ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ
ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ
ΜΗΧΑΝΙΚΩΝ

ΕΚΕΦΕ ΔΗΜΟΚΡΙΤΟΣ

ΙΝΣΤΙΤΟΥΤΟ ΝΑΝΟΕΠΙΣΤΗΜΗΣ
ΚΑΙ ΝΑΝΟΤΕΧΝΟΛΟΓΙΑΣ

ΙΝΣΤΙΤΟΥΤΟ ΠΥΡΗΝΙΚΗΣ ΚΑΙ
ΣΩΜΑΤΙΔΙΑΚΗΣ ΦΥΣΙΚΗΣ

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
«Φυσική και Τεχνολογικές Εφαρμογές»

Μέτρηση της μάζας του top κουάρκ σε γεγονότα παραγωγής μονήρων top κουάρκ από συγκρούσεις πρωτονίων ενέργειας κέντρου μάζας 13 TeV στο πείραμα CMS

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του Νικολάου Βασσάκη

Επιβλέπων: Κουσουρής Κωνσταντίνος
Αναπληρωτής Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος, 2024

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF APPLIED MATHEMATICS AND PHYSICAL SCIENCES

SCHOOL OF MECHANICAL ENGINEERING

NCSR DEMOKRITOS

INSTITUTE OF NANOSCIENCE AND NANOTECHNOLOGY

INSTITUTE OF NUCLEAR AND PARTICLE PHYSICS

Interdepartmental Graduate Program
"Physics and Technological Applications"

Measurement of the top quark mass in single top quark production events from proton collisions at a center of mass energy of 13 TeV in the CMS experiment

---

MASTER'S THESIS

by Nikolaos Vassakis

Supervisor: Konstantinos Kousouris
Associate Professor NTUA

Athens, October, 2024

Περίληψη

Σε αυτήν την εργασία πραγματοποιείται μέτρηση της μάζας του top κουάρκ χρησιμοποιώντας δεδομένα που συλλέχθηκαν σε ενέργεια $\sqrt{s} = 13$ TeV από συγκρούσεις πρωτονίων στο πείραμα CMS με ολοκληρωμένη φωτεινότητα 36.3 fb^{-1}. Η μάζα του top κουάρκ μετράται μέσω του δευτερεύοντος μηχανισμού παραγωγής μονήρων top κουάρκ στο t-κανάλι από ηλεκτρασθενείς αλληλεπιδράσεις ανακατασκευάζοντας την αμετάβλητη μάζα των προϊόντων διάσπασης. Τα υποψήφια γεγονότα επιλέγονται απαιτώντας ένα μιόνιο και ακριβώς δύο πίδακες, εκ των οποίων ο ένας να προέρχεται από bottom κουάρκ, στην τελική κατάσταση. Για να βελτιωθεί η ανακατασκευή της μάζας του top κουάρκ, χρησιμοποιούνται τεχνικές μηχανικής μάθησης, όπως τα BDT, για να εκτιμηθεί με ακρίβεια η διαμήκης ορμή του νετρίνο και η εγκάρσια ορμή του πίδακα από bottom κουάρκ, ενώ εφαρμόζεται κινηματική προσαρμογή για να περιοριστεί η μάζα του W μποζονίου στη διάσπαση W$\to$lν. Ένα ξεχωριστό BDT χρησιμοποιείται για να απομονωθεί το σήμα από το υπόβαθρο. Η μάζας του top κουάρκ κατά τη μέτρηση βρέθηκε 171.25$\pm$0.24 GeV (stat+prof). Αυτό το αποτέλεσμα υποδεικνύει μια πολλά υποσχόμενη μέθοδο για ουσιαστική μείωση των στατιστικών και συστηματικών αβεβαιοτήτων σε μελλοντικές μετρήσεις.

Abstract

In this thesis a measurement of the top quark mass is performed using a data sample collected at $\sqrt{s} = 13$ TeV from proton-proton collisions by the CMS experiment, corresponding to an integrated luminosity of 36.3 fb^{-1}. The top quark mass is measured in the subdominant production process of single top quarks in the t-channel via electroweak interaction by reconstructing the invariant mass of the decay products. Candidate events are selected by requiring a muon and exactly two jets, one of which is identified as originating from a bottom quark, in the final state. To enhance the mass reconstruction of the top quark, machine learning techniques, such as BDTs, are employed to accurately estimate the neutrino's longitudinal momentum and the transverse momentum of the b-tagged jet, while a kinematic fit is applied to constrain the mass of the W boson in the W$\to$lν decay. A separate BDT for classification is used to distinguish the signal from the background. The top quark mass is found to be 171.25$\pm$0.24 GeV (stat+prof). This result demonstrates a promising method for significantly reducing statistical and systematic uncertainties in future measurements.

Ευχαριστίες

Ευχαριστώ θερμά την ομάδα του CMS για τη συνεργασία καθ' όλη τη διάρκεια εκπόνησης της εργασίας μου. Ευχαριστώ ιδιαιτέρως τον επιβλέποντα καθηγητή κ. Κωνσταντίνο Κουσουρή για την πολύτιμη καθοδήγηση και τον υποψήφιο διδάκτορα Θοδωρή Χατζησταύρου για τη συνεχή υποστήριξη.

Εκτεταμένη Περίληψη

Το Καθιερωμένο Πρότυπο (ΚΠ) της φυσικής στοιχειωδών σωματιδίων είναι μια κβαντική θεωρία πεδίου που αποσκοπεί στην περιγραφή των θεμελιωδών σωματιδίων — όπως είναι τα λεπτόνια, τα κουάρκ, τα μποζόνια βαθμίδας και το μποζόνιο Higgs — και των αλληλεπιδράσεων τους. Αποτελεί ένα από τα πιο καλά εδραιωμένα θεωρητικά πλαίσια στη φυσική, πολλαπλώς επιβεβαιωμένο από πειραματικά δεδομένα, παρά τα ανοιχτά, προς το παρόν αναπάντητα, ερωτήματα.

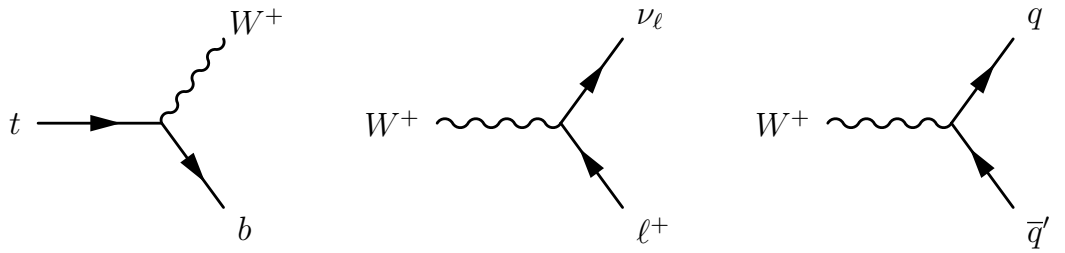
Τα φερμιόνια, δηλαδή τα σωματίδια με σπιν 1/2 που υπακούουν στη στατιστική Fermi-Dirac και στην απαγορευτική αρχή του Pauli, είναι τα βασικά δομικά στοιχεία της ύλης και αποτελούνται από τα κουάρκ και τα λεπτόνια. Το ΚΠ περιλαμβάνει δώδεκα φερμιόνια, κατανεμημένα σε τρεις γενιές, με κάθε γενιά να περιλαμβάνει ένα κουάρκ τύπου up (u,c,t), ένα κουάρκ τύπου down (d,s,b) και ένα φορτισμένο λεπτόνιο (e,μ,τ) με το αντίστοιχο νετρίνο (ν_e,ν_μ,ν_τ). Τα σωματίδια των ανώτερων γενεών διασπώνται σε σωματίδια χαμηλότερης γενιάς, ενώ τα κουάρκ βρίσκονται πάντα δέσμια σε αδρόνια, χωρίς να υπάρχουν ποτέ ως ελεύθερα σωματίδια, λόγω του χρωματικού τους περιορισμού.

Εκτός από τα φερμιόνια, το ΚΠ περιγράφει και τις αλληλεπιδράσεις μεταξύ σωματιδίων μέσω τριών θεμελιωδών δυνάμεων: της ηλεκτρομαγνητικής, της ισχυρής πυρηνικής και της ασθενούς πυρηνικής δύναμης. Αυτές οι δυνάμεις προκύπτουν από συμμετρίες στη δομή του ΚΠ, οι οποίες διέπονται από την τοπικές συμμετρίες βαθμίδας. Η ενοποιημένη ηλεκτρασθενής αλληλεπίδραση βασίζεται στην ομάδα συμμετρίας $\mathrm{SU(2)_L}\times\mathrm{U(1)_Y}$ και περιλαμβάνει τα μποζόνια βαθμίδας, δηλαδή τα φωτόνια, τα $W^\pm$ και τα Z μποζόνια. Η ισχυρή αλληλεπίδραση βασίζεται στην ομάδα $\mathrm{SU(3)_C}$ και περιλαμβάνει τα γκλουόνια. Τα σωματίδια που αποτελούν τους φορείς των δυνάμεων υπακούουν στη στατιστική Bose-Einstein. Το μποζόνιο Higgs, το οποίο είναι ένα σωματίδιο βαθμίδας με σπιν 0, κατέχει μοναδικό ρόλο στο ΚΠ, καθώς αποτελεί το θεμέλιο του μηχανισμού μέσω του οποίου τα υπόλοιπα σωματίδια αποκτούν τη μάζα τους.

Ωστόσο, παρά την επιτυχία στην περιγραφή των θεμελιωδών σωματιδίων και των αλληλεπιδράσεών τους, το ΚΠ δεν περιλαμβάνει τη δύναμη της βαρύτητας, γεγονός που υποδηλώνει ότι αποτελεί μια ελλιπή θεωρία για την ερμηνεία του φυσικού κόσμου. Επίσης, το ΚΠ αδυνατεί μεταξύ άλλων να εξηγήσει φαινόμενα όπως είναι η σκοτεινή ύλη, η φύση του μποζονίου Higgs ή η προέλευση της παραβίασης συμμετρίας CP. Οι απαντήσεις στα παραπάνω ερωτήματα αποτελούν την κινητήρια δύναμη στην αναζήτηση νέας φυσικής και υποδεικνύουν την ανάγκη της επέκτασης του ΚΠ.

Το top κουάρκ είναι ένα φερμιόνιο με σπιν 1/2 και αποτελεί το βαρύτερο θεμελιώδες σωματίδιο στο Καθιερωμένο Πρότυπο. Φέρει ηλεκτρικό φορτίο $(+2/3)e$ και σχηματίζει διπλέτα ασθενούς ισοσπίν μαζί με το bottom κουάρκ έχοντας προβολή ισοσπίν I_3=+1/2. Το top κου-

άρκ ανακαλύφθηκε από τις συνεργασίες CDF και D0 στο Tevatron το 1995 και, λόγω της μεγάλης του μάζας, η διάρκεια ζωής του είναι σημαντικά μικρότερη του χρόνου που απαιτείται για να αλληλεπιδράσει ισχυρά προκειμένου να αδρονοποιηθεί ή για να σχηματίσει δέσμιες αδρονικές καταστάσεις. Το top κουάρκ διασπάται σχεδόν αποκλειστικά μέσω της ασθενούς αλληλεπίδρασης σε έναν b κουάρκ και έναν μποζόνιο W^+, το οποίο με τη σειρά του διασπάται είτε λεπτονικά σε ένα θετικά φορτισμένο λεπτόνιο και το αντίστοιχο αντινετρίνο, είτε αδρονικά σε ένα ζεύγος κουάρκ-αντικουάρκ.



Σχήμα 1: (α) Η κυρία διάσπαση του top κουάρκ σε ένα μποζόνιο W^+ και έναν b κουάρκ. Το μποζόνιο W^+ διασπάται με τη σειρά του είτε (β) λεπτονικά, είτε (γ) αδρονικά.

Όσον αφορά τον τρόπο παραγωγής του top κουάρκ στο Μεγάλο Επιταχυντή Αδρονίων (LHC), ο κυρίαρχος μηχανισμός είναι η παραγωγή ζεύγους $\mathrm{t\bar{t}}$ μέσω της ισχυρής αλληλεπίδρασης. Συγκεκριμένα σε ενέργεια κέντρου μάζας $\sqrt{s} = 13$ TeV σε συγκρούσεις πρωτονίων, τα ζεύγη $\mathrm{t\bar{t}}$ παράγονται κυρίως μέσω συγχώνευσης γκλουονίων (gluon fusion) σε ποσοστό 90%, ενώ η εξαΰλωση κουάρκ-αντικουάρκ ($\mathrm{q\bar{q}}$) συνεισφέρει σε ποσοστό περίπου 10%. Ο λόγος που η κύρια συμβολή στη συνολική ενεργό διατομή προέρχεται από τη συγχώνευση γκλουονίων είναι ότι τα γκλουόνια βρίσκονται σε αφθονία στη συνάρτηση πυκνότητας των παρτονίων (PDF), σε αντίθεση με τα ζεύγη $\mathrm{q\bar{q}}$ σε κλίμακες TeV. Στον LHC σε 13 TeV, υποθέτοντας τη μάζα του top κουάρκ 172.5 GeV, η προβλεπόμενη ενεργός διατομή σε παραγωγή $\mathrm{t\bar{t}}$ (NNLO+NNLL) είναι $\sigma_{t\bar{t}} = 833.9^{+20.5+21.0}_{-30.0-21.0}$ pb, όπου η πρώτη αβεβαιότητα προέρχεται από την ανεξαρτησία κλίμακας και η δεύτερη από τις PDF και τη σταθερά σύζευξης της ισχυρής αλληλεπίδρασης α_s.

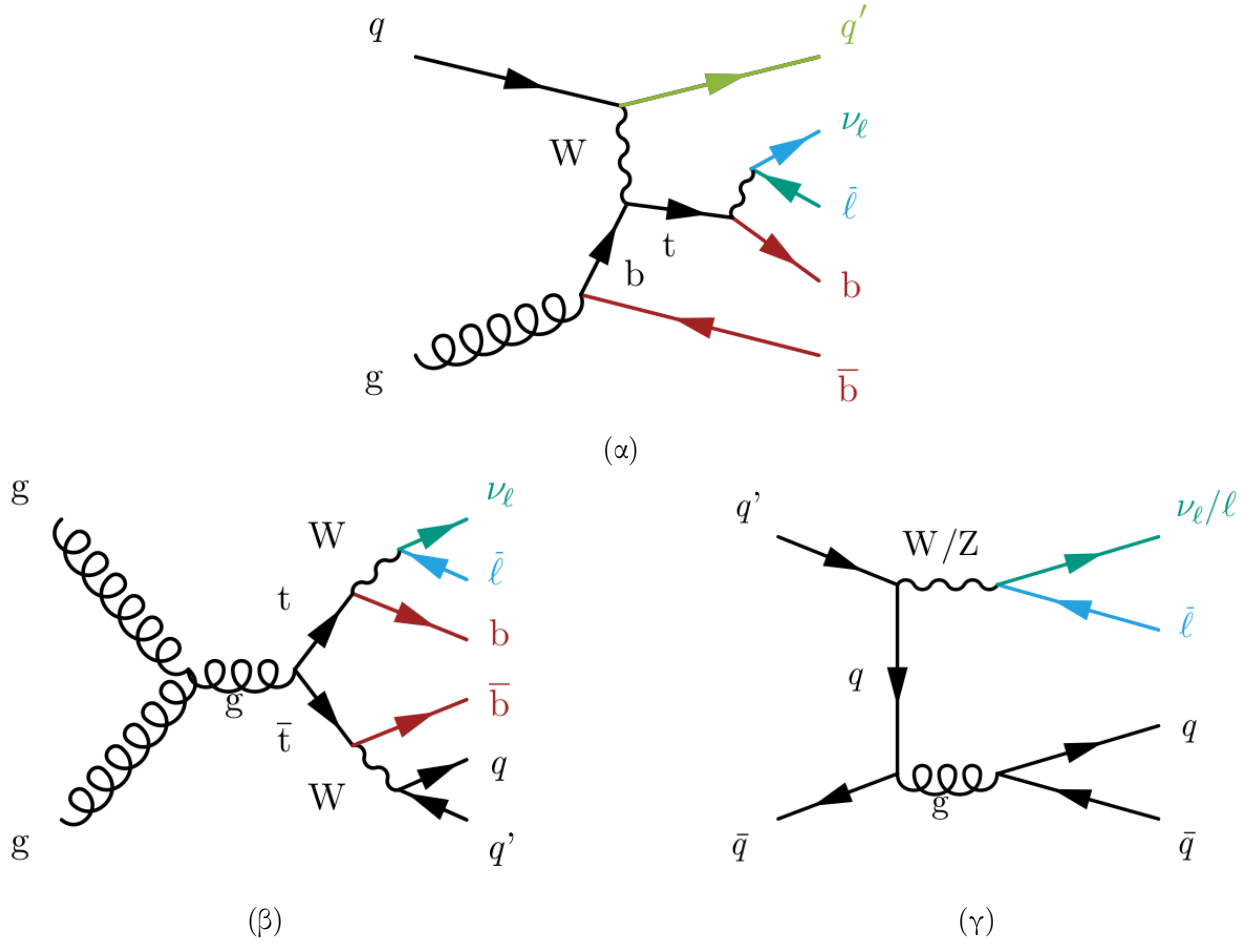
Εναλλακτικά, στον LHC πραγματοποιείται και παραγωγή μονήρων top κουάρκ μέσω της ασθενούς αλληλεπίδρασης και ενός W μποζονίου, με τρείς τρόπους: το t-κανάλι, το s-κανάλι και το tW-κανάλι. Ένα βασικό χαρακτηριστικό είναι η παρουσία του στοιχείου $\mathrm{V_{tb}}$ του πίνακα CMK, που επιτρέπει την άμεση παρατήρηση του κόμβου παραγωγής. Το t-κανάλι είναι ο κυρίαρχος τρόπος, συνεισφέροντας περίπου το 70% της συνολικής ενεργού διατομής. Ένα χαρακτηριστικό του τρόπου παραγωγής στο t-κανάλι είναι η μεγάλη ψευδοωκύτητα του πίδακα αδρονίων που προέρχεται από το ελαφρύ (light) κουάρκ. Στα 13 TeV, η υπολογισμένη διατομή για το t-κανάλι (NLO) είναι $214.2^{+2.4+3.3}_{-1.7-2.0}$ pb, ενώ η ενεργός διατομή παραγωγής top κουάρκ είναι 1.7 φορές μεγαλύτερη από αυτή των top αντικουάρκ, λόγω της μεγαλύτερης αφθονίας των up κουάρκ στο πρωτόνιο. Εκτός από το t-κανάλι, τα μονήρη top κουάρκ παράγονται περίπου 25% των περιπτώσεων μαζί με ένα πραγματικό W μποζόνιο μέσω του tW-καναλιού. Το s-κανάλι, το οποίο συμβάλλει μόνο 5% στην παραγωγή μονήρων top κουάρκ, περιλαμβάνει την παραγωγή

ενός εικονικού W μποζονίου. Όλες οι προβλέψεις των παραπάνω ενεργών διατομών συμφωνούν καλά με τις πειραματικές μετρήσεις από τα πειράματα ATLAS και CMS.

Η μάζα του top κουάρκ, m_{top}, αποτελεί μία σημαντική παράμετρο του ΚΠ. Ακριβείς μετρήσεις της m_{top} είναι απαραίτητες προκειμένου να υπάρξουν συγκρίσεις με τις θεωρητικές προβλέψεις, αποτελώντας, κατά αυτόν τον τρόπο, πύλη σε νέα φυσική. Συνεπώς, υπάρχει πληθώρα πειραματικών αποτελεσμάτων από μετρήσεις μάζας στον LHC, με την πλειοψηφία αυτών να βασίζονται στην κινηματική ανακατασκευή των τελικών προϊόντων της ημιλεπτονικής διάσπασης top κουάρκ-αντικουάρκ. Συνεπείς μετρήσεις από διάφορους τρόπους παραγωγής μπορούν να οδηγήσουν σε υψηλότερη ακρίβεια μέσω του συνδυασμού των αποτελεσμάτων και να συμβάλουν σε καλύτερη κατανόηση των συστηματικών αβεβαιοτήτων. Για το λόγο αυτό, είναι απαραίτητες μετρήσεις σε συμπληρωματικές τοπολογίες γεγονότων. Συγκεκριμένα, το t-κανάλι στην παραγωγή μονήρων top κουάρκ μπορεί να προσφέρει τέτοια μερικώς ανεξάρτητα δείγματα γεγονότων για μετρήσεις. Μία τέτοια πρόσφατη μέτρηση στο CMS από δεδομένα συγκρούσεων pp και ολοκληρωμένη φωτεινότητα 35.9 fb^{-1} στα 13 TeV, κατέγραψε τη μάζα του top κουάρκ ίση με m_{top}=172.13±0.32 GeV. Παρότι αυτή η μέτρηση αποτελεί μία ισχυρή εκτίμηση με μειωμένα σφάλματα, υπάρχει περιθώριο για περαιτέρω βελτίωση και στόχος της παρούσας εργασίας είναι η διερεύνηση εναλλακτικών τεχνικών ανακατασκευής και η πρόταση νέων μεθόδων για μία πιο ακριβή μέτρηση.

Λόγω της εξαιρετικά μικρής διάρκειας ζωής του, το top κουάρκ μελετάται μόνο μέσω των προϊόντων διάσπασής του. Επομένως, απαιτείται βαθιά κατανόηση της τοπολογίας του σήματος και των υποβάθρων. Στην παρούσα ανάλυση, η διαδικασία σήματος περιλαμβάνει την παραγωγή ενός μονήρους top κουάρκ μέσω του t-καναλιού, με χαρακτηριστική τελική κατάσταση που περιλαμβάνει ένα φορτισμένο λεπτόνιο, ένα νετρίνο που οδηγεί σε ανισορροπία εγκάρσιας ορμής, έναν πίδακα από ελαφρύ κουάρκ και έναν πίδακα από την αδρονοποίηση του bottom κουάρκ. Κύρια πηγή υποβάθρου αποτελεί η παραγωγή ζευγών top κουάρκ-αντικουάρκ, η τελική κατάσταση της οποίας συχνά μοιάζει με την τοπολογία του σήματος όταν το ένα top διασπάται λεπτονικά. Ένα άλλο σημαντικό υπόβαθρο περιλαμβάνει γεγονότα από W+jets και Z+jets. Τέλος, τελικές καταστάσεις πανομοιότυπες με αυτή του σήματος ενδέχεται να προκύψουν από και τα υπόλοιπα κανάλια δίασπασης από παραγωγή μονήρους top κουάρκ, από την παραγωγή δύο μποζονίων (diboson) ή από την παραγωγή πιδάκων λόγω QCD.

Η θεωρητική μοντελοποίηση των κατανομών διαφόρων μεταβλητών για τις διαδικασίες σήματος και υποβάθρου είναι απαραίτητη για τη σύγκριση με τα παρατηρούμενα μεγέθη στον ανιχνευτή και για την εξαγωγή εκτιμήσεων των παραμέτρων από τα δεδομένα. Για αυτόν τον σκοπό χρησιμοποιούνται γεννήτριες Monte Carlo (MC) για τη δημιουργία γεγονότων προσομοίωσης από συγκρούσεις υψηλής ενέργειας. Τα γεγονότα που αναλύονται πληρούν τα βασικά κριτήρια που στοχεύουν στην τελική κατάσταση του σήματος, με κύρια κατηγορία την 2J1T, η οποία περιλαμβάνει δύο ανακατασκευασμένους πίδακες, εκ των οποίων ο ένας προέρχεται από bottom κουάρκ. Η παρούσα ανάλυση μελετά μόνο την περίπτωση που το φορτισμένο λεπτόνιο είναι μιόνιο. Τα βασικά αυτά κριτήρια περιλαμβάνουν περικοπές στην εγκάρσια ορμή και την ψευδοωκύτητα: για τα μιόνια $p_T > 35$ GeV και $|\eta| < 2.1$, για τους light πίδακες $p_T > 40$ GeV και $|\eta| < 4.7$, ενώ για τους b-tagged πίδακες $p_T > 35$ GeV και $|\eta| < 2.4$. Φυσικά, ο αριθμός των αναμενόμενων γεγονότων που περνούν τα κριτήρια επιλογής προσαρμόζεται, λαμβάνοντας



Σχήμα 2: Διαγράμματα Feynman διαδικασιών σήματος και υποβάθρου: (α) Μονήρες top κουάρκ στο t-κανάλι. (β) Ζεύγος top κουάρκ-αντικουάρκ και (γ) W/Z+jets .

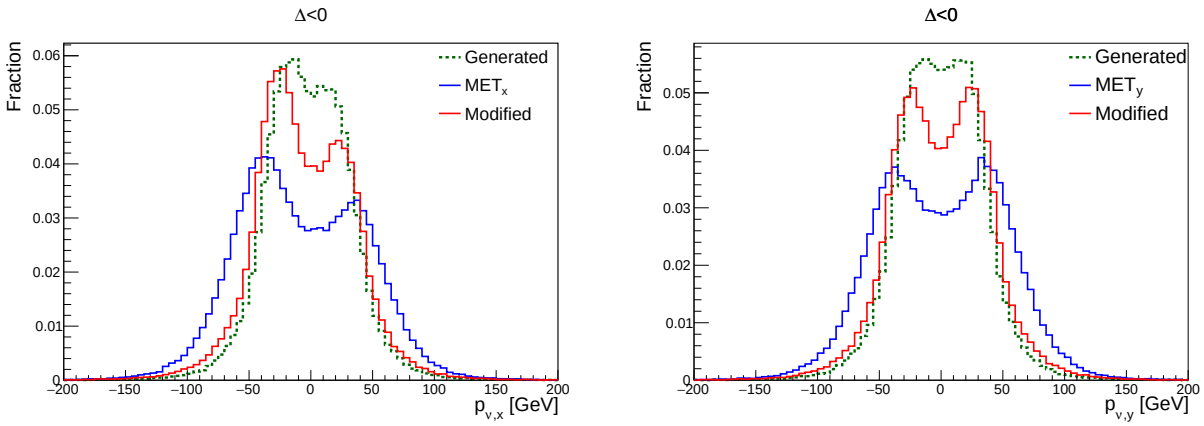
υπόψη την ολοκληρωμένη φωτεινότητα και την ενεργό διατομή κάθε διαδικασίας.

Η τετραορμή του top μπορεί να υπολογιστεί μέσω της ανακατασκευής των προϊόντων διάσπασης από την κινηματική πληροφορία σε κάθε γεγονός. Τα νετρίνα, όμως, δεν αλληλεπιδρούν με τον ανιχνευτή. Ο παραδοσιακός τρόπος με τον οποίο γίνεται η ανακατασκευή του νετρίνου είναι αποδίδοντας την ελλειπούσα εγκάρσια ορμή του γεγονότος εξ ολοκλήρου στην εγκάρσια ορμή του νετρίνου, ενώ η διαμήκης ορμή υπολογίζεται θέτοντας τη μάζα του W σταθερή και ίση με 80.4 GeV. Έπειτα, υποθέτοντας διατήρηση της ενέργειας-ορμής στο βρόχο $W \rightarrow l\nu$, προκύπτει ένα τριώνυμο ως προς το $p_{\nu,z}$:

$$p^{\pm}_{\nu,z} = \frac{\Lambda\, p_{l,z}}{p^2_{l,T}} \pm \sqrt{\frac{\Lambda^2\, p^2_{l,z}}{p^4_{l,T}} - \frac{E^2_l p^2_{\nu,T} - \Lambda^2}{p^2_{l,T}}} \tag{1}$$

όπου $\Lambda = m^2_W/2 + p_{l,T}\, p_{\nu,T}\, cos\Delta\phi$, με $\Delta\phi$ την αζιμουθιακή γωνιακή διαφορά του μιονίου και του νετρίνου. Η υπόρριζη ποσότητα Δ αποτελεί τη διακρίνουσα του τριωνύμου. Εάν η

διακρίνουσα είναι θετική, υπάρχουν δύο πραγματικές λύσεις για το $p_{\nu,z}$ και προτιμάται συνήθως αυτή με τη μικρότερη απόλυτη τιμή. Εάν η διακρίνουσα είναι αρνητική, οι λύσεις του $p_{\nu,z}$ είναι μιγαδικές. Σε αυτήν την περίπτωση τροποποιείται η $p_{\nu,T}$ ώστε να μηδενιστεί η διακρίνουσα, με τρόπο τέτοιο ώστε η μεταβολή του μέτρου της ορμής να είναι η ελάχιστη δυνατή. Με αυτόν τον τρόπο επιτυγχάνεται μια βελτιωμένη ανακατασκευή της εγκάρσιας ορμής του νετρίνου.

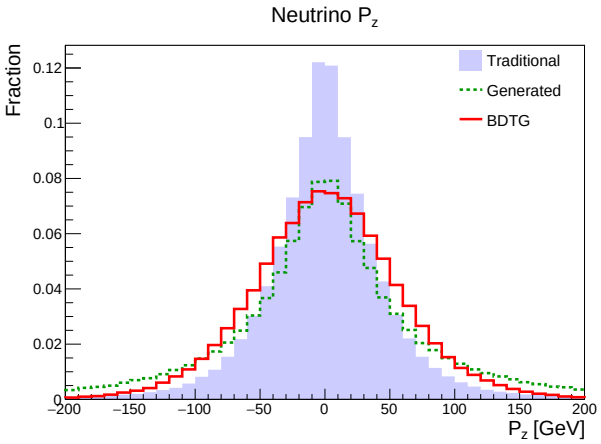


(α) (β)

Σχήμα 3: Σύγκριση των συνιστωσών της εγκάρσιας ορμής του νετρίνου, (α) $p_{\nu,x}$ και (β) $p_{\nu,y}$, μεταξύ των παραγόμενων και ανακατασκευασμένων τιμών, για την περίπτωση αρνητικής διακρίνουσας.

