



UNIVERSITY OF IOANNINA
SCHOOL OF SCIENCES
DEPARTMENT OF PHYSICS



Measurements of jet production cross
sections and studies of Quantum
Chromodynamics with the CMS experiment
at the LHC

PhD THESIS

Polidamas Georgios Kosmoglou Kioseoglou

IOANNINA 2025



ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΦΥΣΙΚΗΣ



Μετρήσεις ενεργών διατομών
παραγωγής πιδάκων σωματιδίων και
μελέτες κβαντικής χρωμοδυναμικής στο
πείραμα CMS του LHC

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Πολυδάμας Γεώργιος Κοσμόγλου Κιοσέογλου



H.F.R.I.
Hellenic Foundation for
Research & Innovation

The research work was supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 3rd Call for HFRI PhD Fellowships (Fellowship Number: 83154).



ΕΛΙΔΕΚ
Ελληνικό Ίδρυμα Έρευνας & Καινοτομίας

Η ερευνητική εργασία υποστηρίχθηκε από το Ελληνικό Ίδρυμα Έρευνας και Καινοτομίας (ΕΛ.ΙΔ.Ε.Κ) στο πλαίσιο της “3ης Προκήρυξης ΕΛ.ΙΔ.Ε.Κ για Υποψήφιος/ες Διδάκτορες” (Αριθμός Υποτροφίας: 83154).

Three-member Advisory Committee

1. Panagiotis Kokkas (Supervisor)
Professor, Physics Department, University of Ioannina
2. Constantinos Fountas
Professor, Physics Department, University of Ioannina
3. Ioannis Papadopoulos
Associate Professor, Physics Department, University of Ioannina

Seven-member Advisory Committee

1. Panagiotis Kokkas (Supervisor)
Professor, Physics Department, University of Ioannina
2. Constantinos Fountas
Professor, Physics Department, University of Ioannina
3. Ioannis Papadopoulos
Associate Professor, Physics Department, University of Ioannina
4. Ioannis Strologas
Assistant Professor, Physics Department, University of Ioannina
5. Constantinos Vellidis
Associate Professor, Physics Department, National and Kapodistrian University of Athens
6. Konstantinos Theofilatos
Associate Professor, Physics Department, National and Kapodistrian University of Athens
7. Konstantinos Kousouris
Associate Professor, School of Applied Mathematical and Physical Sciences, National Technical University of Athens

Abstract

This dissertation presents a measurement of the double differential inclusive dijet cross section using proton-proton collision data collected in 2016, 2017, and 2018 by the CMS experiment at the CERN LHC. Collisions were performed at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ with integrated luminosities of 33.5, 41.5 and, 59.3 fb^{-1} , respectively. Jets were reconstructed using the anti- k_T clustering algorithm with a distance parameter of $R = 0.8$. Cross sections are measured as a function of the dijet invariant mass $m_{1,2}$ and the maximum absolute rapidity $|y|_{max}$ of the dijet system, defined by the two leading jets in transverse momentum p_T . The distributions are unfolded to correct for detector effects and compared to fixed order predictions at next-to-next-to-leading order in perturbative quantum chromodynamics, including corrections for non-perturbative effects and extra electroweak radiation. A global fit is performed using the 2016 dataset to constrain the proton parton distribution functions (PDFs) and determine the strong coupling constant at the scale of the Z -boson mass, yielding a result of $\alpha_S(m_Z) = 0.1179 \pm 0.0019$. The measured cross sections show good agreement with the theoretical predictions, leading to further constraints on PDFs, while the extracted $\alpha_S(m_Z)$ value aligns with the 2022 particle data group world average.

Περίληψη

Αυτή η διατριβή παρουσιάζει μία μέτρηση της διπλά διαφορικής ενεργούς διατομής παραγωγής τουλάχιστον δύο πιδάκων σωματιδίων χρησιμοποιώντας δεδομένα από συγκρούσεις πρωτονίων-πρωτονίων που συλλέχθηκαν τα έτη 2016, 2017, και 2018 από το πείραμα CMS στον επιταχυντή LHC του CERN. Οι συγκρούσεις πραγματοποιούνται σε ενέργεια κέντρου μάζας $\sqrt{s} = 13 \text{ TeV}$ και αναλογούν σε ολοκληρωμένη λαμπρότητα 33.5, 41.5, και 59.3 fb^{-1} , αντίστοιχα. Οι πίδακες σωματιδίων ανακατασκευάζονται μέσω του αλγόριθμου ομαδοποίησης anti- k_T , χρησιμοποιώντας παράμετρο απόστασης $R = 0.8$. Η μέτρηση των ενεργών διατομών διεξάγεται ως συνάρτηση της αναλλοίωτης μάζας $m_{1,2}$ και της μέγιστης απόλυτης ωκύτητας $|y|_{max}$ του ζεύγους των πιδάκων με τη μεγαλύτερη εγκάρσια ορμή p_T . Οι κατανομές αναδιπλώνονται για τη διόρθωσή τους από φαινόμενα του ανιχνευτή και συγκρίνονται με θεωρητικές προβλέψεις σε τρίτη τάξη στη θεωρία διαταραχών της κβαντικής χρωμοδυναμικής, οι οποίες έχουν διορθωθεί για μη διαταραχτικά φαινόμενα και επιπλέον ακτινοβολία λόγω ηλεκτρασθενών αλληλεπιδράσεων. Με χρήση των δεδομένων του έτους 2016, γίνεται μία καθολική προσαρμογή (global fit) για τον προσδιορισμό των συναρτήσεων κατανομής παρτονίων του πρωτονίου (PDFs) και τον υπολογισμό της σταθεράς ζεύξης των ισχυρών αλληλεπιδράσεων στην κλίμακα της μάζας του μποζονίου Z , αποδίδοντας το αποτέλεσμα $\alpha_S(m_Z) = 0.1179 \pm 0.0019$. Οι μετρούμενες κατανομές δείχνουν καλή συμφωνία με τις θεωρητικές προβλέψεις, οδηγώντας σε περαιτέρω περιορισμούς στα PDFs, ενώ η εξαγόμενη τιμή της $\alpha_S(m_Z)$ ταυτίζεται με την παγκόσμια μέση τιμή που δημοσιεύθηκε από την ομάδα σωματιδιακών δεδομένων (particle data group) το έτος 2022.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Panagiotis Kokkas, for giving me the opportunity to work alongside him all these years. He has been an exceptional professor, always proving his deep expertise, standing by his students, and offering unwavering guidance. His dedication to nurturing and supporting us to the best of his abilities has been invaluable.

I was incredibly fortunate to have Costantinos Foudas and Ioannis Papadopoulos as members of my three-member advisory committee. Their absolute dominance and expertise in their respective fields have been a constant source of inspiration. I sincerely thank them for their support whenever I needed it.

A heartfelt thank you goes to Paraskevas Gianneios, an outstanding senior colleague who ceaselessly and generously shared his vast knowledge with me, always offering insightful and unbiased advice in work and in life. I am equally grateful to Kosmas Adamidis, a passionate and free-spirited colleague. Though we never had the chance to collaborate directly, his dedication to his work and his perspective on both physics and life have been truly inspiring. I am fortunate to have shared a workspace with both of them, not just as colleagues but as close friends and role models who have helped shape the person I am today. I look up to them.

I also want to thank Argyro, the newest addition to our research group, with whom I closely collaborated, exchanging ideas and thoughts throughout this journey. My gratitude extends to Ioannis Bestintzanos, Spyros Lontos, and Eustathios Bletsas, all members of our research group at the University of Ioannina, whose companionship and contributions, each in their own way, have been greatly appreciated.

Among my international colleagues, I extend my regards to Patrick L.S. Connor, to whom I am indebted, who exemplified true professionalism, shared his knowledge in both physics and software, and warmly welcomed me to Hamburg, training me as a co-supervisor from the shadows. I am also grateful to Klaus Rabbertz and Daniel Savoiu for their expertise and the many insightful discussions we had.

Lastly, I want to thank all those who are close to me, family and friends alike, who have been an essential part of my daily life. There are too many to name individually, but their support has been invaluable, helping me grow, navigate challenges, and move forward. A special thank you to Vasiliki Pouftsi, whose presence in my life has been transformative. Her love, understanding, and passion for life, inspire me every day, pushing me to shape a better version of myself.

PREFACE

Influenced by my mother, my fascination with the cosmos began in my early years, during primary school. We often engaged in deep conversations, pondering the big questions: Is there a purpose to our existence, or is it merely a coincidence? Could life exist beyond our planetary system? Though we never found definitive answers, nor have I to this day, these discussions sparked my curiosity about the world around me, how it works, and why things appear as they do. Around that time, my mother began telling friends and family, with unwavering conviction, that I would become a scientist, an astrophysicist, she declared.

Strangely enough, this was enough to steer me toward the field of physics. Upon entering university, I was introduced to a vast array of disciplines, many of which I had never known existed. In the end, while I did not become an astrophysicist, studying stars, galaxies and dark matter, I found a different path to pursue my initial curiosity, through the field of High Energy Physics. This field is particularly demanding, requiring expertise among multiple disciplines, mathematics in the form of statistics and statistical analysis, programming and software development, and, of course, the core principles of physics itself. Hopefully, the latter will be sufficiently introduced in the chapters to come.

Over the past few years, as I embarked on my PhD journey, I had the opportunity to gain a vast amount of knowledge and develop new skills in a highly demanding environment. It was an experience like no other, one I would, without hesitation, relive if time travel was possible. I learned that the academic path can be rigorous and often requires sacrifices, and perhaps I was not meant for it. Now, with newly acquired skills, and a transformed mindset, I am eager to explore new fields, and uncover new knowledge that piques my curiosity.

For now, I am content with having lived the life of a scientist, and I am ready to press pause on it. I leave this dissertation, as a record of the knowledge I gained and my findings during my PhD years, hoping it serves future scientists who venture into this field. I also hope it is found well-written and conveys the excitement I felt for High Energy Physics, a truly beautiful discipline.

Until next time!

ΕΚΤΕΤΑΜΕΝΗ ΣΥΝΟΨΗ

Κεφάλαιο 1 Μεγάλος Αδρονικός Επιταχυντής

Το έτος 1949 προτάθηκε για πρώτη φορά η δημιουργία ενός Ευρωπαϊκού οργανισμού με κεντρικό αντικείμενο την έρευνα πάνω στον κλάδο της Πυρηνικής φυσικής και με σκοπό να συγκρατήσει την ευρωπαϊκή ερευνητική κοινότητα, σφηγτά δεμένη. Λίγα χρόνια αργότερα, στον εν λόγω οργανισμό, αποδόθηκε το όνομα Ευρωπαϊκό Κέντρο Πυρηνικών Ερευνών (Conseil Européen pour la Recherche Nucléaire, CERN). Το CERN, χωροταξικά τοποθετείται στα σύνορα μεταξύ Γαλλίας και Ελβετίας, με το γεωγραφικό του αποτύπωμα να μεγαλώνει ανά τα χρόνια με τη δημιουργία νέων επιταχυντικών διατάξεων.

Η πιο πρόσφατη προσθήκη είναι ο Μεγάλος Αδρονικός Επιταχυντής (Large Hadron Collider, LHC), ο οποίος αποτελεί, μέχρι και σήμερα, τον μεγαλύτερο επιταχυντή του είδους του σε παγκόσμια κλίμακα. Ο LHC στεγάζεται σε ένα τούνελ συνολικής περιφέρειας 27 Km , περίπου 100 m κάτω από το έδαφος. Εκεί, επιταχύνονται σωματίδια από πρωτόνια έως και πιο βαρέα ιόντα για τη διεξαγωγή πειραμάτων.

Αναφορικά με τα πρωτόνια, η επιτάχυνση επιτυγχάνεται μέσα από διαδοχικούς σταθμούς (διαφορετικοί επιταχυντές), οι οποίοι σειριακά ανεβάζουν την ενέργεια των σωματιδίων στο ανώτατο επιτρεπτό όριο των 6.5 TeV (6.8 TeV με το πέρας του έτους 2021). Συγκεκριμένα, παραθέτονται όλες οι επιταχυντικές διατάξεις που συνεισφέρουν στην αλυσιδωτή αύξηση της ενέργειας: Linac4, Proton Synchrotron Booster (PSB), Proton Synchrotron (PS), Super Proton Synchrotron (SPS), LHC.

Τα πρωτόνια επιταχύνονται σε ομάδες, οι οποίες στο εξής θα αναφέρονται ως *συστάδες*. Κάθε συστάδα περιέχει περίπου 1.8×10^{11} πρωτόνια, ενώ κάθε δεδομένη χρονική στιγμή ο LHC μπορεί να επιταχύνει 2808 από αυτές. Πρόκειται, δηλαδή για μία συνεχόμενη “ροή” συστάδων πρωτονίων, μία *δέσμη πρωτονίων*. Δύο τέτοιες δέσμες υποστηρίζονται στον LHC, κάθε μία από τις οποίες επιταχύνεται σε ενέργεια 6.5 TeV , επιτρέποντας συγκρούσεις μεταξύ των δύο δεσμών σε συνολική ενέργεια κέντρου μάζας 13 TeV (13.6 TeV με το πέρας του έτους 2021). Ο ρυθμός των συγκρούσεων ανέρχεται στα 40 MHz , τιμή που αντιστοιχεί σε διασταύρωση δύο συστάδων ανά 25 nsec .

Η αρχιτεκτονική του LHC δεν ακολουθεί, απόλυτα, κυκλική συμμετρία. Αντίθετα, αποτελείται από οκτώ τόξα και οκτώ ευθύγραμμα τμήματα. Η επιτάχυνση πραγματοποιείται μέσα από οκτώ, ανά δέσμη, κοιλότητες ραδιοσυχνότητας (Radiofrequency, RF), τοποθετημένες στα ευθύγραμμα τμήματά του, οι οποίες ταλαντώνονται σε συχνότητα 400 MHz , επιστρέφοντας 2 MV ανά πέρασμα, και διατηρώντας τη λειτουργία τους στα 4.5 K . Η στρέψη και εστίαση των δεσμών επιτυγχάνεται με τη χρήση μίας μεγάλης ποικιλομορφίας μαγνητών, κυρίως δίπολων και τετράπολων, αλλά και άλλων πολύπολων.

Χρονικά διαστήματα κατά τα οποία διεξάγονται συγκρούσεις και συλλέγονται δεδομένα

χωρίζονται σε ονομαστικές περιόδους. Ενώ σήμερα, το έτος 2025, ο LHC διανύει την τρίτη περίοδο λειτουργίας του (Run III), τα δεδομένα αυτής της ανάλυσης προέρχονται από τη δεύτερη περίοδο συλλογής δεδομένων, η οποία διήρκεσε από το έτος 2015 έως και το 2018. Στον Μεγάλο Αδρονικό Επιταχυντή στεγάζονται αρκετά πειράματα, τα μεγαλύτερα εκ των οποίων είναι τέσσερα, ονομαστικά: το ATLAS (A Toroidal LHC Apparatus), το CMS (Compact Muon Solenoid), το ALICE (A Large Ion Collider Experiment), και το LHCb (Large Hadron Collider beauty). Κάθε πείραμα εμπεριέχει τη δική του αντζέντα ερευνητικών καθηκόντων και συλλέγει δεδομένα από τις συγκρούσεις που διεξάγονται ανάλογα με τις μελέτες που πραγματοποιούνται. Εμείς θα επικεντρωθούμε στον ανιχνευτή CMS.

Κεφάλαιο 2

Συμπαγές Σωληνοειδές Μιονίων

Το CMS (Compact Muon Solenoid) είναι τοποθετημένο σε ένα από τα οκτώ ευθύγραμμα τμήματα του LHC, συγκεκριμένα στο Σημείο 5 (Point 5, P5). Το κέντρο του ταυτίζεται με το σημείο αλληλεπίδρασης (interaction point, IP), σημείο όπου οι δέσμες ευθυγραμίζονται, εστιάζονται, και προετοιμάζονται για σύγκρουση. Στα πλαίσια του φαινομένου αυτού, με τον όρο σύγκρουση νοείται το πέρασμα δύο συστάδων πρωτονίων, μίας από κάθε δέσμη, τη μία μέσα από την άλλη, διεργασία κατά την οποία δύο ή και περισσότερα πρωτόνια σχεδιάζονται.

Η κατασκευή του CMS είχε ως πρωταρχικό στόχο την παρατήρηση του σωματιδίου Higgs. Έτσι, τα χαρακτηριστικά του ανιχνευτή είχαν επιλεγεί κατάλληλα για την ολοκλήρωση αυτής της αποστολής: καλή ταυτοποίηση μιονίων και καλή διακριτική ικανότητα στην ορμή τους και στη μάζα ζεύγους μιονίων, καλή διακριτική ικανότητα στην ορμή φορτισμένων σωματιδίων και στην ανακατασκευή τους, επίσης, καλή διακριτική ικανότητα στην ενέργεια και στη μάζα από ζεύγος φωτονίων και ηλεκτρονίων, και τέλος, καλή διακριτική ικανότητα στην υπολειπόμενη εγκάρσια ενέργεια και τη μάζα ζεύγους πιδάκων σωματιδίων.

Το σύστημα συντεταγμένων του CMS ακολουθεί κυλινδρική συμμετρία, με το κέντρο του να ταυτίζεται με αυτό του CMS, τον x -άξονα να δείχνει ακτινικά προς το κέντρο του LHC, τον y -άξονα να εκτείνεται κάθετα προς τα πάνω, και τον z -άξονα να δείχνει κατά την αριστερόστροφη κατεύθυνση της δέσμης, προς τα βουνά Jura. Χρησιμοποιώντας σφαιρικές συντεταγμένες, η ορμή ενός σωματιδίου στο χώρο xyz μπορεί να εκφραστεί ως $\vec{r} = (|\vec{r}|, \theta, \phi)$, όπου $|\vec{r}|$ είναι το μέτρο της ορμής του σωματιδίου, θ η πολική γωνία, και ϕ η αζιμουθιακή γωνία, η τελευταία, μετρούμενη από τον x -άξονα προς τον y -άξονα.

Στον υπολογισμό μεγεθών πολλές φορές είναι πιο χρήσιμο να έχουμε ποσότητες που παραμένουν αναλλοίωτες. Έτσι, αντί της πολικής γωνίας προτιμάται η ωκύτητα y ή η ψευδο-ωκύτητα η (βλέπε Εξ. 2.1-2.3), οι διαφορές μεταξύ των οποίων παραμένουν όντως αναλλοίωτες. Παράλληλα, για τον υπολογισμό της γωνιακής απόστασης μεταξύ δύο σωματιδίων μπορεί να υπολογιστεί η ποσότητα $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, ποσότητα επίσης αναλλοίωτη.

Το CMS αποτελείται από ένα σύνολο υποανιχνευτικών συστημάτων, κάθε ένα από τα οποία είναι υπεύθυνο για την ανίχνευση διαφορετικού τύπου σωματιδίων. Ξεκινώντας από το κέντρο του ανιχνευτή, αρχικά, υπάρχει ο ανιχνευτής τροχιών για την ανίχνευση της τροχιάς και κατά επέκταση της ορμής φορτισμένων σωματιδίων. Ο εν λόγω ανιχνευτής χωρίζεται σε δύο τύπους, τον ανιχνευτή ψηφιδών και τον ανιχνευτή μικρολωρίδων. Ομόκεντρα προς τα έξω, ακολουθεί το ηλεκτρομαγνητικό καλορίμετρο, εξάρτημα αφοσιωμένο στη μέτρηση της ενέργειας ηλεκτρονίων και φωτονίων. Στη συνέχεια, υπάρχει το αδρονικό καλορίμετρο, για τη μέτρηση της ενέργειας φορτισμένων και ουδέτερων αδρονίων. Τα προαναφερθέντα ανιχνευτικά συστήματα εμπεριέχονται εντός ενός πελώριου σωληνοειδούς υπεραγωγίμου μαγνήτη, υπεύθυνο για την καμπύλωση των τροχιών φορτισμένων σωματιδίων, διαδικασία απαραίτητη

για τη μέτρηση της ορμής τους. Τέλος, στο εξωτερικό τμήμα του CMS υπάρχουν οι ανιχνευτές μιονίων τριών ειδών: οι θάλαμοι ολίσθησης, οι καθοδικοί θάλαμοι λωρίδων, και οι θάλαμοι αντίστασης παράλληλων πλακών.

Ολόκληρος ο ανιχνευτής συνδέεται με ένα σύστημα σκανδαλισμού, στόχος του οποίου είναι η ρύθμιση του ρυθμού των γεγονότων που περνούν προς συλλογή και καταγραφή. Με τη σειρά του, χωρίζεται σε δύο επίπεδα: το πρώτο επίπεδο σκανδαλισμού (Level-1 Trigger, L1T), το οποίο βασίζεται κυρίως σε ηλεκτρονικές συσκευές (hardware), και το υψηλό επίπεδο σκανδαλισμού (High Level Trigger, HLT), το οποίο αποτελείται από μία τεράστια φάρμα επεξεργαστών που τρέχουν συγκεκριμένο λογισμικό (software). Και τα δύο συστήματα συνεργάζονται, και εφαρμόζουν κατάλληλα κριτήρια για την απόρριψη “κακών” δεδομένων ή δεδομένων εκτός της περιοχής ενδιαφέροντος και την αποδοχή μόνον εκείνων που έχουν μετρηθεί υπό σωστές συνθήκες. Από έναν αρχικό ρυθμό καταγραφής γεγονότων 40 MHz , μετά το L1T, μειώνεται σε $\sim 100\text{ kHz}$, ενώ μετά τον HLT, υφίσταται μία περαιτέρω μείωση στα $\sim 1\text{ kHz}$. Δεδομένα που επιβιώνουν από τη διαδικασία ξεσκαρταρίσματος προωθούνται προς συλλογή και αποθήκευση.

Κεφάλαιο 3 Καθιερωμένο Πρότυπο

Το *Καθιερωμένο Πρότυπο (ΚΠ)* αποτελεί μία ενοποιημένη, μοντέρνα, θεωρία που περιγράφει επιτυχώς τρεις από τις τέσσερις θεμελιώδεις αλληλεπιδράσεις που διακρίνονται στον κόσμο μας, σήμερα. Αυτές είναι: η *ηλεκτρομαγνητική*, η *ασθενής*, και η *ισχυρή*. Μελετώντας τις μακροσκοπικές ιδιότητες της ύλης προκύπτει άλλη μία δύναμη, η *βαρυτική*, η οποία προς το παρόν δεν έχει ενσωματωθεί επιτυχώς σε κάποια, πειραματικά αποδεδειγμένη θεωρία, μαζί με τις υπόλοιπες αλληλεπιδράσεις.

Το ΚΠ προβλέπει τον τρόπο αλληλεπίδρασης μεταξύ των γνωστών, προς στιγμήν, στοιχειωδών σωματιδίων, ονομαστικά: τα κουάρκ (*up, down, strange, charm, top, bottom*), τα λεπτόνια (*electron, muon, tau* και τα νετρίνα αυτών), τους φορείς των τριών αλληλεπιδράσεων (*gluon, photon, W και Z μποζόνια*), και το σωματίο Higgs. Λαμβάνοντας υπόψιν όλες τις εκδοχές που μπορεί να εμφανιστούν τα σωματίδια αυτά, τελικώς, το άθροισμα τους ανέρχεται στον αριθμό 61.

Κάθε μία από τις τρεις βασικές αλληλεπιδράσεις εντός του ΚΠ φαίνεται να διέπεται από τη δική της θεωρία, και κρύβει έναν υποκείμενο νόμο διατήρησης. Το τελευταίο πόρισμα, προκύπτει ως απόρροια του θεωρήματος της Emmy Noether, όπου υποστηρίζει πως κάθε συμμετρία σε ένα φυσικό σύστημα αντιστοιχεί στη διατήρηση κάποιας φυσικής ποσότητας. Κοιτώντας το πρόβλημα από τη μαθηματική σκοπιά, φαίνεται να αποδίδεται η συνέπεια των παραπάνω “υποθέσεων”. Όντως, το ΚΠ είναι μία *κβαντική θεωρία βαθμίδος* (Quantum Field Theory, QFT) όπου τα βασικά της αντικείμενα είναι κβαντικά πεδία. Τα πεδία χωρίζονται σε δύο κατηγορίες, τα *φερμιόνια* που αποτελούν τα πεδία ύλης και τα *μποζόνια* που αντιστοιχούν στα πεδία για τους φορείς των αλληλεπιδράσεων.

Εκμεταλλευόμενοι τα πεδία και τις ιδιότητες που εισάγει ο Λαγκρανζιανός φορμαλισμός, προκύπτει ότι: η ηλεκτρομαγνητική αλληλεπίδραση διέπεται από την *κβαντική ηλεκτροδυναμική* (Quantum Electrodynamics, QED) βασιζόμενη στη συμμετρία βαθμίδος $U(1)$, η ισχυρή διέπεται από την *κβαντική χρωμοδυναμική* (Quantum Chromodynamics, QCD) μέσω της συμμετρίας $SU(3)$, και η ασθενής μέσα από τη “δυναμική των γεύσεων” (Flavordynamics) που εκφράζεται από τη συμμετρία $SU(2)$. Στην πραγματικότητα, η ασθενής αλληλεπίδραση προκύπτει πλήρως, μόνο, μετά την ενοποίηση αυτής με την ηλεκτρομαγνητική, δημιουργώντας έτσι τη θεωρία των ηλεκτρασθενών αλληλεπιδράσεων που εκφράζεται μέσα από τη συμμετρία $SU(2) \times U(1)$ · ενώ η καθαρά ηλεκτρομαγνητική είναι το αποτέλεσμα του αυθόρμητου

σπάσιμο της προηγούμενης συμμετρίας $SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}$.

Σημαντικό κομμάτι στο σύνολο της θεωρίας του ΚΠ είναι το αυθόρμητο σπάσιμο της συμμετρίας, διεργασία γνωστή και ως *μηχανισμός Higgs*. Μέσω αυτού του μηχανισμού γεννιέται το σωματίδιο Higgs αποδίδοντας μάζα στα αρχικά άμαζα μποζόνια βαθμίδος W και Z της ασθενούς αλληλεπίδρασης. Επιπροσθέτως, μέσω της ίδιας ακριβώς διεργασίας τα φερμιόνια αποκτούν μάζα, τα οποία ενώ ενσωματώνονται αρχικά στη θεωρία ως άμαζα, τελικώς, αποκτούν μάζα με την αλληλεπίδρασή τους με το πεδίο Higgs. Καταληκτικά, η ενοποίηση των τριών αλληλεπιδράσεων στα πλαίσια του ΚΠ δίνεται μέσω της συμμετρίας $SU(3)_C \times SU(2)_L \times U(1)_Y$.

Το ΚΠ όμως, δεν είναι μία ολοκληρωμένη θεωρία, αφού αφήνει αρκετά ανοικτά ζητήματα. Για παράδειγμα, δεν περιλαμβάνει τη βαρυτική αλληλεπίδραση, δεν ενσωματώνει κάποιο υποψήφιο σωματίδιο για την περιγραφή της σκοτεινής ύλης και ενέργειας, τα νετρίνα θεωρούνται άμαζα σε αντίθεση με μετρήσεις που τους αποδίδουν μικρή, αλλά μη μηδενική τιμή στη μάζα, ενώ, ακόμη, δεν εξηγεί την ασυμμετρία που παρατηρείται μεταξύ της ύλης και της αντιύλης στο σύμπαν. Αυτά τα στοιχεία μας οδηγούν στο συμπέρασμα ότι η φύση κρύβει μία πιο γενικευμένη θεωρία για την ενοποίηση όλων των γνωστών δυνάμεων και την επεξήγηση των “ανοικτών” φαινομένων. Με τη συνεχόμενη εξέλιξη της επιστήμης η κοινότητα στοχεύει στην εύρεση της πολυπόθητης, πιο γενικευμένης, θεωρίας που συχνά αναφέρεται ως *Θεωρία των Πάντων* (Theory of Everything).

Κεφάλαιο 4

Αντικείμενα Ανάλυσης - Πίδακες Σωματιδίων

Οι πίδακες σωματιδίων είναι ίσως τα πιο πολύπλοκα αντικείμενα που συναντιούνται σε υψηλοενεργειακές συγκρούσεις πρωτονίων-πρωτονίων, όπως αυτές που διεξάγονται στα πλαίσια του LHC και καταγράφονται από το CMS. Για την κατανόησή τους είναι αρκετά χρήσιμο να δοθεί βάση στις υποκείμενες φυσικές διεργασίες που διέπουν τη δημιουργία και εξέλιξή τους.

Η ύλη γύρω μας αποτελείται από τα διάφορα στοιχεία του περιοδικού πίνακα, τα άτομα. Πρωταγωνιστικό ρόλο στον πυρήνα των ατόμων παίζουν τα πρωτόνια και τα νετρόνια που τον αποτελούν. Αυτά, με τη σειρά τους, αποτελούνται από κουάρκς και γλουόνια, στοιχειώδη σωματίδια που θα σχολιαστούν από τη σκοπιά της κβαντικής χρωμοδυναμικής. Τα πρωτόνια και τα νετρόνια ανήκουν στην ευρύτερη κατηγορία σωματιδίων που ονομάζονται *αδρόνια* (hadrons). Τα τελευταία, ορίζονται ως όλα εκείνα τα υποατομικά σωματίδια που αποτελούνται από κουάρκς. Δύο κατηγορίες αδρονίων παρουσιάζονται, τα *μεσόνια* που αποτελούνται από ζεύγη κουάρκ-αντικουάρκ $q_i\bar{q}_i$, και τα *βαρυόνια* που σχηματίζονται, πιο συχνά, από τρία κουάρκ $q_i q_j q_k$ με τα νετρόνια και πρωτόνια να ανήκουν στην τελευταία κατηγορία.

Δύο φυσικές διεργασίες που απορρέουν από την κβαντική χρωμοδυναμική, παίζουν τον ρόλο του ρυθμιστή για τη δημιουργία των αδρονίων. Από τη μία μεριά, υπάρχει το φαινόμενο του εγκλωβισμού των κουάρκς, το οποίο τα αναγκάζει να δημιουργούν δέσμιες καταστάσεις, τα αδρόνια, αποκλείοντας έτσι το ενδεχόμενο να παρατηρηθούν ως ελεύθερα σωματίδια. Από την άλλη μεριά, επειδή ακριβώς τα κουάρκς βρίσκονται σε ένα βασίλειο πολύ μικρής κλίμακας και πολύ μεγάλης ενέργειας, το φαινόμενο της ασυμπτωτικής ελευθερίας αναίρει την προηγούμενη κατάσταση. Τους επιτρέπεται δηλαδή να ταξιδεύουν στιγμιαία ελεύθερα, στα πλαίσια της διαταρακτικής κβαντικής χρωμοδυναμικής (perturbative QCD, pQCD), για πολύ μικρές αποστάσεις και μεγάλες ενέργειες, ενώ όταν αυτές οι συνθήκες υπερβαίνονται, αυτά αναγκάζονται να σχηματίσουν δέσμιες καταστάσεις, σύμφωνα με το φαινόμενο του εγκλωβισμού, διαδικασία μη διαταρακτική.

Σημαντικό ρόλο σε αυτή τη διττότητα της QCD, μεταξύ της διαταρακτικής και μη διαταρακτικής κατάστασης, παίζει η σταθερά των ισχυρών αλληλεπιδράσεων α_S . Σε υψηλές ενέργειες, η τιμή της μειώνεται σημαντικά, όπου για $\alpha_S \ll 1$ μπαίνουμε στο διαταρακτικό κομμάτι της φυσικής, ενώ για χαμηλότερες ενέργειες, η τιμή της μεγαλώνει, και η φυσική είναι μη διαταρακτική, με $\alpha_S \gg 1$.

Η διαταρακτική κβαντική χρωμοδυναμική επιτρέπει τον ορισμό της διαφορικής ενεργούς διατομής για διεργασίες $2 \rightarrow N$ (βλέπε Εξ. 4.8), όπως αυτές που συναντάμε στο LHC, όπου δύο πρωτόνια συγκρούονται στην αρχική κατάσταση και παράγονται N σωματίδια στην τελική. Ο ορισμός αυτός επιμερίζει τον υπολογισμό σε δύο μέρη: το διαταρακτικό σκέλος, που περιγράφεται από τη pQCD μέσω του στοιχείου πίνακα (matrix element, ME), και το μη διαταρακτικό, το οποίο εξαρτάται από την επιλογή ενός PDF set.

Η δημιουργία των πιδάκων σωματιδίων διακρίνεται, ως εξής. Η σύγκρουση δύο πρωτονίων επιφέρει τη δημιουργία, τουλάχιστον, δύο υψηλοενεργειακών *παρτονίων* (κουάρκ ή γλουονίων), τα οποία στιγμιαία κινούνται ελεύθερα. Στη συνέχεια, τα περίσσια αποθέματα ενέργειας των παρτονίων, εναποτίθενται στη δημιουργία περαιτέρω κουάρκ και γλουονίων, με ταυτόχρονη μείωση της αρχικής τους ενέργειας, διαδικασία που ονομάζεται *παρτονικός καταγισμός* (parton shower, PS). Ο καταγισμός θα συνεχιστεί μέχρι το σημείο που η ενέργεια των αρχικών παρτονίων μειωθεί αρκετά, ώστε $\alpha_S \gg 1$ και να παύσει η ισχύς της διαταρακτικής κατάστασης. Τότε, θα αρχίσει ο σχηματισμός δέσμιων καταστάσεων (αδρονίων) μεταξύ των παραγόμενων σωματιδίων, μέσω της διαδικασίας *αδρονοποίησης* (hadronization). Τελικώς, τα σωματίδια προσπίπτουν στον ανιχνευτή όπου ακολουθεί η καταγραφή τους. Κοιτώντας ολόκληρο το φαινόμενο από το σημείο της σύγκρουσης προς τον ανιχνευτή, τα παραγόμενα σωματίδια κινούνται “συγγραμικά” προς την κατεύθυνση των αρχικών παρτονίων, διατεταγμένα εντός μίας κωνικής γεωμετρίας, αντικείμενο το οποίο συνολικά ορίζεται ως *πίδακας σωματιδίων*.

Στην πράξη, πειραματικά, σωματίδια που ανιχνεύονται από το CMS πρέπει πρώτα να ταυτοποιηθούν μέσα από τον *αλγόριθμο ροής σωματιδίων* (particle flow algorithm, PF), ο οποίος λαμβάνει ως είσοδο την ακατέργαστη πληροφορία όπως αυτή παρέχεται από όλα τα υποανιχνευτικά συστήματα της. Στη συνέχεια, η ταυτοποιημένη λίστα των ανιχνευόμενων σωματιδίων προωθείται στον *αλγόριθμο ομαδοποίησης πιδάκων anti- k_T* (anti- k_T jet clustering algorithm), υπεύθυνος για την ανακατασκευή αυτών των σύνθετων αντικειμένων, των πιδάκων σωματιδίων. Τα πειραματικά δεδομένα τελικώς σχηματίζονται με την ομαδοποίηση όλης της πληροφορίας που σχετίζεται με κάθε σύγκρουση, μαζί με τη λίστα των πιδάκων σωματιδίων που προέκυψε. Τα δεδομένα αυτά προωθούνται στην επιστημονική κοινότητα για την επεξεργασία τους και τη διεξαγωγή αναλύσεων.

Μετρήσεις γίνονται τόσο από τα πειραματικά δεδομένα, όσο και από προσομοιωμένα δεδομένα, τα οποία προκύπτουν με χρήση, γεννήτορες γεγονότων Monte Carlo (MC). Τα τελευταία, αποτελούν ένα πολύ χρήσιμο εργαλείο αφού χρησιμοποιούνται για τη συστηματική σύγκριση με το πείραμα, την παραγωγή διορθώσεων και αβεβαιοτήτων για τα πειραματικά δεδομένα και άλλα. Τα δεδομένα MC παράγονται σε δύο στάδια, αρχικά με τη γέννηση των δεδομένων που αντιπροσωπεύουν τη σύγκρουση των πρωτονίων και μεταγενέστερα με την προσομοίωση της απόκρισης και διακριτικής ικανότητας του ανιχνευτή στις μετρούμενες ποσότητες.

Κεφάλαιο 5

Επεξεργασία Δεδομένων

Για να υπάρχει κάποιο ουσιαστικό συμπέρασμα από τα αποτελέσματα μίας ανάλυσης, είναι σημαντικό η μετρούμενη ποσότητα να είναι υπολογίσιμη, τόσο πειραματικά όσο και θεωρητικά, ώστε τελικώς να γίνει μία σύγκριση. Οι πίδακες σωματιδίων επιτρέπουν αυτόν ακριβώς τον

υπολογισμό, και από τα δύο μέτωπα, τον υπολογισμό των ενεργών διατομών. Η επιλογή των πιδάκων ως κεντρικά αντικείμενα ενδιαφέροντος, υπερ της πολύπλοξης φύση τους, γίνεται διότι επιτρέπουν την μελέτη της Κβαντικής Χρωμοδυναμικής. Στα πλαίσια αυτής της διδακτορικής διατριβής, παρουσιάζεται η μέτρηση της διπλά διαφορικής ενεργούς διατομής παραγωγής τουλάχιστον δύο πιδάκων σωματιδίων ως προς την αναλλοίωτη μάζα $m_{1,2}$ και τη μέγιστη απόλυτη ωκύτητα $|y|_{max}$ του ζέγους των πιδάκων σωματιδίων με τη μεγαλύτερη εγκάρσια ορμή (σε κάθε γεγονός), κυρίαρχοι πίδακες ή *leading jets*.

Η μέτρηση βασίζεται σε πειραματικά δεδομένα που αντιστοιχούν σε τρεις διαφορετικές χρονικές περιόδους διεξαγωγής συγκρούσεων στο LHC, τα έτη 2016, 2017, και 2018· με την κεντρική μέτρηση να αποδίδεται από τα δεδομένα του έτους 2016, για τα οποία, επιπροσθέτως, γίνεται η εκτίμηση της σταθεράς των ισχυρών αλληλεπιδράσεων α_S και των συναρτήσεων κατανομής παρτονίων PDFs για το πρωτόνιο. Τα δεδομένα προέρχονται από συγκρούσεις πρωτονίων-πρωτονίων σε ενέργεια κέντρου μάζας $\sqrt{s} = 13 \text{ TeV}$ και έχουν συλλεχθεί από τον ανιχνευτή CMS. Αντιστοιχούν σε ολοκληρωμένη λαμπρότητα $\mathcal{L}_{int} = 33.5, 41.5,$ και 59.3 fb^{-1} , των ετών 2016, 2017, και 2018, αντιστοίχως. Τέλος, η ανακατασκευή των πιδάκων έχει επιτευχθεί μέσω του αλγορίθμου *anti- k_T* για την παράμετρο απόστασης $R = 0.8$.

Υπό το πρίσμα της προσομοίωσης, τα κεντρικά δεδομένα έχουν γεννηθεί από τον γεννήτορα PYTHIA8, με το CUETP8M1 tune για το 2016, ενώ το CP5 tune χρησιμοποιήθηκε για τα έτη 2017 και 2018. Συμπληρωματικά, και μόνον για το έτος 2016, χρησιμοποιήθηκαν και δεδομένα που έχουν παραχθεί με τον συνδυασμό του γεννήτορα MADGRAPH5 και του PYTHIA8 CUETP8M1.

Η ανάλυση πραγματοποιήθηκε με τη χρήση του λογισμικού *DAS Analysis System*, στα πλαίσια του οποίου αναπτύχθηκε κατάλληλος κώδικας για τη διεκπεραίωση των διεργασιών. Η επεξεργασία των δεδομένων ξεκινάει με την κατάλληλη ανακατασκευή και επιλογή γεγονότων στα οποία εφαρμόζονται κεντρικά παρεχόμενες συστάσεις από το πείραμα CMS για την ελαχιστοποίηση των κακώς ανακατασκευασμένων γεγονότων και την αποφυγή θορύβου στα τελικά δεδομένα.

Στα πειραματικά δεδομένα σημαντικό ρόλο παίζουν οι διορθώσεις για την κλίμακα της ενέργειας (Jet Energy Scale, JES) των πιδάκων σωματιδίων. Κατά αυτήν τη διαδικασία, εφαρμόζονται διορθώσεις που αφαιρούν την έξτρα ενέργεια που εναποτίθεται στα γεγονότα λόγω *pileup*, εξισορροπείται η μη ομοιόμορφη και μη γραμμική απόκριση του ανιχνευτή ως προς την ψευδο-ωκύτητα η και την εγκάρσια ορμή p_T αντίστοιχα, διορθώνονται για υπολειπόμενες διαφορές που παρατηρούνται μεταξύ των πειραματικών δεδομένων και αυτών από την προσομοίωση, ενώ, ακόμη, συνυπολογίζεται η διαφορά στην ευαισθησία του ανιχνευτή ως προς την ανίχνευση πιδάκων σωματιδίων που προέρχονται από διαφορετικού τύπου παρτόνια.

Μία άλλη διόρθωση αφορά το φαινόμενο που παρατηρήθηκε κατά τη συλλογή των δεδομένων τα έτη 2016 και 2017. Το υψηλά ραδιενεργό περιβάλλον του ανιχνευτή οδήγησε στην απορρύθμιση του σκανδαλιστή που σχετίζεται με το ηλεκτρομαγνητικό καλορίμετρο. Ως αποτέλεσμα, η απόδοση καταγραφής των δεδομένων μειώθηκε σε συγκεκριμένες περιοχές του φασικού χώρου, φαινόμενο που εξισορροπείται μέσω των διορθώσεων.

Τα προσομοιωμένα δεδομένα διορθώνονται εξίσου για την επαναφορά της κλίμακας ενέργειας των πιδάκων, ενώ επιπρόσθετες διορθώσεις ακολουθούν ώστε οι συνθήκες παραγωγής τους να είναι πιο κοντά σε αυτές των πειραματικών δεδομένων.

Τα δεδομένα από την προσομοίωση χαρακτηρίζονται από μία καλύτερη διακριτική ικανότητα στον ανιχνευτή (Jet Energy Resolution, JER) από ό,τι παρατηρείται στην πράξη. Έτσι, εφαρμόζονται διορθώσεις για τη ρύθμιση του διανύσματος της τετραορμής στους πίδακες στη διαδικασία που ονομάζεται, κατά τη βιβλιογραφία, *smearing*.

Οι *pileup* συγκρούσεις είναι αναπόφευκτες, κατά τη διεξαγωγή του πειράματος. Για τον λόγο αυτό, *pileup* συγκρούσεις προσομοιώνονται ξεχωριστά και προστίθενται στην προσομοίωση που αφορά την κύρια αλληλεπίδραση, με σκοπό να παραχθεί ένα σύνολο δεδομένων που θα αντιπροσωπεύει με μεγαλύτερη ακρίβεια αυτό που παρατηρείται στο πείραμα. Στην πράξη,

συγκρίνοντας το pileup προφίλ μεταξύ της προσομοίωσης και του πειράματος, αυτά διαφέρουν, με τις διορθώσεις αυτές, τελικά, να αντισταθμίζουν τις παρατηρούμενες διαφορές.

Το πρώτο βήμα μετά την ολοκλήρωση της επεξεργασίας και διόρθωσης των δεδομένων είναι ο υπολογισμός των καμπυλών απόδοσης των σκανδαλιστών. Συνολικά, υπάρχουν 10 διαθέσιμοι σκανδαλιστές, καθένας από τους οποίους έχει ένα διαφορετικό κατώφλι ως προς την εγκάρσια ορμή του κυρίαρχου leading πίδακα. Κατά αυτήν τη διαδικασία, υπολογίζεται το σημείο στο οποίο κάθε σκανδαλιστής αποκτά τη μέγιστη απόδοση καταγραφής γεγονότων, με σκοπό τη μετέπειτα εξαγωγή δεδομένων, αποκλειστικά και μόνον, από τις πιο αποδοτικές περιοχές. Στα δεδομένα που απομένουν, εφαρμόζονται μία πιο αυστηρή σειρά κινηματικών κριτηρίων, για τον καλύτερο ορισμό του χώρου φάσεων ενδιαφέροντος. Συγκεκριμένα, οι δύο leading πίδακες σε κάθε γεγονός υποχρεούνται να τηρούν: $p_{T,1} > 100 \text{ GeV}$, $|y_1| < 2.5$, και $p_{T,2} > 50 \text{ GeV}$, $|y_2| < 2.5$, όπου οι δείκτες 1 και 2 αντιπροσωπεύουν τον πίδακα με τη μεγαλύτερη και δεύτερη μεγαλύτερη εγκάρσια ορμή στο γεγονός, αντίστοιχα. Τα δεδομένα συνδυάζονται κατάλληλα, με κάθε σκανδαλιστή να συνεισφέρει σχεδόν σε ολόκληρο το εύρος της αναλλοίωτης μάζας, και με την κατασκευή της διαφορικής ενεργούς διατομής στο επίπεδο του ανιχνευτή (detector level) να γίνεται, σύμφωνα με την Εξ. 5.8.

$$\frac{d^2\sigma}{dy_{max} dm_{1,2}} = \frac{1}{\mathcal{L}_{int}} \frac{N_{eff}}{(2\Delta|y|_{max})\Delta m_{1,2}}$$

Για την αποτελεσματική σύγκριση μεταξύ των πειραματικών μετρήσεων και των θεωρητικών προβλέψεων, οι πρώτες πρέπει να διορθωθούν ώστε να μην εμπεριέχονται τα ενδογενή χαρακτηριστικά του ανιχνευτή στο μετρούμενο φάσμα. Αυτό επιτυγχάνεται μέσα από τη διαδικασία *ανάδιπλωσης* (unfolding), όπου τα MC δεδομένα χρησιμοποιούνται για την προσομοίωση της απόκρισης του ανιχνευτή στη μεταβολή που επιφέρει κατά τη διαδικασία μέτρησης. Με την κατασκευή του πίνακα απόκρισης (response matrix, RM) και την εφαρμογή του στο φάσμα σε επίπεδο ανιχνευτή, αναστρέφονται οι μεταναστεύσεις γεγονότων που έχουν προκληθεί λόγω της πεπερασμένης διακριτικής ικανότητας του ανιχνευτή, διορθώνοντας αποτελεσματικά τη μέτρηση και φέρνοντάς τη στο επίπεδο σωματιδίων (particle level ή truth level).

Ακολουθεί ο υπολογισμός της συνολικής πειραματικής αβεβαιότητας. Εκτός από το ενδογενές στατιστικό σφάλμα που σχετίζεται με την μέτρηση, το οποίο μεταφέρεται μέσα από τη διαδικασία ανάδιπλωσης, συστηματικές αβεβαιότητες υπεισέρχονται στη μέτρηση λόγω των διορθώσεων που εφαρμόζονται στα δεδομένα. Παρατίθενται οι πηγές συστηματικών αβεβαιότητων: αβεβαιότητες από τις διορθώσεις για την κλίμακα της ενέργειας των πιδάκων (JES) και της διακριτικής ικανότητας του ανιχνευτή (JER), αβεβαιότητα στη μέτρηση της ολοκληρωμένης λαμπρότητας για την κανονικοποίηση της διαφορικής ενεργούς διατομής, ενδογενές στατιστικό σφάλμα των προσωμοιωμένων δεδομένων, αβεβαιότητες λόγω της διόρθωσης για την αφαίρεση γεγονότων υποβάθρου (fakes), και χαμένων γεγονότων (misses), και αβεβαιότητα για τη μη αποδοτική συμπεριφορά του ανιχνευτή λόγω prefiring. Τελικά, η συνολική πειραματική αβεβαιότητα υπολογίζεται ως η τετραγωνική ρίζα του αθροίσματος των τετραγώνων της πειραματικής στατιστικής αβεβαιότητας και όλων των επιμέρους συστηματικών αβεβαιότητων.

Κεφάλαιο 6

Θεωρητικές Προβλέψεις Σταθερής Τάξης

Οι θεωρητικές προβλέψεις υπολογίζονται σύμφωνα με την Εξ. 4.8.

$$d\sigma_{(pp \rightarrow N)} = \sum_{i,j} \int dx dx' f_{i/p}(x, \mu_f) \cdot f_{j/p}(x', \mu_f) \times d\sigma_{(ij \rightarrow N)}(x, x', \mu_f, \mu_r, \alpha_S(\mu_r))$$

Το διαταρακτικό κομμάτι, δηλαδή το στοιχείο πίνακα, υπολογίζεται στα πλαίσια της κβαντικής χρωμοδυναμικής σε τρίτη τάξη στη θεωρία διαταραχών (next-to-next-to-leading order, NNLO). Συγκεκριμένα, η πράξη αυτή γίνεται με τη χρήση του προγράμματος NNLOJET το οποίο συνδέεται με το πακέτο FASTNNLO μέσω τις εφαρμογής APPLFAST για τη διαχείριση των πινάκων που περιλαμβάνουν τον διαταρακτικό υπολογισμό. Η χρήση αυτού του λογισμικού επιτρέπει τον συνδυασμό του στοιχείου πίνακα με διαφορετικά PDFs, το μη διαταρακτικό κομμάτι στην Εξ. 4.8, για τον τελικό προσδιορισμό των προβλέψεων στο επίπεδο των παρτονίων (parton level). Επιπλέον, η θεωρητική περιγραφή υιοθετεί την επιλογή της κλίμακας επανακανονικοποίησης μ_R , να είναι ίση με την κλίμακα παραγοντοποίησης μ_F , με τιμή αντίστοιχη της αναλλοίωτης μάζας $\mu_R = \mu_F = m_{1,2}$.

Οι θεωρητικές προβλέψεις διορθώνονται για να περιλαμβάνουν μη διαταρακτικά φαινόμενα, όπως η αδρονοποίηση και οι πολυπαρτονικές αλληλεπιδράσεις. Έτσι, η θεωρία έρχεται στο επίπεδο σωματιδίων, το ίδιο επίπεδο με αυτό των πειραματικών μετρήσεων. Οι κατάλληλες διορθώσεις υπολογίζονται μέσω γεννητόρων γεγονότων, οι οποίοι επιτρέπουν τον προσδιορισμό της μη διαταρακτικής συνεισφοράς. Αυτό επιτυγχάνεται συγκρίνοντας τις προβλέψεις που προκύπτουν από τους γεννήτορες, σε δύο στάδια: αρχικά, λαμβάνοντας υπόψη μόνο το διαταρακτικό κομμάτι και στη συνέχεια, ενσωματώνοντας τόσο τις διαταρακτικές όσο και τις μη διαταρακτικές συνιστώσες. Επιπροσθέτως, οι προβλέψεις διορθώνονται για επιπλέον ακτινοβολία λόγω ηλεκτρασθενούς ακτινοβολίας, οι οποίες γίνονται σημαντικές σε κλίμακες μεγαλύτερες του 1 TeV .

Τρεις πηγές θεωρητικών συστηματικών αβεβαιοτήτων, λαμβάνονται υπόψιν: αβεβαιότητα από την επιλογή της κλίμακας επανακανονικοποίησης και παραγοντοποίησης, η οποία αντανακλά τις εκλείπουσες διορθώσεις υψηλότερης ακρίβειας στον υπολογισμό του στοιχείου πίνακα, αβεβαιότητα από την επιλογή του PDF set, το οποίο συνδυάζεται με το στοιχείο πίνακα για την απόκτηση της τελικής θεωρητικής πρόβλεψης, και τέλος, αβεβαιότητα στον υπολογισμό των μη διαταρακτικών διορθώσεων, όπως αυτές περιγράφηκαν προηγουμένως. Ωστόσο, δεν συνεισφέρουν στη συνολική θεωρητική αβεβαιότητα οι διορθώσεις λόγω ηλεκτρασθενών αλληλεπιδράσεων. Η συνολική αβεβαιότητα προσδιορίζεται ως η τετραγωνική ρίζα του αθροίσματος του τετραγώνου των τριών επιμέρους συνεισφορών.

Οι πειραματικές μετρήσεις συγκρίνονται με τις θεωρητικές προβλέψεις, όπου οι τελευταίες υπολογίζονται χρησιμοποιώντας το CT18 PDF set. Παρατηρείται καλή συμφωνία μεταξύ των δύο, σε όλο το εύρος τιμών της αναλλοίωτης μάζας, και για κάθε μία από τις πέντε περιοχές ωχύτητας, με απόκλιση μικρότερη από 10 %.

Κεφάλαιο 7

Αποτελέσματα

Τα δεδομένα του έτους 2016 χρησιμοποιούνται για την εκτίμηση της σταθεράς ζεύξης των ισχυρών αλληλεπιδράσεων α_S και τον προσδιορισμό των συναρτήσεων κατανομής παρτονίων (PDFs) του πρωτονίου. Η στρατηγική που ακολουθείται είναι η ίδια με αυτήν που εφαρμόστηκε σε παλαιότερες HERAPDF αναλύσεις. Συγκεκριμένα, οι θεωρητικές προβλέψεις προσαρμόζονται σε DIS (Deep Inelastic Scattering) δεδομένα, τα οποία αργότερα συμπληρώνονται από τη μέτρηση της διπλά διαφορικής ενεργούς διατομής που παρουσιάστηκε εδώ. Σημειώνεται, ότι τα DIS δεδομένα περιορίζονται για τιμές στη μεταφορά ορμής πάνω από $Q_{min}^2 = 10 \text{ GeV}^2$.

Η δομή του πρωτονίου εκφράζεται σύμφωνα με την κατανομή του γλουονίου, των up και down κουάρκ σθένους, και τη συνολική κατανομή της θάλασσας των κουάρκ. Οι κατανομές αυτές παραμετροποιούνται ως συνάρτηση του κλάσματος ορμής του πρωτονίου, σε μία αρχική

κλίμακα $\mu_{F,0}^2 = 1.9 \text{ GeV}$, σύμφωνα με την Εξ. 7.1.

$$xf(x, \mu_{F,0}^2) = A_f x^{B_f} (1-x)^{C_f} (1 + D_f x + E_f x^2)$$

Η προσαρμογή ξεκινάει για ένα σύνολο από 10 αρχικές παραμέτρους, ενώ περισσότερες προστίθενται στην πορεία, σύμφωνα με τη διαδικασία ελαχιστοποίησης των τετραγώνων.

Τέσσερις αβεβαιότητες προκύπτουν στην εκτίμηση των PDFs: αβεβαιότητα στην προσαρμογή (fit uncertainty), που αντανακλά τη διάδοση των πειραματικών και θεωρητικών αβεβαιοτήτων της μέτρησης, αβεβαιότητα στην επιλογή μοντέλου (model uncertainty), που προκύπτει από τις επιλογές που έχουν γίνει για τις τιμές παραμέτρων που δεν σχετίζονται με τα PDFs, όπως το κατώφλι στην ορμή Q_{min}^2 , αβεβαιότητα για την τελική παραμετροποίηση (parametrization uncertainty) που χρησιμοποιείται στα PDFs, και τέλος, αβεβαιότητα στην κλίμακα (scale uncertainty) για τις εκλείπουσες διορθώσεις υψηλότερης τάξης στο στοιχείο πίνακα. Τελικώς, η συνολική αβεβαιότητα υπολογίζεται ως η τετραγωνική ρίζα του αθροίσματος των τετραγώνων των fit, model, και scale συνεισφορών, με μοναδική εξαίρεση την αβεβαιότητα parametrization που προστίθεται γραμμικά.

Για να επωφεληθούμε από την αυξημένη συμφωνία μεταξύ των πειραματικών δεδομένων και των θεωρητικών προβλέψεων, στον τελικό υπολογισμό των PDFs, εξαιρείται η χρήση της μέτρησης στην πέμπτη περιοχή ωκύτητας ($2.0 < |y|_{max} < 2.5$), στην οποία παρατηρείται η μεγαλύτερη απόκλιση. Ο Πίνακας 7.2 περιλαμβάνει τις μερικές τιμές χ^2 , με ή χωρίς την ενσωμάτωση των πειραματικών δεδομένων από το CMS.

Πίνακας 1: Partial χ^2 values for HERA DIS data and HERA DIS complemented with the 2016 CMS dijet measurements.

Data set	Partial χ^2/n_{data}	
	HERA DIS	HERA DIS + CMS dijets
CMS dijets		
$ y _{max} < 0.5$		18/22
$0.5 < y _{max} < 1.0$		15/22
$1.0 < y _{max} < 1.5$		16/23
$1.5 < y _{max} < 2.0$		15/12
HERA1+2		
$CC e^-p, E_p = 920 \text{ GeV}$	51/42	51/42
$CC e^+p, E_p = 920 \text{ GeV}$	37/39	37/39
$NC e^-p, E_p = 920 \text{ GeV}$	221/159	222/159
$NC e^+p, E_p = 460 \text{ GeV}$	198/177	197/177
$NC e^+p, E_p = 575 \text{ GeV}$	186/221	186/221
$NC e^+p, E_p = 820 \text{ GeV}$	55/61	55/61
$NC e^+p, E_p = 920 \text{ GeV}$	359/317	364/317
Total χ^2/n_{dof}	1161/1003	1232/1081

Συγκρίνοντας τα PDFs που προκύπτουν μόνο με τη χρήση των DIS δεδομένων, με αυτά, όταν τα DIS δεδομένα συμπληρώνονται από του CMS, συμπεραίνεται η συμφωνία τους. Επίσης, παρατηρείται μείωση στην αβεβαιότητα fit. Το φαινόμενο αυτό, γίνεται πιο εμφανές στην κατανομή του γλουονίου, με το PDF να περιορίζεται στην περιοχή για μεγάλα κλάσματα ορμής, $x > 0.1$. Επιπροσθέτως, η προσαρμογή επαναλαμβάνεται αφήνοντας την α_S ως

ελεύθερη παράμετρο στη διαδικασία, όπου η τιμή που επιστρέφεται στην κλίμακα της μάζας του μποζονίου Z , είναι $\alpha_S(m_Z) = 0.1179 \pm 0.0019$, το οποίο συμβαδίζει πλήρως με τον παγκόσμιο μέσο όρο 0.1179 ± 0.0009 .

“Το λιβάδι του Ποιητή”
Κοσμάς Αδαμίδης

CONTENTS

Acknowledgements	iii
Preface	iv
Εκτεταμένη Σύνοψη	v
List of Tables	xx
Λιστ οφ Φιγυρες	xxii
I Φυνδαμενταλς	1
1 Τηε Λαργε Χαδρον ολλιδερ	3
1.1 Ιντροδυςτιον	3
1.2 Conseil Européen pour la Recherche Nucléaire, CERN	3
1.3 The Large Hadron Collider, LHC	5
1.3.1 Core information	5
1.3.2 Experiments	9
1.3.3 Future aspects	10
Bibliography	14
2 The Compact Muon Solenoid	17
2.1 Introduction	17
2.2 Coordinate system	18
2.3 Detector compartments	20
2.3.1 Superconducting magnet	20
2.3.2 Inner tracking system	21
2.3.3 Electromagnetic calorimeter	25
2.3.4 Hadronic calorimeter	29
2.3.5 The muon system	33
2.4 Trigger	36
Bibliography	41
3 The Standard Model	45
3.1 Introduction	45
3.2 Elementary particles and forces	45
3.3 Symmetries and Lagrangian formalism	48
3.4 Quantum Electrodynamics and the $U(1)$ symmetry	51

3.5	Quantum Chromodynamics and the $SU(3)$ symmetry	53
3.6	Flavordynamics and the $SU(2)$ symmetry	55
3.7	Spontaneous symmetry breaking and the Higgs mechanism	58
3.7.1	Spontaneous breaking of a discrete symmetry	58
3.7.2	Spontaneous breaking of a continuous global symmetry	60
3.7.3	Spontaneous breaking of a continuous local symmetry	60
3.8	Unification – Act I	62
3.8.1	Electroweak interactions and the $SU(2)_L \otimes U(1)_Y$ symmetry	62
3.8.2	Electroweak symmetry breaking $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$	65
3.9	Unification – Act II	67
3.9.1	The Standard Model final Lagrangian	67
3.10	Unification – Act III	68
	Bibliography	69
4	Analysis Objects – Jets	71
4.1	Introduction	71
4.2	Consequences of Quantum Chromodynamics	71
4.2.1	Hadrons	71
4.2.2	Confinement	72
4.2.3	Asymptotic freedom and the strong coupling	73
4.3	Into the pQCD abyss	75
4.3.1	Feynmann diagrams	75
4.3.2	Parton cross sections and decay rates	76
4.3.3	Renormalization	77
4.4	Cross section predictions for pp collisions	79
4.4.1	Initial state adjustments	79
4.4.2	Final state adjustments	80
4.5	Jets	82
4.5.1	Definition	82
4.5.2	From detection to reconstruction	83
4.6	Monte Carlo event generators	90
4.6.1	Full event display	91
4.6.2	A quiver full of MCs	92
	Bibliography	94

II Main Analysis 99

5	Sample handling	101
5.1	Introduction	101
5.2	Data and Monte Carlo samples	102
5.3	Event selection and corrections	104
5.3.1	Analysis software	104
5.3.2	Event selection	105
5.3.3	Corrections	105
5.3.4	Processing stages	112
5.4	Trigger efficiencies	114
5.4.1	Prescale factors and effective luminosity	114
5.4.2	Emulation method and turn-on points	114
5.5	Detector level cross section	119
5.6	Unfolding	122
5.6.1	Probability/Response matrix	123

5.6.2	Closure Test and Backfolding	124
5.6.3	Bottom Line Test	125
5.6.4	Fake and Miss rates	126
5.6.5	Model dependence	128
5.7	Experimental uncertainties	128
5.7.1	Statistical uncertainty	129
5.7.2	Jet energy scale uncertainty	129
5.7.3	Jet energy resolution uncertainty	129
5.7.4	Luminosity uncertainty	130
5.7.5	Unfolding uncertainties	130
5.7.6	Other uncertainties	130
	Bibliography	132
6	Fixed order theory predictions	135
6.1	Introduction	135
6.2	Theory corrections	136
6.2.1	Non-perturbative corrections	136
6.2.2	Electroweak corrections	138
6.3	Theory uncertainties	139
6.3.1	Scale uncertainty	139
6.3.2	PDF uncertainty	139
6.3.3	NP uncertainty	139
6.4	Data to theory comparisons	140
	Bibliography	144
III	QCD Analysis	147
7	Impact studies	149
	Bibliography	156
	Appendices	157
A	Elementary particles extended	158
	Bibliography	159
B	Monte Carlo cross sections	160
C	Prefire maps	162
D	JER studies	163
	Bibliography	170
E	Trigger efficiency curves	171
F	Detector level distributions	175
G	Probability Matrices	176
H	Uncertainties	178
I	Non-perturbative corrections	187

J Complementary data to theory comparisons	190
---------------------------------------------------	------------

LIST OF TABLES

1	Partial χ^2 values for HERA DIS data and HERA DIS complemented with the 2016 CMS dijet measurements.	xiii
1.1	LHC schedule. Run periods and the corresponding cumulative integrated luminosity for each center of mass energy respectively.	8
3.1	Fundamental interactions within the SM along with their respective mediators, effective range, field strength, and particles that interact with.	48
3.2	Symmetries and conservation laws [5].	48
4.1	Quark quantum numbers [1].	72
4.2	List of MC configurations along with the order of the matrix element, the model for the parton shower and the hadronization algorithms, and the tune choice.	93
5.1	DAS dataset names for each year and their integrated luminosity.	102
5.2	Golden JSON files for each year [11].	103
5.3	DAS simulation dataset names for each year and their slices. The P8 abbreviation is used to refer to Pythia8 and MAD to refer to Madgraph.	105
5.4	Summary of the <code>TightLepVeto</code> Jet ID cuts for each year.	106
5.5	List of MET filters applied to each sample.	106
5.6	List of the jet veto map versions for each year [22, 23].	107
5.7	List of JEC versions used for both Data and MC in each year [26].	108
5.8	JER table versions for each year. For 2016 only the scale factors were used, and were combined to homemade resolution tables according to the study in Appendix D.	111
5.9	Effective trigger luminosity for each year.	115
5.10	Trigger turn-on points for all years.	116
5.11	Bin edges for invariant mass and rapidity. The same $m_{1,2}$ binning is used for all $ y _{max}$ bins.	119
7.1	Non-PDF parameter nominal values and their variations to estimate the model uncertainty [1].	151
7.2	Partial χ^2 values for HERA DIS data and HERA DIS complemented with the 2016 CMS dijet measurements [1].	151
B.1	Cross section values for Pythia8 \hat{p}_T sliced (top) and flat (bottom) samples for each year.	160
B.2	Cross section values for the 2016 Madgraph H_T sliced sample.	161

D.1 Binning schemes used in the resolution studies, in terms of p_T^{rec} , $|\eta^{rec}|$ and ρ . 163

LIST OF FIGURES

1.1	CERN accelerator complex	5
1.2	Size comparison between the LHC ring and the surrounding area.	6
1.3	Three-dimensional cut of a LHC's dipole	6
1.4	Beam intersection point	7
1.5	Total integrated luminosity recorded by CMS	8
1.6	Expected schedule for Phase II of the HL-LHC	11
1.7	Luminosity forecasts for Phase-2 of HL-LHC	12
1.8	Possible FCC placements	13
2.1	The CMS detector	18
2.2	CMS coordinate system	19
2.3	Pseudorapidity η to polar angle θ conversion	20
2.4	CMS transverse slice	21
2.5	The CMS superconducting magnet	22
2.6	The CMS inner tracking system	23
2.7	CMS pixel and strip detector sensor modules	23
2.8	CMS tracker in the barrel region	24
2.9	CMS tracker transverse momentum resolution and track reconstruction ef- ficiency	25
2.10	Schematic view of the CMS calorimeters	26
2.11	Examples of electromagnetic particle showers	27
2.12	CMS ECAL specifications	28
2.13	The CMS ECAL	28
2.14	The CMS hadronic calorimeter	30
2.15	Hadronic shower specifics	31
2.16	HCAL wedge numbering schemes	32
2.17	The CMS muon system	34
2.18	Drift tube chamber specifications	34
2.19	Cathode strip chamber specifications	35
2.20	The CMS L1T architecture	38
2.21	The CMS L1T upgrade	40
3.1	The elementary particles of the Standard Model	46
3.2	The potential $V = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ for $\lambda > 0$	59
3.3	The potential $V = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2$ with $\lambda > 0$ and $\mu^2 < 0$	60
4.1	Weight diagrams	72
4.2	Running of the strong coupling α_S	74
4.3	Relative strengths of the four fundamental forces	75

4.4	QCD Feynmann rules	76
4.5	Simple examples of Feynmann diagrams	76
4.6	One loop diagrams (a) and (b), and one leg diagram (c) for $e^-e^+ \rightarrow \mu^-\mu^+$	78
4.7	Signature figures on global PDF fits	81
4.8	Jet evolution, from collision to detection	82
4.9	Particle energy depositions to CMS detector	83
4.10	Sequential-recombination algorithm at work	85
4.11	Jet area for jet clustering algorithms	86
4.12	Squared average shift in jet p_T	87
4.13	Dissociation between main and pile up collision vertices	88
4.14	CMS pp collision event display	89
4.15	PF jet energy fractions	90
4.16	Simulation of a pp collision	91
5.1	Schematic view of the CMS dataset tier system	103
5.2	Schematic view of Monte Carlo production	104
5.3	Consecutive stages of JEC, for data and MC simulation	107
5.4	Prefiring effect	109
5.5	PU profiles	112
5.6	Pileup profile methods	113
5.7	Trigger efficiency curves in central rapidity regions for 2016	117
5.8	Trigger efficiency curves in forward rapidity regions for 2016	118
5.9	Double differential cross section for 2016	121
5.10	Unfolding control plots for Pythia 2016 slices	124
5.11	Two-dimensional probability matrix for 2016 Pythia slices	125
5.12	Closure test and backfolding for 2016 Pythia slices	126
5.13	Bottom Line Test for 2016 Pythia slices and model dependence for	127
5.14	Fake and miss rates for 2016 Pythia Slices	127
5.15	Relative uncertainty (%) for all experimental sources on the measured dijet mass cross section from the 2016 dataset	131
6.1	Theoretical prediction for the double differential cross section as a function of $m_{1,2}$ and k factors	136
6.2	Non-perturbative correction factors	138
6.3	Electroweak correction factors	139
6.4	Data to theory comparison for the 2016 dataset	142
6.5	Individual data to theory comparisons for the 2016 dataset	143
7.1	Parton distribution functions after fitting the HERA DIS data together with the 2016 CMS dijets	153
7.2	Comparison between parton distribution functions after fitting the HERA DIS data alone and together with the 2016 CMS dijets	154
A.1	Extended version of the elementary particles of the Standard Model	159
C.1	Prefiring probability maps	162
D.1	Response distributions for 2016 Pythia8 slices	164
D.2	Resolution curves for 2016 Pythia8 slices	165
D.3	Overview of all resolution curves for 2016 Pythia8 slices	166
D.4	Impact studies on different smearing options	167
D.5	Evolution of response fits	168
D.6	Evolution of resolution fits	169

E.1	Trigger efficiency curves in central rapidity regions for 2017	171
E.2	Trigger efficiency curves in central and forward rapidity regions for 2017 . . .	172
E.3	Trigger efficiency curves in central rapidity regions for 2018	173
E.4	Trigger efficiency curves in forward rapidity regions for 2018	174
F.1	Double differential cross section for 2017 and 2018	175
G.1	Two-dimensional probability matrix for 2017 Pythia slices	176
G.2	Two-dimensional probability matrix for 2018 Pythia slices	177
H.1	Relative uncertainty (%) for all experimental sources on the measured dijet mass cross section from the 2017 dataset	178
H.2	Relative uncertainty (%) for all experimental sources on the measured dijet mass cross section from the 2018 dataset	179
H.3	Relative uncertainty (%) for individual JES sources from the 2016 dataset . .	180
H.4	Relative uncertainty (%) for individual JES sources in the 2016 dataset . . .	181
H.5	Relative uncertainty (%) for individual JES sources in the 2016 dataset . . .	182
H.6	Relative uncertainty (%) for individual JES sources in the 2016 dataset . . .	183
H.7	Relative uncertainty (%) for individual JES sources in the 2016 dataset . . .	184
H.8	Relative uncertainty (%) for individual JES sources in the 2016 dataset . . .	185
H.9	Relative uncertainty (%) for individual JES sources in the 2016 dataset . . .	186
I.1	Non-perturbative correction factors from individual LO MC generators	188
I.2	Non-perturbative correction factors from individual MC generators	189
J.1	Data to theory comparison for the 2017 dataset	191
J.2	Individual data to theory comparisons for the 2017 dataset	192
J.3	Data to theory comparison for the 2018 dataset	193
J.4	Individual data to theory comparisons for the 2018 dataset	194

Part I

Fundamentals

Chapter 1

THE LARGE HADRON COLLIDER

1.1 Introduction

A strong spirit of devotion and sacrifice is nurtured if we first convince ourselves about the significance of the task at hand. During my undergraduate years, while I was aspired to understand the structure of the cosmos and looking to find purpose in our existence, I read my very first article on Elementary Particles – “A model of leptons” by S. Weinberg [1]. The author formulated a model that unifies the electromagnetic and weak forces where the underlying symmetry is spontaneously broken, a mechanism proposed, among others, by P. W. Higgs [2] on 1964. As an outcome a new field would arise, later named as the higgs field, the existence of which would lead to the weak force carriers acquiring their masses. At this point I would kindly ask to not be misjudged for throwing theory-related jargon that the reader might not be familiar with but to find trust in the text as everything will be properly explained in Chap. 3.

Nearly half a century later, on the 4th of July 2012, located in Geneva at the CERN main auditorium, scientists would gather to participate in the forthcoming seminar. That was the day when the two Spokespersons of the ATLAS and CMS collaborations officially announced the discovery of a new particle, the long-awaited missing piece of the Standard Model (SM), the Higgs boson [3, 4]; the media would brand it as “the God particle”.

What led to this discovery, or better, revelation was ATLAS and CMS, two of the larger experiments currently based at the CERN’s Large Hadron Collider. Scientists among different fields and cultures have been coordinating their efforts to achieve this task. To date, Tuesday 27th May, 2025, CERN’s community still grows. More people are joining the collaboration and continue to invest their time in favor of research as a lot of questions remain unresolved. Is it worth it? Answers may vary depending on the perspective of the individual under question.

In this leading thesis chapter preliminary information is shared with the reader on CERN’s evolution as a European laboratory and on the events that drove the High Energy Physics community of “that” time to build the largest particle accelerator in the world.

1.2 Conseil Européen pour la Recherche Nucléaire, CERN

Currently, CERN’s primary focus is High Energy Physics but this was not always the case. As the section title might suggest, it was originally directed to be a European Council for Nuclear Research. After World War II Europe had lost some of the momentum regarding physics research. Due to the aftermath of the war but also due to the evolution and

demanding nature of sciences, laboratory costs were increasing. In an effort to keep the talented individuals from going to North America a solution had to be found. Louis de Broglie, motivated by all that, aiming to reduce costs and to rekindle the edge of the science community, proposed the formation of a European facility on 1949. Three years later at a UNESCO conference the organization name and site was agreed upon between the 12 founding members and the acronym CERN was attributed to it [5, 6]. The founding members were namely Belgium, Britain, Denmark, France, Germany, Greece, Italy, the Netherlands, Norway, Sweden, Switzerland, and Yugoslavia, with more nations joining by the year to this undertaking.

Located in Meyrin Switzerland CERN's main headquarters are very close to the France borders. By 1959 two accelerators had been build, the Synchrocyclotron that reached 600 MeV particle energy focusing on the fields of nuclear physics, astrophysics, and medical physics and the Proton Synchrotron that enabled proton acceleration for the first time, reaching the 28 GeV range, remaining to day among the CERN's accelerator chain. Initially dedicated to research on atomic nuclei CERN's focal point was quickly pivoted towards High Energy Physics for the study of subatomic particles and their interactions. In 1965 it was agreed that the site location will be further extended over the French borders. The following years new accelerators and experiments were continuously built with the evolution of the accelerator complex progressively resembling the current state. Notable mentions are the construction of the Super Proton Synchrotron on 1976 and the Large Electron-Positron collider (LEP) [7, 8] on 1989 hence, signaling the start of the colliding-beam era.

LEP, the largest lepton collider ever built, was based on a 27 Km circular circumference tunnel about 100 m underground below France and Switzerland. It began operation on 1989, lasted about 11 years until it was finally shut down on 2000. Reaching energy values of 209 GeV it broadened the research scope for some interesting phenomena. Allowed the determination of the number of light neutrino families and reported evidence of a weak Higgs boson signal.

Progress was not only made in the physics field. Conversely, technological advancements took place that as it turns out have a huge impact on how we live our day-to-day lives. Due to the nature of the work happening at CERN and how information needed to be processed and fast and safely distributed within the premises, all these conditions led to the development of the World Wide Web and the concept of transfer protocols like the URL and http.

After year 2000, with a budget of 7.4 billion euros, plans initialized for the implementation of LEP's successor, the Large Hadron Collider (LHC) [9]; would occupy the same tunnel and accelerate protons or lead ions instead of leptons. Constructions ended by 2008 and machine commissioning followed right after, yet another era had just began. But CERN is not just the LHC, in reality there is a huge machine chain that slowly builds up the particle's energy and feeds the LHC.

Figure 1.1 shows the CERN accelerator complex (as of 2022). Our focus will remain just on the protons. It all starts from a negatively hydrogen ion source where ions get their first energy boost in the Linear accelerator 4 (Linac4) reaching 160 MeV . The use of quadrupole magnets inside Linac4 ensures a tight and relatively uniform beam. Upon exit ions are stripped of their electrons leaving behind just the protons which are injected into the Proton Synchrotron Booster (PSB); energy is ranked up to 2 GeV . Consecutively, protons are received by the Proton Synchrotron (PS), yes the one that was build back in 1959, and accelerated at 26 GeV , while, later, the Super Proton Synchrotron (SPS) increases the energy at 450 GeV . LHC is the final stage in the acceleration chain where the beam energy is slowly increased at 6.8 TeV ¹. A size comparison can be made between

¹As of 2022, LHC accelerates protons at 6.8 TeV . The actual machine limit, by design, is at

the LHC ring and the surrounding area in Fig. 1.2.

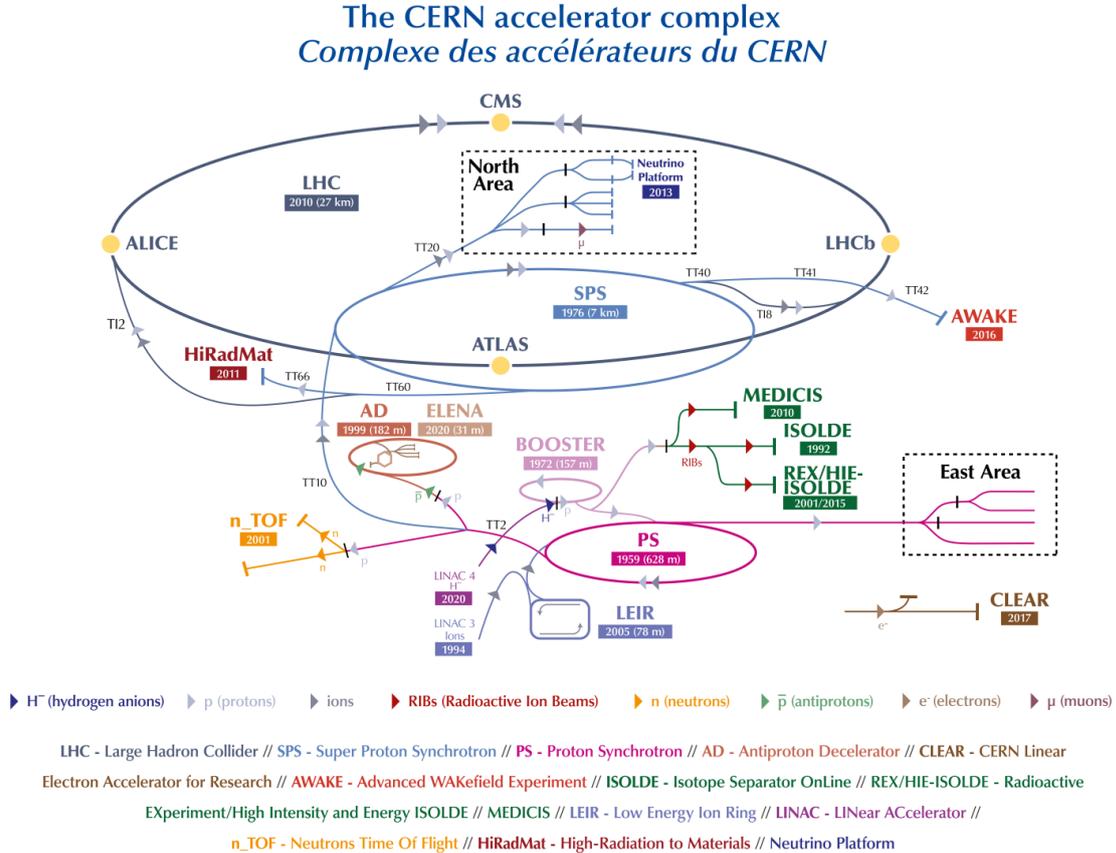


Figure 1.1: CERN accelerator complex [10].

1.3 The Large Hadron Collider, LHC

1.3.1 Core information

The LHC (see Fig. 1.3) currently holds the title of the largest, highest energy particle accelerator in the world. It has the ability to accelerate either protons or lead ions depending on experiment needs. Proton energy is gradually increased from 450 GeV to 6.8 TeV over a period of 20 mins reaching speeds close to the speed of light. In reality, protons are circulated in bunches, each of them containing $\sim 1.8 \times 10^{11}$ of them. This division, into bunches, is happening just before the exit of PSB where it is the last place that a continuous beam exists.

Bunches have a cylinder-like shape and are about 1 mm wide, whereas they are shrank down to $\sim 16 \times 16\ \mu\text{m}$ just before collision time. This way protons are more tightly packed, thus increasing the probability of a collision. Since each bunch is spread from the adjacent ones over 7.5 m the 27 Km circumference can simultaneously hold 3550 of them. However, some extra space is needed for the injection of healthy bunches and dumping of the bad ones. In the end, the effective number of bunches scales down to 2808.

