

Search for New Physics in Dijet Angular Distributions in Proton-Proton
Collisions at 13 TeV

by

Jingyu Zhang
B.A. (Beihang University) 2011
M.S. (Illinois Institute of Technology) 2013

Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Physics
in the Graduate College of the
University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Nikos Varelas, Chair and Advisor
Richard Cavanaugh
David Hofman
James Unwin
Vivian O'Dell, FNAL

Copyright by

Jingyu Zhang

2018

To myself,

and my parents.

ACKNOWLEDGMENTS

First of all, I would like to thank my advisor Prof. Nikos Varelas, for his advice and support during my Ph.D. studies. He suggested working on the trigger project and the dijet angular analysis, which resulted in a very challenging but rewarding experience. He supported me to attend many summer schools and conferences, from which I learned a lot. I also would like to thank him for his very helpful comments on this thesis.

My warmest gratitude goes to Dr. Leonard Apanasevich. He was the person who got me started with my research. He taught me a lot of research skills such as data analysis, programming, and scientific writing techniques. Whenever I had questions, he was always there explaining everything to me with his expertise and patience. Without him, I could not finish what I have accomplished today.

I would also like to thank all the UIC faculty members who are working on the CMS experiment: Prof. Cecilia Gerber, Prof. Richard Cavanaugh, Prof. Mark Adams, and Prof. Corrine Mills. It was my pleasure to work in the UIC CMS group.

In addition, I thank my colleagues in the CMS collaboration. In particular, I thank Dr. Andreas Hinzman and Prof. Klaus Rabbertz for working together on the dijet angular analysis and Prof. Costas Foudas for his support during the time when I was working on the Level-1 Trigger project.

TABLE OF CONTENTS

<u>CHAPTER</u>		<u>PAGE</u>
1	INTRODUCTION	1
1.1	The Standard Model	2
1.1.1	The Elementary Particles	2
1.1.2	The Fundamental Interactions	4
1.2	New Physics Models	11
1.2.1	Quark Contact Interactions	11
1.2.2	Extra Spatial Dimensions	13
1.2.3	Quantum Black Holes	17
1.2.4	The Simplified Model for the Dark Matter	18
2	QUANTUM CHROMODYNAMICS AND DIJET ANGULAR DISTRIBUTIONS	24
2.1	Quantum Chromodynamics	24
2.1.1	The Lagrangian of Quantum Chromodynamics	24
2.1.2	Perturbative Quantum Chromodynamics and the Running of the Strong Coupling Constant	26
2.1.3	The Factorization Theorem	29
2.2	Jet Production	31
2.2.1	The Hard Scattering Process	33
2.2.2	Monte Carlo Event Generators	34
2.3	Jet Algorithms	40
2.4	Dijet Angular Distributions	45
2.4.1	The Kinematics of Dijet Events	45
2.4.2	The Dijet Angular Distributions	48
3	THE LARGE HADRON COLLIDER AND THE COMPACT MUON SOLENOID DETECTOR	53
3.1	The Large Hadron Collider	53
3.2	The Compact Muon Solenoid Detector	57
3.2.1	The Tracker	60
3.2.2	The Calorimeter Detectors	63
3.2.2.1	The Electromagnetic Calorimeter	65
3.2.2.2	The Hadronic Calorimeter	67
3.2.3	The Muon Chambers	69
4	THE CMS TRIGGER SYSTEM	72
4.1	The Level-1 Trigger	75

TABLE OF CONTENTS (Continued)

<u>CHAPTER</u>		<u>PAGE</u>
4.1.1	The Level-1 Calorimeter Trigger	75
4.1.1.1	The Level-1 Trigger e/γ and τ Reconstruction Algorithms . .	77
4.1.1.2	The Level-1 Trigger Jet and Energy Sums Reconstruction Algorithms	80
4.1.2	The Level-1 Muon Trigger	82
4.1.3	The Level-1 Global Trigger	84
4.1.4	Data Quality Monitoring of the Level-1 Trigger	85
4.2	The High-Level Trigger	90
4.3	The Offline Computing Systems	92
5	JET RECONSTRUCTION AND CALIBRATION	94
5.1	The Particle Flow Technique	94
5.1.1	Reconstruction of the Input Elements to the Particle Flow Algorithm	97
5.1.2	Link algorithm	99
5.1.3	Particle Flow Candidates	100
5.2	Particle Flow Jet Reconstruction and Calibration	100
5.2.1	Pileup Mitigation	102
5.2.2	Identification and Rejection of Fake Jets	104
5.2.3	Jet Energy Scale Calibration	105
5.2.3.1	Offset Correction	107
5.2.3.2	Monte Carlo Truth Correction	107
5.2.3.3	Residual Correction	109
5.2.4	Jet Energy Resolution	113
6	MEASUREMENT OF THE DIJET ANGULAR DISTRIBUTIONS	115
6.1	Event Selection	115
6.2	Comparison with the Detector Simulation	119
6.3	Jet Energy Resolution Modeling	132
6.4	Cross Section Unfolding	139
6.5	Experimental Uncertainties	149
6.5.1	Jet Energy Scale Uncertainty	149
6.5.2	Jet Energy Resolution Uncertainty	150
6.5.3	Uncertainty Due to the Modeling of the Jet Energy Resolution	156
6.5.4	Uncertainty Introduced Due to the Unfolding Procedure	156
7	THEORETICAL PREDICTIONS	161
7.1	Theoretical Predictions for Quantum Chromodynamics	161
7.1.1	Theoretical Uncertainties	161
7.1.2	Electroweak Corrections	164
7.1.3	Non-perturbative Corrections	165
7.2	Contact Interaction Theoretical Predictions	165

TABLE OF CONTENTS (Continued)

<u>CHAPTER</u>		<u>PAGE</u>
	7.3	Theoretical Predictions for Extra Spatial Dimensions 167
	7.4	Quantum Black Hole Theoretical Predictions 168
	7.5	Dark Matter Theoretical Predictions 168
8	RESULTS	172
	8.1	Comparisons between Data and Theoretical Predictions 172
	8.2	Determination of Exclusion Limits 175
	8.2.1	Statistical Method 175
	8.2.2	Limits on New Physics Models 181
9	CONCLUSION	189
	APPENDICES	190
	Appendix A	191
	Appendix B	196
	Appendix C	201
	CITED LITERATURE	202

LIST OF TABLES

<u>TABLE</u>		<u>PAGE</u>
I	Summary of the SM fundamental interactions and their strengths.	5
II	Quark CI models.	12
III	LO matrix element for QCD $2 \rightarrow 2$ subprocesses.	36
IV	The square of the LO matrix element for the QCD $2 \rightarrow 2$ subprocesses expressed in χ	50
V	The “tight” PF jet quality criteria.	104
VI	The number of events in each M_{jj} bins after all the analysis selection criteria are applied.	118
VII	The leading experimental uncertainties in the normalized χ_{dijet} distributions, in percent. While in the statistical analysis (see Chapter 8), each systematic uncertainty is represented through a change in the χ_{dijet} distributions correlated across all the M_{jj} and χ_{dijet} bins, representative values which are taken as the relative shifts for the lowest χ_{dijet} bins in the lowest and highest M_{jj} bins, are used to illustrate the relative contributions in this table.	159
VIII	The theoretical systematic uncertainties in the normalized χ_{dijet} distributions, in percent. While in the statistical analysis (see Chapter 8), each systematic uncertainty is represented through a change in the χ_{dijet} distributions correlated across all the M_{jj} and χ_{dijet} bins, representative values which are taken as the relative shifts for the lowest χ_{dijet} bins in the lowest and highest M_{jj} bins, are used to illustrate the relative contributions in this table.	163
IX	Parameters used in the PYTHIA prediction of the ADD virtual graviton exchange process.	167
X	Parameters used in the QBH prediction of the QBH production and decay processes.	168
XI	Agreement between the data distributions and theoretical predictions. The χ^2 in each M_{jj} bin is divided by the number of degrees of freedom (DOF).	175
XII	Significance of the data distributions in number of standard deviations for the QCD+DM models with $g_q = 1.0$	180
XIII	Observed and expected exclusion limits at 95% CL for CI, ADD, and QBH models.	181
XIV	Same as Table XIII but obtained using the unfolded distributions and particle-level predictions.	198
XV	Fitted signal strengths from fits to the data+DM distributions compared to the fitted signal strengths from the fits to the data-only distributions.	199

LIST OF FIGURES

<u>FIGURE</u>		<u>PAGE</u>
1	The elementary particles in the SM.	3
2	Diagrams for a four-fermion contact interaction (left) describing the low energy effective field theory for a new interaction between the fermions (right).	12
3	The Evidence of DM. Left: the measured galactic rotation curve for M 33 galaxy compared to the prediction obtained using only the visible matter in the galaxy. At distances close to the center of the galaxy, the orbital velocities are measured using starlight, and at large radii, the 21 cm hydrogen line (microwave created by electromagnetic radiation of a neutral hydrogen atom) is used to measure the velocities of the interstellar hydrogen. Right: the optical image from Magellan and Hubble Space Telescope for the bullet cluster 1E 0657-558. Also shown are the hot gas (pink) and mass (blue) distributions in the bullet cluster. The measurements of the hot gas and mass clusters are provided by the Chandra X-Ray Observatory and gravitational lensing, respectively. . .	20
4	The interaction between quarks and DM particles (left) and quarks (right) mediated by the DM mediator in the simplified DM model. As DM particles cannot be directly detected by particle detectors, the initial state radiation shown in the left plot provides the mono-jet signature for the DM decay at the LHC.	23
5	Fundamental coupling of QCD: quark-gluon interaction (left) and gluon self interaction terms (middle and right).	25
6	Summary of the measurements of the QCD running coupling.	27
7	The one-loop corrections to the quark (left) and gluon (middle and right) self-energies.	28
8	A pictorial representation of two partons i and j originated from two protons participating a hard scattering process $ij \rightarrow ab$. The momentum fraction distributions of partons i and j within the two protons are parametrized by the PDFs.	30
9	The proton PDF for $u, \bar{u}, d, \bar{d}, s$ and g at 2 (left) and 100 (right) GeV. The PDF for s and \bar{s} are the same.	32
10	Summary of the Feynman diagrams of the QCD $2 \rightarrow 2$ subprocesses.	35
11	Various processes in a collision event including the hard process (shown in red), ISR and FSR (shown in blue), hadronization (shown in green), decays and QCD Bremsstrahlung of unstable hadrons (shown in yellow and dark green, respectively), and UE (shown in purple). . . .	37

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
12	An example of the infrared- and collinear-unsafety behaviors of the cone algorithms. Top: the collinear splitting of the jet seed leads to the seed p_T below seed threshold. Bottom: the emission of a soft gluon leads to the merging of two jets.	41
13	The workflow chart for the sequential recombination algorithm. The algorithm starts with N objects (particles), which are given in the form of the 4-momentum p_i , and produces K jets.	43
14	The jet areas for the k_t , CA, and anti- k_t jet algorithms.	44
15	An illustration of a dijet event in the partonic center-of-mass (left) and lab (right) frames.	46
16	The square of the LO matrix element for the QCD $2 \rightarrow 2$ subprocesses, expressed in χ , and divided by $(1 + \chi)^2$. For subprocesses which have the same initial states, the matrix elements are combined.	51
17	An illustration of the χ distributions from Rutherford scattering, QCD, and NP models. NP models, which have a more isotropic scattering angle distribution than those given by QCD, can be identified as an excess at small values of χ	52
18	The LHC accelerator and detectors at the LHC.	54
19	The instantaneous luminosity (left) and integrated luminosity (right) for the LHC in 2016.	56
20	The layout of the CMS Detector.	59
21	A slice of the CMS detector in the x-y plane, illustrating the signatures for different types of particles in the detector.	60
22	An illustration of a measurement of the momentum of a charged particle.	61
23	Schematic cross section view of the CMS tracker in the x-y plane and positive z direction in LHC Run I. In 2016 and 2017, the endcap and barrel pixel tracker were upgraded to have 3 and 4 layers, respectively.	62
24	Schematic cross-sectional view of the CMS ECAL in the x-y plane and positive z direction.	67
25	Schematic of the CMS Detector in the x-y plane and positive z direction, illustrating the layout of the HCAL detectors.	68
26	Layout of the CMS muons chambers in the x-y plane and positive z direction.	70
27	The cross section and event rate for various physical processes at different collision energies.	73
28	Schematic view of the two-tiered CMS trigger system.	74
29	Schematic view of the L1T.	76
30	The L1 Calorimeter Trigger layout.	78

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
31	An illustration of the L1T electron and photon reconstruction algorithm. A candidate is formed by clustering a seed TT (red) with neighbor TTs (orange and yellow). The isolation criterion is based on the sum of the energies in the isolation region (blue) in ECAL and HCAL. The isolation region is formed as a 5×9 TT region around the seed. The TTs that contain the footprint of the candidate (white) are excluded.	79
32	An illustration of the L1T jet reconstruction algorithm. A jet is reconstructed based on a 9×9 TT sliding window (blue) algorithm centered on a local maximum (green). The pileup energies are mitigated by subtracting the energy sums in the three of the four 3×9 PileUp Subtraction (PUS) regions (magenta) that have the lowest energies. To avoid double-counting of the jets, in cases where two towers in the region have the same energy, the energy of neighboring TTs must satisfy the inequalities shown in the veto (purple) and mask (blue) regions.	81
33	A comparison of the jets E_T from L1T and offline reconstructions in a various of pileup conditions (left).	82
34	The monitoring of ECAL (left) and HCAL (right) TPs in the Layer-1 Calorimeter Trigger.	86
35	The monitoring of jets E_T (top) and H_T (bottom) distributions in the Layer-2 Calorimeter Trigger.	87
36	The summary of bit-by-bit data and emulation comparisons for the E_T of the Layer-1 TTs, the ratio of the ECAL and HCAL energies in the Layer-1 TTs, and the number of active Layer-1 TTs. The bins in the histogram are all empty indicating perfect agreement between data and emulator.	88
37	The ratio of jets E_T (top) and H_T (bottom) distributions from data and the Layer-2 Calorimeter Trigger emulator.	89
38	An illustration of an HLT path.	91
39	An illustration of particles reconstructed by the PF algorithm for a simulated jet containing a π^+ , a π^- , two photons (from a π^0 decay), and a K_L^0 . Event displays in the ECAL and HCAL are shown in the left and right of the figure, respectively. Also shown in the figure are green lines (dots) depicting the reconstructed tracks (and their associated hits) for the charged pions. The calorimeter cells are shown by the squares in the figure. The energy deposits within the cells are displayed in grey with the size of the squares proportional to the energy in the cell.	96
40	PF jet constituents as a function of jet p_T for jets with $ \eta < 1.3$ (left) and jet η for jets with $56 < p_T < 74$	101
41	The jet p_T response (left) and jet p_T resolution (right) for PF jets (red) and calo jets (blue) with $ \eta < 1.3$	102

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
42	An illustration of the pileup effects on jet clustering. The particle originated from adjacent pileup interactions may be clustered into jets associated to the primary interaction.	103
43	The JES calibration workflow applied to data (top row) and simulation (bottom row).	106
44	Average offset energy per additional pileup interaction in data and simulation derived using the RC method.	108
45	The simulated jet response as a function of η for various jet transverse momenta.	109
46	The relative (left) and absolute (jet) jet p_T corrections.	111
47	The JER derived in simulation for jets in $ \eta < 0.5$ (left) and $3.2 < \eta < 4.7$ (right).	114
48	The trigger turn-on curve for the combined single jet $p_T > 450$ GeV and $H_T > 900$ GeV triggers as a function of M_{jj} in the analysis phase space ($y^* < 1.39$ and $y_{\text{boost}} < 1.11$).	116
49	Illustration of the dijet event topologies in the plane of y_{boost} (y_b) and y^*	117
50	Fits of the trigger efficiencies vs. M_{jj} for several χ_{dijet} bins. The trigger is fully efficient for $M_{jj} > 1.9$ TeV and $\chi_{\text{dijet}} < 9$	118
51	Event display for the event with the highest M_{jj} observed in the analysis phase space.	120
52	Comparison of M_{jj} distributions in data (points) and the Full Sim (histograms) from the PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red) event generators.	121
53	Comparison of p_T distributions for the leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	123
54	Comparison of p_T distributions for the 2 nd -leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	124
55	Comparison of rapidity distributions for the leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	125
56	Comparison of rapidity distributions for the 2 nd -leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	126
57	Comparison of y_{boost} distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	127
58	Comparison of χ_{dijet} distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	128

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
59	Comparison of $\frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}$ distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	129
60	Comparison of $\Delta\phi_{1,2}$ distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	130
61	Comparison of $E_T^{\text{miss}} / \sum E_T$ distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO(red).	131
62	Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $ y < 0.5$. The Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response are also shown.	133
63	Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $0.5 < y < 1.0$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.	134
64	Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $1.0 < y < 1.5$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.	135
65	Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $1.5 < y < 1.8$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.	136
66	Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $1.8 < y < 2.1$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.	137
67	Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $2.1 < y < 2.5$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.	138
68	The response matrix, integrated over χ_{dijet} . The numbers in the boxes correspond to the values for R_{ij} in Equation 6.1.	141
69	Comparisons of unfolded and generated χ_{dijet} distributions, where the unfolding correction was determined from the same MC sample. Also shown are the ratios of unfolded to generated χ_{dijet} distributions.	143
70	Comparisons of unfolded and generated χ_{dijet} distributions, where the unfolding correction was determined from an independent sample of events. Also shown are the ratios of unfolded to generated χ_{dijet} distributions.	144
71	Raw and unfolded χ_{dijet} distributions from data. Also shown are the ratios of the unfolded to raw χ_{dijet} distributions.	146
72	Covariance matrices for the M_{ij} bins.	147
73	Correlation coefficients for the M_{ij} bins.	148

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
74	The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.	151
75	The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.	152
76	The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.	153
77	The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.	154
78	The effect of the uncertainty in the width of the core of the JER on the χ_{dijet} distributions.	155
79	χ_{dijet} distributions obtained using the double-sided Crystal Ball parametrization of the JER (blue) compared to those obtained using the Gaussian parametrization (red).	157
80	χ_{dijet} distributions obtained from generated jets (black) smeared with the double-sided Crystal Ball parametrization of the JER (blue) compared to those obtained from the Full Sim (red).	158
81	Comparison of unfolded and generated χ_{dijet} distributions from MADGRAPH5_amc@NLO. In these comparisons the response matrix used in the unfolding procedure is derived from an independent sample of events generated by PYTHIA.	160
82	Scale and PDF uncertainties for the QCD (dashed) and QCD + CI (solid) predictions in the M_{jj} bins of this analysis. The CI prediction corresponds to the CI model with $\Lambda_{\text{LL}}^- = 18$ TeV.	162
83	The electroweak correction factors for dijet production as a function of χ_{dijet} for the M_{jj} bins used in the analysis.	164
84	Non-perturbative corrections derived from PYTHIA (black) and HERWIG++ (green). The y-axis corresponds to the values of the non-perturbative correction factors.	166
85	Relative width of the generated DM mediators as a function of g_{q}	169
86	Inclusive cross section for the generated DM mediators as a function of M_{Med}	170
87	Acceptance ($ y_{\text{boost}} < 1.11$, $\chi_{\text{dijet}} < 16$, and $M_{\text{jj}} > 2.4$ TeV) of the generated DM mediators as a function of M_{Med}	170
88	Normalized χ_{dijet} distributions from data compared to the NLO QCD prediction with EW corrections in the three highest mass bins. Also shown are the predictions from the CI, ADD, QBH, and DM models. The error bars represent statistical and experimental systematic uncertainties added in quadrature. The ticks on the error bars and the gray bands correspond to the experimental and theoretical systematic uncertainty, respectively. The ratios between the unfolded data distributions and the NLO QCD+EW predictions are shown in the lower panels.	173
89	Same as Figure 88 but for the four lower mass bins.	174

LIST OF FIGURES (Continued)

<u>FIGURE</u>		<u>PAGE</u>
90	95% CL exclusion limits for CI, ADD, and QBH models. The observed (expected) limits are indicated by the solid (dashed) vertical lines. The $\pm 1\sigma$ ($\pm 2\sigma$) variations of the expected limits are given by the horizontal gray bands (black lines).	182
91	95% CL upper limits on the quark coupling g_q as a function of the mass of the vector or axial-vector DM mediator M_{Med} with $g_{\text{DM}} = 1.0$ and $m_{\text{DM}} = 1 \text{ GeV}$. Also shown on the vertical axis on the right-hand side of the figure are the corresponding limits on the width of the mediators. The observed limits, expected limits, and the one (two) σ variation of the expected limits are indicated by the solid line, dashed line, and the green (yellow) shaded band, respectively. The dashed horizontal gray line indicates a benchmark DM mediator with $g_q = 1.0$	184
92	The 95% CL observed (red) and expected (blue) excluded regions in the plane of m_{DM} and M_{Med} , for a vector mediator (upper) and an axial-vector mediator (lower) for a DM benchmark model with $g_{\text{DM}} = g_q = 1.0$. Also shown in gray are constraints from astrophysical measurements of the cosmological relic density of DM. The observed and expected lower bounds for M_{Med} overlap with each other. The expected lower bound is not seen because it overlaps with the observed lower bound.	187
93	Limits on the universal coupling g'_q between a leptophobic Z' boson and quarks from various CMS dijet analyses. The expected limits are shown in dashed lines, and the corresponding observed limits are shown in solid lines. The hashed areas show the direction of the excluded area from the observed limits. The grey dashed lines show the g'_q values at fixed values of $\Gamma_{Z'}/M_{Z'}$	188
94	The rates for the L1 single EG η restricted triggers with various E_T thresholds as a function of the values of the η restriction.	193
95	The rate for the L1 single jet (left) and H_T (right) trigger as a function of the trigger threshold estimated from data (red) and simulation (blue).	194
96	A part of the HLT menu with estimated rates.	195
97	Same as Figure 91 but obtained using the unfolded distributions and particle-level predictions.	197
98	Same as Figure 91 but obtained using SM+DM prediction smeared with the double-sided Crystal Ball parameterization of the jet p_T resolution.	197
99	Results of the signal injection test. Same as Figure 91 but replacing the data distributions with the data+DM ($g_q = 0.75$) distributions. . .	200
100	A screenshot of the webpage stating the copyright policy of European Physics Journal C.	201

LIST OF ABBREVIATIONS

ADD	Arkani-Hamed–Dimopoulos–Dvali
ALICE	A Large Ion Collider Experiment
ATLAS	A Toroidal LHC ApparatuS
AOD	Analysis Object Data
BMTF	Barrel Muon Track Finder
CA	Cambridge/Aachen
CERN	European Organization for Nuclear Research
CI	Contact Interactions
CMB	Cosmology Microwave Background
CMS	Compact Muon Solenoid
CL	Confidence Level
CPU	Central Processing Unit
CTP7	Calorimeter Trigger Processor-7
CSC	Cathode Strip Chamber
DAQ	Data AcQuisition system
DOF	Degree Of Freedom
DT	Drift Tube

LIST OF ABBREVIATIONS (Continued)

DQM	Data Quality Monitoring
EB	Electromagnetic Barrel calorimeter
ECAL	Electromagnetic CALorimeter
ED	Extra Dimension
EE	Electromagnetic Endcap calorimeter
EG	Electron/Gamma
EM	ElectroMagnetic
EMTF	Endcap Muon Track Finder
ES	Endcap preShower detector
eV	electronVolt
EW	ElectroWeak
FPGA	Field-Programmable Gate Array
FSR	Final State Radiation
GOF	Goodness Of Fit
GSF	Gaussian Sum Filter
GT	Global Trigger
GUI	Graphical User Interface
HB	Hadronic Barrel calorimeter

LIST OF ABBREVIATIONS (Continued)

HCAL	Hadronic CALorimeter
HE	Hadronic Endcap calorimeter
HF	Hadronic Foward calorimeter
HLT	High-Level Trigger
HO	Hadronic Outer calorimeter
IP	Interaction Point
ISR	Initial State Radiation
JEC	Jet Energy Correction
JER	Jet Energy Resolution
JES	Jet Energy Scale
KK	Kaluza-Klein
LINAC	LINear ACcelerator
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LO	Leading Order
LS1	Long Shutdown 1
LS	Luminosity Section
L1	Level-1

LIST OF ABBREVIATIONS (Continued)

L1A	Level-1 Accept
L1T	Level-1 Trigger
MC	Monte Carlo
μ GMT	micro Global Muon Trigger
μ GT	micro Global Trigger
μ HTR	micro Hadronic calorimeter Trigger and Readout card
MPF	Missing energy Projection Fraction
MPI	Miltiple Parton Interaction
MP7	Master Processor-7
μ TCA	micro Telecommunication Computing Archetec- ture
MTF7	Muon Track Finder-7
NLO	Next-to-Leading Order
NNLO	Next-to-Next-to-Leading Order
OMTF	Overlap Muon Track Finder
oSLB	optical Seriel Link Board
PDF	Parton Distribution Function
PF	Particle Flow

LIST OF ABBREVIATIONS (Continued)

pQCD	perturbative Quantum ChromoDynamics
PS	Parton Shower
PUS	PileUp Subtraction
QBH	Quantum Black Hole
QCD	Quantum ChromoDynamics
QED	Quantum ElectroDynamics
QFT	Quantum Field Theory
RC	Random Cone
RGE	Renormalization Group Equation
RPC	Resistive Plate Chamber
SM	Standard Model
TCDS	Trigger Control and Distribution System
TP	Trigger Primitives
TT	Trigger Tower
UE	Underlying Events
VEV	Vacuum Expectation Value

SUMMARY

A search for quark contact interactions, extra spatial dimensions, quantum black holes, and dark matter is presented in measurements of dijet angular distributions in proton-proton collisions at a center-of-mass energy of 13 TeV. The data were collected with the CMS detector at the LHC and correspond to an integrated luminosity of 35.9 fb^{-1} . The dijet angular distributions are found to be in agreement with predictions from perturbative quantum chromodynamics that include electroweak corrections. Limits for a benchmark model where only left-handed quarks participate, contact interactions are excluded at the 95% confidence level up to a scale of 12.8 and 17.5 TeV, for destructive or constructive interference, respectively. The most stringent lower limits to date are set on the ultraviolet cutoff in the Arkani-Hamed-Dimopoulos-Dvali model of extra dimensions. In the Giudice-Rattazzi-Wells convention, the cutoff scale below 10.1 TeV is excluded. For the first time, lower limits between 2.0 and 4.6 TeV are set on the mass of a dark matter mediator for (axial-)vector mediators with universal quark coupling $g_q = 1.0$.

CHAPTER 1

INTRODUCTION

At the beginning of the 20th century, with the discoveries of the electron, proton, and neutron, and introduction of Quantum Mechanics and Special Relativity, a new era of physics was ushered in, the era of particle physics. In the decades that followed, a remarkable insight into the fundamental structure of matter was achieved: everything in the universe is comprised from a few basic building blocks called elementary particles and the interactions between these elementary particles are governed by four fundamental forces. The known elementary particles and three of the four fundamental interactions are described by the Standard Model (SM) of Particle Physics. The SM was firmly established in the 1970s and 1980s, with a series of experiments that verified the existence of quarks, gluons, and W^\pm and Z bosons. The last elementary particle of the SM, the Higgs Boson, was discovered at the Large Hadron Collider (LHC) on July, 2012 [1, 2].

Although the predictions of the SM have been tested with good precision up to very short distances, or equivalently very high energies, the SM doesn't provide a complete picture of the universe. It describes three of the four fundamental forces: the strong force, the weak force, and the electromagnetic (EM) force, but it doesn't describe the gravitational force. In addition, the SM doesn't explain the Dark Matter (DM), a form of matter which appears to account for approximately 80% of the matter in the universe. There are also questions that have puzzled physicists for years that the SM doesn't answer, such as "why is the gravitational force so weak

compared to the other three fundamental forces?”, “why do the elementary particles have such different masses?”, or “why do neutrinos have mass?”, as well as many more.

In recent decades, physicists have tried to address the shortcomings of the SM by developing new physics (NP) models that extend the SM. This thesis focuses on testing the strong force in the SM and searching for signatures of NP models using dijet angular distributions in high energy proton-proton collisions at the LHC [3]. In the first and second chapters, theoretical models relevant to this thesis will be described. The experimental apparatus and analysis techniques used in this thesis will be discussed in Chapters 3–5. The data analysis procedure and the results will be presented in Chapters 6–8.

1.1 The Standard Model

1.1.1 The Elementary Particles

The elementary particles described in the SM are shown in Figure 1. The particles are classified as either fermions or bosons, depending upon their intrinsic spin. Those with half-integer spin are called fermions while those with integer spin are called bosons. The SM includes 12 fermions, 4 gauge bosons, and a Higgs boson. The fermions are further divided into quarks and leptons. The quarks have an electric charge of either $2/3$ or $-1/3$ of the elementary charge ($e = 1.602 \times 10^{-19}$ coulombs) and participate in all three fundamental interactions, while the leptons have charge $-e$ or 0 and only interact via the EM and weak forces. There are 6 quarks (up u , down d , charm c , strange s , top t , bottom b) and 6 leptons (electron e , electron neutrino ν_e , muon μ , muon neutrino ν_μ , tau τ , tau neutrino ν_τ). The quark (and lepton) types are

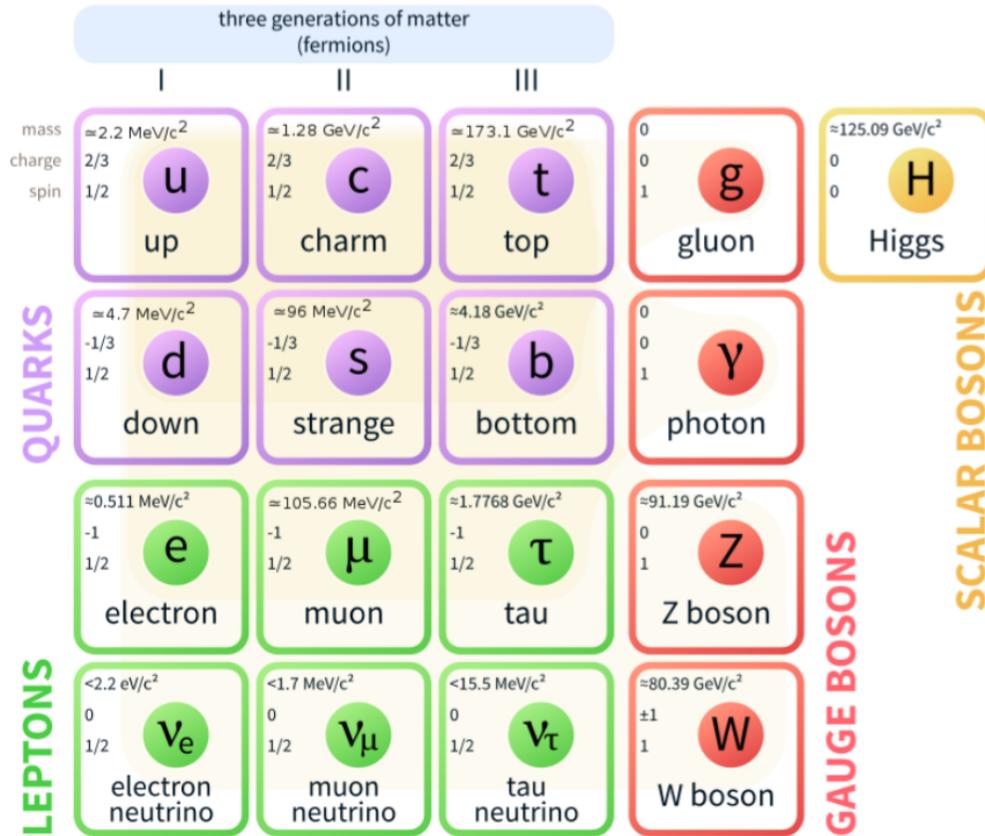


Figure 1: The elementary particles in the SM.

also called the “flavors”. Pairs of quarks and leptons are grouped to form 3 generations (see Figure 1).

While the fermions are the building blocks from which all the visible matter in the universe is comprised, the (spin-1) gauge bosons are the carriers of the fundamental forces in the SM. The force carriers for the strong, EM, and weak forces are the gluon (g), the photon (γ), and the W^\pm and Z bosons, respectively. Besides the gauge bosons, the SM includes a Higgs boson.

The Higgs boson is a scalar (spin-0) particle and is the quantum particle associated with the Higgs field. SM particles acquire mass through their interactions with the Higgs field.

In high energy physics, it is conventional to express energies, momenta, and masses in terms of natural units, in which $\hbar = c = 1$, where \hbar is the reduced Planck constant and c is the speed of light. In this case, the energy, momentum, and mass of the particles are all expressed in unit of electronvolt (eV).

1.1.2 The Fundamental Interactions

The three fundamental interactions in the SM describe all the known subatomic phenomena. The strong force is responsible for binding the quarks to form bound states called hadrons, such as the proton and neutron. The coupling strength of the strong force, denoted as α_s , has a strong dependence on the energy scale of the interaction (see Section 2.1.2). The EM force occurs between electrically charged particles and is responsible for phenomena such as electric fields, magnetic fields, and light. The EM coupling strength α is known as the fine structure constant and it is given by the Coulomb's law with $\alpha = \frac{1}{137}$. The weak force explains phenomena such as the nuclear β decay. The coupling strength of the weak interaction is about 6 orders of magnitude smaller than the strong coupling strength at ~ 1 GeV. The SM interactions and their strengths are summarized in Table I. At energy scales above the so called electroweak (EW) scale, on the order of 10^2 GeV, the EM and weak forces are unified into the EW force (see below).

Gauge Theory The SM interactions are described mathematically by a quantum field theory (QFT) that respects local gauge symmetries, i.e., the Lagrangian of SM is invariant

TABLE I: Summary of the SM fundamental interactions and their strengths.

Force	Force Carrier	Coupling Strength
Strong	\mathbf{g}	1
Electromagnetic	γ	10^{-2}
Weak	W^\pm and Z	10^{-6}

under a local gauge transformation. A local gauge transformation for a SM fermion field is defined as a local phase transformation

$$\psi \rightarrow \psi' = e^{it^a \alpha_a(x)} \psi, \quad (1.1)$$

where t^a are called the generators of the symmetry group; α_a are functions of the spacetime coordinates x ; and a are the indices of the generators of the symmetry group. An $U(N)$ ($SU(N)$) group is a group of $N \times N$ complex matrices that each matrix U in the group satisfies $U^\dagger U = UU^\dagger = \mathbb{1}$ (and the determinant of U equals 1 for $SU(N)$). If all the matrices in a group can be written as $U = e^{it^a \theta_a}$, where θ_a are numerical parameters, t^a are called the generators of the group. The number of generators of an $SU(N)$ group is $N^2 - 1$. The generators of a group have to satisfy the commutation relation $[t^a, t^b] = if^{abc} t^c$, where f^{abc} are called the structure constants (Levi-Civita symbols) of the group. The gauge symmetry in the SM can be written as $SU(3)_C \times SU(2)_L \times U(1)_Y$, to which the strong force is associated with the $SU(3)_C$ symmetry and the EW interaction is associated with the $SU(2)_L \times U(1)_Y$ symmetry.

The gauge invariant Lagrangian in the QFT can be generally written as:

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} + \bar{\psi}_i(i\gamma^\mu D_\mu - m_i)\psi_i \quad (1.2)$$

with the covariant derivative

$$D_\mu = \partial_\mu - ig t_a A_\mu^a \quad (1.3)$$

and field tensor

$$F_{\mu\nu}^a = \frac{[D_\mu, D_\nu]}{ig} = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - ig f^{abc} A_\mu^b A_\nu^c \quad (1.4)$$

where μ and ν represent the space-time indices; ψ_i represent the fermion fields; m_i are the mass of the fermions; γ^μ are the Dirac matrices; $\bar{\psi}_i = \psi_i^\dagger \gamma_0$; A_μ^a are the gauge boson fields; and g is the coupling constant. In the Lagrangian, each gauge group generator t_a is associated with a gauge boson A_μ^a field. The gauge boson fields couple to the fermion fields with the coupling constant g . In order to maintain the gauge invariance in the Lagrangian, the gauge boson fields have to transform as $A_\mu^a \rightarrow e^{it^a \alpha_a(x)} (A_\mu^a - \frac{i}{g} e^{-it^a \alpha_a(x)} \partial_\mu e^{it^a \alpha_a(x)}) e^{-it^a \alpha_a(x)}$. The gauge bosons are massless since if a mass term for the gauge bosons exists $m A_\mu^a A_\mu^a$, the gauge invariance will be violated.

Strong Interaction The theory that describes the strong force is based on the $SU(3)_C$ gauge symmetry. The ‘‘charge’’ (quantum number) of the $SU(3)_C$ symmetry is called color. There are 3 color charges: red (r), blue (b), and green (g). Each quark carries one color charge. Quarks and gluons are called partons. The quark fields, ψ_i , are written as triplets composed

of the color components (e.g., $\psi_u = \begin{pmatrix} u_r \\ u_b \\ u_g \end{pmatrix}$ and $\psi_d = \begin{pmatrix} d_r \\ d_b \\ d_g \end{pmatrix}$ for the up and down quark fields, respectively). The generators in the $SU(3)_C$ group correspond to the 8 Gell-Mann matrices and give rise to the 8 gluon field carriers. Each gluon carries two color charges. Quarks interact with each other and form colorless bound states by exchanging gluons. A more detailed discussion about the strong force will be given in Chapter 2.

Electroweak Interaction The theory describing the EW force is based on the $SU(2)_L \times U(1)_Y$ symmetry, which is known as EW symmetry. For the $SU(2)_L$ symmetry, one left-handed up-type quark or lepton and one left-handed down-type quark or lepton in the same generation form doublets that transform under the $SU(2)_L$ group. The up-type quarks and leptons are u , c , t , ν_e , ν_μ , and ν_τ . The down-type quarks and leptons are d , s , b , e , μ , and τ . The doublets are written as $\psi = \begin{pmatrix} u \\ d \end{pmatrix}_L$, $\psi = \begin{pmatrix} e \\ \nu_e \end{pmatrix}_L$, etc. Weak-isospin T_3 is a quantum number associated with the $SU(2)_L$ symmetry, and takes the values of $\pm\frac{1}{2}$. Left-handed up-type quarks and leptons carry $T_3 = \frac{1}{2}$, and left-handed down-type quarks and leptons carry $T_3 = -\frac{1}{2}$. The right-handed fermions, except for right-handed neutrinos which do not exist in the SM, carry $T_3 = 0$ and do not participate in the weak interaction. The left- and right-hand projections of the fermion fields $\psi_{\text{left/right}}$ are defined as, $\psi_{\text{left/right}} = \frac{1 \mp \gamma^5}{2} \psi$, where γ^5 is the fifth Dirac matrix. The three generators of the $SU(2)_L$ gauge group are the Pauli matrices and correspond to three

gauge bosons: W^a ($a=1,2,3$). The W^a bosons have T_3 values of 1, 0, and -1 for $a=1, 2$, and 3, respectively.

Both left- and right-handed fermions transform under the $U(1)_Y$ symmetry. The $U(1)_Y$ symmetry corresponds to one gauge boson (B), which couples to both left- and right-handed fermions. The weak hypercharge Y is a quantum number associated with the $U(1)_Y$ symmetry and it relates the electric charge Q and the third component of the weak-isospin T_3 by $Q = T_3 + \frac{Y}{2}$.

The Higgs Mechanism and Electroweak Symmetry Breaking The gauge bosons associated to EW symmetry, W^a ($a=1,2,3$) and B are all massless. However, the observation of massive W^\pm and Z bosons suggests that EW symmetry is broken. In the SM, the EW symmetry is a broken symmetry that is broken through the Higgs mechanism [4–9] and the W^\pm and Z bosons and the photon are produced by the mixings of the W^a and B gauge bosons after EW symmetry breaking.

The Higgs mechanism adds a complex doublet field ϕ with $Y = 1$ and $T_3 = -\frac{1}{2}$ to the SM fields. The Lagrangian for the doublet field is invariant under the EW symmetry and is written as,

$$\mathcal{L} = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (1.5)$$

where $\mu^2 < 0$ and $\lambda > 0$. In the Lagrangian, the covariant derivative D_μ is written as

$$D_\mu = \partial_\mu - ig\sigma_a W_\mu^a + ig' \frac{Y}{2} B_\mu, \quad (1.6)$$

where σ_a are the Pauli matrices (generators of the $SU(2)$ group); g is the coupling constant associated to the W^a gauge bosons; and g' is the coupling constant associated to the B gauge boson. In the Lagrangian (Equation 1.5), the first term is the kinematic term of the field and the second and third terms comprise the Higgs potential. The Higgs potential ($V(\phi^\dagger\phi)$) has a minimum value at $\phi^\dagger\phi = -\frac{\mu^2}{\lambda} = v^2$, and v is called the vacuum expectation value (VEV) of the Higgs potential.

In general, the Higgs doublet field is written as $\phi = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}$, where ϕ_1 - ϕ_4 are real scalar fields. At the ground state of the Higgs potential, ϕ can be expressed in terms of a constant part that reflects the non-zero value of the ground state (v) and a variable field ($H(x)$) that is small for perturbations around the ground state: $\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}$. Inserting ϕ into the Higgs potential, we obtain the following expression:

$$V = \frac{\mu^2}{2}H^2 + \frac{\mu^2}{\lambda}H^3 + \frac{\mu^2}{4\lambda^2}H^4. \quad (1.7)$$

The Higgs potential, therefore describes a scalar particle, the Higgs boson, that has mass $\mu\sqrt{2}$.

The W^\pm and Z boson fields and the photon field (A) are expressed by:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (1.8)$$

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}, \quad (1.9)$$

with masses

$$M_{W^\pm} = \frac{1}{2}vg, \quad M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}, \quad M_\gamma = 0, \quad (1.10)$$

where $\cos\theta_W = \frac{M_{W^\pm}}{M_Z} = \frac{g}{\sqrt{g^2 + g'^2}}$ is the Weinberg angle. From the above equations, we see that the SM W^\pm and Z bosons and the photon are given by mixings of the W^a ($a=1,2,3$) and B gauge bosons, and the W^\pm and Z bosons acquire masses from the non-zero VEV of the Higgs potential.

The Mass of the Fermions Although the mass term for the fermions ($m_i\bar{\psi}_i\psi_i$) in Equation 1.2 does not violate the $SU(3)_C$ symmetry, it does violate the $SU(2)_L$ symmetry since it mixes the left- and right-handed fermions. This is addressed by adding to the SM Lagrangian a Yukawa coupling term between the Higgs boson and the fermions:

$$\mathcal{L}_{\text{Yukawa}} = -g_l\bar{L}_L\phi l_R - g_d\bar{Q}_L\phi d_R - g_u\bar{Q}_L\phi^c u_R + h.c., \quad (1.11)$$

where L_L and Q_L are the left-handed lepton and quark doublets, respectively; ϕ^C is the charge conjugate of the Higgs field; and $g_{l,d,u}$ are the Yukawa coupling constants. Inserting $\phi^C = \frac{1}{\sqrt{2}} \begin{pmatrix} v + H \\ 0 \end{pmatrix}$, we get

$$\mathcal{L}_{\text{Yukawa}} = -g_f\frac{v}{\sqrt{2}}\bar{\psi}\psi - \frac{g_f}{\sqrt{2}}\bar{\psi}\psi H. \quad (1.12)$$

Therefore, by interacting with the Higgs field, the fermions acquire mass $m_f = g_f\frac{v}{\sqrt{2}}$ that is proportional to the Yukawa coupling.

1.2 New Physics Models

1.2.1 Quark Contact Interactions

Contact interactions (CIs) were first introduced by Fermi’s theory of β decay in 1933 before the complete electroweak theory was established and the W^\pm bosons were discovered. In Fermi’s theory, the β decay was described by a CI among a proton, a neutron, an electron, and an anti-electron neutrino with interaction strength G_F . Today, we know that Fermi’s CI is a low-energy approximation of the EW interaction between an anti-up (up) quark and a down (anti-down) quark, which is mediated by the W^- (W^+) boson. The W^- (W^+) then decays into an electron (positron) and an electron anti-neutrino (electron neutrino). Similarly, quark CIs can be regarded as the low-energy effective theory approximation of new interactions between quarks at higher energies (see Figure 2).

One example of a new interaction is the quark compositeness model [10–12], in which quarks are bound states of more fundamental constituents. This model is intriguing because it provides possible answers for questions related to the number of quark generations and to why the values of quark masses and charges are what they are. The new interactions between quark constituents occur above a large characteristic energy scale Λ . At energies below the compositeness scale, the new interactions can be integrated out to form a four-fermion contact interaction model.

While one can write down a great variety of different operators contributing to the quark contact interactions with different flavor and color structures, only the most general case—the



Figure 2: Diagrams for a four-fermion contact interaction (left) describing the low energy effective field theory for a new interaction between the fermions (right).

flavor-diagonal color-singlet models [11, 12] will be discussed in this thesis. The Lagrangian describing such CIs can be written as:

$$\mathcal{L}_{\text{qq}} = \frac{2\pi}{\Lambda^2} \left[\eta_{\text{LL}} (\bar{q}_{\text{L}} \gamma^\mu q_{\text{L}}) (\bar{q}_{\text{L}} \gamma_\mu q_{\text{L}}) + \eta_{\text{RR}} (\bar{q}_{\text{R}} \gamma^\mu q_{\text{R}}) (\bar{q}_{\text{R}} \gamma_\mu q_{\text{R}}) + 2\eta_{\text{RL}} (\bar{q}_{\text{R}} \gamma^\mu q_{\text{R}}) (\bar{q}_{\text{L}} \gamma_\mu q_{\text{L}}) \right],$$

where the subscripts L and R refer to the left and right chiral projections of the quark fields, respectively. The variables η_{LL} , η_{RR} , and η_{RL} are taken in various combinations of 0, +1, or -1, with each combination corresponding to a different CI model (see Table II).

TABLE II: Quark CI models.

Model	$(\eta_{\text{LL}}, \eta_{\text{RR}}, \eta_{\text{RL}})$
Λ_{LL}^\pm	$(\pm 1, 0, 0)$
Λ_{RR}^\pm	$(0, \pm 1, 0)$
Λ_{VV}^\pm	$(\pm 1, \pm 1, \pm 1)$
Λ_{AA}^\pm	$(\pm 1, \pm 1, \mp 1)$
$\Lambda_{\text{(V-A)}}^\pm$	$(0, 0, \pm 1)$

The quark CIs have been searched for at the CERN Sp \bar{p} S by the UA1 [13] and UA2 [14] Collaborations, at the Fermilab Tevatron by the CDF [15–17] and D0 [18–21] Collaborations, and at the CERN LHC by the ATLAS [22–27] and CMS [28–33] Collaborations. The most stringent limits on the CI models come from searches using dijet angular distributions [27, 33]. The search performed by the CMS collaboration [33] using 2.7 fb^{-1} of proton-proton collision data at center-of-mass energy $\sqrt{s} = 13 \text{ TeV}$ excluded quark CIs at 95% Confidence Level (CL) up to a scale of 8.4 and 18.6 TeV, depending on the details of the model. The publication from the ATLAS Collaboration [27] using 37 fb^{-1} of data at $\sqrt{s} = 13 \text{ TeV}$ provided lower limits on the quark CI scales from 13.1 to 29.5 TeV.

1.2.2 Extra Spatial Dimensions

An important energy scale that characterizes the gravitational force is the fundamental Planck scale ($M_{\text{Pl}} = G_{\text{N}}^{-1/2} \sim 10^{18} \text{ GeV}$ where G_{N} is the Newton’s gravitational constant), at which gravity is expected to be as strong as the other three fundamental forces. One question that has puzzled physicists for a long time is that why there is a huge difference between the energy scales of the gravity and the electroweak interactions, i.e., why the fundamental Planck scale is 10^{16} times greater than the EW scale ($\sim 10^2 \text{ GeV}$). Arkani-Hamed, Dimopoulos, and Dvali (ADD) proposed the extra spatial dimension model [34, 35] to address this problem.

Before describing the ADD model, it is necessary to discuss the Kaluza-Klein (KK) theory [36, 37]. The KK theory was the first extra dimension (ED) model, in which one extra spatial dimension is added to the regular 4-dimensional space. The KK theory was originally

proposed to unify the gravitational force and the electromagnetic force in the 1920s, but it ran into failure almost immediately since it didn't predict the mass of the electron correctly.

A requirement of the KK theory is that any physical quantity in the ED should be subject to a periodical boundary condition, i.e., a physical quantity at point y in the extra space should be identical to the quantity at $y+2\pi R$, where R is the size of the extra dimension. This requirement is called compactification. Given the periodical boundary condition, the wave function of, for example, a scalar field in the 5-dimensional space $\Phi(x, y)$, where x and y represent the regular 4-dimension and the 5th-dimension coordinates, respectively, can be written in the form of Fourier series:

$$\Phi(x, y) = \sum_{n=0}^{\infty} \phi_n(x) e^{iny/R}. \quad (1.13)$$

The Fourier coefficients, $\phi_n(x)$ are fields over the regular 4-dimensional space. These fields are called the KK modes or KK towers. Applying the 5-dimensional Klein-Gordon equation to the scalar field,

$$(\square - \partial_y^2 + M_0^2)\Phi(x, y) = 0, \quad (1.14)$$

and substituting Φ with the Fourier series, we have,

$$(\square + M_n^2)\phi_n(x) = 0, \quad (1.15)$$

where $n = 1, 2, 3, \dots$, and

$$M_n = \sqrt{M_0^2 + \frac{n^2}{R^2}}. \quad (1.16)$$

This implies that each KK mode by itself is a particle that has mass M_n and satisfies the Klein-Gordon equation. The zero mode ($n = 0$) field has mass equal to the mass of the original 5-dimensional field.

The ADD model speculates the existence of N -dimensional compact extra space with radius R . In it, SM particles are confined in the regular 4-dimensional space and interact with the gravity via KK modes (KK gravitons). The fundamental Planck scale in the $4+N$ dimensional space, M_D , is at the electroweak scale. At distances $r \ll R$, the gravity potential is written as:

$$V(r) = \frac{1}{M_D^{N+2}} \frac{m_1 m_2}{r^{N+1}}. \quad (1.17)$$

At distances $r \gg R$, the gravitational flux in the ED cannot penetrate to the regular 4-dimensional space, and the gravity potential is modified (by applying the Gauss's law) to be

$$V(r) = \frac{1}{M_D^{N+2} R^N} \frac{m_1 m_2}{r}. \quad (1.18)$$

Comparing Equation 1.18 to Newton's law of gravity, we get the relationship between R , M_D , and M_{Pl} :

$$M_{Pl}^2 = \frac{1}{G_N} = M_D^{N+2} R^N. \quad (1.19)$$

Taking $M_{Pl} = 10^{18} \text{ GeV}$, $M_D \sim \text{TeV}$, and $N = 6$, we see that the size of the ED could be as large as $R \sim 0.1 \text{ mm}$. As R is large in the ADD model, the mass differences between the KK modes are small (see Equation 1.16).

In particle collision experiments, signatures of extra spatial dimensions can be searched via the virtual KK graviton exchange process, i.e., KK graviton mediated $2 \rightarrow 2$ scattering processes. The sum of all KK modes in the cross section calculation of this process diverges and therefore needs to be truncated at an energy scale Λ expected to be of order M_D . The effective theory for the ADD model can be parametrized in three ways: Giudice-Rattazzi-Wells (GRW) [38], Hewett [39], and Han-Lykken-Zhang (HLZ) [40]. In this analysis, only the GRW and HLZ parametrizations will be studied. In the GRW convention, the KK graviton exchange process is regularized by the single cutoff scale Λ_T . The HLZ convention describes the process in terms of the parameters M_S and n_{ED} , where M_S represents a scale also expected to be of order M_D , and n_{ED} is the number of ED. The parameters M_S , n_{ED} , and Λ_T are related by the following equations [41]:

$$M_S^4 = \begin{cases} \Lambda_T^4 \log\left(\frac{M_S}{\hat{s}}\right), & n_{ED} = 2 \\ \Lambda_T^4 \frac{2}{n_{ED} - 2}, & n_{ED} > 2 \end{cases} \quad (1.20)$$

In the HLZ convention, we consider scenarios with $n_{ED} = 2-6$. The case of $n_{ED} = 1$ is not considered since it would require extra dimensions with sizes of order the size of the solar system (see below). The case of $n_{ED} = 2$ is special in the sense that the relation between M_S and Λ_T also depends on the parton-parton center of mass energy $\sqrt{\hat{s}}$.

Precision measurements of Newtonian gravity can be used to search for the extra spatial dimensions. By searching for deviations of Newton's law at short distances, the size of the ED in the ADD model with $n_{ED} = 2$ was excluded down to $R > 37 \mu\text{m}$ [42]. Evidence for virtual graviton exchange has previously been searched for at the LHC in diphoton, dilepton, and dijet

final states, among which the dijet search [33] provided the most stringent limits for the cutoff scale in the ADD model.

1.2.3 Quantum Black Holes

A black hole is a region of space that has a gravitational field so intense that matter or radiation cannot escape. Black holes are formed by the gravitational collapse of massive stars that have masses a few times larger than the mass of the sun. However, under the context of extra dimension models, in which the fundamental Planck scale is lowered to a few TeV, microscopic black holes with masses at around the TeV scale can be produced. In this thesis, two large extra dimension models are considered, the ADD model, with $n_{\text{ED}} = 6$, and the Randall–Sundrum (RS) model [43, 44] with a single warped compact extra dimension ($n_{\text{ED}} = 1$) as benchmark models to provide low energy Planck scales for the microscopic black hole productions.

Consider a microscopic black hole with mass $M_{\text{BH}} \geq M_{\text{D}}$ and Schwarzschild radius (radius of the boundary of the black hole) r_{S} . If two partons collide with $\sqrt{\hat{s}} > M_{\text{BH}}$ and impact parameter less than r_{S} , such a black hole can be formed. The production cross section of these black holes is described by the geometrical cross section, πr_{S}^2 , where in the ADD model, r_{S} is given by [45]

$$r_{\text{S}} = \frac{1}{\sqrt{\pi}M_{\text{D}}} \left[\frac{M_{\text{BH}}}{M_{\text{D}}} \frac{8\Gamma(\frac{n_{\text{ED}}+3}{2})}{n_{\text{ED}}+2} \right]^{\frac{1}{n_{\text{ED}}+1}}, \quad (1.21)$$

and in the RS model, r_S is given by [46, 47]

$$r_S = \frac{1}{\pi M e^{-\pi k R_c}} \sqrt{\frac{M_{\text{BH}}}{3\pi e^{-\pi k R_c}}}. \quad (1.22)$$

In Equation 1.22, R_c is the size of the extra dimension, k is the curvature of the 5-dimensional space defined in the RS model, and M is the fundamental 5-dimensional Planck scale in the RS model.

If the microscopic black holes have masses $M_{\text{BH}} \gg M_{\text{D}}$, the decay of the black holes are well understood in the framework of general relativity. These black holes evaporate via Hawking radiation [48] and can be searched for at the LHC using multi-jet final states. If the black holes have masses close to M_{D} , quantum effects have to be taken into account and the black holes will decay predominantly into dijets [49–51]. Such black holes are also called quantum black holes (QBHs). Recent searches for QBHs decaying into dijet final states at the LHC have excluded QBHs with masses below 8.9 TeV [27, 33].

1.2.4 The Simplified Model for the Dark Matter

The earliest evidence for DM came from studying the galactic rotation curves. These analyses studied the orbital velocity of objects within a galaxy as a function of their distance from the center of the galaxy. The velocity was expected to decrease (increase) with increasing distance outside (inside) the region where the majority of the visible matter in the galaxy is clustered. However, the observed galactic rotation curves exhibited large discrepancies from this expectation and suggested that either gravity needs to be modified or that there exists

additional non-luminous matter. As an example, the rotation curve of the galaxy M33 is shown in Figure 3, left. As can be seen in the figure, outside the region containing most of the visible matter the orbital velocity of the objects does not decrease as expected, but instead slightly increases with distance.

Another evidence for DM comes from the study of the bullet cluster (1E 0657-558). The bullet cluster consists of two colliding clusters of galaxies as shown in Figure 3, right. The figure contains three components. The first component, an optical image from the Magellan and the Hubble Space Telescope, shows the galaxies in each of the two colliding clusters in orange and white. The second component (pink), provided by the Chandra X-ray Observatory, shows the distribution of hot gas in the cluster. The hot gas is created mainly by electromagnetic interactions during the collision and contains the bulk of the normal matter in the cluster. The last component (blue) shows the mass distribution of the bullet cluster, as measured by gravitational lensing, which is the distortion of light from the galaxies behind the cluster by the mass in the cluster. The separation between the blue and pink images indicates that the majority of the mass of one cluster simply passed through the other experiencing only the gravitational interaction. This observation provides one of the most striking evidence of DM. Detailed analysis of the observation disfavors an alternative explanation of the images by a model of modified gravity [52, 53].

In the standard thermal “freeze-out” picture of the universe, the DM and SM particles are in thermal equilibrium at the beginning of the universe. As the universe keeps expanding and cooling, the DM density kept dropping because of the annihilation of the DM particles (the

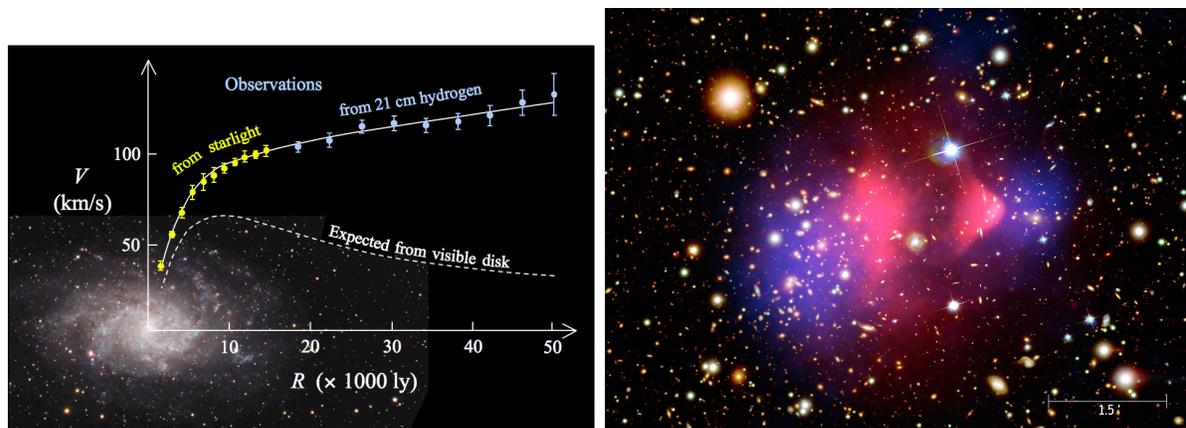


Figure 3: The Evidence of DM. Left: the measured galactic rotation curve for M 33 galaxy compared to the prediction obtained using only the visible matter in the galaxy. At distances close to the center of the galaxy, the orbital velocities are measured using starlight, and at large radii, the 21 cm hydrogen line (microwave created by electromagnetic radiation of a neutral hydrogen atom) is used to measure the velocities of the interstellar hydrogen. Right: the optical image from Magellan and Hubble Space Telescope for the bullet cluster 1E 0657-558. Also shown are the hot gas (pink) and mass (blue) distributions in the bullet cluster. The measurements of the hot gas and mass clusters are provided by the Chandra X-Ray Observatory and gravitational lensing, respectively.

inverse of the process shown on the left in Figure 4). However, at some point, the universe became large enough that the probability of one DM particle finding another to annihilate was very small and the DM density stopped decreasing. The DM particles that were left over are known as the thermal relic of the universe.

