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 ενεργού διατομής παραγωγής ζευγών τοπ-αντι-τοπ κουάρκ με τη βοήθεια αδρονικών πιδάκων μεγάλης εγκάρσιας ορμής στο πείραμα CMS

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της Εμμανουέλας Ανδρεαδάκη

Επιβλέπων: Κώστας Κουσουρής

Αθήνα, Ιούνιος, 2018

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

Contents

References

Abstract

This note presents a first measurement of the top quark pair production cross section in the fully hadronic decay channel at a center-of-mass energy of $\sqrt{s}$ = 13 TeV using data corresponding to an integrated luminosity of 3.7 fb^{-1} taken with the CMS detector. The cross-section is determined from an unbinned maximum likelihood fit to the reconstructed top quark mass. The reconstruction of ttbar candidates is performed after a cut-based event selection using a kinematic fit. A data-driven technique is used to estimate the dominant background from QCD multijet production. The cross-section measurement yields $\sigma incl = 572 \pm 15(stat)^{+126}_{-48}(syst) + 15(lumi) \quad [pb]$. This result is consistent with similar measurements in boosted topologies and within the Standard Model predictions.

Chapter 1

The Standard Model of Particle Physics

Όλες οι δυνάμεις στο Σύμπαν μπορούν να περιγραφούν από την αλληλεπίδραση μερικών στοιχειωδών σωματιδίων. Οι νόμοι που περιγράφουν τις αλληλεπιδράσεις αυτές έχουν περιοριστεί σε ένα σύνολο στοιχειωδών δυνάμεων που όλες μαζί συνθέτουν το Καθιερωμένο Πρότυπο. Ο απώτερος στόχος είναι η δημιουργία μιας θεωρίας σταθερής και αυτόνομης, από την οποία θα μπορούν να εξαχθούν όλοι οι νόμοι της φύσης που μπορούν να περιγράψουν τη συμπεριφορά της μάζας και της ενέργειας. Καθώς μια τέτοια ενοποιημένη θεωρία ακόμα δεν έχει ακόμη υλοποιηθεί, η έρευνα για αυτήν, έχει ωφελήσει αρκετά τη δημιουργία του Καθιερωμένου Προτύπου, μιας θεωρίας που τελειοποιήθηκε στα τέλη της δεκαετίας του '70, μετά από την πειραματική επιβεβαίωση της ύπαρξης των quarks. Παρά τους περιορισμούς του, το SM (i.e. Standard Model), είναι μια θεωρία που έχει ελεγχθεί πειραματικά αρκετές φορές στο πέρασμα του χρόνου. Παρακάτω γίνεται μια αναφορά στη θεωρία αυτή και ειδικότερα στο πιο βαρύ συστατικό της, το top quark.

Το Καθιερωμένο Πρότυπο είναι μια κβαντική θεωρία πεδίου που περιγράφει τα στοιχειώδη σωματίδια και τις αλληλεπιδράσεις τους. Με βάση τη σημερινή κατανόηση της φύσης, όλη η ύλη αποτελείται από 12 στοιχειώδη φερμιόνια: 6 λεπτόνια και 6 quarks και τα αντισωματίδιά τους, τα οποία παρατίθενται στον παρακάτω πίνακα. Επίσης, 4 τύποι δυνάμεων είναι υπεύθυνοι για την αλληλεπίδραση αυτών των σωματιδίων. Το Καθιερωμένο Πρότυπο περιγράφει τα 12 θεμελιώδη συστατικά της ύλης και τις 3 από τις 4 αλληλεπιδράσεις, την ισχυρή, την ασθενή και την ηλεκτρομαγνητική αλληλεπίδραση. Η τέταρτη αλληλεπίδραση, η βαρύτητα, δεν έχει ακόμα διατυπωθεί βάσει μιας επανακανονικοποιήσιμης κβαντικής θεωρίας πεδίου.

Fermions	Generation			electr. charge	spin	weak isospin	colour
Leptons	ν_e	ν_μ	ν_τ	0	$\hbar/2$	$+1/2$	–
	e	μ	τ	$-e$		$-1/2$	
Quarks	u	c	t	$+2/3e$	$\hbar/2$	$+1/2$	r, g, b
	d	s	b	$-1/3e$		$-1/2$	

Table 1.1: Quantum numbers of fundamental particles described by the SM

1.1. Fundamental Particles

Το Καθιερωμένο Πρότυπο είναι μια θεωρία πεδίου βασισμένη στα φερμιονικά πεδία και τις συμμετρίες βαθμίδας (gauge symmetries). Απαιτώντας η Λαγκρανζιανή να είναι αναλλοίωτη κάτω από μετασχηματισμούς βαθμίδας, τοπική και επανακανονικοποιήσιμη οδηγεί σε μια Λαγκρανζιανή της μορφής:

$$\mathcal{L}_{\text{Standard Model}} = \mathcal{L}_{\text{Gauge}} + \mathcal{L}_{\text{Matter}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}}$$

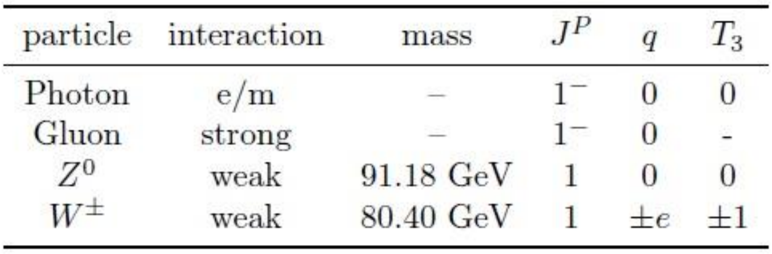
Ο πρώτος όρος περιγράφει την κινητική ενέργεια των πεδίων βαθμίδας και την αλληλεπίδραση με τον εαυτό τους ενώ ο δεύτερος όρος περιγράφει την κινητική ενέργεια των φερμιονίων και την αλληλεπίδρασή τους με τα πεδία βαθμίδας. Όροι μάζας για τα μποζόνια και τα φερμιόνια βαθμίδας είναι απαγορευμένοι καθώς δεν είναι αναλλοίωτοι κάτω από τους μετασχηματισμούς βαθμίδας. Ο τρίτος όρος είναι υπεύθυνος για τη δημιουργία της μάζας των φερμιονίων και την αλληλεπίδρασή τους με το πεδίο Higgs, ενώ ο τέταρτος όρος καθορίζει την κινητική ενέργεια του πεδίου Higgs, τις αλληλεπιδράσεις βαθμίδος και το δυναμικό Higgs. Το ότι η τιμή του πεδίου Higgs στο κενό αναμένεται να είναι διάφορη του μηδενός σπαέι την τοπική ηλεκτρασθενή συμμετρία βαθμίδας και παράγει τη μάζα για τα μποζόνια.

Τα θεμελιώδη σωματίδια του Καθιερωμένου Προτύπου που αναπαρίστανται ως φερμιονικά πεδία, περιγράφονται από μια σειρά κβαντικών αριθμών: το ηλεκτρικό φορτίο, το spin, το ασθενές isospin και το χρώμα. Τα λεπτόνια είναι άχρωμα και φέρουν spin ħ/2, όπου ħ=h/2π, με h τη σταθερά Planck με μονάδες h=4.14*10^{-15} eV στο SI. Υπάρχουν τρεις γεύσεις φορτισμένων λεπτονίων, το ηλεκτρόνιο *e*, το μιόνιο *μ* και το tau λεπτόνιο *τ*, που φέρουν ηλεκτρικό φορτίο e=1.602*10-^{19} C. Είναι κατηγοριοποιημένα σε διπλέτες με τα τρία ουδέτερα λεπτόνια, τη νετρίνο του ηλεκτρονίου, ν_e, το νετρίνο του μιονίου, ν_μ, και το νετρίνο του tau, ν_τ. Τα λεπτόνια αλληλεπιδρούν μόνο μέσω της ασθενής αλληλεπίδρασης στην περίπτωση των νετρίνων, και μέσω της ηλεκτρασθενούς αλληλεπίδρασης στην περίπτωση των φορτισμένων λεπτονίων. Τα νετρίνα έχουν υποτίθεται μηδενική μάζα βάσει των προβλέψεων του Καθιερωμένου Προτύπου. Η μάζα των υπολοίπων φορτισμένων λεπτονίων κυμαίνεται από $m_e = 0.5$ MeV/c^2 για το e, μέχρι $m_\tau = 1776.8$ MeV/c^2 για το tau.

Υπάρχουν 6 quarks, 3 τύπου down quarks, το down, το strange και το bottom quark που φέρουν ηλεκτρικό φορτίο -1/3e, και 3 τύπου up quarks, το up, το charm και το top quark, που φέρουν ηλεκτρικό φορτίο +2/3e.

1.2. Interactions

Τα διαφορετικά σωματίδια αλληλεπιδρούν μεταξύ τους με την ανταλλαγή μποζονίων βαθμίδας spin 1. Μια επισκόπηση των μποζονίων και των κβαντομηχανικών τους ιδιοτήτων παρουσιάζεται στον παρακάτω πίνακα.



particle	interaction	mass	J^P	q	T_3
Photon	e/m	–	1^-	0	0
Gluon	strong	–	1^-	0	-
Z^0	weak	91.18 GeV	1	0	0
$W^\pm$	weak	80.40 GeV	1	$\pm e$	± 1

Table 1.2.: The gauge bosons of the SM and their characteristics. J denotes the angular momentum of the boson, P its parity, q its electrical charge and T3 the third component of the weak isospin

1.2.1. Electromagnetic and Weak Interactions

Φορέας της ηλεκτρομαγνητικής αλληλεπίδρασης είναι το φωτόνιο, το οποίο έχει μηδενική μάζα. Στην ασθενή αλληλεπίδραση γίνεται ανταλλαγή βαθμωτών μποζονίων τα οποία έχουν μάζα, του ηλεκτρικά ουδέτερου *Z* μποζονίου με μάζα 91.1876 ± 0.0021 GeV/c^2 και πλάτος διάσπασης 2.49852 ± 0.0023 GeV/c^2, και του W$^\pm$ μποζονίου με θετικό ή αρνητικό ηλεκτρικό φορτίο, μάζα 80.385 ± 0.015 GeV/c^2 και πλάτος διάσπασης 2.085 ± 0.042 GeV/c^2. Στην ασθενή αλληλεπίδραση οι ιδιοκαταστάσεις της γεύσης δεν αντιστοιχούν στις ιδιοκαταστάσεις της μάζας των quarks. Οι ιδιοκαταστάσεις σχετίζονται με τον πίνακα CKM που προκύπτει από τη διαγωνοποίηση των πινάκων μάζας των φερμιονίων στην προσέγγισή της θεωρίας πεδίου.

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

Table 1.3.: Weak interaction eigenstates connected to mass eigenstates through the unitary transformation, CKM matrix

1.2.2. Strong Interaction

Οι δυνάμεις που συγκρατούν τα quarks εγκλωβισμένα μέσα στα αδρόνια ονομάζονται ισχυρές δυνάμεις. Οι δυνάμεις αυτές πηγάζουν από έναν κβαντικό αριθμό των quarks που ονομάζεται χρώμα, ο οποίος έχει τις εξής ιδιότητες:

i. Υπάρχουν τρία είδη φορτίου χρώμα: R, G, B.

ii. Κάθε quark έχει υποχρεωτικά ένα από τα τρία αυτά φορτία-χρώμα (πχ. uR, uG, uB)

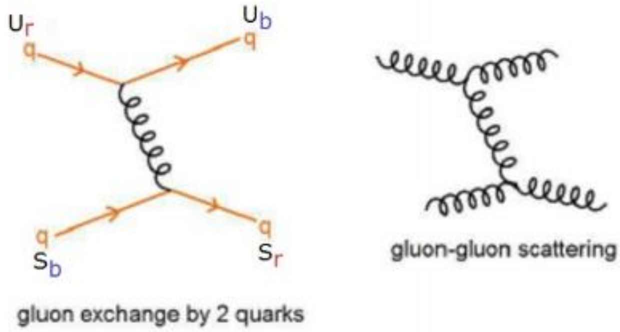
iii. Σε κάθε φορτίο αντιστοιχεί ένα αντιφορτίο/συμπληρωματικό χρώμα.

iv. Κάθε anti-quark έχει υποχρεωτικά ένα από τα τρία αυτά αντιφορτία (πχ. uRbar, uGbar, uBbar)

v. Η συνύπαρξη και των τριών φορτίων/αντιφορτίων χρωμάτων δίνει μη χρώμα (λευκό). Ομοίως και η συνύπαρξη χρώματος-αντιχρώματος.

Έτσι, κάθε quark χαρακτηρίζεται εκτός από το είδος του (u, d, s, c, b, t) και από το φορτίο χρώμα του, με αποτέλεσμα να τριπλασιάζεται ο αριθμός των quarks. Το είδος του quark ονομάζεται γεύση (flavor). Στις ισχυρές αλληλεπιδράσεις πηγή των ισχυρών πεδίων είναι το φορτίο χρώμα που έχει το κάθε quark. Τα κβάντα των ισχυρών πεδίων ονόμαζονται συγκολλητές (gluons) με μηδενική μάζα και spin=1.

Οι ισχυρές δυνάμεις μεταξύ δύο quarks μπορούν να περιγραφούν με την ανταλλαγή gluons., όπως φαίνεται στο παρακάτω σχήμα. Κατά την ανταλλαγή επιτρέπεται να έχουμε αλλαγή του χρώματος των quarks, αν και το συνολικό χρώμα διατηρείται σε κάθε κόμβο.



gluon exchange by 2 quarks

gluon-gluon scattering

Fig. 1.1: qluon exchange between 2 quarks/2 gluons

Υπάρχουν 8 gluons με τα εξής χρώματα:

$$R\overline{G},\ R\overline{B},\ G\overline{R},\ G\overline{B},\ B\overline{R},\ B\overline{G} \qquad (1)$$

$$\frac{1}{\sqrt{2}}(R\overline{R}-G\overline{G}) \qquad \frac{1}{\sqrt{6}}(R\overline{R}+G\overline{G}-2B\overline{B}) \qquad (2)$$

Κατά την ανταλλαγή gluons (1), τα αντίστοιχα quarks αλλάζουν χρώμα ενώ κατά την ανταλλαγή gluons (2) τα αντίστοιχα quαρκς δεν αλλάζουν χρώμα.

Υπάρχει τέλος και ο παρακάτω συνδιασμός (ορθογώνιος και στους 8 (1) & (2),

$$\frac{1}{\sqrt{3}}(R\overline{R}+G\overline{G}+B\overline{B}) \qquad (3)$$

που αποτελεί χρωματικά μοναχική κατάσταση (color singlet), δηλαδή δεν έχει χρώμα και επομένως δε μπορεί να ανταλλάσεται μεταξύ 2 quarks.

1.3. Extension of The Standard Model

Το Καθιερωμένο Πρότυπο είναι μία από τις πιο ελεγμένες θεωρίες και μπορεί να περιγράψει πολλές παρατηρήσεις στον τομέα της ηλεκτρασθενούς και της ισχυρής αλληλεπίδρασης. Έχουν διεξαχθεί πολλές ακριβείς μετρήσεις για να ελεγχθεί η ορθότητα της θεωρίας και, μέχρι σήμερα, έχει αποδειχθεί σωστή. Η πολύ σημαντική επιτυχία του Καθιερωμένου Προτύπου είναι ο καθορισμός της μη ομαλής μαγνητικής ροπής του ηλεκτρονίου, η ανακάλυψη του αξονικού Brout-Enlert-Higgs (BEH) – μποζονίου το 2012 και η πρόβλεψη του top quark το 1995.

- ### Ενοποίηση των δυνάμεων

Η μεγαλύτερη επιτυχία του Καθιερωμένου Προτύπου μέχρι σήμερα, είναι η ενοποίηση της ηλεκτρομαγνητικής και της ασθενούς δύναμης, η λεγόμενη ηλεκτρασθενής. Το 19$^{\text{ο}}$ αιώνα, η ηλεκτρική και η μαγνητική δύναμη ενοποιήθηκαν στη θεωρία του ηλεκτρομαγνητισμού από τον Maxwell. Τον 20$^{\text{ο}}$ αιώνα, αναπτύχθηκε η θεωρία για την ενοποίηση ηλεκτρομαγνητικής και ασθενής δύναμης στην ηλεκτρασθενή, αλλά δεν υπάρχει κάποια αξιόπιστη θεωρία για την ενοποίηση με την ισχυρή αλληλεπίδραση. Παρ' όλα αυτά, αναμένεται σε κάποια πολύ υψηλή ενέργεια οι δυνάμεις αυτές να είναι ενοποιημένες, και η ενοποίηση αυτή σπάει αυθόρμητα στην κλίμακα ενέργειας στην οποία οι δυνάμεις αυτές παρουσιάζονται ξεχωριστά.

- ### Βαρυτική δύναμη

Η Γενική Θεωρία της Σχετικότητας που περιγράφει τη βαρύτητα δε μπορεί να περιγραφεί από το Καθιερωμένο Πρότυπο επειδή μία κβάντωση του πεδίου βαρύτητας οδηγεί σε μη κανονικοποιήσιμες αποκλίσεις. Αυτό είναι το βασικό πρόβλημα που παραμένει και εμποδίζει το ΚΠ να γίνει μία ενοποιημένη θεωρία που να εξηγεί όλα τα φαινόμενα στο Σύμπαν.

- ### Το πρόβλημα της ιεραρχίας

Το γνωστό πρόβλημα της ιεραρχίας αφορά την πολύ μεγάλη διαφορά μεταξύ της ενεργειακής κλίμακας της ηλεκτρασθενούς, της τάξης του 10^2 GeV και της κλίμακας του Planck της τάξης του 10^{19} GeV, όπου η βαρύτητα αρχίζει να παίζει σημαντικό ρόλο. Στο ενεργειακό αυτό κενό μεταξύ των δύο δυνάμεων, δεν έχουν προβλεφθεί νέα φαινόμενα. Συνεπώς, η σχετικά μικρή μάζα του BEH μποζονίου υπονοεί μεγάλες διορθώσεις στο ΚΠ, το οποίο δεν μπορεί να εξηγήσει το παραπάνω παράδοξο.

- Σκοτεινή Ύλη

Το Καθιερωμένο Πρότυπο δεν μπορεί να εξηγήσει τη σκοτεινή ύλη και τη σκοτεινή ενέργεια. Η σκοτεινή ύλη δεν αλληλεπιδρά ηλεκτρομαγνητικά με την ύλη και δεν απορροφά, εκπέμπει ή ανακλά το φως. Έτσι είναι πολύ δύσκολο να εντοπιστεί. Στο Σύμπαν, η ταχύτητα περιστροφής των γαλαξιών δεν μπορεί να υποστηρίξει τη βαρύτητα που δημιουργείται, άρα δε μπορεί να συγκρατούνται μόνο από τη βαρύτητα. Πιστεύεται ότι κάποια μη ανιχνεύσιμη ύλη προσφέρει την επιπλέον μάζα που χρειάζεται σ' αυτούς τους γαλαξίες, δημιουργώντας έτσι την επιπρόσθετη βαρύτητα που απαιτείται για να συγκρατούνται. Αυτή η άγνωστη μορφή ύλης ονομάζεται σκοτεινή ύλη και μετράται μέσω της ασθενούς βαρυτικής καμπύλωσης, και θερωρείται ότι η μάζα των γαλαξιών αποτελείται κατά 80% από σκοτεινή ύλη και το υπόλοιπο 20% από τη συνηθισμένη ύλη. Μελέτες σε Supernova τύπου Ia αποκαλύπτουν τη διαστολή του Σύμπαντος, οδηγώντας έτσι στην ανάγκη ύπαρξης της σκοτεινής ενέργειας. Υπολογίζεται ότι περίπου το 68.3% του Σύμπαντος αποτελείται από σκοτεινή ενέργεια.

- Νετρίνα: μίξη και μάζα

Το Καθιερωμένο Πρότυπο δεν συμπεριλαμβάνει τη μάζα και τις ταλαντώσεις νετρίνων. Περιλαμβάνει μόνο τα αριστερόστροφα άμαζα νετρίνα που κατατάσσονται σε τρεις γενιές γεύσεων. Έτσι, τα νετρίνα δε μπορούν να σχηματίσουν ζεύξη Yukawa και άρα να αποκτήσουν μάζα μέσω του μηχανισμού BEH.

1.4. Proton-proton collisions

Τα πρωτόνια αποτελούνται από δύο up quarks και ένα down quark, τα οποία ονομάζονται quark σθένους. Επιπρόσθετα, τα γκλουόνια συνεχώς εκπέμπονται και απορροφώνται μέσα στο πρωτόνιο το οποίο μπορεί επίσης να διαχωριστεί σε ζεύγη q-qbar, τα quarks θάλασσας. Η διαδικασία αυτή γίνεται σε ένα πολύ μικρό χρονικό διάστημα που επιτρέπεται από την Αρχή Απροσδιοριστίας του Heisenberg. Όλα τα παραπάνω αποτελούν τα πρωτόνια και σε συγκρούσεις πρωτονίων-πρωτονίων, δύο ή παραπάνω συστατικά των πρωτονίων αλληλεπιδρούν και η διαδικασία αυτή ονομάζεται *hard scattering process*. Τα ουδέτερα υπολείμματα της σύγκρουσης που δε φέρουν χρώμα εκπέμπουν ακτινοβολία γκλουονίων που οδηγεί στο σχηματισμό νέων χρωματικά ουδέτερων αδρονίων που συνθέτουν το υποκείμενο γεγονός. Τα γκλουόνια που έχουν χρώμα και συμμετέχουν στην παραπάνω σκέδαση εκπέμπουν επίσης ακτινοβολία πριν και μετά την αλληλεπίδραση η οποία αναφέρεται ως αρχική και τελικά κατάσταση ακτινοβολίας.

Μία σημαντική μεταβλητή για την περιγραφή της παραπάνω σκέδασης είναι η ενέργεια κέντρου μάζας, που ορίζεται ως:

$$\sqrt{s} = \sqrt{(p_1 + p_2)^2}$$

όπου p_1 και p_2 είναι τα τετραδιανύσματα των δύο αλληλοεπιδρώντων σωματιδίων. Με τα δεδομένα του 2016, η ενέργεια στο κέντρο μάζας είναι στα √s = 13 TeV.

Η στιγμιαία φωτεινότητα (Luminosity) L είναι ένα μέτρο του αριθμού των αλληλεπιδράσεων των σωματιδίων στο χρόνο και ορίζεται ως:

$$L = \frac{\gamma\ f\ k_B\ N_p^2}{4\,\pi\,\epsilon_n\,\beta*}\ F$$

όπου γ είναι ο παράγοντας Lorentz, f η συχνότητα, k_B ο αριθμός των 'πακέτων' σωματιδίων (bunches) που έρχονται στον ανιχνευτή, Np ο αριθμός των συγκρούσεων ανά bunch, ϵ_n η κανονικοποιημένη εγκάρσια εκπομπή, β* η συνάρτηση για το βήτατρο στο σημείο αλληλεπίδρασης και F ο παράγοντας ελάττωσης που οφείλεται στη γωνία μεταξύ των δεσμών πρωτονίων στο σημείο αλληλεπίδρασης. Η τιμή που προκύπτει από το ολοκλήρωμα της φωτεινότητας στο χρόνο ∫L dt είναι ένα μέτρο της ποσότητας των δεδομένων που έχουν ληφθεί και καταγραφεί σε ένα συγκεκριμένο χρονικό διάστημα.

Στα πειράματα σύγκρουσης, αντί να χρησιμοποιείται η πολική γωνία ϑ στα συστήματα συντεταγμένων, είναι σύνηθες να χρησιμοποιείται η ωκύτητα (rapidity) η οποία δίνει την ίδια πληροφορία και ορίζεται ως:

$$y = \frac{1}{2}\ln\frac{E + p_z}{E - p_z}$$

ή η ψευδο-ωκύτητα (pseudorapidity):

$$\eta = -\ln\tan\frac{\Theta}{2}$$

1.5. Top quark physics

Το top quark είναι το βαρύτερο στοιχειώδες σωματίδιο που έχει παρατηρηθεί με μάζα ηρεμίας 173.5 ± 0.8 GeV/c^2. Παρατηρήθηκε πρώτη φορά στον επιταχυντή Tevatron (Fermilab, IL, US) το 1995. Είναι το δεύτερο μέρος της ασθενούς διπλέτας του isospin, που σχηματίζεται μαζί με το bottom quark, φέρει ηλεκτρικό φορτίο Q=+2/3 και τρίτη συνιστώσα του Isospin T_3=+1/2. Η πρόβλεψή του είχε γίνει το 1977, αμέσως μετά την ανακάλυψη του bottom quark.

Σε αντίθεση με όλα τα υπόλοιπα quarks που έχουν παρατηρηθεί, ο χρόνος ζωής του top, περίπου $5*10^{-24}$ s, είναι πολύ μικρός για να σχηματιστούν δέσμιες καταστάσεις όπως τα top flavored αδρόνια μέσω της διαδικασίας της αδρανοποίησης, ή καταστάσεις σαν το ttbar quarkonium.

Η μέτρηση των ιδιοτήτων του top quark θεωρείται μία ακριβής μέτρηση των θεωρητικών προβλέψεων του Καθιερωμένου Προτύπου. Η μάζα του top quark είναι μία από τις ελεύθερες παραμέτρους του ΚΠ και γι’ αυτόν το λόγο, δε μπορεί να προβλεφθεί από τη θεωρία. Τα top quarks είναι υπόβαθρό σε αρκετές διαδικασίες πέραν του Καθιερωμένου Προτύπου, για παράδειγμα έρευνες για την υποθετική τέταρτης γενιάς διπλέτα quarks που δεν έχει ακόμα παρατηρηθεί, τα t’ ή b’ ή διασπάσεις υποθετικών υπερσυμμετρικών σωματιδίων (SUSY) όπως η διάσπαση του gluino σε ένα stop quark και ένα top quark: $\tilde{g} \rightarrow \tilde{t}t.$

Property	Value
Mass	172 ± 0.9 (stat) ± 1.3 (syst) GeV
Lifetime	$\sim 5 \times 10^{-26}$s
Charge	$+2e/3$
$\mathcal{B}(t \rightarrow b)$	$\lvert V_{tb} \rvert > 0.998$

Table 1.3.: Top Quark properties

1.5.1. ttbar production

Τα top quarks παράγονται με δύο συγκεκριμένους μηχανισμούς: ζεύγη top-antitop quarks (ttbar) παράγονται μέσω της ισχυρής αλληλεπίδρασης και τα single top quarks μέσω της ηλεκτρασθενούς. Η παραγωγή των ζευγών ttbar κυριαρχεί έναντι της παραγωγής ενός μόνο top.

Τα ζεύγη ttbar παράγονται μέσω της σύντηξης γκλουονίων και της εξαΰλωσης qqbar. Η απαιτούμενη ενέργεια για την παραγωγή ενός ζεύγους ttbar είναι τουλάχιστον η διπλάσια της μάζας του top. Εφόσον στον LHC συγκρούονται πρωτόνια με πρωτόνια, τα anti-quarks είναι τα quarks θάλασσας. Σε ενέργεια κέντρου μάζας 13 TeV, και στον αντίστοιχο παρτονικό χώρο ορμής, η luminosity γκλουονίου-γκλουονίου είναι μεγαλύτερη από τη luminosity των quarks και έτσι η σύντηξη γκλουονίων είναι η κυρίαρχη διαδικασία και το 95% των ζευγών παράγονται έτσι, σε αντίθεση με τον Tevatron (Fermilab, US) στον οποίο συγκρούονταν πρωτόνια με αντιπρωτόνια.

Τα leading order διαγράμματα Feynman (LO) για αυτήν την παραγωγή παρουσιάζονται παρακάτω:

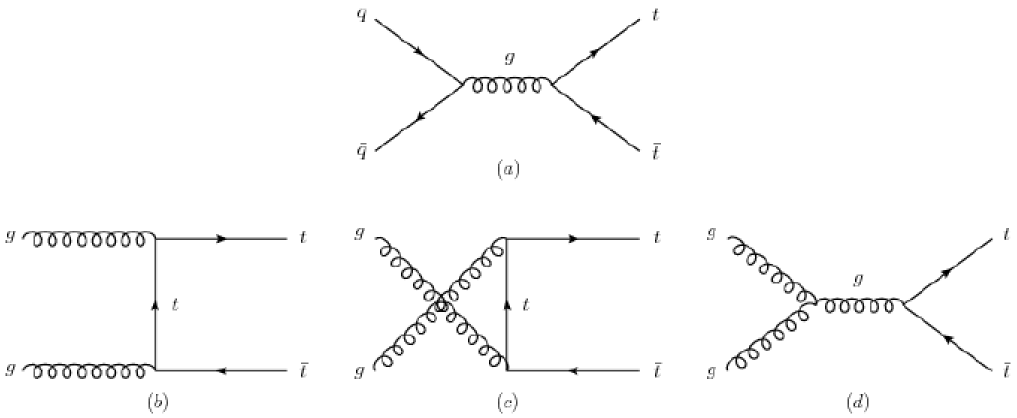


Fig.1.1: LO Feynman diagrams for ttbar production in LHC via (a) qqbar annihilation, (b-d) via gluon fusion

Στην προσέγγιση Born, η ενεργός διατομή της παραγωγής ζευγών top-antitop σε συγκρουστή αδρονίων μπορεί να περιγραφεί με τη βοήθεια των παρτονικών πυκνοτήτων πιθανότητας και της ενεργού διατομής της διαδικασίας αυτής:

$$\sigma(s, m_t^2) = \sum_{i,j} \int_0^1 dx_i \int_0^1 dx_j \; f_i(x_i, \mu_f^2) f_j(x_j, \mu_f^2) \hat{\sigma_{ij}}[\hat{s}, m_t, \alpha_s(\mu_r^2)].$$

Εδώ, m_t είναι η μάζα του top quark, xi ο λόγος της ενέργειας του εισερχόμενου πρωτονίου την οποία φέρει το παρτόνιο, οι δείκτες i,j αθροίζονται πάνω στα εισερχόμενα γκλουόνια και τα ζεύγη quark-antiquark που κυριαρχούνται από τις πυκνότητες πιθανότητας (PDF). Αυτές υπολογίζονται σε κάποια κλασματική κλίμακα, μ_f. Η ενεργός διατομή για τα αλληλοεπιδρώντα παρτόνια, σ_{ij}, με ολοκλήρωση σε όλο το φασικό χώρο, είναι μια συνάρτηση της μάζας του top, της ενέργειας του κέντρου μάζας $\sqrt{s}$ και της σταθεράς ζεύξης της QCD, α_s. Σε υψηλότερες τάξεις, η παρτονική ενεργός διατομή εξαρτάται και από τις μ_r, μ_f. Οι σταθερές μ_r και μ_f συνήθως παίρνουν τιμές σχετικές με τη διαδικασία στην οποία αναφέρονται. Για παραγωγή ttbar η πιο συνήθης επιλογή είναι $\mu_r = \mu_f = m_t$, παρόλο που δεν είναι αναγκαία συνθήκη να είναι ίσες μεταξύ τους οι σταθερές αυτές.

Οι μεταβλητές που αντιπροσωπεύουν φυσικά μεγέθη δεν θα πρέπει να εξαρτώνται από την τάξη μεγέθους σε μια επανακανονικοποιήσιμη θεωρία. Αυτό είναι αληθές μόνο αν κατά τον υπολογισμό των μεταβλητών αυτών συμπεριλαμβάνονται οι όροι διαταραχών σε όλες τις τάξεις, κάτι το οποίο όμως είναι αδύνατον. Συνεπώς το αποτέλεσμα της εφαρμογής υπολογισμών σε μια πεπερασμένη τάξη είναι ουσιαστικά ένας υπολογισμός με μία τεχνητή κλίμακα εξάρτησης. Μεταβάλλοντας τις κλίμακες σε μία συγκεκριμένη τάξη μπορεί να δοθεί εικόνα της εναπομένουσας αβεβαιότητας στον υπολογισμό και αυτή είναι η διαδικασία που ακολουθείται για την ποσοτικοποίηση των συστηματικών σφαλμάτων που απορρέουν από την εξάρτηση από την τάξη μεγέθους.