7 TeV , a value that will be likely obtained in the years to come.



Figure 1.2: Size comparison between the LHC ring (yellow circle) and the surrounding area. The borders between France and Switzerland are illustrated with the dashed white line. White indicators show the location of the four large experiments on top the LHC ring. [11].

The collider is comprised of two rings, each containing its own beam, for clockwise and counterclockwise acceleration. The two beams collide at specific intersection points around which experiments are built. Collision between counter-rotating beams provide a total energy equal to the sum of the energy of each beam. This means that for the LHC the collisions happen at a center of mass energy of 13.6TeV ! The same two beams are used for 10 hours due to the beam's limited lifetime. After that all bunches are dumped and a new cycle of collisions begin with fresh bunches. Taking a closer look at the intersection point (see Fig. 1.4), there, one bunch per beam cross one another every 25nsec , a value referred to as *bunch crossing*. Converting this number to a rate gives 40MHz meaning that every second 40 million crosses occur.

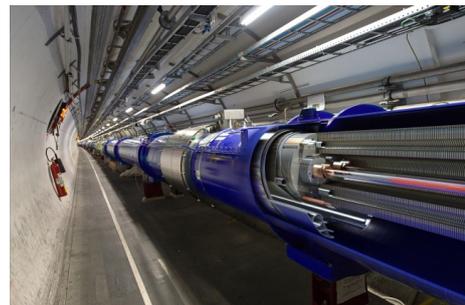


Figure 1.3: Three-dimensional cut of a LHC's dipole [12].

The architecture of the tunnel that the LHC is based is not a perfect circle. The layout is dividend into eight arcs and eight straight sections. The straight regions can serve as possible intersection or utility points. Four of them are occupied by the ATLAS, ALICE, CMS, and LHCb experiments at points 1, 2, 5, and 8 respectively. Superconducting magnets are installed throughout the whole LHC range. They operate at a nominal

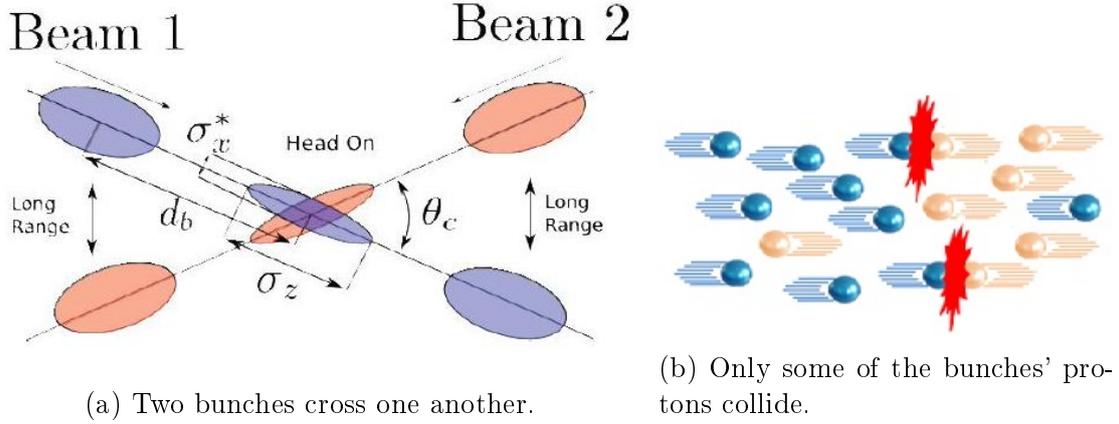


Figure 1.4: Beam intersection point [13].

field value of $8.3T$ and a current of $11.850A$. Dipole magnets are responsible for bending the beam, while quadrupole ones are responsible for focusing it. Other multipole magnets play the role of correcting the beam for small imperfections. To put some numbers into paper there are approximately 9300 magnets out of which 1232 are dipoles, 858 are quadrupoles, leaving the rest 6208 as correcting magnets. The end result is a stable and precisely aligned beam ready for use. Such strong magnetic fields are achieved because of superconductivity. Both the magnets and the coils are baptized into liquid helium at $1.8K$, greatly reducing electrical resistance and minimizing energy-to-heat conversion. This low temperature is obtained first by cooling the $80K$ helium down to $4.5K$ through the use of liquid nitrogen. Liquid helium is filled inside the cavities surrounding the magnets. After a pressure decrease to $15mbars$ the temperature falls below $2.17K$, a critical point where helium undergoes a second phase change. In this superfluid phase helium obtains a very high thermal conductivity and a negligible viscosity, properties that are utilized for heat removal and efficient transportation.

Eight Radiofrequency (RF) cavities are in place, per beam, for proton acceleration, ensuring high density and of good quality bunches. They are positioned in the straight sectors of the LHC ring with each cavity oscillating at $400MHz$, delivering $2MV$, and operating at $4.5K$. Every proton passing through a cavity receives $\sim 16MeV$ of additional energy. Ultimately, as mentioned before, protons will reach maximum energy after $20mins$ worth of revolution. The beam is enclosed in a pipe under ultrahigh vacuum of $10^{-7}Pa$ or $\sim 10^{-9}Pa$ close to the intersection points to minimize friction caused by protons colliding with gas molecules, a scenario that would otherwise negate the beam quality.

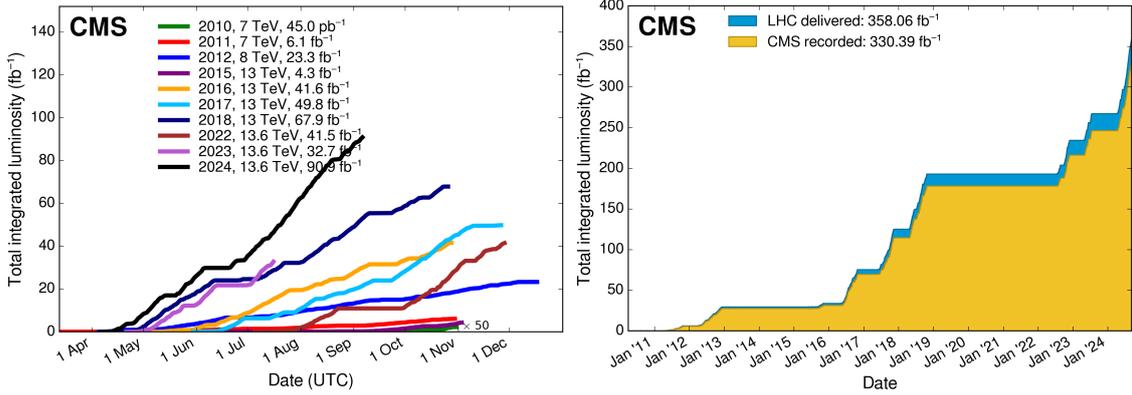
Another great quantity to look at is the *instantaneous luminosity* (\mathcal{L}) [14]. It is considered among the most important properties of an accelerator as it expresses the number of possible collisions per cm^2 and per sec . Naturally, the larger the value the larger the number of collisions will be. Luminosity can be expressed as:

$$\mathcal{L} = f \frac{N_1 N_2}{4\pi \sigma_x^* \sigma_y^*} \mathcal{F} \quad (1.1)$$

where f is the collision frequency, N_1 and N_2 are the number of protons contained within the colliding bunches, σ_x^* and σ_y^* are the transverse dimensions of the bunches and \mathcal{F} is a geometrical factor (≤ 1) that expresses the luminosity reduction caused by the angle at which the bunches collide. This number is calculated to be $\mathcal{L} = 2 \times 10^{34} cm^{-2} sec^{-1}$ for the LHC, 2×10^{34} collisions per second and per cm^2 , a record value compared to all of its predecessors. Interestingly enough taking the integral of instantaneous luminosity over a specified period of time T is, again, an important quantity:

$$\mathcal{L}_{int} = \int_0^T \mathcal{L} dt \quad (1.2)$$

Integrated luminosity (\mathcal{L}_{int}) is a measurement of the bulk volume of data collected in time T , carries the advantage of remaining constant over the period of time T , and is usually expressed in units² of pb^{-1} or fb^{-1} . Figure 1.5 shows the total integrated luminosity recorded by the CMS experiment over time. Particularly, in the left plot the luminosity is



(a) Luminosity collected by year (in different colors) as a function of time (in months). (b) Cumulative delivered (blue) and recorded (yellow) luminosity as a function of time (in years).

Figure 1.5: Total integrated luminosity recorded by CMS [15].

depicted for each year separately over the course of the LHC operation months. It can be seen that already, in September 2024, $90.9 fb^{-1}$ have been collected, a value far exceeding what was observed in the previous years. In the right plot the same information is shared but illustrated in a slightly different manner. There the cumulative luminosity is presented over the course of years. It is interesting to see that over 14 years of operation the LHC has delivered $358.06 fb^{-1}$ worth of data.

Looking at Fig. 1.5b in more detail one can perhaps distinguish the three different periods of LHC operation. Two huge plateaus can be identified, the first between 2013 and late 2015 and the second between 2019 and early 2022. These correspond to the Long Shutdown 1 (LS1) and LS2, respectively, during which LHC was halted and major upgrades were employed. Periods before and after the long shutdowns represent the run periods (Runs) where the LHC had an active beam. This information is better summarized in Table 1.1. At present, in Run III, LHC has surpassed the luminosity that had delivered in Run II, with another year still remaining before the next LS. It is estimated that by the end of Run III $\sim 300 fb^{-1}$ will have been collected.

Table 1.1: LHC schedule. Run periods and the corresponding cumulative integrated luminosity for each center of mass energy respectively.

Run I		LS1	Run II		LS2	Run III	
2010 – 2012			2015 – 2018			2022 – 2025	
7 TeV	8 TeV		13 TeV		13.6 TeV		
$\sim 6 fb^{-1}$	$\sim 23 fb^{-1}$		$\sim 163 fb^{-1}$		$\sim 165 fb^{-1}$		

²1 barn = $10^{-24} cm^2$ and $1 fb^{-1} \approx 70 \times 10^{12}$ proton-proton collisions.

Before we move on, another great quantity to have in mind is that of the *cross section* (σ) which quantifies the probability that a particular event will occur. It is expressed in units of *barn* (b) and makes a good pair when combined with the luminosity. One can calculate the number of expected events (N_{ev}) per *sec* for a particular process by following the relation

$$N_{ev} = \mathcal{L}_{int} \cdot \sigma_{ev} \quad (1.3)$$

where σ_{ev} is the cross section for the process of interest and \mathcal{L}_{int} the available integrated luminosity. We will discuss a little bit more about the cross section definition and its concept, later, in Section 4.4.

1.3.2 Experiments

As mentioned before, the four larger experiments based on the LHC ring are in sequence ATLAS, ALICE, CMS, and LHCb positioned at points 1, 2, 5, and 8 respectively. For ATLAS and CMS new caverns needed to be installed in the LEP tunnel while for the other two experiments the preexisting caverns were used. Besides the big-four, there are five smaller experiments spread over the LHC, namely the TOTEM, LHCf, MoEDAL, FASER, and SND@LHC. The scope of the physical program under study is vast. Apart from the central focus of LHC, the SM of Particle Physics, the research extends beyond it with topics like Supersymmetry, Dark Matter (DM), searches of exotic particles and studies of high density environments.

Below, a brief description will be given about the aforementioned experiments. We will lead with ATLAS and CMS since these two were the driving force for building the LHC, as a very important task had been assigned to them, the discovery of the Higgs boson.

- **ATLAS, A Toroidal LHC ApparatuS** [16]:
the largest general-purpose particle detector in the world with dimensions 46 m long, 25 m wide, and 25 m tall. It weighs 7000 t and is equipped with more than 100 million sensors, having the ability to support both proton-proton (pp) and heavy ion collisions. Its primary target is to test and verify the SM, contributed to the Higgs boson discovery, and has even the ability to discover particles beyond the SM. Lastly, it strives to find answers on the topic of DM and looks for extra dimensions in the universe. More than 5000 members are involved with the experiment.
- **CMS, Compact Muon Solenoid** [17]:
another general-purpose particle detector that shares the same ambitions as its counterpart “colleague”. It follows a cylindrical geometry and is 21 m long has a 16 m diameter and weighs 13,000 t . Its signature compartment is the superconducting solenoid magnet inside of which a 3.8 T magnetic field is achieved. Same as ATLAS, it supports both pp and heavy ion collisions. Over 4000 members are associated with the experiment.
- **ALICE, A Large Ion Collider Experiment** [18]:
despite what the name might imply ALICE supports both heavy ion and pp collisions. Primarily, the experiment runs on lead-lead collisions but lighter ions are also considered. Its dimensions are 16 \times 16 \times 26 m^3 weighting about 10,000 t . ALICE studies the strongly interacting matter at extreme energy densities, a situation which is achieved during the lead nuclei collisions. There the temperature can rise as much as 100,000 times the temperature inside the core of the Sun. Under these conditions new phases of matter can be achieved like the quark-gluon plasma. One can say that in this experiment a simulation of the state of matter is performed as it was assumed to be in the early moments of the universe, right after the Big-Bang. More than 2000 members contribute to this collaboration.

- **LHCb**, Large Hadron Collider beauty[19]: devoted to study the CP violation, in hopes of providing information for the matter-antimatter asymmetry in our universe, and rare decays of beauty or charm hadrons. It extends 21 *m* long, 13 *m* wide, and 10 *m* high, weighing 5600 *t*. Due to the nature of the physical processes of interest LHCb performs well in detecting particles in the forward region, as, such decays mostly happen to small angles, close to the beam line. There are about 1400 members.
- **TOTEM**, **TOTAL** cross section – **Elastic** scattering and diffraction dissociation **M**asurement[20]: placed partially in the center of the CMS detector and about 200 *m* away from the interaction point towards the direction of the beam line. It provides a precise measurement of the total pp cross section and performs mostly studies on elastic and diffractive scattering. It amounts to about 80 scientists.
- **LHCf**, Large Hadron Collider forward [21]: this detector is positioned on either side of the ATLAS experiment about 140 *m* away from the interaction point. It focuses in detecting neutral particles, especially pions, in the forward region, aiming to understand how particle showers are created and evolve in the atmosphere when highly energetic cosmic rays collide with the air’s molecules. It is comprised by about 30 members.
- **MoEDAL**, the **M**onopole and **E**xotics **D**etector **A**t the **LHC** [22]: well there is not much more to explain. Attached to one end of LHCb, MoEDAL is armed with only one true purpose, the detection of exotic particles that lie outside the reach of the SM. A few examples are magnetic monopoles, dyons and black-hole remnants. It “shoots the thrill” for potential new physics. During LS2 the detector was re-branded as MoEDAL-MAPP where a new detector compartment was added, MAPP (MoEDAL Apparatus for Penetrating Particles), in an effort to increase the sensitivity to milli-charged particles and long-lived exotic particles.
- **FASER**, **F**orw**A**rd **S**earch **E**xpe**R**iment [23]: installed about 480 *m* downstream from the ATLAS detector it fills all specifications required for forward particle detection of light and extremely weakly interacting particles. Recently, another detector was paired with FASER, that is, the FASER ν dedicated on neutrino detection, first of its kind in the LHC family!
- **SND@LHC**, **S**cattering and **N**eutrino **D**etector **a**t the **LHC** [24]: the newest addition to the LHC experiments and a complementary detector to FASER ν . It is located about 480 *m* away from ATLAS but contrary to FASER, it is located on the opposite side of ATLAS and slightly off the beam line direction. This way neutrinos are detected in an otherwise unexplored region (pseudorapidity³ $7.2 < \eta < 8.6$).

1.3.3 Future aspects

Let us make a brief pause for a moment as one question still remains in the background. With one more year still remaining before the end of Run III, the obvious question is, what lies ahead? By end of 2025, LHC will be closing the first big chapter of its lifetime. The schedule given in Table 1.1 corresponds only to Phase-1 of the LHC. After that the LHC will enter Phase-2, a new era, the High Luminosity LHC (HL-LHC) [25]. Figure 1.6

³Pseudorapidity η is another quantity to measure the angle of particles relative to the beam axis. It will be further discussed on Chapter 2.

shows the expected schedule for Phase-2, while the actual one remains on the dynamic side, depending on the obstacles or set backs that will be encountered.

During the LS3 (2026-2028) major upgrades will take place to LHC's architecture, design, and technology, of course the experiments will have to comply too. Some notable upgrades are mentioned. First and foremost, the proton energy will be ramped up to the nominal design value of 7TeV , meaning that pp collisions will then be happening at a center of mass energy of 14TeV . Luminosity will be increased by a factor of ~ 5 above the nominal value. This number is expected to produce at least 140 collisions per bunch crossing, compared to the current number which is ~ 40 . As an outcome, a set of new, stronger magnets will be installed for focusing and bending the beam. It is estimated that these will run on a $11-12\text{T}$ magnetic field! Crab cavities are another advancement that will be installed in the LHC ring which will allow bunches to be stretched just before collision, increasing the overlap area so that a higher chance of collision is achieved. It is foreseen that by the end of Phase-2 around 3000fb^{-1} (see Fig. 1.7) will have been delivered.

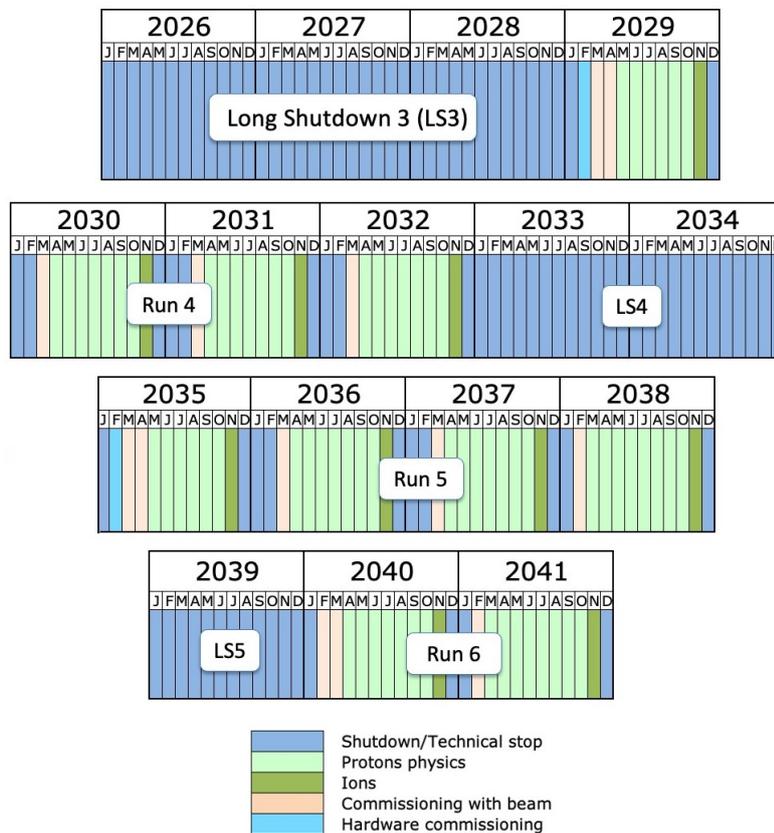


Figure 1.6: Expected schedule for Phase-2 of the HL-LHC [26].

How about a scenario that the LHC is just another machine in the accelerator chain of a larger collider? How this would look like? This is neither a philosophical question nor a matter of fiction. On 2019 a conceptual design report was released featuring a collider that will surpass the LHC, that is the Future Circular Collider (FCC). In reality, three separate proposals were made each considering a different type of collision, the FCC-ee for electron-positron collisions, the FCC-eh for electron-hadron, and the FCC-hh for hadron-hadron (the report for the latter one can be found on [27]).

For pp collisions the target nominal center of mass energy is 100TeV ! While it is not yet fully clear what an optimal value would be, there are arguments that support the 10^2TeV range. No strong evidence or suspicions of new physics drive this number but it is a certainty that at this scale the Higgs self coupling can be studied. At the very least, fully

understanding the properties of the Higgs boson might give the science community hints or ideas for future searches. With gravity still not incorporated in the SM, searches for DM candidates are of great importance. Weakly Interacting Massive Particles (WIMPs) are just one of the many alternatives that currently exist in explaining the origin of DM. While dedicated experiments, spread around and outside the world, search for them, the need for direct confirmation by accelerators is always needed. An upper limit to their mass has been set in the $1 - 3\text{TeV}$ range, making FCC the ideal place for such a search.

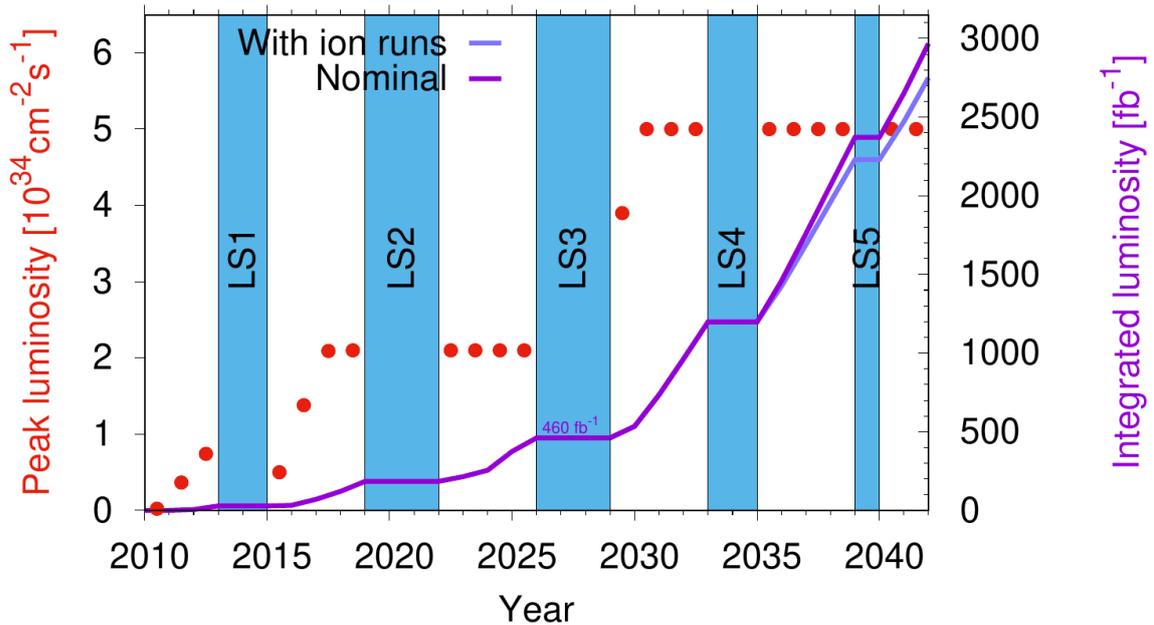


Figure 1.7: Luminosity forecasts for Phase-2 of HL-LHC. The two lines correspond to the integrated luminosity versus time with purple giving the nominal pp values and blue the heavy ion ones. Red points illustrate the expected peak operation luminosity [28].

The 100TeV requirement opens up a new challenge. In order to reach this energy value the experts have estimated that a $90 - 100\text{Km}$ circumference tunnel needs to be excavated. This alone perhaps is the key to FCC’s success or failure. Possible placement scenarios have been analyzed taking into account the geometry of the tunnel, the surface area that needs to be covered, site locations, and resources among other factors. Figure 1.8 illustrates some of the resulted schemes. Colored rings correspond to possible FCC layouts, which will support eight surface sites, while in blue the CERN LHC and SPS accelerators are shown.

It seems that the science community likes to look ahead in time, always prepared to face future challenges and grow. In this non ending game of “information warfare” where one puts a full stop? Venturing in the unknown, people are urged to seek answers with means that in the past years would hardly make any sense. Did CERN and the LHC, this European laboratory, succeeded in its purpose? Thousands of people among different cultures are devoting years of their lives for a common cause, they advance science and technology by requiring unique solutions to their complex problems, find answers that help us understand the structure of our universe, they exchange ideas and philosophies, they collaborate.

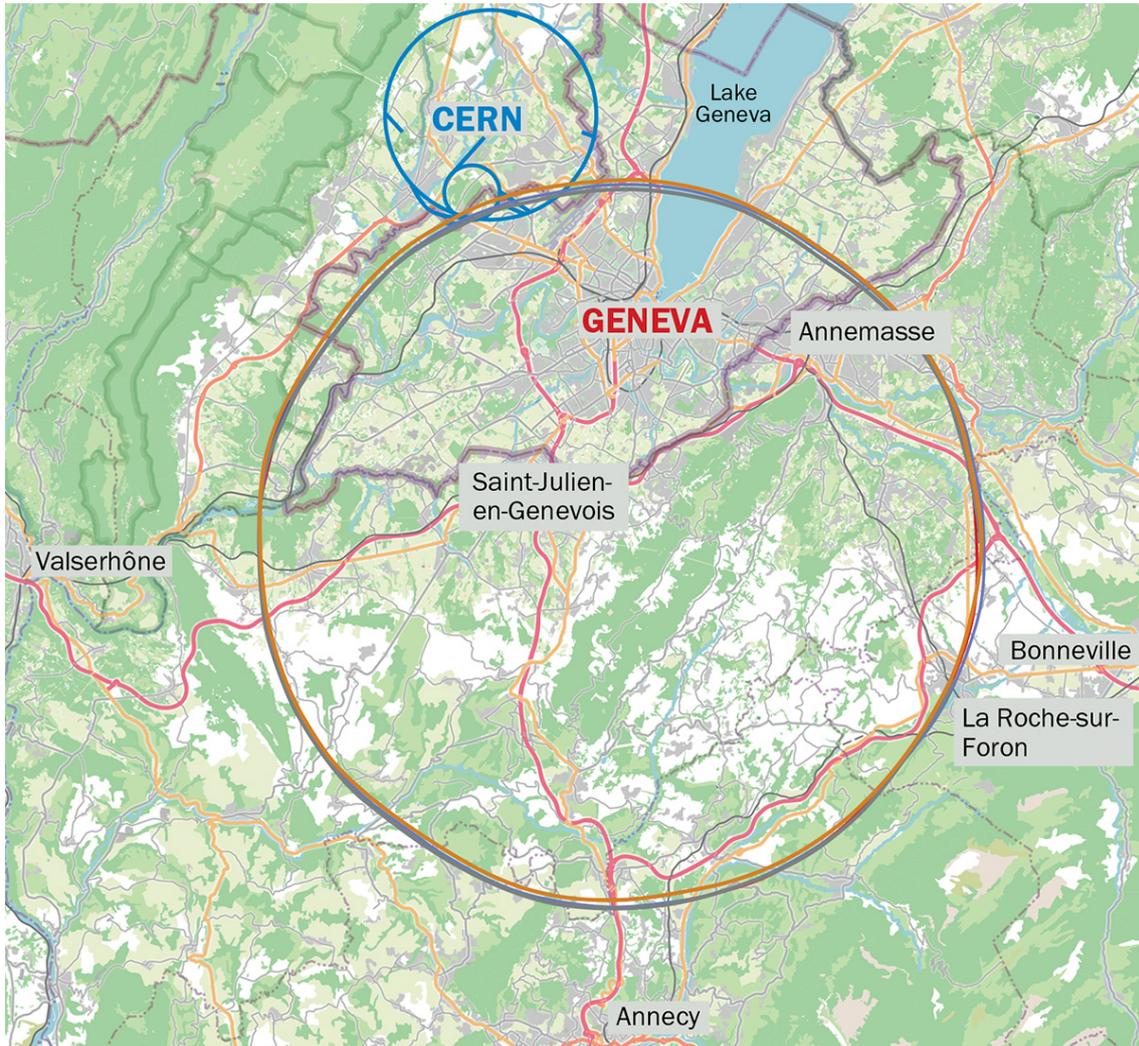


Figure 1.8: Possible FCC placements in colored rings. The LHC and SPS are shown in blue. [29].

Bibliography

- [1] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [2] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [3] G. Aad et al. [ATLAS Collaboration]. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. <https://doi.org/10.1016/j.physletb.2012.08.020>.
- [4] S. Chatrchyan et al. [CMS Collaboration]. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [5] Xabier Cid Vidal and Ramon Cid Manzano. CERN in short. Available: https://lhc-closer.es/taking_a_closer_look_at_lhc/0.cern_in_short. Accessed: 2024-07-26.
- [6] Xabier Cid Vidal and Ramon Cid Manzano. CERN Timeline. Available: https://lhc-closer.es/taking_a_closer_look_at_lhc/0.cern_timeline. Accessed: 2024-07-26.
- [7] C Pellegrini, C Rubbia, and A G Ruggiero. Comments on the desirability and feasibility study for e^+e^- colliding rings with $L 10^{32}sec^{-1}cm^{-2}$ and $E_{cm} = 120GeV$. Technical report, CERN, Geneva, 1975. <https://cds.cern.ch/record/1161844>.
- [8] B. Richter. Very high energy electron-positron colliding beams for the study of weak interactions. *Nuclear Instruments and Methods*, 136(1):47–60, 1976. [https://doi.org/10.1016/0029-554X\(76\)90396-7](https://doi.org/10.1016/0029-554X(76)90396-7).
- [9] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>.
- [10] Fabienne Landua. The CERN accelerator complex layout in 2022. Complexe des accélérateurs du CERN en janvier 2022. 2022. <https://cds.cern.ch/record/2813716>.
- [11] Harriet Jarlett. 25 years of Large Hadron Collider experimental programme. Available: <https://home.cern/news/news/cern/25-years-large-hadron-collider-experimental-programme>. Accessed: 2024-09-04.
- [12] Daniel Dominguez. LHC images gallery. Available: <https://home.cern/resources/image/accelerators/lhc-images-gallery>. Accessed: 2024-09-04.
- [13] Xabier Cid Vidal and Ramon Cid Manzano. LHC p collisions. Available: https://lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_p_collisions. Accessed: 2024-09-04.
- [14] Werner Herr and B Muratori. Concept of luminosity. 2006. <https://cds.cern.ch/record/941318>.
- [15] Anton Babaev. CMS Luminosity - Public Results. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>. Accessed: 2024-09-07.

- [16] G. Aad et al. [ATLAS Collaboration]. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.
- [17] S. Chatrchyan et al. [CMS Collaboration]. The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment. *JINST*, 3:S08004, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08004>.
- [18] K. Aamodt et al. [ALICE Collaboration]. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08002>.
- [19] Jr. Alves et al. [LHCb Collaboration]. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08005>.
- [20] G. Anelli et al. [TOTEM Collaboration]. The TOTEM experiment at the CERN Large Hadron Collider. *JINST*, 3:S08007, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08007>.
- [21] O. Adriani et al. [LHCf Collaboration]. The LHCf detector at the CERN Large Hadron Collider. *JINST*, 3:S08006, 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08006>.
- [22] J. Pinfold et al. [MoEDAL Collaboration]. Technical Design Report of the MoEDAL Experiment. Technical report, 2009. <https://cds.cern.ch/record/1181486>.
- [23] A. Ariga et al. [FASER Collaboration]. FASER: ForwArd Search ExpeRiment at the LHC, 2019. <https://arxiv.org/abs/1901.04468>.
- [24] G. Acampora et al. [SND@LHC Collaboration]. SND@LHC: The Scattering and Neutrino Detector at the LHC. *JINST*, 19(05):P05067, 2024. <https://dx.doi.org/10.1088/1748-0221/19/05/P05067>.
- [25] O. Aberle et al. [CERN LHC]. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2020. <https://cds.cern.ch/record/2749422>.
- [26] Xabier Cid Vidal and Ramon Cid Manzano. LHC Timeline. Available: https://lhc-closer.es/taking_a_closer_look_at_lhc/0.lhc_timeline. Accessed: 2024-09-11.
- [27] A. Abada, M. Abbrescia, S.S. AbdusSalam, et al. FCC-hh: The Hadron Collider. *The European Physical Journal Special Topics*, 228:755–1107, 2019. <https://doi.org/10.1140/epjst/e2019-900087-0>.
- [28] Rogelio Tomas Garcia. Luminosity forecasts. Available: <https://lhc-commissioning.web.cern.ch/schedule/HL-LHC-plots.htm>. Accessed: 2024-09-11.
- [29] Johannes Gutleber CERN. Host states gear up to work on FCC. Available: <https://cerncourier.com/a/host-states-gear-up-to-work-on-fcc/>. Accessed: 2024-09-12.

Chapter 2

THE COMPACT MUON SOLENOID

2.1 Introduction

Mounted with bunches of protons traveling at the speed of light the LHC remains but a plain observer when it comes to matters of measurement and interpretation. In an effort to find meaning in the teachings of the previous chapter the discussion needs to be further extended to the topic of detectors [1]. Collisions at the LHC take place in four designated locations, information that I convey for the last time, I promise. Detectors are then build around the collision points, usually referred to also as *interaction points* (IP), so that all, or most of the particles produced upon collision are captured within them. They play the role of cameras, taking a “photograph” of all particles in a given instance. This chapter gives a detailed description of the Compact Muon Solenoid (CMS) detector located at the LHC Point 5 (P5), its coordinate system, and individual compartments.

Aiming to construct a detector capable of measuring the signal of the Higgs boson, CMS was bound to have very specific requirements (taken from [2]):

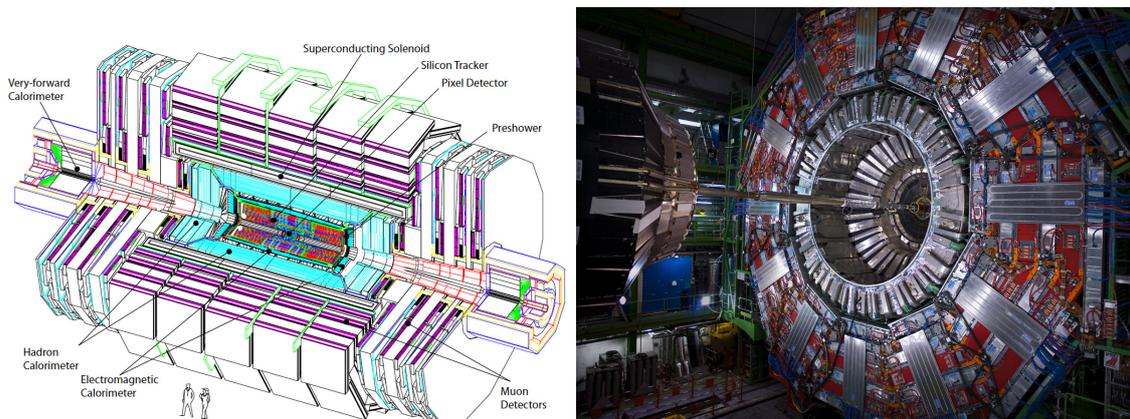
- Good *muon identification* and *momentum* resolution over a wide range of momenta and angles, good *dimuon mass* resolution ($\approx 1\%$ at 100 GeV), and the ability to determine unambiguously the charge of muons with $p < 1\text{ TeV}$.
- Good *charged-particle momentum* resolution and *reconstruction efficiency* in the inner tracker. Efficient triggering and offline tagging of τ 's and b -jets, requiring pixel detectors close to the interaction region.
- Good *electromagnetic energy* resolution, good *diphoton* and *dielectron mass* resolution ($\approx 1\%$ at 100 GeV), wide geometric coverage, π^0 rejection, and efficient *photon* and *lepton isolation* at high luminosities.
- Good *missing-transverse-energy* and *dijet-mass* resolution, requiring hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

And the CMS delivered! Looking at the articles related to the Higgs discovery from ATLAS [3] and CMS [4], it is easy to make an immediate connection between the requirements listed above and the decay channels explored during the offline analysis. The decay channels are given below:

1. $H \rightarrow \gamma\gamma$.
2. $H \rightarrow ZZ \rightarrow 4l$, with l either e or μ .

3. $H \rightarrow W^+W^- \rightarrow l\nu_l l\nu_l$, with l and ν_l either e or μ and ν_e or ν_μ respectively.
4. $H \rightarrow \tau^+\tau^- \rightarrow$ two different decays between $e\mu / \mu\mu / e\tau_h / \mu\tau_h$, with τ_h a hadronic decay.
5. $H \rightarrow b\bar{b} \rightarrow$ dijet resonance in association with a W or Z boson decay of the type $W(l\nu_l)$, $Z(ll)$, and $Z(\nu_l\nu_l)$.

The actual detector is illustrated in Fig. 2.1; (a): a perspective view that individual subdetectors are highlighted and a comparison to human size is made, (b): a real picture where part of the endcap and a wheel are given left and right respectively. CMS is not a single entity. In the contrary, is composed by individual subdetectors, each one excelling in the detection of a specific category of particles and their interactions, are well timely coordinated and combined they work as a unit, the CMS. This is better portrayed in Fig. 2.1a. Subdetectors evolve around the IP (found in the center of CMS) following an "onion"-like structure. First, a tracker system is in place to measure the tracks of charged particles. Calorimeters come right after aiming to quantify the particles' energy. A superconducting solenoid magnet then encloses the aforementioned subdetectors enabling the bending of the charged particles' trajectory. Lastly, outside the superconducting solenoid perimeter lies the muon detector for the efficient and accurate estimation of the muons' momentum and trajectory.



(a) Sketch view of CMS and its components [2]. (b) Actual picture of a transverse slice of CMS [5].

Figure 2.1: The CMS detector in its entirety.

Still, undoubtedly one of the key elements of CMS is missing from Fig 2.1a. From one point of view this is natural as it is not directly attached to the detector itself. This element is the trigger system a concurrent hardware and software based readout system devoted to scrutinize all collision events¹ during run time by selecting only the most interesting ones for storage. A topic which will be revisited in the coming sections.

2.2 Coordinate system

CMS follows a cylindrical geometry. The origin of its coordinate system (see Fig. 2.2) is placed at the center of CMS, exactly at the IP. Having a right-handed coordinate system the x -axis points radially inward to the center of the LHC, the y -axis extends vertically

¹In the realm of collider physics collisions are usually referred to also as *events* or *collision events* which is an instance of two proton bunches passing each other (see Fig. 1.4).

upward, leaving the z -axis to point to the counter-clockwise beam direction toward the Jura mountains. In spherical coordinates the momentum \vec{p} of a particle can be expressed in the xyz space as $\vec{p} = (|\vec{p}|, \theta, \phi)$, where $|\vec{p}|$ is the momentum's norm, θ is the *polar angle* measured from the z -axis, and ϕ is the *azimuthal angle* measured from x -axis to y -axis.

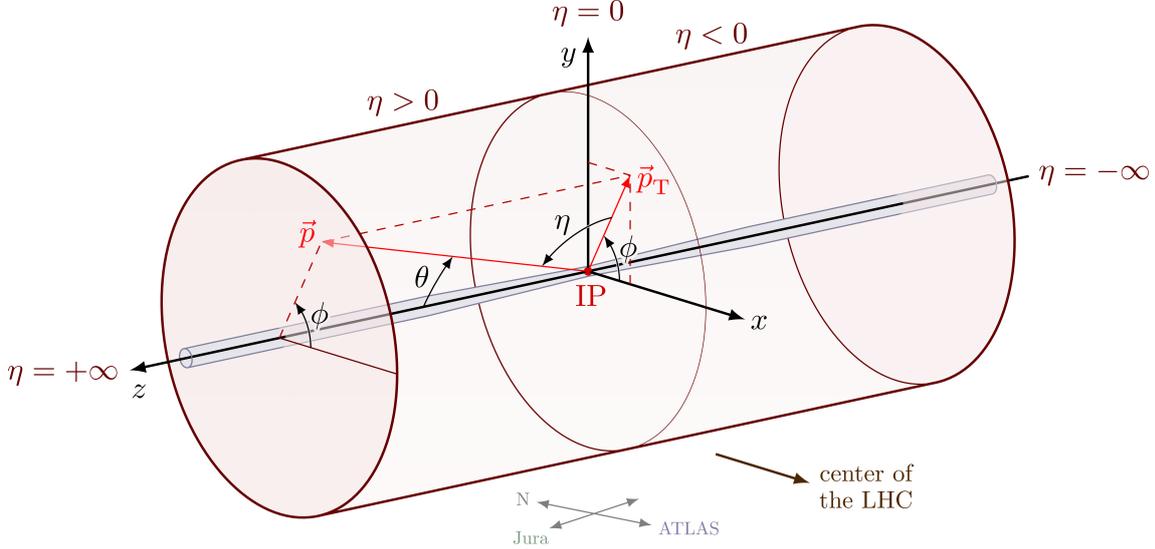


Figure 2.2: CMS coordinate system [6].

In collider physics two new quantities are introduced the *pseudorapidity* η and the *transverse momentum* p_T to replace the polar angle θ and the momentum's norm $|\vec{p}|$. The pseudorapidity η is defined as:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right] \quad (2.1)$$

where $\eta \in (+\infty, -\infty)$ (see Fig. 2.3). It expresses the “verticality” of a particle relative to the beam axis. For $\eta = 0$ the particle is found at 90° relative to the beam axis while for larger values the particle becomes more and more parallel to the beam. When a particle is seen at $|\eta| \geq 2.0$ it is said that it has been found at the *forward* region of the detector, while, otherwise that it has been found at the *central* region. Another alternative is to express η as a function of the particle's momentum:

$$\eta = \frac{1}{2} \ln \left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z} \right) = \tanh^{-1} \left(\frac{p_z}{|\vec{p}|} \right) \quad (2.2)$$

where p_z the longitudinal component of \vec{p} .

In the limit that the particle's speed is close to the speed of light, or, in other words that the mass of the particle can be considered as negligible one can then substitute in Eq. 2.2 the norm of the particle's momentum with its energy as $m \ll |\vec{p}| \Rightarrow E \approx |\vec{p}|$. A new magnitude is formed called *rapidity* y :

$$\eta \approx y \equiv \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \quad (2.3)$$

with $\eta \approx y$ holding only in the relativistic regime. The advantage of working with rapidity is that differences in this quantity are Lorentz invariant under boosts along the beam axis (z -axis). For the exact same reason the transverse momentum, particle momentum in the transverse xy plane, is preferred over $|\vec{p}|$ as p_T depends only on p_x and p_y via:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (2.4)$$

Likewise, the azimuthal angle is, again, invariant as $\phi \equiv \tan^{-1}(p_x/p_y)$. In practice the momentum of a particle is ultimately expressed as $\vec{p} = (p_T, \eta, \phi)$. Reverting back to the cartesian coordinates p_x, p_y, p_z can be always achieved with the following conversion:

$$p_x = p_T \cos \phi \quad (2.5) \quad \eta = 0$$

$$p_y = p_T \sin \phi \quad (2.6) \quad y \uparrow$$

$$p_z = p_T \sinh \eta \quad (2.7) \quad \theta = 90^\circ$$

$$|\vec{p}| = p_T \cosh \eta \quad (2.8)$$

As it turns out a great way to measure the angular separation between two particles is by exploiting both the pseudorapidity (rapidity) and the azimuthal angle:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (2.9)$$

yet another Lorentz invariant quantity which will be mentioned again in Section 4.5.2 when we dive into the jet reconstruction algorithms.

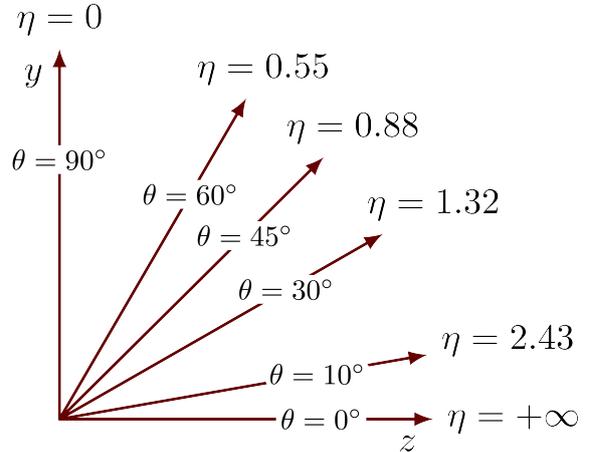


Figure 2.3: Pseudorapidity η to polar angle θ conversion [7].

2.3 Detector compartments

For this section Fig. 2.4 will remain the centerpiece of our focus as we progressively explore all subdetector compartments of CMS. With the exception of the superconducting solenoid, which will be discussed first, the other compartments will be described in successive order moving from the IP outwards, or from left to right as shown in Fig. 2.4.

2.3.1 Superconducting magnet

The CMS superconducting solenoid magnet [9] is of cylindrical shape with a 6 m internal diameter and a 12.5 m length. It is considered the largest solenoid magnet of its kind both in terms of size and maximum stored energy capability of 2.6 GJ at nominal current 19.14 kA. The coil cylinder is composed of five modules each embedded with a four-layer reinforced niobium-titanium (*NbTi*) wire winding (see Fig. 2.5a) weighting 220 t. This structure is found in the literature also as the *cold mass* [2, 10] (see Fig. 2.5b). In the inside a 4 T homogeneous magnetic field is achieved while outside the solenoid this is preserved at about 2 T. In order to enable the superconducting properties of the magnet the temperature is maintained at 4.45 K by enclosing the cold mass inside a cryostat barrel under vacuum. The magnetic flux is returned by an iron *yoke* or better *return yoke* mega structure weighing 10,000 t. The latter, is segmented into five barrel wheels and six endcap disks varying in weight from 400 t to 1920 t. Thus, the solenoid is encapsulated by the iron return yoke as it serves both as a tool to constrain the magnetic flux but also as a structural support for the solenoid itself and the rest of the subdetectors. Figure 2.5c shows the magnet during assembly, the cold mass right before it is inserted into the cryostat barrel. Additionally, one of the five return yoke wheels can be seen in the background which is the red circle-like skeleton surrounding the magnet. On the other hand, Fig. 2.5d shows an intersection of the return yoke and the coil. The color gradient illustrates the value of the magnetic field. Take note that indeed a 4 T leveled field is achieved inside and on the surface of the magnet while it fluctuates around the 2 T value inside the return yoke volume.

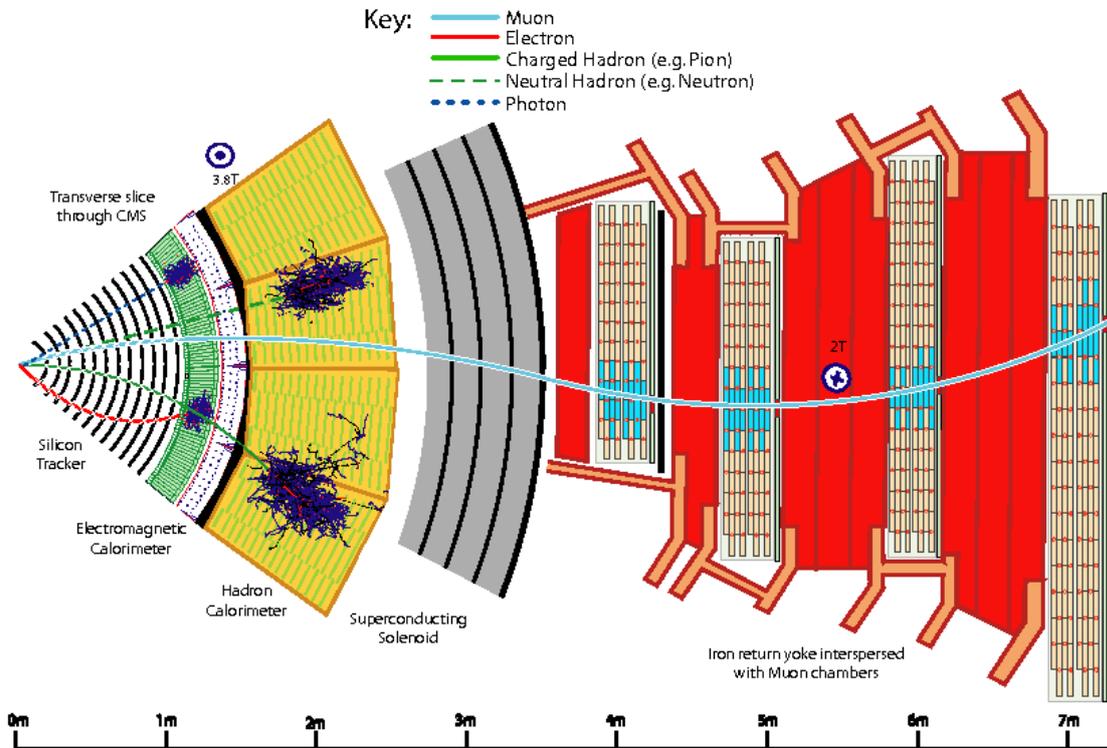


Figure 2.4: CMS transverse slice in the barrel region. Subdetector components are highlighted progressing away from the IP, from left to right. Colored lines illustrate different particles that interact with the subdetectors [8].

The existence of a magnetic field is of great essence in high energy physics experiments and there are two main arguments to back this up. The first reason stems from the need to differentiate the charged particles from the neutral ones, as the former follow a curved path when traveling in a constant magnetic field while the latter ones travel in a straight line. Another benefit of this procedure is that the charge of the particle can be assessed since a sign difference in the charge is observed as particles traveling with an opposite curvature in their trajectory. The second reason is that by measuring the trajectory of a particle it is then straight forward to estimate its transverse momentum. There are two places that permit such a measurement, the tracker and the muon chambers, in a procedure usually termed as *tracking*. The calculation is performed according to the following relationship:

$$p_T = q \cdot R \cdot B \quad (2.10)$$

where q is the charge of the particle, R the radius of the particle's orbit and B the strength of the magnetic field. Looking at Eq. 2.4 it seems that the larger the momentum of a particle the larger the radius of its orbit will be. In reverse, lower energy particles will follow paths with greater curvature. Curvature can be defined as $k \equiv 1/R$. No wonder that the word "solenoid" is included in the CMS name, after all it is recognized as one of the key characteristics of the detector.

2.3.2 Inner tracking system

The inner tracking system [12] is positioned directly around the IP at the center of CMS. It aims to provide a precise and efficient measurement of the charged particle tracks i.e., for muons, electrons, and charged hadrons and efficient reconstruction of secondary vertices.

The total length of the tracker extends to $z = \pm 270 \text{ cm}$ from the IP, has a diameter of about 230 cm , and allows a pseudorapidity coverage of $|\eta| < 2.5$, a range often referred to as *tracker acceptance*. On the performance side, the system is expected to have high granularity, fast response, and operate consistently in a high radiation environment. These requirements drove the scientists to conceive new methods and gain expertise in the field. The end result was a 200 m^2 silicon based detector composed of 1440 pixel and 15,148 strip modules arranged aptly as shown in Fig. 2.6. The whole system is divided into two main components the pixel detector and the silicon strip tracker, hashed band and light pink bands respectively in Fig. 2.6, which will be further described below.



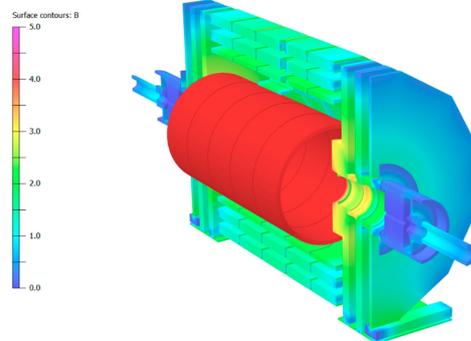
(a) The cold mass four-layer $NbTi$ coil-winding [10].



(b) The cold mass in its entirety with thermal shields [2].



(c) The cold mass just before insertion to the cryostat barrel. In red the iron return for the solenoid and for an intersection of yoke in the CMS barrel region. [10].



(d) Magnet flux density from zero to 5 T for an intersection of the iron return yoke. [11].

Figure 2.5: The CMS superconducting magnet.

Pixel detector

The pixel detector consists of three barrel layers (Barrel PIXel, BPIX) and two endcap² disks (Forward PIXel, FPIX) ensuring at least two hits in the whole tracker acceptance. BPIX layers are positioned at radii 4.4, 7.3, and 10.2 *cm* and are 53 *cm* long. On the other hand, FPIX disks are placed on each side of the barrel at radii of about 6 to 15 *cm* and away from the IP at $z = \pm 34.5$ and $z = \pm 46.5$ *cm*. Each layer contains a number of modular sensors. Consecutively, these sensor modules are equipped with 160×416 pixels each. In turn pixels have a nearly squared shape of about $100 \times 150 \mu m^2$ and provide three-dimensional hit information with respect to the $r\phi$ plane and z direction enabling good secondary vertex reconstruction. Such a module is shown in Fig. 2.7a.

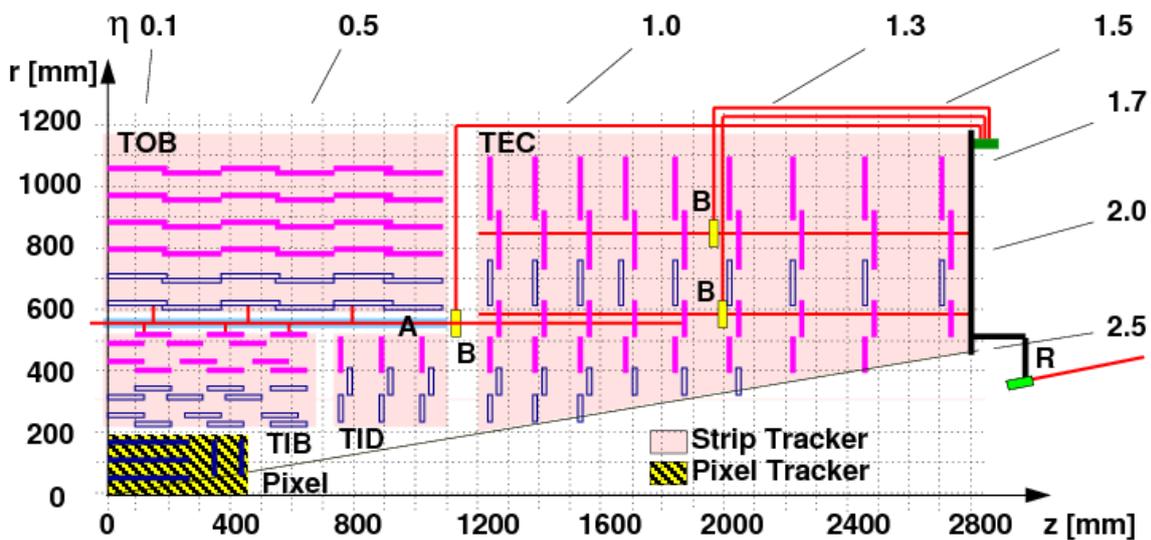
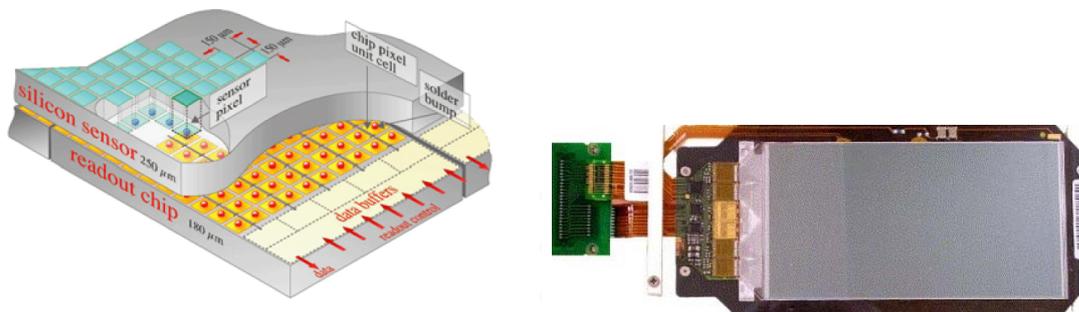


Figure 2.6: A quarter of the CMS inner tracking system in the rz plane. The hashed area shows the pixel detector and the light pink area shows the four different subsystems of the strip detector [13].



(a) Pixel detector sensor module [14].

(b) Strip detector sensor module [15].

Figure 2.7: CMS pixel and strip detector sensor modules.

Silicon strip tracker

Looking back at Fig. 2.6 the silicon strip tracker covers the range between $r = 20 - 116$ *cm* and extends from the IP to $z = \pm 270$ *cm*. It is divided into four subsystems

²Either side of the detector region is called the endcap, while the central cylindrical region is called the barrel.

to efficiently cover the whole space. In the barrel region, right after the pixel detector, there are the Tracker Inner Barrel (TIB) and Tracker Inner Disks (TID) components. The former contains four cylindrical layers of sensors at radii of 255, 339, 418.5, 498 mm and $z = \pm 700$ mm , while the latter is composed of three disks between $z = \pm 800$, mm and $z = 1000$, mm covering about the same radii range. TIB and TID are encapsulated by the Tracker Outer Barrel (TOB) which on each own is a single cylindrical structure made up of six sensor layers and four disks that are joined together via the use of special rods. It covers the radii range between 555 mm and 1160 mm and extends to $z = \pm 1090$ mm . Lastly, the Tracker EndCaps (TEC) are placed at radii of 200 mm to 1135 mm and $z = \pm 1240$ to $z = \pm 2800$ mm and each one contains nine disks.

Sensors mounted on these subsystems are based on six inch silicon wafers. There are 15 different sensor geometries (see Fig. 2.7b for a strip module instance) depending on each system's needs with some typical dimensions being 6×12 cm^2 and 10×9 cm^2 . In total there are 24,244 such sensors in the strip tracker covering an area of 198 m^2 . An average of about nine hits are ensured in the silicon strip tracker for $|\eta| < 2.4$ with at least ≈ 4 of them providing two-dimensional information. Figure 2.8 shows a sketch of all tracker layers, both pixels and strips, in the barrel region (left) and an actual photo of the silicon strip tracker in the barrel region (right).

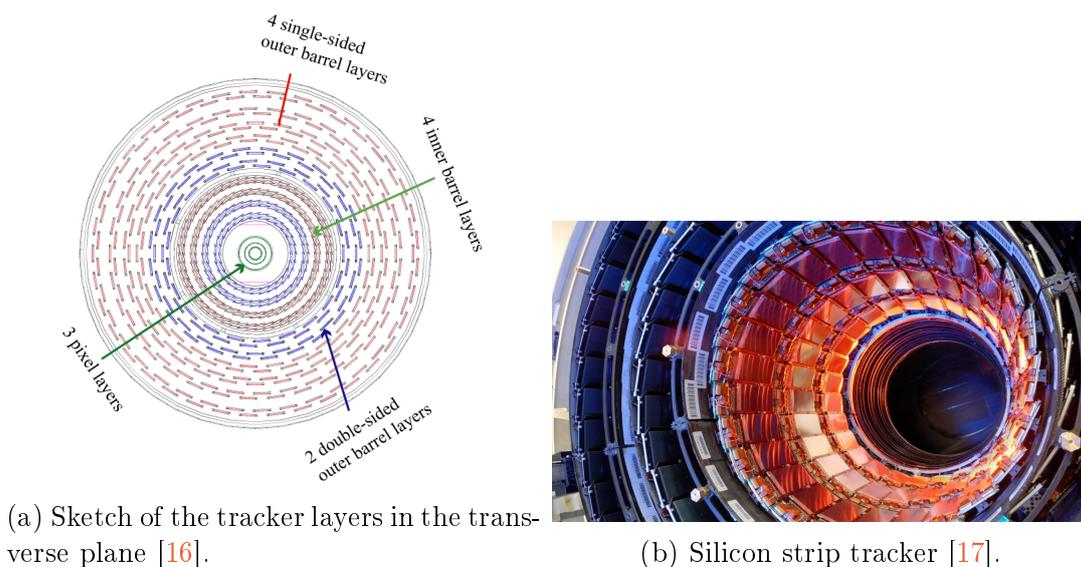
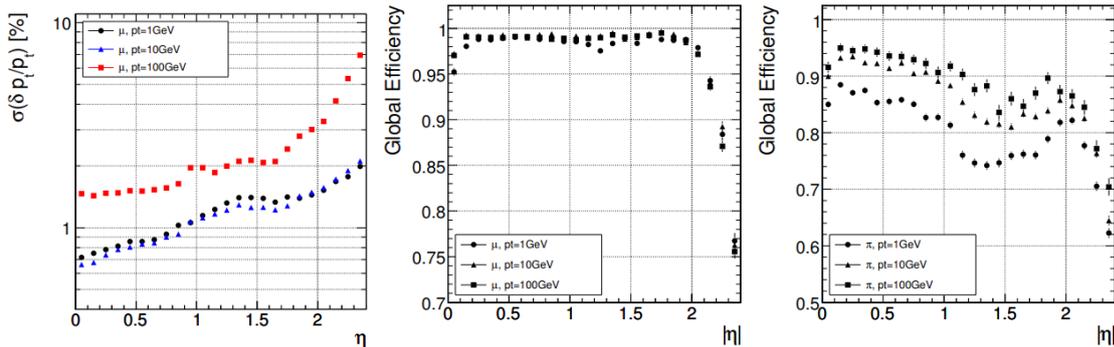


Figure 2.8: CMS tracker in the barrel region.

Figures and upgrades

The tracker as a machine is described by good charged particle spacial and transverse momentum resolution along with high efficiency track reconstruction. This is quantified in Fig. 2.9. Left plot shows the p_T muon resolution versus η inside the tracker acceptance. Different points illustrate the resolution for muons at different energies. It can be seen that the muon p_T in the forward region ($|\eta| > 2.0$) is estimated less accurately compared to the central region. Additionally, estimating the p_T for lower energy muons seem to have a greater resolution compared to higher energetic ones. Middle and right plot show the global reconstruction efficiency for muons and pions respectively, with the muon reconstruction efficiency achieving greater values $\approx 99\%$ over the whole η range. Particles at different energy regimes seem to not affect the muon reconstruction, while for pions, efficiency varies with higher p_T pions having their tracks reconstructed more efficiently.



(a) Resolution on transverse momentum for muons at different energy values. (b) Global track reconstruction efficiency for muons at different energy values. (c) Global track reconstruction efficiency for pions at different energy values.

Figure 2.9: CMS tracker transverse momentum resolution and track reconstruction efficiency [2].

The tracker was originally designed to operate at nominal LHC values of 10^{34} luminosity with an average of 20 pile up collisions at a bunch crossing rate of $25 ns$. Over the course of Run II and III these numbers have more than doubled imposing a detector upgrade. During LS1 and LS2 the tracker underwent several upgrades to cope with the exalted conditions of the LHC. Placed directly around the collision point, the pixel detector had to be totally replaced in both long shutdowns since it was worn of by the high radiation environment. By exploiting this opportunity experts could resolve any hardware or software related bugs encountered during operation and install new chips, more robust, less error prone, with reduced fake signal instances. With LS3 and HL-LHC just around the corner scientists are developing new technologies for the next tracker upgrade. The new tracker [18] will consist of two subsystems, the Inner Tracker and the Outer Tracker both of which will be able to withstand increased radiation exposure and provide higher granularity and response times.

2.3.3 Electromagnetic calorimeter

Ok, up to this point the task of tracking charged particles is covered. What about neutral particles e.g, neutrons (n), pions (π^0) and photons (γ)? Do we care about their tracks? Frankly speaking we do care, but unfortunately they do not leave any trace in the tracker since their trajectories are not bend under a magnetic field but instead follow a straight path. So there is no way to measure their transverse momentum. Or is there? Another way to measure the momentum of a particle of a known mass is to measure its energy. These three quantities (momentum, mass and energy) are connected via the four-vector formalism. This is where calorimeters come into play, and no, they do not measure calories.

The electromagnetic calorimeter [19] (ECAL) is positioned concentrically outside the inner tracker as shown with the green areas in Fig. 2.10. It belongs to the homogeneous calorimeter family meaning that in its entirety the same material is used to both initiate particle showers inside the calorimeter volume (absorbing property) and act as a scintillator to collect and provide a signal output (detection property). Two parts can be identified, the one found in the barrel region (EB) and the one covering the two endcaps (EE).

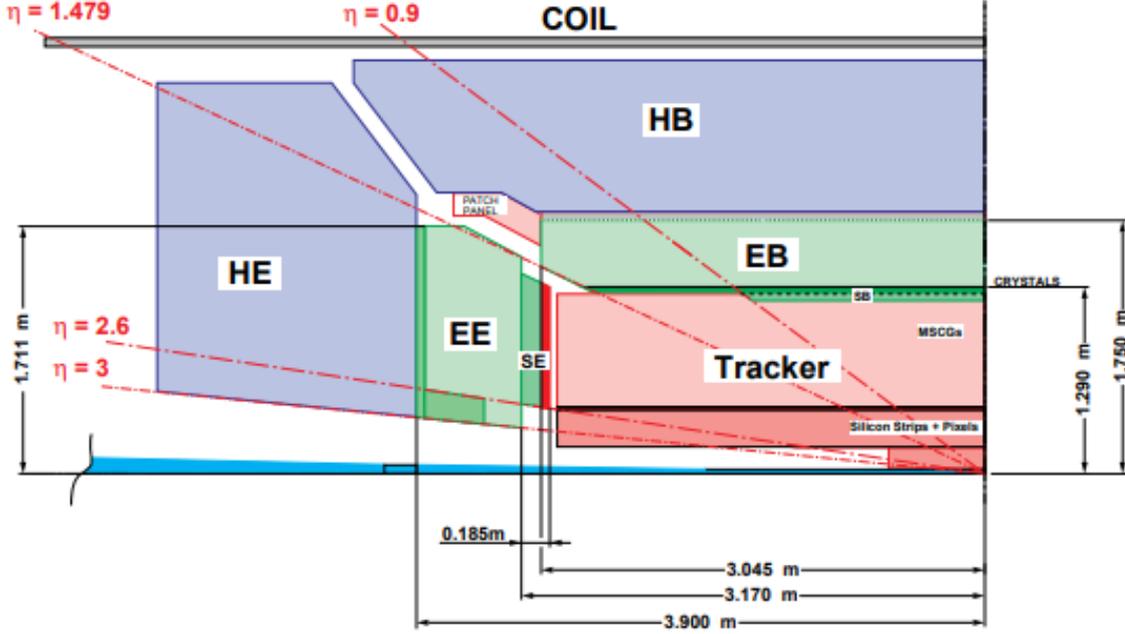


Figure 2.10: Schematic view of a quarter of the CMS calorimeters. In green the electromagnetic ones and in blue the hadronic ones [19].

Basic calorimetry

The ECAL aims at measuring the energy of all particles that mostly interact via the use of the electromagnetic force e.g., photons, electrons and positrons. This is achieved by selecting favored materials so that particles, upon contact with the detector, interact with it and deposit their energy. There are four basic mechanisms that govern this phenomenon and allow a particle shower to be initiated and propagate through the medium. For high energetic photons the prime mechanism is the *pair production*, production of an e^-e^+ pair that occurs in cases where the energy of the initial photon is at least equal or above the mass of the two electrons. The other way around, electrons (and positrons) mostly emit photons due to *bremstrahlung* radiation. In this process when a high energy electron approaches the field of an atomic nuclei, the former is decelerated, resulting to a kinetic energy loss which is attributed to a single photon emission. The cascade will progress over time with the number of produced particles exponentially increasing and their energy constantly diminishing. After a certain energy threshold is reached two additional processes will contribute to the shower evolution, the *photoelectric effect* and *Compton scattering*; an electron emission when an atom absorbs a photon and an electron release caused by photon scattering to an atom respectively.

The electromagnetic (EM) shower will eventually stop. A nice quantity to measure the depth of the cascade into the material is the *radiation length* X_0 . It is defined either as the mean distance over which an electron's energy is reduced to $1/e$ of its initial one due to bremstrahlung radiation or as the mean free path ($9/7 X_0$) for a photon before being absorbed due to pair production. The radiation length can be approximated as:

$$X_0 = 716.4 \text{ g/cm}^2 \frac{A}{Z(Z+1) \ln \frac{287}{\sqrt{Z}}} \sim \frac{1}{Z^2} \quad (2.11)$$

where Z and A are the atomic and mass numbers of the nuclei, assuming a single nucleus type contained in the absorber. In that sense, the evolution itself can be discretized into steps or layers in terms of X_0 . The total length of the cascade can then be parametrized

as a function of X_0 :

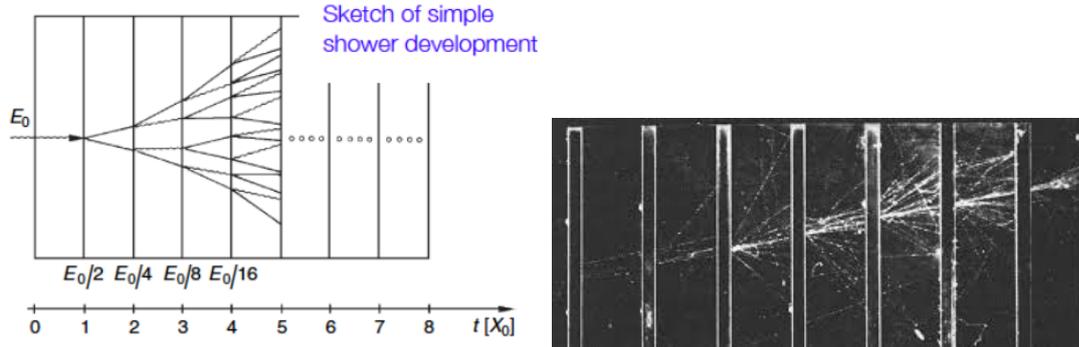
$$X_{max} = X_0 \frac{\ln(E_0/E_c)}{\ln 2} \quad (2.12)$$

where E_0 is the incident energy of the high energy particle that initiated the shower and E_c the critical energy, point at which energy losses due to ionisation become greater than bremsstrahlung ($E_c = 610 \text{ MeV}/(Z + 1.24)$ for liquids or solids). It is observed that the length increases logarithmically with the energy.

The lateral profile of the cascade is contained within a cone of radius less than one radiation length up to the shower maximum. After that the process is dominated by multiple scattering and the lateral spread depends on the *Molière radius* R_M . Ultimately, about 95% of the shower is contained within a cone of radius $2R_M$, where:

$$R_M = \frac{21 \text{ MeV}}{E_c} X_0 \quad (2.13)$$

measured again in g/cm^2 . Figure 2.11a shows a simplified version of an EM shower evolution inside a calorimeter. The shower evolves with respect to time in steps of X_0 . Figure 2.11b on the other hand shows a real picture of an electromagnetic shower evolving inside a cloud chamber with lead absorbers. It is magnificent the fact that the direction of the initial particle and of the shower itself can be distinguished.



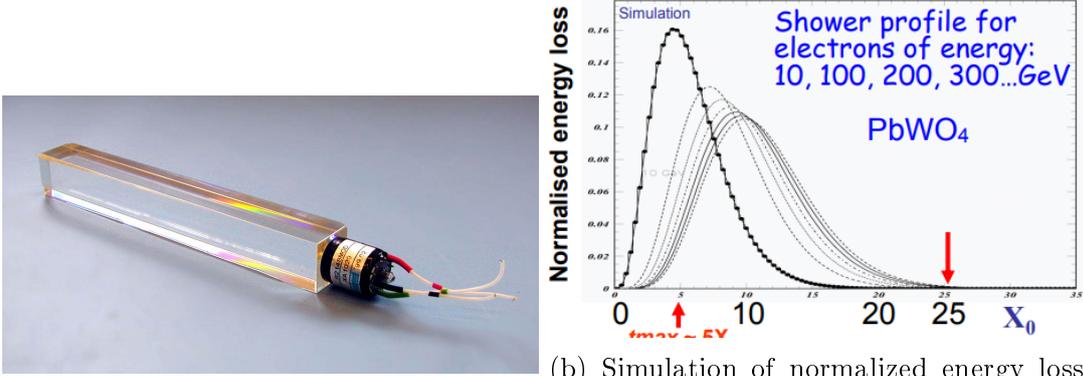
(a) Sketch of simple shower development in steps of X_0 [20]. (b) Electron shower in a cloud chamber with lead absorbers [21].

Figure 2.11: Examples of electromagnetic particle showers.

ECAL specifications

The CMS ECAL is “armed” with lead tungstate crystals ($PbWO_4$) (see Fig. 2.12a) that play the role of the absorber-scintillator. This crystal choice was based on a number of factors. The material had to be radiation-hard, allow a relative compact design to minimize the space occupancy, provide fast response and high granularity, and be as optically clear as possible for an efficient signal collection. All these specifications lie within the reach of the $PbWO_4$ crystals which provide a high density of 8.28 g/cm^3 , a short radiation length of 0.89 cm , and a small Molière radius of 2.2 cm . With that respect the total length of the crystals was decided to be 23 cm ($25.8 X_0$) in the EB and 22 cm ($24.7 X_0$) in the EE, with a rear face of $2.6 \times 2.6 \text{ cm}^2$ and $3 \times 3 \text{ cm}^2$, and a front face of $2.2 \times 2.2 \text{ cm}^2$ and $2.862 \times 2.862 \text{ cm}^2$ respectively. The crystal length choice is justified in Fig. 2.12b. There, simulated results illustrate the normalized energy loss for a particle shower as a function of X_0 for various electron energies. In the region around $25 X_0$ nearly all energy has been absorbed by the $PbWO_4$ based calorimeter. On top of that, the crystal is complemented

by fast response, in 25 ns (LHC bunch crossing rate) about 80% of the light is emitted by the material for collection. Attached to one of the crystal front faces are the light detectors. Avalanche photodiodes (APDs) are used in the barrel and vacuum phototriodes (VPTs) (see Fig. 2.12a) in the endcaps.



(a) $PbWO_4$ crystal in the EE. Attached in one end the VPT [22].

(b) Simulation of normalized energy loss versus radiation lengths for electrons with different energies traversing a $PbWO_4$ crystal [21].

Figure 2.12: CMS ECAL specifications.

There are 61,200 crystals mounted on the barrel region arranged so that there is a total ϕ coverage and extend in the pseudorapidity range of $|\eta| < 1.479$. These are grouped in modules (see Fig. 2.13a) of 400 or 500 crystals, with four such modules forming a supermodule, in the end, amounting to 36 supermodules uniformly distributed in a cylindrical manner. EB weighs 67.4 t and holds a volume of 8.14 m^3 . In turn, the EEs contain 7,324 crystals each and cover the range $1.479 < |\eta| < 3.0$. Crystals are arranged circularly, grouped in 5×5 “mini” modules, called supercrystals or SCs, structured in such a way to form the 24 t endcap disks. Endcap disks are divided into two semicircles, called *Dees* (see Fig. 2.13b), each holding half of the 7,324 crystals.



(a) Barrel 500-module mounted with all crystals.

(b) Endcap Dee mounted with supercrystals.

Figure 2.13: The CMS ECAL [2].

The calorimeter energy resolution can be parametrized as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (2.14)$$

as a function of three separate contributors, the *stochastic S* term; expressing fluctuations in the event to event statistics regarding the number of signal generating processes, the *noise N* term; expressing the noise in the readout electronics, digitization or noise raised due to pile up, the *constant C* term; expressing non-uniform and non-linear behavior on hardware or calibration, lost energy and channel to channel intercalibration errors. Some typical values that were found for these parameters during test performance runs with electrons in the range between 20 and 250 *GeV* are given below:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \quad (2.15)$$

where E is measured in *GeV*. It is observed that resolution will become better, improve, with increasing particle energy.

Preshower detector

In between the endcap calorimeters and the tracker an additional detector is placed which is considered as part of the ECAL system, this is the preshower (SE). This component aims to enhance the Higgs signal collection by efficiently identifying pions. Since both a Higgs and a pion can decay into two photons it is of great necessity to differentiate between the two cases. Specifically for pions, the two photons often decay very close relative to each other, ending up nearly at the same place in the detector, thus making it impossible to distinguish if the photon signal was raised from a single or double photon decay. The endcap preshower covers for this effect as it is very sensitive to detect pion decays and positioned closer to the IP, compared to the EEs, allowing the successful identification of pions.

The SE is a sampling calorimeter (we will discuss more about sampling calorimeters in Section 2.3.4) composed of two alternating layers of lead and silicon strip sensors. The lead layers act as the radiators initiating the particle shower while the silicon layers allow the particles to deposit their energy which is later translated to signal. Its rapidity coverage ranges between $1.653 < |\eta| < 2.6$ and the whole system manages to fit within a 20 *cm* thickness disk, with the disk, as mentioned before, placed directly in front of the endcap crystals, with a 2.5 *m* circumference and a 50 *cm* diameter hole in the middle to leave space for the beam pipe.

2.3.4 Hadronic calorimeter

Sequentially, the CMS hadronic calorimeter [23] (HCAL) measures the energy of those particles that mostly interact via the strong force e.g., protons (p), neutrons (n), pions (π) and kaons (K). Although its primary goal is to measure particle energy deposits, it also favors the indirect observation of non-interacting particles like neutrinos or beyond SM particles through the measurement of the missing transverse energy (MET), and promotes jet-tagging assignment prospects. HCAL is spatially restricted to fit within the space found in between the outer layer of the ECAL system and the inner layer of the solenoid magnet (see blue-shaded areas in Fig. 2.10).

Perhaps a more realistic implementation of the HCAL system can be seen in Fig. 2.14, instead of in Fig 2.10, which shows all four subsystems of the hadronic calorimeter. The hadron barrel (HB) and endcap (HE) are installed within the coil magnet, while the hadron outer (HO) and forward (HF) are placed outside of it, the former touches the outside layer of the solenoid magnet and the latter is found in the very forward region even beyond the muon detectors. In contrast to ECAL, HCAL is a sampling calorimeter.

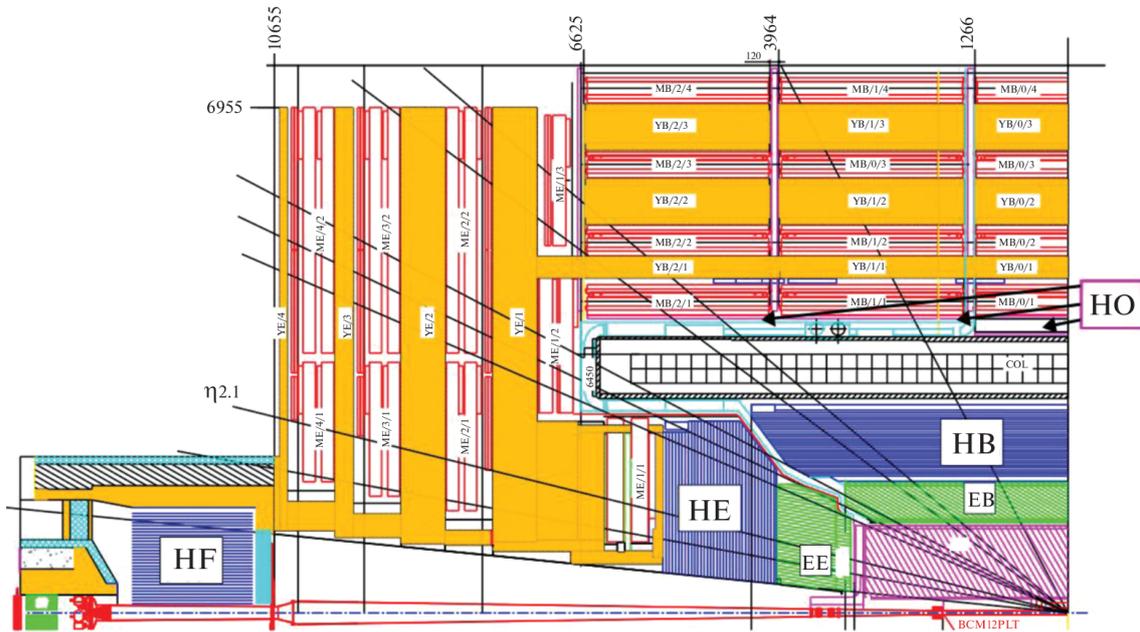


Figure 2.14: A quarter of the CMS hadronic calorimeter. The four subsystems, hadron barrel (HB), endcap (HE), outer (HO), and forward (HF), are illustrated in blue [24].

Basic calorimetry

One basic structural difference between homogeneous and sampling calorimeters is that the latter one is usually composed by two materials and not one. One material is chosen such that it has excellent absorption properties (passive layer), usually a very dense material to produce the cascade, and another one to handle the job of scintillation (active layer), to measure the energy of the particles. So sampling calorimeters consist of alternating layers of these two materials over their whole volume. This contrast, to homogeneous calorimeters, comes with a cost as some of the energy is deposited in the wrong material and, therefore, the total shower energy can not be measured directly but it has to be estimated instead. Subsequently, sampling calorimeters generally have worse resolution compared to the homogeneous ones.

