



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ**

**ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ
ΜΑΘΗΜΑΤΙΚΩΝ
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ
ΜΗΧΑΝΙΚΩΝ**

ΕΚΕΦΕ «ΔΗΜΟΚΡΙΤΟΣ»

**ΙΝΣΤΙΤΟΥΤΟ
ΝΑΝΟΕΠΙΣΤΗΜΗΣ ΚΑΙ
ΝΑΝΟΤΕΧΝΟΛΟΓΙΑΣ**

**ΙΝΣΤΙΤΟΥΤΟ ΠΥΡΗΝΙΚΗΣ ΚΑΙ
ΣΩΜΑΤΙΔΙΑΚΗΣ ΦΥΣΙΚΗΣ**



**Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
«Φυσική και Τεχνολογικές Εφαρμογές»**

**Μέτρηση ενεργού διατομής παραγωγής $W+bb$ στο
πείραμα CMS του LHC.**

**ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
της Ειρήνης Σιαμάρκου**

Επιβλέπων: Κωνσταντίνος Κουσουρής

Αθήνα, Οκτώβριος, 2020

© (2006) Εθνικό Μετσόβιο Πολυτεχνείο. All rights Reserved. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς το συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σ' αυτό έγγραφο εκφράζουν το συγγραφέα και δεν πρέπει να ερμηνευτεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Στόχος της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι η μελέτη της παραγωγής W μποζονίων σε συνδυασμό με δύο πίδακες b ($W + b\bar{b}$), σε συγκρούσεις πρωτονίων-πρωτονίων, με ενέργεια κέντρου μάζας $\sqrt{s} = 13\text{TeV}$.

Abstract

This master thesis aims at the study of W bosons production in association with two b jets ($W + b\bar{b}$), in proton-proton collisions, at centre of mass energy $\sqrt{s} = 13TeV$.

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε, στον τομέα φυσικής της σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών, στα πλαίσια των σπουδών μου στο διατμηματικό μεταπτυχιακό πρόγραμμα Φυσική και Τεχνολογικές Εφαρμογές. Με την ολοκλήρωση της επιθυμώ να εκφράσω ευχαριστίες προς όσους συντέλεσαν στη διεξαγωγή της.

Καταρχάς, θα ήθελα να ευχαριστήσω τον κ. Κωνσταντίνο Κουσουρή, Επίκουρο Καθηγητή της σχολής, υπό την επίβλεψη του οποίου έφερα σε πέρας την εργασία μου, για την εμπιστοσύνη που μου έδειξε με την ανάθεση της διπλωματικής και για την συνεχή καθοδήγηση. Επιπρόσθετα, θέλω να εκφράσω τις ευχαριστίες μου στον ίδιο όσο και στον κ. Γεώργιο Τσιπολίτη, Καθηγητή της σχολής, για το χρόνο που μου αφιέρωσαν, τις γνώσεις που μου μετέδωσαν και για τις συμβουλές.

Συνεχίζοντας, θα ήθελα να ευχαριστήσω όλα τα υπόλοιπα μέλη της ομάδας, για τη συνεργασία και για τις επικοδομητικές συζητήσεις στις εβδομαδιαίες συναντήσεις.

Τέλος, να ευχαριστήσω την οικογένειά μου και τους φίλους μου για τη συμπαράσταση και τη στήριξη κατά τη διάρκεια των σπουδών μου.

Περιεχόμενα

| | |
|--|-----------|
| Περίληψη | iv |
| Absract | v |
| Ευχαριστίες | vii |
| 1 Στοιχειώδη Σωματίδια-Θεωρητικά Στοιχεία | 1 |
| 1.1 Καθιερωμένο Πρότυπο-Standard Model | 1 |
| 1.2 Κβαντική Ηλεκτροδυναμική-QED | 5 |
| 1.3 Κβαντική Χρωμοδυναμική-QCD | 6 |
| 1.4 Ηλεκτρασθενής Αλληλεπίδραση | 7 |
| 2 Πείραμα | 11 |
| 2.1 Ο επιταχυντής LHC-Large Hadron Collider | 11 |
| 2.2 Ο ανιχνευτής CMS (Compact Muon Solenoid) | 12 |
| 2.3 Συγκρούσεις Πρωτονίων Πρωτονίων | 15 |
| 2.4 Μιόνια | 15 |
| 2.5 Ηλεκτρόνια | 16 |
| 2.5.1 Ομαδοποίηση των εναποθέσεων ενέργειας του ηλεκτρονίου στο ηλεκτομαγνητικό θερμιδόμετρο ECAL | 16 |
| 2.5.2 Ανακατασκευή της τροχιάς των ηλεκτρονίων | 18 |
| 2.5.3 Ταυτοποίηση ηλεκτρονίων | 19 |
| 2.5.4 Isolation των ηλεκτρονίων | 19 |
| 2.6 Αλγόριθμος Anti- k_T | 20 |
| 3 Ανάλυση Δεδομένων | 23 |
| 3.1 Μηχανική μάθηση – Machine learning | 23 |
| 3.2 Δέντρο Απόφασης - Decision Tree | 24 |
| 3.3 Ενδυναμωμένο Δέντρο Απόφασης - Boosted Decision Tree | 26 |
| 4 Ανάλυση | 29 |
| 4.1 Εισαγωγή | 29 |
| 4.2 Δείγματα Προσομοίωσης - Monte Carlo samples | 30 |
| 4.3 $W + b\bar{b}$ | 31 |
| 4.4 Διεργασίες υποβάθρου | 33 |

| | | |
|--------|--|----|
| 4.5 | Μεταβλητές | 35 |
| 4.6 | Επιλογή περιορισμών που ορίζουν το φασικό χώρο-Selection cuts | 37 |
| 4.7 | Boosted Decision Tree | 39 |
| 4.7.1 | Μεταβλητές | 40 |
| 4.7.2 | Απόκριση του Boosted Decision Tree - Score | 42 |
| 4.7.3 | Καμπύλη ROC | 43 |
| 4.8 | Σύγκριση της απόκρισης του BDT στα Monte Carlo δείγματα και στα δεδομένα | 44 |
| 4.9 | Data Driven Estimation-QCD multijet | 46 |
| 4.10 | control regions for $t\bar{t}$ | 48 |
| 4.11 | control region for Drell-Yan | 51 |
| 4.12 | BDT 2: ορισμένο σε διευρυμένο φασικό χώρο | 55 |
| 4.12.1 | Μεταβλητές | 55 |
| 4.12.2 | Απόκριση του Boosted Decision Tree - Score | 58 |
| 4.12.3 | Καμπύλη ROC | 58 |
| 4.13 | Fits | 61 |
| 4.13.1 | Toy Monte Carlo | 63 |
| 4.13.2 | Αποτελέσματα του fit | 68 |
| 4.14 | Ενεργός Διατομή - Cross-Section | 71 |
| 4.15 | Συμπεράσματα | 71 |

Κατάλογος Σχημάτων

| | | |
|------|--|----|
| 2.1 | Σχηματική αναπαράσταση του μεγάλου αδρονικού επιταχυντή (LHC) του CERN | 12 |
| 2.2 | α) Το σύστημα συντεταγμένων στο CMS και β) Ενδεικτικές τιμές της ψευδοωκότητας | 13 |
| 2.3 | Διατομή ενός τετάρτου του ανιχνευτή CMS | 14 |
| 2.4 | Απεικόνιση πιδάκων, κυρίως υψηλής ενέργειας, με χρήση του αλγόριθμου anti- k_T | 22 |
| 3.1 | Σχηματική αναπαράσταση ενός δέντρου απόφασης που διαχωρίζει το σήμα (signal-S) από το υπόβαθρο (background-B). | 25 |
| 3.2 | Σχηματική απόδοση ενός Boosted Decision Tree. | 27 |
| 4.1 | Διαγράμματα feynman πρώτης τάξης | 29 |
| 4.2 | Παραγωγή $W + b\bar{b}$ | 32 |
| 4.3 | Παραγωγή W μποζονίου από σύγκρουση πρωτονίων. | 32 |
| 4.4 | Παραγωγή $t\bar{t}$ | 33 |
| 4.5 | Παραγωγή single top διεργασιών | 34 |
| 4.6 | Παραγωγή Drell-Yan | 35 |
| 4.7 | Σχηματική αναπαράσταση της Pt_{rel} | 37 |
| 4.8 | Μεταβλητές Boosted Decision Tree. | 42 |
| 4.9 | Πίνακες συσχετίσεων | 42 |
| 4.10 | Απόκριση του Boosted Decision Tree. | 43 |
| 4.11 | Καμπύλη ROC του Boosted Decision Tree. | 44 |
| 4.12 | Αποτέλεσμα του Boosted Decision Tree για όλες τις διεργασίες. | 45 |
| 4.13 | Η control περιοχή της qcd. | 46 |
| 4.14 | Η control περιοχή της qcd με αφαίρεση του contamination. | 47 |
| 4.15 | Closure test της qcd-multijet διεργασίας στο κανάλι των ηλεκτρονίων. | 48 |
| 4.16 | $t\bar{t}$ control περιοχές στο κανάλι των ηλεκτρονίων | 49 |
| 4.17 | Dilepton $t\bar{t}$ control περιοχή στο κανάλι των μιονίων. | 49 |
| 4.18 | fit στην dilepton control $t\bar{t}$ περιοχή στο κανάλι των ηλεκτρονίων | 50 |
| 4.19 | fit στην dilepton control $t\bar{t}$ περιοχή στο κανάλι των μιονίων | 50 |
| 4.20 | control περιοχές Drell-Yan στο κανάλι των ηλεκτρονίων | 52 |
| 4.21 | control περιοχές Drell-Yan στο κανάλι των μιονίων | 52 |

| | |
|---|----|
| 4.22 fits στην Drell-Yan+ jj control περιοχή | 53 |
| 4.23 Stack ιστόγραμμα του output του BDT με τους scale factors για το κανάλι των ηλεκτρονίων. | 54 |
| 4.24 Stack ιστόγραμμα του output του BDT με τους scale factors για το κανάλι των μιονίων. | 54 |
| 4.25 Μεταβλητές Boosted Decision Tree. | 57 |
| 4.26 Πίνακες συσχετίσεων | 57 |
| 4.27 Απόκριση του Boosted Decision Tree. | 58 |
| 4.28 Καμπύλη ROC του Boosted Decision Tree. | 58 |
| 4.29 Απόκριση του BDT 2 για όλες τις διεργασίες | 59 |
| 4.30 Κατανομές των δύο BDT-ηλεκτρόνια | 60 |
| 4.31 Κατανομές των δύο BDT-μιόνια | 60 |
| 4.32 Toy Monte Carlo για το BDT-14 μεταβλητές στο κανάλι των ηλεκτρονίων | 64 |
| 4.33 Toy Monte Carlo για το BDT-16 μεταβλητές στο κανάλι των ηλεκτρονίων | 65 |
| 4.34 Τιμές του r_{fit} ως προς $r_{injected}$ για toys MC του BDT-14 μεταβλητές. | 66 |
| 4.35 Toy Monte Carlo για το BDT-14 μεταβλητές στο κανάλι μιονίων | 67 |
| 4.36 Toy Monte Carlo για το BDT-16 μεταβλητές στο κανάλι μιονίων | 67 |
| 4.37 Fit του BDT στο κανάλι των ηλεκτρονίων | 68 |
| 4.38 BDT-14 μεταβλητές prefit και postfit | 68 |
| 4.39 Fit του BDT-14 μεταβλητές στο κανάλι των μιονίων | 69 |
| 4.40 BDT-14 μεταβλητές prefit και postfit, κανάλι μιονίων | 70 |

Κεφάλαιο 1

Στοιχειώδη Σωματίδια-Θεωρητικά Στοιχεία

1.1 Καθιερωμένο Πρότυπο-Standard Model

Ο κύριος στόχος της σωματιδιακής φυσικής είναι να ταυτοποιήσει τα σωματίδια που αποτελούν δομικά στοιχεία της ύλης χωρίς εσωτερική δομή και η κατανόηση των δυνάμεων μέσω των οποίων αλληλεπιδρούν μεταξύ τους[1]. Οι θεωρίες που περιγράφουν τα παραπάνω ονομάζονται συνολικά Καθιερωμένο Πρότυπο (Standard Model-SM).

Τα δομικά στοιχεία της ύλης είναι τα φερμιόνια με σπιν $\frac{1}{2}$ (στο φυσικό σύστημα μονάδων). Αυτά χωρίζονται σε δύο είδη: τα λεπτόνια και τα κουάρκ. Και τα δύο αυτά είδη δεν έχουν παρουσιάσει δομή στις μικρότερες αποστάσεις που έχουν επιτευχθεί από τους επιταχυντές υψηλών ενεργειών. Τα λεπτόνια αλληλεπιδρούν μέσω δύο δυνάμεων, την ηλεκτρομαγνητική και την ασθενή, και τα κουάρκ μέσω τριών, την ισχυρή, την ηλεκτρομαγνητική και την ασθενή. Η ηλεκτρομαγνητική και η ασθενής αλληλεπίδραση των λεπτονίων και των κουάρκ περιγράφονται ενοποιημένα από την ηλεκτρασθενή θεωρία των Glashow, Salam και Weinberg που είναι μία γενίκευση της κβαντικής ηλεκτροδυναμικής (QED). Η ισχυρή αλληλεπίδραση των κουάρκ περιγράφεται από την κβαντική χρωμοδυναμική (QCD), που είναι επίσης ανάλογη της QED. Η ομοιότητα αυτή οφείλεται στο γεγονός ότι και οι τρεις αλληλεπιδράσεις είναι θεωρίες βαθμίδας.

- **Λεπτόνια**

Τα λεπτόνια αποτελούνται από e^- , μ^- και τ^- και τα αντισωματίδια τους e^+ , μ^+ και τ^+ και το καθένα εκτός από το σπιν καθορίζεται και από ένα κβαντικό αριθμό που ονομάζεται λεπτονικός αριθμός και θεωρείται ότι διατηρείται. Έτσι υπάρχει ο ηλεκτρονικός λεπτονικός αριθμός, ο μιονικός και ο λεπτονικός αριθμός του ταυ. Αυτοί είναι ορισμένοι έτσι ώστε: $L_l(l^-) = 1$ και $L_l(l^+) = -1$, όπου $l=e, \mu$ ή τ . Και τα τρία λεπτόνια στις ηλεκτρομαγνητικές αλληλεπιδράσεις αντιδρούν το ίδιο. Στις ασθενείς κάθε φορτισμένο λεπτόνιο συνοδεύεται από το αντίστοιχο ουδέτερο

σωματίδιο, το νεutrίνο. Τα νεutrίνο είναι και αυτά τριών ειδών: νεutrίνο ηλεκτρονίου, μιονίου και ταυ (ν_e, ν_μ, ν_τ). Αυτό που εκπέμφθηκε με ένα ηλεκτρόνιο στη β -διάσπαση, αρχικά εισήχθη από τον Pauli το 1930, ως μία προσπάθεια να «σωθούν» οι νόμοι διατήρησης της τετραορμής και της γωνιακής ορμής. Στο Καθιερωμένο Πρότυπο τα νεutrίνο έχουν και αυτά λεπτονικούς αριθμούς, ώστε να διατηρούνται οι γεύσεις των λεπτονίων χωριστά. Επομένως, καταλήγουν σε τρεις γεννιές φορτισμένων και ουδέτερων λεπτονίων σε διπλέτες $(e, \nu_e), (\mu, \nu_\mu), (\tau, \nu_\tau)$ και στα αντισωματίδια αυτών.

Εκτός από τη γεύση, τα νεutrίνο έχουν άλλη μία ενδιαφέρουσα ιδιότητα, τη μάζα τους. Αρχικά θεωρήθηκαν άμαζα, όμως πλέον είναι γνωστό, από τις ταλαντώσεις νεutrίνων, ότι έχουν πολύ μικρές μάζες αλλά όχι μηδενικές. Για τις μάζες αυτές υπάρχουν μόνο άνω όρια και όχι ακριβείς μετρήσεις. Αν τα νεutrίνο ήταν άμαζα τότε νεutrίνο από διαφορετικές γεύσεις δεν θα μπορούσαν να έχουν εναλλαγή.

| Generation | Particle | Mass (MeV) | Q/e | L_e | L_μ | L_τ |
|------------|------------|----------------------|-------|-------|---------|----------|
| 1 | ν_e | $< 2 \times 10^{-6}$ | 0 | 1 | 0 | 0 |
| | e^- | 0.511 | - 1 | 1 | 0 | 0 |
| 2 | ν_μ | < 0.19 | 0 | 0 | 1 | 0 |
| | μ^- | 105.658 | - 1 | 0 | 1 | 0 |
| 3 | ν_τ | < 18.2 | 0 | 0 | 0 | 1 |
| | τ^- | 1777 | - 1 | 0 | 0 | 1 |

ΠΙΝΑΚΑΣ 1.1: Ιδιότητες Λεπτονίων

• Κουάρκ

Τα κουάρκ είναι τα συστατικά των αδρονίων στα οποία είναι δέσμια εξαιτίας των ισχυρών κβαντικών χρωμοδυναμικών δυνάμεων. Τα αδρόνια χωρίζονται σε αυτά με σπιν $\frac{1}{2}, \frac{3}{2}, \frac{5}{2} \dots$, φερμιόνια που είναι βαρυόνια και σε αυτά με σπιν 0,1,2, τα μποζόνια που είναι μεσόνια. Παραδείγματα βαρυονίων είναι τα νουκλεόνια (το ουδέτερο νεutrόνιο και το φορτισμένο πρωτόνιο). Απόδειξη της σύνθετης φύσης των αδρονίων συσσωρεύτηκε τη δεκαετία 1960 με 1970. Η ελαστική σκέδαση των ηλεκτρονίων από πρωτόνια από τον Hofstadter έδειξαν ότι το πρωτόνιο δεν είναι σημειακό σωματίδιο αλλά έχει προσεγγιστικά εκθετική κατανομή φορτίου.

Οι Gell-Mann και Zweig (1964) πρότειναν ότι τα βαρυόνια περιέχουν τρία συστατικά στοιχεία (σωματίδια) με σπιν $\frac{1}{2}$ που ονόμασαν κουάρκ, ενώ τα μεσόνια είναι σύστημα ενός κουάρκ και ενός αντικουάρκ. Αμεσο επακόλουθο είναι ότι τα κουάρκ έχουν κάποιο ποσοστό ηλεκτρομαγνητικού φορτίου. Για παράδειγμα το u κουάρκ έχει $+\frac{2}{3}$ ενώ το d έχει $-\frac{1}{3}$.

Αντίθετα από άλλα σωματίδια, τα κουάρκ δεν έχουν παρατηρηθεί ως σταθερά απομονωμένα σωματίδια. Όταν αδρόνια σε υψηλές ενέργειες συγκρούονται μεταξύ τους, αυτό που παρατηρείται είναι μόνο πολλά παραπάνω αδρόνια. Η εξήγηση για αυτό βρίσκεται στη φύση της δύναμης της QCD. Η χβαντοχρωμοδυναμική υπονοεί τον «περιορισμό» (confinement) των κουάρκ, ότι δηλαδή δεν υπάρχουν ελεύθερα στη φύση μόνο ομαδοποιημένα σχηματίζοντας τα αδρόνια. Υπάρχουν τρεις γεννιές από διπλέτες κουάρκ οι οποίες είναι: (u,d), (c,s), (t,b) μαζί με τα αντισωματίδια τους.

Είναι αξιοσημώτο για το top (t) κουάρκ ότι έχει μεγάλη μάζα. Πρέπει να ειπωθεί ότι αν και διασπάται ασθενώς, η ενέργεια που ελευθερώνεται είναι τόσο μεγάλη που η διάρκεια ζωής του είναι δύο τάξεις μεγέθους μικρότερη από μία τυπική ισχυρή αλληλεπίδραση. Αυτό σημαίνει ότι διασπάται πριν προλάβει να σχηματίσει αδρόνια. Έτσι, όταν ένα top κουάρκ παραχθεί διασπάται ως ελεύθερο σωματίδιο. Η μάζα του μπορεί να καθοριστεί από κινηματική ανάλυση των προϊόντων της διάσπασης.

Τα κουάρκ διαθέτουν και αυτά μία σειρά από χβαντικούς αριθμούς που τα καθορίζουν. Κάθε ένα κουάρκ έρχεται σε τρεις κατηγορίες που διακρίνονται από ένα χβαντικό αριθμό που ονομάζεται «χρώμα». Είναι ακριβώς αυτός ο χβαντικός αριθμός που βασίζεται στη δυναμική της QCD. Το χρώμα είναι στην πραγματικότητα ένα είδος φορτίου για τις ισχυρές αλληλεπιδράσεις. Δηλώνονται τρία χρώματα: το κόκκινο (red), το μπλε (blue) και ο πράσινο (green). δημιουργούνται λοιπόν οι τριπλέτες (q_r, q_b, q_g), όπου q: κάθε ένα από τα έξι κουάρκ.

Επιπλέον τα κουάρκ έχουν χβαντικούς αριθμούς γεύσης, όπως τα λεπτόνια. Τα δύο κουάρκ u και d είναι μία διπλέτα isospin με $T_3 = +\frac{1}{2}$ για το u και $T_3 = -\frac{1}{2}$ για το d. Η γεύση για το s κουάρκ λέγεται «παραξενιά» (strangeness) και έχει τιμή $S = -1$, για το c είναι charm και έχει τιμή $C = +1$. Αντίστοιχα το b κουάρκ έχει $-B = -1$ και το t κουάρκ έχει $T = +1$. Όπου έχει γίνει δεκτή η σύμβαση ότι το πρόσημο είναι το ίδιο με το πρόσημο του φορτίου.

Η ισχυρή και η ηλεκτρομαγνητική αλληλεπίδραση είναι ανεξάρτητες από τις γεύσεις των κουάρκ. Αυτό σημαίνει ότι η γεύση δεν αλλάζει κατά τη διάρκεια ισχυρών αλληλεπιδράσεων ανάμεσα σε αδρόνια αλλά ότι διατηρείται. Η ισχυρή

αλληλεπίδραση εξαρτάται από το φορτίο του χρώματος και η ηλεκτρομαγνητική από το ηλεκτρομαγνητικό φορτίο. Αντίθετα στις ασθενείς αλληλεπιδράσεις, η γεύση των κουάρκ δεν διατηρείται.

Στο καθιερωμένο πρότυπο θεωρείται ότι υπάρχει ίσος αριθμός λεπτονικών και κουάρκ οικογενειών (γιατί χαλάει η επακανονικοποιησιμότητα της ασθενούς αλληλεπίδρασης).

Όσο αφορά στις μάζες των κουάρκ, μόνο η μάζα του top μπορεί να μετρηθεί άμεσα. Όλα τα άλλα είναι δέσμια μέσα σε αδρόνια και δεν είναι προφανές πως να μετρηθούν οι μάζες τους. Η δύναμη της ισχυρής αλληλεπίδρασης είναι πολύ αδύναμη σε κοντινές αποστάσεις. Αυτή η ιδιότητα ονομάζεται asymptotic freedom. Για να διαχωριστούν τα κουάρκ, πρέπει να παρέχεται στο σύστημα ενέργεια, όταν τα δύο κουάρκ έχουν απομακρυνθεί η ενέργεια του συστήματος είναι μεγαλύτερη από το άθροισμα των μαζών των κουάρκ μικρής απόστασης.

| Generation | Particle | Mass | Q/e | S | C | \tilde{B} | T |
|------------|-----------------|------------------|-------|-----|-----|-------------|-----|
| 1 | u_r, u_b, u_g | 1.7 to 3.1 MeV | 2/3 | 0 | 0 | 0 | 0 |
| | d_r, d_b, d_g | 4.1 to 5.7 MeV | - 1/3 | 0 | 0 | 0 | 0 |
| 2 | c_r, c_b, c_g | 1.15 to 1.35 GeV | 2/3 | 0 | 1 | 0 | 0 |
| | s_r, s_b, s_g | 80 to 130 MeV | - 1/3 | - 1 | 0 | 0 | 0 |
| 3 | t_r, t_b, t_g | 172 to 174 GeV | 2/3 | 0 | 0 | 0 | 1 |
| | b_r, b_b, b_g | 4 to 5 GeV | - 1/3 | 0 | 0 | - 1 | 0 |

ΠΙΝΑΚΑΣ 1.2: Ιδιότητες Κουάρκ

Όπως αναφέρθηκε παραπάνω το καθιερωμένο πρότυπο εκτός από τα δομικά στοιχεία της ύλης συμπεριλαμβάνει και τις θεωρίες αλληλεπίδρασης αυτών. Κάθε μία από τις αλληλεπιδράσεις διαθέτει ένα σωματίδιο διαδότη, που είναι μποζόνια.