1.5.2. ttbar decay

Τα top quarks κυρίως διασπώνται σε W μποζόνια και b quarks. Το Καθιερωμένο Πρότυπο προβλέπει B(t →bW) > 0.998 και η πρόβλεψη αυτή είναι συνεπής με ότι έχει παρατηρηθεί πειραματικά. Ο τρόπος αυτός διάσπασης κυριαρχεί στις αμέσως πιο πιθανές διασπάσεις t →Ws και t →Wd, απλά επειδή η τιμή του $|V_{tb}|$ στον πίνακα CKM είναι πολύ κοντά στη μονάδα ($|V_{tb}| = 0.89 \pm 0.07$) ενώ οι τιμές των $|V_{ts}|$ και $|V_{td}|$ είναι πολύ μικρότερες. Δεν υπάρχουν ουδέτερα ρεύματα αλλαγής γεύσης σε επίπεδο δέντρου στο Καθιερωμένο Πρότυπο, συνεπώς είναι μόνο φαινόμενα υψηλότερων τάξεων. Διασπάσεις των ουδέτερων ρευμάτων αλλαγής γεύσης, t →Xq, όπου X: g, γ, Z, H και q: c,, δεν υφίστανται σε επίπεδο δέντρου στο SM, άρα και αυτές υπάρχουν μόνο σε φαινόμενα υψηλότερων τάξεων.

$\sqrt{s}$	Cross section / pb
14 TeV	$920^{+50}_{-39}\,{}^{+33}_{-35}$
10 TeV	$415^{+17}_{-21}\,{}^{+18}_{-19}$
7 TeV	$163^{+7}_{-5}\,{}^{+9}_{-9}$

Fig.1.2: Theoretical production cross sections for ttbar. Calculations are based on a top mass = 173 GeV and calculated at NNLO

Η διάσπαση t →bW θεωρείται ότι πραγματοποιείται με ποσοτό ~100%. Τα W μπορούν να διασπαστούν είτε λεπτονικά είτε αδρονικά με τις παρακάτω διασπάσεις:

- *ημι-λεπτονικές*, όπου το ένα W διασπάται σε ηλεκτρόνιο ή μιόνιο και το άλλο σε ζεύγος qqbar
  $t\bar{t} \rightarrow W^+W^-b\bar{b} \rightarrow \ell\nu\, q\bar{q}\, b\bar{b}.$
- *δι-λεπτονικές*, όπου στην τελική κατάσταση υπάρχουν μόνο ηλεκτρόνια ή και μιόνια από τις διασπάσεις των δύο W

$$t\bar{t} \to W^+W^-b\bar{b} \to \ell\nu\,\ell\nu\,b\bar{b}.$$

- *αδρονικές*, οι διασπάσεις που και τα δύο W διασπώνται σε ζεύγη qqbar και στην τελική κατάσταση υπάρχουν μόνο quarks
  $$t\bar{t} \to W^+W^-b\bar{b} \to q\bar{q}\,q\bar{q}\,b\bar{b}.$$

Στο παρακάτω σχήμα φαίνονται και τα ποσοστά κάθε διάσπασης.

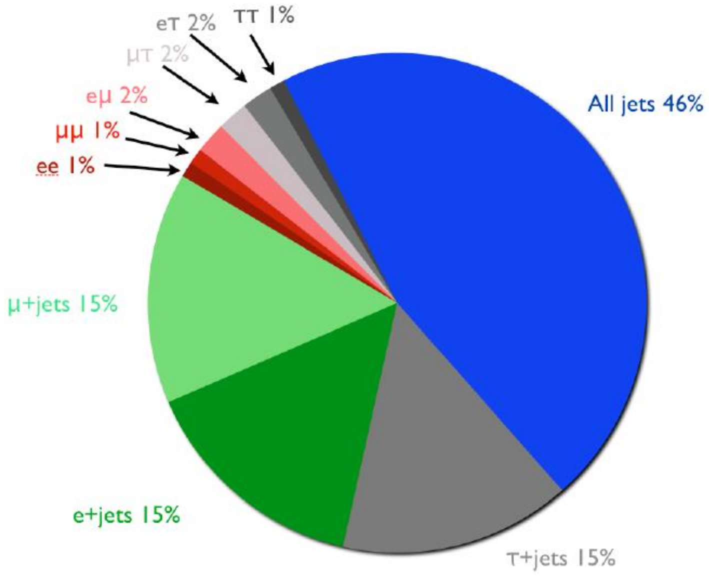


Fig.1.3: Approximate relative BRs for the decays of ttbar pairs

Κάθε ένα από τα παραπάνω κανάλια διάσπασης έχει πλεονεκτήματα και μειονεκτήματα για τη μέτρηση των ιδιοτήτων των ζευγών ttbar. Λαμβάνοντας υπόψιν τις διορθώσεις της QCD, 45.7% όλων των ζευγών ttbar διασπώνται αδρονικά, 43.8% ημι-λεπτονικά και το υπόλοιπο 10.5% δι-λεπτονικά.

Αυτό σημαίνει ότι το κανάλι με τη μεγαλύτερη αφθονία έχει 6 quarks στην υπογραφή τα οποία καταλήγουν σε jets. Αυτό το καθιστά λίγο δύσκολο να ξεχωρίσουν τα top quark γεγονότα από την μεγάλη ποσότητα των multi-jets γεγονότων υποβάθρου που παρατηρούνται σε έναν αδρονικό συγκρουστή. Το δι-λεπτονικό κανάλι έχει μια πολύ καθαρή υπογραφή με δύο λεπτόνια και δύο jets αλλά ο λόγος διάσπασης (branching ratio) στη διάσπαση του top είναι ο μικρότερος από όλα τα κανάλια διάσπασης και η ύπαρξη των δύο νετρίνων τα οποία δε μπορούν να μετρηθούν άμεσα στον ανιχνευτή, καθιστούν δύσκολη την επανακατασκευή της κινηματικής του top quark. Το κανάλι της ημιλεπτονικής διάσπασης συνδιάζει τα πλεονεκτήματα μιας πλήρους αδρονικής διάσπασης και μιας πλήρους λεπτονικής, αφού ο λόγος διάσπασης είναι σχετικά μεγάλος, έχει καθαρή υπογραφή λόγω του φορτισμένου λεπτονίου στο γεγονός και λόγω του μοναδικού neutrino που προκύπτει από τη διάσπαση του W. Αυτό το καθιστά ένα ενδιαφέρον κανάλι προς μέτρηση. Η υπογραφή που προκύπτει από το συγκεκριμένο κανάλι διάσπασης παρατίθεται παρακάτω.

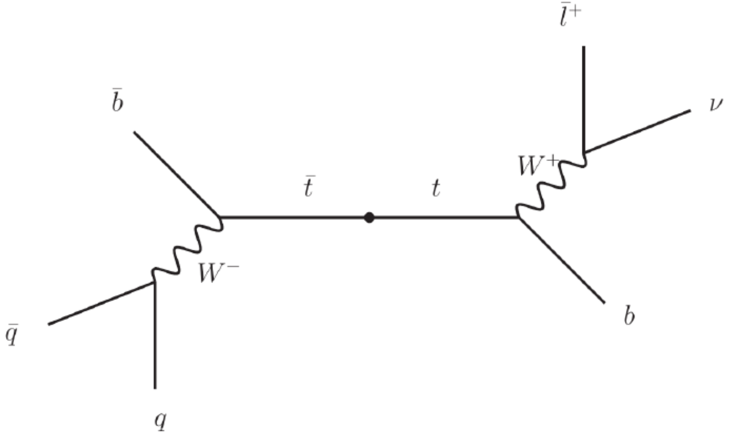


Fig.1.4: Top quark pair decay in the semi-leptonic channel

1.6. Importance of Top Quark Physics

Ο LHC είναι το δεύτερο πείραμα στο οποίο είναι δυνατόν να σχηματιστούν τα top quarks. Η φυσική των top quarks είναι ακόμα ένα σχετικά νέο πεδίο πειραματικής έρευνας και ακόμα τα αποτελέσματα από τον Tevatron έχουν αφήσει αρκετά σημεία τα οποία θα μπορούσαν να βελτιωθούν. Αυτό και μόνο προσδίδει κίνητρο για την περαιτέρω μελέτη των ρυθμών παραγωγής και των ιδιοτήτων τους. Τα top quarks έχουν επιπλέον τη δυνατότητα να χρησιμοποιηθούν ως εργαλείο βαθμονόμησης του ανιχνευτή και η παραγωγή τους θα είναι σημαντική για τη μελέτη της Φυσικής Πέραν του Καθιερωμένου Προτύπου (**B**eyond **S**tandard **M**odel Physics).

1.6.1. Calibration and Commissioning of the CMS detector

Δύο πολύ σημαντικές διαστάσεις της απόδοσης του ανιχνευτή και της επανακατασκευής του για τη μελέτη του top quark είναι το jet energy scale (JES) και η b-tagging performance, η απόδοση δηλαδή στο να αναγνωριστεί με επιτυχία ένα b quark. Το JES περιγράφει την αβεβαιότητα στη μέτρηση της ενέργειας των επανακατασκευασμένων jets, η οποία εξαρτάται από την απόδοση των καλοριμέτρων. Το b-tagging performance αναφέρεται στην αποδοτικότητα της αναγνώρισης των b-jets, το οποίο εξαρτάται ιδιαίτερα από την παρακολούθηση της τροχιάς (tracking). Γεγονότα με top quark μπορούν να χρησιμοποιηθούν για έναν αποτελεσματικό υπολογισμό των παραπάνω παραμέτρων. Επιλέγοντας δι-λεπτονικά γεγονότα με top quark, το ένα b-jet μπορεί να ταυτοποιηθεί και το b-jet που προκύπτει από το άλλο top quark μπορεί να χρησιμοποιηθεί για να εξεταστεί η αποτελεσματικότητα της ταυτοποίησης. Τα γεγονότα με λεπτόνια+jets μπορούν να χρησιμοποιηθούν για να μετρηθεί το JES των ελαφρών quakrs (u,d,s) με περιορισμό τη μάζα του W που διασπάται αδρονικά. Αυτό μπορεί να επεκταθεί χρησιμοποιώντας τη μάζα του top που διασπάται αδρονικά ώστε να υπολογιστεί το Jet Energy Scale των b-jets.

1.6.2. Top Quark as background to New Physics

Για πολλές έρευνες για τη Φυσική Πέραν του Καθιερωμένου Προτύπου, η παραγωγή και η διάσπαση του top quark αποτελεί ένα σημαντικό υπόβαθρο. Μερικές από αυτές τις μελέτες έχουν τοπολογίες γεγονότων που μοιάζουν πολύ με τα γεγονότα που περιέχουν ζεύγη ttbar. Οι μελέτες για SM Higgs με τελικές καταστάσεις όπως $H \to \tau\tau$, *WH, ttbarH* για παράδειγμα, είναι ευαίσθητες σε υπόβαθρο ttbar. Για πολλά κανάλια που κοιτάζουν για SUSY, η ελλείπουσα εγκάρσια ενέργεια είναι μία κρίσιμη υπογραφή και η κατανόηση των ουρών της $\not{E}_T$ στο σύστημα ttbar είναι αναγκαία.

Chapter 2

LHC and CMS Experiment

2.1. The Large Hadron Collider

The LHC is a 27km long superconducting accelerator machine housed at CERN, Geneva, Switzerland. It occupies the tunnel formerly used by the LEP accelerator and is situated between 45m and 170m below the Earth's surface. It is a proton-proton collider with a design center-of-mass energy of p √s = 14TeV. This is an order of magnitude greater than its predecessor the Tevatron, which was a 1.96TeV proton-antiproton collider built at Fermilab. The LHC is also to act as a heavy ion collider, accelerating lead (Pb) ions to 2.8TeV per nucleon. At the design center-of-mass energy of √s = 14TeV the number of bunches per beam will be 2808 with a bunch spacing of 25 ns and 1:15 $*10^{11}$ protons per bunch. The design luminosity is 10^{34} $cm^{-2}s^{-1}$. These specifications would mean that the production of top quark pairs would reach the rate of approximately nine a second, resulting in a true top factory". The production of such high luminosity precludes the use of an antiproton-proton design. This means that to achieve two counter-circulating beams of protons a magnetic system is required that provides opposite lines of flux for each beam. The LHC uses a novel “twin-bore" magnetic dipole design where the windings for both beams are housed in the same cryostat with the lines of flux running in the opposite sense for the two beams.

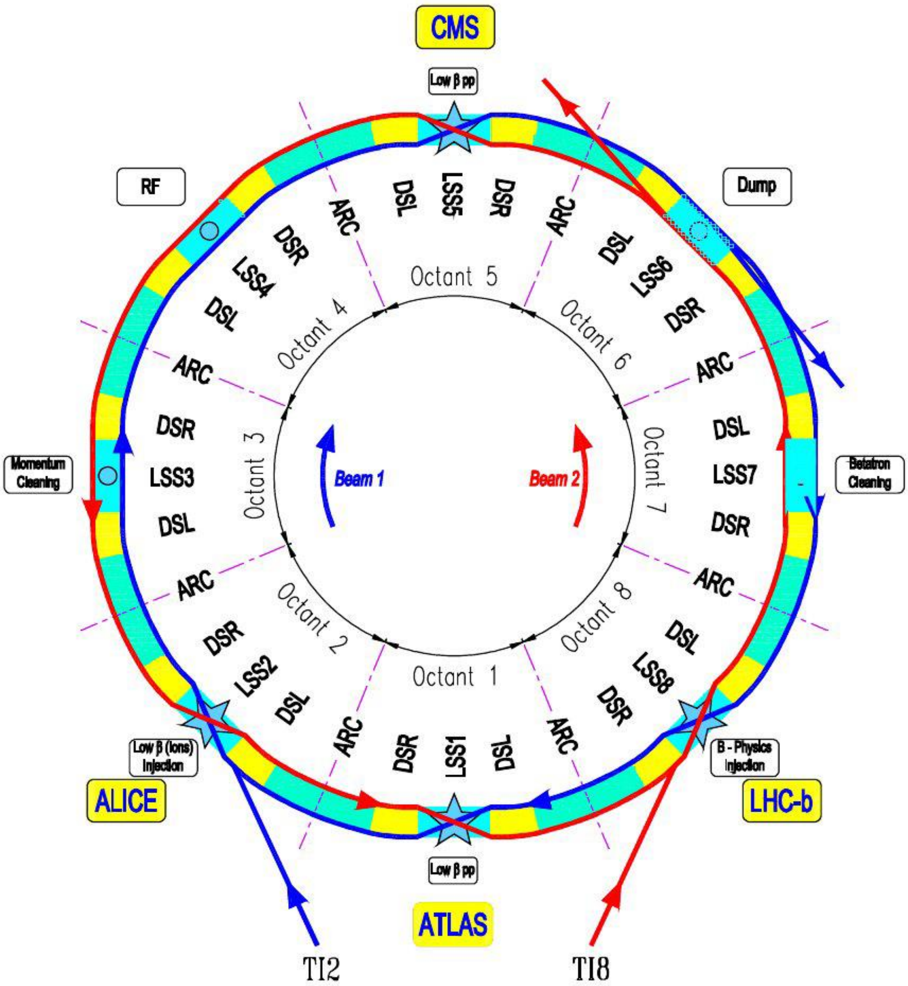


Figure 2.1: Layout of the Large Hadron Collider. In eight arcs and straight sections protons are accelerated, focused and kept on their trajectory along the LHC beam line. Proton-proton collisions take place at four interaction points

2.1.1. Experiments at LHC

There are four main detector experiments based at the LHC. The CMS and ATLAS (A Toroidal LHC ApparatuS) experiments are general purpose detectors. ALICE (A Large Ion Collider Experiment) has been built to study heavy ion collisions during specified heavy ion runs. LHCb (the Large Hadron Collider beauty experiment) is designed to study CP violation and will make precision measurements of rare decays. CMS and ATLAS, situated at diametrically opposite positions of the LHC ring, are built as complementary detectors with the same measurements in mind. They are general purpose detectors meaning they are designed to detect and measure as vast an array of potential particles and signatures as possible. The detectors have full azimuthal coverage with endcap sections closing in very close to the beam line. Their requirements include being able to make high precision measurements of Standard Model processes, such as electro-weak physics or Top physics, but also be able to potentially measure evidence for a potential zoo of particles that may yet lie in wait for discovery. The advantage or requirement of two general purpose detectors is that one experiment may confirm results of the other, whilst achieving similar aims but with different technologies and methodologies.

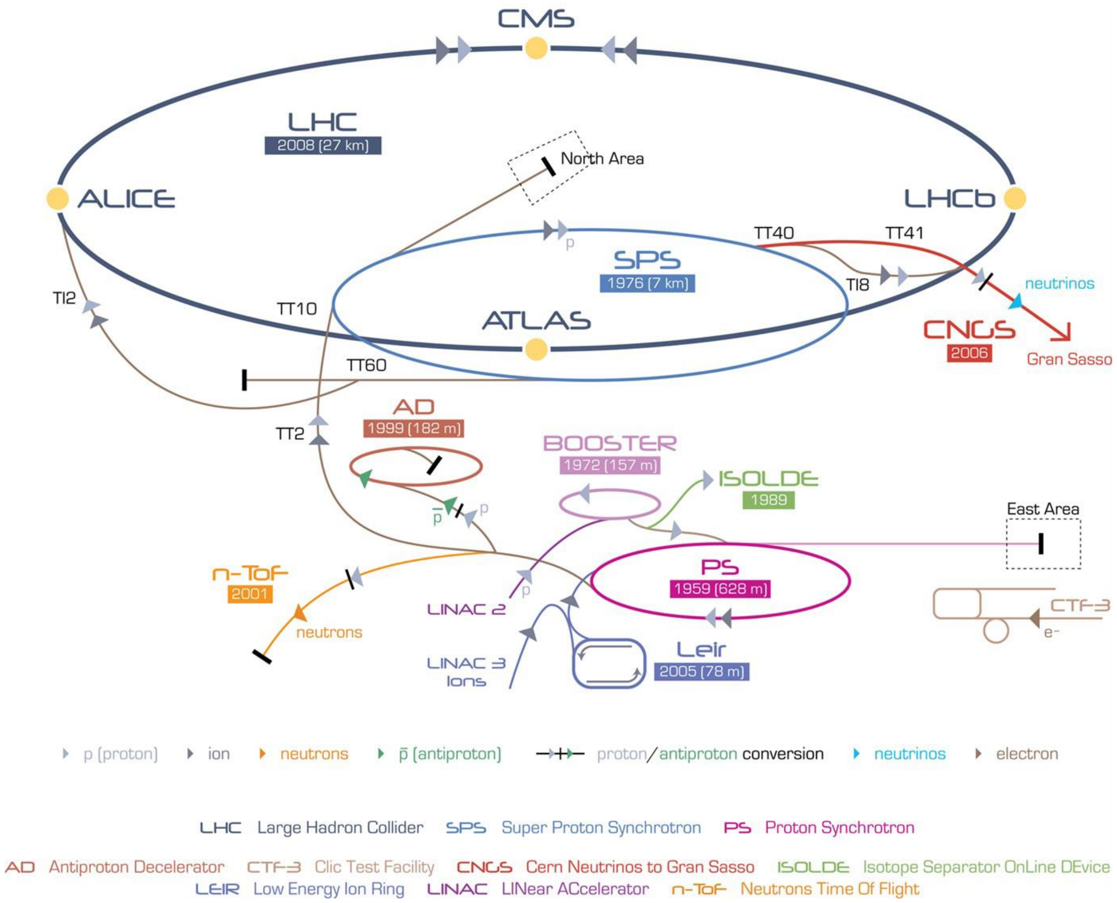


Fig. 2.2: Acceleration chain for protons. Protons undergo subsequent acceleration steps, first, in the Linac2, then in the Proton Synchrotron where bunches are formed and afterwards in the Super Proton Synchrotron from where they are injected into the LHC ring. In the LHC ring the proton bunches are accelerated to the nominal beam energy of currently up to 6.5 TeV

2.2. The Compact Muon Solenoid Detector

The CMS detector is situated at about 100m underground near the French village Cessy between the Jura mountains and Lake Geneva. It is distinguished by its large, high field solenoidal magnet, a fully silicon-based inner tracking system and a homogenous scintillating crystal-based electromagnetic calorimeter system. The coordinate system has the z-axis pointing along the beamline towards the Jura mountains.
The design of CMS is motivated by both the physics program intended for the LHC, primarily the goal of understanding the mechanism for electron-weak symmetry breaking, and also the practical implications of dealing with such high luminosities. At design luminosity up to 20 inelastic collisions can be expected in every 25ns readout window, an issue known as pileup. If the detector and electronics response time is greater than 25 ns, then this becomes a clear problem. A solution is to provide a high granularity detector with good timing resolution, such that low occupancy is achieved in the readout channels. This requires very good synchronization when dealing with millions of detector channels. The high flux of particles also necessitates that the sub-detectors and front-end electronics are radiation-hard.
To achieve the goals of the LHC physics program, the following conditions are required:

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

The design of CMS meets all these requirements.

The overall layout of CMS is shown in Figure 2.3 below.

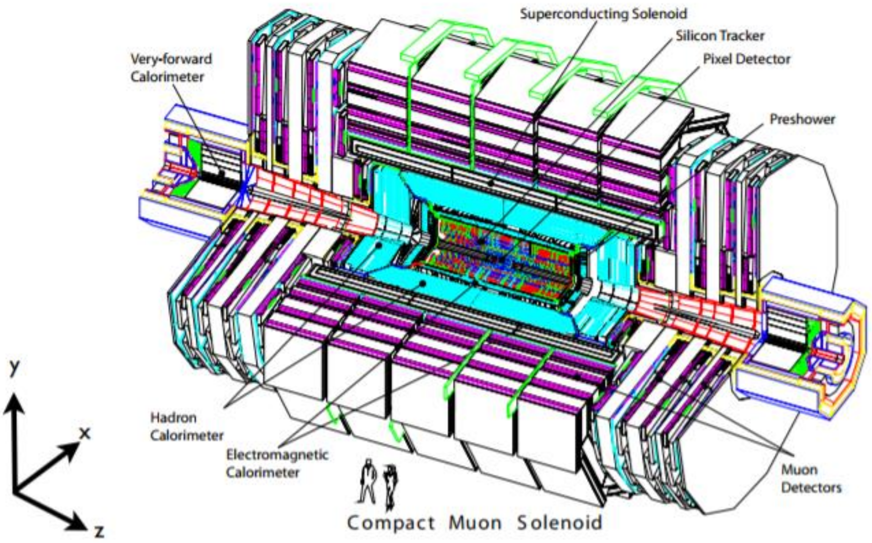


Figure 2.3: Schematic view of the CMS detector

An inner tracking system sits up close to the beamline, providing the tracking of charged particles and vertex reconstruction. This is surrounded by the calorimetry. These are subdetectors whose role is to measure the energy of electrons, photons and hadrons. Due to the different energy loss mechanisms of these particles there are two distinct calorimeters: one for both electrons and photons and one for hadrons. The tracking system and calorimetry is all housed inside the 4 Tesla solenoidal magnet which allows for the measurement of the momentum of charged particles. The muon detection system that comprises three main subdetectors lies outside the magnet.

2.2.1.The Coordinate System

CMS has adopted a coordinate system with the origin centered at the nominal collision point, the y-axis pointing vertically upward, the x-axis pointing radially inward toward the LHC center, and the z-axis pointing along the beam direction from LHC Point 5 toward the Jura mountains. The azimuth angle (ϕ) is measured from the x-axis in the x – y plane. The radial coordinate in this plane is denoted as r. The polar angle (θ) is measured from the z-axis. A quantity related to the polar angle, called the pseudorapidity (η), is defined as

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

Values of η relative to the detector are shown in Figure 2.4. From this, the momentum (P_T) and energy transverse (E_T) to the beam direction are computed from x and y components. The quantity ($\not{E}_T$) measures the imbalance in energy measured in the transverse plane, hypothetically due to neutral, non-interacting particles, by vectorially summing the component of energy from reconstructed objects which is transverse to the beam axis and multiplying by a minus sign. This is so the vector points in the opposite direction of the total reconstructed transverse energy, representing any particle produced in the event which was not reconstructed. Since there is no activity from the incident particles that is transverse to the beam, ($\not{E}_T$) should ideally be zero, requiring a transverse energy balance.

2.2.2. The Superconducting Solenoid Magnet

The solenoid for CMS has been designed to produce a uniform magnetic field of 4 T, though it is limited to 3.8 T during operation to prolong its lifetime. The dimensions of the solenoid are 6 m in diameter and 12.5 m in length, with a full-current stored energy of 2.6 GJ. The magnetic flux is returned through a 10.000 metric ton yoke comprised of 5 wheels and 2 endcaps, which themselves contain three disks each. The cold mass of the solenoid alone is 220 metric tons in mass, containing four winding layers of a stabilized NbTi conductor. The use of four windings are unique to CMS when compared to magnets used at previous experiments, which use only one winding. The greater number of windings are required to produce such a high magnetic field, which requires 4.2 × 107 Amperes / turn to produce. As a result of the size and high field of the magnet, the ratio between the stored energy and the mass is high (11.6 KJ/kg), causing a large mechanical deformation (0.15 %) while energizing the solenoid. This is much larger than values obtained by previous experiments.

2.2.3. The Inner Tracking System

The aim of the CMS tracking system is to resolve the trajectory of charge particles as they traverse the detector and to provide high resolution vertex reconstruction. The tracker consists of inner pixel layers surrounded by layers of silicon strip detectors. A schematic of the tracker is shown below. The entire system has a total length of 5.8 m, is 2.5m in diameter, and covers the pseudorapidity range $|\eta| < 2.5$. With an active silicon area of approximately 200m2 it is the largest silicon tracker ever built.

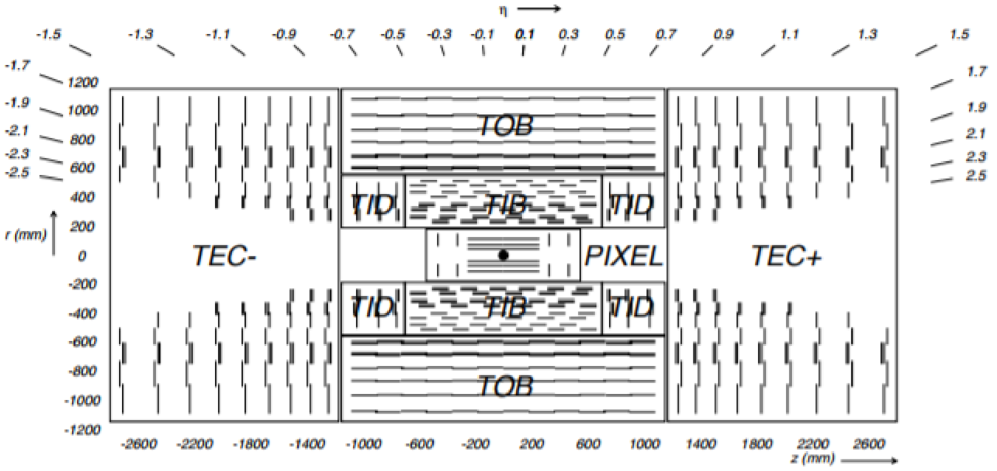


Figure 2.4: A schematic view of the CMS Inner Tracker, showing both pixel and strip components

- Pixel Tracker

Due to the high particle flux in the innermost region of the tracker pixel detectors are used to reduce the occupancy. The pixel tracker has three barrel layers (BPix) situated at radius of 4.4, 7.3 and 10.2 cm. These are complemented by two disks of pixel modules (FPix) at either end, situated at ± 34.5 and ± 46.5 cm from the nominal interaction point in the z direction. The layout is designed to provide a minimum of two hits per track. A schematic of the pixel layers is shown below.

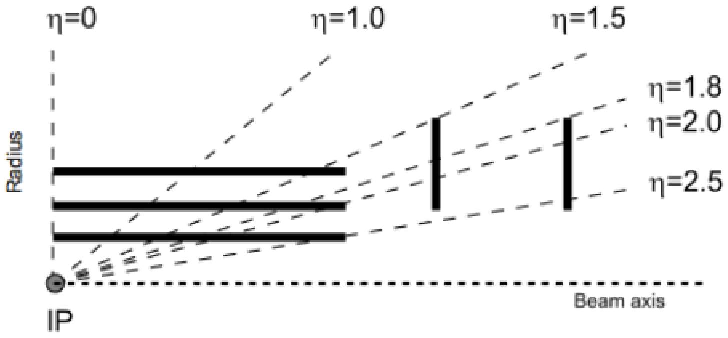


Figure 2.5: Illustration of the pixel detector in the r-z plane

Each pixel cell is 100 * 150 μm^2 in size, providing similar resolution in both the barrel and endcap regions. In total there are 66 million pixels, with a total surface area of $1m^2$. The high granularity means an occupancy of 10^{-4} per pixel per bunch crossing is expected. The Lorentz drift of electrons in the pixels results in the signal charge spreading across more than one pixel. With the analog pulse height being read out an interpolation is possible providing a spatial resolution of 15-20µm. This analog readout also aids the distinction between noise and signal hits.

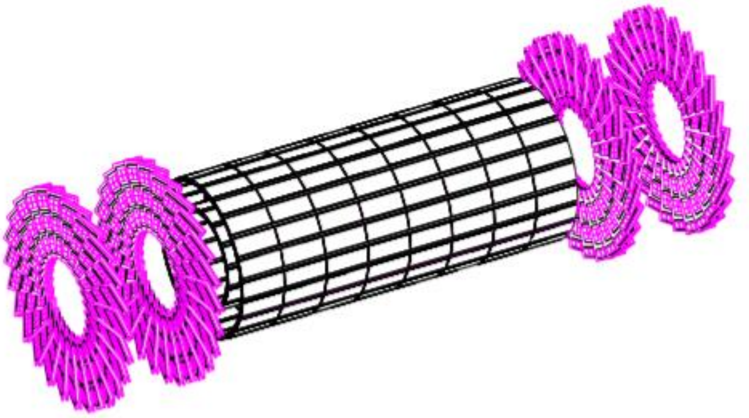


Figure 2.6: The CMS Pixel detector in its default configuration

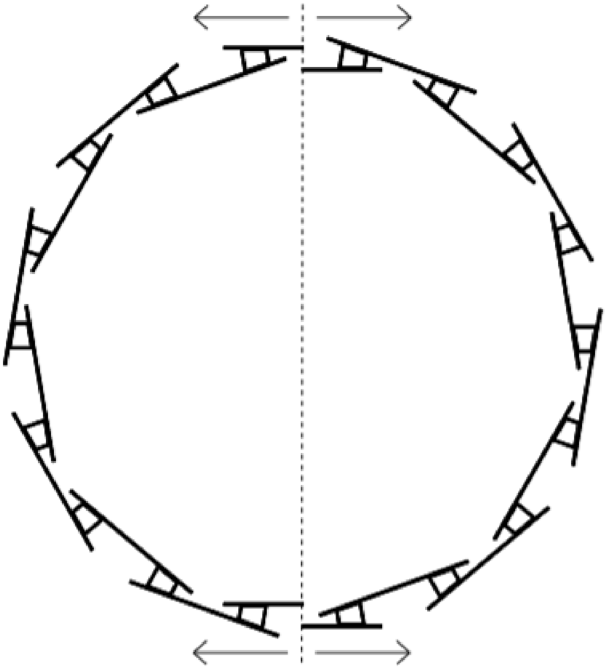


Figure 2.7: Example layer of the CMS pixel barrel detector

- Silicon Tracker

The silicon strip tracker surrounding the pixel detector is composed of three different subsystems. The Tracker Inner Barrel (TIB) and Disks (TID) are composed of four layers of strips in the barrel extending from a radius of 20 cm to a radius of 55 cm, with three disk layers at each end. The strips in the TIB have a pitch of 80 (120) µm in layers 1 and 2 (3 and 4), resulting in a single point resolution of 23 µm (35 µm).
The TIB/TID system is surrounded by the Tracker Outer Barrel (TOB). This contains six layers of strips with a pitch of at most 183 µm. The spatial point resolution is 35-53 _m. The system extends out to a radius of 116 cm, and ± 118 cm in z. Finally, the Tracker Endcaps (TEC) contain nine disks of silicon sensors with radial strips between 97 µm and 184 µm average pitch. This extends out to ± 282 cm in z, and has a radius of 22.5 to 113.5 cm. The occupancy expected in the inner barrel regions is approximately 2-3%. This falls to ~1% in the outer regions.
The operational temperature of the silicon tracker is -10°C, which is required in order to reduce the effects of noise. The total power dissipation in the tracker is nearly 60 kW. To maintain the operational temperature CMS uses a mono-phase liquid cooling system using perfluorohexane, C6F14. This is used both to refrigerate the pixel and silicon sensors, and to screen the tracker from the ECAL which operates at a temperature of 18±4 °C. C_6F_{14} has a low viscosity and high volatility making it ideal in the case of accidental leaks. The cooling system provides 77m3 per hour of C6F14, corresponding to a cooling capacity of 128 kW. In addition, the total tracker volume ($25m^3$) is flushed with pre-chilled nitrogen gas, at a rate of up to $25m^3$ per hour.

