_{\gamma\gamma}$ spectra using GRW with $n_{\text{ED}} = 2$ and $n_{\text{ED}} = 4$ in Fig. 5.5 (right).

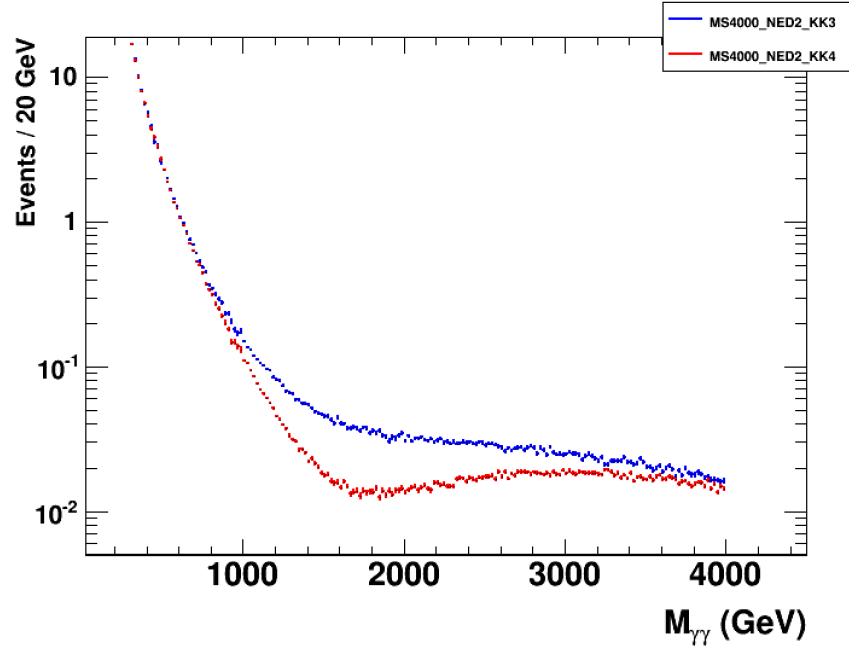


Figure 5.3: Comparison of the $m_{\gamma\gamma}$ spectra between the Hewett+ (blue) and Hewett- (red) conventions using $M_S = 4000$ GeV.

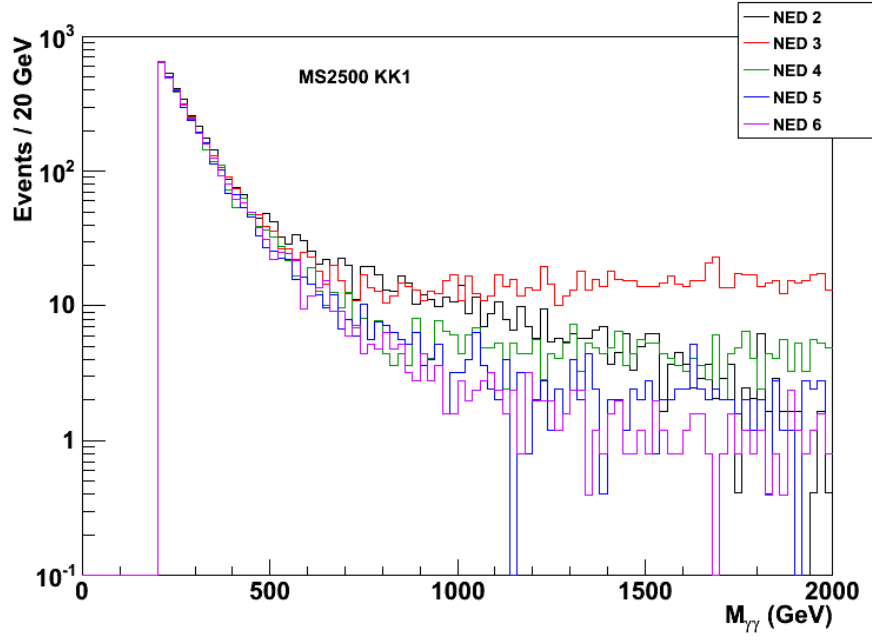


Figure 5.4: The $m_{\gamma\gamma}$ spectra for different signals from the HLZ convention using $M_S = 2500$ GeV for different values of n_{ED} .

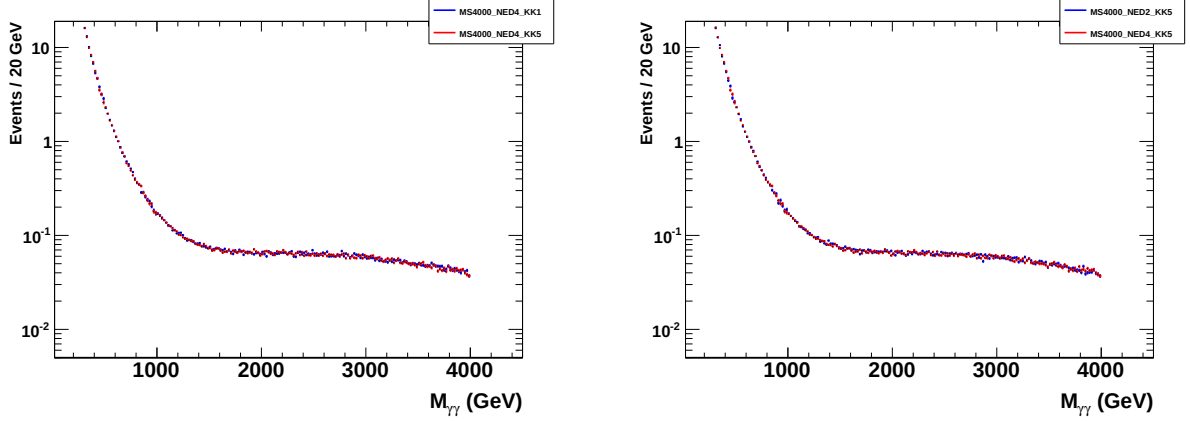


Figure 5.5: The $m_{\gamma\gamma}$ distributions demonstrating that the HLZ convention using $n_{\text{ED}} = 4$ (blue) is equivalent to the GRW convention (red), both for $M_S = 4000$ GeV (right), and those from the GRW convention using $M_S = 4000$ GeV for $n_{\text{ED}} = 2$ (blue) and 4 (red), demonstrating the equivalence of different n_{ED} values (left).

Other conventions can be related to each other through the η_G parameter. This is because the diphoton invariant mass distributions from different conventions are equivalent if they have the same value of η_G . Essentially this means an ADD model point with a particular M_S value in one convention produces the same physical diphoton invariant mass distribution as a specific model point using a different M_S value in another convention; the precise points which will have equivalent spectra are ones which share the same value for η_G . For example, the diphoton spectrum obtained using the Hewett+ convention with $M_S = 4000$ GeV is equivalent to the spectrum obtained for the GRW convention but with $M_S = 4478$ GeV, as illustrated in Fig. 5.6 (left). Using η_G , this arises through the relation $M_S^{\text{GRW}} = (\pi/2)^{1/4} M_S^{\text{Hewett+}}$. Similarly, we can relate the HLZ conventions for $n_{\text{ED}} > 2$ among each other. (The $\mathcal{F}$ parameter for HLZ with $n_{\text{ED}} = 2$ involves $\hat{s}$ and cannot be related to the other conventions.) For example, the HLZ convention with $n_{\text{ED}} = 6$ is related to the HLZ convention using $n_{\text{ED}} = 3$ through the η_G parameter by $M_S^{\text{HLZ}(n_{\text{ED}}=3)} = 4^{1/4} M_S^{\text{HLZ}(n_{\text{ED}}=6)}$. Therefore, the diphoton invariant mass distributions from the HLZ convention with $n_{\text{ED}} = 6$ using $M_S = 4000$ GeV and the HLZ convention with $n_{\text{ED}} = 3$ using $M_S = 5657$ GeV are equivalent, as shown in Fig. 5.6 (right).

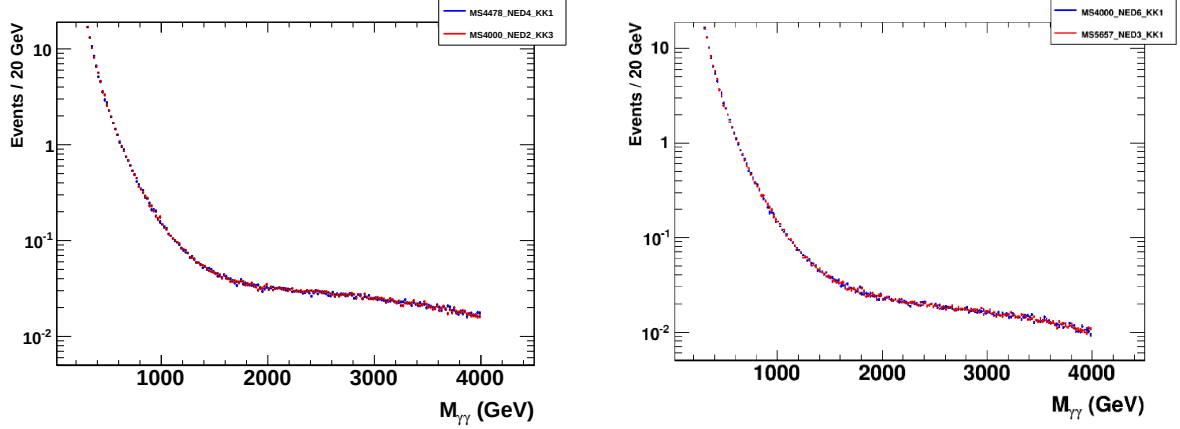


Figure 5.6: The $m_{\gamma\gamma}$ distributions for the Hewett+ convention (red) using $M_S = 4000$ GeV and the GRW convention (blue) using $M_S = 4478$ GeV (left). The $m_{\gamma\gamma}$ distributions for the HLZ convention (blue) using $n_{\text{ED}} = 6$ with $M_S = 4000$ GeV and for the HLZ convention (red) using $n_{\text{ED}} = 3$ with $M_S = 5657$ GeV (right). In each plot, the two conventions have the same value of η_G and their spectra are shown to be equivalent.

5.3 Signal Samples

For this analysis, we consider the ADD parameter space spanning $M_S = 3, 3.5, 4, \dots, 6, 7, \dots, 11$ TeV for each of the following conventions: GRW; HLZ assuming $n_{\text{ED}} = 2, 3, \dots, 7$; Hewett+; and Hewett-. However, the equivalence of signal points demonstrated using the η_G parameterization allows us to significantly reduce the number of MC samples required to span this space for signal generation. Instead of generating all these convention possibilities, consistent with Eq. (1.10), we only need to generate events for the following conventions:

- HLZ (KK = 1) assuming $n_{\text{ED}} = 4$, which has constant positive interference;
- HLZ (KK = 1) assuming $n_{\text{ED}} = 2$, which has positive interference with $\hat{s}$ dependence;
- Hewett- (KK = 4), which has negative interference.

All other conventions listed above have constant positive interference and can be inferred from the samples generated using HLZ assuming $n_{\text{ED}} = 4$. The negative interference and $\hat{s}$ dependence among the conventions needs to be treated separately.

Each of these three conventions requires a separate MC sample for each value of M_S . The example SHERPA run card in Appendix A.2 shows the configuration for HLZ with $n_{\text{ED}} = 4$ using $M_S = 3500$ GeV. However, to improve the statistics for sampling the phase space, each sample is broken into several diphoton invariant mass $m_{\gamma\gamma}$ bins. The bin ranges depend on the particular choice of parameters but are typically 200-500, 500-1000, 1000-2000, 2000-4000, and 4000- M_S GeV, where the last bin is cut off at $m_{\gamma\gamma} = M_S$. Further restrictions are imposed on the phase space for each sample: the kinematic selections requiring photon $p_T > 70$ GeV and photon $|\eta| < 2.8$. The MC samples are generated for pp collisions at $\sqrt{s} = 13$ TeV. The interactions of the generated particles with the CMS detector are simulated using GEANT4, which includes the effects of pileup.

Each sample is prepared and validated privately before internal requests within the CMS Collaboration are submitted for production of the fully simulated samples using the 2016 CMS detector conditions approved for collaboration-wide use. An example, internal dataset path for the MC sample corresponding to the model point for HLZ with $n_{\text{ED}} = 4$ using $M_S = 3500$ GeV in the bin $m_{\gamma\gamma} = 2000\text{--}3000$ GeV has the form `/ADDGravToGG_MS-3500_NED-4_KK-1_M-2000To3000_13TeV-sherpa/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6/MINIAODSIM`. Of its three parts, the first specifies the model point, the second indicates the internal MC campaign used to reproduce the 2016 CMS detector conditions, and the final part signifies the custom data format used within the collaboration. Each sample contains 100,000 fully simulated events. Table 5.1 provides a summary of all the generated ADD samples and their associated cross sections. In total, 122 ADD samples were produced totaling 12,200,000 events.

The LO SM background samples and cross sections used with the above LO ADD+SM samples are tabulated in Table 5.2. Like the ADD+SM samples, these are produced in $m_{\gamma\gamma}$ bins resulting in 6 samples each of 100,000 events. Only the first part of the CMS dataset path is listed as the second and third parts are the same as above. Appendix A.3 shows an example SHERPA run card used to generate this background.

Table 5.1: Cross sections (in units of fb) for the combined ADD+SM samples used in the analysis. For $M_S = 3\text{-}4\text{ TeV}$, the last $m_{\gamma\gamma}$ bin is from $2\text{-}M_S\text{ TeV}$, i.e., there is no bin splitting into $2\text{-}X\text{ TeV}$ and $X\text{-}M_S\text{ TeV}$. For $M_S = 4.5\text{ or }5\text{ TeV}$, the bin edge is at $X = 3\text{ TeV}$. For $M_S > 5\text{ TeV}$, $X = 4\text{ TeV}$.

M_S (TeV)	KK convention	0.2-0.5 TeV	0.5-1 TeV	1-2 TeV	2- X TeV	$X\text{-}M_S$ TeV
3	HLZ ($n_{\text{ED}} = 2$)	-	139.7	91.39	11.6	-
	HLZ ($n_{\text{ED}} = 4$)	898.5	89.66	51.64	55.18	-
	Hewett-	-	73.31	17.52	20.2	-
3.5	HLZ ($n_{\text{ED}} = 2$)	-	102.6	45.1	9.442	-
	HLZ ($n_{\text{ED}} = 4$)	892.4	81.04	21.26	23.65	-
	Hewett-	-	73.98	8.453	8.239	-
4	HLZ ($n_{\text{ED}} = 2$)	-	89.81	25.53	6.604	-
	HLZ ($n_{\text{ED}} = 4$)	889.4	78.21	12.34	10.5	-
	Hewett-	-	74.27	6.494	3.474	-
4.5	HLZ ($n_{\text{ED}} = 2$)	-	83.73	16.61	3.809	0.6004
	HLZ ($n_{\text{ED}} = 4$)	896.6	77.03	9.129	2.688	2.27
	Hewett-	-	74.85	6.033	0.7721	0.7878
5	HLZ ($n_{\text{ED}} = 2$)	-	80.25	12.21	2.423	0.5167
	HLZ ($n_{\text{ED}} = 4$)	894	76.4	7.829	1.382	1.193
	Hewett-	-	74.71	5.999	0.3925	0.3902
5.5	HLZ ($n_{\text{ED}} = 2$)	-	79.11	9.948	1.969	0.04542
	HLZ ($n_{\text{ED}} = 4$)	894.7	75.99	7.209	1.218	0.2419
	Hewett-	-	75.22	6.047	0.3688	0.08207
6	HLZ ($n_{\text{ED}} = 2$)	-	78.19	8.717	1.382	0.0446
	HLZ ($n_{\text{ED}} = 4$)	900.4	75.98	6.887	0.7727	0.1393
	Hewett-	-	75.35	6.09	0.264	0.04472
7	HLZ ($n_{\text{ED}} = 2$)	-	77.18	7.501	0.7842	0.03281
	HLZ ($n_{\text{ED}} = 4$)	-	76.39	6.598	0.4351	0.04842
8	HLZ ($n_{\text{ED}} = 2$)	-	76.27	6.941	0.5179	0.0214
	HLZ ($n_{\text{ED}} = 4$)	-	76.23	6.437	0.3247	0.01974
9	HLZ ($n_{\text{ED}} = 2$)	-	76.27	6.734	0.3959	0.01394
	HLZ ($n_{\text{ED}} = 4$)	-	75.69	6.386	0.2802	0.009672
10	HLZ ($n_{\text{ED}} = 2$)	-	76.08	6.554	0.3313	0.009366
	HLZ ($n_{\text{ED}} = 4$)	-	75.89	6.377	0.259	0.005716
11	HLZ ($n_{\text{ED}} = 2$)	-	75.97	6.527	0.2997	0.006698
	HLZ ($n_{\text{ED}} = 4$)	-	75.59	6.361	0.2504	0.003906

Table 5.2: The LO SM background MC samples. The full, internal CMS path name is specified by appending the dataset path with `/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TracheIV_v6/MINIAODSIM`

Dataset path	Cross section (pb)
/GG_M-200To500_Pt-70_13TeV-sherpa	8.923e-01
/GG_M-500To1000_Pt-70_13TeV-sherpa	7.592e-02
/GG_M-1000To2000_Pt-70_13TeV-sherpa	6.292e-03
/GG_M-2000To4000_Pt-70_13TeV-sherpa	2.315e-04
/GG_M-4000To8000_Pt-70_13TeV-sherpa	1.669e-06
/GG_M-8000To13000_Pt-70_13TeV-sherpa	5.430e-11

The diphoton invariant mass distributions for the three KK conventions considered in this analysis using $M_S = 3, 3.5, 4, \dots, 6$ TeV are shown in Fig. 5.7 (left). Notice the spectra are truncated at their corresponding value of M_S . After subtraction of the SM background, the resulting $m_{\gamma\gamma}$ distributions are shown in Fig. 5.7 (right), for each of these three KK conventions. These are the final $m_{\gamma\gamma}$ distributions containing only the ADD signal and interference between the ADD signal and the SM background. Similar distributions are shown in Fig. 5.8 for the higher values of $M_S = 6, 7, 8, 9$, and 10 TeV using HLZ assuming $n_{\text{ED}} = 4$ compared against the distribution from HLZ with $n_{\text{ED}} = 2$ and $M_S = 8$ TeV, shown in the EBEB and EBEE categories. For these values of M_S , the low-mass region remains dominated by the SM background, while the high-mass region is signal dominated.