Για την ηλεκτρομαγνητική αλληλεπίδραση υπάρχει το φωτόνιο, είναι ουδέτερο ως προς το φορτίο και άμαζο. Επειδή ακριβώς δεν έχει μάζα η εμβέλεια της ηλεκτρομαγνητικής αλληλεπίδρασης είναι άπειρη. Για την ασθενή αλληλεπίδραση υπάρχουν τρία μποζόνια διαδότες, τα W^\pm για τα φορτισμένα ρεύματα και το Z για τα ουδέτερα. Οι διαδότες της ισχυρής αλληλεπίδρασης είναι τα γλουόνια και είναι οχτώ. Τα γλουόνια κουβαλάνε χρώμα, και για αυτό δεν υπάρχουν ελεύθερα στη φύση, όπως ακριβώς και τα κουάρκ. Επίσης, εξαιτίας του χρώματος αλληλεπιδρούν μεταξύ τους χωρίς να είναι απαραίτητη η παρουσία κουάρκ. Τα ανιχνεύουμε μέσα σε αδρόνια ή σε «άχρωμους» συνδυασμούς με άλλα γλουόνια. Πειράματα

(deep inelastic scattering) έδειξαν ότι το 50% της ορμής του πρωτονίου το κουβαλάνε τα γλουόνια.

| Particle | Polarization states | Mass |
|-------------------|---------------------|-------------------------|
| γ (photon) | 2 | 0 (theoretical) |
| g (gluon) | 2 | 0 (theoretical) |
| W^\pm | 3 | 80.399 ± 0.023 GeV |
| Z^0 | 3 | 91.187 ± 0.0021 GeV |

ΠΙΝΑΚΑΣ 1.3: Ιδιότητες Μποζονίων

1.2 Κβαντική Ηλεκτροδυναμική-QED

Η λαγκραντζιανή για ένα σωματίδιο με σπιν $\frac{1}{2}$, θεωρώντας ένα σπινόρα πεδίου Ψ δίνεται:

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi \quad (1.1)$$

είναι αναλλοίωτη κάτω από ολικούς μετασχηματισμούς $\Psi \rightarrow e^{i\theta}\Psi$ (global phase transformation), όπου $\theta \in R$. Τότε $\bar{\Psi} \rightarrow e^{-i\theta}\bar{\Psi}$ και άρα $\bar{\Psi}\Psi = e^{-i\theta}\bar{\Psi}e^{i\theta}\Psi = \bar{\Psi}\Psi$, δηλαδή οι εκθετικοί όροι αλληλοακυρώνονται. Αυτή η συμμετρία ανάγεται σε τοπική αν εφαρμόσουμε το μετασχηματισμό $\theta \rightarrow \theta(x)$:

$$\Psi(x) \rightarrow \Psi'(x) = e^{i\theta(x)}\Psi(x)$$

$$\bar{\Psi}(x) \rightarrow \bar{\Psi}'(x) = e^{-i\theta(x)}\bar{\Psi}(x)$$

δηλαδή η φάση αλλάζει σε διαφορετικά χωροχρονικά σημεία.

Τώρα η λαγκραντζιανή δεν είναι αναλλοίωτη γιατί έχουμε έναν επιπλέον όρο:

$$\partial_\mu(\Psi') = \partial_\mu(e^{i\theta(x)}\Psi) = i(\partial_\mu\theta(x))e^{i\theta(x)}\Psi + e^{i\theta(x)}\partial_\mu\Psi$$

Απαιτώντας ξανά η θεωρία να είναι αναλλοίωτη εισάγουμε την παράγωγο:

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu + ieA_\mu$$

όπου e είναι ελεύθερη παράμετρος που την ταυτοποιούμε με το ηλεκτρικό φορτίο και A_μ είναι διανυσματικό πεδίο βαθμίδας. Το A_μ μετασχηματίζεται:

$$A_\mu(x) \rightarrow A'_\mu(x) \equiv A_\mu(x) - \frac{1}{e} \partial_\mu \theta(x) \quad (1.2)$$

Η πλήρης λαγκραντζιανή όμως πρέπει να περιλαμβάνει ένα ελεύθερο όρο για το διανυσματικό πεδίο A_μ . Χρησιμοποιείται η Proca Lagrangian

$$\mathcal{L} = -\frac{1}{16\pi} F^{\mu\nu} F_{\mu\nu} + \frac{1}{8\pi} \left(\frac{m_A c}{\hbar}\right)^2 A^\nu A_\nu \quad (1.3)$$

Όμως, ενώ ο πρώτος όρος είναι αναλλοίωτος κάτω από το μετασχησμό (1.2), ο όρος $A^\nu A_\nu$ δεν είναι. Θέτωντας τη μάζα του πεδίου ίση με μηδέν ($m_A = 0$) η (1.3) μένει αναλλοίωτη.

Επομένως, αν απαιτήσουμε αναλλοιότητα για τη Dirac λαγκραντζιανή, επιβάλλεται να εισαχθεί ένα άμαζο διανυσματικό πεδίο A_μ . Το πεδίο αυτό είναι το ηλεκτρομαγνητικό δυναμικό και η ολική λαγκραντζιανή δίνεται από τη σχέση:

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi - \frac{1}{16\pi}F^{\mu\nu}F_{\mu\nu} - e\bar{\Psi}\gamma^\mu\Psi A_\mu \quad (1.4)$$

με πυκνότητα ρεύματος $J^\mu = e\bar{\Psi}\gamma^\mu\Psi$.

1.3 Κβαντική Χρωμοδυναμική-QCD

Για την κβαντική χρωμοδυναμική έχουμε τη λαγκραντζιανή:

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi \quad (1.5)$$

όμως εδώ το Ψ έχει τρεις συνιστώσες (μία για κάθε χρώμα) $\Psi \equiv \begin{pmatrix} \Psi_r \\ \Psi_b \\ \Psi_g \end{pmatrix}$ και καθεμία από αυτές είναι ένας σπίνορας τεσσάρων συνιστωσών[2]. Απαιτείται να είναι αναλλοίωτη κάτω από τον τοπικό μετασχηματισμό της μορφής:

$$\Psi \rightarrow \Psi' = e^{i\theta_a(x)} e^{i\frac{\lambda_a}{2}} \Psi \quad (1.6)$$

όπου λ_a με $a = 1, \dots, 8$ είναι οχτώ 3×3 πίνακες Gell-Mann, άιχνοι και γραμμικώς ανεξάρτητοι, οι οποίοι είναι γενίκευση των πινάκων Pauli και υπακούουν στη σχέση μετάθεσης: $[\frac{\lambda_a}{2}, \frac{\lambda_b}{2}] = if_{abc} \frac{\lambda_c}{2}$, όπου f_{abc} είναι πραγματικές σταθερές, οι παράγοντες δομής. Όμως, η παράγωγος είναι:

$$\partial_\mu(\Psi') = \partial_\mu(e^{i\theta(x)\frac{\lambda_a}{2}}\Psi) = i(\partial_\mu\theta(x))\frac{\lambda_a}{2}e^{i\theta(x)\frac{\lambda_a}{2}}\Psi + e^{i\theta(x)\frac{\lambda_a}{2}}\partial_\mu\Psi$$

και η λαγκραντζιανή δεν είναι αναλλοίωτη. Για αυτό εισάγεται η συναλλοίωτη παράγωγος:

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu + ig\frac{\lambda_a}{2}G_\mu^a$$

,έτσι ώστε η παράγωγος να μετασχηματίζεται όπως το πεδίο. Όπου, g η σταθερά σύζευξης και G_μ^a οχτώ πεδία βαθμίδας, τα οχτώ γλουόνια. Αυτά μετασχηματίζονται:

$$G_\mu^a \rightarrow G_\mu^{a'} = G_\mu^a - \frac{1}{g}\partial_\mu\theta_a(x) - f_{abc}\theta_b G_\mu^c$$

Τελικά, προσθέτοντας και ένα αναλλοίωτο κινητικό όρο για κάθε πεδίο G_μ^a , η λαγκραντζιανή για την QCD γίνεται:

$$\mathcal{L} = i\bar{\Psi}\gamma^\mu\partial_\mu\Psi - m\bar{\Psi}\Psi - g\bar{\Psi}\gamma^\mu\frac{\lambda_a}{2}\Psi G_\mu^a - \frac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a \quad (1.7)$$

,όπου περιγράφει την αλληλεπίδραση ενός κουάρκ κάποιου χρώματος με ένα γλουόνιο G_μ με σταθερά σύζευξης g . Ο ταυιστής δύναμης πεδίου $G_{\mu\nu}^a$, που εμφανίζεται στον τελευταίο όρο, δίνεται από τη σχέση:

$$G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - gf_{abc}G_\mu^b G_\nu^c \quad (1.8)$$

Παρατηρείται ότι δεν είναι «καθαρά» κινητικός όρος αλλά περιλαμβάνει έναν όρο αλληλεπίδρασης του γλουονίου με τον εαυτό του. Αυτό δείχνει ότι τα γλουόνια φέρουν φορτίο χρώματος τα ίδια.

1.4 Ηλεκτρασθενής Αλληλεπίδραση

Η συμμετρία $SU(2)$ σαν συμμετρία βαθμίδας, χρησιμοποιείται για να περιγράψει τις ασθενείς αλληλεπιδράσεις. Εισάγεται ο κβαντικός αριθμός του ασθενούς ισοτοπικού σπιν T [3]. Η παραβίαση της ομοτιμίας στις ασθενείς αλληλεπιδράσεις καθορίζει την ταξινόμηση των σωματιδίων. Οι δεξιόστροφες συνιστώσες των λεπτονίων και των κουάρκ αποτελούν τετριμμένες αναπαραστάσεις:

$$\begin{pmatrix} e_R^- \\ \mu_R^- \\ \tau_R^- \end{pmatrix} \quad \begin{pmatrix} u_R \\ c_R \\ t_R \end{pmatrix} \quad \begin{pmatrix} d_R \\ s_R \\ b_R \end{pmatrix}, \text{ με ασθενές ισοτοπικό σπιν } T = 0.$$

Οι αριστερόστροφες συνιστώσες αποτελούν θεμελιώδεις αναπαραστάσεις:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L$$

$$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L, \text{ με ασθενές ισοτοπικό σπιν } T = \frac{1}{2}.$$

Τα πεδία βαθμίδας W_μ^a , με $a = 1, 2, 3$ έχουν $T = 1$ και οι καταστάσεις δίνονται από τις σχέσεις:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \\ W_\mu^0 &= W_\mu^3 \end{aligned} \quad (1.9)$$

Τα ασθενή ρεύματα αφορούν μόνο αριστερόστροφες καταστάσεις της ύλης, και η συμμετρία βαθμίδας συμβολίζεται $SU(2)_L$ με τοπικούς μετασχηματισμούς:

$$\Psi \rightarrow \Psi' = e^{i\frac{\tau_a}{2}\theta_a(x)}\Psi \quad (1.10)$$

και η παράγωγος μετασχηματίζεται:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - ig_2 \frac{\tau_a}{2} W_\mu^a \quad (1.11)$$

Ονομάζεται υπερφορτίο Y ο κβαντικός αριθμός της $U(1)$ συμμετρίας ($U(1)_Y$). Το ηλεκτρικό φορτίο συνδέεται με το ασθενές ισοτοπικό σπιν και το υπερφορτίο μέσω της σχέσης: $Q = T_3 + \frac{Y}{2}$.

Συνολικά έχουμε $SU(2)_L \otimes U(1)_Y$, όπου το γινόμενο αυτό περιγράφει τις ηλεκτρασθενείς αλληλεπιδράσεις.

Όμως τα μποζόνια W^\pm και Z έχουν μάζα. Την αποκτούν μέσω του αυθόρμητου σπασίματος συμμετρίας $SU(2)_L \otimes U(1)_Y$ και του μηχανισμού Higgs. Εισάγεται η διπλέτα για βαθμωτά πεδία:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix}$$

με δυναμικό

$$V(\Phi) = \mu^2(\Phi^\dagger\Phi) + \lambda(\Phi^\dagger\Phi)^2.$$

Ελαχιστοποιώντας το δυναμικό: $|\Phi|^2 = (\Phi^\dagger\Phi) = \frac{1}{2}(\Phi_1^2 + \Phi_2^2 + \Phi_3^2 + \Phi_4^2) = \frac{\mu^2}{2\lambda}$

Ένα από τα απείρα σημεία είναι: $\Phi_1 = \Phi_2 = \Phi_4 = 0$, $\Phi_3^2 = \frac{\mu^2}{\lambda} = u^2$.

Άρα έχουμε για το κενό: $\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ u \end{pmatrix}$ και αφού διαταραχθεί:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ u + h(x) \end{pmatrix}. \quad (1.12)$$

Η επιλογή για το κενό έσπασε τη συμμετρία $SU(2)_L \otimes U(1)_Y$ των ηλεκτρασθενών αλληλεπιδράσεων, αλλά παραμένει η συμμετρία $U(1)_Q$ του ηλεκτρομαγνητισμού. Επομένως, το ηλεκτρικό φορτίο διατηρείται και ο μηχανισμός Higgs θα αφήσει το φωτόνιο άμαζο διανυσματικό πεδίο βαθμίδας.

Η λαγκραντζιανή που περιέχει τις αλληλεπιδράσεις του πεδίου Φ με τα πεδία βαθμίδας είναι:

$$\mathcal{L}_\Phi = (D_\mu \Phi)^\dagger (D_\mu \Phi) - (\mu^2 (\Phi^\dagger \Phi) + \lambda (\Phi^\dagger \Phi)^2) \quad (1.13)$$

όπου $\mathcal{D}_\mu = \partial_\mu + \frac{1}{2}i g_1 B_\mu + \frac{1}{2}i g_2 \tau_a W_\mu^a$ η συναλλοίωτη παράγωγος.

Τα δύο ουδέτερα ρεύματα W_μ^3 και B_μ αναμιγνύονται και δημιουργούν έναν άμαζο γραμμικό συνδυασμό και έναν ορθογώνιο συνδιασμό με μάζα.

$$\begin{aligned} A_\mu &= \sin(\theta_W) W_\mu^3 + \cos(\theta_W) B_\mu \\ Z_\mu &= \cos(\theta_W) W_\mu^3 - \sin(\theta_W) B_\mu \end{aligned} \quad (1.14)$$

Με αντικατάσταση των σχέσεων 1.9, 1.12 και 1.14 στην λαγκραντζιανή 1.13 η λαγκραντζιανή γίνεται:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)(\partial_\mu h) + \frac{1}{4}g_2^2 u^2 W_\mu^+ W_\mu^- + \frac{1}{4}(g_1^2 + g_2^2)u^2 Z_\mu Z^\mu - \lambda u^2 h^2 \quad (1.15)$$

+ όροι αλληλεπίδρασης 3^{ης} και 4^{ης} τάξης + σταθερές

Και από την λαγκραντζιανή 1.15 παρατηρείται ότι τα πεδία παίρνουν μάζα.

$$M_W = \frac{1}{2}g_2 u$$

$$M_Z = \frac{1}{2}u\sqrt{g_1^2 + g_2^2}$$

και ο λόγος είναι:

$$\frac{M_W}{M_Z} = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = \cos \theta_W, \quad (1.16)$$

όπου θ_W είναι η γωνία μίξης των ηλεκτρασθενών αλληλεπιδράσεων.

Ο μηχανισμός Higgs δίνει μάζα στα τρία μποζόνια αφήνοντας το φωτόνιο των ηλεκτρομαγνητικών αλληλεπιδράσεων άμαζο. Επίσης στην λαγκραντζιανή (1.15) παραμένει ένα φυσικό βαθμωτό πεδίο h , το μποζόνιο Higgs με μάζα:

$$M_h = u\sqrt{2\lambda}$$

Οι μάζες όλων των παραπάνω σωματιδίων έχουν μετρηθεί σε πειράματα στο CERN με τελευταία τη μέτρηση του Higgs σωματιδίου το 2012 στα πειράματα CMS και ATLAS του LHC, και οι μάζες τους είναι: $M_W = 80,4\text{GeV}$, $M_Z = 91,2\text{GeV}$ και $M_{Higgs} = 125\text{GeV}$.

Τα φερμιόνια παίρνουν και αυτά μάζα μέσω του μηχανισμού Higgs και του αυθόρμητου σπασίματος συμμετρίας στο μέρος της λαγκραντζιανής που περιέχει τους όρους αλληλεπιδράσεις των φερμιονίων (\mathcal{L}_{YUK}).

Ανάμιξη Γεννιών

Στα λεπτόνια, η σύζευξή τους με τα W^\pm πραγματοποιείται εντός της ίδιας οικογένειας (διατήρηση λεπτονικού αριθμού). Στα κουάρκ, η σύζευξη με τα μποζόνια W^\pm δε σέβεται το διαχωρισμό αυτό. Στα φορτισμένα ρεύματα των ασθενών αλληλεπιδράσεων δεν διατηρείται η γεύση των κουάρκ.

Πρώτα ο Cabibbo (1963) πρότεινε ότι τα d,s κουάρκ αναμιγνύονται έχοντας στραφεί κατά γωνία θ_c . Στην ουσία τα κουάρκ που αλληλεπιδρούν δεν είναι μόνο το d ή το s αλλά ο γραμμικός συνδιασμός αυτών. Αργότερα, οι Kobayashi και Maskawa (1974) γενίκευσαν τα παραπάνω για τις τρεις γενιές των κουάρκ. Συνολικά ο μετασχηματισμός δίνεται από τον πίνακα CKM:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} U_{ud} & U_{us} & U_{ub} \\ U_{cd} & U_{cs} & U_{cb} \\ U_{td} & U_{ts} & U_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (1.17)$$

όπου U_{ij} είναι η σύζευξη από το i στο j κουάρκ.

Τα κουάρκ d,s,b είναι στη φυσική βάση, ενώ τα d',s',b' είναι στη βάση των ρευμάτων. Με τις ιδιοκαταστάσεις της φυσικής βάσης περιγράφεται ένα ελεύθερο κουάρκ που διαδίδεται, ενώ στη βάση των ρευμάτων περιγράφεται ένα κουάρκ που αλληλεπιδρά.

Κεφάλαιο 2

Πείραμα

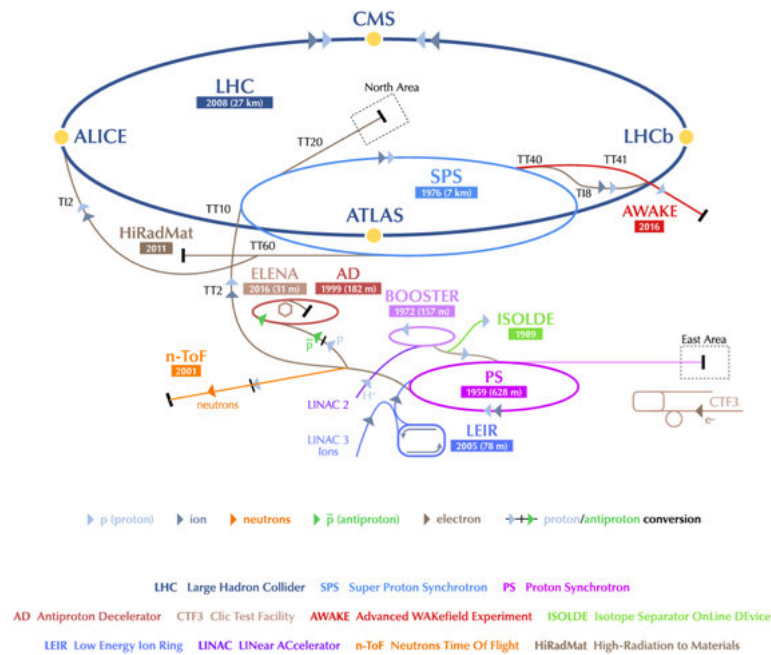
2.1 Ο επιταχυντής LHC-Large Hadron Collider

Ο επιταχυντής αδρονίων Large Hadron Collider (LHC) είναι ο μεγαλύτερος και ισχυρότερος επιταχυντής σωματιδίων του κόσμου [4]. Μπήκε σε λειτουργία το Σεπτέμβριο του 2008. Ο LHC αποτελείται από 27km δακτύλιο, περίπου 100m κάτω από την επιφάνεια της γης, μέσα στον οποίο δύο δέσμες σωματιδίων υψηλής ενέργειας ταξιδεύουν με ταχύτητα που προσεγγίζει την ταχύτητα του φωτός πριν τελικά συγκρουστούν. Η ενεργειακή πυκνότητα και η θερμοκρασία που παράγονται στις συγκρούσεις στον LHC είναι παρόμοιες με εκείνες που υπήρχαν λίγα λεπτά μετά το Big Bang.

Στο πρώτο μέρος του επιταχυντή, ένα ηλεκτρικό πεδίο απογυμνώνει άτομα υδρογόνου (αποτελούμενα από ένα πρωτόνιο και ένα ηλεκτρόνιο) από τα ηλεκτρόνιά τους [5]. Τα ηλεκτρικά πεδία κατά μήκος του επιταχυντή αλλάζουν από θετικό σε αρνητικό σε μια δεδομένη συχνότητα, τραβώντας φορτισμένα σωματίδια προς τα εμπρός κατά μήκος του επιταχυντή. Η συχνότητα της αλλαγής ελέγχεται για να διασφαλιστεί ότι τα σωματίδια επιταχύνονται όχι σε συνεχή ροή, αλλά σε «πακέτα» σωματιδίων.

Χρησιμοποιούνται υπεραγώγιμοι μαγνήτες ώστε να επιταχύνουν και να κατευθύνουν τα σωματίδια των δεσμών. Οι δέσμες πρωτονίων ταξιδεύουν χωριστά σε αντίθετες κατευθύνσεις και διαφορετικούς σωλήνες και διατηρούνται σε εξαιρετικά υψηλό κενό. Για να δουλεύουν οι ηλεκτρομαγνήτες χωρίς αντίσταση ή απώλειες ενέργειας απαιτείται ψύξη των μαγνητών στους $-271,3^{\circ}C$. Για την κατεύθυνση της δέσμης διαφορετικά είδη μαγνητών χρησιμοποιούνται. Υπάρχουν 1232 διπολικοί μαγνήτες μήκους 15m που στρέφουν τις δέσμες μέσα στο δακτύλιο, και 392 τετραπολικοί μαγνήτες μήκους 5-7m που εστιάζουν τις δέσμες ώστε να επιτευχθούν συγκρούσεις. Σύγκρουση των δεσμών πραγματοποιείται σε τέσσερα σημεία γύρω από το δακτύλιο του επιταχυντή, όπου είναι τοποθετημένοι οι ανιχνευτές σωματιδίων. Τα πειράματα αυτά είναι το CMS, το ATLAS, το ALICE και το LHCb.

Στην εικόνα 2.1 φαίνεται σχηματικά ο LHC και τα ανιχνευτικά πειράματα που λαμβάνουν μέρος.



ΣΧΗΜΑ 2.1: Σχηματική αναπαράσταση του μεγάλου αδρονικού επιταχυντή (LHC) του CERN

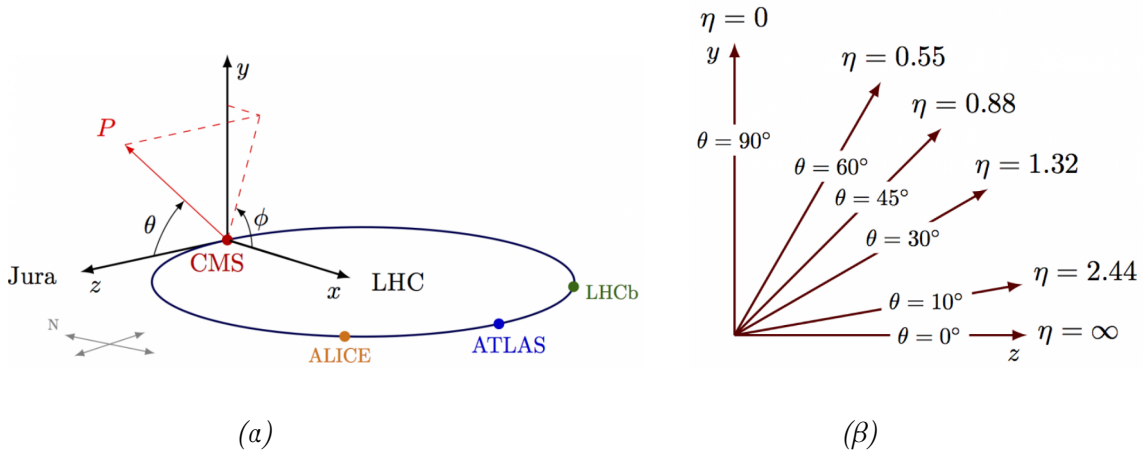
2.2 Ο ανιχνευτής CMS (Compact Muon Solenoid)

Ο ανιχνευτής CMS είναι μήκους 21,6 m, διαμέτρου 14,6m και συνολικού βάρους 12500 τόνων[6]. Το κεντρικό χαρακτηριστικό του CMS είναι ένα υπεραγωγίμο σωληνοειδές με εσωτερική διάμετρο 6 m, που παρέχει ένα μαγνητικό πεδίο 3,8 T [7]. Μέσα στον όγκο του πεδίου είναι ο ανιχνευτής τροχιών. Αυτός αποτελείται από επιμέρους κομμάτια: ένα με πίξελ πυριτίου (silicon pixel) και έναν ανιχνευτή στριπ πυριτίου (strip tracker). Εσωτερικά του μαγνήτη βρίσκεται επιπλέον ένα ηλεκτρομαγνητικό θερμιδόμετρο (ECAL) και ένα αδρονικό θερμιδόμετρο (HCAL). Τα μίονια ανιχνεύονται σε ανιχνευτές ιονισμού αερίου που είναι ενσωματωμένοι in the steel return yoke. Εκτός από τους κυλινδρικούς ανιχνευτές και τους ανιχνευτές στα καπάκια του ανιχνευτή (endcap detectors), το CMS έχει ένα εκτεταμένο θερμιδόμετρο προς τα εμπρός.

