

SEARCH FOR NON-RESONANT NEW PHYSICS IN HIGH-MASS DIPHOTON
EVENTS FROM PP COLLISIONS AT $\sqrt{s} = 13 \text{ TeV}$ WITH THE CMS
DETECTOR AND HIGHLIGHTS FROM HCAL UPGRADES

by

CILICIA UZZIEL PEREZ

CONOR HENDERSON, COMMITTEE CHAIR

SERGEI GLEYZER, CO-CHAIR

PAOLO RUMERIO

DAWN WILLIAMS

NOBUCHIKA OKADA

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

2024



ABSTRACT

Results are presented from a search for non-resonant or broad signatures of beyond the standard model physics in the high-mass diphoton spectrum from proton-proton collisions at $\sqrt{s} = 13$ TeV. The analysis uses data collected in 2016–2018 with the CMS detector at the LHC and corresponds to 138 fb^{-1} . Events with a diphoton invariant mass above 500 GeV are considered. A first-principles calculation of the standard model diphoton spectrum at next-to-next-to-leading order in perturbative quantum chromodynamics calculations was used to estimate the irreducible standard model background. The full data are used to constrain the large extra dimensions model of Arkani-Hamed, Dimopoulos, and Dvali, and the continuum clockwork mechanism. Preliminary studies and limits on the scenario of Unparticles decaying into two photons are also presented. Additionally, work on various hardware and software upgrades are discussed. Software support for improved timing information from the CMS Hadron Calorimeter Barrel for the creation of a Level-1 trigger for Long-Lived Particles are highlighted.

DEDICATION

*To Mama, Roel, Ate Jett, Dorothy, Shirin and Nani the Cat. Sa mga nagsisimula,
kumakayod at nagtatapos. Nawa ay marating mo ang iyong paroroonan.*

ACKNOWLEDGMENTS

The call to adventure to pursue a PhD was not for the faint of heart. It was a serious undertaking, rife with challenges and temptations that could only be surpassed with the support of many others. We have a proverb in the Philippines that says “Ang hindi lumignon sa pinanggalingan, ay hindi makararating sa paroroonan” which in English means, “Those who do not look back at where they came from will not reach their destination”. I am grateful to everyone whom I have crossed paths with along the way and helped me get to the finish line which once seemed so far off or kept moving an inch away with every half-inch I tried to get closer. I have arrived.

First, I’d like to express my deep gratitude to my advisor, Conor Henderson, for going above and beyond his role as an educator. He not only imparted technical and physics knowledge but also opened doors for me within the CMS experiment. He kept me grounded and focused throughout my PhD journey, providing unwavering support during challenging times. Looking back, I couldn’t have asked for a better adviser. I am forever grateful for our honest communication and his commitment to my well-being and success. I’m also thankful to Paolo Rumerio for his kindness and continuous support, despite having many roles to play. His presence and timely support was invaluable to me during my PhD. Sergei Gleyzer, my co-adviser, for his pep talks and sharing his experiences, which played a crucial role in helping me complete my dissertation. I’m grateful for the opportunity to learn machine learning from him and his team. I’ve also had the pleasure of having Dr. Corrinne Mills as my CMS mentor. Her inputs and connections helped guide me find my way when I was lost. She reminded me to maintain a personal life and not simply drown in physics responsibilities.

I owe an enormous debt of gratitude to the members of the diphoton analysis team,

including John Paul Chou, Antonios Agapitos, Hsin-Yeh Wu, and formerly Chris West, whose encouragement, expertise, agile working style and humor were instrumental in ushering the analysis in the last steps of the CMS publication journey. I truly think that I would not have finished the analysis without them. I also appreciate the contributions of Otmane Charaf, Andrew Buccilli, and Seth Cooper, who paved the way for me to make enormous progress in my PhD journey. Without them, there were no building blocks to start with. Working with Grace Cummings, Jay Dittman, Sebastian Yan, and Mary Hadley on the HCAL detector was a dream come true, and I'm thankful for the camaraderie and everything I have learned from them.

I extend my gratitude to the UA Physics Department, Patrick LeClair, Preethi Nair, Adam Hauser, and the rest of my dissertation committee members, Nobu Okada, Dawn Williams for the financial and moral support, physics discussions and feedback. I also owe a great deal to the HCAL Operations, CERN Social Affairs - particularly Sophie Charpie, HUG for their various forms of support during my extended stay in Geneva.

Beyond academia, I'm grateful for the sense of community I found in Tuscaloosa. I appreciate the support from friends and fellow Filipinos, including Ate Jett, Kuya Sonny Reece, Tita Marie, Tita Carms, Tita Doris, Ate E, Tiny, Orville, Carlo, Roel, Prabandha Nakarmi, and the Bengali Soccer Mates. Game nights/afternoons with Colin, Amy, Lars, Sreetama, Kavya, Anna, Eric were important respites. Upama, Anupama, Abhipsa and Ruole were invaluable supportive colleagues. I also thank Jillian Sico for her friendship and for expanding my conception of self. I would also like to thank Jennifer Turner for her counsel since 2019, which in some other universe could turn into a person-of-color version of "Goodwill Hunting" except I don't really beat people up.

My time in Geneva was enriched by the love and appreciation from people like Alexia

Speranza and her circle, Tita Ruby, Tito Ruben, Ilya, Hannah, Ome, Florian, Kathleen Heil, Jay, Mirana Randriambelonoro, Tricia Singh, Maria, Tita Violy. Izza, Tita Butch, Tito Antoine, Tita Mary, and Aaron were my family and gave me my best Christmas in a long time! I also want to thank Aditya and Manya for making me a better honorary Indian by being the best Indians! To Melissa and Christophe Gaillard, thank you for giving me the best summer experiences that I have ever had in my lifetime.

During my return to Tuscaloosa, I have survived a tough transitional period and lived like a vagabond. For this I am also thankful to Tim and Aggie, Butch and Ginger, Linda and Neil, Shirin and Chuck for having extended their homes and lives to a stranger, taught me their ways and let me go on my own path. Through them I met Dorothy Lott, whose kindness touched my heart and renewed my interest in life. Many thanks to your mechanical psychedelic!

Lastly, I want to express gratitude to André David, Delia, Michael Andrews, Afiq, Nuha, Miguel Sulangi, Roel and my dear Bananies - Karl, Bea, Pecier, Maribel, Joey, Tita Nens, Papa, my siblings Lilit, Nicole, and Marc, - for their unwavering support and love, which brought me to this point. My deepest appreciation goes to my Mama, my eternal source of strength, love and humor.

This work was supported in part by the US Department of Energy.

CONTENTS

ABSTRACT	ii
DEDICATION	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	xii
LIST OF FIGURES	xiv
CHAPTER 1 Introduction	1
CHAPTER 2 The Standard Model and Beyond	4
2.1 Standard Model of Particle Physics	4
2.2 Standard Model Pieces	5
2.2.1 QED: Electromagnetic Interaction	6
2.2.2 QCD: Strong Interactions	10
2.2.3 Electroweak Interactions	14
2.2.4 Higgs Mechanism and Spontaneous Symmetry Breaking	18
2.3 Limitations of the Standard Model	24
2.3.1 Hierarchy and Naturalness	25
2.4 Beyond Standard Model Non-resonant Signals	27
2.4.1 ADD Large Extra Dimensions	28
2.4.2 Clockwork/ Linear Dilaton	35
2.4.3 Unparticles Model	38
2.4.4 Previous Searches	42

CHAPTER 3	Experimental Setup: The Large Hadron Collider and CMS	45
3.1	The Large Hadron Collider	45
3.2	LHC Schedule, Beam Performance and Pileup	50
3.3	CMS Experiment and Its Subdetectors	52
3.3.1	ECAL	54
3.3.2	HCAL	57
3.3.3	CMS Data Trigger and Electronics	60
3.3.4	HCAL Online Software and HCAL Upgrades	63
CHAPTER 4	Particle Reconstruction and Event Selection	73
4.1	PF Elements and Physics Objects	74
4.1.1	Tracks and Vertices	75
4.1.2	Photon and Electron Reconstruction	76
4.1.3	Muon Reconstruction	77
4.1.4	Jets Reconstruction	78
4.2	Event Selection	78
4.2.1	Trigger Selection	79
4.2.2	Photon Identification	79
4.2.3	Offline Event Selection	83
4.2.4	Photon ID Efficiency and Scale Factors	84
4.2.5	Datasets	85
4.3	Endcap Region Issues	85
4.3.1	EE L1 Prefiring	90
4.3.2	HEM 15/16: Endcap Minus Power Outage	91
CHAPTER 5	Background Prediction	94
5.1	Monte Carlo Event Generation	94

5.1.1	Background Determination	97
5.2	Nonresonant Prompt Diphoton Background	98
5.3	Background from Fake Photons	105
5.3.1	Subtracting Real Photon Contamination in the Fake Rate	108
5.3.2	Fake Rate Numerator: Templates and Fitting	109
5.3.3	Fake Rate Denominator	113
5.3.4	Extracting the Fake Rate	113
5.3.5	Fake Rate Systematics	117
5.3.6	Fake Rate Application	121
5.4	Photon Fake Rate Closure Test	122
5.4.1	Testing the Fake Template from the Sideband Definition	123
5.4.2	Testing the Template Fit Method for Extracting Fake Yields	123
5.4.3	Testing that Reweighted Denominator Objects Reflect the Kinematics of Known Fake Objects	126
5.4.4	Conclusions from the Closure Test	130
CHAPTER 6	Signal Modelling	132
6.1	ADD Large ED Samples	132
6.1.1	Clockwork Model	136
6.1.2	Unparticles to Diphoton	140
CHAPTER 7	Sources of Systematic Uncertainties	143
7.1	Real Background Systematics	144
7.2	Fake Background Systematics	146
7.3	Signal Systematics	148
CHAPTER 8	Results	152
8.1	Bayesian Confidence Intervals	152

8.2	Exclusion Limits: ADD and Clockwork	155
8.2.1	Prefit Prediction	155
8.2.2	Limits and Postfits	157
8.2.3	ADD Postfit results	158
8.2.4	ADD Limits	161
8.2.5	CW Limits	163
8.2.6	Comparing CW results with ATLAS	165
8.3	Preliminary Exclusion Limits on Unparticles	169
8.3.1	Potential sources of additional sensitivity	170
8.3.2	Comparison with Z+MET search	170
CHAPTER 9	Conclusions	172
REFERENCES		174
APPENDIX A	Scale Factors and Datasets	185
APPENDIX B	Datasets and Trigger Paths	190
B.1	Data Samples	191
B.2	SM Prompt Diphoton Background Datasets	191
B.3	Fake Rate Datasets	194
B.4	Closure Test MC Datasets	194
APPENDIX C	Fake Rate Application	199
C.1	Signal Monte-Carlo	201
APPENDIX D	Sample Cards	213
D.1	ADD Graviton to $\gamma\gamma$	213
D.2	Unparticles to $\gamma\gamma$	214
APPENDIX E	Preliminary results using pseudodata	216
E.0.1	Clockwork Results with Pseudodata	217

E.0.2	ADD Results with Pseudodata	220
-------	---------------------------------------	-----

LIST OF TABLES

2.1	Previous exclusion limits results on the mass scale M_S in TeV for various conventions used in the calculation of the ADD large extra-dimensional scenario using the 2016 CMS detector data corresponding to an integrated luminosity of 35.9 fb^{-1} [71] only.	43
4.1	Selection details for the high- p_T photon ID v2. Numbers in [] are used if the crystal is saturated.	83
5.1	Fake rate definitions, where fake rate = numerator / denominator. The real and fake templates are fit to the numerator candidates using the template variable labeled ‘template’. After the fitting procedure is performed, the cut associated with the template variable is applied. The photon ID is shown for comparison.	107
5.2	There are three n_{PV} bins for 2016. The 28-200 bin are further split into 3 additional bins for 2017 and 2018. The ranges are $0 < n_{PV} \leq 22$ (left), $23 \leq n_{PV} \leq 27$ (center), and $28 \leq n_{PV} \leq 200$ (right). The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.	119
7.1	Systematic uncertainties and nuisances considered in non-resonant search. . .	143
8.1	Full CMS Run II exclusion limits on the mass scale M_S (in units of TeV) for various conventions used in the calculation of the ADD large extra-dimensional scenario using the 2016-2018 CMS detector data corresponding to an integrated luminosity of 138 fb^{-1}	164
8.2	Preliminary limits on the Unparticles	170
A.1	Nominal MC samples used for TnP	186
A.2	Single electron high level trigger	187
A.3	Datasets used for TnP	187
A.4	Tag and Probe Selection for 2017 and 2018	187
A.5	Alternative MC samples used for TnP	188
A.6	Final scale factors and uncertainties for 2016.	188
A.7	Final scale factors and uncertainties for 2017.	189

A.8	Final scale factors and uncertainties for 2018.	189
B.1	MINIAOD Data samples analyzed and the corresponding integrated luminosity, calculated with Double Photon Trigger and the processed luminosity sections.	192
B.2	Good run JSONs used for each data period.	192
B.3	MC background samples used in the analysis.	193
B.4	For each year, the campaign in which MC samples were generated and the global tag used to generate the analysis ntuples.	194
B.5	The JetHT Datasets used for constructing the fake templates and for deriving the fake rates. The fake rate is derived from the average of the rates obtained independently from the JetHT and DoubleMuon dataset	195
B.6	The DoubleMuon datasets for fake rate estimation. The fake rate is derived from the average of the rates obtained independently from the JetHT and DoubleMuon datasets.	196
B.7	2016 MC Fakes	197
B.8	2017 and 2018 MC Fakes	198
C.1	Nonresonant signal samples were generated for the Summer 2016, Fall 2017 and Autumn 2018 campaigns.	201
C.2	Unparticles samples were generated for the 2017/18 campaigns.	201
C.3	ADD signal samples and their associated cross-sections	202
C.4	ADD signal samples and their associated cross-sections	203
C.5	ADD signal samples and their associated cross-sections	204
C.6	ADD signal samples and their associated cross-sections	205
C.7	ADD signal samples and their associated cross-sections	206
C.8	ADD signal samples and their associated cross-sections	207
C.9	ADD signal samples and their associated cross-sections	208
C.10	Unparticles signal samples and their associated cross-sections	209
C.11	Unparticles signal samples and their associated cross-sections	210
C.12	Unparticles signal samples and their associated cross-sections	211
C.13	Unparticles signal samples and their associated cross-sections	212

LIST OF FIGURES

2.1	The Fundamental Particles of the Standard Model of Particle Physics.	7
2.2	The Standard Model of Particle Physics depicting the central role of the Higgs Boson [31].	8
2.3	QED Feynman diagram basic vertex showing an interaction between a fermion and an anti-fermion emitting a photon.	9
2.4	QCD Fundamental vertices where quark-gluon interactions (left) and gluon self-interactions (middle and right) are shown.	11
2.5	Electroweak Interaction Vertices. X are fermions.	15
2.6	Higgs Interaction Vertices.	19
2.7	The infamous “Mexican Hat” Higgs potential for the case of $\mu^2 < 0$ and $\lambda > 0$. There is a circle of minima of radius v	22
2.8	One-loop quantum corrections to the square of the Higgs mass parameter (m_H^2) due to a Dirac fermion f in (a), and a scalar (b) S [39].	26
2.9	LED Scenario Schematic where compactified extra dimension goes beyond our 4-dimensional spacetime where SM particles live. [53]	29
2.10	Mass spectrum for KK gravitons for $n = 2$	33
2.11	A CMS result [56] using data from $\sqrt{s} = 7$ TeV pp-collisions which shows a comparison of the invariant mass spectrum predicted for the Standard Model diphoton production in light blue, overlayed with a continuous spectra for two Large Extra Dimensions signal points shown in dashed lines. The observed data and additional background predictions from γ +jet and dijet events are also shown in various markers indicated in the legend.	34
2.12	Feynman Diagrams for virtual graviton or Unparticle exchange in the diphoton channel $\gamma\gamma$ through quark annihilation (left) and gluon fusion (right).	35
2.13	The “Clockwork gears” schematic representation of the clockwork mechanism where the gears represent the interacting fields. [48].	36
2.14	2016 95% CL exclusion limits results for the continuous clockwork mechanism over the $k - M_5$ parameter space.	43

2.15	A related report on the 95% CL upper limits on unparticle+Z [74] production cross section, as a function of the scaling dimension d_U	44
3.1	A bird's eyeview and and schematic depiction of the underground LHC accelerator complex and its four interaction points in Geneva, Switzerland and France [75].	46
3.2	The CERN Accelerator complex has a series of accelerators that ramp up particle beam energies to 6.5 TeV in Run II.	48
3.3	An overview of the LHC schedule from Run2 onwards before the COVID-19 pandemic delays.	50
3.4	An updated schedule of the LHC accounting for the delays due to the COVID-19 pandemic (Credits: LHC-Commissioning)	51
3.5	The mean number of pileup or inelastic interactions per proton bunch crossing in data for proton-proton (pp) collisions for various years (left). A total inelastic pp collision cross section of 69.2 mb is chosen [84]. The Luminosity is shown for various time with white spaces for periods the LHC is not operating (right).	52
3.6	The relationship between η and θ	53
3.7	A cutaway view of the CMS detector showing its components [86].	55
3.8	A cutaway view of the ECAL and its components [87].	57
3.9	A cutaway view of the HCAL and its subdetectors [87]	59
3.10	An illustration comparing the LHC-produced data without trigger selection to the other Big data players [89].	61
3.11	A cutaway view of the HCAL and its subdetectors [78]	62
3.12	Schematic diagram of the data flow in HB/HE starting from the photomultipliers through the QIE, FPGA and to the backend and AMC13 and to the CMS DAQ [90].	65
3.13	A schematic diagram of an HB RBX containing the 4 RMs, 2 ngCCMs and the backplanes [88]	66
3.14	Image of an RBX at Building 904 where a mock-up of the CMS HCAL Front-End electronics are set up for testing.	66
3.15	On the left is a new AMC13 tested at 904 with the μ HTRs in a selected Back-End Crate shown on the right.	67

3.16	The top picture shows an ngCCM being reworked to add heat sinks to mitigate the VTRx RSSI drift issue which was related to the communications loss in the HCAL endcap. The right shows a reworked ngCCM being tested on a spare backplane.	68
3.17	Reprogramming of the HF Igloo mounted on the HF ngCCMs at the CMS cavern.	69
3.18	TDC Current support. Credits: Tullio Grassi	69
3.19	With the Phase I Upgrade, each QIE channel now has a time-to-digital converter (TDC) which sends out a 6-bit TDC Code. Credit: E.Hughes	70
3.20	Sample LUT. The RM index ranges from 1-5, where the 5th is the Calibration Unit. Labeling credits: Gillian Kopp	71
3.21	An example hardmapping of 6-bit to 2-bit information by Tullio Grassi	71
4.1	The information from the vertices, tracks and calorimeter energy deposits are chained by the PF algorithm to form the various physics objects.	75
4.2	Efficiency of the Double Photon triggers (measured with reference trigger as a function of the p_T of the second-leading photon in 2016 (left), 2017 (center) and 2018 (right) data.	80
4.3	Prompt photons (REAL) and Fake Photons are shown on top and bottom. The figures on the left show the $\sigma_{i\eta i\eta}$ distributions for real and fake. On the right, a representative illustration of a prompt (fake) photon energy deposit in a 5×5 array of crystals is shown on top (bottom).	81
4.4	A comparison of the fitting results between nominal background model (top) and two alternative background models which are exponential (middle) and a second-order polynomial background models (bottom).	86
4.5	Extrapolation plots for 2016. (a,b,c) are the extrapolation fitting results. (d,e,f) are consistency checks. The extrapolated scale factors and uncertainties are drawn on top of the tag-and-probe measured results.	87
4.6	Extrapolation plots for 2017. (a,b,c) are the extrapolation fitting results. (d,e,f) are consistency checks. The extrapolated scale factors and uncertainties are drawn on top of the tag-and-probe measured results.	88
4.7	Extrapolation plots for 2018. (a,b,c) are the extrapolation fitting results. (d,e,f) are consistency checks. The extrapolated scale factors and uncertainties are drawn on top of the tag-and-probe measured results.	89
4.8	Data efficiency and scale factors (bottom) extracted from the tag-and-probe package are plotted vs p_T for three years.	90

4.9	Representative signal samples that land in the pseudorapidity range $2.25 < \eta < 3.0$ which is affected by the prefiring.	91
4.10	Representative background samples that land in the pseudorapidity range $2.25 < \eta < 3.0$, which is affected by the pre-firing.	92
4.11	Numbering scheme for the HCAL Wedges. Each wedge is 20 degrees.	93
4.12	No effects were seen on control region of the diphoton mass spectrum before and after the HEM15/16 issue.	93
5.1	Schematic representation of the structure of an event. The first diagram (a) is a representation of the oncoming protons or “bags of partons”. In (b), the main hard subprocess with the resonance decays are shown. The hard subprocess is represented by matrix elements calculations. Finally, the Parton Shower is shown in (c), where full event is shown with other more components such as the initial and final state radiation and beam remnants. Credits: Sjöstrand, Torbjörn	95
5.2	This image shows the Sherpa Event Generator Framework with the various elements that are needed to produce simulation, from user inputs, to matrix elements to additional interfaces.	96
5.3	Latest results of the global fit done by the NNPDF collaboration showing the PDFs at factorization scales of 10 GeV^2 and 10 GeV^4 . The different PDFs are related by the DGLAP evolution equations [121]	97
5.4	The irreducible and primary background for the analysis comes from the prompt standard model $\gamma\gamma$ production which occur via the $q\text{-}q\bar{q}$ annihilation (Born) or through the gluon fusion (Box) process as shown from left to right.	98
5.5	Sample Feynman Diagrams for (a) $q\bar{q}$ and (b) gq -initiated processes for NLO SM diphoton production are shown. [123]	99
5.6	Representative Feynman Diagrams for NLO diphoton production through quark or gluon fragmentation [123]	100
5.7	Shown here are the out-of-the-box SHERPA prompt simulated $m_{\gamma\gamma}$ prediction reweighted with the K-factor in light teal. The fake prediction is shown with a darker shade of teal.	102
5.8	NNLO k-factor for the $m_{\gamma\gamma}$ distribution for barrel-barrel diphotons (left) and barrel-endcap diphotons (right) for the real diphoton k-factor. The renormalization μ_r and factorization μ_f scales have been varied as follows: $\mu_r = \mu_f = 0.5m_{\gamma\gamma}$ (top), $1.0m_{\gamma\gamma}$ (middle), and $2.0m_{\gamma\gamma}$ (bottom). The cuts on the generator-level photons are the same as the kinematic cuts used offline. . . .	103

5.9	Ratio of diphoton cross sections calculated by MCFM versus the results from SHERPA in the barrel-barrel (left) and barrel-endcap (right) regions.	104
5.10	Schematic of a π^0 decaying into two photons that deposit energy that mimics energy-deposit features of prompt photons.	106
5.11	Representative real templates derived from MC in EB and EE for different p_T bins.	110
5.12	Fake templates comparison measured in various p_T bins extracted from the JetHT data set for EB (left) and EE (right).	111
5.13	Comparison of normalized real template shape among the three analysis years, for EB (left), EE1 ($1.566 < \eta < 2.033$)(center), and EE2 ($2.033 < \eta < 2.5$)(right).	111
5.14	Comparison of normalized fake template shape among the three analysis years, for EB (left), EE1 ($1.566 < \eta < 2.033$)(center), and EE2 ($2.033 < \eta < 2.5$)(right). The top (bottom) row shows the templates derived from the JetHT (DoubleMuon) dataset.	112
5.15	A representative template fit extracted from the 2018 JetHT data set using photon objects from the ECAL barrel in the $130 < p_T < 150$ GeV range, in linear (left) and log scale (right) for the y-axis.	112
5.16	A comparison of the fake rate in EB measured as a function of the number of selected primary vertices (nPV) in the event using the JetHT data set (top) and DoubleMuon data set (bottom). The fake rates are determined separately for 2016 at the (left), 2017 (center), and 2018 (right).	114
5.17	A comparison of the fake rate in EE1 ($1.566 < \eta < 2.033$) measured as a function of the number of selected primary vertices in the event using the JetHT data set (top) and DoubleMuon data set (bottom). The fake rates are determined separately for 2016 (left), 2017 (center), and 2018 (right).	115
5.18	A comparison of the fake rate in EE2 ($2.033 < \eta < 2.5$) measured as a function of the number of selected primary vertices in the event using the JetHT data set (top) and DoubleMuon data set (bottom). The fake rates are determined separately for 2016 (left), 2017 (center), and 2018 (right). The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.	116
5.19	A comparison of the 2016 fake rate in EB (top), EE1 ($1.566 < \eta < 2.033$)(middle), and EE2 ($2.033 < \eta < 2.5$)(bottom) measured as a function of the number of selected primary vertices in the event. The ranges are $0 < n_{PV} \leq 22$ (left), $23 \leq n_{PV} \leq 27$ (center), and $28 \leq n_{PV} \leq 200$ (right). The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.	118

5.20	A comparison of the 2017 fake rate in EB (top), EE1 ($1.566 < \eta < 2.033$)(middle), and EE2 ($2.033 < \eta < 2.5$)(bottom) measured as a function of the number of selected primary vertices in the event. From left to right, the ranges are $0 < n_{PV} \leq 22$, $23 \leq n_{PV} \leq 27$, $28 \leq n_{PV} \leq 32$, $33 \leq n_{PV} \leq 37$, and $38 \leq n_{PV} \leq 200$. The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.	119
5.21	A comparison of the 2018 fake rate in EB (top), EE1 ($1.566 < \eta < 2.033$)(middle), and EE2 ($2.033 < \eta < 2.5$)(bottom) measured as a function of the number of selected primary vertices in the event. From left to right, the ranges are $0 < n_{PV} \leq 22$, $23 \leq n_{PV} \leq 27$, $28 \leq n_{PV} \leq 32$, $33 \leq n_{PV} \leq 37$, and $38 \leq n_{PV} \leq 200$. . .	120
5.22	Representative $\sigma_{i\eta i\eta}$ templates for the ECAL barrel (left) and endcap (right) for 2017 in the 50-70 GeV p_T -bin are shown here. The distributions in gold show the MC truth while the black and red show the fake templates in the $5 < \text{Iso}_{\text{Ch}} < 10$ GeV and $10 < \text{Iso}_{\text{Ch}} < 15$ GeV sidebands.	124
5.23	A representative template fit from the 2018 shows using photon objects from the ECAL barrel that have from 50 to 70 GeV in the Iso_{Ch} sideband $5 < \text{GeV} < 10$	125
5.24	The 2017 MC Truth fake rates and the MC fake rates shows a generally good agreement with each other for both EB and EE. The MC Truth was obtained by a 'counting' procedure whereas the fake rate was obtained through a series of steps which involves a template fitting.	125
5.25	The 2018 MC Truth fake rates and the MC fake rates shows a generally good agreement with each other for both EB and EE. The MC Truth was obtained by a 'counting' procedure whereas the fake rate was obtained through a series of steps which involves a template fitting.	126
5.26	The two columns show the EB (left) and EE (right) single photon kinematic plots for reweighted objects for 2017.	127
5.27	The two columns show the EB (left) and EE (right) single photon kinematic plots for reweighted objects for 2018.	128
5.28	A representative template fit from the 2017.	129
5.29	The Fake Rates were recalculated with the splitting of the barrel region into inner and outer regions. The inner barrel (EB1) corresponds to $ \eta < 0.7221$. The outer barrel (EB2) corresponds to $0.7221 < \eta < 1.4442$. This was done to improve the agreement between the fake prediction and the MC Truth fake prediction for the last step of the closure test.	129
5.30	We also split the endcaps into inner and outer regions to improve the fake prediction. The inner endcap (EE1) $1.566 < \eta < 2.033$ while the outer endcap corresponds to $2.033 < \eta < 2.5$	130

5.31	Splitting the barrel and endcap regions shows some improvements in the agreement between the fake rate reweighted Fake Prediction and the MC Truth Prediction for 2017. A similar result is observed for 2018. The inner barrel (EB1) corresponds to $ \eta < 0.7221$. The outer barrel (EB2) corresponds to $0.7221 < \eta < 1.4442$. The inner endcap (EE1) $1.566 < \eta < 2.033$ while the outer endcap corresponds to $2.033 < \eta < 2.5$	131
6.1	The $m_{\gamma\gamma}$ distribution in signal events for the generated values of Λ_T before (left) and after (right) subtraction of SM background. The upper(lower) plots show signals generated assuming positive (negative) interference with the SM background. The plots were generated using the 2017 ADD samples and are normalized to an integrated luminosity of 1 fb^{-1}	134
6.2	A representative ADD sample signal agreement with the quark-antiquark background (light blue) alone in the low $m_{\gamma\gamma}$ region.	136
6.3	Signal shapes for the clockwork signal after the final selection for a fixed value of $M_5 = 10 \text{ TeV}$ and varying values of k normalized to the total integrated luminosity for EB-EB events in the left and EB-EE events in the right, respectively.	139
6.4	Leading-order Feynman diagrams for the Unparticles to two photons.	140
6.5	The 2D-plane with the parameters of the Unparticles model. The 95% CL upper limits on signal strength (y-axis) vs the Λ_U (x-axis) and the lower limits on Λ_U for the unparticles model. On the left spin 0, on right spin 2. Top to bottom, d_u parameter value: 1.1, 1.5 and 1.9 (Credits to Antonios Agapitos).	141
6.6	The $m_{\gamma\gamma}$ distribution in unparticle signal events for the generated values of Λ_U . The upper (lower) plots show signals generated assuming spin-0 (spin-2) $du = 1.9$. The left (right) plots show before (after) SM background subtraction. These plots were generated using the 2018 Unparticles samples.	142
7.1	The resulting uncertainties from first 25 of the 50 variations of the PDFs eigenvectors. The variation over central estimate is shown versus the $m_{\gamma\gamma}$. (Plots shown stand for BB, while the corresponding for BE, as they are very similar, have been omitted).	146
7.2	The resulting uncertainties from the last 25 of 50 variations of the PDFs eigenvectors. The variation over central estimate is shown versus the $m_{\gamma\gamma}$. (Plots shown stand for BB, while the corresponding for BE, as they are very similar, have been omitted).	147
7.3	The uncertainty on the real background that arises from PDFs (parton distribution function) uncertainties, displayed separately for the barrel-barrel and barrel-endcap regions, for evaluations at LO (left) and NLO (right).	148

7.4	Top: NNLO (MCFM) k-factors for different factorization and renormalization (QCD) scales. Bottom: Ratio of these variations over the central value as a function of $m_{\gamma\gamma}$. Plots for the BB (barrel-barrel) on the left, and BE (barrel-endcap) on the right.	149
7.5	The effect of “fake sample” systematic uncertainty over the fake background $\gamma j, jj$ vs the $m_{\gamma\gamma}$ spectra. Top (bottom) the BB (BE) channels, left to right for years 2016, 2017, and 2018 respectively.	150
7.6	The left shows the systematic uncertainty trends of “variation/central” ratios of real background vs $m_{\gamma\gamma}$ for the sources of: “efficiency”, and right for “energy scale gain”. Top to bottom: the six SRs of BB16, ..., BE18 respectively. . . .	151
8.1	The $m_{\gamma\gamma}$ spectra in the prefit estimate for 2016; left the BB and right the BE channels. These are the six observables used in the fit. The final prediction is derived from a fit to the data, a procedure which absorbs residual mismodeling present at these plots.	156
8.2	The $m_{\gamma\gamma}$ spectra in the prefit estimate for 2017; left the BB and right the BE channels. These are the six observables used in the fit. The final prediction is derived from a fit to the data, a procedure which absorbs residual mismodeling present at these plots.	156
8.3	The $m_{\gamma\gamma}$ spectra in the prefit estimate for 2018; left the BB and right the BE channels. These are the six observables used in the fit. The final prediction is derived from a fit to the data, a procedure which absorbs residual mismodeling present at these plots.	157
8.4	The ADD GRW postfit spectra (with real data) for BB (left) and BE (right) for 2016.	158
8.5	The ADD GRW postfit spectra (with real data) for BB (left) and BE (right) for 2017.	159
8.6	The ADD GRW postfit spectra (with real data) for BB (left) and BE (right) for 2018.	159
8.7	The prefit (top) and the postfit (bottom) spectra with real data for BB (left) and BE (right) for the three years of data taking inclusively. The plot shown in red at the bottom shows a representative signal plot for the ADD model with the GRW convention.	160
8.8	The postfit probability distribution function (pdf) of each nuisance used for the fits for each of the 6 signal regions that have been fitted. The upper right panel corresponds to the statistics panel indicating the mean and standard deviation of the pdfs.	162

8.9	The 95% CL limits on signal strength (y-axis) vs the M_S (x-axis) and the limits on M_S in legend for the ADD GRW model left, and Hewett negative right. These results are using real data.	163
8.10	Exclusion limit for the clockwork framework over the $k - M_5$ parameter space. The shaded region denotes where the theory becomes non-perturbative. The right plot displays a non-logarithmic axis.	165
8.11	The CW prefit (top two rows) and the postfit (bottom two rows) $m_{\gamma\gamma}$ spectra. Columns left to right correspond to the years 2016, 2017, and 2018 respectively.	166
8.12	The postfit pdfs for the 25 nuisance parameters used in the six SRs fit with CW model signal.	167
8.13	The previously published [131] Run II, 2016-only 95% CL exclusion limit for the clockwork model in the $k - M_5$ parameter space. The green (± 1) and yellow (± 2) shaded bands represent the standard deviation uncertainties in the expected limit (dashed lines). The shaded region in purple at the lower right is where the theory becomes nonperturbative.	168
8.14	ATLAS [73] Run II exclusion limits with mass cut and without mass cut shown with flipped CMS results for comparative purposes.	168
8.15	Limits on the Unparticles signal strength as a function of Λ_U . The top plots show the Spin-0 case for $d_u = 1.1, 1.5$, and 1.9 . The bottom shows the Spin-2 case.	169
8.16	Top: Spin-0 (red) and spin-2 (blue), $\mathcal{U} \rightarrow \gamma\gamma$ generator-only histograms with the same $\Lambda_u = 1000$ GeV and $d_u = 1.1$. The middle plot shows various d_u values with a fixed $\Lambda_u = 2500$ GeV. On the bottom, is a plot that has a fixed $d_u = 1.8$ GeV, and Λ is varied. The leading-order Standard Model Prompt Diphoton Background is shown in cyan.	171
8.17	This figure shows $\cos\theta^*$ plots of $U \rightarrow \gamma\gamma$ versus the leading order SM only background (cyan). The right plot shows a different d_u for spin-2, curved differently from the SM background.	171
E.1	CMS Collaboration papers go pre-approval, approval, a Collaboration-Wide Review, and a final reading prior to submission to a peer-reviewed journal [147].	217
E.2	The prefit spectra with the generated pseudodata for BB (top) and BE (bottom) for the tree years of data taking (left to right: 2016, 2017, 2018).	218
E.3	The prefit (top two rows) and the postfit (bottom two rows) $m_{\gamma\gamma}$ spectra for the Clockwork Model using pseudodata. Columns left to right correspond to the years 2016, 2017, and 2018 respectively.	221

E.4	The postfit probability distribution functions (pdfs) for the first 25 nuisances of the parton distribution functions (PDFs) in the six-SRs fit. The statistics panel displays the mean and the std of each pdf. This set of results corresponds to the pseudodata.	222
E.5	The postfit probability distribution functions (pdfs) for the last 25 nuisances of the parton distribution functions (PDFs) in the six-SRs fit. The statistics panel displays the mean and the std of each pdf. This set of results corresponds to the pseudodata.	223
E.6	The postfit pdfs for the 25 nuisance parameters used in the six SRs fit with CW model signal using pseudodata. (We omit the illustration of the other 50 of PDFs pdfs as they are gaussian-like without particular interest.)	224
E.7	The exclusion limit for the continuous graviton model in the clockwork framework over the $k - M_5$ parameter space using pseudodata. The shaded region denotes where the theory becomes non-perturbative. (The right plots is the same as the left one but in non-log axis). The result from the fit over all six SRs (collectively over all years and channels).	225

CHAPTER 1

INTRODUCTION

Science as a core human pursuit, aims to understand the natural world through observation, experimentation and analysis. The Large Hadron Collider, with its 27-kilometer ring of magnets, enable us to accelerate particles to nearly the speed of light and collide them with energy to replicate the conditions of the early universe. It is now known as the largest experiment mankind has ever built to probe the fundamental constituents of matter and the forces governing their interactions. The scientific, computational and engineering challenges bring about benefits to society beyond the understanding of the cosmos through technical innovations from the world-wide web to novel medical devices. Most importantly, it serves as a model for large-scale multinational collaboration that transcends geopolitical boundaries.

On July 4, 2012, two of LHC's independent experiments, ATLAS and CMS, announced the discovery of a new fundamental particle consistent with the Higgs Boson. This discovery, predicted by theorists half a century earlier, filled a critical gap in the Standard Model (SM) of Particle Physics. Despite the LHC's primary mission to discover the Higgs Boson, it continues to operate, supported by global investment in upgrades. With over 20,000 scientists and engineers worldwide, research at the LHC aims to explore new physics. While the Higgs discovery provided insights into particle mass acquisition within the Standard Model, fundamental issues persist. The SM hierarchy problem, highlighted by the disparity between the electroweak and Planck scales, remains unresolved. This discrepancy, along with the significant weakness of gravity compared to the electroweak force, motivates exploration of Beyond-the-Standard-Model (BSM) frameworks to address foundational questions and

provide a quantum description of gravity.

It is in this light that we continue to explore ideas that fundamentally rethink the structure of the universe as we know it. These ideas include extra dimensions with various geometries, and a distinct but related scenario in the form of the existence of a yet-to-be-seen scale invariant sector where strange hazy objects called Unparticles live. In this dissertation, discuss the measurable physical consequences of these new theoretical frameworks. We search for deviations from the predictions of the Standard Model in high-mass $\gamma\gamma$ events that could be characterized by these models. Here we focus on reporting our search for broad, non-resonant signatures which are expected from large extra spatial dimensions, the Continuum Clockwork mechanism and the Unparticles model. Large extra spatial dimensions models proposed by Arkani-Hamed, Dimopoulos and Dvali (ADD) [1–3] modify the fundamental Planck scale to be the same order as the electroweak scale. In their models, the SM fields are constrained to the usual 3+1 spacetime dimensions while gravity can also propagate in the n_{ED} additional, compactified spatial dimensions. With these, the effective strength of gravity is reduced by a Gauss’s law reduction in the flux. Not included in this document is the complementary search for the Randall and Sundrum (RS) [4, 5] model with a 1D-warped extra dimension, where the signal is a resonant bump. Both ADD and RS scenarios, Kaluza–Klein (KK) modes of the graviton couple to the SM through the stress-energy tensor and decay into two SM particles. In ADD, the KK modes are closely spaced and result in a continuum or broad excess of diphoton events over the expected SM background. In the RS graviton model, the KK modes are on-shell and present as resolvable resonances in the diphoton mass spectrum.

Another proposed solution to the SM hierarchy problem is the continuum clockwork mechanism [6]. In the continuum limit of the clockwork, the massless graviton is accompanied by an infinite tower of massive spin-2 graviton KK modes with a characteristic pattern of masses and couplings. Like in the ADD extra dimension scenario, an approximately continuous distribution of KK modes results in a continuum excess [7, 8]; however, in the clockwork

scenario, the KK modes are all on-shell, while in the ADD case, virtual contributions to the diphoton spectrum are significant [9–11].

Searches on these phenomena span over 10 years of work from various independent experiments. Previous searches for BSM physics in high-mass diphoton events from Run 2 of the CERN LHC were performed by the ATLAS and CMS experiments using proton-proton collisions at the same center-of-mass energy $\sqrt{s} = 13$ TeV [12–17]. Prior searches have also been performed by both experiments in Run 1 of the LHC [18–22], at $\sqrt{s} = 7$ and 8 TeV, as well at the Fermilab Tevatron by the CDF [23–26] and D0 [27–30] collaborations using proton-antiproton ($p\bar{p}$) collisions at $\sqrt{s} = 1.96$ TeV.

This dissertation also includes notes on the author’s on-site contributions to upgrading the CMS Hadron Calorimeter (HCAL) from 2020 to 2022. These include upgrades in the HCAL Online Software (HCOS) to support the creation of a trigger for Long-Lived Particles as well as miscellaneous hardware upgrades. This contribution represents just one small part of the broader initiative to enhance the detector’s sensitivity and longevity. Beyond the scope of this dissertation are multiple other efforts to uncover new physics in more complex, unexplored final states, improving detector sensitivity, and ensuring uninterrupted data-taking.

The remainder of this dissertation is organized as follows. Chapter 2 introduces the theoretical basis of the Standard Model, its limitations, and the theoretical frameworks with non-resonant signatures - the ADD Large Extra Dimensions, Clockwork/Linear Dilaton and the Unparticles, that we seek to explore. Chapter 3 gives a description of the experimental setup and the some of the upgrades implemented in the Large Hadron Collider. Chapters 5 and 6, discuss the Monte Carlo Background and Signal Modelling, respectively. We discuss the systematic uncertainties involved in the experiment in Chapter 7. In Chapter 8, we discuss the statistical interpretation of our findings and conclude the document in Chapter 9.

CHAPTER 2

THE STANDARD MODEL AND BEYOND

In this chapter we provide the theoretical basis for our analysis. The first section discusses the standard model (SM), which provides the best description of particle phenomena up to today. Despite its known success, there are energy domains in which the SM shows its limitations. In this light, extensions to the theory collectively known as Beyond Standard Model Physics, are proposed. We talk more about these limitations in Section 2.3. In Section 2.4, we discuss three BSM extensions to the Standard Model, the ADD Large Extra Dimensions 2.4.1, the Clockwork Model 2.4.2 and the Unparticles 2.4.3. These models all phenomenologically exhibit non-resonant excesses over the Standard Model background.

2.1 Standard Model of Particle Physics

The Standard Model of Particle Physics is a theory that describes the dynamics of fundamental particles and their interactions with one another. This quantum field theory is the most successful physics theory to-date with its remarkable consistency with a vast range of experimental results, save for some few issues that imply that it is still work-in-progress.

The mathematical framework of the Standard Model is a product of several decades worth of scientific inquiry, experimental validation and theoretical refinement which accounts for the zoo of particles discovered in laboratory experiments and also predicts the existence of yet-to-be discovered particles. The model, in its current form, categorizes the particles into three generations, each exhibiting distinct properties and masses. It is essentially a quantum field theory (QFT), which is an extension of the quantum mechanics into the relativistic

regime handling a variety of particle spins. Here particles are interpreted as excitations of gauge fields. The Standard Model describes three of the four fundamental forces: the electromagnetic, the weak and strong nuclear forces. Mathematically, this is described as the crossing of three symmetry groups:

$$\text{SU}(3)_c \times \text{SU}(2)_L \times \text{U}(1)_Y \quad (2.1)$$

Gravity, the final force, remains glaringly absent and is among one of the open questions which drives us into beyond Standard Model Physics extensions. These challenges are further described in Section 2.3.

At the very core, these symmetries 2.1 must obey the conservation laws¹ which dictate the rules of particle interactions. This dynamic is encoded in the Lagrangian. A simplified scenario where fundamental forces and particles exhibit a higher degree of symmetry and unity than what we observe in the universe today, the SM Lagrangian has the form:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}}. \quad (2.2)$$

Here $\mathcal{L}_{\text{EW}}$ refers to the electroweak interactions, $\mathcal{L}_{\text{QCD}}$ describes the strong interactions, and $\mathcal{L}_{\text{Higgs}}$ contains the Higgs field interactions with the gauge fermion fields and their acquisition of particle masses. However, our universe no longer exists in this form; it has undergone symmetry-breaking processes leading to seemingly distinct particles, forces, and theoretical components of physics, which we will describe separately in the next section.

2.2 Standard Model Pieces

Fig. 2.1 and fig. 2.2 shows the fundamental particles of the Standard Model. The SM consists of fermions and bosons. Fermions are the matter particles which consist of quarks and leptons. In the representation in fig. 2.1, the fermions are organized in three generations

¹Conservation of Energy and Momentum, electric charge, baryon number, flavor and non-conservation of flavor in weak interactions

in order of increasing mass. Leptons which include the e, μ, τ and their associated neutrinos ν_e, ν_μ, ν_τ are entities that do not experience the strong interaction and they possess a spin of $1/2$. Charged leptons have an electric charge of -1 while the neutrinos are chargeless. In contrast, quarks experience the strong force, and participate in the rest of the SMs fundamental interactions. Quarks also exhibit a unique property known as “flavor”, which distinguishes them into six distinct types: up (u), down (d), charm (c), strange (s), top (t) and bottom (b) quarks. These flavor distinctions play a pivotal role in shaping the properties of hadrons - which are composite subatomic particles made of quarks. The up-type quarks, u, c, t have a $+\frac{2}{3}$ electric charge while the down-type quarks d, s, b have a $-\frac{1}{3}$ electric charge.

In fig. 2.2, the gauge bosons, drawn in orange, have the photon, gluon and the $W^\pm$ and Z bosons which mediate the electromagnetic, strong and the weak forces, respectively. At the center, is the spin-0 Higgs boson which is the only scalar particle in the SM. The 2nd diagram in Fig. 2.2 reflects the central role of the Higgs boson in imbuing mass to the vector W and Z bosons fermions, except for neutrinos, and for the Higgs itself.

2.2.1 QED: Electromagnetic Interaction

QED or Quantum Electrodynamics is a quantum field theory component of the Standard Model that merges the principles of quantum mechanics with electrodynamics. QED at its core describes the interactions between electrically charged particles via photon exchange and it is described by a basic QED vertex shown in Fig. 2.3.

The fundamental principle of QED revolves around the preservation of electric charge which implies a U(1) symmetry. It is also an example of an abelian gauge theory which is a type of QFT where the gauge transformations commute.

We consider the Lagrangian density of a free Dirac field, $\mathcal{L}_D$ given by:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi. \quad (2.3)$$

Figure 2.1: The Fundamental Particles of the Standard Model of Particle Physics.

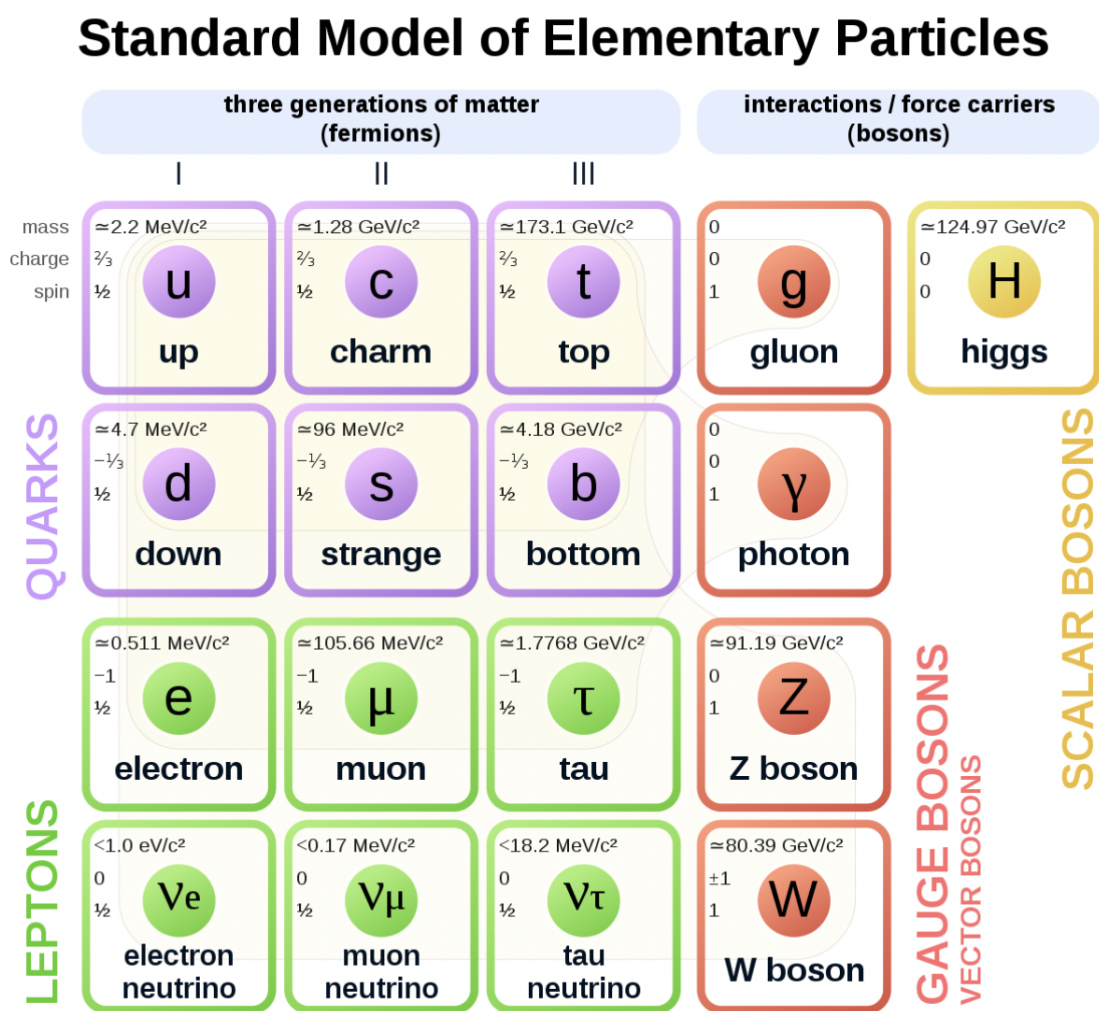


Figure 2.2: The Standard Model of Particle Physics depicting the central role of the Higgs Boson [31].

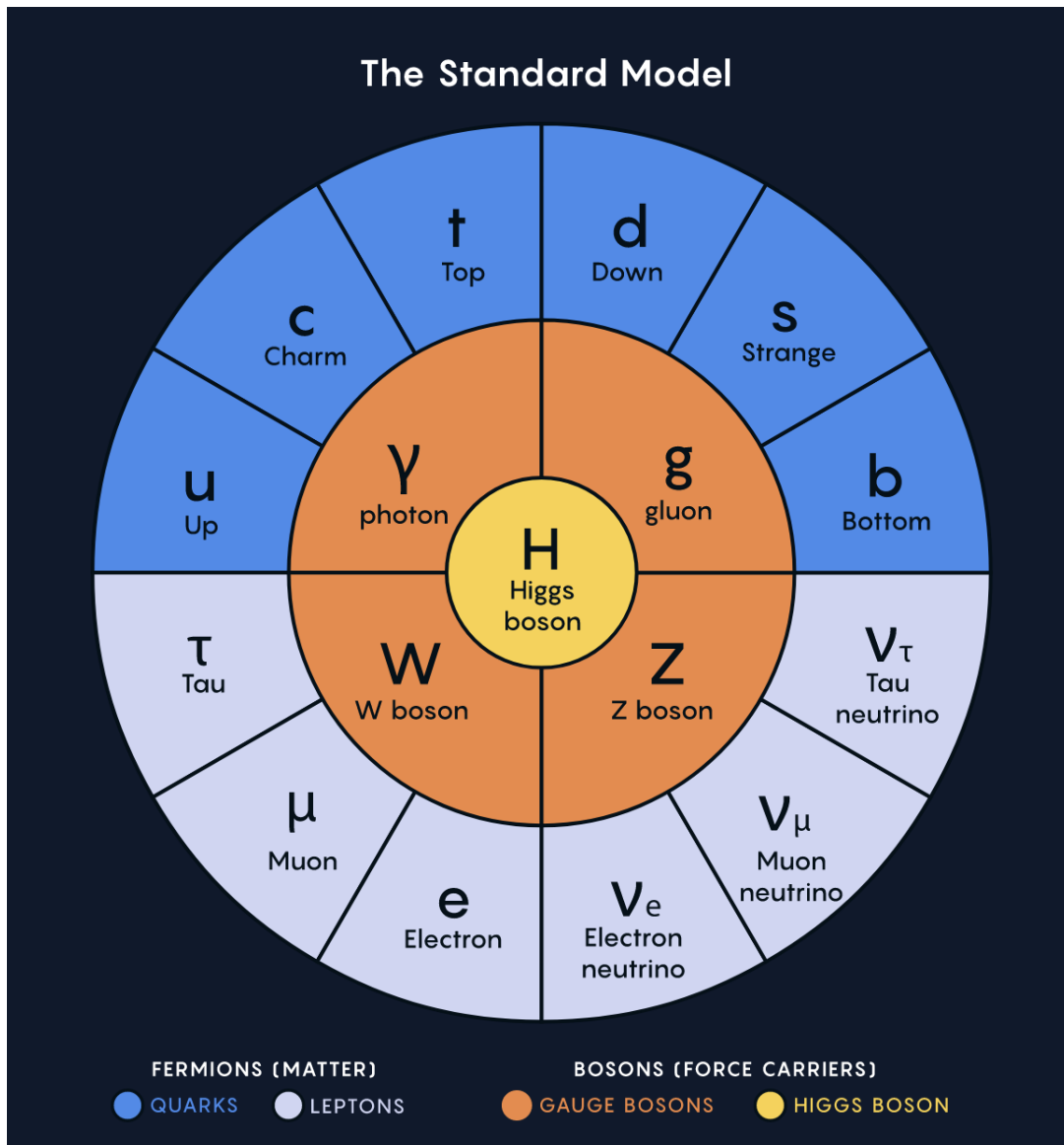
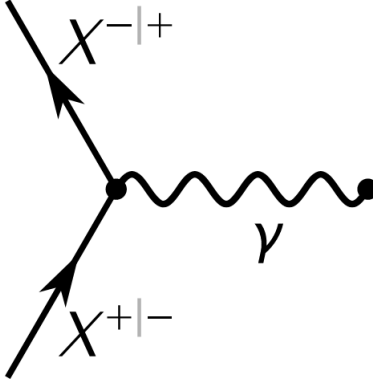


Figure 2.3: QED Feynman diagram basic vertex showing an interaction between a fermion and an anti-fermion emitting a photon.



Here, γ^μ are the Dirac matrices, and ψ are the spin-1/2 fields. This Lagrangian has a global $U(1)$ Abelian symmetry, s.t. it remains invariant under this transformation,

$$\psi(x) \rightarrow e^{i\alpha}\psi(x). \quad (2.4)$$

A local transformation on the other hand, where α is replaced with $\alpha(x)$, requires a change in the derivative ∂_μ with the covariant derivative defined by

$$D_\mu = (\partial_\mu - igA_\mu), \quad (2.5)$$

where ∂_μ is the partial derivative with respect to the spacetime coordinate x^μ , g is the coupling constant, and A_μ is the gauge field. Substituting Eq. 2.5 into Eq. 2.3, we get a resultant Lagrangian that is invariant under the

$$A'_\mu = A_\mu - \frac{1}{g}\partial_\mu\alpha(x), \quad (2.6)$$

and the local transformation in eq. 2.4. Here A_μ is referred to as the photon field. These steps could be extended to non-Abelian symmetries, where transformation operators do not commute, whose treatment is explained in more detail in standard gauge theory texts such as Ref [32].

The expression for the QED Lagrangian Density is as follows:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - Q\bar{\psi}\gamma^\mu\psi A_\mu, \quad (2.7)$$

where ψ is a fermion field and Q is the electric charge. Each term of the Lagrangian can be understood as the Dirac Lagrangian 2.3 plus the Maxwell's terms plus the fermion field and photon field interaction. This new covariant derivative introduces A_μ as a new vector field which we will see as follows is nothing but the electromagnetic four vector potential. Both A_μ and D_μ must be invariant under the $U(1)$ transformation as in Eq. 2.4 and it follows that the commutator of the covariant derivative is also invariant under the same transformation. Expanding the commutator, we find that we can rewrite it in terms of the Electromagnetic field tensor $F^{\mu\nu}$:

$$\begin{aligned} [D_\mu, D_\nu] &= iQ(\partial_\mu A_\nu - \partial_\nu A_\mu) \\ &= iQ(\partial_\mu A_\nu - \partial_\nu A_\mu) \\ &= iQF^{\mu\nu}. \end{aligned}$$

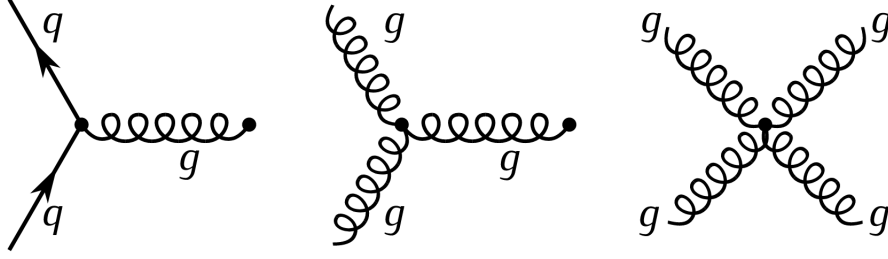
We can now write $\mathcal{L}_{\text{QED}}$ in the gauge invariant form by substituting the partial derivative with the covariant derivative in Eq. 2.5. The QED Lagrangian can be recast simply as:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (2.8)$$

2.2.2 QCD: Strong Interactions

Quantum Chromodynamics (QCD) is the segment within the framework of the Standard model that explains the dynamics of interactions between color-charged particles namely quarks and gluons. This is mathematically captured through transformations within the $SU(3)$ symmetry group to which three distinct charge quantum numbers, namely red, green,

Figure 2.4: QCD Fundamental vertices where quark-gluon interactions (left) and gluon self-interactions (middle and right) are shown.



and blue are introduced. This stands in stark contrast to QED, which involves a single electric charge. Initial experiments unveiled a crucial aspect of the quantum field theory governing quark interactions - the strength of the interactions decreases as the momentum of the interacting particles escalates. This behaviour is a hallmark of non-Abelian gauge theories of which QCD belongs to.

The mathematical structure of quantum chromodynamics can be treated as an extension of the idea of local gauge invariance in QED, where the $U(1)$ gauge group as in Eq. 2.4 is replaced by the $SU(3)$ group of phase transformations on the quark color field:

$$q \rightarrow e^{ig_s T^a} q, \quad (2.9)$$

where q is the quark field, and g_s is the strong coupling constant.

QCD is a non-abelian gauge theory due to the non-commutative generators within this group. The generators of the $SU(3)$ group are as follows:

$$[T^a, T^b] = if^{abc} T^c \quad (2.10)$$

where $T^a = \lambda^a/2$, λ^a are the Gell-Mann matrices and f^{abc} are the structure constants which are antisymmetric coefficients that govern the commutation relations between the Gell-Mann matrices describing how gluons, the force carriers of the strong interaction, interact with each other via the color charges of quarks. The Gell-Mann matrices consist of a

collection of eight 3x3 matrices that are Hermitian and have a trace of zero.

In a similar vein as in Eq. 2.6, we need to introduce the eight gauge fields G_μ^a , each transforming as:

$$G_\mu^a \rightarrow G_\mu^{a'} - \frac{1}{g} \partial_\mu \alpha_a \quad (2.11)$$

From this we can form a covariant derivative,

$$D_\mu = \partial_\mu + igT^a G_\mu^a. \quad (2.12)$$

This introduces 8 gluon fields, G_μ^a , each corresponding to permissible color combinations within the group. Plugging this in the free Lagrangian akin to the Dirac Lagrangian in Eq. 2.3,

$$\mathcal{L} = \bar{q}_j (i\gamma^\mu \partial_\mu - m) q_j, \quad (2.13)$$

where q_i , $i = 1, 2, 3$ denote the three color fields. The QCD analogue of Eq. 2.8 then becomes,

$$\mathcal{L} = \bar{q}(i\gamma^\mu \partial_\mu - m)q - g(\bar{q}\gamma^\mu T_\alpha q)G_\mu^a. \quad (2.14)$$

This is however still not a fully gauge invariant Lagrangian as the last term transformation, since we already from Eq. 2.10 that the Gell-Mann matrices don't commute. To mitigate this problem an extra term must be added to Eq. 2.11,

$$G_\mu^a \rightarrow G_\mu^a - \frac{1}{g} \partial_\mu \alpha_a - f_{abc} \alpha_b - G_\mu^a. \quad (2.15)$$

We then have the final gauge invariant QCD of the form:

$$\mathcal{L} = \bar{q}(i\gamma^\mu \partial_\mu - m)q - g(\bar{q}\gamma^\mu T_\alpha q)G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}, \quad (2.16)$$

or in terms of the covariant derivative,

$$\mathcal{L}_{\text{QCD}} = \bar{q}i\gamma^\mu D_\mu q - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}, \quad (2.17)$$

Gluons can directly partake in interactions involving their own color properties. One can expand the QCD Lagrangian 2.17 in terms of the transformation in Eq. 2.15, and write in symbolic form in terms of increasing orders of “G” [33],

$$\mathcal{L}_{\text{QCD}} = \text{“}\bar{q}q\text{”} + \text{“}G^2\text{”} + g\text{“}\bar{q}qG\text{”} + g\text{“}G^3\text{”} + g^2\text{“}G^4\text{”}. \quad (2.18)$$

The first two terms describe the free propagation of quarks and gluons. The last three terms are represented schematically in Fig. 2.4. The first vertex illustrates the interaction between quarks and gluons, while the subsequent two shows 3- and 4-point self-interactions of gluons.

Gluons’ ability to self-interact directly in Quantum Chromodynamics (QCD) results in a significant QCD feature: the strengthening of the coupling between quarks with increasing distance. This phenomenon leads to the preferred formation of new quark-antiquark pairs from the vacuum when attempting to separate bound quarks, causing the creation of two mesons instead of isolated quarks. This confinement of quarks, known as confinement, prohibits the observation of bare quarks in nature, allowing only color-neutral QCD states. Consequences of confinement are evident at high-energy colliders, where high-energy quark or gluon generation results in chains of hadrons forming hadronic jets through fragmentation or hadronization. These jets are pivotal signatures in particle detectors like CMS. At lower energies, scattering interactions lack the energy for jet formation, yielding a spread of hadrons, prevalent in background processes at colliders like the LHC. The top quark, with a short lifetime, decays weakly, sidestepping hadronization, making it a study candidate for “bare” quarks. Confinement introduces computational challenges, particularly in the non-perturbative QCD regime below the confinement scale, where interaction predictions

rely on heuristic tools. Above this scale, asymptotic freedom manifests, with quarks acting as free particles at high energies, valid for hard scattering interactions at colliders through perturbation theory.

2.2.3 Electroweak Interactions

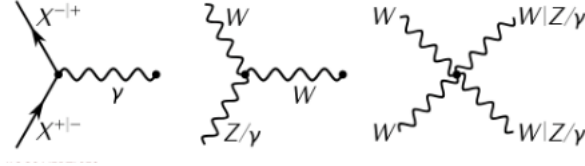
The final piece of the SM, the weak interactions, is an extension of QED which describes how particles charged under weak isospin interact with the $W^\pm$ and Z bosons. QED was initially developed to describe the interactions of electrically charged fermions. Following spontaneous symmetry breaking, the photon gains the ability to interact with the electrically charged $W^\pm$ bosons as well. Additionally, QED on its own, can function independently as a renormalizable, Abelian gauge theory. However the weak interaction involving the $W^\pm$ and Z bosons cannot be adequately described by a standalone gauge theory. It has been determined that these three weak interaction vector bosons and the massless photon in QED need to be collectively addressed within the framework of an $SU(2) \times U(1)$ symmetry in order to explain the observed weak force in the natural world.

The electroweak theory has $SU(2) \times U(1)$ gauge group which have the following fundamental fields: $W_\mu^1, W_\mu^2, W_\mu^3$, associated with the three generators of $SU(2)_L$ and B_μ , associated with $U(1)_Y$. The force mediators, $W^\pm$, Z and γ are orthogonal combinations of these fundamental fields, which we will show later.

As $SU(2)$ is also a non-Abelian theory, so the weak gauge bosons endowed with charges related to weak isospin, like how gluons bear color charges in QCD. Consequently, they exhibit self-interactions including cubic and quartic couplings of the form $WW\gamma$, WWZ , $WWWW$, $WW\gamma\gamma$, $WWZZ$, and $WWZ\gamma$. The mixing among the original states of $SU(2) \times U(1)$ leads to distinct outcomes that distinguishes the electroweak sector with QCD, the other non-Abelian theory. These unique consequences are essential for aligning theoretical predictions with experimental observations. The theory is also chiral in nature and parity-violating [34], i.e. the symmetry transformations act on “left-handed” and “right-handed” fermion fields differently each. This property that the W-boson only couples to fermions of

left-handed chirality is a crucial property of the SM and is encoded in the V-A theory[35] to explain the the weak nuclear force, which is responsible for processes such as beta decay in atomic nuclei. It is responsible for many of the surprising features of the weak interactions, both the most attractive and the most puzzling ones.

Figure 2.5: Electroweak Interaction Vertices. X are fermions.



The mathematical formalism of the electroweak interactions extension of the QED/QCD formalism. To include weak interactions we must write the covariant derivative as:

$$D_\mu = \partial_\mu - ig_w T^i W_\mu^i - ig_{em} Y B_\mu, \quad (2.19)$$

where the operators T_i and $\mathbf{Y}$ are the generators of the $SU(2)_L$ and $U(1)_Y$ groups of gauge transformations, respectively. These generators satisfy

$$Q = T^3 + \frac{Y}{2}. \quad (2.20)$$

The eigenvalues of these generators, T and Y are also known as the weak isospin and the weak hypercharge quantum numbers. This follows that the electromagnetic current is a combination of the neutral currents and we can write:

$$j_\mu^{em} = J_\mu^3 + \frac{1}{2} j_\mu^Y. \quad (2.21)$$

For the interaction lagrangian, we recall from QED that the interaction is mediated by gauge fields as:

$$L_{QED}^{int} = -ie j_\mu^{em} A_\mu \quad (2.22)$$

where $j_\mu^{em} = \bar{\psi}\gamma^\mu\psi$ is the $U(1)_{em}$ current and A_μ is the gauge field photon. We copy this for the case of the EW interaction lagrangian and write it in terms of current triplets and singlets

$$\mathcal{L}_{EW}^{int} = -igj_\mu^i W^{i\mu} - \frac{ig'}{2} j_\mu^Y Y^\mu B_\mu \quad (2.23)$$

where the $SU(2)$ part can be expanded as:

$$-igj_\mu^i W^{i\mu} = -ig\bar{\chi}_L \gamma_\mu T^i \chi_L \quad (2.24)$$

$$\frac{ig'}{2} j_\mu^Y Y^\mu B_\mu = -ig'\bar{\psi}\gamma_\mu \frac{Y^\mu}{2} \psi. \quad (2.25)$$

The vector bosons associated with the particles $W^\pm$, Z and γ , can be constructed as linear combinations of W_μ^i gauge field triplet and the singlet B^μ . The massive vector boson is written as,

$$W^{\pm\mu} = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2). \quad (2.26)$$

The neutral vector bosons A_μ and Z^μ associated with the massless photon and massive Z^0 boson are written as follows:

$$\begin{aligned} A^\mu &= B^\mu \cos \theta_w + W^{3\mu} \sin \theta_w, \\ Z^\mu &= -B^\mu \sin \theta_w + W^{3\mu} \cos \theta_w. \end{aligned} \quad (2.27)$$

Here θ_w is the “weak mixing” or “Weinberg” angle. Substituting the Eqs. 2.27 to the electroweak interaction Lagrangian, grouping together terms with A^μ and Z^μ [33] and using Eq. 2.21, we find that θ_w can be written as

$$e = g \sin \theta_w = g' \cos \theta_w. \quad (2.28)$$

In practice, e , which is the electron charge and $\sin \theta_w$ are used as parameters for the standard model to be measured experimentally. Found in the second term with Z^μ is the weak neutral current:

$$J_\mu^{NC} = \frac{g}{\cos \theta_w} (j_u^3 - \sin^2 \theta_w j_\mu^{em}). \quad (2.29)$$

Here we highlight some key aspects of weak interactions. Firstly, we discuss flavor mixing, which is prominent in quarks but absent in leptons. Weak interactions uphold lepton generation conservation, preventing mixing between different lepton generations. For instance, a W boson interacts with an electron and an electron neutrino but not with a muon and an electron neutrino [36].

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \quad (2.30)$$

In contrast, quark generation is not conserved. Instead, we have the quark doublets that participate in the $SU(2)$ symmetry of the charged-current weak interaction are defined as follows:

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix} \quad (2.31)$$

The primed coordinates are given by the relation involving the CKM matrix in Eq. 2.2.3. This flavor state mixing, non-conservation in the case of quarks is captured in the CKM matrix, denoted as V :

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

In this matrix, the elements V_{ij} represent the complex numbers that describe the mixing between different quark flavor states. The indices (i, j) correspond to the initial and final quark flavor states, where “u” stands for up quark, “d” for down quark, “s” for strange quark, “c” for charm quark, “t” for top quark, and “b” for bottom quark. Since this matrix is mostly diagonal, it results in the suppression of couplings between an up-type quark from one generation and a down-type quark from another to the W boson. In the original Standard Model, neutrinos were considered massless, and thus, there was no mixing in the leptonic sector. However, as we now know that neutrinos have mass, one way to account for this is by introducing right-chiral singlets for neutrinos, allowing them to acquire mass in a manner similar to other Standard Model fermions. This introduces a mixing matrix in the leptonic sector known as the PMNS matrix [37].