These SM background-subtracted ADD signal distributions represent the additional effect of the large extra-dimensional signal and are used in the limit setting procedure, as described in Section 7.2.

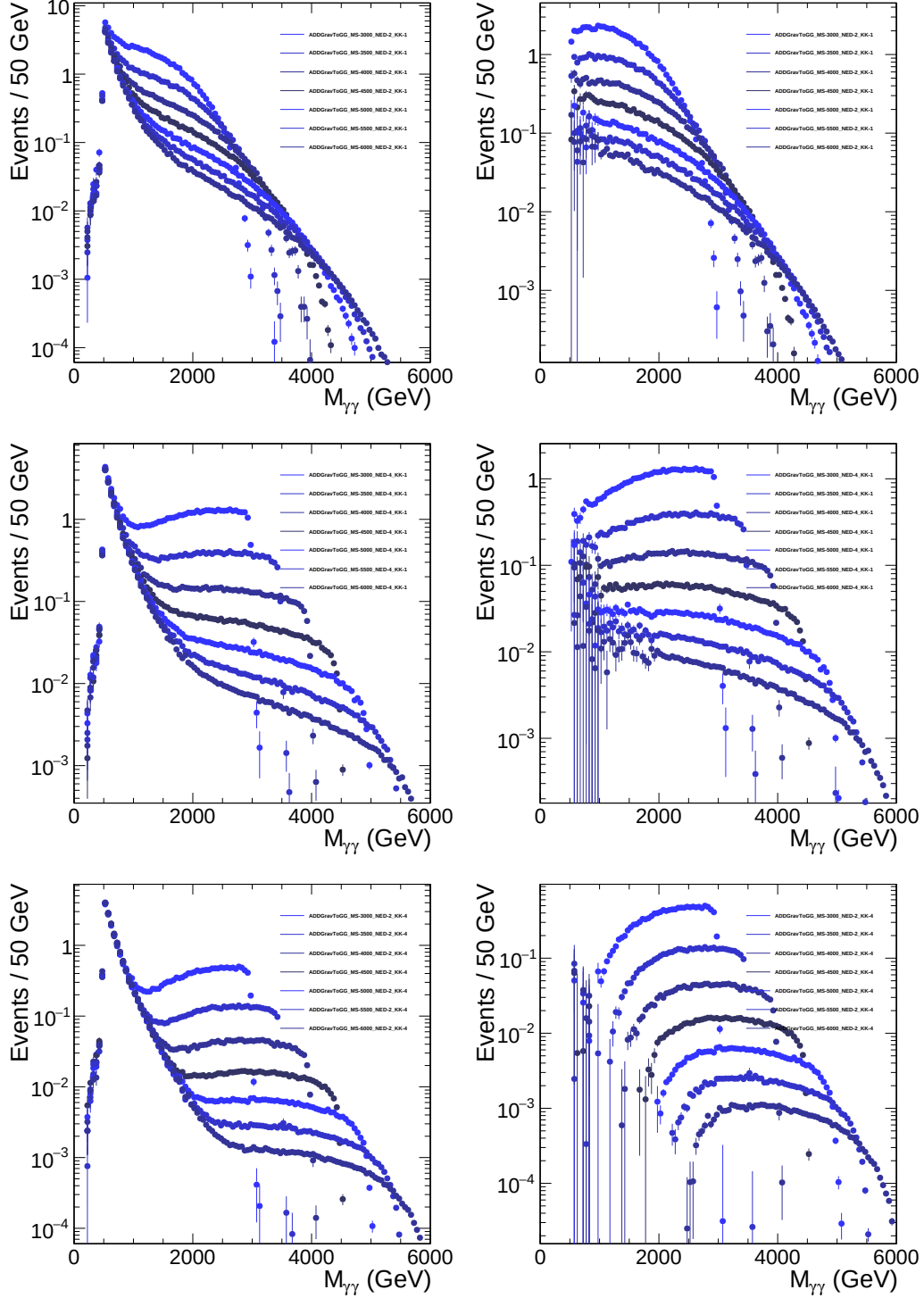


Figure 5.7: The $m_{\gamma\gamma}$ distribution in signal events for the generated values of M_S before (left) and after (right) subtraction of the SM background. The KK conventions used in the plots are HLZ with $n_{\text{ED}} = 2$ (top), HLZ assuming $n_{\text{ED}} = 4$ (middle), and Hewett- (bottom).

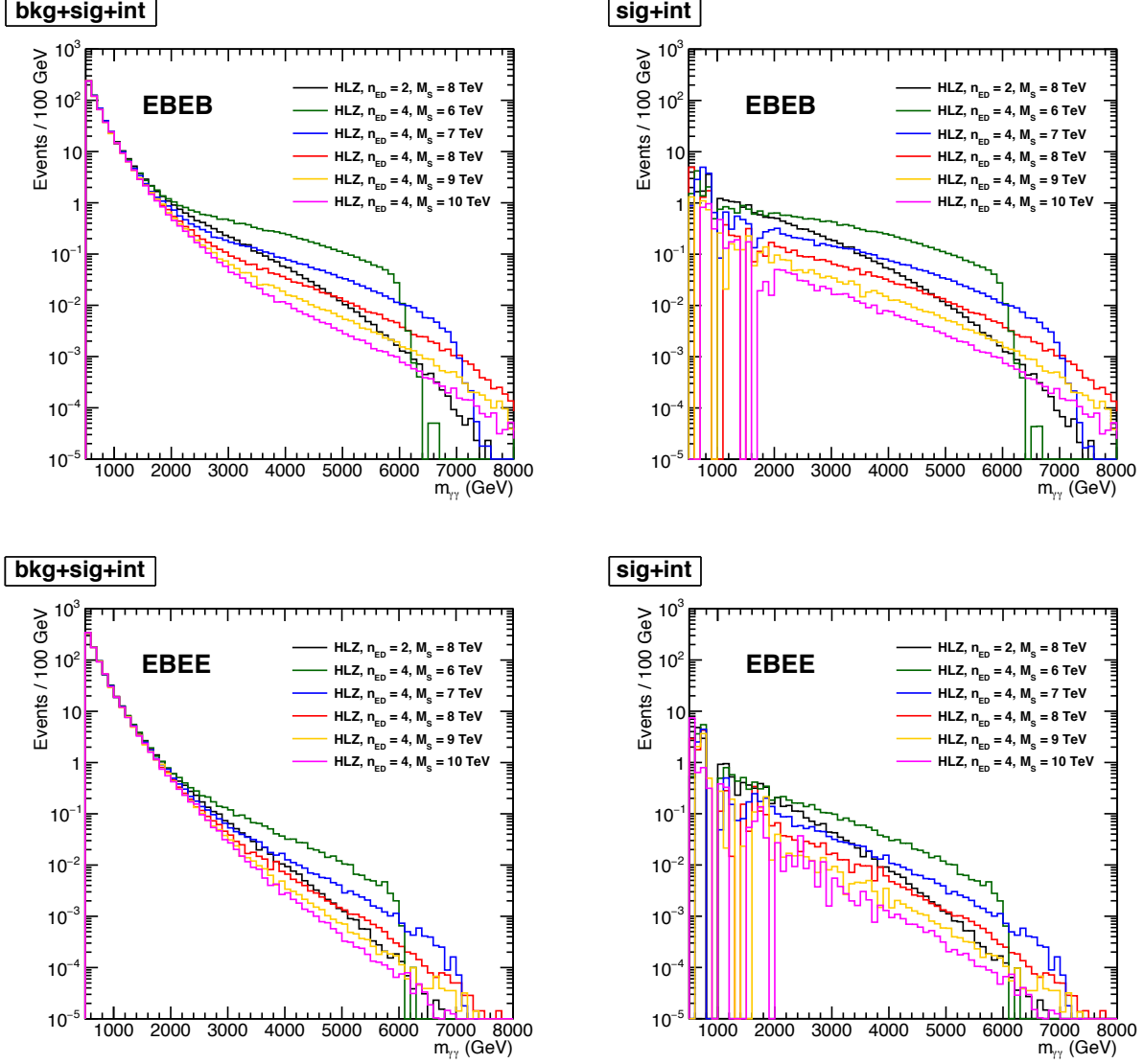


Figure 5.8: The $m_{\gamma\gamma}$ distributions for the ADD signal and interference before (left) and after (right) subtraction of the SM background in the EBEB (top) and EBEE (bottom) categories. The spectra correspond to the KK conventions using HLZ with $n_{\text{ED}} = 2$ and $M_S = 8$ TeV (black) compared against HLZ with $n_{\text{ED}} = 4$ using $M_S = 6, 7, 8, 9$, and 10 TeV.

Chapter 6

Sources of Systematic Uncertainty

Sources of systematic uncertainty associated with the real background prediction, fake background measurement, and signal calculation are considered. These uncertainties either affect the shape or normalization of the corresponding distribution and are applied to the final selection.

6.1 Real Background Systematic Uncertainties

The SM diphoton background prediction is performed at NNLO accuracy, but higher order terms could still contribute. To account for this, we allow the normalization of this predicted background to float freely, constrained only by the data. The fit is primarily constrained at low $m_{\gamma\gamma}$, where the statistical uncertainty on the data is the smallest. Potential signals are expected in the high- $m_{\gamma\gamma}$ region and, provided the signal and background shapes differ significantly, this technique will discriminate between the two, as is the case in this analysis. In addition, allowing the background normalization to float will absorb the systematic uncertainties associated with the integrated luminosity measurement and inefficiencies from the trigger and high- p_T photon ID selection. These uncertainties are small and must be applied to the signal calculation, since its normalization is not allowed to float freely, as described below. The details regarding how the fit is performed will be presented in Chapter 7.

Systematic uncertainties on the shape of the SM background prediction are determined

by varying the scales used in the K factor calculation, separately for the EBEB and EBEE regions. As described in Section 4.1.2, the renormalization, factorization, and fragmentation scales are varied simultaneously between $m_{\gamma\gamma}/2$ and $2m_{\gamma\gamma}$, from their default value of $m_{\gamma\gamma}$. Within the MCFM framework, these scales affect the differential cross section via an estimate of the truncation of the perturbative calculation, the transition point between the hard calculation and the PDFs, and the fragmentation functions, respectively. Fig. 6.1 shows each K factor corresponding to the different choice of scales. The uncertainty increases with $m_{\gamma\gamma}$ and, in general, is larger in the EBEE region.

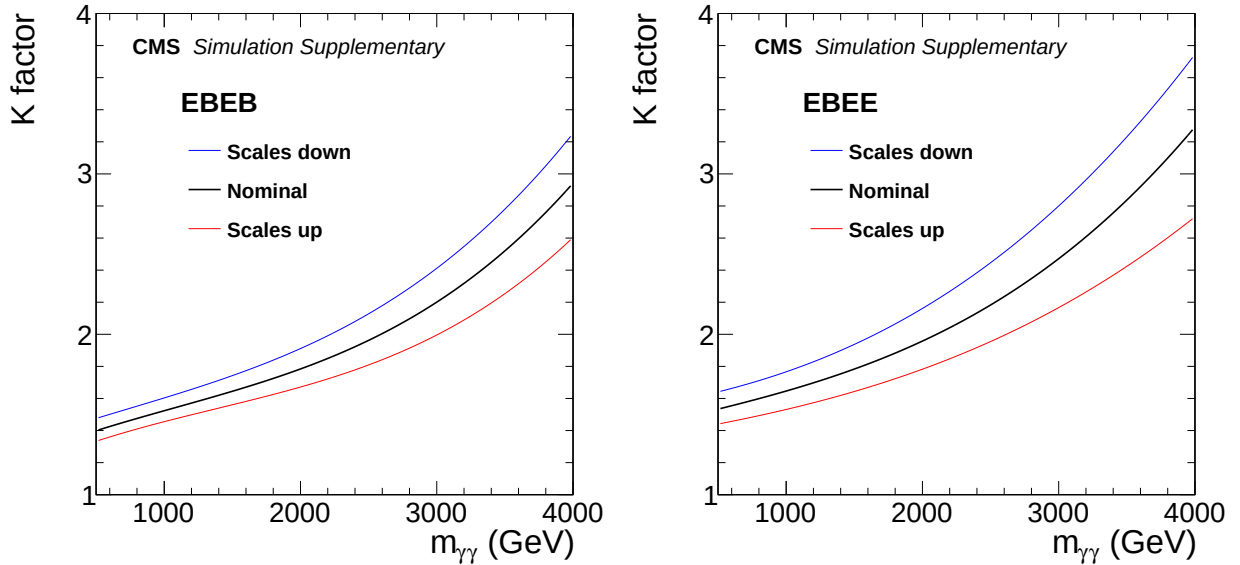


Figure 6.1: The K factor as a function of $m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories. The K factor is defined as the ratio of the predicted MCFM NNLO $pp \rightarrow \gamma\gamma$ cross section to that of SHERPA. The renormalization and factorization scales have been set to $m_{\gamma\gamma}$ (black) and varied simultaneously by factors of 0.5 (blue) and 2.0 (red).

Systematic uncertainties on the SM background shape associated with the PDFs are calculated using the fixed-order, parton-level MC program DIPHOX [128], at NLO with a consistent set of CT10 NLO PDFs as used by MCFM. DIPHOX uses fragmentation functions with a cone-based photon isolation, meaning that the isolation energy within the cone is less than a constant value (5 GeV), yielding results consistent with the Frixione isolation configured in MCFM. DIPHOX only calculates up to NLO and is used primarily for computational

convenience. Checks were performed to ensure that DIPHOX at NLO agreed with MCFM at NLO. The CT10 NLO set of PDFs contains 26 eigenvector pairs, each of which are individually varied by ± 1 standard deviation around its nominal value. Each eigenvector is treated as a separate systematic uncertainty, yielding 52 individual systematic uncertainties, all fully correlated as a function of $m_{\gamma\gamma}$. For each eigenvector variation, the ratio of the $m_{\gamma\gamma}$ distribution obtained with that variation using DIPHOX at NLO to that of the default distribution using MCFM at NNLO is used to reweight the MCFM SM diphoton background prediction, yielding a shape uncertainty. The variations for each eigenvector pair are shown in Figs. 6.2 and 6.3 for the EBEB and EBEE categories, respectively. In general the variations increase and can become large at high- $m_{\gamma\gamma}$ (above 3 TeV). This is expected since experimental data is used as an input to improve PDF calculations, and the LHC is currently probing the highest energy scales, so is yet to provide any data in these very high regions. As a simplified check of this method, the NNLO cross section was calculated using MCFM following the PDF4LHC prescription with the PDF4LHC15_mc set of PDFs [129]. This technique uses the PDF4LHC master formulas [129], which combine the separate uncertainties. The difference between the NNLO cross sections calculated assuming the CT10 and PDF4LHC15_mc sets of PDFs were found to be within the assigned uncertainty from this method.

Another source of systematic uncertainty considered on the background shape is the difference in shape between the NNLO MCFM $m_{\gamma\gamma}$ spectra and of the shapes predicted at NLO by MCFM and DIPHOX. This is determined by comparing the shapes of the MCFM NNLO and the MCFM NLO predictions (used as the positive variation) and between the MCFM NNLO and DIPHOX NLO predictions (used as the negative variation). This is done to provide a conservative bound on the effects of the absent higher order terms.

Uncertainties due to the photon energy scale and resolution have a negligible impact on this nonresonant search.

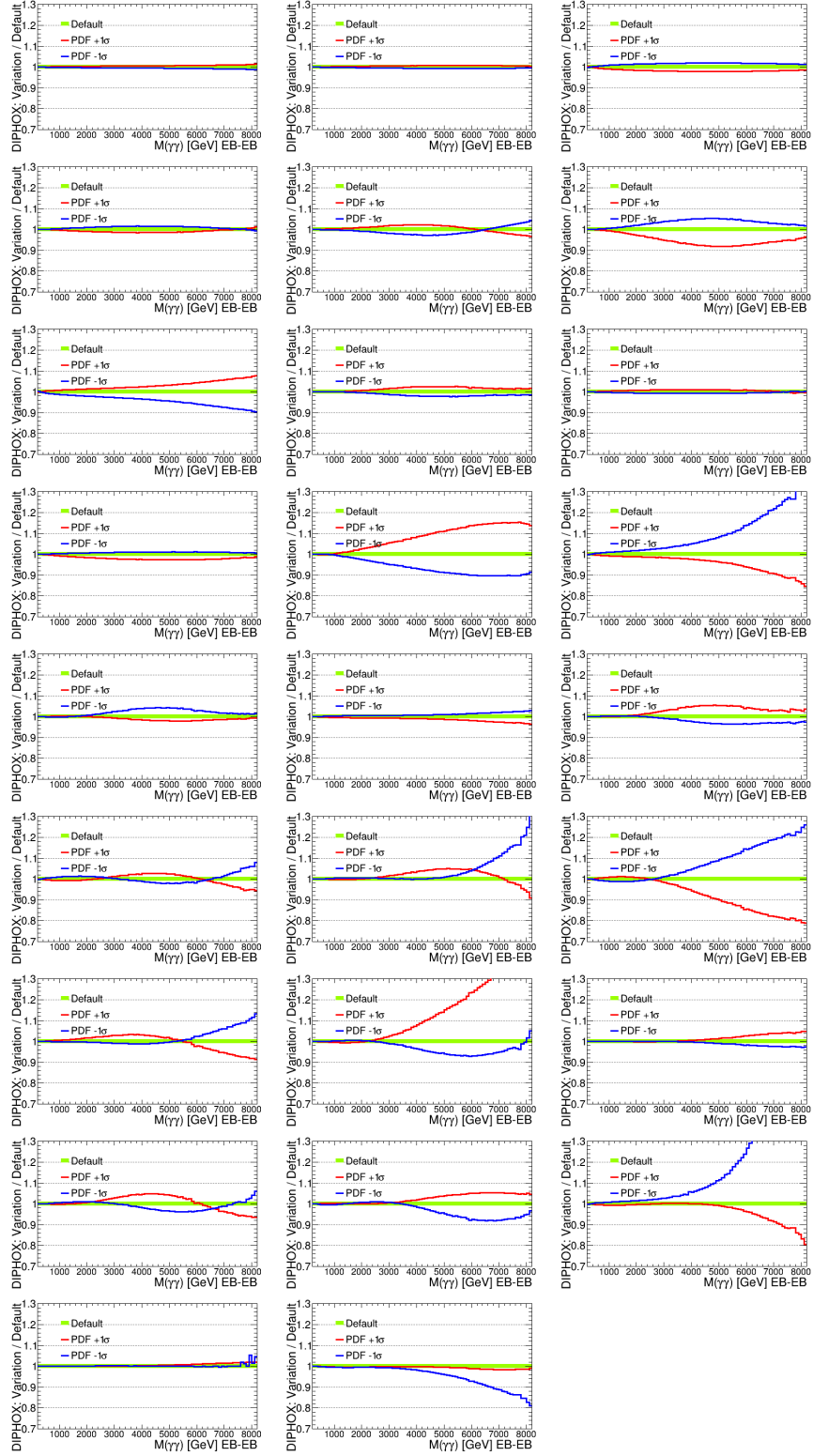


Figure 6.2: The ± 1 standard deviation variations of each of the 26 PDF eigenvectors in the EBEB category. Each distribution is the ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$.