Το σύστημα συντεταγμένων που υιοθετείται από το CMS έχει την αρχή του τοποθετημένη στο ονομαστικό σημείο σύγκρουσης μέσα στο πείραμα, με τον άξονα x να δείχνει ακτινικά προς τα μέσα, προς το κέντρο του LHC. Ο άξονας y προς τα πάνω (κάθετος στο επίπεδο του LHC) και ο z-άξονας κατά μήκος της κατεύθυνσης της δέσμης προς τα αριστερά. Η πολική γωνία θ μετράται από τον θετικό άξονα z και η αζιμουθιακή γωνία φ (σε ακτίνια) μετράται

στο επίπεδο x-y. Η ψευδοωκότητα ορίζεται ως

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



ΣΧΗΜΑ 2.2: α) Το σύστημα συντεταγμένων στο CMS και β) Ενδεικτικές τιμές της ψευδοωκότητας

Ο εσωτερικός ανιχνευτής τροχιών, μετρά τις φορτισμένες τροχιές σωματιδίων στην περιοχή της ψευδοωκότητας $|\eta| < 2,5$. Χρησιμοποιείται για να ανιχνεύσει την ορμή ενός σωματιδίου και τη διαδρομή που ακολουθεί μέσα στο μαγνητικό πεδίο. Όσο περισσότερο κυρτή είναι η διαδρομή που ακολουθεί το σωματίδιο, τόσο μικρότερη είναι και η ορμή του σωματιδίου [8]. Αποτελείται από 1440 πίξελ πυριτίου και 15148 ανιχνευτικές μονάδες από λωρίδες πυριτίου. Παρέχει ανάλυση παραμέτρων χρούσης $\approx 15 \mu\text{m}$ και ανάλυση εγκάρσιας ορμής (p_T) περίπου 1% για φορτισμένα σωματίδια με ορμή $p_T \approx 40 \text{ GeV}$.

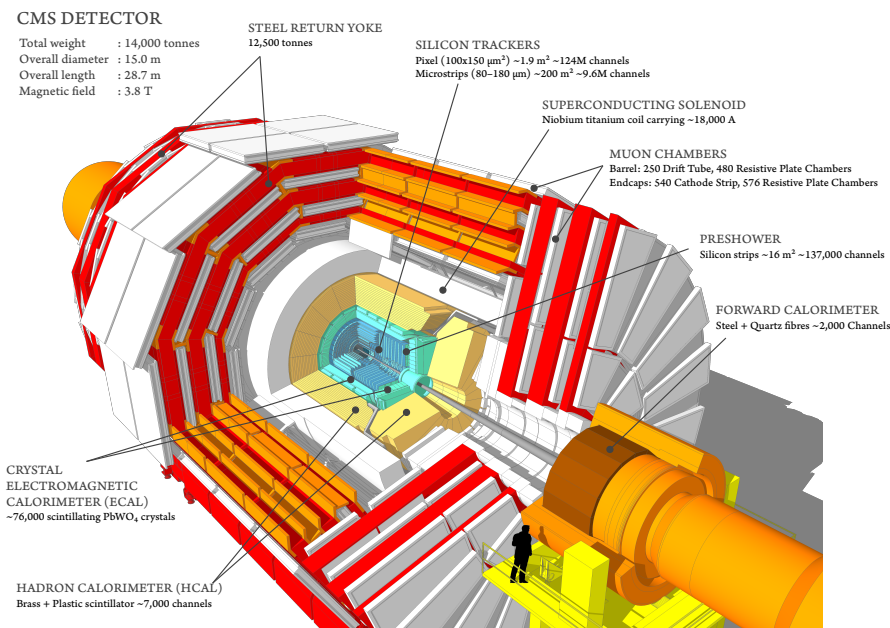
Το ηλεκτρομαγνητικό θερμιδόμετρο (ECAL) αποτελείται από σχεδόν 76000 κρυστάλλους μόλυβδου βολφραμίου, οι οποίοι παρέχουν κάλυψη στο χώρο για τιμές της ψευδοωκότητας $|\eta| < 1,479$ στην κυλινδρική περιοχή του ανιχνευτή (βαρέλι) και για $1,479 < |\eta| < 3,0$ στις δύο περιοχές που υπάρχουν τα καπάκια. Ένας ανιχνευτής ηλεκτρομαγνητικού καταιγισμού, αποτελούμενος από δύο επίπεδα αισθητήρων πυριτίου, στα οποία παρεμβάλλεται μόλυβδος συνολικά τριών μηκών ακτινοβολίας, βρίσκεται μπροστά από τα καπάκια. Χρησιμοποιείται για την ανίχνευση ενεργειακής ροής, φορτισμένων σωματιδίων και ουδέτερων σωματιδίων. Για να μετρηθεί η ολική ενέργεια του φορτισμένου σωματιδίου ή του φωτονίου, το σωματίδιο πρέπει να σταματηθεί πλήρως μέσα στον ανιχνευτή. Το ηλεκτρομαγνητικό θερμιδόμετρο έχει ενεργειακή ανάλυση καλύτερη από 0,5% για τα φωτόνια με εγκάρσιες ενέργειες (E_T) πάνω από 100 GeV. Η ενεργειακή ανάλυση είναι 3% ή καλύτερη για το εύρος των ενεργειών ηλεκτρονίων.

Τα αδρονικά θερμιδόμετρα (HCAL) στο βαρέλι και του μπροστινού καλύμματος είναι συσκευές δειγματοληψίας με ορείχαλκο ως παθητικό υλικό και σπινθηριστή ως το δραστικό

υλικό. Τα συνδυασμένα κομμάτια του θερμιδόμετρου ομαδοποιούνται σε προβολικούς πύργους. Τα αδρονικά καλορίμετρα χρησιμοποιούνται για να μετρήσουν την ενέργεια των αδρονίων, δηλαδή σωματιδίων που αποτελούνται από κουάρκ και γλουόνια. Επίσης, παρέχει μία έμμεση μέτρηση για την ύπαρξη ή μη από μη αλληλεπιδρώντα αφόρτιστα σωματίδια όπως για παράδειγμα τα νετρίνα. Η ενέργεια των φορτισμένων πιονίων και άλλων σχεδόν σταθερών αδρόνων μπορεί να μετρηθεί με τα θερμιδόμετρα (ECAL και HCAL συνδυασμένα)

Για φορτισμένα αδρόνια, η ανάλυση του θερμιδόμετρου βελτιώνεται σε συνδυασμό με την ανάλυση της ορμής από τον ανιχνευτή τροχιών μόνο για ορμή (p_T) που υπερβαίνει τα 500 GeV. Η ενεργειακή ανάλυση για τους πίδακες και την υπολειπόμενη εγκάρσια ενέργεια βελτιώνεται ουσιαστικά σε σχέση με την θερμιδομετρική ανακατασκευή χρησιμοποιώντας τον αλγόριθμο ροής σωματιδίων (particle flow-PF) που συνίσταται στην ανακατασκευή και αναγνώριση κάθε μεμονωμένου σωματιδίου με έναν βελτιστοποιημένο συνδυασμό όλων των πληροφοριών των υποανιχνευτών. Αυτή η προσέγγιση εκμεταλλεύεται την πολύ καλή ανάλυση της ορμής από τον ανιχνευτή τροχιάς, για τη βελτίωση της μέτρησης της ενέργειας των φορτισμένων αδρόνων.

Τα μόνια ανιχνεύονται στην περιοχή της ψευδοωκύτητας $|\eta| < 2,4$, με επίπεδα ανίχνευσης που βασίζονται σε τρεις τεχνολογίες: drift tubes, cathode strip chambers, και resistive plate chambers. Ένα μόνιο υψηλής ορμής που προέρχεται από το σημείο αλληλεπίδρασης παράγει τμήματα τροχιάς συνήθως σε τρεις ή σε τέσσερις σταθμούς στον ανιχνευτή μιονίων. Η αντιστοίχιση αυτών των τμημάτων με κομμάτια που μετρήθηκαν στον εσωτερικό ανιχνευτή τροχιών οδηγεί σε ανάλυση ορμής μεταξύ 1 και 2% για τιμές της ορμής έως 100 GeV.



ΣΧΗΜΑ 2.3: Διατομή ενός τετάρτου του ανιχνευτή CMS

Το πρώτο επίπεδο του συστήματος σκανδαλισμού (level 1 trigger) του CMS, που αποτελείται από hardware επεξεργαστές, έχει σχεδιαστεί για να επιλέγει τα πιο ενδιαφέροντα γεγονότα του πειράματος σε λιγότερο από 1 μs , χρησιμοποιώντας πληροφορίες από τα θερμιδόμετρα και τους ανιχνευτές μιονίων. Ο υψηλού επιπέδου σκανδαλιστής (High Level Trigger-HLT), που αποτελείται από software, μειώνει περαιτέρω τον ρυθμό γεγονότων από 100 kHz σε μερικές εκατοντάδες Hz πριν την αποθήκευση δεδομένων.

2.3 Συγκρούσεις Πρωτονίων Πρωτονίων

Στο πείραμα του LHC έχουμε συγκρούσεις δεσμών πρωτονίων-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} [8]. Η τιμή του s , καθορίζεται από την ορμή των συγκρουόμενων πρωτονίων μέσω της σχέσης:

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

,όπου p_i είναι οι τετραορμές των δύο συγκρουόμενων πρωτονίων, E_i οι ενέργειές τους και \vec{p}_i η τριορμή τους.

Επειδή οι δύο δέσμες πρωτονίων έχουν την ίδια ενέργεια και αντίθετη φορά, ισχύει: $E_1 = E_2 = E_{proton}$ και $p_1 \approx -p_2$. Επομένως, η ενέργεια κέντρου μάζας γίνεται: $\sqrt{s} = 2E_{proton}$.

Τα πρωτόνια είναι σύνθετα σωματίδια και η ακριβής διασπορά περιλαμβάνει κουάρκ και γλουόνια, δηλαδή παρτόνια, που έχουν μόνο ένα πολύ μικρό ποσοστό της αρχικής ορμής της δέσμης των πρωτονίων. Η ενέργεια κέντρου μάζας της διαδικασίας της σκληρής διασποράς (hard scattering) $\sqrt{s'}$ είναι μόνο ένα μικρό ποσοστό από τα 13 TeV. Αν υποθέσει κανείς ότι οι μάζες των παρτονίων είναι μικρές σε σχέση με την ορμή τους, η ενέργεια κέντρου μάζας της διαδικασίας αυτής είναι: $\sqrt{s'} = \sqrt{x_1x_2}\sqrt{s}$, εδώ x_1 και x_2 προσδιορίζουν το ποσοστό της ορμής που έχουν τα δύο 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 που δεν εξαρτώνται από την αρχική ορμή στην κατεύθυνση της δέσμης των πρωτονίων.

2.4 Μιόνια

Ο αλγόριθμος ανακατασκευής μιονίων συνδυάζει πληροφορία από το silicon tracker και το muon spectrometer [9]. Τα μιόνια συλλέγονται εφαρμόζοντας προαπαιτούμενα όσον αφορά την τροχιά τους και την αντίστοιχη ενέργεια που εναποθέτουν στα θερμιδόμετρα. Ανακατασκευάζονται με όριο στην ψευδοωκύτητα $|n| < 2,1$ και με εγκάρσια ορμή $p_t > 30\text{GeV}$.

Τα μόνια από διάσπαση του μποζονίου W είναι συνήθως καλά απομονωμένα (isolated) σε σχέση με άλλες τροχιές στο γεγονός (event) και μία προσέγγιση βασισμένη στον particle flow-PF αλγόριθμο χρησιμοποιείται για να υπολογιστεί ο βαθμός του isolation. Για να καθοριστεί χρησιμοποιούνται όλα τα φορτισμένα σωματίδια, ενώ τα φωτόνια και τα ουδέτερα αδρόνια χρησιμοποιούνται στο άθροισμα μόνο αν έχουν εγκάρσια ενέργεια μεγαλύτερη από $0,5$ ($GeV E_t > 0,5 GeV$). Οι μεταβλητές για το isolation αθροίζουν τη συνεισφορά των παραπάνω σωματιδίων σε ένα κώνο ακτίνας $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0,4$ γύρω από τον άξονα του λεπτονίου. Ο τύπος που βασίζεται στα σωματίδια, για το isolation δίνεται:

$$I_{rel}^{PF} = \frac{[\sum p_T^{charged} + \max(0, \sum E_T^\gamma + \sum E_T^{neutral} - 0,5 \sum E_T^{PU})]}{p_T^l} \quad (2.1)$$

,όπου $\sum p_T^{charged}$ είναι το άθροισμα των εγκάρσιων ορμών των φορτισμένων αδρονίων και $\sum E_T^i$ με $i=\gamma, neutral, PU$ είναι το άθροισμα των εγκάρσιων ενεργειών των φωτονίων, των ουδέτερων αδρονίων και του pileup αντίστοιχα.

Ως pileup εγκάρσια ενέργεια θεωρείται το άθροισμα των συνεισφορών των φορτισμένων σωματιδίων, στο εσωτερικό του κώνου, που δεν προέρχονται από την κύρια σύγκρουση. Ο παράγοντας $0,5$ αντιστοιχεί περίπου στο ρυθμό παραγωγής ουδέτερων προς φορτισμένων αδρονίων στη διαδικασία της αδρονοποίησης από αλληλεπιδράσεις pileup.

Τα μόνια επιλέγονται με απαίτηση να περάσουν το Tight Isolation κριτήριο $I_{rel}^{PF} < 0,15$ και το Loose Isolation κριτήριο $I_{rel}^{PF} < 0,25$ χρησιμοποιείται για δεύτερη απόρριψη λεπτονίων.

2.5 Ηλεκτρόνια

Τα ηλεκτρόνια ανακατασκευάζονται σχετίζοντας μία ανακατασκευασμένη τροχιά στον ανιχνευτή πυριτίου με ένα τμήμα ενέργειας (cluster) στο ηλεκτρομαγνητικό θερμιδόμετρο (ECAL).

Ένας συνδυασμός μίας ανεξάρτητης προσέγγισης και του αλγορίθμου particle-flow συμπληρωματικά χρησιμοποιείται για την μεγιστοποίηση της απόδοσης[10].

2.5.1 Ομαδοποίηση των εναποθέσεων ενέργειας του ηλεκτρονίου στο ηλεκτρομαγνητικό θερμιδόμετρο ECAL

Η ενέργεια ηλεκτρονίων συνήθως απλώνεται σε πολλούς κρυστάλλους του ECAL. Αυτή η εξάπλωση μπορεί να είναι αρκετά μικρή όταν τα ηλεκτρόνια χάνουν λίγη ενέργεια μέσω του bremsstrahlung πριν φτάσουν στο ECAL. Για ένα ηλεκτρόνιο που παράγεται εντός CMS, το αποτέλεσμα που προκαλείται από την ακτινοβολία των φωτονίων μπορεί να είναι μεγάλο: κατά μέσο όρο, το 33% της ενέργειας των ηλεκτρονίων εκπέμπεται πριν φτάσει στο ηλεκτρομαγνητικό θερμιδόμετρο όπου το παρεμβαλλόμενο υλικό είναι ελάχιστο ($n \approx 0$)

και περίπου 86% της ενέργειας του ακτινοβολείται όταν το παρεμβαλλόμενο υλικό είναι το μεγαλύτερο ($|n| \approx 1,4$).

Για να μετρηθεί με ακρίβεια η αρχική ενέργεια του ηλεκτρονίου, είναι απαραίτητο να συλλεχθεί η ενέργεια των φωτονίων που ακτινοβολήθηκαν η οποία απλώνεται κυρίως κατά μήκος της κατεύθυνσης ϕ λόγω της κάμψης της τροχιάς των ηλεκτρονίων στο μαγνητικό πεδίο. Η διασπορά στην κατεύθυνση η είναι συνήθως αμελητέα. Δύο αλγόριθμοι ομαδοποίησης (clustering), ο «hybrid» αλγόριθμος στο βαρέλι (barrel), και ο «multi-5 × 5» στα καπάκια (endcaps), χρησιμοποιούνται για το σκοπό αυτό. Για το βήμα ομαδοποίησης, οι κατευθύνσεις η και ϕ και E_T καθορίζονται σε σχέση με το κέντρο του CMS.

Ο hybrid αλγόριθμος εκμεταλλεύεται τη γεωμετρία του ηλεκτρομαγνητικού θερμιδόμετρου στο βαρέλι και τις ιδιότητες του σχήματος του shower, συλλέγοντας την ενέργεια σε ένα μικρό παράθυρο ως προς την ψευδοωκότητα η και ένα εκτεταμένο παράθυρο ως προς τη γωνία ϕ . Το σημείο εκκίνησης είναι ένας κρύσταλλος seed, που ορίζεται ως αυτός που περιέχει το μεγαλύτερο μέρος της ενέργειας που εναποτίθεται σε οποιαδήποτε εξεταζόμενη περιοχή, ο οποίος έχει ένα ελάχιστο E_T ($E_{T,seed} > 1\text{GeV}$). Συστοιχίες 5×1 κρυστάλλων στον $\eta \times \phi$ χώρο προστίθενται γύρω από τον αρχικό κρύσταλλο, σε ένα εύρος N_{steps} κρυστάλλων και στις δύο κατευθύνσεις του ϕ , εάν οι ενέργειές τους υπερβαίνουν ένα ελάχιστο όριο $E_{array} > 0,1\text{GeV}$. Οι συνεχόμενες συστοιχίες ομαδοποιούνται σε συστάδες (clusters), με κάθε ξεχωριστό cluster να συλλέγεται στο τελικό supercluster-SC αν ικανοποιεί κάποια επιπλέον ενεργειακά κριτήρια.

Ο αλγόριθμος multi-5 × 5 χρησιμοποιείται στο ECAL στα καπάκια του ανιχνευτή, όπου οι κρύσταλλοι δεν είναι διατεταγμένοι σε $\eta \times \phi$ γεωμετρία. Ξεκινά με τους κρυστάλλους "seeds", αυτούς με τοπική μέγιστη ενέργεια σε σχέση με τους τέσσερις άμεσους γείτονές τους, οι οποίοι πρέπει να πληρούν την απαίτηση $E_{T,seed} > 0,18\text{GeV}$. Γύρω από αυτούς τους κρυστάλλους και ξεκινώντας με το μεγαλύτερο E_T , η ενέργεια συλλέγεται σε συστάδες 5×5 κρυστάλλων, που μπορεί εν μέρει να επικαλύπτονται. Αυτά τα σμήνη ομαδοποιούνται έπειτα σε ένα supercluster-SC εάν η συνολική τους εγχάρσια ενέργεια είναι $E_T > 1\text{GeV}$, εντός ενός εύρους $\eta \approx \pm 0,07$, και ένα εύρος σε $\phi \approx \pm 0,3\text{rad}$ γύρω από κάθε αρχικό κρύσταλλο.

Η ενέργεια SC αντιστοιχεί στο άθροισμα των ενεργειών όλων των υποτημημάτων της. Η θέση του supercluster υπολογίζεται ως ο βεβαρυμένος από τις ενέργειες μέσος όρος των θέσεων των clusters.

Επιπλέον, ως μέρος του αλγορίθμου ανακατασκευής particle-flow, εισάγεται ένας άλλος αλγόριθμος ομαδοποίησης που στοχεύει στην ανακατασκευή των shower των σωματιδίων ξεχωριστά. Τα clusters του particle-flow αλγορίθμου ανακατασκευάζονται συγκεντρώνοντας γύρω από τον αρχικό κρύσταλλο όλους τους συνεχόμενους κρυστάλλους με ενέργειες

δύο τυπικών αποκλίσεων (σ) πάνω από τον ηλεκτρονικό θόρυβο που παρατηρήθηκε στην αρχή της διαδικασίας λήψης δεδομένων, με $E_{seed} > 230MeV$ στο βαρέλι και $E_{seed} > 600MeV$ ή $E_{T,seed} > 150MeV$ στα καπάκια. Μια σημαντική διαφορά σε σχέση με την ανεξάρτητη προσέγγιση είναι ότι είναι δυνατό να μοιραστεί η ενέργεια ενός κρυστάλλου μεταξύ δύο ή περισσότερων clusters. Τέτοιες συστάδες χρησιμοποιούνται σε διαφορετικά στάδια ανακατασκευής ηλεκτρονίων.

2.5.2 Ανακατασκευή της τροχιάς των ηλεκτρονίων

Οι τροχιές των ηλεκτρονίων μπορούν να ανακατασκευαστούν στον ανιχνευτή τροχιών χρησιμοποιώντας την Kalman filter (KF) διαδικασία που χρησιμοποιείται για όλα τα φορτισμένα σωματίδια. Ωστόσο, οι μεγάλες απώλειες μέσω ακτινοβολίας των ηλεκτρονίων μέσα στο υλικό του ανιχνευτή τροχιών, οδηγούν σε μειωμένη απόδοση στη συλλογή "hit" και σε κακή εκτίμηση των παραμέτρων τροχιάς. Hit χάνονται όταν η αλλαγή στην καμπυλότητα είναι μεγάλη λόγω του Bremsstrahlung, για αυτό χρησιμοποιείται μια ειδική διαδικασία ανακατασκευής της τροχιάς των ηλεκτρονίων.

Το πρώτο βήμα στην ανακατασκευή τροχιάς των ηλεκτρονίων, ονομάζεται seeding, και αποτελείται από την εύρεση και την επιλογή των δύο ή τριών πρώτων hits στον ανιχνευτή τροχιών από τα οποία μπορεί να ξεκινήσει η τροχιά. Το seeding είναι σημαντικό καθώς η διαδικασία αυτή επηρεάζει την απόδοση της ανακατασκευής. Χρησιμοποιούνται δύο συμπληρωματικοί αλγόριθμοι και τα αποτελέσματά τους συνδυάζονται. Στο ηλεκτρομαγνητικό θερμιδόμετρο το seeding ξεκινά από την ενέργεια και τη θέση του supercluster, που χρησιμοποιήθηκε για να εκτιμηθεί η τροχιά του ηλεκτρονίου και συλλεγει όλα τα seeds ηλεκτρονίων που ανακατασκευάστηκαν. Στον ανιχνευτή τροχιάς το seeding βασίζεται στις τροχιές που ανακατασκευάστηκαν από το γενικό αλγόριθμο για φορτισμένα σωματίδια, και προεκτύνονται προς το ECAL για να συμπέσουν με το SC. Στο seeding που βασίζεται στο ECAL, η ενέργεια του SC και η θέση του χρησιμοποιούνται για να κατασκευαστεί η τροχιά προς τη κορυφή της σύγκρουσης. Η προς τα πίσω αυτή ανακατασκευή της τροχιάς υλοποιείται υποθέτωντας και θετικά και αρνητικά φορτία.

Στη συνέχεια συσχετίζεται η τροχιά στον ανιχνευτή τροχιών με την ενέργεια και τη θέση στο ECAL. Για ηλεκτρόνια από το seeding στο ECAL, το cluster που συσχετίζεται με την τροχιά είναι αυτό που προέκυψε από τον "hybrid" ή από τον "multi-5 × 5" αλγόριθμο. Για ηλεκτρόνια τα οποία προέκυψαν μόνο από τον ανιχνευτή τροχιών, ο συσχετισμός γίνεται με το cluster που προέκυψε από τον Particle Flow αλγόριθμο. Η συσχέτιση ανάμεσα στην τροχιά και το cluster γίνεται μέσω κριτηρίων που διασφαλίζουν υψηλή απόδοση και μειωμένη πιθανότητα λανθασμένης ταυτοποίησης. Για τα ηλεκτρόνια που έγιναν seeded μέσω του ECAL απαιτείται να πληρούνται γεωμετρικά κριτήρια ανάμεσα στην τροχιά και το SC:

- $|\Delta\eta| = |\eta_{SC} - \eta_{in}| < 0,02$, όπου η_{SC} είναι η ψευδοωκότητα της βεβαρυμένης ως προς την ενέργεια θέσης του SC και η_{in} η ψευδοωκότητα από την πιο εσωτερική θέση και κατεύθυνση της τροχιάς ως προς τη θέση του SC.
- $|\Delta\phi| = |\phi_{SC} - \phi_{in}| < 0,15$, όπου ϕ είναι η αζιμουθιακή γωνία και οι δείκτες δηλώνουν ότι και παραπάνω.

Για τα ηλεκτρόνια που έγιναν seeded μέσω του ανιχνευτή τροχιών, μία ολική μεταβλητή ταυτοποίησης καθορίζεται χρησιμοποιώντας μια πολυπαραγοντική τεχνική που συνδυάζει γεωμετρικές και κινηματικές παραμέτρους.