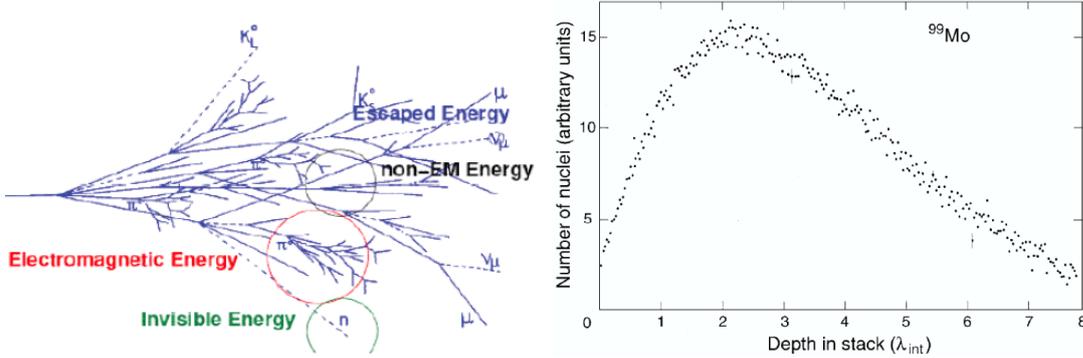
Hadronic showers are quite different, more complex, than the electromagnetic ones, but the concept of their evolution remains the same. When a high energy hadron comes into contact with the HCAL, it encounters a layer of the absorber material which initiates the cascade. Hadrons interact inelastically with the nucleus contained within the absorber resulting in the release of other hadrons (secondary particles p , n , π^\pm , K) which carry a fraction of the initial energy. The secondary particles will then repeat the same process with other nucleus and the shower will evolve. In some cases a neutral pion might be produced which will decay into two photons leading to an EM shower. Another process that can occur is the ionization of the material when a charged particle traverses through its volume, in which case a nuclei may start emitting photons as of de-excitation, again, leading to an EM shower. Figure 2.15a shows a sketch of a hadronic shower evolution. In red, an EM shower. In green, invisible energy due to nuclear binding energy losses or losses due to target recoil. In blue, in the peripheral area, possible escaped energy not contained within the calorimeter. Lastly, in black, energy attributed to hadronic (non-EM) interactions.

Hence, hadronic showers consist of a hadronic and an electromagnetic component both of which evolve in the HCAL simultaneously. In this version the shower's longitudinal

development depends on the *nuclear interaction length* λ_{int} which is the mean free path between hadronic collisions:

$$\lambda_{int} \approx (35 \text{ g/cm}^2) A^{1/3} \quad (2.16)$$

where A is the mass number of the absorber material. About 95% of the cascade is contained within a cylinder that expands less than 10 interaction lengths and has a radius of approximately one λ_{int} . These showers are much larger compared to the EM ones both in terms of reach and spread. To make an actual comparison an EM shower would reach 23 cm in length with a 2.19 cm radius for the $PbWO_4$ crystals (as discussed previously in Section 2.3.3), while a hadronic shower could be of 80 cm length with a 16.7 cm radius for iron. Another difference between the two types of cascades is that in the hadronic flavor the energy is not uniformly distributed across the shower's development. Figure 2.15b shows the longitudinal shower profile for a 300 GeV π^- traversing through a block of uranium. The number of ^{99}Mo radioactive decays are depicted as a function of λ_{int} . It is observed that the number of decays decrease dramatically in a depth of $8 \lambda_{int}$ implying the shower's total absorption.



(a) Sketch of a hadronic shower evolution [20].

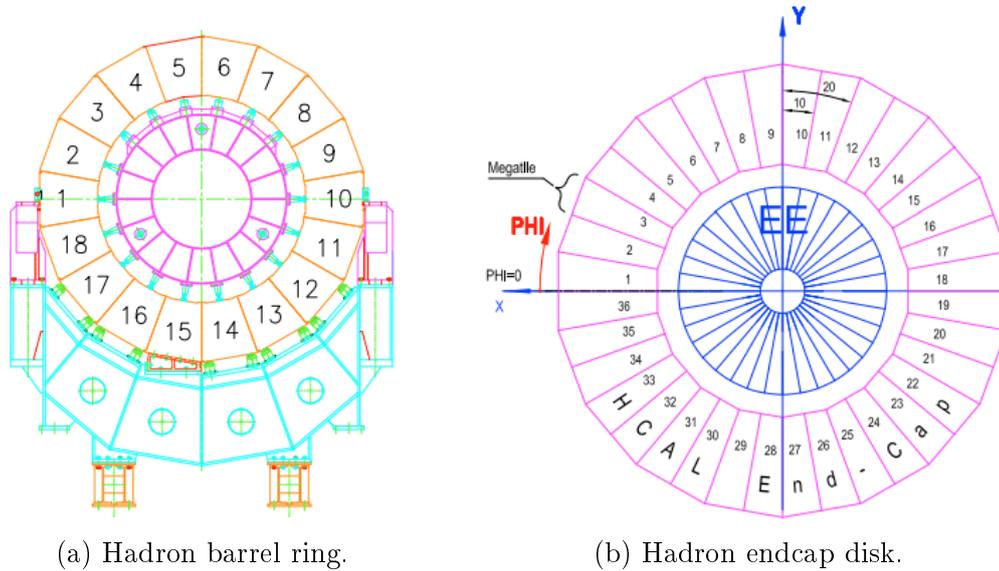
(b) Longitudinal shower profile for a 300 GeV π^- inside a block of uranium. The number of ^{99}Mo radioactive decays versus interaction length [25].

Figure 2.15: Hadronic shower specifics.

HCAL specifications

Starting from the barrel region the HB covers the range of $|\eta| < 1.3$. The barrel is divided into two halves (HB+ and HB-) each of which consists of 18 identical wedges (see Fig. 2.16a). Brass (70% Cu, 30% Zn) was chosen to be the absorber material that has a density of 8.53 g/cm^3 , radiation length X_0 of 1.49 cm, and interaction length λ_{int} of 16.42 cm. For scintillation, silicon tiles are used based mostly on Kuraray SCSN81 and with a few exceptions based on Bicron BC408. Inside the wedges alternating layers between the absorber and scintillator materials are placed. The former exists in eight 50.5 mm and six 56.5 mm thick brass plates, while in between 3.7 mm thick Kuraray SCSN81 tiles are found, with the exception of the first one, which is a 9 mm Bicron BC408. Additionally, the first and last layers in the wedges are made of 40 mm and 75 mm stainless steel plates to provide structural support. In total, each wedge is characterized by $5.82 \lambda_{int}$ at 90° that increases as $1/\sin \theta$ to $10.6 \lambda_{int}$ at $|\eta| = 1.3$. The ECAL component in front of HB adds about $1.1 \lambda_{int}$ worth of absorption material. Light is collected from the active layers via wavelength shifting (WLS) fibers that route it to the photodetectors. The same passive and active materials are used for the hadron endcap (HE) calorimeters with the individual

plates varying in size, only this time, supporting 36 wedges (see Fig. 2.16b). The HEs are mounted on top of the iron return yoke of the endcap muon chambers and cover the rapidity range between $1.3 < |\eta| < 3$. Taking into account the EE and its preshower detector the total endcap calorimeter length sums up to $10 \lambda_{int}$.



(a) Hadron barrel ring.

(b) Hadron endcap disk.

Figure 2.16: HCAL wedge numbering schemes [2].

The two barrel calorimeters (EB and HB) combined do not suffice to measure all variations of particle showers, as in some cases, like late starting showers, the cascade development grows even beyond HB. Experts had to come up with a solution, and they delivered! They added yet another calorimeter layer, this time, outside the solenoid magnet. This way the magnet itself would contribute to the shower's absorption by a factor of $1.4/\sin\theta$ interaction lengths, while the scintillator layer (same active material as in HB) would be placed directly after the solenoid. HO uses as structural support the iron return yoke, which, as mentioned before, is divided into five individual rings (2.536 m wide each) in the $r\phi$ plane, together, covering the whole barrel region. The HO layers exist independently with each other in all of five rings, positioned at $z = 0, \pm 2.686, \pm 5.342\text{ m}$. At $|\eta| = 0$ the central ring holds the minimal absorption length, therefore, two scintillator layers are placed at radii of 3.82 and 4.07 m that sandwich a 19.5 cm thick iron plate (the *tail catcher*). The four other rings contain but a single scintillator layer at $r = 4.07\text{ m}$ thus extending the total absorption length of the barrel region to a minimum of $11.8 \lambda_{int}$.

The last component of the hadron calorimeter, the hadron forward (HF), lies in the very forward region of the CMS detector. The front face of HF is positioned at $z = 11.2\text{ m}$ from the IP. It is a cylindrical structure that expands over 3.47 m to the z direction with the absorber alone having a length of 1.65 m , outer radius of 130 cm , and inner radius of 12.5 cm for the beam pipe. Quartz fibers (fused-silica core and polymer hard-cladding) are used as light collectors, a choice based on their radiation hard quality. Quartz fibers are inserted into the absorber material and run parallel to the beam pipe with the absorber consisting of 5 mm thick grooved steel plates. To mitigate the high particle rate, additional shielding components are added in an attempt to expand the calorimeter's and the electronic's lifespan. The shielding is installed in three discrete layers hermetically surrounding the HF; moving radially from the inside to outside a 40 cm thick steel layer, a 40 cm layer of concrete, and a 5 cm layer of polyethylene.

During LS3 major upgrades will be imposed on the CMS calorimeters. Specifically, the endcap calorimeters (EE and HE) are scheduled to be replaced by a new, uniform system,

under the name High-Granularity Calorimeter (HGCAL) [26, 27]. This new system will incorporate both electromagnetic and hadronic functionalities and trigger capabilities into a single working unit. Its design will be able to cope with the absurd amount of ~ 200 pile up collisions per bunch crossing, withstand extensive radiation exposure and provide a refined energy resolution. Moreover, it will greatly reinforce particle identification for a better insight into the fine structure of showers, improve pile up rejection, enhance jet tagging and introduce new, state-of-the-art, techniques like particle flow calorimetry. HGCAL will be configured as a sampling calorimeter where the active components will consist of hexagonal silicon sensors and plastic tile scintillators, while for the absorber part a mixture of copper, copper-tungsten, steel and lead is chosen. One could say that the whole system will resemble a tracker-like structure that is segmented into 50 layers with over six million readout channels.

2.3.5 The muon system

Since one of the main Higgs boson decays results in four leptons ($H \rightarrow ZZ^* \rightarrow 4l$), out of which, a pair of them or even all of them could end up as muons, a precise and efficient muon detection is of great essence in the context of the LHC physics. Muons are not important only for Higgs physics, in the contrary, they are concerned in the majority of the studies performed (e.g., CP violation, super symmetry, jet physics, cosmic rays). Muons are just like electrons, only 200 times their mass and weakly interact with matter. For this reason, the usual approach of placing a detector with the goal of absorbing them will not suffice as they penetrate matter with very few energy losses and traverse the hole volume of CMS nearly unscathed.

A different approach was taken, and was later incarnated as the CMS muon system [28]. Figure 2.17 shows, again, a transverse view of a quarter of CMS, that this time highlights the three different subsystems utilized by the muon detector. In orange, Drift Tubes (DTs) for the barrel, in green, Cathode Strip Chambers (CSCs) for the endcaps, and in blue, Resistive Plate Chambers (RPCs) for both regions. In this approach, muons leave their mark as they path their way through these subsystems. Muon’s trajectory is deduced by comparing hits of information with one another, to later attribute each found trajectory to the appropriate muon candidate. Once their trajectory has been reconstructed their transverse momentum can be calculated via Eq. 2.10. As a whole, the system allows the accurate measurement of the muon momentum and position while at the same time acts like a trigger by filtering the information based on implemented criteria.

Drift tubes

Drift Tube chambers in the barrel cover the range $|\eta| < 1.2$ and are positioned into four stations in the CMS radial direction (marked as Muon Barrel, MB1–MB4 in Fig. 2.17) in between the iron return yoke layers. Each of the first three stations consists of eight DT rectangular chambers, half of them providing a coordinate in the bend $r\phi$ plane and half of them the z coordinate parallel to the beam line. The last, outer station, only contains chambers for the $r\phi$ measurement. The principal behind their functionality is as follows. The $42 \times 13 \text{ mm}$ DT rectangular cell, shown in Fig. 2.18a, is composed of a cathode aluminum material, embedded on the walls, and a stainless steel anode wire in the center of the tube volume, filled with a mixture of 85% Ar and 15% CO_2 gas. When a charged particle, like a muon, passes through the tube it causes excitation of the gas’s atoms. As a result, the free-traveling electrons will “drift” their way to the positively charged anode where the signal is measured. Taking note of the time that the electrons need to reach the anode can provide precise information on the exact crossing point of the muon.

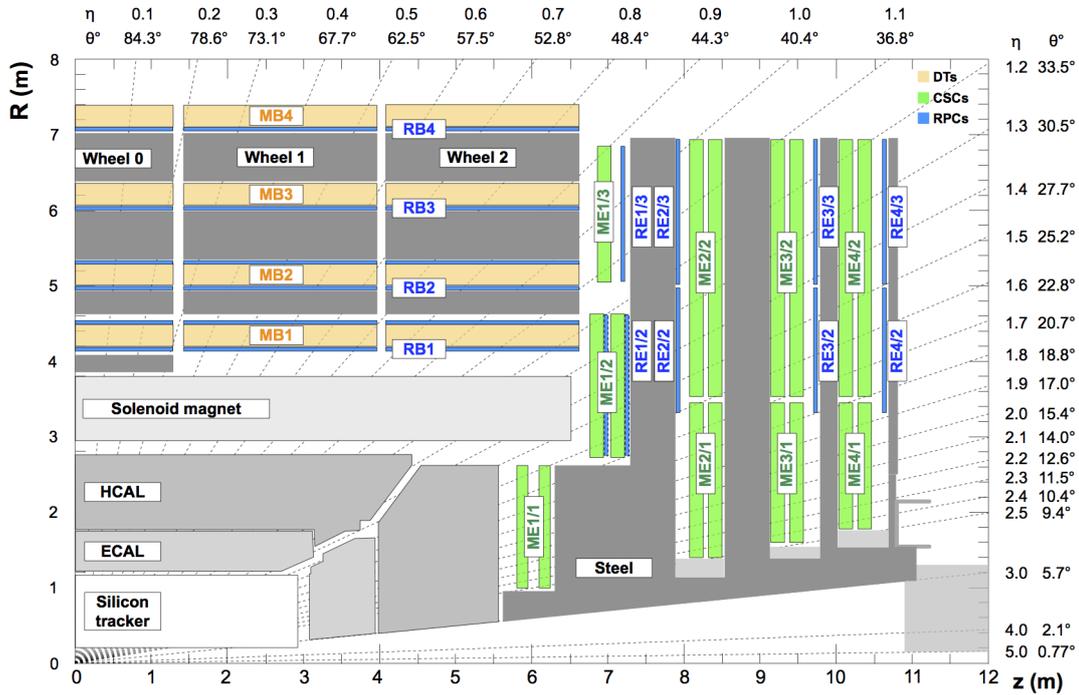
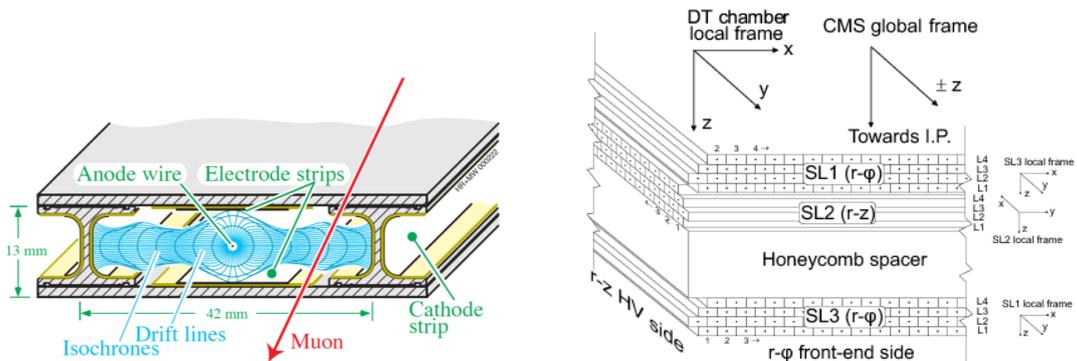


Figure 2.17: The CMS muon system. In orange the DTs (MB1-MB4), in green the CSCs (ME1-ME4), and in blue the RPCs (RB1-RB4 and RE1-RE4) [29].

Now, many of these tiny cells are grouped together, parallel to each other, to form layers. Vertically stacking four such layers form a Super Layer (SL). DT chambers usually contain three SLs, each stacked vertically with respect to the previous layer so that SL1 and SL3 perform a measurement in the $r\phi$ plane and SL2 perform a measurement in the rz plane. The design is better illustrated in Fig. 2.18b. There, the 128 mm thick aluminum honeycomb spacer can be distinguished, which is placed in between SL2 and 3 to provide additional structural rigidity and increase the angular resolution within that muon station. In total, there are 250 such chambers in the DT subsystem with the first three stations containing 60 each and the last one containing the remaining 70.



(a) DT cell unit [2].

(b) DT chamber design [30].

Figure 2.18: Drift tube chamber specifications.

Cathode Strip Chambers

In the endcap region, outside the solenoid volume, trapezoidal-shaped cathode strip chambers are placed to form disks (see Fig. 2.19). Four separate disks can be identified as Muon Endcaps (ME1–ME4) in Fig. 2.17. Each disk is composed of concentric, or nearly concentric rings, that cover the whole azimuthal plane by arranging CSC units one next to another. ME1 includes 216 such units (3 rings, $72 + 72 + 72$), while ME2 to ME4 include 108 each (2 rings, $36 + 72$). Each CSC unit is designed to cover either 10° or 20° in ϕ such that a muon in the range $0.9 < |\eta| < 1.2$ (barrel-endcap overlap range) is detected by both DTs and CSCs and in the range $1.2 < |\eta| < 2.4$ a muon crosses 3 or 4 different CSCs.

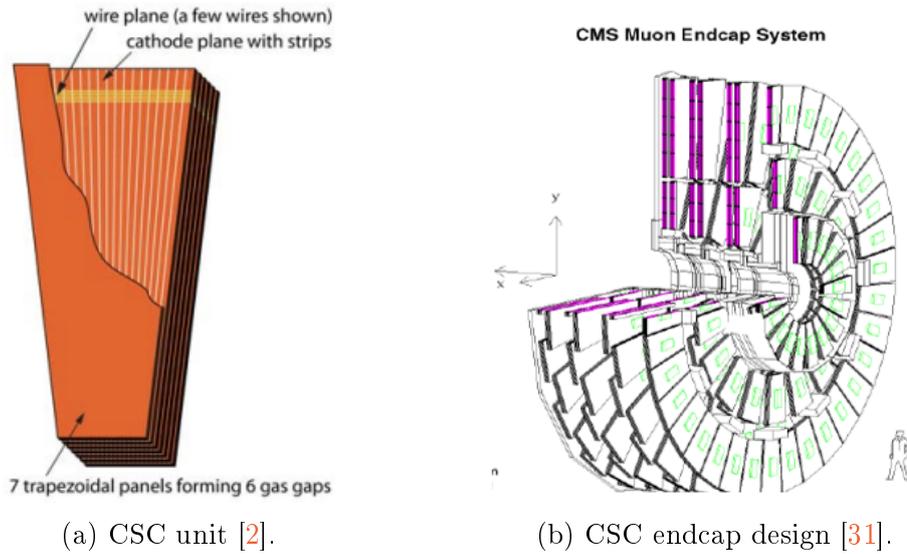


Figure 2.19: Cathode strip chamber specifications.

Figure 2.19a illustrates the layout of the chamber. Seven trapezoidal negatively charged cathode panels are used that each consists of several copper strips that run radially. In between the cathode panels, perpendicular to them, six positively charged anode wire planes are placed that run azimuthally. Additionally, the volume created amid the cathode panels is filled with a mixture of $40\% Ar + 50\% CO_2 + 10\% CF_4$ gas. Likewise, when a muon passes through a chamber it knocks electrons off the gas atoms. The avalanche of electrons will move towards the anode wires, while the ions will move toward the cathode strips. This way two signals can be obtained at a time, the former provides a measurement of the r -coordinate and the latter provides a measurement of the ϕ -coordinate for a given muon instance. To put the size into perspective, the largest cathode strip chambers found in the outer rings of ME2 and ME3 are about $3.4 \times 1.5 m^2$. The CSC architecture provides precise spatial and timing information, allows it to function in a region of a non-uniform magnetic field and of high particle rate, and thus act both as a detector and a trigger by identifying muons and matching their tracks to those in the tracker.

Resistive plate chambers

Both the DTs in the barrel region and the CSCs in the endcaps are complemented by resistive plate chambers. RPCs are classified as trigger devices which are described by very good spatial and time resolution. Particularly on the latter front, RPCs can reach reaction speeds faster than the LHC bunch crossing rate ($25 ns$) down to about $1 ns$. Positioned as shown in Fig. 2.17 with six layers in the barrel and four in the endcaps, they can correctly associate a muon track to the correct bunch crossing even in cases of high pile

up. The redundancy of the two additional layers found in the barrel region, in the first and second muon stations, comes from the fact that in the event of a low- p_T muon this design guarantees the successful track reconstruction based on the 4-layer hit information. In the barrel region (RB1-RB4) rectangular RPCs are arranged in six coaxial cylinders around the beam line, each approximating a dodecagon shape and with their strips running along the beam direction. On the other hand, in the endcaps (RE1-RE4) trapezoidal shaped RPCs are used to form concentric rings with their strips, this time, running radially.

RPCs consist of two parallel plates made of phenolic resin (bakelite) and are separated by a few millimeters gap filled with a non flammable gas mixture of 96.2% $C_2H_2F_4$ + 3.5% iC_4H_{10} + 0.3% SF_6 . The bakelite plates have a resistivity in the range of $10^{10} - 10^{11} \Omega cm$, while they are coated with a conductive graphite paint on their outer surface to form the electrodes. The readout system is then performed through aluminum strips separated from the graphite coating by an insulating PET film. Muons passing through the RPC chambers cause an avalanche of electrons to be collected in both of the resistive plates, making up the total signal as the sum of the two signals of each plate.

Upgrades

In preparation of the the forthcoming HL-LHC era the CMS muon system will receive its own upgrades [32]. New hardware and improved software algorithms are planned for the existing DTs and CSCs that promise higher detection efficiency and increased resolution. Furthermore, RPCs will receive a complete makeover that will use improved technologies with better space resolution in the ϕ -coordinate measurement and a finer time resolution $< 1 ns$ (cosmic tests reveal). Lastly, a new detector component will join the family, that is, Gas Electron Multipliers (GEMs) [33, 34], which will be attached in the endcaps. This type of detector will not only boost the detection and trigger efficiency in the forward region but will also play a complementary role in the forward experiments placed in front of CMS. Already, as of LS2, one out of the three GEMs has already been installed and tested offline during Run III.

2.4 Trigger

We can perhaps take a moment and appreciate Fig. 2.4. At this point particles that have emerged from the pp collision come into contact with the CMS detector with most of them depositing their energy or at least leaving a trace of their existence in one or more of CMS's subsystems. One way or the other this information is extracted so that useful conclusions can be drawn on the importance of the collision event. Lets formulate our next problem. LHC proton-proton collisions occur at a bunch crossing rate of 40 MHz (every 25 ns), meaning that every second about 40 million collisions are delivered. For it all to be stored it would require a tremendous amount of disk space given that each event consists of a couple of megabytes worth of data (or some PB/sec of collisions). Moreover, even if storing all information was viable, a rich amount of resources would need to be disposed of for the offline analysis. Fortunately, to our advantage, most of the events have no contribution if some at all in exploring new physics. Consequently, the vast majority of them can be discarded and only the most interesting among them be kept for storage and later analysis. The CMS trigger system is designed to do exactly that, filter out the non-interesting collisions based on pre-implemented criteria that run during the online event selection. This is a two step process based firstly on the Level-1 Trigger (L1T) and lastly on the High Level Trigger (HLT).

Level-1 trigger

Broadly speaking the Level-1 Trigger [35] has but one true goal, that is to reduce the data rate directed to HLT down to $\approx 100\text{ kHz}$. The L1T system is based on raw hardware electronics, placed partially on top of the detector and inside a control room about 90 m away from the experimental cavern, and on the algorithms that have been programmed to execute. Hardware electronics utilize mostly the technologies of Field Programmable Gate Arrays (FPGAs), Application Specific Integrated Circuits (ASICs), and Look Up Tables (LUTs) that provide the ability to re-program and improve the algorithms depending on the LHC's physical program needs.

The L1T accepts segmented information from both the calorimeter and the muon triggers. To this point local objects are used, like electrons, photons or even jets. Some global quantities are estimated, like the sum of the transverse energy E_T or the missing E_T , whose values are compared to various pre-implemented thresholds. The architecture of the system is displayed in Fig. 2.20. Three major processing layers can be distinguished: the calorimeter trigger, the muon trigger, and the global trigger. The flow of information is quite fast, as the L1T must decide for each single bunch crossing whether it should be kept or discarded before the next collision occurs. This means that all processes run by the trigger are repeated every 25 ns , and should work without encountering any errors or malfunctions. Unfortunately, 25 ns are too few for the whole online analysis to be performed since there are hardware limitations that extend the time needed, i.e., electronics response, and necessary time to transfer information from one system to another. A compromise is made to keep up with the collision rate. Information is stored momentarily in a pipelined fashion before it is fed to the L1T system. There, all data related to one bunch crossing are carefully time-tagged by all systems so that they remain synchronized to be later identified and revoked one at a time. The trigger then has $3.2\text{ }\mu\text{s}$ to make a decision on whether the event should be discarded or produce the L1 accept signal; this process includes virtually no dead time.

In more detail, for the calorimeter trigger it all starts with the Trigger Primitive Generators (TPG) which focuses on the information received by the ECAL, HCAL and HF calorimeter towers³. Then the Regional Calorimeter Trigger (RCT) finds candidate electrons, photons, taus and jets. The list of candidates along with the transverse energy sums are then transferred to the Global Calorimeter Trigger (GCT) which calculates the total E_T and MET sums and sorts all particle families. Lastly, the top four particles from each category are forwarded to the global trigger.

For the muon trigger each subsystem (DTs, CSCs and RPCs) has its own unique nature as they follow different configurations. For this reason, the information is initially processed separately for each subsystem, since a variation in their algorithms is implemented. All three systems try their best to reconstruct possible muon tracks with the available information. The estimation of the p_T comes afterwards. Results are ranked based on the quality of their detector signatures and the best of them are sent to the global muon trigger (GMT). In the GMT all muon tracks are converted into the same η , ϕ and p_T scale before combined. Correlations are found between tracks from different subsystems. Muon signs are deduced and all candidates are sorted, again, based on quality criteria, correlation and p_T . The best four candidates are then sent to the global trigger.

Information is mostly received asynchronously by the global trigger. In that respect, both calorimeter and muon inputs that correspond to the same bunch crossing have to be

³Calorimeter towers or calo towers are the result of the $\eta - \phi$ segmentation followed by ECAL and HCAL. These towers are objects that contain several crystals from ECAL all the way up to some layers from HCAL with their boundaries defined according to the layout followed by the latter one. Calo towers are mostly used in MET calculations and as means to provide a fast and minimal jet reconstruction for the trigger decision.

plus jets for top physics, inclusive electron for calorimeter calibration, samples to measure efficiency and study background, apart from the main datasets. The HLT gains access to the whole readout detector signal for each event and performs similar procedures as those done during the offline global event reconstruction. But time and speed is of essence, so this task is optimized by sequentially reconstructing only necessary information on the event, until the HLT reaches to a conclusion on whether to keep it or discard it. This decision takes about 300 ms per event. Rejected events are lost forever so the system must be efficient and accurate on discovering interesting events. Only a very small fraction of the rejected events is kept for the DAQ monitoring and performance checks. Events selected by the HLT are forwarded to the DAQ system which transfers the raw detector information to the Tier 0 site for storage to tape.

Upgrades

Understandably, the CMS trigger upgrade is partially done as individual subdetector components with trigger functionalities are upgraded. Apart from that, the architecture of L1T was changed during LS1 (see Fig. 2.21) [38] and various algorithms have been improved. Specifically, on the muon trigger side a new algorithm was implemented, the muon track finder, which now attempts to initialize the triggering procedure at an earlier stage by combining information from all three muon detectors. This algorithm comes as a replacement for the older track finders run by the DTs and CSCs and for the pattern comparator run by the RPCs, promising higher performance. On the other hand, for the HLT, as mentioned before, upgrades are being deployed on demand as the market evolves and newer, faster technologies are available. Off course, algorithms have been improved and optimized over the years as this is mostly a software based level but late inclusions to the system is the use of GPUs, in addition to the existing CPU farm, and applications that exploit the neural network potentials on increasing algorithm efficiency and allow implementations to deal with more complex tasks. Farther information on the upcoming HL-LHC upgrade can be found in the related technical design report [39].

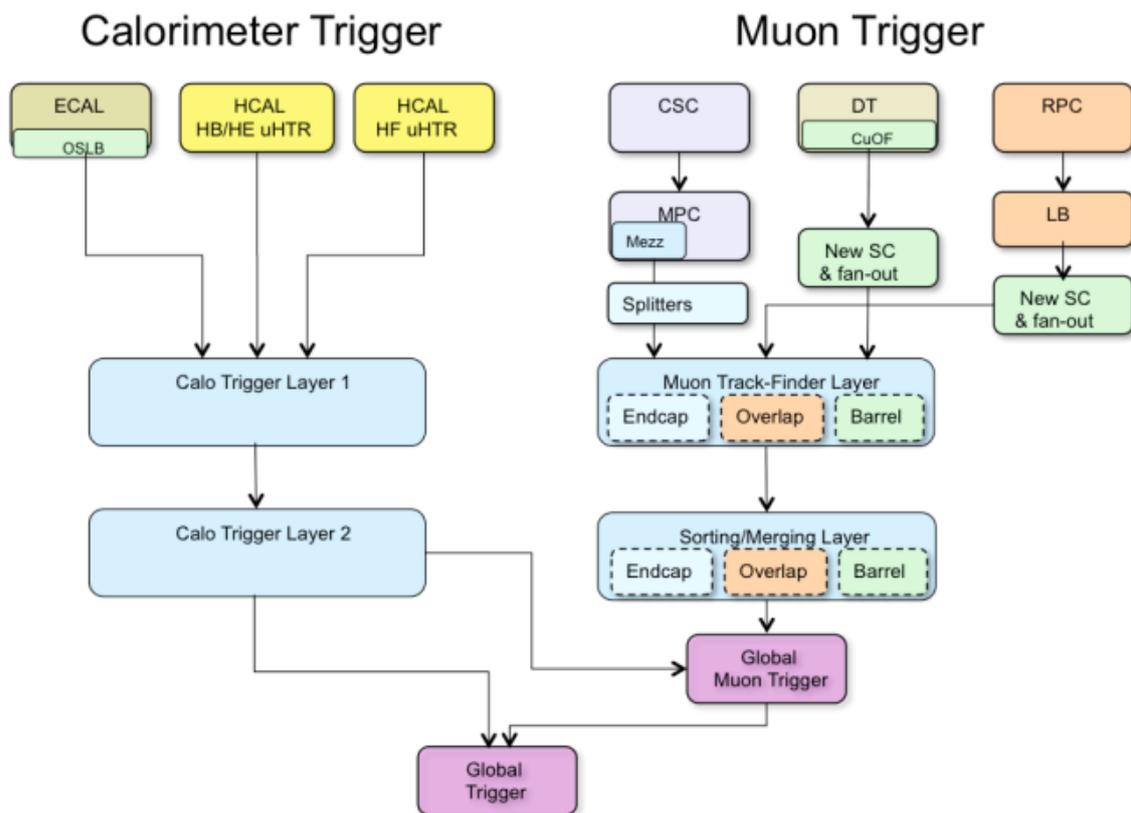


Figure 2.21: The CMS L1T upgrade during LS1 [38].

Bibliography

- [1] Christian W. Fabjan and Herwig Schopper. *Particle Physics Reference Library, volume 2: Detectors for Particles and Radiation*. Springer Cham, 2020. <https://doi.org/10.1007/978-3-030-35318-6>.
- [2] S. Chatrchyan et al. |CMS Collaboration|. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, aug 2008. <https://dx.doi.org/10.1088/1748-0221/3/08/S08004>.
- [3] G. Aad et al. |ATLAS Collaboration|. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. <https://doi.org/10.1016/j.physletb.2012.08.020>.
- [4] S. Chatrchyan et al. |CMS Collaboration|. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [5] Maximilien Brice. CMS images gallery. Available: <https://home.cern/resources/image/experiments/cms-images-gallery>. Accessed: 2024-09-17.
- [6] Izaak Neutelings. CMS coordinate system. Available: <https://tikz.net/axis3d cms/>. Accessed: 2024-09-18.
- [7] Izaak Neutelings. Pseudorapidity. Available: https://tikz.net/axis2d_pseudorapidity/. Accessed: 2024-09-18.
- [8] A. M. Sirunyan et al. |CMS Collaboration|. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017. <https://cds.cern.ch/record/2270046>.
- [9] The CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. <https://cds.cern.ch/record/331056>.
- [10] Domenico Campi and David Stickland. CMS: a super solenoid is ready for business. *CERN Courier*, 47(2):22–25, 2007. <https://cds.cern.ch/record/1734035>.
- [11] Vyacheslav Klyukhin. Design and Description of the CMS Magnetic System Model. *Symmetry*, 13(6), 2021. <https://www.mdpi.com/2073-8994/13/6/1052>.
- [12] V. Karimäki et al. |CMS Collaboration|. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. <https://cds.cern.ch/record/368412>.
- [13] S. Chatrchyan et al. |CMS Collaboration|. Alignment of the CMS tracker with LHC and cosmic ray data. *JINST*, 9:P06009, 2014. <https://cds.cern.ch/record/1667597>.
- [14] The CMS Collaboration. SILICON PIXELS. Available: <https://cms.cern/detector/identifying-tracks/silicon-pixels>. Accessed: 2024-09-25.
- [15] Paolo Azzurri. The CMS Silicon Strip Tracker. *Journal of Physics: Conference Series*, 41(1):127, may 2006. <https://dx.doi.org/10.1088/1742-6596/41/1/011>.

- [16] The CMS Collaboration. SILICON STRIPS. Available: <https://cms.cern/detector/identifying-tracks/silicon-strips>. Accessed: 2024-09-25.
- [17] The CMS Collaboration. TRACKING. Available: <https://cms.cern/detector/identifying-tracks>. Accessed: 2024-09-25.
- [18] A. Rossi et al. [CMS Collaboration]. The CMS Tracker for the High Luminosity LHC. Technical report, CERN, Geneva, 2023. <https://cds.cern.ch/record/2816248>.
- [19] The CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. <https://cds.cern.ch/record/349375>.
- [20] Daniela Bortoletto. Detectors for Particle Physics. Available: https://indico.cern.ch/event/318531/attachments/612850/843143/daniela_l15.pdf. Accessed: 2024-10-03.
- [21] David Cockerill. Introduction to Calorimeters. Available: <https://indico.cern.ch/event/518474/contributions/1198681/attachments/1267581/1877122/Calorimetry-lecture-to-Southampton-students-4May2016-Cockerill-compressed.pdf>. Accessed: 2024-10-03.
- [22] Viatcheslav Grishin. ECAL Endcap Regional Center, CERN/Crystal Palace. Available: <https://cms-ecal-ee-rc.web.cern.ch/index.htm>. Accessed: 2024-10-03.
- [23] The CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. <https://cds.cern.ch/record/357153>.
- [24] P. D. Bunin. Modernization of the CMS Hadron Calorimetry at the LHC. *Phys.Part.Nuclei*, 55:109–115, 2024. <https://doi.org/10.1134/S1063779624010052>.
- [25] G. Gaudio and M. Livan. The Art of Calorimetry, Lecture II. Available: https://static.sif.it/SIF/resources/public/files/va2009/gaudio_0724-2.pdf. Accessed: 2024-10-10.
- [26] The CMS Collaboration. The Phase-2 Upgrade of the CMS Endcap Calorimeter. Technical report, CERN, Geneva, 2017. <https://cds.cern.ch/record/2293646>.
- [27] Dave Barney. A NEW ERA IN CALORIMETRY. Available: <https://cms.cern/news/new-era-calorimetry>. Accessed: 2024-10-15.
- [28] J. G. Layter et al. [The CMS Collaboration]. *The CMS muon project: Technical Design Report*. Technical design report. CMS. CERN, Geneva, 1997. <https://cds.cern.ch/record/343814>.
- [29] A.M. Sirunyan et al. [The CMS Collaboration]. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}=13$ TeV. *Journal of Instrumentation*, 13(06):P06015, jun 2018. <https://dx.doi.org/10.1088/1748-0221/13/06/P06015>.
- [30] S. Chatrchyan et al. [The CMS Collaboration]. Performance of the CMS Drift Tube Chambers with Cosmic Rays. Performance of the CMS Drift Tube Chambers with Cosmic Rays. *JINST*, 5:T03015, 2010. <https://cds.cern.ch/record/1223944>.

- [31] N.P. Ghanathe, A. Madorsky, H. Lam, D.E. Acosta, A.D. George, M.R. Carver, Y. Xia, A. Jyothishwara, and M. Hansen. Software and firmware co-development using high-level synthesis. *Journal of Instrumentation*, 12(01):C01083, jan 2017. <https://dx.doi.org/10.1088/1748-0221/12/01/C01083>.
- [32] Antonello Pellecchia on behalf of the CMS Muon group. The upgrade of the CMS muon system for the High Luminosity LHC, 2023. <https://arxiv.org/abs/2312.05939>.
- [33] A Colaleo, A Safonov, A Sharma, and M Tytgat. *CMS Technical Design Report for the Muon Endcap GEM Upgrade*. CERN, 2015. <https://cds.cern.ch/record/2021453>.
- [34] The CMS Collaboration. GEMS (GAS ELECTRON MULTIPLIERS). Available: <https://cms.cern/detector/detecting-muons/gas-electron-multiplier>. Accessed: 2024-10-22.
- [35] G. L. Bayatyan et al. [The CMS Collaboration]. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. CMS. CERN, 2002. <https://cds.cern.ch/record/706847>.
- [36] S. Cittolin et al. [The CMS Collaboration]. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical design report. CMS. CERN, Geneva, 2002. <https://cds.cern.ch/record/578006>.
- [37] The CMS Collaboration. The CMS high level trigger. *The European Physical Journal C*, 46:605–667, jun 2006. <https://doi.org/10.1140/epjc/s2006-02495-8>.
- [38] A. Tapper et al. [The CMS Collaboration]. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. CERN, 2013. <https://cds.cern.ch/record/1556311>.
- [39] A. Zabi et al. [The CMS Collaboration]. *The Phase-2 Upgrade of the CMS Level-1 Trigger*. CERN, 2020. <https://inspirehep.net/literature/1819968>.

Chapter 3

THE STANDARD MODEL

3.1 Introduction

Most often than not, theory precedes the experiment, especially in cases where experimental costs are in the range of billions of euros/dollars, as was the case for the LHC construction. This chapter provides a brief insight in the Standard Model (SM) of particle physics, that is, the model under which all LHC experiments try to validate or refute. To date, none of the experiments has delivered but a single clue implying a deviation from the SM predictions, on the contrary, all experimental results seem to be in total agreement with it.

Yet again we strive to seek an extension of our current theory so as to incorporate unresolved phenomena within it. A trend appears from back in the days that seems to explain our undying need for a consistent and uniform theory to explain “everything”. It is observed that in physics, the general evolution of the field is usually expressed by a generalization of laws that explain various phenomena in a single equation. An example of such a case is Einstein’s work on general relativity which if taken at the weak field limit degenerates to the Newtonian interpretation of gravity. The SM is just another example of this “trend”, a model that excels in describing interactions between subatomic particles, both in the small scale and low mass limit, by incorporating three out of the four fundamental forces into an elegant formalism. And thus we begin.

3.2 Elementary particles and forces

To our knowledge four fundamental forces are currently distinguished. Fundamental in the way that they cannot be reduced to other, more basic interactions. These are the *electromagnetic*, *weak*, and *strong* interactions, and *gravity*. The weak and strong interactions lie in the subatomic scale, inside the atoms, with the former being responsible for the radioactive decay of atoms, and the latter, being responsible for the quark confinement into stable or nearly stable particles. The remaining two forces, electromagnetism and gravity, exhibit properties even in larger scales, whose effects can be seen directly in everyday life [1]. Particles that carry an electric charge exchange or interact via the electromagnetic force while all things in the microcosm or macrocosm that have a mass interact via the gravity.

The SM takes shape as the theory that succeeded in uniting three out of the four fundamental interactions, the electromagnetic, the weak, and the strong force. In addition to that, it categorizes all known elementary particles that make use of one or more of the aforementioned forces. Note that by elementary we refer to particles that do not contain

any constituents, in other words, these are understood as the basic building blocks of all matter in the universe. Figure 3.1 provides a great illustration of all particle families contained within the SM context along with some of their basic properties.

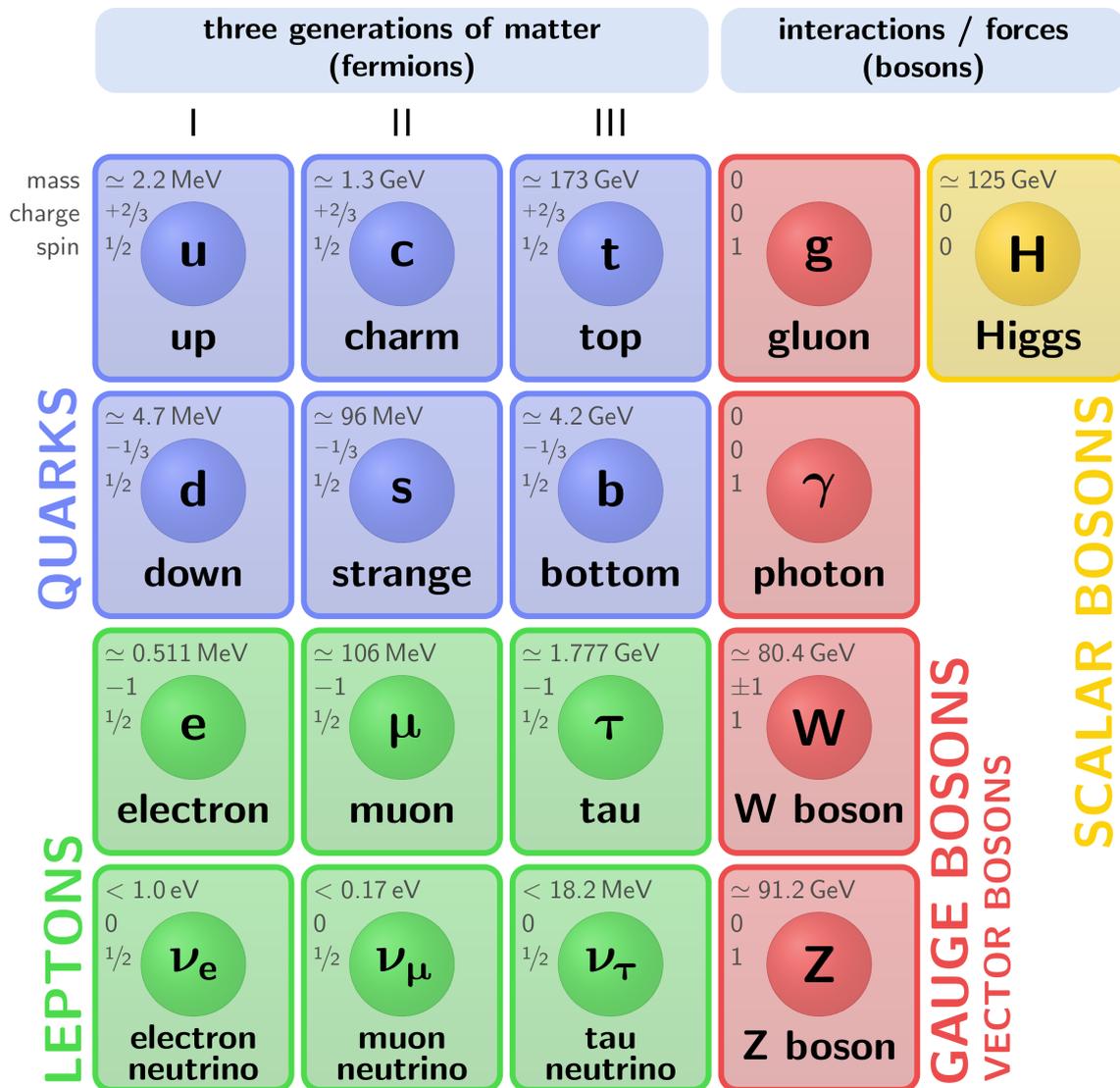


Figure 3.1: The elementary particles of the Standard Model [2].

Let us pause for a moment on this figure. Four individual particle categories can be distinguished, *quarks*, *leptons*, *vector bosons*, and *scalar bosons*. If all particles were to be counted, 61 different elementary particles would be identified:

- There are six **quark** flavors: up (*u*), down (*d*), strange (*s*), charm (*c*), bottom (*b*), and top (*t*). Quarks pairs can be further grouped into three generations of matter with each generation exhibiting an increasing mass. Quarks that fall in the same generation will carry an electric charge of $+2/3$ and $-1/3$ respectively. In addition, each quark type can be found in three different colors, an intrinsic quantum number related to the quark nature, increasing the effective number of quarks to 18. Lastly, for each distinct quark there is its antimatter counterpart¹, an antiquark, in the end amounting to 36.

¹Antimatter particles appear to be exact same copies of their matter selves only having a reverse sign in the charge, parity, and time properties.

- There are six **leptons**: electron (e), electron neutrino (ν_e), muon (μ), muon neutrino (ν_μ), tau (τ), and tau neutrino (ν_τ). Again, three generations are formed by pairing each given lepton with its corresponding neutrino. Each one of the three main leptons (e , μ , τ) appears with considerably larger mass, the higher the generation, while their neutrino pairs all appear massless within the SM formalism (in reality they have a very small mass, currently not described by the SM). Furthermore, the main leptons carry an electric charge of -1 , whilst neutrinos are neutral. If their antimatter partners are taken into account we are left with 12 leptons.
- **Vector bosons** or gauge bosons are the force carriers or mediators and there are different type of mediators depending on the interaction. The strong force is mediated by the gluon (g) which has no electric charge and is massless. Moreover, it has the same color property as the quarks, only this one can exist in 8 different color states. For the weak interaction there are 3 mediators, the negatively and positively charged W^\pm bosons, and the neutral Z^0 boson, while they do present quite a measurable mass at $\simeq 80 \text{ GeV}^2$ and $\simeq 91 \text{ GeV}$ respectively. What remains is the massless and neutral photon (γ) which is the mediator of the electromagnetic interaction, the light!
- Lastly, the **scalar boson** or the Higgs boson is responsible for the mass attribution of all particles like the quarks or the W and Z bosons. It was discovered on 2012 by the ATLAS and CMS experiments [3, 4] and its mass was measured at $\simeq 125 \text{ GeV}$.

Taking into account 36 quarks, 12 leptons, $12(8 + 3 + 1)$ mediators, and the higgs boson we end up with 61 elementary particles. One can find an alternative, more elaborate visualization of Fig. 3.1 in Appendix A.

The SM tries to dictate or predict how these particles interact with one another, how they behave when observed as free particles, if at all allowed, or even how they form more complex particles. A handful of wonderful phenomena and peculiarities lie within the model, and interesting questions can be raised. From the lepton family the electron seems stable, having an infinite lifetime, while its heavier cousins, μ and τ , have a finite lifetime, estimated around at $2.2 \times 10^{-6} \text{ s}$ and $2.9 \times 10^{-13} \text{ s}$ respectively before they decay. The electromagnetic interaction has an infinite range due to the massless nature of the photon. On the other hand, the weak force has a limited scope as it's mediators are heavy. The strong force belongs to the bizarre category since its effective range is even shorter compared to the weak's one even though it possesses a massless mediator; this phenomenon has been understood and it is caused due to the asymptotic freedom of quarks, a phenomenon which will be later discussed in detail in Section 4.2.3. Table 3.1 compiles some interesting information on the fundamental forces under the SM formalism, like their mediators, their effective range and the particles that can interact with.

In practice, each force is directly linked to its own physical theory. Their effects have been bothering many civilizations since ancient times. Physicists have been consistently studying them over the past century, and for a long time they were treated as independent phenomena. Before the 20th century, electricity and magnetism were considered as two separate forces, until Maxwell unified them on 1873 to what we know as the electromagnetism. It was later in the 1940s, when Tomonaga, Feynman, and Schwinger revolutionized

²Note that in High Energy Physics it is usual to work in natural units, here $c = 1$ is assumed i.e., the constant for the speed of light is omitted. Under this notation energy and mass are equal, instead of the standard conversion of $E = mc^2$, thus simplifying equations on that context. Likewise, $\hbar = 1$ is assumed in various places, so do not be prompt to search for it. Physical units can be re-obtained afterwards.

Table 3.1: Fundamental interactions within the SM along with their respective mediators, effective range, field strength, and particles that interact with.

Interaction	Mediator	Range (m)	Strength	Particles
Electromagnetic	γ	∞	10^{-2}	Electr. charged particles
Weak	W^\pm & Z^0	10^{-15}	10^{-13}	Quarks (q) & Leptons (l)
Strong	g	10^{-18}	10	Quarks (q) & Gluons (g)

the field, when they incorporated the concepts of quantum mechanics to the existing theory to introduce the Quantum Electrodynamics (QED) [5]. Similar contributions have led the other fields to evolve and mature, so as to reach a state with more refined theory schemes. Quantum Chromodynamics (QCD) is the theory that describes the strong force, while Flavordynamics is behind the weak interaction. It was not too long before the weak and the electromagnetic force were treated as two aspects of the same, unified interaction, the electroweak interaction. These theories share a lot of similarities, in fact, most of them were inspired by QED. Many people started trying grant unification schemes to incorporate the strong and electroweak interactions into a single unified theory, this concept laid the foundations for the Standard Model [6].

The SM is a gauge quantum field theory (QFT). Put it plainly, the fundamental objects in this realm are the quantum fields, with particles being treated as excited states of those underlying fields [7]. In this formalism fields are categorized based on the value of one of their inner attributes, the spin. *Fermions* have a half-odd-integer spin value and make up the observable matter particles in our universe. Both quarks and leptons are fermions with spin $1/2$. Fields with an integer valued spin are called *bosons* and play the role of the force carriers. So you can imagine that all the mediators are bosons. Indeed, these have a spin value of 1 and specifically belong to the vector boson category, in contrast to the higgs field which has a spin value of 0 and belongs to the scalar boson one. But do not yet tempt yourselves to know about their differences.

3.3 Symmetries and Lagrangian formalism

As it stands, symmetries seem to play an important role in physics. Interesting conclusions can be drawn about a problem even with limited available information just by purely studying its underlying symmetries. They are usually reflected in the equations of motion rather than in particular solutions to those equations [5]. Notably, Emmy Noether's theorem [8], published on 1918, establishes a direct relation between symmetries and conservation laws. In other words, each of the system's symmetries imply the conservation of a given physical quantity and vice versa. Table 3.2 lists some of the most basic symmetries with their associate conservative law.

Table 3.2: Symmetries and conservation laws [5].

Symmetry	Conservation law
Translation in time	Energy
Translation in space	Momentum
Rotation	Angular momentum
Gauge transformation	Charge

Symmetry is a process, or transformation, which when applied to a system it leaves it indistinguishable from the original one. In that respect it is said that the system is *invariant* under this transformation. A mathematical *group* is formed by gathering all symmetry operations for a given system, and conforms to four basic properties [9]:

- *Associativity* – combinational groupings can be interchanged:
If $g_1, g_2, g_3 \in G$ then $g_1 \cdot (g_2 \cdot g_3) = (g_1 \cdot g_2) \cdot g_3$.
- *Closure* – combinations remain in the set:
If $g_1, g_2 \in G$ then $g_1 \cdot g_2 \in G$.
- *Identity* – one element does nothing:
 $\forall g_i \in G \exists I \in G : g_i \cdot I = I \cdot g_i = g_i$.
- *Inverse* – combinations can be undone:
 $\forall g_i \in G \exists g_i^{-1} \in G : g_i \cdot g_i^{-1} = g_i^{-1} \cdot g_i = I$.

One could perhaps say that *group theory* is the systematic study of symmetries. Group elements may or may not commute. In the former case the group is called *Abelian*. There are various group categories. To name a few, they could be listed as *finite* or *infinite* based on the number of elements inside the group or even as *continuous* or *discrete* if the elements depend on one or more continuous variables, or if there is no limit point in it respectively. As it turns out, in High Energy Physics, or general in physics, matrices are a great representation for the symmetry groups. The most commonly used group of matrices is the *Unitary* group $U(n)$, i.e., all $n \times n$ matrices whose inverse is equal to its transpose conjugate $U^{-1} = U^\dagger$. The *Special* subgroup $SU(n)$ is also of interest, that expresses unitary matrices that have a determinant of 1.

Symmetries are greatly displayed in the Lagrangian formalism, so at this point, an introduction is made to it, leaving the rest of the chapter to shine its glory. The Lagrangian was first introduced in classical mechanics and is based on the *stationary-action* or *least-action* principle. It is an energy function built to describe the dynamics of a system and equations of motion can be derived for classical, discrete particles. But let us skip the classical mechanics part and instead dive right into the field realm. In the same way, the Lagrangian, *Lagrangian density* \mathcal{L} this time, can perform calculations with fields which are given as a function of both position and time. Assuming a particle that is described by the field $\phi(\vec{x}, t)$ the Lagrangian density in its general form can be written as:

$$\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi) \quad (3.1)$$

where ∂_μ (∂^μ) is the covariant (contravariant) spacetime partial derivative

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = \left(\frac{1}{c} \frac{\partial}{\partial t}, \vec{\nabla} \right), \quad \partial^\mu \equiv \frac{\partial}{\partial x_\mu} = \left(\frac{1}{c} \frac{\partial}{\partial t}, -\vec{\nabla} \right) \quad (3.2)$$

Note that x^μ is the position four-vector as $x^\mu = (ct, x_0, x_1, x_2) = (ct, x, y, z)$ and $x_\mu = (ct, -x, -y, -z)$. The action is then defined as:

$$S = \int \mathcal{L}(\phi, \partial_\mu \phi) d^4x \quad (3.3)$$

Under the least-action principle the action remains unchanged for small variations $\delta S = 0$. The aftermath are the Euler-Lagrange equations of motion for the ϕ field:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \quad (3.4)$$

Three basic field cases are mentioned, that come to mind if we recall the discussion on the SM, which hopefully will provide the framework needed to continue to the next section.

Spin 0 field

Assuming a real scalar, spin 0 field ϕ that describes a neutral non-interacting particle with mass m , the Lagrangian can be written as:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}m^2\phi^2 \quad (3.5)$$

Substituting \mathcal{L} in Eq. 3.4 leads to the *Klein-Gordon* equation:

$$\partial_\mu\partial^\mu\phi + m^2\phi = 0 \quad (3.6)$$

This equation is nothing more than the differential wave equivalent to the plain relativistic energy-momentum relation $E^2 = (pc)^2 + (mc^2)^2$.

Spin 1/2 field

This time a complex³ spinor, spin 1/2 field ψ is assumed that describes ordinary matter particles with mass m . The Lagrangian is given by:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi \quad (3.7)$$

where $\bar{\psi} = \psi^\dagger\gamma^0$ is the adjoint spinor of ψ , γ^μ ($\mu = 0, 1, 2, 3$) are the Dirac 4×4 matrices written with the aid of the 2×2 Pauli matrices σ_i ($i = 1, 2, 3$). The Dirac and Pauli matrices are given below:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \gamma^i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix} \quad (3.8)$$

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3.9)$$

By substitution in Eq. 3.4 the Dirac equation is obtained that describes all fermions like the quarks and the electron.

$$(i\gamma^\mu\partial_\mu - m)\psi = 0 \quad (3.10)$$

Moreover, it implies the existence of antiparticles as the solutions to this equation return one positive and one negative result for the energy. The negative solution, while not quite understood at first, was later attributed to the existence of antiparticles; fact that was verified with the discovery of the positron, the antimatter counterpart of the electron.

Spin 1 field

A final case remains, that of a massless vector, spin 1 field A_μ . The Lagrangian is formed as:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - j^\mu A_\mu \quad (3.11)$$

where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ the electromagnetic field strength tensor and $j^\mu = (\rho, \vec{j})$ ($c = 1$) the current density four-vector that plays the source of the electromagnetic field. The Euler-Lagrange equation (Eq. 3.4) outputs the following result:

$$\partial_\mu F^{\mu\nu} = j^\nu \quad (3.12)$$

³Complex in the sense that the spinor field can be written in the form $\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ where $\psi_1, \psi_2 \in \mathbb{C}$.

where if the dual antisymmetric tensor $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$ is defined, we also get:

$$\partial_\mu \tilde{F}^{\mu\nu} = 0 \quad (3.13)$$

both of which incorporate all four of Maxwell's equations, Gauss and Ampere law, and no monopoles and Faraday law respectively, from classical electromagnetism into two elegant expressions.

But so far there is no sign of symmetries, and no, you have not been tricked. The SM is based on three individual symmetries bringing together the three fundamental forces discussed previously. In the following sections each symmetry will be studied separately as we slowly build the SM to its final form. Similarly, but in a more detail to what has been done just above the SM Lagrangian will be constructed inside the quantum field theory framework so as to provide a sophisticated theory that will describe all known elementary particles, their properties, and their interactions.

3.4 Quantum Electrodynamics and the $U(1)$ symmetry

In a pursuit of a theory that describes the electromagnetic interactions a choice is made to focus on the Lagrangian density, to start from fundamental level and slowly try to reach our goal constructively. Since the single simplest entity in electromagnetism is the electron, which nonetheless is a particle (spin 1/2), perhaps a good place to start is the Dirac Lagrangian in Eq. 3.7 that describes free fermions.

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (3.14)$$

The time has finally come to choose a symmetry group, apply a transformation inspired from the symmetry group on top of the Lagrangian, and deduce whether it will remain invariant or not. The following transformation is chosen for the field ψ :

$$\psi \rightarrow \psi' = e^{ia}\psi \quad (3.15)$$

where a is an arbitrary real constant across all points in spacetime. The ψ field undergoes a phase transition. Historically this type of transformation falls under the *gauge⁴ symmetry* category. As a matter of fact, because this change happens at the same time in all space this transformation is considered as a *global gauge symmetry*, while the phase element itself (e^{ia}) belongs to the unitary Abelian⁵ group with dimension one $U(1)$. By applying this transformation to Eq. 3.14 the Lagrangian remains unchanged. Note that:

$$\bar{\psi}' = e^{-ia}\bar{\psi} \quad (3.16)$$

$$\partial_\mu\psi' = \partial_\mu(e^{ia}\psi) = e^{ia}\partial_\mu\psi \quad (3.17)$$

And all is well, the Lagrangian remains invariant according to this global transformation under the $U(1)$ symmetry. So far nothing has been gained.

The more general approach is to apply the same transformation but this time having a spacetime dependence in the parameter $a = a(x^\mu) = a(x)$:

$$\psi \rightarrow \psi' = e^{ia(x)}\psi \quad (3.18)$$

⁴Gauge symmetry implies a type of transformation that does not affect the fundamental structure of the field, here the ψ field, in a way that it maintains the same physical attributes, while it incorporates some arbitrariness to the way that it can be selected.

⁵Again, Abelian means that the group has the property to perform multiplication commutatively. Given $U(a) = e^{ia} \Rightarrow U(a_1)U(a_2) = U(a_2)U(a_1)$.

Since now there is a spacetime dependence on the phase that will be applied to ψ it is called a *local transformation*. This phenomenally “simple” change introduces some implications to the matter. Upon replacement in Eq 3.14, an additional term is created as the derivative now acts both in ψ and $a(x)$. Note that:

$$\bar{\psi}' = e^{-ia(x)}\bar{\psi} \quad (3.19)$$

$$\partial_\mu\psi' = \partial_\mu\left(e^{ia(x)}\psi\right) = e^{ia(x)}\partial_\mu\psi + ie^{ia(x)}\psi\partial_\mu a(x) \quad (3.20)$$

the additional term in Eq 3.20, that contains the derivative of the parameter a , $\partial_\mu a(x)$, breaks the invariance of \mathcal{L} . But not all hope is lost, as there are various mathematical tricks up our sleeves. Invariance can be recovered, rather we are obliged to recover it, that is, if we would like to derive a theory that is both mathematically and physically consistent.

Ideally, an alternative formalism for the derivative is desired, call it D_μ , so that it transforms covariantly and in the end no extra terms remain:

$$D_\mu\psi' = e^{ia(x)}D_\mu\psi \quad (3.21)$$

The *covariant derivative* D_μ is introduced with the following form:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu \quad (3.22)$$

where the above definition incorporates a new field A_μ in the theory and dictates how it transforms:

$$A_\mu \rightarrow A'_\mu = A_\mu - \frac{1}{q}\partial_\mu a(x) \quad (3.23)$$

The A_μ field is called *gauge field* and remember that it was introduced by our theory as we demanded invariance under the local transformation. The process which a symmetry is generalized from global to a local one is some times referred to as *gauging*. Once again, making use of Eq. 3.22 in Eq. 3.14 we obtain our "new" Lagrangian:

$$\begin{aligned} \mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu\psi A_\mu \end{aligned} \quad (3.24)$$

By trying now out the transformation in Eq. 3.18 in the above Lagrangian we see that the invariance has been recovered:

$$\begin{aligned} \mathcal{L} \rightarrow \mathcal{L}' &= i\bar{\psi}'\gamma^\mu\partial_\mu\psi' - m\bar{\psi}'\psi' - q\bar{\psi}'\gamma^\mu\psi'A'_\mu \\ &= i\left(e^{-ia(x)}\bar{\psi}\right)\gamma^\mu\partial_\mu\left(e^{ia(x)}\psi\right) - m\left(e^{-ia(x)}\bar{\psi}\right)\left(e^{ia(x)}\psi\right) \\ &\quad - q\left(e^{-ia(x)}\bar{\psi}\right)\gamma^\mu\left(e^{ia(x)}\psi\right)\left[A_\mu - \frac{1}{q}\partial_\mu a(x)\right] \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu\psi A_\mu \\ &= \mathcal{L} \end{aligned} \quad (3.25)$$

It is time to discuss about the last term in Eq. 3.25 in more detail, and no, I did not neglect to comment about the q parameter that was introduced with it. In this last term both the new vector field A_μ and the ψ field are observed. This is a coupling term, expressing the interaction of the the A_μ field with the electron field (the electron field, since on this assumption we initiated our syllogism). It turns out that A_μ is the electromagnetic potential and q is originally introduced as a multiplicative factor to measure the strength of this interaction. Later, it will be re-branded as the *coupling constant* for the electromagnetic interaction and will be identified to be equal to the electric charge ($-e$). Equation 3.24 can be rewritten as:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - j^\mu A_\mu \quad (3.26)$$

where $j^\mu = q\bar{\psi}\gamma^\mu\psi$ is the current density or the *conserved current* that is foreseen by Noether's theorem and acts as a source to the gauge field A_μ .

Aiming to approach all aspects of electromagnetism, for A_μ to be regarded as the photon field another term needs to be included in the Lagrangian that will account for its kinetic energy. The electromagnetic field strength tensor $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ comes to mind and since on its own it is invariant, adding such a term will preserve the Lagrangian invariance. For the last time, Eq. 3.26 is written as:

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - j^\mu A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (3.27)$$

The importance of Eq. 3.11 is emphasized here as it is realized in the last two terms above. Another interesting consequence of this Lagrangian is that no mass term of the form $\frac{1}{2}A_\mu A^\mu$ is permitted since it would break the invariance. Off course the photon is observed massless in nature so this works in favor of our theory. Equation 3.27 is then recognized as the Lagrangian of Quantum Electrodynamics.

3.5 Quantum Chromodynamics and the $SU(3)$ symmetry

In fact, the concepts shown in the previous section are part of the *Lie Algebra* which is a mathematical framework devised to study continuous symmetries, usually the later referred to as *Lie groups*. The elements of the $U(1)$ symmetry depend on the continuous parameter $a(x)$ hence fall under the Lie group category. Any element of any such group given a particular representation can be written in the following general form [10]:

$$D_n(a_i) = e^{ia_i X_i} \quad (3.28)$$

where a_i are the parameters and X_i the generators of the group D_n of dimension n . Lie algebra imposes that the commutator between different generators must be proportional to some linear combination of them:

$$[X_i, X_j] = if_{ijk}X_k \quad (3.29)$$

where f_{ijk} the *structure constants* of the group. For each group category (e.g., $U(n)$, $SU(n)$, $SO(n)$) there is an explicit relation that connects the dimension of the group with the number of generators required for that group. For example the $U(n)$ group has n^2 generators so the $U(1)$ symmetry has only one, the photon. We will shortly see how this generalizes for other groups.

Having QED to govern electromagnetic interactions the science community started applying the same ideas that led to QED to other symmetry groups, in an effort of finding befitting theories to describe the remaining forces. It would be then straightforward to expand the same syllogism to higher dimension groups. It seems that a special role is played by the $SU(3)$ symmetry group, that is, the special unitary 3×3 matrices, reminding that "special" in this case means that these matrices possess a determinant of 1. Local transformations under the $SU(3)$ symmetry are of the form:

$$\psi \rightarrow \psi' = e^{ia_a(x)T_a}\psi = e^{ia_a(x)\frac{\lambda_a}{2}}\psi \quad (3.30)$$

where $a_a(x)$ are the spacetime dependent continuous group parameters and T_a are the 3×3 matrix generators of the group. For $SU(n)$ groups the number of generators is given as $n^2 - 1$, eight for our case, the eight gluons, but lets not be hasty. This means that the indice a found in the exponent runs from one to eight, $a = 1, 2, \dots, 8$, while summation

is implied over all eight components by the indice repetition. The conventional choice for the T_a matrices are the *Gell-Mann* $\lambda_a/2$ matrices [11], given below:

$$\begin{aligned}\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\ \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \\ \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}\end{aligned}\quad (3.31)$$

which are the generalization of the Pauli matrices (Eq. 3.9). Gell-Mann matrices commute according to Eq. 3.29 as:

$$\left[\frac{\lambda_a}{2}, \frac{\lambda_b}{2} \right] = if_{abc} \frac{\lambda_c}{2} \quad (3.32)$$

where the non-zero structure constants are:

$$\begin{aligned}f_{123} &= 1, f_{458} = f_{678} = \frac{\sqrt{3}}{2}, \\ f_{147} &= f_{516} = f_{246} = f_{257} = f_{345} = f_{637} = \frac{1}{2}\end{aligned}\quad (3.33)$$

Since there are non-zero commutations, i.e., not all generators commute, the $SU(3)$ group is categorized as non-Abelian, the consequences of which will become clear later in the section.

The starting point is, again, the Dirac free Lagrangian (Eq. 3.14). Invariance is demanded under the local transformation of Eq. 3.30 only this time we consider infinitesimal phase transformations as to:

$$\psi \rightarrow \psi' = \left[1 + ia_a(x) \frac{\lambda_a}{2} \right] \psi \quad (3.34)$$

$$\partial_\mu \psi \rightarrow \partial_\mu \psi' = \left[1 + ia_a(x) \frac{\lambda_a}{2} \right] \partial_\mu \psi + i \frac{\lambda_a}{2} \psi \partial_\mu a_a(x) \quad (3.35)$$

For the sake of efficiency the degenerate case of global transformation is skipped as proving that the Lagrangian remains invariant is quite trivial. Same as before, the last term in Eq. 3.35 breaks the invariance. In accordance with the QED treatment the covariant derivative D_μ is defined:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_3 \frac{\lambda_a}{2} G_\mu^a \quad (3.36)$$

where g_3 the coupling constant of the $SU(3)$ gauge theory. This time eight gauge fields G_μ^a are incorporated in the theory from the covariant derivative definition and are required to transform as (considering infinitesimal transformations $a_a(x) \ll 1$):

$$G_\mu^a \rightarrow G'^a = G_\mu^a - \frac{1}{g_3} \partial_\mu a_a(x) - f_{abc} a_b(x) G_\mu^c \quad (3.37)$$

By replacing ∂_μ with D_μ from Eq. 3.36 to the Dirac free Lagrangian we obtain:

$$\begin{aligned}\mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - g_3 \left(\bar{\psi}\gamma^\mu \frac{\lambda_a}{2} \psi \right) G_\mu^a\end{aligned}\quad (3.38)$$

The final Lagrangian is obtained with the inclusion of an additional term that will account for the kinetic energy of the G_μ^a fields, similar to what was done in the QED case:

$$\mathcal{L}_{QCD} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - g_3\left(\bar{\psi}\gamma^\mu\frac{\lambda_a}{2}\psi\right)G_\mu^a - \frac{1}{4}G^{\mu\nu a}G_{\mu\nu}^a \quad (3.39)$$

where in order to preserve the invariance the strength field tensor $G_{\mu\nu}^a$ is constructed as:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - g_3 f_{abc} G_\mu^b G_\nu^c \quad (3.40)$$

Quarks are found in three colors in nature (red, blue, green) but there antimatter particles can exist in their anti-color states as well (anti-red, anti-blue, anti-green). So the way the strong interaction works is that it always preserves the color. This is why gluons carry pairs of colors so that upon interacting with a quark can “negate” their original color and attribute to it a new one. For example, a red quark will interact with an anti-red, blue gluon to change to a blue quark state. Counting all possible color states they amount to nine, meaning that to transition from one state to the other it is required to have eight distinct gluons. In that respect, gluons carry color and color can be conserved even in the case that gluons interact with other gluons. This fact is reflected in the Lagrangian of Eq. 3.39 if it is rewritten by expanding the field tensor $G_{\mu\nu}^a$ [11].

$$\mathcal{L}_{QCD} = \text{“}\bar{\psi}\psi\text{”} + \text{“}G^2\text{”} + g_3\text{“}\bar{\psi}\psi G\text{”} + g_3\text{“}G^3\text{”} + g_3^2\text{“}G^4\text{”} \quad (3.41)$$

The first three terms mirror the situation seen in QED, expressing the propagation of quarks, gluons, and the quark-gluon interaction, while the absence of a mass term for the gauge bosons is still required by the invariance of \mathcal{L} . On the other hand, the last two terms are observed to contain the third and the fourth power of G meaning that gluons can form self-interacting states of three and four gluons, states that are called *gluon balls*. This is a direct consequence of the non-Abelian nature of the $SU(3)$ symmetry. There is no analogue in QED, being an Abelian gauge theory it is forbidden. The extends of Eq. 3.39 provides the final Lagrangian for Quantum Chromodynamics \mathcal{L}_{QCD} , the theory that describes the strong force.

This contradiction can be reinforced if we apply the Euler-Lagrange equations (Eq. 3.4) once in the kinetic term for the $F_{\mu\nu}$ field tensor from Eq. 3.27 and once in the kinetic term for the $G_{\mu\nu}^a$ field tensor from Eq. 3.39. In the former case, it is shown again that the Maxwell equations are obtained:

$$\partial_\nu \tilde{F}^{\mu\nu} = 0 \quad (3.42)$$

while in the latter case, the result is:

$$\partial_\nu G^{\mu\nu a} = g_3 f_{abc} G_\nu^b G^{\mu\nu c} \quad (3.43)$$

By comparing these two equations we can see the difference between an Abelian and a non-Abelian theory. In the QCD case (Eq. 3.43) the gluon field G_μ^a can self-interact and can act as a source to itself; right hand side of the equation, as opposed to the photon field which has no such property.

3.6 Flavordynamics and the $SU(2)$ symmetry

Let us first address the name of *Flavordynamics*. From the past, the suffix “dynamics” was connected to be used as a naming convention to theories that describe some kind of interaction e.g., Quantum Electrodynamics, Quantum Chromodynamics, Thermodynamics, and so on. Over the years, as scientists tried to construct a theory that describes the weak

interactions they needed a name to collectively refer to their findings. Weak interactions are related to the flavor⁶ change of particles so the name Flavordynamics was eventually adopted.

Similar to what was done for QED, and especially for QCD, the $SU(2)$ symmetry is related to weak interactions. The mathematical treatment shown in the previous section for $SU(3)$ falls under a more general category, that is the treatment of non-Abelian $SU(n)$ gauge theories referred to as *Yang-Mills* theories. Actually Chen Ning Yang and Robert Mills first established their framework based on $SU(2)$ on 1954 [12] which was later generalized to all $SU(n)$ groups. In this section, it will be briefly shown how the $SU(2)$ symmetry is directly related to weak interactions and how it fails to describe some key aspects of it, bearing the need for an additional breakthrough to happen.

Originally the $SU(2)$ symmetry had been used to describe the nucleon in the context of nuclear physics. The neutron n and the proton p would be attributed to the strong isospin representation of $I = 1/2$ for values of $I_3 = +1/2$ and $I_3 = -1/2$ respectively. The neutron could then be written in the form of an $SU(2)$ doublet:

$$N = \begin{pmatrix} p \\ n \end{pmatrix} \quad (3.44)$$

Under this theory protons and neutrons would be treated similarly by the strong nuclear force [5, 13]. But this was of little hope as the theory was based in two particles that have the same mass, not quite the case for n and p . Some thought that their tiny mass difference of 1.29 MeV could be attributed to the electromagnetic symmetry breaking. What's more the theory required the existence of an isotriplet of massless vector bosons. The closest candidate at that time was the ρ mesons, which regrettably do come with a sizeable mass of 770 MeV , a fact that could not be neglected. A lot of efforts were made to allow massive gauge bosons inside the theory, none of which seemed to get the job done..

In the context of weak interactions and by exploiting the $SU(2)$ symmetry any fermionic field can be expressed as a function of two components, a *left-handed* field ψ_L , and a *right-handed* field ψ_R , such that $\psi = \psi_R + \psi_L$, where

$$\psi_R = \frac{1 + \gamma^5}{2} \psi, \quad \psi_L = \frac{1 - \gamma^5}{2} \psi \quad (3.45)$$

and $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ the *chirality* or *handedness* operator. Under a chiral theory such as this (the $SU(2)$) left and right handed components behave differently. Specifically, the right-handed components are but trivial $I = 0$ representations:

$$f_R : \begin{matrix} e_R^- & u_R & d_R \\ \mu_R^- & c_R & s_R \\ \tau_R^- & t_R & b_R \end{matrix} \quad (3.46)$$

while left-handed components compose fundamental $I = 1/2$ representations, for values of $I_3 = +1/2$ and $I_3 = -1/2$ respectively as shown below:

$$\begin{aligned} f_{eL} &= \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, & f_{\mu L} &= \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, & f_{\tau L} &= \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \\ f_{qL} &= \begin{pmatrix} u \\ d \end{pmatrix}_L, & f_{cL} &= \begin{pmatrix} c \\ s \end{pmatrix}_L, & f_{bL} &= \begin{pmatrix} t \\ b \end{pmatrix}_L \end{aligned} \quad (3.47)$$

⁶In particle physics, flavor refers to the number of variations that a given particle of a particular category can exist. For example, quarks are found in six different flavors (u, d, s, c, t, b), while leptons are found in three (e, μ, τ).

This notation introduces a new feature in the theory, the three generations of matter, three generations for leptons and three for quarks, exactly how they are divided in pairs above (discussed also in Section 3.2). The three gauge fields of the group ($n^2 - 1$, here $n = 2$) are then denoted as W_μ^i where $i = 1, 2, 3$ and form an isospin triplet for $I = 1$:

$$W_\mu = \begin{pmatrix} W_\mu^1 \\ W_\mu^2 \\ W_\mu^3 \end{pmatrix} \quad (3.48)$$

Electrically charged states can be formed by mixing the first and second components of W_μ :

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \\ W_\mu^0 &= W_\mu^3 \end{aligned} \quad (3.49)$$

But let us take back a step and try to arrive at the same conclusion from the Lagrangian perspective. The following local transformations are assumed which are related to the left-handed spinor component:

$$\psi_L \rightarrow \psi'_L = e^{i\frac{\sigma_i}{2}\theta^i(x)}\psi_L \quad (3.50)$$

where $\theta^i(x)$ are the three parameters and $\frac{\sigma_i}{2}$, the Pauli matrices (from Eq. 3.9), the three generators of the $SU(2)_L$ group (denoted as $SU(2)_L$ instead of $SU(2)$ since we are only concerned with left-handed particles). Working in a non-Abelian gauge theory the Pauli matrices commute according to:

$$\left[\frac{\sigma_i}{2}, \frac{\sigma_j}{2}\right] = i\epsilon_{ijk}\frac{\sigma_k}{2} \quad (3.51)$$

where ϵ_{ijk} is the Levi-Civita symbol, the structure constants of the group. Equation 3.50 is rewritten for infinitesimal transformations as:

$$\psi_L \rightarrow \psi'_L = \left[1 + i\theta^i(x)\frac{\sigma_i}{2}\right]\psi_L \quad (3.52)$$

For the last time, but I do not swear on it, invariance is demanded upon the Dirac free Lagrangian (Eq. 3.7). The covariant derivative is defined in the form:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_2\frac{\sigma_i}{2}W_\mu^i \quad (3.53)$$

where g_2 the coupling constant of the $SU(2)_L$ gauge theory and W_μ^i the three gauge bosons, required to transform as:

$$W_\mu^i \rightarrow W_\mu^{i'} = W_\mu^i - \frac{1}{g_2}\partial_\mu\theta^i(x) - \epsilon_{ijk}\theta^j(x)W_\mu^k \quad (3.54)$$

Replacing Eq. 3.53 in the Dirac free Lagrangian we obtain:

$$\begin{aligned} \mathcal{L} &= i\bar{\psi}_L\gamma^\mu D_\mu\psi_L - m\bar{\psi}_L\psi_L \\ &= i\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L - m\bar{\psi}_L\psi_L - g_2\left(\bar{\psi}_L\gamma^\mu\frac{\sigma_i}{2}\psi_L\right)W_\mu^i \end{aligned} \quad (3.55)$$

Additionally inserting a kinetic term for the W_μ^i gauge fields, the above Lagrangian is reshaped to:

$$\mathcal{L} = i\bar{\psi}_L\gamma^\mu\partial_\mu\psi_L - m\bar{\psi}_L\psi_L - g_2\left(\bar{\psi}_L\gamma^\mu\frac{\sigma_i}{2}\psi_L\right)W_\mu^i - \frac{1}{4}W^{\mu\nu i}W_{\mu\nu}^i \quad (3.56)$$

where

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_2 \epsilon_{ijk} W_\mu^j W_\nu^k \quad (3.57)$$

The same conclusions can be drawn for the $SU(2)_L$ case. The $W_{\mu\nu}^i$ strength field tensor acts as a source to itself, a property enabled due to the non-Abelian nature of the symmetry group. Still, no mass term is allowed for the tensor in the Lagrangian as it would break the invariance. But if we revisit Section 3.2 and Fig. 3.1, it is understood that for the weak interactions the three mediators do carry a measurable mass, a quality which does not yet seem to appear in the theory. Moreover, the Z^0 boson is nowhere to be found. These are some of the reasons that make $SU(2)_L$ to fail in describing weak interactions. Fear not, as in the end, weak interactions emerge through a more subtle and elegant mechanism, which in addition, lays the foundation to incorporate mass terms for both the mediators of the weak force and all the fermions.

