

Search for Quark Compositeness in γ +jet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS Detector at the Large Hadron Collider

THESIS SUBMITTED TO THE UNIVERSITY OF DELHI
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY



ROCKY BALA GARG

DEPARTMENT OF PHYSICS & ASTROPHYSICS
UNIVERSITY OF DELHI
DELHI 110 007
INDIA

APRIL, 2018



DECLARATION

It is certified that the work presented in this thesis entitled “**Search for Quark Compositeness in γ +jet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS Detector at the Large Hadron Collider**” has been carried out at the Department of Physics & Astrophysics, University of Delhi, Delhi, under the supervision of Prof. Brajesh Chandra Choudhary. This thesis deals with the search of excited states of light and heavy flavor quarks in the photon + jet final states at a center-of-mass energy of 13 TeV using 35.9 fb^{-1} of proton–proton collision data collected by the CMS Experiment at the Large Hadron Collider at European Organization for Nuclear Research (CERN) in Geneva, Switzerland.

This work has been done by the candidate herself and to the best of her knowledge, no part of this work has earlier been submitted for any degree or diploma of this or any other university.

Rocky Bala Garg

(Candidate)

Prof. Brajesh Chandra Choudhary

(Supervisor)

Prof. Sanjay Jain

Head, Department of Physics & Astrophysics

University of Delhi

Delhi 110 007



Department of Physics & Astrophysics

University of Delhi
Delhi-110007, India

Date: _____

Certificate of Originality

The research work embodied in this thesis entitled “Search for Quark Compositeness in γ +jet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS Detector at the Large Hadron Collider” has been carried out by me at the Department of Physics & Astrophysics, University of Delhi, Delhi, India. The manuscript has been subjected to plagiarism check by Turnitin software. The work submitted for consideration of award of Ph.D is original.

Rocky Bala Garg

Name and Signature of the Candidate

Student Approval Form

Name of the Author	Rocky Bala Garg
Department	Physics & Astrophysics
Degree	Doctor of Philosophy
University	University of Delhi
Guide	Prof. Brajesh Chandra Choudhary
Thesis Title	Search for Quark Compositeness in γ +jet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS Detector at the Large Hadron Collider
Year of Award	

Agreement

1. I hereby certify that, if appropriate, I have obtained and attached hereto a written permission/statement from the owner(s) of each third party copyrighted matter to be included in my thesis/dissertation, allowing distribution as specified below.
2. I hereby grant to the university and its agents the non-exclusive license to archive and make accessible, under the conditions specified below, my thesis/dissertation, in whole or in part in all forms of media, now or hereafter known. I retain all other ownership rights to the copyright of the thesis/dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis, dissertation, or project report.

Conditions:

1. Release the entire work for access worldwide	
2. Release the entire work for 'My University' only for 1 Year 2 Year 3 Year and after this time release the work for access worldwide	
3. Release the entire work for 'My University' only while at the same time releasing the	

following parts of the work (e.g. because other parts relate to publications) for worldwide access. a) Bibliographic details and Synopsis only. b) Bibliographic details, synopsis and the following chapters only. c) Preview/Table of Contents/24 page only.	
4. View Only (No Downloads) (worldwide)	

Signature of the Scholar

Signature and seal of the Guide

Place: _____

Date: _____

Dedicated To My Family

ACKNOWLEDGEMENTS

This thesis is the culmination of my journey of Ph.D which was just like climbing a high peak step by step accompanied with encouragement, hardship, trust, and frustration. When I found myself at the top experiencing the feeling of fulfillment, I realized though only my name appears on the cover of this thesis, a great many people have contributed to accomplish this huge task.

Foremost, I want to offer this endeavour to **GOD Almighty** for the wisdom he bestowed upon me, the strength, peace of mind and good health, in order to finish this research.

I would like to express my deepest gratitude towards my mentor and advisor Prof. Brajesh C. Choudhary for his constant guidance, overwhelming support and patience, though mixed with occasional shouting. His great insight into the subject has always been valuable whenever I found myself directionless. He has always understood my constraints and encouraged me to perform better. It was his faith and motivation that helped me in sustaining my efforts which many a times did oscillate. I am really thankful to him for providing me the opportunity to be a part of one of the world's biggest high energy physics experiments. He has also taught me the importance of being organized and having a professional attitude in work.

I also wholeheartedly thank Prof. Debajyoti Choudhury who is solely responsible for creating and enhancing my interest in the theoretical aspect of high energy physics. I consider myself blessed and privileged for his significant guidance during the initial years of my Ph.D life. He has always been available and guided me with best of his knowledge whenever I needed his help during any stage of my research.

I owe a sincere thanks to Dr. Varun Sharma with whom I have worked closely since the beginning of my Ph.D tenure. I am very grateful for his help and guidance at both the fundamental and technical levels. His suggestions have always been of invaluable importance during each stage of my research work. I also thank Dr. Sushil Singh Chauhan for his fruitful suggestions and discussions during the final stages of this study. My heartfelt thanks also goes to Dr. Daniele Del Re and Prof. Robert Brown for their thorough reviews and important suggestions which have played a vital role in improving this analysis. A special thanks to Dr. Dinko Ferencek and Dr. Javier M. Duarte for their availability and providing a helping hand on the technical front. A vote of thanks to all the members of CMS collaboration and the people involved in this analysis directly or indirectly. I am also thankful to LHC Physics Center, Fermilab, USA for providing computational resources for this analysis.

I acknowledge the members of my advisory committee, Dr. Nivedita Deo and Dr. Supriya K. Kar for their valuable advice and suggestions. I am truly obliged to Prof. Sanjay Jain, Head of the Department of Physics & Astrophysics, for providing adequate facilities and comfortable environment in the department. I also thank Mr. Pradeep Kundra for handling the administrative matters properly.

My thanks and appreciations also go to all the members of Experimental High Energy Physics group at the University of Delhi, including Prof. R. K. Shivpuri, Prof. Kirti Ranjan, Dr. Ashutosh Bhardwaj, Dr. Md. Naimuddin and Dr. Ashok Kumar. I also thank all the members of India-CMS collaboration for their encouragement and support during the various meetings. A sincere thank to the Council of Scientific and Industrial Research (CSIR) for providing the fellowship in order to carry out this research. I also would want to thank DAE-DST (Govt. of India) and UGC for the financial support for international travels.

I am highly indebted to my seniors and colleagues in the department, Dr. Sudha Ahuja, Dr. Varun Sharma, Dr. Arun Kumar, Dr. Ajay Kumar, Dr. Daljeet Kaur, Dr. Shivali Malhotra, Dr. Ranjeet Dalal, Ramkrishan Sharma, Arjun Chettri, Kanwaljeet Singh Channey, Sumit Kesri, Zubair Ahmed Dar, Ankit Gaur, Geetika Jain, Aashaq Hussain Shah, Mohit Gola, Priyanka Phogat, Probhjot Singh, Aman Phogat, Prajesh Sharma, Ashar, Rijwan, Chakaresh, for their goodwill and constant encouragement. My appreciation also extends to the members of laboratory staff, Md. Yunus, Aamir Hussain and Atul, for their innumerable help.

This acknowledgement will never be complete without the mention of my bosom friends, Reena Ghadwal and Rachna Vats who have always loved me for who I am. I can truly be myself when I am with them. Its Reena, who truly laid the first stone of my Ph.D journey. I will always be indebted to her for this.

I revere the patronage and moral support extended with love, by my family (specially, my in-laws family), whose encouragement, support and patience has made it possible for me to complete this journey. I owe this success to my mother-in-law who has always understood me and cared for me like her own daughter. I am truly grateful to my brother-in-law for supporting me and giving me the love and care of an elder brother. A big thank to my dear brother, Yogesh Singla who has always been my morale booster. My deep appreciations to Amit and Indu Goel for always being around. Much love to my nephew Naman Garg, for his innocent affections.

It is impossible to acknowledge, in words, the unconditional love and support that I got from my husband during this journey. This thesis has become a reality only because of his continuous efforts. He was always around at times when I

thought it is impossible to continue. He helped me to keep things in perspective. I greatly value his contribution and deeply appreciate his belief in me. My heart goes to my baby, my little girl Kriti for her selfless love and positivity that she brought in my life. She has truly been my lucky charm. I appreciate her for abiding my ignorance and for the patience she showed during this time.

Rocky Bala Garg

LIST OF PUBLICATIONS

Journal Publications:

1. CMS Collaboration, “Search for excited quarks of light and heavy flavor in γ + jet final states in proton–proton collisions at $\sqrt{s} = 13$ TeV”.
Accepted in Physics Letters B (Ref. No. PLB-D-17-01732R1).

Other Public Results:

1. CMS Collaboration, “Search for excited states of light and heavy flavor quarks in γ + jet final state in proton–proton collisions at $\sqrt{s} = 13$ TeV”.
CMS-PAS-EXO-17-002 (Public Analysis Summary).
2. CMS Collaboration, “Search for excited quarks in γ + jet final state in proton–proton collisions at $\sqrt{s} = 13$ TeV”.
CMS-PAS-EXO-16-015 (Public Analysis Summary).

CMS Analysis and Detector Notes:

1. “Search for excited states of light and heavy flavor quarks using CMS data collected at center of mass energy of 13 TeV in photon+jet final state”, Brajesh Choudhary, **Rocky Bala Garg**, and Varun Sharma.
CMS Analysis Note-16-462 (2016).
2. “Search for excited quark resonance using the photon+jet mass spectrum with 2.5 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV in the CMS experiment”, Brajesh Choudhary, **Rocky Bala Garg**, and Varun Sharma.
CMS Analysis Note-15-262 (2015).
3. “Search for excited b-quarks in b-tagged gamma+jet final state at 8 TeV”, Brajesh Choudhary, **Rocky Bala Garg**, and Varun Sharma.
CMS Analysis Note-15-039 (2015).
4. “Comparison of Single Particle Response in Data and MC at $\sqrt{s} = 8$ TeV”, **Rocky Bala Garg**, Ruchi Gupta, Suananda Banerjee.
CMS Detector Note-14-030 (2014).

Conference Proceedings/Poster Presentations:

1. “Quark substructure at CMS”, **Rocky Bala Garg**. Poster presented at the **XXII DAE-BRNS HIGH ENERGY PHYSICS SYMPOSIUM**, University of Delhi, Delhi, India, 12th – 16th December, 2016.
2. “Particle physics at the crossroads”, attended a symposium organized by the University of Delhi and University of Edinburgh, India International Center, New Delhi, India, 15th – 17th February, 2013.

Search for Quark Compositeness in γ +jet final states in proton-proton collisions at $\sqrt{s} = 13$ TeV with the CMS Detector at the Large Hadron Collider

ABSTRACT

The main aim of the particle physics is to investigate the fundamental building blocks of matter and the forces governing them. Within the context of the standard model of particle physics, quarks and leptons are considered to be fundamental. Their existence have been verified by a number of experiments and their properties are extensively studied. In spite of its great success, the standard model is not considered as a complete theory. There are a number of unexplained phenomena which predict the existence of some new physics beyond the standard model. The existence of three nearly identical generations of quarks and leptons strongly suggest the possibility of an underlying sub-structure. A variety of models, referred to as, the compositeness models, are theorized which consider the quarks and leptons to be made up of more fundamental particles, known as “preons”. A convincing signature of the quarks sub-structure is the observation of their excited states. These excited quarks are supposed to interact with their standard model counterparts through gauge interaction for the compositeness scale, $\Lambda \leq \sqrt{s}$. Various searches have been performed in the past to observe excited states but no evidence has been found.

This thesis presents a search for the excited states of light u, d and heavy b flavor quarks in the γ + jet and γ + b-jet final states respectively, using the pp collision data at $\sqrt{s} = 13$ TeV, collected by the CMS detector in 2016. This data correspond to a total integrated luminosity of 35.9 fb^{-1} . The excited quark signals, if exist, will show their presence in the form of bumps over the continuous invariant mass distribution spectrums of γ + jet and γ + b-jet states respectively. Many standard model processes can also imitate these final states, thereby, forming the backgrounds of this study. This analysis has two main backgrounds coming from SM γ + jet and QCD dijet processes, with a small contribution from W/Z + jet processes. The MC samples corresponding to these backgrounds are used to validate the data as per the standard model expectations. The actual background of this study has been obtained by using a data driven technique. The excited quarks signals are generated and simulated using the event generator PYTHIA to scan a broad mass range.

In the absence of any expected signal, upper limits on the cross-sections or lower limits on the masses of excited quarks have been set at the 95% confidence level using the frequentist approach in the asymptotic approximation. A comparison of

cross-section upper limits has been made with the theoretical predictions of excited quarks in order to set the lower mass limits on the excited states. The excited light quarks within the mass range $1.0 < M_{q^*} < 5.5 \text{ TeV}$ and excited b-quarks within the mass range $1.0 < M_{b^*} < 1.8 \text{ TeV}$ are excluded at the 95% confidence level, corresponding to standard model like coupling strengths. The limits presented in this thesis are the most sensitive limits for excited quark searches in the $\gamma + \text{jet}$ and $\gamma + \text{b-jet}$ final states. The search for the excited b-quarks has been presented for the first time in any final state at $\sqrt{s} = 13 \text{ TeV}$.

The work presented in this thesis has been accepted for publication in Physics Letters B (Ref. No. PLB-D-17-01732R1).

CONTENTS

1	Theoretical Perspective	1
1.1	Evolution Of Particle Physics	1
1.2	Standard Model	8
1.2.1	Matter particles	9
1.2.2	Forces and messenger particles	10
1.2.3	SM as a gauge theory	12
1.2.4	Challenges faced by SM: unknown mysteries	22
1.3	Beyond The Standard Model	23
1.3.1	Model of excited quarks	23
1.3.2	A follow-up of previous searches	26
2	The Experimental Apparatus	27
2.1	The Large Hadron Collider	27
2.1.1	LHC tunnel layout	29
2.1.2	Main components of the LHC	30
2.1.3	Proton's journey to maximum acceleration	32
2.1.4	Luminosity measurements at LHC	33
2.2	The CMS Detector	35
2.2.1	Layout of the CMS detector	36
2.2.2	CMS Trigger and Data Acquisition System	52
2.2.3	CMS data management	56
2.2.4	CMS software and computing	57
3	Event Reconstruction And Simulation	59
3.1	Event Reconstruction	60
3.1.1	Track reconstruction	62
3.1.2	Vertex reconstruction	64

3.1.3	Photon reconstruction and identification	65
3.1.4	Jet reconstruction and identification	72
3.2	Event Simulation	81
3.2.1	Event generation	83
3.2.2	Detector simulation	86
4	Search for light and heavy flavor excited quarks	87
4.1	Analysis strategy	87
4.1.1	Excited quark analysis	88
4.2	Production of Ntuples	91
4.2.1	Samples processed	91
4.3	Re-weighting of MC events	97
4.3.1	Luminosity re-weighting	97
4.3.2	Pile-up re-weighting	97
4.4	Event selection	98
4.4.1	Trigger selection	99
4.4.2	MET filters	100
4.4.3	Primary vertex selection	101
4.4.4	Analysis object selection: Photon	101
4.4.5	Analysis object selection: Jet	105
4.4.6	Analysis object selection: b-Jet	106
4.4.7	Kinematical restrictions between γ and jet/b-jet	108
4.4.8	Optimization of event selection	117
4.5	Highest invariant mass events in data	118
4.6	Signal spectrum	121
4.7	Signal interpolation	122
4.8	Parametric background	124
4.8.1	Fisher test	125
4.8.2	Goodness-of-fit	127
4.8.3	Bias due to the choice of background fit function	129
4.9	Systematic Uncertainties	132
4.9.1	Luminosity	132
4.9.2	Pile-up Uncertainty	132
4.9.3	HLT trigger inefficiency	133
4.9.4	Photon and jet energy scale	133
4.9.5	Photon and jet energy resolution	133
4.9.6	Photon ID inefficiency	134

4.9.7	b-Tag scale factor	134
4.9.8	Signal shape interpolation	135
4.9.9	Background Shape Uncertainty	135
4.10	Limit computation	135
4.11	Results	138
4.11.1	Significance of local excesses in data	138
4.11.2	Exclusion limits	139
5	Summary and conclusion	145

LIST OF FIGURES

1.1	A timeline of particle discoveries.	6
1.2	$W^{\mu 3}$ - B^{μ} plane	17
1.3	A potential with real scalar field $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ before and after spontaneous symmetry breaking.	19
1.4	Feynman diagrams for $q^* \rightarrow \gamma + \text{jet}$ final state with (a) quark-gluon fusion (b) quark-antiquark annihilation (c) gluon-gluon fusion processes. Whereas the light quarks will show their presence in all the three modes, heavy bottom quark will appear only in quark-gluon fusion and quark-antiquark annihilation.	25
2.1	Big-Bang timeline.	28
2.2	A schematic layout of the LHC tunnel.	30
2.3	A Radio Frequency cavity.	30
2.4	A sketch showing formation of a bunch.	30
2.5	The LHC twin-bore magnet system	32
2.6	A schematic view of CERN accelerator complex representing the different accelerators deployed for providing acceleration to the proton beam.	33
2.7	Total integrated luminosity delivered by the LHC and recorded by CMS as a function of time, during the year 2016 [108].	35
2.8	A layout of the CMS experiment at CERN.	36
2.9	Longitudinal contour of a quadrant of the CMS detector.	37
2.10	A cross-sectional view of the CMS tracker system in r - z plane, with each line representing a silicon detector module.	39
2.11	A schematic layout of the CMS pixel detector.	40
2.12	A schematic representation of the upgrades in pixel detector barrel (left) and endcaps (right) during Phase1 Pixel Upgrade.	41

2.13	Performance measurement of the older pixel detector (blue) and the upgrade pixel detector (red) using $t\bar{t}$ events, showing tracking efficiency (left) and fake rate (right) as a function of per event pileup [115].	42
2.14	A layout of the CMS Electromagnetic calorimeter displaying the arrangement of crystals in barrel and endcaps with the preshower present in front.	44
2.15	A schematic view of a quadrant of the CMS Hadron calorimeter presenting different sub-detectors (in yellow color).	47
2.16	A schematic view of the solenoidal magnet used in the CMS detector, showing the coil and the central return yoke.	49
2.17	A schematic view of the solenoidal magnet used in CMS, showing the coil and the central return yoke. The trajectory of a muon through the muon stations is also presented.	50
2.18	CMS Data Acquisition and trigger system.	53
2.19	An architecture of CMS Level-1 trigger system.	54
3.1	A diagrammatic representation of an event processing chain.	59
3.2	A graphical representation of the trajectories of various particles through the CMS detector.	61
3.3	Different steps of CMS reconstruction program.	61
3.4	Various steps followed in reconstruction of particles inside ECAL.	66
3.5	An illustration of the Hybrid algorithm used for clustering in ECAL barrel region.	67
3.6	An illustration showing overlapping of two Multi5×5 clusters. Clusters in yellow are eligible to seed other Multi5×5 clusters if they form the local energy maxima. The mechanism of formation of a supercluster using many multi5×5 clusters is also shown.	69
3.7	A pictorial representation of a jet.	73
3.8	The behaviour of different jet clustering algorithms.	75
3.9	Different jet energy correction levels applied to data and MC.	77
3.10	A depiction of a b-jet emerging from the primary vertex.	79
3.11	An illustration of b-tagging algorithms.	79
3.12	Event simulation and data analysis chain.	82
3.13	Sequential steps involved in the formation of an event.	84
4.1	Feynman diagrams for quark-gluon fusion process in s and t channel.	89
4.2	Standard model background contribution from γ + jet processes.	89

4.3	Standard model background contribution from QCD dijet processes.	90
4.4	Bremsstrahlung contribution to SM QCD dijet production.	90
4.5	Background contributions from electroweak processes.	91
4.6	Signal shapes for q^* (left) and b^* (right) for different masses and couplings.	95
4.7	Data-MC comparison of the number of vertices in an event, showing the effect of pile-up re-weighting.	98
4.8	Relative and absolute trigger efficiencies for the single photon trigger, HLT_Photon165_HE10.	100
4.9	Data-MC comparison of photon identification and isolation variables after final selection.	104
4.10	ΔR distribution between prompt gen photon matched to selected reco photon and other gen particles: (a) QCD before overlap removal, (b) γ + jet before overlap removal, (c) Stitching of QCD and γ + jet after overlap removal.	105
4.11	p_T^γ vs. p_T^{jet} in 2-dimensional representation for q^* signals.	106
4.12	Data-MC comparison of jet particle flow identification variables after final selection.	107
4.13	Data-MC comparison of CSVv2 discriminator distribution before and after applying the b-tagging.	108
4.14	Turn-on due to various kinematical selections, plotted for different jet p_T cuts.	109
4.15	The product of acceptance and efficiency as a function of q^* and b^* mass for SM coupling, $f = 1.0$, obtained using MC simulation. . . .	111
4.16	Data-MC comparison of p_T distribution of photons and jets for three categories: no b-tag, 1 b-tag and 0 b-tag.	112
4.17	Data-MC comparison of η distribution of photons and jets for three categories: no b-tag, 1 b-tag and 0 b-tag.	113
4.18	Data-MC comparison of ϕ distribution of photons and jets for three categories: no b-tag, 1 b-tag and 0 b-tag.	114
4.19	Distributions of photons and jets in $\Delta\eta$ (above left), $\Delta\phi$ (above right), and ΔR (below).	115
4.20	Number of photons (above left), jets (above right), and b-jets (below) per event in Data-MC.	115
4.21	Data-MC comparison of invariant mass distribution of γ + jet in three categories: no b-tag, 1 b-tag and 0 b-tag.	116

4.22	Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different photon ID cuts corresponding to SM coupling $f = 1.0$	117
4.23	Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different $\Delta\eta$ cuts corresponding to SM coupling $f = 1.0$	118
4.24	Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different $\Delta\phi$ cuts corresponding to SM coupling $f = 1.0$	119
4.25	Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different CSVv2 cuts corresponding to SM coupling $f = 1.0$	119
4.26	Event display for the highest invariant mass event at $M_{\gamma+\text{jet}} = 4616.69 \text{ GeV}$. This event has the jet tagged as a b-jet, thereby, forming a highest q^* as well as highest b^* event.	120
4.27	Event display for the 2 nd highest invariant mass event at $M_{\gamma+\text{jet}} = 4595.43 \text{ GeV}$	120
4.28	Fitting the core of invariant mass of q^* signal at different masses for $f = 1.0$, using Breit-Wigner (Blue), Crystal-Ball (Red), and Gaussian (Green) functions to obtain the resonance width of the signal.	122
4.29	Polynomial fitting of mass resolution vs. resonance mass to obtain variable binning parameters.	122
4.30	Alternate generated (solid line) and interpolated (dotted line) shapes for different q^* mass points.	123
4.31	Comparison of generated and interpolated shapes at different q^* mass points.	124
4.32	The F-test distributions for fitting of function $F1$ with q^* invariant mass, for 3 and 4 parameters, showing that $F1$ with 4 parameters provide a better fit compared to 3 parameters.	126
4.33	The F-test distributions for fitting of function $F1$ with b^* invariant mass, for 3 and 4 parameters, showing that $F1$ with 4 parameters provide a better fit compared to 3 parameters.	126
4.34	Toy distributions for the goodness-of-fit study for q^*	128
4.35	Toy distributions for the goodness-of-fit study for b^*	128
4.36	Invariant mass distribution spectrum of $\gamma + \text{jet}$ fitted with 4-parameter polynomial fit function for three categories: (a) no b-tag (b) 1 b-tag and (c) 0 b-tag.	130
4.37	Pulls distribution as a function of q^* (left) and b^* (right) mass corresponding to the four test functions.	131

4.38	Effect of jet and photon energy correction uncertainties on the invariant mass of q^* signal.	133
4.39	Effect of jet and photon energy resolution on the invariant mass of q^* signal.	134
4.40	Effect of b-tag scale factor errors on the invariant mass of b^* signal.	135
4.41	The observed significance in data for q^* signal corresponding to the coupling multipliers (a) $f = 1.0$, (b) $f = 0.5$, and (c) $f = 0.1$ as a function of the resonance mass for $\gamma + \text{jet}$ resonance.	139
4.42	The observed significance in data for b^* signal corresponding to the coupling multipliers (a) $f = 1.0$, (b) $f = 0.5$, and (c) $f = 0.1$ as a function of the resonance mass for $\gamma + b\text{-jet}$ resonance.	139
4.43	The observed and expected 95% CL upper limits on $\sigma \times B \times A \times \epsilon$ (left column) and $\sigma \times B$ (right column) as a function of the mass of q^* resonance, for the coupling multipliers $f = 1.0$ (upper row), $f = 0.5$ (middle row), and $f = 0.1$ (lower row). The limits are also compared with the theoretical predictions for excited quark production. The inner (green) and outer (yellow) bands indicate the regions containing the 68% and 95% of the expected limits. . . .	141
4.44	The observed and expected 95% CL upper limits on $\sigma \times B \times A \times \epsilon$ (left column) and $\sigma \times B$ (right column) as a function of the mass of b^* resonance, for the coupling multipliers $f = 1.0$ (upper row), $f = 0.5$ (middle row), and $f = 0.1$ (lower row). The limits are also compared with the theoretical predictions for excited b-quark production. The inner (green) and outer (yellow) bands indicate the regions containing the 68% and 95% of the expected limits. . . .	142
4.45	The observed and expected 95% CL upper limits on $\sigma \times B$ as a function of the mass of the excited quark, for $f = 1.0$. The limits are also compared with theoretical predictions for excited quark production for the three couplings. The inner (green) band and the outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the expected limits.	143
4.46	The observed and expected 95% CL upper limits on $\sigma \times B$ as a function of the mass of the excited b-quark, for $f = 1.0$. The limits are also compared with theoretical predictions for excited b-quark production for the three couplings. The inner (green) band and the outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the expected limits.	143

4.47	The observed and expected regions excluded at 95% CL for q^* and b^* production and decay, as a function of M_{q^*} , M_{b^*} and f	144
------	---	-----

LIST OF TABLES

1.1	Standard model fundamental fermion's three generations.	9
1.2	Important properties of fundamental forces of universe.	10
1.3	Assignment of the various quantum numbers including T_3 and Y to the electroweak singlet and doublet fundamental fermions.	15
1.4	Approximate estimate of branching fractions (BR) of q^* decay in different channels, assumptions made are: $\Lambda = M_{q^*}$ and $f = f' = f_s$	25
2.1	Design values of the various parameters of LHC and a comparison with the values obtained during the 2016 run.	34
4.1	List of data samples along with corresponding run ranges and inte- grated luminosity. The JSON file used to select certified runs and luminosity sections is also listed.	92
4.2	MC samples for q^* signal, simulated and reconstructed at $\sqrt{s} = 13$ TeV.	93
4.3	MC samples for b^* signal, simulated and reconstructed at $\sqrt{s} = 13$ TeV.	94
4.4	The background MC samples used in this analysis, simulated and reconstructed at $\sqrt{s} = 13$ TeV.	96
4.5	Different triggers used in the trigger efficiency study.	100
4.6	Effective area for different photon isolations.	102
4.7	Selection cut flow table for signal, different backgrounds and data.	110
4.8	Selection efficiency (%) table for the q^* signal samples in the mass range 1.0–7.0 TeV with SM Coupling $f = 1.0$	110
4.9	Selection efficiency (%) table for the b^* signal samples in the mass range 1.0–4.0 TeV with SM Coupling $f = 1.0$	110
4.10	Photon ID cut optimization corresponding to different mass expected limits.	117
4.11	$\Delta\eta$ cut optimization corresponding to different mass expected limits.	118
4.12	$\Delta\phi$ cut optimization corresponding to different mass expected limits.	119

4.13 CSVv2 cut optimization corresponding to different mass expected limits.	119
4.14 Summary of the dominant sources of systematic uncertainties and their effect on the signal yield.	136
4.15 Constraint terms associated with different systematic uncertainty nuisance parameters.	137

THEORETICAL PERSPECTIVE

We, humans have always been curious about the physical world around us. Our, possible, sole existence in this vast universe has fascinated us since time immemorial. Humanity's interest in the self-actualization has been universal and enduring. Since ages, we are driven to explore the unknown to find out the answers to the questions about our origin. These fascinations and curiosity has given rise to the birth of the field of particle physics. Particle physics is a branch of physics that tries to answer the fundamental questions like what the world is made of and how it came into existence? It allows us to explore the secrets of the universe at the levels unmatched by any other field. It is perhaps the most esoteric and mystical-sounding scientific field. It is also termed as "High energy physics" because it requires microscopic particles to interact at tremendously high energies to observe new and interesting phenomenon. The theoretical model that defines the properties of fundamental particles and their interactions is known as "Standard Model" (SM) [1–5]. The subject of particle physics primarily revolves around the SM and its various possible extensions, mainly investigating the postulates and assumptions posed by it.

1.1 Evolution Of Particle Physics

The first evidences of an understanding of the world in terms of fundamental elements date back to ancient Indian philosophy that categorizes the world in five elements (called Pancha Mahabhootas): earth, fire, air, water and space. An

Indian philosopher, named Kanada who lived around 6th century B.C., is credited to be the first to develop the atomic theory of matter. He founded a school of philosophy, named Vaisheshika. Kanada suggested that everything around us can be subdivided, but the subdivision can not go on forever. There exists smallest entities, named parmanu, that are eternal and can not be further divided. Also, the Greek philosophers, Leucippus and his disciple Democritus around 5th century B.C., speculated that all the matter occupying the universe is essentially made up of myriads of minuscule particles, called atoms, coming from the Greek word “atomos”, meaning “indivisible”. Faiths of similar forms are also found to exist in other civilizations. However, these early theories of atoms were greatly conceived through abstract, philosophical reasoning without any profound experimental basis.

The concept of atom was then, revisited and embellished by many scientists and philosophers, like Galileo, Newton, Boyle and Lavoisier during 16th and 17th centuries. In 1802, John Dalton, through his experiments on atmospheric gases, modernized the ancient Greek ideas of atoms and led the foundation of the modern atomic theory of matter [6, 7]. The idea of Dalton was further taken forward by Mendeleev, who developed a periodic table in 1869, by classifying the atoms with similar properties into groups and predicting various new elements. This classification led to the idea that the atoms can be made up of simpler building blocks.

The modern era of particle physics began with the discovery of first sub-atomic particle in the cathode-ray tube experiment by J. J. Thomson in 1897 [8], when he observed that the cathode rays bent in the presence of magnetic field. He demonstrated that these rays are made up of fundamental subatomic particles called “corpuscles”, later known as “electrons”. Overwhelmed by this discovery, Thomson proposed a model of atom in 1904, known as the “plum-pudding model” [9], in which he hypothesized that the atom consists of both positive and negative charges with positive charges spread uniformly in the form of a spherical cloud and negative charges embedded in it in the form of electrons. This model was, however, overturned by Rutherford [10, 11], in 1911, when he performed an experiment of scattering of alpha particles from a thin gold foil. He observed that a very small fraction of alpha particles scattered through large angles, an observation impossible to be explained by Thomson’s model. Rutherford, therefore, proposed that the atom consists of a small positively charged nucleus at the center with

electrons revolving around. Subsequent discovery of proton by Rutherford in 1919 (earlier proposed by Goldstein in 1886) and neutron by J. Chadwick in 1932 [12] as fundamental ingredients of nucleus, completed the picture of atom.

The inter-atomic phenomenon could not be explained by the laws of classical physics. Therefore, to understand the characteristics of subatomic world, concepts of “Quantum Mechanics” (QM) were developed by Planck, Einstein, Bohr, Schrodinger, Heisenberg along with others. Max Planck, in 1900, in an attempt to explain the nature of black body radiations, proposed the concept of energy quantization, assuming it a peculiar nature of emission mechanism. Einstein, in 1905 through Photoelectric effect [13], argued that quantization is an essential feature of electromagnetic (EM) radiation. Though his conception was widely disagreed until its experimental verification in 1923 by Arthur Compton in the famous Compton effect [14]. The quanta of electromagnetic radiation came to be known as “photon”, a term coined by Gilbert N. Lewis in 1926. The quantization of energy is a consequence of wave-particle duality, put forward by Louis De Broglie in 1924, that laid the foundation of QM. It says that each particle having mass m and velocity v has an associated wave of wavelength, $\lambda = \frac{h}{mv}$, h being the Planck’s constant. Einstein’s proposal of special theory of relativity (1905) [15] explained the dynamics of particles moving at very high speed. The reconciliation of special relativity with quantum mechanics resulted into the birth of “Quantum Field Theory” (QFT) which became a perfect theoretical tool to explain the behaviour of microscopic particles moving at relativistic speeds.

The first implication of QFT was the relativistic theory of electrons, proposed by Paul Dirac in late 1920s, commonly known as “Quantum Electrodynamics” (QED) [16]. Dirac’s theory successfully explained the behaviour of relativistic electrons but these electrons were found to occupy both positive and negative energy states. Dirac proposed that these negative energy states correspond to anti-particles of electrons with same mass and opposite charge. This most anticipated particle of Dirac, the positron (anti-electron), was discovered two years later, in 1932, by Carl Anderson [17] and became a spectacular triumph of Dirac’s theory. A more compelling interpretation of Dirac’s theory was presented by Feynman [18–20], Schwinger [21, 22] and Tomonaga [23], in 1948, by acknowledging the negative energy states of Dirac’s theory as the energy states of a different particle, thereby placing particle and anti-particle on same ground. They demonstrated that the

electromagnetic processes at the basic level occur by the exchange of photons. Feynman also introduced the pictorial representations named “Feynman diagrams”, to visualize the various processes.

The compactness of nucleus ($\sim$ fm or 10^{-15} m) puzzled physicists during 1930s since they could not understand: “How is it possible for so many positively charged protons to remain in such a close proximity inside the nucleus, in spite of the electrostatic repulsions?” Evidently, nucleons (protons and neutrons) inside the nucleus are bound together by a force more stronger than electrostatic repulsive force, thereby, given the name “strong force”. Since such a force was nowhere observed, it was considered that this force is of very short range, probably that of the size of a nucleus. The first serious theoretical background for strong force was staged by a Japanese physicist Hideki Yukawa in 1934 [24] when he suggested that the nucleons exhibit strong force through the mechanism of exchange of mediator particles, called “mesons”. Owing to the short range of strong force, Yukawa predicted that these mesons would be massive particles having mass $\sim 300m_e$, m_e being the electron’s mass. Soon after Yukawa’s prediction, in 1937, Anderson and Neddermeyer while performing cosmic ray experiments, discovered a particle with a mass of around $200m_e$ and acknowledged it as the Yukawa’s meson [25]. Further studies showed that this particle does not participate in strong interaction and behaves as a heavier version of electron, later named as “muon” (μ). The true Yukawa’s mesons were discovered, in 1947, in cosmic ray studies by Cecil Powell et. al. [26] and given the name “pions” (π ’s).

Around the same time in 1930s, there was one more puzzle making rounds in high energy physics community. It was the continuous electron energy spectrum obtained in nuclear β decay (first observed in 1914). The β decay process was noted to be a two-body decay in which a nucleus transformed into a lighter one while emitting an electron. Using the laws of conservation of energy and momentum, the energy of decay products could be computed and the electron was expected to acquire a discrete energy value, in contrast to the observation. This apparent perversities of β decay even threatened scientists of that time with the abandonment of law of energy conservation. A solution was sought by Wolfgang Pauli in 1930 who proposed that the β decay process might be a three-body decay and the outgoing electron is accompanied by an “invisible” elementary particle having vanishingly small mass and no charge. This particle was later named as “neutrino” by Enrico

Fermi in 1932. The experimental observation of neutrino was achieved quite late, in 1956, by Cowan and Reines in Savannah river nuclear reactor [27]. They looked for the inverse beta decay process: $\bar{\nu} + p \rightarrow n + e^+$. Detection of outgoing neutrons and positrons was a signature for existence of anti-neutrinos (then considered as neutrinos). Attempts were made to understand the nature of β decay process as it could not be explained by either the strong or EM interactions. Fermi came up with a theory of β decay in 1934 [28], in which he introduced a new interaction, now known as weak interaction, responsible for this decay. The theory posits interaction of four fermions directly at one point, known as the contact interaction. He also explained that the underlying process occurring in all β^- decays is $n \rightarrow p + e + \nu$.

The elementary particle's world seemed quite simple until late 1940s. The situation changed soon with the discovery of a lot new unconventional particles in the cosmic ray studies. During the end of 1947, Rochester and Butler found evidence for a new particle named kaon (K^+) [29], soon followed by the discovery of more particles which includes Λ^0 [30], Σ^+ [31, 32], Ξ^- [33] etc. by other groups. A picture depicting the timeline of various particle discoveries presented in Fig. 1.1 clearly shows the enormosity of particles discovered in fifties and sixties. These particles were termed as “strange particles” due to their strange features like long lifetime, production in pairs etc. In order to account for the quirkiness of these particles, Gell-Mann and Nishijima in 1953 proposed a new quantum number called “strangeness” for these particles. This quantum number was conserved in strong interactions but violated in weak interactions. Advent of particle accelerators during this time resulted into the appearance of many more strange and non-strange particles in the laboratory. These particles were classified based on their properties into different groups: leptons (e^- , ν_e etc.), mesons (π 's, K 's etc.) and baryons (p , n , Λ etc.). Gell-Mann and Ne'eman introduced the “Eightfold way” in 1961, that arranged the hadrons (mesons and baryons) having same spin-parity (J^P) but different charge (Q) and strangeness (S) into octets and decuplets. This classification was regarded as “periodic table” of particle physics and it led to the prediction of a new particle Ω^- having strangeness -3. This particle was discovered later on in 1964 at Brookhaven National Laboratory (BNL) [34] and thereby, became an experimental proof of the authenticity of eightfold patterns.

The astonishing patterns observed in eightfold classification were signalling

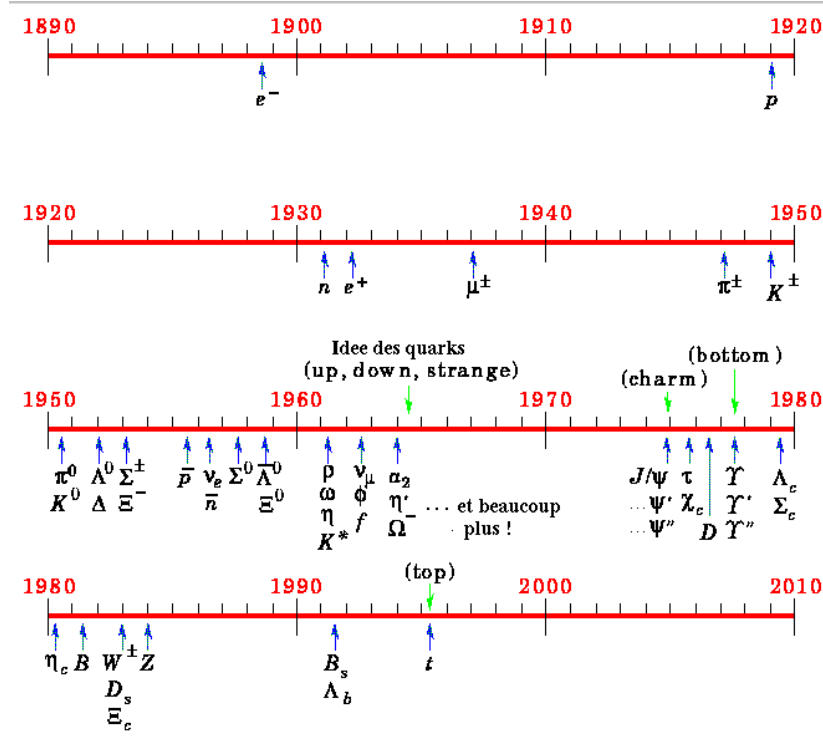


Figure 1.1: A timeline of particle discoveries.

towards something striking. The explanation arrived in the form of “Quark Model”, proposed independently by Gell-Mann [35] and Zweig [36] in 1964. They postulated that hadrons are not elementary but composed of fundamental entities, called “quarks”. These quarks are present in three flavours: d (down), u (up) and s (strange). Each baryon is made up of three quarks while each meson is made up of a quark and an anti-quark. Every member of the eightfold supermultiplets got a quark substructure, however, the states like Δ^{++} (uuu), Δ^{-} (ddd) and Ω^{-} (sss) were appeared to violate the Pauli’s exclusion principle. A way out was proposed by O. W. Greenberg in the same year by suggesting that each flavour quark further consists of three different color states (red, green and blue), known as color quantum number. Quarks of different colors combined together in forming the hadrons and all the particles observed in nature are colorless. This color hypothesis, though did not have any experimental proof, was able to describe the various aspects. Indirect searches which includes scattering of electrons with nuclear targets known as “deep inelastic scattering” [37] (similar to Rutherford’s scattering of atom) were performed in 1967 at SLAC and the first ever evidence of existence of proton substructure was observed. The Yukawa’s meson theory was not applicable for the strong force between quarks, so a new theory of strong

force, known as “Quantum Chromodynamics” (QCD) was developed in 1973 by H. Fritzsch and H. Leutwyler, along with Gell-Mann [38]. They proposed a gauge theory with SU(3) gauge group that predicted the existence of strong force carriers among quarks, termed as “gluons”. The first evidence of gluons was obtained at PETRA in 1979 in three-jet events [39].

The elementary matter particles known at this stage were: four leptons (e^- , ν_e , μ^- , ν_μ) and three quarks (u , d , s). The existence of a fourth quark, named as “charm” (c) quark, was predicted by Bjorken and Glashow around 1964 [40], in order to maintain a symmetry between number of quarks and leptons. In 1970, more technical aspects were provided for the existence of the charm quark by Glashow, Iliopoulos, and Maiani through GIM mechanism [41]. The first evidence of charm quark were obtained in 1974 in the form of $c\bar{c}$ (J/ψ) bound state simultaneously by C. C. Ting’s group at BNL [42] and B. Richter’s group at SLAC [43]. The discovery of J/ψ is known as *November Revolution* in the history. The Glashow’s symmetry that was restored by the discovery of charm quark was soon broken in 1975 with the observation of a new lepton, known as “tau” (τ) lepton, by M. Perl et. al. at SLAC [44]. This naturally led scientists to expect for a new family of quarks. Two years later, in 1977, a new heavy meson Upsilon (Υ) was discovered by L. Lederman et. al. at Fermilab [45] and was immediately recognized as the bound state of fifth quark, named as bottom/beauty (b) quark. After the bottom quark discovery, the search of its partner, named as truth/top (t) quark was very much anticipated. However, it took a really long time of 18 years to be finally discovered in 1995 at Tevatron’s CDF [46] and D0 [47] detectors at fermilab.

The discovery of the strange particle K^+ was a startling one. On its first observation in 1947, it appeared to decay in two different final states, initially thought as two different particles: τ^+ (decaying into three π ’s) and θ^+ (decaying into two π ’s). The difference in masses and lifetimes of τ and θ were observed to be within experimental errors, but they had different parities [48]. Since parity was considered a universal symmetry then, it led to the famous τ - θ puzzle which was solved by the observation of parity violation in weak interactions by C. S. Wu and collaborators in 1957 [49]. The experimental verification of parity violation led to the reformulation of Fermi’s theory of β decay in terms of Gamov-Teller transitions. Also, since Fermi’s theory was only effective at low energies and found to violate unitarity at high energies, an attempt was made to devise the

theory in terms of intermediate vector particles acting as mediators. The first step was a proposal from C. N. Yang and R. Mills who in 1954, developed the theory in terms of massless vector particles [50]. But weak interactions required its mediators to be massive, any attempt made to incorporate masses led to the breaking of local gauge symmetry, thereby making the theory inconsistent. A complete description of weak interactions was provided by S. Glashow [1], A. Salam [2] and S. Weinberg [3] in terms of electroweak unification, by considering $SU(2) \times U(1)$ gauge group, having three weak gauge bosons denoted by $W^\pm$ and Z . They invoked the Higgs mechanism proposed by Higgs [51–53], Englert & Brout [54], Guralnik, Kibble, & Hagen [55] in 1964, that provided masses to gauge bosons by spontaneous symmetry breaking of local gauge symmetries. The Higgs mechanism also predicted the existence of a new particle named Higgs boson, the mediator of Higgs field. The experimental verification of electroweak unification first arrived with the discovery of neutral currents by Gargamelle collaboration in 1973 [56] while performing neutrino scattering experiments, followed by the discovery of weak bosons, $W^\pm$ [57, 58] and Z [59–61] in $p\bar{p}$ collisions at SPS, CERN by UA1 and UA2 collaborations in 1983. The much awaited Higgs boson was finally discovered in 2012 by ATLAS [62] and CMS [63] collaborations at CERN, thereby, verifying the source of mass generation of universal particles. Both the collaborations have extensively studied the properties of this newly found boson [64, 65] and measured the mass to be $m_H \approx 125 \text{ GeV}/c^2$ within experimental errors.

The SM of particle physics that incorporates six leptons ($e, \nu_e, \mu, \nu_\mu, \tau, \nu_\tau$), six quarks (u, d, s, c, b, t) and three fundamental interactions (strong, EM, and weak) alongwith their carriers, was developed in 1960s & 1970s. At present, all the particles hypothesized by SM have been found and are well within the predictions.

1.2 Standard Model

The “Standard Model” is the theoretical model of particle physics that encapsulates the fundamental particles and their interactions to provide an ontological basis to describe the organization of all matter forms except the ones related to gravity. The fundamental particles are classified into two categories: the matter particles, known as “fermions”, and the mediators of interactions, known as “bosons”. In total, SM consists of 17 fundamentals including 12 fermions and 5 bosons. This model, over the years, has been found to be immensely successful in all of its

experimental testings.

1.2.1 Matter particles

All the universal matter is made up of fermions, one set of fundamental particles. These fermions carry spin- $\frac{1}{2}$ and can be further classified into leptons and quarks; each group consisting of six particles related in pairs, thereby forming three generations.

The three quark generation involves: the up (u) and down (d) quarks, forming 1st generation, followed by the strange (s) and charm (c) quarks as 2nd generation and the top (t) and bottom (b) quarks as 3rd generation. These quarks carry fractional charges of $+\frac{2}{3}e$ and $-\frac{1}{3}e$, e being the electronic charge, and exhibit electromagnetic, strong as well as weak forces. Quarks also carry color charge, a property which does not allow them to exist freely in nature.

The three generation of leptons are: electron (e) and electron neutrino (ν_e), muon (μ) and muon neutrino (ν_μ), and tau (τ) and tau neutrino (ν_τ). Each lepton (e, μ, τ) carry one unit of negative charge while corresponding neutrinos are chargeless. SM predicts the neutrinos to be massless, however, recent discoveries of neutrino oscillations [66–71] assure that they have small non-zero mass. The leptons take part in electromagnetic and weak interactions while neutrinos involve in weak interactions only. The three generations of quarks and leptons along with their crucial properties [72] are summarized in Table 1.1.

	Leptons			Quarks		
	Flavor	Charge	Mass (MeV/C ²)	Flavor	Charge	Mass (MeV/C ²)
1 st Generation	e	-1	0.511	u	$+\frac{2}{3}$	~ 2.2
	ν_e	0	$< 2.2 \times 10^{-6}$	d	$-\frac{1}{3}$	~ 4.7
2 nd Generation	μ	-1	105.7	c	$+\frac{2}{3}$	$\sim 1.27 \times 10^3$
	ν_μ	0	< 0.17	s	$-\frac{1}{3}$	~ 96
3 rd Generation	τ	-1	1776.9	t	$+\frac{2}{3}$	$\sim 173.21 \times 10^3$
	ν_τ	0	< 18.2	b	$-\frac{1}{3}$	$\sim 4.18 \times 10^3$

Table 1.1: Standard model fundamental fermion’s three generations.

For each particle listed in table above, corresponding anti-particle having same mass and spin but opposite charge exists.

1.2.2 Forces and messenger particles

All universal phenomena exhibit and experience four different kind of forces: strong, weak, electromagnetic and gravitational. The gravity and electromagnetic interaction are the only two forces that can be experienced in day-to-day life. Strong and weak forces are applicable at sub-nuclear level only. Standard model describes each force, except gravity, in the form of a physical theory, known as the Quantum Field Theory (QFT). These theories are based on the analogy of exchange forces via mediators, the spin-1 bosons; another set of fundamental particles. These mediators for electromagnetic interaction are “photons” whereas for strong force are “gluons” and for weak force are “intermediate vector bosons” $W^\pm$ and Z . These forces operate between particles due to different charges: EM $\rightarrow$ electric charge, Strong $\rightarrow$ color charge, Weak $\rightarrow$ flavor charge. The Table 1.2 summarize the important properties [72] of all the forces.

Force	Mediator	Charge	Mass (GeV/C ²)	Range (m)	Rel. Strength	Long distance behavior
Strong	Gluons (g)	0	0	10^{-15}	1	$\sim r$
Electromagnetic	Photon (γ)	0	0	∞	$1/137$	$1/r^2$
Weak	$W^\pm$	± 1	80.38	10^{-18}	10^{-5}	$1/r e^{-mr}$
	Z^0	0	91.18			
Gravity	Graviton (G)	0	0	∞	10^{-38}	$1/r^2$

Table 1.2: Important properties of fundamental forces of universe.

The electromagnetic interaction is the unification of electric and magnetic forces, first formulated in 1864 by Maxwell. The QFT of electromagnetic forces is known as Quantum Electrodynamics (QED) [18, 19, 21–23, 73, 74], perfected by Feynman, Tomonaga, and Schwinger in the 1940s. EM interactions are of infinite range, thereby requiring its mediator (photon) to be massless. The magnitude of this interaction is defined by means of a coupling constant, $\alpha(Q^2) = \frac{\alpha(\mu^2)}{1 - \frac{\alpha(\mu^2)}{3\pi} \log\left(\frac{Q^2}{\mu^2}\right)}$ where μ is the typical scale of renormalization. The interaction strength increases with the increase of interaction energy and decrease of interaction length.

The weak force is answerable for existence of life on earth owing to its radioactive decays. It is actually the only force responsible for decay of particles. It is also called Quantum Flavourdynamics (QFD) because of its capability of changing the flavour of quarks. The first ever theory of weak forces was represented by Fermi in 1934; later refined by Feynman and Gell-mann and finally, presented in

its current scheme by Glashow, Weinberg, and Salam, in 1960s [1–3], in the form of Electroweak theory. This theory envision the unification of electromagnetic and weak forces at some higher energy regime, thereby, predicting the existence of massive weak bosons, $W^\pm$ and Z . The mechanism that is responsible for providing mass to weak bosons is known as Higgs mechanism that also predict the existence of one more fundamental boson called Higgs boson. The magnitude of weak forces is found to depend on the coupling strength of weak bosons as well as their masses.

The strong force is the mechanism responsible for the binding of quarks inside protons and neutrons, thereby, ensuring the stability of nucleus. Most of the mass of a proton (or neutron) arises from the field energy provided by the strong force, the contribution of quark mass is almost negligible in total proton’s mass. The field theory of strong force is known as Quantum Chromodynamics (QCD) that explains the interaction mechanism of strong force in the form of exchange of massless gluons. The gluons interact with quarks as well as other gluons by means of a charge known as color charge. Color charge is analogous to electric charge but it comes in three different types. The behaviour of partons¹ within the nucleons are determined by the properties of asymptotic freedom and confinement. The coupling constant [75] determining the QCD interaction strength is given by,

$$\alpha_s(Q^2) = \frac{12\pi}{(11c - 2n_f) \ln\left(\frac{Q^2}{\Lambda^2}\right)} \quad (1.1)$$

where Q is the momentum transferred, c is the total number of quark colors (i.e. 3), n_f is the total number of quark flavors (i.e. 6), Λ is the chromodynamics scale, which is given by:

$$\Lambda^2 = \mu^2 \left[\frac{-12\pi}{(11c - 2n_f)\alpha_s(\mu^2)} \right] \quad (1.2)$$

where μ is the renormalization scale. For $Q^2 \rightarrow \infty$ and hence distance $\rightarrow 0$, $\alpha_s(Q^2) \rightarrow 0$, which means that at high energies with large momentum transfers in the interactions ($Q^2 \gg \Lambda^2$), the coupling strength between the partons decreases, thereby making the quarks and gluons to behave as free particles inside the nucleus. This property of quarks is referred to as “asymptotic freedom” [76–78]. For low Q^2 , partons form hadronic bound states and can not be observed in isolated form. This attribute is referred to as “quark confinement”.

¹Together quarks and gluons are referred to as partons

1.2.3 SM as a gauge theory

The standard model has been established on the notion of “symmetry”. As per the statement of the Noether’s theorem: the continuous symmetries of a system are associated with conservation laws. For example, invariance of the laws of physics under space and time transformations, refer to as the symmetries of nature, leads to the conservation of momentum and energy. These symmetry transformations of a system can be easily modelled in the form of mathematical group representations, like the rotational symmetry of a system in n -dimension form the group representation for the special orthogonal group $SO(n)$.

A symmetry can be labelled as a global or local symmetry depending upon whether the transformation is same or different for each space-time point. A physical theory of a system constructed by requiring local symmetries is known as “gauge theory”, while the corresponding symmetry transformations are called “gauge transformations”. The corresponding associated group representation is known as “gauge group”. The gauge theory is called non-abelian if the underlying gauge group is non-commutative. In SM, a physical system is described in terms of a Lagrangian which is then required to remain invariant under a local gauge transformation. This requirement leads to the appearance of additional fields (known as “gauge fields”) in the Lagrangian that couples with the particles of the system, thereby generating the fundamental interactions. The number of gauge fields in a gauge theory is determined by the dimension of the gauge group and the quanta of these gauge fields are known as the “gauge bosons”.

The standard model portray the three fundamental forces in the form of gauge theories and unifies them as a non-abelian gauge theory under $SU(3)_{color} \otimes SU(2)_{flavor} \otimes U(1)_{charge}$ gauge group. A more detailed description of this formalism is outlined as:

1.2.3.1 Quantum electrodynamics

Quantum electrodynamics is the abelian gauge theory of electromagnetic interactions characterized by the symmetry group $U(1)$. The dynamics of a free particle of spin- $\frac{1}{2}$ can be represented in terms of Dirac Lagrangian,

$$\mathcal{L}_\Psi = \bar{\Psi}(x)(i\gamma^\mu\partial_\mu - m)\Psi(x) \tag{1.3}$$

where $\Psi(x)$ is the Dirac field and γ^μ are the 4×4 Dirac matrices. We define a symmetry operation of the global phase transformations:

$$\Psi(x) \rightarrow \Psi'(x) = e^{i\alpha}\Psi(x) \quad ; \quad \bar{\Psi}(x) \rightarrow \bar{\Psi}'(x) = e^{-i\alpha}\bar{\Psi}(x) \quad (1.4)$$

where phase α is independent of the spacetime variation. The Lagrangian remains invariant under this transformation, $\Delta\mathcal{L} = 0$. As per the statement of Noether's theorem, this invariance corresponds to a current j^μ , known as Noether's current, defined by,

$$j^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\Psi)}\delta\Psi \propto \bar{\Psi}\gamma^\mu\Psi, \quad (1.5)$$

which remains conserved, that is, $\partial_\mu j^\mu = 0$. Now, if the phase is made to vary locally as a function of spacetime points, such that

$$\Psi(x) \rightarrow \Psi'(x) = e^{i\alpha(x)}\Psi(x) \quad ; \quad \bar{\Psi}(x) \rightarrow \bar{\Psi}'(x) = e^{-i\alpha(x)}\bar{\Psi}(x) \quad (1.6)$$

then under this symmetry operation, the Lagrangian transforms as:

$$\mathcal{L}_\Psi \rightarrow \mathcal{L}'_\Psi = \mathcal{L}_\Psi - \bar{\Psi}\gamma^\mu\Psi\partial_\mu\alpha(x) \quad (1.7)$$

which is no longer invariant. In an attempt to achieve invariance, we introduce a covariant derivative D_μ and a new gauge field A_μ such as

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - ieA_\mu \quad ; \quad A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{e}\partial_\mu\alpha(x) \quad (1.8)$$

Under this transformation, the Lagrangian changed as

$$\begin{aligned} \mathcal{L}_\Psi &\rightarrow \mathcal{L}'_\Psi = \bar{\Psi}'(i\gamma^\mu D_\mu - m)\Psi' \\ &= \bar{\Psi}e^{-i\alpha(x)}(i\gamma^\mu(\partial_\mu - ieA'_\mu) - m)e^{i\alpha(x)}\Psi \\ &= \bar{\Psi}e^{-i\alpha(x)}(i\gamma^\mu\partial_\mu + e\gamma^\mu(A_\mu + \frac{1}{e}\partial_\mu\alpha(x)) - m)e^{i\alpha(x)}\Psi \\ &= \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi + e\bar{\Psi}\gamma^\mu A_\mu\Psi \\ &= \mathcal{L}_\Psi + e\bar{\Psi}\gamma^\mu A_\mu\Psi \end{aligned} \quad (1.9)$$

which is invariant under the local gauge transformation defined in Eq. 1.6.