2.5.3 Ταυτοποίηση ηλεκτρονίων

Στο CMS χρησιμοποιούνται διάφορες τεχνικές για να ταυτοποιηθούν τα απομονωμένα ηλεκτρόνια του σήματος και να διαχωριστούν από αυτά που προέρχονται από πηγές υποβάθρου. Τέτοιες είναι, τα ηλεκτρόνια να προέρχονται από φωτόνια, από πίδακες που λανθασμένα ταυτοποιήθηκαν ως ηλεκτρόνια, ή από ηλεκτρόνια από ημιλεπτονικές διασπάσεις των b ή c κουάρκ. Έχουν αναπτυχθεί απλοί και ισχυροί αλγόριθμοι για το διαχωρισμό αυτό αλλά και πιο σύνθετοι οι οποίοι συνδυάζουν μεταβλητές σε πολυπαραγοντικές μεθόδους για να επιτευχθεί καλύτερος διαχωρισμός.

2.5.4 Isolation των ηλεκτρονίων

Ένα μέρος του υποβάθρου για τα απομονωμένα ηλεκτρόνια της κύριας σύγκρουσης οφείλεται σε εσφαλμένη ταυτοποίηση πιδάκων ή σε ηλεκτρόνια μέσα σε ένα πίδακα που προκύπτει από ημιλεπτονικές διασπάσεις των κουάρκ b ή c . Και στις δύο περιπτώσεις, τα υποψήφια ηλεκτρόνια έχουν σημαντική ροή ενέργειας κοντά στις τροχιές τους και απαιτώντας την απομόνωση των ηλεκτρονίων (isolation) από μια τέτοια κοντινή δραστηριότητα μειώνει σημαντικά αυτές τις μορφές υποβάθρου. Οι απαιτήσεις για απομονωμένα ηλεκτρόνια διαχωρίζονται από την ταυτοποίηση των ηλεκτρονίων, καθώς η αλληλεπίδραση μεταξύ τους τείνει να εξαρτάται από την ανάλυση. Επιπλέον, η αντιστροφή των απαιτήσεων του isolation, ανεξάρτητα από εκείνες που χρησιμοποιούνται για την ταυτοποίηση, παρέχει έλεγχο σε διαφορετικές πηγές τέτοιων υποβάθρων στα δεδομένα. Δύο τεχνικές χρησιμοποιούνται στο CMS. Η απλούστερη αναφέρεται ως απομόνωση με βάση τον ανιχνευτή και βασίζεται στο άθροισμα των εναποθέσεων ενέργειας είτε στο ECAL είτε στο HCAL γύρω από κάθε τροχιά ηλεκτρονίων ή στο βαθμωτό άθροισμα της εγκάρσιας ορμής όλων των τροχιών που ανακατασκευάστηκαν από την κορυφή σύγκρουσης. Αυτά τα αθροίσματα υπολογίζονται συνήθως σε ακτίνες κώνου $\Delta R = 0,3$ ή $0,4$ γύρω από την κατεύθυνση ηλεκτρονίων και αφαιρούν τις συνεισφορές από το ηλεκτρόνιο μέσω μικρότερων κώνων αποκλεισμού. Αυτή η διαδικασία, που έχει καλή απόδοση στην απόρριψη πιδάκων που ταυτοποιούνται εσφαλμένα ως ηλεκτρόνια, χρησιμοποιείται από το HLT-High Level Trigger, και σε ορισμένες αναλύσεις στις οποίες επιτυγχάνεται μόνο ήπια απόρριψη υποβάθρου. Οι περισσότερες αναλύσεις,

ωστόσο, ωφελούνται από την τεχνική PF για τον προσδιορισμό ποσοτήτων για το isolation. Αντί να χρησιμοποιούν μετρήσεις ενέργειας σε ανεξάρτητους υπο-ανιχνευτές, η απομόνωση καθορίζεται χρησιμοποιώντας τα σωματίδια του Particle Flow (PF) που ανακατασκευάστηκαν με κατεύθυνση ορμής μέσα σε έναν επιλεγμένο κώνο. Με αυτόν τον τρόπο, μπορεί να χρησιμοποιηθεί η σωστή βαθμονόμηση και αποφεύγεται η πιθανή διπλή μέτρηση της ενέργειας που αποδίδεται σε αυτά τα σωματίδια. Όταν ένα υποψήφιο ηλεκτρόνιο αναγνωρίζεται εσφαλμένα από τον PF ως άλλο σωματίδιο, εισέρχεται στο άθροισμα για το isolation, και αυξάνει τεχνητά το μέγεθος της τιμής του αθροίσματος. Αυτό το φαινόμενο αυξάνεται όταν μειώνεται η αποτελεσματικότητα του PF. Η ταυτοποίηση ηλεκτρονίων χρησιμοποιώντας τον PF αλγόριθμο αποδίδει πολύ καλά για ηλεκτρόνια στο βαρέλι ECAL, όπου δεν απαιτούνται επιπρόσθετες διορθώσεις για την αφαίρεση της συνεισφοράς ηλεκτρονίων στο άθροισμα. Ο τύπος για τον βαθμό του isolation των ηλεκτρονίων δίνεται από τη σχέση (2.1) που ισχύει και για τα μίονια.

2.6 Αλγόριθμος Anti- k_T

Οι αλγόριθμοι ανακατασκευής πιδάκων αποτελούν ένα από τα κύρια εργαλεία για την ανάλυση δεδομένων από αδρονικές συγκρούσεις [11]. Μία σημαντική κατηγορία είναι οι αλγόριθμοι Sequential Clustering. Οι αλγόριθμοι που ανήκουν στην κατηγορία αυτή είναι ο αλγόριθμος k_T , ο Cambridge/Aachen και ο anti- k_T . Οι αλγόριθμοι αυτοί ξεκινούν με την υπόθεση ότι τα σωματίδια μέσα στους πίδακες έχουν μικρές διαφορές στην εγκάρσια ορμή τους (transverse momentum) και τα σωματίδια ομαδοποιούνται με βάση τον χώρο των ορμών. Αυτό έχει σαν αποτέλεσμα τη δημιουργία περιοχών πιδάκων οι οποίες έχουν κυμαινόμενη επιφάνεια. Ενώ αρχικά οι αλγόριθμοι αυτοί είχαν αρκετά αργή απόδοση, ήταν όμως αποτελεσματικοί, πλέον χρησιμοποιείται το πρόγραμμα FastJet μέσω του οποίου ο υπολογισμός των πιδάκων γίνεται πολύ γρηγορότερα. Οι αλγόριθμοι Sequential Clustering είναι όλοι απλοί και δεν επηρεάζονται από σωματίδια χαμηλής ενέργειας. Τέλος, σε αυτούς τους αλγόριθμους το παρατηρούμενο γεγονός παραμένει αμετάβλητο αναφορικά με συγγραμμικές εκπομπές γλουονίων χαμηλής ενέργειας. Μέσω της ιδιότητας αυτής, ο αλγόριθμος εγγυάται ακύρωση πραγματικών και εικονικών αποκλίσεων σε υπολογισμούς υψηλότερων όρων (higher order calculations). Σε αυτή την εργασία περιγράφεται ο αλγόριθμος anti- k_T .

Εισάγονται οι αποστάσεις, d_{ij} μεταξύ των σωματιδίων (ή ψευδοπιδάκων) i και j και d_{iB} μεταξύ του σωματιδίου i και της δέσμης B . Η ομαδοποίηση προχωρά προσδιορίζοντας τη μικρότερη από τις αποστάσεις και εάν είναι η d_{ij} , συνδυάζονται τα σωματίδια i και j , ενώ αν είναι η d_{iB} , ονομάζεται το i ως ένας πίδακας και απομακρύνεται από τη λίστα των σωματιδίων. Οι αποστάσεις επανυπολογίζονται και η διαδικασία επαναλαμβάνεται μέχρι να μην απομείνει κανένα σωματίδιο.

Τα μέτρα των αποστάσεων για τον anti- k_T αλγόριθμο ορίζονται:

$$d_{ij} = \min\left(\frac{1}{k_{ti}^2}, \frac{1}{k_{tj}^2}\right) \frac{\Delta R_{ij}^2}{R^2}$$

και

$$d_{iB} = \frac{1}{k_{ti}^2}$$

όπου, k_{ti} είναι η εγκάρσια ορμή, $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ είναι η απόσταση των σωματιδίων i, j στον (η, ϕ) χώρο (η : η ψευδοωκότητα και ϕ : η αζιμουθιακή γωνία).

Η λειτουργικότητα του αλγορίθμου anti- k_T μπορεί να γίνει κατανοητή εάν θεωρήσουμε ένα γεγονός με μερικά καλά διαχωρισμένα σωματίδια υψηλής ενέργειας (hard) με εγκάρσια ορμή k_{t1}, k_{t2}, \dots και πολλά σωματίδια χαμηλής ενέργειας (soft). Η απόσταση $d_{1i} = \min\left(\frac{1}{k_{t1}^2}, \frac{1}{k_{ti}^2}\right) \frac{\Delta R_{1i}^2}{R^2}$ μεταξύ ενός σωματιδίου υψηλής ενέργειας 1 και ενός σωματιδίου χαμηλής ενέργειας i καθορίζεται αποκλειστικά από την εγκάρσια ορμή του σωματιδίου με υψηλή ενέργεια και της απόστασης ΔR_{1i} . Το d_{ij} μεταξύ ομοίως διαχωρισμένων σωματιδίων χαμηλής ενέργειας θα είναι αντίθετα πολύ μεγαλύτερο. Επομένως αυτά τα σωματίδια θα τείνουν να ομαδοποιούνται με εκείνα της υψηλής ενέργειας πολύ πριν ομαδοποιηθούν μεταξύ τους. Εάν ένα σωματίδιο υψηλής ενέργειας δεν έχει ομοιά του σωματίδια 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 σωματίδιο μεγάλης ενέργειας 2, έτσι ώστε $R < \Delta R_{12} < 2R$ τότε θα υπάρχουν δύο πίδακες υψηλής ενέργειας. Δεν είναι δυνατόν και οι δύο να είναι απόλυτα κωνικοί. Αν $k_{t1} \gg k_{t2}$ 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 $k_{t1} = k_{t2}$, 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 περιχοπούν, με το όριο b μεταξύ τους να ορίζεται από τη σχέση:

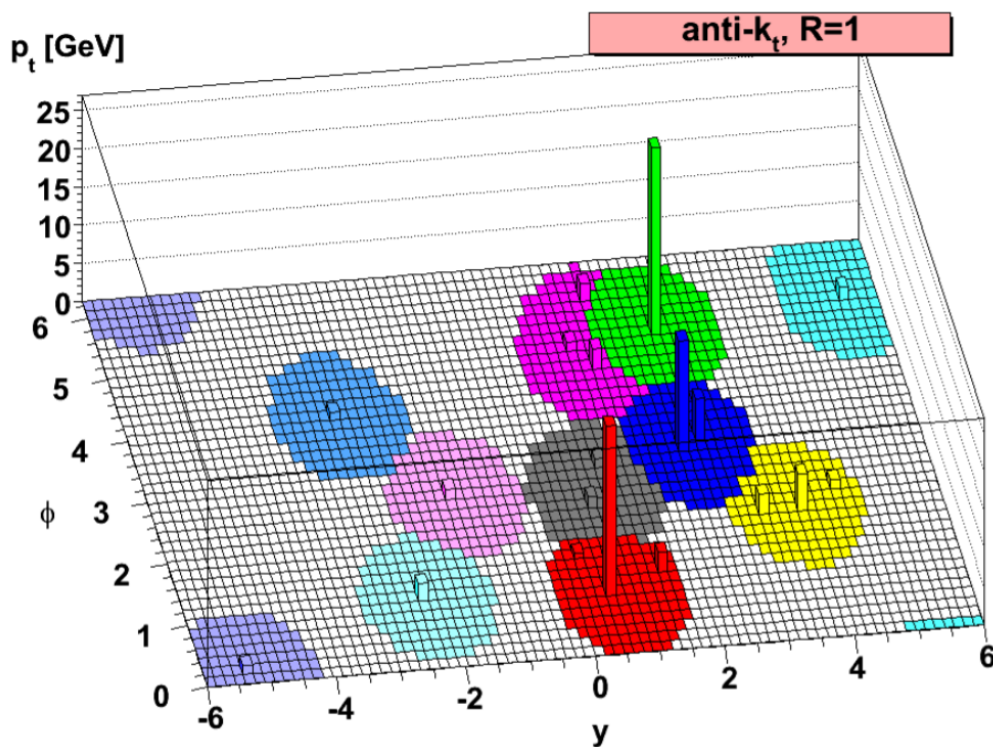
$$\frac{\Delta R_{1b}}{k_{t1}} = \frac{\Delta R_{2b}}{k_{t2}}$$

Όταν το $\Delta R_{12} < R$, τα σωματίδια 1 και 2 θα σχηματίσουν μαζί ένα μόνο πίδακα. Αν $k_{t1} \gg k_{t2}$ τότε θα είναι ένας κωνικός πίδακας με κέντρο στο k_1 . Για τη γενική περίπτωση, το σχήμα θα είναι πιο περίπλοκο, θα είναι η ένωση των κώνων (ακτίνα $< R$) γύρω από κάθε ενεργειακά υψηλά σωματίδιο, συν τον κώνο (ακτίνας R) με κέντρο τον τελικό πίδακα. Το

βασικό χαρακτηριστικό παραπάνω είναι ότι τα σωματίδια χαμηλής ενέργειας δεν καθορίζουν το σχήμα του πίδακα, ενώ αυτά με υψηλή το κάνουν.

Οι παραπάνω ιδιότητες του αλγορίθμου anti- k_T μεταφράζονται σε συγκεκριμένα αποτελέσματα για διάφορες ποσοτικές ιδιότητες των πιδάκων.

Στην εικόνα που ακολουθεί, φαίνεται ότι με τον αλγόριθμο anti- k_T , οι πίδακες υψηλής ενέργειας είναι όλοι κυκλικοί με ακτίνα R , και μόνο οι πίδακες χαμηλότερης ενέργειας έχουν πιο πολύπλοκα σχήματα.



ΣΧΗΜΑ 2.4: Απεικόνιση πιδάκων, κυρίως υψηλής ενέργειας, με χρήση του αλγορίθμου anti- k_T .

Κεφάλαιο 3

Ανάλυση Δεδομένων

3.1 Μηχανική μάθηση – Machine learning

Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούν υπολογιστικές μεθόδους για να «μαθαίνουν» τις πληροφορίες απευθείας από τα δεδομένα χωρίς να στηρίζονται σε μια προκαθορισμένη εξίσωση ως μοντέλο [12]. Έχουν την ικανότητα να βελτιώνουν προσαρμοστικά την απόδοσή τους καθώς αυξάνεται ο αριθμός των διαθέσιμων δειγμάτων για εκπαίδευση. Οι αλγόριθμοι μηχανικής μάθησης βρίσκουν φυσικά μοτίβα στα δεδομένα που οδηγούν στη λήψη καλύτερων αποφάσεων και στη διεξαγωγή ακριβέστερων προβλέψεων. Με την ολοένα αυξανόμενη ανάγκη για αντιμετώπιση προβλημάτων μεγάλων δεδομένων, η μηχανική μάθηση έχει καταστεί βασική τεχνική για την επίλυση τους.

Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούν δύο τύπους τεχνικών: την επιβλεπόμενη μάθηση, η οποία εκπαιδεύει ένα μοντέλο σε γνωστά δεδομένα εισόδου και εξόδου, έτσι ώστε να μπορεί να παράγει λογικές προβλέψεις στο μέλλον για την απάντηση σε νέα δεδομένα, και στη μη επιβλεπόμενη μάθηση, η οποία βρίσκει κρυμμένα μοτίβα ή εγγενείς δομές στα δεδομένα εισόδου.

Η επιβλεπόμενη μηχανική μάθηση χτίζει ένα μοντέλο που κάνει τις προβλέψεις βασιζόμενο σε γνωστά στοιχεία με την ύπαρξη κάποιας αβεβαιότητας. Ένας τέτοιος αλγόριθμος εκμάθησης, χρησιμοποιεί τεχνικές ταξινόμησης (classification) είτε παλινδρόμησης (regression) για την ανάπτυξη προγνωστικών μοντέλων.

Οι τεχνικές ταξινόμησης έχουν στόχο την εύρεση μίας βέλτιστης επιφάνειας διαχωρισμού σε ένα πολυδιάστατο χώρο χαρακτηριστικών [13]. Προβλέπουν διακριτές απαντήσεις, επομένως τα δεδομένα μας πρέπει να μπορούν να επισημανθούν, να κατηγοριοποιηθούν ή να διαχωριστούν σε συγκεκριμένες ομάδες ή κλάσεις. Μερικοί συνήθεις αλγόριθμοι για την εκτέλεση της ταξινόμησης περιλαμβάνουν τον support vector machine (SVM), τα ενδυναμωμένα δέντρα απόφασης, τον πλησιέστερο γείτονα, τον Naïve Bayes και τα νευρωνικά δίκτυα. Παραδείγματα εφαρμογών που χρησιμοποιούν ταξινόμηση περιλαμβάνουν την ιατρική απεικόνιση και την αναγνώριση ομιλίας.

Η παλινδρόμηση, κάνει προσαρμογή ενός μοντέλου σε ένα σύνολο μετρήσεων. Προβλέπει συνεχείς αποκρίσεις. Συνήθεις αλγόριθμοι παλινδρόμησης περιλαμβάνουν το γραμμικό μοντέλο, το μη γραμμικό μοντέλο, τη σταδιακή παλινδρόμηση, τα ενδυναμωμένα δέντρα αποφάσεων, τα νευρωνικά δίκτυα και την προσαρμοζόμενη νευρο-ασαφή μάθηση.

Η μη επιβλεπόμενη μάθηση όπως έχει προαναφερθεί, βρίσκει κρυμμένα μοτίβα στα δεδομένα. Χρησιμοποιείται για την εξαγωγή συμπερασμάτων από σύνολα δεδομένων εισόδου χωρίς προκαθορισμένες απαντήσεις.

Η εύρεση σημνών είναι η πιο κοινή τεχνική μάθησης χωρίς επίβλεψη. Χρησιμοποιείται για διερευνητική ανάλυση δεδομένων και κάνει ανακατασκευή δομών σε ένα πολυδιάστατο χώρο χαρακτηριστικών. Ένα παράδειγμα είναι η ανακατασκευή εικόνας. Στην επεξεργασία εικόνας, οι τεχνικές αναγνώρισης μοτίβων χωρίς επίβλεψη χρησιμοποιούνται για ανίχνευση αντικειμένων και τμηματοποίηση εικόνας.

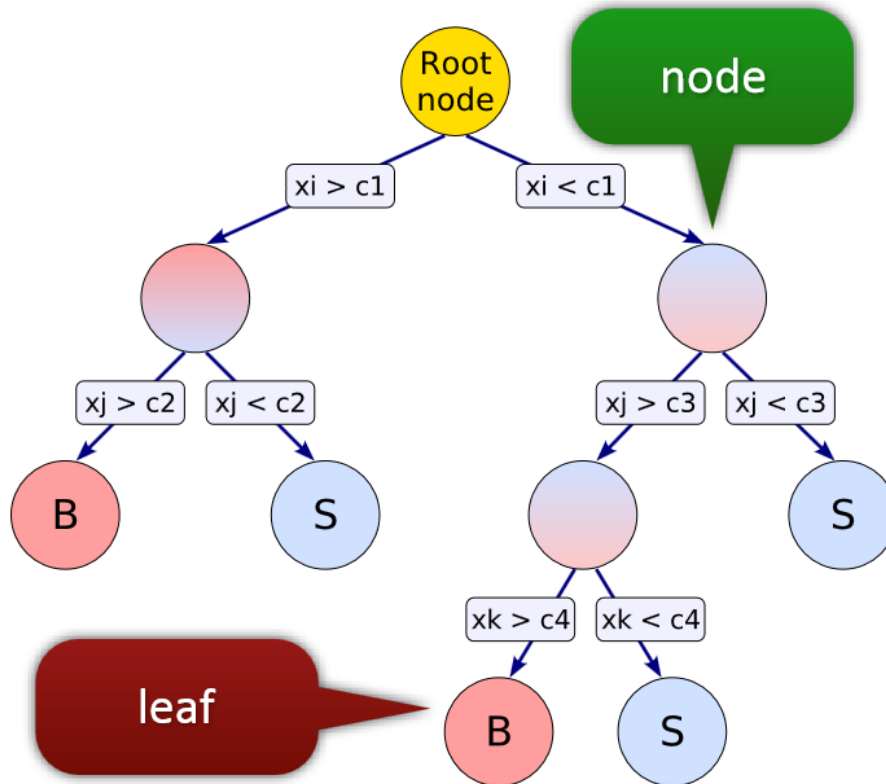
Στην ανάλυση δεδομένων που πραγματοποιείται σε αυτή την εργασία, χρησιμοποιείται η τεχνική της ταξινόμησης (classification), που ανήκει στην κατηγορία της επιβλεπόμενης μάθησης, και ειδικότερα χρησιμοποιείται ενδυναμωμένο δέντρο απόφασης (Boosted Decision Tree-BDT).

3.2 Δέντρο Απόφασης - Decision Tree

Πριν τη μέθοδο της ενδυνάμωσης θα αναφερθεί πως λειτουργεί ένα απλό δέντρο απόφασης. Στη φυσική υψηλών ενεργειών οι ερευνητές καλούνται συχνά να διαχωρίσουν το σήμα από το υπόβαθρο, επομένως μιας και αποτελεί το αντικείμενο αυτής της διπλωματικής, από αυτή τη σκοπιά θα περιγραφεί και το δέντρο απόφασης.

Για το διαχωρισμό του σήματος από το υπόβαθρο, με επιβλεπόμενη μάθηση, είναι απαραίτητο να υπάρχουν δεδομένα προσομοίωσης για το καθένα. Τα δεδομένα αυτά χωρίζονται σε δύο μέρη, ένα για να χρησιμοποιηθεί για εκπαίδευση (training) και ένα για τη μετέπειτα δοκιμή (testing). Για κάθε γεγονός, υπάρχει ένας αριθμός μεταβλητών ταυτοποίησης σωματιδίων (particle identification –PID), που είναι χρήσιμες για το διαχωρισμό μεταξύ σήματος και υποβάθρου [14]. Αρχικά, για κάθε PID μεταβλητή, κατατάσσονται τα γεγονότα σύμφωνα με την τιμή της μεταβλητής. Στη συνέχεια, επιλέγεται η πρώτη μεταβλητή και για κάθε μία τιμή γίνεται διαχωρισμός σε δύο μέρη. Επιλέγεται η τιμή διαίρεσης που δίνει τον καλύτερο διαχωρισμό, έχοντας στη μία πλευρά ως επί το πλείστον σήμα και στην άλλη κυρίως υπόβαθρο. Στη συνέχεια, επαναλαμβάνεται αυτό για κάθε μεταβλητή με τη σειρά. Ξεκινώντας από τη μεταβλητή και την τιμή διαίρεσης που δίνει τον καλύτερο διαχωρισμό, το αρχικό δείγμα γεγονότων σε έναν «κόμβο», χωρίζεται σε δύο μέρη που ονομάζονται «κλάδοι». Για κάθε κλάδο, επαναλαμβάνεται αυτή η διαδικασία, δημιουργώντας

έτσι νέους κόμβους και νέους κλάδους, έως ότου να υπάρχει ένας ορισμένος αριθμός τελικών κλάδων, τα λεγόμενα «φύλλα» ή μέχρι το κάθε «φύλλο» να είναι καθαρό σήμα ή καθαρό υπόβαθρο ή να έχει λίγα γεγονότα για να συνεχιστεί η διαδικασία.



ΣΧΗΜΑ 3.1: Σχηματική αναπαράσταση ενός δέντρου απόφασης που διαχωρίζει το σήμα (signal-S) από το υπόβαθρο (background-B).

Συνοψίζοντας λοιπόν, ένα δέντρο απόφασης είναι ένα πολυεπίπεδο σύστημα που αποτελείται από αλληλουχία δυαδικών αποφάσεων, που «τεμαχίζει» το χώρο των χαρακτηριστικών σε άνισα «κουτιά». Σε κάθε στάδιο επιλέγεται ο επόμενος κλάδος να χωριστεί, έτσι ώστε να δώσει την καλύτερη αύξηση της ποιότητας του διαχωρισμού. Το κριτήριο που χρησιμοποιείται για τον καθορισμό της ποιότητας του διαχωρισμού είναι:

$$\Delta I = I - (N_1/N) * I_1 - (N_2/N) * I_2 > 0$$

όπου I η ομοιογένεια του αρχικού διαχωρισμού και I_i των κόμβων που ακολουθούν, και N_i ο συνολικός αριθμός δεδομένων και των κόμβων i αντίστοιχα. Ως ομοιογένεια ορίζεται:

$$\text{Gini index: } I = p^*(1-p)$$

ή

$$\text{Cross entropy: } I = -p \cdot \ln p - (1-p) \cdot \ln(1-p)$$

και αποτελεί ένα μέτρο της καθαρότητας του κόμβου.

Στο τέλος, 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 να λυθεί αυτό επιλέγουμε την τεχνική ενδυνάμωσης που περιγράφεται στην επόμενη παράγραφο.