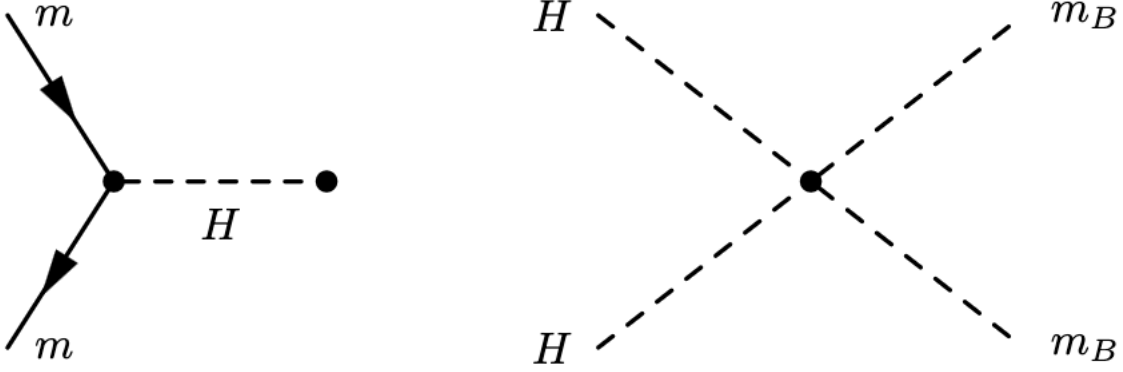
Historically, the masses of the weak force carriers had been a significant puzzle. The SM of the 1970s or also known as Glashow-Salam-Weinberg theory predicted the existence of the W and Z bosons but did not provide exact predictions of their masses. They were measured later at CERN with the UA1 experiment (W) and the UA2 collaboration (Z-boson). The origin of their masses was elusive and couldn’t be incorporated back then into the SM theory in a gauge-invariant matter. This posed as a challenge for achieving true unification. We will see in the next section the significance of the Higgs Mechanism in how particles acquire mass through spontaneous symmetry breaking and preserving the gauge invariance of the theory.

2.2.4 Higgs Mechanism and Spontaneous Symmetry Breaking

The Higgs mechanism was conceived originally to provide mass to the weak gauge bosons. However, it also offers a means of imparting mass to fermions through the Yukawa coupling. We will show both in this section. Consequently, the Higgs boson forms direct connections with all massive particles within the Standard Model. Furthermore, the Higgs boson itself possesses mass, though it is not a gauge boson. It acquires a mechanism for self-interaction through the Higgs potential. The interaction vertices involving the Higgs are illustrated in

Figure ???. The first of these (Figure 3.6a) delineates the interaction between the Higgs boson and massive particles.

Figure 2.6: Higgs Interaction Vertices.



An ad hoc introduction of vector boson mass terms in the Lagrangian leads to a violation of gauge invariance. For the QED Lagrangian in Eq. 2.8, a mass term of the form

$$\mathcal{L} = \frac{1}{2} M_A^2 A^\mu A_\mu \quad (2.32)$$

would destroy gauge invariance under the transformation of A_μ (Eq. 2.6). To generate mass for the gauge bosons, the local gauge symmetries must be broken.

Standard introductory particle physics texts and lectures [33] usually begin with a toy model where a single complex scalar field is coupled to the U(1) gauge field

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + |D_\mu \phi|^2 - V(\phi) \quad (2.33)$$

where $D_\mu = \partial_\mu - ieA_\mu$ and $V = \mu^2 |\phi|^2 - \lambda |\phi|^4$. For $\mu^2 < 0$, this is just the QED Lagrangian for a charged complex scalar particle of mass μ save for the ϕ^4 self-interaction term. It has a unique minimum at zero and does not break the symmetry. To generate the mass terms we take $\mu^2 > 0$. This breaks the $U(1)$ symmetry and has a minimum at

$$\langle \phi \rangle = \frac{v}{\sqrt{2}} = \sqrt{-\frac{\mu^2}{2\lambda}}. \quad (2.34)$$

This is a circle of minimum with radius v and is shown in Fig. 2.7.

For convenience we can rewrite ϕ in terms of real fields,

$$\phi = \frac{v+h}{\sqrt{2}} e^{i\frac{\theta}{v}}, \quad (2.35)$$

here h and χ are the Higgs and Goldstone bosons respectively

$$\begin{aligned} L = & -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - evA_\mu\partial^\mu\chi + \frac{e^2v^2}{2}A_\mu A^\mu \\ & + \frac{1}{2}(\partial_\mu h\partial^\mu h - 2\mu^2 h^2) \\ & + \frac{1}{2}\partial_\mu\chi\partial^\mu\chi + \text{interaction terms.} \end{aligned} \quad (2.36)$$

With this new Lagrangian, the photon is imparted a mass $m_A = ev$. The theory now has a Higgs boson h of mass $m_h = \sqrt{2\mu} = \sqrt{2\lambda v^2}$, and a massless Goldstone θ . If we choose the following gauge transformation:

$$A_\mu \rightarrow A_\mu + \frac{1}{ev}\partial_\mu\theta \quad (2.37)$$

the theory will be independent of θ . The theory will not just have two interacting massive particles which are A_μ and h .

The procedure could be generalized for an $SU(n)$ gauge symmetry [36].

$$\begin{aligned} \mathcal{L} &= D_\mu\phi^\dagger D^\mu\phi - V(\phi) \\ V(\phi) &= \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \end{aligned} \quad (2.38)$$

This is invariant under the following transformations:

$$\begin{aligned}\phi_i &\rightarrow (1 - \epsilon^a \tau^a)_{ij} \phi_j \\ D_\mu \phi &= (\partial_\mu - ig\tau^a A_\mu^a) \phi\end{aligned}\tag{2.39}$$

where $i, j = 1, 2, \dots, n$; $a = 1, \dots, n^2 - 1$, g is a coupling constant, and ϵ^a are small parameters and τ^a are the group generators of the theory.

Now that the concept of spontaneous symmetry breaking is introduced, we can introduce the Weinberg formulation for a Higgs Lagrangian, similar to Eq. 2.38 such that the $W^\pm$ and Z bosons are massive and the photon remains massless. Here we introduce four real scalar fields which is equivalent to a Higgs $SU(2)$ doublet:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}\tag{2.40}$$

The covariant derivative would have the form:

$$D_\mu = \partial_\mu - i\frac{g}{2}\sigma^a W_\mu^a - i\frac{g'}{2}B_\mu,\tag{2.41}$$

which indicates a coupling of the Higgs to $SU(2)$ gauge bosons, $W^a, a = 1, 2, 3$ and the $U(1)$ gauge boson, B . Once again, $\mu^2 < 0$ for spontaneous symmetry breaking to occur, with the following choice of a minimum or vacuum expectation value (VEV):

$$\langle \phi \rangle = \phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}.\tag{2.42}$$

Substituting this vacuum expectation value Eq. 2.42 to the Lagrangian in Eq. 2.38,

$$D_\mu \phi^\dagger D^\mu \phi = \left(\frac{1}{2}vg\right)^2 W_\mu^+ W^{-\mu} + \frac{1}{8}v^2 (W_\mu^3, B_\mu) \begin{pmatrix} g^2 & -gg' \\ -gg' & g'^2 \end{pmatrix} \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix}.\tag{2.43}$$

Multiplying the matrices, and identifying the mass terms, we have:

$$\begin{aligned}
M_w &= \frac{gv}{2}, & W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2) \\
M_z &= \frac{v}{2}\sqrt{g^2 + g'^2}, & Z_\mu &= \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2 + g'^2}} \\
M_A &= 0, & A_\mu &= \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2 + g'^2}}.
\end{aligned} \tag{2.44}$$

If we reexpress the results here with the weak mixing angle in eq. 2.28, we can relate the $W^\pm$ and Z masses with

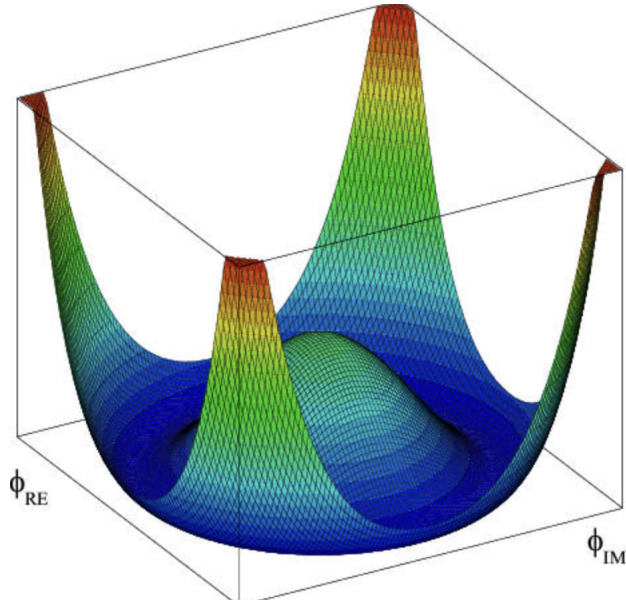
$$\frac{M_W}{M_Z} = \cos \theta_W. \tag{2.45}$$

Experimentally, the masses are measured to be,

$$\begin{aligned}
M_W &\approx 81 \text{ GeV} \\
M_Z &\approx 93 \text{ GeV},
\end{aligned} \tag{2.46}$$

which are in impressive agreement with the predictions of SM.

Figure 2.7: The infamous “Mexican Hat” Higgs potential for the case of $\mu^2 < 0$ and $\lambda > 0$. There is a circle of minima of radius v .



As mentioned earlier, the fermions gain mass via the Yukawa coupling of the single Higgs doublet ϕ with the fermions [38]. In a nutshell, one can write a Yukawa coupling term in the lagrangian for a fermion:

$$\mathcal{L}_{\text{Yukawa}} = -g_f \bar{\psi}_f \phi \psi_f + h.c. \quad (2.47)$$

where g is a dimensionless coupling constant for some fermion f , ψ_f is the fermion field and ϕ is the Higgs field. After spontaneous symmetry breaking, the Higgs field takes on the vacuum expectation value in Eq. 2.42 the Lagrangian in Eq. 2.47. For illustration, we do this for the first generation of fermions:

$$\mathcal{L}_{\text{Yukawa}} = \frac{f_e v}{\sqrt{2}} (\bar{e}_L e_R + \bar{e}_R e_L) + \frac{f_u v}{\sqrt{2}} (\bar{u}_L u_R + \bar{u}_R u_L) + \frac{f_d v}{\sqrt{2}} (\bar{d}_L d_R + \bar{d}_R d_L), \dots \quad (2.48)$$

From here one can read off the masses as:

$$m_i = \frac{f_i v}{\sqrt{2}}, \quad i = e, u, d. \quad (2.49)$$

One thing to note is that the neutrinos do not have right-handed partners in the SM and therefore could not acquire a mass term through Yukawa coupling [39].

On July 4, 2012, the CMS and ATLAS experiments independently announced the discovery of a particle consistent with the Higgs boson. This landmark achievement, while significant, did not “complete” the Standard Model in the sense of resolving all fundamental questions in particle physics. Rather, it confirmed one of the final pieces of the SM puzzle, highlighting that numerous unresolved questions and challenges remain which requires further the exploration of new theoretical frameworks to expand our understanding of the universe.

2.3 Limitations of the Standard Model

In this section, we briefly give an itemized overview of a number of open questions in the Standard Model and discuss in more detail how the BSM models we are exploring play a role in addressing these open problems.

- **Hierarchy Problem:** The Standard Model struggles to explain the vast disparity in the strengths of the fundamental forces, particularly why gravity is incredibly weaker than the other forces in our four-dimensional spacetime.
- **Parameter Fine-tuning:** The Standard Model involves several “unnatural” parameters, such as the low mass of the Higgs boson. Adjusting these parameters to match experimental observations appears contrived, lacking a deeper theoretical explanation.
- **Quantum Theory of Gravity:** While the Standard Model successfully describes the other fundamental forces through quantum field theories, it falls short in providing a consistent quantum field theory for gravity, a long-standing challenge in theoretical physics.
- **Dark Matter and Dark Energy:** The Standard Model accounts for only a tiny fraction, approximately 4%, of the total matter-energy content of the universe. It fails to address the presence of dark matter and the mysterious dark energy, both of which dominate the universe’s composition.
- **Baryon Asymmetry:** There is an unexplained imbalance between matter and anti-matter in the universe, known as the baryon asymmetry, which the Standard Model does not elucidate.
- **Strong CP Problem:** The mathematical formulation of Quantum Chromodynamics (QCD) within the Standard Model predicts a violation of CP-symmetry in strong interactions, which contradicts experimental observations. CP-symmetry dictates that

the laws of physics should remain unchanged when particles are replaced with their antiparticles (charge reversal), followed by a left-right flip (parity reversal).

- **Neutrino Masses:** Neutrinos, initially assumed to be massless in the Standard Model, have been observed to have tiny but non-zero masses through experimental evidence. The SM does not naturally incorporate these neutrino masses and mixing angles, necessitating an extension of the theory to accommodate them.

2.3.1 Hierarchy and Naturalness

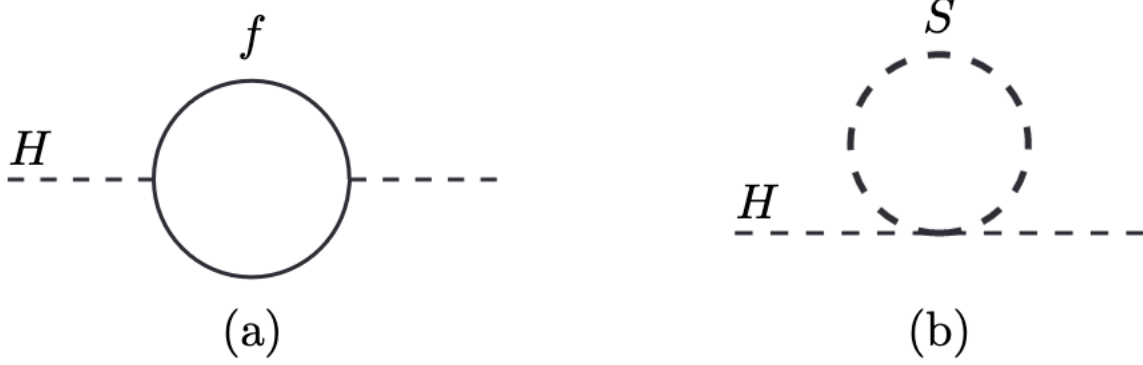
It is well-established at this point that the Standard Model is a work-in-progress specially at higher energy regimes. Extensions to the standard model are guided by the principle of the desert hypothesis which has the fundamental premise that there is no new physics between the electroweak unification scale at $m_{EW} \approx 1$ TeV and near the Planck mass or the reduced Planck Scale $M_P \approx 10^{18}$ GeV where quantum gravitational effects become significant. This large gap between the two scales refers to the well-known “hierarchy problem”.

Theorists however think that this desert, 16 orders of magnitude wide, might conceivably host a plethora of successive layers of new effective field theories (EFTs). Among these are the ADD Large Extra Dimensions 2.4.1, the Clockwork Model 2.4.2 and the Unparticles 2.4.3, which could serve diverse purposes, including the initiation of dynamical symmetry breakings or providing explanations for the intricate patterns observed in fermion masses and mixings [40]. While the hierarchy problem might not be a direct issue with the Standard Model it reflects unsettling consequences related to the “naturalness” of the Higgs mass.

The Higgs boson, being the only scalar particle in the Standard Model, possesses a distinct property: its quantum corrections are quadratically sensitive to high energy scales. This issue arises from the substantial quantum corrections originating from the virtual effects of every particle or phenomena directly or indirectly that couples to the Higgs field [39]. The general form for the Higgs boson mass calculation is written as:

$$m_H^2 = m_{bare}^2 + \Delta m_H^2, \quad (2.50)$$

Figure 2.8: One-loop quantum corrections to the square of the Higgs mass parameter (m_H^2) due to a Dirac fermion f in (a), and a scalar S [39].



where m_H is the Higgs boson mass measured to be 125 GeV, m_{bare} is the Higgs boson bare mass, and Δm_H^2 are the said quantum corrections. To illustrate the problem, consider a correction from a loop described by the Feynman diagram in Fig. 2.8 containing a Dirac fermion of mass m_f coming from a term in the lagrangian $-\lambda_f H \bar{f} f$,

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2} \Lambda_{\text{cutoff}}^2 + \dots \quad (2.51)$$

Here, $\Lambda_{\text{cutoff}} = \Lambda_{UV}$ refers to the ultraviolet momentum cutoff that is used to regulate the loop integral. It is also the energy scale at which new physics emerges. We can surmise that the SM is an EFT or a low-energy QFT approximation of some higher-energy theory which may include gravity and it is valid up to some cutoff scale. When bridging the hierarchy gap, $\Lambda_{\text{cutoff}} \rightarrow \Lambda_{Planck}$, this contribution becomes too large, notwithstanding that these contributions add up for each fermion in the SM. At the Λ_{Planck} scale, the bare mass and the quantum corrections would need to cancel over 30 orders of magnitude to match the measured value of the Higgs at 125 GeV. It seems that these quantum corrections have to be finely tuned and canceled out with an incredible degree of precision. This fine-tuning seems

unnatural because it suggests that the observed Higgs mass is a delicate balance between the tree-level mass and the quantum corrections. In other words, it appears unlikely that nature would naturally set the Higgs mass to this specific value, given the sensitivity of the corrections to it. This is known as the Higgs naturalness problem.

2.4 Beyond Standard Model Non-resonant Signals

The hierarchy and naturalness problem necessitated extensions to the Standard Model of which the primary and uncontested early candidate was Supersymmetry (SUSY). SUSY posits the existence of yet-to-be-discovered fermionic (bosonic) superpartners to the known bosons and fermions. With this framework, the strong, electromagnetic and weak coupling constants could be extrapolated and converge at a point just below the Planck Scale [39].

In 1998, authors Nima Arkani-Hamed, Savas Dimopoulos, and Georgi Dvali (ADD), proposed a new framework[40, 41] that has a diametrically opposing premise from SUSY. They asserted that there is no desert between the electroweak scale and the Planck scale. Instead, they hypothesize the existence of additional n_{ED} spatial dimensions with a radius R which modifies the fundamental Planck Scale, which transforms the fine-tuning dilemma into a question of dynamics and geometry [7]. With this approach, they also showed that the weakness of gravity could be explained by the existence of these extra dimensions in which only gravity propagates. Additionally, it has been shown [42, 43] that these extra dimensions naturally lead to the unification of the coupling strengths at scales substantially below the usual GUT/ Planck scale.

The idea of extra spatial dimensions is not entirely new. It has a historical precedent dating back to the 1920s, notably explored by Kaluza and Klein as an attempt to unify the fundamental forces [44–46]. In the 1980s, string theory revisited this concept as a requirement for a consistent theory of quantum gravity, with the dimensions compactified near the Planck scale and therefore not testable experimentally [47]. While the initial endeavour faltered, the Kaluza-Klein (KK) formalism remains valuable today, albeit with modified forms and

assumptions.

A year after the ADD model was proposed, Randall and Sundrum introduced a groundbreaking approach wherein the extra dimension has a warped geometry in a five-dimensional negatively curved, Anti-de Sitter (AdS) spacetime, with a compactification scale near 1/TeV. This model explained the electroweak scale's smallness relative to the Planck scale by invoking a gravitational redshift factor that is intrinsic to the warped AdS metric. Initially conceived for gravity alone, it was soon evident that the Standard Model's gauge fields and fermions could also propagate within these five-dimensional spacetime regions [4, 5, 47].

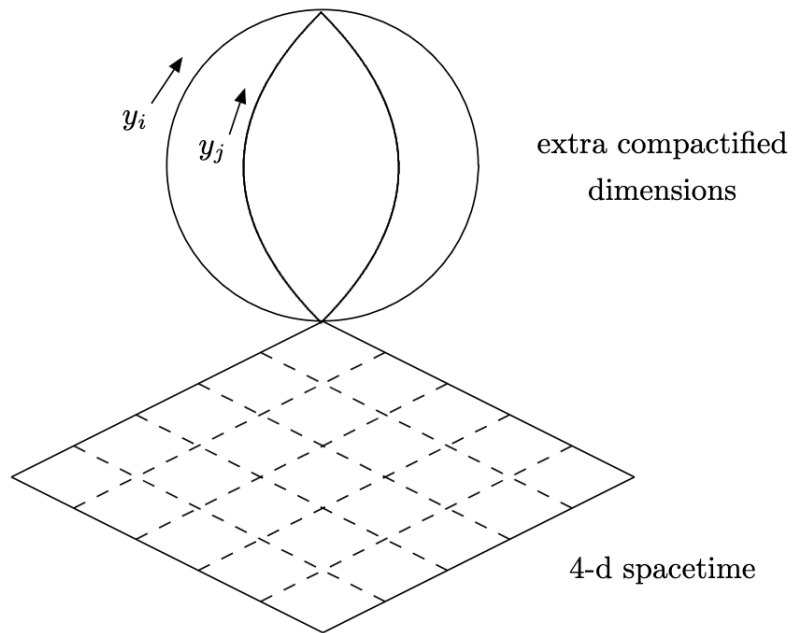
Around a decade later, additional frameworks that try to solve the hierarchy problem through various geometries and or field dynamics were proposed. One model is the Continuum Clockwork Model [48] which has a remarkable resemblance to the Linear Dilaton theory of Little String Theory [7]. At some limit, we show that it is equivalent to the ADD LED and the RS model. Another is the Unparticles Scenario which explores the possible existence of unconventional scale-invariant fields and interactions in particles. While the Unparticles [49–51] framework is distinctly different from the other aforementioned models, they mimic the phenomenology of extra dimensions which look like an almost continuum of broad or closely-spaced resonances [52]. In the succeeding parts of this chapter, we explain explain some salient mathematical and phenomenological features of the ADD Large Extra Dimensions 2.4.1, the Clockwork/Linear Dilaton 2.4.2 and the Unparticles 2.4.3 - the non-resonant models we explore in this analysis.

2.4.1 ADD Large Extra Dimensions

In the ADD scenario, the hierarchy problem is resolved via the existence of n extra dimensions with a finite radius R . This is illustrated in Fig. 2.9. The apparent weakness of gravity is also explained because it propagates in the bulk or the extradimensional volume. In effect, the enormity of the Planck scale becomes simply a consequence of the large size of the extra dimensions.

With this as a guiding principle, a new fundamental scale, much closer than the Planck

Figure 2.9: LED Scenario Schematic where compactified extra dimension goes beyond our 4-dimensional spacetime where SM particles live. [53]



scale and therefore experimentally accessible, is established. The Planck scale in terms of this new scale M_D can be written as:

$$M_{Pl}^2 = M_D^{2+n} R^n. \quad (2.52)$$

To illustrate this, we follow the ideas given in the first paper on the ADD LED scenario [40]. Here, all SM particles and gauge interactions are confined our 4-dimensional brane which is embedded in a larger $4+N$ -dimensional bulk space, with N extra dimensions that are compactified in a sphere or torus of radius R , with volume $V_n = (2\pi R)^n$. In this $4+N$ world, the gravitational potential between two masses m_1 and m_2 , a distance $r \ll R$ from one another will feel a gravitational potential dictated by Gauss's law:

$$V(r) \approx \frac{m_1 m_2}{M_{Pl(4+N)}^{N+2}} \frac{1}{r^{n+1}}, (r \ll R). \quad (2.53)$$

If these two masses are separated with a distance $r \gg R$ from one another, their gravitational flux is too weak to penetrate the the extra dimensions so the usual $1/r$ potential is obtained,

$$V(r) \approx \frac{m_1 m_2}{M_{Pl(4+N)}^{N+2} R^N} \frac{1}{r}, (r \gg R). \quad (2.54)$$

Comparing the two equations, we get the effective 4-D M_{Pl} ,

$$M_{Pl}^2 \approx M_{Pl(4+N)}^{2+N} R^N. \quad (2.55)$$

This is equivalent to Eq. 2.52.

For the hierarchy problem to be solved, we set $M_D \approx M_{EW} \approx 1$ TeV, and demand that R the observed M_{Pl} yields,

$$R \sim 10^{\frac{30}{n}-17} \text{cm} \times \left(\frac{1 \text{ TeV}}{m_{EW}} \right)^{1+\frac{2}{n}}. \quad (2.56)$$

When $n = 1$, the radius R is approximately 10^{13} centimeters which leads to observable deviations to Newtonian gravity within the solar system. This is clearly ruled out. For all cases $n \geq 2$, R goes at sub-mm distances. These are cases currently being explored in collider experiments and astrophysical and cosmological constraints are being set on the reduced Planck scale M_D [47].

It's important to acknowledge that while the existence of large extra dimensions can address the hierarchy problem, it introduces another challenge in the form of a finely-tuned parameter: the size of the extra dimensions, R . As discussed earlier, R must be carefully chosen to match the 4D Planck scale and ensure that M_D is on the Electroweak (EW) scale. This requirement replaces the need for fine-tuning of the bare Higgs boson mass. Regardless of our confidence in the ADD model specifically, it inspires a range of experiments to detect potential deviations from the Standard Model (SM).

ADD Phenomenology

The Einstein-Hilbert Action for the ADD model can be pieces [53]:

$$S = S_{bulk} + S_{brane} \quad (2.57)$$

where we are assuming that there is only one brane where the SM fields exist. The bulk action is just the Einstein-Hilbert Action for $4 + n$ dimensional gravity:

$$S_{bulk} = - \int d^{(4+n)}x \sqrt{g^{4+n}} M_D^{n+2} R^{(4+n)}, \quad (2.58)$$

where $R^{(4+n)}$ is the higher-dimensional Ricci curvature scalar.

In terms of the lagrangian density for the lagrangian density of the bulk fields $\phi(x, \vec{y})$, we have:

$$S_{bulk} = \int d^4x d^n y \sqrt{|g^{4+n}|} \mathcal{L}(\phi(x, \vec{y})), \quad (2.59)$$

,

Here x stands for the (3+1) coordinates of the brane and y is for the n extra dimensions [54]. On the other hand the action for the Brane Fields, $\xi(x)$ is written as

$$S_{Brane} = \int d^4x d^n y \sqrt{|g^{4+n}|} \mathcal{L}(\xi(x)) \delta^n(\vec{y} - y_0), \quad (2.60)$$

where the use of the delta density promotes the fields to a higher dimensional expression. The brane is taken to be at a position $\vec{y} = \vec{y}_0$.

The line element in the bulk is

$$ds^2 = g_{\mu\nu}^{4+n} dx^\mu dx^\nu. \quad (2.61)$$

The ADD model assumes spacetime is flat, in contrast to the RS model where it is warped. Hence, the metric tensor $g_{\mu\nu}$, the quantity that encodes the curvature of spacetime, can be expanded about flat spacetime with 4D metric fluctuations $h_{\mu\nu}$,

$$ds^2 = (\eta_{\mu\nu} + h_{\mu\nu}) dx^\mu dx^\nu - r^2 d\Omega^2(n), \quad (2.62)$$

where $d\Omega_{(n)}$ are n -dimensional toroidal coordinates. Following the steps in ref. [53], we can integrate out the higher dimensions and arrive at the result

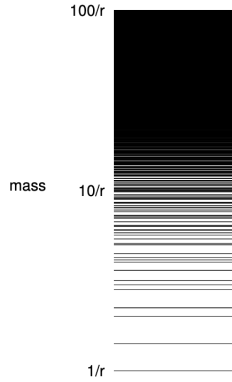
$$M^2 = M_D^{n+2} (2\pi r)^n, \quad (2.63)$$

which is the same result in Eq. 2.55 we quote in the previous section. The arbitrary metric fluctuations about the flat space time, $h_{\mu\nu}$ could be interpreted as the higher dimensional graviton which are laid out in more detail Refs. [53–55]. Defining higher dimensional coordinates y_i , and imposing a periodic boundary conditions, $y_i \rightarrow y_i + 2\pi r$ due to the compactifications of the extra dimensions, we have the expansion of the higher dimensional graviton in terms of the 4D Kaluza-Klein (KK) fields:

$$h_{\mu\nu} = \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_n=-\infty}^{\infty} \frac{h_{\mu\nu}^{(m)}(x)}{\sqrt{V_n}} e^{i \frac{m_j y_j}{R}}, \quad (2.64)$$

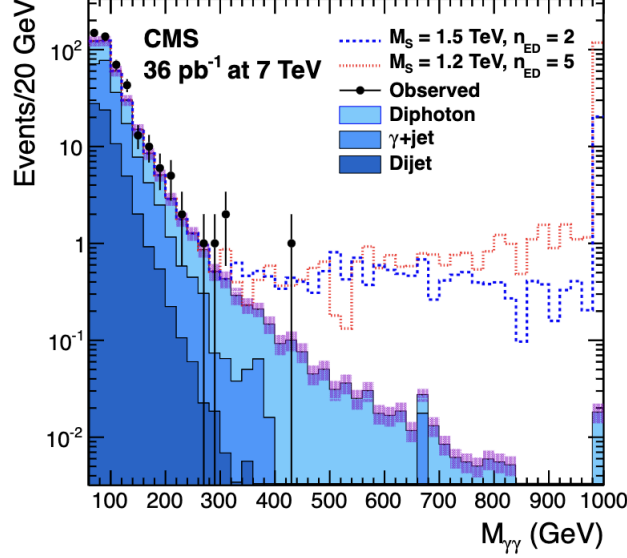
where h^m is shorthand for $h^{(m_1, m_2, \dots, m_n)}$. These KK graviton excitations or "tower" results from the coupling of the graviton to SM fields. Each of these KK modes couple to the stress-energy tensor of the SM field by the strength of gravity. The KK mass is given by $m_n^2 = m^2 + \frac{k^2}{r^2}$. This mass spectrum, illustrated in Fig. 2.10, can be much less than the detector resolution. These discrete KK states are closely-spaced together and get closer together and appear as a continuum. In this CMS analysis, we look at the invariant diphoton mass spectrum illustrated by a result from a previous analysis in Fig. 2.11 [56]. The closely spaced KK modes are therefore predicted to appear as a continuum excess in the diphoton mass spectrum.

Figure 2.10: Mass spectrum for KK gravitons for $n = 2$.



Since there are an infinite number of KK modes, an ultraviolet (UV) cutoff M_s is introduced to get a non-diverging cross-section. The cross-sections for the virtual graviton exchange (see Fig. 2.12) for various parameters are shown in in Appendix C.1. Any contributions above the cutoff scale M_s is set to zero. The total cross-section is a combination of the contributions from the SM, ADD as well as the interference between the SM and ADD:

Figure 2.11: A CMS result [56] using data from $\sqrt{s} = 7$ TeV pp-collisions which shows a comparison of the invariant mass spectrum predicted for the Standard Model diphoton production in light blue, overlaid with a continuous spectra for two Large Extra Dimensions signal points shown in dashed lines. The observed data and additional background predictions from γ +jet and dijet events are also shown in various markers indicated in the legend.



$$\sigma_{total} = \sigma_{SM} + \frac{\mathcal{F}}{M_S^4} \sigma_{int} + \frac{\mathcal{F}}{M_S^8} \sigma_{ADD} \quad (2.65)$$

Here, $\mathcal{F}$ is parameter which allows us to translate our results in the different conventions for the ADD model that are found in literature. Here we consider the GRW [55], HLZ [57] and the Hewett [58] conventions where we have,

$$\mathcal{F} = \begin{cases} 1 & \text{if GRW} \\ \log\left(\frac{M_s^2}{\hat{s}}\right) & \text{if } n_{ED} = 2 \\ \frac{2}{n_{ED}-2} & \text{if } n_{ED} > 2 \\ \pm \frac{2}{\pi} & \text{if Hewett.} \end{cases} \quad (2.66)$$

Here $\hat{s}$ is the centre-of-mass energy of the colliding partons. It is important to note that the GRW and the HLZ conventions interfere constructively while the Hewett could either

interfere constructively (+) or destructively (-), according to the signs in Eq. 2.66. Later in Ch. 6, M_s , is replaced by Λ_T whose equivalence varies by convention.

Figure 2.12: Feynman Diagrams for virtual graviton or Unparticle exchange in the diphoton channel $\gamma\gamma$ through quark annihilation (left) and gluon fusion (right).



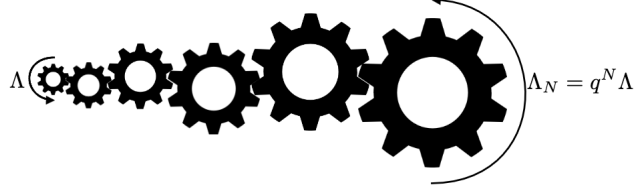
2.4.2 Clockwork/ Linear Dilaton

A more recent model that has been proposed is the Continuum Clockwork model which is equivalent to the Linear Dilaton theory in 5D. The CW/LD model is distinct but related to the more known extra-dimensional models which are the ADD LED and the RS Graviton scenarios. In the clockwork picture, the hierarchy is solved by introducing a continuum of clockwork fields which interact with each other. Inspired by gears transmitting energy, they are arranged in a way that the fields/particles associated have increasing masses/energy scales as you move along the chain as seen in Fig. 2.13. This general mechanism can take large effective interaction scales from dynamics occurring at much lower energies. These “gears” could also be interpreted as N copies of some particle content laid out on a one-dimensional lattice in theory space. In the continuum limit this lattice interpreted as a physical extra dimension and the gears are the KK modes [48].

The basic mathematical setup of the CW/LD follows from Ref. [59]. A 5D space is considered in which the extra dimension is a circle parametrized by y where $-\pi R \leq y \leq \pi R$. The SM or TeV brane is set at $y = 0$, whereas the Planck brane is set at $y = y_p = \pi R$. With a Z_2 orbifold symmetry, $y \leftrightarrow -y$ are identical. The full action in the Einstein frame is a sum of the SM action, the Bulk action and additional terms account for the extrinsic curvature of the two boundaries and the scalar potentials at the TeV and Planck branes near their local

minima.

Figure 2.13: The “Clockwork gears” schematic representation of the clockwork mechanism where the gears represent the interacting fields. [48].



In this scenario, Einstein’s equations and the dilaton equation of motion are solved by the metric,

$$ds^2 = e^{\frac{4}{3}ky} g_{\mu\nu} dx^\mu dx^\nu + dy^2. \quad (2.67)$$

Taking into account the metric in Eq.3.34 of Ref. [48] which describes the graviton fluctuations around the 4D Minkowski space, we can write the decomposed graviton states as:

$$h_{\mu\nu} = \sum_{n=0}^{\infty} \frac{\tilde{h}_{\mu\nu}(x) \Psi_n(y)}{\sqrt{\pi R}}. \quad (2.68)$$

If we suppose that the SM fields localized at $y = 0$, we can write the interaction lagrangian as

$$\mathcal{L} = -\frac{h_{\mu\nu}(x, 0) T_{\mu\nu}^{SM}(x)}{M_5^{3/2}} = -\sum_{n=0}^{\infty} \frac{\tilde{h}_{\mu\nu}(x) T_{\mu\nu}^{SM}(x)}{\Lambda_n}, \quad (2.69)$$

where $T_{\mu\nu}^{SM}(x)$, is the SM energy-momentum tensor and Λ_n is the effective scale of interaction. In more detail these are written as,

$$T_{\mu\nu}^{SM}(x) = -2\frac{\partial \mathcal{L}^{SM}}{\partial g_{\mu\nu}} + g_{\mu\nu} \mathcal{L}^{SM}; \quad \Lambda_n \equiv \frac{\sqrt{\pi R M_5^{3/2}}}{\psi_n(0)}. \quad (2.70)$$

With this metric and the action, (see Appendix A of Ref. [48]) one can derive the analogue to Eq. 2.52,

$$M_P^2 \approx \frac{M_5^3}{k} (e^{2\pi k R - 1}), \quad (2.71)$$

where M_5 is the fundamental scale or the 5D reduced Planck Mass M_D . This is the scale at which the theory must be UV-completed and it is assumed to be close to M_{EW} . Here, M_P is an illusion caused by exponential factor, similar to the RS scenario. Using the Eq. 2.70, Eq. 2.71, as well as Eq.3.16 and 3.17 of Ref. [48], the effective scales of gravitational interaction are:

$$\Lambda_0 = M_P; \quad \Lambda_n = \sqrt{M_5^3 \pi R \left(1 + \frac{k^2 R^2}{n^2}\right)}. \quad (2.72)$$

To account for hierarchy, one needs to choose an appropriate kR and expand the natural logarithm

$$kR \approx \frac{1}{\pi} \ln \left(\frac{M_P}{M_5} \sqrt{\frac{k}{M_5}} \right) \approx 10 + \frac{1}{2\pi} \ln \left(\frac{k}{\text{TeV}} \right) - \frac{3}{2\pi} \ln \left(\frac{M_5}{10\text{TeV}} \right). \quad (2.73)$$

Here, k is a mass parameter called the “clockwork spring” which measures the effectiveness of the clockwork mechanism. A flat metric corresponds to $k = 0$. As described in the original paper, the extra dimension is discretized by choosing $y_j = ja$, where $j = 0, \dots, N$. The lattice spacing a is such that $Na = \pi R$. While this construction resembles fields in a deconstructed flat extra dimension, there is a critical difference which comes from how each link breaks the symmetry of a single gear site. A parameter q treats the site $j + 1$ and the site j (with $j = 0, \dots, N - 1$) asymmetrically. It’s important to note that we can establish a total of N connections among the $N + 1$ available sites. As a consequence, one site retains its masslessness. For q to remain y -independent and for q^N to give a finite but non-trivial clockwork behavior, as the number of sites approaches infinity, we set

$$q^N = e^{k\pi R}. \quad (2.74)$$

The excitations have a similar form with the large extra dimensions, where at the large

N limit, we have

$$m_0^2 = 0; \quad m_n^2 = k^2 + \frac{n^2}{R^2} + \frac{O(1)}{N}, \quad n = 1, \dots, N \quad (2.75)$$

which appears as a continuous spectrum like the ADD LED but with evenly distributed energy levels and mass-squared splitting equal to the inverse radius-squared. However, we will see later in Ch. 6 that the spectrum shifts by an amount equal to the clockwork spring constant k , introducing a mass gap [48].

2.4.3 Unparticles Model

The two previous models propose scenarios which address the hierarchy problem via the existence of extra spatial dimensions. The Unparticles [49] model on the other hand, addresses the hierarchy problem by introducing a new energy scale related to a yet-to-be-seen scale-invariant Banks-Zaks ($\mathcal{BZ}$) sector where “unparticle stuff” live. If this sector exists, there are measureable effects coming from the interactions between the Standard model particles and the these hidden sector Unparticles. This has particular consequences with the Higgs naturalness problem - a dilemma where the Higgs mass receives quantum corrections that are quadratically sensitive to the energy scale of new Physics. Instead of the scale being at M_P , this new scale could be much closer via the existence of the $\mathcal{BZ}$ and the unparticle-SM interactions could help redistribute the quantum corrections and potentially reduce the fine-tuning problem [60].

While not directly related to this dissertation, the Unparticles model can also potentially address issues related to flavor physics and CP violation via new interactions through the hidden sector and modifications in the flavor-changing and CP-violating processes that could still be consistent with experimental observations [49, 61–63].

In the first paper by Georgi [49], the scale-invariant $\mathcal{BZ}$ sector must interact weakly with the Standard Model. Such a sector could not have particles of definite non-zero mass as mass terms explicitly breaks scale invariance. It only holds “Unparticles” which is a

hazy concept that is different from our familiar notion of particulate matter. However, using a low-energy “Effective field theory” (EFT), some phenomenological exploration is possible.

In general, direct searches are only possible if the form of the BSM theory is established, but an EFT [64] approach allows us to quantify deviations from SM predictions of known particles without introducing additional assumptions about the BSM theory. The only assumption made is that the BSM theory will faithfully reproduce the SM in the regime where it is already empirically validated. As mentioned earlier, we can surmise that the SM is an effective field theory of some higher-energy theory which is valid to some cut-off energy scale Λ .

For the Unparticles scheme, the new higher energy theory contains the SM fields and the $\mathcal{BZ}$ fields, which has a nontrivial IR fixed point. The two sets of fields interact via the exchange of particles with a large Mass Scale M_U . Below this scale the interaction can be described by non-renormalizable couplings suppressed by powers of M_U and we can write the interacting lagrangian in the following form

$$\mathcal{L}_{int} = \frac{1}{M_U^k} \mathcal{O}_{SM} \mathcal{O}_{\mathcal{BZ}} \quad (2.76)$$

where $\mathcal{O}_{SM}$ is an operator with mass dimension d_{SM} that is built out of SM fields and $\mathcal{O}_{\mathcal{BZ}}$ is an operator $d_{\mathcal{BZ}}$ built out of $\mathcal{BZ}$ fields. The exponent $k = d_{SM} + d_{\mathcal{BZ}} - 4$.

At energy Λ_U , the $\mathcal{BZ}$ sector emerges and it is expected to be conformal. However, this coupling transmutes as the energy increases to Λ_U as the scale invariance or conformal invariance is expected. In the effective theory below Λ_U , the $\mathcal{O}_{\mathcal{BZ}}$ operators match unto Unparticle operators and the interactions match onto the following form

$$\mathcal{L}_{int} = \frac{C_U \Lambda_U^{d_{\mathcal{BZ}} - d_{SM}}}{M_U^k} \mathcal{O}_{SM} \mathcal{O}_U = \frac{\lambda}{\Lambda_U^{d_U + d_{SM} - 4}} \mathcal{O}_U \mathcal{O}_{SM} \quad (2.77)$$

Here $\mathcal{O}_U$ is the Unparticle operator with dimension d_U , λ is the coupling constant between

SM and $\mathcal{BZ}$. Georgi observed that the Unparticle mass spectrum looked like a $d_{\mathcal{U}}$ number of invisible massless particles. Following the arguments from Ref. [49, 65, 66], scale invariance can be used to fix the two-point functions of these unparticle operators to show this effect. Consider a the vacuum matrix element, $\mathcal{O}_{\mathcal{U}}$,

$$\begin{aligned}\langle 0|\mathcal{O}_{\mathcal{U}}(x)\mathcal{O}_{\mathcal{U}}^{\dagger}(0)|0\rangle &= \langle 0|e^{i\hat{P}\cdot x}\mathcal{O}_{\mathcal{U}}(0)e^{-i\hat{P}\cdot x}\mathcal{O}_{\mathcal{U}}^{\dagger}(0)|0\rangle \\ &= \int d\lambda \int d\lambda' \langle 0|\mathcal{O}_{\mathcal{U}}(0)|\lambda'\rangle \times \langle \lambda'|e^{i\hat{P}\cdot x}|\lambda\rangle \langle \lambda|\mathcal{O}_{\mathcal{U}}^{\dagger}(0)|0\rangle \\ &= \int \frac{d^4P}{(2\pi)^4} e^{i\hat{P}\cdot x} \rho_{\mathcal{U}}(P^2),\end{aligned}\tag{2.78}$$

where $\rho_{\mathcal{U}}(P^2)$ is a spectral density given by:

$$\rho_{\mathcal{U}}(P^2) = A_{d_{\mathcal{U}}} \theta(P^0) \theta(P^2) (P^2)^{d_{\mathcal{U}}-2}.\tag{2.79}$$

Here P is the invariant Unparticle mass P , θ is the Heaviside step function, and $A_{d_{\mathcal{U}}}$ is the normalization factor. The heaviside step functions $\theta(P^0)$, $\theta(P^2)$ ensure that the only positive energy solutions and that the unparticle stuff is not tachyonic, respectively [67]. As shown in ref. [65], under scale transformations $x \rightarrow sx$ and $\mathcal{O}_{\mathcal{U}}(sx) \rightarrow s^{-d_{\mathcal{U}}}\mathcal{O}_{\mathcal{U}}(x)$, $A_{d_{\mathcal{U}}}$ becomes

$$A_{d_{\mathcal{U}}} = \frac{16\pi^2\sqrt{\pi}}{(2\pi)^{2d_{\mathcal{U}}}} \frac{\Gamma(d_{\mathcal{U}} + \frac{1}{2})}{\Gamma(d_{\mathcal{U}} - 1)\Gamma(2d_{\mathcal{U}})}.\tag{2.80}$$

This phase space is the same as that for $d_{\mathcal{U}}$ massless particles, where $d_{\mathcal{U}}$, can be a fractional number.

Unparticle Physics to Diphotons

There are several papers [68–70] in literature that explore the phenomenology of the Unparticle model decaying into two photons. In this dissertation, we consider scalar (Spin-0) and

tensor (Spin-2) Unparticles. The virtual exchange of Unparticles, shown in Fig. 2.12, have the following associated matrix elements. For Spin-0, we have,

$$|\bar{M}_{q\bar{q}}|^2 = \frac{1}{8N_c} \lambda^4 \chi_U^2 \left(\frac{s}{\Lambda_U^2} \right)^{2d_U-1},$$

$$|\bar{M}_{gg}|^2 = \frac{1}{8(N_c^2 - 1)} \frac{1}{4} \lambda^4 \chi_U^2 \left(\frac{s}{\Lambda_U^2} \right)^{2d_U}.$$

And for Spin-2 Unparticles,

$$|M_{q\bar{q}}|^2 = \frac{1}{8N_c} \left[e^4 Q_f^4 8 \left(\frac{u}{t} + \frac{t}{u} \right) \right]$$

$$- 8e^2 Q_f^2 \lambda_t^2 \chi_U \cos(d_U \pi) \left(\frac{s}{\Lambda_U^2} \right)^{d_U} \frac{1}{s^2} (u^2 + t^2)$$

$$+ 2\lambda_t^4 \chi_U^2 \left(\frac{s}{\Lambda_U^2} \right)^{2d_U} \frac{1}{s^4} (u^4 + t^4),$$

$$|\bar{M}_{gg}|^2 = \frac{1}{8(N_c^2 - 1)} 2\lambda_t^4 \left(\frac{s}{\Lambda_U^2} \right)^{2d_U} \chi_U^2 \frac{1}{s^4} (u^4 + t^4)$$

where Q_f is the electric charge of the parton with flavor f , $\chi_U = \frac{A_{d_U}}{2 \sin(d_U \pi)}$ and N_c is the number of colors. The variables s, t, u are the Mandelstam variables. It can be seen in the above equations that only Spin-2 tensor Unparticles interfere with the SM subprocess amplitudes. As the authors of ref. [68] argued, that only $1 < d_u < 2$ is the viable range that could be explored in the LHC. It is also important to note that due to the factor $\Lambda_U^{-d_U}$ in Eqs. 2.4.3 and 2.4.3, the cross sections are suppressed with increasing Λ_U . The cross section also scales with the Wilson coefficient, $\frac{\lambda}{\Lambda_u}$.

Unparticles and Extra Dimensions

The Unparticles model and Extra Dimensions have been shown to have similar collider signatures. The calculation of their cross-sections are similar as they share analogous phase-

space integrations [65, 70]. In a simple case where a scalar field is identified as a massless scalar bulk field permeating into the LED. With this setup, the dimension of the Unparticle field d_U is related to the number of extra dimension [65] as follows

$$d_U = \frac{n}{2} + 1 \quad (2.81)$$

In this context, incorporating the Unparticles model interpretation into the Search for Non-resonant Signal Models appears to hold promise as a scientifically valuable endeavour.

2.4.4 Previous Searches

The search for signatures of Large extra dimensions, CW/LD and Unparticles go beyond collider searches. However, we limit the discussion of previous results to collider searches here. The methods and results in this study has a direct precedent from the work published in Ref. [71]. Limits have been placed on the relevant model parameters for the ADD Large Extra Dimensions as shown on Table 2.1, and the Clockwork model using high-mass diphoton events from the 2016 data from the CMS experiment (see Fig. 2.14) which was at the time the most stringent limits on the ADD Large Extra Dimensions and the Clockwork Model. The total asymmetric uncertainties are shown on the expected limits for Table 2.1. In Figure 2.14, the shaded green and yellow bands represent the 1 and 2 standard deviation uncertainty in the expected limit. The theory becomes nonperturbative in the $k > M_5$ region denoted by the light purple shaded region. The region below and to the left of the solid line constitutes the excluded region. A complementary search has also been done on dilepton events [72]. Both searches showed no significant deviation from the SM expectation of unity.

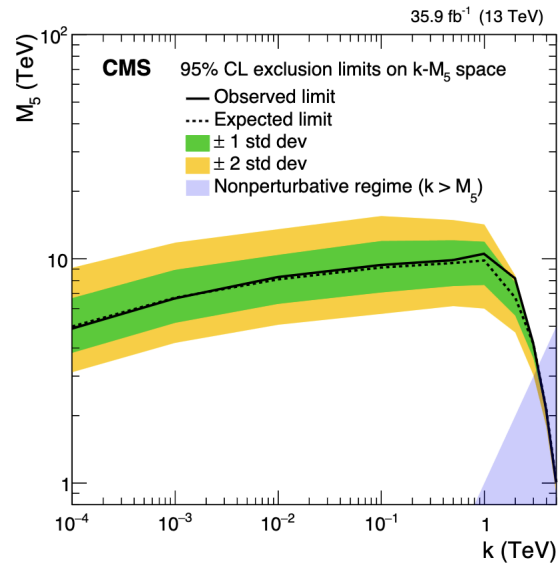
More recently, the ATLAS experiment [73] recently placed exclusion limits at 95% on the two-dimensional parameter space of the clockwork gravity model in their search for periodic signals in the dielectron and diphoton mass spectrum. We compare these recent results with our latest findings in Ch. 8.

Both the Unparticles and Large Extra dimensions have also been used as additional model

Table 2.1: Previous exclusion limits results on the mass scale M_5 in TeV for various conventions used in the calculation of the ADD large extra-dimensional scenario using the 2016 CMS detector data corresponding to an integrated luminosity of 35.9 fb^{-1} [71] only.

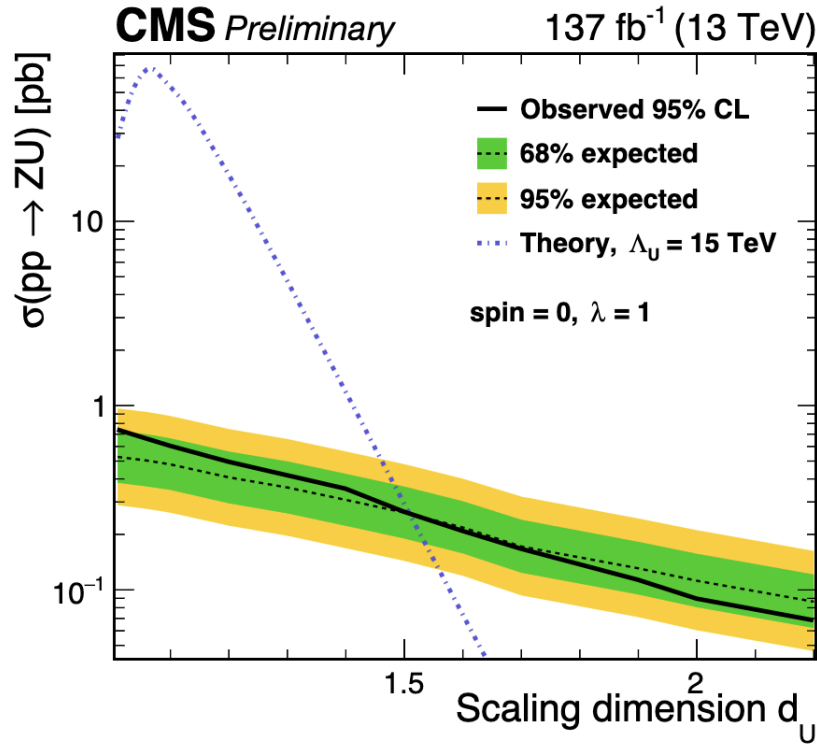
Signal	GRW	Hewett		HLZ					
		negative	positive	$n_{\text{ED}} = 2$	$n_{\text{ED}} = 3$	$n_{\text{ED}} = 4$	$n_{\text{ED}} = 5$	$n_{\text{ED}} = 6$	$n_{\text{ED}} = 7$
Expected	$7.1^{+0.7}_{-0.5}$	$5.5^{+0.1}_{-0.3}$	$6.3^{+0.6}_{-0.4}$	$8.4^{+1.3}_{-1.1}$	$8.4^{+0.8}_{-0.6}$	$7.1^{+0.7}_{-0.5}$	$6.4^{+0.6}_{-0.5}$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

Figure 2.14: 2016 95% CL exclusion limits results for the continuous clockwork mechanism over the $k - M_5$ parameter space.



interpretations of Searches for Dark Matter that are produced in association with Z boson plus missing E_T decays using the full Run II CMS data [74]. Figure 2.15 shows limits applied to fixed values of the effective cutoff scale $\Lambda_U = 15$ TeV and the coupling constant $\lambda = 1$. While this analysis involves a direct production rather than a virtual exchange of Unparticles, which is what is investigated here. The diphoton channel provides a complementary search where there is potentially more sensitivity for some values of d_U and Λ_U .

Figure 2.15: A related report on the 95% CL upper limits on unparticle+Z [74] production cross section, as a function of the scaling dimension d_U .



CHAPTER 3

EXPERIMENTAL SETUP: THE LARGE HADRON COLLIDER AND CMS

“It doesn’t matter how beautiful your theory is, it doesn’t matter how smart you are. If it doesn’t agree with experiment, it’s wrong.”

-Richard Feynman

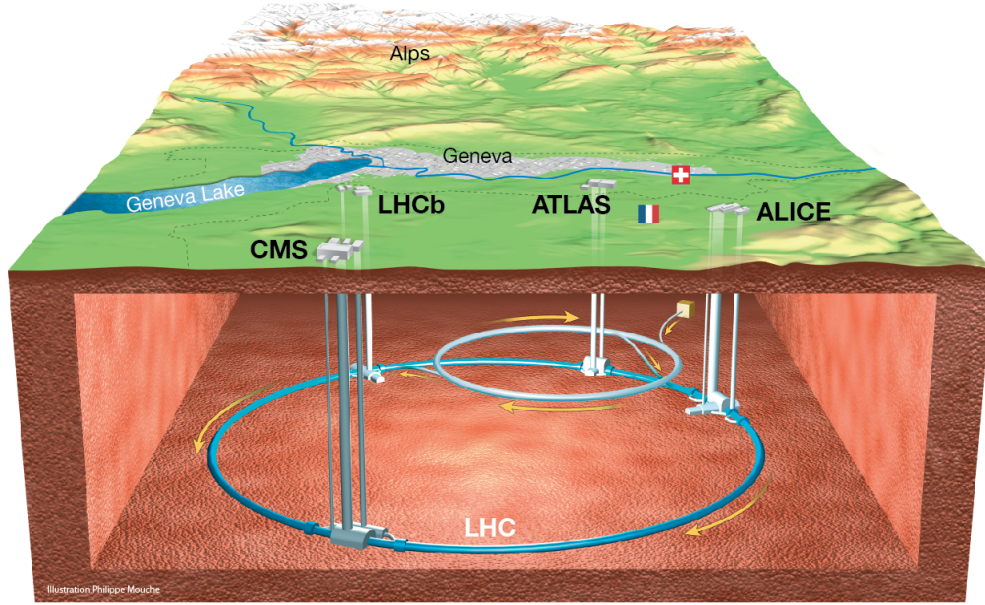
In this chapter, we discuss the experimental setup. First, in Sec 3.1 we discuss the accelerator complex that accelerates protons into counter-rotating beams which are then collided to produce the two-photon events which are collected by the CMS detector 3.3.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) depicted in Figure 3.1, is the world’s largest and most powerful man-made particle accelerator as of date. It is nestled approximately 100 meters underground, beneath the Jura Mountains and Lac Leman, straddling Geneva, Switzerland and the neighbouring territory of France.

The LHC was built as a general-purpose collider and the data it produces serve various physics programs from precision measurements in the Standard Model to searches for exotic, beyond standard model physics. In this section, we will briefly describe the LHC accelerator complex, the engineering motivations for the collider design and the key physics performance in various runs.

Figure 3.1: A bird's eyeview and schematic depiction of the underground LHC accelerator complex and its four interaction points in Geneva, Switzerland and France [75].



LHC Ring of Magnets

The full 27-km LHC ring is made up of segments of cryogenically-cooled superconducting dipole and quadrupole magnets that house two counter-rotating proton beam pipes. The dipole magnets steer the particle beams into an approximately circular orbit. Superconducting quadrupole magnets focus the particle beams while higher order moment magnets are used for beam corrections, all in the effort to maximize collision rates [76]. A total of 1248 dipole and 400 quadrupole magnets constitute the full ring, each of them 15-m and 3-m in length respectively. They are constructed with niobium-titanium (NbTi) which produce an operational field strength of 7.7 T. To maintain their superconductive state, they are cooled to 1.9 K with liquid helium.

Experiments Overview

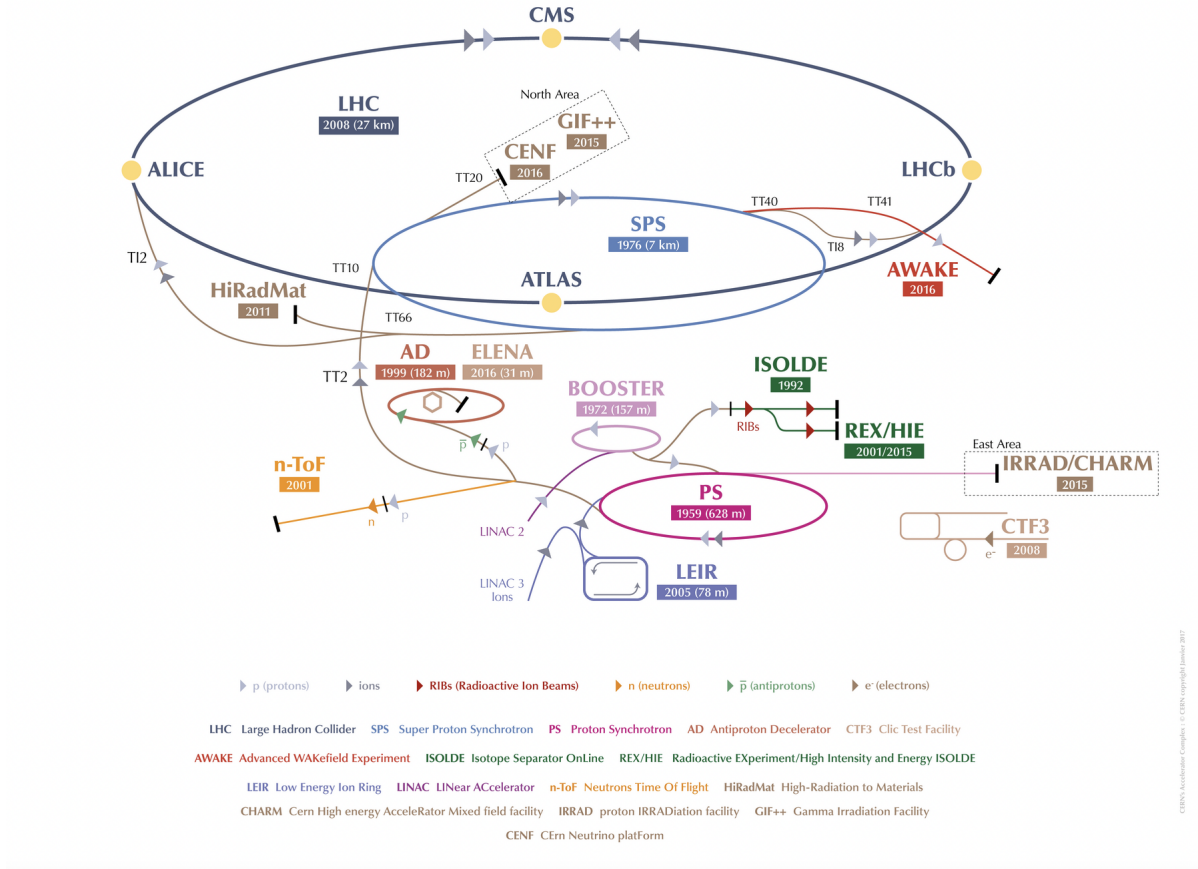
Protons are sourced from a bottle of hydrogen and end their journey when they collide at four interaction points in the LHC (see Figure 3.2). Around these interaction points are the four main experiments: ATLAS [77], CMS [78], LHCb [79], and ALICE [80]. ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) are two of the general-purpose detectors that have similar physics programs where their results are often benchmarked against each other's. ALICE (A Large Ion Collider Experiment), focuses on heavy ion collisions and primarily studies the quark-gluon plasma. LHCb specializes on studying CP violation, or the investigating slight differences in matter and antimatter using “beauty” quark decays. Four additional, auxiliary experiments also operate through these points, TOTEM, LHCf, MoEDAL and FASER.

The CERN Accelerator Complex

Before the protons or lead ions reach the collision points, they first have to go through a series of accelerators shown in Figure 3.2 to reach the desired total collision energy of 13 TeV for Run II. For the sake of brevity, we will only consider proton beams at this point but lead ions follow the same route but start from vaporized lead. Protons, on the other hand, are sourced from molecular hydrogen which are ionized. Ionized hydrogen (H^+) are then extracted and injected to Linac4 where Radiofrequency (RF) cavities are used to accelerate them to 160 MeV. They are then prepared to enter the first of the circular accelerators called the Proton Synchrotron Booster (PSB). Before they enter the PSB, the ions are stripped of their electrons and the protons are accelerated to 2 GeV. This prepares them for injection to the Proton Synchrotron (PS) which increases the beam energy to 26 GeV. The final step prior to injection to the final LHC ring is in the Super Proton Synchrotron (SPS) where the beams reach up to 450 GeV. The proton beams take around 20 minutes to reach their maximum energy of 6.5 TeV in the LHC ring and they can circulate for many hours inside the LHC beam pipes during normal operations. During collisions, the counter-rotating beams

meat at the interaction points inside the four primary detectors described earlier.

Figure 3.2: The CERN Accelerator complex has a series of accelerators that ramp up particle beam energies to 6.5 TeV in Run II.



Collider Design Fundamentals

Creating new particles and probing the underlying structure of matter requires energy and momenta - quantities whose magnitude are dictated by Einstein's equation on Mass-energy equivalence $E^2 = (mc^2)^2 + (pc)^2$ and de-Broglie's relation. The LHC achieves high energies through its near perfectly circular design with a large radius, $r \approx 4$ -km. In this section we discuss the advantages and disadvantages of these design choices to guide future colliders.

Circular colliders have a distinct advantage over linear ones. The first one is high collision rates. Particle beams could circulate over the 27-km ring at a rate of 11 kHz, for many hours

allowing it to achieve the highest nominal collision rates at 40 MHz. This collision rate is calculated by multiplying the number of bunches times their revolution frequency - which are experimentally modified. Linear colliders, on the other hand, have the disadvantage of having to be refilled with bunches at each collision.

The drawbacks [81, 82] for circular colliders mainly come from synchrotron radiation loss. Circulating charged particles means there are energy losses with the following proportionality rate:

$$P \propto \frac{q^2}{m^4 R^2}. \quad (3.1)$$

Second, a large magnetic field is also required to maintain the circular path of the particles. Assuming an idealized homogeneous dipole oriented along the particle orbit, the magnetic field strength required is:

$$|\vec{B}| = \frac{|\vec{p}|}{qcR}. \quad (3.2)$$

For the LHC, these radiation losses are mitigated by having a large-radius accelerator. This also simultaneously contributes to having a larger magnetic field. Interestingly, the LHC uses the repurposed tunnels of the Large Electron Positron (LEP) Collider, the largest radius of any collider in the world.

The LHC from the name itself, collides hadrons which are composite particles made of two or more quarks. Most hadrons are unstable, in contrast with Protons, or heavy elements like lead (**Pb**) or gold that have long lifetimes making them good choices for particle collisions. In our analysis, we only examine proton-proton collisions. A complication that arises from colliding a composite particle, is that the full momentum is shared by the actual interacting particles. Electrons, which are point-particles are much simpler, i.e. we know exactly the energy of the initial state, i.e. if we were to collide, an electron and a positron. In their simplest representation, protons (uud), are composed of two up quarks and a down quark and gluons which bind these components together. In addition, sea quarks arise from the energy of gluon splitting. The constituent quarks or gluons, collectively known as partons carry only

a fraction of the energy and this fraction is determined by the parton distribution functions. Parton distribution functions give the probability to find partons carrying a fraction x of the proton's total momentum. Apart from the primary collision event or the hard scattering event, there are also softer scattering events which are called underlying events.

The LHC program transitions to the High-Luminosity (HL-LHC) phase beginning in Run3 and ends by around 2040. At this time, a Future Circular Collider (FCC) is expected to push the energy and intensity frontier of particle colliders, with the aim of reaching collision energies up to 100 TeV [83]. Figure 3.3 shows the LHC and HL-LHC scheduled in different phases, while Figure 3.4 shows the updated LHC schedule accounting for the delays due to the COVID-19 pandemic.

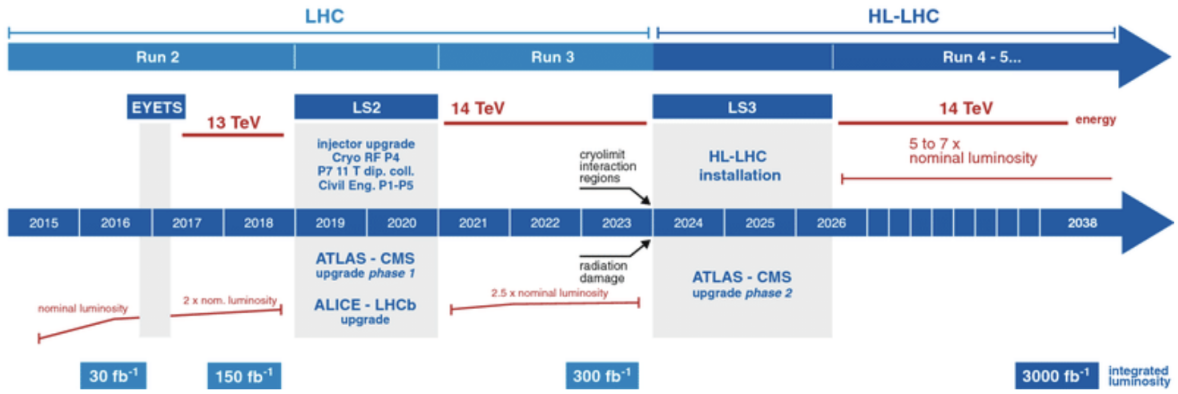
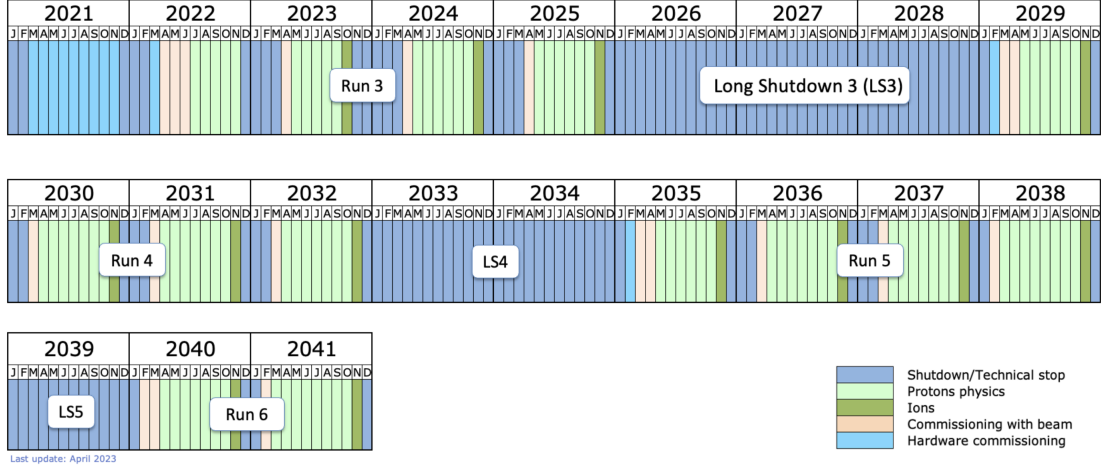


Figure 3.3: An overview of the LHC schedule from Run2 onwards before the COVID-19 pandemic delays.

3.2 LHC Schedule, Beam Performance and Pileup

The LHC operates in production cycles consisting of Runs and Long Shutdowns (LS). In Run 1 of the LHC, which took place from 2010-2012, the LHC operated at 7-8 TeV centre-of-mass energy. This was followed by a 3-year Long Shutdown period where upgrades and repairs were performed. Run 2 started in the summer of 2015 and ended in 2018. During this time, the LHC delivered 163.54 fb^{-1} of data at 13 TeV which surpassed the design luminosity goal of $L = 10^{34} \text{ cm}^{-2}\text{s}^{-1}$. This number is generally referred to as the integrated luminosity,

Figure 3.4: An updated schedule of the LHC accounting for the delays due to the COVID-19 pandemic (Credits: LHC-Commissioning)



$$L = \int \mathcal{L} dt.$$

The luminosity is different from the collision rate. The proton beam bunches cross at 40 MHz, but not all of the protons within the beam collide as the proton bunches have to be focused into a smaller cross-section. In terms of the beam parameters, the instantaneous luminosity is defined as

$$\mathcal{L} = \frac{N_b^2 * f n_b}{4\pi \sigma_x^* \sigma_y^*} F, \quad (3.3)$$

where N_b is the beam intensity or the number of protons/bunch; n_b is the number of bunches, f is the revolution frequency of the protons, $\sigma_{x,y}^*$ are the transverse RMS beam widths at the interaction point; and F is a geometric loss factor that accounts for crossing angle, shape, and longitudinal beam size.

