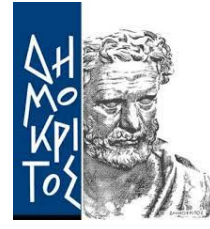




Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών
και Φυσικών Επιστημών



ΕΚΕΦΕ "Δημόκριτος"
Ινστιτούτο Πυρηνικής και
Σωματιδιακής Φυσικής

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
" Φυσική και Τεχνολογικές Εφαρμογές "

Χαρακτηρισμών αδρονικών πιδάκων από συγκρούσεις πρωτονίων στο πείραμα CMS του LHC

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΤΗΣ
ΑΝΝΑΣ ΖΑΧΑΡΟΠΟΥΛΟΥ

Επιβλέπων: Κωσταντίνος Κουσουρής
Επικ. Καθηγητής ΕΜΠ

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2019

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών
και Φυσικών Επιστημών

ΕΚΕΦΕ "Δημόκριτος"
Ινστιτούτο Πυρηνικής και
Σωματιδιακής Φυσικής

Χαρακτηρισμών αδρονικών πιδάκων από συγκρούσεις πρωτονίων στο πείραμα CMS του LHC

Μεταπτυχιακή Διπλωματική Εργασία
της

Άννας Ζαχαροπούλου

Επιβλέπων:

Κ. Κουσουρής
Καθηγητής ΕΜΠ

Τριμελής Επιτροπή

Κ. Κουσουρής
Επικ. Καθηγητής
ΕΜΠ
(επιβλέπων)

Γ. Τσιπολίτης
Καθηγητής
ΕΜΠ

Ν. Τράκας
Καθηγητής
ΕΜΠ

Αθήνα, Φεβρουάριος 2019

Περίληψη

Η παρούσα διπλωματική εργασία είχε ως σκοπό τη μελέτη των αδρονικών και pile up πιδάκων που σχηματίζονται κατά τις συγκρούσεις πρωτονίων, στα δεδομένα του 2017 που συλλέχθηκαν από τον ανιχνευτή CMS στα 13 TeV στο κέντρο μάζας.

Στις τρέχουσες συνθήκες λειτουργίας του Large Hadron Collider οι pile up πίδακες (jets) αποτελούν ένα σταθερό υπόβαθρο στα δεδομένα. Η παρουσία τους οφείλεται στην επικάλυψη των jets με χαμηλό p_T και αυξάνεται τετραγωνικά με τον αριθμό των pile up συγκρούσεων. Εξαιτίας του σχηματισμού τους εμφανίζουν χαρακτηρισικά βάσει των οποίων είναι εφικτός ο διαχωρισμός τους από τα jets τα οποία προέρχονται από quarks ή gluons.

Σε αυτήν την ανάλυση έγινε συνδυασμός 14 συνολικά μεταβλητών σε ένα πολυπαραγοντικό ταξινομητή (BDT) με σκοπό την ταυτοποίηση και την αφαίρεση των pile up jets στις δύο περιοχές του ανιχνευτή. Το αποτέλεσμα του ταξινομητή χρησιμοποιήθηκε για τη μεγιστοποίηση της αποδοτικότητας αυτού ως συνάρτηση της εγκάρσιας ορμής των jets. Τέλος πραγματοποιήθηκε επικύρωση της μεθόδου που αναπτύχθηκε με τα δεδομένα του 2017, που συλλέχθηκαν από τον ανιχνευτή CMS.

Abstract

The aim of this thesis was the study of hadronic and pile up jets during proton collisions in 2017 data that has been collected from CMS detector at 13 TeV.

Pileup jets are a ubiquitous background under the current running conditions of the Large Hadron Collider. Their presence typically arises from overlapping low p_T jets and grows roughly quadratically with the number of pileup collisions. Due to their unusual formation, pileup jets exhibit distinct features that allow them to be separated from real jets that have originated from either quarks or gluons.

Identification and removal of pileup jets is performed either through the use of vertex information or through the use of shape information. The former can be exploited in the central region of CMS detector where tracking is available. Shape information, although less effective than vertexing, extends throughout the whole detector volume and in conjunction with vertex information enhances the ability to identify pileup jets. Even though PUPPI (Pile Up Per Particle) algorithm is an efficient way to remove pileup it is not always due to the fact that random pileup clusters might have created incidentally. That is the reason why further removal of pileup must be done through Pileup jet Identification Algorithm.

In this analysis fourteen in total number vertex and shape variables were combined through a multivariate BDT that gave the pileup id. In addition, this pileup id was used in order to maximize and stabilize the efficiency as a function of jet p_T in central and forward region. Finally, the validation of this method was performed with 2017 data of CMS detector.

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κύριο Κωνσταντίνο Κουσουρή, Επίκουρο καθηγητή του Εθνικού Μετσόβιου Πολυτεχνείου, που με εμπιστεύτηκε και μου έδωσε την ευκαιρία να ασχοληθώ με το συγκεκριμένο θέμα, καθώς και για τη συνεχή βοήθεια και καθοδήγηση που μου παρείχε όλο αυτό το διάστημα. Στη συνέχεια, θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής Γεώργιο Τσιπολίτη και Νίκο Τράκα καθηγητές του Εθνικού Μετσόβιου Πολυτεχνείου, και ιδιαιτέρως τον κύριο Γ. Τσιπολίτη για τη συνεργασία και τη συνεχή υποστήριξή του τα τελευταία χρόνια.

Επιπλέον, θα ήθελα να ευχαριστήσω τους Γιάννη Παπακριθόπουλο και Γιώργο Μπάκα, διδακτορικούς φοιτητές για τη βοήθειά τους και τις επικοινωνητικές παρατηρήσεις τους κατά τη διάρκεια της διεκπεραίωσης της διπλωματικής μου εργασίας, όπως επίσης και την Ειρήνη Σιαμάρκου, μεταπτυχιακή φοιτήτρια για τη συνεργασία μας όλο αυτό το διάστημα. Επιπλέον, ευχαριστώ από καρδιάς την οικογένειά μου και τους φίλους μου παλιούς και καινούργιους, που είχα τη χαρά να γνωρίσω κατά τη διάρκεια αυτής της χρονιάς, για τη σημαντική στήριξη και την υποστήριξη τους.

Περιεχόμενα

Κατάλογος Σχημάτων

Κεφάλαιο 1

Φυσική Στοιχειωδών Σωματιδίων

Η Φυσική Στοιχειωδών Σωματιδίων έχει ως στόχο τη μελέτη και κατανόηση των θεμελιωδών στοιχείων της ύλης. Για πάρα πολλά χρόνια στον τομέα αυτό επιστήμονες προσπαθούν να περιγράψουν θεωρητικά αλλά και να επιβεβαιώσουν πειραματικά από τι αποτελείται η ύλη γύρω μας και ποια είναι τα βασικά συστατικά του σύμπαντος. Στο κεφάλαιο αυτό γίνεται μια σύντομη αναφορά στο Καθιερωμένο Πρότυπο που περιγράφει με έναν κομψό μαθηματικά τρόπο τα στοιχειώδη σωματίδια, κατηγοριοποιώντας τα σε ομάδες καθώς και τον τρόπο με τον οποίο αυτά αλληλεπιδρούν. Επίσης, σκιαγραφούνται δυο μηχανισμοί που περιγράφουν τις αλληλεπιδράσεις των στοιχειωδών σωματιδίων, η Κβαντική Ηλεκτροδυναμική (QED) και η Κβαντική Χρωμοδυναμική (QCD) και τέλος τα βασικά χαρακτηριστικά της ασθενούς και ισχυρής αλληλεπίδρασης.

1.1 Το Καθιερωμένο Πρότυπο

Το πληρέστερο μοντέλο που υπάρχει για την περιγραφή των στοιχειωδών σωματιδίων είναι το καθιερωμένο πρότυπο (Standard Model, SM). Μέχρι στιγμής το SM περιγράφει με πολύ καλή ακρίβεια όλα τα πειραματικά αποτελέσματα που έχουν παρατηρηθεί.

Σύμφωνα με το SM τα στοιχειώδη σωματίδια χωρίζονται σε 2 κατηγορίες όπως φαίνεται και στην εικόνα ??, τα μποζόνια και τα φερμιόνια. Τα φερμιόνια είναι τα σωματίδια από τα οποία αποτελείται η ύλη και χωρίζονται σε 2 κατηγορίες τα λεπτόνια και τα κουάρκς. Η κατηγορία των λεπτονίων αποτελείται από το ηλεκτρόνιο, το μιονίο και το ταυ. Κάθε ένα από αυτά τα σωματίδια μπορεί να βρεθεί ελεύθερο στη φύση. Επίσης για κάθε ένα από τα σωματίδια αυτά, υπάρχει και ένα αντίστοιχο νεutrino που τα "συνοδεύει". Τα κουάρκς είναι 6 και δεν μπορούν να βρεθούν ελεύθερα, αλλά σε συνδυασμούς των 2 ή των 3 δημιουργούν άλλα σωματίδια, τα αδρόνια.

Εκτός από την ύπαρξη των στοιχειωδών σωματιδίων, το SM περιγράφει και τον τρόπο με τον οποίο αυτά αλληλεπιδρούν. Στη φύση υπάρχουν 4 στοιχειώδεις δυνάμεις οι οποίες διέπουν τις αλληλεπιδράσεις των σωματιδίων. Η ισχυρή, η ασ-

θενής, η ηλεκτρομαγνητική και η βαρύτητα. Η βαρύτητα καθώς είναι πολύ ασθενής δεν επηρεάζει τα στοιχειώδη σωματίδια και γι' αυτό δεν συμπεριλαμβάνεται στο Καθιερωμένο Πρότυπο. Για κάθε μια από αυτές τις δυνάμεις το SM προβλέπει την ύπαρξη σωματιδίων (μποζονίων) μέσω των οποίων εκφράζεται η κάθε δύναμη.

	mass →	charge →	spin →					
QUARKS	$\approx 2.3 \text{ MeV}/c^2$	$2/3$	$1/2$	u up	$\approx 1.275 \text{ GeV}/c^2$	$2/3$	$1/2$	c charm
					$\approx 173.01 \text{ GeV}/c^2$	$2/3$	$1/2$	t top
								g gluon
								H Higgs boson
					$\approx 4.8 \text{ MeV}/c^2$	$-1/3$	$1/2$	d down
					$\approx 95 \text{ MeV}/c^2$	$-1/3$	$1/2$	s strange
LEPTONS					$\approx 4.18 \text{ GeV}/c^2$	$-1/3$	$1/2$	b bottom
								γ photon
	$0.511 \text{ MeV}/c^2$	-1	$1/2$	e electron	$105.7 \text{ MeV}/c^2$	-1	$1/2$	μ muon
					$1.777 \text{ GeV}/c^2$	-1	$1/2$	τ tau
								Z Z boson
					$< 2.2 \text{ eV}/c^2$	0	$1/2$	ν_e electron neutrino
GAUGE BOSONS					$< 0.17 \text{ MeV}/c^2$	0	$1/2$	ν_μ muon neutrino
					$< 15.5 \text{ MeV}/c^2$	0	$1/2$	ν_τ tau neutrino
					$80.4 \text{ GeV}/c^2$	± 1	1	W W boson

Figure 1.1: Ο πίνακας των στοιχειωδών σωματιδίων σύμφωνα με το καθιερωμένο πρότυπο. Τα σωματίδια χωρίζονται σε 2 κατηγορίες τα φερμιόνια και τα μποζόνια. Τα φερμιόνια είναι τα σωματίδια από τα οποία αποτελείται η ύλη και χωρίζονται σε δύο κατηγορίες, τα λεπτόνια και τα κουάρκς. Τα λεπτόνια μπορούν να βρεθούν ελεύθερα στη φύση ενώ τα κουάρκς όχι. Τα κουάρκς σε συνδυασμό 2 ή 3 σχηματίζουν άλλα σωματίδια, τα αδρόνια.

Επιπλέον, το μοντέλο αυτό προβλέπει την ύπαρξη ενός ακόμα σωματιδίου το οποίο ευθύνεται για την μάζα των υπολοίπων. Το σωματίδιο αυτό είναι το μποζόνιο του Higgs και δεν είχε παρατηρηθεί ποτέ. Ένας από τους στόχους των πειραμάτων που διεξάγονται στο LHC είναι η παρατήρηση του σωματιδίου αυτού. Το καλοκαίρι του 2012 ανακοινώθηκε από τα πειράματα του LHC η παρατήρηση του συγκεκριμένου σωματιδίου και έτσι ο LHC κατάφερε να επιτύχει έναν από τους σημαντικότερους στόχους του. Όπως αναφέρθηκε και παραπάνω, το SM δεν περιγράφει τη βαρύτητα. Είναι δεδομένο συνεπώς πως σαν μοντέλο είναι ελλιπές καθώς δεν περιλαμβάνει όλες τις αλληλεπιδράσεις που συμβαίνουν στη φύση. Ένας από τους στόχους του LHC είναι να προσπαθήσει να παρατηρήσει φυσικές διεργασίες που δεν περιγράφονται από το καθιερωμένο πρότυπο. Να ψάξει δηλαδή για ενδείξεις για καινούργια φυσική ή αλλιώς για φυσική πέρα από το καθιερωμένο πρότυπο (Beyond the Standard Model, BSM). Μια από τις πιο υποσχόμενες θεωρίες η ύπαρξη της οποίας θα προσπαθήσει να ελεγχθεί από τον LHC είναι αυτή της υπερσυμμετρίας (Supersymmetry).

1.2 Κβαντική Ηλεκτροδυναμική- Quantum Electrodynamics QED

Η Κβαντική Ηλεκτροδυναμική (QED) είναι ο πρώτος μηχανισμός που περιέγραψε τις γνωστές αλληλεπιδράσεις σε μια κβαντική θεωρία πεδίου. Το φορτίο της QED είναι το ηλεκτρομαγνητικό φορτίο και ανάγεται συνήθως σε μονάδες του θεμελιώδους φορτίου του ηλεκτρονίου. Στην QED απαιτείται η Λαγκρανζιανή να είναι αναλλοίωτη κάτω από έναν τυπικό μετασχηματισμό βαθμίδας της $U(1)$ ομάδας συμμετρίας. Η Λαγκρανζιανή που περιγράφει ένα ελεύθερο πεδίου Dirac (όπως του ηλεκτρονίου) με μάζα m είναι

$$\mathcal{L}_{free} = \bar{\psi}(x)(i\partial_\mu\gamma^\mu - m)\psi(x) \quad (1.1)$$

όπου ο πρώτος όρος είναι ο κινηματικός όρος και ο δεύτερος είναι ο όρος της μάζας. Κάτω από έναν τοπικό $U(1)$ μετασχηματισμό φάσης έχουμε για την κυματοσυνάρτηση ψ , που είναι ένας σπίνορας Dirac

$$\psi \rightarrow \psi' = \exp(i\theta(x)a)\psi \quad (1.2)$$

όπου το $\theta(x)$ αντιστοιχεί σε μια δύναμη σύζευξης και το a αναπαριστά μια φάση που εξαρτάται από τις χωροχρονικές συντεταγμένες. Ο όρος της μάζας είναι αναλλοίωτος κάτω από τον μετασχηματισμό αυτό. Από την άλλη για την κλίση $\partial^\mu\psi(x)$ υπάρχει ένας επιπλέον όρος

$$\partial^\mu\psi(x) \rightarrow \exp(i\theta(x)a)\partial^\mu\psi(x) - i\exp(-i\theta(x)a)\partial^\mu\psi(x)\partial^\mu\theta(x) \quad (1.3)$$

οπότε εισάγεται το διανυσματικό πεδίο που μετασχηματίζεται

$$A^\mu(x) \rightarrow A^\mu(x) - \frac{1}{e}\partial^\mu\theta(x) \quad (1.4)$$

Επίσης, ορίζεται η συναλλοίωτη παράγωγος

$$D^\mu\psi(x) = (\partial^\mu + ieA^\mu(x))\psi(x) \quad (1.5)$$

που μετασχηματίζεται με τον ίδιο τρόπο όπως το πεδίο. Έτσι έχουμε για την Λαγκρανζιανή

$$\mathcal{L}_{free} = \bar{\psi}(x)(iD_\mu\gamma^\mu - m)\psi(x) = \mathcal{L}_{free} - eA_\mu\bar{\psi}(x)\gamma^\mu\psi(x) \quad (1.6)$$

Ο δεύτερος όρος δίνει την αλληλεπίδραση του πεδίου Dirac με το διανυσματικό πεδίο. Η πλήρης Λαγκρανζιανή της QED επιτυγχάνεται αν προσθέσουμε έναν κινηματικό όρο για το ηλεκτρομαγνητικό πεδίο και έναν όρο βαθμίδας (βαθμίδα Lorentz) ώστε να εξασφαλίζεται η ύπαρξη του διαδότη του φωτονίου

$$\mathcal{L}_\gamma = \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{2}\partial_\mu A_\mu \partial_\nu A^\nu \quad (1.7)$$

όπου $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ είναι ο ηλεκτρομαγνητικός τανυστής με αποτέλεσμα

$$\mathcal{L}_{QED} = \mathcal{L} + \mathcal{L}_\gamma \quad (1.8)$$

Στην παραπάνω εξίσωση δεν θα μπορούσαμε να προσθέσουμε έναν όρο μάζας mAA_μ διότι παραβιάζει το αναλλοίωτο βαθμίδας, γεγονός που οδηγεί στο συμπέρασμα ότι το φωτόνιο δεν έχει μάζα.

1.3 Κβαντική Χρωμοδυναμική- Quantum Chromodynamics QCD

Η περίπτωση της QCD είναι είναι παρόμοια με αυτήν της QCD αν και κάπως πιο περίπλοκη λόγω της μη αβελιανής δομής της $SU_C(3)$ που αντιστοιχεί στη συμμετρία χρώματος. Στην QCD το φορτίο είναι το χρώμα και υπάρχει σε 3 μορφές, πράσινο, κόκκινο και μπλε και κάθε χρώμα έχει και το αντιχρώμα. Η Λαγκρανζιανή πυκνότητα γράφεται

$$\mathcal{L}_{QCD} = -\frac{1}{4}G_{\alpha}^{\mu\nu}G_{\mu\nu}^{\alpha} + i\bar{\psi}_j^{\alpha}\gamma^{\mu}(D_{\mu})_{\alpha\beta}\psi_j^{\beta} - m_j\bar{\psi}_j^{\alpha}\psi_{j,\alpha} - \frac{1}{2}\partial^{\mu}A_{\mu}^{\alpha}\partial_{\mu}A_{\mu}^{\alpha} - \partial_{\mu}\bar{\varphi}_{\alpha}D^{\mu}\varphi^{\alpha} \quad (1.9)$$

όπου $G_{\mu\nu}^{\alpha} = \partial_{\mu}A_{\nu}^{\alpha} - \partial_{\nu}A_{\mu}^{\alpha} + gf_{abc}A_{\mu}^bA_{\nu}^c$ ($a = 1, 2, \dots, 8$) είναι ο τανυστής έντασης του πεδίου των γκλουονίων ψ_j είναι το πεδίο quark γεύσης j , ενώ φ^{α} είναι τα 8 βαθμωτά πεδία που ακολουθούν στατιστική φερμιονίων και ονομάζονται Faddeev-Popov ghosts particles τα οποία είναι αναγκαία για μια συνεπή κβάντωση της θεωρίας, μέσω των ολοκληρωμάτων διαδρομών, αφού εξαλείφουν αφύσικες καταστάσεις στον διαδότη των γκλουονίων (εξαλείφουν το διαμήκες μέρος που εμφανίζεται σε βρόχους γκλουονίων). $(D_{\mu})_{\alpha\beta} = \delta_{\alpha\beta}\partial_{\mu} - ig\sum_{\alpha}\frac{1}{2}\lambda_{\alpha\beta}^{\alpha}A_{\mu}^{\alpha}$ είναι η συναλλοίωτη παράγωγος που δρα στο χρώμα των quark, g είναι ο όρος που περιγράφει τη σύζευξη των κουαρκς με την ισχυρή δύναμη, ενώ $\lambda_{\alpha\beta}^{\alpha}$ είναι οι 8 3×3 πίνακες χρώματος και f_{abc} είναι οι πραγματικές σταθερές δομής της $SU(3)$ άλγεβρας Lie. Ο προτελευταίος όρος είναι για την επιλογή βαθμίδας (βαθμίδα Lorentz), που είναι απαραίτητος για την συναλλοίωτη κβάντωση του πεδίου των γκλουονίων.

Χρησιμοποιώντας τους γεννήτορες T_{α} οι οποίοι εκφράζουν τις περιστροφές στον τρισδιάστατο χώρο των χρωμάτων που περιγράφεται από την $SU_C(3)$. Για τους οποίους ισχύει:

$$[T_{\alpha}, T_{\beta}] = if_{abc}T_c\kappa_{\alpha\beta}(T^{\alpha})_{\alpha\beta} = \frac{1}{2}\lambda_{\alpha\beta}^{\alpha} \quad (1.10)$$

Η συναλλοίωτη παράγωγος γίνεται $D_{\mu} = \partial_{\mu} + igT^{\alpha}G_{\mu}^{\alpha}$. Επειδή, οι γεννήτορες της $SU_C(3)$ δεν μετατίθενται μεταξύ τους, επιτρέπονται και αλληλεπιδράσεις μεταξύ γκλουονίων. Η Λαγκρανζιανή επομένως (σε μια μορφή που μοιάζει με τη Λαγκρανζιανή της QED για πιο εύκολη σύγκριση) είναι

$$\mathcal{L}_{QCD} = \bar{\psi}(i\partial_{\mu}\gamma^{\mu} - m)\psi - g\bar{\psi}(\gamma^{\mu}T^{\alpha}G_{\mu}^{\alpha})\psi - \frac{1}{4}G_{\alpha}^{\mu\nu}G_{\mu\nu}^{\alpha} \quad (1.11)$$

όπου ο πρώτος όρος είναι ο κινηματικός όρος και ο όρος της μάζας, ο δεύτερος όρος είναι η αλληλεπίδραση μεταξύ κουάρκ-κουάρκ και κουάρκ-γκλουονίων και ο τρίτος όρος είναι ο όρος αλληλεπίδρασης γκλουονίου-γκλουονίου.

Το νέο στοιχείο σε σχέση με την αβελιανή QED είναι η αλληλεπίδραση του γκλουονίου-γκλουονίου και λόγω αυτού, η εισαγωγή του πεδίου Faddeev-Popov. Η \mathcal{L}_{QCD} είναι τοπικά αναλλοίωτη κάτω από το μετασχηματισμό BRST. Έτσι, αποδεικνύεται ότι σε όλες τις τάξεις της θεωρίας διαταραχών το μη εγκάρσιο τμήμα του διαδότη του γκλουονίου παραμένει το ίδιο με τον ελεύθερο διαδότη και δεν συμβάλλει στην ενεργό διατομή διαφόρων σκεδάσεων.

1.4 Αλληλεπιδράσεις

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

Gauge Boson	Interaction	mass	J^P	Q	T_3
γ	EM	-	1^-	0	0
gluon	Strong	-	1^-	0	-
Z^0	weak	91.18GeV	1	0	0
W	weak	80.40GeV	1	$\pm e$	± 1

table 1.1: Τα μποζόνια βαθμίδας τα οποία αποτελούν τους φορείς των αντιστοιχών αλληλεπιδράσεων μεταξύ των στοιχειωδών σωματιδίων και μερικά βασικά χαρακτηριστικά τους. Το J δηλώνει την angular momentum του φορέα, P την parity, Q το ηλεκτρικό φορτίο του και T_3 την τρίτη συνιστώσα του weak isospin.

1.4.1 Ασθενής και Ηλεκτρασθενής Αλληλεπίδραση

Φορέας της ηλεκτρομαγνητικής αλληλεπίδρασης είναι το φωτόνιο, το οποίο έχει μηδενική μάζα. Στην ασθενή αλληλεπίδραση γίνεται ανταλλαγή βαθμωτών μποζονίων τα οποία έχουν μάζα και επομένως μικρό εύρος. Η μάζα του ηλεκτρικά ουδέτερου Z μποζονίου είναι $91.1876 \pm 0.0021 \text{ GeV}/c^2$ και έχει πλάτος διάσπασης $2.49852 \pm 0.0023 \text{ GeV}/c^2$, ενώ του W^\pm μποζονίου με θετικό ή αρνητικό ηλεκτρικό φορτίο είναι $80.385 \pm 0.015 \text{ GeV}/c^2$ και έχει πλάτος διάσπασης $2.085 \pm 0.042 \text{ GeV}/c^2$. Το φορτίο της ισχυρής αλληλεπίδρασης είναι το ισοσπίν, που συναντάται στα μποζόνια αυτά. Η ομάδα συμμετρίας που περιγράφει την ασθενή αλληλεπίδραση είναι η $SU(2)_L$ όπου L μας δείχνει την προτίμηση ζεύξης με αριστερόστροφα σωματίδια και δεξιόστροφα αντισωματίδια. Τα μποζόνια W^\pm πραγματοποιούν ζεύξη μόνο με αριστερόστροφα σωματίδια, σε αντίθεση με το ουδέτερο μποζόνιο Z το οποίο πραγματοποιεί ζεύξη και με δεξιόστροφα αντισωματίδια. Τα αριστερόστροφα κουαρκ και λεπτόνια είναι διατεταγμένα σε διπλέτες (doublets) ενώ τα δεξιόστροφα σωματίδια αναπαριστώνται από singlets. Τέλος, τα νεutrino εμφανίζονται μόνο αριστερόστροφα. Επομένως, συνοπτικά το σύνολο των φερμιονίων περιγράφεται ως εξής:

$$\begin{pmatrix} u \\ d \end{pmatrix}_L, \begin{pmatrix} c \\ s \end{pmatrix}_L, \begin{pmatrix} t \\ b \end{pmatrix}_L, \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L,$$

$$u_R, d_R, c_R, s_R, b_R, t_R, e_R, \mu_R, \tau_R$$

Στην ασθενή αλληλεπίδραση οι ιδιοκαταστάσεις της γεύσης δεν αντιστοιχούν στις ιδιοκαταστάσεις της μάζας των quarks. Τα φερμιόνια αλλάζουν γεύση μέσω των

μποζονίων W^\pm . Η αλλαγή γεύσης μεταξύ των κουαρκ περιγράφεται από τον πίνακα μίξης των Cabibbo-Kobayash-Maskawa (CKM), οποίος είναι ο εξής:

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

Η πιθανότητα ένα κουαρκ τύπου i να μετατραπεί σε ένα κουαρκ τύπου j , εκλύοντας ένα μποζόνιο W είναι ανάλογη της ποσότητας $|V_{ij}|$. Λόγω των καταστάσεων μίξης, παρατηρείται παραβίαση της ομοτιμίας αλλά και του φορτίου (CP-violation).

Το πρώτο βήμα για την ενοποίηση όλων των δυνάμεων που περιγράφονται από το ΚΠ, γίνεται ενώνοντας την QED με την ασθενή αλληλεπίδραση. Το φορτίο Q της ηλεκτρομαγνητικής και το ισοσπίν (η τρίτη προβολή αυτού) T_3 συνδυάζονται σχηματίζοντας το υπερφορτίο Y :

$$Y = 2(Q - T_3) \quad (1.12)$$

Η Λαγκρατζιανή πρέπει να είναι αναλλοίωτη κάτω από τους μετασχηματισμούς των ομάδων $SU(2)_L \times U(1)_Y$. Το αποτέλεσμα είναι τέσσερα άμαζα μποζόνια.

1.4.2 Ισχυρή Αλληλεπίδραση

Η ισχυρή αλληλεπίδραση ή ισχυρή δύναμη είναι η αλληλεπίδραση μεταξύ των κουαρκ και των γκλουονίων, η οποία περιγράφεται από τη θεωρία της κβαντικής χρωμοδυναμικής (quantum chromodynamics - QCD). Η ισχυρή δύναμη είναι η θεμελιώδης δύναμη η οποία διαδίδεται από τα γκλουόνια, και ασκείται πάνω στα κουαρκ, τα αντικουάρκ, καθώς και στα ίδια τα γκλουόνια για να σχηματίσουν τα πρωτόνια, τα νετρόνια και τα άλλα αδρόνια. Επίσης, σε μεγαλύτερη κλίμακα, είναι η δύναμη που κρατά τα πρωτόνια και τα νετρόνια μαζί για να σχηματίσουν τους πυρήνες των ατόμων. Δηλαδή, αν και η ισχυρή δύναμη δρα μόνο σε στοιχειώδη σωματίδια απευθείας, η δύναμη παρατηρείται μεταξύ των αδρονίων, ως η πυρηνική δύναμη. Τέλος, το φορτίο της ισχυρής δύναμης είναι το φορτίο χρώμα, το οποίο φέρει σημαντικές ιδιότητες στις οποίες θα γίνει αναφορά στη συνέχεια.