Οι αναλυτικές μέθοδοι παρέχουν λογικές εκτιμήσεις για την ορμή του νετρίνου, αλλά εισάγουν ασυνέπειες λόγω του σταθερού περιορισμού στη μάζα του W μποζονίου. Έτσι, υπάρχει περιθώριο βελτίωσης. Στην παρούσα ανάλυση, προτείνεται η μέθοδος Regressed Kinematic Fit (KinFitReg) για την ενίσχυση της ανακατασκευής του top κουάρκ. Πιο συγκεκριμένα, η διαμήκης συνιστώσα της ορμής του νετρίνου, $p_{\nu,z}$, εκτιμάται αρχικά με χρήση μηχανικής μάθησης και συγκεκριμένα με Ενδυναμωμένα Δέντρα Αποφάσεων (BDT) για παλινδρόμηση. Οι μεταβλητές εισόδου στα BDT περιλαμβάνουν τις κινηματικές ιδιότητες του b-jet, του light πίδακα, του μιονίου και των συνδυασμών τους και επιλέγονται βάσει της ισχυρής τους συσχέτισης με την πραγματική τιμή $p_{\nu,z}^{\text{true}}$ που παράγεται από τις προσομοιώσεις. Συνολικά χρησιμοποιούνται 22 μεταβλητές. Ο μεγάλος αριθμός μεταβλητών δεν επηρεάζει την απόδοση, καθώς τα BDT είναι ανθεκτικά στην παρουσία εισόδων με χαμηλή διαχωριστική ικανότητα. Έτσι, προσφέρεται μία εκτίμηση για το $p_{\nu,z}$ η οποία περιγράφει με μεγαλύτερη ακρίβεια την κατανομή των παραγόμενων τιμών.

Το επόμενο στάδιο περιλαμβάνει την ταυτόχρονη εκτίμηση όλων των συνιστωσών της ορμής του νετρίνου μέσω κινηματικής προσαρμογής. Οι αρχικές εκτιμήσεις για την εγκάρσια ορμή έχουν ήδη προκύψει μέσω της ελλειπούσας εγκάρσιας ορμής p_T^{miss}, ενώ για τη διαμήκη ορμή μέσω της παλινδρόμησης. Αυτές οι εκτιμήσεις εισάγονται στη διαδικασία ελαχιστοποίησης χ^2,



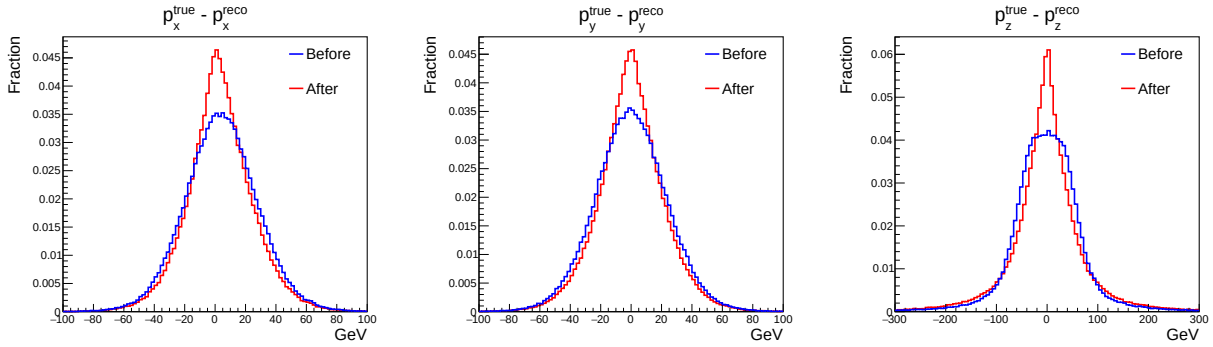
Σχήμα 4: Σύγκριση της παραδοσιακής μεθόδου με την παλινδρόμηση για τη διαμήκη ορμή του νετρίνου $p_{z,\nu}$.

το οποίο εφαρμόζει τον περιορισμό μάζας του W μποζονίου:

$$\chi^2(\tilde{p}_x^{(\nu)}, \tilde{p}_y^{(\nu)}, \tilde{p}_z^{(\nu)}) = \left(\frac{m_w^{reco} - m_w^{true}}{\Gamma w}\right)^2 + \sum_{i\in\{x,y,z\}} \left(\frac{\tilde{p}_i^{(\nu)} - p_i^{(\nu)}}{\sigma_i}\right)^2 \tag{2}$$

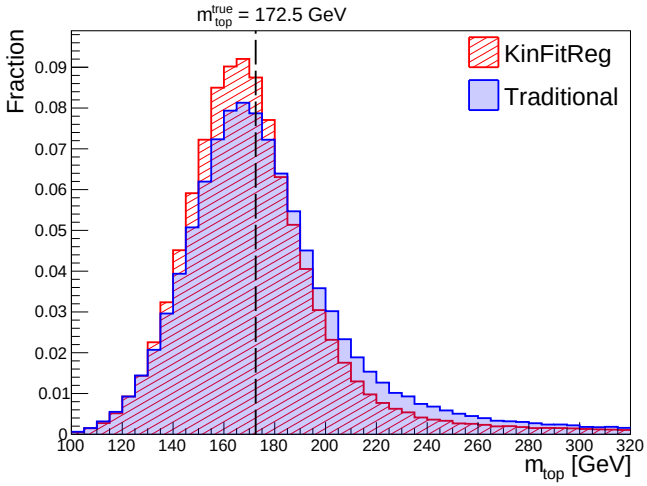
Οι τιμές των σ_i και Γ_w υποδεικνύουν την επιτρεπτή απόκλιση από τις μετρημένες τιμές. Αυτή η προσέγγιση μειώνει τις διαφορές της ανακατασκευασμένης ορμής με την πραγματική τιμή, οδηγώντας σε πιο συνεπείς προβλέψεις για κάθε συνιστώσα της ορμής μετά την ελαχιστοποίηση του χ^2. Συνολικά, η κατάλληλη προσαρμογή σε συνδυασμό με την παλινδρόμηση βελτιώνει σημαντικά την ευκρίνεια, οδηγώντας σε υψηλότερη κορυφή και μειώνοντας τις υψηλές ουρές που προκαλούνται από λανθασμένη ανακατασκευή του top κουάρκ, αν και παραμένει μια μεροληψία (bias) προς χαμηλότερες ενέργειες. Η υπεροχή της μεθόδου KinFitReg έγκειται στο γεγονός ότι η λύση για την ορμή του νετρίνου δεν επιβάλλεται αλλά προκύπτει από αρχικές εκτιμήσεις, παρέχοντας έτσι ένα πιο φυσικό σχήμα για τη μάζα του W μποζονίου σε σύγκριση με τις παραδοσιακές τεχνικές.

Η διάσπαση του top κουάρκ παράγει ένα W μποσόνιο και ένα b κουάρκ, με το καθένα να λαμβάνει περίπου τη μισή αρχική ενέργεια του top κουάρκ. Η ενεργειακή βαθμονόμηση των πιδάκων στο CMS γίνεται κατά μέσο όρο με βάση το p_T, χωρίς να λαμβάνονται υπόψη οι διαφορές γεύσης, με τους b-tagged πίδακες να χάνουν ενέργεια σε νετρίνα από λεπτονικές διασπάσεις B μεσονίων, οδηγώντας σε συστηματική υποεκτίμηση έως 10%. Για την ανάκτηση της χαμένης ενέργειας, εφαρμόζεται ξεχωριστή παλινδρόμηση με BDT στο p_T των b-jets σε επίπεδο parton, βελτιώνοντας την εκτίμηση της εγκάρσιας ορμής του αρχικού bottom κουάρκ. Μετά την παλινδρόμηση, η απόκριση ακολουθεί κατανομή Gauss με μέση τιμή 1.1 και τυπική απόκλιση 0.12, διορθώνοντας την αρχική υποεκτίμηση. Οι διορθωμένες τιμές του p_T επηρεάζουν την τελική κατανομή μάζας του top κουάρκ, με την παλινδρόμηση να συμβάλλει στην αναίρεση της υποβόσκουσας μεροληψίας, διατηρώντας όμως τη βελτίωση στην ευκρίνεια.



(α) (β) (γ)

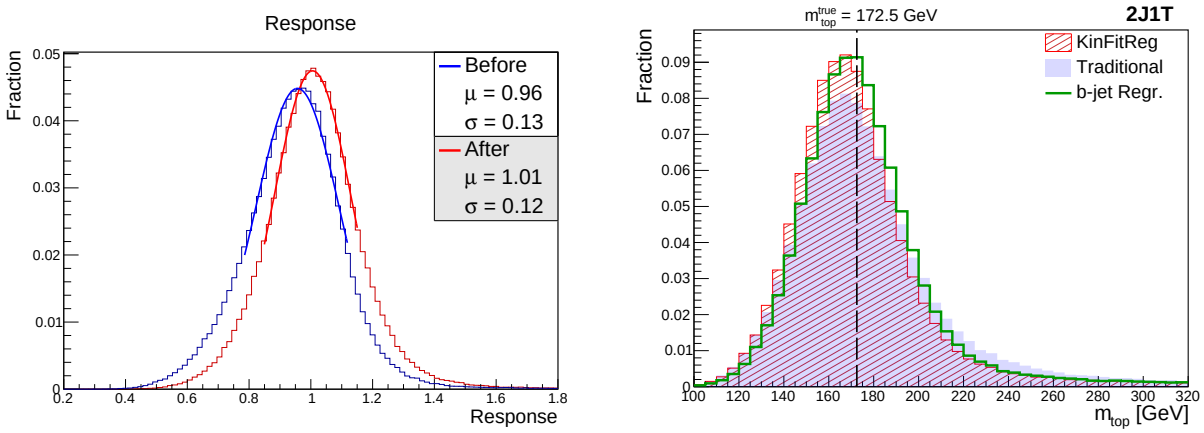
Σχήμα 5: (α)-(γ) Σύγκριση των διαφορών μεταξύ των παραγόμενων και ανακατασκευασμένων τιμών κάθε συνιστώσας της ορμής του νετρίνου, πριν και μετά την ελαχιστοποίηση του χ^2.



Σχήμα 6: Κατανομές της μάζας του top κουάρκ χρησιμοποιώντας τη παραδοσιακή μέθοδο και τη μέθοδο KinFitReg.

Όπως αναφέρθηκε, παραγωγή του top κουάρκ σε γεγονότα μονήρων top στο t-κανάλι δεν είναι ο κυρίαρχος τρόπος παραγωγής top κουάρκ. Για την καλύτερη διάκριση μεταξύ σήματος και υποβάθρου, χρησιμοποιούνται BDT ως διαχωριστές πολλαπλών μεταβλητών. Ως διαδικασία σήματος λαμβάνεται υπόψη και η παραγωγή top αντικουάρκ, ενώ ως υπόβαθρο θεωρούνται μόνο οι διεργασίες παραγωγής ζευγών $t\bar{t}$ και τα γεγονότα από W+jets. Οι υπόλοιπες διεργασίες υποβάθρου θεωρείται ότι καταστέλλονται αποτελεσματικά από τα αρχικά κριτήρια επιλογής. Δύο χαρακτηριστικές μεταβλητές με υψηλή διαχωριστική ικανότητα για το σήμα αποτελούν:

1. Η απόλυτη ψευδοωκύτητα του light πίδακα ($|\eta_j|$), καθώς η τροχιά του κουάρκ που έρχεται κατευθείαν από το πρωτόνιο μεταβάλλεται ελαφρώς, παράγοντας με αυτό τον τρόπο πίδακες με μεγάλες ψευδοωκύτητες.



(α) (β)

Σχήμα 7: (α) Η απόκριση της ανακατασκευασμένης p_T του b-jet πριν και μετά την παλινδρόμηση και (β) η κατανομή της μάζας του top κουάρκ χρησιμοποιώντας την παραδοσιακή μέθοδο και τη μέθοδο KinFitReg, με και χωρίς εφαρμογή παλινδρόμησης.

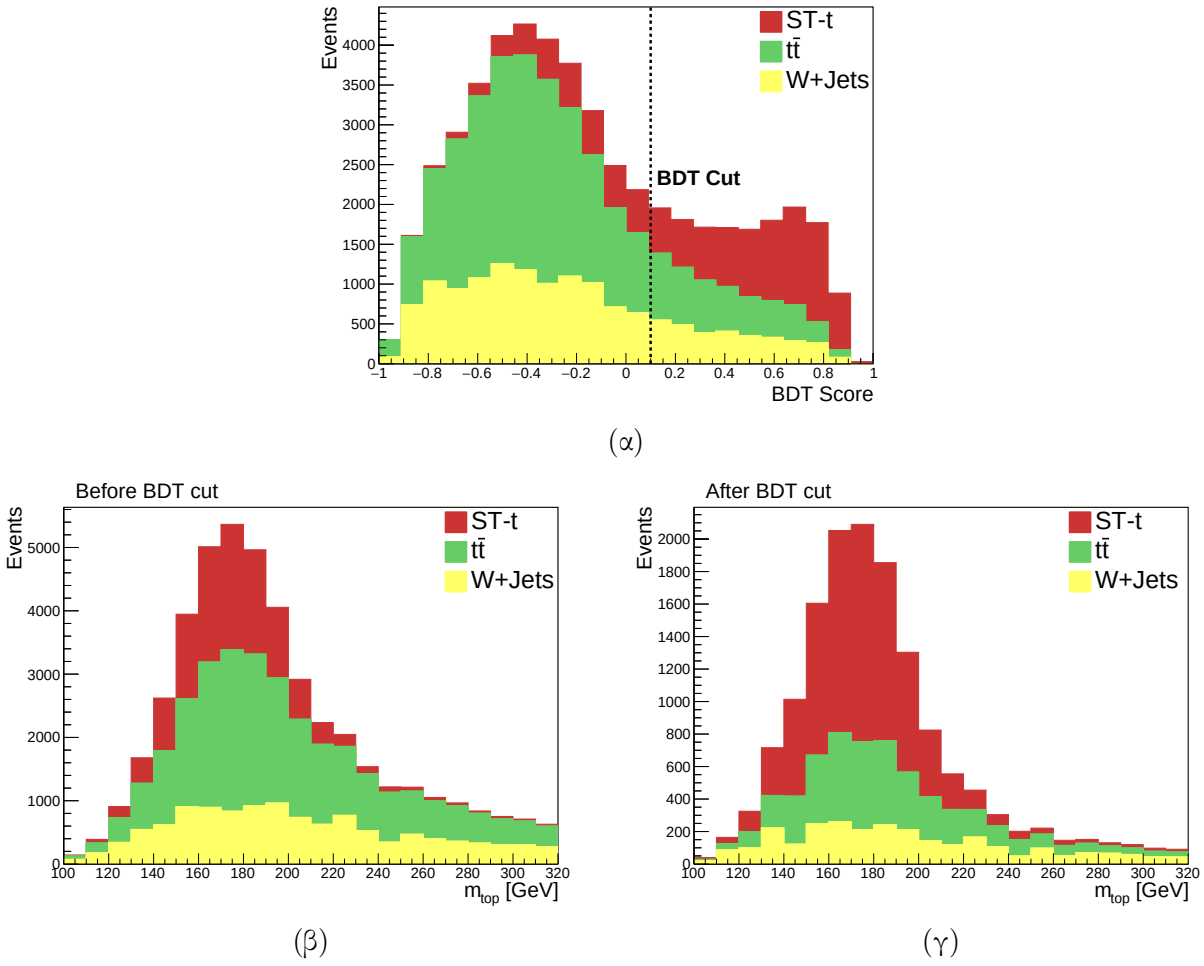
2. Το συνημίτονο της γωνίας μεταξύ του light πίδακα και του λεπτονίου στο σύστημα ηρεμίας του top κουάρκ ($\cos\theta^*$). Με αυτόν τον τρόπο αξιοποιείται η υψηλή πόλωση των top κουάρκ η οποία επιβάλλει συγκεκριμένα σπιν στα σωματίδια τελικής κατάστασης, επηρεάζοντας τις γωνιακές τους κατανομές.

Στην επιλογή μεταβλητών εισόδου αποφεύγονται μεταβλητές με υψηλή συσχέτιση με τη μάζα του top, καθώς σε διαφορετική περίπτωση, η εφαρμογή κατωφλίου για το υπόβαθρο θα οδηγήσει σε μορφοποίηση της κατανομής της ανακατασκετασμένης μάζας, γεγονός που καθιστά τη διάκριση σήματος και υποβάθρου δύσκολη στο στάδιο της προσαρμογής. Στη συνέχεια, ο διαχωριστής αποδίδει ένα BDT σκορ σε κάθε γεγονός, με τα γεγονότα σήματος να λαμβάνουν υψηλότερη βαθμολογία σε σύγκριση με τα γεγονότα υποβάθρου. Με την εφαρμογή ενός κατωφλίου στο BDT σκορ, ο διαχωριστής είναι σε θέση να καταστείλει τις διεργασίες υποβάθρου αποτελεσματικά, διατηρώντας το σχήμα τους και, συνολικά, να ενισχύει σημαντικά την καθαρότητα του σήματος.

Προκειμένου να επαληθευτεί ότι είναι δυνατό να εξαχθούν ουσιαστικά αποτελέσματα από τον αλγόριθμο που αναπτύχθηκε, είναι απαραίτητη η διεξαγωγή μελέτης Toy Monte Carlo. Η εκτίμηση των συνεισφορών των διαφορετικών διεργασιών σε ένα σύνολο δεδομένων D επιτυγχάνεται με την ανάλυση των διαφορών στις κατανομές μάζας (templates), οι οποίες προκύπτουν από προσομοιώσεις MC. Οι αναμενόμενοι αριθμοί γεγονότων για κάθε template θεωρούνται ως παράμετροι του μοντέλου και προσαρμόζονται για τη βέλτιστη συμφωνία με τα δεδομένα. Για μια παρατηρούμενη μάζα top x, το μοντέλο προσαρμογής δίνεται ως:

$$D(x) = N_{st}F_{st}(x; m_{top}^{true}) + N_{t\bar{t}}F_{t\bar{t}}(x; m_{top}^{true}) + N_{wj}F_{wj}(x) \quad (3)$$

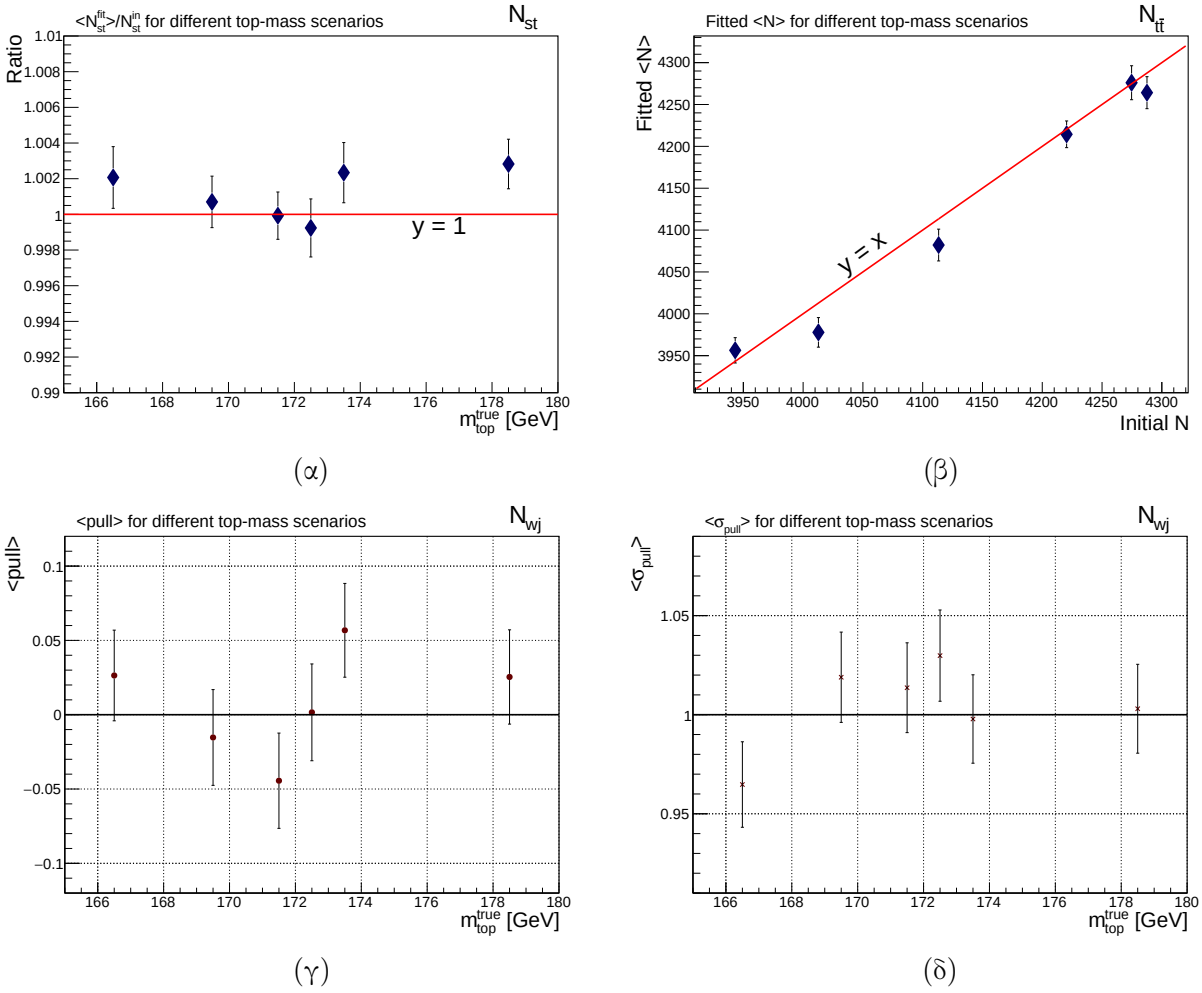
όπου οι F_{st}, $F_{t\bar{t}}$ και F_{wj} είναι οι συναρτήσεις πυκνότητας πιθανότητας των αντίστοιχων διεργασιών. Για την επικύρωση του μοντέλου, δημιουργούνται ψευδοδεδομένα και εκτελείται προσαρμογή



Σχήμα 8: (α) Η απόκριση του BDT σε stack ιστόγραμμα και οι κατανομές μάζας του top κουάρκ (β) πριν και (γ) μετά το BDT cut.

ML σε κάθε σετ για έλεγχο σταθερότητας των παραμέτρων. Προκειμένου να ληφθεί υπόψη η εξάρτηση των προσομοιωμένων δειγμάτων από τη μάζα του top κουάρκ, χρησιμοποιήθηκαν πολλαπλά δείγματα, το καθένα υποθέτοντας διαφορετική μάζα top. Για τον έλεγχο γραμμικότητας, ο λόγος της μέσης εκτιμηθείσας παραμέτρου N_{st} ως προς την αρχική τιμή οφείλει να είναι κοντά στη μονάδα, οι αναμενόμενες τιμές της εκτιμηθείσας παραμέτρου $N_{t\bar{t}}$ οφείλουν να έχουν γραμμική εξάρτηση ως προς τις αρχικές με κλίση μονάδα και μηδενική μετατόπιση, ενώ, τέλος, η κατανομή του μέσου pull για την παράμετρο N_{wj} οφείλει να είναι κοντά στο μηδέν εντός των σφαλμάτων με την τυπική απόκλιση κοντά στη μονάδα. Συνολικά, αυτή η εκτενής μελέτη διασφαλίζει την αξιοπιστία των μοντέλων προσαρμογής και την ακρίβεια στην εκτίμηση των παραμέτρων σε διαφορετικά σενάρια μάζας του top κουάρκ.

Έχοντας διαθέσιμα έξι μοντέλα προσαρμογής, καθίσταται δυνατή η εκτίμηση της μάζας του top κουάρκ μέσω σάρωσης πιθανοφάνειας. Πραγματοποιείται, δηλαδή, προσαρμογή των



Σχήμα 9: Έλεγχος γραμμικότητας: (α) ο λόγος της fitted παραμέτρου N_{st} προς την αρχική τιμή, (β) η fitted $N_{t\bar{t}}$ ως συνάρτηση των αρχικών τιμών και (γ) η μέση τιμή του pull και (δ) της τυπικής απόκλισης της fitted N_{wj} ως συνάρτηση της m_{top} για διαφορετικά σενάρια μάζας top κουάρκ.

δεδομένων σε κάθε μοντέλο, και έπειτα αναζητάται το μοντέλο στο οποίο η (αρνητική) πιθανοφάνεια, συγκεκριμένα η negative log-likelihood, είχε τη μικρότερη τιμή. Στη συνέχεια, γίνεται προσαρμογή πολυωνύμου τρίτου βαθμού στη συνάρτηση πιθανοφάνειας και υπολογίζεται μέσω παραβολικής προσέγγισης το σημείο ελαχίστου. Η εκτιμώμενη τιμή της μάζας αντιστοιχεί στο ελάχιστο της παραβολής, ενώ το στατιστικό σφάλμα προσδιορίζεται από το πλάτος. Αξίζει να σημειωθεί ότι, επειδή τα δείγματα μάζας καλύπτουν ευρύ φάσμα, τα templates μάζας παρεμβάλλονται με βήμα 0.25 GeV ούτως ώστε να καλύπτεται επαρκώς η περιοχή γύρω από το ελάχιστο, υποθέτοντας μια γραμμική σχέση ανάμεσα σε δύο γειτονικά δείγματα,

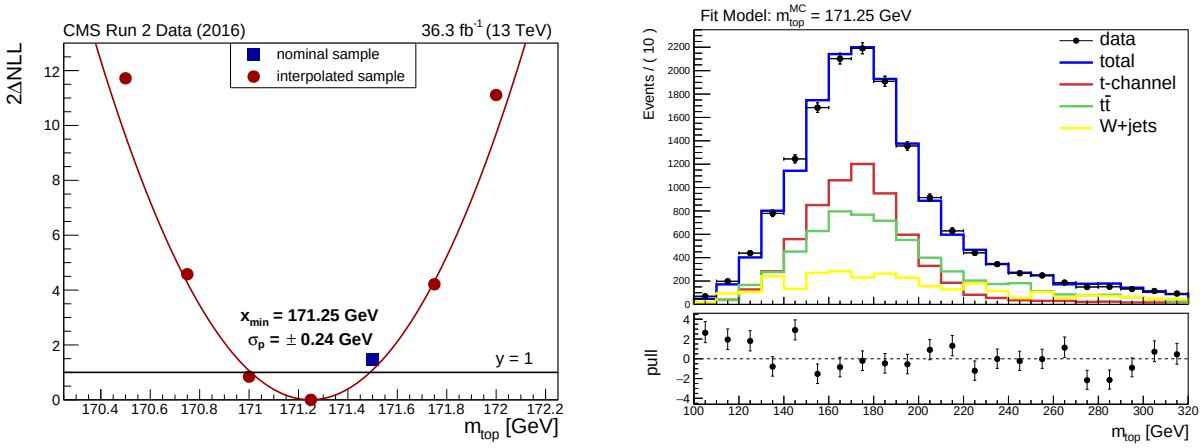
Στη συνέχεια, η επιλογή του κατωφλίου του BDT καθορίζεται με βάση την επίδρασή του στο σφάλμα της μέτρησης. Συγκεκριμένα, το BDT cut επιλέγεται έτσι ώστε να οδηγεί σε πιο

στενό παραβολικό σχήμα στην κατανομή της πιθανοφάνειας, δηλαδή με μικρότερο πλάτος. Παράγοντας πολλά ψευδοδεδομένα από ένα δείγμα μάζας MC υπολογίζεται το μέσο σφάλμα μέσω σάρωση πιθανοφάνειας. Επαναλαμβάνοντας τη διαδικασία για διαφορετικά BDT cuts εντοπίζεται το βέλτιστο κατώφλι το οποίο οδηγεί στο ελάχιστο μέσο σφάλμα. Για μέγιστη ακρίβεια, επομένως, χρησιμοποιήθηκε κατώφλι BDT σε σκορ 0.1, με μέσο σφάλμα $<\sigma_p>$=0.113.

Έχοντας επαληθεύσει ότι ο αλγόριθμος που αναπτύχθηκε λειτουργεί σωστά με τα δείγματα MC, εφαρμόζεται στα πραγματικά δεδομένα για τη μέτρηση. Στην ανάλυση αυτή χρησιμοποιούνται δεδομένα του CMS από το Run 2 του 2016 με ενέργεια κέντρου μάζας 13 TeV και συνολική φωτεινότητα 36.3 fb^{-1}. Η τιμή της μάζας του top, m_{top}, χρησιμοποιώντας γεγονότα παραγωγής μονήρων top κουάρκ στο t-κανάλι, μετρήθηκε και βρέθηκε να είναι:

$$m_{top}^{t-ch} \;=\; 171.25 \,\pm\, 0.24 \text{ (stat+prof) GeV} \tag{4}$$

Το αποτέλεσμα αυτό είναι συμβατό με τις προηγούμενες μετρήσεις από ATLAS και CMS εντός των αβεβαιοτήτων. Είναι σημαντικό, ωστόσο, να σημειωθεί ότι οι στατιστικές αβεβαιότητες, σε σύγκριση με την αντίστοιχη μέτρηση του CMS που αναφέρθηκε προηγουμένως, μειώνονται κατά 25%. Αυτό αποτελεί σημαντική πρόοδο και επιτυγχάνει τον αρχικό στόχο, ο οποίος είναι είναι η ελαχιστοποίηση των αβεβαιοτήτων. Άρα, λοιπόν, η βελτίωση της ευκρίνειας στην κατανομή της ανακατασκευασμένης μάζας του top κουάρκ μέσω της μεθόδου KinFitReg που εισήχθη σε αυτήν την ανάλυση έχει ως αποτέλεσμα τη μείωση των στατιστικών σφαλμάτων, κάτι που οδηγεί εν γένει και στη μείωση των συστηματικών αβεβαιοτήτων.



(α) (β)

Σχήμα 10: (α) Εκτίμηση της μάζας του top κουάρκ μέσω σάρωσης πιθανοφάνειας. Η παραβολική προσέγγιση δείχνει ένα ελάχιστο στα 171.25 GeV με συμμετρική αβεβαιότητα ±0.24 GeV και (β) προσαρμογή των δεδομένων στο μοντέλο με υποτιθέμενη μάζα top κουάρκ 171.25 GeV.