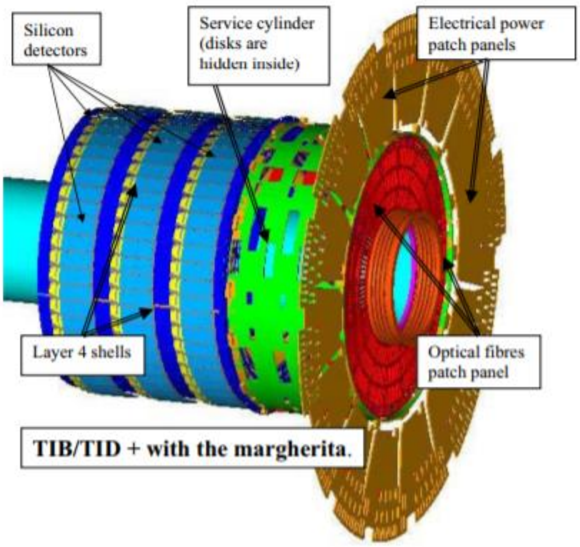


Figure 2.8: The Silicon Tracker Inner Barrel and Inner Disk detectors with service cylinder

2.2.4. The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a detector designed to measure the energy of particles produced in electromagnetic interactions, specifically electrons and photons. It is designed to be as hermetic and homogeneous as possible, and contains 61,200 lead tungstate (PbWO4) crystals in the barrel, and 7,324 crystals in each of two end caps on either end of the barrel. To facilitate the discrimination between neutral pions and photons in the endcap region, a preshower detector is placed in front of the endcaps. Particles striking the detector produce photons in the crystals that are collected using avalanche photodiodes in the barrel and vacuum photodiodes in the endcaps. Using high density crystals allows for a detector which is fast, has fine granularity in the $\eta - \phi$ plane, and is radiation hard. The ECAL central barrel covers a pseudorapidity range of $|\eta| < 1.479$. while the ECAL endcaps cover a range of $1.479 < |\eta| < 3.0$.

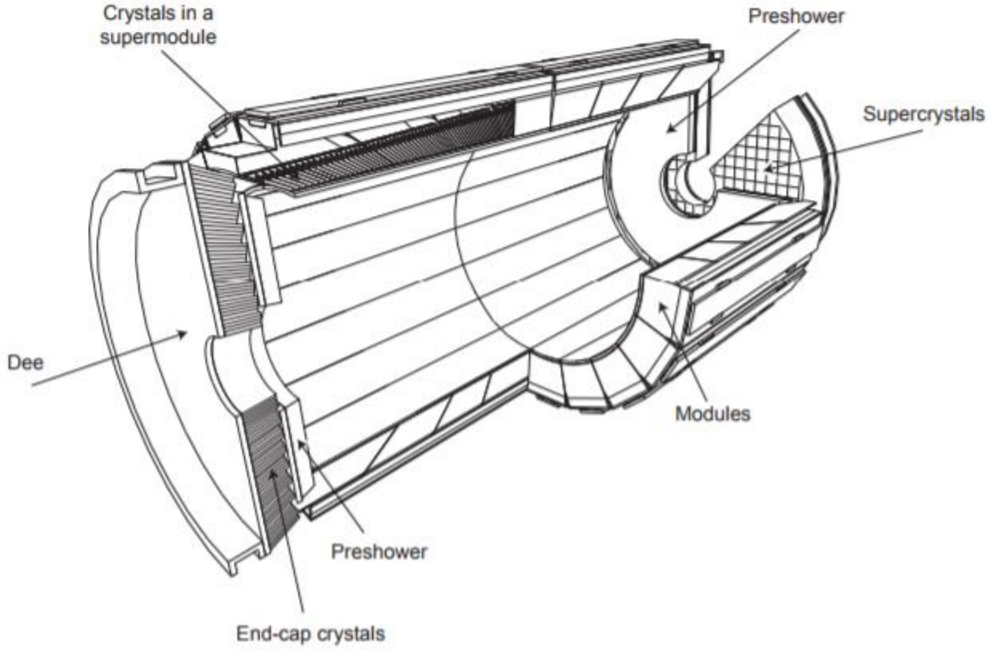


Figure 2.7: A schematic view of the CMS ECAL

The photodetectors used are the avalanche photodiodes for the barrel region and vacuum phototriodes for the endcaps. The use of high density crystals makes this calorimeter fast, with a fine granularity and good radiation resistance. Its good energy resolution (as provided by a homogeneous crystal calorimeter) enhances its chances to detect the decay of Higgs Boson to two photons.

The energy resolution for this Electromagnetic Calorimeter system can be described by equation:

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

where S is a stochastic term which depends on photostatistics, lateral shower containment and fluctuation in energy deposited in the preshower absorber. N is the noise term, with contributions coming from electronic noise, digitization noise and pileup. C is a constant term due to non-uniformity in the longitudinal light collection, intercalibration errors and leakage of energy from the back of the crystals.

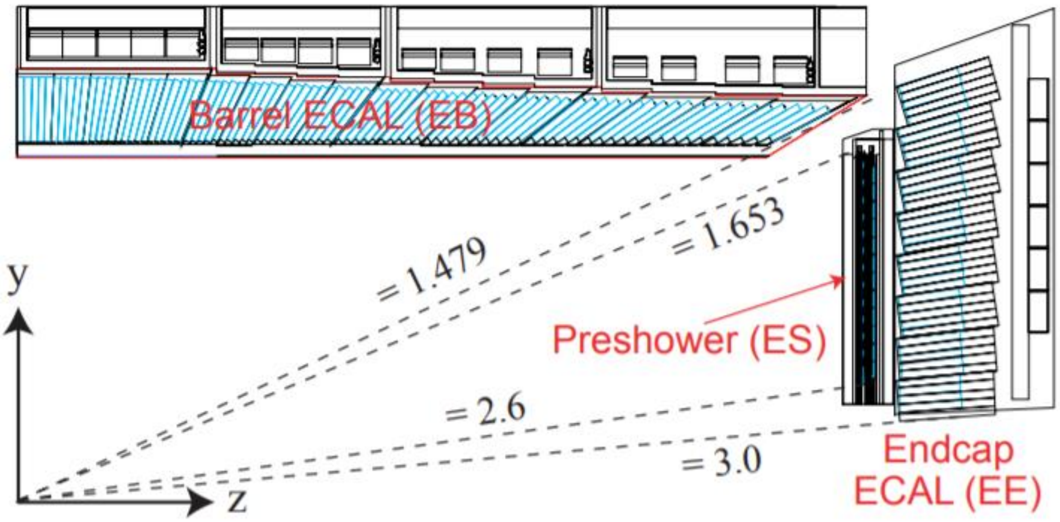


Figure 2.8: Transverse section of the ECAL

ECAL Preshower

The principle aim of the preshower is to identify the decays of neutral pions to photons which could mimic a $H \rightarrow \gamma\gamma$ signal if two neutral pions are incident in the forward region. The photons from a neutral pion decay can be very close together in the forward region and with the crystals alone they can look like one very high energy photon (as might be expected from a possible Higgs decay) rather than two lower energy photons. The preshower provides the extra resolution to tell if a high energy photon is actually two lower energy photons. The pre-shower is situated in the fiducial region $1.653 < |\eta| <$

2.6. It is a sampling calorimeter using a layer of lead radiators, to initiate electromagnetic showers, and silicon strip sensors, with a width of 2mm situated behind the radiators, to measure the energy deposited. There are two such layers of lead and strip sensors in order to provide a two-coordinate measurement of incident particles. The total width of the preshower is 20cm.

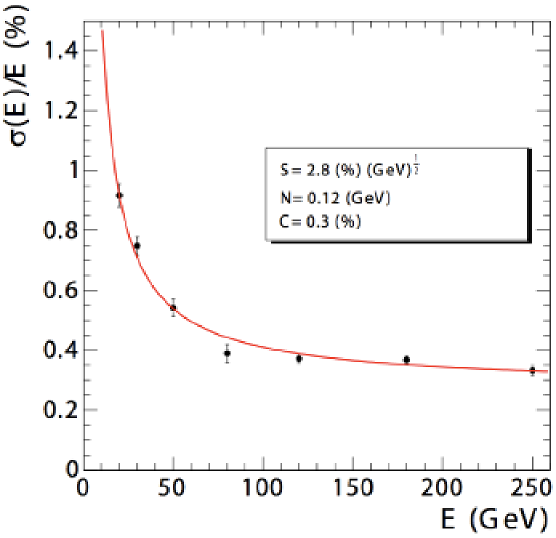


Figure 2.9: Energy resolution measured on a test beam of electrons

2.2.5. The Hadronic Calorimeter

The Hadron Calorimeters (HCAL) are very important for the measurement of hadron jets and transverse missing energy due to neutrinos or exotic particles. The Hadron Calorimeter Barrel and Endcaps are placed between the Electromagnetic Calorimeter and the inner of the magnet coil. This collocation limits the total amount of material which can be put to absorb the hadronic shower. This is the reason why an outer calorimeter is placed outside the solenoid. The η range covered by the Hadron Barrel (HB) calorimeter is up to $|\eta| < 1.3$. To extend this range to $|\eta| < 5.2$ a Forward Hadronic calorimeter is added outside the Barrel calorimeter Endcaps. The HB consists of 36 azimuthal wedges which forms two half barrels HB+ and HB-. Each Wedge is composed of absorber plates (one 40-mm-thick front steel plate, eight 50.5-mm-thick brass plates and one 75-mm-thick back steel plate) and active scintillating medium. The total absorber thickness is 5.82 interaction lengths (λ_I) at 90^0 and increases with polar angle θ as $1/\sin\theta$ and reaches a value of $10.6\lambda_I$ at $|\eta| = 1.3$. The plastic scintillator is divided into 16 η sectors, resulting in a segmentation $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The energy resolution of HCAL is:

$$\frac{\sigma_E}{E} \approx (65\%/\sqrt{E} \oplus 5\%)$$

in the barrel and

$$\frac{\sigma_E}{E} \approx (83\%/\sqrt{E} \oplus 5\%)$$

in the endcaps.

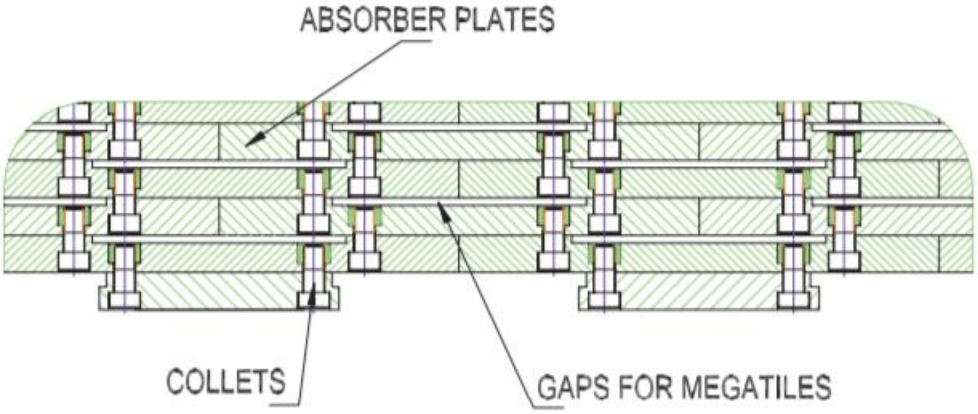


Figure 2.10: A schematic view of the HCAL Endcap detector.

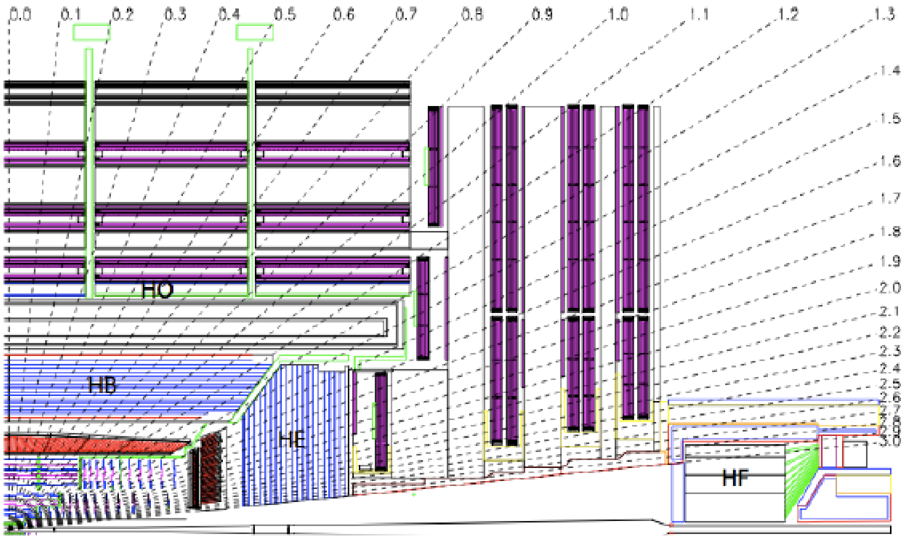


Figure 2.11: Longitudinal view of CMS. Marked are the barrel and endcap hadronic calorimeters (HB, HE) and the outer (HO) and forward (HF) sections.

2.2.6. The Muon System

One of the central themes in the design of CMS is the muon detection system. One of the key decay modes of the postulated Standard Model Higgs boson is a decay to ZZ or ZZ*, in which the Zs(Z*) then decay to muons. The resulting 4-muon final state is key because, in addition to the relative ease in reconstructing muons, muons suffer less radiation losses than electrons. This and other searches necessitated wide angular coverage for muon detection. The muon system has three functions: muon identification, muon momentum measurement, and triggering. To achieve these functions three different technologies are used to cope with the different background rates and magnetic fields the subdetectors have to withstand. These technologies are Drift Tube Chambers, Cathode Strip Chambers, and Resistive Plate Chambers. High momentum resolution and triggering are enabled by the high-field magnet and the flux-return yoke. The yoke also acts as an absorber of hadrons to aid the identification of muons. The muon system is the outermost detector at CMS, with the idea being that muons, which are minimum ionizing particles, will pass through the bulk of inner material minimally perturbed, while other particles will be contained within their respective detectors. As with other detectors, the muon system uses both a barrel and endcap design. The barrel, covering a pseudorapidity range of $|\eta| < 1.2$, contains drift tube chambers organized into four stations. These stations are interspersed among the layers of the flux return plates, meant to act as absorbent material for muon identification. The first three stations contain eight chambers (divided into two groups of four), oriented parallel to the beam axis to perform measurements in the $r - \phi$ bending plane, and an additional four chambers oriented orthogonally to the beam axis, which provide measurements along the z direction. The fourth station is without the z-measuring chambers. The chambers are separated as much as possible to provide the best angular resolution.

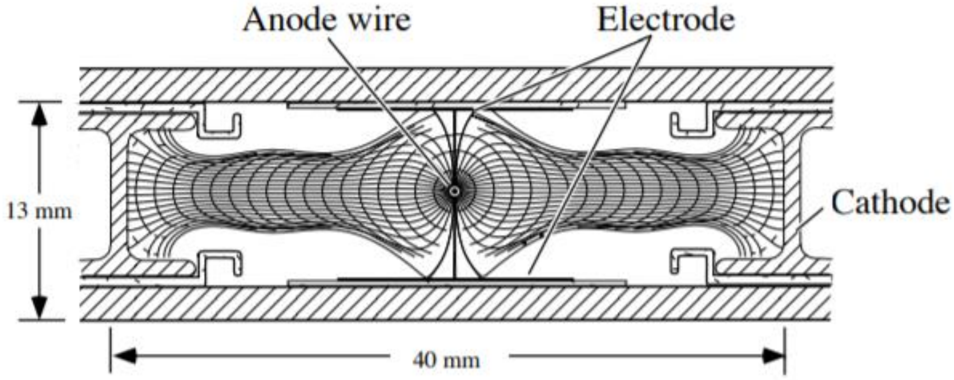


Figure 2.12: A schematic of a drift tube chamber with drift lines and isochrones

The endcaps of the muon system ($0.9 < |\eta| < 2.4$) use cathode strip chambers (CSC, Figure 2.20) to perform measurements. This is due to the higher expected rates of both muons and background, and the non-uniform magnetic field in this region. The CSCs provide fast response times with fine segmentation and resistance to radiation. Each endcap contains four stations of CSCs aligned perpendicularly to the beam line and interspersed among the flux return plates. The CSCs provide precision measurements in the $r - \phi$ bending plane. The anode wires of the CSCs run approximately perpendicular to the strips, and are read out in order to provide measurements of both η and beam-crossing time for the muon.

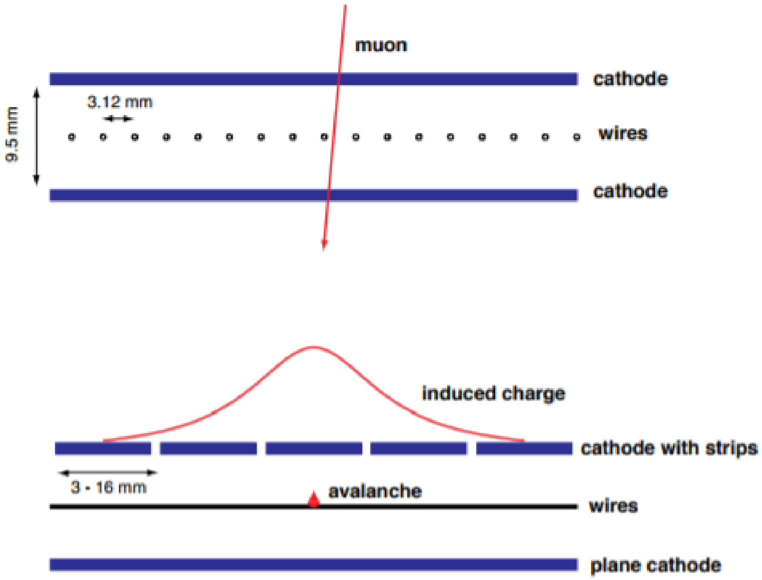


Figure 2.13: A schematic of a cathode strip chamber

An additional system utilizing resistive plate chambers (Fig. 2.14) is also installed in the muon system which acts as a dedicated muon trigger system. The RPCs are useful for triggering as they provide fast, highly-segmented measurements with a sharp P_T threshold over $|\eta| < 1.6$. RPCs are double-gap chambers which ensure good operation even at high rates. A total of six-barrel layers contain RPCs, two in each of the first two stations, and one in each of the last two. The redundancy in the first two station allows for the ability to trigger on low-P_T muons that may stop before reaching the outer stations. In the endcap region, a plane of RPCs is placed in each of the first three stations to allow the trigger to use coincidences between the RPCs and CSCs in order to reduce background and improve time resolution, as well as P_T resolution.

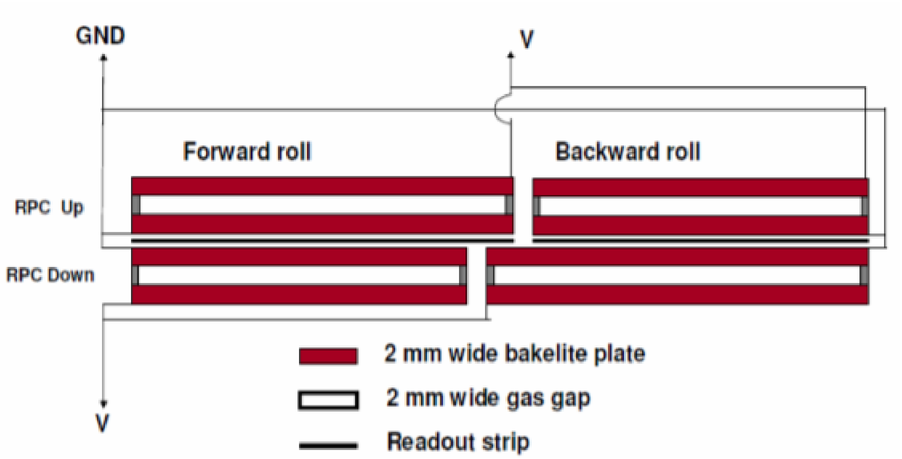


Figure 2.14: A schematic of a resistive plate chamber

2.2.7. Triggering and Data Acquisition

LHC provides high rate proton-proton collisions. There will be one bunch crossing every 25 ns (40 MHz). At a luminosity of L = 10^{34} cm^{-2} s^{-1} there will be an average of 20 interactions per bunch crossing. Since it is impossible to store such an amount of data (100 TByte/s), a drastic rate reduction must be achieved. The CMS trigger system is responsible for this rate reduction.

The triggering system at CMS utilizes a two-step approach:

- A set of custom-designed, programmable electronics called the Level-1 (L1) trigger, which reduces the event rate to tens of kHz.
- A software system operated on a filter farm containing a vast number of processors, called the High-Level Trigger (HLT), which reduces the event rate to approximately 100 Hz.

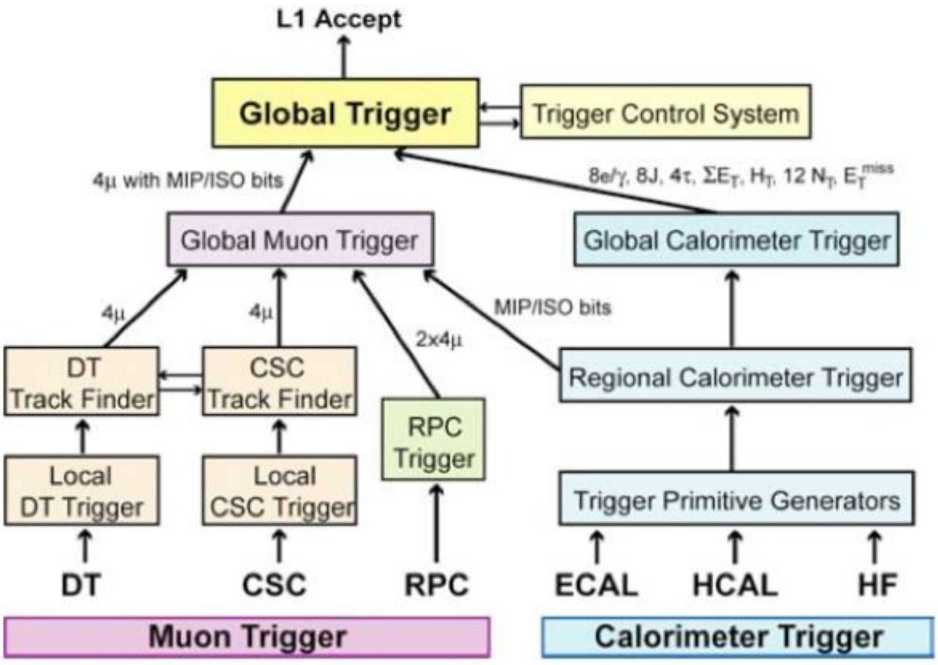


Figure 2.15: Diagram illustrating the workflow of L1 Trigger

The Level-1 Trigger System

The Level-1 trigger system is a hardware-based triggering system. Custom-made integrated circuits and electronics provide low-level event and object reconstruction mechanisms. Muons are reconstructed from hits in the muon system while electrons and jets are reconstructed from energy depositions in the calorimeter above a certain threshold. The Level-1 trigger system is very fast because it evaluates the detector information only locally. There are no correlations between the subdetectors and the time-consuming track reconstruction in the inner tracking system is not used.

The High-Level Trigger System

At a rate of 100 kHz the full detector information for a given event can be transferred through the readout system to a local computing farm. There, the events are processed by the High-Level Trigger system. All sub-detector channels are read out for the High-Level Trigger and the information between the sub-detectors is correlated. The time-consuming tracking which is not used in the Level-1 trigger is available and used in the High-Level Trigger. The event reconstruction algorithms that are used by the High-Level Trigger are more sophisticated than those of the Level-1 Trigger and close to the offline event reconstruction. For a given run a certain list of trigger paths is defined which specify different sets of software modules for the event reconstruction and filters which decide based on pre-defined conditions if an event is written out to the storage system or is discarded. An example for a top quark physics specific trigger path is discussed in Section 5.1.1. At an event rate of 300 Hz events are selected by the High-

Level Trigger and then transferred for full event reconstruction and permanent storage to the Cern Tier-0 centre, a computing infrastructure.

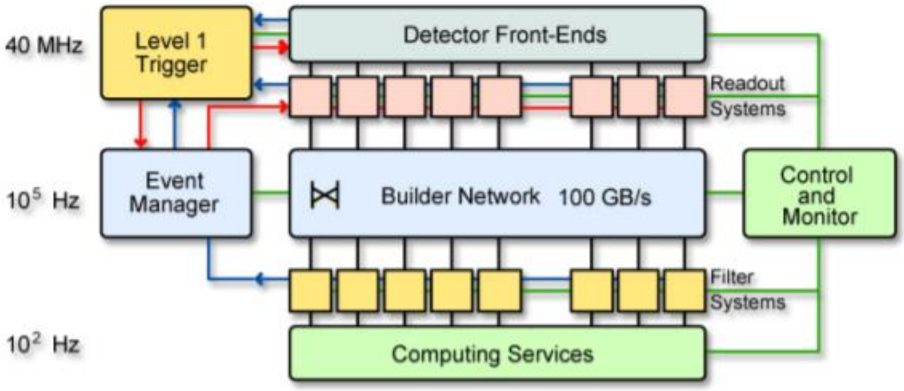


Figure 2.16: Architecture of the CMS DAQ System

2.2.8. Computing at CMS

CMS has adopted a distributed computing model in order to cope with the requirements for storage, processing and analysis of the huge amount of data the experiment will collect. In the CMS computing model, resources are geographically distributed and operated by means of Grid Software. The CMS Event Data Model (EDM) is centered around the concept of an Event. This is a C++ container that provides access to all information recorded from an individual bunch crossing, as well access to new data derived from it. This includes raw digitized data such as deposits in the ECAL crystals, subdetector base elements such as ECAL superclusters, and higher level elements such as electron objects. The origins of raw data are stored, as well as the provenance of all derived data types. In addition to physics data, detector conditions such as temperature, calibration constants and alignment conditions are also stored.

2.2.9. CMS Data Model

Various levels of data format are available in CMS which differ primarily in the amount of information available in each; in each case their physical format is still the ROOT file type. This is to allow flexibility while providing data formats small enough to allow for practical storage at many storage centers. CMS makes use of several different event formats:

- **RAW format**: contains fully recorded detector readout information along with trigger information. The RAW format is used for offline reconstruction to convert detector information into physics objects, such as electrons, muons, photons, and tracks. The RAW data is permanently archived in safe storage, designed to occupy 1.5 MB / event. An extension of the RAW format is also used for simulated datasets, which occupies 2 MB / event, due to additional Monte Carlo truth information.
- **RECO format**: Reconstructed data produced by applying pattern recognition modules to RAW data. The result is physics objects which may be used in physics analyses in addition to reconstructed inputs (detector hits, energy clusters) to these objects.
- **AOD format**: Analysis Object Data, which is a compact form of the RECO format. It is meant to be easily transportable and storable, while providing all the necessary information for a typical physics analysis. An extension of this format, AODSIM, is used for simulated datasets, and provides pertinent Monte Carlo truth information needed by the average user.
- **DQM format**: Data Quality Monitoring, which is information used to determine the quality of both simulated events and real collision data by comparing the information, in histogram form, to some centrally defined reference.

Chapter 3

Trigger Selection and Event Reconstruction

3.1. Trigger Selection and Datasets

For current Analysis, the data of 2016 in Integrated Lumonisity of 37 fb^{-1} was used.

The Monte Carlo sample of powheg-pythia8 was used as nominal signal file. Additional Monte Carlo simulation samples are used to estimate systematic uncertainties due to variations in the factorization scale and in the matrix element to parton shower matching threshold. Finally, the QCD multijet background was estimated using a data-driven technique.

The triggers used for selecting events require high P_T , multijet events with at least two jets having medium to high b-tag discriminator selection criteria.

Details about Datasets and Triggers are presented later, in 4.3.1.

3.2. Object Reconstruction

To perform physics analysis on CMS data the EDM framework produces several objects with close representations to their physical counterparts. Electron and muon objects represent the physical leptons. Neutrinos, or any non/weakly interacting particle which escapes the detector are closely related to the missing transverse energy objects available in the event. Tauons and hadonic particles manifest themselves as jet objects. Any charged particle can also be associated with track objects from the Tracker.

3.2.1.Kinematic Variables

For particles and reconstructed objects, we define some kinematic variables:

- Transverse momentum P_T

$$p_T = \sqrt{{p_x}^2 + {p_y}^2},$$

where px and py are the momentum projections of the particle/object on the axes perpendicular to the direction of the particle beams;

- Rapidity *y*

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z},$$

where pz is the component of the momentum of the particle/object along the axis of the particle beams and E is the energy;

- Pseudorapidity *η*

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

where θ is the azimuth angle measured with respect to the axis perpendicular to the horizontal plane passing through LHC. It can be shown that rapidity distributions, i. e. the number of particles per unit rapidity, dN/dy, are invariant under Lorentz boosts along the z direction. In the limit of momenta much larger than the mass of a particle, the rapidity converges to pseudorapidity:

$$\lim_{|\vec{p}| \gg m} y = \eta.$$

The pseudorapidity of a particle is a purely geometrical quantity, it only depends on the polar angle θ, but not on the particle mass.

- Jet invariant mass m_{jet}

$$m_{jet} = \sqrt{\left(\sum_i P_i\right)^2},$$

where Pi are the four-vectors of all i particles clustered into the jet, assuming a pion mass for all charged hadrons.

For this analysis, a common practice will be followed; work in natural units where the reduced Planck constant ~ and the speed of light in vacuum c are set to 1:

$$\hbar = c = 1.$$

Factors of ħ and c can always be restored by dimensional analysis.

3.2.2. Hadronic States

Quarks and gluons produced in hard scattering of partons in pp collisions manifest themselves as hadronic jets. A detailed understanding of the jet energy calibration and resolution is of crucial importance and is a leading source of uncertainty for many analyses with jets in the final state. A brief description of CMS jet reconstruction algorithms, jet energy calibration techniques and the jet energy scale uncertainties are presented below.

3.2.2.1. Reconstruction of Hadronic Jets

Four types of jets are reconstructed at CMS depending on the input to the jet clustering algorithm: calorimeter jets, Jet-Plus-Track (JPT) jets, Particle-Flow (PF) jets, and track jets. Jets are reconstructed using the anti-kT clustering algorithm with the size parameter R = 0.5. To evaluate their performance, in Monte Carlo simulations, generated jets (GenJets) or particle jets are reconstructed as well by applying the same jet clustering algorithm to all stable generated particles.

Calorimeter jets are reconstructed using energy deposits in the calorimeter towers, where calorimeter tower consists of one or more hadronic calorimeter (HCAL) cells and the geometrically corresponding electromagnetic (ECAL) crystals. The Jet-Plus-Track algorithm exploits the excellent performance of the CMS tracking detectors to improve the P_T response and resolution of calorimeter jets. For each track in the jet, the average expected calorimeter energy is subtracted and the momentum measured in the tracker is added to the jet. For the tracks which are bent out of the jet cone due to magnetic field the momentum of the track is added to the jet. PF jets are reconstructed from the list of particles reconstructed using particle-flow algorithm. The jet momentum and spacial resolutions are improved with respect to the calorimeter jets, since the use of tracking detectors and excellent ECAL granularity allows to resolve and precisely measure charged hadrons and photons inside jets.