3.7 Spontaneous symmetry breaking and the Higgs mechanism

But why are we so fixated on acquiring a theory that is described by an invariant Lagrangian? It was mentioned that such a theory is mathematically and physically consistent, but this can be interpreted as arbitrary and general drivel. The reason why invariance is so important is that it makes a theory renormalizable, i.e., it is a theory that can provide finite predictions which can be later verified through the experiment. In contrast, a non-renormalizable theory can not provide any experimentally meaningful predictions. *Renormalization* is a mathematical fabrication aiming to deal with the infinities that can be encountered at times in Quantum Field Theory calculations and other fields. Without diving too much into this topic, for now, it is just noted that mass terms break the invariance and incorporate infinities that are difficult or not at all able to be handled. Other means must be exploited such that, in the end, the resulting theory can be renormalized.

In favor of avoiding cryptic sentences and meanings to only later reach a grand finale, let us plainly state that there is a very elegant mechanism that allows the introduction of mass terms for the gauge vector bosons into the Lagrangian. It is a well studied property of the gauge symmetry space and is referred to as *Spontaneous Symmetry Breaking (SSB)*. Following the guidelines paved by the literature, three cases are considered here, the breaking of a discrete, a continuous global, and a continuous local symmetry.

3.7.1 Spontaneous breaking of a discrete symmetry

We start with a reformed version of the Klein-Gordon Lagrangian (Eq. 3.5) for a scalar field ϕ :

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\mu^2 \phi^2 - \frac{1}{4}\lambda \phi^4 \quad (3.58)$$

where a choice was made to spawn an additional quartic term (ϕ^4 term). The last two terms can emerge if we imagine some kind of an exponential expansion ($e^{-(\alpha \phi)^2} = 1 - \phi^2 + \phi^4 - \dots$) in terms of ϕ^2 where the constant term has been dropped, as this trivial term leaves the Lagrangian invariant and only the first and second orders in terms of ϕ^2 are kept. Yes, it might be argued that some shenanigans is going on with the signs of these two terms. But this can be justified since we wanted to construct something that resembles a mass term, and the parameters were chosen in a way to simplify matters.

It is observed that the above Lagrangian remains invariant under the symmetry operation $\phi \rightarrow -\phi$, a discrete symmetry that forms the $\mathbb{Z}_2 = \{1, -1\}$ group. Borrowing the

notation from classical mechanics, the above Lagrangian is rewritten so that it contains a kinetic and a potential term $\mathcal{L} = T - V$. The potential is then given by:

$$V = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4 \quad (3.59)$$

where $\lambda > 0$ so that the whole second term remains always positive. Two cases can be explored here, for $\mu^2 > 0$ and for $\mu^2 < 0$, shown in Fig. 3.2a and Fig. 3.2b respectively. The trivial $\mu^2 > 0$ case ($\mu \in \mathbb{R}$) describes a scalar field ϕ with mass μ that has the ability to form a four particle vertex with coupling λ . It is said that the symmetry of the system

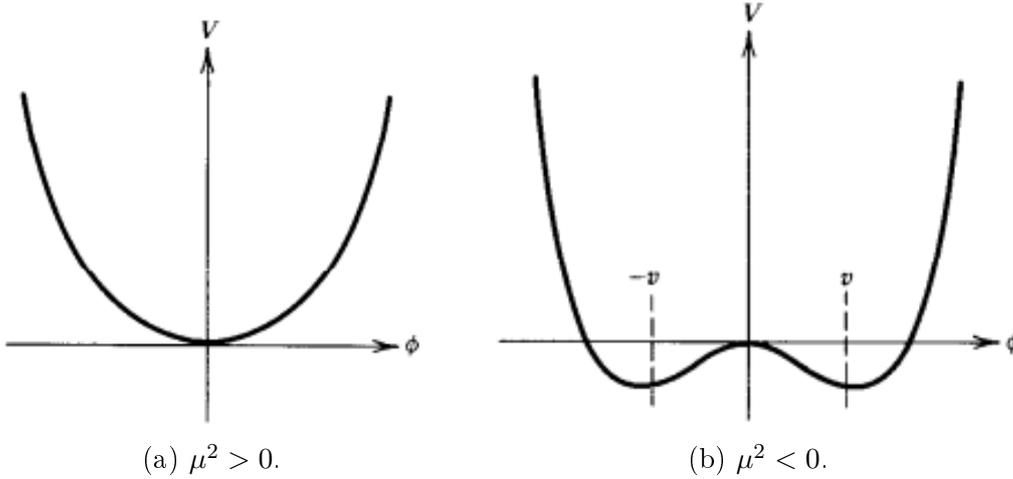


Figure 3.2: The potential $V = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ for $\lambda > 0$ [11].

is reflected upon the *ground state* (the *vacuum*) which corresponds to $\phi = 0$. The $\mu^2 < 0$ case ($\mu \in \mathbb{C}$) is far more interesting. Now the potential has two minima, either of which can express the ground state:

$$\phi = \pm v \quad \text{with} \quad v = \sqrt{\frac{-\mu^2}{\lambda}} \quad (3.60)$$

In QFT it is usual practice to expand the fields around the vacuum value in order to estimate the excited states of the fields i.e., the particles of the theory [13]. When fields are quantized they are treated as operators and the above minima values are usually referred to as *Vacuum Expectation Values*, *VEV* which express the average value of a field in its lowest energy state, the ground state. A non-zero VEV is a strong indication that a symmetry of the theory is spontaneous broken! So the field ϕ is expanded around the VEV:

$$\phi(x) = v + \eta(x) \quad (3.61)$$

where $\eta(x)$ represents a very small quantum (perturbative) fluctuation around the chosen minimum point. Note that here a choice is made, to expand around the $+v$ value. By substitution to the Lagrangian of Eq. 3.58 we obtain:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \lambda v^2\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4 + \text{const.} \quad (3.62)$$

The Lagrangian now describes the field η with the first two terms expressing the kinetic energy and the mass for the field, while higher order terms express the ability of η to interact with itself. Comparing the mass terms between Eq. 3.62 and Eq. 3.58 the mass for the η can be written as:

$$m_\eta = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2} \quad (3.63)$$

At this point the initial $\phi \rightarrow -\phi$ has been lost due to the presence of the cubic term η^3 . The symmetry has been spontaneously broken by our choice to expand the field around the VEV. In other words, the initial symmetry has been concealed by our arbitrary selection of a particular ground state [5]. In return, the theory now describes a new scalar field η with mass m_η . In perturbation theory, this is the correct way to proceed in finding the solutions for the Lagrangian, to avoid solving the equation at the unstable equilibrium point $\phi = 0$ and instead expand the fields around the potential minimum.

3.7.2 Spontaneous breaking of a continuous global symmetry

Still, the goal is to generate a mass term for the currently massless vector fields of the $SU(2)_L$ group. Having this in mind we transpose to a more realistic scenario where the same practices are applied only this time to a continuous global symmetry. The Lagrangian from Eq. 3.58 is rewritten as:

$$\mathcal{L} = (\partial_\mu \phi)^* (\partial^\mu \phi) - \mu^2 \phi^* \phi - \lambda (\phi^* \phi)^2 \quad (3.64)$$

so that it now describes a complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$. Trying out a transformation in the form of Eq. 3.15 (a $U(1)$ continuous local symmetry), \mathcal{L} is invariant. Minimizing the potential this time provides an infinite number of ground states in the shape of a circle, shown also in Fig. 3.3.

$$|\phi|^2 = \phi_1^2 + \phi_2^2 = v^2 \quad \text{with} \quad v^2 = \sqrt{\frac{-\mu^2}{\lambda}} \quad (3.65)$$

Again, any value can be chosen for the ground state. We proceed by choosing $\phi_1 = v$, $\phi_2 = 0$ and expand ϕ as a function of the fields $\eta(x)$, $\xi(x)$.

$$\phi(x) = \frac{1}{\sqrt{2}}[v + \eta(x) + i\xi(x)] \quad (3.66)$$

Substitution to back to \mathcal{L} (Eq. 3.64) gives:

$$\begin{aligned} \mathcal{L} = & \frac{1}{2}(\partial_\mu \xi)(\partial^\mu \xi) + \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) + \mu^2 \eta^2 \\ & + \text{const.} \\ & + \text{cubic and quartic terms in } \eta, \xi \end{aligned} \quad (3.67)$$

The Lagrangian stands to describe the two new fields η , ξ . There is a subtlety here. The η field, as before, has a mass of $m_\eta = \sqrt{-2\mu^2}$, whereas the field ξ is massless. This goes under the *Goldstone theorem* [14] which states that the spontaneous breakage of a continuous symmetry arms the theory with a massless scalar gauge boson, the Goldstone boson. But the existence of a massless scalar would be susceptible to some kind of observation, given that is massless, so it would be within the reach of current experiment setups. Yet, no evidence exist to support this claim within this theory framework. Lets make it local..

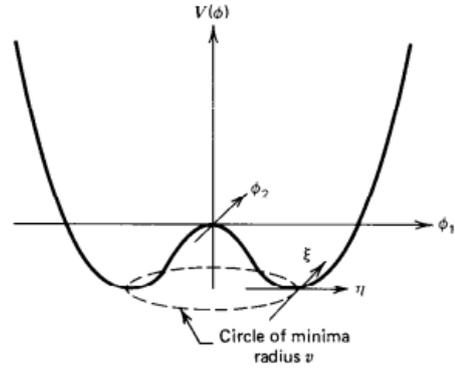


Figure 3.3: The famous Mexican hat. The potential $V = \mu^2 \phi^* \phi + \lambda (\phi^* \phi)^2$ with $\lambda > 0$ and $\mu^2 < 0$ [11].

3.7.3 Spontaneous breaking of a continuous local symmetry

Revisiting the conclusions drawn in Section 3.4, the Lagrangian of Eq. 3.64 is rewritten so as to remain invariant under local $U(1)$ transformations by including the covariant

derivative D_μ (definition given in Eq. 3.22, note also that A_μ transforms as in Eq. 3.23):

$$\mathcal{L} = (\partial^\mu + ieA^\mu)\phi^*(\partial_\mu - ieA_\mu)\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (3.68)$$

where we are still considering the scalar field ϕ and the coupling constant g , found in Eq. 3.22 and Eq. 3.23, has been replaced by $-e$ as promised. The case where $\mu^2 > 0$ is the complete QED analogue (except of the $(\phi^*\phi)^2$ term) for a scalar field compared to what was shown in Section 3.4 for a spinor field ψ . Thus, we dive into the $\mu^2 < 0$ case to exploit the mechanics of spontaneous symmetry breaking.

By substitution of Eq. 3.66 into the above \mathcal{L} we obtain:

$$\begin{aligned} \mathcal{L} = \frac{1}{2}(\partial_\mu\xi)(\partial^\mu\xi) + \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \lambda v^2\eta^2 + \frac{1}{2}e^2v^2A_\mu A^\mu - evA_\mu\partial^\mu\xi \\ - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \text{interaction terms} \end{aligned} \quad (3.69)$$

Just by looking the fourth term in \mathcal{L} we see that we have successfully attributed mass to the vector boson A_μ . Nevertheless, there is still a catch. This will become obvious if the number degrees of freedom (*ndof*) are counted once for the initial Lagrangian of Eq. 3.68 and once for the final one in Eq. 3.69:

- **Initial *ndof*:** 4
 - 1 for the scalar ϕ field
 - 1 for the scalar ϕ^* field
 - 2 transverse components for the massless vector A_μ field
- **Final *ndof*:** 5
 - 1 for the massless scalar ξ field
 - 1 for the massive scalar η field
 - 3 = 2 + 1, 2 transverse components and 1 longitudinal component for the massive vector A_μ field

It is observed that the final *ndof* have been increased by one! The additional longitudinal degree of freedom to the vector boson was expected as it gained mass, so the evidence point that one of the other particles must not have a real physical meaning.

In the effort of finding a gauge that can eliminate one of the fields to restore the final *ndof*, the gauge transformation from Eq. 3.66 is rewritten by inverting the expansion on the $i\xi$ term.

$$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2}}[v + \eta(x) + i\xi(x)] \\ &\simeq \frac{1}{\sqrt{2}}[v + \eta(x)]e^{i\xi(x)/v} \end{aligned} \quad (3.70)$$

Notation is changed so that a different set of real fields (h , θ , A_μ) is used:

$$\phi \rightarrow \phi' = \frac{1}{\sqrt{2}}[v + h(x)]e^{i\theta(x)/v} \quad (3.71)$$

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{ev}\partial_\mu\theta(x) \quad (3.72)$$

and by substitution to the original \mathcal{L} (Eq. 3.68) we obtain:

$$\begin{aligned} \mathcal{L} = \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \lambda v^2 h^2 + \frac{1}{2}e^2 v^2 A_\mu A^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \\ - \lambda v h^3 - \frac{1}{4}\lambda h^4 + \frac{1}{2}e^2 h^2 A_\mu A^\mu + v e^2 h A_\mu A^\mu \end{aligned} \quad (3.73)$$

Now, the unwanted massless Goldstone boson θ does not appear in the theory. The Lagrangian just describes the interaction between the massive vector boson A_μ and the massive scalar boson h . The missing massless Goldstone boson that occurred after the spontaneous symmetry breaking of the local $U(1)$ symmetry is said to be absorbed by the vector boson of the theory and is manifested as an additional longitudinal degree of freedom to it. This is the result of the *Higgs mechanism* (proposed on 1964 by P. Higgs [15] and others like R. Brout, F. Englert [16], and G. S. Guralnic, C. R. Hagen, T. W. B. Kibble [17]) for spontaneous local symmetry breaking and the massive scalar boson is the Higgs particle. At last, a vector boson from a renormalizable theory acquired mass.

3.8 Unification – Act I

Two aspects need to be addressed in this section. On one side, there is still the open item of giving mass to the weak vector bosons. On the other side, there is an unresolved matter regarding weak interactions. It was mentioned that fermionic fields can be written as a function of two components, a left-handed one and a right-handed one. But the $SU(2)_L$ symmetry only concerns the left-handed treatment, and while two electrically charged vector boson fields can be identified (Eq. 3.49), the neutral one seems to have not revealed itself yet. The latter matter is addressed first.

3.8.1 Electroweak interactions and the $SU(2)_L \otimes U(1)_Y$ symmetry

We start by recalling the discussion on electromagnetic interactions. In fact, it is admitted that we “cheated” a bit in favor of getting a simple and meaningful result. The $U(1)$ transformation (Eq. 3.18) used to derive QED in Section 3.4 was written on its simplest form so that it only acts on fermions, specifically having in mind only the electron. But quarks do carry an electric charge themselves and the previous transformation can be rewritten as:

$$U(1)_{em} \equiv U(1)_Q = e^{iQa(x)} \quad (3.74)$$

where the operator for the electric charge Q (with eigenvalue of -1 for the electron) is introduced and the discussion can now be generalized also to quarks. It is then understood that the $U(1)_Q$ symmetry is the one that fully describes the electromagnetic force and corresponds to the conservation of the electric charge. In order for quark charges to emerge, the following definition is required for Q :

$$Q = I_3 + \frac{Y}{2} \quad (3.75)$$

where this relation associates the electric charge Q , with the weak isospin I and the hypercharge Y , the latter being a new quantum number. Before diving to the details of the above equation, some important equations from QED are rewritten in accordance with

Eq. 3.74:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iQeA_\mu \quad (3.76)$$

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{eQ} \partial_\mu a(x) \quad (3.77)$$

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - e\bar{\psi}\gamma^\mu Q\psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (3.78)$$

$$\text{where } (j^Q)^\mu = \bar{\psi}\gamma^\mu Q\psi \quad (3.79)$$

$$\text{i.e., an interaction of the form } -e(j^Q)^\mu A_\mu = -e(\bar{\psi}\gamma^\mu Q\psi)A_\mu \quad (3.80)$$

The charge definition in Eq. 3.75 tells us that actually electromagnetic and weak interactions can be unified into a single theory which goes under the name of *electroweak* interactions. The basis for this new gauge theory is set by the $SU(2)_L \otimes U(1)_Y$ symmetry, where the isospin I is conserved under the $SU(2)_L$ symmetry (discussed in Section 3.6) and the hypercharge Y is conserved under the $U(1)_Y$ symmetry. The field transformations for this symmetry group can be written as:

$$\chi_L \rightarrow \chi'_L = e^{i\vec{\theta}(x) \cdot \frac{\vec{\sigma}}{2} + i\beta(x) \frac{Y_L}{2}} \chi_L \quad (3.81)$$

$$\psi_R \rightarrow \psi'_R = e^{i\beta(x) Y_R} \psi_R \quad (3.82)$$

where χ_L is an isospin doublet and ψ_R an isospin singlet. It is observed that the $U(1)_Y$ symmetry acts on both the left and right-handed components of the fields, while the $SU(2)_L$ symmetry still acts only on the left-handed fields (see Eq. 3.46 and 3.47). To approach our goal, it is of great interest to isolate the interaction terms for the respective symmetries. To further simplify our case we focus only on the electron interaction terms.

$$\begin{aligned} \mathcal{L}(U(1)_Y) &= \bar{f}_{eL} i\gamma^\mu (ig_1 \frac{Y_L}{2} B_\mu) f_{eL} + \bar{e}_R i\gamma^\mu (ig_1 \frac{Y_R}{2} B_\mu) e_R \\ &= -\frac{g_1}{2} [Y_L (\bar{\nu}_L \gamma^\mu \nu_L + \bar{e}_L \gamma^\mu e_L) + Y_R \bar{e}_R \gamma^\mu e_R] B_\mu \end{aligned} \quad (3.83)$$

$$\begin{aligned} \mathcal{L}(SU(2)_L) &= \bar{f}_{eL} i\gamma^\mu (ig_2 \frac{\sigma_i}{2} W_\mu^i) f_{eL} \\ &= -\frac{g_2}{2} (\bar{\nu}_L \gamma^\mu \nu_L W_\mu^0 + \sqrt{2} \bar{\nu}_L \gamma^\mu e_L W_\mu^+ + \sqrt{2} \bar{e}_L \gamma^\mu \nu_L W_\mu^- - \bar{e}_L \gamma^\mu e_L W_\mu^0) \end{aligned} \quad (3.84)$$

As per usual, for each symmetry group invariance is demanded and new vector fields are incorporated into the theory through the definition of the covariant derivatives. A brief summary is given. For $U(1)_Y$, g_1 is the coupling constant of the gauge theory, $Y_{L,R}$ is the hypercharge operator, the generator of the group, acting on either left or right-handed fields, and B_μ (not the photon field) the sole vector gauge boson of the theory. Similarly, for $SU(2)_L$ ⁷, g_2 is the coupling constant of the theory, σ_i the Pauli matrices, and W_μ^i the three vector gauge bosons of the theory. Take note that in Eq. 3.84 the charged W^\pm bosons are formed (definition in Eq. 3.49).

The next step is to combine terms from both Eq. 3.83 and 3.84 so as to separate the electron from the neutrino interactions.

$$\text{electron: } -\bar{e}_L \gamma^\mu e_L \left(\frac{g_1}{2} Y_L B_\mu - \frac{g_2}{2} W_\mu^0 \right) + \bar{e}_R \gamma^\mu e_R \left(\frac{g_1}{2} Y_R B_\mu \right) \quad (3.85)$$

$$\text{neutrino: } -\bar{\nu}_L \gamma^\mu \nu_L \left(\frac{g_1}{2} Y_L B_\mu + \frac{g_2}{2} W_\mu^0 \right) \quad (3.86)$$

while the corresponding term for $U(1)_Q$ in Eq. 3.80 can be reformulated as:

$$e\bar{e}\gamma^\mu e A_\mu = e(\bar{e}_L \gamma^\mu e_L + \bar{e}_R \gamma^\mu e_R) A_\mu \quad (3.87)$$

⁷Exactly the same notation is used as in Section 3.6, when $SU(2)_L$ was initially introduced.

Comparing the electron to the neutrino interactions and to the electromagnetic one, a linear combination is observed between the B_μ and W_μ^0 fields. This fact is exploited to define two new fields which will be given as an orthogonal combination of B_μ and W_μ^0 . Thus, Eq. 3.85 and 3.86 are yet again reformed as a function of these two new fields:

$$\begin{aligned} \text{electron:} \quad & - \left[\bar{e}_L \gamma^\mu e_L \left(\frac{g_1 g_2 Y_L}{\sqrt{g_2^2 + g_1^2 Y_L^2}} \right) + \bar{e}_R \gamma^\mu e_R \left(\frac{g_1 g_2 Y_R}{2\sqrt{g_2^2 + g_1^2 Y_L^2}} \right) \right] A_\mu \\ & - \left[\bar{e}_L \gamma^\mu e_L \left(\frac{g_1^2 Y_L^2 - g_2^2}{2\sqrt{g_2^2 + g_1^2 Y_L^2}} \right) + \bar{e}_R \gamma^\mu e_R \left(\frac{g_1 Y_L Y_R}{2\sqrt{g_2^2 + g_1^2 Y_L^2}} \right) \right] Z_\mu \end{aligned} \quad (3.88)$$

$$\text{neutrino:} \quad - \bar{\nu}_L \gamma^\mu \nu_L \left(\frac{\sqrt{g_2^2 + g_1^2 Y_L^2}}{2} \right) Z_\mu \quad (3.89)$$

where

$$Z_\mu = \frac{1}{\sqrt{g_2^2 + g_1^2 Y_L^2}} (g_1 Y_L B_\mu + g_2 W_\mu^0) \quad (3.90)$$

$$A_\mu = \frac{1}{\sqrt{g_2^2 + g_1^2 Y_L^2}} (g_2 B_\mu - g_1 Y_L W_\mu^0) \quad (3.91)$$

These two fields seem quite familiar! Evidently, the photon field A_μ and the neutral weak boson Z_μ are given as a function of the B_μ and W_μ^0 fields. While the actual mechanism which causes the A_μ and Z_μ fields to emerge will be discussed in the next subsection, it is nice to already see that all three of the weak vector bosons have been assembled at last. Comparing the photon term of Eq. 3.88 with Eq. 3.87 it is deduced that:

$$e = \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}}, \quad \text{where} \quad Y_L = \frac{Y_R}{2} = -1 \quad (3.92)$$

a relationship between the coupling constant of the electromagnetic interactions ($U(1)_Q \rightarrow e$) and those of the electroweak interactions ($U(1)_Y, SU(2)_L \rightarrow g_1, g_2$). Inspired by this relationship the *weak mixing angle* θ_W is defined:

$$\begin{aligned} \sin \theta_W &= \frac{g_1}{\sqrt{g_1^2 + g_2^2}} \\ \cos \theta_W &= \frac{g_2}{\sqrt{g_1^2 + g_2^2}} \end{aligned} \quad (3.93)$$

Using the weak mixing angle from above, Eq. 3.90, 3.91, and 3.92 can be redefined as:

$$e = g_1 \cos \theta_W = g_2 \sin \theta_W \quad (3.94)$$

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^0 \quad (3.95)$$

$$Z_\mu = -\sin \theta_W B_\mu + \cos \theta_W W_\mu^0 \quad (3.96)$$

The θ_W angle constitutes a free parameter of the SM which is estimated only experimentally and found to be slightly below $30^\circ \Rightarrow \sin^2 \theta_W \approx 0.23$. Equation 3.94 relates the g_1, g_2 couplings, with the electromagnetic coupling e , and with the QED fine-structure constant $\alpha \equiv e^2/4\pi = 1/137$. Lastly, interactions of the type as in Eq. 3.88 and Eq. 3.89 that involve the neutral Z_μ boson are usually referred to as *Neutral Current (NC)* interactions, compared to interactions that involve the charged W^\pm bosons as in Eq. 3.84 which fall under the *Charged Current (CC)* type.

3.8.2 Electroweak symmetry breaking $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$

How can it be that from two distinct scenarios, the $U(1)_Q$ and the $SU(2)_L \otimes U(1)_Y$, the exact same theory is derived, i.e., the description of the electromagnetic interactions? In that respect, remember that by forming the electroweak symmetry we have successfully managed to in addition reveal the neutral Z_μ^0 boson. Having the $SU(2)_L \otimes U(1)_Y$ gauge symmetry, which now contains four vector bosons, we have one but last wish, that is to give mass to the three weak vector bosons while preserving the photon in a massless state for the electromagnetic interaction. Following the usual practice we proceed in breaking the electroweak symmetry.

The discussion starts with four scalar fields arranged in an $SU(2)$ complex doublet ($T = 1/2 \Rightarrow T_3 = \pm 1/2$), a choice originally made by Weinberg [18] in parallel with Glashow [19] and Salam [20].

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (3.97)$$

where ϕ is constructed in such a way so that ϕ^+ is positively charged $Q = +1$ and ϕ^0 is neutrally charged $Q = 0$. Substitution to Eq. 3.75 returns $Y = 1$ for both the “up” and “down” states of ϕ . The minimum for the potential $V(\phi) = \mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2$ is found for $\lambda > 0$ and $\mu^2 < 0$:

$$\phi^\dagger\phi \equiv (\phi^+)^*\phi^+ + (\phi^0)^*\phi^0 = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{-\mu^2}{2\lambda} \equiv \frac{v^2}{2} \quad (3.98)$$

a hypersphere in four dimensions. Without loss of generality we chose the vacuum for the theory to be $\phi_1 = \phi_2 = \phi_4 = 0$, $\phi_3^2 = -\frac{\mu^2}{\lambda} \equiv v^2$, resulting in the choice:

$$\phi_0 \equiv \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (3.99)$$

Expanding the ϕ field around the VEV now gives:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (3.100)$$

Substitution of Eq. 3.100 into the electroweak Lagrangian below (Lagrangian that contains interactions terms between the ϕ field and the gauge boson fields of $SU(2)_L \otimes U(1)_Y$):

$$\mathcal{L}_\phi = (D_\mu\phi)^\dagger(D^\mu\phi) - [\mu^2(\phi^\dagger\phi) + \lambda(\phi^\dagger\phi)^2] \quad (3.101)$$

$$\text{with } D_\mu = \partial_\mu + ig_1 \frac{Y}{2} B_\mu + ig_2 \frac{\sigma_i}{2} W_\mu^i = \partial + \frac{1}{2} ig_1 B_\mu + \frac{1}{2} ig_2 \sigma_i W_\mu^i \quad (3.102)$$

returns:

$$\begin{aligned} \mathcal{L}_h = & \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{8}g_2^2 v^2 [(W_\mu^1)^2 + (W_\mu^2)^2] + \frac{1}{2}v^2 (g_1 B_\mu - g_2 W_\mu^3)^2 - \lambda v^2 h^2 \\ & + \text{cubic interaction terms} + \text{quartic interaction terms} + \text{const} \end{aligned} \quad (3.103)$$

while this is rewritten by incorporating the four gauge bosons of the electroweak interaction according to Eq. 3.49 for W_μ^\pm and Eq. 3.90 and 3.91 for Z_μ , A_μ as:

$$\begin{aligned} \mathcal{L}_h = & \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{4}g_2^2 v^2 W_\mu^+ W_\mu^- + \frac{1}{4}(g_1^2 + g_2^2)v^2 Z_\mu Z^\mu - \lambda v^2 h^2 \\ & + \text{cubic interaction terms} + \text{quartic interaction terms} + \text{const} \end{aligned} \quad (3.104)$$

Three mass terms emerge in the Lagrangian for the weak gauge bosons W_μ^\pm , Z_μ in addition to the mass term attributed to the Higgs field, the photon field remaining massless ($M_A = 0$):

$$M_W^2 = \frac{1}{4}g_2^2v^2 \quad (3.105)$$

$$M_Z^2 = \frac{1}{4}(g_1^2 + g_2^2)v^2 = \frac{M_W}{\cos\theta_W} \quad (3.106)$$

$$M_h^2 = 2\lambda v^2 \quad (3.107)$$

The effect is understood as follows: the choice of the new vacuum breaks the electroweak interactions i.e., the $SU(2)_L \otimes U(1)_Y$ is spontaneously broken, both of the individual symmetries are actually broken in a way that the Lagrangian is no longer invariant under their respective field transformations. From the initial four generators, three are spontaneously broken, leading to three massless scalar Goldstone bosons that are eaten by the weak force mediators (for each gauge boson an additional longitudinal degree of freedom is added in the existing two transverse ones) thus acquiring their mass. One degree of freedom remains for the Higgs field, the Higgs particle that was discovered in 2012 by the ATLAS and CMS collaborations at CERN. There is an infinite amount of choices for the vacuum, all of which lie within a spherical plane, another symmetry. When the electroweak interactions are spontaneously broken, not all symmetry is lost, in the contrary, the system collapses in the internal symmetry shared by the vacuum states which is respected by the resulting Lagrangian. This is the $U(1)_Q$ symmetry of the electromagnetic interactions and it is usually written as $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$. The $U(1)_Q$ of electromagnetism is just the low energy approximation of the system.

The weak mixing angle θ_W encountered in Subsection 3.8.2 (Eq. 3.93) is said to be the angle by which the $W_\mu^3 - B_\mu$ plane is rotated upon the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$ leading to the production of the Z_μ boson and the photon A_μ . The coupling of the $SU(2)_L$ symmetry (g_2) is related with the known Fermi constant G_F and with the VEV v as:

$$\frac{G_F}{\sqrt{2}} = \frac{g_2^2}{8M_W^2} = \frac{1}{2v^2} \quad (3.108)$$

with the latter approximately estimated at $v \approx 246 \text{ GeV}$. Combining this value and the value of θ_W , the W_μ^\pm and Z_μ masses are predicted by the SM according to Eq. 3.105 and 3.106 at $M_W \approx 80 \text{ GeV}$ and $M_Z \approx 91 \text{ GeV}$ which were later verified from their discovery at CERN in 1953 [21–23]. The mass of the Higgs boson is not predicted by the SM due to the unknown free parameter λ in Eq. 3.107 and is only experimentally measured as in [3, 4] ($M_h \approx 125 \text{ GeV}$).

The final item in the list is to attribute mass to the fermions which is also a direct consequence of the spontaneous symmetry breaking mechanism. An interaction term between a fermion and a scalar field looks like this:

$$\mathcal{L}_{f,\phi} = -g_f \bar{f} f \phi \quad (3.109)$$

where this type of interaction is called a *Yukawa interaction* and the constants g_f as *Yukawa coupling constant*. Expanding the above Lagrangian for the first generation of fermions (same for the other generations):

$$\begin{aligned} \mathcal{L}_{f,\phi} = & -g_e(\bar{f}_{eL}\phi e_R + \bar{e}_R\phi^\dagger f_{eL}) \\ & -g_d(\bar{f}_{dL}\phi d_R + \bar{d}_R\phi^\dagger f_{dL}) -g_u(\bar{f}_{uL}\phi^c u_R + \bar{u}_R\phi^{c\dagger} f_{uL}) \end{aligned} \quad (3.110)$$

where to be able to attribute mass to the up quark the charge conjugate field ϕ^c is introduced.

$$\phi^c = i\tau_2\phi^* = \begin{pmatrix} \phi^{0*} \\ -\phi^- \end{pmatrix} \quad (3.111)$$

The charge conjugate field ϕ^c is an $SU(2)_L$ doublet, just like the ϕ field, only this time corresponding to an opposite hypercharge state $Y = -1$. The ϕ^c field is expanded around the VEV as previously done for ϕ (Eq. 3.100):

$$\phi^c(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} v + h(x) \\ 0 \end{pmatrix} \quad (3.112)$$

The result after replacing both Eq. 3.100 and Eq. 3.112 to Eq. 3.110 is:

$$\begin{aligned} \mathcal{L}_{f,h} = & -m_e \bar{e}e - m_d \bar{d}d - m_u \bar{u}u \\ & - \frac{m_e}{v} \bar{e}eh - \frac{m_d}{v} \bar{d}dh - \frac{m_u}{v} \bar{u}uh \end{aligned} \quad (3.113)$$

with the fermion masses revealed as:

$$m_e = \frac{g_e v}{\sqrt{2}}, \quad m_d = \frac{g_d v}{\sqrt{2}}, \quad m_u = \frac{g_u v}{\sqrt{2}} \quad (3.114)$$

It is observed that the fermion masses are proportional to the Yukawa coupling constants which are arbitrary and unknown. Due to this definition, fermion masses are left as free SM parameters which can only be experimentally estimated. Another conclusion that can be drawn from Eq. 3.113 is that the Higgs field interacts with the fermions and the strength of that interaction depends on their mass. The heavier the mass of a fermion, the stronger the interaction to the Higgs field.

3.9 Unification – Act II

To obtain the final coveted theory that will cover all aspects of the story there is a need to write a symmetry such that will preserve the characteristics found by the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$ while still introducing the quark dynamics. Luckily, this final step is but trivial as the unification of electroweak and strong interactions is given by the plain and simple combination of the respective symmetries $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$. Quark dynamics are introduced exactly as seen in Section 3.5 and invariance under the $SU(3)_C$ symmetry remains even after the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_Q$ so that the gluons remain massless. A lot more could be said for the QCD part of our theory but for now we settle for the basics as the full Standard Model Lagrangian unravels itself.

3.9.1 The Standard Model final Lagrangian

The full, but compact, version of the Standard Model Lagrangian is given below, that is consistent with the symmetries implied upon the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry

group. The Lagrangian can be split in four fundamental sectors:

$$\begin{aligned}
 \mathcal{L}_{SM} = & \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa} \\
 & - \frac{1}{4} G^{\mu\nu\alpha} G_{\mu\nu}^\alpha - \frac{1}{4} W^{\mu\nu i} W_{\mu\nu}^i - \frac{1}{4} B^{\mu\nu} B^{\mu\nu} \quad \left\{ \begin{array}{l} \text{Kinetic energy and self} \\ \text{coupling terms for the} \\ \text{gluons, } W^\pm, Z \text{ and } \gamma \end{array} \right. \\
 & + \sum_{\psi} \bar{\psi} i \gamma^\mu D_\mu \psi \quad \left\{ \begin{array}{l} \text{Kinetic energy for leptons} \\ \text{and quarks and their inter-} \\ \text{action terms to gluons, } W^\pm, \\ Z \text{ and } \gamma \end{array} \right. \\
 & + (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi) \quad \left\{ \begin{array}{l} \text{Higgs and mediator masses} \\ \text{and couplings} \end{array} \right. \\
 & - \sum_f g_f (\bar{\psi}_L \phi \psi_R + h.c.) \quad \left\{ \begin{array}{l} \text{Fermion masses and} \\ \text{coupling to Higgs} \end{array} \right.
 \end{aligned} \tag{3.115}$$

where,

$$\begin{aligned}
 \psi & : \text{Represents all fermionic fields} \\
 D_\mu = \partial_\mu - ig_3 \frac{\lambda_a}{2} G_\mu^a - ig_2 \frac{\sigma_i}{2} W_\mu^i - ig_1 \frac{Y}{2} B_\mu & : \text{The covariant derivative} \\
 V(\phi) = \mu^2 (\phi^\dagger \phi) + \lambda (\phi^\dagger \phi)^2 & : \text{The Higgs potential} \\
 g_f & : \text{The Yukawa coupling constants} \\
 \psi_L \text{ and } \psi_R & : \text{Left and right-handed fermions} \\
 h.c. & : \text{The hermitian conjugate}
 \end{aligned} \tag{3.116}$$

3.10 Unification – Act III

But the story does not end here. While the Standard Model itself comprises a great achievement of the science community of the 20th and 21st centuries and has shaped the modern particle physics as we know it today, it is yet an incomplete theory. Still, unresolved questions remain, with the scientists trying their best as usual to power through this plateau of knowledge.

To name a few cases, gravity is not included in the SM. Dark matter and dark energy though showing promising evidence that they do exist, no such particle candidates are incorporated in the current theory. Neutrinos are assumed massless while their mass is finite, though small, and has been experimentally estimated. Matter over antimatter asymmetry is not explained. There is an exhaustive list of unmodeled phenomena that are not treated within the SM and efforts are being made to compose new theories or extensions of the SM to incorporate them. Some examples are string theory, supersymmetry, Grand Unified Theories (GUTs), extra dimensions or quantum gravity frameworks.

More importantly, experiments that can explore new regions of the energy scale spectrum are of great need as such experiments can provide more evidence and guide the theory community towards a more concrete direction. Greater scale experiments can help narrow down the list of the theory frameworks that are being developed in hopes of finding new physics. What lies into the future is quite interesting, with each new generations of scientists pushing, advancing and exploring the unknown.

Bibliography

- [1] Wikipedia. Fundamental Interactions. Available: https://en.wikipedia.org/wiki/Fundamental_interaction#References. Accessed: 2024-11-05.
- [2] Izaak Neutelings. Standard Model. Available: https://tikz.net/sm_particles/. Accessed: 2024-11-04.
- [3] G. Aad et al. [ATLAS Collaboration]. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012. <https://doi.org/10.1016/j.physletb.2012.08.020>.
- [4] S. Chatrchyan et al. [CMS Collaboration]. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [5] D. Griffiths. *Introduction to Elementary Particles*. Wiley, 2008. 2nd ed.
- [6] A. Pais and S. B. Treiman. How Many Charm Quantum Numbers are There? *Phys. Rev. Lett.*, 35:1556–1559, Dec 1975. <https://link.aps.org/doi/10.1103/PhysRevLett.35.1556>.
- [7] Wikipedia. Mathematical formulation of the Standard Model. Available: https://en.wikipedia.org/wiki/Mathematical_formulation_of_the_Standard_Model. Accessed: 2024-11-08.
- [8] Emmy Noether. Invariant variation problems. *Transport Theory and Statistical Physics*, 1(3):186–207, jan 1971. <http://dx.doi.org/10.1080/00411457108231446>.
- [9] R. Mann. *An Introduction to Particle Physics and the Standard Model*. Taylor & Francis Group, 2010. <https://www.taylorfrancis.com/books/oa-mono/10.1201/9781420083002/introduction-particle-physics-standard-model-robert-mann>.
- [10] M. Robinson. *Symmetry and the Standard Model*. Springer New York, NY, 2011. <https://link.springer.com/book/10.1007/978-1-4419-8267-4>.
- [11] F. Halzen and A. D. Martin. *Quarks & Leptons: An Introductory Course in Modern Particle Physics*. John Wiley & Sons, 1984. <https://www.wiley.com/en-gb/Quarks+and+Leptones%3A+An+Introductory+Course+in+Modern+Particle+Physics-p-9780471887416>.
- [12] C. N. Yang and R. L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Phys. Rev.*, 96:191–195, Oct 1954. <https://link.aps.org/doi/10.1103/PhysRev.96.191>.
- [13] K. E. Vagionakis. *Particle Physics, An Introduction to the Basic Structure of Matter*. NTUA, 2013.
- [14] J Goldstone. Field theories with "superconductor" solutions. *Nuovo Cimento*, 19:154–164, 1961. <https://doi.org/10.1007/BF02812722>.
- [15] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [16] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964. <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.

-
- [17] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964. <https://link.aps.org/doi/10.1103/PhysRevLett.13.585>.
- [18] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [19] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, 1961. <https://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [20] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968. <https://inspirehep.net/literature/53083>.
- [21] G. Arnison et al. [UA1 Collaboration]. Experimental observation of isolated large transverse energy electrons with associated missing energy at $s=540$ GeV. *Physics Letters B*, 122(1):103–116, 1983. <https://www.sciencedirect.com/science/article/pii/0370269383911772>.
- [22] G. Arnison et al. [UA1 Collaboration]. Experimental observation of lepton pairs of invariant mass around $95 \text{ GeV}/c^2$ at the CERN SPS collider. *Physics Letters B*, 126(5):398–410, 1983. <https://www.sciencedirect.com/science/article/pii/0370269383901880>.
- [23] M. Banner et al. [UA2 Collaboration]. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider. *Physics Letters B*, 122(5):476–485, 1983. <https://www.sciencedirect.com/science/article/pii/0370269383916052>.

Chapter 4

ANALYSIS OBJECTS – JETS

4.1 Introduction

With this chapter we continue to delve deeper into the goal of this dissertation. All following sections shed light on the analysis's primary objects, the jets. Jets are among the most frequently used objects in high energy collider experiment analyses. A lot of times they are considered the signature objects of a collision and they are quite complex. Hopefully, within this chapter some of their underlying mechanisms and properties will be shared that will help the reader gain some familiarity on their use and importance. Each chapter provides a different insight into the aspects of jets covering both the theoretical and experimental point of view, by the end, promising a cohesive perspective of the topic.

4.2 Consequences of Quantum Chromodynamics

4.2.1 Hadrons

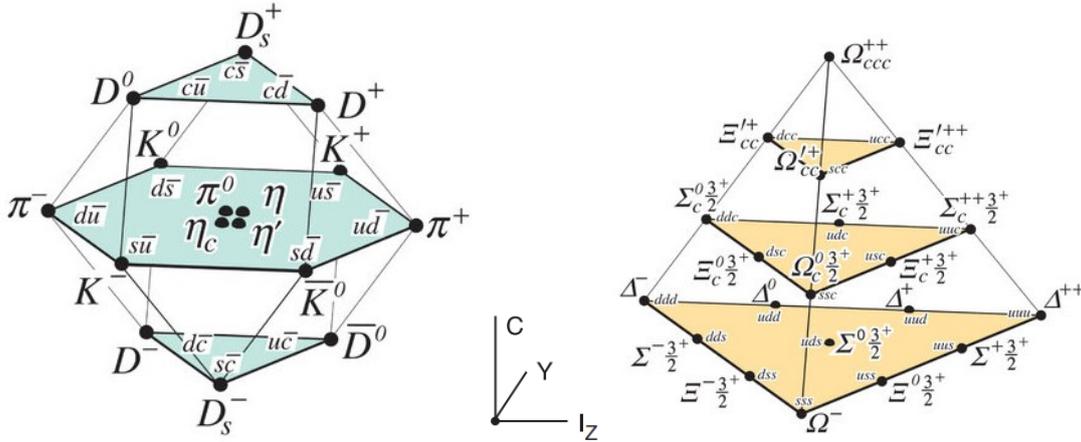
We start perhaps with the most fundamental quantity of QCD, or better, of the $SU(3)_C$ symmetry, the color. As mentioned in Section 3.5, the color is a new quantum number intrinsic to quarks and gluons (particles that interact primarily via the strong force) and color conservation is assumed in any given process that these participate. Our focus is slightly favored towards QCD since quarks and gluons are seen in abundance in particle colliders, like the LHC, and the mechanisms that are introduced here pave the way on the topic of jet formation.

Within the Standard Model it is dictated that there are three colors, red, green, and blue. This particular choice of colors is but a plain convention adapted by the community, which in addition, is supported by color theory. Having this argument in mind and based on the mathematical framework provided by $SU(3)_C$ the observed matter particles can only form singlet colorless states (states that will remain invariant under color transformations, if any). In that sense quarks can be arranged into two distinct configurations to form bound states. They can either form symmetric combinations of quark-antiquark pairs $q_i\bar{q}_i$ known as *mesons* that have a baryon number of zero $\mathcal{B} = 0$, or form antisymmetric combinations, usually of three quarks $q_iq_jq_k$ and rarely of three quarks plus any number of quark-antiquark pairs, known as *baryons* that have a baryon number of one $\mathcal{B} = 1$. Together, mesons and baryons, constitute the *hadrons*, basically all subatomic particles that are made of two or more quarks fall in that category. Table 4.1 summarizes all quark related quantum numbers. By exploiting this table scientists were able to arrange the hadrons into groups based on their properties and according to their quark constituents, quite similar to what was done with the periodic table of elements in chemistry. One such

example is given for both mesons and baryons in Fig. 4.1. Of course, not all hadrons known today had been discovered at that time, but this systematic arrangement provided a foundation for making predictions.

Table 4.1: Quark quantum numbers [1].

	d	u	s	c	b	t
Electric charge Q	$-\frac{1}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$	$-\frac{1}{3}$	$+\frac{2}{3}$
Isospin I	$\frac{1}{2}$	$\frac{1}{2}$	0	0	0	0
Isospin z-component I_z	$-\frac{1}{2}$	$+\frac{1}{2}$	0	0	0	0
Strangeness S	0	0	-1	0	0	0
Charm C	0	0	0	+1	0	0
Bottomness B	0	0	0	0	-1	0
Topness T	0	0	0	0	0	+1



(a) 16-plet of pseudoscalar mesons $\mathcal{B} = 0$. (b) (c) 20-plet of spin 1/2 baryons $\mathcal{B} = 1$.

Figure 4.1: Weight diagram examples of a meson and baryon multiplet. The coordinate system (b) is given as a function of isospin I_z , charm C and hypercharge $Y = \mathcal{B} + S - \frac{C-B+T}{3}$ [1].

4.2.2 Confinement

Hadrons comprise the observable particles be it in collider experiments or cosmic rays, but quarks (and gluons) are never observed as isolated free particles in nature, only in stable bound states. This phenomenon is called *confinement* or *quark confinement*. Indeed, looking at calculations in perturbative QCD (pQCD) it can be seen that the potential between a quark-antiquark pair for very small distances has a dominant Coulomb-like term ($\sim 1/r$) just as in QED for an electron-positron pair. Bringing too close a quark and an antiquark can result in a binding force. In contrast, when going to larger distances non perturbative calculations show that the potential becomes linear with the distance ($\sim r$). This means that as the two quarks are separated from each other the system requires

continuously more energy [2]. Both arguments lead to the confinement of quarks into colorless stable bound states, the hadrons, which are experimentally observed. This single effect results in the creation of hadrons after the hard collision in high energy experiments through a process known as *hadronization*, ultimately giving rise to jets.

4.2.3 Asymptotic freedom and the strong coupling

Interestingly enough, another core idea can be revealed about QCD just by studying the strong coupling constant α_S (found as α_3 in the context of Chapter 3) which determines the strength of the interaction between quarks and gluons. In reality α_S is not a constant in the usual kind of meaning, but its value depends on the energy scale that it is measured. This is also referred as *running of the strong coupling constant* and in perturbative QCD is estimated as:

$$\alpha_S(Q^2) = \frac{\alpha_S(\mu_r^2)}{1 + b_0 \ln(Q^2/\mu_r^2)\alpha_S(\mu_r^2)} \quad (4.1)$$

where

$$b_0 = \frac{33 - 2N_F}{12\pi} \quad (4.2)$$

The above equation relates the value of the coupling at a scale Q to the one at scale μ_r given that both scales are in the perturbative regime [3]. N_F expresses the number of quark flavors with masses smaller than the scale μ_r . Note that b_0 remains positive as long as $N_F \leq 16$, thus the positive sign remains in the α_S equation and the strength of the coupling decreases with Q^2 , i.e. the higher the energy (or for small distances) the smaller the value of α_S . Conversely, the smaller the energy, or as the distance increases, the higher the strength of the interaction.

This effect comes in contrast to the one observed in the QED analogue for the electron. In QED when two charges are brought in close range the strength of the interaction increases, i.e. the effective electric charge increases as the *screening* of the electric charge weakens. QCD on the other hand falls under the opposite situation as there is the *anti-screening* of the color charge. The self-interacting nature of gluons negates the screening done by opposite color charges. This contradiction between the two theories comes from the non-Abelian property of QCD that allows gluons to carry color and self-interact, in contrast to photons that are electrically neutral. Inevitably, in the higher energy scales quarks and gluons interact weakly and can be treated as asymptotically free traveling particles, a phenomenon referred as *asymptotic freedom*. In the asymptotically free regime the value of the strong coupling is very small $\alpha_S \ll 1$ and perturbation theory can be applied. Asymptotic freedom enabled the correct translation of the experimental data found from Deep Inelastic Scattering (DIS) experiments at SLAC which reported a point-like behavior of loosely moving particles residing within the proton after electron scattering. These were the valence quarks and the quark-gluon sea that is assumed to exist within the protons.

The running of the strong coupling is shown as a function of the energy scale in Fig. 4.2. Markers represent the evaluation of α_S at different scales Q as given by a large variety of signature experiments. Vertical error bars around the markers indicate the total uncertainty (experimental and theoretical) for each given measurement. In the background, the continuous black dashed line that spans over the whole energy range represents the evolved α_S value as estimated from the initial world average value at the scale of the Z boson mass m_Z , $\alpha_S(m_Z) = 0.1180 \pm 0.0009$. Lastly, the yellow band around the dashed line expresses the associated uncertainty. The behavior shown in Fig. 4.2 is in complete agreement with the asymptotic freedom hypothesis discussed previously.

The strong coupling grows very large, very fast, especially as $Q \rightarrow 0$. It holds great interest to find a reference point below which pQCD calculations can be applied and above which non-perturbative (NP) effects become dominant. Usually, the scale Λ_{QCD} is defined

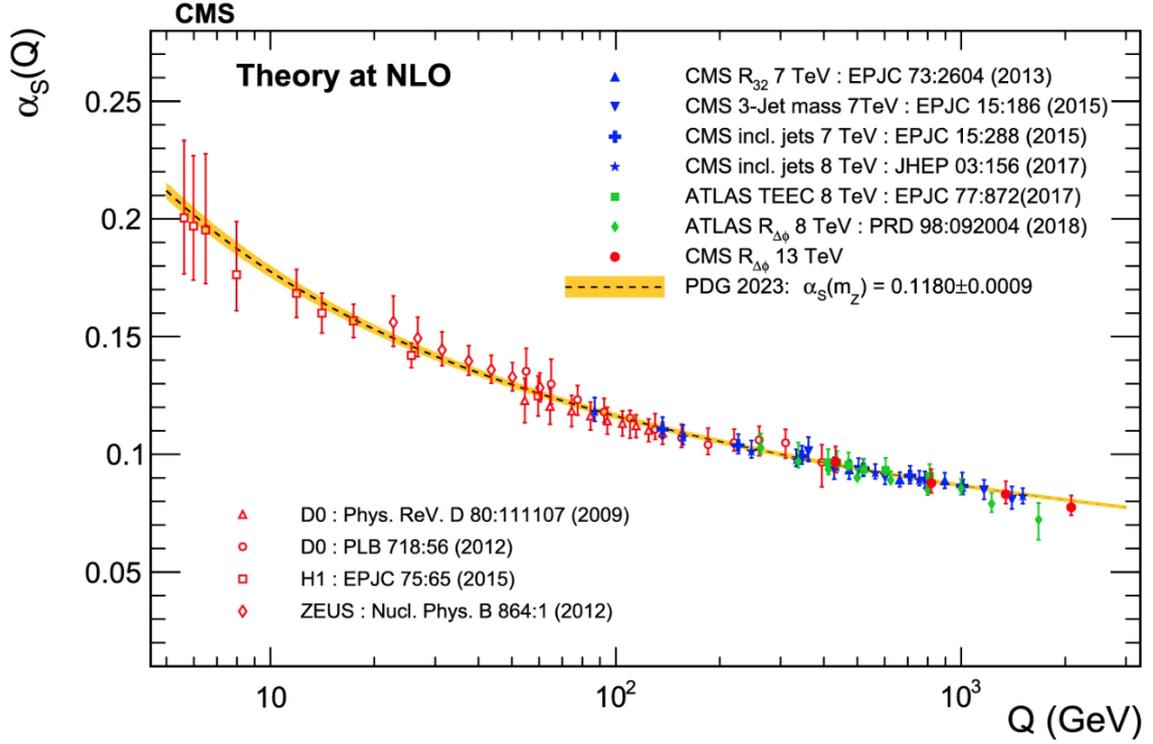


Figure 4.2: Running of the strong coupling α_S as a function of the energy scale Q [4].

to express exactly that, the point at which $\alpha_S(Q)$ diverges [3]. Equation 4.1 is then rewritten as:

$$\alpha_S(Q^2) = \frac{1}{b_0 \ln(Q^2/\Lambda_{QCD}^2)} \quad (4.3)$$

Close to the scale Λ_{QCD} there is the formation of the hadron bound states. This parameter is not predicted by QCD and is model dependent based on the renormalization scheme choice (more about this topic is shared in the following section). It is experimentally estimated, with its exact value being on the order of a few hundred MeV , $\Lambda_{QCD} \approx 200 MeV$, notably close to a typical mass of a hadron.

Given the opportunity, it is worth of mention to return briefly to the topic of Grand Unified Theories (GUTs). Presumably, all coupling constants from the three gauge theories contained within the SM converge at very large energy scales. This unification is assumed to be similar to what was done for weak and electromagnetic interaction to obtain the electroweak theory, only this time including the strong force. A guesstimate point for this grand unification scheme is estimated at $\sim 10^{15} GeV$, an energy scale so large that at first sight seems hopeless to be verified in the foreseeable future (compare the aforementioned value with the x -axis scale of Fig. 4.2). Hypotheses continue with even greater unification schemes where gravity is also incorporated into the picture, labeled this time as Theory of Everything (TOE), which would happen to even larger energy regimes. A nice representation of the above concepts is given in Fig. 4.3.

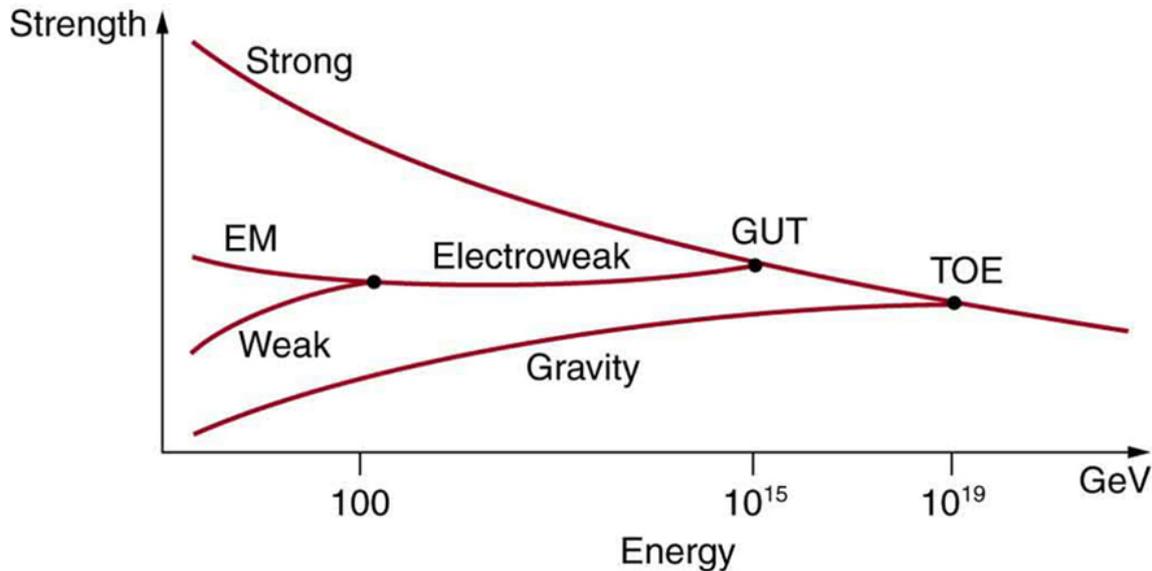


Figure 4.3: Relative strength of the four fundamental forces as a function of energy [5].

4.3 Into the pQCD abyss

4.3.1 Feynmann diagrams

Already, the concept of perturbative QCD has been mentioned a number of times throughout the text. The time has come where we can delve a bit deeper into this subject. Richard Phillips Feynmann invented a pictorial representation for particle interactions in the context of Quantum Field Theory (QFT), the *Feynmann diagrams* [6]. Over the years, their use on QFT perturbative calculations has been proved invaluable. The exact derivation of these rules is omitted as it goes beyond the scope of this thesis, instead the reader is encouraged to look up the specifics in the dedicated QFT textbooks like Ref. [7, 8]. It suffices to say that the *Feynmann rules*, that govern the construction of the diagrams, are extracted from none other than the Lagrangian, with some modifications to it, like quantization of the fields and introduction of ghost terms, anyhow.. The QCD Feynmann rules dictate some acceptable basic vertices which are shown in Fig. 4.4. Starting from left to right there is the quark-gluon interaction vertex, the 3-leg gluon self-interaction vertex, and the 4-leg gluon self-interaction vertex. At this point, the vertices shown in Fig. 4.4 do not represent any physical process on their own, rather they can be thought as the building blocks used to depict physical interactions.

A more realistic example is perhaps shown in Fig. 4.5a where there is a gluon exchange between two quarks. Looking at the top vertex of the figure, there is a green quark which changes into a blue quark by emitting a green anti-blue gluon. In the bottom vertex, we can imagine a blue quark absorbing the green anti-blue gluon, in the end, changing to a green quark. Color is conserved in both vertices. Again, this process is not observed in nature as quarks are not seen as free traveling particles but they appear inside the hadrons.

Similar rules exist also for the rest of the interactions but for the sake of brevity are not mentioned here. Instead, we directly discuss on the purely QED process $e^-e^+ \rightarrow \mu^-\mu^+$ depicted in Fig. 4.5b. Looking at the diagram from left to right it illustrates an electron positron annihilation resulting in the emission of a photon which in turn decays into a muon anti-muon pair. Note that Feynmann diagrams are read from left to right, which is defined as the direction at which time t evolves. It is then clear, that in this process

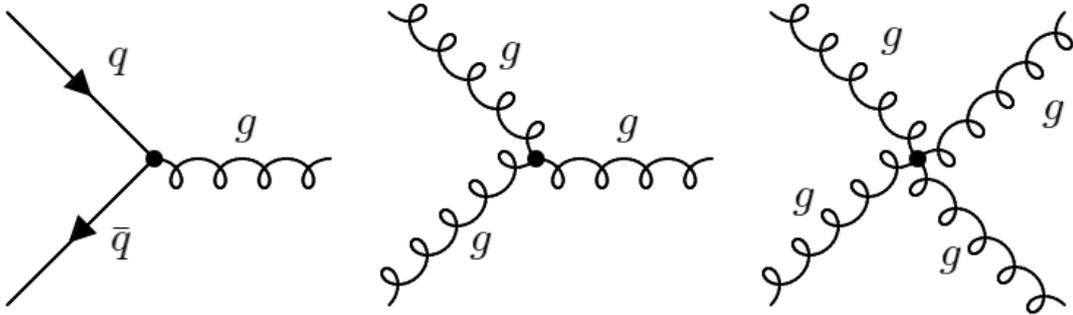
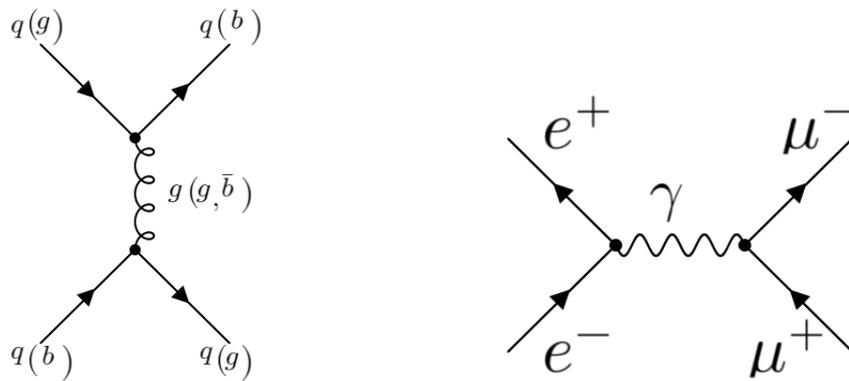


Figure 4.4: QCD Feynmann rules (made with the Feynmann diagram maker from Ref. [9]).



(a) Gluon exchange between two quarks.

(b) The process $e^-e^+ \rightarrow \mu^-\mu^+$.

Figure 4.5: Simple examples of Feynmann diagrams [9].

the initial state particles are the two electrons and the final state particles are the two muons. Additionally, particles are depicted with straight lines that point towards the same direction as time evolves (towards the right). On the other hand, antiparticles are expressed with the straight lines that point to the opposite direction, as if they travel backwards in time. Lastly, any 90° rotation of a diagram results in the representation of a different physical process. A clockwise 90° rotation of the diagram in Fig. 4.5b for example results in the electron muon scattering process $e^-\mu^- \rightarrow e^-\mu^-$.

4.3.2 Parton cross sections and decay rates

Situation requires a physical quantity that can be measured both experimentally and theoretically. Comparison between the two, initiates a feedback loop in both directions. Any differences between theory and experiment translate to either miss treatment on the detection and analysis front or to physical aspects not included in the theoretical model. Conversely, a good agreement can lead to estimation of other, more fundamental, quantities like α_S and the fermion masses, or as it will be later discussed the Parton Distribution Functions.

There are two famous choices that fulfill the aforementioned requirement, *parton cross sections* σ and *decay rates* Γ . The former, in the context of collider experiments, in a way, expresses the probability that a given process will happen. Experimentally, it is directly related to the number of observed events and the luminosity of a collider, as discussed in

Eq. 1.3. Theoretically, it can be estimated given the *Matrix Element (ME)* \mathcal{M} or *amplitude*, and the available *phase space*. The ME gives the probability amplitude to obtain a desired final state; it encodes the dynamics of a given physical process and can be calculated by taking into consideration all possible Feynmann diagrams for the process, including all higher order correction diagrams up to a certain point in pQCD. The phase space element is the kinematic restrictions implied to the problem by the particles' initial and final state 4-momenta. For a $2 \rightarrow N$ process, i.e. scattering of two initial particles that produce N particles in the final state, the differential cross section is given by the formula [7]:

$$d\sigma = \frac{1}{2E_{\mathcal{A}}2E_{\mathcal{B}}|v_{\mathcal{A}} - v_{\mathcal{B}}|} \left(\prod_f \frac{d^3p_f}{(2\pi)^3} \frac{1}{2E_f} \right) \times |\mathcal{M}(p_{\mathcal{A}}, p_{\mathcal{B}} \rightarrow \{p_f\})|^2 (2\pi)^4 \delta^{(4)} \left(p_{\mathcal{A}} + p_{\mathcal{B}} - \sum p_f \right) \quad (4.4)$$

where $E_{\mathcal{A}}, p_{\mathcal{A}}$ and $E_{\mathcal{B}}, p_{\mathcal{B}}$ are the 4-momenta of the initial state colliding beams, $|v_{\mathcal{A}} - v_{\mathcal{B}}|$ the relative velocity of the beams as viewed from the laboratory frame, E_f, p_f the 4-momenta of the final state particles, and δ the Kronecker delta function. Note that the differential cross section $d\sigma$ given by Eq.4.4 is calculated at parton¹ level. This restriction comes from the applicability reach of perturbation theory; beyond parton level, non-perturbative effects need to be incorporated in the previous calculation through additional calculus steps.

Particles are often found in unstable bound states, and if so, they decay into another final state, producing more particles that may or may not further decay. The total mass of the final state particles has to, off course, be less than the mass of the initial state particle due to mass conservation. The *lifetime*² τ of a particle is then defined as the inverse of its decay rate. One particle can have multiple decay modes, each with a different probability. In this case, all possible decays have to be taken into account. The formula to obtain the differential decay rate is very similar to that of the differential cross section just above:

$$d\Gamma = \frac{1}{2m_{\mathcal{A}}} \left(\prod_f \frac{d^3p_f}{(2\pi)^3} \frac{1}{2E_f} \right) \times |\mathcal{M}(m_{\mathcal{A}} \rightarrow \{p_f\})|^2 (2\pi)^4 \delta^{(4)} \left(p_{\mathcal{A}} - \sum p_f \right) \quad (4.5)$$

where the discussion now holds for a single particle of mass $m_{\mathcal{A}}$ in the initial state, that decays into one or more particles in the final state with E_f, p_f . The amplitude \mathcal{M} remains relevant in the equation.

4.3.3 Renormalization

It was mentioned that for the evaluation of the amplitude \mathcal{M} , all process-specific diagrams up to a certain degree in pQCD should be considered. But what is the exact stopping point in the calculation? This is directly related to the level of precision that is required in each case. You see, the simplest form of Feynmann diagrams are such that only contain the raw process, like the example given in Fig. 4.5b. This is just the lowest possible order diagram that can be drawn for the $e^-e^+ \rightarrow \mu^-\mu^+$ process and is considered to be a diagram at *tree-level* or *tree diagram*. Calculation of the ME at this level in pQCD is said to be at first order or Leading Order (LO). More complex diagrams can be formed that result in the same final state but include more realistic scenarios on how to achieve this case. The method here is that in the previous diagram (Fig. 4.5b) closed loops can be

¹Partons are said to be the constituents of the proton, the quarks and gluons.

²Note that the half-life of a particle is defined as $\tau \cdot \ln 2$.

added in various places that represent extra corrections. These are called *loop diagrams* and contribute to the precision of the calculation, with each loop added increasing the order by one. Including one loop results in a second order calculation or Next-to-Leading Order (NLO), two loops in third order or Next-to-Next-to-Leading Order (NNLO) and so on. Two such examples, of one loop diagrams for the $e^-e^+ \rightarrow \mu^-\mu^+$ process, are shown in Fig. 4.6a and 4.6b. Another way that the diagrams can be varied so that the precision of the calculation increases is by adding additional legs (particles) in the final state, as done in Fig. 4.6c. This approach, again, increments the order of the correction by one for each additional leg.

But the amount of possible diagrams increases exponentially with the number of loops included, and there can be an infinite number of loops. Analogously, the complexity of the calculation also increases. This is where the cut-off point comes in play, for two reasons. Firstly, making such a calculation, to infinite precision, is on its own impossible. Secondly, as higher order terms are included (i.e. beyond LO) infinities arise in the calculations quite early. The latter infinities are caused because of the closed loops, as the momentum within the loop is not fully constrained and can range to infinity. Such cases are called *ultraviolet divergences* and are avoided through the renormalization process. Under this treatment a regulator is introduced that helps absorb the infinities which is called *renormalization scale* μ_r . To be exact, in pQCD any observable \mathcal{X} can be expressed as an expansion in powers of the strong coupling constant α_S given that $\alpha_S \ll 1$ [3]:

$$\mathcal{X} = c_0 + c_1\alpha_S + c_2\alpha_S^2 + \dots = \sum_{i=0}^n c_i \cdot \alpha_S^i \quad (4.6)$$

where c_i are the perturbative coefficients that emerge in the theory as a price for the renormalization, calculated using the Feynmann diagrams.

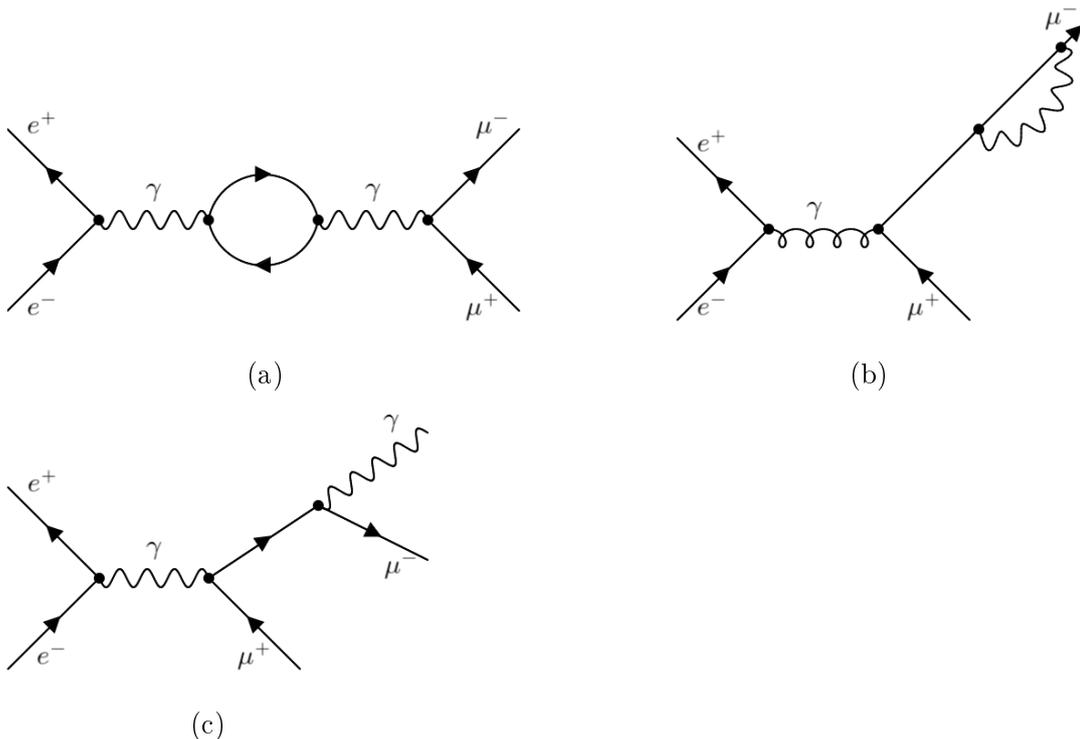


Figure 4.6: One loop diagram examples for $e^-e^+ \rightarrow \mu^-\mu^+$ (inspiration taken from [10]).

In fact, apart from the amplitude \mathcal{M} , other theory parameters can be expressed as a function of μ_r , like α_S or the fermion masses, so as to avoid the infinities by extracting the finite part. This was actually done in a previous section, Section 4.2.3, for α_S . Equation 4.1 is only but a solution to the *Renormalization Group Equations* (RGE) which govern the exact dependence of the strong coupling to the renormalization scale. The RGE for the strong coupling constant looks something like this:

$$\mu_r^2 \frac{\partial \alpha_S(\mu_r^2)}{\partial \mu_r^2} = \beta(\alpha_S(\mu_r^2)) \quad (4.7)$$

where $\beta(\alpha_S) = -\alpha_S^2 \cdot (b_0 + b_1 \alpha_S + b_2 \alpha_S^2 + \mathcal{O}(\alpha_S^3))$ the β function of QCD, expanded according to Eq. 4.6. The coefficients are renormalization-scheme dependent and b_0 is the one shared already in Eq. 4.2 according to the *modified minimal subtraction* $\overline{\text{MS}}$ scheme. Equation 4.1 then is acquired by considering only the first term b_0 [3].

In the end, the renormalization scale μ_r is the scale at which physical quantities from our theory are redefined directly after removing the ultraviolet divergencies that arise in their calculations. The renormalized version of the physical quantities, stripped from their infinities, hopefully remain with a meaningful and finite part. As it will be later seen, usually μ_r is taken to be at the same scale as the one of the physical process under investigation.

4.4 Cross section predictions for pp collisions

The subject of this dissertation centers on differential cross sections, prompting a more detailed exploration of this topic. Recall that the differential cross section for a $2 \rightarrow N$ process (two partons in the initial state) is expressed by Eq. 4.4. Shifting to the experimental perspective, particularly at the LHC, such a direct connection cannot be established. Since protons, rather than individual partons, are accelerated for collision, the cross section formula requires modifications. These adjustments are necessary to define a quantity that can be consistently measured and compared between theory and experiment.

4.4.1 Initial state adjustments

Realistically, as already mentioned numerous times, partons are not observed free in nature, but confined within hadrons. Therefore, the initial state has to be modified in order to describe protons. This treatment is performed by the *factorization theorem* [11] which rewrites the cross section definition as:

$$d\sigma_{(pp \rightarrow N)} = \sum_{i,j} \int dx dx' f_{i/p}(x, \mu_f) \cdot f_{j/p}(x', \mu_f) \times d\sigma_{(ij \rightarrow N)}(x, x', \mu_f, \mu_r, \alpha_S(\mu_r)) \quad (4.8)$$

The calculation has now been split into two separate steps. There is the “standard” perturbative part denoted as $d\sigma_{(ij \rightarrow N)}$ which corresponds to Eq. 4.4, except that a summation is performed over all possible contributing initial state quarks and gluons i, j . And there is a non-perturbative contribution related to the colliding protons denoted with the $f_{i,j/p}$ terms.

That’s right, at the initial state, as protons are accelerated before collision into beams, the energy scale Q of the process is low, resulting in $\alpha_S \gg 1$. As such, perturbative calculations are rendered useless. The change to the non-perturbative domain is justified as the total energy from the acceleration is distributed to all of the proton’s constituents (valence quarks and quark-gluon sea). In contrast to the moment right after collision,

where the proton energy is channeled into a single valence quark or gluon, armed for the hard collision.

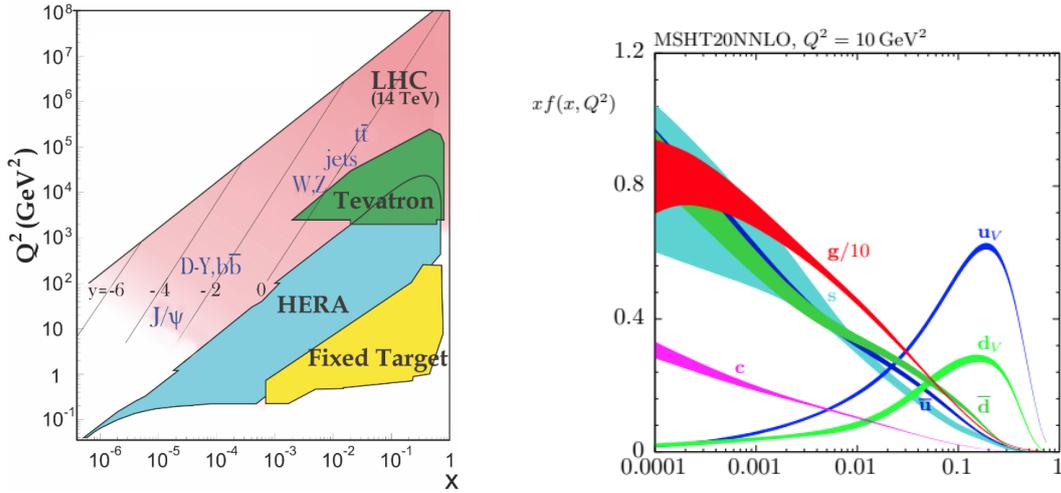
The question then remains on what exactly are the $f_{i,j/p}$ terms? These are the *Parton Distribution Functions* or PDFs [12] in short, $f_{i/h}(x, \mu_f)$, and parametrize the probability density to find a parton i , with momentum fraction $x_i = \vec{p}_i/\vec{p}_h$, inside a hadron h (in our case inside the proton), traveling at the same direction as the hadron. It makes sense to have two PDFs, one for each of the two incoming protons. Both the PDFs and the partonic cross section depend on a new scale, the *factorization scale* μ_f . The introduction of this scale is necessary for the treatment of *infrared divergencies* in the non-perturbative, long-distance physics. When the energy of a massless particle, like the gluon, becomes very small or when the gluon is emitted in very small angles, parallel to the parent particle, infinities arise that make the calculations unsafe. The two previous processes are referred to as *soft* and *collinear emissions* respectively, and can cause problems similar to those discussed for the ultraviolet divergencies. The factorization scale is defined such that these spurious effects get separated from entering the pQCD calculations, so that long-distance, non-perturbative effects are absorbed in the PDFs while short-distance, perturbative effects are absorbed in the partonic cross section. The evolution of PDFs according to the scale μ_F is described by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [13–16], just like RGE equations describe the running of the strong coupling constant.

Contrary to the matrix element part which is evaluated from theory, PDFs are calculated from the experiment. Specifically, they are extracted from global fits (Part III is dedicated to such a study) to data obtained from a large range of experiments and a variety of processes under study. To their advantage, PDFs are universal, in the way that once determined they can be used for any calculation. Exploiting data at different energy scales Q ensures their reliability across the whole phase space. Figure 4.7a shows the accessible phase space in the $x - Q^2$ plane for different experiments. Dedicated groups are involved in this process, the process of determining the PDF sets, with the goal of further constraining the existing version of PDFs by including newly obtained data. To name a few, there is the Coordinated Theoretical-Experimental Project on QCD (CTEQ), the Neural Network PDF (NNPDF), the Martin-Stirling-Thorne-Watt (MSTW), the HERAPDF group and others. As an example, the MSHT20 PDFs are shown in Fig. 4.7b at NNLO accuracy in pQCD. It is observed that there is a higher probability for a gluon to appear from the hard process in the low momentum fraction region, while for the up and down valence quarks this probability peaks around the 0.2 value. In other words, during a hard scattering process, valence quarks carry a significant fraction of the proton's momentum, as opposed to the gluon, whose contribution becomes dominant in lower momentum transfers.

Before we move on, it should be mentioned that both calculations required for the total for Eq. 4.8, the amplitude and the PDFs, should be performed at the same order in pQCD. The standard practice is that once the ME is determined in a given order in pQCD, it is then convoluted with a PDF set of the same order to obtain the final cross section.

4.4.2 Final state adjustments

Regrettably, the story does not end here. So far the final state for the $pp \rightarrow N$ process has been deliberately neglected. The high energetic partons produced from the hard process (hard scattering between the two protons) will eventually lead to the formation of colorless singlet states, i.e. to numerous hadrons. These are the particles that actually impinge on the detector, and whose energy and momentum is collected. The formation of hadrons is referred to as the *hadronization process*, which is a long-distance and therefore a non-perturbative effect that has to be incorporated in the cross section formula of Eq 4.8 nonetheless. Situation is treated similarly as to what was done for the NP effects in the



(a) Parton phase space available as a function of the momentum fraction x and the energy scale Q^2 [1]. Shaded areas in different colors illustrate the phase space coverage for different experiments, fixed target (yellow), HERA (cyan), Tevatron (green), and LHC (pink). Some final states are depicted for the LHC to the appropriate regions and as a function of rapidity y .

(b) MSHT20 NNLO PDF sets at $Q^2 = 10 \text{ GeV}^2$ as a function of the momentum fraction x [17]. The gluon (red), up valence quark (blue), down valence quark (green), strange quark (cyan), charm quark (magenta), up antiquark (dark blue), and down antiquark (dark green).

Figure 4.7: Signature figures on global PDF fits.

initial state. A new function is introduced that resembles the characteristics of PDFs, this is the *fragmentation function* (FF) $D_{k \rightarrow h}(z, \mu_F)$ which expresses the probability of finding a hadron h within the fragmentation products of an initial parton k , carrying a momentum fraction z of the parton's momentum. They are experimentally estimated and by using them, Eq. 4.8 is rewritten as [3]:

$$\begin{aligned}
 d\sigma_{(pp \rightarrow N)} = \sum_{i,j,k} \int dx dx' dz f_{i/p}(x, \mu_f) \cdot f_{j/p}(x', \mu_f) \times d\sigma_{(ij \rightarrow k)}(x, x', z, \mu_f, \mu_r, \alpha_S(\mu_r)) \\
 \times D_{k \rightarrow N}(z, \mu_F)
 \end{aligned}
 \tag{4.9}$$

where the *fragmentation scale* μ_F (not to be confused with the factorization scale which is denoted with the small letter f , μ_f) is incorporated into the calculation along with the definition of the FF. Likewise, the μ_F scale defines the transition point above which the long-distance effects of the hadronization process will be described by the FF and below which perturbation theory holds.

The reformed cross section definition presented just above is only one of the possible ways that the treatment of the final state can be handled. An approach like this is usually adapted by analyses that are very sensitive to the particle content of the final state. As a simplification, precise identification and measurement of the particles' four-momentum, which is a time consuming procedure, can be avoided by the definition of more complex objects. The most commonly used objects in high energy physics analyses is that of *particle jets*. These are also the main objects related to the work presented in this dissertation. Jets, in short, are defined as a collection of closely related particles, usually contained within a cone like structure, that travel together towards the same direction, from the collision

point, to the detector. The advantage of this approach is that the cross section can still be given by Eq. 4.8 without the need of FFs to describe the final state. Nevertheless, additional corrections are still necessary for such prediction that model the NP effects. These can be calculated using Monte Carlo techniques that parametrize the physical aspects of the long-distance effects, and will be later discussed.

4.5 Jets

4.5.1 Definition

In this section, we enter the experimental territory. Figure 4.8 illustrates the evolution of a jet, moving from the parton-level to its detection. Starting from the bottom, proton beams are accelerated close to the speed of light and collide at the interaction point within a detector, such as CMS. When two protons undergo a head-on collision, also called a *hard process*, they do not interact as whole objects. Instead, the actual collision occurs between individual partons, quarks or gluons, one from each proton. As a result, at least two energetic partons are produced in the scattering process, though additional partons may be generated, as shown in the figure.

At this stage, the interaction occurs at extremely short distances ($< 10^{-18} m$) and at a large energy scale, where perturbative QCD is valid ($\alpha_S \ll 1$). Due to asymptotic freedom, the partons are instantaneously free and carry the full energy of the collision. However, they subsequently radiate additional partons, a process known as *parton shower* (PS). This perturbative cascade continues until the energy of the emitted partons drops below a certain threshold, where α_S becomes large ($\alpha_S \gg 1$), and the system transitions into the non-perturbative regime.