Όπως αναφέρθηκε, η κβαντική χρωμοδυναμική, η οποία είναι μέρος του Καθιερωμένου Προτύπου της σωματιδιακής φυσικής, είναι μια τυπική μη Αβελιανή θεωρία βαθμίδας βασισμένη σε μια τοπική ομάδα συμμετρίας $SU(3)$. Όλα τα σωματίδια σε αυτήν τη θεωρία αλληλεπιδρούν μεταξύ τους, μέσω της ισχυρής αλληλεπίδρασης. Η ένταση της αλληλεπίδρασης καθορίζεται από την ισχυρή σταθερά ζευξης (a_s). Η ένταση επομένως αλλάζει από το φορτίο χρώματος του σωματιδίου. Τα κουαρκ και τα γκλουόνια είναι τα μόνα στοιχειώδη σωματίδια που φέρουν μη μηδενικό φορτίο χρώματος, και κατά συνέπεια είναι τα μόνα που συμμετέχουν στις ισχυρές αλληλεπιδράσεις.

Υπάρχουν τρία είδη φορτίου χρώμα, κόκκινο (R), πράσινο (G) και μπλε (B) και σε κάθε φορτίο αντιστοιχεί ένα αντιφορτίο (\bar{R} , \bar{G} , \bar{B}). Επίσης, κάθε κουαρκ έχει υποχρεωτικά ένα από αυτά τα τρία φορτία χρώμα ενώ στα αντικουάρκ έχει ένα

αντιφορτίο. Η συνύπαρξη και των τριών φορτίων ή των τριών αντιφορτίων δίνει μη χρώμα, όπως και η συνύπαρξη φορτίου χρώμα με αντιφορτίο. Όλα τα αδρόνια είναι άχρωμα, καθώς στα βαρυόνια τα τρία κουάρκ έχουν διαφορετικό φορτίο χρώμα και στα μεσόνια το κουάρκ με το αντικουάρκ που τα συγκροτούν έχουν συμπληρωματικά φορτία χρώμα. Οι ισχυρές δυνάμεις μεταξύ δύο κουάρκς ή των γκλουονίων μπορούν να περιγραφούν με την ανταλλαγή γκλουονίων, όπως φαίνεται στην εικόνα ???. Κατά την ανταλλαγή επιτρέπεται να έχουμε αλλαγή του χρώματος των κουάρκς ή των γκλουονίων, αν και το συνολικό χρώμα διατηρείται σε κάθε κόμβο.

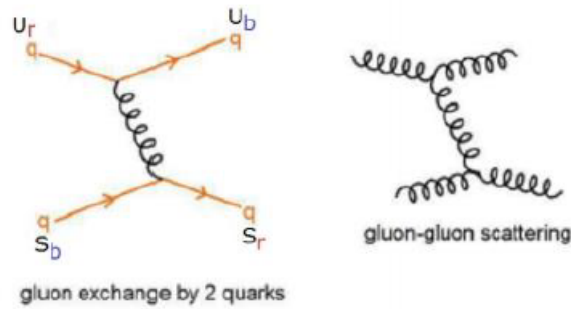


Figure 1.2: Ανταλλαγή γκλουονίων και αλλαγή του φορτίου χρώματος μεταξύ δύο κουάρκς και 2 γκλουονίων

Η αλλαγή χρώματος των κουάρκς διέπεται από την ισχυρή αλληλεπίδραση και γίνεται μέσω της ανταλλαγής των 8 γκλουονίων που αποτελούν την οχταπλή αναπαράσταση της $SU(3)$. Όσον αφορά την ισχυρή αλληλεπίδραση, η κατάσταση ενός σωματιδίου δίνεται από ένα διάνυσμα σε κάποιο διανυσματικό χώρο επί του οποίου τα στοιχεία της $SU(3)$ δρουν ως γραμμικοί (στην πραγματικότητα μοναδιαίοι) τανυστές. Το σωματίδιο επομένως, μετασχηματίζεται κάτω από κάποια αναπαράσταση της $SU(3)$. Για παράδειγμα, δεδομένου ότι τα στοιχεία της $SU(3)$ είναι 3×3 πίνακες, μπορούν να δράσουν σε διανύσματα στήλες με πολλαπλασιασμό πινάκων, αυτό δίνει μια τρισδιάστατη αναπαράσταση της $SU(3)$. Τα κουάρκς μετασχηματίζονται κάτω από αυτή την αναπαράσταση της $SU(3)$, και επειδή είναι τρισδιάστατα προκύπτει ότι τα κουάρκς φέρουν 3 χρώματα: κόκκινο, πράσινο και μπλε.

Τα γκλουόνια δεν έχουν μάζα, έχουν σπιν 1 και συνολικά είναι 8 και μπορεί να βρεθεί στις ακόλουθες καταστάσεις με τα εξής χρώματα

$$R\bar{G}, R\bar{B}, G\bar{R}, G\bar{B}, B\bar{R}, B\bar{G} \quad (1.13)$$

$$\frac{1}{\sqrt{2}}(R\bar{R} - G\bar{G}), \frac{1}{\sqrt{6}}(R\bar{R} + G\bar{G} - 2B\bar{B}) \quad (1.14)$$

Όταν ανταλλάσσονται γκλουόνια από την ??, τα αντίστοιχα κουάρκς αλλάζουν χρώμα ενώ κατά την ανταλλαγή των ?? τα αντίστοιχα κουάρκς δεν αλλάζουν χρώμα. Επίσης υπάρχει και ο τελευταίος συνδυασμός που αποτελεί φορτίο χρώμα γκλουονίου, είναι

ορθογώνιος σε κάθε έναν από τις 8 καταστάσεις $??, ??$

$$\frac{1}{\sqrt{3}}(R\bar{R} + G\bar{G} + B\bar{B}) \quad (1.15)$$

Η κατάσταση αυτή αποτελεί color single κατάσταση και δεν φέρει χρώμα, επομένως δεν μπορεί να διαδοθεί μεταξύ δυο κουάρκς.

1.5 Συγκρούσεις πρωτονίων

Τα πρωτόνια αποτελούνται από δύο up κουάρκς (u, \bar{u}) και ένα down κουάρκ (d), τα οποία ονομάζονται κουάρκ σθένους. Επίσης, γκλουόνια συνεχώς εκπέμπονται και απορροφώνται μέσα στο πρωτόνιο το οποίο μπορεί να διαχωριστεί σε ζεύγη $q\bar{q}$, τα κουάρκ 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 φέρουν χρώμα εκπέμπουν ακτινοβολία γκλουονίων, που οδηγεί στο σχηματισμό νέων χρωματικά ουδέτερων αδρονίων που συνθέτουν το γεγονός. Τα γκλουόνια που έχουν χρώμα και συμμετέχουν στην παραπάνω σκέδαση εκπέμπουν επίσης ακτινοβολία πριν και μετά την αλληλεπίδραση.

Στον LHC έχουμε συγκρούσεις δεσμών πρωτονίων με σύγκρουση δέσμης ανά 25 ns. Η ενέργεια που είναι διαθέσιμη για την παραγωγή σωματιδίων και την κινητική ενέργεια που ενδέχεται να αποκτήσουν, είναι η ενέργεια κέντρου μάζας ίση με $\sqrt{s} = 13TeV$. Η τιμή του s , καθορίζεται από την ορμή των συγκρουόμενων πρωτονίων στο κέντρο μάζας του συστήματος (CM):

$$s = (p_1 + p_2)^2 = p_1^2 + p_2^2 + 2p_1p_2 = (E_1^2 - \vec{p}_1^2) + (E_2^2 - \vec{p}_2^2) + 2(E_1E_2 - \vec{p}_1\vec{p}_2) \quad (1.16)$$

όπου p_i είναι οι τετραορμές των δύο συγκρουόμενων πρωτονίων, E_i οι ενέργειές τους και \vec{p}_i το διάνυσμα της ορμής στον τρισδιάστο χώρο. Επειδή οι δύο δέσμες πρωτονίων έχουν την ίδια ενέργεια και αντίθετη φορά, ισχύει: $E_1 = E_2 = E_{proton}$ και $\vec{p}_1 = -\vec{p}_2$. Επομένως, η ενέργεια κέντρου μάζας γίνεται: $\sqrt{s} = 2E_{proton}$.

Τα πρωτόνια είναι σύνθετα σωματίδια, η ακριβής διασπορά περιλαμβάνει κουάρκς και γκλουόνια, δηλαδή παρτόνια, που περιλαμβάνουν μόνο ένα πολύ μικρό ποσοστό της αρχικής ορμής της δέσμης των πρωτονίων x_{proton} . Η ενέργεια κέντρου μάζας της διαδικασίας hard scattering $\sqrt{s'}$ αποτελεί μόνο ένα μικρό ποσοστό από τα 13TeV. Αν υποθέσει κανείς ότι οι μάζες των παρτονίων είναι μικρές σε σχέση με την ορμή τους ως σχετικιστικά σωματίδια, η ενέργεια κέντρου μάζας της διαδικασίας αυτής είναι:

$$\sqrt{s'} = \sqrt{x_1x_2}\sqrt{s} \quad (1.17)$$

όπου x_1, x_2 προσδιορίζουν το ποσοστό της ορμής που έχουν τα συγκρουόμενα παρτόνια. Επειδή αυτό το ποσοστό δεν είναι γνωστό, η ενέργεια κέντρου μάζας στη σκληρή διασπορά δεν είναι επίσης γνωστή. Επομένως για να κάνει κανείς ανάλυση δεδομένων, πρέπει να χρησιμοποιήσει μεταβλητές, οι οποίες δεν εξαρτώνται από την αρχική ορμή στην κατεύθυνση της διαδρομής των πρωτονίων. Στους ανιχνευτές συγκρουόμενων αδρονίων η χρήση μια συνάρτησης πυκνότητας παρτονίων είναι πολύ σύνηθες. Οι συναρτήσεις αυτές μας δίνουν την πιθανότητα να βρεθεί ένα παρτόνιο σε ένα συγκεκριμένο διάστημα ορμής σε ένα πρωτόνιο.

Κεφάλαιο 2

Ο Ανιχνευτής Compact Muon Solenoid (CMS)

Στο κεφάλαιο αυτό γίνεται μια σύντομη αναφορά στο μεγάλο επιταχυντή αδρονίων Large Hardon Collider (LHC) και στη συνέχεια μια περιγραφή στον ανιχνευτή Compact Muon Solenoid (CMS), ο οποίος είναι ένας ανιχνευτής γενικού ενδιαφέροντος και αποτελεί έναν από τους 4 ανιχνευτές που βρίσκονται στην περίμετρο του LHC. Τέλος, θα γίνει μια σύντομη αναφορά στα επιμέρους επίπεδα του ανιχνευτή CMS, κάθε ένα από τα οποία αποτελεί ένα επιμέρους ανιχνευτικό σύστημα (subdetectors). Περιγράφεται ο ρόλος τους στο γενικότερο πείραμα αλλά και ο τρόπος λειτουργίας τους.

2.1 Ο μεγάλος επιταχυντής αδρονίων LHC

Ο μεγάλος ανιχνευτής αδρονίων αποτελεί τον ισχυρότερο επιταχυντή στοιχειωδών σωματιδίων, ο οποίος μετρά 10 χρόνια λειτουργίας στο ευρωπαϊκό κέντρο έρευνας στην πυρηνική φυσική (CERN). Είναι ένας κυκλικός επιταχυντής σωματιδίων με περίμετρο 27 km, οποίος βρίσκεται 100 μέτρα κάτω από την επιφάνεια της γης στα σύνορα Ελβετίας και Γαλλίας.

Ο LHC κατασκευάστηκε στο τούνελ όπου προϋπήρχε και λειτουργούσε ο μεγάλος επιταχυντής ηλεκτρονίων-ποζιτρονίων (Large Electron-Positron, LEP). Κύριος σκοπός του LHC είναι η μελέτη της δομής και της συμπεριφοράς των στοιχειωδών σωματιδίων καθώς και των μεταξύ τους αλληλεπιδράσεων. Αυτό επιτυγχάνεται με την επιτάχυνση 2 δεσμών πρωτονίων που κινούνται σε αντίθετες κατευθύνσεις με σκοπό την συγκρούσή τους σε 4 προκαθορισμένα σημεία, στα οποία είναι τοποθετημένοι οι ανιχνευτές των στοιχειωδών σωματιδίων. Βάσει των προδιαγραφών του, ο LHC είναι ικανός να επιταχύνει ταυτόχρονα 2 δέσμες πρωτονίων προσφέροντας 14TeV στο κέντρο μάζας. Αυτή η λειτουργία του ανιχνευτή τον καθιστά τον ισχυρότερο επιταχυντή που έχει κατασκευαστεί μέχρι στιγμής.



Figure 2.1: Αεροφωτογραφία του ευρωπαϊκού κέντρου έρευνας στην φυσική στοιχειωδών σωματιδίων (CERN). Στη φωτογραφία απεικονίζονται σχηματικά τα τούνελ στα οποία επιταχύνονται οι δέσμες καθώς και τα τέσσερα σημεία στα οποία βρίσκονται οι ανιχνευτές. Όλες οι διατάξεις βρίσκονται 100 μέτρα κάτω από την επιφάνεια της γης.

2.1.1 Επιταχυντικές Διατάξεις του LHC

Το επιταχυντικό σύστημα του LHC αποτελείται από ένα σύνολο υπεραγωγίων μαγνητών, οι οποίοι χρησιμοποιούνται για την επιτάχυνση και καμπύλωση των δεσμών καθώς και την εστίαση αυτών με σκοπό τη σύγκρουσή τους στα προκαθορισμένα σημεία πάνω στην περίμετρο του επιταχυντή. Για την επιτάχυνση της δέσμης και την εστίαση της χρησιμοποιούνται περίπου 1200 και 400 μαγνήτες αντίστοιχα. Κατά τη διάρκεια λειτουργίας του επιταχυντή οι μαγνήτες διατηρούνται σε θερμοκρασία 1.9K με χρήση υγρού ηλίου.

Για την επιτάχυνση της δέσμης στο CERN υπάρχει ένα σύνολο μικρότερων επιταχυντικών διατάξεων στις οποίες πραγματοποιείται το πρωταρχικό στάδιο επιτάχυνσης της δέσμης, πριν αυτή εισέλθει στον LHC, στον οποίο γίνεται το τελικό στάδιο της επιτάχυνσης των δεσμών.

Τα πρωτόνια προέρχονται από αέριο υδρογόνο, το οποίο καθώς εισέρχεται σε ισχυρό ηλεκτρικό πεδίο τα ηλεκτρόνια του πυρήνα απομακρύνονται. Το πρώτο στάδιο της επιτάχυνσης των πρωτονίων που προέκυψαν πραγματοποιείται στο γραμμικό επιταχυντή LINAC 2, ο οποίος τους αποδίδει ενέργεια 50MeV. Στη συνέχεια η δέσμη πρωτονίων εισέρχεται διαδοχικά στο Proton Synchrotron Booster (PSB) και στο Proton Synchrotron, όπου επιταχύνεται σε ενέργεια 26 GeV. Από εκεί οχηγείται στο Super Proton Synchrotron (SPS), όπου επιταχύνεται σε ενέργεια 450 GeV. Τέλος, η δέσμη εισέρχεται στον LHC όπου γίνεται το τελικό στάδιο της επιτάχυνσης. Η δέσμη του LHC αποτελείται από πακέτα πρωτονίων (bunches) κάθε ένα από τα οποία περιέχει 1.15×10^{11} πρωτόνια.

Οι δύο βασικές παράμετροι που καθορίζουν τη λειτουργία του LHC είναι οι ενέργειες των δύο συγκρουόμενων δεσμών και η φωτεινότητά τους (Luminosity L).

Ορυθμός των γεγονότων (συγκρούσεων) που προκύπτουν από τη λειτουργία δίνεται από τη σχέση:

$$R = \sigma L \quad (2.1)$$

όπου σ είναι η ενεργός διατομή της φυσικής διεργασίας και L η φωτεινότητα. Για να διατηρηθεί η αποδοτικότητα του πειράματος για δεδομένη ενέργεια E , η φωτεινότητα του ανιχνευτή πρέπει να αυξάνεται ανάλογα με το τετράγωνο της ενέργειας (E^2), εφόσον η ενεργός διατομή είναι αντιστρόφως ανάλογη της ποσότητας αυτής ($\sigma \propto \frac{1}{E^2}$). Η luminosity ενός επιταχυντή ο οποίος συγκρούει 2 πακέτα σωματιδίων κάθε ένα εκ των οποίων αποτελείται από n_1 και n_2 το πλήθος σωματίδια αντίστοιχα με συχνότητα f , δίνεται από τη σχέση:

$$L = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} \quad (2.2)$$

όπου σ_x, σ_y είναι τα εγκάρσια προφίλ της δέσμης.

Η προβλεπόμενη luminosity του LHC είναι $L = 10^{34} \text{cm}^{-2} \text{s}^{-1}$. Αξίζει να σημειωθεί πως για την επίτευξη της τόσο μεγάλης luminosity, θα πρέπει η συχνότητα συγκρούσεων να είναι πολύ μεγάλη, καθώς επίσης και ο αριθμός των πρωτονίων που περιέχονται σε κάθε πακέτο. Επομένως, για το λόγο αυτό κάθε δέσμη πρωτονίων αποτελείται από 2080 πακέτα σε πολύ κοντινή απόσταση μεταξύ τους στα οποία περιέχονται περίπου 10^{11} πρωτόνια και συγκρούονται κάθε 25ns, δηλαδή ο LHC έχει συχνότητα συγκρούσεων ίση με 40MHz.

2.1.2 Ανιχνευτικές Διατάξεις του LHC

Σκοπός της επιτάχυνσης των δεσμών πρωτονίων σε αντίθετες κατευθύνσεις είναι οι συγκρούσεις αυτών σε 4 προκαθορισμένα σημεία, όπου βρίσκονται οι 4 ανιχνευτές σωματιδίων πάνω στην περιμέτρο του LHC. Οι ανιχνευτές είναι οι εξής:

- **A Toroidal LHC ApparatuS (ATLAS)** point 1
- **Compact Muon Solenoid (CMS)** point 5
- **LHC beauty (LHCb)** point 2
- **A Large Ion Collider Experiment (ALICE)** point 8

Ως "point" καθορίζονται τα σημεία πρόσβασης του LHC, οποίος έχει συνολικά 8 σημεία πρόσβασης και στα 4 από αυτά είναι τοποθετημένοι οι ανιχνευτές.

Οι ανιχνευτές CMS και ATLAS αποτελούν πειράματα γενικού ενδιαφέροντος στη φυσική υψηλών ενεργειών, ενώ αντίθετα οι ανιχνευτές LHCb και ALICE είναι σχεδιασμένοι και κατασκευασμένοι για στοχευμένη λειτουργία. Το πείραμα LHCb έχει ως σκοπό την μελέτη του b quark (beauty) και την παραβίαση της συμμετρίας CP (CP violation). Από την άλλη το πείραμα ALICE στοχεύει στη μελέτη συγκρούσεων βαρέων ιόντων αλλά και συγκρούσεων βαρέων ιόντων και πρωτονίων. Εκτός από τα 4 αυτά πειράματα υπάρχουν και άλλα μικρότερα, όπως το TOTEM και το CASTOR, τα οποία μοιράζονται το σημείο σύγκρουσης με το CMS. Τέλος στην εικόνα ?? φαίνονται

και άλλα μικρότερα πειράματα τα οποία τροφοδοτούνται με δέσμες από κάποιον από τους μικρότερους επιταχυντές του συνολικού συμπλέγματος επιταχυντών του CERN. Αναφορικά ορισμένα από αυτά είναι το **ISOLDE** (Isotope Separator On Line Device), το **nTOF** (neutron Time Of Flight) το **ELENA** (Extra Low ENergy Antiproton) και άλλα.

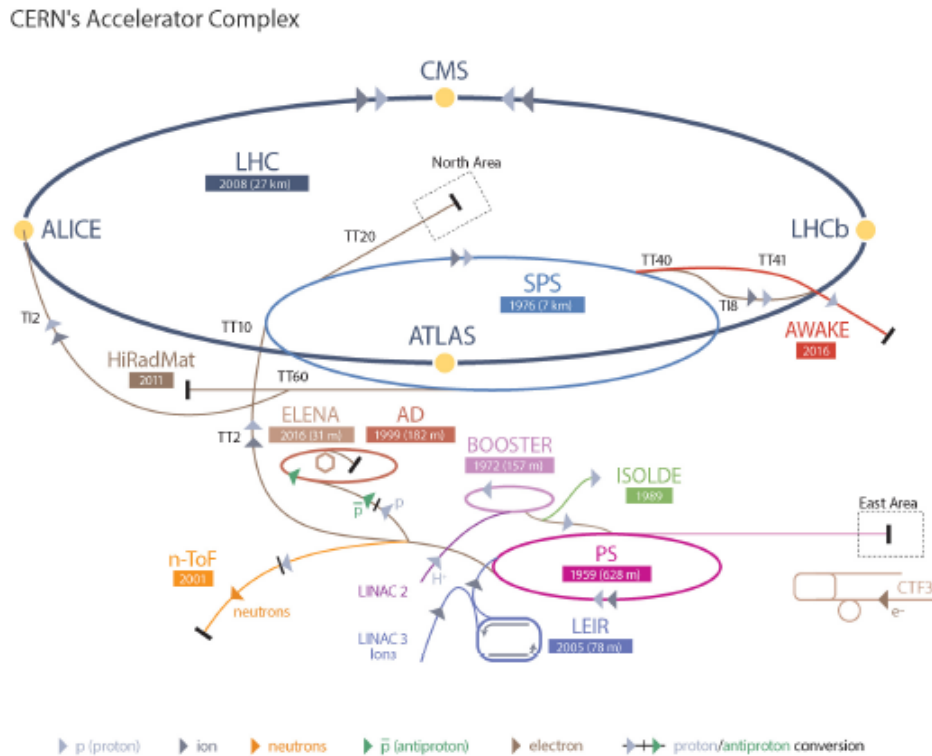


Figure 2.2: Το σύμπλεγμα επιταχυντών του LHC. Οι δέσμες πρωτονίων εισέρχονται αρχικά στον γραμμικό επιταχυντή LINAC 2, στη συνέχεια περνούν διαδοχικά από τους επιταχυντές Proton Synchrotron Booster (BSB), Proton Synchrotron και Super Proton Synchrotron (SPS), πριν καταλήξουν στον LHC για το τελικό στάδιο επιτάχυνσης. Εκτός από τον LHC, με δέσμη από τα προηγούμενα στάδια επιτάχυνσης τροφοδοτούνται μικρότερα πειράματα στο CERN όπως το n-TOF, ISOLDE, ELENA κ.α.

2.2 Compact Muon Solenoid (CMS)

Ο ανιχνευτής CMS είναι ένας ανιχνευτής γενικού ενδιαφέροντος, ο οποίος είναι σχεδιασμένος και κατασκευασμένος ώστε να ανιχνεύει όλα τα στοιχειώδη σωματίδια. Αποτελείται από ένα σύνολο επιμέρους ανιχνευτικών διατάξεων (subdetectors) οι οποίοι τοποθετημένοι σε διάφορα επίπεδα, σχηματίζουν έναν κύλινδρο με κύριο άξονα τον άξονα της δέσμης, σαν ένα "βαρέλι" (barrel). Εφόσον στόχος του CMS είναι η ανίχνευση όλων των στοιχειωδών σωματιδίων που παράγονται κατά τις συγκρούσεις, πρέπει να είναι ερμητικά κλειστός. Για το λόγο αυτό στα δυο άκρα του είναι τοποθετημένοι δυο δίσκοι κάθετοι στον άξονα της δέσμης, οι οποίοι κλείνουν τον ανιχνευτή (endcaps).

Η λέξη compact στο όνομα του ανιχνευτή προέρχεται από τις διαστάσεις αυτού. Έχει μήκος 28.7m και διάμετρο 15m, ενώ ζυγίζει περίπου 14.000 τόνους. Επομένως, είναι αρκετά μικρός (compact) για το βάρος του και την ποσότητα υλικού που περιέχει.

Ένα από τα σωματίδια με το μεγαλύτερο ενδιαφέρον ως προς την ανίχνευση τους είναι τα μίονια. Αυτό οφείλεται στο γεγονός ότι η ανίχνευση τους αποτελεί μια αρκετά ξεκάθαρη ένδειξη για την πραγματοποίηση διεργασιών οι οποίες είναι υψίστης σημασίας για την λειτουργία του πειράματος. Για το λόγο αυτό ο ανιχνευτής CMS είναι κατασκευασμένος ώστε να ανιχνεύει με μεγάλη ακρίβεια μίονια και από εκεί προκύπτει η λέξη muon στο όνομα του ανιχνευτή.

Τέλος πολύ σημαντικό μέρος του ανιχνευτή, είναι ο μαγνήτης ο οποίος δίνει τη δυνατότητα για ακριβή μέτρηση της ορμής των σωματιδίων που παράγονται. Στο CMS υπάρχει ένας σωληνοειδής μαγνήτης, εξού και το solenoid, ο οποίος παράγει μαγνητικό πεδίο ύψους 4T παράλληλα στον άξονα της δέσμης με αποτέλεσμα να καμπυλώνει τις τροχιές των σωματιδίων κάθετα στον άξονα αυτής.

Μπορεί να αποδειχθεί πως η ανακατασκευής της ορμής ενός μιονίου από έναν ανιχνευτή, εξαρτάται άμεσα από την ισχύ του μαγνητικού αλλά και το μήκος αυτού, σύμφωνα με τη σχέση:

$$\frac{\sigma_{P_T}}{P_T} = \frac{\sigma_s}{s} = \sigma_s \frac{8P_T}{0.3BL^2} \quad (2.3)$$

όπου P_T είναι η εγκάρσια συνιστώσα της ορμής του σωματιδίου, s η καμπύλωση της τροχιάς του από το μαγνητικό πεδίο έντασης B , που παράγει ο μαγνήτης του ανιχνευτή και L το μήκος του ανιχνευτή. Ο ανιχνευτής CMS είναι κατασκευασμένος έτσι ώστε να έχει μικρό μέγεθος και υψηλή ένταση μαγνητικού πεδίου, σε αντίθεση με τον ανιχνευτή ATLAS, ο οποίος έχει μεγαλύτερο μήκος και μικρότερης έντασης μαγνητικό πεδίο.

Όπως αναφέρθηκε ο ανιχνευτής CMS έχει κατασκευαστεί σε διάφορα επίπεδα και αποτελείται από ένα σύνολο επιμέρους ανιχνευτικών διατάξεων. Κάθε επίπεδο αποτελεί έναν υποανιχνευτή (subdetector). Κάθε subdetector είναι υπεύθυνος για την συλλογή ενός συγκεκριμένου κομματιού πληροφορίας. Όλα αυτά τα κομμάτια συλλέγονται και ο συνδυασμός τους δίνει τη δυνατότητα ταυτοποίησης των σωματιδίων

που δημιουργήθηκαν κατά τη σύγκρουση των πρωτονίων. Αναφορικά, οι επιμέρους αυτές διατάξεις είναι οι εξής:

- **Tracker** Ένα υψηλής ποιότητας σύστημα ανίχνευσης τροχιών, το οποίο δίνει πληροφορίες για το είδος και την ορμή των σωματιδίων που προέρχονται από τις συγκρούσεις των δεσμών πρωτονίων .
- **Electromagnetic Calorimeter, ECAL** Ένα ηλεκτρομαγνητικό θερμιδόμετρο υψηλής ανάλυσης για την ανίχνευση και μελέτη φωτονίων και ηλεκτρονίων
- **Hadronic Calorimeter, HCAL** Ένα αδρονικό καλορίμετρο για την ανίχνευση των ουδέτερων και φορτισμένων αδρονίων που δημιουργούνται από τις συγκρούσεις που πραγματοποιούνται στο κέντρο του ανιχνευτή.
- **Muon Detectors** Ένα υψηλής απόδοσης μιονικό σύστημα ικανό να ανιχνεύσει τα μόνια αλλά και να μετρήσει τα χαρακτηριστικά τους, όπως η ορμή τους κα.