3.3 Ενδυναμωμένο Δέντρο Απόφασης - Boosted Decision Tree

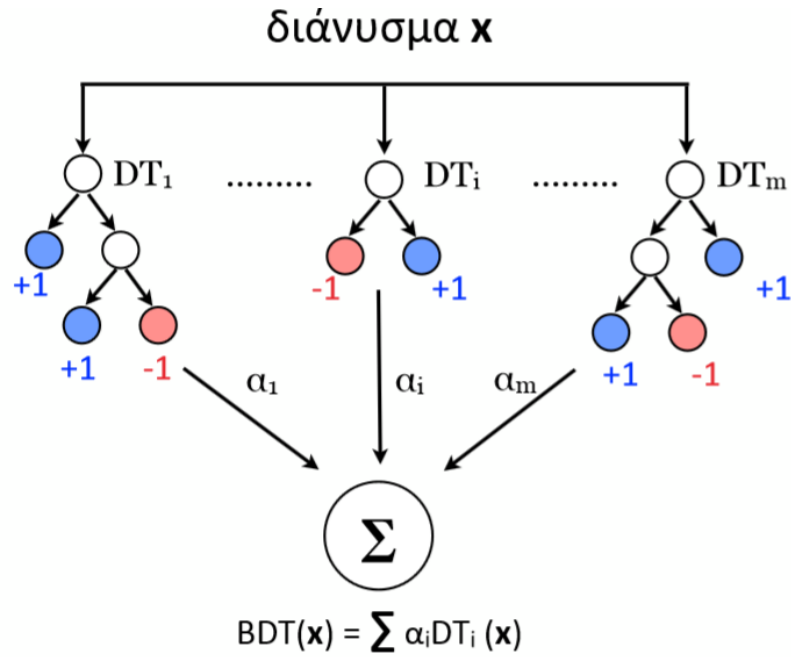
Με τον όρο ενδυνάμωση (boosting) εννοούμε αλληλουχία πολλών ασθενών ταξινομητών, εδώ απλών δέντρων απόφασης. Έτσι μπορούμε να απαλαχθούμε από τις στατιστικές διακυμάνσεις. Η λογική είναι ότι χρησιμοποιούμε τα ίδια δεδομένα εκπαίδευσης, πολλές φορές αλλά με διαφορετικό βάρος κάθε φορά, εκπαιδεύοντας ξεχωριστά δέντρα απόφασης μικρού βάθους (με λίγες διακλαδώσεις). Στο τέλος παίρνουμε το βεβαρυμένο μέσο όρο όλων των επιμέρους δέντρων απόφασης σύμφωνα με τη σχέση:

$$F(\vec{x}) = \sum_k a_k \phi(\vec{x}; \theta_k) \quad (3.1)$$

, όπου ϕ το αποτέλεσμα ενός απλού ταξινομητή, \vec{x} είναι ένα διάνυσμα εκπαίδευσης χαρακτηριστικών μεταβλητών, θ είναι ένα διάνυσμα παραμέτρων και k ο αριθμός των ταξινομητών που έχουμε [15]. Κάθε διάνυσμα εκπαίδευσης χαρακτηρίζεται από μία τιμή $y = +1$ ή $y = -1$. Σωστή θεωρείται η ταξινόμηση όταν αυτό που προβλέπει έχει το ίδιο πρόσημο με αυτό που είναι γνωστό ότι ισχύει (δηλαδή ϕ και y το ίδιο πρόσημο) σύμφωνα με τη συνάρτηση κόστους:

$$J(\alpha_k; \theta_k) = \sum_{i=0}^N e^{-y_i \phi(\vec{x}_i)} \quad (3.2)$$

Με αυτή τη συνάρτηση κόστους, τα λάθος ταξινομημένα διανύσματα ($y_i \varphi(x_i) < 0$) βαρύνονται πολύ περισσότερο (penalty) από αυτά με σωστή ταξινόμηση. 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 συνδιασμό όλων των επιμέρους δέντρων με τα κατάλληλα βάρη.



ΣΧΗΜΑ 3.2: Σχηματική απόδοση ενός Boosted Decision Tree.

Κεφάλαιο 4

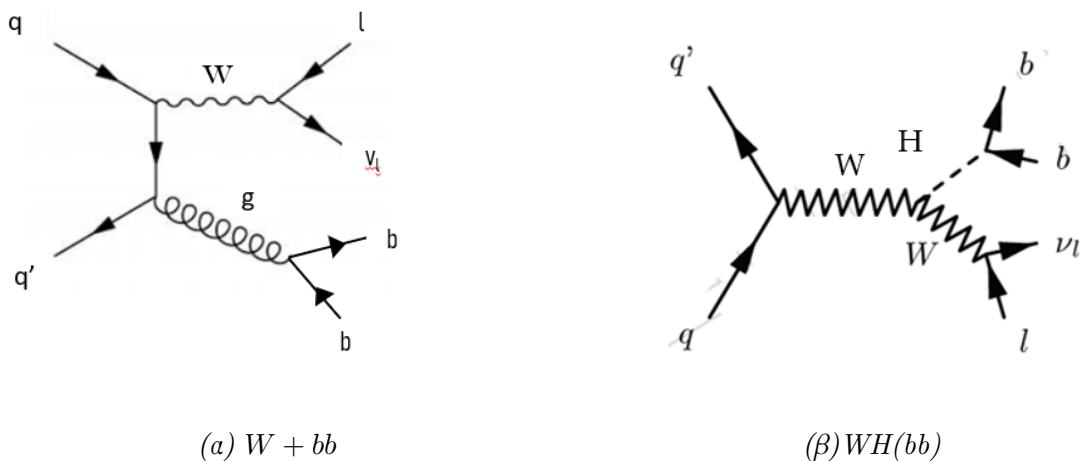
Ανάλυση

4.1 Εισαγωγή

Στην ανάλυση χρησιμοποιήθηκαν δεδομένα από συγκρούσεις πρωτονίων-πρωτονίων με ενέργεια κέντρου μάζας $\sqrt{s} = 13\text{TeV}$, που συλλέχθηκαν από το πείραμα CMS στον LHC, με φωτεινότητα (luminosity) $35,920\text{fb}^{-1}$.

Η μελέτη παραγωγής ενός W μποζονίου σε σχέση με δύο b κουάρκ (quarks) σχετίζεται με αρκετές ερευνητικές μελέτες. Καταρχάς, αποτελεί έλεγχο για τις προβλέψεις του Καθιερωμένου Προτύπου (Standard Model) [9]. Συγκεκριμένα, τέτοιες διεργασίες αποτελούν υπόβαθρο στη μελέτη διάφορων άλλων διεργασιών, όπως το $t\bar{t}$ ή το Single-top. Επιπρόσθετα, η $W + b\bar{b}$ διεργασία αποτελεί υπόβαθρο της παραγωγής του μποζονίου Higgs, όπου το Higgs μποζόνιο διασπάται σε ένα ζεύγος $b\bar{b}$. Αυτό δίνει επιπλέον κίνητρο για τη μελέτη, ώστε να συμβάλει στην κατανόηση της φύσης του μποζονίου και να καθορίσει τη σύζευξή του με τα b quarks. Επιπρόσθετα, παρέχει σημαντική πληροφορία για τη βελτίωση των υπολογισμών στην διαταραχτική κβαντική χρωμοδυναμική.

Το διάγραμμα πρώτης τάξης (leading order-LO) που περιγράφει την παραγωγή $W + b\bar{b}$ παρατίθεται στο Σχήμα 4.1 (α), όπου το ζεύγος $b\bar{b}$ παράγεται από ακτινοβολία γλουονίων (gluon radiation) και στο Σχήμα 4.1 (β) η παραγωγή του Higgs μποζονίου $WH(b\bar{b})$.



ΣΧΗΜΑ 4.1: Διαγράμματα feynman πρώτης τάξης

4.2 Δείγματα Προσομοίωσης - Monte Carlo samples

Στη φυσική υψηλών ενεργειών χρησιμοποιούνται μέθοδοι προσομοίωσης Monte Carlo. Η γενική ιδέα των μεθόδων αυτών, είναι αντί να πραγματοποιούνται τόσο πολύπλοκοι υπολογισμοί αναλυτικά, υλοποιείται ένας μεγάλος αριθμός πειραμάτων χρησιμοποιώντας ένα γεννητόρα τυχαίων αριθμών. Η χρήση Monte Carlo υποθέτει ότι το σύστημα περιγράφεται από πυκνότητες πιθανότητας PDFs, οι οποίες μοντελοποιούνται [16].

Αρχικά γίνεται επιλογή θεωρητικού μοντέλου, κινηματική ανάλυση γνωστών σωματιδίων και προσομοίωση ανιχνευτή. Στη σωματιδιακή φυσική χρησιμοποιούνται Monte Carlo μέθοδοι για να σχεδιάσουν και να βελτιστοποιήσουν την ακριβή περιγραφή του ανιχνευτή. Επιπλέον, χρησιμοποιούνται για να προσομοιώσουν τις αλληλεπιδράσεις των σωματιδίων με το υλικό του ανιχνευτή.

Οι πιο βασικοί αλγόριθμοι που χρησιμοποιούνται για την παραγωγή Monte Carlo στις υψηλές ενέργειες είναι οι Herwig, Pythia, Sherpa και POWHEG. Όλα τα δείγματα που χρησιμοποιήθηκαν στην ανάλυση παρατίθενται στον παρακάτω πίνακα μαζί με την ενεργό διατομή του κάθε δείγματος αντίστοιχα.

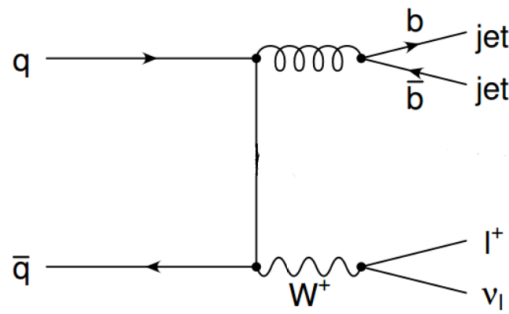
| Δείγματα Προσομοίωσης Monte Carlo | Ενεργός Διατομή (pb) |
|---|----------------------|
| WJetsToLNu_HT-70To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1353 |
| WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1346 |
| WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 359,7 |
| WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 48,91 |
| WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 12,05 |
| WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 5,501 |
| WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1,329 |
| WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0,03216 |
| WToLNu_0J_13TeV-amcatnloFXFX-pythia8 | 49670 |
| WToLNu_1J_13TeV-amcatnloFXFX-pythia8 | 8264 |
| WToLNu_2J_13TeV-amcatnloFXFX-pythia8 | 2628 |

| Δείγματα Προσομοίωσης Monte Carlo (συνέχεια) | Ενεργός Διατομή (pb) |
|---|----------------------|
| DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 169,9 |
| DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 147,4 |
| DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 40,99 |
| DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 5,678 |
| DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 1,367 |
| DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0,6304 |
| DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0,1514 |
| DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 | 0,003565 |
| TT_TuneCUETP8M2T4_13TeV-powheg-pythia8 | 832 |
| ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 | 136,02 |
| ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1 | 80,95 |
| ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | 35,6 |
| ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1 | 35,6 |
| ST_s-channel_4f_InclusiveDecays_13TeV-amcatnlo-pythia8 | 10,32 |
| WW_TuneCUETP8M1_13TeV-pythia8 | 115,00 |
| WZ_TuneCUETP8M1_13TeV-pythia8 | 47,13 |
| ZZ_TuneCUETP8M1_13TeV-pythia8 | 16,523 |

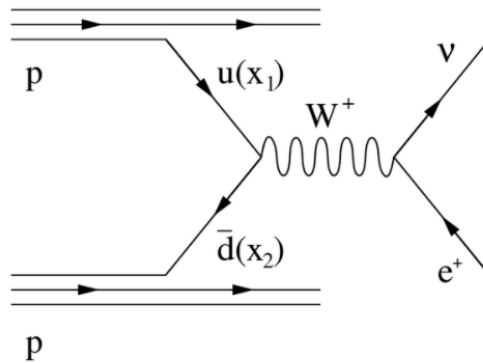
ΠΙΝΑΚΑΣ 4.1: Πίνακας δειγμάτων προσομοίωσης Monte Carlo και των αντίστοιχων ενεργών διατομών

4.3 $W + b\bar{b}$

Το σήμα αυτής της μελέτης είναι η $W + b\bar{b}$ διεργασία, όπου ένα W μποζόνιο παράγεται μαζί με δύο πίδακες b που προκύπτουν από gluon splitting. Στην τελική κατάσταση υπάρχουν δύο b πίδακες με ένα λεπτόνιο και ένα νεutrino.

ΣΧΗΜΑ 4.2: Παραγωγή $W + b\bar{b}$

Γενικά το μποζόνιο W μπορεί να παραχθεί λεπτονικά ή μέσω κουάρκ. Στις συγκρούσεις πρωτονίων το W μποζόνιο παράγεται από την αλληλεπίραση δύο κουάρκ. Ένα κουάρκ από το ένα πρωτόνιο και ένα αντικουάρκ από το άλλο αλληλεπιδρούν και παράγουν το μποζόνιο. Κυρίως αυτό συμβαίνει μέσω ενός u και ενός \bar{d} και παράγεται ένα W^+ (ή $\bar{u}d$ και παράγεται ένα W^-) [17]. Για αυτό στις συγκρούσεις πρωτονίων πρωτονίων υπερισχύει η παραγωγή W^+ . Όμως το φορτίο του μποζονίου δεν συγκεντρώνει το ενδιαφέρον της ανάλυσης αυτής.

ΣΧΗΜΑ 4.3: Παραγωγή W μποζονίου από σύγκρουση πρωτονίων.

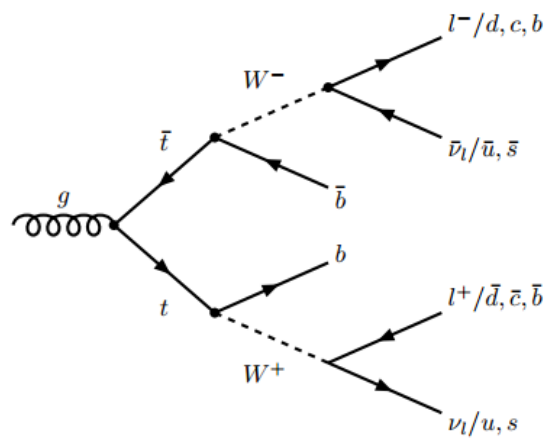
Το W μποζόνιο έχει μικρό χρόνο ημιζωής ($3 \times 10^{-25} \text{ sec}$) και δεν παρατηρείται άμεσα στον ανιχνευτή. Αντί αυτού αναζητούνται τα προϊόντα στα οποία διασπάστηκε. Το W μποζόνιο διασπάται αδρονικά ή λεπτονικά. Τα αδρόνια είναι δύσκολο να ταυτοποιηθούν αν προήλθαν από διασπάσεις W ή από άλλες διεργασίες που πραγματοποιούνται κατά τη διάρκεια του πειράματος. Για αυτό το ενδιαφέρον συγκεντρώνεται στη λεπτονική διάσπασή του. Λεπτονικά διασπάται σε e , μ ή τ και το αντίστοιχο νεutrino ν_e, ν_μ ή ν_τ . Το νεutrino δεν ανιχνεύεται, υπολογίζεται όμως η ελλειπούσα εγκάρσια ορμή E_T^{miss} . Στην τελευταία περίπτωση το τ διασπάται γρήγορα αδρονικά ή λεπτονικά σε e ή μ . Αυτά τα λεπτόνα θα είναι χαμηλότερης ενέργειας αφού προκύπτουν από δευτερεύουσα διάσπαση και ένας τρόπος να εξαληφθεί αυτή η συνεισφορά στο σήμα είναι θέτοντας ένα κάτω περιορισμό στην ορμή του λεπτονίου [7].

Στα πλαίσια αυτής της ανάλυσης υλοποιήθηκαν τα βήματα και για τα δύο κανάλια, εστιάζοντας το ενδιαφέρον πρωτίστως στο κανάλι των ηλεκτρονίων. Για αυτό και παρακάτω παρατίθενται αναλυτικά οι μέθοδοι για τα ηλεκτρόνια ενώ για το μιονικό κανάλι παρουσιάζονται κυρίως τα τελικά διαγράμματα.

4.4 Διεργασίες υποβάθρου

Οι κυριότερες διεργασίες που αποτελούν υπόβαθρο στην μελέτη της $W + b\bar{b}$ είναι:

- $t\bar{t}$

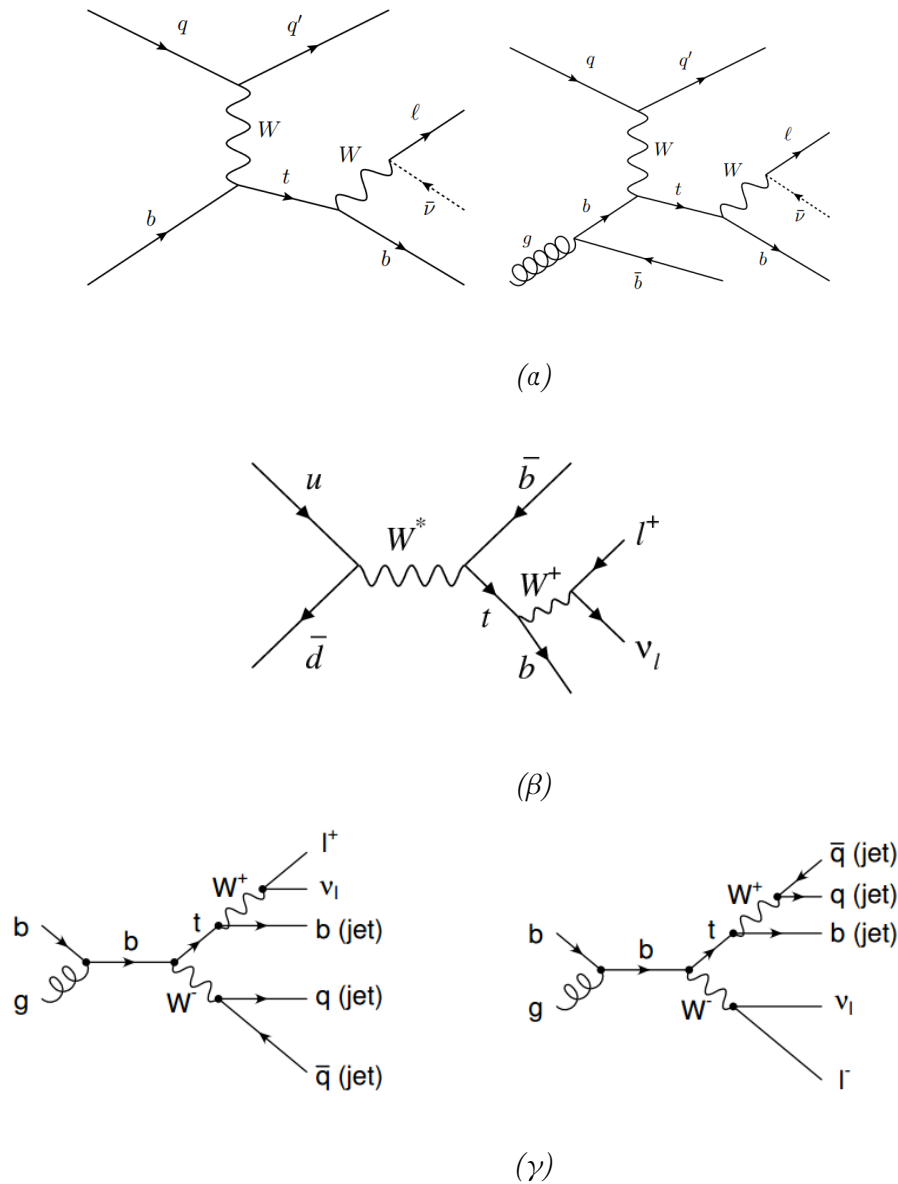


ΣΧΗΜΑ 4.4: Παραγωγή $t\bar{t}$

Το κυρίαρχο υπόβαθρο προέρχεται από διεργασίες $t\bar{t}$, όπου ένα ζεύγος top-antitop κουάρκ παράγεται μέσω της ισχυρής αλληλεπίδρασης. Το top κουάρκ διασπάται κυρίως σε W μποζόνιο και b κουάρκ ($t \rightarrow bW^+$). Στην τελική κατάσταση έχει δύο b πίδακες, όπως και το σήμα μας. Αυτό περνάει στο σήμα εάν ένα τουλάχιστον από τα W μποζόνια, που παράγονται από διάσπαση του top, διασπαστεί σε λεπτόνιο και νεutrino. Είτε αν διασπαστούν και τα δύο και το ένα λεπτόνιο δεν ανιχνευτεί.

- **single-top**

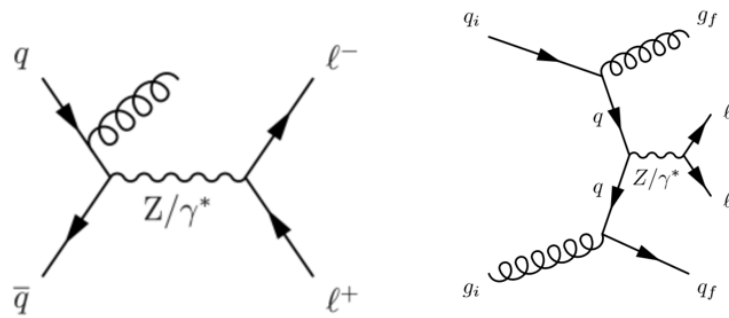
Single-top διεργασίες παράγονται μέσω t-channel, s-channel και Wt [18]. Η κυρίαρχη διεργασία είναι το t-channel όπου ένα top κουάρκ παράγεται με ένα ελαφρύ κουάρκ ή με ένα ελαφρύ και ένα b-jet. Ακολουθεί το Wt όπου ένα top κουάρκ παράγεται σε συνδιασμό με ένα W μποζόνιο. Τελευταίο έρχεται το s-κανάλι στο οποίο το t κουάρκ παράγεται με ένα b κουάρκ. Έχουν συνεισφορά ως υπόβαθρο λόγω του W που παράγεται είτε από διάσπαση του top είτε άμεσα.



ΣΧΗΜΑ 4.5: Παραγωγή single top διεργασιών: α) t-channel , β) s-channel και γ) Wt

- **Drell-Yan $Z \rightarrow ll$**

Η παραγωγή ενός μποζονίου Z σε συνδυασμό με πίδακες σε συγκρούσεις πρωτονίων-πρωτονίων κυριαρχείται από ένα μίγμα ηλεκτρασθενών και ισχυρών διεργασιών με ενεργό διατομή της τάξης α_{EW}^2 επί ένα παράγοντα α_S για κάθε επιπλέον πίδακα που παράγεται [19]. Λεπτονικές διασπάσεις, μέσω μποζονίου Z αναφέρονται ως Drell-Yan. Όπου ένα κουάρκ και ένα αντικουάρκ εξαυλώνονται και μέσω ενός Z μποζονίου παράγουν ένα λεπτόνιο και ένα αντιλεπτόνιο. Αυτά μπορεί να είναι ηλεκτρόνια, μίονια ή ταυ.



ΣΧΗΜΑ 4.6: Παραγωγή Drell-Yan

Διεργασίες Drell-Yan ικανοποιούν τις απαιτήσεις του φασικού χώρου του σήματος, εάν ένα από τα λεπτόνια διαφύγει από το εύρος του ανιχνευτή ή δεν περάσει τους περιορισμούς για την ταυτοποίηση ή το isolation. Και στις δύο περιπτώσεις έχουμε λανθάνουσα ελλειπούσα εγκάρσια ενέργεια E_T^{miss} και ένα λεπτόνιο στην τελική κατάσταση όπως και για $W + b\bar{b}$ διεργασίες.

- **Dibosons**

Συνεισφέρουν στο σήμα κυρίως μέσω WW διεργασιών, όπου το W διασπάται λεπτονικά. Οι WZ , ZZ διεργασίες έχουν μικρότερη συνεισφορά και ειδικά η τελευταία που δεν διαθέτει W μποζόνιο.

Το σύνολο των υποβάθρων που αναφέρθηκαν μέχρι τώρα εκτιμώνται από τα δείγματα προσομοίωσης.

- **QCD-multijet**

QCD-multijet γεγονότα μπορούν να περάσουν τα κριτήρια του σήματος όταν ένας πίδακας ταυτοποιείται λανθασμένα ως λεπτόνιο, και αυτή η λάθος μέτρηση στην ενέργεια του πίδακα οδηγεί σε ψεύτικη αύξηση της ελλειπούσας εγκάρσιας ορμής. Αν και η πιθανότητα αυτή είναι μικρή, η μεγάλη ενεργός διατομή παραγωγής multijet οδηγεί σε μία μη αμεληταία συνεισφορά αυτού του υποβάθρου. Οι QCD-multijet διεργασίες εκτιμώνται από τα δεδομένα.