Demanding local gauge invariance under the gauge group $U(1)$ has resulted into the introduction of gauge field A_μ which is none other than the required electromagnetic field, also called the photon field, and it couples with Dirac

particles (e^- s) through the term $e\bar{\Psi}\gamma^\mu A_\mu\Psi$. Since A_μ is a physical field, we must introduce a gauge invariant kinetic energy term for this, given by

$$\mathcal{L}_{K.E.} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (1.10)$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, known as the electromagnetic strength tensor. Adding this term, the complete gauge invariant Lagrangian for electromagnetic interactions can be written as:

$$\begin{aligned} \mathcal{L}_{total} &= \mathcal{L}_{free} + \mathcal{L}_{int} + \mathcal{L}_{K.E} \\ \mathcal{L}_{total} &= \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi + e\bar{\Psi}\gamma^\mu A_\mu\Psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \end{aligned} \quad (1.11)$$

If a mass term, like $m^2 A_\mu A^\mu$, is considered for field A_μ , it will again spoil the gauge invariance. So this term is prohibited in the QED Lagrangian and corresponding gauge particle, the photon, remain massless. The interaction term in the Lagrangian, $e\bar{\Psi}\gamma^\mu A_\mu\Psi$ can also be written as $-j^\mu A_\mu$ where $j^\mu = -e\bar{\Psi}\gamma^\mu\Psi$, known as the current density, $-e$ being the electric charge. The current density j^μ is a conserved quantity of electromagnetic interactions.

1.2.3.2 Electroweak

The electroweak theory is a unified theory of electromagnetic and weak interactions. This theory came into existence in an attempt to derive a self-consistent invariant gauge theory for weak interactions, analogous to QED. The weak interactions were first proposed by Fermi in 1934 as an explanation of β -decay. He came up with the following interaction Lagrangian for β -decay process:

$$\mathcal{L}_w = G_F(\bar{\Psi}_p\gamma_\mu\Psi_n)(\bar{\Psi}_e\gamma^\mu\Psi_\nu) \quad (1.12)$$

where G_F is the Fermi coupling constant and $\Psi_p, \Psi_n, \Psi_e, \Psi_\nu$ are the particle fields. In his proposal, Fermi considered the interaction of four fermions at one point, known as contact interaction. The most remarkable difference in Fermi Lagrangian was the presence of charged currents, contrary to neutral current of QED.

The observation of parity violation in weak interactions emphasized that weak interactions involve only left-handed fermions and right-handed anti-fermions (observed experimentally in 1958 by Goldhaber et. al. [79]) and led to the introduction

of axial-vector terms in the Fermi Lagrangian,

$$\mathcal{L}_w = \frac{G_F}{\sqrt{2}} (\bar{\Psi}_p \gamma_\mu (1 - g_A \gamma_5) \Psi_n) (\bar{\Psi}_e \gamma^\mu (1 - \gamma_5) \Psi_\nu) \quad (1.13)$$

where $g_A = |\frac{G_A}{G_V}|$, G_V and G_A being the constants for vector and axial-vector couplings respectively. However, this $V - A$ theory was still incomplete in the sense that it could not be renormalized and violated unitarity at high energies.

Later on, in 1960s, it was observed independently by Glashow, Salam and Weinberg [1–3] that a gauge-invariant theory for the weak interactions can be constructed if electromagnetic interactions are also included under the gauge group $SU(2)_L \times U(1)_Y$. The group $SU(2)_L$, where L refers to left-handed, forms the gauge group of weak interactions and considers the doublet of left-handed lepton (or an up-type quark) and lepton neutrino (or a down-type quark) as its fundamental representation under the weak isospin, $T = \frac{1}{2}$, which remains conserved under this group transformation. The group $U(1)_Y$ form the gauge group under the weak hypercharge, Y . The weak isospin's 3^{rd} component, T_3 and weak hypercharge, Y are related to the electric charge, Q by the Gell-Mann-Nishijima relation [80–82]: $Q = T_3 + \frac{Y}{2}$. The Table 1.3 summarize the electroweak singlets and doublets and the various quantum numbers associated with them.

Fermions: Singlets & Doublets			Q	T	T_3	Y
$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$	0 -1	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $-\frac{1}{2}$	-1 -1
$\begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} c \\ s \end{pmatrix}_L$	$\begin{pmatrix} t \\ b \end{pmatrix}_L$	$\frac{2}{3}$ $-\frac{1}{3}$	$\frac{1}{2}$ $\frac{1}{2}$	$\frac{1}{2}$ $-\frac{1}{2}$	$\frac{1}{3}$ $\frac{1}{3}$
e_R	μ_R	τ_R	-1	0	0	-2
u_R	c_R	t_R	$\frac{2}{3}$	0	0	$\frac{4}{3}$
d_R	s_R	b_R	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$

Table 1.3: Assignment of the various quantum numbers including T_3 and Y to the electroweak singlet and doublet fundamental fermions.

If we consider only the first generation quarks and leptons, then

$$\Psi_L = \begin{pmatrix} \nu_{e^-} \\ e^- \end{pmatrix}_L \quad \text{or} \quad \begin{pmatrix} u \\ d \end{pmatrix}_L \quad ; \quad \Psi_R = e_R^-, u_R, d_R \quad (1.14)$$

form the left-handed and right-handed particle fields. The free particle Lagrangian

for these fields can be written as,

$$\mathcal{L}_{free} = \bar{\Psi}_L(i\gamma^\mu\partial_\mu - m)\Psi_L + \bar{\Psi}_R(i\gamma^\mu\partial_\mu - m)\Psi_R \quad (1.15)$$

which remains invariant under the following global gauge transformations:

$$\Psi_L \rightarrow \Psi'_L = e^{i\alpha_a T_a + i\beta \frac{Y}{2}} \Psi_L \quad ; \quad \Psi_R \rightarrow \Psi'_R = e^{i\beta \frac{Y}{2}} \Psi_R \quad (1.16)$$

where T_a and Y are the generators of $SU(2)$ and $U(1)$ transformations with $T_a = \sigma_a/2$, σ_a are the Pauli spin matrices² for $a = 1, 2, 3$. However, the invariance of Lagrangian is carried away when the phases α and β are made to vary as a function of spacetime points. In order to restore the invariance, the following covariant derivatives and fields are introduced:

$$\begin{aligned} \partial_\mu \rightarrow D_\mu &= \partial_\mu - igT_a W_\mu^a - ig' \frac{Y}{2} B_\mu \quad \text{for } SU(2) \text{ doublets} \\ \partial_\mu \rightarrow D_\mu &= \partial_\mu - ig' \frac{Y}{2} B_\mu \quad \text{for } U(1) \text{ singlets} \end{aligned} \quad (1.17)$$

where the fields W_μ and B_μ varies as

$$W_\mu \rightarrow W'_\mu = W_\mu + \frac{1}{g} \partial_\mu \alpha(x) - \alpha(x) \times W_\mu \quad ; \quad B_\mu \rightarrow B'_\mu = B_\mu + \frac{1}{g'} \partial_\mu \beta(x) \quad (1.18)$$

Under these covariant transformations, the Lagrangian achieve invariance as:

$$\begin{aligned} \mathcal{L}_{free} \rightarrow \mathcal{L}'_{free} &= \mathcal{L}_{fermion} = \mathcal{L}_{free} + g\bar{\Psi}_L \gamma^\mu T_a W_\mu^a \Psi_L + g' \frac{Y}{2} B_\mu (\bar{\Psi}_L \gamma^\mu \Psi_L + \bar{\Psi}_R \gamma^\mu \Psi_R) \\ &= \mathcal{L}_{free} + g\bar{\Psi}_L \gamma^\mu T_a W_\mu^a \Psi_L + g' \bar{\Psi} \gamma^\mu \frac{Y}{2} B_\mu \Psi \\ &= \bar{\Psi}_L (i\gamma^\mu D_\mu - m) \Psi_L + \bar{\Psi}_R (i\gamma^\mu D_\mu - m) \Psi_R \end{aligned} \quad (1.19)$$

where $T_a W_\mu^a = \frac{\sigma_a}{2} \cdot W_\mu^a = \frac{1}{2}(\sigma_1 W_\mu^1 + \sigma_2 W_\mu^2 + \sigma_3 W_\mu^3)$. Replacing the Pauli matrices with the matrix form, and using $W_\mu^\pm = (W_\mu^1 \mp iW_\mu^2)/\sqrt{2}$,

$$T_a W_\mu^a = \frac{1}{2} \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} W_\mu^3 & \sqrt{2}W_\mu^+ \\ \sqrt{2}W_\mu^- & -W_\mu^3 \end{pmatrix} \quad (1.20)$$

Thus, the invariance under local gauge transformations led to the introduction of four gauge bosons into the theory: a positively charged $W^{\mu+}$, a negatively

²Pauli matrices are a set of three 2×2 hermitian and unitary complex matrices, given as:
 $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$

charged $W^{\mu-}$ and two neutrals $W^{\mu3}$, and B^μ where W_μ 's interacts with only the left-handed fermions while B_μ interacts with both left and right handed. The kinetic energy terms for the gauge fields can be written as,

$$\begin{aligned}\mathcal{L}_{K.E.} &= -\frac{1}{4}W^{\mu\nu a}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \quad \text{where,} \\ W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon_{abc}W_\mu^b W_\nu^c, \quad \epsilon_{abc} \text{ is totally anti-symmetric tensor} \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu\end{aligned}\tag{1.21}$$

The corresponding weak current densities that will remain conserved under $SU(2) \times U(1)$ are represented as:

$$j_\mu^T = -g\bar{\Psi}_L\gamma_\mu T\Psi_L \quad ; \quad j_\mu^Y = -g'\bar{\Psi}\gamma_\mu Y\Psi\tag{1.22}$$

Using the general form of electromagnetic current, $j_\mu^{em} = -e\bar{\Psi}\gamma_\mu Q\Psi$, where Q is the charge operator for EM interactions, and Gell-Mann-Nishijima relation, $Q = T_3 + \frac{Y}{2}$,

$$j_\mu^{em} = j_\mu^3 + \frac{j_\mu^Y}{2}\tag{1.23}$$

Thus, the electromagnetic current is a combination of weak isospin and weak hypercharge current. Likewise, the two neutral physical gauge fields, Z^μ and A^μ are the orthogonal combinations of gauge fields $W^{\mu3}$ and B^μ with a mixing angle θ_W ,

$$\begin{aligned}Z^\mu &= W^{\mu3} \cos \theta_W - B^\mu \sin \theta_W \\ A^\mu &= W^{\mu3} \sin \theta_W + B^\mu \cos \theta_W\end{aligned}\tag{1.24}$$

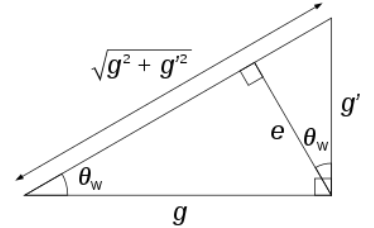


Figure 1.2: $W^{\mu3}$ - B^μ plane

obtained by considering a plane in $W_\mu^3 - B_\mu$ with W_μ^3 along x-axis and B_μ along y-axis. The relation among θ_W , g and g' is obtained as

$$\cos \theta_W = \frac{g}{\sqrt{g^2 + g'^2}} \quad ; \quad \sin \theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}\tag{1.25}$$

Also, using Eq. 1.23, the following relation between e , g and g' holds,

$$e = g \sin \theta_W = g' \cos \theta_W\tag{1.26}$$

Thus, the gauge invariance has been restored, but no required mass term appears for the gauge bosons. Any mass term like $m^2 W^\mu W_\mu$ will break the gauge invariance. This hints towards an underlying mechanism, which provide the masses to gauge bosons alongwith keeping the theory renormalizable and gauge invariant. This mechanism is known as Higgs mechanism [51–55].

Higgs mechanism: The higgs mechanism to endow the gauge bosons with mass has been developed on the idea of “spontaneous symmetry breaking” [83] which considers the breaking of a symmetry of a physical system such that the underlying laws remains invariant under the symmetry transformation, but the physical system itself changes. It is a spontaneous action that changes the system from a symmetrical state to an asymmetrical state.

To realize this, a complex scalar field Lagrangian, gauge invariant under $SU(2) \times U(1)$ has been introduced into the electroweak theory using the same covariant derivative defined in Eq. 1.17.

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger D^\mu \Phi - \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2 \quad (1.27)$$

where Φ is a $SU(2) \times U(1)$ multiplet of complex scalar fields having weak isospin, $T = 1/2$ and weak hypercharge, $Y = 1$:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.28)$$

The potential term of the Lagrangian, $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$ has two parameters, μ^2 and λ , that decide its nature. The four particle coupling λ specifies the self interactions among the ϕ fields and is always positive. We have two choices for μ^2 : $\mu^2 > 0$ and $\mu^2 < 0$. For positive μ^2 , $V(\Phi)$ has a minimum at $\Phi = 0$ and the Lagrangian describes the QED interaction among scalar particles, but for negative μ^2 , the potential behaves differently with ground state exhibited by, $\Phi^\dagger \Phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2}$, in a multi-dimensional regime. The corresponding potential for real scalar fields before and after symmetry breaking is presented in Fig. 1.3.

Choosing one particular minimum will break the symmetry spontaneously. We choose a ground state where $\phi_3 = v$, $\phi_1 = \phi_2 = \phi_4 = 0$ so that

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad \text{with } v = \frac{\mu}{\sqrt{\lambda}} \quad (1.29)$$

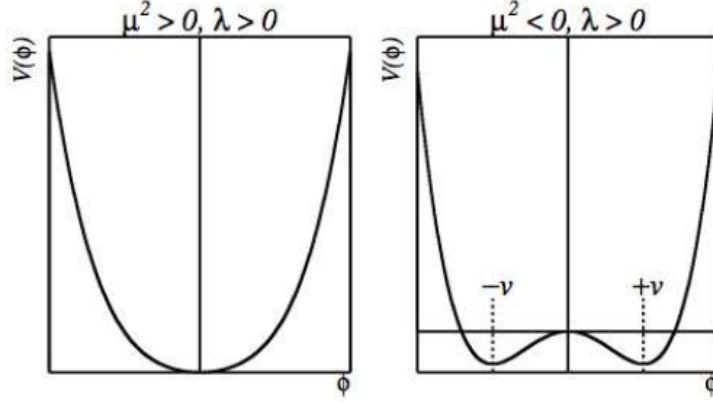


Figure 1.3: A potential with real scalar field $V(\phi) = \mu^2\phi^2 + \lambda\phi^4$ before and after spontaneous symmetry breaking.

This particular choice will break both $SU(2)_L$ and $U(1)_Y$ gauge symmetries but maintain the $U(1)_{em}$ gauge symmetry. Thus the vacuum state remain invariant under $U(1)_{em}$ and photon does not acquire any mass. We perturb this field by considering small quantum fluctuations $h(x)$, $h_1(x)$, $h_2(x)$ and $h_4(x)$ corresponding to ϕ_3 , ϕ_1 , ϕ_2 and ϕ_4 fields respectively, such that

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} h_1(x) + ih_2(x) \\ v + h(x) + ih_4(x) \end{pmatrix} \quad (1.30)$$

Substituting this value of Φ in the Lagrangian and using the relations defined in Eqs. 1.24, 1.25 and 1.26 results into,

$$\mathcal{L}_\Phi = \frac{1}{2}(\partial_\mu h)^2 + \frac{1}{4}g^2(v+h)^2(W^{\mu+}W_\mu^-) + \frac{1}{8}(g^2+g'^2)(v+h)^2Z^\mu Z_\mu - \lambda v^2h^2 - \lambda v h^3 - \frac{\lambda h^4}{4} + \frac{\lambda v^4}{4} \quad (1.31)$$

This Lagrangian consists of gauge invariant massive higgs field h with mass, $m_h = \sqrt{2\lambda}v$, along with the mass terms for weak gauge bosons $W_\mu^\pm$ and Z_μ . It can be seen that there is no mass term for photon field A_μ .

$$M_{W^\pm} = \frac{1}{2}gv, \quad M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}; \quad M_A = 0 \quad (1.32)$$

Thus, the spontaneous breaking of gauge symmetry has resulted into the generation of masses of weak gauge bosons. The other terms of the Lagrangian represents the interaction terms of the higgs field with gauge bosons. The same higgs field also provide the masses to fermions through Yukawa coupling defined by,

$$\mathcal{L}_{Yukawa} = -y_l \bar{L} \phi e_R - y_u \bar{Q} (-i\sigma_2) \phi^* u_R - y_d \bar{Q} \phi d_R + h.c. \quad (1.33)$$

with $y_{l,u,d}$ being Yukawa coupling constants and u_R, d_R (e_R) being the left-handed doublet and right-handed singlet quark (lepton) fields. Now again we break the symmetry and perturb the system around the minima to obtain a value for field Φ defined in Eq. 1.30. Using it in the Lagrangian gives,

$$\mathcal{L}_{Yukawa} = -\frac{1}{\sqrt{2}} y_l \bar{l}_L (v + h) l_R - \frac{1}{\sqrt{2}} y_q \bar{q}_L (v + h) q_R, \quad (1.34)$$

where $l_{L,R}$ and $q_{L,R}$ denote the lepton and quark fields. We see that a mass term for fermions appears in this lagrangian given by, $m_f = \frac{1}{\sqrt{2}} y_f v$, showing that the coupling strength y_f of fermions with higgs field is proportional to the mass of fermion.

Thus, the full electroweak Lagrangian is a sum of four terms,

$$\mathcal{L}_{EWK} = \mathcal{L}_{fermion} + \mathcal{L}_{K.E.} + \mathcal{L}_{\Phi} + \mathcal{L}_{Yukawa} \quad (1.35)$$

1.2.3.3 Quantum chromodynamics

The theory of strong force, known as Quantum chromodynamics, has also been developed on the steps of QED. Since quarks come in three colors, in contrast to the single charge of QED photon, the gauge group considered is $SU(3)$, an extension of $U(1)$ gauge group in 3-dimension. It is a non-abelian gauge theory with quark's color fields forming the fundamental representation. The Lagrangian for free quark fields can be written as,

$$\mathcal{L}_q = \sum_{j=1,2,3} \bar{q}_j(x) (i\gamma^\mu \partial_\mu - m) q_j(x) \quad (1.36)$$

This Lagrangian is conserved under the global phase transformation with the conserved currents, $j_\mu^c = \sum_{j=1,2,3} g \bar{q}_j \gamma^\mu T_a q_j$, known as color charge densities. Considering the invariance of Lagrangian under the local gauge transformations,

$$q_j(x) \rightarrow q'_j(x) = e^{i\alpha_a(x) T_a} q_j(x) \quad (1.37)$$

where T_a , $a = 1, 2, \dots, 8$ are the generators of $SU(3)$ transformations with $[T_a, T_b] = if_{abc} T_c$, f_{abc} are anti-symmetric real structure constants. To make Lagrangian

invariant under this local phase, we introduce the following covariant derivatives and fields,

$$D_\mu = \partial_\mu + igT_a G_\mu^a \quad ; \quad G_\mu^a \rightarrow G_\mu^a - \frac{1}{g} \partial_\mu \alpha_a(x) - f_{abc} \alpha_b(x) G_\mu^c \quad (1.38)$$

where last term of G_μ^a signifies the non-abelian nature of the theory. Substituting $\partial_\mu \rightarrow D_\mu$ in the Lagrangian,

$$\mathcal{L}_q \rightarrow \mathcal{L}'_q = \mathcal{L}_q - \sum_{j=1,2,3} g \bar{q}_j \gamma^\mu T_a G_\mu^a q_j \quad (1.39)$$

Also, considering the kinetic energy term for the gauge fields, G_μ^a , we can write the complete QCD Lagrangian as,

$$\mathcal{L}_q = \sum_{j=1,2,3} \bar{q}_j(x) (i \gamma^\mu \partial_\mu - m) q_j(x) - g \bar{q}_j \gamma^\mu T_a G_\mu^a q_j - \frac{1}{4} G_{\mu\nu}^a G^{\mu\nu a} \quad (1.40)$$

where,

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g f_{abc} G_\mu^b G_\nu^c \quad (1.41)$$

Thus, we can see that the invariance of Lagrangian under local gauge has resulted into the appearance of interaction fields into the theory with the mediators of interaction, gluons, remaining massless.

1.2.4 Challenges faced by SM: unknown mysteries

The standard model is acknowledged to be the most successful theoretical model till today. Over the years, it has been built up owing to the lively conversations between theoretical predictions and experimental proofs. Even after being able to describe so many natural phenomena, this model is not considered perfect. The reason being several existing questions that standard model can not adequately explain. Some of these include:

- **Gravity:** Standard model does not include gravity and provides no explanation for it.
- **Three generations:** No particular justification has been given in SM as to why only three particle generations exist and why there is so much variation in masses of particles from different generations.
- **Neutrino mass:** In SM, neutrinos are regarded as massless. However, there are many experimental results [66, 67, 69, 70] which prove that neutrinos have

small tiny non-zero mass.

- **Matter-Antimatter asymmetry:** The imbalance of matter and anti-matter within the observable universe, also referred to as baryon asymmetry, has been one of the greatest mysteries of particle physics for which standard model has no answer.
- **Dark energy and dark matter:** It has been observed in various cosmological observations that the visible matter accounts for only 5% of the total energy-mass content of the universe. The SM gives no explanation for the rest of about 25% of the non-luminous dark matter and about 70% of dark energy.
- **Short range of strong force:** Usually, the range of a force is inversely proportional to the mass of its mediator. But this is not true for strong force. Its mediators, gluons are massless, still, this force is of limited range. No explanation for this is provided in SM.

So the picture depicted by SM is not perfect, it has been, therefore, regarded as a piece of a bigger picture that will be explained by some beyond the standard model theories.

1.3 Beyond The Standard Model

Many theoretical models have been developed to overcome the deficiencies observed in standard model. These models are collectively known as “Beyond the Standard Model” (BSM) theories. The most famous among these are String theory, Supersymmetry, Technicolor, compositeness etc.

1.3.1 Model of excited quarks

The study represented in this thesis is based on a model that considers the excited state of quarks (q^*) and leptons (l^*), also known as compositeness models [84–87]. The prime motivation for these models are the presence of three quark and lepton generations with similar properties. This replication points toward some underlying structure identical in all the families. The most implicit signature of this sub-structure is the observation of a fermion (quark or lepton) in the excited state. The analogy with excited states has been made on the basis of a very basic physics phenomena. We know that a composite system, e.g. a hydrogen atom, consists of

various energy levels arising due to the interactions among its constituents. Thus, the observation of a system in excited state is the confirmation of its composite structure. The excited state of fermions is assumed as a resonance state which then eventually decay into SM particles.

The fundamental particles that quarks and leptons are considered to be composed of, are known as “preons”. These preons are postulated to experience a new kind of force that becomes very strong at a particular scale Λ , known as the compositeness scale, forming quarks and leptons bound states. Excited quarks and leptons can interact with their ordinary counterparts through contact interactions for $\Lambda \gg \sqrt{\hat{s}}$ and through gauge mediation for $\Lambda < \sqrt{\hat{s}}$ where $\sqrt{\hat{s}}$ is the center of mass energy. The model used in this study assumes compositeness scale $\Lambda < \sqrt{\hat{s}}$ and mass scale of excited quarks $M_{q^*} = \Lambda$, gauge interactions dominate over contact interactions in this regime. The excited quark model used in this study [87] considers a $SU(3) \times SU(2) \times U(1)$ symmetry with ground state quarks in the form of left-handed doublets and right-handed singlets while the excited quarks in the form of left- and right-handed doublets.

$$\text{Ground state: } \begin{pmatrix} u \\ d \end{pmatrix}_L, u_R, d_R \quad \text{Excited state: } \begin{pmatrix} u^* \\ d^* \end{pmatrix}_L, \begin{pmatrix} u^* \\ d^* \end{pmatrix}_R \quad (1.42)$$

The excited quarks interact with gauge bosons via vectorlike coupling defined by the Lagrangian:

$$\mathcal{L}_{gauge} = \bar{q}^* \gamma^\mu \left[g_s \frac{\lambda^a}{2} G_\mu^a + g \frac{\tau^b}{2} W_\mu^b + g' \frac{Y}{2} B_\mu \right] q^* \quad (1.43)$$

where g_s , g , and g' are the coupling constants for strong and electroweak interactions; G_μ^a , W_μ , and B_μ denote the gauge fields corresponding to $SU(3)$, $SU(2)$, and $U(1)$ gauge groups. The excited quarks attain spin and isospin value of $\frac{1}{2}$ and weak hypercharge Y value of $\frac{1}{3}$ under this model. The gauge interaction mediated between ordinary and excited quarks by gauge bosons is obtained through the requirement of gauge invariance and is of magnetic-moment type, defined by the effective Lagrangian:

$$\mathcal{L}_{int} = \frac{1}{2\Lambda} \bar{q}_R^* \sigma^{\mu\nu} \left[g_s f_s \frac{\lambda^a}{2} G_{\mu\nu}^a + g f \frac{\tau^b}{2} W_{\mu\nu}^b + g' f' \frac{Y}{2} B_{\mu\nu} \right] q_L + h.c., \quad (1.44)$$

Here $G_{\mu\nu}^a$, $W_{\mu\nu}$, and $B_{\mu\nu}$ are corresponding field-strength tensors while λ_a , τ_b , Y are the corresponding generators for $SU(3)$, $SU(2)$, and $U(1)$ fields. The

constants f_s, f, f' determine the strength of a coupling and are evaluated using the compositeness dynamics, these are considered to be equal to unity to allow for the standard model like coupling strengths.

A composite quark can acquire an excited state by the absorption of a gluon and will radiate a photon, gluon or weak boson on returning to the ground state. The branching fractions of excited quark decays into various final states is given in the Table 1.4 below.

Decay Channel	BR
$q^* \rightarrow qg$	$\sim 83\%$
$q^* \rightarrow q\gamma$	$\sim 2\%$
$q^* \rightarrow qZ^0$	$\sim 5\%$
$q^* \rightarrow qW^-$	$\sim 10\%$

Table 1.4: Approximate estimate of branching fractions (BR) of q^* decay in different channels, assumptions made are: $\Lambda = M_{q^*}$ and $f = f' = f_s$.

This study considers an excited light (u^*, d^*) and heavy (b^*) flavor quark decaying into a quark and a photon final state. The physics processes contributing to this final state are represented in the form of Feynman diagrams in the Fig. 1.4 with q referring to u, d and b quark.

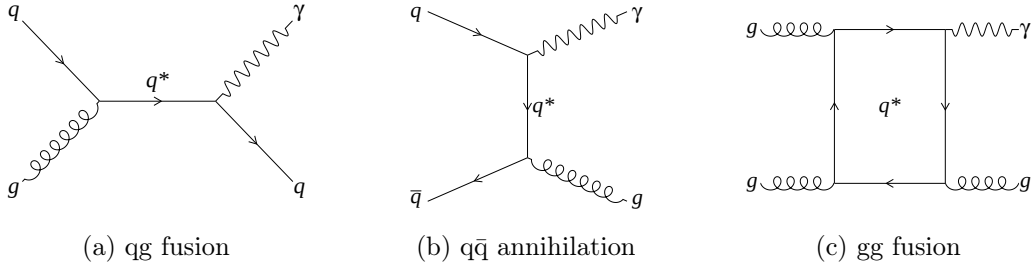


Figure 1.4: Feynman diagrams for $q^* \rightarrow \gamma + \text{jet}$ final state with (a) quark-gluon fusion (b) quark-antiquark annihilation (c) gluon-gluon fusion processes. Whereas the light quarks will show their presence in all the three modes, heavy bottom quark will appear only in quark-gluon fusion and quark-antiquark annihilation.

Under the assumption $\Lambda < \sqrt{\hat{s}}$, gauge interactions will be dominating and the excited quarks will be produced mainly via the s-channel³ quark-gluon fusion

³ s, t and u are called Mandelstam variables. For a 2-2 scattering process of initial momenta p_1 and p_2 and final momenta p_3 and p_4 , the variables are defined as, $s = (p_1 + p_2)^2 = (p_3 + p_4)^2$, $t = (p_1 - p_3)^2 = (p_4 - p_2)^2$ and $u = (p_1 - p_4)^2 = (p_3 - p_2)^2$. Each one represents a different Feynman diagram of a process.

process. The scattering of an ordinary quark and gluon in s-channel will produce an excited quark resonance that will tower over the continuous SM γ +jet background. Appearance of excited quark in other channels will be captured in the form of excess of events over the background distribution.

1.3.2 A follow-up of previous searches

Many attempts have been made in the past to look for the evidences of quark compositeness. But no success has been gained yet. A summary of the various experiments that deployed their potential in this important search are reported here.

The ZEUS detector at HERA looked for excited quarks of heavy flavor in e^+p collisions at $\sqrt{s} = 300$ GeV and excluded the excited quarks in the mass range 40-169 GeV at the 95% confidence level (CL) [88].

The CDF detector collaboration at Tevatron also searched for excited quarks in γ +jet and dijet channels in $p\bar{p}$ collisions at center of mass energy of 1.8 TeV and excluded the ranges $80 < M_{q^*} < 540$ GeV [89], $200 < M_{q^*} < 520$ GeV & $580 < M_{q^*} < 760$ GeV [90], respectively. The D0 collaboration at Tevatron looked for evidence in dijet final state [91] and excluded M_{q^*} below 775 GeV at 95% CL. The CMS and ATLAS Collaborations have also performed the searches in pp collisions at $\sqrt{s} = 7$ TeV, 8 TeV as well as 13 TeV. At $\sqrt{s} = 8$ TeV, exclusions have been reported by both the collaborations in different channels. At $\sqrt{s} = 13$ TeV, ATLAS had excluded q^* below 5.2 TeV with 95% CL in the dijet channel [92] and below 4.4 TeV in the γ +jet channel [93]. The CMS Collaboration has also excluded excited quarks below 5.0 TeV in the dijet channel [94].

Searches for excited states of heavy bottom quarks have also been performed earlier in different channels. The ATLAS collaboration searched for b^* in tW final state at $\sqrt{s} = 7$ TeV [95] and excluded M_{b^*} for masses below 870 GeV. The CMS collaboration have looked for excited b-quarks in tW and dijet final state at $\sqrt{s} = 8$ TeV. In tW final state, it excluded b^* upto the masses of 1.53 TeV for vector-like couplings [96] at 95% CL while in dijet final state, it claimed an exclusion in the range of 1.20-1.56 TeV [97]. At $\sqrt{s} = 13$ TeV, no search for b^* has been performed yet, this study, therefore, presents the first search for b^* at $\sqrt{s} = 13$ TeV.

THE EXPERIMENTAL APPARATUS

The “Large Hadron Collider” (LHC) accelerator and the “Compact Muon Solenoid” (CMS) detector, situated across Swiss-French border in Geneva, Switzerland at the European Organization of Nuclear Research (CERN), has been used to accumulate data for this study. This chapter presents a brief introduction and description of the design and performance of these accelerator and detector facilities.

2.1 The Large Hadron Collider

The Big-Bang theory [98] provides an explanation of the origin of the universe. It says that the universe initially started from a small singularity and an explosion (called Big-Bang) put the universe into a continuous state of inflation, so it took around 13.7 billion years for the universe to reach in the present state (Fig. 2.1). During the initial seconds after the birth of universe, the temperature was tremendously high and the universe consists of plasma of vast number of fundamental particles interacting with each other with extremely high energies. In order to study the properties of early universe, scientists use particle accelerators in which sub-atomic particles travelling at relativistic speeds are made to undergo inelastic collisions to reproduce the conditions that existed at the time of Big-Bang or soon there after.

The Large Hadron Collider [99, 100] as the name suggests is a large circular

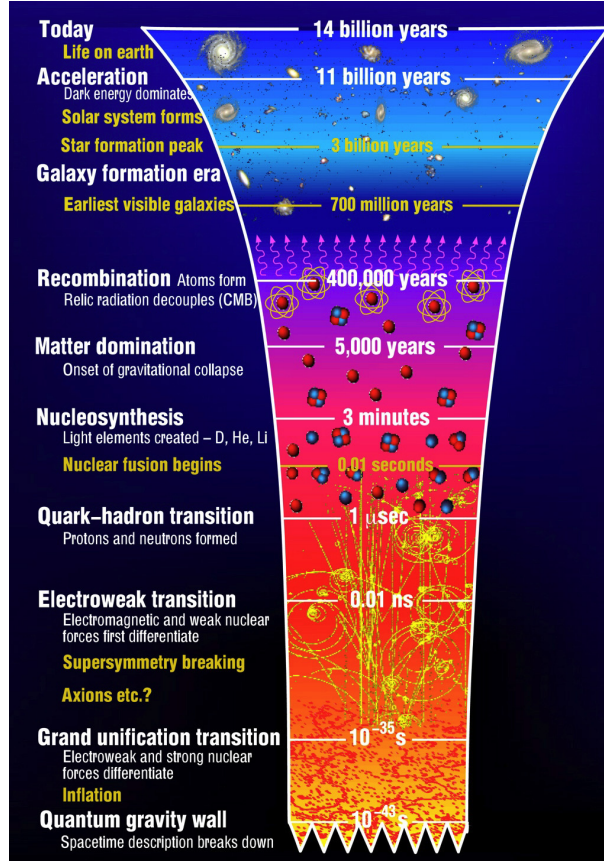


Figure 2.1: Big-Bang timeline.

accelerator having a circumference of ~ 27 km which accelerates counter rotating beams of hadrons and collides them to study the fundamental properties and interactions of quarks, leptons and bosons. It is the world's largest and most complex particle accelerator ever built. Earlier particle accelerators like “Large Electron Positron” collider (LEP) [101] at CERN (in the same tunnel where LHC has been built) and Tevatron [102] at Fermilab, Batavia, IL, USA, have been extensively used to confirm the predictions and test the precision of Standard Model parameters. Both of these were circular colliders. The LEP was an e^+e^- collider with a center of mass energy ($\sqrt{s}$) ranging from 90 GeV to 209 GeV while Tevatron was a $p\bar{p}$ collider with $\sqrt{s} = 1.8$ TeV and 1.96 TeV. Studying a collision data obtained from electron-positron beam is far more easier as compared to proton beams due to the absence of hadronic background in the former. However, in a circular ring of a particular radius, the loss of energy caused by synchrotron radiations is given as:

$$-\Delta E = \frac{4\pi\alpha}{3r}\beta^3\gamma^4 \quad (\beta = v/c \simeq 1 \text{ and } \gamma = E/mc^2 \text{ for relativistic speeds}), \quad (2.1)$$

which is very large due to small mass of e^+ and e^- and further increases as the speed v approaches the speed of light c . The choice is to either increase the radius of the circular rings or to use linear colliders or use heavy particles in collision beams. The first two instances are not very easy to meet either due to technical or financial constraints, so in order to achieve very high energies, proton beams are used. Since $m_p \simeq 2000 m_e$, use of proton beam reduces synchrotron losses by a factor of $(2000)^4 \simeq 10^{13}$.

The requirement of very high energies for particle accelerators can be explained by the fact that in scattering experiments, the wavelength of the probe should be smaller than the target particle size in order to observe the particle. Since, in particle physics, we look for extraordinarily small entities, we need probes of smaller wavelengths or higher energies (as per the de broglie's relation, $\lambda = \frac{h}{mv}$).

2.1.1 LHC tunnel layout

The LHC is situated in an underground tunnel having a circumference of 26.7 km and a diameter of 3.8 m at a depth of around 45 to 170 m beneath the France-Switzerland border. The LHC tunnel is divided in eight straight sections, each of 528 m separated by eight arcs. The straight sections are utilized as experimental or utility insertions. A schematic diagram representing the layout of LHC tunnel is shown in Fig. 2.2.

The eight sections are numbered from 1 to 8. The two general purpose detectors, namely, A Toroidal LHC ApparatuS (ATLAS) [103] and Compact Muon Solenoid (CMS) [104] are located at point 1 and point 5 respectively. Point 2 is occupied by the heavy ion detector, A Large Ion Collider Experiment (ALICE) [105] while the detector for b-physics study, Large Hadron Collider beauty (LHCb) [106] is placed at point 8. These two points also have the beam injection utility for the two beams. Collimation systems utilized for beam cleaning are placed at point 3 and 7 while the radio frequency (RF) systems for each beam are present at point 4. The point 6 has been used for the beam dump where the two beams are extracted vertically using horizontally and vertically deflecting magnets. The LHC is primarily a proton-proton collider, but it also uses heavy ion beams to perform proton-lead and lead-lead collisions for some short periods of time.

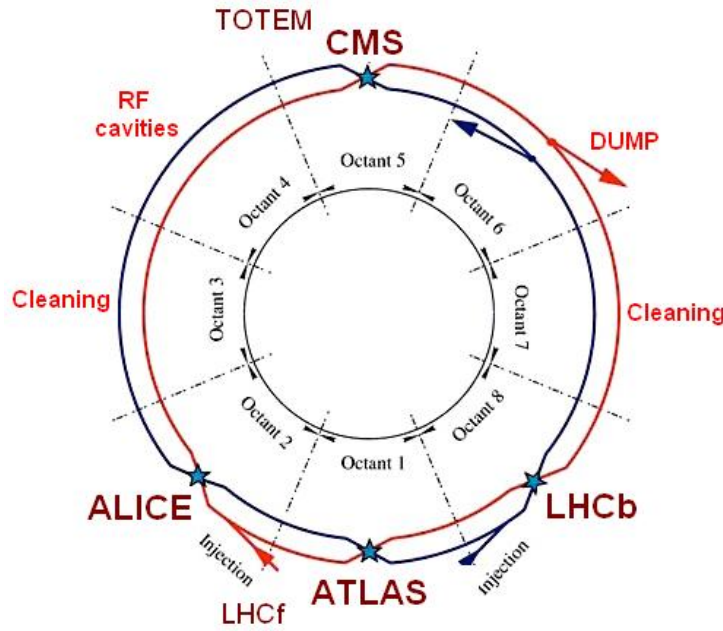


Figure 2.2: A schematic layout of the LHC tunnel.

2.1.2 Main components of the LHC

The design of LHC depends on very basic physics principles associated with latest technology. The three main components of LHC are:

1. **Radio Frequency (RF) cavities:** The RF cavities are located along the beam pipe and are used to provide acceleration to the proton beams using alternating electric field. There are two independent RF systems, one for each beam. A sketch of a RF cavity is shown in the Fig. 2.3 below.

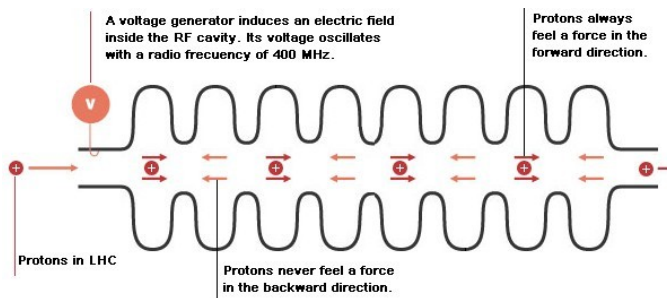


Figure 2.3: A Radio Frequency cavity.

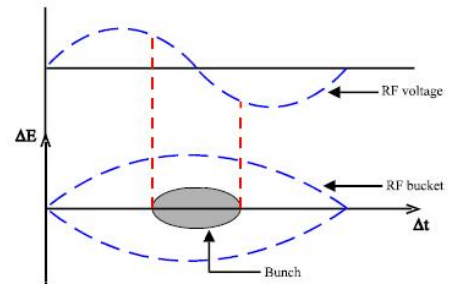


Figure 2.4: A sketch showing formation of a bunch.

The sinusoidal varying alternating electric field of RF cavity changes polarity after each half cycle which implies that it accelerates the protons during one half cycle and decelerates during the other half. This results into the formation of bunches of protons. As can be seen in Fig. 2.4, when a proton arrives inside

a RF cavity at a time when the voltage is zero, it will see no force and is called synchronous proton. Any proton arriving before the synchronous proton will see a negative polarity and hence will decelerate while any proton arriving later than the synchronous proton will see positive polarity and will get accelerated, so these protons will come closer to the synchronous proton and form a bunch. At LHC, the RF cavities resonate at a frequency of 400 MHz. The LHC is designed to have a total of 2808 bunches at a gap of 25 ns. The main goal of RF cavities is to maintain the tightness of proton bunches to provide high luminosity at collision points. These cavities also transfer the radio frequency power to the beam to provide acceleration. The LHC has eight RF cavities per beam operating at 4.5° K, each delivering a power of 2 MV at 400 MHz.

2. **Vacuum chamber:** The beam pipe in which proton beams travel are kept at a low pressure of $\sim 10^{-13}$ bar, much less than the outer space, forming an ultrahigh vacuum. This is required to minimize the interactions with parasitic particles and subsequent loss of accelerated protons.
3. **Magnet system:** In order to provide proper bending and focusing of the proton beams, an intense magnetic field varying in the range of 0.54 T (for 450 GeV proton energy) to 8.33 T (for 7 TeV proton energy) has been used. This high magnetic field has been achieved by the use of superconducting magnets made of copper plated Niobium-Titanium (NbTi) superconductors. The superconducting state has been attained by using many tones of superfluid Helium cooled to a temperature of 1.9° K. Due to the constraints of small tunnel diameter, it was not possible to make completely separate rings for both the proton beams. So an alternate way was adopted in the form of twin-bore magnet design (a front view is shown in Fig. 2.5) which comprises of two coils and beam pipes at a separation of 180 mm within the same mechanical framework. This magnet structure is constructed in such a way that it provides opposite magnetic fields in the two beam pipes. The copper plated NbTi filaments are joined together to make the cables and two layers of these cables are wrapped around each beam pipe. The current is made to flow in opposite directions through the cables around the two pipes, which produces an intense magnetic field of opposite polarity in both the beam pipes. The copper provides an insulation among the NbTi filaments in superconducting state and act as a conductor upon reaching the conducting state.

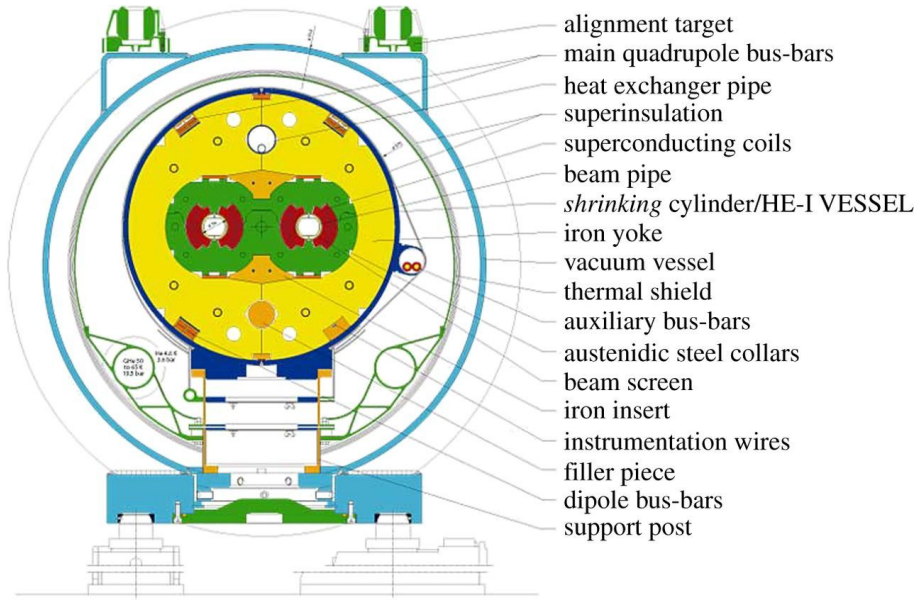


Figure 2.5: The LHC twin-bore magnet system

The LHC consists of different type of magnets that include, dipoles used for bending of the particle's beam, quadrupoles used for focusing of the beam, sextupoles and octupoles used for correcting the size and position of the beam. The LHC ring consists of 1232 dipoles, 392 quadrupoles and many sextupoles and octupoles. To enhance the probability of collisions, the bunches are squeezed by the use of special magnets positioned near the collision chambers. These magnets reduce the beam diameter to a minimum of $16 \mu\text{m}$.

2.1.3 Proton's journey to maximum acceleration

Making the proton beams travel closer to the speed of light is a very intricate task. At LHC, it has been accomplished in a number of steps as can be seen in the schematic representation of CERN accelerator complex [107] in Fig. 2.6. First of all, protons are extracted from hydrogen gas by stripping off the electrons using an electric discharge. Then, the protons are directed towards a Radio Frequency Quadrupole (RFQ) that provides an initial acceleration of 750 keV to the beam. After that, the protons are sent to LINAC2, a linear accelerator that makes use of RF cavities and accelerates the proton beam upto 50 MeV . The beam from LINAC2 is made to enter Proton Synchrotron Booster (PSB) which increases the energy of protons to 1.4 GeV and divides the proton beam into 6 bunches. The output of PSB is then injected into Proton Synchrotron (PS) that accelerates the protons upto 25 GeV . The PS is also responsible for preparing the LHC bunch

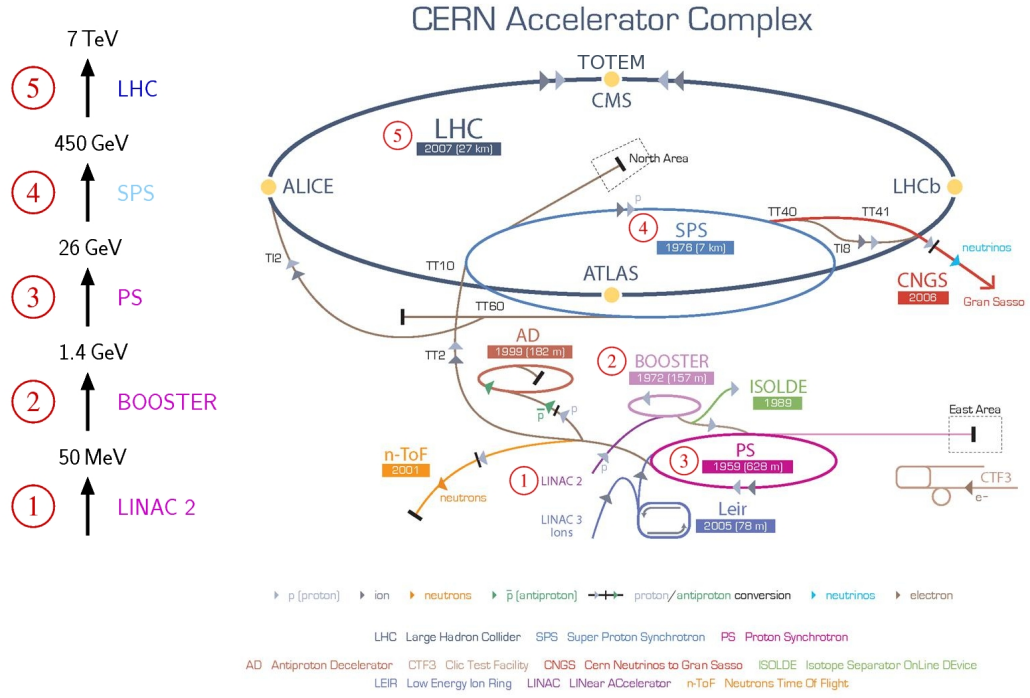


Figure 2.6: A schematic view of CERN accelerator complex representing the different accelerators deployed for providing acceleration to the proton beam.

structure. A total of 72 bunches with a spacing of 25 ns are formed in PS. The proton bunches from PS then enter into Super Proton Synchrotron (SPS) where they are further bunched into 288 bunches and are accelerated upto the LHC injection energy of 450 GeV. The protons from SPS are then finally made to enter into LHC where they accelerated upto the highest energy required.

2.1.4 Luminosity measurements at LHC

At LHC, the rate of collision N is expressed in terms of the instantaneous luminosity L as $N = \sigma L$ where σ is the cross section of the process under study. For a given process, the higher the luminosity, the chances are more of a collision to happen. The luminosity of a beam depends upon many machine parameters and is given by,

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \beta^* \epsilon_n} F, \quad (2.2)$$

where,

N_b : number of particles in a bunch,

n_b : number of bunches in each beam,

f_{rev} : beam revolution frequency,
 γ_r : relativistic gamma factor,
 ϵ_n : normalized transverse beam emittance¹,
 β^* : betatron function², and
 F : luminosity reduction factor due to the presence of crossing angle between the two beams.

In order to achieve highly luminous beams, one need to provide large number of highly populated bunches having low value of beam emittance and betatron function colliding at high frequencies. The various parameters related to LHC machine and proton beams are summarized in Table 2.1.

LHC Parameters	Design Value	2016 Value
Total number of magnets	9600	–
Number of dipole magnets	1232	–
Beam energy (TeV)	14	13
Bunch spacing (ns)	25	25
No. of protons per bunch	1.1×10^{11}	1.1×10^{11}
No. of bunches per beam	2808	2220
Beam crossing angle (μrad)	285	370
Beam emittance, ϵ_n (μm)	3.75	3.4
Betatron function, β^* (m)	0.55	0.4
Instantaneous luminosity ($\text{cm}^{-2}\text{sec}^{-1}$)	1.0×10^{34}	1.0×10^{34}

Table 2.1: Design values of the various parameters of LHC and a comparison with the values obtained during the 2016 run.

The sum of instantaneous luminosities obtained over a period of time is referred to as the integrated luminosity. The Fig. 2.7 shows the total integrated luminosity delivered by the LHC and recorded by CMS [108] over a period of time for the year 2016.

During the 13 TeV run in 2016, the LHC delivered a total of 40.8 fb^{-1} of pp collision data, out of which, CMS recorded a total of 37.8 fb^{-1} . The CMS has certified 35.9 fb^{-1} of recorded data to be used for physics analysis. The study presented in this thesis has been performed using this 35.9 fb^{-1} of data.

¹The beam emittance is defined as the transverse size of the beam. The narrower the beam, the lesser is the beam emittance and more are the chances of interaction.

²Betatron function is defined as the distance from the collision point at which the beam width in transverse plane is twice compared to the beam width at collision point.

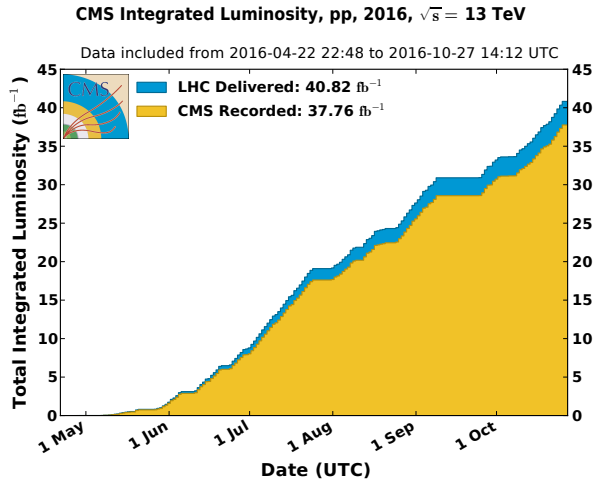


Figure 2.7: Total integrated luminosity delivered by the LHC and recorded by CMS as a function of time, during the year 2016 [108].

2.2 The CMS Detector

The CMS detector [109] is one of the main particle detectors designed at the LHC to observe different kind of particles and processes produced in the pp collisions. The prime objective of the CMS detector is to explore the physics at Terascale, the energy region at which the physicists believe the experiment has the potential to answer some of the existing key questions. The detector has a cylindrical onion like structure having a coverage of 4π steradian, with each layer representing a different sub-detector used to measure different type of particles. The information from different layers is used to build up the complete picture of an event. It is situated in a cavern, 100m underground, located at point 5 of the LHC tunnel in France near village Cessy. It has a length of 21.6 m and a diameter of 14.6 m, it weights around 14,000 tonnes. CMS is a large scientific collaboration involving around 3,000 people from 198 scientific institutions and universities belonging to 45 countries.

Experimental challenges: As discussed in last section, the LHC has a very high collision rate which is achieved by colliding bunches of more than 100 billion protons each, 40 million times every second, resulting into a pp interaction rate of around 10^9 per second. The high track density arising due to these interactions create severe radiation environment. The detector needs to remain unaffected from these radiations along with providing efficient particle identification. The detector has to face the challenge of adjusting between occupancy and granularity. Additionally, a

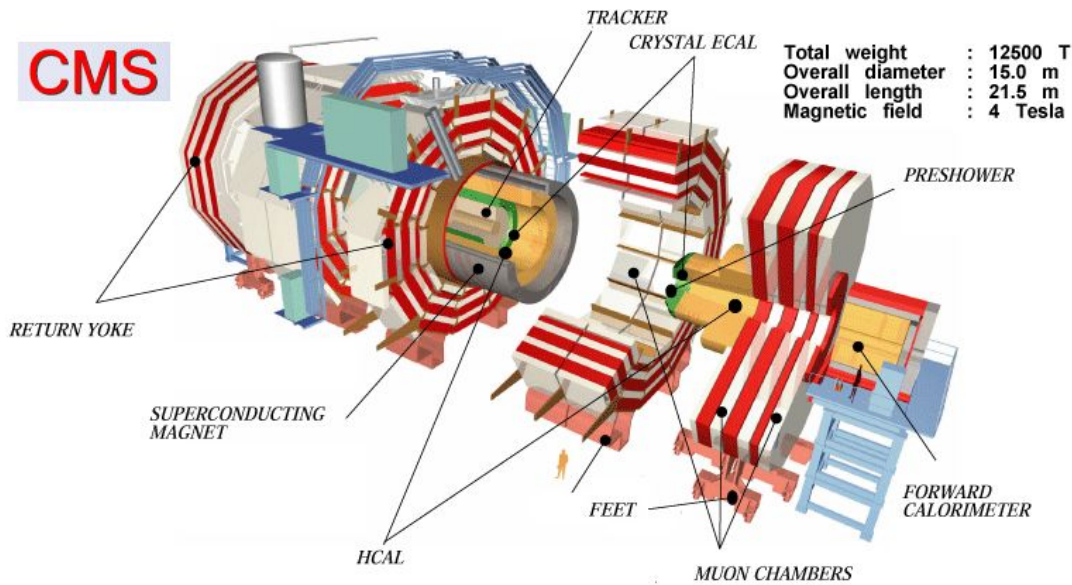


Figure 2.8: A layout of the CMS experiment at CERN.

majority of collisions contain uninteresting events, a fast filtering trigger system is required to reject these unwanted events. Also, the interesting pp events are superimposed by many uninteresting events occurring in the same bunch crossing (known as pile-up). Dealing with these events require high granularity and good time resolution detectors, which further demand to have large number of detector channels with good synchronization. Even storing the data after huge reduction require about 10 million gigabytes per year which in itself is a big challenge.