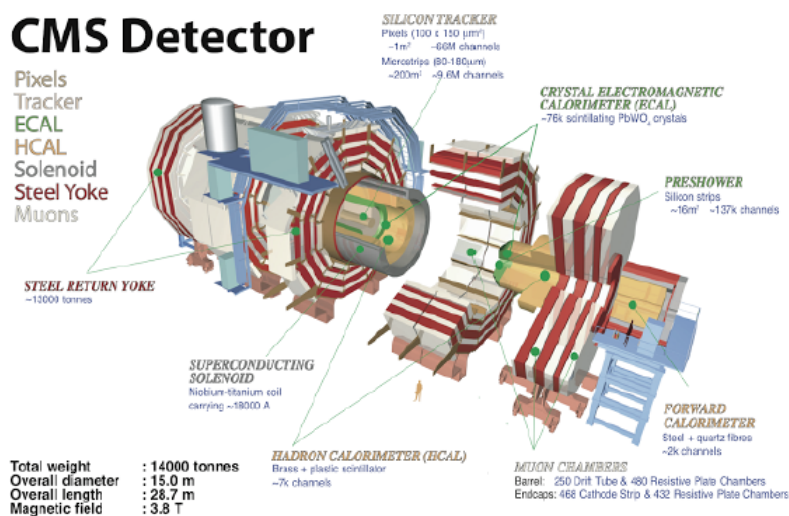


Figure 2.3: Ο ανιχνευτής Compact Muon Solenoid. Ο ανιχνευτής αποτελείται από 4 επιμέρους υποανιχνευτές, τον Tracker, το ηλεκτρομαγνητικό θερμιδόμετρο, το αδρονικό θερμιδόμετρο και τους ανιχνευτές μιονίων. Για τη λειτουργία του ανιχνευτή χρησιμοποιείται επίσης ένας σωληνοειδής μαγνήτης ο οποίος παράγει ένα μαγνητικό πεδίου έντασης 4 T.

Η διάταξη που ακολουθείται στο CMS είναι πως τα 3 ανιχνευτικά συστήματα Tracker, ECAL και HCAL βρίσκονται στο εσωτερικό του μαγνήτη ενώ το μιονικό σύστημα βρίσκεται έξω από τον μαγνήτη. Κάθε ένα από τα συστήματα αυτά, διαθέτει

ένα κομμάτι του στο κύριο μέρος του ανιχνευτή, στο barrel δηλαδή, αλλά και ένα στα καπάκια αυτού. Στις επόμενες ενότητες, θα γίνει ξεχωριστή αναφορά και περιγραφή του καθενός από τα ανιχνευτικά συστήματα του CMS.

Αξίζει να γίνει μια σύντομη αναφορά στο σύστημα συντεταγμένων που χρησιμοποιείται στα πειράματα του LHC. Για την περιγραφή του ανιχνευτή, μπορεί να χρησιμοποιηθεί ένα καρτεσιανό σύστημα συντεταγμένων σύμφωνα με το οποίο, με κέντρο αξόνων το σημείο της κύριας σύγκρουσης ο άξονας x δείχνει προς το κέντρο του δακτυλίου του LHC, ο άξονας y να δείχνει προς τα πάνω κάθετα στο άξονα της δέσμης, ενώ άξονας z είναι παράλληλος προς τον άξονα της δέσμης. Ωστόσο επειδή το σχήμα του ανιχνευτή είναι κυλινδρικό, η περιγραφή του γίνεται πιο εύκολη αν χρησιμοποιηθεί ένα άλλο σύστημα αναφοράς το οποίο χαρακτηρίζεται από την τριπλέτα (r, φ, η) . Όπου r είναι η απόσταση από τον άξονα z , φ είναι η αζιμουθιακή γωνία από τον άξονα x και η είναι μια ποσότητα που ονομάζεται pseudo-rapidity, η οποία δίνεται από τη σχέση

$$\eta = -\ln \tan \frac{\theta}{2} \quad (2.4)$$

όπου θ είναι η γωνία από το θετικό ημιάξονα z . Στην εικόνα () φαίνεται η pseudo-rapidity για διάφορες τιμές της γωνίας θ .

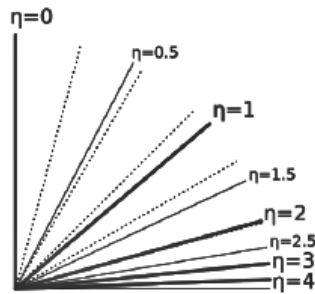


Figure 2.4: Η pseudo-rapidity ($\eta = -\ln \tan \frac{\theta}{2}$) ως συνάρτηση της γωνίας θ . Η γωνία θ είναι η γωνία από τον θετικό ημιάξονα z . Ως άξονας z έχει οριστεί ο άξονας της δέσμης του LHC

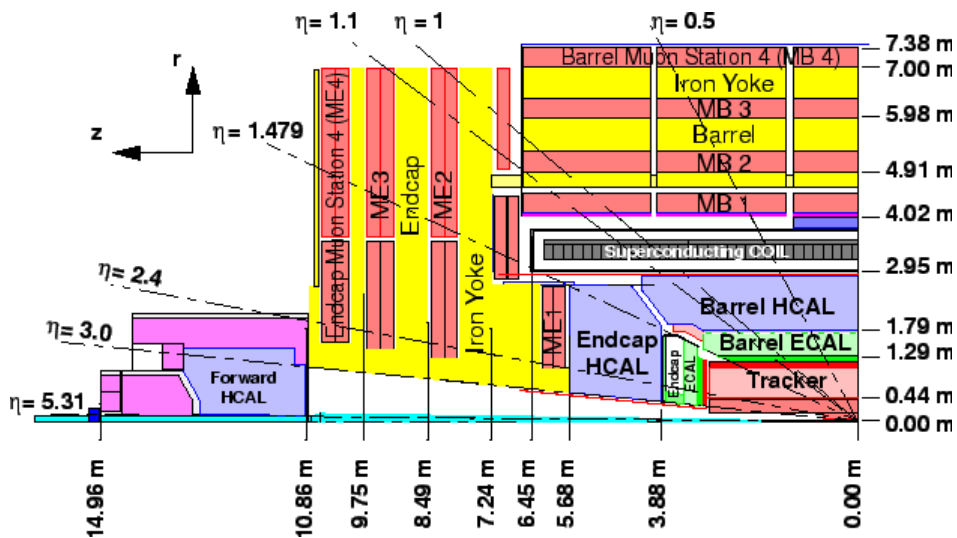


Figure 2.5: Διάφορες τιμές της pseudo-rapidity ως προς τον ανιχνευτή CMS

2.2.1 Ο ανιχνευτής τροχιών (Tracker)

Ο ανιχνευτής τροχιών του CMS είναι το πρώτο ανιχνευτικό σύστημα που διαθέτει το πείραμα και βρίσκεται πιο κοντά στο σημείο αλληλεπίδρασης των 2 δεσμών. Ο ρόλος του είναι η ανίχνευση της ορμής ενός σωματιδίου και της διαδρομής που αυτό ακολούθησε μέσα στο μαγνητικό πεδίο. Όσο πιο κυρτή είναι τροχιά του, τόσο μικρότερη είναι η ορμή του. Είναι κατασκευασμένος με σκοπό να μετρήσει με μεγάλη ακρίβεια την ορμή μιονίων, ηλεκτρονίων αλλά και φορτισμένων αδρονίων. Επίσης πέρα από την ανακατασκευή της τροχιάς των σωματιδίων, έχει τη δυνατότητα ανακατασκευής των κορυφών (vertices) στις οποίες γίνεται η παραγωγή δευτερευόντων σωματιδίων. Ο ανιχνευτής αυτός, έχει μήκος 5.4 m και καλύπτει μια περιοχή $r < 1.2\text{m}$ και $|\eta| < 2.5$ και είναι εξ'ολοκλήρου κατασκευασμένος από πυρίτιο.

Λόγω της μεγάλης φωτεινότητας του επιταχυντή και της θέσης του, ο ανιχνευτής τροχιών διαθέτει πολύ καλή χωρική διακριτική ικανότητα αλλά και μεγάλη αντοχή στην ακτινοβολία. Επίσης πρέπει να έχει πολύ καλό χρόνο απόκρισης για να καταγράφει όσο το δυνατόν περισσότερα γεγονότα.

Για τους ανιχνευτές τροχιές, είναι πολύ σημαντικό να είναι κατασκευασμένοι με τέτοιο τρόπο ώστε να περιέχουν τη μικρότερη δυνατή ποσότητα υλικού, καθώς τα σωματίδια αλληλεπιδρούν με το υλικό του ανιχνευτή και χάνουν ενέργεια. Αν ο ανιχνευτής περιέχει μεγάλη ποσότητα υλικού, είναι πολύ πιθανό τα σωματίδια να εναποθέσουν όλη τους την ενέργεια στον ανιχνευτή, να σταματήσουν σε αυτόν και να μην προχωρήσουν στα υπόλοιπα ανιχνευτικά συστήματα χάνοντας έτσι πολύτιμη πληροφορία για την ανίχνευση τους. Για την κατασκευή του ανιχνευτή τροχιών στο CMS χρησιμοποιούνται δυο είδη ανιχνευτών τροχιές, οι Silicon Pixel και οι Silicon Microstrip. Αποτελείται από 1440 silicon pixels detectors και 15148 ανιχνευτικές μονάδες από silicon stripes

Οι ανιχνευτές Silicon pixel είναι ανιχνευτές στερεάς κατάστασης και έχουν μέγεθος ενός κουτιού από παπούτσια, Το κάθε κουτί περιέχει 65 εκατομύρια pixels για την ανίχνευση της διαδρομής των σωματιδίων που προκύπτουν από τις συγκρούσεις πρωτονίων στον επιταχυντή. Οι ανιχνευτές Silicon pixels μπορούν να ανιχνεύσουν $\frac{10 \times 10^{16} \text{ particles}}{\text{cd}^2 \text{ s}}$. Στη συνέχεια, τα σωματίδια περνούν μέσα από 10 στρώματα από ανιχνευτές Silicon Strips, οι οποίοι έχουν σχεδόν την ίδια λειτουργία με τους pixel detectors.

Στον ανιχνευτή CMS χωρικά ο Tracker χωρίζεται στα εξής τμήματα, pixel detector, στους ανιχνευτές που βρίσκονται στο εσωτερικό του κυλίνδρου (Tracker Inner Barrel, TIB), σε αυτούς που βρίσκονται στο εξωτερικό στρώμα του κυλίνδρου (Tracker Outer Barrel TOB), στους εσωτερικούς δίσκους (Tracker Inner Disks, TID) και στους ανιχνευτές που βρίσκονται στα endcaps (Tracker Endcaps TEC). Κάθε τμήμα πυριτίου είναι σχεδιασμένο διαφορετικά για τη θέση που έχει στο σύνολο του Tracker. Στην εικόνα () φαίνονται τα διάφορα τμήματα του ανιχνευτή τροχιών καθώς και η περιοχή που αυτός καλύπτει.

Τέλος, η φυσική διαδικασία στην οποία βασίζεται η λειτουργία των ανιχνευτών πυρητίου είναι η ακόλουθη. Όταν ένα φορτισμένο σωματίδιο διαπερνά τον ανιχνευτή, αποβάλλονται ηλεκτρόνια από τα άτομα του πυριτίου δημιουργώντας ζεύγη e^- -οπών. Τα ηλεκτρόνια που αποβάλλονται παράγουν έναν παλμό, ο οποίος στη συνέχεια ενισχύεται από chips APV25 (Analogue Pipeline Voltage Chip) δίνονται ένα hit κάθε φορά που ένα σωματίδιο περνάει. Έτσι παρέχεται η δυνατότητα ανακατασκευής της τροχιάς του σωματιδίου. Επίσης, για την ομαλή λειτουργία του ανιχνευτή τροχιών και την ελαχιστοποίηση της διαταραχής του πυρητίου λόγω της υψηλής ακτινοβολίας που εκπέμπεται από τις συγκρούσεις, διατηρείται σαφής θερμοκρασία -20°C με σκοπό την αποφυγή ή την αποτροπή διαιώνισης τυχόν βλαβών που μπορεί να προκληθούν.

2.2.2 Θερμιδόμετρα

Βασικός σκοπός των θερμιδόμετρων είναι η μέτρηση της συνολικής ενέργειας των σωματιδίων που διαδίδονται στο υλικό του ανιχνευτή. Η αρχή λειτουργίας τους βασίζεται στο ότι τα σωματίδια αλληλεπιδρούν με το υλικό του ανιχνευτή με αποτέλεσμα αφήνουν την ενέργεια τους σε αυτόν, δημιουργώντας καταιγισμούς. Στη συνέχεια, η ενέργεια που εναποτέθηκε στον ανιχνευτή μπορεί να μετρηθεί και έτσι να προσδιοριστεί η ενέργεια του αρχικού σωματιδίου. Για να προσδιοριστεί όμως η ολική ενέργεια του σωματιδίου, πρέπει αυτό να εναποθέσει όλη του την ενέργεια και να σταματήσει μέσα στον ανιχνευτή. Για το λόγο αυτό, στα πειράματα φυσικής χρησιμοποιούνται δύο ειδών θερμιδόμετρα, ένα ηλεκτρομαγνητικό και ένα αδρονικό, καθώς η δημιουργία αλλά και η ανίχνευση των δύο ειδών καταιγισμών είναι διαφορετικές διαδικασίες.

Γενικά τα θερμιδόμετρα χωρίζονται σε δύο είδη ανάλογα με τα είδη των υλικών που χρησιμοποιούνται σαν απορροφητές και σαν ενεργό υλικό του ανιχνευτή, τα **ομοιογενή** και τα **δειγματοληπτικά**. Τα πρώτα, αποτελούνται από ένα και μόνο

υλικό το οποίο είναι υπεύθυνο τόσο για τη δημιουργία καταίγισμών (για την απορρόφηση του σωματιδίου), όσο και για την μετατροπή της ενέργειας που εναποτέθηκε σε μετρήσιμο σήμα. Οι ανιχνευτές αυτού του είδους έχουν πολύ καλή διακριτική ικανότητα. Ένα παράδειγμα ομοιογενούς θερμιδόμετρου είναι αυτό του ηλεκτρομαγνητικού θερμιδόμετρου του CMS το οποίο αποτελείται από κρυστάλλους μολύβδου βολφραμίου ($PbWO_4$). Από την άλλη, στα δειγματοληπτικά θερμιδόμετρα η απορρόφηση και ανίχνευση, γίνεται από διαφορετικά υλικά. Το υλικό-απορροφητής είναι υπεύθυνο για την δημιουργία των καταίγισμών και επομένως χρησιμοποιούνται πολύ πυκνά υλικά όπως μόλυβδος και σίδηρος. Το ενεργό υλικό του ανιχνευτή είναι υπεύθυνο για την παραγωγή μετρήσιμου σήματος. Σαν ενεργό υλικό χρησιμοποιούνται οργανικοί ή ανόργανοι σπινθηριστές, πυρίτιο και πολλά άλλα. Η επιλογή αυτή εξαρτάται από τα τεχνικά χαρακτηριστικά που πρέπει να έχει ο ανιχνευτής όπως διακριτική ικανότητα, αντοχή στην ακτινοβολία και άλλα.

Η διακριτική ικανότητα των θερμιδομέτρων αποτελεί μια πολύ σημαντική ιδιότητα αυτών. Η ιδιότητα αυτή ορίζεται ως η ικανότητα του ανιχνευτή να διαχωρίσει δυο διαφορετικές αλλά πολύ κοντά μεταξύ τους ενέργειες. Τα θερμιδόμετρα ανιχνεύουν σωματίδια βάσει των αλληλεπιδράσεων αυτού με το υλικό του ανιχνευτή. Λόγω της στατιστικής φύσεως των αλληλεπιδράσεων αυτών, το αποτέλεσμα της μέτρησης της ενέργειας είναι μια κατανομή Gauss αντί για μια συνάρτηση-δ όπως θα περίμενε κανείς. Το εύρος της καμπύλης αυτής καθορίζει τη διακριτική ικανότητα του ανιχνευτή. Η πραγματική διακριτική του ικανότητα δίνεται από το εύρος της καμπύλης στο μέσο του μέγιστου της (Full Width at Half Maximum (FWHM)) το οποίο για μια κατανομή Gauss είναι ίσο με $FWHM = 2.35\sigma$. Πρακτικά, αυτό σημαίνει ότι ενέργειες που απέχουν μεταξύ τους λιγότερο από αυτό τον αριθμό δεν μπορούν να ταυτοποιηθούν σαν διαφορετικές από τον ανιχνευτή. Η ενέργεια που μετράται από τον ανιχνευτή εξαρτάται από τον αριθμό των σωματιδίων (N), τα οποία παράχθηκαν κατά των καταίγισμό. Ο αριθμός αυτός είναι ανάλογος της ενέργειας του εισερχόμενου σωματιδίου, καθώς ο αριθμός των παραγόμενων σωματιδίων ακολουθεί κατανομή Poisson. Επομένως, αποδεικνύεται πως η διακριτική ικανότητα του ανιχνευτή, $\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{E}}$

2.2.2.1 Το Ηλεκτρομαγνητικό Καλορίμετρο (ECAL)

Το δεύτερο σύστημα ανιχνευτών του CMS είναι το ηλεκτρομαγνητικό θερμιδόμετρο. Σκοπός του είναι η μέτρηση της ενέργειας των ηλεκτρονίων και των φωτονίων που παράγονται από τις συγκρούσεις. Για τη μέτρηση της ολικής ενέργειας ενός φορτισμένου σωματιδίου ή ενός φωτονίου, όπως αναφέρθηκε ήδη πρέπει το σωματίδιο να σταματήσει πλήρως στο υλικό του ανιχνευτή, να απορροφηθεί πλήρως. Η έναρξη της διαδικασίας σηματοδοτείται όταν ένα φορτισμένο σωματίδιο ή ένα φωτόνιο εισέλθει στον ανιχνευτή και ξεκινήσει η διαδικασία του ηλεκτρομαγνητικού καταίγισμού.

Ο ηλεκτρομαγνητικός καταίγισμός είναι μια έννοια ύψιστης σημασίας για τη θερμιδομετρία. Τα σωματίδια με μεγάλη ενέργεια τα οποία διαδίδονται στον ανιχνευτή, χάνουν ενέργεια κυρίως μέσω του μηχανισμού εκπομπής ακτινοβολίας

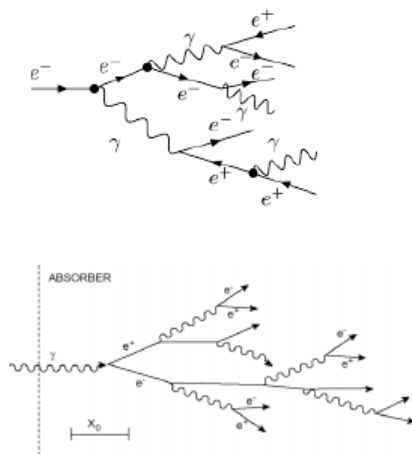


Figure 2.6: Απλή αναπαράσταση ενός ηλεκτρομαγνητικού καταιγισμού ο οποίος ξεκινάει με την εκπομπή ακτινοβολίας Bremsstrahlung από ένα ηλεκτρόνιο και ένα φωτόνιο.

πέδησης (ακτινοβολία Brehmstrahlung), εκπέμποντας δηλαδή υψηλοενεργειακά φωτόνια. Τα φωτόνια αυτά στη συνέχεια μέσω του μηχανισμού της δίδυμης γένεσης εκπέμπουν ζεύγη ηλεκτρονίων-ποζιτρονίων. Η διαδικασία αυτή διεξάγεται διαδοχικά με αποτέλεσμα τη δημιουργία μιας "χινοστοιβάδας" σωματιδίων, που ονομάζεται ηλεκτρομαγνητικός καταιγισμός. Η όλη διαδικασία επαναλαμβάνεται μέχρις ότου τα φωτόνια να μην έχουν αρκετή ενέργεια να πραγματοποιήσουν δίδυμη γένεση και τα φορτισμένα σωματίδια να χάνουν ενέργεια μόνο μέσω της διαδικασίας του ιονισμού. Δηλαδή, τα φορτισμένα σωματίδια έχουν φτάσει στην κρίσιμη ενέργεια E_c , η οποία είναι η ενέργεια, στην οποία οι απώλειες ενέργειας λόγω ακτινοβολίας Brehmstrahlung εξισώνονται με τις απώλειες λόγω ιονισμού. Η κρίσιμη ενέργεια εξαρτάται κυρίως από τον ατομικό αριθμό του υλικού του ανιχνευτή (Z).

Η ανίχνευση της ενέργειας των φορτισμένων σωματιδίων από τον ανιχνευτή βασίζεται στο φαινόμενο του σπινθηρισμού. Όταν ένα άτομο διεγείρεται, απορροφά δηλαδή ενέργεια, ένα ηλεκτρόνιο πηγαίνει σε υψηλότερη ενεργειακή στοιβάδα και όταν αποδιεγερθεί, επιστρέφοντας στην αρχική ενεργειακή κατάσταση, απελευθερώνει ενέργεια μέσω ενός φωτονίου. Επομένως, όταν ένα σωματίδιο που αλληλεπιδρά ηλεκτρομαγνητικά περάσει μέσα από το υλικό παράγει φως το οποίο ανιχνεύεται και έτσι μπορεί να μετρηθεί η ενέργεια του σωματιδίου. Επομένως, όταν πραγματοποιείται ο ηλεκτρομαγνητικός καταιγισμός και τα άτομα στο υλικό του ανιχνευτή λαμβάνουν ενέργεια από τα σωματίδια που περνούν και διεγείρονται. Όταν αποδιεγείρονται εκπέμπουν φωτόνια μπλε φωτός. Τα φωτόνια αυτά που παράγονται από σπινθηρισμό συλλέγονται από μια συσκευή φωτοπολλαπλασιαστή (PMT). Από το ηλεκτρικό σήμα που παράγεται από την όλη διαδικασία υπολογίζεται η ενέργεια του εισερχόμενου φορτισμένου σωματιδίου ή φωτονίου στον ανιχνευτή

Στο ECAL του CMS χρησιμοποιούνται κρύσταλλοι μολύβδου-βολφραμίου ($PbWO_4$). Το υλικό αυτό επιλέχθηκε καθώς είναι πολύ πυκνό ($8.28g/cm^3$) και διαθέτει μικρό μήκος ακτινοβολίας¹ ($X_0 = 0.89cm$) και ακτίνα Moliere² ($R_m = 0.89cm$). Επομένως, λόγω της υψηλής πυκνότητας του υλικού παράγονται σπινθηρισμοί οι οποίοι οδηγούν σε γρήγορους, μικρούς και καλά ορισμένους ηλεκτρομαγνητικούς καταγισμούς σωματιδίων. Ένας επιπλέον λόγος επιλογής αυτών των κρυστάλλων, είναι πως η απόκριση τους είναι πολύ γρήγορη. Πιο συγκεκριμένα, η παραγωγή του 80% του συνολικού φωτός που παράγει το εισερχόμενο σωματίδιο γίνεται σε διάστημα 25ns, σε χρόνο δηλαδή συγκρίσιμο με τη συχνότητα που συγκρούονται οι δέσμες των πρωτονίων.

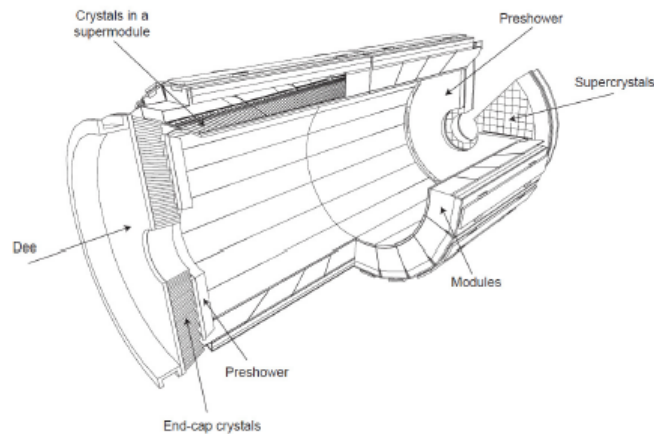


Figure 2.7: Το ηλεκτρομαγνητικό θερμιδόμετρο (ECAL) του ανιχνευτή CMS κατασκευασμένο από κρύσταλλους μολύβδου-βολφραμίου ($PbWO_4$). Το ECAL αποτελείται από το ECAL barrel (EB), το ECAL endcap (EE) και τον ανιχνευτή preshower.

Όπως όλοι οι υποανιχνευτές του CMS έτσι και το ECAL χωρίζεται σε 2 περιοχές. Το κομμάτι που βρίσκεται στο barrel (ECAL Barrel, EB) και αυτό που βρίσκεται στα endcaps (EE). Συνολικά το ECAL καλύπτει την περιοχή $1.2m < r < 1.8m$ και $|\eta| < 3$. Το EB καλύπτει την περιοχή $|\eta| < 0.1479$ και αποτελείται από επιμέρους κομμάτια τα οποία ονομάζονται supermodules. Κάθε supermodule αποτελείται από 4 modules, στα οποία το πρώτο περιέχει 500 κρυστάλλους και τα υπόλοιπα 400 κρυστάλλους. Συνολικά υπάρχουν 26 supermodules και 61200 κρύσταλλοι στο EB. Τέλος το κομμάτι του ανιχνευτή στο endcap καλύπτει την περιοχή $0.1479 < |\eta| < 3$. Το EE αποτελείται από 2 ημικύκλια η δομική μονάδα του καθενός, περιέχει 5×5

¹Μήκος ακτινοβολίας (radiation length) X_0 ορίζεται το μήκος εκείνο στο οποίο η ενέργεια του σωματιδίου μειώνεται στο $\frac{1}{e}$ της αρχικής λόγω εκπομπής ακτινοβολίας

²Η ακτίνα Moliere είναι ένα μέγεθος το οποίο χαρακτηρίζει τον ηλεκτρομαγνητικό καταγισμό που δημιουργείται από ένα φωτόνιο ή φορτισμένο σωματίδιο. Ορίζεται ως η ακτίνα του κυλίνδρου μέσα στον οποίο περιέχεται το 90% της ενέργειας καταγισμού

κρυστάλλους που ονομάζονται "supercrystals", έτσι κάθε EE περιέχει 7324 κρυστάλλους.

Τέλος αξίζει να σημειωθεί πως το ECAL σχεδιάστηκε ώστε να μπορεί να μετρήσει με μεγάλη ακρίβεια τις ενέργειες των φωτονίων, καθώς ένας από τους επικρατέστερους τρόπους διάσπασης του μποζονίου του Higgs είναι μέσω της αντίδρασης $H \rightarrow \gamma\gamma$ σε δύο πολύ ενεργητικά φωτόνια. Επίσης, δύο φωτόνια παράγονται και από τη διάσπαση του ουδέτερου πιονίου $\pi^0 \rightarrow \gamma\gamma$. Συνεπώς, το θερμιδόμετρο αυτό σχεδιάστηκε ώστε να διαθέτει πολύ καλή διακριτική ικανότητα για την ανίχνευση των δύο φωτονίων, καθώς ο διαχωρισμός των δύο αυτών διεργασιών είναι πολύ σημαντικός για την ανίχνευση του μποζονίου Higgs. Έτσι, όπως φαίνεται και στην εικόνα (), στο ECAL περιλαμβάνεται και ένα επιπλέον θερμιδόμετρο, το Preshower, που βρίσκεται μόνο στα endcaps του CMS. Ο λόγος ύπαρξής του είναι η ταυτοποίηση και απόρριψη των γεγονότων που προέρχονται από τη διάσπαση του ουδέτερου πιονίου, καθώς εαν τα δύο φωτόνια που προέρχονται από τη διάσπασή του φύγουν σε πολύ μικρές γωνίες, μπορεί να μην ταυτοποιηθούν σαν ξεχωριστά σωματίδια αλλά σαν ένα πολύ ενεργητικό.

2.2.2.2 Το Αδρονικό Καλορίμετρο (HCAL)

Το τρίτο επίπεδο ανίχνευσης του CMS αποτελείται από το αδρονικό θερμιδόμετρο (HCAL). Το ανιχνευτικό σύστημα αυτό είναι υπεύθυνο για την μέτρηση της ενέργειας των αδρονίων, δηλαδή σωματιδίων που αποτελούνται από quarks και gluons καθώς και των προϊόντων τους. Επίσης, παρέχει τη δυνατότητα έμμεσης μέτρησης για την ύπαρξη ή μη σωματιδίων που δεν αλληλεπιδρούν καθόλου ή αλληλεπιδρούν ελάχιστα με την ύλη, όπως τα νετρίνα. Αυτό γίνεται μέσω της μέτρησης της έλλειψης ορμής, μέσω της οποίας μπορεί να πιστοποιηθεί κάποιο σωματίδιο σαν προϊόν της αντίδρασης το οποίο δεν αλληλεπίδρασε με την ενεργό ύλη του ανιχνευτή και επομένως δεν ανιχνεύθηκε.