Με μια οπτική επιθεώρηση των προσαρμοσμένων δεδομένων στο μοντέλο με μάζα top 171.25 GeV, παρατηρείται ότι ενώ το μοντέλο περιγράφει με ακρίβεια τα δεδομένα στην περιοχή της κορυφής, στην περιοχή χαμηλότερων ενεργειών οι κατανομές pull είναι μεγαλύτερες από το 0,

αποκλίνοντας μέχρι και 3σ. Ο συνολικός αριθμός των αναμενόμενων γεγονότων είναι σημαντικά μικρότερος από τον παρατηρούμενο αριθμό στα δεδομένα, υποδεικνύοντας ότι σε χαμηλότερες ενέργειες μπορεί να υπάρχουν διεργασίες, όπως το υπόβαθρο QCD, που αγνοήθηκαν στις μελέτες Toy Monte Carlo. Αυτές οι συνεισφορές υποβάθρου μπορούν να εξηγήσουν την αύξηση του $t\bar{t}$, καθώς η διαδικασία προσαρμογής αποδίδει λανθασμένα τα επιπλέον γεγονότα στην παραγωγή ζεύγους top κουάρκ-αντικουάρκ. Πράγματι, παρατηρούνται σημαντικές αποκλίσεις στις τιμές pull για το υπόβαθρο $t\bar{t}$ με τιμή 4.06 και για το σήμα μονήρων top κουάρκ με τιμή -2.35. Συνεπώς, απαιτείται περαιτέρω έρευνα για τα υπόλοιπα υποβάθρα, προκειμένου να προσδιοριστεί η επίδρασή τους στο τελικό αποτέλεσμα.

	Αρχικά	Προσαρμογή	Σφάλμα	Pull
N_{st}	7342	6522	350	-2.35
$N_{t\bar{t}}$	3978	6142	533	4.06
N_{wj}	2884	2981	270	0.36

Πίνακας 1: Αρχικές και προσαρμοσμένες τιμές των ελεύθερων παραμέτρων, τα σφάλματά τους και οι τιμές των pulls.

Εκ των πραγμάτων, στα πλαίσια μίας διπλωματικής εργασίας επιβλήθηκαν αρκετοί περιορισμοί και, επομένως, η ανάλυση αυτή δεν αποτελεί μια ολοκληρωμένη μέτρηση. Μια πιο λεπτομερής ανάλυση θα περιλάμβανε: τη συμπερίληψη καναλιών διάσπασης σε τελική κατάσταση με ηλεκτρόνια, την εξέταση επιπλέον πηγών υποβάθρου όπως τα υπόλοιπα κανάλια μονήρων top tW και s, την QCD, τα Z+Jets και τα VV, τη χρήση περισσότερων δειγμάτων μάζας, τη βελτιστοποίηση της απόκρισης των BDT και την εκτίμηση των συστηματικών αβεβαιοτήτων. Ωστόσο, παρά τους περιορισμούς, η φαινόμενη μείωση των στατιστικών αβεβαιοτήτων που αποδίδεται στη βελτίωση της ανακατασκευής της μάζας του top κουάρκ υποδεικνύει ότι είναι εφικτή μια ακριβέστερη μέτρηση η οποία θα οδηγήσει στη μείωση σφαλμάτων συνολικά. Συνεπώς, αυτό το αποτέλεσμα αναδεικνύει μια πολλά υποσχόμενη μέθοδο για ουσιαστική μείωση των στατιστικών και συστηματικών αβεβαιοτήτων σε μελλοντικές μετρήσεις της μάζας του top κουάρκ.

Contents

List of Figures

Chapter 1

Introduction

In the Standard Model (SM) the top quark is a fundamental fermion with spin 1/2. It carries an electric charge of $(+2/3)e$ and is a color triplet. It also forms a weak isospin doublet together with the bottom quark, where the top quark is the up-type quark with the third component of the weak isospin $I_3 = +1/2$. It was discovered by the CDF and D0 collaborations at the Tevatron in 1995 and is the most massive elementary particle currently known. The uniqueness of the top quark lies in its decay time; it is considerably shorter than the time of hadronization, hence resulting in a shorter lifetime than the time needed for strong interactions to modify its properties or to bind it into a hadron. Therefore, the top quark provides a unique laboratory to test our understanding of matter and fundamental interactions at the Electroweak symmetry-breaking scale and beyond.

1.1 The Standard Model of Particle Physics

The SM of particle physics is a quantum field theory that describes the fundamental set of particles - leptons, quarks, gauge bosons and Higgs boson - and their interactions [1, 2]. Although there remain open questions that cannot be answered by it yet, the SM constitutes perhaps the most well-established theoretical construct in physics, enjoying a multitude of observational confirmations throughout the last century.

The elementary particles of spin 1/2 obeying the Fermi-Dirac statistics and the Pauli exclusion principle are called fermions. These particles include all quarks and leptons and are regarded as the fundamental building blocks of matter. The SM contains twelve fermions in total, six leptons and six quarks distributed across three generations. As shown in Table 1.1, each generation consists of an up-type quark (u,c,t), a down-type quark (d,s,b), a charged lepton (e,μ,τ) and the corresponding neutrino (ν_e,ν_μ,ν_τ). Charged leptons carry an electric charge of one elementary charge $-1e$, whereas quarks carry a fractional electric charge of either $(+2/3)e$ or $(-1/3)e$. For each of those fermions, there exists an antiparticle with identical physical properties except for opposite charges.

All regular matter in the universe is exclusively composed from particles of the first generation. Nevertheless, despite the decaying of higher-generation particles into the first

Flavor	Fermion	Symbol	El. charge $[e]$	Mass $[MeV/c^2]$
up	quark	u	$+\frac{2}{3}$	$2.16^{+0.49}_{-0.26}$
down	quark	d	$-\frac{1}{3}$	$4.67^{+0.48}_{-0.17}$
electron	lepton	e	-1	0.511
e-neutrino	lepton	ν_e	0	$< 0.8 \cdot 10^6$
charm	quark	c	$+\frac{2}{3}$	$(1.27 \pm 0.2) \cdot 10^3$
strange	quark	s	$-\frac{1}{3}$	$94.4^{+8.6}_{-3.4}$
muon	lepton	μ	-1	105.658
μ-neutrino	lepton	ν_μ	0	< 0.19
top	quark	t	$+\frac{2}{3}$	$(172.69 \pm 0.3) \cdot 10^3$
bottom	quark	b	$-\frac{1}{3}$	$(4.18^{+0.03}_{-0.02}) \cdot 10^3$
tau	lepton	τ	-1	1776.86 ± 0.12
τ-neutrino	lepton	ν_τ	0	< 18.2

Table 1.1: The three generations of fermions, their electric charges and masses [3].

generation, these massive fermions are studied through high energy interactions, such as collider experiments. Additionally, quarks are never observed as free particles; they are always confined in bound states, known as hadrons. These hadronic states are either quark-antiquark pairs (mesons) or three quarks (baryons) or three antiquarks (antibaryons). The formation of hadronic bound states is realised in accordance with Pauli exclusion principle through the introduction of an additional quantum number for quarks; the color. A quark's color can take one of three charges: red, blue, green (or antired, antiblue and antigreen for an antiquark respectively). The mixing of all three (anti)colors, or any color with its complement, is considered colorless and has a net color charge of zero. Due to color confinement, free particles must have color charge of zero, thus directly implying that free quarks cannot exist.

Besides fermions, the SM also provides the theoretical framework for the interactions between particles by describing three of the four fundamental interactions: the electromagnetic force, the strong and weak nuclear forces. Each of these forces originates from a symmetry by a mechanism called local gauge invariance, which requires that the Lagrangian of a physical system remains invariant under a certain group of local transformations. In particular, the unified electroweak sector of the SM is based on the $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry group and is mediated by the so-called gauge bosons: the massless photon (γ) and the $W^\pm$, Z bosons, whereas the strong sector, based on the $\mathrm{SU}(3)_C$ group, is mediated by the massless gluons (g). These force carriers are particles of spin 1 obeying Bose-Einstein statistics, which allows them to occupy the same quantum state, contrary to fermions.

The final element of the SM is the Higgs boson, which differs from all other particles. Unlike the fundamental fermions and gauge bosons, the Higgs boson is a spin 0 scalar particle and the only fundamental scalar discovered to date. In the context of quantum field theory, the Higgs boson is regarded as an excitation of the Higgs field, which has a non-zero expec-

Force	Bosons	Symbol	El. charge [e]	Mass [MeV/c^2]
strong	8 gluons	g	0	0
electromagnetic	photons	γ	0	0
weak	W	$W^{\pm}$	± 1	80.369 ± 0.013
weak	Z	Z^0	0	91.188 ± 0.002

Table 1.2: The gauge bosons, their electric charges and masses [3].

tation vacuum value. Its pivotal role in the SM stems from the fact that the interaction with this non-zero Higgs field is the mechanism through which all other particles acquire their mass.

The remarkable success of the SM lies in its ability to describe a wide range of precise experimental measurements, thus establishing its validity up to the electroweak scale and providing in this way a robust understanding of the constituents of the universe and their interactions. However, its inability to incorporate the gravitational force in the existing quantum-field framework, although it can be neglected on subnuclear scales, strongly indicates that in its current form it cannot account as a theory of everything. Additionally, open questions related to dark matter, the nature of the Higgs boson, the origin of CP violation, among others, underscore the need for exploring physics Beyond the Standard Model.

1.2 Electroweak theory

The single framework describing the electromagnetic and the weak nuclear forces is a gauge theory called electroweak theory and is based on the group $SU(2)_L \times U(1)_Y$. At low energies the two forces are completely different, the weak force mediated by massive $W^{\pm}$ and Z^0 bosons is short range, while electromagnetism, mediated by massless photons, is long range. Nevertheless, there is an energy scale, at the order of $\sim$100 GeV, above which the two forces are unified.

The weak force is described by the $SU(2)_L$ part, while $U(1)_Y$ represents the hypercharge, which combines with the weak force account for electromagnetism. When acting on an isospin doublet, the generators σ^{α} are represented by $\frac{\sigma_i}{2}$, where σ_i are the Pauli matrices. The linear combination of the gauge fields W_1^{μ} and W_2^{μ} is responsible for the charged current interactions corresponding to $W^{\pm}$ bosons:

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}} \left(W_{\mu}^{1} \mp i W_{\mu}^{2} \right) \tag{1.1}$$

Accordingly, the electromagnetic field A_{μ} can be expressed in terms of the remaining fields by performing a rotation:

$$\begin{pmatrix} A_{\mu} \\ Z_{\mu} \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_{\mu} \\ W_{\mu}^{3} \end{pmatrix} \tag{1.2}$$

under the condition $g \sin\theta_W = g' \cos\theta_W = e$, where e is the electric charge of the electron. The parameter θ_W is the electroweak mixing angle and the neutral field Z_μ corresponds to the field of the Z boson.

The left-handed chirality states of the fermionic fields are organized in isospin doublets of $SU(2)$:

$$L_L^j = \begin{pmatrix} \nu_{lL}^j \\ e_{lL}^j \end{pmatrix} \quad \text{and} \quad Q_L^j = \begin{pmatrix} q_{uL}^j \\ q_{dL}^j \end{pmatrix} \tag{1.3}$$

where index $j = 1, 2, 3$ runs over the fermion generations, ν_{lL}^j, e_{lL}^j, q_{uL}^j and q_{dL}^j indicate the Dirac fields of lepton neutrinos and charged leptons, up-type quarks and down-type quarks respectively. The right-handed fields are treated as SU(2) singlets. The Lagrangian density describing the dynamics of the fermionic fields is given by:

$$\mathcal{L} = i\bar{L}_L \gamma^\mu D_\mu L_L + i\bar{\nu}_{lR} \gamma^\mu D_\mu \nu_{lR} + i\bar{l}_R \gamma^\mu D_\mu l_R \tag{1.4}$$

$$D_\mu = \partial_\mu - ig\sigma^\alpha W_\mu^\alpha(x) - ig' \frac{Y_\Psi}{2} B_\mu(x) \tag{1.5}$$

where g, g' are real dimensionless parameters, σ^α are the three generators of $SU(2)$, Y_Ψ is the hypercharge operator, and $W_\mu^i(x)$ and $B_\mu(x)$ are the corresponding gauge fields.

1.3 Quantum Chromodynamics

The theory of strong interactions, known as Quantum Chromodynamics (QCD), is a gauge theory based on the $SU(3)_C$ symmetry group and describes the fundamental force binding quarks together into hadronic particles, such as protons and neutrons. The quark fields are organized in SU(3) triplets $\psi_q = (q_a, q_b, q_c)^T$, which contain three color states for each of the quark flavors. The QCD term of the SM Lagrangian can be written as:

$$\mathcal{L}_{QCD} = -\frac{1}{4} F_{\mu\nu}^a F^{a\,\mu\nu} + \sum_q \bar{\psi}_q \left(i\gamma^\mu D_\mu - m_q \right) \psi_q \tag{1.6}$$

$$\text{where} \quad D_\mu = \partial_\mu - ig_s t_a A_\mu^\alpha(x) \quad \text{and} \quad F_{\mu\nu}^\alpha = \partial_\mu A_\nu^\alpha - \partial_\nu A_\mu^\alpha + g_s f^{abc} A_\mu^b A_\nu^c \tag{1.7}$$

The index q runs over the quark flavors and the symbols γ^μ and t_a denote the Dirac and Gell-Mann matrices respectively. $A_\mu^\alpha(x)$ correspond to the eight massless gauge fields of the SU(3) group and the f^{abc} are the completely antisymmetric SU(3) structure constants defined by the commutation relation $[t_a, t_b] = if^{abc} t_c$. The last term in the second equation of 1.7 shows that gluons, besides mediating the strong interaction, are self-interacting particles.

The coupling constant of QCD, denoted by g_s, determines the strength of the strong interaction. Commonly expressed as $\alpha_s = g_s^2/(4\pi)$ in quantum field theory, the strong coupling constant is dependent on the energy scale Q of a given physical process, as a direct

consequence of the renormalization procedure. Therefore, in one-loop approximation, the dependence is given by:

$$\alpha_s(Q^2) \;=\; \frac{1}{b_0 \ln\left(Q^2/\Lambda^2_{QCD}\right)} \quad \text{where} \quad b_0 \;=\; \frac{33 - 2n_f}{12\pi} \tag{1.8}$$

The number of quark flavors with mass $m_q \ll Q$ is denoted by n_f, and the energy scale at which the QCD coupling constant is expected to become large is denoted by the parameter Λ_{QCD}. This implies that the regime of fixed-order perturbative calculations remains valid up to a certain point. Nevertheless, as observed in equation 1.8, increasing the energy transfer Q leads to a decrease in the value of the strong coupling constant. As a result, in high-energy QCD processes, colored particles interact weakly, allowing for a perturbative approach to QCD. This property, known as asymptotic freedom, is essential for predictions relevant experimental high-energy physics. In contrast, at low-energy scales, the strong interaction gives rise to color confinement, namely the mechanism that prevents quarks and gluons from isolating. In the context of collision experiments, this confinement causes high-energy quarks and gluons to form multiple bound states, producing jets in a process known as hadronization.

1.4 The Higgs Mechanism

The requirement of gauge invariance forbids explicit mass terms for the fermionic fields and for the gauge bosons in the Lagrangian. To account for the generation of the masses of elementary particles, the Higgs mechanism is introduced for spontaneous symmetry breaking. The Higgs field, a complex scalar field that permeates all of space, is a doublet under the electroweak gauge symmetry $SU(2)_L \times U(1)_Y$ that can be written as:

$$\phi \;=\; \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.9}$$

Even though the Lagrangian is invariant under the $SU(2)_L \times U(1)_Y$ symmetry, the vacuum state is not. The Higgs field takes on a non-zero vacuum expectation value (VEV) and breaks the symmetry to U(1), the gauge symmetry of electromagnetism. In this way, W and Z boson acquire mass, while the photons remain massless. The Higgs potential can be written as:

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.10}$$

For $\mu^2 < 0$ the Higgs potential receives the characteristic "hat" shape and leads to a non-zero VEV with a form:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.11}$$

where $v = \sqrt{-\mu^2/\lambda} \approx 246$ GeV is the VEV. Thereby, the electroweak gauge bosons acquire their mass by interacting with the Higgs field. The covariant derivative in the Lagrangian describing those interactions is:

$$\mathcal{L} = |D_\mu \phi|^2 - V(\phi) \tag{1.12}$$

$$\text{where} \quad D_\mu \phi = \left(\partial_\mu - \frac{ig}{2} \sigma^a W_\mu^a - \frac{ig'Y}{2} B_\mu \right) \phi \tag{1.13}$$

and the masses of the W and Z bosons are:

$$M_W = \frac{gv}{2}, \quad M_Z = \frac{\sqrt{g^2 + g'^2} v}{2} \tag{1.14}$$

where g and g' are the gauge couplings for the $\mathrm{SU}(2)_L$ and $\mathrm{U}(1)_Y$ fields respectively. In its vacuum state, the Higgs field can be expressed in terms of a perturbation around its non-zero vacuum expectation value (in unitary gauge):

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.15}$$

where h is the field corresponding to a physical state, named Higgs particle, a scalar chargeless boson with a mass equal to $m_h = v\sqrt{2\lambda}$. The existence of the Higgs boson was experimentally confirmed at the Large Hadron Collider in 2012.

Moreover, Higgs mechanism accounts also for the masses of the SM fermions through the Yukawa interactions, which take the general form of:

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{f}_L \phi f_R + \text{h.c.} \tag{1.16}$$

where y_f is the Yukawa coupling constant that determines the mass of the fermion after the spontaneous symmetry breaking:

$$m_f = \frac{y_f v}{\sqrt{2}} \tag{1.17}$$

However, mass eigenstates do not correspond to flavor eigenstates. The superposition of flavor eigenstates in the basis of mass eigenstates suggests the existence of flavor-changing charged currents mediated by charged $\mathrm{W}^\pm$ bosons. The relative strength of quark transisions to different flavors are encoded in the Cabibbo-Kobayashi-Maskawa (CKM) matrix ($\mathrm{V_{CKM}}$):

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{\text{CKM}} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.18}$$

The CKM matrix describes the probabilities of transitions between different quark flavors, where the transition from a flavor j quark to a flavor i quark is proportional to $|V_{ij}|^2$. In contrast, an analogous flavor mixing for leptons is not observed if neutrinos are considered massless.

1.5 The Top Quark

1.5.1 Decay of Top Quark

Top quarks decay via the weak interaction almost exclusively into a b quark and a W boson, since the decay rates into other down-type quarks, namely into d and s quarks, are suppressed due to the small CKM matrix elements of these decays. The partial decay width for the decay channel $t \to W^+b$, $\Gamma(t \to W^+b)$, is proportional to the CKM matrix element $|V_{tb}|^2$. The branching ratio for the channel $t \to W^+b$ can be calculated in terms of the elements of the CKM matrix:

$$\mathcal{B}(t \to W^+b) \;=\; \frac{|V_{tb}|^2}{|V_{td}|^2 + |V_{ts}|^2 + |V_{tb}|^2} \tag{1.19}$$

where, by using the independently measured CKM elements, its value is approximately $\mathcal{B}(t \to W^+b) \approx 0.95$ [3]. By assuming $|V_{tb}| \gg |V_{td}|, |V_{ts}|$ and by neglecting terms of order m_b^2/m_t^2, α_s^2 and $(\alpha_s/\pi)M_W^2/m_t^2$, the total decay width is perturbatively calculable and the prediction within the SM at NLO is:

$$\Gamma_t \;=\; \frac{G_F m_t^3}{8\pi\sqrt{2}} \left(1 - \frac{M_W^2}{m_t^2}\right) \left(1 + 2\frac{M_W^2}{m_t^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right] \tag{1.20}$$

where m_t is the top quark pole mass, α_s the strong coupling constant and G_F the Fermi constant. At the reference value of $m_t = 172.5$ GeV, the total decay width is $\Gamma_t = 1.326$ GeV and by inverting it, the top quark mean lifetime is $\tau_t = 5 \times 10^{-24}$ s. This lifetime is considerably shorter than the typical time scale of hadronization which can be estimated from the inverse of the energy scale Λ_{QCD} at which QCD becomes non-perturbative: $1/\Lambda_{QCD} \approx 1/(200\text{MeV}) \approx 3\times10^{-24}$ s. As a result, top quarks are expected to decay before hadronization and do not form top-flavored hadrons or $t\bar{t}$-quarkonium-bound states. Additionally, in this way, the top quark spin polarization and the correlation between spins are largely preserved as well. Consequently, any polarization of the top-quark or any spin correlations in top quark pair production are reflected in angular correlations of the decay products.

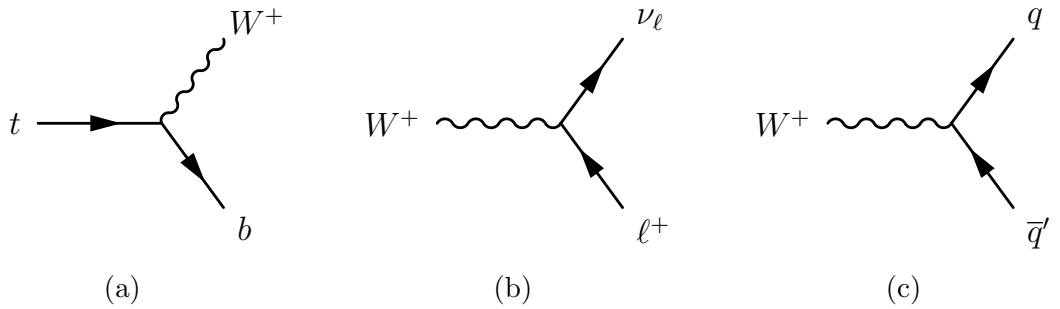


Figure 1.1: (a) The dominant decay process of the top quark into a W^+ boson and a b quark. The W^+ boson, in its turn, decays either (b) leptonically or (c) hadronically.

The dominant decay process of the top quark, shown in Figure 1.1, consists of a W^+ and a b quark. While the b quark hadronizes, the W^+ boson can decay either leptonically, namely into a positively charged lepton and the corresponding neutrino, or hadronically into a quark-antiquark pair. Although the universality of the weak interaction suggests equal probabilities, the hadronic decays of W boson are favored by a factor of 3 due to the three different color charges. As a consequence, each of the three leptonic decay modes has a branching fraction of 1/9, whereas each one of the hadronic 3/9:

$$BF\left(W^+ \to \ell^+\nu_\ell\right) \;=\; \frac{1}{9}, \quad \ell^+ \;=\; \left\{e^+, \mu^+, \tau^+\right\} \tag{1.21}$$

$$BF\left(W^+ \to q\overline{q}'\right) \;=\; \frac{3}{9}, \quad q\overline{q}' \;=\; \left\{u\overline{d}, c\overline{s}\right\} \tag{1.22}$$

It is worth noting that, in general, the leptonic channel, although includes a neutrino which escapes experimental setups, is highly favored for reconstruction because the imprint of charged high-energy leptons in detectors are much clearer than that of two light-quark jets from the hadronic channel, which mix with the multi-jet background of QCD.

1.5.2 Production of Top Quarks

Top Quark-Antiquark Pair

At the LHC the dominant production mode for the top quark is the top pair production mechanism ($t\bar{t}$) via the strong interaction. More specifically, at center-of-mass energy of $\sqrt{s}$=13 TeV in pp collisions, $t\bar{t}$ pairs are predominantly produced via gluon-gluon fusion (gg) approximately 90% of the time, with a subdominant contribution from quark-antiquark annihilation ($q\bar{q}$) at 10% [3]. Figure 1.2 below depicts the Feynman diagrams for different top quark pair production modes at LO in QCD.

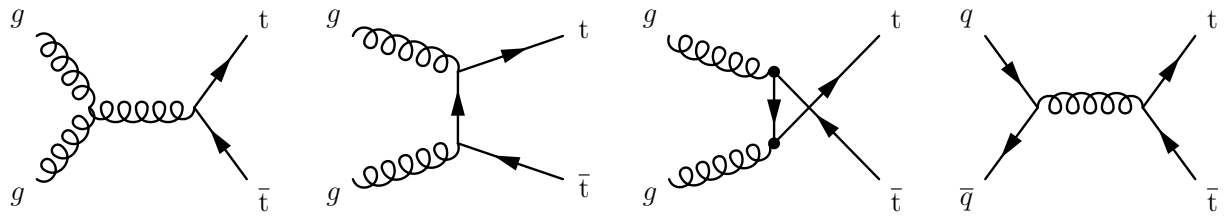


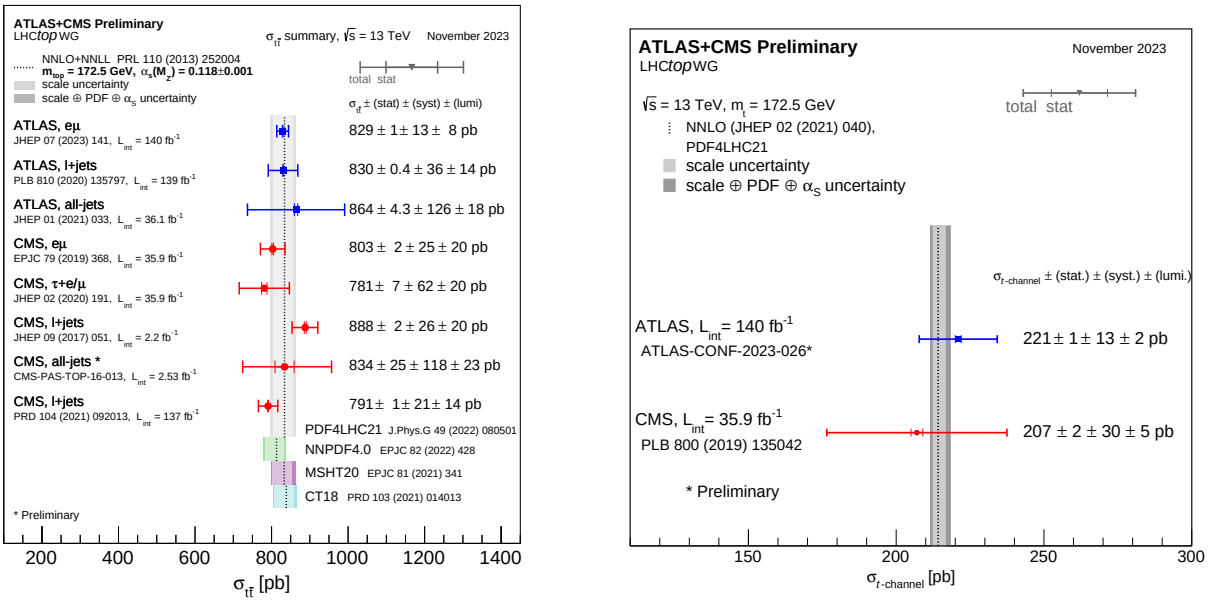
Figure 1.2: Feynman diagrams for $t\bar{t}$ production in QCD at LO: (a)-(c) gg fusion in s, t and u-channel respectively and (d) $q\bar{q}$ annihilation.

The main contribution to the total cross section stems from gg fusion is due to the fact that gluons are abundant in the parton distribution function (PDF) of a proton, in contrast to $q\bar{q}$ pairs at the TeV scales. These relative fractions depend on the PDFs of the initial-state hadrons (pp or $p\bar{p}$ collisions) and on the center-of-mass energy as well. For example, in $p\bar{p}$

collisions, $q\bar{q}$ annihilation is enhanced because in that case the annihilation of valence quarks from the proton with the valence antiquarks of the antiproton is now feasible [4]. At the LHC at 13 TeV, assuming a top quark mass of 172.5 GeV, the predicted $t\bar{t}$ production cross section (NNLO+NNLL) is:

$$\sigma_{t\bar{t}} = 833.9^{+20.5\,+21.0}_{-30.0\,-21.0} \text{ pb} \tag{1.23}$$

where the first uncertainty originates from scale independence and the second from PDFs and α_s [5]. The measurements displayed in Figure 1.3a are in agreement with this theoretical prediction.



(a) (b)

Figure 1.3: Summary of the ATLAS and CMS Collaboration measurements of (a) the top-pair production cross-section and (b) the single top production cross-section in the t-channel at 13 TeV. The measurements are compared to theoretical calculations assuming a top mass of 172.5 GeV [6].

Single Top Quark

Instead of forming pairs, top quarks can also be singly produced at the LHC through the electroweak interaction via the exchange of a W boson. The single top quark production at LO can be realized in three modes, depending on the virtuality of the W boson involved in the process; the t-channel (spacelike), the s-channel (timelike) and the tW-channel (on-shell), as displayed in the Figure 1.4 below. A unique feature in single top quark production is the presence of the CKM matrix element V_{tb} in the production vertex, which allows its direct

measurement. Since the W boson interacts only with left-handed quarks and right-handed antiquarks, the top quarks produced are fully polarized.

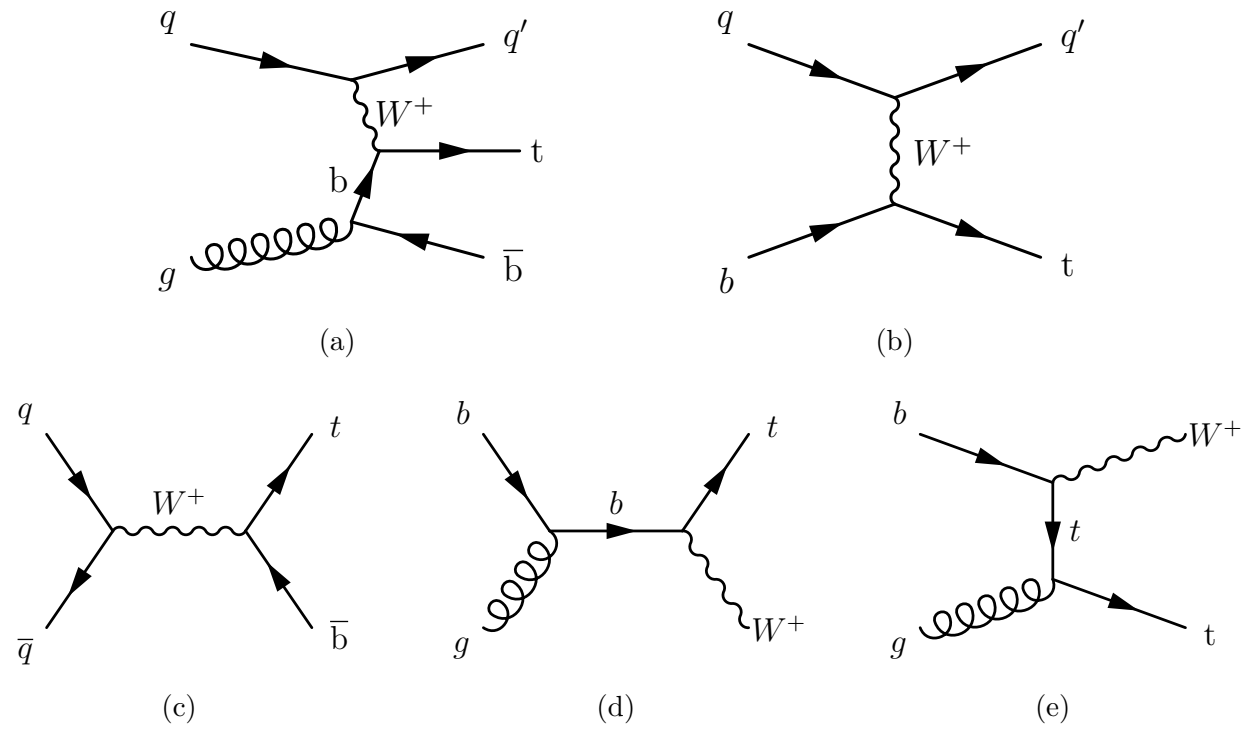


Figure 1.4: Feynman diagrams for electroweak single top production at LO: (a)-(b) t-channel 4FS and 5FS scheme, (c) s-channel, (d)-(e) tW-channel.