2.2.1 Layout of the CMS detector

A general layout of the CMS detector is presented in Fig. 2.8. It is based on four principle sub-systems: A high granularity central tracker, a homogeneous electromagnetic calorimeter, a sampling hadron calorimeter and a muon detector. The heart of the CMS is its very high superconducting magnet that is situated between the calorimeters and the muon detector. All the sub-systems of the detector are divided into one cylindrical barrel region and two opposite facing endcap regions, complimented with very forward regions. This arrangement provides a very good coverage of the collision fragments required for event building. A longitudinal view of the CMS detector system, showing the barrel and endcap sections of various sub-detectors is presented in Fig. 2.9, with the origin denoting the interaction point.

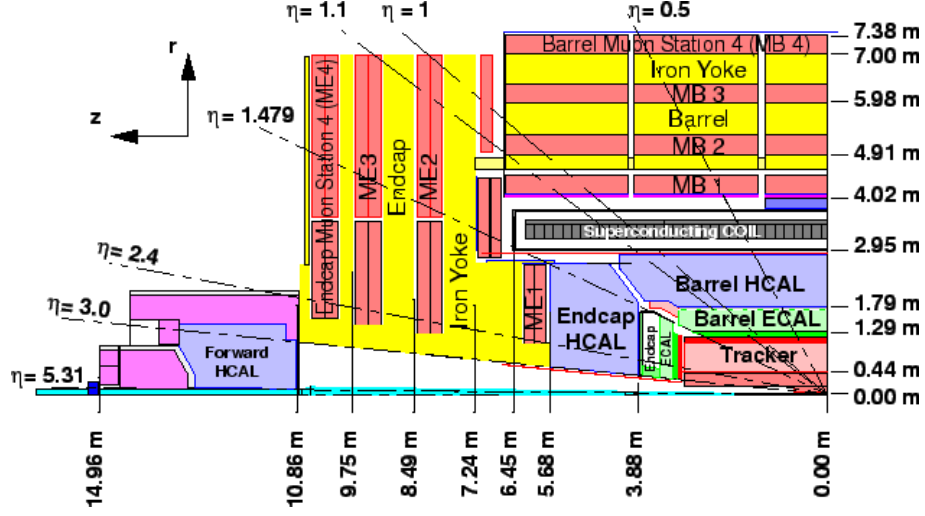


Figure 2.9: Longitudinal contour of a quadrant of the CMS detector.

Co-ordinate system: The CMS has adopted a right-handed spherical co-ordinate system with origin at the center of the detector which is also the intended collision point. The x -axis points towards the center of the LHC tunnel, the y -axis is vertically upwards and the z -axis is along the beam pipe. The polar angle θ is taken from the z -axis to the x - y plane and the azimuthal angle ϕ is taken from the x -axis to the y -axis in the x - y plane. The angle θ denotes the particle's angle w.r.t the beam axis while angle ϕ denotes the particle's orientation in the transverse plane. The co-ordinate r denotes the radial distance and is defined by $r = \sqrt{x^2 + y^2 + z^2}$. It is conventional to use pseudorapidity η instead of polar angle θ as seen in Fig. 2.9. The pseudorapidity is defined in terms of angle θ as

$$\eta = -\ln \left[\tan \frac{\theta}{2} \right]. \quad (2.3)$$

On going from $-z$ to $+z$ axis, θ goes from $-\pi$ to $+\pi$ and η goes from $-\infty$ to $+\infty$. In the non-relativistic limit, it is called rapidity and is defined as,

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \quad (2.4)$$

where E and p_z are the particle's energy and z -component of the momentum. The rapidity, y requires the measurement of both E and p_z and in relativistic regime, it is quite difficult to measure the z -component of momentum very precisely. The reason of choosing η over θ is that the difference between the pseudorapidity of two particles ($\Delta\eta$) is Lorentz invariant under boosts along the beam axis. This means that an observer in the lab frame, boosted along z -axis w.r.t the rest frame, will

measure the same $\Delta\eta$ between two particles as an observer in the rest frame. The importance of this feature comes from the fact that in the pp collisions, the hard interaction takes place between proton constituents (quarks or gluons), usually having different momenta, which means the different interactions have different boosts of their center-of-mass frames w.r.t the detector rest frame. Invariance of $\Delta\eta$ under boosts allows us to treat the different interactions in a similar way as the interaction happening in the center-of-mass rest frame. The various histograms binned on the angular separation of particles will remain undistorted by these arbitrary boosts of different events. The angular distance R between two particles in $\eta - \phi$ plane, as observed from the origin of CMS detector, is written in terms of the differences in η and ϕ as, $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

In colliders, two particle beams colliding head-on consist of only longitudinal component of momentum. So the particles in the beam do not have a transverse component of momentum before collision, thus by conservation of momentum, the total transverse momentum after the collision should also add up to zero. This feature is extensively used in collider experiments and it requires to consider the particle's energy and momentum in the transverse direction, known as transverse energy (E_T) and transverse momentum (p_T), defined as $E_T = E \sin \theta$ and $p_T = \sqrt{p_x^2 + p_y^2}$ respectively. Any imbalance in the transverse momentum is caused by either the invisible particles like neutrinos or by the energy lost in the unavoidable nuclear processes. This imbalance is known as missing transverse energy and is given by, $E_T^{\text{miss}} = -\sum p_T$.

A short description of all the components of the CMS detector from innermost to outermost has been provided as follows:

2.2.1.1 CMS tracking system

The tracker [109, 110] is the innermost sub-detector, installed close to the interaction point. It has been designed to build tracks of charged particles coming out of the collisions and reconstruct primary as well as secondary vertices, with utmost precision and accuracy.

The tracker has a total length of 5.8 m, a diameter of 2.5 m and covers a region upto $|\eta| < 2.5$. A uniform magnetic field of 3.8 T is present over its entire volume. A cross-sectional view of the CMS tracker has been presented in Fig. 2.10. The tracker is made up of a central barrel and endcap parts consisting of pixel and

strip layers. Both pixel and strip layers are made up of silicon sensors of different dimensions. The CMS tracker is the world's largest silicon tracker with over 200 m² of active silicon area. Whenever a charged particle passes through a silicon sensor, it results into generation of electron-hole pairs, which are measured in an outer circuit as a hit. Subsequent information of the large number of hits from different sensors leads to the reconstruction of the particle's track. Since the energy required to generate an electron-hole pair is small, the energy/momentum resolution of these sensors is very large. Also, since the electrons travel very fast to the outer circuit, the time resolution is better. The charged particles of different momentum curved differently under the effect of the magnetic field and the tracker records the particle's tracks and momentum by finding their positions and curvature at various key points.

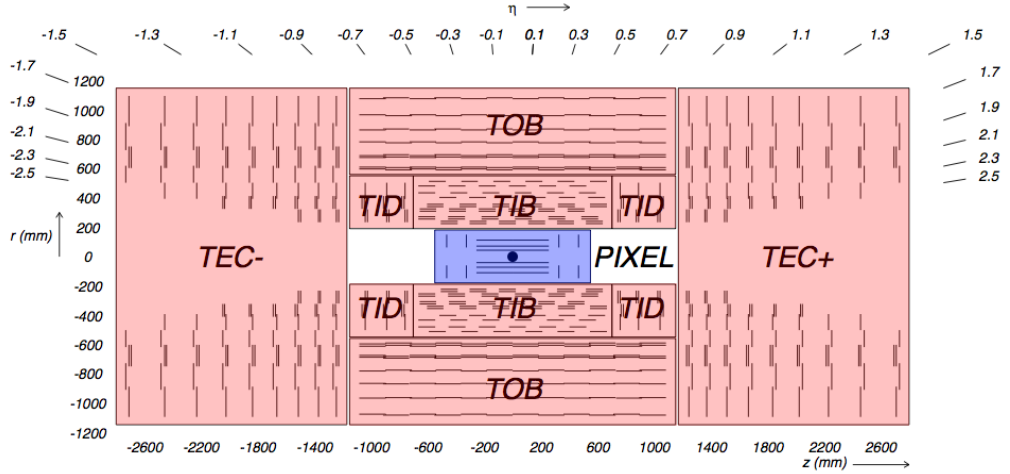


Figure 2.10: A cross-sectional view of the CMS tracker system in r - z plane, with each line representing a silicon detector module.

The tracker has to face around 1000 particles coming from each bunch crossing at every 25 ns which lead to high radiation doses in the detector. So it should be able to operate in such a harsh environment in addition to dealing with high occupancy rate as well as providing high granularity and quick response [111]. The silicon detector technology meets all these goals very efficiently. The number of tracker layers is a tradeoff among the tracking efficiency and material budget [112]. On one side, more layers leads to more hits and precise track reconstruction. While on other side, large amount of material inside the tracker leads to multiple scattering which in turn can spoil proper tracking too.

In order to reconstruct a track using hits, the simplest approach is to start

from a hit in the innermost layer and in radial direction, project a cone on the next layer. If the next layer contains no hit within the cone, leave the starting point considering it a noise or a particle of very low energy. Otherwise, keep on repeating the procedure until last layer is reached. A more advanced approach start the search from the outermost layer along with taking additional information from other sub-detectors. A description of pixel and strip layers of CMS tracker has been provided below.

2.2.1.1.1 Pixel detector

The high precision pixel detector [113] is the innermost detector of the CMS tracker. Since the particle flux near the collision point is very large, a small-scale pixel geometry is constructed to achieve unambiguous track reconstruction and precise vertex determination. The particles with relatively long lifetime, such as b and c quarks, arise from primary vertex but decay after travelling small distances, forming secondary vertices. The pixel detector is designed to distinguish these secondary vertices from the primary collision point.

The pixel detector consists of three concentric barrel layers and two endcap disks on both sides (shown in Fig. 2.11), covering a pseudorapidity region of $|\eta| < 2.5$. The barrel layers are 53 cm long and are present at the radii of 4.4, 7.3 and 10.2 cm from the beam line. The two endcap disks are extended in radius from 6 cm to 15 cm and are present at ± 34.5 and ± 46.5 cm from the collision point. The detector is equipped with a total of about 66 million pixels, each having a size of $100 \times 150 \mu m^2$. The occupancy rate of each pixel is of the order of 0.1% per LHC bunch crossing.

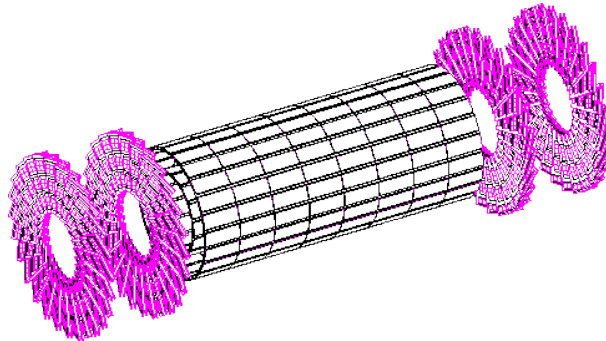


Figure 2.11: A schematic layout of the CMS pixel detector.

The pixel detector has performed exceedingly well during Run I, providing a

resolution of around $10\mu\text{m}$ in $r-\phi$ and $20\text{--}40\mu\text{m}$ in z direction along with an overall efficiency of more than 99%. The LHC Run II started in 2015 with an increased center of mass energy of 13 TeV and a lower bunch crossing frequency of 25 ns. During overall Run II phase, a peak instantaneous luminosity of $2 \times 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ was achieved. It also saw a pileup rate of more than 50. Under these conditions, the current pixel detector will suffer from increased fake rates and reduced resolution due to higher occupancy. So in order to maintain the performance of pixel detector, an upgrade known as “Phase1 Pixel Upgrade” [114, 115] was performed during the technical stop at the end of 2016. The upgraded pixel detector features 4 barrel layers and a turbine-like module arrangement for endcap disks (as shown in Fig. 2.12). Light-weight support structures and 2-phase CO_2 cooling has been introduced to keep the optimum material budget. A comparison of the two detector configurations using tracking efficiency and fake rate as a function of pileup are presented in Fig. 2.13.

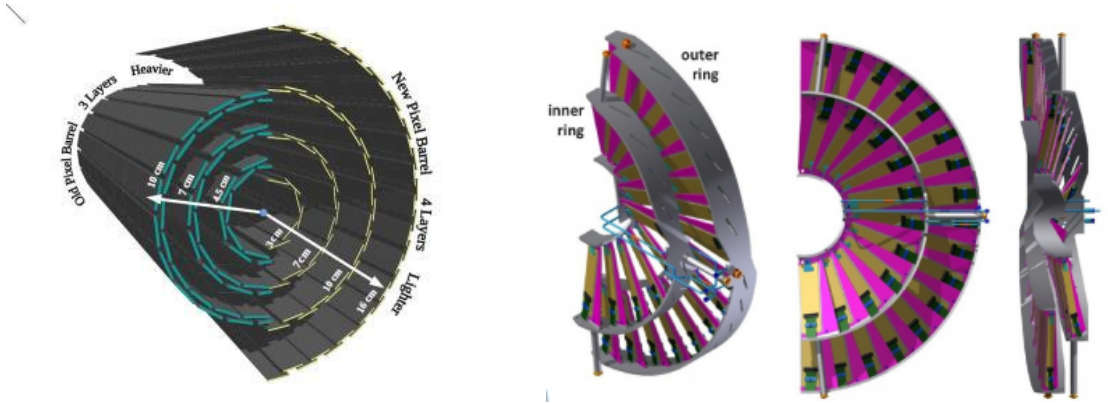


Figure 2.12: A schematic representation of the upgrades in pixel detector barrel (left) and endcaps (right) during Phase1 Pixel Upgrade.

2.2.1.1.2 Silicon strip detector

The silicon strip tracker [109, 110] is present outside the pixel detector and consists of more than 15000 single or double sided (shown in Fig. 2.10 by single and double lines) modules. Since the particle flux reduces in strip tracker, the dimensions of strip modules are kept larger than that of pixel detector modules. The size of modules also increases with distance to keep the occupancy level at around 1%. The modules are of rectangular shape in barrel region while that of trapezoidal shape in the endcap region. The strip tracker consists of four subsystems: The Tracker Inner Barrels (TIB), Tracker Inner Disks (TID), Tracker Outer Barrels (TOB) and Tracker EndCaps (TEC).

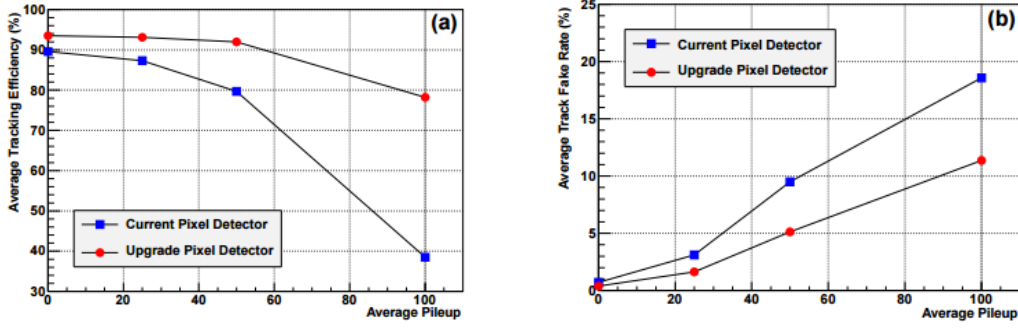


Figure 2.13: Performance measurement of the older pixel detector (blue) and the upgrade pixel detector (red) using $t\bar{t}$ events, showing tracking efficiency (left) and fake rate (right) as a function of per event pileup [115].

The TIB and TID are extended up to a radius of 55 cm and consists of 4 barrel + 3 endcap layers. The modules in TIB/TID are $320\mu\text{m}$ thick and delivers up to 4 measurements in $r - \phi$ direction, providing a position resolution of $23\mu\text{m}$ in the inner two layers and $35\mu\text{m}$ in the outer two layers. The outer barrel layer TOB is made up of 6 layers of $500\mu\text{m}$ thick sensors extended up to an outer radius of 116 cm which provides 6 more measurements in $r - \phi$ direction, corresponding to a resolution of $53\mu\text{m}$ in the initial four layers and $35\mu\text{m}$ in the last two layers. The TOB covers a z range up to ± 118 cm. Outside this range, $\pm\text{TEC}$ occupies the region from 124 – 282 cm in $\pm z$ and 22.5 – 113.5 cm in $\pm r$. Each of the TEC is made up of 9 disks, thereby providing 9 measurements in $r - \phi$ per trajectory.

The transverse momentum resolution of the tracker varies in the range from 0.7% at 1 GeV to 1.5% at 100 GeV.

2.2.1.2 Calorimeter

A calorimeter is a device that is used to measure energies of different type of particles. The incident particle interacts with the material of the calorimeter and produces a cascade (shower) of secondary particles of progressively lower energies, until the energies fall below a critical value which is then dissipated in ionization or excitation of the atoms of the material. The subsequent energy released by the atoms is captured by the photo-detector which through photoelectric effect, yield an electronic signal that provides a measurement of the particle's energy. In case of the sampling calorimeters, the medium that activate the particle shower is known as the passive medium (absorber) while the one that absorb the shower

particles and lead to the energy measurement, is known as the active medium (scintillator). Calorimeters have the capability to stop almost all particles except muons and neutrinos.

Calorimeters are broadly divided into two categories: Hadronic and Electromagnetic, depending on the particle type. They can also be classified on the basis of their composition into sampling and homogeneous calorimeters. A sampling calorimeter is the one that consists of alternate layers of two materials acting as passive and active mediums while a homogeneous calorimeter is built up of only one type of material that accomplishes both the tasks. An advantage of sampling configuration is that both the materials used are well-suited for their tasks. A very high density material is used as passive medium which results in the quick evolution of the shower into a limited space, so sampling calorimeters can be made compact as compared to homogeneous calorimeters. However, it has a disadvantage that part of the shower energy gets deposited in the passive medium and left unmeasured.

The CMS consists of Electromagnetic calorimeter (ECAL) just outside the tracker, followed by Hadronic calorimeter (HCAL), both contained inside the solenoidal magnet. A short description of each one is considered below.

2.2.1.2.1 Electromagnetic calorimeter

The CMS ECAL measures the energy of electromagnetically interacting particles, primarily photons and electrons. The photons interact with the material of the detector mainly through pair production for energies above 1 MeV and through Compton scattering and photoelectric effect for energies below 1 MeV. On the other hand, electrons mainly interact via bremsstrahlung for energies above 10 MeV and through ionization and excitation for energies below it. As a consequence, high energy electrons and photons, inside ECAL, produce a number of secondary particles, thereby generating electromagnetic (EM) showers. The EM showers are characterized by a material dependent parameter, the radiation length denoted as X_0 , which is the average distance x travelled by an electron before losing an energy equivalent to $1/e$ times its original energy i.e. $E = E_0 e^{-\frac{x}{X_0}}$. It is also equivalent to $7/9$ times the free path travelled by photon before converting into an e^+e^- pair.

As a result of particle multiplication and faster interaction, the amount of material required to contain an EM shower is rather small. Therefore, the

ECAL [109, 116] (shown in Fig. 2.14) is constructed as a homogeneous calorimeter made up of lead tungstate (PbWO_4) scintillating crystals. It is a highly dense material (8.28 g/cm^3) with a radiation length of 0.89 cm and Molière radius of 2.2 cm , allowing a design of fast, granular and radiation resistant calorimeter. The crystals are designed to provide very good energy resolution of electrons and photons required for CMS physics program. These are arranged in carbon fibre matrices to provide optical isolation.

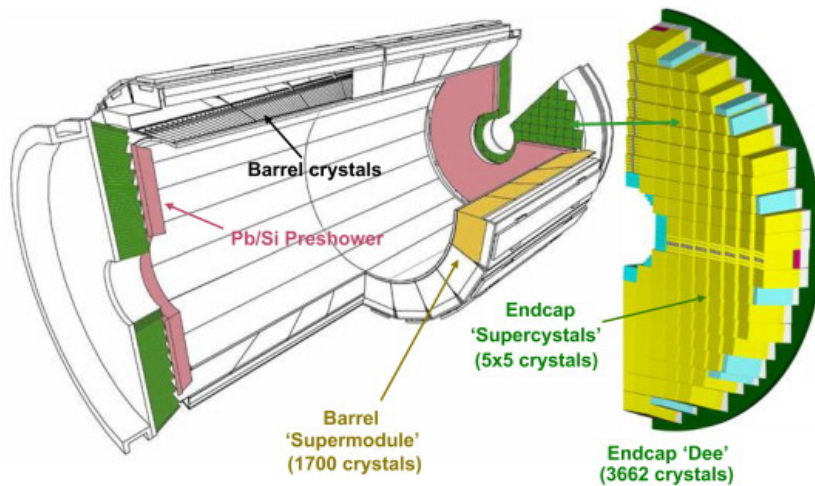


Figure 2.14: A layout of the CMS Electromagnetic calorimeter displaying the arrangement of crystals in barrel and endcaps with the preshower present in front.

The ECAL barrel (EB) is extended upto $|\eta| < 1.479$ and consists of 61,200 crystals providing 360-fold granularity in ϕ and (2×85) -fold granularity in η direction. The crystals are of trapezoidal shape having a front cross section of $22 \times 22 \text{ mm}^2$, a rear cross section of $26 \times 26 \text{ mm}^2$ and a length of 230 mm , equivalent to $25.8 X_0$ implying that it is able to contain 99.999...% of e/γ energy. These crystals are assembled into 36 supermodules, each weighing around 3 tonnes and containing 1700 crystals. Each supermodule covers 20° in ϕ and half barrel in η . The avalanche photodiodes (APDs) are used as photo-detectors to capture the scintillating light generated in the crystals upon interaction. These are made up of semiconducting silicon and have an active area of $5 \times 5 \text{ mm}^2$. The relatively low light yield of PbWO_4 crystals is compensated by the signal amplification by APDs through avalanche process. In addition, insensitivity of APDs towards high magnetic field and high radiation environment, makes them convenient for EB.

The flat circular ECAL endcaps (EE) cover the barrel at both the ends in the pseudorapidity range $1.479 < |\eta| < 3.0$, each consisting of 7,324 crystals arranged

in 5×5 crystal units, known as supercrystals. The trapezoidal crystals have a front face of $28.62 \times 28.62 \text{ mm}^2$, a rear face of $30 \times 30 \text{ mm}^2$ and a length of 220 mm ($24.7X_0$). Endcaps use vacuum phototriodes (VPTs) as photo-detectors since these can work in high radiation environment compared to APDs. Each VPT has a diameter of 25 mm and an active area of 280 mm^2 . In front of each ECAL endcap is present a high granularity sampling calorimeter, known as ECAL Preshower (ES), made up of alternating layers of lead (passive medium) and silicon strips (active medium). It covers a pseudorapidity range of $1.653 < |\eta| < 2.6$ and forms a disc of 2.5 m circumference. The disc is about 20 cm thick and has a 50 cm diameter hole in the middle for the passage of beam pipe. The principle aim of ES is to identify and stop overlapping photons coming from the decay of π^0 so as to prevent false signals. The high density of lead stimulate quick showering, thereby resulting into the shower containment in a very small volume. Around 95% of incident photons start showering in the first lead plane.

Many contributions deteriorate the actual energy resolution of any realistic calorimeter. The energy resolution of ECAL (for energies below 500 GeV) can be written in general as:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \quad (2.5)$$

where $\oplus$ corresponds to quadratic sum which implies that the contributions of different terms are independent from each other. The first term S is known as stochastic or sampling term. It considers the statistical fluctuations due to the shower fluctuations, lateral shower leakage, photo-detector inefficiencies, efficiency loss due to the shower containment in the passive medium of preshower etc. The second term N corresponds to the noise term arising due to instrumental effects, light-attenuation, pileup effects etc. The last term C is a constant term and it includes effects from imperfections in calorimeter construction, inter-calibration errors, longitudinal shower leakage, fluctuation in energy deposition in the dead areas of calorimeter etc.

The test beam studies [117] were performed using a block of 3×3 crystals and various terms involved in energy resolution were computed. These terms for barrel region are: $S = 2.8\%$, $N = 124 \text{ MeV}$ and $C = 0.3\%$ while for endcap region are: $S = 5\%$, $N = 500 \text{ MeV}$ and $C = 0.3\%$.

2.2.1.2.2 Hadronic calorimeter

The CMS HCAL has been designed to measure energy of particles (hadrons) interacting via strong force. Hadrons interact with the material of the detector through strong nuclear interactions, resulting into the production of a large number of secondary hadrons, thereby activating a hadronic shower. A majority of these shower particles, around 90%, are pions. Among these, the neutral pions (π^0) decay electromagnetically into two photons, thereby generating a component of EM shower within a hadronic shower. The fraction of energy of a hadron taken off by the EM component is denoted as f_{EM} and it varies from event-to-event. This fraction depends on the energy of initial hadron and varies from 30% for a 10 GeV hadron to 50% for a 100 GeV hadron. Also, a significant fraction of shower particles lose their energy as nuclear binding energy in nuclear reactions. This energy does not appear as a signal in the calorimeter and therefore, dissipate as invisible energy. As a consequence, the calorimeter signal for a hadron is usually smaller than an electron having same energy, a phenomenon known as non-compensation, also denoted as e/h ratio. These two features of hadron showers result into a non-linearity in the energy measurement of hadrons by HCAL.

The shower profiles of hadrons are characterized by nuclear interaction length, λ_{int} (expressed in g/cm^2) which is the mean distance traversed by a hadron before undergoing a nuclear interaction. Usually, the interaction length of a hadron is much larger than the radiation length of an electron of same energy, e.g. in copper, $X_0 = 1.4\text{ cm}$ while $\lambda_{\text{int}} = 15\text{ cm}$. This implies that a large amount of material is required to contain a hadron of a particular energy compared to an electron of same energy, not possible to achieve with homogeneous calorimeter. Therefore, the hadron calorimeter always use sampling configuration. A properly designed sampling detector achieves compensation of hadronic signals by the use of various techniques and sampling fractions [118].

The HCAL [119] is a sampling calorimeter constructed of a dense absorber material made up of brass/steel plates interleaved with plastic scintillator tiles, which are read out via embedded wavelength-shifting (WLS) fibres. This configuration is chosen for both barrel and endcap regions of HCAL due to the non-magnetic behaviour, high absorption length and good mechanical strength of brass while long term stability and radiation hardness of plastic scintillators. The detector is assembled in such a way that it does not have any cracks or dead areas in

ϕ . The scintillator tiles emit blue-violet light upon particle interaction, which is then converted into green light by WLS fibres. These optical signals are then converted into electronic signals by the use of Hybrid Photodiodes (HPDs). The HCAL constitutes of four sections: HCAL Barrel (HB) [120], HCAL Endcaps (HE) [121], HCAL Forward (HF) [122], and HCAL Outer (HO) [123]. The HCAL, in association with ECAL, provides the energy/direction measurements of hadrons. A schematic view of a section of HCAL is presented in Fig. 2.15.

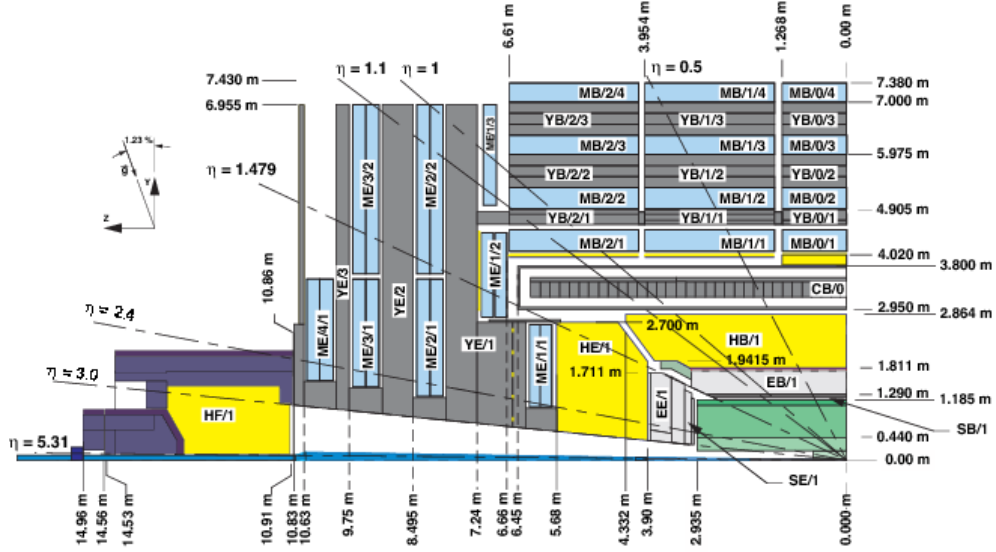


Figure 2.15: A schematic view of a quadrant of the CMS Hadron calorimeter presenting different sub-detectors (in yellow color).

The HCAL barrel covers a pseudorapidity range of $|\eta| < 1.3$ and is made up of two half barrels, each made of 18 azimuthal wedges, each covering 20° in ϕ . Each wedge is sliced into 16 absorber plates, with front and back plates made of steel and rest made of brass (a Cu-Zn alloy). The plastic scintillating tiles with WLS fibres are arranged into trays containing many tiles and are inserted between each plate to construct a sampling configuration. Each wedge is further partitioned into 4ϕ sectors while the plastic scintillators are further partitioned into 16η sectors, resulting into a segmentation of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The HB is situated between EB (at $r = 1.77$ m) and superconducting magnet (at $r = 2.95$ m), resulting into highly constrained material amount. The total thickness of HB corresponds to $5.82\lambda_{\text{int}}$ at $\eta = 0$ and $10.6\lambda_{\text{int}}$ at $\eta = 1.3$, ECAL also adds a thickness of $1.1\lambda_{\text{int}}$ to HB material. In order to ensure the proper energy containment of hadronic showers, a complementary detector layer of same configuration known as HO, is positioned outside the magnet coil which acts as a tail-catcher for any late starting

showers. It provides an additional thickness of $1-2\lambda_{\text{int}}$ to HB showers. The high pseudorapidity region, $1.3 < |\eta| < 3.0$ has been instrumented using HE on both sides of HB. The endcaps are made of 17 brass plate layers with plastic scintillator tiles accommodated in between. Including with EE, it corresponds to a total thickness of around $10\lambda_{\text{int}}$.

The high radiation environment ($3.0 < |\eta| < 5.0$) near the beam pipe is armed with a more robust hadron forward (HF) calorimeter made up of quartz fibres as active medium and steel as the absorber medium. The front face of HF is located at a distance of 11.2 m from the interaction point. The structure is divided into 36 azimuthal wedges of steel, 18 wedges on each side of collision point and composed of 5 mm thick scintillator plates along with WLS fibres. A signal in the form of Cherenkov radiation is produced in HF when a charged particle from the shower enters the active medium (quartz) with a velocity greater than that of light in the medium. The HF has been utilized to improve the $E_{\text{T}}^{\text{miss}}$ measurements and identify/reconstruct very forward jets which can contain distinguishing features of many physics processes. The HF is also used in the measurement of online luminosity of CMS.

The energy resolution of CMS HCal is parametrized as: $\frac{\sigma}{E} = \frac{A}{\sqrt{E}} \oplus B$, with the parameters, $A = 0.847 \text{ GeV}$ and $B = 0.074$ for HB and HE and $A = 1.98 \text{ GeV}$ and $B = 0.09$ for HF.

2.2.1.3 Solenoidal magnet

The CMS uses a large solenoid magnet [124] as its central device around which the detector has been constructed. In order to achieve precise measurements and resolutions of energy/momenta of charged particles coming from the collisions, CMS has deployed a large magnetic field of 3.8 T. A charged particle in a uniform magnetic field $\vec{B}$ experience a force equal to $F = qvB \sin \theta$, where q is the particle's charge, $\vec{v}$ is the particle's velocity and θ is the angle between $\vec{v}$ and $\vec{B}$. In this magnetic field, the particle will traverse a curve of radius $\frac{mv}{qB \sin \theta}$, m being the mass of particle, which implies that the particles of different initial velocities or momenta will be curved differently inside field $\vec{B}$. Therefore, measurement of the angle and radius of curvature of charged particles inside $\vec{B}$ can be used to get an estimate of particle's momentum. Also, the momentum resolution depends on the magnitude of magnetic field $\vec{B}$ and length of solenoid L as $\frac{\Delta p}{p} = \frac{p}{BL^2}$, stating that a large

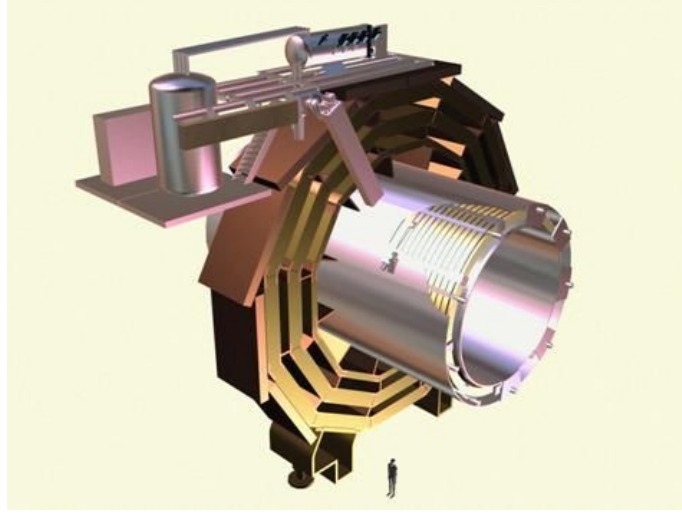


Figure 2.16: A schematic view of the solenoidal magnet used in the CMS detector, showing the coil and the central return yoke.

resolution (less Δp) can be achieved by the use of a high magnetic field of large size.

The CMS magnet as shown in Fig. 2.16 is a solenoidal magnet made of superconducting Nb-Ti coils which produce an intense magnetic field of 3.8 T when a current of 19,500 Ampere is passed through them. The magnet with a length of 13 m and an internal diameter of 5.9 m fit snugly the tracker and calorimeters inside. It consists mainly of three parts: the coil, vacuum tank and yoke. The vacuum tank acts as a cryostat and provide the essential cooling to solenoid using liquid He. The return yoke is made up of iron and is present outside the magnet coil reaching an outer diameter of 14 m. It is responsible for return of magnetic flux with a field of 2 T. The yoke is a 12-sided structure made up of 11 discs, 5 in barrel and 6 in endcaps, the muon chambers are interleaved in this. The total energy stored inside the magnet amounts to around 2.6 GJ.

2.2.1.4 Muon detector

The CMS has been specifically designed for precise identification, momentum resolution and triggering of muons within a range $|\eta| < 2.4$ and $p_T \leq 1$ TeV. Muons, though interact electromagnetically, do not lose much of their energy via bremsstrahlung owing to their high mass as compared to electrons. They interact primarily through ionization losing an energy on the average of 1-2 MeV/g/cm². Due to their less interacting nature, they penetrate the material of ECAL and HCAL, so the muon chambers for their detection are installed at the very edge of

the detector.

The CMS muon [125] system has been divided into four stations interleaved with the return yoke and having a magnetic field of 2 T. The barrel muon detector is also longitudinally divided into five wheels labelled as YB0, YB $\pm$ 1, YB $\pm$ 2, with each wheel being further divided in the azimuth direction into 12 sectors of 30° each. It uses three different kind of detectors to observe muons: Drift Tubes (DTs), present in the barrel region upto $|\eta| < 1.2$; Cathode Strip Chambers (CSCs), present in the endcap region for $0.9 < |\eta| < 2.4$; and Resistive Plate Chambers (RPCs), present in barrel as well as endcaps. Whenever a particle passes through these chambers, it leaves its signature in the form of hits. Tracking the particle's position in multiple layers and combining the information from tracker, leads to a precise determination of the position and momentum of the particle. The DTs/CSCs are good in position measurements while the RPCs provide fast response, so a combination of the three result into a robust muon detector. Diagrams representing the mechanical layout of CMS muon detector and the trajectory of muon through muon stations is shown in Fig. 2.17.

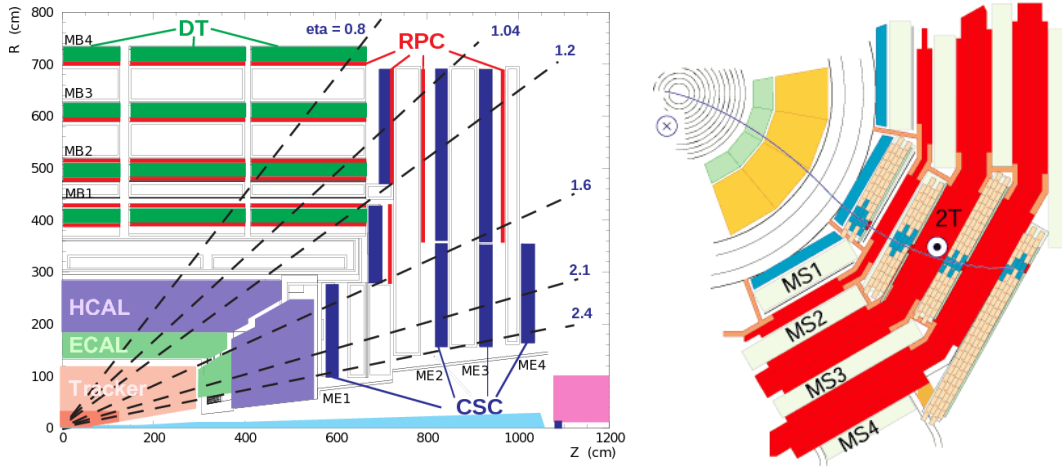


Figure 2.17: A schematic view of the solenoidal magnet used in CMS, showing the coil and the central return yoke. The trajectory of a muon through the muon stations is also presented.

The drift tubes are organized into 4 stations labelled as MB1, MB2, MB3 and MB4, which forms concentric cylinders around the beam pipe. Each of the inner three stations are composed of 12 drift chambers, with each chamber containing many drift tubes. These chambers are further grouped into three superlayers denoted as SL1, SL2 and SL3, each consisting of four chambers. SL1 and SL3 provides the co-ordinates of incident muons in $r - \phi$ plane while SL2 measures the

co-ordinates in $r - z$ plane. The outermost station consists of two superlayers only. The choice of drift chambers is made in barrel due to the low particle rate and uniform magnetic field. Each drift tube contains a anode wire of $50\mu m$ diameter and two electrode plates to create electric field. The tube walls are grounded and act as cathodes. The drift tube is filled with a gas containing a mixture of Ar (85%) and CO_2 (15%). The voltage difference between wire and walls is about 1.8kV. When a muon pass through the gas volume, it knocks off electrons which travel to the anode wire, carrying along the information of incident particle. The DTs provide a signal gain of 10^5 and a drift time of 380 ns.

The CSCs have been deployed in the high radiation and uneven magnetic field environment in the endcaps. Owing to their quick response time, high granularity and radiation hardness, the CSCs are well qualified to identify muons in the range $0.9 < |\eta| < 2.4$, overlapping with DTs in $0.9 < |\eta| < 1.2$. The CSC chambers are of trapezoidal shape placed perpendicular to the beam line and cover 10° - 20° in ϕ . These are organized in 4 stations per endcap, labelled as ME $\pm$ 1, ME $\pm$ 2, ME $\pm$ 3 and ME $\pm$ 4, each station containing a number of CSCs. Each CSC consists of an array of positive anode wires made of tungsten held perpendicular to the negative cathode strips made of copper, in a gas mixture of Ar (40%), CO_2 (50%) and CF_4 (10%). When a muon enters a CSC, it knocks off electrons from the gas atoms, which then flock towards the anode wires resulting into an electron avalanche. The positive gas ions move towards cathode strips and induce a charge pulse. Since the anode wires and cathode strips are present in perpendicular direction, these provide two position co-ordinates corresponding to each particle. This arrangement in CSCs also results into fast response making them convenient for triggering. Each CSC is composed of six anode wires interleaved with seven cathode strips, providing the measurements in $r - \phi$ plane.

The RPCs are parallel plate gaseous detectors with fast response time and adequate position resolution making them suitable for muon triggering. The RPCs are installed both in barrel (480 chambers) and endcaps (432 chambers) with two chambers of rectangular shape per DT chamber in barrel and one chamber of trapezoidal shape per CSC chamber in endcaps. The RPCs have the ability to tag a particle in a time frame much lesser than the bunch crossing time, making them ideal for triggering. The RPC is composed of two gaps, each operating in avalanche mode and read-out strips are inserted in-between. The total signal induced by the

RPC is sum of signal induced in each gap. However, in order to ensure proper stability and operation of the RPC, an intensive control of humidity, temperature and pressure is required.

The muon system of CMS is made up of around 25000 m² area and more than a million readout channels. Considering only the muon system without any information from tracker (stand-alone muon), a momentum resolution of around 9% is obtained for a muon of $p_T \leq 200$ GeV in low η region. For a high p_T muon (~ 1 TeV), the resolution varies from 15-40% for different η . Adding tracker information (global muon) results into an improvement in momentum resolution of around 1% for low p_T muons and around 5% for high p_T muons.

2.2.1.5 CMS upgrade during Long Shutdown 1

During LS1 in 2013-2015, minor repair and maintenance work were performed on most of the CMS sub-systems, except for some major improvements that include: Installation of new pixel tracker layer (explained in Section 2.2.1.1) which was actually done after LS1 and during the technical stop at the end of 2016, installation of new photo-detectors to improve signal-to-noise ratio in HCAL outer barrel, installation of the fourth RPC disk in the endcap region of the muon detector which was actually the part of the first CMS technical design report and installation of around 72 new chambers in CSC along with 468 old chambers. The electronics have also been updated to handle high collision rates.

2.2.2 CMS Trigger and Data Acquisition System

Most of the LHC collision data consists of “soft” particles which do not participate in any interesting process. Also, at each bunch crossing, the collisions amount to a raw data of approximately 1 MB, which at the crossing rate of 40 MHz (25 ns), results into around 40 TB data per second. It is almost impossible to store and process such a large amount of data. Therefore, to reduce the data rate while keeping the potentially interesting events for further analysis, a hardware and software based automated system, referred to as the CMS trigger system [126] has been designed together with a Data Acquisition (DAQ) system [127], as shown in figure Fig. 2.18. The trigger system consists of two layers: a hardware based trigger known as Level-1 (L1) trigger and a software based trigger known as the High Level Trigger (HLT), reducing the event rate to an average of a few 100 Hz

to a maximum of 1 kHz. The data passing through both the trigger levels is stored for offline physics analysis. In order to obtain the required data rate, sometimes the triggers are prescaled at L1 or at HLT level, thereby storing only a fraction of selected events. DAQ acts as a bridge between the two trigger layers, it gathers data from first trigger layer, digitizes and build it and pass it to second trigger layer.

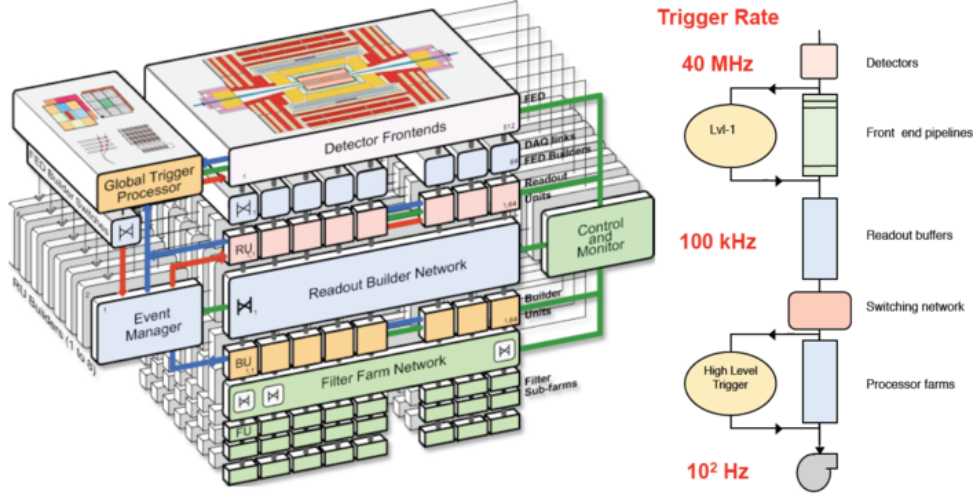


Figure 2.18: CMS Data Acquisition and trigger system.

2.2.2.1 Level-1 trigger

The Level-1 trigger of CMS is a set of hardware based triggers that rely on custom electronics. It has been designed to produce an output data rate limit of 100 kHz. The L1 triggers are implemented as custom hardware processors and considers only low resolution, coarsely segmented data from ECAL, HCAL and muon systems while keeping the high resolution data in the pipelined memories present in the front-end electronics. The L1 stores data in pipeline for $3.2\mu s$ during which it transmit its decision to detector electronics. The L1 trigger consists of a number of hardware components including field-programmable gate-array (FPGA) technology, ASICs and programmable lookup tables (LUTs). A block diagram representing the architecture of L1 trigger system is shown in Fig. 2.19.

The L1 trigger consists of local, regional and global components. At the top are the local triggers, also known as Trigger Primitive Generators (TPGs), which takes the information from the energy deposits in calorimeter towers and hit patterns in muon chambers. The information from TPGs are then passed on to

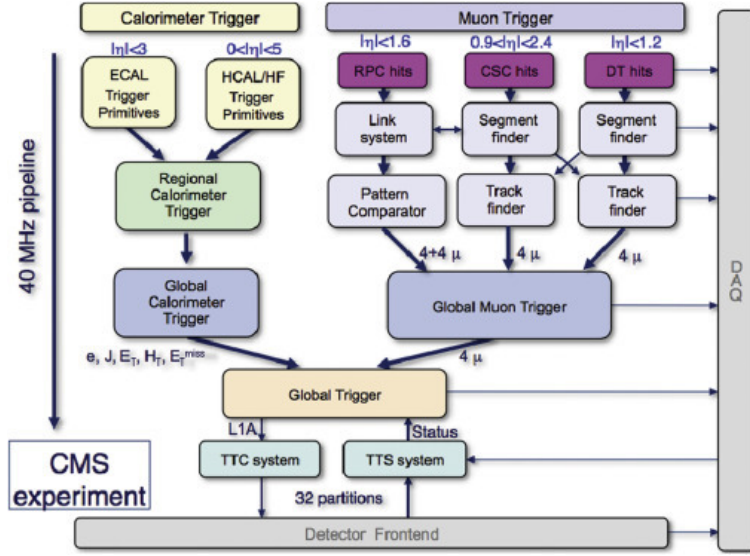


Figure 2.19: An architecture of CMS Level-1 trigger system.

the regional triggers. The regional calorimeter triggers (RCTs) makes use of this information and forms sorted and ranked trigger objects like electrons, photons and jets. The regional muon triggers (RMTs) join the segments and complete the tracks along with assigning physical parameters to them. The regional triggers passes these objects along with various information to the Global Calorimeter Triggers (GCTs) and Global Muon Triggers (GMTs) which sort out the highest ranked calorimeter and muon candidates along with determining various physics parameters like E_T , E_T^{miss} , number of jets, H_T etc. This information from GCTs and GMTs are then passed on to the Global triggers (GTs), the top level entity of L1 hierarchy, which makes the final decision of keeping or rejecting the event by applying the programmable topological cuts and energy thresholds on the objects and issues the final Level-1 Accept (L1A) decision. The L1A is then passed on to the Trigger Timing and Control (TTC) system for transferring it further to the detector front-end, from where it goes into the DAQ system. The TTC also interfaces with the LHC machine to provide clock and orbit information. The GT sends a command by the use of TTC in order to keep all the detectors and their electronics well in synchronization.

2.2.2.2 High level trigger

The CMS high level trigger accepts the output of L1 trigger and process it further by using a software (written mostly in C++) based trigger system implemented in a filter farm consisting of more than thousand commercial processors. All major

aspects of offline reconstruction code are included in the filter farm software and is designed to reduce data rate to a minimum of few 100 Hz. The average time taken in HLT processing is around 100 ms. It employs the use of high resolution data and utilize many sophisticated algorithms to identify interesting events for further storage. The HLT is based on highly flexible programming environment such that the changes in the algorithms can be easily done to improve the various selections and to deal with unexpected experimental conditions. A set of algorithms that reconstruct the physics candidates (known as producers) and apply various selections to the reconstructed objects (known as filters) is known as a “trigger path”; a number of trigger paths together constitute a “HLT trigger menu”. Within a trigger path, the various algorithms are arranged in an increasing order of complexity, with the most complex algorithms like the b-tagging, are executed in the end. Events that pass any of the trigger paths within a trigger menu are considered for further analysis and are sent to storage manager where these are stored on disks and are sent to CMS Tier-0 system present at CERN for offline processing and permanent storage.

2.2.2.3 Data acquisition system

The CMS DAQ mainly work as a carrier that gathers the data stored in detector’s front-end modules at the Level-1 accept rate and deliver it to HLT filter farms for further processing. It operates at an input data rate of 100 kHz, the output rate of L1 trigger. The complete processing of DAQ is summarized as follows: First of all, the front end systems (FESs) store data continuously in 40 MHz pipelined buffers. As the synchronous L1 trigger comes, data are extracted from FES via TTC and pushed into front end drivers (FEDs). The FEDs digitize, convert and zero suppress the data and then transfer it to front end read out links (FRLs) that merge data from different FEDs. These FRLs are known as event builders as these built a complete event by assembling the event fragments from all FEDs belonging to same L1. The data from FRLs are then transmitted into filter units (FUs). The HLT is mounted over FUs and use its computing power to perform its operations. These FESs, FEDs, FRLs, FUs all forms the different parts of DAQ system. A Trigger-Throttling System (TTS) has also been designed to protect DAQ against any back-pressures.

2.2.3 CMS data management

The raw data after passing the various triggering stages, is managed further by the use of the LHC computing grid [128]. The grid is constituted of four levels/tiers, referred to as Tier-0 (T0), Tier-1 (T1), Tier-2 (T2) and Tier-3 (T3). Each tier consists of a number of computer facilities and provides a set of dedicated services. The Tier-0 is the data center present at CERN, it contributes less than 20% of total grid computing capacity but it is responsible for maintaining a full copy of raw data, for performing calibration and prompt reconstruction (prompt-reco) and for distributing the raw and prompt-reco data to Tier-1 centers. Tier-1 is a group of around 13 computer centers located in various part of the world and joined together via optical-fibre links operating at a speed of 10 gigabytes/sec. The T1 centers together maintain a second copy of raw data, produce various Monte Carlo (MC) samples³ and re-reconstruct the raw data along with distributing it to various Tier-2 centers. The T2 centers are the sites where physicists can log-in and perform their analysis within the CMSSW environment. These centers also produce the MC samples and share them with T1 centers for storage and distribution to other sites. There are around 155 T2 sites distributed all over the world. The Tier-3 consists of local computing resources at various universities and institutes which individual users can access from any place around the world using grid, and can use for storage and data analysis.

The data present at various tiers are managed by the use of three data management tools [129]: Physics Experiment Data Export (PhEDEx) [130,131], Data Bookkeeping Service (DBS) [132] and Data Aggregation Service (DAS) [133]. The PhEDEx service provides data information access through the central PhEDEx database, perform operations such as requesting data transfer or data deletion and transport the data among the different CMS sites along with keeping a track. The DBS service keeps a metadata catalog of all the CMS data used by physicists as well as the various production and analysis systems. The DAS service provides the user an overview of all the available datasets and keep a mapping between the datasets and corresponding file blocks.

³simulated data samples, more information provided in Chapter 3

2.2.4 CMS software and computing

The CMS physics analyst perform various analyses using a collection of softwares based on C++ and python, referred to as the “CMS Software frameWork” (CMSSW) [104], installed at various T2 and T3 centers. A number of services including simulations, calibration corrections or reconstruction etc. are supported by this framework. It consists of one main executable, known as cmsRun, along with many other plug-in modules, containing all the required codes for event processing. The cmsRun operates over a python configuration file that contains all the important information regarding the data and modules. The CMSSW software has been built over the concept of an “Event”. An Event contains the information of a particular collision in different data formats, a physics dataset consists of many events and are stored in the form of ROOT⁴ [134] files. A RAW data format contains the output of online trigger decisions and has a size of 1.5 MB/event, the size of RAW data in a simulated sample is around 2 MB/event due to additional information. A RECO data format contains the offline reconstructed data that includes the high level physics objects such as electrons, photons, jets, b-jets, muons, taus E_T^{miss} etc., in convenient format and has a size of 250 kB/event. Usually, a RECO data format contains a large amount of unwanted information, so more compact data formats are obtained by filtering the RECO data. These data formats includes Analysis Object Data (AOD) and mini Analysis Object Data (mini-AOD), used in almost all physics analyses.

⁴ROOT is a data analysis framework

EVENT RECONSTRUCTION AND SIMULATION

The detector provides information about the collisions and the underlying physics processes in the form of various electronic signals. A number of steps have to be followed in order to obtain a meaningful physics result, by making use of scientific and theoretical knowledge, computing and man-power. A diagrammatic representation of an event processing chain is presented in Fig. 3.1.

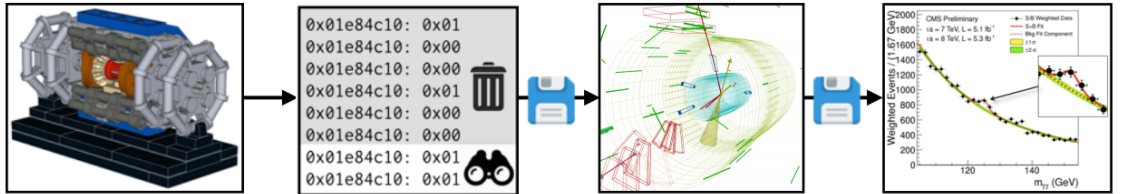


Figure 3.1: A diagrammatic representation of an event processing chain.

The electronic signals from various sub-detectors are stored, after digitization, in the form of RAW data using online trigger and DAQ system. This RAW data is further processed using a software based multi-level process, known as “Event Reconstruction” [135], that refines the data, applies various calibration corrections and construct high-level physics objects¹. At CMS, two levels of reconstruction are performed, one is a fast reconstruction, known as prompt reconstruction, done alongwith the data collection by using less time consuming algorithms and the

¹High-level physics objects are the reconstructed particles like photons, electrons, muons, taus etc.

second one is a detailed reconstruction, known as re-reconstruction, performed at a later time. The reconstructed data is written in various data formats (e.g. RECO, AOD, mini-AOD) that differ in size, technicalities, compactness etc.