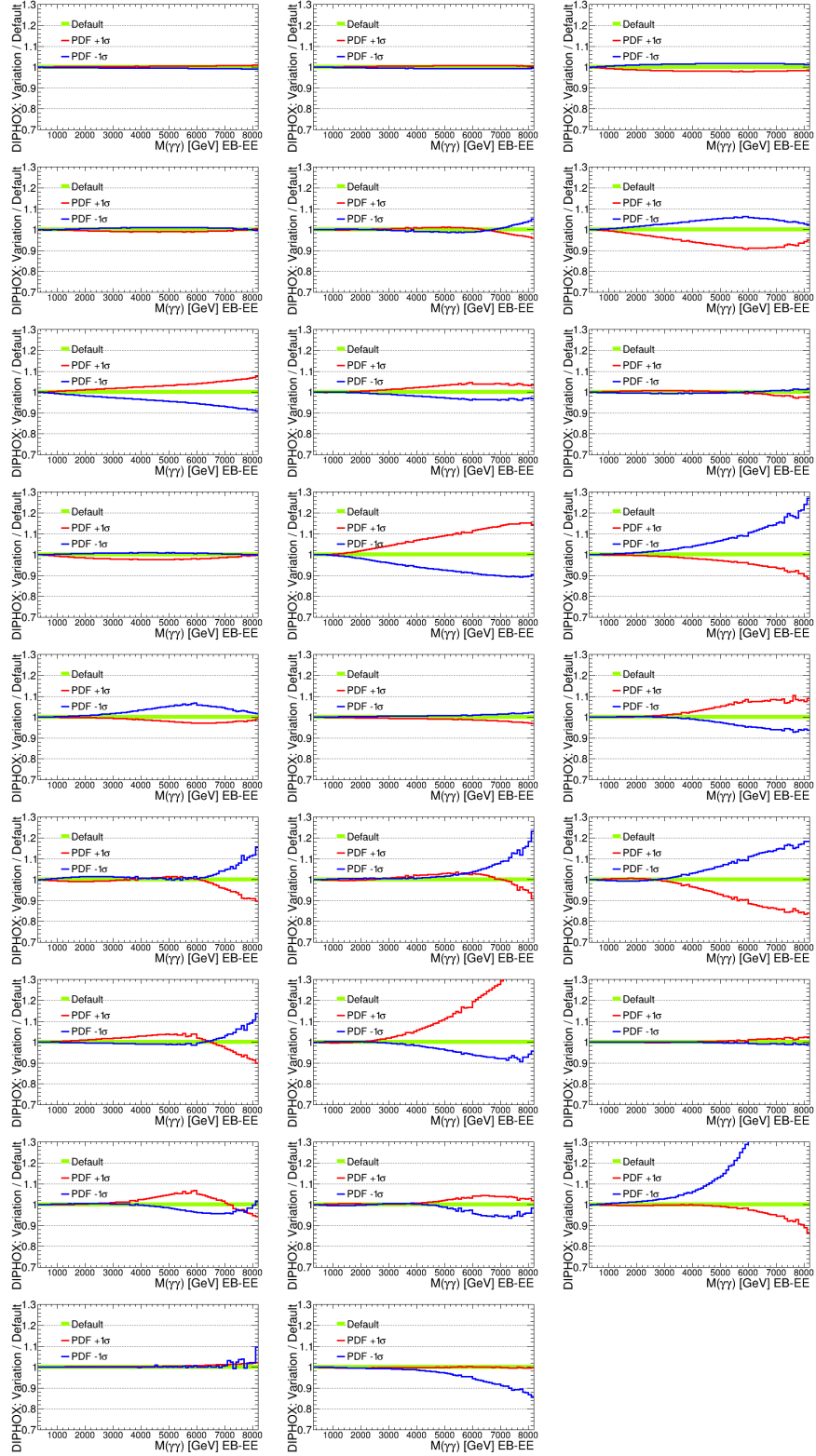


Figure 6.3: The ± 1 standard deviation variations of each of the 26 PDF eigenvectors in the EBEE category. Each distribution is the ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$.

6.2 Fake Background Systematic Uncertainties

A systematic uncertainty of 30% is assigned to the normalization of the fake background measurement from the variation of the fake rate as a function of pileup, photon η , a change of the sideband definition, and a change of template variable. In addition, this covers the degree of variation observed in the fake rate method from the closure test, including any potential real photon contamination. A separate systematic uncertainty on the shape of the fake background is applied to account for the difference in shape between the fake rate measurements constructed using the jet- and muon-triggered datasets. This shape uncertainty makes a subdominant contribution. The details of these studies were presented in Section 4.2.

6.3 Signal Systematic Uncertainties

Systematic uncertainties are assigned to the normalization of the signal calculation to account for the uncertainties in measurements of the integrated luminosity and overall diphoton selection efficiency, corresponding to 2.5% [83] and 6%, respectively. The selection efficiency includes the trigger and high- p_T photon ID efficiency, which are 3% per photon. The PDF uncertainties affecting the signal shape are determined using the same procedure as is done for the background, by separately varying the 26 eigenvalue pairs using DIPHOX at NLO. These PDF uncertainties are assumed to be 100% correlated between the signal and background. The effect of the PDF uncertainty on the signal cross section is considered a theoretical uncertainty and not treated as a systematic uncertainty.

6.4 Summary of the Systematic Uncertainties

A summary of the dominant systematic uncertainties as a function of $m_{\gamma\gamma}$ in the EBEB and EBEE categories is shown in Fig. 6.4. Here, the uncertainties are added in quadrature

and shown as fraction of the total yields. The colored bands represent the dominant pre-fit systematic uncertainties, i.e., before the fit of the background normalization, constrained by the data. The pre-fit systematic uncertainties do not include the normalization of the SM diphoton prediction or the NLO shapes, and, therefore, the black and red curves cannot be directly compared. At high- $m_{\gamma\gamma}$, the PDF uncertainties are dominant. The total post-fit systematic uncertainty is also shown in each category. At high- $m_{\gamma\gamma}$, these are driven upward primarily due to their influence under the shape and normalization of the data. A decomposition of the total post-fit uncertainty into its individual components is limited by the choice of fitting framework used. However, the variation of each post-fit systematic uncertainty will be described and shown in Chapter 7, during the discussion of the fitting procedure.

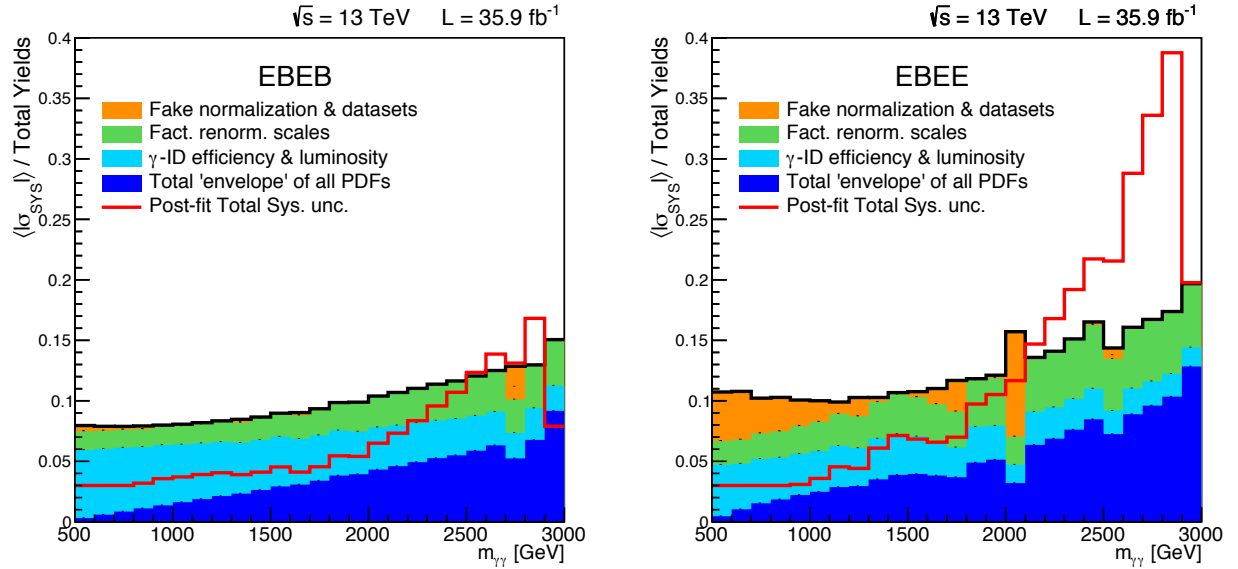


Figure 6.4: The dominant pre-fit systematic uncertainties and total post-fit uncertainty as a function of $m_{\gamma\gamma}$ in the EBEB (left) and EBEE (right) categories.

Chapter 7

Results and Limit Setting

The results of the ADD large extra-dimensional model search are presented. This signal search is for a nonresonant excess of data over the total background prediction. In the absence of a significant excess, limits will be set on the ADD large extra-dimensional model parameters.

7.1 Results of the High-Mass Diphoton Search

The variable $m_{\gamma\gamma}$ is sensitive to the ADD extra-dimensional signal search and is used as the primary discriminating variable. This search initially stayed blind to data above $m_{\gamma\gamma} > 1$ TeV, using the control region $500 < m_{\gamma\gamma} < 1000$ GeV to validate the search procedure. The fully unblinded data is presented.

The kinematic variables associated with individual photons and the diphoton system are shown in Fig. 7.1, for photons in the EB and diphotons in the EBEB category, and Fig. 7.2, for photons in the EE regions and diphotons in the EBEE category. The variables p_T , η , and ϕ are shown separately for the leading (γ_1) and subleading (γ_2) photons. For the diphoton objects, the angle between the photons $\cos(\theta^*)$ and the difference between each photon's η and ϕ coordinates, $\Delta\eta_{\gamma\gamma}$ and $\Delta\phi_{\gamma\gamma}$, respectively, are shown. The comparison presented is between the final data selection compared to the SM background prediction (with K factor applied) and fake background measurement. The background is scaled to the full 2016

data luminosity of 35.9 fb^{-1} . The ratios between the data and total background show good agreement among all variables. There is some discrepancy, primarily for low- p_T photons with high values of η . This effect was observed in simulation, consistent with the kinematic studies performed as part of the closure test to the photon fake rate method. These low- p_T photons in the EE region do not affect the high- $m_{\gamma\gamma}$ signal region.

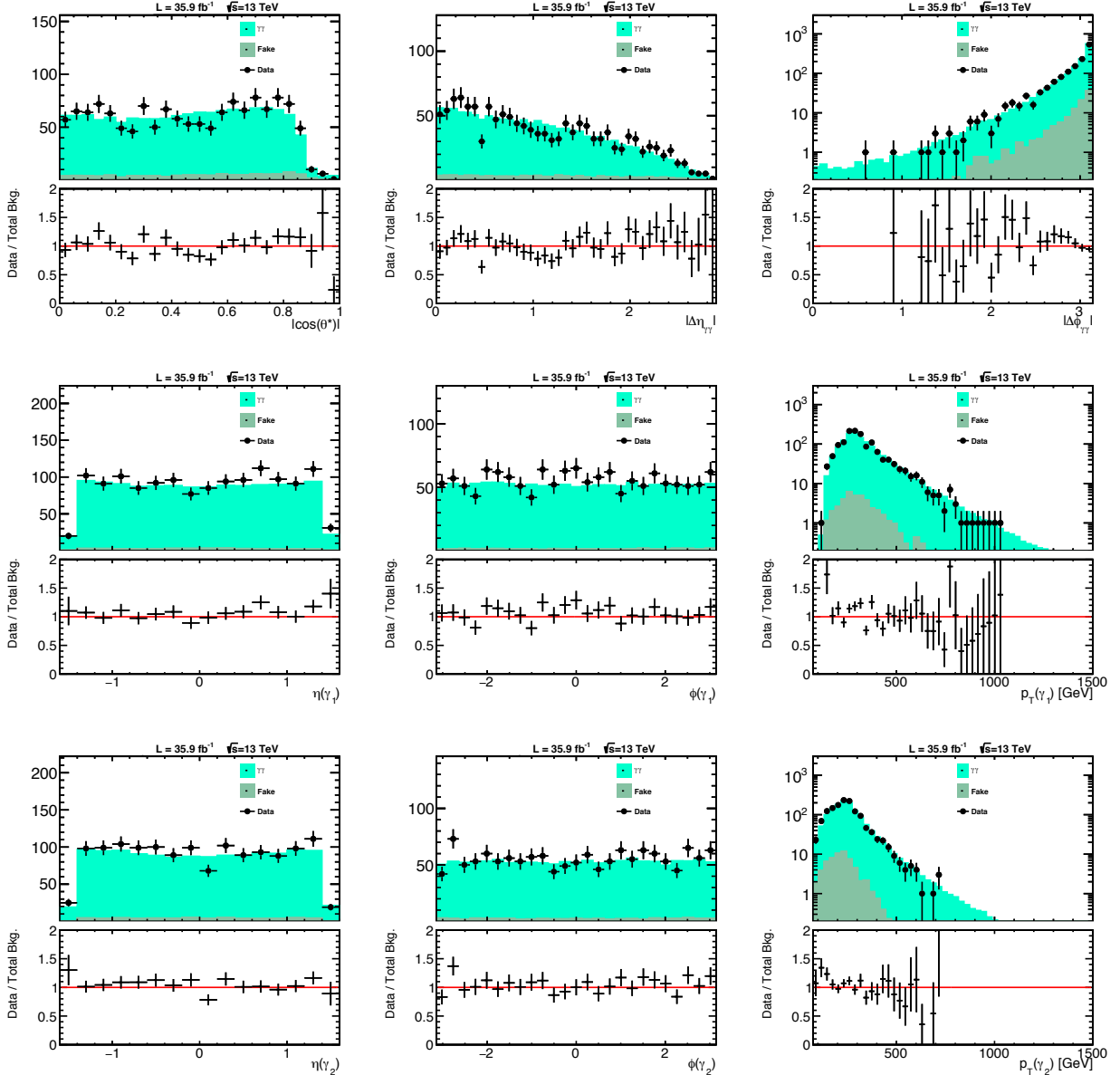


Figure 7.1: A comparison of various kinematic distributions between the SM background prediction, fake background measurement, and data. The photons are restricted to the EB and diphotons to the EBEB.

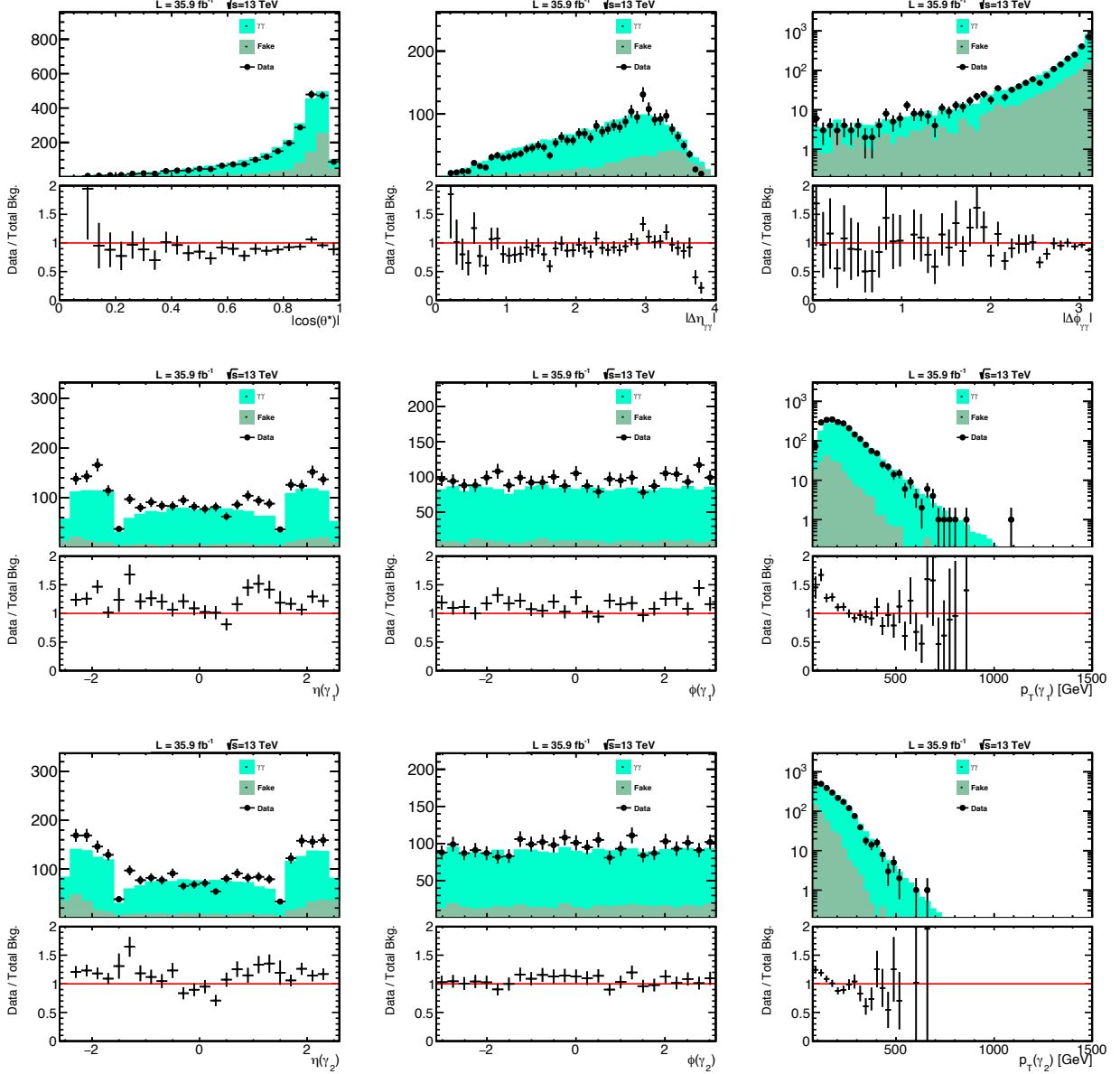


Figure 7.2: A comparison of various kinematic distributions between the SM background prediction, fake background measurement, and data. The photons are restricted to the EE and diphotons to the EBEE.