The dominant process in single top quark production in pp collisions at the LHC is the t-channel, contributing approximately 70% of the total cross section. Single top quark production in the t-channel can be calculated in two alternate ways: the four-flavor scheme (4FS) and the five-flavor scheme (5FS). In the 4FS, the initial state gluon splits into a $b\bar{b}$ pair from which the resulting b quark interacts with W boson to produce the top quark; these bottom quarks are not considered part of the proton. In contrast, in the 5FS, the initial state b quark originates from flavor excitation occurring inside the proton as part of it [4]. Furthermore, a distinguishing feature of the t-channel production mode is the subtle alteration of the direction of the spectator quark, namely the quark which originates from one of the interacting protons, during the scattering process. During the reconstruction this is translated as a jet at large pseudorapidity $|\eta|$. Calculations for the t-channel cross section at the LHC at 13 TeV at NLO accuracy yield:

$$\sigma^{t}_{t-ch} = 134.2^{+1.5\,+2.1}_{-1.1\,-1.3}\ \text{pb} \tag{1.24}$$

$$\sigma^{\bar{t}}_{t-ch} = 80.0^{+0.8\,+1.6}_{-0.8\,-1.2}\ \text{pb} \tag{1.25}$$

$$\sigma^{t+\bar{t}}_{t-ch} = 214.2^{+2.4\,+3.3}_{-1.7\,-2.0}\ \text{pb} \tag{1.26}$$

where the uncertainties are from scale dependence and from PDFs and α_s [7]. Here, the cross section for producing top quarks is about 1.7 times larger than the cross section for producing top antiquarks due to the fact that (valence) up quarks outnumber down quarks in the proton. These predictions, as shown in Figure 1.3b agree with the recent ATLAS and CMS measurements.

Besides the t-channel, top quarks are singly produced in association with a real W boson around 25% of the time, where the initial b quark is a sea quark originating from the splitting of virtual gluons inside the proton. The predicted cross section (t+$\bar{t}$) of the tW-channel at 13 TeV, computed at NLO in QCD with the addition of third-order corrections of soft-gluon emissions by resuming NNLL terms and assuming top quark mass 172.5 GeV [8], is:

$$\sigma_{tW} \;=\; 179.3^{+1.9\,+2.2}_{-1.8\,-2.2}\ \text{pb} \tag{1.27}$$

The final mode in single top quark production, the s-channel, involves an incoming quark and an incoming antiquark producing a virtual W boson, which in turn decays into a bottom and a top quark approximately 5% of the time. The s-channel has the smallest cross section among the single top production modes, with NNLO approximated calculations at 13 TeV (t+$\bar{t}$) yielding:

$$\sigma_{s-ch} \;=\; 6.84^{+0.06}_{-0.03}\ \text{pb} \tag{1.28}$$

where the uncertainty is from scale dependence only. All of the SM predictions for the single top production cross sections in various channels over a wide range of center of mass energies, as depicted in Figure 1.5, are in good agreement with the respective measurements so far.

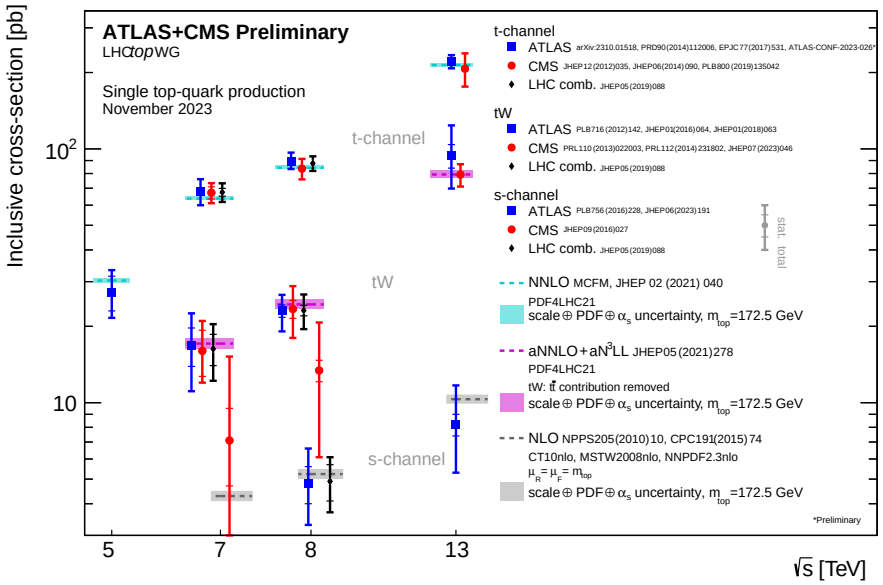


Figure 1.5: Summary of ATLAS and CMS measurements of the single top production cross-sections in various channels as a function of the center of mass energy. The measurements are compared to theoretical calculations assuming a top mass of 172.5 GeV [6].

Chapter 2

The CMS Experiment at the LHC

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a circular particle accelerator operating at CERN laboratories in Geneva, near the Franco-Swiss border. It is composed of a 27-kilometre ring of superconducting magnets with a number of accelerating structures that boost the energy of charged particles, such as protons. These particles are accelerated to high speeds in order to collide with each other. Besides protons, ions of lead (Pb) are also injected in the LHC, hence p-Pb and Pb-Pb collisions can take place too. Collisions are achieved at four interactions points, where the main four experiments are located: CMS, ATLAS, LHCb and ALICE.

In the quest for precision measurements and probing new physics at the TeV scale, the LHC is designed to provide center-of-mass collision energies up to $\sqrt{s} = 14$ TeV. Inside the collider, two particle beams move in opposite directions in distinct beam pipes, approaching the velocity of light. Magnetic fields along the accelerator are used to keep beams on track inside the pipes. In order to achieve the highest possible efficiency with the least energy loss, the magnets have to be superconducting. Therefore, the electromagnets in the LHC are chilled to -271.3°C by using liquid helium and the beam pipes are also kept in ultra-high vacuum to reduce the amount of heat that seeps from the surrounding room-temperature environment [9]. Indicatively, 1232 dipole magnets of 15m length and 392 quadrupole magnets of 5-7m length that bend and focus the beams respectively, form the LHC.

Nevertheless, the LHC is just the final element of a sophisticated acceleration structure in which the beams reach their highest energies, up to 6.8 TeV per beam. The CERN acceleration complex, shown in Figure 2.1, is a succession of machines, each of which accelerates the particles further before injecting them into the next machine in the chain. The first element of the acceleration chain consists of a linear accelerator, the Linac4 [10], operational since 2020, that accelerates negative hydrogen ions to 160 MeV to prepare them to enter the Proton Synchrotron Booster [11], which is also part of the LHC injection chain. Radiofrequency cavities are employed to charge cylindrical conductors, which accelerate the ions by pulling them forward and pushing them from behind as they pass through. The tightness of the hydrogen ion beams is ensured with the assistance of quadrupole magnets. Linac4 boosts ions

in stages: first to 3 MeV by a radio-frequency quadrupole (RFQ), then to 50 MeV by drift tube linacs (DTLs), then to 100 MeV by coupled-cavity drift tube linacs (CCDTLs), and finally to 160 MeV by Pi-mode structures (PIMS). Subsequently, the ions are stripped of their two electrons during the injection from Linac4 into the Proton Synchrotron Booster in order to accelerate up to 2 GeV before being delivered to the Proton Synchrotron (PS) [12], which pushes the beam up to 26 GeV. In this way, the particle accumulation and therefore the beam quality are enhanced. Then, protons are transferred to the Super Proton Synchrotron (SPS) [13], the second-largest machine in CERN's accelerator complex, reaching energies up to 450 GeV and finally up to 6.8 TeV in the two beam pipes of the LHC.

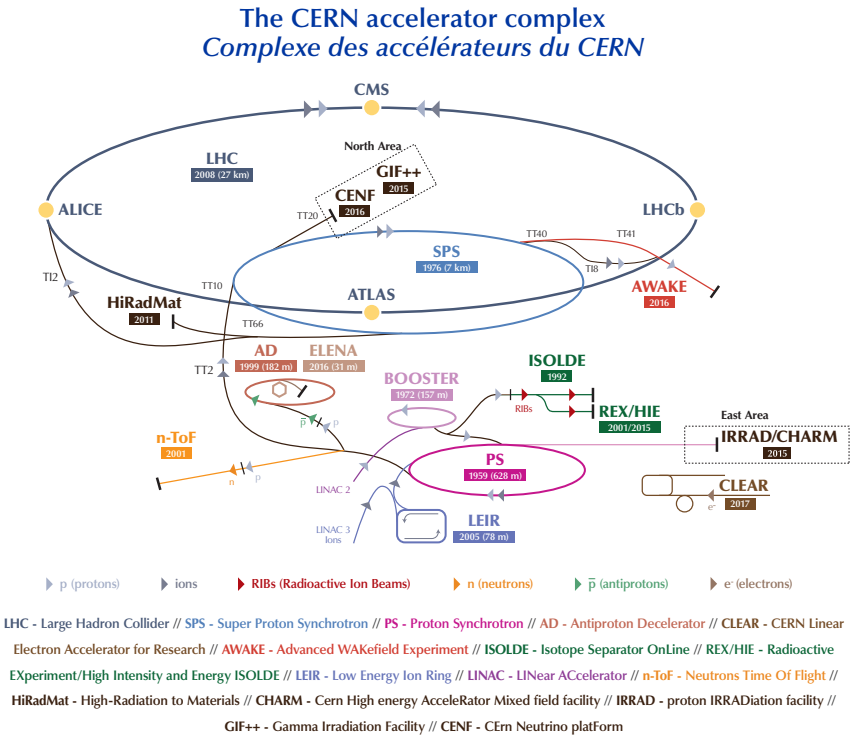
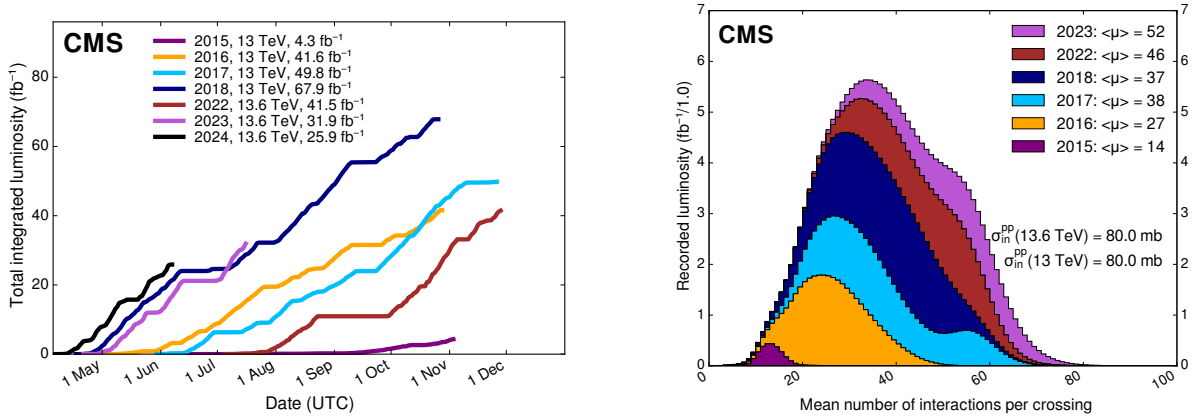


Figure 2.1: The CERN accelerator complex [14].

It is worth noting that at the LHC the beams are not a constant flux of particles, but they are separated in bunches of particles. A typical bunch at the start consists of $1.2 \cdot 10^{11}$ protons and each proton beam contains 2808 bunches. This multitude of protons per bunch leads often to the so-called in-time pileup effect, namely the phenomenon of more than one proton-proton interaction taking place per bunch crossing. This effect is quantified by the average number of interactions per bunch crossing, as shown in Figure 2.2b, where the pileup distributions in CMS for each year are stacked above the previous years over the Run 2 and Run 3 (until 2023) period. In addition to in-time pileup, there can be out-of-time pileup, in which the readout time of the signal in the detector is longer than the time between two consecutive bunch crossings, resulting in the mixing of the two signals.



(a) (b)

Figure 2.2: (a) Integrated luminosity collected by the CMS experiment and (b) average number of interactions per bunch crossing by year of data taking [15].

What is more, besides the number of simultaneous collisions within the same bunch crossing, machine-dependent parameters of particular interest for physics analyses include the center-of-mass energy of the collisions and the instantaneous and integrated luminosity as well. The production rate of a physical process depends on the cross section σ for a specific process and on the instantaneous luminosity $\mathcal{L}$:

$$\frac{dN}{dt} = \sigma\mathcal{L} \tag{2.1}$$

Subsequently, the total number of events produced is derived by integrating equation 2.1 over time. Since the cross section is dependent only upon the center-of-mass energy, which is constant over the time of a fill, the number of events is given by:

$$N = \sigma \int dt\mathcal{L} = \sigma L \tag{2.2}$$

where L is the integrated luminosity. The instantaneous luminosity depends on the parameters of the LHC:

$$\mathcal{L} = \frac{kN_p^2 f\gamma}{4\pi\beta^*\epsilon}F \tag{2.3}$$

where γ is the relativistic gamma factor, N_p is the number of particles per bunch, k is the number of bunches per beam and f is the revolution frequency of the bunches. The normalized beam emittance ϵ gives the spread of the beam in the phase space and the beta function β^* describes the focusing of the beams at the interaction point. Finally, the form factor F accounts for the fact that the two beams are collided at a non-zero angle, thus reducing the effective area. Luminosity, measured in $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, is an important indicator of

the performance of an accelerator: the higher the luminosity, the more data can be gathered to allow observations of rare processes. For this reason, the High-Luminosity LHC (HL-LHC) project [16] aims to increase the integrated luminosity by a factor of 10 beyond the LHC's design value from the beginning of 2029.

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) experiment [17] is one of two large general-purpose detectors built at the LHC. It is 21.6 m in length, 14.6 m in diameter and weighs 14000 tonnes, covering a pseudorapidity range of $|\eta| \leq 5$ with a cylindrical geometry. The main CMS subsystems, in a radial order from the interaction point, are the silicon pixel and strip tracker, the electromagnetic and hadron calorimeters, the superconducting solenoid, and the muon system. A brief overview of each subsystem, along with the trigger and data acquisition system, follows in the next sections.

As shown in the vertical slice of the CMS detector in Figure 2.3, the particle tracks originating from collisions can be traced back to the interaction point. Initially, charged particles leave hits in the tracking system. Particles that interact electromagnetically are absorbed by the electromagnetic calorimeter, while hadrons deposit their energy in the hadronic calorimeter. On the other hand, muons travel through the entire detector and are measured in the muon system. The trajectories of the charged particles are bent by the powerful magnetic field created by the superconducting solenoid, which produces a field of 3.8 T in the inner part of the detector and 2 T in the muon system. In this way, particles are accurately identified and measured, allowing for precise reconstruction with the CMS detector.

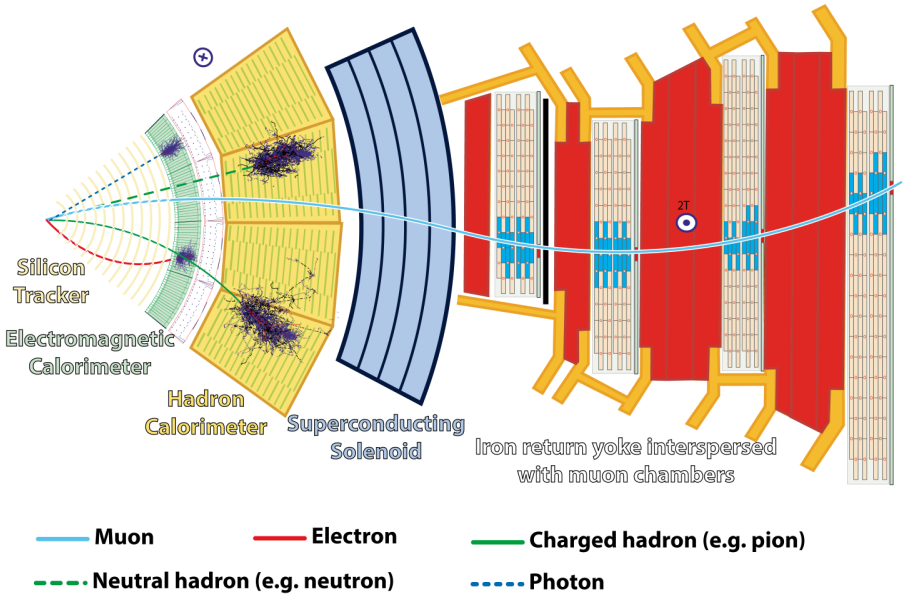


Figure 2.3: Slice of the CMS detector [18].

2.3 The Tracking System

The CMS inner tracking system is designed to accurately reconstruct the trajectories of charged particles emerging from collisions. As these particles move through a powerful magnetic field, their paths curve. Consequently, the momentum of the particles can be inferred; greater momentum results in less curvature. With an accuracy of 10 μm, only a few position measurements are needed for a satisfactory reconstruction. However, surrounding the interaction point, the tracker receives intense flux of particles. This poses challenges to the tracking system. Firstly, high granularity and fast response are necessary features of the detector to avoid attributing trajectories to particles from different bunch crossings. Secondly, the detector components must be capable of withstanding the harsh environment of increased radiation. Therefore, the tracker detector is made exclusively of silicon.

The CMS tracker with a length of 5.8 m and diameter of 2.5 m covers a pseudorapidity range of $|\eta| < 2.5$ [19]. It consists of two subdetectors; a silicon pixel detector built around the interaction point and a silicon microstrip detector around the pixel detector. Incoming particles passing through the silicon detectors that excite the silicon atoms create electron-holes pairs. By using an external electric field, these charges are collected by p-n junctions, shaped in pixels or strips, resulting in a small pulse of electrical signal which is regarded as "hit". The combination of hits along the cylindrical structure of the tracking detectors provides the information for reconstructing the trajectories of the charged particles. In this way, precise reconstruction of the secondary vertices, namely the points where particles other than the initial protons decayed, is feasible as well.

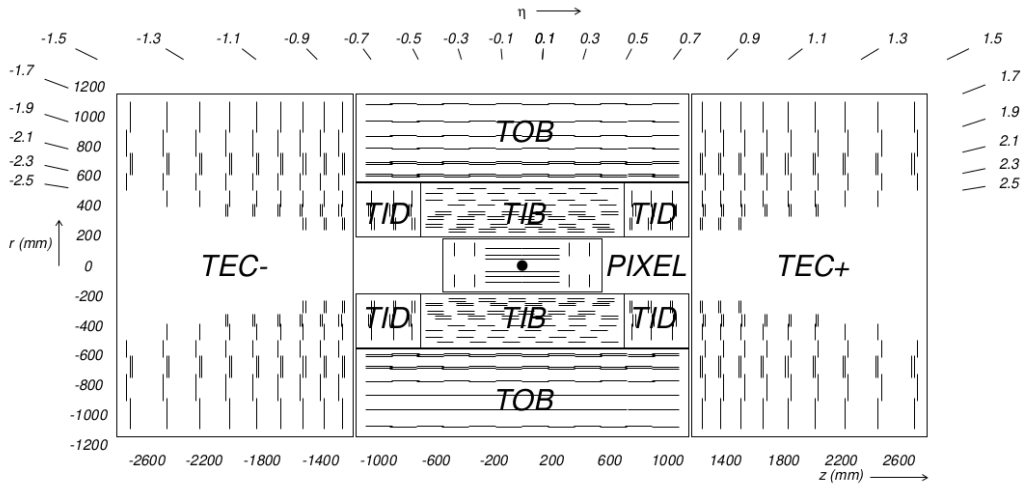


Figure 2.4: Layout of the CMS tracker [19].

As shown in Figure 2.4, the tracker is composed of five components: the Pixel Detector (PIXEL) around the collision point , the Tracker Inner Barrel and Disks (TIB/TID) consisting of 4 barrels and 3 disks at each end, the Tracker Outer Barrel (TOB) consisting of 6 barrel layers surrounding the TIB/TID and the Tracker EndCaps (TEC+/TEC−) composed

of 18 disks in total with each endcap carrying up to 7 rings of silicon micro-strip detectors.

Pixel Detector

The pixel system is the part of the tracking system closest to the beam pipe consisting of 4 barrel layers at radii of 3, 7, 11 and 16 cm from the beam line and 3 disks on each side of the interaction region at a distance of 29, 40, and 52 cm from the interaction point [20]. The detector contains around 124 million pixels with pixel cell size of 100×150 μm^2 providing high-precision tracking points in the $r-\phi$ plane and the z coordinate. This is crucial for achieving small impact parameter resolution to reconstruct secondary vertices. As previously mentioned, due to its proximity to the interaction point, the detector operates in a region with high track multiplicity and severe irradiation. As a result, the sensors were designed with a n+ pixel on a n−substrate. This design increases radiation tolerance by allowing for partially depleted operation. Additionally, to maintain high performance of the CMS tracking system, the original pixel detector (Phase 0) was replaced in 2017 with an upgraded system (Phase 1) during the extended LHC Shutdown.

Silicon Strip Tracker

The next parts of the tracker that particles pass through are the layers of silicon strip detectors. With a total radius of 130 cm, these layers consist of 15200 modules, housing approximately 10 million detector strips and 72000 readout microelectronic chips. Each module comprises three main parts: one or two silicon sensors, the structure that provides mechanical support and the read-out electronics. The strip tracker operates similarly to the pixel detector. The sensor elements in the strip tracker are single-sided silicon microstrip sensors of the p-on-n type and the signals from the silicon sensors are amplified, shaped, and stored by a custom integrated circuit, the APV25 [19]. The front-end readout electronics for a detector module are housed within a multichip module called hybrid. Because the tracker and its electronics are exposed to high levels of radiation, the system is kept at -20°C to minimize potential damage to the silicon and to prevent it from spreading.

2.4 The Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (ECAL) is a homogeneous, fine-grained lead tungstate ($PbWO_4$) crystal calorimeter designed to measure the energies of electrons and photons with high precision. The selection of $PbWO_4$ with a density of 8.28 g/cm^3 was driven by its high radiation tolerance, short radiation length of $X_0 = 0.89$ cm, small Molière radius of $r_m = 2.19$ cm, and its fast response, with 99% of the light collected within 100 ns [21]. When high-energy particles that interact electromagnetically traverse the ECAL, a cascade of less energetic secondary particles is produced. Specifically, photons generate this cascade through electron-positron pair production, while charged particles do so via Bremsstrahlung. This process continues until the secondary particles are fully absorbed. The scintillation

light produced by the crystals, which is proportional to the energy of the initial particle, is detected by avalanche photodiodes. In this way, the energy of photons and electrons is accurately measured.

The ECAL, Figure 2.5, consists of 75848 crystals arranged in the central barrel section (EB), which has pseudorapidity coverage up to $|\eta| = 1.48$. The coverage is extended up to $|\eta| = 3.0$ by the two endcaps (EE) that close the barrel. The barrel crystals measure 23 cm in length, corresponding to 26 radiation lengths, and have a front-face cross-section of 2.2×2.2 cm^2, whereas the endcap crystals are 22 cm long, corresponding to 25 radiation lengths, with a front-face cross-section of 2.86×2.86 cm^2 [22]. The total crystal volume is 11 m^3, with a combined weight of 92 tons. The barrel calorimeter is organized into 36 supermodules, each containing 1700 crystals, while the endcaps consist of two dees each comprising 3662 crystals. Finally, a preshower detector (ES) composed of lead absorbers and silicon strip sensors, featuring 4,288 sensors and approximately 3 radiation lengths in thickness, is positioned in front of the endcaps with pseudorapidity coverage $1.65 < |\eta| < 2.6$ to assist in the separation of photons and π^0 mesons.

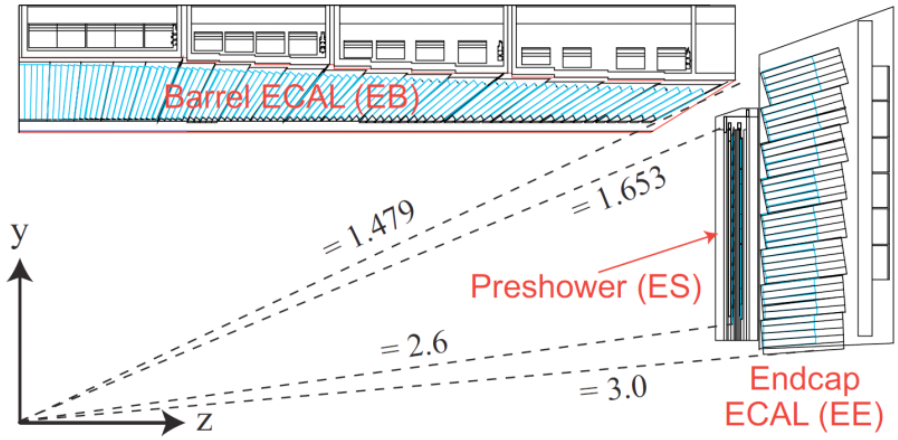


Figure 2.5: Geometric view of the ECAL at CMS [21].

The performance of the ECAL has been tested with electron beams in a setup without magnetic fields nor any other materials in front of the calorimeters. The energy (E) resolution of the barrel is typically parametrized as:

$$\frac{\sigma}{E} = \frac{2.8\%}{\sqrt{E}} \oplus \frac{12\%}{E} \oplus 0.3\% \tag{2.4}$$

where the energy E is measured in GeV [21]. The initial term is referred to as stochastic and represents the statistical variations in the showering process and the number of photons detected by the photomultipliers. The second term is the noise term representing the electronic noise in the readout chain. In addition, there is the last term which is constant and not energy dependent. It can arise from various factors such as the geometry of the detector,

the temperature gradients or the radiation damage. It stems from several factors, such as the detector's geometry, temperature variations, or radiation-induced damage.

2.5 The Hadronic Calorimeter

The CMS hadronic calorimeter (HCAL) is a sampling calorimeter that, together with the ECAL, forms the CMS calorimetry system. As its name implies, the HCAL is designed to measure the energy of charged and neutral hadrons, while also providing information for indirect measurements of particles that do not interact with the detectors, such as neutrinos. The HCAL consists of alternating layers of brass absorbers and plastic scintillators and it determines the position, energy and time of arrival of a particle in the detector. Incoming particles passing through the HCAL interact with the brass material, creating hadronic cascades. The secondary particles then interact with the alternating layers of active scintillation material, producing rapid light pulses. These pulses are collected by special light fibers and transmitted to photodetectors, which amplify the signal. In this way, the total amount of light collected in a given region provides the measure of the initial particle's energy.

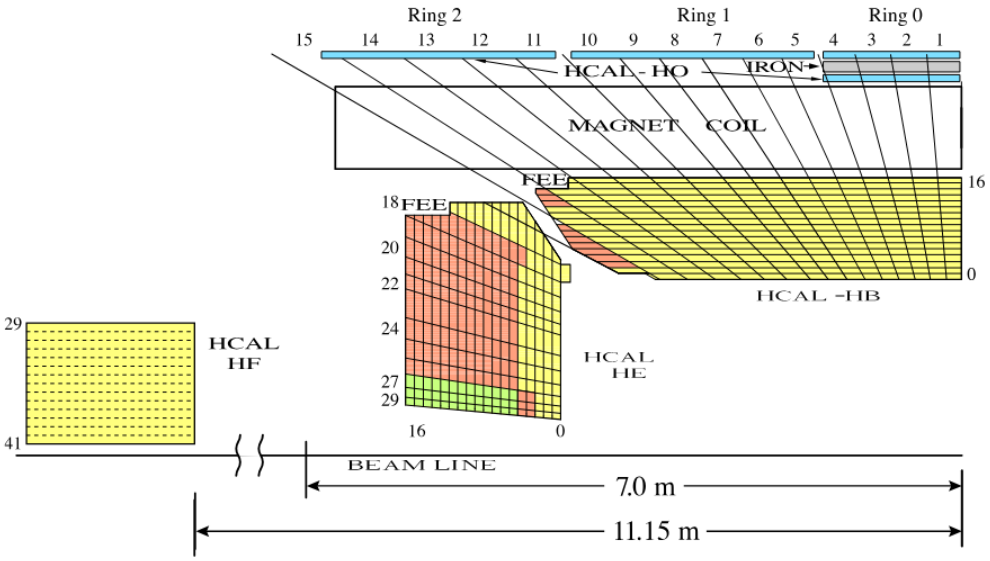


Figure 2.6: Layout of the HCAL at CMS [23].

The HCAL includes four distinct subsystems: the barrel (HB), the endcap (HE), the outer (HO) and the forward (HF) calorimeters, as shown in Figure 2.6. The HB covers the pseudorapidity range $|\eta| < 1.4$ and its effective thickness depends on the polar angle. It contains 5.82 nuclear interaction lengths (the mean distance travelled by a hadron before undergoing inelastic nuclear interaction, λ_l) at $|\eta| = 0$ and 10.6 λ_l at $|\eta| = 1.4$. To contain the entire hadronic shower, the HO is placed outside the solenoid. At $|\eta| = 0$, where the HB's effective thickness is the lowest, two HO layers are placed around a 19.5 cm thick

tail catcher iron piece. This increases the HCAL thickness in the barrel region to at least 11.8 λ_I. The pseudorapidity range of $1.3 < |\eta| < 3.0$ is covered by the HE detector. To achieve maximum coverage, a Cherenkov-based forward calorimeter, the HF, is positioned 11.15 meters from the interaction point, extending the pseudorapidity coverage to $|\eta| = 5.2$. Due to the high particle flux, a design tolerant to high radiation was necessary. Therefore, a Cherenkov detector was chosen. The HF is composed of steel absorbers and quartz fibers as the active medium.