Based on the theoretical knowledge, it is possible to simulate the behaviour of various particles within the detectors using event generators, also called Monte Carlo (MC) generators [136, 137]. Event generators make use of some computer programs based on Monte Carlo techniques [138] to generate physics events corresponding to different physics processes. This simulated data form an integral part of any physics analysis, used in determining various SM parameters, validating/rejecting any new theory, forming expected background for experimental data, evaluating various uncertainties and trigger parameters etc. In the language of statistics, it forms the null hypothesis for the test of any new theory. The simulated data is also written in different data formats like RECO, AODSIM, mini-AODSIM, in correspondence with reconstructed data.

This chapter details the different steps and procedures used in the reconstruction and simulation of the events for CMS physics analyses.

3.1 Event Reconstruction

Reconstruction is the process of constructing physics quantities using the raw data collected from the experiment. Inside the CMS detector, starting from the collision point, the particles first arrive in the tracker where the charged particle's trajectories, their momenta and interaction points are reconstructed using the particle's signal and bending angle in the sensitive layers of tracker. If the particle is either electron or photon, it will be absorbed in ECAL. The EM showers generated by such particles are detected as energy clusters in ECAL cells and are used in the determination of energy and direction of these particles. Hadrons (charged and neutral) will be absorbed in HCAL, sometimes initiating their showers in ECAL. The associated clusters are used in measuring the energies and directions of hadrons. Only muons and neutrinos pass through the calorimeters with negligible interactions. While neutrinos escape unidentified, muons leave their signature in muon detectors, used to reconstruct the muon direction and momentum. This behaviour of different particles is graphically depicted in Fig. 3.2 in a slice of CMS detector.

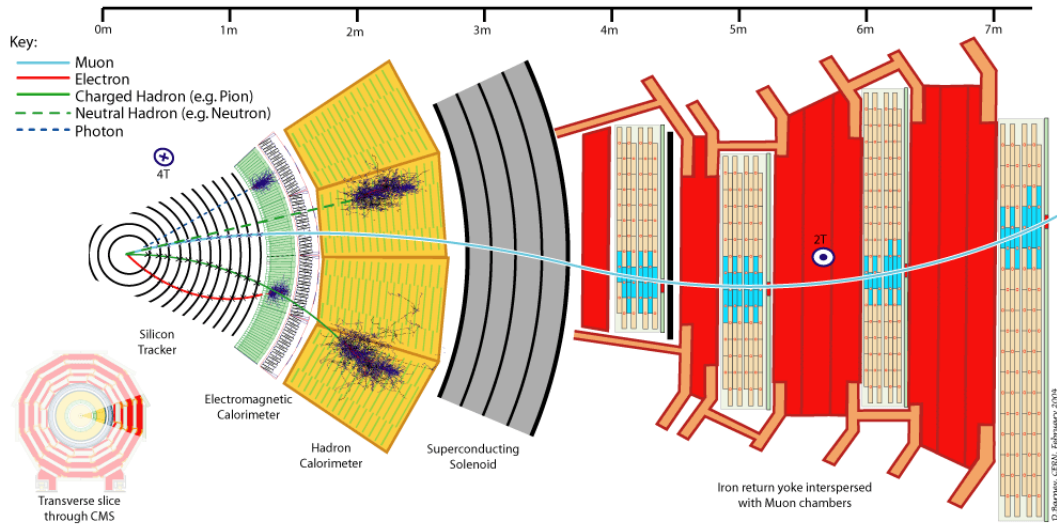


Figure 3.2: A graphical representation of the trajectories of various particles through the CMS detector.

The CMS reconstruction software [135] is a collection of several independent units, each consisting of many reconstruction algorithms and provides one set of reconstructed objects at the output. The software has been developed and validated by making use of simulated data. The complete reconstruction process is divided into three steps as shown in Fig. 3.3.

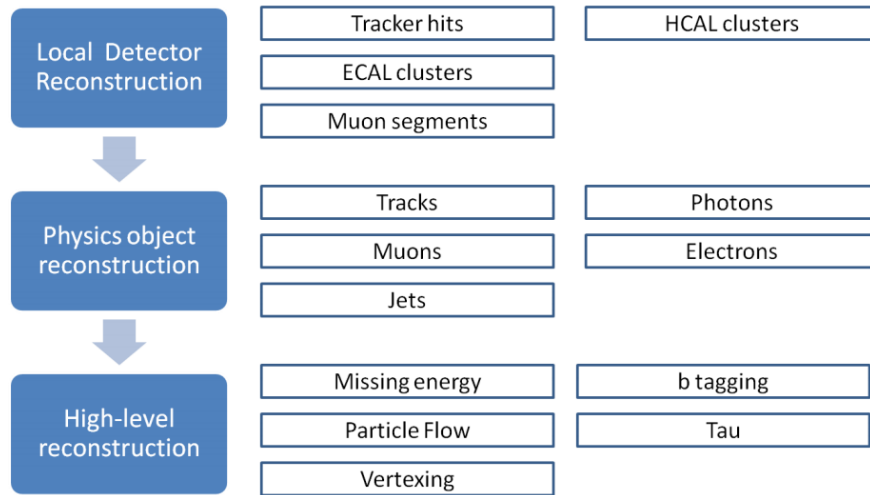


Figure 3.3: Different steps of CMS reconstruction program.

At first, the local reconstruction takes place within a particular detector module of a sub-detector. It uses raw data as input and provide in output the reconstructed units, also called rechits. These rechits are corrected for any uncertainties by the application of different calibration corrections. The form of these rechits depends

on the sub-detector under consideration. These are the position measurements inside tracker and muon system while energy clusters inside the calorimeters. These reconstructed hits act as input to the next reconstruction level, known as global reconstruction where the physics object reconstruction takes place. At this level, the information obtained from different detector modules within a sub-detector is combined to form sub-detector level objects, like high/low p_T charged particle tracks and displaced vertices in the tracker, candidate muon tracks in the muon system, photon/electrons in ECAL and jets in ECAL+HCAL. These objects from different sub-detectors are then combined to reconstruct high-level physics analyses objects using an optimally efficient Particle Flow (PF) reconstruction algorithm [139], as a final stage of event reconstruction. The high-level objects like E_T^{miss} , tau, b-tagging etc. that require the additional knowledge of tracks, displaced vertices, $\sum p_T$, are also reconstructed at this level.

Since this thesis is based on an analysis that requires a photon and a jet or a b-jet in the final state, so only the steps needed to reconstruct photons and jets (b-jets) are considered in detail in the following sections.

3.1.1 Track reconstruction

The tracker is the first detector in which the charge particles leave their signatures, hence reconstruction of tracks is one of the most important task of the CMS reconstruction program. More than a thousand particles pass through the tracker every 25 ns, thereby, making the track reconstruction immensely challenging. The track reconstruction is a combinatorial kind of problem, as for a partial track formed by hits in the internal layers, there can be many compatible hits forming different tracks in the outer layers, so the reconstruction algorithm should be able to properly locate the equitable hits without losing any tracks, over a broad range of p_T from 100 MeV to 1 TeV. This insurmountable feat is achieved by the use of the “Combinatorial Track Finder” (CTF) [111, 140] algorithm which is an extension of the Kalman filter [141–143] algorithm, an algorithm that allows single framework modelling of complete track reconstruction program.

At first, the individual hits in pixel and strip tracker are reconstructed at the local reconstruction level by clustering the zero-suppressed² signals passing the

²Zero-suppression is the default noise cleaning performed in CMS detector, in which all signals below a particular threshold are rejected, the value of the threshold depends on the sub-detector part under study.

specified thresholds. The corresponding hit positions and associated uncertainties are also estimated with an average efficiency of more than 99%. These hits are then used as input into the CTF algorithm to reconstruct the charged particle tracks and their position/momentum. The CTF algorithm is based on a process known as “iterative tracking” that reconstruct the track collection in a number of iterations. The basic idea is to first sort out the easily accessible tracks, like the high p_T ones near the interaction region and do not consider the hits corresponding to these tracks in the next iterations, thereby, successively reducing the complexity of subsequent iterations. This strategy is very useful in building up of more difficult tracks, like low p_T or displaced tracks. The algorithm has a total of 6 iterations, starting with iteration 0, that reconstruct the prompt tracks (close to pp interaction point) with $p_T > 0.8 \text{ GeV}$ and having at least three pixel hits. Iteration 1 takes care of all other high p_T prompt tracks with two pixel hits. Iteration 2 has been configured to locate low p_T ($p_T \leq 0.8 \text{ GeV}$) prompt tracks. Lastly, the iterations 3 – 5 are intended to recover tracks originated outside the luminous region of pp interactions and not considered in previous iterations. The first three iterations of CTF has a prompt track reconstruction efficiency of $\sim 90\%$ for charged hadrons and $\sim 99.5\%$ for muons, the efficiency for hadrons deteriorate due to their nuclear interactions with the tracker material. The track reconstruction is the most CPU extensive task with each iteration step constituted of four further steps as detailed below:

- **Seed generation:** In this step, the track seeds are generated using the hits in the internal pixel layers of tracker. A set of parameters are used to determine the compatible hits that has the capability of track formation. Usually, a window around a hit in a layer is used to find the compatible hit in the adjacent layer. A seed provides an initial estimate of track parameters and their associated uncertainties. The seed generation algorithm requires the knowledge of center of beam spot and location of primary vertices including the ones coming from pileup events. Since the charged particles traverse a helical path under the effect of the uniform magnetic field, a minimum of five parameters are required to define a trajectory.
- **Pattern recognition/track finding:** The track seeds originating in the first step are used to determine the complete track by looking for other hits in the outer pixel and strip layers. The corresponding algorithm is based on the Kalman

filter technique [141–143] that makes use of coarsely estimated track parameters by the trajectory seed. As it finds new hits in successive layers, it keeps on updating the track parameters. A χ^2 -test is performed to check the compatibility of hits with the extrapolated trajectory.

- **Track fitting:** The candidate track estimated by the pattern recognition can be biased by the constraints applied to the trajectory at the seeding step. Therefore, the trajectory is fitted again by initiating a Kalman filter from the innermost hit and proceeding iteratively through all the hits in the outward direction, the track parameters are updated sequentially at each hit. The covariance matrix corresponding to the fit has been scaled by a large factor to avoid any biases. This filter is then followed by a smoothing step, in which a second filter is initiated in the outermost hit using the results obtained by the first filter and proceeds backwards toward the beam-line. The final track parameters are then obtained by the weighted average of the two filters. In order to get best precision, a Runge-Kutta propagator is used to extrapolate the trajectory. The χ^2 approach is also followed to check the presence of and removal of any spurious hits (outliers), that do not belong to the track.
- **Track selection:** In the high luminosity environment of LHC, there is a significant probability of reconstructing fake tracks (tracks not associated with charged particles). So track selection step imposes some quality requirements on the reconstructed tracks to reduce the fraction of fakes, these quality checks include the requirements on the number of hit layers in the track, the $\chi^2/\text{d.o.f.}$ of the track fit, the impact parameters of track in transverse and longitudinal directions and the compatibility with the primary vertex. A track is considered as a final track candidate if it passes these quality criteria.

3.1.2 Vertex reconstruction

The vertex reconstruction [111] involves determination of all the pp interaction vertices in an event, coming from the signal as well as the pileup collisions. The algorithm is constituted of three steps, which include the track selection, clustering of the tracks that seem to be originating from the same interaction point and fitting of the tracks to locate the vertex position.

The track selection step sort out the prompt tracks that are appeared to be produced in the primary interaction region by requiring the maximum transverse

impact parameter significance relative to the beam spot [144] position to be < 5 , more than 2 pixel or more than 5 pixel+strip hits and normalized χ^2 from the trajectory fit to be < 20 . No requirements are imposed on track p_T in order to ensure high reconstruction efficiency. The clustering of the selected tracks is done by the use of a deterministic annealing (DA) algorithm [145] which considers the z-coordinates of the tracks at the point of closest approach from the center of beam spot and their corresponding uncertainties. This algorithm considers many prototype vertices initially and assign a number of vertices to each track with different weights which reflects the track consistency to be coming from the beam spot. A priori, each of the configurations is considered to be equally likely. A χ^2 minimization is performed to obtain the most probable vertex corresponding to each configuration which finally results into a number of effective vertices at distinct positions.

3.1.3 Photon reconstruction and identification

Photons are one of the most abundant particles produced in pp collisions either as the direct products of pp interactions (known as prompt or signal photons) or as the decay products of other particles. These, along with electrons, are detected and reconstructed using ECAL signals. A detailed description of the procedure used has been provided as follows.

3.1.3.1 Photon reconstruction

Photons are absorbed in ECAL by depositing their energies, in the form of showers of secondary particles, in a number of ECAL crystals. The energy reconstruction of photons takes place by grouping the energy depositions in different crystals in the form of clusters and superclusters (clusters of clusters). If no material is present between the collision point and the calorimeter, then electrons/photons deposit around 97% (94%) of their energy in 5×5 (3×3) crystal matrix. But the presence of tracker in-between, results into showering even before reaching ECAL, with a spread in ϕ direction arising due to the effect of magnetic field. Similar reconstruction algorithms are used for both electrons and photons; electrons also use further directional information from the tracker.

The photon reconstruction [146,147] in CMS takes place in a number of steps illustrated in the Fig. 3.4. The first step is the local reconstruction and calibration

of EcalRecHits. This is then followed by a second step that involves clustering of various rechits to form energy clusters and superclusters. The third and final step considers the energy scale correction to the reconstructed superclusters. These steps are explained one by one in the coming sections.

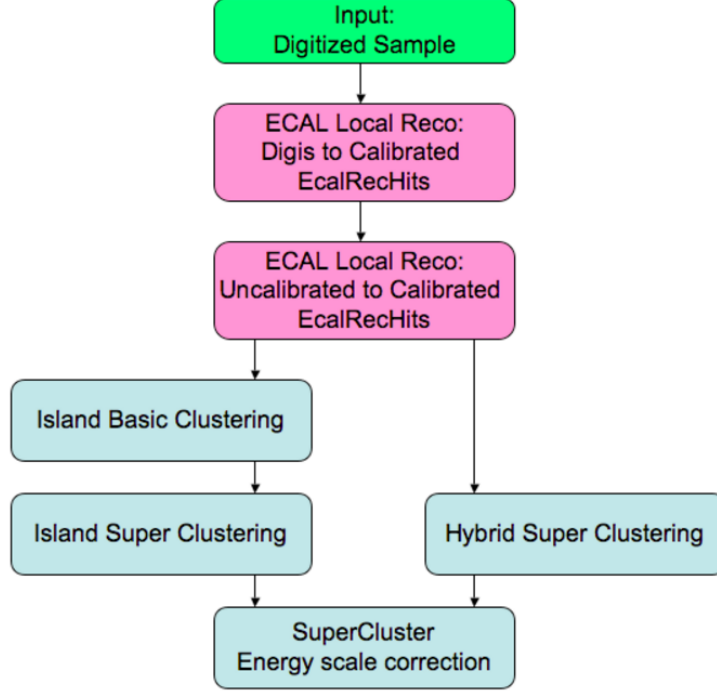


Figure 3.4: Various steps followed in reconstruction of particles inside ECAL.

3.1.3.1.1 Calibration of Ecal channels

The e^-/γ signals in ECAL are calibrated and corrected to account for a number of detector effects [117]. The ECAL crystals have a tendency to degrade light response if illuminated continuously for a long time, therefore, the transparency of these crystals is monitored during data taking by using a synchronous laser system and the observed changes are corrected for individual crystals during event reconstruction. The relative calibration of individual channels is done by deploying the knowledge of ϕ -symmetry of various energy depositions, the invariant mass of π^0/η mesons in diphoton channel and the momentum measurement of electrons, coming from decays of $W^\pm$ and Z bosons, in tracker.

3.1.3.1.2 Clustering

The clustering algorithms group together the crystals associated with different electromagnetic showers into clusters. Since the showers have a spread in ϕ -direction

due to the magnetic field, the clusters which are close in η -direction are further grouped together into superclusters using a larger ϕ window. Due to different crystal geometries in ECAL barrel and endcap, two different algorithms, namely, Hybrid algorithm in barrel and Multi5 $\times$ 5 (a modified island) algorithm in endcap, described in detail below are used.

- **Hybrid algorithm:** The hybrid algorithm is a super-clustering algorithm deployed in the ECAL barrel that makes use of the η – ϕ geometry of crystals and the corresponding knowledge of lateral shower shape in η -direction to construct superclusters of electron and photon energy deposits. This algorithm considers fixed length window in η while variable length window in ϕ -direction. An illustration of the algorithm has been provided in Fig. 3.5.

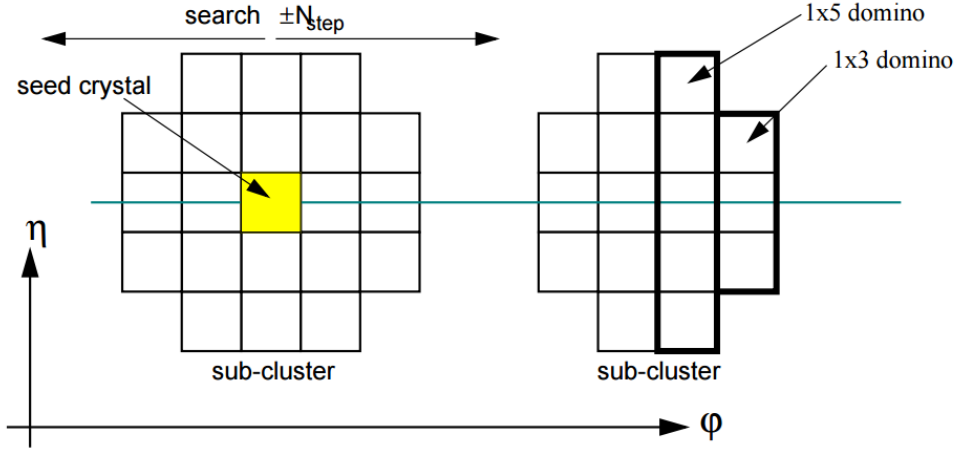


Figure 3.5: An illustration of the Hybrid algorithm used for clustering in ECAL barrel region.

The complete step-by-step procedure is outlined as follows:

1. Start by searching for a seed crystal, that is the maximum energy crystal satisfying $E_T > E_{T,seed}^{thres}$ (1 GeV). The crystal should not belong to any other cluster. Terminate the process in case no such crystal is found.
2. Construct a 1 $\times$ 3 domino of crystals in ϕ – η direction. If $E_{domino} > 0$, then extend it to a 1 $\times$ 5 domino. Add other 1 $\times$ 5 dominoes to it by moving in ϕ -direction on either side of seed domino upto $\pm N_{step}$ crystal length in ϕ where $N_{step} = 17$. If energy of a domino falls below E_{domino}^{thres} (0.1 GeV), do not consider it.
3. Group the selected adjacent dominoes into local energy clusters. If energy of

a local cluster falls below $E_{\text{cluster}}^{\text{thres}}$ (0.35 GeV), then remove the corresponding dominoes from the final supercluster.

- **Multi5×5 algorithm:** Since the geometry of crystals in ECAL endcap is not the same as in the barrel, a different algorithm named Multi5×5 has been utilized in endcaps to collect the energy deposits within a window in $\eta - \phi$ plane. The Multi5×5 algorithm first naively creates the basic clusters and then group them together into superclusters on the basis of their proximity in η/ϕ direction. The complete description of the procedure is provided as follows:

1. At first, the seed crystals that do not belong to any cluster and that satisfy $E_T > E_{T,\text{seed}}^{\text{thres}}$, required to be equivalent to 0.18 GeV, are selected. Terminate the process if no such seed is found.
2. The seeds are required to be local maxima when compared to the four direct neighbours. If this is not the case, repeat the previous step again. Arrange all the seeds in descending order of E_T .
3. Start with the highest E_T seed and move towards the lower ones while constructing 5×5 crystal matrices around each. In this process, do not include the crystals in a matrix that have already been clustered in a previous matrix. The matrices in Multi5×5 algorithm can partially overlap with each other but the crystals will only belong to the one with high E_T seed. The overlapping matrices represent the close showers due to bremsstrahlung etc.
4. In order to construct a supercluster, all the 5×5 clusters whose total energy satisfies, $E_{T,\text{cluster}} > E_{T,\text{cluster}}^{\text{min}}$ (1 GeV), are considered within a search road in η and ϕ around each seed having the range $\eta \pm \Delta\eta$ and $\phi \pm \Delta\phi$ where $\Delta\eta = 0.07$ and $\Delta\phi = 0.3\text{rad}$. An illustration of this is presented in Fig. 3.6.

Since the pre-shower detector is present in front of ECAL endcaps in the η range $1.65 < |\eta| < 2.6$, the particles deposit a fraction of their energy in the pre-shower before reaching the endcap. In order to obtain this energy, the energy weighted positions of all the 5×5 clusters belonging to a supercluster in endcap are taken and extrapolated to the preshower planes, while taking the most energetic cluster as reference point. The maximum distance between the clusters and the reference point in ϕ , along with an extension of $\pm 0.15\text{rad}$, is taken as the pre-shower clustering range in ϕ . The corresponding range in η has been set to ± 0.15 . The

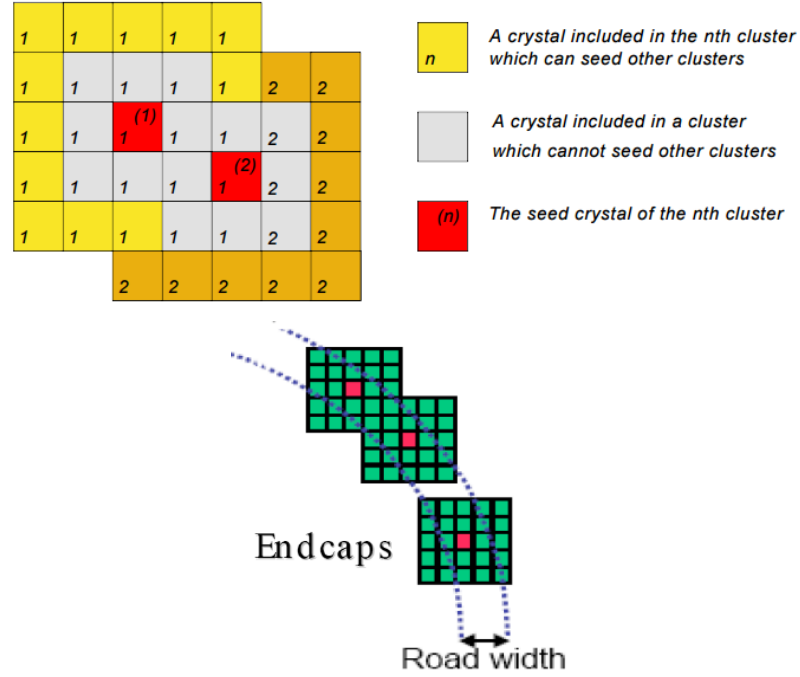


Figure 3.6: An illustration showing overlapping of two Multi5x5 clusters. Clusters in yellow are eligible to seed other Multi5x5 clusters if they form the local energy maxima. The mechanism of formation of a supercluster using many multi5x5 clusters is also shown.

energy deposited in pre-shower in this range is computed and added into the endcap supercluster energy.

Each supercluster corresponds to a potential photon candidate. The position of a supercluster in all the clustering algorithms is estimated via an energy weighted average of the positions of all the crystals belonging to the supercluster. Each crystal is assigned a weight equivalent to, $w_i = \max(0, 4.7 + \ln(E_i/E_{SC}))$ [148]. The supercluster position is then connected to the primary vertex of the event to obtain a direction of photon's momentum. If many primary vertices exist in an event, then the one with highest track sum p_T^2 is considered.

3.1.3.1.3 Energy scale corrections of superclusters

There are many factors that affect the resolution of the energy measured in ECAL using clustering algorithms. These mainly include the energy losses due to the bremsstrahlung and photon conversions in the material of the tracker, leading to an underestimation of the energy. The aim of supercluster energy corrections is to compensate for various energy losses and achieve a homogeneous response

throughout the ECAL. The corrected energy of a reconstructed particle in ECAL is given by

$$E = F \times G \times \sum_{\text{cluster}} c_i \times A_i \quad (3.1)$$

where A_i refers to the signal amplitude given by the ADC counts; c_i is the inter-calibration term that equalizes the different crystal's response (as discussed under 3.1.3.1.1); G is the scale representing the global correction term defined such that the product of G and energy amplitude of a 5×5 crystal matrix is equivalent to the energy of an unconverted photon. The factor F refers to the supercluster correction term and it includes corrections for three type of effects in barrel while two type of effects in endcap, described as below:

- The factor $C_{\text{EB}}(\eta)$ used in barrel only to compensate for the lateral energy leakage from the exposed sides of the EB crystals.
- The correction term f_{brem} that is used to correct for the response of the clustering algorithms with different SC topologies towards the shower.
- The residual correction term, $f(E_{\text{T}}, \eta)$, applied to all the reconstructed superclusters. It corrects for the non-linear matter distribution in the detector and the corresponding energy dependence.

3.1.3.2 Photon identification

As photons are an important ingredient of many prominent physics searches, a precise identification and selection of photons is a crucial task of CMS physics program. The photons selected using various reconstruction algorithms can have a significant fraction of fake photons coming from different sources. In addition, the signal photons (directly coming from pp interactions) are contaminated with a large fraction of background photons (coming from decay of other particles), arising mainly from the decay of neutral mesons (mainly, π^0 and η) produced in the jets. These mesons decay into two photons which can get reconstructed as a single photon if the initial p_{T} of mesons is quite high. Two different algorithms for photon identification have been deployed in CMS: one approach considers the selections applied to the individual photon variables; while second approach uses a multivariate technique.

The analysis documented in this thesis adopts the first approach that makes use

of the different photon variables utilizing the knowledge of properties of photons. These variables can be divided into two different categories as mentioned below:

3.1.3.2.1 Shower shape variables

The direct photons are distinguished from background photons and other particles on the basis of their lateral shower patterns within the calorimeter cells. A number of variables are utilized under this category, including

- **Shower profile:** Studying the lateral extension of energy deposits in the η -direction, is a very sophisticated and powerful way to distinguish the showers due to a single photon or two overlapping photons. A variable, $\sigma_{i\eta i\eta}$, is used that provides the transverse shape of EM cluster in a particular η direction and is defined as the energy weighted standard deviation of a single crystal η in a matrix of 5×5 crystals centered around the maximum energy crystal. It is computed using logarithmic weights and is defined as:

$$\sigma_{i\eta i\eta}^2 = \frac{\sum_{5 \times 5} (\eta_i - \bar{\eta})^2 w_i}{\sum w_i}, \quad (3.2)$$

$$\text{where } \bar{\eta} = \frac{\sum w_i \eta_i}{\sum w_i} \text{ and } w_i = \max \left[0 ; 4.7 + \log \left(\frac{E_i}{E_{5 \times 5}} \right) \right].$$

Here the distances η are measured in units of crystal size in the η -direction. This variable has a symmetric and narrow distribution for a prompt photon while has a long tail on the lower side of the distribution for misidentified or overlapping photons. Another variable, $\sigma_{i\phi i\phi}$, a counterpart of $\sigma_{i\eta i\eta}$ in ϕ direction is also used, however, it is less effective compared to $\sigma_{i\eta i\eta}$.

- **Single tower hadronic over electromagnetic ratio (H/E):** It is defined as the ratio of energy deposited in the single HCAL tower, closest to the photon supercluster position and centered in the photon direction, within a cone of radius $R = 0.15$ in $\eta - \phi$ plane and the energy deposited by that supercluster in the ECAL. Due to the large depth of ECAL ($\sim 26X_0$), the probability of EM shower leakage in HCAL is relatively small, so this variable proves to be a good discriminator.
- **R₉:** It is defined as the ratio of total energy deposited in a 3×3 crystal matrix around the most energetic crystal in a supercluster and the total supercluster energy. The energy spread due to a pair of overlapping photons is larger as compared to a single photon, therefore, this variable also provides a good

discrimination.

- **Conversion safe electron veto:** This variable is used to distinguish between electrons and converted photons. It requires that for a photon cluster in ECAL, there should not be any charged particle track with pixel hits, matching with the conversion vertex, in the tracker pointing towards the photon direction.

3.1.3.2.2 Isolation variables

If a photon candidate is a direct descendant of a pp collision, it will be quite away from other particles produced in the same collision. On the other hand, if a photon is a decay product of a hadronic jet, it will be associated with hadronic activity in its vicinity. Therefore, requiring limited amount of activity around the photon results into extraction of good photon candidates. Various isolation sums, expressed as the total energy/momentum carried by various particles within a specific volume around the photon candidate, are estimated using the particle flow (PF) algorithms [139], making use of all the sub-detectors. These sums are evaluated within a cone of radius $R = 0.3$ around the photon candidate while neglecting the energy/momentum within a veto region of smaller radius, to avoid the inclusion of the candidate photon energy. Three different isolation sums depending upon the type of particles under consideration, are defined:

- **PF charged hadron isolation:** $\sum p_T$ of all the charged hadrons within a hollow cone of $0.02 < R < 0.3$ around the photon supercluster.
- **PF neutral hadron isolation:** $\sum p_T$ of all the neutral hadrons within a cone of $R = 0.3$ around the photon supercluster.
- **PF photon isolation:** Scalar $\sum p_T$ of all the photons within a cone of $R = 0.3$ while excluding a strip in η of 0.015 around the photon supercluster.

3.1.4 Jet reconstruction and identification

Jets in the detector represent the experimental signature of quarks and gluons emerging from the high energy pp collision. Since, quarks and gluons can not survive individually owing to their colour confinement, they hadronize in the form of jets soon after their production. Therefore, jets provide a way to access and study the underlying processes at partonic level. A precise determination of jets

and understanding of their properties is of utmost importance for many physics searches.

Jets are made up of a large number of particles including leptons, hadrons as well as bosons (mainly photons). Within the detector, these appear as a collimated spray of particles in a particular direction as seen in Fig. 3.7. As they travel through the detector, they leave their signals in different detector components. These signals are then combined by the jet reconstruction algorithms to build a jet. A number of jet energy corrections are also considered to equalize the reconstructed jet energy to the true particle-level energy.

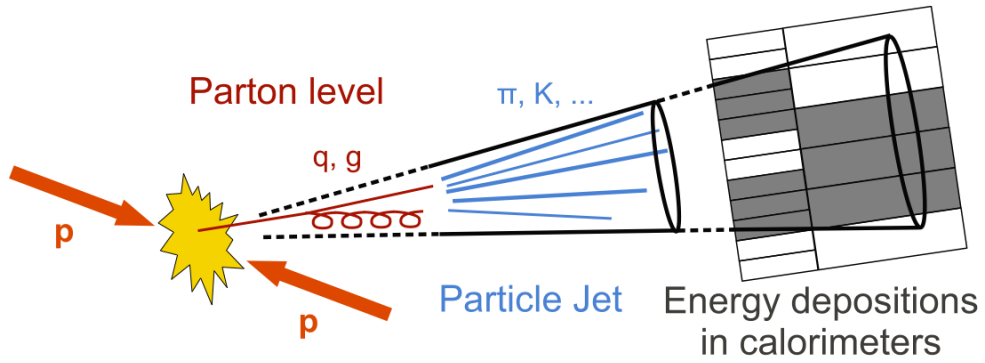


Figure 3.7: A pictorial representation of a jet.

3.1.4.1 Jet reconstruction

The aim of jet reconstruction is to reconstruct and identify jets coming from the hadronization of partons. This process can be divided into two parts: The first part reconstructs the individual jet constituents and the second part clusters them together to produce a complete jet.

For reconstruction of jet constituents, three different approaches are considered at CMS: a calorimeter based approach which utilizes the information of the calorimeter energy deposits only; a Jet-Plus-Track (JPT) approach that improves the calorimeter information by using the additional knowledge from the associated tracks; a Particle-Flow (PF) approach that attempts to combine the information from all the relevant sub-detectors. A highly improved spatial and momentum resolution can be achieved by the use of PF approach as compared to calorimeter or JPT approach.

The jet clustering algorithms are used to cluster individual jet constituent particles to form a jet. These algorithms can be broadly divided into two classes:

the sequential recombination algorithms and the cone algorithms. The sequential algorithms work by combining a pair of particles based on the closeness in some distance measure and then repeating the procedure for other particles until some stopping condition is reached. The main sequential algorithms employed at CMS include, k_T [149], anti- k_T [150] and Cambridge/Aachen (CA) [151, 152], described by the distance measurements,

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad \text{and} \quad (3.3)$$

$$d_{iB} = k_{Ti}^{2p},$$

where,

d_{ij} : distance between i th and j th particle,

d_{iB} : distance of i th particle from the beam line,

k_T : particle's transverse momentum, and

ΔR_{ij} : spatial distance between i th and j th particles, $\sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$.

The parameter R determines the angular reach of a jet and is usually called the jet radius. The parameter, $p = 1$ for k_T algorithm, implies that this algorithm clusters the soft particles (low p_T) first, $p = -1$ for anti- k_T algorithm means that it clusters hard particles (high p_T) first and $p = 0$ for CA algorithm performs an energy independent clustering. On the other hand, the cone algorithms define specific conical regions around the particles and put together all the particles inside the cone in the form of a jet such that the total momentum of the particles coincide with the cone axis. The famous cone algorithms include, Iterative cone, Midpoint cone, SIScone etc. These algorithms mainly differ from each other in the strategy used to search for the stable conical region. The Fig. 3.8 represents the behaviour of different jet clustering algorithms.

The sequential algorithms have a number of advantages over the cone algorithms, like: simplicity and detector independence, can be easily implemented in experimental analyses and theoretical calculations, insensitive to hadronic effects, infrared and collinear safe³ etc. The analysis in this thesis considers the particle flow jets clustered using the anti- k_T algorithm. A detailed description of these has been provided below:

³Infrared safety requires that the clustering algorithms should not be affected by the inclusion of an extra soft gluon and collinear safety requires that the jet clustering should remain same for splitting of a parton into two partons. Both of these lead to finite perturbative calculations upto any order.

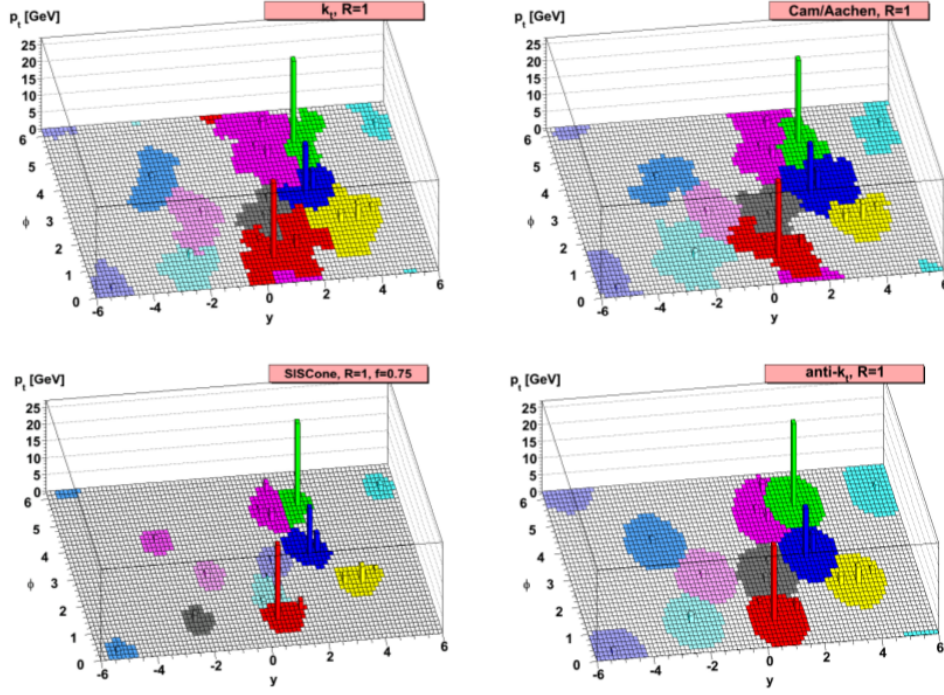


Figure 3.8: The behaviour of different jet clustering algorithms.

Particle flow reconstruction algorithm: The particle flow algorithm of CMS [139, 153] relies mainly on the high efficiency of track reconstruction and fine granularity of ECAL and HCAL. This algorithm works by linking the energy clusters in the calorimeters with the tracker signals to create a complete picture of the event. Since, this algorithm reconstructs individual particles corresponding to each sub-detector first, a dedicated calibration can be done and the impact of non-compensating nature of HCAL can be mitigated to get an overall good response. Also due to the availability of tracking and vertex information, the impact of pile-up collisions can be reduced. The list of individual particles from each sub-detector is then used to reconstruct jets, E_T^{miss} , τ , b-jets etc.

This algorithm is performed in three main steps. The first step builds the basic elements of PF algorithm, like the reconstruction of charged particles tracks in the tracker and energy clusters in the calorimeters. The energy clustering in HCAL starts by searching for a seed channel, a local energy maxima. The topological clusters are then formed by adding neighbouring channels to it. The energy of each calorimeter channel is required to be above some predefined threshold, in order to avoid any contributions from electronic noise etc. The energies and positions of the final clusters are determined using an iterative method by re-weighting the

contributions of individual channels on the basis of their distance from the seed channel. This step is then followed by a second step that involves the building of PF blocks using a mechanism known as link algorithm, which performs the linking of the tracks with the energy deposits in ECAL and HCAL, compatible with the electromagnetic and hadronic shower profiles. A link between ECAL and HCAL energy deposits is also made if the ECAL cluster position lies within the envelope of HCAL cluster. The tracks are also linked to the muon hits in the muon system, thereby forming global muons. The resultant PF block consists of a maximum of three elements. The quality of these blocks is determined by the compatibility of its elements, either in the form of $\eta - \phi$ distance in case of track-to-cluster and cluster-to-cluster links or in the form of χ^2 fit in case of track-to-muon links. The third step consists of the reconstruction of final PF objects on the basis of the PF blocks.

The performance of PF algorithm has been studied using simulated events [153]. A great precision is achieved, using tracker and ECAL, for the measurement of around 90% of jet energy taken by the charged hadrons and photons. The remaining 10% of jet energy shared by neutral hadrons is measured solely in HCAL and has an energy resolution of $120\%/\sqrt{E}$.

Anti- k_T clustering algorithm: The default jet clustering algorithm used in CMS is one of the sequential recombination algorithm, the anti- k_T [150] algorithm. It determines two distance measures for a pair of particles (i, j) as

$$d_{ij} = \min(k_{Ti}^{-2}, k_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \quad \text{and} \quad (3.4)$$

$$d_{iB} = k_{Ti}^{-2}.$$

This algorithm works by finding the distance of i th particle from the closest j th particle (d_{ij}) and from the beam line (d_{iB}), by using the definition of distance measures given above. If i th particle is closer to j th particle as compared to beam line (d_{ij} is smallest), it replaces i th and j th particle with a single new particle of momentum $k_{Ti} + k_{Tj}$, often called a pseudojet, as it is neither a single particle nor a full jet. However, if the i th particle is closer to beam line as compared to j th particle (d_{iB} is smallest), then it removes the i th particle from the list of particles under study and considers i th particle as an inclusive jet. The algorithm then repeats this procedure with a new pair of closest particles/pseudojets, until it includes all the particles. The CMS uses anti- k_T algorithm corresponding to two

values of jet radius, viz 0.4 and 0.8, forming AK4 and AK8 jets.

3.1.4.2 Jet energy corrections

The energy of a reconstructed jet is not the same as that of the true energy of a particle level jet. The primary reason being the non-linearity and non-uniform response of calorimeters, especially HCAL. The jet energy corrections (JEC) [154] are a set of tools that performs the mapping of the reconstructed jet energy to the actual particle-level energy. At CMS, a multi level factorized approach to JEC has been considered where each level correspond to a different correction of specific type. These corrections are extracted using the data-driven as well as the simulation methods and are provided as *in-situ* calibrations in data and MC. Each correction level is dependent on various jet quantities and provides a scaling to the jet four momentum⁴. These corrections are always applied sequentially in a fixed order. The different JEC levels are represented in Fig. 3.9 and are described below:

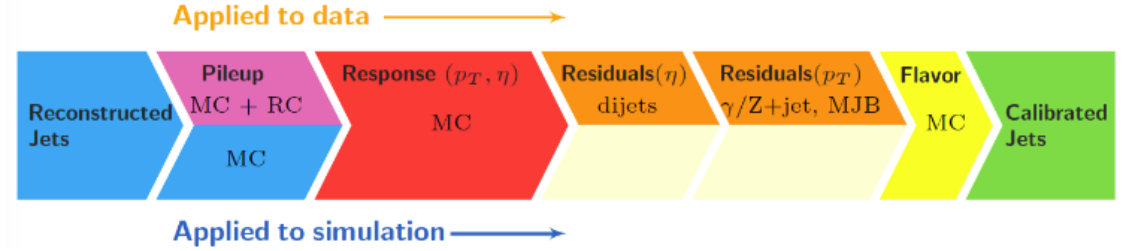


Figure 3.9: Different jet energy correction levels applied to data and MC.

- **L1 Offset:** The purpose of the L1 correction is to subtract any energy contribution coming from the pile-up and underlying events [155]. This correction, in principle, removes the dependence of data sample on the luminosity, so that the further corrections are applied on a luminosity independent sample. These corrections are estimated using the events processed with and without considering pile-up overlay in a simulated sample of QCD dijet and parameterized as a function of energy density ρ , jet area, η^{jet} and p_T^{jet} . The residual differences between data and simulation are corrected as a function of η using a random cone method implemented using zero-bias events, resulting into different L1 corrections for data and MC.

⁴(p_x, p_y, p_z, E)

- **L2 Relative:** These corrections are applied to both data and MC to make the jet response uniform as a function of pseudorapidity, η .
- **L3 Absolute:** After correcting the jet response in η , it is also made uniform in p_T using L3 corrections. The L2 and L3 corrections are now computed together using QCD dijet samples and are collectively called L2L3 MC-truth corrections.
- **L2L3 Residual:** This correction is determined by the *in-situ* measurements of absolute and relative jet energy scales and is meant to correct the small differences between data and MC, of the order of a few %, in jet response. A correction factor, equivalent to the residual difference between data and MC, is applied only to data.

Apart from these, an optional L5 flavor correction derived using Z+jet and γ +jet simulated samples, is also applied in some cases. The analysis in this thesis considers the essential L1, L2, L3 and L2L3 corrections only. Each correction level also has the related uncertainties, forming the biggest source of systematic error in a jet measurement.

3.1.4.3 b-tagging of a jet

The jets arising from the hadronization of b-quarks are known as b-jets. The b-jets have many distinctive signatures (Fig. 3.10) like, long lifetime ($\sim 10^{-12}$ s), large mass (4.2 GeV), high track multiplicity ($\sim 4-5$), relatively large semi-leptonic branching ratio ($\sim 20\%$ for each of e^- and μ^-), hard fragmentation function⁵ etc., which leads to their precise measurement. The CMS with its rigorous charged particle tracking, powerful lepton identification and highly segmented calorimetry, is well matched for the task of b-jet identification [156, 157]. The b-tagging is a kind of reconstruction procedure that utilizes various b-hadron properties and based on these, assigns a “likelihood” to each jet, of the presence of b-hadrons.

A number of b-tagging algorithms have been deployed at CMS that mainly differs in their complexity and input information. All these algorithms provide a numerical discriminator for a jet in the output whose value determines the b-tagging status of the jet. The higher is the value of discriminator, the larger is the probability of the jet being a b-jet. These b-tagging algorithms are constructed on the basis of the information gathered from either the tracks or secondary vertex

⁵A significant fraction of initial b-quark’s momentum is taken away by b-hadrons.

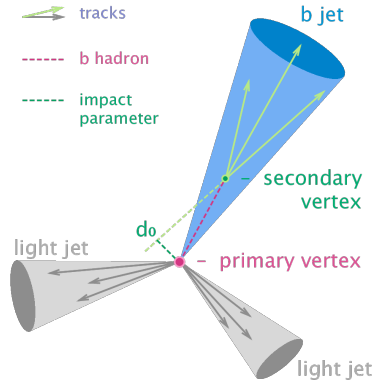


Figure 3.10: A depiction of a b-jet emerging from the primary vertex.

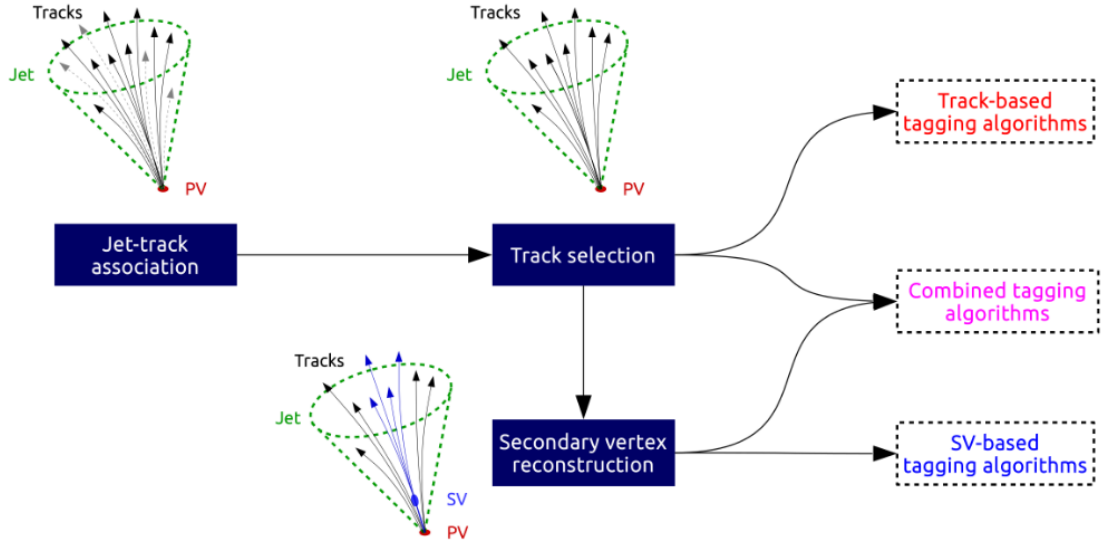


Figure 3.11: An illustration of b-tagging algorithms.

or soft lepton or any combination of these, as presented in Fig. 3.11. The main b-tagging algorithms employed in CMS are:

- **Track Counting (TC) algorithm:** This approach exploits the long lifetime of b-hadrons and considers the charged tracks associated with it. It computes the significance of the signed impact parameter⁶ of all the good tracks, arrange them in decreasing significance and assign the significance of N th track to the discriminator. It leads to an algorithm of high efficiency for $N = 2$ and high purity for $N = 3$.
- **Jet Probability (JP) algorithm:** This algorithm uses information from all

⁶Impact Parameter (IP) is defined as the minimum distance between the track and primary vertex, the sign of IP is determined by the scalar product of IP segment with the jet direction.

selected tracks in a jet. Using a sample of tracks with negative impact parameters, it builds a probability distribution function (PDF) that considers the probability of a track to be originating from the primary vertex. Based on this PDF, a probability pointer is assigned to each track and a discriminator is built using full set of tracks. A version of JP algorithm, known as jet B probability is also used that considers only the four highest displaced tracks matching with the reconstructed multiplicity of the b-hadron decay vertex.

- **Soft Lepton (SL) algorithm:** This algorithm identifies the b-hadrons on the basis of their semi-leptonic decays as the leptons from b-decay usually tend to have high p_T relative to the jet axis. The corresponding b-tag discriminator is the result of the neural network⁷ build up using lepton p_T , impact parameters and some other variables.
- **Simple Secondary Vertex (SSV) algorithm:** This algorithm considers the reconstruction of secondary vertex from the b-hadron decay using an adaptive vertex finder technique and uses the variables associated with it to compute the discriminator value.
- **Combined Secondary Vertex (CSV) algorithm:** This is the most sophisticated and complex approach making use of all the known information like secondary vertices, lifetime, IP significances, decay lengths etc. This algorithm has the capability to provide a discrimination even when no secondary vertex is reconstructed. It uses a list of variables and feeds them as input into a likelihood ratio which is then used to compute the discriminator value. A “version 2” of this algorithm known as “CSVv2” has also been introduced during LHC Run II that makes use of the artificial neural networks instead of likelihood ratio.

These algorithms have a b-tag identification efficiency varying in a range of 40–70%. These are available in three working points: loose, medium and tight, corresponding to light flavour mis-tagging rate of 10%, 1% and 0.1%, respectively. At CMS, these algorithms are determined using the inclusive multijet, muon enriched jets and dilepton $t\bar{t}$ data samples, the jets are required to be PF jets clustered using anti- k_T algorithm having a $p_T > 20 \text{ GeV}$ and lying within the tracker acceptance region, $|\eta| < 2.4$.

⁷A neural network is a hardware/software system designed on the basis of working of neurons in the human brain.

3.1.4.4 Jet identification

The reconstructed and calibrated jets contain a significant fraction of fake jets coming mainly from the electronic noise and other detector artifacts. In order to identify and locate a good collection of jets to be used in an analysis, a set of identification variables are defined [158], listed as below:

- **Charged hadron fraction:** The fraction of energy deposited in HCAL by charged hadrons.
- **Neutral hadron fraction:** The fraction of energy deposited in HCAL by neutral hadrons.
- **Charged electromagnetic fraction:** The fraction of energy deposited in ECAL by charged constituents of the jet.
- **Neutral electromagnetic fraction:** The fraction of energy deposited in ECAL by neutral constituents of the jet.
- **Number of constituents:** The number of particles in a jet.
- **Charged multiplicity:** The number of charged particles in a jet.

3.2 Event Simulation

Event simulation is an integral part of any high energy physics experiment right from the designing stage to the final analysis. It is almost impossible to perform a physics analysis without the use of simulated data. It acts as a reference to the detector output as well as allows us to study the behaviour and consequences of any new theory without having observed it directly. The various efficiencies and correction factors used by the analyses are computed by comparing experimental and simulated data while a number of systematic uncertainties are computed using simulated data alone. In general, the quality of any physics result is strongly influenced by the quality of simulation involved.

Event simulation produces events that perfectly mimic the output from an actual detector. The various steps involved in a simulation chain are presented schematically in Fig. 3.12.

As the first step of the simulation chain, the “Physics Model” in which the calculation has to be performed, is selected. The physics model can be the standard

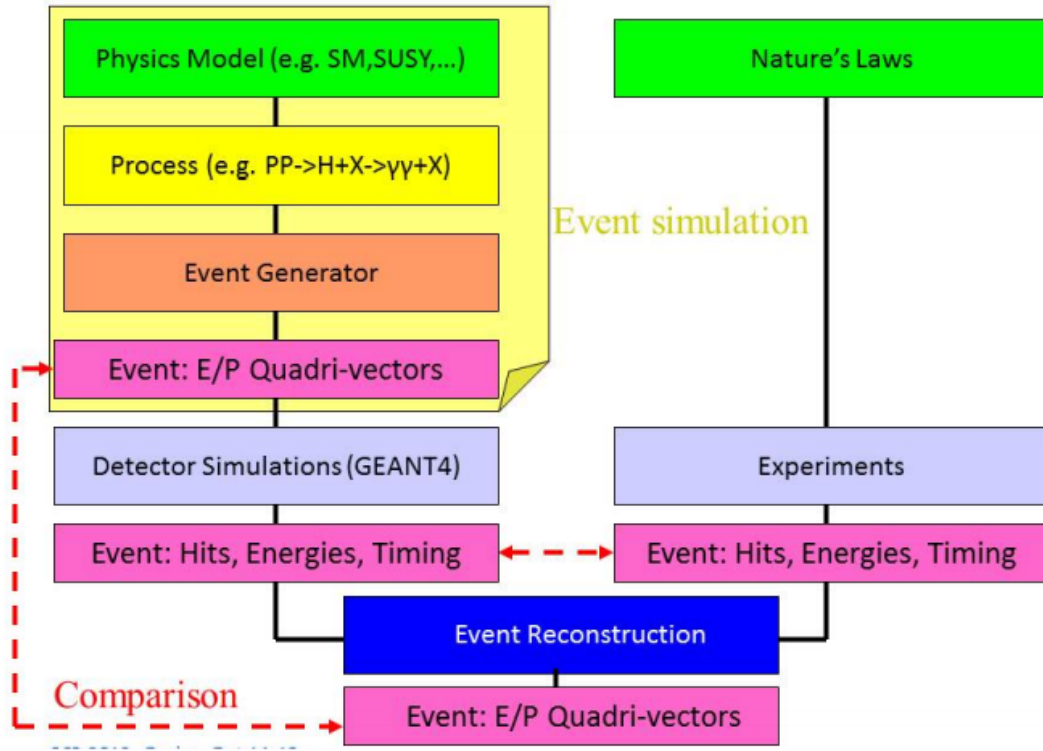


Figure 3.12: Event simulation and data analysis chain.

model (to simulate background processes) or any model beyond the standard model (to simulate signal processes of new physics). The simulation process is defined at the next level by specifying the input/output particles and the order of calculation. The expressions for matrix element and probability cross-section are prepared automatically on the basis of the selected physics model. At the “Event Generator” level, the matrix element is integrated over the entire phase space and the random samples of all final state particles, along with the information of their energy-momentum, are generated. The generated events are then made to propagate through a “Detector Simulation” model that represents the simulation of an actual detector. The huge amount of information produced in output by the detector simulation exactly replicate the output from a real detector. At this level, both the theoretical simulated data and experimental detector data are at equal footing and can be compared with each other. The simulated data is reconstructed using the same reconstruction programme as used for real data. Comparison of reconstructed simulated data with real data provides a test of self-consistency of the underlying physics model, though any discrepancy between the two may hint towards some new physics. The final analysis is performed using a data analysis toolkit, known

as ROOT [134]. The simulated and observed events passing various filters and cuts are studied and compared by organizing them in the form of histograms. At CERN, the simulation and analysis of data is performed using a data distribution grid system known as WLCG (Worldwide LHC Computing Grid) [159]. This infrastructure is based on the tiers computing centers, distributed all around the world.

The two main components of event simulation process, viz, event generation and detector simulation are described in more detail in the sections below.

3.2.1 Event generation

Event generation [136, 137] is the first stage of the simulation process. It is based on software programs that use Monte Carlo (MC) methods [138, 160] to simulate events. A Monte Carlo technique is basically an integration tool based on random sampling⁸. It is mainly used to obtain approximate numerical solutions of the analytical problems which are otherwise too complicated to solve.

The main objective of event generators is to accurately provide the description of an actual particle collision. In a high energy pp collision, the formation of a single event involve several steps as depicted in Fig. 3.13. At the start, the two partons are selected, in flavor as well as in color, from each of the two colliding protons. These partons are allowed to radiate gluons, known as initial state radiation, before interacting. The hard scattering process occurs upon interaction and results into the production of two or more final state particles, probably via a resonance state. These final state particles can further decay into other particles producing several daughters. The remaining quarks and gluons undergo the emission of parton showers before the final hadronization stage occurs, resulting into the recombination of partons to form hadrons. In addition to these, the left over partons (known as beam remnants) of a proton, after the parton underwent hard scattering is pulled off, can interact with each other, leading to the multiple parton interactions (MPI). These MPIs result into creation of multiple vertices as well as addition of extra tracks and energy deposits interfering with the main collision, referred to as the underlying events (UE).