A list of the highest diphoton invariant mass $m_{\gamma\gamma}$ events is provided in Table 7.1. No events containing photons causing saturation in the ECAL were observed. Fig. 7.3 shows an event display of the highest invariant mass diphoton event recorded in 2016 at 1840 GeV in the EBEB category. The two photons in this event display are represented by the two large, red ECAL energy deposits. The highest diphoton invariant mass event in the EBEE category was recorded at 2444 GeV in 2016.

Table 7.1: The three highest $m_{\gamma\gamma}$ events observed in data for the EBEB and EBEE categories. The photon p_T is given in units of GeV. The dataset paths begin with /DoubleEG and end with /MINIAOD, and include X = 03Feb2017. Here, run, LS, and event correspond to the run number, luminosity section, and event number, respectively. Each event was recorded in 2016, where pp collisions occurred (the majority of the time) from May-October.

	$m_{\gamma\gamma}$ (GeV)	γ_1 (p_T , η_{SC})	γ_2 (p_T , η_{SC})	Run:LS:Event	Dataset path	Date
EBEB	1840.38	(964.15, -0.881)	(429.01, 0.916)	278820:832:1528641148	/Run2016G-X-v1	Aug 14
	1816.31	(833.60, -0.160)	(679.74, -1.424)	276810:249:448659898	/Run2016D-X-v1	Jul 14
	1777.39	(508.64, -1.053)	(451.83, 1.405)	276437:452:753595848	/Run2016D-X-v1	Jul 6
EBEE	2444.38	(415.88, -2.484)	(272.98, 1.442)	279794:445:700510844	/Run2016G-X-v1	Aug 30
	2134.42	(570.82, -0.695)	(515.46, 1.916)	284039:18:31148565	/Run2016H-X_ver3-v1	Oct 26
	1985.68	(577.69, -2.443)	(397.34, 0.273)	276585:21:32975214	/Run2016D-X-v1	Jul 10

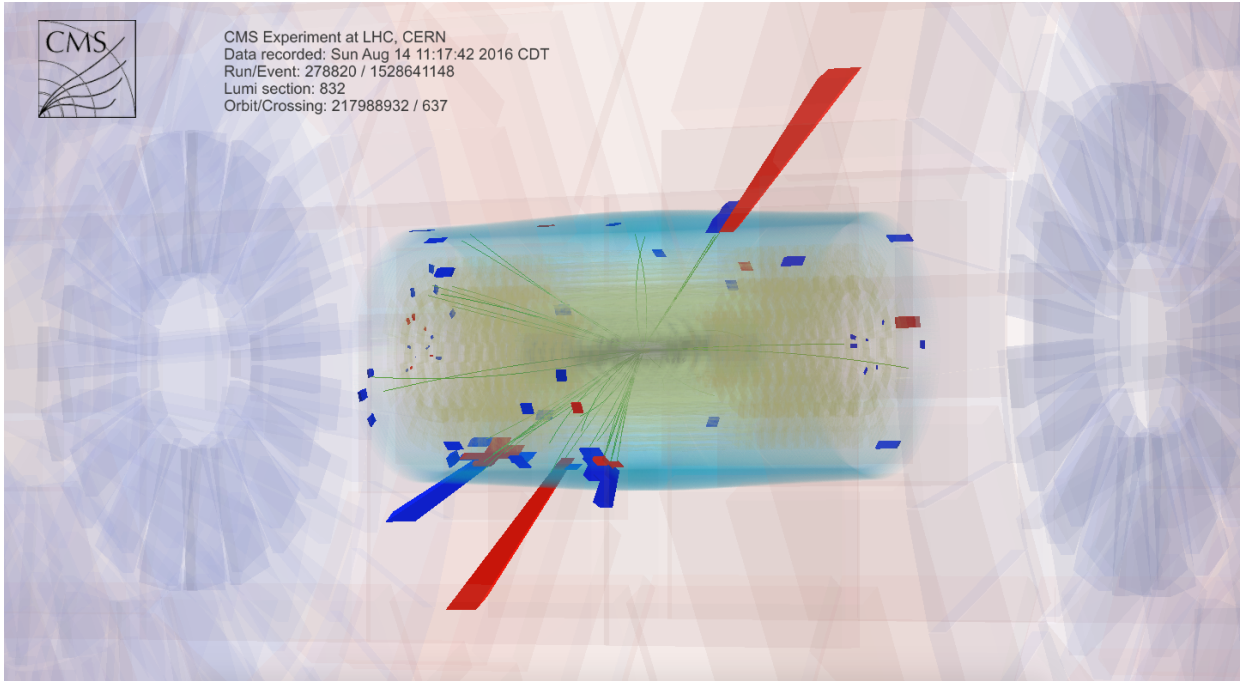


Figure 7.3: An event display of the highest invariant mass diphoton event recorded in 2016 at 1840 GeV in the EBEB category. The two photons are represented by the two large, red ECAL energy deposits.

Further checks were performed on the high-mass diphoton events to ensure good health (e.g., to confirm that there was not large detector noise). The (η, ϕ) spatial distribution of the events with $m_{\gamma\gamma} > 1.5$ TeV is shown in Fig. 7.4, split between the leading and subleading photons in the EBEB and EBEE categories. The angular separation between the two photons is directly proportional to the $m_{\gamma\gamma}$ of their associated diphoton object. Hence, these high- $m_{\gamma\gamma}$ events tend to be back-to-back in the ϕ direction, as seen in the EBEB region. The R_9 and $\sigma_{i\eta i\eta}$ distributions of these events are shown in Fig. 7.5, also split between the leading and subleading photons, but with both categories combined. Also, shown in this plot are the reference distributions obtained using all photons from events with $m_{\gamma\gamma} < 1.5$ TeV. These checks show that the high-mass diphoton events are compatible with the rest of the diphoton events, ensuring that they are genuine.

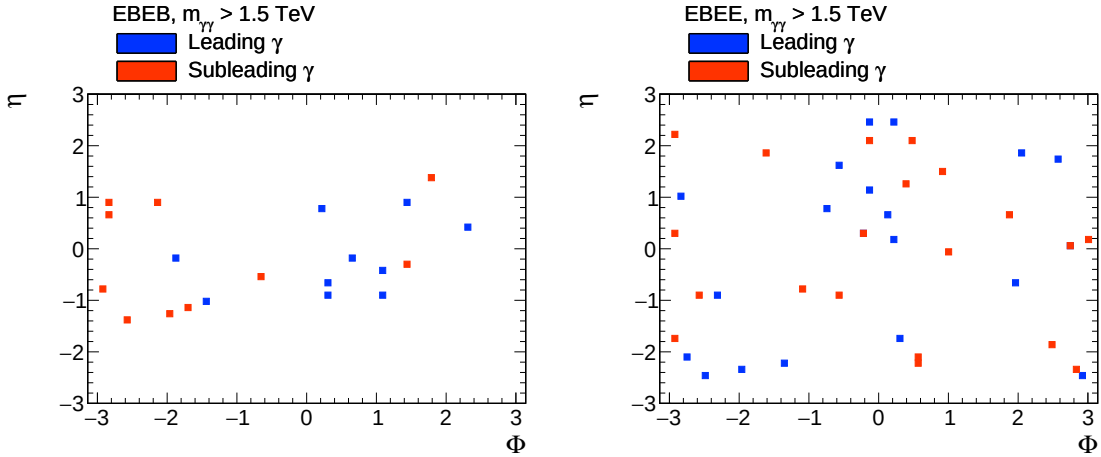


Figure 7.4: The (η, ϕ) spatial distribution of the leading (blue) and subleading (red) photons in the diphoton events with $m_{\gamma\gamma} > 1.5$ TeV for the EBEB (left) and EBEE (right) categories.

The final results are obtained in two stages: using the pre- and post-fit $m_{\gamma\gamma}$ distributions from before and after the background normalization is allowed to float, respectively. These distributions utilize 100 GeV binning, starting at 500 GeV. Fig. 7.6 shows the $m_{\gamma\gamma}$ spectra using the pre-fit SM background prediction and the fake background measurement compared to the data in the EBEB and EBEE categories. The systematic uncertainties neglect the (unbounded) normalization of the diphoton prediction and the NLO shapes.

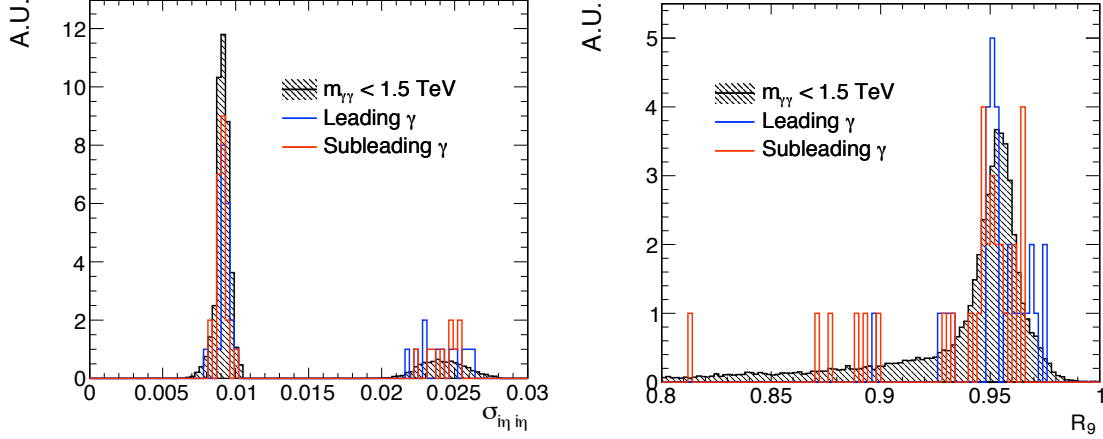


Figure 7.5: The $\sigma_{\eta\eta}$ (left) and R_9 (right) distributions of the leading (blue) and subleading (red) photons in the diphoton events with $m_{\gamma\gamma} > 1.5$ TeV compared to the photons in events with $m_{\gamma\gamma} < 1.5$ TeV. The y -axis is labeled using arbitrary units.

Even though this does not constitute the final background prediction, excellent agreement is found, using the full NNLO background calculation. No significant excess of data over background is observed, as indicated in the bottom panels. These results are consistent with the background-only hypothesis for the signal search.

The final post-fit prediction is performed simultaneously with the limit extraction on the ADD model parameters. A Bayesian statistical approach is adopted and performed by the THETA framework [130]. The systematic uncertainties described in Chapter 6 are treated as nuisance parameters and are marginalized using a Markov chain MC method [131] within THETA. After the marginalization of the nuisance parameters, posterior predictions for the SM background $m_{\gamma\gamma}$ spectra are derived, using the data as a constraint. Since this fit is performed simultaneously with the limit setting procedure, the process will be detailed in the next section, during the limit setting discussion.

Fig. 7.7 shows the final post-fit predictions of the SM background and the fake background measurement compared to the data in the EBEB and EBEE categories. As is the case with the pre-fit prediction, excellent agreement is found within the total uncertainty. As expected, the uncertainties on the post-fit prediction are smaller. The invariant mass distributions from the ADD signal scenario using the HLZ convention assuming $n_{\text{ED}} = 2$ and $M_S = 8$ TeV are

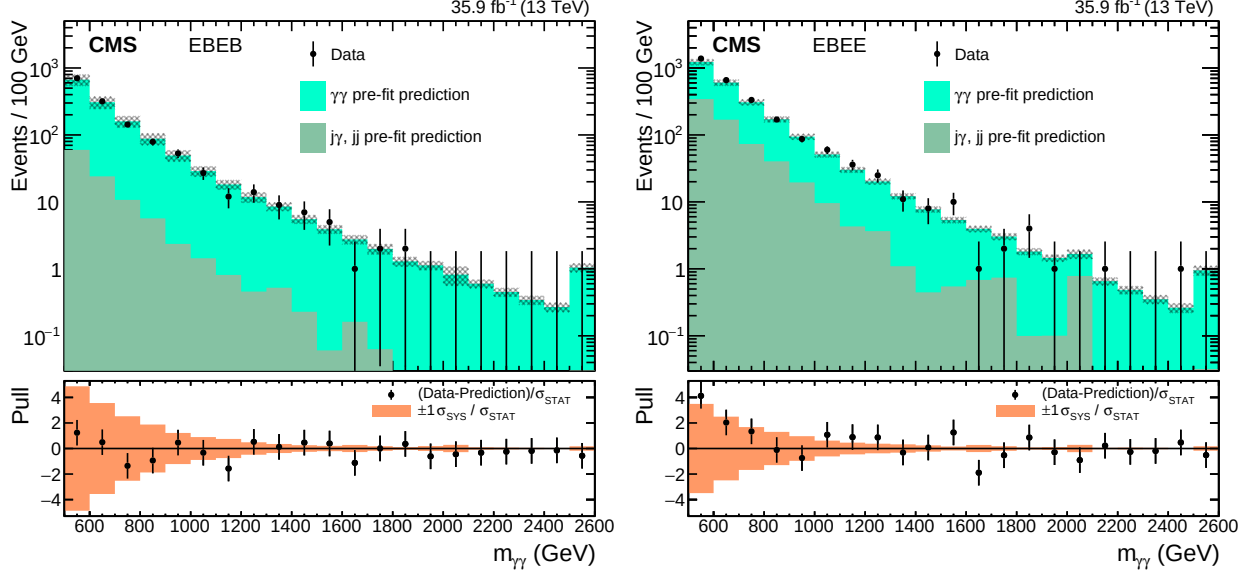


Figure 7.6: The diphoton invariant mass distributions in the EBEB (left) and EBEE (right) categories for the pre-fit SM diphoton background prediction and the fake background measurement compared to the data. The last bin includes the overflow. The error bars on the points indicate the statistical uncertainty, and the hatched bands indicate the total pre-fit systematic uncertainties. The bottom panels show the pull distributions, indicating the difference between the data and background prediction, divided by the uncertainty in the background, with error bars representing the statistical uncertainty and shaded bands showing the one standard deviation systematic uncertainty, normalized by the statistical uncertainty.

superimposed. The highest diphoton invariant mass events probed are 1840 and 2444 GeV in the EBEB and EBEE categories, respectively. A breakdown of the event yields for the total pre- and post-fit background predictions compared to the data are listed in Table 7.2, using different ranges of $m_{\gamma\gamma}$.

Table 7.2: Event yields for the total pre- and post-fit background predictions compared to the data for different ranges of $m_{\gamma\gamma}$. The total systematic uncertainties are included on the post-fit yields.

Category	Type	0.5-1 TeV	1-1.5 TeV	1.5-2 TeV	2-13 TeV	0.5-13 TeV
EBEB	Total pre-fit background	1275.7	73.0	11.1	3.5	1363.3
	Total post-fit background	1285 ± 40	73.7 ± 2.9	11.0 ± 0.5	3.30 ± 0.22	1373.0 ± 40.1
	Data	1296	69	10	0	1375
EBEE	Total pre-fit background	2406.1	121.4	15.6	4.4	2547.5
	Total post-fit background	2645 ± 60	128 ± 6	16.1 ± 1.2	4.1 ± 0.5	2793.2 ± 60.3
	Data	2631	140	18	2	2791

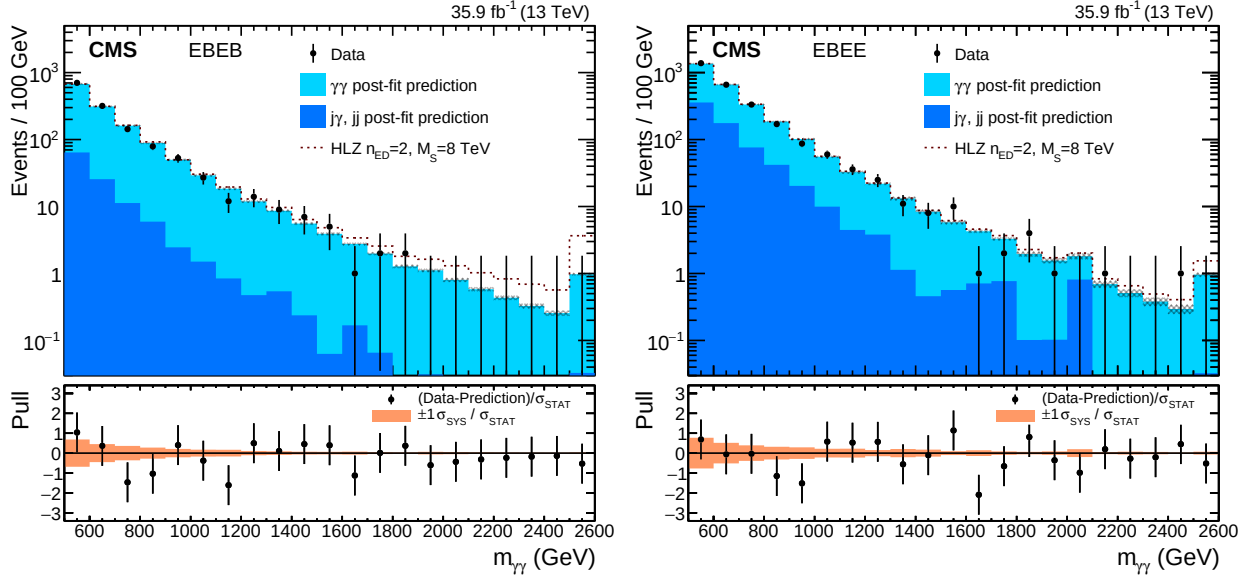


Figure 7.7: The diphoton invariant mass distributions in the EBEB (left) and EBEE (right) categories for the post-fit SM diphoton background prediction and the fake background measurement compared to the data. The last bin includes the overflow. The error bars on the points indicate the statistical uncertainty, and the hatched bands indicate the total post-fit systematic uncertainties. Invariant mass distributions from one signal scenario is superimposed. The bottom panels show the pull distributions, indicating the difference between the data and background prediction, divided by the uncertainty in the background, with error bars representing the statistical uncertainty and shaded bands showing the one standard deviation systematic uncertainty, normalized by the statistical uncertainty.