Since most particles begin to shower in the ECAL, the hadronic energy resolution is dependent on both the ECAL and the HCAL [23]. The resolution is customarily parametrized as

$$\frac{\sigma}{E} = \frac{84.7\%}{\sqrt{E}} \oplus 7.4\% \tag{2.5}$$

Because the HCAL is a sampling calorimeter, its energy resolution is generally less precise than that of the ECAL.

2.6 The Superconducting Solenoid

The next layer of the CMS detector, surrounding the tracking system and the calorimeters, is the superconducting solenoid. This solenoid has an inner diameter of 5.9 m and a length of 12.9 m. This solenoid has an inner diameter of 5.9 meters and a length of 12.9 meters. It is constructed using niobium-titanium coils, which generate a powerful magnetic field of 3.8 T and storing an energy of 2.6 GJ. The coil of the solenoid is cooled to 4.7 K using 6000 l of liquid helium. As previously discussed, the magnet plays a crucial role in bending the trajectories of charged particles, thereby enabling the determination of their momentum and electric charge. Additionally, the solenoid is surrounded by yokes made from common structural steel, weighing a total of 12500 tonnes. These yokes consist of 5 wheels and 2 endcaps and are used to reduce stray magnetic fields by returning the solenoid's magnetic flux. The relative ease with which these elements can be moved simplifies the assembly of the subdetectors and provides mechanical support to the experiment's overall structure. Beyond the solenoid, the magnetic field strength is 2 T in the opposite direction, encompassing the muon system. The steel yokes also function as absorbers for the four interleaved layers of muon chambers, which allow for independent measurements of muon momentum, separate from the inner tracking system.

2.7 The Muon System

The precise measurement of muons is vital for CMS, as implied by the experiment's name. Specifically, the muon system is utilized for muon identification, momentum measurement, and triggering. Muons register hits in four stations in the muon chambers, unlike all other particles which stop at the calorimeters. Subsequently, the hits are combined to create high-purity muon tracks. The muon system has a pseudorapidity coverage up to $|\eta|$ <2.4. The

solenoid's shape dictates a cylindrical design for the muon system, which includes a barrel section and two planar endcap regions. The solenoid magnet and the flux-return yoke provide robust momentum resolution and triggering capabilities. Additionally, the yoke acts as a hadron absorber, enabling muon identification.

Three types of gaseous detectors are employed by CMS for effective muon identification: Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). The muon chambers, covering 25000 m^2 of detection planes, require cost-effective and reliable components [19]. The operating principle of gaseous detectors, broadly speaking, involves particles with sufficient energy ionizing the gas between two electrodes within the detector, resulting in an electric signal generated by the drift of the electrons toward the electrodes.

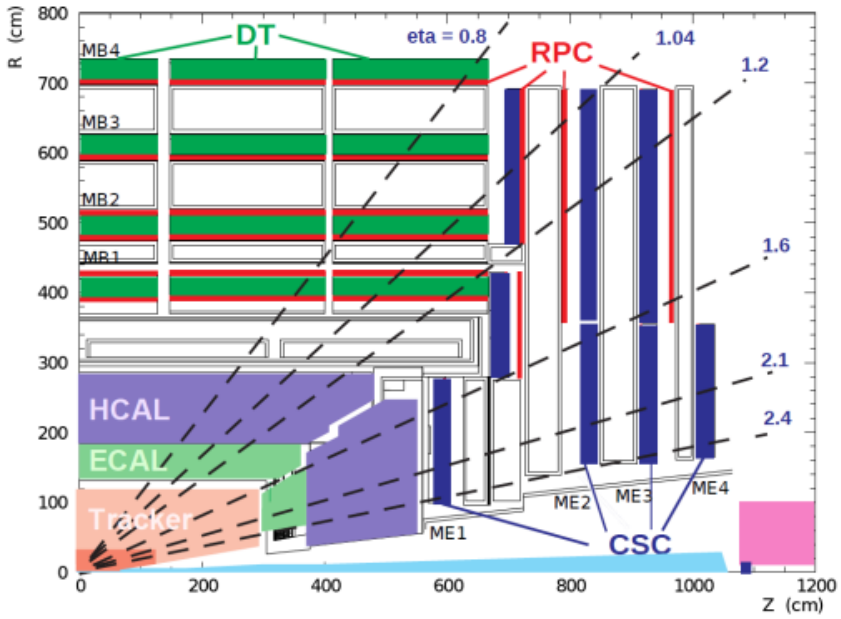


Figure 2.7: The layout of the CMS detector, highlighting the four DT stations in the barrel (MB1–MB4 in green), the four CSC stations in the endcap (ME1–ME4 in blue) and the RPC stations (in red) [24].

Figure 2.7 shows the CMS detector, highlighting the muon system. The DTs are filled with a gas mixture of Ar and CO_2 and have gold-plated anode wire at the center. The drift tubes consist of four stations, covering the barrel region up to $|\eta| < 1.2$. The CSCs are placed at the endcaps, covering the $|\eta|$ region from 0.9 to 2.4, where the magnetic field is strong and not uniform. These chambers are filled with a gas mixture of Ar, CO_2 and CF_4 and have a fast timing response. CSCs consist of multiple anode wires and cathode strips arranged orthogonally to each other, enabling the simultaneous triggering and precise measurement of the momenta of muons in the endcap region as the DTs do in the barrel region. The RPCs, the third component of the muon system, cover the $|\eta|$ region up to 1.6. Designed to provide excellent timing resolution, the RPCs consist of two parallel plates with high resistivity and

are primarily used for triggering.

2.8 The Trigger and Data Acquisition System

At the LHC, each beam crossing at the design luminosity of 10^{34} cm^{-2} s^{-1} results in approximately 50 inelastic proton-proton collisions producing a data rate of several terabytes per second. These collision rates and the corresponding data rates exceed the current storage capabilities. Furthermore, the majority of these events are not of interest to high-energy physics analyses, as they are dominated by soft multi-jet scattering processes. Therefore, a trigger system is essential to optimize the data storage rate by selecting physically significant events and discarding the rest.

The CMS Trigger and Data Acquisition System (TriDAS) [25] inspects the the detector information and accomplishes the event selection for archiving and later offline analysis in two steps. The Level 1 Trigger (L1), as shown in Figure 2.8, reduces the incoming average data rate from approximately 40 MHz to a maximum of 100 kHz by processing fast trigger information coming from the calorimeters and the muon chambers while holding the high-resolution data in pipelined memories in the front-end electronics. The second step is the High-Level Trigger (HLT), which is designed to reduce the rate of events even further to a final output rate of 100 Hz. The HLT accesses the full read-out data and performs sophisticated computations comparable to those conducted by the offline analysis software.

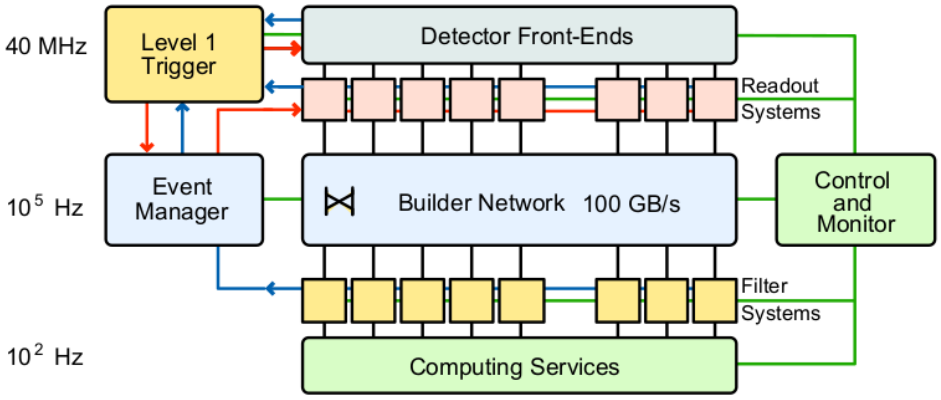


Figure 2.8: Architecture of the CMS DAQ system [19].

Furthermore, apart from forwarding the accepted events, the CMS DAQ/HLT system, it also reserves a small sample from the rejected events to monitor the performance of the CMS detector. Additionally, another important function of the DAQ system is operating the Detector Control System (DCS), which supervises all detector components and the experiment's general infrastructure, thereby ensuring the collection of high-quality physics data.

Chapter 3

Event Reconstruction at the LHC

3.1 Hard Processes

In proton-proton collisions a multitude of particles is produced, and these subatomic particles are studied through their interactions with the detector. In this high-energy environment, the interactions often involve a large transfer of momentum, known as hard scattering [26]. Hard processes enable the probing of distance scales much smaller than the proton's radius, allowing investigation at the parton level, specifically its constituents: quarks and gluons, as shown in Figure 3.1. The cross section calculations of such events consist of a term describing the partonic scattering with cross section $\hat{\sigma}$ and factors accounting for the incoming flux of partons, described by the parton density functions (PDFs) $f_{i,p}$. A general expression for $pp \to X$ can be written as:

$$\sigma(pp \to X) \;=\; \sum_{i,j} \int dx_1 dx_2 \, f_{i,p}(x_1, \mu_F^2) \, f_{j,p}(x_2, \mu_F^2) \, \hat{\sigma}_{ij}(x_1 x_2 s, \mu_R^2, \mu_F^2) \tag{3.1}$$

where the summation is over the possible initial-state partons with longitudinal momentum fractions χ_1 and χ_2 giving rise to the final state X at a center-of-mass energy $\sqrt{\chi_1 \chi_2 s}$. Additionally, the cross section depends on the two energy scales μ_F and μ_R, referred to as the factorization and renormalization scales. Under the assumption of factorization, PDFs are universal at a specific momentum scale, and their evolution at hard scales can be calculated using perturbative QCD. This is indeed true for certain processes, and as a result, PDFs are well understood across a broad range of momentum fractions and scales.

As expected, production of particles in a hadron-hadron collision involves parton-parton scatterings, initial-state radiation (ISR), final-state radiation (FSR), and beam-beam remnants (BBR), thus increasing the complexity of a collision event. In addition, the large parton densities available in pp collisions often result in more than one parton-parton scattering, an effect referred to as multiple parton interaction (MPI). The overall combination of particle production from BBR, ISR, FSR, and MPI is called the underlying event (UE) [27]. The UE activity cannot be uniquely separated from initial and final state radiation and therefore

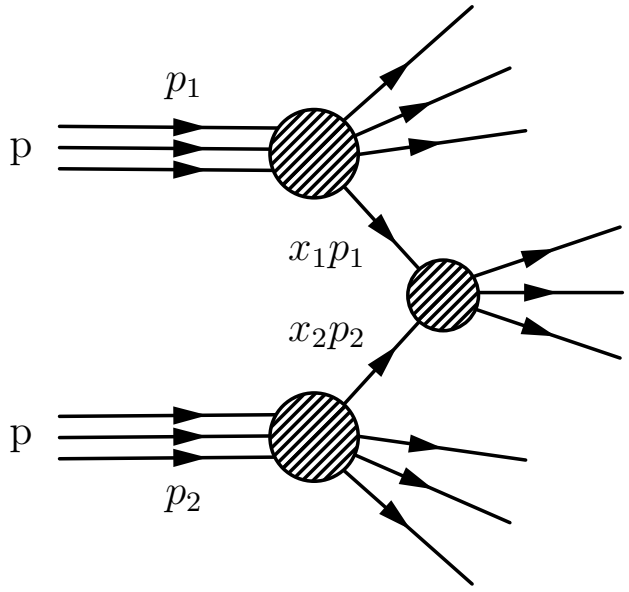


Figure 3.1: Schematic representation of a proton-proton collision with a hard process.

a good description of its properties is crucial for precision measurements of the SM in the LHC.

To perform such measurements, simulations assuming a specific physical process taking place in the hard interaction of the collision are necessary. In particular, the gluons emitted from the additional radiation split into further secondary particles, creating a parton shower. The evolution of the parton shower is modeled using Monte Carlo simulations, which account for the probabilistic nature of quantum processes involved. Once the energy of the partons reaches the order of $\Lambda_{\text{QCD}} \sim 200$ MeV, they start to form color-neutral bound states due to QCD confinement. At this non-perturbative regime for QCD at low energy scales, hadronization is described using phenomenological models. After the four-vectors of the final-state particles have been generated by the Monte Carlo generators, a detailed simulation of the detector is performed with GEANT4, which accurately models the interaction of particles with the detector's material and geometry. This process covers a wide range of electromagnetic, hadronic, and optical interactions over an extensive energy range, from a few hundred eV to the TeV scale.

3.2 The Particle Flow Algorithm

The produced particles from the collisions can be identified based on their interaction signatures with the different subsystems: charged particles "hit" the tracking system, electrons and photons deposit energy in the ECAL, charged and neutral hadrons are absorbed by the HCAL and muons uniquely leave an imprint on the muon chambers. The information

obtained from these electronic signals from all subdetectors needs to be combined optimally to access the elementary hard process that occurred at the collision point. For this purpose, CMS uses the particle flow (PF) algorithm to reconstruct events [28]. The main goal of the PF algorithm is, firstly, to perform efficient and pure track reconstruction of the PF elements, such as tracks and calorimeter clusters, by disentangling overlapping showers, and secondly, to link the energy deposits of each particle efficiently across the subdetectors. In overview, the algorithm iterates over the possible PF elements, and after reconstruction, removes them from the PF block. The order of reconstruction is fixed: muons are identified first, followed by isolated photons and electrons, and lastly hadrons and non-isolated photons.

3.2.1 Tracks and Vertices

Tracks are reconstructed in an iterative way in four steps: seeding, pattern recognition, final fitting, and selection. The first step, called seeding, provides initial track candidates from hits in the tracker, typically two or three hits and the position of the primary vertex. In the second step, for pattern recognition, the initial seeds are extrapolated outward to search for new hits in the outer layers of the tracker that are compatible with the original track. In the third step, the best estimate of the track parameters is determined by fitting the final collection of hits associated with the track using a Kalman Filter (KF) algorithm. Missing hits or outliers that are not compatible with the fitted trajectory may be removed from the track. The final step, the selection process, includes retaining only the track candidates that satisfy specific quality criteria. Then, this algorithm is iterated multiple times, each time using different seeds and quality criteria. In order to reduce the complexity of the combination process, after each iteration, the hits associated with determined tracks are discarded in subsequent iterations.

Once tracks are reconstructed, the primary vertex is determined, along with any other vertices arising from pileup effects in the same LHC bunch crossing. This is achieved by clustering the tracks that appear to originate from the same interaction vertex in each event, based on a deterministic annealing algorithm. The spatial coordinates of the reconstructed vertex are then specified by the adaptive vertex fitter algorithm. This algorithm determines the best vertex parameters, discards any outlier tracks, and assigns a weight to each of the remaining tracks based on how well the track's kinematics align with the vertex position.

3.2.2 Muon Reconstruction and Identification

Muon reconstruction is based on combining the information from the inner tracking system with that of the muon chambers. There are three different types of reconstructed muons:

- *Standalone muons* are reconstructed based only on the information provided by the muon system
- *Tracker muons* correspond to the muons whose track in the tracker system, when extrapolated outward to the muon system, is compatible with the position of at least

one segment in the muon segment

- *Global muons* are obtained by simultaneously fitting the hits from the inner tracker and muon chambers inwards with a Kalman Filter

Typically, over 99% of the muons with sufficient p_T are reconstructed either as global or tracker muons or both. In the case muon candidates in the two categories share the same inner track, they are merged into a single candidate. On the other hand, muons with inadequate p_T (less than 10 GeV) are usually reconstructed as tracker muons, since they are less energetic and are scattered in the iron return yoke.

The PF algorithm, after muon reconstruction, proceeds to muon identification by requiring the muon candidates to pass quality identification criteria based on the global and tracker muon properties. Firstly, global muons are considered isolated if the sum of p_T of additional inner tracks and energy deposits in the calorimeter with a distance $\Delta R < 0.3$ to the muon direction are less than 10%. In this way, charged hadrons mistakenly considered as muons are adequately rejected. The relative isolation variable is required to be small and is defined as:

$$I_{rel} = \frac{I^{ch} + \max\left(I^{\gamma} + I^{nh} - 0.5 \times I^{pu}, 0\right)}{p_T^{\mu}} \tag{3.2}$$

where I^{ch} is the energy of charged hadrons, I^{γ} the energy photons and I^{nh} the energy of neutral hadrons within a cone of fixed radius R around the muon. The charged hadron energy includes a correction for pileup (PU) effects, denoted as I^{pu}. The last term accounts for an estimate of the neutral pileup energy by subtracting half of the charged PU energy from the neutral hadron energy.

However, muons inside jets, originating from semileptonic heavy-flavor decays or from charged-hadron decays in flight, need additional criteria for proper identification. To address the contributions from charged hadrons misidentified as muons or, conversely, unidentified muons considered as charged hadrons, tight ID criteria are usually applied to non-isolated muons. This ID selection requires that the muon candidate be reconstructed as both a PF muon and a global muon, and the normalized χ^2 of the global muon track fit must be $\chi^2/ndf < 10$. Moreover, the associated muon track must have more than ten hits in the tracker, with at least one in the pixel detector, and should be matched with two muon segments in at least two muon stations. Additionally, it should have a transverse impact parameter $|d_{xy}| < 2$mm and a longitudinal impact parameter $|dz| < 0.5$cm with respect to the primary vertex. In total, different ID criteria might be applied, depending on the specific physics analysis in question.

Electron Reconstruction

On the other hand, electron reconstruction is based on combining information from both the silicon tracker and the energy deposits in the ECAL. There are two complementary methods for this purpose. The first method identifies the energy deposits by grouping neighboring crystals with considerable energy into superclusters (SCs). The second method utilizes the

PF algorithm, where the initiated electron tracks in the silicon detectors are extrapolated to the ECAL and matched to the corresponding superclusters. In this way, the efficiency for low p_T and non-isolated electrons is enhanced.

The reason electrons are more challenging to reconstruct than muons is their energy losses via bremsstrahlung; they emit photons as they pass through the tracker material before the ECAL, resulting in alterations to the electron's momentum and trajectory. Since the direction of the bremsstrahlung photons generally does not coincide with the electron's ϕ direction, the energy of the electron reaches the ECAL with a significant spread in ϕ. Therefore, a modified Kalman Filter, known as the Gaussian Sum Filter algorithm (GSF), is applied to account for the electron energy losses due to photon emissions.

3.2.3 Jets Reconstruction

A jet is a collection of hadrons and other particles produced from the process of hadronization of a quark or a gluon within a well-defined cone of fixed radius R. Quarks and gluons produced from inelastic scattering, which are colored partonic states, cannot exist freely due to QCD confinement. Therefore, they hadronize into colorless bound states, creating a shower of secondary particles. These particles, reconstructed by the PF algorithm, are clustered by identifying the values of the distances d_{ij} between two candidates (i and j) and the distance d_{iB} assigned to each individual candidate in relation to the beam axis, as defined by the anti-kT jet clustering algorithm [29]. Between these quantities, if dij has the smallest value, candidate i is merged with candidate j by summing their four-momenta. Conversely, if d_{iB} has the smallest value, candidate i is identified as a jet and removed from the list of particles. This procedure is iterated until all particles are clustered into jets. The distance metrics d_{ij} between two PF objects and d_{iB} between the PF object and the beam axis are defined as:

$$d_{ij} = \min\left(p_{Ti}^{2k}, p_{Tj}^{2k}\right) \cdot \frac{\Delta_{ij}^2}{R^2} \quad \text{and} \quad d_{iB} = p_{Ti}^{2k} \tag{3.3}$$

where $\Delta_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$, and p_{Ti}, y_i, and ϕ_i are, respectively, the transverse momentum, rapidity, and azimuth of particle i. In addition to the radius parameter R, there is a parameter k that defines the type of jet-clustering algorithm. In the case of the anti-k$_T$ algorithm, the value is $k = -1$. This algorithm is preferred in most analyses because the reconstructed jets are resilient to both the emission of soft radiation (infrared safety) and the presence of collinear decay products (collinear safety), while also maintaining an intuitive conical structure in the η-ϕ plane.

Efficient jet reconstruction is necessary since jets carry information about the initial parton's energy and momentum. Depending on the stage of the hadronization process, jet reconstruction can be divided into three stages: the parton level, which consists of the initial parton (quark or gluon) from which the jet originated; the particle level, which includes the particles produced from the initial parton after its hadronization; and the detector level, which refers to the jet we observe due to the interaction of the shower particles with the detector systems. To link the measured properties from the calorimeters with the underlying

parton, jet energy corrections are applied to the energy deposits to account for contamination from pileup effects and detector noise, the non-linearity of the calorimeter response in jets' p_T and η, and for residual differences between data and simulation responses. These corrections are used to rescale the four-momentum of the initial jets.

Jet Identification from b-quarks

Jets originating from b-quark decays are typically referred to as b-jets, and the process of identifying them is known as b-tagging. A distinct feature of b hadrons is their relatively long lifetime of approximately 1.5 ps, which allows them to travel a measurable distance in the detector before decaying, often resulting in the formation of secondary vertices. This property is a powerful tool for isolating physics processes involving b-quark production, and several methods have been developed for b-tagging at CMS, such as the Combined Secondary Vertex (CSV) and Deepjet algorithms. The most prominent method, Deepjet [30], is a multiclass flavor tagging algorithm that exploits low level variables and deep neural network architecture for high b-tagging efficiency. The Deepjet algorithm defines three working points: loose, medium and tight corresponding to cut values in the discriminator distribution. These cuts are set to achieve misidentification rates for light jets as b-jets of 10%, 1%, and 0.1% respectively.

3.3 Kinematic Variables

In collisions at the LHC at 13 TeV, each proton comes in with back-to-back momenta of E_p=6.5 TeV, conventionally taken to be in the z-direction. At the TeV scale, the mass of the protons is negligible, so the four-momentum of the protons can be written as:

$$P_1^\mu = (E_p, 0, 0, E_p) \quad \text{and} \quad P_1^\mu = (E_p, 0, 0, -E_p) \tag{3.4}$$

In hadron-hadron collisions, the center-of-mass frame of the pp system is not the center-of-mass frame of the colliding partons. Taking into account the fact that the partons, possessing an unknown fraction of the proton's momentum, are the ones participating in the collisions, there is an a priori unknown longitudinal momentum component of the parton hard scattering decay products. Therefore, it is preferable to choose variables for expressing quantities that are independent of the selection of the lab or parton center-of-mass reference frame, namely those that are invariant under Lorentz longitudinal boosts. In proton collisions, a Cartesian coordinate system is employed in which the z axis is aligned with the beam direction, while the x and y axes define the transverse plane orthogonal to the beam, as explained in [31]. The origin of the coordinate system is defined as the point where the beams collide. Due to the cylindrical shape of the detectors, it is customary to use the angles of the cylindrical coordinates (r, ϕ, θ), where r is the distance from the z-axis, ϕ is the azimuthal angle measured in the xy-plane starting from the x-axis, and θ is the polar angle measured in the rz-plane starting from the z-axis, as shown in Figure 3.2a.

The momentum of each particle, $\vec{p} = (p_x,p_y,p_z)$, can be decomposed into the longitudinal component along the z-axis and the transverse component p_T. The transverse momentum vector and its magnitude are given by the following relations:

$$\vec{p_T} \equiv (p_x, p_y, 0) = (p_T \cos\phi, p_T \sin\phi, 0) \tag{3.5}$$

$$p_T = |\vec{p_T}| = \sqrt{p_x^2 + p_y^2} \tag{3.6}$$

where the azimuthal angle ϕ is expressed as a function of the momentum components as:

$$\phi = \tan^{-1}\left(\frac{p_x}{p_y}\right) \tag{3.7}$$

Both the transverse momentum p_T and the azimuthal angle ϕ are Lorentz invariant under longitudinal boosts, unlike the component p_z. Additionally, the polar angle θ is also not invariant:

$$\theta = \tan^{-1}\left(\frac{p_T}{p_z}\right) \tag{3.8}$$

As a result, the introduction of a Lorentz-invariant variable requires replacing θ with the quantity:

$$y \equiv \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{3.9}$$

where E and p_z are the measured energy and z-component of momentum of the particle. The quantity y is called rapidity, and it serves as a measure of the relativistic velocity of the particle. It has the property that under a longitudinal boost, it changes only by a constant, so the differences Δy remain invariant.

In the case of massless particles, it becomes $p_z = |\vec{p}| = E\cos\theta$, so equation 3.9 takes the form:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) = \frac{1}{2}\ln\left(\frac{1 + \cos\theta}{1 - \cos\theta}\right) = \frac{1}{2}\ln\left(\frac{2\cos^2(\theta/2)}{2\sin^2(\theta/2)}\right) = \ln\left(\cot(\theta/2)\right) \tag{3.10}$$

From this equation, the geometric quantity η is defined, which is called pseudorapidity:

$$\eta \equiv -\ln\left(\tan(\theta/2)\right) \tag{3.11}$$

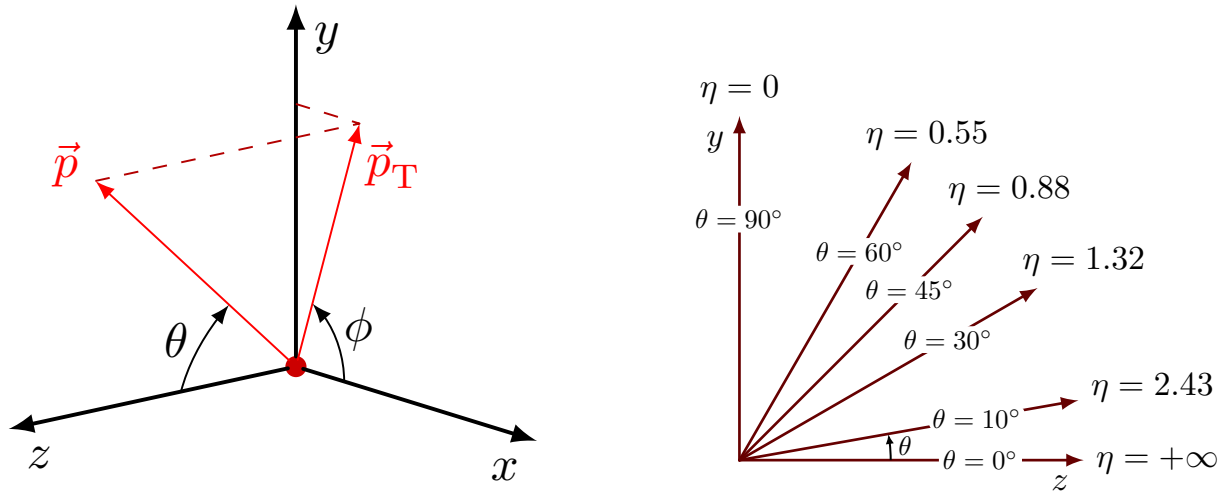
For massless particles, pseudorapidity is equal to rapidity. Figure 3.2b shows the values of η for different angles θ.

Again for massless particles, the anglular metric can be written as a function of η:

$$(d\Omega)^2 = (d\theta)^2 + \sin^2\theta(d\phi)^2 = \frac{1}{\cos^2\eta}\left[(d\eta)^2 + (d\phi)^2\right] \tag{3.12}$$

Thus, the angular distance can be defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.13}$$



(a) (b)

Figure 3.2: (a) The coordinate system showing momentum components and angles in proton collisions and (b) The relationship between pseudorapidity η and polar angle θ.

which is Lorentz-invariant under boosts. For a fixed ΔR, as the value of η increases, the angular separation $\Delta\phi$ decreases. This implies that the phenomena of underlying events and pileup tend to cause greater dilution in the forward region, namely the region with large η, in contrast to the central region, which is approximately uniform in the $\eta - \phi$ plane.

In detectors, the quantities observed are inferred through the interaction of particles with the detector components. However, some particles cannot be observed, such as neutrinos, which exit the detector without interacting at all. For this reason, their momentum is calculated indirectly through the missing transverse energy, which is equal to the negative vector sum of all the individual transverse momenta $\vec{p_{T}}_i$ of an event:

$$\vec{p_T}^{\,\mathrm{miss}} = \sum_i \vec{p_{T}}_i \quad \text{for} \quad i = \text{leptons, jets} \tag{3.14}$$

In a similar way, the missing transverse energy, E_T^{miss}, is defined as the energy lost from the event due to undetectable particles and is equal to the absolute value of the negative sum over all detected particles:

$$E_T^{\mathrm{miss}} = \left|\vec{p_T}^{\,\mathrm{miss}}\right| = \left|\sum_i \vec{p_{T}}_i\right| \tag{3.15}$$

Another useful quantity can be the sum of the transverse momenta of hadronic jets:

$$H_T = \left|\sum_j \vec{p_{T}}_i\right| \quad \text{where} \quad j = \text{jets} \tag{3.16}$$

In addition, an important quantity is the invariant mass, m, which is equal to the rest mass of a particle with four-momentum $p^\mu = (E, \vec{p})$ and is defined as:

$$m \;=\; \sqrt{p^\mu p_\mu} \;=\; \sqrt{E^2 - |p|^2} \tag{3.17}$$

For a system of n particles, perhaps coming from the decay of an initial particle, the invariant mass can be calculated as:

$$m = \sqrt{\left(\sum_{i=1}^{n} E_i\right)^2 - \left(\sum_{i=1}^{n} \vec{p}_i\right)^2}, \tag{3.18}$$

Finally, the transverse mass, m_T, can be calculated using only the transverse components of the momentum, according to the relation:

$$m_T = \sqrt{(E_{T,1} + E_{T,2})^2 - (\vec{p}_{T,1} + \vec{p}_{T,2})^2}. \tag{3.19}$$

where $E_T = \sqrt{m^2 + p_T^2}$ is correspondingly referred to as the transverse energy. If $m_T = m$ then the momenta p_1 and p_2 are purely in the transverse plane ($\eta = 0$), while if $m_T = 0$, the momenta p_1 and p_2 are longitudinal. In all other cases, $0 \leq m_T \leq m$.

Chapter 4

Analysis

In this chapter, a novel method for reconstructing top quarks using events with a single top in the t-channel is introduced to measure the mass. For this purpose, machine learning techniques, such as Boosted Decision Trees with gradient boosting, are employed to accurately estimate the neutrino's longitudinal momentum and the transverse momenta of b-tagged jets, both observed in the final state of the relevant process. In addition, a kinematic fit is applied to constrain the mass of the W boson in the $W \to l\nu$ decay. Finally, since the t-channel is a sub-dominant process, the effect of this method on background processes is studied to set the framework for the mass measurement in the next chapter.