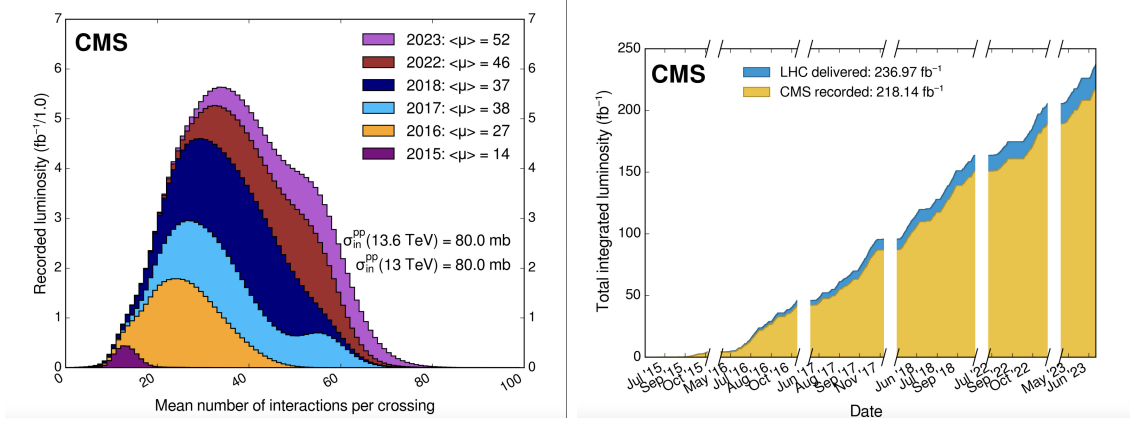
For a given physics cross-section, the total number of events produced is related to the Luminosity as such:

$$N_{\text{events}} = L\sigma. \quad (3.4)$$

While higher luminosities provide us more data, it comes with challenges. In each of the

proton bunch crossing, multiple simultaneous interactions occur. This problem is known as **pileup**. This problem can be mitigated by attempts to reconstruct the primary collision vertex but there are also other spurious events that could lead to problems in the reconstruction of physics objects. Additionally, out-of-time pileup can also occur. These arise from collisions coming from prior or later bunch crossings. These factors are major driving forces in the design of CMS algorithms and future detector upgrades.

Figure 3.5: The mean number of pileup or inelastic interactions per proton bunch crossing in data for proton-proton (pp) collisions for various years (left). A total inelastic pp collision cross section of 69.2 mb is chosen [84]. The Luminosity is shown for various time with white spaces for periods the LHC is not operating (right).

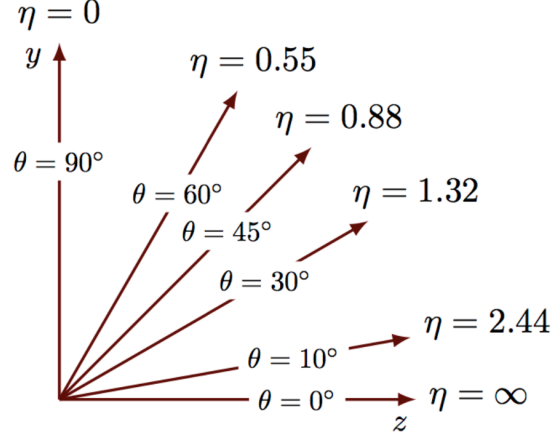


3.3 CMS Experiment and Its Subdetectors

The CMS detector (Figure 3.7), is one of general-purpose detectors of the LHC which collects the data generated at the CMS interaction point found in P5 in Cessy, France. The detector has a series of concentric cylindrical layers which comprise the Barrel region of the detector. To capture collision products going in the more forward regions, at angles moving closer to the beam line, Endcap detectors are also constructed. These layers hermetically seal the central collision point.

The CMS uses a coordinate system that is centered at nominal interaction point which is found halfway through the cylinder axis of the detector. The z-axis goes through the beam

Figure 3.6: The relationship between η and θ .



pipe, and the radial coordinate, r is measured from this axis. The angular coordinates ϕ and η are defined as follows, ϕ corresponds to the azimuthal angle or the horizontal angle going clockwise around the z -axis, and η is the pseudorapidity which is related to the angle θ by the equation (see Figure 3.6),

$$\eta = -\ln(\tan(\theta/2)). \quad (3.5)$$

From the interaction point, collision products traverse layers of materials that are designed to identify, and absorb the energy of most of the particles. These layers are assembled in the following order: 1.) Tracker System, 2.) Electromagnetic Calorimeter, 3.) Hadron Calorimeter, 4.) Muon System. The particles first traverse the Silicon Tracker system which records the position or “hits” of charged particles.

The CMS silicon tracker system, records particle paths accurately but are of lightweight material such that they disturb the particle trajectory as little as possible [78]. In addition, they tolerate high radiation fluxes since they are close to the interaction point. Combining the information collected one could reconstruct charged particle p_T , and the particles could be traced back to a specific vertex. This information is used in pileup mitigation.

The next two layers of the CMS detector are the calorimeters that measure the energies of electromagnetic and hadronic particles. Particles going through matter produce a shower, where each interaction with atomic nuclei produces secondary particles, which goes into a repetitive process of cascading particles until the full energy of the initial particle is spent. Ideally, full shower should be captured by the calorimeter so that a more accurate energy value could be reconstructed. We discuss them in more detail in the next two subsections.

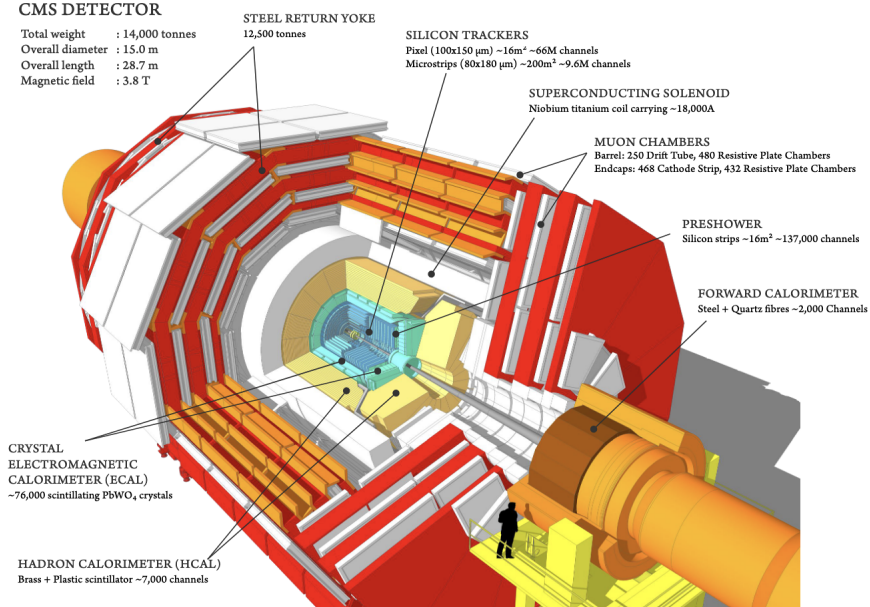
Finally, the Muon system characterizes the muons which are minimally ionizing and can escape the calorimeters due to their long life-times and low-energy deposition characteristics. The muons identify muons, measure their momenta, and provide signals for triggering on them via four complementary systems arranged in the steel flux-return yoke of the CMS solenoid. These systems cover a large range of pseudorapidity. The location of the magnetized steel behind the calorimeters and solenoid ensures a low probability of penetration by non-muon or non-neutrino candidates.

The drift tube system (DT) in the barrel covering $|\eta| < 1.2$ has the task of providing precise spatial measurements and trigger information through drift chambers with rectangular cells. Next, a cathode strip chamber (CSC) system in the endcap covers the forward region with pseudorapidity $0.9 < |\eta| < 2.4$. It has the same task as the DT but is comprised of a different material because of the higher flux of particles in the endcaps, and it also has a faster response time. Additionally, RPCs or resistive plate chambers – double-gap chambers operated in avalanche mode are placed in both the barrel and endcap to complement the DTs and CSCs to unambiguously identify the bunch crossing corresponding to a muon trigger candidate. It also has a rapid response time [85].

3.3.1 ECAL

The Electromagnetic calorimeter or ECAL, is the second layer of subdetectors that particles go through after the tracker. The ECAL is designed to measure the energies of particles that interact electromagnetically, particularly photons and electrons. It is made of a homogeneous and hermetic lead tungstate (PbWO_4) crystal calorimeter. Homogeneous calorimeters

Figure 3.7: A cutaway view of the CMS detector showing its components [86].



act both as absorbers and scintillators. As the incident particle interacts with the atomic nuclei of the crystal, electromagnetic showers are induced. Ionization is induced in the atoms and light is emitted as they de-excite. In short, the scintillators convert energy from the showers to light that is in proportion to the incident particle's energy.

There are about 75000 lead tungstate crystals, 61200 of those are spread across the barrel section (EB), located 1.29 meters from the beam line, and the rest are placed at the circular endcaps (EE) which are at ± 3.14 m from the detector origin. Each endcap seals the barrel, with 7324 crystals which gives coverage up to $|\eta| < 3.0$. The EB covers a pseudorapidity range of $|\eta| < 1.479$ where the crystals are laid out in a 170×360 η - ϕ grid. In the EE, the crystals are placed in an x-y grid, with a pseudorapidity coverage of $1.479 < |\eta| < 2.5$.

There are dimensional differences in the EB and EE crystals. The EB crystals have dimensions of about 22×22 mm² in the front-and-rear-facing sides and are 230 mm in length, corresponding to 25.8 radiation lengths X_0 . The radiation length of a material is the mean length for an electron to lose all but $1/e$ of its original energy. The EB gives a

$\Delta\eta\text{-}\Delta\phi = 0.174 \times 0.0174$ granularity. To prevent particles from slipping through the spaces in between, the crystals are tilted at 3° with respect to the interaction point normal. The endcap crystals have dimensions of $28.6 \times 28.6 \text{ mm}^2$ in x-y and has a length of 220 mm which corresponds to $24.7X_0$.

To facilitate assembly of the detector, the crystals are arranged into groups supermodules in the EB and supercrystals in the EE. There are 36 identical supermodules in the EB and each of these covering half of the barrel length. The supermodules are made up of 4 modules that contain 2 rows of 5 crystals. In the EE, the supercrystals are 5×5 structures of crystals, where the inter-crystal gaps are $320 \text{ }\mu\text{m}$. Each supercrystals have a gap of 2 mm between them.

Additionally, a preshower detector (ES) is placed in front of the EE with a pseudorapidity coverage of $1.653 < |\eta| < 2.6$. These are 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 absorbers and 2 planes of silicon strip detectors which act mainly as scintillators. These provide about $3X_0$ of material which impacts the energy resolution in the EE relative to EB.

The energy resolution as a function of the energy E is as follows:

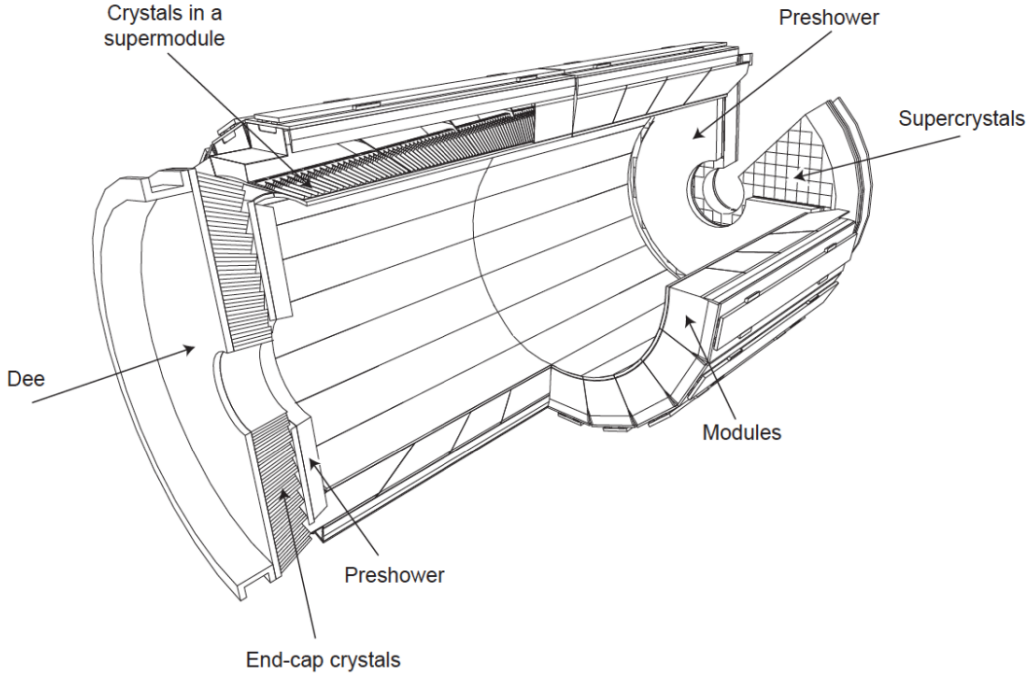
The energy resolution of the ECAL as a function of energy E is characterized by separate terms added in quadrature of the form

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \quad (3.6)$$

where a is a stochastic term related to the statistical nature of shower containment, b is a noise term accounting for electronic terms, and c is a constant term dependent on the intercalibration of the readout channels. These terms are measured in test beam studies to be $a = 2.8 \%$, $b = 12 \%$, and $c = 0.3\%$. At high energies, the constant term dominates but is corrected for by the laser monitoring system.

To amplify the light yield from the crystals, solid-state avalanche photodiodes (APDs) and vacuum phototriodes (VPTs) are fitted to the rear ends of the EB and EE crystals respectively. VPTs, a type of photomultiplier are chosen for the EE since magnetic field lines are bent there and higher radiation levels are present. There are two APDs that serve each crystal and each are operated at a gain of around 50. The VPTs are also operated at a gain of around 50 but require significantly higher bias voltages. From these photodetectors, the signals are passed off to the detector electronics for processing. These are discussed further in the Trigger and Electronics subsection.

Figure 3.8: A cutaway view of the ECAL and its components [87].



3.3.2 HCAL

The Hadronic Calorimeter (HCAL) is designed to measure the energy of hadrons that are typically contained in jets. Hadronic showers are more complicated than electromagnetic showers, i.e. the cascade of different particles coming from strong interactions with the

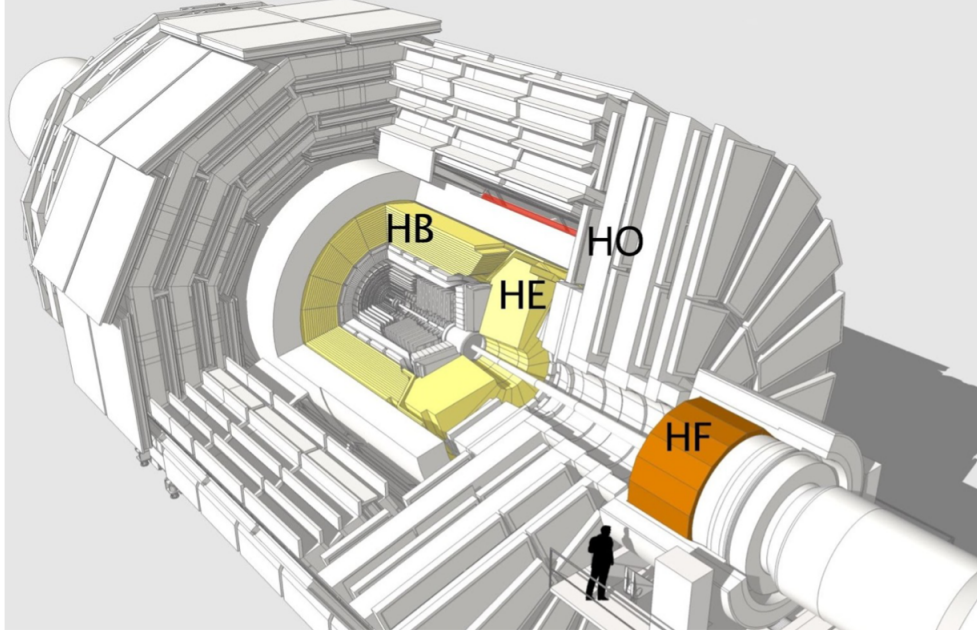
nuclei contain charged or neutral pions or neutrinos. This fact poses challenges with the energy resolution and providing coverage for missing energy. For these reasons, the HCAL is substantially longer than the ECAL to fully contain the energy of the incident hadrons.

Due to budgetary and material constraints, the HCAL cannot be made with the same homogeneous calorimeter material like the ECAL is made from. Instead, it is built as a sampling calorimeter with alternating layers of steel and brass absorber plates and plastic scintillator tiles for most sections of the HCAL. The scintillator tiles carries light through wavelength-shifting fibers to photodetectors that convert the light pulses to electric signals. The sum of the light across the entire shower is proportional to the total energy of the incident hadrons.

The HCAL is comprised of four subdetectors: The HCAL Barrel (HB) the HCAL Endcaps (HE), the HCAL Forward (HF) and the HCAL Outer (HO). The first two cover up to $|\eta| < 3$. Beyond the HE region, a complementary Cherenkov-light-based system called the HF (Forward), extends the coverage up to $|\eta| < 5$. Outside of the magnet, the HO system, with coverage of $|\eta| < 1.26$ has additional scintillators that are installed to complement and improve the energy resolution in the HB.

The barrel region of the HCAL consists of the HCAL Barrel (HB) and the HCAL Outer (HO). The HB is situated between the CMS solenoid magnet and the ECAL and spans a radial distance of $1.77 < r < 2.96$ m, with a pseudorapidity coverage $|\eta| < 1.3$. The first layer of the HCAL consists of a plastic scintillator made of 9 mm Bicron BC408 which is different from the succeeding layers of scintillators. This upgrade was made to enhance the depth-segmented readout [88] which was made to measure the hadronic showers that began in the ECAL. The information obtained from this layer directly affects our photon identification strategy. The first and last layers of are made of stainless steel absorbers for the purposes of structural support. In between are pairs of brass absorber plates and plastic scintillators. Brass was chosen for cost-effectiveness and robustness against deformation under the strong 4T solenoid field. While these material provide excellent shower containment, the

Figure 3.9: A cutaway view of the HCAL and its subdetectors [87]



HO was placed behind the solenoid magnet to provide additional detection for the shower tails extending beyond HB. It is made of 10 mm the same scintillator material found behind the ECAL and an additional layer of iron absorber. The HB+HO combination provide an effective thickness of 11.8 nuclear interaction lengths λ_I . Nuclear interaction length is defined as the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction.

The HCAL endcaps are located between $3.9 < |z| < 5.7$ and covers the range $1.3 < |\eta| < 3.0$. As mentioned earlier, the HF extends coverage in the forward region to $|\eta| = 5.2$, located 11.2 m from the detector origin. Instead of plastic scintillators, it uses quartz fibers as the active material to produce Cherenkov light. The forward region experiences the highest particle fluxes in the detector, about 8 times as much as other subdetectors and therefore the switch in material. Steel absorbers are used in the HF.

The HB+HE region provides a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, a similar coverage

of 5×5 EB crystals for $|\eta| < 1.6$ while for $|\eta| > 1.6$ the granularity is coarser $\Delta\eta \times \Delta\phi = 0.174 \times 0.174$. Across the depth of a single $\Delta\eta$ region, towers of scintillation light are formed. The plastic scintillators of the towers are connected to wavelength-shifting (WLS) fibres which are routed to hybrid photodiodes (HPD) which are then read out to the electronics for processing. We discuss more on the electronics in later sections.

In terms of resolution, the HCAL performance is much better evaluated with the jet and p_T resolutions of the ECAL + HCAL system instead of just HCAL alone. The results are similar to that described in Eq. 3.6.

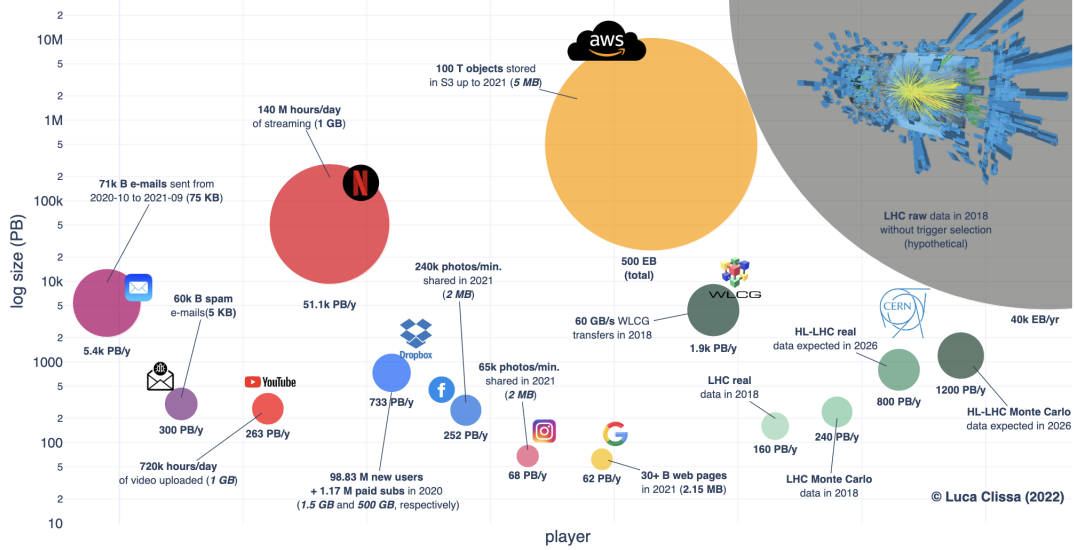
3.3.3 CMS Data Trigger and Electronics

The LHC proton-proton collisions occur every 25 ns, or 40 MHz. This corresponds to a 1 MB average event size, with 40 PB of data if every event was to be recorded. In one year alone, the data produced by the LHC is around one order of magnitude bigger than the total size of objects ever stored on commercial cloud storage services making the LHC one of the most prominent players in the Big Data era [89]. The strategy then is to determine in advance only the events that are of interest. CMS does this with a two-level filtering scheme employed to progressively reduce the data collected. The first stage of event rate reduction, called Level 1 (L1) trigger, is performed by high-speed electronics and uses coarser-grained detector information. The event rate is reduced from 40 MHz to about 100 kHz. In the second stage, the high-level trigger (HLT), is software-based menus to further reduce the rate even further from 100 kHz to 2kHz. Unlike L1, HLT uses the full detector granularity, i.e. the information from all the subdetectors including the inner tracks. Only events that pass through both the L1 and HLT are stored for offline physics analysis. Figure 3.10 shows how untriggered LHC data compares other Big Data players.

L1 Trigger

The L1 Trigger system makes a decision to accept an event within $3.2 \mu s$, with every single step in the decision-making chain made before the next proton-proton bunch crossing, i.e.

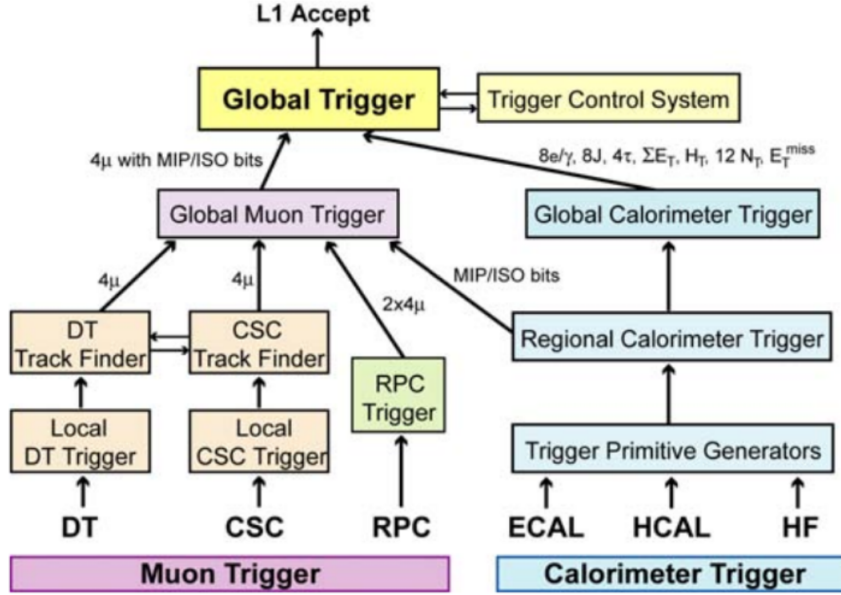
Figure 3.10: An illustration comparing the LHC-produced data without trigger selection to the other Big data players [89].



within less than 25 ns. This speed of decision-making requires custom-programmed FPGAs or ASICs that are situated in the front-end or what is called on-detector electronics. These hardware are characterized for their radiation tolerance being situated so close to the detector components.

The architecture of the L1 system is shown in Figure 3.11. An event is “L1 Accepted” based on aggregated information from the calorimeter and muon systems, both of which are organized into a hierarchy of local, regional and global components. For the Calorimeter Triggers, the first step are the local triggers also known as Trigger Primitive Generators (TPGs) which collect E_T deposits from the ECAL or the HCAL trigger towers that have a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The local TPGs also correct bunch crossing information. In the next step, the TPGs are passed on to the Regional Calorimeter Trigger (RCT) which sums the TPG trigger towers into RCT towers (4×4 TPG trigger towers). At this step, crude electromagnetic particle candidates are formed. Finally, a Global Calorimeter Trigger receives the information from the RCT to construct jet candidates, and

Figure 3.11: A cutaway view of the HCAL and its subdetectors [78]



missing E_T candidates and more complex electromagnetic particle candidates. The Muon Trigger has a similar pipeline at L1 begins with information coming from the DT, and CSC muon trackers which provide track hits. The aggregated information is passed on to the Regional Muon Trigger (RMT) which performs track reconstruction. The RPC Trigger information is able to reconstruct complete tracks without any intermediate steps because of its superior timing information. The CSC, DT, and RPC Trigger Chain are then collectively passed on to the Global Muon Trigger (GMT). The combined Global Muon Trigger (GMT) and Global Calorimeter Triggers (GCT) are then passed on to the Global Trigger which makes the final decision on whether to trigger on the event and pass it on to the HLT.

Events that make it through the HLT must further pass a software-based list of cuts called an HLT menus, to reduce the event rate to 1 kHz. An on-site commercial processing farm of more than 10,000 cores, performs the event reconstruction. In this case, there is more time flexibility with triggering using the full detector granularity. Events that make it through this final step are then recorded for physics analysis.

The HLT menus consists of a set of simple cuts on single objects or events based on topology. For example, in this analysis, we are interested in events with a two-photon final state. The algorithms used by the HLT are similar to that of the offline Physics Reconstruction which we will discuss in the next chapter.

3.3.4 HCAL Online Software and HCAL Upgrades

A series of upgrades have been planned for CMS with an eye to prepare the detector for the challenges of the High-Luminosity (HL-LHC) era. The first of the of the series termed Phase 1, in between Run 2 and Run 3, in a period called Long Shutdown (2). In this section we discuss the HCAL Phase I upgrade which was completed in October 2019 in time for Run 3 which was unfortunately pushed later to 2022 due to the COVID-19 pandemic. My contributions to HCAL encompassed upgrades in both HCAL Online Software and both Front-End (FE) and Back-End (BE) Electronics. My major contribution on the HCAL Online Software was work on providing software support for the improved timing information to create an L1 trigger on Long-Lived Particles (LLP), under the supervision of Seth Cooper and Aleko Khukhunaishvili. I also participated in the refurbishing and accounting of Front-End (FE) and Back-end (BE) electronics which was performed with the team consisting primarily of Grace Cummings, Mary Hadley, Sebastian Yan, Candan Isik and Jay Dittman.

In the first two parts, we first give an overview of the FE and BE electronics in Section 3.3.4 to guide the reader into the idea behind the L1 trigger on Long-Lived Particles (LLP).

Front-End and Back-End Electronics

The Front-End and Back-End electronics data flow diagram is shown in Figure 3.12. The Front-End electronics are composed of units called read-out boxes (RBXs) that contain the following: 4 read out modules or (RMs), 1 calibration unit (CU), Clock, Control and Monitoring Modules (CCMs) and a backplane that connects all the components within an RBX. Each of the RMs cover about a 5 degree division in ϕ . Each RM is a card pack

that has the QIE or charge integration boards and light detectors. The RMs combine light into readout towers and converted into digital data and are transmitted to the back-end electronics. Figure 3.13 shows a schematic of the RBXs and Figure 3.14 shows an example RBX set-up at the HCAL test facility at Building 904.

The HCAL Back-end electronics are responsible for receiving the continuous readout information coming from the front-end electronics, calculating and transmitting trigger information, storing and sending the front-end trigger data when a L1-accept has been received from the CMS global trigger [90]. Figure 3.12 shows a schematic diagram for the backend. It also has a local data acquisition and triggering mode, and also handles fast clock and control signals. Back-end crates are called μ TCA crates which contain an Advanced Mezzanine Card (AMC13) which is responsible for event building and linking the back-end to the CMS central data acquisition (cDAQ) system. The μ HTR performs two functions via the two Xilinx FPGAs. The front FPGA takes input from the front-end data, synchronizes the data and pipelines and calculates and transmits the trigger information. The back FPGA is responsible for buffering and applying zero suppression and transmits data to cDAQ. The upgraded electronics include increased granularity and feature bits based on TDC information which will be relevant in our work on the software support for the HB TDC LUTs.

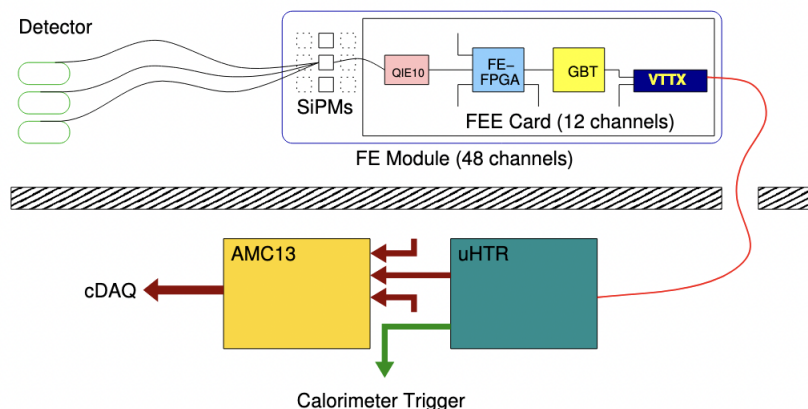
I was part of the last phase of the ngCCM refurbishing project and I also helped coordinate the start of the front-end and back-end electronic spares bookkeeping and preparation. The refurbishing of the ngCCMs is largely because of a communication issue loss which was attributed to a manufacturing weakness in the VTRxs. The VTRx optical receivers were found to be temperature-sensitive with temperature increases causing outgassing material to obscure the central part of the optical fibers where light is transmitted [91]. To mitigate the issue, heat sinks were added within the ngCCMs.

I also began the preparation of the new AMC13s which were production-tested by Chris Cosby from Boston University. I also updated the firmware to match the latest firmware at P5 which was $T1, T2 = 0 \times 2266 \ 0 \times 34$. As there was not a good criteria as to what would

qualify as a good spare, I went through several testing steps such as local event building tests and taking local runs with the new AMC13s loaded in the mock up test area of Bldg. 904 at the CERN Preveessin site. We determined that the final step before swapping them at P5¹ was to do a 904 miniDAQ testing. Local runs do not test cDAQ links. As the miniDAQ software was not well configured before my stay at CERN ended, I delegated the tasks to the rest of the team who would take over the project. Figure 3.15 and Figure 3.16 shows the reworking and testing setups for the AMC13 and ngCCM testing. Credits to Jay Dittman and Grace Cummings for the set-up.

In the first half 2021, I have also participated DAQ On-call and DCS Technical shifts at P5 as well as in the updating of the firmwares for the igloo2 FPGAs in the HF ngCCMs located at the CMS cavern shown in Figure 3.17.

Figure 3.12: Schematic diagram of the data flow in HB/HE starting from the photomultipliers through the QIE, FPGA and to the backend and AMC13 and to the CMS DAQ [90].



HB TDC LUTs Support for Long-lived Particles

CMS can take full picture events at 40 MHz but can only record 1000 full events per second. The idea is to construct a Level-1 (L1) Hardware trigger using the HCAL TDC information within each bunch crossing of 25 ns. Currently, hadronic Long-Lived Particle (LLP) trigger

¹Jay Dittman also proposed that we could swap the five spares with the AMC13s that are known to be working well at P5 so we have 'good hot spares'.

Figure 3.13: A schematic diagram of an HB RBX containing the 4 RMs, 2 ngCCMs and the backplanes [88]

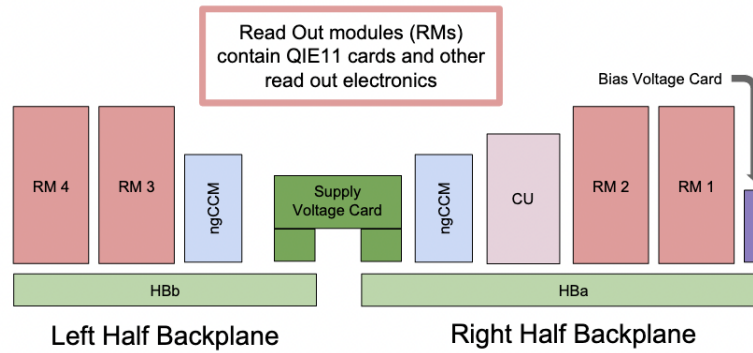


Figure 3.14: Image of an RBX at Building 904 where a mock-up of the CMS HCAL Front-End electronics are set up for testing.

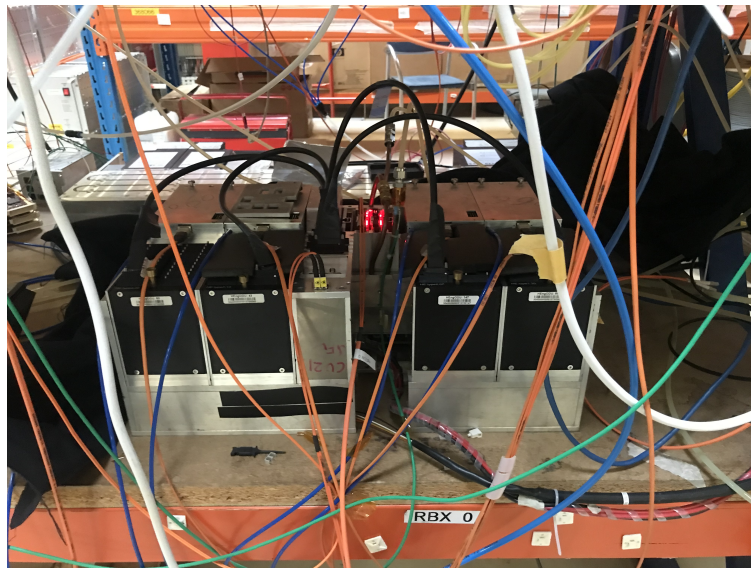


Figure 3.15: On the left is a new AMC13 tested at 904 with the μ HTRs in a selected Back-End Crate shown on the right.



can cover lifetimes greater than ns or microseconds. It is currently at L1 trigger level where the heaviest constraint lies when it comes to distinguishing prompt particles from long-lived ones.

One way to workaround this constraint is to take advantage of the limited bandwidth information. In the HB, only 2 bits/channel are allowed to be sent from the Front-End (FE) electronics to the Back-End (BE) (see Figure 3.18). This constraint does not exist for both the HE and the HF where the full 6 bit/ channel are transmitted through the data links from FE to the BE. In the Phase I upgrade, each QIE channel now has a time-to-digital converter (TDC), that sends out a 6 bit code. This is illustrated in a schematics shown in Figure 3.19. The Igloo2 FPGAs pack the 6-bit information into meaningful 2-bit information through physics Look-up tables (LUTs). In computer science, Look Up Tables maps keys to some value. In this context, the LUTs map the available TDC codes to the time windows within the bunch crossing of interest. Previously, these LUTs were only hard-coded mappings from

Figure 3.16: The top picture shows an ngCCM being reworked to add heat sinks to mitigate the VTRx RSSI drift issue which was related to the communications loss in the HCAL endcap. The right shows a reworked ngCCM being tested on a spare backplane.



Figure 3.17: Reprogramming of the HF Igloo mounted on the HF ngCCMs at the CMS cavern.

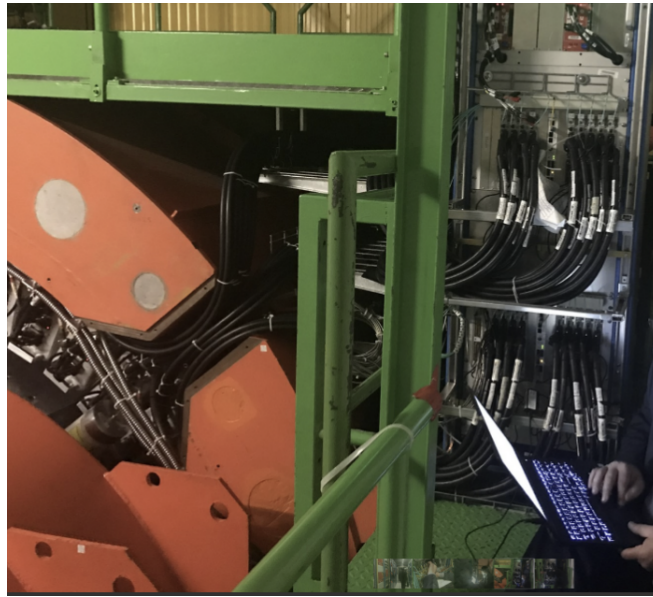


Figure 3.18: TDC Current support. Credits: Tullio Grassi

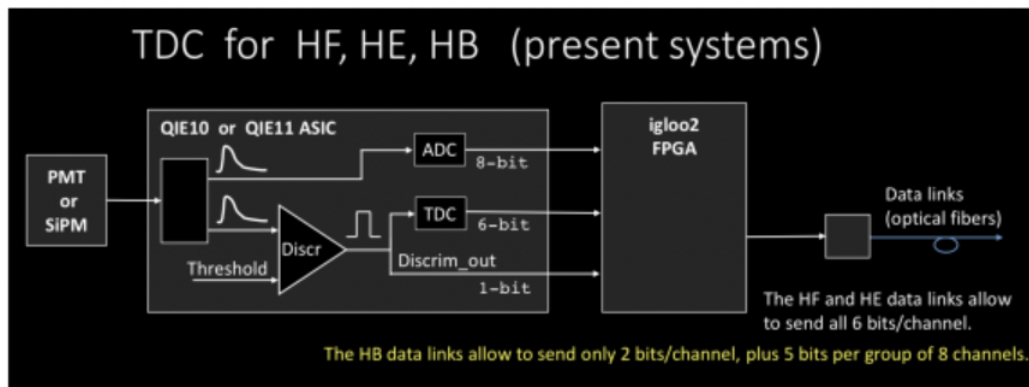
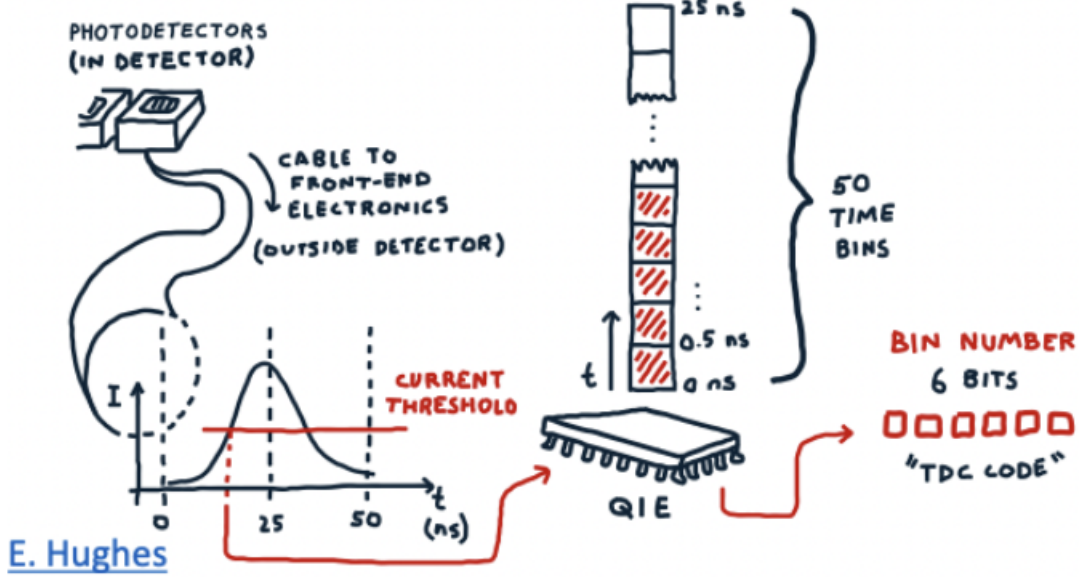


Figure 3.19: With the Phase I Upgrade, each QIE channel now has a time-to-digital converter (TDC) which sends out a 6-bit TDC Code. Credit: E.Hughes



the 6-bit to 2-bit information. There are several disadvantages to this. One, the FPGA Igloo2 Firmware would need to be updated each time a new Look Up Table is to be used or investigated. Updating the firmware could be time consuming and also could run into the danger of turning the hardware into “bricks” when the updating process is interrupted. Not only are the FPGAs expensive to replace, one would also have to wait for a time-window to be able to replace them manually underground. The idea then is to make the LUTs configurable and loadable prior to each data taking run.

Currently, the software support we implemented for the HB TDC Look Up Tables, takes in xml files. An example xml file containing the LUTs is shown in Figure 3.20. They are indexed by RBX, QIE, RM which has a one-to-one map to the detector coordinates indexed by η , ϕ , and depth.

An example of a 6-bit to 2-bit mapping as shown in Figure 3.21. The 6-bit TDC code information comes from the QIE11, which is the processed output of the silicon photomul-

Figure 3.20: Sample LUT. The RM index ranges from 1-5, where the 5th is the Calibration Unit. Labeling credits: Gillian Kopp

```

RBX <Parameter name="INFOTYPE" type="string">HBTCLUT</Parameter>
<Parameter name="CREATIONSTAMP" type="string">2021-10-12</Parameter>
<Parameter name="CREATIONTAG" type="string">HBTestTag</Parameter>
<Parameter name="RBX" type="string">HBP01</Parameter>
<Data elements="1" encoding="hex" qie="1" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0xaaa50000</Data>
<Data elements="1" encoding="hex" qie="2" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0xaaa50000</Data>
<Data elements="1" encoding="hex" qie="3" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0xaaa50000</Data>
<Data elements="1" encoding="hex" qie="4" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0xaaa50000</Data>
<Data elements="1" encoding="hex" qie="5" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0x94000000</Data>
<Data elements="1" encoding="hex" qie="6" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0x94000000</Data>
<Data elements="1" encoding="hex" qie="7" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0x94000000</Data>
<Data elements="1" encoding="hex" qie="8" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0x94000000</Data>
<Data elements="1" encoding="hex" qie="9" rm="1">0xffffffff 0xaaaaaaaa 0xaaaaaaaa 0x94000000</Data>

```

QIE RM 64 2-bit TDC codes

Figure 3.21: An example hardmapping of 6-bit to 2-bit information by Tullio Grassi

6-bit TDC	2-bit code	meaning
0	0	Valid TDC
1	0	Valid TDC
...		...
7	0	
8	1	
...		
15	1	
16	2	
...		
31	2	
32	3	Valid TDC
...		Valid TDC
49	3	Valid TDC
50	0	Unused
...		
63	0	special

tiplier pulse coming from the detector. Each bunch crossing happens every 25 ns, so the TDC values range from 0-49 in half ns steps to cover the whole time range. Additional error codes are also encoded for example in cases where the pulse starts high (62) or starts low and stays low (63).

The 6-bit TDC address from QIE11 is packed into 2-bits in the IGLOO2 firmware LUT, to encode the following 4 states:

- 00 prompt pulse: $TDC \leq t_p$
- 01 slightly delayed pulse: $t_p < TDC \leq t_p + 2$
- 10 very delayed pulse: $t_p + 2 < TDC < 50$
- 11 invalid pulse: $TDC > 50$

To determine exactly the time boundary between the states is ascertained from the the distribution of TDC values in a prompt QCD sample and is currently being studies as of time of writing [92]. One of the potential downsides of loading LUTs was the effect on the configuration time. Prior to every run, the configuration of the various electronics components are set. The new software version was tested at P5 and only an average of 7s was added to the Configuration time.

CHAPTER 4

PARTICLE RECONSTRUCTION AND EVENT SELECTION

The CMS Detector is composed of layers of subdetectors each of which provide information from the footprints left behind by the particles that come out of the collisions. The data across all the subdetectors are complementary and are linked to reconstruct Physics Objects [93]. For most CMS analyses, the Particle Flow Algorithm is used to construct idealized representations of physics objects such as photons, electrons, muons and jets. In the algorithm, tracks are first reconstructed through an efficient track reconstruction algorithm. If these tracks fall within a defined boundaries of one or several clusters, these tracks are associated with the cluster via a linking procedure. A separate clustering algorithm is used to disentangle the overlapping showers. Muons that go all the way to the end of detector are determined first as their tracks do not give rise to a charged hadron. Multiple clusters across various subdetectors and have an associated track are considered charged hadrons. The electrons, due to Bremsstrahlung photon emissions, uses a tailor-made track reconstruction to properly attach these photon cluster to the electron it originated from. This also avoids energy double counting. Once all the tracks have been considered, the remaining clusters without any associated tracks are used to determine the neutral particles. Energy clusters concentrated in the ECAL and without associated tracks, are categorized as photons, the primary physics objects used for this analysis. In the case of neutral hadrons, most of their energies are contained in the HCAL. Once the once tracks are linked to appropriate clusters, for charged particles, and the neutral objects are determined, they are linked to the best fit primary vertex, the nature of the particles are assessed and a four-momenta is calculated.

Further details on the Particle Flow Algorithm and other physics objects are found in [94].

In this analysis, the events of interest are those with two high-energy, prompt photons. The events containing this signature are first required to pass a trigger selection and an offline reconstruction is performed. To obtain even higher purity and higher search sensitivity, these diphoton candidates are required to pass additional pre-selection and various levels of identification criteria called photon IDs.

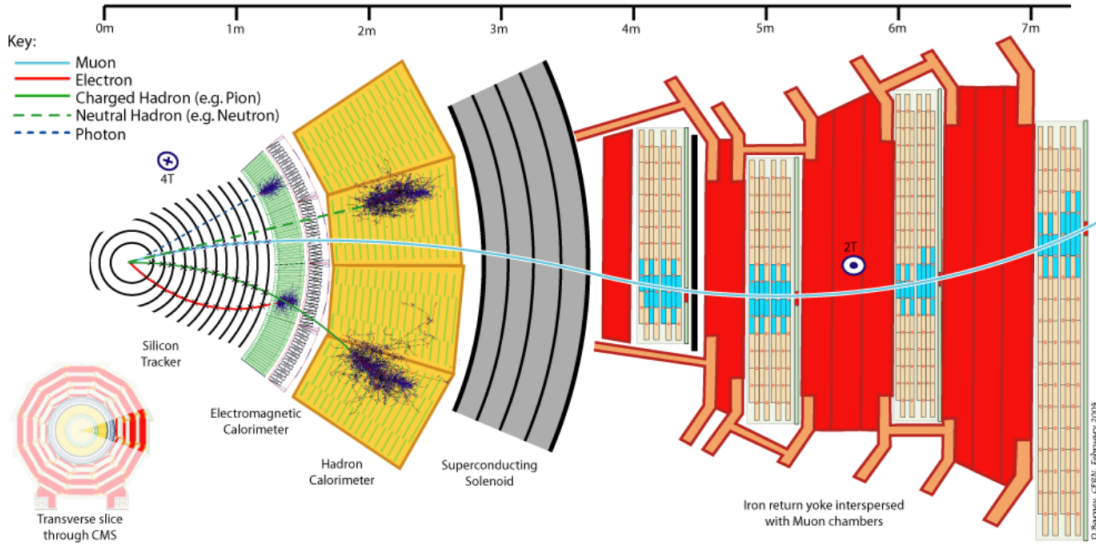
In the succeeding pages we discuss the Particle Flow elements and Particle Reconstruction, event selection, triggers used in this analysis. The second half of the chapter will discuss the Photon ID 4.2.2 used for high-mass diphoton events, the data sets used for the analysis as well as the known detector issues that might affect the results of the analysis.

4.1 PF Elements and Physics Objects

The Particle Flow (PF) elements are the basic building blocks used for particle reconstruction for CMS. These are mainly the *Tracks and Vertices* and the *Calorimeter Clusters*. The PF algorithm makes use of the track and vertices to identify and measure the momentum of charged particles and also identify displaced vertices, particularly from b-quark jets. The energy deposits in the calorimeters are clustered by the algorithm and used to identify neutral particles such as photons. These are also used to augment the reconstruction of physics objects like electrons and charged hadrons in jets. The information from the vertices, tracks and clusters are then chained, to the best estimate of the algorithm to form the basic physics objects. This is shown in a schematic in Figure 4.1. More detailed information on the algorithm is found in reference [94]. While not used in this dissertation, it is interesting to note that novel machine learning methods have recently been developed to directly exploit detector level data without going through the hand-engineered features that the particle flow algorithm has to categorize particles [95, 96].

For this analysis, we are mostly interested in photons, muons and jets. We are interested in signal events containing two photons and some of our background estimates are derived

Figure 4.1: The information from the vertices, tracks and calorimeter energy deposits are chained by the PF algorithm to form the various physics objects.



from jet and muon datasets.

4.1.1 Tracks and Vertices

Only charged particles can leave discernible tracks in the Silicon Tracker layers due to their interactions with the doped silicon material where impurities in the silicon crystal lattice regions that have excess positive or negative charge. The proper reconstruction of these tracks are fundamental as they are essential for reconstructing the full event topology, as well as the identification of other particles. They contribute towards the subsequent reconstruction of other physics objects in an event. To reconstruct the trajectory of the particles, the tracker hits are fit into the path that likely belong to the same particle track simultaneously discarding irrelevant hits while taking into consideration the hit position uncertainty, the effects of multiple Coulomb scattering as well as energy loss. To reconstruct the tracks, CMS makes use of the Combinatorial Track Finder algorithm (CTF) based on a Kalman filter. The latter is a technique used to recursively estimate quantities such as position, momentum, etc. based on its previous state and updating this prediction using new measurements from the tracker hits. This entire process is described in more detail in

in Ref. [97].

Since there are approximately 10^{11} protons in each bunch crossing, there are numerous proton-proton collisions occurring simultaneously resulting in pileup interactions, in addition to other interactions coming from particle decays after the hard interaction and interactions with detector materials and cosmic rays. This makes the determination of the primary vertex a challenge. To reconstruct the primary vertex, first, the tracks that are compatible with the LHC beam spot are selected and clustered. This beam spot corresponds to the 3D luminous region where the LHC beams collide. This track clustering algorithm is described in more detail in reference [97]. Multiple primary vertices can be reconstructed in an event due to pileup but the largest summed charge-particle track p_T^2 , plus some constraints, is chosen as the main primary vertex. Ref. [98] gives more details on the deterministic annealing algorithm used for vertex reconstruction. The paper also achieved a position resolution in each of the 3 spatial dimensions of 10-12 μm for the reconstructed primary vertices.

For this analysis, the primary interaction vertex associated with the diphoton system is selected using Ref. [99, 100], which uses a multivariate classifier to enhance the probability of correct vertex assignment. Input variables to the classifier include the p_T^2 sum of the charged-particle tracks associated with the vertex, and two variables that quantify the vector and scalar balance of the transverse momentum between the diphoton system and the tracks associated with the vertex. Additionally, if either photon has an associated track originating from a photon conversion to an $e - e+$ pair, the conversion information is used. For signal events above 500 GeV, a 90% correct vertex assignment was observed.

4.1.2 Photon and Electron Reconstruction

Photon candidates, being neutral particles, are reconstructed from energy deposits in the ECAL that typically do not have any associated tracks in the tracker. They are evolved from a fixed 5×5 matrix of crystals. The "seed crystal" which contains an energy signal greater than all of its immediate neighbors, form as the central piece of a photon cluster [101]. The aggregated energy deposits in individual crystals beginning from the seed crystal forms the

clusters and superclusters (clusters of clusters) [102] which are compatible with the expected shower shape extending along the azimuthal (ϕ) direction. This shower shape is similar to electrons but there is a slight difference due to the effect of the magnetic field on the electron trajectory however, the clustering algorithm does not make this distinction. For this reason, the same algorithm for photon reconstruction can be applied to $Z \rightarrow e^+e^-$ events to measure the efficiency of the photon selection criteria and of the photon energy scale and resolution. About 94% of the photon’s energy is deposited in a 3×3 group of ECAL crystals and 97% in 5×5 crystals. Any candidate with HCAL energy deposit exceeding 10% of its energy from the ECAL supercluster are rejected. Photon Energy reconstruction includes energy corrections that account for detector effects. For the endcap photons, energy from the ECAL preshower are also added to the photon energy measurement. Pileup contamination and photon shower loss are also accounted for. Photon showers may be lost in the gaps between crystals or ECAL modules. To correct for these losses, the photon efficiency is calculated and the differences between efficiencies in data and simulation are corrected using scale factors (see Sec. 4.2.4). More details on the full process are found in [103].

Electrons being charged particles are reconstructed by associating an ECAL cluster with a track in the tracker. The ECAL clusters are reconstructed by summing up the energy deposits in the crystals around the “seed” crystal which has largest energy deposit. These clusters are matched to hits in the inner tracker using the Gaussian sum filter (GSF) algorithm. Both the summing and the GSF algorithms take into account the energy deposits arising from the bremsstrahlung emissions in the tracker [103].

4.1.3 Muon Reconstruction

Muon candidates are reconstructed from information from the dedicated muon system and inner tracker. These particles ionize the gas in the chambers and leave hits whose precise locations are determined more accurately. The hits in DT and the CSC are used to reconstruct track segments in the muon system. The requirements for these hits include the following: an angular cone of $\Delta R < 0.3$, and the summed p_T of any additional tracks and

calorimeter deposits should not exceed 10% of the muon track's p_T . These reconstructed muon tracks must also be compatible with a track in the inner tracker. A best track fit between the inner and outer track including information from the RPCs is employed [104].

4.1.4 Jets Reconstruction

Jets are collimated spray of hadrons that appear as a cluster of energy in the ECAL and HCAL subdetectors. These candidates are only reconstructed after isolated photon, electron and muon candidates have already been identified from the pool of particle flow candidates. There are multiple algorithms used to reconstruct jets. The anti- k_t algorithm [105] is a typical method used and is suitable for reconstructing jets with well-defined boundaries. These candidates are further classified into neutral (e.g. K^0), charged hadrons (e.g. $\pi^\pm$, $K^\pm$, protons) and non-isolated photons(π^0). Non-isolated photons have ECAL clusters that are not linked to any track. Neutral hadrons are also not linked to any track but have energy clusters in the HCAL. Reconstructed tracks matching to ECAL and or HCAL clusters form charged hadrons. Their energy deposits are corrected to account for the nonlinear response of the calorimeters. For $|\eta| > 2.5$, neutral hadrons cannot be distinguished from charged hadrons since that range is beyond the tracker acceptance. For non-isolated photons or neutral hadrons, their energies are determined by the corrected ECAL or HCAL cluster energy or the combined cluster energy of ECAL and HCAL if $|\eta| > 2.5$. Charged hadron energies are derived from whichever is the larger value between the calorimeter cluster energies or the sum of the matched track momenta [106].

4.2 Event Selection

This analysis selects events containing two high-energy, prompt photons. To select these events, an online trigger selection, and further offline selection and photon identification criteria are imposed. These further selections ensure that the photons are well-measured and isolated from other particles in the event, help reduce background from other processes and thereby increase the sensitivity of this analysis.

4.2.1 Trigger Selection

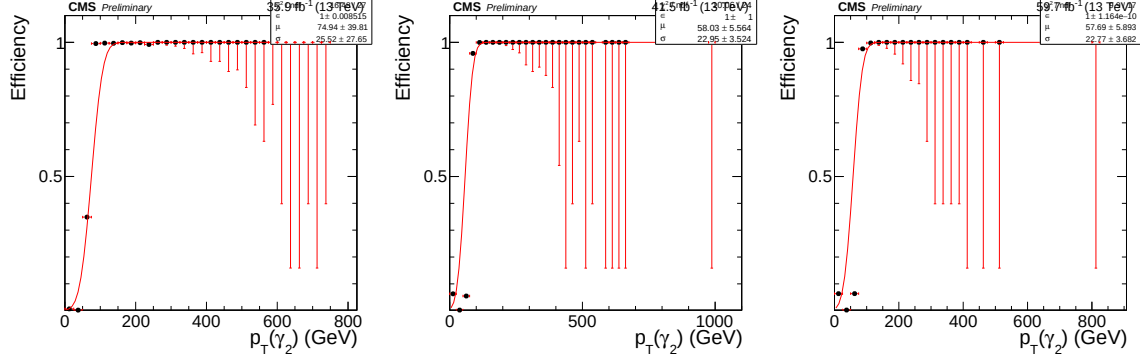
Events for this analysis are selected online by an online trigger (HLT) containing at least two reconstructed photon candidates, each with transverse momentum $p_T > 60(70)$ GeV, for the 2016 (2017/2018). We refer to these as the Double Photon Triggers. The two leading photons in p_T are chosen in events with more than two photon candidates. The trigger photon candidates are also required to have a ratio of hadronic/electromagnetic energy of less than 0.15.

To quantify the efficiency of the Double Photon triggers, we use a reference for normalization. This reference trigger selects events that contains two photons each with transverse momentum greater than 33 GeV and satisfies the Calorimeter Identification (CaloId) criteria. The efficiency is primarily a function of the p_T of the second-leading photon, the photon which has the second highest p_T among the collection of photons. The results are shown in Figure 4.2. The precise form of the fit is irrelevant as the efficiency plateaus well before the offline cut value. This procedure only probes the turn-on of the p_T leg [107] have shown that there is no additional source of efficiency and this study serves as a cross-check. An additional backup trigger was also used to mitigate potential efficiency losses at high diphoton invariant masses. We calculated an additional 0.2% yield over the full analysis region with several thousand events, without viewing any distributions to avoid bias during the blinded phase of the analysis. The trigger efficiency is fully efficient at $p_T > 125$ GeV, which is the photon p_T selection used in this analysis. For details on the exact trigger path names see Appendix B.

4.2.2 Photon Identification

The reconstructed photon candidates passing the trigger selection must pass additional identification criteria to suppress background events coming from misidentified jets and electrons while maintaining high efficiency. These criteria are based on observables sensitive to the electromagnetic shower shape, shower containment within ECAL and any extra activ-

Figure 4.2: Efficiency of the Double Photon triggers (measured with reference trigger as a function of the p_T of the second-leading photon in 2016 (left), 2017 (center) and 2018 (right) data.



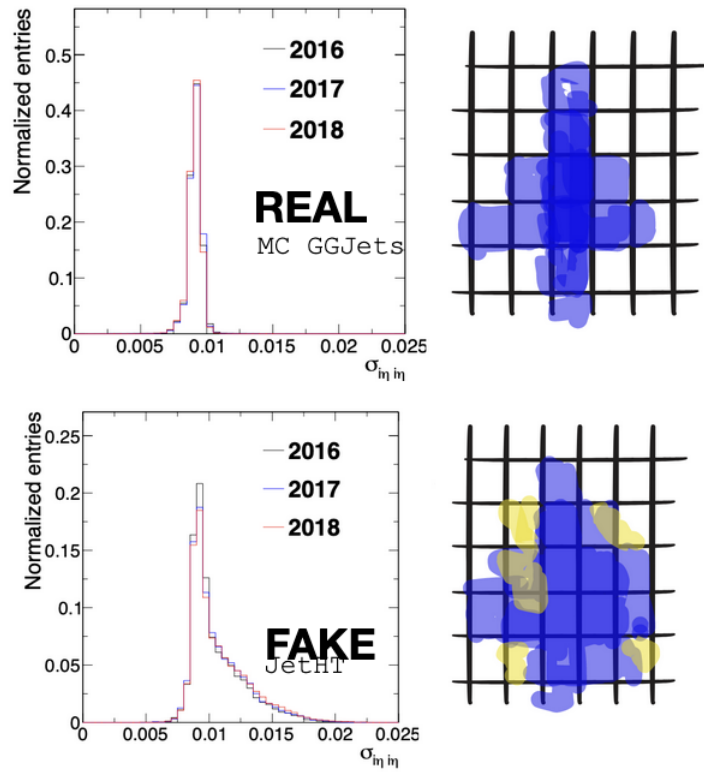
ity surrounding the shower. The electromagnetic shower is measured using $\sigma_{i\eta i\eta}$. The $\sigma_{i\eta i\eta}$ variable [102], is the spatial second-order moment of the photon candidate with coordinates $(\eta_\gamma, \phi_\gamma)$. It is defined mathematically as follows:

$$\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i w_i (\eta - \bar{\eta})^2}{\sum_i w_i}} \quad (4.1)$$

where $w_i = \max(0.0, 4.7 + \ln E_i/E_{5x5})$, η_i is the coordinate of crystal i in the 5×5 array, $\bar{\eta}$ is the average supercluster position in η , i runs over a 5×5 matrix of crystals centered on the crystal with maximum energy deposit and the sums are taken over all the crystals in the matrix. This variable is a measure of the spread of the energy distribution within the crystal matrix, along the η direction of the detector. It is a useful quantity for identifying electromagnetic showers initiated by electrons or photons, which typically have a narrow, localized distribution of energy, in contrast to showers initiated by hadrons, which are broader and more diffuse. Prompt photons have smaller values of $\sigma_{i\eta i\eta}$ corresponding to narrower showers. The variable $\sigma_{i\phi i\phi}$ is defined analogously and together with $\sigma_{i\eta i\eta}$ they form a covariance matrix which describes the spatial extension of the electromagnetic showers in the (η, ϕ) space. Additionally, a selection on the shower shape variable $R_9 < 0.8$ is imposed to reduce the contribution of jets that fake photons. R_9 is defined as the ratio of the energy

deposited in a 3×3 array of crystals centered around the most energetic crystal in the supercluster, to the total energy deposited in the supercluster. It is a measure of the degree of lateral containment of the shower. Values closer to 1 indicate a more compact shower and values closer to 0 indicate a more spread-out shower.

Figure 4.3: Prompt photons (REAL) and Fake Photons are shown on top and bottom. The figures on the left show the $\sigma_{i\eta i\eta}$ distributions for real and fake. On the right, a representative illustration of a prompt (fake) photon energy deposit in a 5×5 array of crystals is shown on top (bottom).



Photons in the event should also be well-isolated hence isolation variables which describes the degree to which a photon is isolated from other particles in an event are also used in the identification criteria. These variables are based on the total transverse momentum of the particle candidates with (η, ϕ) coordinates that are reconstructed within a cone size of $\Delta R < 0.3$ around the photon candidate with coordinates $(\eta_\gamma, \phi_\gamma)$. Here, $\Delta R = \sqrt{(\eta - \eta_\gamma)^2 + (\phi - \phi_\gamma)^2}$. Charged hadrons (Iso_{Ch}) and Photon candidates (Iso_γ) are defined separately. It was found that prompt photons are well-isolated for both variables.

The photon isolation variable is also found to be dependent on photon transverse momentum and pileup. To mitigate the effect, we define a corrected photon isolation variable that has been pileup subtracted with p_T dependence as follows,

$$\overline{\text{Iso}}_\gamma = \alpha + \text{Iso}_\gamma - \rho A + \kappa p_T \quad (4.2)$$

where κ is a p_T coefficient, ρ is the average pileup p_T flow density per unit area in the event obtained from the jet area method [108, 109], A is the effective area of the isolation region multiplied by an η -dependent correction factor, and α is a constant measured from experiment chosen to make the $\overline{\text{Iso}}_\gamma$ peak near zero. More details on the pileup removal are found in Ref. [84].

The cuts on these variables are summarized in Table 4.1. In the case of saturated crystals, a situation where the energy deposited by a particle exceeds the crystal's capacity to accurately measure that energy, a different $\sigma_{i\eta i\eta}$ cut is used to recover efficiency. Note that it is not the crystal itself that gets saturated but the electronic readout of the energy reaches a maximum. Since only a certain number of bits encode the readout, energies above maximum are given the same readout value. ECAL readout electronics saturate at around 1.7 (3.0) TeV in the barrel (endcaps). Objects with very high energy electromagnetic showers which deposit energy in a single crystal must take into account this saturation [110, 111]. For our purposes, this impact is negligible. We counted saturated photons passing a preselection of $p_T > 125$ GeV and $M_{\gamma\gamma} > 500$ GeV in the full 2016-2018 dataset. Only one saturated photon was observed in both the barrel region and one in the endcap. This counting allows to avoid simply relying entirely on simulation to validate the assumption that saturated photons are negligible. We did not check any other details of these photons or whether they passed the final selection, to avoid bias during the blinded phase of the analysis.

Finally, we apply the energy scale and smearing corrections recommended by the CMS Electron/Photon Physics Object Working Group [112]. These residual corrections scale the data to the MC and smear the MC to the resolution matched in the data. The photon

energy scale correction is a multiplicative correction applied to the photon energy, which ensures that the energy measured in the detector corresponds to the true energy of the photon. The correction is typically determined using a combination of simulation and data-driven techniques, such as measuring the energy of photons produced in the decay of the Z boson 4.2.4. The correction is typically a function of the photon energy and other variables, such as the position of the photon in the detector. The photon energy resolution, or smearing, correction is a Gaussian smearing applied to the photon energy to account for the detector resolution. The smearing is typically determined from simulation studies and depends on the photon energy and other variables, such as the angle of the photon with respect to the beamline. The energy resolution is often parameterized as a function of the photon energy and the pseudorapidity of the photon.

The Conversion-safe electron veto is designed as a variable that rejects direct electrons and not electrons that might come from photon conversions. This veto requires that a photon cluster in the ECAL is not matched to a reconstructed conversion vertex, and a hit in the inner layer of the pixel detector associated with a charged-particle track indicating a direct and real electron [101].

Table 4.1: Selection details for the high- p_T photon ID v2. Numbers in [] are used if the crystal is saturated.

Photon category	Iso _{Ch} (GeV)	Iso _γ (GeV)	H/E (tower-based)	R_9	$\sigma_{i\eta i\eta}$	Conversion-safe electron veto
EB	<5	<2.75	<0.05	< 0.8	<0.0105 [0.0112]	applied
EE	<5	<2.0	<0.05	< 0.8	<0.0280 [0.0300]	applied

4.2.3 Offline Event Selection

Events are selected which contain two photons passing the above ID requirements and having $p_T > 125$ GeV. This p_T selection is set above the trigger threshold described in Sec. 4.2.1 to ensure uniform efficiency. Each of the diphoton candidate is assigned a primary vertex from which the photon candidate's kinematic properities like invariant mass, $m_{\gamma\gamma}$ are

computed. We consider two diphoton event categories: the first category (denoted EBEB) is where both photons are reconstructed in the fiducial region of the ECAL barrel (EB) with $|\eta| < 1.4442$, while the second category (denoted by EBEE) consists of events with one photon in the EB and the other photon in the fiducial region of the EE, $1.566 < |\eta| < 2.5$. Photons landing in the overlapping region between $1.44 < |\eta| < 1.57$ have low reconstruction efficiency and are excluded in the analysis. Events where both photons are in the EE are also not considered since this category is dominated by SM backgrounds and has negligible sensitivity to the target signal models. The EBEE category on the other hand, contributes an additional 10% sensitivity to the EBEB category. Additionally, photon pairs must angularly separated by $\Delta R > 0.45$ to be consistent with the MC SM diphoton production calculations to be described in the next chapter.

4.2.4 Photon ID Efficiency and Scale Factors

The efficiency of the high- p_T photon ID was measured using the tag-and-probe (TnP) technique [113]. It is a technique used to get a high purity set of physical objects from known resonances such as J/ψ and Z . Due to the similarity between the photon and electron IDs, the $Z \rightarrow e^+e^-$ channel is chosen to compute the photon ID scale factors which are corrections to the efficiency difference between data and MC. The determination of these corrections is a critical ingredient in any physics analysis as it accounts for particles missed by reconstructions and any inefficiencies in the algorithms used.

In the generic tag-and-probe technique, one of the electrons in the $Z \rightarrow e^+e^-$ resonance decay, labeled as tag, is required to pass a tight identification criteria while the probe passes a looser identification. For this analysis, we define the efficiency as the ratio of signal yield of the passing probe and the sum of the signal yields of the passing and failing probe:

$$\text{Efficiency} = \frac{\text{Signal Yield}_{\text{passing probe}}}{\text{Signal Yield}_{\text{passing probe}} + \text{Signal Yield}_{\text{failing probe}}}. \quad (4.3)$$

These signal yields were obtained by fitting the passing and failing probe distributions for

each $\eta - p_T$ bins with the signal + background model. Representative fitting results are shown in Figure 4.4. The scale factors obtained from the ratio between the Data and MC efficiencies for different years are shown in Figure 4.8. They are calculated by dividing the efficiency in data and the MC in $\eta - p_T$ bins as shown in this equation.

$$\text{Scale Factor}(\eta, p_T) = \frac{\text{Efficiency}_{\text{Data}}(\eta, p_T)}{\text{Efficiency}_{\text{MC}}(\eta, p_T)}. \quad (4.4)$$

Above 200 GeV, the uncertainties are larger and the scale factors are extrapolated. The results are shown in Figures. 4.5, 4.6, 4.7. The datasets used for the calculation of the scale factors are shown in Appendix A. The results are consistent within uncertainties. The figures (g,h,i) are systematics comparison plots between a constant 6% uncertainties and the current extrapolated uncertainties.

4.2.5 Datasets

In this analysis we used data collected by the CMS experiment from 2016 to the end of 2018, corresponding to a total integrated luminosity of 138 fb^{-1} . These samples are listed in Table B.1, fulfill the standard data quality criteria for all of the subdetectors of CMS as specified by the good run JSONs listed in Table B.2. These JSONs or JavaScript Object Notation are lightweight file formats which contains the details about specific run periods that are considered to be of high quality and good for analysis.

4.3 Endcap Region Issues

There were two known issues in the endcaps which had to be accounted for in CMS analyses. These issues were related to the L1 trigger prefiring [114, 115] and the HEM15/16 failures. We discuss the problem in more detail in this section and show that in both cases, the effects on this diphoton analysis were negligible.