7.2 Limits on the ADD Model

Bayesian inference is used for setting upper limits on the ADD signal model parameters based on the background measured and data. An overview of general statistical methods can be found in Section 39 of Ref. [16] and in Ref. [132], and a general prescription of statistical applications to CMS searches can be found in Ref. [133]. The Bayesian approach used here follows the procedure described in Ref. [134].

Given some prior degree of belief in a hypothesis for an experimental measurement, the probability for a given event to occur can be inferred through Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (7.1)$$

for some hypothesis A given event B with $P(B) \neq 0$. $P(A)$ is the prior probability of A , $P(A|B)$ is the posterior probability of A given B , $P(B|A)$ is the likelihood of B given A , and $P(B)$ is sometimes called the marginal likelihood of B . For this limit-setting procedure, a binned likelihood is constructed such that the total likelihood $\mathcal{L}$ is equal to the product of the individual likelihoods $\mathcal{L}_i$ in each bin i . Given the counting nature of the experiment, the likelihoods take on the following Poissonian form:

$$\mathcal{L}(n|s, b) = \prod_i \mathcal{L}_i(n_i|s_i, b_i) = \prod_i \frac{(s_i + b_i)^{n_i} e^{-(s_i + b_i)}}{n_i!} \quad (7.2)$$

where n , s , and b are the total number of data, signal, and background events, respectively, and the subscript on these quantities denotes the number of them in bin i . Substituting this into Bayes' theorem, we recast its form as

$$P(s, b|n) = \frac{\mathcal{L}(n|s, b)\pi(s, b)}{P(n)} = \mathcal{L}(n|s, b)\pi(s, b) \quad (7.3)$$

where $P(n)$ is treated as a normalization constant and has been absorbed into the likelihood.

By convention in the field, a flat prior $\pi(s, b)$, bounded below by zero, is assumed for each ADD signal model considered.

The signal and background depend on their associated systematic uncertainties, which are each assigned a separate nuisance parameter θ . For each nuisance parameter, a choice of prior $\pi(s, b, \theta)$ is made. These are handled by integrating them out in the following way:

$$P(s, b|n) = \int P(s, b, \theta|n) d\theta = \int \mathcal{L}(n|s, b, \theta) \pi(s, b, \theta) d\theta \quad (7.4)$$

The prior shapes for each systematic uncertainty are either lognormal or Gaussian, depending on whether or not the uncertainty is bounded below by zero, except for the diphoton background normalization, which has a flat prior, motivated by our lack of prior knowledge on its uncertainty and due to this being a shape analysis. The posterior distributions for all nuisance parameters together with their ± 1 standard deviation variations in the EBEB and EBEE categories are shown in Figs. 7.8, 7.9, and 7.10. As seen in these figures, when the prior is Gaussian, its posterior distribution is largely unmodified and remains Gaussian, especially for the nuisance parameters associated with the PDF eigenvectors.

The nuisance parameters are assumed to be fully correlated across the $m_{\gamma\gamma}$ bins. If a nuisance parameter is uncorrelated between the EBEB and EBEE regions, then separate nuisance parameters are used. For the PDF uncertainties, the eigenvectors are assumed to be fully correlated not only across the EBEB and EBEE regions, but also between the signal and background. The correlation matrix among all nuisance parameters is shown in Fig. 7.11. No strong correlation is observed between the PDF eigenvectors. The technique for marginalizing the nuisance parameters was described in the previous section, on the search results. All $m_{\gamma\gamma}$ spectra used as input to marginalization process are unnormalized, except on the PDF uncertainties for the signal points. After marginalization, a maximum likelihood estimate is used to determine the posterior background prediction, as well as to provide a posterior prediction on the signal.

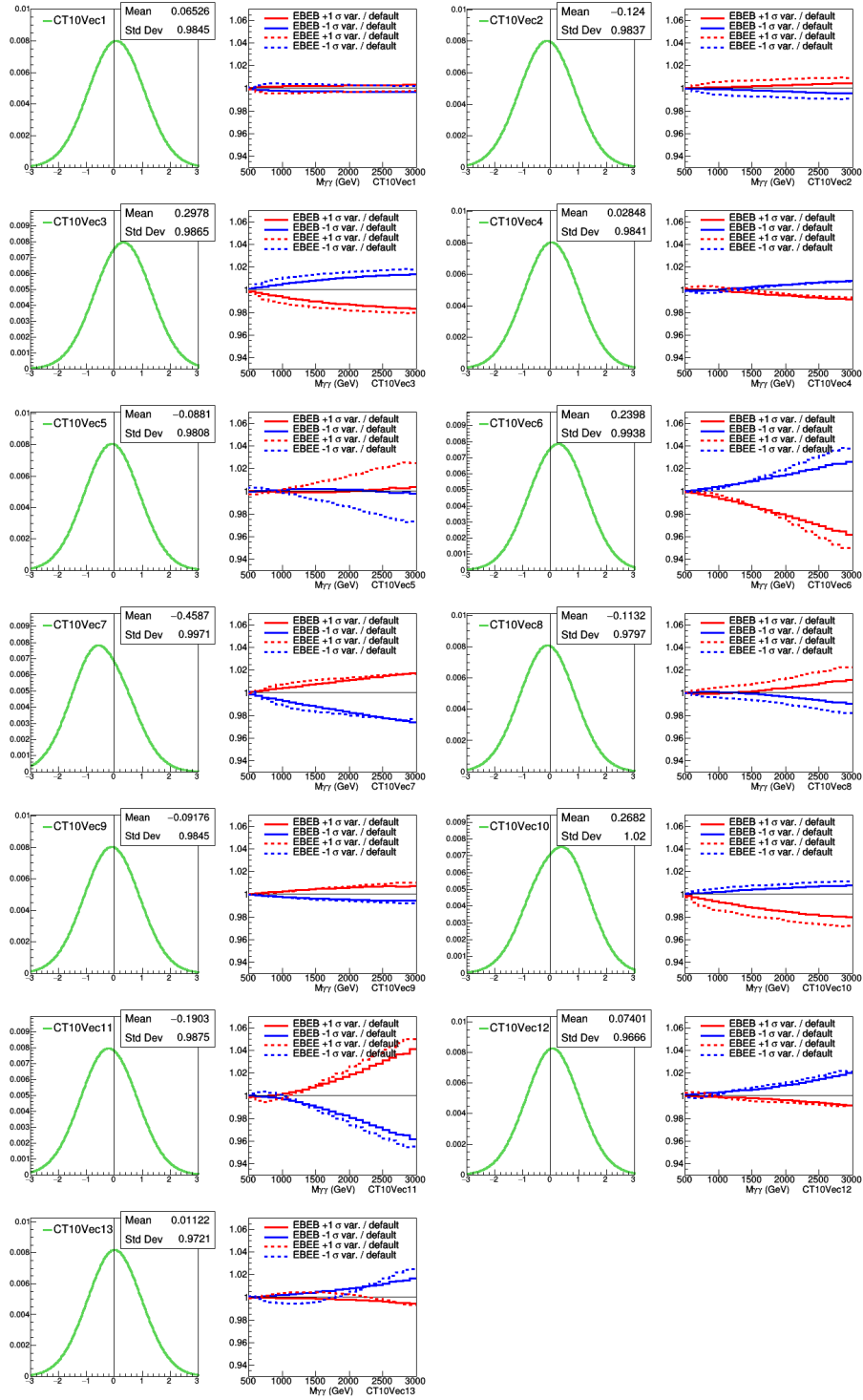


Figure 7.8: Posterior Gaussian distributions (using a mean of 0 and standard deviation of 1) of the nuisance parameters assigned to PDF eigenvectors 1-13 (first and third columns). Their ± 1 standard deviation variations are shown as a ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$ in the EBEB (solid) and EBEE (dashed) categories (second and fourth columns).

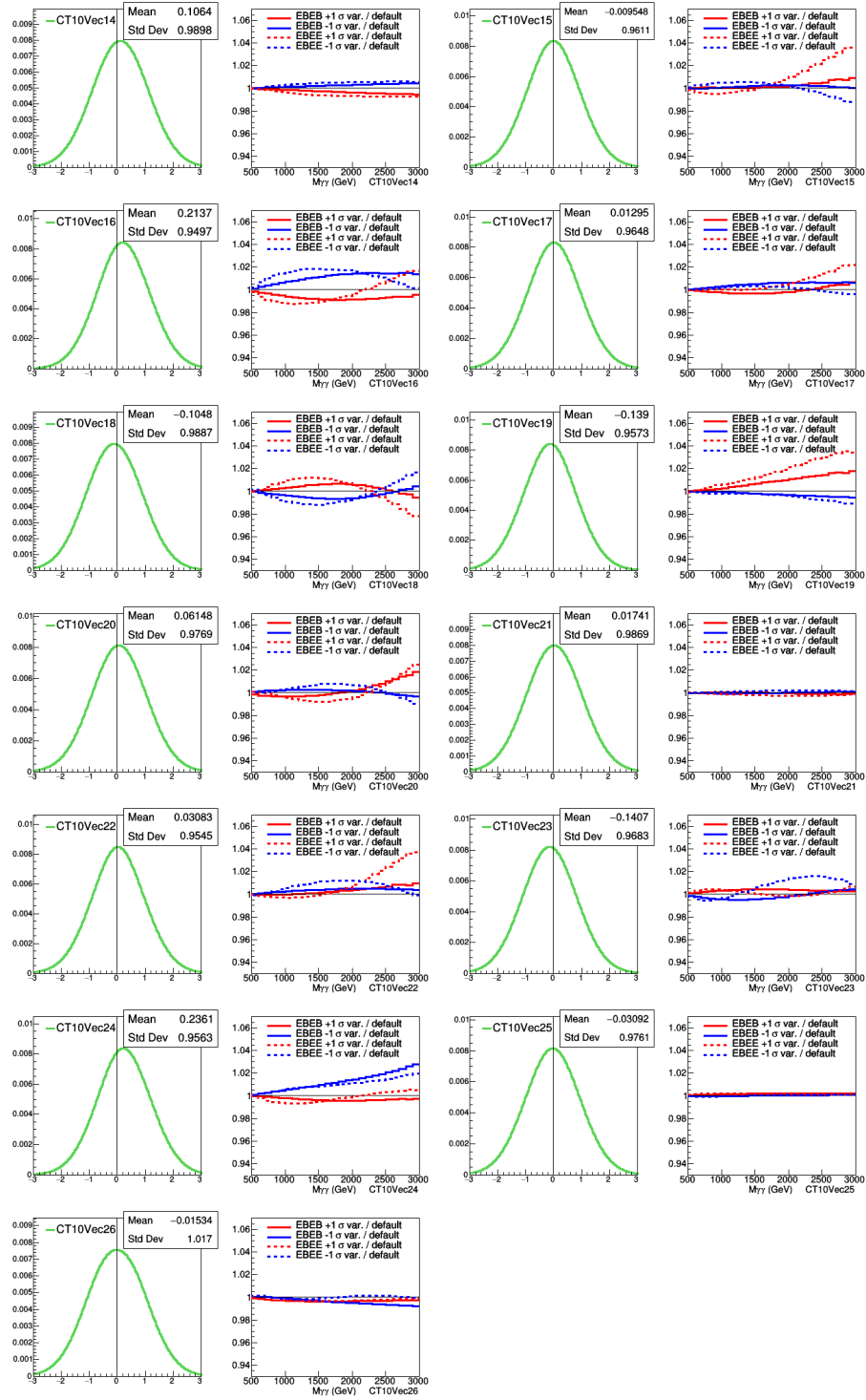


Figure 7.9: Posterior Gaussian distributions (using a mean of 0 and standard deviation of 1) of the nuisance parameters assigned to PDF eigenvectors 14-26 (first and third columns). Their ± 1 standard deviation variations are shown as a ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$ in the EBEB (solid) and EBEE (dashed) categories (second and fourth columns).

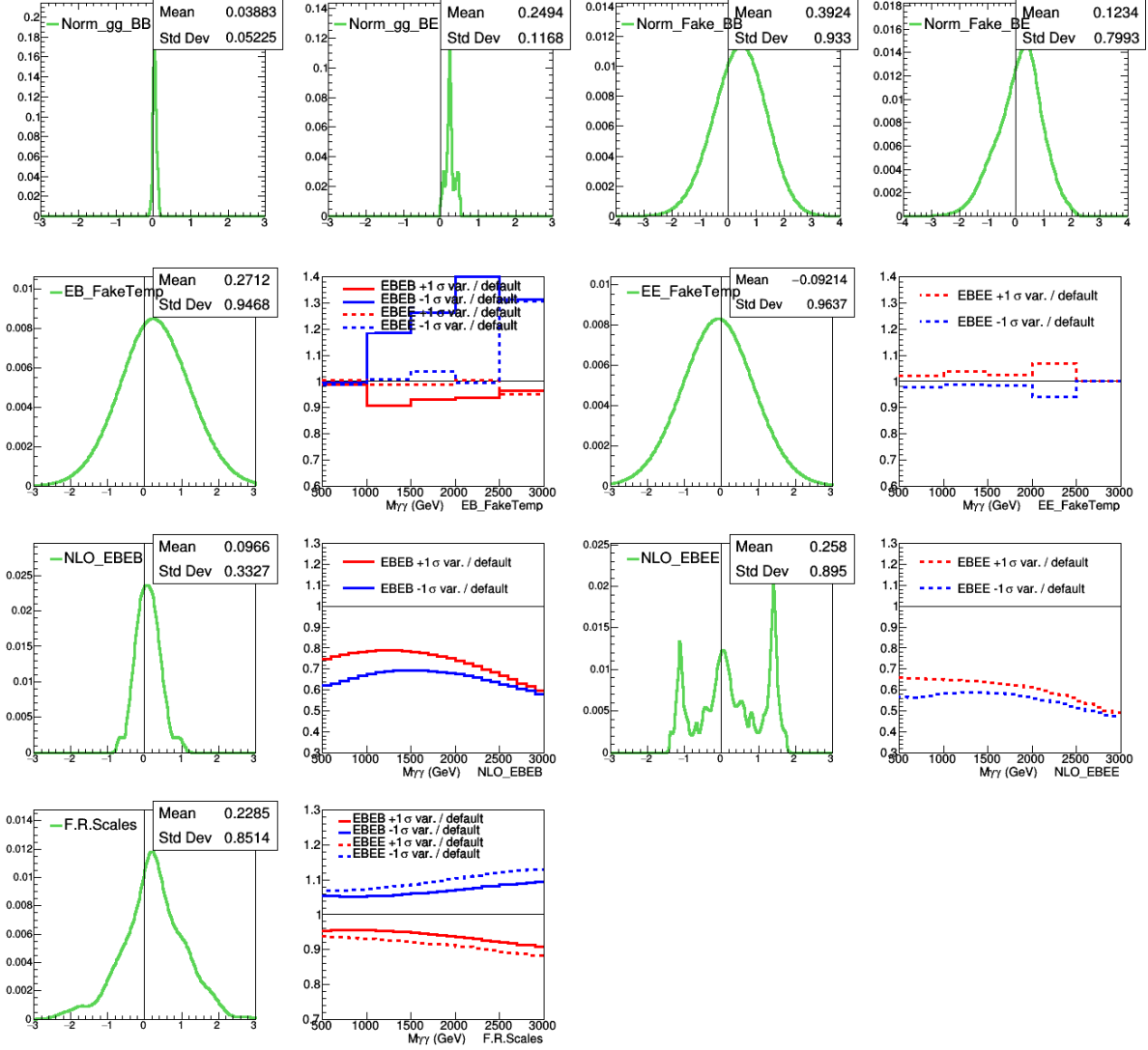


Figure 7.10: Posterior distributions for all non-PDF nuisance parameters associated with the real and fake backgrounds. Some of their associated ± 1 standard deviation variations are shown as a ratio of the variation over the default prediction as a function of $m_{\gamma\gamma}$ in the EBEB and EBEE categories. These are shown for the normalization of the real and fake background in the EBEB and EBEE categories (first row), the fake templates (second row), the NLO uncertainties (third row), and the factorization and renormalization scale variations (fourth row).