Η λειτουργία του αδρονικού θερμιδόμετρου βασίζεται στη δημιουργία αδρονικών καταγισμών, οι οποίοι είναι αποτέλεσμα της αλληλεπίδρασης των αδρονίων με τον πυρήνα του υλικού του ανιχνευτή. Η διαδικασία τώρα η οποία είναι υπεύθυνη για τη διάδοση του αδρονικού καταγισμού είναι πολύ διαφορετική από αυτή του ηλεκτρομαγνητικού καταγισμού. Γενικά τα αδρόνια μπορεί να σκεδαστούν:

- ελαστικά: $p + N \rightarrow p + N$ (σ_e)
- ανελαστικά: $p + N \rightarrow X$ (σ_{incl})

με τους πυρήνες του υλικού στο οποίο διαδίδονται. Κατά τις ελαστικές σκεδάσεις, τα αδρόνια μεταφέρουν ένα μέρος της ενέργειας τους στον πυρήνα και αλλάζουν την πορεία τους. Για την μέτρηση της ενέργειας ενός αδρονίου από τον ανιχνευτή, οι ανελαστικές σκεδάσεις παίζουν σημαντικότερο ρόλο από τις ελαστικές. Κατά τις ανελαστικές σκεδάσεις, το αδρόνιο χτυπά τον πυρήνα και ένα από αυτά ή και τα δύο αλλάζουν ταυτότητα δηλαδή δημιουργούνται καινούργια σωματίδια. Καθώς αδρόνια υψηλής ενέργειας περνούν μέσα από το υλικό του ανιχνευτή σκεδάζονται πολλές

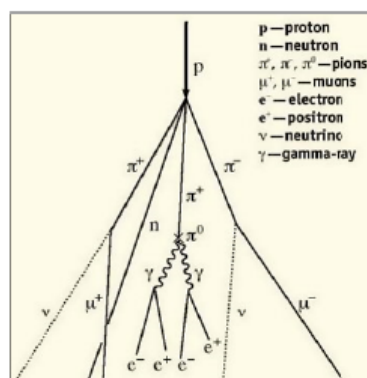


Figure 2.8: Σχηματική αναπαράσταση της ανάπτυξης ενός αδρονικού καταίγισμου. Όπως φαίνεται και στην εικόνα κάποια στιγμή κατά της ανάπτυξη του καταίγισμου, δημιουργείται ένα ουδέτερο πιόνιο (π^0). Αυτό έχει σαν συνέπεια την ύπαρξη ενός ηλεκτρομαγνητικού κομματιού μέσα στον αδρονικό καταίγισμό.

φορές, δημιουργώντας αδρονικούς καταίγισμούς () όπως ακριβώς συμβαίνει και τους ηλεκτρομαγνητικούς καταίγισμούς.

Οι αδρονικοί καταίγισμοί είναι πολύ πιο πολύπλοκοι από τους ηλεκτρομαγνητικούς, μεγαλύτεροι σε μήκος και πιο πυκνοί και επομένως απαιτούν περισσότερο χρόνο για να εξελιχθούν. Οι διαδικασίες που λαμβάνουν χώρα κατά της δημιουργία τους υπόκεινται σε στατιστικές διακυμάνσεις και η περιγραφή των καταίγισμών είναι εξαιρετικά πολύπλοκη. Τα φαινόμενα τα οποία παίζουν καθοριστικό ρόλο στην ανάπτυξη ενός αδρονικού καταίγισμου είναι: η παραγωγή αδρονίων, η πυρηνική αποδιέγερση και η αποσύνθεση πιονίων και μιονίων. Τα ουδέτερα πιόνια αποτελούν το 1/3 των συνολικά παραγόμενων πιονίων και η ενέργεια τους διασκορπίζεται σε μορφή ηλεκτρομαγνητικού καταίγισμου, καθώς όπως αναφέρθηκε διασπάται σε 2 φωτόνια ($\pi^0 \rightarrow 2\gamma$). Μια ποσότητα που χρησιμοποιείται για την διαμήκη ανάπτυξη των αδρονικών καταίγισμών είναι το μήκος αλληλεπίδρασης το οποίο ορίζεται ως η μέση απόσταση που πρέπει να διανύσει ένα αδρόνιο για να αντιδράσει μέσω ισχυρής αλληλεπίδρασης, και δίνεται από τον τύπο: $\lambda_{had} = \frac{A}{N_A \sigma_{abs}}$, όπου A ο αριθμός Avogadro και σ_{abs} η απόλυτη ενεργός διατομή της διαδικασίας.

Το HCAL είναι ένα δειγματοληπτικό θερμιδόμετρο το οποίο μπορεί να προσδιορίσει τη θέση, την ενέργεια και τη χρονική στιγμή που έφτασε το σωματίδιο. Η μέτρηση αυτή πραγματοποιείται με τη χρήση εναλλασσόμενων στρωμάτων ενεργού υλικού που χρησιμοποιείται για ανίχνευση και απορροφητών. Σαν απορροφητής χρησιμοποιείται κυρίως ορείχαλκος γιατί δεν είναι μαγνητικό υλικό, αλλά και ατσάλι. Όταν ένα σωματίδιο διέλθει του ενεργού υλικού αυτό φθορίζει και παράγει φως μπλε χρώματος το οποίο συλλέγεται από ειδικές ίνες μετατόπισης κύματος και μετατρέπεται σε φως πράσινου χρώματος. Στη συνέχεια οδηγείται σε ειδικούς ανιχνευτές, τους

Hybrid Photodiodes (HPD) με σκοπό την ενίσχυσή του και την μετατροπή του σε ψηφιακό σήμα. Τέλος, μεταφέρεται στη μονάδα συλλογής δεδομένων.

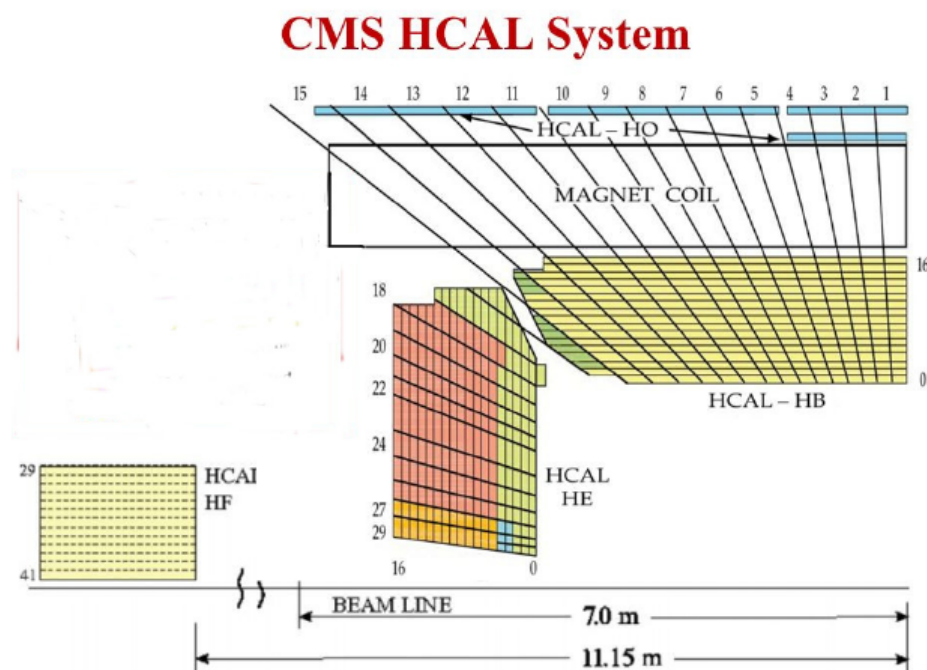


Figure 2.9: Το αδρονικό θερμιδόμετρο (HCAL) του ανιχνευτή CMS. Το κομμάτι στο barrel αποτελείται από το HCAL barrel και το HCAL outer. Στα endcaps υπάρχει το HCAL endcap καθώς και το HCAL forward detector.

Το HCAL (Εικόνα) καλύπτει μια περιοχή $1.8 < r < 2,9$ και $|\eta| < 5$. Όπως και οι υπόλοιποι ανιχνευτές έτσι και αυτός αποτελείται από 2 κομμάτια, αυτό που βρίσκεται στο κυρίως μέρος του CMS, το HCAL barrel (HB) και αυτό που βρίσκεται στα endcaps, HCAL endcaps (HE). Τα δύο αυτά κομμάτια, δίνουν κάλυψη $|\eta| < 3$. Το HCAL συμπληρώνεται από το HF, HCAL forward detector, το οποίο έχει το ρόλο ανίχνευσης των αδρονίων που φεύγουν σε πολύ μικρές γωνίες σε σχέση με τον άξονά της δέσμης. Όπως είναι λογικό αυτό το κομμάτι του CMS δέχεται πολύ μεγάλη ακτινοβολία καθώς όταν οι δέσμες διασταυρώνονται το μεγαλύτερο μέρος των πρωτονίων δεν συγκρούονται μεταξύ τους και απλά εκτρέπονται με σκοπό να καταλήγουν πάνω στους ανιχνευτές των endcaps. Για το λόγο αυτό, στο HF χρησιμοποιούνται οπτικές ίνες που είναι πολύ ανθεκτικές στην ακτινοβολία. Τέλος, ενώ το HCAL βρίσκεται στο εσωτερικό του μαγνήτη τα αδρόνια που φεύγουν σε πολύ μεγάλες γωνίες, σχεδόν κάθετα στον άξονα της δέσμης, δεν διανύουν μεγάλη απόσταση μέσα στο θερμιδόμετρο. Έτσι η πληροφορία για μέτρηση της ενέργειας τους δεν είναι αρκετή. Για το λόγο αυτό, ένα ακόμα επίπεδο του HCAL έχει τοποθετηθεί εκτός του μαγνήτη,

το οποίο ονομάζεται HCAL outer (HO) και καλύπτει μια περιοχή $|\eta| < 1.26$.

2.2.3 Ο Μαγνήτης

Ένα από τα κυριότερα θέματα στην κατασκευή του CMS ήταν η επιλογή του μαγνήτη, το είδος καθώς και η ισχύς του. Στο CMS επιλέχθηκε η χρήση ενός υπεραγωγίσιμου σωληνοειδούς ($L = 13\text{m}$) παράλληλα στον άξονα της δέσμης, το οποίο παράγει μαγνητικό πεδίο 4T. Ο μαγνήτης είναι αρκετά μεγάλος, ώστε τα 3 από τα 4 υποανιχνευτικά συστήματα του CMS, εκτός από τον ανιχνευτή μιονίων, να βρίσκονται στο εσωτερικό του. Ο ρόλος του είναι η καμπύλωση των τροχιών των σωματιδίων με σκοπό να μετρηθεί με ακρίβεια η ορμή τους. Ο σωληνοειδής μαγνήτης του CMS καλύπτει μια περιοχή $|\eta| < 1.5$

2.2.4 Ο ανιχνευτής μιονίων

Στο τελευταίο επίπεδο του CMS βρίσκεται το ανιχνευτικό σύστημα μιονίων. Η ανίχνευση των μιονίων είναι πολύ σημαντική στα πειράματα που διεξάγονται στον LHC. Επίσης, τα μίονια είναι ιδιαίτερα διεισδυτικά σωματίδια και μπορούν να διανύσουν πολλά μέτρα μέσα σε κάποιο υλικό χωρίς να αλληλεπιδράσουν και άρα να μην ανιχνευτούν. Επομένως, για την ανίχνευση τους χρησιμοποιείται ένα ολόκληρο ξεχωριστό σύστημα ανιχνευτών. Το μιονικό σύστημα εκτός από την ανίχνευση των μιονίων, παίζει κύριο ρόλο στο σύστημα σκανδαλισμού του CMS.

Το μιονικό σύστημα του CMS, αποτελείται από 4 σταθμούς και καλύπτει συνολικά μια έκταση 25000 m^2 ενεργής περιοχής, η περιοχή δηλαδή στην οποία μπορεί να ανιχνευθεί ένα σωματίδιο και καλύπτει μια περιοχή $4 < r < 7.4$ και $|\eta| < 2.4$. Λόγω της μεγάλης έκτασης που καλύπτει και της διαφορετικής έντασης της ακτινοβολίας σε κάθε περιοχή, το μιονικό σύστημα αποτελείται από τρία διαφορετικά είδη ανιχνευτών, οι οποίοι είναι ανιχνευτές αερίου και κάθε ένας από τους οποίους έχει διαφορετικό ρόλο. Η τεχνολογία στην οποία βασίζονται οι ανιχνευτές μιονίων είναι απλή. Όταν ένα φορτισμένο σωματίδιο, στη συγκεκριμένη περίπτωση μίονιο, διέρχεται από έναν ανιχνευτή αερίου εάν έχει αρκετή ενέργεια μπορεί να ιονίσει τα άτομα του αερίου, δηλαδή προσδίδεται ενέργεια στα ηλεκτρόνια σθένους του ατόμου του αερίου και εκείνα αποδεδμεύονται από το άτομο. Το τελικό αποτέλεσμα είναι η δημιουργία χιονοσοιβάδας ηλεκτρονίων. Η χιονοσοιβάδα ηλεκτρονίων που έχουν ξεφύγει από τα άτομα συλλέγονται με χρήση ηλεκτρικού πεδίου που παράγεται από μια άνοδο και μια κάθοδο (θετικά φορτισμένο καλώδιο και αρνητικά φορτισμένο καλώδιο αντίστοιχα), καθώς τα ηλεκτρόνια επιταχύνονται από την άνοδο.

Οι 3 επιμέρους ανιχνευτές του μιονικού συστήματος είναι οι Drift Tubes, Cathode Strip Chambers και Resistive Plate Chambers, οι δύο πρώτοι έχουν πολύ καλή χωρική διακριτική ικανότητα και οι τελευταίοι καλή χρονική διακριτική ικανότητα.

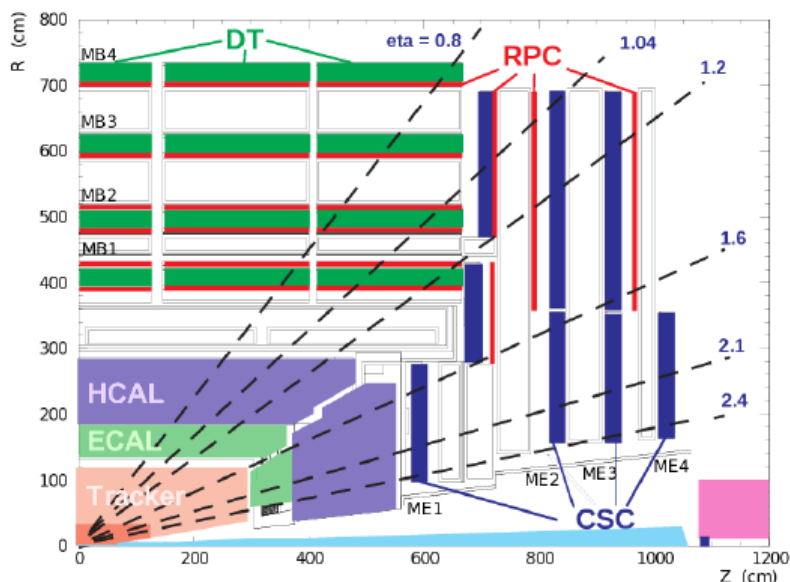


Figure 2.10: Το μιονικό σύστημα του CMS. Συνολικά χρησιμοποιούνται τρία διαφορετικά είδη ανιχνευτών για την ανίχνευση των μιονίων στο CMS. Οι drift tubes (DT) οι οποίοι είναι τοποθετημένοι στο barrel, οι cathode strip chambers (CSC) οι οποίοι είναι στα endcaps και τέλος οι resistive plate chambers (RPC) οι οποίοι είναι τοποθετημένοι και στα δύο κομμάτια του ανιχνευτή.

• Drift Tubes (DT)

Βρίσκονται στο κύριο σώμα του CMS, στο barrel, όπου η ροή σωματιδίων είναι χαμηλή και το μαγνητικό πεδίο ομογενές και όχι τόσο ισχυρό. Το σύστημα των DFs αποτελείται από 5 κυλίνδρους ο κύριος άξονας των οποίων συμπίπτει με τον άξονα της δέσμης. Κάθε κύλινδρος αποτελείται από 4 stations με διαφορετική ακτίνα το καθένα (MB1- MB4). Κάθε σωλήνας (tube) έχει διάμετρο 4 εκατοστά και στο κέντρο του υπάρχει ένα σύρμα και μείγμα αερίων αργού και διοξειδίου του άνθρακα. Όταν ένα μόνιο περάσει μέσα από το αέριο, ιονίζει τα μόρια του αερίου. Τα ηλεκτρόνια που δημιουργούνται από τον ιονισμό κατευθύνονται προς το σύρμα λόγω του ηλεκτρικού πεδίου που υπάρχει στο εσωτερικό του σωλήνα. Ανιχνεύοντας το σημείο που τα ηλεκτρόνια έπεσαν πάνω στο σύρμα καθώς και το χρόνο ολίσησης (drift) μέχρι να φτάσουν στο σύρμα υπολογίζεται η τροχιά του μιονίου. Κάθε ανιχνευτής τύπου DT, έχει μέγεθος περίπου 2×2.5 m και αποτελείται από 12 στρώματα χωρισμένα σε 3 ομάδες με συνολικά 60 tubes. Η μεσαία ομάδα, ανιχνεύει την συντεταγμένη της τροχιάς κατά τον άξονα της δέσμης, ενώ οι άλλες 2 ομάδες τις συντεταγμένες στο επίπεδο κάθετο στον άξονα της δέσμης.

- **Cathode Strip Chambers (CSC)**

Βρίσκονται στα endcaps κάθετα στη διεύθυνση της δέσμης, όπου η ροή σωματιδίων είναι μεγάλη και το μαγνητικό πεδίο ανομοιογενές. Οι CSC έχουν πολύ μικρό χρόνο απόκρισης και έτσι μπορούν να χρησιμοποιηθούν σε περιβάλλον με τόσο μεγάλη ακτινοβολία. Επίσης έχουν τη δυνατότητα να παρέχουν πληροφορίες σχετικά με το χρόνο διέλευσης του σωματιδίου. Η αρχή λειτουργίας τους βασίζεται σε ένα πλέγμα συρμάτων. Διαθέτουν θετικά φορτισμένα σύρματα τα οποία λειτουργούν σαν άνοδοι και αρνητικά φορτισμένα κάθετα στις ανόδους που λειτουργούν σαν κάθοδοι. Το πλέγμα βρίσκεται μέσα σε ένα μείγμα αερίου. Όταν ένα φορτισμένο σωματίδιο περάσει μέσα από την ενεργή περιοχή του ανιχνευτή τα μόρια του αερίου ionίζονται και τα ηλεκτρόνια που παράγονται από τον ionισμό οδηγούνται προς τις καθόδους δημιουργώντας χιονοστιβάδα. Το σήμα που παράγεται από την χιονοστιβάδα δίνει πληροφορία για την τροχιά του σωματιδίου που πέρασε.

- **Resistive Plate Chambers (RPC)**

Βρίσκονται τόσο στο barrel αλλά και στα endcaps του CMS και λειτουργούν συμπληρωματικά για τους DT και τους CSC παρέχοντας πληροφορίες για το σκανδαλισμό. Αποτελούνται από δύο πλάκες, μια θετικά φορτισμένη που λειτουργεί σαν άνοδος και μια αρνητικά φορτισμένη που έχει το ρόλο της καθόδου και στο ενδιάμεσο (2 mm περίπου) τους υπάρχει αέριο. Όταν ένα μόνιο περάσει από το αέριο προκαλεί ionισμό. Λόγω των φορτισμένων πλακών στην περιοχή του αερίου υπάρχει ηλεκτρικό πεδίο. Έτσι τα ηλεκτρόνια που παράγονται από τον ionισμό επιταχύνονται και δημιουργούν χιονοστιβάδα. Τα προϊόντα της χιονοστιβάδας οδηγούνται προς την άνοδο όπου και συλλέγονται δίνοντας έτσι πληροφορία για το μόνιο που πέρασε. Το ηλεκτρικό πεδίο είναι τέτοιο ώστε το σήμα να έρχεται σε πολύ μικρότερο χρόνο από την συχνότητα των συγκρούσεων. Λόγω της πολύ γρήγορης απόκρισης αλλά και της καλής διακριτικής τους ικανότητας οι RPC χρησιμοποιούνται για τη διαδικασία του triggering.

2.2.5 Σύστημα σκανδαλισμού(Trigger)

Κατά τη διάρκεια λειτουργίας του LHC οι ανιχνευτές θα καταγράφουν γεγονότα με συχνότητα 109Hz. Κατά τη διαδικασία αυτή, κάθε ένας από τους ανιχνευτές του LHC θα παράγει περίπου 100 TeraByte δεδομένων το δευτερόλεπτο. Όλη αυτή η πληροφορία είναι αδύνατο να αποθηκευθεί πρώτον διότι δεν υπάρχουν τα μέσα για την αποθήκευση τόσο μεγάλου όγκου δεδομένων και δεύτερον διότι ένα πολύ μικρό κομμάτι από αυτά τα δεδομένα είναι ενδιαφέροντα από πλευράς φυσικής ώστε να χρησιμοποιηθούν σε περαιτέρω ανάλυση. Για το λόγο αυτό όλοι οι ανιχνευτές χρησιμοποιούν συστήματα με σκοπό τη μείωση του όγκου των δεδομένων.

Στο CMS το συγκεκριμένο σύστημα ονομάζεται Trigger and Data Acquisition System (TRIDAS) και η διαδικασία αυτή πραγματοποιείται σε 2 στάδια. Πρώτα τα δεδομένα περνούν από τον Level-1 trigger και στη συνέχεια από τον High Level

Trigger (HLT). Ο Level-1 Trigger είναι το πρώτο στάδιο του συστήματος σκανδαλισμού του CMS. Ο συνολικός χρόνος συλλογής των δεδομένων και της ανάλυσης τους από το σύστημα είναι περίπου 3.2 μs . Λόγω προβλημάτων καθυστέρησης, ο trigger έχει στην πραγματικότητα περίπου 1 μs για να "αποφασίσει". Τα δεδομένα αποθηκεύονται σε buffers μέχρι ο trigger να "αποφασίσει" αν θα κρατήσει ή όχι τα δεδομένα. Ο Level-1 trigger αποτελείται αποκλειστικά από hardware το οποίο βρίσκεται τοποθετημένο όσο το δυνατόν πιο κοντά στον ανιχνευτή για να μειωθούν φαινόμενα καθυστέρησης λόγω μεταφοράς της πληροφορίας. Εξαιτίας του πολύ μικρού χρόνου που έχει στη διάθεση του το σύστημα αυτό, τα δεδομένα τα οποία χρησιμοποιεί προέρχονται αποκλειστικά από τα θερμιδόμετρα και το σύστημα μιονίων. Μετά το στάδιο αυτό ο συνολικός όγκος των δεδομένων έχει μειωθεί σε 50 kHz και τα δεδομένα που διαλέγονται μεταφέρονται στον HLT για περαιτέρω ανάλυση.

Ο HLT αποτελείται αποκλειστικά από software. Το σύστημα αυτό χρησιμοποιεί αλγορίθμους ανάλογους με αυτούς που χρησιμοποιούνται από τους φυσικούς για την offline ανάλυση και το φιλτράρισμα των γεγονότων γίνεται σε 3 στάδια. Στο πρώτο στάδιο χρησιμοποιούνται δεδομένα μόνο από τα θερμιδόμετρα και το μιονικό σύστημα ανάλογα με τον L1 trigger αλλά με μεγαλύτερη λεπτομέρεια. Στη συνέχεια χρησιμοποιούνται δεδομένα και από τον tracker. Τέλος χρησιμοποιείται όλη η πληροφορία και γίνεται ανακατασκευή ολόκληρου του γεγονότος. Μόλις η διαδικασία του triggering έχει ολοκληρωθεί τα δεδομένα θα έχουν μειωθεί σε 100 Hz. Τα γεγονότα που έχουν περάσει και τα δύο επίπεδα του trigger αποθηκεύονται και χρησιμοποιούνται για offline ανάλυση.

Κεφάλαιο 3

Ανακατασκευή Φυσικών Αντικειμένων

Ο ανιχνευτής CMS καταγράφει τεράστιο πλήθος μετρήσεων επομένως ο αριθμός των δεδομένων που συλλέγονται, από τους subdetectors είναι εξίσου τεράστιος. Οι ενεργειακές εναποθέσεις των σωματιδίων σε κάθε στρώμα του ανιχνευτή που μετράται όπως αναλύθηκε στο προηγούμενο κεφάλαιο με διαφορετικές τεχνολογίες πρέπει να "μεταφραστεί" σε αντικείμενα τα οποία θα μπορούμε να διαχειριστούμε και να επεξεργαστούμε. Έτσι, τα δεδομένα ανακατασκευάζονται με τη βοήθεια αλγορίθμων με σκοπό την μετέπειτα ανάλυση τους και τελικά την ταυτοποίηση και ανίχνευση των προϊόντων των συγκρούσεων πρωτονίων.

Στην ενότητα αυτή γίνεται αναφορά στον αλγόριθμο ροής σωματιδίων (Particle flow algorithm), που χρησιμοποιείται κατα κόρον στο CMS για την ανακατασκευή και ταυτοποίηση σωματιδίων που προέρχονται από τις συγκρούσεις πρωτονίων. Στην συνέχεια γίνεται μια σύντομη επισκόπηση στους αδρονικούς πίδακες σωματιδίων και σκιαγραφούνται οι βασικοί αλγόριθμοι που χρησιμοποιούνται στο CMS για την ανακατασκευή των jets. Τέλος, γίνεται αναφορά στις διαδικασίες που αφορούν στην ενεργειακή διόρθωση και βαθμονόμηση αυτών, και κυρίως στον αλγόριθμο Pile Up Per Particle Identification (PUPPI), που χρησιμοποιείται για την ενεργειακή διόρθωση των jets και την απομάκρυνση των pileup jets.

3.1 Αδρονικοί πίδακες και το φαινόμενο Pileup

Κατά τις συγκρούσεις πρωτονίων παράγονται παρτόνια (κουάρκς και γκλουόνια) τα οποία λόγω των περιορισμών της QCD ως έγχρωμες καταστάσεις της ύλης δεν μπορούν να υπάρξουν ως ελεύθερες καταστάσεις, επομένως δεν μπορούν να ανιχνευθούν, αυτό που τελικά μετράται στον ανιχνευτή είναι το ενεργειακό αποτύπωμα της

αδρανοποίησης ¹ αυτών των παρτονίων. Μετράται δηλαδή το αποτέλεσμα της διαδικασίας μετατροπής έγχρωμων παρτονίων σε καταιγιισμό άγχρωμων καταστάσεων, τα αδρόνια. Το σύνολο αυτών των σωματιδίων, των τελικών δηλαδή καταστάσεων ονομάζονται **πίδακες, jets**. Τα jets αποτελούν το ενεργειακό "αποτύπωμα" των παρτονίων και κάθε ένα αποτελεί έναν στενό κώνο που περιέχει κυρίως σωματίδια όπως πιόνια, καόνια ακόμα και πρωτόνια και νετρόνια. Τέλος, τα jets μπορούν να θεωρηθούν ως ένα τοπικό cluster ενέργειας συνοδευόμενα από τροχιές σωματιδίων.

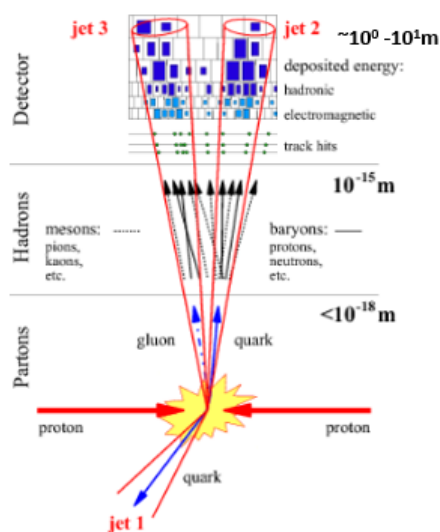


Figure 3.1: Κατά τις συγκρούσεις πρωτονίων παράγονται παρτόνια (κουάρκς και γκλουόνια), τα οποία δεν μπορούν να υπάρξουν ως ελεύθερες καταστάσεις, επομένως δεν μπορούν να ανιχνευθούν. Αυτό που τελικά μετράται στον ανιχνευτή είναι το αποτέλεσμα της διαδικασίας μετατροπής έγχρωμων παρτονίων σε καταιγιισμό άγχρωμων καταστάσεων, τα αδρόνια (αδρανοποίηση). Το σύνολο αυτών των σωματιδίων ονομάζονται jets. Τα jets αποτελούν το ενεργειακό "αποτύπωμα" των παρτονίων και κάθε ένα αποτελεί έναν στενό κώνο που περιέχει κυρίως σωματίδια όπως πιόνια, καόνια ακόμα και πρωτόνια και νετρόνια. Τέλος, τα jets μπορούν να θεωρηθούν ως ένα τοπικό cluster ενέργειας συνοδευόμενα από τροχιές σωματιδίων.