4.1 Motivation

The mass of the top quark, m_{top}, is a significant free parameter of the Standard Model. As the heaviest known particle, the top quark is related to other parameters of the SM through its large contributions to quantum corrections, such as the mass of the Higgs boson. Therefore, precise measurements of m_{top} are essential for testing the validity of the SM and potentially probing new physics beyond it. There is a wealth of experimental results for the measurement of m_{top} at the LHC, as shown in Figure 4.1, which presents a summary of ATLAS and CMS measurements from top quark decays. For each measurement, the statistical uncertainty and the sum of the remaining uncertainties are reported separately.

The majority of these measurements are based on the kinematic reconstruction of semileptonic final states of top quark-antiquark pair production, as it constitutes the dominant production mode. However, decay channels other than $t\bar{t}$ and different observables sensitive to m_{top}, are subject to different systematic uncertainties. Consistent measurements from different channels can lead to higher precision when combining the results and provide a better understanding of the systematic effects. As a result, measurements on complementary event topologies are necessary. Specifically, the t-channel in single top production can offer such partially independent event samples for measurements, as it occurs at lower energy scales.

A recent measurement at CMS in the t-channel using single top events, based on *pp* collision data corresponding to an integrated luminosity of 35.9 fb^{-1} at 13 TeV [32], reported

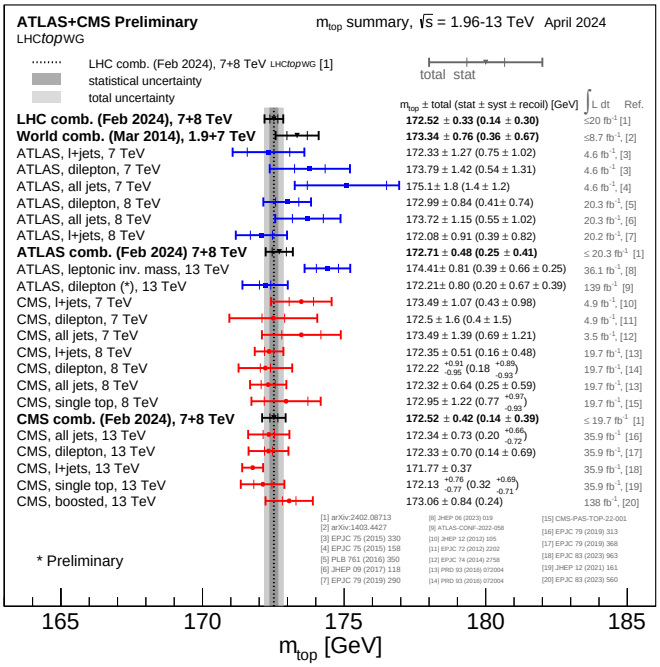


Figure 4.1: Summary of the ATLAS and CMS measurements of the top quark mass [6].

a top quark mass of:

$$m_{\text{top}}^{\text{t-ch}} = 172.13 \pm 0.32\ (\text{stat+prof})\ ^{+0.69}_{-0.71}\ (\text{ext})\ \text{GeV} = 172.13^{+0.76}_{-0.77}\ \text{GeV} \qquad (4.1)$$

Although this measurement, performed by reconstructing the top quark directly from its final products, serves as a solid first approximation with reduced errors, there is still room for improvement. Therefore, the primary goal of this analysis is to explore alternative reconstruction techniques and propose novel methods to achieve a more accurate measurement.

4.2 Event Topology

Due to its extremely short lifetime, the top quark can only be studied through the decay products in its final state. As a result, measuring the mass of the top quark using single top events at the t-channel requires a good understanding of the event topology. A detailed knowledge of the physical processes that share the same final states allows defining a kinematic phase-space region enriched with signal and suppressed background. A precise definition of the signal and background processes is provided below.

4.2.1 Signal Process

The signal process in this analysis is the t-channel production of single top quarks. The LO Feynman diagram for the t-channel final state is shown in Figure 4.2. The t-channel

signature includes an isolated, high-momentum charged lepton, a neutrino which leads to an overall transverse momentum imbalance, a light-quark jet produced in the forward direction and a jet from the hadronization of a b quark from the top quark decay. The second b jet, arising from gluon splitting, typically escapes detection due to its softer p_T spectrum and broader pseudorapidity η distribution compared to the b jet originating from the top quark. Consequently, the expected final state requires the identification of one electron or muon and exactly two jets, on of which is b-tagged.

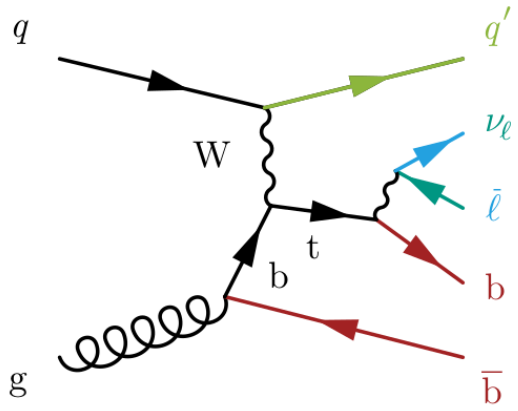


Figure 4.2: Feynman diagram of single top quark production in the t-channel final state.

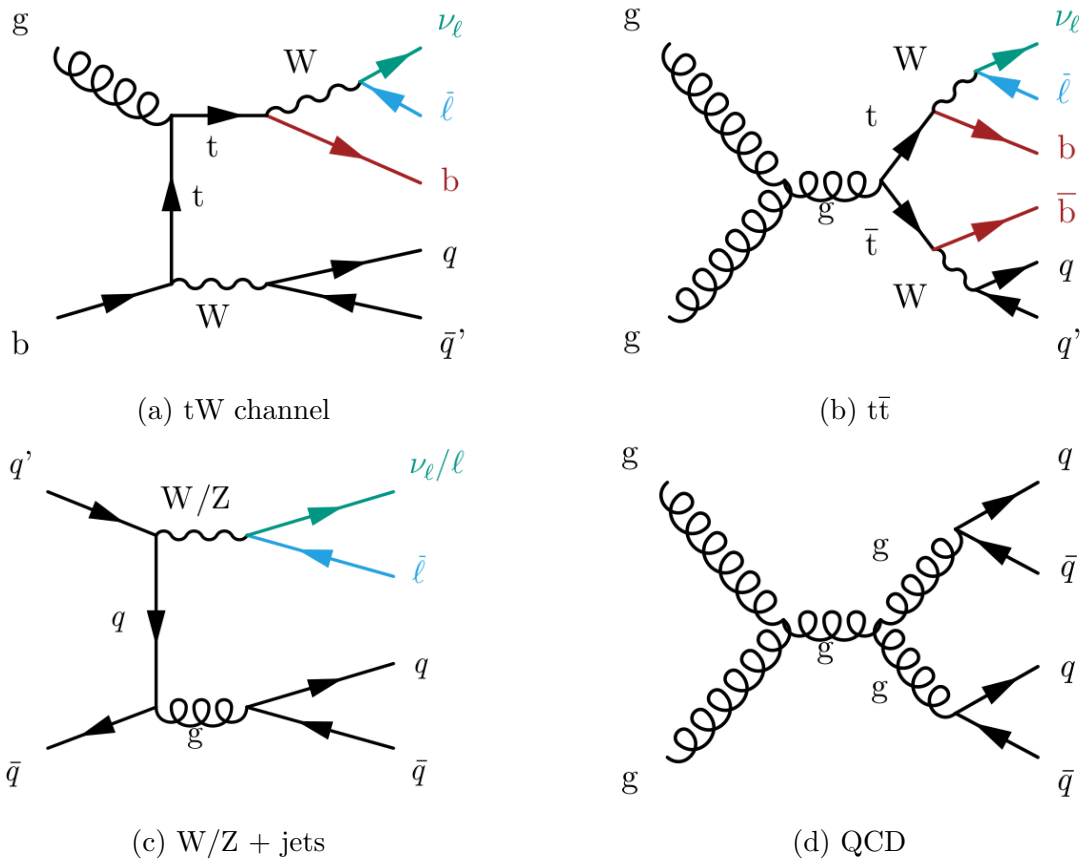
4.2.2 Background Processes

Additional Single Top Quark Channels

There are two additional subdominant production modes of single top quarks: associated production with a W boson and the s-channel. The tW-channel, shown in Figure 4.3a, has a cross section approximately three times smaller than that of the t-channel, but it can contribute to the background if one of the two W bosons decays hadronically and the other leptonically. In this case, the final state becomes identical to that of the signal process. However, since the preferred tW final state includes three jets, the requirement of exactly two jets in the event selection significantly reduces this background. On the other hand, the s-channel has a much smaller cross section, and its contributions are typically negligible in analyses.

Top Quark-Antiquark Pairs

Top quark-antiquark pair production consists the primary source of background in this analysis. As the dominant mode of top quark production, $\text{t}\bar{\text{t}}$ events often resemble the signal event topology. In particular, as shown in Figure 4.3b, one of the two top quarks produced must decay leptonically, resulting in a final state with a charged lepton, missing transverse momentum, two b-tagged jets and (likely) two additional jets. If one of these additional jets



(a) tW channel

(b) $t\bar{t}$

(c) W/Z + jets

(d) QCD

Figure 4.3: Feynman diagrams in the final states for various background processes.

escapes detection, the final state of semileptonic $t\bar{t}$ decay becomes indistinguishable from single top t-channel production. Nevertheless, by considering the number of jets in the event, this background is effectively reduced.

W and Z Bosons in Association with Jets

This background involves the production of a W or Z boson from a purely QCD process, along with an arbitrary number of jets from gluon splitting, as shown in Figure 4.3c. It has a cross section approximately an order of magnitude larger than that of single top production. In W+jets events the W boson may decay leptonically and if one of the produced jets is identified as a b-jet, the final state can resemble the signal. However, since these events do not involve a top quark, the resulting kinematic characteristics differ from those of single top events, leading to broader mass distributions in the reconstruction. In Z+jets events, there is no missing transverse momentum because the Z boson decays into a lepton-antilepton pair. While it is possible for one lepton to escape detection, mimicking the signal's final state, this process contributes negligibly to the background.

Diboson

The diboson background involves the production of two bosons (WW, WZ, or ZZ), often accompanied by additional jets. If one boson decays hadronically and the other leptonically, the final state can resemble that of single top production in the t-channel. However, because of the small cross section in comparison with the other backgrounds, diboson contributions are typically negligible.

QCD Multijet

The event topology of QCD multijet production, as shown in Figure 4.3d, is significantly different from that of the signal process. However, the large total cross section for QCD interactions makes it a considerable source of background. The final state consists of many jets originating from initial or final state radiation. Misidentification of jets can result in the production of "fake" leptons which may mimic the final state of single top t-channel events.

4.3 Event Selection

4.3.1 Monte Carlo Samples

Theoretical modeling of the distributions of various variables for signal and background processes is necessary for a comparison to the measured quantities at the detector and for extracting parameter estimations from the data. For this purpose, Monte Carlo (MC) generators are employed to simulate events from high-energy collisions. Table 4.1 presents the MC samples used in this analysis, along with their corresponding generators and cross sections. Each sample contains a number of generated events, which is subject to statistical fluctuations. As a result, an accurate description of the kinematic distributions requires a sufficiently large number of generated events within the relevant phase space.

4.3.2 Baseline Cuts

The events analyzed in this study are those that meet the baseline criteria applied to target the final state of the signal process. The selected events can be categorized based on the number of jets n and the number of b-tagged jets m (nJmT). The specific event category examined in this thesis is the 2J1T category, which contains exactly two well-reconstructed jets, with one of them being b-tagged, retaining only events that include isolated muons in the final state that fire the HL triggers for tight isolation (HLT_IsoMu27_v and HLT_IsoTkMu24_v). Overall, the baseline selection cuts encompass:

- muons with $p_T > 35$ GeV and $|\eta| < 2.1$
- light jets with $p_T > 40$ GeV and $|\eta| < 4.7$
- b-tagged jets with $p_T > 40$ GeV and $|\eta| < 2.4$

Process		Generator	σ [pb]
Single Top	top t-ch	PowhegV2	136.02
	anti-top t-ch	Powheg	80.95
Top-Antitop pair	$t\bar{t}$	”	832
W+Jets (NLO)	0 jets	Madgraph@NLO	49670
	1 jet	”	8264
	2 jets	”	2628
W+Jets (LO)	$H_T \in [70, 100]$ GeV	Madgraph	1353
	$H_T \in [100, 200]$ GeV	”	1346
	$H_T \in [200, 400]$ GeV	”	359.7
	$H_T \in [400, 600]$ GeV	”	48.91
	$H_T \in [600, 800]$ GeV	”	12.05
	$H_T \in [800, 1200]$ GeV	”	5.501
	$H_T \in [1200, 2500]$ GeV	”	1.329
	$H_T \in [2500, +\infty]$ GeV	”	0.03216

Table 4.1: List of MC samples used in this analysis with their corresponding generators and cross sections.

In addition, events were required to have zero non-isolated leptons and zero veto leptons. The b-tag score of the b-jet needed to exceed the tight working point (b-tag score > 0.9535, while the b-tag score for light jets had to be below the loose working point (b-tag score < 0.5426) to boost b-jet identification efficiency. Finally, a cut on the transverse mass of the W boson $m_T^W > 50$ GeV was also applied to eliminate contributions from QCD multijet production.

The parameter of interest from the MC samples is the the expected number of events N_{exp} in the selected phase space. Therefore, the number of events that pass the selection criteria, N_{pass}, needs to be adjusted by rescaling the integral of the distributions to match the expected yields using the integrated luminosity, L_{int}, and the number of generated events, N_{gen}, in the specific sample. By multiplying with the cross section of each process, the expected yields can be calculated as follows:

$$N_{exp}^{j} = \sum_{i \in j} \sigma_i \, L_{int} \, \frac{N_{pass}^{i}}{N_{gen}^{i}} \tag{4.2}$$

where L_{int} at 36.3 fb^{-1} is the integrated luminosity at the time the data sample was recorded. The index j runs over the signal and background processes, while the index i corresponds to the different samples used for a specific process.

4.4 Top Quark Reconstruction

The four-momentum of the top quark can be calculated from the kinematic information available in the event. The decay products of the top quark are a b-tagged jet and a W boson which decays further into a charged lepton and a neutrino. Nevertheless, the neutrino interacts solely via the weak interaction and therefore cannot be observed in the detector. The transverse momentum of the neutrino $p_{\nu,T}$ can be estimated from the event's missing transverse energy p_T^{miss}, which is usually fully attributed to the neutrino. However, the longitudinal component of the neutrino momentum $p_{\nu,z}$ remains completely unknown. For this reason, the constraint that the W boson is on-shell is imposed, with its mass fixed at the pole value of 80.4 GeV [3], to enable the derivation of an equation for determining the z-component of the neutrino's momentum. Assuming energy-momentum conservation at the $W \to l\nu$ vertex, the constraint can be expressed as:

$$m_W^2 \;=\; p_W^2 \;=\; (p_l + p_\nu)^2 \;=\; p_l^2 + p_\nu^2 + 2p_l p_\nu \tag{4.3}$$

and by neglecting the invariant masses of the lepton and of the neutrino:

$$\begin{aligned} m_W^2 \;&=\; 2p_l p_\nu \;=\; 2\cdot(E_l E_\nu - \vec{p_l}\vec{p_\nu}) && (4.4)\\ &=\; 2\cdot(E_l E_\nu - \vec{p}_{l,T}\,\vec{p}_{\nu,T} - p_{l,z}\,p_{\nu,z}) && (4.5)\\ &=\; 2\cdot(E_l E_\nu - p_{l,T}\,p_{\nu,T}\,cos\Delta\phi - p_{l,z}\,p_{\nu,z}) && (4.6) \end{aligned}$$

where $\Delta\phi$ is the azimuthal angle difference between the charged lepton and the neutrino. By expressing the energy of the massless neutrino in terms of its momentum components and by introducing the abbreviation as following:

$$E_\nu = \sqrt{p_{\nu,T}^2 + p_{\nu,z}^2} \qquad \text{and} \qquad \Lambda \;=\; \frac{m_W^2}{2} \;+\; p_{l,T}\,p_{\nu,T}\,cos\Delta\phi$$

a quadratic equation in $p_{z,\nu}$ can be obtained by squaring 4.6 and rearranging the terms:

$$E_l\sqrt{p_{\nu,T}^2 + p_{\nu,z}^2} \;=\; \Lambda + p_{l,z}\,p_{\nu,z} \;\;\Rightarrow\;\; E_l^2(p_{\nu,T}^2 + p_{\nu,z}^2) \;=\; (\Lambda + p_{l,z}\,p_{\nu,z})^2 \;\;\Rightarrow \tag{4.7}$$

$$\Rightarrow\;\; p_{\nu,z}^2 \;-\; 2\frac{\Lambda\,p_{l,z}}{E_l^2 - p_{l,z}^2}\cdot p_{\nu,z} + \frac{E_l^2\,p_{\nu,T}^2 - \Lambda^2}{E_l^2 - p_{l,z}^2} \;=\; 0 \tag{4.8}$$

By substituting $E_l^2 - p_{l,z}^2$ to $p_{l,T}^2$, the two solutions of equation 4.8 can be written as:

$$p_{\nu,z}^{\pm} \;=\; \frac{\Lambda\,p_{l,z}}{p_{l,T}^2} \;\pm\; \sqrt{\frac{\Lambda^2\,p_{l,z}^2}{p_{l,T}^4} \;-\; \frac{E_l^2 p_{\nu,T}^2 - \Lambda^2}{p_{l,T}^2}} \tag{4.9}$$

The solutions are either real or complex depending on the sign of the radicand (Δ):

- If the radicand is positive, there are two real solutions for $p_{\nu,z}$. Typically, the solution with the smaller absolute value $|p_{\nu,z}|$ is preferred because it provides a better description of the steeply falling distribution of $p_{\nu,z}$. This choice is correct approximately 65% of the time.

- If the radicand is negative, the solutions for $p_{\nu,z}$ are complex. Although not physically acceptable, this case occurs approximately 35% of the time and therefore the imaginary solutions cannot be simply disregarded. The standard treatment is to modify the values of $p_{\nu,T}$ minimally to nullify the radicand, thus leading to a unique solution for $p_{\nu,z}$. This case is explained in detail below.

The negative case

Typically, assuming the leptons are accurately measured, the negative value of the radicand is attributed to poor reconstruction of $p_{\nu,x}$ and $p_{\nu,y}$ due to imperfect resolution of the missing transverse energy. As a result, the transverse momentum of the neutrino should not be directly linked to p_T^{miss}. To avoid the complications of complex solutions for $p_{\nu,z}$, the components of $\vec{p}_{\nu,T}$ are modified accordingly to ensure $\Delta = 0$. This requirement leads to a quadratic relation of the two transverse components of neutrino momentum:

$$p_{\nu,y}^{\pm}(p_{\nu,x}) \;=\; \frac{m_W^2 p_{l,y} + 2p_{l,x}\, p_{l,y}\, p_{\nu,x}}{2p_{l,x}^2} \;\pm\; \frac{m_W\, p_{l,T}}{2p_{l,x}^2}\sqrt{m_W^2 + 4p_{l,x}\, p_{\nu,x}} \tag{4.10}$$

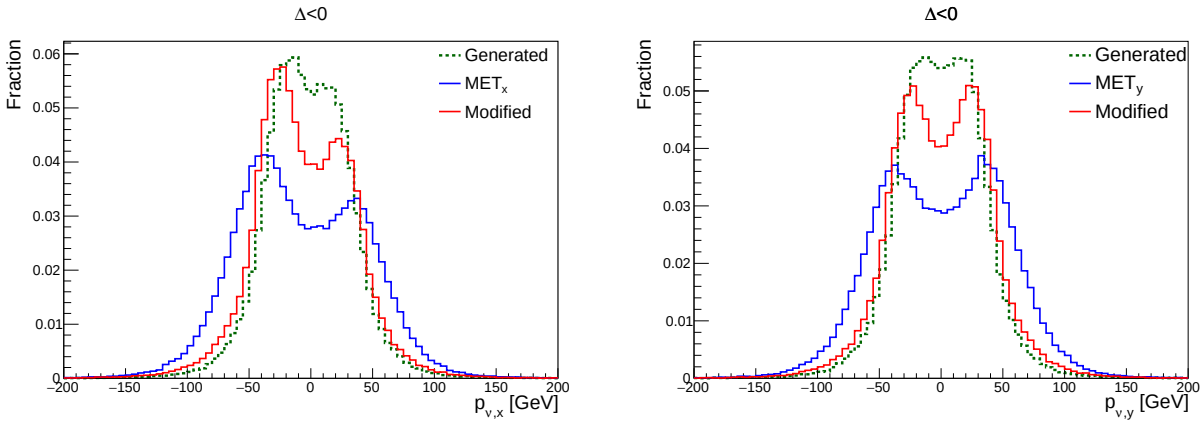
The measured p_T^{miss}, however, can still be considered a reasonably good estimate. Consequently, the modified values $p_{\nu,x}$ and $p_{\nu,y}$ should be as close as possible to the respective measured values $p_{T,x}^{\text{miss}}$ and $p_{T,y}^{\text{miss}}$. For this reason, quantity δ is defined:

$$\delta^{\pm} \;=\; \sqrt{\left(p_{T,x}^{\text{miss}} - p_{\nu,x}\right)^2 \;+\; \left(p_{T,y}^{\text{miss}} - p_{\nu,y}^{\pm}\right)^2} \tag{4.11}$$

In practice, since $p_{\nu,y}$ can be expressed as a function of $p_{\nu,x}$, one iteratively scans over a range of $p_{\nu,x}$ values that keep the radicand equal to zero, searching for the value that minimizes the quantity δ. As there are two possible solutions $p_{\nu,y}^{\pm}$, that satisfy this condition, two corresponding distance functions $\delta^{\pm}$ are minimized. The solution with the smaller absolute value of δ is preferred, as it results in the smallest deviation from the measured p_T^{miss}. The effects of this treatment are clearly illustrated in Figure 4.4. A comparison between the distributions obtained directly from p_T^{miss} and those after modification shows the latter better describes the generated neutrino momentum distributions. The tendency to retrieve smaller values for $p_{\nu,x}$ and $p_{\nu,y}$ after the modification suggests that only a portion of p_T^{miss} is attributed to the neutrino's transverse momentum.

Thus, by construction, the longitudinal component $p_{\nu,z}$ becomes equal to $\Lambda p_{l,z}/p_{l,T}^2$ from equation 4.9, thereby fully defining the complete neutrino momentum that satisfies this equation. With all the essential elements in place, namely the four-vectors of the b-tagged jet, the charged lepton, and the neutrino, the four-vector of the top quark, and hence its mass,

can be determined. This analytic method of reconstructing the top quark will be referred to as "Traditional" throughout this thesis and will serve as a baseline for comparison with the new method introduced next.



(a) (b)

Figure 4.4: Comparison of the transverse momentum components of the neutrino, (a) $p_{\nu,x}$ and (b) $p_{\nu,y}$, between generated and reconstructed values, for the case of a negative radicand.

4.5 Regressed Kinematic Fit Method

While standard analytical techniques provide reasonable estimates, they introduce inconsistencies due to the fixed constraint imposed on the W boson mass. Consequently, there is potential for further refinement and improvement. In this analysis, a novel method, referred to as Regressed Kinematic Fit (KinFitReg), is proposed to enhance the reconstruction of the top quark. The method involves two steps: first, a regression is used to predict the longitudinal component of the neutrino momentum, followed by a chi-square minimization to estimate all components of the neutrino momentum.

4.5.1 Neutrino Longitudinal Momentum Regression

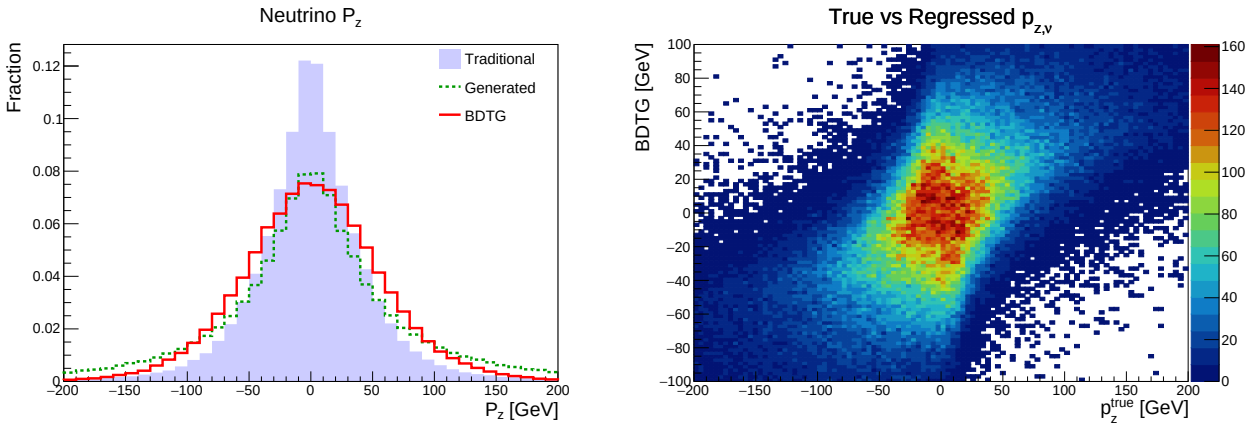
The longitudinal component of the neutrino momentum $p_{z,\nu}$ is initially estimated using a Boosted Decision Tree (BDT) for regression. The input variables for the BDT include the kinematic properties of the b-jet, the light jet, the lepton, and their combinations. These variables are selected based on their high correlation with the target value, which is the true $p_{z,\nu}^{\text{true}}$ generated from simulations. A list of the 22 variables is provided in Table 4.2. The large number of variables does not affect the performance of the regression, as decision trees are insensitive to the inclusion of input variables with low discriminative power.

The regression is implemented through the TMVA package [33] with gradient boosting, combining in total 100 decision trees with a depth of 4 layers into a forest. The shrinkage

Variable	Description
p_T^{miss}	Missing transverse energy (MET)
$\Delta\phi$(lep,met)	Azimuthal angle difference between lepton and MET
η_l, η_b, $\eta_{j'}$	Pseudorapidity of the lepton, b-tagged and light jet
p_T^l, $p_T^{j'}$	Transverse momentum of the lepton and light jet
$\Delta R_{l,j'}$	Angular separation in (η,ϕ) space between lepton and light jet
$\Delta\eta_{lj}$, $\Delta\eta_{lb}$, $\Delta\eta_{bj}$	Pseudorapidity difference between lepton,b-tagged and light jet
m_{lj}, m_{lb}, m_{jb}	Invariant mass of systems comprising the lepton, b-tagged and light jet
p_T^{lb}, p_T^{lj}, p_T^{jb}	Scalar sum of the p_T of the system of lepton, b-tagged and light jet
E_{lb}, E_{lj}, E_{jb}	Ratios of lepton, b-tagged and light jet energies
E_l	Energy of the lepton
$m_{j'}$	Mass of the light jet

Table 4.2: List of BDT input variables

parameter was set to 0.1. The outcome of the regression is illustrated in Figure 4.5a by comparing the two approaches, the traditional and the regression. It is evident that the BDT can provide an estimate for $p_{z,\nu}$ describing more accurately the generated distribution. The effectiveness of the BDT regression is also indicated in Figure 4.5b , which shows a positive correlation between the true and the regressed $p_{z,\nu}$.



(a) (b)

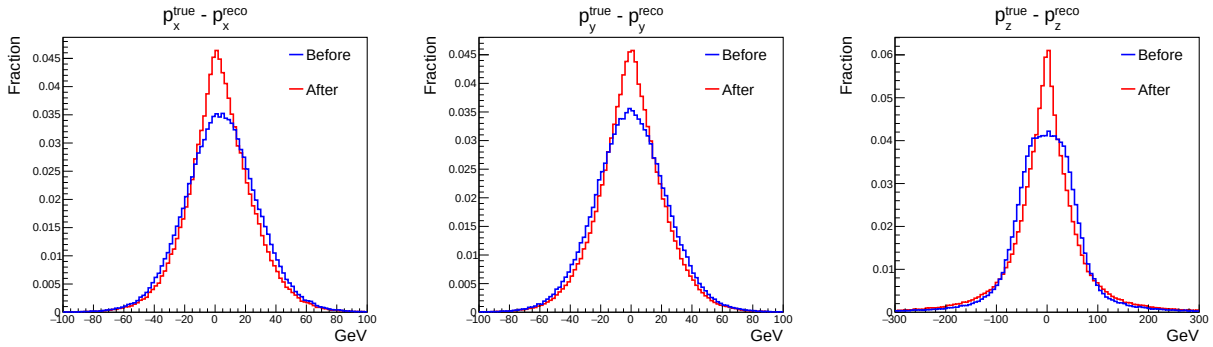
Figure 4.5: (a) Comparison of traditional and regression approaches for estimating the longitudinal neutrino momentum and (b) correlation between true and regressed neutrino $p_{z,\nu}$.

4.5.2 Chi-Square Minimization

The next step is to estimate all components of the neutrino momentum simultaneously through kinematic fitting. The separate estimates for the transverse momentum are already obtained from p_T^{miss} measurements, and the longitudinal momentum is estimated from the regression. These are then provided as inputs for the minimization of a chi-square, which applies the W boson mass constraint as follows:

$$\chi^2(\tilde{p}_x^{(\nu)}, \tilde{p}_y^{(\nu)}, \tilde{p}_z^{(\nu)}) = \left(\frac{m_w^{\text{reco}} - m_w^{\text{true}}}{\Gamma w}\right)^2 + \sum_{i\in\{x,y,z\}} \left(\frac{\tilde{p}_i^{(\nu)} - p_i^{(\nu)}}{\sigma_i}\right)^2 \tag{4.12}$$

The σ values essentially indicate the allowable deviation from the measured values. The value of Γ_w is 1.99 GeV, determined by fitting a Breit-Wigner distribution to the W boson at parton level, while the parameters $\sigma_x = 22.38$, $\sigma_y = 22.48$ and $\sigma_z = 72.62$ (GeV) were obtained by fitting a Gaussian distribution to the differences between the true and reconstructed momentum components, $p_{\text{i}}^{\text{true}} - p_{\text{i}}^{\text{reco}}$, as shown in blue in Figure 4.6 for each momentum component. By penalizing significant discrepancies, this approach reduces the differences, leading to more consistent predictions for each momentum component after the chi-square minimization.