3.2.2.2. Energy Calibration of Hadronic Jets/Jet Energy Correction

Due to the non-uniform and non-linear response of the CMS calorimeters the jet energy measured in the detector is typically different from the corresponding particle jet energy. Furthermore, electronic noise and additional pp interactions in the same bunch crossing (event pile-up) leads to extra unwanted energy. The purpose of the jet energy correction is to relate, on average, the energy measured for the detector jet to the energy of the corresponding particle jet. CMS has developed a factorized multi-step procedure for the jet energy calibration (JEC). The correction is applied as a multiplicative factor to each component of the raw jet four momentum vector p raw µ as shown in Equation below:

$$p_\mu^{corrected} = p_\mu^{raw} \cdot C_{offset}(p_T^{raw}) \cdot C_{MC}(p_T^{'}, \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p_T^{''})$$

where P_T' is the transverse momentum of the jet after applying offset correction and P_T'' is the transverse momentum of the jet after all previous corrections. C_{offset} is the offset correction derived using the jet area method. For each event, an average P_T density ρ per unit area is estimated which characterizes the soft jet activity and is contamination of the underlying event, the electronic noise and the pile-up. The MC calibration, CMC, is based on the simulation and corrects the energy of the reconstructed jets such that it is equal to the energy of generated MC particle jets. It removes the bulk of the non-uniformity in η and the non-linearity in P_T. The residual corrections C_{rel} and C_{abs} for the relative and absolute energy scales, respectively, are derived using data driven method, using dijet and γ/Z+jets events, to account for the minor differences between data and simulation.

3.2.2.3. CSV Algorithm

The presence of a secondary vertex and the kinematic variables associated with this vertex can be used to discriminate between b and non-b jets. Two of these variables are the flight distance and direction, using the vector between primary and secondary vertices. The other variables are related to various properties of the system of associated secondary tracks such as the multiplicity, the mass, or the energy. Secondary-vertex candidates must meet the following requirements to enhance the b purity:

- secondary vertices must share less than 65% of their associated tracks with the primary vertex and the significance of the radial distance between the two vertices has to exceed 3σ
- secondary vertex candidates with a radial distance of more than 2.5 cm with respect to the primary vertex, with masses compatible with the mass of K0 or exceeding 6.5 GeV/c 2 are rejected, reducing the contamination by vertices corresponding to the interactions of particles with the detector material and by decays of long-lived mesons
- the flight direction of each candidate has to be within a cone of ΔR < 0.5 around the jet direction.

The Combined Secondary Vertex (CSV) algorithm – used in this analysis – involves the use of secondary vertices, together with track-based lifetime information. By using these additional variables, the CSV algorithm provides discrimination also in cases when no secondary vertices are found, increasing the maximum efficiency with respect to the so-called “Simple Secondary Vertex” algorithms – these using only the flight distance as discriminating variable. In many cases, tracks with an impact parameter significance SIP – that is the ratio of the IP to its estimated uncertainty – that is > 2 can be combined into a “pseudo vertex”, allowing the computation of a subset of secondary-vertex-based quantities even without an actual vertex fit. Finally, when even this is not possible, a “no vertex” category reverts to track-based variables and the discrimination is conducted in a way similar to that of the track-based algorithms. Therefore, the CSV algorithm uses the following set of variables with high discriminating power and low correlations (of course, in the “no vertex” category only the last two variables are available): • the vertex category (real, “pseudo,” or “no vertex”);

- the flight distance significance in the transverse plane (“2D”);

- the vertex mass;
- the number of tracks at the vertex;
- the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- the pseudorapidities of the tracks at the vertex with respect to the jet axis;
- the 2D IP significance of the first track that raises the invariant mass above the charm threshold of 1.5 GeV/c 2 (tracks are ordered by decreasing IP significance and the mass of the system is recalculated after adding each track);
- the number of tracks in the jet;
- the 3D IP significances for each track in the jet.

Two likelihood ratios are built from these variables. They are used to discriminate between b and c jets and between b and light-parton jets. They are combined with prior weights of 0.25 and 0.75, respectively. The CSV algorithm has evolved into the CSVv2 (Combined Secondary Vertex version 2) algorithm in Run 2. Just like the CSV, the CSVv2 is based on secondary vertex and track-based lifetime information. Despite this, the new version of the CSV algorithm combines the variables using a neural network instead of a likelihood ratio to produce a discriminator csv, and the secondary vertex information is obtained with the Inclusive Vertex Finder algorithm. The operating point values for the loose, medium and tight tagging criteria are set to 0.460, 0.800, 0.935, respectively. The b-tagging efficiency measured in MC events is corrected using scale factors in order to reproduce the efficiency measure in data.

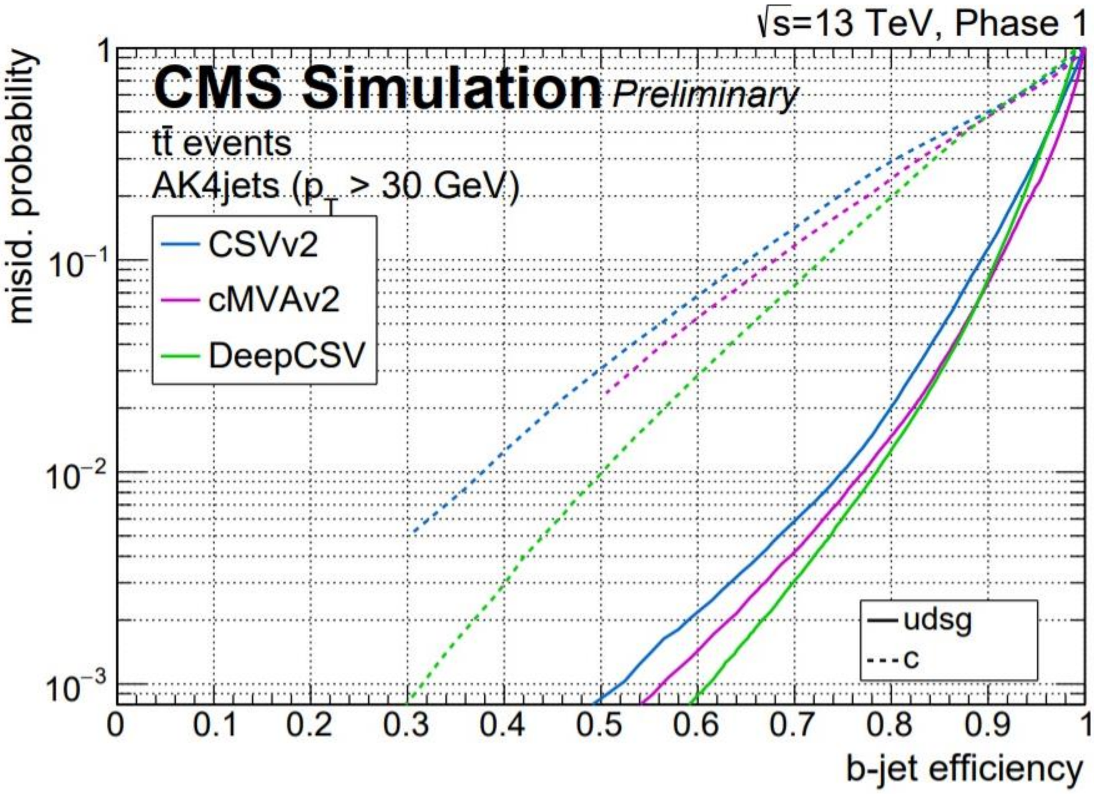


Fig. 3.1: Performance of the b jet identification algorithms demonstrating the probability for non-b jets to be misidentified as b jet, as a function of the efficiency to correctly identify b jets. The absolute performance in this figure serves as an illustration since the b jet identification efficiency depends on the pT and η distribution of the jets in the topology as well as the amount of b jets from gluon splitting in the sample.

3.2.3. Reconstruction of Missing Transverse Energy

The missing transverse momentum, E_{miss}^{T}, is reconstructed as the negative of the vector sum of the transverse momenta of all final-state particles reconstructed in the detector. There are three distinct algorithms developed in CMS to reconstruct E_{miss}^{T}: PF E_{miss}^{T}, Calo E_{miss}^{T}, and TC E_{miss}^{T}. PF E_{miss}^{T} is calculated from the reconstructed PF particles, Calo E_{miss}^{T} is calculated using the energies contained in calorimeter towers and their direction, relative to the center of the detector, to define pseudo-particles, and TC E_{miss}^{T} is based on Calo E_{miss}^{T}, but the response and resolution is improved using tracks reconstructed in the inner tracker. A three-step correction is devised to remove the bias in the E_{miss}^{T} scale due to the non-linearity of the response of the calorimeter for neutral and charged hadrons, caused by event pile-up, large bending of low P_T tracks due to strong magnetic field in CMS, etc.. The correction procedure relies on the fact that E_{miss}^{T} can be factorized into contributions from jets, isolated high P_T photons, electrons, muons and unclustered energies. The jet energy scale corrections are propagated E_{miss}^{T} using the so called "type-I" correction:

$$E_{x,y}^{miss,corrected} = E_{x,y}^{miss,raw} - \sum_{i} \left(p_{i,(x,y)}^{corrected\ jets} - p_{i,(x,y)}^{raw\ jets} \right)$$

In order to correct for the soft jets below the threshold used for "type-I" correction and energy deposits not clustered in any jet, a second correction can be applied to the unclustered energy, which is called "type-II" correction. This correction is obtained from Z→ee events. To reduce the dependency of E_{miss}^{T} on event pile-up, a so called "type-0" correction has been developed only for PF E_{miss}^{T}. For each pile-up vertex the expected missing neutral momentum is calculated using an improved PF candidate to vertex association technique and added it vectorially to PF E_{miss}.

3.3. Particle-Flow Algorithm

Stable particles are identified with the Particle Flow (PF) algorithm that reconstructs each individual particle with an optimized combination of information from the various elements of the CMS detector. The energy of photons is directly obtained from the ECAL measurement, corrected for zero-suppression effects. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as determined by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is determined from a combination of their momentum measured in the tracker and the matching ECAL and HCAL energy deposits, corrected for zero-suppression effects and for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energy. For the measurements presented jets are reconstructed by clustering PF particles using the anti-kT (AK) jet clustering algorithm, with a distance parameter R = 0.5 (AK5 jets). For the **boosted topologies**, jets are clustered with a larger opening angle corresponding to R = 0.8 (AK8 jets). When clustering the particles in jets, isolated electrons and muons as well as charged particles associated with other interaction vertices are removed. Jet momentum is determined as the vectoral sum of all particle momenta in the jet. Jet energies are calibrated to correct for the different detector response as a function of the transverse momentum and pseudorapidity of the jets. Furthermore, an offset correction is applied to jet energies to consider the contribution from additional proton-proton interactions within the same bunch crossing. Jets should lie within the tracker acceptance, hence pseudorapidity $|\eta| < 2.4$, and have a transverse momentum exceeding 20 GeV.

The flavor for jets in the simulated events is determined by re-clustering the jet constituents including also the generator-level hadrons and partons. The re-clustering is performed in such a way that the re-clustered jet four-momenta are identical to the original jets. The jet flavor is then determined based on the flavor of the clustered hadrons (or partons) inside a jet giving priority to the b flavor when at least one b hadron is present. In the absence of a clustered b hadron, priority is given to the c flavor in case a c hadron is found. If there are no b and c hadrons clustered in the jet, it is considered as light flavor unless a b (or c) quark is clustered in the jet in which case the jet will be considered as b (or c) jet.

Particle Flow Event Reconstruction

The particle-flow event reconstruction algorithm aims at reconstructing all stable particles in the event by combining information from all CMS sub-detectors. The algorithm optimizes the determination of particle types, directions and their energies. The resulting list of particles are then used to reconstruct higher level objects such as jets, taus, missing transverse energy, to compute charged lepton and photon isolation, etc. The basic elements of the particle-flow event reconstruction are the charged particle tracks reconstructed in the central tracker and the energy clusters reconstructed in electromagnetic and hadronic calorimeters. The charged particle tracks are reconstructed using an iterative tracking strategy, with both a high efficiency and a low fake rate for charged particle momentum as low as 150 MeV/c. The energy clustering is performed in each sub-detector of the calorimeters separately using a specific clustering algorithm, developed for particle-flow event reconstruction, which aims for a high detection efficiency even for low energy particles and separation of close energy deposits. These basic elements are then connected to each other using a link algorithm to fully reconstruct each single particle, while removing any possible double counting from different detectors. The algorithm produces “blocks” of elements linked directly or indirectly. The particle-flow algorithm is finally used to reconstruct and identify a set of particles from each block of elements. Charged hadrons are reconstructed from the tracks in the central tracker. Photons and neutral hadrons are reconstructed from energy clusters in calorimeters. Clusters separated from the extrapolated position of tracks in the calorimeters constitute a clear signature of these neutral particles. A neutral particle overlapping with charged particles in the calorimeters can be detected as a calorimeter energy excess with respect to the sum of the associated track momenta. The resulting list of reconstructed particles constitute a global description of each event, available for subsequent physics analysis.

3.4. B-jet Identification

The properties of the b-hadrons are used to identify hadronic jets originating from the fragmentation of b-quarks. These hadrons have relatively large masses, long lifetimes, and daughter particles with hard momentum spectra. Their semileptonic decays can be exploited as well. The reconstruction and identification of a set of particles from each block of elements is finally performed by the particle-flow algorithm. The resulting list of reconstructed particles constitutes a global description of each event, available for subsequent physics analysis.

3.4.1. CMS Algorithms

CMS has developed a variety of algorithms to identify b-quarks based on variables such as the impact parameter of charged particle tracks, the properties of reconstructed decay vertices, and the presence of a lepton, or the combination of the above information]. Each of these algorithms produces a single discriminator value for each jet. The minimum thresholds on these discriminators define loose (“L"), medium (“M"), and tight (“T") working points corresponding to the mis-identification probability for light parton jets of approximately 10%, 1%, and 0.1%, respectively, at an average jet pT of 80 GeV/c. The impact parameter (IP) of a track with respect to the primary vertex is calculated in three dimensions by taking the advantage of the excellent resolution of the pixel detector along the z axis. The sign of the IP is defined as the sign of the scalar product of the vector pointing from the primary vertex to the point of closest approach with the jet direction. While the IP values of the tracks originating from the decay of particles traveling along the jet axis tend to have positive values, the IP of prompt tracks can have positive or negative values. The impact parameter significance SIP, defined as the ratio of the IP to its estimated uncertainty, is used as a discriminating observable. The simplest algorithm based on the track impact parameter is called Track Counting (TC) algorithm which sorts tracks in a jet by decreasing values of IP significance. The Track Counting High Efficiency (TCHE) and Track Counting High Purity (TCHP) algorithms use the SIP of second and third ranked track as the discriminator value. The IP information of several tracks in a jet are also combined to provide better discriminating power. The Jet Probability (JP) algorithm uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex. The Jet B Probability (JBP) algorithm gives more weight to the tracks with the highest IP significance, up to a maximum of four such tracks, which matches the average number of reconstructed charged particles from the b-hadron decays. The presence of a secondary vertex provides the most powerful discrimination between b and non-b jets. The kinematic variables of the secondary vertex such as flight distance, direction, track multiplicity, mass or the energy are used in the b-tagging algorithms. The Simple Secondary Vertex (SSV) algorithm uses the significance of the flight distance, the ratio of flight distance to its estimated uncertainty, as the discriminating variable. A more complex algorithm, the Combined Secondary Vertex (CSV) algorithm, involves the use of secondary vertices together with track based lifetime information to provide the most efficient discrimination between b and non-b jets.

3.4.2. Reconstructed objects used in b-jet identification

Jets are clustered from objects reconstructed by the particle-flow algorithm. This algorithm combines information from all subdetectors to create a consistent set of reconstructed particles for each event. The particles are then clustered into jets using the anti-kT clustering algorithm with a distance parameter of 0.5. The raw jet energies are corrected to obtain a uniform response in η and an absolute calibration in pT. Although particle-flow jets are used as the default, the b-jet tagging algorithms can be applied to jets clustered from other reconstructed objects. Each algorithm described in the next section uses the measured kinematic properties of charged particles, including identified leptons, in a jet. The trajectories of these particles are reconstructed in the CMS tracking system in an iterative procedure using a standard Kalman filterbased method. Details on the pattern recognition, the track-parameter estimation. A “global” muon reconstruction, using information from multiple detector systems, is achieved by first reconstructing a muon track in the muon chambers. This is then matched to a track measured in the silicon tracker. A refit is then performed using the measurements on both tracks. Primary vertex candidates are selected by clustering reconstructed tracks based on the z coordinate of their closest approach to the beam line. An adaptive vertex fit is then used to estimate the vertex position using a sample of tracks compatible with originating from the interaction region. Among the primary vertices found in this way, the one with the highest ∑(p track T) 2 is selected as a candidate for the origin of the hard interaction, where the p track T are the transverse momenta of the tracks associated to the vertex. The b-jet tagging algorithms require a sample of well-reconstructed tracks of high purity. Specific requirements are imposed in addition to the selection applied in the tracking step. The fraction of misreconstructed or poorly reconstructed tracks is reduced by requiring a transverse momentum of at least 1 GeV/c. At least eight hits must be associated with the track. To ensure a good fit, χ 2/n.d.o.f. < 5 is required, where n.d.o.f. stands for the number of degrees of freedom in the fit. At least two hits are required in the pixel system since track measurements in the innermost layers provide most of the discriminating power. A loose selection on the track impact parameters is used to further increase the fraction of well-reconstructed tracks and to reduce the contamination by decay products of long-lived particles, e.g. neutral kaons. The impact parameters dxy and dz are defined as the transverse and longitudinal distance to the primary vertex at the point of closest approach in the transverse plane. Their absolute values must be smaller than 0.2 cm and 17 cm, respectively. Tracks are associated to jets in a cone $\Delta R < 0.5$ around the jet axis, where the jet axis is defined by the primary vertex and the direction of the jet momentum. The distance of a track to the jet axis is defined as the distance of closest approach of the track to the axis. In order to reject tracks from pileup this quantity is required to be less than 700 µm. The point of closest approach must be within 5 cm of the primary vertex. This sample of associated tracks is the basis for all algorithms that use impact parameters for discrimination. Properties of the tracks and the average multiplicity after the selection (except for the variable plotted) are shown in Fig. 1. The uncertainties shown in this and all following figures are statistical unless otherwise stated. The data were recorded with a prescaled jet trigger in the second part of 2011 when the number of pileup events was highest. The jet pT threshold was 60 GeV/c. The distributions show satisfactory agreement with the expectations from simulation. The track multiplicity and the lower part of the momentum spectrum are particularly sensitive to the modelling of the particle multiplicity and kinematics by the Monte Carlo generator, as are other distributions such as the number of hits in the innermost pixel layers. Detector effects that are not modelled by the simulation, such as the dynamic readout inefficiency in the pixel system, can also contribute to the remaining discrepancies. The b jets in these events tend to be close in space and can be inadvertently merged by the clustering algorithm, resulting in a higher average track multiplicity per jet. The combinatorial complexity of the reconstruction of the decay points (secondary vertices) of b or c hadrons is more challenging in the presence of multiple proton-proton interactions. In order to minimize this complexity a different track selection is applied. Tracks must be within a cone of $\Delta R = 0.3$ around the jet axis with a maximal distance to this axis of 0.2 cm and pass a “high-purity” criterion. The “high-purity” criterion uses the normalized χ 2 of the track fit, the track length, and impact parameter information to optimize the purity for each of the iterations in track reconstruction. The vertex finding procedure begins with tracks defined by this selection and proceeds iteratively. A vertex candidate is identified by applying an adaptive vertex fit, which is robust in the presence of outliers. The fit estimates the vertex position and assigns a weight between 0 and 1 to each track based on its compatibility with the vertex. All tracks with weights > 0.5 are then removed from the sample. The fit procedure is repeated until no new vertex candidate can be found. In the first iteration the interaction region is used as a constraint in order to identify the prompt tracks in the jet. The subsequent iterations produce decay vertex candidates.

3.4.3. B-tagging in Boosted Topology

High-mass resonances with a final state containing b quarks are predicted by various models of new physics. They may decay into top-quark pairs or Higgs bosons, and if they have a large enough momentum (” boosted topologies”), their decay products are very collimated, resulting in a small angular distance ΔR between them, and ending up clustered in a single fat jet. Boosted topologies are usually reconstructed and interpreted using jet substructure reconstruction methods such as top/W/Z-tagging algorithms. Algorithms of b-tagging in the jet substructure can significantly improve the sensitivity of these methods.

3.4.3.1. b-tagging in jet substructure

One important reconstruction parameter is the size of the jet, which needs to be optimised to include all decay products, depending on the jet pT . Two cases have been studied in detail: for top-tagging, the use of the HEPTopTaggeralgorithm, which is based on Cambridge/Aachen jets of size R = 1.5 (CA15), is investigated. The fat-jet substructure is identified by undoing the CA algorithm clustering. For Higgs-tagging, the focus is on CA jets of size R=0.8 and the jet substructure is described by pruned jets. Algorithms of b-tagging can then be applied on the fat jet or on its substructure components, the second option giving the best performance.

3.4.3.2. Performance measurement

Measurement of b-tagging efficiency in boosted topologies is challenging, and needs specific treatment since results on standard jets are not necessarily applicable to boosted objects. For Higgs-tagging, efficiency is measured using LT method on different control samples to study the performance of b-tagging both on fat jets and subjets. The agreement found between data and simulation is compatible with what is observed in non-boosted topologies. A modified implementation of the FTC method has been developed to measure the btagging efficiency in boosted top-quark events and results show that the simulation reproduces the b-tagging efficiencies in data equally well in boosted and in nonboosted top-quark events.

3.4.4. Soft Drop Technique

Here a tagging/grooming method called soft drop declustering is described. Like any grooming method, soft drop declustering removes wide-angle soft radiation from a jet in order to mitigate the effects of contamination from initial state radiation (ISR), underlying event (UE), and multiple hadron scattering (pileup). Given a jet of radius R_0 with only two constituents, the soft drop procedure removes the softer constituent unless

$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$

where pti are the transverse momenta of the constituents with respect to the beam, Δ R12 is their distance in the rapidity-azimuth plane, zcut is the soft drop threshold, and β is an angular exponent. By construction, Equation above fails for wideangle soft radiation. The degree of jet grooming is controlled by zcut and β, with β → ∞ returning back an ungroomed jet. For β > 0, soft drop declustering removes soft radiation from a jet while still maintaining a fraction (controlled by β) of the soft-collinear radiation. For β < 0, soft drop declustering can remove both soft and collinear radiation. For a jet to pass the soft drop procedure, it must have at least two constituents satisfying the Equation. Thus, in this regime, soft drop acts like a "tagger", since it vetoes jets that do not have two well-separated hard prongs.

The soft drop declustering procedure depends on two parameters, a soft threshold zcut and an angular exponent β, and is implemented as follows:

1. Break the jet j into two subjets. Label the resulting two subjets as j1 and j2.
2. If the subjets pass the soft drop condition equation above, then deem j to be the final soft-drop jet.
3. Otherwise, redefine j to be equal to subjet with larger pt and iterate the procedure.

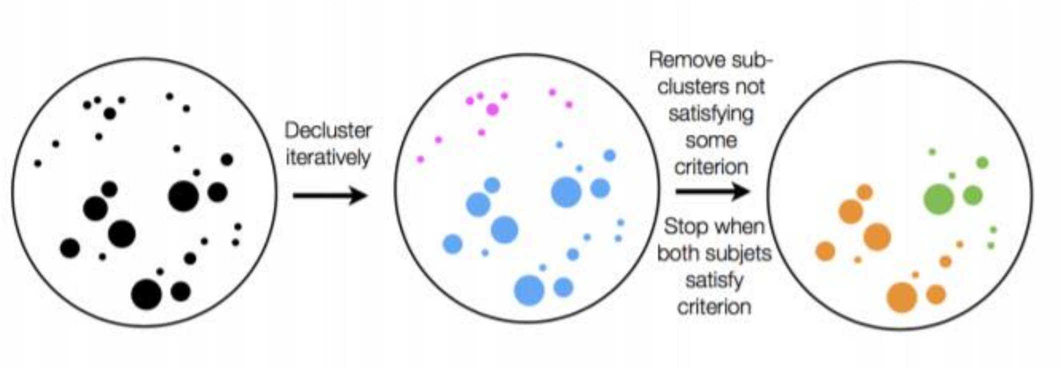


Fig. 3.1: Scheme of the soft drop declustering method.

This algorithm helps to get the true mass and eliminate the extra particles not coming from the original jet.

Chapter 4

Analysis

Aim of this analysis is the total and fiducial cross-section measurement of total top quark production in the all hadronic final state.

4.1. Observables and Measurement Techniques

4.1.1. Inclusive cross-section

The most inclusive observable to measure particle production is the inclusive (also: total) production cross section. The inclusive cross section is a measure of the production probability in the full kinematic phase space of the production processes. As phase space, the space of all possible final state configurations in top-quark events with all possible four-momenta consistent all conservation laws is implied. The first cross section measurements performed in the top-quark sector, both at the Tevatron and the LHC, were inclusive ttbar cross sections σ_{ttbar}.

All cross-section measurements start with a basic selection of candidate events to suppress background while retaining a large fraction of signal events in the data sample. After the event selection the simplest way to extract σ_{tt} is to perform a 'counting experiment':

$$\sigma_{t\bar{t}} = \frac{N_{\text{top}} - N_{\text{bkg}}}{\int \mathcal{L}\,\mathrm{d}t \cdot \epsilon}$$

where N_{top} and N_{bkg} are the number of top-quark events and background events, $\int Ldt$ is the integrated luminosity and ϵ the efficiency to detect top-quark events in the full phase space. While counting experiments are simple and robust, the need for absolute predictions of the signal efficiency and background level limits their precision. More precise cross section results can be obtained by exploiting the kinematic properties of the final state particles. The shapes of kinematic distributions are determined for the signal and all background processes and stored in discretized form in template histograms. The sum of template histograms for the signal and background processes is then fitted to the data using maximum-likelihood (ML) methods. Fitting kinematic distributions in signal-enriched and background-enriched regions simultaneously allows for better constraints on the background level, resulting in reduced statistical uncertainty of the result. At the same time additional systematic uncertainties arise due to the limited knowledge of the shapes of kinematic distributions. In more sophisticated fitting procedures, also the shapes of kinematic distributions are allowed to vary within their uncertainties. The top-quark production cross section and its uncertainties may then be determined either from a multi-parameter profile likelihood ratio fit, or the uncertainties are estimated using pseudo-experiments.

4.1.2. Fiducial Cross-Section

The need to know the absolute efficiency makes measurements of inclusive cross sections model-dependent. The efficiency ϵ in Equation stated in previous section, may be factorized into the detector acceptance and the detection efficiency of final state particles within the detector acceptance. While the detection efficiency can be calibrated using data to high accuracy, a determination of the detector acceptance, i.e. the ratio of detectable events to all events, requires an (often large) extrapolation to the full phase space of the ttbar final state. The extrapolation is usually performed using simulated data samples and hence depends on the MC model on which the simulation is based. This model dependence can be reduced by measuring the cross section in a restricted (“fiducial") region of the phase space that closely resembles the detector acceptance. Typical phase space requirements include the detector's pseudorapidity range and a minimum transverse momentum of analysis objects. The fiducial phase space is usually defined in the MC simulation on the particle level, after the particles hadronize but before they decay. The fiducial cross section can then be extrapolated to the full phase space by employing the predictions of different MC models.

4.2. Statistical Methods

Analysis of top-quark events often relies on advanced statistical methods, many of which are based on the maximum-likelihood (ML) principle. Sophisticated statistical methods are employed in several analysis steps, in order to maximize the precision of measurements or the sensitivity of searches for BSM physics. Examples include the selection of physics objects and entire events, the classification of the selected events as either signal or background, and the extraction of physics information such as model parameters from the data.

4.2.1. Maximum-Likelihood Method

A typical task in physics data analysis is to extract model parameters from data. The ML method solves this task by first constructing the likelihood function $\mathcal{L}(\vec{\mu})$ as the product probability densities $P\ (\vec{x_i}|\vec{\mu})$ for single measurements $\vec{x_i}$ given the true parameter vector $\vec{\mu}$:

$$\mathcal{L}(\vec{\mu}) = \prod_i P(\vec{x}_i|\vec{\mu})$$

The ML estimator of $\vec{\mu}$ is the maximum of the likelihood function, usually determined by minimizing -ln $\mathcal{L}(\vec{\mu})$. For discretized (“binned") data distributions, P is the Poisson distribution of the number of events in each bin given the number of events expected from the model. The model expectation is usually obtained from simulated data and represented as bins of a template histogram. The model parameters $\vec{\mu}$ estimated with the ML method are the normalizations of the histograms, which are in turn proportional to the total number of events expected from the model. Unbinned data distributions can be fitted with continuous probability density distributions, e.g. constructed by kernel-density estimates.

In more sophisticated ML models used with binned data further parameters are added that describe the influence of systematic uncertainties on the normalizations and shapes of the template histograms. The model parameters are then split into the parameter(s) of interest” $\vec{\beta}$ and additional “nuisance parameters” $\vec{\delta}$. In a Bayesian approach, a-priori knowledge, for example from auxiliary measurements, is used to constrain $\vec{\delta}$. To obtain an estimate of the parameters of interest and their uncertainties, the nuisance parameters can be either profiled or marginalized. Profiling means that the profile likelihood ratio

$$\lambda(\vec{\beta}) = \frac{\mathcal{L}(\vec{\beta}, \hat{\hat{\vec{\delta}}})}{\mathcal{L}(\hat{\vec{\beta}}, \hat{\vec{\delta}})}$$

is minimized instead of the original likelihood. The numerator of the profile likelihood ratio is the minimum of the likelihood function at a fixed value of $\vec{\beta}$, where the nuisance parameters assume the values $\vec{\delta}$, the denominator is the global minimum of the likelihood function, with parameter values $\vec{\beta}$ and $\vec{\delta}$. In the marginalization approach the likelihood function is integrated numerically, typically with MC methods. The parameters of interest are then extracted from the projections of the likelihood function on these parameters (“marginal distributions").

A frequentist method to deal with systematic uncertainties is to perform ensemble tests by drawing pseudo-experiments (also: “toy experiments", “MC experiments"). Many random variations of distributions are generated and the entire analysis chain is performed on each variation. The variance of the results is a measure of the uncertainty. Examples of software tools used in top-quark physics that include the above sophisticated ML methods are RooFit/RooStats shipped with ROOT, which were used for this analysis.

One way of interpreting measurements of top-quark properties is to compare the BSM physics prediction for an observable with the corresponding SM prediction. The statistical method applied in the comparison is called hypothesis test. First the null hypothesis H_0 (e.g. SM) and the alternative hypothesis H_1 (e.g. BSM) are formulated and a test statistic is constructed that is able to discriminate between H_0 and H_1. A popular choice of the test statistic is the ratio of likelihoods for the vector of measurements $\vec{x}$, given H_0 or H_1:

$$r(\vec{x}) = \frac{L(\vec{x}|H_0)}{L(\vec{x}|H_1)}$$

From the observed value of the likelihood ratio r_{obs}, the significance for the hypotheses is obtained.

4.3. Event Selection

4.3.1. Triggers, Datasets and Selection Criteria

The data used for this analysis were collected during the 2016 LHC run of pp collisions at 13 TeV, corresponding to an integrated luminosity of 37 fb^{-1}. They were stored in ROOT files as trees that contain all the relevant information on the major physics objects reconstructed in the detector (jets, leptons, photons, tracks). For the scope of this analysis, two triggers were used.

HLT_AK8DiPFJet280_200_TrimMass30_BTagCSV_p20_v trigger was used as Signal Trigger and HLT_AK8DiPFJet250_200_TrimMass30 _v trigger was used as Reference Trigger.

More specifically, Signal Trigger:

- is a “High-Level Trigger” (HLT), which means it is able to reduce to about 1 kHz the collision rate of the proton bunches. It is used after another trigger, called “Level 1” trigger, which previously reduces by a factor of 1000 the original collision rate, which is of 40 MHz at LHC. A HLT trigger is implemented in software running on a farm of commercial computers which includes about 16,000 CPU cores and exploits the same sophisticated software used for offline reconstruction and analysis, optimized in order to comply with the strict time requirements of the online selection
- selects two “AK8 jets”, i.e. wide jets reconstructed with a larger than usual clustering parameter (R = 0:8), whose pT is at least 280 GeV for the leading one and at least 200 GeV for the second. These jets are reconstructed through the “particle-flow” (PF) algorithm, which aims at identifying and reconstructing individually each particle arising from the LHC proton-proton collision
- selects a value of the so-called “trimmed mass” greater than 30 GeV. Jet trimming can achieve significant improvements in event reconstruction mitigating sources of contamination in jets initiated by light partons such as the initial state radiation, multiple interactions and pileup
- requires at least one b-tagged jet.

In the Figure below, the triggers used are shown.