The relic density of the DM can be obtained from studying the fluctuations in the Cosmic Microwave Background (CMB), the blackbody electromagnetic radiation left over from the Big Bang. The CMB was formed in the early universe, once it was sufficiently cooled to allow baryons to form and photons to propagate freely without constantly being scattered by electrons

and protons. One source of the fluctuations in the CMB comes from the competing effects of photon pressure pushing matter outward and gravity pushing matter inward causing oscillations in the photon and baryon densities. If DM is present, the oscillation pattern will be changed since DM is unaffected by the photon pressure, allowing the DM relic density to be obtained. The Planck Collaboration [54] provided the most precise observation on the CMB to date. The relic density for the DM extracted by the Planck Collaboration is $\Omega_c h^2 = 0.118 \pm 0.003$, where Ω_c is the relic density and h is the Hubble constant and equals to 0.7. The uncertainty in the Planck result corresponds to 68% CL.

While the existence of DM has been well established, the particle physics property of DM remains unknown. Many NP models provide DM candidates. In principle, any particle that satisfies the following conditions can be a DM candidate: must interact very weakly (otherwise would have been discovered already), must be stable over cosmological time scales (otherwise would have decayed by now), and must give the right relic density. These candidates include axions, sterile neutrinos, weakly interacted massive particles, etc., and have masses range from a few keV to several TeV, depending on the model. Although the searches for the DM candidates can be model dependent, it is possible to create simplified DM models for the DM searches at the LHC [55–57].

In the simplified model of the DM, the DM particle (χ) with mass m_{DM} is assumed to interact with SM particles via some new spin-0 or spin-1 mediator with mass M_{Med} (see Figure 4). The

coupling strengths of the mediator and quarks or DM are denoted as g_q and g_{DM} , respectively.

The spin-1 mediator (Z') model has the following Lagrangian

$$\mathcal{L}_{\text{vector}} = -g_{\text{DM}} Z'_\mu \bar{\chi} \gamma^\mu \chi - g_q \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q} \gamma^\mu q,$$

and

$$\mathcal{L}_{\text{axial-vector}} = -g_{\text{DM}} Z'_\mu \bar{\chi} \gamma^\mu \gamma_5 \chi - g_q \sum_{q=u,d,s,c,b,t} Z'_\mu \bar{q} \gamma^\mu \gamma_5 q,$$

for vector and axial-vector mediators, respectively. In collider experiments, the primary search channels for spin-1 DM mediators are the dijet channel (Figure 4 right), in which the mediators decay to a pair of quarks, and the mono-jet channel (Figure 4 left), in which the mediators decay to a pair of DM particles. As the DM particles cannot be directly detected by particle detectors, the initial state radiation shown in Figure 4 left provides a mono-jet signature for the DM decay. The fact that g_q is assumed to be the same for all quark flavors makes the spin-1 mediators particularly suitable for searches with a dijet final state. The spin-0 (scalar and pseudo-scalar) mediators have properties similar to the Higgs boson, for which the coupling to an SM particle is proportional to the mass of the particle. Therefore, the spin-0 mediators are searched for using final states that contain heavy particles, such as top or bottom quarks. In a scenario where $g_q = 0.25$ and the relative widths in the DM decay channels are negligible, values of M_{Med} below 3.0 TeV were excluded by narrow dijet resonance searches [27, 58] for vector and axial-vector mediators. Mono-jet searches [59–61] have excluded vector and axial-vector mediators below 1.8 (2.1) TeV for $g_q = 0.25$ ($g_q = 1.0$) and $g_{\text{DM}} = 1.0$ [60].



Figure 4: The interaction between quarks and DM particles (left) and quarks (right) mediated by the DM mediator in the simplified DM model. As DM particles cannot be directly detected by particle detectors, the initial state radiation shown in the left plot provides the mono-jet signature for the DM decay at the LHC.

CHAPTER 2

QUANTUM CHROMODYNAMICS AND DIJET ANGULAR DISTRIBUTIONS

As mentioned earlier, the quarks carry color charges and interact with each other by exchanging gluons. The theory describing these interactions is called Quantum Chromodynamics (QCD). Experimentally, free partons have never been observed directly. They always bound together to form colorless (color-neutral) hadrons. This phenomenon is called color confinement. In high energy particle collision experiments, collimated sprays of hadrons created in the final states are called jets. Jets are the experimental signature of partons. As the LHC is a proton-proton collider, the most common interaction that takes place during the collision is the strong interaction described by QCD and which produces jets final states. This chapter describes QCD and jet physics in detail.

2.1 Quantum Chromodynamics

2.1.1 The Lagrangian of Quantum Chromodynamics

The QCD Lagrangian that describes the dynamics of the quarks and gluons can be written as:

$$\mathcal{L} = \sum_{q=u,d,s,c,b,t} \bar{q}_a (i\gamma^\mu (D_\mu)_{ab} - m) q_b - \frac{1}{4} G_{\mu\nu}^A G_A^{\mu\nu} \quad (2.1)$$

where,

$$G_{\mu\nu}^A = \partial_\mu A_\nu^A - \partial_\nu A_\mu^A - if^{ABC} A_\mu^B A_\nu^C \quad (2.2)$$

$$(D_\mu)_{ab} = \partial_\mu \delta_{ab} + ig_s T_{ab}^A A_\mu^A \quad (2.3)$$

and $A, B, C = 1 \dots 8$ are the gluon indices; $a, b = 1, 2, 3$ are the color indices; A_μ is the gluon field; and g_s is the strong coupling and related to α_s by $\alpha_s = \frac{g_s^2}{4\pi}$. The color group $SU(3)_C$ generators are written as $T^A = \lambda^A/2$, where λ^A are the Gell-Mann matrices.

The first term in Equation 2.1 represents the free quark fields and quark-gluon interactions. The second term of Equation 2.1 describes the free gluon fields and the cubic and quartic gluon self-interaction vertices. These gluon self-interaction terms mark an important difference between QCD and the theory describing the electromagnetic force, the quantum electrodynamics (QED). In QED, photon self-interactions are not allowed since QED is described by the $U(1)$ symmetry that has a structure constant of zero. Gauge theories that have a non-zero structure constant are also called Non-Abelian gauge theories. The fundamental QCD couplings are shown in Figure 5.

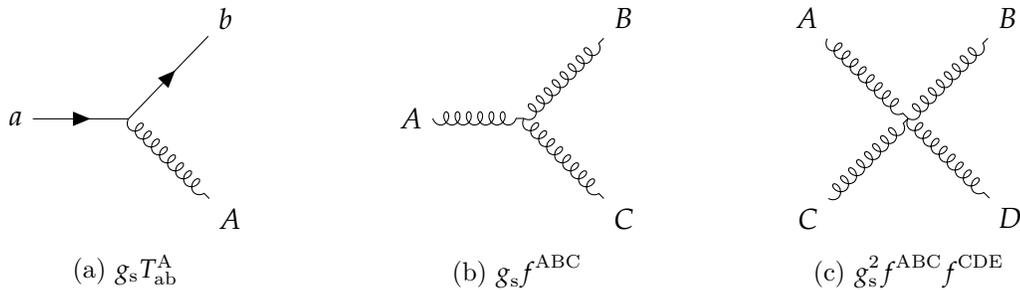


Figure 5: Fundamental coupling of QCD: quark-gluon interaction (left) and gluon self interaction terms (middle and right).

2.1.2 Perturbative QCD and the Running of the Strong Coupling Constant

In perturbative QCD (pQCD), the cross section for a QCD parton-level process can be expanded as a perturbative series in the strong coupling constant α_s ,

$$\sigma = \alpha_s^2(\sigma_0 + \sigma_1\alpha_s + \sigma_2\alpha_s^2 + \dots), \quad (2.4)$$

where σ_0 , σ_1 , and σ_2 are leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO) cross sections in α_s , respectively. With the assumption of α_s being small, an observable (such as a cross section) can be approximated by the first few terms in the perturbative expansion. The coupling α_s is not truly constant, but is dependent on the momentum transfer Q . This energy dependence is due to vacuum fluctuations, and as it will be described later, it leads to a property of strong interactions called asymptotic freedom.

Using Feynman rules, the LO quantities can be calculated rather easily. Calculations for higher order terms become complicated when loop diagrams come into play since the integrals over the four-momenta in a loop are logarithmically divergent. The technique that overcomes this divergence is called renormalization. During renormalization, α_s is redefined such that the infinities in the QCD observable calculations are absorbed by redefining α_s that needs to be constrained by experiments. In this way, finite order pQCD calculations can be achieved.

The renormalized α_s is defined as $\alpha_s(\mu_R^2)$, where μ_R is an (unphysical) renormalization scale. When μ_R is close to the scale of the momentum transfer Q for a given process, $\alpha_s(\mu_R^2 \simeq Q^2)$ then indicates the effective strength of the strong interaction in that process. The dependence

of the renormalized coupling on the scale of the momentum transfer is referred to as the running coupling and is calculable in pQCD. A summary of measurements of the running coupling as a function of Q is given in Figure 6 [62]. The average value of the strong coupling at the energy scale equal to the mass of the Z boson is $\alpha_s(M_Z) = 0.1181 \pm 0.0011$.

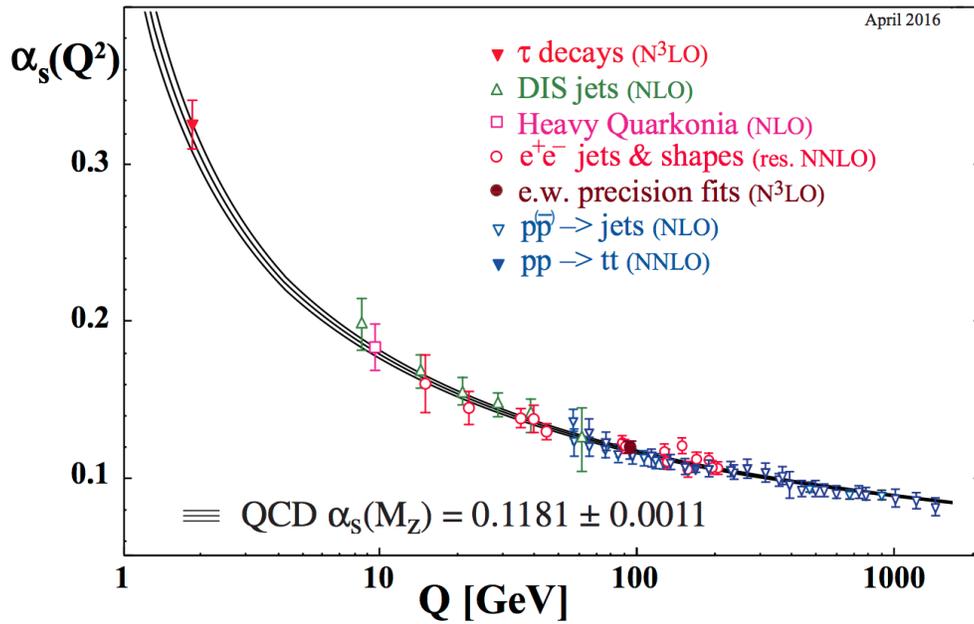


Figure 6: Summary of the measurements of the QCD running coupling [62].

The running of the coupling constant $\alpha_s(\mu_R^2)$ is given by the renormalization group equation (RGE):

$$\mu_R^2 \frac{\partial \alpha_s(\mu_R^2)}{\partial \mu_R^2} = \beta(\alpha_s(\mu_R^2)). \quad (2.5)$$

In pQCD, the right-hand side of Equation 2.5 can be expanded in a series of terms of $\alpha_s(\mu_R^2)$

$$\beta(\alpha_s(\mu_R^2)) = -\alpha_s(\mu_R^2)^2(b_0 + b_1\alpha_s(\mu_R^2) + b_2\alpha_s^2(\mu_R^2) + \dots). \quad (2.6)$$

The coefficients in the expansion are determined using higher order (loop) corrections to the vertices of the QCD interactions and the self-energy of the gluons and quarks. As an example, the Feynman diagrams for the one-loop corrections to the self-energies of quarks and gluons are shown in Figure 7. The first two coefficients in the expansion, given by the one- and two-loop corrections, are $b_0 = \frac{33 - 2n_f}{12\pi}$ and $b_1 = \frac{153 - 19n_f}{24\pi^2}$, respectively, where n_f is the number of flavors for quarks that have a mass smaller than μ_R . The first term in b_0 , $\frac{33}{12\pi}$, comes from the diagram in Figure 7 right, and causes the lowering of the coupling constant as Q increases since $n_f \leq 6$. Retaining only the leading term b_0 and solving Equation 2.5, we can get the analytic expression of the running coupling

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu_R^2)}{1 + b_0 \ln(Q^2/\mu_R^2)\alpha_s(\mu_R^2)}. \quad (2.7)$$

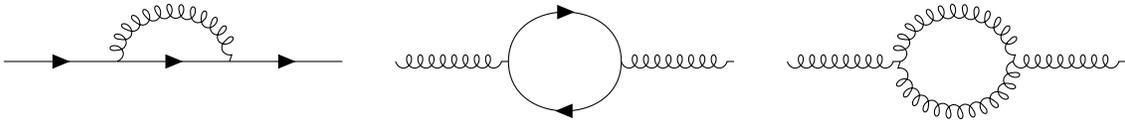


Figure 7: The one-loop corrections to the quark (left) and gluon (middle and right) self-energies.

The running of the strong coupling provides one of the key properties of the QCD: asymptotic freedom. It can be seen in Figure 6 or Figure Equation 2.7, the value of the strong coupling becomes smaller for processes involving large momentum transfers. Such processes probe small distances and are called “hard processes”. For processes involving small momentum transfers (or large distances), the strong coupling is large such that pQCD will not be applicable. Denoting the energy scale where the running coupling formally diverges as Λ_{QCD} , we have

$$\alpha_s(Q) = \frac{1}{b_0 \ln(Q^2/\Lambda_{\text{QCD}}^2)}. \quad (2.8)$$

Experimental measurements give $\Lambda_{\text{QCD}} \simeq 200$ MeV. This value can be considered as the boundary of the region where pQCD is not valid anymore.

2.1.3 The Factorization Theorem

The factorization theorem of QCD states that the cross section calculation in high energy proton-proton collisions can be separated into two parts. A short distance part which describes the underlying hard process cross section and is calculable with pQCD, and a manifestly long distance non-perturbative part that describes the momentum distributions of the partons within the colliding protons through universal functions that have to be extracted from experimental measurements¹.

¹Universal functions describing the probability of partons to fragment to hadrons are also extracted. These functions are called fragmentation functions.

In the factorization theorem, the cross section for a generic process in high energy proton-proton collisions can be written as

$$\sigma_{pp \rightarrow X} = \sum_{i,j} \int dx dx' f_{i/p}(x, \mu_F) f_{j/p}(x', \mu_F) d\sigma_{ij \rightarrow X}(x, x', \mu_R, \alpha_s(\mu_R)) \quad (2.9)$$

where i and j represent the initial state partons within the colliding protons, x and x' are the momentum fractions carried by the colliding protons, and $\sigma_{ij \rightarrow X}$ is the cross section for the hard process $ij \rightarrow X$. The functions $f_{i/p}(x, \mu_F)$ and $f_{j/p}(x', \mu_F)$ describe the momentum fraction distributions of partons within a proton and are called the parton distribution functions (PDFs). A pictorial representation of the factorization theorem can be seen in Figure 8.

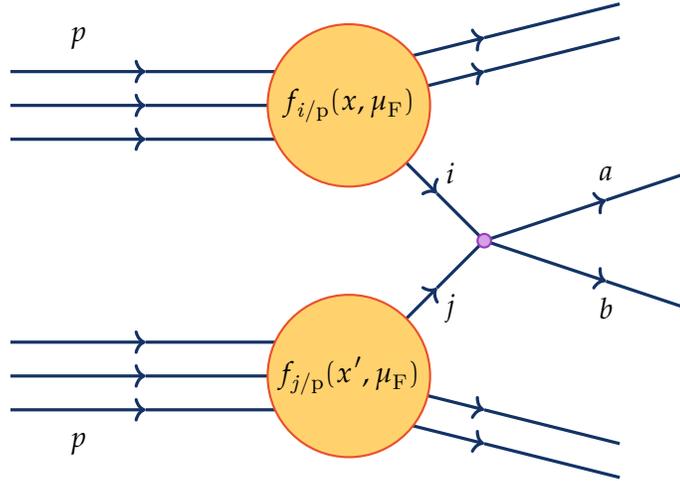


Figure 8: A pictorial representation of two partons i and j originated from two protons participating a hard scattering process $ij \rightarrow ab$. The momentum fraction distributions of partons i and j within the two protons are parametrized by the PDFs.

The PDFs, $f_{i/h}(x, \mu_F)$, quantify the probability to find a parton i with longitudinal momentum fraction x within a hadron h at the factorization scale μ_F , which characterizes the momentum transfer scale at which the hadron is probed. The proton PDFs for $u, \bar{u}, d, \bar{d}, s$ and g at $\mu_F = Q = 2$ and 100 GeV extracted by the CTEQ Collaboration are shown in Figure 9 [63]. The PDFs at an arbitrary scale are related to the PDFs at a given scale Q via the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi evolution equations

$$\mu_F^2 \frac{\partial f_{i/p}(x, \mu_F)}{\partial \mu_F} = \sum_{j=q,q,g} \int_x^1 \frac{dz}{z} \frac{\alpha_s(\mu_F)}{2\pi} P_{ij}(z) f_{j/p}(x/z, \mu_F) \quad (2.10)$$

where the $P_{ij}(z)$ are the splitting functions describing the probability for a parton i originating from a parton j and carrying the momentum fraction z of the original proton. The analytical formula of the splitting functions can be calculated using pQCD and are written at LO as:

$$P_{qq}(z) = \frac{4}{3} \left(\frac{1+z^2}{(1-z)_+} + \frac{3}{2} \delta(1-z) \right) = P_{qq}(z) \quad (2.11)$$

$$P_{qg}(z) = \frac{1}{2} (z^2 + (1-z)^2) \quad (2.12)$$

$$P_{gq}(z) = \frac{4}{3} \left(\frac{1+(1-z)^2}{z} \right) = P_{gq}(z) \quad (2.13)$$

$$P_{gg}(z) = 6 \left(\frac{z}{(1-z)_+} + \frac{1-z}{z} + z(1-z) \right) + \frac{33-2n_F}{6} \delta(1-z). \quad (2.14)$$

2.2 Jet Production

As mentioned before, in high energy collisions, the outgoing partons are observed as collimated sprays of colorless hadrons, called jets, that travel in roughly the same direction as the

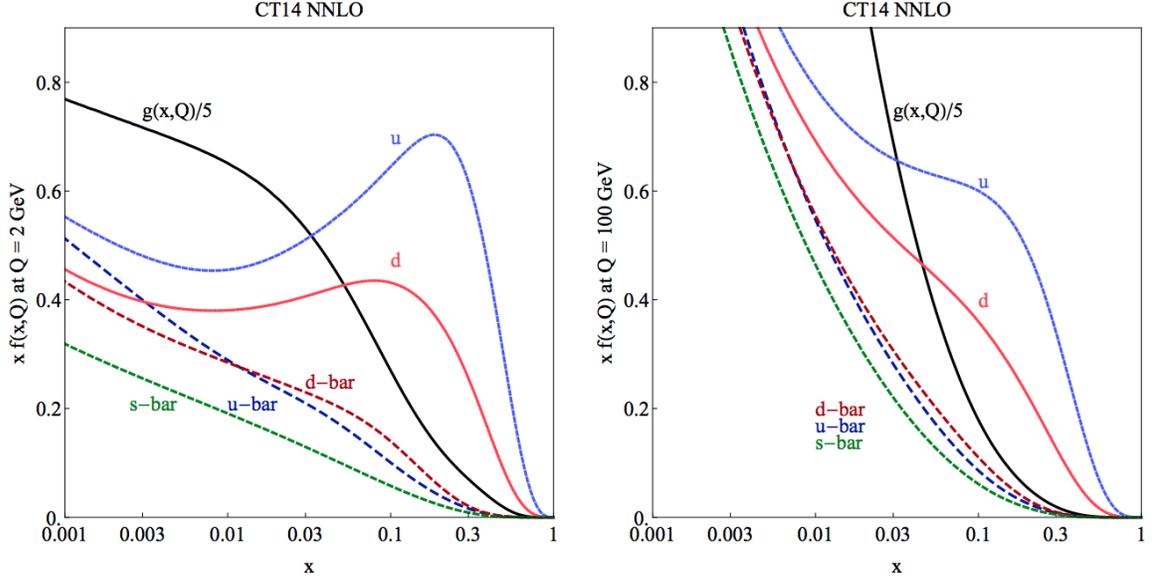


Figure 9: The proton PDF for u , \bar{u} , d , \bar{d} , s and g at 2 GeV (left) and 100 (right) GeV [63]. The PDF for s and \bar{s} are the same.

parent partons. At the start of the jet production process, the initial-state or final-state partons can either radiate gluons or split into quark and antiquark pairs. These secondary emission processes are known as parton branching. As the distance between the outgoing partons increases, the strength of their coupling increases rapidly and virtual quark and antiquark pairs are created from the vacuum which merge with the outgoing partons to form colorless hadrons. This process is called hadronization.

In the following sections, the production of jets will be discussed. Included in the discussion is the calculation of the cross section of the hard process and the simulation of jet production in high energy proton-proton collisions.

2.2.1 The Hard Scattering Process

Consider the $2 \rightarrow 2$ hard scattering of high energy partons. At LO pQCD, the differential cross section for this process is proportional to the square of the LO matrix element $|M|^2$ and is written as:

$$\frac{d\sigma}{d\Omega} = \frac{\alpha_s^2}{4\hat{s}} |M|^2, \quad (2.15)$$

where Ω is the solid angle of the scattering partons, and \hat{s} is one of the Mandelstam variables \hat{s} , \hat{t} , and \hat{u} , defined as:

$$\hat{s} = (p_A + p_B)^2 \quad (2.16)$$

$$\hat{t} = (p_A - p_C)^2 \quad (2.17)$$

$$\hat{u} = (p_A - p_D)^2 \quad (2.18)$$

where p_A , p_B , p_C , and p_D are the 4-momentum of the incoming and outgoing partons for the $AB \rightarrow CD$ process. As the energies and momenta of the partons in the hard process are large, the parton masses can be assumed to be zero. Given the massless parton approximation, the Mandelstam variables satisfy the following relation

$$\hat{s} + \hat{t} + \hat{u} = 0. \quad (2.19)$$

The QCD $2 \rightarrow 2$ hard scattering includes multiple processes, which are defined according to the species of the initial and final state partons. The Feynmann diagrams for these processes are

summarized in Figure 10. These diagrams can be categorized into three types: the s-, t-, and u-channel diagrams. As an example, the amplitude M for the three diagrams shown from left to right in Figure 10-d is proportional to $\frac{1}{\hat{s}}$, $\frac{1}{\hat{t}}$, and $\frac{1}{\hat{u}}$, respectively, representing the s-, t-, and u-channel diagrams. To calculate $|M|^2$ for a given process, all the diagrams that contribute to that process must be added together. For example, for the $qq \rightarrow qq$ process that has Feynmann diagrams as shown in Figure 10 b, $|M|^2$ is written as

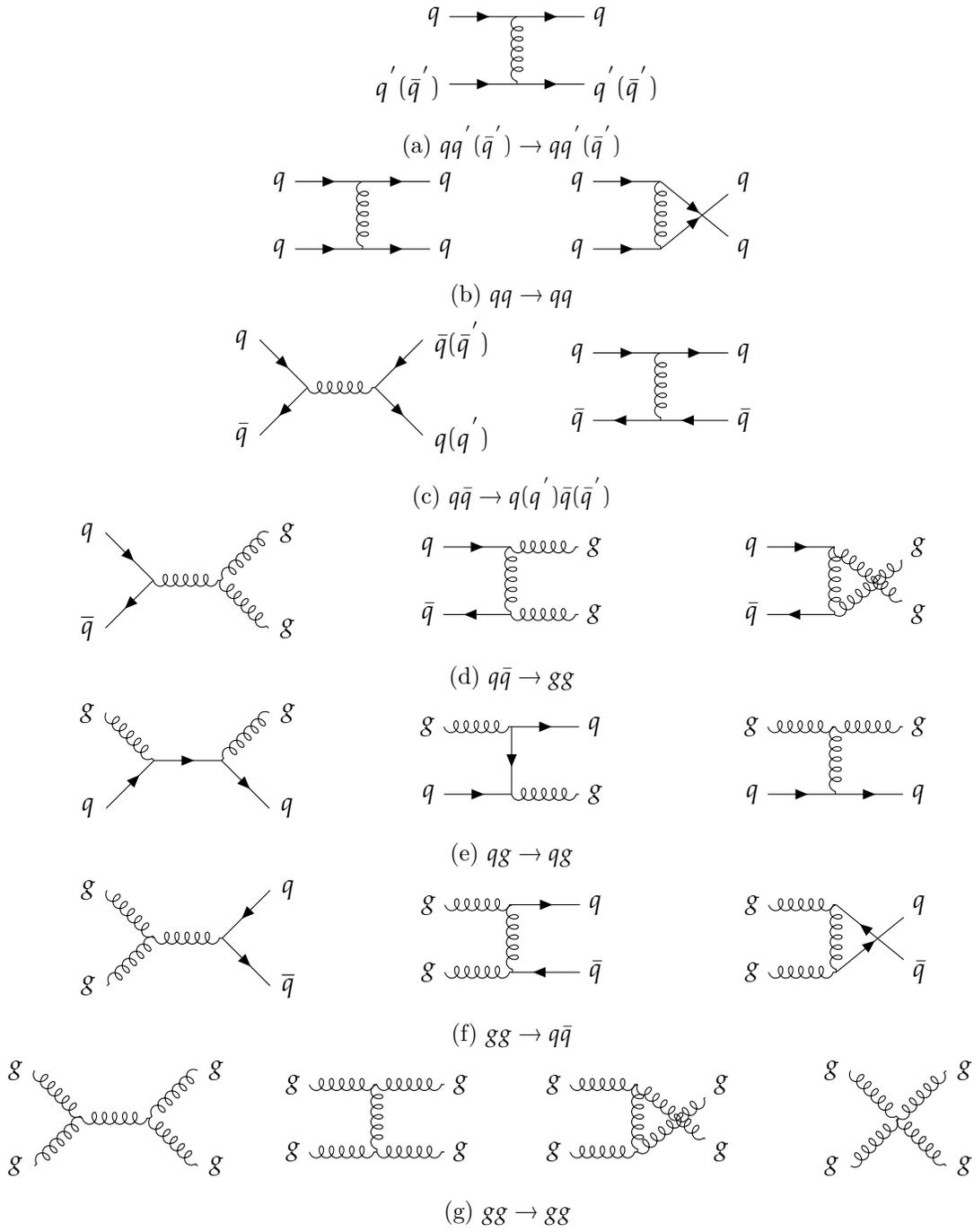
$$\begin{aligned}
 |M_{qq \rightarrow qq}|^2 &= |M_{t\text{-channel}} + M_{u\text{-channel}}|^2 \\
 &= |M_{t\text{-channel}}|^2 + |M_{u\text{-channel}}|^2 + |M_{t\text{-channel}}^* M_{u\text{-channel}}| + |M_{u\text{-channel}}^* M_{t\text{-channel}}|
 \end{aligned}
 \tag{2.20}$$

The last two terms in Equation 2.20 are the interference terms. The LO $|M|^2$ for all the processes, summed over final state color and spins and averaged over initial state color and spins, are summarized in Table III.

Higher-order pQCD calculations for $2 \rightarrow N$ ($N > 2$) processes get complicated very quickly due to the number of sub-processes contributing in the final state. For this reason, high order calculations do not exist for many processes, and predictions of jet production have to rely on Monte Carlo (MC) event generators.

2.2.2 Monte Carlo Event Generators

MC event generators are software packages used to simulate high-energy proton-proton collision events. The simulation of high energy hadronic collision events is important in high energy experiments for comparing experimental data to theoretical predictions, estimating the

Figure 10: Summary of the Feynman diagrams of the QCD $2 \rightarrow 2$ subprocesses.

Process	$ M ^2$
$qq'(\bar{q}') \rightarrow qq'(\bar{q}')$	$\frac{4}{9} \left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2} \right)$
$qq \rightarrow qq$	$\frac{4}{9} \left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2} + \frac{\hat{s}^2 + \hat{t}^2}{\hat{u}^2} - \frac{2}{3} \frac{\hat{s}}{\hat{u}\hat{t}} \right)$
$q\bar{q} \rightarrow q'\bar{q}'$	$\frac{4}{9} \left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2} \right)$
$q\bar{q} \rightarrow q\bar{q}$	$\frac{4}{9} \left(\frac{\hat{s}^2 + \hat{u}^2}{\hat{t}^2} + \frac{\hat{u}^2 + \hat{t}^2}{\hat{s}^2} - \frac{2}{3} \frac{\hat{u}^2}{\hat{s}\hat{t}} \right)$
$q\bar{q} \rightarrow gg$	$\frac{32}{27} \left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{t}\hat{u}} - \frac{9}{4} \frac{\hat{u}^2 + \hat{t}^2}{\hat{s}^2} \right)$
$qg \rightarrow qg$	$\frac{4}{9} \left(-\frac{\hat{s}^2 + \hat{u}^2}{\hat{s}\hat{u}} + \frac{9}{4} \frac{\hat{u}^2 + \hat{s}^2}{\hat{t}^2} \right)$
$gg \rightarrow q\bar{q}$	$\frac{1}{6} \left(\frac{\hat{t}^2 + \hat{u}^2}{\hat{t}\hat{u}} - \frac{9}{4} \frac{\hat{u}^2 + \hat{t}^2}{\hat{s}^2} \right)$
$gg \rightarrow gg$	$\frac{9}{2} \left(3 - \frac{\hat{s}\hat{u}}{\hat{t}^2} - \frac{\hat{t}\hat{u}}{\hat{s}^2} - \frac{\hat{t}\hat{s}}{\hat{u}^2} \right)$

TABLE III: LO matrix element for QCD $2 \rightarrow 2$ subprocesses.

various reconstructed object efficiencies, understanding the calibration of detectors, estimating the background processes to various observables, calculating trigger rates for various running conditions, etc. There are a number of MC event generators available, however, they all work in a similar manner. The event simulation begins with the hard scattering process, which is typically calculated to LO in QCD, and then initial-state radiation (ISR) and final-state radiation (FSR) effects are added. Furthermore, other partons from the colliding protons can interact as well and these secondary interactions form the underlying event (UE) which is also

taken into account by the event generators. After all these processes have been modeled, the resultant partons are grouped together to form the final state hadrons based on a hadronization model. The event generator can then either allow the hadrons to decay, or leave this step for later in the detector simulation.

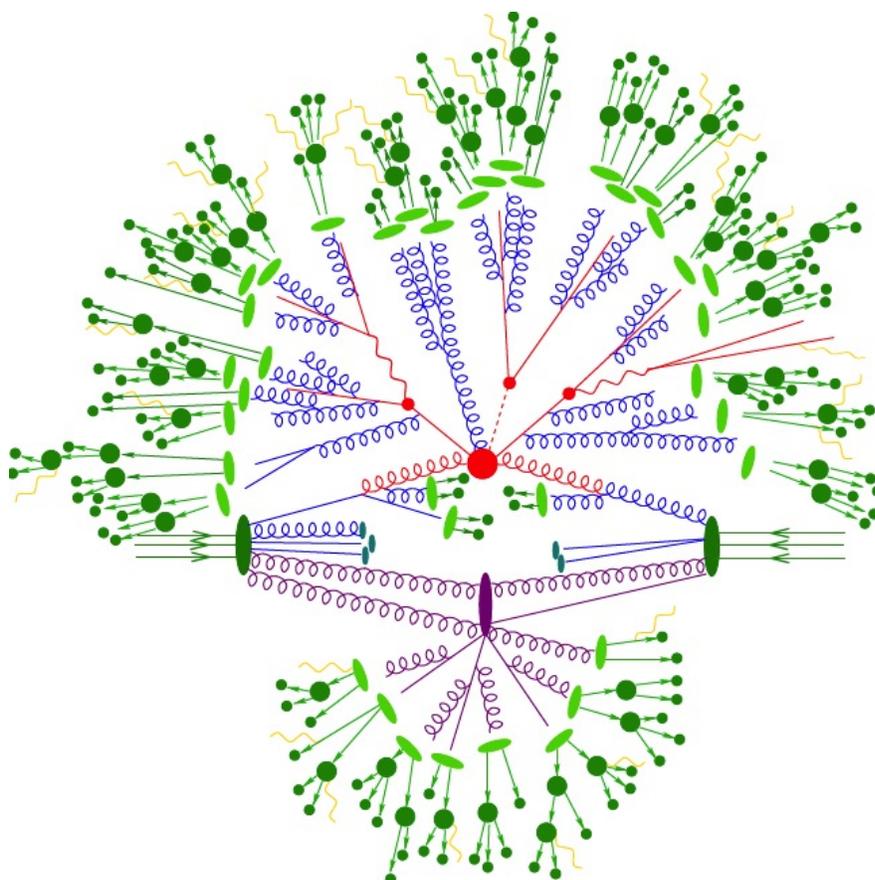


Figure 11: Various processes in a collision event including the hard process (shown in red), ISR and FSR (shown in blue), hadronization (shown in green), decays and QCD Bremsstrahlung of unstable hadrons (shown in yellow and dark green, respectively), and UE (shown in purple) [64].

An illustration of the processes involved in a proton-proton collision is shown in Figure 11. In the figure, the hard process is shown in red, ISR and FSR are shown in blue, the hadronization is shown in green, decays and QED Bremsstrahlung of unstable hadrons are shown in yellow and dark green, respectively, and interactions between the proton remnants are shown in purple.

MC programs that can simulate all these processes are called general purpose MC event generators. Two general purpose MC generators are used in this thesis: PYTHIA [65, 66] and HERWIG++ [67, 68]. These generators calculate the matrix elements of the hard processes of the SM and various NP models at LO. For QCD, these generators calculate the LO matrix elements of the $2 \rightarrow 2$ process, which are then combined with the parton shower (PS) to include the ISR and FSR effects.

The Parton Shower Method: Parton Shower (PS) is a method used in MC event generators to simulate the FSR and ISR effects. It makes use of the splitting functions $P_{ab}(z)$ as described in Section 2.1.3. The probability for the parton branching process $a \rightarrow bc$, $P_{a \rightarrow bc}$, can be written as

$$dP_{a \rightarrow bc} = \frac{d\theta}{\theta} dz P_{ab}(z) \quad (2.21)$$

where the two partons are separated by an angle θ . Using this probability, a cascade of splittings can be constructed, where at each splitting the 4-momentum of the incoming parton is shared by the outgoing partons. The PS method is different from a full high-order pQCD calculation in that in the PS method loop diagrams are not included. The PS method continues to produce partons from the secondary emissions until the partons reach the energy scale of ~ 1 GeV, at which point, α_s becomes too large and the PS calculation are no longer applicable. The PS for

the ISR and FSR are treated separately. Since the matrix elements are calculated first, ISR is typically propagated backward in time from the hard process, while FSR is propagated forward in time.

Hadronization: At energy scales below 1 GeV, the hadronization process takes over. This process is parametrized using phenomenological models in the MC event generators. The hadronization models employed by PYTHIA and HERWIG++ are the Lund string model [69] and the cluster model [70], respectively.

Underlying Event: UE is included using phenomenological models based upon multiple parton interactions (MPI) [71–74]. The parameters in the MPI models cannot be determined from first principles, and must be determined using data. Each set of these parameters is called a Tune. The Tunes used in this thesis are Tune CUETP8M1 [75, 76] for PYTHIA and Tune CUETHS1 [76, 77] for HERWIG++.

While the general purpose MC generators only calculate the LO matrix elements for the simplest processes (e.g., the QCD $2 \rightarrow 2$ process), there exist generators that can calculate higher-order matrix elements. However, these generators need to interface with general purpose MC generators to include the effects of FSR, ISR, hadronization, and UE. These generators are called matrix element generators. The matrix element generators used in this thesis are MADGRAPH5_aMC@NLO [78, 79], NLOJET++ [80], CIJET [81], and QBH [82]. MADGRAPH5_aMC@NLO can calculate the matrix elements for various SM and NP processes at LO or NLO, depending on the process. For QCD, MADGRAPH5_aMC@NLO can calculate the LO matrix elements for $2 \rightarrow 2$ to $2 \rightarrow 7$ processes. NLOJET++ is able to calculate the NLO matrix

element for QCD $2 \rightarrow 3$ process. The CIJET and QBH programs are generators dedicated to the CI and QBH models.

2.3 Jet Algorithms

In order to compare theoretical predictions from pQCD with experimental results, jet algorithms need to be employed. The jet algorithms link together the final state stable particles produced through the parton branching and hadronization processes to form a single object, a jet, whose 4-momentum represents the 4-momentum of the outgoing parton.

An important requirement of a jet algorithm is to be collinear-safe, meaning that the splitting of a parton inside a jet into two collinear partons will not affect the jet properties. Jet algorithms developed in the 1980s and 1990s, such as the cone algorithms [83] which cluster particles based upon their locations with respect to an angular cone around a seed particle that has energy above a given threshold, are not collinear-safe because the collinear splitting of the jet seed could cause the seed p_T to fall below the threshold, resulting in no jet being found. This is illustrated in Figure 12 top.

Another important property of a jet algorithm is infrared-safety, which means the emission of a low-energy parton from partons inside a jet will not affect the jet properties. The cone algorithms are not infrared-safe as illustrated in Figure 12 bottom. As can be seen in the figure, if the emitted gluon is above the seed threshold, the two jets from the outgoing partons can be merged into one.

The sequential recombination algorithms [85–87] used in this thesis were developed in the late 1990s and early 2000s, and satisfy the requirements of being collinear- and infrared-safe.

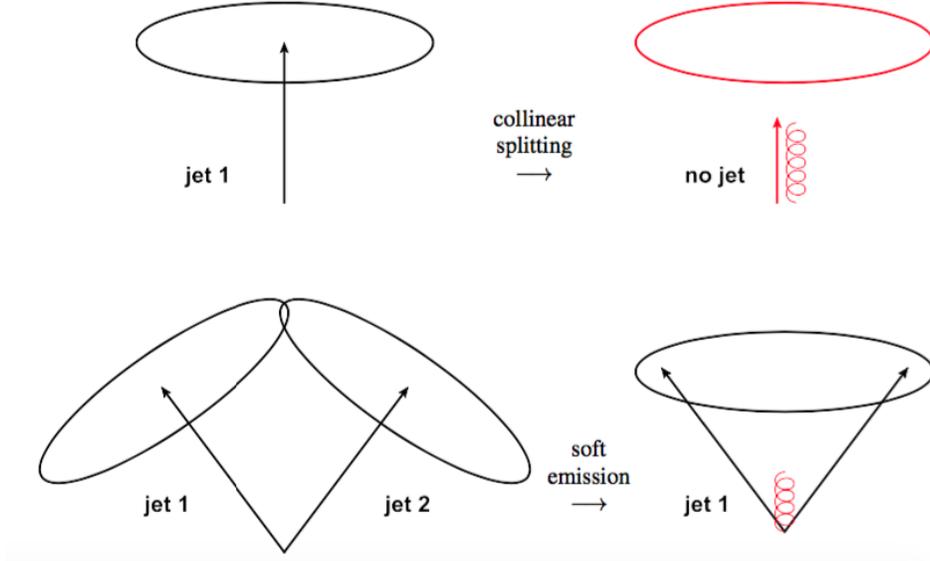


Figure 12: An example of the infrared- and collinear-unsafety behaviors of the cone algorithms. Top: the collinear splitting of the jet seed leads to the seed p_T below seed threshold. Bottom: the emission of a soft gluon leads to the merging of two jets [84].

These algorithms are based on an iterative method, in which the closest pair of objects (particles) are combined sequentially. In these algorithms, the following distance variables are defined:

$$d_{ij} = \min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (2.22)$$

and

$$d_{iB} = p_{T,i}^{2p} \quad (2.23)$$

where d_{ij} represents the distance between two objects labeled i and j , d_{iB} is the distance between object i and the beam, and

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2. \quad (2.24)$$

The distance d_{ij} is scaled by the parameter R , which is referred to as the jet size parameter. p_T , η , and ϕ are the jet's transverse momentum, rapidity, and azimuthal angle, respectively (see next section). The value of the power p in Equation 2.22 and Equation 2.23 can be $+1$ for the k_t [85], 0 for the Cambridge/Aachen (CA) [86], or -1 for the anti- k_t [87] algorithm.

The sequential recombination jet algorithm proceeds as follows: first, d_{ij} values for each (i,j) pair of the N input objects and d_{iB} values for each input object are computed. Then, the minimum from all the d_{ij} and d_{iB} is found. If the minimum value is among the d_{iB} values, the object i will be declared as a jet and the algorithm will start again using the rest of the objects that have not been declared as jets. If the minimum value is not among the d_{iB} values, the objects i and j with the minimum value of d_{ij} will be combined into a new object and the algorithm will start again, with the new object replacing the objects i and j . The 4-momentum of the new object will be the sum of the 4-momenta of objects i and j . The algorithm continues to run until all the objects are clustered into jets. The workflow of the algorithm is illustrated in Figure 13.

It is interesting to compare the shape of jets reconstructed by the k_t , CA, and anti- k_t algorithms. The k_t algorithm first clusters low- p_T objects and create jets that have irregular-shaped areas. The anti- k_t algorithm, with $p = -1$, starts with the highest- p_T objects and produces jets that have much rounder shapes, similar to the jets from cone algorithms. The

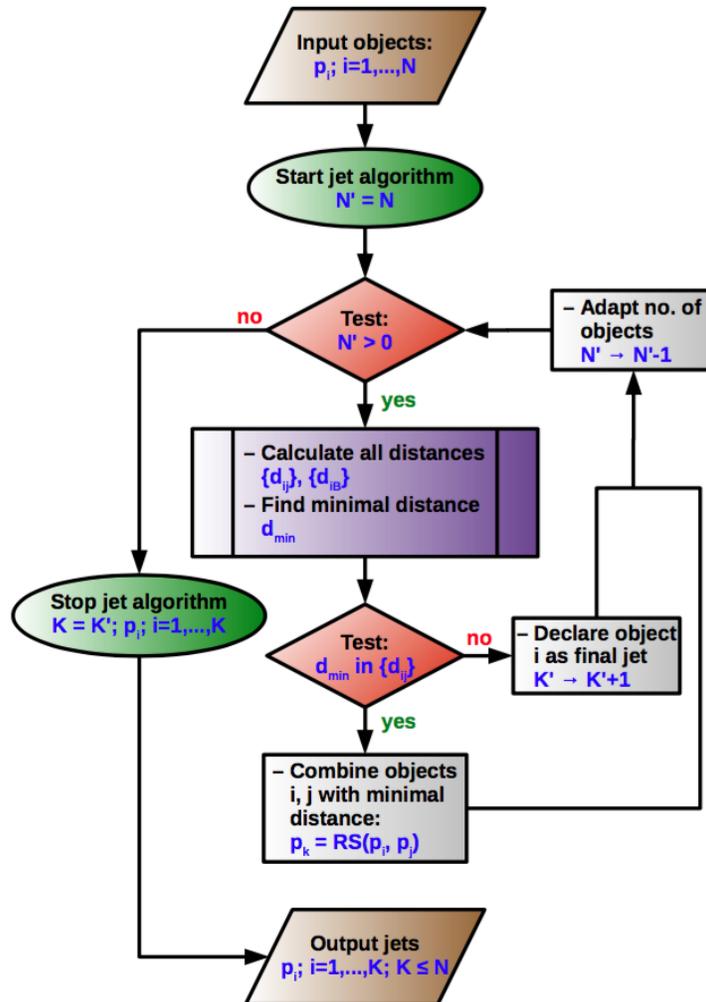


Figure 13: The workflow chart for the sequential recombination algorithm. The algorithm starts with N objects (particles), which are given in the form of the 4-momentum p_i , and produces K jets [84].

CA algorithm, with $p = 0$, combines the particles which are geometrically nearest to each other and produces jets that are more round than those from the k_t algorithm, but not as round as those from the anti- k_t algorithm. An example of the jet shapes produced by the k_t , CA, and anti- k_t algorithms is shown in Figure 14. In the figure, the shaded jet areas are determined by clustering additional large numbers of input objects which have negligible momentum and uniformly distributed in the y - ϕ plane (“ghost particles”).

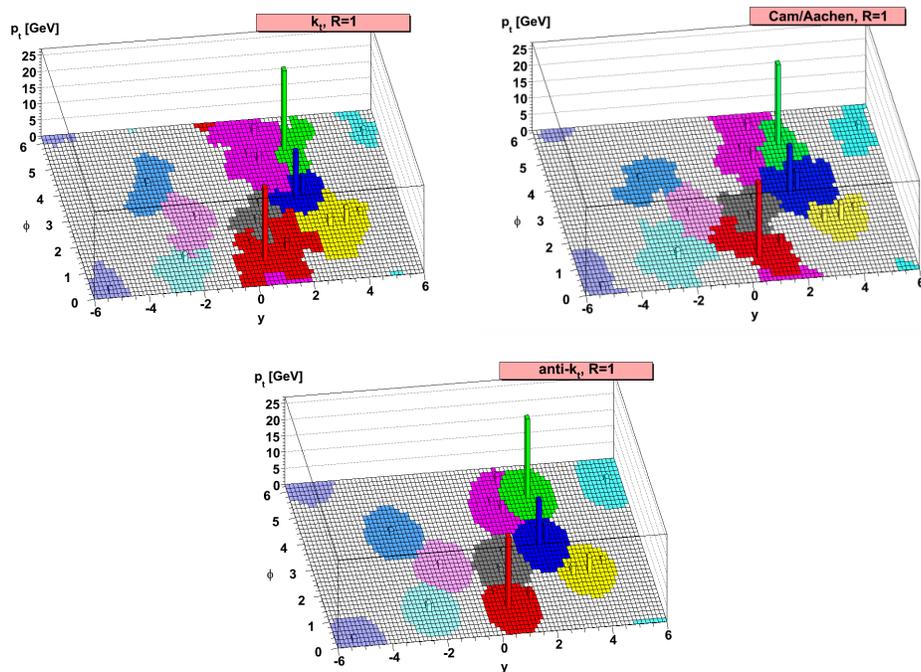


Figure 14: The jet areas for the k_t , CA, and anti- k_t jet algorithms [87].

2.4 Dijet Angular Distributions

In this thesis, measurement of the dijet angular distributions is presented. Dijets are defined as the two highest- p_T jets produced in the collision. The dijet angular distributions probe the dynamics of the underlying hard processes and can be used to confront the predictions of pQCD and to search for signatures of NP. This section discusses the kinematics of dijet events in proton-proton collisions.

2.4.1 The Kinematics of Dijet Events

An illustration of a dijet event in the partonic center-of-mass and lab (proton-proton) frames is shown in Figure 15 left and right, respectively. In the figure, the z-axis is coincident with the beam direction. The kinematics of such events can be described by either the Cartesian components of the momenta, $\{p_x, p_y, p_z\}$, or by the spherical components $\{\theta, \phi, p\}$ of the momenta, where ϕ is the azimuthal angle, θ is the polar angle with respect to the beam line, and p is the magnitude of the momentum.

Assuming the transverse motion of the initial state partons and the parton radiations are negligible, the two jets have to be back-to-back in the parton-parton center-of-mass frame. The value of the polar angle θ^* will be the same for both jets. However, since the longitudinal momenta of the initial state partons are unknown and do not have to be equal to each other, the center-of-mass frame is boosted in the z-direction. While p_T and ϕ are invariant under Lorentz boosts, but θ is not. Therefore, it is typical to replace θ with a quantity which has better properties under Lorentz boosts when describing the kinematics of the proton-proton

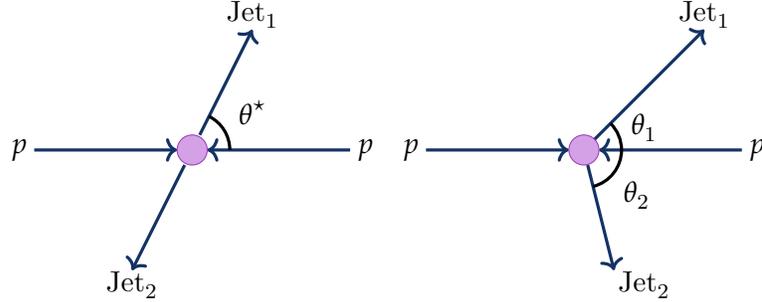


Figure 15: An illustration of a dijet event in the partonic center-of-mass (left) and lab (right) frames.

collision events. Such a quantity is the rapidity, denoted as y . For a particle with 4-momentum (p_x, p_y, p_z, E) , y is defined as

$$y = \frac{1}{2} \ln \frac{E - p_z}{E + p_z} = \tanh^{-1} \left(\frac{p_z}{E} \right). \quad (2.25)$$

The 4-momentum of the particle can then be expressed as

$$p = (m_T \cosh y, p_T \cos \phi, p_T \sin \phi, m_T \sinh y). \quad (2.26)$$

The transverse mass m_T is defined as $m_T^2 = p_T^2 + m^2$, where m is the mass of the particle (or jet). Although the rapidity itself is still not invariant under Lorentz boosts in the z -direction, the difference between the rapidities for the two particles (or jets) is invariant.

Related to the rapidity is the psuedo-rapidity η defined as:

$$\eta = -\ln \tan \frac{\theta}{2}. \quad (2.27)$$

If the particle has zero mass, y is the same as η . Also, for highly relativistic particles, we can approximate y with η . In such cases, the four-momentum of each outgoing parton (jet) can be expressed as

$$p = p_T(\cosh \eta, \cos \phi, \sin \phi, \sinh \eta). \quad (2.28)$$

As this thesis considers only high- p_T jets, the parton (jet) masses can be assumed to be negligible. For the scattering of massless partons which are traveling collinear with the beam, the longitudinal boost of the partonic center-of-mass frame with respect to the lab frame, y_{boost} , and θ^* are obtained from the rapidities of the two jets from the two scattered partons by:

$$y_{\text{boost}} = \frac{1}{2}(y_1 + y_2) \quad \text{and} \quad |\cos \theta^*| = \tanh y^*, \quad (2.29)$$

where $y^* = \frac{1}{2}|y_1 - y_2|$, and $\pm y^*$ are the rapidities of the two jets in the partonic center-of-mass frame.

The center-of-mass energy for the colliding partons is equivalent to the invariant mass of the dijets, i.e.,

$$\hat{s} = M_{\text{jj}}^2 = p_1^\mu p_{2,\mu}. \quad (2.30)$$

Using Equation 2.28, we have

$$M_{jj}^2 = 4p_T^2 \cosh^2 y^*. \quad (2.31)$$

The Mandelstam variables (Equation 2.16) \hat{t} and \hat{u} can then be written in terms of M_{jj} and $\cos \theta^*$ as

$$\hat{t} = -\frac{1}{2}M_{jj}^2(1 - \cos \theta^*) \quad (2.32)$$

$$\hat{u} = -\frac{1}{2}M_{jj}^2(1 + \cos \theta^*). \quad (2.33)$$

2.4.2 The Dijet Angular Distributions

The dijet angular distribution is studied using the variable χ_{dijet} , defined as:

$$\chi_{\text{dijet}} = e^{|y_1 - y_2|} = \frac{1 + |\cos \theta^*|}{1 - |\cos \theta^*|} = \frac{\hat{u}}{\hat{t}}. \quad (2.34)$$

The choice of χ_{dijet} , rather than θ^* , to measure the dijet angular distribution is motivated by the fact that in Rutherford scattering, which is dominated by small angle scatterings, the χ distribution is uniform. This follows from the fact that in Rutherford scattering, the angular distribution has the following form:

$$\frac{d\sigma}{d \cos \theta^*} \propto \frac{1}{\sin^4 \frac{\theta^*}{2}} = \frac{1}{(1 - |\cos \theta^*|)^2} \quad (2.35)$$

which can be expressed in terms of χ as:

$$\frac{d\sigma}{d\cos\theta^*} \propto (\chi + 1)^2. \quad (2.36)$$

The χ distribution can then be written as

$$\frac{d\sigma}{d\chi} = \frac{d\sigma}{d\cos\theta^*} \frac{d\cos\theta^*}{d\chi} = \left(\frac{1}{\chi + 1} - \frac{\chi - 1}{(\chi + 1)^2} \right) \frac{d\sigma}{d\cos\theta^*} = \text{Constant}. \quad (2.37)$$

Similarly, the χ distributions for the QCD $2 \rightarrow 2$ process are determined as

$$\frac{d\sigma}{d\chi} = \frac{d\sigma}{d\cos\theta^*} \frac{d\cos\theta^*}{d\chi} = 2\pi \frac{d\sigma}{d\Omega} \frac{d\cos\theta^*}{d\chi} = \frac{1}{(1 + \chi)^2} \frac{\pi\alpha_s^2}{2\hat{s}} |M|^2. \quad (2.38)$$

The LO $|M^2|$ of the QCD $2 \rightarrow 2$ subprocesses, shown in Table III, can be expressed in terms of χ as shown in Table IV. The $|M^2|$ distributions divided by $(1 + \chi)^2$, as a function of χ are shown in Figure 16. In Figure 16, processes that have the same initial states are combined. The χ distributions for nearly all the QCD subprocesses are relatively uniform in χ . The exception is the $gg \rightarrow jj$ process, in which the contribution from the s-channel diagrams are large. This is because s-channel diagram has an isotropic scattering angle distribution, i.e., $\frac{d\sigma}{d\cos\theta^*} \sim 1$, and when multiplied by $\frac{1}{(1 + \chi)^2}$, a peak is produced at low χ since $\frac{1}{(1 + \chi)^2}$ peaks at low χ .

The cross sections for NP processes are typically dominated by the contribution from the s-channel diagram (see the previous chapter) and therefore will produce an excess in the χ_{dijet} distribution at low χ_{dijet} relative to the prediction from QCD. This excess signifies the presence

Subprocess	$ M ^2$
$qq'(\bar{q}') \rightarrow qq'(\bar{q}')$	$\frac{4}{9}[(\chi+1)^2 + \chi^2]$
$qq \rightarrow qq$	$\frac{4}{9}[(2\chi^2 + 2\chi + 2 + \frac{2}{\chi} + \frac{2}{\chi^2}) - \frac{2}{3}(\chi + 2 + \frac{1}{\chi})]$
$q\bar{q} \rightarrow q'\bar{q}'$	$\frac{4}{9}[\frac{1+\chi^2}{(1+\chi)^2}]$
$q\bar{q} \rightarrow q\bar{q}$	$\frac{4}{9}[(\chi+1)^2 + \chi^2 + \frac{1+\chi^2}{(1+\chi)^2} + \frac{2}{3}\frac{\chi^2}{(1+\chi)}]$
$q\bar{q} \rightarrow gg$	$\frac{32}{27}[\chi + \frac{1}{\chi} - \frac{9}{4}\frac{1+\chi^2}{(1+\chi)^2}]$
$qg \rightarrow qg$	$\frac{4}{9}[1 + \frac{1}{\chi} + \frac{\chi}{1+\chi} + \frac{9}{4}(2\chi+1+2\chi^2)]$
$gg \rightarrow q\bar{q}$	$\frac{1}{6}[\chi + \frac{1}{\chi} - \frac{9}{4}\frac{1+\chi^2}{(1+\chi)^2}]$
$gg \rightarrow gg$	$\frac{9}{2}[3 + \chi^2 + \chi + \frac{1}{\chi} + \frac{1}{\chi^2} - \frac{\chi}{(1+\chi)^2}]$

TABLE IV: The square of the LO matrix element for the QCD $2 \rightarrow 2$ subprocesses expressed in χ .

of NP. This is illustrated in Figure 17. The MC generator predictions of the χ_{dijet} distributions in QCD and NP models will be presented in Chapter 7.

Previous Experimental Results Measurements of dijet angular distributions at the Sp \bar{p} S by the UA1 Collaboration [13], at the Fermilab Tevatron by the CDF [16, 17] and D0 [18, 20, 21] Collaborations, and at the LHC by the ATLAS [22–27] and CMS [29, 30, 32, 33] Collaborations have previously been reported. The most recent measurement performed by the CMS

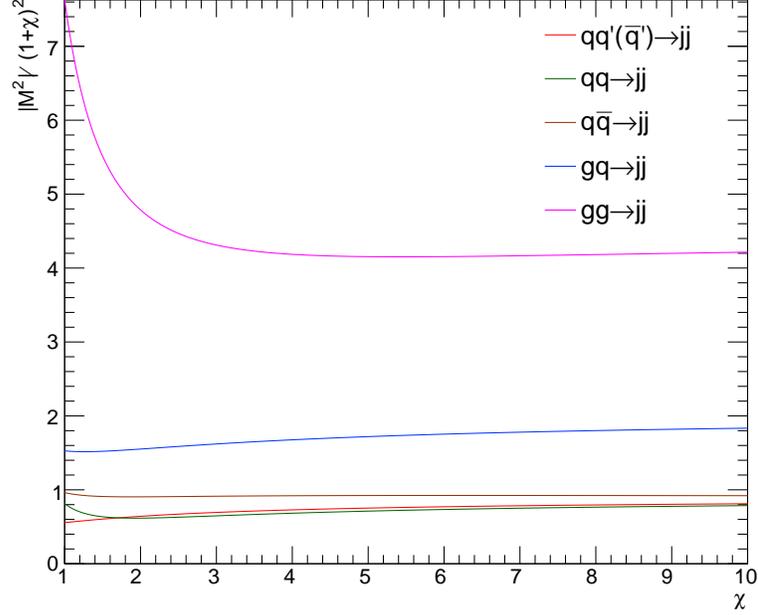


Figure 16: The square of the LO matrix element for the QCD $2 \rightarrow 2$ subprocesses, expressed in χ , and divided by $(1 + \chi)^2$. For subprocesses which have the same initial states, the matrix elements are combined.