Figure 4.4: A comparison of the fitting results between nominal background model (top) and two alternative background models which are exponential (middle) and a second-order polynomial background models (bottom).

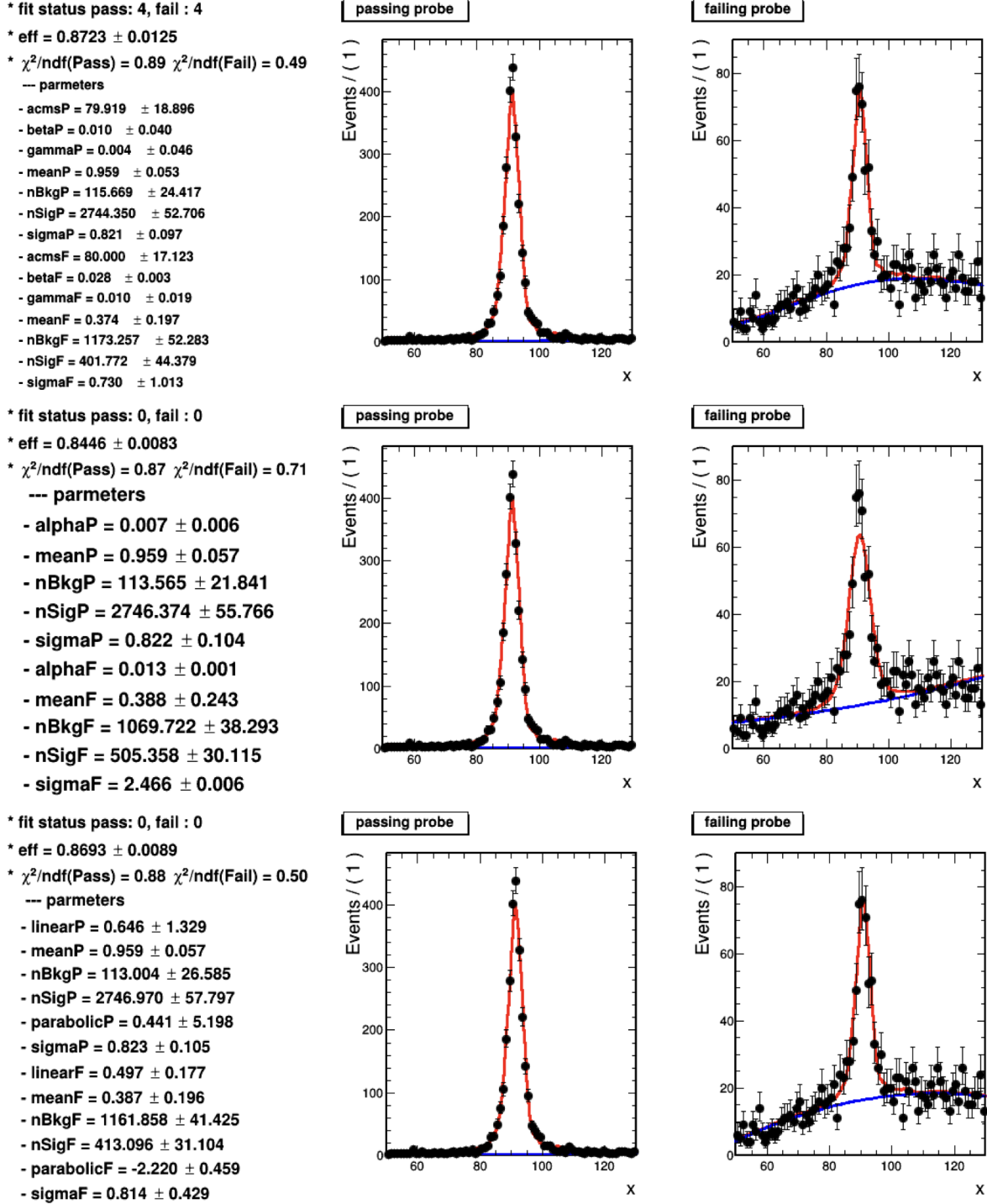


Figure 4.5: Extrapolation plots for 2016. (a,b,c) are the extrapolation fitting results. (d,e,f) are consistency checks. The extrapolated scale factors and uncertainties are drawn on top of the tag-and-probe measured results.

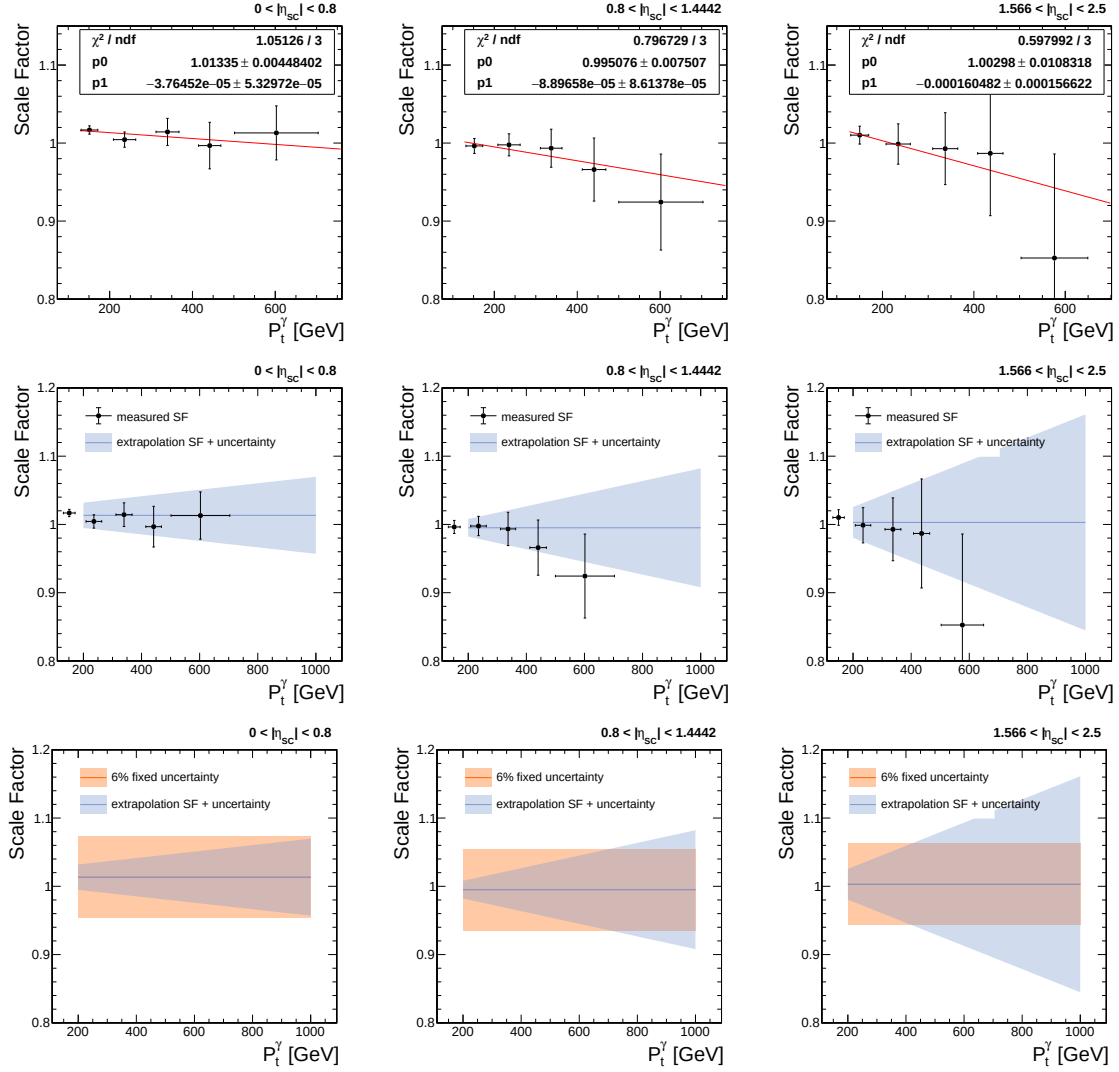


Figure 4.6: Extrapolation plots for 2017. (a,b,c) are the extrapolation fitting results. (d,e,f) are consistency checks. The extrapolated scale factors and uncertainties are drawn on top of the tag-and-probe measured results.

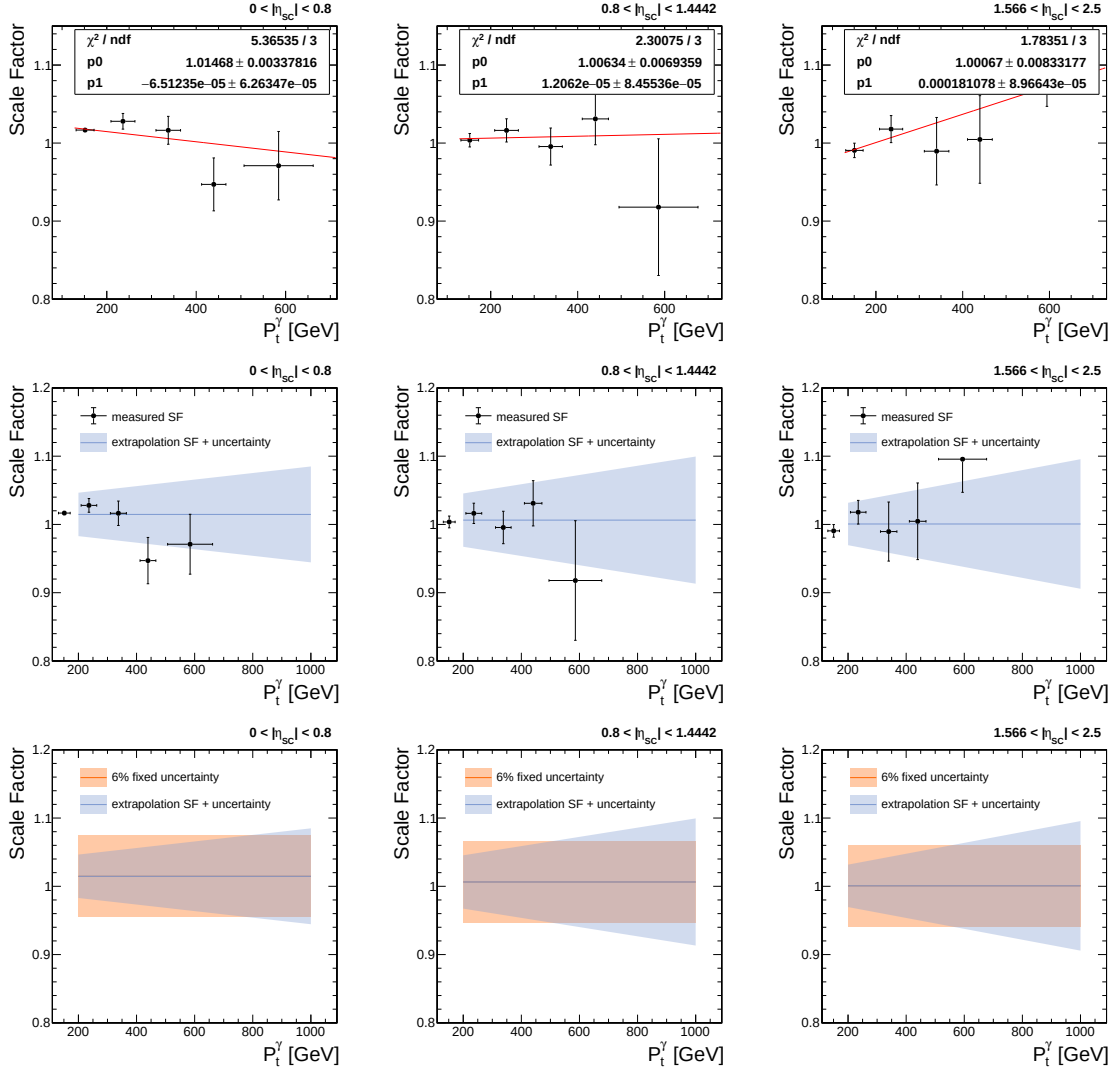


Figure 4.7: Extrapolation plots for 2018. (a,b,c) are the extrapolation fitting results. (d,e,f) are consistency checks. The extrapolated scale factors and uncertainties are drawn on top of the tag-and-probe measured results.

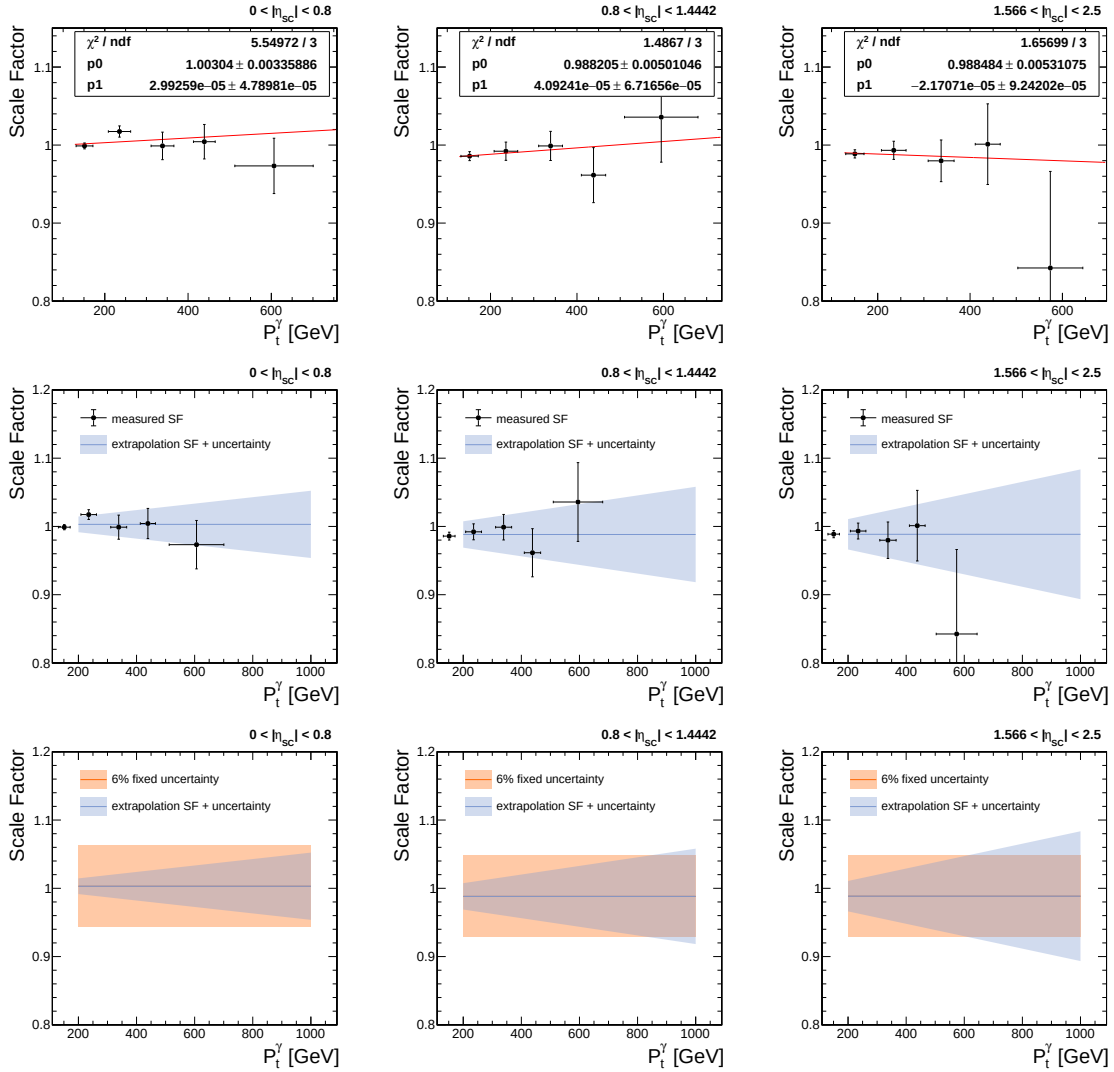
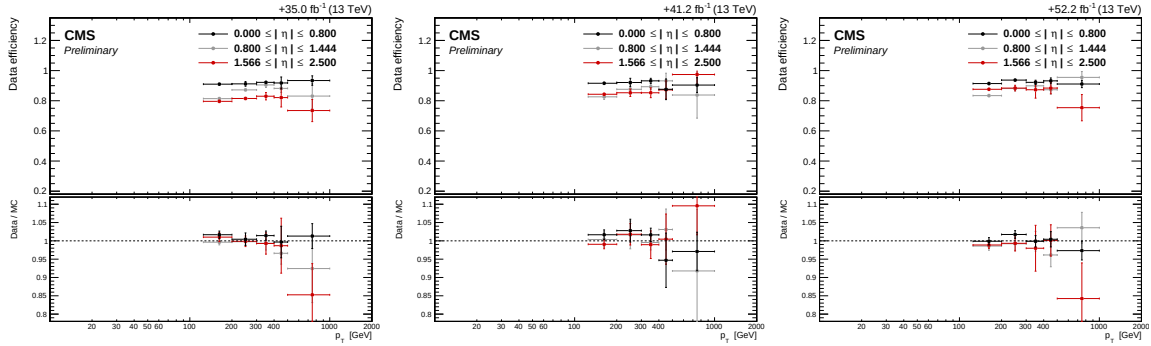


Figure 4.8: Data efficiency and scale factors (bottom) extracted from the tag-and-probe package are plotted vs p_T for three years.



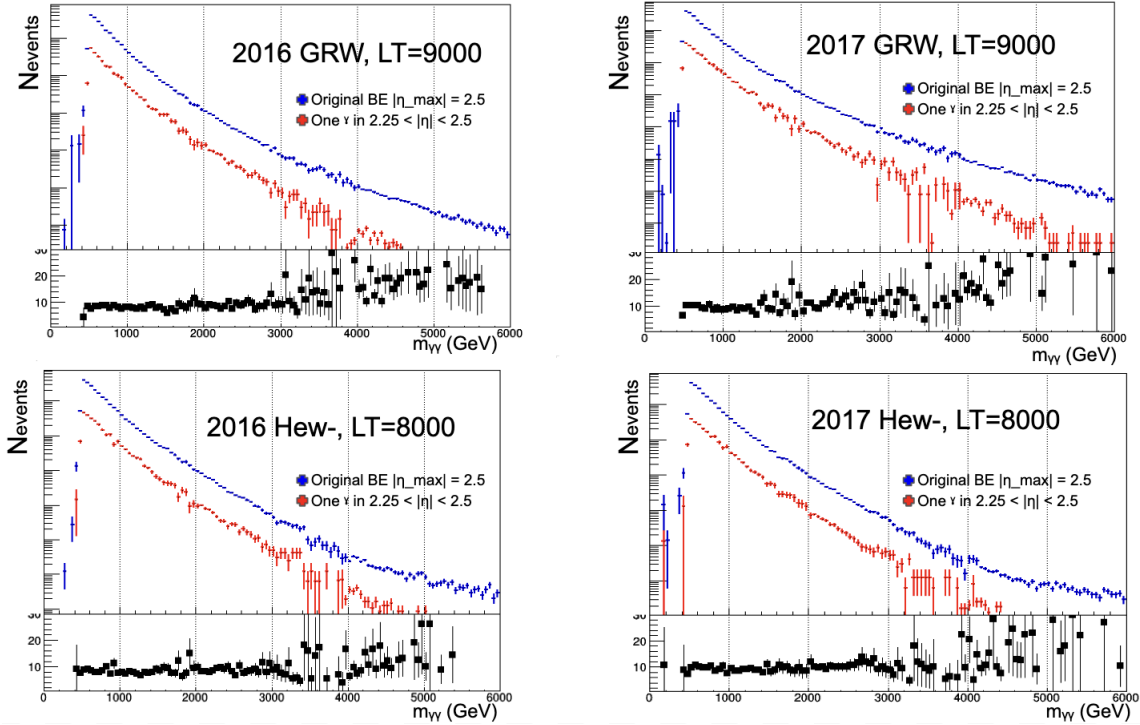
4.3.1 EE L1 Prefiring

From the end of the 2016 data taking and the beginning of Run 2, there was a slowly developing shift in the shape of the ECAL pulses. The effect showed as an increasing offset in the timing calibration of the pulses which resulted from a transparency loss of the ECAL crystals. It is assumed that this loss is radiation-induced as the endcap region is subject to higher fluxes of radiation compared to the ECAL barrel. The problem was realized in early 2018. For this reason, the trigger primitives of the L1 trigger located in the ECAL endcaps were not matched to the correct bunch crossing (BX). Since there is also a rule which prevents the L1 trigger to trigger on two consecutive bunch crossings, an event which should have been accepted by the L1 trigger could have been discarded inappropriately. To mitigate this issue a reweighting procedure was designed in Ref. [115].

The affected region of ECAL was in the pseudorapidity range $\eta > 2.5$. Here the L1 trigger system “prefired” and accepted the earlier collision in BX-1, whereas BX 0 is the one of interest. This effect is not accounted for in Monte Carlo simulations. To mitigate the issue a weight is assigned to each object that can cause the prefiring which are photons and jets. The weight is a probability for the object to cause the prefiring and the total weight assigned to the event is the probability that none of the objects causes a prefiring. We

checked photons that land in the $2.25 < |\eta| < 3.0$ affected region for both signal (Figure 4.9 and background (Figure 4.10) Monte Carlo Events. The pre-firing has a 10% effect on the samples but the overall effect on the analysis is less than 1% as the endcap region only contributes 10% additional sensitivity to the analysis. Since the combined EBEE region only constitutes 10% of our sensitivity, and the issue is only present at the tail-end of 2016 and parts of 2017 data-taking, the overall effect is smaller than 1%. Hence, no reweighting was necessary in accordance to the guidelines stipulated in Ref. [115].

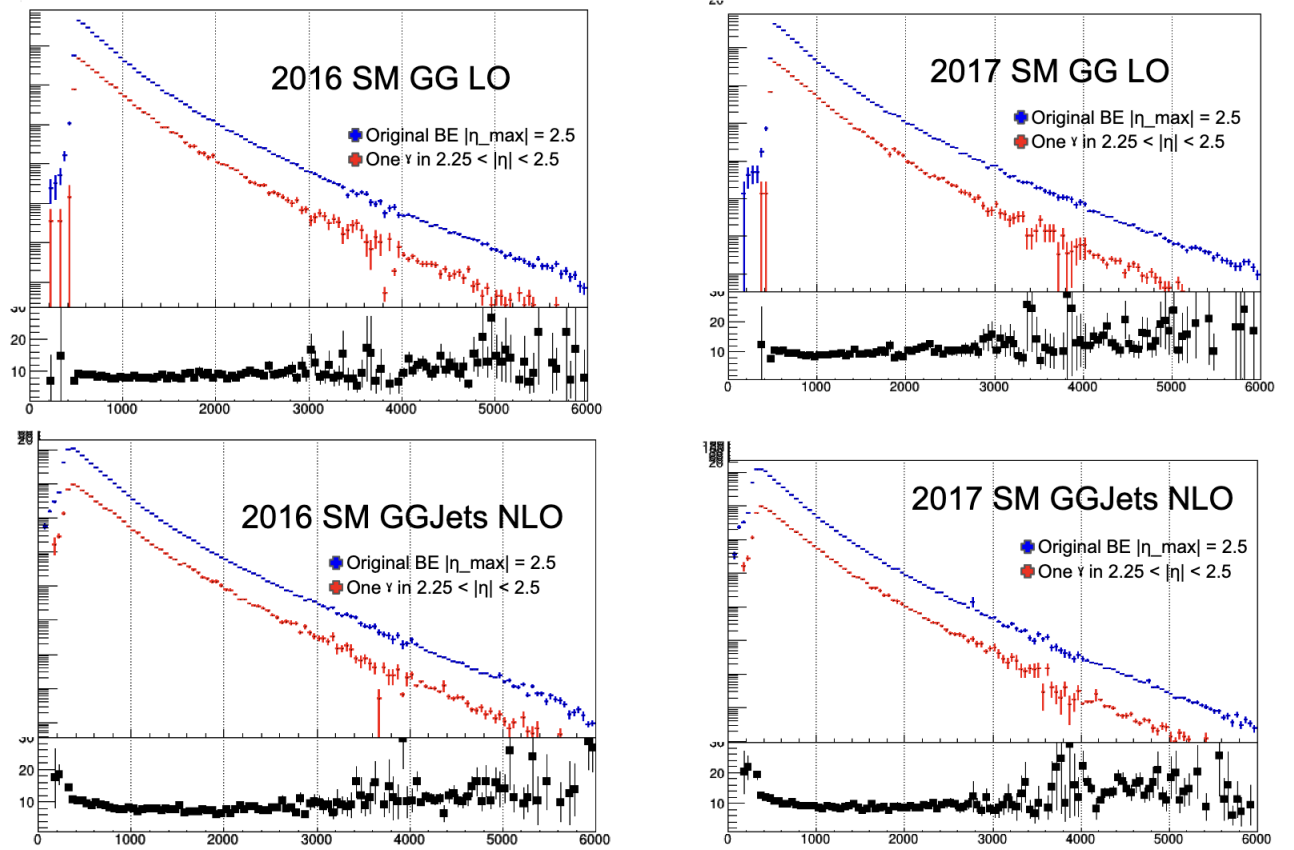
Figure 4.9: Representative signal samples that land in the pseudorapidity range $2.25 < |\eta| < 3.0$ which is affected by the prefiring.



4.3.2 HEM 15/16: Endcap Minus Power Outage

From May 2018 to December 2018, the HCAL endcap minus side experienced power issues in wedges 30-31 known as HEM15 and wedges 32-33 or HEM16 which is shown in the numbering scheme shown in Figure 4.11. This issue then is isolated for run era C of the 2018 data taking. We compare in Figure 4.12 the diphoton mass spectrum in the control region before run A and B and after, run C and D the HEM15/16 issue. From the plots, we see no

Figure 4.10: Representative background samples that land in the pseudorapidity range $2.25 < |\eta| < 3.0$, which is affected by the pre-firing.



significant effects were seen. This is expected as HCAL is mostly seen as a veto for $e - \gamma$ events.

Figure 4.11: Numbering scheme for the HCAL Wedges. Each wedge is 20 degrees.

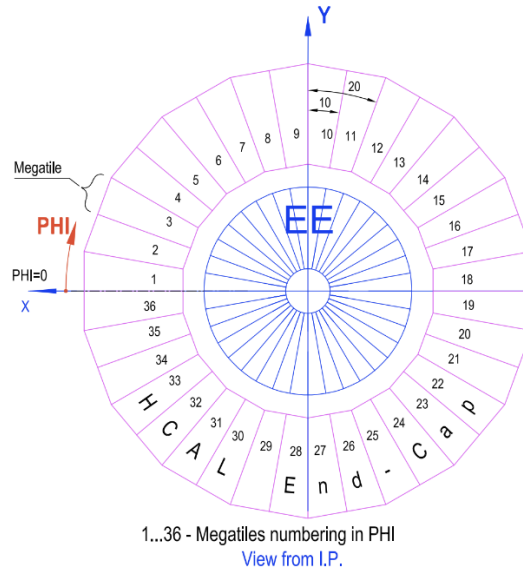
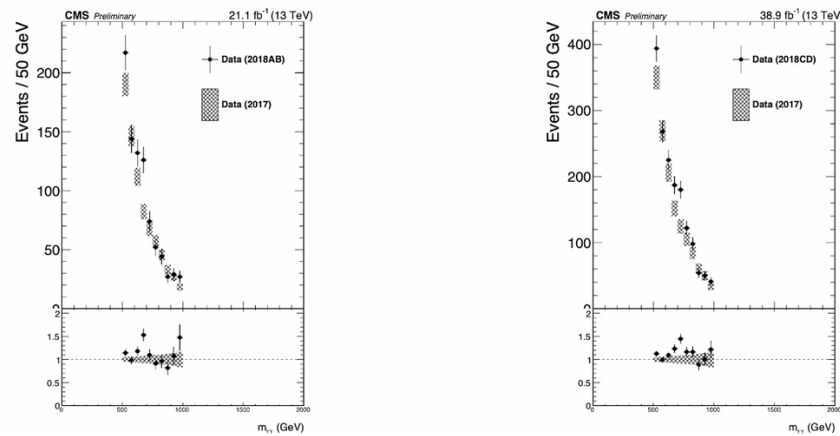


Figure 4.12: No effects were seen on control region of the diphoton mass spectrum before and after the HEM15/16 issue.



CHAPTER 5

BACKGROUND PREDICTION

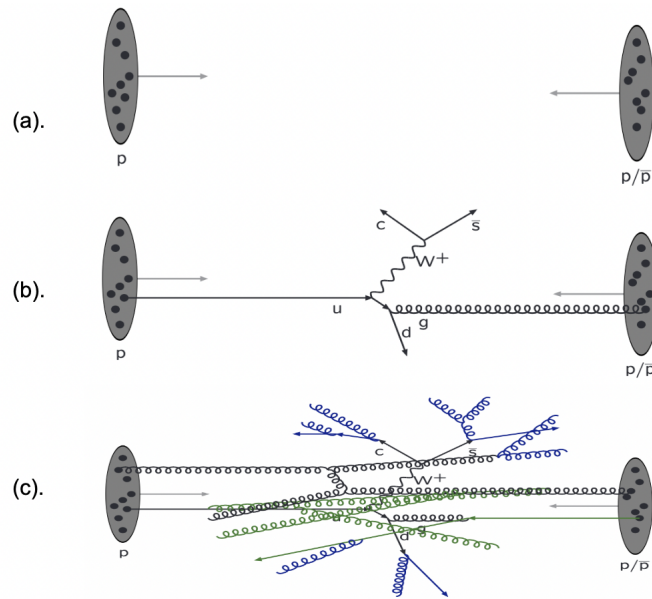
In this Chapter, we give a brief introduction to Monte Carlo Event generation in high-energy physics and the tools that are commonly used. In the succeeding sections, we will then discuss the modelling of the SM Prompt Diphoton Background and the process for including the next-to-next-to-leading-order corrections to the overall prediction.

5.1 Monte Carlo Event Generation

To model and interpret the results coming from proton-proton collisions, Monte Carlo event generators are indispensable tools for high-energy physicists. Their results, together with detector simulation are used to estimate signals and backgrounds in high-energy processes [116]. To model a complicated event or the collision of two protons, which are composed of quarks and gluons, a “divide and conquer” strategy must be employed.

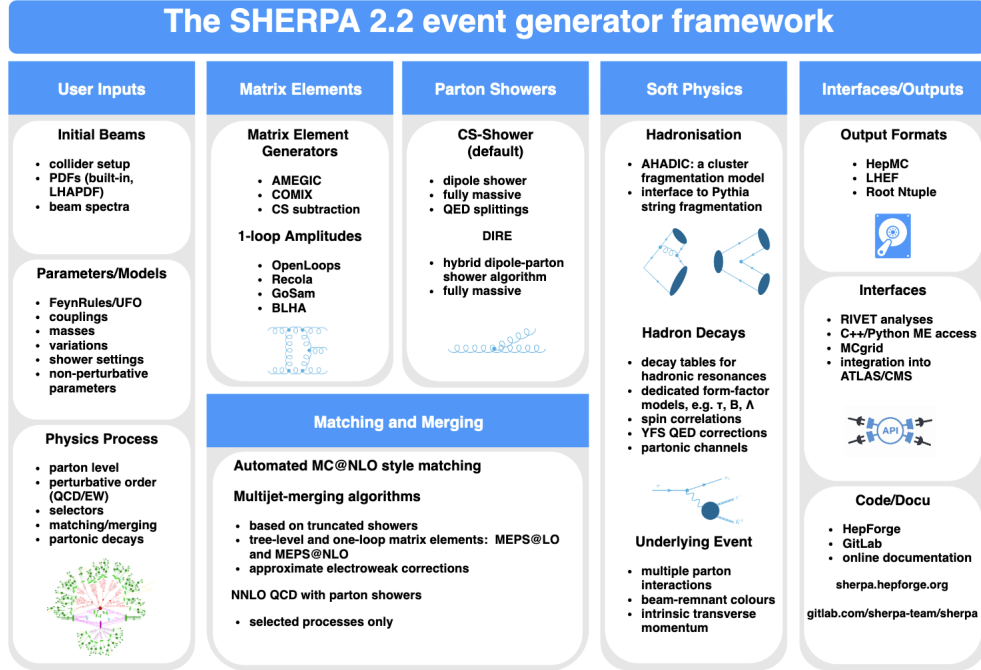
We use two well-known general purpose event generators in this analysis: SHERPA [117] and PYTHIA 8.2 [118, 119]. In both generators, the phase space is split into domains, each having their own technologies. See for example SHERPA’s Event Generator Framework in Figure 5.2. These domains include the hard process interaction, the parton shower and the hadronization part. The hard process, which occurs at $\mathcal{O}(1)$ TeV, is represented by the matrix element (ME). At $\mathcal{O}(1)$ TeV, softer scattering particles emerge from parton-parton interactions, beam remnants and initial and final state radiation. These come from quarks and gluons not participating in the main hard process. They radiate gluons and split into quark-antiquark pairs in a splitting process called fragmentation. Fragmentation

Figure 5.1: Schematic representation of the structure of an event. The first diagram (a) is a representation of the oncoming protons or “bags of partons”. In (b), the main hard subprocess with the resonance decays are shown. The hard subprocess is represented by matrix elements calculations. Finally, the Parton Shower is shown in (c), where full event is shown with other more components such as the initial and final state radiation and beam remnants. Credits: Sjöstrand, Torbjörn



is dominated by the emission of soft and collinear radiation until the $\Lambda_{QCD} \approx 1\text{GeV}$ scale is reached, and via the strong interaction, these fragments snap back into confinement within hadrons. This process is called Hadronization. These domains taken together comprise a full event picture which are represented schematically in Figure 5.1.

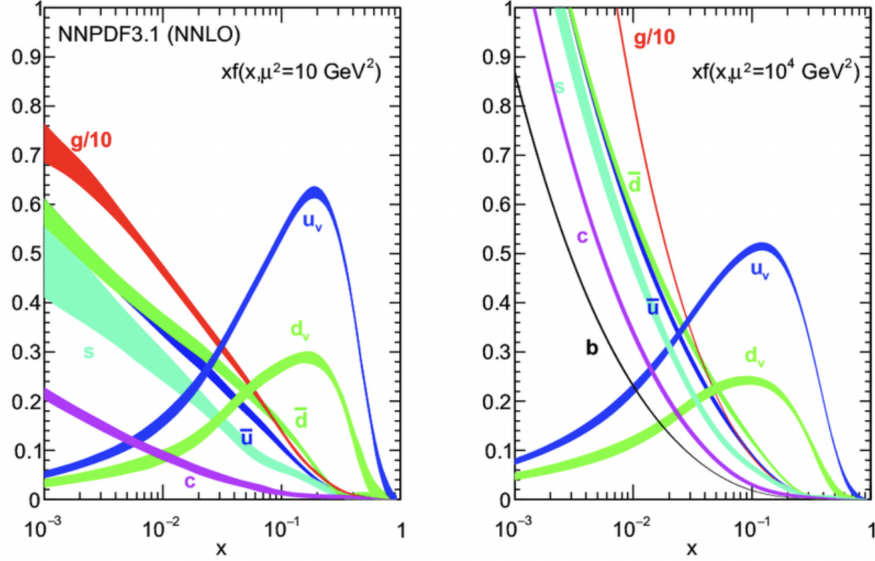
Figure 5.2: This image shows the Sherpa Event Generator Framework with the various elements that are needed to produce simulation, from user inputs, to matrix elements to additional interfaces.



The Monte Carlo generators also makes use of parton distribution functions (PDFs) which have the descriptions of the probability of some struck parton within the proton to carry a fraction x of the proton's momentum p . Both SHERPA and PYTHIA 8 include the NNPDF [120] set of parton distribution functions. NNPDF makes use of neural networks to extract PDFs from data to describe the structure of protons in terms of their quark and gluon constituents. These PDFs cannot be computed from first principles and hence must be extracted from data by carefully matching data with their corresponding theoretical predictions dictated by non-perturbative QCD. The PDFs are also a function of the factorization scale μ_f which roughly separates the long-distance (lower energy) and short-distance (higher

energy) processes. This is typically set at $\mu_f \approx Q^2$, the hard scattering scale where Q^2 represents the scale of the collision momentum transfer. At short-distances, QCD is perturbative but for longer-distances non-perturbative techniques are employed.

Figure 5.3: Latest results of the global fit done by the NNPDF collaboration showing the PDFs at factorization scales of 10 GeV^2 and 10 GeV^4 . The different PDFs are related by the DGLAP evolution equations [121]



The generators allow users to configure the collider setup, the PDFs, other underlying-event parameters [122], as well as the main physics model and their corresponding decays to a final state.

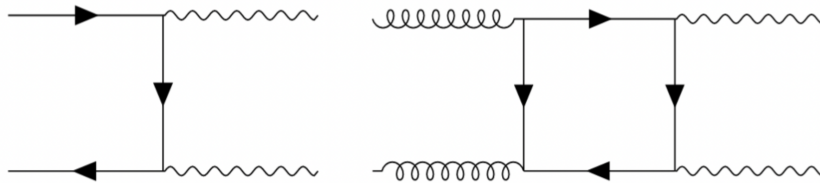
5.1.1 Background Determination

This dissertation searches for non-resonant BSM signatures decaying into the two-photon final state. In this analysis, we account for two main sources of background, 1.) Real background: SM Prompt Diphoton, 2.) Fake Background, which comes from jets that fragment and are misidentified as photons in the detector. The SM diphoton real background is the dominant, irreducible background which is determined up to next-to-next-to-leading (NNLO) accuracy. The subdominant background coming from “fakes” is on the other hand, reducible and is estimated from control samples in data.

5.2 Nonresonant Prompt Diphoton Background

The main source of the background arises from the prompt SM diphoton production, $pp \rightarrow \gamma\gamma$, which occurs via quark annihilation (“Born”) or the gluon fusion process (“Box”). These are illustrated in Figure 5.4. The Born process dominates over the box process due to large quark PDFs. These processes constitute the irreducible background for this search. We use the SHERPA event-generator to prepare diphoton+jets (GGJets) Monte Carlo sample which include the effects of detector simulation.

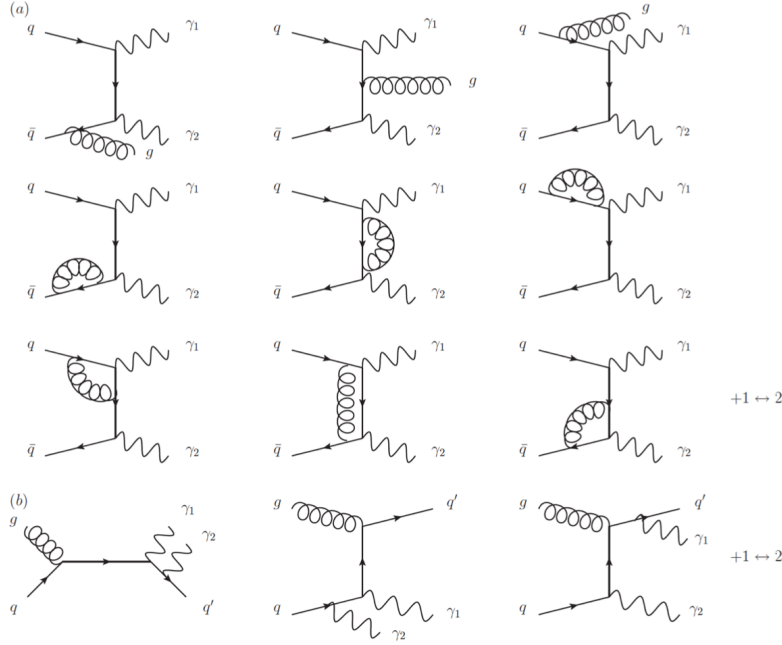
Figure 5.4: The irreducible and primary background for the analysis comes from the prompt standard model $\gamma\gamma$ production which occur via the $q\text{-}q\bar{q}$ annihilation (Born) or through the gluon fusion (Box) process as shown from left to right.



Additionally, up to three additional final state partons, i.e. $(\gamma\gamma + nj)$ are included in SHERPA. Here, j stands for jets and n stands for $n = 0, 1, 2, 3$. While gq -initiated processes are also possible, it is formally a higher-order term and is suppressed compared to the box process due to the same reason. These contributions arise at NLO are not present at LO and scale variations of the quark-antiquark contributions that are present at LO do not provide a reliable estimate of these contributions. Similarly, the gluon-gluon-initiated contributions in the box diagram at NNLO are absent at LO.

It must be noted that SHERPA only includes the real radiation component of these higher order processes. The virtual corrections will later on be accounted in the K-factor corrections calculated from MCFM which will be discussed in more detail in the next section. A distinguishing trait of calculators versus event generators is that, the former’s output cannot

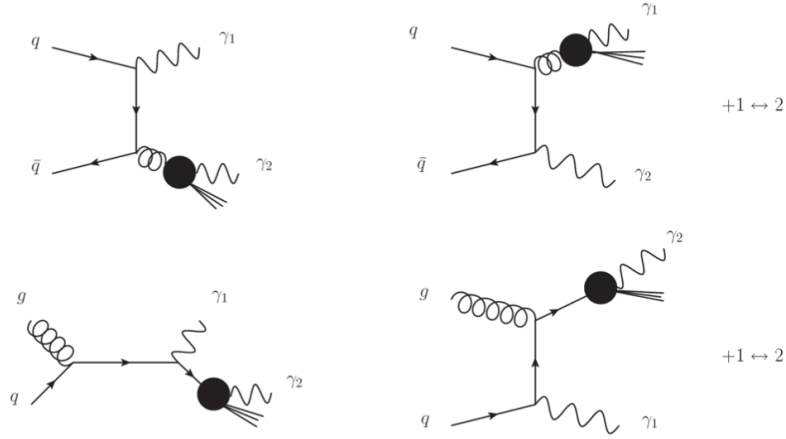
Figure 5.5: Sample Feynman Diagrams for (a) $q\bar{q}$ and (b) gq -initiated processes for NLO SM diphoton production are shown. [123]



be interfaced with a program such as GEANT 4 which accounts for detector effects on the outgoing particles in a collision. However, they can calculate the diphoton cross-section at generator-level in a specified phase space and yield various kinematic distributions.

Higher order corrections have been shown to substantially reduce scale uncertainties estimated from scale variations on total cross section predictions [124]. Additional final state jets are important in sampling the diphoton phase space as shown in Ref. [125]. It was found that the separation in ϕ of the two photons are sensitive to the presence of the additional jets and using an leading-order LO simulation was insufficient for calculating the process. SHERPA applies event weights to the additional final state partons. These weights are accounted for in the analysis. Final state photons could also come from non-perturbative contributions, i.e. fragmentations from a quark or gluon, where soft or collinear photons are emitted from harder jets through a QED $q \rightarrow q\gamma$ splittings. SHERPA interlaces photons into the QCD + QED parton shower. This fragmentation is represented in Figure 5.6

Figure 5.6: Representative Feynman Diagrams for NLO diphoton production through quark or gluon fragmentation [123]



The K factor

As mentioned earlier, our background determination from SHERPA which has up to 3 final state jets in the simulated events includes most of the QCD real emission component of the higher order corrections to the basic diphoton process. However, virtual corrections are not included. To attempt to remedy this, we use MCFM v8.0 [126] which is capable of performing a full next-to-next-to-leading order (NNLO) calculation. The MCFM diphoton calculation is based on the Q_T -subtraction procedure using N-jettiness as a phase-space-slicing variable [127] and it also includes virtual processes. MCFM and SHERPA outputs generator-level distributions which are distributions that are not interfaced with a detector-level simulator like Geant4. An attempt to provide a full NNLO diphoton background prediction is done by calculating the K factor which is the ratio of the MCFM and SHERPA outputs. Reweighting the SHERPA out-of-the-box background prediction with the K-factor accounts for the missing higher order terms and virtual corrections. The K-factor is specifically calculated as a function of the diphoton invariant mass from the ratio of the MCFM NNLO prediction to the SHERPA MC. The K-factor-reweighted SHERPA MC events that passed through detector simulation becomes the final prediction for the real SM diphoton invariant mass spectrum.

This is shown in Figure 5.7.

To avoid bias, K-factor is calculated separately for the diphoton barrel-barrel (BB) and barrel-endcap (BE) acceptance categories. The NNLO cross-section from MCFM is calculated separately in four bins of $m_{\gamma\gamma}$:

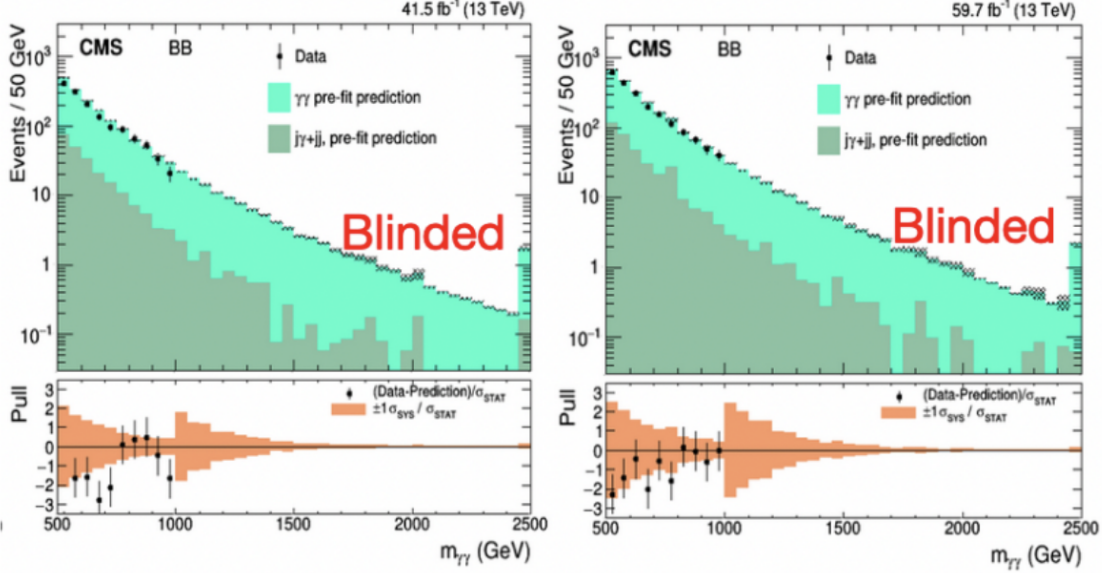
- $500 < m_{\gamma\gamma} < 750 \text{ GeV}$
- $750 < m_{\gamma\gamma} < 1000 \text{ GeV}$
- $1000 < m_{\gamma\gamma} < 1500 \text{ GeV}$
- $1500 < m_{\gamma\gamma} < 4000 \text{ GeV}$

If the invariant mass range is too large, larger invariant mass bins within the range wouldn't be adequately sampled. This is because the stopping criteria for the estimated accuracy of the numerical integration will be satisfied early. These separate bins, weighted by their cross-sections are ultimately stitched together to form a continuous and smooth distribution.

Figure 5.8, shows the K factor distributions for both barrel-barrel and barrel-endcap diphotons. The effects of simultaneously varying the renormalization μ_r and factorization μ_f scales by a factor of two in each direction, up and down around the nominal prediction which is the diphoton invariant mass. The parameters, μ_r and μ_f are introduced in QCD calculations to remedy the UV and infrared divergences in the calculation of the amplitude squared of any given partonic process. The UV divergences come from large momentum in the loops of Feynman diagrams while infrared divergences happen due to virtual or real particle that reach zero momentum. The coupling constant α_s becomes a function of the renormalization scale μ_r . The PDF, introduced earlier in the chapter, and fragmentation functions will become a function of the factorization scale. The choice of varying the scale factors as done here is often sufficient to know how observables vary within those scales.

Figure 5.9 shows a summary of the ratios of the diphoton cross section from MCFM to that from SHERPA. Different orders of the MCFM calculation are shown and the dashed lines

Figure 5.7: Shown here are the out-of-the-box SHERPA prompt simulated $m_{\gamma\gamma}$ prediction reweighted with the K-factor in light teal. The fake prediction is shown with a darker shade of teal.



show the variation of the renormalization and factorization scales as shown in Figure 5.8. The nominal ratios are shown in full lines while the dashed lines show the effect of the simultaneous renormalization and factorization variations. This effect is treated as a systematic uncertainty and will be discussed more in Ch. 7.

In MCFM, the fragmentation scale is set to the renormalization scale by default. The fragmentation scale is the scale of the fragmentation functions that describe the probability for a parton to fragment into a hadron. This contribution is negligible when using the Frixione isolation which is discussed more in [128]. The fragmentation scale is a parameter that describes the energy scale at which a high-energy quark or gluon (parton) fragments into a jet of hadrons, which are composite particles made up of quarks and gluons. Note that the fragmentation scale and the renormalization scale are both important concepts in theoretical particle physics, but they refer to different aspects of the theory. The fragmentation process is a non-perturbative phenomenon, which means it cannot be calculated using standard perturbative techniques. Instead, it is described using phenomenological models that are

Figure 5.8: NNLO k-factor for the $m_{\gamma\gamma}$ distribution for barrel-barrel diphotons (left) and barrel-endcap diphotons (right) for the real diphoton k-factor. The renormalization μ_r and factorization μ_f scales have been varied as follows: $\mu_r = \mu_f = 0.5m_{\gamma\gamma}$ (top), $1.0m_{\gamma\gamma}$ (middle), and $2.0m_{\gamma\gamma}$ (bottom). The cuts on the generator-level photons are the same as the kinematic cuts used offline.

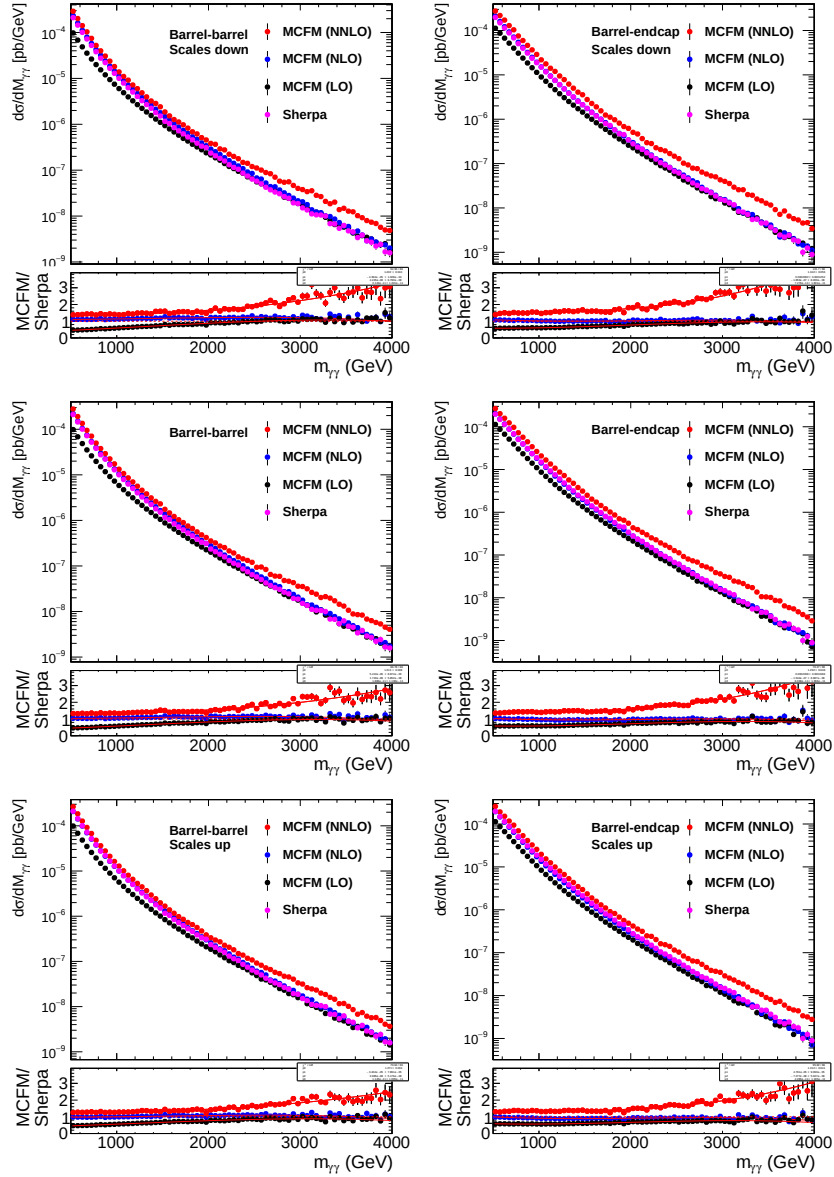
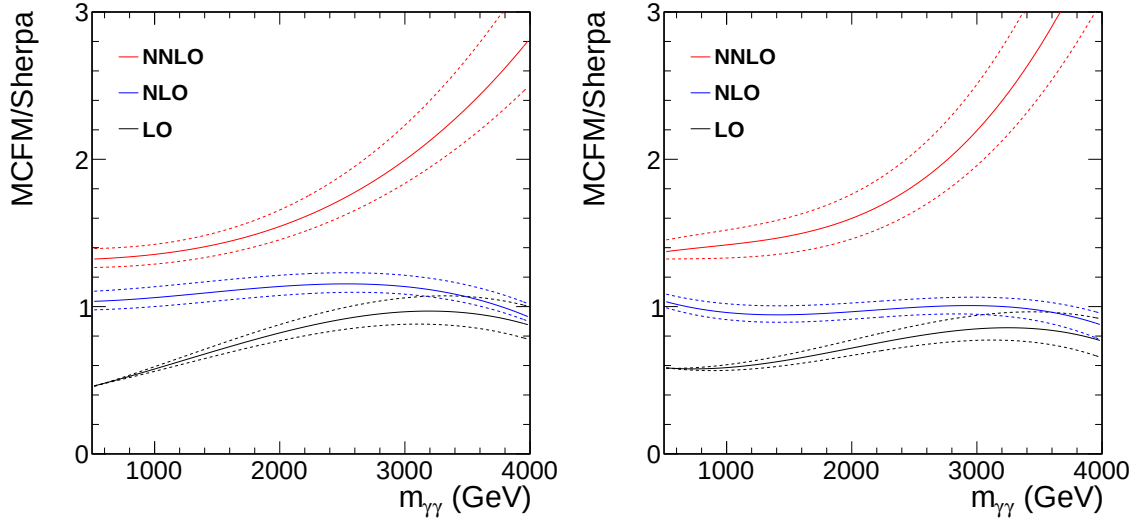


Figure 5.9: Ratio of diphoton cross sections calculated by MCFM versus the results from SHERPA in the barrel-barrel (left) and barrel-endcap (right) regions.



tuned to experimental data. The fragmentation scale is typically on the order of a few GeV to tens of GeV, depending on the process being studied. The renormalization scale, on the other hand, is a parameter that describes the energy scale at which the coupling constants in the theory are defined. In quantum field theory, the coupling constants are typically dependent on the energy scale at which they are probed. This is due to the effects of virtual particles that are created and annihilated in the vacuum. To account for these effects, the coupling constants are "renormalized", which means they are adjusted at different energy scales to remove divergences in the theory. The renormalization scale is typically chosen to be on the order of the energy scale at which the process of interest is taking place. In summary, the fragmentation scale and the renormalization scale are two different energy scales that are used in different contexts in theoretical particle physics. The fragmentation scale is used to describe the energy scale at which partons fragment into hadrons, while the renormalization scale is used to describe the energy scale at which the coupling constants in the theory are defined.

The NNPDF30_NNLO_as_0118 set of parton distribution functions (PDFs) are used [129].

SHERPA includes the contribution of prompt photons from gluon and quark partons fragmenting to photons, which avoids the necessity of imposing an arbitrary ΔR selection between photons and jets to avoid infrared divergences [130]. Infrared divergence is a phenomenon that arises in quantum field theory when the energies of the particles involved in a scattering process become very low. In particular, infrared divergence occurs when the energy of a particle becomes much smaller than the mass of the particle. The photons are required to satisfy $p_T > 50 \text{ GeV}$ at generator level.

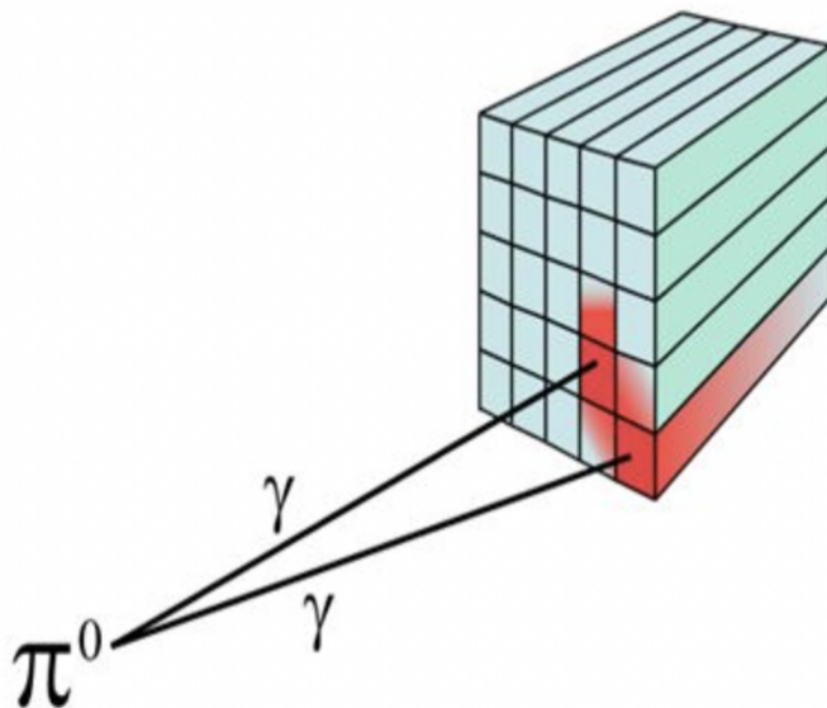
As described later in Section 5.2, these simulated diphoton+jets events will be reweighted by the NNLO k-factor to provide our prediction to the irreducible diphoton background. The photon+jet samples listed in Table B.3 are used only for studies such as the MC closure test, cross-checking of fake predictions and deriving templates for real photons (helped by large statistics), but are not part of the final background prediction, which is data-driven for fakes. We list them here only for completeness. Also listed in Table B.3 are a separate set of leading-order-only MC background samples which were also generated with PYTHIA 8.2 [119] with the NNPDF3.1 LO set of parton distribution functions as found in the event tune, **TuneCP2** [122]. This LO diphoton set is used only for subtracting the background from the signal samples that are generated with the same generator and PDF. This takes into account the effect of interference between signal and background.

While minor backgrounds arise from other SM processes, we have not included them in this analysis as their effects are negligible. The processing configuration (Campaign and Global Tag, by year) for all MC samples in this analysis is listed in Table B.4.

5.3 Background from Fake Photons

Another source of background in the diphoton signal region comes from 'fake photons'. This term is assigned to events where one or two jets are misidentified as 'real' photons by virtue of passing the high p_T ID. This occurs when jets fragments in such a way that when they decay to two photons; and at sufficiently high momentum they overlap in the detector

Figure 5.10: Schematic of a π^0 decaying into two photons that deposit energy that mimics energy-deposit features of prompt photons.



and resemble a single photon signature. Such jet fragmentation typically happens with π^0 or another neutral meson decaying into two photons 5.10. A neutral meson that has little hadronic activity around it can also be reconstructed that can pass the photon isolation criteria and be mislabeled as a photon.

The method used to estimate the 'fake' contribution assumes that dijet or $\gamma + j$ events where these jets fragmented in such a way to produce a fake photon signature could be modeled by other such events where the jet fragmented differently. Here we define two distinct categories of jet fragmentation which we measure. The ratio of these two categories provide us the 'fake rate'.

The first category, the numerator, are jets that pass the high- p_T photon ID requirements specified in Table 4.1. The denominator category, consist of 'photon-like' jets which pass an less strict version of the photon ID requirements, and are additionally required to fail at least one of the other photon ID requirements. The latter requirement ensures, that the numerator and denominator are orthogonal, i.e. there is no contamination from real photons in the denominator. The complete set of denominator selection criteria is given in Table 5.1.

Table 5.1: Fake rate definitions, where fake rate = numerator / denominator. The real and fake templates are fit to the numerator candidates using the template variable labeled 'template'. After the fitting procedure is performed, the cut associated with the template variable is applied. The photon ID is shown for comparison.

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
$\text{Hadronic}/E < 0.1$	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

What we seek to measure, termed the *fake rate*, is defined as

$$f = \frac{\text{number of jets passing high } p_T \text{ ID}}{\text{number of jets passing denominator definition}} \quad (5.1)$$

Note that this “fake rate” is just a relative ratio between two distinct categories of jet fragmentation, not the rate at which a generic jet fakes a photon, which is actually much smaller. The “fake rate” is measured in a reference data sample as a function of the photon p_T , η and the number of primary vertices. Events containing a denominator jet are taken to be representative of other possible events where the jets could have fragmented in a way that they pass the photon ID. To obtain the “fake contribution”, we take the data events containing objects that pass the pass the looser “photon-like jet” denominator definition and reweight them by the fake rate.

5.3.1 Subtracting Real Photon Contamination in the Fake Rate

In measuring the fake rate in Eq. 5.1, a complication arises in the determination of the numerator. The set of objects selected from data which pass the photon ID criteria will include both jets and real photons. This real photon contribution must be subtracted since our definition of the fake rate are from jets-faking-photons counts. The general procedure for subtracting this real photon contamination in the numerator is as follows. We consider a photon ID variable (such as shower-shape, σ_{imin} which is sensitive to the difference between real and fake photons. We relax the cut on this variable, specified in Table 5.1 as **template** and plot the distribution of this variable for all objects which otherwise pass the rest of the numerator ID. Shower-shape templates are constructed for both real and fake photons and then these templates are fit to the observed data. The fit allows us to determine the fraction of photons which are real/fake after applying the ID cut on the fitting variable. We then calculate the number we should use for our numerator in the fake rate. Because the variable distributions typically vary with p_T , this fitting and subtractions is done per p_T bin.

The following sections describe this procedure in more detail. Closure tests that validate this procedure are described further in Section 5.4. In the closure test, we found that

splitting the barrel and endcap regions into inner and outer regions further improved the fake prediction. This is an improvement over the previous analysis in Ref. [131].

5.3.2 Fake Rate Numerator: Templates and Fitting

The fake rate in our analysis is defined as a ratio of strictly jets-which-fragmented-to-pass-photon-ID to jets-which-fragmented-to-pass-a-looser-ID. When we run through the data to measure the numerator, i.e. jets-which-fragmented-to-pass-photon-ID, we inevitably select real photons. To subtract this real photon contamination, we use a template fit to a variable that is sensitive to the difference between real and fake photons. Here we use the $\sigma_{i\eta i\eta}$ variable to form real and fake template distributions.

We construct the real photon $\sigma_{i\eta i\eta}$ distributions from MC **GGjets** samples that are listed in Section 5.3.6). The reconstructed photon object is required to match in ΔR to a GenParticle photon before being used to form the template.

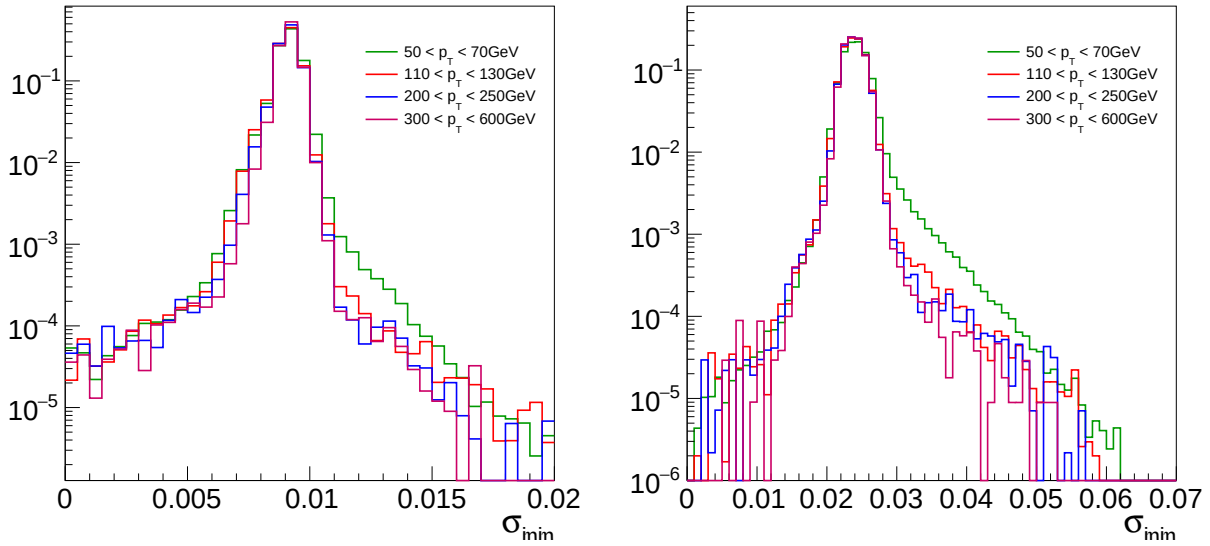
We derive the fake template from the **JetHT** dataset which provides a large supply of ‘fake photons’ from jets. These constructed fake templates are then used to fit separately to the **JetHT** or **DoubleMuon** datasets. Table B.5 and B.6 show the **JetHT** and **DoubleMuon** data sets, respectively.

We select a sideband sample of objects that pass most of the Photon ID, but are required to have a charged hadron isolation value in the sideband of $5 < \text{Iso}_{\text{Ch}} < 10 \text{ GeV}$. The idea behind the sideband is to select objects which still has similar properties to the “fake” objects that pass the Photon ID. By requiring the objects to fail Iso_{Ch} cut, makes the sample fake-enriched. The Iso_{Ch} variable is chosen to define the sideband sample because it is the ID variable which is *least* correlated with the $\sigma_{i\eta i\eta}$ template variable. There is still some unavoidable correlation between these variables. This is because of the physics of jet fragmentation which tends to produce high-EM content (via decays of neutral mesons) also tends to produce charged hadrons which contribute to the isolation variable. Using a different sideband definition will unavoidable sculpt the extracted $\sigma_{i\eta i\eta}$ template to some degree. To minimize the bias, we choose a sideband which is as close as possible to the cut

value for passing the real photon ID. We then explore the variation in template shape that is obtained by using different sideband definitions. This is a source of systematic uncertainty on the fake rate method which will be discussed later in Ch. 7.

Figure 5.11 and Figure 5.12 show the p_T dependence or variation among the real and fake templates, respectively in both the EB and EE. The template fit to the **JetHT** or **DoubleMuon** datasets are then performed separately in each p_T bin.

Figure 5.11: Representative real templates derived from MC in EB and EE for different p_T bins.



There is very little dependence of the template shape on the year. This is shown for the real template in Figure 5.13 for photons in each of the three calorimeter regions. Figure 5.14 shows the same for the fake templates, separated by the data set in which they are derived.

An illustration of the template fit for a given p_T bin is shown in Figure 5.15. Real photons are primarily found near the peak for the $\sigma_{in\eta}$ template variable, while the fake photons generally populate the broad tail at large $\sigma_{in\eta}$ values. This indicates a broader shower-shape in ECAL, arising from the presence of more than one photon in the shower from neutral meson decays. The overall scale factors for the real and fake templates serve as the only two free parameters for the fit. After fitting, the post-fit fake template is integrated

Figure 5.12: Fake templates comparison measured in various p_T bins extracted from the JetHT data set for EB (left) and EE (right).

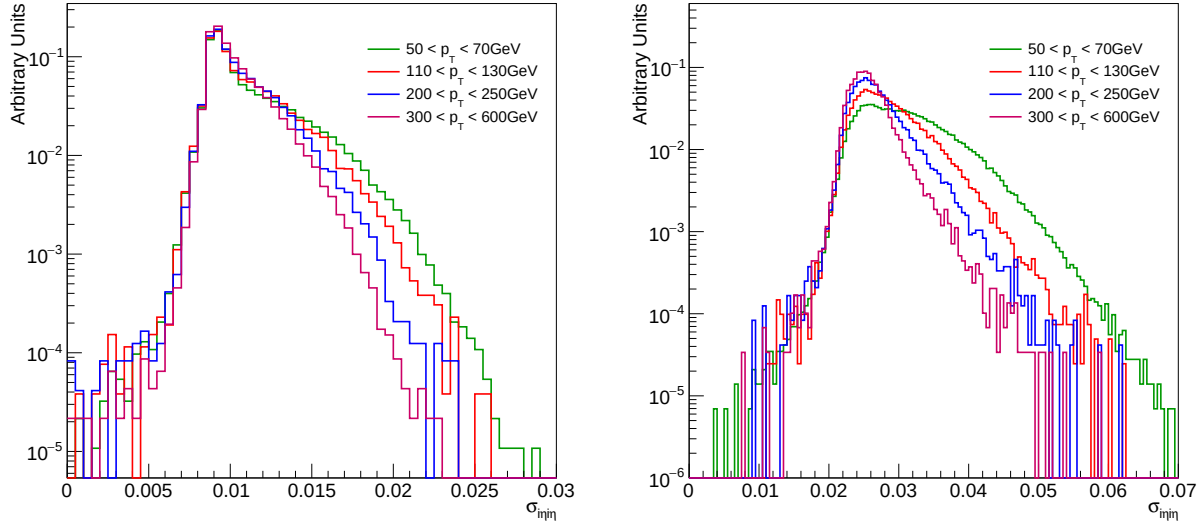


Figure 5.13: Comparison of normalized real template shape among the three analysis years, for EB (left), EE1 ($1.566 < |\eta| < 2.033$)(center), and EE2 ($2.033 < |\eta| < 2.5$)(right).