The event generators also formulate an event in the same sequence through

⁸A method of selecting a sample from a large statistical population such that each sample in the population has some predefined probability of being selected.

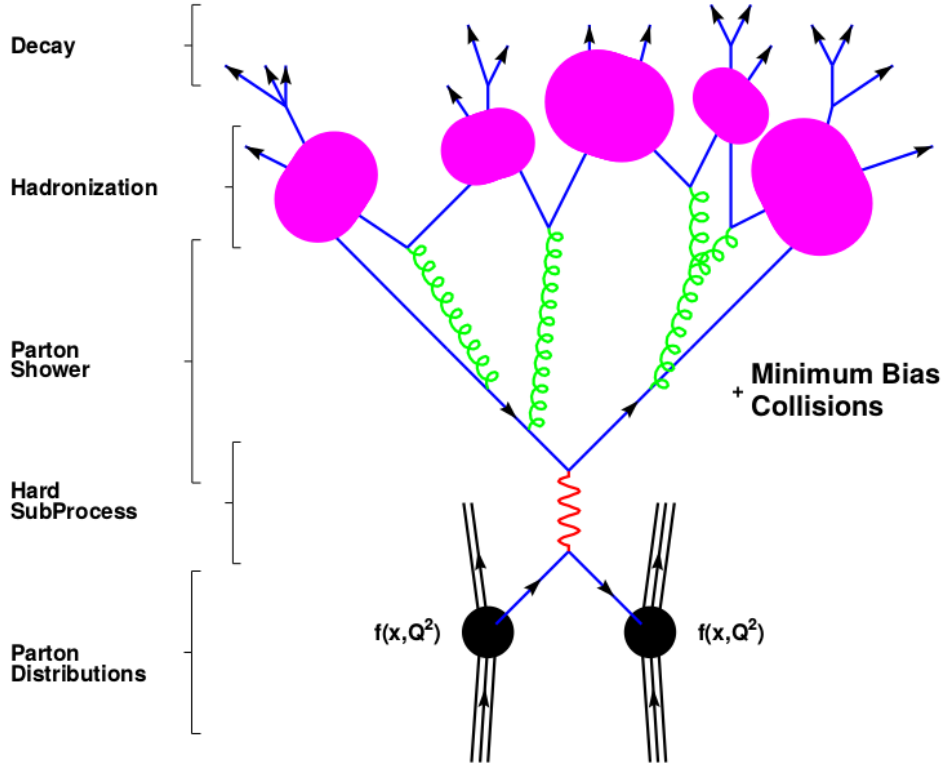


Figure 3.13: Sequential steps involved in the formation of an event.

the use of different theoretical models for different physics aspects, like hard and soft interactions, initial/final state parton showers, parton distribution functions, multiple interactions, underlying events, parton hadronization and decay etc. Simulating these aspects of a process require the calculation of relevant matrix element involving loop diagrams up to many orders, achieved by the use of dedicated MC techniques.

The cross-sections for various processes are stored within the event generators which are solved by the use of MC integration techniques. The cross-section of a $2 \rightarrow 2$ process at tree-level is given by,

$$\sigma(ab \rightarrow cd) = \sum_{ab} \int_0^1 \int_0^1 \hat{\sigma}_{ab \rightarrow cd} \left(\hat{s}, \alpha_s(\mu_R^2), \frac{Q^2}{\mu_F^2}, \frac{Q^2}{\mu_R^2} \right) f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) dx_1 dx_2 \quad (3.5)$$

Here Q is the scale of hard interaction while μ_F and μ_R are factorization and renormalization scale. The quantity, $\hat{s} = x_1 x_2 s$, s is the center-of-mass energy of collision, x_1 and x_2 are the fractions of total proton's momentum, carried by two interacting partons. The functions $f_a(x_1, \mu_F^2)$ and $f_b(x_2, \mu_F^2)$ are the probability

density function for two partons, also known as parton distribution function (PDF) [161], representing the probability of carrying a momentum fraction x_1 and x_2 by parton a and b respectively. The initial partons for a hard interaction are selected by the use of PDF. These PDF's are determined experimentally from deep inelastic scattering processes, Drell-Yan processes, and jet production at high energies. The $\hat{\sigma}$ is the parton level cross-section of the sub-process $ab \rightarrow cd$. The expression is integrated using MC techniques over the entire phase space. If the expression contains loop corrections, then integration over loop propagators is also performed. The multi-dimensional integration yield the total cross-section of the process. The event after passing through the various event generation levels is then passed to the detector simulation stage.

3.2.1.1 MC event generators

A number of event generators based on MC techniques are used in present day colliders. These can be divided into general-purpose generators and specific generators. The specific generators are good for simulating any particular physics aspect but need to be interfaced with general-purpose generators to simulate a complete event. Examples of some of the general-purpose generators are: PYTHIA [162], MADGRAPH [163], HERWIG++ [164], ALPGEN [165] and SHERPA [166]. The analysis documented in this thesis uses PYTHIA and MADGRAPH as the generators to simulate various signal and background processes. A brief description of these are given below.

3.2.1.1.1 PYTHIA

It is one of the oldest general-purpose event generator used for lepton as well as hadron colliders. Its hadronization model is based on the Lund string model [167], consisting of around 240 hard scattering processes. It is a leading order (LO) event generator and considers only two particles in initial and final states. It consists of models for initial and final state parton showers, hard processes matching and merging, multi parton interactions, underlying events etc. It also has the facility to interface with external programs.

3.2.1.1.2 MADGRAPH

PYTHIA is an ideal event generator for simulating $2 \rightarrow 2$ processes, but in actual collisions, most of the processes contain more than two particles in final state.

The MADGRAPH [163] matrix element generator has the capacity to produce the processes with up to nine particles in the final state. It uses ALOHA package [168] to automatically generate the amplitudes of relevant sub-processes. The events simulated using MADGRAPH are put in some standard event format files known as Les Houches format files (LHE files) [169] which are then passed to PYTHIA to simulate parton showering and hadronization. The PYTHIA also performs matching and merging of the matrix elements (ME) with the partons showers (PS) using CKKW algorithm [170] and MLM scheme [171].

3.2.2 Detector simulation

In order to simulate the behaviour of particles within the detector, a software based model representing a virtual detector has been used. The event generator output is made to pass through this detector simulation model to obtain a replica of actual collision data. With increase in complexity and resolution of the detectors, the detailed simulation of an experimental setup has become a challenging task. It requires efficient and sophisticated tools to accurately reproduce the behaviour of detector in the presence of collision particles.

In CMS, the simulation of the detector is performed using an object oriented toolkit known as GEANT4 [172]. The GEANT4 consists of a set of C++ libraries which can be used to simulate any detector system. It simulates the particle interactions in matter on the basis of Monte Carlo methods. The GEANT4 package provide accurate description of various detector aspects like, detector geometry, material of different detector components and the dead material such as bolts, cables, support etc. It consists of a set of physics models that are able to handle the various particle's interactions with the matter over a wide range of energy, ranging from few hundred eVs to TeV. Starting from the event generator output, the events are made to propagate through the detector simulation while taking into account the magnetic field for charged particles as well as the various interactions of particles with matter like bremsstrahlung, multiple scattering, photon conversions, scintillation, ionization, cherenkov light etc. Finally, the raw output similar to that of the actual detector data is obtained.

SEARCH FOR LIGHT AND HEAVY FLAVOR EXCITED QUARKS

This chapter documents in great detail the various steps and procedures followed in the analysis of CMS data to search for any signature of excited light and heavy quark presence in $\gamma + \text{jet}$ and $\gamma + \text{b-jet}$ final states respectively.

4.1 Analysis strategy

In high energy physics, the invariant mass is a very important parameter to determine the nature of a process. It is a Lorentz invariant quantity and is defined as the system mass constructed using the Lorentz invariant energy and momentum. For a single particle, the invariant mass is equivalent to its rest mass, given by $m^2 = E^2 - p^2$, in natural units¹. This relation is also true for a system of masses with E and p representing the total energy and momentum of the system. In a process that occurs through a resonance, the invariant mass of final state particles is equivalent to the rest mass of the resonance. The process has a larger cross-section corresponding to the energy of resonance mass.

If we plot the invariant mass distribution of final state particles for a large sample of events, the number of events corresponding to the resonance mass will be relatively larger as compared to the neighbouring masses. The resonance will

¹In natural units, $\hbar = c = 1$

appear as a peak over the continuous distribution, arising due to the standard model processes referred to as the background processes, with the peak-width proportional to the resonance's life time. The background due to various SM processes is subtracted either by the use of Monte Carlo simulated background samples or by the use of data driven technique where the background is evaluated by fitting the invariant mass distribution with a smooth parameterization. Any significant deviation from the smooth distribution at any point in data is considered as a signal of new physics. In case of absence of any significant deviation, the resonance signal model, simulated on the basis of the underlying theory, is used to evaluate a signal cross-section (mass), above (below) which the presence of the signal is discarded with some statistical significance.

4.1.1 Excited quark analysis

In this analysis, the resonance signals considered are the excited states of light $q = u, d$ and heavy b quarks. The underlying processes are $pp \rightarrow q^* \rightarrow \gamma + \text{jet}$ and $pp \rightarrow b^* \rightarrow \gamma + b\text{-jet}$. The invariant mass distribution spectra of $\gamma + \text{jet}$ and $\gamma + b\text{-jet}$ are the most important distributions for this study, which have been analyzed to investigate the presence or absence of excited quark signals. A detailed overview of the signals and backgrounds of this study is presented in sections below.

4.1.1.1 Signals

This analysis considers the q^* and b^* signals decaying into $\gamma + \text{jet}$ and $\gamma + b\text{-jet}$ final states respectively, where the jet can be a light quark, a gluon or a b -quark. The Feynman diagrams representing the various physics processes contributing to this channel are presented in Fig. 1.4. Among these, the most dominant quark-gluon fusion process has been considered for this analysis. This process contributes through two Feynman diagrams in s- and t-channel as shown in Fig. 4.1.

The resonance will show its presence in the form of a bump in the s-channel and in the form of an excess of events in t-channel, over the continuum invariant mass distribution of $\gamma + \text{jet}$.

4.1.1.2 Backgrounds

In pp collisions, the $\gamma + \text{jet}$ channel has a rather clean and simple signature within the detector. Still, many known SM processes can mimic this final state

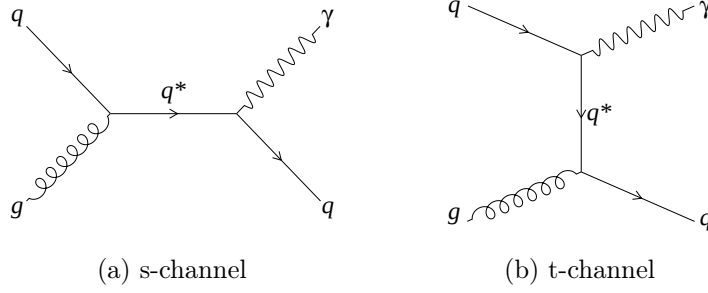


Figure 4.1: Feynman diagrams for quark-gluon fusion process in s and t channel.

and form the backgrounds of this study. These potential physics backgrounds are reported below as per their dominance:

4.1.1.2.1 SM γ + jet

The SM γ + jet processes where a photon and a jet are produced in the hard collision, form the most dominant and irreducible background for this study. A number of processes including quark-gluon compton scattering ($qg \rightarrow q\gamma$), quark-antiquark annihilation ($q\bar{q} \rightarrow g\gamma$) and gluon-gluon fusion ($gg \rightarrow g\gamma$) contribute to this background. The Feynman diagrams representing these processes are shown in Fig. 4.2. The qg scattering process dominates the total SM γ + jet cross-section over the entire p_T range. The contribution from the $q\bar{q}$ annihilation process is lesser, though it increases with increasing p_T . The gg fusion process, being a NLO process, contributes negligibly.

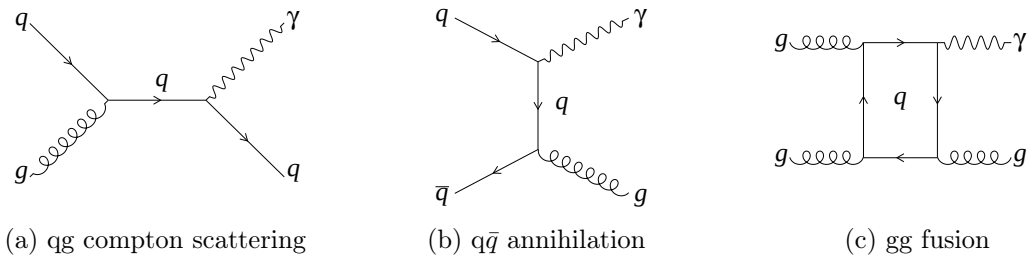


Figure 4.2: Standard model background contribution from γ + jet processes.

4.1.1.2.2 QCD dijet

The SM dijet processes where two jets are produced in the hard scattering, can form a background to a γ + jet process if one of the jets fragments into a highly energetic π^0 which then decays into a pair of overlapping photons and gets reconstructed as a single photon in the detector. A number of processes like the

ones shown in Fig. 4.3, contribute to this background. The production cross-section of these processes is around 10^4 times larger than the SM $\gamma + \text{jet}$ processes. But, since the probability of a jet faking a photon is about $10^{-4} - 10^{-3}$, this background form the second largest background of this study. Also, the cross-section for these processes falls very swiftly ($\sim p_T^{-4}$) with increasing p_T , so this background has significantly lower contribution at high p_T .

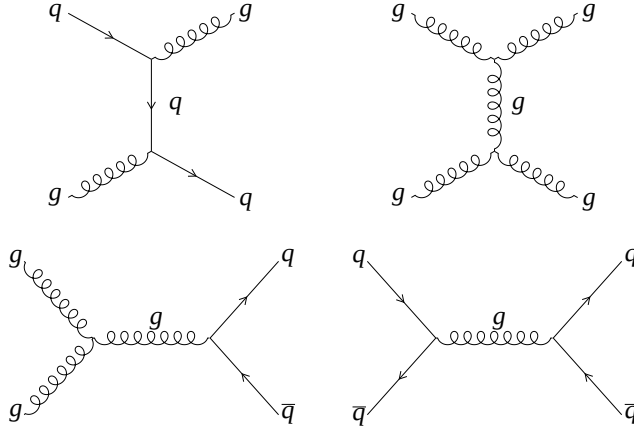


Figure 4.3: Standard model background contribution from QCD dijet processes.

This background is also supplemented by the bremsstrahlung process where a photon is radiated by one of the initial or final state quarks as shown in Fig. 4.4. This results into one photon and two jets in the final state and can mimic a $\gamma + \text{jet}$ process if one of the jets is lost or mismeasured. However, since these photons do not emerge from primary interaction vertex and have hadronic activity in close vicinity, these can be easily removed by the photon isolation requirements.

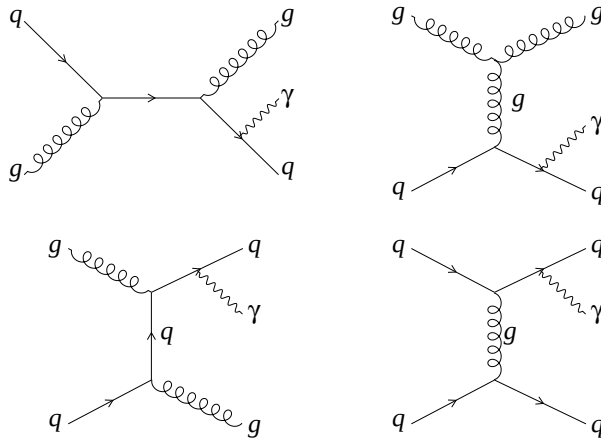


Figure 4.4: Bremsstrahlung contribution to SM QCD dijet production.

4.1.1.2.3 EWK

The SM electroweak processes, like $W/Z + \text{jet}$ production, can contribute a small fraction to the total background when the heavy electroweak boson decays leptonically and one of the leptons is either lost or goes away in the form of missing transverse energy. The Feynman diagrams corresponding to these processes are shown in Fig. 4.5

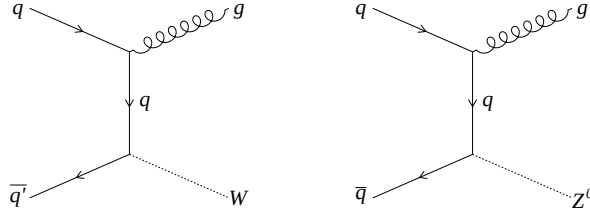


Figure 4.5: Background contributions from electroweak processes.

4.2 Production of Ntuples

The analysis chain starts with the production of Ntuples² of required data and MC samples. This step makes use of an analyzer, a code written mostly in C++ and python within the CMSSW environment, that extracts the information, related to different physics objects, triggers, vertices, pile-up etc., from the reconstructed or simulated datasets and store them in ROOT files in the form of Ntuples. Since the analyzer has to run over billions of events contained within a dataset, this step is almost impossible to perform using interactive computing. A technique known as grid computing [128], that considers a number of computer resources located at multiple places all over the world, is used along with CMS Remote Analysis Builder (CRAB)³ [173] to achieve this task. The Ntuples consist of organized information of all the quantities required for further analysis. This study has used the Tier-3 center present at Fermilab, USA, referred to as T3_US_FNALLPC, for the production, storage, and analysis of data and MC Ntuples.

4.2.1 Samples processed

The recorded detector data and simulated MC samples used in this analysis are detailed below. The MC samples include the signal as well as the background

²Analysis specific event information extracted from the complete dataset and stored in ROOT files in the form of tree/branch structure.

³An official CMS tool that allows an end user to access the data distributed at various sites as well as automates the task of job submission.

samples. All the samples are used in mini-AOD format and have been processed using a CMSSW version 8_0_26_patch1.

4.2.1.1 Data

This analysis used the pp collision data collected by the CMS detector at $\sqrt{s} = 13$ TeV in 2016, with a bunch crossing frequency of 25 ns. This data corresponds to a total integrated luminosity of 35.9 fb^{-1} . The reconstructed data in mini-AOD format, with only certified runs and luminosity sections selected using an official JSON⁴ file, have been used. The list of data samples split in different run era ranging from B to H and the name of JSON file are presented in Table 4.1.

Data Samples	Run Range	Luminosity (fb^{-1})
/SinglePhoton/Run2016B-03Feb2017_ver2-v2/MINIAOD	273150 – 275376	5.788
/SinglePhoton/Run2016C-03Feb2017-v1/MINIAOD	275656 – 276283	2.573
/SinglePhoton/Run2016D-03Feb2017-v1/MINIAOD	276315 – 276811	4.248
/SinglePhoton/Run2016E-03Feb2017-v1/MINIAOD	276831 – 277420	4.009
/SinglePhoton/Run2016F-03Feb2017-v1/MINIAOD	277932 – 278808	3.102
/SinglePhoton/Run2016G-03Feb2017-v1/MINIAOD	278820 – 280385	7.540
/SinglePhoton/Run2016H-03Feb2017_ver2-v1/MINIAOD	281613 – 284035	8.391
/SinglePhoton/Run2016H-03Feb2017_ver3-v1/MINIAOD	284036 – 284044	0.215
Total	273150 – 284044	35.866
JSON File		
Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt		

Table 4.1: List of data samples along with corresponding run ranges and integrated luminosity. The JSON file used to select certified runs and luminosity sections is also listed.

4.2.1.2 Signal MC

The signal MC samples for $q^* \rightarrow \gamma + \text{jet}$ and $b^* \rightarrow \gamma + b\text{-jet}$ are generated and simulated at the leading order (LO) using PYTHIA 8.212 event generator [162] and GEANT4 [172] detector simulator, for three different values of the coupling multiplier, viz. $f = 1.0, 0.5$ and 0.1 . These resonance signals are generated at different masses in the range from 0.5 to 9.0 TeV at an interval of 1 TeV for q^* and 0.5 to 5.0 TeV at an interval of 0.5 TeV for b^* , to explore the entire mass range. The simulation is performed using the CUETP8M1 underlying event tune [174, 175], with renormalization and factorization scale set to, $\mu = p_T$ of hard-scattered

⁴A JSON file is a special java format file that lists the good runs and luminosity sections collected with optimum detector conditions. The JSON files are located at the link: <https://cms-service-dqm.web.cern.ch/cms-service-dqm/CAF/certification/Collisions16/13TeV/ReReco/>

partons, and NNPDF2.3LO parton distribution function (PDF) [176]. The list of all the samples along with the number of generated events and cross-sections is presented in Table 4.2 for q^* and in Table 4.3 for b^* .

Sample Name	Events	Cross section (in pb)	Coupling
QstarToGJ_M-500_f-1p0_TuneCUETP8M1_13TeV-pythia8	299,760	3.033E+2	1.0
QstarToGJ_M-1000_f-1p0_TuneCUETP8M1_13TeV-pythia8	296,364	1.632E+1	1.0
QstarToGJ_M-2000_f-1p0_TuneCUETP8M1_13TeV-pythia8	99,499	5.213E-1	1.0
QstarToGJ_M-3000_f-1p0_TuneCUETP8M1_13TeV-pythia8	74,879	4.272E-2	1.0
QstarToGJ_M-4000_f-1p0_TuneCUETP8M1_13TeV-pythia8	73,482	4.8E-3	1.0
QstarToGJ_M-5000_f-1p0_TuneCUETP8M1_13TeV-pythia8	74,918	5.835E-4	1.0
QstarToGJ_M-6000_f-1p0_TuneCUETP8M1_13TeV-pythia8	75,000	7.076E-5	1.0
QstarToGJ_M-7000_f-1p0_TuneCUETP8M1_13TeV-pythia8	50,000	8.66E-6	1.0
QstarToGJ_M-8000_f-1p0_TuneCUETP8M1_13TeV-pythia8	49,814	1.283E-6	1.0
QstarToGJ_M-9000_f-1p0_TuneCUETP8M1_13TeV-pythia8	50,000	2.985E-7	1.0
QstarToGJ_M-500_f-0p5_TuneCUETP8M1_13TeV-pythia8	300,000	7.378E+1	0.5
QstarToGJ_M-1000_f-0p5_TuneCUETP8M1_13TeV-pythia8	99,723	4.129	0.5
QstarToGJ_M-2000_f-0p5_TuneCUETP8M1_13TeV-pythia8	100,000	1.328E-1	0.5
QstarToGJ_M-3000_f-0p5_TuneCUETP8M1_13TeV-pythia8	75,000	1.095E-2	0.5
QstarToGJ_M-4000_f-0p5_TuneCUETP8M1_13TeV-pythia8	72,816	1.212E-3	0.5
QstarToGJ_M-5000_f-0p5_TuneCUETP8M1_13TeV-pythia8	75,000	1.437E-4	0.5
QstarToGJ_M-6000_f-0p5_TuneCUETP8M1_13TeV-pythia8	74,776	1.62E-5	0.5
QstarToGJ_M-7000_f-0p5_TuneCUETP8M1_13TeV-pythia8	49,520	1.672E-6	0.5
QstarToGJ_M-8000_f-0p5_TuneCUETP8M1_13TeV-pythia8	49,427	1.647E-7	0.5
QstarToGJ_M-9000_f-0p5_TuneCUETP8M1_13TeV-pythia8	50,000	2.329E-8	0.5
QstarToGJ_M-500_f-0p1_TuneCUETP8M1_13TeV-pythia8	99,681	2.955	0.1
QstarToGJ_M-1000_f-0p1_TuneCUETP8M1_13TeV-pythia8	99,254	1.655E-1	0.1
QstarToGJ_M-2000_f-0p1_TuneCUETP8M1_13TeV-pythia8	75,000	5.315E-3	0.1
QstarToGJ_M-3000_f-0p1_TuneCUETP8M1_13TeV-pythia8	75,000	4.356E-4	0.1
QstarToGJ_M-4000_f-0p1_TuneCUETP8M1_13TeV-pythia8	72,312	4.861E-5	0.1
QstarToGJ_M-5000_f-0p1_TuneCUETP8M1_13TeV-pythia8	75,000	5.715E-6	0.1
QstarToGJ_M-6000_f-0p1_TuneCUETP8M1_13TeV-pythia8	49,942	6.241E-7	0.1
QstarToGJ_M-7000_f-0p1_TuneCUETP8M1_13TeV-pythia8	49,910	5.973E-8	0.1
QstarToGJ_M-8000_f-0p1_TuneCUETP8M1_13TeV-pythia8	49,650	4.515E-9	0.1
QstarToGJ_M-9000_f-0p1_TuneCUETP8M1_13TeV-pythia8	49,814	2.655E-10	0.1

Table 4.2: MC samples for q^* signal, simulated and reconstructed at $\sqrt{s} = 13$ TeV.

Sample Name	Events	Cross section (in pb)	Coupling
BstarToGJ_M-500_f-1p0_TuneCUETP8M1_13TeV-pythia8	100,000	6.236	1.0
BstarToGJ_M-1000_f-1p0_TuneCUETP8M1_13TeV-pythia8	99,736	2.148E-1	1.0
BstarToGJ_M-1500_f-1p0_TuneCUETP8M1_13TeV-pythia8	73,990	2.204E-2	1.0
BstarToGJ_M-2000_f-1p0_TuneCUETP8M1_13TeV-pythia8	75,000	3.585E-3	1.0
BstarToGJ_M-2500_f-1p0_TuneCUETP8M1_13TeV-pythia8	75,000	7.488E-4	1.0
BstarToGJ_M-3000_f-1p0_TuneCUETP8M1_13TeV-pythia8	73,170	1.766E-4	1.0
BstarToGJ_M-3500_f-1p0_TuneCUETP8M1_13TeV-pythia8	50,000	4.517E-5	1.0
BstarToGJ_M-4000_f-1p0_TuneCUETP8M1_13TeV-pythia8	48,100	1.202E-5	1.0
BstarToGJ_M-4500_f-1p0_TuneCUETP8M1_13TeV-pythia8	50,000	3.289E-6	1.0
BstarToGJ_M-5000_f-1p0_TuneCUETP8M1_13TeV-pythia8	49,608	9.216E-7	1.0
BstarToGJ_M-500_f-0p5_TuneCUETP8M1_13TeV-pythia8	99,544	1.574	0.5
BstarToGJ_M-1000_f-0p5_TuneCUETP8M1_13TeV-pythia8	100,000	5.438E-2	0.5
BstarToGJ_M-1500_f-0p5_TuneCUETP8M1_13TeV-pythia8	75,000	5.666E-3	0.5
BstarToGJ_M-2000_f-0p5_TuneCUETP8M1_13TeV-pythia8	74,404	9.162E-4	0.5
BstarToGJ_M-2500_f-0p5_TuneCUETP8M1_13TeV-pythia8	74,409	1.884E-4	0.5
BstarToGJ_M-3000_f-0p5_TuneCUETP8M1_13TeV-pythia8	72,730	4.393E-5	0.5
BstarToGJ_M-3500_f-0p5_TuneCUETP8M1_13TeV-pythia8	49,150	1.112E-5	0.5
BstarToGJ_M-4000_f-0p5_TuneCUETP8M1_13TeV-pythia8	50,000	2.909E-6	0.5
BstarToGJ_M-4500_f-0p5_TuneCUETP8M1_13TeV-pythia8	50,000	7.705E-7	0.5
BstarToGJ_M-5000_f-0p5_TuneCUETP8M1_13TeV-pythia8	50,000	2.027E-7	0.5
BstarToGJ_M-500_f-0p1_TuneCUETP8M1_13TeV-pythia8	99,033	6.357E-2	0.1
BstarToGJ_M-1000_f-0p1_TuneCUETP8M1_13TeV-pythia8	99,040	2.175E-3	0.1
BstarToGJ_M-1500_f-0p1_TuneCUETP8M1_13TeV-pythia8	75,000	2.278E-4	0.1
BstarToGJ_M-2000_f-0p1_TuneCUETP8M1_13TeV-pythia8	73,936	3.674E-5	0.1
BstarToGJ_M-2500_f-0p1_TuneCUETP8M1_13TeV-pythia8	75,000	7.595E-6	0.1
BstarToGJ_M-3000_f-0p1_TuneCUETP8M1_13TeV-pythia8	74,911	1.768E-6	0.1
BstarToGJ_M-3500_f-0p1_TuneCUETP8M1_13TeV-pythia8	50,000	4.454E-7	0.1
BstarToGJ_M-4000_f-0p1_TuneCUETP8M1_13TeV-pythia8	50,000	1.155E-7	0.1
BstarToGJ_M-4500_f-0p1_TuneCUETP8M1_13TeV-pythia8	46,590	3.031E-8	0.1
BstarToGJ_M-5000_f-0p1_TuneCUETP8M1_13TeV-pythia8	49,388	7.703E-9	0.1

Table 4.3: MC samples for b^* signal, simulated and reconstructed at $\sqrt{s} = 13$ TeV.

The signal shapes of q^* and b^* signals for different masses and different coupling multipliers, obtained by taking the probability distribution of the invariant mass of $\gamma + \text{jet}$ and $\gamma + b\text{-jet}$ respectively, are shown in Fig. 4.6.

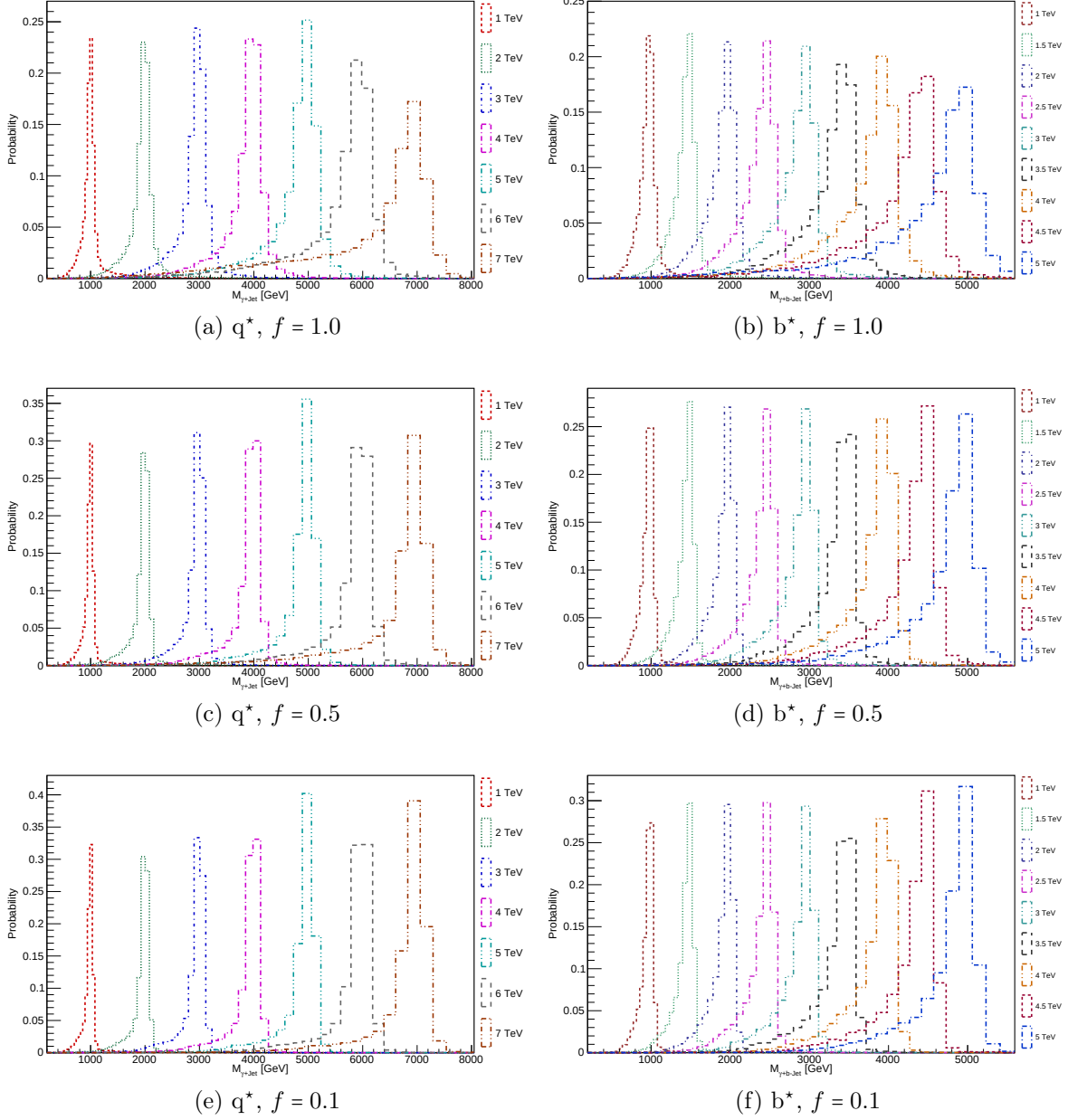


Figure 4.6: Signal shapes for q^* (left) and b^* (right) for different masses and couplings.

4.2.1.3 Background MC

The list of MC samples, along with the number of generated events and process cross-section, corresponding to the three backgrounds for this study are presented in Table 4.4 below.

Sample Name	Events	Cross section (in pb)
GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	4,467,985	20820
GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	5,131,873	9201
GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	10,036,487	2308
GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2,529,729	275.2
GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2,463,946	93.3
QCD_Pt_120to170_TuneCUETP8M1_13TeV_pythia8	6,708,572	471100
QCD_Pt_170to300_TuneCUETP8M1_13TeV_pythia8	6,958,708	117276
QCD_Pt_300to470_TuneCUETP8M1_13TeV_pythia8	4,150,588	7823
QCD_Pt_470to600_TuneCUETP8M1_13TeV_pythia8	3,959,986	648.2
QCD_Pt_600to800_TuneCUETP8M1_13TeV_pythia8	3,896,412	186.9
QCD_Pt_800to1000_TuneCUETP8M1_13TeV_pythia8	3,992,112	32.29
QCD_Pt_1000to1400_TuneCUETP8M1_13TeV_pythia8	2,999,069	9.418
QCD_Pt_1400to1800_TuneCUETP8M1_13TeV_pythia8	396,409	8.426E-1
QCD_Pt_1800to2400_TuneCUETP8M1_13TeV_pythia8	397,660	1.149E-1
QCD_Pt_2400to3200_TuneCUETP8M1_13TeV_pythia8	399,226	6.830E-3
QCD_Pt_3200toInf_TuneCUETP8M1_13TeV_pythia8	391,735	1.654E-4
DYJetsToLL_Pt-100To250_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	2,040,596	83.12
DYJetsToLL_Pt-250To400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	423,976	3.047
DYJetsToLL_Pt-400To650_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	432,056	3.921E-1
DYJetsToLL_Pt-650ToInf_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	430,691	3.636E-2
WJetsToLNu_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	29,705,748	61526.7

Table 4.4: The background MC samples used in this analysis, simulated and reconstructed at $\sqrt{s} = 13$ TeV.

The SM $\gamma + \text{jet}$ background has been generated using the MADGRAPH5_amc@NLO 2.2.2 [163] event generator in different H_T^5 bins, in the range of 40 – 100, 100 – 200, 200 – 400, 400 – 600 and 600 – ∞ GeV with showering and hadronization carried out by the PYTHIA program. The MLM [177] matching scheme has been used to remove the double counting of the partons generated with MADGRAPH and PYTHIA.

The SM QCD dijet background has been generated using PYTHIA, with the same conditions as signal MC, in different p_T bins, in the range of 120 – 170, 170 – 300, 300 – 470, 470 – 600, 600 – 800, 800 – 1000, 1000 – 1400, 1400 – 1800, 1800 – 2400, 2400 – 3200 and 3200 – ∞ GeV.

The EWK Z+jet background has been generated in different p_T bins in the range of 100 – 250, 250 – 400, 400 – 650 and 650 – ∞ GeV using the AMCATNLO event generator with FFX [177] matching scheme and PYTHIA for showering and hadronization. The corresponding W+jet background is generated using the MADGRAPH event generator with MLM matching scheme and PYTHIA for showering and hadronization. All the samples use CUETP8M1 as the underlying event tune.

⁵ $H_T = \sum p_T$ of all the jets in an event

4.3 Re-weighting of MC events

The detector data is compared with the MC samples to validate it w.r.t. the theoretical description. In order to have a fair comparison, both data and MC are kept at equal footing, by weighting the MC events corresponding to the data.

4.3.1 Luminosity re-weighting

The MC events are re-weighted to reflect the data luminosity. The relevant luminosity weight per event applied to MC is:

$$w_{\text{lumi}} = \frac{\sigma_{\text{process}} \times \mathcal{L}_{\text{data}}}{N_{\text{process}}} \quad (4.1)$$

where σ_{process} and N_{process} are the theoretical cross-section and the number of generated events for the simulated process and $\mathcal{L}_{\text{data}}$ is the total integrated luminosity of data.

4.3.2 Pile-up re-weighting

In high-luminosity colliders, each bunch crossing may result into many inelastic interactions, with the number of interactions being proportional to the instantaneous luminosity of the collision. These additional interactions are attributed to as the pile-up (PU) interactions. These pile-up events can affect the reconstruction efficiency and can appear in the p_T distribution of the reconstructed objects. The presence of pile-up can affect the current analysis in two ways: additional energy from PU can be added to the reconstructed jet energy, thereby, worsening the jet energy measurement and tracks and calorimeter energy deposits from PU interactions can get added to the isolation energy sum of photons, thereby, making the isolation cuts less efficient. However, a number of techniques are developed and centrally validated in CMS to alleviate the deterioration of object reconstruction due to PU effects. To minimize the impact of PU interactions, various cuts applied for object selection and energy corrections are optimized so that in the end, the effect is almost negligible.

The simulated samples are also generated with additional pile-up to match with the actual data taking conditions. The samples used in this analysis are generated with 2015_25ns_Startup_PoissonOOTPU [178] pile-up scenario. Still, the residual differences in the pile-up distributions of data and MC are corrected by means of

pile-up re-weighting. In this process, the pileup distribution for data is taken from the instantaneous luminosity per bunch crossing for each luminosity section, and the total pp inelastic cross-section of 69.2 mb, using the officially recommended procedure [178]. A Poissonian smearing is used to model the statistical fluctuations in the actual pileup events present in the data. For MC, the true number of interactions per event are considered as the pileup distribution. These pile-up distributions in data and MC are compared corresponding to the number of primary vertices and a weight factor is calculated on the event by event basis. This weight factor is then multiplied to the luminosity weight factor to get the total event weight applied to the Monte Carlo events. To give an account of the effect, a data-MC comparison of the number of vertices in an event has been shown in Fig. 4.7, before and after applying the pile-up re-weighting.

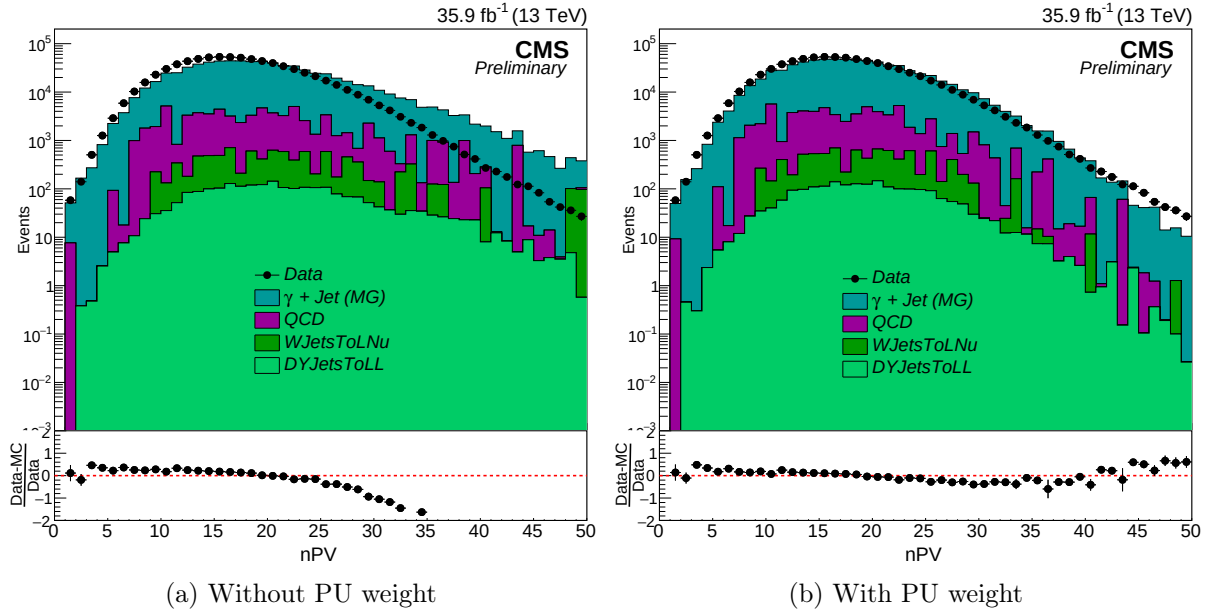


Figure 4.7: Data-MC comparison of the number of vertices in an event, showing the effect of pile-up re-weighting.

4.4 Event selection

In this analysis, we are searching for the possible signatures of excited quarks decaying into a photon and a jet final state. So the experimental technique of event selection starts with the measurement of the inclusive process, $pp \rightarrow \gamma + \text{jet} + X$ where we require at least one hard photon and one hard jet in the event, X refers to the additional photons or jets or anything else that can be present in

the event. The events containing these additional particles are not vetoed as it would unnecessarily restrict our signal to a narrow topology. The event statistics for the search of excited b-quark is formed by requiring the selected jet to be a b-jet. Based on the type of the jet, three different event categories are made:

- **Inclusive category:** Events containing photons and jets where jets can be the light jets or the b-jets. This forms the event statistics for q^* search.
- **Exclusive 1 b-tag category:** Events containing photons and b-jets, forming the event statistics for b^* search.
- **Exclusive 0 b-tag category:** Events containing photons and non b-jets. This category is also used in b^* search in order to overcome any loss of statistical power due to migration of events from one category to another.

The selection criteria used in this analysis is described in detail in the sections below:

4.4.1 Trigger selection

The trigger system of the CMS experiment is designed to select interesting physics events of good quality. As mentioned in Section 2.2.2, it consists of two levels: the L1 trigger and the HLT trigger. The data sample used in this analysis is required to pass the L1_SingleEG40 trigger, which selects the events that have at least one e/γ candidate with $E_T > 40$ GeV and HLT_Photon165_HE10 trigger, which is a single photon trigger requiring the events to have at least one photon candidate with $p_T > 165$ GeV and H/E fraction $< 10\%$. The HLT path is required to be un-prescaled for the entire data taking period, so that all events passing the trigger are stored for further analysis.

The offline trigger efficiency of HLT trigger, also referred to as the trigger turn-on, has been studied as a function of the leading photon p_T in order to observe any inefficiencies. The trigger turn-on measured w.r.t. the same object dataset is referred to as the relative trigger efficiency while the trigger turn-on measured w.r.t. an orthogonal dataset is referred to as the absolute trigger efficiency, as the relative trigger efficiency can not reveal the presence of any systematic detector effect. The idea is to identify unbiased photon candidates in data and measure the frequency of photons activating the trigger decision. The trigger efficiency in this analysis has been evaluated w.r.t. the lower p_T prescaled photon, jet and muon

reference triggers, as listed in Table 4.5. This efficiency is measured by selecting the events passing the lower p_T threshold triggers and computing the rate at which these events also pass the main analysis trigger. The trigger turn-on curve for all the three reference triggers is presented in Fig. 4.8.

Analysis trigger	Photon ref. triggers	Jet ref. triggers	Muon ref. triggers
HLT_Photon165_HE10	HLT_Photon75	HLT_PFJet60	HLT_Mu50
	HLT_Photon90	HLT_PFJet80	
	HLT_Photon120	HLT_PFJet140	

Table 4.5: Different triggers used in the trigger efficiency study.

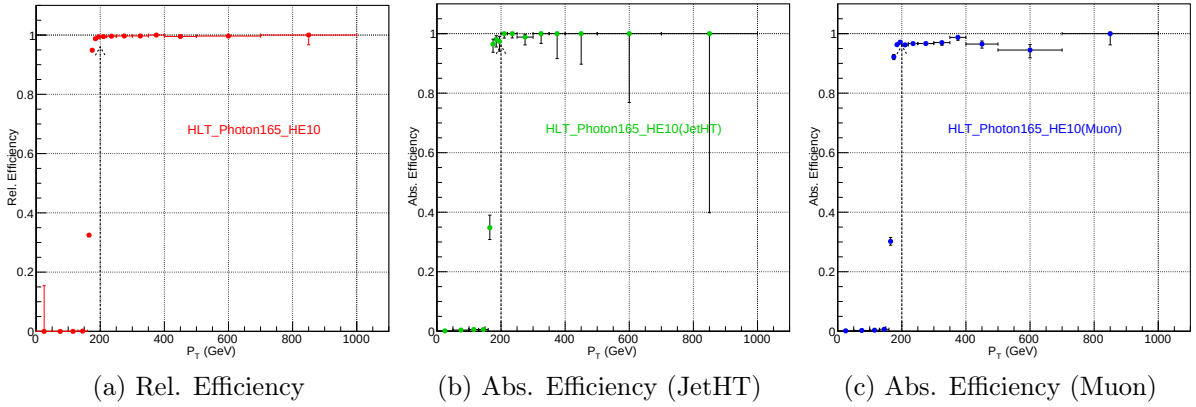


Figure 4.8: Relative and absolute trigger efficiencies for the single photon trigger, HLT_Photon165_HE10.

It can be seen that the trigger become fully efficient and reaches the plateau region at around 200 GeV, for photon and jet reference triggers. However, the turn-on w.r.t. the muon reference trigger is found to be around 95% efficient with a sizable uncertainty which is taken into account as a systematic uncertainty.

4.4.2 MET filters

While monitoring the quality of detector data during offline processing, a number of noisy events were reported, which are removed by applying a set of officially recommended noise cleaning filters, listed below:

- **CSC tight beam halo filter:** rejects muons from beam halo.
- **HBHE noise filter with isolated noise rejection:** rejects noisy events in HCAL.

- **HCAL laser filter:** removes events that contains HCAL calibration laser firing within the collision bunch-crossing.
- **ECAL dead cell trigger primitive filter:** rejects events containing large missing energy coming from masked crystals.
- **Bad EE supercrystal filter:** removes events that contains two EE crystals with anomalously high energies.

4.4.3 Primary vertex selection

A typical pp collision event may contain more than one reconstructed primary vertex coming from in-time and out-of-time bunch crossings, pile-up events etc. But, the events that have at least one primary vertex coming from in-time collision are kept for further study. These events are selected by requiring the vertex to be within 24 cm of the nominal collision point along the z-direction and within 2 cm in the radial direction. The vertex should have at least 4 associated tracks, along with the largest value of $\sum p_T^2$ of all the tracks and have a proper 3d vertex fit with the tracks.

4.4.4 Analysis object selection: Photon

In each event passing the primary vertex selection, we search for a well identified and isolated photon candidate by using the criteria explained in Section 3.1.3.2. The cut-off value used for each variable is listed below:

- $\sigma_{i\eta i\eta} < 0.01022$.
- Single tower H/E < 0.0396 .
- Conversion safe electron veto = true.
- PF charged hadron isolation < 0.441 .
- PF neutral hadron isolation $< 2.725 + 0.0148 \times p_T^\gamma + 0.000017 \times (p_T^\gamma)^2$.
- PF photon isolation $< 2.571 + 0.0047 \times p_T^\gamma$.

The photon isolation cone is susceptible to a diffuse background of soft particles coming from the underlying events or pile-up events. So the isolation variables

are corrected for the presence of these additional interactions using the following relation:

$$\text{Iso}^{\text{corrected}} = \max.(\text{Iso}^{\text{original}} - \rho_{\text{event}} \times A_{\text{eff}}, 0.0) \quad (4.2)$$

where, ρ_{event} is known as the energy density computed using the FastJet package [179] and is defined as the median energy density per unit area corresponding to the particles associated with the pile-up vertices. This quantity gives a measure of the pile-up activity in the event. The variable, A_{eff} is the ratio of the slopes obtained from linear fitting of $\rho - N_{\text{vtx}}$ and Isolation- N_{vtx} . The values of A_{eff} for three isolations, in different η regions covering barrel and endcaps, are shown in Table 4.6.

η -bin	EA charged hadrons	EA neutral hadrons	EA photons
$ \eta < 1.0$	0.0360	0.0597	0.1210
$1.0 < \eta < 1.479$	0.0377	0.0807	0.1107
$1.479 < \eta < 2.0$	0.0306	0.0629	0.0699
$2.0 < \eta < 2.2$	0.0283	0.0197	0.1056
$2.2 < \eta < 2.3$	0.0254	0.0184	0.1457
$2.3 < \eta < 2.4$	0.0217	0.0284	0.1719
$ \eta > 2.4$	0.0167	0.0591	0.1998

Table 4.6: Effective area for different photon isolations.

In order to avoid non-collision backgrounds coming from the anomalous calorimetric signals resulting due to the direct ionization of Avalanche Photo-diodes (APDs) by a highly ionizing particle such as a proton and depositing all the energy within a single ECAL crystal (known as “spikes”), a minimal requirement on the rectangular ratios and shower shape profiles, as mentioned below, are also placed:

- $R_9 < 1.0$.
- $\sigma_{i\eta i\eta} > 0.001$.
- $\sigma_{i\phi i\phi} > 0.001$.

This selection results into a collection of photons in each event, out of which, the highest p_T photon (known as leading photon) is selected and is required to be greater than 200 GeV, the p_T value at which the trigger turn-on curve becomes 100% efficient. The selected photon is also required to be in the central barrel region of the detector, that is, $|\eta| < 1.4442$. This requirement is placed since the q^*

or b^* resonances (if exists), produced via hard scattering, would decay mostly into back-to-back γ and jet and are expected to be detected in the central region of the detector. A comparison of Data vs. MC for different photon variables, after the final selection, is presented in Fig. 4.9.

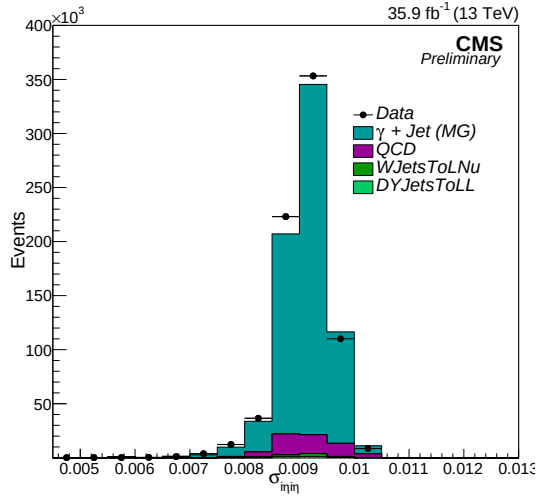
4.4.4.1 Overlap removal between γ + jet and QCD dijet background

The QCD sample used in this analysis has been generated using PYTHIA with the process: HardQCD::all = ON, that is, at the generator level, all the processes like $gg \rightarrow gg$, $gg \rightarrow q\bar{q}$, $qg \rightarrow qg$, $qq \rightarrow qq$, $q\bar{q} \rightarrow gg$, $q\bar{q} \rightarrow q\bar{q}$ are allowed and are associated with initial and final state parton showers. Since quarks tend to hadronize as soon as they are produced, the particles in parton showers appear to come from the primary vertex. Thus, the photons present in these parton showers form the source of prompt photons in the QCD sample and can mimic a γ + jet event.

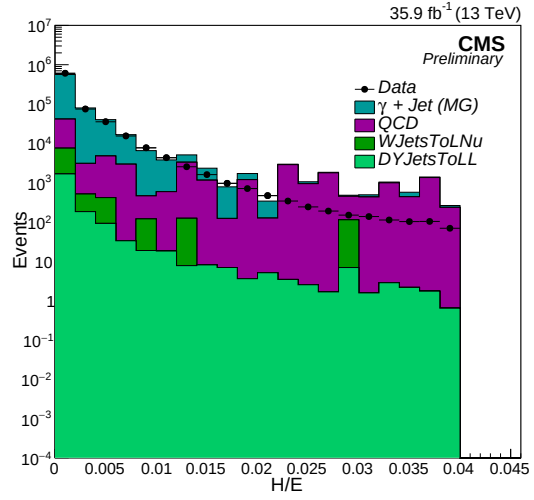
However, in QCD, since the prompt photons come from the parton showers, they have a small ΔR separation from other gen particles. The same is visible in Fig. 4.10a which shows the ΔR distribution between genPhotons and genPartons in the QCD sample. On the other hand, in γ + jet sample, the prompt photons come from the hard process and hence are largely separated from other particles as can be seen in Fig. 4.10b.

The events containing these prompt photons in QCD sample are removed in order to avoid double counting of events between γ + jet and QCD dijet samples. The procedure followed to remove the overlap is summarized as follows:

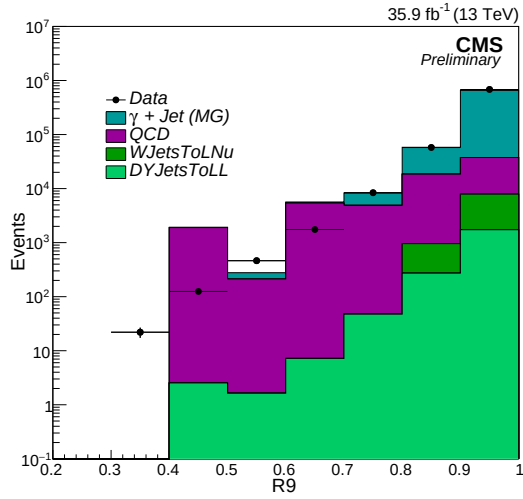
- In a QCD event, a gen photon corresponding to the selected reco photon is searched by requiring, gen particle ID = 22, $\Delta R(\text{Gen}, \text{Reco}) < 0.1$ and $\frac{\Delta p_T}{p_T}(\text{Gen}, \text{Reco}) < 0.1$. Once the gen photon is found, its prompt status is checked using its gen level properties. If it is prompt, the event is removed from the event collection, otherwise not.
- Since γ + jet MC sample has a gen level cut of $\Delta R(\gamma, \text{Parton}) > 0.05$, so the events with only $\Delta R(\gamma, \text{Parton}) > 0.05$ are removed from the QCD sample.
- The events with $\Delta R(\gamma, \text{Parton}) < 0.05$ (if any) are, then, removed from the γ + jet MC sample.