Collaboration used 3 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ [33], and was used to constrain the quark CI, ADD, and QBH models.

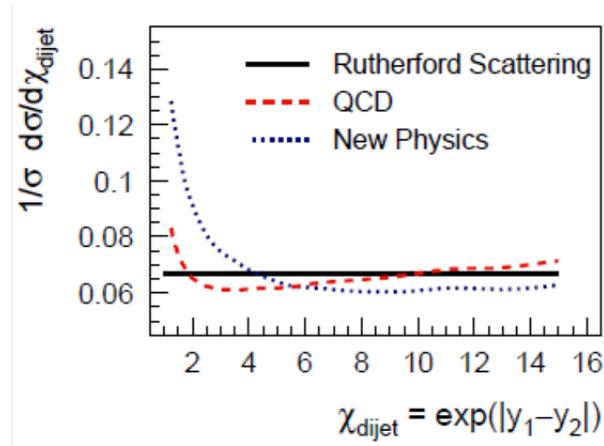


Figure 17: An illustration of the χ distributions from Rutherford scattering, QCD, and NP models. NP models, which have a more isotropic scattering angle distribution than those given by QCD, can be identified as an excess at small values of χ .

CHAPTER 3

THE LARGE HADRON COLLIDER AND THE COMPACT MUON SOLENOID DETECTOR

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [3], built by the European Organization for Nuclear Research (CERN), is the largest and most powerful particle accelerator in the world. It is located near Geneva, Switzerland and is contained in a 27 km circular tunnel buried 40–175 m beneath the ground (see Figure 18).

Inside the tunnel, two beams of charged particles travel in separate beampipes and opposite directions with velocities close to the speed of light. The beampipes are surrounded by superconducting magnets, mainly dipoles and quadrupoles, which focus and accelerate the particles and guide them around the accelerator. Depending on the mode of operation, the particle beams consist of either protons or lead-ions. The design beam energies are 7 TeV and 2.76 TeV per nucleon for proton beams and Pb beams, respectively. For head-on collisions of two beams with equal energy and traveling in opposite directions, the center-of-mass energy for the collisions (\sqrt{s}) is twice the beam energy.

The proton beams are created by putting hydrogen gas in an electric field and stripping away the electrons. The beams are then accelerated to the energy of 450 GeV in a series of pre-acceleration steps before they are finally injected into the LHC and brought to 7 TeV. The

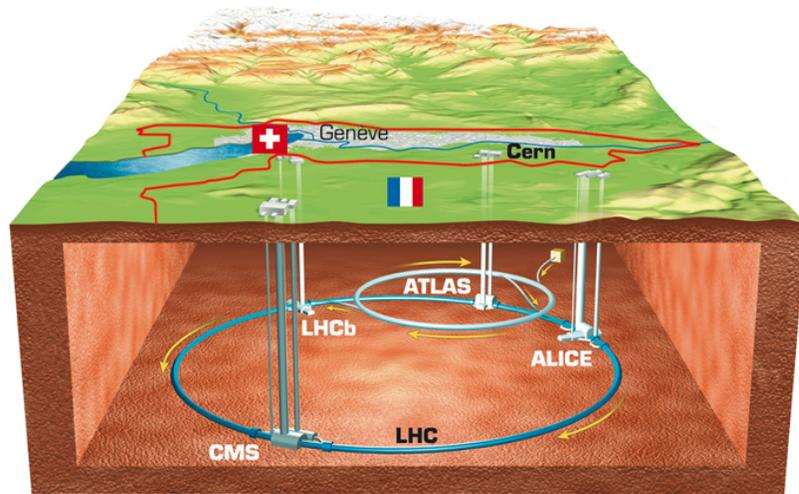


Figure 18: The LHC accelerator and detectors at the LHC.

pre-acceleration steps are performed in accelerators including the linear accelerator (LINAC), which focuses protons to form proton beams and accelerates the beams to the energy of 50 MeV, the Proton Synchrotron Booster, which accelerates the beams to the energy of 1.4 GeV, the Proton Synchrotron, which brings the beams to the energy of 26 GeV, and the Super Proton Synchrotron, where the beams reach the energy of 450 GeV.

There are four points along the accelerator ring where the beampipes cross, and particles are brought to collide. These points are called the interaction points (IPs). Four major experiments are installed at the 4 IPs (as can be seen in Figure 18). The ATLAS (A Toroidal LHC

ApparatuS) and CMS (Compact Muon Solenoid) experiments are general-purpose detectors which are designed to study a broad range of physical phenomena. The LHCb (Large Hadron Collider beauty) Experiment aims to study the decays of B-mesons with high precision and high statistics. ALICE (A Large Ion Collider Experiment) focuses on the investigation of heavy ion collisions.

The proton beams are split up into discrete packets called “bunches”, which are separated by 7.5 m. This implies that the collisions happen every 25 ns (i.e., $25\text{ns} \sim 7.5\text{m}/c$ where c is the speed of light), and the maximum number of bunches in the LHC tunnel is 2808. The bunches are formed in the Radio Frequency Cavities in the LINAC. For proton beams, each bunch contains about 10^{11} protons.

The instantaneous luminosity, \mathcal{L} , characterizes the proton-proton collision rate, and is expressed as:

$$\mathcal{L} = \frac{N_1 N_2 N_b \gamma f F}{4\pi\sigma^2}, \quad (3.1)$$

where N_1 and N_2 are the number of particles in beam 1 and beam 2, N_b is the number of bunches per beam, f is the orbit frequency, F is the geometric luminosity reduction factor due to the crossing angle at the IPs, and σ is the cross sectional area of the beam. The instantaneous luminosity is expressed in units of inverse cross section per second, typically Hz/nb, Hz/pb, or Hz/fb.

For a process with cross section σ , the interaction rate can be expressed as:

$$\frac{dN_{\text{ev}}}{dt} = \mathcal{L}\sigma. \quad (3.2)$$

The total number of interaction events in a period of running time of the collider is called the integrated luminosity L and is given by:

$$L = \int \mathcal{L} dt. \quad (3.3)$$

To increase the event rate for rare processes, the LHC operates in high luminosity conditions. An important consideration in the high luminosity environment is the number of instantaneous collisions per bunch crossing (pileup). The number of pileup events N_{pileup} can be determined as:

$$N_{\text{pileup}} = \mathcal{L} \sigma_{\text{pp}} \cdot d, \quad (3.4)$$

where \mathcal{L} is the instantaneous luminosity, $\sigma_{\text{pp}} = 70 \text{ mb}$ (10^{-27} cm^2) is the inelastic proton-proton collision cross section, and d is the bunch spacing.

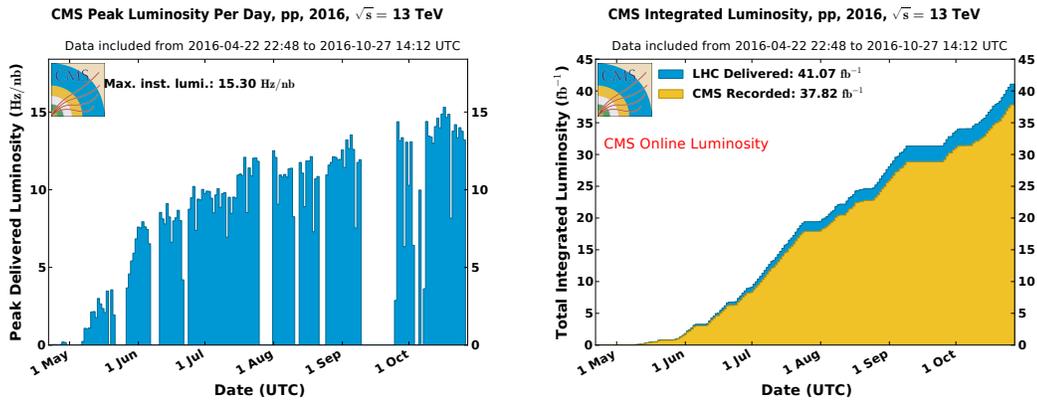


Figure 19: The instantaneous luminosity (left) and integrated luminosity (right) for the LHC in 2016.

The first LHC proton-proton collision run (Run I) took place from March 2010 to April 2013. In the first part of the run, the collision energy was 7 TeV, and later it was increased to 8 TeV. The instantaneous luminosity ranged from about 0.1 to 5 $\text{Hz} \cdot \text{nb}^{-1}$. In Run I, the LHC delivered data corresponding to the integrated luminosities of 6 fb^{-1} at $\sqrt{s} = 7$ and 23 fb^{-1} at $\sqrt{s} = 8$ TeV. Using the Run I data, the CMS and ATLAS collaborations independently made one of the most significant scientific discovery in the last several decades, the Higgs Boson [1, 2, 88]. After Run I, the LHC went into a two-year long shutdown (LS1). During the shutdown, various upgrades were made to the accelerator which allowed the LHC to increase the collision energy to 13 TeV and double the instantaneous luminosity achieved in Run I. The LHC began the Run II in April 2015. The instantaneous and integrated luminosities for the LHC in 2016 are shown in Figure 19. The analysis for this thesis is performed using data collected by the CMS experiment in 2016 that correspond to an integrated luminosity of 35.9 fb^{-1} .

3.2 The Compact Muon Solenoid Detector

The CMS Detector [89], located in Cessy France, is a general purpose detector optimized to discover the Higgs Boson and search for new physics at the LHC. It is 21.6 m long and 14.6 m in diameter and weighs 14,000 tonnes.

A central element of the CMS detector is a superconducting solenoid of 12.5 m long and 6 m internal diameter, providing an axial magnetic field with a magnitude of 3.8 T, and an iron yoke to return the flux of the magnetic field. It is the most powerful magnetic solenoid in the world. The trajectory of charged particles emerging from high-energy collisions at the

LHC will be bent in the magnetic field. The high field strength allows CMS to use a relatively compact detector design in combination with the high-granularity tracker and muon detectors to achieve an accurate measurement of the momentum of high-energy charged particles.

The CMS sub-detectors consist of a pixel and strip tracker, an electromagnetic calorimeter (ECAL), a hadronic calorimeter (HCAL), and muon detectors. These detectors are arranged as concentric cylinders around the beam line (“barrel”), with “endcap” disks arranged orthogonally to the beam at both ends of the barrel.

The CMS detector is designed to identify electrons, muons, photons, and charged and neutral hadrons, and measure the energy and momentum of these particles with good precision. The layout of the CMS detector can be seen in Figure 20. The signatures for different types of particles in the CMS detector is shown in Figure 21. The trajectories of charged particles are recorded by the tracker. The electrons and photons deposit most of their energies in ECAL, while hadrons deposit most of their energies in the HCAL. Muons are identified in the tracker and muon chambers. Neutrinos and other weakly interacting particles, (e.g., DM particles), do not interact with the CMS detectors and therefore cannot be identified directly. Instead, their presence can be inferred indirectly by measuring the total unconserved transverse momentum in each event (E_T^{miss}), given the fact that the energies and momentums for other particles produced in the event can be well determined and the nearly 4π solid angle coverage of the detector. The details of the trackers, calorimeters, and muon detectors are described in the following subsections.

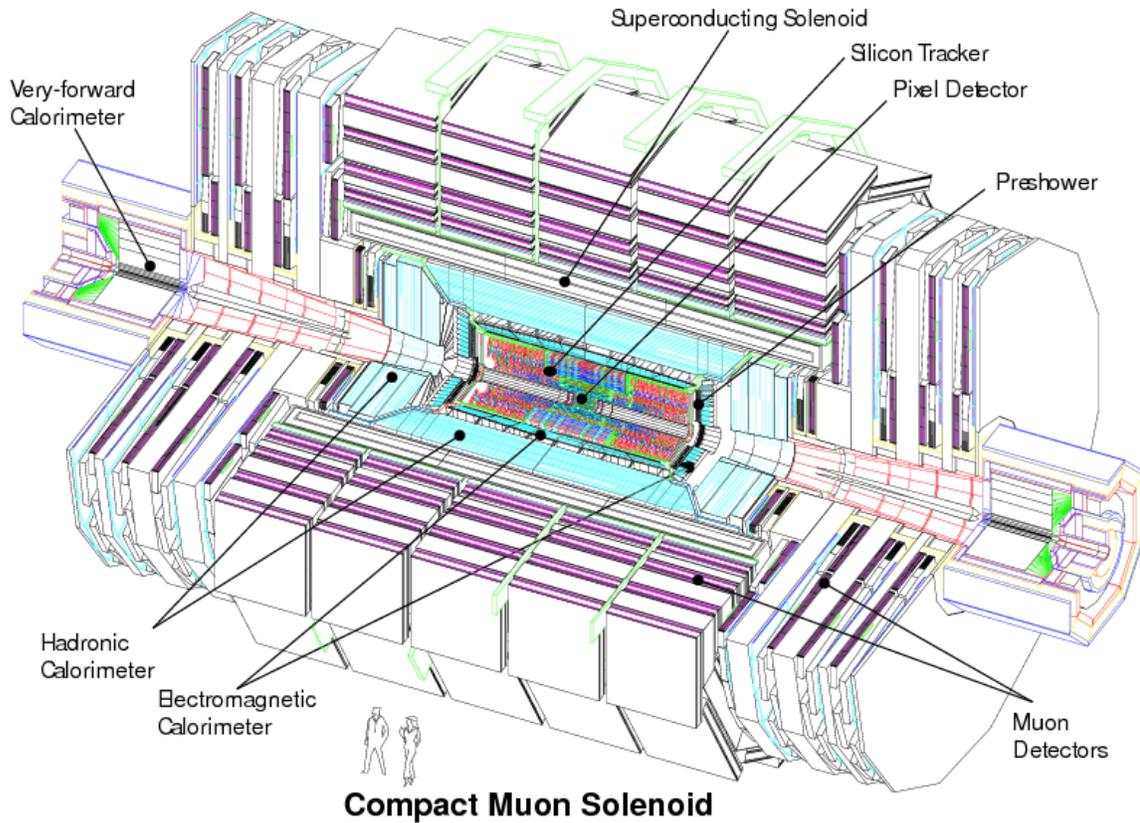


Figure 20: The layout of the CMS Detector.

The CMS coordinate system is centered at the nominal collision point with the z -axis parallel to the beam. The x -axis and y -axis are in the plane transverse to the z -axis. The x -axis points towards the center of the LHC ring and the y -axis points in the upward direction. The azimuthal angle ϕ is measured from the x -axis in the x - y plane and the polar angle θ is measured from the z -axis. The pseudorapidity, η , is commonly used to map the polar angle, where $\eta = -\ln[\tan(\theta/2)]$ (see Chapter 2).

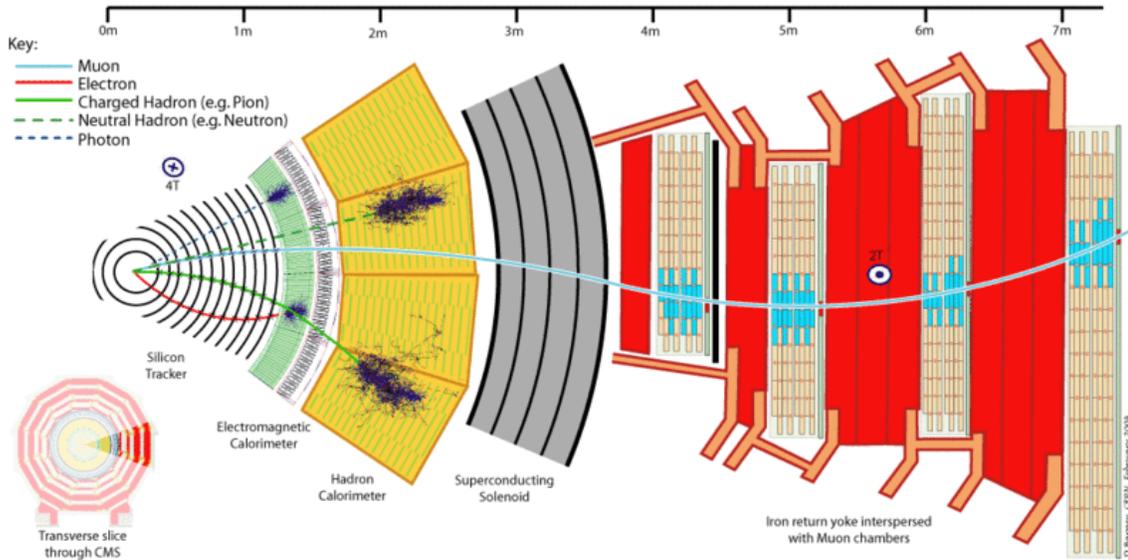


Figure 21: A slice of the CMS detector in the x-y plane, illustrating the signatures for different types of particles in the detector.

3.2.1 The Tracker

The CMS tracker is designed to measure the position and momentum of charged particles as well as to determine the vertex position for each collision. It is the detector located closest to the beam. And as such it has to deal with high particle flux and radiation damages. In addition, the tracker has to be lightweight such that particles passing through the tracker will only lose little amount of energy before they reach the calorimeters, where their energies will be measured. Because of these considerations, the CMS tracker is made up of layers of a silicon pixel detector located nearest to the IP, followed by the silicon strip detector. A charged particle

passing through the detectors will produce “hits” in the modules where the sensors register a signal induced by the ionization electrons from the particle.

The determination of the momentum of charged particles can be performed by measuring the bending of the particle trajectories (tracks) in the magnetic field. As illustrated in Figure 22, the momentum of a charged particle in the transverse plane, p_T , can be expressed as $p_T = eB \cdot R$, where B is the magnetic field and R is the bending radius, which is obtained by fitting the hits in each layer of the tracker. The momentum resolution of the measurement is therefore proportional to the spatial resolution of the detector. This is another reason that the CMS tracker is built to have very high granularity.

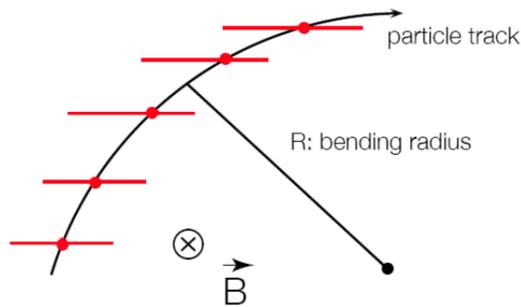


Figure 22: An illustration of a measurement of the momentum of a charged particle.

As mentioned above, the inner part of the CMS tracker is the pixel detector. It had 1440 pixel detector modules distributed in 3 (2) layers in LHC Run I, and it was upgraded in 2016

and 2017 to have 1856 modules in 4 (3) layers in the barrel (endcap) region. The total number of pixels in the upgraded detector is 124 million. The 3 barrel layers are located at radii of 4 to 11 cm from the beam line. Each pixel module has a size of $100 \times 150 \mu\text{m}^2$, and a thickness of $250 \mu\text{m}^2$, providing a position resolution of 10–20 μm , depending on η .

The strip tracker is located just outside of the pixel tracker, at radii of 20 cm to 110 cm. The size of each strip module is 10–25 cm in length and 80–180 μm in width, depending on the distance between the module and the interaction point. There are ~15000 strip modules distributed in 13 layers in the barrel and 14 layers in the endcaps. The position resolution of the strip tracker is 20–40 μm , depending on η .

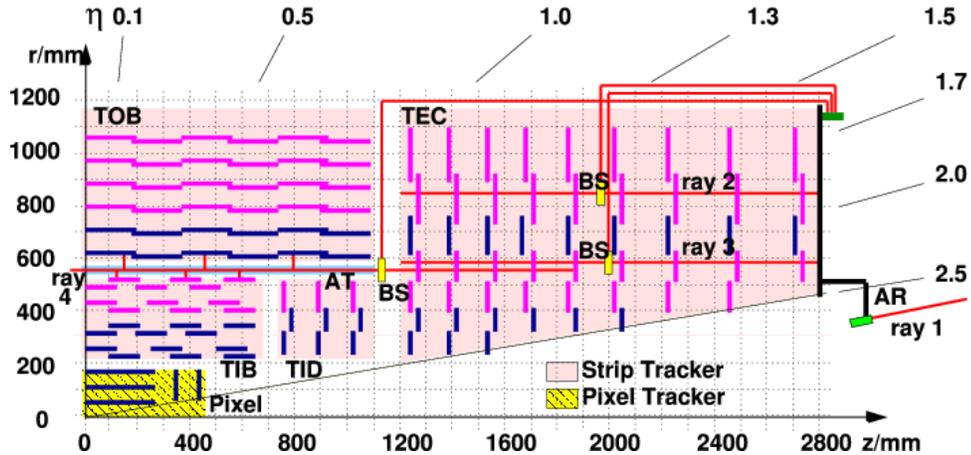


Figure 23: Schematic cross section view of the CMS tracker in the x-y plane and positive z direction in LHC Run I. In 2016 and 2017, the endcap and barrel pixel tracker were upgraded to have 3 and 4 layers, respectively.

Overall, the tracker covers $|\eta| < 2.5$. The layout of the CMS tracker can be seen in Figure 23. The p_T resolution of reconstructed tracks [90] at $|\eta| = 0$ is 1.5% at 10 GeV and 2.0% at 100 GeV [91]. The resolution deteriorates with increasing $|\eta|$. The vertex [90] position resolution is 20 (25) μm for vertices with 30 associated tracks in x or y (z) direction [91].

3.2.2 The Calorimeter Detectors

Calorimeters are destructive detectors that are designed to measure the energy and position of the incident particle by absorbing all the energy of the particle. When the particles interact with the material of the detector, cascades of secondary particles (particle showers) are produced until all the secondary particles are fully stopped and absorbed in the material. To ensure that all the particle energy is deposited within the detector volume, the calorimeters are made of dense materials.

There are two types of particle showers, electromagnetic and hadronic showers. Electromagnetic showers are produced when electrons, positrons, and photons interact with the calorimeter detector. The showers develop through the processes of bremsstrahlung and pair production, since in most materials, bremsstrahlung and pair production are the dominant processes for electrons (and positrons) and photons to lose energy, respectively, with energies above a few tens MeV. The radiation length of a given material is defined as the mean length needed to reduce the energy of a high energy electron by the factor $1/e$ due to bremsstrahlung, or $7/9$ of the mean free path for photon pair production in that material. It characterizes the longitudinal development of the electromagnetic shower. The lateral spread of an electromagnetic shower is governed mainly by the multiple scattering of the low-energy electrons in the shower and is

characterized by Moliere radius, which is defined as the radius of a cylinder which has the same axis as the shower axis and contains on average 90% of the electromagnetic shower's energy.

The development of hadronic showers is initiated by nuclear inelastic collisions of the incoming hadrons with the detector material. Hadronic showers are produced when hadronic particles such as pions, protons, and neutrons interact with the detector. An important property of the detector material is the nuclear interaction length, which is the mean length needed to reduce the energy of a hadron by a fraction of $1/e$. Like the radiation length, it is used to characterize the longitudinal development of a hadronic shower in the material. The lateral spread of hadron showers are caused by large transverse momentum transfers in nuclear interactions and is much larger than the lateral spread of electromagnetic showers.

There are two types of calorimeters, homogeneous calorimeters, in which a single material serves as both the active material (the material which produces the electronic signal) and absorber, and sampling calorimeters, which are made of repeating layers of absorbing and active materials. The shower energy is measured using the active material (e.g., liquid argon, plastic scintillator, etc.) in the detector. The CMS calorimeter uses scintillators as the active material. When ionizing particles pass through scintillator material, they excite the electrons in the atoms of the material and then photons are emitted, typically near the blue end of the visible spectrum, when the atoms return to their ground states. This process is called scintillation. The scintillation photons are then collected by photodetectors, and the number of photons are proportional to the amount of energy deposited by the ionizing particles.

The intrinsic energy resolution of the calorimeters is typically parametrized by:

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{S\sqrt{E}}{E}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \quad (3.5)$$

where the first term is the stochastic term, the second term is the noise term, and the third term is the constant term. As the energy detected is proportional to the number of particles produced in a shower, the stochastic term represents the statistical fluctuations of the number of particles in the shower and is proportional to square root of the energy. The stochastic term for homogeneous calorimeters is in general smaller compared to those for sampling calorimeters since in sampling calorimeters, many of the particles in the shower are produced in the absorbing material and they are not detected. The main contribution to the noise term comes from readout electronic noise. The contribution to the constant term comes from the nonuniformity of the detectors due to imperfections in the detector mechanical structure, radiation damage, detector aging, etc.

The CMS calorimeter consists of an ECAL which is optimized to measure the energy of electrons and photons, and an HCAL, which is optimized to measure the energy of hadrons. The components of the detectors are described below.

3.2.2.1 The Electromagnetic Calorimeter

The ECAL is made up of 61200 lead tungstate crystals (PbWO_4) in the barrel region and 7324 crystals in each of the two endcap disks. The inner surface of the electromagnetic barrel (EB) calorimeter is located 129 cm away from the beam line, and the electromagnetic endcap

(EE) calorimeter is located 315 cm away from the nominal interaction point. It is a homogeneous detector, where each of the PbWO_4 crystals serve as both absorber and scintillator. A photodetector is attached to the back of each crystal to collect scintillator light.

The PbWO_4 crystals have a radiation length of 0.89 cm, Moliere radius of 2.2 cm, and nuclear interaction length of 19.5 cm. Each of the crystals has a length of ~ 23 cm, equivalent to ~ 25 radiation lengths. This is sufficient to contain 98% of the depth of electromagnetic showers created by electrons of 1 TeV energy. The crystals have cross-sectional areas of 2.2×2.2 and $2.9 \times 2.9 \text{ cm}^2$ in EB and EE, respectively. These cross-sectional areas correspond to a granularity in $\eta - \phi$ plane of $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$.

To discriminate isolated photons from high energy neutral pions, which often decay into two photons in close spatial proximity and can mimic an isolated photon signal in the ECAL, a high granularity endcap preshower (ES) detector made of lead and silicon sensors is installed in front of the EE.

Overall, the CMS ECAL covers $|\eta| < 3.0$. A schematic view of the CMS ECAL can be seen in Figure 24. Because the CMS ECAL is a homogeneous calorimeter, it has good energy resolution. In a test beam study, S , N , and C in Equation 3.5 for a 3×3 crystals array in EB were measured to be 2.8%, 12%, and 0.3%, respectively, for electrons with energy less than 500 GeV [89]. In a study using the $Z \rightarrow e^+e^-$ events at $\sqrt{s} = 13$ TeV, the single electron energy resolution is found to be between 1 and 5 %, depending on η and the energy [92].

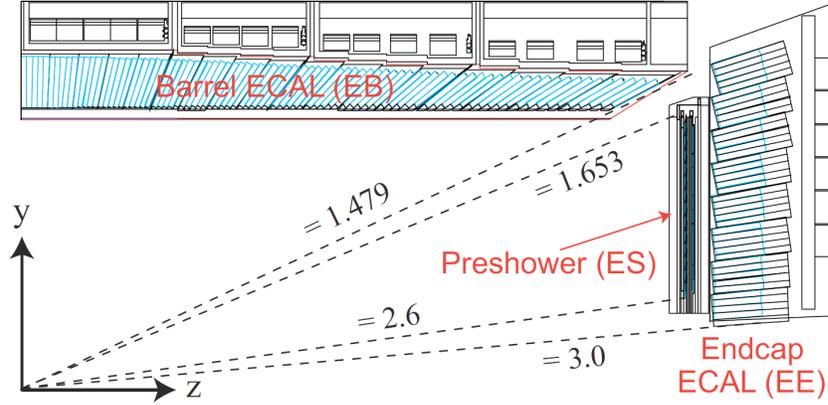


Figure 24: Schematic cross-sectional view of the CMS ECAL in the x-y plane and positive z direction.

3.2.2.2 The Hadronic Calorimeter

The HCAL is important for the identification of jets and the measurement of jet energies and event E_T^{miss} . For most materials, the nuclear interaction length is much larger than the radiation length. Therefore, to keep the construction cost reasonable, HCALs are usually built as sampling calorimeters. The CMS HCAL consists of 4 sub-detectors, the hadronic barrel (HB) calorimeter, hadronic endcap (HE) calorimeter, hadronic outer (HO) calorimeter, and hadronic forward (HF) calorimeter. The layout of the HCAL in CMS detector can be seen in Figure 25.

The HB is located just outside of the ECAL and inside the magnet with inner radius 177 cm from the beamline and outer radius 295 cm and it covers $|\eta| < 1.3$. The two HE disks, placed 4 m away from the interaction point, cover $1.3 < |\eta| < 3.0$. The HB and HE use brass as absorber and plastic tiles as scintillator. Brass has a radiation length of 1.5 cm and interaction length

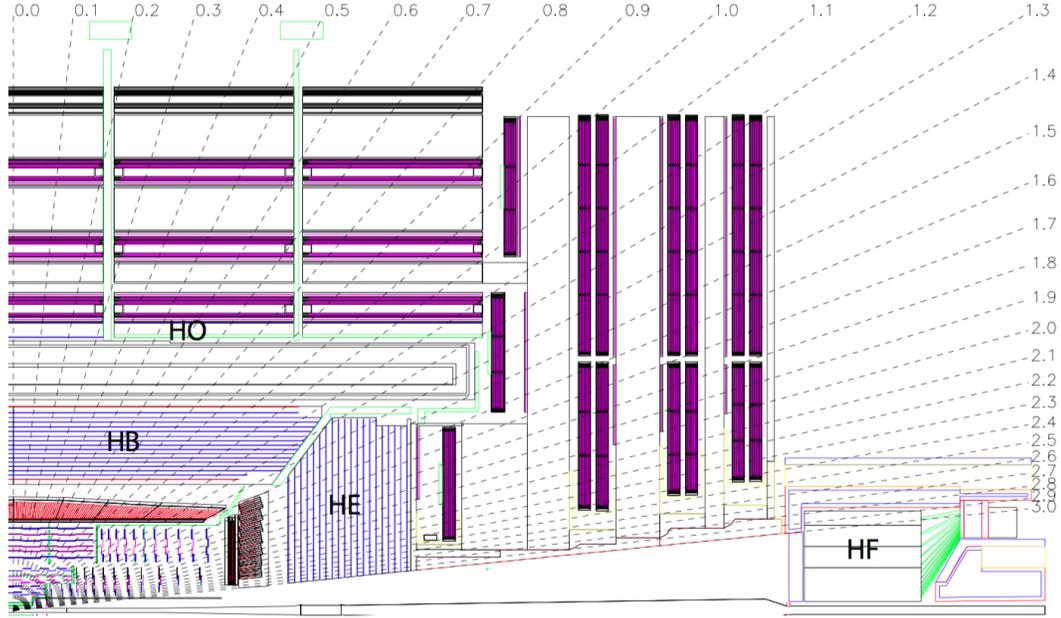


Figure 25: Schematic of the CMS Detector in the x-y plane and positive z direction, illustrating the layout of the HCAL detectors.

of 16.4 cm. The total absorber thickness is 5.6 interaction lengths at small- $|\eta|$ and increases to 10.3 with increasing $|\eta|$. The granularities for HB and HE are about $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in $|\eta| < 1.6$ and $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$ in $1.6 < |\eta| < 3.0$. Each segment in HCAL is aligned with 5×5 crystals in ECAL.

The HO is placed outside of the magnet. It is built to detect the energy of showers that leak out of the back of the HB. The HO uses the magnet as the absorbing medium and plastic tiles as scintillator, extending the total absorber thickness in the barrel region to 10 interaction lengths.

The HF calorimeter extends the η coverage of the CMS calorimeters to $3.0 < |\eta| < 5.0$. It uses steel as absorber and quartz fibers as the sensitive material. When high-energy charged particles penetrate the HF, Cherenkov light is emitted in the quartz fiber and then collected by photodetectors that create electronic signals. Two different lengths of quartz fibers are used in the detector, long fibers with length 165 cm, or approximately 10 interaction lengths, and short fibers with length 143 cm and start at a depth of 22 cm from the front of the detector. The two fiber lengths provide the ability to distinguish electromagnetic and hadronic showers: electromagnetic showers typically deposit most of their energy in the long fibers, while hadronic showers deposit their energy in both the long and short fibers. The HF calorimeter has a granularity of $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$, similar to the granularity of the HE.

The hadronic energy resolution can be parametrized in a similar fashion as the electromagnetic energy resolution. In a test beam study which contained both HCAL and ECAL modules [93], S and N in Equation 3.5 for the hadronic energy resolution are 85% (198%) and 7% (9%) in the barrel (endcaps), respectively. The noise term is negligible compared to the stochastic and constant terms.

3.2.3 The Muon Chambers

Muons are elementary particles that are very similar to electrons but have a mass that is 207 times that of the electron. Because of its heavy mass, muon bremsstrahlung is heavily suppressed compared to electrons bremsstrahlung, and therefore muons mainly interact with the detector via ionization and can penetrate the CMS calorimeters with very little energy loss.

The muon detectors can therefore be placed outside of the calorimeters, and with the 1.8 T field in the return yolk of the magnet, the muon momenta can be measured.

The muon detector consists of 1400 gaseous chambers: 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) track the particles' positions in barrel and endcap, respectively. Resistive plate chambers (RPCs) are installed in between DT and CSC layers to provide precision timing measurements. The muon chambers cover $|\eta| < 2.4$ and are located within the return yolk. The overall layout of the CMS muon chambers can be seen in Figure 26.

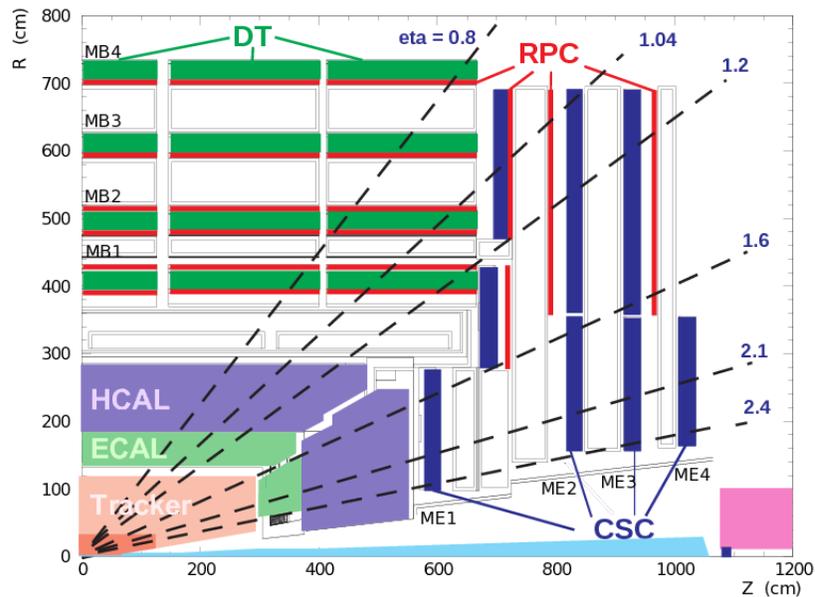


Figure 26: Layout of the CMS muons chambers in the x-y plane and positive z direction.

The algorithm used to measure the muon momentum in the muon chambers is similar to the algorithm used for the tracker. The best muon momentum resolution is obtained by combining the tracks in muon chambers and tracker.

CHAPTER 4

THE CMS TRIGGER SYSTEM

At design luminosity, the proton-proton collision rate at the LHC exceeds 1 GHz. As each recorded event has a size of ~ 1 MB, it is not feasible to save every event for later processing. However, only a small fraction of these events are of interest to the CMS physics program, and therefore, the experiment employs a sophisticated trigger system that identifies the interesting events for later processing at runtime and thereby reduces the rate of events written to tape to a manageable level. As indicated in Figure 27, the event rate for interesting physics processes (e.g., production of Higgs bosons, high- p_T jets, etc.) are orders of magnitudes lower than the total event rate at the LHC.

The CMS trigger is a two-tiered system, with a Level-1 Trigger (L1T) and a high-level trigger (HLT) [94]. A schematic view of the CMS trigger system is shown in Figure 28. During data taking, the CMS sub-detectors read out collision events every 25 ns. The signals in the sub-detectors are digitized in the front-end electronics, and the data are temporarily stored in the readout buffers. At the same time, a copy of the data from the calorimeters and muon detectors is sent to the L1T, and the L1T runs its selection algorithms on the data to decide if the event worth saving or not. If so, the event, including the data from the tracker, is sent to the HLT via a switch network system for further processing. Events that pass the HLT selection algorithms are written to the CMS storage disks for offline physics analysis. Events which don't pass the L1T or HLT selections are dumped from the buffers immediately.

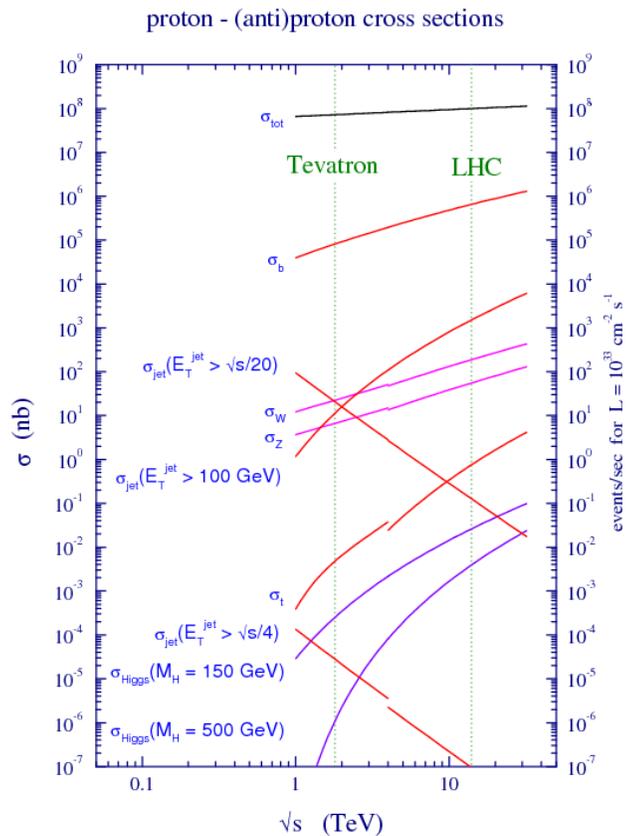


Figure 27: The cross section and event rate for various physical processes at different collision energies.

The L1T is a hardware system based on customized electronics with a fixed latency of $3.2 \mu\text{s}$. It has to decide if an event is interesting or not within this latency. The latency is limited by the size of the readout buffer. The L1T reduces the data rate from the total collision rate of 1 GHz to about 100 kHz.

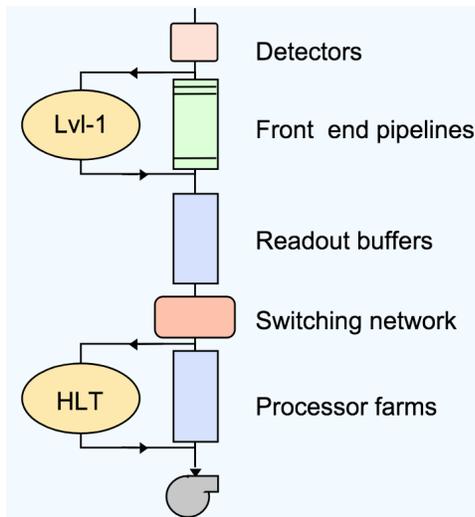


Figure 28: Schematic view of the two-tiered CMS trigger system.

In 2015 and 2016, CMS performed an upgrade on the L1T [95] in order to accommodate the high luminosity environment expected by LHC in Run II. The upgrade was done to maintain the acceptance rate of interesting events below the 100 kHz rate limit, and it was completed in two stages, called Stage-1 and Stage-2 in early 2015 and early 2016, respectively. The Stage-1 upgrade only updated the Calorimeter Trigger, and was used only for data taking in 2015. The Stage-2 upgrade updated all the trigger sub-systems, including the Stage-1 Calorimeter Trigger, and was used for data taking from 2016 onward.

The HLT hardware is made of a processor farm composed of $\sim 20,000$ Central Processing Unit (CPU) cores. It has a latency of ~ 100 ms and further reduces the event rate from 100 kHz to ~ 1 kHz. The HLT latency and rate are limited by the offline mass storage and computing resources.

Detailed descriptions of the Stage-2 upgraded L1T and the HLT systems are presented in the following sections. The Stage-1 system will not be discussed as it is not used in this analysis. A brief discussion of the CMS offline computing system will also be provided at the end of this chapter.

4.1 The Level-1 Trigger

At the first step in the trigger processing chain, trigger primitives (TPs) are generated from the energy deposits in the ECAL and HCAL and from the hit patterns in the muon system. The TPs are processed to form trigger objects (jets, e/γ 's, etc.), which are then sent to the Global Trigger (GT) and a decision is made on whether to accept the event or not. A schematic view of the upgraded L1T is in Figure 29.

The upgraded L1T hardware is based on the μ TCA (Micro Telecommunication Computing Architecture) telecommunication standard. The trigger boards include the Calorimeter Trigger Processor-7 (CPT7), Master Processor-7 (MP7), and Muon Track Finder-7 (MTF7) boards featured with Xilinx Virtex-7 Field-Programmable Gate Arrays (FPGAs). These FPGAs allow the implementation of sophisticated reconstruction algorithms to improve the position and energy resolutions for the reconstructed physics objects in L1T. The reconstruction algorithms lead to significant reductions in the trigger rates in Run II. More details of the L1T sub-systems and the L1T reconstruction algorithms are discussed in the following sub-sections.

4.1.1 The Level-1 Calorimeter Trigger

The L1 Calorimeter Trigger uses the ECAL and HCAL TPs to reconstruct jets, e/γ objects, τ leptons, and energy sums. The ECAL TPs are formed from the transverse energy detected

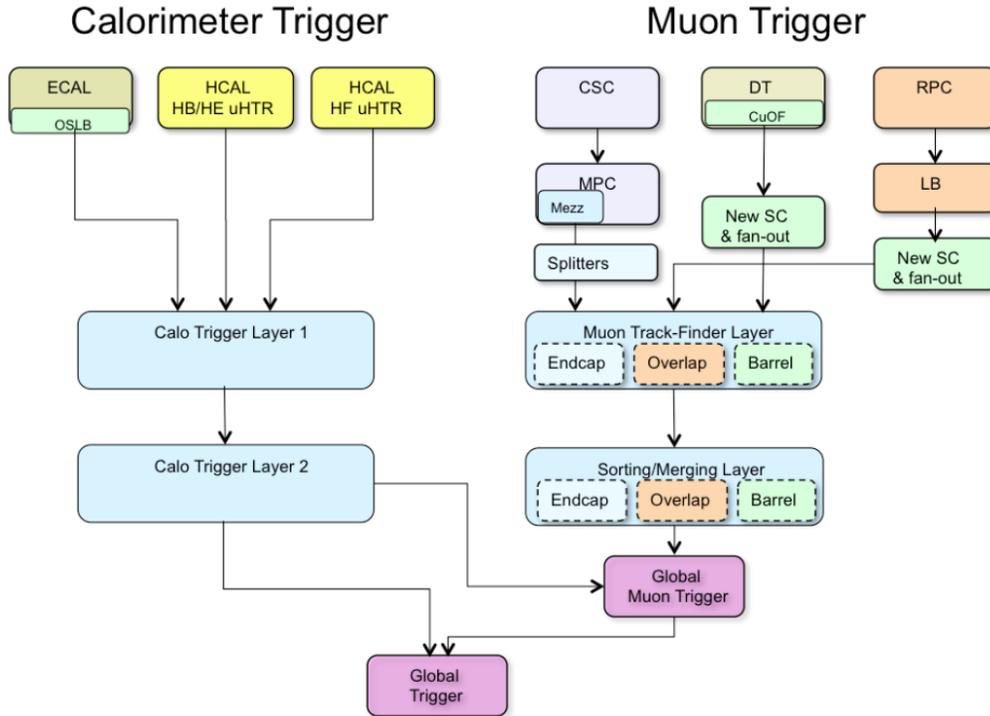


Figure 29: Schematic view of the L1T.

in 5×5 blocks of crystals to form a trigger tower (TT) and converted to optical format by the optical Serial Link Board (oSLB) at the ECAL front-end. Similarly, the transverse energy detected in the HCAL segments forms the HCAL TPs. The HCAL TPs are received by the L1 Calorimeter Trigger on fibers at the Micro HCAL Trigger and Readout (μ HTR) boards. Each HCAL segment corresponds to a TT. The ECAL and HCAL TTs have a η - ϕ coverage of 0.087×0.087 in the barrel region ($|\eta| < 1.5$) and larger in the endcaps.

The L1 Calorimeter Trigger has a time-multiplexing architecture as illustrated in Figure 30. The first layer, Layer-1, consists of 18 CTP7 boards, performs pre-processing and data formatting. Each pre-processing board requires only a regional view of the detector. The outputs of the Layer-1 corresponding to one event are sent to a single processing node, an MP7 board, in the second layer, Layer-2. The Layer-2 consists of 9 MP7 boards. Each MP7 board has access to information at trigger tower granularity. It finds particle candidates and computes the transverse energy (E_T) for each of the candidates, as well as the global energy sums and E_T^{miss} . The outputs of the Layer-2 are sent to a demultiplexer board, an MP7, which formats the data for the muon trigger and GT. As mentioned above, CPT7 and MP7 boards are μ TCA modules featuring Xilinx Virtex-7 FPGAs.

4.1.1.1 The Level-1 Trigger e/γ and τ Reconstruction Algorithms

The e/γ reconstruction algorithm is based on clusters of TTs as illustrated in Figure 31. The clusters are seeded by a single TT that has a local maximum of ECAL transverse energy above a predetermined threshold (the red TT in Figure 31). The local maximum is found in an 11 TT region within a 3×5 TT window in the $\eta \times \phi$ directions, excluding the four TTs in the corners of the window. The first and second neighboring towers (the orange and yellow TTs in Figure 31) surrounding the seed are grouped to form a cluster. To reduce the impact of the energy deposits from pileup interactions, the maximum size of each cluster is limited to 8 TTs within a 2×5 window in the $\eta \times \phi$ directions. The motivation of restricting the cluster within the 2×5 window is that electromagnetic showers do not spread more than 2 TTs in the η -direction, while they can spread up to 5 TTs in the ϕ -direction due to the magnetic field.

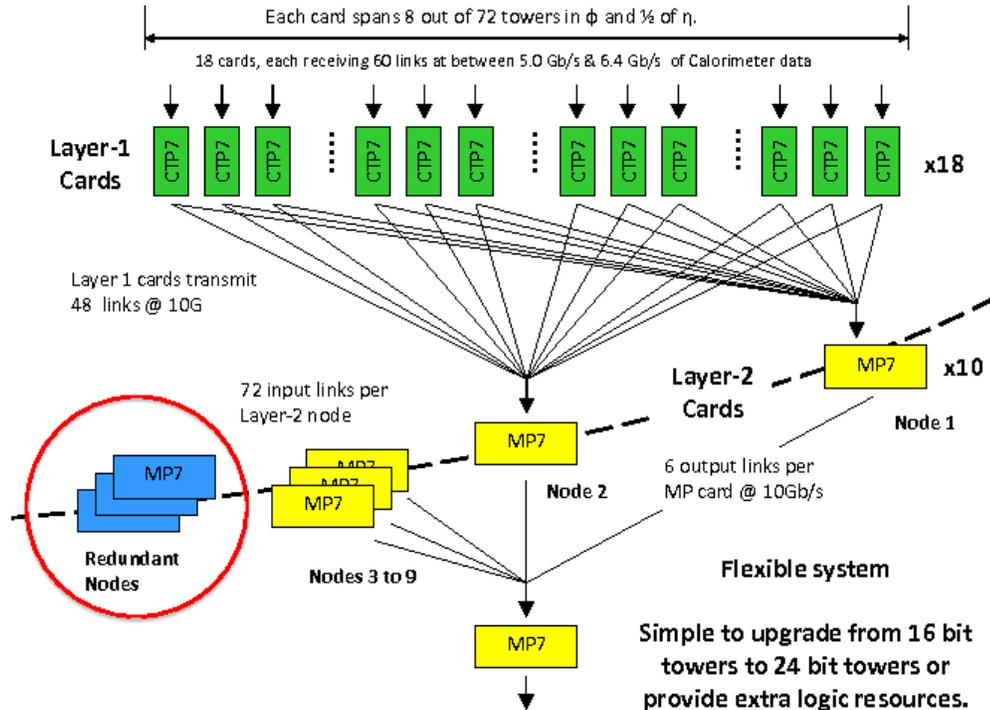


Figure 30: The L1 Calorimeter Trigger layout.

Additional background rejection is achieved by applying vetos to the shape of the clusters. The energy of each reconstructed cluster is calculated as the sum of all the energies in the cluster. The cluster position is determined from the energy-weighted average center of the TTs that make up the cluster. The e/γ candidates are classified as either isolated or non-isolated.

The isolation of the e/γ candidates is based on the energy within a 5×9 TT region in $\eta \times \phi$ directions, around the seed in both ECAL and HCAL, excluding the TTs that contain the footprint of the candidate. If the energy sum in the region is below a given threshold, the

candidate is defined as isolated. The threshold is a function of η and the number of pileup interactions and is determined from studies of zero bias data, which contain events selected by a zero bias trigger that has no selection criteria imposed on the trigger objects.

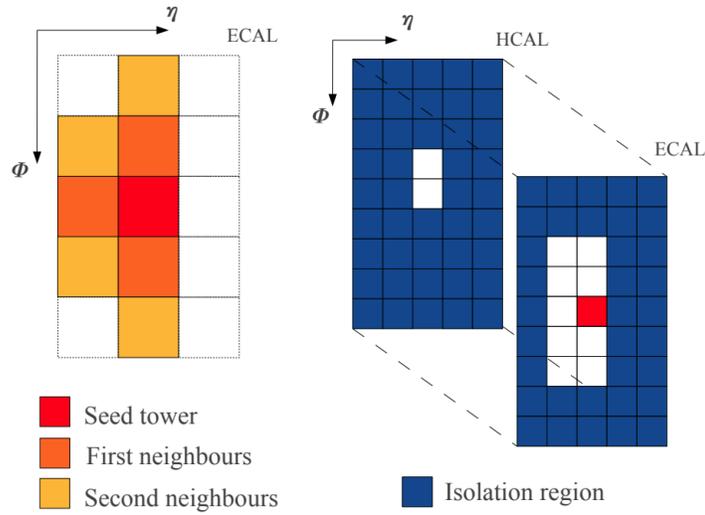


Figure 31: An illustration of the L1T electron and photon reconstruction algorithm. A candidate is formed by clustering a seed TT (red) with neighbor TTs (orange and yellow). The isolation criterion is based on the sum of the energies in the isolation region (blue) in ECAL and HCAL. The isolation region is formed as a 5×9 TT region around the seed. The TTs that contain the footprint of the candidate (white) are excluded.

Hadronically decaying τ leptons can be identified and reconstructed efficiently by the L1T. The identification and reconstruction algorithm for τ leptons is similar to the algorithms used for the e/γ candidates. Two neighboring clusters in the ϕ -direction are allowed to merge to

reconstruct multi-prong decaying τ leptons. The decay products of the τ leptons produce clusters that have different shapes to those for electrons and photons, allowing τ leptons being identified. The isolation criterion is also developed in the same manner but larger isolation region as those for the e/γ candidates.

4.1.1.2 The Level-1 Trigger Jet and Energy Sums Reconstruction Algorithms

The jet reconstruction algorithm is based on a 9×9 TT sliding window algorithm centered on a local maximum as illustrated in Figure 32. The size of the windows is chosen so that the jet size roughly matches the size from the offline jet reconstruction algorithm. To avoid double-counting of the jets, in cases where two towers in the region have the same energy, the energy of neighboring TTs must satisfy the inequalities shown in Figure 32. The jet energy is calculated as the sum of the energies in all the TTs in the 9×9 TT window. The position of the jet is assigned to the center of the window. Additional energy due to pileup is removed using the “chunky donut” technique. In this technique, the energy of each of the four 3×9 TT regions surrounding the jet window is calculated, as illustrated in Figure 32, and the total energy constructed from the energies of the three regions with the lowest 3×9 TT energies is calculated and subtracted from the jet energy. The energy from the most energetic 3×9 TT region is not used to prevent overestimating the pileup energy in cases where there is another high-energy jet nearby.

The jet reconstruction algorithm is implemented to have performance as close to those used in offline as possible. A comparison of the E_T 's reconstructed from L1T and offline in various



Figure 32: An illustration of the L1T jet reconstruction algorithm. A jet is reconstructed based on a 9×9 TT sliding window (blue) algorithm centered on a local maximum (green). The pileup energies are mitigated by subtracting the energy sums in the three of the four 3×9 PileUp Subtraction (PUS) regions (magenta) that have the lowest energies. To avoid double-counting of the jets, in cases where two towers in the region have the same energy, the energy of neighboring TTs must satisfy the inequalities shown in the veto (purple) and mask (blue) regions.

pileup conditions are shown in Figure 33. As can be seen in the figure, the E_T of the jets reconstructed in the L1T agrees well with the E_T of the jets reconstructed in offline.

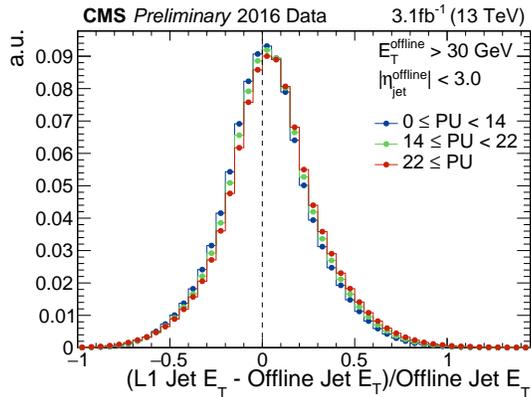


Figure 33: A comparison of the jets E_T from L1T and offline reconstructions in a various of pileup conditions (left).

Three types of energy sum algorithms are computed at the L1 Calorimeter Trigger: H_T , the scalar sum of jet E_T 's for jets with $|\eta| < 3.0$ and $E_T > 30$ GeV; E_T , the scalar sum of the trigger towers with $|\eta| < 5.0$; and E_T^{miss} , the vector sum of the trigger towers with $|\eta| < 3.0$.

4.1.2 The Level-1 Muon Trigger

Similar to the calorimeter trigger, the upgraded L1 muon trigger consists of two layers: the muon track finding layer and the Micro Global Muon Trigger (μ GMT). The muon track finding layer is partitioned into three components which are dedicated to finding muons in the barrel, endcap, and overlap regions. The track finders receive TPs from the muon detectors and assign

p_T values to the muon tracks. The track finders then send the muon candidates to the μ GMT which removes duplicates across the track finder boundaries and ranks the muon candidates according to criteria based upon p_T and quality before sending them to the GT.

The barrel muon track finder (BMTF) covers $|\eta| < 0.8$. It receives TPs from the DT and RPC. The endcap muon track finder (EMTF) receives TPs from CSC and RPC and covers $1.2 < |\eta| < 2.5$. The overlap muon track finder (OMTF) receives TPs from all the three muon detectors and covers $0.8 < |\eta| < 1.2$. The processor boards used for track finders are MP7s for BMTF and MTF7s for OMTF and EMTF. All the processing boards are connected to the detector front-end electronics via optical links. The μ GMT uses CPT7 boards as the processing boards.

The track finders create muon candidates by searching for track segments in adjacent muon detector modules compatible with a muon originating from the interaction point. Quality values based upon the number of track segments found are assigned to each of the muon candidates. The p_T values of the muon candidates are assigned based on the output from a Boosted Decision Tree (BDT). The BDT takes the kinematic information of the muon candidates and estimates the muon transverse momenta using regression analysis. The BDT is trained using Monte Carlo (MC) events and implemented via the lookup tables in the MTF7 boards.

The μ GMT receives muon candidates from the track finders, as well as values of energies for each 2×2 and 5×5 TT regions from the demultiplexer board of the Layer-2 Calorimeter Trigger. These values are used to determine the muon isolations.

4.1.3 The Level-1 Global Trigger

The upgraded L1 GT, which is called the Micro GT (μ GT), is the final step in the L1 Trigger decision chain. The μ GT hardware consists of two MP7 boards. The μ GT takes as inputs the energy and position information of the physics objects reconstructed in the calorimeter and muon triggers and makes the final decision on whether or not to send the event to the HLT based on the results from a set of trigger selection algorithms. The trigger algorithms include single-object and multi-object algorithms. An example of a single-object algorithm is SingleJet450 which requires a jet with p_T greater than 450 GeV. An example of a multi-object algorithm is EG30_Mu45 which requires an e/γ candidate with p_T greater than 30 GeV and a muon with p_T greater than 45 GeV. The η , ϕ , isolation, and quality information of the physics objects can also be used in the algorithms. More complicated selection algorithms that are based on the η and ϕ spatial separation between two objects, like invariant mass and/or transverse mass of two objects, can also be constructed. The μ GT can run a maximum number of 512 trigger algorithms for an event.

The μ GT issues a logical signal called the L1 Accept (L1A), which is set to true if any of the 512 algorithms run by the μ GT is satisfied. The L1A is sent to the Trigger Control and Distribution System (TCDS). The TCDS receives control signals from all CMS sub-detectors. If the L1A is satisfied and all the sub-detectors are ready to take data, the TCDS issues the signal for Data Acquisition (DAQ) to read out the data. The TCDS is also responsible for the synchronization of the L1T. It delivers the LHC timing reference and L1T decisions with the corresponding bunch and event numbers.

The μ GT also runs a number of selection algorithms that have low thresholds on the trigger objects or have minimal selection criteria imposed on the trigger objects. These are needed to study efficiencies and possible biases in the selection algorithms. The μ GT controls the rate from these algorithms through a mechanism called prescaling. If a trigger has a prescale N , 1 of N events that pass the trigger selection will be marked as accepted. The L1 trigger algorithms and prescales are proposed by the various physics analysis groups in the experiment and integrated into a trigger menu. The trigger menu contains the trigger algorithms and sets of prescale values that are optimized for various instantaneous luminosity conditions.

4.1.4 Data Quality Monitoring of the Level-1 Trigger

To certify the quality of data recorded by the CMS detector including the L1T, a Data Quality Monitoring (DQM) system is implemented by the CMS collaboration. The system provides visualizations of the recorded data in the form of histograms as the collisions are occurring in real-time for shifters to monitor. The visualization is achieved using an interactive website: the DQM Graphical User Interface (GUI). Besides the online monitoring on the incoming data, the DQM system is also used to provide more detailed data validation offline and to validate new releases of the CMS software.