At this point, the partons can no longer exist as free particles, and they undergo *hadronization*, forming color-neutral bound states such as pions, protons, and neutrons. This effect takes place at distances approximately $10^{-15} m$. The hadrons produced in this way travel outwards from the collision point until they interact with the detector. Looking at the entire process, the hadrons are emitted in collimated streams, roughly aligned with the original directions of the scattered partons. These structures, known as *jets*, provide an experimental signature of the underlying parton dynamics in high energy collisions (a nice overview on jets can be found in Ref. [18]).

There are three notable sources that contribute to high energetic parton production which can in turn evolve into jets. Firstly, there is the hard collision between protons, as mentioned above. Secondly, the hadronic decay of a heavy particle e.g. $t \rightarrow Wb$. Finally, the gluon emission from another parton, for example during the parton shower process. In reality, in the pp collisions observed at CERN, most of the time at least two of these

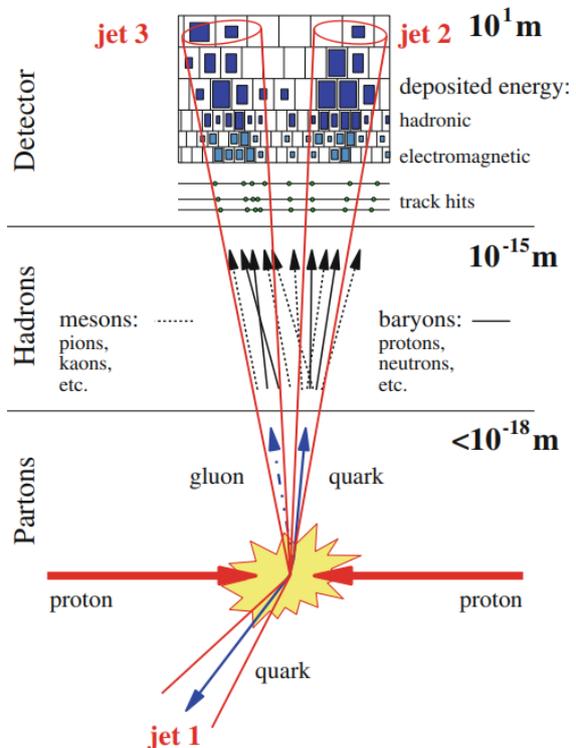


Figure 4.8: Jet evolution, from collision to detection [3].

processes contribute to jet formation in the final state, justifying the high jet multiplicity detected in the actual collision events.

4.5.2 From detection to reconstruction

Particle flow algorithm

No unique way exists to define a jet. Instead, over the years, scientists have developed algorithms that perform the tasks of jet detection and reconstruction. Such algorithms are based on mathematical prescriptions that can be varied and adapted depending on analysis needs. Nonetheless, for such algorithms to function, some knowledge of the particle content in the collision events is needed. Therefore, we now return the discussion to the detector.

As it has already been extensively discussed particles that emerge from the hard collision eventually interact with the detector, depositing their energy throughout its volume. A graph illustration for this effect is presented in Fig. 4.9 for a variety of particles. Each particle is shown to traverse horizontally different layers of the CMS subdetectors, starting with the silicon tracker, positioned at about 2.9 cm from the interaction point, all the way out to the muon chambers. The first objective towards jet reconstruction is efficient and accurate particle identification. This task is performed by the *particle flow* (PF) algorithm [20], a task usually referred to as global event reconstruction.

A typical workflow for the PF algorithm goes as follows (it is encouraged to have in mind Fig. 4.9). Firstly, energy clusters in the calorimeters are reconstructed. Then track reconstruction begins for all particles. Charged particles are expected to have a bend trajectory as opposed to neutral ones, which travel in a straight line. Reconstructed tracks are linked to energy clusters. From the topology of the PF elements (e.g. energy clusters and tracks), particle identification is achieved to a satisfactory level. As shown in the figure, different particle families are expected to leave their trace in specific regions of the detector. Muon identification follows shortly after. In the end, the PF algorithm delivers a comprehensive list of all identified (reconstructed) final state particles paired with some key information, like their momentum four-vectors and collision to detector angles.

Jet clustering algorithm

Once the list of reconstructed particles has been successfully obtained for an event it is fed as an input to the next algorithm in line, responsible for the jet reconstruction. Nowadays, selecting the appropriate clustering algorithm is typically straightforward. However, this was not always the case, especially in the past where expertise in the field was still developing. For this reason, before we dive into the specifics of the algorithm used in this work, we start by giving a brief overview of the topic.

Over the years, after a lot of algorithmic variations, a general set of requirements has been established for the desirable properties of a jet reconstruction algorithm [3]:

- **General properties:**

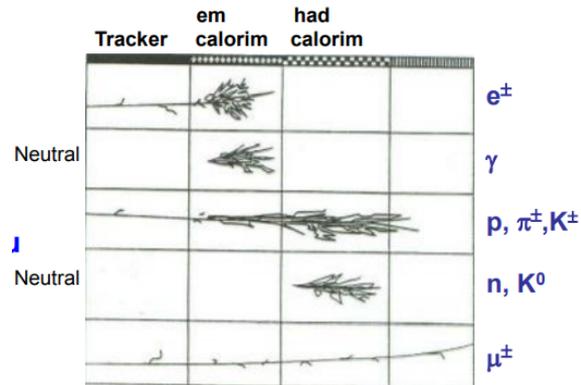


Figure 4.9: Particle energy depositions to CMS detector [19].

1. *Order dependence*, equal applicable to partons, particles, or measured tracks and depositions.
 2. *Simple implementation*, avoidance of complex code developments.
- **Theoretical properties:**
 1. *Collinear and infrared safe*, for finite pQCD calculations.
 2. *Longitudinal boost invariance*, for jet observables.
 3. *Boundary stability*, insensitivity of jet kinematic boundaries from details of the hadronic final state.
 4. *Insensitivity to non-perturbative effects*, limitation of impact of hadronization and the underlying event.
 - **Experimental properties:**
 1. *Detector independence*, no dependence on detector details.
 2. *Computational efficiency and predictability*, predictable computing times that mildly increase with growing numbers and input objects.
 3. *Maximal reconstruction efficiency*, lossless treatment of all input objects.
 4. *Insensitivity to pile-up collisions*, accurate correction for additional energy not coming from the primary vertex.
 5. *Ease of calibration*, accurate and straightforward estimation of diverse detector effects on jet response.
 6. *Minimal resolution smearing and angular biasing*, avoidance of algorithmic distortions in addition to detector effects.

From this extensive list of requirements perhaps the most crucial ones are those of collinear and infrared safety, and of detector independence (the first points from both theoretical and experimental desired properties). In more detail, compliance to these conditions is necessary, as emission of a particle to small angles (collinear splitting), can lead to the disappearance of a jet. Similarly, emission of a soft gluon can lead to the merging of two jets into one. Finally, detector independence ensures that comparisons can be achieved between different experiments.

The algorithm used for this work complies with all of the above requirements and belongs to the broad category³ of *sequential-recombination algorithms* that perform iterative comparisons between the closest pair of objects until completion. The stages to this iterative procedure are presented below:

1. Given an input list of N particles, also referred to as *PF candidates*, define the relative distance between each possible pair of particles d_{ij} and the distance between each object and the beam direction d_{iB} according to

$$d_{ij} = \min \left(p_{T,i}^{2p}, p_{T,j}^{2p} \right) \frac{\Delta R_{ij}^2}{R^2} \quad (4.10)$$

$$d_{iB} = p_{T,i}^{2p} \quad (4.11)$$

$$\text{with } (\Delta R_{ij})^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \quad (4.12)$$

where ΔR_{ij} the angular distance between the objects i and j .

³Another category is the *cone algorithms* that groups particles based on geometrical criteria. The grouping is performed in a well defined cone radius around high-energy seed particles. These algorithms were popular in the past but were found to be prone to collinear and infrared unsafety.

2. Deduce whether the minimum distance d_{min} originates from the $\{d_{ij}\}$ set of values or from the $\{d_{iB}\}$ ones.
3. If $d_{min} \in \{d_{ij}\}$, it is assumed that objects i and j are spatially close to each other and are merged into a new object (a protojet) by summing their momentum four-vectors. The new object is added to the PF list, whilst the original objects i and j are removed from it.
4. If $d_{min} \in \{d_{iB}\}$, it is assumed that no other objects are spatially close to object i and is labeled as a final jet. This object is removed from the PF candidates list and is added to the final list of K jets, which at the start of the procedure is empty.
5. Repeat from step 1 for the updated PF candidate list until no more objects are left, in which case the sequential procedure comes to a halt.

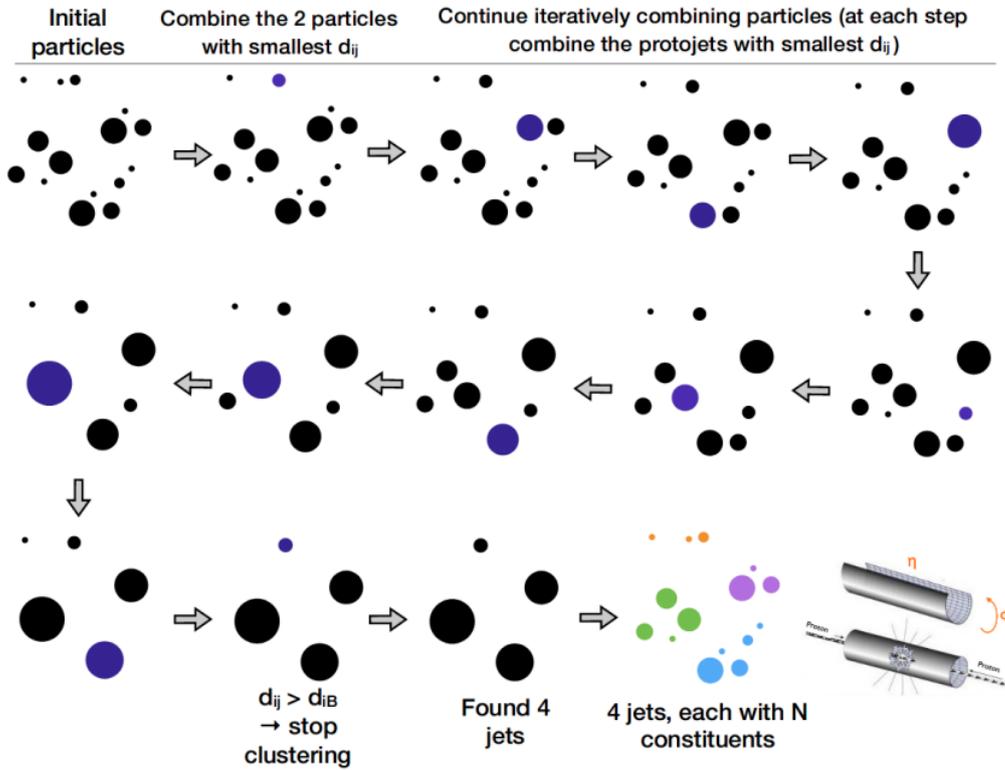


Figure 4.10: Sequential-recombination algorithm at work [21].

Figure 4.10 provides a great illustration of the sequential-recombination algorithm described above. From an initial list of PF candidates (top left corner figure), the objects are gradually merged into protojets until no extra iterations can be performed. In the second to last, and third to last pictures it is seen that the algorithm has deduced to a final state with four jets.

Looking at Eq. 4.10 the power p can be selected such that it changes the behavior of the iteration. There are three possibilities to choose from, each corresponding to an alternative algorithm. For $p = 1$ there is the k_T [22], $p = 0$ the Cambridge/Aachen [23, 24], and $p = -1$ the anti- k_T [25] algorithm. Breaking apart each case, we can see that when $p = 1$ the d_{ij} distance is proportional to the transverse momentum, thus priority to clustering is given to lower- p_T objects and leads to larger and non-uniformed jet shapes. Taking $p = -1$ on the other hand, makes the equation proportional to $1/p_T$, meaning that high- p_T objects are prioritized in the clustering and merging is performed uniformly around them,

resulting to a consistent cone-like jet shape. Lastly, $p = 0$ leaves the d_{ij} dependence solely on angular quantities, thus the clustering is performed according to the angular distance, which again results to larger and non-uniform shapes, just like for $p = 1$. Figure 4.11 shows a comparison between the jet area topologies that result from each jet clustering algorithm for $R = 1$. Surely, the anti- k_T algorithm ($p = -1$) provides circular jet shapes centered around the highest p_T object in each area.

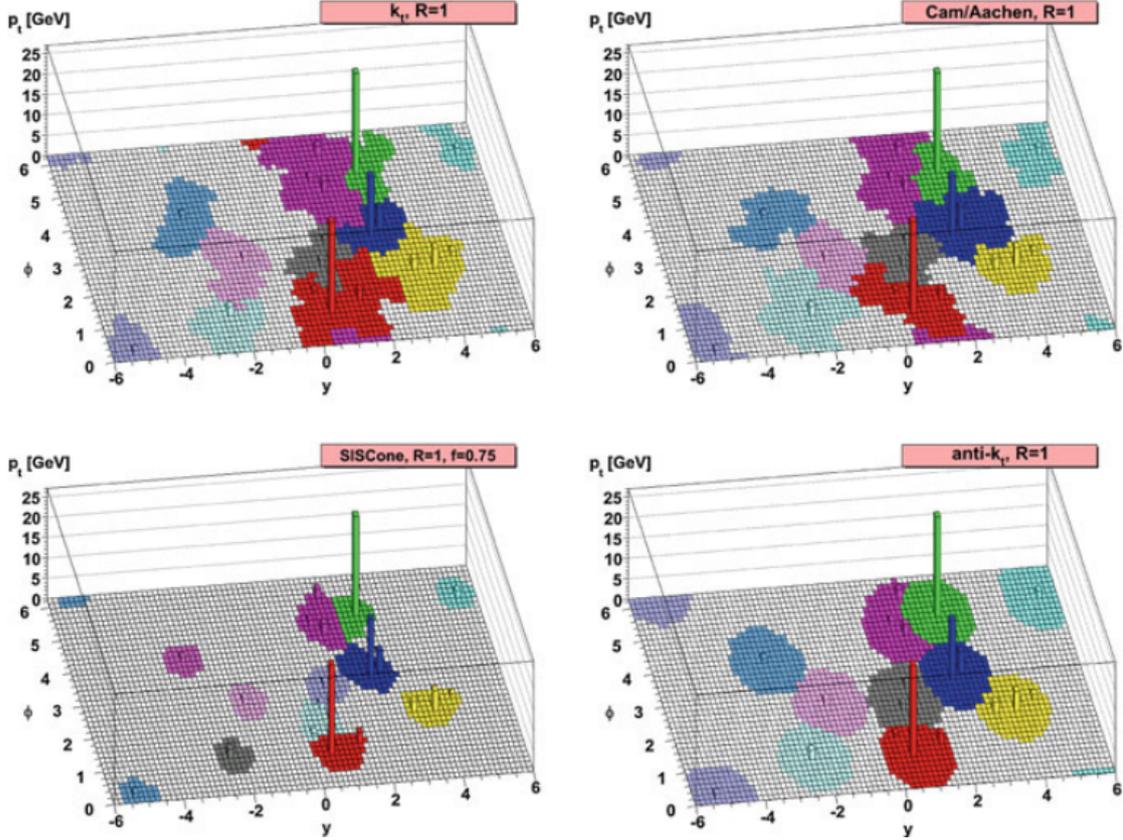


Figure 4.11: Jet area for the k_T , Cambridge/Aachen, SIScone⁴, and anti- k_T jet clustering algorithms [3].

Needless to say, the jet topologies in this thesis have been reconstructed using the anti- k_T clustering algorithm. By prioritizing the clustering for high- p_T objects, the algorithm ensures that most pairwise mergings include at least one hard particle. Moreover, since calibration corrections often depend on the jet area, the well defined jet shape by anti- k_T offers distinct advantage over other algorithms. Jet reconstruction at CMS is achieved via the Jet Toolbox⁵ [27] interface which utilizes the FASTJET [28] package, a tool that supports various clustering algorithms including the anti- k_T .

Jet radius

This brings us to the next topic of discussion the *jet* or *cone radius* R as seen in Eq. 4.10. By definition, R carries no physical units and is dimensionless, with typical values $0.4 \leq R \leq 1.0$. Traditionally, in LHC Run I ATLAS had been using $R = 0.4$ and 0.6 jets, while CMS $R = 0.5$ and 0.7 . From Run II and onward CMS has adapted to 0.4 , for the smaller jet radius, instead of 0.5 , to enable comparison of the results between the two experiments.

⁴A *seedless infrared-safe cone* (SIScone) algorithm [26]

⁵CMS Jet Toolbox GitHub page: <https://github.com/cms-jet/JetToolbox>

The larger radius has been rather increased for both experiments with ATLAS changing it to 1.0 and CMS to 0.8.

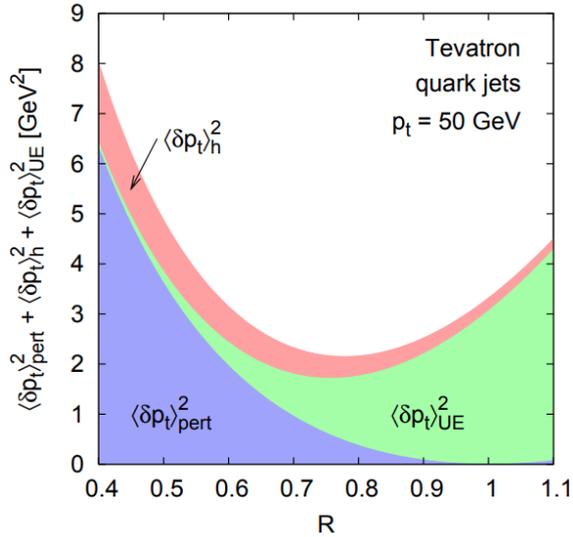


Figure 4.12: Squared average shift in jet p_T caused by perturbative radiation (pert), hadronization (h), and underlying event (UE) [29].

through the notation *AK8 jets*.

Jet types

There are three main jet categories, namely *PFJets*, *CaloJets*, and *GenJets*. Depending on the input given to the Jet Toolbox, and by extension to the FASTJET package, jets can be obtained at different levels.

- *PFJets* are the bread and butter. These are the standard, typical, jets used in the majority of the CMS analyses, constructed by the full list of input particles as provided by the PF algorithm. They effectively represent the jets observed in the experiment.
- *CaloJets* can be thought as a minimalistic version of the PFJets. They are the result of a fast reconstruction, usually performed at the HLT level, to make a decision on the event accept/reject signal. Other times they are used as an intermediate step to calibrate the actual jets, the PFJets. For these reasons, CaloJets are obtained by using only a small fragment of the global event information, just from the summation of the energy depositions in the ECAL and HCAL calorimeter towers.
- *GenJets* are the product of simulation, thus they do not represent real data. Event generators are often exploited to simulate the collision conditions. Given a list of generated particles one can obtain jets at the generator level, the GenJets.

A fourth, hidden, category emerges as an aftermath of GenJets. As mentioned above, the latter, are the pure product of the generated final state particles of a hypothetical collision. Situation requires to gain access to the same information at the level of the detector.

- *RecJets* are the counterparts to GenJets, representing the same information but after undergoing the additional step of detector simulation. This process introduces

The choice of the radius matters. Studies at the Tevatron experiment [29] showed that the measured jet p_T can vary with the radius R . Figure 4.12 summarizes the results with respect to three different physical phenomena, the perturbative radiation (blue), the hadronization (red), and the soft particle production from the underlying event (green). The shift caused by perturbative radiation seems to disappear for $R \sim 1$, but there the shift of the underlying event starts to become dominant. A sweet spot seems to appear between the 0.7 to 0.8 values but the hadronization process appears constrained. There is no best choice, as this is based on the measurement at hand. One can always choose to be (in)sensitive to a given effect. For the analysis presented here (analysis in Part II) the larger cone size jets are used, for $R = 0.8$. A concise way to convey that the anti- k_T algorithm is used with a radius of 0.8 is

detector effects, such as the finite resolution, which impact the reconstructed jets. As a result, RecJets provide a realistic representation of jets measured at detector level (as do the PFJets).

Pile up

Jets can be further classified into various subflavors. An interesting aspect that needs to be considered is the choice of mitigation technique. So far, we have deliberately avoided discussing the possibility of simultaneous proton collisions. When two proton bunches cross, multiple proton-proton interactions can occur within the same event that influence the jet reconstruction process. This effect goes under the name of *pile up*, it is encountered in every pp collision at the LHC, and it increases with the luminosity \mathcal{L} .

The usual approach to address this phenomenon is to first define the *Primary Vertex* (PV), i.e. the hardest, most energetic, pp collision in the event. All other collisions are labeled as *pile up vertices* (pile up collisions). Firstly, tracks from all particles are extrapolated backwards, from the detector to the interaction point, to locate their point of origin. This way all vertices are identified, and ordered according to the quadratic sum of the particles' p_T . The vertex at the top of the list is labeled as the PV. Jets associated with the PV offer the most potential for analysis purposes.

Pile up mitigation algorithms have been designed to remove jets related to pile up vertices that contaminate the collision event. Additionally, filtering is performed also to particles originating from pile up vertices that end up inside PV jets, leading to an overestimation of their measured energy. An illustration of this effect is given in Fig. 4.13. Most importantly, in this example we can see in red, that a particle from the pile up vertex, ends up inside the jet from the main vertex. There are multiple sources of pile up, like cosmic ray radiation and the underlying event, to name a few, indicated with the orange arrows.

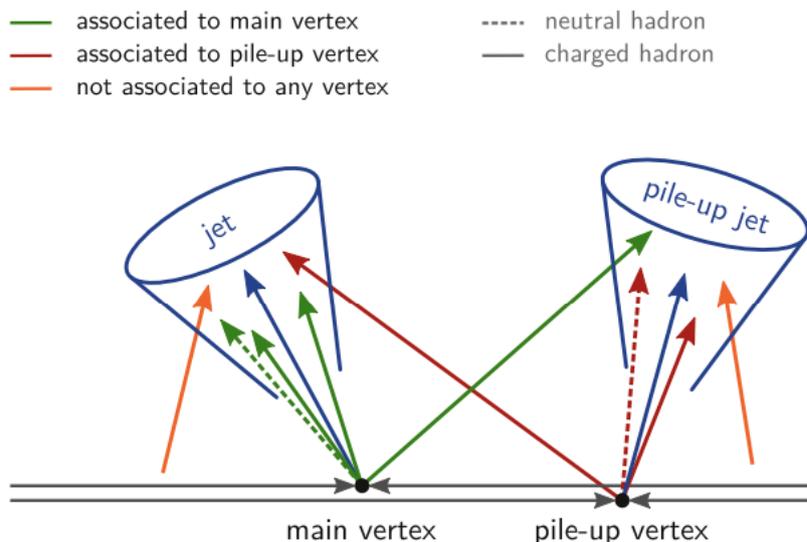


Figure 4.13: Dissociation between main and pile up collision vertices [3].

Available choices for pile up removal algorithms are *Soft Killer* (SK), *Constituent Subtraction* (CS), *Pile up Per Particle Identification* (PUPPI), and *Charge Hadron Subtraction* (CHS), with the latter two having the most popularity among them. We give the specifics for the CHS algorithm [30], the predominant choice for CMS analyses in Run II, like this one. In this approach, charged hadrons originating from pile up vertices are identified and removed from the jets associated to the PV. As for the neutral hadrons, an additional

amount of energy is subtracted on average from the jets under the assumption that energy depositions from neutral particles is uniformly distributed across the detector. So the drawback to this scenario is that single collisions are not optimally reconstructed [31]. This type of correction is applied directly to the PF candidates, before the list of particles is given as an input to the clustering algorithm. Revisiting the notation of AK8 jets, we can refine it to also indicate the jet type. Within the scope of this thesis the jets can be explicitly labeled as *AK8 PFchs jets*. A realistic view of a CMS pp collision event display with an average pile up of ~ 100 can be seen in Fig. 4.14. Orange points in the middle represent the vertices.

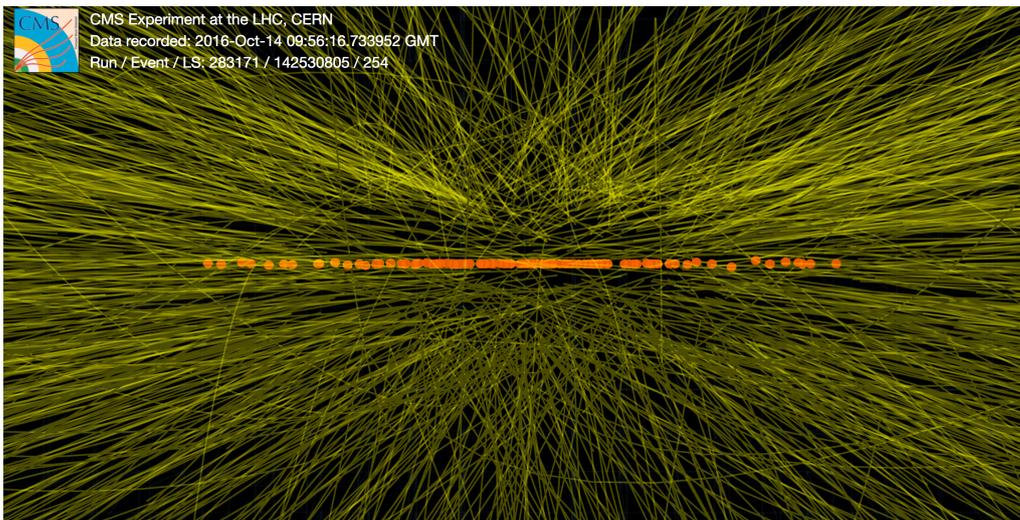


Figure 4.14: CMS pp collision event display with an average pile up of ~ 100 [32].

Particle content

We are ready to look some actual results. The top plot of Fig. 4.15 shows the PF energy fractions of fully reconstructed and corrected jets as a function of the tagged jet p_T . The study was performed on dijet events with the 2016 dataset collected by CMS for AK4 PFchs jets. The same results were obtained from simulation by Pythia8 and were overlaid to the ones from data. Experimental results are indicated by the different markers, while histograms with plain continuous lines display the results from simulation. It is observed that jet particle content is dominated by charged hadrons (brown and red bands). Another large contribution comes from photons (in blue). Combining the two, we can see that around 90% of the energy originates from charged hadrons and photons alone, with the remaining 10% mostly composed from neutral hadrons (in green). Leptons (in cyan) have the smallest contribution. Specifically, the jet particle content is occupied mostly by pions, either charged (π^\pm) or neutral (π^0), the latter, in the form of photon pairs. The importance of pile up mitigation algorithms is highlighted by the fact that charged hadrons constitute the majority of a jet's particle content, making precise energy measurement crucial.

Bottom plot in the same figure shows the ratio of data to simulation. The agreement between the two seems to be good across the whole p_T range and mostly varies within 0.02%.

The shape, and therefore, the particle content of a jet, can also change based on the parton that initiates the jet. The mathematical structure of $SU(3)$ introduces the *color factors* which describe the relative strength (the probability) of some QCD processes.

For this subject we are interested in three of them, the *Fundamental Casimir operator* (C_F), the *Adjoint Casimir operator* (C_A), and the *Normalization of the Fundamental representation* (T_F). C_F expresses the interaction strength for a quark to emit a gluon $C_F \equiv (N_C^2 - 1)/2N_C = 4/3$. Respectively, C_A , represents the probability for a gluon to emit a gluon $C_A \equiv N_C = 3$. By taking the ratio $C_A/C_F = 9/4$ one can deduce that gluon emissions happen more than twice as much for gluons than for quarks. Lastly, the quark-antiquark emission from a gluon is given by $T_F = 1/2$, so the ratio $C_A/T_F = 6$ translates to a six times higher probability for a gluon pair to be emitted from a gluon than a quark-antiquark pair. Furthermore, gluon-initiated jets appear wider compared to quark-initiated. Consequently, the former, usually appear with a higher particle multiplicity.

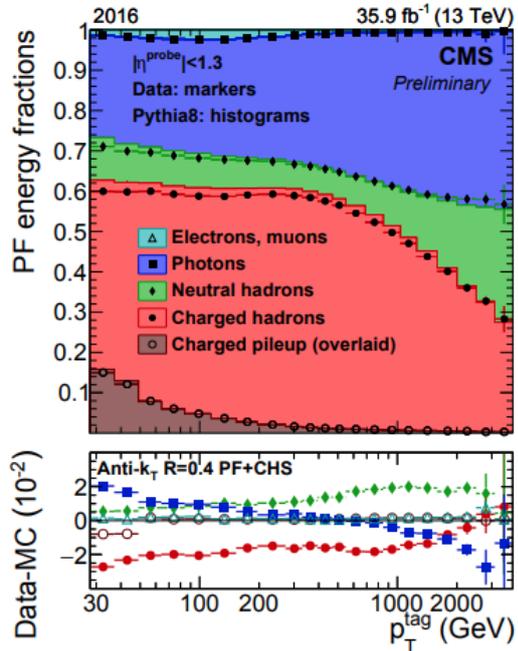


Figure 4.15: PF jet energy fractions for anti- k_T $R = 0.4$ PFchs jets from the 2016 CMS dataset and Pythia8 simulation. [33].

4.6 Monte Carlo event generators

Monte Carlo (MC) is a numerical technique that relies mostly on pseudo-random numbers, rather than truly random ones, and incorporates various algorithms and statistical methods to solve complex problems. Originally developed by mathematician Stanislaw Ulam [34], the method was inspired by the use of random numbers in casinos, particularly the Monte Carlo casino, hence the name. Its applications extend to numerous fields such as chemistry, statistics, finance, artificial intelligence, and off course physics.

In the context of High Energy Physics, MC techniques are employed through their application on event generators. Their use offers several advantages. They provide predictions, such that experimental data can be compared to. Conclusions can be drawn on whether the theory lacks on the modeling of some physical aspects. Through simulation they help on the construction of future experiments. Moreover, they provide corrections for non-perturbative and electroweak effects on fixed-order theory predictions. And many more.

Usually, the overall process is divided into two separate steps, the event generation and the detector simulation. Event generation amounts for the union of all the physical processes that result in a definite, explicit, final state of particles. If these particles are given as an input to a clustering algorithm, GenJets are constructed. In the follow-up stage, detector simulation is performed where the machine details are encoded in the input information giving rise to RecJets. Detector simulation in CMS is performed via the GEANT4 [35] package, a dedicated toolkit for simulating the interaction of particles with matter. In principle, if precise knowledge of detector effects could be obtained, a straightforward simulation could provide sufficient results, eliminating the need for actual experiments.

Before giving the details on the MCs used for the analysis in Part II, we first take a closer look on the physical aspects that contribute to the evolution of a high energy pp collision, that need to be considered during the event generation stage. Up to this point,

most of the individual processes discussed below have already been addressed to some extent. Here, we offer a more inclusive perspective.

4.6.1 Full event display

Figure 4.16 illustrates all physical processes involved in a pp collision that must be considered in an event generator simulation. Its deconstruction follows:

1. *Hard process*. At the core of a pp collision lies the hard process, which represents the single most energetic interaction in the event, characterized by the highest momentum transfer between partons. Its differential cross section is determined by Eq. 4.8, where pQCD provides the process amplitude, while the PDFs, evaluated at a factorization scale close to that of the hard interaction, describe the behavior of the incoming protons. The hard process, occurring at an energy scale of approximately $\mathcal{O}(1\text{TeV})$, is visually represented in the figure by the large red circle at the center. It leads to the production of four highly energetic partons - three quarks (depicted as smaller red circles) and a gluon (illustrated by a red swirling spring-like line).

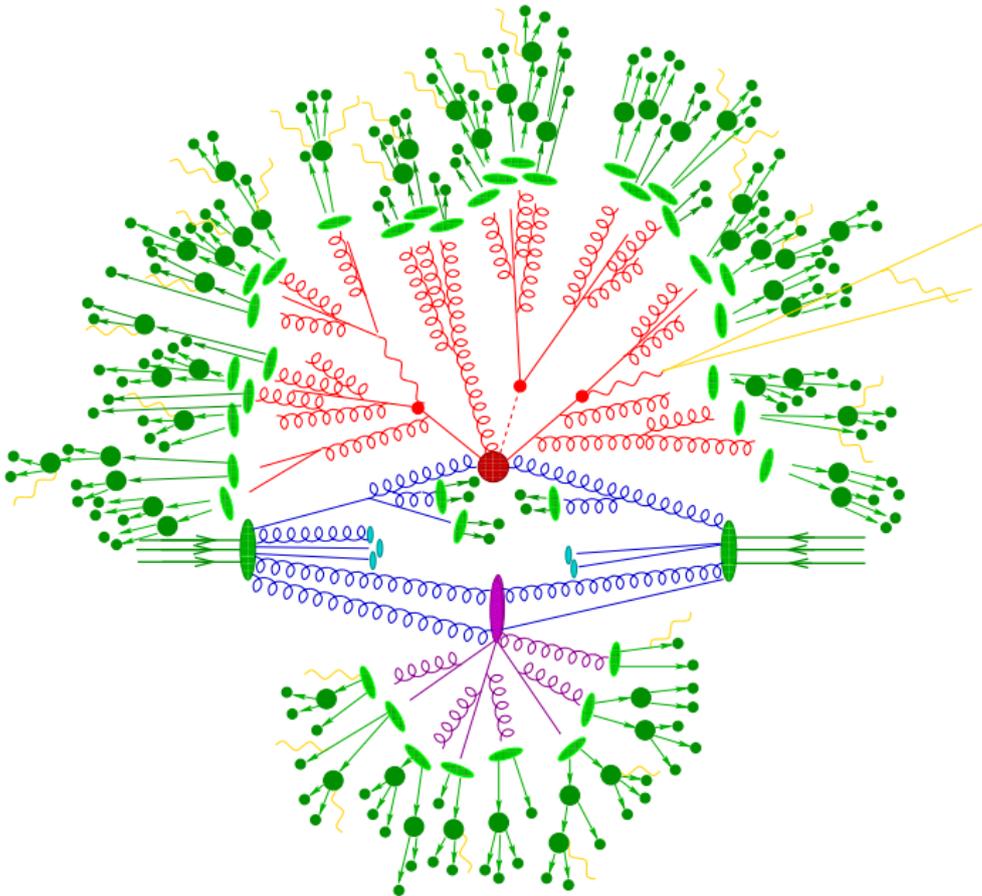


Figure 4.16: Simulation of a pp collision from an event generator point of view [36].

2. *Parton shower (PS)*. The high energetic partons that resulted from the hard interaction cause particle cascades by the emission of more quarks and gluons in a process known as parton shower. This process occurs within the $\mathcal{O}(1\text{TeV} - 1\text{GeV})$ energy scale and is calculable by pQCD. In simulations this effect is parametrized by algorithms in the form of a correction to the initial matrix element, known as

- leading-logarithm approximation (LLA)*. Different approaches are adapted depending on the generator, such as the *angular ordering* of successive emissions and the p_T or Q^2 ordering of PS. The PS is represented by the red-lined tree-like structure in the figure.
3. *Hadronization (HAD)*. When the energy scale falls below a certain threshold, $\mathcal{O}(< 1 \text{ GeV})$, one at which pQCD is not valid anymore, the formation of bound states will begin under the hadronization/fragmentation process. Again, this non-perturbative process is modeled within the generators. Two variations are popular, *Lund string fragmentation* and *cluster fragmentation*. The high complexity of these processes and the absence of a concrete definition requires the tuning of the algorithm’s parameters to real data. Hadronization effects are illustrated by the bright green ellipses.
 4. *Initial state radiation (ISR)*. Partons within protons may radiate before the hard interaction takes place. This additional emissions are modeled within the PS algorithms and are treated as further corrections to the pQCD calculation. Such emissions are seen with blue swirling lines at the center of the figure prior to the red circle.
 5. *Final state radiation (FSR)*. Similarly to ISR, final state radiation amounts for additional radiation caused by the particles in the final state. Bound states have a limited lifetime and a fraction of them may decay before they reach the detector material, resulting in further quark and gluon emissions. FSR is modeled within the PS algorithm and is perturbately calculable. FSR is illustrated by the dark green arrows and circles, and the yellow lines, that emerge from bright green ellipses.
 6. *Underlying event (UE)*. Once a parton from each proton undergoes a hard interaction, the remaining proton constituents continue to evolve, eventually forming bound states through the mechanisms described earlier. This evolution gives rise to multiple parton interactions (MPI) - softer secondary collisions that generate additional particles, radiation, and jets. All activity not directly associated with the primary hard interaction is collectively referred to as the underlying event (UE). In the figure, the UE is represented in the lower section, all processes surrounding the purple ellipse.
 7. *Pile up (PU)*. Pile up interactions (not included in the figure) are simulated separately and are based on minimum bias events (lower energy events). Only afterwards are combined with the simulated events that contain the hard process to better reflect the situation observed in actual data. PU is divided into two categories, the *in-time pile up (IT PU)*, which corresponds to additional pp collisions that originate from the same bunch crossing, and the *out-of-time pile up (OOT PU)*, which amounts for collisions originating from the adjacent bunch crossings, that are present due to the finite response time of the detector electronics.

4.6.2 A quiver full of MCs

MCs for this work are divided into two classes. On one hand, there are the official MC samples provided centrally by CMS for analyses, just like the data samples. Details on the official samples are shared in Section 5.2. On the other hand, “homemade” samples are generated using the Robust Independent Validation of Experiment and Theory toolkit or RIVET [37, 38] in short. The difference between the two is that, the former, includes a full simulation, providing the event content information in both the generator level via GenJets and the reconstructed level with RecJets. They are used during the offline analysis

to calibrate the data samples and perform comparisons. The latter, the rivet outputs, only contain the information at generator level, contain less statistics and are used to derive NP corrections for the fixed-order theory predictions (a procedure which is well documented in Section 6.2.1). Hopefully, the context itself will make the distinction clear, or will be explicitly stated otherwise.

In total, eight different MC configurations are used for the derivation of the NP corrections, listed in Table 4.2. Information on the order of the ME calculation, the PS and hadronization simulation models, and the tune choice are shared in individual columns for each event generator. For leading order ME calculation PYTHIA8 [39] (version 8.240) is employed with two different tunes, CUETP8M1 [40] and CUETP8M2T4 [41], and HERWIG++ [42] (version 2.7.1) with the EE5C [43] tune. These generators do not currently support calculations at NLO, for this reason, the POWHEG [44, 45] event generator is used which can reach the desired precision. The disadvantage here is that Powheg can not model the PS, hadronization, and MPI processes, but only provide the ME at NLO. Consequently, it is interfaced with the LO generators, to use their algorithms (and tunes) to simulate them. Lastly, a newer version of Herwig++, HERWIG7 [46], is used to obtain both LO and NLO predictions via the CH3 [47] tune.

Table 4.2: List of MC configurations along with the order of the matrix element, the model for the parton shower and the hadronization algorithms, and the tune choice.

MC	Matrix element	Parton shower	Hadronization	Tune
PYTHIA8	$2 \rightarrow 2$ (LO)	p_T ordering	Lund string fragmentation	CUETP8M1
PYTHIA8	$2 \rightarrow 2$ (LO)	p_T ordering	Lund string fragmentation	CUETP8M2T4
HERWIG++	$2 \rightarrow 2$ (LO)	Angular ordering	Cluster fragmentation	EE5C
HERWIG7	$2 \rightarrow 2$ (LO)	Angular ordering	Cluster fragmentation	CH3
POWHEG + PYTHIA8	$2 \rightarrow 2$ (NLO) $2 \rightarrow 3$ (LO)	PYTHIA8	PYTHIA8	CUETP8M1
POWHEG + PYTHIA8	$2 \rightarrow 2$ (NLO) $2 \rightarrow 3$ (LO)	PYTHIA8	PYTHIA8	CUETP8M2T4
POWHEG + HERWIG++	$2 \rightarrow 2$ (NLO) $2 \rightarrow 3$ (LO)	HERWIG++	HERWIG++	EE5C
HERWIG7	$2 \rightarrow 2$ (NLO)	Angular ordering	Cluster fragmentation	CH3

Bibliography

- [1] S. Navas et al. [Particle Data Group]. Review of Particle Physics. *Phys. Rev. D*, 110:030001, Aug 2024. <https://link.aps.org/doi/10.1103/PhysRevD.110.030001>.
- [2] K. E. Vagionakis. *Particle Physics, An Introduction to the Basic Structure of Matter*. NTUA, 2013.
- [3] K. Rabbertz. *Jet Physics at the LHC*. Springer Cham, 2017. <https://doi.org/10.1007/978-3-319-42115-5>.
- [4] A. Hayrapetyan et al. [CMS Collaboration]. Measurement of multijet azimuthal correlations and determination of the strong coupling in proton-proton collisions at $\sqrt{s} = 13 \text{ TeV}$. *Eur. Phys. J. C*, 84, 842, Aug 2024. <https://doi.org/10.1140/epjc/s10052-024-13116-7>.
- [5] P. P. Urone and R. Hinrichs. *College Physics*. OpenStax, 2012. <https://openstax.org/books/college-physics/pages/1-introduction-to-science-and-the-realm-of-physics-physical-quantities-and-units>.
- [6] R. P. Feynman. Space-Time Approach to Quantum Electrodynamics. *Phys. Rev.*, 76:769–789, Sep 1949. <https://link.aps.org/doi/10.1103/PhysRev.76.769>.
- [7] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.
- [8] S. Weinberg. *The Quantum Theory of Fields, Volume 1: Foundations*. Cambridge University Press, 1995.
- [9] A. Randle-Conde. Aidan’s website – Feynmann diagram maker. Available: <https://www.aidansean.com/feynman/>. Accessed: 2025-01-10.
- [10] P. L. S. Connor. Precision Measurement of the Inclusive b Jet Production in Proton-Proton Collisions with the CMS Experiment at the LHC at $\sqrt{s} = 13 \text{ TeV}$. PhD thesis, Hamburg (2018).
- [11] J. Collins. *Foundations of Perturbative QCD*, volume 32. Cambridge University Press, 2011. <https://inspirehep.net/literature/922696>.
- [12] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Parton distributions for the LHC. *Eur. Phys. J. C*, 63:189–285, 2009. <https://inspirehep.net/literature/810127>.
- [13] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.
- [14] L. N. Lipatov. The parton model and perturbation theory. *Yad. Fiz.*, 20:181–198, 1974.
- [15] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298–318, 1977. <https://www.sciencedirect.com/science/article/pii/0550321377903844>.
- [16] Y. L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

- [17] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. *The European Physical Journal C*, 81(4), Apr 2021. <http://dx.doi.org/10.1140/epjc/s10052-021-09057-0>.
- [18] Gavin P. Salam. Towards jetography. *The European Physical Journal C*, 67(3–4):637–686, May 2010. <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6>.
- [19] David Cockerill. Introduction to Calorimeters. Available: <https://indico.cern.ch/event/518474/contributions/1198681/attachments/1267581/1877122/Calorimetry-lecture-to-Southampton-students-4May2016-Cockerill-compressed.pdf>. Accessed: 2024-10-03.
- [20] A.M. Sirunyan et al. [CMS Collaboration]. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003, Oct 2017. <https://dx.doi.org/10.1088/1748-0221/12/10/P10003>.
- [21] P.L.S. Connor, Q. Wang, R. Zlebick. CMS Data Analysis School, Short exercise on Jets. Slides obtained from private conversation with one of the authors, Accessed: 2025-01-20.
- [22] S.D. Ellis and D.E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev. D*, 48:3160–3166, Oct 1993. <https://link.aps.org/doi/10.1103/PhysRevD.48.3160>.
- [23] Y.L. Dokshitzer and G.D. Leder and S. Moretti and B.R. Webber. Better jet clustering algorithms. *Journal of High Energy Physics*, 1997(08):001, Sep 1997. <https://dx.doi.org/10.1088/1126-6708/1997/08/001>.
- [24] M. Wobisch and T. Wengler. Hadronization Corrections to Jet Cross Sections in Deep-Inelastic Scattering, 1999. <https://arxiv.org/abs/hep-ph/9907280>.
- [25] M. Cacciari and G.P. Salam and G. Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, Apr 2008. <https://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [26] G.P. Salam and G. Soyez. A practical seedless infrared-safe cone jet algorithm. *Journal of High Energy Physics*, 2007(05):086, May 2007. <https://dx.doi.org/10.1088/1126-6708/2007/05/086>.
- [27] A. Gomez. Jet Toolbox. Available: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetToolbox>. Accessed: 2025-01-21.
- [28] M. Cacciari, and G.P. Salam, and G. Soyez. FastJet user manual. *The European Physical Journal C*, 72(3):1896, Mar 2012. <http://link.springer.com/10.1140/epjc/s10052-012-1896-2>.
- [29] M. Dasgupta, and L. Magnea, and G.P. Salam. Non-perturbative QCD effects in jets at hadron colliders. *Journal of High Energy Physics*, 2008(02):055, Feb 2008. <https://dx.doi.org/10.1088/1126-6708/2008/02/055>.
- [30] A.M. Sirunyan et al. [CMS Collaboration]. Pileup mitigation at CMS in 13 TeV data. *Journal of Instrumentation*, 15(09):P09018, Sep 2020. <https://dx.doi.org/10.1088/1748-0221/15/09/P09018>.

- [31] CMS Collaboration. How CMS weeds out particles that pileup. Available: <https://cms.cern/news/how-cms-weeds-out-particles-pile>. Accessed: 2025-01-22.
- [32] CMS Collaboration and T. Mc Cauley. Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill. <https://cds.cern.ch/record/2231915>, 2016.
- [33] CMS Collaboration. Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018, 2020. <https://cds.cern.ch/record/2715872>.
- [34] N. Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, 1949. <https://www.tandfonline.com/doi/abs/10.1080/01621459.1949.10483310>.
- [35] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003. <https://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [36] A. Schälicke et al. An event generator for particle production in high-energy collisions. *Progress in Particle and Nuclear Physics*, 53(1):329–338, 2004. <https://www.sciencedirect.com/science/article/pii/S014664100400047X>.
- [37] C. Bierlich et al. Robust Independent Validation of Experiment and Theory: Rivet version 3. *SciPost Phys.*, 8:026, 2020. <https://scipost.org/10.21468/SciPostPhys.8.2.026>.
- [38] Hosted by HEPForge. Rivet — the particle-physics MC analysis toolkit. Available: <https://rivet.hepforge.org/>. Accessed: 2025-01-27.
- [39] T. Sjöstrand et al. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, 2015. <https://www.sciencedirect.com/science/article/pii/S0010465515000442>.
- [40] V. Khachatryan et al. |CMS Collaboration|. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3):155, 2016. <https://doi.org/10.1140/epjc/s10052-016-3988-x>.
- [41] CMS Collaboration. Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV. Technical report, CERN, Geneva, 2016. <http://cds.cern.ch/record/2235192>.
- [42] M. Bähr et al. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, Nov 2008. <http://dx.doi.org/10.1140/epjc/s10052-008-0798-9>.
- [43] M.H. Seymour and A. Siódmok. Constraining MPI models using σ_{eff} and recent Tevatron and LHC Underlying Event data. *Journal of High Energy Physics*, 2013(10), Oct 2013. [http://dx.doi.org/10.1007/JHEP10\(2013\)113](http://dx.doi.org/10.1007/JHEP10(2013)113).
- [44] S. Frixione and P. Nason and C. Oleari. Matching NLO QCD computations with parton shower simulations: the POWHEG method. *Journal of High Energy Physics*, 2007(11):070, Nov 2007. <https://dx.doi.org/10.1088/1126-6708/2007/11/070>.

-
- [45] S. Alioli and P. Nason and C. Oleari and E. Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *Journal of High Energy Physics*, 2010(6), Jun 2010. [http://dx.doi.org/10.1007/JHEP06\(2010\)043](http://dx.doi.org/10.1007/JHEP06(2010)043).
- [46] J. Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *The European Physical Journal C*, 76(4), Apr 2016. <http://dx.doi.org/10.1140/epjc/s10052-016-4018-8>.
- [47] A.M. Sirunyan et al. |CMS Collaboration|. Development and validation of HERWIG 7 tunes from CMS underlying-event measurements. *The European Physical Journal C*, 81(4), Apr 2021. <http://dx.doi.org/10.1140/epjc/s10052-021-08949-5>.

Part II

Main Analysis

Chapter 5

SAMPLE HANDLING

5.1 Introduction

With this section we now enter the main part of this dissertation, dedicated to a detailed presentation of the offline analysis conducted throughout my PhD years. In order, all sample processing stages are outlined, for both data and simulated events, leading to the construction of the final observable. Additionally, the production of fixed-order theory predictions is discussed, followed by comparisons with data distributions. Finally, the results are leveraged to evaluate the proton's PDFs and the strong coupling constant α_s as part of the complementary QCD analysis in Part III.

Hopefully, the nature of jets and their properties have been sufficiently highlighted. Seen in abundance in every proton-proton collision, they can be exploited to construct meaningful observables that offer fundamental insights into the underlying physics. The analysis is based on the measurement of the double differential inclusive dijet production cross section as a function of the invariant mass $m_{1,2}$ and the largest absolute rapidity $|y|_{max}$ of the two leading jets in the event.

Data samples from three separate years 2016, 2017, and 2018 are investigated, covering nearly the entire Run II period of the LHC (see Table 1.1). These datasets were collected by the CMS detector from proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. Respectively, each dataset amounts to 33.5, 41.5 and 59.3 fb^{-1} of integrated luminosity, with a progressive increase over time. Among the three, the primary dataset is set to be that of 2016, which is the only one for which a QCD analysis is performed. Consequently, all plots contained within the main text body refer to this sample, while results from other years are shared in dedicated appendices. It is noted that results for the 2017 and 2018 were obtained with distinct software versions and occasionally some of the processing stages may slightly differ. Therefore, plots originating from these samples should not be used to draw strong conclusions and only a qualitative comparison should be done to the main results derived from 2016. Jet reconstruction is achieved by the anti- k_T clustering algorithm [1] with a radius of $R = 0.8$.

The dijet data cross sections are, for the first time, compared to fixed-order theory predictions at NNLO accuracy in pQCD. Past measurements of this exact observable have been conducted from both CMS and ATLAS at the lower center-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$ [2–4]. While similar measurements have been performed by ATLAS, by exploiting the dijet rapidity separation y^* , instead of the largest absolute rapidity $|y|_{max}$, at a center-of-mass energy of $\sqrt{s} = 7 \text{ TeV}$ [5] and $\sqrt{s} = 13 \text{ TeV}$ [6].

5.2 Data and Monte Carlo samples

As mentioned in Section 2.4, events selected by the CMS trigger system are transferred to the Tier 0 site at CERN for storage, collectively forming the RAW dataset. Various dataset flavors exist, depending on the set of triggers used for event selection. Furthermore, each dataset is divided into different run periods (eras), labeled alphabetically, with each letter corresponding to a distinct time frame of data collection. At this stage, the data contain unprocessed, raw, information as recorded by the CMS subdetectors. Representing the largest data format, they are not yet suitable for physics analyses, as event reconstruction has yet to be performed.

Event reconstruction is applied for each run period separately, resulting to the RECO dataset (reconstructed dataset). Reconstructed objects are now included like tracks, primary and secondary vertex candidates and particle IDs. This data layer is still too detailed, and therefore slow and heavy, making it again unfitting for offline analyses. For this reason, the next data tier available is used, which is a compact or “distilled” version of the RECO data, that is the *Analysis Object Data* (AOD), keeping only necessary high level information useful for analyses. In reality, an even finer version is used for our case, this is the MINIAOD format, which contains reconstructed objects with additional, more complex variables, and up-to-date calibrations [7, 8]. This was the standard data format officially recommended by CMS for RUN II analyses. A schematic view of the CMS data tier system is given in Fig. 5.1. The dataset names, as found in the Data Aggregation System (DAS) [9, 10], are collectively organized in Table 5.1, covering all eras within a given year alongside the corresponding total integrated luminosity for each year.

Table 5.1: DAS dataset names for each year and their integrated luminosity.

Year	$\mathcal{L}_{int} (fb^{-1})$	Era	Data Aggregation System names
2016	33.5	B	/JetHT/Run2016B-17Jul2018-ver2-v1/MINIAOD
		C	/JetHT/Run2016C-17Jul2018-v1/MINIAOD
		D	/JetHT/Run2016D-17Jul2018-v1/MINIAOD
		E	/JetHT/Run2016E-17Jul2018-v1/MINIAOD
		F	/JetHT/Run2016F-17Jul2018-v1/MINIAOD
		G	/JetHT/Run2016G-17Jul2018-v1/MINIAOD
		H	/JetHT/Run2016H-17Jul2018-v1/MINIAOD
		2017	41.5
C	/JetHT/Run2017C-UL2017_MiniAODv2-v1/MINIAOD		
D	/JetHT/Run2017D-UL2017_MiniAODv2-v1/MINIAOD		
E	/JetHT/Run2017E-UL2017_MiniAODv2-v1/MINIAOD		
F	/JetHT/Run2017F-UL2017_MiniAODv2-v1/MINIAOD		
2018	59.3		
		B	/JetHT/Run2018B-UL2018_MiniAODv2_GT36-v1/MINIAOD
		C	/JetHT/Run2018C-UL2018_MiniAODv2_GT36-v1/MINIAOD
		D	/JetHT/Run2018D-UL2018_MiniAODv2_GT36-v1/MINIAOD

It can be seen that the number of run periods varies with the year, with 2016 having seven eras, 2017 five, and 2018 just four. Another observation is that for 2016 the *Legacy* ReReco version of the samples is used while for 2017 and 2018 the *Ultra Legacy (UL)* ReReReco. The latter case, presenting a finer overall event reconstruction and newer

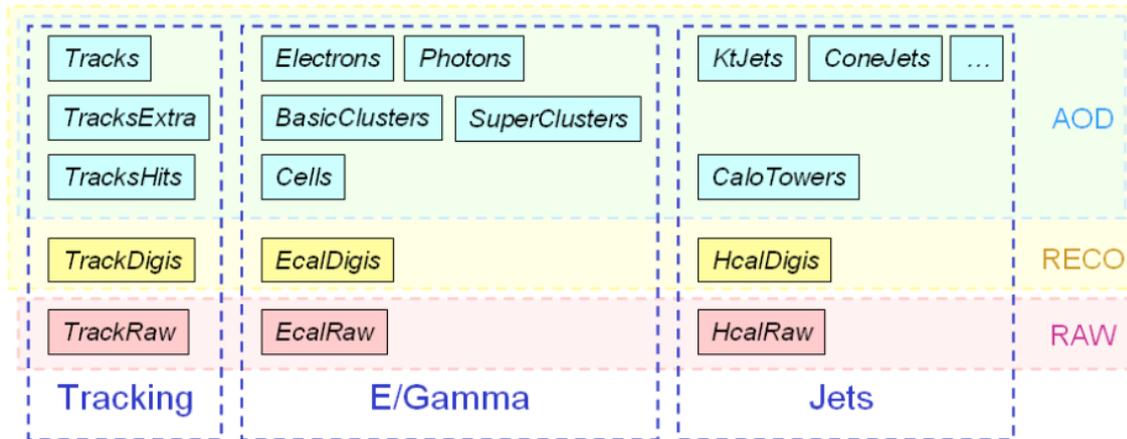


Figure 5.1: Schematic view of the CMS dataset tier system. RECO data contain objects from all reconstruction stages, in contrast to AOD which provide a compact format convenient for analyses [8].

calibrations. What all samples have in common is that they are obtained using the *JetHT* HLT path, an exclusive set of triggers.

Solely *certified data* are used, that not only have passed the first quality checks applied by the trigger, but on top of that, an overall good detector performance is ensured. Events that correspond to good *lumisections*¹, periods at which the CMS subdetectors and generally all data conditions are optimal, are marked in specified files called *Golden JSON* files. These are provided officially by CMS for the analyzers, and are produced for each sample independently, presented in Table 5.2.

Table 5.2: Golden JSON files for each year [11].

Year	Golden JSON file names
2016	Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt

A similar data tier structure is found on the Monte Carlo side, shown in Fig. 5.2. Simulated samples are originally obtained at GEN level (generator level), and contain the pure generation of the particle four-vectors that represent a target physical process. Next in order comes the simulation of the detector through the GEANT4 package, returning as an output the SIM dataset. After the digitization of the detector signals the DIGI tier occurs. Once this is done, reconstruction algorithms are applied, pile-up content is added so that it better reflects realistic data taking conditions, and the trigger menu is imported, in the end, resulting in the RECO format. The rest of the procedure remains the same, just as done for the experimental side. Depending on the level of information kept inside the samples, skimmed versions of RECO will result to the construction of AODSIM or MINIAODSIM datasets ready for analyses. Note that NANO AOD is an even lighter version of MINIAOD and is slowly becoming the most popular dataset format for RUN III analyses. Table 5.3 shows all MC samples used in the analysis for each year separately.

¹One lumisection is equivalent to 23 s worth of collisions.

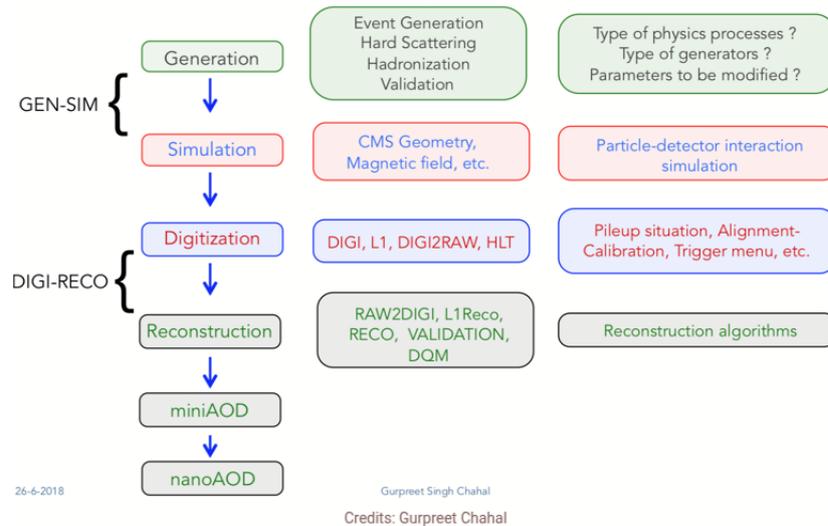


Figure 5.2: Schematic view of Monte Carlo production. DIGI represents an identical format to that of the actual detector output [12].

For 2016 three individual MCs were used - PYTHIA8 [13] with the CUETP8M1 tune [14] in a sliced and a flat version, and MADGRAPH5 [15, 16] interfaced to Pythia8 CUETP8M1, to provide an alternative estimation of the matrix element. For 2017 and 2018 just one MC was used, the sliced version of Pythia8, this time with the newer CP5 tune [17]. Parallel to how the experimental samples were divided into different sub-periods, simulated samples are usually obtained in a definite number of slices. These are expressed in terms of the generator variable used to define the hard process, with Pythia having 14 \hat{p}_T slices, and Madgraph having 9 in terms of H_T . Exact slice ranges and their individual cross section values can be found at Appendix B.

The advantage of a sliced sample over a flat one is that the former populates individual phase space regions with events uniformly, in a controlled manner, resulting in a higher statistics. Pythia in slices was set as the primary sample for all three years, to perform comparisons, and deduce corrections and systematics. Especially for 2016, a flat sample was used in addition, so that cross checks can be made throughout the offline analyses. On the other hand, the Madgraph sample was exploited to derive a model uncertainty for the measurement, thus taking into account that no unique way exists currently to define the correct approach for calculating such processes.

5.3 Event selection and corrections

5.3.1 Analysis software

A customized version of the MINIAOD samples was produced, reducing their size by retaining only the information relevant to the measurement performed here. This optimization allowed for a faster and more efficient analysis. From homemade ntuplization², to production of the final results, the whole analysis was based on the *DAS Analysis System*, or just DAS in short. The software code is publicly available on GitLab (see Ref. [18]) to the whole collaboration. It is based mostly on custom software developed by analyzers for their studies. Continuously evolving, it already serves as the foundation for multiple

²Ntuplization refers to the process of creating structured datasets. The term originates from the historical use of n-tuples as the standard data format for organizing and storing information in high energy physics analyses.

Table 5.3: DAS simulation dataset names for each year and their slices. The P8 abbreviation is used to refer to Pythia8 and MAD to refer to Madgraph.

Year	Name	Slices	Data Aggregation System names
2016	P8 Sliced	14	/QCD_Pt_*to*_TuneCUETP8M1_13TeV_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6*/MINIAODSIM
	P8 Sliced	1	/QCD_Pt-15to7000_TuneCUETP8M1_Flat_13TeV_pythia8/RunIISummer16MiniAODv3-PUMoriond17_magnet0n_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	MAD+P8 Sliced	9	/QCD_HT*to*_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3*/MINIAODSIM
2017	P8 Sliced	14	/QCD_Pt_*to*_TuneCP5_13TeV_pythia8/RunIISummer20UL17MiniAODv2-106X_mc2017_realistic_v9-v1/MINIAODSIM
2018	P8 Sliced	14	/QCD_Pt_*to*_TuneCP5_13TeV_pythia8/RunIISummer20UL18MiniAODv2-106X_upgrade2018_realistic_v16_L1v1-v1/MINIAODSIM

analyses. Different modules exist to manage high-level objects and arm users with tools to conduct their research. The code itself sits on top of CMSSW³ [19] and makes a heavy use of the ROOT framework [20]. Processing of the 2016 (2017 and 2018) datasets was performed through the CMSSW_10_6_X (CMSSW_12_4_X) version.

5.3.2 Event selection

After jet reconstruction (see Section 4.5.2), and during the ntuplization stage, events are scrutinized, to discard spurious events that may have slipped through previous quality checks. Events are required to contain at least one Primary Vertex (PV). If so, its z component must fulfill the following condition $|z(PV)| < 24\text{ cm}$, i.e. should not be too far away from the interaction point. Additionally, the radius of the PV on the $x - y$ plane should satisfy $r_{xy}(PV) < 2\text{ cm}$, while the vertex itself should be the product of a fit composed by at least 5 points in space, $ndof > 4$. These ensure a well-defined, good PV.

A stricter cut on the jet phase space is applied later during the construction of the observable, as discussed in Section 5.5.

5.3.3 Corrections

Jet Identification

Further criteria are applied to minimize noise - possible artifacts from bad event reconstruction, and reject fake events, that retain 98 – 99% of real, well reconstructed ones. These fall under the category of *Particle Flow Jet Identification Criteria*, *PFjetID*, and are officially recommended by CMS [21]. Particularly the **TightLepVeto** Jet ID cuts are applied, a summary of which is presented in Table 5.4.

³The official CMS offline SoftWare, CMSSW

Table 5.4: Summary of the TightLepVeto Jet ID cuts for each year.

PF Jet ID	2016	2017 & 2018
	$ \eta \leq 2.7$	$ \eta \leq 2.7$
Neutral Hadron Fraction	< 0.90	< 0.90
Neutral EM Fraction	< 0.90	< 0.90
Number of Constituents	> 1	> 1
Muon Fraction	< 0.80	< 0.80
Charged Hadron Fraction	> 0	> 0
Charged Multiplicity	> 0	> 0
Charged EM Fraction	< 0.90	< 0.80

MET filters

Measurement of the Missing Transverse Momentum (MET) is crucial to various analyses, most importantly to searches for new physics. Dedicated filters exist to correct this observable by rejecting events affected by detector malfunctions, beam-related backgrounds, and reconstruction failures that would otherwise overestimate this quantity. These go under the name of *MET filters*, and are applied to all jet based analysis, presented in Table 5.5.

Table 5.5: List of MET filters applied to each sample.

MET filters	2016	2017 & 2018
goodVertices	✓	✓
globalSuperTightHalo2016Filter	✓	✓
HBHENoiseFilter	✓	✓
HBHENoiseIsoFilter	✓	✓
EcalDeadCellTriggerPrimitiveFilter	✓	✓
BadPFMuonFilter	✓	✓
BadPFMuonDzFilter	—	✓
hfNoisyHitsFilter	—	✓
BadChargedCandidateFilter	✓	✓
eeBadScFilter	✓	✓
ecalBadCalibFilter	—	✓

Hot zones

Problems can arise anywhere, anytime, especially in complex systems such as the 14t, 21m long CMS detector. During online data taking, for specified and well identified time periods, it was observed that some regions of the calorimeter were producing abnormally high jet rates, these were later labeled as *hot zones*. The effect was caused by sub-optimal calibration for jet measurement performance. In order to avoid introducing any bias due to this systematic effect to the measurements, dedicated efficiency maps have been released by CMS to veto events in the specified “hot” regions. The exact versions to these maps are given in Table 5.6. Note that to keep the phase space symmetric, this correction is applied both on data and MC with the exact same versions for each year.

Table 5.6: List of the jet veto map versions for each year [22, 23].

Year	Version
2016	2016EOY_LegacyReReco/V2
2017	Summer19UL17_V2
2018	Summer19UL18_V1

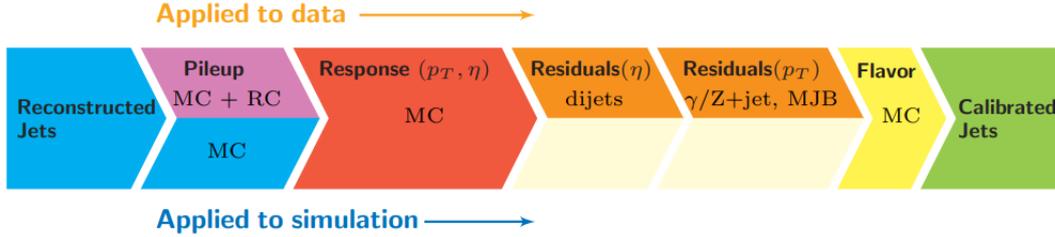


Figure 5.3: Consecutive stages of JEC, for data and MC simulation. All corrections marked with MC are derived from simulation, RC stands for random cone, and MJB refers to the analysis of multiple events. [24].

Jet Energy Corrections

During the offline analysis, jets are calibrated so that they better reflect the correct *jet energy scale* (*JES*). This step is crucial for both Data and MC, as the detector’s response to particles is inherently nonlinear. Consequently, an initial direct measurement does not yield the true particle energy, necessitating these adjustments. *Jet Energy Corrections* (*JECs*) are applied directly to the jet four-vector in the form of multiplicative factors. They address multiple detector effects, each presenting each own correction factor, are applied sequentially, one after the other, in the end, resulting in the calibrated jet four-vector. An illustration of this procedure can be found in Fig. 5.3, where the top part corresponds to corrections applied to data, and the bottom to those applied in MC.

The procedure is led by the *pileup offset correction* or *L1 correction* which removes additional energy caused by neutral particles due to pileup effects. In principle, such energy deposits have been already “dealt with” through the pileup mitigation algorithm, the CHS algorithm. In practice, while CHS properly removes contributions from charged hadrons, for neutral ones the correction is slightly overestimated. The L1 correction restores the energy scale in that regard, improving the resolution, and reducing systematics. Note that in the case of PUPPI jets, there is no need for such corrections, as they are accurately treated by the algorithm itself. This correction is derived by comparing MC QCD dijet events with and without pileup. Any residual differences between data and the simulation are treated using the random cone (RC) method in Minimum Bias (MB) events, resulting in a bidirectional correction that is applied in both the data and simulation.

MC-truth corrections follow, which account for the non-uniform and non-linear response of the detector with respect to the pseudorapidity η and the transverse momentum p_T respectively. They are obtained by comparing the reconstructed level p_T to that of the particle level in QCD dijet events. Applied both on data and MC.

MC samples are created according to a more ideal image of the detector, resulting in a better overall response compared to what is actually observed in data. *Residual corrections* or *L2L3 corrections* are applied only on data in that respect, to treat imperfections of the real detector so that the response “uniformity” is restored. This is achieved by measuring the p_T balance in the barrel ($|\eta| < 1.3$) between a jet and a reference object, usually a well measured object, like a muon or a photon. For this reason, $Z(\mu\mu, ee)+$ jet, Wqq decays

for $t\bar{t}$ events, γ + jets, and multijet event topologies are exploited.

Lastly, there are the optional *Flavor corrections*. These amount for the differences observed in the detector response between quark and gluon-initiated jets. As mentioned before, gluon initiated jets are of lower momenta and appear in wider shapes, compared to up and down quark initiated jets that result in a higher energy content, and a narrower shape, leading to a detection with better resolution. The response for charm and beauty quark jets lies in between that of the gluon and u/d jets. This correction is applied to both data and MC and is based on QCD gluon enriched, and quark enriched Z + jet, and γ + jet events.

Each sample is accompanied by a unique *Global Tag* which dictates what version of any type of corrections should be applied. Jet energy corrections specifically are produced centrally by CMS, from the JERC subgroup [25]. The JEC versions used in the analysis can be found in Table 5.7.

Table 5.7: List of JEC versions used for both Data and MC in each year [26].

Year	Data	MC
2016	Summer16_07Aug2017 (Run) _V11_DATA	Summer16_07Aug2017_V11_MC
2017	Summer19UL17_Run (Run) _V5_DATA	Summer19UL17_V5_MC
2018	Summer19UL18_Run (Run) _V6_DATA	Summer19UL18_V5_MC

Prefire

Another kind of correction applied only to data is related to an issue spotted at the L1 ECAL trigger. Apparently, the L1 ECAL trigger primitives were prematurely firing (prefiring) due to a misadjusted time offset which was originally introduced to negate this already existing effect of ECAL. The problem originates from the high radiation environment inside CMS, which causes ECAL crystals to be prone to radiation damage. For this reason, a calibration over time is needed to mitigate the prefiring effect.

This correction concerns only the 2016 and 2017 data. Prefiring caused some jets in the forward region $2.0 < |\eta| < 3.0$ to be wrongly considered by the L1 trigger as belonging to the previous bunch crossing, overrunning the L1 rule that “forbids two consecutive bunch crossings to fire”. As a result, an additional inefficiency is introduced in the measurement and needs to be addressed. Two main approaches exist to treat this issue, either to apply new weights in the simulated samples so that the prefiring effect is reproduced, or to apply the inverted weights in the data to negate it. Here the second approach is followed, as foreshadowed at the start of the paragraph. Dedicated maps, containing the prefiring probability as a function of p_T and η , are provided by the JetMET group [27].

JetMET provides the correction tables in two formats, the *Maps per Era* and the *Average Maps*. In the former case, different maps exist for each run period (era), while in the latter case, these are maps averaged over time, over all eras. Both carry advantages and disadvantages over one another. The main advantage of maps per era is that they account for the time dependence of the prefiring effect, but contain less statistics since they are split into different files, therefore are provided in a coarser binning scheme, leading to a loss of resolution. On the contrary, average maps contain more statistics and are provided in a finer binning scheme. A comparison of the two use cases can be found in Fig. 5.4 for the 2016 data.

The opportunity is taken to discuss on the plot style. The ratios in Fig. 5.4 are shown as a function of the invariant dijet mass $m_{1,2}$. While a precise definition of the quantity has

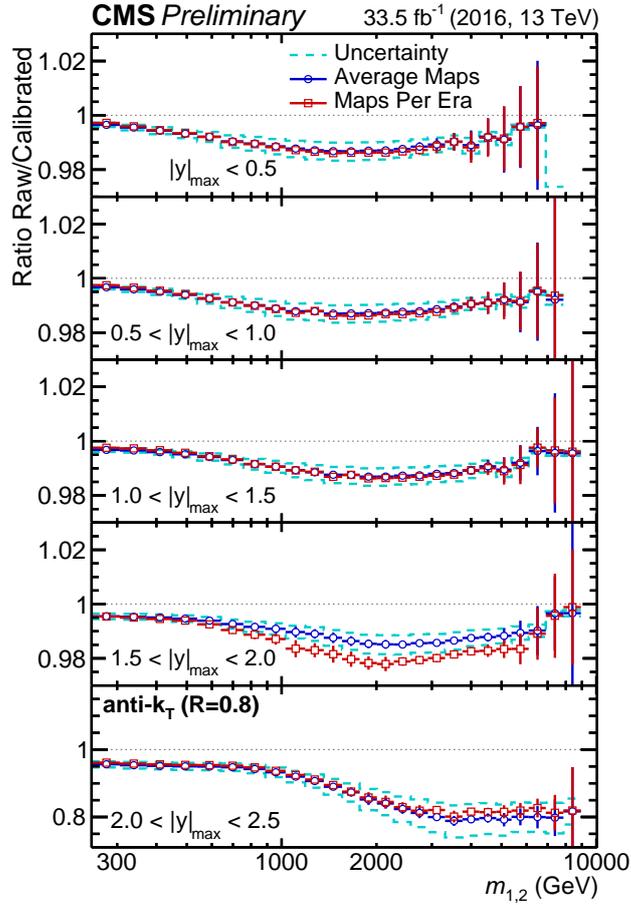


Figure 5.4: Prefiring effect for the 2016 dataset. Ratio between the raw (uncorrected) and the calibrated (corrected) data cross section as a function of the invariant dijet mass $m_{1,2}$. Each cell corresponds to a different rapidity $|y|_{max}$ region. Maps per era are indicated in red and average maps in blue. The cyan dashed lines represent the uncertainty variation of the average maps use case.

not yet been provided, it serves as the central quantity for most of the plots presented in this work. Similarly, the explored phase space is divided in five rapidity regions in terms of the $|y|_{max}$ variable, so most of the plots will contain five cells each representing a different corner of the phase space.

Looking at Fig 5.4 in more detail, it is observed that the two use cases result in compatible results, within respective uncertainties, nearly across all phase space regions. Small deviations from the previous conclusion are only observed in the fourth rapidity bin ($1.5 < |y|_{max} < 2.0$) for central values of $m_{1,2}$. The prefiring effect is estimated to have a low impact in the central⁴ rapidity regions $\sim 1.5\%$ at maximum, while this effect becomes significant when going in the forward⁵ ones, up to $\sim 20\%$. In the context of this analysis the average maps are used to compensate for the loss of efficiency caused by the prefiring effect. The correction is applied directly in the event weight according to the recommendations, while the maps themselves can be found in Appendix C. An uncertainty is attributed to the effect which is calculated by taking the maximum between 20% of the

⁴Central rapidity regions are usually referred to as the ones that lie within $|y| < 1.5$, i.e. approximately the tracker and barrel calorimeter regions of the CMS detector.

⁵Forward rapidity regions on the other hand are usually referred to as the ones that lie within $1.5 < |y| < 5$, i.e. the endcap and forward calorimeter regions of the CMS detector.

prefiring probability and the corresponding statistical uncertainty.

Jet Energy Resolution

The detector has a finite jet energy resolution and Monte Carlo samples are produced before the actual detector is well understood. As a result, simulated samples are created with a too optimistic detector resolution. To account for this phenomenon, MC samples need to be smeared, i.e. worsen the situation, so that their resolution better reflects that found in data. According to recommendations [28] two approaches exist to treat this phenomenon, presented below:

- **Scaling method.** In this approach, the reconstructed jet four-vector is rescaled by a factor

$$c_{JER} = 1 + (s_{JER} - 1) \frac{p_T - p_T^{ptcl}}{p_T} \quad (5.1)$$

where p_T and p_T^{ptcl} are the jet transverse momentum at reconstructed and generator level (or particle level) respectively, and s_{JER} is a scale factor, provided by JetMET, that expresses the data-to-simulation resolution differences. The key point here is that this type of correction can only be applied if matching has been performed, meaning a generator level jet must be successfully matched to the corresponding reconstructed jet, before applying the smearing to its p_T . For the matching the following requirements are imposed

$$\Delta R < R_{cone}/2, \quad \frac{|p_T - p_T^{ptcl}|}{p_T} < 3 \sigma_{JER} \quad (5.2)$$

where R_{cone} is the jet cone size, in our case 0.8, $\frac{|p_T - p_T^{ptcl}|}{p_T}$ short of, the response⁶, and σ_{JER} the relative p_T resolution as measured in simulation.

- **Stochastic method.** This approach comes as complementary to the previous one, as it does not require the presence of a matching generator level jet. Since in such cases there is no quantitative way to obtain an estimate for the response, one can resort to sampling, i.e. the use of random numbers. The scaling factor can then be written as

$$c_{JER} = 1 + \mathcal{N}(0, \sigma_{JER}) \sqrt{\max(s_{JER}^2 - 1, 0)} \quad (5.3)$$

where $\mathcal{N}(0, \sigma_{JER})$ denotes a random number sampled from a normal distribution with a zero mean and variance σ_{JER}^2 .

In reality, a combination of the two approaches is recommended, called the *hybrid method*. When a matching particle level jet is found, and their response lies within the Gaussian core assumption (requirement of $\Delta < 3 \sigma_{JER}$), the *scaling method* should be used, otherwise, for unmatched jets or matched jets that their response lies at the tails of the distribution ($\Delta > 3 \sigma_{JER}$), the *stochastic smearing* should be applied. In the end, for 2016, a modified version of the recommended hybrid approach is applied, a decision made as an aftermath of the study shown in Appendix D. Homemade resolution tables were produced for 2016, while the original scale factors by JetMET were preserved and used. On the other hand, for 2017 and 2018, the plain stochastic method is followed, where both the resolution and SFs are taken as provided. The resolution tables and SFs are taken from Ref. [28], while the exact versions are shown in Table 5.8.

Table 5.8: JER table versions for each year. For 2016 only the scale factors were used, and were combined to homemade resolution tables according to the study in Appendix D.

Year	Data
2016	Summer16_25nsV1_MC
2017	Summer19UL17_JRV2_MC
2018	Summer19UL18_JRV2_MC

Pileup profile reweighting

As already mentioned, PU interactions are simulated separately from the generation of the hard interaction events, and only later are combined into actual samples. A difference is observed when making a comparison between the PU profile contained in MC and the one observed in real data. This difference is shown in Fig.5.5 for the 2016 data and Pythia samples, and is caused by the overestimation of the effect during the simulation stage on the MC side. The top plot displays the pileup profiles, or better, the number of pileup interactions of both samples. On the data front these estimates are based on a 69.2mb inelastic pp cross section [29]. In detail, the shaded pink area corresponds to the Pythia profile while the lines in different colors represent the individual profiles of each trigger that contributed to the data sample. The bottom plot then contains a direct comparison between the PU profiles of each trigger to the MC one.

To redeem the situation, the PU profile found in simulated samples is corrected to match that found in data, through a process called *Pileup profile reweighting*. Two approaches exist within the software:

- *Per trigger PU profile reweighting*. This approach results in a trigger dependent treatment of the effect, exploiting the individual trigger profiles to derive appropriate weights for jets found on the MC sample.
- *Global PU profile reweighting*. In this case, the total data PU profile is used, in which no trigger information is exploited, resulting in a trigger independent approach.

A comparison of the two methods in the dijet mass spectrum is shown in Fig. 5.6. There, the ratio is displayed between the corrected over the uncorrected differential dijet mass cross section. The trigger independent method is given in red, while the trigger dependent in blue, along with its uncertainty variations, illustrated with the cyan lines. Both of the results lie within the uncertainty band, indicating small deviations. The effect of the correction seems to be of the same order for all rapidity bins and relatively uniform. In the context of this analysis, since no strong deviations are observed, the trigger dependent method is preferred since it provides a finer overall correction by incorporating trigger information.

Pileup cleaning

Another kind of correction related to the pileup found in simulated samples is the removal of over-weighted events. We refer to this type of correction as *PU cleaning*. The problem

⁶Originally, the response is defined as $\frac{p_T - p_T^{ptcl}}{p_T^{ptcl}}$ rather than normalizing by p_T . Nevertheless, for brevity, the quantity obtained by dividing with p_T will be referred to as the response throughout the main text.