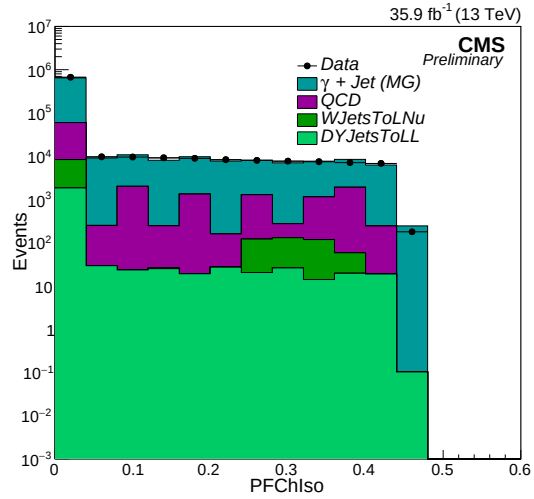
(a) $\sigma_{i\eta i\eta}$



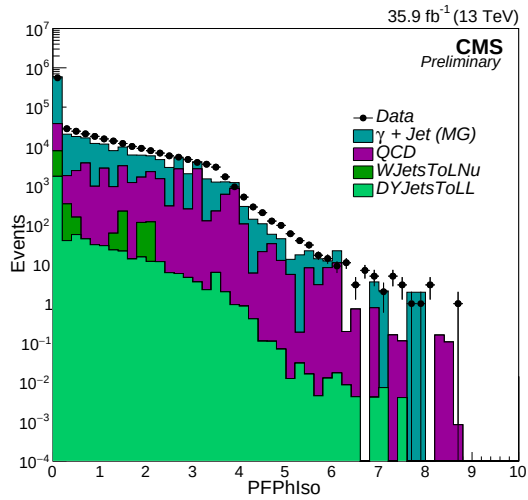
(b) Single Tower H/E



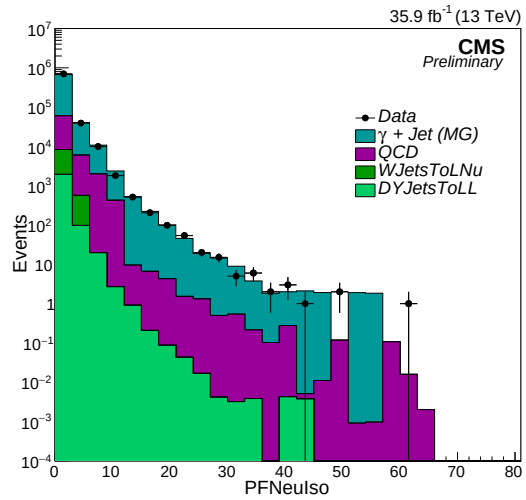
(c) R9



(d) PF Charged Hadron Isolation



(e) PF Photon Isolation



(f) PF Neutral Hadron Isolation

Figure 4.9: Data-MC comparison of photon identification and isolation variables after final selection.

The Fig. 4.10c shows the stitching of the two samples in ΔR after overlap removal. It can be seen that the stitching behaves nicely at the ΔR cut.

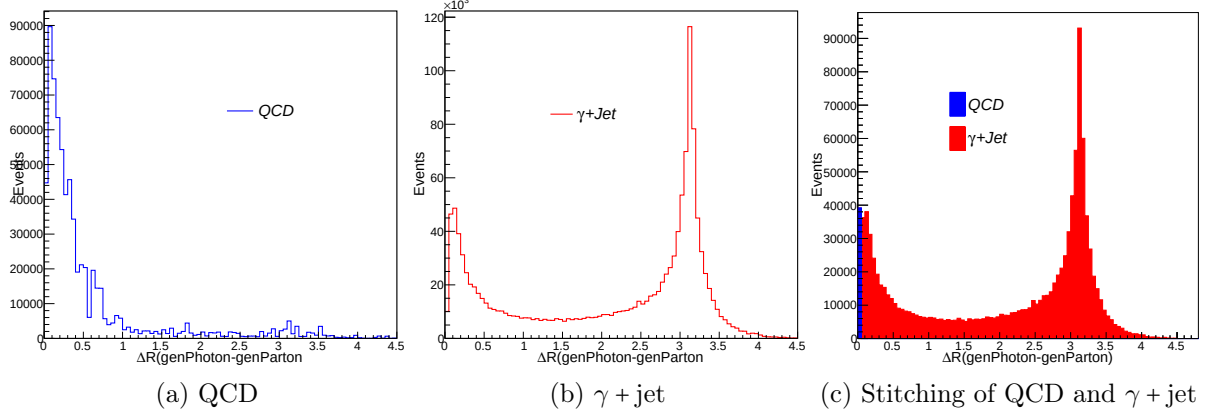


Figure 4.10: ΔR distribution between prompt gen photon matched to selected reco photon and other gen particles: (a) QCD before overlap removal, (b) $\gamma + \text{jet}$ before overlap removal, (c) Stitching of QCD and $\gamma + \text{jet}$ after overlap removal.

4.4.5 Analysis object selection: Jet

Events containing a good photon candidate are further inspected for the presence of an AK4 jet using the identification criteria mentioned in Section 3.1.4.4. The cut-off values used for different variables is listed below. The jets qualifying this selection are required to be away from the selected photon by $\Delta R > 0.5$, to ensure that they are different objects and not the fragments of the same particle.

- Charged hadron fraction > 0 .
- Neutral hadron fraction < 0.90 .
- Charged EM fraction < 0.99 .
- Neutral EM fraction < 0.90 .
- Number of constituents > 1 .
- Charged multiplicity > 0 .

Out of the selected jet collection, the jet with highest p_T is considered and is required to have a $p_T > 170 \text{ GeV}$ and $|\eta| < 2.4$. A nearly symmetric cut on the p_T of photon and jet is considered since the photons and jets in the signal samples are distributed symmetrically in p_T as can be seen for q^* samples in Fig. 4.11.

In order to overcome the non-linear response of the detector, L1, L2, L3 and L2L3 corrections as mentioned in Section 3.1.4.2 are applied. A comparison of Data and MC for different jet variables after final selection is shown in Fig. 4.12.

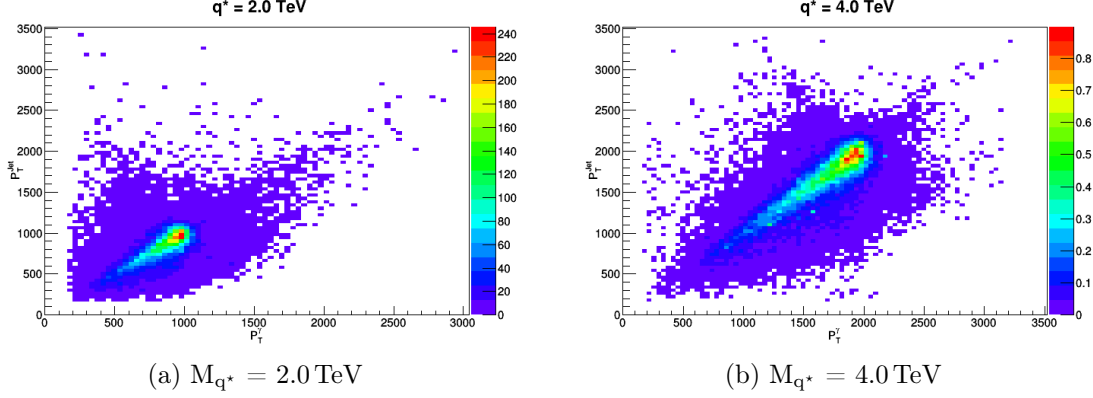


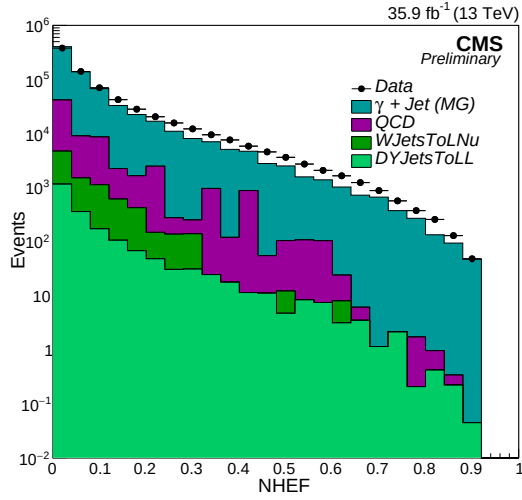
Figure 4.11: p_T^γ vs. p_T^{jet} in 2-dimensional representation for q^* signals.

4.4.6 Analysis object selection: b-Jet

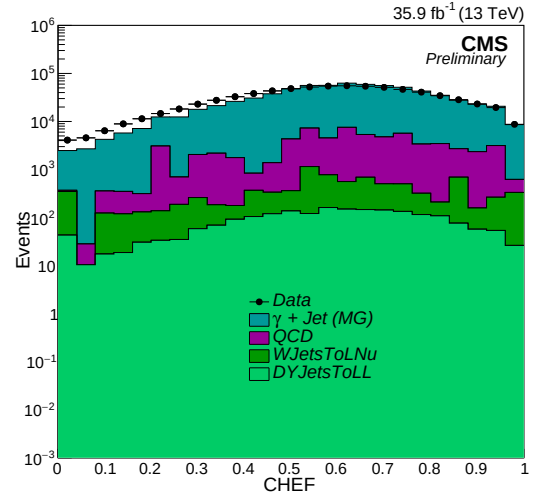
The selected jet in each event is examined for its b-tagging status using the “CSVv2” (explained in Section 3.1.4.3) discriminator. A cut-off value of 0.5426, corresponding to the loose working point, has been used and all the jets with b-tag discriminator value greater than the cut-off value are considered as b-jets.

In general, the b-jet identification efficiency and c /light-jet misidentification probability of a tagger is not the same in data and MC. These observed differences are taken into account by means of the b-tag scale factors (SF) defined as, $\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{MC}}}$, where ϵ_{data} and ϵ_{MC} are the b-tagging efficiencies of CSVv2 discriminator in data and MC simulation, respectively. These SFs, in CMS, are determined using the dedicated techniques, described in [156, 157]. The SFs are dependent on jet flavor, jet p_T and jet η . The official recommendations for computing these scale factors and errors on them are provided in [180]. The SFs are available for the p_T range of 20–1000 GeV and $|\eta| < 2.4$. For values above the p_T limit, the recommendations are to use the SF values at the limits with double the uncertainty while for values above the η limits, the SF values are taken equal to zero.

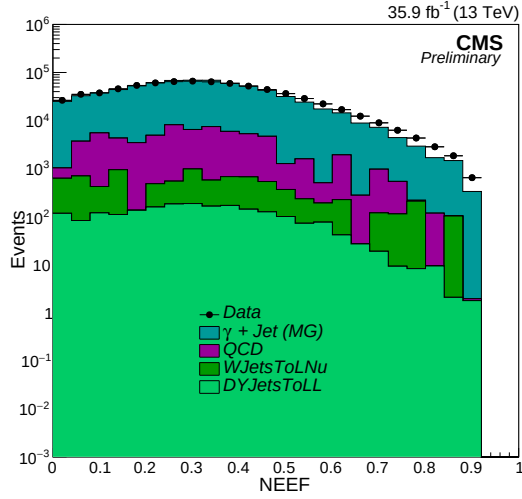
Errors on b-tag scale factors arise from various systematic uncertainties, like the jet energy scale uncertainty, the sample purity which is the contamination of jets with a flavor complementary to the one for which the scale factor is being measured, the uncertainty on the scale factor for c -jets and the uncertainty due to



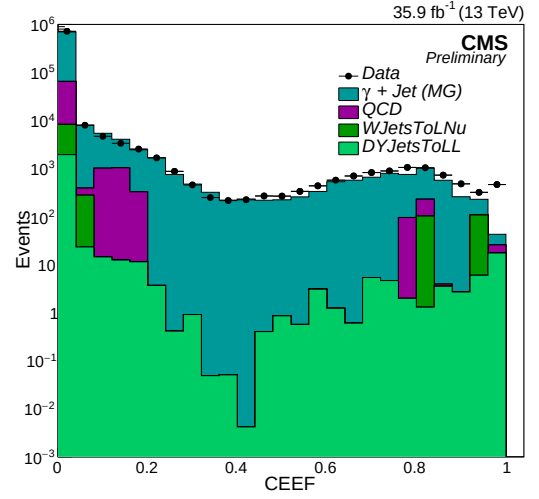
(a) Neutral Hadron Energy Fraction



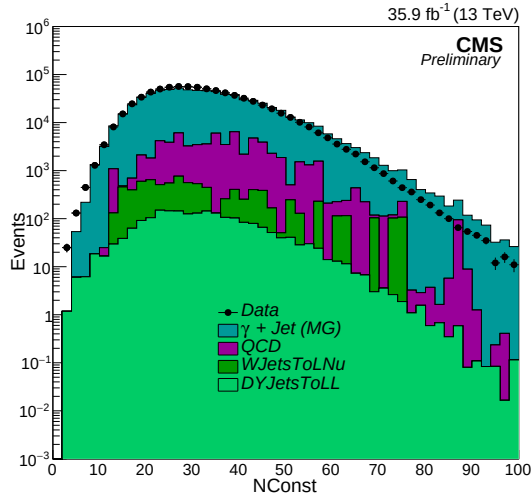
(b) Charged Hadron Energy Fraction



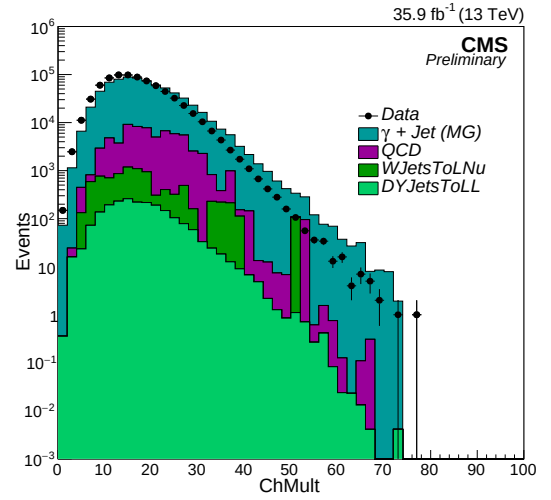
(c) Neutral EM Energy Fraction



(d) Charged EM Energy Fraction



(e) Number of Constituents



(f) Charged Multiplicity

Figure 4.12: Data-MC comparison of jet particle flow identification variables after final selection.

finite number of events in the simulation which results into statistical fluctuations etc. These errors are evaluated as the differences in the scale factors for central and up/down systematic type as mentioned in [180] and are used as a source of systematic uncertainty due to the b-tagging.

In principle, the SF value provides the fraction of actual b-jets among the jets passing the discriminator cut. Based on this terminology, two b-tag categories are formed:

- 1 b-tag category = Jets passing the discriminator $\times$ SF.
- 0 b-tag category = Jets passing the discriminator $\times$ (1 - SF) + Jets failing the discriminator.

So the 1 b-tag category essentially gives the fraction of b-jets and the 0 b-tag category gives the fraction of non b-jets, called as the 1 and 0 b-tag efficiencies of discriminator. A Data-MC comparison of the CSVv2 discriminator before and after applying the b-tagging is shown in Fig. 4.13. The 0 b-tag category may contain the jets passing the discriminator, with a weight (1 - SF), as is seen in Fig. 4.13b.

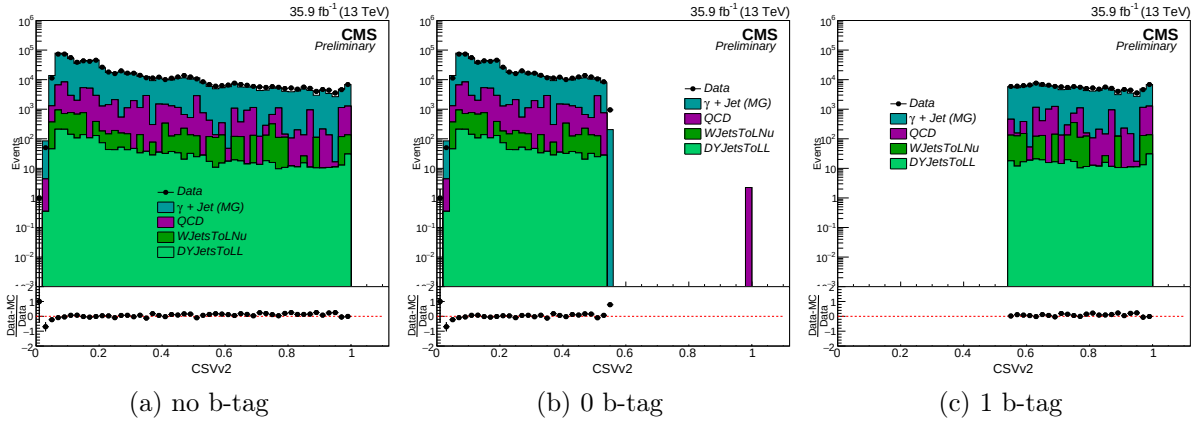


Figure 4.13: Data-MC comparison of CSVv2 discriminator distribution before and after applying the b-tagging.

4.4.7 Kinematical restrictions between γ and jet/b-jet

The γ + jet final state from excited quark decays would be produced through a s-channel process and has an isotropic distribution of final state particles. On the other hand, the γ + jet final state from background processes would take place through a t-channel process and result into high density of particles in

forward and backward directions. Hence, to reduce these background contributions while maintaining high acceptance for the signal, $|\Delta\eta(\gamma, \text{jet/b-jet})| < 1.5$ has been applied. Also, in order to avoid the turn-on region due to various kinematical selections, as shown in Fig. 4.14, a cut on the invariant mass of $\gamma + \text{jet/b-jet}$, $M_{\gamma+\text{jet}} = \sqrt{(E^\gamma + E^{\text{jet}})^2 - (\vec{p}^\gamma + \vec{p}^{\text{jet}})^2}$, has been applied and is required to be greater than 700 GeV.

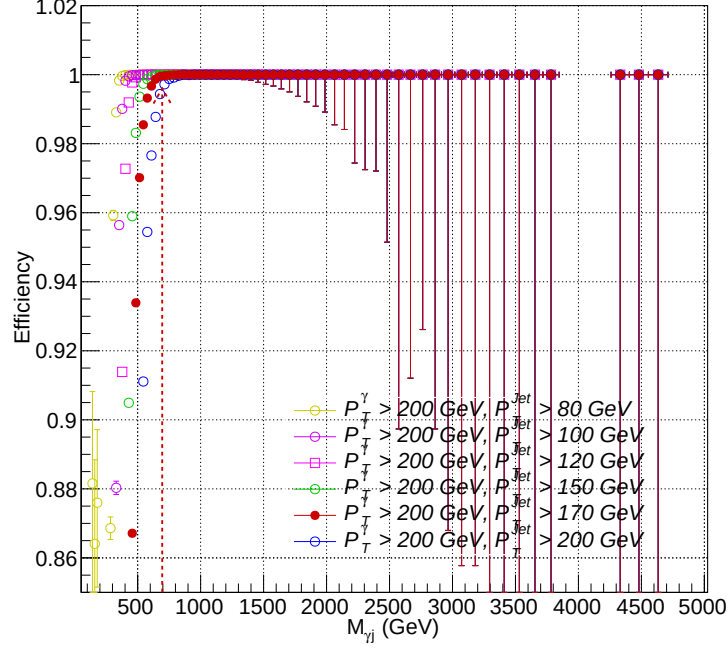


Figure 4.14: Turn-on due to various kinematical selections, plotted for different jet p_T cuts.

The theoretical predictions for $\gamma + \text{jet}$ and dijet backgrounds are generated at the LO level using PYTHIA and MADGRAPH, so in order to include the NLO effects, the MC events are scaled with a K-factor of 1.3 [181, 182], along with the luminosity and pile-up scale factors. The cut flow table for signal, backgrounds and data, after application of all selection cuts and scale factors, is presented in Table 4.7. The SM $\gamma + \text{jet}$ contributes about 95% of the total background while the contribution from QCD dijet background is around 4% and from electroweak background is around 1%.

The signal efficiencies for q^* and b^* signals at different mass points for $f = 1.0$ are presented in Table 4.8 and Table 4.9 respectively. The variation of q^* and b^* selection efficiencies as a function of mass is presented in Fig. 4.15.

Cuts	q* 2 TeV	SM γ + jet Bkg	QCD dijet Bkg	EWK Bkg	Total Bkg	Data
Total	18389.6	1.50357E+09	2.66074E+10	2.87352E+09	3.09845E+10	1.1584E+08
HLT	18389.6	1.50357E+09	2.66074E+10	2.87352E+09	3.09845E+10	3.93895E+07
MET filters	18389.6	1.50357E+09	2.66074E+10	2.87352E+09	3.09845E+10	3.74467E+07
Photon ID	12868.4	6.38738E+08	6.70708E+07	7.31732E+07	7.78982E+08	4.20307E+06
$p_T^\gamma > 200$ GeV & $ \eta^\gamma < 1.4442$	11552.4	1.06063E+06	481691	9231.4	1.55155E+06	1.26456E+06
QCD overlap removal	11552.4	1.06007E+06	91298.3	9231.4	1.1606E+06	1.26456E+06
Jet ID	11548.1	1.05854E+06	91244.5	8953.46	1.15874E+06	1.26033E+06
$p_T^{\text{jet}} > 170$ GeV	11496.6	814187	74359.6	6393.9	894940	930191
$ \eta^{\text{jet}} < 2.4$	11193	782291	62086.8	6179.82	850558	905346
$ \Delta\eta < 1.5$	8270.48	652420	37343.4	5201.72	694965	748663
$M_{\gamma+\text{jet}} > 700$ GeV	8237.03	122831	4950.45	1379.37	129160	126770

Table 4.7: Selection cut flow table for signal, different backgrounds and data.

Cuts	1 TeV	2 TeV	3 TeV	4 TeV	5 TeV	6 TeV	7 TeV
Total	100	100	100	100	100	100	100
Photon ID	69.4811	69.3168	69.696	70.0057	70.2649	69.0685	67.0689
$p_T^\gamma > 200$ GeV & $ \eta^\gamma < 1.4442$	54.6384	62.2149	63.6378	64.1612	64.5831	63.3158	61.0851
Jet ID	54.5833	62.1878	63.6138	64.149	64.5657	63.3051	61.0571
$p_T^{\text{jet}} > 170$ GeV	53.1439	61.9043	63.5149	64.0551	64.4989	63.2264	60.977
$ \eta^{\text{jet}} < 2.4$	50.0591	60.2276	62.7361	63.6412	64.2839	63.061	60.7929
$ \Delta\eta < 1.5$	39.2106	44.5	45.5168	45.5196	45.7801	44.6649	43.2637
$M_{\gamma+\text{jet}} > 700$ GeV	37.4718	44.317	45.4661	45.472	45.7347	44.6102	43.2077

Table 4.8: Selection efficiency (%) table for the q^* signal samples in the mass range 1.0–7.0 TeV with SM Coupling $f = 1.0$.

Cuts	1 TeV	1.5 TeV	2 TeV	2.5 TeV	3 TeV	3.5 TeV	4 TeV
Total	100	100	100	100	100	100	100
Photon ID	69.8006	69.6057	69.8508	70.3567	70.4957	71.1226	70.9883
$p_T^\gamma > 200$ GeV & $ \eta^\gamma < 1.4442$	54.6467	58.3353	60.0603	61.3592	62.1837	62.5457	63.0047
Jet ID	54.6361	58.3307	60.0561	61.3549	62.1784	62.5411	62.9972
$p_T^{\text{jet}} > 170$ GeV	53.1637	57.8604	59.8709	61.2866	62.1325	62.4845	62.9699
$ \eta^{\text{jet}} < 2.4$	52.5537	57.3699	59.4802	61.0701	61.952	62.4065	62.9059
CSVv2	35.5183	35.6018	34.2311	30.9516	28.901	28.1197	27.3777
$ \Delta\eta < 1.5$	28.4642	26.5476	25.3583	22.5024	20.4847	19.8818	19.5085
$M_{\gamma+\text{jet}} > 700$ GeV	27.3834	26.4028	25.3119	22.4633	20.4636	19.8627	19.4598

Table 4.9: Selection efficiency (%) table for the b^* signal samples in the mass range 1.0–4.0 TeV with SM Coupling $f = 1.0$.

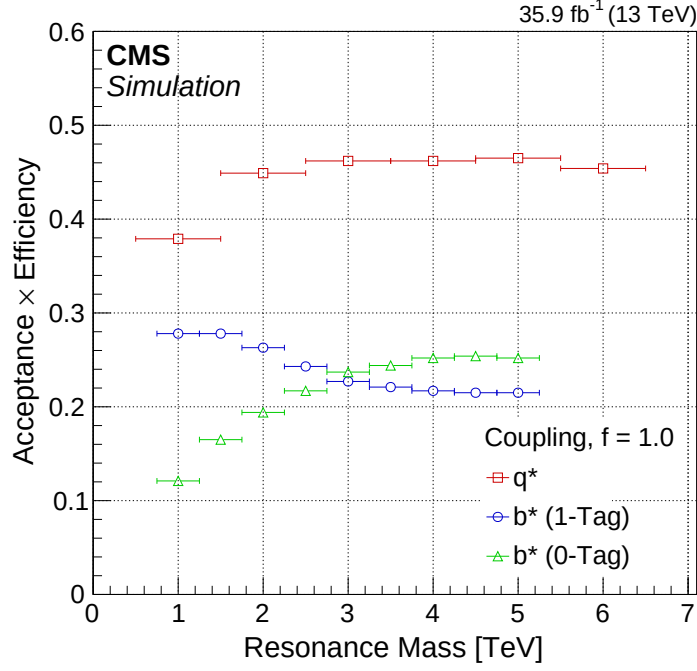


Figure 4.15: The product of acceptance and efficiency as a function of q^* and b^* mass for SM coupling, $f = 1.0$, obtained using MC simulation.

A comparison between Data and MC for different kinematical variables for photons and jets, after final selection, corresponding to the three categories viz. no b-tag, 1 b-tag and 0 b-tag, are presented in Fig. 4.16, Fig. 4.17 and Fig. 4.18 respectively. These distributions are found to be in reasonable agreement with the theoretical predictions. The p_T distribution for data falls steeply with the increasing p_T and agrees well with MC within the uncertainties. Likewise, the η distribution is in good agreement with the shape anticipated by MC.

The distributions of photons and jets in $\Delta\eta$, $\Delta\phi$ and ΔR are also presented in Fig. 4.19. The number of photons, jets and b-jets per event are shown in Fig. 4.20. The Data-MC comparison of the invariant mass distributions of $\gamma + \text{jet}$ and $\gamma + \text{b-jet}$ in three categories is displayed in Fig. 4.21.

The various kinematical distributions depicted here confirm that the kinematics of analysis objects in detector data comply with the theoretical expectations. However, the theoretical MC simulations, in this analysis, are used just to check the behaviour of the data. The actual backgrounds for the processes under study, have been computed using a data driven technique by fitting the invariant mass distribution spectra with a smooth parameterization.

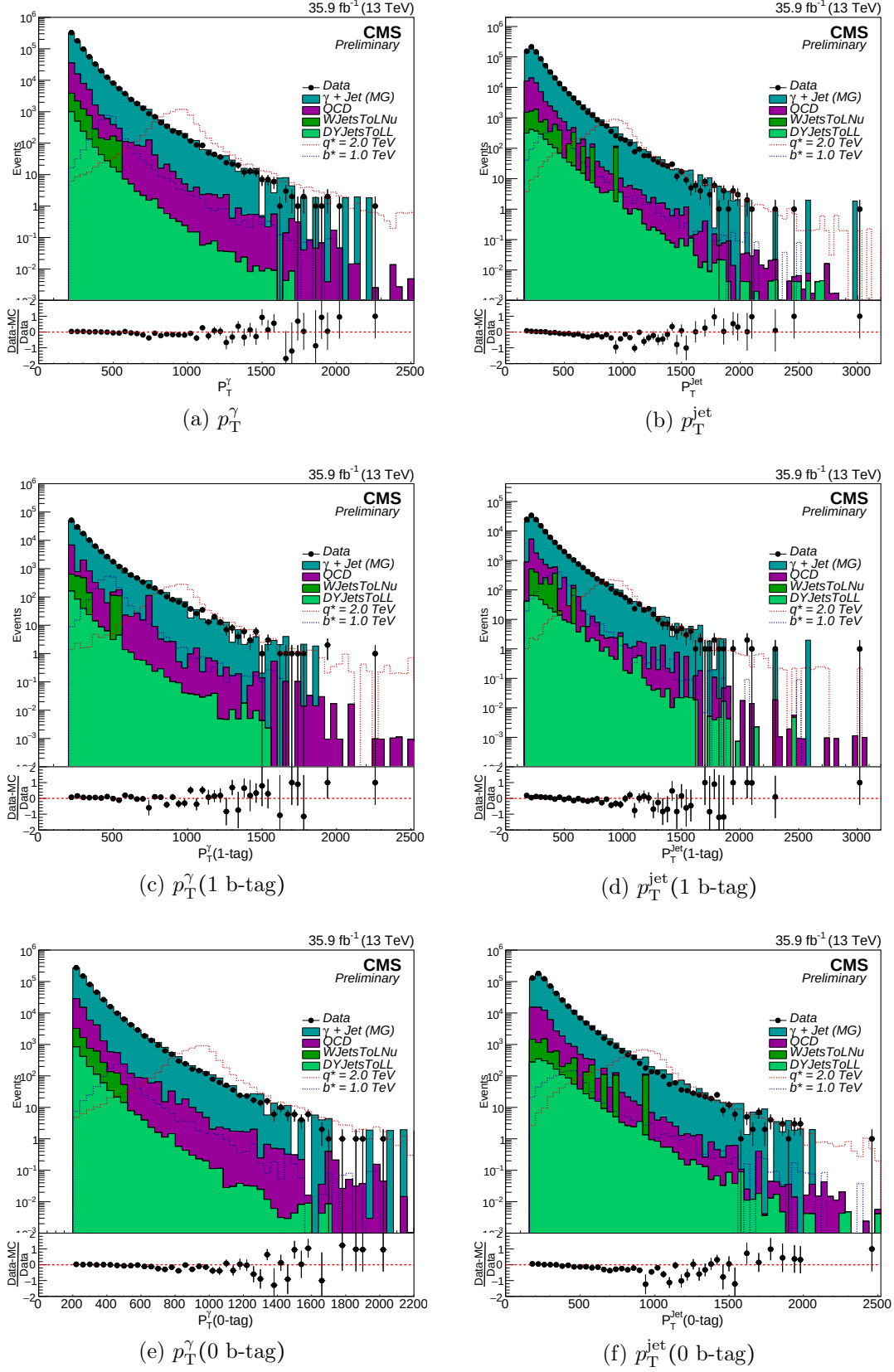


Figure 4.16: Data-MC comparison of p_T distribution of photons and jets for three categories: no b-tag, 1 b-tag and 0 b-tag.

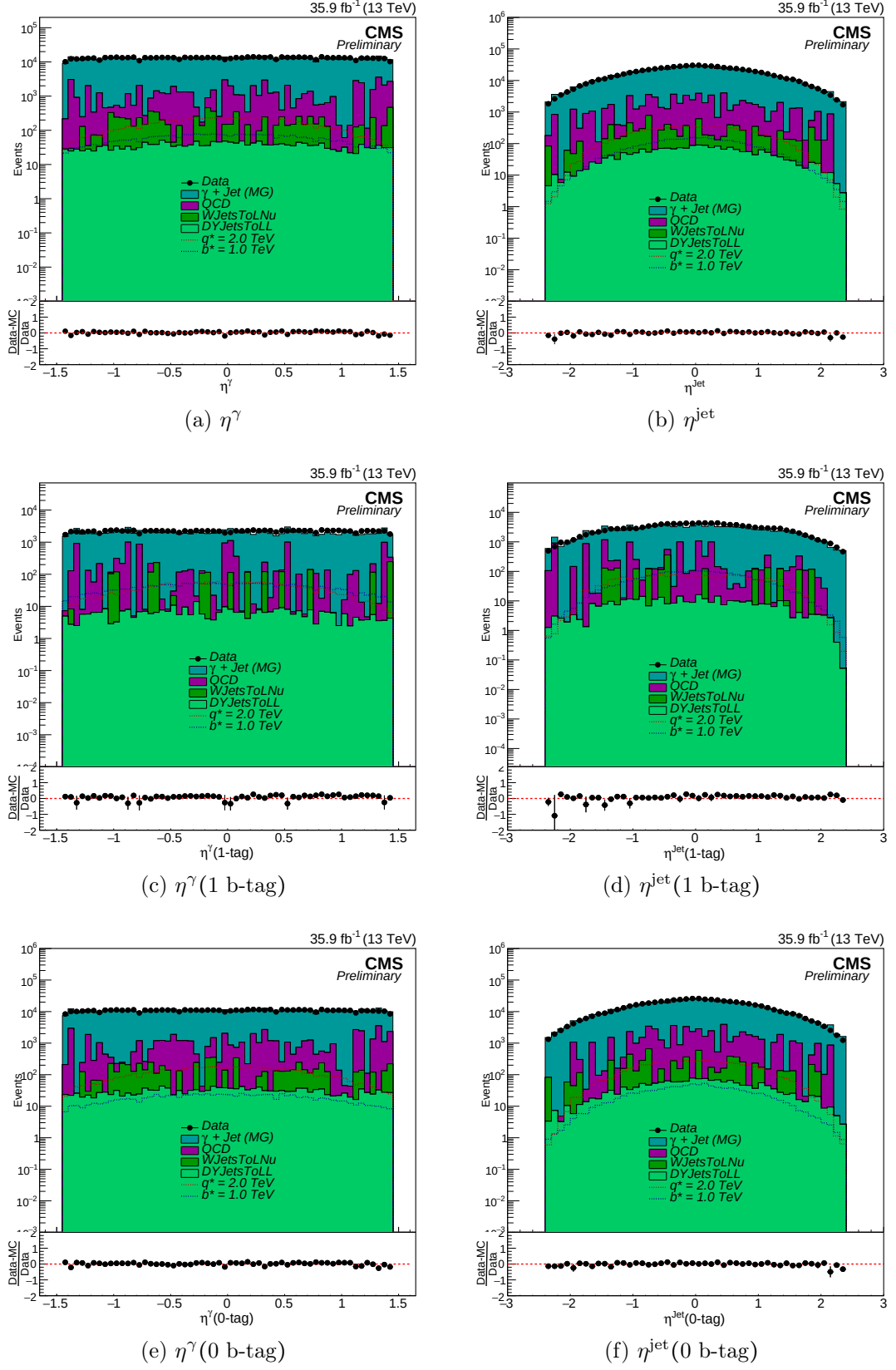


Figure 4.17: Data-MC comparison of η distribution of photons and jets for three categories: no b-tag, 1 b-tag and 0 b-tag.

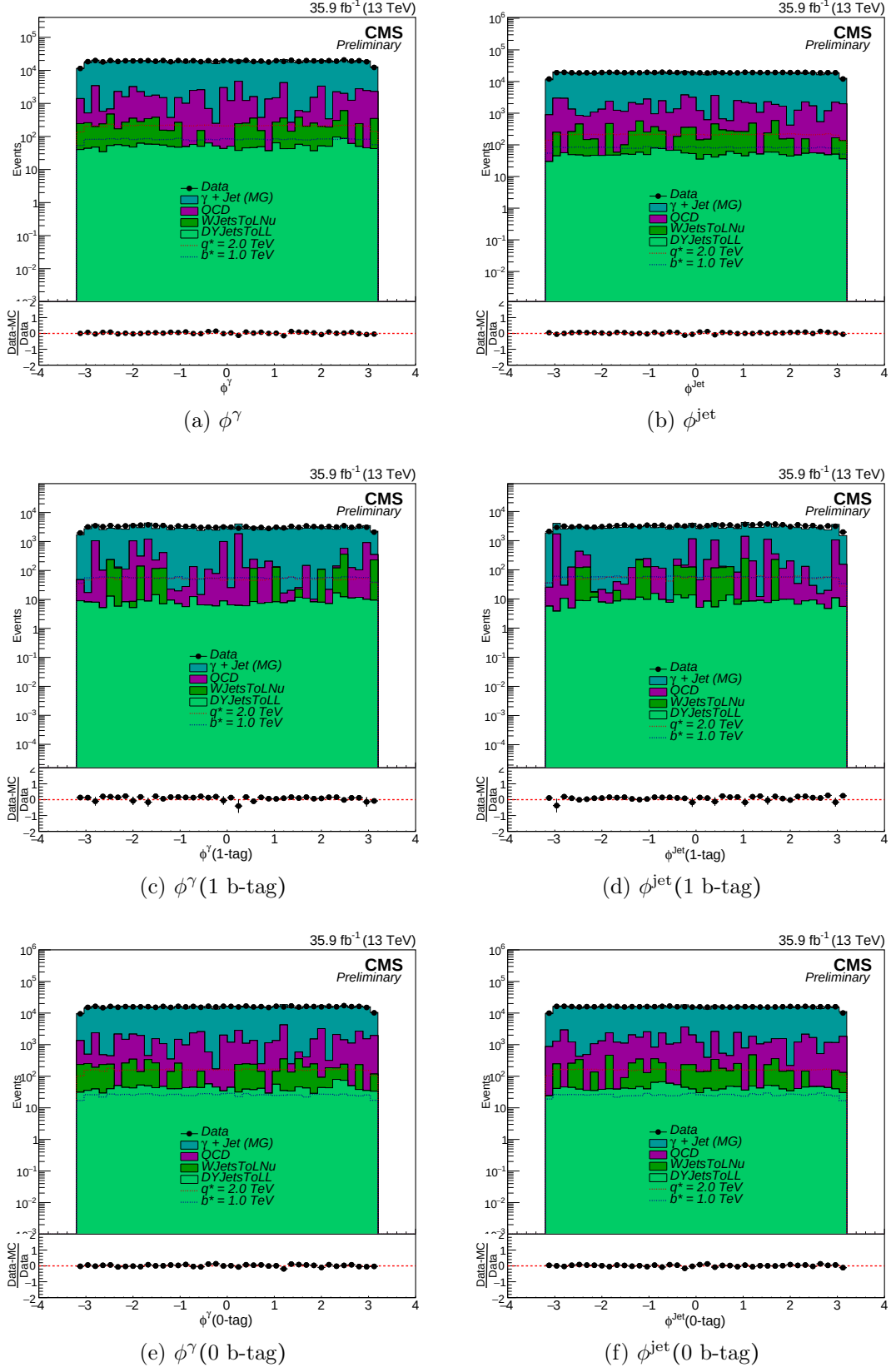


Figure 4.18: Data-MC comparison of ϕ distribution of photons and jets for three categories: no b-tag, 1 b-tag and 0 b-tag.

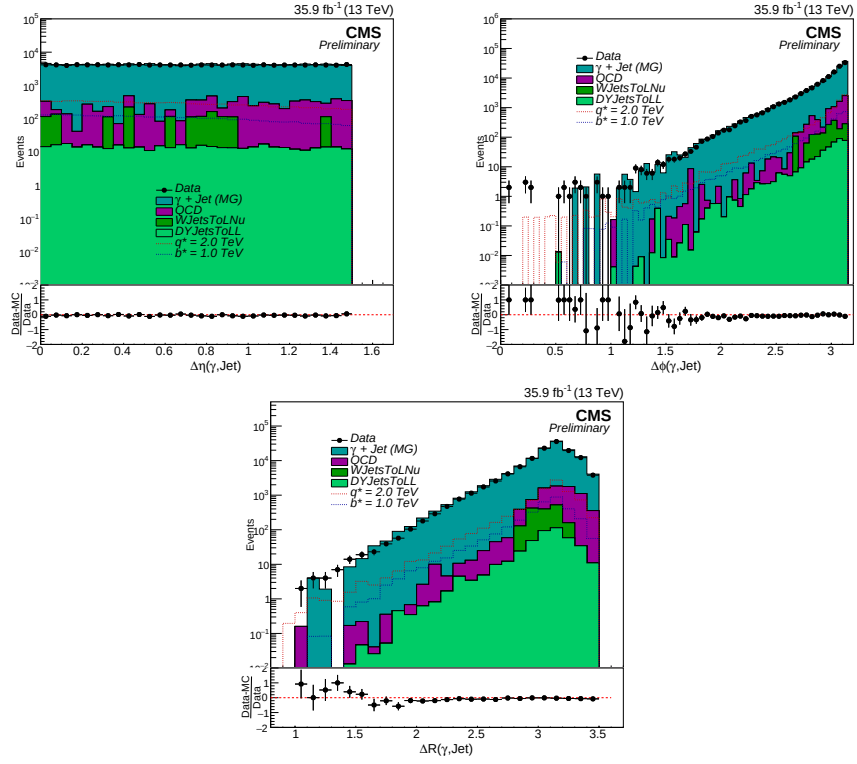


Figure 4.19: Distributions of photons and jets in $\Delta\eta$ (above left), $\Delta\phi$ (above right), and ΔR (below).

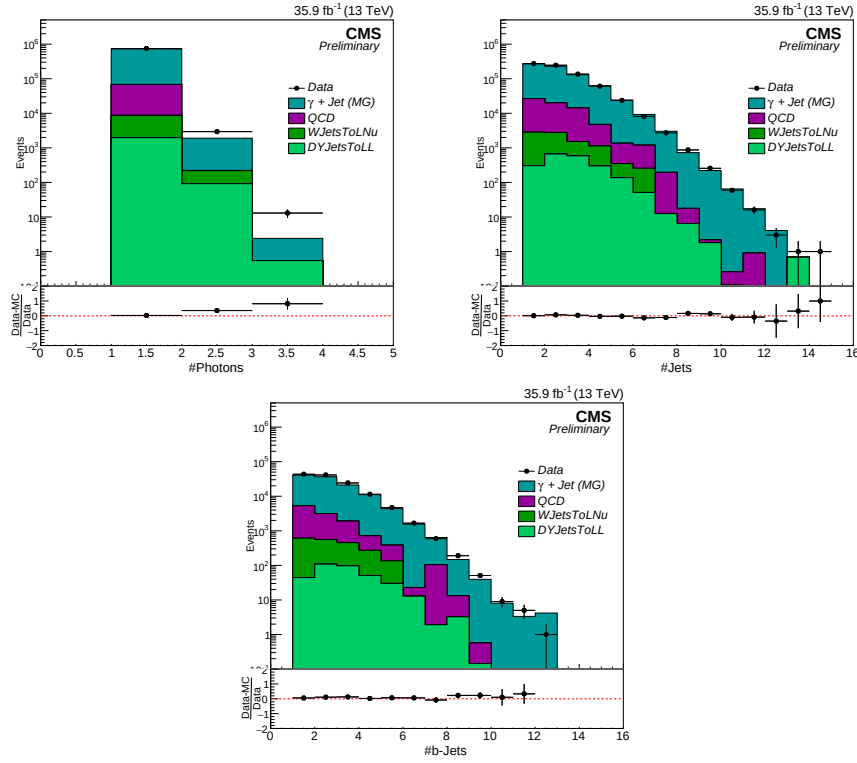
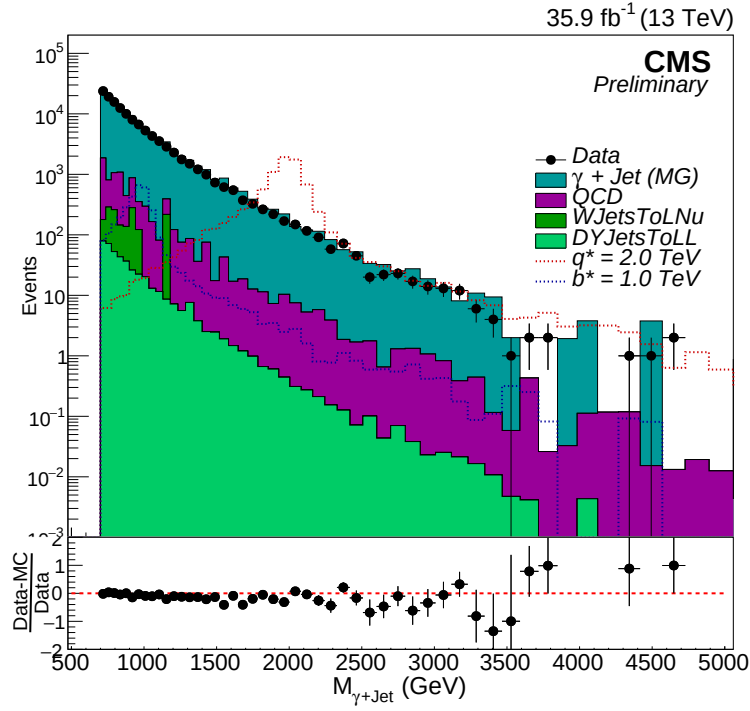
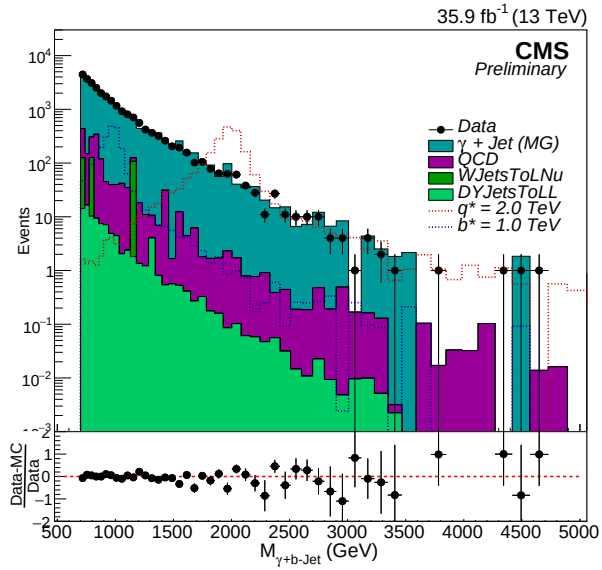


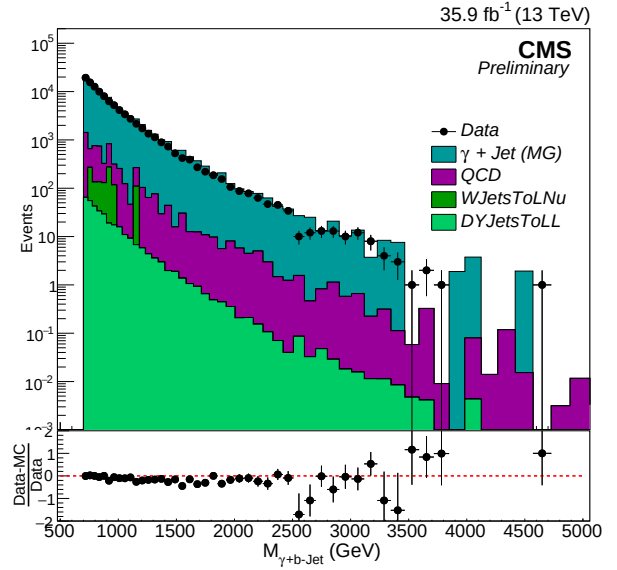
Figure 4.20: Number of photons (above left), jets (above right), and b-jets (below) per event in Data-MC.



(a) $M_{\gamma+\text{jet}}$



(b) $M_{\gamma+\text{jet}}$ (1 b-tag)



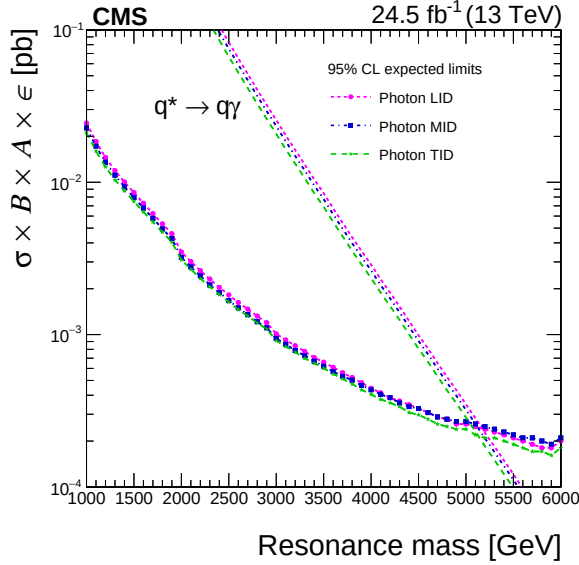
(c) $M_{\gamma+\text{jet}}$ (0 b-tag)

Figure 4.21: Data-MC comparison of invariant mass distribution of $\gamma + \text{jet}$ in three categories: no b-tag, 1 b-tag and 0 b-tag.

4.4.8 Optimization of event selection

The event selection considered in this analysis has been optimized for best expected mass limits on q^* and b^* resonances (the procedure to obtain expected limits is explained in the upcoming Section 4.10). The expected background obtained from the Poisson fluctuation of total background MC has been used in evaluating limits. The standard selection used for optimization study is “loose” photon ID, “tight” jet ID, $\Delta\eta < 1.8$, $\Delta\phi > 1.5$ and “medium” CSVv2 cut.

For different photon ID working points (WP), the expected limits at 95% CL are presented in Fig. 4.22 and Table 4.10. The optimization study tells that the “loose” WP is the most efficient one. However, it has been found that the trigger efficiency w.r.t muon reference trigger improves on going from loose to medium WP, therefore, “medium” WP for photon ID has been chosen.



Photon ID	Exp. mass limit (TeV)
loose	5.17
medium	5.11
tight	5.13

Table 4.10: Photon ID cut optimization corresponding to different mass expected limits.

Figure 4.22: Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different photon ID cuts corresponding to SM coupling $f = 1.0$.

The optimization studies for different $\Delta\eta$ cuts is presented in Fig. 4.23 and Table 4.11 and a $\Delta\eta(\gamma, \text{jet})$ cut of 1.5 is found to be the most efficient. One more advantage of a $\Delta\eta(\gamma, \text{jet})$ cut is that it helps in achieving mass turn-on at a lower mass point (as shown in Fig. 4.14) compared to no $\Delta\eta$ cut.

The corresponding optimization results for various $\Delta\phi$ cuts are presented in

Fig. 4.24 and Table 4.12 while for various CSVv2 cuts, it is presented in Fig. 4.25 and Table 4.13. A selection without any $\Delta\phi$ cut and a “loose” CSVv2 cut is found optimum.

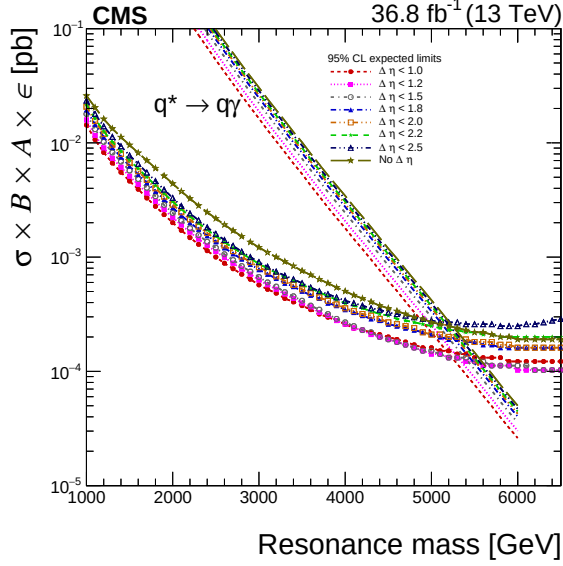


Figure 4.23: Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different $\Delta\eta$ cuts corresponding to SM coupling $f = 1.0$.

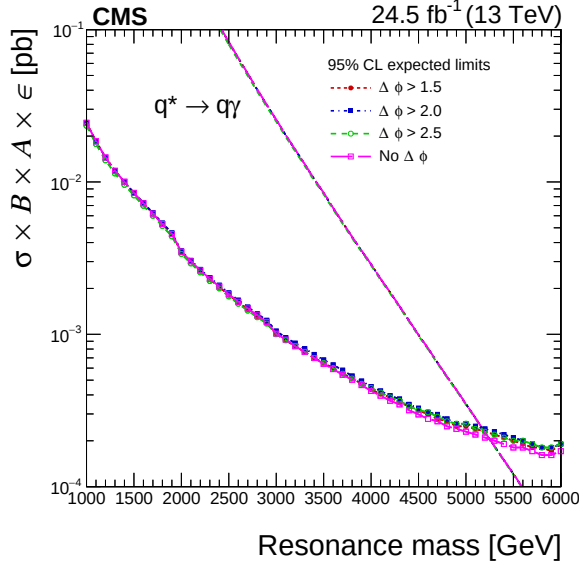
$\Delta\eta$ cuts	Exp. mass limit (TeV)
$\Delta\eta < 1.0$	5.17
$\Delta\eta < 1.2$	5.31
$\Delta\eta < 1.5$	5.39
$\Delta\eta < 1.8$	5.27
$\Delta\eta < 2.0$	5.28
$\Delta\eta < 2.2$	5.24
$\Delta\eta < 2.5$	5.19
No $\Delta\eta$	5.26

Table 4.11: $\Delta\eta$ cut optimization corresponding to different mass expected limits.

4.5 Highest invariant mass events in data

The event with highest invariant mass is observed in data at $M_{\gamma+\text{jet}} = 4616.69 \text{ GeV}$. The event display for this event has been made using the cmsShow [183] package in different detector orientations and is presented in Fig. 4.26. The photon candidate in this event is found to have a $p_T = 1253 \text{ GeV}$ and $\eta = -0.857733$ while the jet candidate has a $p_T = 3022.47 \text{ GeV}$ and $\eta = 0.395303$. The jet candidate is tagged as a b-jet, therefore, this event form the highest q^* as well as the highest b^* invariant mass event. This event was recorded by the CMS detector on 29th August, 2016 in the run number 279716.

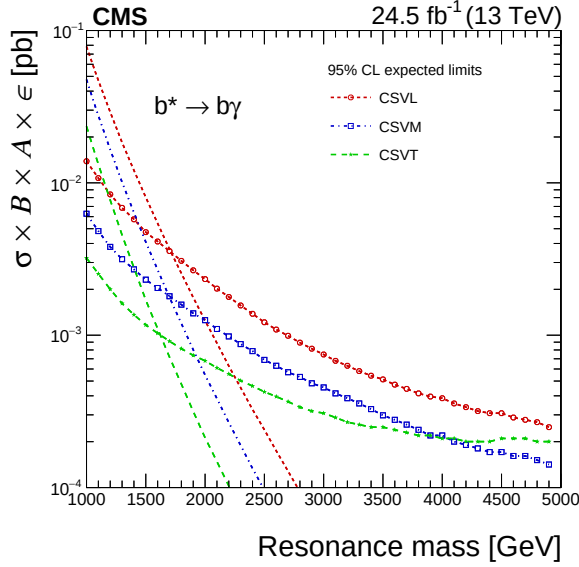
The event with 2nd highest invariant mass is observed at $M_{\gamma+\text{jet}} = 4595.43 \text{ GeV}$ with the photon candidate having $p_T = 1872.51 \text{ GeV}$, $\eta = 0.927873$ and the jet candidate having $p_T = 1945.66 \text{ GeV}$, $\eta = -0.299033$. Since the jet is not tagged as a b-jet, so this event form the part of q^* statistics only. The event display for this event is presented in Fig. 4.27.



$\Delta\phi$ cuts	Exp. mass limit (TeV)
$\Delta\phi > 1.5$	5.19
$\Delta\phi > 2.0$	5.17
$\Delta\phi > 2.5$	5.18
No $\Delta\phi$	5.24

Table 4.12: $\Delta\phi$ cut optimization corresponding to different mass expected limits.