4.5 Μεταβλητές

Για τη διεξαγωγή της ανάλυσης χρησιμοποιήθηκε ένας μεγάλος αριθμός μεταβλητών. Οι κυριότερες αναφέρονται:

- $lepId$: η μεταβλητή που διευκρινίζει την ταυτότητα του λεπτονίου. Η τιμή 13 και -13 αντιστοιχεί σε μιονίο και αντιμιονίο αντίστοιχα και η τιμή 11 και -11 σε ηλεκτρόνιο και ποζιτρόνιο αντίστοιχα.
- $nJets$: ο αριθμός των πιδάκων (jets) σε κάθε γεγονός (event)

- nFwdJets: ο αριθμός των πιδάκων με ψευδοωκυτότητα $|\eta| > 2.4$ σε κάθε γεγονός
- nLeptons: ο αριθμός των απομονωμένων (isolated) λεπτονίων σε κάθε γεγονός
- nNonIsoLeptons: ο αριθμός των μη απομονωμένων λεπτονίων σε κάθε γεγονός
- nVetoLeptons: ο αριθμός των λεπτονίων που απορρίφθηκαν σε κάθε γεγονός
- nBJets: ο αριθμός των πιδάκων που ταυτοποιήθηκαν ως πίδακες από b κουάρκ
- lepη: η τιμή της ψευδοωκυτότητας του απομονωμένου λεπτονίου.
- jetη: η τιμή της ψευδοωκυτότητας του πίδακα.
- jetPt₀: η εγκάρσια ορμή του πίδακα με την μεγαλύτερη ορμή (leading jet)
- jetPt₁: η εγκάρσια ορμή του πίδακα με την δεύτερη μεγαλύτερη ορμή (subleading jet)
- lepPt₀: η εγκάρσια ορμή του leading lepton.
- met: η ελλειπούσα εγκάρσια ενέργεια. Ορίζεται ως το αρνητικό διανυσματικό άθροισμα των εγκάρσιων ορμών όλων των ανακατασκευασμένων σωματιδίων στο γεγονός.

$$\vec{E}_T^{miss} = - \sum_i \vec{P}_{Ti} \quad (4.1)$$

- lmetDPhi: η αζιμουθιακή γωνία του απομονωμένου λεπτονίου και της ελλειπούσας εγκάρσιας ορμής.
- lMetMT: ή M_{TW} είναι εγκάρσια μάζα που ανακατασκευάστηκε από το απομονωμένο λεπτόνιο και την ελλειπούσα εγκάρσια ορμή, δηλαδή η εγκάρσια μάζα του W. Αυτή ορίζεται ως:

$$m_T^2 = (E_T^l + E_T^{miss})^2 - (\vec{P}_T^l + \vec{P}_T^{miss})^2 = m_l^2 + 0 + 2(E_T^l E_T^{miss} - \vec{P}_T^l \cdot \vec{P}_T^{miss}) \quad (4.2)$$

όπου θεωρώντας το λεπτόνιο άμαζο $m_l = 0$ γίνεται:

$$m_T^2 = 2E_T^l E_T^{miss} (1 - \cos\Delta\phi) \quad (4.3)$$

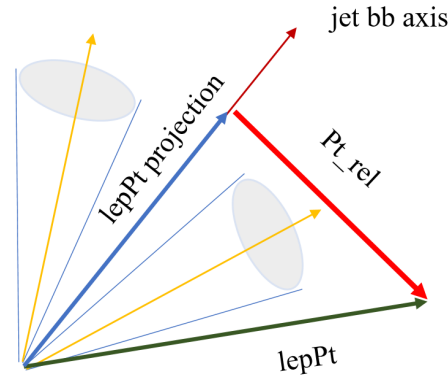
όπου $\Delta\phi$ είναι η αζιμουθιακή γωνία μεταξύ λεπτονίου και ελλειπούσας ενέργειας στο εγκάρσιο επίπεδο. Η m_T είναι μεταβλητή με μεγάλη διαχωριστική ικανότητα όσο αφορά σε τελικές καταστάσεις χωρίς W, όπως QCD-multijet γεγονότα που έχουν πολύ χαμηλή m_T .

- dRlBMin: η απόσταση του απομονωμένου λεπτονίου και του κοντινότερου b πίδακα σε αυτό.

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \quad (4.4)$$

όπου $\Delta\eta$ είναι η διαφορά της ψευδοωκυτότητας (η) του απομονωμένου λεπτονίου και της ψευδοωκυτότητας του πίδακα και $\Delta\phi$ η διαφορά των αζιμουθιακών γωνιών αντίστοιχα.

- mBB: η μάζα που ανακατασκευάζεται από τα δύο b-jet του γεγονότος.
- dPhiBB: η αζιμουθιακή γωνία ανάμεσα στους δύο b-πίδακες του γεγονότος.
- lepP_proj_jetP: η προβολή της ορμής του λεπτονίου στον άξονα των δύο b-jets.
- Pt_rel: η κάθετη απόσταση της εγκάρσιας ορμής του λεπτονίου από τον κοινό άξονα των δύο b-jets.



ΣΧΗΜΑ 4.7: Σχηματική αναπαράσταση της Pt_rel

- topmass_2: η ανακατασκευασμένη εγκάρσια μάζα του t (top) κουάρκ από το W μποζόνιο και το κοντινότερο b-jet, στην ουσία δηλαδή ανακατασκευασμένη από το λεπτόνιο, την ελλειπούσα εγκάρσια ορμή και το κοντινότερο b-jet

$$M_T^{top2} = (E_T^W + E_T^{jet})^2 - (\vec{P}_T^W + \vec{P}_T^{jet})^2 =$$

$$= E_T^{W2} + E_T^{jet2} + 2E_T^W E_T^{jet} - P_T^{W2} - P_T^{jet2} - 2\vec{P}_T^W \vec{P}_T^{jet}$$

και τελικά:

$$M_T^{top2} = M_T^{W2} + M_T^{jet2} + 2E_T^W E_T^{jet}$$

$$- 2 P_T^{lep} P_T^{jet} \cos(\Phi_{lep} - \Phi_{jet}) - 2E_T^{miss} P_T^{jet} \cos(\Phi_{E^{miss}} - \Phi_{jet}) \quad (4.5)$$

όπου αντί της M_T^W στον πρώτο προσθετικό όρο χρησιμοποιήθηκε η θεωρητική τιμή της μάζας του W μποζονίου ίση με 80,4 GeV.

4.6 Επιλογή περιορισμών που ορίζουν το φασικό χώρο-Selection cuts

Για να είναι καλά ορισμένος ο φασικός χώρος του σήματος $W + b\bar{b}$ απαιτούνται περιορισμοί σε ορισμένο αριθμό μεταβλητών, ώστε να επιλεγούν τα γεγονότα που αφορούν στην

διεργασία. Οι περιορισμοί αυτοί είναι τέτοιοι ώστε να μειώσουν όσο το δυνατόν περισσότερο τη συνεισφορά των επιμέρους υποβάθρων.

Από τις πρώτες απαιτήσεις είναι τα γεγονότα να αποτελούνται από ακριβώς ένα ανακατασκευασμένο λεπτόνιο με εγκάρσια ορμή $p_T > 35\text{GeV}$ και ψευδοωκότητα $|\eta| < 2,1$. Ο περιορισμός της εγκάρσιας ορμής καθορίζεται από τη διαλογή του trigger με κατώφλι ενέργειας 27GeV για το ηλεκτρονίο και 24GeV για το μόνιο. Η ψευδοωκότητα καθορίζεται από τη περιοχή του ανιχνευτή που λειτουργεί ο trigger μιονίων. Η επιλογή ακριβώς ενός λεπτονίου, συμβάλλει στην μείωση της συνεισφοράς στο υπόβαθρο από τις διεργασίες που έχουν παραπάνω από ένα λεπτόνια στην τελική κατάσταση. Τέτοιες είναι οι Drell-Yan, diboson, $t\bar{t}$. Επιπρόσθετα απαιτείται το λεπτόνιο να είναι καλά απομονωμένο. Η απαίτηση αυτή οδηγεί στη μείωση της συνεισφοράς από QCD-multijet διεργασίες.

Οι πίδακες που έχουν ανακατασκευαστεί από τον anti- k_T αλγόριθμο απαιτείται να έχουν ακτίνα $R=0,5$ και τιμές ψευδοωκότητας $|\eta| < 2,4$. Ο περιορισμός για την ψευδοωκότητα συμβάλλει στη μείωση της συνεισφοράς στο υπόβαθρο της διεργασίας single-top, όπου το s-channel αναμένεται να έχει πίδακα στο μπροστινό (forward) τμήμα του ανιχνευτή. Οι πίδακες που βρίσκονται σε μικρότερη ακτίνα από $0,5$ με κάποιο απομονωμένο λεπτόνιο αφαιρούνται από τα γεγονότα, ώστε να αποφευχθεί το ενδεχόμενο τέτοια λεπτόνια να επισημανθούν λανθασμένα ως πίδακες. Τα γεγονότα απαιτείται να έχουν ακριβώς δύο πίδακες. Ο λόγος είναι η μείωση της διεργασίας παραγωγής $t\bar{t}$, η οποία αποτελεί και το κύριο υπόβαθρο και κυριαρχεί ακόμα περισσότερο όταν τα γεγονότα έχουν μεγαλύτερη πολλαπλότητα σε πίδακες από δύο. Σε συνδυασμό με τον αριθμό των πιδάκων να είναι ίσοι με δύο, υπάρχει επιπλέον απαίτηση να είναι και οι δύο b-tagged, όπως καθορίζεται από την τελική κατάσταση της διεργασίας $W + b\bar{b}$ (σήματος). Έτσι σχεδόν εξαλείφεται η συνεισφορά από παραγωγή μποζονίων (W ή Z) σε συνδυασμό με άλλους πίδακες (u, d, s, c), εκτός από πίδακες που γίνονται misstaggering ως πίδακες b. Το b-tagging αφορά μόνο στο κεντρικό τμήμα του ανιχνευτή, δηλαδή σε περιοχή της ψευδοωκότητας $|\eta| < 2,4$. Τέλος, για τους δύο πίδακες τον leading και τον subleading απαιτείται ορμή μεγαλύτερη από 30GeV .

Συνοπτικά όλες οι προϋποθέσεις για τον καθορισμό του φασικού χώρου είναι:

- nJets = 2
- nBJets = 2
- nFwdJets = 0
- nLeptons = 1
- nNonIsoLeptons = 0
- nVetoLeptons = 0

- $lepPt > 35 \text{ GeV}$
- $jetPt_0 > 30 \text{ GeV}$
- $jetPt_1 > 30 \text{ GeV}$

4.7 Boosted Decision Tree

Στα πλαίσια αυτής της ανάλυσης διερευνήθηκε ένας εκλέπτυσμένος τρόπος ανάδειξης του σήματος μέσω πολυπαραγοντικών μεθόδων διαχωρισμού (MVA). Μέσω του εργαλείου του ROOT TMVA υλοποιήθηκε ένα ενδυναμωμένο δέντρο απόφασης (Boosted Decision Tree-BDT). Η εκπαίδευση (training) έγινε χρησιμοποιώντας για την κλάση του σήματος τα Monte Carlo δείγματα με $W+Jets$ διεργασίες σε LO τάξη και για την κλάση του υποβάθρου τα δείγματα με διεργασίες $t\bar{t}$, single top, Drell-Yan και Diboson. Τα οχτώ δείγματα του σήματος και τα δεκαέξι δείγματα του υποβάθρου προστέθηκαν στις αντίστοιχες κλάσεις με κατάλληλο βάρος το καθένα που ανταποκρίνεται στη φύση, δηλαδή στην ενεργό διατομή. Το βάρος αυτό ορίζεται ως ο λόγος της ενεργού διατομής προς το συνολικό αριθμό γεγονότων στο κάθε δείγμα αντίστοιχα.

Υπάρχουν διαφορετικοί αλγόριθμοι ενδυνάμωσης (boosting) ενός BDT. Οι κυριότεροι που υποστηρίζονται από το πακέτο ROOT TMVA, που χρησιμοποιήθηκε για την υλοποίηση αυτής της ανάλυσης, είναι οι: AdaBoost, Gradient Boost και Bagging. Σε αυτό το BDT χρησιμοποιήθηκε ο Gradient Boost.

Gradient Boost

Η διαδικασία της ενδυνάμωσης προσπαθεί να ελαχιστοποιήσει την απόκλιση μεταξύ του μοντέλου και της πραγματικής τιμής που είναι γνωστή για το σύνολο εκπαίδευσης. Η απόκλιση καθορίζεται από τη συνάρτηση απώλειας (loss function). Ο GradientBoost αλγόριθμος χρησιμοποιεί ως συνάρτηση απώλειας τη διωνυμική log-likelihood:

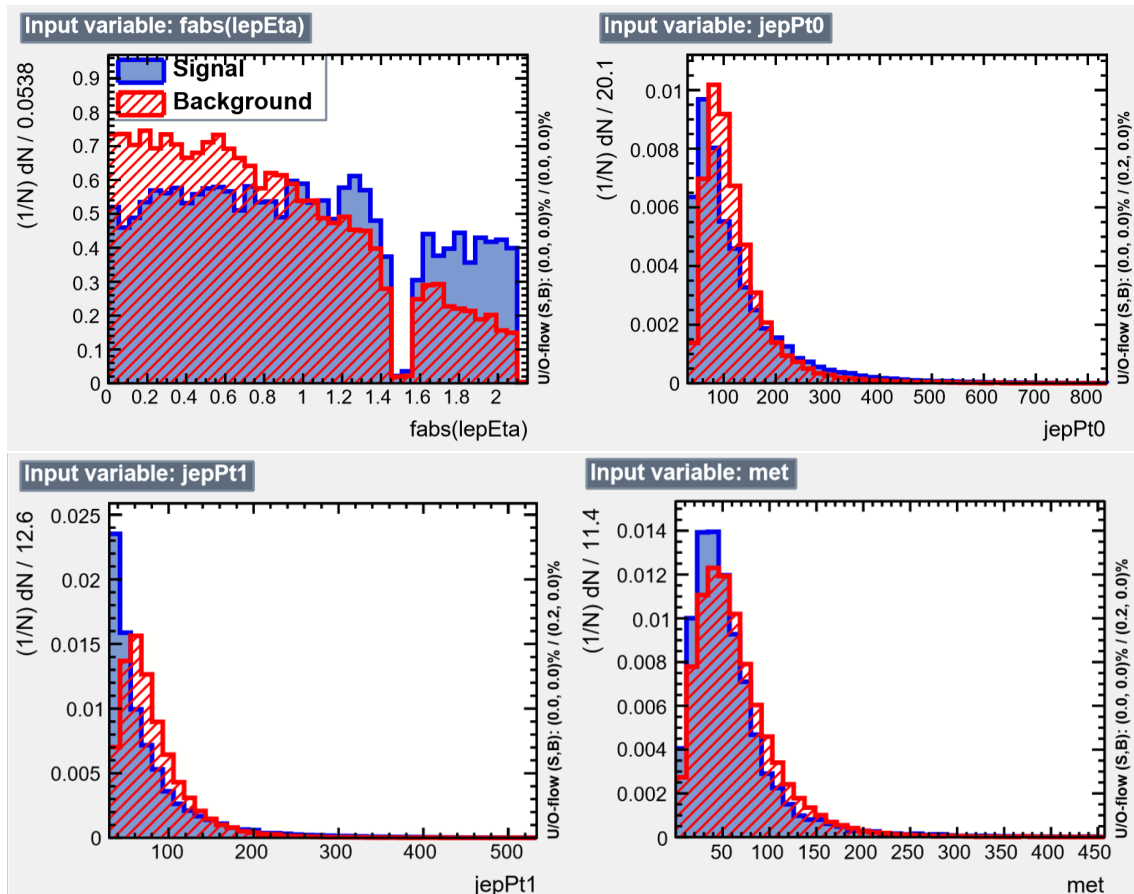
$$L(F, y) = \ln(1 + e^{-2F(x)y}), \quad (4.6)$$

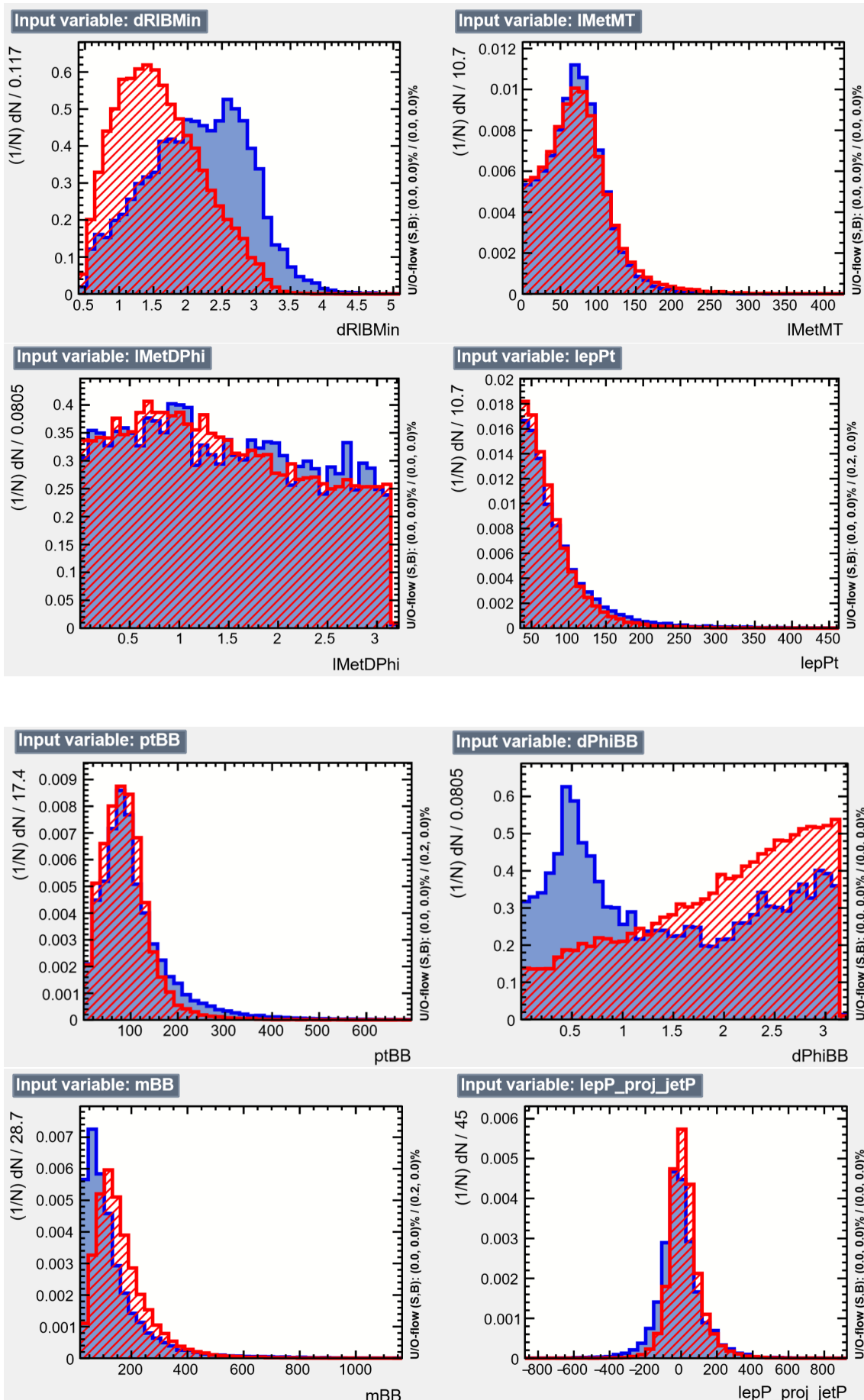
όπου y είναι η πραγματική τιμή και $F(x)$ είναι το αποτέλεσμα του BDT. Για την ελαχιστοποίηση αυτής της συνάρτησης απώλειας χρησιμοποιείται η μέθοδος της απότομης καθόδου. Η επιλογή αυτής της συνάρτησης απώλειας κάνει τον GradientBoost αλγόριθμο πιο σταθερό από τον AdaBoost αλγόριθμο, ως προς τις ακραίες τιμές ή λαθος μοντελοποιημένα γεγονότα.

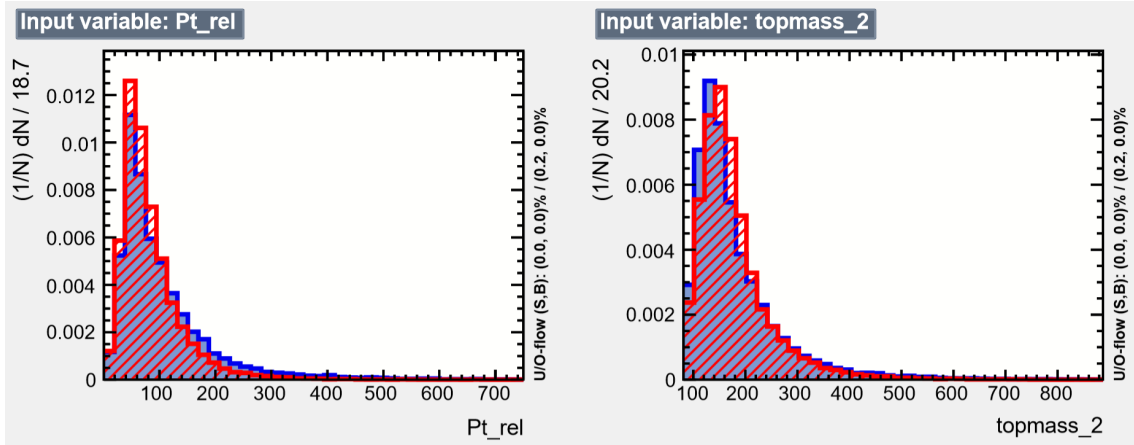
Τα επιμέρους δέντρα απόφασης οφείλουν να είναι, όπως έχει ήδη αναφερθεί ασθενείς ταξινομητές. Επομένως, έχουν διαμορφωθεί δέντρα μικρού βάθους, με μέγιστο αριθμό διακλαδώσεων ίσο με τρία. Αυτή η επιλογή κάνει το BDT λιγότερο ευπαθή ως προς την υπερεκπαίδευση. Για να ενισχυθεί αυτό και η σταθερότητα του BDT επιλέγεται η παράμετρος Shrinkage μικρή, δηλαδή στην παρούσα ανάλυση ίση με 0,1. Η παράμετρος Shrinkage είναι ο ρυθμός εκμάθησης και ελέγχει το βάρος των επιμέρους δέντρων απόφασης. Όπως είναι αναμενόμενο, χαμηλός ρυθμός εκμάθησης απαιτεί μεγαλύτερο αριθμό δέντρων απόφασης. Ο αριθμός των δέντρων αυξήθηκε μέχρι να μην επιφέρει σημαντική επιπλέον βελτίωση στον ταξινομητή. Αυτό συνέβει για 100 δέντρα. Τέλος, μία σημαντική παράμετρος που αξίζει να αναφερθεί είναι ο αριθμός των cuts. Για κάθε μεταβλητή ο διαχωρισμός γίνεται έτσι ώστε να επιτευχθεί μεγαλύτερη καθαρότητα ανάμεσα στον αρχικό κόμβο και στο άθροισμα των επιμέρους κόμβων (στους οποίους διαχωρίζεται) βεβαρυμένο με το σχετικό κλάσμα των γεγονότων. Για το διαχωρισμό αυτό ελέγχεται ένας επιλεγμένος αριθμός από cuts. Στο BDT που υλοποιήθηκε επιλέχθηκε αριθμός cuts ίσος με 200.

4.7.1 Μεταβλητές

Οι κατανομές των μεταβλητών που χρησιμοποιήθηκαν στο training και testing του BDT για τις δύο κλάσεις (σήμα και υπόβαθρο) απεικονίζονται στα σχήματα που ακολουθούν.

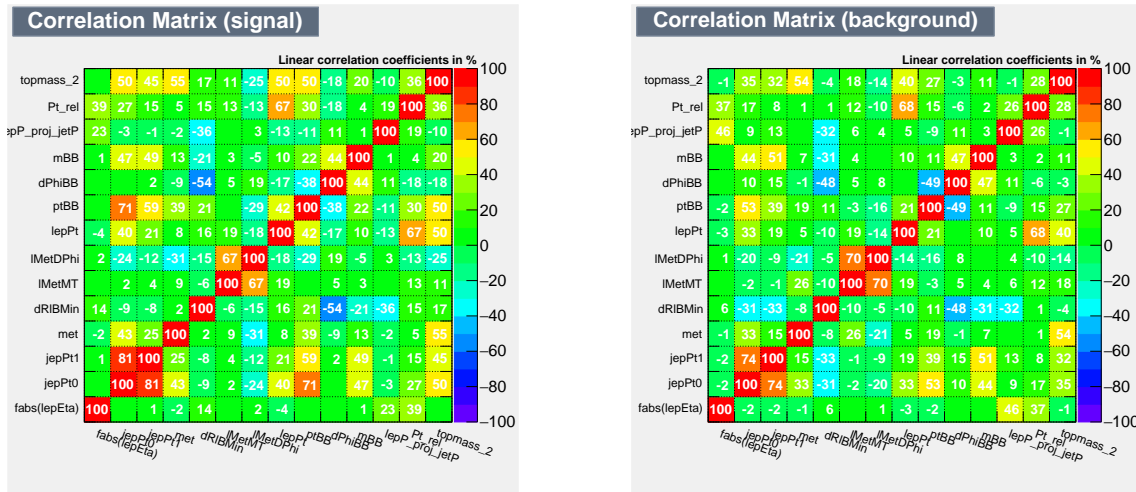






ΣΧΗΜΑ 4.8: Μεταβλητές Boosted Decision Tree.