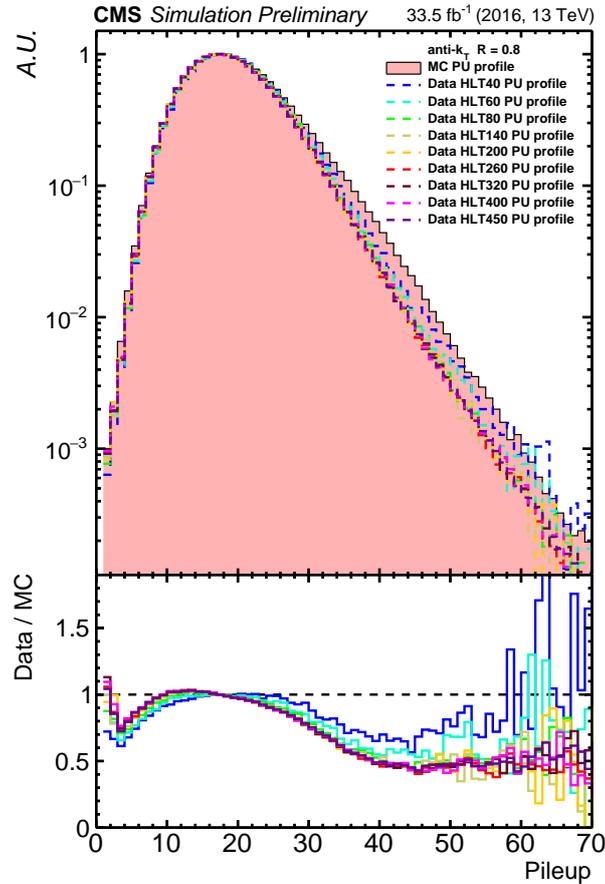


Figure 5.5: PU profile comparison between the 2016 data and the 2016 Pythia sliced samples (top), and their ratio (bottom).

arises when large MC samples are split in slices (in terms of \hat{p}_T or H_T , see Table.5.3) so that events are sufficiently generated in smaller phase space intervals. As mentioned above, PU events are simulated separately and do not use the slicing method. So when the main events are combined with the PU ones and slices are normalized with the appropriate cross section, PU jets may end up in phase space regions that correspond to a lower \hat{p}_T or H_T value. Events then that are found with an unphysical large weight in misplaced regions are dropped.

5.3.4 Processing stages

This section provides an overview of the individual processing steps for both Data and MC samples. These steps will be explicitly listed, with dedicated sections, later, addressing each topic in detail, while most of them have already been introduced in the previous section, which focused on the applied corrections.

Data processing stages

1. Ntuplization
2. Jet Identification
3. MET filters
4. Hot zones

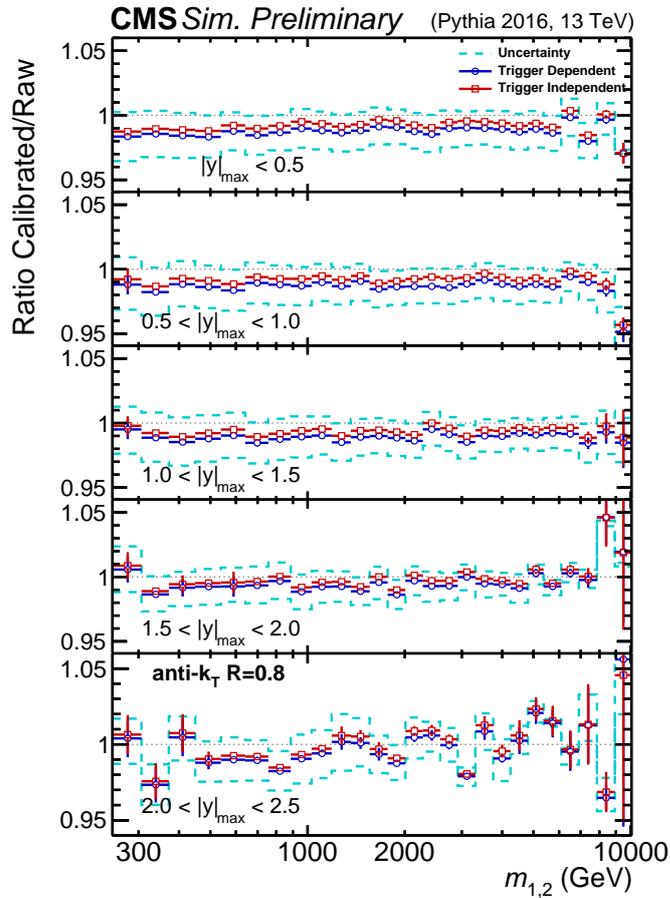


Figure 5.6: Impact of the PU profile reweighting options on the dijet mass spectrum obtained from the Pythia sliced sample.

5. Jet Energy Corrections
6. Normalization (see Section 5.5)
7. Prefire (not applied for 2018 samples)

MC processing stages

1. Ntuplization
2. Jet Identification
3. MET filters
4. Hot zones
5. Normalization (see Section 5.5)
6. PU cleaning
7. Jet Energy Corrections
8. Jet Energy Resolution
9. PU profile reweighting

5.4 Trigger efficiencies

At the center of CMS, every 25 ns , the accelerated bunches of protons collide. The available amount of information to be recorded and processed from just one bunch crossing (BX) is approximately 1 MB/event . This number is only getting larger if all $\sim 10^9$ interactions per second (40 MHz) are taken into account. There is no feasible way of physically storing all these data. Apart from that, only a small fraction of them contain events of interest for physics analysis. This is where the two-tier trigger system comes into play.

For this work only data collected by a specific trigger path are used. These are the *High Level single PF Jet Triggers* (HLT_AK8PFJet_X), specified for AK8 jets, where each of them requires the presence of at least one well reconstructed jet in the event. In total, there are 10 such triggers, armed with different p_T thresholds. Moreover, for an event to be recorded it is essential that the p_T of the leading⁷ jet is larger than that of the trigger threshold value.

5.4.1 Prescale factors and effective luminosity

Another property of the triggers is the prescale factor k , which answers to the question of why bothering using different trigger thresholds. In the hypothetical case of only one threshold and therefore one trigger, this trigger would be solely responsible for covering the entire phase space. It would be impossible to keep up with the high event production rate, which is the initial problem to begin with. For this reason, different triggers are used to uniformly cover all phase space regions. What's more, the differential cross section measured here is a steeply falling spectrum, meaning that at lower energies event rates are larger compared to those at higher regions. The prescale factor comes to the rescue, which is introduced as a calibration mechanism, to reduce the high event rates encountered at the lower end of the spectrum. As an example, a trigger with a prescale factor $k = 1000$ records 1 event for every 1000. Events recorded this way are chosen randomly and appropriate weights are attributed to them during the offline analysis. This way the initially large rates are kept under control, while at the same time events are filtered since only a small fraction of the lower energy, not so interesting events are kept. Triggers with higher thresholds have a lower prescale value, so as to keep increasingly more events at higher energy regimes, which have a higher probability of unraveling any interesting physics phenomena. Of course, unprescaled triggers, i.e. a prescale value of $k = 1$, are the ones that record each event that passes their criteria.

Table 5.9 shows the effective luminosity collected by the triggers for each year. The effective luminosity is estimated by multiplying the active luminosity times HLT and L1 prescales by using the *brilcalc* tool [30]. Only triggers that carry the highest p_T threshold for each year are unprescaled, HLT_AK8PFJet_450 for 2016, and HLT_AK8PFJet_500 for 2017 and 2018. Since different triggers record events that belong to different sub-ranges of the phase space, selected information has to be carefully extracted and combined during the offline analysis to successfully reconstruct the desired distributions.

5.4.2 Emulation method and turn-on points

Triggers are used when they are fully efficient ($> 99.5\%$) and only events that lie in these efficient regions are exploited to assemble the observable. The analysis continues with

⁷Leading jet is considered the jet with the highest transverse momentum p_T in the event. Likewise, the jet with the second largest p_T in an event is referred to as subleading.

Table 5.9: Effective trigger luminosity for each year.

HLT paths	$\mathcal{L}_{eff}^{2016} (fb^{-1})$	$\mathcal{L}_{eff}^{2017} (fb^{-1})$	$\mathcal{L}_{eff}^{2018} (fb^{-1})$
HLT_AK8PFJet_40	0.0496663	0.182566	0.014916
HLT_AK8PFJet_60	0.328065	0.504795	0.416271
HLT_AK8PFJet_80	1.00466	2.527472	2.134792
HLT_AK8PFJet_140	10.1074	26.6014	46.7846
HLT_AK8PFJet_200	85.7619	188.957	201.048
HLT_AK8PFJet_260	518.048	469.357	461.592
HLT_AK8PFJet_320	1525.55	1226.89	1229.30
HLT_AK8PFJet_400	4591.04	7690.57	3686.44
HLT_AK8PFJet_450	33534.8	9663.22	7326.59
HLT_AK8PFJet_500	—	41471.4	59262.1

the construction of the efficiency trigger curves, so as to evaluate the exact *turn-on* point for each trigger, point at which their efficiency is greater than 99.5%. The efficiency is calculated according to the *emulation* method given below. As an example, suppose the we want to calculate the efficiency curve of HLT_AK8PFJet_80:

1. First, a histogram is filled with all objects that fired the trigger with a lower p_T threshold, in this case, the HLT_AK8PFJet_60. This is just a reference histogram to HLT_AK8PFJet_80 since it will be used as a handle for a comparison in the following steps.
2. L1 and HLT objects of the reference histogram are accessed.
3. We compare which of these objects fulfill the criteria to be recorded by the trigger of interest, here of the HLT_AK8PFJet_80.
4. Events that have passed step 3 are filled in a second histogram. This is a histogram that emulates the efficiency of HLT_AK8PFJet_80 since it contains events “supposedly” recorded by it. Note that in the whole procedure, indeed, no actual events recorded by HLT_AK8PFJet_80 are used; its behavior is only emulated.
5. Lastly, dividing the emulated histogram, from step 4, to the reference one, from step 1, the efficiency curve of HLT_AK8PFJet_80 is obtained. This is also represented below

$$ef_{80} = \frac{Emul_{80}}{Ref_{60}} \quad (5.4)$$

Efficiency curves are computed with respect to the leading jet p_T for all trigger paths, separately for each rapidity region. Since transverse momentum is not sensitive to rapidity, no significant deviations in the turn-on points between rapidity regions are expected, a priori. Figure 5.7 and 5.8 show the curves obtained for the 2016 dataset. Central rapidity regions are contained in the plots of the former figure and forward ones in the latter. Efficiency curves for different triggers are distinguished using various colors and marker styles within the same plot. Lines are overlaid on the markers to enhance visualization and aid differentiating the trigger curves. Parentheses in the legend display the turn-on points for the specific rapidity bin, while vertical colored lines illustrate their exact position in the x -axis. Trigger curves for the 2017 and 2018 data can be found in Appendix E.

All efficiency curves are obtained via the emulation method discussed above, with the exception of the 2016 curve of HLT_AK8PFJet_40 where the *Tag & Probe* method was used,

as there is no trigger with a lower p_T threshold to emulate his behavior. In the end, as it will be shown in the next section, `HLT_AK8PFJet_40` does not contribute to the construction of the dijet mass spectrum, so the specifics for the *Tag & Probe* method are omitted. For this reason, since there is no contribution from `HLT_AK8PFJet_40` its curves are utterly dropped and are not shown for 2017 and 2018 in Appendix E. Three sets of turn-on points are determined, one for each year. Within a given year, the highest value among the five rapidity bins for a specific trigger is selected as the final turn-on point. Table 5.10 lists the turn-on points among all relevant triggers for each year separately. Once more, unrescaled triggers are highlighted by having their turn-on points presented in bold.

Table 5.10: Trigger turn-on points for all years.

HLT paths	Turn-on points (GeV)		
	2016	2017	2018
HLT_AK8PFJet_40	74	—	—
HLT_AK8PFJet_60	97	133	133
HLT_AK8PFJet_80	114	153	153
HLT_AK8PFJet_140	196	220	245
HLT_AK8PFJet_200	272	300	300
HLT_AK8PFJet_260	330	430	395
HLT_AK8PFJet_320	395	507	468
HLT_AK8PFJet_400	507	592	592
HLT_AK8PFJet_450	592	686	638
HLT_AK8PFJet_500	—	737	790

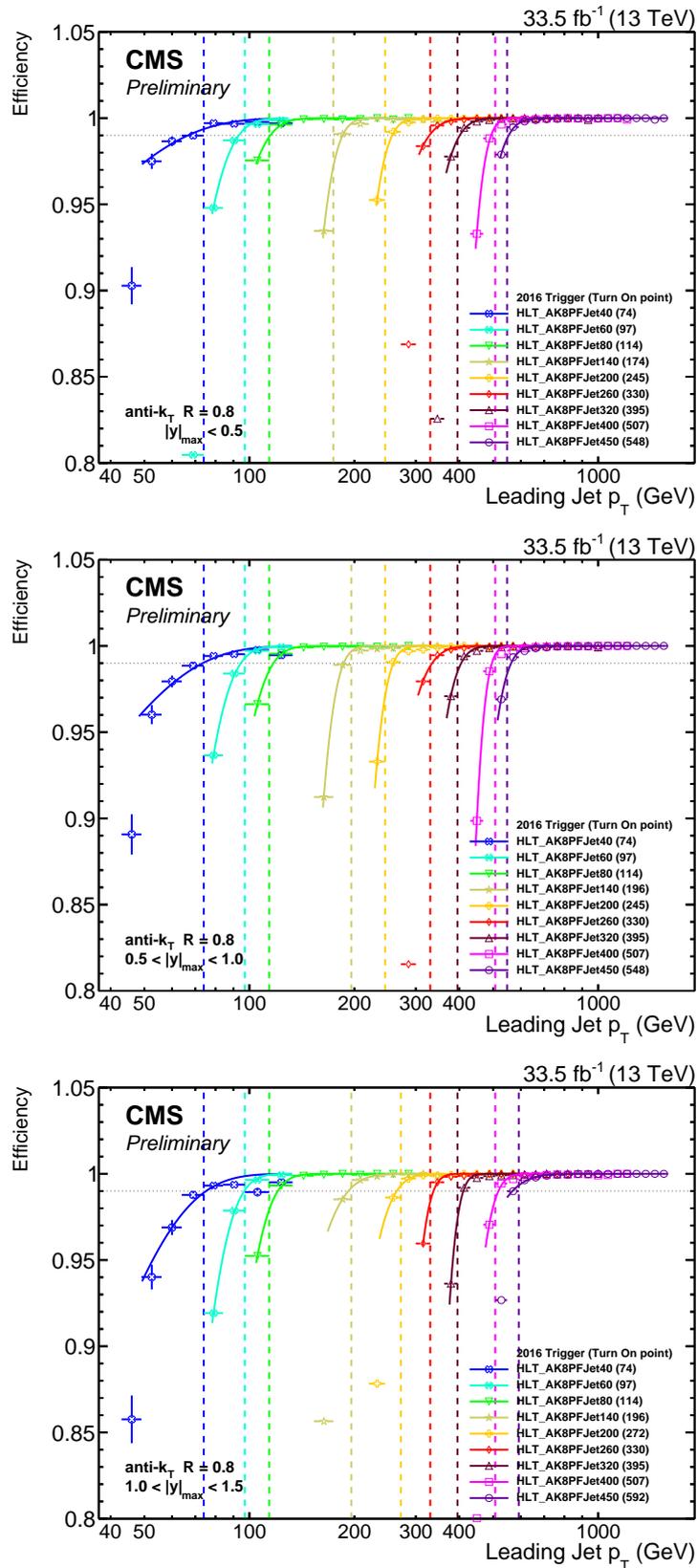


Figure 5.7: Trigger efficiency curves as a function of the leading jet p_T in central rapidity regions, $|y|_{\max} < 0.5$ (top), $0.5 < |y|_{\max} < 1.0$ (middle), and $1.0 < |y|_{\max} < 1.5$ (bottom), for 2016.

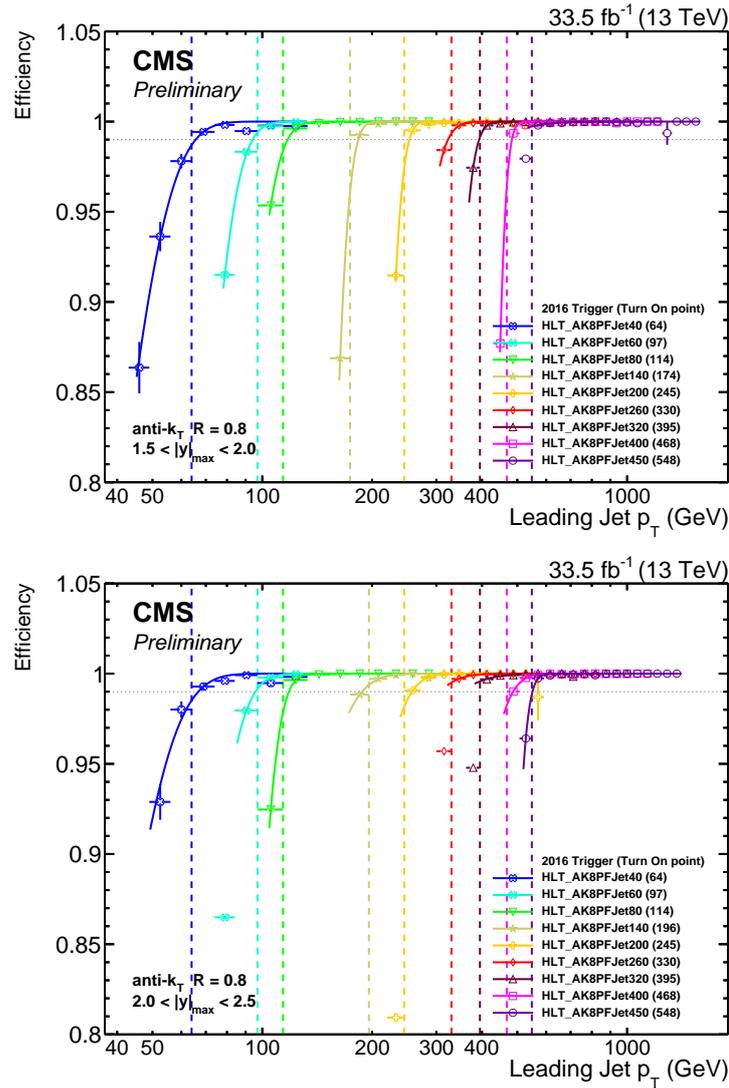


Figure 5.8: Trigger efficiency curves as a function of the leading jet p_T in forward rapidity regions, $1.5 < |y|_{max} < 2.0$ (top), and $2.0 < |y|_{max} < 2.5$ (bottom), for 2016.

5.5 Detector level cross section

After obtaining the necessary information from the trigger studies discussed in the previous section, it is time to move on with the construction of the detector level spectrum. Events that pass the trigger selection are further restricted so that the leading and subleading jets comply with the following requirements of $p_{T,1} > 100 \text{ GeV}$, $|y_1| < 2.5$, and $p_{T,2} > 50 \text{ GeV}$, $|y_2| < 2.5$, where the subscripts 1 and 2 denote attributes required from the leading and subleading jets respectively.

The double differential inclusive dijet production cross section is measured as a function of the invariant mass $m_{1,2}$ and the largest absolute rapidity $|y|_{max}$ of the two leading jets in the event. Their definition is provided:

$$m_{1,2} = \sqrt{(E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2} \quad (5.5)$$

$$y_{max} = \text{sign}(|\max(y_1, y_2)| - |\min(y_1, y_2)|) \times \max(|y_1|, |y_2|) \quad (5.6)$$

$$|y|_{max} = |y_{max}| = \max(|y_1|, |y_2|) \quad (5.7)$$

where y_{max} corresponds to the rapidity of the jet closer to the beam line (outermost jet).

The phase space is divided into five equal-width bins in $|y|_{max}$ and in 28 variable-width bins in $m_{1,2}$, with detailed bin edges provided in Table 5.11. The same $m_{1,2}$ binning is applied within each rapidity bin. The dijet mass binning scheme is an extended version of that used in a similar analysis of the triple differential dijet cross section measurement at 13 TeV with the CMS detector. Notably, both that measurement and the analysis presented in this thesis were published in a shared paper found in Ref. [31]. The $m_{1,2}$ bin widths have been chosen such that they are approximately equal to or greater than four times the resolution (4σ), so that event migrations between neighboring bins are minimized.

Table 5.11: Bin edges for invariant mass and rapidity. The same $m_{1,2}$ binning is used for all $|y|_{max}$ bins.

Binning schemes			$ y _{max}$
$m_{1,2} \text{ (GeV)}$			
[160 – 200)	[200 – 249)	[249 – 306)	[0.0, 0.5)
[306 – 372)	[372 – 449)	[449 – 539)	[0.5, 1.0)
[539 – 641)	[641 – 756)	[756 – 887)	[1.0, 1.5)
[887 – 1029)	[1029 – 1187)	[1187 – 1361)	[1.5, 2.0)
[1361 – 1556)	[1556 – 1769)	[1769 – 2008)	[2.0, 2.5)
[2008 – 2273)	[2273 – 2572)	[2572 – 2915)	
[2915 – 3306)	[3306 – 3754)	[3754 – 4244)	
[4244 – 4805)	[4805 – 5374)	[5374 – 6094)	
[6094 – 6908)	[6908 – 7861)	[7861 – 8929)	
[8929 – 10050)			

From the effective total number of events N_{eff} one can transpose to the differential cross section through:

$$\frac{d^2\sigma}{dy_{max} dm_{1,2}} = \frac{1}{\mathcal{L}_{int}} \frac{N_{eff}}{(2\Delta|y|_{max})\Delta m_{1,2}} \quad (5.8)$$

where $\Delta|y|_{max}$ and $\Delta m_{1,2}$ denote the bin widths in the respective quantities. While the phase space is defined in terms of absolute rapidity $|y|_{max}$ bins, the signed version is retained on the left side of the cross section definition. To account for both the negative and positive parts of the phase space, a factor of 2 is included in the denominator in front of $\Delta|y|_{max}$. As expected, the entire dataset is normalized to the total integrated luminosity \mathcal{L}_{int} . On the Data side, events contributing to the cross section are weighted according to trigger prescales and selection efficiency. On the MC side, each \hat{p}_T (or H_T) slice is normalized to the corresponding cross section, obtained from the generator, and the entire sample is further normalized according to \mathcal{L}_{int} .

The resulted data distributions at detector level are shown in the top plot of Fig. 5.9. For each of the five rapidity bins the corresponding spectrum is displayed with a different color and marker style, illustrated in the legend. Each distribution is scaled by an increasing factor, in powers of 10^n , to increase visibility. The x -axis extends from $249 GeV$ up to $8929 GeV$ in the higher end of the spectrum. The differential cross section is observed to fall rapidly over multiple orders of magnitude, having the expected behavior of a steeply falling spectrum, as more energetic events are encountered rarely in comparison to lower energetic ones.

The bottom plot in the same figure conveys the exact same information but, this time, event contributions coming from individual triggers are highlighted. Distributions between rapidity bins are shown in different cells within the same plot. It is observed that multiple triggers contribute to a given $m_{1,2}$ bin. This is natural since triggers record events based on the leading jet p_T , quite a different quantity compared to the dijet mass as defined in Eq. 5.5. Trigger contributions are illustrated with different colors, in the end, resulting in a stacked plot (the total event content for a given mass bin can be estimated by the sum of the individual contributions for that bin). Looking at the legends, next to the trigger names, inside the parentheses the respective turn-on points are displayed. Note that no contribution is observed by the lowest threshold trigger denoted as HLT 40 in the plot, which is justified due to the high phase space selection cuts imposed on the transverse momentum of the dijet pair. The detector level spectrums for the 2017 and 2018 can be found in Appendix F.

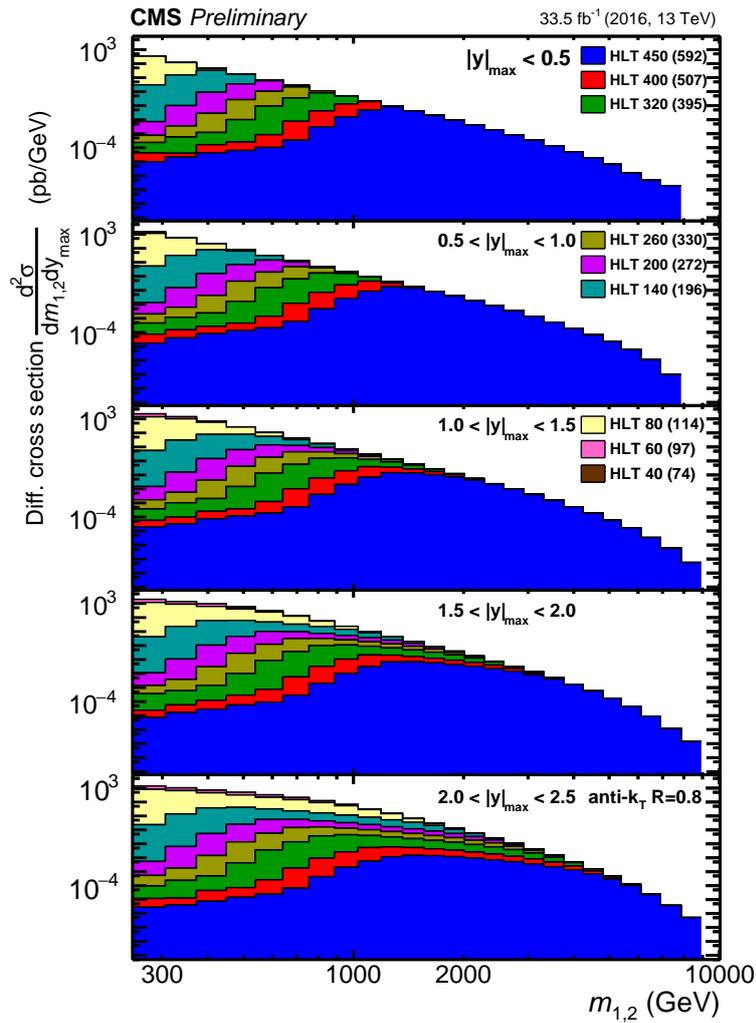
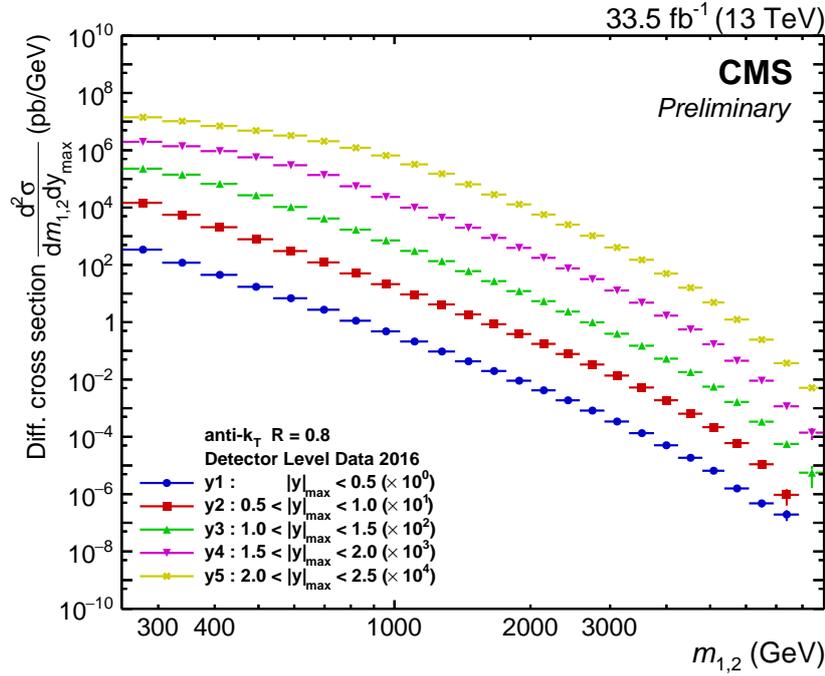


Figure 5.9: Detector level double differential cross section as a function of $m_{1,2}$ and $|y|_{max}$ for the 2016 dataset (top), and trigger contributions (bottom).

5.6 Unfolding

All measurements based on event counting inherently carry statistical uncertainty, meaning that repeating the experiment under identical conditions will not yield exactly the same results. This variation arises from the fundamental nature of random processes. Similarly, this element of randomness is also present in measurements performed by the CMS detector, such as the jet energy or transverse momentum. Due to the detector's finite response and resolution, the measured values are expected to deviate from the true values. A huge machine like this is actually susceptible to various sources that can lead to this effect like reconstruction inefficiencies, hardware failures, and pileup to name a few.

As a more realistic example, consider a jet produced at 1 TeV in the hard process. Upon colliding with the detector, its energy is deposited in the calorimeters, information which is afterwards collected and fed to the reconstruction algorithm. However, the reconstructed jet energy will be slightly shifted either to a higher or a lower value, an offset attributed to the detector resolution, which causes a misidentification of the jet's true energy value. From an event-level perspective, this effect manifests as a distortion in the measured spectrum when compared to its true distribution (or the generated spectrum in the case of MC). Specifically, the measured distribution appears shifted due to event migrations between neighboring bins. This occurs as a direct consequence of dividing the phase space into discrete bins in binned measurements.

Some of the events near mass (rapidity) bin edges will most likely migrate to neighboring bins because of their shifted reconstructed value. Each bin of the measured/smear spectrum contains events that have migrated in from adjacent bins, but also, is missing events that have migrated out of it. The purpose of *unfolding* is to correct for these distortions caused by the detector resolution in the measured cross sections (see Fig. 5.9), effectively restoring or *unsmearing* the detector level spectrum. Unfolded distributions offer two key advantages: (1) they can be directly compared to fixed order theoretical predictions, which do not account for detector effect to begin with, and (2) they allow for meaningful comparisons with unfolded results from other experiments, as detector-specific influences have been removed. A more detail insight in the methodology of unfolding can be found in Ref. [32, 33].

Two-dimensional unfolding is performed to account for event migrations happening between both mass and rapidity bins. In addition, effects like spurious reconstructions, background events and migrations inside the phase space due to noise or other fluctuations are considered as fakes entries, while the limited detector acceptance, reconstruction inefficiencies and smearing pushing events outside the phase space are treated as misses. The former, are subtracted from measured distributions prior to unfolding, while for the latter, dedicated correction factors are derived from MC and applied to unfolded distributions.

The task at hand is addressed according to the *matrix pseudo-inversion* method which requires the minimization of the following quantity:

$$\chi^2 = \min [(Ax + b - y)^T V^{-1} (Ax + b - y)] \quad (5.9)$$

where x represents the vector containing the desired unfolded distribution, corresponding to the true event spectrum, y the measured distribution at detector level, b the background contribution, A the probability matrix (or response matrix), essentially containing the probability of an event generated in bin i to be observed in the reconstructed bin j , and V the covariance matrix containing the detector level data statistical uncertainties. The unfolding procedure is performed using the TUnfold package [34]. While Tikhonov regularization is not explicitly applied, a carefully chosen binning scheme is implemented to ensure numerical stability and smooth behavior of the unfolding process, preventing the response matrix from becoming ill-conditioned.

5.6.1 Probability/Response matrix

The *probability matrix* (PM), a normalized version of the *response matrix* (PM), is constructed from simulation, specifically the Pythia sliced sample. The dijet mass $m_{1,2}$ is computed twice, once at the generator level and once at the reconstructed level. This calculation is performed only for jets that are successfully matched at both levels and remain within the defined phase space. Events that enter the phase space from outside are classified as fake entries, while those that exit the phase space are classified as miss entries.

In the matrix pseudo-inversion method employed here, the measured distribution is defined with twice the number of bins compared to the unfolded one (particle level). This choice is imposed from the start during the RM construction, ensuring the the final $m_{1,2}$ binning scheme (see Table 5.11) is preserved at the particle level. Additionally, an even coarser binning scheme is applied in the two outermost rapidity regions, where mass bins are merged in pairs, essentially reducing there the final number of bins two only half of the original.

This approach ensures that the *purity* and *stability* remain stable across all regions and stay above 50%. The former, quantifies how many events in a given reconstructed bin j originate from the same true bin i , while the latter, measures kind of the reverse effect, how many events in a given true bin i remain in the same reconstructed bin j . Their definitions are:

$$Purity^{(j)} = \frac{R_{jj}}{\sum_i R_{ji}} \quad (5.10)$$

$$Stability^{(i)} = \frac{R_{ii}}{\sum_j R_{ij}} \quad (5.11)$$

where R_{ii} , R_{jj} are diagonal elements of the RM, representing events that were generated and reconstructed in the same bin, $\sum_i R_{ji}$ the total number of reconstructed events in bin j , and $\sum_j R_{ij}$ the total number of generated events in bin i . A high purity and stability translates to minimal event migrations while a low value can point to strong event migrations causing implications in the unfolding procedure. These are shown in Fig 5.10 (left plot) as obtained from the 2016 Pythia sliced sample. Purity is displayed by the solid blue markers while stability by the empty red ones. It is observed that both of them indeed remain above 50% in all phase space regions, preserving a relatively smooth behavior. The coarser binning in terms of $m_{1,2}$ can be seen in action in the two forward rapidity regions.

In turn the probability matrix is presented in Fig. 5.11. The matrix is structured in two dimensions to illustrate event migrations between both mass and rapidity bins, from generator level to reconstructed level. Each cell represents the migration probability of events between these two stages, with the x -axis expressing the dijet mass at RECO level and the y -axis the dijet mass and GEN level. In one-dimensional unfolding scenario, only the diagonal elements would be relevant, one for each rapidity bin, corresponding to migrations between adjacent mass bins. However, since the analysis also accounts for migrations between neighboring rapidity bins, off-diagonal elements emerge, indicating significant event migrations in both dimensions. Matrices in the cells are diagonal illustrating that the majority of the events end up in the same reconstructed bin as the one that they were generated. This PM corresponds to A in Eq. 5.9. The probability matrices for the 2017 and 2018 Pythia sliced samples can be found in Appendix G.

Looking at the PM of Fig. 5.11 one can deduce whether regularization is needed by computing the *condition number*. For its calculation, first, the probability matrix needs to be diagonalized. Then a ratio is taken between the largest and the lowest of its eigenvalues. As a general principle, and according to the CMS statistics committee, a conditions number < 10 is considered good, implying that no regularization is needed and the overall procedure

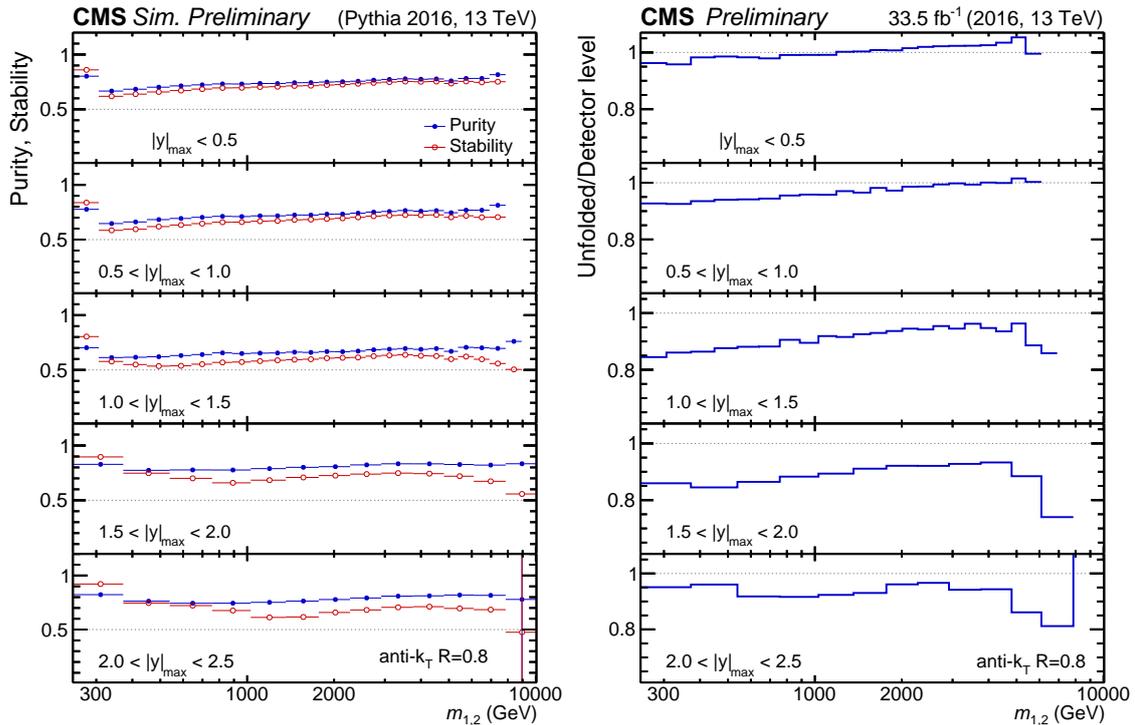


Figure 5.10: Purity and stability (left), and ratio between the unfolded and the detector level distributions (right) for Pythia 2016 slices.

is more or less stable. In our case, the value of the condition number is 3.16 for the 2016 Pythia slices, and is displayed in the top left corner of Fig. 5.11.

The obtained distribution, after unfolding, is compared to the detector level one in Fig. 5.10 (right plot). This comparison should only be taken as a qualitative figure of merit as it does not provide any strong conclusions on the procedure itself, it is only presented to illustrate the effect of unfolding on the initial distribution. The ratio remains mostly below unity and differences vary depending on the region, ranging from $\sim 1\%$ up to 30% in extreme cases. For a steeply falling spectrum such as the dijet mass one, the unfolded distribution is expected to be shifted leftward. This happens because lower mass bins typically contain more events than those in the higher end of the spectrum, so when the distribution is measured at detector level more events move to the right direction, than left, due to detector resolution. After unfolding this shift is partially reverted.

5.6.2 Closure Test and Backfolding

To validate the unfolding process and ensure that no bias is introduced in the results, a series of sanity checks are constructed to monitor the procedure. The primary objective is to verify the the RM has been correctly inverted and applied. Two complementary test address this verification from different perspectives: the *Closure Test* and the *Backfolding Test*, illustrated in Fig. 5.12, left and right, respectively.

In the Closure Test, the reconstructed dijet mass spectrum obtained from Pythia is treated as pseudo-data and is unfolded using the RM derived from same generator. The resulting distribution is then compared to the original generator level distribution within the simulation sample. Their ratio should be exactly at 1, indicating perfect agreement. Similarly, in the Backfolding Test, the generator level distribution is folded (or smeared), using the same RM, to construct the “detector level” spectrum. This folded spectrum is then compared to the originally reconstructed distribution. Again, a perfect agreement

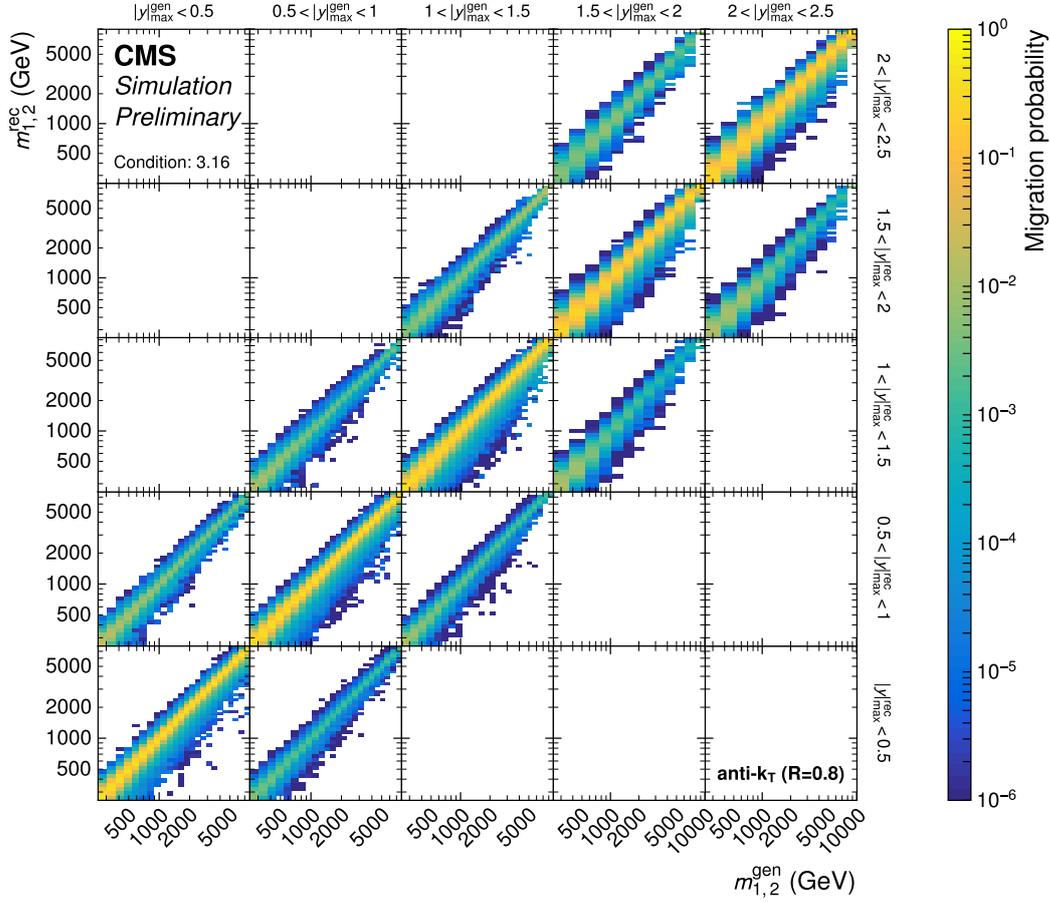


Figure 5.11: Two-dimensional probability matrix for 2016 Pythia slices. Event migration probability between rapidity and mass bins from generator to reconstructed level.

confirms the consistency of the response matrix and the unfolding process.

5.6.3 Bottom Line Test

Another sanity check is provided by the *Bottom Line Test*, which tests whether the data to simulation agreement remains the same after the unfolding procedure. This is presented in Fig. 5.13 (left plot), where two ratios are compared with each other. In blue, the data detector level distribution is compared to the reconstructed one obtained from Pythia. It represents the difference between data and simulation before unfolding. In red, the same comparison is displayed but this time after the unfolding, i.e., the data distribution at particle level is compared to the one from Pythia at generator level. It is observed that their shapes display a very good agreement across the whole phase space ensuring that no bias is introduced after unfolding. Paying close attention to the binning, it can be seen that for the first three rapidity bins two reconstructed bins correspond to one bin at generator level. This 2:1 ratio becomes 4:1 in the two forward regions, where a coarser binning is applied at particle level.

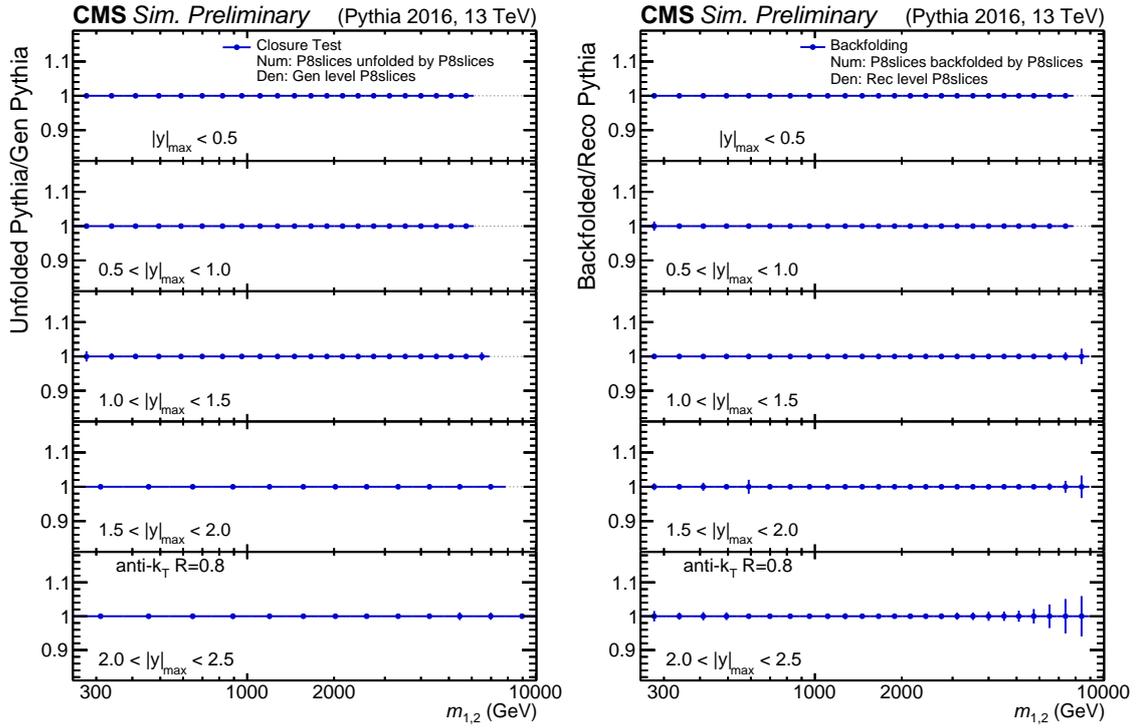


Figure 5.12: Closure test (left) and backfolding (right) for 2016 Pythia slices.

5.6.4 Fake and Miss rates

As previously discussed, estimating fake and miss event rates is crucial for jet-based differential cross section measurements, particularly for unfolding. These rates, are given in Fig. 5.14 as obtained from the 2016 Pythia sliced sample.

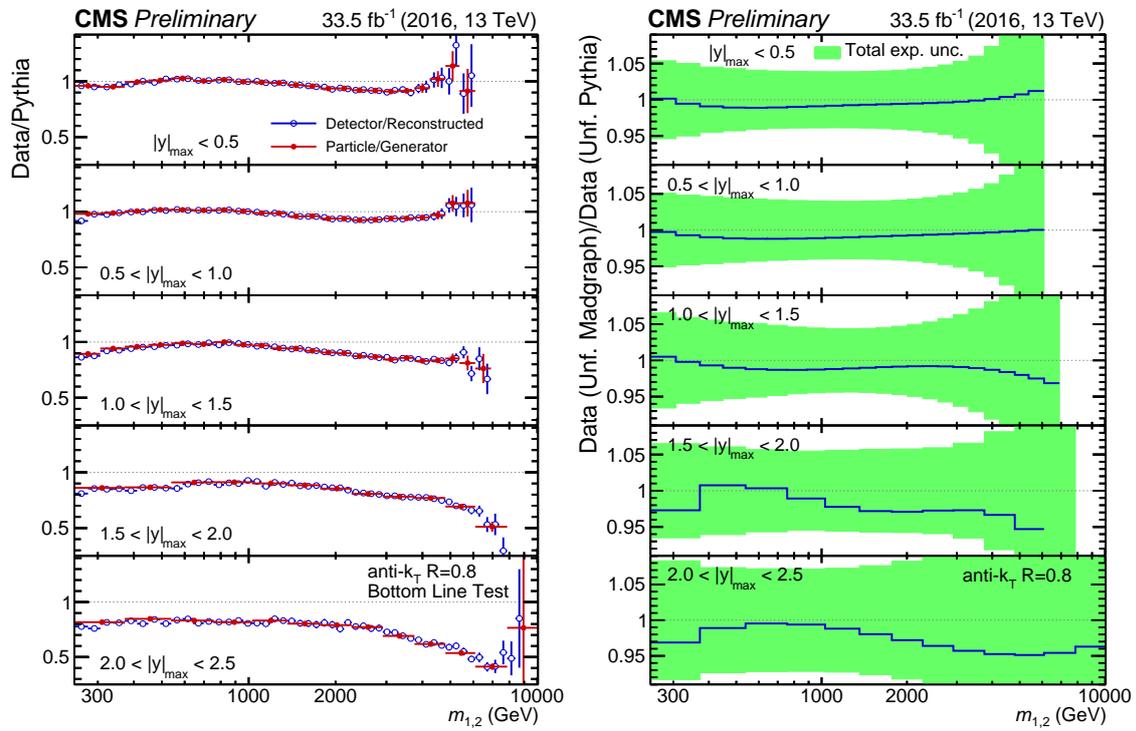


Figure 5.13: Bottom Line Test for 2016 Pythia slices (left). Model dependence (right), the 2016 Madgraph sample used in addition to Pythia.

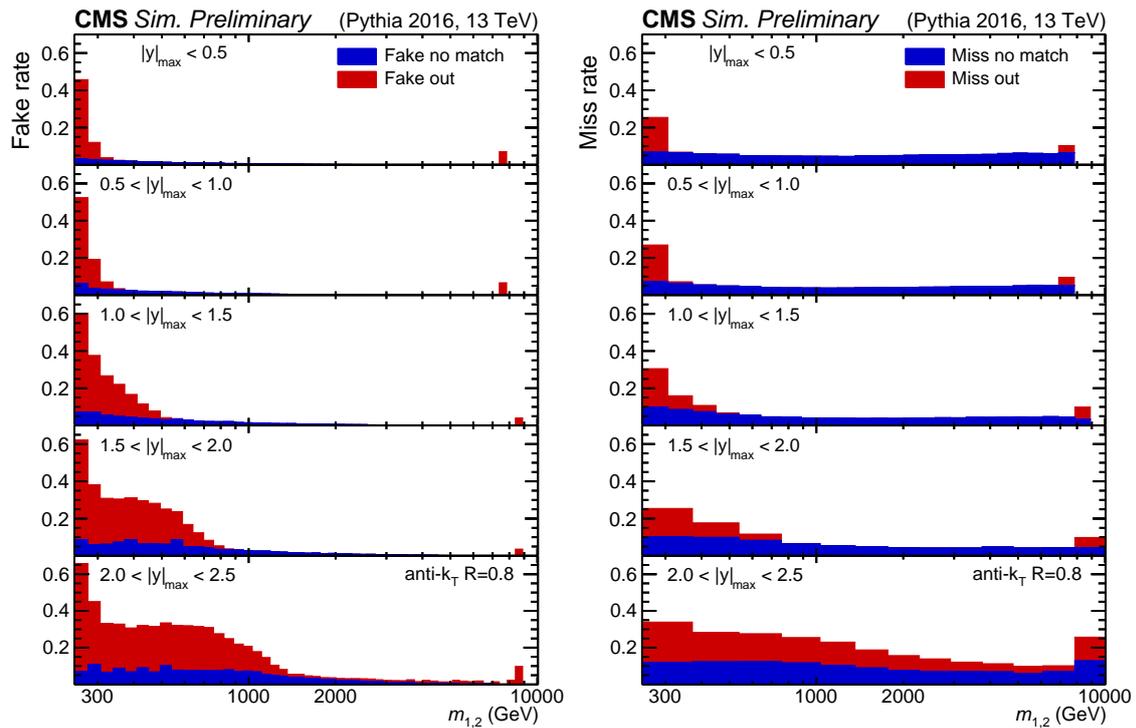


Figure 5.14: Fake (left) and miss (right) rates for 2016 Pythia Slices.

The fake rate (left plot) is shown as a function of dijet mass across all rapidity regions. A significant increase is observed in the low mass bins, which is expected since this region is dominated by pileup, soft radiation and detector noise. Additionally, the fake rate rises in the forward rapidity regions due to the reduced detector resolution.

The miss rate (right plot) exhibits a similar enhancement at low dijet mass but also shows a notable increase at the high end of the spectrum. This effect is mostly attributed to reconstruction inefficiencies, particularly in the forward region.

Furthermore, increased rates near the phase space boundaries can be explained by event migrations between neighboring bins. Both rates are decomposed into two contributions: *no match* refers to dijet events that lack a corresponding match between GEN and RECO levels, while *out* represents events where the dijet pair is matched but falls outside the defined phase space at either level. Lastly, attention should be given to the binning convention used to express these rates. The fake rate is measured at the RECO level, whereas the miss rate is determined at the GEN level, reflected with a coarser binning scheme.

5.6.5 Model dependence

The choice of the default MC sample used for unfolding the data, is somewhat arbitrary, as no single option can be considered definitively superior. Differences between simulated samples arise because each MC generator employs its own algorithms to produce events and model physical processes. To assess the model dependence on the choice of MC, an alternative sample, the 2016 Madgraph sliced sample, is used for unfolding. The data are unfolded again, this time using a RM derived from Madgraph. By comparing the resulting distribution to the one obtained with Pythia-based unfolding, the impact of the MC choice can be estimated from their difference. The result is shown in Fig.5.13 (right plot), where the blue line represents the model dependence. The green band surrounding it illustrates the total experimental uncertainty. The effect of picking a particular MC is covered by the experimental uncertainties. No model uncertainty is considered for the 2017 and 2018 datasets.

5.7 Experimental uncertainties

This section is dedicated on the evaluation of the experimental uncertainty sources associated with the differential dijet cross section. The following subsections provide a detailed description of each uncertainty source. A breakdown of the individual contributions is presented in Fig 5.15, where the relative uncertainty (%) is shown as a function of the dijet mass for each $|y|_{max}$ bin. A logarithmic scale is used on the plot's y -axis to enhance visibility and facilitate differentiation between uncertainty sources. Each source is represented by a distinct line color or style.

Uncertainties that are originally asymmetric are *symmetrized* by averaging the up and down variations for each source. Among all sources, the jet energy corrections (JES) is the most dominant one across all rapidity bins. The gray band represents the total experimental uncertainty, obtained by combining all statistical and systematic uncertainties in quadrature. The total uncertainty ranges from approximately 5% at low dijet masses and increases to 40% in the higher end of the spectrum.

All uncertainty contributions are computed within the DAS framework as a separate step performed parallel to the unfolding procedure. Data natured corrections, like JES, are applied on the detector level distribution one at a time. Each variation of the spectrum is then unfolded by the RM obtained from simulation. Differences to the nominal distribution define the respective uncertainty for each source. Similarly, corrections applied on the MC

samples, like JER, result in variations that produce reshaped RMs. In this case, nominal data distributions are unfolded once for each modified RM. The resulting deviations from the nominal unfolded distribution quantify the associated uncertainty.

In summary, uncertainties are estimated by shifting experimental parameters by $\alpha \pm 1\sigma$, allowing for independent calculation of the up and down variations of each systematic source. This inherently leads to asymmetric uncertainties. To standardize the uncertainty estimates, all uncertainties are first symmetrized by averaging the up and down variations of each source. They are then smoothed by fitting them with Chebyshev polynomials. The final smoothed relative uncertainties are presented in Fig. 5.15. Experimental uncertainties for the 2017 and 2018 can be found in Figs. H.1 and H.2 respectively of Appendix H. Note that for 2017 and 2018 the raw uncertainties are given just by averaging them, they are not smoothed.

5.7.1 Statistical uncertainty

Data samples contain a finite number of events, limiting the available statistics for each observable bin. Consequently, cross section measurements from such samples suffer from statistical uncertainties, particularly in certain regions of the phase space. In the case of the double differential dijet cross section, statistical limitations become significant at the higher end of the spectrum, where the probability of producing a highly energetic dijet pair, at the scale of several TeV , is substantially lower.

To quantify the statistical uncertainty in each mass bin, it is assumed that event counts follow a Poisson distribution, where the standard deviation is given by the square root of the quadratic sum of all event weights within a bin. The resulting statistical uncertainty is represented by the black solid line in Fig. 5.15. It starts at the sub-percent level in the low mass region and increases up to 30% at high dijet mass.

While the statistical uncertainties are naturally uncorrelated at the detector level, the situation changes after unfolding. Due to event migrations induced by the detector resolution, correlations arise when events are redistributed back to their original neighboring bins. These non-trivial correlations are propagated through unfolding via the covariance matrix V , as found in Eq. 5.9. Mass bins that end up with a statistical uncertainty larger than 50% are dropped, ensuring a moderate definition of the cross section. The dijet mass ranges found in Fig. 5.15 are the final ranges that theoretical predictions are produced and later compared to data.

5.7.2 Jet energy scale uncertainty

The jet energy corrections (JECs), discussed in Section 5.3.3, are perhaps the most important corrections of them all. As previously mentioned, JECs correct for a plethora of phenomena including the nonlinear response of the detector, differences between data and MC, pileup, and even the change in response in gluon versus quark initiated jets. Actually there are 27 individual JES-related corrections applied to data [35, 36] that when taking their uncertainties in quadrature result in the pink solid line of Fig. 5.15. It is the most dominant uncertainty, ranging between 5% and 20%. The up and down variations of the individual JES sources for 2016 can be found in Figs. H.3-H.9 of Appendix H. Each source represents a fully correlated uncertainty between all observable bins and is considered to be independent of the other contributions.

5.7.3 Jet energy resolution uncertainty

The same rule applies again; invasive processes come with a cost, usually in the form of an uncertainty. By mimicking the jet energy resolution observed in data, into the simulation,

a bias is introduced inside the measurement. Scale factors are provided with an associated uncertainty, which are propagated through the unfolding. The resulting JER uncertainty is illustrated with the teal solid line in Fig. 5.15 which remains mostly at the sub-percent level, with the exception of the very forward rapidity region that increases approximately to 5%. This uncertainty is considered fully correlated between all data points.

5.7.4 Luminosity uncertainty

Equation 5.8 transforms the effective number of events in a particular $|y|_{max}$ and $m_{1,2}$ bin into a cross section value. The integrated luminosity, which is used as a scaling factor in this process, plays an important role since it affects the global shape of the distribution. The corresponding uncertainty presented with a yellow solid line in Fig. 5.15 is estimated at 1.2% for 2016 [37] (2.3% for 2017 [38] and 2.5% for 2018 [39]) across all observable bins and is fully correlated.

5.7.5 Unfolding uncertainties

Monte Carlo samples are likewise susceptible to their own statistical uncertainty, as do the data, due to their finite size. The corresponding uncertainty is referred to as “Unf. (stat.)” in Fig. 5.15, illustrated by the black dashed line. This source arises from the limited statistics of the simulated samples used to construct the RMs. The impact of the uncertainty is propagated through the unfolding and correlations are contained in the corresponding covariance matrix. It remains at the sub-percent level across all bins, with the exception of some high mass bins in the forward region where it increases to a few percent and exceptionally reaches the value of 10% for a single bin.

Another uncertainty related to unfolding is that of the model dependence, denoted as “Unf. (model)” in Fig. 5.15, displayed by a blue dashed-dotted line. This uncertainty is introduced to the measurement upon making the choice of picking a particular MC to construct the RMs for unfolding. Its magnitude is estimated by comparing the nominal data distributions at particle level, unfolded by Pythia, to the ones obtained after unfolding via Madgraph. It results in a symmetric uncertainty that is taken as fully correlated between all bins. In the central region it remains relatively flat $\sim 1\%$, while in the two forward regions it displays an increase of a few percent, reaching 10% in maximum cases.

5.7.6 Other uncertainties

A cocktail of four individual uncertainty sources are grouped into a single contribution due to their small size, represented by the red solid line, labeled as “Other” in Fig. 5.15. A 5% normalization uncertainty is introduced before unfolding with the call of the TUNFOLD package because of the background subtraction of fake entries. Another 5% normalization uncertainty is introduced by hand due to the correction factors applied to account for the inefficiency caused by miss entries. Additionally, a prefiring uncertainty is incorporated to account for inefficiencies caused by triggers that were firing prematurely (see Section 5.3.3). Lastly, an uncertainty related to the pileup profile reweighting correction that is estimated by varying the total inelastic pp cross section used in the procedure around its associated uncertainty of 4.6% [29].

In summary, the “Other” uncertainty remains at the level of 1 – 2% in all phase space regions with the exception of the very forward rapidity bin where it ranges between 3% and 5%. This increase is related to the prefiring uncertainty that dominates that region, in contrast to the rest of the phase space, which is mostly negligible. Note again that, the prefiring correction and the corresponding uncertainty is only relevant for 2016 and 2017, since for 2018 there was no such effect observed at the detector.

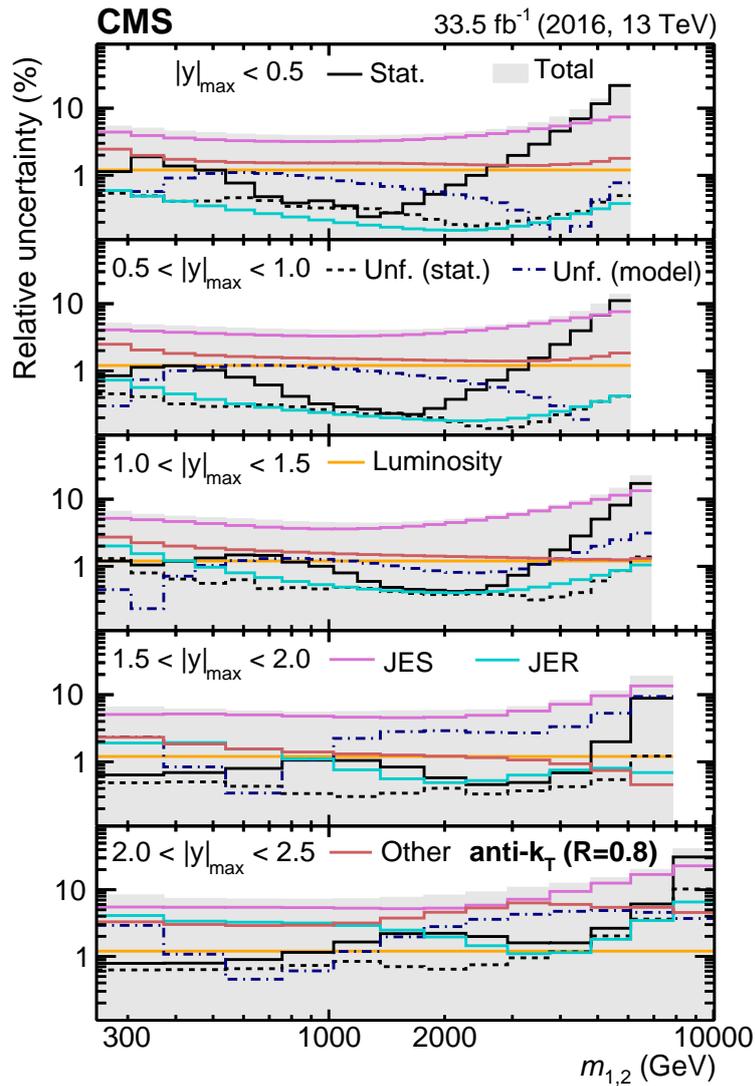


Figure 5.15: Relative uncertainty (%) for all experimental sources on the measured dijet mass cross section from the 2016 dataset. Individual contributions are explained in Section 5.7. The abbreviation “Unf.” refers to the unfolding uncertainties. The total uncertainty is estimated as the sum in quadrature of all statistical and systematic uncertainty sources.

Bibliography

- [1] M. Cacciari and G.P. Salam and G. Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, Apr 2008. <https://dx.doi.org/10.1088/1126-6708/2008/04/063>.
- [2] G. Aad et al. |ATLAS Collaboration|. Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector. *The European Physical Journal C*, 71(2):1512, 2011. <https://doi.org/10.1140/epjc/s10052-010-1512-2>.
- [3] S. Chatrchyani et al. |CMS Collaboration|. Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7\text{ TeV}$. *Physics Letters B*, 700(3):187–206, 2011. <https://www.sciencedirect.com/science/article/pii/S0370269311005338>.
- [4] S. Chatrchyan et al. |CMS Collaboration|. Measurements of differential jet cross sections in proton-proton collisions at $\sqrt{s} = 7\text{ TeV}$ with the CMS detector. *Physical Review D*, 87(11), Jun 2013. <http://dx.doi.org/10.1103/PhysRevD.87.112002>.
- [5] G. Aad et al. |ATLAS Collaboration|. Measurement of dijet cross-sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector. *Journal of High Energy Physics*, 2014(5):59, 2014. [https://doi.org/10.1007/JHEP05\(2014\)059](https://doi.org/10.1007/JHEP05(2014)059).
- [6] M. Aaboud et al. |ATLAS Collaboration|. Measurement of inclusive jet and dijet cross-sections in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector. *Journal of High Energy Physics*, 2018(5):195, 2018. [https://doi.org/10.1007/JHEP05\(2018\)195](https://doi.org/10.1007/JHEP05(2018)195).
- [7] T. VanLaer. CMS Computing Model. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookComputingModel>. Accessed: 2025-02-14.
- [8] A. Pozdnyakov. Data Formats and Data Tiers. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookDataFormats>. Accessed: 2025-02-14.
- [9] A. Pozdnyakov. Locating Data Samples. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookLocatingDataSamples>. Accessed: 2025-02-17.
- [10] CMS Collaboration. Data Aggregation System (DAS). Available: <https://cmsweb.cern.ch/das/>. Accessed: 2025-02-17.
- [11] A. Brinkerhoff. How to work with files for Good Luminosity Sections in JSON format. Available: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideGoodLumiSectionsJSONFile>. Accessed: 2025-02-17.
- [12] CMS Collaboration. Monte Carlo production tools. Available: <https://cms-pdmv.gitbook.io/project>. Accessed: 2025-02-18.
- [13] T. Sjöstrand et al. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, 2015. <https://www.sciencedirect.com/science/article/pii/S0010465515000442>.
- [14] V. Khachatryan et al. |CMS Collaboration|. Event generator tunes obtained from underlying event and multiparton scattering measurements. *Eur. Phys. J. C*, 76(3):155, 2016. <https://doi.org/10.1140/epjc/s10052-016-3988-x>.

- [15] F. Maltoni and T. Stelzer. MadEvent: automatic event generation with MadGraph. *Journal of High Energy Physics*, 2003(02):027, Feb 2003. <https://dx.doi.org/10.1088/1126-6708/2003/02/027>.
- [16] J. Alwall and M. Herquet and F. Maltoni and O. Mattelaer and T. Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):128, 2011. [https://doi.org/10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128).
- [17] A. M. Sirunyan et al. |CMS Collaboration|. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1):4, 2020. <https://doi.org/10.1140/epjc/s10052-019-7499-4>.
- [18] DAS Analysis System. Available: <https://gitlab.cern.ch/cms-analysis/genera1/DasAnalysisSystem>. Accessed: 2025-02-19.
- [19] CMS Offline Software. Available: <https://github.com/cms-sw/cmssw>. Accessed: 2025-02-19.
- [20] ROOT Data Analysis Framework. Available: <https://root.cern/>. Accessed: 2025-02-19.
- [21] E. Tziaferi. Jet Identification. Available: https://twiki.cern.ch/twiki/bin/view_auth/CMS/JetID. Accessed: 2025-02-19.
- [22] CMS JME group. Jet Veto Maps. Available: https://cms-jerc.web.cern.ch/Recommendations/#2022_1. Accessed: 2025-02-22.
- [23] CMS JME group. JECDatabase on GitHub Jet Veto Maps. Available: https://github.com/cms-jet/JECDatabase/tree/master/jet_veto_maps. Accessed: 2025-02-22.
- [24] V. Khachatryan, et al. |CMS Collaboration|. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014, Feb 2017. <https://dx.doi.org/10.1088/1748-0221/12/02/P02014>.
- [25] CMS JME group. Jet Energy Resolution and Corrections. Available: <https://cms-jerc.web.cern.ch/JEC/#factorized-approach>. Accessed: 2025-02-21.
- [26] G. Agarwal. Recommended Jet Energy Corrections and Uncertainties For Data and MC. Available: <https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC>. Accessed: 2025-02-21.
- [27] Reweighting recipe to emulate Level 1 ECAL and Muon prefiring. Available: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1PrefiringWeightRecipe>. Accessed: 2025-02-24.
- [28] R. Pradhan. Jet Energy Resolution. Available: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetResolution>. Accessed: 2025-02-24.
- [29] A. M. Sirunyan et al. |CMS Collaboration|. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2018(7):161, 2018.
- [30] BRIL Work Suite. Available: <https://cms-service-lumi.web.cern.ch/cms-service-lumi/>. Accessed: 2025-02-27.

- [31] A. Hayrapetyan et al. [CMS Collaboration]. Measurement of multidifferential cross sections for dijet production in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$. *The European Physical Journal C*, 85(1), Jan 2025. <http://dx.doi.org/10.1140/epjc/s10052-024-13606-8>.
- [32] L. Moureaux. Recommendations on Unfolding. Available: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/ScrecUnfolding>. Accessed: 2025-03-03.
- [33] G. Cowan. *Statistical Data Analysis*. Oxford University Press, Oxford, UK, 1998.
- [34] S Schmitt. TUnfold, an algorithm for correcting migration effects in high energy physics. *Journal of Instrumentation*, 7(10):T10003, Oct 2012. <https://dx.doi.org/10.1088/1748-0221/7/10/T10003>.
- [35] CMS Collaboration. *Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018*. , 2020. <https://cds.cern.ch/record/2715872>.
- [36] G. Agarwal. Jet energy scale uncertainty sources. Available: <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources>. Accessed: 2025-03-05.
- [37] A.M. Sirunyan et al. [CMS Collaboration]. Precision luminosity measurement in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$ in 2015 and 2016 at CMS. *The European Physical Journal C*, 81(9), Sep 2021. <http://dx.doi.org/10.1140/epjc/s10052-021-09538-2>.
- [38] CMS Collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13\text{TeV}$* . , 2018. <https://inspirehep.net/literature/1677076>.
- [39] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\text{TeV}$* . , 2019. <https://inspirehep.net/literature/1736713>.

Chapter 6

FIXED ORDER THEORY PREDICTIONS

6.1 Introduction

The core of this analysis lies in comparing data distributions to fixed order theory predictions. This comparison serves as a critical mechanism for both experimental and theoretical aspects. Discrepancies can drive refinements in either domain, prompting improvements in theoretical modeling or experimental techniques, while strong agreement provides a foundation for extracting fundamental parameters, such as the strong coupling constant, and for constraining the proton's partons distribution functions (PDFs).

Theoretical predictions are calculated at NNLO in perturbative QCD (pQCD). These are obtained with the NNLOJET program (revision 5918) [1], using the APPLFAST interface (version 0.0.46) [2, 3] to the FASTNLO package (version 2.3) [4, 5]. The FASTNLO package provides interpolation grids that can be combined with alternative PDF sets and renormalization μ_R and factorization μ_F scales to obtain the final prediction. It is common practice that these two scales are chosen to the same value, a value that should express the order of the process's hard scale under investigation. For this reason, they were both chosen to be equal to the dijet mass $\mu_R = \mu_F = m_{1,2}$, as done in Ref. [6]. The statistical precision of the predictions ranges between 0.5% and 2% depending on the phase space region.

As a representative plot, Fig. 6.1 left, shows the theoretical predictions for the double differential cross section as a function of $m_{1,2}$ in the $1.0 < |y|_{max} < 1.5$ bin (taken from Ref. [7]). The top panel contains the cross section itself at hadron level as combined with the CT18NNLO PDF set [8]. Different markers styles and colors illustrate the calculation at different orders in pQCD, with LO in green, NLO in blue, and NNLO in red. The band around the points corresponds to the scale uncertainty (discussed later), which is reduced for higher orders. While the bottom panel shows the ratios to the LO prediction, also known as *K factors*. The NNLO scale uncertainty band overlaps with the NLO one across all $m_{1,2}$ bins showing a good perturbative convergence.

In the bottom panel the NLO and NNLO *K factors*, in blue and red respectively, display a sharp increase for lower dijet mass values. According to $m_{1,2} \sim 2p_{T,1}p_{T,2} \cosh(\Delta y)$ events with low- p_T jets are eventually rejected as they are found at outer rapidities. This is caused by the minimum p_T requirements imposed on the leading jets essentially restricting the accessible phase space for LO processes in favor of higher-order contributions [7].

Figure 6.1 right, presents the *K factors* for all rapidity bins. This time, the ratios compare predictions at a given perturbative order to those at the preceding order. Specifically, the NLO/LO ratio is shown in red, while the NNLO/NLO ratio is depicted in blue. The same effect discussed earlier is observed here. Focusing on the NLO/LO *K factor*, one can see that for higher rapidity values, an increasing number of mass bins in the lower part

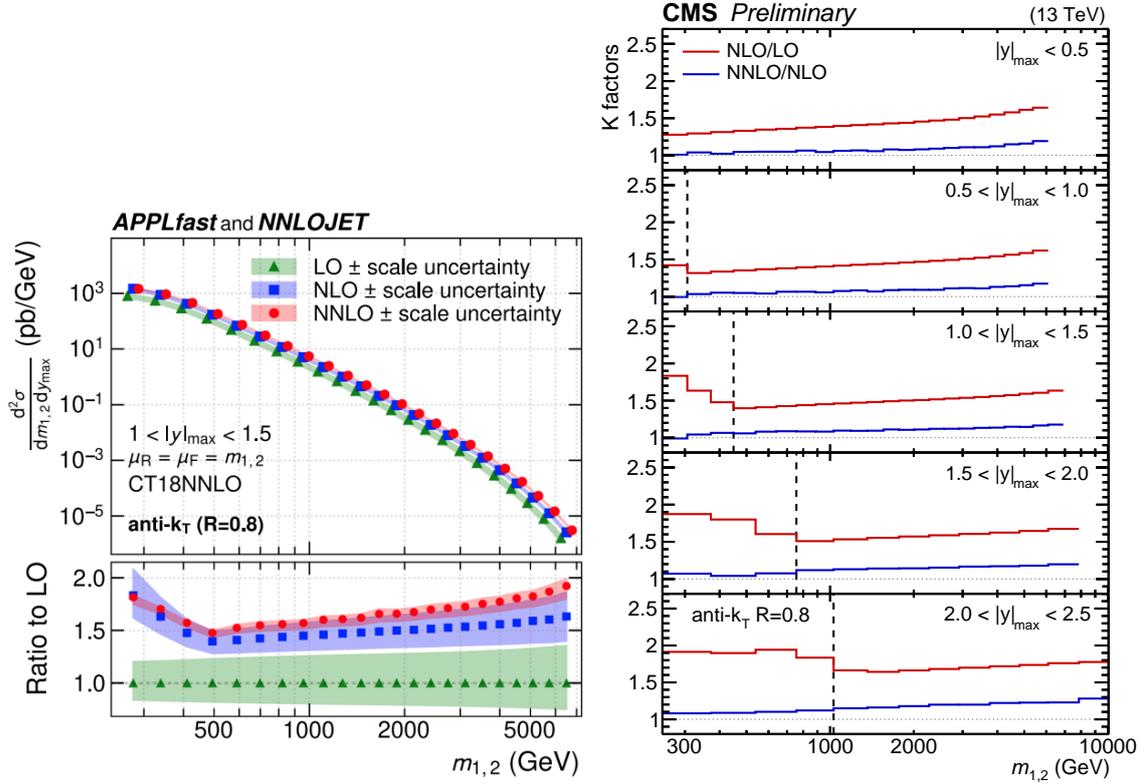


Figure 6.1: Theoretical prediction for the double differential cross section as a function of $m_{1,2}$ in $1.0 < |y|_{max} < 1.5$ for different orders in pQCD [7] (left). Prediction k factors, i.e. ratio between the same prediction at different orders in pQCD, for all rapidity bins (right).

of the spectrum exhibit a steep rise, confirming that the phase space limitations become more pronounced in that region. Vertical black dashed lines indicate the regions below which this effect is observed in each rapidity bin.

NNLO predictions are based on the leading-color and leading-flavor-number approximations [6, 9]. Several studies have been performed that examine the subleading-color contribution at NLO and conclusions vary depending on the observable [10–13]. So far, no verdict has been reached regarding their impact on the observable presented here.

6.2 Theory corrections

6.2.1 Non-perturbative corrections

Theoretical predictions are produced in two distinct stages, as discussed in Section 4.4 and described by Eq. 4.8. First, the Matrix Element (ME) is computed at a given order in pQCD. This result is then combined with a PDF set of the same order, to incorporate non-perturbative, long-distance effects arising from the initial state protons.

However, the direct output of this procedure is at parton level, meaning that final state particle interactions are not yet included. For a meaningful comparison with measured cross sections, theoretical predictions must be corrected for non-perturbative (NP) effects, specifically for multi-parton interactions and hadronization. Only after, applying these corrections do both experimental and theoretical cross sections correspond to the same level, the particle level, allowing for a direct comparison.

Non-perturbative correction factors are derived using MC event generators. A set of

eight different event generators is utilized, with samples produced via the RIVET toolkit, as discussed in Section 4.6.2. For a comprehensive overview of the individual MCs, their tunes, and algorithms, Table 4.2 provides relevant details. For each MC, only the generator level information is produced. The production process is carried out twice: once with the full jet evolution, incorporating parton shower (PS), multi parton interactions (MPI), and hadronization (HAD); and once with MPI and HAD turned off, isolating the perturbative component.

To extract non-perturbative corrections, correction factors are formed by taking the ratio of cross sections obtained from the two different setups:

$$C_{NP} = \frac{\sigma^{PS+MPI+HAD}}{\sigma^{PS}} \quad (6.1)$$

where the numerator represents the cross section estimate at particle level, including the full process, and the denominator corresponds to the cross section obtained by turning off the MPI and HAD processes, leaving only the PS contribution. This formulation ensures that the perturbative contribution effectively “cancels out”, isolating the NP component in the correction factor C_{NP} .

The calculation is repeated for all MC configurations, as estimated from Eq. 6.1, with individual results presented in Figs. I.1 and I.2 of Appendix I. Physical processes are expected to follow a smooth behavior. For this reason, obtained values are fitted, so that any statistical fluctuations are absorbed. The fit function is given below:

$$C_{NP}(m_{1,2}) = A + B * m_{1,2}^C \quad (6.2)$$

Under usual circumstances, fitting is performed in the entire dijet mass range. However, the present situation demands a more refined approach. Except for the very central rapidity bin ($|y|_{max} < 0.5$), all others exhibit a sudden drop in the C_{NP} factor at low mass values. Comparing the bins that show this behavior with those affected by the phase space restriction, displayed in the K factors of Fig. 6.1 (right), a clear connection is revealed between the two effects. In fact, the observed drop in the NP correction factors arises from the same underlying cause. The fit is then performed only in the unaffected mass bins, while for the affected lower mass bins the correction factor is taken by value from the bin content itself, without performing any fit.

Fit results are collectively presented in the left plot of Fig. 6.2. The NP correction factor is displayed as a function of the dijet invariant mass. Each cell inside the plot, again corresponds to a different rapidity bin, while estimations from individual MCs are illustrated by colored lines. Solid lines represent MC event generators that compute the cross section at LO in pQCD, while dashed lines correspond to those performing the calculation at NLO. Vertical black dashed lines indicate the region above which a fit is performed and below which factors are taken by value from the bin content.

Each MC configuration is armed with different settings and models physical processes in varying ways. This diversity necessitates using a broad set of them, not only to determine a central value for the factors but also estimate the associated uncertainty. At low to mid $m_{1,2}$ bins, NLO MCs (dashed lines) are largely contained within an envelope formed by the LO ones (solid lines). However, this trend reverses in the high dijet mass region, where NLO estimations now appear to enclose the LO ones. This effect is primarily driven by limited statistics on the NLO side, as event generation in our setup was constrained on that front, whereas LO samples were produced without a strict restriction on the number of events; an effect clearly displayed in Figs. I.1 and I.2.