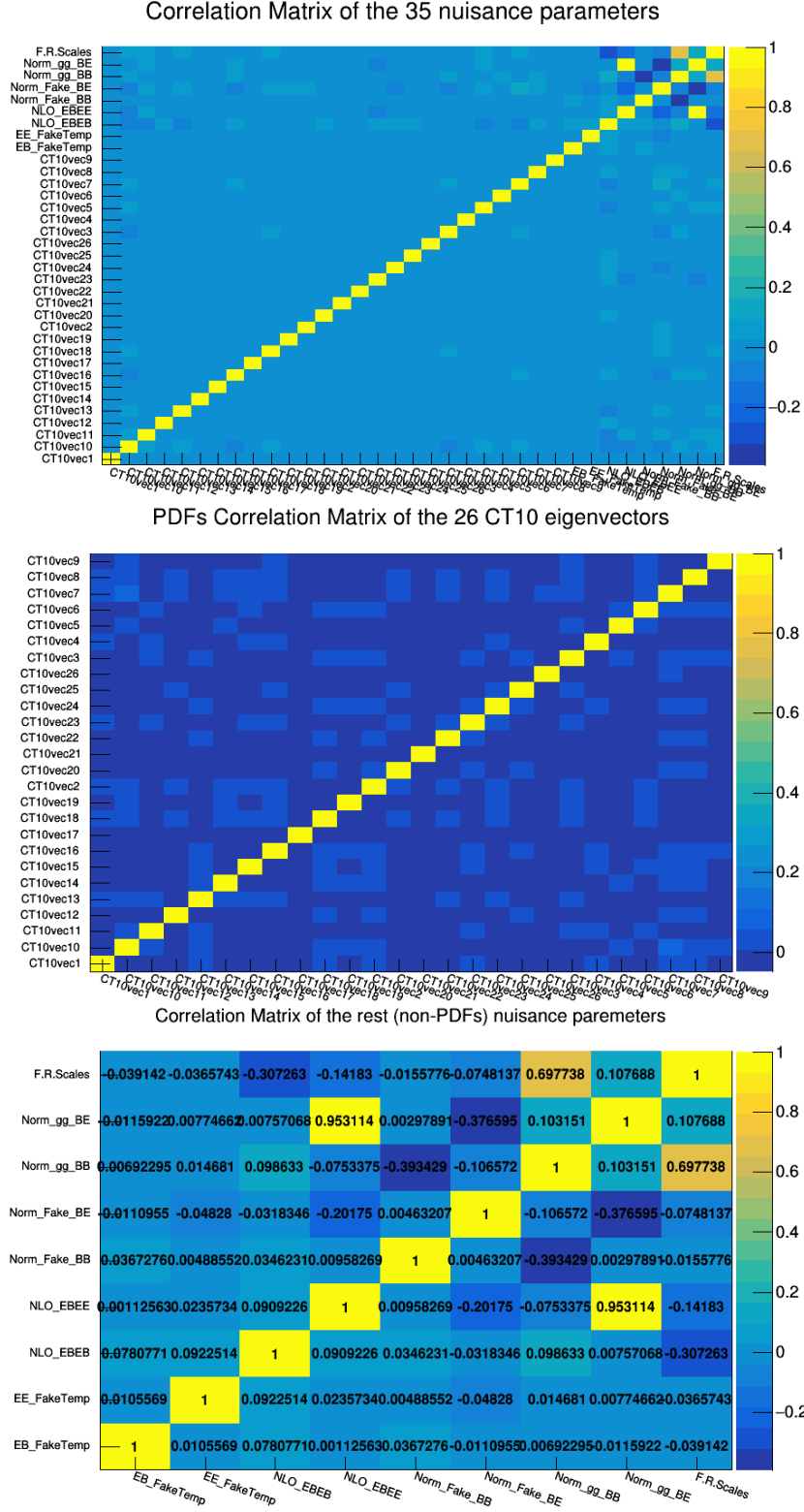


Figure 7.11: The correlation matrix for all 35 independent nuisance parameters used in marginalization of the real and fake background (top), the 26 PDF nuisance parameters (middle), and the remaining 9 non-PDF nuisance parameters (bottom).

The posterior predictions on the signal are used to derive the Bayesian 95% confidence level (CL) exclusion limits on the ADD model parameters, in particular, on the parameter M_S associated with the GRW, HLZ assuming $n_{\text{ED}} = 2$, and Hewett- model conventions. To determine the 95% CL upper limits on the signal, we need to integrate the posterior prediction to 95% of the total integral:

$$0.95 = \frac{\int_0^{s^{\text{up}}} P(s, b|n) ds}{\int_0^{\infty} P(s, b|n) ds} = \frac{\int_0^{s^{\text{up}}} \mathcal{L}(n|s, b) ds}{\int_0^{\infty} \mathcal{L}(n|s, b) ds} \quad (7.5)$$

where s^{up} is the value of s needed to satisfy this equation. These upper limits are translated into lower limits on the mass scale M_S by interpolating the value of M_S that has a signal of 1.0 excluded.

One caveat with this approach is that the interference effects between the SM background and ADD signal amplitudes are non-negligible. Even though the signal is assumed to scale with the normalization, the interference term actually scales as $1/M_S^4$, while the direct signal term scales as $1/M_S^8$. Since we are only interested in the signal for a fixed value of M_S , this approximation has a negligible effect, but, for values of the excluded signal away from 1.0, this cannot be naively extrapolated.

The 95% CL exclusion limits on the strength of the signal for the GRW, HLZ assuming $n_{\text{ED}} = 2$, and Hewett- conventions are shown in Fig. 7.12. These limits are expanded to the other ADD model conventions, namely, Hewett+ and HLZ assuming $n_{\text{ED}} = 3-7$, using the η_G parameterization in order to find equivalent conventions corresponding to a specific value of M_S , as described in Chapter 5. The expected and observed lower limits on M_S for the full ADD model parameter space considered in this analysis are shown in Table 7.3. The expected limits are determined assuming that the data follow the background exactly, while the observed limits are based on the actual data, which can differ. From the observed limits, the excluded values of M_S range from 5.6 to 9.7 TeV, depending on the model convention.

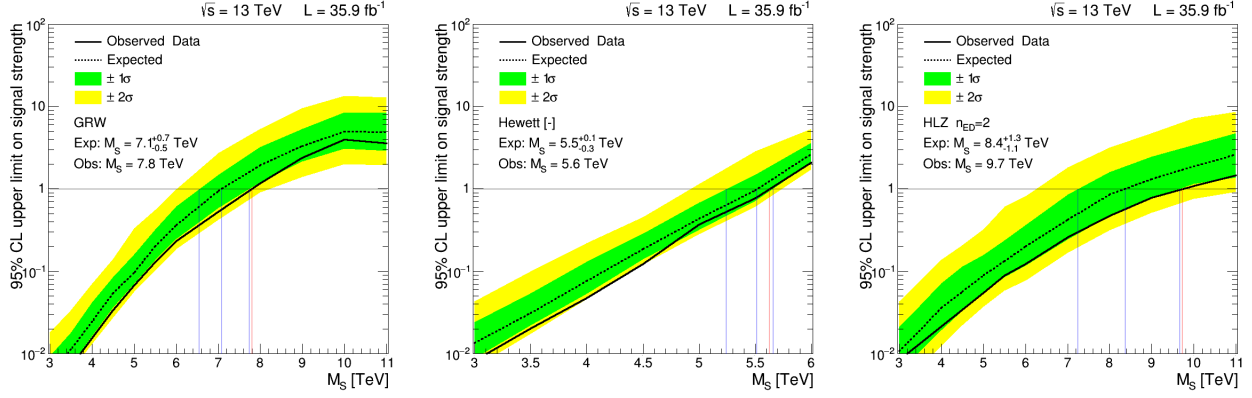


Figure 7.12: The 95% CL exclusion limits on the ADD model signal strength for the GRW (left), Hewett- (middle), and HLZ assuming $n_{ED} = 2$ (right) conventions. A 100 GeV binning is used.

Table 7.3: Exclusion limits on the mass scale M_S (in units of TeV) for various conventions used to calculate the ADD large extra-dimensional scenario. The total asymmetric uncertainties are shown on the expected limits.

Signal	GRW	Hewett		HLZ					
		negative	positive	$n_{ED} = 2$	$n_{ED} = 3$	$n_{ED} = 4$	$n_{ED} = 5$	$n_{ED} = 6$	$n_{ED} = 7$
Expected	$7.1^{+0.7}_{-0.5}$	$5.5^{+0.1}_{-0.3}$	$6.3^{+0.6}_{-0.4}$	$8.4^{+1.3}_{-1.1}$	$8.4^{+0.8}_{-0.6}$	$7.1^{+0.7}_{-0.5}$	$6.4^{+0.6}_{-0.5}$	$6.0^{+0.6}_{-0.4}$	$5.6^{+0.6}_{-0.4}$
Observed	7.8	5.6	7.0	9.7	9.3	7.8	7.0	6.6	6.2

Chapter 8

Conclusion

This dissertation presented a search for a nonresonant excess of high-mass diphoton events over the Standard Model background prediction produced from proton-proton collisions at a center-of-mass energy of 13 TeV. The data sample used corresponded to an integrated luminosity of 35.9 fb^{-1} , collected by the CMS detector at the Large Hadron Collider, and constituted the full 2016 dataset. Such an excess could be a signature of large extra dimensions, as described by the scenario of Arkani-Hamed, Dimopoulos, and Dvali. This model proposes a solution to the Standard Model hierarchy problem by modifying the fundamental Planck scale M_D to be on the order of the electroweak scale.

The large-extra dimensional signal was modeled using the SHERPA Monte Carlo event generator, taking into account the non-negligible interference effects between the signal and background. This model is parameterized by the the number of extra dimensions n_{ED} and the string mass scale M_S , which is directly proportional to M_D . Various conventions for constraining the model were considered, spanning a parameter space ranging between $3 < M_S < 11 \text{ TeV}$ for $n_{\text{ED}} = 2-7$. The diphoton invariant mass region above 500 GeV was used for this signal search.

A next-to-next-to-leading order Monte Carlo calculation was used to predict the dominant, irreducible Standard Model diphoton background using the SHERPA Monte Carlo event generator combined with the MCFM Monte Carlo parton-level calculator. To account

for additional higher order terms absent in this calculation, the normalization of predicted Standard Model background was allowed to float, under constraint from the data. A technique using control samples in data was used to measure the subdominant, reducible fake background, primarily arising when one or two jets fragment in such a way as to mimic a photon signature, causing them to be misreconstructed as diphoton events in the CMS detector. Templates representing real and fake photon distributions were fit to data to yield a measurement of the total fake contribution.

The data are found to be in agreement with the total background predicted from Standard Model sources, consistent with the background-only hypothesis. No evidence for new physics is observed. For the large extra-dimensional model of Arkani-Hamed, Dimopoulos, and Dvali, exclusion limits at 95% confidence level are set in the range $5.6 < M_S < 9.7 \text{ TeV}$, depending on the specific model convention. This range starts to surpass the range of the electroweak scale. This result extends the current best lower limits on M_S from the diphoton channel, as presented in Ref. [55].

This is the first search by the CMS Collaboration for a nonresonant excess in the high-mass diphoton channel performed during the LHC Run 2 data taking period. The previous CMS result used 2.2 fb^{-1} of Run 1 data collected in 2011 at a center-of-mass energy of 7 TeV [58]. The techniques used in Run 1 formed the basis of this search and were improved, in particular, by utilizing a next-to-next-to-leading order Monte Carlo calculation of the prompt Standard Model background and incorporating data from the CMS detector endcaps. Extensions of this result are expected by a planned combination with the CMS dilepton channel, as well as a future combination with the full 2017 and 2018 diphoton datasets, which would constitute the CMS Run 2 legacy result in the high-mass diphoton channel.

The results of this dissertation, along with an analogous search for a resonant excess in the high-mass diphoton channel, have been published in Ref. [64]. The existence of extra spatial dimensions in the universe has not been ruled out, but their hope of solving the Standard Model hierarchy problem has been narrowed by this effort.

References

- [1] C.-N. Yang and R. L. Mills, “Conservation of Isotopic Spin and Isotopic Gauge Invariance”, *Phys. Rev.* **96** (1954) 191–195, doi:10.1103/PhysRev.96.191.
- [2] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [3] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132–133, doi:10.1016/0031-9163(64)91136-9.
- [4] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.
- [6] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons”, *Phys. Rev.* **145** (1966) 1156–1163, doi:10.1103/PhysRev.145.1156.
- [7] T. W. B. Kibble, “Symmetry breaking in non-Abelian gauge theories”, *Phys. Rev.* **155** (1967) 1554–1561, doi:10.1103/PhysRev.155.1554.
- [8] Y. Nambu, “Quasiparticles and Gauge Invariance in the Theory of Superconductivity”, *Phys. Rev.* **117** (1960) 648–663, doi:10.1103/PhysRev.117.648.
- [9] J. Goldstone, “Field Theories with Superconductor Solutions”, *Nuovo Cim.* **19** (1961) 154–164, doi:10.1007/BF02812722.
- [10] J. Goldstone, A. Salam, and S. Weinberg, “Broken Symmetries”, *Phys. Rev.* **127** (1962) 965–970, doi:10.1103/PhysRev.127.965.
- [11] 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) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [12] 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) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

- [13] ATLAS and CMS Collaborations, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.
- [14] ATLAS and CMS Collaborations, “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”, *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.
- [15] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Submitted to: Eur. Phys. J.* (2018) arXiv:1809.10733.
- [16] Particle Data Group Collaboration, “Review of Particle Physics”, *Phys. Rev. D* **98** (2018), no. 3, 030001, doi:10.1103/PhysRevD.98.030001.
- [17] F. Halzen and A. D. Martin, “QUARKS AND LEPTONS: An Introductory Course in Modern Particle Physics”. John Wiley & Sons, Inc., New York, USA, 1984. ISBN 0471887412, 9780471887416.
- [18] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Nonabelian Gauge Theories”, *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.
- [19] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.
- [20] G. ’t Hooft and M. J. G. Veltman, “Regularization and Renormalization of Gauge Fields”, *Nucl. Phys. B* **44** (1972) 189–213, doi:10.1016/0550-3213(72)90279-9.
- [21] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.
- [22] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.
- [23] A. Salam, “Weak and Electromagnetic Interactions”, *Conf. Proc. C* **680519** (1968) 367–377.
- [24] Wikimedia Commons, “Standard Model Feynman Diagram Vertices”, Sept, 2018. https://en.wikipedia.org/wiki/Standard_Model.
- [25] C. Quigg, “Gauge Theories of the Strong, Weak, and Electromagnetic Interactions”. Princeton University Press, USA, 2013. ISBN 9780691135489, 9781400848225,
- [26] E. Witten, “Dynamical breaking of supersymmetry”, *Nucl. Phys. B* **188** (1981) 513, doi:10.1016/0550-3213(81)90006-7.
- [27] S. Dimopoulos and H. Georgi, “Softly broken supersymmetry and SU(5)”, *Nucl. Phys. B* **193** (1981) 150, doi:10.1016/0550-3213(81)90522-8.

- [28] CMS Collaboration, “Summaries of CMS cross section measurements”, Jul, 2018.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsCombined>.
- [29] S. P. Martin, “A Supersymmetry primer”,
doi:10.1142/9789812839657_0001, 10.1142/9789814307505_0001,
arXiv:hep-ph/9709356.
- [30] T. Kaluza, “Zum Unitätsproblem der Physik”, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1921** (1921) 966–972, arXiv:1803.08616.
- [31] O. Klein, “The Atomicity of Electricity as a Quantum Theory Law”, *Nature* **118** (1926) 516, doi:10.1038/118516a0.
- [32] O. Klein, “Quantum Theory and Five-Dimensional Theory of Relativity”, *Z. Phys.* **37** (1926) 895–906, doi:10.1007/BF01397481.
- [33] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “The hierarchy problem and new dimensions at a millimeter”, *Phys. Lett. B* **429** (1998) 263,
doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.
- [34] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “New dimensions at a millimeter to a fermi and superstrings at a TeV”, *Phys. Lett. B* **436** (1998) 257, doi:10.1016/S0370-2693(98)00860-0, arXiv:hep-ph/9804398.
- [35] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “Phenomenology, astrophysics and cosmology of theories with submillimeter dimensions and TeV scale quantum gravity”, *Phys. Rev. D* **59** (1999) 086004, doi:10.1103/PhysRevD.59.086004, arXiv:hep-ph/9807344.
- [36] 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.
- [37] 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.
- [38] J. L. Hewett and M. Spiropulu, “Particle physics probes of extra space-time dimensions”, *Ann. Rev. Nucl. Part. Sci.* **52** (2002) 397–424,
doi:10.1146/annurev.nucl.52.050102.090706, arXiv:hep-ph/0205106.
- [39] T. Han, J. D. Lykken, and R.-J. Zhang, “On Kaluza-Klein states from large extra dimensions”, *Phys. Rev. D* **59** (1999) 105006, doi:10.1103/PhysRevD.59.105006, arXiv:hep-ph/9811350.
- [40] T. Gleisberg et al., “Helicity formalism for spin-2 particles”, *JHEP* **0309** (2003) 001, doi:10.1088/1126-6708/2003/09/001, arXiv:hep-ph/0306182.
- [41] G. F. Giudice, R. Rattazzi, and J. D. Wells, “Quantum gravity and extra dimensions at high-energy colliders”, *Nucl. Phys. B* **544** (1999) 3–38,
doi:10.1016/S0550-3213(99)00044-9, arXiv:hep-ph/9811291.

- [42] J. L. Hewett, “Indirect collider signals for extra dimensions”, *Phys. Rev. Lett.* **82** (1999) 4765–4768, doi:10.1103/PhysRevLett.82.4765, arXiv:hep-ph/9811356.
- [43] J. Bijnens et al., “QCD signatures of narrow graviton resonances in hadron colliders”, *Phys. Lett. B* **503** (2001) 341–348, doi:10.1016/S0370-2693(01)00238-6, arXiv:hep-ph/0101316.
- [44] Y. Tang, “Implications of LHC Searches for Massive Graviton”, *JHEP* **08** (2012) 078, doi:10.1007/JHEP08(2012)078, arXiv:1206.6949.
- [45] CMS Collaboration, “Search for new physics in final states with a single photon and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Submitted to: JHEP* (2018) arXiv:1810.00196.
- [46] T. Banks and W. Fischler, “A Model for high-energy scattering in quantum gravity”, arXiv:hep-th/9906038.
- [47] S. Dimopoulos and G. L. Landsberg, “Black holes at the LHC”, *Phys. Rev. Lett.* **87** (2001) 161602, doi:10.1103/PhysRevLett.87.161602, arXiv:hep-ph/0106295.
- [48] S. B. Giddings and S. D. Thomas, “High-energy colliders as black hole factories: The End of short distance physics”, *Phys. Rev. D* **65** (2002) 056010, doi:10.1103/PhysRevD.65.056010, arXiv:hep-ph/0106219.
- [49] CMS Collaboration, “Search for black holes and sphalerons in high-multiplicity final states in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11** (2018) 042, doi:10.1007/JHEP11(2018)042, arXiv:1805.06013.
- [50] ATLAS Collaboration, “Search for extra dimensions using diphoton events in 7 TeV proton-proton collisions with the ATLAS detector”, *Phys. Lett. B* **710** (2012) 538–556, doi:10.1016/j.physletb.2012.03.022, arXiv:1112.2194.
- [51] ATLAS Collaboration, “Search for Extra Dimensions in diphoton events using proton-proton collisions recorded at $\sqrt{s} = 7$ TeV with the ATLAS detector at the LHC”, *New J. Phys.* **15** (2013) 043007, doi:10.1088/1367-2630/15/4/043007, arXiv:1210.8389.
- [52] ATLAS Collaboration, “Search for high-mass diphoton resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Phys. Rev. D* **92** (2015), no. 3, 032004, doi:10.1103/PhysRevD.92.032004, arXiv:1504.05511.
- [53] ATLAS Collaboration, “Search for resonances in diphoton events at $\sqrt{s}=13$ TeV with the ATLAS detector”, *JHEP* **09** (2016) 001, doi:10.1007/JHEP09(2016)001, arXiv:1606.03833.
- [54] ATLAS Collaboration, “Search for scalar diphoton resonances with 15.4 fb^{-1} of data collected at $\sqrt{s}=13$ TeV in 2015 and 2016 with the ATLAS detector”, Technical Report ATLAS-CONF-2016-059, CERN, Geneva, Aug, 2016.