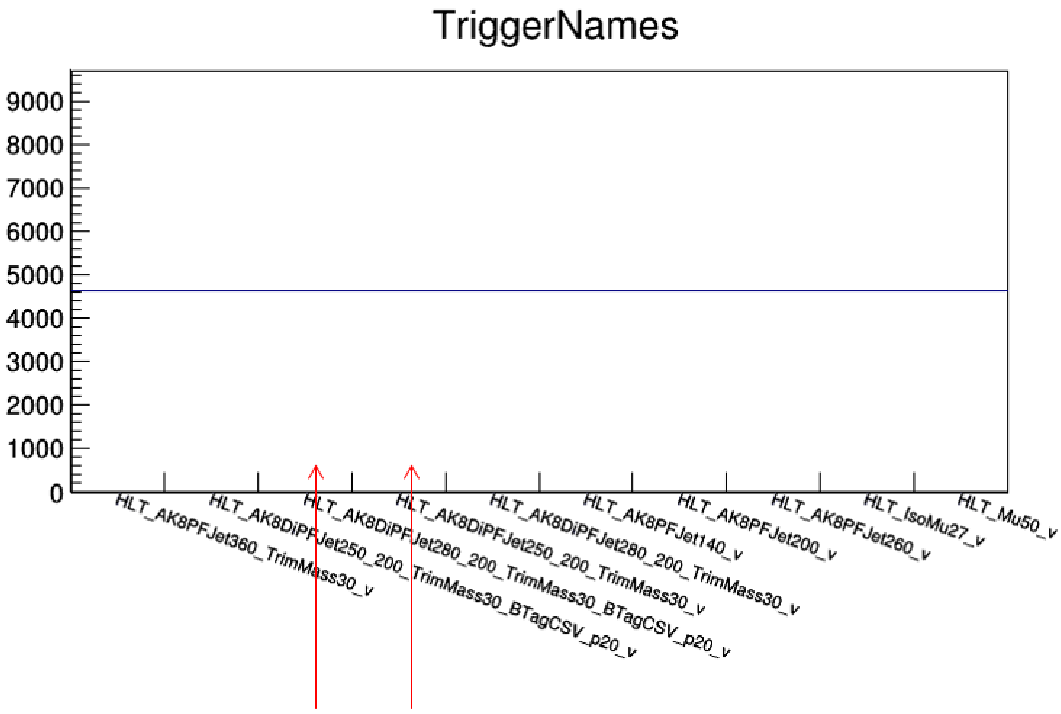


Fig. 4.1: Signal and Reference trigger (i.e. the selected ones) used for this analysis. Info regarding trigger paths is stored in TriggerBit array in dataset used.

For the initial vast number of events to be sized down, so that only interesting ones should be saved, a simple selection should be made. The set of cuts used in creating the initial root file are stated below.

Pre-Selection Criteria:

- Soft Drop Mass of the two leading jets is > 50 GeV
- JetPt of first and second leading jet is > 200 GeV
- At least one of the sub-jets in both leading jets is b-tagged

No trigger cuts were used in pre-selection of events.

For the scope of this analysis, a set of specific selection cuts was used.

Selection Criteria:

Events should pass the following selection criteria in order to be accepted:

- at least two jets must be present
- full hadronic final state is needed
- High PT dijet events with at least two jets with Pt>400GeV are accepted so that top decay products are collimated
- Both leading jets must be in the Top Mass window, 70<jetMassSoftDrop<300 GeV (i.e. jetMassSoftDrop is jetMass after removal of soft radiation)
- mva > 0.8 for the scope of this multi multivariate analysis. Mva is a fisher discriminant demanding jets to have a sort of ‘substructure’, since top quarks decay into W bosons and b quarks.

All the above events should also pass the reference or signal trigger.

B-Tagging Selection Criteria:

Events with 0 btag, 1 btag and 2 btags are distinguished and stored separately for further analysis; an additional cut of category=0/1/2 is used for selecting jets with 0/1/2 b quarks included.

The selection criteria are presented in detail below.

- At least 2 jets present in final state
- No isolated leptons in final state

- PT of sub-leading jet must be over 400 GeV
- Trigger selection (signal or reference)
- Soft drop jet’s mass for both leading jets in 70-00 GeV window
- Category for identifying the number of btagged subjets (0,1 or 2)
- mva cut is applied on 0.8

By choosing the selection criteria listed above, the Electroweak background and the single-top background are efficiently decreased.

Datasets

The Monte Carlo datasets used in this analysis are shown below. Each dataset has been produced using different parameters so that the process of ttbar decay will be simulated. The scope of this analysis is the extrapolation of data from the fiducial phase space to the full phase space.

TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root (Nominal)
Smeared_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root
SmearedUp_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root
SmearedDown_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root
ShiftedUp_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root
ShiftedDown_TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root
TT_TuneCUETP8M2T4up_13TeV-powheg-pythia8.root
TT_TuneCUETP8M2T4down_13TeV-powheg-pythia8.root
TT_TuneCUETP8M2T4_13TeV-powheg-isrup-pythia8.root
TT_TuneCUETP8M2T4_13TeV-powheg-isrdown-pythia8.root
TT_TuneCUETP8M2T4_13TeV-powheg-fsrup-pythia8.root
TT_TuneCUETP8M2T4_13TeV-powheg-fsrdown-pythia8.root

Table 4.1. Data and Monte Carlo files (nominal and variations) used for this Analysis

4.3.2. Trigger Efficiency

Trigger Efficiency is stated as the fraction of number of events that pass the offline selection criteria along with the reference trigger over the ones that only pass the offline selection criteria.

$$\text{efficiency} = \frac{\text{events passing the trigger \& offline criteria}}{\text{events passing the offline criteria}}$$

If offline reconstruction was identical to the online one, the efficiency would be close to 1. The existence of a trigger means though that in the end, the collected events are the ones that successfully pass the trigger; a redefinition of trigger efficiency should be considered.

The final trigger efficiency used is as shown below:

$$\text{efficiency} = \frac{\text{events passing the trigger \& offline criteria \& and the reference trigger}}{\text{events passing the offline criteria \& the reference trigger}}$$

Reference and Signal Triggers used in this Analysis, are described earlier in 4.1.

More specifically, trigger efficiency *e* is stated as:

$$e = \frac{N_{pass}}{N_{tot}}$$

where Npass is the fraction of events that pass the selection criteria, while Ntot is the whole sample of events, so, *e* should be as close to 1 as possible.

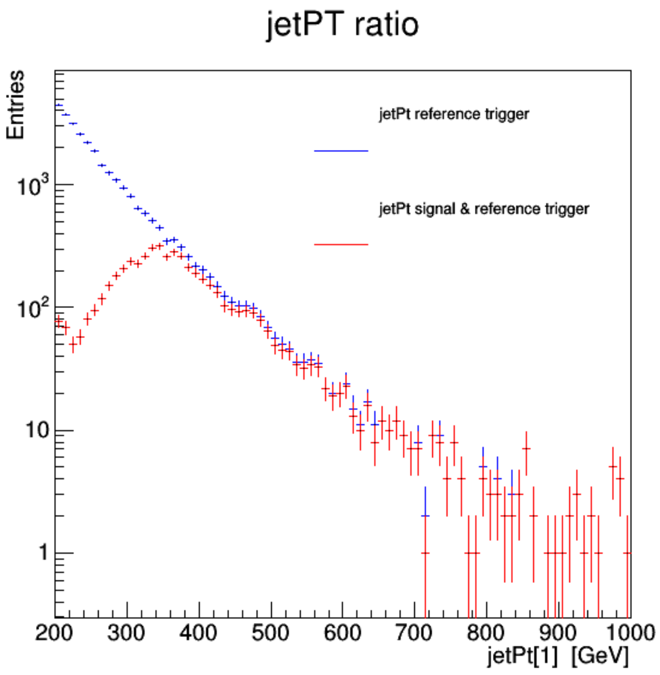


Fig.4.2: ratio of PT of 2nd Jet of events passing the reference and signal triggers over the ones passing the reference trigger

The trigger efficiency of the above triggers is presented below.

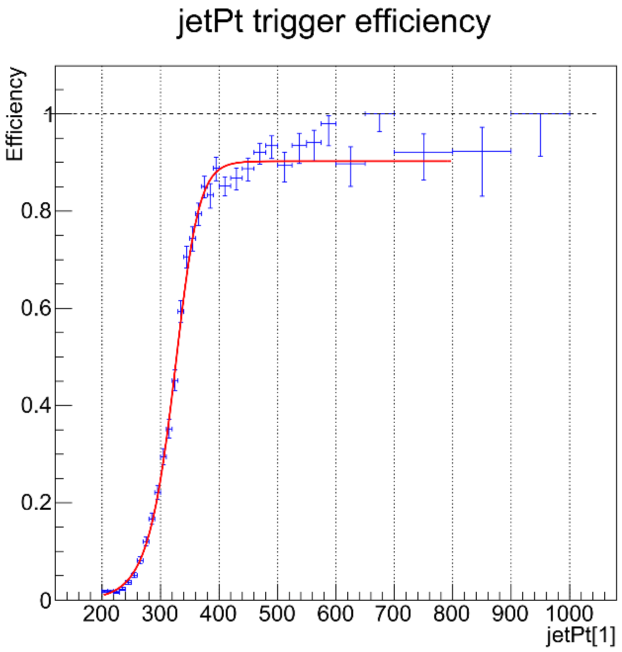
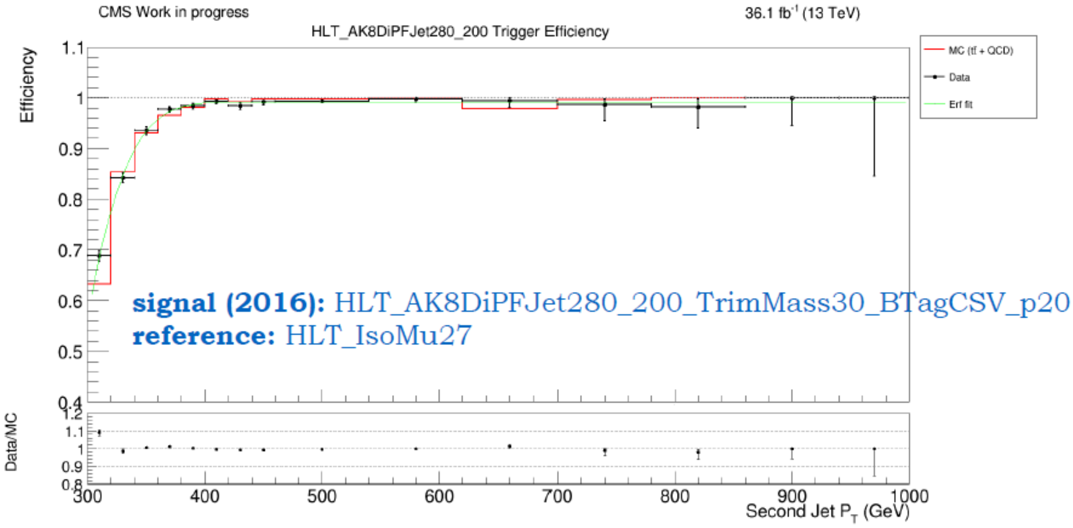


Fig. 4.3: Trigger efficiency of signal trigger used for this Analysis. The fitting function used was $1/(1+\exp((x-m)/\sigma)^{-1}$ with values m=300 and σ=50

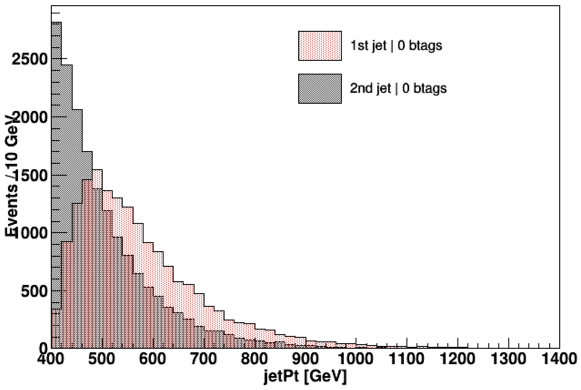


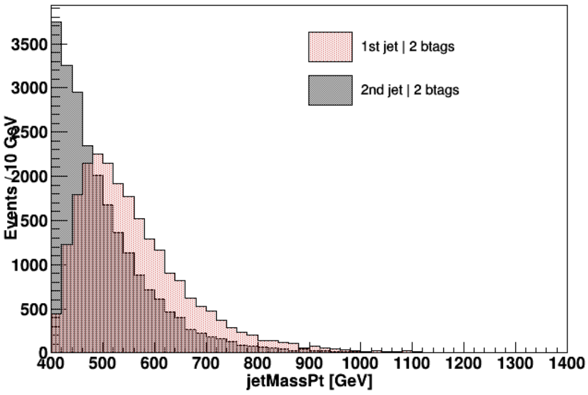
signal (2016): HLT_AK8DiPFJet280_200_TrimMass30_BTagCSV_p20
reference: HLT_IsoMu27

Fig. 4.4: Trigger efficiency of signal trigger used for this Analysis

Kinematic Variables Plots

The distributions of jetPt and eta are displayed below.





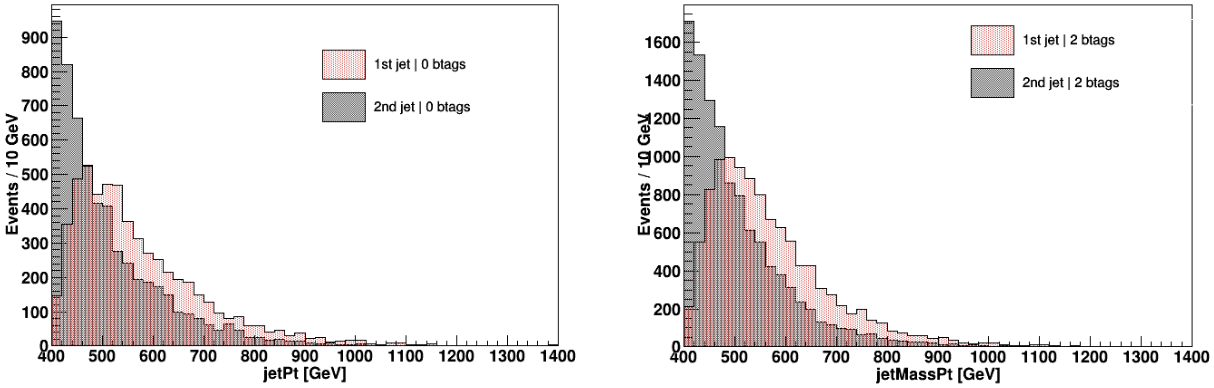


Fig 4.5: jetPT distributions for the two leading jets for nominal MC sample (up) and data (down) for 0 b tags (left) and 2 b tags (right) respectively

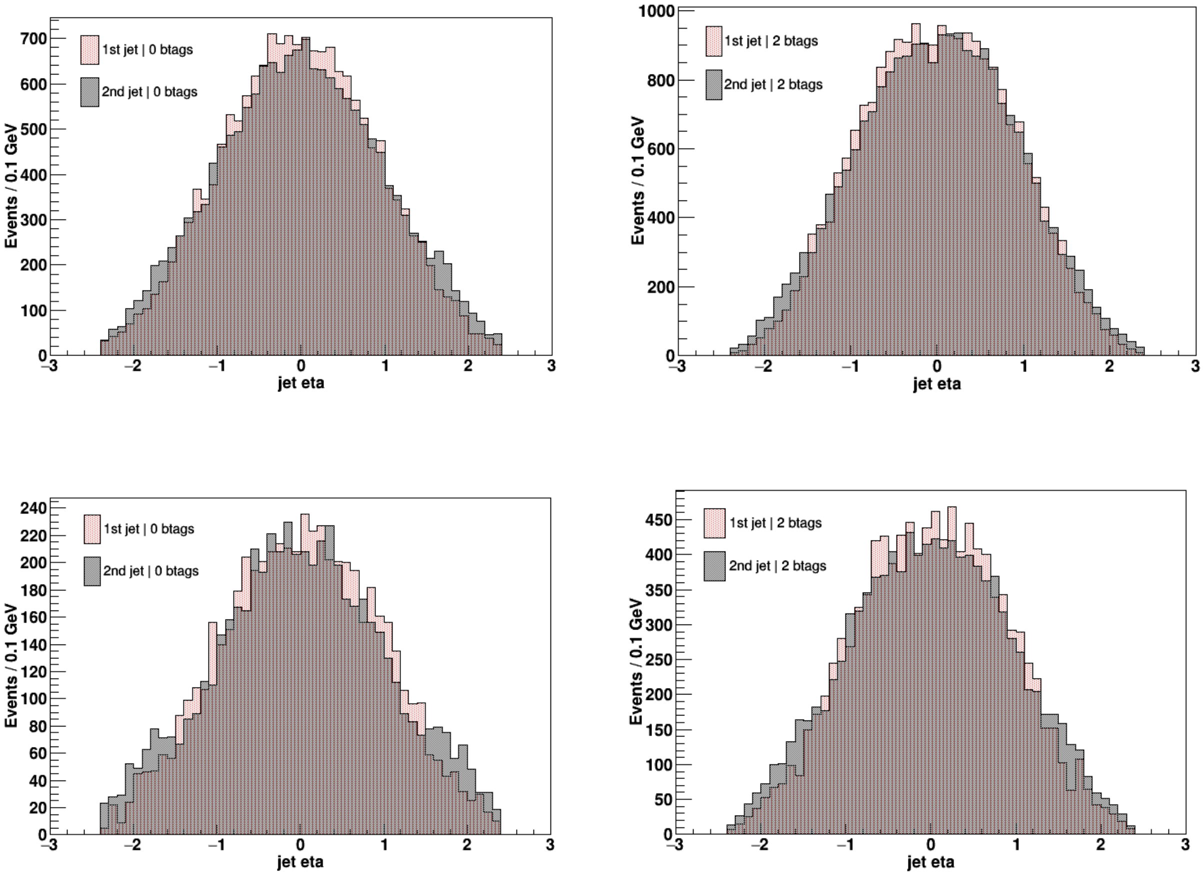


Fig 4.6: eta distributions for the two leading jets

All Monte Carlo simulated files and dataset as displayed in Table 4.2 have been processed in order for the needed plots to be created. The plots were stored in extra root files for further investigation. The selection criteria are the ones stated in Table 4.1 before.

- Soft Drop Jet Mass of first two leading jets: jetMassSoftDrop
- Jet mass for first two leading jets
- mva variable
- jetTau 1,2,3

All the above plots have been stored separately if 0, 1 or 2 bjets are included.

4.4. Multivariate Analysis

In high-energy physics, with the search for ever smaller signals in ever larger data sets, it has become essential to extract the most of the available information from the data. Multivariate classification methods based on machine learning techniques have become a fundamental ingredient to most analyses. Also the multivariate classifiers themselves have significantly evolved in recent years. Artificial neural networks (ANNs) are computing systems inspired by the biological neural networks that constitute animal brains. The human brain has approximately 100 billion neurons, which communicate through electro-chemical signals. The neurons are connected through junctions called synapses. Each neuron receives thousands of connections with other neurons, constantly receiving incoming signals to reach the cell body. If the resulting sum of the signals surpasses a certain threshold, a response is sent through the axon. The ANN attempts to recreate the computational mirror of the biological neural network, learning to do tasks by considering examples. An ANN is based on a collection of connected units called artificial neurons. Each connection between neurons can transmit a signal to another neuron. The receiving neuron can process the signals and then signal downstream neurons connected to it. Neurons generally have states represented by real numbers between 0 and 1. Typically, neurons are organized in layers. Different layers may perform different kinds of transformations on their inputs. Signals travel from the first input to the last output layer, possibly after traversing the layers multiple times, as shown in Fig. 4.4. Each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another.

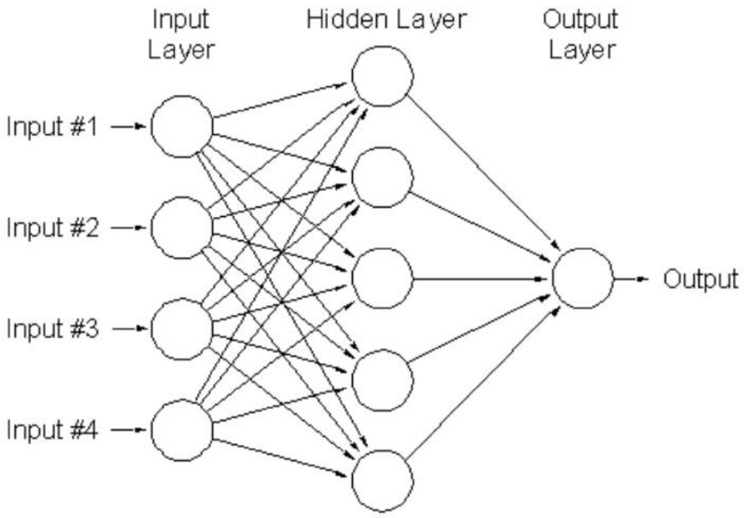


Fig. 4.7: The ANN structure

The original goal of the neural network approach was to solve problems in the same way that a human brain would. Over time, attention focused on matching specific mental abilities, leading to deviations from biology such as backpropagation, or passing information in the reverse direction and adjusting the network to reflect that information.
Inputs of our neural network as shown in Fig. 4.5 below.

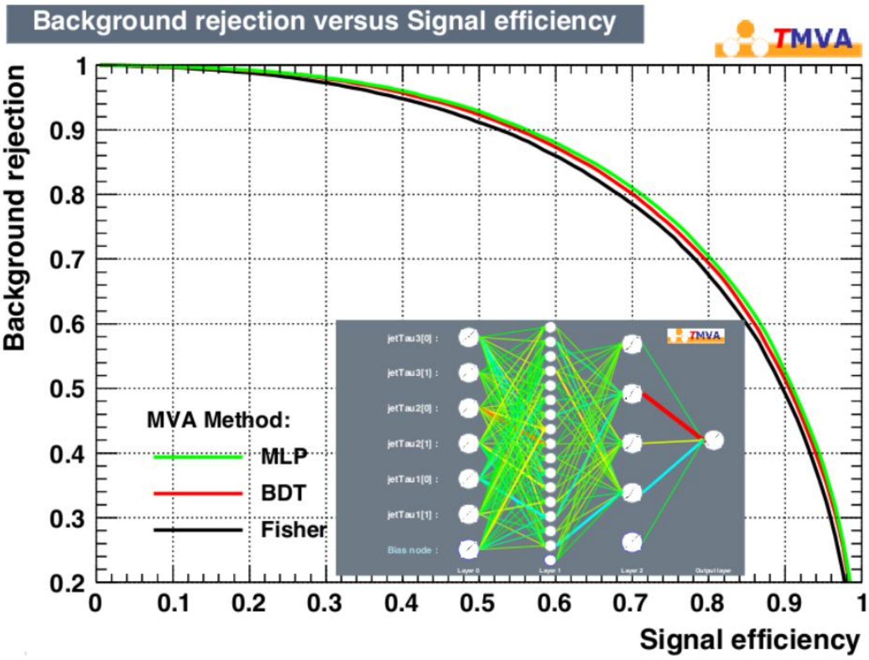


Fig. 4.8: ANN architecture and comparison between ROC curves for three different analysis methods.

4.4.1. What is mva?

The ANN training consists in finding transfer functions from one node to another in a way to provide separation between signal and background. During the training the network is told which events are to be considered signal (i.e. ttbar) and which events are to be considered background (i.e. QCD), then ANN is trained to recognize signal and background events based on a neural network output mva which ranges between 0 and 1, with signal events clustered towards mva = 1 and background events shifted towards mva = 0. Fig. 4.6. shows the output of the multivariate training.

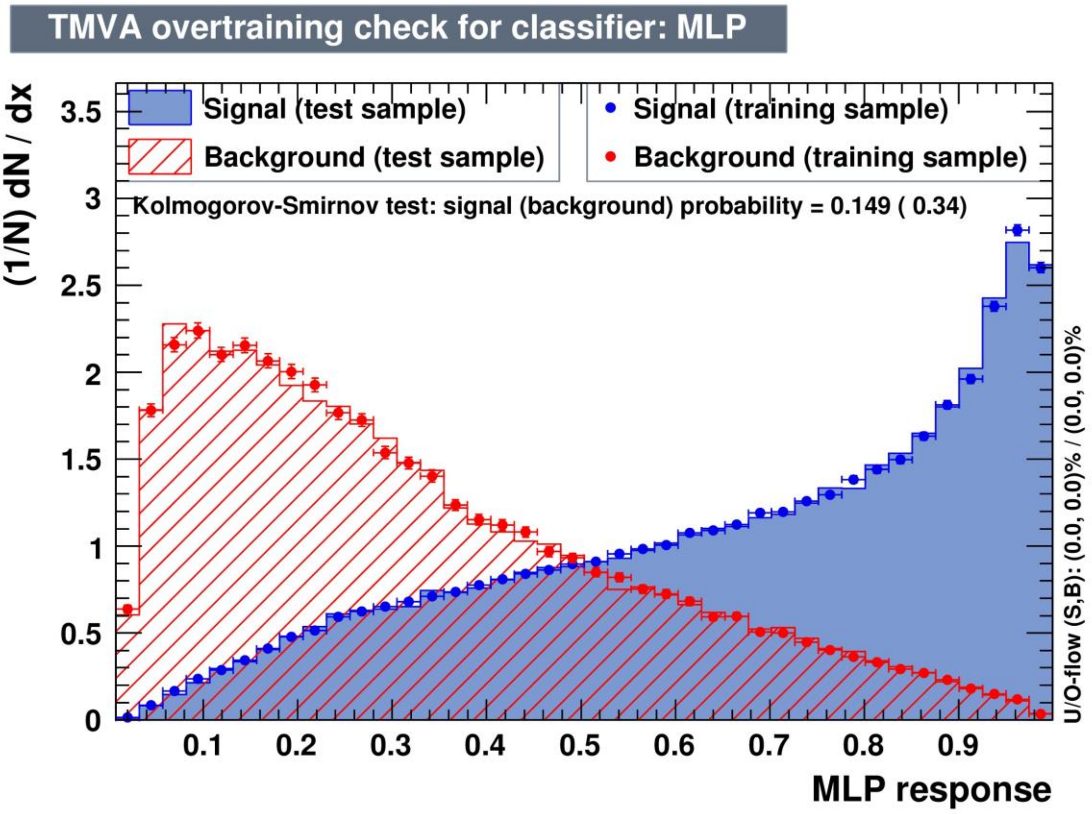


Fig. 4.9: mva distributions after the training.

Both signal and background events are requested to satisfy the previous selection, except for the b-tagging. Our analysis aims to the mva region close to 1, where the ttbar signal significantly emerges from the background, and with the additional request for two b-tagged jets, as occurs in the all-jets boosted topology.

The MVA output layer of the neural network is thus a variable which considers the specific characteristic of the signal events and can be used as the last request for our events. Therefore, we need to evaluate a minimum value of the mva beyond which we accept the events, and this is performed by a maximization approach of the ratio between S and √B (S√B). This optimal minimum value of mva should allow us to have a good signal-to-noise ratio S=B but also a sufficient amount of candidate events that passed the request.

4.4.2. N-Subjettiness

Substructure information can be further leveraged by using a parameter called Nsubjettiness. The N-subjetiness algorithm defines variables τN, where N is the number of subjet axes, as follows:

$$\tau_N = \frac{1}{d_0} \sum_i p_{T_i} \times min(\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i})$$

where $\Delta R_{j,i}$ is the distance between the subjet axis j and the candidate jet i. Normalizing term d_0, takes p_T into account, with $d_0 = \Sigma P_{T_i} R_0$, and R_0 is the distance parameter (R) used in the jet clustering algorithm (R = 0.8 for CA8 jets).
N-subjettiness can be used to distinguish between top and non-top jets because hadronic top jets should have three subjet prongs whereas the QCD background will preferentially have one prong. A τ_N value near 0 means the jet likely has N subjets or fewer; a value near 1 means the jet likely has more subjets. The ratio of τ_3/τ_2 (known as τ_{32} and itself bound between 0 and 1) has been shown to provide much better discrimination between top and QCD than either τ_3 or τ_2 and is used in this analysis.

The N-subjettiness distributions for $\tau_{1,2,3}$ (i.e. jetTau1,2,3 for the two leading jets) are shown in Fig. 4.10 below.

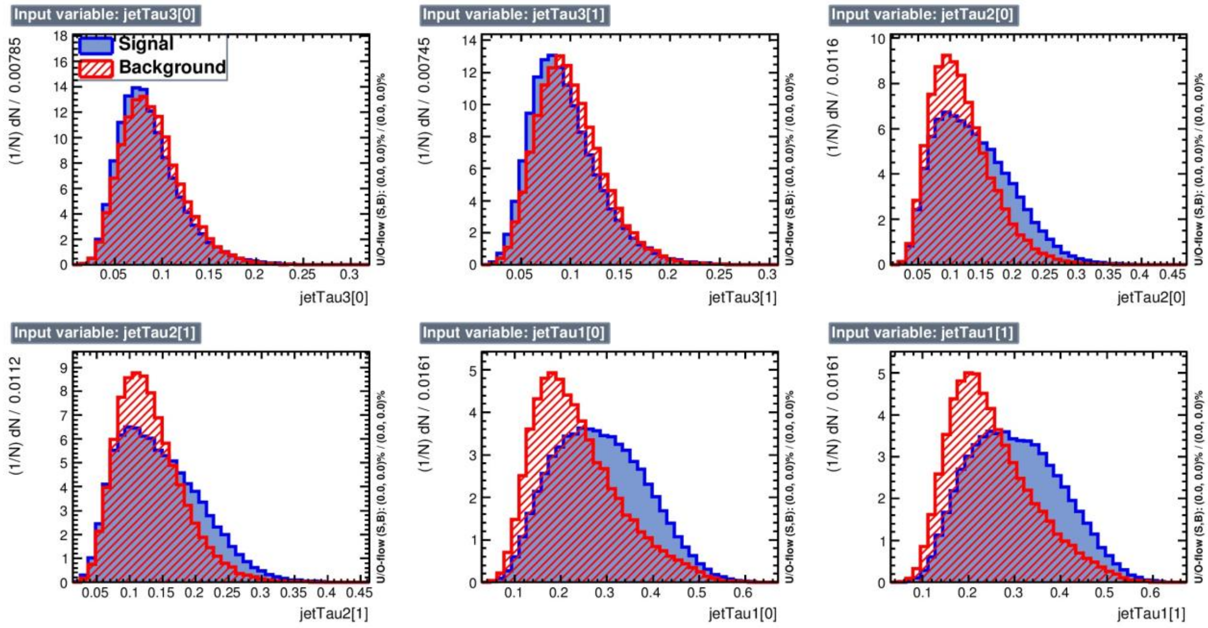


Fig. 4.10: jetTau1,2,3 distributions for two leading jets

4.4.3. Defining optimum mva cut

A usual technique for systematic uncertainties to be tested and verified is the generation of ensemble tests (i.e. 'pseudo-experiments'), called 'toys'. The kinematic variable used for this Analysis is the Soft Drop mass of the two leading jets, so since the Maximum Likelihood fit is applied on this variable, it has to be tested for any possible unhandled systematic bias. Randomly generated distributions of a variable can be created for the entire Analysis to be based on that. The variety of outcomes is an index that represents uncertainty.

When generating pseudo-experiments to test the properties of a fitting algorithm that includes constraints, it is necessary to understand which parameters to fluctuate, and how to fluctuate them. The method used is to fluctuate the event rate according to a Poisson distribution with fixed mean, and separately to fluctuate the constraining value according to its Gaussian distribution.

Generating a variable x as a Gaussian distribution for N times, with mean = m and sigma = σ, a standard Gaussian distribution will be extracted with mean = 0 and σ = 1, i.e. 'Pull' distribution. The pull distributions actually represent the existence or not of a possible systematic bias which will eventually affect all following measurements, since they are based on the specific kinematic fit. So, pull distributions must be verified.