Looking at Fig. 6.2 (left), two cases can be distinguished. An envelope can be formed by considering the factors either from all MCs (both LO and NLO) or by considering just the LO ones. This comparison, between the two envelopes is illustrated in Fig. 6.2

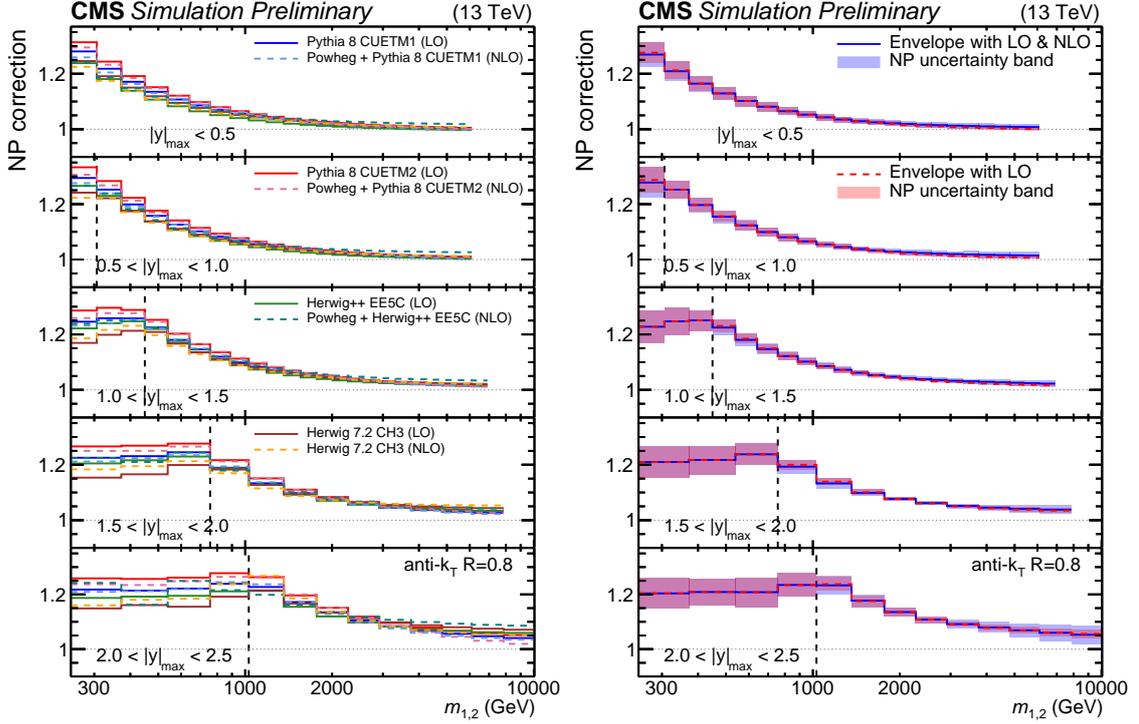


Figure 6.2: Non-perturbative correction factors from all eight MC generators (left), and final factors with their uncertainty (right).

(right), where the red envelope corresponds to LO estimations and blue to the LO plus NLO ones. Envelopes are formed by taking the maximum and minimum C_{NP} value in a given mass bin, while the central NP correction is defined as the middle point of the envelope, represented by a dashed (solid) line for LO (LO+NLO). The NP uncertainty, due to the model dependence of the generators, is taken as half the envelope spread, resulting in a symmetric uncertainty. Correction factors are nearly identical between the two cases. Uncertainties for low mass bins appear very similar, while in the higher end of the spectrum the LO+NLO uncertainty is overestimated due to the increased fluctuations caused by limited statistics. To remain on the conservative side the final NP correction factors and their uncertainty is taken as derived by the envelope considering both LO and NLO MCs. NP factors are mostly dominant in the lower mass region, where MPIs dominate the event content and reach values of the order of 20%, while for higher energies they approach unity, as MPI and HAD contributions largely cancel out.

6.2.2 Electroweak corrections

At the TeV scale, electroweak (EW) contributions become important and dedicated corrections need to be applied to the final predictions. At the higher end of the spectrum, additional Feynman diagrams from EW processes (virtual exchange of soft or collinear W or Z bosons) contribute to the same final state as those originating purely from QCD interactions. These contributions are estimated and applied to the partonic cross section. The final factors are presented in Fig. 6.3 [14], where different colors and marker styles illustrate the respective correction for each $|y|_{max}$ bin. The corrections exhibit a strong dependence on $m_{1,2}$, becoming increasingly important above $1 TeV$, where they reach values of 15%. The factors also vary with rapidity, showing an upward trend for the two lowest $|y|_{max}$ bins, but transition to a downward trend in more forward regions. No uncertainty is attributed to this correction as their value was estimated to be negligible compared to the

experimental and other theory uncertainties in the respective regions.

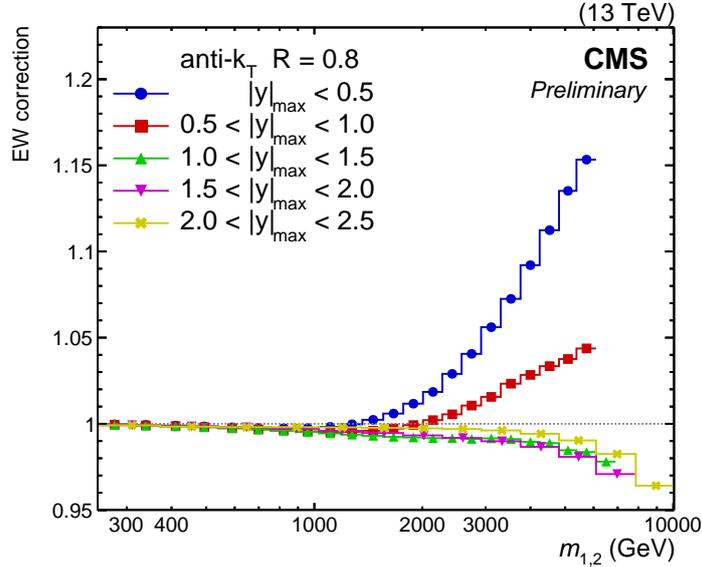


Figure 6.3: Electroweak correction factors [14].

6.3 Theory uncertainties

Uncertainties follow right after whenever a restrictive choice is being made.

6.3.1 Scale uncertainty

The scale uncertainty is introduced to account for missing higher order corrections in the calculation of the ME at NNLO. The six-point variation method is applied, a recipe taken from Ref. [15–17]. The predictions are derived six times, aside from the nominal ($\mu_R = \mu_F = m_{1,2}$) variation, for different pair values of μ_R and μ_F :

$$(\mu_R/m_{1,2}, \mu_F/m_{1,2}) = (1/2, 1/2), (1/2, 1), (1, 1/2), (2, 1), (1, 2), (2, 2) \quad (6.3)$$

Resulted distributions form an envelope with boundaries defined by the maximum and minimum values between the respective six distributions for each mass bin. The scale uncertainty is then defined as the difference between the maximum (minimum) value of the envelope and the nominal one, finally, resulting in an asymmetric uncertainty.

6.3.2 PDF uncertainty

The matrix element is combined with a reference PDF set to obtain the final prediction at parton level. As previously mentioned, the default choice for the analysis presented here is the CT18 [8], derived from global fits while keeping the value of the strong coupling constant fixed at $\alpha_S(m_Z) = 0.118$. An asymmetric uncertainty is attributed to the PDF choice, calculated as 68% confidence interval according to the prescriptions in Ref. [8].

6.3.3 NP uncertainty

A symmetric NP uncertainty to account for the model dependence in deriving the relevant correction factors that incorporate the non-perturbative component to the FO predictions. Their derivation is discussed in Section 6.2.1.

6.4 Data to theory comparisons

Finally, the stage has been set for a meaningful comparison between experimental results and theoretical predictions. At this point, data have been unfolded and fixed order predictions have been corrected, so that the differential cross section calculation from both sides is brought to the same level, the particle level.

The top plot in Fig. 6.4 displays the double differential cross section as estimated by the 2016 dataset as a function of the dijet invariant mass. Blue markers with different styles illustrate the cross section values for each rapidity region. On top, fixed order predictions are overlaid, represented by the red solid lines. Theory is obtained for $\mu_R = \mu_F = m_{1,2}$, combined with the CT18 reference PDF set and corrected for extra EW radiation, and NP effects. To increase visibility among distributions, each of them is scaled by a factor in powers of 10^n , $n = 0, 1, 2, 3, 4$. The spectrum falls rapidly, over multiple orders of magnitude, where a good agreement is displayed between the experimental results and the theory across the entire $m_{1,2}$ phase space. The final dijet mass reach spans from 249 GeV to well beyond several TeV , depending on the rapidity bin. Specifically, it extends up to 6094 GeV for the two central rapidity regions, 6908 GeV for the third bin, and reaches 7861 GeV and 10050 GeV for the two forward bins, respectively. Nonetheless, y -axis being in the logarithmic scale it is not ideal for a detailed inspection of their agreement.

To best validate the agreement between data and theory, their ratio is shown in the bottom plot of Fig. 6.4. The purpose of this plot is twofold. First, it represents the data to theory ratio, depicted by black markers. Second, it acts as a template for comparing predictions derived from different PDF sets against the reference prediction from CT18. The comparison is made for each rapidity bin separately, distributed among different cells within the same plot.

Vertical black lines around the markers illustrate the relative data statistical uncertainty, while the yellow band enclosing them shows the relative total experimental uncertainty. Around unity the relative total theory uncertainty is displayed with the cyan colored band, calculated as the square root of the quadratic sum of the scale, PDF, and NP uncertainty sources. A good agreement is observed in most phase space regions that remains mostly below 10%.

Individual comparisons between different PDF sets and the CT18 are illustrated with continuous lines. PDF results presented here were taken in their most recent versions from the LHAPDF library (version 6.3.0) [18], namely comparison to ABMP16 [19] is shown with the dotted green line, MSHT20 [20] with the dashed red line, and NNPDF3.1 [21] with the solid purple line. All PDFs have been derived from global fits with a fixed value for the strong coupling constant at $\alpha_S(m_Z) = 0.118$, except for ABMP16 for which it was chosen at $\alpha_S(m_Z) = 0.1147$. The best agreement to the CT18 result is displayed by the predictions obtained with MSHT20 (red dashed line). On the contrary, the worst agreement is observed by ABMP16, most likely due to the different value in α_S . Despite that, predictions seem to be compatible within the theoretical PDF uncertainty.

An alternative representation of the information presented in the bottom plot of Fig. 6.4 can be found in Fig. 6.5. In the latter figure, comparisons for each rapidity bin are presented in entirely separate canvases where increased visibility is achieved. Additionally, in these plots the originally total theory uncertainty around unity, has been split to its three individual sources, presenting separately the scale, PDF, and NP uncertainties with black dashed-dotted lines, gray dashed band, and cyan band respectively. Similar plots, that show the data to theory agreement, for the 2017 and 2018 datasets can be found in Appendix J.

The results presented in this dissertation, including the sample processing stages, theoretical calculations, final comparisons, and the QCD analysis (in Part III) have been pub-

lished in the paper “Measurement of multidifferential cross sections for dijet production in proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$ ” [7] exploiting the 2016 CMS dataset. Notably, this paper also includes five additional measurements, conducted by other colleagues within the CMS Collaboration. While each measurement focuses in slightly different observables, they all share a common goal, measuring the inclusive dijet cross section. Together these results provide a comprehensive overview of dijet production, offering valuable insights across multiple observables.

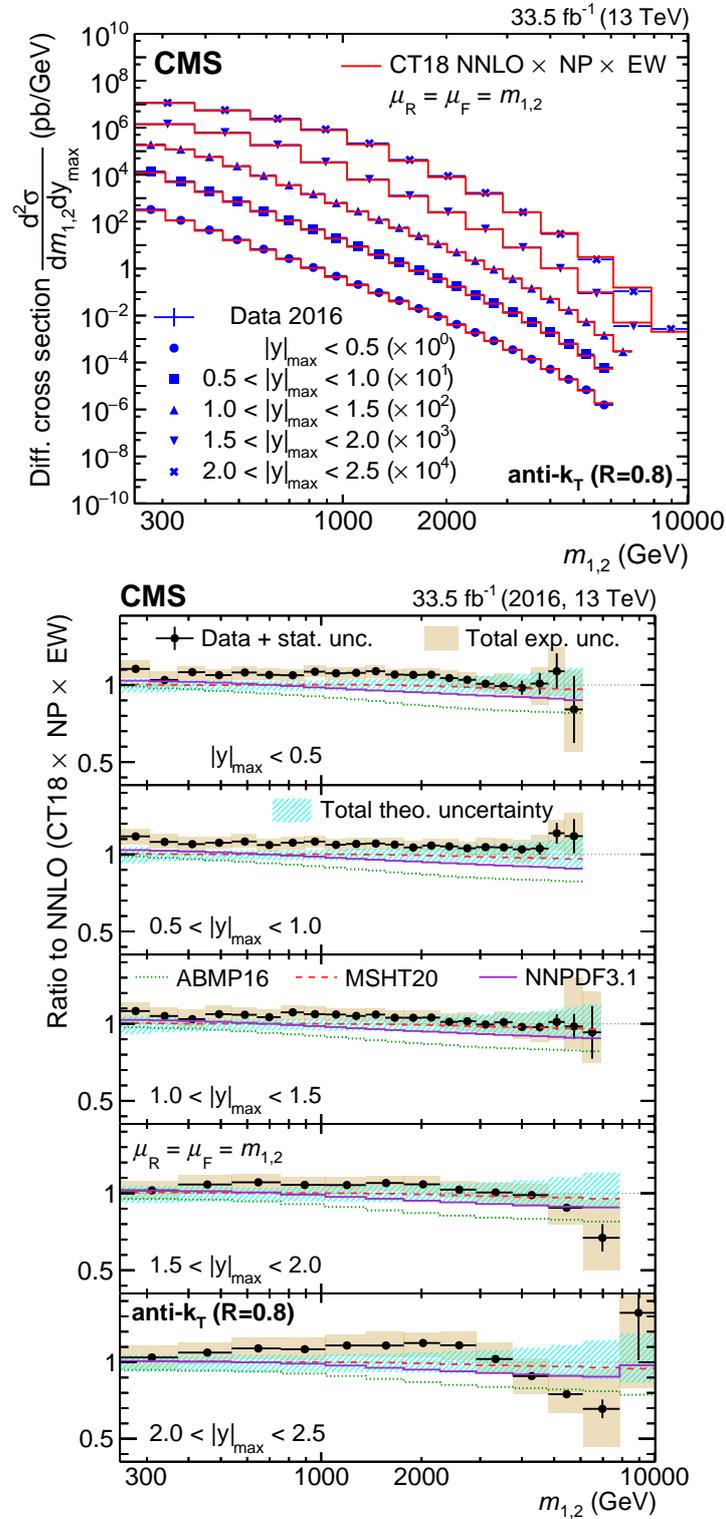


Figure 6.4: Data to theory comparison. Experimental differential cross sections from the 2016 dataset overlaid on top of the fixed order predictions (top), and their ratio with respective uncertainties (bottom).

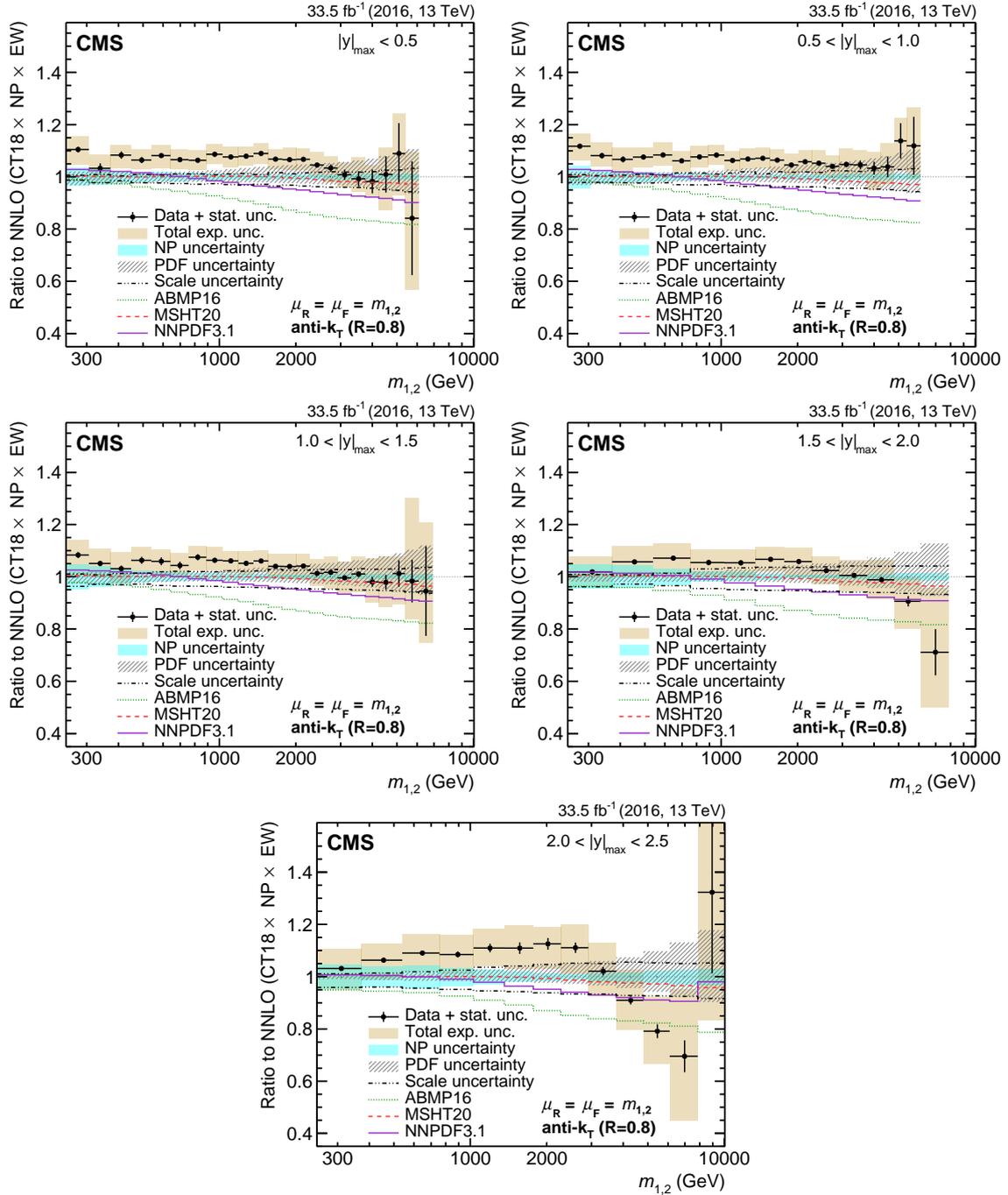


Figure 6.5: Individual data to theory comparisons. Experimental differential cross sections from the 2016 dataset compared to fixed order theory predictions with respective uncertainties. Total theory uncertainty is decomposed into contributions from different sources. Each plot corresponds to a different rapidity region.

Bibliography

- [1] T. Gehrmann et al. Jet cross sections and transverse momentum distributions with NNLOJET. *PoS*, RADCOR2017:074, 2018. <https://pos.sissa.it/290/074>.
- [2] D. Britzger et al. Calculations for deep inelastic scattering using fast interpolation grid techniques at NNLO in QCD and the extraction of α_S from HERA data. *The European Physical Journal C*, 79(10), Oct 2019. <http://dx.doi.org/10.1140/epjc/s10052-019-7351-x>.
- [3] D. Britzger et al. NNLO interpolation grids for jet production at the LHC, 2022. <https://arxiv.org/abs/2207.13735>.
- [4] T. Kluge and K. Rabbertz and M. Wobisch. Fast pQCD calculations for PDF fits. https://www.worldscientific.com/doi/abs/10.1142/9789812706706_0110.
- [5] D. Britzger and K. Rabbertz and F. Stober and M. Wobisch. New features in version 2 of the fastNLO project. In *Proc. 20th Intern. Workshop on Deep-Inelastic Scattering and Related Subjects*, page 217, 2012. <https://indico.cern.ch/event/153252/contributions/1396811/>.
- [6] J. Currie et al. Precise Predictions for Dijet Production at the LHC. *Phys. Rev. Lett.*, 119:152001, Oct 2017. <https://link.aps.org/doi/10.1103/PhysRevLett.119.152001>.
- [7] A. Hayrapetyan et al. [CMS Collaboration]. Measurement of multidifferential cross sections for dijet production in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$. *The European Physical Journal C*, 85(1), Jan 2025. <http://dx.doi.org/10.1140/epjc/s10052-024-13606-8>.
- [8] T.J. Hou et al. New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC. *Phys. Rev. D*, 103:014013, Jan 2021. <https://link.aps.org/doi/10.1103/PhysRevD.103.014013>.
- [9] A. Ridder et al. Triple Differential Dijet Cross Section at the LHC. *Phys. Rev. Lett.*, 123:102001, Sep 2019. <https://link.aps.org/doi/10.1103/PhysRevLett.123.102001>.
- [10] J. Currie, E. W. N. Glover, and J. Pires. Next-to-Next-to Leading Order QCD Predictions for Single Jet Inclusive Production at the LHC. *Phys. Rev. Lett.*, 118:072002, Feb 2017. <https://link.aps.org/doi/10.1103/PhysRevLett.118.072002>.
- [11] X. Chen, T. Gehrmann, E. W. N. Glover, A. Huss, and J. Mo. NNLO QCD corrections in full colour for jet production observables at the LHC. *Journal of High Energy Physics*, 2022(9), Sep 2022. [http://dx.doi.org/10.1007/JHEP09\(2022\)025](http://dx.doi.org/10.1007/JHEP09(2022)025).
- [12] V. Khachatryan et al. [CMS Collaboration]. Measurement of the double-differential inclusive jet cross section in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$. *The European Physical Journal C*, 76(8), Aug 2016. <http://dx.doi.org/10.1140/epjc/s10052-016-4286-3>.
- [13] G. Aad et al. [ATLAS Collaboration]. Measurement of dijet cross-sections in pp collisions at 7 TeV centre-of-mass energy using the ATLAS detector. *Journal of High Energy Physics*, 2014(5), May 2014. [http://dx.doi.org/10.1007/JHEP05\(2014\)059](http://dx.doi.org/10.1007/JHEP05(2014)059).

- [14] S. Dittmaier, A. Huss, and C. Speckner. Weak radiative corrections to dijet production at hadron colliders. *Journal of High Energy Physics*, 2012(11), Nov 2012. [http://dx.doi.org/10.1007/JHEP11\(2012\)095](http://dx.doi.org/10.1007/JHEP11(2012)095).
- [15] M. Cacciari, S. Frixione, G. Ridolfi, M. L. Mangano, and P. Nason. The $t\bar{t}$ cross-section at 1.8 and 1.96 TeV: a study of the systematics due to parton densities and scale dependence. *Journal of High Energy Physics*, 2004(04):068, May 2004. <https://dx.doi.org/10.1088/1126-6708/2004/04/068>.
- [16] S. Catani, D. de Florian, M. Grazzini, and P. Nason. Soft-gluon resummation for Higgs boson production at hadron colliders. *Journal of High Energy Physics*, 2003(07):028, Jul 2003. <https://dx.doi.org/10.1088/1126-6708/2003/07/028>.
- [17] A. Banfi, G. P. Salam, and G. Zanderighi. Phenomenology of event shapes at hadron colliders. *Journal of High Energy Physics*, 2010(6), Jun 2010. [http://dx.doi.org/10.1007/JHEP06\(2010\)038](http://dx.doi.org/10.1007/JHEP06(2010)038).
- [18] A. Buckley et al. LHAPDF6: parton density access in the LHC precision era. *The European Physical Journal C*, 75(3), Mar 2015. <http://dx.doi.org/10.1140/epjc/s10052-015-3318-8>.
- [19] S. Alekhin, J. Blümlein, S. Moch, and R. Plačakytė. Parton distribution functions, α_s , and heavy-quark masses for LHC Run II. *Phys. Rev. D*, 96:014011, Jul 2017. <https://link.aps.org/doi/10.1103/PhysRevD.96.014011>.
- [20] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs. *The European Physical Journal C*, 81(4), Apr 2021. <http://dx.doi.org/10.1140/epjc/s10052-021-09057-0>.
- [21] R. D. Ball et al. Parton distributions from high-precision collider data: NNPDF Collaboration. *The European Physical Journal C*, 77(10), Oct 2017. <http://dx.doi.org/10.1140/epjc/s10052-017-5199-5>.

Part III
QCD Analysis

Chapter 7

IMPACT STUDIES

A QCD analysis is conducted to assess the impact of the 2016 dataset on the determination of the proton's PDFs and the strong coupling constant α_S . Notably, the results of this study were published as part of a collaborative effort alongside five additional measurements related to the dijet cross section [1]. This final chapter of the dissertation, along with the subsequent results, originates from the work of a fellow colleague who contributed to the publication. Consequently, only a concise overview is produced here to offer the reader a complete perspective on the analysis. For a more in-depth discussion, the reader is encouraged to refer to the original publication.

The procedure applied here is the same as the one followed from past HERAPDF analyses [2–4]. DIS data, obtained in $e^\pm p$ collisions at the HERA collider experiments H1 and ZEUS [2, 3] are complemented by the double differential cross section measurements presented here. To ensure a good theoretical description, the phase space of the former measurements is restricted to momentum transfer values above $Q_{min}^2 = 10 \text{ GeV}^2$. The theoretical predictions used in the procedure are the ones discussed earlier, in Chapter 6, produced at NNLO in pQCD, for $\mu_R = \mu_F = m_{1,2}$, and corrected for NP and EW effects.

Theoretical predictions are fit to the “complemented” data (HERA DIS + CMS dijets) with the XFITTER program (version 2.0.1) [5, 6] while access to the theory is provided via FASTNLO. Evolution of the PDFs is performed via the DGLAP equations using the QCDNUM package (version 1701/15) [7]. Contributions from heavy quarks are treated in the Thorne-Roberts optimal variable flavor number scheme (RTOPT) [8–10], with the masses of the charm and bottom quarks set to $m_c = 1.43 \text{ GeV}$ and $m_b = 4.5 \text{ GeV}$ respectively.

According to the HERAPDF approach the proton structure is expressed in terms of the gluon distribution $g(x)$, the up and down valence quark distributions $u_v(x)$ and $d_v(x)$, and the up- and down- type sea antiquark distributions $\bar{U}(x)$ and $\bar{D}(x)$, the latter two defining the total sea quark distribution as $\Sigma(x) = 2(\bar{U}(x) + \bar{D}(x))$. These are parametrized as a function of the proton momentum fraction x at a starting scale $\mu_{F,0}^2 = 1.9 \text{ GeV}^2$ as:

$$xf(x, \mu_{F,0}^2) = A_f x^{B_f} (1-x)^{C_f} (1 + D_f x + E_f x^2) \quad (7.1)$$

where A_f are the normalization parameters, B and C are parameters to control the shape of the distribution as x approaches the edges of its domain at 0 and 1, respectively, and D and E parameters introduce additional degrees of freedom related to the functional forms.

Constraints imposed by normalization conditions and sum rules (refer to the publication for the exact conditions) result in a total of ten A , B , and C parameters, whose values are determined during the fit. A minimization scan is performed where additional D and E parameters are included whenever an improvement in the χ^2 is observed. The final

parametrization for the fits with the inclusion of CMS data is presented below:

$$\begin{aligned}
xg(x, \mu_{F,0}^2) &= A_g x^{B_g} (1-x)^{C_g} \\
xu_{\nu}(x, \mu_{F,0}^2) &= A_{u_{\nu}} x^{B_{u_{\nu}}} (1-x)^{C_{u_{\nu}}} (1 + \mathbf{D}_{u_{\nu}} x + \mathbf{E}_{u_{\nu}} x^2) \\
xd_{\nu}(x, \mu_{F,0}^2) &= A_{d_{\nu}} x^{B_{d_{\nu}}} (1-x)^{C_{d_{\nu}}} \\
x\bar{U}(x, \mu_{F,0}^2) &= A_{\bar{U}} x^{B_{\bar{U}}} (1-x)^{C_{\bar{U}}} (1 + \mathbf{D}_{\bar{U}} x) \\
x\bar{D}(x, \mu_{F,0}^2) &= A_{\bar{D}} x^{B_{\bar{D}}} (1-x)^{C_{\bar{D}}}
\end{aligned} \tag{7.2}$$

where parameters in bold ($\mathbf{D}_{u_{\nu}}$, $\mathbf{E}_{u_{\nu}}$, and $\mathbf{D}_{\bar{U}}$) are the ones added from the minimization procedure.

Four uncertainty sources are attributed to the fits, namely the *fit*, *model*, *scale* and, *parametrization* uncertainties, based on the procedure followed in Ref. [4]:

- Uncertainties represented by the experimental measurement and the theory are propagated to the output PDFs as a fit uncertainty. It is estimated using the MC method, outlined in Refs [11, 12]. The input data are varied according to their statistical and systematic uncertainties and alternative fits are performed for each case. The standard deviation of the ensemble of PDFs is taken as the final fit uncertainty.
- A model uncertainty to account for the choices in made for the values of non-PDF parameters, like the minimum Q^2 value discussed above, the strangeness fraction, or the charm and bottom quark masses. It is determined by varying their nominal values as indicated in Table 7.1, in the end, adding their differences to the nominal fit in quadrature, resulting in an asymmetric uncertainty.
- The choice in the PDF parametrization introduces yet another source of uncertainty, which is determined by performing additional fits after including one more D or E parameter compared to the nominal parametrization. The maximum deviation between the obtained variations to the nominal one is taken as the final parametrization uncertainty.
- A scale uncertainty to account for missing higher orders in pQCD, estimated by the six-point variation method and the resulting envelope as discussed in Section 6.3.1.

The total uncertainty on the PDF derivation is calculated by taking the sum in quadrature of the fit, model and scale contributions, while adding the parametrization uncertainty linearly.

To profit more from the improved data-to-theory agreement observed in the first four rapidity bins, the forward rapidity bin ($2.0 < |y|_{max} < 2.5$) is excluded from the fit procedure. This decision is motivated by the fact that, when initially included in the QCD analysis, the partial χ^2 per number degrees of freedom in that region was significantly large, indicating a poorer agreement. More information on the comparison of the full versus restricted phase space for the fits is given in Ref. [1]. The partial and total χ^2 values are listed in Table 7.2.

Figure 7.1 shows the resulted PDFs after fitting the HERA DIS data together with the CMS dijets. Two measurements complement the HERA data, the double differential (2D) measurement presented in this thesis (red line), and the triple differential (3D) measurement contained within Ref. [1] (blue dashed line). Each plot corresponds to a different PDF with the up valence quark distribution shown in the top left, the down valence quark top right, the gluon bottom left, and the sea quark bottom right. Additionally, individual uncertainty contributions are given with different band styles. PDFs are displayed as a function of the proton momentum fraction x , evolved from the starting scale $\mu_{F,0}^2 = 1.9 \text{ GeV}^2$ to the scale of the top quark mass $\mu_F^2 = m_t^2$. The upper part of each

plot illustrates the central PDF value, while the lower part corresponds to a comparison between the uncertainties and central values of the results as obtained from the inclusion of the 2D and 3D CMS dijets. Results seem to be compatible within their total uncertainties.

Table 7.1: Non-PDF parameter nominal values and their variations to estimate the model uncertainty [1].

Parameter	Nominal value	Variations	
		down	up
$Q_{min}^2 (GeV^2)$	10	7.5	12.5
f_s	0.4	0.3	0.5
$m_c (GeV)$	1.43	1.37*	1.49
$m_b (GeV)$	4.5	4.25	4.75
$\mu_{F,0}^2 (GeV^2)$	1.9	1.6	2.2*

Table 7.2: Partial χ^2 values for HERA DIS data and HERA DIS complemented with the 2016 CMS dijet measurements [1].

Data set	Partial χ^2/n_{data}	
	HERA DIS	HERA DIS + CMS dijets
CMS dijets		
$ y _{max} < 0.5$		18/22
$0.5 < y _{max} < 1.0$		15/22
$1.0 < y _{max} < 1.5$		16/23
$1.5 < y _{max} < 2.0$		15/12
HERA1+2		
$CC e^-p, E_p = 920 GeV$	51/42	51/42
$CC e^+p, E_p = 920 GeV$	37/39	37/39
$NC e^-p, E_p = 920 GeV$	221/159	222/159
$NC e^+p, E_p = 460 GeV$	198/177	197/177
$NC e^+p, E_p = 575 GeV$	186/221	186/221
$NC e^+p, E_p = 820 GeV$	55/61	55/61
$NC e^+p, E_p = 920 GeV$	359/317	364/317
Total χ^2/n_{dof}	1161/1003	1232/1081

To better evaluate whether the inclusion of CMS data, results in further constrained PDF sets, fits are performed one more time, by only using the HERA DIS data alone. The exact same setup and parametrization is imposed, as the one used in the case where dijet data. A one-to-one comparison between PDFs can now be achieved, illustrated in Fig. 7.2. This time the upper part of each plot makes a direct comparison between the distributions as derived by fitting the HERA DIS data alone (gray), together with the 2D dijets (red), and together with the 3D dijets (blue). Colored Bands represent the fit uncertainty of the

PDFs in the respective cases. The lower part of the plots shows a direct comparison to the plain HERA DIS results.

PDFs obtained with and without the CMS dijet data seem to be compatible. Moreover, a reduction in the fit uncertainty is observed. Specifically, looking at the gluon distribution (bottom left plot), the PDF is considerably constrained in the high x region, for momentum fractions $x > 0.1$.

During the procedure presented above the strong coupling constant is kept fixed during the fit procedure. To extract, in addition to the PDFs, the value of α_S , fits are repeated yet another time. This time the coupling is released, taken as a free parameter in the fit, thus ensuring a consistent treatment of correlations between $\alpha_S(m_Z)$ and the PDF parameters. The setup remains exactly the same. The extracted value with the inclusion of the double differential measurement is:

$$\begin{aligned}
 \alpha_S(m_Z) &= 0.1179 \pm 0.0015 \text{ (fit)} \\
 &\quad \pm 0.0008 \text{ (scale)} \\
 &\quad \pm 0.0008 \text{ (model)} \\
 &\quad \pm 0.0001 \text{ (param.)} \\
 &= 0.1179 \pm 0.0019 \text{ (total)}
 \end{aligned} \tag{7.3}$$

The obtained value of $\alpha_S(m_Z)$ with the inclusion of the 2D CMS dijets is in agreement with the value of 0.1166 ± 0.0017 obtained from the measurement of the double differential inclusive jet cross section in Ref. [13], and the world average value of 0.1179 ± 0.0009 [14].

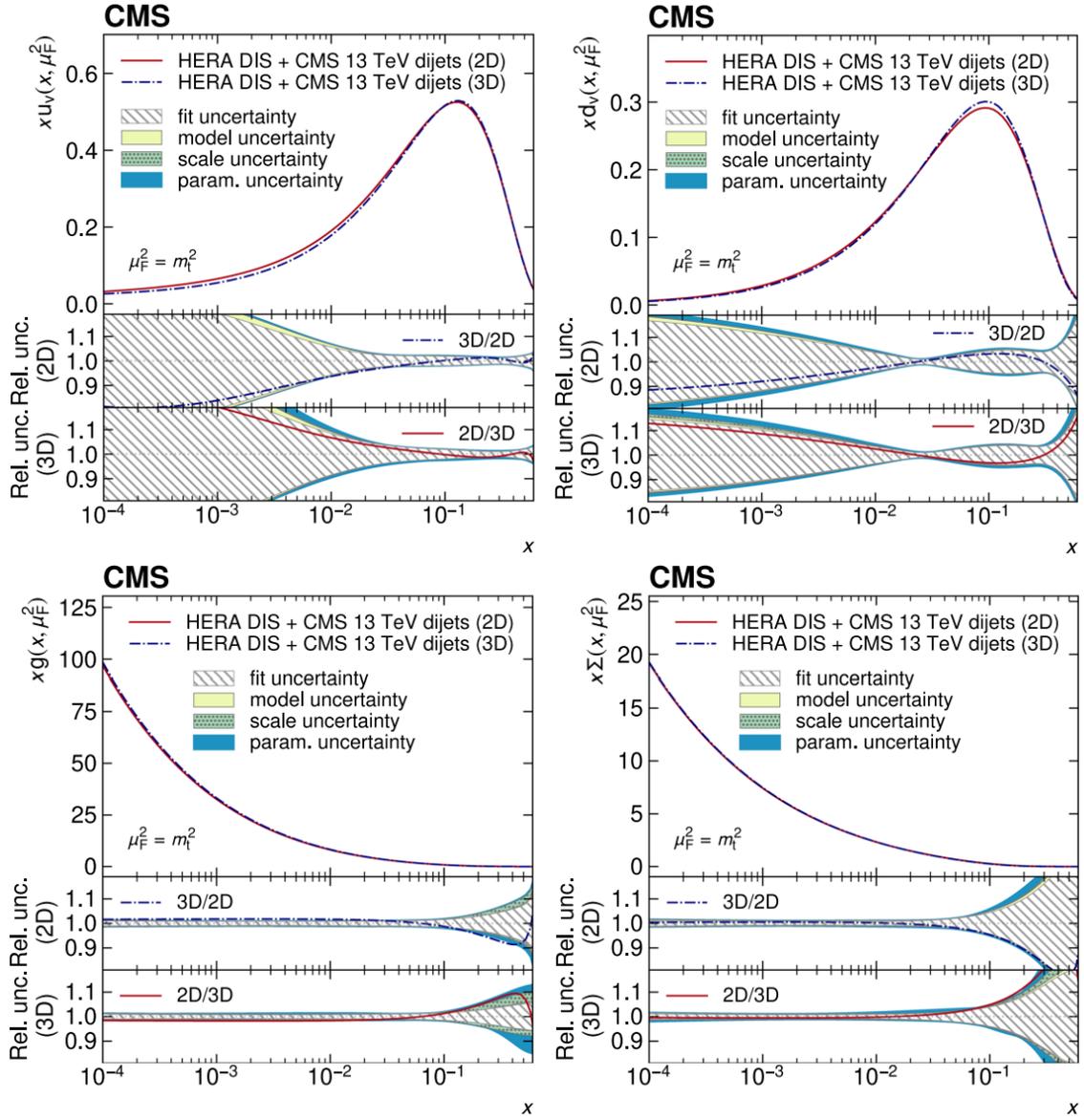


Figure 7.1: Parton distribution functions after fitting the HERA DIS data together with the 2016 CMS dijets. The up valence quark distribution (top left), down valence quark (top right), gluon (bottom left), and sea quark (bottom right). The upper part of each plot displays the central PDF value, while the lower part makes a comparison between the distributions as obtained from the inclusion of double and triple differential CMS dijets, the latter, another measurement being part of the publication in Ref. [1].

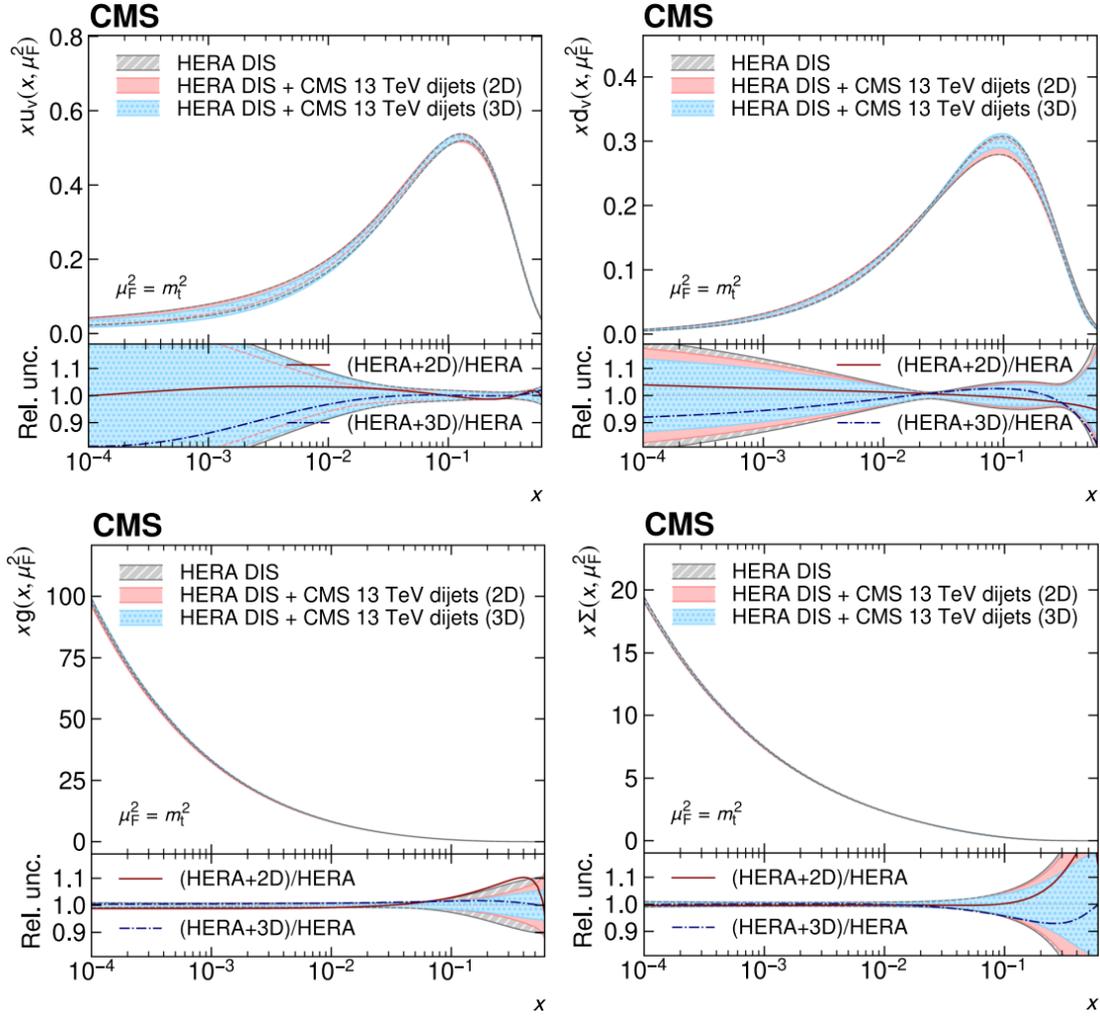


Figure 7.2: Comparison between parton distribution functions after fitting the HERA DIS data alone and together with the 2016 CMS dijets. The up valence quark distribution (top left), down valence quark (top right), gluon (bottom left), and sea quark (bottom right). The upper part of each plot displays the central PDF value around which the fit uncertainty is displayed with a band. The lower part makes a comparison between the distributions as obtained from fitting the HERA data alone to the inclusion of double and triple differential CMS dijets, the latter, another measurement being part of the publication in Ref. [1].

Summary

This dissertation concludes the measurement of the double differential inclusive dijet production cross section as a function of the invariant mass $m_{1,2}$ and the maximum absolute rapidity $|y|_{max}$ of the two leading jets in the event. The data were collected by CMS in 2016, 2017, and 2018, after proton-proton collisions at a center-of-mass energy of $13 TeV$. Each dataset corresponds to 33.5, 41.5, and $59.3 fb^{-1}$ of integrated luminosity respectively. Jet reconstruction was performed with the anti- k_T clustering algorithm for a distance parameter $R = 0.8$.

Data cross sections are compared to fixed order theoretical prediction at NNLO in pQCD. The latter, are corrected for NP effects and additional contributions coming from electroweak radiation, while they are combined with the CT18 PDF set. The agreement is found to be good across the whole $m_{1,2}$ range, in all five rapidity regions, with any divergence remaining mostly below 10%.

Specifically for the 2016 dataset a QCD analysis is performed for the determination of the strong coupling constant and the proton's PDFs. Theoretical predictions at NNLO in pQCD are fit to DIS data which are complemented by the CMS dijet cross sections. The obtained PDFs seem to be compatible with the ones when repeating the QCD analysis by fitting only the DIS data. A reduction in the fit uncertainty is observed, particularly in the gluon PDF in the high x region, for values $x > 0.1$. The QCD analysis is repeated once again leaving, this time, α_S as a free parameter in the procedure and is estimated at the scale of the Z boson mass $\alpha_S(m_Z) = 0.1179 \pm 0.0019$, a value in agreement with the 2022 world average value provided by PDG $\alpha_S(m_Z) = 0.1179 \pm 0.0009$.

Bibliography

- [1] A. Hayrapetyan et al. [CMS Collaboration]. Measurement of multidifferential cross sections for dijet production in proton–proton collisions at $\sqrt{s} = 13\text{TeV}$. *The European Physical Journal C*, 85(1), Jan 2025. <http://dx.doi.org/10.1140/epjc/s10052-024-13606-8>.
- [2] F. D. Aaron et al. [H1 and ZEUS Collaborations]. Combined measurement and QCD analysis of the inclusive $e^\pm p$ scattering cross sections at HERA. *Journal of High Energy Physics*, 2010(1), Jan 2010. [http://dx.doi.org/10.1007/JHEP01\(2010\)109](http://dx.doi.org/10.1007/JHEP01(2010)109).
- [3] H1 and ZEUS Collaborations. Combination of Measurements of Inclusive Deep Inelastic $e^\pm p$ Scattering Cross Sections and QCD Analysis of HERA Data, 2015. <https://arxiv.org/abs/1506.06042>.
- [4] H1 and ZEUS Collaborations. Impact of jet-production data on the next-to-next-to-leading-order determination of HERAPDF2.0 parton distributions, 2021. <https://arxiv.org/abs/2112.01120>.
- [5] S. Alekhin et al. [HERAFitter]. HERAFitter, Open Source QCD Fit Project, 2015. <https://arxiv.org/abs/1410.4412>.
- [6] V. Bertone et al. [xFitter]. xFitter 2.0.0: An Open Source QCD Fit Framework. *PoS, DIS2017:203*, 2017. <https://pos.sissa.it/297/203>.
- [7] M. Botje. QCDNUM: Fast QCD evolution and convolution. *Computer Physics Communications*, 182(2):490–532, 2011. <https://www.sciencedirect.com/science/article/pii/S0010465510004212>.
- [8] R. S. Thorne and R. G. Roberts. Ordered analysis of heavy flavor production in deep-inelastic scattering. *Phys. Rev. D*, 57:6871–6898, Jun 1998. <https://link.aps.org/doi/10.1103/PhysRevD.57.6871>.
- [9] R. S. Thorne. Variable-flavor number scheme for next-to-next-to-leading order. *Phys. Rev. D*, 73:054019, Mar 2006. <https://link.aps.org/doi/10.1103/PhysRevD.73.054019>.
- [10] R. S. Thorne. Effect of changes of variable flavor number scheme on parton distribution functions and predicted cross sections. *Phys. Rev. D*, 86:074017, Oct 2012. <https://link.aps.org/doi/10.1103/PhysRevD.86.074017>.
- [11] W. T. Giele and S. A. Keller. Implications of hadron collider observables on parton distribution function uncertainties. *phys. rev. d*, 58:094023, Oct 1998. <https://link.aps.org/doi/10.1103/PhysRevD.58.094023>.
- [12] W. T. Giele, S. A. Keller, and D. A. Kosower. Parton Distribution Function Uncertainties, 2001. <https://arxiv.org/abs/hep-ph/0104052>.
- [13] A. Tumasyan et al. [CMS Collaboration]. Addendum to: Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s} = 13\text{TeV}$, 2022. [https://link.springer.com/article/10.1007/JHEP12\(2022\)035](https://link.springer.com/article/10.1007/JHEP12(2022)035).
- [14] R. L. Workman et al. [Particle Data Group]. Review of Particle Physics. *Progress of Theoretical and Experimental Physics*, 2022(8):083C01, 08 2022. <https://doi.org/10.1093/ptep/ptac097>.

Appendices

Appendix

A

ELEMENTARY PARTICLES EXTENDED

An extended version of the elementary particles and force carriers within the Standard Model is given below in Fig. A.1. Note that this figure does not contribute any new information on the constituents of the SM compared to Fig. 3.1 (found in Chapter 3) but rather illustrates the same exact points in a more clear and elaborate manner. Four main categories can be identified each containing a different family of particles:

- **Quarks** are illustrated in blue color, where quark pairs make up the three generations of matter. Quarks are found in three colors, illustrated with the three colored spheres in each quark cell. Lastly, the distinction between matter and antimatter is depicted by presenting, in addition, all antiquarks. In total there are 36 different quarks.
- **Leptons**, in green, are in turn divided into three generations of matter. This time each generation is comprised from a particular lepton and each associated neutrino partner. Again, for each lepton exists its equal counterpart the antilepton, in the end amounting to 12 different leptons.
- **Gauge bosons** or vector bosons, in red, are the force carriers. Eight gluons are found, indicated with eight colored spheres inside the gluon cell. Three carriers are found for the weak interaction, each having each own cell, W^+ , W^- , Z^0 . Finally, the photon mediator is left for the electromagnetic force. These sum up to 12.
- **Scalar boson**, in yellow, the pinnacle of the SM, the Higgs boson.

In its entirety, the 61 constituents of the SM.

Appendix B

MONTE CARLO CROSS SECTIONS

Table B.1: Cross section values for Pythia8 \hat{p}_T sliced (top) and flat (bottom) samples for each year.

\hat{p}_T slices	Cross section (pb)		
	2016	2017	2018
30 \rightarrow 50	138,800,000	107,200,000	106,700,000
50 \rightarrow 80	19,110,000	15,770,000	15,710,000
80 \rightarrow 120	2,735,000	2,341,000	2,338,000
120 \rightarrow 170	466,200	406,800	406,300
170 \rightarrow 300	117,200	103,700	103,400
300 \rightarrow 470	7,763	6,833	6,852
470 \rightarrow 600	641.0	552.9	551.8
600 \rightarrow 800	185.7	156.0	156.7
800 \rightarrow 1000	32.02	26.21	26.15
1000 \rightarrow 1400	9.375	7.476	7.483
1400 \rightarrow 1800	0.8384	0.6466	0.6493
1800 \rightarrow 2400	0.1133	0.08763	0.08765
2400 \rightarrow 3200	0.006746	0.005238	0.005218
3200 \rightarrow ∞	0.0001623	0.0001354	0.0001345
15 \rightarrow 7000 (Flat sample)	1,975,000,000	—	—

Table B.2: Cross section values for the 2016 Madgraph H_T sliced sample.

H_T slices	Cross section (pb) 2016
50 \rightarrow 100	246,400,000
100 \rightarrow 200	27,940,000
200 \rightarrow 300	1,712,000
300 \rightarrow 500	347,700
500 \rightarrow 700	32,150
700 \rightarrow 1000	6,828
1000 \rightarrow 1500	1,200
1500 \rightarrow 2000	120.0
2000 \rightarrow ∞	25.34

Appendix C

PREFIRE MAPS

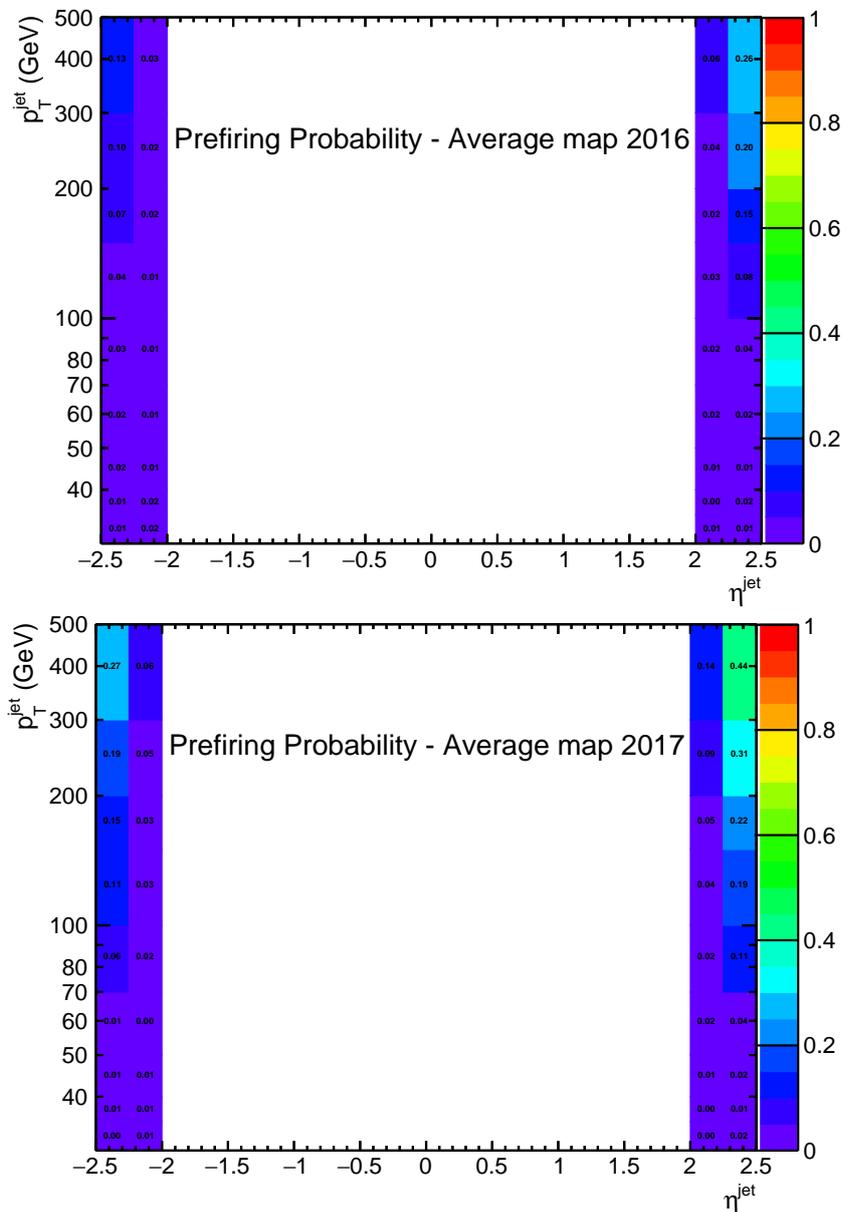


Figure C.1: Prefiring probability maps for 2016 (top) and 2017 (bottom). Maps are expressed as a function of the jet p_T and η .

Appendix D

JER STUDIES

The study undertaken here is twofold. Firstly, it critically examines the validity of the hybrid method, which, as will be discussed later, exhibits several shortcomings. Secondly, this effort contributes to the development of software tools [1] designed to generalize and enhance the methodology for deriving jet resolution tables, thereby refining resolution studies and offering deeper insights into their systematic approach.

The study begins with the construction of response distributions, exploiting the 2016 Pythia8 sliced sample (see Tab 5.3). The response Δ is defined as

$$\Delta = \frac{p_T^{rec} - p_T^{gen}}{p_T^{gen}} \quad (\text{D.1})$$

where the jet transverse momentum at reconstructed level is now denoted as p_T^{rec} , and at generator level as p_T^{gen} . The phase space is divided according to three rec level quantities, p_T^{rec} , the jet absolute pseudo-rapidity $|\eta^{rec}|$, and the soft activity ρ (a quantity related to pileup). The exact bin edges for each quantity are outlined in Table D.1.

Table D.1: Binning schemes used in the resolution studies, in terms of p_T^{rec} , $|\eta^{rec}|$ and ρ .

$p_T^{rec} \text{ (GeV)}$	$ \eta^{rec} $	ρ
15, 18, 21, 24, 28,	0.000, 0.261,	0,
32, 37, 43, 49, 56,	0.522, 0.783,	6.69,
64, 74, 84, 97, 114,	1.044, 1.305,	12.39,
133, 153, 174, 196, 220,	1.566, 1.740,	18.09,
245, 272, 300, 330, 362,	1.930, 2.043,	23.79,
395, 430, 468, 507, 548,	2.172, 2.322,	29.49,
592, 638, 686, 737, 790,	2.500, 2.650,	35.19,
846, 905, 967, 1032, 1101,	2.853, 2.964,	40.9,
1172, 1248, 1327, 1410, 1497,	3.139, 3.489,	∞
1588, 1684, 1784, 1890, 2000,	3.839, 5.191	—
2116, 2238, 2366, 2500, 2640,	—	—
2787, 2941, 3103, 3273, 3450,	—	—
3637, 3832, 4037	—	—

In detail, each ρ bin is divided into 19 $|\eta^{rec}|$ bins, each of which, in turn, is then subdivided into 62 p_T^{rec} bins. Figure D.1 is a representative plot that shows the collection of all response distributions that correspond to the $18.09 < \rho < 23.79$ & $|\eta^{rec}| < 0.261$ bin. This grid includes all p_T^{rec} bins for the aforementioned $[\rho, \eta^{rec}]$ values, with each cell containing the normalized response centered around zero. In total, there are 152 (8ρ bins \times 19 $|\eta^{rec}|$ bins = 152) such grids.

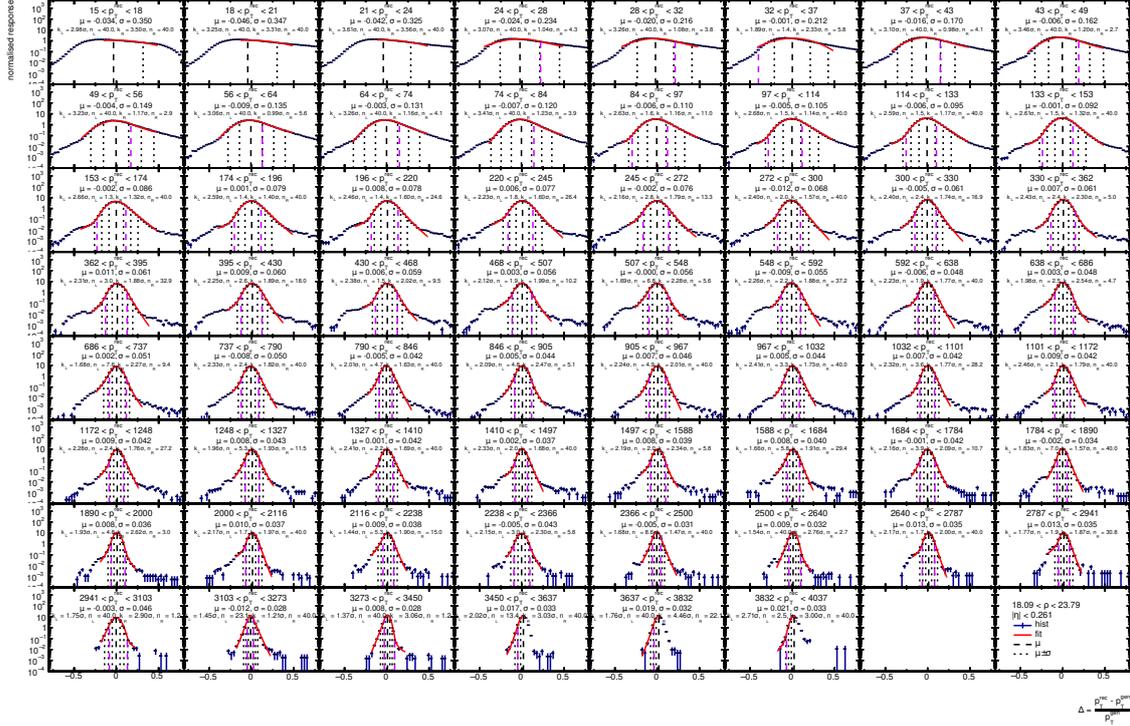


Figure D.1: Response distributions for 2016 Pythia8 slices. Responses shown here correspond to all p_T^{rec} slices related to the $[18.09 < \rho < 23.79, |\eta^{rec}| < 0.261]$ bin.

The resulted response distributions are then fitted with a Double sided Crystal Ball (DCB) function. This function is a very good candidate for the fit procedure since it combines a Gaussian function, to parametrize the core of the response distributions, and two Exponent terms, to parametrize separately each of the two tails. The fit is illustrated with the red line on top the response distributions. Inside each cell, additional information is printed on the fit performance. Notably, the mean (μ) and half-width (σ) of the distribution are displayed. Moreover, vertical black dotted lines indicate the 1, 2, and 3 σ variations around the mean, the single vertical black dashed line shows the position of the mean, and the vertical purple dashed lines display the left and right “transition points”, the points at which the tails of the distribution start. The goal of the fit is to extract the width (half-width) of the response distribution which represents the desired resolution (σ).

Next, the resolution is plotted for each $[\rho, \eta^{rec}]$ bin, as a function of p_T^{rec} . This is displayed in Fig. D.2. This time a grid of 19 cells is displayed, that contains the resolution curves for all $|\eta^{rec}|$ bins inside a given ρ bin. Looking at the very first cell (top left), this resolution curve is obtained from the sole response grid presented in Fig. D.1. Each of the blue points there corresponds to the width of the response distributions as extracted by the fit. In principle, there are 9 resolution grids in total, one for each separate ρ bin. The fit of the resolution itself is displayed by the red curve. Fit parameters and the chi square divided by the number degrees of freedom are displayed on top of each cell in red.

The resolution fit is performed via the “NSC” (Noise, Stochastic, Constant) function

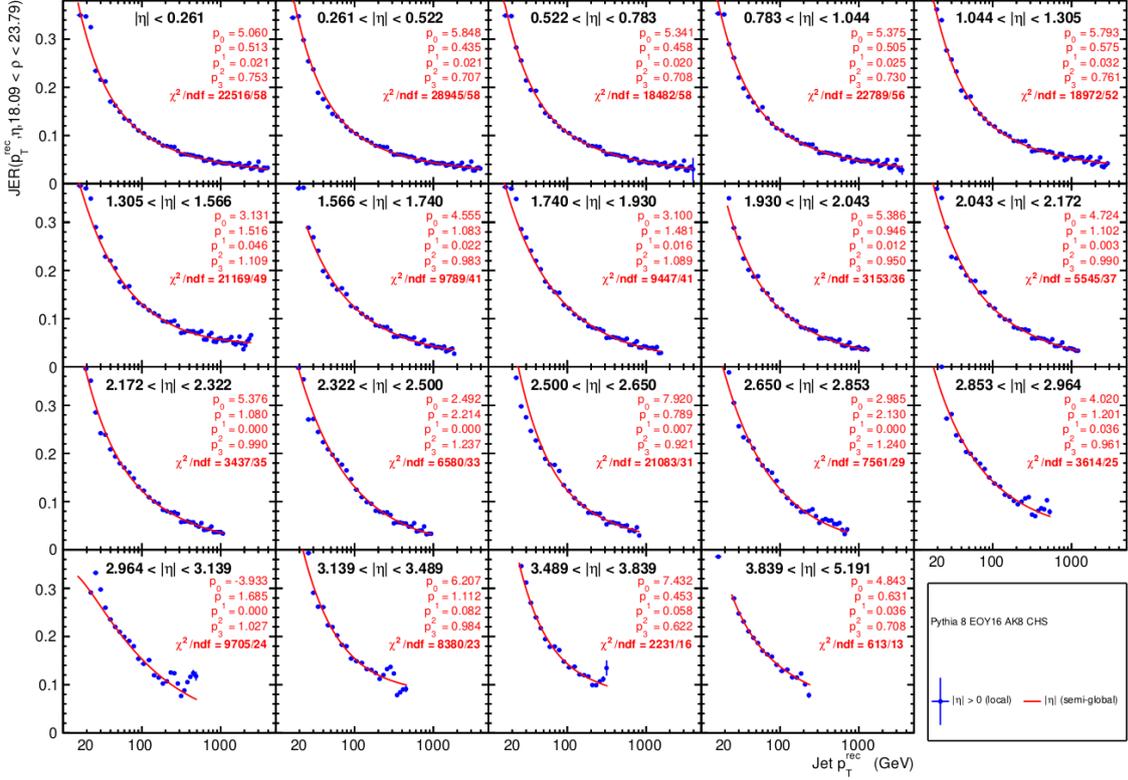


Figure D.2: Resolution curves for 2016 Pythia8 slices. All curves shown here correspond to the $18.09 < \rho < 23.79$ bin.

according to the standard procedure recommended by the JetMET group.

$$\frac{\sigma}{p_T} = \sqrt{N \frac{|N|}{p_T^2} + \frac{S^2}{p_T^n} + C^2} \equiv \sqrt{p_0 \frac{|p_0|}{p_T^2} + \frac{p_1^2}{p_T^{p_3}} + p_2^2} \quad (\text{D.2})$$

where N , S , C , and n , or respectively, p_i for $i = 0, 1, 2, 3$, the fit parameters. An overview of all resolution fits across all $[\rho, \eta^{rec}]$ bins is shown in Fig. D.3. There, each $|\eta^{rec}|$ cell contains the resolution curves for all 8 ρ bins, illustrated with the blue-gradient lines. The shape of the resolution curves is observed to follow the expected behavior. For lower p_T values the resolution is large, at the order of 30 – 40%, while in the higher end of the spectrum the resolution from all ρ bins in a given $|\eta^{rec}|$ bin converge to the same small value, mostly well below 5%. A smaller value for the resolution is interpreted as “better” resolution, i.e. p_T measurements can be performed with finer precision with increasing energy. Any deviations from the expected shape in Fig. D.3 stem from subtle fit misadjustments. These discrepancies are generally observed in the very forward rapidity regions, which, within the scope of this analysis, lie outside the phase space of interest, as we focus only on the $|\eta^{rec}| < 2.5$ region.

From the curves (Fig. D.3) obtained after the resolution fit, homemade resolution tables are constructed. The Pythia sample is smeared according to the hybrid method using the homemade resolution tables and the scale factors provided by JetMET. In reality, different smearing options are applied and the dijet mass spectrum at reconstructed level is evaluated for each case. Collectively, the results of the study are displayed in Fig. D.4. Five cells exist, each representing one of the five rapidity regions of the available phase space. The x -axis runs in the complete dijet mass range ranging from 249 GeV up to 10050 GeV . Inside each cell, different ratios are illustrated where a comparison between different smearing options is made to the *stochastic only* option. The software build-in

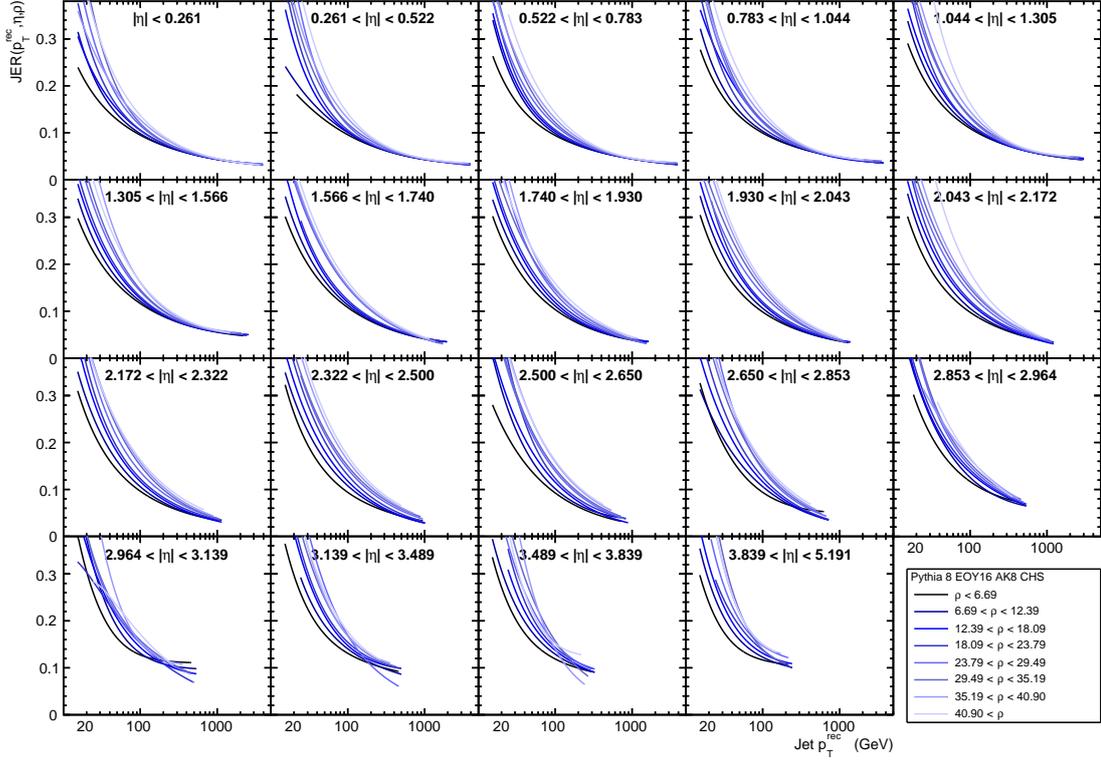


Figure D.3: Overview of all resolution curves for 2016 Pythia8 slices.

smearing options are explained:

- *Stochastic Only*. Smearing according to the stochastic approach (see Section 5.3.3) is applied uniformly to all jets, without distinctions.
- *Scaling All Matched*. Means that the scaling approach is followed for all jets that are matched, while unmatched jets are smeared with the stochastic method.
- *Scaling Core Only* (*purple, red, green*). Scaling is applied only in the core of the resolution, otherwise stochastic method is used. For this option three variations exist that define the core of the resolution at 1, 2, and 3σ . The 3σ case is then one-to-one equivalent to the *hybrid* approach recommended by JetMET.

At this point, some topics of discussion can be raised. Firstly, for the derivation of resolution tables the JetMET group performs only a Gaussian fit of the response, not taking into consideration the tails of the distribution, thus significantly underestimating the extracted resolution. Secondly, it is assumed that the Gaussian core is contained within the $[-3\sigma, 3\sigma]$ interval, something that is not exactly reflected in the responses presented in Fig. D.1 (compare the locations of the vertical purple dashed lines to that of the third - off-centered - vertical dotted lines). Two novelties are found here, on one hand, the response fits are performed with the DCB function, taking into consideration the contribution from the tails and achieving a more accurate estimation of the resolution. On the other hand, a conservative approach is taken regarding the assumption of the Gaussian core's position, which is considered to lie predominantly within the $[-1\sigma, 1\sigma]$ interval when applying the hybrid method. The impact of this choice is compared to that of applying only the stochastic method, which imposes no strong assumptions on each definition. This comparison is illustrated by the purple colored ratio in Fig. D.4. While all options appear to agree within uncertainties with the results obtained from the plain stochastic approach, the hybrid method with a fixed core at 1σ demonstrates slightly better

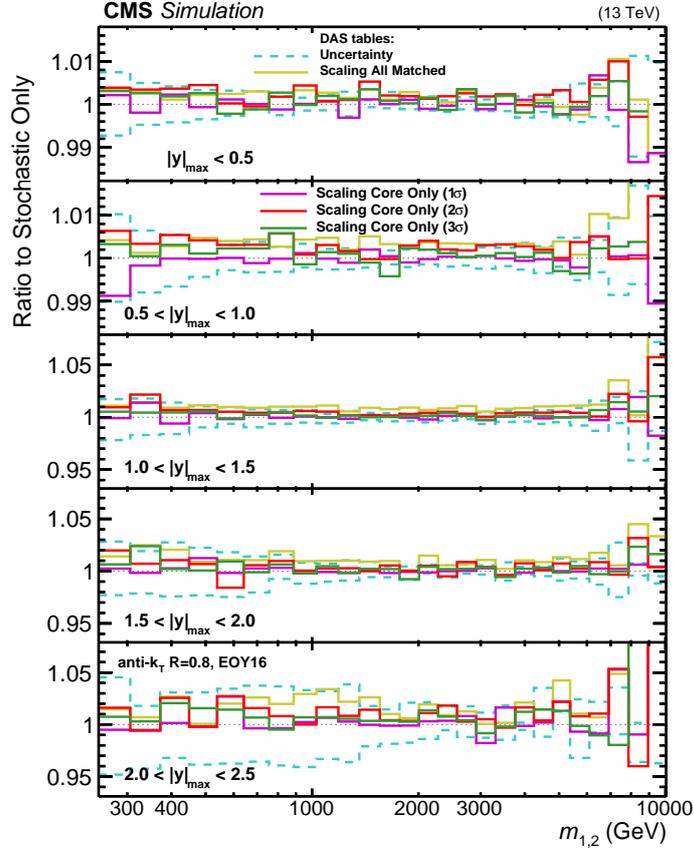


Figure D.4: Impact studies on the dijet mass observable after different smearing options.

agreement. Consequently, it is adopted as the default smearing option for the 2016 Pythia slices sample.

No such dedicated studies were performed for 2017 and 2018 simulation samples, so the provided resolution tables are taken, and smearing with the stochastic approach is applied to all jets in the phase space.

For completeness, additional plots (Fig D.5 and D.6) are provided to illustrate the software's evolution over time. While the results presented in these plots are not directly related to this work, they are nonetheless relevant to other projects within the JERC subgroup. These plots were obtained from a presentation given directly to the group [2].

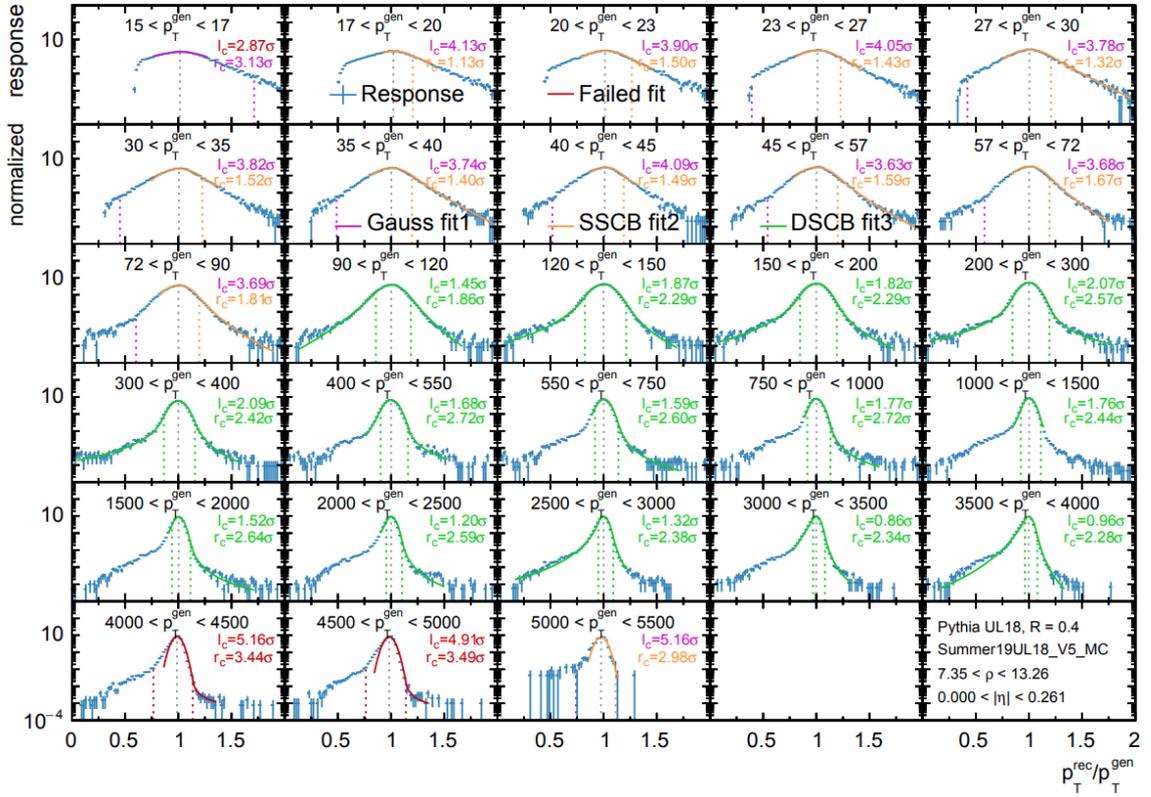


Figure D.5: Evolution of response fits. Response is centered around unity, and each grid now contains several p_T^{gen} bins. Fit is performed in three stages, Gauss, Gauss + one tail, Gauss + two tails, and fit success/failure is displayed depending on the level that it evolved. Fit ranges are chosen in a dynamic approach, extending the effective number of possible fits.

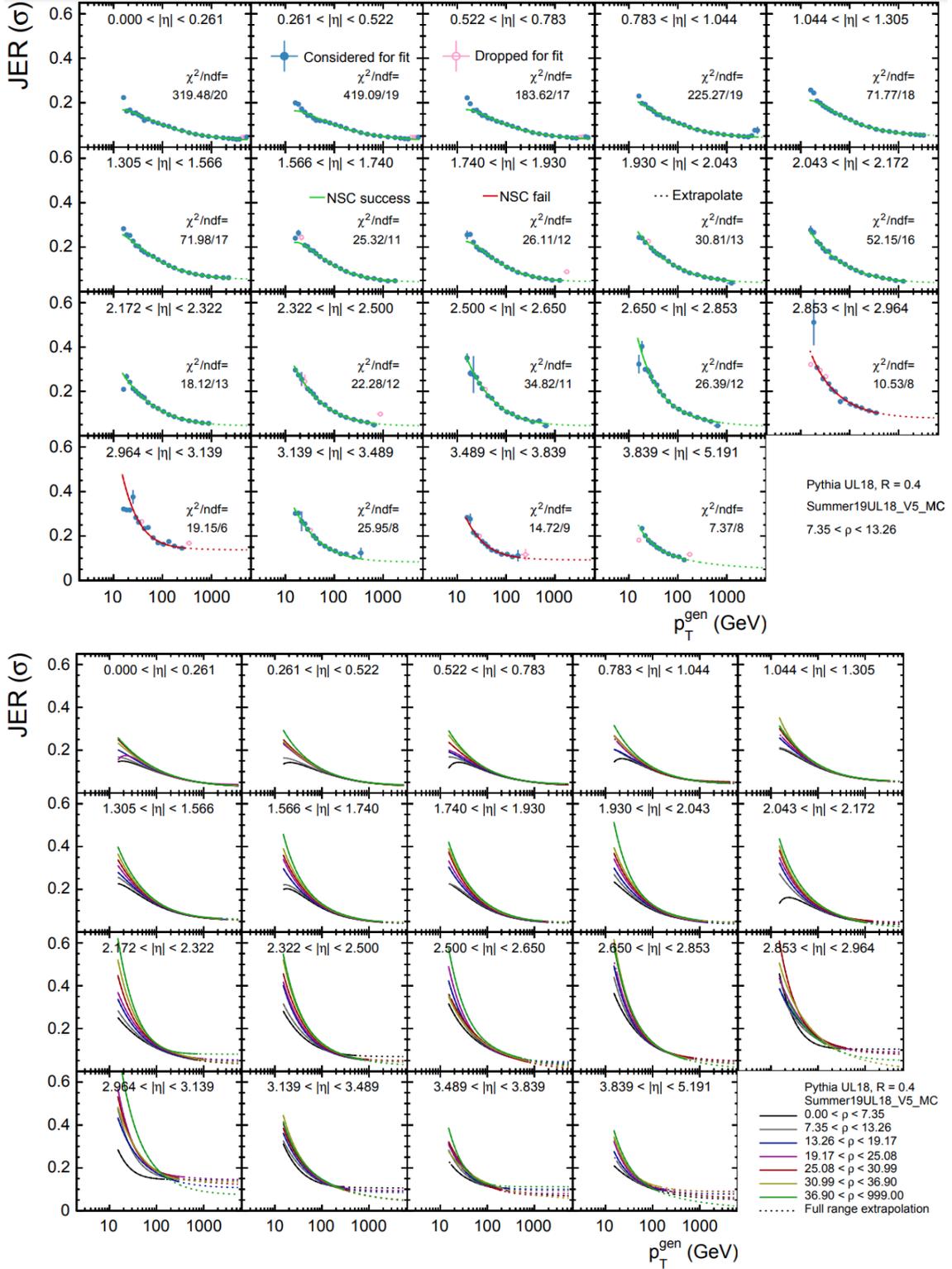


Figure D.6: Evolution of resolution fits. Example of a detailed grid with resolution fits (top). Fit success (in green) or failure (in red) is displayed. Empty markers illustrate points dropped from the fit procedure, while a dashed line indicates the extrapolation of the fit function. An overview of all resolution fits (bottom), with multiple colors expressing fits in different cases.

Bibliography

- [1] DAS Analysis System. Available: <https://gitlab.cern.ch/cms-analysis/genera1/DasAnalysisSystem>. Accessed: 2025-02-19.
- [2] A. Ziaka P.L.S. Connor, P.G. Kosmoglou Kioseoglou. Jet Energy Resolution studies. Available: <https://indico.cern.ch/event/1399194/#3-jer-studies>. Accessed: 2025-02-26.