- [55] ATLAS Collaboration, “Search for new phenomena in high-mass diphoton final states using 37 fb^{-1} of proton–proton collisions collected at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector”, *Phys. Lett. B* **775** (2017) 105–125, doi:10.1016/j.physletb.2017.10.039, arXiv:1707.04147.
- [56] CMS Collaboration, “Search for Large Extra Dimensions in the Diphoton Final State”, Technical Report CMS-PAS-EXO-09-004, CERN, Geneva, Jul, 2009.
- [57] CMS Collaboration, “Search for Large Extra Dimensions in the Diphoton Final State at the Large Hadron Collider”, *JHEP* **05** (2011) 085, doi:10.1007/JHEP05(2011)085, arXiv:1103.4279.
- [58] CMS Collaboration, “Search for signatures of extra dimensions in the diphoton mass spectrum at the Large Hadron Collider”, *Phys. Rev. Lett.* **108** (2012) 111801, doi:10.1103/PhysRevLett.108.111801, arXiv:1112.0688.
- [59] CMS Collaboration, “Search for new physics in high mass diphoton events in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”, Technical Report CMS-PAS-EXO-15-004, CERN, Geneva, 2015.
- [60] CMS Collaboration, “Search for diphoton resonances in the mass range from 150 to 850 GeV in pp collisions at $\sqrt{s} = 8 \text{ TeV}$ ”, *Phys. Lett. B* **750** (2015) 494–519, doi:10.1016/j.physletb.2015.09.062, arXiv:1506.02301.
- [61] CMS Collaboration, “Search for High-Mass Diphoton Resonances in pp Collisions at $\sqrt{s} = 8 \text{ TeV}$ with the CMS Detector”, Technical Report CMS-PAS-EXO-12-045, CERN, Geneva, 2015.
- [62] 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.
- [63] CMS Collaboration, “Search for Resonant Production of High-Mass Photon Pairs in Proton-Proton Collisions at $\sqrt{s} = 8$ and 13 TeV”, *Phys. Rev. Lett.* **117** (2016), no. 5, 051802, doi:10.1103/PhysRevLett.117.051802, arXiv:1606.04093.
- [64] CMS Collaboration, “Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ”, *Phys. Rev. D* **98** (2018), no. 9, 092001, doi:10.1103/PhysRevD.98.092001, arXiv:1809.00327.
- [65] CDF Collaboration, “Search for New Dielectron Resonances and Randall-Sundrum Gravitons at the Collider Detector at Fermilab”, *Phys. Rev. Lett.* **107** (2011) 051801, doi:10.1103/PhysRevLett.107.051801, arXiv:1103.4650.
- [66] D0 Collaboration, “Search for Randall-Sundrum gravitons in the dielectron and diphoton final states with 5.4 fb^{-1} of data from $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ ”, *Phys. Rev. Lett.* **104** (2010) 241802, doi:10.1103/PhysRevLett.104.241802, arXiv:1004.1826.

- [67] G. Landsberg, “Searches for Extra Spatial Dimensions with the CMS Detector at the LHC”, *Mod. Phys. Lett. A* **30** (2015), no. 15, 1540017, doi:10.1142/S0217732315400179, arXiv:1506.00024.
- [68] ATLAS Collaboration, “Search for contact interactions and large extra dimensions in the dilepton channel using proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **74** (2014), no. 12, 3134, doi:10.1140/epjc/s10052-014-3134-6, arXiv:1407.2410.
- [69] CMS Collaboration, “Search for physics beyond the standard model in dilepton mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **04** (2015) 025, doi:10.1007/JHEP04(2015)025, arXiv:1412.6302.
- [70] CMS Collaboration, “Z’ to tau tau - emu final state”, Technical Report CMS-PAS-EXO-12-046, CERN, Geneva, 2015.
- [71] CMS Collaboration, “Search for new physics in dijet angular distributions using proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models”, *Eur. Phys. J. C* **78** (2018), no. 9, 789, doi:10.1140/epjc/s10052-018-6242-x, arXiv:1803.08030.
- [72] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [73] P. Mouche, “Overall view of the LHC. Vue d’ensemble du LHC”, Jun, 2014. <https://cds.cern.ch/record/1708847>.
- [74] E. Mobs, “The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018”, Aug, 2018. <http://cds.cern.ch/record/2636343>.
- [75] “LHC Machine Outreach”, 2005. <https://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/>.
- [76] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [77] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.
- [78] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [79] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, doi:10.1088/1748-0221/3/08/S08005.
- [80] AC Team, “Diagram of an LHC dipole magnet. Schéma d’un aimant dipôle du LHC”, Jun, 1999. <https://cds.cern.ch/record/40524>.
- [81] O. S. Brüning et al., “LHC Design Report Vol.1: The LHC Main Ring”. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.

- [82] CMS Collaboration, “Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2018) 161, doi:10.1007/JHEP07(2018)161, arXiv:1802.02613.
- [83] CMS Collaboration, “CMS Luminosity Measurements for the 2016 Data Taking Period”, Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
- [84] A. G. et al., “High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1”. CERN Yellow Reports: Monographs. CERN, Geneva, 2017.
- [85] D. Contardo et al., “Technical Proposal for the Phase-II Upgrade of the CMS Detector”, Technical Report CERN-LHCC-2015-010. LHCC-P-008. CMS-TDR-15-02, Geneva, Jun, 2015.
- [86] CMS Collaboration, “CMS Phase II Upgrade Scope Document”, Technical Report CERN-LHCC-2015-019. LHCC-G-165, CERN, Geneva, Sep, 2015.
- [87] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter”, Technical Report CERN-LHCC-2017-023. CMS-TDR-019, CERN, Geneva, Nov, 2017.
- [88] CMS Collaboration, “CMS Luminosity - Public Results”, 2018.
<https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>.
- [89] T. Sakuma and T. McCauley, “Detector and Event Visualization with SketchUp at the CMS Experiment”, *J. Phys. Conf. Ser.* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032, arXiv:1311.4942.
- [90] CMS Collaboration, “Interactive Slice of the CMS detector”, Aug, 2016.
<https://cds.cern.ch/record/2205172>.
- [91] M. Schott and M. Dunford, “Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV. Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV”, *Eur. Phys. J. C* **74** (May, 2014) 60 p.
- [92] CMS Collaboration, “Precision measurement of the structure of the CMS inner tracking system using nuclear interactions”, *JINST* **13** (2018), no. 10, P10034, doi:10.1088/1748-0221/13/10/P10034, arXiv:1807.03289.
- [93] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [94] CMS Collaboration, “CMS ECAL with 2017 data”, May, 2018.
<https://cds.cern.ch/record/2319285>.
- [95] CMS Collaboration, “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV”, *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016.

- [96] CMS Collaboration, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.
- [97] CMS Collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [98] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018), no. 06, P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- [99] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.
- [100] 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.
- [101] CMS Collaboration, “Performance of tau-lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001, arXiv:1109.6034.
- [102] CMS Collaboration, “Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018), no. 10, P10005, doi:10.1088/1748-0221/13/10/P10005, arXiv:1809.02816.
- [103] CMS Collaboration, “Performance of quark/gluon discrimination in 8 TeV pp data”, Technical Report CMS-PAS-JME-13-002, CERN, Geneva, 2013.
- [104] ATLAS Collaboration, “Light-quark and gluon jet discrimination in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *Eur. Phys. J. C* **74** (2014), no. 8, 3023, doi:10.1140/epjc/s10052-014-3023-z, arXiv:1405.6583.
- [105] L. Asquith et al., “Jet Substructure at the Large Hadron Collider : Experimental Review”, Technical Report FERMILAB-PUB-18-123-PPD, 2018. arXiv:1803.06991.
- [106] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018), no. 05, P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [107] CMS Collaboration, “Measurement of the top quark mass using single top quark events in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **77** (2017), no. 5, 354, doi:10.1140/epjc/s10052-017-4912-8, arXiv:1703.02530.
- [108] 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.
- [109] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

- [110] CMS Collaboration, “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, *Eur. Phys. J. C* **74** (2014), no. 10, 3076, doi:10.1140/epjc/s10052-014-3076-z, arXiv:1407.0558.
- [111] CMS Collaboration, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Submitted to: JHEP* (2018) arXiv:1804.02716.
- [112] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.
- [113] M. Cacciari, G. P. Salam, and G. Soyez, “The Catchment Area of Jets”, *JHEP* **04** (2008) 005, doi:10.1088/1126-6708/2008/04/005, arXiv:0802.1188.
- [114] CMS Collaboration, “Pileup Removal Algorithms”, Technical Report CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- [115] CMS Collaboration, “Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **10** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789.
- [116] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.
- [117] B. R. Webber, “A QCD Model for Jet Fragmentation Including Soft Gluon Interference”, *Nucl. Phys. B* **238** (1984) 492–528, doi:10.1016/0550-3213(84)90333-X.
- [118] J.-C. Winter, F. Krauss, and G. Soff, “A Modified cluster hadronization model”, *Eur. Phys. J. C* **36** (2004) 381–395, doi:10.1140/epjc/s2004-01960-8, arXiv:hep-ph/0311085.
- [119] H.-L. Lai et al., “New parton distributions for collider physics”, *Phys. Rev. D* **82** (2010) 074024, doi:10.1103/PhysRevD.82.074024, arXiv:1007.2241.
- [120] J. Gao et al., “CT10 next-to-next-to-leading order global analysis of QCD”, *Phys. Rev. D* **89** (2014) 033009, doi:10.1103/PhysRevD.89.033009, arXiv:1302.6246.
- [121] J. M. Campbell, J. W. Huston, and W. J. Stirling, “Hard Interactions of Quarks and Gluons: A Primer for LHC Physics”, *Rept. Prog. Phys.* **70** (2007) 89, doi:10.1088/0034-4885/70/1/R02, arXiv:hep-ph/0611148.
- [122] L. D’Errico and P. Richardson, “Next-to-Leading-Order Monte Carlo Simulation of Diphoton Production in Hadronic Collisions”, *JHEP* **02** (2012) 130, doi:10.1007/JHEP02(2012)130, arXiv:1106.3939.
- [123] S. Hoeche, S. Schumann, and F. Siegert, “Hard photon production and matrix-element parton-shower merging”, *Phys. Rev. D* **81** (2010) 034026, doi:10.1103/PhysRevD.81.034026, arXiv:0912.3501.

- [124] GEANT4 Collaboration, “GEANT4—A Simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250–303, doi:10.1016/S0168-9002(03)01368-8.
- [125] J. M. Campbell, R. K. Ellis, Y. Li, and C. Williams, “Predictions for diphoton production at the LHC through NNLO in QCD”, *JHEP* **07** (2016) 148, doi:10.1007/JHEP07(2016)148, arXiv:1603.02663.
- [126] S. Frixione, “Isolated photons in perturbative QCD”, *Phys. Lett. B* **429** (1998) 369–374, doi:10.1016/S0370-2693(98)00454-7, arXiv:hep-ph/9801442.
- [127] R. Barlow and C. Beeston, “Fitting using finite Monte Carlo samples”, *Computer Physics Communications* **77** (1993) 219–228.
- [128] T. Binoth, J. P. Guillet, E. Pilon, and M. Werlen, “A Full next-to-leading order study of direct photon pair production in hadronic collisions”, *Eur. Phys. J. C* **16** (2000) 311–330, doi:10.1007/s100520050024, arXiv:hep-ph/9911340.
- [129] 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.
- [130] T. Müller, J. Ott, and J. Wagner-Kuhr, “theta – a framework for template-based modeling and inference”, 2010.
<http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta.pdf>,
<http://www-ekp.physik.uni-karlsruhe.de/~ott/theta/theta-auto>.
- [131] N. Metropolis et al., “Equation of state calculations by fast computing machines”, *J. Chem. Phys.* **21** (1953) 1087, doi:10.1063/1.1699114.
- [132] R. D. Cousins, “Lectures on Statistics in Theory: Prelude to Statistics in Practice”, arXiv:1807.05996.
- [133] ATLAS and CMS Collaborations, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, Aug, 2011.
- [134] D0 Collaboration, “A Recipe for the construction of confidence limits”, Technical Report FERMILAB-TM-2104, D0-NOTE-3476, D0-NOTE-2775-A, 2000. doi:10.2172/753775.

Appendix A

Example MC Run Cards

The following sections contain example run cards, each called *Run.dat*, used to configure the MC event generator SHERPA v.2.1.1 for use in modeling the signal and background needed for this analysis. Particles are specified using the PDG ID code numbering scheme, as described in Section 43 of Ref. [16]. Photons correspond to the PDG ID code 22. The proton beams are specified using the PDG ID code 2212 and are set to $\sqrt{s} = 13$ TeV. The PDG ID code 21 corresponds to gluons, and SHERPA uses an internal code of 93 to indicate particles in its jet container.

A.1 Higher Order SM Background Run Card

The following run card is used to configure SHERPA for generation of the higher order SM background used for the real background calculation, as described in Section 4.1. The box process is calculated with an internal loop generator using a variable scale to set both the renormalization and factorization scales to the diphoton invariant mass. The Born process includes up to three additional jets in its final state and uses a 20 GeV scale for the ME+PS merging scheme. This example configuration uses the following kinematic selection: photon $p_T > 50$ GeV, photon $|\eta| < 2.8$, and $60 < m_{\gamma\gamma} < 200$ GeV.

```
(run){  
  EVENTS = 100;  
  EVENT_MODE = HepMC;  
  WRITE_MAPPING_FILE 3;  
}(run)
```

```
(beam){  
  BEAM_1 = 2212; BEAM_ENERGY_1 = 6500.;  
  BEAM_2 = 2212; BEAM_ENERGY_2 = 6500.;  
}(beam)
```

```
(processes){  
  Process 21 21 -> 22 22;  
  Scales VAR{Abs2(p[2]+p[3])};  
  Loop_Generator gg_yy;  
  End process;  
  Process 93 93 -> 22 22 93{3};  
  Order_EW 2;  
  Enhance_Factor 2 {3};  
  Enhance_Factor 5 {4};  
  Enhance_Factor 10 {5};  
  CKKW sqr(20./E_CMS);  
  Integration_Error 0.005 {3};  
  Integration_Error 0.03 {4};  
  Integration_Error 0.05 {5};  
  End process;  
}(processes)
```

```
(selector){  
  Mass 22 22 60. 200.;  
  PT 22 50. E_CMS;
```

```

PseudoRapidity 22 -2.8 2.8;
}(selector)

(shower){
    CSS_EW_MODE = 1;
}(shower)

(integration){
    FINISH_OPTIMIZATION = Off;
}(integration)

(isr){
    PDF_LIBRARY = LHAPDFSherpa;
    PDF_SET = CT10;
    PDF_SET_VERSION = 0;
}(isr)

(me){
    ME_SIGNAL_GENERATOR = Internal Comix;
    EVENT_GENERATION_MODE = Unweighted;
}(me)

(mi){
    MI_HANDLER = Amisic;
}(mi)

```

A.2 An ADD Signal Run Card

This run card is used to configure SHERPA for generation of the ADD signal, as described in Chapter 5. In this example configuration, the signal model parameters $KK = 1$, $n_{\text{ED}} = 4$, and $M_S = 3500$ GeV are specified. The kinematic selection of $p_T > 70$ GeV and $|\eta| < 2.8$ is imposed for photons and $m_{\gamma\gamma}$ is restricted to be between 1000-2000 GeV.

Run.dat

```
(run){  
  EVENTS = 100;  
  EVENT_MODE = HepMC;  
  WRITE_MAPPING_FILE 3;  
}(run)  
  
(beam){  
  BEAM_1 = 2212; BEAM_ENERGY_1 = 6500.;  
  BEAM_2 = 2212; BEAM_ENERGY_2 = 6500.;  
}(beam)  
  
(model){  
  MODEL = ADD;  
  M_S = 3500;  
  N_ED = 4;  
  KK_CONVENTION = 1;  
  G_NEWTON = 6.707e-39;  
}(model)  
  
(processes){  
  Process 21 21 -> 22 22;  
  Scales VAR{Abs2(p[2]+p[3])};  
  Loop_Generator gg_yy;  
  End process;
```

```

Process 93 93 -> 22 22;
Order_EW 2;
CKKW sqr(20./E_CMS);
Print_Graphs : Process;
End process;
}(processes)

(selector){
Mass 22 22 1000. 2000.;
PT 22 70. E_CMS;
PseudoRapidity 22 -2.8 2.8;
}(selector)

(shower){
CSS_EW_MODE = 1;
}(shower)

(integration){
FINISH_OPTIMIZATION = Off;
}(integration)

(me){
ME_SIGNAL_GENERATOR = Amegic;
}(me)

(mi){
MI_HANDLER = Amisic;
}(mi)

```

A.3 Leading Order SM Background Run Card

The below run card is used to configure SHERPA for generation of the LO SM background, which is used with the LO ADD+SM signal samples in order to extract the ADD signal with interference samples, as described in Chapter 5. This run card is identical to the one used to produce the corresponding ADD signal, such as the one shown above in Appendix A.2, but has the model section removed. This example configuration restricts $m_{\gamma\gamma}$ between 500-1000 GeV.