Regarding this Analysis, a number of 10.000 toy experiments was generated with specific values for Nsignal and Nbackground, for different values on mva cut. Eventually, a pull distribution is expected assuming no bias is present.

mva >	N_{tt-}	N_{QCD}
0.4	11200	25300
0.5	10500	18100
0.6	9500	12100
0.7	8400	7200
0.8	6900	3600
0.9	4600	1050

Table 4.2: Number for Signal and Background used in toys

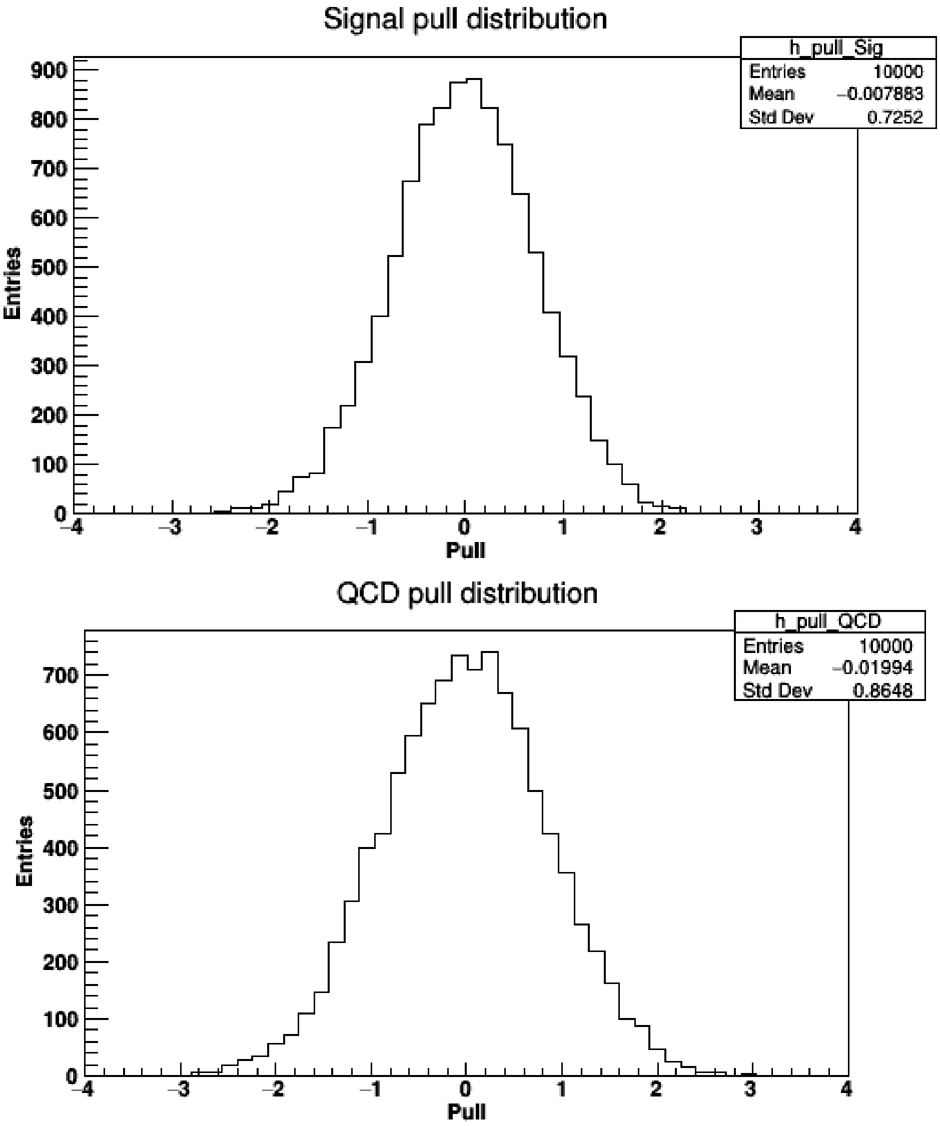


Fig 4.11: Pull distributions for ttbar signal and QCD background, for mva>0.8

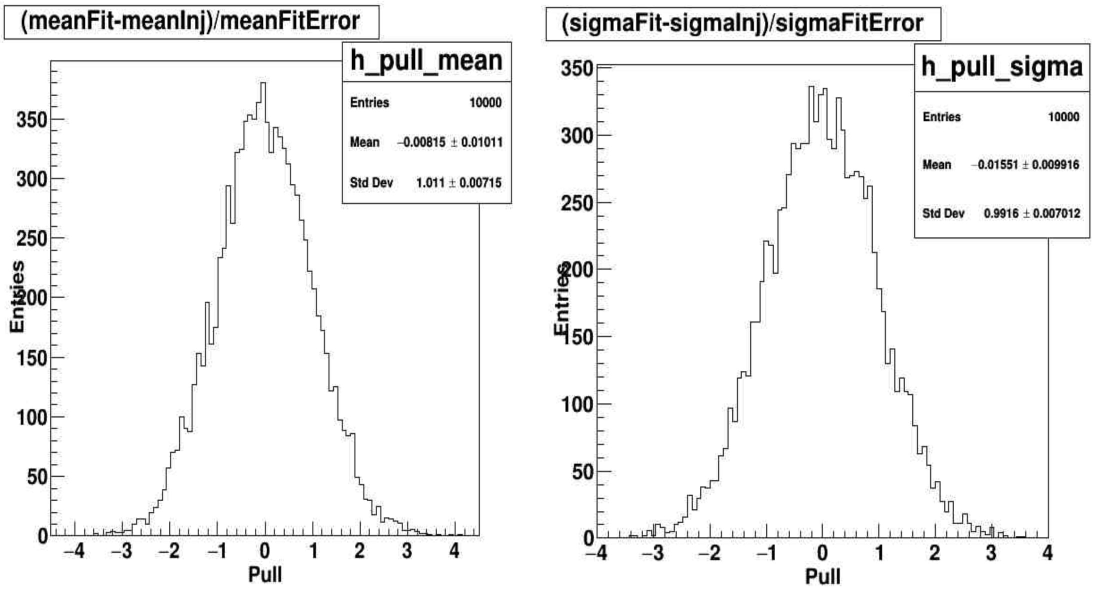
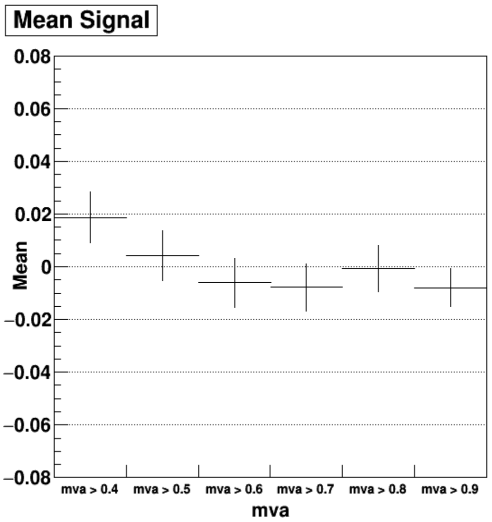
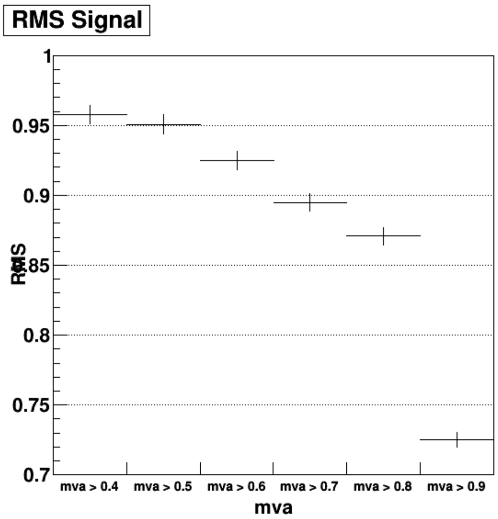


Fig 4.12: Pull distributions’ relative error mean and sigma, for mva>0.8

As expected, all distributions are standard Gaussian with mean = 0 and standard deviation = 1; a significant bias is absent. The same process was followed for all cuts on mva; final results are presented below.





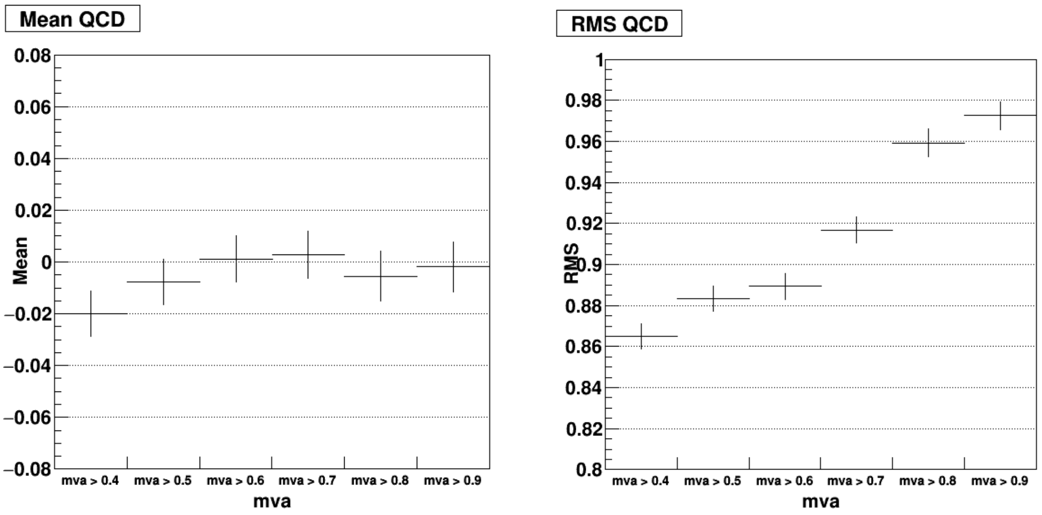


Fig 4.13: Pull distributions’ mean and sigma for all mva cuts for ttbar Signal (top) and QCD background (bottom)

An optimum value for mva cut would require the least Relative Error; such plot is presented below.

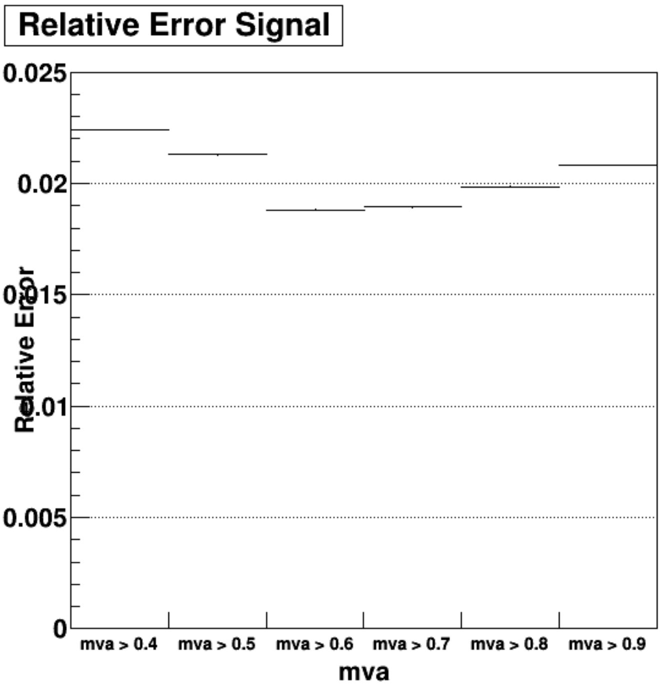


Fig 4.14: Signal’s Relative Error for all cuts on mva

From all above distributions, an optimum cut on mva would imply a Pull distribution of mean as close to 0 and sigma as close to 1 as possible, along with the least relative error. A combination of the above is complete after also maintaining the best Signal over Background ratio, as shown in following plot.

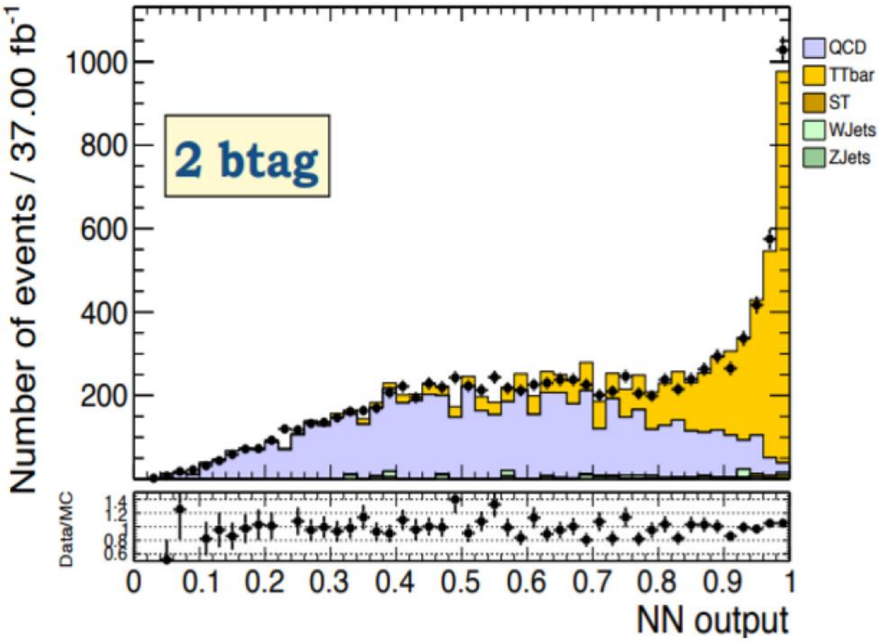


Fig 4.15: ttbar and QCD multijet yield for the nominal selection cuts after requiring 2 btagged subjets present in final state

Considering all the above information, the mva cut on 0.8 is decided and applied for the whole Analysis.

4.5. Background Modelling

In order to measure the ttbar production cross section, both the signal and the background distributions are needed. While the signal distribution can be directly obtained from the ttbar sample with the selection of two b-jets, the estimation of the background distribution is a more difficult issue. First of all, background events can be extracted from data or reconstructed from QCD simulations, but it is preferable to use the actual data rather than QCD samples, thus avoiding possible simulation biases. However, in doing so, it is necessary to ensure that the background model we extract from the data is compatible with the expected QCD one. The estimation of QCD background is uncertain because of the uncertainty in the cross section of QCD multijet processes, so we need a proper technique that is able not only to reject as much as possible QCD multijet processes, but also to extract the absolute normalization of the background yield. This can be performed using the soft-drop mass of the two leading boosted jets. The background distribution is derived from data with the selection of no b-tag, in such a way to suppress the presence of signal events as much as possible, and should have a shape similar to the one derived from QCD events with the same no b-tagging request.

A QCD closure test is perfomed and shown in Fig. 4.8 below. The QCD multijet background's shape doesn't change under special requirements regarding the existance of a b-tagged subjet.

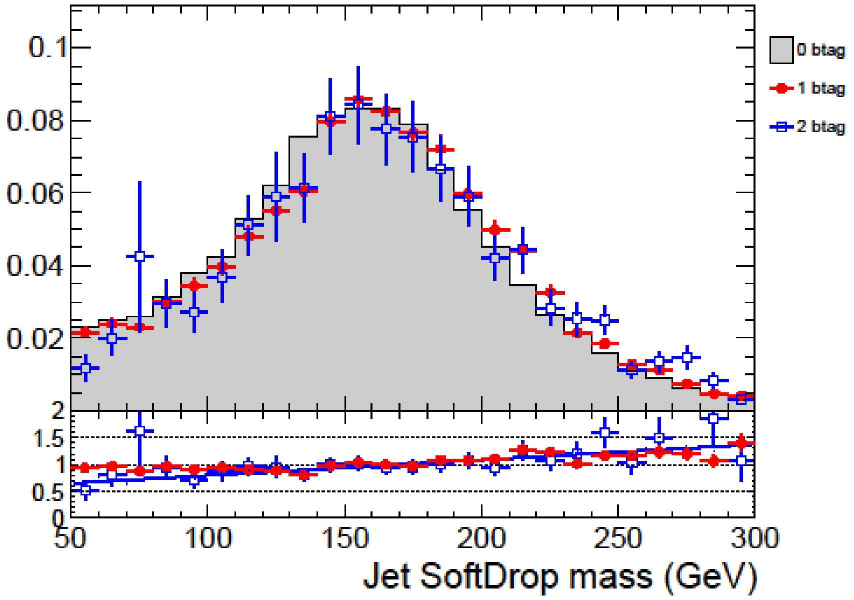


Fig. 4.16: QCD yield for 0, 1 and 2 btags required. Shape remains unchangeable.

Background shape changes while applying a cut to mva as shown in Figure 4.17 below. A more distinct peak is shown at mva>0.9, while a slight right shift is noticed as mva cut increases. Mva cut in figure below is ranged between 0.4 and 0.9.

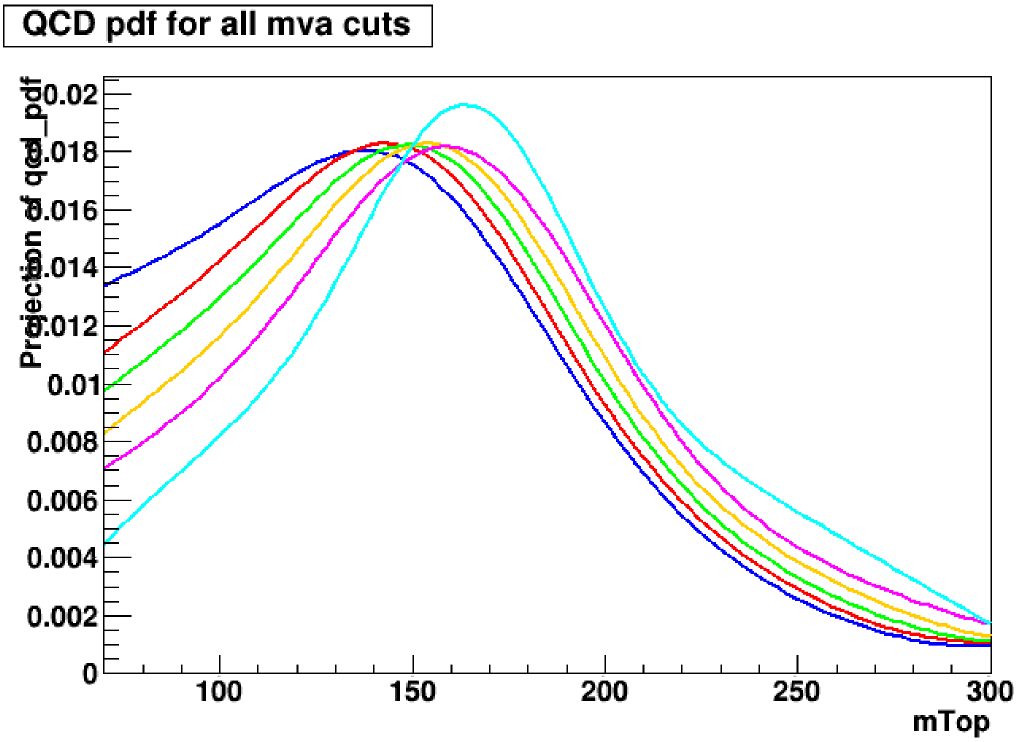


Fig. 4.17: QCD background shape is modified as mva cut increases.

4.6. Templates

Templates are Probability Density Functions (PDF)s constructed from full simulations in reconstructed quantities (eg. M^{reco}_{top}, kinematic endpoints, etc), for many top quark masses (eg. M^{gen}_{top}). They are different for signal (i.e. ttbar) and background (i.e. QCD), the templates are parametrized and the parameters are fitted linearly to varied quantities (M_{top} in this analysis), and Likelihood uses the fitted Template functions.

The distribution of MSD for 0 btags, where we extract the background shape from and 2 btags, where we extract signal shape from respectively are shown below.

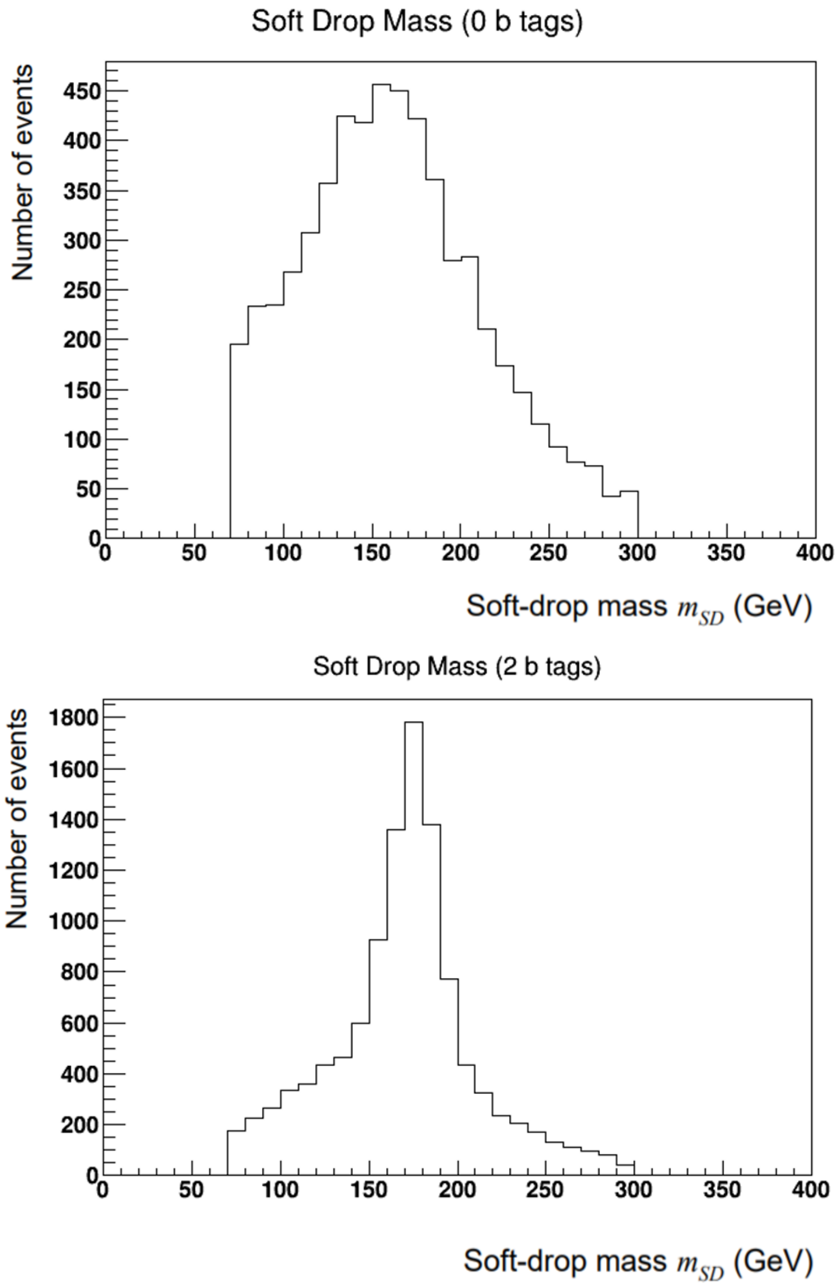


Fig. 4.18: Jet's soft drop mass distributions for 0 (up) and 2 (down) b tags requirement.

A ROOT macro has been used to extract the soft-drop mass M_{SD} for the ttbar signal and QCD background expected after the events selection.

These two distributions are called 'template distributions' because they have been normalized to unit areas, and therefore they represent the two probability distribution functions (pdf) for the ttbar signal and QCD background. Our choice for the background modeling is to use the background as inferred from data with no b-jets. The signal is characterized by a peak in the area around 173 GeV, corresponding to the mass of the quark top. The background is distributed uniformly, growing and decreasing smoothly.

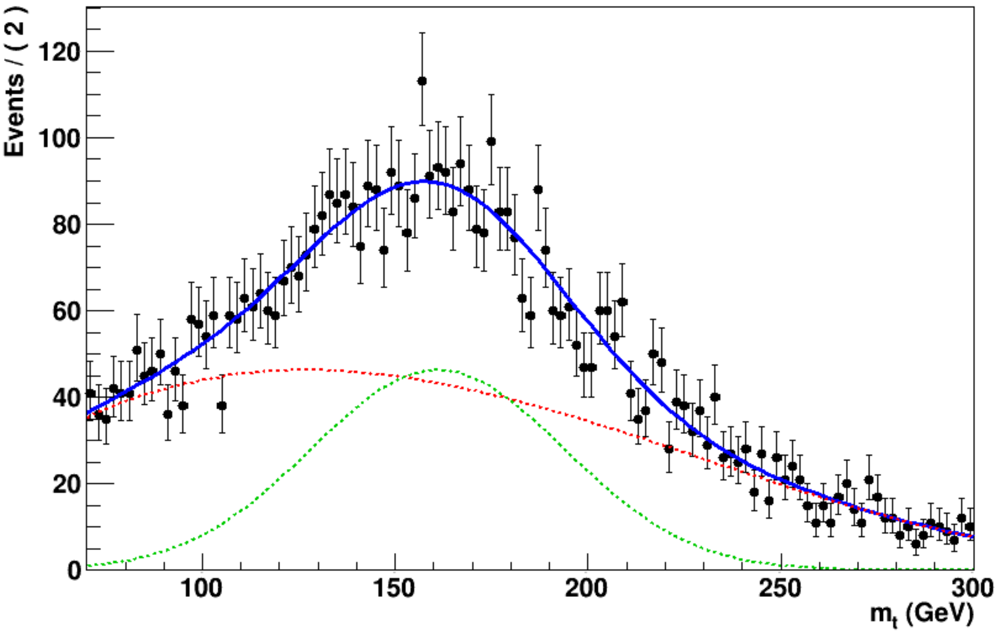


Fig. 4.19: Template distribution for the QCD background, extracted from data with no b-tagging requirement (green is for background, red is for signal and blue is the product of the two shapes, processed with Roofit)

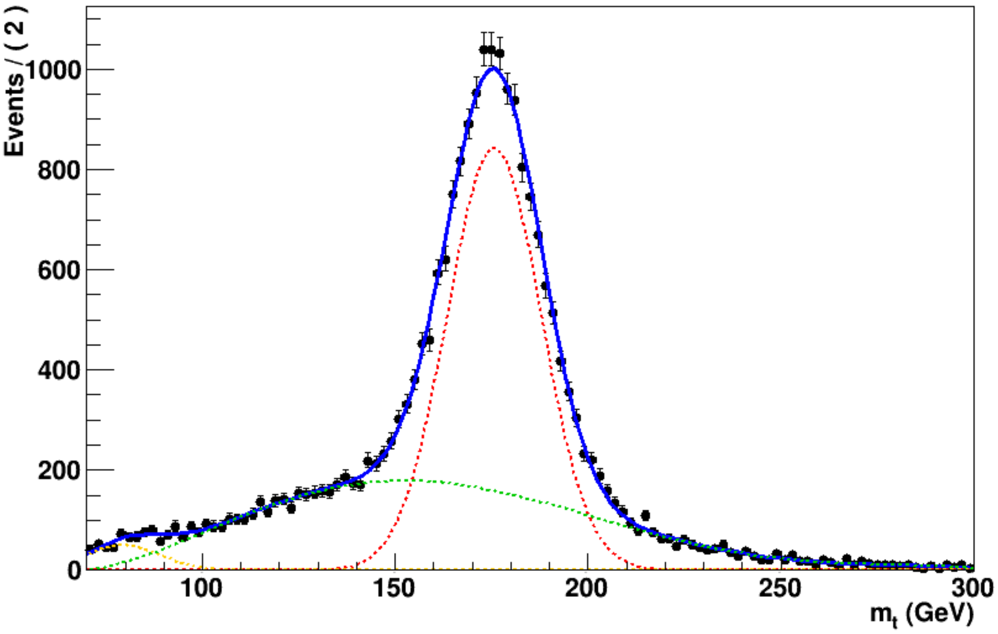


Fig. 4.20: Template distribution for ttbar Signal with a requirement of 2 btagged jets (green is for background, red is for signal and blue is the product of the two shapes, processed with Roofit)

4.7. Fit to Data

Data have been fit to nominal Monte Carlo sample along with the QCD that has previously been extracted from data with no btags required.

The result of the above fit is presented below.

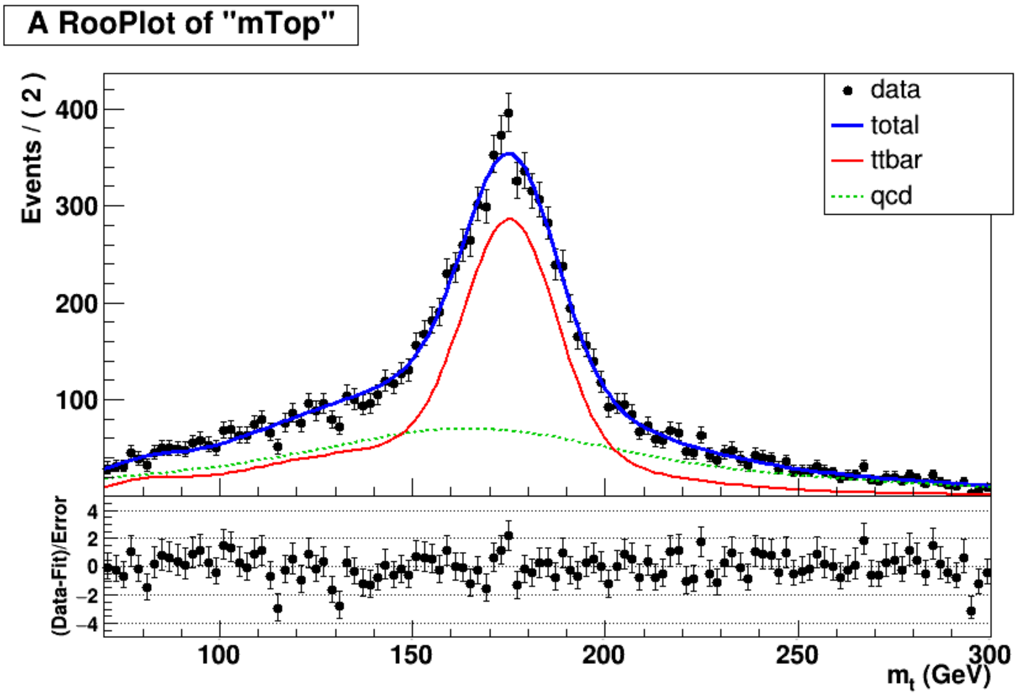


Fig 4.21: Fit Data on MC signal and data-driven QCD background using nominal Monte Carlo signal sample.

The required info has been extracted from the Likelihood Fit in order for final cross-section result to be calculated.

Nominal Sample:

Ngen = $7.7*10^7$ events
Npass = 23067 events
ϵ = Npass/Ngen = $2.99*10^{-4}$

While *after Fit* the following info is extracted:

N_{tt-} = 6486 ± 167 events
N_{QCD} = 4423 ± 160 events
chi2/ndof = 0.87

where Ngen is the total numbers of the generated events present in the nominal Monte Carlo sample, Npass is the number of events that pass the trigger and selection criteria, and ϵ is the efficiency.

Using the numbers above, for the integrated luminosity of $\int Ldt = 37\ fb^{-1}$, the cross-section has been calculated to:

$$\boxed{\sigma_{incl}^{nom}=N_{ttbar}/\epsilon^{*}L\ \ = 585.74\ pb}$$

4.8. Efficiency Correction

The efficiency _is computed from signal simulated events and refers to the full selection. However, both the trigger and b-tagging might not be well simulated. For this reason, we would need to introduce two MC-to-data scale factors: SF_{trig} and $SF_{b\text{-}tag}$, which improve the estimated MC efficiency as to reflect the value it has on data. From our study earlier, we do not see much difference between the efficiency trigger as evaluated from MC and seen in data, so we assume $SF_{trigger}$ = 1.

4.8.1. b-tagging efficiency

The CMS simulated b tag efficiency, for correctly identifying a b tagged jet, along with the simulated fake rate for misidentifying light jets coming from u,d,s or c quarks or gluons for b jets, is shown below.