The online DQM of the L1T provides a live monitoring on the inputs and outputs of each layer of the trigger processing chain, as well as a live comparison between the data and the software emulation of the trigger algorithms. Some examples of the L1T DQM histograms are shown in Figure 34, which shows the ECAL and HCAL TP occupancies in the Layer-1 Calorimeter Trigger, and in Figure 35, which shows the E_T and H_T distributions in the Layer-2

Calorimeter Trigger for central jets (i.e., jets with $|\eta| < 3.0$). Dead (white) and hot (red) spots in Figure 34 or non-smoothness in the distributions plotted in Figure 35 indicate potential problems with the incoming data.

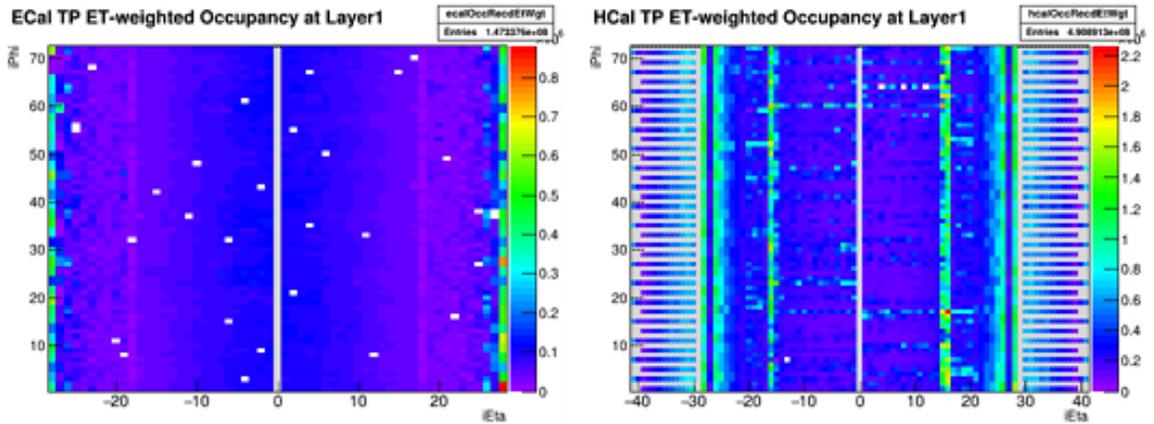


Figure 34: The monitoring of ECAL (left) and HCAL (right) TPs in the Layer-1 Calorimeter Trigger.

An important quantity to monitor is the comparison between the output from the L1T firmware and the output from the L1T emulator. The L1T emulator is a detailed simulation of the firmware algorithms and, in most cases, is expected to give perfect bitwise agreement with the output from the firmware for all component of the L1T. As an example, the implementation of the data and emulation comparisons for the Layer-1 and Layer-2 Calorimeter Triggers are presented. In the Layer-1 comparison, the output data from the Layer-1 Calorimeter Trigger is

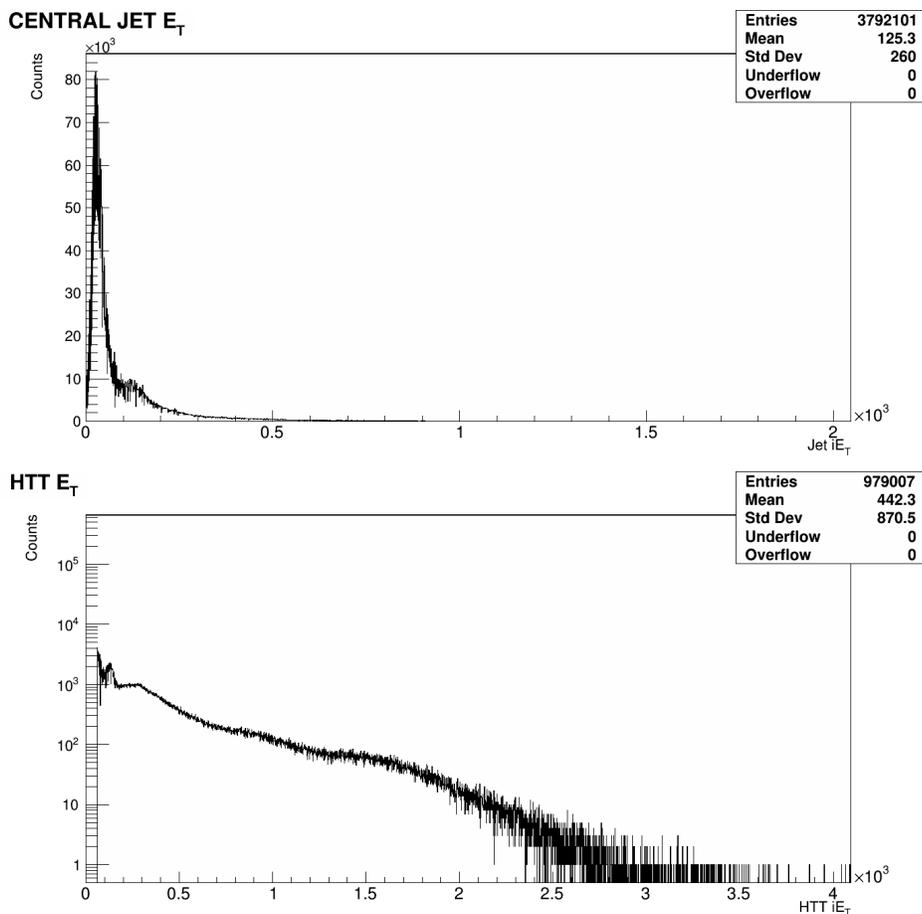


Figure 35: The monitoring of jets E_T (top) and H_T (bottom) distributions in the Layer-2 Calorimeter Trigger.

compared to the output data from the Layer-1 emulation which takes data from the ECAL and HCAL TPs as input. In the Layer-2 comparison, the output data from the Layer-2 Calorimeter Trigger is compared to the Layer-2 emulation which takes the output data from the Layer-1 as input. The results of the comparisons are displayed in DQM histograms. The summary histogram for the comparisons between the E_T of the Layer-1 TTs, the ratio of the ECAL and

HCAL energies in the Layer-1 TTs, and the number of active Layer-1 TTs in emulator and firmware are shown in Figure 36. As can be seen in the figure, these comparisons reached perfect agreement. At Layer-2, the ratio of the kinematic distributions for the L1 reconstructed objects in data and emulation are also provided. For example, the ratio of central jet E_T and event H_T distributions in data and emulation are shown in Figure 37. Perfect agreement is found between central jet E_T and H_T distributions in data and emulation.

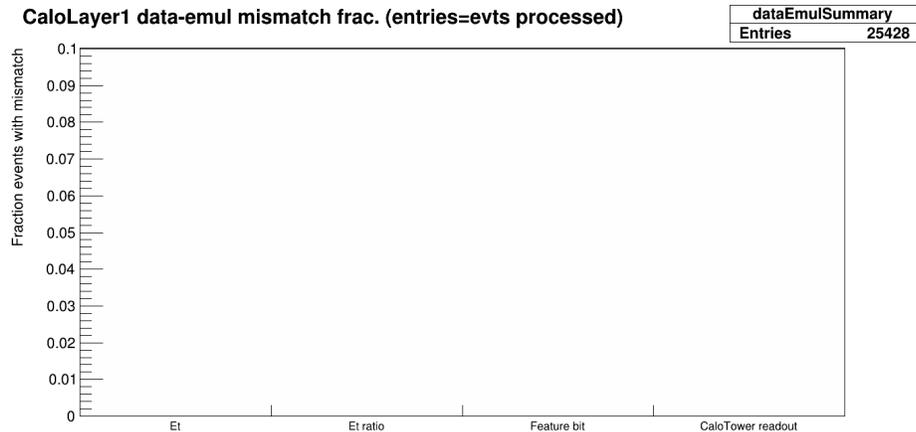


Figure 36: The summary of bit-by-bit data and emulation comparisons for the E_T of the Layer-1 TTs, the ratio of the ECAL and HCAL energies in the Layer-1 TTs, and the number of active Layer-1 TTs. The bins in the histogram are all empty indicating perfect agreement between data and emulator.

Automatic alarm functionality is also implemented for the L1T DQM. The DQM GUI will raise the alarm when histograms fail quality tests. For example, if 1% of the jets fail the data

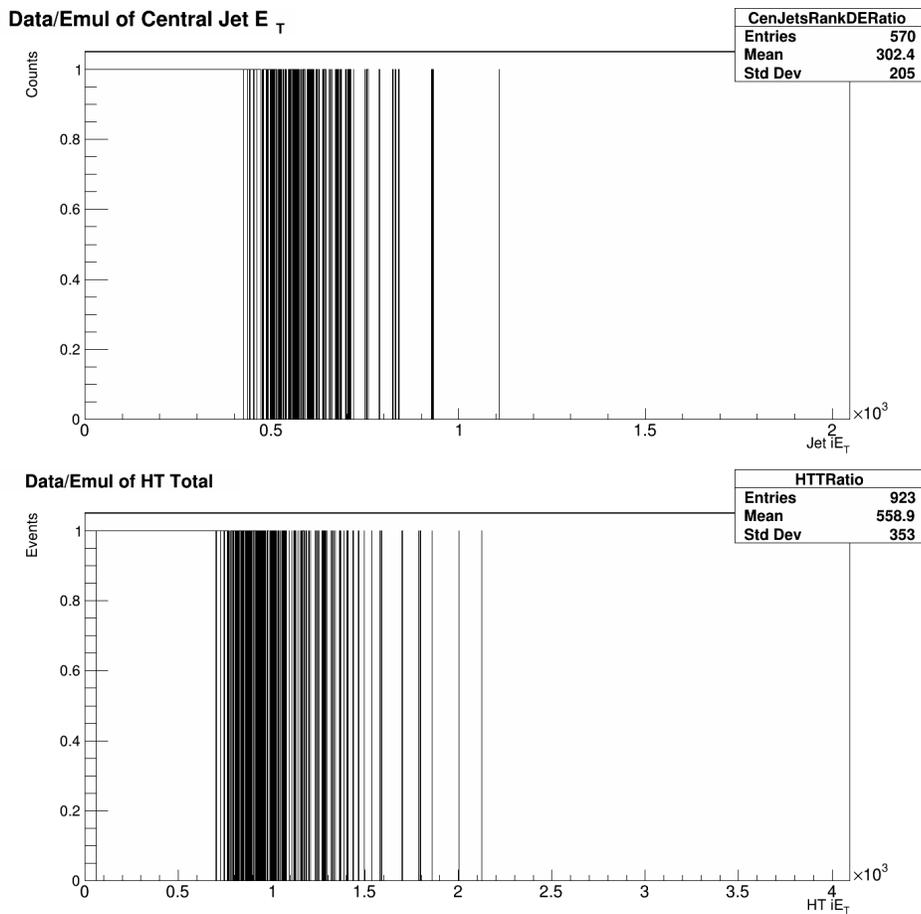


Figure 37: The ratio of jets E_T (top) and H_T (bottom) distributions from data and the Layer-2 Calorimeter Trigger emulator.

and emulation comparison or 5% of the ECAL TPs channels are empty, the shifters at the CMS control room will be alerted.

The L1T offline DQM and release validation systems produce the same histograms as the online monitoring but run on prompt reconstructed events and simulated events at the CMS offline computing centers (see Section 4.3).

4.2 The High-Level Trigger

The HLT has access to the full detector information (raw data) and processes up to 100 kHz of events, as selected by the L1 system. Each event is sent to a single CPU and the decision to keep the event is made by running a series of selection algorithms, called paths, which are independent of each other but contain common sequences that are shared by various algorithms. An example of an HLT path is illustrated in Figure 38. Each HLT path is seeded by one or more L1 algorithms. The events are processed through a series of producers, which reconstruct the candidates, and filters which run the selection algorithms. At the start of the series, the selection algorithm typically uses only information from the calorimeters or muon systems to quickly filter out uninteresting events before passing them through to the selection algorithm that use the tracking system. The HLT runs a version of the Particle Flow (PF) algorithm that is optimized for speed, to reconstruct particles. The PF algorithm is the reconstruction algorithm used in the offline analyses, which will be discussed in detail in the next chapter. The reconstructed particles include electrons, photons, muons, tau leptons, jets, E_T^{miss} , H_T as well as b-tagged jets.

The selected events are sent to different data streams, which are organized in categories such as Physics, Scouting, DQM, etc. The Physics stream contains events used for most of the physics analyses and full event information is stored. The Scouting stream contains events selected with lower thresholds than the events in the physics stream and are used for physics analyses that only need the HLT-reconstructed information. The events in the physics stream

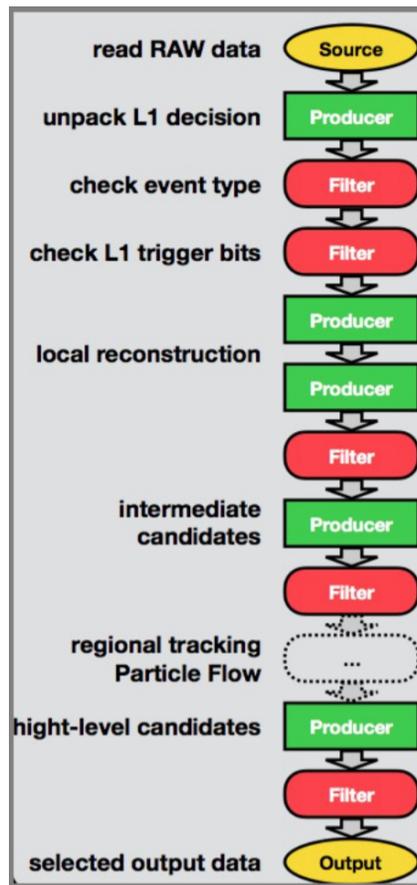


Figure 38: An illustration of an HLT path.

are further split into different datasets, depending on the trigger algorithms. This analysis uses the JetHT dataset, which contains events triggered by the jet and H_T triggers.

Similarly to the L1T, the HLT paths and prescales are proposed by the physics analysis groups and integrated into the HLT menu. The trigger rate of the HLT menu, as well as the L1T menu, need to be estimated to make sure the rate do not exceed the allocated limit before the menus are deployed online for data taking. The rate estimation can be done using either data or simulated events. More details regarding the rate estimate can be found in Appendix A.

4.3 The Offline Computing Systems

The CMS computing infrastructure consists of a series of computing centers located at various locations throughout the world. The central computing center is located at CERN and is called the Tier-0. In addition, there are 8 large computing centers distributed throughout the world, called Tier-1's, and more than 50 smaller scale Tier-2 and Tier-3 centers worldwide. The Tier-0 receives the events accepted by the HLT in RAW format, which contains just the low-level signals from the various detectors, and performs prompt event reconstruction (RECO). The RECO dataset contains the 4-vector information for all the reconstructed particles, as well as other useful information such as reconstructed vertices and collections of the L1T and HLT output. The RAW and RECO datasets are transferred to the various Tier-1 centers for permanent storage. Fractions of the data in Tier-1 centers are sent to the Tier-2 centers so that analyzers from all over the world can access them easily. In addition, CMS makes Analysis Object Data (AOD) dataset, a slimmed version of the RECO dataset, to reduce computing time for the offline physics analyses. The AOD dataset contains only the most important RECO

information for the offline physics analyses. The Tier-2 and Tier-3 centers are also responsible for the production and storage of the simulation events.

The CMS software framework is maintained in a package called CMSSW. The CMSSW software is written in C++ and provides all the reconstruction, simulation, calibration, and data analysis software tools for CMS. Many of these software tools use software provided by the ROOT [96] data analysis framework. In CMSSW, the independent C++ plugin modules are executed using instructions and parameters provided by configuration files written in Python.

CHAPTER 5

JET RECONSTRUCTION AND CALIBRATION

As described in Chapter 2, jets are the experimental signature of the outgoing quarks and gluons in high energy collisions. At CMS, jets are clustered using collinear- and infrared-safe sequential recombination algorithms (see Section 2.3). The input to these algorithms can be taken from a variety of different sources. For example, jets can be reconstructed using only the information from the calorimeters or the trackers. These jets are called calorimeter jets (CaloJets) or track jets (TrackJets), respectively. Jets can also be reconstructed using the combined information from both calorimeters and trackers, and such jets are called jet-plus-tracks jets. In this analysis, jets are reconstructed based on the output of the Particle Flow (PF) technique [97]. The following sections describe the PF reconstruction technique, the PF jet reconstruction algorithm, and the jet energy calibration procedure.

5.1 The Particle Flow Technique

To optimize the jet reconstruction performance, CMS employs the PF technique [97], which combines information from all the subdetectors to identify and reconstruct each final-state particle in the event. The reconstructed particles are called the PF candidates and include photons, electrons, muons, and charged and neutral hadrons. The PF technique takes tracks and calorimeter clusters as inputs. These input elements are then linked together to form objects called PF blocks and by analyzing each PF block, the PF candidates are identified and

reconstructed. An example showing how the PF algorithm reconstructs particles is illustrated in Figure 39, which depicts a simulated jet containing 5 particles: a π^+ , a π^- , two photons (from a π^0 decay), and a K_L^0 . Event displays for the ECAL and HCAL are shown on the left and right of the figure, respectively. Also shown in the figure are green lines (dots) which represent the reconstructed tracks (and their associated hits) for the charged pions. The calorimeter cells are indicated by the squares in the figure. The energy deposits within the cells are displayed in grey with the size of the squares proportional to the energy in the cell. To reconstruct these 5 particles, the PF algorithm proceeds as follows. First, 2 tracks T_1 and T_2 , 4 ECAL clusters E_1, E_2, E_3 , and E_4 , and 2 HCAL clusters H_1 and H_2 are reconstructed. Then, links between the tracks and calorimeters are established to form PF blocks. T_1 is linked to E_1 and H_1 . T_2 is linked to H_2 . E_1 is also linked to H_1 . The other 3 clusters in ECAL are not linked to anything. Finally, each PF block is analyzed to reconstruct the particles. The PF block which contains T_1 , E_1 , and H_1 is reconstructed as the π^- . The block which contains T_2 and H_2 is reconstructed as the π^+ . E_2 and E_3 are close to each other and are therefore reconstructed as two photons from the π^0 decay. And E_4 is reconstructed as the K_L^0 . The PF technique provides a global event description, and the reconstructed PF candidates are then used to reconstruct higher-level objects such as jets, hadronic decaying τ leptons, and missing transverse momenta.

The CMS detector is well suited for PF reconstruction. The fine granularity of the tracker and calorimeter systems provide the resolution necessary to link the tracks to the calorimeter objects, as well as the ability to efficiently separate overlapping showers from closely spaced particles. In the following sub-sections, more details about the reconstruction of the input

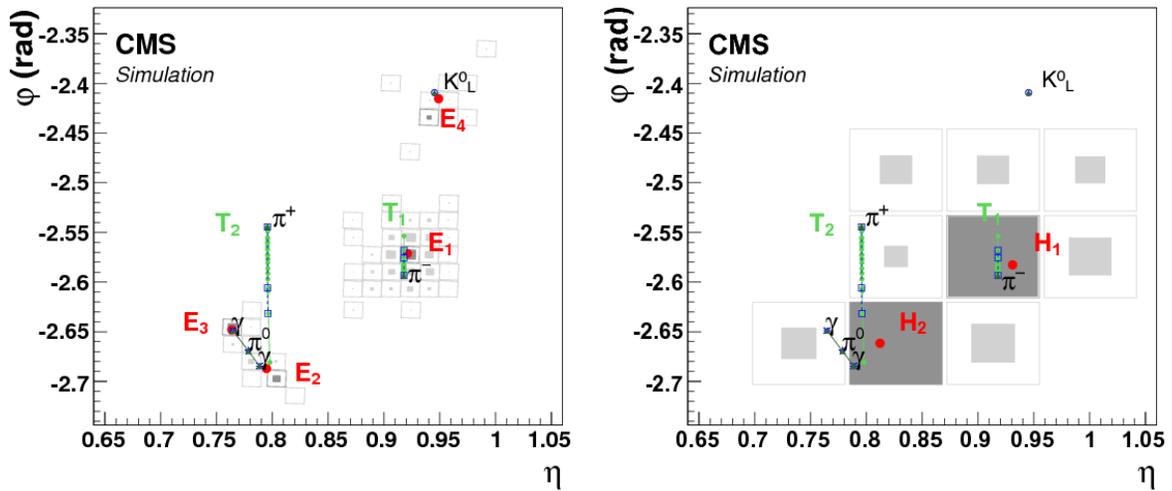


Figure 39: An illustration of particles reconstructed by the PF algorithm for a simulated jet containing a π^+ , a π^- , two photons (from a π^0 decay), and a K_L^0 [97]. Event displays in the ECAL and HCAL are shown in the left and right of the figure, respectively. Also shown in the figure are green lines (dots) depicting the reconstructed tracks (and their associated hits) for the charged pions. The calorimeter cells are shown by the squares in the figure. The energy deposits within the cells are displayed in grey with the size of the squares proportional to the energy in the cell.

elements to the PF algorithm, the link algorithm, and the reconstruction of the PF candidates are provided.

5.1.1 Reconstruction of the Input Elements to the Particle Flow Algorithm

Tracks are reconstructed using a multi-stage approach based on the Kalman filter [98, 99]. In the first stage, track seeds are formed requiring only a few hits (2 or 3) in the pixel detector. These seeds form a list of track candidates. Next, the trajectories from these seeds are extrapolated to each layer of the tracker, searching for additional hits that can be assigned to the track candidate. As additional hits are identified, the track trajectories are updated accordingly. In the last stage, a final fitting is performed using all reconstructed hits to determine the charged-particle properties: origin, transverse momentum, and direction. To improve the track finding efficiency without losing the purity (i.e., increasing the misreconstruction rate), several iterations of the track finding algorithm are run before the final list of tracks is obtained. The motivation for iterative track finding comes from the fact that the initial iterations search for tracks that are easiest to find (have relatively large p_T , and are produced near the interaction region). After each iteration, hits associated with identified tracks are removed based on quality criteria such as the χ^2 of the fit, thereby reducing the combinatorial complexity, and simplifying subsequent iterations when searching for tracks that are more difficult to find (low- p_T , or greatly displaced tracks).

Unlike charged hadrons, electrons can also interact in the tracker through bremsstrahlung. This can result in a loss in efficiency for finding electrons in the tracker because the Kalman filter may not be able to accommodate the changes in the electron momenta when the electrons

undergo bremsstrahlung. In these cases, the tracks tend to be lost due to insufficient number of hits and/or too large χ^2 in the fits. To recover the efficiency, the fits for these tracks are performed using a modified version of the Kalman filter, called the Gaussian Sum Filter [100] (GSF), which allows for sudden and substantial energy losses along the trajectory of the track candidates.

The energy deposits in the ES, ECAL, and HCAL are clustered as follows. First, cluster seeds are identified. A seed cell is defined as a cell with an energy larger than a given threshold, and larger than the energy of the neighboring cells. For HCAL clusters, the cells considered as neighbors are the four closest cells, which share a side with the seed cell. The cells considered as neighboring cells for ECAL clusters are the eight closest cells, which share a side or a corner with the seed cell. Then, topological clusters are constructed based on the seeds by aggregating cells with an energy at least twice the noise energy threshold and with at least a corner shared with a cell already in the topological cluster. With this definition, a topological cluster may contain multiple seeds. Finally, clusters within a topological cluster are reconstructed using a Gaussian-mixture model. In the Gaussian-mixture model, the energy deposits within a topological cluster are described by the summation of N Gaussian functions, where N is the number of seeds in the topological cluster. The amplitudes and means for the Gaussian functions represent the energies and positions of the clusters, and are obtained by fitting the N Gaussian functions to the energy deposits in the topological cluster. With this construction, energy deposits from two nearby particles can be well separated.

5.1.2 Link algorithm

The links between tracks and calorimeter clusters are established as follows. Each track is extrapolated to the ECAL and HCAL to depths corresponding to the expected maximum of a typical longitudinal electron shower profile in the ECAL, and one interaction length in the HCAL. If the extrapolated track falls within the boundaries of a cluster, the cluster is associated to the track. With this construction, several clusters can be linked to the same track, and several tracks can be linked to the same cluster. To remove combinatorial complexities in these cases, the distances between any of the extrapolated track positions and the cluster positions in the η - ϕ plane are calculated and only the link that has the smallest distance is kept. To recover the energy of photons emitted by electron bremsstrahlung, the tangential lines to the GSF tracks at the intersection points between the track and each of the tracker layers are extrapolated to the ECAL. If the extrapolated tangential position is within the boundaries of a cluster, and the distance between the cluster and the extrapolated track in η is smaller than 0.05, the cluster is linked to the track as a potential bremsstrahlung photon. The tracking system is relatively thick in terms of radiation length and therefore bremsstrahlung photons, as well as prompt photons, have a good possibility to convert to e^+e^- pairs. A dedicated conversion finder [101] is used to recover these photons.

In addition to the links between tracks and calorimeter clusters, links are also formed between HCAL and ECAL clusters, and also ECAL and ES clusters. A link is established when the cluster position in the more granular calorimeter (ES or ECAL) is within the cluster envelope of the less granular calorimeter (ECAL or HCAL). When multiple HCAL clusters are

linked to the same ECAL cluster, or when multiple ECAL clusters are linked to the same ES clusters, the link distances between any of the two calorimeter clusters are calculated and only the link with the smallest distance is kept.

5.1.3 Particle Flow Candidates

As mentioned earlier, after the input elements are linked together, PF blocks are formed, and the identification and reconstruction of the PF candidates proceeds starting from each of the PF blocks. Muon candidates are identified first by matching the tracks reconstructed by the muon detectors with those reconstructed by the inner tracker. Then, electrons and charged hadrons are identified using PF blocks that have a link between a track and calorimeter clusters. When reconstructing the electron 4-momenta, the 4-momentum of the Bremsstrahlung photons are combined with the electron 4-momenta. As 90% of the charged hadrons are charged pions, when reconstructing the charged hadron 4-momentum from tracks, the charged hadrons are assumed to be charged pions and the transverse energy of the charged hadrons are calculated using the track p_T of the charged hadrons.

After muons, electrons, and charged hadrons are reconstructed, the associated tracks and calorimeter clusters are removed from the list of PF blocks. Photons and neutral hadrons are then reconstructed using the remaining energy deposits in ECAL and HCAL, respectively.

5.2 Particle Flow Jet Reconstruction and Calibration

PF jets are formed by clustering the muons, electrons, photons, and charged and neutral hadrons as found by the PF algorithm. At CMS, the most commonly used jet clustering

algorithm is the anti- k_t jet clustering algorithm. The energy composition of the PF constituents in jets are shown in Figure 40 for data and simulation as a function of p_T and η .

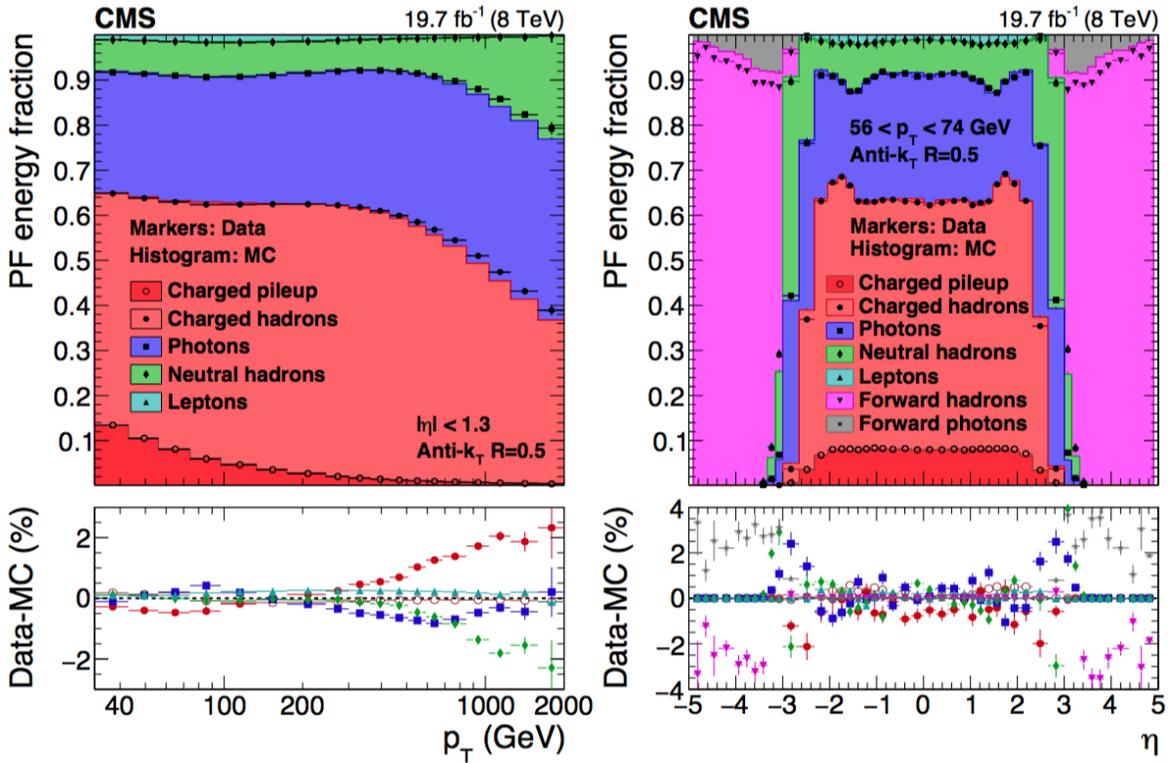


Figure 40: PF jet constituents as a function of jet p_T for jets with $|\eta| < 1.3$ (left) and jet η for jets with $56 < p_T < 74$ [97].

Because the PF candidates are reconstructed using the combined information from all the CMS sub-detectors, the measurement of the jet momenta obtained from PF jets is more reliable than the momenta obtained from calorimeter jets. To illustrate this, the jet p_T response, defined

as the ratio of the reconstructed jet p_T to the reference jet p_T , and the jet p_T resolution, defined as the fluctuations in the jet p_T response, from PF jets and calorimeter jets are compared in Figure 41. Both the response and the resolution for simulated events are significantly better for PF jets with jet momenta below ~ 1 TeV.

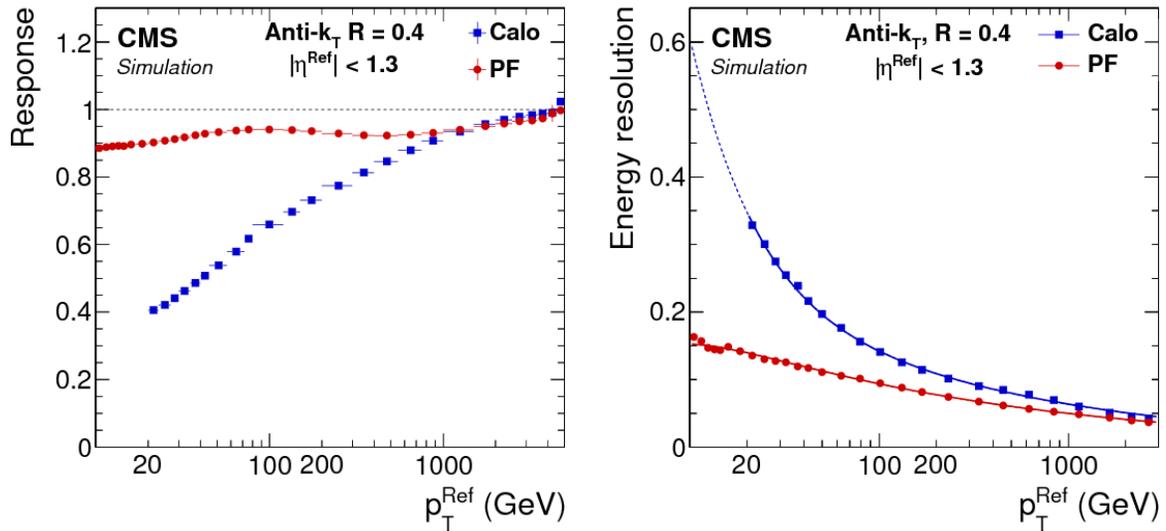


Figure 41: The jet response (left) and jet momentum resolution (right) for PF jets (red) and calo jets (blue) with $|\eta| < 1.3$ [97].

5.2.1 Pileup Mitigation

As mentioned before, the LHC runs in a very high luminosity environment, which results in as many as 50 proton-proton collisions per bunch crossing. This leads to challenges in the event reconstruction at the LHC experiments. The kinematics of the reconstructed jets will be

affected by the energies from adjacent pileup interactions being clustered into jets associated with the primary interaction as illustrated in Figure 42.

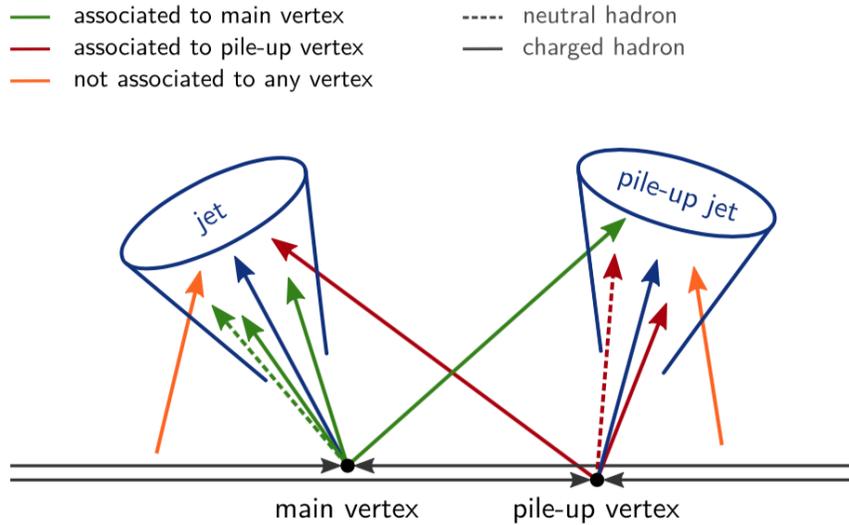


Figure 42: An illustration of the pileup effects on jet clustering. The particle originated from adjacent pileup interactions may be clustered into jets associated to the primary interaction.

The pileup effects on the jet momentum measurement are mitigated using a technique called Charged Hadron Subtraction (CHS) [102]. In this approach, charged hadrons that are unambiguously not associated with the event primary vertex [90] are removed from the list of the PF candidates to be included in the jet reconstruction. The primary vertex is defined as the reconstructed vertex with the highest sum of the squares of all associated physics objects p_T .

The physics objects are the jets returned by the anti- k_t algorithm, using all tracks associated with the vertex as input, plus the corresponding associated missing transverse momentum.

The residual pileup energy, which is mainly due to photons and neutral hadrons originating from pileup interactions, will be corrected using a method described Section 5.2.3.1.

5.2.2 Identification and Rejection of Fake Jets

Noise in calorimeter cells can lead to fake jets and also to misreconstruction of the jet energies and directions. A set of jet identification criteria has been developed by CMS to remove jets created by detector electronic noise. These criteria are based on the neutral hadron fraction (NHF), neutral electromagnetic (photon) fraction (NEF), charged hadron fraction (CHF), charged electromagnetic (electron) fraction (CEF), number of constituents ($n_{\text{constituents}}$) and charged multiplicity (n_{charged}). For this analysis, the “tight” criteria have been applied, and the requirements are summarized in Table V. In addition to these requirements, jets with a muon energy fraction greater than 0.8 are rejected to suppress noise from misreconstructed high energy muons.

TABLE V: The “tight” PF jet quality criteria.

variable	$ \eta $	tight JetID
<i>NHF</i>		< 0.9
<i>NEF</i>		< 0.9
<i>CHF</i>	< 2.4	> 0
<i>CEF</i>	< 2.4	< 0.99
$n_{\text{constituents}}$		> 1
n_{charged}	< 2.4	> 0

5.2.3 Jet Energy Scale Calibration

The reconstructed jet p_T needs to be corrected to match the p_T of the particles that initiated the jet. There are a number of reasons for this. As mentioned before, additional energy may be clustered into jets due to pileup interactions and detector electronic noise. Energy may also be absorbed in uninstrumented regions of the detector and therefore fail to be reconstructed. The jet p_T also needs to be corrected for the non-linear and non-uniform response of the detectors themselves. The jet p_T correction is known as the jet energy correction (JEC) or jet energy scale calibration (JES).

The JEC is performed in a series of factorized steps as illustrated in Figure 43. The first step, described in Section 5.2.3.1, is called the offset correction. This step corrects the jet p_T from energy contributions coming from detector electronic noise and residual pileup effects. The second step, called the MC truth correction and described in Section 5.2.3.2, corrects the jet p_T response and is derived from simulations. The residual differences in the jet p_T response between data and simulation are taken into account using a data-driven approach in the third step (discussed in Section 5.2.3.3). The third step is called residual correction and is performed in two stages: first a relative correction that takes into account the η dependence of the difference is applied. This is followed by an absolute correction that takes into account the p_T dependence of the difference. A fourth step, which accounts for differences in jet response between jets with different quark flavor compositions, may be applied for certain analyses (e.g., analyses using b-jets) if necessary. This analysis does not apply the fourth step.

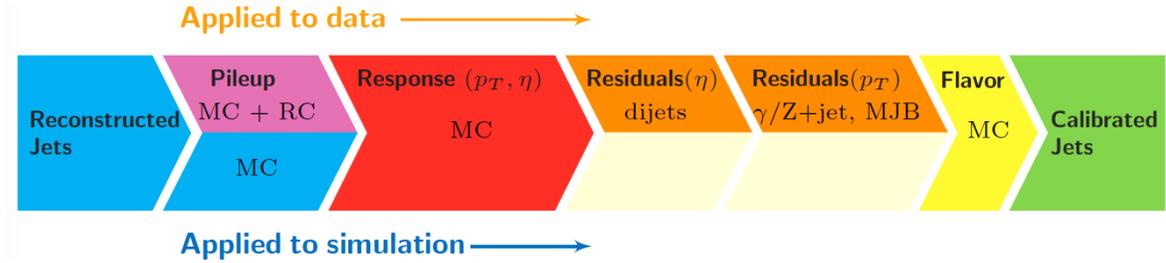


Figure 43: The JES calibration workflow applied to data (top row) and simulation (bottom row).

The JEC is applied as a multiplicative factor C to the raw jet p_T and it is expressed as:

$$p_T^{\text{corr}} = C \cdot p_T^{\text{raw}}. \quad (5.1)$$

The correction factor C is obtained by multiplying together the correction factors for each of the JEC steps and can be expressed as

$$C = C_{\text{offset}}(p_T^{\text{raw}}) \cdot C_{\text{MC}}(p_T', \eta) \cdot C_{\text{relative}}(\eta) \cdot C_{\text{absolute}}(p_T'') \quad (5.2)$$

where C_{offset} , C_{MC} , C_{relative} , and C_{absolute} are the correction factors for the offset, MC truth, relative, and absolute corrections, respectively. The offset correction is a function of jet η and number of pileup vertices in the event, and is applied to the raw jet p_T . The MC truth correction is a function of jet η and p_T and is applied to the jet p_T after the offset correction is applied (p_T'). The relative correction is a function of jet η and the absolute correction is a function of

jet p_T . The absolute correction is applied to the jet p_T after the MC truth correction is applied (p_T''). A detailed description for the JES in CMS can be seen in Ref. [103].

5.2.3.1 Offset Correction

The offset correction is the first step in the JEC and it subtracts the energy contribution coming from detector electronic noise and pileup interactions from the jet p_T . The correction factor is expressed as

$$C_{\text{offset}} = 1 - \frac{\rho \cdot A}{p_T^{\text{raw}}} \quad (5.3)$$

where A is the jet area and ρ is the offset energy density.

The jet area is determined using the technique described in Section 2.3. The offset density ρ characterizes the energy density from pileup interactions and electronic noise and it is obtained using the random cone (RC) method. In the RC method, jet cones are randomly placed in the η - ϕ plane for data events that are selected with the zero bias trigger in various luminosity conditions. Since these events typically contain only low energy pileup collision products and detector electronic noise, the average p_T of the jets measured in the RC method can be used to represent the offset energy in a high- p_T jet reconstructed with the same jet algorithm and same jet size parameter. A comparison of the average offset energy per additional pileup interaction in data and simulation derived from the RC method is shown in Figure 44 [104].

5.2.3.2 Monte Carlo Truth Correction

The MC truth correction is the second step of the JEC and it corrects the reconstructed jet p_T to match the true (generated) jet p_T . The correction takes advantage of the fact that

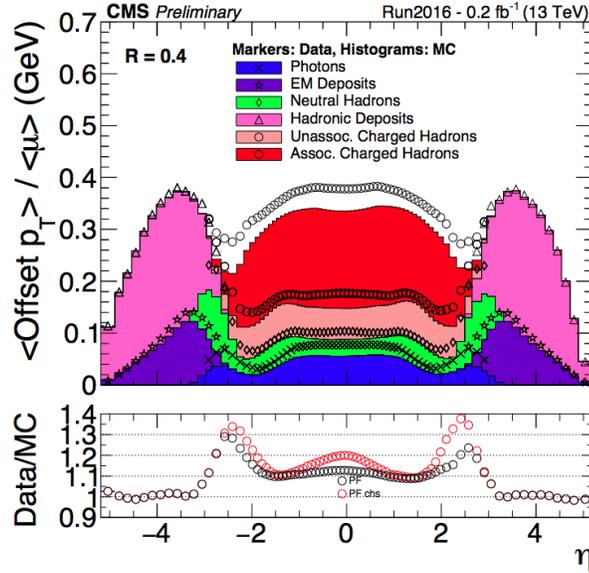


Figure 44: Average offset energy per additional pileup interaction in data and simulation derived using the RC method [104].

both the true jet p_T and the reconstructed jet p_T are available in the MC simulations. The simulated samples are obtained from dijet events generated by PYTHIA that include a detailed simulation of the CMS detector. More details about the detector simulation will be described in the next chapter.

The jet response $R = \frac{p_T^{\text{reco}}}{p_T^{\text{truth}}}$ is obtained using reconstructed jets and generated jets that are matched in η and ϕ , and it is calculated in bins of the generated jet η and p_T . The MC truth correction factor C_{MC} is then obtained as the inverse of the average jet response in each η and p_T bin. The jet response as a function of η for various jet transverse momenta are shown in Figure 45.

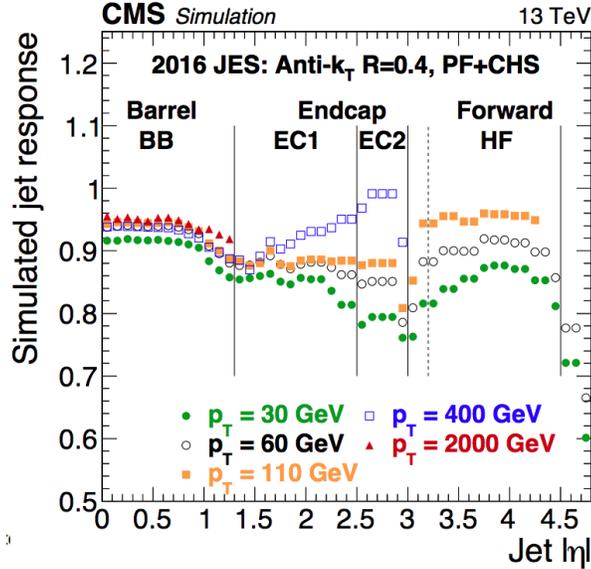


Figure 45: The simulated jet response as a function of η for various jet transverse momenta [104].

5.2.3.3 Residual Correction

The residual correction corrects for the differences in jet response in data and simulation. It is performed in two steps: first a relative correction is applied to take into account the η dependence of the difference. Then, an absolute correction that takes into account the p_T dependence of the difference is applied.

Relative Correction The relative correction is obtained from a sample of dijet events [105] using the “tag” and “probe” method. In dijet events, the p_T of the two leading jets must balance in the limit of no radiation. To make sure that the p_T ’s of the two leading jets are truly balanced, the dijet sample is selected with the requirements that the third leading jet p_T is much smaller than the average p_T of the two leading jets and that the azimuthal angle between the two

leading jets, $\Delta\phi$, is close to π . In this method, the “tag” (reference) jet is restricted to the barrel region of $|\eta| < 1.3$, since the η -dependence of the jet response is small in this region, and the “probe” jet is unconstrained in η .

The relative correction factor C_{relative} is based on the relative response R_{relative} , which is defined as $R_{\text{relative}} = \frac{\langle p_{\text{T}}^{\text{probe}} \rangle}{\langle p_{\text{T}}^{\text{tag}} \rangle}$. To reduce biases due to the jet p_{T} resolution, the relative response is binned with respect to the average p_{T} of the “probe” and “tag” jets, $p_{\text{T}}^{\text{ave}}$, instead of the “tag” jet p_{T} . The relative response, binned in $p_{\text{T}}^{\text{ave}}$ and “probe” jet η , is expressed as

$$R_{\text{relative}} = \frac{1 + \langle A \rangle}{1 - \langle A \rangle} \quad (5.4)$$

where

$$A = \frac{p_{\text{T}}^{\text{prob}} - p_{\text{T}}^{\text{tag}}}{2p_{\text{T}}^{\text{ave}}}, \quad (5.5)$$

and is calculated for both data and simulation. The relative correction factor C_{relative} is then obtained as the ratio between the relative response in data and simulation $\langle \frac{R_{\text{relative}}^{\text{data}}}{R_{\text{relative}}^{\text{MC}}} \rangle$ in various η bins.

A more advanced method, the Missing energy Projection Fraction (MPF) technique, has also been developed to take into account all the hadronic activity in the events. The MPF is based on the following relation:

$$\vec{p}_{\text{T}}^{\text{tag}} + R_{\text{relative}}^{\text{MPF}} \vec{p}_{\text{T}}^{\text{probe}} = -\vec{E}_{\text{T}}^{\text{miss}}. \quad (5.6)$$

where $E_T^{\vec{\text{miss}}}$ is defined as vector sum of all PF candidates p_T 's and $R_{\text{relative}}^{\text{MPF}}$ is the relative response defined in the MPF method. $R_{\text{relative}}^{\text{MPF}}$ is also binned in “probe” jet η and p_T^{ave} , and is expressed as:

$$R_{\text{relative}}^{\text{MPF}} = \frac{1 + \langle B \rangle}{1 - \langle B \rangle}, \quad (5.7)$$

where

$$B = \frac{E_T^{\vec{\text{miss}}} \cdot (p_T^{\text{tag}} / p_T^{\text{ave}})}{2p_T^{\text{ave}}}. \quad (5.8)$$

The relative correction in the MPF method is then obtained as $C_{\text{relative}} = \langle \frac{R_{\text{relative}}^{\text{data,MPF}}}{R_{\text{relative}}^{\text{MC,MPF}}} \rangle$ in various η bins. C_{relative} derived using the MPF method is shown in Figure 46 (left).

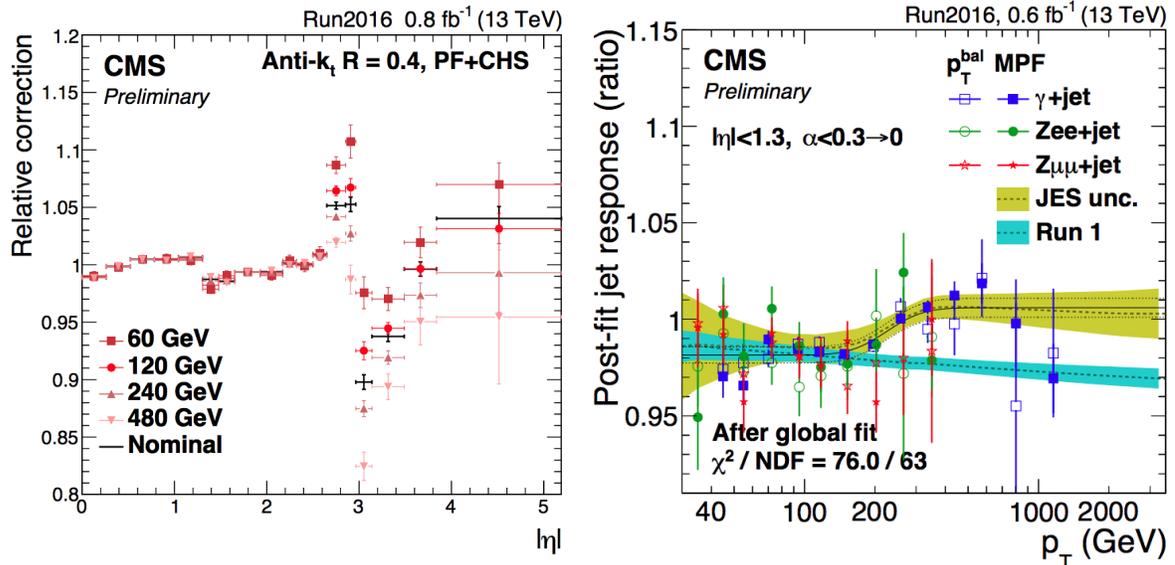


Figure 46: The relative (left) and absolute (right) jet p_T corrections [104].

Absolute Correction The absolute correction, is obtained in the rapidity region $|\eta| < 1.3$ using events that have a photon or Z boson recoiling against a jet. Similar to the dijet sample used to derive the relative correction, the γ/Z +jet sample is selected with the requirements that the second leading jet p_T is much smaller than the p_T of the photon or Z boson and that the azimuthal angle between the leading jet and the recoil γ/Z is close to π . The Z bosons are reconstructed in $Z \rightarrow e^-e^+$ and $Z \rightarrow \mu^-\mu^+$ channels. This method takes advantage of the fact that the reconstructed energies for single e^\pm 's and μ^\pm 's are very close to their true energies.

As with the relative correction, the absolute correction can be obtained using two methods, the p_T -balance and MPF methods. In the p_T -balance method, the absolute response is defined as $R_{\text{absolute}} = \frac{p_T^{\text{jet}}}{p_T^{\gamma/Z}}$. In the MPF method, the absolute response satisfies the relation:

$$\vec{p}_T^{\gamma/Z} + R_{\text{absolute}}^{\text{MPF}} \vec{p}_T^{\text{jet}} = -E_T^{\text{miss}} \quad (5.9)$$

and is expressed as

$$R_{\text{absolute}}^{\text{MPF}} = 1 + \frac{E_T^{\text{miss}} \cdot \vec{p}_T^{\gamma/Z}}{(p_T^{\gamma/Z})^2}. \quad (5.10)$$

The absolute correction is then obtained as the ratio between the mean absolute response in data and simulation $\frac{\langle R_{\text{absolute}}^{\text{data}} \rangle}{\langle R_{\text{absolute}}^{\text{MC}} \rangle}$ ($\frac{\langle R_{\text{absolute}}^{\text{data,MPF}} \rangle}{\langle R_{\text{absolute}}^{\text{MC,MPF}} \rangle}$) in bins of the Z/γ p_T for the p_T -balance (MPF) method. The absolute correction can be seen in Figure 46 (right).

5.2.4 Jet Energy Resolution

As mentioned before, the jet p_T resolution is defined as the width of the jet p_T response distribution, $\sigma(\frac{p_T^{\text{reco}}}{p_T^{\text{truth}}})$. Since the jet angular resolution is practically negligible [106], the jet p_T resolution is equivalent to the jet energy resolution (JER). The JER in simulation can be obtained by comparing reconstructed jets that are matched in η and ϕ with truth jets in bins of p_T and η and then measuring the widths of the resultant jet p_T response distributions. The JER can also be obtained from data using the dijet asymmetry, A , which is defined as:

$$A = \frac{p_T^1 - p_T^2}{p_T^1 + p_T^2}, \quad (5.11)$$

where p_T^1 and p_T^2 are the p_T 's of the two leading jets. If the extra jet activity is negligible and the two leading jets have the same η value, i.e., $p_T^1 = p_T^2$ and $|\eta^1| = |\eta^2|$, the width of the asymmetry distribution,

$$\sigma_A^2 = \left| \frac{\partial A}{\partial p_T^1} \right|^2 \sigma^2(p_T^1) + \left| \frac{\partial A}{\partial p_T^2} \right|^2 \sigma^2(p_T^2), \quad (5.12)$$

is related to the JER through the expression,

$$\sigma(A) = \frac{1}{\sqrt{2}} \frac{\sigma(p_T)}{p_T}. \quad (5.13)$$

The additional effects on the JER due to ISR, FSR, UE, and pileup are also included in the JER measurements using the methods detailed in Ref. [105]. The JER derived in simulation for two η regions, $|\eta| < 0.5$ and $3.2 < |\eta| < 4.7$, and various pileup conditions are shown in

Figure 47. The JER is seen to degrade as the average pileup increases at low p_T , but the degradation becomes less and less with increasing jet p_T , and is independent of pileup for jet $p_T > 100$ GeV.

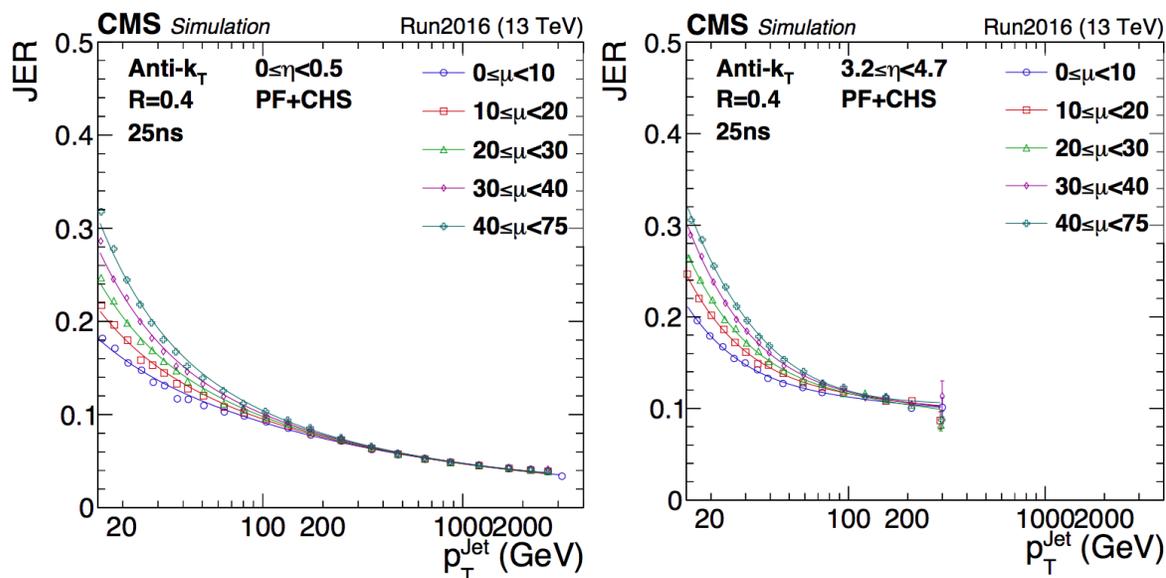


Figure 47: The JER derived in simulation for jets with $|\eta| < 0.5$ (left) and $3.2 < |\eta| < 4.7$ (right) [104].

CHAPTER 6

MEASUREMENT OF THE DIJET ANGULAR DISTRIBUTIONS

6.1 Event Selection

The events used in this analysis are selected by either the HLT single jet or HT triggers. The HLT selection requires that the event contains a jet with $p_T > 450$ GeV or that the H_T has a value greater than 900 GeV. These thresholds were the lowest unrescaled thresholds used during the 2016 run. The trigger efficiency is defined as the ratio between the total number of events selected by the trigger and the total number of events and is shown in Figure 48 as a function of M_{jj} . The efficiency is measured using a single muon trigger as the reference trigger since the single muon trigger is used as the reference trigger because the single muon trigger is uncorrelated with the single jet and H_T triggers. As seen in the figure, the trigger is fully efficient above $M_{jj} = 2.0$ TeV in the analysis phase space (see below).

The χ_{dijet} distribution is measured over the range $1 < \chi_{\text{dijet}} < 16$, which implies $y^* = \frac{1}{2}|y_1 - y_2| < 1.39$, where y_1 and y_2 are the rapidities of the two jets with the highest p_T in the event (leading jets). In this analysis, the value of y_{boost} , where $y_{\text{boost}} = \frac{1}{2}(y_1 + y_2)$, is also required to be less than 1.11. These two requirements constrain $|y_1|$ and $|y_2|$ to have values less than 2.5. This is done in order to restrict the jets to the central region of the calorimeter, where the jet performance is maximized. It is interesting to see how the topologies of the dijets depend on the values of y^* and y_{boost} . This is illustrated in Figure 49. As can be seen in the

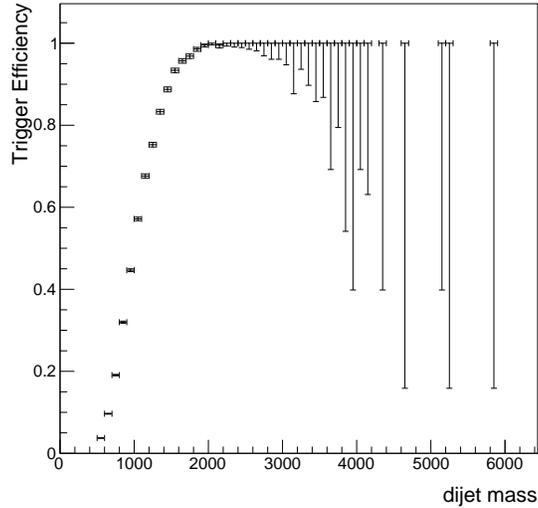


Figure 48: The trigger turn-on curve for the combined single jet $p_T > 450$ GeV and $H_T > 900$ GeV triggers as a function of M_{jj} in the analysis phase space ($y^* < 1.39$ and $y_{\text{boost}} < 1.11$).

figure, the most centrally produced jets have y^* and y_{boost} values less than one, while the most forward dijet topologies have low values of y^* and high values of y_{boost} .

The dijet angular distributions are measured in seven M_{jj} bins: 2.4–3.0, 3.0–3.6, 3.6–4.2, 4.2–4.8, 4.8–5.4, 5.4–6.0, 6.0–13.0 TeV. The minimum M_{jj} threshold is above the point where the triggers are fully efficient. The highest M_{jj} range was chosen by optimizing the search sensitivity to the NP signals considered in this analysis.