(a) (b) (c)

Figure 4.6: (a)-(c) Comparison of the differences between the generated and reconstructed values of each component of the neutrino momentum, before and after the minimization of the chi-square.

The improvement in neutrino reconstruction impacts the mass distribution of the top quark. In Figure 4.7, this effect is visible by comparing different methods. Overall, the kinematic fit combined with the regression substantially improves resolution, resulting in a higher peak while reducing the high tails caused by improper top quark reconstruction, although a remaining bias towards lower m_{top} values is still evident, as indicated by the vertical dashed line. The superiority of KinFitReg lies in the fact that the solution for the neutrino momentum is not imposed but inferred from initial estimates, thus providing a more natural shape for the W boson mass compared to existing standard techniques.

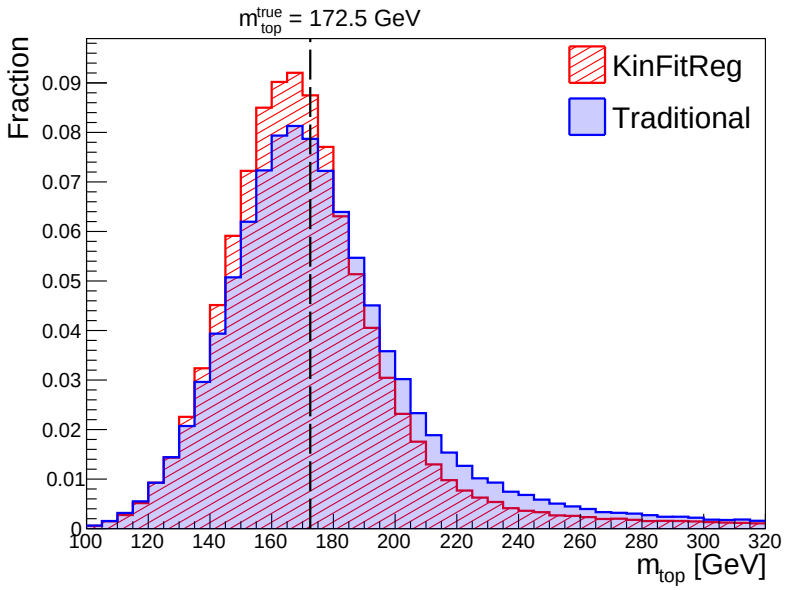


Figure 4.7: Top quark mass distributions using the Traditional method and KinFitReg.

4.5.3 B-jet Energy Scale Regression

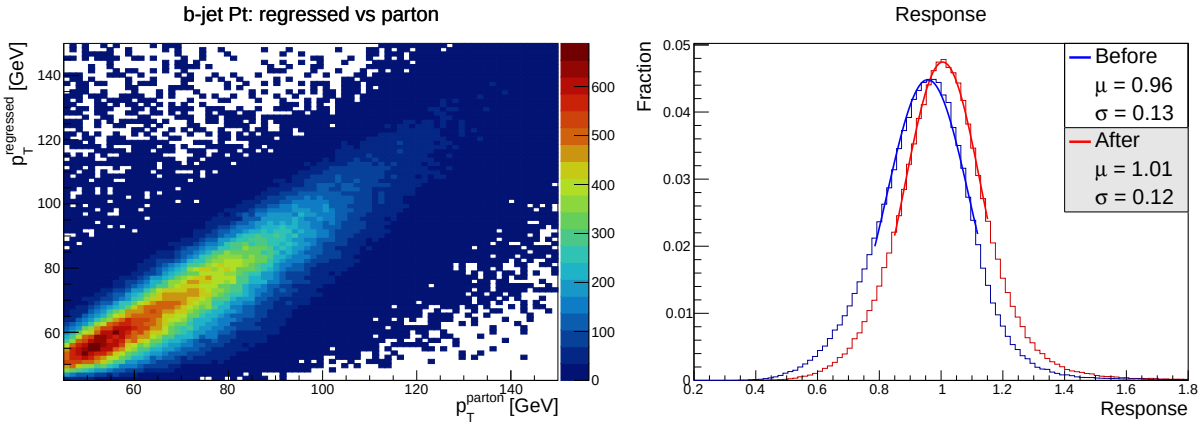
The top quark decay results in a W boson and a b quark, with each receiving roughly half of the top quark's initial energy. In this way, the showering and hadronization initiated by the b quark typically consist of the most energetic product in an event, thereby contributing significantly to the top quark mass distribution. Generally, jet-energy calibration in CMS is done on average in terms of p_T, without taking into account the difference of flavour [34]. For instance, jets with high charged-hadron fraction are expected to have a higher response than those with low charged-hadron fraction. In addition, a significant fraction of b-tagged jets lose energy to neutrinos from B meson leptonic decays, and since the corrections target only the visible energy, this leads to energy loss and a lower response. B-tagged jets are sytematically under-calibrated up to 10%.

To account for the missing energy, a separate regression is employed similar to the one for the neutrino but this time targeted to the b-jet p_T at parton level. This regression analyzes the energy distribution of particles within the b-tagged jets to improve the estimation for the transverse momentum of the original b quark. The regression is implemented again through the TMVA package and the method is BDT with gradient boosting, combining in total 100 decision trees with a depth of 3 layers into a forest and shrinkage parameter equal to 0.1. The training process was conducted using a sample of single top events, where the reconstructed b-tagged jet is matched to the generated b-jet using a ΔR(reco,gen)<0.4 criterion. The input variables for the regression are listed at Table 4.3 below.

The outcome of the regression is visualised in Figure 4.8a, which shows a linear correlation between the regressed transverse momentum of the b-tagged jet and the generated momentum at the parton level. Moreover, Figure 4.8b illustrates the response of the re-

Variable	Description
p_T^b	Missing transverse momentum of the b-jet
η_b	Pseudorapidity of the b-jet
ϕ_b	Azimuthal angle of the b-jet
m_b	Invariant mass of the b-jet
b-tag	B-tagging score of the b-jet
f_{ch}^b	Charged hadron p_T fraction
f_{neu}^b	Neutral hadron p_T fraction
f_{ph}^b	Photon p_T fraction
f_{mu}^b	Muon p_T fraction
f_e^b	Electron p_T fraction

Table 4.3: List of BDT input variables

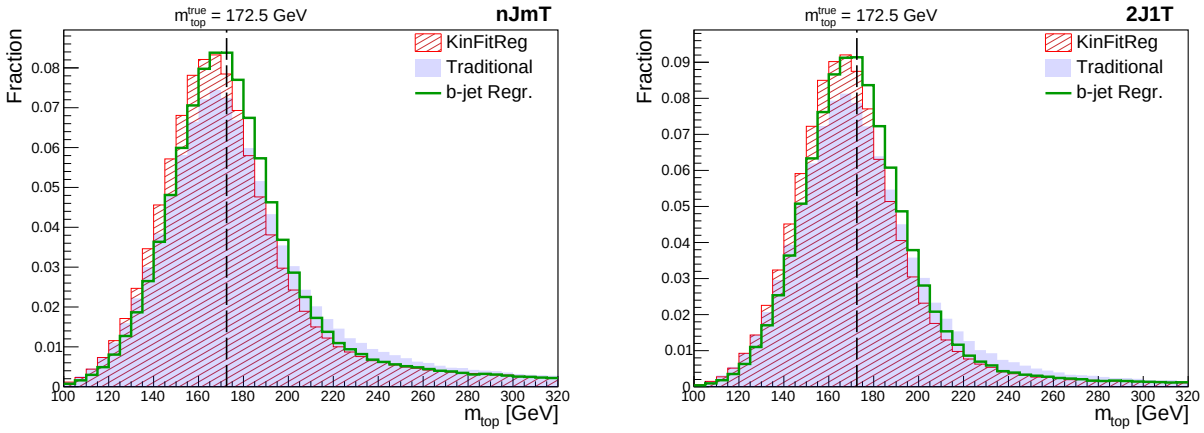


(a) (b)

Figure 4.8: (a) Correlation between the generated and regressed b-jet p_T and (b) the response of the reconstructed p_T before and after the regression.

constructed p_T before and after the regression. After the regression, the response forms a Gaussian distribution with a mean of 1 and standard deviation of 0.12, thus successfully regaining the missing energy from the initial under-calibration.

The corrected values of the b-jet p_T have an impact on the final top quark mass distribution. In Figure 4.9, the mass distributions of the top quark are compared using the method KinFitReg, with and without the regression applied to the b-tagged jet, in nJmT and 2J1T event categories. While preserving the improved resolution, the regression aids in recovering any existing residual biases in distributions, thus enhancing the overall accuracy of the reconstruction in both cases.



(a) (b)

Figure 4.9: The mass distributions of the top quark using the Traditional method and the KinFitReg method, with and without applying regression to the b-jet p_T in (a) nJmT and (b) 2J1T event categories.

4.6 Background Suppression

The single top t-channel process is a subdominant mechanism for top quark production. To enhance signal and background separation, BDTs are employed as multivariate analysis (MVA) discriminators for classification. Since single top antiquarks can also be produced in collisions, they are included as part of the signal as well. For the background processes, only top quark pair $t\bar{t}$ production and W+jets events are considered. Other background processes, such as single top tW and s-channels, Z+jets, dibosons, and QCD multijets, are effectively suppressed by the baseline selection criteria and are thus safely neglected. However, these assumptions should be reevaluated if discrepancies arise between data and Monte Carlo predictions.

The classification was implemented with gradient boosting, combining a total of 400 trees into a forest and shrinkage parameter equal to 0.2. For the W+jets process, leading-order MC samples were used for BDT training to avoid the use of negative event weights in the training process. However, the modeling of this background in the analysis is based on NLO samples because of their considerable contributions. To account for the difference in cross sections between the LO and NLO samples, a scale factor of k=1.78 was applied, as explained in detail in Appendix A. A list with the selected input variables for the BDT is presented in Table 4.4. These variables offer strong discrimination power for different processes. Two key variables unique to the single top t-channel are:

1. The absolute pseudorapidity of the light jet ($|\eta_j|$). The spectator quark originating from one of the interacting protons has its direction subtly altered during the scattering process. This leads to very forward jets with large pseudorapidity, which is characteristic of single top t-channel events.

2. The cosine of the angle between the light jet and the lepton in the top quark rest frame ($\cos\theta^*$). This variable takes advantage of the fact that in single top production, top quarks are highly polarized due to the weak interaction involved. This polarization translates into specific spin configurations in the final state particles, affecting their angular distributions. In contrast, this effect is absent in top quark-antiquark pair production, which is primarily driven by QCD.

Variable	Description
dR_{bj}	Angular separation between the b-jet and the light jet
m_T^W	Transverse mass of the W boson
$\lvert\eta_j\rvert$	Absolute pseudorapidity of the light jet
$\lvert\eta_l\rvert$	Absolute pseudorapidity of the lepton
m_{jb}	Invariant mass of the b-jet and the light jet
$\lvert\Delta\eta_{lb}\rvert$	Absolute pseudorapidity difference between the lepton and the b-jet
p_T^{jb}	Scalar sum of the transverse momenta of the b-jet and the light jet
$\cos\theta^*$	Cosine of the angle between the light jet and the lepton in the top quark RF

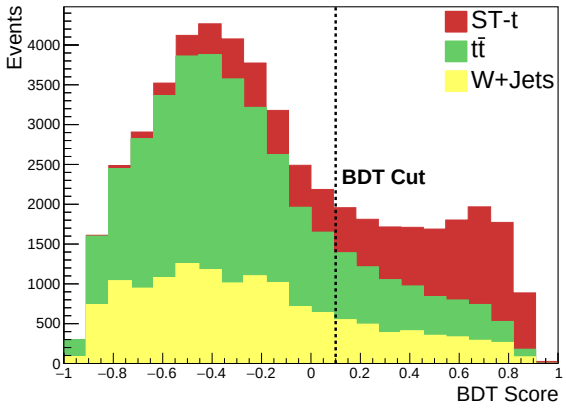
Table 4.4: List of BDT input variables

An important consideration when selecting variables is their correlation with the top quark mass. High correlations can shape the mass distribution after applying the BDT cut in the reconstructed mass. This is not desired, since the goal is to discriminate signal from background, which becomes challenging if the mass distributions are too similar. Therefore, it is crucial to select an orthogonal set of variables to avoid this issue. As a result, top mass ingredients, such as the p_T of the b-jet, which greatly influences the mass reconstruction, will be dismissed.

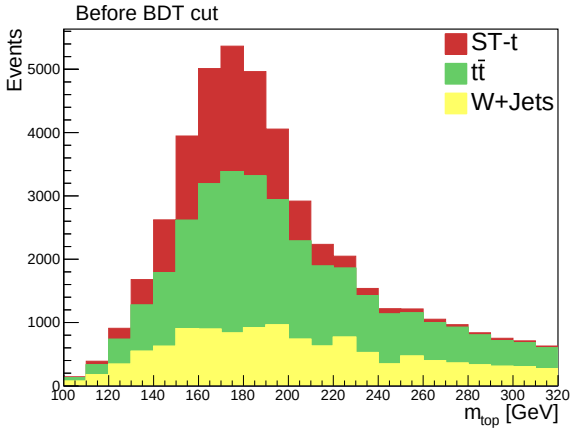
The discriminator assigns a BDT score on a event-by-event basis to classify signal and background events. The BDT response is illustrated in Figure 4.10a as a stacked histogram, where the number of expected events is normalized according to the cross-section of each process. Signal events receive higher BDT scores compared to background events. By applying a threshold on the BDT score, the classifier is able to separate the signal (shown in red, on the right) from the background (shown in green and yellow, on the left). The selection of the exact value of the BDT cut will be based on its impact on the error in the fitting procedure used to measure the mass of the top quark in the subsequent chapter.

In Figure 4.10b, the top mass distribution is shown before applying the BDT cut. As expected, the $t\bar{t}$ process is the dominant source of background and forms a peak in the region of interest, as this process includes the production of a top quark and is reducible to a single top. In contrast, the W+jets background exhibits a broad distribution since no top quark is involved. After applying the BDT threshold, Figure 4.10c shows that the background processes are effectively suppressed while preserving their shape. In this case, the larger tails dominated by background will be useful for the measurement, as they help anchor the fit and

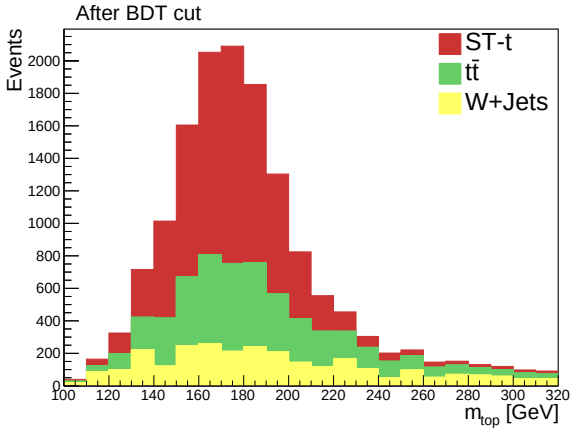
accurately define the background. Overall, the application of the BDT significantly enhances the signal purity.



(a)



(b)



(c)

Figure 4.10: (a) The response of the BDT as a stacked histogram and top mass distributions (b) before and (c) after the BDT cut.

Chapter 5

Top Quark Mass Measurement

5.1 Toy Monte Carlo Study

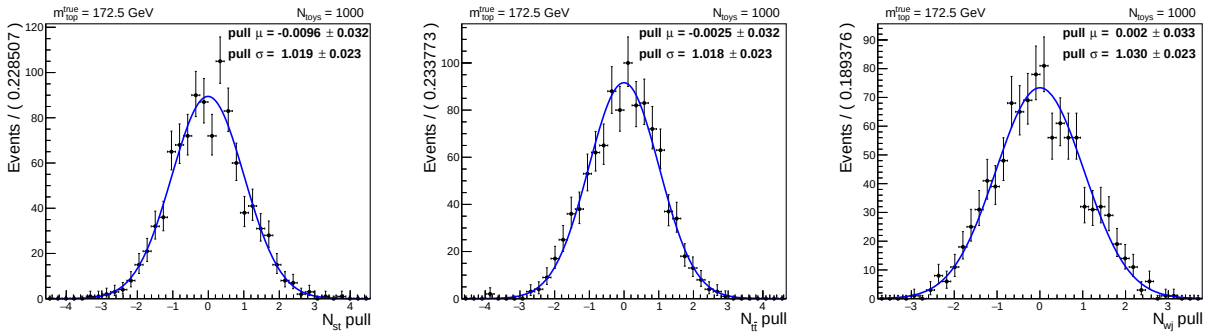
The estimation of the contributions that different processes have in an observed dataset D is achieved by exploiting the shape differences in the mass distributions. These distributions, referred to as templates, are obtained from the MC simulations. The number of expected events of each template are interpreted as model parameters and are adjusted to optimize the agreement between the data and the sum of the templates. Denoting the reconstructed top quark mass as x, the fit model is given by:

$$D(x) = N_{\mathrm{st}} F_{\mathrm{st}}(x; m_{\mathrm{top}}^{\mathrm{true}}) + N_{\mathrm{t\bar{t}}} F_{t\bar{t}}(x; m_{\mathrm{top}}^{\mathrm{true}}) + N_{\mathrm{wj}} F_{\mathrm{wj}}(x) \tag{5.1}$$

where F_{st}, $F_{t\bar{t}}$ and F_{wj} represent the probability density functions for the mass distribution of the corresponding process. Since the associated production of a W boson with jets does not involve a top quark, F_{wj} does not depend on the assumed true top mass in the simulation.

To validate the fit model and ensure it produces meaningful results, a large number of pseudo-data sets were generated from the model, and, then, a ML fit was performed on each toy dataset to evaluate the stability of the parameter estimates. Specifically, by allowing nuisance parameters to vary from 0.2 to 2 times their initial values and following a Poisson distribution, 1000 toys were generated. Since the pseudodata sample is large, the data are binned to increase computational efficiency when calculating the likelihood function in the fitting process, commonly referred to as (Extended) Binned Maximum Likelihood. The fit range is chosen to include both the mass-peak region and the sidebands, allowing both signal and background contributions to be well determined. Additionally, the bin size is chosen at 10 Gev so that no relevant information regarding the signal is lost; it is large enough to minimize statistical fluctuations without obscuring the discrimination of the different processes.

The most prominent tool to evaluate the performance of the toys is the analysis of pull distribution, namely the difference between the estimated parameters and their initial values divided by the estimated uncertainty, as shown in Figure 5.1. For estimators with Gaussian uncertainties, the distributions of the pulls are Gaussian distributions themselves, centered at zero with a standard deviation of one. In this way, it is confirmed that no significant biases



(a) (b) (c)

Figure 5.1: Pull distributions of the estimated parameters for 1000 simulated toy experiments.

are introduced and the initial parameters are retrieved successfully. A visual inspection of the estimated distribution of the top quark mass alongside the pseudodata points generated for a specific toy, as illustrated in Figure 5.2, can support this conclusion. As expected, the values of the estimated parameters are equal to the initial parameters within uncertainties demonstrating the desired behavior of the toy model.

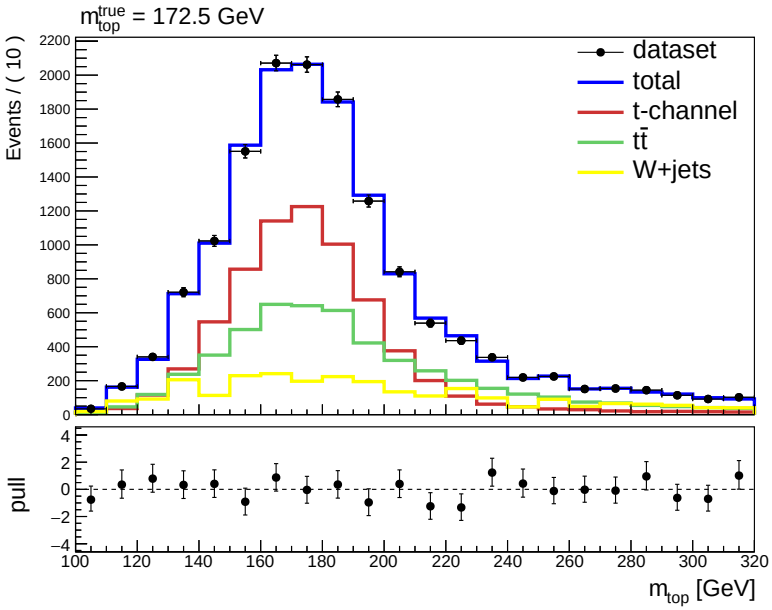
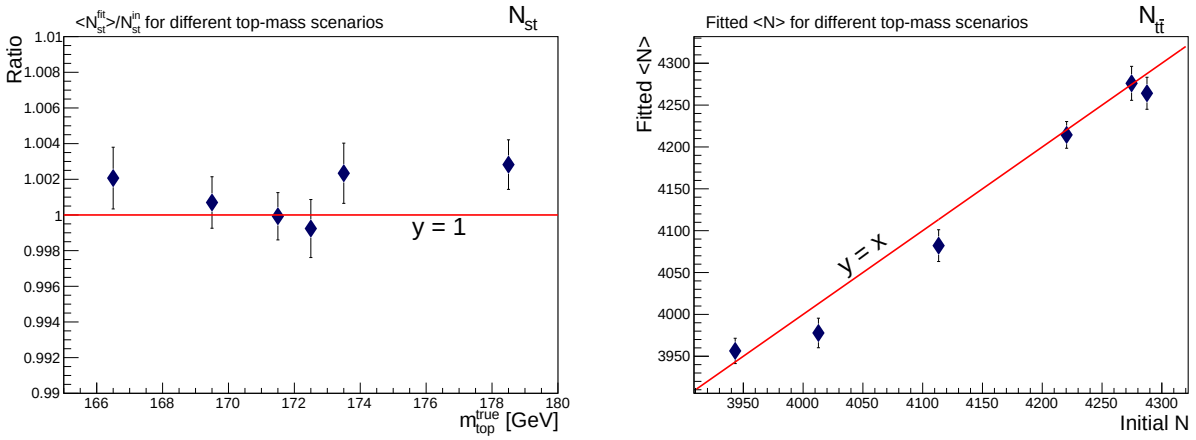


Figure 5.2: Estimated top quark mass distribution with pseudodata for a specific toy.

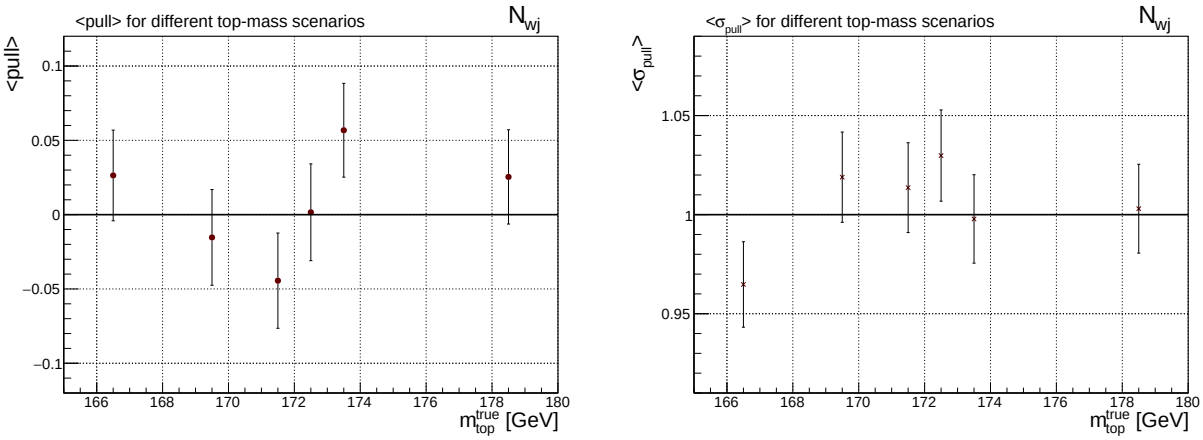
To account for the dependence of the simulated samples on the mass of the top quark, which is the observable being measured, multiple samples were used, each assuming a different top mass. In particular, the generated masses for the five additional samples were $m_{top} = \{166.5, 169.5, 171.5, 173.5, 178.5\}$ GeV. By repeating the procedure of generating 1000 pseudodatasets for each sample, the same ML fitting is performed on each toy, and the

pull distributions of the fitted parameters and their errors are obtained to evaluate possible biases.



(a) (b)

Figure 5.3: Linearity check: (a) The ratio of the fitted parameter N_{st} to the initial value for different top-mass samples, and (b) the fitted $N_{t\bar{t}}$ as a function of the initial values.



(a) (b)

Figure 5.4: Mean of (a) the pull and (b) standard deviation of the fitted parameter N_{wj} as a function of m_{top} for different top-mass scenarios.

For the linearity check, Figure 5.3a shows the ratio of the average fitted parameter N_{st} to the initial value for each mass sample, with results close to unity with precision at the thousandth decimal place. Furthermore, in Figure 5.3b, the expectation values of the fitted parameter $N_{t\bar{t}}$ are also plotted as a function of the initial parameters, using the standard deviation of the averages as the uncertainty. It is evident that the estimators are unbiased,

since they exhibit a linear trend, with a slope of one and no offset from zero. Additionally, Figure 5.4 shows the average pull distribution on the left with its standard deviation on the right for the N_{wj} parameter. Once again, the average pull is close to zero within errors and the standard deviation is close to one. Overall, this extensive study ensures the robustness of the fit model and the reliability of the top mass estimation across different top mass scenarios.

5.2 Top Quark Mass Extraction

With a set of six fit models available covering a mass range of 12 GeV, it is feasible to estimate the mass of the top quark by performing a likelihood scan. Specifically, a given dataset can be fitted to each model, and then a search through the calculated likelihoods can be conducted to identify the model with the minimum value. Afterward, a third-order polynomial is fitted to the likelihood distribution, and, next, a parabolic approximation is made around this minimum. The estimated mass value corresponds to the minimum of the parabola, while the statistical error is determined by the width of the parabola. It is worth noting that the mass samples cover a wide range to achieve a parabolic shape and hence achieve sub-GeV measurement accuracy. Typically, more mass samples would be needed; however, in their absence, the mass templates are interpolated with a step of 0.25 GeV by assuming a linear relation between two neighboring samples to cover the area around the minimum.

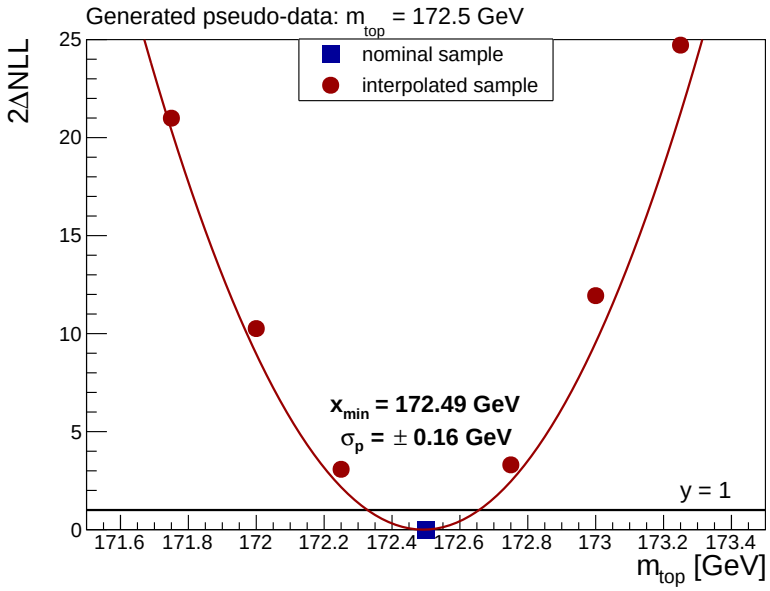


Figure 5.5: Toy experiment of measuring the mass of the top quark.

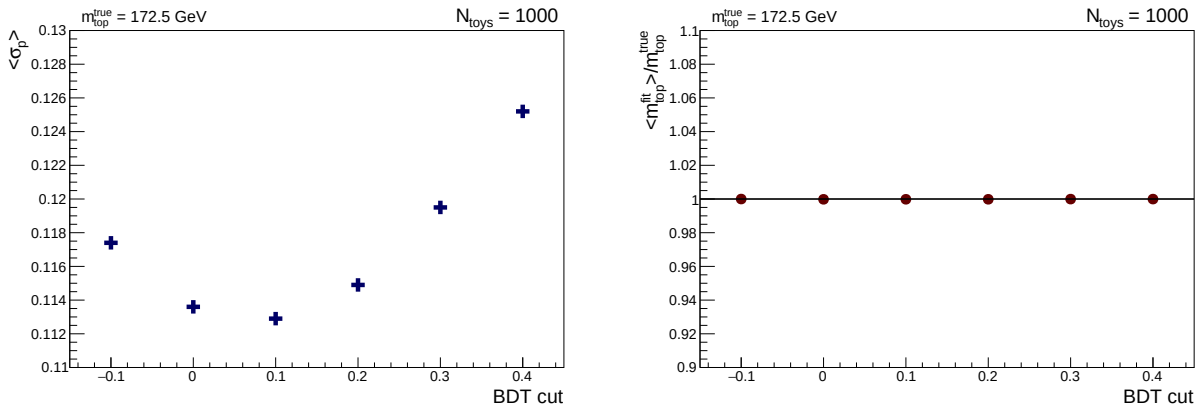
As a closure test, a sample with an assumed mass of 172.5 GeV is used to generate a pseudodata set. In this toy experiment, the same algorithm used to extract the mass from

real data is employed to verify that the model most compatible with the pseudodata is indeed the one used to generate it. The result is presented above in Figure 5.5, where the metric used is the negative log-likelihood multiplied by two, with the smallest value set to zero for reference. The value of the minimum is found at 174.49 GeV, and the width of the parabola, which is determined in this case by going one unit above the minimum, is by construction symmetrical and found to be ± 0.16 GeV. This toy experiment favors the desired mass sample and accurately estimates the value within the statistical uncertainties, thus confirming the validity of this methodology for extracting the mass.