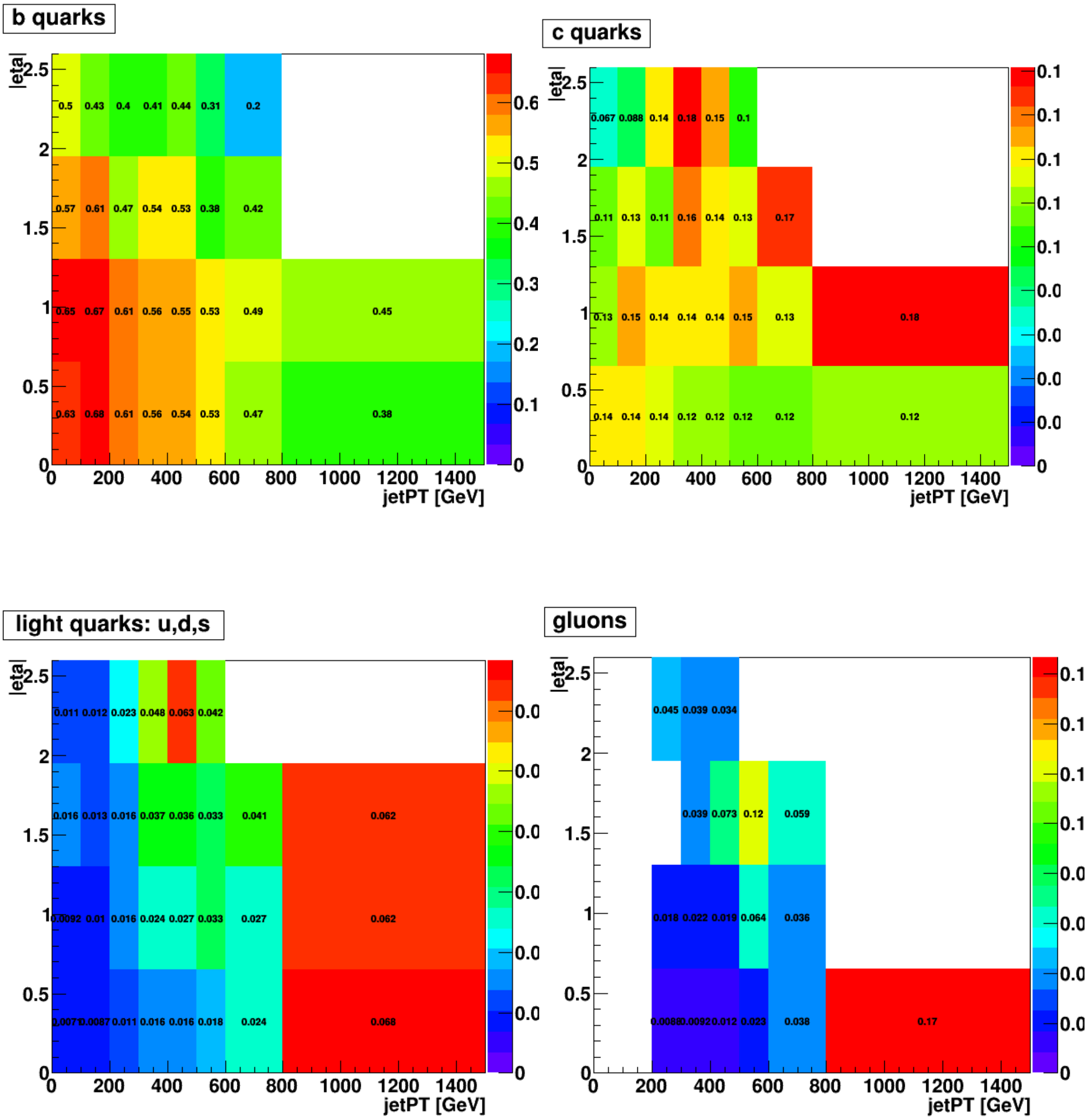


Fig.4.22: CMS simulated b tag efficiency (upper left) and fake rate for c (upper right), light quarks (down left) and gluons (down right)

From the above, a b-tag scale factor has been calculated and applied to nominal Monte Carlo sample’s efficiency.

4.9. Systematic Uncertainties

The cross-section measurement is affected by many sources of systematic uncertainties; experimental and theoretical.

As for experimental uncertainties, they originate from:

- *Jet Energy Scale (JES)* and *Jet Energy Resolution (JER)*:

JES is the main source of experimental uncertainty. n the MC signal sample the energy scale and resolution of jets have been corrected in order to better describe the data. However, the uncertainties on these corrections have impact on the cross-section measurement. To evaluate these sources of uncertainty the jet energy and resolution have been shifted and smeared up and down by one JES and JER standard deviation respectively. These corrections impact both on the shape of the signal templates and on the efficiency of the selection.

- *Trigger Efficiency*:

As explained earlier in this Analysis, this relates to whether the offline reconstruction is close to the online one. From Fig. 4.4 it is clear that this systematic plays a minor role in this Analysis, as trigger efficiency is close to 1 for the signal trigger used here.

- *B-tagging Efficiency*:

A b-tagging Scale Factor has been extracted from the study of simulated b-tag efficiency and fake rate of CMS. This scale factor has been calculated and applied to nominal Monte Carlo sample's efficiency, being the base for the current Analysis.

The theoretical uncertainties originate from:

- *Initial State Radiation (ISR)* and *Final State Radiation (FSR)*:

during a collision at the LHC, the initial and final state particles may emit radiation in the form of gluons, which will then hadronize resulting in jets of particles. The average production of initial and final state radiation can be estimated, but it will be inevitably affected by an uncertainty, which will in turn become a source of systematic uncertainty in the cross-section measurement. This contribution can be evaluated by increasing and decreasing by one standard deviation the amount of radiation produced in the initial and final state and then compute the cross section in each case.

- *Underlying Event (UE):*

In a pp collision, the interaction is more complicated than the simple “hard scattering”, i.e. the process where one parton from the first proton collides with another parton from the second proton. The probability of such interaction between two specified partons is given by the PDFs. In addition to the remnants of the hadronization of the non-interacting (“spectator”) quarks, more than just one parton from each proton may interact. These multiple parton interactions usually result in the radiation of soft particles with low energies. This is because the probability for a second hard interaction within the same proton-proton collision is smaller, which means that in the detector the signatures of the hard interaction are spoiled by soft energy deposits throughout the detector making the jet reconstruction itself more challenging.

- *Pile-up*:

At LHC the beam does not consist of individual protons but large bunches of protons. As a result, instead of only one pp interaction, two or more are happening simultaneously, and the contribution of these additional interactions (“pileup”) increases with the instantaneous luminosity. The soft radiation originating from pileup will also result in energy deposits all over the detector. A single pileup interaction would be less severe for a measurement than one from UE.

The systematic uncertainty related to the integrated luminosity measurement is determined by x-y beam scans. For the 2016 data taking period, it amounts to 2.5%.

Chapter 5

Results

The above process of fitting data has been performed for all Monte Carlo signal samples. The Monte Carlo ttbar samples have been fit with the data-driven QCD background in order for the signal yield to be extracted.

The efficiency is computed from ttbar simulated events using the formula:

$$\epsilon = \frac{Npass}{Ngen}\bigg|_{\text{MC.}}$$

The calculation for all Monte Carlo samples along with info regarding Nsig events derived from fit are presented below. As nominal, the efficiency with b-tag Scale Factor applied is used.

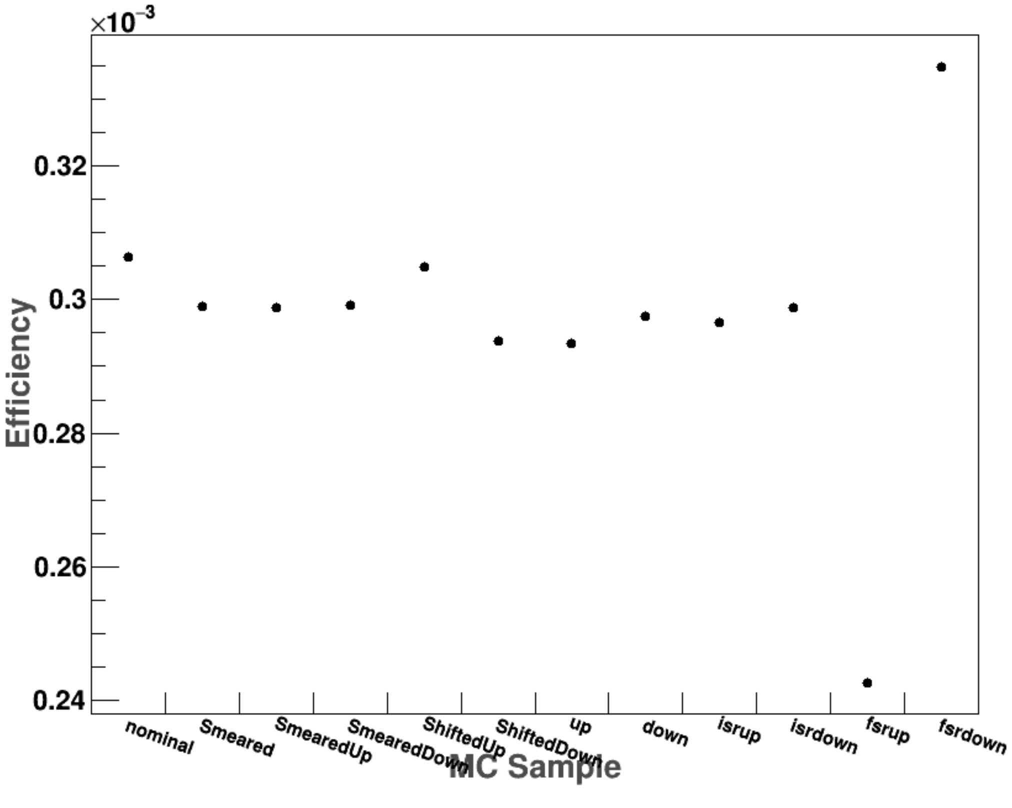


Fig. 5.1: Efficiency results for all Monte Carlo samples used

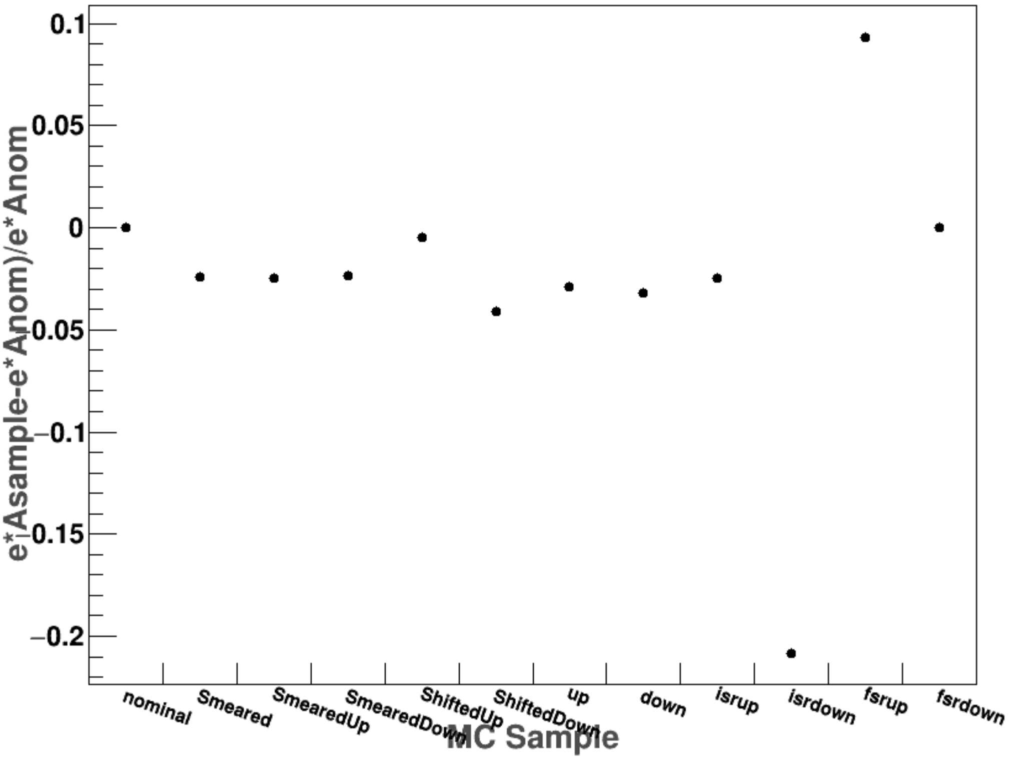


Fig. 5.2: Fractional difference of efficiency from nominal Monte Carlo sample

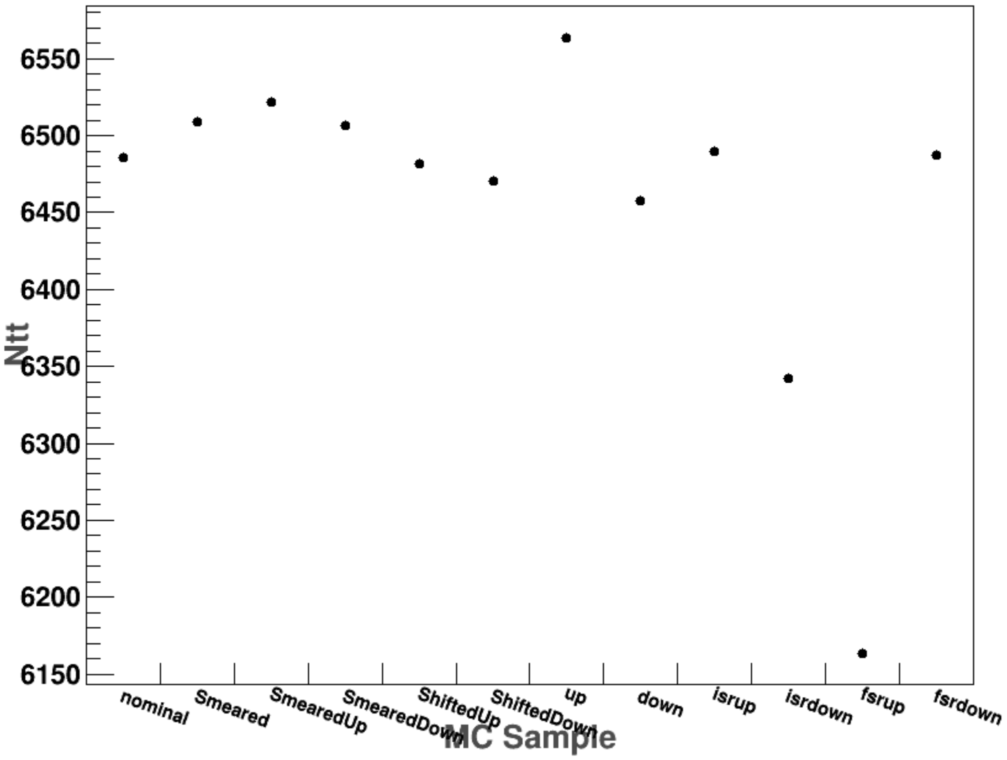


Fig. 5.3: Number of signal events derived from fit, from all Monte Carlo samples used

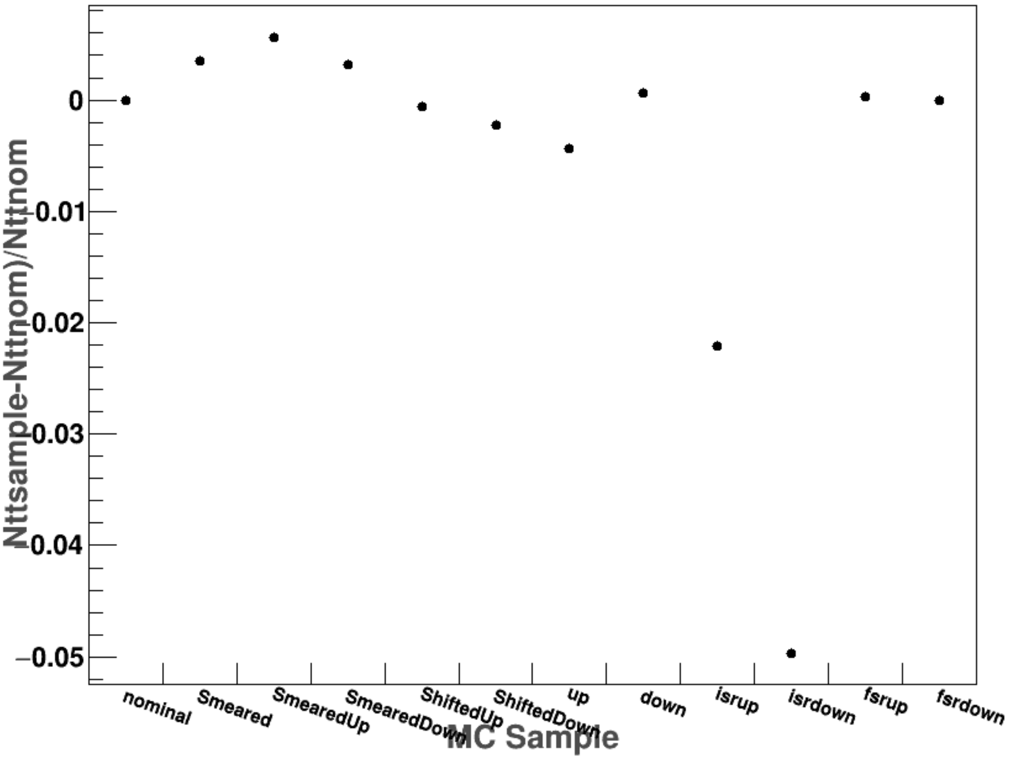


Fig. 5.4: Fractional difference of signal events compared from nominal Monte Carlo sample

The derived cross-section calculations for all Monte Carlo samples are shown below.

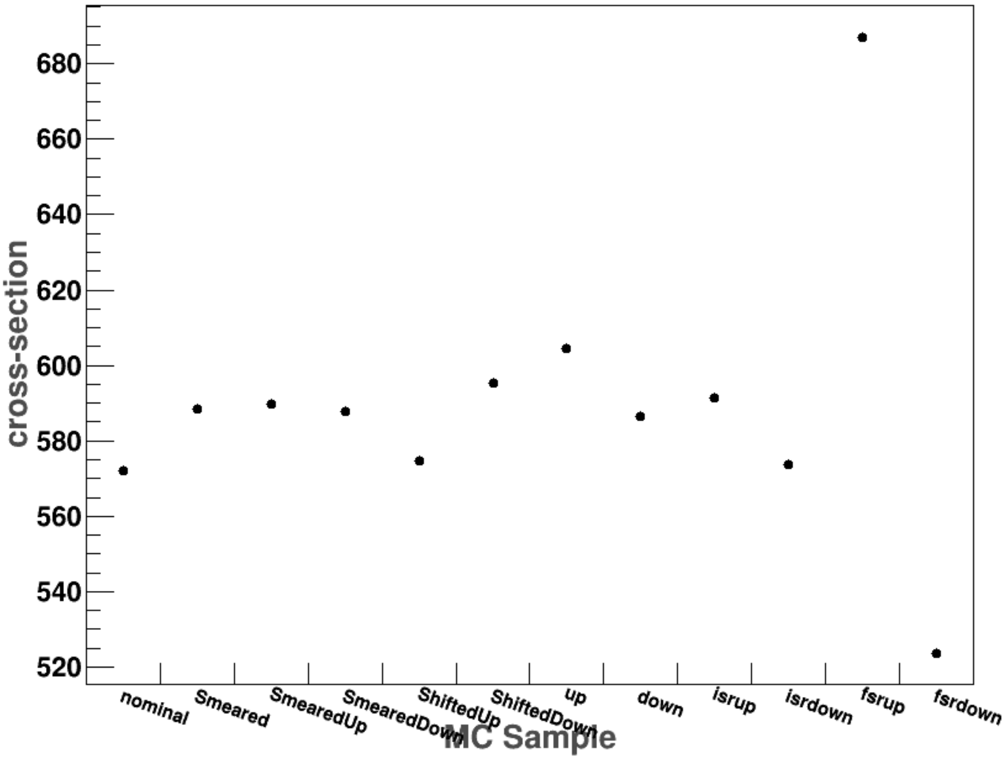


Fig. 5.5: Cross-section final calculations for all Monte Carlo samples used

The final cross-section measurements are

$$\sigma fid = 175.29 \pm 4.5\ (stat)^{+2.46}_{-9.58}(syst) + 0.15\big(lumi\big)\quad [fb]$$

for fiducial cross-section,

and by extrapolating to the full phase space, the inclusive (total) cross-section is calculated to be:

$$\sigma incl = 572 \pm 15(stat)^{+126}_{-48}(syst) + 15(lumi)\quad [pb]$$

to be compared with the theoretical prediction of
σfid = 384± 36 [fb]
and
$\sigma incl = 832^{+40}_{-46}\ [pb]$ respectively.

5.1 Comparison of Monte Carlo Signal Samples

A significant asymmetry between positive and negative shift is noticed and must be investigated further.
The first step is to see how calculated efficiency is affected by the application of the nominal selection cuts.

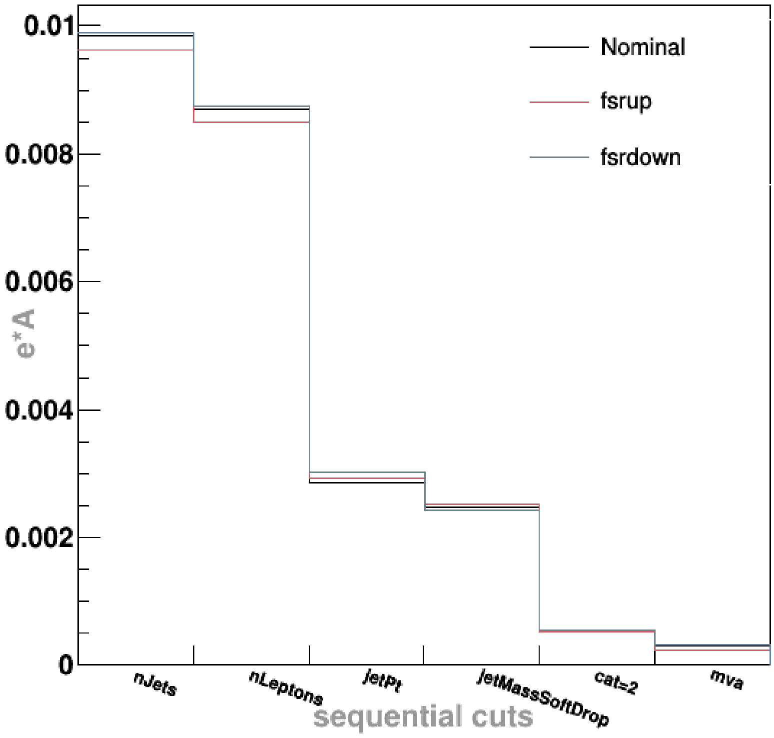


Fig. 5.6: Efficiency decreases while applying the sequential cuts

More info can be extracted from the plot showing the fractional difference of the two Monte Carlo samples (i.e. fsrup, fsrdown) that behave differently from the nominal sample. This plot is shown below.

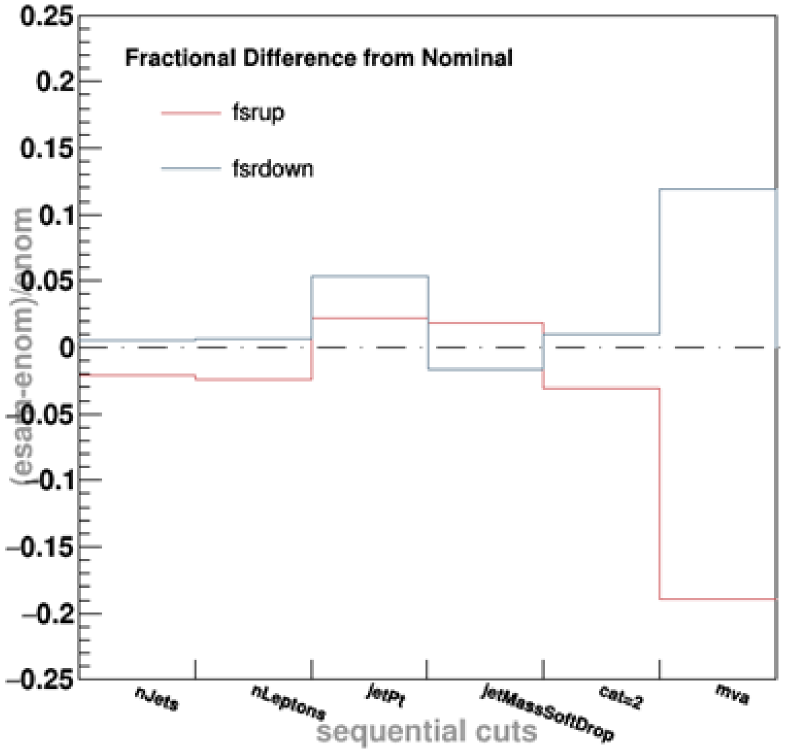


Fig. 5.7: Variation of fractional difference of efficiency while applying the sequential cuts

As can be seen from the plot before, the cut on mva is the one that makes these two samples display such difference from nominal sample.

A need to further look at any difference shown in 'N-subjetiness' variables between nominal and the two Monte Carlo samples is obvious.
Info is extracted from the following Templates, for nominal selection criteria apart but jet's Soft Drop mass in range 120-220 GeV.

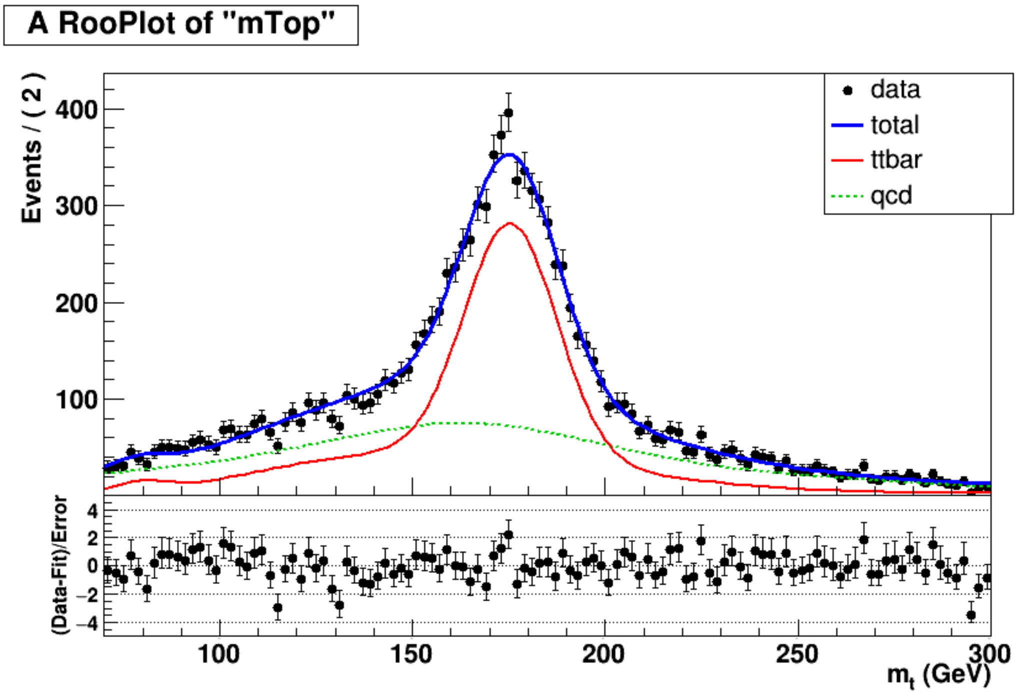


Fig 5.8: Template distribution for fsrup Monte Carlo sample.

After Fit info extracted:

N_{tt-} = 6163 ± 160 events
N_{QCD} = 4746 ± 155 events
chi2/ndof = 0.93

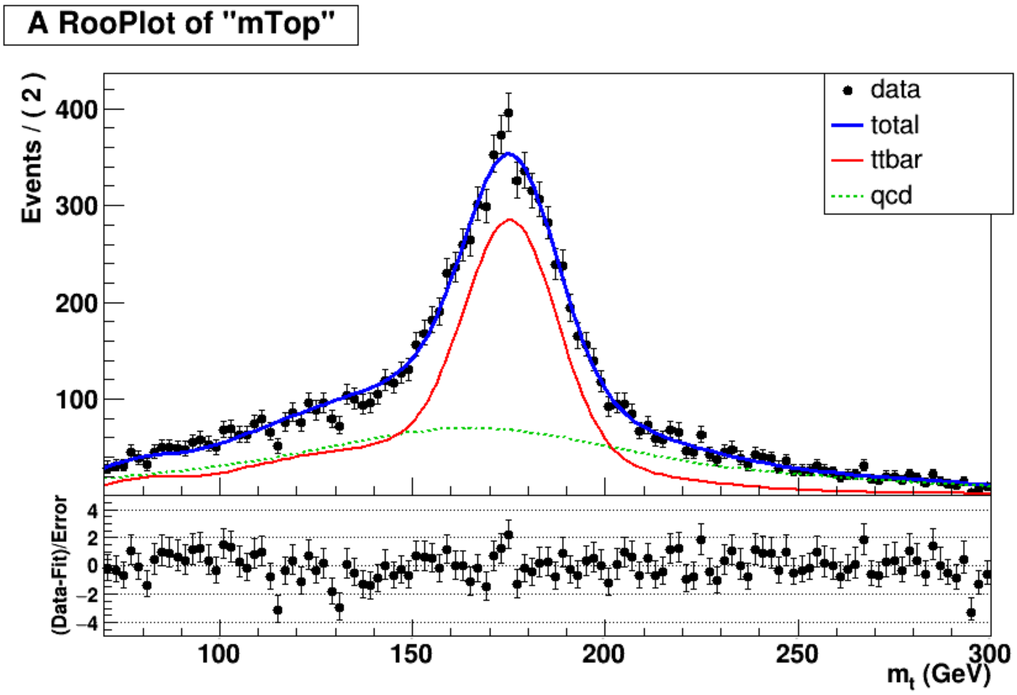


Fig 5.9: Template distribution for fsrdown Monte Carlo sample.

After Fit info extracted:

N_{tt-} = 6487 ± 167 events
N_{QCD} = 4422 ± 160 events
chi2/ndof = 0.94

In distributions presented below, the QCD background yield is extracted from data with 0 btag requirement while Signal yield is extracted from Monte Carlo fit (shown above).