Although the lower M_{jj} bound of this measurement is 2.4 TeV, events in the 1.9–2.4 TeV M_{jj} range also need to be included in the analysis to account for the effect of event migrations due to the finite JER (see details in Section 6.4) between lower M_{jj} values and the 2.4–3.0 TeV dijet mass bin. Since the trigger is not fully efficient in the 1.9–2.4 TeV mass bin, trigger corrections

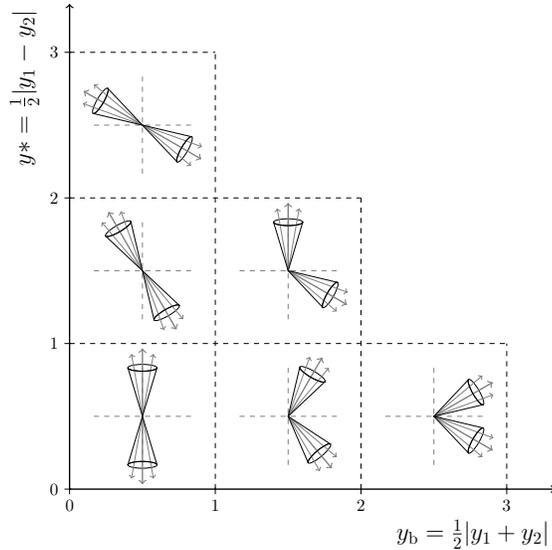


Figure 49: Illustration of the dijet event topologies in the plane of y_{boost} (y_b) and y^* [107].

are applied to these events to account for the inefficiency. The trigger corrections are obtained as the inverse of parametric fits of the trigger turn-on curve vs. M_{jj} in various χ_{dijet} bins. An inverse tangent function is used to fit the trigger turn-on curves. The fits of the trigger turn-on curves in various χ_{dijet} ranges are shown in Figure 50.

In the analysis, events are required to contain a reconstructed primary vertex that lies within ± 24 cm of the nominal detector center along the beam line, and within 2 cm of the detector center in the plane transverse to the beam. Table VI shows the number of data events in each mass bin after all selection criteria are applied: $y^* < 1.39$, $y_{\text{boost}} < 1.11$, “tight” jet criteria (see Section 5.2.2), and primary vertex selection.

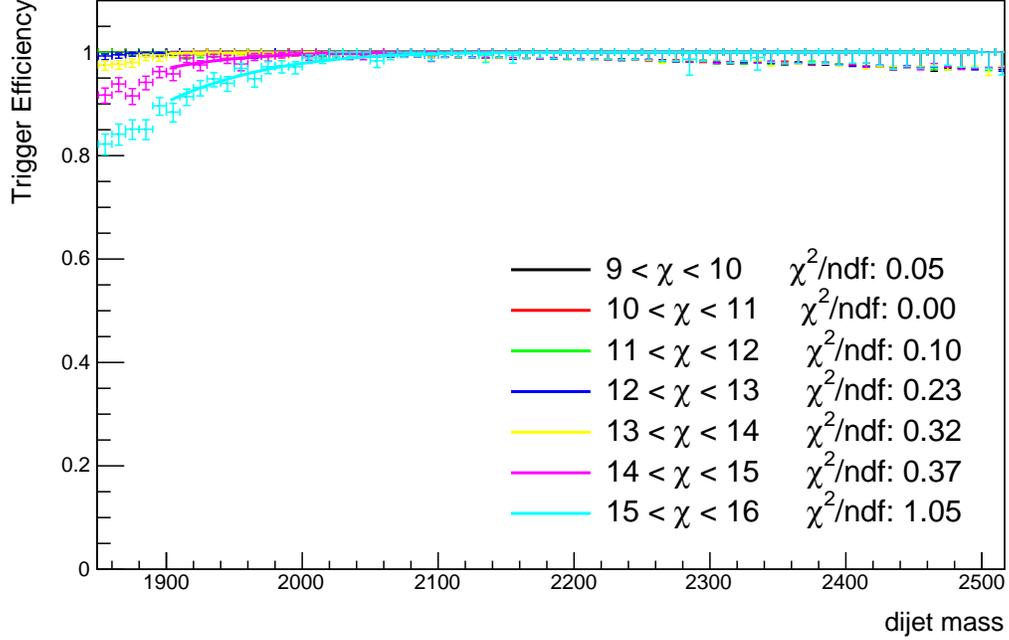


Figure 50: Fits of the trigger efficiencies vs. M_{jj} for several χ_{dijet} bins. The trigger is fully efficient for $M_{jj} > 1.9$ TeV and $\chi_{\text{dijet}} < 9$.

TABLE VI: The number of events in each M_{jj} bins after all the analysis selection criteria are applied.

Mass bin [TeV]	# events
2.4–3.0	353083
3.0–3.6	71840
3.6–4.2	16714
4.2–4.8	4287
4.8–5.4	1154
5.4–6.0	330
6.0–13.0	95

The highest value of M_{jj} observed among these events is 8,153.3 GeV. The event display for this event can be seen in Figure 51. The display shows that the event has properties typical of what one would expect in a real dijet scattering event, rather than an event that was triggered by detector noise or some other non-collision process. The p_T and rapidity of the leading jets for this event are 2,805.3 GeV and -1.1 and 2,535.6 GeV and 0.9. The χ_{dijet} for this event is 7.15. The event displays from all the events in the highest M_{jj} bin were individually scanned to ensure that only hard scattering events were included in the sample.

6.2 Comparison with the Detector Simulation

The CMS detector simulation was done using the GEANT [108] software toolkit developed at CERN. The toolkit is used to simulate the passage of particles through the magnetic field and detector volume, the interactions of the particles with the detector material, and the digitization and readout of the sensitive detector elements, and includes a detailed description of each sub-detector's geometry and material. The GEANT simulation is called the Full Sim. The inputs of the GEANT simulation are the events generated by the MC event generators described earlier. The outputs of the simulation are written to disk in RAW format, the same format used for the data. The Full Sim events are processed and reconstructed with the same reconstruction software as the data. By running the Full Sim events through the same reconstruction algorithms as the data, any inefficiencies or biases in the reconstruction algorithms are present in the Full Sim samples and therefore accounted for in the ensuing corrections.

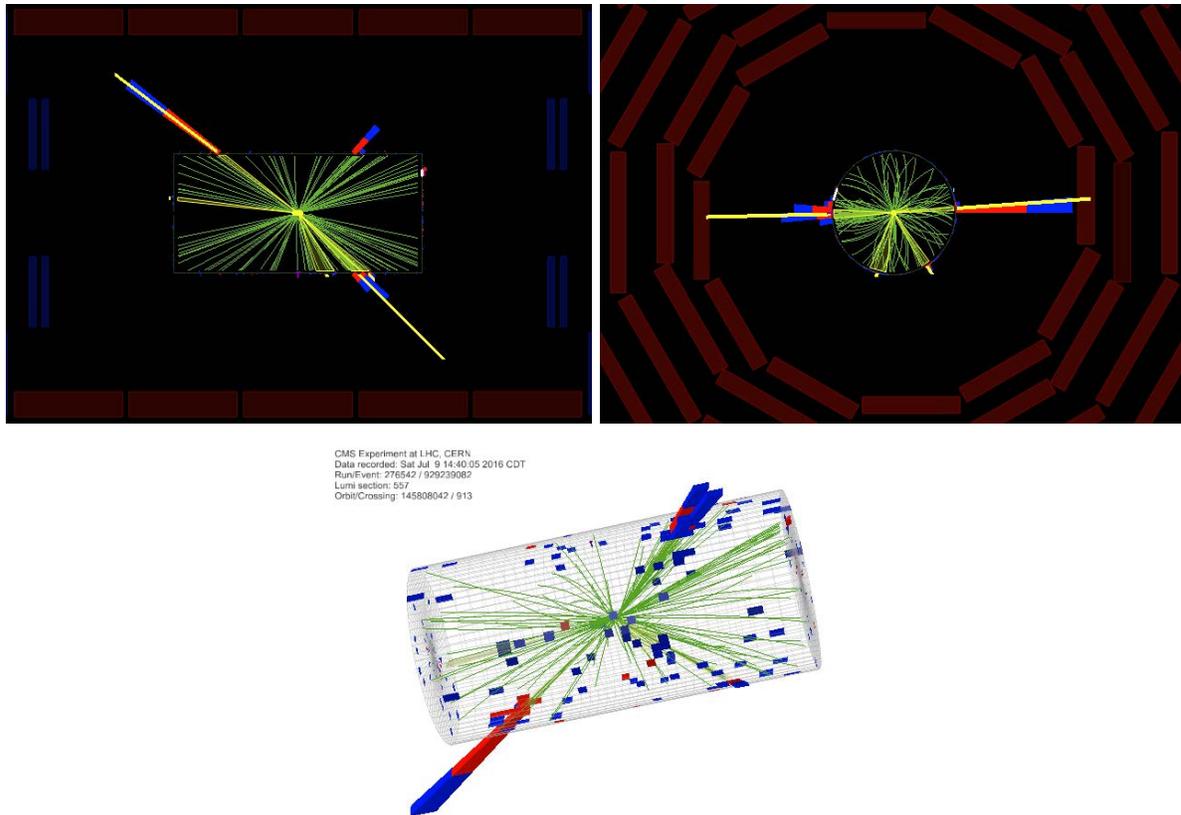


Figure 51: Event display for the event with the highest M_{jj} observed in the analysis phase space.

A variety of comparisons are made between the data and Full Sim MC events to verify that the simulation provides an adequate description of the QCD events observed in the data. The MC events include those from PYTHIA and MADGRAPH5_aMC@NLO, which use Tune CUETP8M1 [75, 76], and HERWIG++, which uses Tune CUETHS1 [76, 77].

The comparison between M_{jj} distributions in data and Full Sim is shown in Figure 52. In this comparison, the total number of events in the simulations are normalized to the total number of

events in data. The PYTHIA and HERWIG++ predictions describe the M_{jj} distributions in data much better than those from MADGRAPH5_aMC@NLO. The modeling of the M_{jj} distributions is important to this measurement since the simulated samples will be used to characterize and correct the effect of event migrations between M_{jj} bins due to the finite JER (see the following sections).

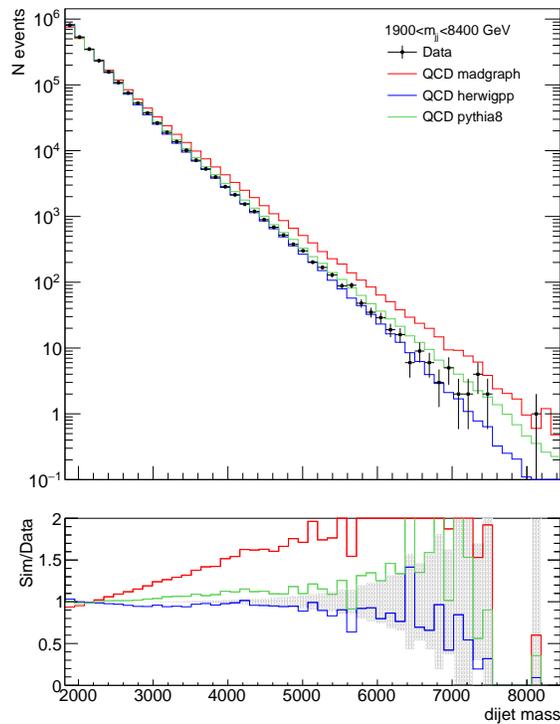


Figure 52: Comparison of M_{jj} distributions in data (points) and the Full Sim (histograms) from the PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO (red) event generators.

In addition to the M_{jj} comparison, comparisons of several other kinematic variables are also performed to check the quality of the simulation. These comparisons are made for all the M_{jj} bins used in the analysis. In each mass bin, the total number of events in the simulated samples is normalized to the total number of events in data. In Figures 53 and 54, the data p_T distributions for the leading and 2nd-leading jets are compared to the MC predictions, respectively. In Figures 55 and 56, the data to MC comparisons of the rapidities for the leading and 2nd-leading jets are shown. The comparisons of y_{boost} , χ_{dijet} , $\frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}$, $\Delta\phi_{1,2}$ (the azimuthal angle separation of the two leading jets), and $\frac{E_T^{\text{miss}}}{\sum E_T}$ (where $\sum E_T$ is the sum of all reconstructed particle E_T 's in the event) are shown in Figures 57 to 61. As can be seen in these figures, the simulations describe the data well for all the distributions. In addition, the comparisons of $\frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}$, $\frac{E_T^{\text{miss}}}{\sum E_T}$, and $\Delta\phi_{1,2}$ distributions show that no significant noise effects are present in the data, as noise would significantly broaden the tails of these distributions in the data as opposed to the MC simulations.

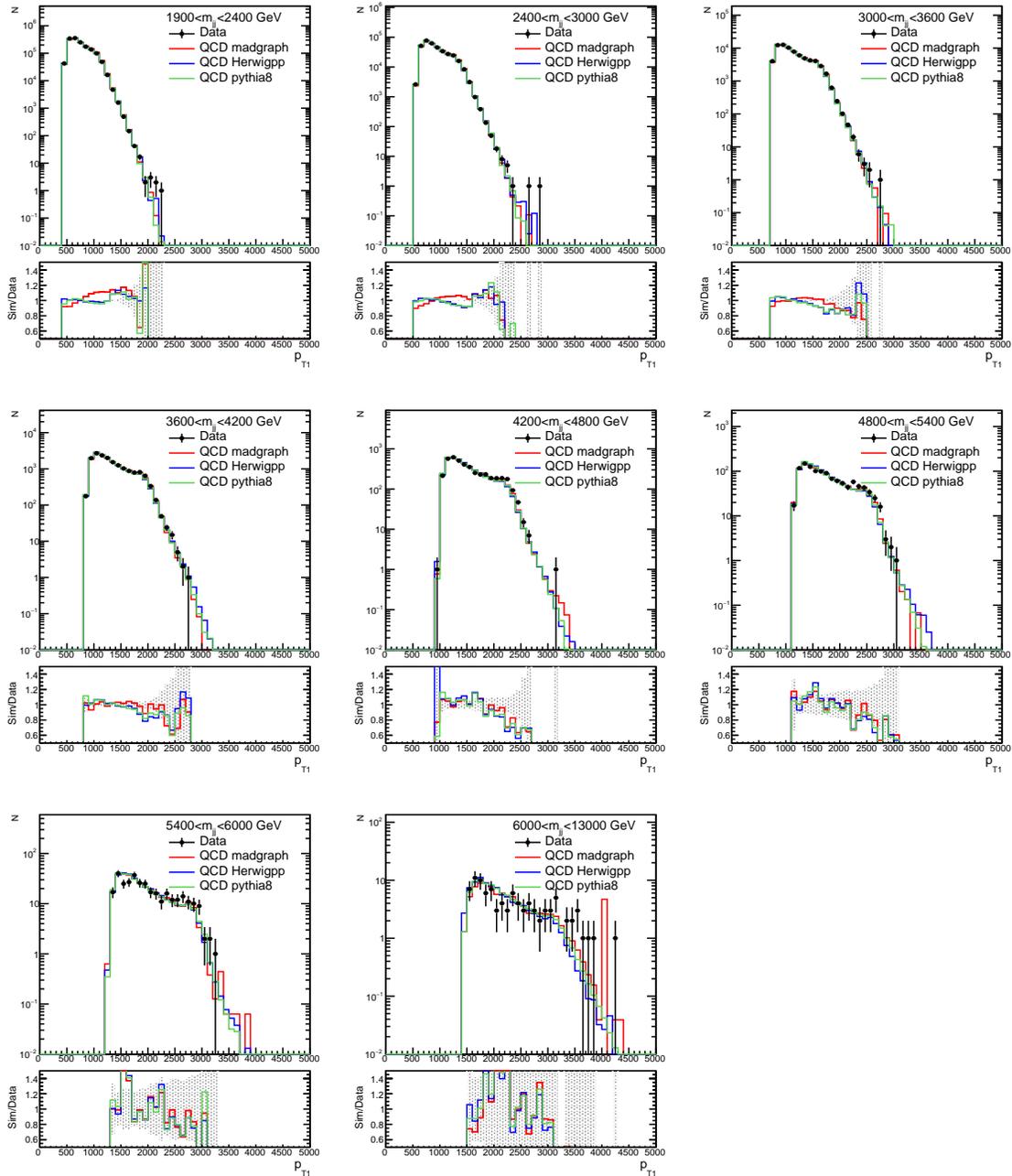


Figure 53: Comparison of p_T distributions for the leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO (red).

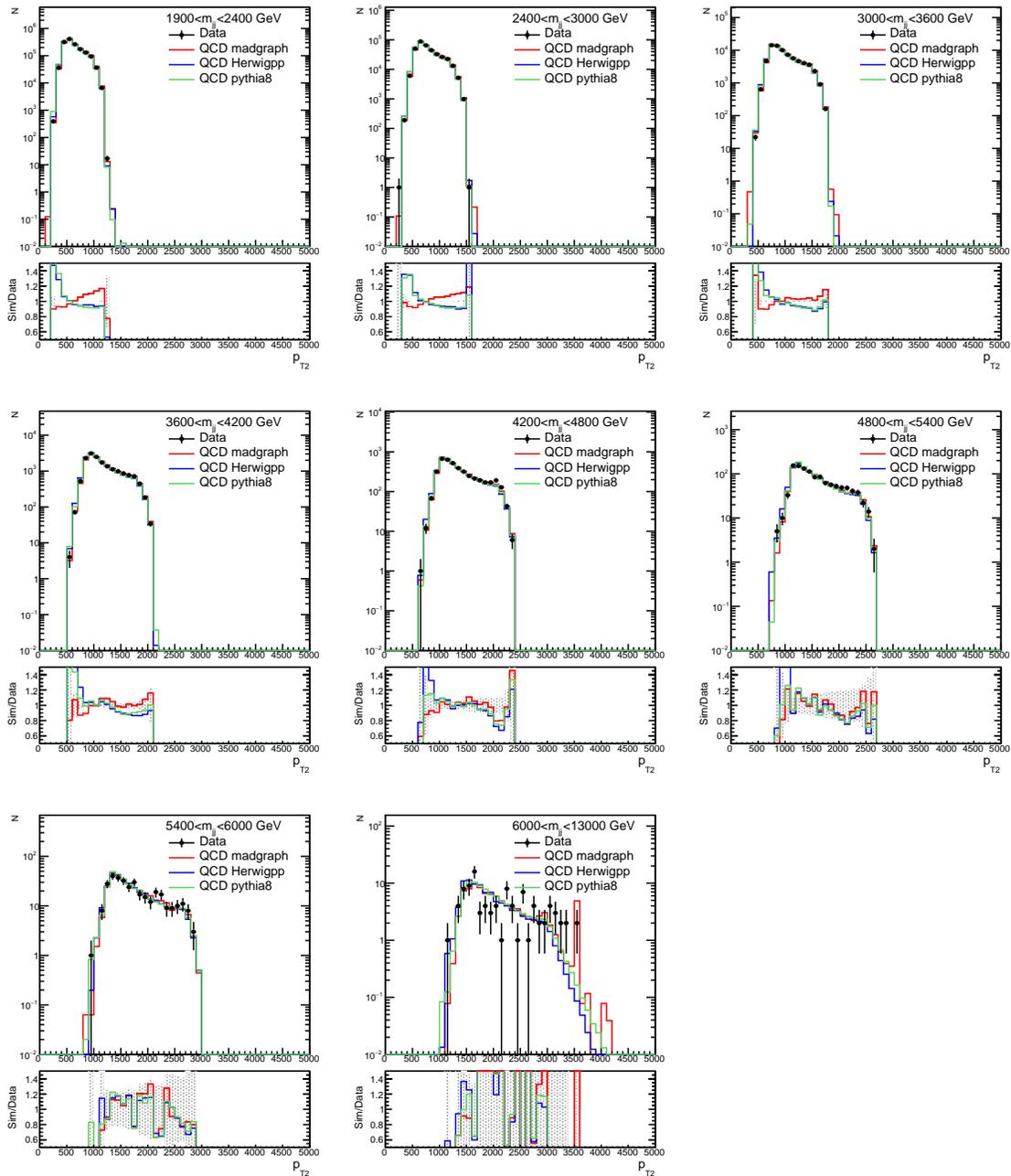


Figure 54: Comparison of p_T distributions for the 2nd-leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_AMC@NLO (red).

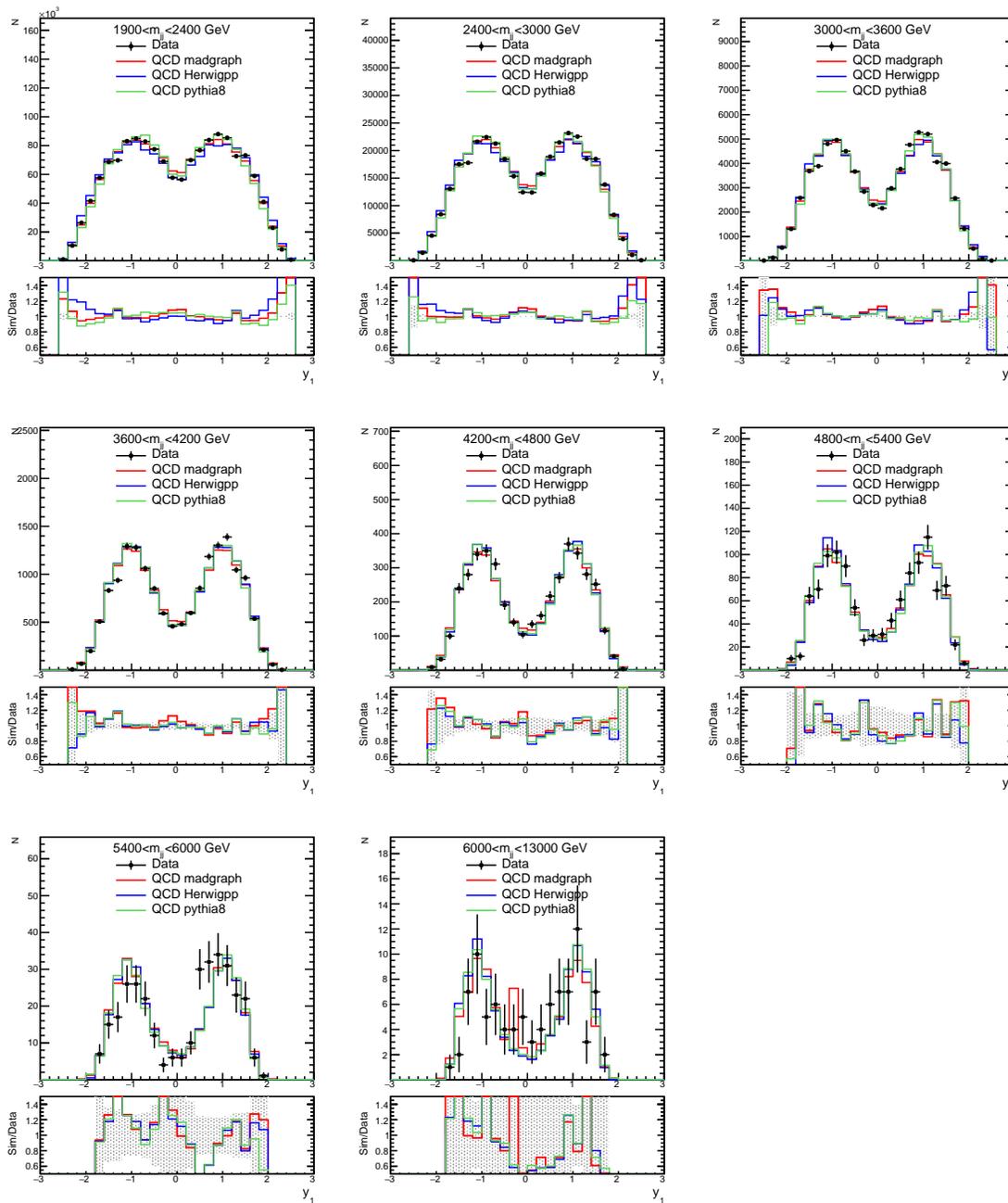


Figure 55: Comparison of rapidity distributions for the leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO (red).

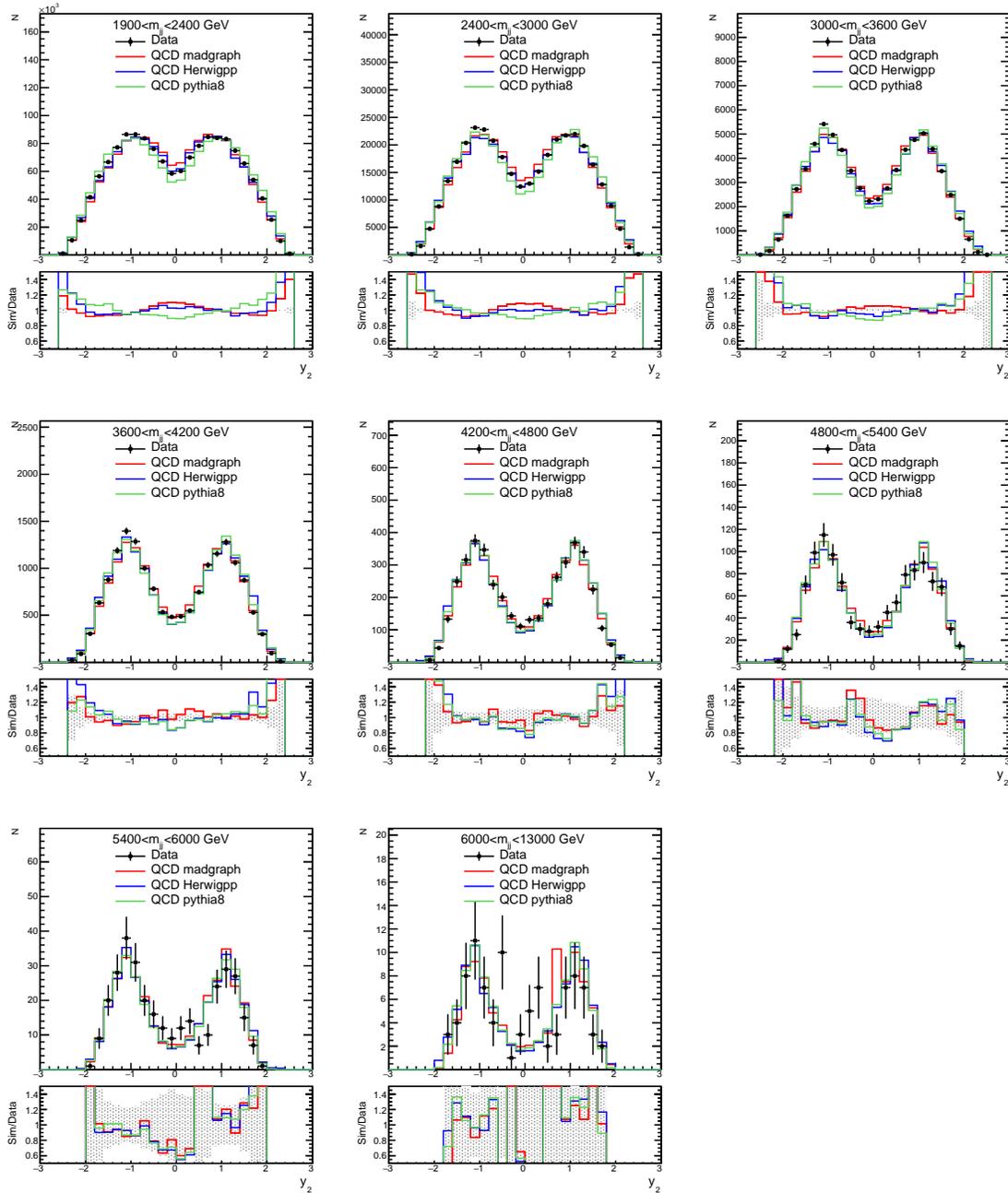


Figure 56: Comparison of rapidity distributions for the 2nd-leading jets in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_amc@NLO (red).

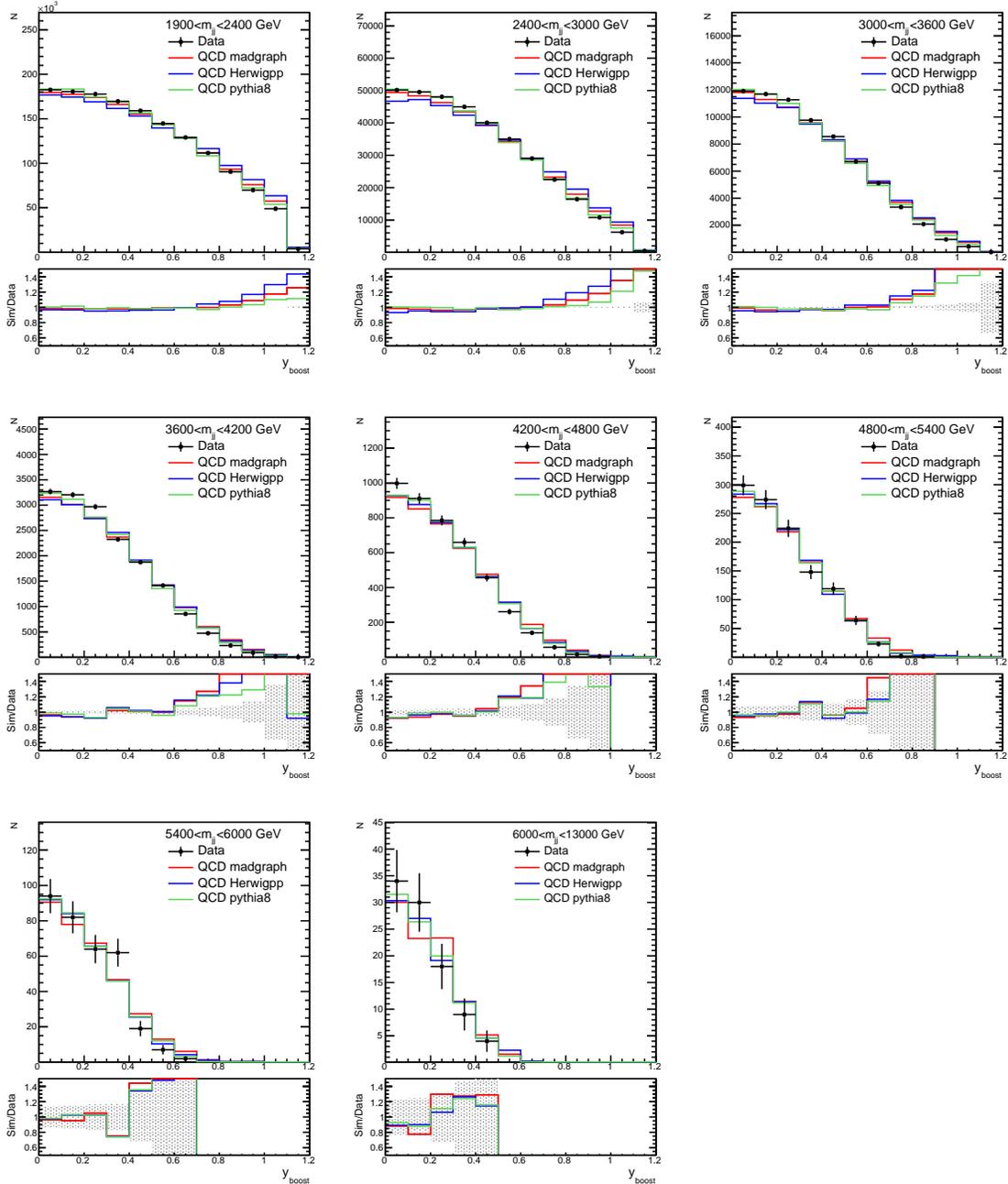


Figure 57: Comparison of y_{boost} distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_AMC@NLO (red).

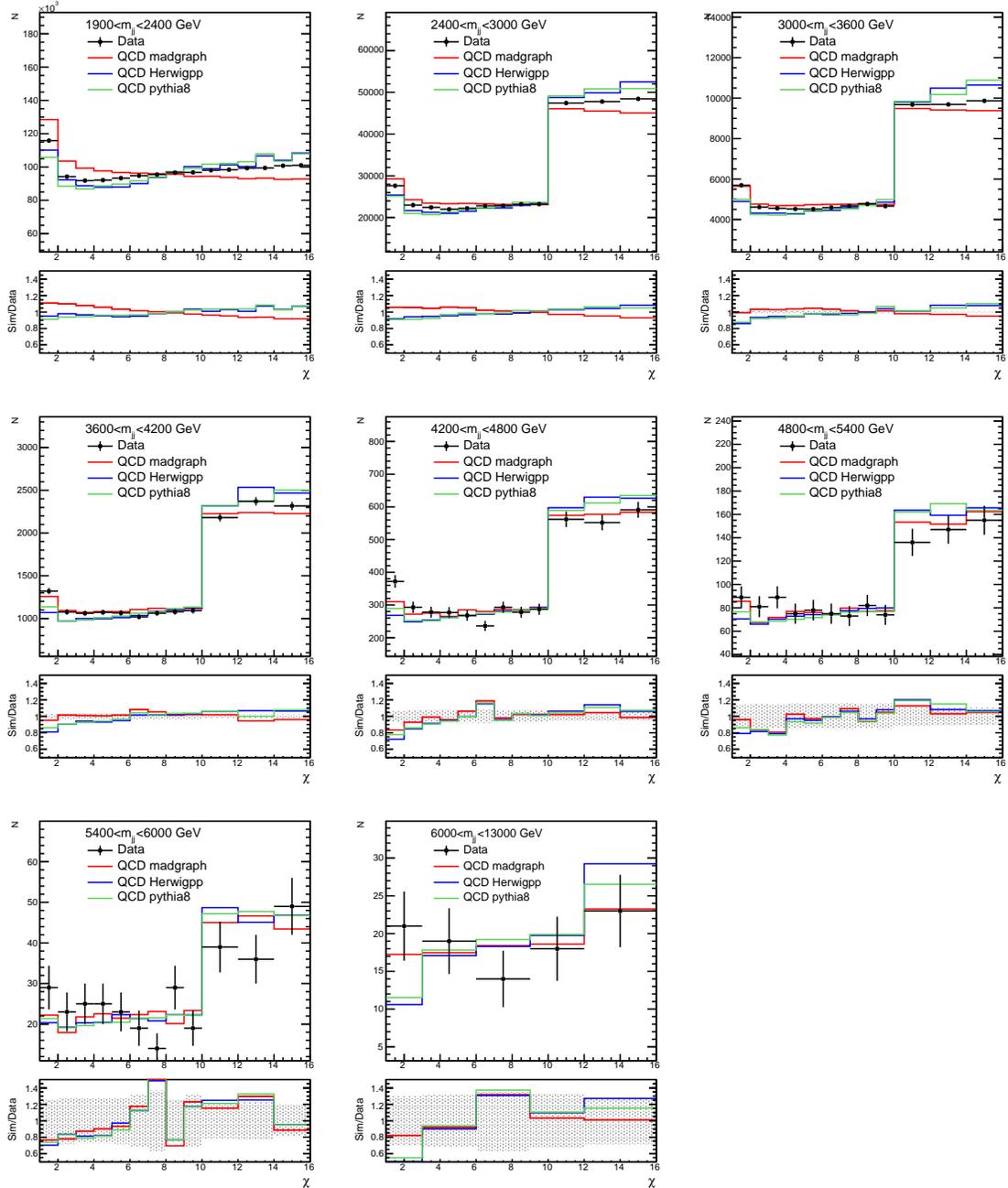


Figure 58: Comparison of χ_{dijet} distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_AMC@NLO (red).

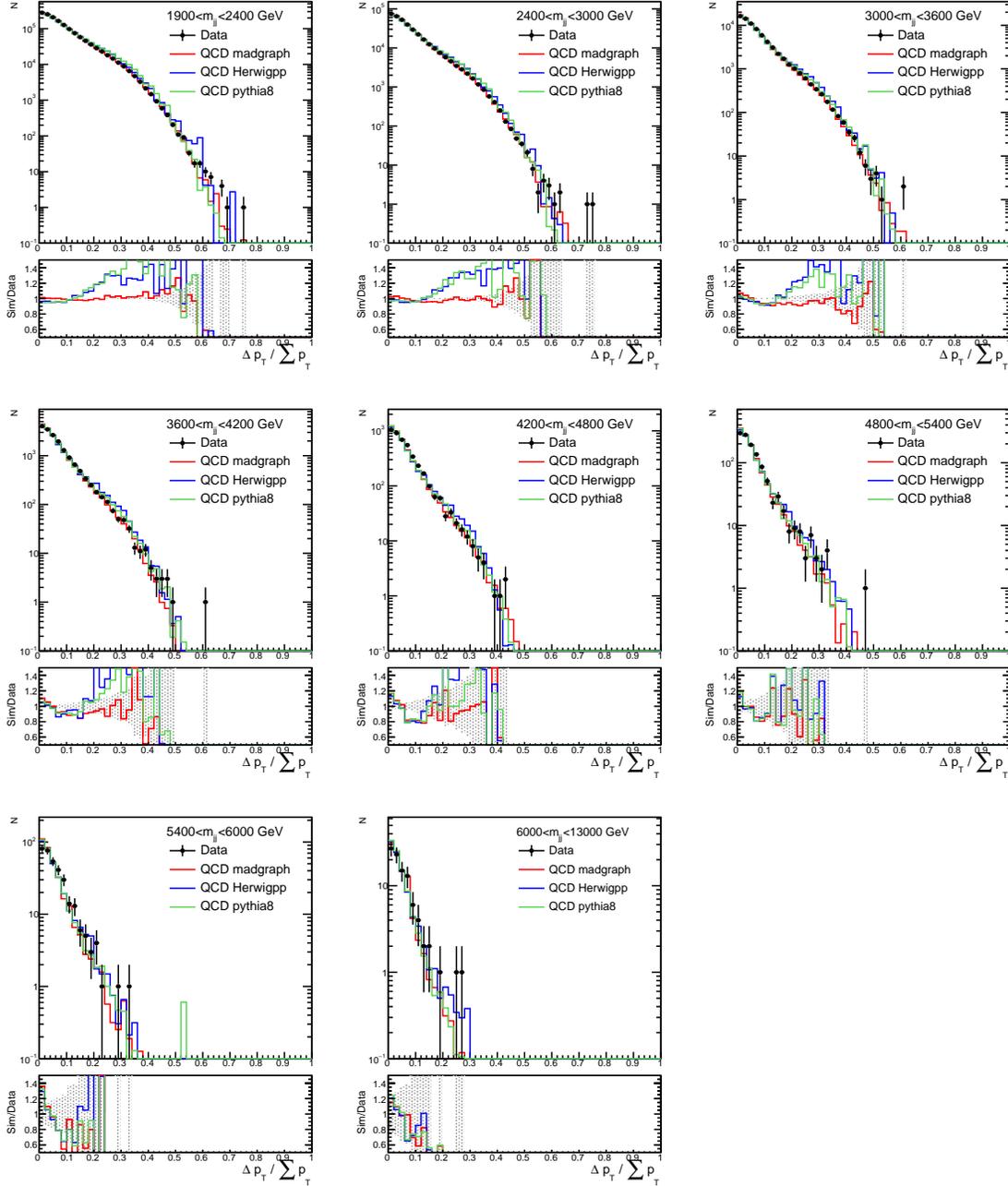


Figure 59: Comparison of $\frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}$ distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_aMC@NLO (red).

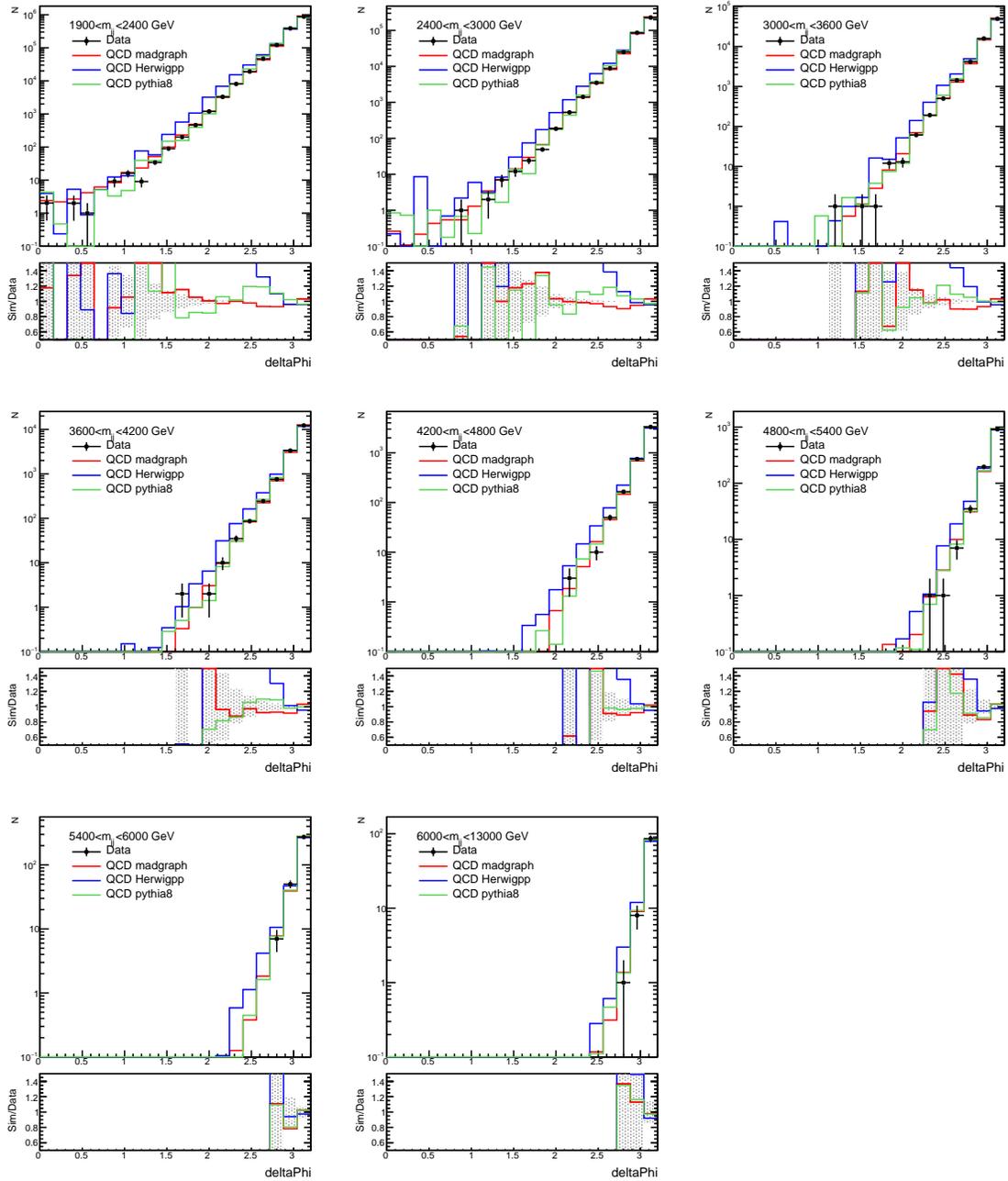


Figure 60: Comparison of $\Delta\phi_{1,2}$ distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_AMC@NLO (red).

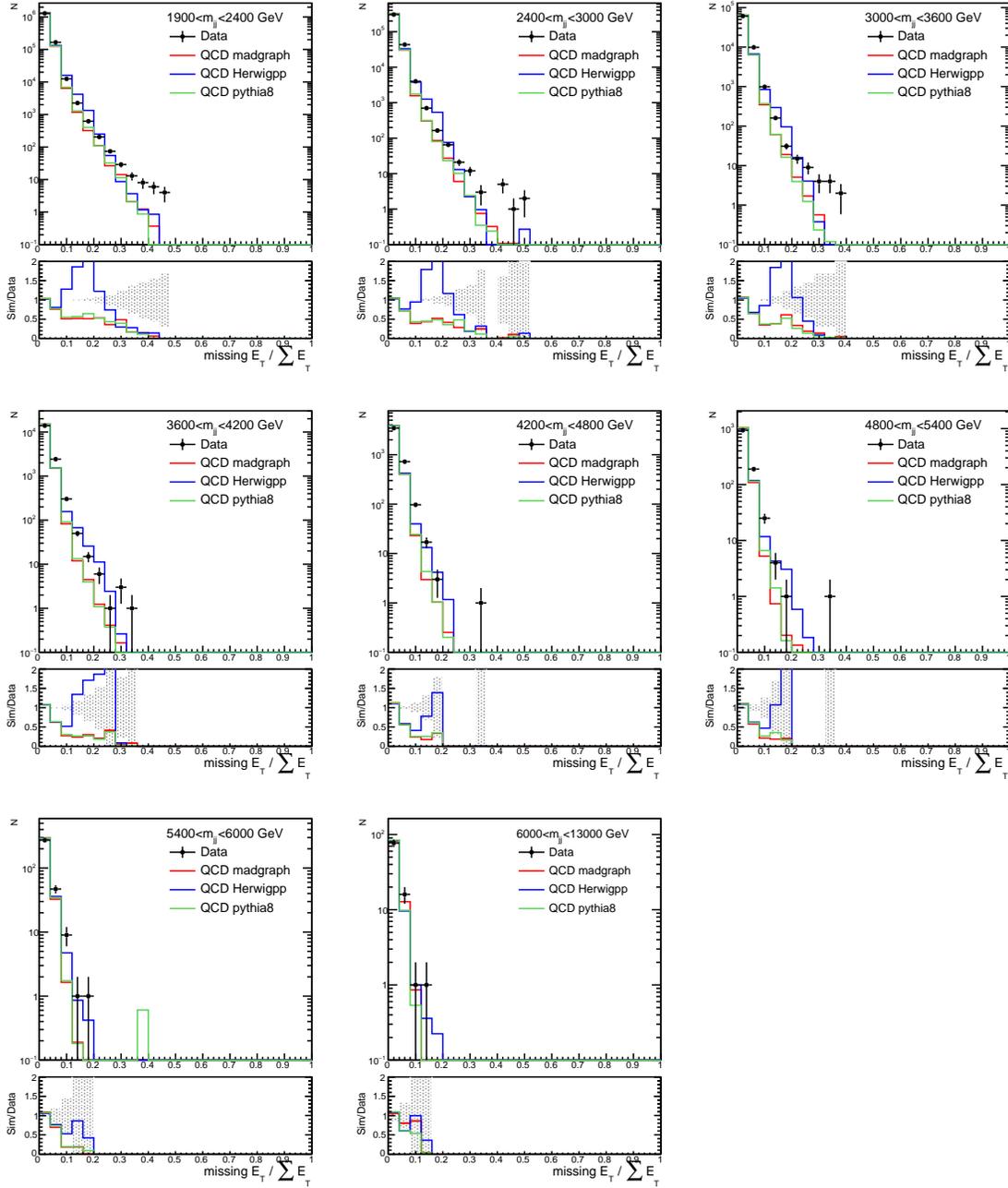


Figure 61: Comparison of $E_T^{\text{miss}} / \sum E_T$ distributions in data (points) and the Full Sim (histograms) from PYTHIA (green), HERWIG++ (blue), and MADGRAPH5_AMC@NLO (red).

6.3 Jet Energy Resolution Modeling

The largest detector effect on the measurement of the χ_{dijet} distributions comes from the JER, which arises from energy fluctuations due to electronic noise, pile-up, energy sampling, and non-uniformities and non-linearities in the detector response, and causes event migrations between bins of M_{jj} and χ_{dijet} . The jet angular resolution [106] is much smaller than the width of the χ_{dijet} bins, and therefore has a negligible effect on the measurement of the χ_{dijet} distributions.

The JER in various p_{T} and rapidity ranges determined using the Full Sim PYTHIA simulation are shown in Figures 62 to 67. Also shown in the figures are the distributions from Gaussian and double-sided Crystal Ball [109] parametrizations of the JER. The Crystal Ball function consists of a Gaussian core and a power-law low-end (high-end) tail, below (above) a certain threshold of the jet p_{T} response. The Gaussian parametrization describes the core of the resolution well, while the double-sided Crystal Ball parametrization describes both the core and the high-end part of the non-Gaussian tail. As the M_{jj} spectrum is steeply falling, the effect of events migrating from higher M_{jj} bins to lower M_{jj} bins due to JER is much smaller than those from lower M_{jj} bins to higher M_{jj} bins. Therefore, the modeling of the low-end tail is not critical for this measurement, though it is nevertheless accounted for in the systematic uncertainties. The Crystal Ball parametrization is finally adjusted to match the JER measured in data [103] (also see Section 5.2.4).

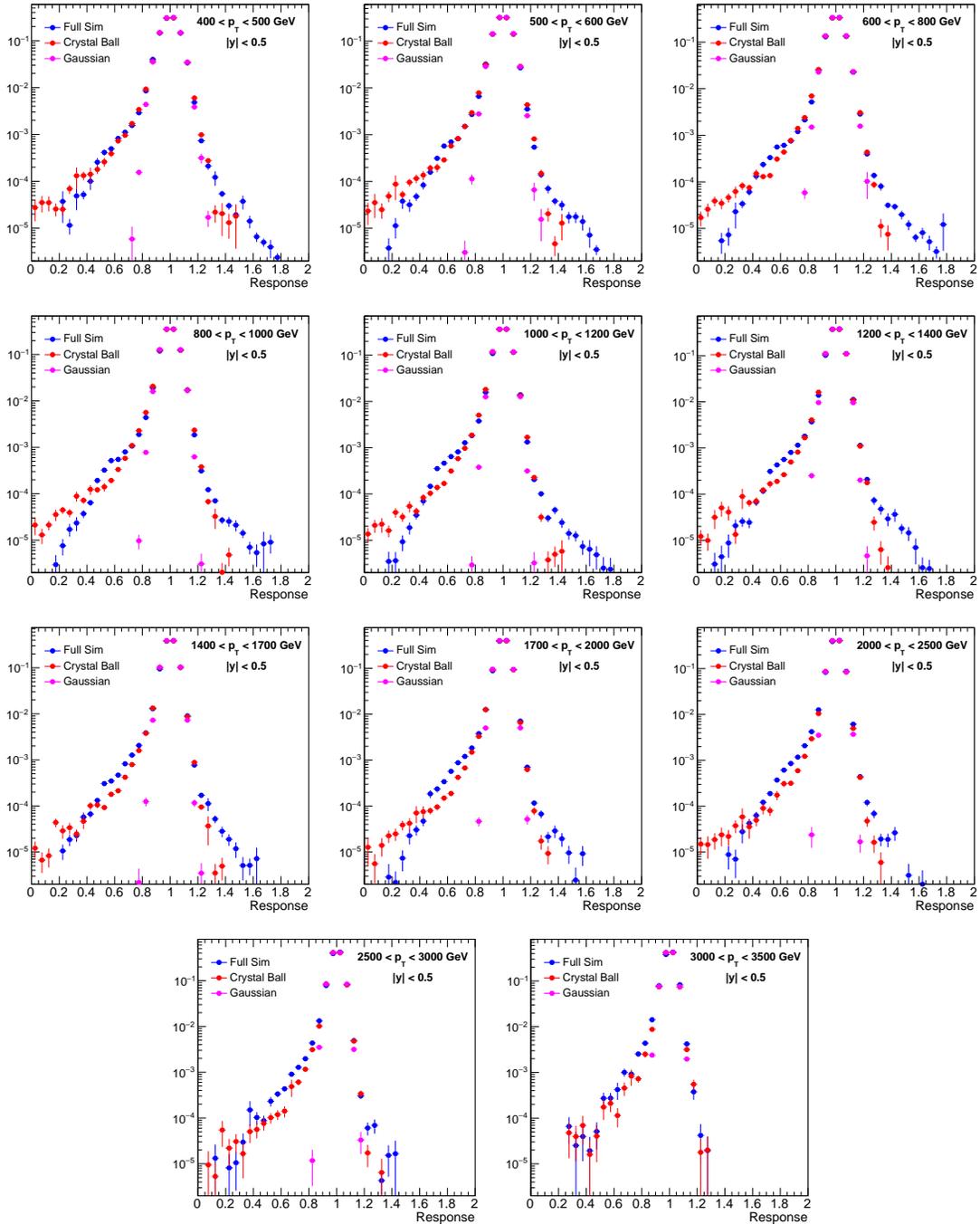


Figure 62: Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $|y| < 0.5$. The Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response are also shown.

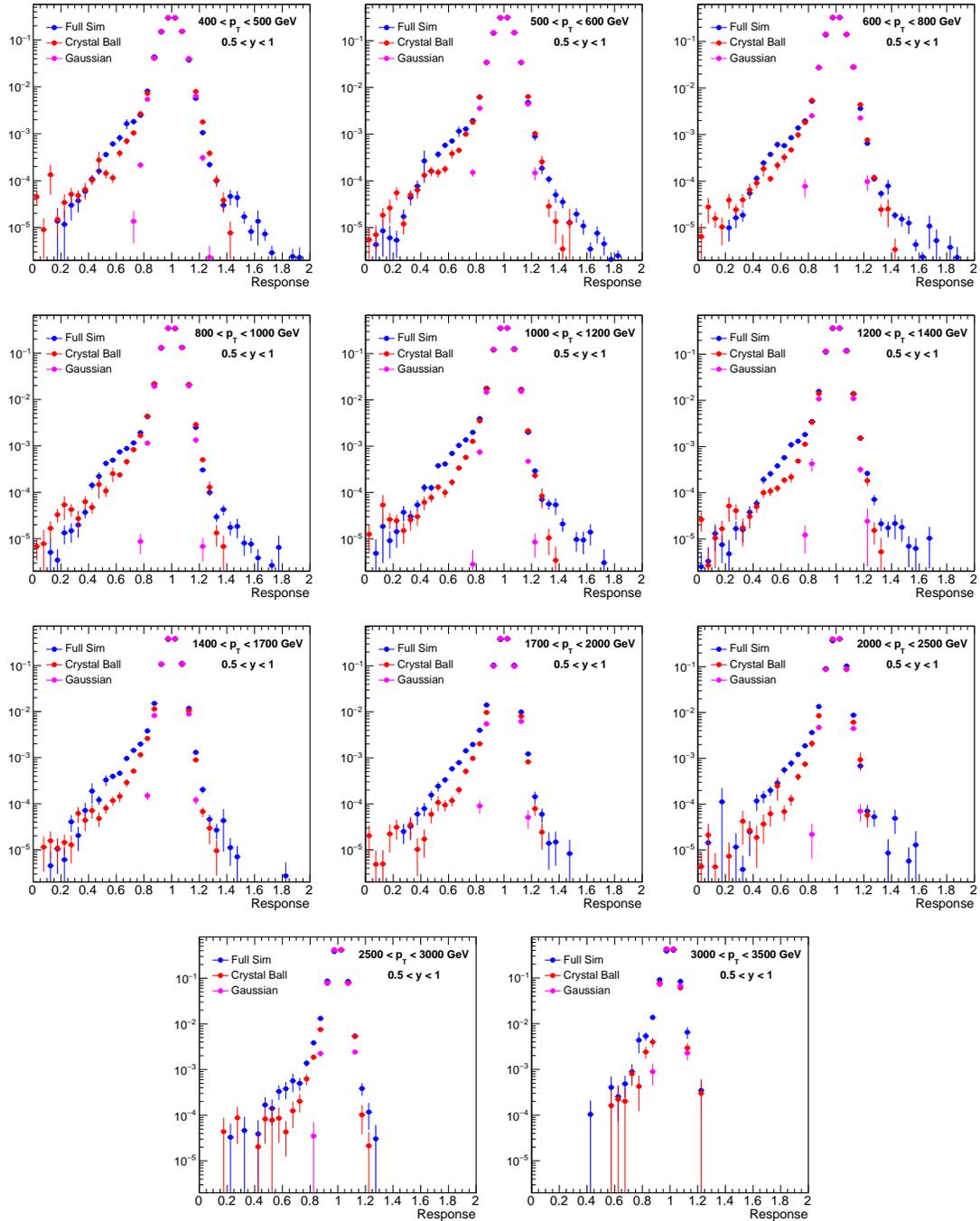


Figure 63: Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $0.5 < |y| < 1.0$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.

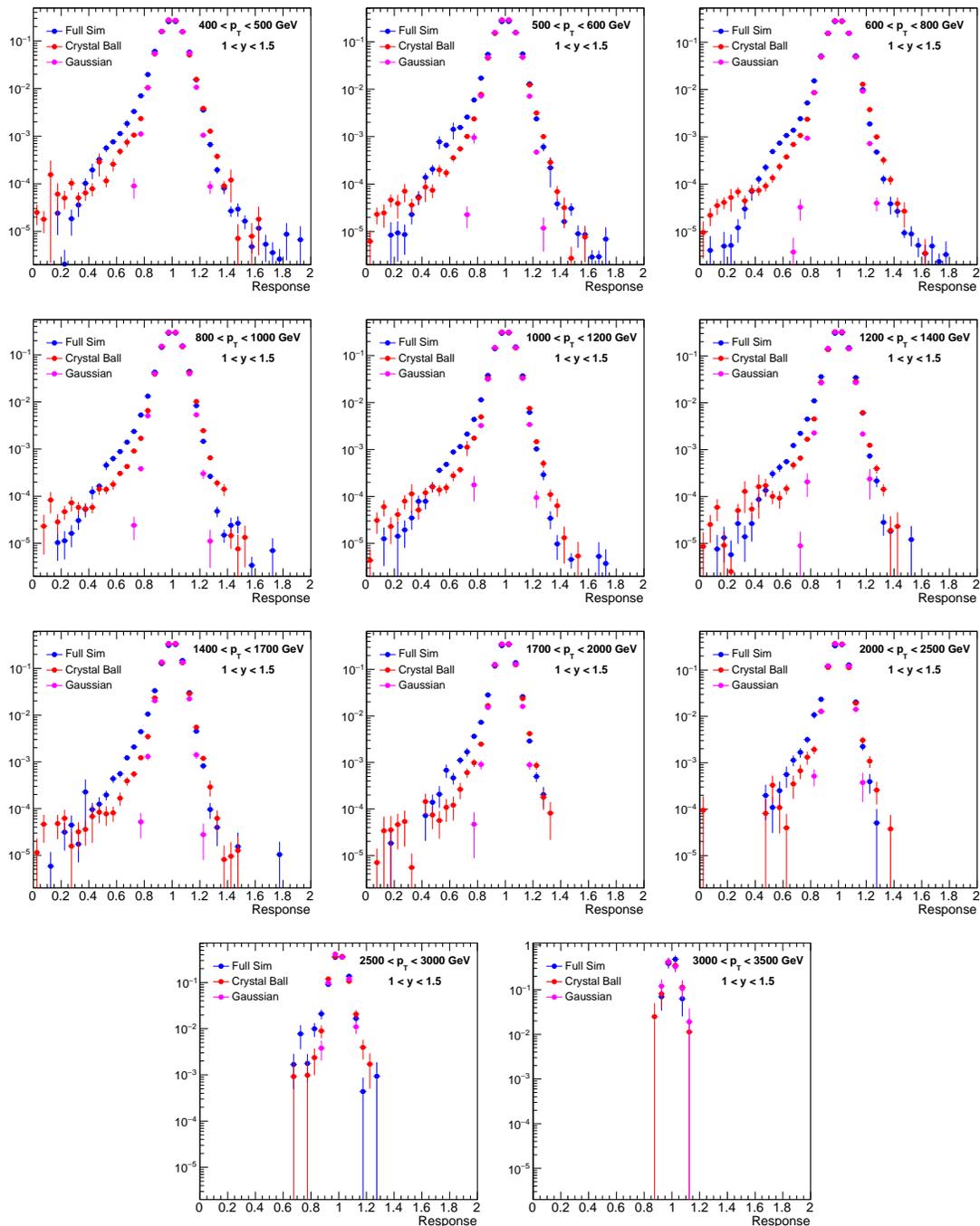


Figure 64: Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $1.0 < |y| < 1.5$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.