Παρακάτω παρατίθενται οι πίνακες των συσχετίσεων των μεταβλητών για κάθε κλάση. Το BDT ως ταξινομητής εκμεταλεύεται τις διαφορές των συσχετίσεων ανάμεσα στο σήμα και το υποβάθρο για να ενισχύσει το διαχωρισμό.



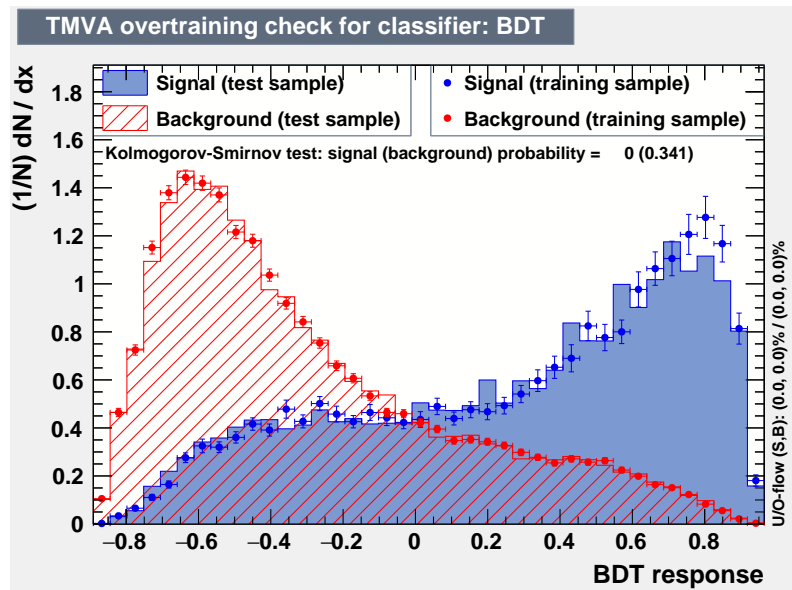
(α)

(β)

ΣΧΗΜΑ 4.9: Πίνακες συσχετίσεων α)για το σήμα και β)για το υπόβαθρο.

4.7.2 Απόκριση του Boosted Decision Tree - Score

Ο όρος score δηλώνει την απόκριση του boosted decision tree. Στα στοιχεία της κλάσης του σήματος αντιστοιχεί η τιμή 1 ενώ σε αυτά του υποβάθρου η τιμή -1. Εφόσον όμως είναι συνεχής κατανομή, όσο πλησιέστερα είναι στην τιμή 1 η κλάση του σήματος και στην τιμή -1 η κλάση του υποβάθρου τόσο καλύτερος θεωρείται ο διαχωρισμός.



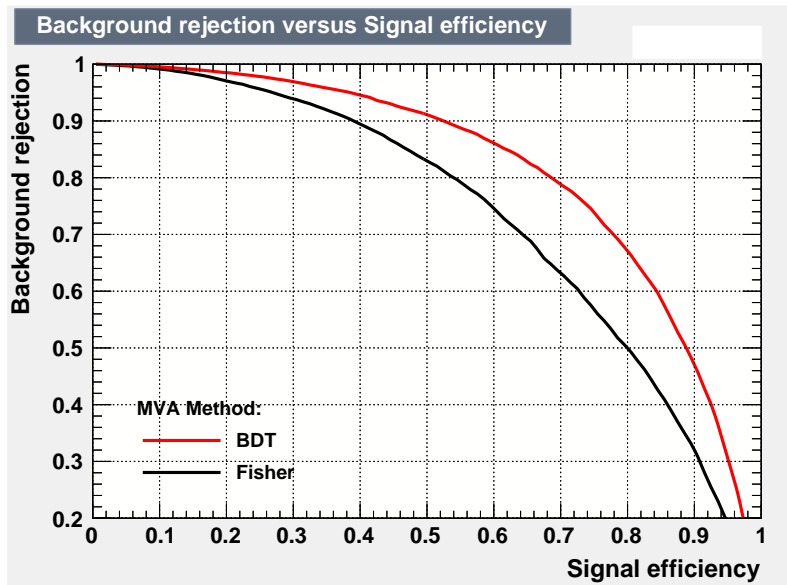
ΣΧΗΜΑ 4.10: Απόκριση του Boosted Decision Tree.

Στο διάγραμμα του σχήματος (4.10) απεικονίζεται η απόκριση από το δείγμα εκπαίδευσης και από το δείγμα ελέγχου (test sample). Αυτός είναι ένας έλεγχος του ταξινομητή για την υπερεκπαίδευση. Με μία πρώτη ματιά φαίνεται η περιοχή του σήματος να έχει υποστεί overtraining. Οι συνθήκες αρχικοποίησης του BDT (shrinkage rate, neuts, maxdepth) όπως εξηγείται παραπάνω, έχουν επιλεγθεί έτσι ώστε να ευνοούν τη σταθερότητα και τη μείωση της πιθανότητας για υπερεκπαίδευση. Είναι γεγονός ότι τα Monte Carlo δείγματα του σήματος σε leading order υποφέρουν από χαμηλή στατιστική, γεγονός που γίνεται εμφανές και από τα σφάλματα του δείγματος εκπαίδευσης. Θεωρώντας ότι το δείγμα του testing έχει αντίστοιχα μεγάλα σφάλματα, αυτή η απόκλιση είναι μέσα στα όρια αυτών. Επίσης, η εφαρμογή του BDT στα δεδομένα, που είναι δείγμα γεγονότων ανεξάρτητο από το δείγμα της εκπαίδευσης, δείχνει ικανοποιητική συμφωνία με την απόκριση του ταξινομητή. Τέλος, πολύ βασικό επιχείρημα εδώ αποτελεί ότι η απόκριση του BDT δεν θα χρησιμοποιηθεί για την επιλογή ενός working point σύμφωνα με το οποίο θα καθοριστεί μία efficiency του σήματος. Σκοπός είναι να εκτιμηθεί μέσω fit ο αριθμός των γεγονότων του σήματος. Η διεργασία του σήματος αφήνεται χωρίς περιορισμούς στο fit, όπως είναι αναμενόμενο. Επομένως, τυχόν μικρό mismodeling στο σχήμα του καλύπτεται από αυτή την αβεβαιότητα.

4.7.3 Καμπύλη ROC

Η καμπύλη ROC (Receiver Operating Characteristic) αποτελεί ένα μέτρο της απόδοσης του boosted decision tree. Για το σχεδιασμό της σαρώνεται κάθε τιμή της απόκρισης (score) του BDT, και υπολογίζεται η πιθανότητα να απορριφθεί το υπόβαθρο (background rejection) και η πιθανότητα να διατηρηθεί το σήμα (signal efficiency). Οι πιθανότητες αυτές δίνονται

από την ολοκλήρωση της καμπύλης της απόκρισης για την ποσότητα που απορρίπτεται ή διατηρείται αντίστοιχα.



ΣΧΗΜΑ 4.11: Καμπύλη ROC του Boosted Decision Tree.

4.8 Σύγκριση της απόκρισης του BDT στα Monte Carlo δείγματα και στα δεδομένα

Μετά την εκπαίδευση και την αξιολόγηση, η μέθοδος BDT χρησιμοποιήθηκε για την ταξινόμηση στα πλήρη δείγματα Monte Carlo και στο δείγμα των δεδομένων με άγνωστο το σήμα και τη σύνθεση του υποβάθρου. Για το W+Jets, στην εφαρμογή του BDT χρησιμοποιήθηκαν τα δείγματα προσομοίωσης ανώτερης τάξης (Next to leading order-NLO) λόγω μειωμένης στατιστικής στα δείγματα LO.

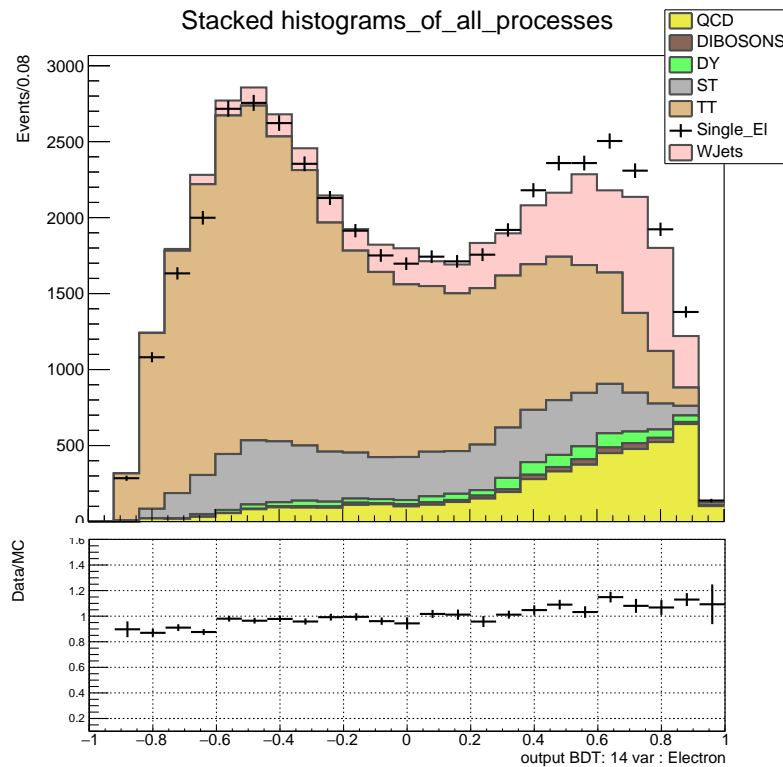
Για το συνολικό σχεδιασμό των επιμέρους διεργασιών αλλά και για το σύνολο τους χρησιμοποιήθηκαν τα κατάλληλα βάρη για την ανθροισή τους. Τα βάρη αυτά είναι η luminosity (φωτεινότητα) επί το λόγο της ενεργής διατομής του κάθε δείγματος προς το συνολικό αριθμό των γεγονότων του δείγματος αυτού.

$$w = \text{luminosity} \frac{\text{cross section}}{\text{number of generated events}} \quad (4.7)$$

Οι αλγόριθμοι για την ταυτοποίηση και ανακατασκευή των λεπτονίων έχουν μία διαφορά απόδοσης μεταξύ πραγματικών δεδομένων και Monte Carlo. Συνεπώς, πρέπει να εφαρμοστούν στα δείγματα προσομοίωσης πολλαπλασιαστικοί παράγοντες, ώστε να υπάρχει ένα πιο συμβατό αποτέλεσμα. Οι διορθωτικοί αυτοί παράγοντες, δηλαδή οι efficiencies των λεπτονίων, εφαρμόστηκαν ανά γεγονός. Τα ηλεκτρόνια έχουν τρεις efficiencies: μία για την

ανακατασκευή, μία για την ταυτοποίηση και μία για τον trigger. Τα μόνια έχουν δύο γιατί δεν ανακατασκευάζονται.

Στο Σχήμα 4.12 απεικονίζεται το stack ιστόγραμμα των διεργασιών, όπως προκύπτουν από τα Monte Carlo, και τα δεδομένα σε συνδιασμό με το διάγραμμα του λόγου δεδομένων προς προσομοίωση. Και στον πίνακα παρακάτω παρατίθεται ο αριθμός των γεγονότων ανά κάθε διεργασία.



ΣΧΗΜΑ 4.12: Boosted Decision Tree για όλες τις διεργασίες και για τα δεδομένα. Στο ιστόγραμμα είναι αθροισμένες όλες οι διεργασίες και τα σημεία με τα σφάλματα δείχνουν το αποτέλεσμα στα δεδομένα.

| Yields | |
|--------------|---------|
| $W+b\bar{b}$ | 5999,7 |
| Drell-Yan | 954,2 |
| $t\bar{t}$ | 27044,1 |
| Single top | 6310,4 |
| Diboson | 345,9 |
| Qcd | 4562,8 |
| Data | 45217 |

ΠΙΝΑΚΑΣ 4.2: Πίνακας με τον αριθμό γεγονότων από κάθε διεργασία για το ιστόγραμμα της εικόνας 4.12.

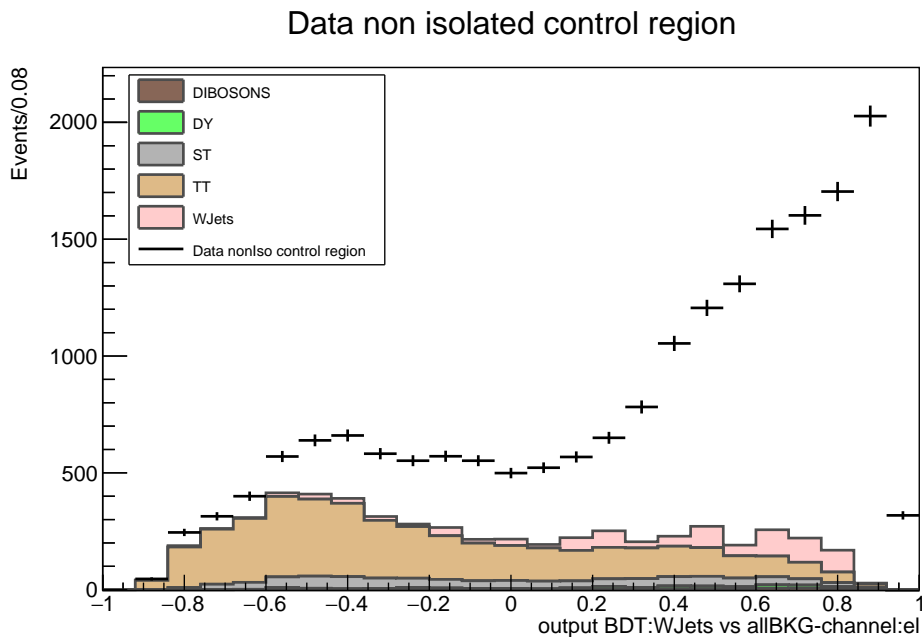
Όλες οι διεργασίες υπολογίστηκαν από τα δείγματα Monte Carlo εκτός από την QCD η οποία υπολογίστηκε από τα δεδομένα. Εκτενέστερη περιγραφή ακολουθεί στην επόμενη παράγραφο.

4.9 Data Driven Estimation-QCD multijet

Η QCD-multijet είναι δύσκολο υπόβαθρο να προσομοιωθεί επαρκώς [20]. Δεδομένης της πολυπλοκότητας στη μοντελοποίηση τέτοιων διεργασιών, η τεχνική εκτίμησης της από τα δεδομένα (Data Driven method) αποτελεί βασική μέθοδο που χρησιμοποιείται συχνά σε μελέτες στη φυσική υψηλών ενεργειών.

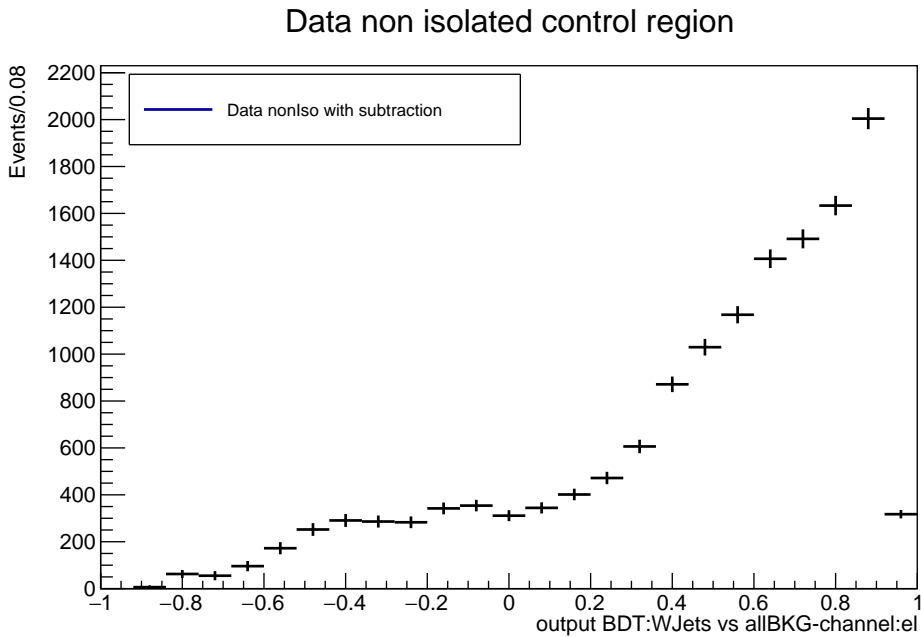
Για να υλοποιηθεί η μέθοδος πρέπει πρωτίστως να καθοριστεί μία control περιοχή, δηλαδή μία περιοχή του φασικού χώρου εμπλουτισμένη με QCD-multijet γεγονότα που θα έχει χαμηλή «μόλυνση» (contamination) από άλλες διεργασίες.

Σύμφωνα με τα selection cuts για το καθορισμό της περιοχής του σήματος, βασικής σημασίας για τη μείωση της συνεισφοράς της QCD ήταν το κριτήριο για το isolation του λεπτονίου. Αντιστρέφοντας την απαίτηση για ένα καλώς απομονωμένο λεπτόνιο είναι το κυριότερο βήμα για τον καθορισμό της control περιοχής. Συνεπώς, πρέπει να ελεγχθεί ότι για να γίνει δεκτό ένα γεγονός πρέπει να μην υπάρχει καλώς απομονωμένο λεπτόνιο ενώ υπάρχει ένα όχι καλά απομονωμένο. Μία άλλη προϋπόθεση είναι η επιλογή των triggers. Οι επιλεγμένοι triggers για QCD γεγονότα είναι loose. Αυτές οι αλλαγές στα selection cuts εφαρμόστηκαν στα δεδομένα και στα Monte Carlo και παρακάτω απεικονίζεται η κατανομή των δεδομένων στην control περιοχή και η συνεισφορά σε αυτή από άλλες διεργασίες.



ΣΧΗΜΑ 4.13: Κατανομή των data στην control περιοχή.

Παρατηρείται ότι η συνεισφορά άλλων διεργασιών στην control περιοχή δεν μπορεί να θεωρηθεί αμελητέα. Κυρίως το contamination οφείλεται στη διεργασία $t\bar{t}$ και στην συνέχεια στη single top. Η συνεισφορά των διεργασιών $t\bar{t}$, single top και Drell-Yan αφαιρείται. Η διεργασία Diboson θεωρήθηκε αμελητέα, ενώ το σήμα $W+b\bar{b}$ δεν αφαιρέθηκε καθώς αποτελεί τη διεργασία που έχει στόχο να καθορίσει η ανάλυση αυτή. Μετά την αφαίρεση η κατανομή των δεδομένων δίνεται στο διάγραμμα που ακολουθεί.



ΣΧΗΜΑ 4.14: Κατανομή των data στην control περιοχή με αφαίρεση του contamination.

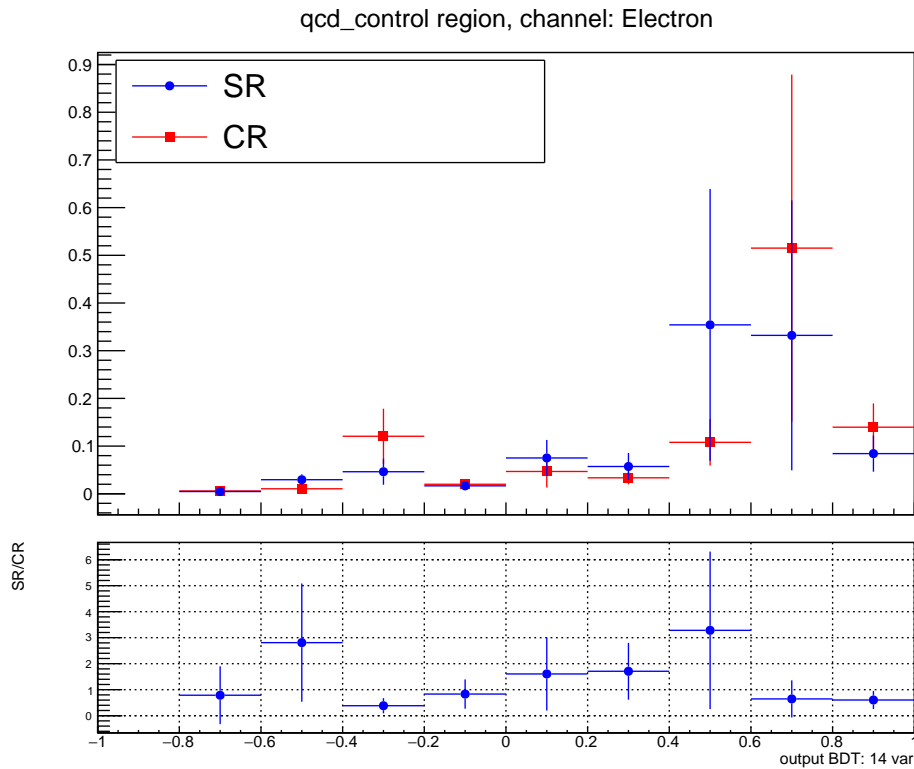
Στη συνέχεια αφού εκτιμηθεί το σχήμα της QCD, χρησιμοποιείται ένας πολλαπλασιαστικός παράγοντας ώστε να καθορισθεί ο αριθμός των γεγονότων της QCD στην περιοχή του σήματος. Αυτό έχει σαν αποτέλεσμα η άθροιση του yield της QCD και των άλλων διεργασιών που προέχονται από την προσομοίωση να δίνουν το yield των δεδομένων. Ο παράγοντας κανονικοποίησης της QCD δίνεται από τη σχέση:

$$QCD^{SF} = \frac{N_{data,SR} - \sum_i N_{MC,SR,i}}{N_{QCD,CR}} \quad (4.8)$$

όπου $N_{data,SR}$ είναι ο αριθμός των γεγονότων στην περιοχή του σήματος για τα δεδομένα και $N_{MC,SR,i}$ για το δείγμα προσομοίωσης i , ενώ $N_{QCD,CR}$ είναι ο αριθμός των QCD γεγονότων στην control περιοχή, όπως προέκυψε μετά την αφαίρεση του contamination.

Θεωρητικά θεωρούμε ότι το σχήμα της QCD στην control περιοχή θα είναι ίδιο με αυτό στην περιοχή του σήματος. Για να ελεγχθεί αυτός ο ισχυρισμός υλοποιήθηκαν closure tests σε Monte Carlo δείγματα της QCD. Θεωρούνται δείγματα μεγάλης καθαρότητας σε γεγονότα QCD. Επομένως, η περιοχή του σήματος και η control περιοχή αναμένεται να έχουν ίδιο σχήμα

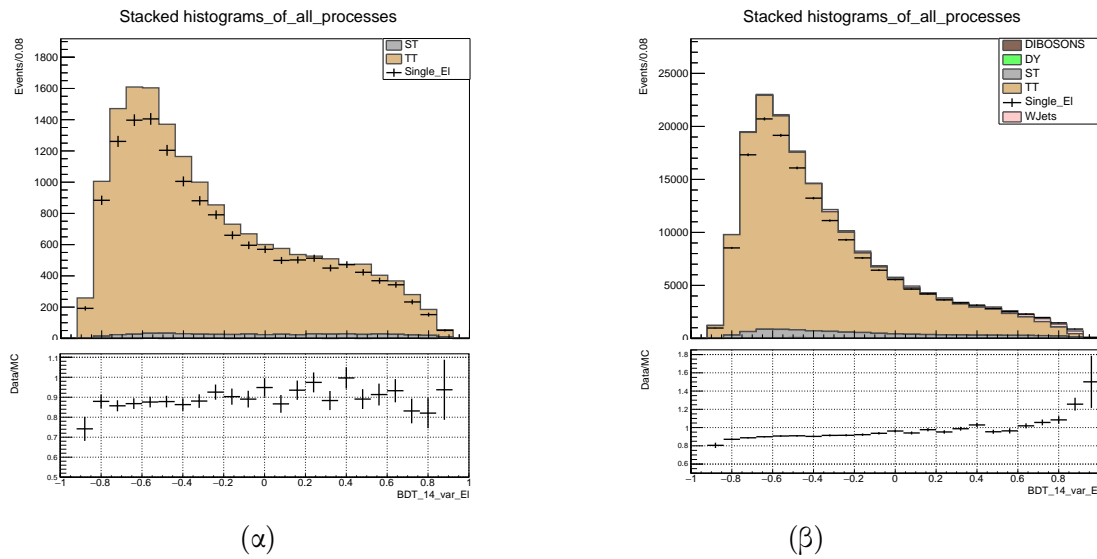
κατανομής. Στο Σχήμα 4.15 αυτό επιβεβαιώνεται και για αυτό μπορεί να χρησιμοποιηθεί η μέθοδος εκτίμησης της QCD από τα δεδομένα.