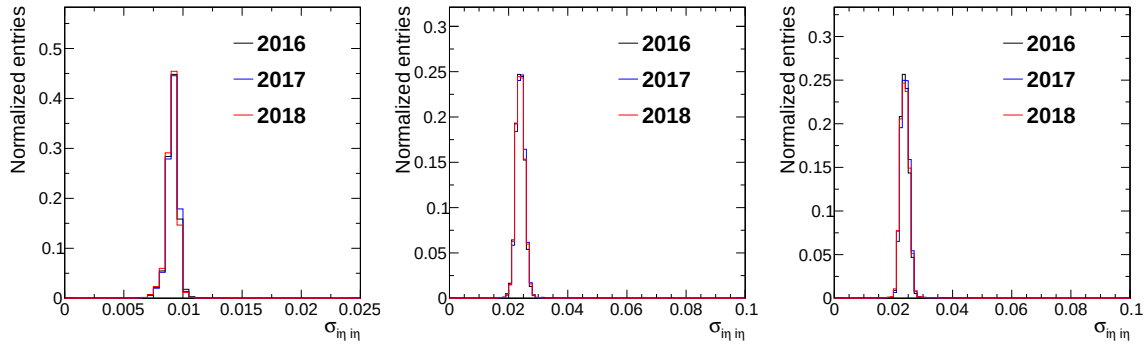
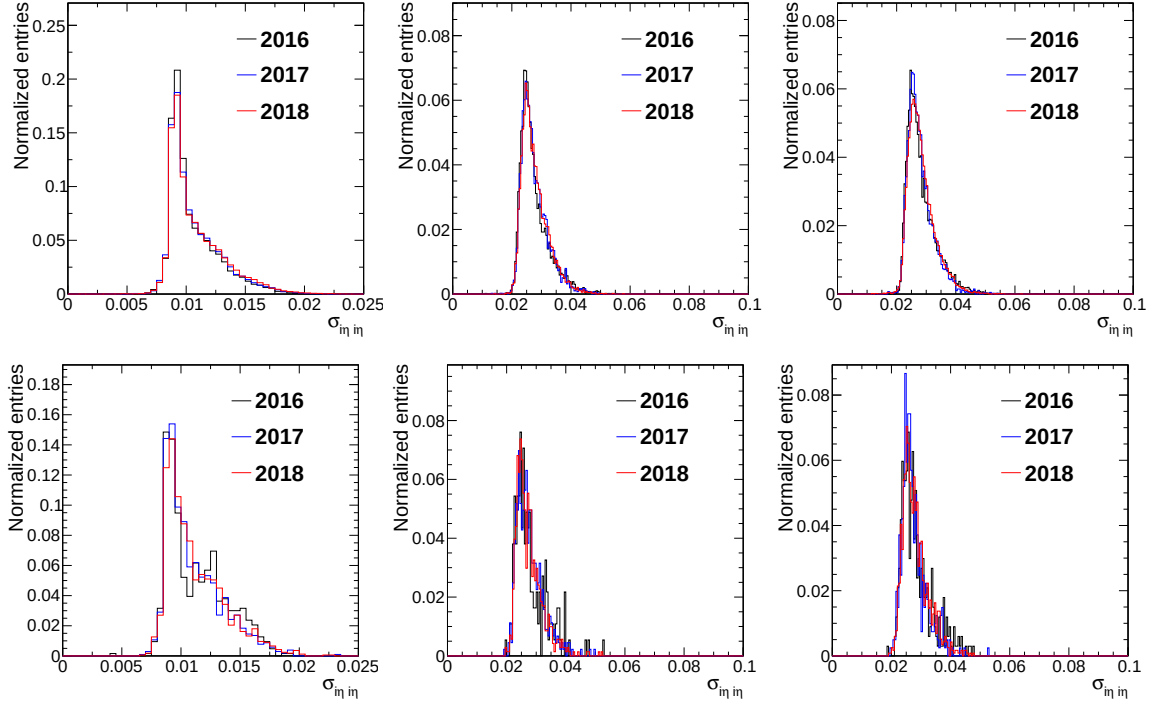
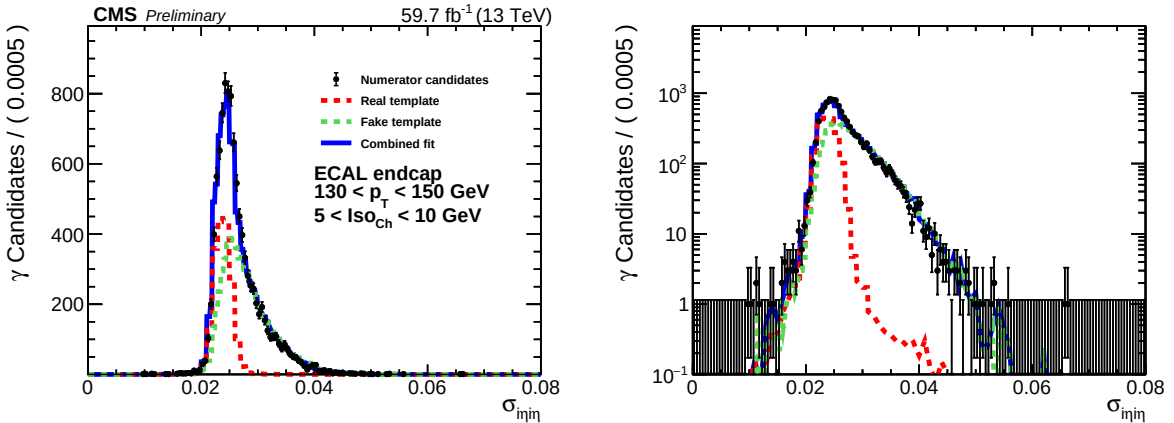


Figure 5.14: Comparison of normalized fake template shape among the three analysis years, for EB (left), EE1 ($1.566 < |\eta| < 2.033$)(center), and EE2 ($2.033 < |\eta| < 2.5$)(right). The top (bottom) row shows the templates derived from the JetHT (DoubleMuon) dataset.



from $\sigma_{i\eta i\eta} = 0$ up to the appropriate cut value from Table 4.1 to give the estimated number of fake-photons-from-jets in the sample. This number serves as the value for the *numerator* in our fake rate for the particular p_T bin.

Figure 5.15: A representative template fit extracted from the 2018 JetHT data set using photon objects from the ECAL barrel in the $130 < p_T < 150$ GeV range, in linear (left) and log scale (right) for the y-axis.



5.3.3 Fake Rate Denominator

We define the denominator of the fake rate as a count of the number of objects passing our denominator definition in each p_T bin. While this definition is arbitrary, the denominator objects obtained from the `JetHT` sample and the `DoubleEG` samples must be similar with this choice of definition which is defined as follows:

1. Denominator object must *FAIL* one of the high- p_T ID cuts
2. Must *PASS* the conversion-safe electron veto (CSEV)
3. Both Iso_{Ch} and $\overline{\text{Iso}}_{\gamma}$ must be < 0.2 times the object's p_T
4. Cone-based H/E < 0.1
5. $R_9 > 0.8$ to match the photon preselection cut.

labeldenom_definition

The first criteria requires that the denominator object fails at least one of the Photon ID cuts. This ensures that the denominator sample is orthogonal to the numerator, and is not contaminated by real photons. Additionally, the object must pass the conversion-safe electron veto variable which is a variable designed to reject direct electrons but not electrons that originate from photon conversions. To ensure that the denominator objects are still photon-like-jets, that is they are still reasonably isolated, we require that the isolation variables are not too large ($< 20\%$ maximum). This is important since we are reweighting the denominator by the fake rate to estimate our fake prediction. The cone-based H/E cut is used in the photon ID at HLT level while the tower-based Had/EM is the more typical offline variable used in our photon ID. Since we are selecting from data, we use the cone-based H/E cut.

5.3.4 Extracting the Fake Rate

The fake rate is just the ratio of the numerator and denominator which we have counted using the procedure described above. The fake rate is primarily a function of p_T and is

determined for EB and EE separately. The EE is further split into two bins, which we show later in Sec. 5.4 to be an improvement in our fake prediction for the η distributions. The fake rates are also dependent on the number of selected primary vertices. This is shown in Figure 5.16, Figure 5.17, and Figure 5.18 for EB, EE1 and EE2, respectively. EE1 covers the region $1.566 < |\eta| < 2.033$ while EE2 covers $2.033 < |\eta| < 2.5$. We take the average of the fake rates derived from the `JetHT` and `DoubleMuon` data sets. As a consequence of the lower statistics in the `DoubleMuon` dataset, the fake rate derivation have a different p_T binning. We apply the fake rate separately for each of the two data sets at the given p_T , then the two fake rates are averaged. We chose ranges for comparison that have a roughly uniform level of statistics between them.

Figure 5.16: A comparison of the fake rate in EB measured as a function of the number of selected primary vertices (nPV) in the event using the `JetHT` data set (top) and `DoubleMuon` data set (bottom). The fake rates are determined separately for 2016 at the (left), 2017 (center), and 2018 (right).

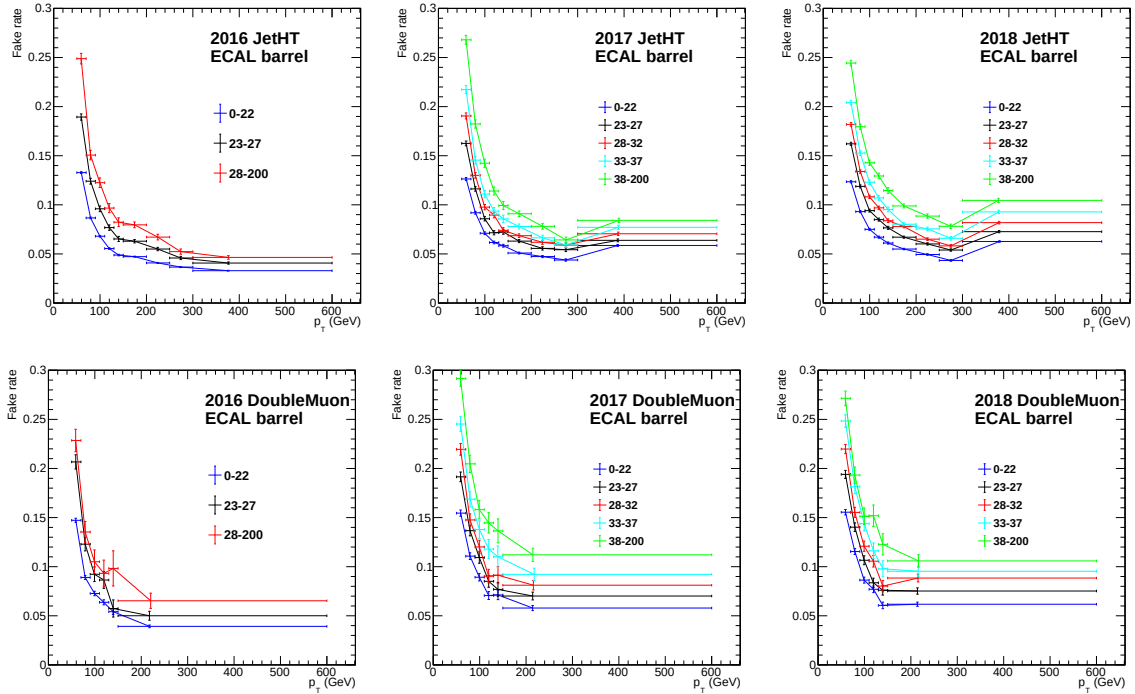


Figure 5.17: A comparison of the fake rate in EE1 ($1.566 < |\eta| < 2.033$) measured as a function of the number of selected primary vertices in the event using the **JetHT** data set (top) and **DoubleMuon** data set (bottom). The fake rates are determined separately for 2016 (left), 2017 (center), and 2018 (right).

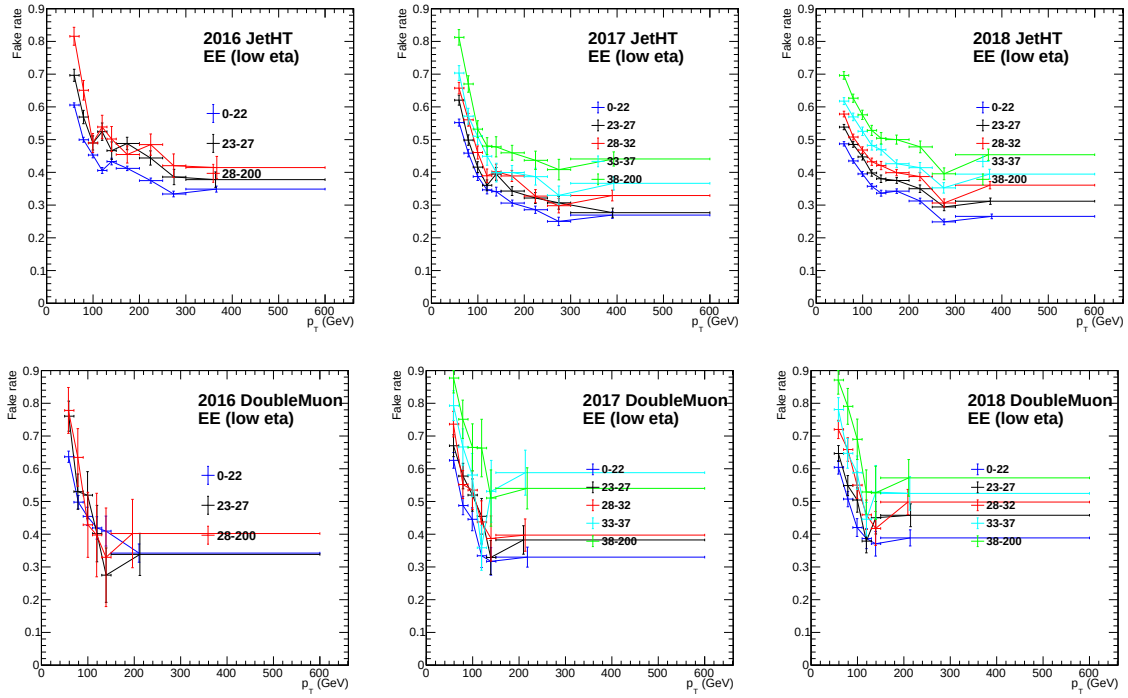
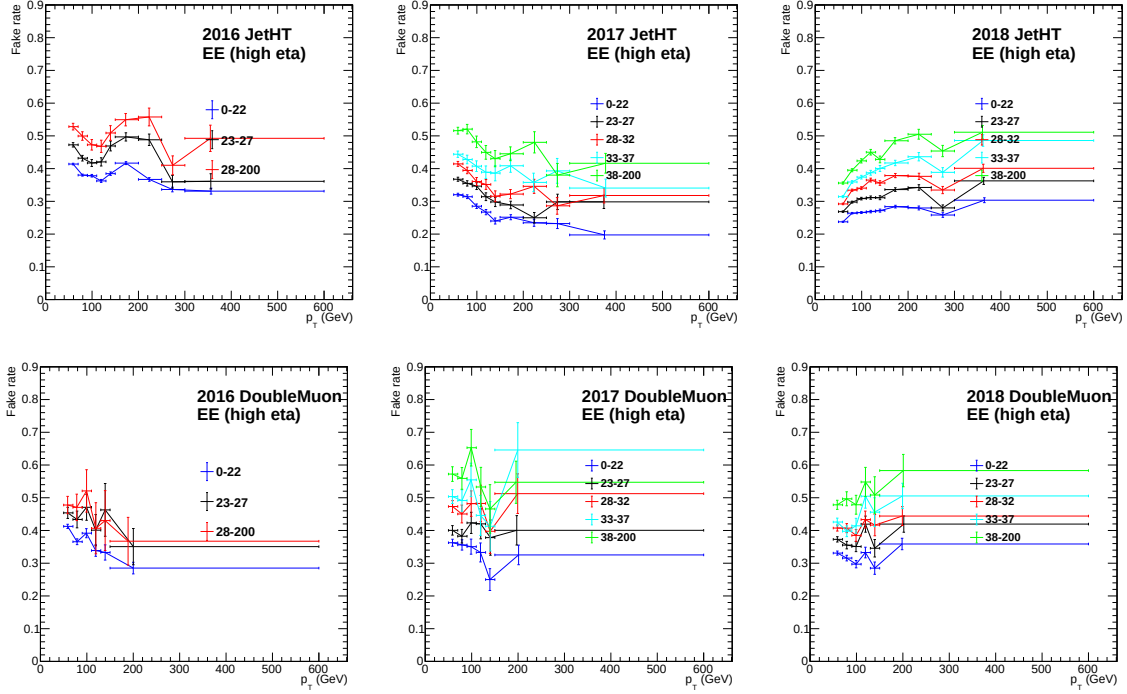


Figure 5.18: A comparison of the fake rate in EE2 ($2.033 < |\eta| < 2.5$) measured as a function of the number of selected primary vertices in the event using the **JetHT** data set (top) and **DoubleMuon** data set (bottom). The fake rates are determined separately for 2016 (left), 2017 (center), and 2018 (right). The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.



5.3.5 Fake Rate Systematics

Quarks and gluons hadronize differently. Gluon jets in general contain a larger number of particles and are therefore less likely to pass the isolation criteria of the photon ID. They therefore have a lower probability of faking a photon than quark jets. While taggers exist in data, it is difficult to distinguish between quark and gluon jets in data in an unbiased way. This represents a source of systematic uncertainty on our fake background estimation. The use of the **JetHT** and the **DoubleMuon** data sets are motivated by the need to estimate the dependence between the fake rate on the ratio of quark to gluon flavor in the jets used in the fake rate measurement. The **JetHT** dataset is an admixture of both quark and gluon jets while the **DoubleMuon** dataset has a higher quark content. This higher quark content is due to the prevalence of EW bosons (V) which means there is a higher proportion of V+jet events. The difference between the fake rates extracted from these two datasets is taken as an estimate of the systematic uncertainty due to the quark/gluon flavor of jets.

We emphasize that the **JetHT** and **DoubleMuon** are not MC samples but actual data sets. In the **JetHT** data set, all photons are matched within $\Delta R < 0.6$ to the leading jet in the event - this removes any trigger bias that could arise if the fake photon is not the leading jet and therefore some other object has to fire the trigger. As much as possible we are trying to select photons that are comparable to the target sample composition, i.e. we would like our jet composition to be biased toward quark-initiated rather than gluon-initiated jets. By focusing on the photons near the leading jets, we achieve this goal. In principle, the **DoubleMuon** dataset covers this difference as explained earlier, but it is better to try to get the central value as close as possible.

Multiple interactions in each proton bunch crossings also affect our estimates of the fakes. The fake rate is then sensitive to the pileup conditions in the CMS detector as shown in Figure 5.19, Figure 5.20, and Figure 5.21. These fake rates are binned by the number of primary vertices which are summarized in Table 5.2.

Figure 5.19: A comparison of the 2016 fake rate in EB (top), EE1 ($1.566 < |\eta| < 2.033$) (middle), and EE2 ($2.033 < |\eta| < 2.5$) (bottom) measured as a function of the number of selected primary vertices in the event. The ranges are $0 < n_{PV} \leq 22$ (left), $23 \leq n_{PV} \leq 27$ (center), and $28 \leq n_{PV} \leq 200$ (right). The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.

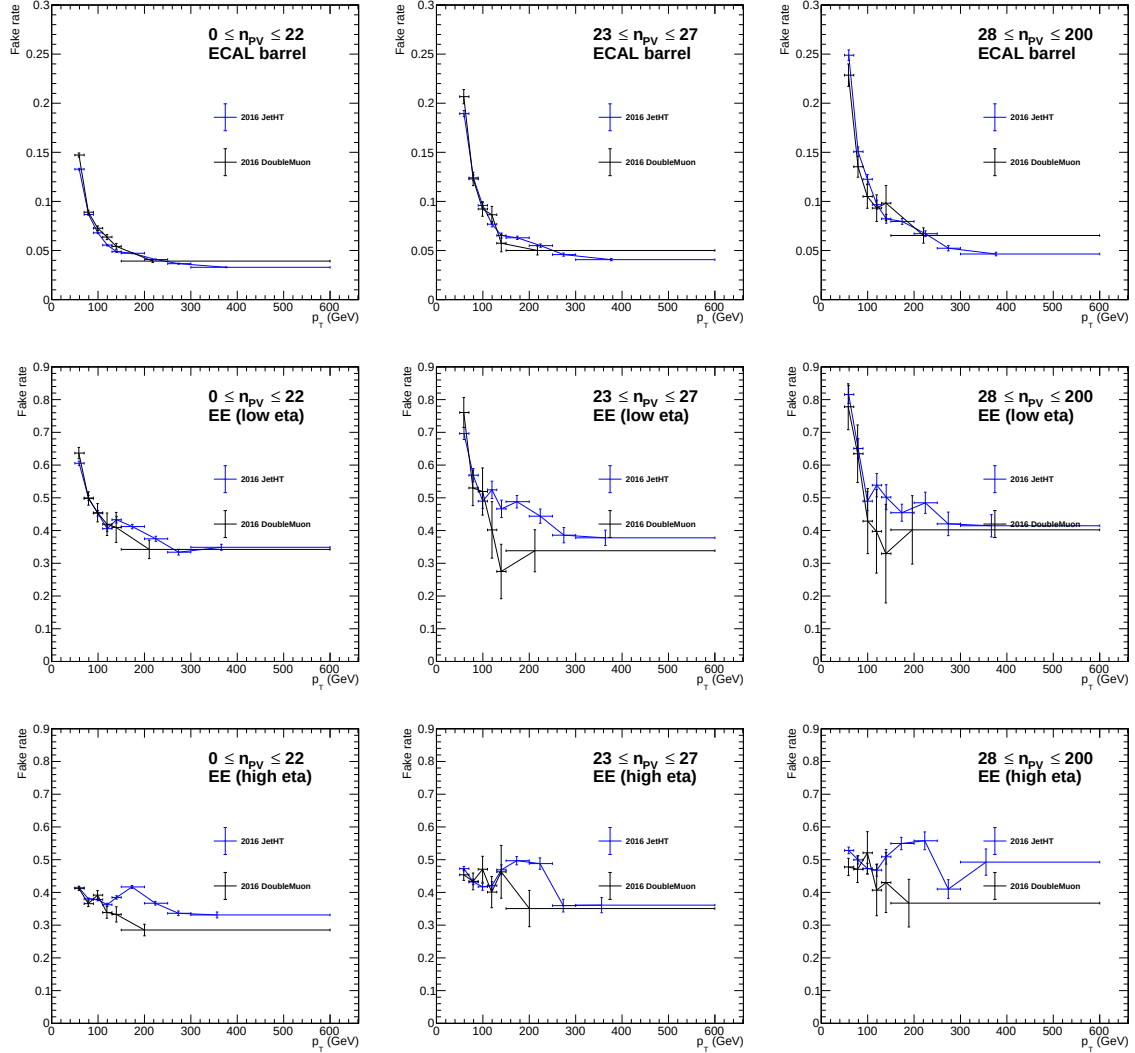


Table 5.2: There are three n_{PV} bins for 2016. The 28-200 bin are further split into 3 additional bins for 2017 and 2018. The ranges are $0 < n_{PV} \leq 22$ (left), $23 \leq n_{PV} \leq 27$ (center), and $28 \leq n_{PV} \leq 200$ (right). The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.

2016	2017-2018
0-22	0-22
23-27	23-27
28-200	28-32
-	33-37
-	38-200

Figure 5.20: A comparison of the 2017 fake rate in EB (top), EE1 ($1.566 < |\eta| < 2.033$) (middle), and EE2 ($2.033 < |\eta| < 2.5$) (bottom) measured as a function of the number of selected primary vertices in the event. From left to right, the ranges are $0 < n_{PV} \leq 22$, $23 \leq n_{PV} \leq 27$, $28 \leq n_{PV} \leq 32$, $33 \leq n_{PV} \leq 37$, and $38 \leq n_{PV} \leq 200$. The ranges used for comparison were chosen to reflect a roughly uniform level of statistics between all ranges.

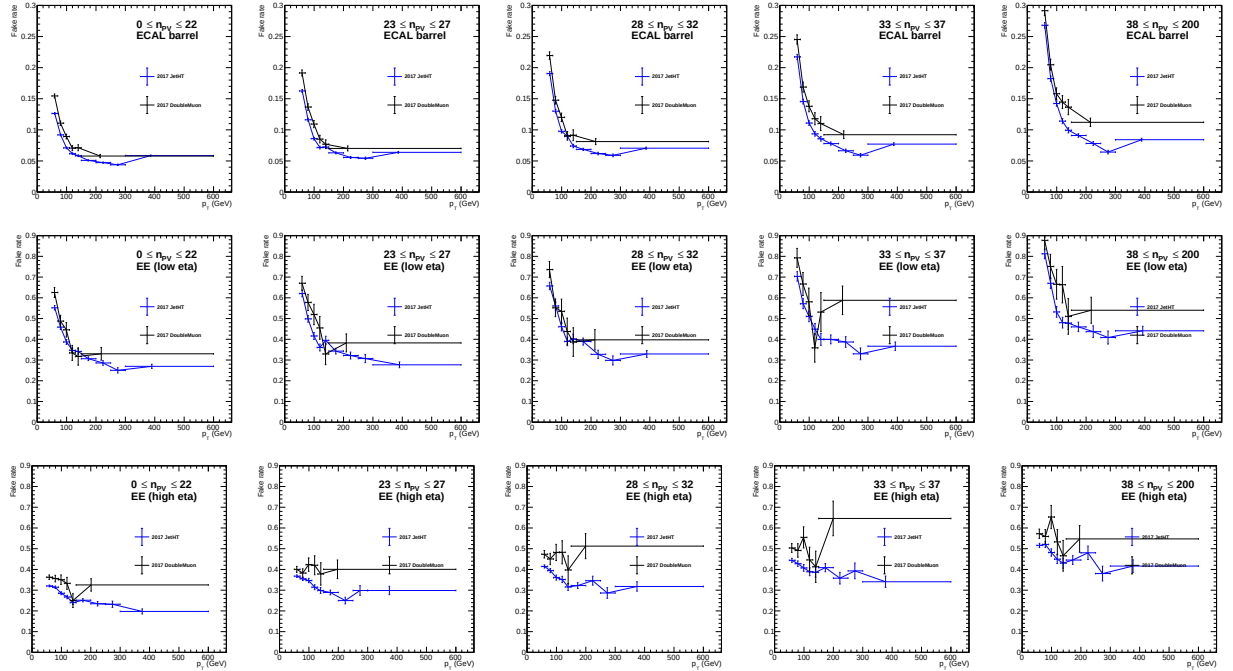
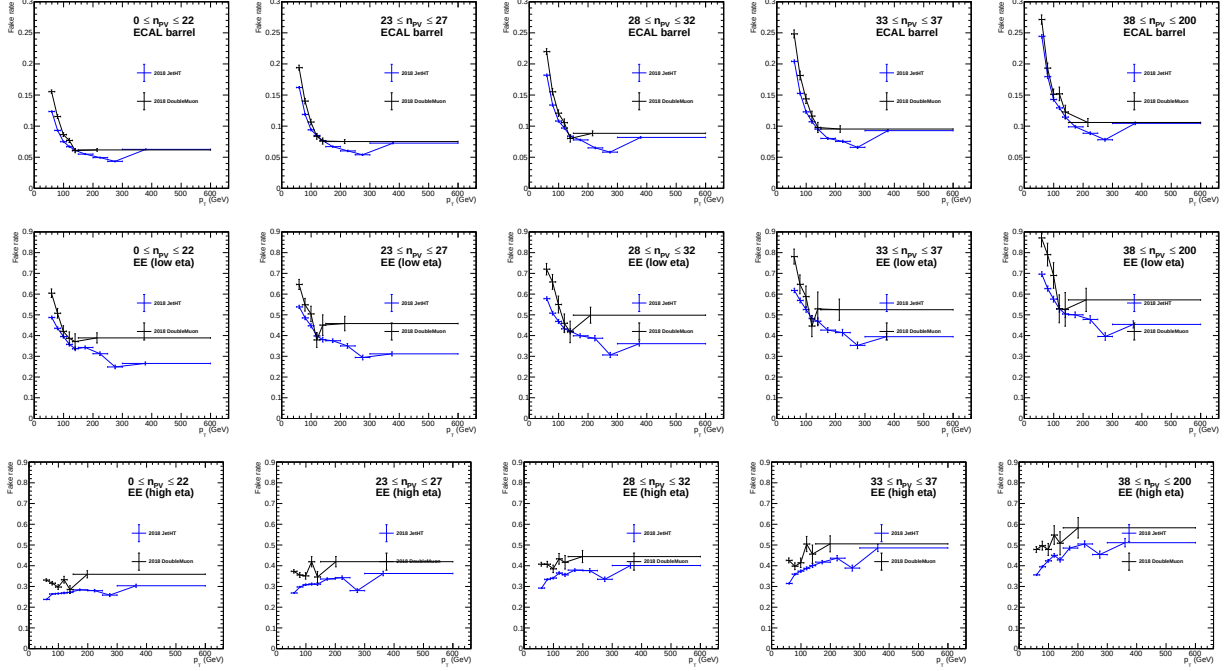


Figure 5.21: A comparison of the 2018 fake rate in EB (top), EE1 ($1.566 < |\eta| < 2.033$) (middle), and EE2 ($2.033 < |\eta| < 2.5$) (bottom) measured as a function of the number of selected primary vertices in the event. From left to right, the ranges are $0 < n_{PV} \leq 22$, $23 \leq n_{PV} \leq 27$, $28 \leq n_{PV} \leq 32$, $33 \leq n_{PV} \leq 37$, and $38 \leq n_{PV} \leq 200$.



5.3.6 Fake Rate Application

The “fake rate” obtained is used technically as a transfer function. We identify objects in the data which pass the denominator definition, then use the ‘fake rate’ to reweight those objects to reflect the relative probability that the object might have passed the Photon ID but is actually a jet that fragmented to look like a photon as defined by the numerator in the fake rate definition in Table 5.1. The data we use comes from the **DoubleEG** and **EGamma** datasets discussed in . These data sets are listed in Table B.1. We identify objects in these datasets which pass the Denominator selection in Table 5.1 which is broadly speaking ‘Loose-but-not-tight’. We call these objects Fake. There is no overlap with the ‘Tight’ objects, i.e. Numerator objects which pass the photon ID. Events that have an object pair where one candidate that passes the Denominator definition (‘Fake’) or Loose, and another candidate passes the actual photon ID (called ‘Tight’) are categorized as Tight-Fake (TF) or Fake-Tight (FT) depending on which object has a higher p_T . Events with objects that pass the Denominator definition are called Fake-Fake (FF). For these categories, we reweight the pair based on evaluating the appropriate fake rate function¹ depending on the eta region (EB central, EE forward, etc), pileup, and p_T of the Fake object(s).

There are then two sources of fake background events. These are the jj and $\gamma + j$, which are two jets and a photon+jet backgrounds. The jj and $\gamma + j$ backgrounds are estimated by reweighting each of the two F objects in the data sample and by reweighting the TF and FT samples minus twice the FF samples, respectively. This is succinctly expressed as,

$$\begin{aligned} N_{jj} &= N_{LL} f_1 f_2, \\ N_{\gamma j} &= N_{LT} f - 2N_{LL} f_1 f_2. \end{aligned} \tag{5.2}$$

More details on the mathematical derivation of Eqs. 5.2 is shown in Appendix C. It is important to note that the jj contribution was 100 times smaller than the gj contribution and hence was dropped in the final limits calculation.

¹The code for application of the fake rate is found in this link.

5.4 Photon Fake Rate Closure Test

We validated the fake rate method using an MC-based approach. The general idea is that if we apply the same method to MC, by treating it as if it were data, the results would yield a fake prediction comparable to the results from MC truth information. The MC samples that were treated as data were QCD dijets and GJets listed in Tables B.7 and B.8. These were chosen such that there is a sufficient number of fake photons present. The truth information on the actual number of fakes in the MC sample is obtained by matching each `pat::Photon` that passed our ID to the best final state (status 1) gen particle within $\Delta R \leq 0.1$ in the pruned collection of generated particles, if possible. In this section, we refer to photons that have been reconstructed and identified by the Particle Flow algorithm as `pat::Photons`. If the `pat::Photon` is matched to a hadron or when the final state gen photon's comes from a hadron then they are fakes. A complication arises when the photon has multiple mothers, which is the usual case. The sole mother is chosen by finding the best match in ΔR among all of its mothers. In summary, fake photons are tagged as such if they come from hadron decays. Comparing the fake prediction from the fake rate method and using a simple counting method from the truth information, we validate our fake rate.

The closure test has three steps. First we determine the accuracy of the fake templates produced using the sideband definition (treating MC as data) and comparing them with the MC truth templates as in Figure 5.22. Second, we perform a fitting procedure where we extract the fake yields and compare with the actual yields from the MC truth information as a function of p_T and η . Third, we use the fake rates from the previous step to reweight denominator objects, or objects that pass a looser photon ID, to predict the kinematics of the actual known fakes. These three steps are sequentially described in more detail below. Here we show that by splitting the EB and EE regions into inner and outer regions, the fake prediction improves as the fake rate is no longer averaged over the entire barrel or endcap region.

5.4.1 Testing the Fake Template from the Sideband Definition

The first step of the closure test is to create fake and real templates and validate that the chosen sideband-defined template does a reasonable job of replicating the truth template for known fakes. The MC fake templates are filled from the MC sample listed in Tables B.7 and B.8, using the same selection and sideband definitions as done with the data (see Table B.1). These templates are p_T -dependent so we prepared templates for p_T bins of 50–70, 70–90, 90–110, 110–130, 130–150, 150–200, 200–250, 250–300, 300–600 GeV. We use coarser binning compared to the actual fake rate method due to the limited statistics in the MC samples.

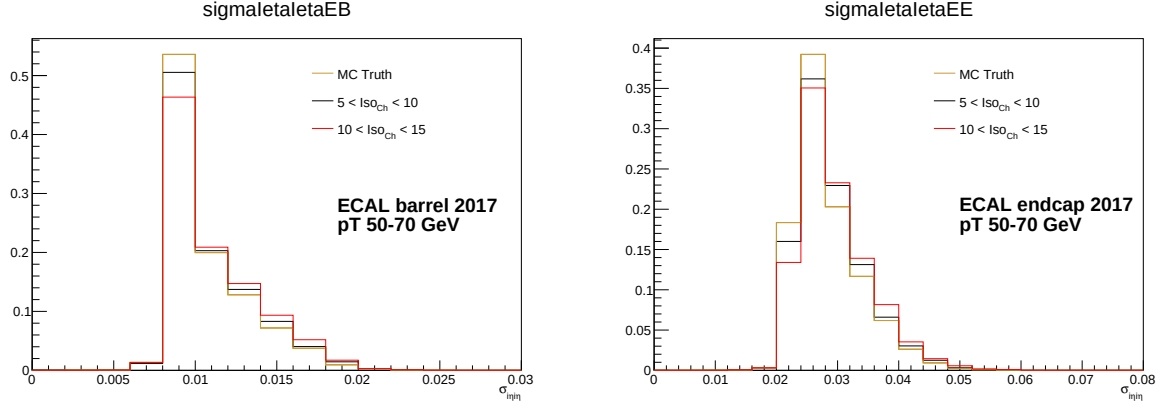
In Figure 5.22, we see that the fake templates for each sideband Iso_{Ch} choice does a reasonable job of replicating the MC truth templates. The comparison is shown for the lowest p_T bin separated in EB and EE. We see that the templates don't fully replicate the MC truth because of some bias. This bias is unavoidable because fakes in a sideband are intrinsically different from fakes that pass the ID. This biasing is because of the correlation of the Iso_{Ch} variable with $\sigma_{i\eta i\eta}$. This bias is the main source of systematic uncertainty on this fake rate method.

The least bias is seen for the $5 < \text{Iso}_{\text{Ch}} < 10$ GeV sideband making it our nominal sideband choice for the analysis.

5.4.2 Testing the Template Fit Method for Extracting Fake Yields

After preparing and validating the fake templates in the previous step, we fit these templates together with the real templates (same as selection as the data) to the numerator $\sigma_{i\eta i\eta}$ distribution. Figure 5.23 illustrates the fit for p_T bin $90 < p_T < 130$ GeV. The numerator templates are filled with objects that pass the photon ID and it contains both real and fake photons. The real and fake templates are fitted to the numerator templates which are by definition contains both real and fake photons that pass the photon ID. In this step we extract the fake yields by fitting the fake templates prepared and validated in the previous step.

Figure 5.22: Representative $\sigma_{i\eta i\eta}$ templates for the ECAL barrel (left) and endcap (right) for 2017 in the 50-70 GeV p_T -bin are shown here. The distributions in gold show the MC truth while the black and red show the fake templates in the $5 < \text{Iso}_{\text{Ch}} < 10$ GeV and $10 < \text{Iso}_{\text{Ch}} < 15$ GeV sidebands.



The template fit in $\sigma_{i\eta i\eta}$ is performed using a negative log-likelihood function with Barlow-Beeston errors [132]. The fits are done for each p_T bin in EB, EE and then we split the EB and EE into inner and outer regions to fine-tune the fake yield prediction. After the fit, the $\sigma_{i\eta i\eta}$ ID cut is applied yielding a fake prediction for each separate p_T bin. We then divide this fake prediction by the denominator in the MC sample, where we then get an MC fake rate as a function of p_T . This same procedure was repeated using two different Iso_{Ch} sideband choices for the MC fake templates. We then compare the obtained MC fake rate to the MC truth fake rate which was obtained by simply counting the number of fake photons in each p_T bin and dividing by the denominator to get an MC truth fake rate.

We see from Figure 5.29 and Figure 5.30 a generally good agreement between the MC Truth fake rates and the MC fake rates. The "non-closure" or the failure to fully replicate the truth MC fake rates is attributed to the bias in the fake templates constructed from the sideband definition. The underprediction is minimized by choosing the sideband that is close as possible to the ID region ($\text{Iso}_{\text{Ch}} < 5 \text{ GeV}$).

In general, we observe an underprediction within an uncertainty of about 30% as shown in Figs. 5.24 and 5.25. Note that we see larger statistical fluctuations at higher p_T due to the lack of statistics for fake photons.

Figure 5.23: A representative template fit from the 2018 shows using photon objects from the ECAL barrel that have from 50 to 70 GeV in the Iso_{Ch} sideband $5 < \text{GeV} < 10$.

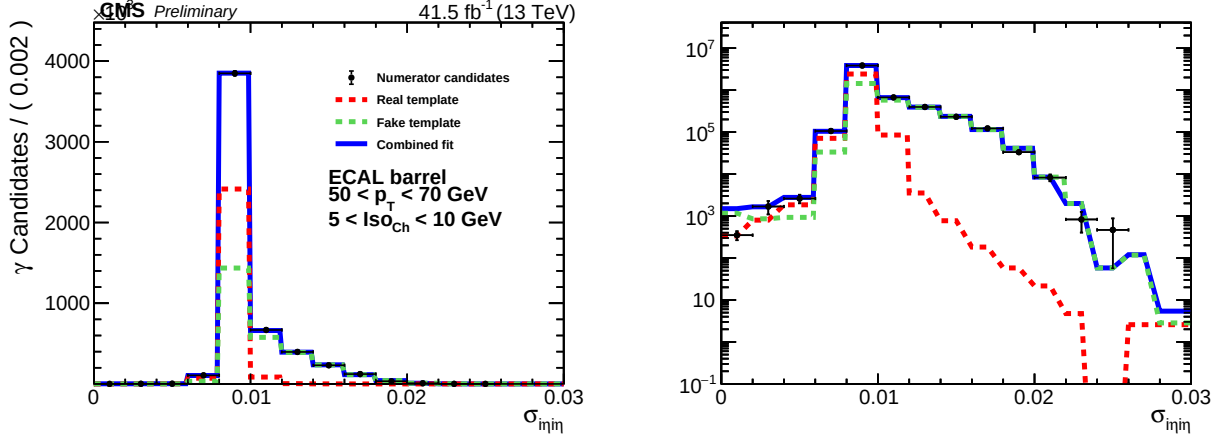


Figure 5.24: The 2017 MC Truth fake rates and the MC fake rates shows a generally good agreement with each other for both EB and EE. The MC Truth was obtained by a 'counting' procedure whereas the fake rate was obtained through a series of steps which involves a template fitting.

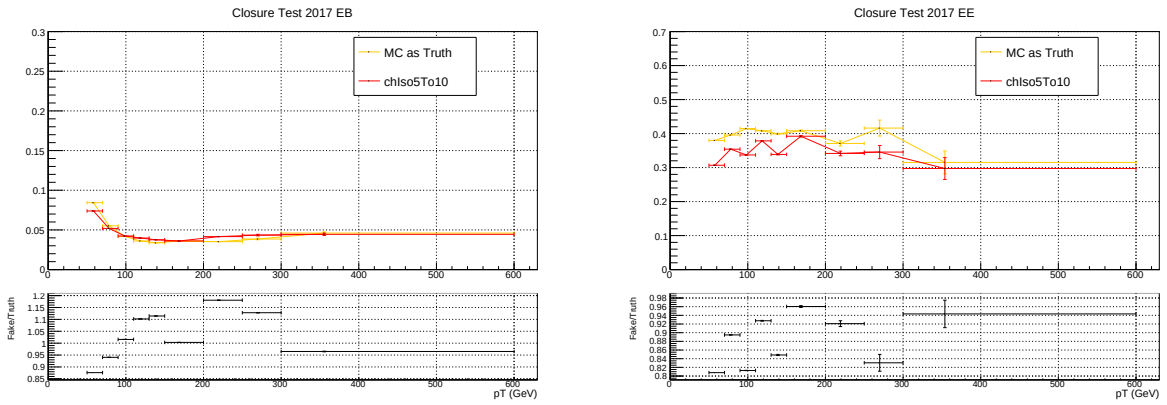
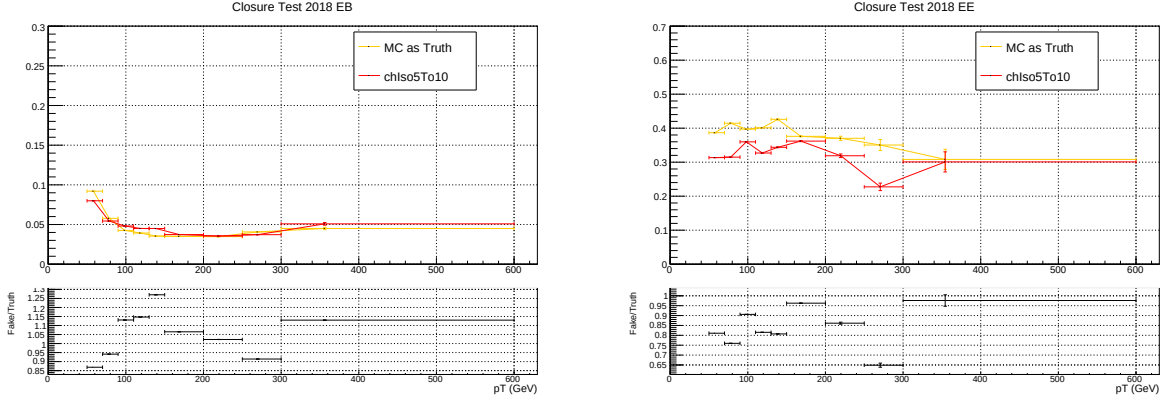


Figure 5.25: The 2018 MC Truth fake rates and the MC fake rates shows a generally good agreement with each other for both EB and EE. The MC Truth was obtained by a ‘counting’ procedure whereas the fake rate was obtained through a series of steps which involves a template fitting.



5.4.3 Testing that Reweighted Denominator Objects Reflect the Kinematics of Known Fake Objects

The final step in the MC closure test is to verify that we can reproduce an accurate representation of the actual fake background estimate used in the analysis. We do this by reweighting the denominator objects (‘photon-like’ jets) with the fake rate as a function of p_T for different eta regions (EB or EE). This gives us the kinematic distributions of known fake events as seen in Figure 5.26 and Figure 5.27. The fake rate used to reweight the denominator distributions were obtained by fitting separate functions to the MC fake rates using the sideband choice that reflects the best “closure” or agreement with the truth. This is the nominal sideband choice which is $5 < \text{Iso}_{\text{Ch}} < 10$ GeV.

Figure 5.26 and Figure 5.27 shows single photon kinematics for p_T , η , and ϕ for the reweighted object compared against the distributions. Agreement in the barrel is excellent but in the endcap, we see some disagreement that indicates that endcap fake rate has more eta-dependence. To verify this we split the barrel and endcap regions into inner and outer regions and went through steps 1 to 3 of the closure test. The inner barrel (EB1) corresponds to $|\eta| < 0.7221$. The outer barrel (EB2) corresponds to $0.7221 < |\eta| < 1.4442$. The inner endcap (EE1) $1.566 < |\eta| < 2.033$ while the outer endcap corresponds to $2.033 < |\eta| < 2.5$.

We can see from the figure 5.31 some improvement in the agreement in the fake prediction and the MC truth fakes.

Figure 5.26: The two columns show the EB (left) and EE (right) single photon kinematic plots for reweighted objects for 2017.

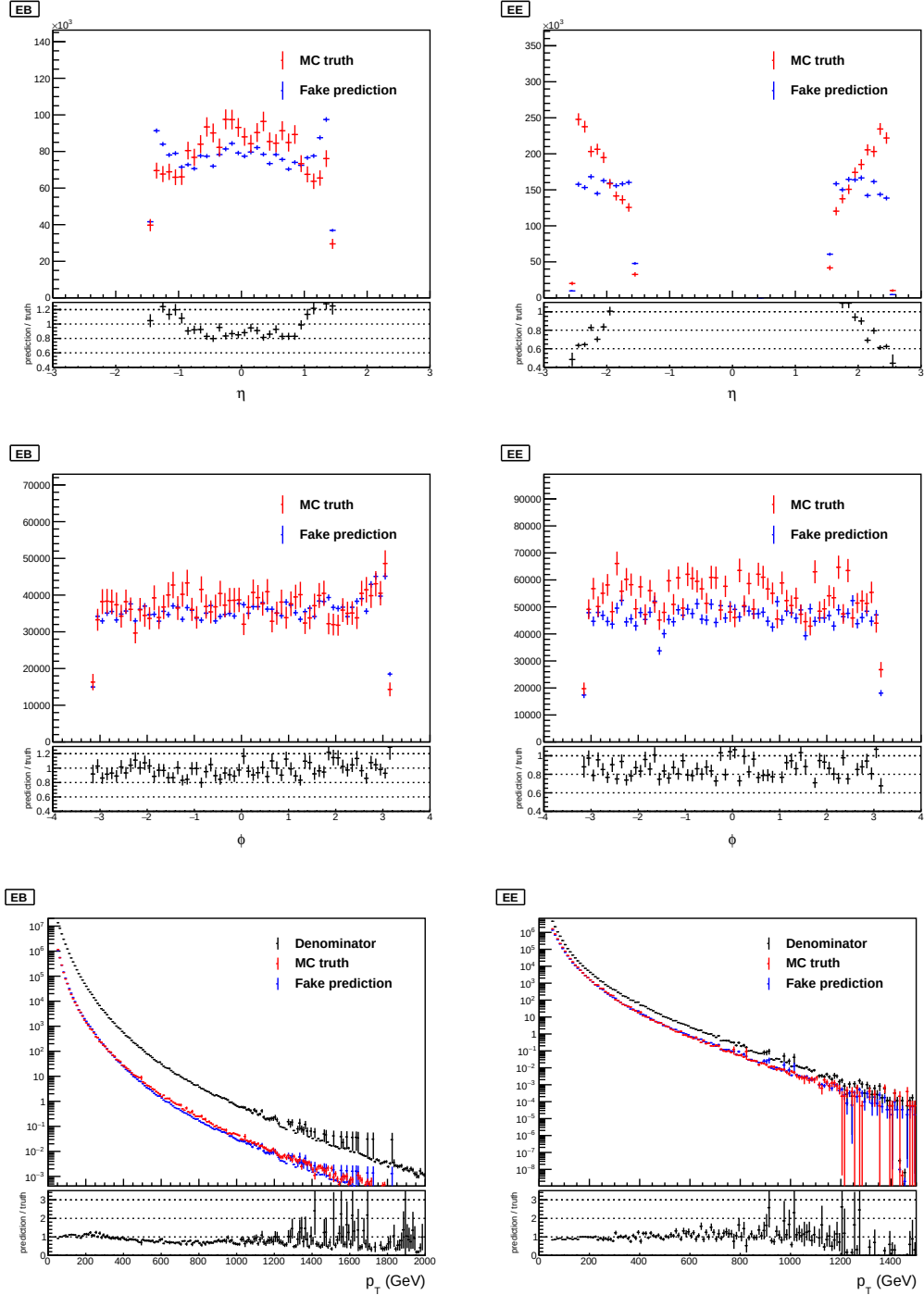


Figure 5.27: The two columns show the EB (left) and EE (right) single photon kinematic plots for reweighted objects for 2018.

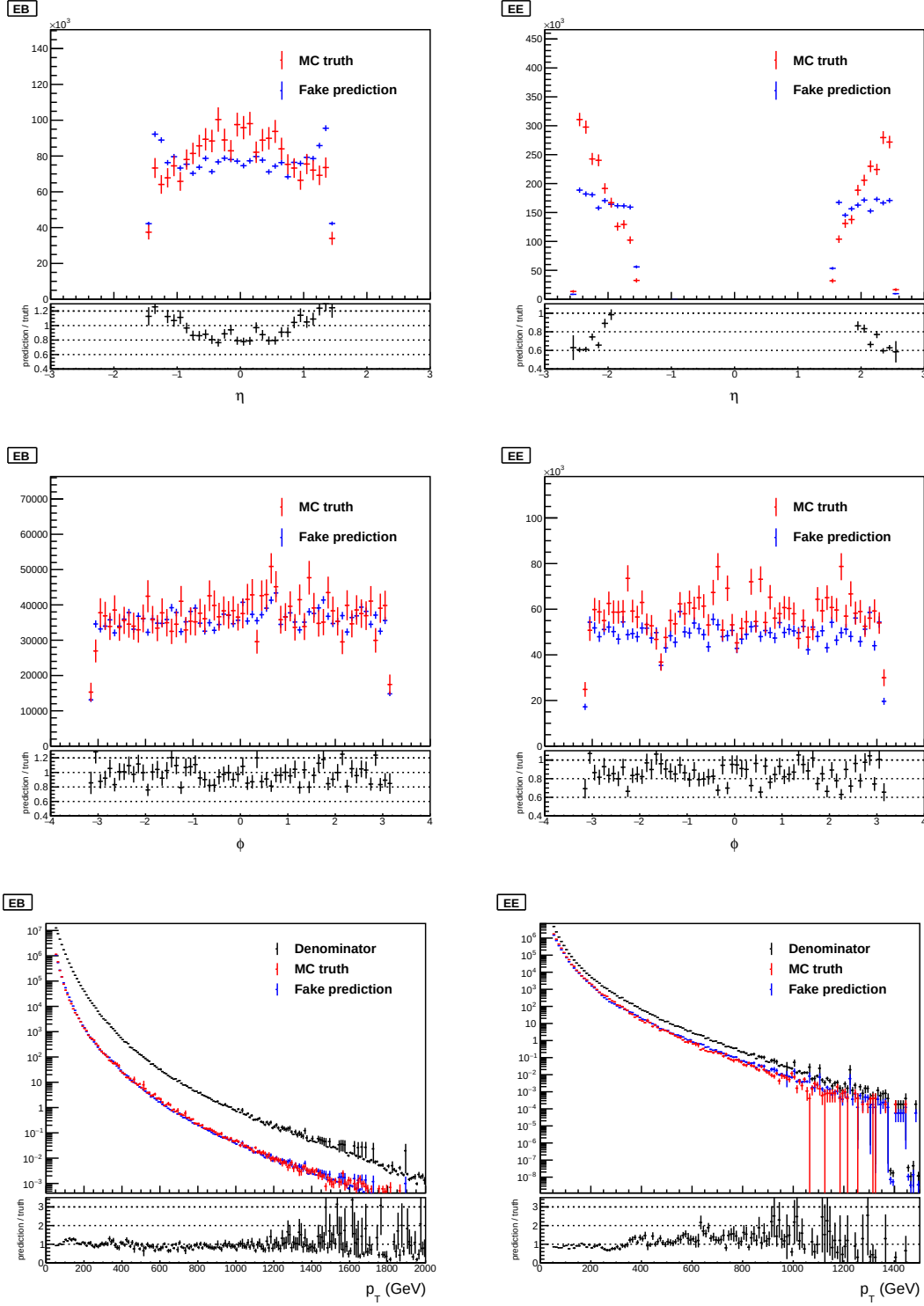


Figure 5.28: A representative template fit from the 2017.

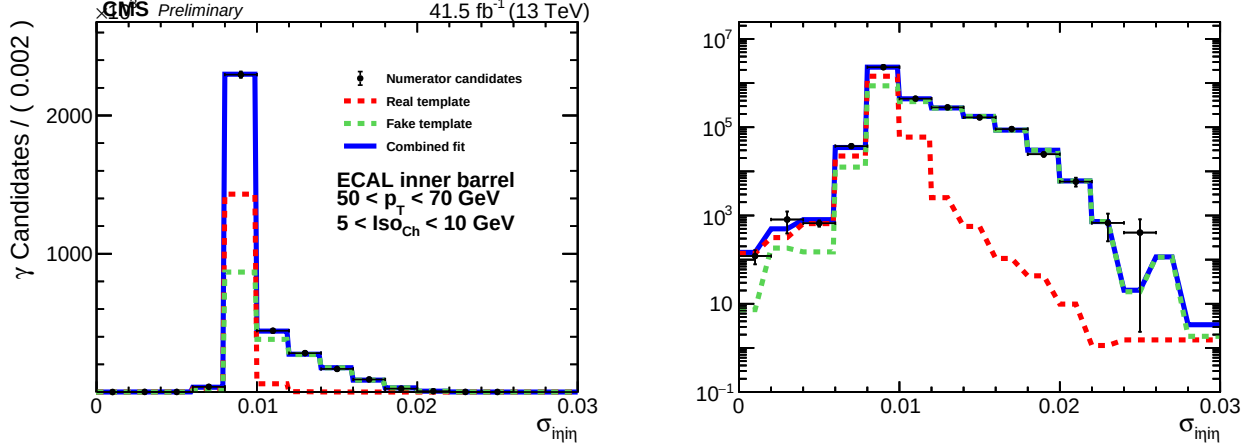


Figure 5.29: The Fake Rates were recalculated with the splitting of the barrel region into inner and outer regions. The inner barrel (EB1) corresponds to $|\eta| < 0.7221$. The outer barrel (EB2) corresponds to $0.7221 < |\eta| < 1.4442$. This was done to improve the agreement between the fake prediction and the MC Truth fake prediction for the last step of the closure test.

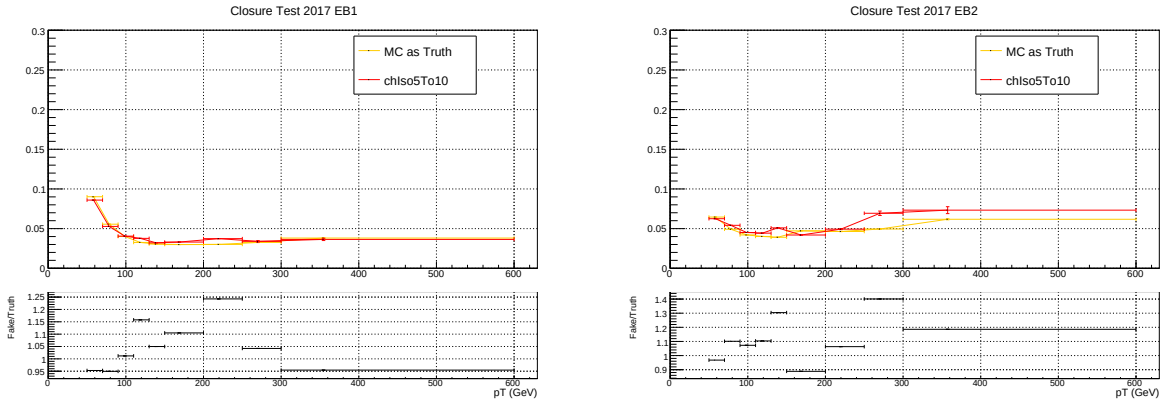
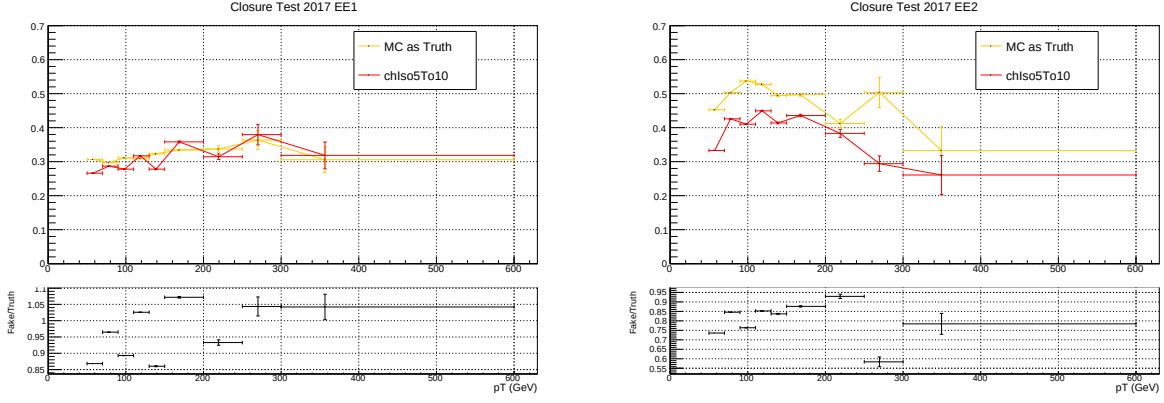


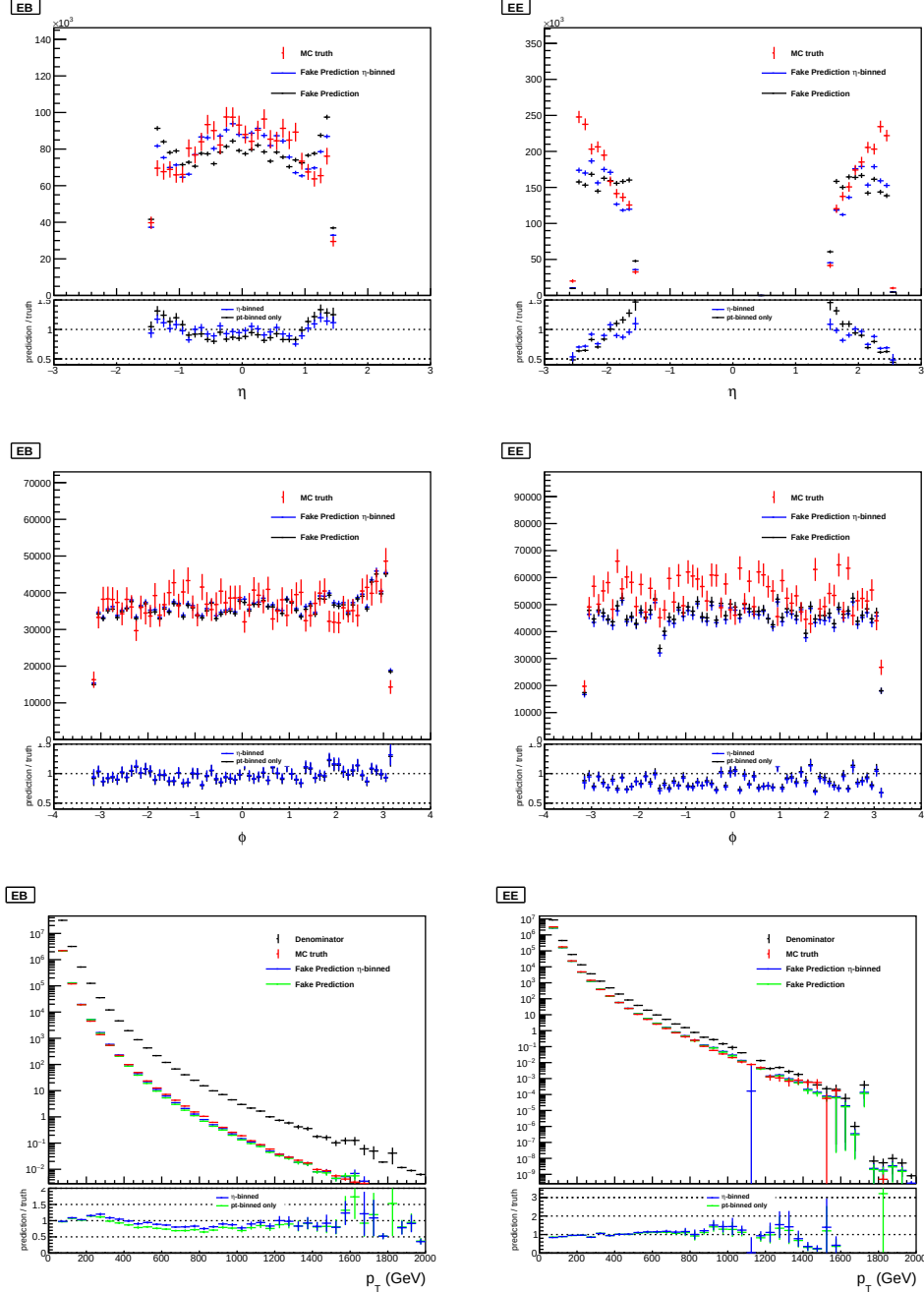
Figure 5.30: We also split the endcaps into inner and outer regions to improve the fake prediction. The inner endcap (EE1) $1.566 < |\eta| < 2.033$ while the outer endcap corresponds to $2.033 < |\eta| < 2.5$.



5.4.4 Conclusions from the Closure Test

The fake rate method was validated via the Closure Test by applying the same steps on MC. The fake rate extracted from the method does a reasonable job of replicating the MC Truth fake rate as shown in Figure 5.24 and 5.25. Approximately 30% non-closure across the 0 – 600 GeV pT range was observed. It was also observed that splitting the endcap and barrel into inner and outer regions improved the agreement between the kinematic plots of the Fake Prediction and the MC Truth. The inner barrel (EB1) corresponds to $|\eta| < 0.7221$. The outer barrel (EB2) corresponds to $0.7221 < |\eta| < 1.4442$. The inner endcap (EE1) $1.566 < |\eta| < 2.033$ while the outer endcap corresponds to $2.033 < |\eta| < 2.5$. In principle, implementing a further finer binning strategy would likely enhance the agreement between the MC truth and the Fake prediction but it comes at the cost of reduced statistical reliability within individual bins.

Figure 5.31: Splitting the barrel and endcap regions shows some improvements in the agreement between the fake rate reweighted Fake Prediction and the MC Truth Prediction for 2017. A similar result is observed for 2018. The inner barrel (EB1) corresponds to $|\eta| < 0.7221$. The outer barrel (EB2) corresponds to $0.7221 < |\eta| < 1.4442$. The inner endcap (EE1) $1.566 < |\eta| < 2.033$ while the outer endcap corresponds to $2.033 < |\eta| < 2.5$.



CHAPTER 6

SIGNAL MODELLING

In this chapter we discuss the MC generation of the ADD Large Extra Dimensions, Clockwork and the Unparticles - the three model interpretations used for our statistical analysis.

6.1 ADD Large ED Samples

The large extra dimensions model of Arkani-Hamed, Dimopoulos and Dvali (ADD) [40, 41] can produce diphoton events through virtual graviton exchange, leading to an enhancement of the diphoton invariant mass spectrum at large mass. Since the spacing between Kaluza-Klein graviton states is typically small in the ADD scenario, this effectively gives a continuum, non-resonant enhancement.

We consider two conventions for the ADD signal, one having a positive interference with the Standard Model diphoton production process and another with negative interference. These conventions are equivalent to conventions by Giudice, Ratazzi and Wells (GRW) [55], and the one by Hewett [58] which were explored in a previous paper [131]. These are labeled as **NegInt-1** and **NegInt-0**, respectively ¹.

To span the entire mass spectrum, samples were produced in split mass bins from $m_{\gamma\gamma} = 500$ GeV up to the value of Λ_T . Note that the ADD model is an effective theory that is only valid below Λ_T . The mass binning are split in such a way that a large effective luminosity

¹We found the convention choice in PYTHIA a little odd for this parameter, so we note here that PYTHIA default for the ADD model is to apply negative interference (parameter **NegInt** defaults to 0), and to turn this off, you have to set **NegInt=1**. This mislabeling was confirmed with Steve Mrenna. It results from the fact that the formula implemented within PYTHIA for the interference already has a minus sign built into it.

is generated at higher masses so as to get better statistics. For illustration, the samples generated with $\Lambda_T = 9000$ have the following mass bins, 500–1000, 1000–2000, 2000–4000, 4000–9000.

The amplitude for a process involving virtual graviton exchange involves a sum over the Kaluza-Klein tower of graviton mass states. This results in a higher dimension operator with coefficients suppressed by some mass scale [133], which can be parametrized by $\eta_G = \mathcal{F}/M_S^4$, where $\mathcal{F}$ is a dimensionless parameter for which several conventions exist in the literature:

$$\mathcal{F} = \begin{cases} 1 & \text{(GRW),} \\ \log\left(\frac{M_S^2}{\hat{s}}\right), & \text{if } n_{ED} = 2 \\ \frac{2}{n_{ED}-2}, & \text{if } n_{ED} > 2 \\ \pm \frac{2}{\pi} & \text{(Hewett),} \end{cases} \quad \text{(HLZ),} \quad (6.1)$$

where $\sqrt{\hat{s}}$ is the center-of-mass energy of the colliding partons. HLZ and GRW denote the names of the authors who introduced the convention: Han, Lykken and Zhang [57]; and Guidice, Rattazzi and Wells [55].

The results of the 2016 nonresonant diphoton analysis were interpreted in each of these three scenarios. The results presented in this note use the GRW convention, with $\Lambda_T = M_S$, or the equivalent with negative interference, $\mathcal{F} = -1$ (as explained earlier, this change in convention is driven by the switch to ADD signal generation using PYTHIA, which uses a different ADD convention than the prior SHERPA usage in 2016).

The $m_{\gamma\gamma}$ distributions for the signal samples are displayed in Figure 6.1 both before and after the subtraction of the SM background. These plots were generated using the 2017 signal samples but the 2018 signal samples are similar.

The ADD signal samples (Table C.1) used for the analysis were produced at leading order (LO) using PYTHIA 8.2 [119] with the NNPDF 3.1 LO set of parton distribution functions included in the underlying event tune, **TuneCP2**. In the context of Pythia, “Tunes” refers

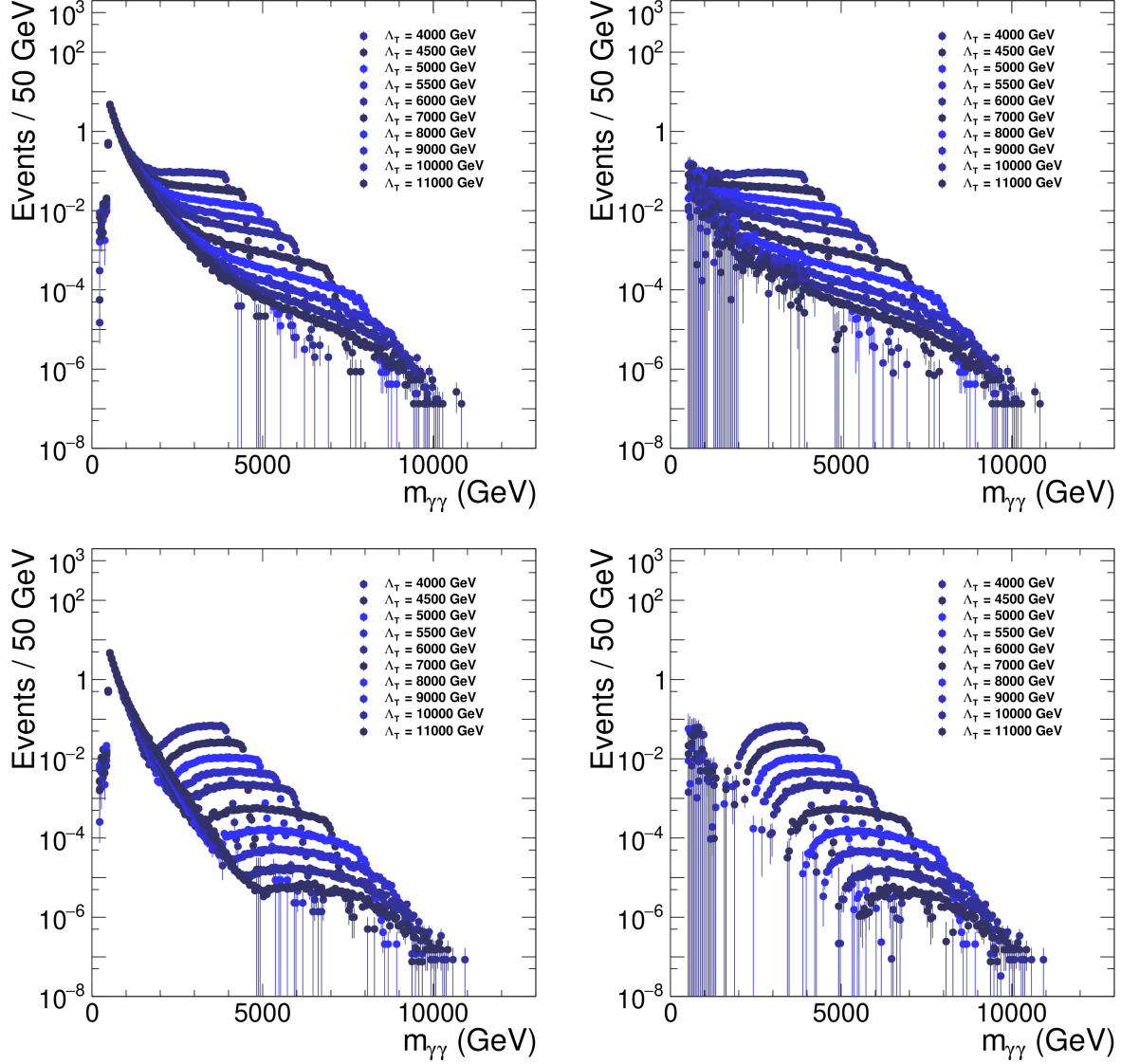


Figure 6.1: The $m_{\gamma\gamma}$ distribution in signal events for the generated values of Λ_T before (left) and after (right) subtraction of SM background. The upper(lower) plots show signals generated assuming positive (negative) interference with the SM background. The plots were generated using the 2017 ADD samples and are normalized to an integrated luminosity of 1 fb^{-1} .

to specific sets of parameters used to configure the event generator for simulating particle collision events. Previously, for our 2016 analysis [12], the samples were generated with a built-in implementation of the ADD model in SHERPA 2.1.1; but later SHERPA versions moved to using Universal FeynRules Output (UFO) interface [134] and removed the built-in ADD implementation. The UFO interface makes use of the FeynRules package in Mathematica, which assists in the generation of Feynman diagrams for specific particle physics models. The output is then saved to a format that can be used as an input to Monte Carlo generators such as SHERPA to generate events associated with the physics models. The Large Extra Dimensions (LED) implementation with the UFO presented some difficulties in interfacing with SHERPA, so we switched to using PYTHIA 8 which already has the ADD model built-in. Other related analyses, for example the dilepton [72, 135] and dijet channels [136] have also used PYTHIA 8. In PYTHIA 8, the ADD model is parametrized by the Λ_T variable which is the ultraviolet cutoff parameter for the virtual graviton exchange process.