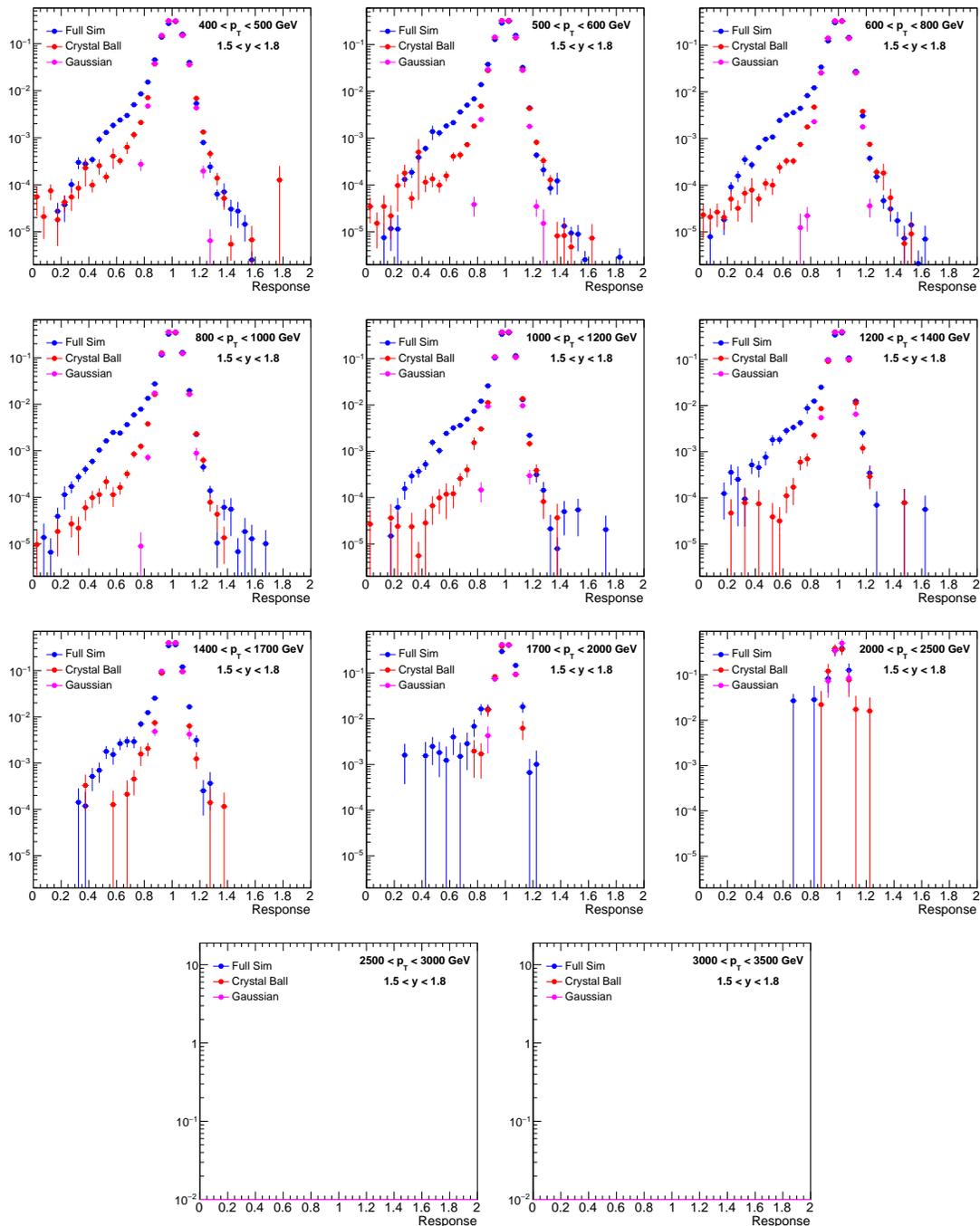


Figure 65: Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $1.5 < |y| < 1.8$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.

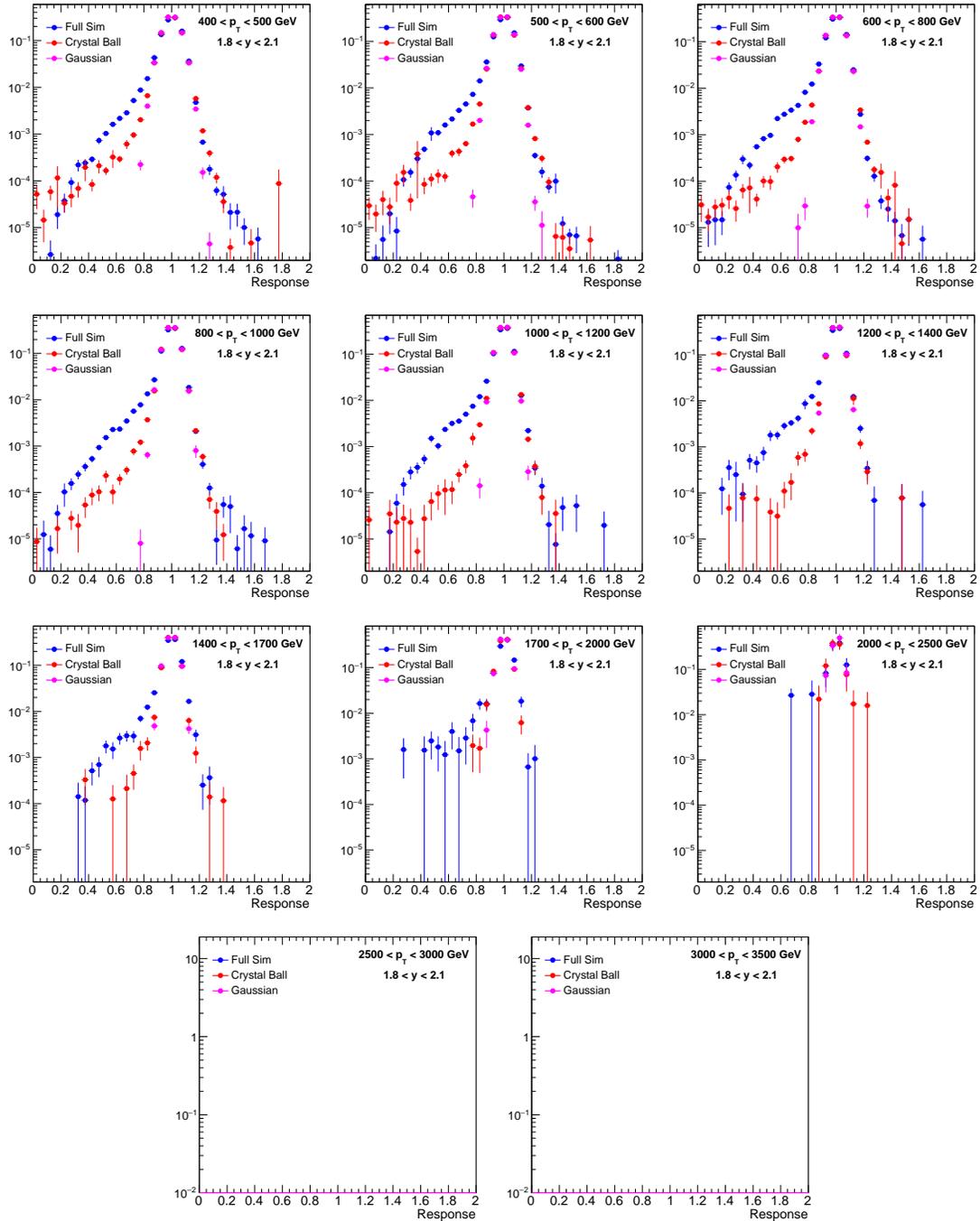


Figure 66: Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $1.8 < |y| < 2.1$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.

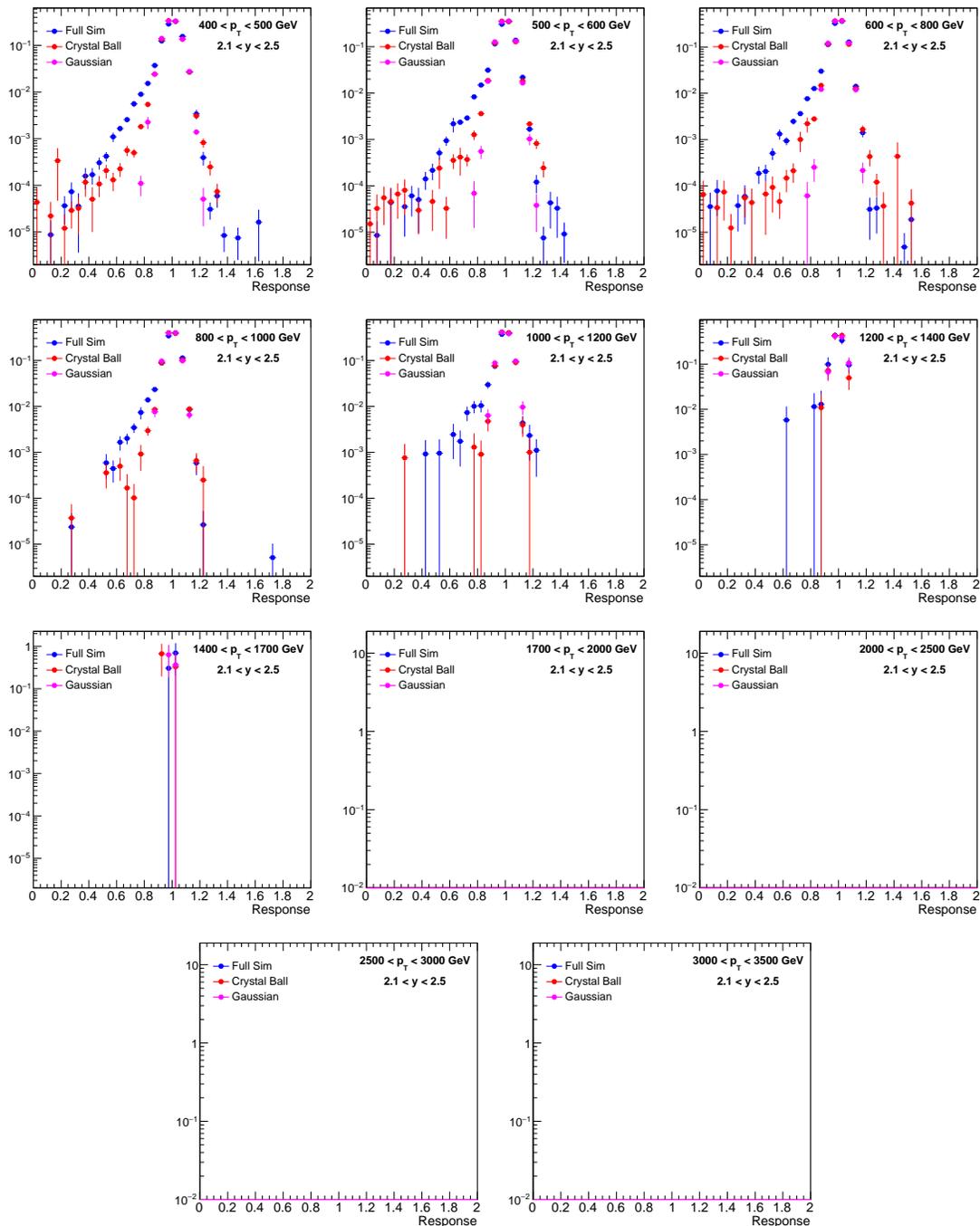


Figure 67: Jet p_T response from the Full Sim PYTHIA sample (blue) for jets with $2.1 < |y| < 2.5$. Also shown are the Gaussian (magenta) and double-sided Crystal Ball (red) parametrizations of the response.

6.4 Cross Section Unfolding

As previously mentioned, the finite JER can cause migrations between bins of M_{jj} and/or χ_{dijet} . It can cause event migrations in χ_{dijet} bins because lower energy jets can be misidentified as leading jets resulting in values of χ_{dijet} different from the true values. It can also cause event migrations in M_{jj} bins because fluctuations in the p_T of the leading jets can result in different values of M_{jj} . The migration effects are characterized using a 2D response matrix that maps the particle-level distributions¹ to detector-level ones. The 2D response matrix is binned in the variables M_{jj} and χ_{dijet} and is derived using particle-level jets that are generated using PYTHIA and converted to detector-level (smeared) using the double-sided Crystal-Ball parametrization of the JER. The response matrix is derived from smeared generator-level MC rather than from the Full Sim because the Full Sim is extremely CPU intensive and it is infeasible to generate enough events to sufficiently populate the 2D response matrix. The response matrix, integrated over χ_{dijet} is shown in Figure 68. In the figure, the detector-level (particle-level) M_{jj} are denoted as “reconstructed” (“generated”) on the y-axis (x-axis). In the response matrix, the migration effects between M_{jj} bins are expressed as

$$N_{\text{reconstructed},i} = \sum_j R_{ij} N_{\text{generated},j} \quad (6.1)$$

¹The distributions that include the detector effects are called “detector-level” or “raw” distributions, while the distributions that have been corrected for detector effects or come from simulations that do not include detector effects, are called “particle-level distributions”.

where i and j are the bin indices for detector-level and particle-level M_{ij} distributions, respectively, N is the number of events the corresponding bin, and R_{ij} is the response matrix. The migration effects between M_{ij} bins are 15% to 25%, depending on the M_{ij} values. The response matrix integrated over M_{ij} is not shown because the migration effects between χ_{dijet} bins are very small.

The data distributions are unfolded to particle level by simply applying the inverted 2D response matrix to the detector-level data distributions:

$$y_{\text{particle-level}} = R^{-1}_{2D} y_{\text{raw}}, \quad (6.2)$$

where y represents the data distributions. The unfolding is done using the software package ROOUNFOLD [110]. A problem that is often encountered in the unfolding procedure is that the inverse of the response matrix is ill-conditioned, meaning that the unfolded distributions are very sensitive to small changes in the input distributions. For example, if the inverse of the response matrix is ill-conditioned or even singular, small statistical fluctuations in the detector-level distributions can be amplified and exhibit high-frequency oscillations in the unfolded distributions. In such cases, the unfolding procedure needs to be regularized. The degree to which the response matrix is ill-conditioned can be inferred from its condition number, which is defined as the ratio of the largest and smallest singular values obtained by the singular value decomposition of the matrix. Typically, a condition number less than 10 implies that the response matrix is a well-conditioned matrix and the unfolding procedure does

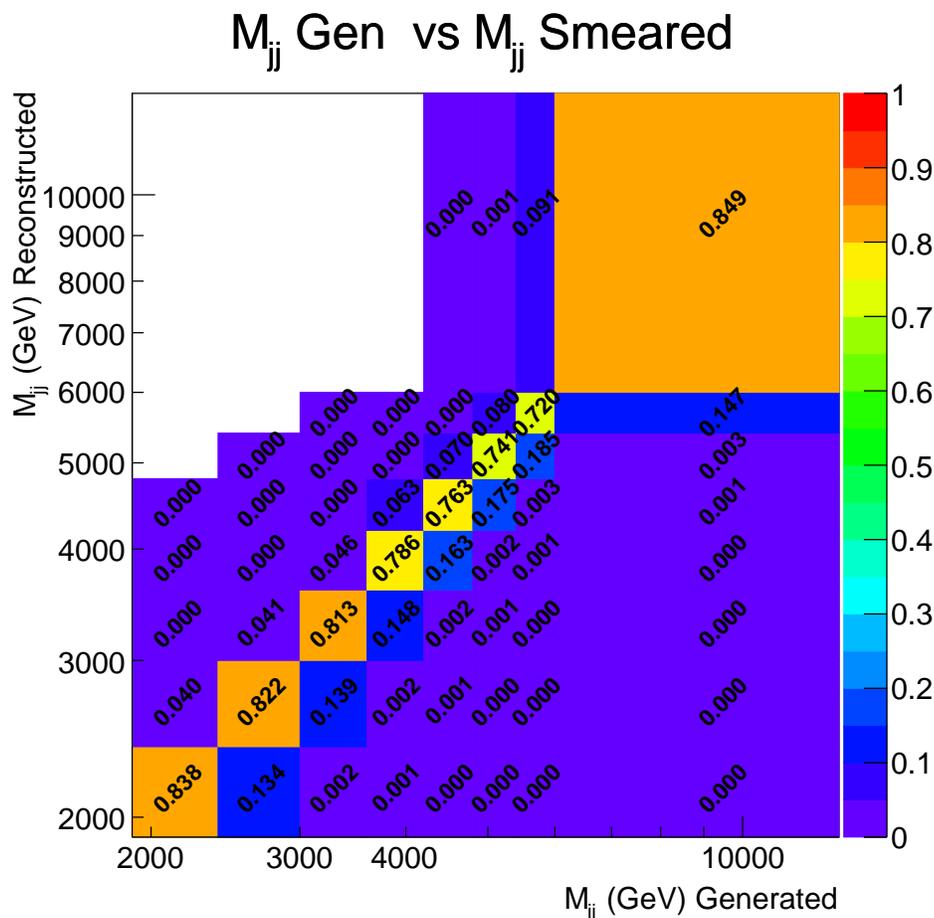


Figure 68: The response matrix, integrated over χ_{dijet} . The numbers in the boxes correspond to the values for R_{ij} in Equation 6.1.

not need regularization. The condition number for the response matrix used in this thesis is 2.89. As a cross check, the unfolded distributions from the simple matrix inversion method without regularization is compared to the distributions obtained from the D'Agostini iterative method [111] which is regularized by the number of iterations, n_{iter} . With $n_{\text{iter}} = 4$, which is

the default regularization strength for the D'Agostini method, the two results are found to be consistent to within 5%.

A trivial closure test is performed by unfolding the χ_{dijet} distributions from the same sample that is used to derive the response matrix. Perfect closure is obtained as shown in Figure 69. Another closure test is performed by splitting the simulated samples into independent training and testing samples. An unfolding matrix is determined from the training sample and then used to unfold the χ_{dijet} distributions from the testing sample. The result of this closure test is shown in Figure 70 and no significant discrepancies are observed in the results.

A comparison between the raw and unfolded data distributions is shown in Figure 71. The unfolding correction changes the shape of the χ_{dijet} distributions by less than 1% across χ_{dijet} in the lowest M_{jj} range and less than 8% in the highest M_{jj} range.

The ROOUNFOLD package provides the covariance matrix of the unfolded data distributions. This covariance matrix, given by the simple matrix inversion unfolding procedure, is expressed as

$$V_{\text{unfold}} = R_{2\text{D}}^{-1} V_{\text{raw}} (R_{2\text{D}}^{-1})^{\text{T}}, \quad (6.3)$$

where the covariance matrix for the raw distribution V_{raw} is a diagonal matrix where each diagonal element equals the variance of each bin in the raw distribution. The covariance matrices for M_{jj} bins in various χ_{dijet} ranges are shown in Figure 72. As shown in the covariance matrices, the off-diagonal elements are not zero, indicating that the unfolding procedure introduces

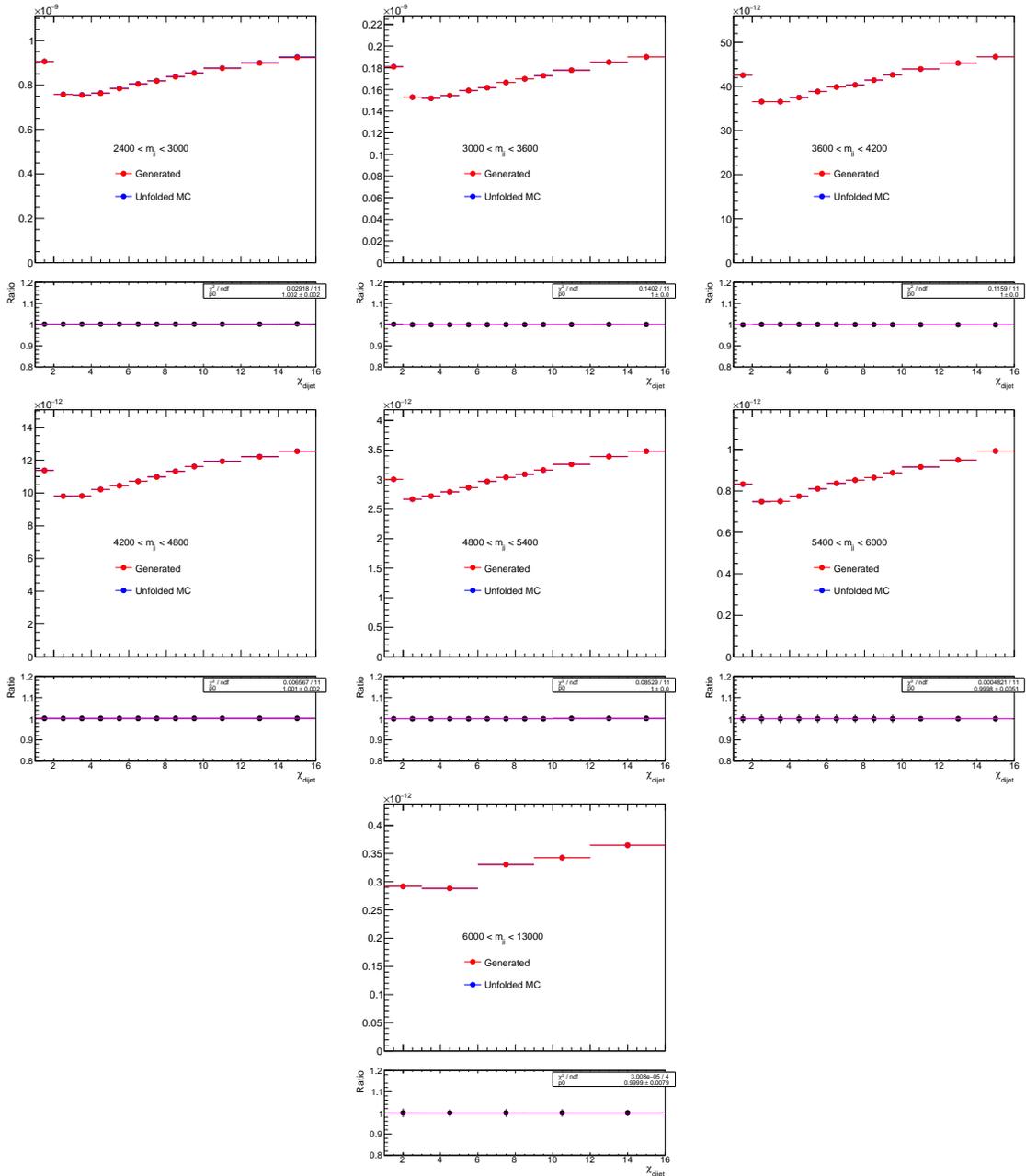


Figure 69: Comparisons of unfolded and generated χ_{dijet} distributions, where the unfolding correction was determined from the same MC sample. Also shown are the ratios of unfolded to generated χ_{dijet} distributions.

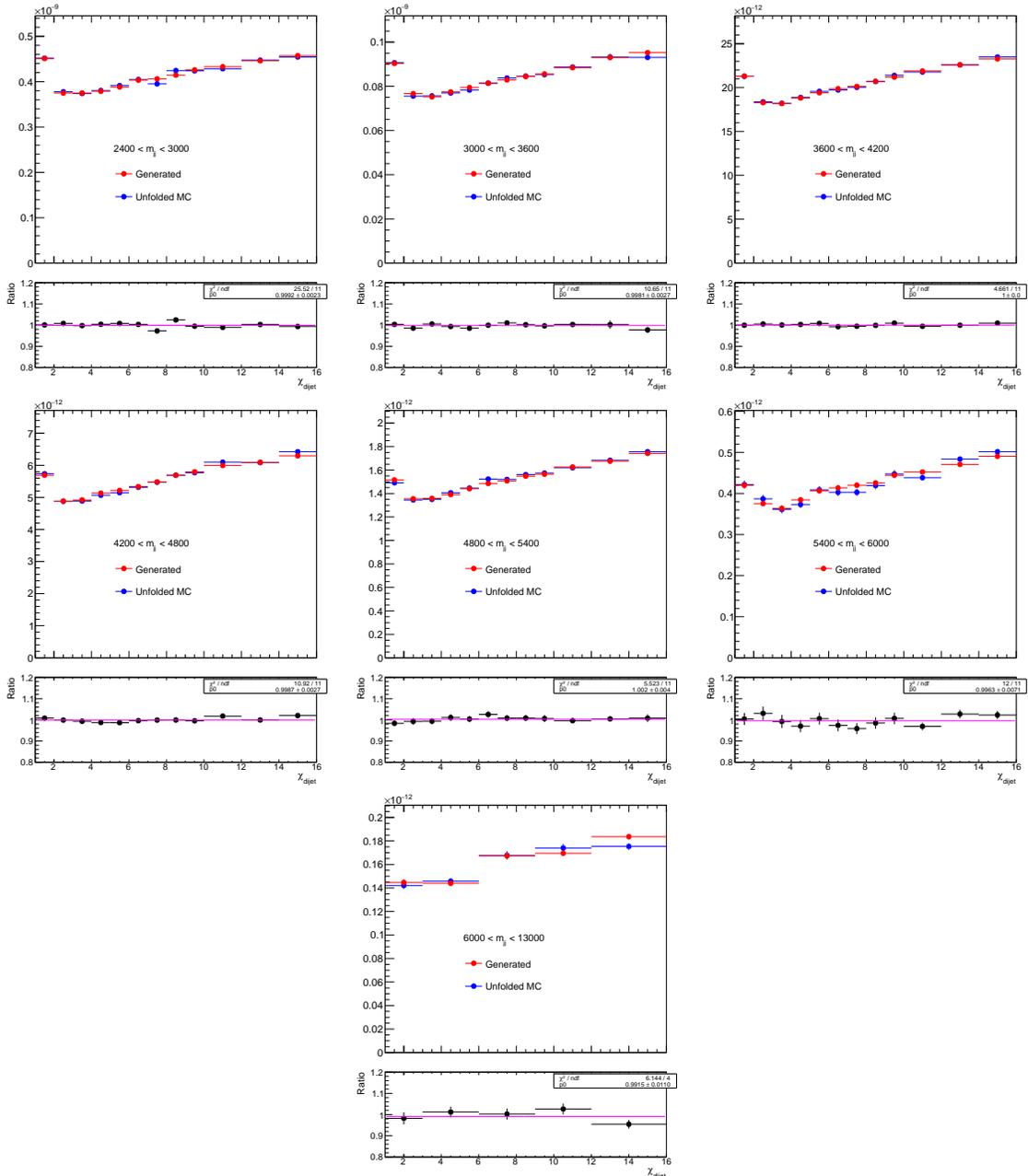


Figure 70: Comparisons of unfolded and generated χ_{dijet} distributions, where the unfolding correction was determined from an independent sample of events. Also shown are the ratios of unfolded to generated χ_{dijet} distributions.

correlations between M_{jj} bins. The correlations are characterized by the correlation coefficients of the covariance matrix. The correlation coefficients are defined as:

$$\rho = \frac{V(i, j)}{\sigma(i)\sigma(j)} = \frac{V(i, j)}{\sqrt{V(i, i)}\sqrt{V(j, j)}}, \quad (6.4)$$

where i and j are the bin indices and σ is the variance for each of the bins in the distributions. The correlation coefficients between M_{jj} bins are shown in Figure 73. As shown in the figure, the correlations between M_{jj} bins are 15–20% in the lowest M_{jj} bin and 25–40% in the highest M_{jj} bin, depending on the χ_{dijet} values. The correlations between χ_{dijet} bins are negligible.

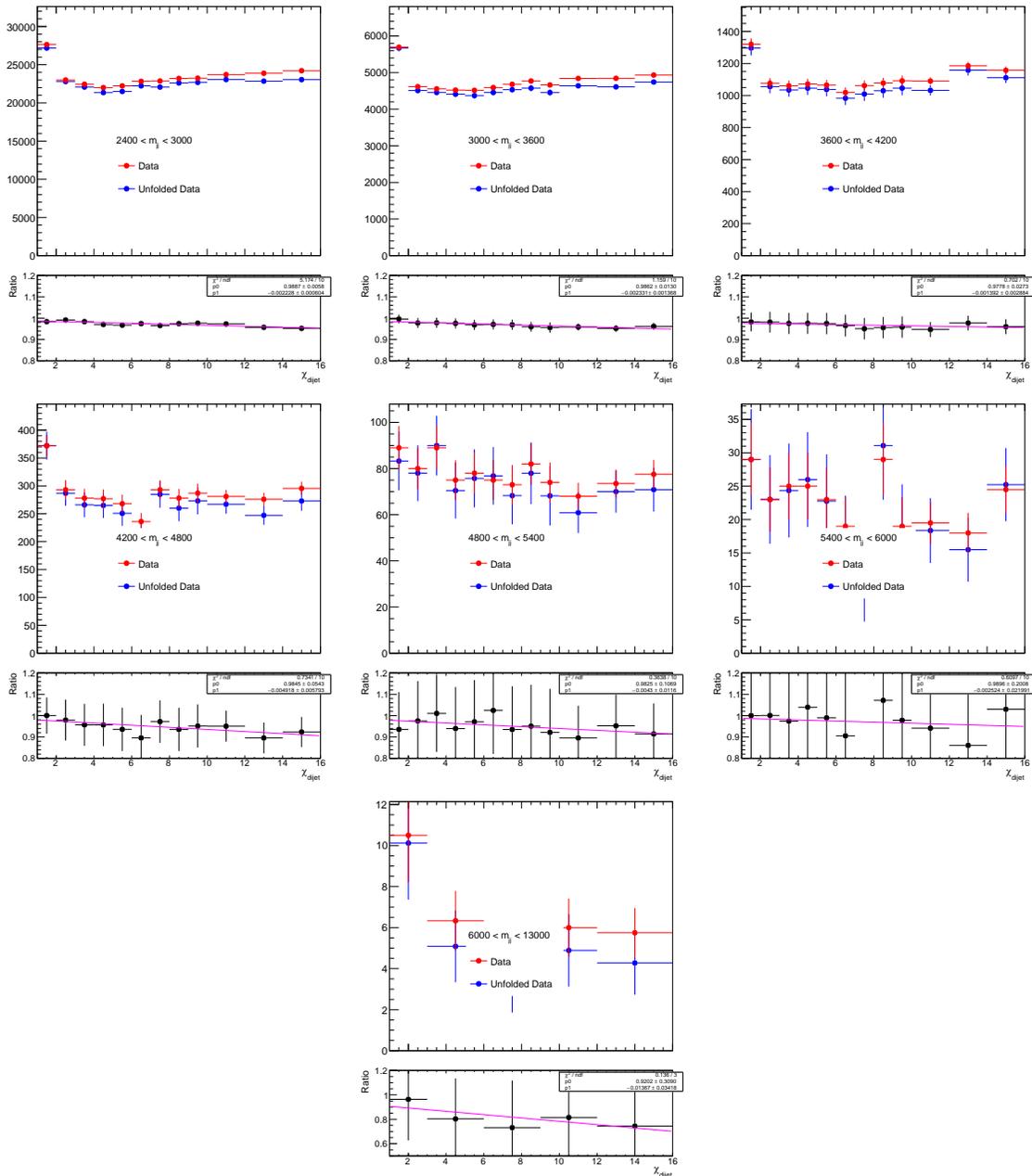
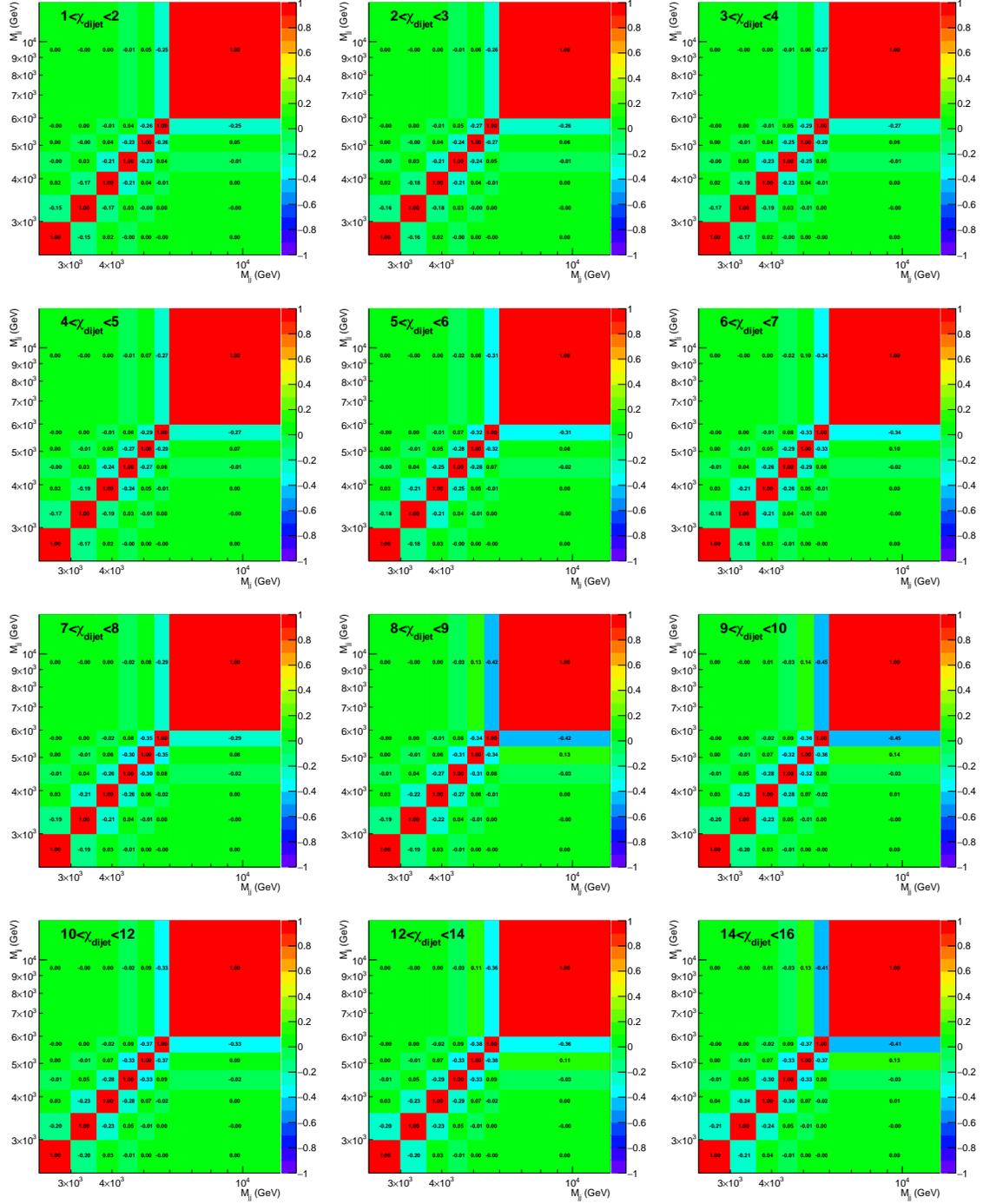


Figure 71: Raw and unfolded χ_{dijet} distributions from data. Also shown are the ratios of the unfolded to raw χ_{dijet} distributions.

Figure 72: Covariance matrices for the M_{jj} bins.

Figure 73: Correlation coefficients for the M_{jj} bins.

6.5 Experimental Uncertainties

The dijet angular distributions are normalized to unity in each M_{jj} range, denoted

$$(1/\sigma_{\text{dijet}})(d\sigma_{\text{dijet}}/d\chi_{\text{dijet}}), \quad (6.5)$$

where σ_{dijet} is the dijet cross section in the M_{jj} range. An advantage to using normalized distributions is that they are relatively insensitive to many systematic effects. In particular, they are not sensitive to uncertainties that affect the overall normalization of the distributions such as the luminosity uncertainty. Instead, the largest systematic uncertainties come from detector effects that change the shape of the distributions. The main systematic uncertainties in this analysis come from the jet energy scale, the jet energy resolution, and the unfolding correction. The effects of these uncertainties on the dijet angular distributions are discussed below.

6.5.1 Jet Energy Scale Uncertainty

JES uncertainties are derived for each of the factorized corrections in the JES calibration process and are provided in the form of mutually uncorrelated sources as functions of jet p_T and η . In the p_T - η space covered by this analysis, the overall JES uncertainty is found to be less than 1% [103, 104]. The variation of the JES as a function of η is less than 1% per unit η . To calculate the effects of the JES uncertainties on the χ_{dijet} distributions, the p_T of the jets are varied according to the p_T and η dependent uncertainties of the various JES uncertainty sources and the normalized χ_{dijet} distributions are evaluated for all these variations. Ratios

between the normalized χ_{dijet} distributions obtained using the nominal JES corrections and the ones derived with increased (or decreased) jet p_T 's are shown in Figures 74 to 77. The JES uncertainty sources are taken into account independently in the NP limit setting procedure since this accounts for the correlations among the M_{jj} and χ_{dijet} bins for each JES uncertainty source. However, for the display of errors in the figures and tables, the total JES uncertainty is obtained from the quadratic sum over all of these uncertainty sources and is found to be 3.6% in the lowest M_{jj} range and 9.2% in the highest M_{jj} range.

6.5.2 Jet Energy Resolution Uncertainty

The main sources of the uncertainty on the JER come from the width of the JER core and the size of the JER tails. The uncertainty in the width of the core is found to range from 1% to 5%, depending on η [103, 104]. To evaluate the effect of the uncertainty in the core on the χ_{dijet} distributions, the generated jet energies are smeared using the Gaussian parametrization of the JER with width given by the nominal JER. Then, additional jet collections are obtained by smearing the generated jets using the same Gaussian parametrization with widths varied from the nominal values by the JER uncertainties. The uncertainty is obtained from the ratios of the normalized χ_{dijet} distributions obtained with the JER shifted widths in the Gaussian width to those obtained with the widths given by the nominal JER. These ratios (before normalization) are shown in Figure 78. This uncertainty on the χ_{dijet} distributions is found to be less than 1% for all M_{jj} .

The size of tails in the jet resolution has been constrained from data only to a limited extent [112]. Therefore, to evaluate the uncertainty due to the uncertainty in the tails, χ_{dijet}

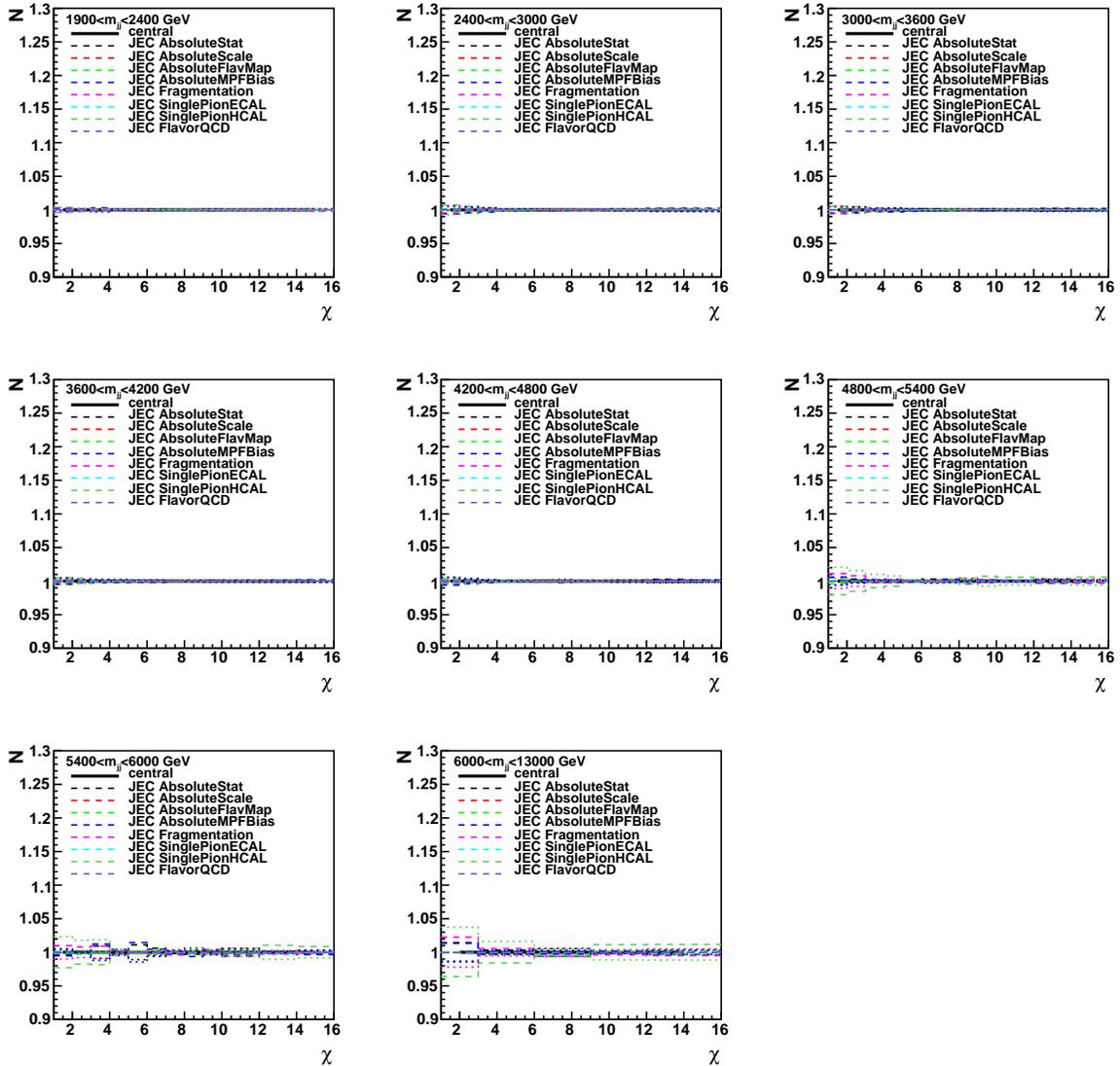


Figure 74: The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.

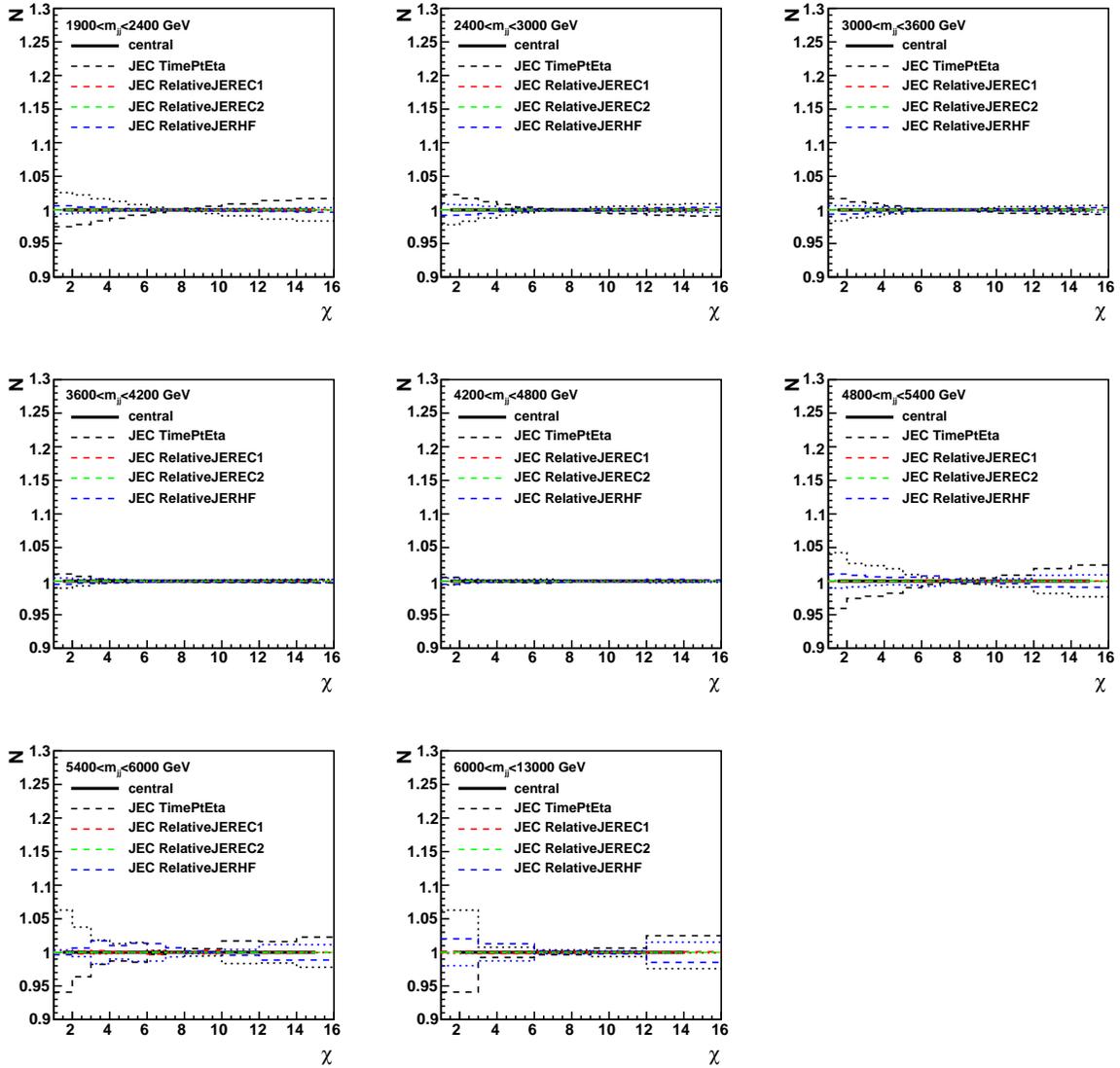


Figure 75: The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.

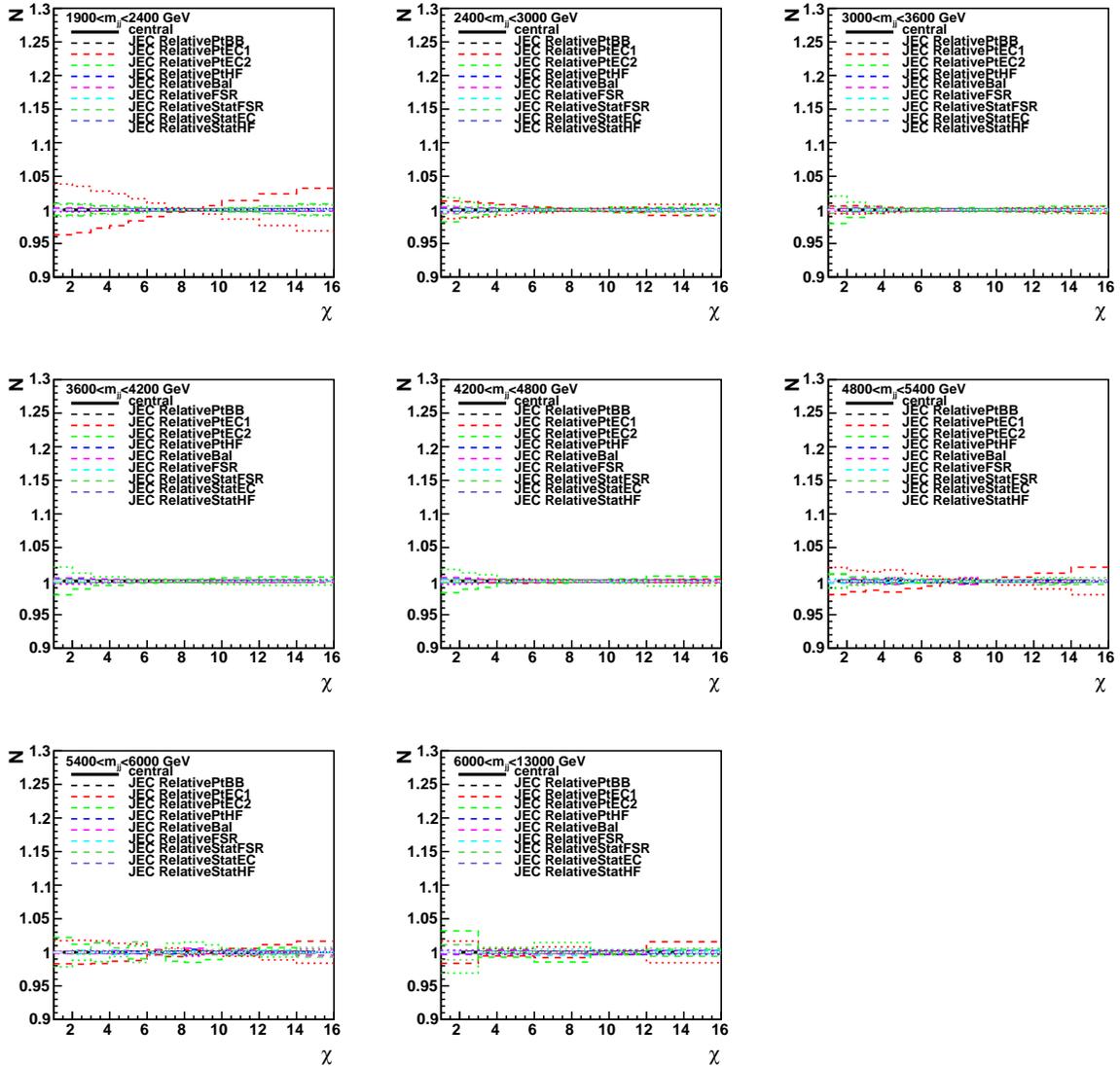


Figure 76: The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.

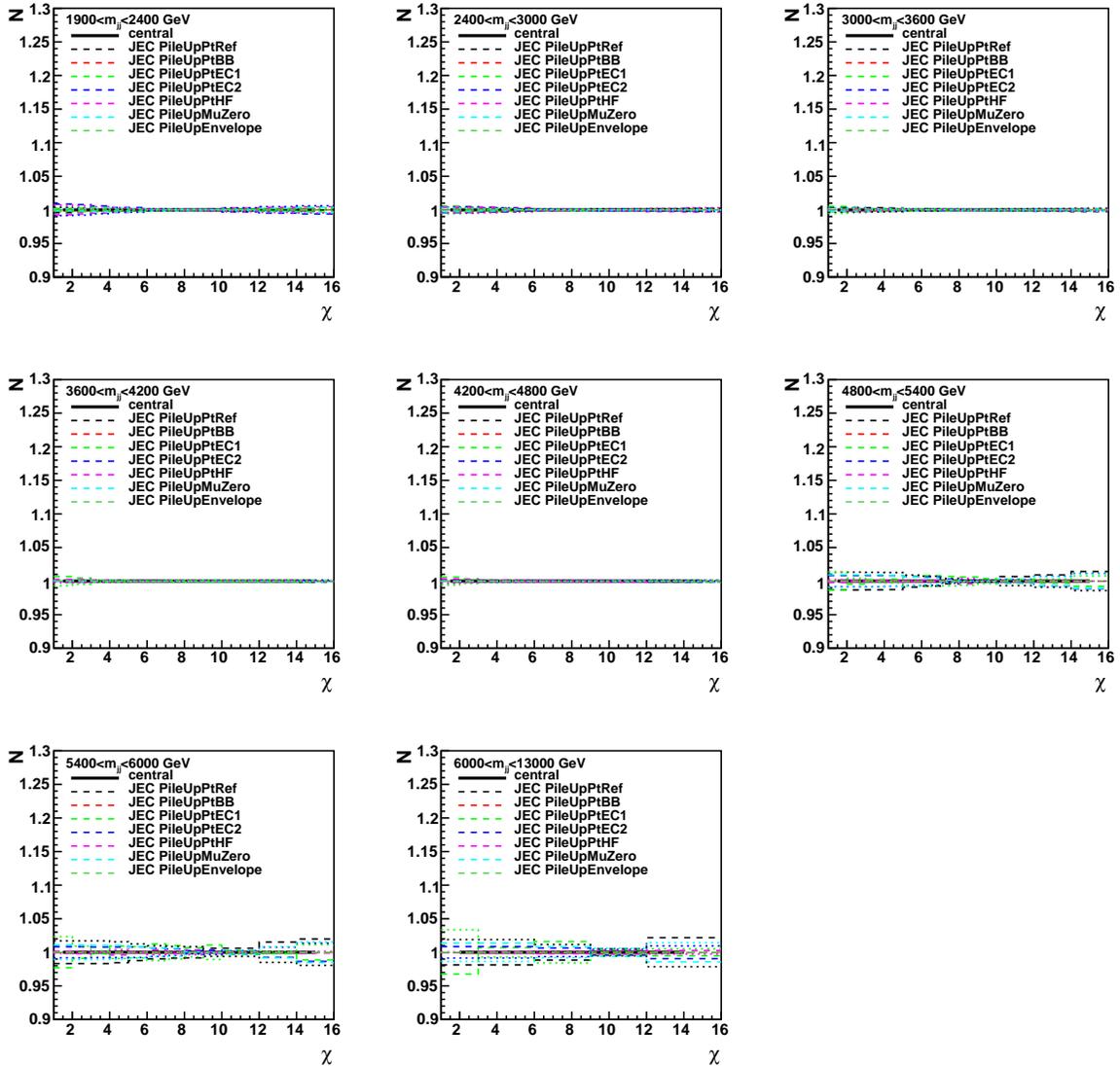


Figure 77: The effect of various JES uncertainty sources on the normalized χ_{dijet} distributions.

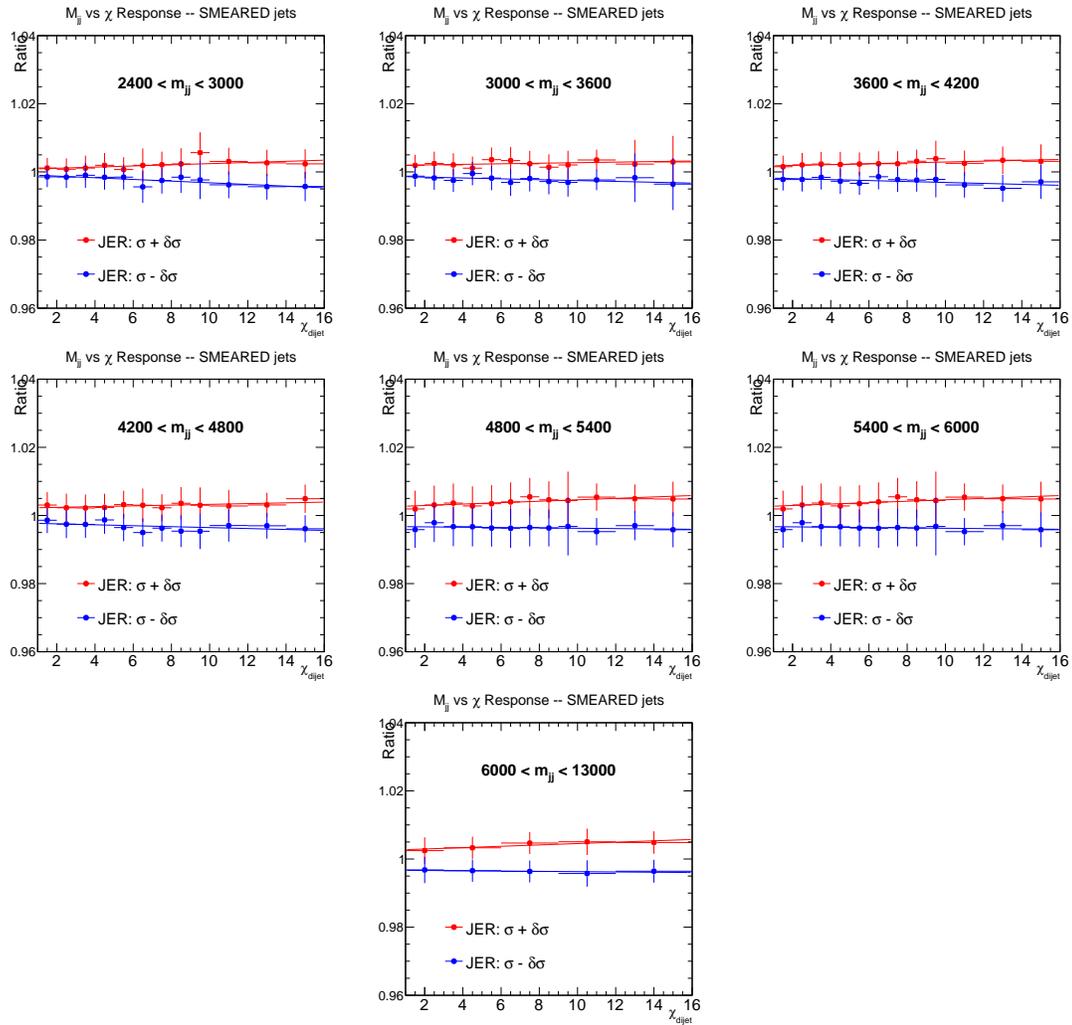


Figure 78: The effect of the uncertainty in the width of the core of the JER on the χ_{dijet} distributions.

distributions obtained using the double-sided Crystal Ball parametrization of the JER are compared to the χ_{dijet} distributions obtained using the Gaussian parametrization. These comparisons are shown in Figure 79. 50% of this difference on the normalized χ_{dijet} distributions is assigned as the uncertainty due to the JER tails. The size of this uncertainty is less than 1.5% for all M_{jj} .

6.5.3 Uncertainty Due to the Modeling of the Jet Energy Resolution

Another source of systematic uncertainty that was evaluated is the use of a parametric model to simulate the jet p_{T} resolution of the detector. The ratio between the normalized χ_{dijet} distributions obtained using smeared jets and Full Sim jets is considered as a systematic uncertainty. These ratios, along with the corresponding χ_{dijet} distributions are shown in Figure 80. The uncertainty is found to be 0.5% and 1% in the lowest and highest M_{jj} ranges, respectively.