Τα jets διαχωρίζονται σε δύο γενικές κατηγορίες:

- **Prompt jet:** jet που προέρχονται από την ίδια κύρια σύγκρουση (primary vertex)
- **Pileup:** jet που προέρχονται από δευτερογενείς κύριες συγκρούσεις στο ίδιο

¹Αδρανοποίηση (Hadronization): η διαδικασία σχηματισμού αδρονίων από παρτόνια (κουάρκς ή γκλουόνια)

bunch crossing, μετατοπισμένες συγκρούσεις στον άξονα της δέσμης (secondary primary vertices)

Ως κύρια σύγκρουση (primary vertex PV) ορίζεται ως η κορυφή με το μεγαλύτερο άθροισμα του τετραγώνου των ορμών ($\sum p_T^2$) φορτισμένων τροχιών που συνδέονται με αυτή. Για να ικανοποιείται αυτή η συνθήκη απαιτείται επίσης στην κορυφή να περιέχονται τουλάχιστον τέσσερις τροχιές και η μέγιστη απόσταση από το ονομαστικό σημείο αλληλεπίδρασης να είναι μικρότερο από 24 cm κατά μήκος του άξονα z.

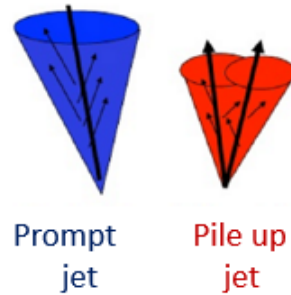


Figure 3.2: τα jets διαχωρίζονται σε δύο γενικές κατηγορίες τα prompt jets και τα pileup jets. Ως prompt jet θεωρείται το jet που προέρχονται από την ίδια κύρια σύγκρουση (primary vertex) ενώ pileup jet το jet που προέρχονται από δευτερογενείς κύριες συγκρούσεις στο ίδιο bunch crossing, μετατοπισμένες συγκρούσεις στον άξονα της δέσμης (secondary primary vertices)

Για τη λήψη υψηλού αριθμού δεδομένων στον LHC απαιτείται υψηλός ρυθμός συγκρούσεων, οι οποίες πραγματοποιούνται σε συγκρούσεις δεσμών (bunch crossings) με πολλή υψηλή συχνότητα όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, κάθε 25ns. Όμως η υψηλή στιγμιαία φωτεινότητα που απαιτείται στον LHC έχει ως αποτέλεσμα πολλαπλές συγκρούσεις πρωτονίων οι οποίες λαμβάνουν χώρα ανα bunch crossing, όσες δεν προέρχονται από την κύρια σύγκρουση οφείλονται κυρίως σε soft QCD διεργασίες και παράγουν πρόσθετες τροχιές στον ανιχνευτή τροχιών (Tracker) αλλά και εναποθέτουν ενέργεια στα θερμιδόμετρα. Έτσι έχουμε το φαινόμενο του **pileup contamination**. Η συνεισφορά του φαινομένου που προέρχεται από τις επιπρόσθετες συγκρούσεις στο ίδιο bunch crossing ονομάζεται εντός χρόνου (in-time) pileup (IT PU). Από την άλλη, λόγω του χρόνου εναποθέσης του τελικού σήματος στα θερμιδόμετρα, οι συγκρούσεις πρωτονίων που συμβαίνουν στο προηγούμενο και στα επόμενα bunch crossings συμβάλλουν επίσης στην θερμιδομετρική ενέργεια στο ίδιο χρονικό παράθυρο με την κύρια σκληρή αλληλεπίδραση. Αυτή η συνεισφορά ονομάζεται pileup εκτός χρόνου (out-of-time) (OOT PU).

3.2 Ανακατασκευή αδρονικών πιδάκων

Η ανακατασκευή των πιδάκων είναι ένα πολύ κρίσιμο κομμάτι της ανάλυσης των αδρονικών τελικών καταστάσεων (hadronic final states). Λόγω της διαδικασίας καταϊγισμού, μικρές αλλαγές κατά την αρχή της αδρανοποίησης, μπορούν να οδηγήσουν σε αρκετά μεγάλες αποκλίσεις στο τελικό αποτέλεσμα. Επίσης, ένας άλλος παράγοντας που κάνει τη διαδικασία της ταυτοποίησης δύσκολη είναι ο ηλεκτρονικός θόρυβος. Για την αντιμετώπιση των προβλημάτων αυτών, έχουν δημιουργηθεί διάφοροι αλγόριθμοι για τη ταυτοποίηση και μελέτη των πιδάκων.

Η διαδικασία ανακατασκευής των jets ξεκινά διαβάζοντας την αρχική πληροφορία που συλλέχθηκε από τα θερμιδόμετρα. Τα δεδομένα που συλλέγονται από το ECAL και το HCAL μας δίνουν πληροφορίες οι οποίες χρησιμοποιούνται ως είσοδος στην όλη διαδικασία ανακατασκευής, η οποία εξελίσσεται σε 4 στάδια. Στο πρώτο στάδιο εφαρμόζονται σε κάθε κανάλι από κάθε θερμιδόμετρο, πρώτα κατώτα όρια καταστολής. Για το HCAL το όριο αυτό είναι 2-4ADC counts με μετρούμενο θόρυβο στα 0.75 ADC counts. Ενώ στο ECAL, η ελάχιστη ενέργεια είναι 90MeV στο βαρέλι και 450MeV στα endcaps του ανιχνευτή, ενέργεια που αντιστοιχεί σε 2-3σ πάνω από το μετρούμενο θόρυβο.

Στο δεύτερο στάδιο, οι ξεχωριστές πληροφορίες που συλλέγονται από τα θερμιδόμετρα συνδυάζονται σε "πύργους" ECalPlusHCAL Towers, οι οποίοι αποτελούνται από 1 HCal και 3×3 ECal κελιά. Στο στάδιο αυτό επιβάλλεται ένα cut στη διαμήκη ενέργεια κάθε πύργου στα $E_T = 0.5 MeV$, με σκοπό την μείωση συλλογής ενέργειας που προέρχεται από pileup και underlying events στα θερμιδομετρα. Στη συνέχεια, στο τρίτο στάδιο οι ECalPlusHCAL Towers αποτελούν την είσοδο στον πραγματικό αλγόριθμο ανακατασκευής των jets, οποίος συγκεντρώνει όλη την ενέργεια που βρίσκεται στα jet. Πριν το τελικό στάδιο εφαρμόζεται ένα cut στη διαμήκη ενέργεια $E_T = 10 MeV$, καθώς jets με χαμηλές ενέργειες τείνουν να μην είναι καλά ορισμένα. Τελικά τα ανακατασκευασμένα jets βαθμονομούνται με τη βοήθεια καλά ορισμένων φυσικών γεγονότων.

3.2.1 Ο Αλγόριθμος ροής σωματιδίων - Particle Flow Algorithm

Ο αλγόριθμος ροής σωματιδίων (PF) χρησιμοποιείται σε όλες τις αναλύσεις του CMS και στοχεύει στην αναγνώριση αλλά και την ανακατασκευή όλων των σωματιδίων, τα οποία προέρχονται από τις συγκρούσεις πρωτονίων, συνδυάζοντας άριστα τις πληροφορίες των διαφορετικών υποανιχνευτών. Το προκύπτον ανακατασκευασμένο συμβάν από τον αλγόριθμο ροής σωματιδίων οδηγεί σε ανακατασκευή των jets και την ταυτοποίηση ηλεκτρονίων, μιονίων και ταυ σωματιδίων με ικανοποιητικά βελτιωμένη απόδοση.

Για την αποτελεσματικότητα του αλγορίθμου PF πρωταρχικής σημασίας είναι οι ακόλουθες αρχές. Αρχικά, βασίζεται στη μεγιστοποίηση του διαχωρισμού μεταξύ φορτισμένων και ουδέτερων αδρόνων, απαιτείται ένα ισχυρό ενιαίο πεδίο και τέλος

είναι απαραίτητη η μεγάλη διακριτική ικανότητα των θερμιδομέτρων. Επίσης, ένα βασικό στοιχείο είναι η ύπαρξη ενός αποδοτικού ανιχνευτή τροχιάς. Ο ανιχνευτής CMS ικανοποιεί αρκετές από αυτές τις απαιτήσεις, με field integral περισσότερο από δύο φορές μεγαλύτερο από ό,τι σε άλλα προηγούμενα ή υπάρχοντα πειράματα και ένα ηλεκτρομαγνητικό θερμιδόμετρο με εξαιρετική απόδοση και λεπτομερή ανάλυση.

Ο αλγόριθμος CMS PF βασίζεται σε μια αποδοτική και καθαρή ανακατασκευή τροχιάς, σε έναν αλγόριθμο ομαδοποίησης (clustering) ικανό να αποσαφινίσει αλληλεπικαλύπτομενους καταγισμούς σωματιδίων και τέλος σε μια αποτελεσματική διαδικασία σύνδεσης των ενεργειακών εναποθέσεων κάθε σωματιδίου στους υποανιχνευτές. Στη συνέχεια, περιγράφεται συνοπτικά η λειτουργία του αλγόριθμου CMS PF. Οι τροχίες εξάγονται μέσω των θερμιδομέτρων, αν εμπίπτουν στα όρια ενός ή περισσότερων cluster, έτσι τα clusters συσχετίζονται με την τροχιά του σωματιδίου. Το σύνολο των τροχιών και των clusters αποτελούν ένα φορτισμένο αδρόνιο και τα δομικά στοιχεία του αδρονίου αυτού δεν λαμβάνονται υπόψη πλέον στον υπόλοιπο αλγόριθμο. Τα μίονια αναγνωρίζονται εκ των προτέρων, έτσι ώστε η τροχιά τους να μην δημιουργεί σύγχυση με τις τροχίες των φορτισμένων αδρονίων. Τα ηλεκτρόνια είναι πιο δύσκολο να διαχειριστούν, καθώς εκπέμπουν ακτινοβολία πέδησης. Επομένως, για τα ηλεκτρόνια χρησιμοποιείται μια συγκεκριμένη διαδικασία ανακατασκευής της τροχιάς τους αλλά και για την σύνδεση αυτών με τα clusters φωτονίων ώστε η ενέργειά τους να μην υπολογιστεί πάνω από μία φορά. Όταν, όλες οι παραπάνω τροχίες έχουν ανακατασκευαστεί τα εναπομείναντα clusters αντιστοιχούν σε φωτόνια για τα clusters από το ECal και σε αδρόνια για αυτά από το HCal.

Όταν συσχετιστούν όλες οι ενεργειακές εναποθέσεις των σωματιδίων, μπορεί να εκτιμηθεί η φύση του κάθε σωματιδίου και οι πληροφορίες των υποανιχνευτών συνδυάζονται για να καθοριστεί με τον καλύτερο δυνατό τρόπο η τετραορμή του. Στην περίπτωση που η βαθμονομημένη ενέργεια θερμιδομέτρου των clusters, που είναι απλώς ένας γραμμικός συνδυασμός των ενεργειακών αποθέσεων στα ECal και HCal, οποίος σχετίζεται με μια τροχιά, βρίσκεται σε περίσσεια σε σχέση με την ορμή της τροχιάς σε περισσότερα από ένα cluster, η περίσσεια αποδίδεται σε ένα επικαλυπτόμενο ουδέτερο σωματίδιο (φωτόνιο ή αδρόνιο), που φέρει ενέργεια που αντιστοιχεί στη διαφορά των δύο μετρήσεων.

Η προκύπτουσα λίστα σωματιδίων, δηλαδή τα φορτισμένα αδρόνια, τα φωτόνια, τα ουδέτερα αδρόνια, τα ηλεκτρόνια και τα μίονια, στη συνέχεια χρησιμοποιούνται για την ανασυγκρότηση των jets, της ελλείπουσας εγκάρσιας ενέργειας (E_T^{miss} , MET), καθώς και για την ανακατασκευή και τον εντοπισμό των ταυ σωματιδίων από τα προϊόντα των διασπάσεων αυτών.

3.2.2 Αλγόριθμος anti- k_t

Δύο είναι οι κύριες κατηγορίες αλγορίθμων που χρησιμοποιούνται για την ανακατασκευή των jets που παράγονται κατά τις συγκρούσεις πρωτονίων, οι αλγόριθμοι **Cone Type** και οι αλγόριθμοι **Sequential Clustering**. Στην ενότητα αυτή θα γίνει μια σύντομη αναφορά στη δεύτερη κατηγορία και πιο συγκεκριμένα στον αλγόριθμο **anti- k_t** ο οποίος ανήκει σε αυτήν.

Γενικά στην κατηγορία των Sequential Clustering αλγορίθμων, στην οποία ανήκουν και οι αλγόριθμοι k_t και Cambridge/Aachen, τα σωματίδια ομαδοποιούνται στο χώρο των ορμών. Επίσης, γίνεται η υπόθεση πως τα σωματίδια στα jets εμφανίζουν μικρές διαφορές στην εγκάρσια ορμή τους. Αυτό έχει σαν αποτέλεσμα τη δημιουργία περιοχών jets οι οποίες έχουν κυμαινόμενη επιφάνεια. Στην αρχή οι αλγόριθμοι αυτοί είχαν αρκετά αργή απόδοση, ήταν όμως αποτελεσματικοί, πλέον χρησιμοποιείται το πρόγραμμα FastJet μέσω του οποίου ο υπολογισμός των jets γίνεται πολύ πιο γρήγορα. Οι αλγόριθμοι Sequential Clustering είναι σχετικά απλοί και δεν επηρεάζονται από σωματίδια χαμηλής ενέργειας. Τέλος, οι αλγόριθμοι αυτοί είναι IRC Safe. Ο όρος αυτός δηλώνει πως ένα γεγονός είναι infrared και collinear safe εαν στο όριο μιας διάσπασης ή κατα την εκπομπή ενός soft σωματιδίου, το παρατηρούμενο γεγονός παραμένει αμετάβλητο. Έτσι, σε αυτούς τους αλγόριθμους το παρατηρούμενο γεγονός παραμένει αμετάβλητο αναφορικά με συγγραμμικές εκπομπές γκλουονίων χαμηλής ενέργειας. Μέσω της ιδιότητας αυτής, ο αλγόριθμος εγγυάται την ακύρωση πραγματικών και εικονικών αποκλίσεων σε υπολογισμούς υψηλότερων όρων (higher order calculations). Για την ανακατασκευή ενός jet σε διαταρακτική QCD ο αλγόριθμος που χρησιμοποιείται πρέπει να είναι IRC safe, καθώς εκπομπή soft σωματιδίων αλλά και collinear διασπάσεις δεν πρέπει να τροποποιούν τα hard jets.

Όσον αφορά συνοπτικά τον αλγόριθμο, εισάγονται δύο αποστάσεις. Αρχικά η πρώτη απόσταση d_{ij} ορίζεται ως η απόσταση μεταξύ 2 σωματιδίων i και j

$$d_{ij} = \min\left\{\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right\} \frac{\Delta R_{ij}^2}{R^2} \quad (3.1)$$

όπου $p_{T,i}$, $p_{T,j}$ οι εγκάρσιες ορμές, $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ είναι η απόσταση των σωματιδίων i και j στον χώρο $\eta - \phi$ και R η παράμετρος ακτίνας η οποία καθορίζει το τελικό μέγεθος του jet (συνήθως 0.4-0.7). Η δεύτερη απόσταση είναι η απόσταση στο χώρο των ορμών μεταξύ του σωματιδίου i με τη δέσμη

$$d_{iB} = \frac{1}{p_{T,i}^2} \quad (3.2)$$

Ο αλγόριθμος συγκρίνει αυτές τις δύο αποστάσεις και υπολογίζει την ελάχιστη από αυτές κάθε φορά, $\min\{d_{ij}, d_{iB}\}$. Εάν η ελάχιστη είναι η d_{ij} τότε τα δύο σωματίδια i και j συγχωνεύονται σε ένα σωματίδιο με πρόσθεση των τετραορμών τους, ενώ εαν ελάχιστη είναι η απόσταση d_{iB} τότε το σωματίδιο i ταυτοποιείται ως jet και απομακρύνεται απο τη λίστα των σωματιδίων. Οι αποστάσεις επανυπολογίζονται και η διαδικασία επαναλαμβάνεται μέχρι να μην απομείνει κανένα σωματίδιο, κάθε

σωματίδιο επομένως θα αποτελεί μέρος ενός jet (Inclusive clustering) ή μέχρι να υπολογιστεί ο επιθυμητός αριθμός jets (Exclusive clustering).

Η λειτουργία και η διαδικασία ομαδοποίησης των σωματιδίων σε jets του αλγορίθμου anti- k_T μπορεί να γίνει κατανοητή εάν θεωρήσουμε ένα γεγονός με ορισμένα καλά διαχωρισμένα σωματίδια υψηλής ενέργειας, σωματίδια που προέρχονται από hard process (hard σωματίδια), με εγκάρσια ορμή p_{T1}, p_{T2}, \dots και πολλά σωματίδια χαμηλής ενέργειας, σωματίδια που προέρχονται από soft διεργασίες (soft σωματίδια). Η απόσταση $d_{1i} = \min\{\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\} \frac{\Delta R_{ij}^2}{R^2}$ μεταξύ ενός hard σωματιδίου 1 και ενός soft σωματιδίου i καθορίζεται αποκλειστικά από την εγκάρσια ορμή του σωματιδίου με υψηλή ενέργεια και της απόστασης ΔR_{1i} .

Αντίθετα, η απόσταση d_{ij} μεταξύ ομοίως διαχωρισμένων soft σωματιδίων θα είναι πολύ μεγαλύτερο. Επομένως, αυτά τα σωματίδια θα τείνουν να ομαδοποιούνται με εκείνα της υψηλής ενέργειας πολύ πριν ομαδοποιηθούν μεταξύ τους. Εάν για ένα υψηλοενεργειακό σωματίδιο δεν υπάρχει όμοιο του σε απόσταση $2R$, τότε θα υπάρχει συσσώρευση όλων των σωματιδίων χαμηλής ενέργειας μέσα σε έναν κύκλο ακτίνας R , με αποτέλεσμα έναν τέλεια κωνικό πίδακα. Εάν υπάρχει όμως ένα άλλο hard σωματίδιο (σωματίδιο 2), έτσι ώστε $R < \Delta R_{12} < 2R$, τότε θα υπάρχουν δύο πίδακες υψηλής ενέργειας. Δεν είναι δυνατόν όμως και οι δύο να είναι απόλυτα κωνικοί.

Αν $p_{T1} \gg p_{T2}$ τότε ο πίδακας 1 θα είναι κωνικός και ο πίδακας 2 θα είναι εν μέρει κωνικός, αφού θα χάσει το τμήμα που επικαλύπτεται με τον πίδακα 1. Αντίθετα, αν $p_{T1} = p_{T2}$, κανένας από τους δύο πίδακες δεν θα είναι κωνικός και το επικαλυπτόμενο μέρος θα είναι απλά διαιρούμενο, σε ίσα μέρη, με ευθεία γραμμή. Στη γενική περίπτωση, και οι δύο κώνοι θα περικοπούν, με το όριο b μεταξύ τους να ορίζεται από τη σχέση:

$$\frac{\Delta R_{2b}}{p_{T1}} = \frac{\Delta R_{2b}}{p_{T2}} \quad (3.3)$$

Στην περίπτωση $\Delta R_{12} < R$, τα σωματίδια 1 και 2 θα σχηματίσουν μαζί ένα μόνο πίδακα. Αν $p_{T1} \ll p_{T2}$ τότε θα είναι ένας κωνικός πίδακας με κέντρο στο p_{T1} . Στη γενική περίπτωση, το σχήμα είναι πιο περίπλοκο, καθώς είναι η ένωση των κώνων (ακτίνα $< R$) γύρω από κάθε υψηλοενεργειακό σωματίδιο, συνυπολογίζοντας τον κώνο ακτίνας R , έχοντας ως κέντρο τον τελικό πίδακα.

Το βασικό χαρακτηριστικό του αλγορίθμου anti- k_t είναι ότι τα σωματίδια χαμηλής ενέργειας (soft σωματίδια) δεν καθορίζουν το σχήμα του πίδακα, καθώς η απόσταση d_{ij} καθορίζεται από hard σωματίδια (υψηλής ενέργειας) και ο αλγόριθμος τείνει να συλλέγει (cluster) hard σωματίδια, τα οποία εν τέλει καθορίζουν και το σχήμα του jet.

3.2.3 Ενεργειακή βαθμονόμηση και διόρθωση των jets

Ένα πολύ σημαντικό σημείο κάθε ανάλυσης στο CMS που χρησιμοποιεί τελικές αδρονικές καταστάσεις είναι η ενεργειακή βαθμονόμηση των jets. Η ενέργεια των σωματιδίων που εναποτίθεται στα θερμιδόμετρα είναι ανάλογη με την τελική απόκριση του ανιχνευτή. Όμως η συσχέτιση αυτή ενέργειας-σήματος δεν είναι άμεση

και πρέπει να ρυθμιστεί από το πείραμα αλλά και τα πειραματικά δεδομένα. Αρχικά απαιτείται μια βασική βαθμονόμηση του ανιχνευτή που επιτυγχάνεται από test beams (δέσμες δοκιμής) με γνωστά σωματίδια και τις ενέργειες αυτών. Για τη βαθμονόμηση κατά τη διάρκεια του τρέχοντος πειράματος χρησιμοποιούνται γνωστές και αρκετά μελετημένες φυσικές διεργασίες. Για τον ECal χρησιμοποιείται η διάσπαση $Z \rightarrow e^-e^+$ ενώ για τον HCal χρησιμοποιούνται $\gamma + jets$ γεγονότα.

Όσον αφορά τη $\gamma + jets$ βαθμονόμηση, στη leading order εικόνα ενός $\gamma + jet$ γεγονότος υπάρχουν δυο παρτόνια ($q\bar{q}$ ή qg τα οποία συγκρούονται μεταξύ τους. Μέσα από αυτή τη διαδικασία μπορεί να παραχθεί ένα φωτόνιο (γ) και παρτόνιο το οποίο αντιστοιχεί σε ένα jet μέσω της διαδικασίας αδρανοποίησης, όπως φαίνεται και στα διαγράμματα Feynmann (εικόνα ??) που αντιστοιχούν στη leading order εικόνα μιας $\gamma + jet$ διαδικασίας. Η ενεργειακή βαθμονόμηση του αδρονικού καλορίμετρου από

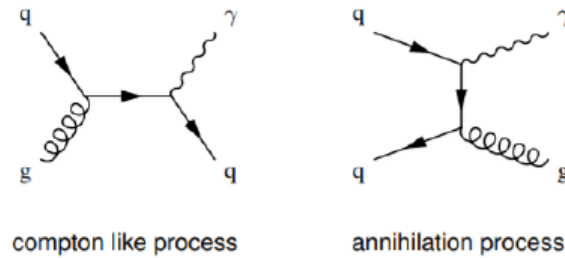


Figure 3.3: Στη leading order εικόνα ενός $\gamma + jet$ γεγονότος υπάρχουν δυο παρτόνια ($q\bar{q}$ ή qg τα οποία συγκρούονται μεταξύ τους. Μέσα από αυτή τη διαδικασία μπορεί να παραχθεί ένα φωτόνιο (γ) και παρτόνιο το οποίο αντιστοιχεί σε ένα jet μέσω της διαδικασίας αδρανοποίησης.

τις συγκεκριμένες διαδικασίες βασίζεται στο ότι η εγκάρσια ορμή του φωτονίου που παράγεται από τη διαδικασία είναι κατά βάση αντίθετη από το ανατρέπον (recoiling) jet. Έτσι λόγω της καλύτερης ενεργειακής διακριτικής ικανότητας του ECal (1%) η ενέργεια του jet μπορεί να διορθωθεί γνωρίζοντας την ενέργεια που εναποθέτησε στο ECal το παραγόμενο φωτόνιο. Έτσι το ανακατασκευασμένο jet βαθμονομείται και διορθώνεται ενεργειακά. Αξίζει να σημειωθεί πως με τη μέθοδο αυτή, η ενέργεια του jet μπορεί να βρεθεί με σφάλμα της τάξης του 3 – 10%, αναλόγως με την εγκάρσια ενέργεια του jet.

Στο CMS έχει αναπτυχθεί ένα σύστημα διόρθωσης των jets πολλαπλών επιπέδων. Οι διορθωτικοί παράγοντες λαμβάνονται από προσομοιώσεις βασισμένα σε δεδομένα από test beams, από δεδομένα που προέρχονται κατευθείαν από συγκρούσεις και από προσομοιώσεις συντονισμένες με δεδομένα συγκρούσεων. Οι διορθώσεις αυτές εφαρμόζονται με συγκεκριμένο τρόπο, πχ offset με σκοπό την ενεργειακή διόρθωση και μείωση του pileup, όπως ο αλγόριθμος PUPPI (Pile Up Per Particle Identification) που θα αναλυθεί στην πορεία. Επίσης, οι εκάστοτε διόρθώσεις εφαρμόζονται με σκοπό την σταθεροποίηση της απόκρισης του jet σε συνάρτηση του η_{jet} για μια συγκεκριμένη τιμή του $p_{T_{jet}}$ είτε σε συνάρτηση του $p_{T_{jet}}$

3.3 Pile Up Per Particle Identification Algorithm

Στην ενότητα αυτή γίνεται μια σύντομη αναφορά στον αλγόριθμο Pile Up Per Particle Identification (PUPPI), ο οποίος αποτελεί σημαντικό εργαλείο στις αναλύσεις του CMS τα τελευταία χρόνια καθώς αποσκοπεί όχι απλά στην ενεργειακή διόρθωση διόρθωση και μείωση της επιρροής του φαινομένου pileup αλλά στην αφαίρεση αυτού.

3.3.1 Εισαγωγή

Το φαινόμενο του pileup, που οφείλεται στις αλληλοεπικαλυπτόμενες δευτερογενείς συγκρούσεις στο ίδιο bunch crossing, θα αποτελέσει μείζονα πρόκληση στα υψηλής φωτεινότητας runs του LHC. Για την εξάλειψη του φαινομένου αυτού και την προσπάθεια για την ταυτοποίηση και αφαίρεση των pileup jets στις αναλύσεις του CMS τα τελευταία χρόνια χρησιμοποιούνται διάφοροι αλγόριθμοι. Αναφορικά αυτοί είναι οι Four-vector area subtraction, Grooming, Pileup jet identification, Topological clustering, Charged hadron subtraction (CHS), Cleansing, Constituent subtraction. Οι μέθοδοι αυτοί προχωρούν προοδευτικά από μια global προσέγγιση προς μια πιο τοπική.

Σε γενικές γραμμές, οι μέθοδοι αυτοί χρησιμοποιούν τρία βασικά στοιχεία πληροφοριών για τον εντοπισμό του pileup: την πυκνότητα pileup σε όλο το event, πληροφορίες κορυφής (vertex information) από φορτισμένες διαδρομές και την τοπική κατανομή του pileup σε σχέση με τα σωματίδια από την κορυφαία κορυφή (leading vertex, LV). Καθώς κάθε τεχνική έχει πλεονεκτήματα και μειονεκτήματα, είναι απίθανο μια μεμονωμένη μέθοδος από μόνη της να απομακρύνει βέλτιστα το pileup. Είναι επομένως ζωτικής σημασίας να υπάρχει ένα ευέλικτο πλαίσιο για την ενσωμάτωση των διαφόρων πληροφοριών. Ο αλγόριθμος Pile Up Per Particle Identification (PUPPI) συνδυάζει αυτές τις πληροφορίες. Δηλαδή, αυτή η μέθοδος χρησιμοποιεί τόσο τις global πληροφορίες από το event, όσο και τις τοπικές πληροφορίες για τον εντοπισμό του pileup στο επίπεδο των σωματιδίων. Να σημειωθεί πως ο αλγόριθμος αυτός αποσκοπεί στην απομάκρυνση του pileup, και όχι μόνο στο να διορθώνει τις ποσότητες του jet, με σκοπό να παράγει τελικά μια συνεπή ερμηνεία του event.