As the ADD signal results in the same final state as the SM background, interference effects are large in the regions of phase space where the signal and background amplitudes are comparable. To take this effect into account, the signal samples are generated including both the signal and SM background amplitudes. Additional samples generated identically, but excluding the signal contribution, are used to subtract the SM background, leaving the contributions of signal and the interference between signal and background. These samples are listed in a table in the previous section.

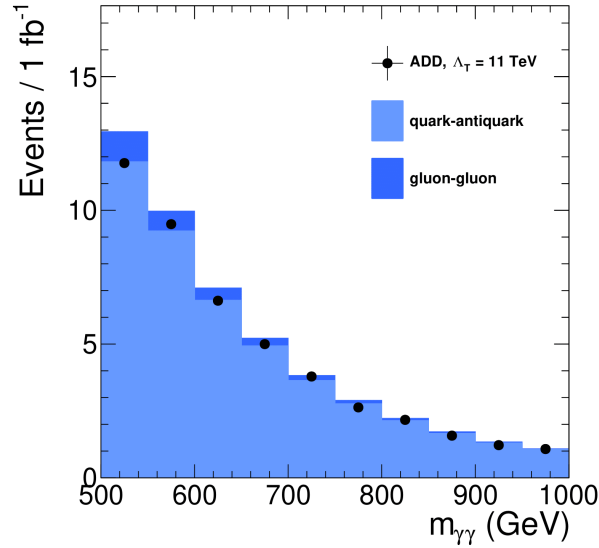
It is not computationally feasible to include additional partons in the matrix element in the signal plus background samples, and so additional partons are not included in either the signal+background or background-only samples.

Though the GG samples used to subtract the background from the ADD signal (which includes the SM interference and background) are generated with PYTHIA 8 while the GGJets samples are generated with SHERPA, it is not necessary to use the same generator; the background prediction is not used in constructing the signal distributions and vice versa, so

there is no need for consistency. The only other difference between these background samples and those used in the background prediction is an increase of the photon p_T requirement at generator level from 50 GeV to 70 GeV, which does not affect the analysis since the actual p_T threshold used in the analysis exceeds both.

These background samples include a small contribution from the gluon-gluon-initiated box diagram. This box diagram is not included in the simulation of the ADD signal and so, in order to avoid over-subtracting the background, we exclude events from the $\gamma\gamma$ background subtraction samples that have PDG ID equal to 21. Figure 6.2 demonstrates, in the low $m_{\gamma\gamma}$ region, for a signal sample with negligible signal contribution in this region, the agreement between the ADD signal and quark-antiquark background alone.

Figure 6.2: A representative ADD sample signal agreement with the quark-antiquark background (light blue) alone in the low $m_{\gamma\gamma}$ region.



6.1.1 Clockwork Model

To generate the Clockwork signal, we make use of the ADD Large Extra Dimensions samples and translated them to produce the Clockwork Prediction. In the Clockwork model, the KK modes are all on shell, so there is no interference effect. The direct term for the

ADD expression for the total cross section can be rescaled by

$$\theta(m_{\gamma\gamma} - k) \frac{30\Lambda_T^8}{283\pi M_5^3} \sqrt{1 - \frac{k^2}{m_{\gamma\gamma}^2}} \frac{1}{m_{\gamma\gamma}^5} \left[1 + \frac{5^2 \cdot 7 \cdot 17}{283 \cdot 2^8} \left(1 - \frac{k}{m_{\gamma\gamma}} \right)^9 \sqrt{\frac{m_{\gamma\gamma}}{k}} \right] \quad (6.2)$$

where Λ_T is equivalent to M_s in the GRW convention, M_5 is the fundamental scale of the gravitational interactions analogous to the fundamental Planck scale M_D in the ADD model, and k is called the clockwork spring. The clockwork spring sets the inverse size of the extra dimension. Phenomenologically, it also controls the energy scale at which the KK modes can be excited. Perturbativity constrains $k < M_5$. The direct term for the ADD expression is the third term in the equation for the total cross section for the SM plus the ADD:

$$\sigma_{total} = \sigma_{SM} + \frac{\mathcal{F}}{M_S^4} \sigma_{int} + \frac{\mathcal{F}}{M_S^8} \sigma_{ADD} \quad (6.3)$$

where $\mathcal{F}$ is defined in Eq. 6.1.

We can take the linear combination of the SM-background subtracted positive interference GRW and the negative interference Hewett samples, given by the equations:

$$\sigma_{GRW} = \frac{1}{M_s^4} \sigma_{int} + \frac{1}{M_s^8} \sigma_{ADD} \quad (6.4)$$

$$\sigma_{Hewett} = \frac{-2/\pi}{M_s^4} \sigma_{int} + \frac{4/\pi}{M_s^8} \sigma_{ADD} \quad (6.5)$$

We can use Eq. 2.66 to simplify and rewrite Eqs. 6.4 and 6.5 in terms of Λ_T :

$$\sigma_{GRW} = \frac{1}{\Lambda_T^4} \sigma_{int} + \frac{1}{\Lambda_T^8} \sigma_{ADD} \quad (6.6)$$

$$\sigma_{Hewett} = \frac{-1}{\Lambda_T^4} \sigma_{int} + \frac{1}{\Lambda_T^8} \sigma_{ADD} \quad (6.7)$$

Combining these two we get the direct term:

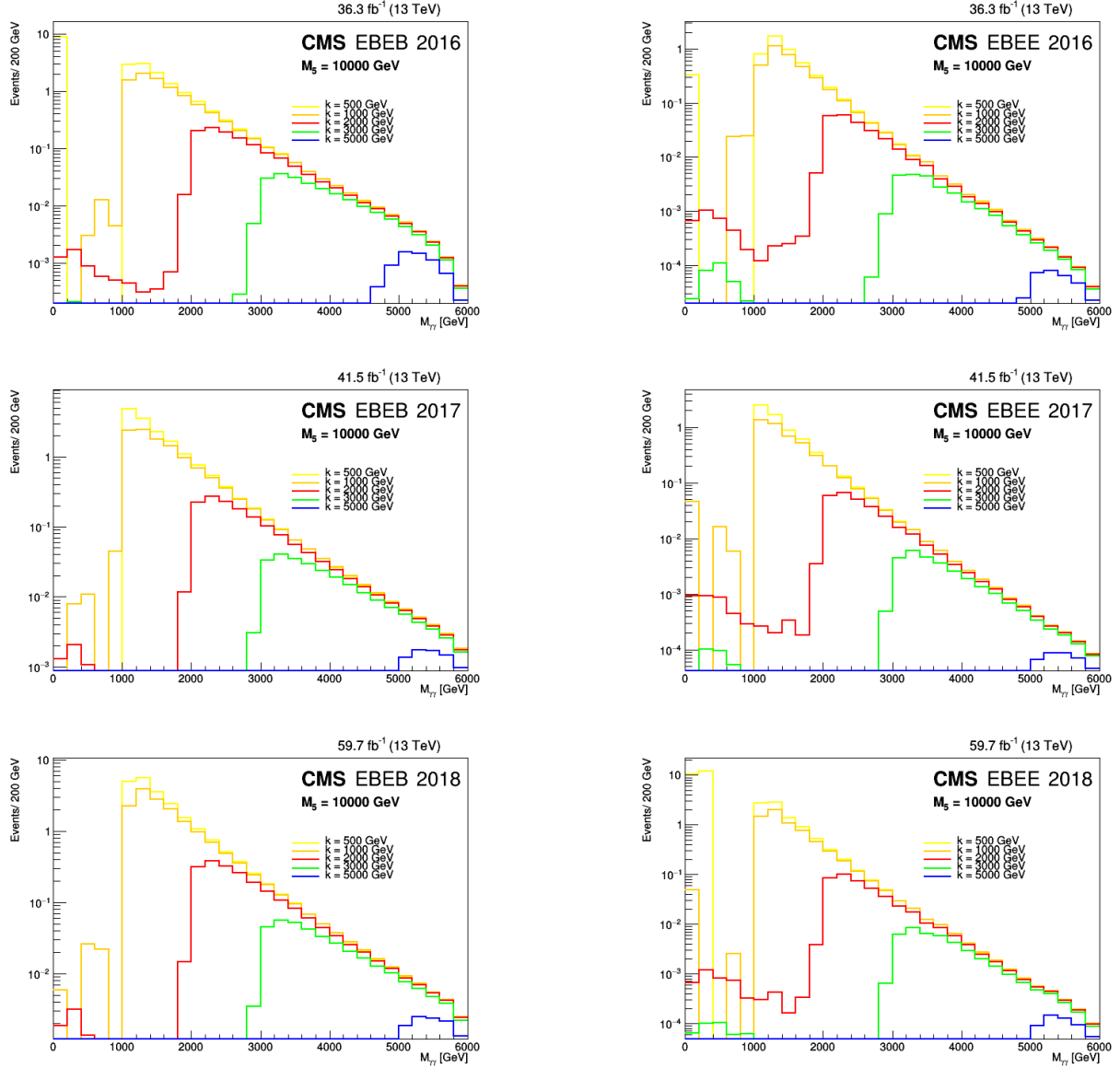
$$\sigma_{ADD} = \frac{\sigma_{GRW} + \sigma_{Hewett-}}{2} \quad (6.8)$$

It is this equation that is rescaled by Eq. 6.2 to produce the clockwork spectrum. We are careful to use the generator-level value of $M_{\gamma\gamma}$ rather than the reconstructed $m_{\gamma\gamma}$ to do the rescaling.

The signal yields for the clockwork model after the final selection and normalized to the total luminosity are shown in Figure 6.3 for EB-EB and EE-EE events. It is immediately noticeable that the shapes differ substantially from those predicted by the ADD model as shown in Figure 6.1. When k is small, the distributions are falling as much in the same way that the SM background does, just not as steeply. This provides us with an important experimental challenge in controlling our irreducible background while still providing sufficient discrimination.

Because the ADD samples are generated together with the SM contribution to account for the interference, the background-subtracted ADD prediction suffers from large statistical uncertainties at low $m_{\gamma\gamma}$. For the ADD model, this is not a problem because the SM contributions completely dominate, but for the clockwork model, this is problematic because the clockwork can predict a large excess even at low values of $m_{\gamma\gamma}$ when k is small. To mitigate this issue, we stitch together positive interference GRW and negative interference Hewett samples with different values of Λ_T . ADD samples with low Λ_T have larger ADD contributions at low $m_{\gamma\gamma}$ so this helps with the statistics. Even then, we are only able to make statistically reliable predictions for the clockwork signals for $m_{\gamma\gamma} > 900$ GeV with the samples we have. Higher Λ_T samples are used at higher $m_{\gamma\gamma}$ to deal with the unitarity truncation when $\sqrt{\hat{s}} > \Lambda_T$.

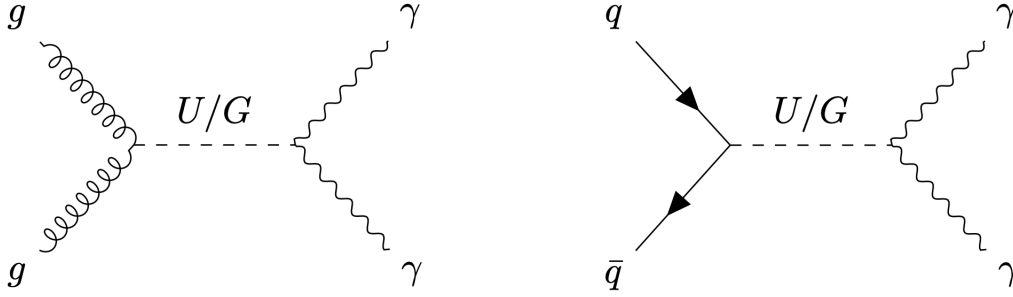
Figure 6.3: Signal shapes for the clockwork signal after the final selection for a fixed value of $M_5 = 10$ TeV and varying values of k normalized to the total integrated luminosity for EB-EB events in the left and EB-EE events in the right, respectively.



6.1.2 Unparticles to Diphoton

In this dissertation we also show our preliminary studies on both the spin-0 (scalar) and spin-2 (tensor) Unparticles to diphoton production. As shown in Figure 6.4, the Unparticles can be produced through $q\bar{q}$ annihilation and gg fusion subprocess with spin-0 and spin-2 unparticles appearing as propagators.

Figure 6.4: Leading-order Feynman diagrams for the Unparticles to two photons.



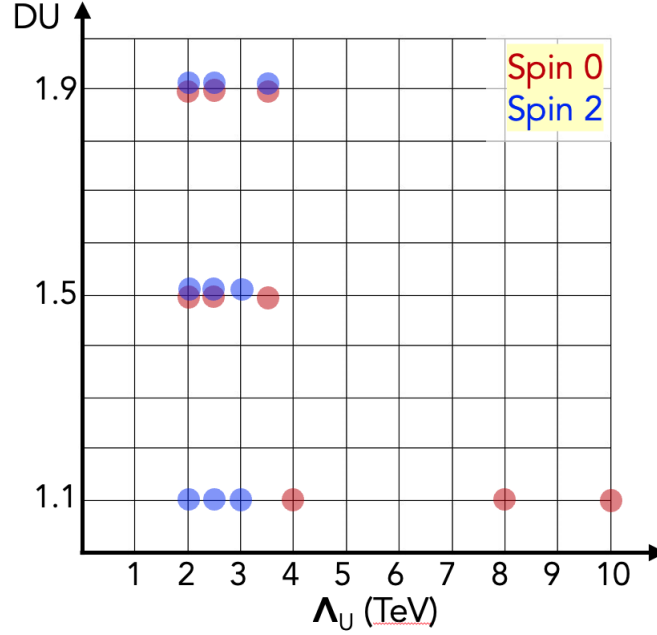
Our search space is a two-dimensional parameter space, where d_u is the scale dimension parameter and Λ_U , is the Unparticle renormalization scale. To simplify our search, we set the strength of the Unparticle Coupling to the Standard Model $\lambda = 1$.

The Unparticles model produce diphoton events via virtual graviton exchange and appear as a non-resonant excess in the diphoton invariant mass spectrum similar to the ADD model [70]. The Unparticles signal samples listed in Table C.2 were produced at leading order (LO) using PYTHIA 8.2 [119], with the NNPDF3.1 LO set of parton distribution functions as found in the event tune, **TuneCP2** [122].

For our first look, we span the 2-dimensional parameter space as follows: for each spin we chose du values 1.1, 1.5, 1.9. For numerical reasons[137], it is recommended that du be restricted from 1.01 to 1.9 as values beyond this search space cannot be probed due to some coefficients in the interference and matrix-element rapidly varying. The choices for Λ_U vary for each du . For spin-0, the Λ_U choices for each du are as follows: $du = 1.1$, $\Lambda_U = 4000, 8000, 10000$ GeV, $du = 1.5$ and 1.9 , $\Lambda_U = 2000, 2500, 3500$ GeV. For spin-2, $du = 1.1$, 1.5 , $\Lambda_U = 2000, 2500, 3000$ GeV, and 1.9 , $\Lambda_U = 2000, 2500, 3500$ GeV, with the mass ranges

M 500 – 1000, 1000 – 2000, 2000 – ∞ for Λ_U less than 3000 GeV. For $\Lambda_U > 3000$ GeV, the mass bins are 500 – 1000, 1000 – 2000, 2000 – 3000 and 3000 – ∞ in units of GeV.

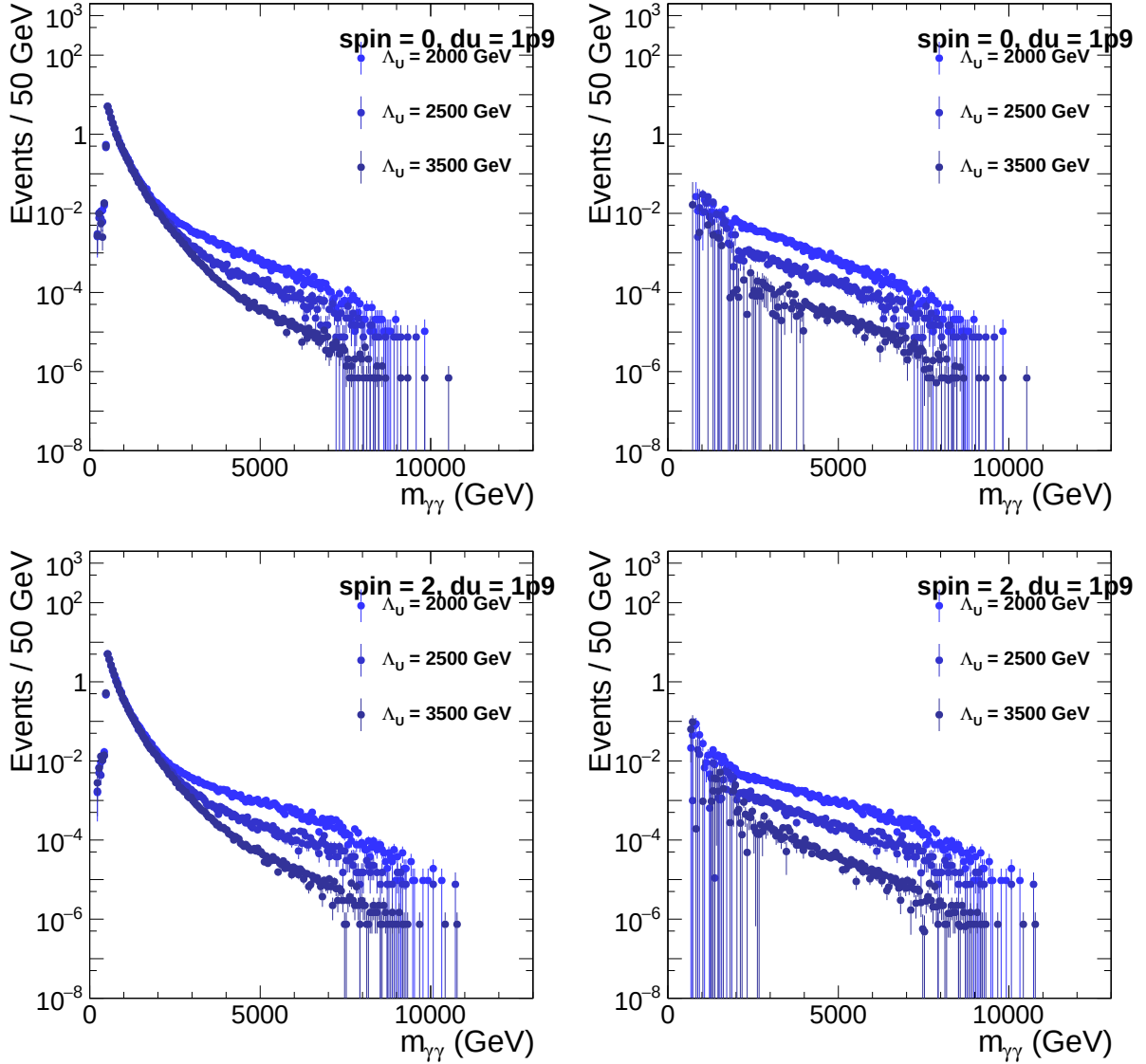
Figure 6.5: The 2D-plane with the parameters of the Unparticles model. The 95% CL upper limits on signal strength (y-axis) vs the Λ_U (x-axis) and the lower limits on Λ_U for the unparticles model. On the left spin 0, on right spin 2. Top to bottom, d_u parameter value: 1.1, 1.5 and 1.9 (Credits to Antonios Agapitos).



We used PYTHIA 8.3 [119] to generate the events with Unparticles decaying into two photons (see Appendix D for a sample PYTHIA card.).

The $m_{\gamma\gamma}$ distributions for the signal samples are displayed in Figure ?? both before and after the subtraction of the SM background.

Figure 6.6: The $m_{\gamma\gamma}$ distribution in unparticle signal events for the generated values of Λ_U . The upper (lower) plots show signals generated assuming spin-0 (spin-2) $du = 1.9$. The left (right) plots show before (after) SM background subtraction. These plots were generated using the 2018 Unparticles samples.



CHAPTER 7

SOURCES OF SYSTEMATIC UNCERTAINTIES

In this chapter we discuss the various sources of systematic uncertainties arising from limitations in the experimental setup as well as the modelling of both the signal and the background. These systematics together with the bin-by-bin statistical uncertainties are considered as nuisance parameters in the fit for the final limit setting. These uncertainties which affect either the shape or normalization of the distributions are summarized in Table 7.1.

Table 7.1: Systematic uncertainties and nuisances considered in non-resonant search.

Name	Number of Nuisances	Shape or Rate	On Signal(S) or BKG(B)	Constrained over SRs
Signal strength	1	Rate	S	6 SRs
Luminosity	1	Rate (1.6%)	S	6 SRs
Normal. real	6 uncorrelated	Rate (5%)	B ()	1 for each SR
Normal. fake $\gamma j, jj$	6 uncorrelated	Rate (30%)	B ($\gamma j, jj$)	1 for each SR
kFactor QCD scales	2, each corr. across years	Shape	B ()	1 for each of BB, BE
Photon efficiency	1 corr. across SRs	Shape	S & B ()	6 SRs
Energy scale gain	1 corr. across SRs	Shape	S & B ()	6 SRs
Energy scale sigma	1 corr. across SRs	Shape	S & B ()	6 SRs
Fake sample	6 corr. across SRs	Shape	B ($\gamma j, jj$)	1 for each SR
kFactors stat.par.0-3	4 each corr. across SRs	Shape	B ($\gamma j, jj$)	6 SRs
PDFs 1-50 hessians	50, each corr. across 6 SRs	Shape	B ()	6 SRs
Pileup	1 corr. across SRs	Shape	S & B ()	6 SRs

The systematic uncertainties are treated introducing nuisance parameters. The first column of Table 7.1 show the names of these sources. The second column presents the number of nuisances together with the correlation information. The third column shows whether the nuisance parameter affects the overall normalization or the shape of the distributions, while the fourth column indicates whether the parameter is constrained for the signal, background or both. Finally, the fifth column shows which signal regions (SRs) the nuisance is constrained over. In this analysis, there is a total of 6 signal regions corresponding to BB,

which are barrel-barrel events or BE, which are barrel-endcap events, for each of the three years: 2016, 2017, 2018.

7.1 Real Background Systematics

Since the Real SM Prompt diphoton background was only performed at NNLO accuracy and higher order terms are unaccounted for, we allow the normalization of the background to float freely, constrained only by the data, at a rate of 5%. This rate also absorbs the systematic uncertainties associated with the integrated luminosity measurement and inefficiencies from the trigger and high- p_T photon ID selection. In total, six nuisances are assigned, one for each of the six signal regions.

Two sources of shape uncertainties on the SM background come from the variation of the K factor scales and the Parton Distribution Function (PDF) uncertainties. We account for the theoretical uncertainty in the factorization and renormalization in QCD scales. This uncertainty is determined by varying the scales used in the K factor calculation, separately for BB and BE. The renormalization and factorization scales are varied simultaneously by a factor of two in each direction, and taking the resultant variation as a measure of the systematic uncertainty on the k-factor. This is found to give an uncertainty varying from $^{+5.40\%}_{-4.3\%}$ ($^{+4.10\%}_{-3.6\%}$) to $^{+7.3\%}_{-5.8\%}$ ($^{+11.20\%}_{-6.9\%}$) for BB (BE) at 0.5 TeV and 2.0 TeV, respectively. In MCFM, these scales affect the differential cross-section calculations through the estimation of the truncation point of the perturbative calculation. Figure 7.4 shows the k-factor variations as well as the ratio of these variations over the default value or nominal value of $m_{\gamma\gamma}$. From these figures, we see that the uncertainty increases with increasing $m_{\gamma\gamma}$.

The PDF uncertainties are obtained from a new feature from MCFM 9.0 which allows us to use PDF weights to calculate uncertainties. MCFM constructs a Hessian matrix which contains information about how variations in the parameters of the PDF model affect the goodness-of-fit to the experimental data, which is a way to estimate the uncertainties associated with the PDF. In particular, the PDF uncertainties on the real background $m_{\gamma\gamma}$ are

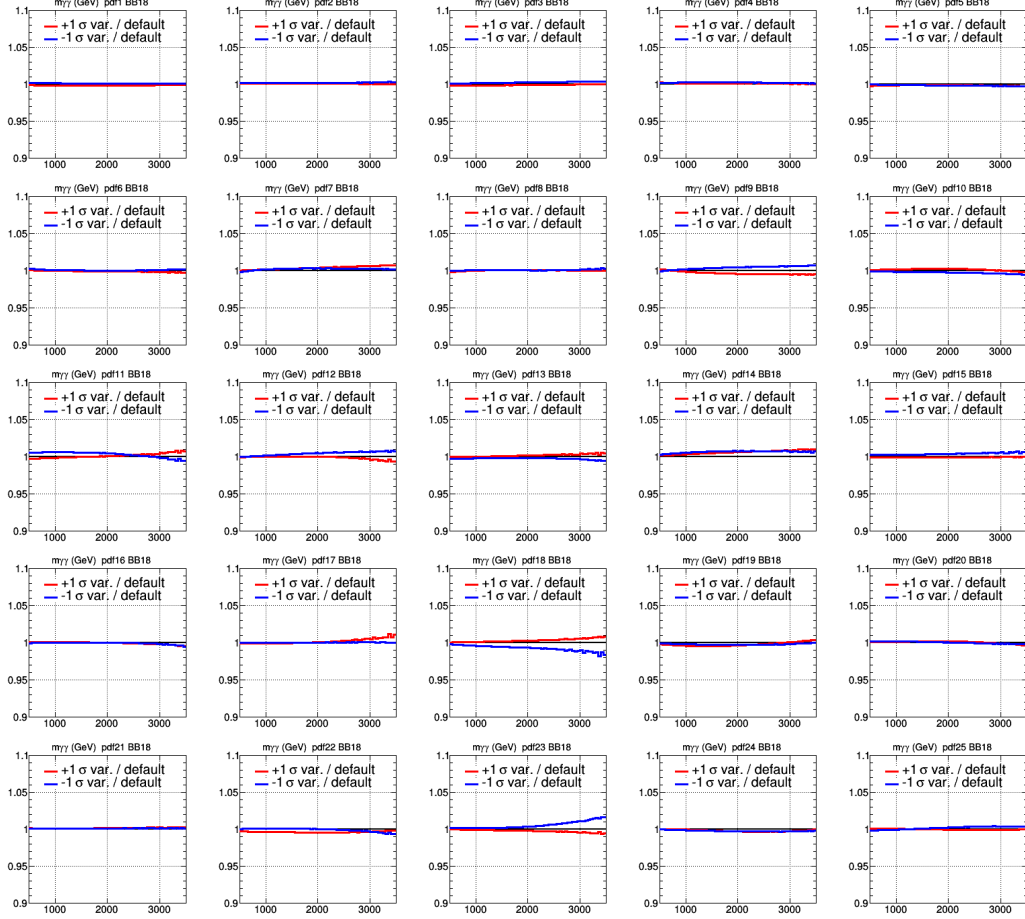
estimated using the `NNPDF30_nnlo_as_0118_hessian` PDF set. By construction, this PDF set has the same central values as the `NNPDF30_nnlo_as_0118` set used in the SHERPA MC samples and the MCFM cross-section calculations with uncertainties provided by eigenvectors rather than PDF replicas. This procedure simplifies the inclusion of PDF uncertainties as the eigenvector variations are uncorrelated, in contrast with the standard PDF set. Fifty (50) eigenvectors are provided in the Hessian PDF set and these are shown in Figure 7.1 and Figure 7.2. These plots are BB plots showing the resulting uncertainties from the 50 variations of the PDFs eigenvectors which refer to variations in the sets of PDFs obtained from global fits to experimental data. MCFM 9.0 allows the evaluation of uncertainties due to individual eigenvectors and in principle should allow for calculating uncertainties on the full NNLO cross section. In practice, we found that it was impossible to complete a significant number of the NNLO jobs within the maximum allowed CPU/wall clock time limits. These eigenvectors represent directions in the parameter space of the PDF model. Figure 7.3 shows the uncertainty on the real background arising from the quadratic combination of all 50 individual variations.

The exact treatment of the PDFs uncertainties are described as follows. For each of the 50 eigenvector variations, a pair of different $m_{\gamma\gamma}$ spectra is derived. These pairs correspond to ± 1 variations of eigenvectors. A nuisance parameter with a gaussian prior pdf for each pair is assigned. Each of these is constrained simultaneously over all the six signal regions. We observe that the resulting shape uncertainties are small and affect less than 2% the spectra for $m_{\gamma\gamma} < 3.5$ TeV.

Uncertainties due to the photon energy scale and resolution shown in Figure 7.6 have also been considered in the analysis. For each of these uncertainties, two different $m_{\gamma\gamma}$ shapes are formed (corresponding to the plus and minus shifts) for each of the six signal regions. We assign a single nuisance parameter for each which is constrained over all six SRs. For the “energy scale statistical” and the “energy sigma” there were neither shape nor rate effect in the variations and so these source of uncertainty has been omitted to preserve computing

resources.

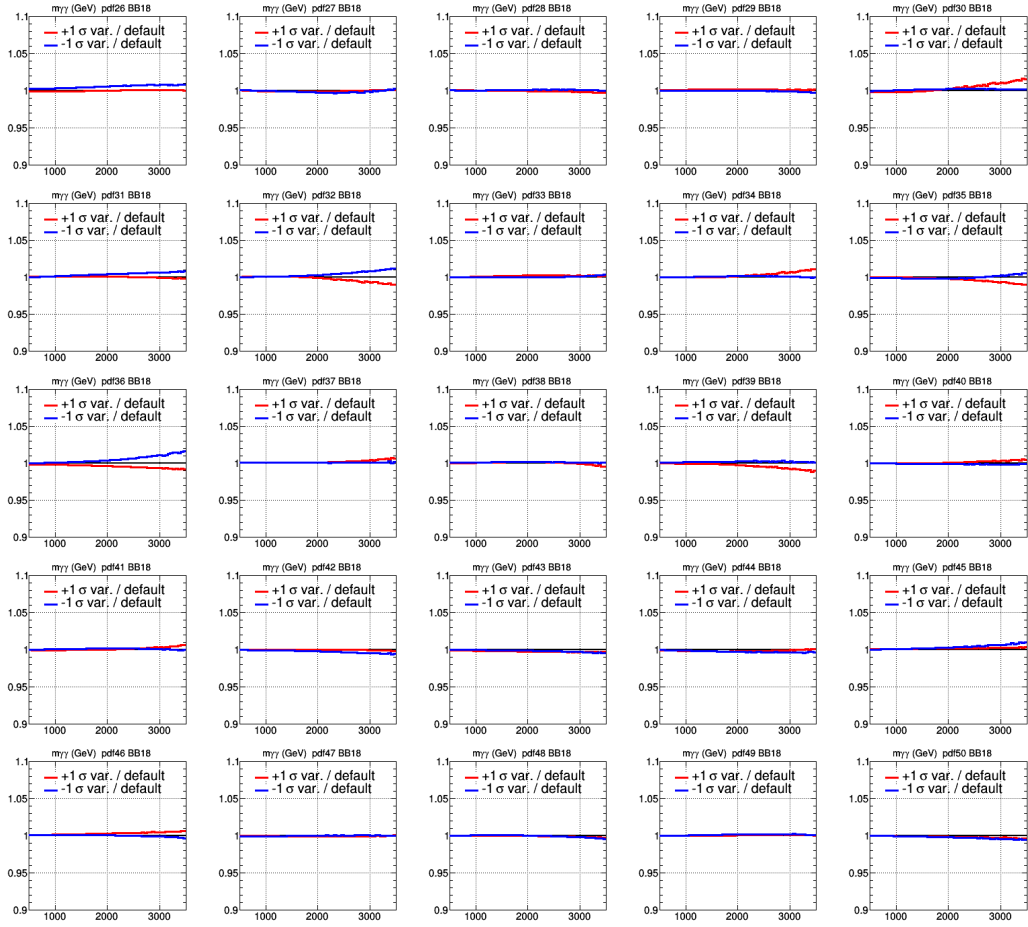
Figure 7.1: The resulting uncertainties from first 25 of the 50 variations of the PDFs eigenvectors. The variation over central estimate is shown versus the $m_{\gamma\gamma}$. (Plots shown stand for BB, while the corresponding for BE, as they are very similar, have been omitted).



7.2 Fake Background Systematics

We assign a 30% systematic uncertainty to the normalization of the fake $g\bar{g}$ background covering the degree of variation in the fake rate in bins of photon η , number of primary vertices, pileup, and the choice of the sideband definition and template variable. Six different nuisances have been assigned, one for each of the six signal regions, with a 30% magnitude and a lognormal prior. Figure 7.5 shows the effect of the "fake sample" systematic uncertainty listed in Table 7.1. We've decided to drop the $j\bar{j}$ contribution since it is 2 orders of magnitude

Figure 7.2: The resulting uncertainties from the last 25 of 50 variations of the PDFs eigenvectors. The variation over central estimate is shown versus the $m_{\gamma\gamma}$. (Plots shown stand for BB, while the corresponding for BE, as they are very similar, have been omitted).



smaller than the gj contribution. The 30% rate also covers the non-closure seen in the fake rate closure test, and the differences between the fake rate derived from the jet-triggered and muon-triggered data samples. The jet-triggered and the muon-triggered data samples resulted in slightly different fake rate estimates and the mean of the two was taken as the central value for the fake rate estimate. The maximum and minimum extremes were used to assign two (plus and minus) alternative predictions for the fake background. This source of uncertainty accounts for the difference in the quark and gluon contents of the two data sets. The details of these studies are found in Sec. 5.3 and Sec. 5.4.

7.3 Signal Systematics

The signal distributions are normalized to the integrated luminosity. We assign a rate of 1.6% systematic uncertainty related to the uncertainties in the measurement of the integrated luminosity. The photon selection efficiency, energy scale and resolution, as well as the effect of pileup have been considered as shape-based uncertainties.

Figure 7.3: The uncertainty on the real background that arises from PDFs (parton distribution function) uncertainties, displayed separately for the barrel-barrel and barrel-endcap regions, for evaluations at LO (left) and NLO (right).

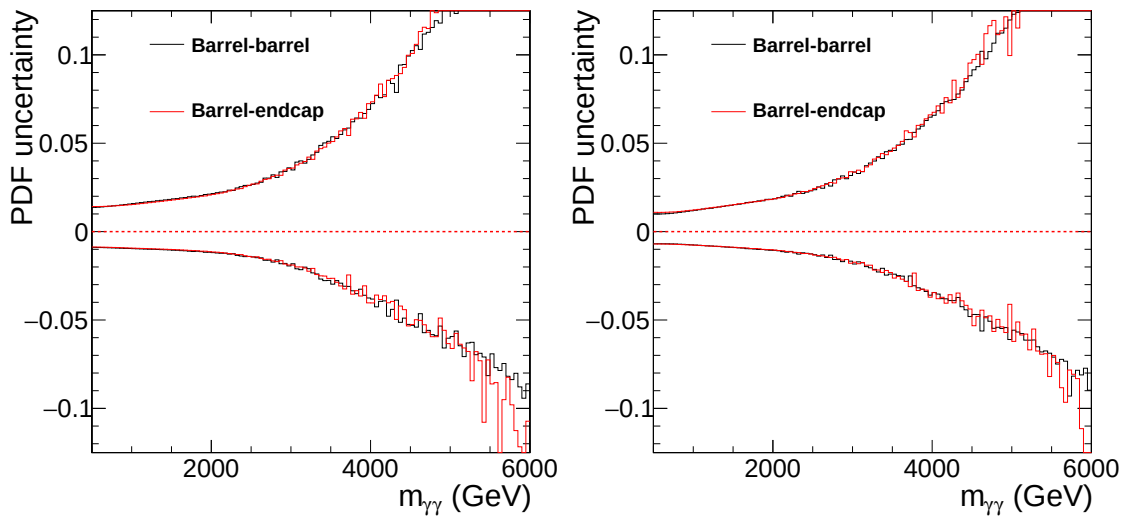


Figure 7.4: Top: NNLO (MCFM) k-factors for different factorization and renormalization (QCD) scales. Bottom: Ratio of these variations over the central value as a function of $m_{\gamma\gamma}$. Plots for the BB (barrel-barrel) on the left, and BE (barrel-endcap) on the right.

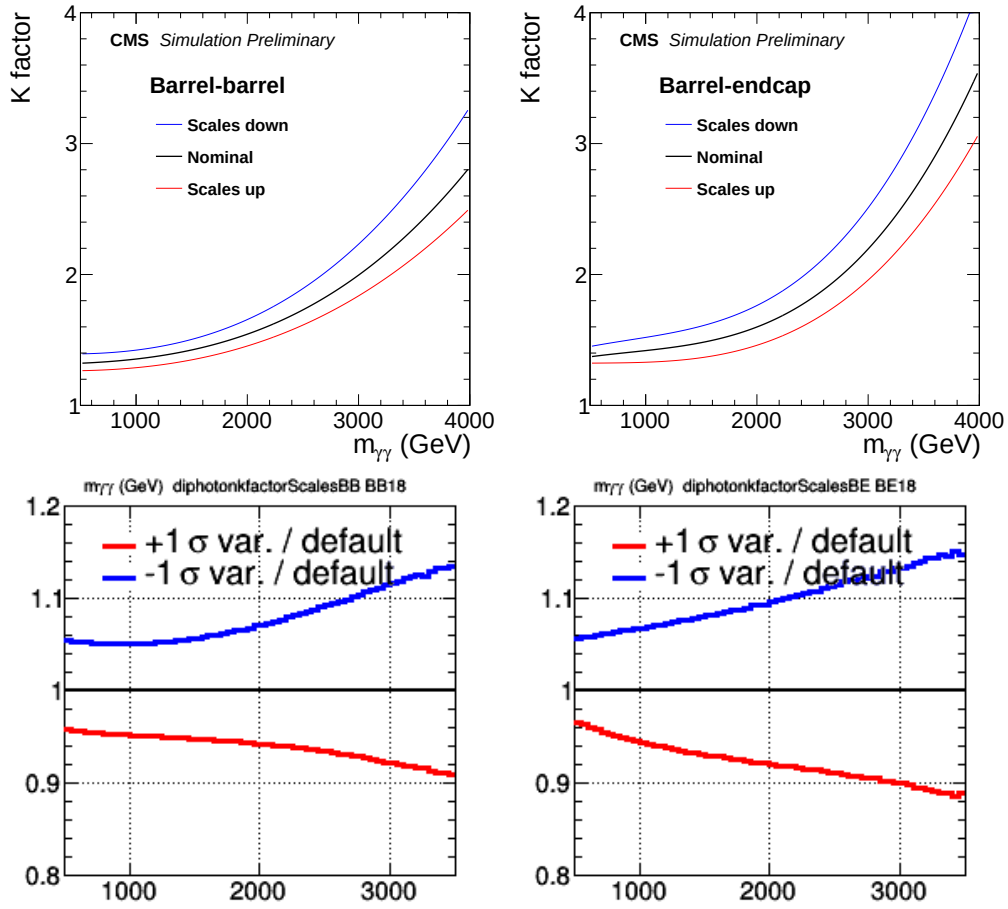


Figure 7.5: The effect of “fake sample” systematic uncertainty over the fake background $\gamma j, jj$ vs the $m_{\gamma\gamma}$ spectra. Top (bottom) the BB (BE) channels, left to right for years 2016, 2017, and 2018 respectively.

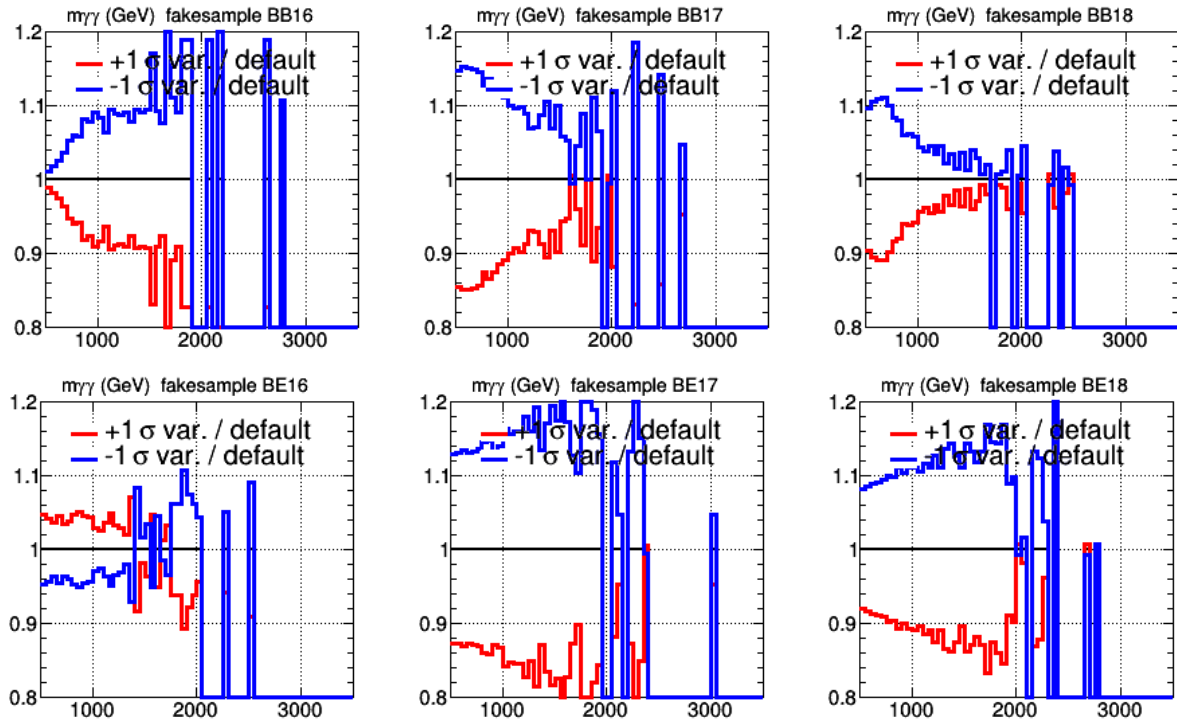
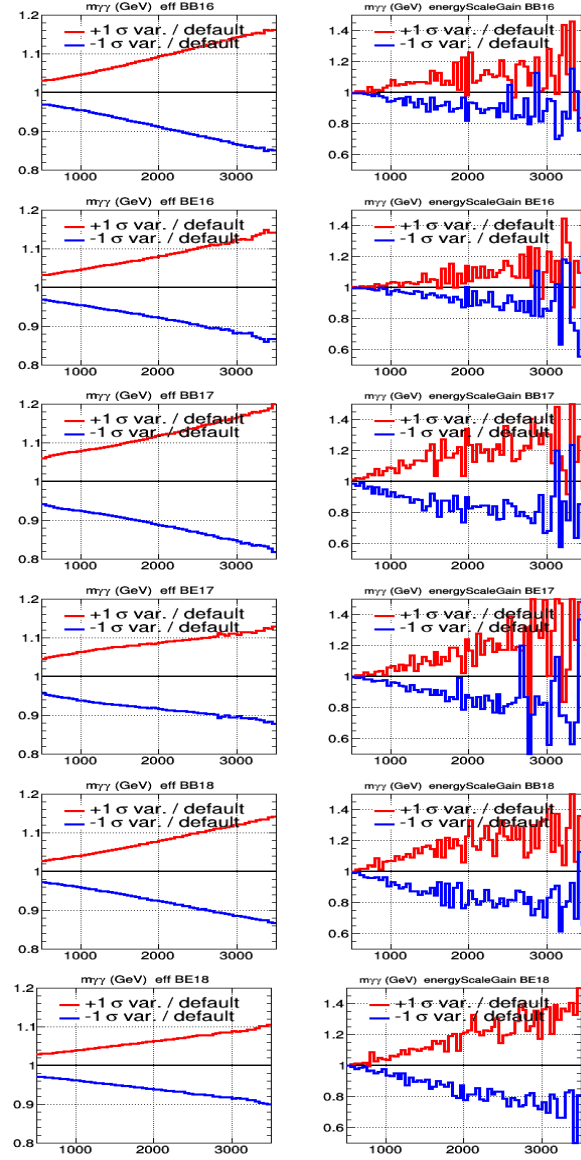


Figure 7.6: The left shows the systematic uncertainty trends of “variation/central” ratios of real background vs $m_{\gamma\gamma}$ for the sources of: “efficiency”, and right for “energy scale gain”. Top to bottom: the six SRs of BB16, ..., BE18 respectively.



CHAPTER 8

RESULTS

This chapter discusses the results comparing the compatibility of the observed data with the background-only versus various signal-plus-background hypotheses using a shape-based Bayesian Marginalization procedure for the ADD Large Extra Dimensions and the Clockwork Framework. The Unparticles model uses a simpler cut-and-count approach (rather than a shape-based approach), with the modified frequentist method, also known as the CLs method, is used according to the prescriptions of Ref. [138], together with the use of asymptotic formulas [139].

8.1 Bayesian Confidence Intervals

A bayesian statistical approach was used to constrain the signal parameters of the ADD Large Extra Dimensions and the Clockwork Framework using the **THETA** framework [140], with a flat prior on the signal strength assumed and constrained to be positive. By convention in particle physics, a flat prior constrained to be positive is motivated by our lack of prior knowledge on the signal uncertainty and the fact that this is a shape analysis. It is often considered a good choice due to simplicity, ease of implementation, and avoids the introduction of an additional assumption in the analysis. **THETA** uses a Markov-Chain method to marginalize the nuisance parameters. In the next parts of this text, we briefly introduce Bayes theorem, and the inputs used to set the limits on the aforementioned models.

Bayes theorem provides a formal mechanism for updating the probability of a hypothesis in light of new evidence which corresponds to an experimental measurement. Mathematically,

it is expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \quad (8.1)$$

Here, $P(A|B)$ is the posterior probability of the hypothesis A given B, which in this case is the data. $P(B|A)$ is called the “marginal likelihood” which indicates the probability of the event/data being true, given A. $P(A)$ corresponds to our knowledge, or what is called the *prior*, which is the measure of the probability of A being true. $P(B)$, is the marginal likelihood, which is the probability of the data being true.

For the shape-based analysis, a binned likelihood is constructed where the total likelihood $\mathcal{L}$ is equal to the product of the individual likelihoods $\mathcal{L}_i$ in each bin i . Since we count the number of events occurring in each bin, the likelihood in each bin can be written using the Poisson Distribution formula,

$$\mathcal{L}(N|\mu_i) = \prod_i \mathcal{L}_i(N_i|\mu_i) = \prod_i \frac{\mu_i^{N_i} e^{-(\mu_i)}}{N_i!}, \quad (8.2)$$

where the index i runs over all the bins of the histogram. N_i is the number events in the bin i and μ_i is the sum of the signal strength β times the number of signal events in bin i plus the number of background events b_i in the same bin:

$$\mu_i = \beta s_i + b_i. \quad (8.3)$$

We can now rewrite Eq. 8.1 as follows,

$$P(\beta, N) = \frac{\mathcal{L}(N_i|\mu_i(\beta)\pi(\beta))}{P(N)} = \mathcal{L}(N_i|\mu_i)\pi(\mu_i). \quad (8.4)$$

Treating $P(N)$ as a normalization constant absorbed into the likelihood shows that the posterior is simply proportional to the likelihood.

Both the signal and background have associated systematic uncertainties that affect

our knowledge of their associated yields. We assign a separate nuisance parameter θ for each of these systematic uncertainties which are detailed in Table 7.1. For each of these nuisance parameters, a different choice of prior $\pi(s, b, \theta)$ is made. These constraints allow the signal and background yields to deviate from their nominal values, subject to some cost or penalty on the likelihood in the form of a normal or lognormal distribution. These individual parameters are described in Sec. 7. They are either assumed to be fully correlated across the $m_{\gamma\gamma}$ bins across all the 6 signal regions, i.e. the 3 years for BB and BE. Otherwise, separate nuisance parameters are used for each signal region.

We use constant bin widths of 100 GeV for both the signal and the background $m_{\gamma\gamma}$ distributions. These, together with the nuisance parameters and additional shape-based uncertainties, are extracted from `root` files to construct the statistical model. The framework then performs a maximum likelihood fit to get estimates for all model parameters. At this step a simultaneous fit to the $m_{\gamma\gamma}$ spectrum in six event categories (EBEB) and (EBEE) for each of the three years of data collection, is performed along with the marginalization of the nuisance parameters. This produces updated posterior predictions for the fake and prompt diphoton background distributions.

Mathematically, the marginalization procedure for the nuisance parameters is done by the following integration:

$$P(\beta|N) = \int P(\beta\theta|N_i)d\theta = \int \mathcal{L}(n|\beta, \theta)\pi(\beta, \theta)d\theta, \quad (8.5)$$

where $\pi(s, b, \theta)$ is the chosen prior for each nuisance parameter.

A commonly employed statistical method to express confidence is through the computation of a parameter's upper or lower limit at a 95% confidence level. This involves excluding any hypotheses where the parameter surpasses (or falls below) a predefined threshold. For determining the 95% upper limit on a signal hypothesis, characterized by a signal strength denoted as β , we integrate the posterior prediction up to 95% of the total integral, as illus-

trated below:

$$\int_0^\beta P(\beta|N) = 0.95 \times \int_0^\infty P(\beta|N). \quad (8.6)$$

8.2 Exclusion Limits: ADD and Clockwork

We employ the general bayesian marginalization procedure described in the previous section on the ADD and the Clockwork models. For the ADD interpretation, the $m_{\gamma\gamma}$ distribution begins at 500 GeV while the threshold was raised to 1 TeV in the clockwork mechanism to avoid statistical fluctuations in the template construction described in Ch. 6.

8.2.1 Prefit Prediction

The unblinded spectra of $m_{\gamma\gamma}$ are shown in Figs.8.1,8.2, and 8.3, together with the prefit prediction. The predictions for the real $\gamma\gamma$ background and the fake $\gamma j, jj$ background are represented by light- and dark-filled histograms, respectively. We observe that the fake background, in the range of high sensitivity $m_{\gamma\gamma} > 1$ TeV, amounts approximately only 5% (15%) for BB (BE) channels. The bottom pads present the pulls - the difference between the data and the MC prediction divided by the standard deviation σ_{STAT} , as well as the magnitude of prefit rate systematic uncertainty, σ_{SYS} , where for the latter only the 5% of real and the 30% of the fake backgrounds uncertainties on rate have been considered.

The data exhibit over all a fair agreement with the prefit background estimate withing the systematics uncertainties, while in some case (for low masses) the data are slightly off the $\pm 1\sigma$ uncertainty band for a few sequential bins. The spectra are fit within the range of 500–4000 GeV, where the last bin includes the overflow events. Additionally, the $m_{\gamma\gamma}$ background normalization is allowed to float freely within 5% in the BB and BE categories separately. The low-mass data will easily constrain the nuisance parameters.

It is important to clarify at this point that the NNLO k-factor-corrected diphoton estimate together with the data-driven fake background estimate do not correspond to the final prediction of the total SM background. Instead, the final prediction is derived from the fit

Figure 8.1: The $m_{\gamma\gamma}$ spectra in the prefit estimate for 2016; left the BB and right the BE channels. These are the six observables used in the fit. The final prediction is derived from a fit to the data, a procedure which absorbs residual mismodeling present at these plots.

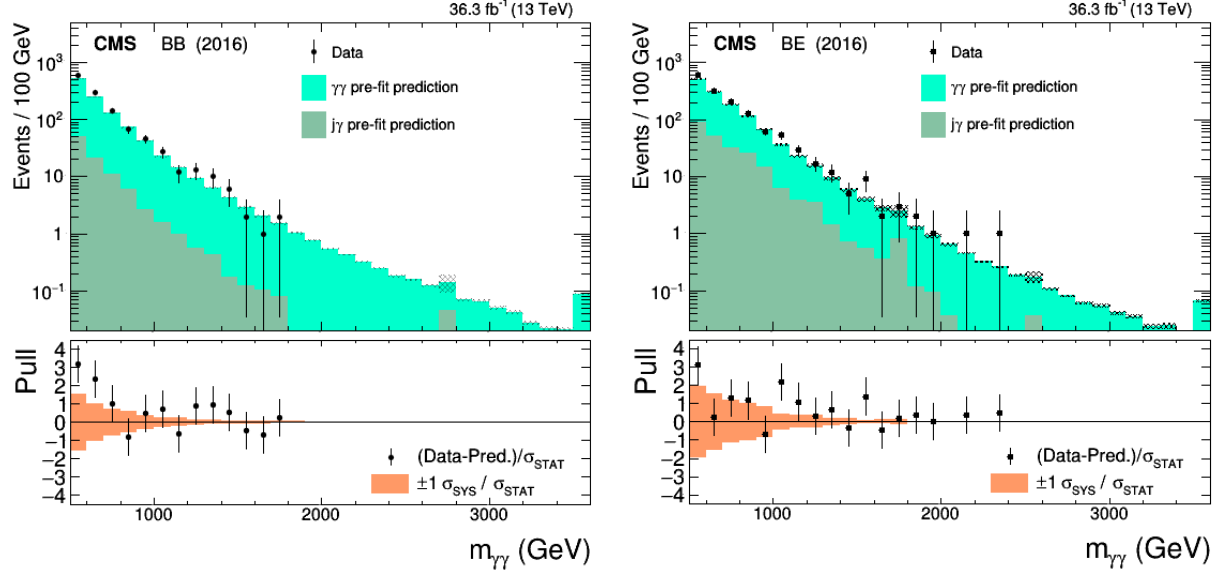


Figure 8.2: The $m_{\gamma\gamma}$ spectra in the prefit estimate for 2017; left the BB and right the BE channels. These are the six observables used in the fit. The final prediction is derived from a fit to the data, a procedure which absorbs residual mismodeling present at these plots.

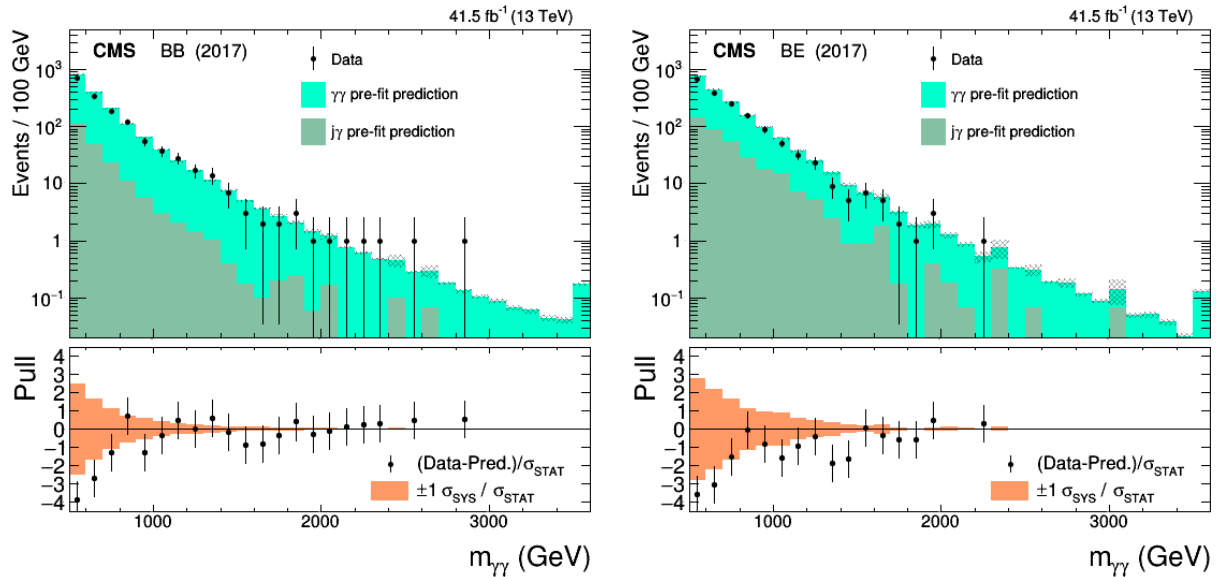
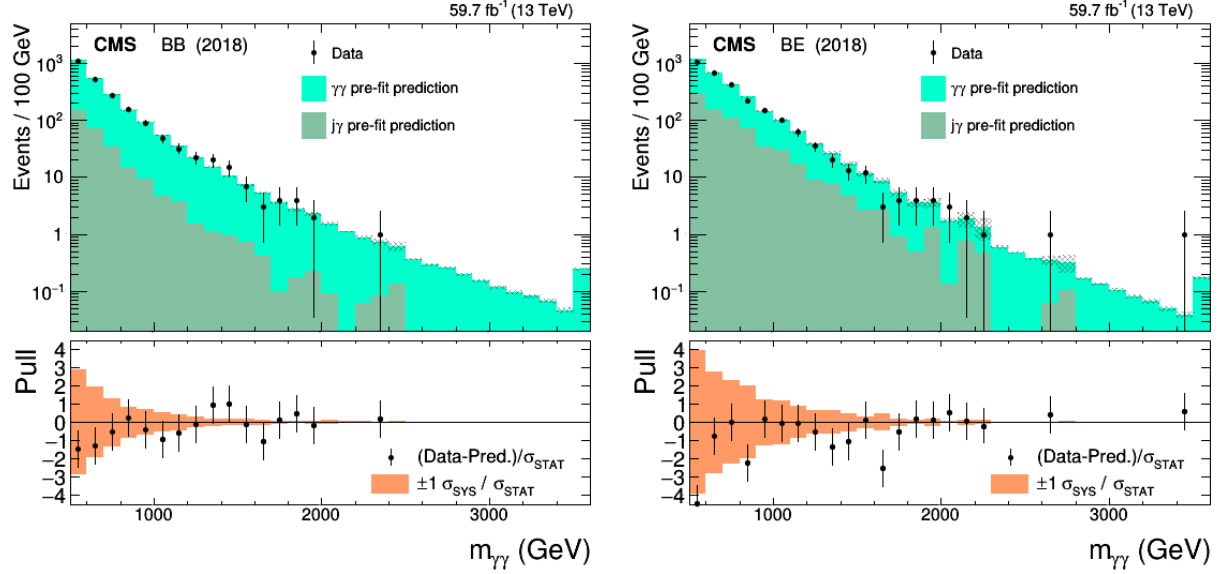


Figure 8.3: The $m_{\gamma\gamma}$ spectra in the prefit estimate for 2018; left the BB and right the BE channels. These are the six observables used in the fit. The final prediction is derived from a fit to the data, a procedure which absorbs residual mismodeling present at these plots.



which simultaneously constrains nuisances and the signal strength. Additionally, the $m_{\gamma\gamma}$ spectra presented in Figs. 8.1, 8.2, and 8.3 are the six observables used in the fit and termed as the prefit prediction together with all systematic uncertainties discussed in the previous Ch. 7. The nuisance parameters are then marginalized with the THETA framework to derive posterior predictions for the $m_{\gamma\gamma}$ spectra and signal strength, from which we can derive the 95% CL (Confidence Level) exclusion limits on the BSM models probed.

8.2.2 Limits and Postfits

After the marginalization procedure, THETA returns postfit distributions, the nuisance postfit pulls, and the limits results for the models in question. In the next few pages, we will discuss more details on the shape-based analysis for the CW and ADD, respectively. In this shape-based approach there is a presupposition that the signal and background differ in their *shape* and not only in the rate. A benefit from this procedure is that we are able to absorb and exploit higher-order corrections to the SM diphoton background prediction and their differences with the signal models.

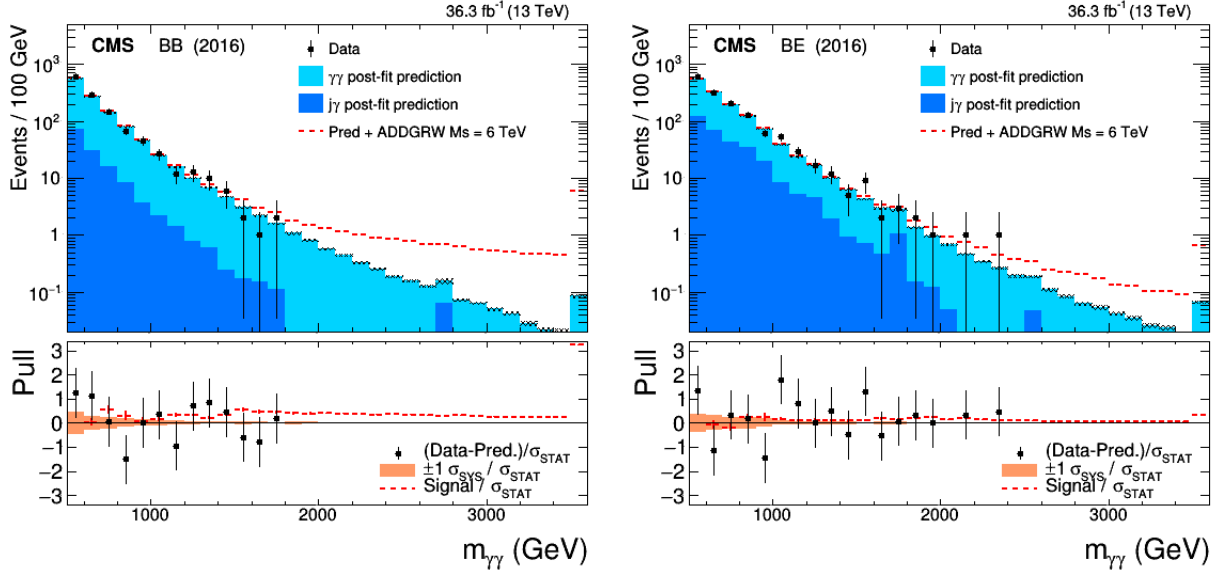
Note that prior to unblinding the $m_{\gamma\gamma}$ region, we made use of pseudodata to run the

full THETA limits machinery code and produce the expected results. The pseudodata are generated using the sum of the real $\gamma\gamma$ and fake γj and jj spectra, which are the prefit background estimates, are used as the mean for poisson random number (see: TRandom) generator from ROOT. In this way, the pseudodata feature realistic systematic structures and fluctuations and we can check whether prefit pull trends are covered by our current parametrization. The results of the pseudodata used are found in Appendix E.

8.2.3 ADD Postfit results

The ADD postfit results are shown in Figs. 8.4, 8.5, 8.6, where we use light and dark blue for the real and fake background estimates respectively. We observe from the bottom pads that various systematic trends initially present in prefit pulls are smoothened out in the postfit version. The post fit pulls exhibits only statistical fluctuations around one, and thus the $m_{\gamma\gamma}$ parametrization absorbs the systematic effects. Figure 8.7 illustrates prefit and postfit results for all three years of data taking inclusively for the ADD scenario.

Figure 8.4: The ADD GRW postfit spectra (with real data) for BB (left) and BE (right) for 2016.



The nuisances are constrained and their priors which are either gaussian or lognormal probability distribution function transforms into their corresponding postfit version. While

Figure 8.5: The ADD GRW postfit spectra (with real data) for BB (left) and BE (right) for 2017.

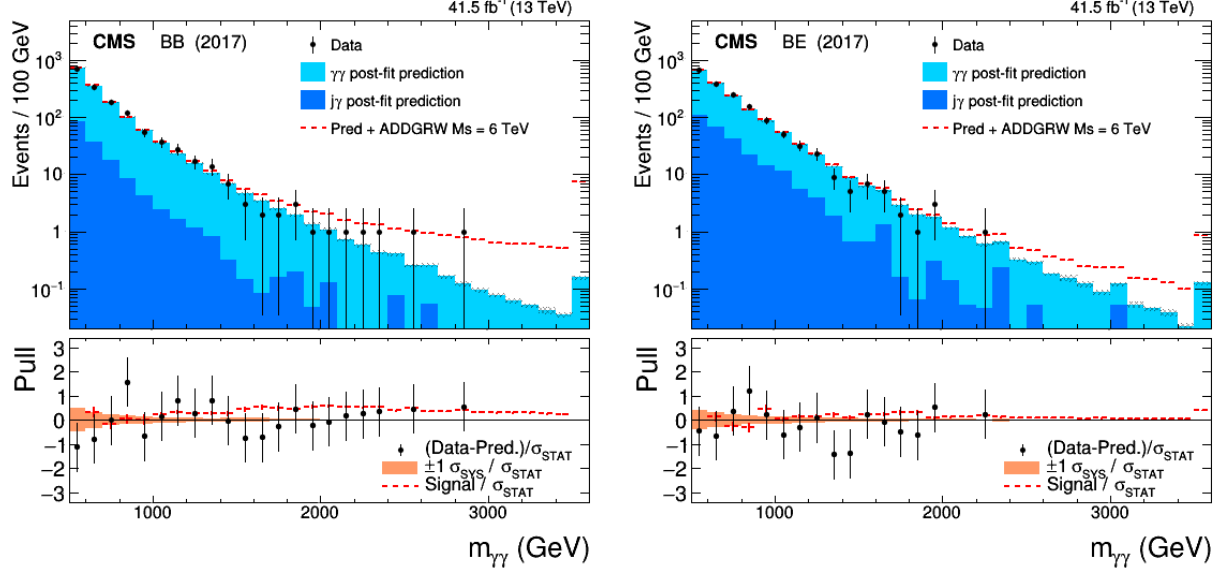


Figure 8.6: The ADD GRW postfit spectra (with real data) for BB (left) and BE (right) for 2018.