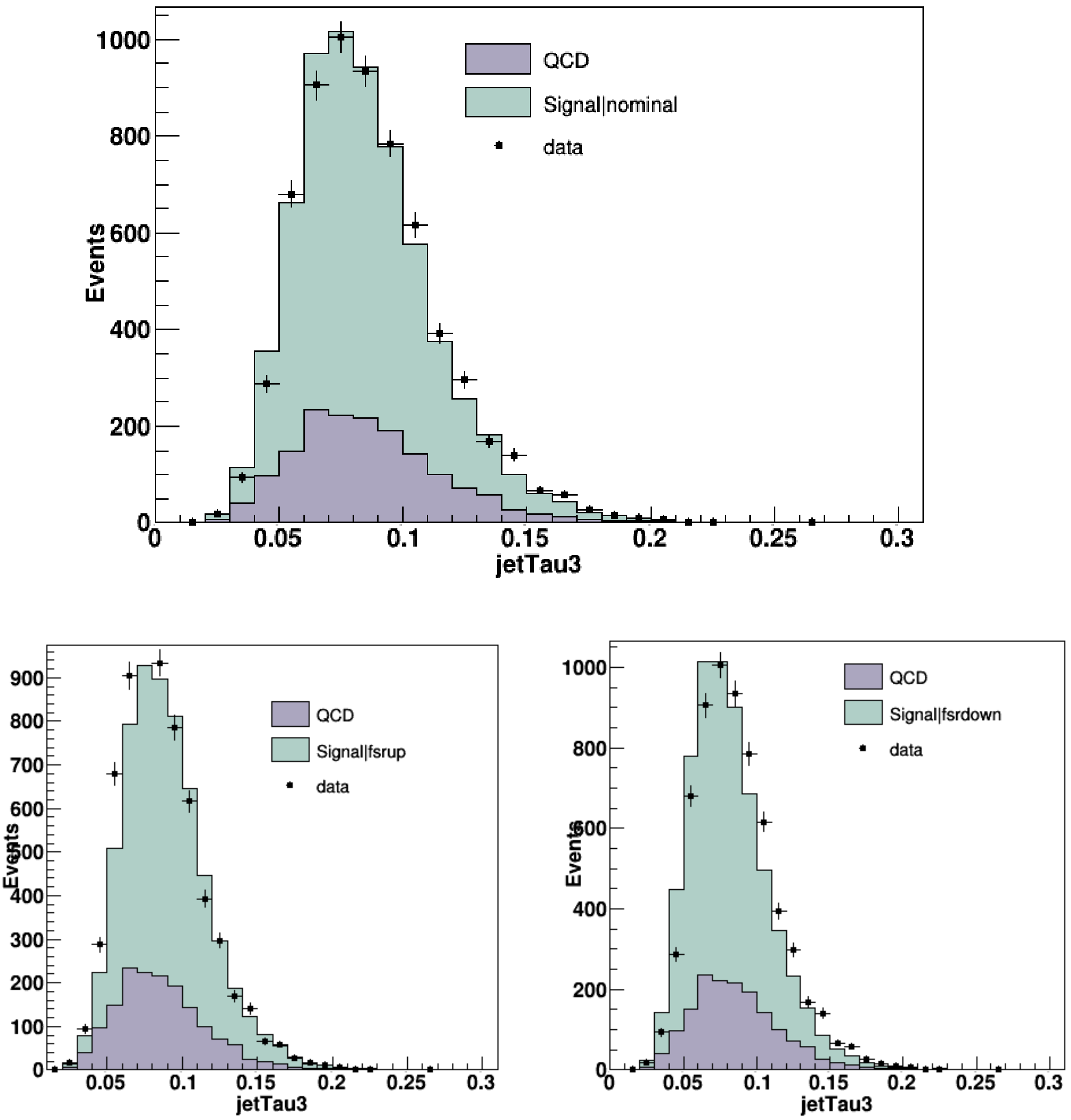


Fig 5.10: distibutions for 3 subjets present in fat jet for nominal (up), fsrup (down left) and fsrdown (down right) Monte Carlo samples

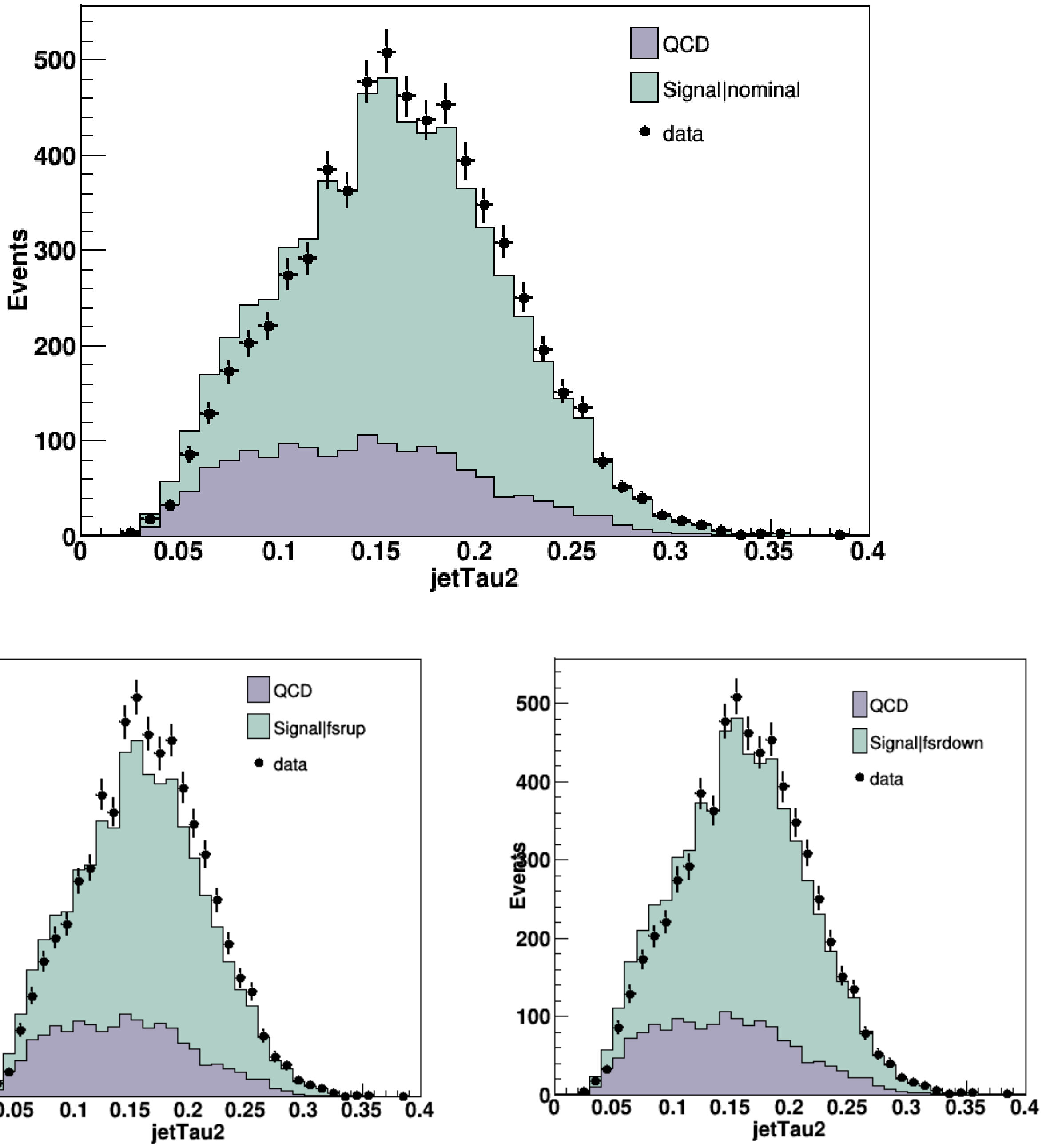


Fig 5.11: distibutions for 2 subjets present in fat jet for nominal (up), fsrup (down left) and fsrdown (down right) Monte Carlo samples

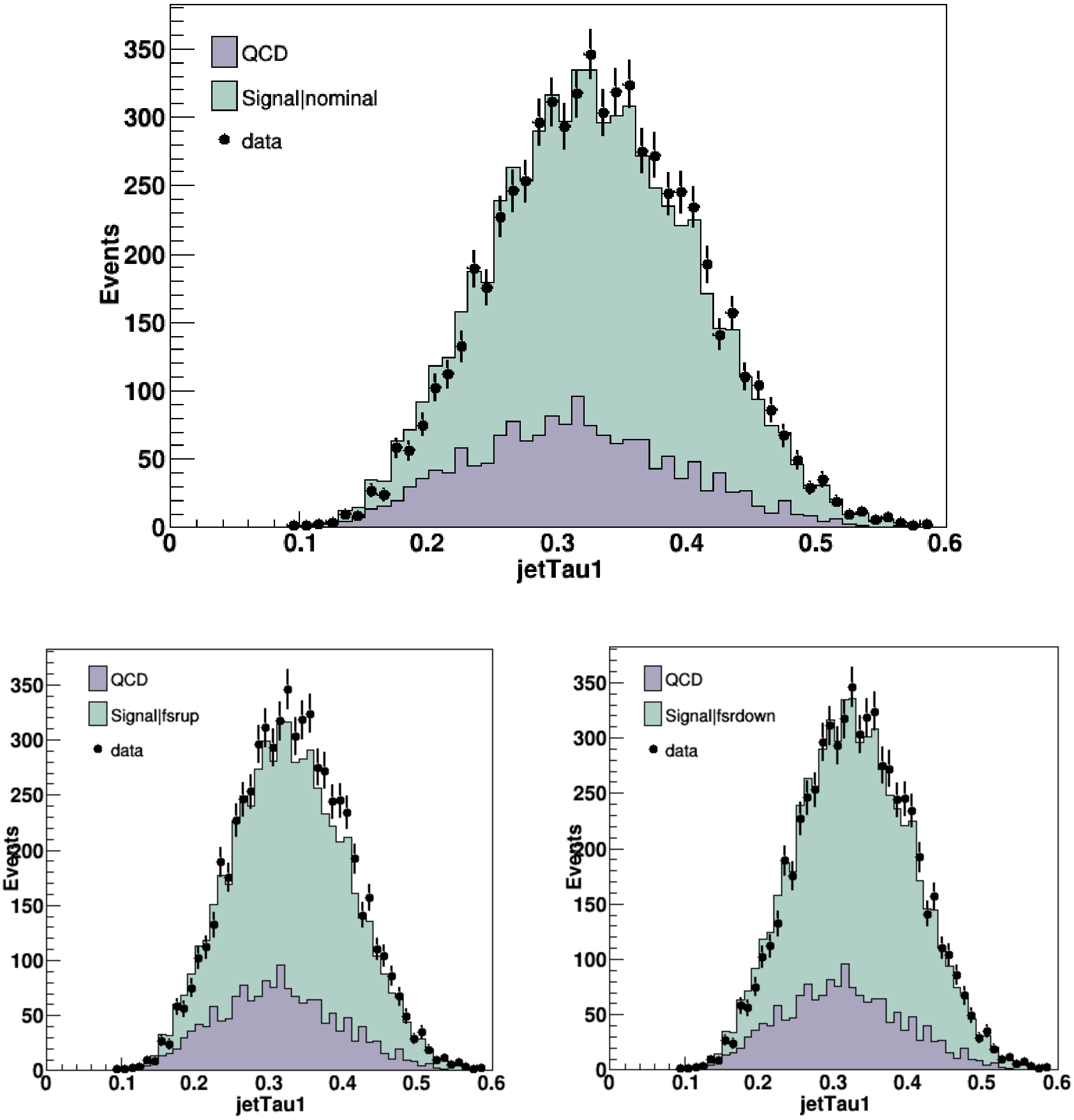


Fig 5.12: distibutions for 1 subjet present in fat jet for nominal (up), fsrup (down left) and fsrdown (down right) Monte Carlo samples

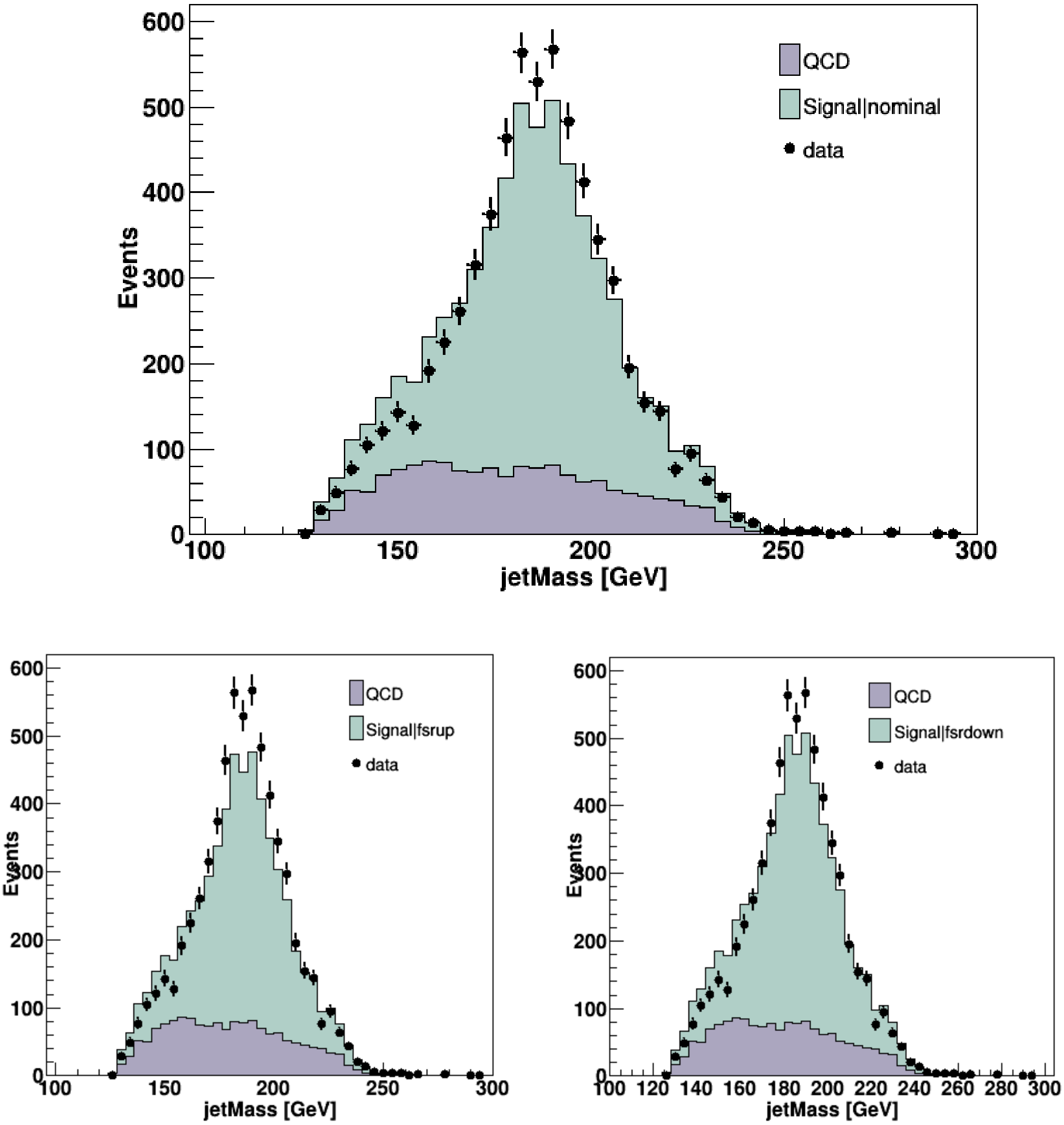


Fig 5.13: distibutions for jet’s mass for nominal (up), fsrup (down left) and fsrdown (down right) Monte Carlo samples

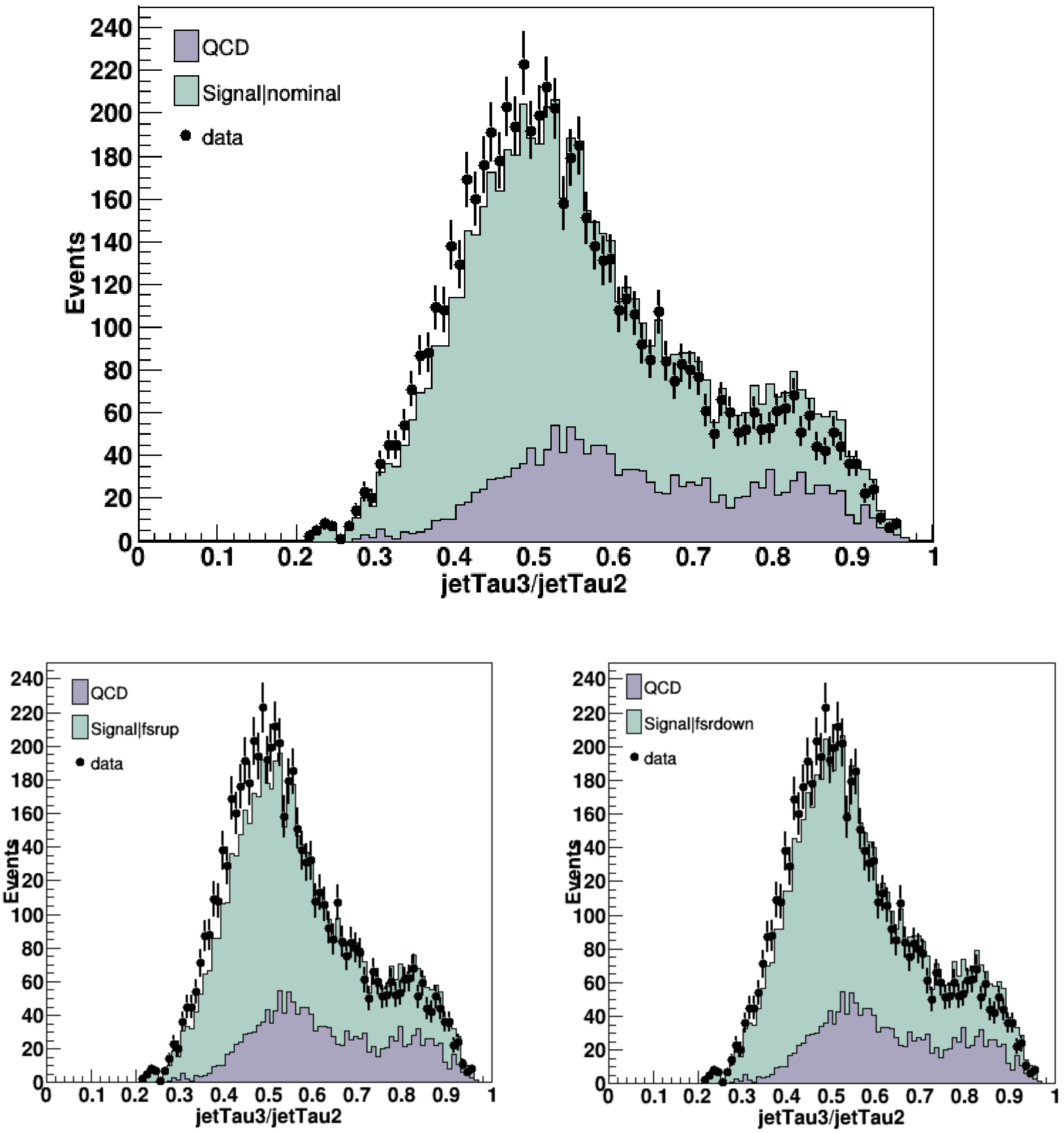


Fig 5.14: distibutions for 3 present subjets over 2 ratio for nominal (up), fsrup (down left) and fsrdown (down right) Monte Carlo samples

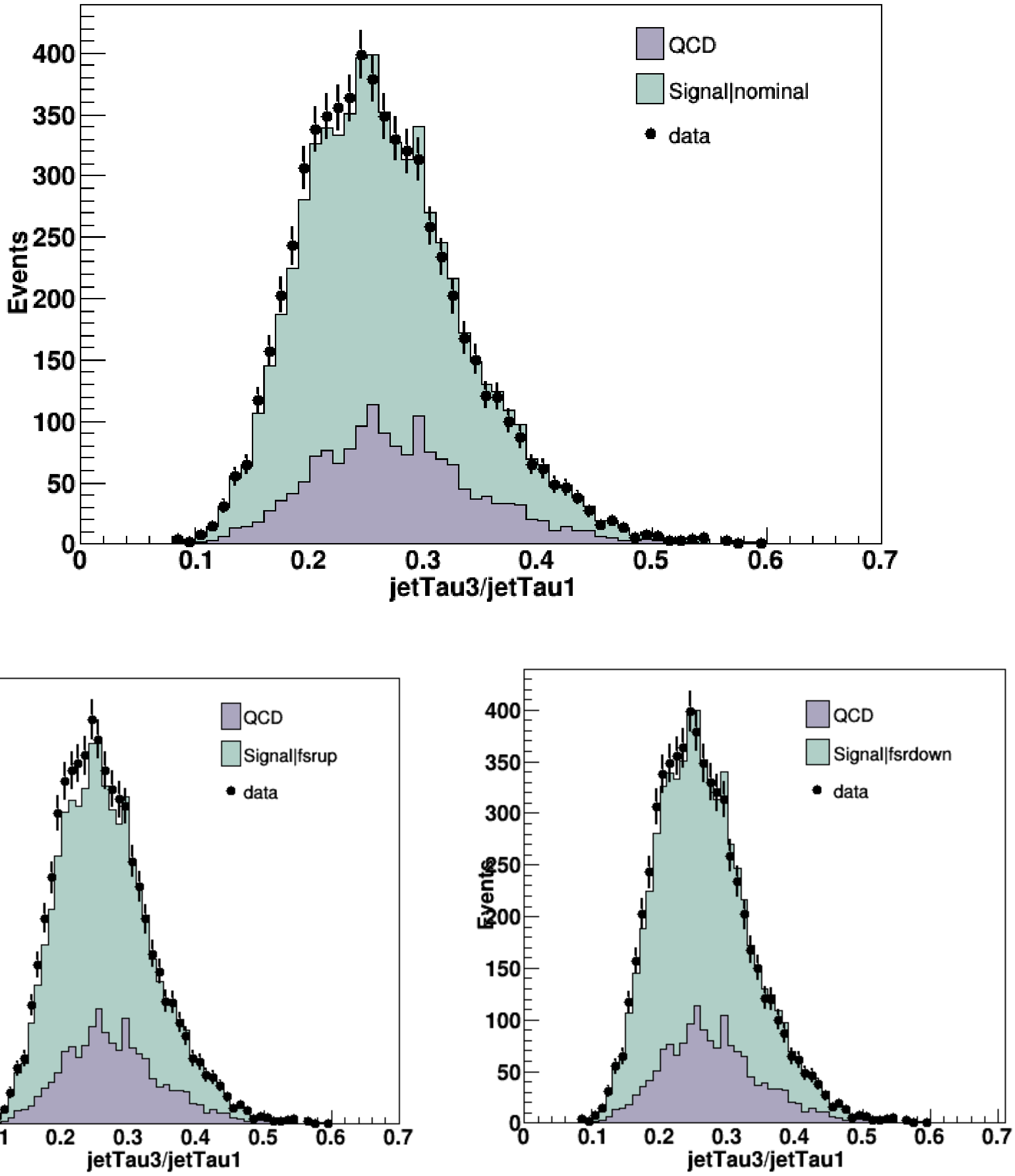
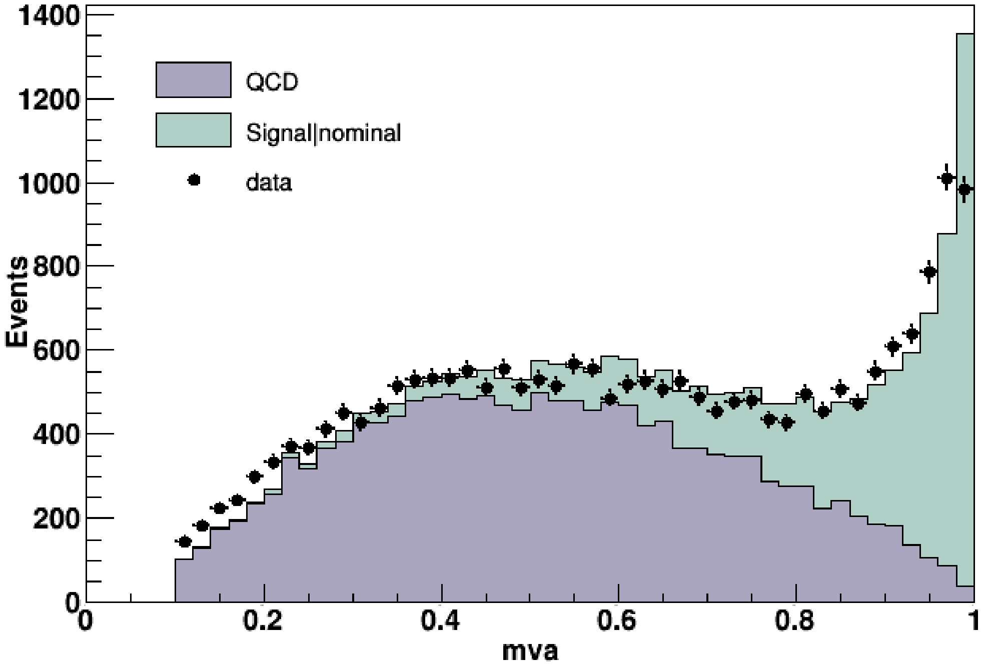
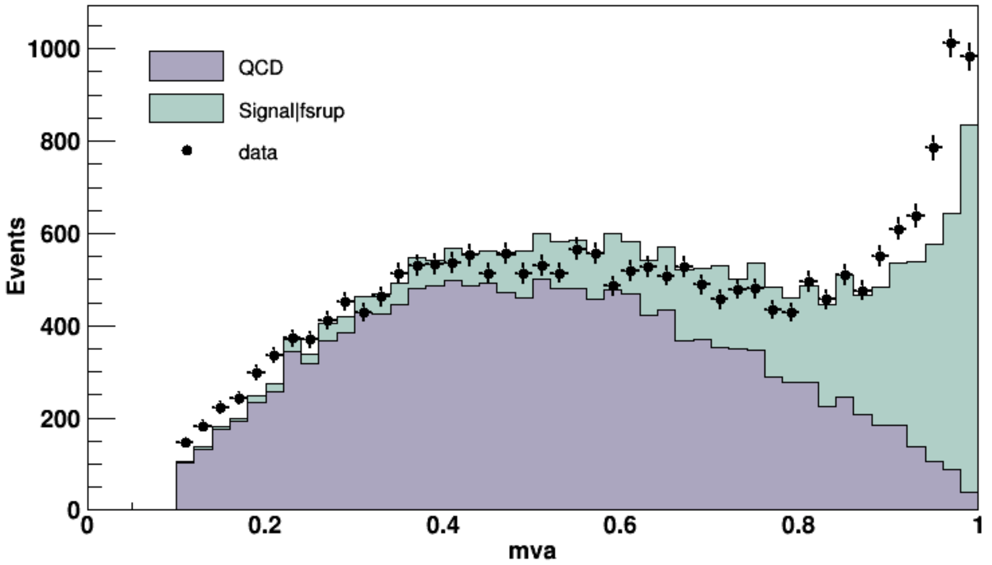


Fig 5.15: distibutions for 3 present subjets over 1 ratio for nominal (up), fsrup (down left) and fsrdown (down right) Monte Carlo samples

In the following plots, no mva cut is applied.


QCD
Signal|nominal
data
Events
mva
0
200
400
600
800
1000
1200
1400
0.2
0.4
0.6
0.8
1


QCD
Signal|fsrup
data
Events
mva
0
200
400
600
800
1000
0.2
0.4
0.6
0.8
1

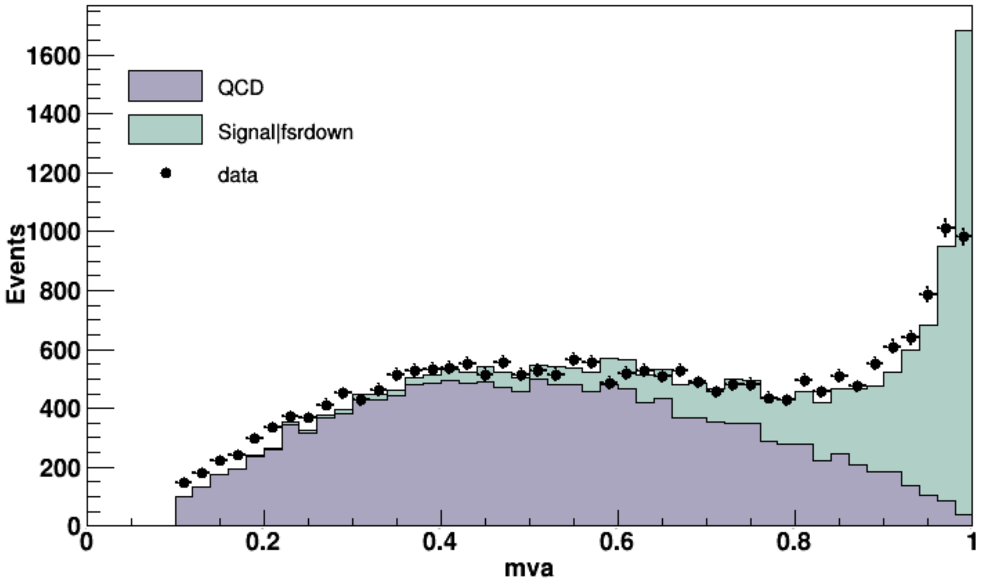


Fig 5.16: mva distribution for nominal (up), fsrup (middle) and fsrdown (down) Monte Carlo samples

From the distributions above, it is obvious that fsrup Monte Carlo sample predicts less signal events than nominal Monte Carlo. This leads to a higher cross-section measurement. The difference between Signal yield predicted by fsrup and fsrdown samples explain the positive or negative systematic shift accordingly.

5.2. Improvements

For final cross-section measurement to be closer to the theoretical prediction, based in NNLO+NNLL calculations, a few improvements could be made.

- Monte Carlo NNLO samples to be used; a ~ 30% improvement in final calculations would be achieved
- fsrup Monte Carlo Sample could be improved by first looking at data before deciding which settings of nominal powheg-pythia8 sample should be altered and to what extend; the way this sample has been created present a much different behavior than data, which is also obvious from final measurements.

Applying the suggestions above, a more consistent to the theoretical prediction final measurement would be possible.

Chapter 6

Conclusions

In this Analysis, the measurement of inclusive ttbar production cross section, using all-hadronic ttbar events in boosted topology has been presented. The data sample that was used was collected in 2016 by CMS experiment, amounting to an integrated luminosity of 37^{-1} fb.

For extracting signal events from data, the signal trigger HLT AK8DiPFJet280 200 TrimMass30 BTagCSV, was used, applying an extended selection to the event passing this trigger criteria. Using the CSVv2 algorithm for b-tagging, each one of the two leading jets were required to contain one b-tagged subjet. Starting from the two leading jet n-subjettiness jetTau1,2,3 a neural network was used, being trained to recognize background events from signal ones, using Monte Carlo samples.

Signal and Background template distributions were extracted, with QCD yield extracted by data, requiring 2 and 0 b-tagged subjets accordingly; then these distributions were evaluated by fitting the leading jet Soft Drop mass distribution. The Signal template was evaluadted by a fit on simulated Monte Carlo samples and QCD background template was fitted on a QCD-enriched control sample, extracted from data as stated above.

Finally, by fitting these templates on 2-btag data samples, the signal yield was obtained with along with the selection efficiency extracted from the ttbar simulated signal sample, allowed the calculation of the fiducial cross-section of: $\boldsymbol{\sigma fid = 175.29 \pm 4.5\ (stat)^{+2.46}_{-9.58}(syst) + 0.15(lumi)\ \ [fb]}$ and then inclusive ttbar production cross-section, which eventually lead to: $\boldsymbol{\sigma incl = 572 \pm 15(stat)^{+126}_{-48}(syst) + 15(lumi)\ \ [pb]}$.

This measurement presents a final cross-section much lower than the theoretical prediction as also described earlier in this Analysis. However, variations of ~ 30% have been observed since the nominal Monte Carlo that was used was Powheg+Pythia8, which overestimates the selection efficiency when compared to other simulations; a NNLO+NNLL Monte Carlo sample would improve the final measurements, making them more compatible to the theoretical predictions.

References

[1] S. Moch, “Expectations at LHC from hard QCD,” J. Phys., vol. G35, p. 073001, 2008

[2] N. Kidonakis, “Theoretical results for top-quark cross sections and distributions,” in 38th International Conference on High Energy Physics (ICHEP 2016) Chicago, IL, USA, August 03-10, 2016, 2016

[3] A. Blondel, E. Graverini, N. Serra, M. Shaposhnikov., Search for Heavy Right Handed Neutrinos at the FCC-ee. arXiv:1411.5230

[4] D0 Collaboration, S. Abachi et al., \Observation of the Top Quark," Phys. Rev. Lett. 74 (Apr, 1995) 2632{2637. doi:10.1103/PhysRevLett.74.2632

[5] S. W. Herb, D. C. Hom, L. M. Lederman, J. C. Sens, H. D. Snyder, J. K.Yoh, J. A. Appel, B. C. Brown, C. N. Brown, W. R. Innes, K. Ueno,T. Yamanouchi, A. S. Ito, H. J•ostlein, D. M. Kaplan, and R. D. Kephart, \Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions," Phys. Rev. Lett. 39 (Aug, 1977) 252{255. doi:10.1103/PhysRevLett.39.252.

[6] D. Rainwater D. Chakraborty, J. Konigsberg. Top Quark Physics. Annu. Rev. Nucl. Part. Sci. 2003. 53:301-51.

[7] K. Nakamura et al. (particle data group).

[8] Akira Shibata. Top Quark Physics at the LHC. arXiv:0808.0031v3 [hep-ex].

[9] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, Jesse Thaler, Soft Drop. arXiv:1402.2657 [hep-ph]

[10] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS-PAS-PFT-09-001 (2009).

[11] CMS Collaboration, “Measurement of the tt production cross section in the fully hadronic decay channel in pp collisions at √ s = 7 TeV", CMS PAS TOP-11-007

[12] CMS Collaboration, “Performance of b-Jet Identification in CMS”, CMS Physics Analysis Summary CMS-PAS-BTV-11-001 (2011)

[13] CMS Collaboration, “Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-002 (2010).

[14] Ulrich Husemann, ‘Top-Quark Physics: Status and Prospects’, arXiv:1704.01356v2 [hep-ex] 19 Apr 2017

[15] CMS Collaboration, ‘The CMS Particle Flow Algorithm’, arXiv:1401.8155 [hep-ex]

[16] Leandro G. Almeidaa,b, Seung J. Leec , Gilad Perezc , George Stermana , Ilmo Sung, “Template Overlap Method for Massive Jets”, arXiv:1006.2035 [hep-ph]

[17] Jose Juknevich , “Heavy resonances and top production Overview of top tagging techniques”, http://dx.doi.org/10.3204/DESY-PROC-2014-02/35

[18] Luc Demortier, Louis Lyons, “Everything you always wanted to know about pulls”, CDF/ANAL/PUBLIC/5776, Version 2.10, August 20, 2002

[19] Ulrich Husemann , “Top-Quark Physics: Status and Prospects”, Institut fur Experimentelle Kernphysik

[20] Jos´e Juknevich, “Heavy resonances and top production overview of top tagging techniques” http://dx.doi.org/10.3204/DESY-PROC-2014-02/35