6.5.4 Uncertainty Introduced Due to the Unfolding Procedure

The potential mismodeling of the dijet kinematic distributions by the PYTHIA event generator was considered as another source of systematic uncertainty. In practice, a mismodeling of the M_{jj} distribution can affect the size of the unfolding correction. To estimate the size of this uncertainty, distributions from MADGRAPH5_aMC@NLO generator were used, as the M_{jj} distribution from MADGRAPH5_aMC@NLO is found to be considerably different from the M_{jj} distribution obtained from PYTHIA (see Section 6.2). The inverted response matrix derived using the PYTHIA sample is applied to the χ_{dijet} distributions obtained from the smeared MADGRAPH5_aMC@NLO sample and the results are compared to the corresponding MADGRAPH5_aMC@NLO particle-level χ_{dijet} distributions. The unfolded and generated distribu-

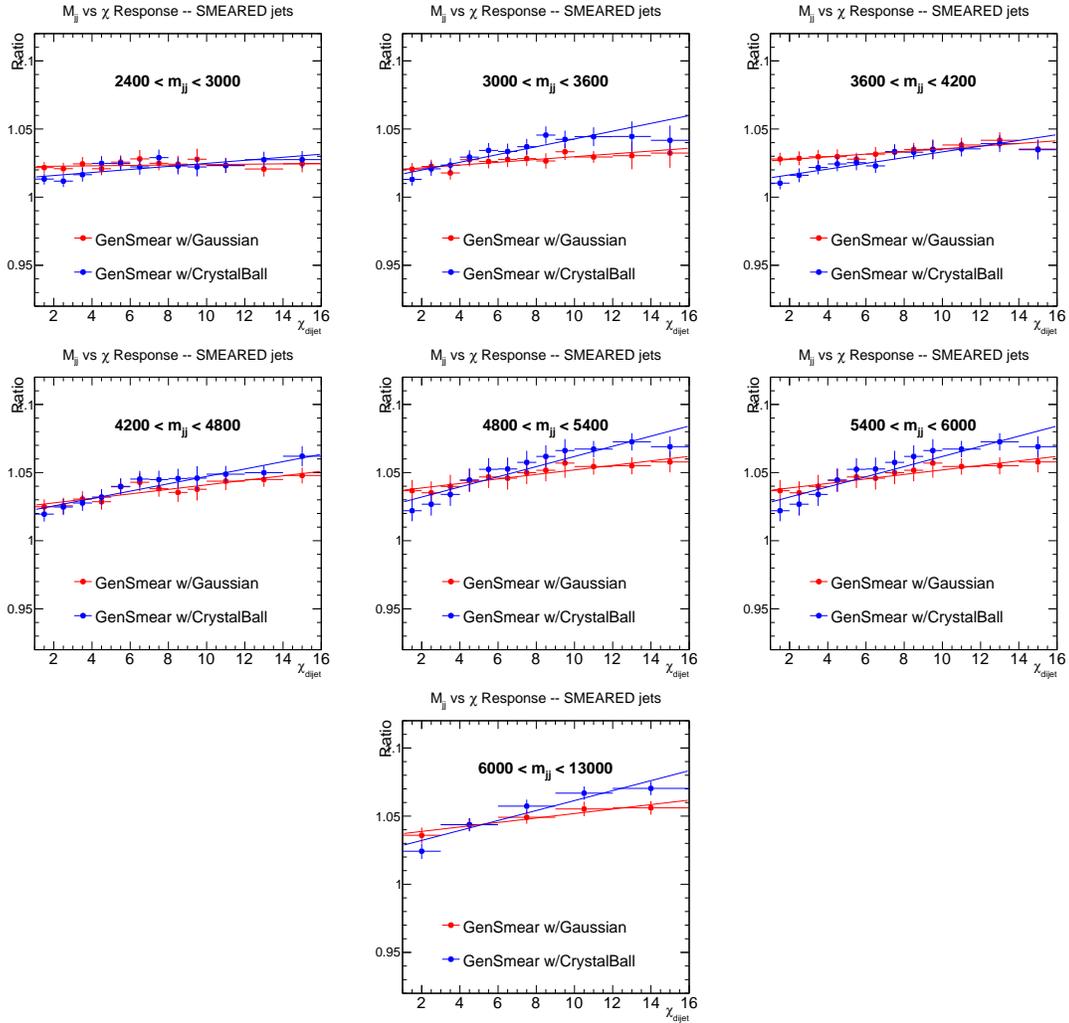


Figure 79: χ_{dijet} distributions obtained using the double-sided Crystal Ball parametrization of the JER (blue) compared to those obtained using the Gaussian parametrization (red).

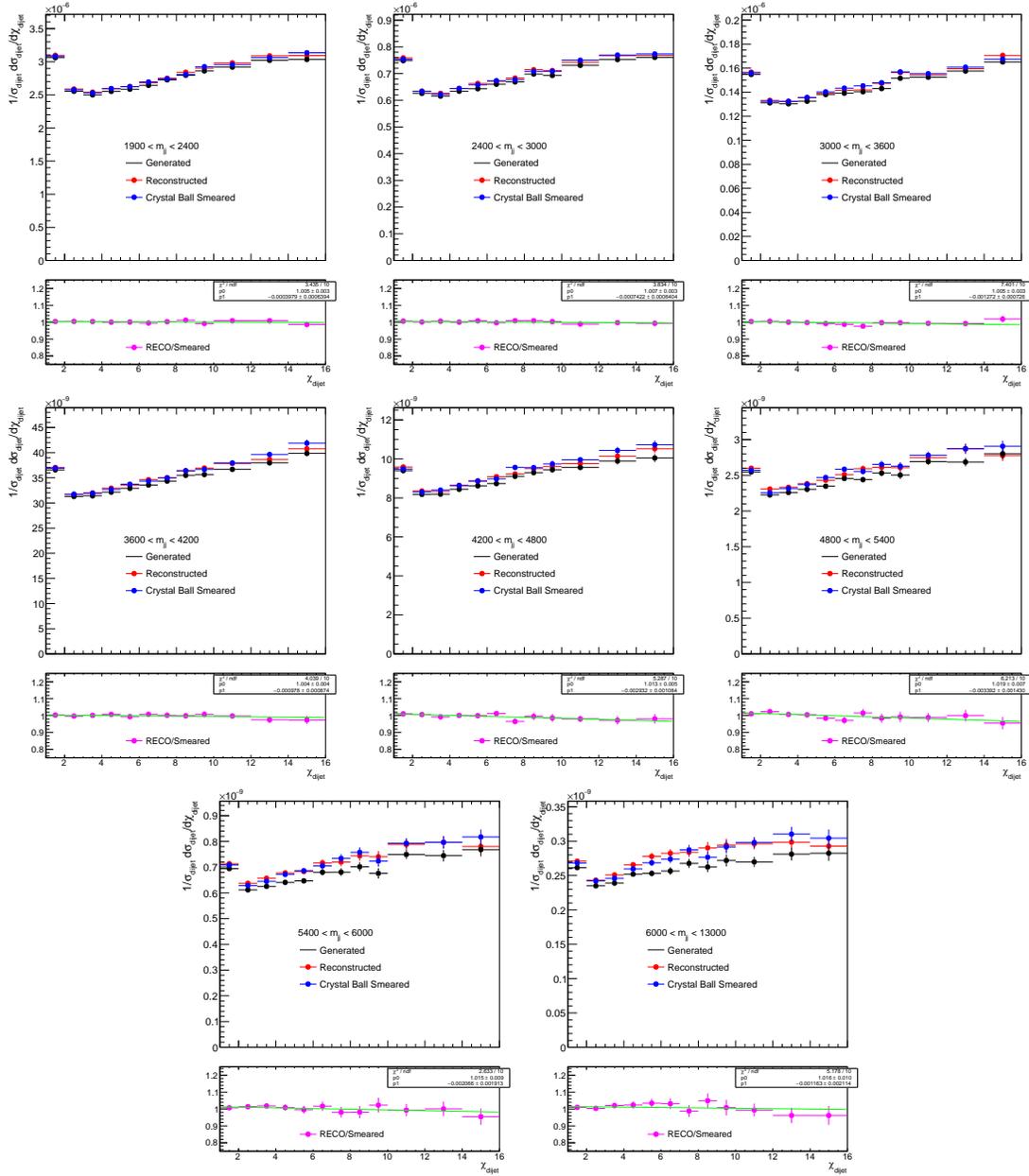


Figure 80: χ_{dijet} distributions obtained from generated jets (black) smeared with the double-sided Crystal Ball parametrization of the JER (blue) compared to those obtained from the Full Sim (red).

tions are shown in Figure 81. The effect on the resultant normalized χ_{dijet} distributions is found to be less than 1.5% for all M_{jj} .

A summary of the experimental uncertainties is shown in Table VII. While in the statistical analysis (see Chapter 8), each systematic uncertainty is represented through a change in the χ_{dijet} distributions correlated across all the M_{jj} and χ_{dijet} bins, representative values, which are taken as the relative shifts for the lowest χ_{dijet} bins in the lowest and highest M_{jj} bins, are used to illustrate the relative contributions in this table. As implied in the table, the leading experimental uncertainty is JES uncertainty for low M_{jj} bins and statistical uncertainty for high M_{jj} bins.

TABLE VII: The leading experimental uncertainties in the normalized χ_{dijet} distributions, in percent. While in the statistical analysis (see Chapter 8), each systematic uncertainty is represented through a change in the χ_{dijet} distributions correlated across all the M_{jj} and χ_{dijet} bins, representative values which are taken as the relative shifts for the lowest χ_{dijet} bins in the lowest and highest M_{jj} bins, are used to illustrate the relative contributions in this table.

Source of uncertainty	$2.4 < M_{\text{jj}} < 3.0 \text{ TeV}$	$M_{\text{jj}} > 6.0 \text{ TeV}$
Statistical	0.7	27
JES	3.6	9.2
Jet p_{T} resolution (core)	1.0	1.0
Jet p_{T} resolution (tails)	1.0	1.5
Detector response model	0.5	1.0
Unfolding, model dependence	0.2	1.5

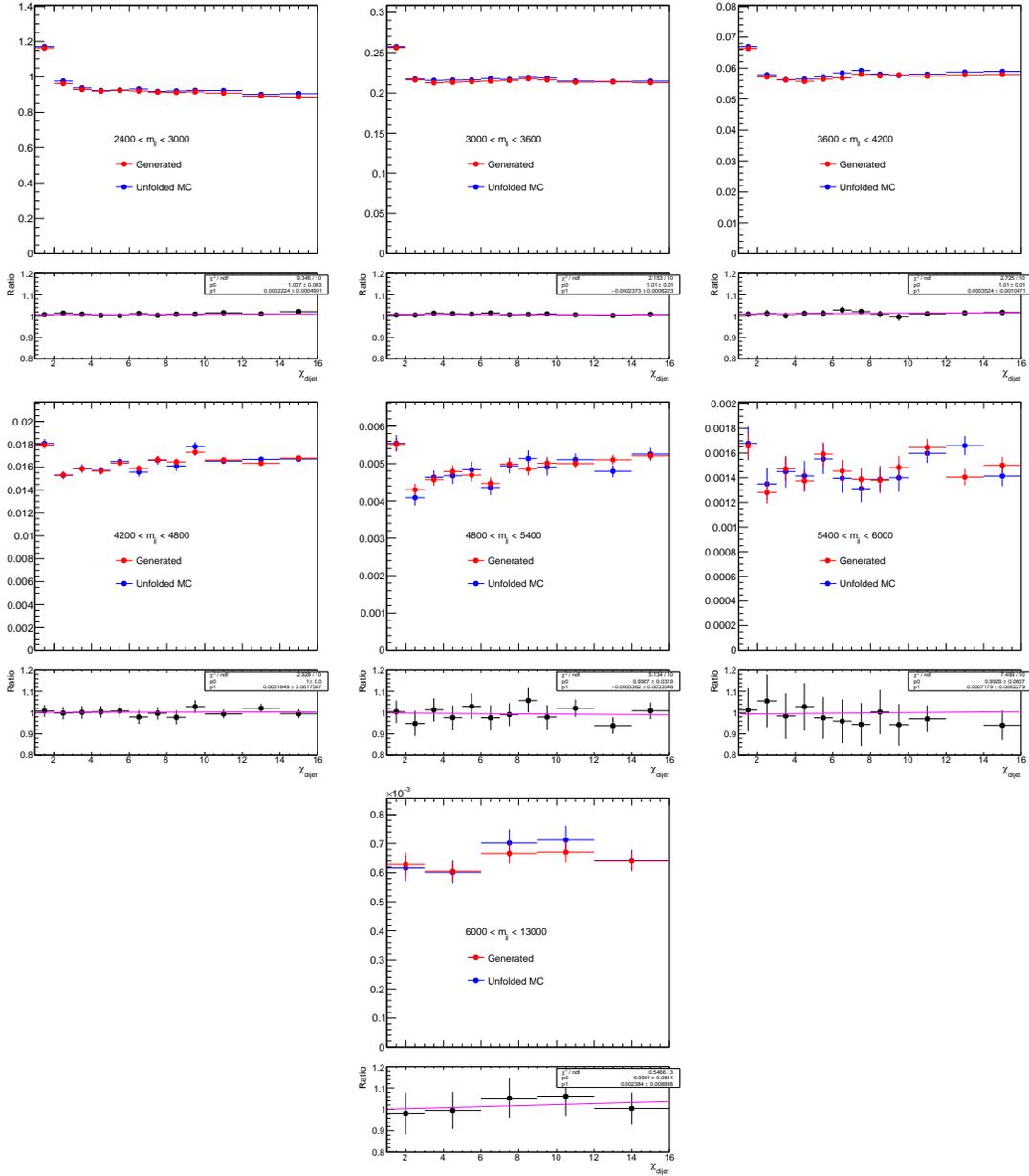


Figure 81: Comparison of unfolded and generated χ_{dijet} distributions from MAD-GRAPH5_amc@NLO. In these comparisons the response matrix used in the unfolding procedure is derived from an independent sample of events generated by PYTHIA.

CHAPTER 7

THEORETICAL PREDICTIONS

7.1 Theoretical Predictions for Quantum Chromodynamics

The NLOJET++ [80] generator is used to calculate the predictions for the NLO QCD χ_{dijet} distributions. NLOJET++ is integrated into the fastNLO package [113], which can calculate the uncertainties due to PDFs and μ_{R} and μ_{F} scales with good CPU performance. The nominal calculation is performed using the CT14 NLO PDFs [63], with the μ_{F} and μ_{R} scales to be the average p_{T} of the two leading jets ($\mu_{\text{F}} = \mu_{\text{R}} = \langle p_{\text{T}} \rangle$).

7.1.1 Theoretical Uncertainties

The scale uncertainty in the predicted χ_{dijet} distributions is obtained by varying both μ_{F} and μ_{R} in six combinations: $(\mu_{\text{F}}, \mu_{\text{R}}) = (\langle p_{\text{T}} \rangle/2, \langle p_{\text{T}} \rangle/2)$, $(\langle p_{\text{T}} \rangle/2, \langle p_{\text{T}} \rangle)$, $(\langle p_{\text{T}} \rangle, \langle p_{\text{T}} \rangle/2)$, $(2\langle p_{\text{T}} \rangle, 2\langle p_{\text{T}} \rangle)$, $(2\langle p_{\text{T}} \rangle, \langle p_{\text{T}} \rangle)$ and $(\langle p_{\text{T}} \rangle, 2\langle p_{\text{T}} \rangle)$ and taking the maximal downwards and upwards deviations of the χ_{dijet} distributions from the central values. This recipe for determining the scale uncertainties is taken from Refs. [114, 115]. The scale uncertainty for the various M_{jj} regions defined in this analysis is presented in Figure 82 (red dashed line). It varies from 8.5% in the lowest M_{jj} bin, to 19% in the highest mass bin and is dominated by the variation of μ_{R} .

The uncertainty in the predicted χ_{dijet} distributions due to the PDF uncertainties is evaluated using the set of uncertainty eigenvectors provided with the CT14 PDF set. The CT14 PDF uncertainty is obtained using the Hessian method [116, 117], where the correlated exper-

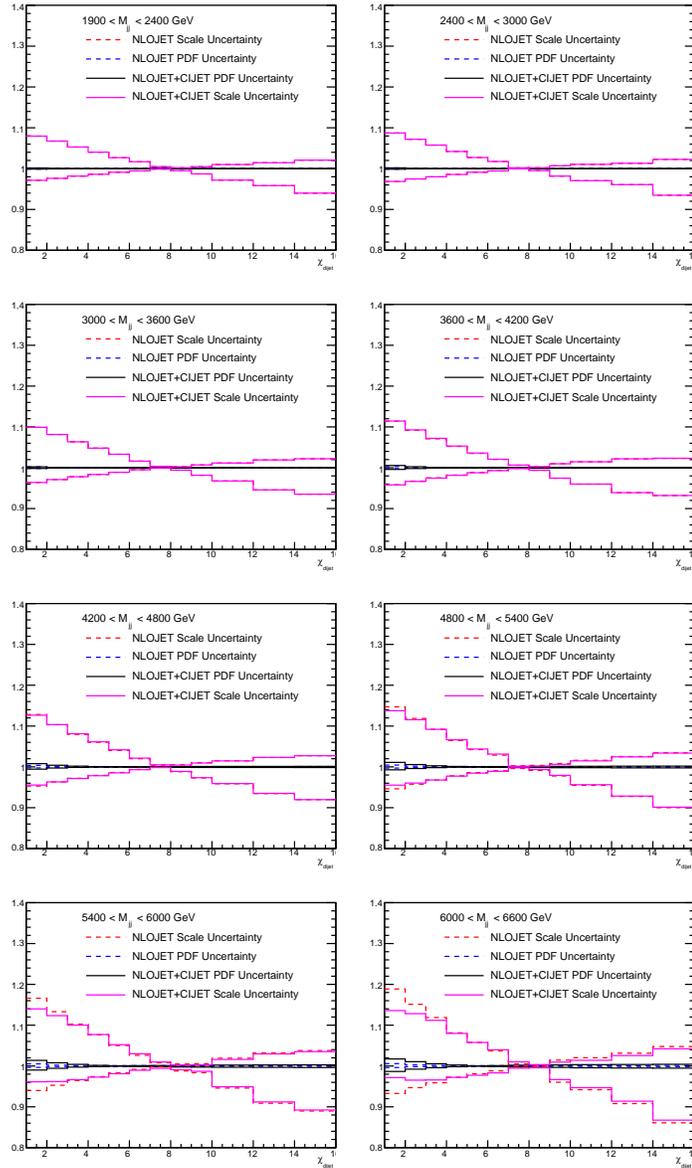


Figure 82: Scale and PDF uncertainties for the QCD (dashed) and QCD + CI (solid) predictions in the M_{jj} bins of this analysis. The CI prediction corresponds to the CI model with $\Lambda_{\overline{LL}} = 18$ TeV.

imental data uncertainties on the fitted parameters in the PDFs are provided as uncertainty eigenvectors within an orthonormalized parameter space. Since each uncertainty eigenvector in the Hessian method is uncorrelated with all other uncertainty eigenvectors, the total impact on χ_{dijet} is evaluated as the quadratic sum of the impact of each uncertainty eigenvector. The PDF uncertainty is shown in Figure 82 (blue dashed line). The magnitude of the PDF uncertainty is relative small because most of the QCD $2 \rightarrow 2$ processes have a uniform χ_{dijet} distribution.

As a crosscheck, the PDF uncertainty is also evaluated using a more flexible statistical combination of multiple PDF sets as proposed by the PDF4LHC15_100 [63, 118–122] method. The results obtained from this method exhibit only small differences as compared to the results obtained using the CT14 PDF set as described above and have a negligible impact on the limits for the NP models described in the next chapter.

A summary of the theoretical uncertainties is shown in Table VIII. As can be seen in the table, the μ_{R} and μ_{F} scale uncertainty dominates the theoretical systematic uncertainties.

TABLE VIII: The theoretical systematic uncertainties in the normalized χ_{dijet} distributions, in percent. While in the statistical analysis (see Chapter 8), each systematic uncertainty is represented through a change in the χ_{dijet} distributions correlated across all the M_{jj} and χ_{dijet} bins, representative values which are taken as the relative shifts for the lowest χ_{dijet} bins in the lowest and highest M_{jj} bins, are used to illustrate the relative contributions in this table.

Source of uncertainty	$2.4 < M_{\text{jj}} < 3.0 \text{ TeV}$	$M_{\text{jj}} > 6.0 \text{ TeV}$
QCD NLO scale (6 changes in μ_{r} and μ_{f})	$\begin{matrix} +8.5 \\ -3.0 \end{matrix}$	$\begin{matrix} +19 \\ -5.8 \end{matrix}$
PDF (CT14 eigenvectors)	0.2	0.6

7.1.2 Electroweak Corrections

Electroweak corrections for dijet productions have been applied [123]. In these corrections, the electroweak interactions between quarks, e.g., the $u\bar{d} \rightarrow W^+ \rightarrow u\bar{d}$ process, are included in the prediction of dijet productions. The correction factors are provided by the authors of Ref. [123] and are shown in Figure 83. The correction factors are shown in Figure 83.

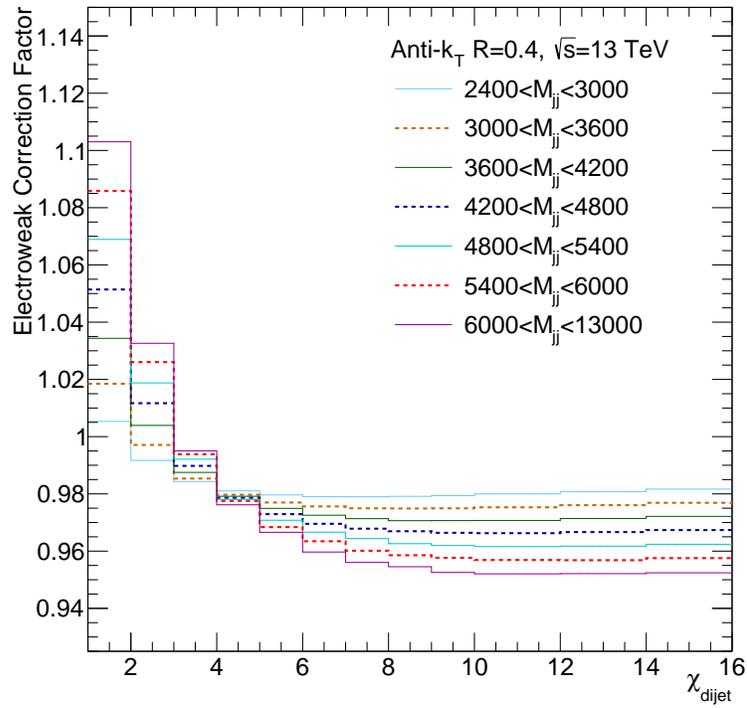


Figure 83: The electroweak correction factors for dijet production as a function of χ_{dijet} for the M_{jj} bins used in the analysis.

7.1.3 Non-perturbative Corrections

The NLO QCD calculations are performed at parton-level and do not include non-perturbative effects due to hadronization and multiple parton interactions. The impact of non-perturbative effects on the predictions are evaluated using the PYTHIA (PartonLevel:MPI, HadronLevel:Hadronize) and HERWIG++ (LHCHandler:MultipleInteractionHandler, LHCHandler:HadronizationHandler) event generators. The simulated jets are reconstructed with and without non-perturbative effects enabled in the event generators and the ratio of the χ_{dijet} distributions obtained with and without the non-perturbative effects are taken as the non-perturbative correction factors. The resulting correction factors are shown in Figure 84, and they are found to be negligible for both MC generators. Therefore, no correction needs to be applied to the QCD predictions for these non-perturbative effects.

7.2 Contact Interaction Theoretical Predictions

The CI models considered were described earlier in Section 1.2.1. The models with positive (negative) η_{LL} or η_{RR} lead to destructive (constructive) interference with the QCD terms and consequently a lower (higher) cross section. In the CI models studied in this thesis, NLO QCD corrections are applied to the predictions. In proton-proton collisions, the Λ_{LL}^{\pm} and Λ_{RR}^{\pm} models have identical tree level diagrams and NLO corrections, leading to same cross sections. For Λ_{VV}^{\pm} and Λ_{AA}^{\pm} , as well as for $\Lambda_{(V-A)}^{\pm}$, the CI predictions are identical at tree level, but exhibit different NLO corrections and therefore yield different cross sections. For the calculation of the CI terms, as well as the interference between QCD and CI terms at NLO in QCD, the CIJET program version 1.0 [81] is used. The CI predictions are then combined with the NLO

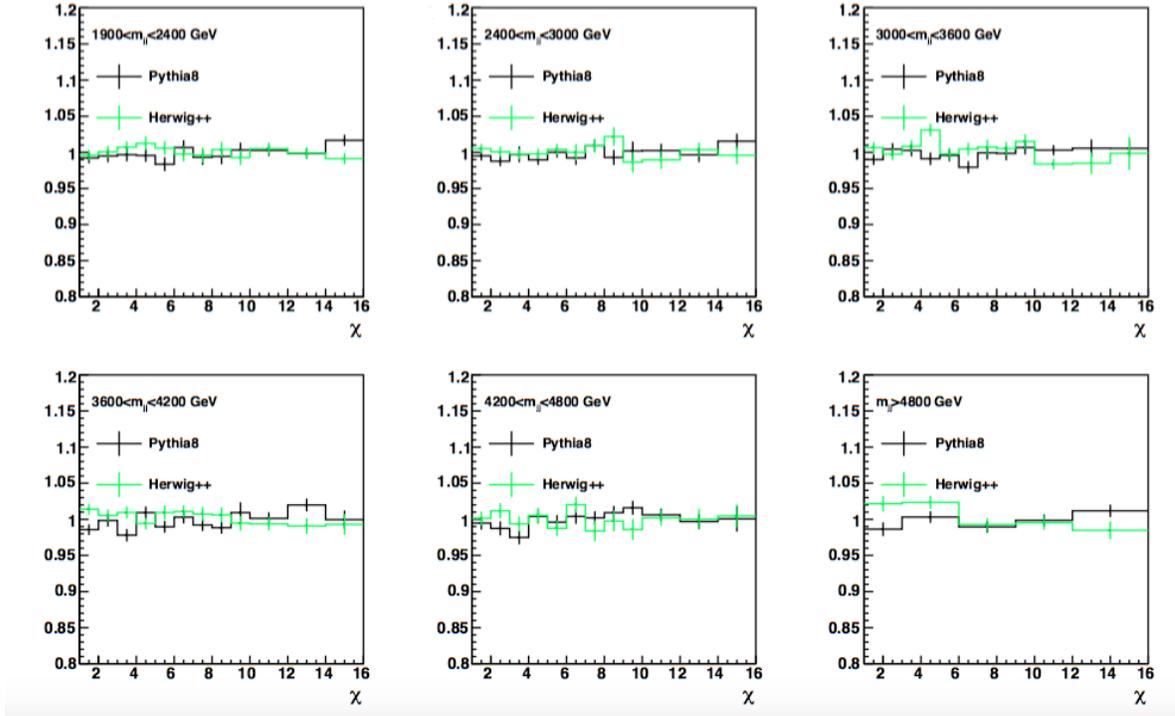


Figure 84: Non-perturbative corrections derived from PYTHIA (black) and HERWIG++ (green). The y-axis corresponds to the values of the non-perturbative correction factors.

QCD+EW predictions to obtain the final predictions. The PDF set used in the CI predictions is the CT14 NLO set.

PDF and μ_R and μ_F scale choice uncertainties for the CI predictions are calculated in the same manner as for the QCD predictions. The CI+QCD scale and PDF uncertainties for $\Lambda_{\text{LL}}^- = 18$ TeV are shown in Figure 82 (pink and black solid lines). The uncertainties for the scale and PDF choices in the CI+QCD predictions in the highest mass bin are estimated to be 14% and 2%, respectively. The fact that the scale uncertainties for the CI+QCD prediction is

smaller than the scale uncertainties for the QCD-only predictions is due to cancellations that occur in the normalization of the χ_{dijet} distributions.

7.3 Theoretical Predictions for Extra Spatial Dimensions

The LO predictions for the ADD model are calculated using PYTHIA in the GRW convention using the parameters shown in Table IX. The predictions for the HLZ convention can be obtained from the GRW predictions using the relations shown in Equation 1.20. The interference between the ADD and QCD terms are included in PYTHIA. To obtain the best estimate for the QCD+ADD process, NLO QCD and EW corrections for the QCD term ($\sigma_{\text{NLO+EW}}^{\text{QCD}} - \sigma_{\text{LO}}^{\text{QCD}}$) are added bin-by-bin to the LO QCD+ADD prediction:

$$\sigma^{\text{QCD+ADD}} = (\sigma_{\text{NLO+EW}}^{\text{QCD}} - \sigma_{\text{LO}}^{\text{QCD}} + \sigma_{\text{LO}}^{\text{QCD+ADD}}). \quad (7.1)$$

TABLE IX: Parameters used in the PYTHIA prediction of the ADD virtual graviton exchange process.

parameter	value	description
HardQCD:all	off	custom dijet processes
ExtraDimensionsLED:dijets	on	use extra dimensions
ExtraDimensionsLED:CutOffmode	0	GRW convention
ExtraDimensionsLED:LambdaT	Λ_T	energy scale parameter
ExtraDimensionsLED:nQuarkNew	5	outgoing mass-less quark flavours
ExtraDimensionsLED:opMode	1	use GRW convention

7.4 Quantum Black Hole Theoretical Predictions

The QBH 3.0 generator [82] is used to calculate the predictions for QBH production. The QCD terms do not interfere with QBH terms since gravity has no interference with the QCD processes. Because QBHs have mass close to the fundamental Planck scale, the minimum QBH mass, M_{QBH} , is set to be equal to the fundamental Planck scale M_{D} for simplicity. The other parameters used in the QBH generator are shown in Table X.

TABLE X: Parameters used in the QBH prediction of the QBH production and decay processes.

parameter	value	description
setQscale	true	QCD scale for PDFs to be inverse gravitational radius
setLHAglue	10042	CTEQ6L1 PDF
setPlanckdef	3	definition of Planck Scale to be PDF definition
setSM	false	Allow violation of standard model global symmetry
setChiral	false	Neutrinos are only left-handed
setMajorana	false	Neutrinos are Dirac particles

7.5 Dark Matter Theoretical Predictions

Interactions of quarks with dark matter particles mediated by the vector and axial-vector mediators in the simplified model of DM [55, 56] are generated using MADDM version 2.0.6 [124, 125] at LO. The dark matter predictions are generated at fixed g_{DM} and m_{DM} values, where $g_{\text{DM}} = 1.0$ and $m_{\text{DM}} = 1$ GeV.

The width of the mediators as a function of g_q is shown in Figure 85. Scenarios with $g_q > 0.5$ are of particular interest to this search, as they predict resonances with a relative width larger than 15%. And in these scenarios, the sensitivity of searches using dijet resonances is limited. The inclusive cross sections and the overall acceptances in the analysis phase space of various generated mediator models as a function of the mediator mass are shown in Figures 86 and 87, respectively. As shown in the figures, the differences in the cross sections and acceptances between vector and axial-vector mediators are negligible for the generated samples.

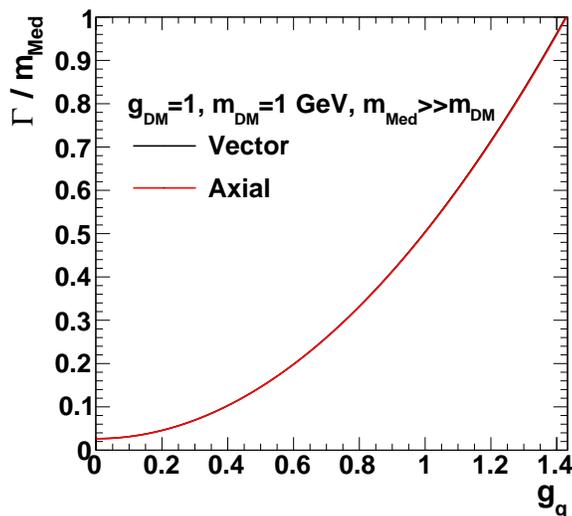


Figure 85: Relative width of the generated DM mediators as a function of g_q .

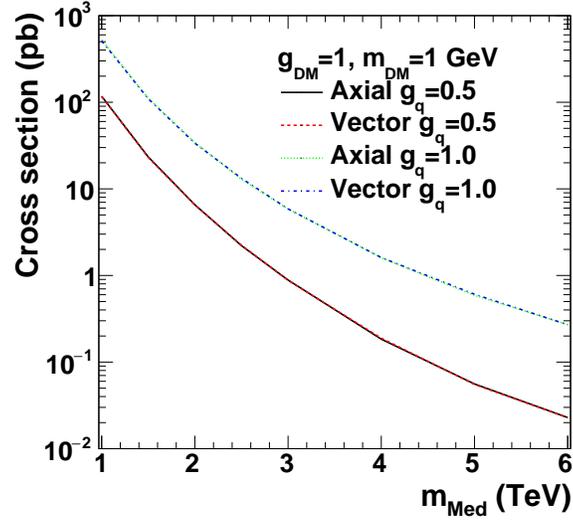


Figure 86: Inclusive cross section for the generated DM mediators as a function of M_{Med} .

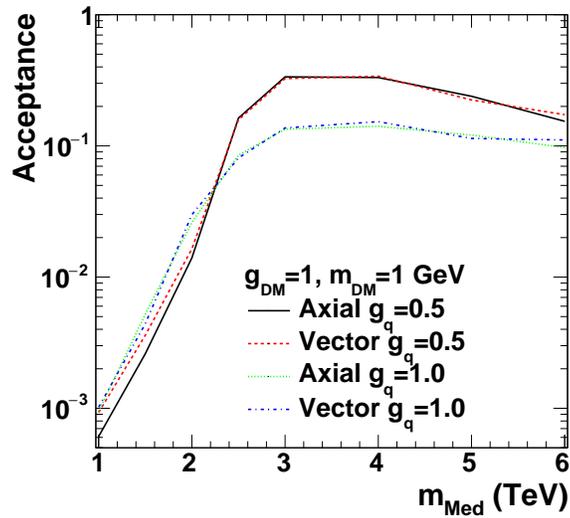


Figure 87: Acceptance ($|y_{\text{boost}}| < 1.11$, $\chi_{\text{dijet}} < 16$, and $M_{\text{jj}} > 2.4$ TeV) of the generated DM mediators as a function of M_{Med} .

The PDF uncertainty on the DM predictions is evaluated using the NNPDF3.0 PDF set [120] and found to be non-negligible for DM mediators that have large mass and coupling for $M_{jj} > M_{\text{Med}}$. For example, for an axial-vector mediator with $M_{\text{Med}} = 6 \text{ TeV}$ and $g_q = 1.0$, which corresponds to a resonance with a relative width of 50%, the PDF uncertainty is 14% in the $M_{jj} > 6.0 \text{ TeV}$ bin. The impact of the scale uncertainties on the DM models is found to be negligible.

CHAPTER 8

RESULTS¹

8.1 Comparisons between Data and Theoretical Predictions

The measured normalized χ_{dijet} distributions, unfolded to particle level, are compared to NLO QCD predictions with EW corrections in Figures 88 and 89. Also shown are the ratios between the unfolded data distributions and the NLO QCD+EW predictions. The level of agreement between the data distributions and SM predictions in each M_{jj} bin is quantified using: 1), a simple χ^2 test, which combines the systematic and statistical uncertainties in quadrature and assumes that the uncertainties are uncorrelated between χ_{dijet} bins, and 2), a goodness-of-fit (GOF) test, which takes into account all uncertainties and their correlations based on the saturated model [127] and estimates the significance with respect to the expected goodness of fit. The results of these tests are summarized in Table XI and indicate no significant deviation of the data from the SM predictions.

The distributions shown in Figures 88 and 89 are also compared to predictions for several NP models: QCD+CI with $\Lambda_{\text{LL}}^{\pm} = 14$ TeV, QCD+ADD with $\Lambda_{\text{T}}(\text{GRW}) = 10$ TeV, QCD+QBH with $M_{\text{QBH}}(\text{ADD6}) = 8$ TeV, and QCD+DM with $M_{\text{Med}} = 2, 3$ and 5 TeV and $g_{\text{q}} = 1.0$. In

¹The results of this thesis have been published in European Physics Journal C [126]. The copyright policy of the European Physics Journal C can be found in Appendix C.

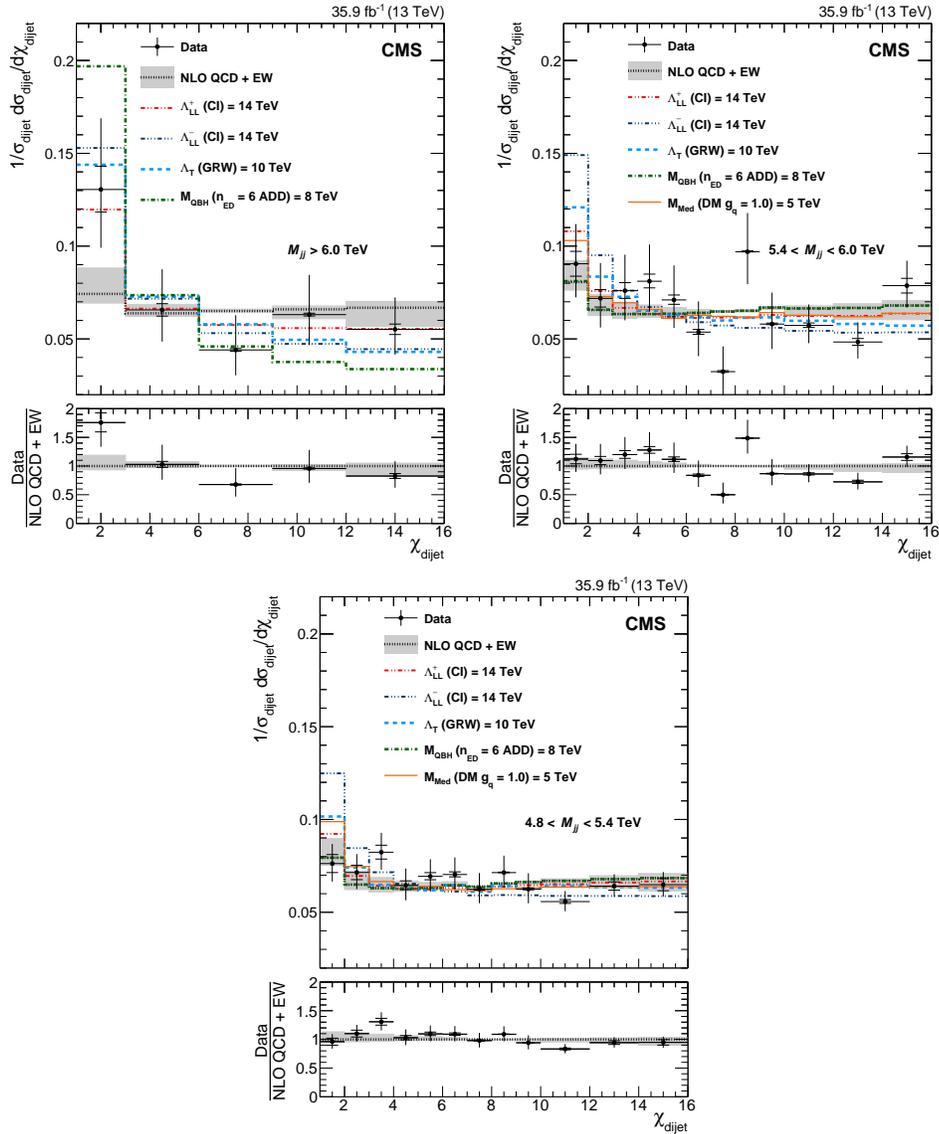


Figure 88: Normalized χ_{dijet} distributions from data compared to the NLO QCD prediction with EW corrections in the three highest mass bins. Also shown are the predictions from the CI, ADD, QBH, and DM models. The error bars represent statistical and experimental systematic uncertainties added in quadrature. The ticks on the error bars and the gray bands correspond to the experimental and theoretical systematic uncertainty, respectively. The ratios between the unfolded data distributions and the NLO QCD+EW predictions are shown in the lower panels.

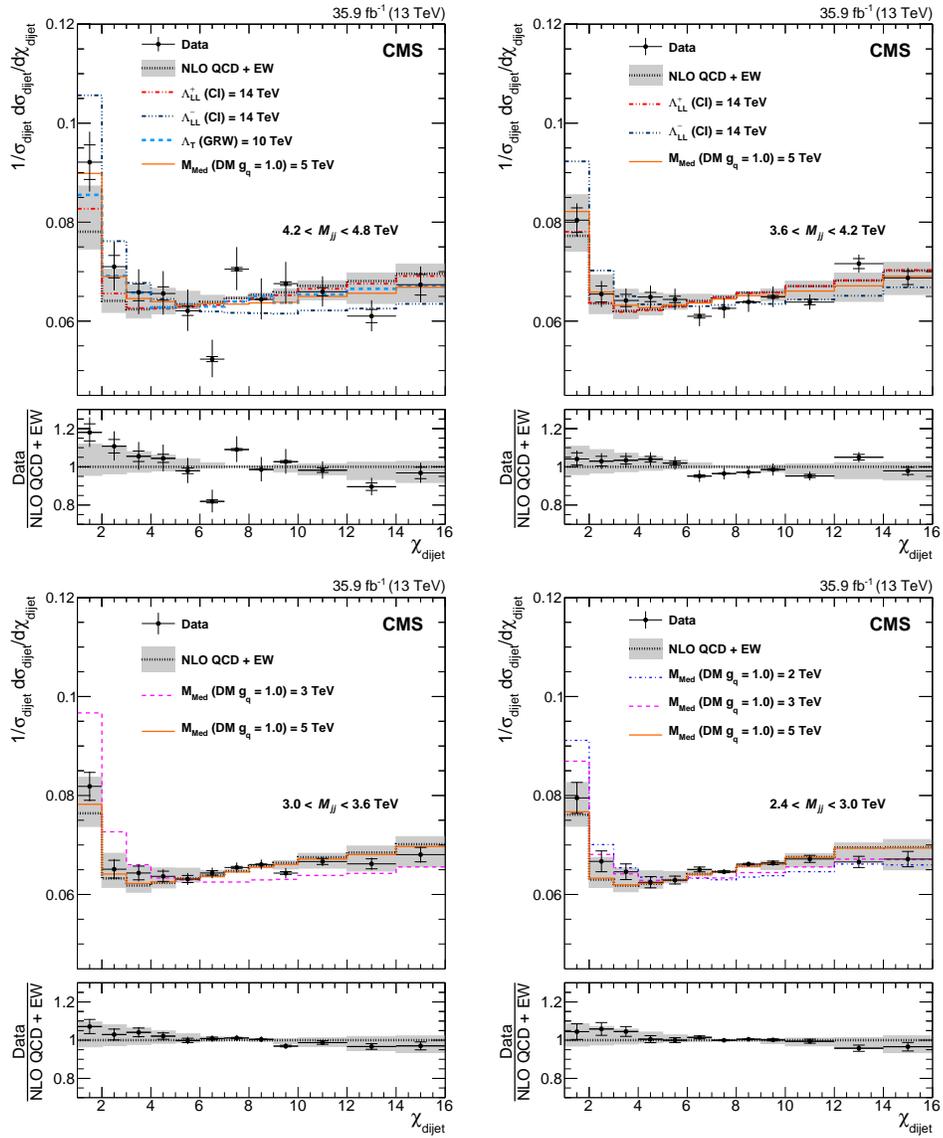


Figure 89: Same as Figure 88 but for the four lower mass bins.

TABLE XI: Agreement between the data distributions and theoretical predictions. The χ^2 in each M_{jj} bin is divided by the number of degrees of freedom (DOF).

Mass bins [TeV]	χ^2/DOF	GOF
6.0–13.0	1.2	0.4
5.4–6.0	1.0	0.4
4.8–5.4	1.9	1.3
4.2–4.8	0.7	0.3
3.6–4.2	2.2	0.2
3.0–3.6	3.4	1.5
2.4–3.0	3.9	1.7

these figures, distributions from the NP models are shown only for those M_{jj} ranges that have sensitivity to the signal.

8.2 Determination of Exclusion Limits

8.2.1 Statistical Method

The exclusion limits presented in this thesis are determined using the modified frequentist approach [128, 129]. The test statistic, which discriminates the QCD-only (background) and QCD+NP (signal+background) hypotheses, is defined as

$$q = -2 \ln \frac{L_{\mu s+b}}{L_{\mu_0 s+b}}, \quad (8.1)$$

where μ is the signal strength and is equal to 1 (0) for the QCD+NP (QCD-only) hypothesis. The likelihoods L_{s+b} and L_b are defined for the respective QCD+NP and QCD-only hypotheses as a product of Poisson likelihood functions for each bin in the χ_{dijet} distributions,

$$L(n_i|\mu s_i + b_i) = \prod_i \frac{(\mu s + b)_i^{n_i}}{n_i!} \cdot e^{-(\mu s + b)_i} \quad (8.2)$$

where n_i is the number of events in bin i in the data distributions and s and b are the predictions for the signal and background yields in bin i . In the statistical tests, the predictions for each M_{jj} range are normalized to the number of observed events in the M_{jj} range.

The theoretical and experimental systematic uncertainties are treated as nuisance parameters $\theta = (\theta_0, \theta_1, \dots, \theta_k)$ in the likelihood model:

$$L(n|\mu s + b, \theta) = L(n|(\mu s(\theta) + b(\theta))) \cdot \rho(\theta), \quad (8.3)$$

where $\rho(\theta)$ is the probability density function of the nuisance parameters. It is described by the normal distribution:

$$\rho(\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta^2}{2}\right). \quad (8.4)$$

The nuisance parameters are profiled with respect to the background-only and signal+background models by maximizing the corresponding likelihood functions. After the maximum likelihood fit for the nuisance parameters, the profile likelihoods are written as

$$L(n|\mu s + b, \hat{\theta}), \quad (8.5)$$

where $\hat{\theta}$ are the maximum likelihood estimators of the nuisance parameters. Profiling the nuisance parameters helps constrain the systematic uncertainties because of the large number of events in the low- M_{jj} ranges. This gives 2–30% better observed limits on the NP scales and masses than the limits obtained using the method in the predecessor of this search reported in Ref. [33], where the nuisance parameters were allowed to float freely according to the systematic uncertainties.

The p -values, $P_{\text{QCD+NP}}(\hat{q} \geq \hat{q}_{\text{obs}})$ and $1 - P_{\text{QCD}}(\hat{q} \leq \hat{q}_{\text{obs}})$, where \hat{q} is the profile log-likelihood ratio and \hat{q}_{obs} is determined from the data, are evaluated for the QCD+NP and QCD-only hypotheses, respectively. The p -value $P_{\text{QCD+NP}}(\hat{q} \geq \hat{q}_{\text{obs}})$ quantifies the compatibility of the data with the signal+background hypothesis, while the p -value $1 - P_{\text{QCD}}(\hat{q} \leq \hat{q}_{\text{obs}})$ quantifies the compatibility of the data with the background-only hypothesis. These p -values can be determined by generating an ensemble of pseudo-data samples according to the QCD-only and QCD+NP predictions and comparing the q value obtained from each set of pseudo-data with q_{obs} . However, as generating the pseudo-data ensembles is very CPU intensive, another method is often used to evaluate the p -values. This is called the asymptotic method [130]. If the number of observed events is large enough, the asymptotic method gives the same results as the method using pseudo-data. However, when the number of events is small, the asymptotic method fails. A cross-check has been performed using the ADD model to verify that the asymptotic method is applicable to this analysis. The difference between the results from the two methods was found to be negligible and therefore this analysis uses the asymptotic method. The ADD model

was chosen for this test since the ADD model has signal only in the highest mass bin, and in this bin the total number of events is small.

Limits on the QCD+NP models are set based on the quantity

$$\text{CL}_s = \frac{\text{CL}_{s+b}}{\text{CL}_b} = \frac{P_{\text{QCD+NP}}(\hat{q} \geq \hat{q}_{\text{obs}})}{(1 - P_{\text{QCD}}(\hat{q} \leq \hat{q}_{\text{obs}}))}, \quad (8.6)$$

which by convention is required to be less than 0.05 for a 95% confidence level (CL) of exclusion. This method is called the modified frequentist method and is different from the classical frequentist method, where limits are set based on CL_{s+b} , in that the modified method gives more sensible results for signals that have small yield. This can be understood by the following argument. If the signal yield is small, then $s + b \sim b$ and $\text{CL}_{s+b} \sim \text{CL}_b$. When we exclude the signal+background hypothesis using the classical frequentist method at 95% CL (i.e., $\text{CL}_{s+b} < 0.05$), we also exclude the background-only hypothesis at approximately the same CL (i.e., $\text{CL}_b \sim 0.05$). However, in the modified frequentist method, this situation can be avoided since we require $\text{CL}_{s+b}/\text{CL}_b < 0.05$ for a 95% CL exclusion.

To quantify the sensitivity of the analysis independent of statistical fluctuations in the data, one can calculate expected limits. This calculation can be done by replacing q_{obs} with \hat{q} at the median of $P_{\text{QCD}}(\hat{q})$ in the CL_s . With this construction, the expected limits are equivalent to the observed limits for the case where the observed data overlap with the expectation from the QCD-only hypothesis. In addition to the expected limits, one can also calculate the 1σ (2σ)

variations of the expected limits by replacing \hat{q} at the median of $P_{\text{QCD}}(\hat{q})$ with \hat{q} at 16% (2.2%) and 84% (97.7%) values of $P_{\text{QCD}}(\hat{q})$.

In the limit calculation for this analysis, only the M_{jj} ranges that improve the expected limits by more than 1% are included. For the CI models, ADD models, and QBH models, mass bins with $M_{\text{jj}} > 3.6 \text{ TeV}$, $M_{\text{jj}} > 4.2 \text{ TeV}$, and $M_{\text{jj}} > 4.8 \text{ TeV}$ are used, respectively. For the DM mediators, mass bins that cover the M_{jj} range of $0.5M_{\text{Med}}-1.2M_{\text{Med}}$ are used.

Detector-level χ_{dijet} distributions are compared to theoretical predictions that have been folded to detector-level to determine the exclusion limits on the NP models. This is done because for the detector-level χ_{dijet} distributions, each χ_{dijet} bin can be modeled by a Poisson likelihood function, while in the unfolded distributions, bins are correlated among the M_{jj} ranges and therefore, strictly speaking, cannot be modeled by Poisson likelihood functions. The particle-level predictions are folded to detector-level by multiplying the particle-level distributions by the response matrix. A cross-check was performed to verify that the limits for SM+NP predictions folded with the response matrix derived from the PYTHIA QCD sample (see Section 6.4) smeared with the double-sided Crystal Ball parametrization of the jet p_{T} resolution agree with the SM+NP predictions smeared with the same parameterization. The details of this cross-check is described in Appendix B.

Although the limits are extracted using detector-level theoretical predictions and data distributions, it is also useful to extract the limits using particle-level theoretical predictions and unfolded χ_{dijet} distributions approximated by Poisson likelihood functions since the limits obtained from the two methods are not expected to be very different. The resulting observed

limits on the NP scales and masses determined at particle level are found to be $\sim 1\%$ ($\sim 10\%$) more stringent than the observed limits determined at detector level for models which have signal in the highest mass bin (across all mass bins) (see Appendix B).

Although the agreement between data and the NLO QCD+EW predictions has been quantified using χ^2 and GOF tests, it can also be quantified using the profile likelihood ratio by calculating the p-value $P_{\text{QCD}}(q < q_{\text{obs}})$ (CL_b) for each mass bin separately. The largest excess is found in the $M_{\text{jj}} > 6.0$ TeV bin, with a significance of 1.8 standard deviations. In this calculation, $\Lambda_{\text{LL/RR}}^+$ model is used in the likelihood ratio (Equation 8.1). As mass bins are combined in the QCD+NP models under study, the significances of the data distributions in combined mass bins are also calculated and the results are shown in Table XII. The largest significances are found to be between 2.7–2.8 standard deviations for the QCD+DM models with M_{Med} between 4.5–6.0 TeV and $g_q = 1.0$.

TABLE XII: Significance of the data distributions in number of standard deviations for the QCD+DM models with $g_q = 1.0$.

M_{Med} [TeV]	M_{jj} ranges [TeV]	Significance
2.0	2.4–3.0	0.2
2.5	2.4–3.0	0.6
3.0	2.4–3.6	2.1
3.5	2.4–4.2	2.4
4.0	2.4–4.2	2.0
4.5	2.4–4.8	2.8
5.0	2.4–6.0	2.8
6.0	3.0–13.0	2.7

8.2.2 Limits on New Physics Models

The observed and expected 95% CL exclusion limits for the different CI, ADD, and QBH models studied in this analysis are shown in Figure 90. They are also displayed in Table XIII. The observed limits are less stringent than the expected limits because of the upward fluctuation of the data at low χ_{dijet} in the $M_{\text{jj}} > 6.0$ and the $4.2 < M_{\text{jj}} < 4.8$ TeV bins compared to the theoretical predictions.

TABLE XIII: Observed and expected exclusion limits at 95% CL for CI, ADD, and QBH models.

Model	Observed lower limit (TeV)	Expected lower limit (TeV)
$\Lambda_{\text{LL/RR}}^+$ (NLO)	12.8	14.6 ± 0.8
$\Lambda_{\text{LL/RR}}^-$ (NLO)	17.5	23.5 ± 3.0
Λ_{VV}^+ (NLO)	14.6	16.4 ± 0.8
Λ_{VV}^- (NLO)	22.4	30.7 ± 3.7
Λ_{AA}^+ (NLO)	14.7	16.5 ± 0.8
Λ_{AA}^- (NLO)	22.3	30.6 ± 3.8
$\Lambda_{\text{(V-A)}}^+$ (NLO)	9.2	11.5 ± 1.0
$\Lambda_{\text{(V-A)}}^-$ (NLO)	9.3	11.8 ± 1.1
ADD Λ_{T} (GRW)	10.1	11.5 ± 0.7
ADD M_{S} (HLZ) $n_{\text{ED}} = 2$	10.7	12.5 ± 0.8
ADD M_{S} (HLZ) $n_{\text{ED}} = 3$	12.0	13.7 ± 0.8
ADD M_{S} (HLZ) $n_{\text{ED}} = 4$	10.1	11.5 ± 0.7
ADD M_{S} (HLZ) $n_{\text{ED}} = 5$	9.1	10.4 ± 0.6
ADD M_{S} (HLZ) $n_{\text{ED}} = 6$	8.5	9.7 ± 0.6
QBH M_{QBH} (ADD $n_{\text{ED}} = 6$)	8.2	8.5 ± 0.4
QBH M_{QBH} (RS $n_{\text{ED}} = 1$)	5.9	6.3 ± 0.7

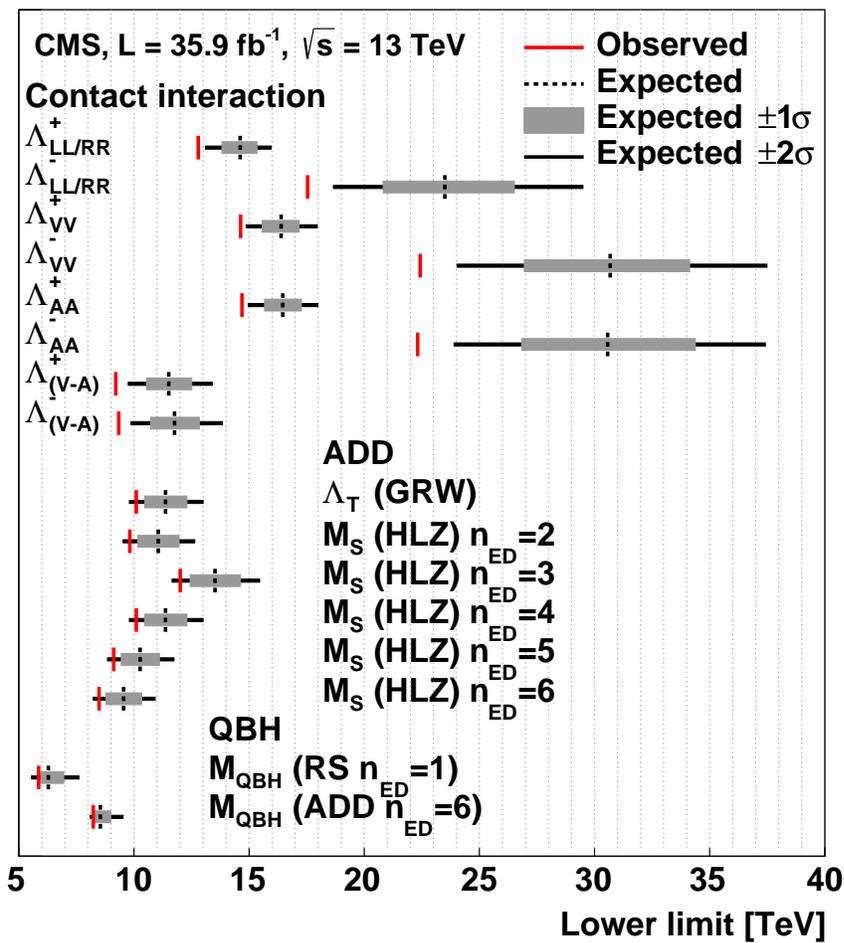


Figure 90: 95% CL exclusion limits for CI, ADD, and QBH models. The observed (expected) limits are indicated by the solid (dashed) vertical lines. The $\pm 1\sigma$ ($\pm 2\sigma$) variations of the expected limits are given by the horizontal gray bands (black lines).

95% CL upper limits on g_q as a function of the mass of the vector or axial-vector DM mediator with $g_{\text{DM}} = 1.0$ and $m_{\text{DM}} = 1 \text{ GeV}$ are shown in Figure 91. Also shown on the vertical axis on the right-hand side of the figure are the corresponding limits on the width of the mediators. The limits degrade below $M_{\text{Med}} = 2.5 \text{ TeV}$ because the lower M_{jj} boundary of the analysis is 2.4 TeV . This causes the acceptance losses to grow as the resonance mass moves below this boundary. The limits degrade quickly above $M_{\text{Med}} > 4 \text{ TeV}$ because resonances with large mass and width are predominantly produced off-shell with a mass less than 2.4 TeV . This results in a loss of acceptance for these resonances as a function of increasing resonance width. The observed limit above 5 TeV is at $\Gamma/M_{\text{Med}} \geq 1$, which is in a region where the simplified model of a mediator particle is no longer valid.