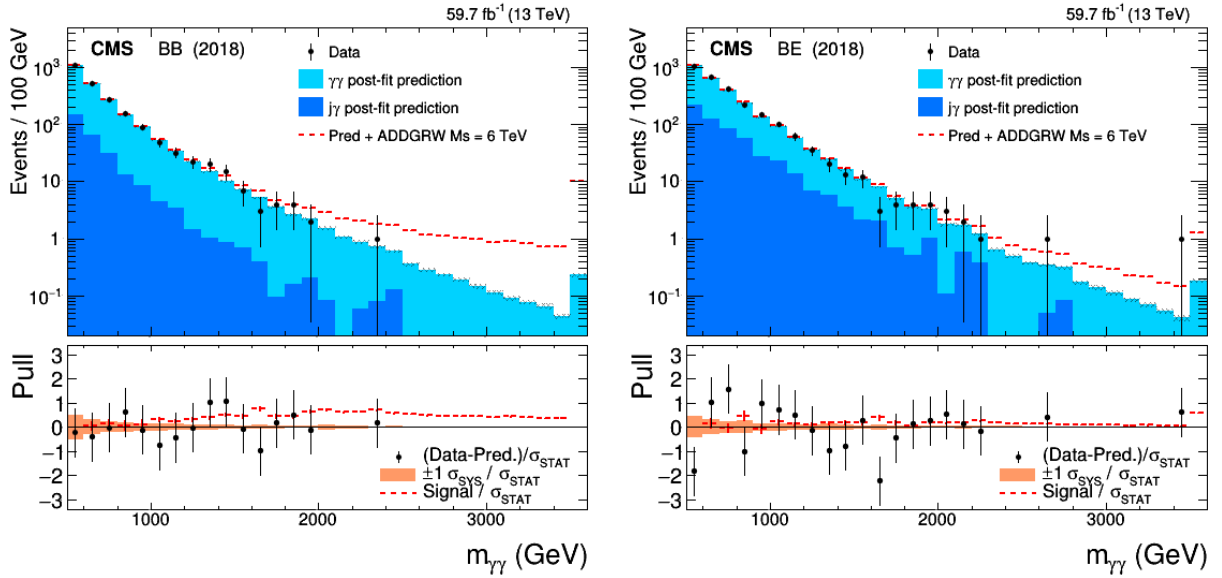
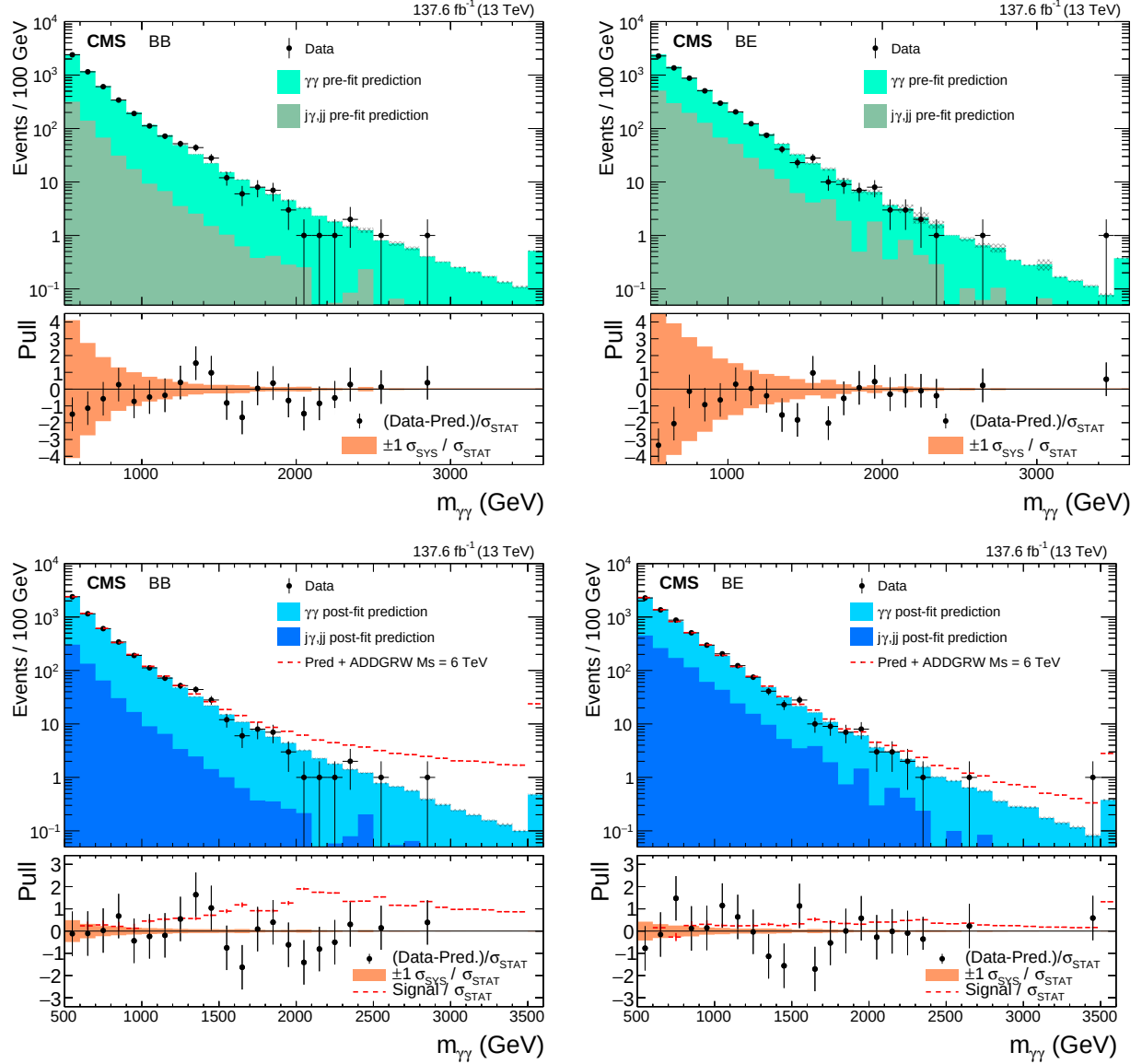


Figure 8.7: The prefit (top) and the postfit (bottom) spectra with real data for BB (left) and BE (right) for the three years of data taking inclusively. The plot shown in red at the bottom shows a representative signal plot for the ADD model with the GRW convention.



COMBINE provides summary tables with the means and the standard deviations of the nuisances postfit pdfs, THETA provides the full postfit pdfs spectra of each nuisance. The postfit probability distribution function (pdfs) of nuisances are presented in Figure 8.8. Each plot corresponds to the sources (nuisances) presented of the table 7.1 (names in labels). The statistics panel of each plot contain the mean and the standard deviation of the pdf. Text indicates the name of each nuisance, and the parenthesis after that indicates the regions in which this is constrained over. This information can also be found in table 7.1. In this figure, we see that signal strength named `beta_signal` on the top-left, is strongly constrained near zero for GRW, $M_S = 6$ TeV as expected. Normalization nuisances of real and fake $\gamma j, jj$ background events are constrained and their postfit pdfs are close to zero or some other value dictated by the fit on the data correcting prefit trends. The Postfit nuisances corresponding to the 50 pdfs of the PDFs nuisances; are all close to the prefit, that is it is a gaussian with mean zero and standard deviation of one, as this set of data can not constrain them further.

8.2.4 ADD Limits

For the ADD model, we have the positive and the negative interference conventions which have significant shape differences. These are referred to as the GRW and Hewett- conventions respectively. We ran the limit code for all available Λ_T values, which are equivalent to M_s depending on the model convention. THETA then provides us the upper limit on the signal strength at 95% CL for each M_S scenario. Instances of M_s with postfit signal strength < 1 are excluded at 95% CL. The results for the ADD GRW model after fitting all six signal regions from BB16,..., BE18 are presented in Figure 8.9.

From Figure 8.9, we can read the limits on the M_S parameter as intersections with the horizontal line, where the signal strength is equal to 1. The green and yellow bands indicate the $\pm 1\sigma$ and $\pm 2\sigma$ standard deviation bands, respectively. The limits results from the GRW and Hewett- conventions could be translated to other ADD conventions as explained in Section 6.1. These are the HLZ conventions of positive interference for $n_{ED} = 3, 4, 5, 6, 7$ which are obtained by scaling the GRW and the Hewett+ is obtained by scaling the Hewett-

Figure 8.8: The postfit probability distribution function (pdf) of each nuisance used for the fits for each of the 6 signal regions that have been fitted. The upper right panel corresponds to the statistics panel indicating the mean and standard deviation of the pdfs.

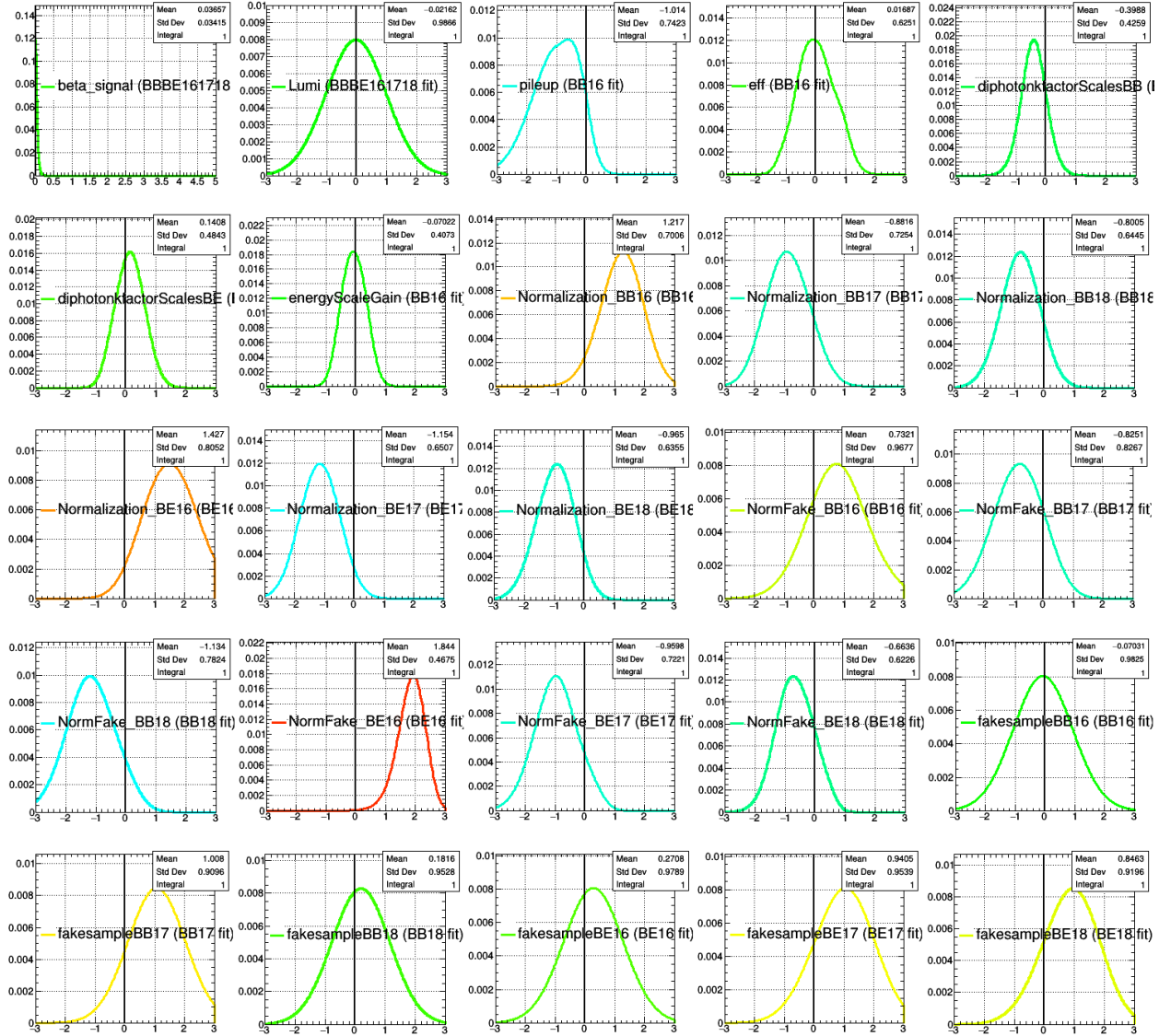
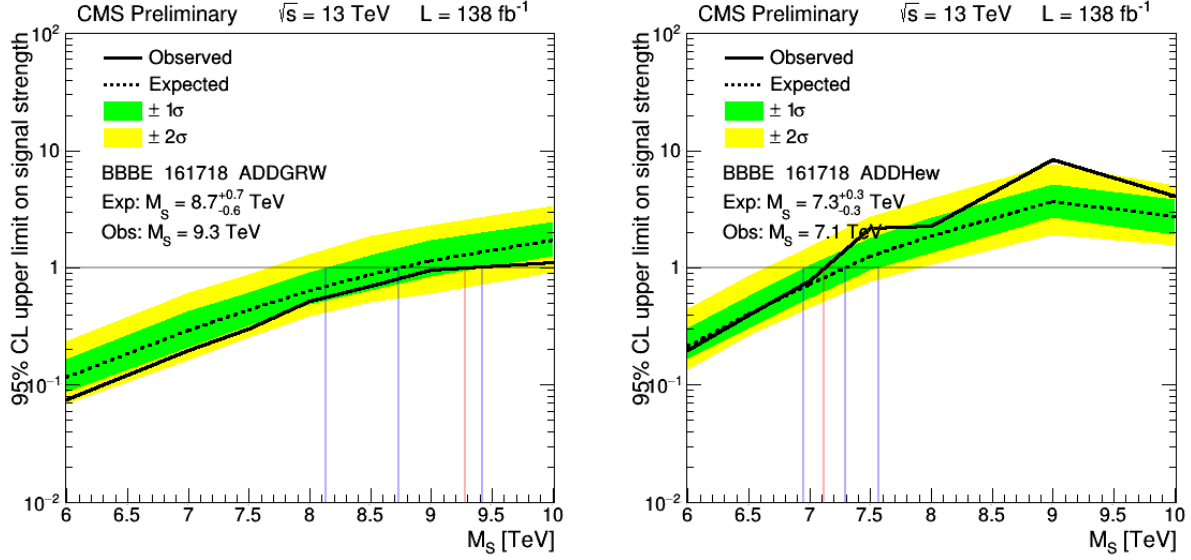


Figure 8.9: The 95% CL limits on signal strength (y-axis) vs the M_S (x-axis) and the limits on M_S in legend for the ADD GRW model left, and Hewett negative right. These results are using real data.



results. Table 8.1 presents the lower limits on M_S scale for all available ADD signal scenarios. The total asymmetric uncertainties are appended to the expected limits. The limit code was previously tested with pseudodata and the results were consistent with the 2016 data search. With the unblinded results, we have set the most stringent limits up-to-date on the ADD interpretation in the diphoton channel using the full Run-2 CMS data from 2016–2018.

The different trends that GRW and Hewett exhibit are artifacts that the former model predicts an excess while the latter deficit of events in high masses. As the data in the BB region where we are most sensitive indicate small deficit with respect to SM postfit estimate, observed limits behave this way in accordance with our expectations.

8.2.5 CW Limits

The strategy for limit setting on the clockwork model (CW) is similar to the ADD model as described in the previous section. We use the data $m_{\gamma\gamma}$ spectra to constrain the resonant excess due to the clockwork model by marginalizing over the aforementioned nuisance parameters assuming a flat prior for the signal strength. One difference with the ADD model is that for the CW we can only use the spectra $m_{\gamma\gamma} > 1000$ GeV. Since the CW

Table 8.1: Full CMS Run II exclusion limits on the mass scale M_S (in units of TeV) for various conventions used in the calculation of the ADD large extra-dimensional scenario using the 2016-2018 CMS detector data corresponding to an integrated luminosity of 138 fb⁻¹.

Signal	GRW	Hewett		HLZ				
		negative	positive	$n_{\text{ED}} = 3$	$n_{\text{ED}} = 4$	$n_{\text{ED}} = 5$	$n_{\text{ED}} = 6$	$n_{\text{ED}} = 7$
Expected	$8.7^{+0.7}_{-0.6}$	$7.3^{+0.3}_{-0.3}$	$7.8^{+0.6}_{-0.5}$	$10.3^{+0.8}_{-0.7}$	$8.7^{+0.7}_{-0.6}$	$7.9^{+0.6}_{-0.5}$	$7.3^{+0.6}_{-0.5}$	$6.9^{+0.6}_{-0.5}$
Observed	9.3	7.1	8.3	11.1	9.3	8.4	7.8	7.4

distributions are combinations of the positive and negative interference ADD samples, we are constrained by the samples that are readily available. Below $m_{\gamma\gamma} > 1000$ GeV, statistical fluctuations dominate such that meaningful conclusions could not be made.

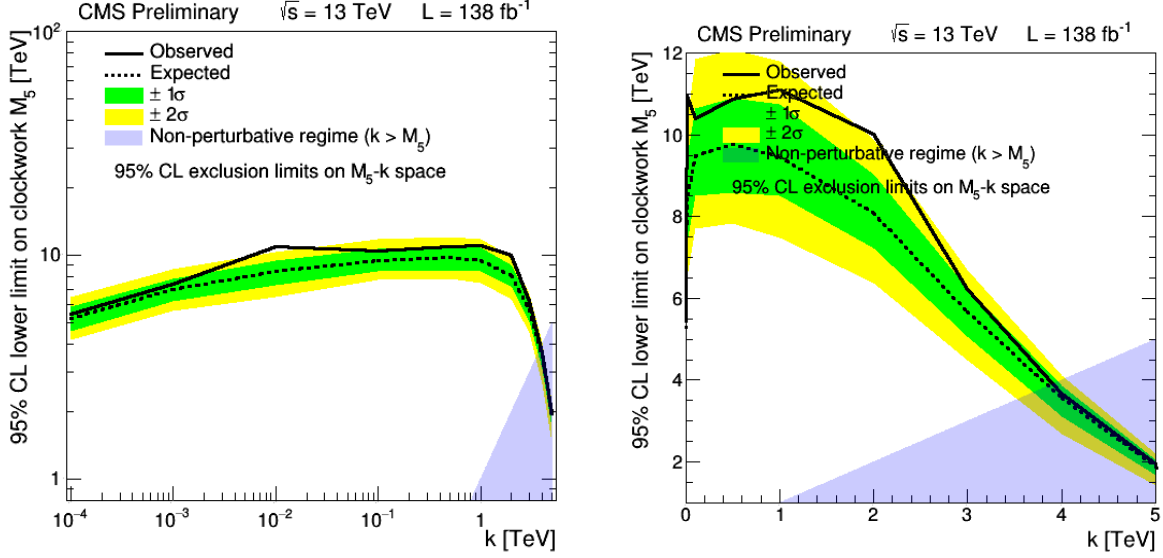
Eq. 6.2 shows that the CW signal strength normalization scales as M_5^{-3} , hence the upper limit on the signal strength can be directly translated as a lower limit on M_5 for a fixed k . Figure 8.10 shows the updated 95% CL exclusion in the k - M_5 plane with the full Run II data.

The signal shapes are generated for values of k as low as 0.1 GeV and up to 5 TeV. The $k > M_5$ space is non-perturbative so we denote this region of phase-space with blue color in the k - M_5 plane.

Figure 8.11 presents the $m_{\gamma\gamma}$ spectra in the fitted range (> 1 TeV) in their prefit and postfit ranges. In these plots we also present the plain signal with the background unsubtracted so its shape is visible. We show the $k = 1$ TeV scenario in this plot which has a turn on at around 1 TeV, thus the bulk of the signal events coincides with the bulk of the background. Similarly for $k < 1$ TeV, the signal events are spread in low masses and the sensitivity degrades as the shape difference with the SM background is small. For signal with $k \geq 2$ TeV the bulk of signal events are in high masses and in the overflow $m_{\gamma\gamma}$ bin. The sensitivity is then limited by the statistics at the tails mostly at BB.

Figure 8.12 present the postfit pdfs for the 25 nuisance parameters used in the six SRs

Figure 8.10: Exclusion limit for the clockwork framework over the $k - M_5$ parameter space. The shaded region denotes where the theory becomes non-perturbative. The right plot displays a non-logarithmic axis.



fit with CW model as signal. We observe that only a few of them are pulled to correct the prefit trends. The pull effects are milder compared to the ADD as we fit $m_{\gamma\gamma} > 1000$ GeV where the prefit is already in good agreement with the data.

8.2.6 Comparing CW results with ATLAS

The CMS clockwork exclusion results shown in Figure 8.13 show Run 1 results which excludes values of M_5 lower than 5 TeV for k values in the range of 0.2 GeV to 2.0 TeV. With increased data in Run II, the limits were pushed further with values of M_5 lower than 5.5 TeV for k values from 0.1 to 2 TeV, and M_5 values lower than 11 TeV[8 TeV expected] to 1 TeV for k values from 100 GeV to 5 TeV. ATLAS [73], have published their results on the CW model using a different signal generation and limit setting techniques. Figure 8.14 show that the ATLAS limits results without mass threshold is comparable results with ours.

Figure 8.11: The CW prefit (top two rows) and the postfit (bottom two rows) $m_{\gamma\gamma}$ spectra. Columns left to right correspond to the years 2016, 2017, and 2018 respectively.

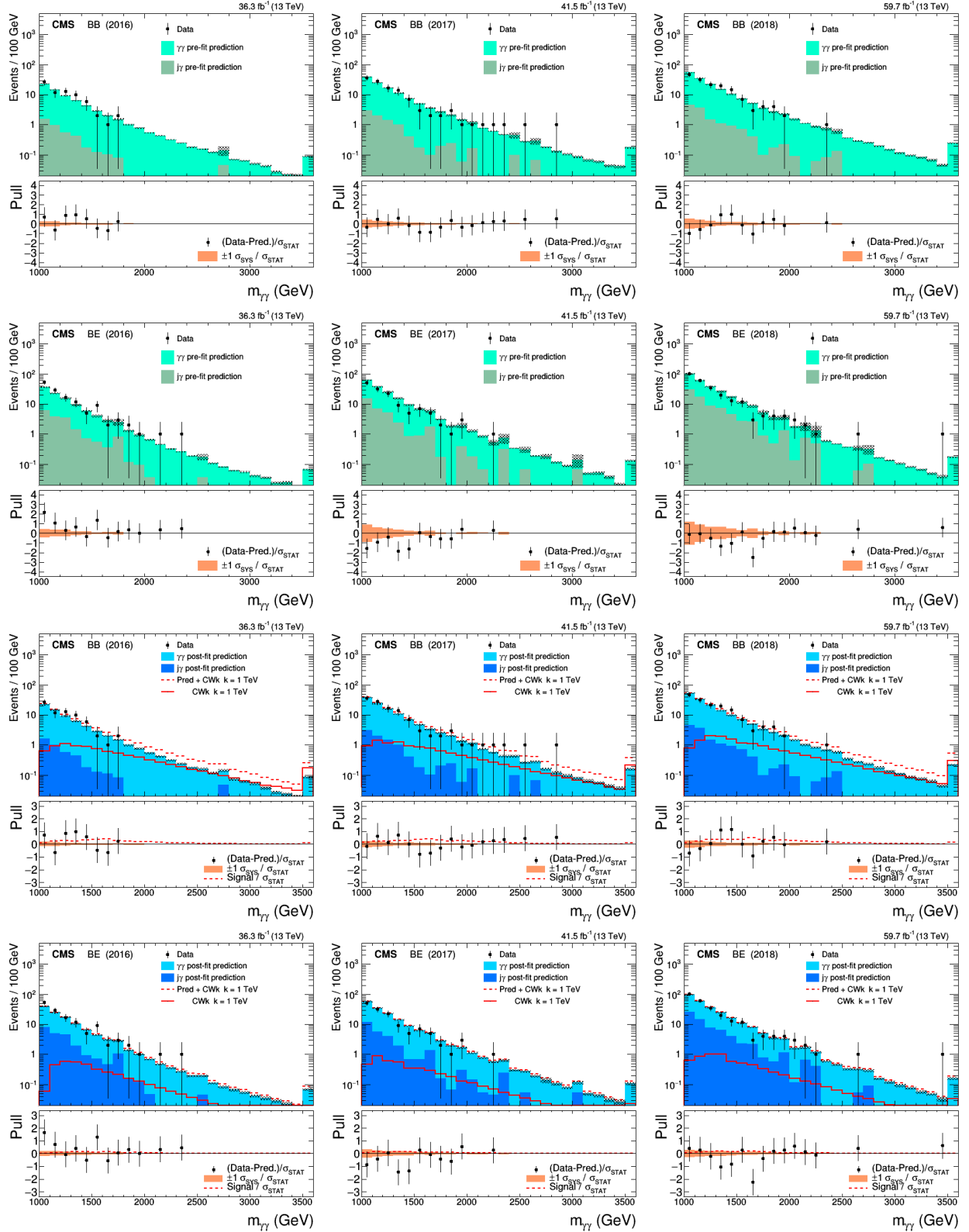
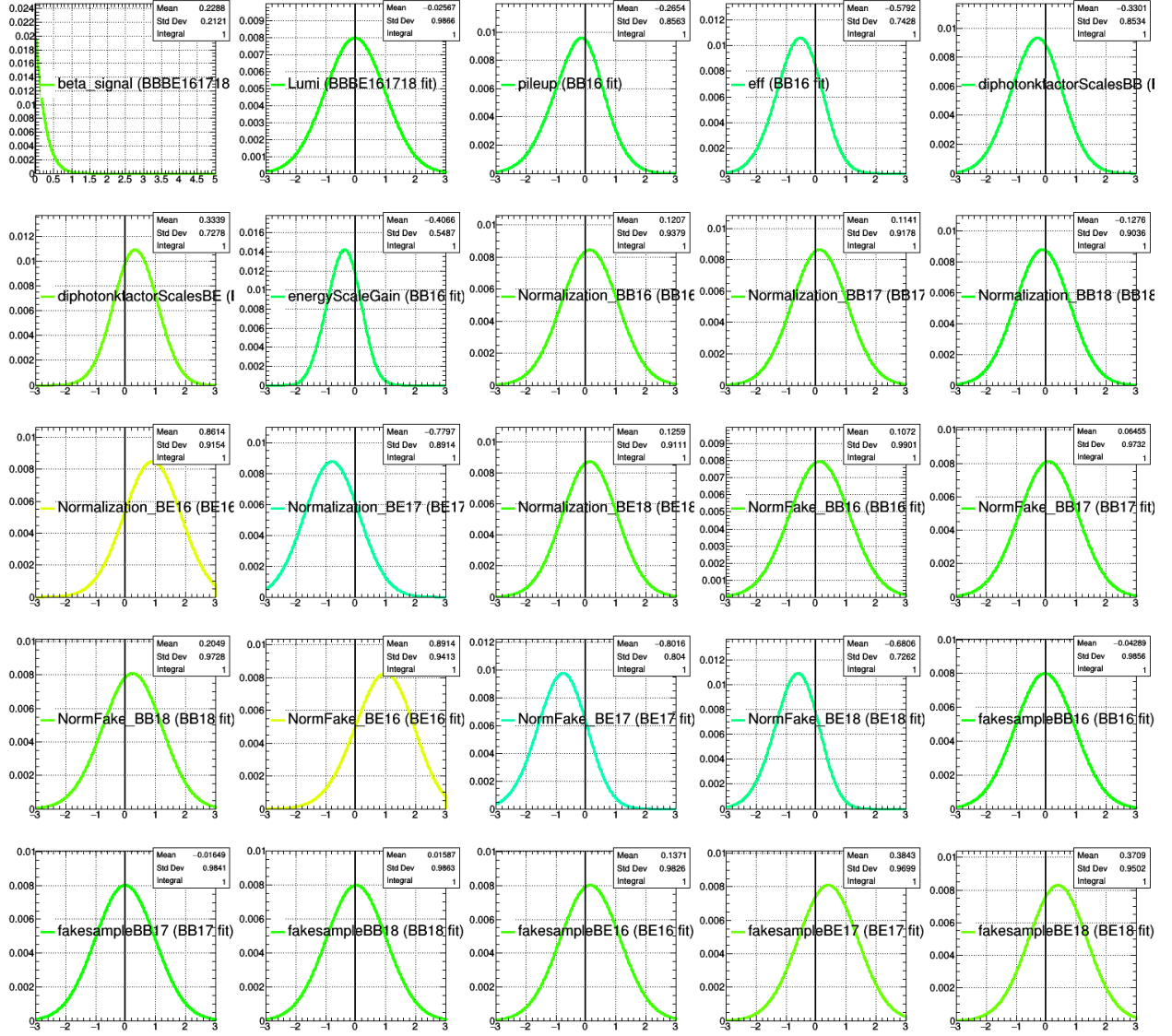


Figure 8.12: The postfit pdfs for the 25 nuisance parameters used in the six SRs fit with CW model signal.



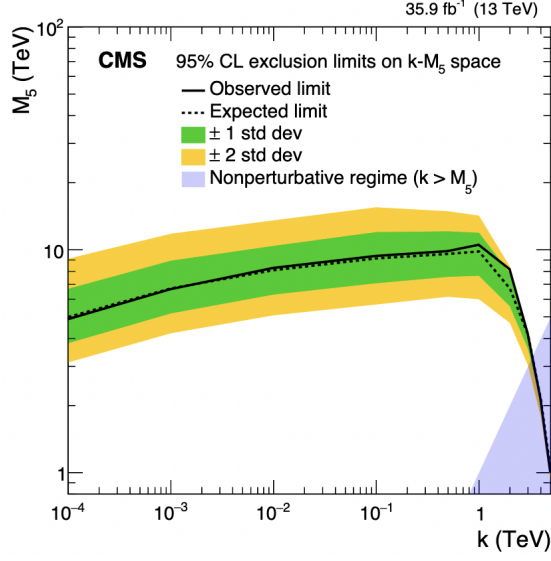
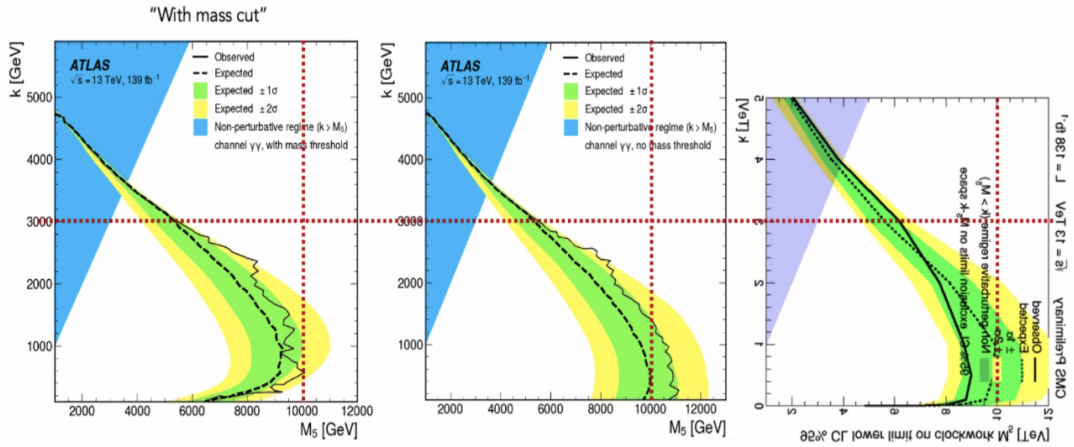


Figure 8.13: The previously published [131] Run II, 2016-only 95% CL exclusion limit for the clockwork model in the k – M_5 parameter space. The green (± 1) and yellow (± 2) shaded bands represent the standard deviation uncertainties in the expected limit (dashed lines). The shaded region in purple at the lower right is where the theory becomes nonperturbative.

Figure 8.14: ATLAS [73] Run II exclusion limits with mass cut and without mass cut shown with flipped CMS results for comparative purposes.



8.3 Preliminary Exclusion Limits on Unparticles

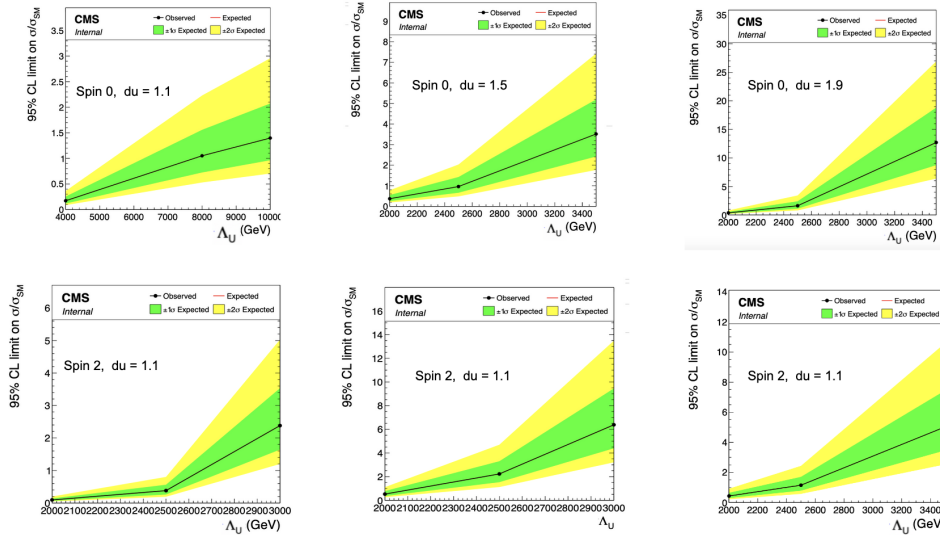
For the Unparticles, we use a cut-and-count approach using the CLs prescription in COMBINE. The inputs to COMBINE are datacards which contain the details of the experiment such as the signal and the background yields above $m_{\gamma\gamma} > 2000$ GeV.

The CLs Method [141, 142] is also known as the modified frequentist method which is more conservative than the standard p-value procedure. It is often used in particle physics experiments as the mathematical form shown in Eq. 8.7 avoids excluding a signal model when the sensitivity is low and also protects against exclusions due to underfluctuations in the data. The CLs prescription takes the ratio of the p-values under signal plus background, p_μ and background-only p_b hypotheses

$$CL_s = \frac{p_\mu}{1 - p_b}. \quad (8.7)$$

To quickly estimate the expected limits we use the Asymptotics Limits statistical method which is implemented in the COMBINE framework.

Figure 8.15: Limits on the Unparticles signal strength as a function of Λ_U . The top plots show the Spin-0 case for $d_u = 1.1, 1.5$, and 1.9 . The bottom shows the Spin-2 case.



For this preliminary study, we only make use of the 2017 centrally produced samples listed

in Appendix C.1 and normalized them to the total 2016-2018 CMS luminosity. Table 8.2 and Figure 8.15 show the preliminary limits results. Note that these expected limits results do not yet consider systematic uncertainties as well as the real data.

Table 8.2: Preliminary limits on the Unparticles

	Spin-0	Spin-2
d_U	Λ_U (TeV)	Λ_U (TeV)
1.1	7.8	2.7
1.5	2.5	2.4
1.9	2.2	2.5

8.3.1 Potential sources of additional sensitivity

We also performed additional sensitivity studies on the $U \rightarrow \gamma\gamma$ channel by looking at various parameters and kinematic variables. For these initial studies, we have used GEN-only events. From Figure 8.16, we can see that we are more sensitive to spin-2 than spin-0 if Λ_u and d_u are fixed. We also find in the same figure that for a fixed d_u (Λ_u) the sensitivity goes up with lower(higher) values of Λ_u (d_u). In future extensions of this work, we could conduct a multivariable analysis (MVA) or machine learning techniques to exploit other variables such as $\cos\theta^*$. This could provide additional sensitivity to the Unparticles model in the diphoton channel.

8.3.2 Comparison with Z+MET search

The search for Unparticles have also been done in the Z + MET channel where the unparticle appears as the missing energy undetected by CMS. In the latest results from Ref. [74], upper limits were set at 95% CL on the cross section with $\Lambda_U = 15$ TeV. Their observed (expected) limits are 0.5 (0.7) pb, 0.24 (0.26) pb, and 0.09 (0.07) pb for $d_u = 1$, $d_u = 1.5$, and $d_u = 2$. The limits depend on the choice of λ and Λ_u since the cross section scales with the Wilson coefficient written as $\frac{\lambda}{\Lambda_u}$. To compare this with other results, future limit plots should show how the limit varies with d_u .

Figure 8.16: Top: Spin-0 (red) and spin-2 (blue), $\mathcal{U} \rightarrow \gamma\gamma$ generator-only histograms with the same $\Lambda_u = 1000$ GeV and $d_u = 1.1$. The middle plot shows various d_u values with a fixed $\Lambda_u = 2500$ GeV. On the bottom, is a plot that has a fixed $d_u = 1.8$ GeV, and Λ is varied. The leading-order Standard Model Prompt Diphoton Background is shown in cyan.

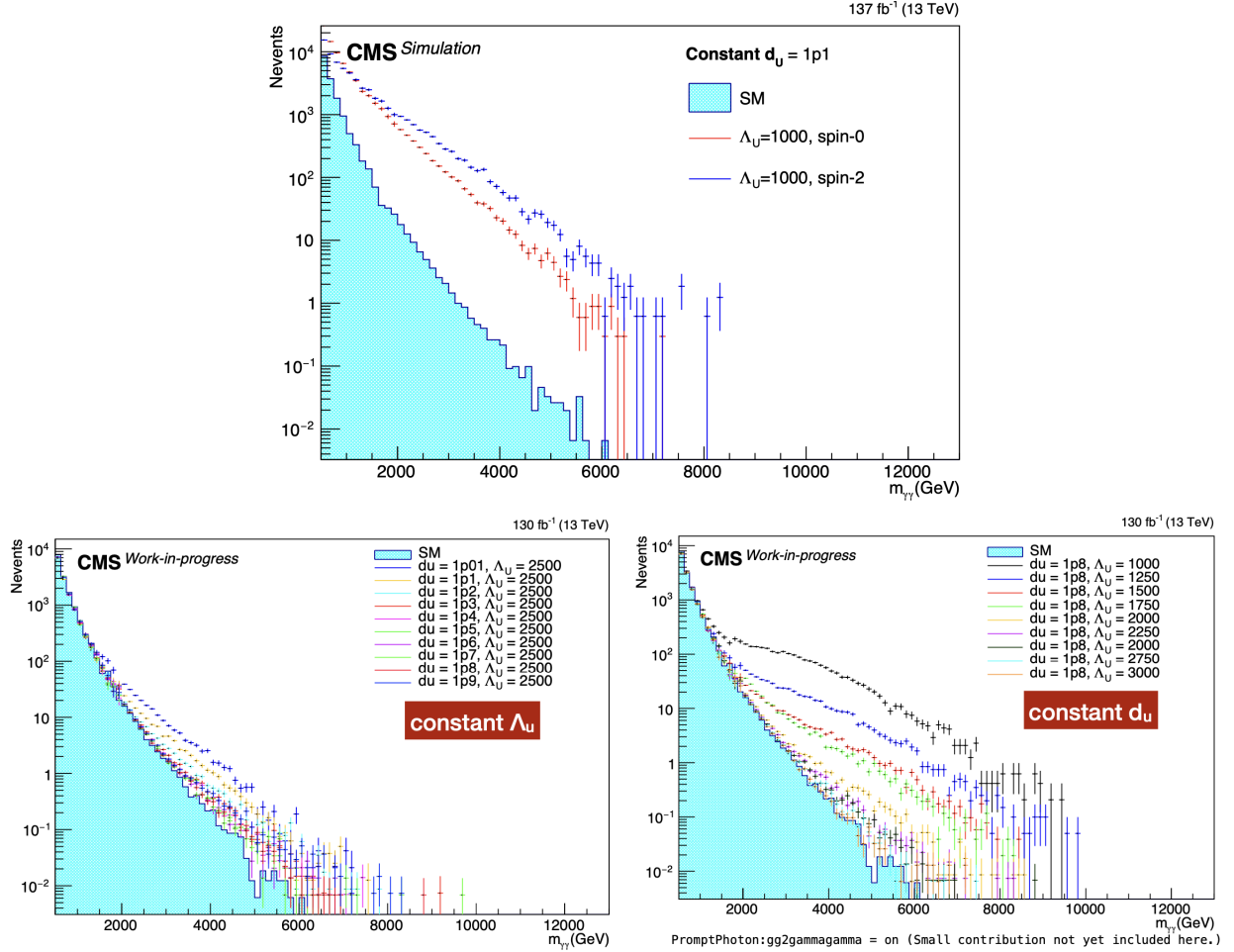
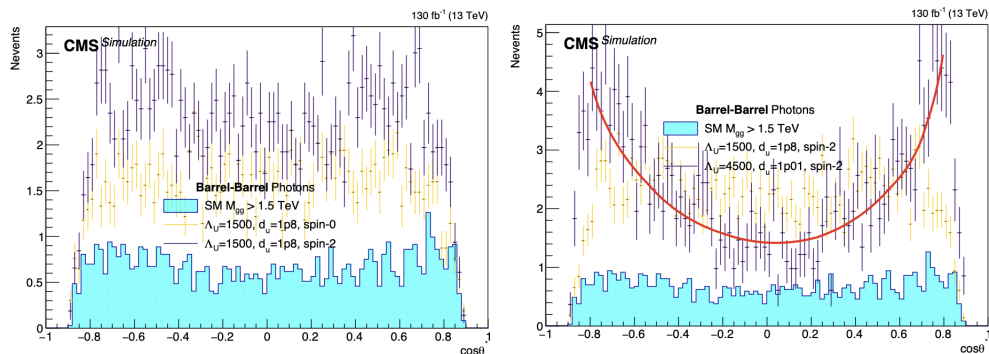


Figure 8.17: This figure shows $\cos\theta^*$ plots of $U \rightarrow \gamma\gamma$ versus the leading order SM only background (cyan). The right plot shows a different d_u for spin-2, curved differently from the SM background.



CHAPTER 9

CONCLUSIONS

This dissertation presented the CMS Run 2 legacy results for a search for non-resonant signature of Beyond the Standard Model (BSM) physics in high-mass diphoton events produced from pp collisions at a center-of-mass energy of 13 TeV using CMS data corresponding to 138 fb^{-1} . We considered three model interpretations which are the Arkani-Hamed, Dimopoulos, and Dvali (ADD) Large Extra Dimensions, the Clockwork/Linear Dilaton and the Unparticles. These models propose solutions to the hierarchy problem and the Higgs Naturalness Problem through the existence of extra spatial dimensions or a scale invariant sector that closes the gap between the electroweak and the illusory Planck Scale.

The data are consistent with the predictions of the Standard model and exclusion limits are placed on the ADD and CW model at 95% confidence level. For the ADD model the limits are further pushed to the range of $7.1 < M_S < 11.1 \text{ TeV}$ for the various model conventions. Limits are also set in the two-dimensional space of the Clockwork model, with the fundamental scale M_5 level below 8.0 TeV for k values in the range between 0.2 GeV and 2.0 TeV. These results extend the current best lower limits on M_S and M_5 by CMS in the diphoton channel, as presented in Ref. [131]. These results also add to over 10 years worth of work from various independent experiments at the LHC [12–17] and at the Fermilab Tevatron (USA) [23–30]. Additionally, preliminary studies on the expected limits for the Unparticles are also presented using a cut-and-count approach with the modified frequentist method (CLs) [139] to set conservative confidence intervals on the parameters of interest which are scaling dimension d_U and the energy scale Λ_U .

Various upgrades in the Hadronic Calorimeter (HCAL) have also been highlighted in this document. These upgrades include the reworking of the next-generation clock and control Modules, the creation of testing procedures for the back-end electronics and the software upgrades for the creation of the a high-level trigger for long-lived particles. These efforts are to prepare CMS for future data taking periods. Run 3 commenced in July 2022 after more than three years of upgrades and maintenance across the various experiments at the LHC.

Extensions of this analysis result for Run 3 are expected. First, a model-independent search could be performed using an autoencoder-based anomaly detection procedure to search for generic non-resonant deviations from the SM background. Additionally, a multivariate analysis (MVA) can be used to enhance the cut-based photon ID, as well as considering other kinematic variables. The CMS dijet Searches for Extra Dimensions incorporated angular information into the analysis as an additional signal discriminator [143]. For the fake prediction, a 30% non-closure contributes to a large uncertainty in our knowledge of the fake background normalization. The systematic uncertainties could also be reduced by using reprocessed data sets which include improved Electromagnetic Calorimeter (ECAL) calibration with better energy resolution [144]. Finally, the ATLAS collaboration [73] also suggested that since the non-resonant enhancements are narrowly spaced resonances undetected by even CMS’s excellent ECAL detector resolution, the continuous wavelet transform method (CWT) could allow us to examine the LHC spectra in the frequency domain and not just on the mass space alone. The outputs of the CWT in the form of scalograms can then be used as inputs to neural networks such as a binary classifier and an autoencoder to improve the limits on model-dependent and model-independent analyses. These are just some of the possibilities that may be realized by future analyzers.

At the time of writing this dissertation, the analysis presented here, in combination with a related search for resonant excesses in the high-mass diphoton spectrum, is in the final reading stage within the CMS collaboration and is on-track for publication in the Journal of High Energy Physics (JHEP). The publication-related data for the paper is found in HEPData.

REFERENCES

- [1] N. Arkani-Hamed, S. Dimopoulos, and G. 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.
- [2] I. Antoniadis, N. Arkani-Hamed, S. Dimopoulos et al., “New dimensions at a millimeter to a Fermi and superstrings at a TeV”, *Phys. Lett. B* **436** (1998) 257–263, doi:10.1016/S0370-2693(98)00860-0, arXiv:hep-ph/9804398.
- [3] 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.
- [4] L. Randall and R. Sundrum, “A Large mass hierarchy from a small extra dimension”, *Phys. Rev. Lett.* **83** (1999) 3370–3373, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.
- [5] L. Randall and R. Sundrum, “An Alternative to compactification”, *Phys. Rev. Lett.* **83** (1999) 4690–4693, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.
- [6] G. F. Giudice and M. McCullough, “A Clockwork Theory”, *JHEP* **02** (2017) 036, doi:10.1007/JHEP02(2017)036, arXiv:1610.07962.
- [7] M. Baryakhtar, “Graviton phenomenology of linear dilaton geometries”, *Phys. Rev. D* **85** (2012) 125019, doi:10.1103/PhysRevD.85.125019, arXiv:1202.6674.
- [8] G. F. Giudice, Y. Kats, M. McCullough et al., “Clockwork / linear dilaton: structure and phenomenology”, *JHEP* **06** (2018) 009, doi:10.1007/JHEP06(2018)009, arXiv:1711.08437.
- [9] G. F. Giudice and A. Strumia, “Constraints on extra-dimensional theories from virtual-graviton exchange”, *Nucl. Phys. B* **663** (2003) 377, doi:10.1016/S0550-3213(03)00404-8, arXiv:hep-ph/0301232.
- [10] G. F. Giudice, T. Plehn, and A. Strumia, “Graviton collider effects in one and more large extra dimensions”, *Nucl. Phys. B* **706** (2005) 455, doi:10.1016/j.nuclphysb.2004.10.056, arXiv:hep-ph/0408320.
- [11] R. Franceschini, P. P. Giardino, G. F. Giudice et al., “LHC bounds on large extra dimensions”, *JHEP* **05** (2011) 092, doi:10.1007/JHEP05(2011)092, arXiv:1101.4919.
- [12] CMS Collaboration, “Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **98** (2018) 092001, doi:10.1103/PhysRevD.98.092001, arXiv:1809.00327.

- [13] ATLAS Collaboration, “Search for resonances decaying into photon pairs in 139 fb⁻¹ of pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector”, *Phys. Lett. B* **822** (2021) 136651, doi:10.1016/j.physletb.2021.136651, arXiv:2102.13405.
- [14] 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 TeV with the ATLAS detector”, *Phys. Lett. B* **775** (2017) 105, doi:10.1016/j.physletb.2017.10.039, arXiv:1707.04147.
- [15] 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, doi:10.1016/j.physletb.2017.01.027, arXiv:1609.02507.
- [16] 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.
- [17] 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) 051802, doi:10.1103/PhysRevLett.117.051802, arXiv:1606.04093.
- [18] 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) 032004, doi:10.1103/PhysRevD.92.032004, arXiv:1504.05511.
- [19] CMS Collaboration, “Search for diphoton resonances in the mass range from 150 to 850 GeV in pp collisions at $\sqrt{s}$ = 8 TeV”, *Phys. Lett. B* **750** (2015) 494, doi:10.1016/j.physletb.2015.09.062, arXiv:1506.02301.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] CDF Collaboration, “Limits on Extra Dimensions and New Particle Production in the Exclusive Photon and Missing Energy Signature in $p\bar{p}$ Collisions at $\sqrt{s}$ = 1.8

- TeV”, *Phys. Rev. Lett.* **89** (2002) 281801, doi:10.1103/PhysRevLett.89.281801, arXiv:hep-ex/0205057.
- [25] CDF Collaboration, “Search for Randall–Sundrum Gravitons in the Diphoton Channel at CDF”, *Phys. Rev. D* **83** (2011) 011102, doi:10.1103/PhysRevD.83.011102, arXiv:1012.2795.
 - [26] 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.
 - [27] 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.
 - [28] D0 Collaboration, “Search for large extra dimensions in dielectron and diphoton production”, *Phys. Rev. Lett.* **86** (2001) 1156, doi:10.1103/PhysRevLett.86.1156, arXiv:hep-ex/0008065.
 - [29] D0 Collaboration, “Search for large extra spatial dimensions in the dielectron and diphoton channels in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\text{ TeV}$ ”, *Phys. Rev. Lett.* **102** (2009) 051601, doi:10.1103/PhysRevLett.102.051601, arXiv:0809.2813.
 - [30] D0 Collaboration, “Search for Randall–Sundrum gravitons in dilepton and diphoton final states”, *Phys. Rev. Lett.* **95** (2005) 091801, doi:10.1103/PhysRevLett.95.091801, arXiv:hep-ex/0505018.
 - [31] L. R. I. Natalie Wolchover, Samuel Velasco, “A New map of the Standard Model”, 2020.
 - [32] Peskin, Michael E. and Schroeder, Daniel V., “An Introduction to quantum field theory”. Addison-Wesley, Reading, USA, 1995.
 - [33] F. Halzen and A. D. Martin, “QUARKS AND LEPTONS: AN INTRODUCTORY COURSE IN MODERN PARTICLE PHYSICS”. 1984.
 - [34] M. Kobayashi and T. Maskawa, “CP Violation in the Renormalizable Theory of Weak Interaction”, *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.
 - [35] Particle Data Group Collaboration, “Review of Particle Physics”, *Phys. Rev. D* **98** (2018), no. 3, 030001, doi:10.1103/PhysRevD.98.030001.
 - [36] M. Klute, “Introduction to Nuclear and Particle Physics”, 2020.
 - [37] Z. Maki, M. Nakagawa, and S. Sakata, “Remarks on the unified model of elementary particles”, *Prog. Theor. Phys.* **28** (1962) 870–880, doi:10.1143/PTP.28.870.
 - [38] K. Nguyen, “The Higgs Mechanism”.

- [39] S. P. Martin, “A Supersymmetry primer”, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1–98, doi:10.1142/9789812839657_0001, arXiv:hep-ph/9709356.
- [40] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, “The Hierarchy problem and new dimensions at a millimeter”, *Phys. Lett. B* **429** (1998) 263–272, doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.
- [41] 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.
- [42] K. R. Dienes, E. Dudas, and T. Gherghetta, “Grand unification at intermediate mass scales through extra dimensions”, *Nucl. Phys. B* **537** (1999) 47–108, doi:10.1016/S0550-3213(98)00669-5, arXiv:hep-ph/9806292.
- [43] K. R. Dienes, E. Dudas, and T. Gherghetta, “Extra space-time dimensions and unification”, *Phys. Lett. B* **436** (1998) 55–65, doi:10.1016/S0370-2693(98)00977-0, arXiv:hep-ph/9803466.
- [44] T. Kaluza, “Zum Unitätsproblem der Physik”, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1921** (1921) 966–972, doi:10.1142/S0218271818700017, arXiv:1803.08616.
- [45] O. Klein, “The Atomicity of Electricity as a Quantum Theory Law”, *Nature* **118** (1926) 516, doi:10.1038/118516a0.
- [46] O. Klein, “Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)”, *Z. Phys.* **37** (1926) 895–906, doi:10.1007/BF01397481.
- [47] Particle Data Group Collaboration, “Review of Particle Physics”, *PTEP* **2020** (2020), no. 8, 083C01, doi:10.1093/ptep/ptaa104.
- [48] G. F. Giudice and M. McCullough, “A clockwork theory”, *Journal of High Energy Physics* (2017) doi:10.1007/JHEP02(2017)036.
- [49] H. Georgi, “Unparticle Physics”, *Phys. Rev. Lett* **98** (2007) doi:10.1103/physrevlett.98.221601, arXiv:hep-ph/0709.2478.
- [50] H. Georgi, “Unparticle Physics”, *Phys. Lett. B* doi:10.1103/PhysRevLett.98.221601, arXiv:0703.3260.
- [51] H. Georgi, “Another odd thing about unparticle physics”, *Phys. Lett. B* **650** (2007) 275–278, doi:10.1016/j.physletb.2007.05.037, arXiv:0704.2457.
- [52] M. G. Folgado and V. Sanz, “On the Interpretation of Nonresonant Phenomena at Colliders”, *Adv. High Energy Phys.* **2021** (2021) 2573471, doi:10.1155/2021/2573471, arXiv:2005.06492.

- [53] G. D. Kribs, “TASI 2004 lectures on the phenomenology of extra dimensions”, in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in D* 4, pp. 633–699. 5, 2006. [arXiv:hep-ph/0605325](#).
- [54] A. Perez-Lorenzana, “An Introduction to extra dimensions”, *J. Phys. Conf. Ser.* **18** (2005) 224–269, doi:10.1088/1742-6596/18/1/006, [arXiv:hep-ph/0503177](#).
- [55] 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](#).
- [56] 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](#).
- [57] 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](#).
- [58] 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](#).
- [59] G. F. Giudice, Y. Kats, M. McCullough et al., “Clockwork/linear dilaton: structure and phenomenology”, *JHEP* **06** (2018) 009, doi:10.1007/JHEP06(2018)009, [arXiv:1711.08437](#).
- [60] T. Kikuchi and N. Okada, “Unparticle physics and Higgs phenomenology”, *Phys. Lett. B* **661** (2008) 360–364, doi:10.1016/j.physletb.2008.02.041, [arXiv:0707.0893](#).
- [61] C.-H. Chen and C.-Q. Geng, “Unparticle physics on direct CP violation”, *Phys. Rev. D* **76** (2007) 115003, doi:10.1103/PhysRevD.76.115003, [arXiv:0705.0689](#).
- [62] A. Freitas and D. Wyler, “Astro Unparticle Physics”, *JHEP* **12** (2007) 033, doi:10.1088/1126-6708/2007/12/033, [arXiv:0708.4339](#).
- [63] R. Zwicky, “Unparticles at heavy flavour scales: CP violating phenomena”, *Phys. Rev. D* **77** (2008) 036004, doi:10.1103/PhysRevD.77.036004, [arXiv:0707.0677](#).
- [64] H. Georgi, “Effective field theory”, *Ann. Rev. Nucl. Part. Sci.* **43** (1993) 209–252, doi:10.1146/annurev.ns.43.120193.001233.
- [65] K. Cheung, W.-Y. Keung, and T.-C. Yuan, “Collider Phenomenology of Unparticle Physics”, *Phys. Rev. D* **76** (2007) 055003, doi:10.1103/PhysRevD.76.055003, [arXiv:0706.3155](#).
- [66] K. Cheung, W.-Y. Keung, and T.-C. Yuan, “Collider phenomenology of unparticle physics”, *Physical Review D* **76** (2007) doi:10.1103/physrevd.76.055003, [arXiv:hep-ph/07063155](#).

- [67] M. Kenzie, “A Review of Unparticle Physics”, 2010.
- [68] M. C. Kumar, P. Mathews, V. Ravindran et al., “Unparticle physics in diphoton production at the CERN LHC”, *Phys. Rev. D* **77** (2008) 055013, doi:10.1103/PhysRevD.77.055013, arXiv:0709.2478.
- [69] M. C. Kumar, P. Mathews, V. Ravindran et al., “Unparticles in diphoton production to NLO in QCD at the LHC”, *Phys. Rev. D* **79** (2009) 075012, doi:10.1103/PhysRevD.79.075012, arXiv:0804.4054.
- [70] S. Ask, I. V. Akin, L. Benucci et al., “Real Emission and Virtual Exchange of Gravitons and Unparticles in Pythia8”, *Comput. Phys. Commun.* **181** (2010) 1593–1604, doi:10.1016/j.cpc.2010.05.013, arXiv:0912.4233.
- [71] CMS Collaboration, “Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **98** (2018), no. 9, 092001, doi:10.1103/PhysRevD.98.092001, arXiv:1809.00327.
- [72] CMS Collaboration, “Search for resonant and nonresonant new phenomena in high-mass dilepton final states at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2021) 208, doi:10.1007/JHEP07(2021)208, arXiv:2103.02708.
- [73] ATLAS Collaboration, “Search for periodic signals in the dielectron and diphoton invariant mass spectra using 139 fb⁻¹ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, arXiv:2305.10894.
- [74] CMS Collaboration, “Search for dark matter produced in association with a leptonically decaying Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Eur. Phys. J. C* **81** (2021), no. 1, 13, doi:10.1140/epjc/s10052-020-08739-5, arXiv:2008.04735. [Erratum: *Eur.Phys.J.C* 81, 333 (2021)].
- [75] P. Mouche, “Overall view of the LHC. Vue d’ensemble du LHC”, Jun, 2014.
- [76] L. Rossi, “The LHC main dipoles and quadrupoles toward series production”, in *Applied Superconductivity Conference*. 12, 2002.
- [77] ATLAS Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003.
- [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] ALICE Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002.

- [81] M. Ferrario and B. J. Holzer, “Introduction to Particle Accelerators and their Limitations”, in *CAS - CERN Accelerator School 2019: High Gradient Wakefield Accelerators*. 7, 2020. [arXiv:2007.04075](#).
- [82] M. Andrews, “Search for exotic Higgs boson decays to merged photons employing a novel deep learning technique at CMS”. PhD thesis, Carnegie Mellon U., 2022. [doi:10.1184/R1/18412073.v1](#).
- [83] A. Blondel and P. Janot, “FCC-ee overview: new opportunities create new challenges”, *Eur. Phys. J. Plus* **137** (2022), no. 1, 92, [doi:10.1140/epjp/s13360-021-02154-9](#), [arXiv:2106.13885](#).
- [84] CMS Collaboration, “Pileup mitigation at CMS in 13 TeV data”, *JINST* **15** (2020), no. 09, P09018, [doi:10.1088/1748-0221/15/09/P09018](#), [arXiv:2003.00503](#).
- [85] CMS Collaboration, “Development of the CMS detector for the CERN LHC Run 3”, [arXiv:2309.05466](#).
- [86] 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](#).
- [87] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, [doi:10.1088/1748-0221/3/08/S08004](#).
- [88] G. Cummings, “A Search for Resonant Cascades to Neutral Standard Model Bosons and p_T miss and Selected Stories of the CMS Hadron Calorimeter Phase 1 Upgrade”. PhD thesis, 2022. Presented on October 26, 2022.
- [89] ATLAS Collaboration, “Survey of Big Data Sizes in 2021”, [arXiv:2202.07659](#).
- [90] CMS Collaboration, “Phase I Upgrade of the CMS Hadron Calorimeter”, *Nucl. Part. Phys. Proc.* **273-275** (2016) 1002–1007, [doi:10.1016/j.nuclphysbps.2015.09.157](#).
- [91] CMS Collaboration, “CMS HCAL VTRx-induced communication loss and mitigation”, *JINST* **17** (2022), no. 05, C05020, [doi:10.1088/1748-0221/17/05/C05020](#).
- [92] G. Kopp et al., “Specifications for HCAL uHTR Firmware”,.
- [93] CMS Collaboration, F. Beaudette, “The CMS Particle Flow Algorithm”, in *International Conference on Calorimetry for the High Energy Frontier*, pp. 295–304. 2013. [arXiv:1401.8155](#).
- [94] 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](#).

- [95] M. Andrews, J. Alison, S. An et al., “End-to-end jet classification of quarks and gluons with the CMS Open Data”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **977** (oct, 2020) 164304, doi:10.1016/j.nima.2020.164304.
- [96] M. Andrews, M. Paulini, S. Gleyzer et al., “End-to-End Physics Event Classification with CMS Open Data: Applying Image-Based Deep Learning to Detector Data for the Direct Classification of Collision Events at the LHC”, *Computing and Software for Big Science* **4** (mar, 2020) doi:10.1007/s41781-020-00038-8.
- [97] 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.
- [98] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [99] CMS Collaboration, “Observation of the diphoton decay of the Higgs boson and measurement of its properties”, *Eur. Phys. J. C* **74** (2014) 3076, doi:10.1140/epjc/s10052-014-3076-z, arXiv:1407.0558.
- [100] CMS Collaboration, “Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11** (2018) 185, doi:10.1007/JHEP11(2018)185, arXiv:1804.02716.
- [101] 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.
- [102] 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.
- [103] CMS Collaboration, “Electron and Photon Performance in CMS in Run-2 and Prospects for Run-3”, *Phys. At. Nucl.* **84** (2021), no. 4, 571–578, doi:10.1134/S1063778821040207.
- [104] CMS Collaboration, “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **13** (2018) P06015, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.
- [105] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-ik/isubit/i/subjet clustering algorithm”, *Journal of High Energy Physics* **2008** (apr, 2008) 063–063, doi:10.1088/1126-6708/2008/04/063.
- [106] CMS Collaboration, “Performance of Jet reconstruction in CMS at 13 TeV”, *PoS ICHEP2016* (2016) 736, doi:10.22323/1.282.0736.

- [107] C. D. working group, “Search for new physics in the diphoton spectrum at high mass at 13 TeV (2016 data)”, CMS Note 2016/167, CERN, (2016).
- [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] M. Raymond, J. Crooks, M. French et al., “The MGPA electromagnetic calorimeter readout chip for CMS”, *IEEE Transactions on Nuclear Science* **52** (2005), no. 3, 756–763, doi:10.1109/TNS.2005.850979.
- [111] B. Clerbaux, S. Elgammal, T. Mahmoud et al., “Saturation and energy corrections for TeV electrons and photons”,.
- [112] CMS Collaboration Collaboration, “Guide to E/gamma MiniAOD V2 Format”.
- [113] The CMS Collaboration, “Generic Tag and Probe Tool for Measuring Efficiency at CMS with Early Data”, *CMS Analysis note* **2009/111** (2009).
- [114] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.
- [115] C. Collaboration, “Reweighting recipe to emulate Level 1 ECAL prefiring”, 2019-04-08.
- [116] Particle Data Group Collaboration, “Review of Particle Physics”, *Phys. Rev. D* **98** (2018), no. 3, 030001, doi:10.1103/PhysRevD.98.030001.
- [117] T. Gleisberg, S. Hoeche, F. Krauss et al., “Event generation with SHERPA 1.1”, *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.
- [118] C. Bierlich, S. Chakraborty, N. Desai et al., “A comprehensive guide to the physics and usage of PYTHIA 8.3”, arXiv:2203.11601.
- [119] T. Sjöstrand, S. Mrenna, and P. Skands, “An Introduction to PYTHIA 8.1”, *Comput. Phys. Comm.* **178** (2008) 852–867, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- [120] NNPDF Collaboration, “An open-source machine learning framework for global analyses of parton distributions”, *Eur. Phys. J. C* **81** (2021), no. 10, 958, doi:10.1140/epjc/s10052-021-09747-9, arXiv:2109.02671.
- [121] L. Del Debbio, “Parton distributions in the LHC era”, *EPJ Web Conf.* **175** (2018) 01006, doi:10.1051/epjconf/201817501006.
- [122] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, arXiv:1903.12179.

- [123] 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.
- [124] S. Badger, A. Guffanti, and V. Yundin, “Next-to-leading order QCD corrections to di-photon production in association with up to three jets at the Large Hadron Collider”, *JHEP* **03** (2014) 122, doi:10.1007/JHEP03(2014)122, arXiv:1312.5927.
- [125] CMS Collaboration, “Search for High-Mass Diphoton Resonances in pp Collisions at $\sqrt{s} = 8$ TeV with the CMS Detector”, Technical Report CMS-PAS-EXO-12-045, CERN, Geneva, (2015).
- [126] J. M. Campbell, R. K. Ellis, Y. Li et al., “Predictions for diphoton production at the LHC through NNLO in QCD”, *JHEP* **07** (2016) 148, doi:10.1007/JHEP07(2016)148, arXiv:1603.02663.
- [127] J. Campbell and T. Neumann, “Precision Phenomenology with MCFM”, arXiv:1909.09117.
- [128] 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.
- [129] NNPDF Collaboration, “Parton distributions for the LHC Run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [130] 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.
- [131] CMS Collaboration, “Search for physics beyond the standard model in high-mass diphoton events from proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. D* **98** (2018), no. 9, 092001, doi:10.1103/PhysRevD.98.092001, arXiv:1809.00327.
- [132] R. Barlow and C. Beeston, “Fitting using finite monte carlo samples”, *Computer Physics Communications* **77** (1993) 219–228.
- [133] T. Gleisberg, F. Krauss, K. T. Matchev et al., “Helicity formalism for spin-2 particles”, *JHEP* **09** (2003) 001, doi:10.1088/1126-6708/2003/09/001, arXiv:hep-ph/0306182.
- [134] C. Degrande, C. Duhr, B. Fuks et al., “UFO - The Universal FeynRules Output”, *Comput. Phys. Commun.* **183** (2012) 1201–1214, doi:10.1016/j.cpc.2012.01.022, arXiv:1108.2040.
- [135] CMS Collaboration, “Search for contact interactions and large extradimensions in the dilepton mass spectra from proton-proton collisions at $s = 13$ TeV”, *Journal of High Energy Physics* (2019) doi:10.001/JHEP04(2019)114, arXiv:1812.10443v2.

- [136] CMS Collaboration, “Search for new physics with dijet angular distributions in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2017) 013, doi:10.1007/JHEP07(2017)013, arXiv:1703.09986.
- [137] M. C. Kumar, P. Mathews, V. Ravindran et al., “Unparticle physics in diphoton production at the CERN LHC”, *Physical Review D* **77** (2008) doi:10.1103/physrevd.77.055013, arXiv:hep-ph/0709.2478.
- [138] ATLAS and CMS Collaborations, and LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-11, (2011).
- [139] G. Cowan, K. Cranmer, E. Gross et al., “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: 10.1140/epjc/s10052-013-2501-z].
- [140] J. Ott, “Practical Statistics”.
- [141] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [142] A. L. Read, “Presentation of search results: The CL_s technique”, *J. Phys. G* **28** (2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.
- [143] CMS Collaboration, “Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles”, *JHEP* **08** (2018) 130, doi:10.1007/JHEP08(2018)130, arXiv:1806.00843.
- [144] CMS Collaboration, “Calibration and Performance of the CMS Electromagnetic Calorimeter in LHC Run2”, *EPJ Web Conf.* **245** (2020) 02027, doi:10.1051/epjconf/202024502027.
- [145] M. Doneg, P. Musella, S. Pigazzini et al., “Search for high-mass resonances in the diphoton channel (full 2016 dataset)”, CMS Note 2017/206, (2017).
- [146] S. M. Kaplan, “A Search for Large Extra Dimensions in the non-resonant Diphoton Channel with 35.9 fb^{-1} of $\sqrt{s} = 13$ TeV LHC Data ”. PhD thesis, Rutgers University, 2017. doi:10.1184/R1/18412073.v1.
- [147] T. C. Experiment, “Publishing an Analysis CMS Experiment”.

APPENDIX A

SCALE FACTORS AND DATASETS

In this chapter we discuss the datasets used for the tag-and-probe method and the final scale factors used in the analysis.

The tag-and-probe method begins by obtaining a set of high purity $Z \rightarrow e^+e^-$ samples from data shown in Table A.3. Events are required to pass an unprescaled single electron trigger, listed on Table A.2. These single electron triggers are used to assure that the probe electrons selected are unbiased. Both the tag electrons and the probe electrons are further selected with the following cuts listed in Table A.4. The tag-and-probe system must be within the invariant mass range from 70 GeV to 110 GeV. The tag electrons have p_T cuts greater than the HLT requirements at to avoid ambiguity. The selected probe electrons are further categorized into whether they fail or pass the photon ID. The distributions are binned and p_T and η as follows: $p_T = [125, 200, 300, 400, 500, 1000]$ GeV and $|\eta| = [0, 0.8, 1.4442]$ in the EB and $|\eta| = [1.566, 2.5]$ in the endcap region.

In the next step, we apply the same selection procedure on MC samples instead of data. The MC samples are listed in Table A. The choice of using HT-binned Drell-Yann samples are governed by the need for more statistics in the high p_T region.

The passing and failing probe distributions for each $\eta - p_T$ bins are fitted to get the signal yields. The nominal fitting signal model is the template from MC, Zlineshape, the lineshape of the Z resonance peak is extracted and smeared with the Gaussian. The nominal fitting background model used is the RooCMSShape function which is a special case of the general

Table A.1: Nominal MC samples used for TnP

Year	Nominal MC samples
2016	/DYJetsToLL_M-50_HT-*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2
2017	/DYJetsToLL_M-50_HT-*_TuneCP5_13TeV-madgraphMLM-pythia8 /RunIIFall17MiniAODv2-PU2017_12Apr2018_new_pmx_94X_mc2017_realistic_v14-v2
2018	/DYJetsToLL_M-50_HT-*_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8 RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v2

Crystal Ball function, a pdf which combines a Gaussian core with a power-tail law used to model the invariant mass distribution of a particle resonance decay. We extract the signal yield for data. The MC, which serves as our ground truth, we use a simple cut and count. Using these yields we can subsequently calculate the efficiency which is defined as follows:

$$\text{Efficiency} = \frac{\text{Signal Yield}_{\text{passing probe}}}{\text{Signal Yield}_{\text{passing probe}} + \text{Signal Yield}_{\text{failing probe}}} \quad (\text{A.1})$$

Alternative MC samples listed in Table A.5 are used to evaluate the generator dependence and the fitting model systematics. The signal model and the background models are alternately chosen as the nominal one. The alternative signal models were analytically fit with a Crystal Ball function convoluted with a Gaussian. The alternative background model is fitted with a second order polynomial. The default alternative background model in the tag-and-probe package is an exponential function. Comparing the fit results between the two alternative background models, the second order polynomial has a better fit and gives us smaller systematics. The representative fit shown in Figure 4.4 shows that the exponential function does not fit very well due to the sign of its curvature. The second order polynomial function has more freedom to flex its curvature to fit the data and results in a better fit.

Finally, the scale factors are calculated by dividing the efficiency in data and the MC in $\eta - p_T$ bins. The results are shown in Figure 4.8. The large uncertainties in the high-pt region dominated by the statistics, and this leads us to extrapolate the scale factors above 200 GeV.

$$\text{Scale Factor}(\eta, p_T) = \frac{\text{Efficiency}_{\text{Data}}(\eta, p_T)}{\text{Efficiency}_{\text{MC}}(\eta, p_T)} \quad (\text{A.2})$$

Table A.2: Single electron high level trigger

Year	TnP trigger path
2016	HLT_Ele27_WPTight_Gsf
2017	HLT_Ele32_WPTight_Gsf_L1DoubleEG
2018	HLT_Ele32_WPTight_Gsf

Table A.3: Datasets used for TnP

Year	Dataset
2016	/SingleElectron/Run2016B-17Jul2018-ver2-v1/MINIAOD /SingleElectron/Run2016C-17Jul2018-v1/MINIAOD /SingleElectron/Run2016D-17Jul2018-v1/MINIAOD /SingleElectron/Run2016E-17Jul2018-v1/MINIAOD /SingleElectron/Run2016F-17Jul2018-v1/MINIAOD /SingleElectron/Run2016G-17Jul2018-v1/MINIAOD /SingleElectron/Run2016H-17Jul2018-v1/MINIAOD
2017	/SingleElectron/Run2017B-31Mar2018-v1/MINIAOD /SingleElectron/Run2017C-31Mar2018-v1/MINIAOD /SingleElectron/Run2017D-31Mar2018-v1/MINIAOD /SingleElectron/Run2017E-31Mar2018-v1/MINIAOD /SingleElectron/Run2017F-31Mar2018-v1/MINIAOD
2018	/EGamma/Run2018B-17Sep2018-v1/MINIAOD /EGamma/Run2018C-17Sep2018-v1/MINIAOD /EGamma/Run2018D-22Jan2019-v2/MINIAOD

Table A.4: Tag and Probe Selection for 2017 and 2018

Tag Selection	Probe Selection
$ \eta < 1.4442$ or $1.566 < \eta < 2.1$ $p_T > 30.0$ (35.0) Pass Tight electronCutBasedID	$ \eta < 1.4442$ or $1.566 < \eta < 2.5$

At high p_T , we extrapolate the scale factors as a function of p_T . The scale factors are found to be relatively flat and we take the central value and extrapolate to the entire high p_T region. The detailed procedure to obtain the extrapolated scale factors and uncertainties are described as follows See Appendix A for final scale factors.