Αποτελεί μια τοπική προσέγγιση κατά την οποία δεν πραγματοποιείται clustering αλλά αποδίδεται ένα βάρος σε κάθε σωματίδιο βάσει του οποίου γίνεται ανακατασκευή της τετραορμής του κάθε σωματιδίου. Στην ιδανική περίπτωση, τα σωματίδια που προέρχονται από το pileup θα πάρουν βάρος μηδέν και τα σωματίδια που προέρχονται από τη σκληρή διασπορά θα πάρουν βάρος ενός. Αυτό οδηγεί σε διορθωμένο event, όπου μπορεί κανείς να προχωρήσει στην εύρεση jet χωρίς να χρειάζεται περαιτέρω διόρθωση pileup. Στην πράξη αυτό όμως είναι δύσκολο καθώς η αφαίρεση του pileup είναι αρκετά αποτελεσματική αλλά όχι καθολική και στις περισσότερες περιπτώσεις η ταυτοποίηση του pileup και η περαιτέρω αφαίρεση του είναι απαραίτητη.

3.3.2 Ο αλγόριθμος PUPPI

Για κάθε σωματίδιο ορίζεται ένα τοπικό μέγεθος α , το οποίο στοχεύει στον τοπικό διαχωρισμό της ακτινοβολίας parton-like έναντι pileup-like στη γειτονιά κάθε σωματιδίου. Βασικό στοιχείο που εκμεταλεύεται ο αλγόριθμος για το διαχωρισμό σωματιδίων που προέρχονται από κύρια κορυφή και αυτών που προέρχονται από pileup, είναι το γεγονός πως το φάσμα της εγκάρσιας ορμής (p_T) του pileup μειώνεται πολύ πιο γρήγορα. Στη συνέχεια, το μέγεθος απροσπαθεί να εκμεταλλευτεί πρόσθετες πληροφορίες σχετικά με την p_T του κάθε μεμονωμένου σωματιδίου. Στην περιοχή όπου η ανίχνευση τροχιών είναι εφικτή, εντός του tracker δηλαδή, είναι εφικτή η ταυτοποίηση του αν ένα φορτισμένο σωματίδιο προέρχεται από κύρια κορυφή (LV) ή από pileup κορυφή (pileup vertex, PV). Επομένως, η μέση τιμή και η τιμή RMS του μεγέθους αγια τα φορτισμένα pileup χρησιμοποιείται για τον χαρακτηρισμό της κατανομής pileup σε επίπεδο γεγονότος (event level).

Στη συνέχεια, συγκρίνοντας τη τιμή του μεγέθους ακάθε σωματιδίου που με τη μέση τιμή τη κατανομής του φορτισμένου pileup, αποδίδεται ένα βάρος σε κάθε σωματίδιο. Το βάρος αυτό μπορεί να πάρει τιμές μεταξύ μηδέν και ένα και αποτελεί ένα μέτρο που υποδεικνύει κατά πόσο ένα σωματίδιο επιτρέπεται να συνεισφέρει στο event. Ιδανικά, σωματίδια που προέρχονται από hard processes θα τους αποδοθεί βάρος ίσο με τη μονάδα, ενώ σε σωματίδια pileup, που προέρχονται κυρίως από soft QCD διεργασίες θα τους αποδοθεί βάρος ίσο με μηδέν.

Το μέγεθος α για όλα σχεδόν τα pileup σωματίδια έχει τιμές μέσα σε μερικές τυπικές αποκλίσεις της διάμεσης και επομένως, τους αποδίδονται μικρά βάρη. Από την άλλη, σε σωματίδια για τα οποία οι τιμές του μεγέθους αποκλίνουν κατά πολύ από τη μέση τιμή της κατανομής του φορτισμένου pileup, αποδίδονται μεγάλα βάρη, καθώς αυτό είναι ασύνηθες για pileup. Τέλος, τα βάρη αυτά χρησιμοποιούνται για την ανακατασκευή της τετραορμής του κάθε σωματιδίου. Σωματίδια επομένως με μικρό βάρος ή μικρή ανακατασκευασμένη p_T απορρίπτονται.

Το τοπικό μέγεθος α

Πιο αναλυτικά, το τοπικό μέγεθος αγια κάθε σωματίδιο i ορίζεται ως εξής

$$\alpha_i = \log \sum_{j \in \text{event}} \xi_{ij} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0) \quad (3.4)$$

όπου

$$\xi_{ij} = \frac{p_{T_j}}{\Delta R_{ij}} \quad (3.5)$$

Επίσης, η συντόμηση $\Theta(R_{min} \leq \Delta R_{ij} \leq R_0)$ αντιστοιχεί στη $\Theta(\Delta R_{ij} - R_{min}) \times \Theta(R_0 - \Delta R_{ij})$, όπου Θ είναι η Heaviside βηματική συνάρτηση, ΔR_{ij} είναι η απόσταση μεταξύ των σωματιδίων i και j στον $\eta - \phi$ χώρο, p_{T_j} είναι η εγκάρσια ορμή του σωματιδίου j , R_0 καθορίζει έναν κώνο γύρω από κάθε σωματίδιο i , έτσι ώστε μόνο τα σωματίδια που βρίσκονται μέσα σε αυτόν τον κώνο παίρνουν μέρος στον υπολογισμό του μεγέθους α . Επιπλέον, η απόσταση R_{min} ρυθμίζει τη συλλογή συγγραμμικών

σωματιδίων, δηλαδή σωματίδια που βρίσκονται σε απόσταση μικρότερη από R_{min} από το σωματίδιο i απορρίπτονται και αυτά από το άθροισμα.

Λόγω της συγγραμμικότητας του παρτονικού καταγισμού, ένα σωματίδιο που προέρχεται από hard process είναι πιο πιθανό να βρίσκεται σε κοντινή απόσταση με άλλα σωματίδια που προέρχονται από την ίδια διαδικασία. Επομένως, το μέγεθος α_i για ένα τέτοιο σωματίδιο i έχει υψηλές τιμές. Από την άλλη, αναμένεται ότι τα pileup σωματίδια δεν ακολουθούν κάποια δομή καταγισμού, δεν σχετίζονται με σωματίδια που προέρχονται από LV και άρα μόνο "κατά τύχη" μπορεί να βρίσκονται χωρικά σε σωματίδια που προέρχονται από LV. Αυτό οδηγεί στο συμπέρασμα πως οι τιμές του μεγέθους α_i για ένα τέτοιο σωματίδιο i έχει μικρές τιμές.

Καθώς δεν είμαστε σε θέση να γνωρίζουμε a priori ποια σωματίδια προέρχονται από LV και ποιά από PV, μπορούμε να εκμαιεύσουμε πληροφορίες για τα φορτισμένα σωματίδια στην κεντρική περιοχή του ανιχνευτή ($|\eta| < 3$). Σε αυτήν την περιοχή, έχουμε στη διάθεση μας πληροφορίες για τις τροχιές των φορτισμένων σωματιδίων οι οποίες παρέχουν τη δυνατότητα διάκρισης των τροχιών που προέρχονται από LV και από PV.

Η συσχέτιση των τροχιών με σωματίδια γίνεται με τον αλγόριθμο Particle Flow, ο οποίος περιγράφεται στην ενότητα ?? και συνδυάζει μετρήσεις από τους subdetectors για τον προσδιορισμό μεμονωμένων σωματιδίων. Βάσει του αλγόριθμου PF τα σωματίδια που προσδιορίζονται μπορούν να ταξινομηθούν σε τρεις κατηγορίες: ουδέτερα σωματίδια, φορτισμένα αδρόνια από LV και φορτισμένα αδρόνια από PV. Επομένως, χρησιμοποιώντας τα φορτισμένα σωματίδια από LV ως έναν "αντιπρόσωπο" (proxy) για το σύνολο των σωματιδίων που προέρχονται από LV. Επεκτείνοντας το άθροισμα της εξίσωσης ?? για κάθε περίπτωση σωματιδίου στην κεντρική περιοχή έχουμε

$$\sum_j = \sum_{j \in Ch, PU} + \sum_{j \in Ch, LV} + \sum_{j \in Neutral} \quad (3.6)$$

όπου

- Ch,PU αναφέρεται σε φορτισμένα pileup σωματίδια
- Ch,LV αναφέρεται σε φορτισμένα σωματίδια που προέρχονται από LV
- Neutral αναφέρεται σε όλα τα ουδέτερα σωματίδια (και pileup και αυτά που προέρχονται από LV)

Αυτό οδηγεί στον ορισμό μεγέθους για κάθε περιοχή του ανιχνευτή, την κεντρική (C) και την forward (F), κάνοντας τη γενική υπόθεση πως στη forward περιοχή, όπου δεν υπάρχουν πληροφορίες για την τροχιά των σωματιδίων, ότι όλα τα σωματίδια προέρχονται από LV

$$\alpha_i^C = \log \sum_{j \in Ch, LV} \xi_{ij} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0) \quad (3.7)$$

$$\alpha_i^F = \log \sum_{j \in event} \xi_{ij} \times \Theta(R_{min} \leq \Delta R_{ij} \leq R_0) \equiv \alpha_i \quad (3.8)$$

Ενώ υπάρχουν συμβολές θορύβου από το pileup, αυτές καταστέλλονται σε σχέση με τις συνεισφορές των σωματιδίων από LV λόγω του p_{T_j} στον αριθμητή. Έτσι, ο αλγόριθμος μπορεί να αποδίδει βάρη σε περιοχές όπου δεν υπάρχει δυνατότητα ταυτοποίησης τροχιών.

Στην εικόνα ?? φαίνονται οι κατανομές του μεγέθους για τις δυο περιοχές του ανιχνευτή (α^C , α^F) για σωματίδια που προέρχονται από LV και pileup. Επίσης, από την κατανομή α^C μπορεί να γίνει εμφανές πως εαν ένα σωματίδιο i δεν βρίσκεται σε περιοχή σωματιδίων που προέρχονται από LV, το άθροισμα της εξίσωσης ?? έχει πολύ μικρή τιμή με αποτελέσματα τελικά $\alpha_i \rightarrow -\infty$. Σε αυτή την περίπτωση το σωματίδιο i χαρακτηρίζεται ως pileup και απορρίπτεται από το event. Τέλος, φαίνεται πως το μέγεθος α^C έχει πιο ισχυρή διαχωριστική ικανότητα από το α^F .

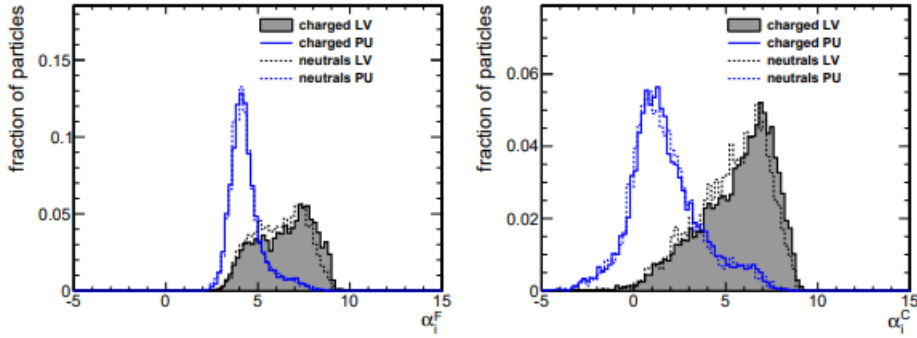


Figure 3.4: Κατανομή του μεγέθους α_i για σωματίδια i που προέρχονται από LV και σωματίδια που προέρχονται από pileup. Αριστερά το μέγεθος α_i^F για την forward περιοχή που υπολογίζεται από την εξίσωση ?? με άθροισμα σε όλα τα σωματίδια απο το LV. Δεξιά το μέγεθος α_i^C για την κεντρική περιοχή που υπολογίζεται από την ?? με άθροισμα σε όλα τα φορτισμένα σωματίδια που προέρχονται από το LV.

Για ένα δεδομένο event τις πληροφορίες προέλευσης (LV ή pileup) των φορτισμένων σωματιδίων υπολογίζονται οι αντίστοιχες κατανομές τα δυο μεγέθη. Όσον

αφορά τα ουδέτερα σωματίδια, για τα οποία δεν έχουμε πληροφορίες για την τροχιά τους, γίνεται η υπόθεση πως ανοίκουν στην κατανομή με τις ίδιες ιδιότητες. Δηλαδή γίνεται η υπόθεση πως οι κατανομές α^C και α^F είναι οι ίδιες για φορτισμένα και ουδέτερα σωματίδια και για τις δυο περιοχές του ανιχνευτή. Οι ποσότητες που χρησιμοποιούνται για τον χαρακτηρισμό των κατανομών σε μια βάση event-by-event, είναι οι μέση τιμή και η τιμή RMS.

$$\begin{aligned}\bar{\alpha}_{PU}^F &= \text{median}\{\alpha_{i \in Ch, PU}^F\} \\ \sigma_{PU}^F &= \text{RMS}\{\alpha_{i \in Ch, PU}^F\} \\ \bar{\alpha}_{PU}^C &= \text{median}\{\alpha_{i \in Ch, PU}^C\} \\ \sigma_{PU}^C &= \text{RMS}\{\alpha_{i \in Ch, PU}^C\}\end{aligned}$$

Αξίζει να σημειωθεί πως καθώς ο υπολογισμός των ποσοτήτων $\bar{\alpha}_{PU}^F$ και σ_{PU}^F πραγματοποιείται για τα φορτισμένα σωματίδια pileup, πρέπει να γίνει στην κεντρική περιοχή, παρά το γεγονός πως αυτά τα μεγέθη χρησιμοποιούνται για τον υπολογισμό των βαρών που θα αποδοθούν στα σωματίδια στην forward περιοχή του ανιχνευτή.

Τα βάρη που αποδίδονται σε κάθε σωματίδιο

Τα βάρη που αποδίδονται σε κάθε σωματίδιο μπορούν να πάρουν τιμές μεταξύ 0 και 1 και στην ιδαική περίπτωση 0 θα πάρουν τα σωματίδια pileup και 1 τα σωματίδια που προέρχονται από LV. Στην πραγματικότητα όμως γίνεται προσπάθεια εκτίμησης της προέλευσης των σωματιδίων βάσει των διαθέσιμων πληροφοριών, επομένως τα βάρη αποτελούν μια συνεχή κλασματική συνάρτηση.

Ορίζεται η ποσότητα

$$\chi_i^2 = \Theta(\alpha_i - \bar{\alpha}_{PU}) \times \frac{\alpha_i - \bar{\alpha}_{PU}}{\sigma_{PU}^2} \quad (3.9)$$

η οποία υπολογίζει τις διακυμάνσεις του μεγέθους α_i από την μέση τιμή της κατανομής του pileup. Διακυμάνσεις υπό της μέσης τιμής υποδεικνύουν pileup και επομένως τους αποδίδεται βάρος κοντά στο μηδέν. Αντίθετα μεγάλες διακυμάνσεις πάνω από τη μέση τιμή υποδεικνύουν σωματίδιο που προέρχεται από LV και επομένως του αποδίδεται βάρος κοντά στη μονάδα. Οποιαδήποτε ενδιάμεση κατάσταση διακύμανσης το βάρος θα έχει τιμή μεταξύ 0 και 1. Στην εξίσωση ?? οπουδήποτε μπορεί να χρησιμοποιηθούν οι μεταβλητές επισυμμεσμένες με C προτιμούνται έναντι των αντίστοιχων με F.

Από την εικόνα ?? η κατανομή του pileup φαίνεται χοντρικά να είναι μια Gaussian. Για το λόγο αυτό η εξίσωση ?? μοιάζει με την κατανομή $\chi_{NDF=1}^2$. Τα βάρη που αποδίδονται στα σωματίδια δίνονται απο τη σχέση

$$w_i = F_{\chi_{NDF=1}^2}(\chi_i^2) \quad (3.10)$$

όπου F_{χ^2} είναι η αθροιστική συνάρτηση της κατανομής χ^2 . Ορίζοντας μια τιμή κατωφλίου (threshold) στο βάρος w_{cut} είναι δυνατή η απόρριψη του pileup

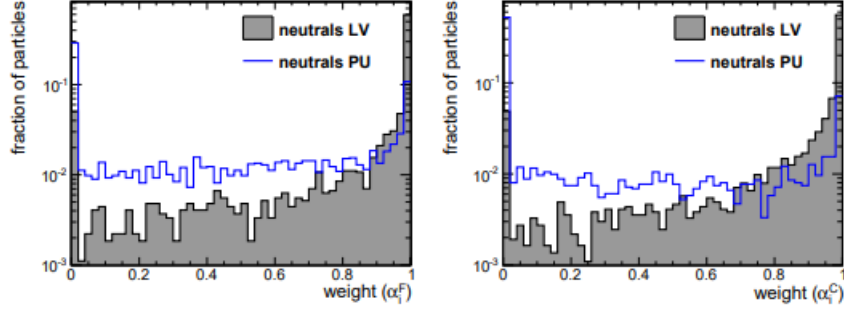


Figure 3.5: Κατανομή του βάρους που αποδίδεται σε κάθε ουδέτερο σωματίδιο που προέρχεται από LV ή pileup. Αριστερά τα βάρη έχουν υπολογιστεί από τη σχέση ?? χρησιμοποιώντας το μέγεθος α_i^F ενώ δεξιά χρησιμοποιώντας το μέγεθος α_i^C

από το event. Στην εικόνα ?? φαίνονται οι κατανομές του βάρους για τα ουδέτερα σωματίδια που έχουν υπολογιστεί από τη σχέση ?? χρησιμοποιώντας τα μεγέθη α_i^F και α_i^C .

Τέλος, τα βάρη χρησιμοποιούνται για την ανακατασκευή της τετραομής του σωματιδίου i , $p_i^{mu} \rightarrow w_i \times p_i^{mu}$. Άρα σωματίδια με μικρό βάρος, $w_i < w_{cut}$ ή ανακατασκευασμένη p_T μικρότερη από την τιμή κατωφλίου που προσδιορίζεται βάσει του w_{cut} , δηλαδή $p_T < p_{Tcut}$ απορρίπτονται και δε συμβάλλουν στο event.

Το εναπομείναν σύνολο ανακατασκευασμένων σωματιδίων αποτελεί ένα event διορθωμένο από το pileup.

Κεφάλαιο 4

Πολυπαραγωγντική Ταξινόμηση στη Μηχανική Εκμάθηση

Στη Φυσική Υψηλών Ενεργειών, η αναζήτηση για ολοένα και μικρότερα σήματα σε ολοένα και μεγαλύτερα σύνολα δεδομένων καθιστά ύψιστης σημασίας την εξαγωγή της μέγιστης δυνατής διαθέσιμης πληροφορίας από το σύνολο των δεδομένων. Οι πολυπαραγωγντικές μέθοδοι ταξινόμησης (multivariate classification) που βασίζονται στις τεχνικές μηχανικής εκμάθησης (Machine Learning), αποτελούν πλέον σημαντικό συστατικό στις περισσότερες αναλύσεις. Επίσης, οι ίδιοι οι πολυπαραγωγντικοί ταξινομητές έχουν σημειώσει σημαντική εξέλιξη τα τελευταία χρόνια. Οι επιστήμονες που ασχολούνται με την ανάλυση των δεδομένων έχουν βρει νέους τρόπους να συνδυάζουν τους ταξινομητές, με αποτέλεσμα να έχουν τη δυνατότητα να βελτιώσουν περαιτέρω τις επιδόσεις τους.

Η ανάλυση που πραγματοποιήθηκε στα πλαίσια αυτής της διπλωματικής εργασίας, η οποία θα αναλυθεί στο επόμενο κεφάλαιο, βασίστηκε στην τεχνική της πολυπαραγωγντικής ταξινόμησης, που ανήκει στην κατηγορία της επιβλεπόμενης μηχανικής εκμάθησης. Πιο συγκεκριμένα, χρησιμοποιήθηκε ένα ενδυναμομένο δέντρο απόφασης (Boosted Decision Tree, BDT). Στην ενότητα αυτή γίνεται μια σύντομη αναφορά στις τεχνικές μηχανικής εκμάθησης και τις τεχνικές ταξινόμησης. Τέλος, περιγράφεται το μοντέλο του ταξινομητή BDT.

4.1 Μηχανική Εκμάθηση - Machine Learning

Οι αλγόριθμοι μηχανικής εκμάθησης χρησιμοποιούν υπολογιστικές μεθόδους για να λάβουν και να μάθουν τις πληροφορίες απευθείας από τα δεδομένα χωρίς να στηρίζονται σε μια προκαθορισμένη εξίσωση ως μοντέλο. Με την ολοένα αυξανόμενη ανάγκη για αντιμετώπιση προβλημάτων μεγάλων δεδομένων, η μηχανική μάθηση έχει καταστεί βασική τεχνική για την επίλυση τους. Μια βασική ικανότητα των αλγόριθμων μηχανικής εκμάθησης είναι το γεγονός πως βελτιώνεται προσαρμοστικά η απόδοσή τους, καθώς αυξάνεται ο αριθμός των διαθέσιμων δειγμάτων προς εκ-

παίδευση. Επίσης, για τη λήψη καλύτερων αποφάσεων και τη διεξαγωγή ακριβέστερων προβλέψεων, οι αλγόριθμοι αυτοί βρίσκουν φυσικά μοτίβα στα δεδομένα.

Με βάση τον τρόπο εκμάθησης του αλγορίθμου μηχανικής εκμάθησης υπάρχουν τρεις γενικές κατηγορίες. Η πρώτη κατηγορία είναι η **επιβλεπόμενη εκμάθηση (Supervised Learning)**, όπου ένα μοντέλο εκπαιδεύεται σε γνωστά αποτελέσματα εισόδου και εξόδου, έτσι ώστε να μπορεί να παράγει λογικές προβλέψεις στο μέλλον για την απάντηση σε νέα δεδομένα. Δηλαδή, στο σύνολο δεδομένων περιέχονται τα στοιχεία εισόδου αλλά και τα επιθυμητά αποτελέσματα για αυτά. Τα αποτελέσματα επομένως έχουν τον ρόλο του “δασκάλου”, δηλαδή συντελούν στην εύρεση ενός γενικού κανόνα αντιστοίχισης εισόδων με αποτελέσματα. Η δεύτερη κατηγορία είναι η **μη επιβλεπόμενη εκμάθηση (Unsupervised Learning)**, η οποία βρίσκει κρυμμένα μοτίβα ή εγγενείς δομές στα δεδομένα εισόδου. Σε αυτήν την περίπτωση το σύνολο δεδομένων δεν περιέχει απαντήσεις, δηλαδή δεν υπάρχει εμπειρία στον αλγόριθμο. Τέλος, η τρίτη κατηγορία είναι η **ενισχυτική εκμάθηση (Reinforcement Learning)**, όπου περιλαμβάνεται η αλληλεπίδραση του αλγορίθμου με ένα δυναμικό περιβάλλον, με αποτέλεσμα να έχουμε συνεχώς νέα δεδομένα χωρίς να είναι σαφές εάν ο αλγόριθμος προσεγγίζει τον στόχο του.

Για την ανάπτυξη των προγνωστικών μοντέλων τα οποία κάνουν προβλέψεις βασισμένα σε γνωστά στοιχεία με την ύπαρξη κάποιας αβεβαιότητας ένας αλγόριθμος επιβλεπόμενης μηχανικής εκμάθησης χρησιμοποιεί τεχνικές ταξινόμησης (classification) ή τεχνικές παλινδρόμησης (regression).

Οι τεχνικές ταξινόμησης στοχεύουν στην εύρεση της βέλτιστης δυνατής διαχωριστικής επιφάνειας διαχωρισμού σε ένα πολυδιάστατο χώρο χαρακτηριστικών. Οι απαντήσεις που προβλέπουν είναι διακριτές και για αυτό το λόγο τα δεδομένα που αποτελούν την είσοδο του αλγορίθμου πρέπει να μπορούν να επισημανθούν, να κατηγοριοποιηθούν ή να διαχωριστούν σε συγκεκριμένες ομάδες ή κλάσεις. Οι αλγόριθμοι k-nearest neighbors (kNN), support vector machine (SVM), Naïve Bayes, καθώς και τα ενδυναμωμένα δέντρα αποφάσεων (Boosted Decision Trees BDT) και τα νευρωνικά δίκτυα (Neural Networks NN) ανήκουν στην κατηγορία των αλγορίθμων που εφαρμόζουν την τεχνική της ταξινόμησης. Από την άλλη, οι τεχνικές παλινδρόμησης προσαρμόζουν το μοντέλο σε ένα σύνολο μετρήσεων και οι απαντήσεις που προβλέπουν είναι ένα συνεχές σύνολο. Στην κατηγορία των αλγορίθμων που εφαρμόζουν την τεχνική της παλινδρόμησης ανήκουν το γραμμικό και μη γραμμικό μοντέλο, τα ενδυναμωμένα δέντρα αποφάσεων (BDT) και τα νευρωνικά δίκτυα (NN).

Η μη επιβλεπόμενη μάθηση βρίσκει κρυμμένα μοτίβα στα δεδομένα. Χρησιμοποιείται για την εξαγωγή συμπερασμάτων από σύνολα δεδομένων εισόδου χωρίς προκαθορισμένες απαντήσεις. Η πιο κοινή τεχνική εκμάθησης χωρίς επίβλεψη είναι η τεχνική clustering, η οποία χρησιμοποιείται για τη διερευνητική ανάλυση δεδομένων και κάνει ανακατασκευή δομών σε ένα πολυδιάστατο χώρο χαρακτηριστικών.

Για την ανάλυση των δεδομένων της συγκεκριμένης διπλωματική εργασίας έγινε εκπαίδευση ενός ταξινομητή και πιο συγκεκριμένα ενός ενδυναμωμένου δέντρου απόφασης (BDT), που αναλύεται στην επόμενη ενότητα.

4.2 Δέντρα απόφασης - Decision Trees

Πριν γίνει η ανάλυση των ενδυναμωμένων δέντρων απόφασης (boosted Decision Trees) και της μεθόδου της ενδυνάμωσης (boosting) θα γίνει μια σύντομη αναφορά στους απλούς ταξινομητές δέντρα απόφασης (Decision Trees), που όπως αναφέρθηκε ανήκουν στην κατηγορία της επιβλεπόμενης μηχανικής εκμάθησης και στην οικογένεια των μη γραμμικών ταξινομητών. Επίσης αποτελούν σημαντικό εργαλείο στις αναλύσεις που γίνονται στον τομέα της Φυσικής Υψηλών Ενεργειών για το διαχωρισμό του σήματος από το υπόβαθρο.

Συνοπτικά, ένα δυαδικό δέντρο απόφασης είναι ένα πολυεπίπεδο σύστημα που αποτελείται από μία αλληλουχία δυαδικών αποφάσεων, που τεμαχίζει το χώρο των χαρακτηριστικών σε άνισα κουτιά, στις κλάσεις-περιοχές R_i . Σε κάθε στάδιο επιλέγεται ο επόμενος κλάδος να χωριστεί, έτσι ώστε να δώσει την καλύτερη αύξηση της ποιότητας του διαχωρισμού.

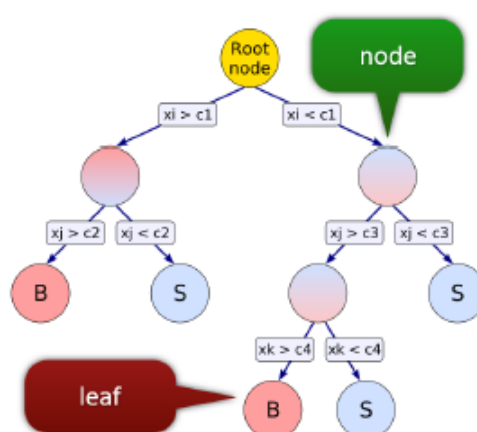


Figure 4.1: Ένα δέντρο απόφασης είναι ένα πολυεπίπεδο σύστημα που αποτελείται από μία αλληλουχία δυαδικών αποφάσεων. Έχοντας σαν αφετηρία κάθε μεταβλητή και την τιμή διαίρεσης που δίνει τον καλύτερο διαχωρισμό, το αρχικό δείγμα γεγονότων σε έναν κόμβο (node), χωρίζεται σε δύο μέρη που ονομάζονται κλάδοι. Αυτή η διαδικασία επαναλαμβάνεται για κάθε κλάδο δημιουργώντας έτσι νέους κόμβους και νέους κλάδους. Η διαδικασία επαναλαμβάνεται μέχρι να υπάρξει ένας συγκεκριμένος αριθμός τελικών κλάδων, τα λεγόμενα φύλλα (leaf), ή μέχρι το κάθε φύλλο να αποτελεί καθαρό σήμα ή καθαρό υπόβαθρο ή να έχει λίγα γεγονότα για να συνεχιστεί η διαδικασία.