Run.dat

```
(run){
  EVENTS = 100;
  EVENT_MODE = HepMC;
  WRITE_MAPPING_FILE 3;
}(run)

(beam){
  BEAM_1 = 2212; BEAM_ENERGY_1 = 6500.;
  BEAM_2 = 2212; BEAM_ENERGY_2 = 6500.;
}(beam)

(processes){
  Process 21 21 -> 22 22;
  Scales VAR{Abs2(p[2]+p[3])};
  Loop_Generator gg_yy;
  End process;
  Process 93 93 -> 22 22;
  Order_EW 2;
  CKKW sqr(20./E_CMS);
  Print_Graphs : Process;
  End process;
```

```

}(processes)

(selector){
  Mass 22 22 500. 1000.;
  PT 22 70. E_CMS;
  PseudoRapidity 22 -2.8 2.8;
}(selector)

(shower){
  CSS_EW_MODE = 1;
}(shower)

(integration){
  FINISH_OPTIMIZATION = Off;
}(integration)

(me){
  ME_SIGNAL_GENERATOR = Amegic;
}(me)

(mi){
  MI_HANDLER = Amisic;
}(mi)

```

Appendix B

Additional Fake Rate Material

B.1 The MC Fake Rate Fits

The $\sigma_{i\eta i\eta}$ template fits used in the MC based closure test to the photon fake rate method are shown in Fig. B.1 for the EB and EE regions. The fits are performed in photon p_T bins of 50-70, 70-90, 90-130, 130-200, and 200-600 GeV. The nominal sideband definition of $5 < \text{Iso}_{\text{Ch}} < 10$ GeV is used for both categories. These same plots were produced after inverting the template and sideband variables. The Iso_{Ch} template fits are shown in Fig. B.2 for the EB and EE regions. The fits are performed in the same photon p_T bins, using the nominal sideband of $0.0105 < \sigma_{i\eta i\eta} < 0.0150$ in the EB and $0.0280 < \sigma_{i\eta i\eta} < 0.0400$ in the EE category.

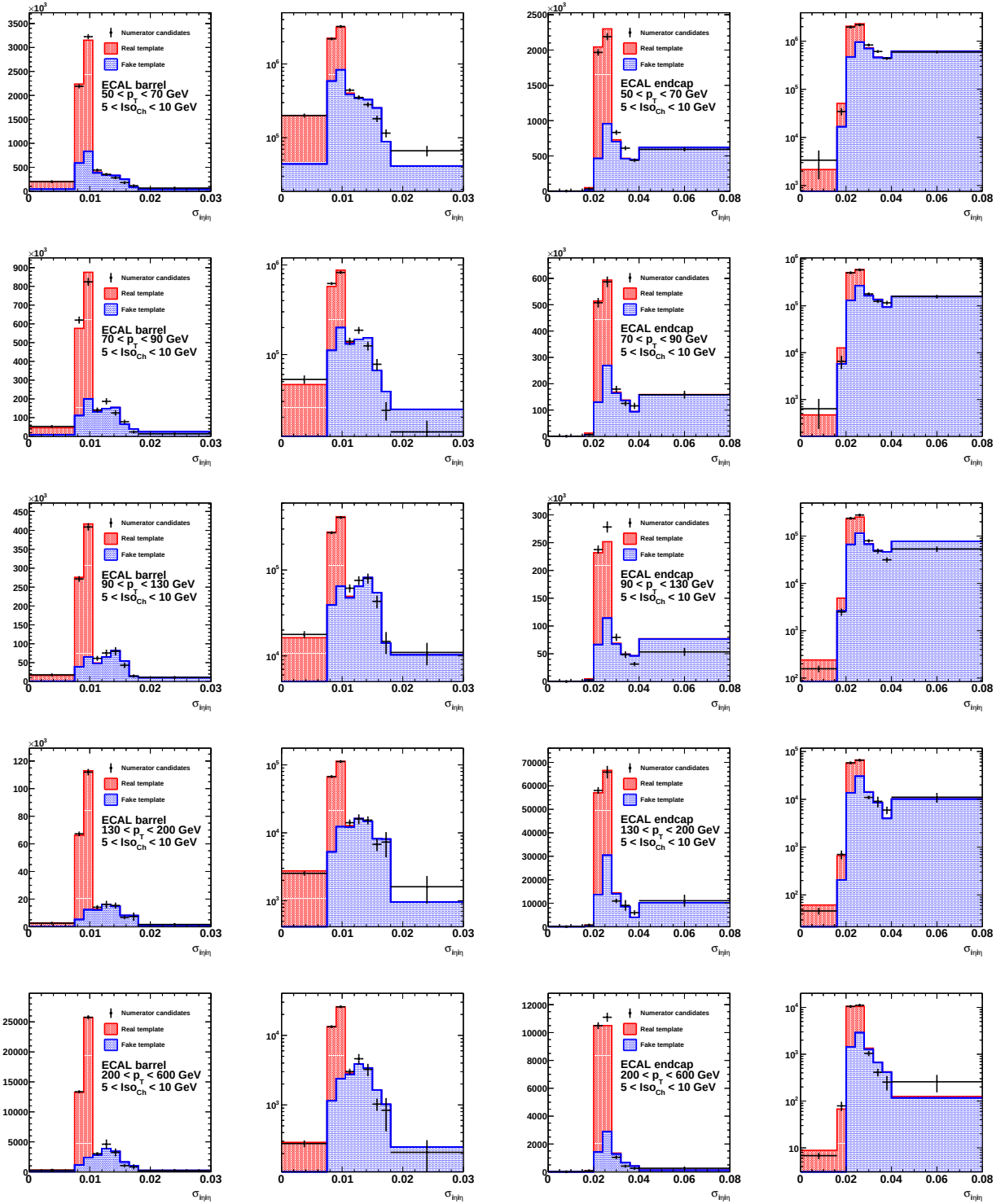


Figure B.1: Template fits for the MC closure test using the $\sigma_{ii\eta\eta}$ template variable, shown in linear and log scales, for the EB (left) and EE (right) regions. The different p_T bins are shown in the rows.

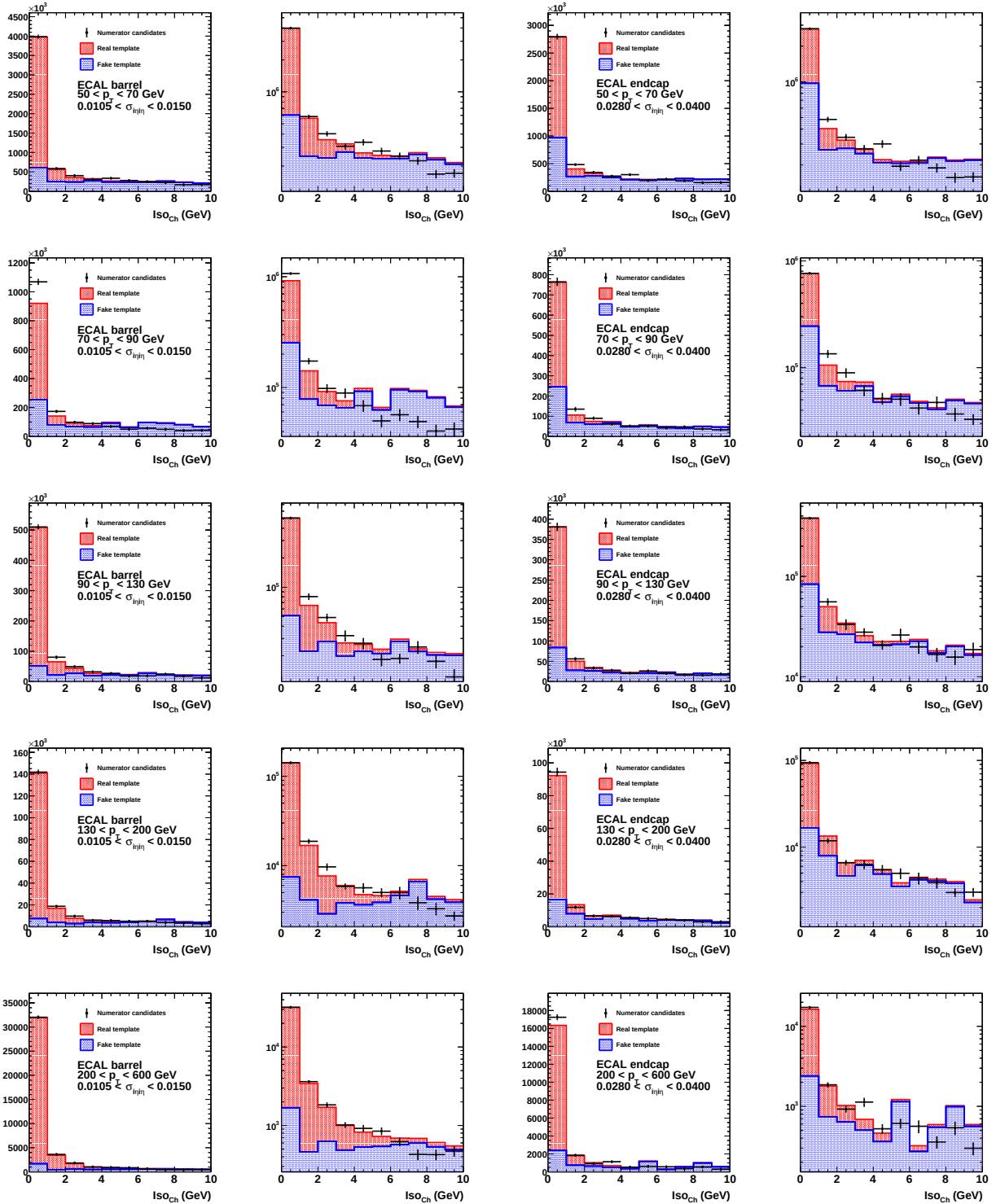


Figure B.2: Template fits for the MC closure test using the Iso_{Ch} template variable, shown in linear and log scales, for the EB (left) and EE (right) regions. The different p_{T} bins are shown in the rows.

B.2 The MC Fake Templates

The fake templates in $\sigma_{i\eta i\eta}$ used in the MC based closure test to the photon fake rate method are shown in Fig. B.3 for the EB and EE categories. The templates are shown in p_T bins of 50-70, 70-90, 90-130, 130-200, and 200-600 GeV. For each p_T bin, the templates constructed using the nominal Iso_{Ch} sideband of 5-10 GeV are compared to those derived using the MC truth information, which can be considered the ideal fake template corresponding in that p_T bin. The fake templates in Iso_{Ch} using these same p_T bins are shown in Fig. B.4 for the EB and EE categories. The templates constructed using different $\sigma_{i\eta i\eta}$ sidebands are compared against the fake templates derived using the MC truth information in the simulated sample for the EB and EE categories.

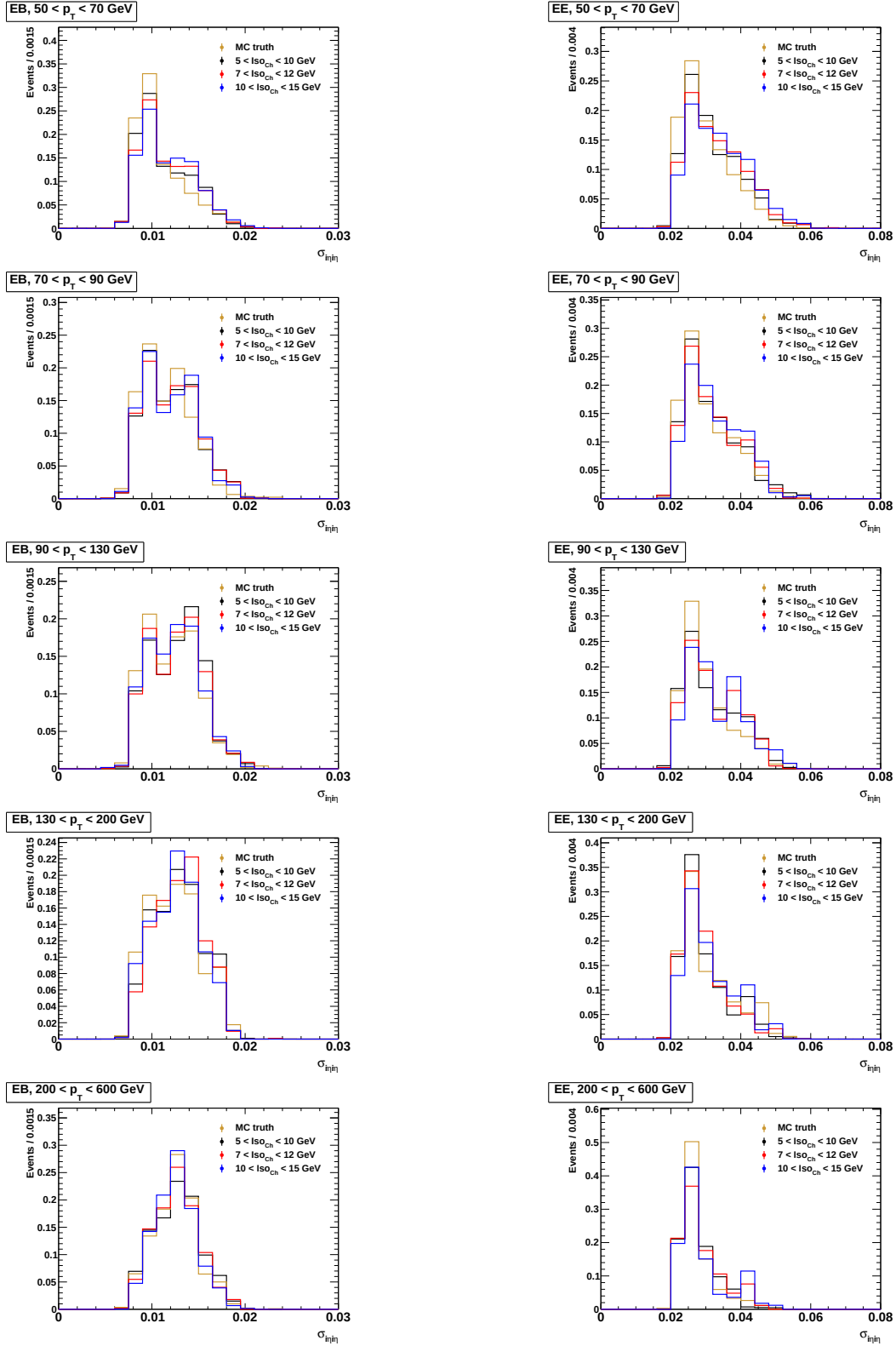


Figure B.3: Fake templates in σ_{inj} from simulation in the EB (left) and EE (right) categories for various p_T bins, shown in the rows.

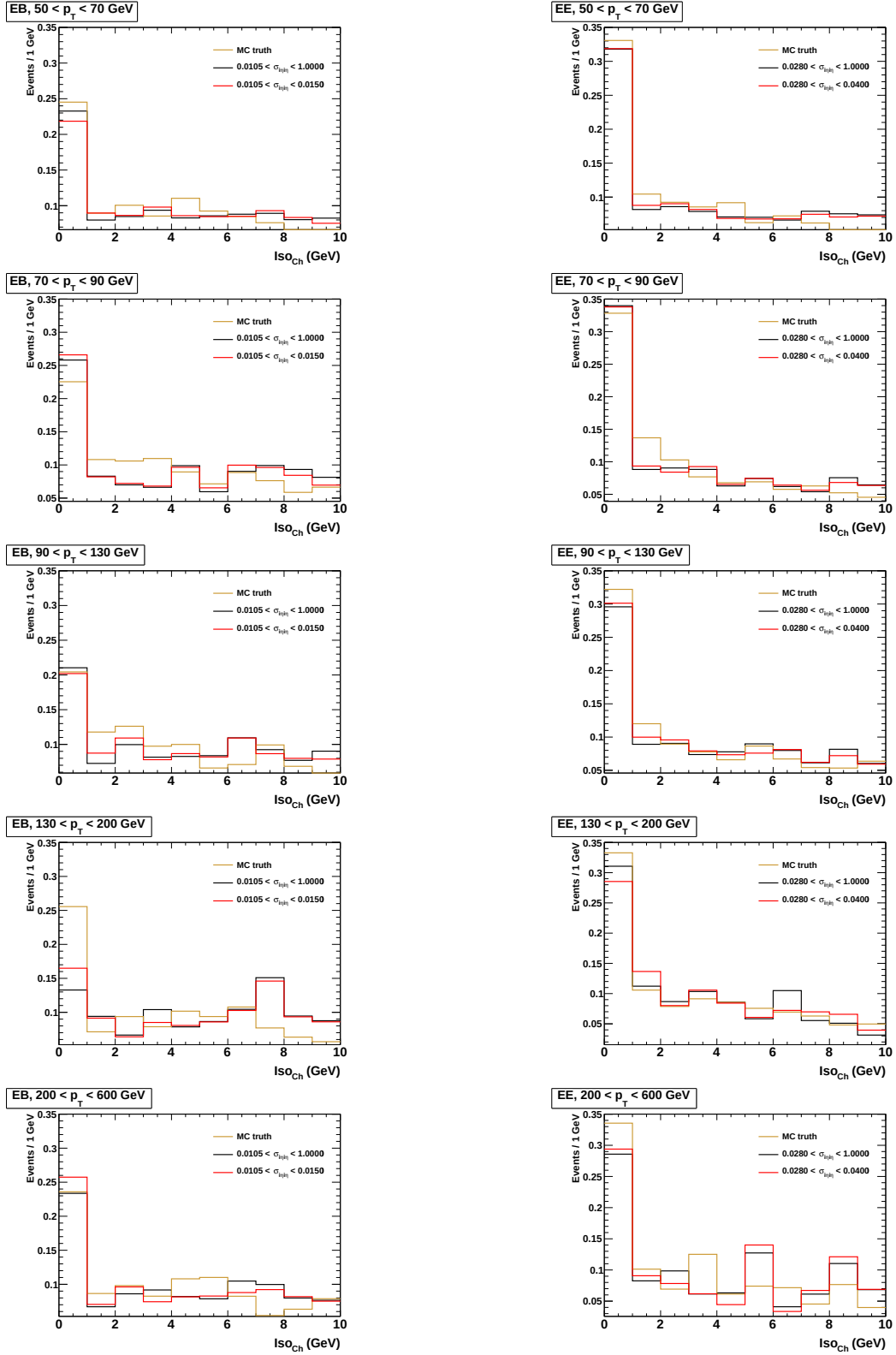


Figure B.4: Fake templates in Iso_{Ch} from simulation in the EB (left) and EE (right) categories for various p_T bins, shown in the rows.