Table A.5: Alternative MC samples used for TnP

Year	Nominal MC samples
2016	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8 /RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3_ext2-v1
2017	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8 /RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1*
2018	/DYToEE_M-50_NNPdF31_TuneCP5_13TeV-powheg-pythia8 /RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1*

For each η binning defined in the tag-and-probe package, the scale factors vs p_T are fitted with a first order polynomial function, defined as $p_0 + p_1(p_T - 200)$. The fitted value of p_0 would be taken as the constant scale factor for $p_T > 200$ GeV. The fitting error of p_1 would be taken as the 68% limit on the gradient of SF to reflect how confident we are to believe the scale factors are flat, and this would be an additional systematics for extrapolation.

The final systematics for $p_T > 200$ GeV would be the systematics of 200-300 GeV p_T bin obtained from tag-and-probe method $\oplus$ extra extrapolation systematics. A consistency check between the extrapolated and tag-and-probe measured results are done, and the tag-and-probe scale factors are within uncertainty. The final uncertainties are also compared with the constant 6% systematics quoted in the 2016 analysis [145]. The current uncertainties are smaller in most p_T region due to more statistics from HT-binned DY samples, a better alternative background function choice, and the extrapolation. Towards very high p_T region the extrapolation systematics start to dominate because its p_T dependent behaviour, which reflects some non-flatness of the scale factors.

Table A.6: Final scale factors and uncertainties for 2016.

η -bin	p_T bin	Scale Factor	Uncertainty	Extrapolation Uncertainty (GeV)
$0 < \eta < 0.8$	$125 < p_T < 200$	1.0168	0.0108	-
$0 < \eta < 0.8$	$200 < p_T$	1.0133	0.0186	5.33e-05
$0.8 < \eta < 1.4442$	$125 < p_T < 200$	0.9963	0.0098	-
$0.8 < \eta < 1.4442$	$200 < p_T$	0.9951	0.0129	8.61e-01
$1.566 < \eta < 2.5$	$125 < p_T < 200$	1.0102	0.0145	-
$1.566 < \eta < 2.5$	$200 < p_T < 300$	1.0030	0.0225	1.57e-04

Table A.7: Final scale factors and uncertainties for 2017.

η -bin	p_T bin	Scale Factor	Uncertainty	Extrapolation Uncertainty (GeV)
$0 < \eta < 0.8$	$125 < p_T < 200$	1.0166	0.0134	-
$0 < \eta < 0.8$	$200 < p_T$	1.0147	0.0318	6.26e-05
$0.8 < \eta < 1.4442$	$125 < p_T < 200$	1.0036	0.0214	-
$0.8 < \eta < 1.4442$	$200 < p_T$	1.0063	0.0390	8.46e-05
$1.566 < \eta < 2.5$	$125 < p_T < 200$	0.9906	0.0137	-
$1.566 < \eta < 2.5$	$200 < p_T < 300$	1.0007	0.0311	8.97e-05

Table A.8: Final scale factors and uncertainties for 2018.

η -bin	p_T bin	Scale Factor	Uncertainty	Extrapolation Uncertainty (GeV)
$0 < \eta < 0.8$	$125 < p_T < 200$	0.9989	0.0103	-
$0 < \eta < 0.8$	$200 < p_T$	1.0030	0.0115	4.79e-05
$0.8 < \eta < 1.4442$	$125 < p_T < 200$	0.9858	0.0119	-
$0.8 < \eta < 1.4442$	$200 < p_T$	0.9982	0.0193	6.72e-05
$1.566 < \eta < 2.5$	$125 < p_T < 200$	0.9887	0.0097	-
$1.566 < \eta < 2.5$	$200 < p_T < 300$	0.9885	0.0223	9.24e-05

APPENDIX B

DATASETS AND TRIGGER PATHS

Events containing two prompt photon events are selected online with `HLT_DoublePhoton60` or `HLT_DoublePhoton70` trigger paths for 2016 and 2017/2018 data respectively. These two triggers require at least two reconstructed photon candidates with a transverse momentum $p_T > 60(70)$ GeV across the range $|\eta| < 3.0$. At this level, the primary selection requires that each photon candidate must have a ratio of the hadronic to the electromagnetic energy (H/E) < 0.15 . The electromagnetic energy here is defined as the corrected supercluster energy while the hadronic energy is the total HCAL energy within a cone of $\Delta R < 0.15$ from the supercluster position. This ratio is also the amount of the HCAL energy in the tower directly behind the photon seed crystal. This trigger is seeded by the logical `OR` of a suite `SingleEG`, `DoubleEG`, `SingleJet`, and `SingleTau` L1 seeds with varying trigger thresholds.

To quantify the efficiency of the `HLT_DoublePhoton60/70` triggers, we used the reference trigger `HLT_DoublePhoton33_CaloIdL` for normalization. The latter trigger selects events that contain two photons, each with transverse momentum greater than 33 GeV and satisfying the Calorimeter Identification (`CaloId`) criteria. The efficiency is primarily a function of the p_T of the second-leading photon, the photon which has the second highest p_T among the collection of photons. The results are shown in Figure 4.2. This procedure only probes the turn-on of the p_T leg [107] have shown that there is no additional source of efficiency and this study serves as a cross-check.

Additionally, we use `HLT_ECALHT800` as a backup trigger which adds approximately 0.2% in yield across the entire 2016-2018 data set. Although this comprises only a small fraction

of the total event yield, we choose to use it anyway in cases where the diphoton triggers lose efficiency specifically at high diphoton invariant masses. We calculate only the yield over the full analysis region with several thousand events, without viewing any distributions to avoid bias during the blinded phase of the analysis. The trigger efficiency is fully efficient at $p_T > 125$ GeV, which is the photon p_T selection used in this analysis.

B.1 Data Samples

The data samples considered for this analysis are listed in Table B.1 and correspond to an integrated luminosity of 138 fb^{-1} collected by the CMS experiment from 2016 to the end of 2018. The 2016 and 2017 data sets were processed in `CMSSW_9_4_13`, while the 2018 data sets are processed using `CMSSW_10_2_16`. The global tags used to analyze the data are listed in Table B.1. These samples also fulfill standard data quality criteria for all the subdetectors of CMS as specified by the good run JSONs listed in Table B.2. Calculation of the luminosity using the `HLT ECALHT800` trigger gives the same luminosity. The integrated luminosity of the MINIAOD Data samples were calculated with the `--hltpath HLT DoublePhoton60/70` option of `brilcalc` and the processed luminosity sections given by `crab --report`.

B.2 SM Prompt Diphoton Background Datasets

The Standard Model prompt diphoton background are listed in Tab. B.3. The campaigns and global tags used to generate the ntuples are listed in Tab. B.4. Here XXX stands for `RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3` for 2016 samples, `RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14` for 2017 samples or `RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1` for 2018 samples.

Table B.1: MINIAOD Data samples analyzed and the corresponding integrated luminosity, calculated with Double Photon Trigger and the processed luminosity sections.

Data set	Int. lumi (fb^{-1})	Global tag
/DoubleEG/Run2016B-17Jul2018_ver2-v1	5.781	94X_dataRun2_v10
/DoubleEG/Run2016C-17Jul2018_v1	2.573	94X_dataRun2_v10
/DoubleEG/Run2016D-17Jul2018_v1	4.248	94X_dataRun2_v10
/DoubleEG/Run2016E-17Jul2018_v1	4.009	94X_dataRun2_v10
/DoubleEG/Run2016F-17Jul2018_v1	3.102	94X_dataRun2_v10
/DoubleEG/Run2016G-17Jul2018_v1	7.540	94X_dataRun2_v10
/DoubleEG/Run2016H-17Jul2018_v1	8.606	94X_dataRun2_v10
2016 total integrated luminosity	35.86	
/DoubleEG/Run2017B-31Mar2018_v1	4.79	94X_dataRun2_v11
/DoubleEG/Run2017C-31Mar2018_v1	9.63	94X_dataRun2_v11
/DoubleEG/Run2017D-31Mar2018_v1	4.25	94X_dataRun2_v11
/DoubleEG/Run2017E-31Mar2018_v1	9.31	94X_dataRun2_v11
/DoubleEG/Run2017F-31Mar2018_v1	13.54	94X_dataRun2_v11
2017 total integrated luminosity	41.53	
/EGamma/Run2018A-17Sep2018-v2	13.70	102X_dataRun2_v11
/EGamma/Run2018B-17Sep2018-v1	7.06	102X_dataRun2_v11
/EGamma/Run2018C-17Sep2018-v1	6.89	102X_dataRun2_v11
/EGamma/Run2018D-22Jan2019-v2/MINIAOD	31.74	102X_dataRun2_Prompt_v11
2018 total integrated luminosity	59.40	

Table B.2: Good run JSONs used for each data period.

Data period	Good run JSON
2016	Collisions16/13TeV/ReReco/Final/Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt
2017	Collisions17/13TeV/ReReco/Final/Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt
2018	Collisions18/13TeV/ReReco/Final/Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt

Table B.3: MC background samples used in the analysis.

Object or process Cross-section (pb)	Data set path
/GG_M-500To1000_Pt70_TuneCP2_13TeV-pythia8/XXX	1.344e-01
/GG_M-1000To2000_Pt70_TuneCP2_13TeV-pythia8/XXX	1.359e-02
/GG_M-2000To4000_Pt70_TuneCP2_13TeV-pythia8/XXX	6.696e-04
/GG_M-4000To6000_Pt70_TuneCP2_13TeV-pythia8/XXX	8.757e-06
/GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8/XXX	1.862e+04
/GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/XXX	8.625e+03
/GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/XXX	2.194e+03
/GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/XXX	2.583e+02
/GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/XXX	8.520e+01
/GGJets_M-60To200_Pt-50_13TeV-sherpa/XXX	5.785e+00
/GGJets_M-200To500_Pt-50_13TeV-sherpa/XXX	2.244e+00
/GGJets_M-500To1000_Pt-50_13TeV-sherpa/XXX	1.510e-01
/GGJets_M-1000To2000_Pt-50_13TeV-sherpa/XXX	1.084e-02
/GGJets_M-2000To4000_Pt-50_13TeV-sherpa/XXX	3.690e-04
/GGJets_M-4000To6000_Pt-50_13TeV-sherpa/XXX	2.451e-06
/GGJets_M-6000To8000_Pt-50_13TeV-sherpa/XXX	1.753e-08
/GGJets_M-8000To13000_Pt-50_13TeV-sherpa/XXX	7.053e-11

Table B.4: For each year, the campaign in which MC samples were generated and the global tag used to generate the analysis ntuples.

Year	Campaign	Global tag
2016	RunIISummer16MiniAODv3	94X_dataRun2_v10
2017	RunIIFall17MiniAODv2	94X_dataRun2_v11
2018	RunIIAutumn18MiniAOD	102X_dataRun2_v11

B.3 Fake Rate Datasets

Table B.5 and Table B.6 lists the jet-triggered and muon-triggered datasets, respectively, that were used for the derivation of the fake rate.

B.4 Closure Test MC Datasets

The Closure test makes use of MC simulations rather than data in validating the Fake Rate method. The MC datasets used are listed in Table B.7 and Table B.8 for the 2016 and 2017-18, respectively.

Table B.5: The **JetHT** Datasets used for constructing the fake templates and for deriving the fake rates. The fake rate is derived from the average of the rates obtained independently from the **JetHT** and **DoubleMuon** dataset

Data set
/JetHT/Run2016B-17Jul2018_ver1-v1/MINIAOD
/JetHT/Run2016B-17Jul2018_ver2-v2/MINIAOD
/JetHT/Run2016C-17Jul2018-v1/MINIAOD
/JetHT/Run2016D-17Jul2018-v1/MINIAOD
/JetHT/Run2016E-17Jul2018-v1/MINIAOD
/JetHT/Run2016F-17Jul2018-v1/MINIAOD
/JetHT/Run2016G-17Jul2018-v1/MINIAOD
/JetHT/Run2016H-17Jul2018-v1/MINIAOD
/JetHT/Run2017B-31Mar2018-v1/MINIAOD
/JetHT/Run2017C-31Mar2018-v1/MINIAOD
/JetHT/Run2017D-31Mar2018-v1/MINIAOD
/JetHT/Run2017E-31Mar2018-v1/MINIAOD
/JetHT/Run2017F-31Mar2018-v1/MINIAOD
/JetHT/Run2018A-17Sep2018-v1/MINIAOD
/JetHT/Run2018B-17Sep2018-v1/MINIAOD
/JetHT/Run2018C-17Sep2018-v1/MINIAOD
/JetHT/Run2018D-PromptReco-v2/MINIAOD

Table B.6: The **DoubleMu**on datasets for fake rate estimation. The fake rate is derived from the average of the rates obtained independently from the **JetHT** and **DoubleMu**on datasets.

Data set
/DoubleMuon/Run2016B-17Jul2018_ver1-v1/MINIAOD
/DoubleMuon/Run2016B-17Jul2018_ver2-v1/MINIAOD
/DoubleMuon/Run2016C-17Jul2018-v1/MINIAOD
/DoubleMuon/Run2016D-17Jul2018-v1/MINIAOD
/DoubleMuon/Run2016E-17Jul2018-v1/MINIAOD
/DoubleMuon/Run2016F-17Jul2018-v1/MINIAOD
/DoubleMuon/Run2016G-17Jul2018-v1/MINIAOD
/DoubleMuon/Run2016H-17Jul2018-v1/MINIAOD
/DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD
/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD
/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD
/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD
/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD
/DoubleMuon/Run2018A-17Sep2018-v2/MINIAOD
/DoubleMuon/Run2018B-17Sep2018-v1/MINIAOD
/DoubleMuon/Run2018C-17Sep2018-v1/MINIAOD
/DoubleMuon/Run2018D-PromptReco-v2/MINIAOD

Table B.7: 2016 MC Fakes

Data set	Cross section (pb)
/QCD_Pt_5to10_TuneCUETP8M1	61018300000.
/QCD_Pt_10to15_TuneCUETP8M1	5887580000.
/QCD_Pt_15to30_TuneCUETP8M1	1837410000.
/QCD_Pt_30to50_TuneCUETP8M1	140932000.
/QCD_Pt_50to80_TuneCUETP8M1	19204300.
/QCD_Pt_80to120_TuneCUETP8M1	2762530.
/QCD_Pt_120to170_TuneCUETP8M1	471100.
/QCD_Pt_170to300_TuneCUETP8M1	117276.
/QCD_Pt_300to470_TuneCUETP8M1	7823.
/QC_Pt_470to600_TuneCUETP8M1	648.2
/QCD_Pt_600to800_TuneCUETP8M1	186.9
/QCD_Pt_800to1000_TuneCUETP8M1	32.293
/QCD_Pt_1000to1400_TuneCUETP8M1	9.4183
/QCD_Pt_1400to1800_TuneCUETP8M1	0.84265
/QCD_Pt_1800to2400_TuneCUETP8M1	0.114943
/QCD_Pt_2400to3200_TuneCUETP8M1	0.00682981
/QCD_Pt_3200toInf_TuneCUETP8M1	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

Table B.8: 2017 and 2018 MC Fakes

Data set	Cross section (pb)
/QCD_Pt_15to30_TuneCP5_13TeV_pythia8	1246000000.0
/QCD_Pt_30to50_TuneCP5_13TeV_pythia8	106900000.0
/QCD_Pt_50to80_TuneCP5_13TeV_pythia8	15710000.0
/QCD_Pt_80to120_TuneCP5_13TeV_pythia8	2336000.0
/QCD_Pt_120to170_TuneCP5_13TeV_pythia8	407300.0
/QCD_Pt_170to300_TuneCP5_13TeV_pythia8	103500.0
/QCD_Pt_300to470_TuneCP5_13TeV_pythia8	6830.0
/QCD_Pt_470to600_TuneCP5_13TeV_pythia8	552.1
/QCD_Pt_600to800_TuneCP5_13TeV_pythia8	156.5
/QCD_Pt_800to1000_TuneCP5_13TeV_pythia8	26.28
/QCD_Pt_1000to1400_TuneCP5_13TeV_pythia8	7.47
/QCD_Pt_1400to1800_TuneCP5_13TeV_pythia8	0.6484
/QCD_Pt_1800to2400_TuneCP5_13TeV_pythia8	0.08743
/QCD_Pt_2400to3200_TuneCP5_13TeV_pythia8	0.005236
/QCD_Pt_3200toInf_TuneCP5_13TeV_pythia8	0.0001357
/GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8	1.862e+04
/GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	8.625e+03
/GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2.194e+03
/GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	2.583e+02
/GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	8.520e+01

APPENDIX C

FAKE RATE APPLICATION

In Sec. 5.3 we defined the photon fake rate as

$$f = \frac{\text{number of jets passing high } p_T \text{ ID}}{\text{number of jets passing denominator definition}}. \quad (\text{C.1})$$

The derivations for Eq. 5.2, here is followed from Appendix A of Ref. [146]. Knowing that the fake rate is defined as in eq. C.1, the probability that a known jet is tight (T) or loose (L) are respectively given by

$$\begin{aligned} P(T|j) &= \frac{f}{1+f}, \\ P(L|j) &= \frac{1}{1+f}. \end{aligned} \quad (\text{C.2})$$

Here Tight (T) refers to objects passing the photon ID, while Loose (L) refers to objects passing the denominator criteria. If the object is a known photon, then $P(T|\gamma) = 1$ and $P(L|\gamma) = 0$.

Using the conditional probabilities above, we can write the number of Tight-Tight, Tight-Loose, Loose-Tight, Loose-Loose object pairs as follows,

$$\begin{aligned} N_{TT} &= N_{\gamma\gamma} + N_{\gamma j}P_2(T|j) + N_{j\gamma}P_1(T|j) + N_{jj}P_1(T|j)P_2(T|j), \\ N_{TL} &= N_{\gamma j}P_2(T|j) + N_{jj}P_1(T|j)P_2(T|j), \\ N_{LT} &= N_{j\gamma}P_1(T|j) + N_{jj}P_1(T|j)P_2(T|j), \\ N_{LL} &= N_{jj}P_1(T|j)P_2(T|j). \end{aligned} \quad (\text{C.3})$$

The subscripts 1 and 2 denote whether the object is leading or subleading, respectively. From the previous equations, it becomes clear that the number of jj events are:

$$N_{jj} = N_{LL}(1 + f_1)(1 + f_2). \quad (\text{C.4})$$

Making a couple more substitutions in the set of Eqs. C.3, we have

$$\begin{aligned} N_{\gamma j} &= N_{TL}(1 + f_2) - N_{LL}f_1(1 + f_2), \\ N_{j\gamma} &= N_{LT}(1 + f_1) - N_{LL}f_2(1 + f_1). \end{aligned} \quad (\text{C.5})$$

To get the number of diphotons $N_{\gamma\gamma}$, we use the equation for N_{TT} which is a function of Eqs. C.5 and get

$$\begin{aligned} N_{\gamma\gamma} &= N_{TT} - N_{TL}f_2 - N_{LT}f_1 + N_{LL}f_1f_2 \\ &= N_{TT} - N_{\text{fake}}. \end{aligned} \quad (\text{C.6})$$

This equation shows the terms that must be subtracted from the tight-tight contributions and we denote them as N_{fake} . From here we will take N_{TL} as either Loose-Tight or Tight-Loose and write N_{fake} from Eq. as follows:

$$\begin{aligned} N_{\text{fake}} &= N_{TL}f_2 - N_{LT}f_1 + N_{LL}f_1f_2, \\ &= N_{TL}f - N_{LL}f_1f_2. \end{aligned} \quad (\text{C.7})$$

Here, f is the fake rate evaluated at the p_T of the Loose object. The fake background can be estimated as a sum of the $\gamma + j$ and jj components. So, $N_{\text{fake}} = N_{jj} + N_{\gamma j}$. We arrive at the final expression used for our background estimation for γj and jj events.

$$\begin{aligned} N_{jj} &= N_{LL}f_1f_2, \\ N_{\gamma j} &= N_{LT}f - 2N_{LL}f_1f_2. \end{aligned} \quad (\text{C.8})$$

C.1 Signal Monte-Carlo

This appendix contains the full list of signal sample names and their associated cross sections. The same information can also be found on the project's GitHub repository in the file CrossSections.h.

The Dataset paths for the signal interpretations used in this analysis are shown below. Nonresonant signal samples were generated for the following campaigns:

1. Summer 2016: RunIISummer16MiniA0Dv3
2. Fall 2017: RunIIFall17MiniA0Dv2
3. Autumn 2018: RunIIAutumn18MiniA0D.

To span the entire mass spectrum, samples were produced in split mass bins from $m_{\gamma\gamma} = 500 \text{ GeV}$ up to the value of Λ_T . The associated cross sections for the Unparticles are found in this link.

Table C.1: Nonresonant signal samples were generated for the Summer 2016, Fall 2017 and Autumn 2018 campaigns.

Data set
/ADDGravToGG_NegInt-0_LambdaT-XXX_M-XXXToXXXX_TuneCP2_13TeV-pythia8
/ADDGravToGG_NegInt-1_LambdaT-XXX_M-XXXToXXX_TuneCP2_13TeV-pythia8

Table C.2: Unparticles samples were generated for the 2017/18 campaigns.

Data set
/UnparToGG_Spin0_duXXX_LambdaU-XXX_pT70_MXXX-XXX_TuneCP2_13TeV-pythia8
/UnparToGG_Spin2_duXXX_LambdaU-XXX_pT70_MXXX-XXX_TuneCP2_13TeV-pythia8

Table C.3: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-0_LambdaT-6000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.272e-01
/ADDGravToGG_NegInt-0_LambdaT-6000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.284e-02
/ADDGravToGG_NegInt-0_LambdaT-6000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.759e-04
/ADDGravToGG_NegInt-0_LambdaT-6000_M-4000To6000_TuneCP2_13TeV-pythia8/XXX	1.383e-04
/ADDGravToGG_NegInt-0_LambdaT-6500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.272e-01
/ADDGravToGG_NegInt-0_LambdaT-6500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.301e-02
/ADDGravToGG_NegInt-0_LambdaT-6500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.195e-04
/ADDGravToGG_NegInt-0_LambdaT-6500_M-4000To6500_TuneCP2_13TeV-pythia8/XXX	8.238e-05
/ADDGravToGG_NegInt-0_LambdaT-7000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.272e-01
/ADDGravToGG_NegInt-0_LambdaT-7000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.305e-02
/ADDGravToGG_NegInt-0_LambdaT-7000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.033e-04
/ADDGravToGG_NegInt-0_LambdaT-7000_M-4000To7000_TuneCP2_13TeV-pythia8/XXX	4.982e-05
/ADDGravToGG_NegInt-0_LambdaT-7500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.273e-01
/ADDGravToGG_NegInt-0_LambdaT-7500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.315e-02
/ADDGravToGG_NegInt-0_LambdaT-7500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.057e-04
/ADDGravToGG_NegInt-0_LambdaT-7500_M-4000To7500_TuneCP2_13TeV-pythia8/XXX	3.128e-05

Table C.4: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-0_LambdaT-6000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.272e-01
/ADDGravToGG_NegInt-0_LambdaT-6000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.284e-02
/ADDGravToGG_NegInt-0_LambdaT-6000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.759e-04
/ADDGravToGG_NegInt-0_LambdaT-6000_M-4000To6000_TuneCP2_13TeV-pythia8/XXX	1.383e-04
/ADDGravToGG_NegInt-0_LambdaT-6500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.272e-01
/ADDGravToGG_NegInt-0_LambdaT-6500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.301e-02
/ADDGravToGG_NegInt-0_LambdaT-6500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.195e-04
/ADDGravToGG_NegInt-0_LambdaT-6500_M-4000To6500_TuneCP2_13TeV-pythia8/XXX	8.238e-05
/ADDGravToGG_NegInt-0_LambdaT-7000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.272e-01
/ADDGravToGG_NegInt-0_LambdaT-7000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.305e-02
/ADDGravToGG_NegInt-0_LambdaT-7000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.033e-04
/ADDGravToGG_NegInt-0_LambdaT-7000_M-4000To7000_TuneCP2_13TeV-pythia8/XXX	4.982e-05
/ADDGravToGG_NegInt-0_LambdaT-7500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.273e-01
/ADDGravToGG_NegInt-0_LambdaT-7500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.315e-02
/ADDGravToGG_NegInt-0_LambdaT-7500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.057e-04
/ADDGravToGG_NegInt-0_LambdaT-7500_M-4000To7500_TuneCP2_13TeV-pythia8/XXX	3.128e-05

Table C.5: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-0_LambdaT-8000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.274e-01
/ADDGravToGG_NegInt-0_LambdaT-8000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.319e-02
/ADDGravToGG_NegInt-0_LambdaT-8000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.115e-04
/ADDGravToGG_NegInt-0_LambdaT-8000_M-4000To8000_TuneCP2_13TeV-pythia8/XXX	2.061e-05
/ADDGravToGG_NegInt-0_LambdaT-9000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.277e-01
/ADDGravToGG_NegInt-0_LambdaT-9000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.325e-02
/ADDGravToGG_NegInt-0_LambdaT-9000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.269e-04
/ADDGravToGG_NegInt-0_LambdaT-9000_M-4000To9000_TuneCP2_13TeV-pythia8/XXX	1.120e-05
/ADDGravToGG_NegInt-0_LambdaT-10000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.277e-01
/ADDGravToGG_NegInt-0_LambdaT-10000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.328e-02
/ADDGravToGG_NegInt-0_LambdaT-10000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.403e-04
/ADDGravToGG_NegInt-0_LambdaT-10000_M-4000To10000_TuneCP2_13TeV-pythia8/XXX	8.330e-06
/ADDGravToGG_NegInt-0_LambdaT-11000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.274e-01
/ADDGravToGG_NegInt-0_LambdaT-11000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.332e-02
/ADDGravToGG_NegInt-0_LambdaT-11000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.500e-04
/ADDGravToGG_NegInt-0_LambdaT-11000_M-4000To11000_TuneCP2_13TeV-pythia8/XXX	7.580e-06

Table C.6: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-0_LambdaT-13000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.277e-01
/ADDGravToGG_NegInt-0_LambdaT-13000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.331e-02
/ADDGravToGG_NegInt-0_LambdaT-13000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.594e-04
/ADDGravToGG_NegInt-0_LambdaT-13000_M-4000To13000_TuneCP2_13TeV-pythia8/XXX	7.635e-06
/ADDGravToGG_NegInt-0_LambdaT-13000_M-4000To12990_TuneCP2_13TeV-pythia8/XXX	7.623e-06
/ADDGravToGG_NegInt-1_LambdaT-4000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.311e-01
/ADDGravToGG_NegInt-1_LambdaT-4000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.903e-02
/ADDGravToGG_NegInt-1_LambdaT-4000_M-2000To3000_TuneCP2_13TeV-pythia8/XXX	5.572e-03
/ADDGravToGG_NegInt-1_LambdaT-4000_M-3000To4000_TuneCP2_13TeV-pythia8/XXX	3.809e-03
/ADDGravToGG_NegInt-1_LambdaT-4500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.296e-01
/ADDGravToGG_NegInt-1_LambdaT-4500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.630e-02
/ADDGravToGG_NegInt-1_LambdaT-4500_M-2000To3000_TuneCP2_13TeV-pythia8/XXX	2.800e-03
/ADDGravToGG_NegInt-1_LambdaT-4500_M-3000To4500_TuneCP2_13TeV-pythia8/XXX	2.166e-03
/ADDGravToGG_NegInt-1_LambdaT-5000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.291e-01
/ADDGravToGG_NegInt-1_LambdaT-5000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.507e-02

Table C.7: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-1_LambdaT-5000_M-2000To3000_TuneCP2_13TeV-pythia8/XXX	1.707e-03
/ADDGravToGG_NegInt-1_LambdaT-5000_M-3000To5000_TuneCP2_13TeV-pythia8/XXX	1.239e-03
/ADDGravToGG_NegInt-1_LambdaT-5500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.287e-01
/ADDGravToGG_NegInt-1_LambdaT-5500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.443e-02
/ADDGravToGG_NegInt-1_LambdaT-5500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	1.643e-03
/ADDGravToGG_NegInt-1_LambdaT-5500_M-4000To5500_TuneCP2_13TeV-pythia8/XXX	3.124e-04
/ADDGravToGG_NegInt-1_LambdaT-6000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.282e-01
/ADDGravToGG_NegInt-1_LambdaT-6000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.409e-02
/ADDGravToGG_NegInt-1_LambdaT-6000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	1.234e-03
/ADDGravToGG_NegInt-1_LambdaT-6000_M-4000To6000_TuneCP2_13TeV-pythia8/XXX	2.006e-04
/ADDGravToGG_NegInt-1_LambdaT-6500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.284e-01
/ADDGravToGG_NegInt-1_LambdaT-6500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.386e-02
/ADDGravToGG_NegInt-1_LambdaT-6500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	1.028e-03
/ADDGravToGG_NegInt-1_LambdaT-6500_M-4000To6500_TuneCP2_13TeV-pythia8/XXX	1.296e-04

Table C.8: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-1_LambdaT-7000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.280e-01
/ADDGravToGG_NegInt-1_LambdaT-7000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.372e-02
/ADDGravToGG_NegInt-1_LambdaT-7000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	9.073e-04
/ADDGravToGG_NegInt-1_LambdaT-7000_M-4000To7000_TuneCP2_13TeV-pythia8/XXX	8.497e-05
/ADDGravToGG_NegInt-1_LambdaT-7500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.280e-01
/ADDGravToGG_NegInt-1_LambdaT-7500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.362e-02
/ADDGravToGG_NegInt-1_LambdaT-7500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	8.364e-04
/ADDGravToGG_NegInt-1_LambdaT-7500_M-4000To7500_TuneCP2_13TeV-pythia8/XXX	5.808e-05
/ADDGravToGG_NegInt-1_LambdaT-8000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.280e-01
/ADDGravToGG_NegInt-1_LambdaT-8000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.354e-02
/ADDGravToGG_NegInt-1_LambdaT-8000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	7.903e-04
/ADDGravToGG_NegInt-1_LambdaT-8000_M-4000To8000_TuneCP2_13TeV-pythia8/XXX	4.137e-05
/ADDGravToGG_NegInt-1_LambdaT-8500_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.277e-01
/ADDGravToGG_NegInt-1_LambdaT-8500_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.350e-02
/ADDGravToGG_NegInt-1_LambdaT-8500_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	7.583e-04
/ADDGravToGG_NegInt-1_LambdaT-8500_M-4000To8500_TuneCP2_13TeV-pythia8/XXX	3.088e-05
/ADDGravToGG_NegInt-1_LambdaT-9000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.278e-01

Table C.9: ADD signal samples and their associated cross-sections

Data set path	Cross section (pb)
/ADDGravToGG_NegInt-1_LambdaT-9000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.351e-02
/ADDGravToGG_NegInt-1_LambdaT-9000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	7.384e-04
/ADDGravToGG_NegInt-1_LambdaT-9000_M-4000To9000_TuneCP2_13TeV-pythia8/XXX	2.410e-05
/ADDGravToGG_NegInt-1_LambdaT-10000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.280e-01
/ADDGravToGG_NegInt-1_LambdaT-10000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.347e-02
/ADDGravToGG_NegInt-1_LambdaT-10000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	7.144e-04
/ADDGravToGG_NegInt-1_LambdaT-10000_M-4000To10000_TuneCP2_13TeV-pythia8/XXX	1.679e-05
/ADDGravToGG_NegInt-1_LambdaT-11000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.275e-01
/ADDGravToGG_NegInt-1_LambdaT-11000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.342e-02
/ADDGravToGG_NegInt-1_LambdaT-11000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.984e-04
/ADDGravToGG_NegInt-1_LambdaT-11000_M-4000To11000_TuneCP2_13TeV-pythia8/XXX	1.336e-05
/ADDGravToGG_NegInt-1_LambdaT-13000_M-500To1000_TuneCP2_13TeV-pythia8/XXX	1.278e-01
/ADDGravToGG_NegInt-1_LambdaT-13000_M-1000To2000_TuneCP2_13TeV-pythia8/XXX	1.340e-02
/ADDGravToGG_NegInt-1_LambdaT-13000_M-2000To4000_TuneCP2_13TeV-pythia8/XXX	6.859e-04
/ADDGravToGG_NegInt-1_LambdaT-13000_M-4000To13000_TuneCP2_13TeV-pythia8/XXX	1.059e-05
/ADDGravToGG_NegInt-1_LambdaT-13000_M-4000To12990_TuneCP2_13TeV-pythia8/XXX	1.061e-05

Table C.10: Unparticles signal samples and their associated cross-sections

Data set path	Cross section (pb)
UnparToGG_Spin0_dulp1_LambdaU-10000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.349e-02
UnparToGG_Spin0_dulp1_LambdaU-10000_pT70_M2000-4000_TuneCP2_13TeV_pythia8/XXX	7.881e-04
UnparToGG_Spin0_dulp1_LambdaU-10000_pT70_M4000_TuneCP2_13TeV_pythia8/XXX	1.138e-05
UnparToGG_Spin0_dulp1_LambdaU-10000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.313e-01
UnparToGG_Spin0_dulp1_LambdaU-4000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.736e-02
UnparToGG_Spin0_dulp1_LambdaU-4000_pT70_M2000-4000_TuneCP2_13TeV_pythia8/XXX	1.356e-03
UnparToGG_Spin0_dulp1_LambdaU-4000_pT70_M4000_TuneCP2_13TeV_pythia8/XXX	3.384e-05
UnparToGG_Spin0_dulp1_LambdaU-4000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.464e-01
UnparToGG_Spin0_dulp1_LambdaU-8000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.378e-02
UnparToGG_Spin0_dulp1_LambdaU-8000_pT70_M2000-4000_TuneCP2_13TeV_pythia8/XXX	7.663e-04
UnparToGG_Spin0_dulp1_LambdaU-8000_pT70_M4000_TuneCP2_13TeV_pythia8/XXX	1.301e-05
UnparToGG_Spin0_dulp1_LambdaU-8000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.345e-01
UnparToGG_Spin0_dulp5_LambdaU-2000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.346e-02
UnparToGG_Spin0_dulp5_LambdaU-2000_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	9.939e-04
UnparToGG_Spin0_dulp5_LambdaU-2000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.309e-01
UnparToGG_Spin0_dulp5_LambdaU-2500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.483e-02
UnparToGG_Spin0_dulp5_LambdaU-2500_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	7.909e-04
UnparToGG_Spin0_dulp5_LambdaU-2500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.308e-01
UnparToGG_Spin0_dulp5_LambdaU-3500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.477e-02

Table C.11: Unparticles signal samples and their associated cross-sections

Data set path	Cross section (pb)
UnparToGG_Spin0_du1p5_LambdaU-3500_pT70_M2000-3000_TuneCP2_13TeV_pythia8/XXX	6.394e-04
UnparToGG_Spin0_du1p5_LambdaU-3500_pT70_M3000_TuneCP2_13TeV_pythia8/XXX	7.517e-05
UnparToGG_Spin0_du1p5_LambdaU-3500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.308e-01
UnparToGG_Spin0_du1p9_LambdaU-2000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.483e-02
UnparToGG_Spin0_du1p9_LambdaU-2000_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	1.043e-03
UnparToGG_Spin0_du1p9_LambdaU-2000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.308e-01
UnparToGG_Spin0_du1p9_LambdaU-2500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.477e-02
UnparToGG_Spin0_du1p9_LambdaU-2500_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	7.480e-04
UnparToGG_Spin0_du1p9_LambdaU-2500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.308e-01
UnparToGG_Spin0_du1p9_LambdaU-3500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.326e-02
UnparToGG_Spin0_du1p9_LambdaU-3500_pT70_M2000-3000_TuneCP2_13TeV_pythia8/XXX	6.198e-04
UnparToGG_Spin0_du1p9_LambdaU-3500_pT70_M3000_TuneCP2_13TeV_pythia8/XXX	6.923e-05

Table C.12: Unparticles signal samples and their associated cross-sections

Data set path	Cross section (pb)
UnparToGG_Spin0_dulp9_LambdaU-3500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.308e-01
UnparToGG_Spin2_dulp1_LambdaU-2000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.574e-02
UnparToGG_Spin2_dulp1_LambdaU-2500_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	9.181e-04
UnparToGG_Spin2_dulp1_LambdaU-2500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.270e-02
UnparToGG_Spin2_dulp1_LambdaU-2000_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	1.819e-03
UnparToGG_Spin2_dulp1_LambdaU-2000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.260e-01
UnparToGG_Spin2_dulp1_LambdaU-3000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.309e-01
UnparToGG_Spin2_dulp1_LambdaU-3000_pT70_M3000_TuneCP2_13TeV_pythia8/XXX	9.096e-05
UnparToGG_Spin2_dulp1_LambdaU-3000_pT70_M2000-3000_TuneCP2_13TeV_pythia8/XXX	6.176e-04
UnparToGG_Spin2_dulp1_LambdaU-2500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.373e-01
UnparToGG_Spin2_dulp1_LambdaU-3000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.163e-02
UnparToGG_Spin2_dulp5_LambdaU-2000_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	9.209e-04
UnparToGG_Spin2_dulp5_LambdaU-2000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.298e-01
UnparToGG_Spin2_dulp5_LambdaU-3000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.381e-02
UnparToGG_Spin2_dulp9_LambdaU-2000_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	9.490e-04
UnparToGG_Spin2_dulp5_LambdaU-2500_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	7.299e-04
UnparToGG_Spin2_dulp5_LambdaU-2000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.301e-02
UnparToGG_Spin2_dulp5_LambdaU-2500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.317e-02

Table C.13: Unparticles signal samples and their associated cross-sections

Data set path	Cross section (pb)
UnparToGG_Spin2_du1p5_LambdaU-3000_pT70_M2000-3000_TuneCP2_13TeV_pythia8/XXX	6.250e-04
UnparToGG_Spin2_du1p9_LambdaU-3500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.323e-02
UnparToGG_Spin2_du1p5_LambdaU-3000_pT70_M3000_TuneCP2_13TeV_pythia8/XXX	7.064e-05
UnparToGG_Spin2_du1p5_LambdaU-2500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.337e-01
UnparToGG_Spin2_du1p9_LambdaU-2500_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.334e-02
UnparToGG_Spin2_du1p9_LambdaU-2000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.311e-01
UnparToGG_Spin2_du1p9_LambdaU-2000_pT70_M1000-2000_TuneCP2_13TeV_pythia8/XXX	1.422e-02
UnparToGG_Spin2_du1p9_LambdaU-3500_pT70_M2000-3000_TuneCP2_13TeV_pythia8/XXX	6.346e-04
UnparToGG_Spin2_du1p5_LambdaU-3000_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.337e-01
UnparToGG_Spin2_du1p9_LambdaU-3500_pT70_M3000_TuneCP2_13TeV_pythia8/XXX	7.395e-05
UnparToGG_Spin2_du1p9_LambdaU-2500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.308e-01
UnparToGG_Spin2_du1p9_LambdaU-2500_pT70_M2000_TuneCP2_13TeV_pythia8/XXX	7.575e-04
UnparToGG_Spin2_du1p9_LambdaU-3500_pT70_M500-1000_TuneCP2_13TeV_pythia8/XXX	1.338e-01

APPENDIX D

SAMPLE CARDS

D.1 ADD Graviton to $\gamma\gamma$

An example Pythia Card for the ADD Graviton to $\gamma\gamma$. Additional cards are found in GitHub.

```
import FWCore.ParameterSet.Config as cms

from Configuration.Generator.Pythia8CommonSettings_cfi import *
from Configuration.Generator.MCTunes2017.PythiaCP2Settings_cfi import *

generator = cms.EDFilter("Pythia8GeneratorFilter",
    maxEventsToPrint = cms.untracked.int32(1),
    pythiaPylistVerbosity = cms.untracked.int32(1),
    filterEfficiency = cms.untracked.double(1.0),
    pythiaHepMCVerbosity = cms.untracked.bool(False),
    comEnergy = cms.double(13000.),
    PythiaParameters = cms.PSet(
        pythia8CommonSettingsBlock,
        pythia8CP2SettingsBlock,
        processParameters = cms.vstring(
            'ExtraDimensionsLED:LambdaT = 10000.0',
            'ExtraDimensionsLED:n = 2',
            'ExtraDimensionsLED:ffbar2gammagamma = on',
            'ExtraDimensionsLED:gg2gammagamma = on',
            'ExtraDimensionsLED:CutOffmode = 2',
            'ExtraDimensionsLED:NegInt= 0',
            'PhaseSpace:pTHatMin = 70.0',
            'PhaseSpace:mHatMin = 1000.0',
            'PhaseSpace:mHatMax = 2000.0',
        ),
    ),
```

```

        parameterSets = cms.vstring('pythia8CommonSettings',
                                    'pythia8CP2Settings',
        'processParameters',
        )
    )
)

```

D.2 Unparticles to $\gamma\gamma$

For the PYTHIA implementation of the Spin-0 case we include both the gluon fusion (gg2gammagamma) and the quark-antiquark (ffbar2gammagamma) annihilation process on. In the Spin-2 case, the (ffbar2gammagamma) is already included in the Unparticles implementation and could be removed from the pythia card.

```

import FWCore.ParameterSet.Config as cms

from Configuration.Generator.Pythia8CommonSettings_cfi import *
from Configuration.Generator.MCTunes2017.PythiaCP2Settings_cfi import *
from Configuration.Generator.....

generator = cms.EDFilter("Pythia8GeneratorFilter",
    maxEventsToPrint = cms.untracked.int32(1),
    pythiaPylistVerbosity = cms.untracked.int32(1),
    filterEfficiency = cms.untracked.double(1.0),
    pythiaHepMCVerbosity = cms.untracked.bool(False),
    comEnergy = cms.double(13000.),
    PythiaParameters = cms.PSet(
        pythia8CommonSettingsBlock,
        pythia8CP2SettingsBlock,
        pythia8PSweightsSettingsBlock,
        processParameters = cms.vstring(
            'ExtraDimensionsUnpart:ffbar2gammagamma = on',
            'ExtraDimensionsUnpart:gg2gammagamma = on',
            '\textbf{'PromptPhoton:gg2gammagamma = on'},}
            '\textbf{'PromptPhoton:ffbar2gammagamma = on'},}
        'ExtraDimensionsUnpart:LambdaU = 2500.0',
        'ExtraDimensionsUnpart:lambda = 1.0',
        'ExtraDimensionsUnpart:dU = 1.9',
        'ExtraDimensionsUnpart:spinU = 0',
        'PhaseSpace:pTHatMin = 70',
    )
)

```

```

        'PhaseSpace:mHatMin = 2000',
        'PhaseSpace:mHatMax = 1',
    ),
    parameterSets = cms.vstring('pythia8CommonSettings',
                                'pythia8CP2Settings',
                                'processParameters',
                                'pythia8PSweightsSettings',
                                )
)

ProductionFilterSequence = cms.Sequence(generator)

```

APPENDIX E

PRELIMINARY RESULTS USING PSEUDODATA

CMS Collaboration papers go through a series of steps summarized in Fig. E.1 before they get submitted to a journal. When an analysis has reached a certain maturity, the Physics Working group and the analyzers make sure that the simulations had taken into account detector malfunctions/resolution in the preliminary conclusions, the physics working group greenlights the project for “unblinding” in the preapproval stage. Prior to this stage, the analysis results remained blinded for $m_{\gamma\gamma}$ above 1 TeV. In lieu of the real CMS data, pseudodata have been generated to test and validate our prediction.

The pseudodata have been generated using the sum of the $\gamma + \gamma j + jj$ spectra or what we call the prefit background estimate. For each bin, a poisson distribution with a mean equal to the prefit value, is used to generate random integer yields. In this way, the pseudodata feature realistic systematic structures and fluctuations. Thus, we can check whether prefit trends in pulls are covered by the current parametrization.

For the limit setting, the $m_{\gamma\gamma}$ spectra shown in Fig. E.2 are used. The six $m_{\gamma\gamma}$ spectra are binned in 100 GeV sizes, starting at 500 GeV and ending at 4000 GeV. Last bin contains the overflow above 4000 GeV.

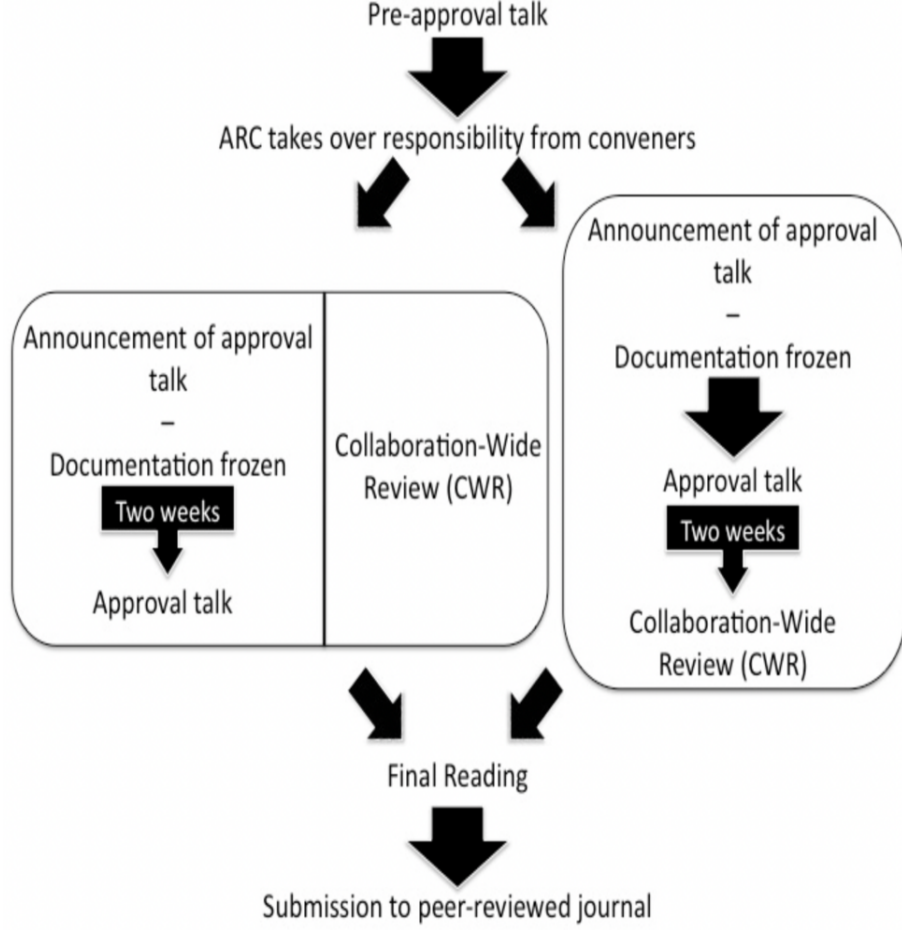


Figure E.1: CMS Collaboration papers go pre-approval, approval, a Collaboration-Wide Review, and a final reading prior to submission to a peer-reviewed journal [147].

E.0.1 Clockwork Results with Pseudodata

For the Continuum Clockwork model, we start the fit from 1 TeV as we are constrained by the available ADD samples. Below this, statistical fluctuations are too high to arrive at any meaningful conclusions. In running THETA, the nuisance parameters are simultaneously marginalized in a bayesian marginalization constrain fit, and a posterior prediction for the combined diphoton+fake $m_{\gamma\gamma}$ spectra produced using the pseudodata as the constraint. The signal strength is given a flat prior bounded below by zero, although the background normalization is allowed to float freely (within 5%). This approach presupposes that the

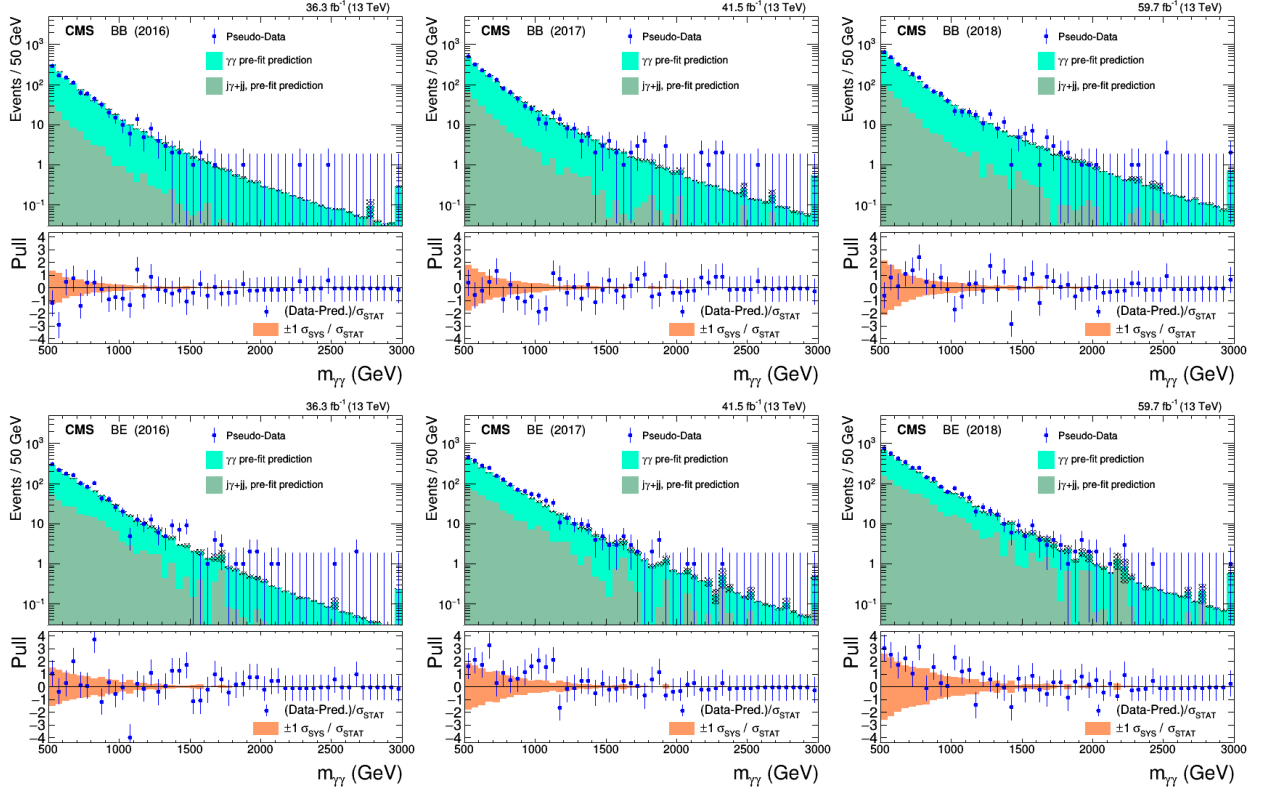


Figure E.2: The prefit spectra with the generated pseudodata for BB (top) and BE (bottom) for the tree years of data taking (left to right: 2016, 2017, 2018).

signal and background differ in their shape and not only in the rate. A benefit from the shape-based analysis is that we are able to absorb higher-order corrections to the diphoton prediction that are (in principle) unknown directly into the background model under the assumption that those higher-order corrections resulted in a relatively flat k-factor.

The bayesian marginalization gives out the postfit results which are presented in Fig. E.3. The bottom pads show that the pulls or the various systematic trends initially present in the prefit pulls are smoothened out in the postfit version. For pseudodata, the origin of these trends is random and arbitrary. The pseudodata happened to deviate in the same direction for a few sequential bins. The postfit pulls exhibit only statistical fluctuations around one, and thus the $m_{\gamma\gamma}$ parametrization absorbs the systematic effects.

In this portion, we make additional commentaries on the nuisance parameters considered in the marginalization procedure. In Fig. E.3, we see that the signal strength β , named as **beta_signal**, shown in the top left is strongly constrained near zero as expected. Normalization nuisances of real $\gamma\gamma$ and fake background events are constrained and their postfit pdfs are close to zero some value dictated by the fit on pseudodata. Figures E.4 and E.5, correspond to the 50 pdfs of the PDF nuisances. They are all close to the prefit and appear as a gaussian with mean zero and standard deviation of one as this set of pseudodata cannot constrain them further.

Note here that the nuisances are constrained and their priors which are either gaussian or lognormal probability distribution functions transform into their corresponding postfit versions. While COMBINE, provides summary tables with the means and the standard deviations of the nuisances postfit pdfs, THETA provides the full postfit pdf spectra of each nuisance. Overall, the postfit results suggest a healthy performance of the limit code and fit

procedure which will be used in the actual data. Finally, Fig. E.7 shows a comparable shape trend with the 2016 result [131] which is shown in Fig. 2.14.

E.0.2 ADD Results with Pseudodata

For want of space, we exclude the pseudodata results for the ADD. But the procedure is the same as that for the Clockwork model except we begin below 1 TeV.

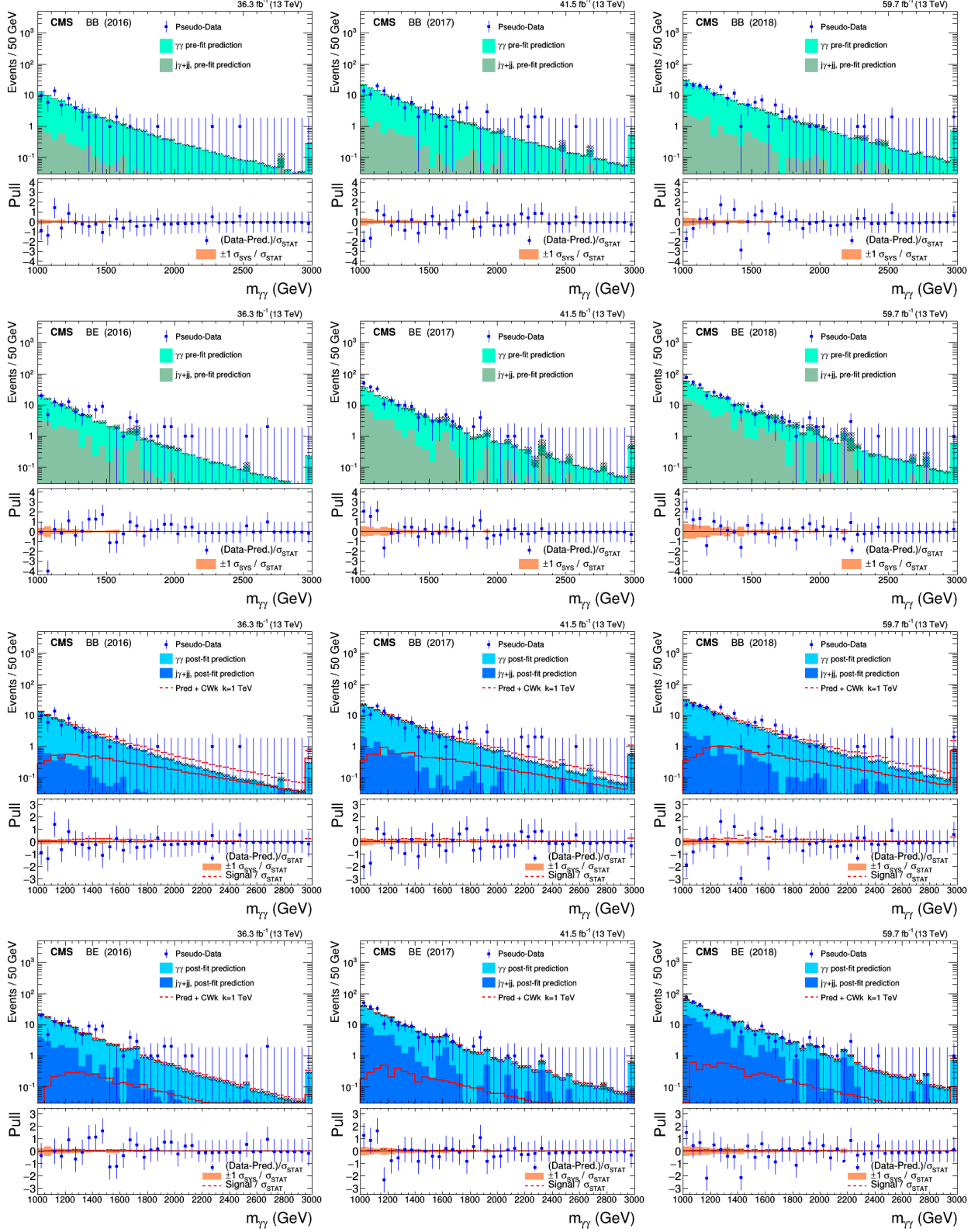


Figure E.3: The prefit (top two rows) and the postfit (bottom two rows) $m_{\gamma\gamma}$ spectra for the Clockwork Model using pseudodata. Columns left to right correspond to the years 2016, 2017, and 2018 respectively.

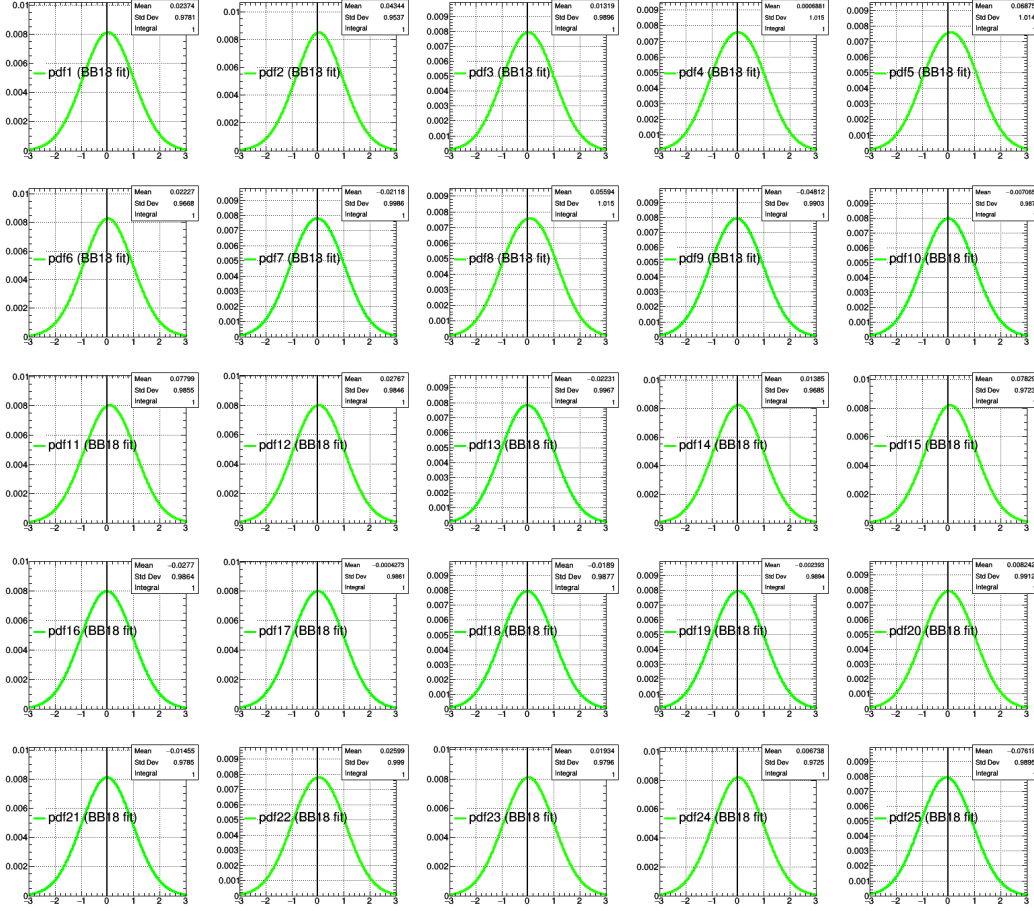


Figure E.4: The postfit probability distribution functions (pdfs) for the first 25 nuisances of the parton distribution functions (PDFs) in the six-SRs fit. The statistics panel displays the mean and the std of each pdf. This set of results corresponds to the pseudodata.

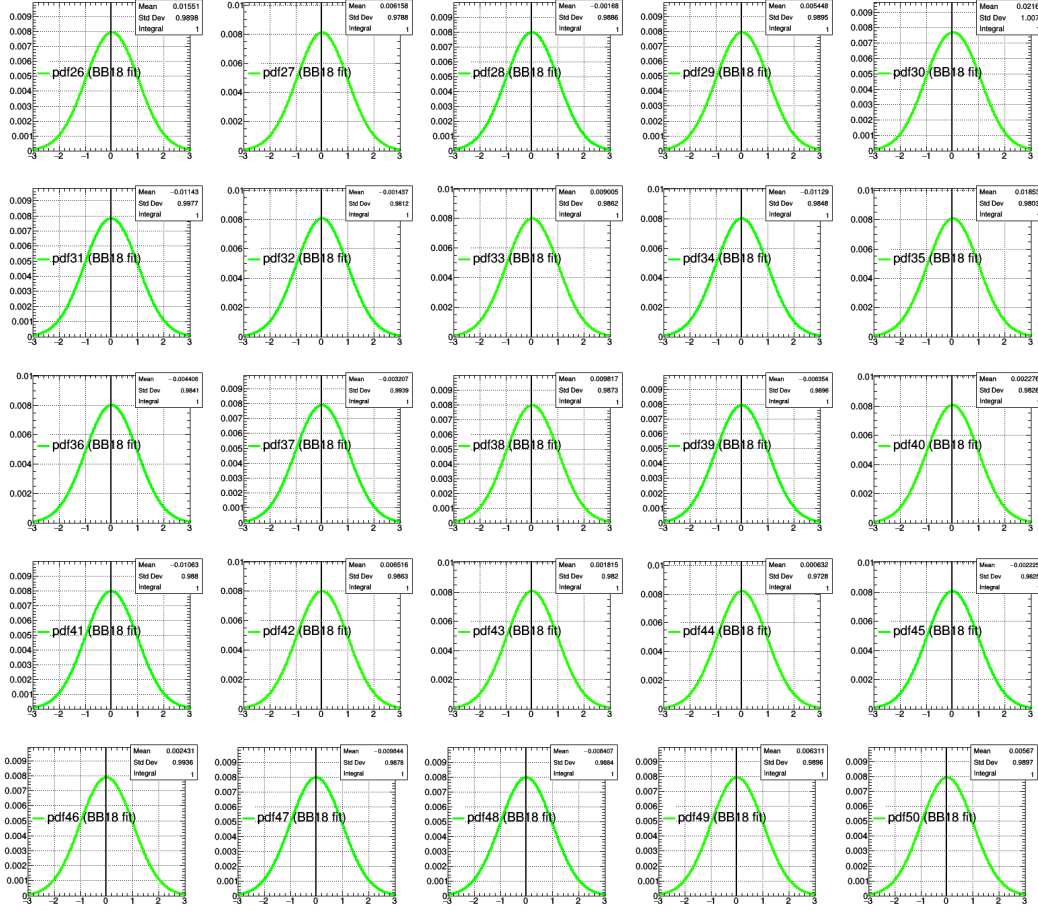


Figure E.5: The postfit probability distribution functions (pdfs) for the last 25 nuisances of the parton distribution functions (PDFs) in the six-SRs fit. The statistics panel displays the mean and the std of each pdf. This set of results corresponds to the pseudodata.

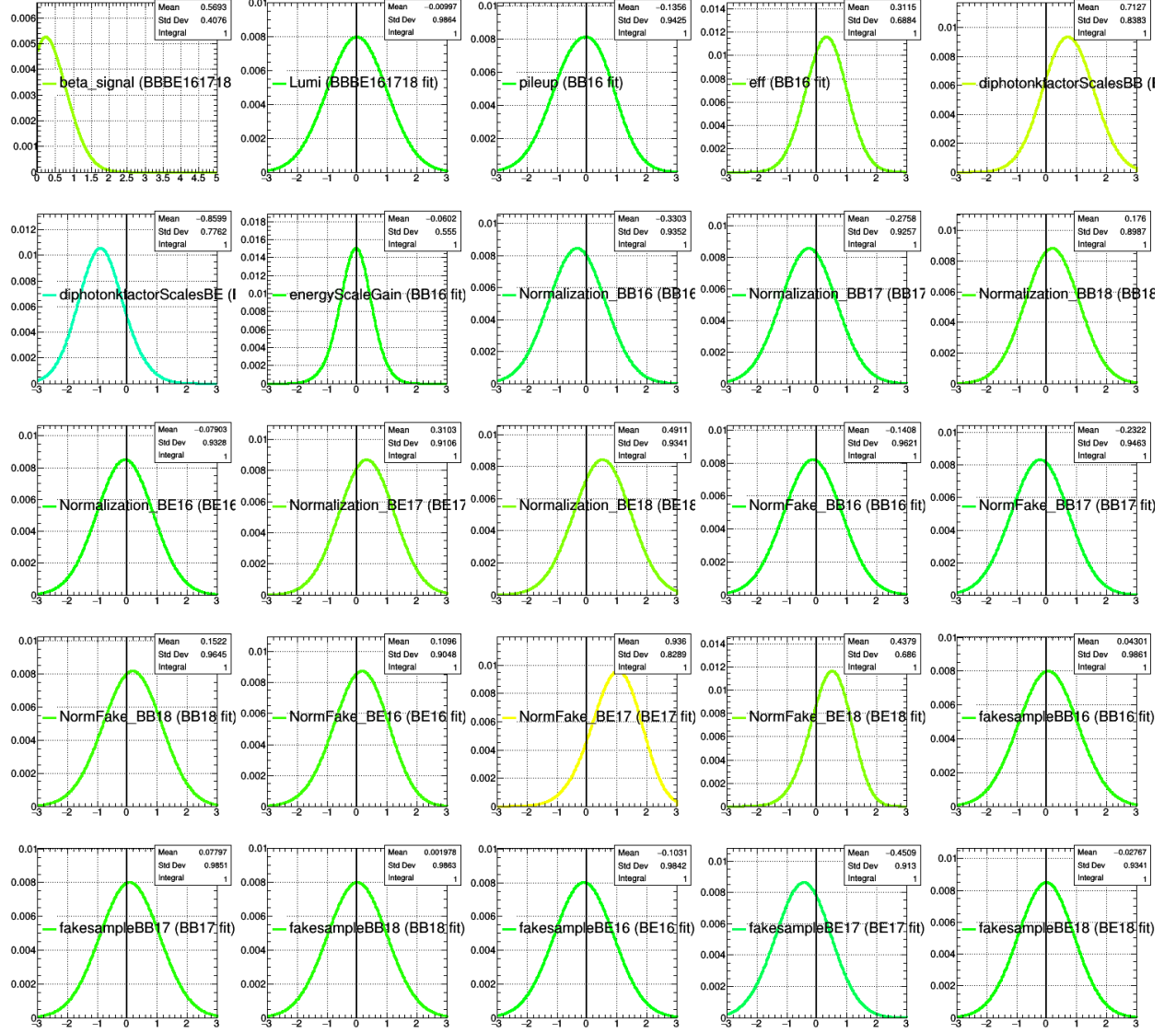


Figure E.6: The postfit pdfs for the 25 nuisance parameters used in the six SRs fit with CW model signal using pseudodata. (We omit the illustration of the other 50 of PDFs pdfs as they are gaussian-like without particular interest.)

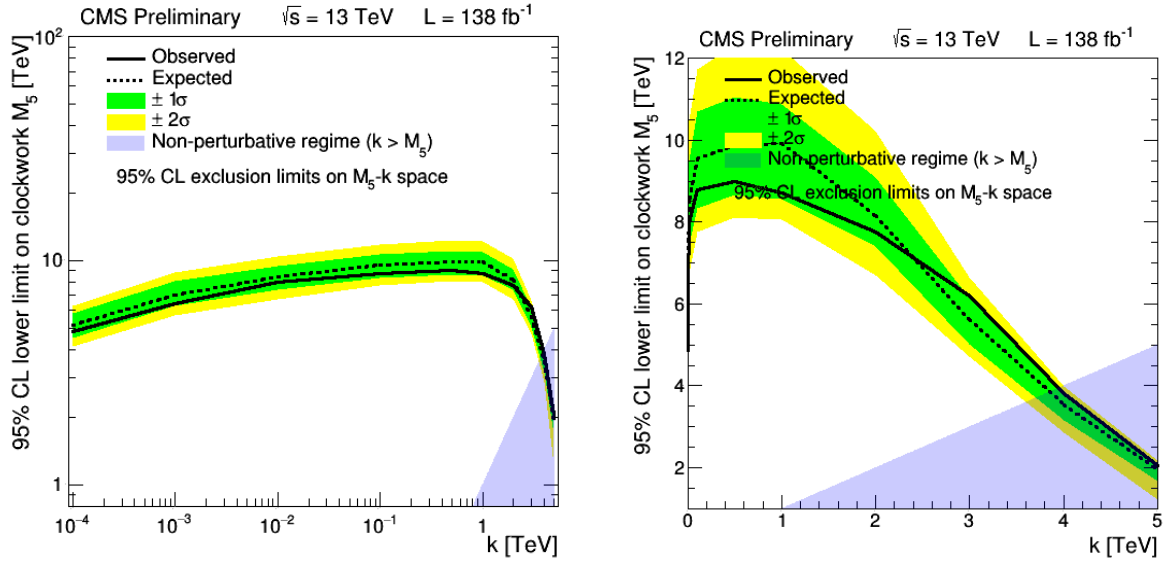


Figure E.7: The exclusion limit for the continuous graviton model in the clockwork framework over the $k - M_5$ parameter space using pseudodata. The shaded region denotes where the theory becomes non-perturbative. (The right plots is the same as the left one but in non-log axis). The result from the fit over all six SRs (collectively over all years and channels).