Για το πρόβλημα ταξινόμησης δυο κλάσεων που αντιστοιχούν σε σήμα και υπόβαθρο τα δεδομένα που προέρχονται από προσομοίωση, χωρίζονται στη μέση σε δύο μέρη. Το ένα θα χρησιμοποιηθεί για την εκπαίδευση και το άλλο για

τη μετέπειτα εκπαίδευση. Για κάθε γεγονός, υπάρχει ένας αριθμός μεταβλητών ταυτοποίησης σωματιδίων (particle identification -PID), που είναι χρήσιμες για το διαχωρισμό μεταξύ των δύο κλάσεων. Για κάθε μια από αυτές τις μεταβλητές τα γεγονότα ταξινομούνται με βάση την τιμή της μεταβλητής, επιλέγοντας ως τιμή διαίρεσης αυτή που δίνει τον καλύτερο δυνατό διαχωρισμό, έχοντας στη μία πλευρά ως κυρίως σήμα και στην άλλη κυρίως υπόβαθρο. Η διαδικασία επαναλαμβάνεται διαδοχικά για κάθε μεταβλητή.

Επομένως, έχοντας σαν αφετηρία κάθε μεταβλητή και την τιμή διαίρεσης που δίνει τον καλύτερο διαχωρισμό, το αρχικό δείγμα γεγονότων σε έναν κόμβο, χωρίζεται σε δύο μέρη που ονομάζονται κλάδοι. Αυτή η διαδικασία επαναλαμβάνεται για κάθε κλάδο δημιουργώντας έτσι νέους κόμβους και νέους κλάδους. Η διαδικασία επαναλαμβάνεται μέχρι να υπάρξει ένας συγκεκριμένος αριθμός τελικών κλάδων, τα λεγόμενα φύλλα, ή μέχρι το κάθε φύλλο να αποτελεί καθαρό σήμα ή καθαρό υπόβαθρο ή να έχει λίγα γεγονότα για να συνεχιστεί η διαδικασία.

Σκοπός ενός δυαδικού δέντρου απόφασης είναι η αύξηση της ομοιογένειας των υποσυνόλων που δημιουργούνται. Η ομοιογένεια αποτελεί ένα μέτρο της καθαρότητας του κάθε κόμβου και περιγράφεται από δυο συναρτήσεις από τις οποίες γίνεται επιλογή της κατάλληλης σε κάθε εκπαίδευση.

- Gini index: $I = p(1 - p)$
- Cross Entropy: $I = -p \ln p - (1 - p) \ln(1 - p)$

Στις σχέσεις αυτές ως p ορίζεται η πιθανότητα ενός υποσυνόλου να ανήκει στην κλάση-περιοχή R_i και είναι ίση με $\frac{N_i}{N}$, όπου N_i το σύνολο των διανυσμάτων εισαγωγής στον κόμβο i ενώ N το σύνολο των δεδομένων.

Το κριτήριο για τον καθορισμό της ποιότητας του διαχωρισμού είναι:

$$\Delta I = I - \frac{N_1}{N} I_1 - \frac{N_2}{N} I_2 \quad (4.1)$$

Εάν ένα φύλλο έχει καθαρότητα μεγαλύτερη από $\frac{1}{2}$ (ή ό, τι έχει οριστεί), τότε καλείται ένα φύλλο σήματος και αν η καθαρότητα είναι μικρότερη από $\frac{1}{2}$ είναι ένα φύλλο υποβάθρου.

Στα πλεονεκτήματα του Decision Tree είναι η διαφάνεια της διαδικασίας απόφασης και ότι ως ταξινομητής είναι αδιάφορος ως προς ασθενείς μεταβλητές. Έχει όμως και ένα σημαντικό μειονέκτημα, το ότι είναι ευαίσθητο σε στατιστικές διακυμάνσεις του δείγματος εκπαίδευσης. Για την απαλλαγή από τυχαίες διακυμάνσεις του συνόλου εκπαίδευσης, γίνεται επιλογή των ενδυναμωμένων δέντρων απόφασης (Boosted Decision Trees), τα οποία αναλύονται στην επόμενη ενότητα.

4.3 Η τεχνική ενδυνάμωσης (Boosting)

Η τεχνική της ενδυνάμωσης (boosting) είναι μια ισχυρή μέθοδος που προκύπτει από τον συνδυασμό πολλών απλών και συχνά ασθενών ταξινομητών με σκοπό να παράγει ένα καλύτερο αποτέλεσμα. Η πιο ευρέως διαδεδομένη μορφή της τεχνικής αυτής είναι ο αλγόριθμος Adaboost. Για την επεξήγηση της τεχνικής αυτής θα χρησιμοποιηθεί ως παράδειγμα ένα πρόβλημα ταξινόμησης με εκπαίδευση ενός ταξινομητή τύπου δέντρο απόφασης, εφόσον αυτή η τεχνική εφαρμόστηκε στα πλαίσια αυτή της διπλωματικής εργασίας.

Σε ένα πρόβλημα ταξινόμησης με εκπαίδευση ενός ταξινομητή τύπου δέντρο απόφασης, σκοπός της τεχνικής boosting είναι η τροποποίηση του δέντρου απόφασης ώστε να επιτευχθεί απαλλαγή από τυχαίες διακυμάνσεις του συνόλου εκπαίδευσης. Αυτό επιτυγχάνεται με την κατασκευή ενός “δάσους” από δέντρα, τα οποία εκπαιδεύτηκαν από τα ίδια δεδομένα με διαφορετικά βάρη κάθε φορά.

4.3.1 Ενδυναμωμένα Δέντρα Απόφασης - Boosted Decision Trees

Τα ενδυναμωμένα δέντρα απόφασης αποτελούν μια αλληλουχία πολλών ασθενών και μικρού βάθους, δηλαδή με λίγες διακλαδώσεις, δέντρων απόφασης. Η λογική είναι ότι χρησιμοποιούνται τα ίδια δεδομένα για την εκπαίδευση κάθε ενός από αυτά τα δέντρα αλλά με διαφορετικό βάρος κάθε φορά.

Ο αλγόριθμος Adaptive Boost (AdaBoost)

Ο πιο διαδεδομένος αλγόριθμος ενδυνάμωσης (boosting) είναι ο αλγόριθμος προσαρμοστικής ενδυνάμωσης, Adaptive Boost (AdaBoost). Σε ένα πρόβλημα ταξινόμησης, τα γεγονότα τα οποία ταξινομήθηκαν λάθος κατά την εκπαίδευση ενός δέντρου απόφασης, λαμβάνουν μεγαλύτερο βάρος κατά την εκπαίδευση του επόμενου.

Ξεκινώντας από τα αρχικά βάρη των events κατά την εκπαίδευση του πρώτου δέντρου απόφασης, τα επόμενα δέντρα απόφασης εκπαιδεύονται χρησιμοποιώντας ένα τροποποιημένο δείγμα από events, όπου τα βάρη των events που είχαν ταξινομηθεί λάθος πολλαπλασιάζονται με ένα απλό boost βάρος α_i .

$$\alpha_i = \frac{1 - err}{err} \quad (4.2)$$

όπου err είναι ο ρυθμός λάθος ταξινόμησης του προηγούμενου δέντρου απόφασης. Το βάρος ολόκληρου του event sample στη συνέχεια επανκανονικοποιείται ώστε το άθροισμα των βαρών να παραμένει σταθερό.

Έστω $\varphi(x; \theta_k)$ το αποτέλεσμα ενός μεμονωμένου ταξινομητή, όπου x είναι το διάνυσμα εκπαίδευσης που αποτελείται από χαρακτηριστικές μεταβλητές και θ_k ένα διάνυσμα παραμέτρων. Στόχος είναι η κατασκευή ενός άλλου ισχυρού ταξινομητή

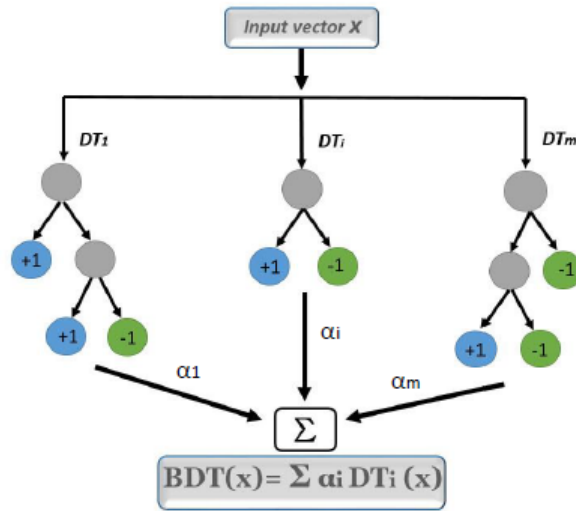


Figure 4.2: Τα ενδυναμωμένα δέντρα απόφασης αποτελούν μια αλληλουχία πολλών ασθενών και μικρού βάθους, δηλαδή με λίγες διακλαδώσεις, δέντρων απόφασης. Χρησιμοποιούνται τα ίδια δεδομένα για την εκπαίδευση κάθε ενός από αυτά τα δέντρα αλλά με διαφορετικό βάρος κάθε φορά. Το αποτέλεσμα είναι ο βεβαρυμένος μέσος όρος όλων των επιμέρους δέντρων απόφασης.

$f(x) = \text{sign}(F(x))$, όπου η συνάρτηση $F(x)$ δίνεται ως εξής:

$$F(x) = \sum_k \alpha_k \varphi_k(x; \theta_k) \quad (4.3)$$

που αποτελεί το βεβαρυμένο μέσο όρο όλων των επιμέρους ταξινομητών. Το αποτέλεσμα κάθε επιμέρους ταξινομητή λαμβάνει την τιμή $y_i = +1$ όταν έχει ταξινομηθεί ως σήμα και $y_i = -1$ όταν έχει ταξινομηθεί ως background σε κάθε βήμα εκπαίδευσης. Ο αλγόριθμος AdaBoost έχει την εξής συνάρτηση κόστους, για τον ταξινομητή k :

$$J_k = \sum_{i=1}^N \exp - [y_{k_i} F_k(x_i)] \quad (4.4)$$

Από τη συγκεκριμένη συνάρτηση κόστους φαίνεται πως στα λάθος ταξινομημένα διανύσματα ($y_i F(x_i) < 0$) επιβάλλεται μεγαλύτερη ποινή (penalty) από αυτά που ταξινομούνται σωστά ($y_i F(x_i) > 0$). Έτσι κάθε φορά υπάρχει μείωση του συνολικού σφάλματος, αν και παραμένουν ασθενείς ταξινομητές καθώς υπάρχει η πιθανότητα να κάνουν ένα νέο σφάλμα. Ακόμα και το τελευταίο δέντρο παραμένει ασθενής ταξινομητής γιατί διαμερίζει το φασικό χώρο αδρά. Το ότι μειώνεται το σφάλμα δεν σημαίνει ότι κάποιο δέντρο από μόνο του δίνει σωστό αποτέλεσμα, χρειάζεται το συνδυασμό όλων των επιμέρους δέντρων με τα κατάλληλα βάρη.

Η διαδικασία που ακολουθείται για τη βελτιστοποίηση αυτής της συνάρτησης είναι αυτή που ακολουθεί. Αρχικά ορίζεται η βοηθητική συνάρτηση του μερικού αθροίσματος, καθώς η απευθείας βελτιστοποίηση της συνάρτησης κόστους δεν είναι εύκολη. Η βοηθητική συνάρτηση μερικού αθροίσματος F_m είναι η εξής:

$$F_m = \sum_{k=1}^m \alpha_k \varphi(x; \theta_k) \quad (4.5)$$

οπότε προκύπτει η αναδρομική σχέση:

$$F_m = F_{m-1} + \alpha_m \varphi(x; \theta_k) \quad (4.6)$$

Επομένως η αναδρομική συνάρτηση κόστους προκύπτει πως είναι η εξής

$$\begin{aligned} J_m &= \sum_{i=1}^N \exp[-y_i(F_{m-1}(x) + \alpha_m \varphi(x; \theta_k))] \Rightarrow \\ J_m &= \sum_{i=1}^N w_i^{(m)} \exp[\alpha_m \varphi(x; \theta_k)] \end{aligned} \quad (4.7)$$

όπου

$$w_i^{(m)} = \exp[-y_i(F_{m-1}(x))] \quad (4.8)$$

το βάρος που αποκτά κάθε διάνυσμα εκπαίδευσης, το οποίο σχετίζεται με το αποτέλεσμα της ταξινόμησης στο προηγούμενο βήμα. Η εκπαίδευση του ταξινομητή στο βήμα m ισοδυναμεί με την εύρεση των παραμέτρων θ_k που ισοδυναμεί με την ελαχιστοποίηση του σφάλματος ταξινόμησης P_m . Το βάρος του δινύσματος εκπαίδευσης που αντιστοιχεί στο βήμα m εκφρασμένο με την πιθανότητα λάθος και σωστής ταξινόμησης αντίστοιχα είναι

$$\begin{aligned} P_m &= \sum_{y_i \varphi(x; \theta_k) < 0} w_i^{(m)} \\ 1 - P_m &= \sum_{y_i \varphi(x; \theta_k) > 0} w_i^{(m)} \end{aligned} \quad (4.9)$$

Επίσης απαιτείται τα βάρη να είναι κανονικοποιημένα σε κάθε βήμα m για να ερμηνευτούν ως πιθανότητες. Η συνάρτηση κόστους γίνεται

$$J_m = e^{-\alpha_m} (1 - P_m) + e^{\alpha_m} P_m \quad (4.10)$$

Η οποία ελαχιστοποιείται για

$$\alpha_m = \frac{1}{2} \ln\left(\frac{1 - P_m}{P_m}\right) \quad (4.11)$$

Τα βάρη w_i^{m+1} , τα βάρη δηλαδή για το επόμενο βήμα υπολογίζονται χρησιμοποιώντας τις ελάχιστες τιμές των πιθανοτήτων P_m και των α_m .

Συνοψίζοντας, ο αλγόριθμος AdaBoost είναι πιο αποδοτικός σε αδύναμους ταξινομητές, δηλαδή μικρά (με 2 ή 3 διακλαδώσεις δηλαδή) μεμονωμένα δέντρα απόφασης, τα οποία δεν έχουν ικανοποιητική διαχωριστική ικανότητα από μόνα τους. Η μέθοδος ταξινόμησης με ενδυναμωμένα δέντρα απόφασης είναι ισχυρή και αποτελεσματική. Σε γενικές γραμμές είναι γρήγορη και δεν απαιτεί τεράστια υπολογιστική ισχύ. Τέλος, όπως αναφέρθηκε δεν επηρεάζεται από στατιστικές διακυμάνσεις, όμως το μεγαλύτερο πλεονέκτημα του ενδυναμωμένου δέντρου απόφασης είναι ότι δεν υπάρχει κίνδυνος για υπερεκπαίδευση (overtraining), αφού το διαχωριστικό σύνορο είναι απλό.

Κεφάλαιο 5

Analysis

5.1 Introduction

In the high pileup environment a phenomenon is present whereby pileup can increase the incidence of jets by several large factors. To reduce the incidence of jets from pileup and to preserve the rate of good jets, a jet identification based on both vertex information and jet shape information has been developed. The construction of this jet identifier is described and the performances are evaluated using both $Z + jets$ MC simulated samples and $Z + jets$ data collected in the 2017 $\sqrt{s} = 13TeV$ run.

The current running of the large hadron collider (LHC) is at such high intensities that multiple proton-proton collisions per bunch interaction occur with high likelihood. In this instance, one is typically concerned about identifying and reconstructing a single primary collision where a physics event of interest occurs amongst the background of the additional proton-proton collisions. Such backgrounds are due to processes that occur with high likelihood like low p_T jet production. These additional collisions are known as pileup.

In the current CMS detector, some of the sub-detectors also read data in an extended window about the time of the current collision. This allows for pileup from both previous and following proton bunches to affect the reconstructed event. This effect is known as out-of-time pileup (as opposed to in-time-pileup). The influence of out-of-time pileup on the event is much smaller. In this analysis both effects are combined and referred to generically as pileup.

The origin of pileup deposits are varied, however most pileup jets are built from low p_T QCD jet production resulting from pileup collisions. This implies that the pileup itself is clustered. Additionally, it is known from extrapolations of the inclusive jet cross sections down to low p_T that a single jet with a $p_T > 5$ GeV occurs with nearly every collision. Such a large incidence of low p_T jets induces a phenomenon whereby the low p_T jets combine to form one single high p_T jet. The resulting jet formed from overlapping jets is known as a pileup jet.

5.1.1 Incidence of Pileup Jets

Consider a numerical model for the rate of two overlapping jets. The probability of two overlapping jets with added total p_T give by p_T , while integrating over both η and ϕ , can be written by

$$p(\text{overlap}|p_T) = N_{pu}(N_{pu} - 1)\alpha_{jet}^2 \int_0^{p_T} dp'_T \frac{d\sigma}{dp_T}(p'_T) \frac{d\sigma}{dp_T}(p_T - p'_T) \quad (5.1)$$

where α_{jet} represents the area of a jet, N_{pu} is the number of pileup events, and the rightmost integral represents the convolution of the inclusive differential cross section as a function of p_T for two sub-jets having p_T values of p'_T and $p_T - p'_T$. The measured cross section can be expressed in the form of a falling exponential as

$$\frac{d\sigma}{dp_T}(p'_T) = \frac{A}{p_T^5} \quad (5.2)$$

where the term A is a constant roughly equal to 300 mb. Expanding out the full form of the convolution integral numerically gives an expression of the form

$$p(\text{overlap}|p_T) = N_{pu}^2 \alpha_{jet}^2 \frac{A^2}{p_T^{6.2}} \quad (5.3)$$

The key result from this calculation is that the rate of overlapping jets grows quadratically with pileup. If one considers the rate of three overlapping jets or more, this rate grows even more rapidly with pileup. Taking the full form of the convolution, the p_T distribution falls more rapidly than the inclusive p_T spectrum, making it such that for higher p_T objects the rate of overlapping pileup is small. However, the fact that overlapping jets combine to make a larger jet with the equivalent sum p_T of all the internal jets allows for a mechanism of pileup jets which can lead to large p_T pileup jets. One last observation is that the rate of overlapping jets grows quadratically with the area of the jet cone size. Reducing the area would thus allow for a smaller incidence of pileup jets.

Due to the fact that pileup jets primarily come from overlapping jets incurred during pileup interactions, pileup jets exhibit two characteristic features: they are both diffuse and, where charged particle identification is possible, some fraction of the charged particles will not point to the primary vertex. These characteristics allow for the identification of pileup jets in both regions where charged particle tracking is available and regions where jet shape identification is possible. Both vertex and shape information are combined through a multivariate analysis technique, to give a single discriminator targeting the identification of pileup jets. This technique is known as the pileup jet identification.

5.2 Data Samples and Object Definition

The analysis is performed using samples of $Z + jets$ events, with the Z boson decaying to muons. This allows for a clean definition of the recoiling pT, for which jets can be balanced against. $Z \rightarrow \mu\mu$ events are selected by requiring two isolated muons ($\mu^+\mu^-$) with $p_T > 20\text{GeV}$ and $|\eta| < 5$, with an invariant mass in a window of 30GeV around the nominal Z mass. So the invariant $mass_{\mu^+\mu^-}$ must be between 70 and 110GeV. The muon isolation is computed as the sum of the transverse energy of the particles inside a cone of radius $\Delta R = 0.3$ around the muon direction divided by the muon transverse momentum. Then a correction for the pileup contribution¹ to the energy inside the cone is applied.

In this analysis the method that is used to mitigate the effect of pileup is the method pileup per particle identification (PUPPI) which attempts to use local shape information, event pileup properties and tracking information together to mitigate the effect of pileup on jet observables. PUPPI thus operates at the PF candidate level, before any jet clustering is performed. A local variable α is computed which contrasts the collinear structure of QCD with the soft diffuse radiation coming from pileup interactions. The α variable is used to calculate a weight which encodes the probability that an individual particle originates from a pileup collision. These per-particle weights are used to rescale the particles four-momenta to correct for pileup, superseding the need for jet-based pileup corrections.

After applying the PUPPI algorithm for pileup mitigation jets are reconstructed using the CMS Particle Flow (PF) algorithm. This algorithm reconstructs and identifies single particles produced in a collision with an optimized combination of all subdetector information. The particles are classified into mutually exclusive categories: charged hadrons, photons, neutral hadrons, muons, and electrons. These objects are then clustered into jets with the anti-kT algorithm with a distance parameter $R = 0.5$. Jet energy corrections are applied to account for the non-linear response of the calorimeters to the particle energies and other instrumental effects. In this analysis jets with $p_T > 20\text{GeV}$ and $|\eta| < 5$ are considered. In the following figures there are the distribution of the transverse momentum of jet for $p_T > 20\text{GeV}$ and the distribution of jet η .

5.3 Pileup Jet Identification Algorithm

The identification of pileup jets is based on two observations:

1. The majority of tracks associated with pileup jets come from non-primary vertices.
2. Pileup jets originate from overlapping particles from pileup collisions and

¹multiple interactions in the same bunch crossing

therefore tend to be more broad and diffuse than jets originating from one single quark or gluon from the hard scatter.

Pileup jet identification relies on two distinct classes of variables:

- vertexing related variables (track-based)
- shape related variables

Track-based variables are defined based on observation (1), while jet shape observables are defined based on observation (2).

Charged PF candidates with tracks contribute to roughly half of the total pileup. Two thirds of the pileup in the tracker volume is charged, the other half of the pileup originates from either neutral candidates from charged particles which are outside of the tracker volume or true neutral candidates where no track is linked. Inside or near the tracker volume a distinct enhancement in the ability to discriminate against pileup is possible by exploiting the compatibility of the jet tracks to come from the PV. Outside the tracker volume, this use of vertexing is not possible, thus jet shower shapes are the only handle to distinguish pileup jets. Since characteristically overlapping pileup jets tend to result in wider jets, shape related variables are precisely designed to target the diffuseness of a jet.

To perform the identification of pileup jets fourteen distinct variables, are combined in a boosted decision tree (BDT) yielding a single discriminator which can be cut on to give jets of varying pileup contamination. The training of the BDT and optimization of the jet id working points are done separately in two regions corresponding to the two regions of CMS detector, the central region where are the tracker and the calorimeters with $|\eta| < 3$ and the forward region with $3 < |\eta| < 5$

5.3.1 Input Variables

The track and shape based variables that were used in training of the BDT are:

- β : fraction of tracks associated to the primary vertex.

The variable β is a track based variable and is defined as the sum of p_T of all PF charged candidates originating from the LV divided by the sum of the p_T of all charged candidates in the jet

$$\beta = \frac{\sum_{i \in LV} p_{Ti}}{\sum_i p_{Ti}} \quad (5.4)$$

To be identified as coming from the LV, the charged PF candidate must have a $|\Delta Z| < 0.2cm$ where ΔZ is the distance with respect to the LV along the z axis.

- manW and maxW: major and minor axis of the jet ellipsoid in the $\eta - \phi$ space.

- $A < (\Delta R) < A + 0.1$ (frac01-04) consist in the fractional energy deposits in four annuli (concentered rings) about the jet axis, where A is in the 0.1 intervals from 0 to 0.3 about the jet cone axis.

$$A < (\Delta R) < A + 0.1 = \frac{1}{p_T^{jet} \sum_{i \in A < (\Delta R) < A + 0.1} p_{T_i}} \quad (5.5)$$

- nParticles: number of PF candidates
- nCharged: number of charged PF candidates

The charged and neutral multiplicities, $N_{charged}$ and $N_{neutrals}$, are also added to the pileup jet id so as to play the dual role of separately enhancing the quark versus pileup and gluon versus pileup separation by allowing for splitting of quarks and gluons into categories and also by further enhancing the pileup separation.

- ptD: p_T -weighted average p_T of constituents.

The variable ptD is considered in the construction of the pileup jet id to enhance the ability to separate quark and gluon jets from pileup jets. Is defined as

$$p_T^D = \frac{\sqrt{\sum_i p_{T_i}^2}}{\sum_i p_{T_i}} \quad (5.6)$$

the sums run over all the PF constituents inside the jet. As pileup jets tend to have lower p_T^D than gluon jets, the addition of this variable enhances the gluon-pileup separation, particularly at high η

- pull: magnitude of the pull vector.
- jetRchg: transverse momentum fraction carried by the leading charged PF candidate.

Is defined as

$$jetRchg = \frac{p_{T_j}}{\sum_i p_{T_i}} \quad (5.7)$$

j is the constituent with the greatest transverse momentum and the sum runs over all the PF constituents inside the jet.

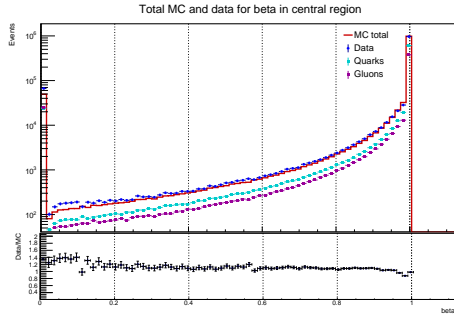
Finally, the transverse momentum (p_T) as well as the pseudirapidity (η) of the jet are also added in the list of variables.

These variables are combined in order to perform the training of the BDT. The training is performed separately for the two regions of the detector. In the central region ($|\eta| < 3$ both track-based and shape based variables are combined due to the fact that in this region there is information about the tracks of particle

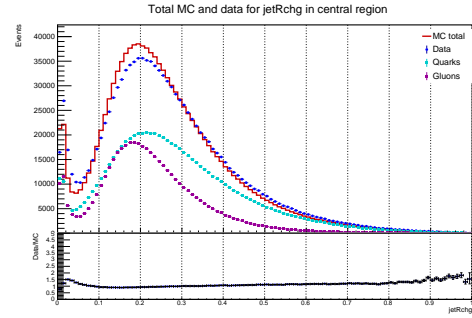
candidates from the tracker. In the following figures ??-?? there are the distribution of these variables for the reconstructed PUPPI jets, in the specific region of η for jet $p_T > 20\text{GeV}$.

In the forward region $3 < |\eta| < 5$ only shape-based variables are used, as this region is out of range of the tracker and tracking is not available. In the following figures ??-?? there are the distribution of these variables for the reconstructed PUPPI jets, in the specific region of η for jet $p_T > 20\text{GeV}$.

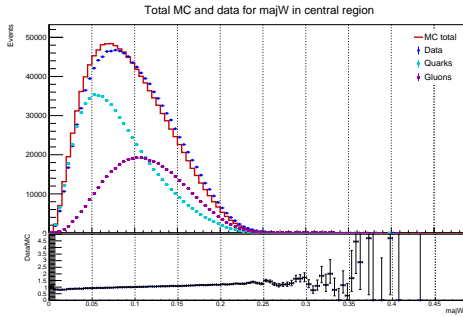
In the figures ??-?? there the distributions of transeverse momentum of the reconstructed PUPPI jets in the two regions of the detector as well as the pseudorapidity of them.