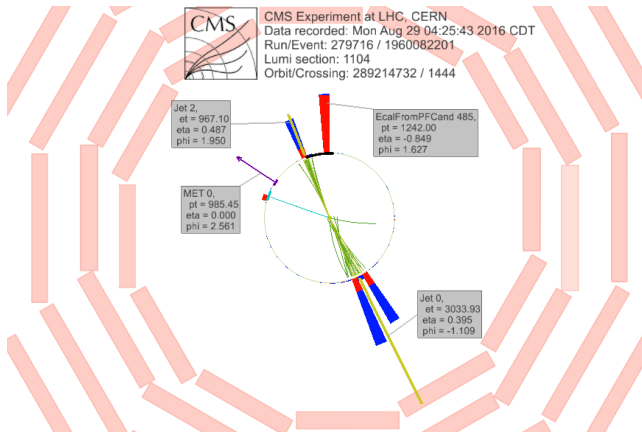
Figure 4.24: Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different $\Delta\phi$ cuts corresponding to SM coupling $f = 1.0$.



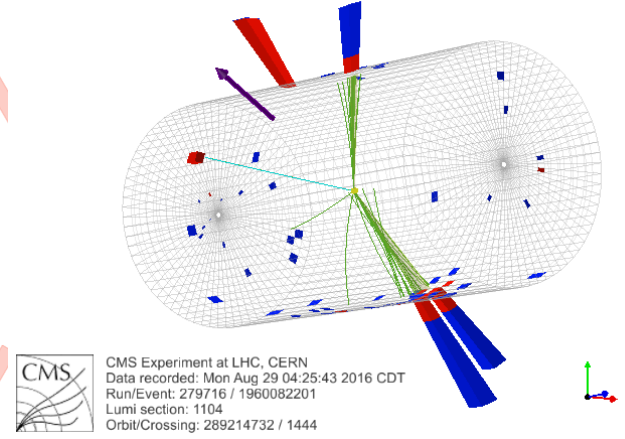
CSV2	Exp. mass limit (TeV)
loose	1.73
medium	1.70
tight	1.50

Table 4.13: CSV2 cut optimization corresponding to different mass expected limits.

Figure 4.25: Expected 95% CL upper limits on $\sigma \times \text{BR} \times A \times \epsilon$ versus q^* mass for different CSV2 cuts corresponding to SM coupling $f = 1.0$.

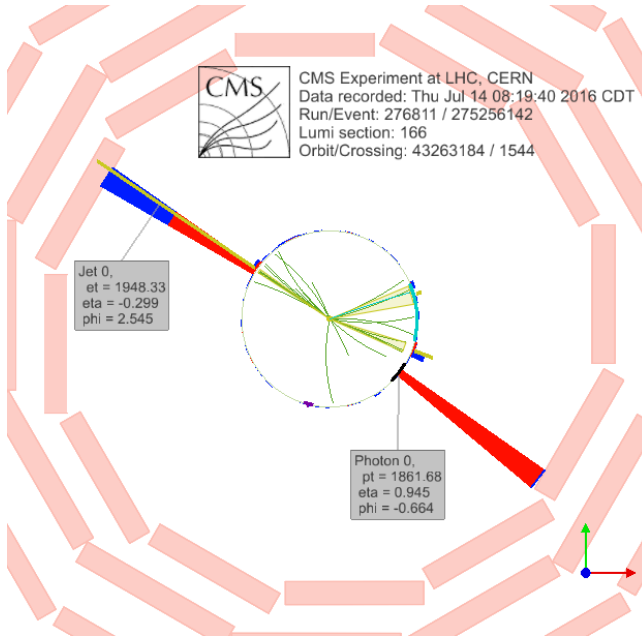


(a) ρ - ϕ plane

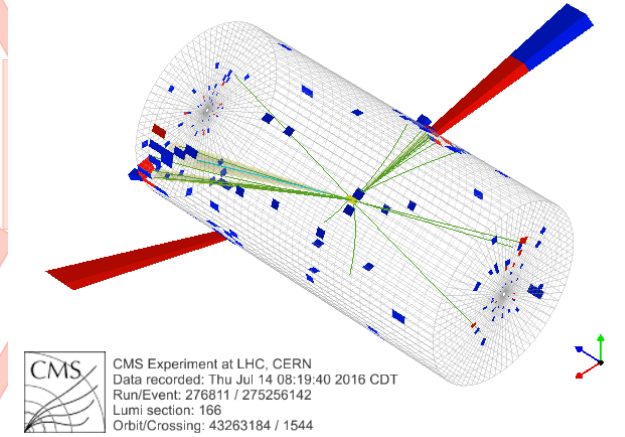


(b) 3-dimensional plane

Figure 4.26: Event display for the highest invariant mass event at $M_{\gamma+jet} = 4616.69 \text{ GeV}$. This event has the jet tagged as a b-jet, thereby, forming a highest q^* as well as highest b^* event.



(a) ρ - ϕ plane



(b) 3-dimensional plane

Figure 4.27: Event display for the 2nd highest invariant mass event at $M_{\gamma+jet} = 4595.43 \text{ GeV}$.

4.6 Signal spectrum

The γ + jet/b-jet invariant mass distribution spectrum falls steeply with increasing mass. The q^* and b^* signals, if exist, will appear as narrow resonances over the continuous spectrum. In order to observe these narrow resonances, the bin widths of invariant mass distributions should match with the mass resolutions of the q^* and b^* resonances. Since the resonance width and hence the mass resolution is directly proportional to the resonance mass, it increases with increase of resonance mass. So the invariant mass spectra are required to have a variable binning with increasing bin widths.

To obtain variable binning, the signal shape at different masses are fitted with three different functions, viz. Gaussian, Breit-Wigner and Crystal-Ball (CB), as shown in Fig. 4.28 for some of q^* mass points. This is required to retrieve the mass resolutions. The Crystal-Ball function is found to best describe the shapes and its functional form is given by,

$$CB(x; \alpha, n, \mu, \sigma) = N_{CB} \cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & \text{for } \frac{x-\mu}{\sigma} > -\alpha \\ \left(\frac{n}{|\alpha|}\right)^n \cdot \exp\left(-\frac{|\alpha|^2}{2}\right) \cdot \left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\mu}{\sigma}\right)^{-n}, & \text{for } \frac{x-\mu}{\sigma} \leq -\alpha, \end{cases} \quad (4.3)$$

where N_{CB} is the normalization factor, σ is the width, μ is the mean and α, n are the parameters determining the functional fitting. The resonance widths and means obtained from CB fitting are then used in a functional form,

$$\frac{\sigma}{\text{Mean}} = A + \frac{B}{M_{\text{Res}}} \quad (4.4)$$

where $\frac{\sigma}{\text{Mean}}$ is known as the mass resolution with σ , the resonance width and Mean, the resonance mean obtained upon fitting and M_{Res} is the resonance mass. This functional form is plotted between $\frac{\sigma}{\text{Mean}}$ and M_{Res} for different masses and fitted to obtain the values of parameters A and B as shown in Fig. 4.29. These parameters are then used to obtain variable mass binning for q^* and b^* invariant mass distributions and is presented below:

```
nMassBin[120] = { 1, 3, 6, 10, 16, 23, 31, 40, 50, 61, 73, 86, 100, 115, 132, 150, 169,
                  189, 210, 232, 252, 273, 295, 318, 341, 365, 390, 416, 443, 471, 500,
                  530, 561, 593, 626, 660, 700, 731, 768, 806, 846, 887, 929, 972, 1017,
                  1063, 1110, 1159, 1209, 1261, 1315, 1370, 1427, 1486, 1547, 1609, 1673,
                  1739, 1807, 1877, 1950, 2025, 2102, 2182, 2264, 2349, 2436, 2526, 2619,
                  2714, 2812, 2913, 3018, 3126, 3237, 3352, 3470, 3592, 3718, 3847, 3980,
```

4117, 4259, 4405, 4556, 4711, 4871, 5036, 5206, 5381, 5562, 5748, 5940,
6138, 6342, 6552, 6769, 6993, 7223, 7461, 7706, 7959, 8219, 8487, 8764,
9049, 9343, 9646, 9958, 10280, 10612, 10954, 11307, 11671, 12046, 12432,
12830, 13241, 13664, 14000};

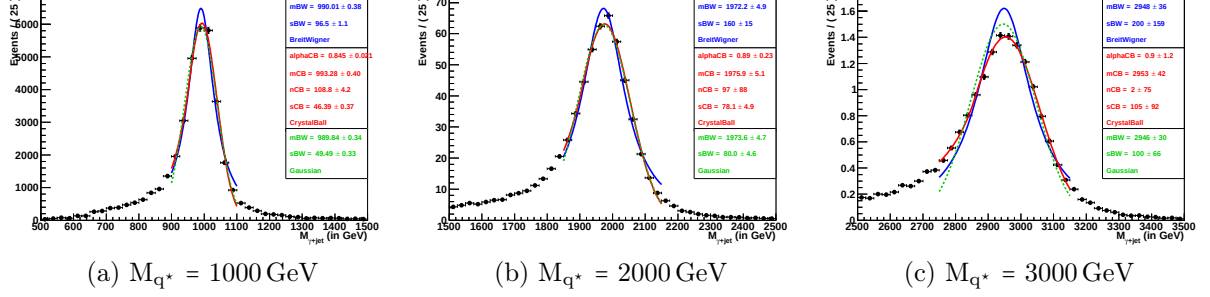


Figure 4.28: Fitting the core of invariant mass of q^* signal at different masses for $f = 1.0$, using Breit-Wigner (Blue), Crystal-Ball (Red), and Gaussian (Green) functions to obtain the resonance width of the signal.

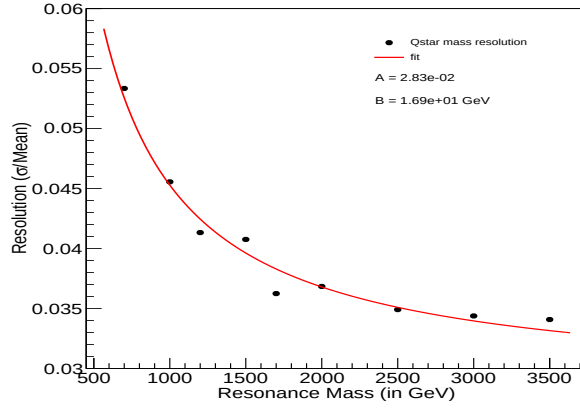


Figure 4.29: Polynomial fitting of mass resolution vs. resonance mass to obtain variable binning parameters.

4.7 Signal interpolation

While searching for new particles, with unknown masses, it is essential to scan the entire possible mass range. Due to limitation of resources, only a limited number of signal masses over a large range are generated. For q^* (b^*) resonances, only 10 mass points per coupling multiplier are generated in the mass range from 0.5–9.0 TeV (0.5–5.0 TeV). The signal shapes at intermediate masses are obtained by using an interpolation technique [184] which utilizes the information from neighbouring generated masses.

In order to interpolate a signal shape at an intermediate mass M between two generated masses M_A and M_B using this technique, a probability distribution in a parameter X is constructed which is defined as $\text{Prob.}(X) = \frac{M_{\gamma+\text{jet}}}{M_{\text{Res}}}$, where M_{Res} is the resonance mass and $M_{\gamma+\text{jet}}$ is the invariant mass distribution at the resonance mass. This distribution for masses M_A and M_B are obtained and used to get the $\text{Prob.}(X)$ distribution for mass M using the relation,

$$\text{Prob}_M(X) = \text{Prob}_{M_A}(X) + [\text{Prob}_{M_B}(X) - \text{Prob}_{M_A}(X)] \cdot \frac{M - M_A}{M_B - M_A} \quad (4.5)$$

The X -distribution is then converted back to $\gamma + \text{jet}$ mass distribution using the `ROOT::Math::Interpolator` [134] class to get the resonance shape at the interpolated resonance mass. The generated and interpolated signal shapes for different q^* mass points are plotted alternatively for comparison in the Fig. 4.30.

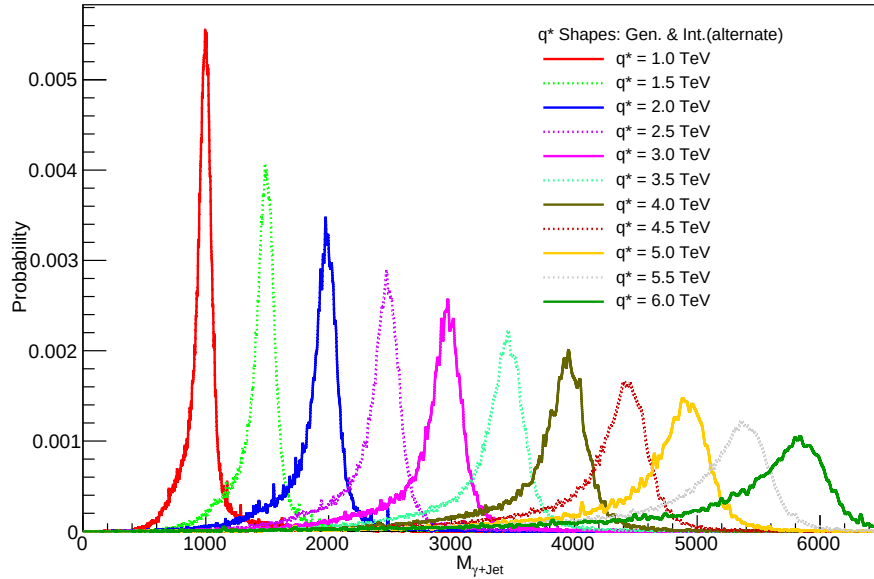


Figure 4.30: Alternate generated (solid line) and interpolated (dotted line) shapes for different q^* mass points.

The interpolation procedure has been authenticated by performing a closure test in which already generated shapes from PYTHIA are interpolated again using the immediate neighbours and a comparison is made as shown in Fig. 4.31 for three different q^* mass points. The interpolated shapes are found to be in good agreement with the generated ones. Only a small difference is observed on going towards higher masses, which is considered as a systematic uncertainty.

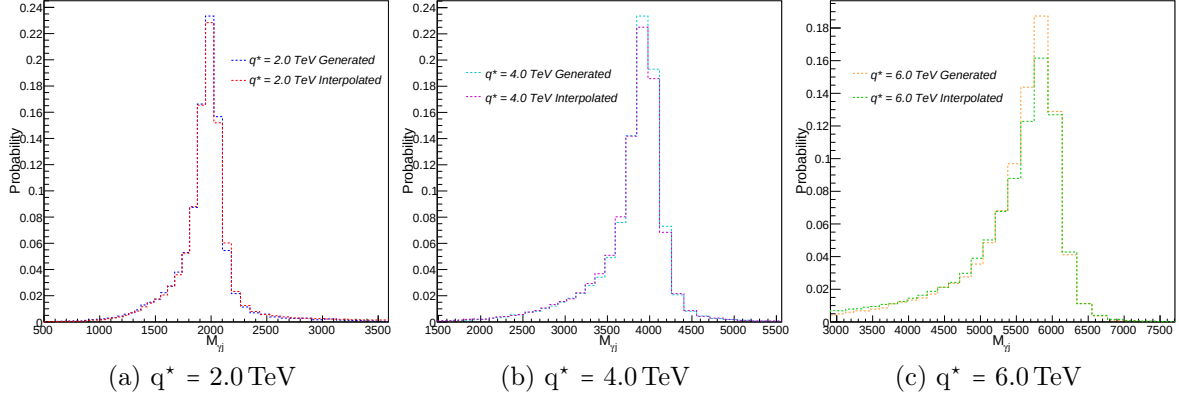


Figure 4.31: Comparison of generated and interpolated shapes at different q^* mass points.

4.8 Parametric background

The invariant mass distributions of γ +jet and γ +b-jet are well described by MC simulation. The binning of invariant mass distribution is chosen to have a bin width almost equal to the expected q^* and b^* mass resolutions, that varies from about 4.5% at 1 TeV to 3.3% at 6 TeV. However, in order to model the background of this study, a data driven method using a smooth parameterization has been considered. The usage of a fit function is advantageous over MC simulation in background modelling as MC simulation has many theoretical (like PDFs, renormalization scale etc.) and experimental (like jet energy scale, energy resolutions etc.) uncertainties associated with it. The methodology of smooth parameterization is considered because the background due to γ + jet processes is smooth and monotonically falling and can be extracted by fitting the distribution with a smooth polynomial function.

Several functional forms inspired from previous similar searches [185–190] have been considered, which are listed below:

$$1. F_1 = \frac{P_0(1-x)^{P_1}}{x^{P_2+P_3} \ln x}$$

$$2. F_2 = \frac{P_0(1+x)^{P_1}}{x^{P_2+P_3} \ln x}$$

$$3. F_3 = \frac{P_0}{(P_1+P_2 \cdot x+x^2)^{P_3}}$$

$$4. F_4 = \frac{P_0}{(P_1+x)^{P_2}}$$

$$5. F_5 = P_0 \cdot \exp(1 + P_1 \cdot x + P_2 x^2 + P_3 x^3)$$

where $x = m/\sqrt{s}$, m being the γ + jet invariant mass and $\sqrt{s}$ being the center of mass energy, P_0 , P_1 , P_2 and P_3 are the fit parameters.

4.8.1 Fisher test

The order of the various functions for q^* and b^* scenarios has been chosen on the basis of the Fisher-test. This test is used to determine whether for two fit models, 1 and 2, where model 1 with p_1 parameters is “nested” within model 2 with p_2 parameters ($p_2 > p_1$), the model 1 provide *significantly* better fit to data compared to model 2.

A F-test is any statistical test which has the test statistic distributed according to the F-distribution under the null hypothesis. We use the saturated model goodness-of-fit test statistics [191, 192]:

$$-2 \log \lambda = 2 \sum_j^{n_{\text{bins}}} = (b_j - x_j + x_j \log(x_j/b_j)) \quad (4.6)$$

where x_j is the number of data events in the j^{th} bin and b_j is the prediction in the j^{th} bin. This test statistic asymptotically follows a χ^2 -distribution with $(n_{\text{bins}} - p_i)$ degrees of freedom, p_i being the number of parameters for model i . This statistic forms an appropriate metric for the goodness-of-fit of maximum likelihood fits using a Poissonian likelihood. Based on this, a F-statistics has been constructed as

$$F = \frac{-2 \log(\lambda_1/\lambda_2)/(p_2 - p_1)}{-2 \log \lambda_2/(n_{\text{bins}} - p_2)}, \quad (4.7)$$

where n_{bins} is the number of bins and $-2 \log \lambda_i$ is the “saturated model” goodness-of-fit test statistic for model i .

The null hypothesis states that the model 2 does not provide a significantly better fit compared to model 1 and under this hypothesis, F will have a F-distribution with $(p_2 - p_1, p_2 - n)$ degrees of freedom. The null hypothesis will be rejected if the F computed from data is greater than the critical value of F-distribution for the false rejection probability of $\alpha = 0.05$.

To perform this test, the pseudo invariant mass distribution from background MC has been considered and fitted using the same function with p_1 and p_2 parameters, where $p_2 > p_1$, to obtain the best-fit parameters. Based on the fitting, the χ^2 test statistics (eq. 4.6) for each set of parameters is constructed and a

F -value is determined using the definition above. Now assuming that the null hypothesis is true, that is function with p_2 parameters does not provide a better fit compared to function with p_1 parameters, toys are generated using fit results corresponding to the function with p_1 parameters. A F -value is computed for each toy and a F -distribution is constructed. If the F -value obtained from initial best fit parameters is found to be greater than the F -value corresponding to a p-value of 0.05, null hypothesis is rejected, otherwise accepted.

We have sequentially tested each order of different functions ($F1 - F5$) and obtained the best orders (as listed above) for which a F -value corresponding to a p-value > 0.05 is obtained. The F -distributions for function $F1$ corresponding to 3 and 4 parameters has been presented in Fig. 4.32 for q^* and in Fig. 4.33 for b^* . A p-value > 0.05 is obtained for $F1$ with 4 parameters in each case.

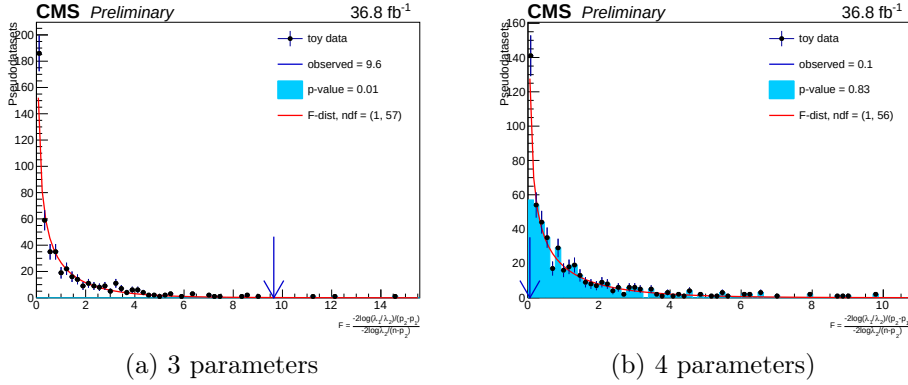


Figure 4.32: The F -test distributions for fitting of function $F1$ with q^* invariant mass, for 3 and 4 parameters, showing that $F1$ with 4 parameters provide a better fit compared to 3 parameters.

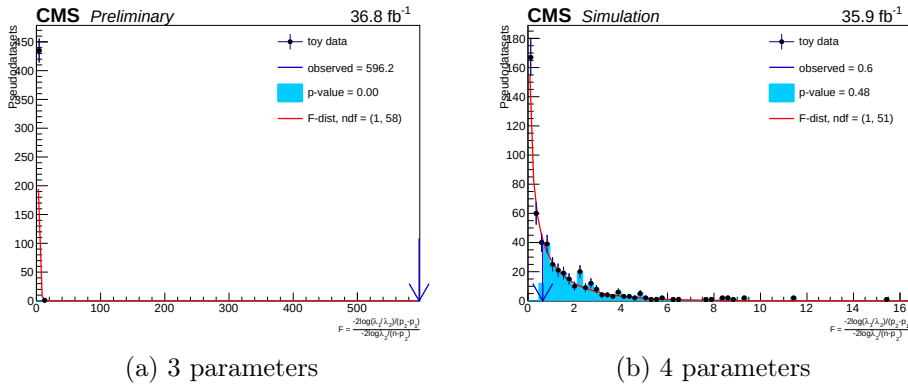


Figure 4.33: The F -test distributions for fitting of function $F1$ with b^* invariant mass, for 3 and 4 parameters, showing that $F1$ with 4 parameters provide a better fit compared to 3 parameters.

4.8.2 Goodness-of-fit

In order to test the goodness-of-fit (GOF) of different functional forms $F1 - F5$ with data, we have used the modified χ^2 test statistics,

$$\chi^2 = \sum_{i=1}^{n_b} \left(\frac{x_i - b_i}{\sigma_{x_i}} \right)^2, \quad (4.8)$$

where the “uncertainty” σ_{x_i} is the 68% CL region of a Poisson distribution and is defined as follows, setting $\alpha = 1 - 0.687$,

$$\sigma_{x_i} = \begin{cases} D_c^{-1}(\alpha/2, x_i + 1), & \text{if } b_i > x_i \\ D^{-1}(\alpha/2, x_i), & \text{if } b_i < x_i \end{cases} \quad (4.9)$$

where $D^{-1}(\alpha/2, x_i)$ is the quantile function of gamma distribution, which is also the inverse of the cumulative distribution function of the lower tail of the gamma distribution,

$$D(\alpha/2, x_i) = \int_{-\infty}^{\alpha/2} \frac{1}{\Gamma(x_i)} z^{x_i-1} e^{-z} dz, \quad (4.10)$$

and, $D_c^{-1}(\alpha/2, x_i + 1)$ is the inverse of the cumulative distribution function of the upper tail of the gamma distribution,

$$D_c(\alpha/2, x_i + 1) = \int_{\alpha/2}^{+\infty} \frac{1}{\Gamma(x_i + 1)} z^{x_i} e^{-z} dz. \quad (4.11)$$

We generated around 1000 pseudo datasets using the best-fit model parameters on data, refit each pseudo dataset with the maximum likelihood fit and save the test statistic value. The distribution of the test statistics from these pseudo datasets as well as the value observed in data for different functions $F1 - F5$ is presented in Fig. 4.34 for q^* and in Fig. 4.35 for b^* . The χ^2/ndf values for different functions is also reported.

Since almost all the functions provide reasonable fit to data, one of the functions, $F1$ has been chosen as the background parameterization for this study, written as

$$\frac{d\sigma}{dm} = \frac{P_0(1 - m/\sqrt{s})^{P_1}}{(m/\sqrt{s})^{P_2+P_3 \ln(m/\sqrt{s})}}. \quad (4.12)$$

This parameterization is motivated by the LO QCD. The term x^{P_2} (or m^{P_2}) in the denominator refers to the mass dependence of the QCD matrix element while

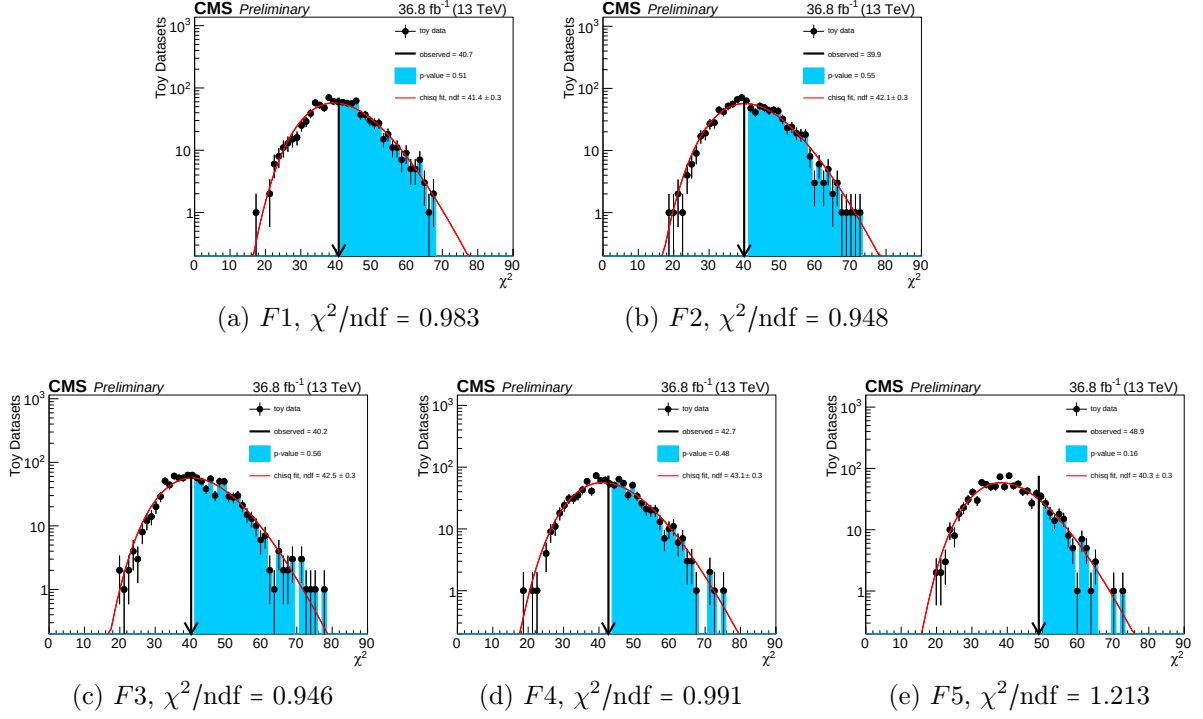


Figure 4.34: Toy distributions for the goodness-of-fit study for q^* .

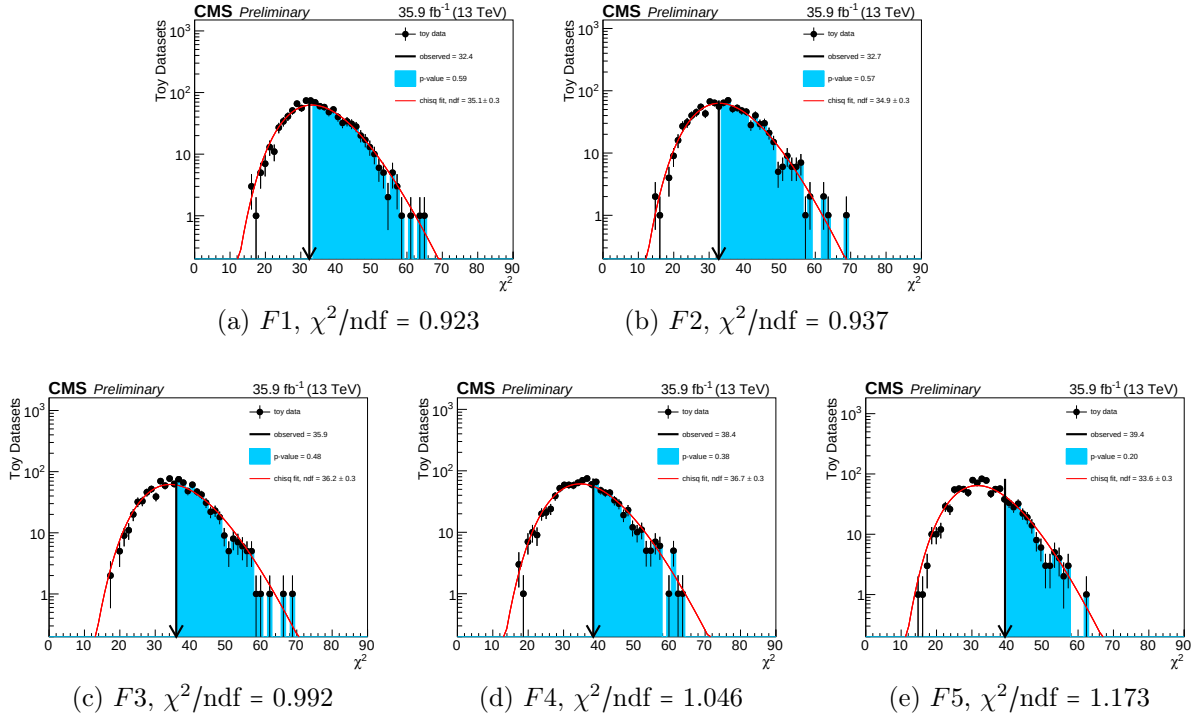


Figure 4.35: Toy distributions for the goodness-of-fit study for b^* .

the term $(1-x)^{P_1}$ in the numerator refers to the mass dependence of the parton distributions, the term P_3 models the data at higher invariant masses.

We have performed an extended, background-only, binned, maximum-likelihood fit to the data by using the following likelihood,

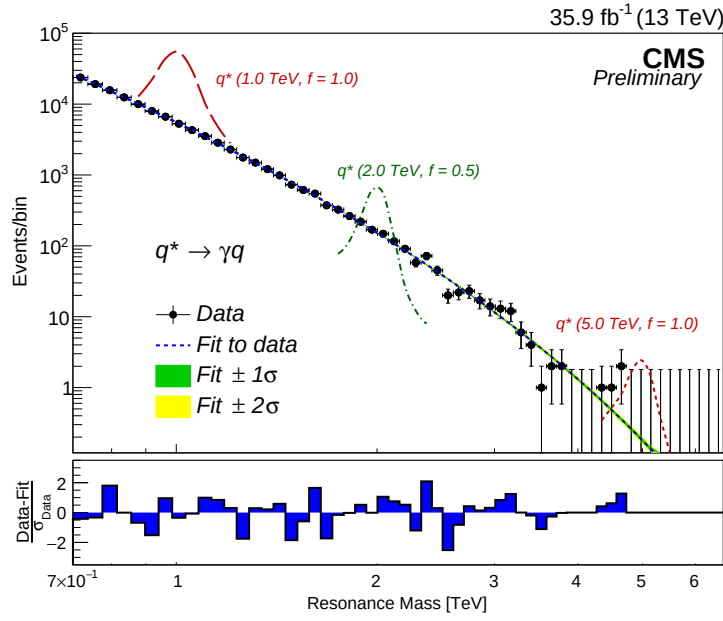
$$\mathcal{L}(\text{data}|\theta) = \prod_{i=1}^{n_b} \text{Poisson}(x_i|b_i(\theta)) = \prod_{i=1}^{n_b} \frac{b_i(\theta)^{x_i} e^{-b_i(\theta)}}{x_i!} \quad (4.13)$$

where n_b corresponds to the number of bins, θ is the vector of nuisance parameters (P_0, P_1, P_2, P_3) (or, equivalently (N_b, P_1, P_2, P_3)), x_i is the data yield in the i^{th} bin, and b_i is the integral of the fit function in i^{th} bin multiplied by the expected events N_b . The result of the background only fit to the data for q^* and b^* in three categories is presented in Fig. 4.36.

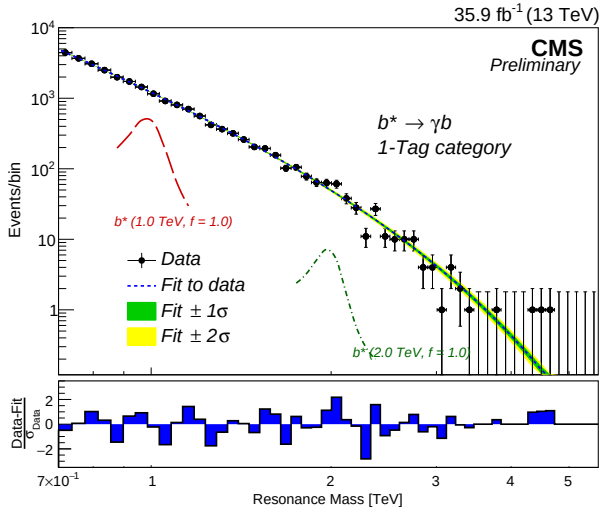
4.8.3 Bias due to the choice of background fit function

The background fit function defined in Eq. 4.12 has been considered to model the $\gamma + \text{jet}$ and $\gamma + \text{b-jet}$ invariant mass distributions. However, the uncertainties associated with the choice of a particular fit function can impact the final result. Therefore, a bias study has been performed to quantify the amount of potential bias that can be present given a certain choice of functional form. In this study, the function $F1$ is taken as the default function while the functions $F2 - F5$ are taken as the alternate functions. The procedure implemented to study the bias is described as follows:

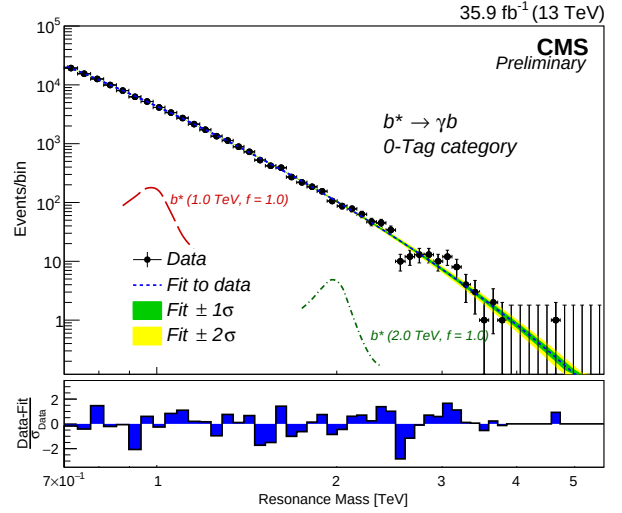
- An underlying distribution $h(m_{\gamma+\text{jet}})$ in the $\gamma + \text{jet}$ (and $\gamma + \text{b-jet}$) invariant mass distribution is obtained using the total MC background.
- S+B fit to $h(m_{\gamma+\text{jet}})$ with an input signal of mass m and signal strength $\mu = 1$ is performed using the test functions $F2 - F5$ to get the best fit parameters.
- These best fit parameters are then used to generate N number of toys corresponding to each function.
- The toys associated with each test function are then fitted with the default background function $F1$ and strength of the signal is predicted for each toy using maximum likelihood approach, resulting into a distribution $\hat{f}_i(m_{\gamma+\text{jet}})$, $i = 2, 3, 4, 5$, of μ_i
- The signal strength $\hat{\mu}_i$ predicted by $\hat{f}_i(m_{\gamma+\text{jet}})$ is compared with input signal



(a) $M_{\gamma+\text{jet}}$



(b) $M_{\gamma+\text{jet}}$ (1 b-tag)



(c) $M_{\gamma+\text{jet}}$ (0 b-tag)

Figure 4.36: Invariant mass distribution spectrum of $\gamma + \text{jet}$ fitted with 4-parameter polynomial fit function for three categories: (a) no b-tag (b) 1 b-tag and (c) 0 b-tag.

strength $\mu = 1$ in several mass windows m_j and a pull test statistics is constructed as:

$$p_i^j = \frac{\hat{\mu}_i^j - \mu}{\sigma_{\mu_i^j}} \quad (4.14)$$

- This pull distribution obtained for functions $F2 - F5$ for different masses are fitted with Gaussian to obtain the predicted value of μ .
- For the ideal case, the pull distribution should be a normal distribution with mean close to zero and rms close to 1. The deviation from zero is the measure of the bias introduced on choosing default function to describe the data when the true distributions are described by $F2$ to $F5$.

A background parameterization is considered to be acceptable if for different mass regions m_j , the following relation holds:

$$b^j = |\text{median}(p_i^j)| < 0.5, \quad (4.15)$$

where p_i^j is the Gaussian mean for i^{th} toy and j^{th} mass point. The bias is considered as negligible if for each mass point, the median of all the toys is found to be within 0.5. Fig. 4.37 shows that the above relation holds for all the four test functions corresponding to q^* and b^* scenarios. The observed maximum bias is found to be less than or close to 50%.

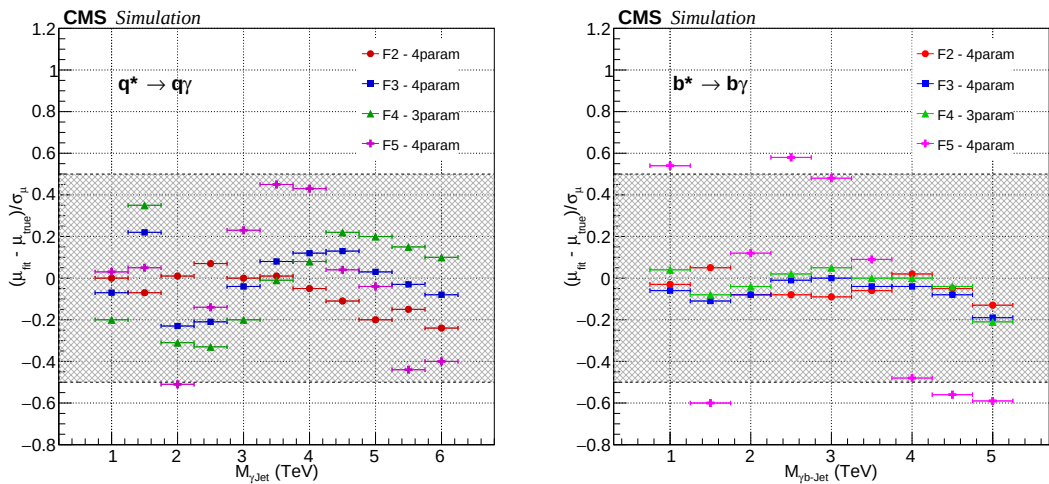


Figure 4.37: Pulls distribution as a function of q^* (left) and b^* (right) mass corresponding to the four test functions.

The pull distributions for q^* and b^* modelling are thus found to be consistent with a normal distribution with medians deviating by no more than 0.5 from zero and widths consistent with unity. If added in quadrature with the statistical uncertainty, the contribution of bias uncertainty is found to be only 10% of the total. Therefore, it has been concluded that the systematic uncertainty associated with the choice of background fit function is negligible and the statistical uncertainty of the fit is the only uncertainty that need to be considered.

4.9 Systematic Uncertainties

A number of systematic uncertainties due to the various sources, like, object identification, energy scales, luminosity measurements etc., are associated with this analysis. This section summarizes the most significant of these, as listed below:

- Luminosity
- Pile-up uncertainty
- HLT trigger inefficiency
- Photon and jet energy scale
- Photon and jet energy resolution
- Photon Id inefficiency
- b-tag scale factor
- Signal shape interpolation
- Background shape

The background shape uncertainty is considered for the background, the rest are considered for signal shapes only.

4.9.1 Luminosity

The uncertainty on the total integrated luminosity is provided centrally by CMS and is taken to be 2.5%.

4.9.2 Pile-up Uncertainty

The uncertainty associated with the modelling of pile-up is estimated by performing a variation of $\pm 5\%$ in the number of interactions while using a central

value of total inelastic cross-section to be 69 mb. Effect of pileup uncertainty on $A \times \epsilon$ of signal samples is found to $\sim 1 - 2\%$.

4.9.3 HLT trigger inefficiency

The trigger turn-on w.r.t the muon reference trigger is found to be around 95% efficient and hence an uncertainty of 5% has been assigned to this effect.

4.9.4 Photon and jet energy scale

The measured energy of the particle differs from its true energy due to the non-uniformity and non-linear response of the calorimeters. This difference is corrected in the photon and jet energy scale (PES & JES) at the reconstruction level. The uncertainties associated with these corrections form the source of systematic uncertainty for this analysis. The photon [117] and jet [193] energy scale uncertainties are extracted on an event by event basis and their effects are transmitted to the invariant mass distribution of the signal, to compute the amount of systematic effect. The effect of JES and PES uncertainties on the signal mass can be seen in Fig. 4.38. The effect of JES uncertainty comes out to be around 1% while that of PES is around 0.6%.

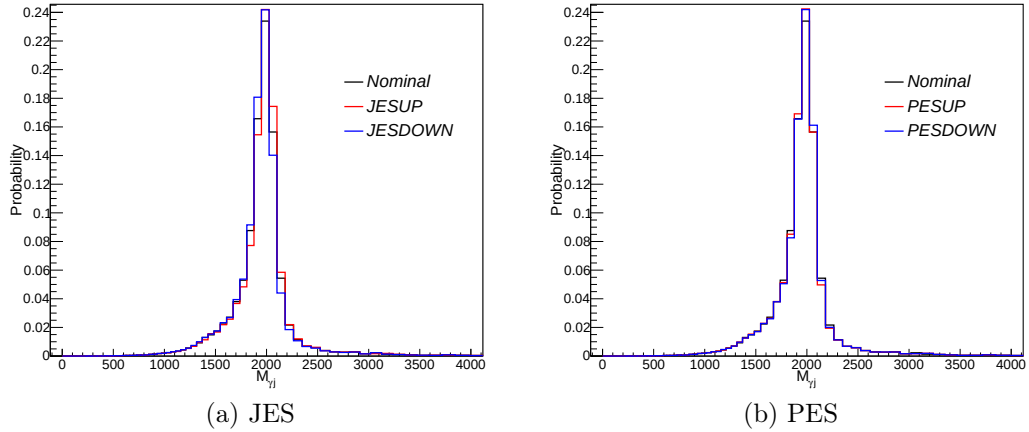


Figure 4.38: Effect of jet and photon energy correction uncertainties on the invariant mass of q^* signal.

4.9.5 Photon and jet energy resolution

The photon and jet energy measurements have some associated resolution which is different in data and MC and hence, can impact the energy/mass resolution

of the expected signal in data, resulting into the shrinking or stretching of the signal shape which can further impact the search. So, the resolution effect need to be taken into account in the form of a systematic uncertainty. This effect as photon [117] and jet [193] energy resolution (JER & PER) is computed and propagated to the signal invariant mass. The effect of JER and PER on signal shape is shown in Fig. 4.39 and the magnitude of both these effects is found to be around 0.2–0.4%.

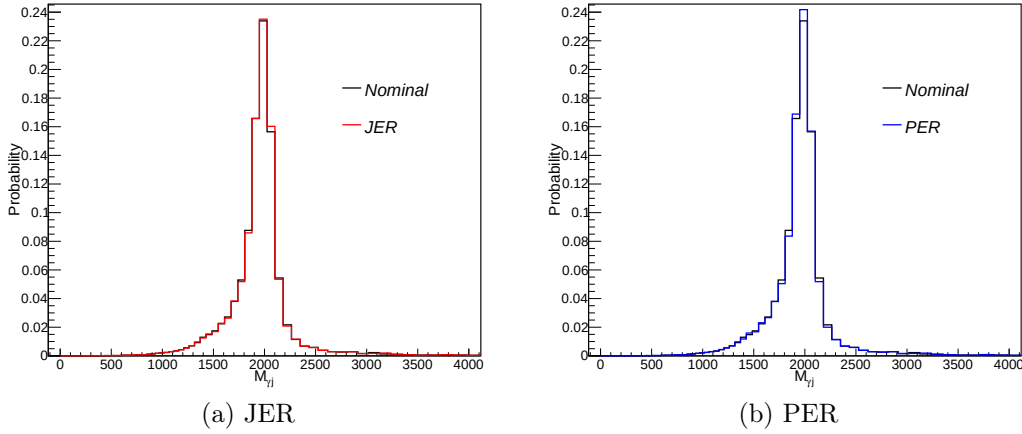


Figure 4.39: Effect of jet and photon energy resolution on the invariant mass of q^* signal.

4.9.6 Photon ID inefficiency

A systematic of 2% due to photon id inefficiency has been considered in this analysis.

4.9.7 b-Tag scale factor

The efficiency of b-tag discriminator is not the same in data and MC and to overcome this difference, b-tag scale factors are provided centrally in CMS. These scale factors have some associated errors which are propagated as a systematic uncertainty on the signal shape and normalization. The effect of these errors on the invariant mass spectrum of b^* signal can be seen in Fig. 4.40. The b-tag scale factors are varied up and down by the scale factor errors and the invariant mass with up and down variation are constructed, to get an estimate of the systematic effect. This effect is found to be around 2% on signal normalization and around 1% on signal shape.

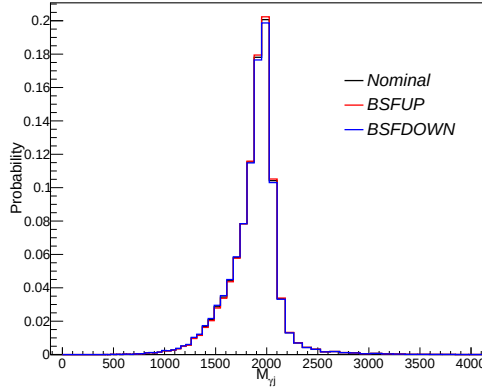


Figure 4.40: Effect of b-tag scale factor errors on the invariant mass of b^* signal.

4.9.8 Signal shape interpolation

The uncertainty due to the interpolation of signal shapes for intermediate masses is accounted to be around 0.5 – 1.0%.

4.9.9 Background Shape Uncertainty

The background shape uncertainty arises due to the difference in the shapes of data and background fit function and is evaluated in the form of uncertainty on the parameters of the analytical fit function. These parameters are considered as the nuisance parameters with uniform priors. A frequentist approach has been considered in which the systematic uncertainties related to these nuisance parameters are taken into account by maximizing the likelihood. A signal-plus-background fit to the data is initially performed to identify a reasonable starting point for the parameter values and to get an estimate of errors associated with them.

The Table 4.14 summarizes the effect of various systematic uncertainties on the signal yield.

4.10 Limit computation

A widely used and recommended procedure to establish discovery or exclusion in the particle physics is based on a frequentist significance test which uses a likelihood ratio as the test statistic. In CMS, this test is performed using a widely recommended and centrally adopted approach, referred to as “Higgs COMBINE Tool” [194]. In this approach, along with the parameters of interest such as the

Source	Effect on signal yield(%)
Integrated luminosity	2.5
Jet energy scale	~ 1
Jet energy resolution	0.2–0.4
Photon energy scale	~ 0.6
Photon energy resolution	0.2–0.4
Pileup	1–2
Photon ID efficiency	~ 2
Trigger efficiency	~ 5
Signal interpolation	0.5–1
PDF choice	1.5–3
b tag SF (only b*)	~ 1
b tag SF normalization (only b*)	~ 2

Table 4.14: Summary of the dominant sources of systematic uncertainties and their effect on the signal yield.

cross-section of the signal process, the models for signal and backgrounds also contain the nuisance parameters whose values are not considered as known a priori but either determined by fitting the data or computed by some other means. In the absence of any significant deviations in the data compared to the standard model backgrounds, the upper limit on the cross-section or lower limit on the signal mass are set.

In order to set limits, a multi-bin counting experiment likelihood which is a product of Poisson distributions corresponding to different bins, have been used:

$$\mathcal{L}(\text{data}|\mu, \theta) = \prod_{i=1}^{n_b} \text{Poisson}(x_i | s_i(\mu, \theta) + b_i(\theta)) \cdot \text{Constraint}(\theta | \bar{\theta}, \delta\theta), \quad (4.16)$$

where μ is the resonance signal strength (also known as the parameter of interest), θ is the vector of nuisance parameters, x_i is the data yield in the i^{th} bin, $s_i(\mu, \theta)$ is the corresponding signal yield, $b_i(\theta)$ is the corresponding background yield and the product runs over n_b , the number of bins. Different constraint terms are used for different type of nuisance parameters as summarized in the Table 4.15.

Using this likelihood, a test statistic as the profile likelihood ratio following the LHC Confidence Levels (CLs) procedure [195], is defined as follows:

$$\tilde{q}_\mu = -2 \log \frac{\mathcal{L}(\text{data}|\mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \quad 0 \leq \hat{\mu} \leq \mu, \quad (4.17)$$

where $\hat{\theta}_\mu$ is the conditional maximum likelihood estimator of θ , assuming that the

Systematic uncertainty	Constraint
Jet energy scale	Gaussian
Photon energy scale	Gaussian
Jet energy resolution	Gaussian
Photon energy resolution	Gaussian
b-Tag scale factors	Gaussian
Luminosity	LogNormal
Trigger	LogNormal
P ₀	Uniform
P ₁	Uniform
P ₂	Uniform
P ₃	Uniform

Table 4.15: Constraint terms associated with different systematic uncertainty nuisance parameters.

signal strength parameter μ and the data yield x_i 's refer to the actual experimental observation or pseudo-data. The pair of parameter estimators $\hat{\mu}$ and $\hat{\theta}$ correspond to the global maximum of the likelihood. The lower constraint $0 \leq \hat{\mu}$ requires the signal rate to be positive while the upper constraint $\hat{\mu} \leq \mu$ is imposed by hand to guarantee a one-sided confidence interval.

In the frequentist paradigm, the systematic uncertainties associated with the nuisance parameters θ are taken into account through likelihood maximization, also known as profiling. In order to derive the observed 95% confidence level upper limit on the signal strength parameter using the modified frequentist CLs method [196, 197] in the asymptotic approximation [198], we consider a value of μ which satisfies

$$\text{CL}_s \equiv \frac{\text{CL}_{s+b}}{\text{CL}_b} = \frac{1 - \Phi(\sqrt{\tilde{q}_\mu})}{\Phi(\sqrt{\tilde{q}_{\mu,A}} - \sqrt{\tilde{q}_\mu})} = \alpha, \quad (4.18)$$

where $\alpha = 0.05$ and $\tilde{q}_{\mu,A}$ is the test statistic evaluated on the Asimov dataset [199] which corresponds exactly to the expected background and the nominal nuisance parameters obtained by setting all fluctuations to zero. The function $\Phi(x)$ is the cumulative distribution function of the standard normal distribution.

A similar expression is used to derive the median expected 95% CL upper limit,

$$\Phi(\sqrt{\tilde{q}_{\mu,A}}) = 1 - 0.5\alpha, \quad (4.19)$$

and the $N\sigma$ uncertainty band around the expected limit is obtained as,

$$\Phi(\sqrt{\tilde{q}_{\mu,A}}) = 1 - \alpha\Phi(N) + N. \quad (4.20)$$

On the contrary, in case of discovery, one tests $\mu = 0$ and measures the “local significance” by the use of a modified test statistic,

$$q_0 = -2 \log \frac{\mathcal{L}(\text{data}|0, \hat{\theta}_0)}{\mathcal{L}(\text{data}|\hat{\mu}, \hat{\theta})}, \quad \hat{\mu} \geq 0. \quad (4.21)$$

In this case, the observed significance is simply given by $Z = \sqrt{q_0}$.

4.11 Results

The results are computed using the “Higgs COMBINE frequentist approach” with systematic uncertainties treated as nuisance parameters. The input signal is used in the form of shapes taken from MC simulation while the background is taken from the parametric fit using a fit function. The results are not computed for masses below 1 TeV since in this range, the uncertainties are associated with signal efficiency owing to the invariant mass selection, $M_{\gamma+\text{jet}} > 700 \text{ GeV}$.

4.11.1 Significance of local excesses in data

The significance of local excesses have been measured in data using the following significance estimator:

$$\text{Sig.} = \sqrt{-2 \ln \left(\frac{\mathcal{L}_b}{\mathcal{L}_{s+b}} \right)}, \quad (4.22)$$

where $\mathcal{L}_b$ and $\mathcal{L}_{s+b}$ respectively are the maximum likelihoods corresponding to the best background-only and signal+background fits to the data.

The significance values are determined only for the upward fluctuations considering positive signal strengths. The systematic uncertainties are included while evaluating the significance. The result of the significance scan for q^* signal for three coupling multipliers, $f = 1.0, 0.5$ and 0.1 as a function of the resonance mass for $\gamma + \text{jet}$ resonance hypothesis is presented in Fig. 4.41 while the corresponding results for b^* signal for all the three couplings are presented in Fig. 4.42.

The largest fluctuation observed in data has a significance of 2.6σ for b^* signal

hypothesis for $f = 0.5$, very small compared to the 5σ requirement of discovery.

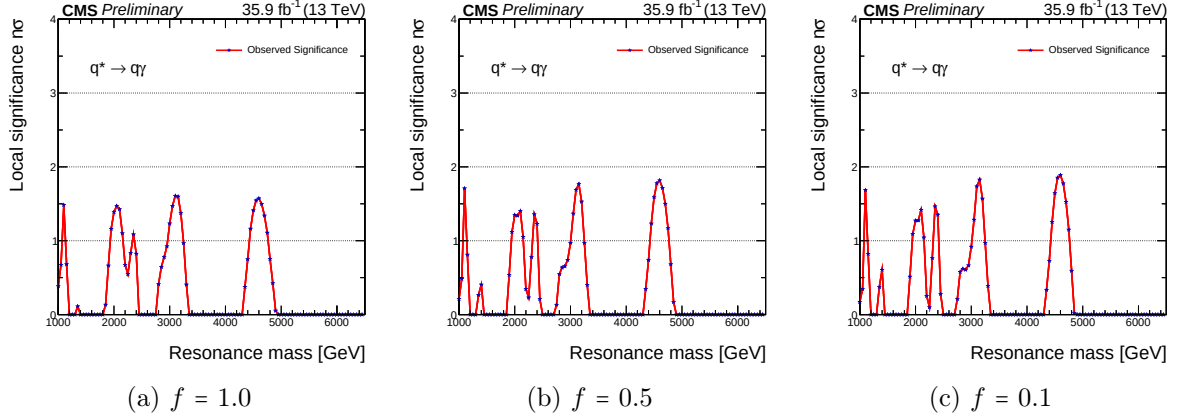


Figure 4.41: The observed significance in data for q^* signal corresponding to the coupling multipliers (a) $f = 1.0$, (b) $f = 0.5$, and (c) $f = 0.1$ as a function of the resonance mass for $\gamma + \text{jet}$ resonance.