5.3 BDT Cut Selection

As mentioned in the previous chapter, the BDT classifier separates signal from background by applying a specific cut to the assigned BDT score on an event-by-event basis. In this analysis, the exact selection of the cut is based on its impact on the error in the measurement procedure. In other words, the BDT cut that yields the narrowest parabolic shape of the likelihood distribution, and thus the smallest width, is chosen. To avoid personal biases, a blind analysis method is applied, meaning the BDT threshold is determined in advance, before being applied to the data.

The optimization process involves generating 1000 toy datasets using the fit model of the nominal mass sample at 172.5 GeV for a specific BDT cut, followed by performing a likelihood scan to calculate the width of the parabola σ_p. This procedure is then repeated for a range of different BDT cuts, and the one yielding the smallest average σ_p is selected. As Figure 5.6a suggests, the BDT cut should be set to 0.1 to maximize accuracy, with an average $<\sigma_p>$=0.113. In any case, as shown in Figure 5.6b, the generated mass is, on average, accurately retrieved regardless of the BDT cut selection.



(a) (b)

Figure 5.6: (a) The average mass measurement error and (b) the ratio of the measured mass to the generated m_{top} for 1000 toys across various BDT cuts.

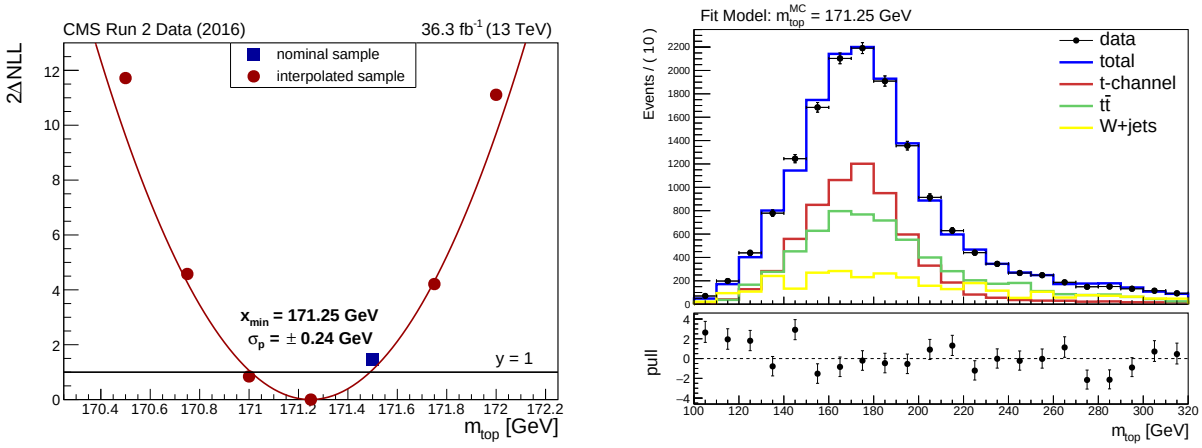
5.4 Results

In this analysis, the CMS Run 2 dataset from 2016, recorded at a center-of-mass energy of 13 TeV, is used. The dataset corresponds to an integrated luminosity of 36.3 fb^{-1}, with a 2.5% uncertainty. The developed algorithm has been tested and shown to function correctly with MC samples in extracting the mass of the top quark. The next step is its application to real data.

The measured value of top quark mass, m_{top}, using events from the t-channel single top quark process, inclusive of the lepton charge in the final state, is:

$$m_{top}^{\text{t-ch}} = 171.25 \pm 0.24 \text{ (stat+prof) GeV} \tag{5.2}$$

The result is consistent with previous measurements by the ATLAS and CMS within given uncertainties. No systematic uncertainties were studied in this analysis. It is noteworthy, nevertheless, that the statistical uncertainties, when compared to the CMS measurement using the same data sample presented in 4.1, are reduced by 25%. This is a significant advancement toward the initial goal, which is minimizing uncertainties. Improving the resolution in the reconstructed mass distribution through the KinFitReg method introduced in this analysis results in reduced statistical errors, which will unambiguously lead to a decrease in systematic uncertainties as well. Consequently, this method is very promising for achieving a more accurate measurement of the top quark mass.



(a) (b)

Figure 5.7: (a) Estimation of the top quark mass through a likelihood scan. The parabolic approximation indicates a minimum at 171.25 GeV, with a symmetric uncertainty of ±0.24 GeV. (b) Fitting the data to the model with an assumed top quark mass of 171.25 GeV.

A visual inspection of the fitted data to the model with an assumed top quark mass of 171.25 GeV is provided in Figure 5.7b, along with the various subprocesses. Despite the accurate description of the data by the fit model in the region of the peak, it can be observed that in the lower mass region, the pull distributions are higher, reaching up to 3σ. Specifically,

the total number of expected events is considerably smaller than the number observed in the data. This suggests that at lower energies, processes such as the QCD background, which were ignored in the Toy Monte Carlo studies, may be present. Significant contributions from these neglected backgrounds could also explain the apparent increase in the $t\bar{t}$ background, as the fitting procedure mistakenly attributes the excess events to top quark-antiquark pair production.

This is rigorously demonstrated in the Table 5.1 showing the initial values of the free parameters along with their post-fit values, their errors, and the respective pull values. Remarkably, significant discrepancies are observed by the pull values of 4.06 for the $t\bar{t}$ background and -2.35 for the single top process. Consequently, further investigation into the remaining backgrounds should be conducted to determine their impact on the final result.

	Initial	Fitted	Error	Pull
N_{st}	7342	6522	350	-2.35
$N_{t\bar{t}}$	3978	6142	533	4.06
N_{wj}	2884	2981	270	0.36

Table 5.1: Pre-fit and post-fit values of the free parameters, their errors and pull values.

Pre-Fit and Post-Fit Comparison

In what follows, several key kinematic variables are presented pre-fit and post-fit for comparison, as shown in Figures 5.8, 5.9 and 5.10. Since no sample is available at the measured mass value, the nominal sample with an assumed top quark mass of 171.5 GeV is used to compare the data with Monte Carlo predictions. The ratio of MC to data at lower energies is consistently below unity, strongly suggesting the presence of a low p_T background not accounted for in the analysis. Nevertheless, it is evident that the post-fit MC expectations agree with the observed distributions from the data.

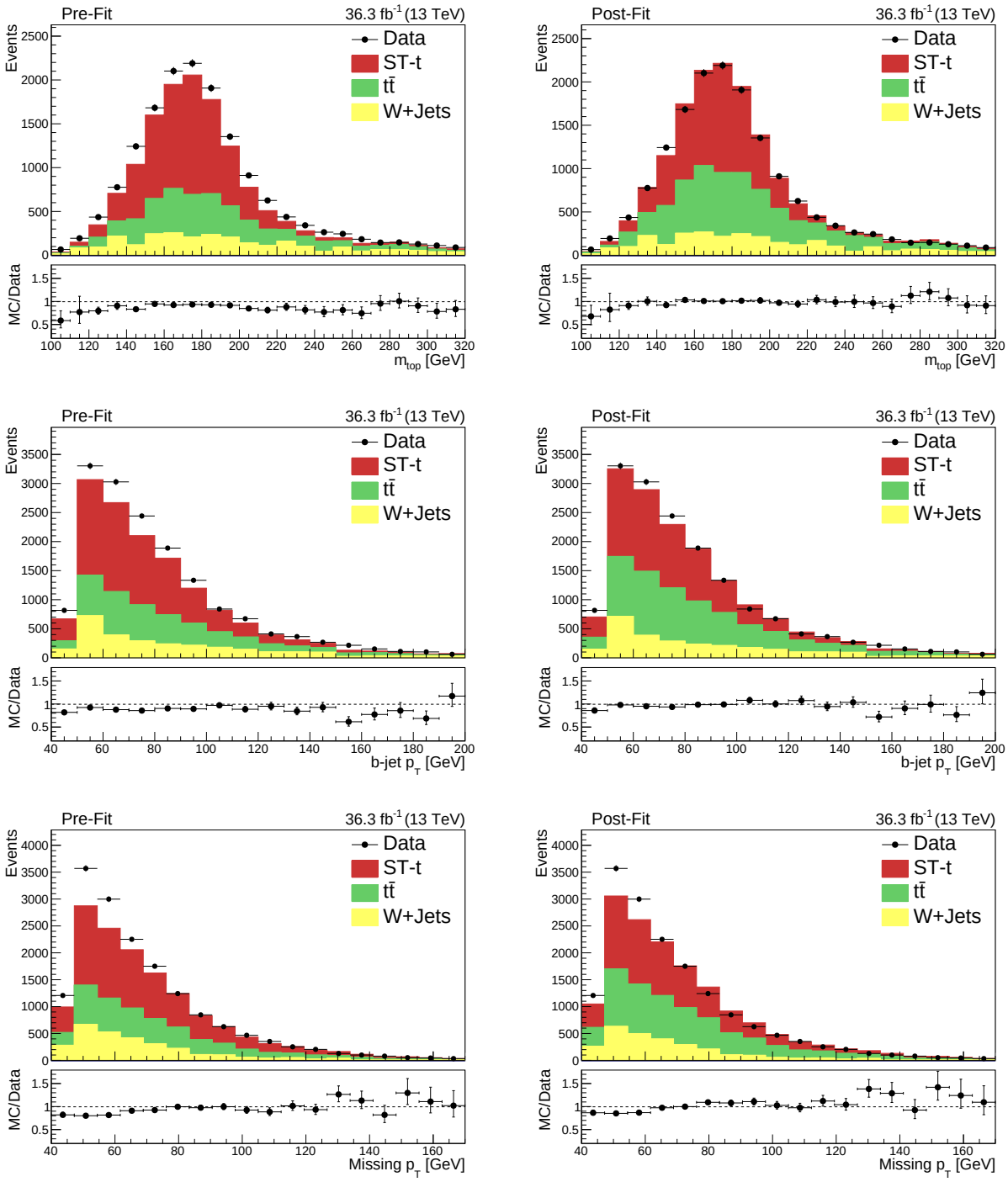


Figure 5.8: Distributions of m_{top} (upper row), b-jet p_T (middle row) and missing p_T (lower row) compared to data before and after the fit.

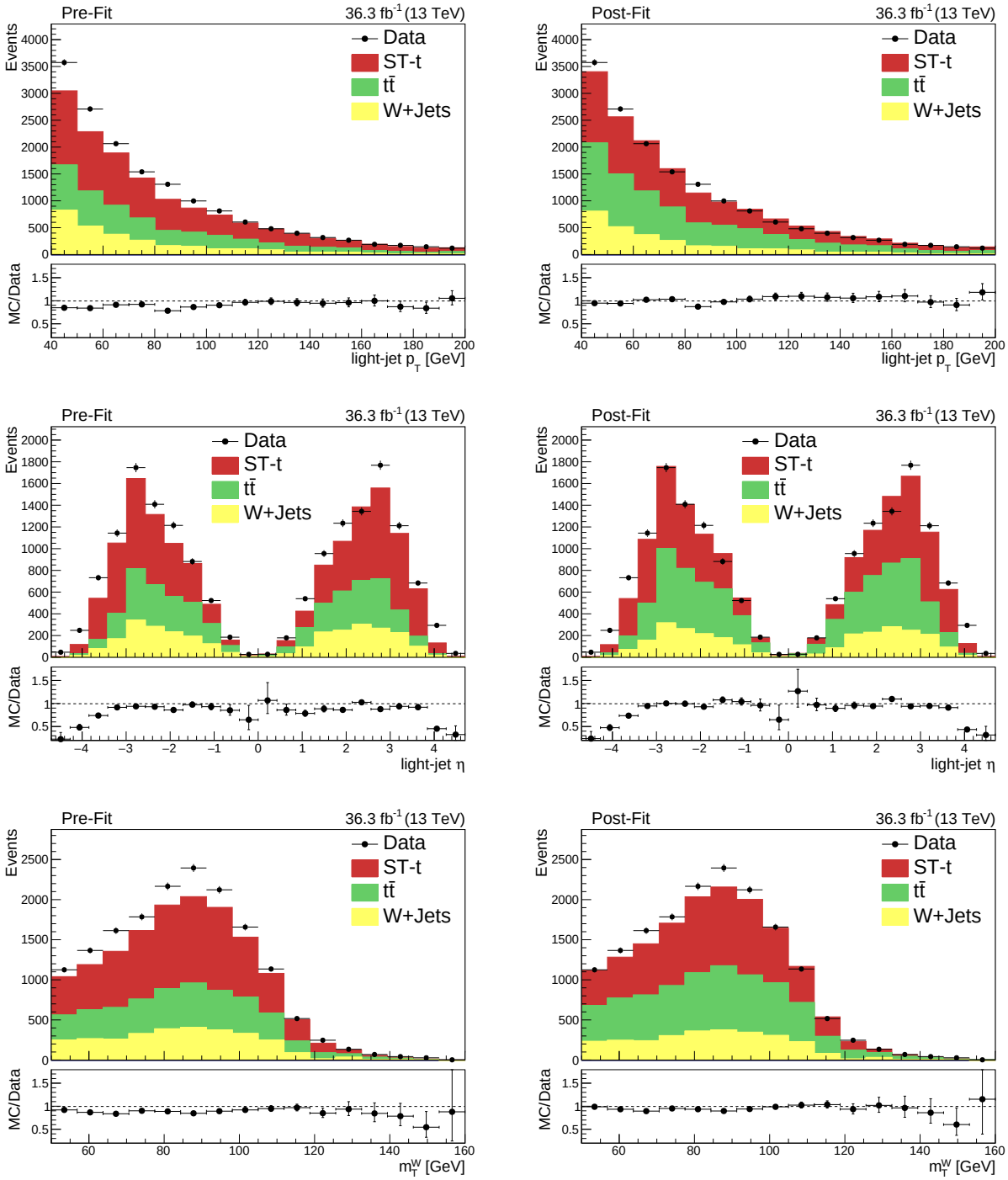


Figure 5.9: Distributions of light jet p_T (upper row), light jet pseudorapidity η (middle row) and transverse mass of W boson m_T^W (lower row) compared to data before and after the fit.

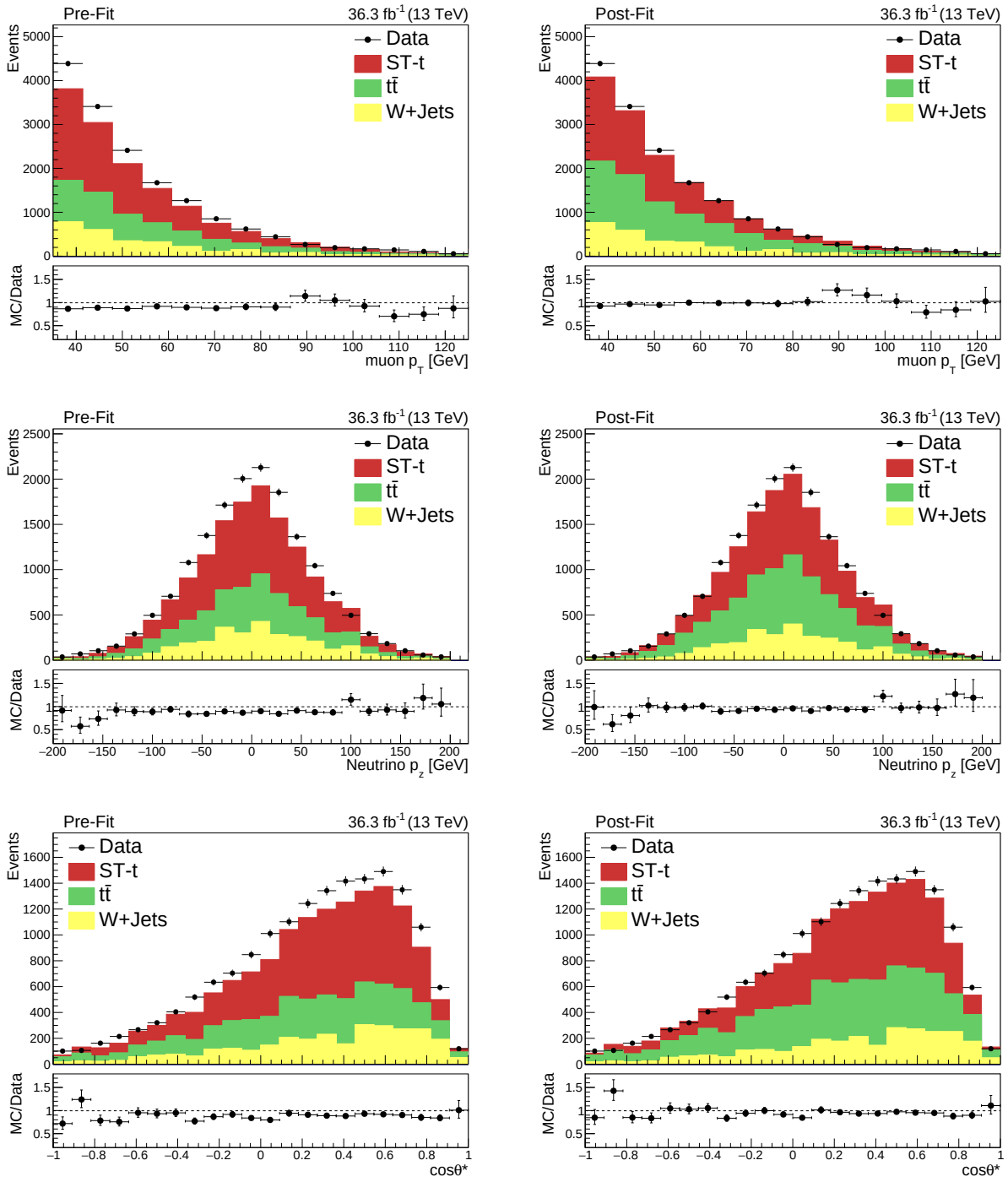


Figure 5.10: Distributions of muon p_T (upper row), neutrino p_z (middle row) and $\cos\theta^*$ (lower row) compared to data before and after the fit.

5.5 Conclusions and Outlook

The top quark mass was successfully measured using events with a single reconstructed top quark, and the precision of the measurement was improved by introducing the KinFitReg method, as opposed to traditional approaches. However, in the context of a master's thesis, several limitations were imposed, meaning this analysis is not yet a complete measurement. Notably, the next steps that would contribute to a more rigorous study include:

- Analyzing electron decay channels as well, since only muons were studied. In particular, the combination of data from both electron and muon decay channels will lead to a more accurate estimate of the mass.
- Considering additional background sources, such as single top tW and s-channels, QCD, Z+Jets, and VV backgrounds. As previously mentioned, the most prominent background, QCD multijet production, requires a dedicated data-driven treatment to account for the large discrepancies in the fitted parameters.
- Using more mass samples. The interpolation approach followed in this analysis was based on the assumption of a linear relation of the mass distributions, which can be a source of uncertainty.
- Adjusting BDTs (input variables and hyperparameters) for optimal response. Although adequate outputs were obtained from the regressions and classification processes, the overall performance can be characterized as sub-optimal. In-depth optimization was not attempted due to time restrictions.
- Estimating the systematic uncertainties (experimental and modeling).

In spite of the aforementioned limitations the results obtained in this analysis align well with previous results. The apparent reduction in statistical uncertainties, attributed to the improved resolution of top mass distributions, strongly suggests that an overall accurate measurement, with systematic errors also reduced, is feasible. Therefore, this method is highly promising for measuring the top quark mass with unprecedented precision.

Appendix A

W+JETS; Comparing LO/NLO

To account for the W+Jets background, NLO order contributions should be considered. However, due to limited statistics, the training of the BDT for signal and background classification was performed using LO, based on the assumption that the kinematic differences between LO and NLO distributions are negligible. This assumption is justified by observing the differences of the input variable distributions presented in the plots below. In this way, the need for TMVA to handle the negative weights present in NLO samples is also removed.

This hybrid method, which involves training on LO samples and applying the results to NLO samples, suffers from the reduced cross section in LO. To account for this, a scale factor of $k = 1.78$ is applied during training. This factor was obtained by dividing the integral of the NLO distributions by that of the LO distributions.

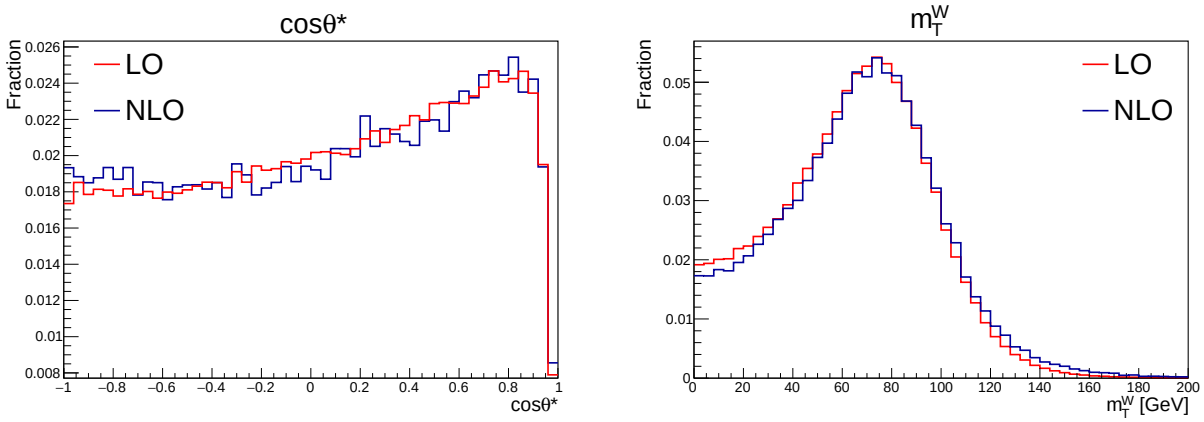


Figure A.1: Comparison of the LO and NLO distributions for the input variables.

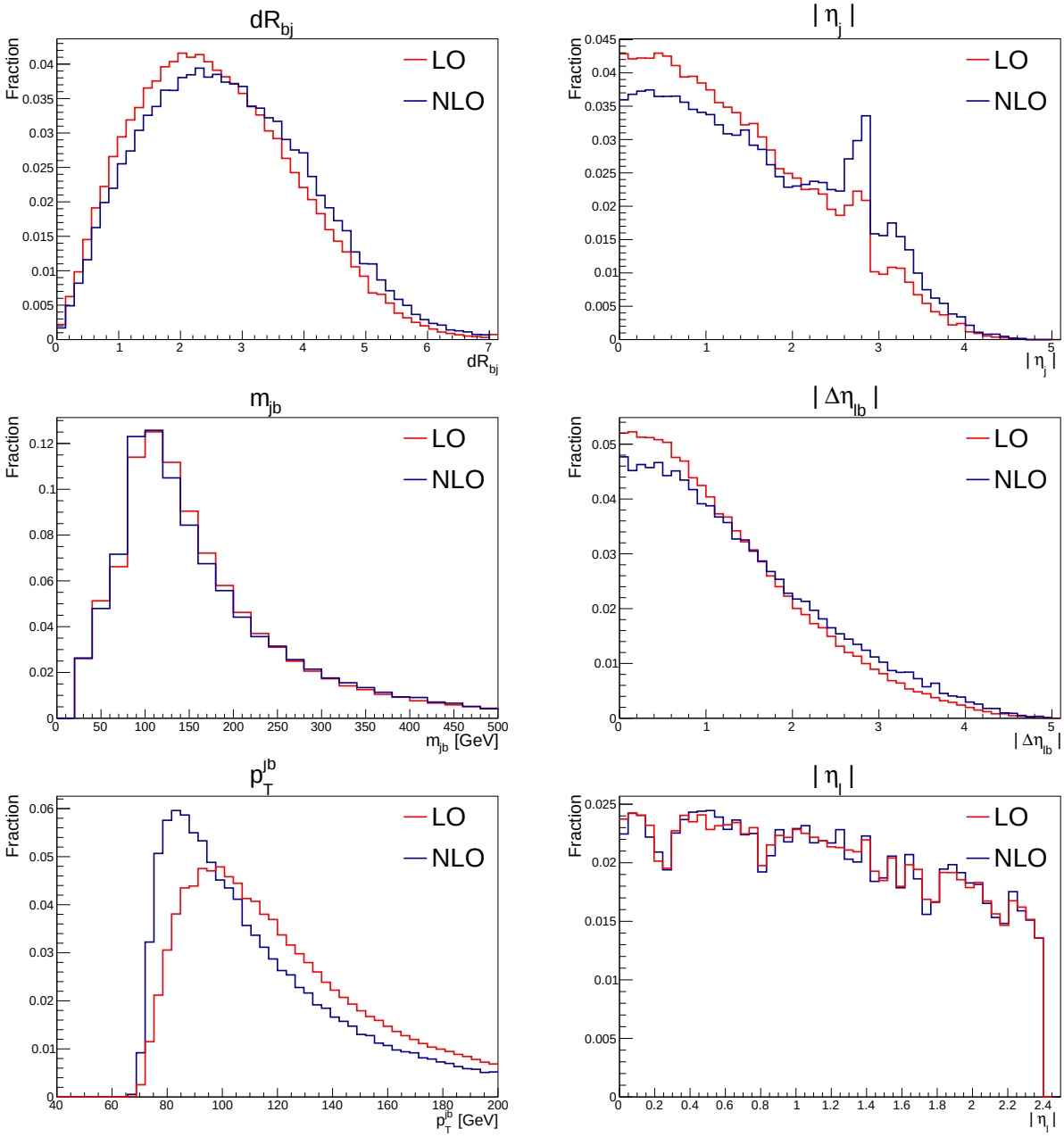


Figure A.2: Comparison of the LO and NLO distributions for the rest of the input variables.

References

[1] M. Thomson, *Modern Particle Physics* (Cambridge University Press, 2013).

[2] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics* (Wiley, 1984).

[3] P. D. Group, R. L. Workman, *et al.*, "Review of Particle Physics," Progress of Theoretical and Experimental Physics, Volume 2022, Number 8, Pages 083C01 (2022).

[4] U. Husemann, "Top-quark physics: Status and prospects," Progress in Particle and Nuclear Physics, Volume 95, Pages 48–97 (2017).

[5] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders through $o(\alpha_s^4)$," Physical Review Letters, Volume 110, Number 25 (2013).

[6] LHC Top Working Group, "LHCTopWG Summary Plots," CC-BY-4.0 license (2024).

[7] J. Campbell, T. Neumann, and Z. Sullivan, "Single-top-quark production in the t-channel at nnlo," JHEP, Volume 02, Pages 040 (2021), arXiv:2012.01574 [hep-ph] .

[8] N. Kidonakis and N. Yamanaka, "Higher-order corrections for tw production at high-energy hadron colliders," JHEP, Volume 05, Pages 278 (2021), arXiv:2102.11300 [hep-ph] .

[9] CERN, "Facts and figures about LHC," `https://home.cern/resources/faqs/facts-and-figures-about-lhc` (Accessed: 07/2024).

[10] CERN, "Linear accelerator 4," `https://home.cern/science/accelerators/linear-accelerator-4` (Accessed: 07/2024).

[11] CERN, "Proton synchrotron booster," `https://home.cern/science/accelerators/proton-synchrotron-booster` (Accessed: 07/2024).

[12] CERN, "Proton synchrotron," `https://home.cern/science/accelerators/proton-synchrotron` (Accessed: 07/2024).

[13] CERN, "Super proton synchrotron," `https://home.cern/science/accelerators/super-proton-synchrotron` (Accessed: 07/2024).

[14] E. Mobs, "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018," (2018), general Photo.

[15] CMS Collaboration, "Public CMS luminosity information," `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults` (Accessed: 07/2024).

[16] CERN, "High-luminosity LHC," `https://home.cern/resources/faqs/high-luminosity-lhc` (Accessed: 07/2024).

[17] G. L. Bayatian *et al.* (CMS), *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*, Tech. Rep. (CERN, 2006).

[18] D. Barney, "CMS Detector Slice," (2016), CMS Collection.

[19] S. Chatrchyan *et al.* (CMS), "The CMS Experiment at the CERN LHC," Journal of Instrumentation (JINST), Volume 3, Page S08004 (2008).

[20] L. Caminada (CMS Collaboration), "Performance and operation of the CMS Phase 1 pixel detector," Nuclear Instruments and Methods in Physics Research Section A, Volume 936, Pages 688–690 (2019).

[21] A. Benaglia, "The CMS ECAL performance with examples," CERN, CMS Collaboration, Geneva (2014).

[22] A. Martelli (CMS), "The CMS Electromagnetic Calorimeter: lessons learned during LHC run 1, overview and future projections," Proceedings of the TIPP 2014 Conference, Volume TIPP2014, Page 029 (2014).

[23] CMS, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data," Journal of Instrumentation, Volume 5, Number 03, Pages T03012 (2010).

[24] CMS, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV," Journal of Instrumentation, Volume 7, Number 10, Pages P10002 (2012).

[25] CMS, *CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger*, Tech. Rep. (CERN, 2002).

[26] J. M. Butterworth, G. Dissertori, and G. P. Salam, "Hard Processes in Proton-Proton Collisions at the Large Hadron Collider," Annual Review of Nuclear and Particle Science, Volume 62, Pages 387–405 (2012), arXiv:1202.0583 [hep-ex] .

[27] R. D. Field, "The underlying event in hard scattering processes," (2002), arXiv:hep-ph/0201192 [hep-ph] .

[28] A. M. Sirunyan *et al.* (CMS), "Particle-flow reconstruction and global event description with the CMS detector," Journal of Instrumentation, Volume 12, Number 10, Page P10003 (2017), arXiv:1706.04965 [physics.ins-det] .

[29] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{t} jet clustering algorithm," Journal of High Energy Physics, Volume 2008, Number 04, Page 063 (2008).

[30] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, "Jet Flavour Classification Using DeepJet," Journal of Instrumentation, Volume 15, Number 12, Page P12012 (2020), arXiv:2008.10519 [hep-ex] .

[31] M. D. Schwartz, "Tasi lectures on collider physics," (2017), arXiv:1709.04533 .

[32] A. Tumasyan *et al.* (CMS), "Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13$ TeV," Journal of High Energy Physics, Volume 12, Page 161 (2021), arXiv:2108.10407 [hep-ex] .

[33] A. Hocker *et al.* (TMVA), "TMVA - Toolkit for Multivariate Data Analysis," CERN-OPEN-2007-007 (2007), arXiv:physics/0703039 .

[34] N. Chernyavskaya (CMS), "Search for the standard model Higgs boson produced in vector boson fusion and decaying to bottom quarks using the Run1 and 2015 Run2 data samples with the CMS experiment," Proceedings of Science, Volume LHCP2016, Page 186 (2016).