Appendix E

TRIGGER EFFICIENCY CURVES

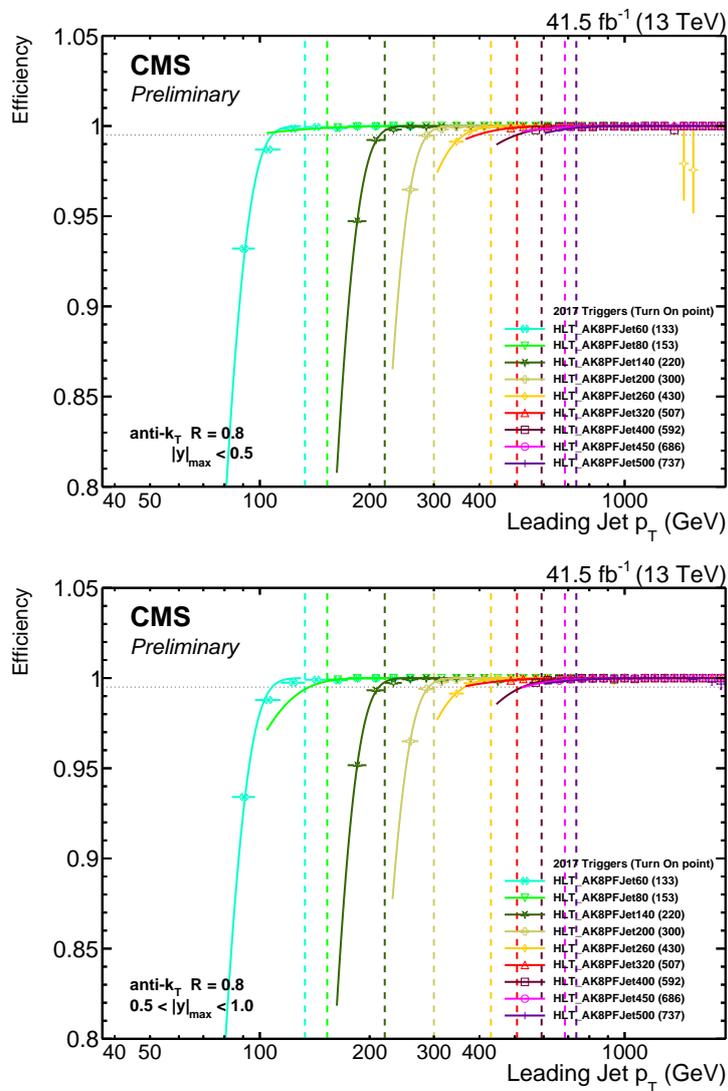


Figure E.1: Trigger efficiency curves as a function of the leading jet p_T in central rapidity regions, $|y|_{\max} < 0.5$ (top), and $0.5 < |y|_{\max} < 1.0$ (bottom), for 2017.

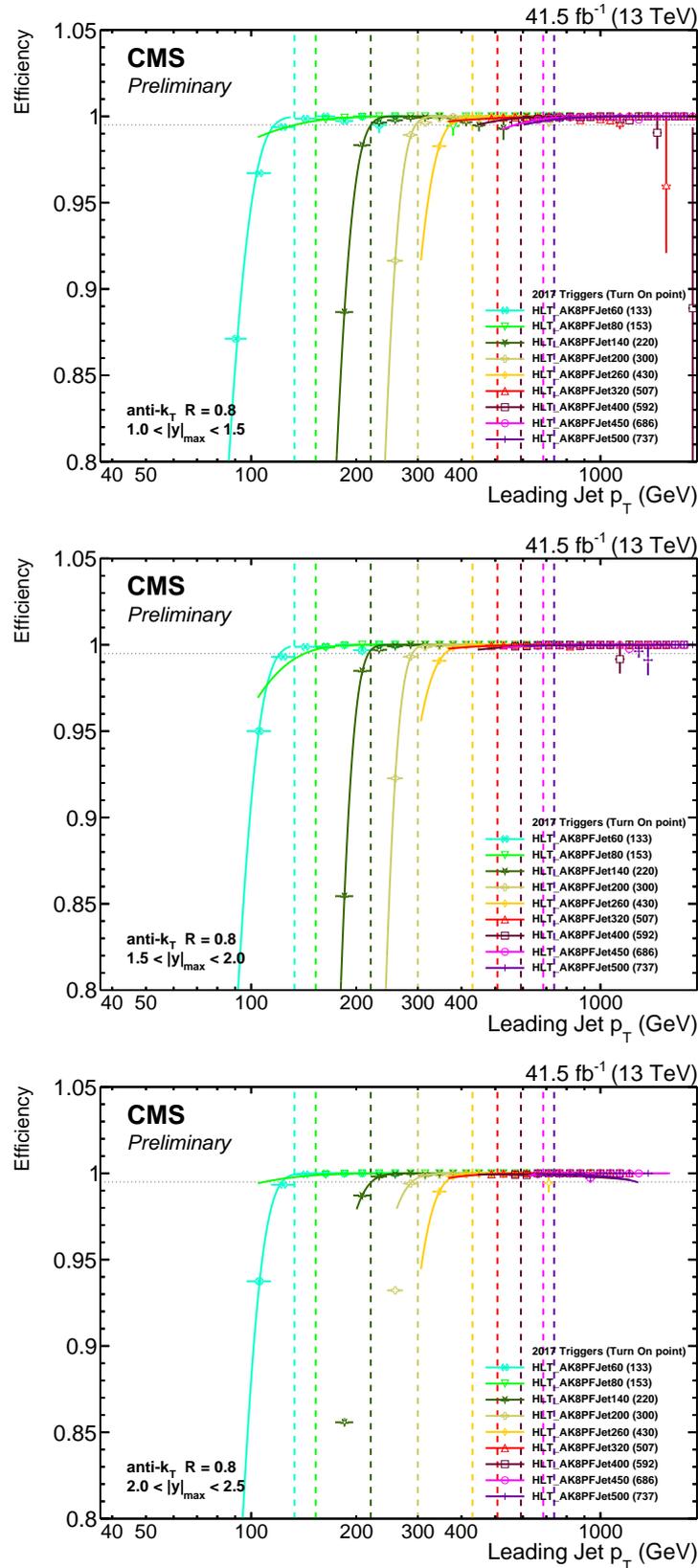


Figure E.2: Trigger efficiency curves as a function of the leading jet p_T in the central rapidity region $1.0 < |y|_{\max} < 1.5$ (top), and in the forward ones $1.5 < |y|_{\max} < 2.0$ (middle), and $2.0 < |y|_{\max} < 2.5$ (bottom), for 2017.

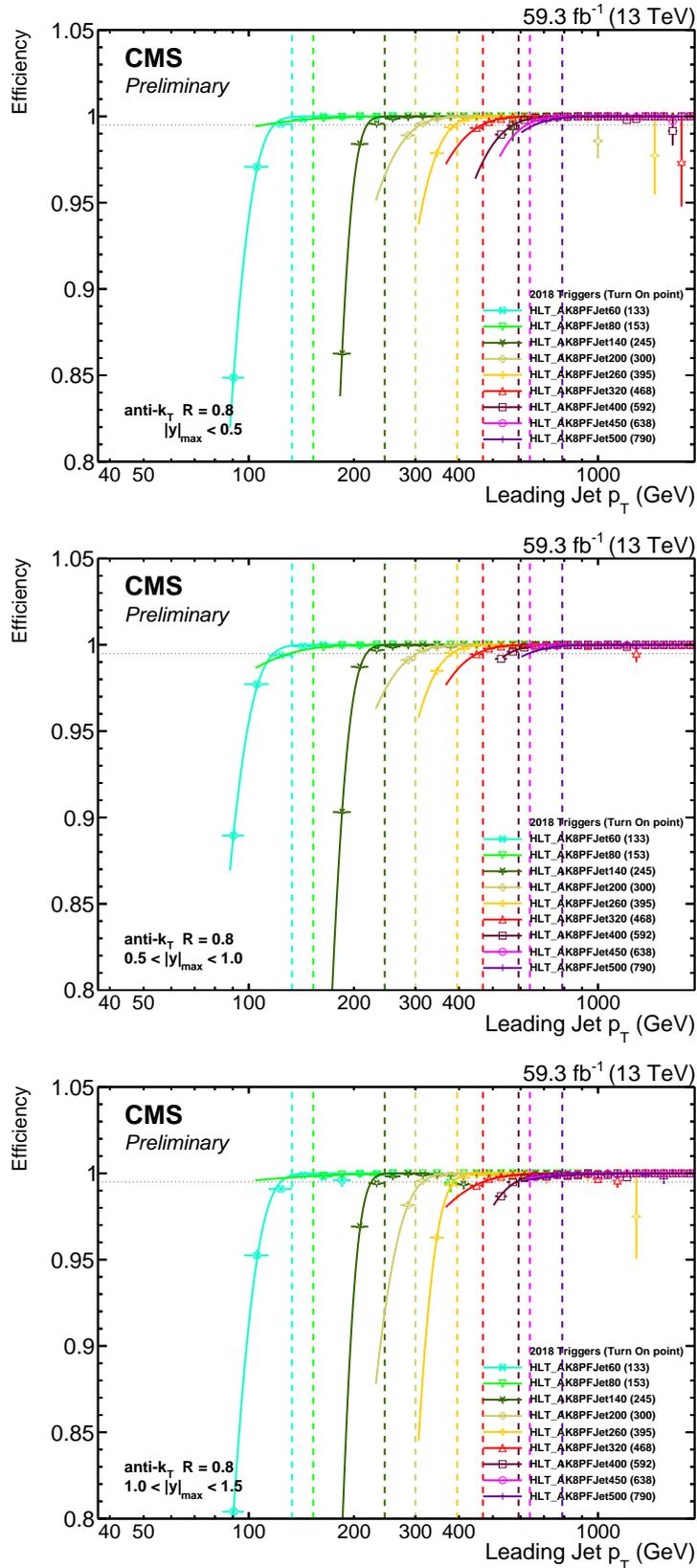


Figure E.3: Trigger efficiency curves as a function of the leading jet p_T in central rapidity regions, $|y|_{\max} < 0.5$ (top), $0.5 < |y|_{\max} < 1.0$ (middle), and $1.0 < |y|_{\max} < 1.5$ (bottom), for 2018.

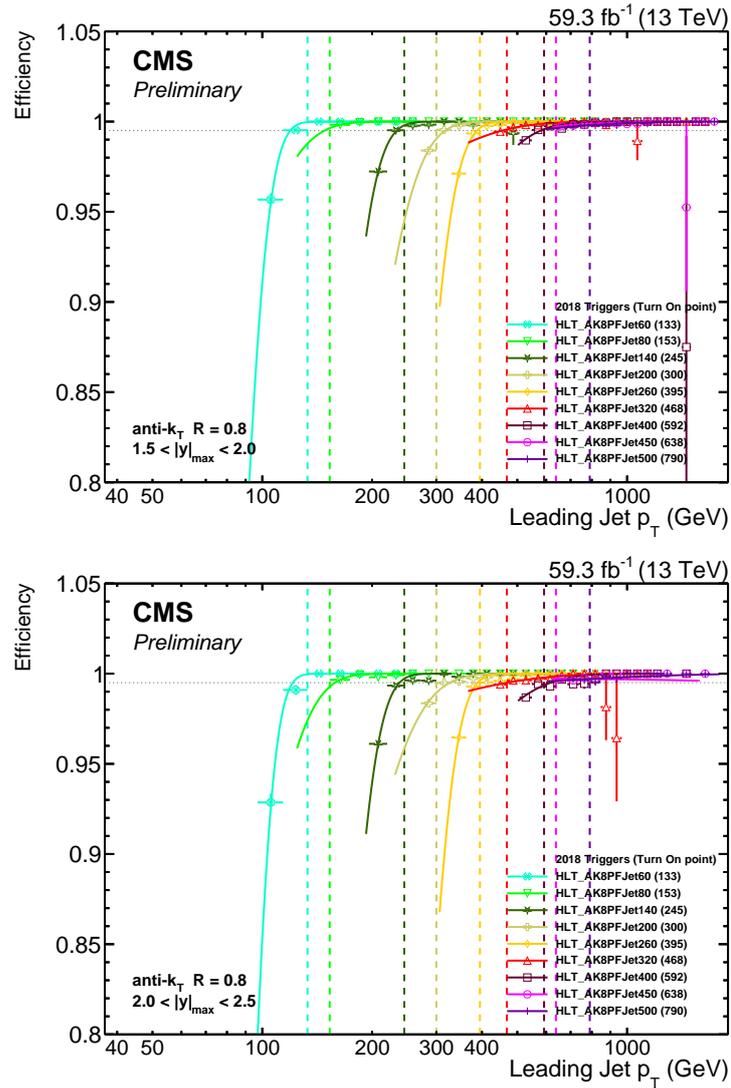


Figure E.4: Trigger efficiency curves as a function of the leading jet p_T in forward rapidity regions, $1.5 < |y|_{max} < 2.0$ (top), and $2.0 < |y|_{max} < 2.5$ (bottom), for 2018.

Appendix F

DETECTOR LEVEL DISTRIBUTIONS

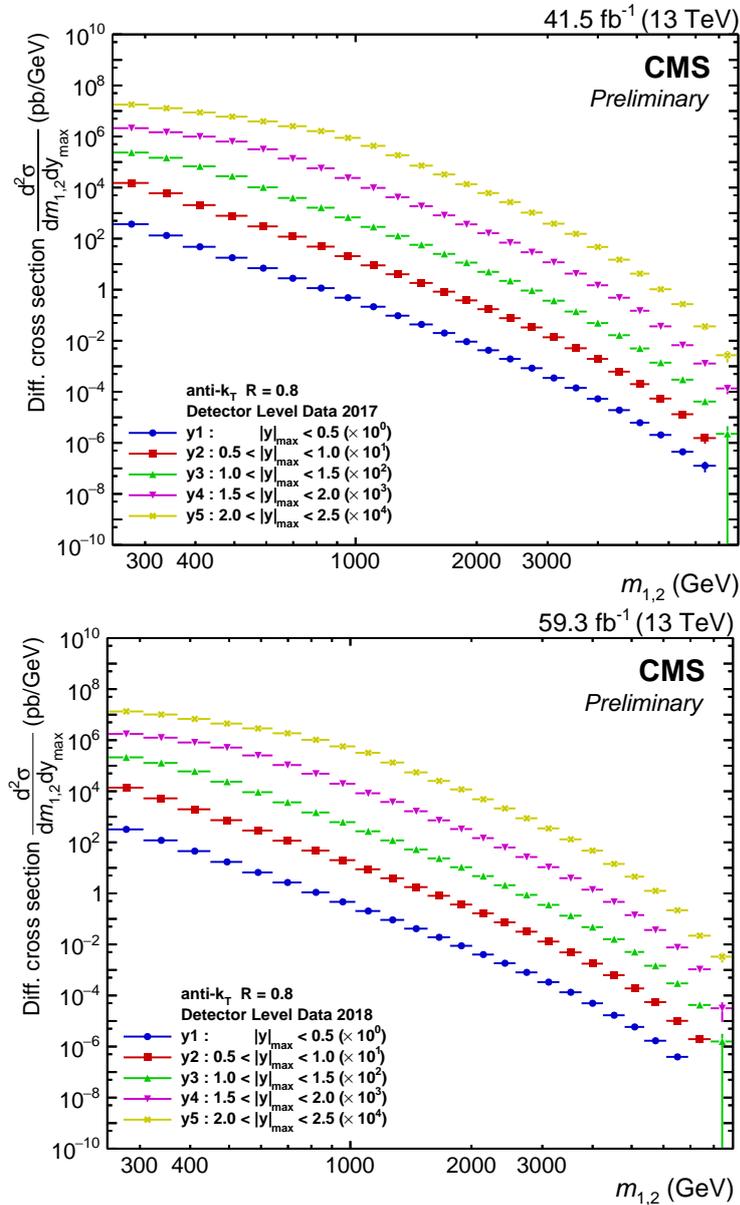


Figure F.1: Detector level double differential cross section as a function of $m_{1,2}$ and $|y|_{\max}$ for the 2017 (top) and 2018 (bottom) datasets.

Appendix G

PROBABILITY MATRICES

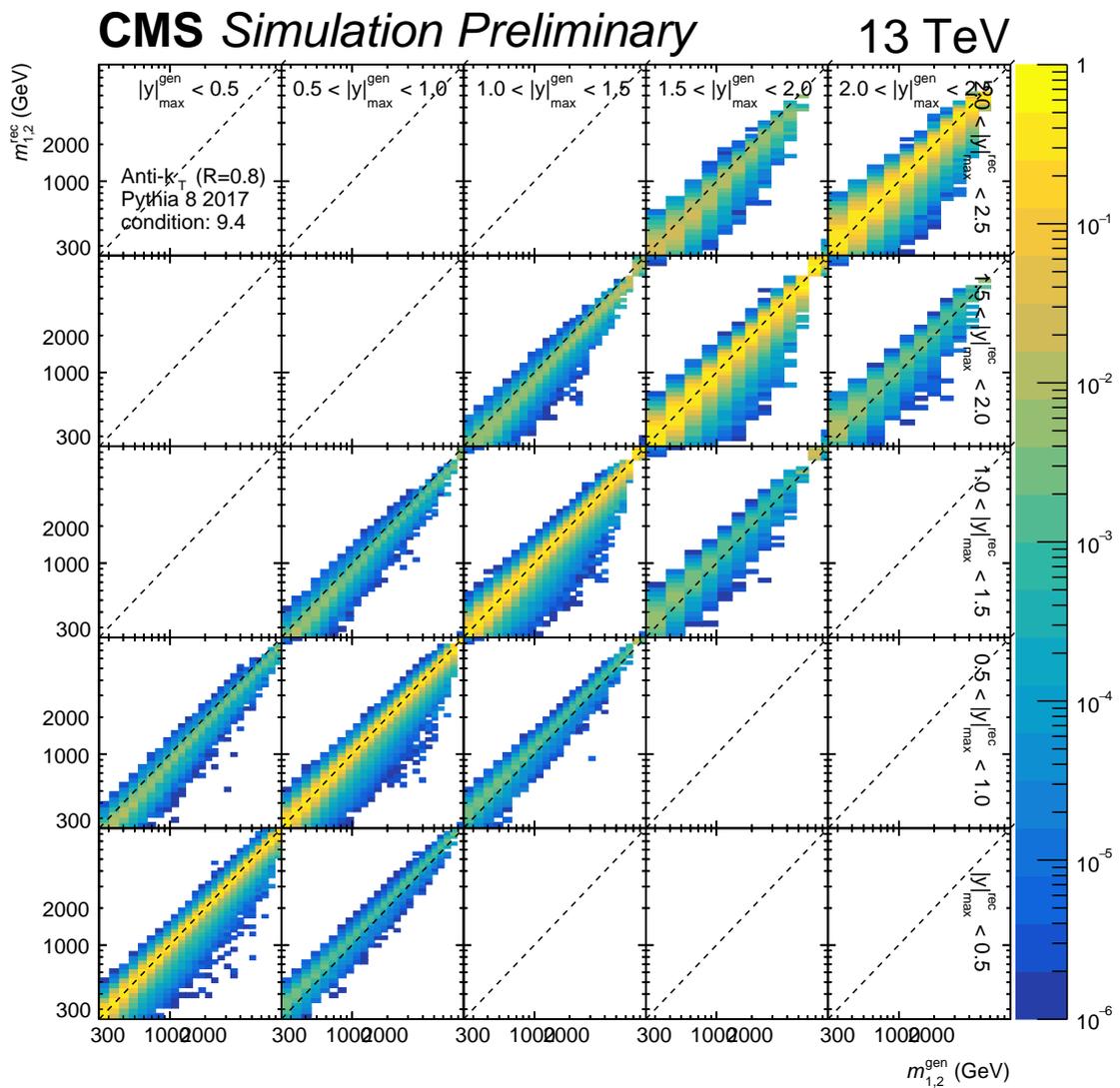


Figure G.1: Two-dimensional probability matrix for 2017 Pythia slices. Event migration probability between rapidity and mass bins from generator to reconstructed level.

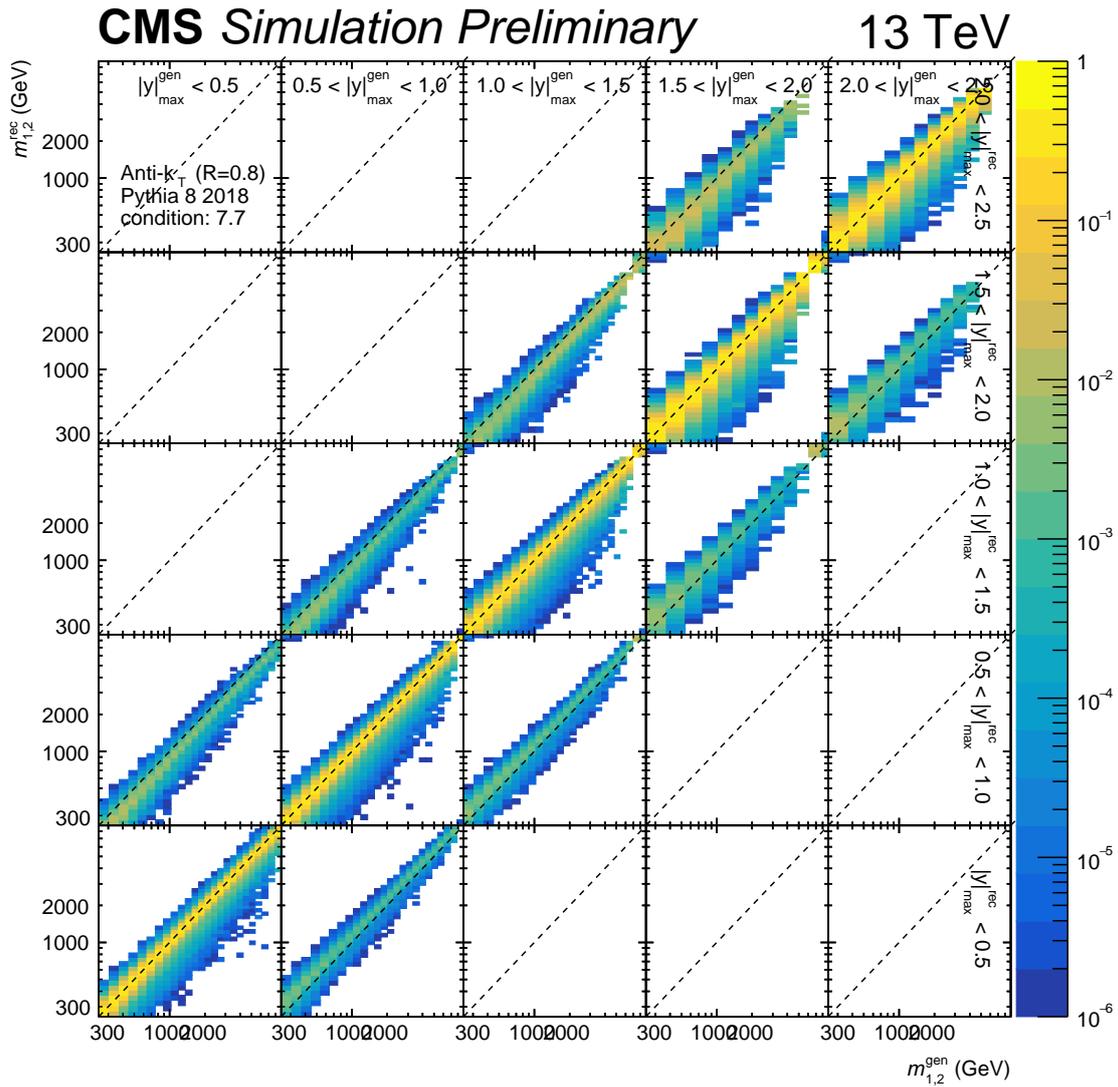


Figure G.2: Two-dimensional probability matrix for 2018 Pythia slices. Event migration probability between rapidity and mass bins from generator to reconstructed level.

Appendix H

UNCERTAINTIES

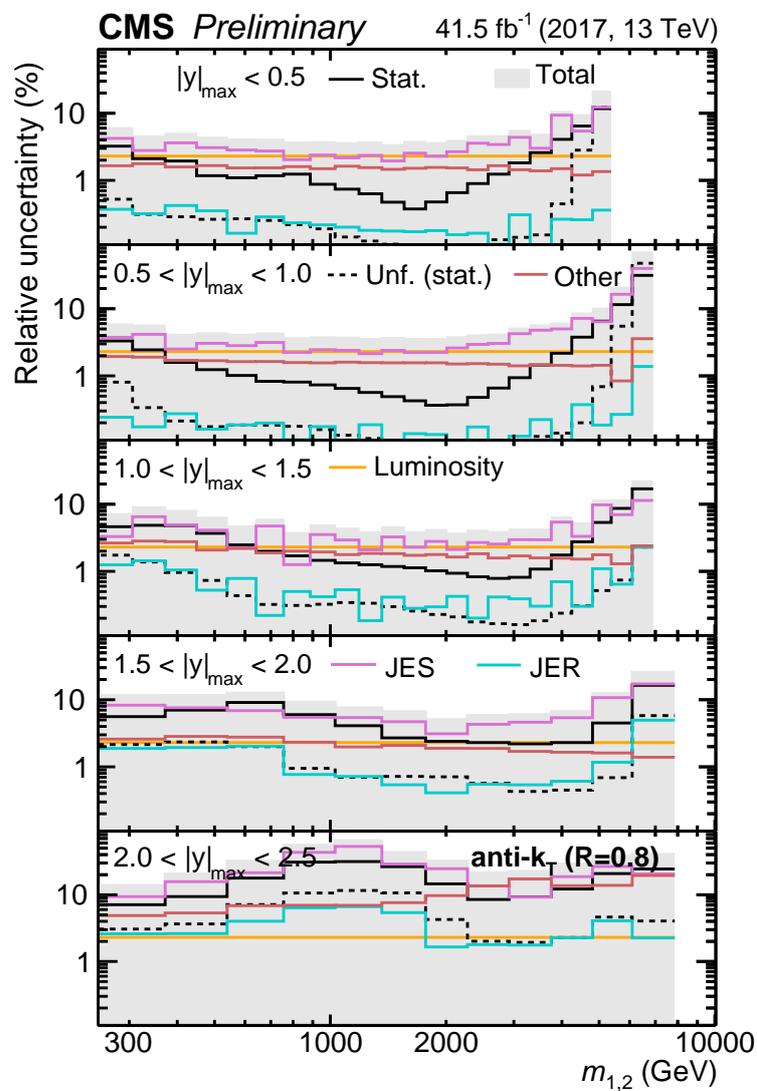


Figure H.1: Relative uncertainty (%) for all experimental sources on the measured dijet mass cross section from the 2017 dataset. Individual contributions are explained in Section 5.7. The abbreviation “Unf.” refers to the unfolding uncertainty. The total uncertainty is estimated as the sum in quadrature of all statistical and systematic uncertainty sources.

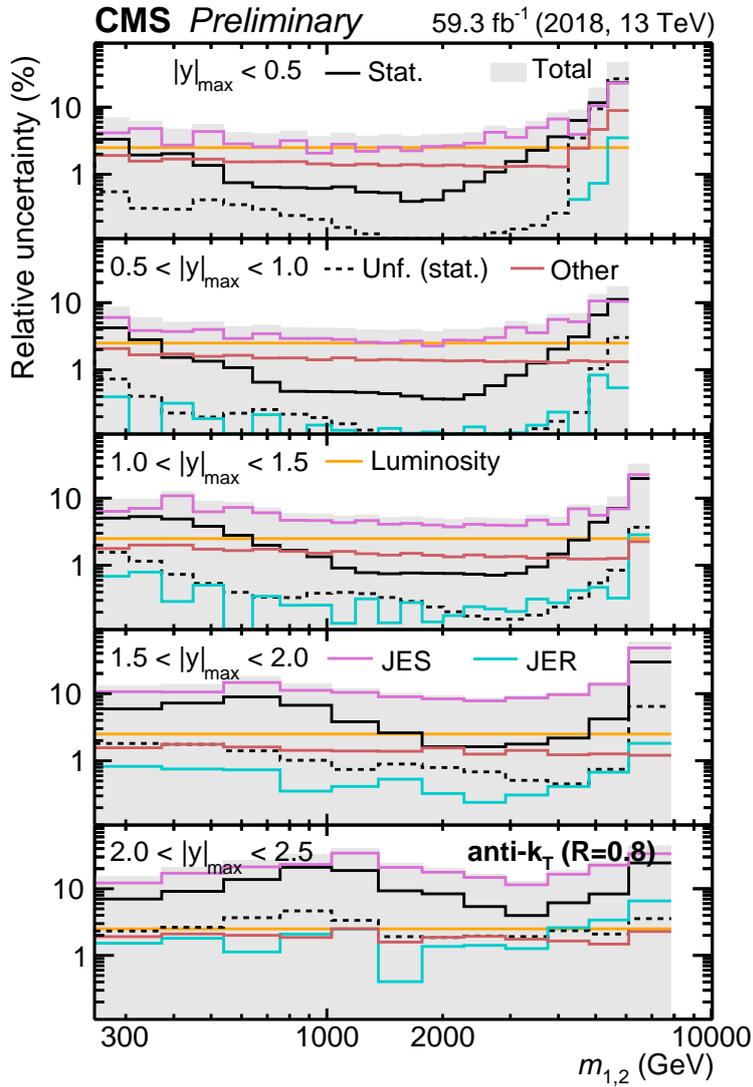


Figure H.2: Relative uncertainty (%) for all experimental sources on the measured dijet mass cross section from the 2018 dataset. Individual contributions are explained in Section 5.7. The abbreviation “Unf.” refers to the unfolding uncertainty. The total uncertainty is estimated as the sum in quadrature of all statistical and systematic uncertainty sources.

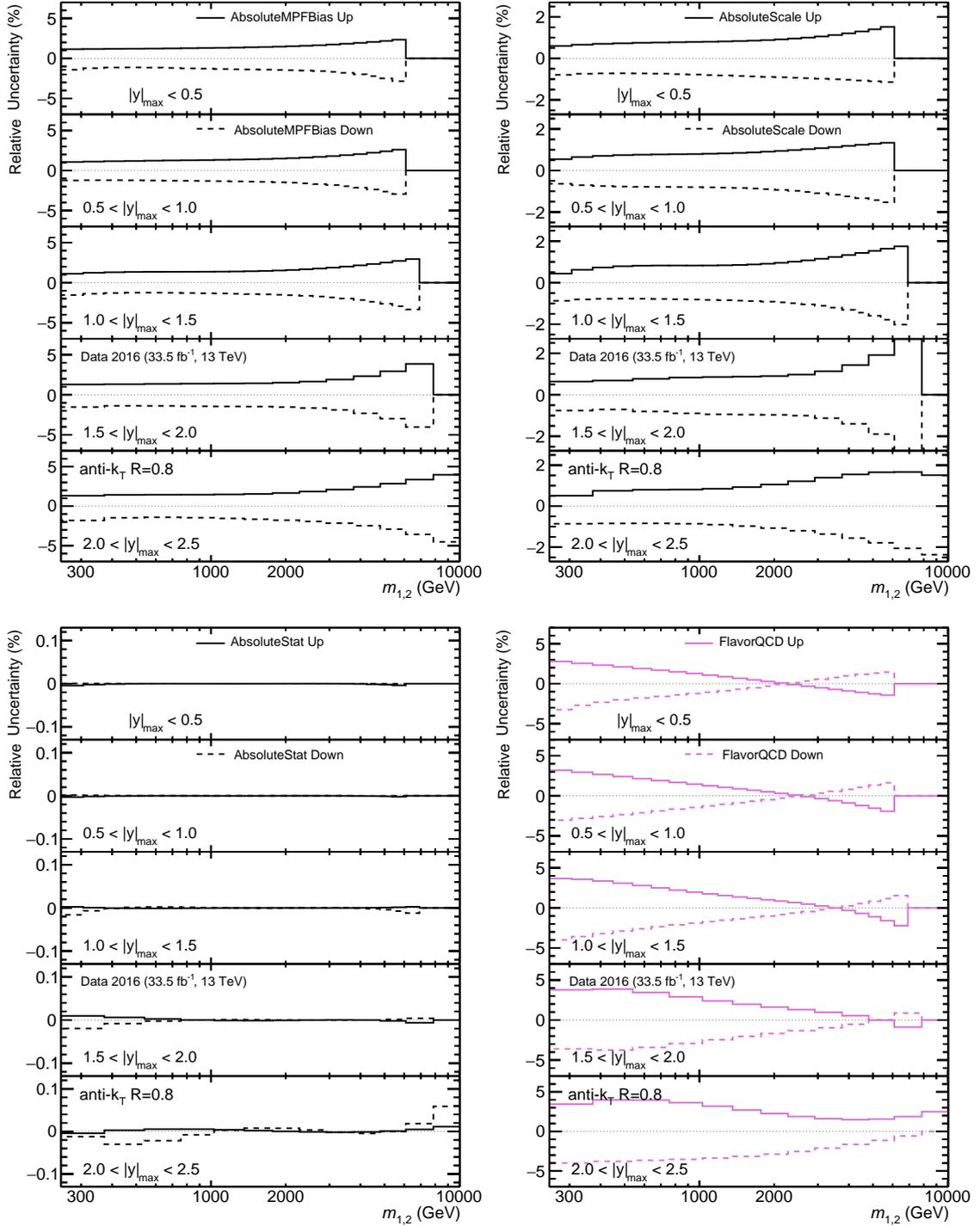


Figure H.3: Relative uncertainty (%) for individual JES sources from the 2016 dataset: AbsoluteMPFBias (top left), AbsoluteScale (top right), AbsoluteStat (bottom left), and FlavorQCD (bottom right). Both up and down variations are shown.

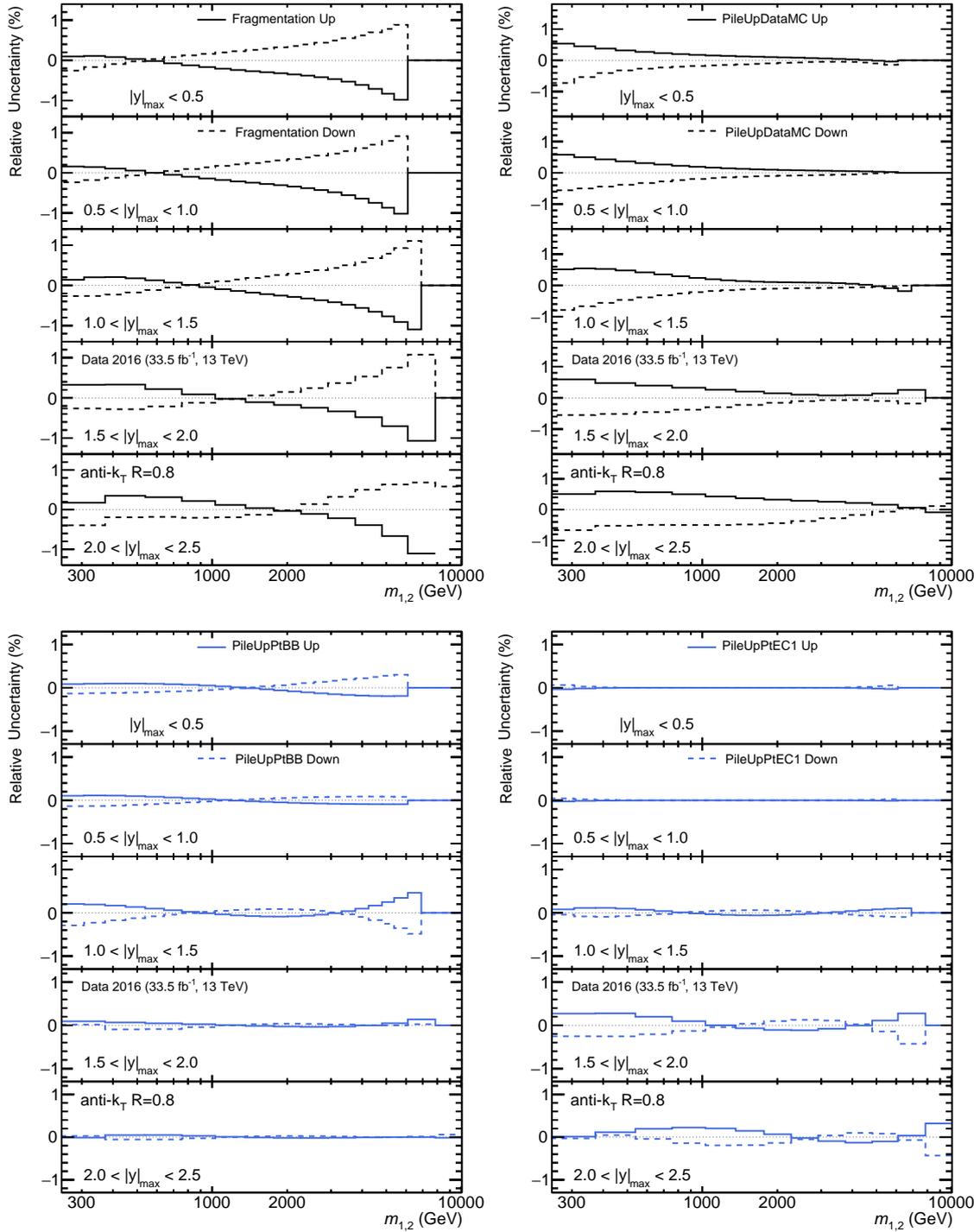


Figure H.4: Relative uncertainty (%) for individual JES sources in the 2016 dataset: Fragmentation (top left), PileUpDataMC (top right), PileUpPtBB (bottom left), and PileUpPtEC1 (bottom right). Both up and down variations are shown.

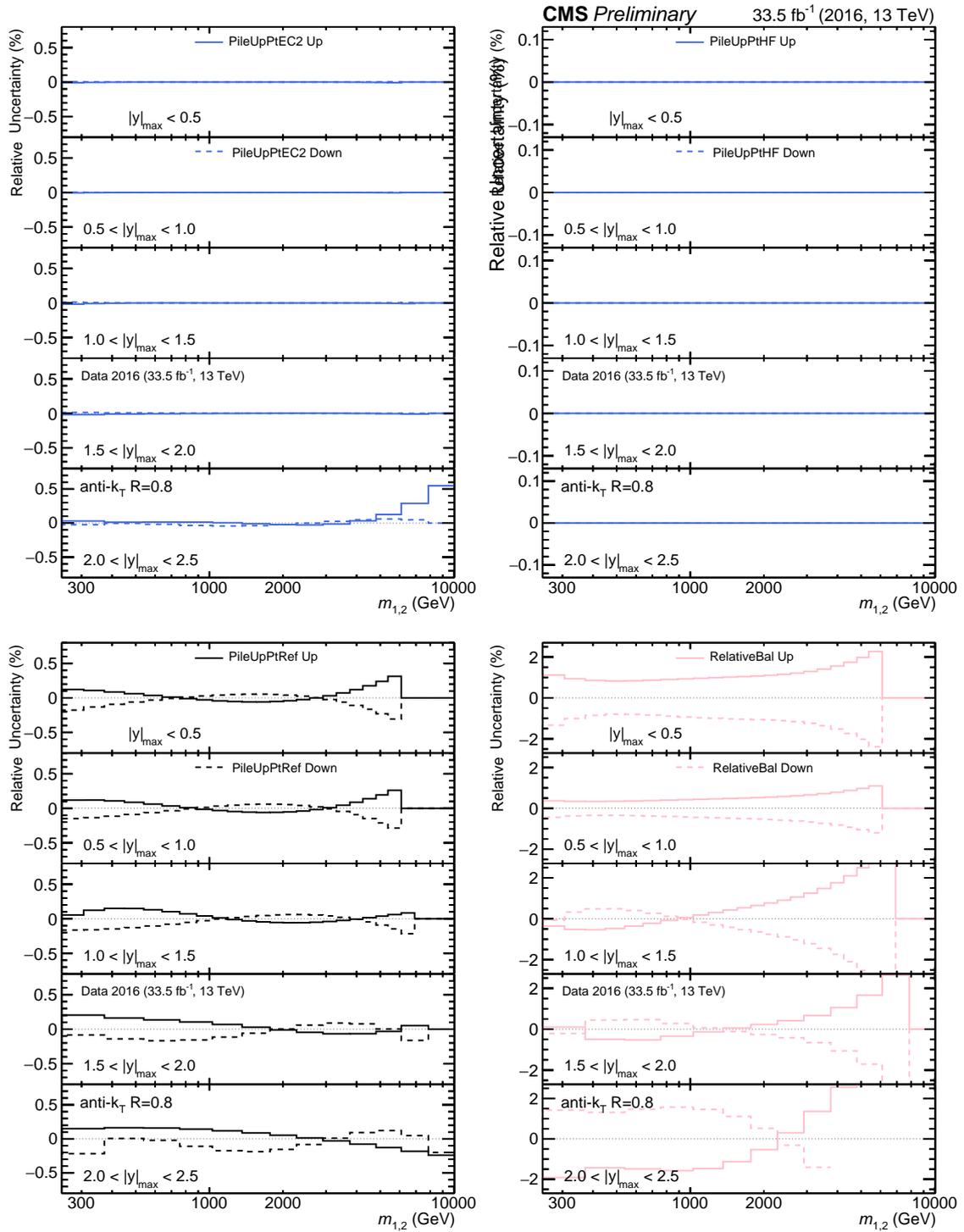


Figure H.5: Relative uncertainty (%) for individual JES sources in the 2016 dataset: PileUpPtEC2 (top left), PileUpPtHF (top right), PileUpPtRef (bottom left), and RelativeBal (bottom right). Both up and down variations are shown.

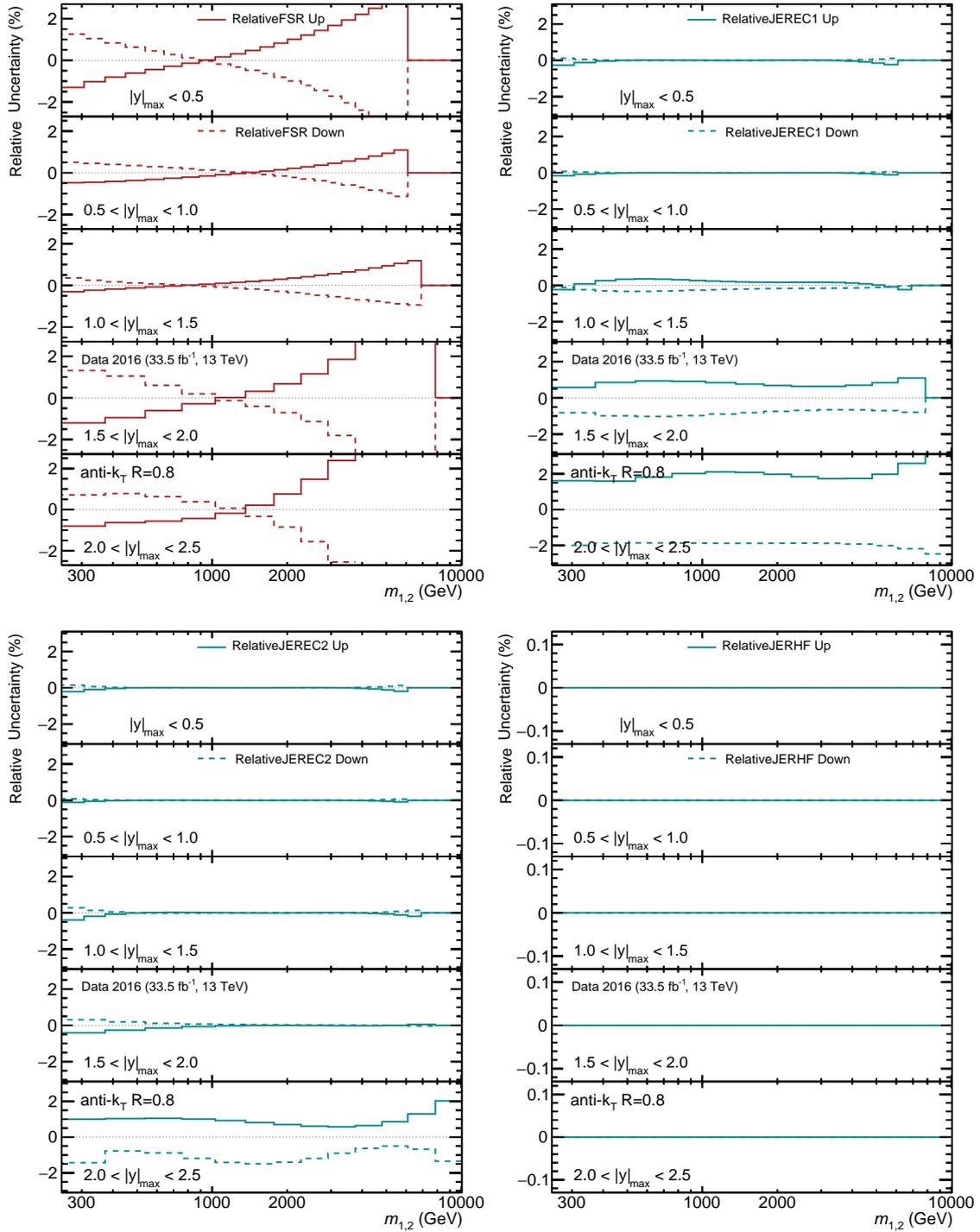


Figure H.6: Relative uncertainty (%) for individual JES sources in the 2016 dataset: RelativeFSR (top left), RelativeJEREC1 (top right), RelativeJEREC2 (bottom left), and RelativeJERHF (bottom right). Both up and down variations are shown.

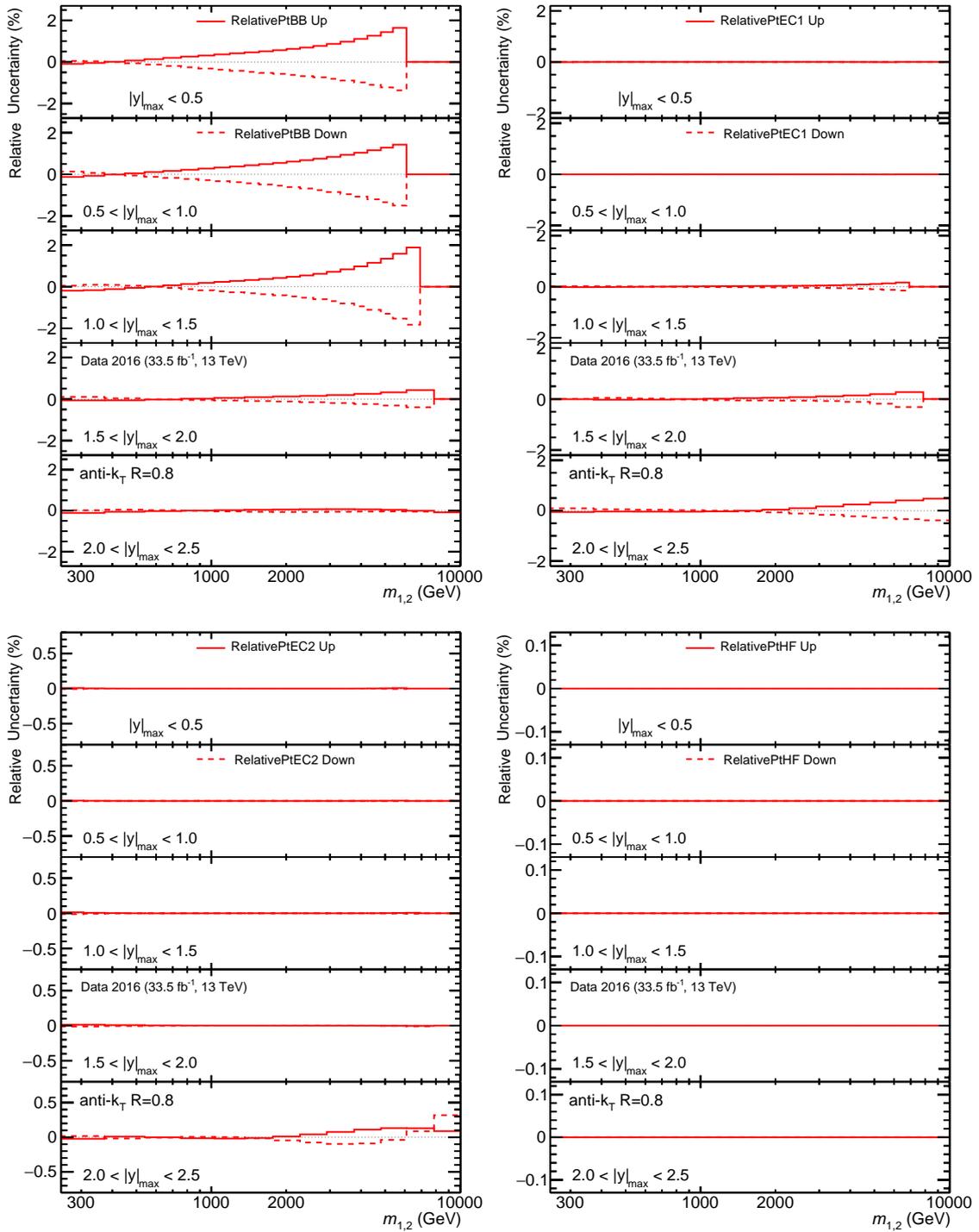


Figure H.7: Relative uncertainty (%) for individual JES sources in the 2016 dataset: RelativePtBB (top left), RelativePtEC1 (top right), RelativePtEC2 (bottom left), and RelativePtHF (bottom right). Both up and down variations are shown.

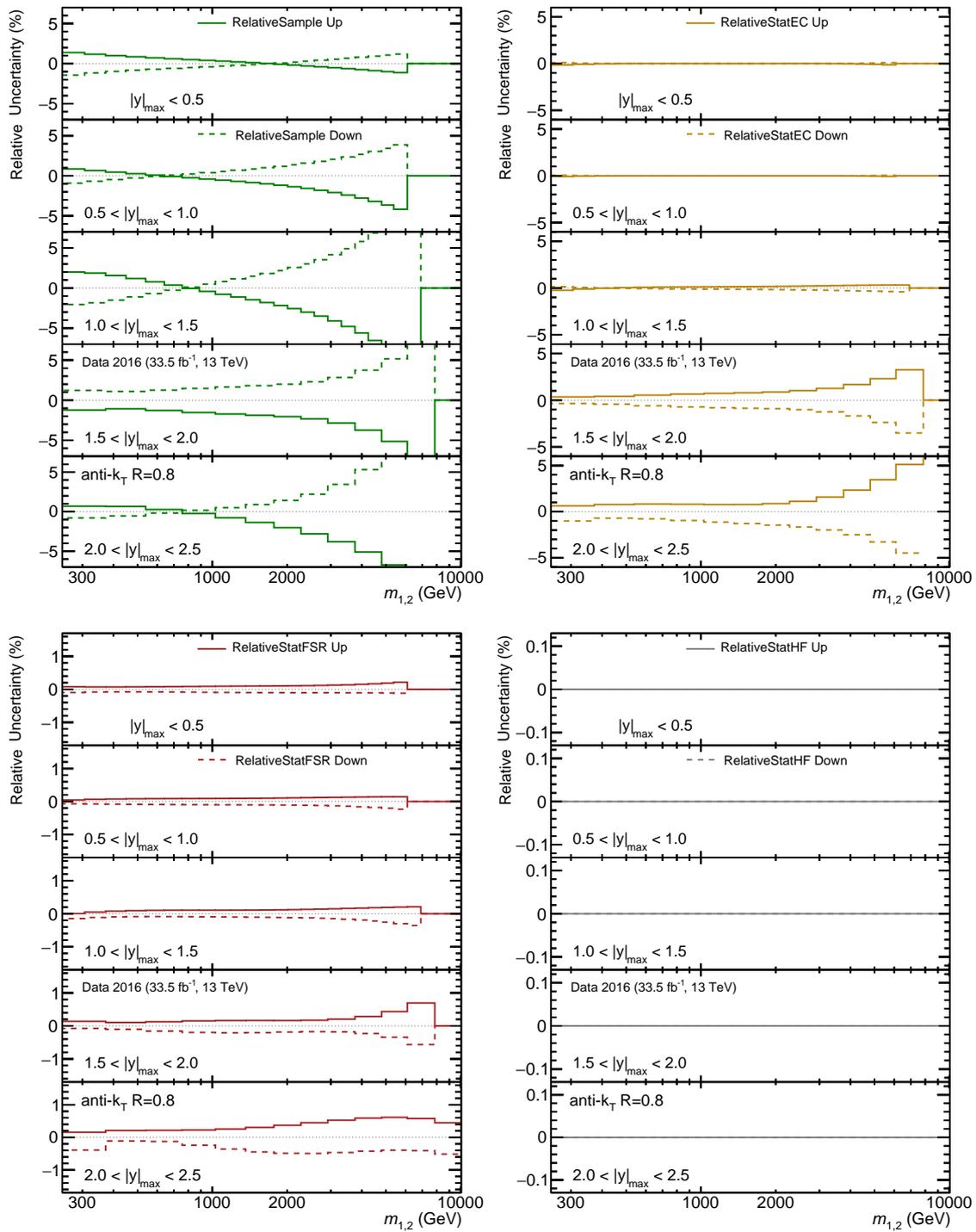


Figure H.8: Relative uncertainty (%) for individual JES sources in the 2016 dataset: RelativeSample (top left), RelativeStatEC (top right), RelativeStatFSR (bottom left), and RelativeStatHF (bottom right). Both up and down variations are shown.

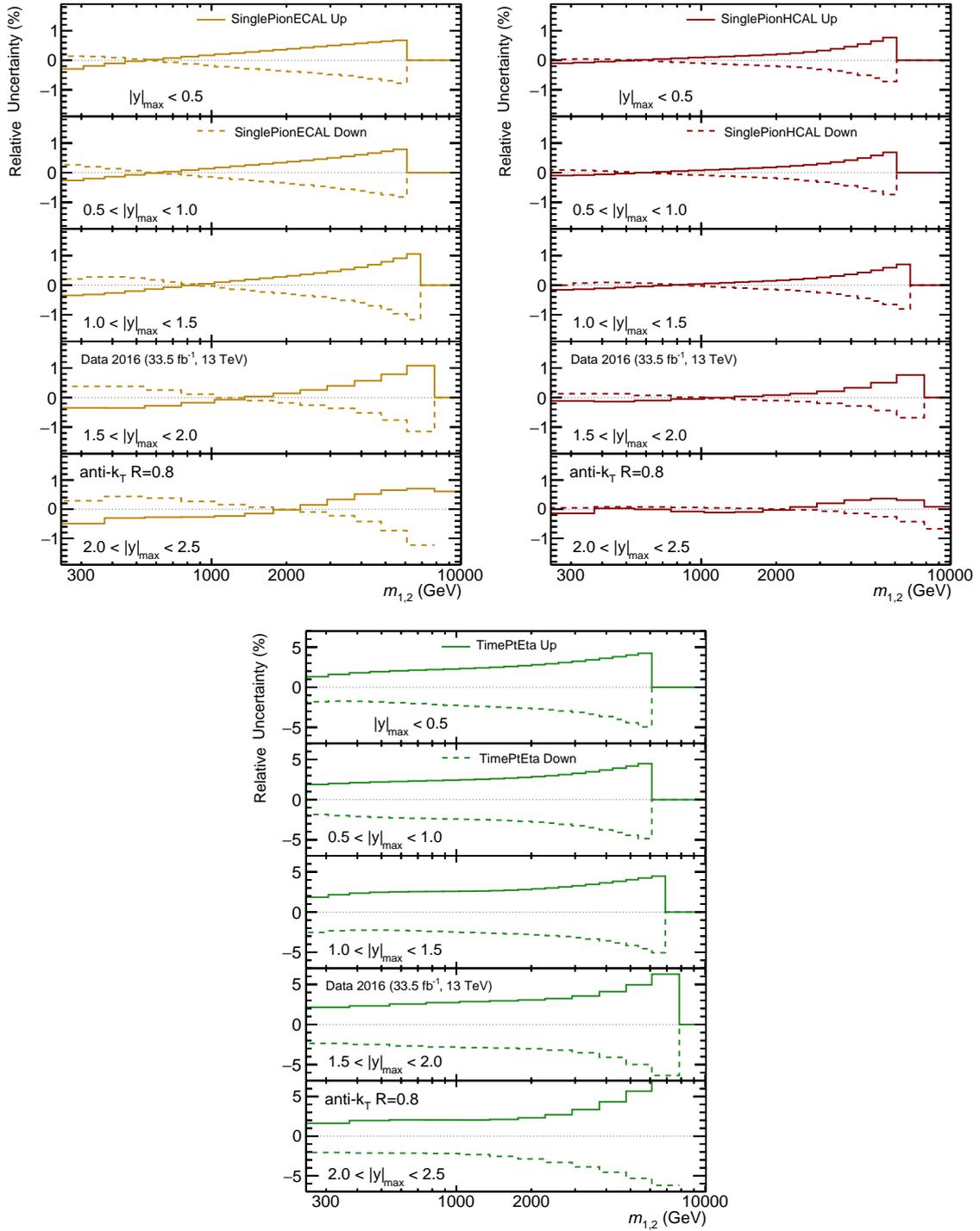


Figure H.9: Relative uncertainty (%) for individual JES sources in the 2016 dataset: SinglePionECAL (top left), SinglePionHCAL (top right), and TimePtEta (bottom). Both up and down variations are shown.

Appendix I

NON-PERTURBATIVE CORRECTIONS

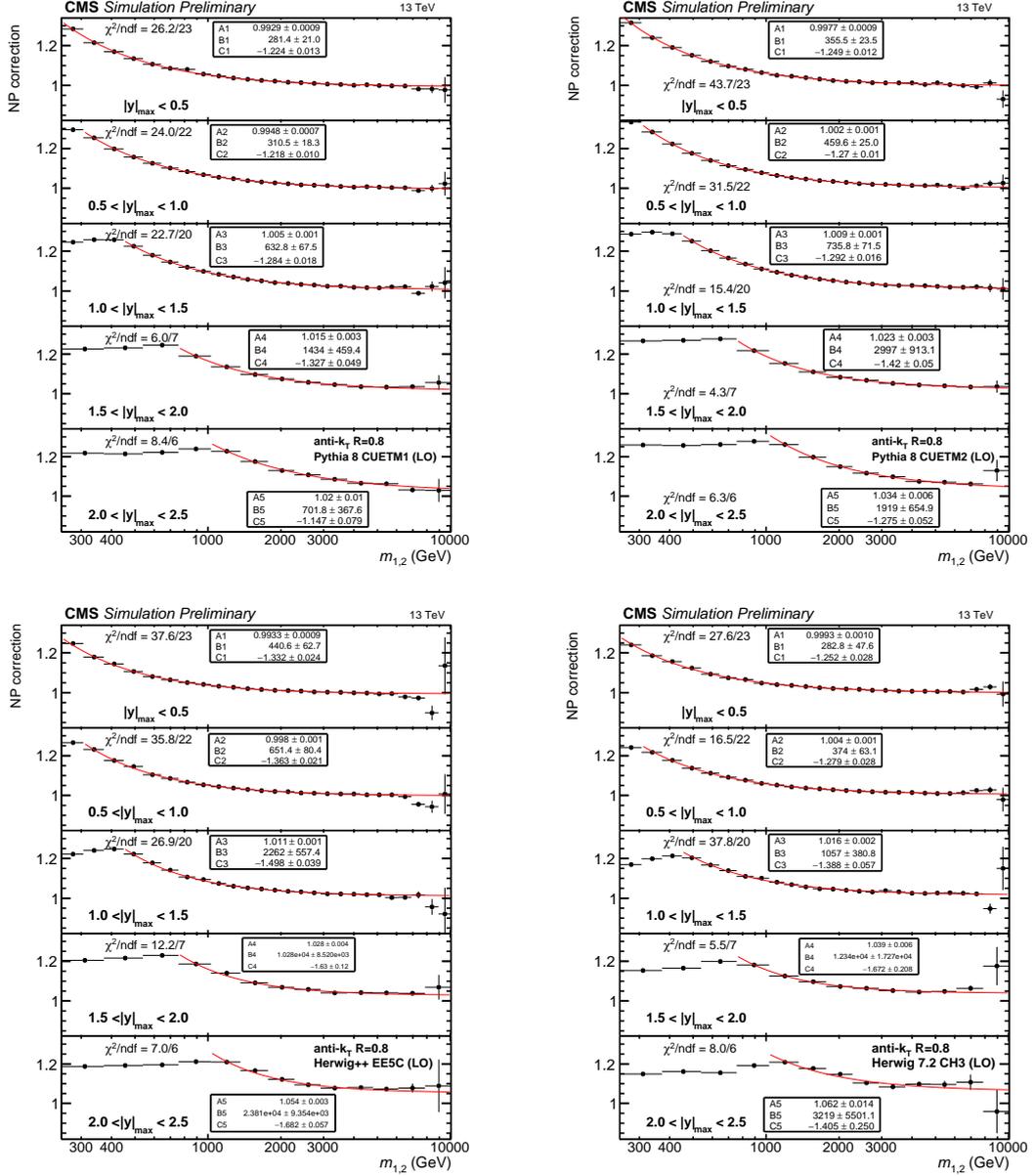


Figure I.1: Individual non-perturbative correction factors from individual LO MC generators: Pythia 8 CUETP8M1 (top left), Pythia 8 CUETP8M2T4 (top right), Herwig++ EE5C (bottom left), and Herwig 7 CH3 (bottom right).

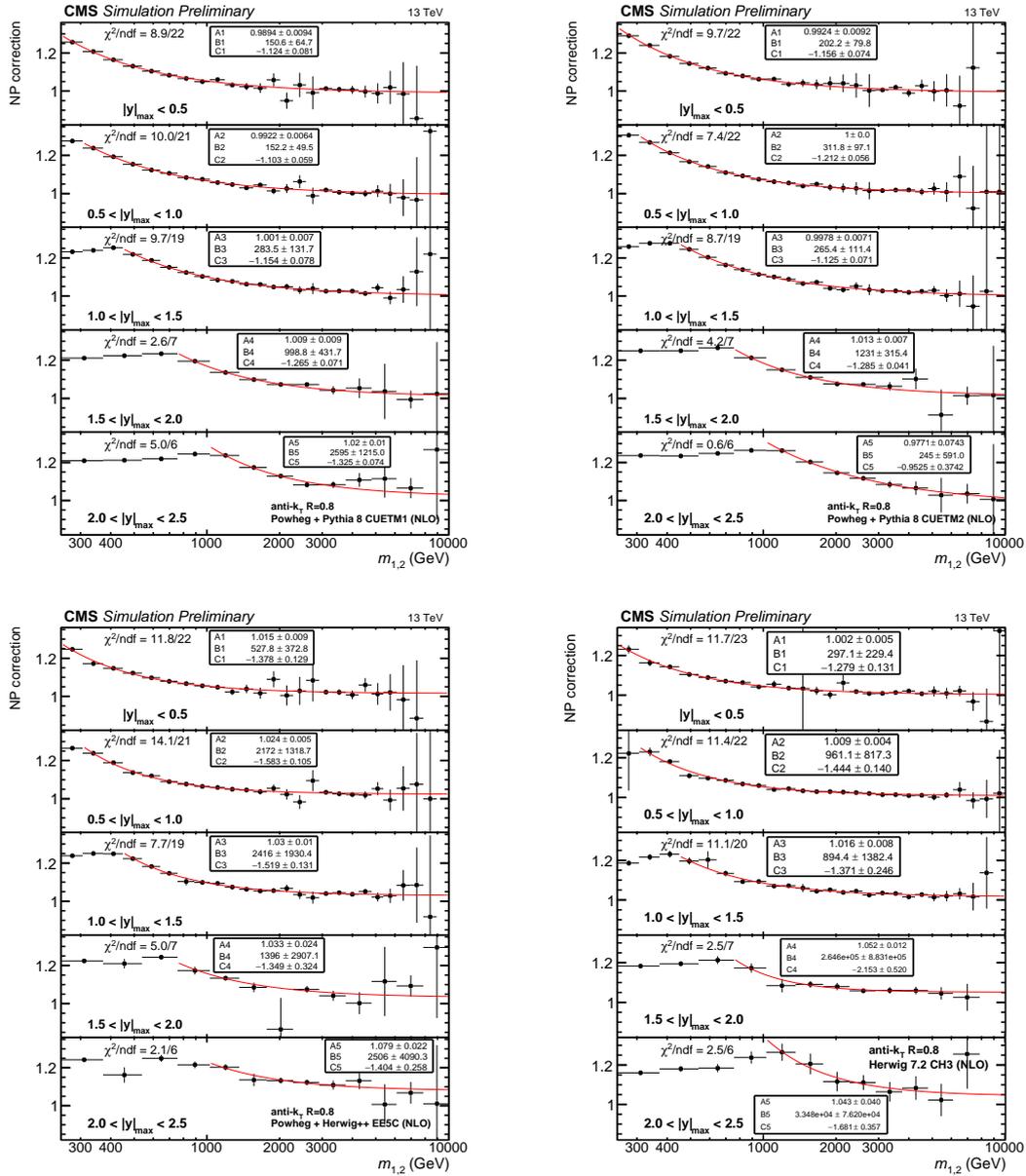


Figure I.2: Individual non-perturbative correction factors from individual MC generators: Powheg + Pythia 8 CUETP8M1 (top left), Powheg + Pythia 8 CUETP8M2T4 (top right), Powheg + Herwig++ EE5C (bottom left), and Herwig 7.2 CH3 (bottom right).

Appendix J

COMPLEMENTARY DATA TO THEORY COMPARISONS

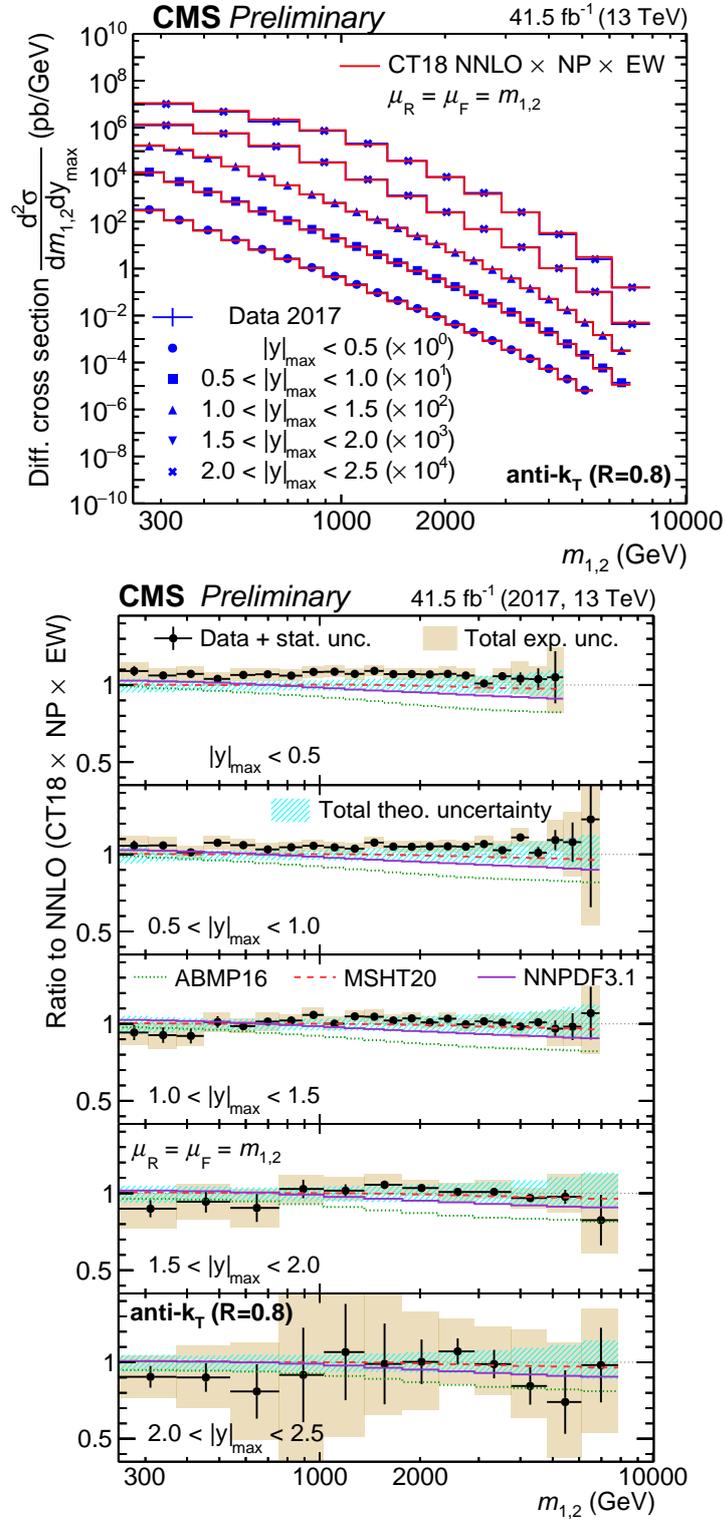


Figure J.1: Data to theory comparison. Experimental differential cross sections from the 2017 dataset overlaid on top of the fixed order predictions (top), and their ratio with respective uncertainties (bottom).

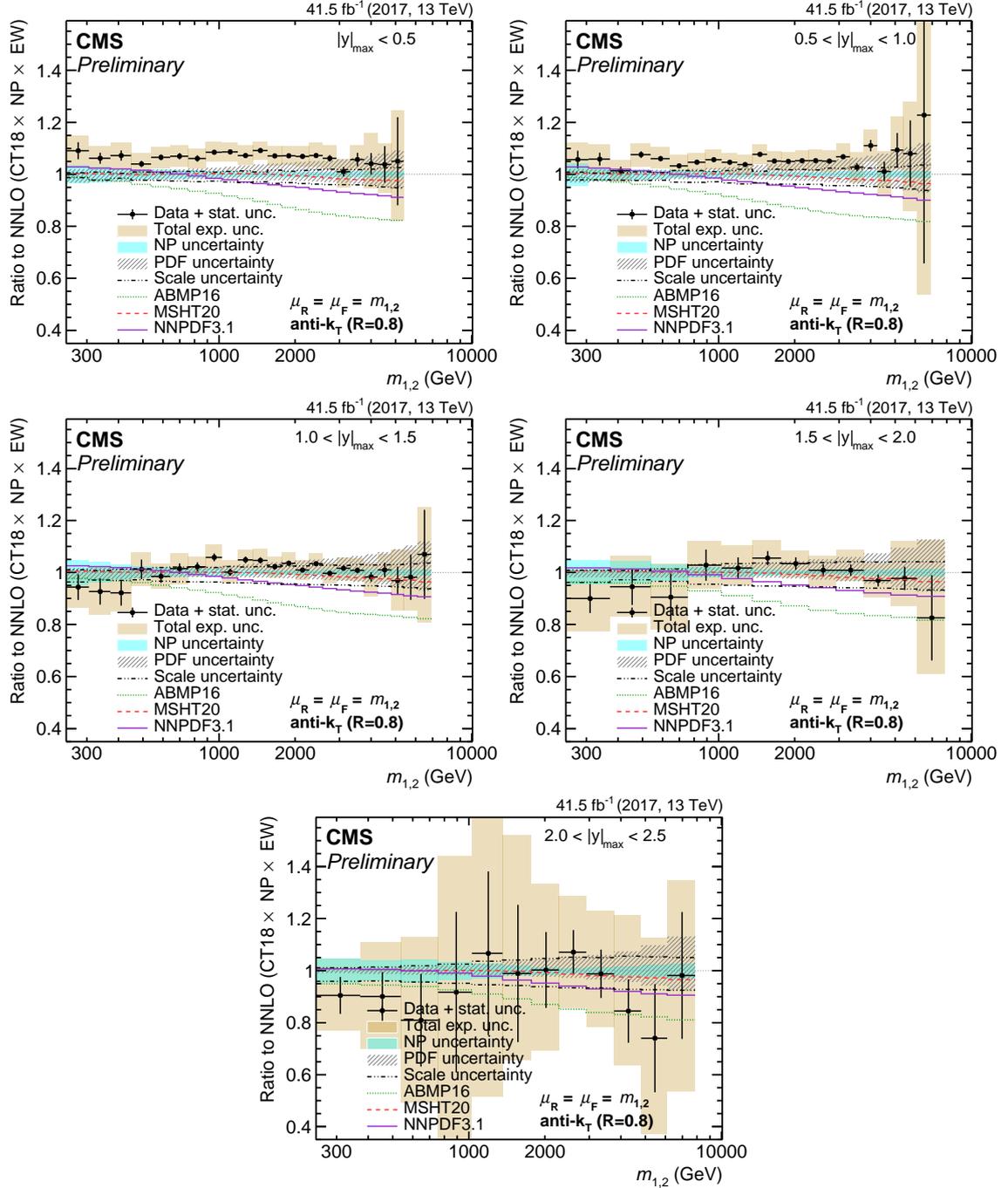


Figure J.2: Individual data to theory comparisons. Experimental differential cross sections from the 2017 dataset compared to fixed order theory predictions with respective uncertainties. Total theory uncertainty is decomposed into contributions from different sources. Each plot corresponds to a different rapidity region.

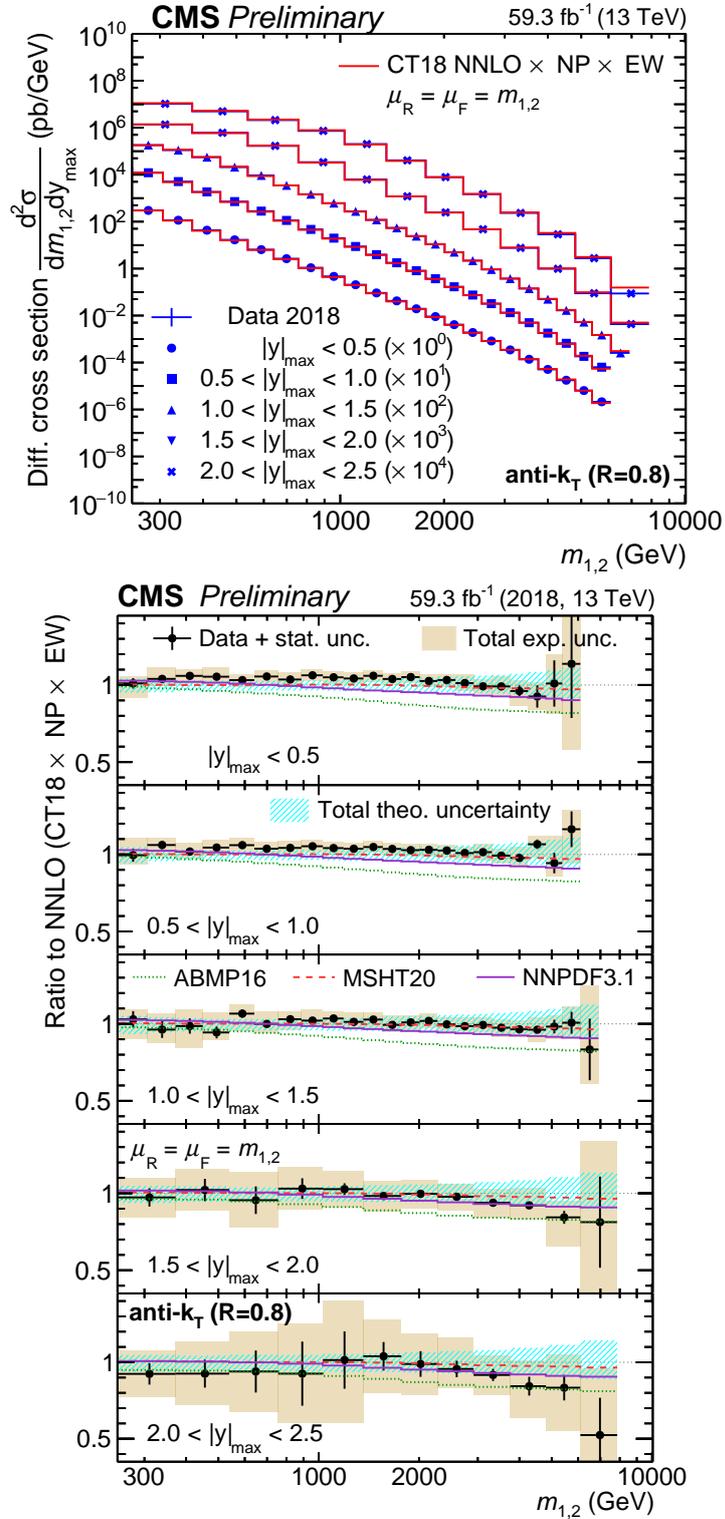


Figure J.3: Data to theory comparison. Experimental differential cross sections from the 2018 dataset overlaid on top of the fixed order predictions (top), and their ratio with respective uncertainties (bottom).

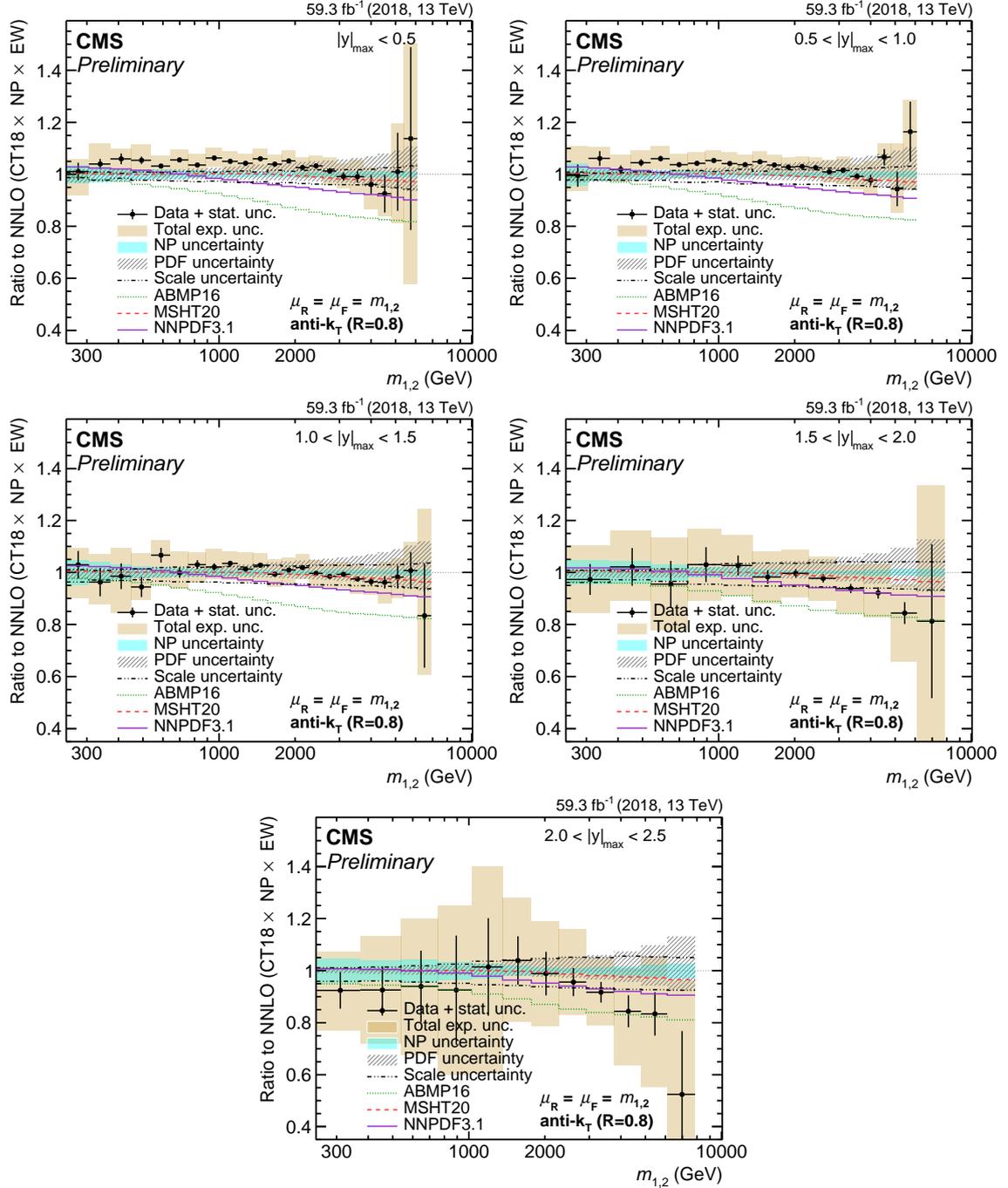


Figure J.4: Individual data to theory comparisons. Experimental differential cross sections from the 2018 dataset compared to fixed order theory predictions with respective uncertainties. Total theory uncertainty is decomposed into contributions from different sources. Each plot corresponds to a different rapidity region.