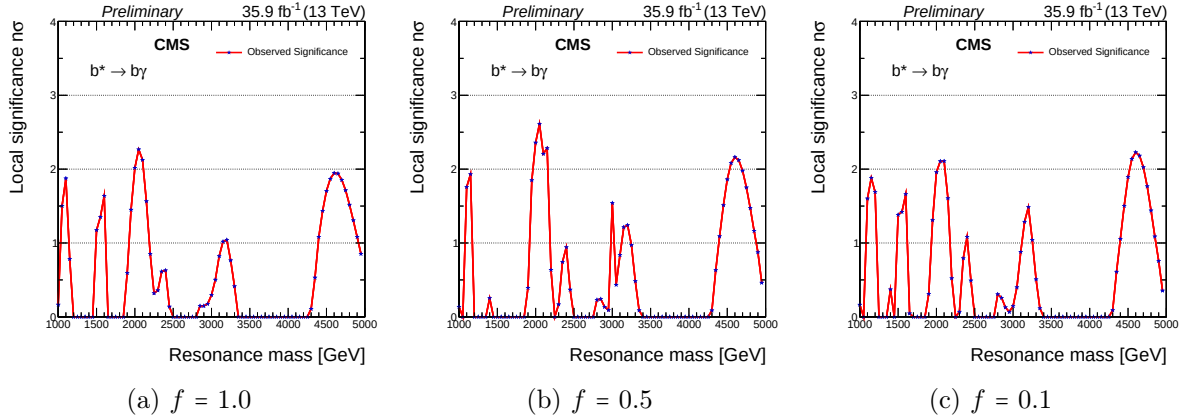


Figure 4.42: The observed significance in data for b^* signal corresponding to the coupling multipliers (a) $f = 1.0$, (b) $f = 0.5$, and (c) $f = 0.1$ as a function of the resonance mass for $\gamma + \text{b-jet}$ resonance.

4.11.2 Exclusion limits

Since no significant excess has been observed in the mass region studied, a 95% CL upper limit on the production cross-sections and lower limit on the masses of q^* and b^* decaying to $q\gamma$ and $b\gamma$ respectively, has been set. The limits are calculated by evaluating the likelihood independently at successive values of resonance mass in the range from 1-6 TeV for q^* and 1-5 TeV for b^* , in steps of 50 GeV. The limits for b^* are evaluated by combining the likelihoods for 1 b-tag and 0 b-tag categories together.

The observed and expected 95% CL upper limits on the product of $\sigma \times B \times A \times \epsilon$ and $\sigma \times B$ for q^* and b^* resonances has been presented in Fig. 4.43 and Fig. 4.44, respectively, for all the three coupling multipliers. Here, σ corresponds to the production cross-section and B corresponds to the branching fraction of $q^* \rightarrow q\gamma$ and $b^* \rightarrow b\gamma$ processes. The product $A \times \epsilon$ refers to the acceptance times efficiency of the q^* and b^* signals as presented in Fig. 4.15 and it varies in the range from 38% to 45% for q^* and 28% to 22% for b^* in the full mass range explored. The limits on $\sigma \times B \times A \times \epsilon$ take into account the detector effects in the form of $A \times \epsilon$ and thus, are referred to as detector dependent limits. On the other hand, the limits on $\sigma \times B$ refers to the detector independent limits. Since the variation of $A \times \epsilon$ over the explored mass range is not very large, the detector dependent and independent limits come out to be almost the same. The observed limits for both q^* and b^* resonances are found to be consistent with the expected limits in the absence of signal.

The observed cross-section upper limits are compared to the LO theoretical predictions, for the three coupling multipliers, in order to estimate the lower mass limits on the q^* and b^* resonances. The observed lower bounds obtained are 5.5 and 1.8 TeV for q^* and b^* respectively, for $f = 1.0$. The corresponding expected mass limits are 5.4 and 1.8 TeV. The observed (expected) limits for $f = 0.5$ and $f = 0.1$ for the q^* resonances have been evaluated to be 4.4 TeV (4.3 TeV) and 2.0 TeV (1.9 TeV) respectively, while the observed (expected) limit for $f = 0.5$ for b^* resonances comes out to be 1.4 TeV (1.3 TeV). No exclusion has been observed for b^* for $f = 0.1$.

It has been observed that $\sigma \times B$ (or $\sigma \times B \times A \times \epsilon$) has very little dependence on f , of the order of 10-20%, so the comparison of $\sigma \times B$ limits corresponding to $f = 1.0$ with the theoretical predictions of lower couplings, yield approximately the same mass limit as can be seen in Fig. 4.45 for q^* and in Fig. 4.46 for b^* . The variation of the excluded mass as a function of the coupling strength, as presented in Fig. 4.47, can be obtained by evaluating the mass exclusions for intermediate couplings through the interpolation of signal efficiencies of the three signal MC samples corresponding to $f = 1.0, 0.5$ and 0.1 . This result can also be interpreted in terms of M_{Res}/Λ , that is, if we left the assumption $\Lambda = M_{\text{Res}}$, the production cross-section of excited quarks is proportional to f as well as M_{Res}/Λ .

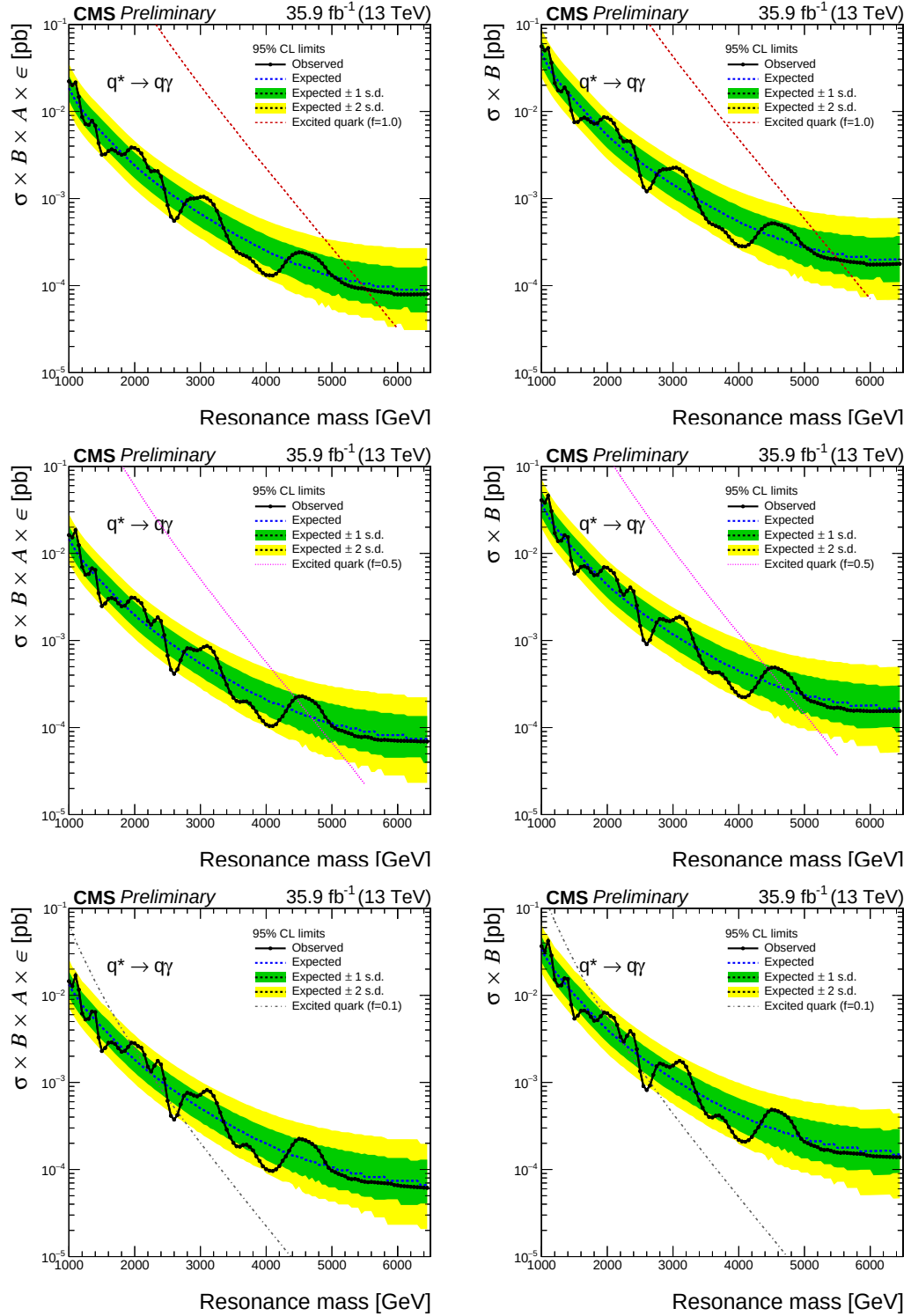


Figure 4.43: The observed and expected 95% CL upper limits on $\sigma \times B \times A \times \epsilon$ (left column) and $\sigma \times B$ (right column) as a function of the mass of q^* resonance, for the coupling multipliers $f = 1.0$ (upper row), $f = 0.5$ (middle row), and $f = 0.1$ (lower row). The limits are also compared with the theoretical predictions for excited quark production. The inner (green) and outer (yellow) bands indicate the regions containing the 68% and 95% of the expected limits.

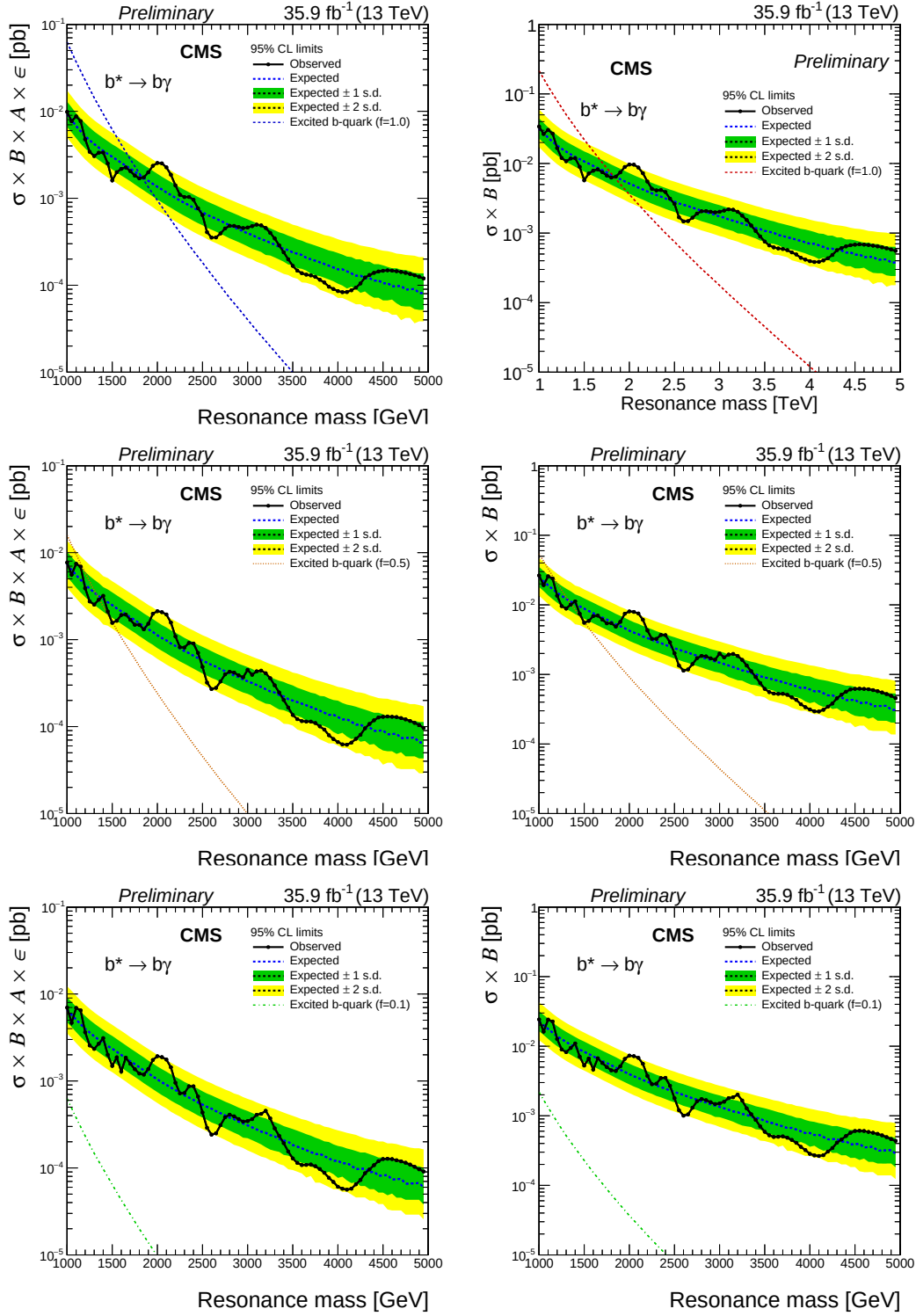


Figure 4.44: The observed and expected 95% CL upper limits on $\sigma \times B \times A \times \epsilon$ (left column) and $\sigma \times B$ (right column) as a function of the mass of b^* resonance, for the coupling multipliers $f = 1.0$ (upper row), $f = 0.5$ (middle row), and $f = 0.1$ (lower row). The limits are also compared with the theoretical predictions for excited b-quark production. The inner (green) and outer (yellow) bands indicate the regions containing the 68% and 95% of the expected limits.

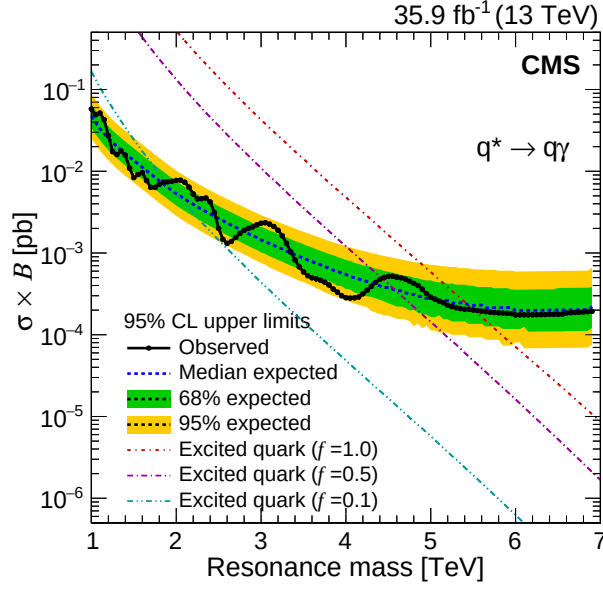


Figure 4.45: The observed and expected 95% CL upper limits on $\sigma \times B$ as a function of the mass of the excited quark, for $f = 1.0$. The limits are also compared with theoretical predictions for excited quark production for the three couplings. The inner (green) band and the outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the expected limits.

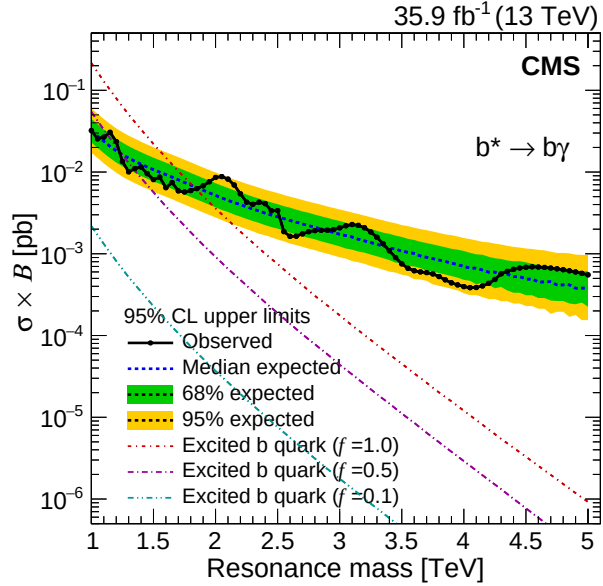


Figure 4.46: The observed and expected 95% CL upper limits on $\sigma \times B$ as a function of the mass of the excited b-quark, for $f = 1.0$. The limits are also compared with theoretical predictions for excited b-quark production for the three couplings. The inner (green) band and the outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the expected limits.

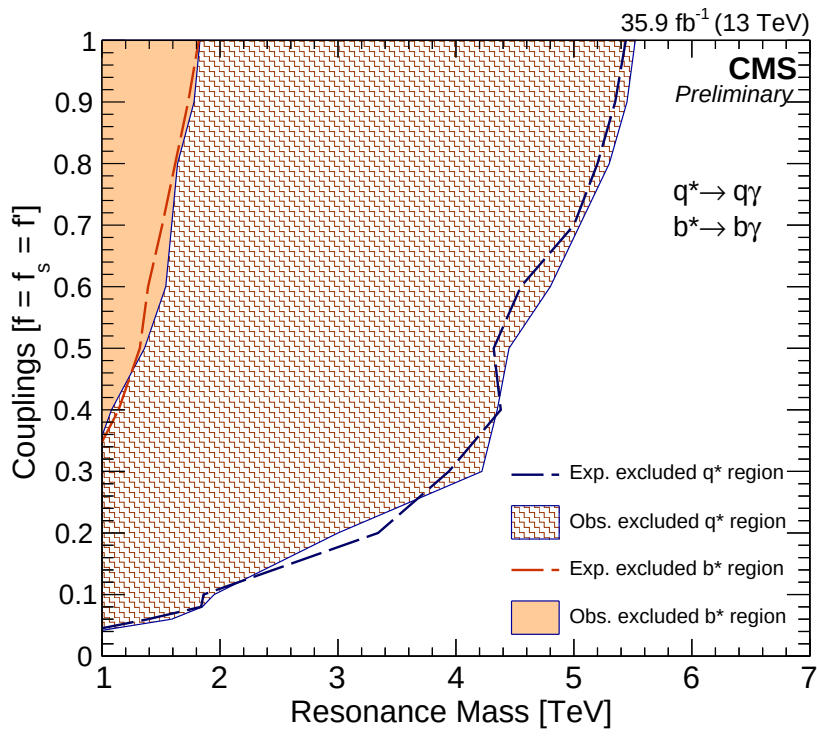


Figure 4.47: The observed and expected regions excluded at 95% CL for q^* and b^* production and decay, as a function of M_{q^*} , M_{b^*} and f .

SUMMARY AND CONCLUSION

This thesis presents a search for the excited states of light u,d and heavy b flavor quarks in the $\gamma + \text{jet}$ and $\gamma + \text{b-jet}$ final states respectively, using the pp collision data at $\sqrt{s} = 13 \text{ TeV}$, collected by the CMS detector in 2016. This data correspond to a total integrated luminosity of 35.9 fb^{-1} .

The underlying processes which have been studied are $pp \rightarrow q^* \rightarrow \gamma + \text{jet}$ and $pp \rightarrow b^* \rightarrow \gamma + \text{b-jet}$. The resonance signals for q^* and b^* , if exist, will show their presence in the form of a bump over the continuous invariant mass distribution spectrum of $\gamma + \text{jet}$ and $\gamma + \text{b-jet}$ respectively. Therefore, the invariant mass distributions are the most important distribution for this study. The signal regions in the mass range from 1 TeV to 9 TeV for q^* and from 1 TeV to 5 TeV for b^* with a small gap of 50 GeV, have been scanned for the presence of any expected signature of excited quarks. Variable mass binning with bin width comparable to the signal mass resolution has been considered for the invariant mass spectra. The event statistics for q^* search has been formed by requiring a high quality photon and a high quality jet. The kinematical restrictions on the selected photons and jets are put in order to optimize the signal selection while reducing the background. In order to form event statistics for b^* search, the b-flavor of selected jets has been checked using a b-tag discriminator. The fraction of jets passing the discriminator form the 1 b-tag category while the rest failing the discriminator form the 0 b-tag category. The likelihoods of two categories have been combined together in b^*

search to avoid any statistical power loss due to the event migration between the two categories. The data events are corrected for pile-up effects and various object energy corrections are applied. The event selection has been optimized for different analysis cuts in order to obtain the best signal discrimination. The highest mass event in data is observed at $M_{\gamma+\text{jet}} = 4.6 \text{ TeV}$.

This analysis has two main backgrounds coming from SM γ +jet and QCD dijet process. The MC samples corresponding to these backgrounds are used to validate the data as per the SM expectations. At every stage of this analysis, the data are found to be in good agreement with the SM background predictions. However, the actual background for this search has been obtained by fitting the data with a smooth parameterization. A significance test has been performed in order to look for any significant deviation in the data compared to the SM background expectations. The most significant deviation observed is at 2.6σ , which can be safely considered as a statistical fluctuation.

In the absence of any expected signal, upper limits on the cross-sections or lower limits on the masses of q^* and b^* have been set at the 95% confidence level using the “Higgs COMBINE Tool” with the frequentist approach in the asymptotic approximation. The cross-section upper limits are compared with the theoretical predictions in order to set the lower mass limits on the excited states of quarks. The excited light quarks within the mass range $1.0 < M_{q^*} < 5.5 \text{ TeV}$ and excited b-quarks within the mass range $1.0 < M_{b^*} < 1.8 \text{ TeV}$ are excluded at the 95% confidence level, corresponding to standard model like coupling strengths ($f = 1.0$).

The limits presented in this thesis are the most sensitive limits for q^ and b^* searches in the γ +jet and γ +b-jet final states. The search for the excited b-quarks has been presented for the first time in any final state at $\sqrt{s} = 13 \text{ TeV}$.*

The results of this study have been accepted for publication in Physics Letters B (Ref. No. PLB-D-17-01732R1).

BIBLIOGRAPHY

- [1] S. L. Glashow, “Partial Symmetries of Weak Interactions”, *Nucl. Phys.* **22** (1961) 579–588, doi:10.1016/0029-5582(61)90469-2.
- [2] A. Salam and J. C. Ward, “Electromagnetic and weak interactions”, *Phys. Lett.* **13** (1964) 168–171, doi:10.1016/0031-9163(64)90711-5.
- [3] S. Weinberg, “A Model of Leptons”, *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.
- [4] D. J. Gross and F. Wilczek, “Ultraviolet Behavior of Nonabelian Gauge Theories”, *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.
- [5] H. D. Politzer, “Reliable Perturbative Results for Strong Interactions?”, *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.
- [6] J. Dalton, “On the absorption of gases by water and other liquids”, *Memoirs of the Literary and Philosophical Society of Manchester* **1** (1805) 271–287, doi:10.1080/14786440608563325.
- [7] A. W. Thackray, “The Origin of Dalton’s Chemical Atomic Theory: Daltonian Doubts Resolved”, *Isis* **57** (1966) 35–55, doi:10.1086/350077.
- [8] J. J. Thomson, “Cathode Rays”, *Philosophical Magazine* **44** (1897) 293.
- [9] J. J. Thomson, “On the Structure of the Atom: an Investigation of the Stability and Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle, with Application of the Results to the Theory of Atomic Structure”, *Philosophical Magazine* **7** (1904) 237–265, doi:10.1080/14786440409463107.

- [10] E. Rutherford, “The scattering of alpha and beta particles by matter and the structure of the atom”, *Phil. Mag.* **21** (1911) 669–688, doi:10.1080/14786440508637080.
- [11] E. Rutherford, “The structure of the atom”, *Nature* **92** (1913) 423.
- [12] J. Chadwick, “Possible Existence of a Neutron”, *Nature* **129** (1932) 312.
- [13] A. Einstein, “On a heuristic viewpoint concerning the production and transformation of light”, *Annalen Phys.* **17** (1905) 132–148.
- [14] A. H. Compton, “A Quantum Theory of the Scattering of X-rays by Light Elements”, *Phys. Rev.* **21** (1923) 483–502, doi:10.1103/PhysRev.21.483.
- [15] A. Einstein, “On the electrodynamics of moving bodies”, *Annalen Phys.* **17** (1905) 891–921, doi:10.1002/andp.200590006.
- [16] P. A. M. Dirac, “The Quantum Theory of the Emission and Absorption of Radiation”, *Proc. R. Soc. Lond. A114* **243** (1927) 710.
- [17] C. D. Anderson, “The Positive Electron”, *Phys. Rev.* **43** (1933) 491–494, doi:10.1103/PhysRev.43.491.
- [18] R. P. Feynman, “Space-time approach to nonrelativistic quantum mechanics”, *Rev. Mod. Phys.* **20** (1948) 367–387, doi:10.1103/RevModPhys.20.367.
- [19] R. P. Feynman, “A Relativistic cutoff for classical electrodynamics”, *Phys. Rev.* **74** (1948) 939–946, doi:10.1103/PhysRev.74.939.
- [20] R. P. Feynman, “The Theory of positrons”, *Phys. Rev.* **76** (1949) 749–759, doi:10.1103/PhysRev.76.749.
- [21] J. S. Schwinger, “On Quantum electrodynamics and the magnetic moment of the electron”, *Phys. Rev.* **73** (1948) 416–417, doi:10.1103/PhysRev.73.416.
- [22] J. S. Schwinger, “Quantum electrodynamics. I A covariant formulation”, *Phys. Rev.* **74** (1948) 1439, doi:10.1103/PhysRev.74.1439.
- [23] S.-I. Tomonaga and J. R. Oppenheimer, “On Infinite Field Reactions in Quantum Field Theory”, *Phys. Rev.* **74** (1948) 224–225, doi:10.1103/PhysRev.74.224.

- [24] H. Yukawa, “On the Interaction of Elementary Particles I”, *Proc. Phys. Math. Soc. Jap.* **17** (1935) 48–57, doi:10.1143/PTPS.1.1. [Prog. Theor. Phys. Suppl.1,1(1935)].
- [25] S. H. Neddermeyer and C. D. Anderson, “Note on the Nature of Cosmic Ray Particles”, *Phys. Rev.* **51** (1937) 884–886, doi:10.1103/PhysRev.51.884.
- [26] C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, “Observations on the Tracks of Slow Mesons in Photographic Emulsions. 1”, *Nature* **160** (1947) 453–456, doi:10.1038/160453a0.
- [27] C. L. Cowan and Reines, “Detection of the free neutrino: A Confirmation”, *Science* **124** (1956) 103, doi:10.1126/science.124.3212.103.
- [28] E. Fermi, “Fermi’s Theory of Beta Decay”, *Z. Phy.* **88** (1934) 161.
- [29] G. D. Rochester and C. C. Butler, “Evidence for the Existence of New Unstable Elementary Particles”, *Nature* **160** (1947) 855–857, doi:10.1038/160855a0.
- [30] V. D. Hopper and S. Biswas, “Evidence Concerning the Existence of the New Unstable Elementary Neutral Particle”, *Phys. Rev.* **80** (Dec, 1950) 1099, doi:10.1103/PhysRev.80.1099.
- [31] A. Bonetti, R. L. Setti, M. Panetti, and G. Tomasini, “On the existence of unstable charged particles of hyperprotonic mass”, *Il Nuovo Cimento* **10** (1953) 1736–1743, doi:10.1007/BF02781667.
- [32] C. M. York, R. B. Leighton, and E. K. Bjornerud, “Direct Experimental Evidence for the Existence of a Heavy Positive V Particle”, *Phys. Rev.* **90** (1953) 167–168, doi:10.1103/PhysRev.90.167.
- [33] E. Cowan, “A V -Decay Event with a Heavy Negative Secondary, and Identification of the Secondary V -Decay Event in a Cascade”, *Physical Review* **94** (1954) 161.
- [34] V. E. Barnes et al., “Observation of a Hyperon with Strangeness -3”, *Phys. Rev. Lett.* **12** (1964) 204–206, doi:10.1103/PhysRevLett.12.204.
- [35] M. Gell-Mann, “A Schematic Model of Baryons and Mesons”, *Phys. Lett.* **8** (1964) 214–215, doi:10.1016/S0031-9163(64)92001-3.

- [36] G. Zweig, “An SU_3 model for strong interaction symmetry and its breaking; Version 2”,. Version 1 is CERN preprint 8182/TH.401, Jan. 17, 1964.
- [37] Y. L. Dokshitzer, “Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics”, *Zh. Eksp. Teor. Fiz* **73** (1977) 1216.
- [38] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, “Advantages of the Color Octet Gluon Picture”, *Phys. Lett.* **47B** (1973) 365, doi:10.1016/0370-2693(73)90625-4.
- [39] TASSO Collaboration, “High energy trends in e^+e^- physics”, *Invited talk at Geneva Conference* (1979).
- [40] J. D. Bjorken and S. L. Glashow, “Elementary Particles and $SU(4)$ ”, *Phys. Lett.* **11** (1964) 255–257, doi:10.1016/0031-9163(64)90433-0.
- [41] S. L. Glashow, J. Iliopoulos, and L. Maiani, “Weak Interactions with Lepton-Hadron Symmetry”, *Phys. Rev.* **D2** (1970) 1285–1292, doi:10.1103/PhysRevD.2.1285.
- [42] E598 Collaboration, “Experimental Observation of a Heavy Particle J”, *Phys. Rev. Lett.* **33** (1974) 1404–1406, doi:10.1103/PhysRevLett.33.1404.
- [43] SLAC-SP-017 Collaboration, “Discovery of a Narrow Resonance in e^+e^- Annihilation”, *Phys. Rev. Lett.* **33** (1974) 1406–1408, doi:10.1103/PhysRevLett.33.1406. [Adv. Exp. Phys.5,141(1976)].
- [44] M. L. Perl et al., “Evidence for Anomalous Lepton Production in e^+e^- Annihilation”, *Phys. Rev. Lett.* **35** (1975) 1489–1492, doi:10.1103/PhysRevLett.35.1489.
- [45] S. W. Herb et al., “Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions”, *Phys. Rev. Lett.* **39** (1977) 252–255, doi:10.1103/PhysRevLett.39.252.
- [46] CDF Collaboration, “Observation of top quark production in $\bar{p}p$ collisions”, *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

- [47] D0 Collaboration, “Observation of the top quark”, *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.
- [48] T. D. Lee and C.-N. Yang, “Question of Parity Conservation in Weak Interactions”, *Phys. Rev.* **104** (1956) 254–258, doi:10.1103/PhysRev.104.254.
- [49] C. S. Wu et al., “Experimental Test of Parity Conservation in Beta Decay”, *Phys. Rev.* **105** (1957) 1413–1414, doi:10.1103/PhysRev.105.1413.
- [50] C.-N. Yang and R. L. Mills, “Conservation of Isotopic Spin and Isotopic Gauge Invariance”, *Phys. Rev.* **96** (1954) 191–195, doi:10.1103/PhysRev.96.191.
- [51] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132–133, doi:10.1016/0031-9163(64)91136-9.
- [52] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- [53] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons”, *Phys. Rev.* **145** (1966) 1156–1163, doi:10.1103/PhysRev.145.1156.
- [54] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- [55] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.
- [56] Gargamelle Neutrino Collaboration, “Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment”, *Phys. Lett.* **B46** (1973) 138–140, doi:10.1016/0370-2693(73)90499-1.
- [57] UA1 Collaboration, “Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV”, *Phys. Lett.* **B122** (1983) 103–116, doi:10.1016/0370-2693(83)91177-2.

- [58] UA2 Collaboration, “Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $p\bar{p}$ Collider”, *Phys. Lett.* **B122** (1983) 476–485, doi:10.1016/0370-2693(83)91605-2.
- [59] UA1 Collaboration, “Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c² at the CERN SPS Collider”, *Phys. Lett.* **B126** (1983) 398–410, doi:10.1016/0370-2693(83)90188-0.
- [60] UA1 Collaboration, “Further Evidence for Charged Intermediate Vector Bosons at the SPS Collider”, *Phys. Lett.* **B129** (1983) 273–282, doi:10.1016/0370-2693(83)90860-2.
- [61] UA2 Collaboration, “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN $p\bar{p}$ Collider”, *Phys. Lett.* **B129** (1983) 130–140, doi:10.1016/0370-2693(83)90744-X.
- [62] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett.* **B716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [63] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.* **B716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [64] CMS Collaboration, “Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV”, (2013). CMS-PAS-HIG-13-005.
- [65] ATLAS Collaboration, “Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25 fb⁻¹ of proton-proton collision data”, (2014). ATLAS-CONF-2014-009.
- [66] Super-Kamiokande Collaboration, “Evidence for oscillation of atmospheric neutrinos”, *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.
- [67] Super-Kamiokande Collaboration, “Constraints on neutrino oscillations using 1258 days of Super-Kamiokande solar neutrino data”, *Phys. Rev. Lett.* **86** (2001) 5656–5660, doi:10.1103/PhysRevLett.86.5656, arXiv:hep-ex/0103033.

- [68] SNO Collaboration, “Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory”, *Phys. Rev. Lett.* **89** (2002) 011301, doi:10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.
- [69] KamLAND Collaboration, “Measurement of neutrino oscillation with KamLAND: Evidence of spectral distortion”, *Phys. Rev. Lett.* **94** (2005) 081801, doi:10.1103/PhysRevLett.94.081801, arXiv:hep-ex/0406035.
- [70] K2K Collaboration, “Evidence for muon neutrino oscillation in an accelerator-based experiment”, *Phys. Rev. Lett.* **94** (2005) 081802, doi:10.1103/PhysRevLett.94.081802, arXiv:hep-ex/0411038.
- [71] MINOS Collaboration, “Observation of muon neutrino disappearance with the MINOS detectors and the NuMI neutrino beam”, *Phys. Rev. Lett.* **97** (2006) 191801, doi:10.1103/PhysRevLett.97.191801, arXiv:hep-ex/0607088.
- [72] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C38** (2014) 090001, doi:10.1088/1674-1137/38/9/090001.
- [73] F. J. Dyson, “The Radiation theories of Tomonaga, Schwinger, and Feynman”, *Phys. Rev.* **75** (1949) 486–502, doi:10.1103/PhysRev.75.486.
- [74] M. E. Peskin and D. V. Schroeder, “An Introduction to quantum field theory”. Addison-Wesley (1995) 842 p, Reading, USA, 1995.
- [75] S. Kluth, “Review of alpha(s) Measurements”, *Conf. Proc.* **C060726** (2006) 449–452, arXiv:hep-ex/0609020. [,449(2006)].
- [76] D. J. Gross and F. Wilczek, “Asymptotically Free Gauge Theories. 1”, *Phys. Rev.* **D8** (1973) 3633–3652, doi:10.1103/PhysRevD.8.3633.
- [77] D. J. Gross and F. Wilczek, “Asymptotically Free Gauge Theories. 2.”, *Phys. Rev.* **D9** (1974) 980–993, doi:10.1103/PhysRevD.9.980.
- [78] H. D. Politzer, “Asymptotic Freedom: An Approach to Strong Interactions”, *Phys. Rept.* **14** (1974) 129–180, doi:10.1016/0370-1573(74)90014-3.
- [79] M. Goldhaber, L. Grodzins, and A. W. Sunyar, “Helicity of Neutrinos”, *Phys. Rev.* **109** (1958) 1015, doi:10.1103/PhysRev.109.1015.

- [80] T. Nakano and K. Nishijima, “Charge Independence for V-particles”, *Prog. Theor. Phys.* **10** (1953) 581, doi:10.1143/PTP.10.581.
- [81] K. Nishijima and Kazuhiko, “Charge Independence Theory of V Particles”, *Prog. Theor. Phys.* **13** (1955) 285, doi:10.1143/PTP.13.285.
- [82] M. Gell-Mann, “The interpretation of the new particles as displaced charge multiplets”, *Nuovo Cim.* **4** (1956) 848, doi:10.1007/BF02748000.
- [83] A. Djouadi, “The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model”, *Phys. Rept.* **457** (2008) 1, doi:10.1016/j.physrep.2007.10.004, arXiv:hep-ph/0503172.
- [84] J. C. Pati, A. Salam, and J. A. Strathdee, “Are Quarks Composite?”, *Phys. Lett.* **B59** (1975) 265, doi:10.1016/0370-2693(75)90042-8.
- [85] E. Eichten, K. D. Lane, and M. E. Peskin, “New Tests for Quark and Lepton Substructure”, *Phys. Rev. Lett.* **50** (1983) 811, doi:10.1103/PhysRevLett.50.811.
- [86] U. Baur, I. Hinchliffe, and D. Zeppenfeld, “Excited Quark Production at Hadron Colliders”, *Int.J.Mod.Phys.* **A2** (1987) 1285, doi:10.1142/S0217751X87000661.
- [87] U. Baur, M. Spira, and P. Zerwas, “Excited Quark and Lepton Production at Hadron Colliders”, *Phys.Rev.* **D42** (1990) 815, doi:10.1103/PhysRevD.42.815.
- [88] ZEUS Collaboration, “A Search for excited fermions in e^+p collisions at HERA”, *Z. Phys.* **C76** (1997) 631, doi:10.1007/s002880050585, arXiv:hep-ex/9708007.
- [89] CDF Collaboration, “Search for excited quarks in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”, *Phys.Rev.Lett.* **72** (1994) 3004, doi:10.1103/PhysRevLett.72.3004.
- [90] CDF Collaboration, “Search for new particles decaying to dijets at CDF”, *Phys.Rev.* **D55** (1997) 5263, doi:10.1103/PhysRevD.55.R5263, arXiv:hep-ex/9702004.
- [91] D0 Collaboration Collaboration, “Search for new particles in the two jet decay channel with the D0 detector”, *Phys.Rev.* **D69** (2004) 111101, doi:10.1103/PhysRevD.69.111101, arXiv:hep-ex/0308033.

- [92] ATLAS Collaboration, “Search for new phenomena in dijet mass and angular distributions from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Lett.* **B754** (2016) 302, doi:10.1016/j.physletb.2016.01.032, arXiv:1512.01530.
- [93] ATLAS Collaboration, “Search for new phenomena with photon+jet events in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *JHEP* **03** (2016) 041, doi:10.1007/JHEP03(2016)041, arXiv:1512.05910.
- [94] CMS Collaboration, “Search for narrow resonances decaying to dijets in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *Phys. Rev. Lett.* **116** (2016), no. 7, 071801, doi:10.1103/PhysRevLett.116.071801, arXiv:1512.01224.
- [95] ATLAS Collaboration, “Search for single b^* -quark production with the ATLAS detector at $\sqrt{s} = 7$ TeV”, *Phys. Lett.* **B721** (2013) 171, doi:10.1016/j.physletb.2013.03.016, arXiv:1301.1583.
- [96] CMS Collaboration, “Search for the production of an excited bottom quark decaying to tW in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *JHEP* **01** (2016) 166, doi:10.1007/JHEP01(2016)166, arXiv:1509.08141.
- [97] CMS Collaboration, “Search for resonances and quantum black holes using dijet mass spectra in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev.* **D91** (2015) 052009, doi:10.1103/PhysRevD.91.052009, arXiv:1501.04198.
- [98] D. S. Gorbunov and V. A. Rubakov, “Introduction to the theory of the early universe: Hot big bang theory”. World Scientific, Hackensack, 2011.
- [99] O. S. Brüning et al., “LHC Design Report”. CERN, Geneva, 2004. CERN-2004-003-V-1.
- [100] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [101] S. Myers and E. Picasso, “The design, construction and commissioning of the CERN large Electron–Positron collider”, *Contemporary Physics* **31** (1990) 387, doi:10.1080/00107519008213789.

- [102] S. Holmes, R. S. Moore, and V. Shiltsev, “Overview of the Tevatron Collider Complex: Goals, Operations and Performance”, *JINST* **6** (2011) T08001, doi:10.1088/1748-0221/6/08/T08001, arXiv:1106.0909.
- [103] ATLAS Collaboration, A. Airapetian et al., “ATLAS detector and physics performance: Technical Design Report, 1”. TDR ATLAS. CERN, Geneva, 1999. ATLAS-TDR-014; CERN-LHCC-99-014.
- [104] CMS Collaboration, G. L. Bayatian et al., “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. TDR CMS. CERN, Geneva, 2006. CERN-LHCC-2006-001.
- [105] ALICE Collaboration, P. Cortese et al., “ALICE physics performance: Technical Design Report”. TDR ALICE. CERN, Geneva, 2005. ALICE-TDR-13; CERN-LHCC-2005-030.
- [106] LHCb Collaboration, R. Antunes-Nobrega et al., “LHCb reoptimized detector design and performance: Technical Design Report”. TDR LHCb. CERN, Geneva, 2003. LHCb-TDR-9; CERN-LHCC-2003-030.
- [107] “CERN Experiments and Facilities”. <http://home.web.cern.ch/about>. Accessed, 2015.
- [108] CMS Collaboration, S. Chatrchyan et al., “CMS Luminosity - Public Results”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>. Accessed: 2014-09-30.
- [109] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [110] CMS Collaboration, V. Karimäki et al., “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [111] 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.
- [112] CMS Collaboration, “Studies of Tracker Material”, (2010). CMS-PAS-TRK-10-003.

- [113] R. Baur and W. H. Bertl, “The CMS pixel vertex detector”, *Nucl. Phys. Proc. Suppl.* **78** (1999) 293, doi:10.1016/S0920-5632(99)00560-5.
- [114] CMS Collaboration, M. Dominguez et al., “CMS Technical Design Report for the Pixel Detector Upgrade”. Technical Design Report CMS. CERN, Geneva, 2012.
- [115] CMS Collaboration, “Upgrade of the CMS tracker”, *JINST* **9** (2014) C03041, doi:10.1088/1748-0221/9/03/C03041.
- [116] CMS Collaboration, “The CMS electromagnetic calorimeter project: Technical Design Report”. TDR CMS. CERN, Geneva, 1997. CERN-LHCC-97-033.
- [117] CMS Collaboration, “Energy Calibration and Resolution of the CMS Electromagnetic Calorimeter in pp Collisions at $\sqrt{s} = 7$ TeV”, *JINST* **8** (2013) P09009, doi:10.1088/1748-0221/8/09/P09009, arXiv:1306.2016.
- [118] R. Wigmans, “Advances in hadron calorimetry”, *Ann. Rev. Nucl. Part. Sci.* **41** (1991) 133, doi:10.1146/annurev.ns.41.120191.001025.
- [119] CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”. TDR CMS. CERN, Geneva, 1997. CERN-LHCC-97-031.
- [120] CMS HCAL Collaboration, “Design, performance, and calibration of CMS hadron-barrel calorimeter wedges”, *Eur.Phys.J.* **C55** (2008) 159–171, doi:10.1140/epjc/s10052-008-0573-y.
- [121] CMS HCAL Collaboration, “Design, performance, and calibration of CMS hadron endcap calorimeters”, (2008). CERN-CMS-NOTE-2008-010.
- [122] G. Bayatian et al., “Design, performance and calibration of the CMS forward calorimeter wedges”, *Eur.Phys.J.* **C53** (2008) 139–166, doi:10.1140/epjc/s10052-007-0459-4.
- [123] CMS HCAL Collaboration, “Design, performance, and calibration of the CMS Hadron-outer calorimeter”, *Eur.Phys.J.* **C57** (2008) 653–663, doi:10.1140/epjc/s10052-008-0756-6.
- [124] CMS Collaboration, “The CMS magnet project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.

- [125] CMS Collaboration, “The CMS muon project: Technical Design Report”. TDR CMS. CERN, Geneva, 1997. CERN-LHCC-97-032.
- [126] CMS Collaboration, G. L. Bayatyan et al., “CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems”. TDR CMS. CERN, Geneva, 2000. CMS-TDR-6-1; CERN-LHCC-2000-038.
- [127] CMS Collaboration, S. Cittolin et al., “CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project”. TDR CMS. CERN, Geneva, 2002. CMS-TDR-6; CERN-LHCC-2002-026.
- [128] C. Eck et al., “LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)”. Technical Design Report LCG. CERN, Geneva, 2005.
- [129] M. Giffels et al., “The CMS Data Management System”, *J. Phys. Conf. Ser.* **513** (2014) 042052, doi:10.1088/1742-6596/513/4/042052.
- [130] R. Egeland, T. Wildish, and S. Metson, “Data transfer infrastructure for CMS data taking”, *Proceedings of Science* **ACAT08** (2008) 033.
- [131] R. Egeland, T. Wildish, and C.-H. Huang, “PhEDEx Data Service”, Technical Report CMS-CR-2009-071, CERN, Geneva, (2009). <https://cds.cern.ch/record/1196164>.
- [132] CMS Collaboration, “The CMS dataset bookkeeping service”, *J. Phys. Conf. Ser.* **119** (2008) 072001, doi:10.1088/1742-6596/119/7/072001.
- [133] V. Kuznetsov, D. Evans, and S. Metson, “The CMS data aggregation system”, *Procedia Computer Science* **1** (2010) 1535, doi:10.1016/j.procs.2010.04.172.
- [134] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework”, *Nucl. Instrum. Meth.* **A389** (1997) 81, doi:10.1016/S0168-9002(97)00048-X.
- [135] CMS Collaboration, “The CMS reconstruction software”, *J. Phys. Conf. Ser.* **331** (2011) 032020, doi:10.1088/1742-6596/331/3/032020.
- [136] A. Buckley et al., “General-purpose event generators for LHC physics”, *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

- [137] M. Dobbs et al., “Les Houches guidebook to Monte Carlo generators for hadron collider physics”, [arXiv:hep-ph/0403045](#).
- [138] C. P. Robert, “Monte Carlo Methods”,
[doi:10.1002/9781118445112.stat03876.pub2](#).
- [139] 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](#).
- [140] CMS Collaboration, “Track Reconstruction in the CMS tracker”, Technical Report CMS-NOTE-2006-041, CERN, Geneva, (2006).
<https://cds.cern.ch/record/934067>.
- [141] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting”,
Nucl.Instrum.Meth. **A262** (1987) 444–450,
[doi:10.1016/0168-9002\(87\)90887-4](#).
- [142] P. Billoir, “Progressive track recognition with a Kalman like fitting procedure”, *Comput. Phys. Commun.* **57** (1989) 390–394,
[doi:10.1016/0010-4655\(89\)90249-X](#).
- [143] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the Kalman filtering method”, *Nucl. Instrum. Meth.* **A294** (1990) 219,
[doi:10.1016/0168-9002\(90\)91835-Y](#).
- [144] CMS Collaboration, “Beam Position Determination using Tracks”, Technical Report CMS-NOTE-2007-021, CERN, Geneva, (2007).
<https://cds.cern.ch/record/1061285>.
- [145] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression, and Related Optimization Problems”,
Proceedings of the IEEE **86** (1998) [doi:10.1109/5.726788](#).
- [146] CMS Collaboration, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, [arXiv:1502.02702](#).
- [147] CMS Collaboration, “Review of clustering algorithms and energy corrections in ECAL”, technical report, CERN, Geneva, (2010).
<https://cds.cern.ch/record/1365024>.

- [148] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, “Electron Reconstruction in the CMS Electromagnetic Calorimeter”, Technical Report CMS-NOTE-2001-034, CERN, Geneva, (2001).
<https://cds.cern.ch/record/687345>.
- [149] S. Catani, Y. L. Dokshitzer, M. Seymour, and B. Webber, “Longitudinally invariant k_t clustering algorithms for hadron hadron collisions”, *Nucl.Phys.* **B406** (1993) 187, doi:10.1016/0550-3213(93)90166-M.
- [150] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti- k_t jet clustering algorithm”, *JHEP* **0804** (2008) 063,
doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [151] CMS Collaboration, “A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging”, Technical Report CMS-PAS-JME-09-001, CERN, Geneva, (2009). <http://cds.cern.ch/record/1194489>.
- [152] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, “Better jet clustering algorithms”, *JHEP* **9708** (1997) 001,
doi:10.1088/1126-6708/1997/08/001, arXiv:hep-ph/9707323.
- [153] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, (2009).
<https://cds.cern.ch/record/1194487>.
- [154] CMS Collaboration, “Plans for Jet Energy Corrections at CMS”, Technical Report CMS-PAS-JME-07-002, CERN, 2008. Geneva, (2008).
<https://cds.cern.ch/record/1194485>.
- [155] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys.Lett.* **B659** (2008) 119, doi:10.1016/j.physletb.2007.09.077,
arXiv:0707.1378.
- [156] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013,
doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.
- [157] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, (2017). arXiv:1712.07158. Submitted to *JINST*.

- [158] CMS Collaboration, “Jet Identification in the CMS detector”.
<https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetID>. Accessed:
 2016.
- [159] J. Shiers, “The Worldwide LHC Computing Grid (worldwide LCG)”,
Computer Physics Communications **177** (2007) 219,
 doi:10.1016/j.cpc.2007.02.021. Proceedings of the Conference on
 Computational Physics 2006.
- [160] C. P. Robert, “Monte Carlo Methods”. Wiley Online Library, 2004.
- [161] A. Martin, W. Stirling, R. Thorne, and G. Watt, “Parton distributions for
 the LHC”, *Eur.Phys.J.* **C63** (2009) 189–285,
 doi:10.1140/epjc/s10052-009-1072-5, arXiv:0901.0002.
- [162] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys.*
Commun. **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024,
 arXiv:1410.3012.
- [163] J. Alwall et al., “The automated computation of tree-level and
 next-to-leading order differential cross sections, and their matching to
 parton shower simulations”, *JHEP* **1407** (2014) 079,
 doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [164] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur.Phys.J.* **C58** (2008)
 639–707, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [165] M. L. Mangano et al., “ALPGEN, a generator for hard multiparton
 processes in hadronic collisions”, *JHEP* **0307** (2003) 001,
 doi:10.1088/1126-6708/2003/07/001, arXiv:hep-ph/0206293.
- [166] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *JHEP* **0902**
 (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.
- [167] T. Sjöstrand and M. Bengtsson, “The Lund Monte Carlo for Jet
 Fragmentation and e+ e- Physics. Jetset Version 6.3: An Update”, *Comput.*
Phys. Commun. **43** (1987) 367, doi:10.1016/0010-4655(87)90054-3.
- [168] P. de Aquino et al., “ALOHA: Automatic Libraries Of Helicity Amplitudes
 for Feynman Diagram Computations”, *Comput. Phys. Commun.* **183**
 (2012) 2254–2263, doi:10.1016/j.cpc.2012.05.004, arXiv:1108.2041.

- [169] J. Alwall et al., “A Standard format for Les Houches event files”, *Comput.Phys.Commun.* **176** (2007) 300–304, doi:10.1016/j.cpc.2006.11.010, arXiv:hep-ph/0609017.
- [170] F. Krauss, “Matrix elements and parton showers in hadronic interactions”, *JHEP* **0208** (2002) 015, doi:10.1088/1126-6708/2002/08/015, arXiv:hep-ph/0205283.
- [171] M. L. Mangano, M. Moretti, and R. Pittau, “Multijet matrix elements and shower evolution in hadronic collisions: $Wb\bar{b} + n$ jets as a case study”, *Nucl.Phys.* **B632** (2002) 343–362, doi:10.1016/S0550-3213(02)00249-3, arXiv:hep-ph/0108069.
- [172] GEANT4 Collaboration, “GEANT4 — a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [173] CMS Collaboration, “CRAB: Distributed analysis tool for CMS”, Technical Report CMS-CR-2009-318, CERN, Geneva, (2009). <https://cds.cern.ch/record/1358821>.
- [174] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [175] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.
- [176] R. D. Ball et al., “Parton distributions with LHC data”, *Nucl. Phys. B* **867** (2013) 244, doi:10.1016/j.nuclphysb.2012.10.003, arXiv:1207.1303.
- [177] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J. C* **53** (2008) 473, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [178] CMS Collaboration, S. Chatrchyan et al., “Pile-up scenario for Run-II”. <https://twiki.cern.ch/twiki/bin/view/CMS/PdmVPileUpDescription>, https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJSONFileforData#Pileup_JSON_Files_For_Run_II.

- [179] 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.
- [180] CMS Collaboration, S. Chatrchyan et al., “B-tag Recommendation”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation80X>.
- [181] S. D. Ellis, Z. Kunszt, and D. E. Soper, “Two jet production in hadron collisions at order α_s^3 in QCD”, *Phys.Rev.Lett.* **69** (1992) 1496–1499, doi:10.1103/PhysRevLett.69.1496.
- [182] W. Giele, E. N. Glover, and D. A. Kosower, “Higher order corrections to jet cross-sections in hadron colliders”, *Nucl.Phys.* **B403** (1993) 633–670, doi:10.1016/0550-3213(93)90365-V, arXiv:hep-ph/9302225.
- [183] CMS Collaboration, S. Chatrchyan et al., “Physics Analysis Oriented Event Display”. <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookFireworks>. Accessed: 2015.
- [184] S. Ozturk et al., “Plans to Search for New Particles Decaying to Dijets in pp Collisions at $\sqrt{s} = 10$ TeV”, (2009). CMS AN-2009/070.
- [185] CDF Collaboration, “Search for new particles decaying into dijets in proton-antiproton collisions at $\sqrt{s} = 1.96$ TeV”, *Phys. Rev. D* **79** (2009) 112002, doi:10.1103/PhysRevD.79.112002, arXiv:0812.4036.
- [186] R. M. Harris and K. Kousouris, “Searches for Dijet Resonances at Hadron Colliders”, *Int. J. Mod. Phys.* **A26** (2011) 5005, doi:10.1142/S0217751X11054905, arXiv:1110.5302.
- [187] ATLAS Collaboration, “Search for production of resonant states in the photon-jet mass distribution using pp collisions at $\sqrt{s} = 7$ TeV collected by the ATLAS detector”, *Phys. Rev. Lett.* **108** (2012) 211802, doi:10.1103/PhysRevLett.108.211802, arXiv:1112.3580.
- [188] CMS Collaboration, “Search for narrow resonances using the dijet mass spectrum in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev.* **D87** (2013) 114015, doi:10.1103/PhysRevD.87.114015, arXiv:1302.4794.
- [189] ATLAS Collaboration, “Search for new phenomena in photon+jet events collected in proton–proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS

- detector”, *Phys. Lett.* **B728** (2014) 562,
doi:10.1016/j.physletb.2013.12.029, arXiv:1309.3230.
- [190] CMS Collaboration, “Search for excited quarks in the γ +jet final state in proton proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett.* **B738** (2014) 274,
doi:10.1016/j.physletb.2014.09.048, arXiv:1406.5171.
- [191] S. Baker and R. D. Cousins, “Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms”, *Nucl. Instrum. Meth.* **221** (1984) 437–442, doi:10.1016/0167-5087(84)90016-4.
- [192] P. D. G. Collaboration, “Review of Particle Physics”, *Chin. Phys.* **C40** (2016) no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.
- [193] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS”, *JINST* **6** (2011) P11002,
doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [194] CMS Collaboration, “Higgs Combine Toolkit”.
<https://cms-hcomb.gitbooks.io/combine/content/>. Accessed: 2016.
- [195] The ATLAS Collaboration, The CMS Collaboration, The LHC Higgs Combination Group Collaboration, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, CERN, Geneva, (Aug, 2011). <https://cds.cern.ch/record/1379837>.
- [196] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435,
doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [197] A. L. Read, “Presentation of search results: The CL_s technique”, *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [198] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, “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: *Eur. Phys. J. C* **73** (2013) 2501, 10.1140/epjc/s10052-013-2501-z].
- [199] I. Asimov, “Isaac Asimov: The Complete Stories, Vol. 1”. Broadway Books, New York City, USA, 1990.