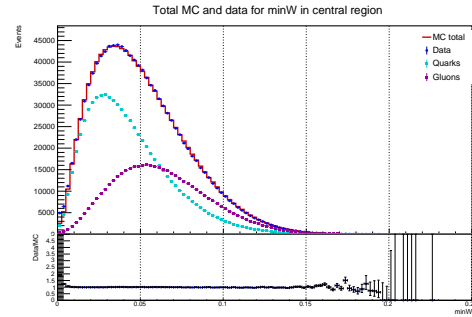
(a) Distribution of the fraction of tracks associated to the primary vertex



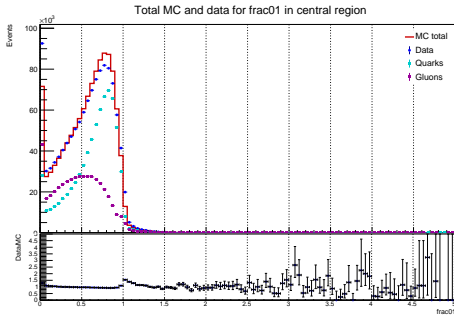
(b) Distribution of the transverse momentum fraction carried by the leading charged PF candidate



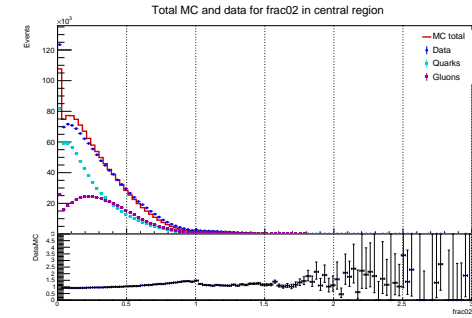
(c) Distribution of the major axis of the jet ellipsoid in the $\eta - \phi$ space



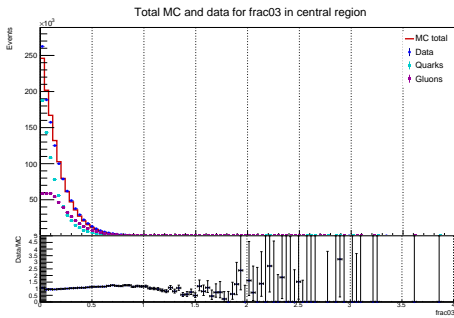
(d) Distribution of the minor axis of the jet ellipsoid in the $\eta - \phi$ space



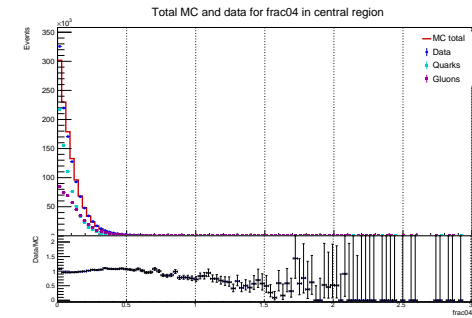
(e) Distribution of the concentric ring about the jet axis $0.1 < (\Delta R) < 0.2$



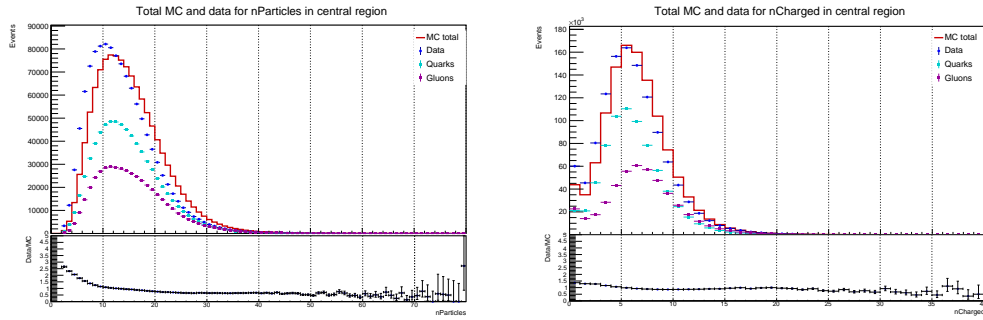
(f) Distribution of the concentric ring about the jet axis $0.2 < (\Delta R) < 0.3$



(g) Distribution of the concentric ring about the jet axis $0.2 < (\Delta R) < 0.3$

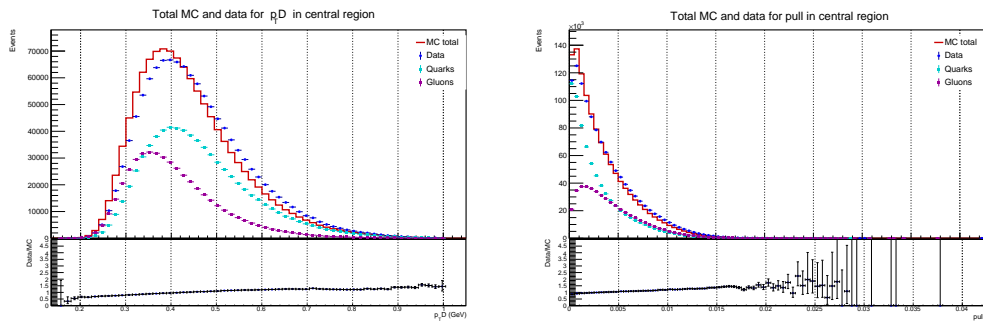


(h) Distribution of the concentric ring about the jet axis $0.3 < (\Delta R) < 0.4$



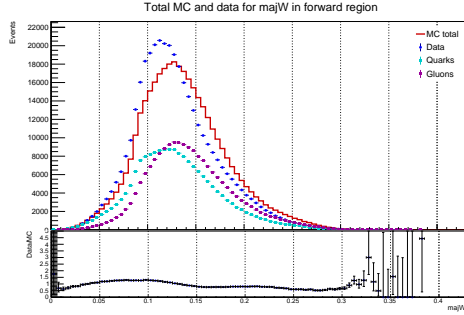
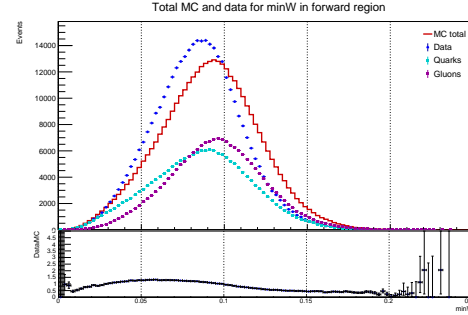
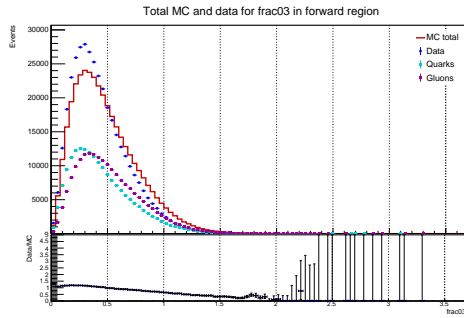
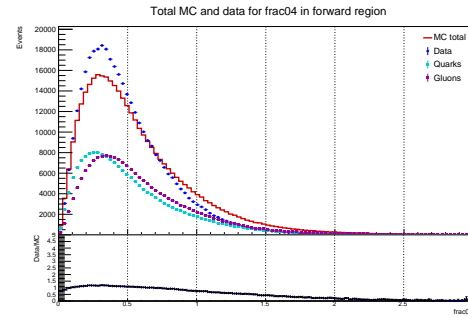
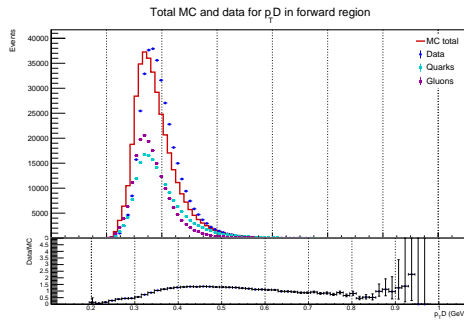
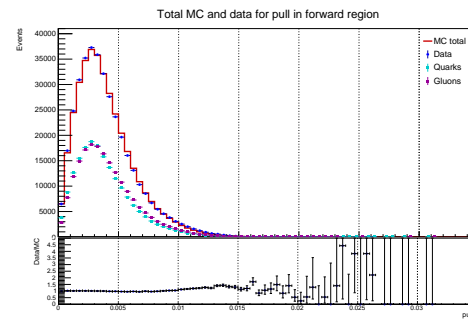
(i) Distribution of the number of PF candidates

(j) Distribution of the number of charged PF candidates

(k) Distribution of the p_T -weighted average p_T of constituents

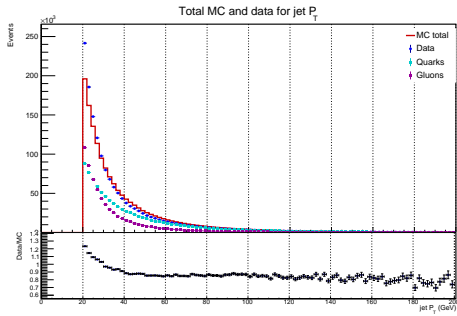
(l) Distribution of the magnitude of the pull vector

Figure 5.1: Distribution of the variables for the reconstructed PUPPI jets for $p_T > 20 \text{ GeV}$ in the central region $|\eta| < 3$.

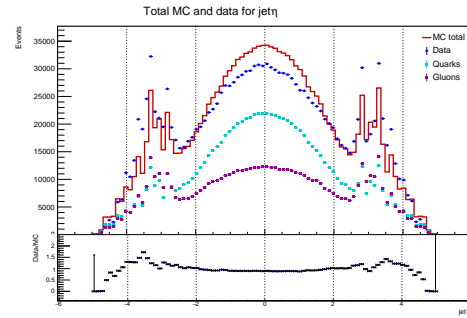
(a) Distribution of the major axis of the jet ellipsoid in the $\eta - \phi$ space(b) Distribution of the minor axis of the jet ellipsoid in the $\eta - \phi$ space(c) Distribution of the concentric ring about the jet axis $0.2 < (\Delta R) < 0.3$ (d) Distribution of the concentric ring about the jet axis $0.3 < (\Delta R) < 0.4$ (e) Distribution of the p_T -weighted average p_T of constituents

(f) Distribution of the magnitude of the pull vector

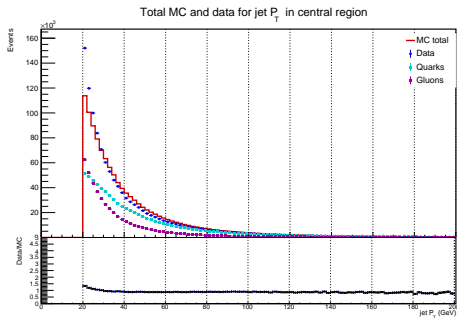
Figure 5.2: Distribution of the variables for the reconstructed PUPPI jets for $p_T > 20 \text{ GeV}$ in the forward region $3 < |\eta| < 5$.



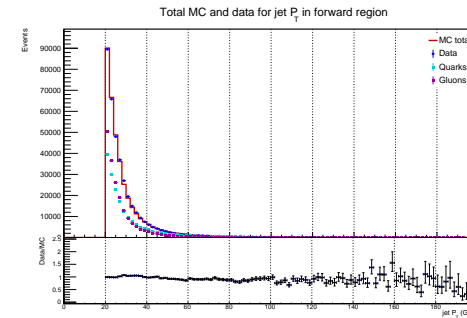
(a) Distribution of the total transverse momentum of jet in the whole region of the detector $|\eta| < 5$



(b) Distribution of the total pseudorapidity η of jet

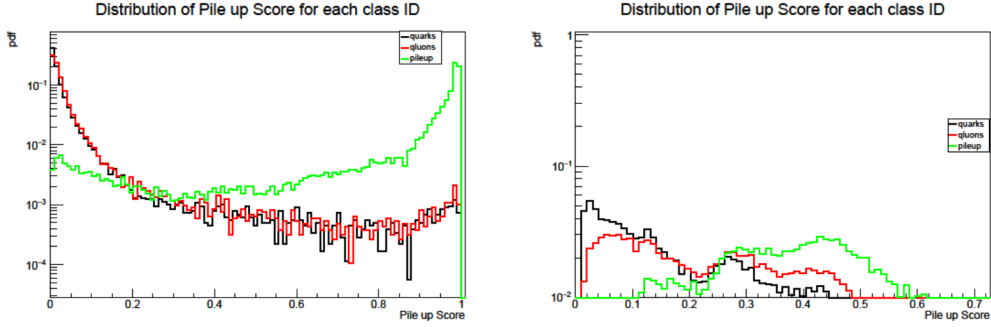


(c) Distribution of the transverse momentum of jet in the central region $|\eta| < 3$



(d) Distribution of the total transverse momentum of jet in the forward region $3 < |\eta| < 5$

Figure 5.3: Distribution of the transverse momentum and pseudorapidity for the reconstructed PUPPI jets for $p_T > 20 \text{ GeV}$.



(a) Distribution of the output scores of the BDT for the central region

(b) Distribution of the output scores of the BDT for the forward region

Figure 5.4: Distribution output scores of the BDT in both regions of η .

5.3.2 Pileup Jet Identification Training

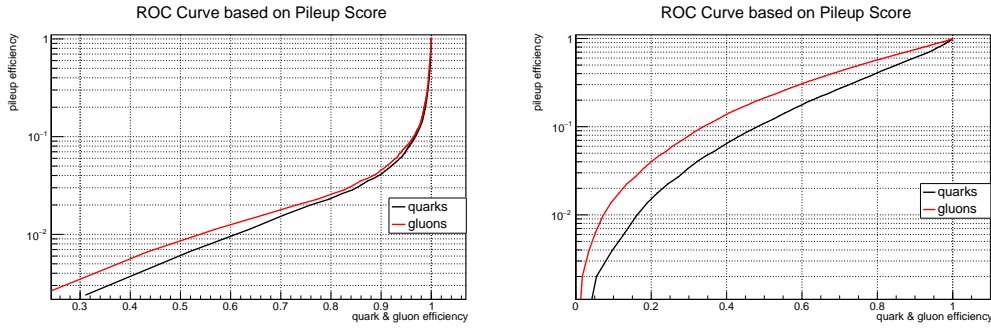
To perform the training all the aforementioned shape and tracking variables are added to a boosted decision tree. The training is performed separately for the two different bins in η for all jets with $p_T > 20\text{GeV}$. Figure ?? shows the pileup jet id BDT output distribution for PUPPI jets with $p_T > 20\text{GeV}$.

The training is done using a Drell-Yan process simulation, since the quark-gluon composition of the jets in these samples is representative of for the most important use case of pileup jet identification. Pileup jets, quark jets, and gluon jets are defined based on matching the jet to generated partons from the hard scattering and simulated jets. Quark and gluon jets are defined as jets which are with $\Delta R < 0.2$ of a hard process quark or gluon. Pileup jets are defined as jets which are not matched to a simulated jet ($\Delta R > 0.3$), nor are they matched with a parton within $\Delta R < 0.2$. Any jet not falling into these categories is defined as “rest” and not used for the training of the BDT.

5.4 Performance

The performance of the pileup jet identification algorithms is evaluated with simulated Drell-Yan processes $q\bar{q} \rightarrow l^+l^- + jets$ events. The MVA output values are used to classify the events as either good jets or pileup jets. For each MVA output value, the probability for a good jet to have a higher value defines the signal efficiency $\epsilon(\text{signal})$, whereas the probability for a pileup jet to have a higher value gives the background efficiency $\epsilon(\text{background})$.

The performance is characterized by the ROC curves for the MVA classifiers. The results are derived yielding working points for the two different jet η regions. The working points are defined as a function of jet p_T in these two regions in order to maximize and stabilise the signal efficiency $\epsilon(\text{signal})$. Finally,



(a) ROC Curve based on the Pileup score of the BDT for the central region (b) ROC Curve based on the Pileup score of the BDT for the forward region

Figure 5.5: ROC Curve based on the Pileup score of the BDT in both regions of η .

the efficiencies are determined separately for quark and gluon jets to get a hold of potential efficiency differences due to differences in the jet shapes.

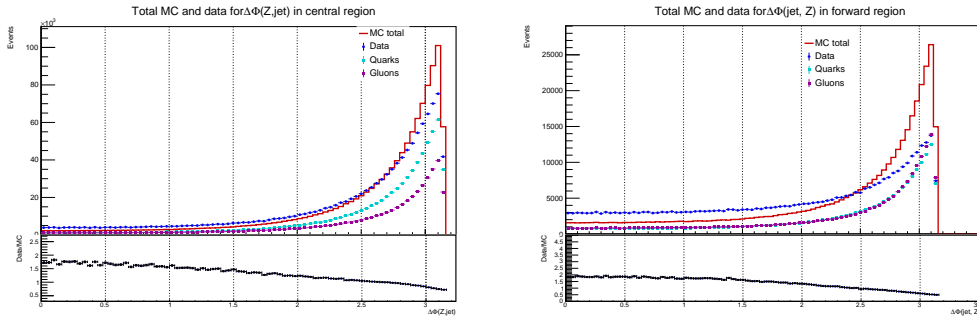
Quark and gluon jets have different properties that affect the discrimination from pileup jets. Most importantly, gluon jets are less collimated than quark jets, and they have a higher charged multiplicity as well as a softer fragmentation function. For the shape-based variables, this implies that gluon jets exhibit more pileup-like properties than quark jets. However, the larger charged multiplicity in conjunction with the softer fragmentation function leads to narrower distributions of the b variable for gluon jets, resulting in a higher discrimination between gluon and pileup jets at low values of β .

The performance for the different detector regions is given by the ROC curves in figure ??.

Based on pileup efficiency working points are defined, twain each area central and forward.

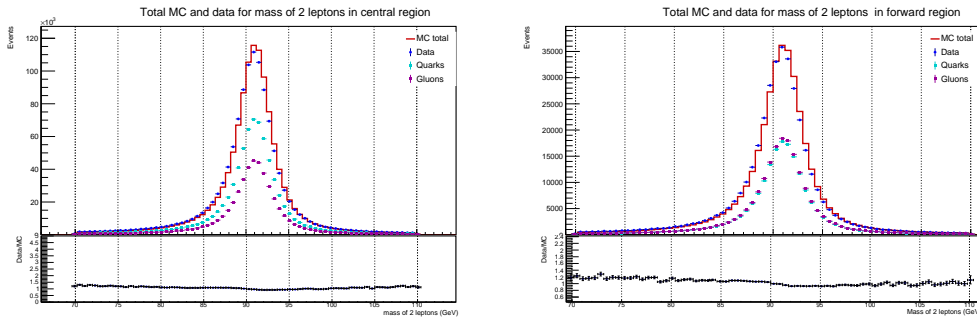
η bin	Pileup	Quark	Gluon
$ \eta < 3$	10%	98%	97%
	20%	99%	98%
$3 < \eta < 5$	20%	62%	50%
	40%	80%	65%

In order to maximize and stabilize the signal efficiency (quarks and gluons efficiency) a linear association between the working points and transverse momentum of jet (jet p_T) is defined. The same method is applied on the Data and Monte Carlo samples. The results are discussed in next section.



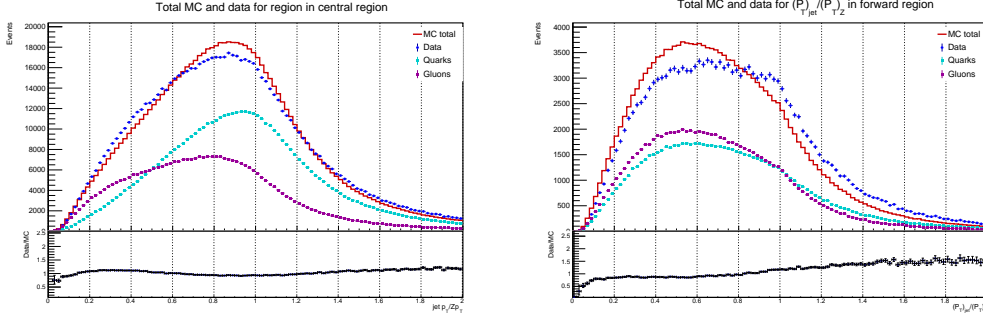
(a) Distribution of $\Delta\phi(Z, jet)$ for the cen- (b) Distribution of $\Delta\phi(Z, jet)$ for the for-
tral region ward region

Figure 5.6: Distribution of variable $\Delta\phi(Z, jet)$ in both regions of η .



(a) Distribution of the mass of two lep- (b) Distribution of the mass of two lep-
tons for the central region tons for the forward region

Figure 5.7: Distribution of mass of two leptons that peaks at the mass of Z boson in both regions of η .



(a) Distribution of the ratio $(p_T)_{jet}/(p_T)_Z$ for the central region
 (b) Distribution of the ratio $(p_T)_{jet}/(p_T)_Z$ for the forward region

Figure 5.8: Distribution of the ratio between the transverse momentum of the jet over the transverse momentum of Z boson ($(p_T)_{jet}/(p_T)_Z$) in both regions of η .

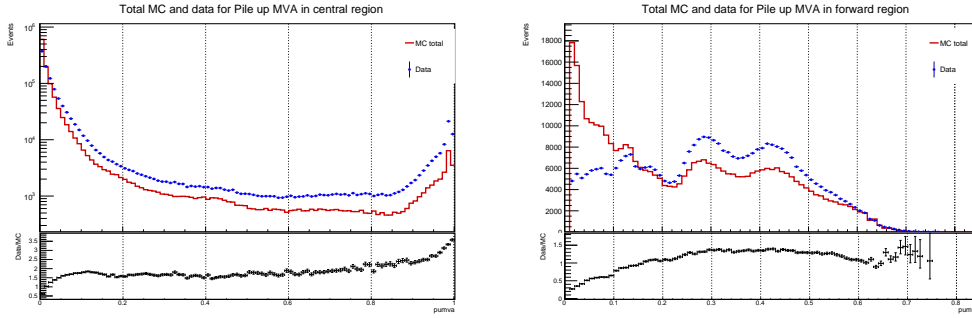
5.5 Performance in data

The efficiency of the pileup jet identification criteria on real jets is checked using a tag-and probe method on a control sample of $Z(\rightarrow \mu\mu) + jets$ events, where the jet recoiling against the Z is used as a probe. The efficiency in data is measured by utilizing the fact that the pileup jets are not correlated to the Z candidate. Requirements on the balancing between the Z and the hardest jet momenta are applied. We select events with a Z candidate and an accompanying jet with the ratio between the jet p_T and the Z p_T must be between 0.5 and 1.5 ($0.5 < p_{T,jet}/p_{T,Z} < 1.5$). Also two regions of interest were defined according to the absolute azimuthal separation $|\Delta\Phi(Z, jet)|$ of the two objects: the signal (prompt jet) region, with $|\Delta\Phi(Z, jet)| > 2.5$ (back-to-back) and the background (pileup jet) dominated region $|\Delta\Phi(Z, jet)| < 1.5$. Assuming that the distribution of pileup jets is flat in $\Delta\phi(Z, jet)$, the efficiency of a given pileup id working point is given by the following formula:

$$\epsilon = \frac{N_{passId,sig} - kN_{passId,bkg}}{N_{all,sig} - kN_{all,bkg}} \quad (5.8)$$

where $N_{all,sig}$ is the total number of jets in the control region ($|\Delta\Phi(Z, jet)| > 2.5$), $N_{all,bkg}$ is the total number of jets in the pileup enriched region ($|\Delta\Phi(Z, jet)| < 1.5$), $N_{passId,sig}$ is the number of jets in the control region passing the jet identification, $N_{passId,bkg}$ is the number of jets passing the jet identification in the pileup enriched region and, finally, $k = (\pi - 2.5)/1.5$ is the scaling factor to extrapolate the number of pileup jets from the pileup enriched region to the control sample.

The classification and efficiency calculations are made based on the MVA discriminator for jets in the two regions that are demonstrating in figure ??.



(a) MVA discriminator for jets with $p_T > 20 \text{ GeV}$ for the central region

(b) MVA discriminator for jets with $p_T > 20 \text{ GeV}$ for the forward region

Figure 5.9: Distribution of MVA discriminator for particle flow jets with $p_T > 20 \text{ GeV}$ in both regions of η .

The results of the efficiency measured in data and MC simulation and of their ratio are reported in figure ?? for central region and in figure ?? for forward region. As shown, the agreement between data and simulation is within 2-10% depending on the jet pseudorapidity (η) and transverse momentum range (p_T). The ratios of data to simulation in bins of jet p_T and η can be used as scale factors to correct for mismodeling in the simulation. The largest data/MC scale factors are observed in lower jet p_T bins.

It can be seen that the method that was developed is biased, while in central region in low p_T bins the efficiency calculated for data and MC samples is systematically $\sim 5\% - 10\%$ lower than the simulated efficiency. For $p_T > 40 \text{ GeV}$ the developed method seems to be more effective in general while efficiencies for data and MC samples converge to the simulated one for gluons and quarks.

In forward region the efficiency seems to get steady after $p_T > 40 \text{ GeV}$, while there is better agreement between the calculated efficiencies for data and MC samples and the simulated one for gluons and quarks.

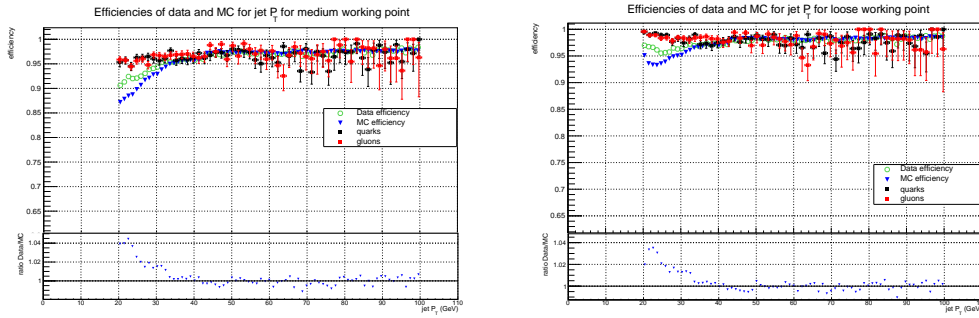


Figure 5.10: Efficiency as a function of transverse momentum for reconstructed PUPPI jets with $p_T > 20\text{GeV}$ that passes the medium (left) and the loose (right) working point in central region.

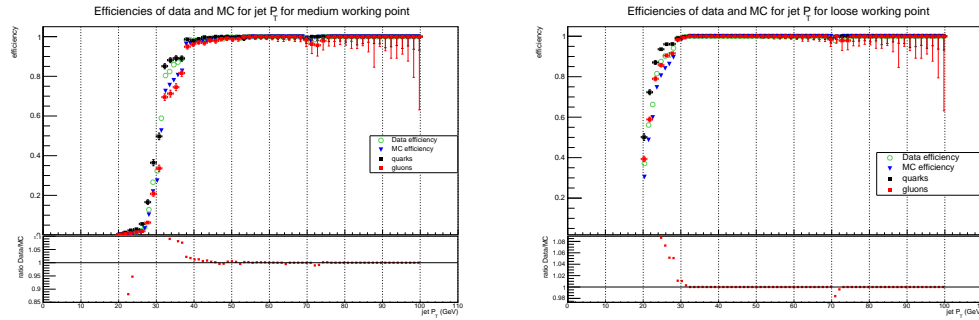


Figure 5.11: Efficiency as a function of transverse momentum for reconstructed PUPPI jets with $p_T > 20\text{GeV}$ that passes the medium (left) and the loose (right) working point in forward region.

Κεφάλαιο 6

Conclusions

Pileup jets are a ubiquitous background under the current running conditions of the Large Hadron Collider. Their presence typically arises from overlapping low p_T jets and grows roughly quadratically with the number of pileup collisions. Due to their unusual formation, pileup jets exhibit distinct features that allow them to be separated from real jets that have originated from either quarks or gluons.

Identification and removal of pileup jets is performed either through the use of vertex information or through the use of shape information. The former can be exploited in the central region of CMS detector where tracking is available. Shape information, although less effective than vertexing, extends throughout the whole detector volume and in conjunction with vertex information enhances the ability to identify pileup jets. Even though PUPPI (Pile Up Per Particle) algorithm is an efficient way to remove pileup it is not always due to the fact that random pileup clusters might have created incidentally. That is the reason why further removal of pileup must be done through Pileup jet Identification Algorithm.

In this analysis fourteen in total number vertex and shape variables were combined through a multivariate BDT that gave the pileup id. In addition, this pileup id was used in order to maximize and stabilize the efficiency of jet p_T in central and forward region, through the definition of working points a linear function of jet p_T . In fact, this method was produced to identify and remove the pileup that have been left after the application of PUPPI Algorithm. Regarding the efficiency of transverse momentum of jet is a biased method for the low p_T bins. Comparing the true efficiency of Quarks and Gluons with the efficiency of Monte-Carlo samples and Data, in the central region for jet $p_T > 30GeV$ there is good agreement whereas for $20 < p_T < 30GeV$ Data and MC-samples are systematically lower than true efficiency independently of the working point. Also in low p_T efficiency of Data is $\sim 10\%$ greater than the efficiency of Monte-Carlo samples. On the other hand, in forward region, there is better agreement between the efficiencies.

Βιβλιογραφία

- [1] D. Griffiths, "Introduction to Elementary Particles, Revised Edition"
- [2] CMS Collaboration, "The CMS Experiment at the CERN LHC", JINST 3 (2008) S08004, doi:10.1088/1748-0221/3/08/S08004.
- [3] Particle Data Group Review of Particle Physics 2016, Chinese Physical Society, 1674-1137
- [4] Luca Lista Statistical Methods for Data Analysis in Particle Physics, Springer
- [5] A. Pich, "The Standard Model of Electroweak Interactions", 2005.
- [6] F. Beaudette, "The CMS Particle Flow Algorithm", arXiv:1401.8155
- [7] O. Behnke, K. Kroninger, G. Schott, T. Schorner-Sadenius "Data Analysis in High Energy Physics" WILEY-VCH
- [8] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-kt jet clustering algorithm", doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189
- [9] CMS Collaboration "Pileup Jet Identification" CMS AN AN-13-186
- [10] CMS Collaboration "Jet algorithms performance in 13 TeV data" arXiv:1712.07158v2
- [11] Daniele Bertolini, Philip Harris, Matthew Low, Nhan Tran, "Pileup Per Particle Identification" arXiv:1407.6013