The limits for arbitrary DM mass (m'_{DM}) as a function of M_{Med} at fixed couplings can be calculated based on the fact that at fixed mediator production cross sections, changes in the width of the DM decay channel will lead to changes in the width of the quark decay channel. The cross section of the DM mediator can be written as:

$$\sigma = \frac{g_q^4}{\Gamma} \frac{C}{M_{\text{Med}}} \quad (8.7)$$

where C is a constant and Γ is the total width of the DM mediator. The partial widths for the quark and DM decay channels for vector and axial-vector mediators can be written as:

$$\Gamma_{\text{vector}}^{\text{qq}} = \frac{g_q^2 M_{\text{Med}}}{4\pi} \left(1 - 4 \frac{m_q^2}{M_{\text{Med}}^2}\right)^{1/2} \left(1 + 2 \frac{m_q^2}{M_{\text{Med}}^2}\right) \quad (8.8)$$

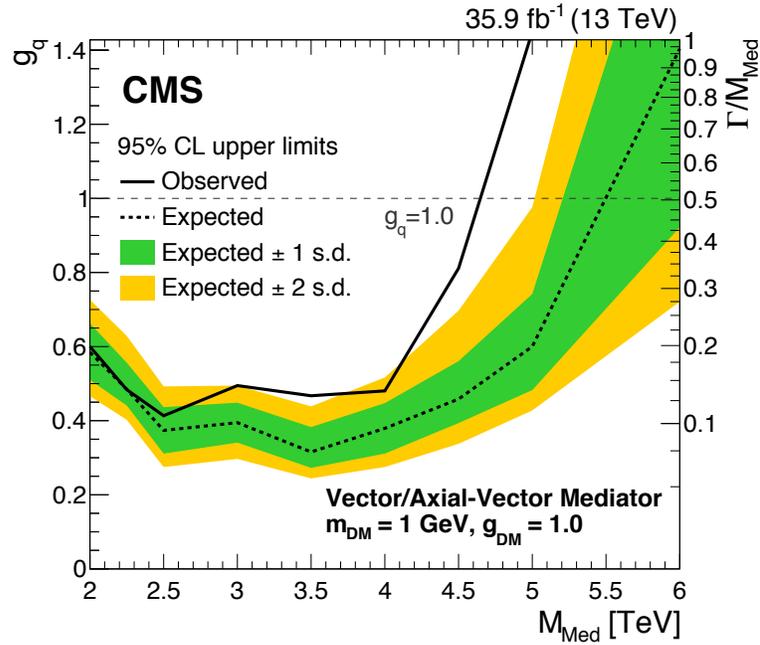


Figure 91: 95% CL upper limits on the quark coupling g_q as a function of the mass of the vector or axial-vector DM mediator M_{Med} with $g_{\text{DM}} = 1.0$ and $m_{\text{DM}} = 1$ GeV. Also shown on the vertical axis on the right-hand side of the figure are the corresponding limits on the width of the mediators. The observed limits, expected limits, and the one (two) σ variation of the expected limits are indicated by the solid line, dashed line, and the green (yellow) shaded band, respectively. The dashed horizontal gray line indicates a benchmark DM mediator with $g_q = 1.0$.

$$\Gamma_{\text{vector}}^{\text{DM}} = \frac{g_{\text{DM}}^2 M_{\text{Med}}}{12\pi} \left(1 - 4 \frac{m_{\text{DM}}^2}{M_{\text{Med}}^2}\right)^{1/2} \left(1 + 2 \frac{m_{\text{DM}}^2}{M_{\text{Med}}^2}\right) \quad (8.9)$$

and

$$\Gamma_{\text{axial-vector}}^{\text{qq}} = \frac{g_{\text{q}}^2 M_{\text{Med}}}{4\pi} \left(1 - 4 \frac{m_{\text{q}}^2}{M_{\text{Med}}^2}\right)^{3/2} \quad (8.10)$$

$$\Gamma_{\text{axial-vector}}^{\text{DM}} = \frac{g_{\text{DM}}^2 M_{\text{Med}}}{12\pi} \left(1 - 4 \frac{m_{\text{DM}}^2}{M_{\text{Med}}^2}\right)^{3/2}, \quad (8.11)$$

respectively. From these equations, it follows that at fixed mediator production cross section, two models with the same M_{Med} and g_{DM} but different m_{DM} (m'_{DM}) and g_{q} (g'_q) are related to each other by:

$$\sigma = \frac{g_{\text{q}}^4}{\Gamma^{\text{qq}}(g_{\text{q}}) + \Gamma^{\text{DM}}(m_{\text{DM}}, g_{\text{DM}})} \frac{C}{M_{\text{Med}}} = \frac{g'_{\text{q}}{}^4}{\Gamma^{\text{qq}}(g'_q) + \Gamma^{\text{DM}}(m'_{\text{DM}}, g_{\text{DM}})} \frac{C}{M_{\text{Med}}}. \quad (8.12)$$

Solving Equation 8.12 for g_{q} , the following expression is obtained:

$$g_{\text{q}} = \sqrt{\frac{A + \sqrt{A^2 + 4B}}{2}} \quad (8.13)$$

where

$$A = \frac{g'_{\text{q}}{}^2 \Gamma^{\text{qq}}(g'_q)}{\Gamma^{\text{qq}}(g'_q) + \Gamma^{\text{DM}}(m'_{\text{DM}}, g_{\text{DM}})} \quad \text{and} \quad B = \frac{g'_{\text{q}}{}^4 \Gamma^{\text{DM}}(m_{\text{DM}}, g_{\text{DM}})}{\Gamma^{\text{qq}}(g'_q) + \Gamma^{\text{DM}}(m'_{\text{DM}}, g_{\text{DM}})}. \quad (8.14)$$

With these relations, the limits for models with arbitrary m'_{DM} and g'_q can be obtained using the limits from the model with a given m_{DM} and g_{q} .

In Figure 92, exclusion regions for vector (top) and axial-vector (bottom) mediators are shown in the m_{DM} and M_{Med} plane for the model where $g_{\text{q}} = g_{\text{DM}} = 1.0$. As can be seen in the figure, the exclusion range for M_{Med} is practically independent of m_{DM} . This behavior is expected because in the model with $g_{\text{q}} = g_{\text{DM}} = 1.0$, the total width of the mediator is dominated by the width of quark decay channel due to the large number of possible quark flavors and colors.

The DM limits can also be transformed into mass limits for a leptophobic Z' boson. The limits for the Z' boson are equivalent to the limits of the vector and axial-vector DM mediators with $\Gamma^{\text{DM}} = 0$. The limits on the universal coupling g'_{q} between the Z' boson and quarks as a function of the Z' mass are shown in Figure 93. The limits are also compared to limits obtained by the dijet resonance searches at CMS [58, 131–133]. Most of the analyses, with the exception of χ_{dijet} and broad dijet, assume that the intrinsic width is negligible compared to the experimental resolution, and hence are valid for $\Gamma_{Z'}/M_{Z'} \leq 10\%$. The sensitivity of the narrow dijet resonance searches begins to degrade at the region where the widths in the quark decay channels increase to more than 10%, where the data-driven background estimation in the analyses becomes biased by the signal. The broad dijet analysis is valid for $\Gamma_{Z'}/M_{Z'} \leq 30\%$, and the χ_{dijet} analysis is valid for $\Gamma_{Z'}/M_{Z'} \leq 100\%$. For Z' mass between 2.0 and 4.6 TeV, this search excludes couplings with $1.0 \leq g'_{\text{q}} \leq 1.4$, which are not accessible via dijet resonance searches. The results of this thesis have been published in Ref. [126].

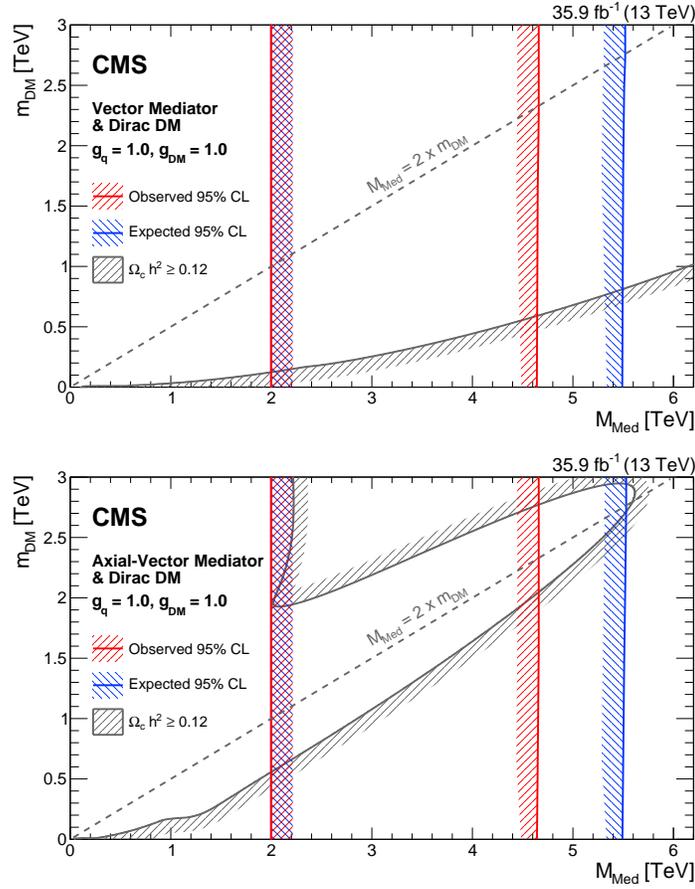


Figure 92: The 95% CL observed (red) and expected (blue) excluded regions in the plane of m_{DM} and M_{Med} , for a vector mediator (upper) and an axial-vector mediator (lower) for a DM benchmark model with $g_{\text{DM}} = g_{\text{q}} = 1.0$. Also shown in gray are constraints from astrophysical measurements [54] of the cosmological relic density of DM. The expected lower bound is not seen because it overlaps with the observed lower bound.

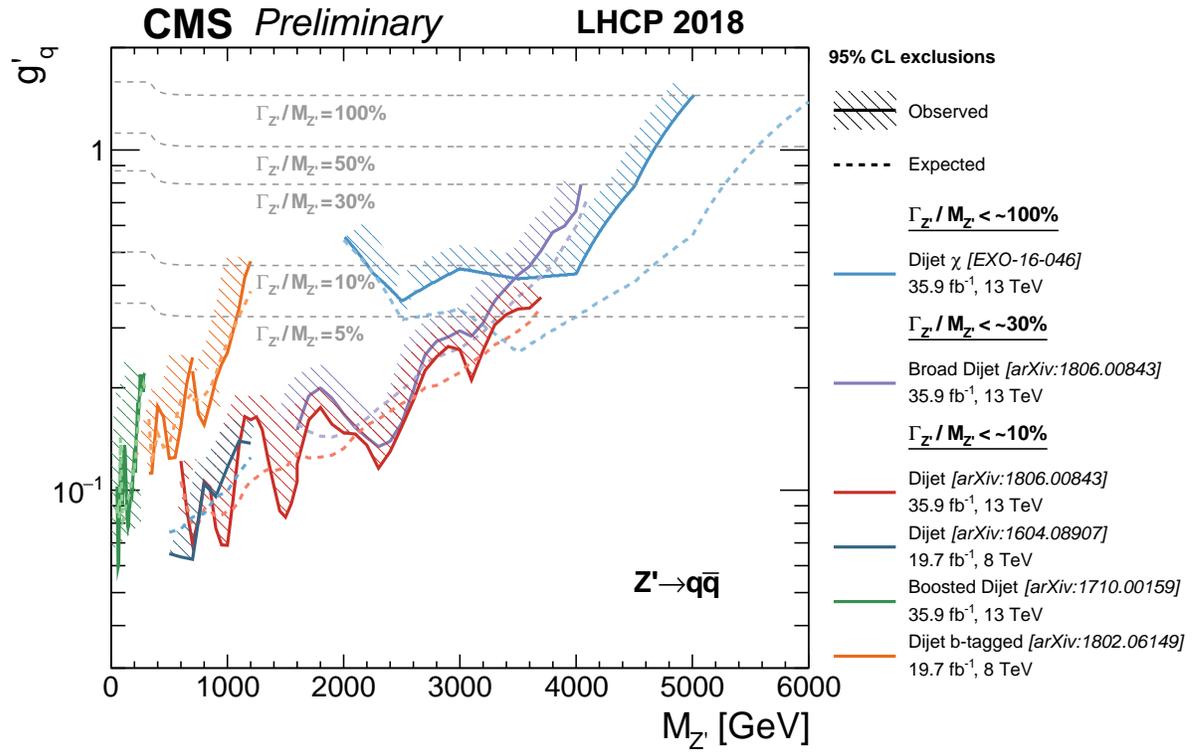


Figure 93: Limits on the universal coupling g'_q between a leptophobic Z' boson and quarks from various CMS dijet analyses. The expected limits are shown in dashed lines, and the corresponding observed limits are shown in solid lines. The hashed areas show the direction of the excluded area from the observed limits. The grey dashed lines show the g'_q values at fixed values of $\Gamma_{Z'}/M_{Z'}$.

CHAPTER 9

CONCLUSION

Normalized dijet angular distributions have been measured at $\sqrt{s} = 13$ TeV with the CMS detector over a wide range of dijet invariant masses. The distributions are found to be in agreement with predictions of perturbative QCD that include electroweak corrections and are used to set 95% confidence level lower limits on various new physics models containing quark contact interactions, extra spatial dimensions, quantum black holes, or dark matter. For the first time, lower limits between 2.0 and 4.6 TeV are set on the mass of a dark matter mediator for (axial-)vector mediators, for the universal quark coupling $1.0 \leq g_q \leq 1.4$. The lower limits for the contact interaction scale Λ range from 9.2 to 22.4 TeV. The lower limits on the ultraviolet cutoff in the Arkani-Hamed-Dimopoulos-Dvali model are in the range of 8.5–12 TeV, and are the most stringent limits available. Quantum black hole masses below 8.2 TeV are excluded in the model with six large extra spatial dimensions, and below 5.9 TeV in the Randall-Sundrum model with a single, warped extra dimension. To facilitate comparisons with the predictions of other models, the angular distributions, corrected to particle level, are published in HEPData.

APPENDICES

Appendix A

TRIGGER RATE ESTIMATIONS

During data taking, trigger rates exceeding the limit of 100 kHz in L1 and ~ 1500 Hz in HLT can result in a loss of data in the DAQ. It is therefore crucial for CMS to estimate the trigger rates before trigger menus are deployed online.

The trigger rates can be estimated using data and/or simulation. For the method that uses data, the trigger rate can be written as:

$$R_{\text{data}} = \frac{N_{\text{ev}}}{N_{\text{LS}} \times 23.3} N_{\text{prescale}}. \quad (\text{A.1})$$

where 23.3 seconds is the length of a luminosity section (LS), N_{LS} is the number of LSs in the data sample, N_{prescale} is the prescale factor of the zero bias trigger, and N_{ev} is the number of events triggered in the sample. The data sample used to estimate the trigger rates is the zero bias sample, which contains events triggered by the zero bias trigger that has no selection criteria imposed on the trigger objects. Since the zero bias trigger is prescaled, N_{prescale} needs to be applied in the rate estimation. N_{ev} is obtained by running the trigger emulation on the zero bias data sample and counting the number of triggered events.

Both the cumulative rate of a trigger menu and the rates for the individual triggers need to be estimated. The cumulative rate for a trigger menu does not equal to the sum of the rates of the individual triggers in the menu since one event can fire multiple triggers. The

Appendix A (Continued)

quantity that takes into account this multiple-counting effect is called the pure rate. If an event passes multiple triggers, it will be weighted by the inverse number of triggers it passes in the pure rate calculation. As the LHC instantaneous luminosity was increasing every week, the trigger rates had to be predicted in higher luminosity conditions using data collected at lower luminosities. This is achieved by linearly (quadratically) extrapolating the pure rate for single-object (multi-object) triggers.

For the method using the simulated events, the trigger rate can be written as:

$$R_{\text{MC}} = \frac{\sigma_{\text{MC}} \times i\mathcal{L} \times N_{\text{pass}}}{N_{\text{total}}}. \quad (\text{A.2})$$

where σ_{MC} is the cross section of the simulated process, $i\mathcal{L}$ is the desired instantaneous luminosity, N_{pass} is the number of events that pass the trigger, and N_{total} is the total number of events in the simulated sample. The zero bias events are commonly used for the L1T rate estimation since there are no trigger thresholds applied, while events from QCD, Drell-Yan, and W boson decay processes are used for the HLT. The MC samples are produced with a variety of pileup scenarios to ensure that the effects of pileup are properly accounted for in the simulations.

Some examples of the trigger rates are shown in Figures 94 to 96. In Figure 94, the rates for the L1 single electron/photon (EG) triggers with various E_{T} thresholds as a function of the values of the η restriction are shown. In Figure 95, the rates for the L1 single jet and H_{T} triggers are estimated from simulation and compared to data from a run with a similar number of pileup vertices. As can be seen in the figure, the rates in data and simulation agree within

Appendix A (Continued)

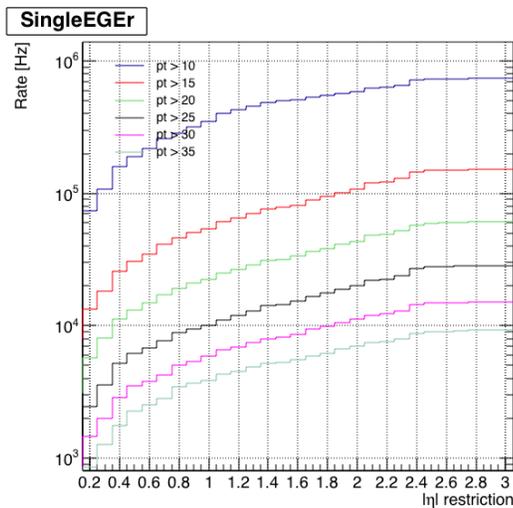


Figure 94: The rates for the L1 single EG η restricted triggers with various E_T thresholds as a function of the values of the η restriction.

25%. In Figure 96, a screenshot of the estimated HLT trigger rates for a subset of the full menu is presented. These rates are used for the HLT trigger menu development.

Appendix A (Continued)

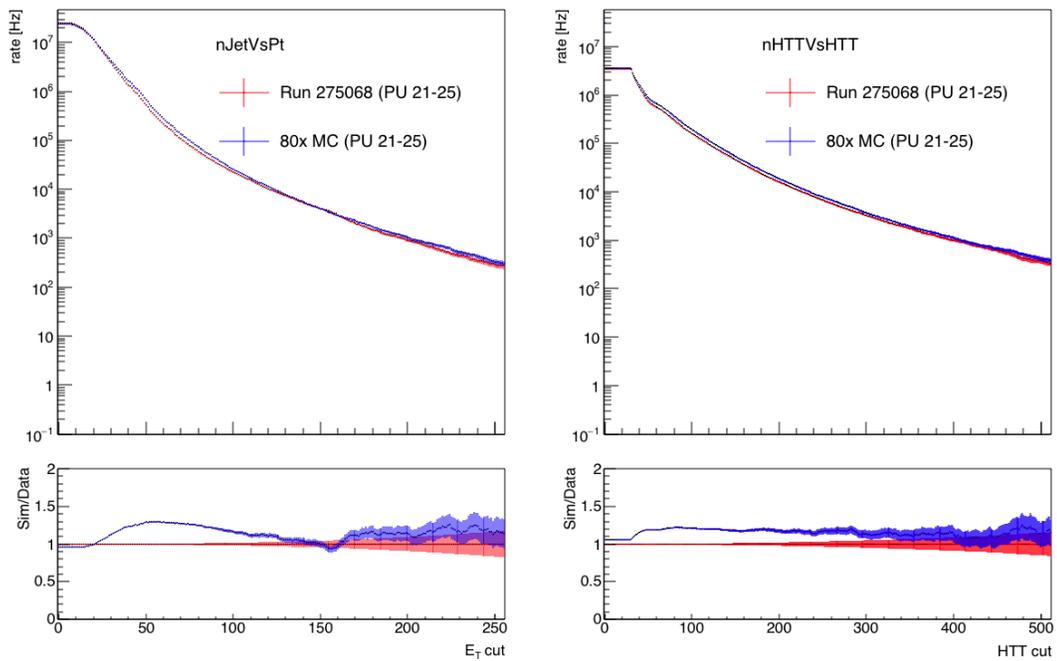


Figure 95: The rate for the L1 single jet (left) and H_T (right) trigger as a function of the trigger threshold estimated from data (red) and simulation (blue).

Appendix A (Continued)

HLT Paths	HLT PreScale	Spring 15 Rates
HLT_IsoMu24_eta2p1_v2	1	161.17 ± 5.00
HLT_Ele32_eta2p1_WPLoose_Gsf_v1	1	153.97 ± 5.97
HLT_IsoTkMu24_eta2p1_v2	1	144.68 ± 0.96
HLT_IsoMu27_v2	1	141.65 ± 4.08
HLT_IsoTkMu27_v2	1	130.09 ± 0.78
HLT_DoubleMu4_LowMassNonResonantTrk_Displaced_v2	1	107.72 ± 6.55
HLT_Ele32_eta2p1_WPTight_Gsf_v1	1	104.24 ± 5.10
HLT_PFHT800_v1	1	85.59 ± 6.95
HLT_AK8DiPFJet280_200_TrimMass30_BTagCSV0p45_v2	1	83.89 ± 6.35
HLT_Mu8_v2	2	78.50 ± 8.59
HLT_Mu45_eta2p1_v2	1	67.89 ± 2.86
HLT_DoubleJet90_Double30_TripleBTagCSV0p67_v2	1	61.09 ± 0.70
HLT_PFHT400_SixJet30_BTagCSV0p55_2PFBTagCSV0p72_v2	1	60.68 ± 0.79
HLT_DoubleMu4_JpsiTrk_Displaced_v2	1	58.57 ± 5.09
HLT_QuadPFJet_DoubleBTagCSV_VBF_Mqq240_v2	1	53.80 ± 0.81
HLT_Mu8_TrkIsoVVL_v2	2	53.32 ± 8.07
HLT_PFHT350_PFMET100_NoiseCleaned_v1	1	48.22 ± 6.35
HLT_PFHT750_4JetPt50_v1	1	47.62 ± 0.28
HLT_QuadJet45_TripleBTagCSV0p67_v2	1	47.59 ± 0.68
HLT_Mu50_v2	1	47.54 ± 2.85
HLT_DiCentralPFJet55_PFMET110_NoiseCleaned_v1	1	45.20 ± 6.97
HLT_AK8PFHT700_TrimR0p1PT0p03Mass50_v2	1	42.26 ± 0.45
HLT_PFJet450_v2	1	41.25 ± 6.94
HLT_AK8PFJet360_TrimMass30_v2	1	40.70 ± 6.94
HLT_Dimuon20_Jpsi_v2	1	37.47 ± 6.38
HLT_PFHT650_WideJetMJJ900DEtaJ1p5_v2	1	36.60 ± 6.94
HLT_PFHT450_SixJet40_PFBTagCSV0p72_v2	1	36.18 ± 0.49
HLT_RsqMR270_Rsq0p09_MR200_v1	1	35.52 ± 4.91
HLT_PFMETNoMu120_NoiseCleaned_PFMHTNoMu120_IDTight_v2	1	33.57 ± 6.42
HLT_CaloJet500_NoJetID_v2	1	32.38 ± 6.94
HLT_PFHT650_WideJetMJJ950DEtaJ1p5_v2	1	32.19 ± 6.94
HLT_Ele30WP60_SC4_Mass55_v2	2	31.63 ± 4.09
HLT_PFHT200_DiPFJetAve90_PFAAlphaT0p63_v1	1	31.58 ± 0.86
HLT_PFJet500_v2	1	31.11 ± 6.94
HLT_IsoMu16_eta2p1_CaloMET30_v2	1	30.50 ± 0.92
HLT_Mu17_Photon30_CaloldL_L1ISO_v2	1	29.41 ± 6.34

Figure 96: A part of the HLT menu with estimated rates.

Appendix B

CROSS-CHECKS FOR THE LIMITS

Several cross-checks have been performed to ensure that the final limits presented are correct. One cross-check was to compare the final limits extracted using detector-level χ_{dijet} distributions for the data and theoretical predictions to the limits extracted using the unfolded data distributions and particle-level theoretical predictions. The particle-level limits are shown in Figure 97 for the DM model and Table XIV for the CI, ADD, and QBH models. In the calculation, the correlations between M_{jj} bins in the unfolded distributions are ignored. Comparing Figure 97 and table Table XIV with Figure 91 and Table XIII, we see that for most models, the difference in the limits for the NP masses and scales are less than 3%. For the wide resonances (DM mediators with large g_q), the difference is larger (10%). This is due to the fact that wide resonances have signals across several M_{jj} ranges, while the CI, ADD, and QBH models only have signals in the highest mass bin(s).

In another cross-check, DM limits are extracted using particle-level predictions that are smeared using the double-sided Crystal Ball parameterization of the jet p_T resolution and compared to limits extracted using particle-level theoretical predictions folded using the analysis response matrix derived from PYTHIA (see Section 6.4). This is another check of the robustness of the limit extraction procedure against differences in the input M_{jj} spectrum. The results are shown in Figure 98. Comparing Figure 98 with Figure 91, we see that the two sets of limits agree within 5%.

Appendix B (Continued)

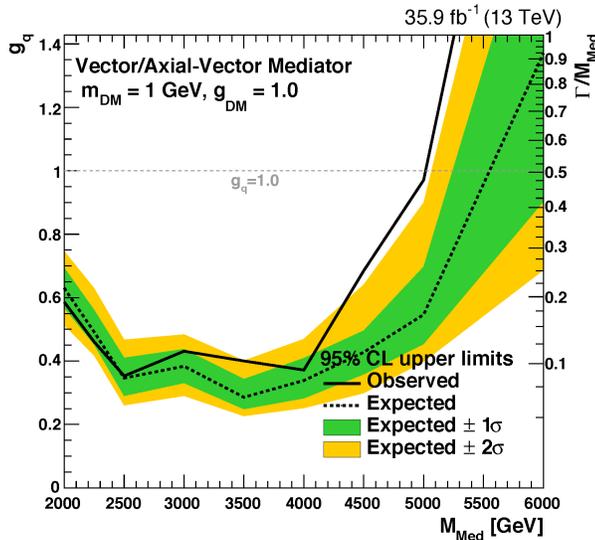


Figure 97: Same as Figure 91 but obtained using the unfolded distributions and particle-level predictions.

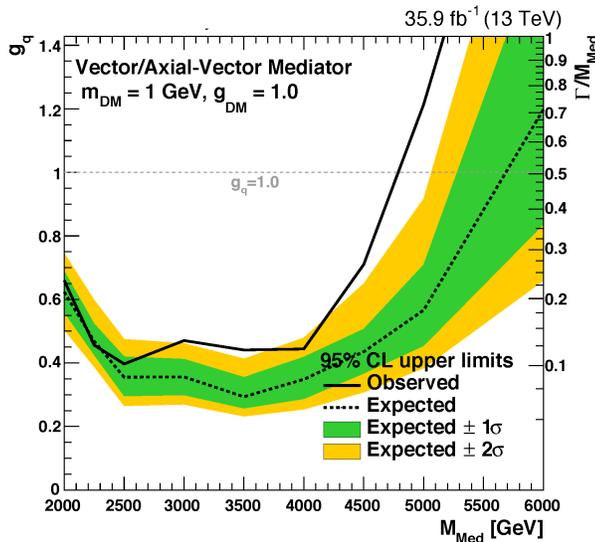


Figure 98: Same as Figure 91 but obtained using SM+DM prediction smeared with the double-sided Crystal Ball parameterization of the jet p_T resolution.

Appendix B (Continued)

TABLE XIV: Same as Table XIII but obtained using the unfolded distributions and particle-level predictions.

Model	Observed lower limit (TeV)	Expected lower limit (TeV)
$\Lambda_{\text{LL/RR}}^+$ (NLO)	12.7	14.8 ± 0.8
$\Lambda_{\text{LL/RR}}^-$ (NLO)	17.6	23.8 ± 3.1
Λ_{VV}^+ (NLO)	14.6	16.6 ± 0.9
Λ_{VV}^- (NLO)	22.5	31.2 ± 3.9
Λ_{AA}^+ (NLO)	14.7	16.7 ± 0.9
Λ_{AA}^- (NLO)	22.4	30.8 ± 3.6
$\Lambda_{\text{(V-A)}}^+$ (NLO)	9.2	11.6 ± 1.0
$\Lambda_{\text{(V-A)}}^-$ (NLO)	9.4	11.9 ± 1.1
ADD Λ_{T} (GRW)	10.3	11.7 ± 0.8
ADD M_{S} (HLZ) $n_{\text{ED}} = 2$	11.0	12.8 ± 0.9
ADD M_{S} (HLZ) $n_{\text{ED}} = 3$	12.2	13.7 ± 0.9
ADD M_{S} (HLZ) $n_{\text{ED}} = 4$	10.3	11.7 ± 0.8
ADD M_{S} (HLZ) $n_{\text{ED}} = 5$	9.3	10.5 ± 0.7
ADD M_{S} (HLZ) $n_{\text{ED}} = 6$	8.7	9.8 ± 0.5
QBH M_{QBH} (ADD6)	8.3	8.7 ± 0.3
QBH M_{QBH} (RS1)	6.0	6.5 ± 0.4

Appendix B (Continued)

Finally, a signal injection test for the DM model is performed. This test is a sanity check to make sure that if a signal is present, it can be found by the statistical method used in this analysis. DM signals corresponding to $g_q = 0.75$ and $M_{\text{Med}} = 2.0, 2.25, 3.0, 3.5, 4.0,$ and 4.5 TeV are added to the data. Maximum likelihood fits for the signal+background distributions to the modified data distributions are performed. The resultant fitted signal strengths (see Section 8.2.1) are listed in Table XV and are compared to those from the fits to the unmodified data distributions. The fitted signal strengths of the modified data distributions converge to 1 as expected. The limits for the modified data distributions are shown in Figure 99. As expected, for the modified data distributions, the observed limits on g_q are greater than 0.75.

TABLE XV: Fitted signal strengths from fits to the data+DM distributions compared to the fitted signal strengths from the fits to the data-only distributions.

Mediator Mass (TeV)	Fitted μ (Data)	Fitted μ (Data+NP)
2.0	0.21	1.02
2.25	0.14	1.01
2.5	0.29	1.06
3.0	0.51	1.12
3.5	0.55	1.14
4.0	0.58	1.15
4.5	0.80	1.27

Appendix B (Continued)

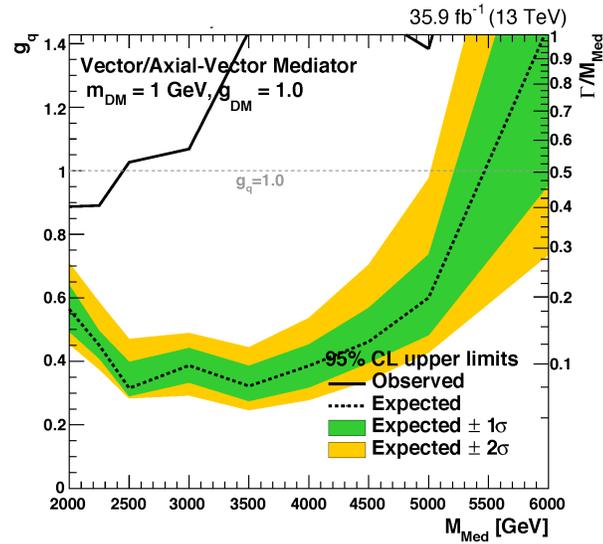


Figure 99: Results of the signal injection test. Same as Figure 91 but replacing the data distributions with the data+DM ($g_q = 0.75$) distributions.

Appendix C

COPYRIGHT POLICY OF EUROPEAN PHYSICS JOURNAL C

The copyright policy of European Physics Journal C is stated in the following webpage: <https://epjc.epj.org/epjc-copyright-form>. A screenshot of the webpage is provided in Figure 100.

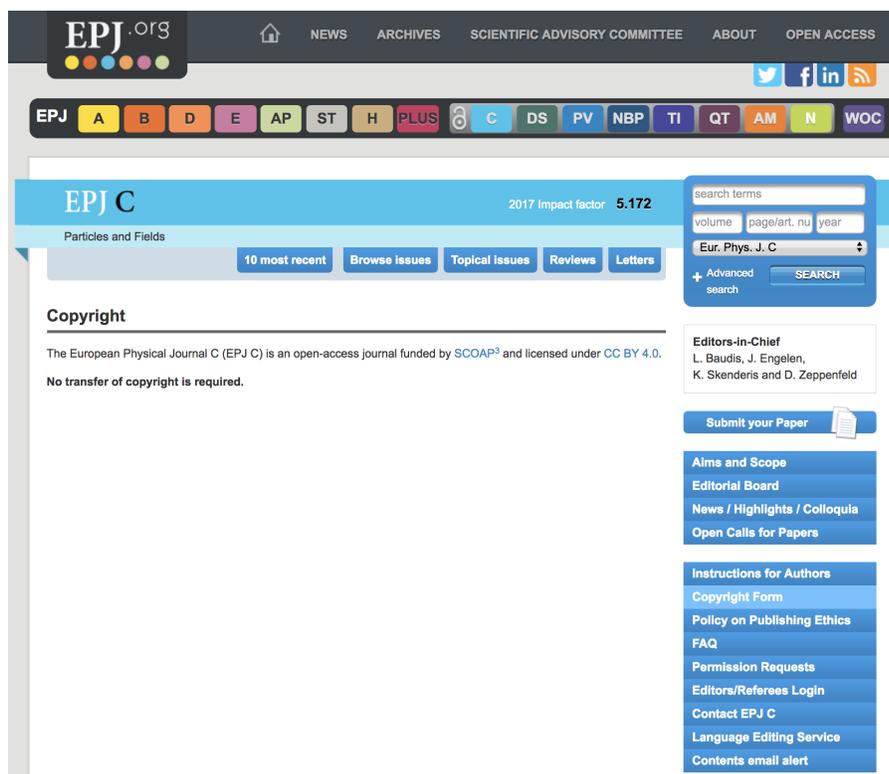


Figure 100: A screenshot of the webpage stating the copyright policy of the European Physics Journal C.

CITED LITERATURE

- [1] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett.* **B 716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett.* **B 716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- [3] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001.
- [4] 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. [,157(1964)].
- [5] P. W. Higgs, “Broken symmetries, massless particles and gauge fields”, *Phys. Lett.* **12** (1964) 132–133, doi:10.1016/0031-9163(64)91136-9.
- [6] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508. [,160(1964)].
- [7] 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. [,162(1964)].
- [8] P. W. Higgs, “Spontaneous Symmetry Breakdown without Massless Bosons”, *Phys. Rev.* **145** (1966) 1156–1163, doi:10.1103/PhysRev.145.1156.
- [9] T. W. B. Kibble, “Symmetry breaking in nonAbelian gauge theories”, *Phys. Rev.* **155** (1967) 1554–1561, doi:10.1103/PhysRev.155.1554. [,165(1967)].
- [10] H. Terazawa, “Subquark model of leptons and quarks”, *Phys. Rev. D* **22** (1980) 184, doi:10.1103/PhysRevD.22.184.

- [11] E. Eichten, K. Lane, and M. Peskin, “New tests for quark and lepton substructure”, *Phys. Rev. Lett.* **50** (1983) 811, doi:10.1103/PhysRevLett.50.811.
- [12] E. Eichten, I. Hinchliffe, K. Lane, and C. Quigg, “Supercollider physics”, *Rev. Mod. Phys.* **56** (1984) 579, doi:10.1103/RevModPhys.56.579.
- [13] UA1 Collaboration, “Angular distributions for high mass jet pairs and a limit on the energy scale of compositeness for quarks from the cern $p\bar{p}$ collider”, *Phys. Lett. B* **177** (1986) 244, doi:10.1016/0370-2693(86)91065-8.
- [14] UA2 Collaboration, “Measurement of the $s^{**}(1/2)$ Dependence of Jet Production at the CERN anti-p p Collider”, *Phys. Lett.* **160B** (1985) 349, doi:10.1016/0370-2693(85)91341-3.
- [15] CDF Collaboration, “Measurement of the inclusive jet cross-section in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV”, *Phys. Rev. Lett.* **62** (1989) 613, doi:10.1103/PhysRevLett.62.613.
- [16] CDF Collaboration, “The Dijet angular distribution in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”, *Phys. Rev. Lett.* **69** (1992) 2896–2900, doi:10.1103/PhysRevLett.69.2896.
- [17] CDF Collaboration, “Measurement of dijet angular distributions at CDF”, *Phys. Rev. Lett.* **77** (1996) 5336, doi:10.1103/PhysRevLett.77.5336, arXiv:hep-ex/9609011. [Erratum:10.1103/PhysRevLett.78.4307].
- [18] D0 Collaboration, “Measurement of dijet angular distributions and search for quark compositeness”, *Phys. Rev. Lett.* **80** (1998) 666–671, doi:10.1103/PhysRevLett.80.666, arXiv:hep-ex/9707016.
- [19] D0 Collaboration, “The dijet mass spectrum and a search for quark compositeness in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV”, *Phys. Rev. Lett.* **82** (1999) 2457–2462, doi:10.1103/PhysRevLett.82.2457, arXiv:hep-ex/9807014.
- [20] D0 Collaboration, “High- p_T jets in $\bar{p}p$ collisions at $\sqrt{s} = 630$ and 1800 gev”, *Phys. Rev. D* **64** (2001) 032003, doi:10.1103/PhysRevD.64.032003, arXiv:hep-ex/0012046.
- [21] D0 Collaboration, “Measurement of dijet angular distributions at $\sqrt{s} = 1.96$ tev and searches for quark compositeness and extra spatial dimensions”, *Phys. Rev. Lett.* **103** (2009) 191803, doi:10.1103/PhysRevLett.103.191803, arXiv:0906.4819.

- [22] ATLAS Collaboration, “Search for new physics in dijet mass and angular distributions in pp collisions at $\sqrt{s} = 7$ TeV measured with the ATLAS detector”, *New J. Phys.* **13** (2011) 053044, doi:10.1088/1367-2630/13/5/053044, arXiv:1103.3864.
- [23] ATLAS Collaboration, “Search for quark contact interactions in dijet angular distributions in pp collisions at $\sqrt{s} = 7$ TeV measured with the ATLAS detector”, *Phys. Lett. B* **694** (2010) 327, doi:10.1016/j.physletb.2010.10.021, arXiv:1009.5069.
- [24] ATLAS Collaboration, “ATLAS search for new phenomena in dijet mass and angular distributions using pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **01** (2013) 029, doi:10.1007/JHEP01(2013)029, arXiv:1210.1718.
- [25] ATLAS Collaboration, “Search for new phenomena in dijet angular distributions in proton-proton collisions at $\sqrt{s} = 8$ TeV measured with the ATLAS detector”, *Phys. Rev. Lett.* **114** (2015) 221802, doi:10.1103/PhysRevLett.114.221802, arXiv:1504.00357.
- [26] 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. B* **754** (2016) 302, doi:10.1016/j.physletb.2016.01.032, arXiv:1512.01530.
- [27] ATLAS Collaboration, “Search for new phenomena in dijet events using 37 fb⁻¹ of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. D* **96** (2017) 052004, doi:10.1103/PhysRevD.96.052004, arXiv:1703.09127.
- [28] CMS Collaboration, “Search for quark compositeness with the dijet centrality ratio in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* **105** (2010) 262001, doi:10.1103/PhysRevLett.105.262001, arXiv:1010.4439.
- [29] CMS Collaboration, “Measurement of dijet angular distributions and search for quark compositeness in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* **106** (2011) 201804, doi:10.1103/PhysRevLett.106.201804, arXiv:1102.2020.
- [30] CMS Collaboration, “Search for quark compositeness in dijet angular distributions from pp collisions at $\sqrt{s} = 7$ TeV”, *JHEP* **05** (2012) 055, doi:10.1007/JHEP05(2012)055, arXiv:1202.5535.

- [31] CMS Collaboration, “Search for contact interactions using the inclusive jet p_T spectrum in pp collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. D* **87** (2013) 052017, doi:10.1103/PhysRevD.87.052017, arXiv:1301.5023.
- [32] CMS Collaboration, “Search for quark contact interactions and extra spatial dimensions using dijet angular distributions in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett. B* **746** (2015) 79, doi:10.1016/j.physletb.2015.04.042, arXiv:1411.2646.
- [33] CMS Collaboration, “Search for new physics with dijet angular distributions in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2017) 013, doi:10.1007/JHEP07(2017)013, arXiv:1703.09986.
- [34] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, “The hierarchy problem and new dimensions at a millimeter”, *Phys. Lett. B* **429** (1998) 263, doi:10.1016/S0370-2693(98)00466-3, arXiv:hep-ph/9803315.
- [35] N. Arkani-Hamed, S. Dimopoulos, and G. Dvali, “Phenomenology, astrophysics and cosmology of theories with sub-millimeter dimensions and TeV scale quantum gravity”, *Phys. Rev. D* **59** (1999) 086004, doi:10.1103/PhysRevD.59.086004, arXiv:hep-ph/9807344.
- [36] T. Kaluza, “Zum Unitätsproblem der Physik”, *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1921** (1921) 966–972, arXiv:1803.08616.
- [37] O. Klein, “Quantum Theory and Five-Dimensional Theory of Relativity. (In German and English)”, *Z. Phys.* **37** (1926) 895–906, doi:10.1007/BF01397481. [,76(1926)].
- [38] G. F. Giudice, R. Rattazzi, and J. D. Wells, “Quantum gravity and extra dimensions at high-energy colliders”, *Nucl. Phys. B* **544** (1999) 3, doi:10.1016/S0550-3213(99)00044-9, arXiv:hep-ph/9811291.
- [39] J. L. Hewett, “Indirect collider signals for extra dimensions”, *Phys. Rev. Lett.* **82** (1999) 4765–4768, doi:10.1103/PhysRevLett.82.4765, arXiv:hep-ph/9811356.
- [40] T. Han, J. D. Lykken, and R.-J. Zhang, “Kaluza-Klein states from large extra dimensions”, *Phys. Rev. D* **59** (1999) 105006, doi:10.1103/PhysRevD.59.105006, arXiv:hep-ph/9811350.

- [41] K. Cheung and G. Landsberg, “Drell-Yan and diphoton production at hadron colliders and low scale gravity model”, *Phys. Rev. D* **62** (2000) 076003, doi:10.1103/PhysRevD.62.076003, arXiv:hep-ph/9909218.
- [42] J. Murata and S. Tanaka, “A review of short-range gravity experiments in the LHC era”, *Class. Quant. Grav.* **32** (2015), no. 3, 033001, doi:10.1088/0264-9381/32/3/033001, arXiv:1408.3588.
- [43] L. Randall and R. Sundrum, “Large mass hierarchy from a small extra dimension”, *Phys. Rev. Lett.* **83** (1999) 3370, doi:10.1103/PhysRevLett.83.3370, arXiv:hep-ph/9905221.
- [44] L. Randall and R. Sundrum, “An alternative to compactification”, *Phys. Rev. Lett.* **83** (1999) 4690, doi:10.1103/PhysRevLett.83.4690, arXiv:hep-th/9906064.
- [45] R. C. Myers and M. J. Perry, “Black Holes in Higher Dimensional Space-Times”, *Annals Phys.* **172** (1986) 304, doi:10.1016/0003-4916(86)90186-7.
- [46] T. G. Rizzo, “Warped phenomenology of higher-derivative gravity”, *Journal of High Energy Physics* **2005** (2005), no. 01, 028.
- [47] T. G. Rizzo, “TeV-scale black holes in warped higher-curvature gravity”, arXiv:hep-ph/0510420.
- [48] S. W. Hawking, “Particle creation by black holes”, *Commun. Math. Phys.* **43** (1975) 199, doi:10.1007/BF02345020. [Erratum: <https://projecteuclid.org/euclid.cmp/1103899590>].
- [49] X. Calmet, W. Gong, and S. D. H. Hsu, “Colorful quantum black holes at the LHC”, *Phys. Lett. B* **668** (2008) 20, doi:10.1016/j.physletb.2008.08.011, arXiv:0806.4605.
- [50] P. Meade and L. Randall, “Black holes and quantum gravity at the LHC”, *JHEP* **05** (2008) 003, doi:10.1088/1126-6708/2008/05/003, arXiv:0708.3017.
- [51] D. M. Gingrich, “Quantum black holes with charge, colour, and spin at the LHC”, *J. Phys. G* **37** (2010) 105008, doi:10.1088/0954-3899/37/10/105008, arXiv:0912.0826.

- [52] D. Clowe et al., “A direct empirical proof of the existence of dark matter”, *Astrophys. J.* **648** (2006) L109–L113, doi:10.1086/508162, arXiv:astro-ph/0608407.
- [53] G. W. Angus, B. Famaey, and H. Zhao, “Can MOND take a bullet? Analytical comparisons of three versions of MOND beyond spherical symmetry”, *Mon. Not. Roy. Astron. Soc.* **371** (2006) 138, doi:10.1111/j.1365-2966.2006.10668.x, arXiv:astro-ph/0606216.
- [54] Planck Collaboration, “Planck 2015 results. XIII. Cosmological parameters”, *Astron. Astrophys.* **594** (2016) A13, doi:10.1051/0004-6361/201525830, arXiv:1502.01589.
- [55] J. Abdallah et al., “Simplified Models for Dark Matter Searches at the LHC”, *Phys. Dark Univ.* **9-10** (2015) 8–23, doi:10.1016/j.dark.2015.08.001, arXiv:1506.03116.
- [56] G. Busoni et al., “Recommendations on presenting LHC searches for missing transverse energy signals using simplified s-channel models of dark matter”, (2016). arXiv:1603.04156.
- [57] A. Albert et al., “Recommendations of the LHC Dark Matter Working Group: Comparing LHC searches for heavy mediators of dark matter production in visible and invisible decay channels”, arXiv:1703.05703.
- [58] CMS Collaboration, “Search for narrow and broad dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter mediators and other new particles”, *JHEP* **08** (2018) 130, doi:10.1007/JHEP08(2018)130, arXiv:1806.00843.
- [59] ATLAS Collaboration, “Search for new phenomena in final states with an energetic jet and large missing transverse momentum in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector”, *Phys. Rev. D* **94** (2016) 032005, doi:10.1103/PhysRevD.94.032005, arXiv:1604.07773.
- [60] CMS Collaboration, “Search for dark matter produced with an energetic jet or a hadronically decaying W or Z boson at $\sqrt{s} = 13$ TeV”, *JHEP* **07** (2017) 014, doi:10.1007/JHEP07(2017)014, arXiv:1703.01651.
- [61] CMS Collaboration, “Search for new physics in final states with an energetic jet or a hadronically decaying W or Z boson and transverse momentum imbalance at

- $\sqrt{s} = 13 \text{ TeV}$ ", *Phys. Rev.* **D97** (2018), no. 9, 092005, doi:10.1103/PhysRevD.97.092005, arXiv:1712.02345.
- [62] Particle Data Group Collaboration, "Review of Particle Physics", *Chin. Phys.* **C40** (2016), no. 10, 100001, doi:10.1088/1674-1137/40/10/100001.
- [63] S. Dulat et al., "New parton distribution functions from a global analysis of quantum chromodynamics", *Phys. Rev.* **D93** (2016), no. 3, 033006, doi:10.1103/PhysRevD.93.033006, arXiv:1506.07443.
- [64] T. Gleisberg et al., "Event generation with SHERPA 1.1", *JHEP* **02** (2009) 007, doi:10.1088/1126-6708/2009/02/007, arXiv:0811.4622.
- [65] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:10.1088/1126-6708/2006/05/026, arXiv:hep-ph/0603175.
- [66] T. Sjöstrand, S. Mrenna, and P. Skands, "A brief introduction to PYTHIA 8.1", *Comput. Phys. Comm.* **178** (2008) 852, doi:10.1016/j.cpc.2008.01.036, arXiv:0710.3820.
- [67] M. Bähr et al., "Herwig++ physics and manual", *Eur. Phys. J. C* **58** (2008) 639, doi:10.1140/epjc/s10052-008-0798-9, arXiv:0803.0883.
- [68] J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J.* **C76** (2016), no. 4, 196, doi:10.1140/epjc/s10052-016-4018-8, arXiv:1512.01178.
- [69] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics", *Phys. Rept.* **97** (1983) 31–145, doi:10.1016/0370-1573(83)90080-7.
- [70] R. D. Field and S. Wolfram, "A QCD Model for $e^+ e^-$ Annihilation", *Nucl. Phys.* **B213** (1983) 65–84, doi:10.1016/0550-3213(83)90175-X.
- [71] T. Sjostrand and M. van Zijl, "A Multiple Interaction Model for the Event Structure in Hadron Collisions", *Phys. Rev.* **D36** (1987) 2019, doi:10.1103/PhysRevD.36.2019.
- [72] T. Sjostrand and P. Z. Skands, "Multiple interactions and the structure of beam remnants", *JHEP* **03** (2004) 053, doi:10.1088/1126-6708/2004/03/053, arXiv:hep-ph/0402078.

- [73] T. Sjostrand and P. Z. Skands, “Transverse-momentum-ordered showers and interleaved multiple interactions”, *Eur. Phys. J.* **C39** (2005) 129–154, doi:10.1140/epjc/s2004-02084-y, arXiv:hep-ph/0408302.
- [74] M. Bahr, S. Gieseke, and M. H. Seymour, “Simulation of multiple partonic interactions in Herwig++”, *JHEP* **07** (2008) 076, doi:10.1088/1126-6708/2008/07/076, arXiv:0803.3633.
- [75] 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.
- [76] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [77] M. H. Seymour and A. Siodmok, “Constraining MPI models using σ_{eff} and recent tevatron and LHC underlying event data”, *JHEP* **10** (2013) 113, doi:10.1007/JHEP10(2013)113, arXiv:1307.5015.
- [78] J. Alwall et al., “MadGraph 5 : Going Beyond”, *JHEP* **06** (2011) 128, doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.
- [79] 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* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [80] Z. Nagy, “Three-jet cross sections in hadron hadron collisions at next-to-leading order”, *Phys. Rev. Lett.* **88** (2002) 122003, doi:10.1103/PhysRevLett.88.122003, arXiv:hep-ph/0110315.
- [81] J. Gao et al., “Next-to-leading QCD effect to the quark compositeness search at the LHC”, *Phys. Rev. Lett.* **106** (2011) 142001, doi:10.1103/PhysRevLett.106.142001, arXiv:1101.4611.
- [82] D. M. Gingrich, “Monte Carlo event generator for black hole production and decay in proton-proton collisions”, *Comput. Phys. Commun.* **181** (2010) 1917, doi:10.1016/j.cpc.2010.07.027, arXiv:0911.5370.

- [83] B. Flaughner and K. Meier, “A Compilation of jet finding algorithms”, in *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90) Snowmass, Colorado, June 25-July 13, 1990*, pp. 0128–133. 1990.
- [84] K. Rabbertz, “Jet Physics at the LHC”, *Springer Tracts Mod. Phys.* **268** (2017) pp.1–214, doi:10.1007/978-3-319-42115-5.
- [85] S. D. Ellis and D. E. Soper, “Successive combination jet algorithm for hadron collisions”, *Phys. Rev.* **D48** (1993) 3160–3166, doi:10.1103/PhysRevD.48.3160, arXiv:hep-ph/9305266.
- [86] M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, in *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, pp. 270–279. 1998. arXiv:hep-ph/9907280.
- [87] M. Cacciari, G. P. Salam, and G. Soyez, “The Anti-k(t) jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [88] ATLAS and CMS Collaboration, “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”, *Phys. Rev. Lett.* **114** (2015) 191803, doi:10.1103/PhysRevLett.114.191803, arXiv:1503.07589.
- [89] CMS Collaboration, “The CMS Experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [90] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [91] CMS Collaboration, “2017 tracking performance plots”, Technical Report CMS-DP-2017-015, CERN, Apr, 2017.
- [92] CMS Collaboration, “2015 ECAL detector performance plots”, technical report, Dec, 2015.
- [93] CMS Collaboration, “Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data”, *JINST* **5** (2010) T03012, doi:10.1088/1748-0221/5/03/T03012, arXiv:0911.4991.

- [94] CMS Collaboration, “The CMS trigger system”, *JINST* **12** (2017), no. 01, P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.
- [95] CMS collaboration Collaboration, “CMS Technical Design Report for the Level-1 Trigger Upgrade”, Technical Report CERN-LHCC-2013-011, CMS-TDR-12, Jun, 2013.
- [96] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework”, *Nucl. Instrum. Meth.* **A389** (1997) 81–86, doi:10.1016/S0168-9002(97)00048-X.
- [97] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017), no. 10, P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [98] 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.
- [99] P. Billoir and S. Qian, “Simultaneous pattern recognition and track fitting by the Kalman filtering method”, *Nucl. Instrum. Meth.* **A294** (1990) 219–228, doi:10.1016/0168-9002(90)91835-Y.
- [100] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC”,.
- [101] CMS Collaboration, “Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 08, P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.
- [102] CMS Collaboration Collaboration, “Pileup Removal Algorithms”, Technical Report CMS-PAS-JME-14-001, CERN, Geneva, 2014.
- [103] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [104] CMS Collaboration, “Jet energy scale and resolution performances with 13 TeV data”, CMS Detector Performance Summary CERN-DP-2016-020, 2016.

- [105] CMS Collaboration, “Jet algorithms performance in 13 TeV data”, CMS Physics Analysis Summary CMS-PAS-JME-16-003, 2017.
- [106] CMS Collaboration, “Jet performance in pp collisions at 7 tev”, *CMS Physics Analysis Summary* **CMS-PAS-JME-10-003** (2010).
- [107] CMS Collaboration, “Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV and constraints on parton distribution functions”, *Eur. Phys. J.* **C77** (2017), no. 11, 746, doi:10.1140/epjc/s10052-017-5286-7, arXiv:1705.02628.
- [108] GEANT4 Collaboration, “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [109] M. J. Oreglia, “A study of the reactions $\psi' \rightarrow \gamma\gamma\psi$ ”. PhD thesis, Stanford University, 1980. SLAC Report SLAC-R-236.
- [110] T. Auye, “Unfolding algorithms and tests using RooUnfold”, in *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN, Geneva, Switzerland 17-20 January 2011*, pp. 313–318, CERN. CERN, Geneva, 2011. arXiv:1105.1160. doi:10.5170/CERN-2011-006.313.
- [111] G. D’Agostini, “A multidimensional unfolding method based on Bayes’ theorem”, *Nucl. Instrum. Meth. A* **362** (1995) 487, doi:10.1016/0168-9002(95)00274-X.
- [112] 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.
- [113] T. Kluge, K. Rabbertz, and M. Wobisch, “Fast pQCD calculations for PDF fits”, in *DIS 2006, 20-24 Apr 2006*, p. 483. Tsukuba, Japan, April, 2006. arXiv:hep-ph/0609285.
- [114] M. Cacciari et al., “The t anti-t cross-section at 1.8-TeV and 1.96-TeV: A Study of the systematics due to parton densities and scale dependence”, *JHEP* **04** (2004) 068, doi:10.1088/1126-6708/2004/04/068, arXiv:hep-ph/0303085.

- [115] A. Banfi, G. P. Salam, and G. Zanderighi, “Phenomenology of event shapes at hadron colliders”, *JHEP* **06** (2010) 038, doi:10.1007/JHEP06(2010)038, arXiv:1001.4082.
- [116] D. Stump et al., “Uncertainties of predictions from parton distribution functions. 1. The Lagrange multiplier method”, *Phys. Rev.* **D65** (2001) 014012, doi:10.1103/PhysRevD.65.014012, arXiv:hep-ph/0101051.
- [117] J. Pumplin et al., “Uncertainties of predictions from parton distribution functions. 2. The Hessian method”, *Phys. Rev.* **D65** (2001) 014013, doi:10.1103/PhysRevD.65.014013, arXiv:hep-ph/0101032.
- [118] J. Butterworth et al., “PDF4LHC recommendations for LHC run ii”, *J. Phys. G* **43** (2016) 023001, doi:10.1088/0954-3899/43/2/023001, arXiv:1510.03865.
- [119] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne, “Parton distributions in the LHC era: MMHT 2014 PDFs”, *Eur. Phys. J. C* **75** (2015) 204, doi:10.1140/epjc/s10052-015-3397-6, arXiv:1412.3989.
- [120] NNPDF Collaboration, “Parton distributions for the LHC run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [121] J. Gao and P. Nadolsky, “A meta-analysis of parton distribution functions”, *JHEP* **07** (2014) 035, doi:10.1007/JHEP07(2014)035, arXiv:1401.0013.
- [122] S. Carrazza et al., “An unbiased hessian representation for monte carlo PDFs”, *Eur. Phys. J. C* **75** (2015) 369, doi:10.1140/epjc/s10052-015-3590-7, arXiv:1505.06736.
- [123] S. Dittmaier, A. Huss, and C. Speckner, “Weak radiative corrections to dijet production at hadron colliders”, *JHEP* **11** (2012) 095, doi:10.1007/JHEP11(2012)095, arXiv:1210.0438.
- [124] M. Backovic, K. Kong, and M. McCaskey, “MadDM v.1.0: Computation of dark matter relic abundance using MadGraph5”, *Phys. Dark Univ.* **5-6** (2014) 18, doi:10.1016/j.dark.2014.04.001, arXiv:1308.4955.
- [125] M. Backovic et al., “Direct detection of dark matter with MadDM v.2.0”, *Phys. Dark Univ.* **9-10** (2015) 37, doi:10.1016/j.dark.2015.09.001, arXiv:1505.04190.

- [126] CMS Collaboration, “Search for new physics in dijet angular distributions using proton–proton collisions at $\sqrt{s} = 13$ TeV and constraints on dark matter and other models”, *Eur. Phys. J. C* **78** (2018), no. 9, 789, doi:10.1140/epjc/s10052-018-6242-x, arXiv:1803.08030.
- [127] R. Cousins, “Saturated model for goodness of fit”, doi:http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.
- [128] 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.
- [129] A. L. Read, “Presentation of search results: the cl_s technique”, *J. Phys. G: Nucl. Part. Phys.* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [130] 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: 10.1140/epjc/s10052-013-2501-z].
- [131] CMS Collaboration, “Search for narrow resonances in dijet final states at $\sqrt{s} = 8$ TeV with the novel CMS technique of data scouting”, *Phys. Rev. Lett.* **117** (2016), no. 3, 031802, doi:10.1103/PhysRevLett.117.031802, arXiv:1604.08907.
- [132] CMS Collaboration, “Search for low mass vector resonances decaying into quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **01** (2018) 097, doi:10.1007/JHEP01(2018)097, arXiv:1710.00159.
- [133] CMS Collaboration, “Search for narrow resonances in the b-tagged dijet mass spectrum in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev. Lett.* **120** (2018), no. 20, 201801, doi:10.1103/PhysRevLett.120.201801, arXiv:1802.06149.