ΣΧΗΜΑ 4.15: Closure test της qcd-multijet διεργασίας στο κανάλι των ηλεκτρονίων.

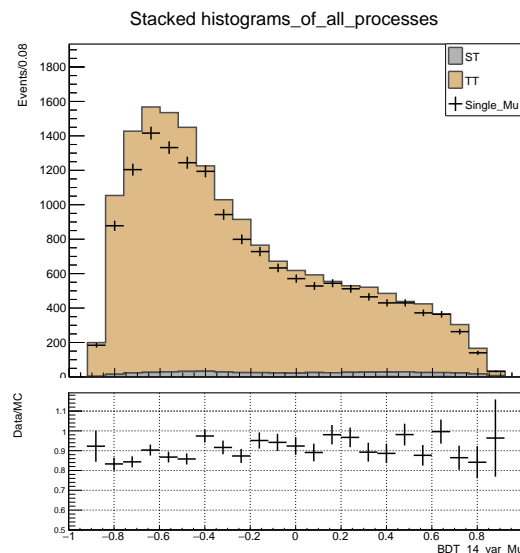
4.10 control regions for $t\bar{t}$

Για να επιβεβαιωθεί ότι η συνεισφορά του υποβαθρού $t\bar{t}$ περιγράφεται καλά από τα δείγματα Monte Carlo, ορίζονται δύο control περιοχές. Η πρώτη, dilepton, καθορίζεται διατηρώντας όλες τις υπόλοιπες απαιτήσεις που ορίζουν το φασικό χώρο του σήματος $W + b\bar{b}$, εκτός ότι ζητούνται δύο καλά απομονωμένα λεπτόνια, διαφορετικής γεύσης, ώστε να μειωθεί η συνεισφορά από τη Drell-yan διεργασία, και αντίθετου φορτίου και διευρύνεται ο περιορισμός για τον αριθμό των πιδάκων από ίσο με δύο σε μεγαλύτερο ή ίσο με δύο. Για τη δεύτερη, multijet control περιοχή, απαιτείται περισσότερη δραστηριότητα από πίδακες. Συγκεκριμένα, απαιτήθηκε πολλαπλότητα σε πίδακες τουλάχιστον ίση με τέσσερα. Οι περιοχές αυτές κυριαρχούνται από γεγονότα $t\bar{t}$ με μικρό contamination, κυρίως από διεργασίες single-top.



ΣΧΗΜΑ 4.16: $t\bar{t}$ control περιοχές στο κανάλι των ηλεκτρονίων: (α)dilepton και (β)multijet

Παρατηρείται ότι συστηματικά και στις δύο περιοχές η προσομοίωση υπερτερεί των δεδομένων, γεγονός που δείχνει ότι σε αυτό το φασικό χώρο ενδεχομένως το $t\bar{t}$ δεν έχει μοντελοποιηθεί καλά από την προσομοίωση, αλλά έχει υπερεκτιμηθεί. Το ίδιο παρατηρείται και για τα μόνια στο Σχήμα 4.17.

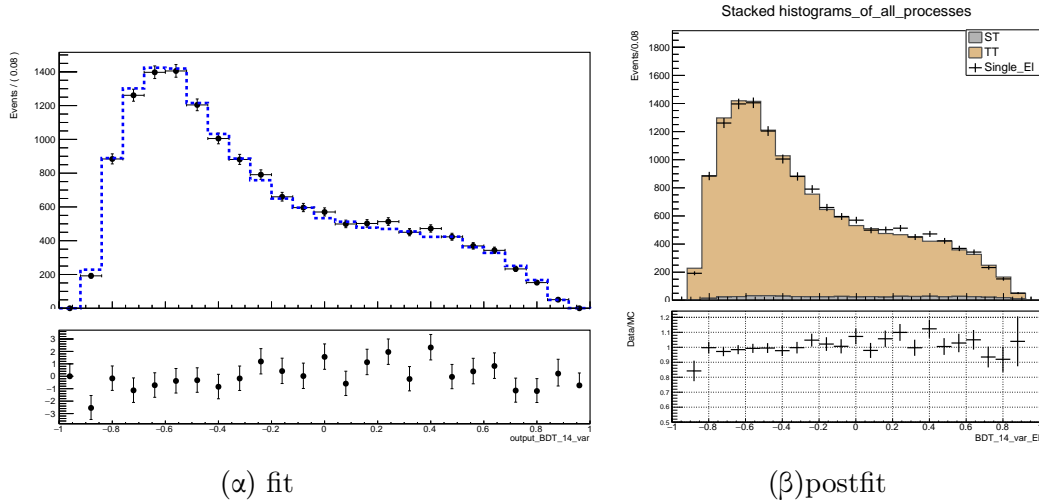


ΣΧΗΜΑ 4.17: Dilepton $t\bar{t}$ control περιοχή στο κανάλι των μιονίων.

Υλοποιήθηκε fit για να προσδιοριστεί ένας διορθωτικός πολλαπλασιαστικός παράγοντας για την διεργασία $t\bar{t}$. Το fit, που περιγράφεται αναλυτικότερα σε επόμενη παράγραφο, υλοποιήθηκε για την dilepton περιοχή που έχει μεγαλύτερη καθαρότητα σε γεγονότα $t\bar{t}$.

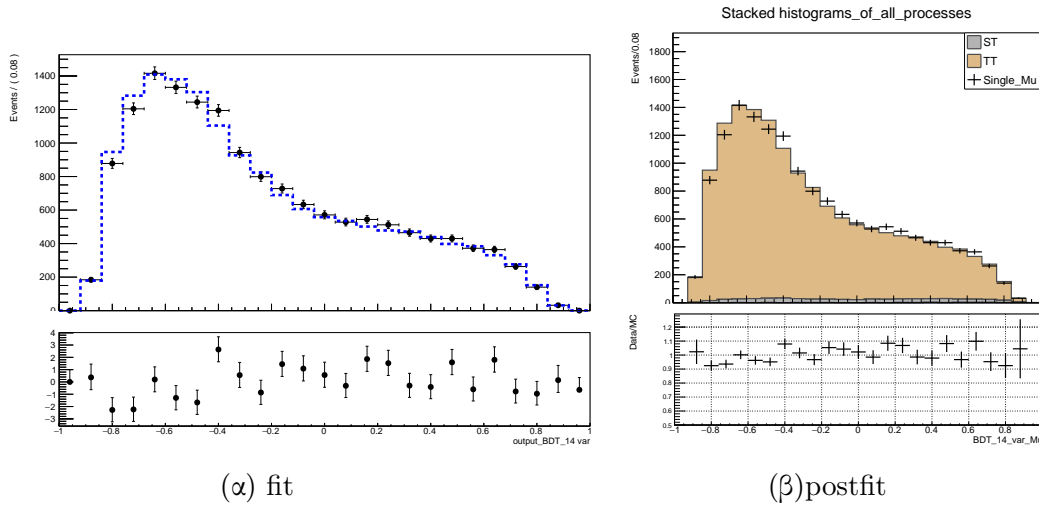
Η διεργασία του single-top αφαιρέθηκε να κυμανθεί μέσα στο όριο του $\pm 10\%$ της αρχικής τιμής του. Παρατίθενται τα διαγράμματα από τα fit στα δύο κανάλια διάσπασης.

Ηλεκτρόνια



ΣΧΗΜΑ 4.18: fit στην dilepton control $t\bar{t}$ περιοχή στο κανάλι των ηλεκτρονίων.

Μιόνια



ΣΧΗΜΑ 4.19: fit στην dilepton control $t\bar{t}$ περιοχή στο κανάλι των μιόνιων.

Ο νέος αριθμός γεγονότων για το $t\bar{t}$ δίνεται από τη σχέση:

$$N_{t\bar{t}}^{final} = SF_{t\bar{t}} * N_{t\bar{t}}^{initial}$$

όπου ο πολλαπλασιαστικός παράγοντας βρέθηκε:

$$SF_{t\bar{t}}^e = 0.88 \pm 0.01$$

για τα ηλεκτρόνια, και για τα μίονια:

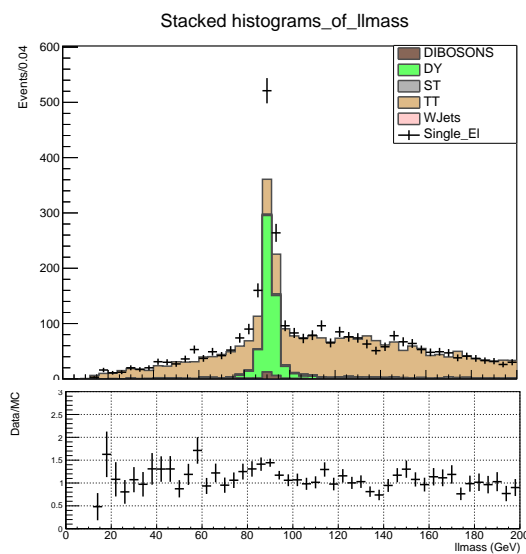
$$SF_{t\bar{t}}^\mu = 0.90 \pm 0.01$$

4.11 control region for Drell-Yan

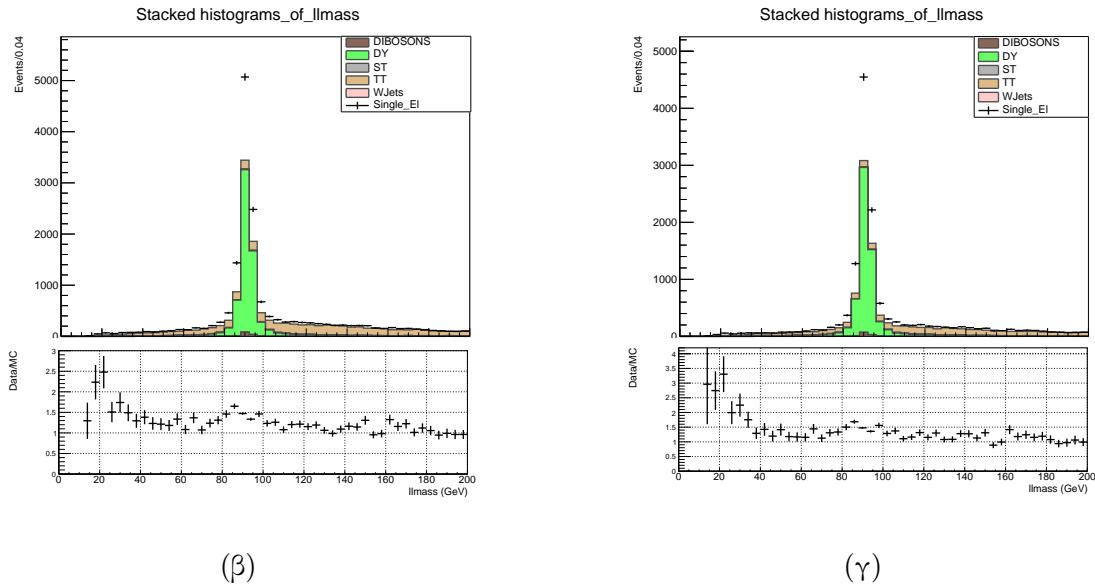
Το υπόβαθρο Drell-Yan καθορίζεται σε μία control περιοχή όπου διατηρούνται όλες οι προϋποθέσεις που απαιτήθηκαν για το σήμα $W+b\bar{b}$, εκτός του ότι απαιτήθηκαν δύο, αντί ενός, απομονωμένα λεπτόνια ίδιου flavor και αντίθετου φορτίου. Η περιοχή αναφέρεται ως $DY+b\bar{b}$ περιοχή. Η μεταβλητή που χρησιμοποιήθηκε είναι η ανακατασκευασμένη μάζα των δύο λεπτονίων. Για Drell-Yan διεργασίες η μεταβλητή αυτή εμφανίζει κορυφή γύρω από τη μάζα του Z , στα 91.2GeV , ενώ για άλλες διεργασίες δίνει μία σχεδόν επίπεδη κατανομή.

Η $DY+b\bar{b}$ περιοχή εμφανίζει contamination από $t\bar{t}$ διεργασίες, και για αυτό υλοποιήθηκαν δύο «καθαρότερες» control περιοχές για το DY υπόβαθρο. Η μία καθορίστηκε απαιτώντας δύο πίδακες, χωρίς καμία απαίτηση για b-tagged και αναφέρεται ως $DY+jj$ περιοχή. Η άλλη απαιτώντας δύο πίδακες, όπου μόνο ο ένας να είναι b-tagged και αναφέρεται ως $DY+1b$ -tagged. Αυτές οι περιοχές μειώνουν τη συνεισφορά του $t\bar{t}$ σε σχέση με DY γεγονότα.

Ηλεκτρόνια

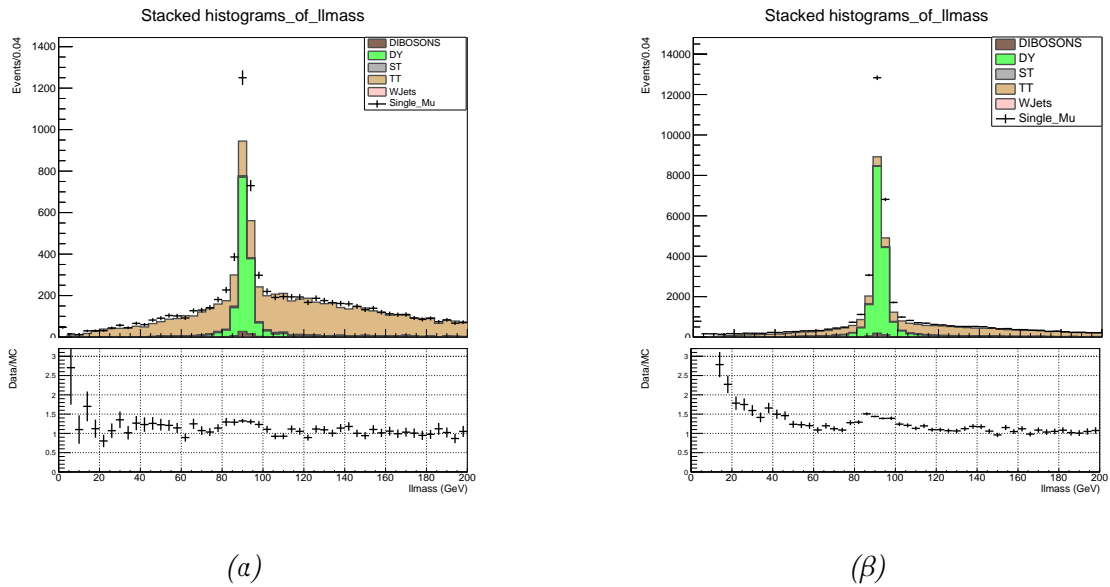


(α)



ΣΧΗΜΑ 4.20: control περιοχές για τη διεργασία Drell-Yan στο κανάλι των ηλεκτρονίων: (α) DY+bb (β)DY+jj και (γ) DY+1b-tagged

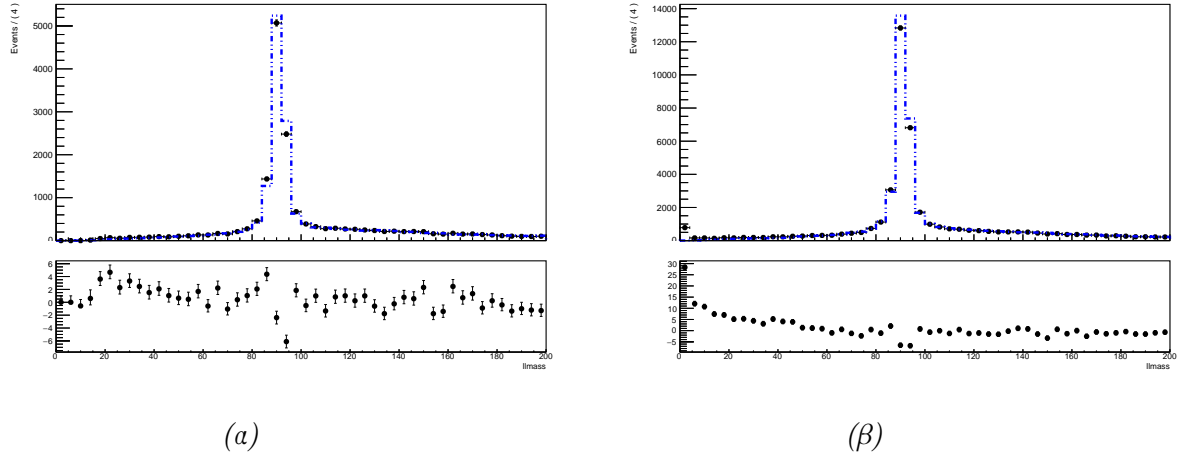
Μιόνια



ΣΧΗΜΑ 4.21: control περιοχές για τη διεργασία Drell-Yan στο κανάλι των μιόνων: (α) DY+bb και (β) DY+jj

Πραγματοποιήθηκε fit στην περιοχή DY+jj, ως περιοχή αρκετά κοντά στο φασικό χώρο του σήματος και μεγάλης καθαρότητας σε γεγονότα Drell-Yan. Στη διεργασία $t\bar{t}$ έχει εφαρμοστεί

ο πολλαπλασιαστικός παράγοντας που βρέθηκε από την control $t\bar{t}$ περιοχή (όπως αναφέρεται στην προηγούμενη παράγραφο) και το εύρος διακύμανσής του καθορίστηκε $\pm 5\%$.



ΣΧΗΜΑ 4.22: fits control περιοχή Drell-Yan+jj: (α) ηλεκτρόνια και (β) μίονια

Ο νέος αριθμός γεγονότων για την Drell-Yan δίνεται από τη σχέση:

$$N_{DY}^{final} = SF_{DY} * N_{DY}^{initial}$$

όπου ο πολλαπλασιαστικός παράγοντας βρέθηκε:

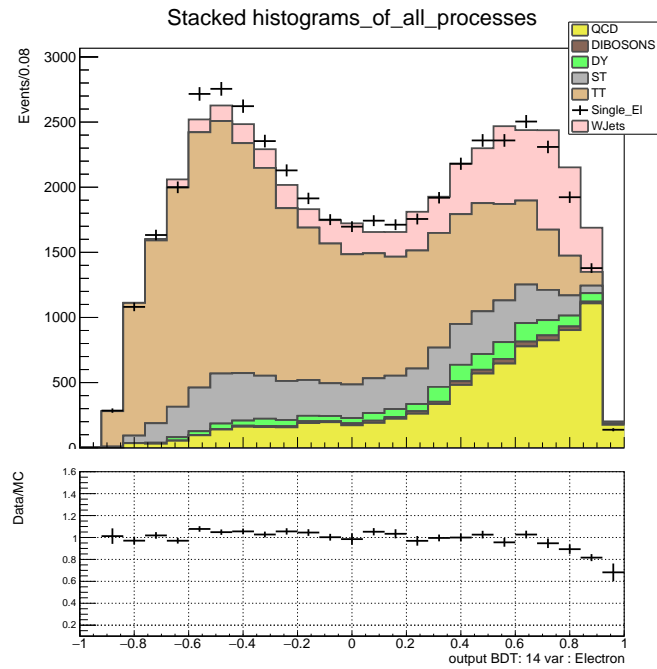
$$SF_{DY}^e = 1,52 \pm 0,02$$

για τα ηλεκτρόνια, και για τα μίονια:

$$SF_{DY}^\mu = 1,48 \pm 0,01$$

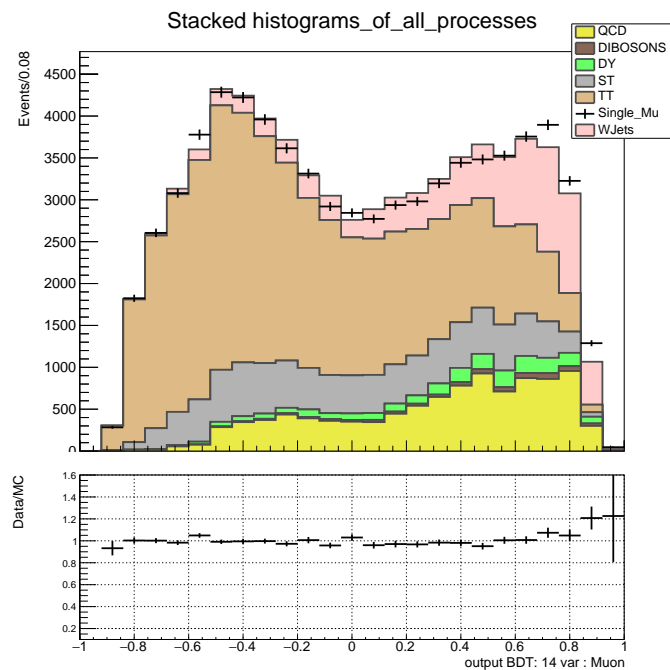
Τελικό stack ιστόγραμμα μετά την εφαρμογή των $SF_{t\bar{t}}$ και SF_{DY}

Για την διεργασία Single-top χρησιμοποιήθηκε ένας πολλαπλασιαστικός παράγοντας 0,90 που υπολογίστηκε από άλλη ανάλυση μέλους της ομάδας. Μετά την εκτίμηση και την εφαρμογή των παραπάνω πολλαπλασιαστικών παράγοντων το διορθωμένο stack ιστόγραμμα όλων των διεργασιών για το BDT φαίνεται παρακάτω.



ΣΧΗΜΑ 4.23: Stack ιστόγραμμα του output του BDT με τους scale factors για το κανάλι των ηλεκτρονίων.

Όμοια η απόκριση του αντίστοιχου BDT για τα μόνια μετά την εκτίμηση της qcd διεργασίας και των control περιοχών $t\bar{t}$ και Drell-Yan δίνεται στο διάγραμμα της εικόνας που ακολουθεί.



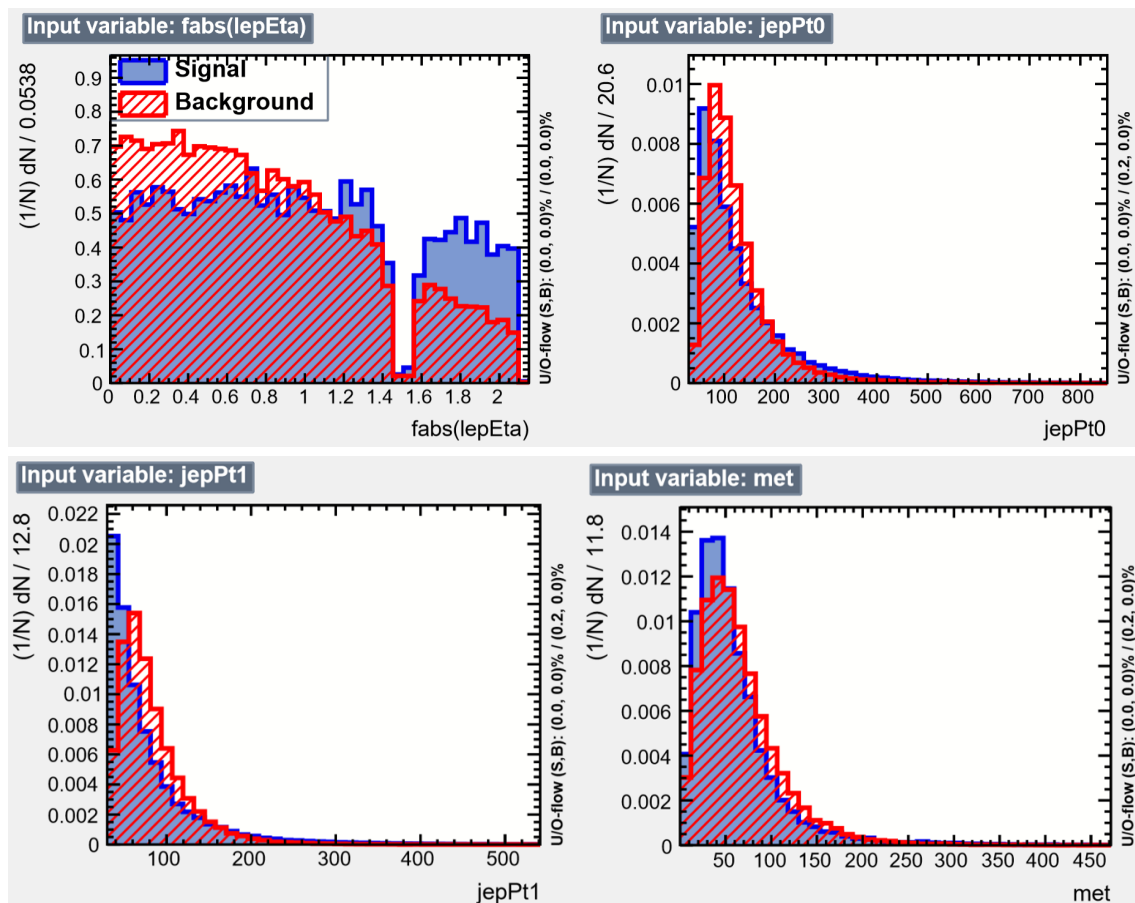
ΣΧΗΜΑ 4.24: Stack ιστόγραμμα του output του BDT με τους scale factors για το κανάλι των μιονίων.

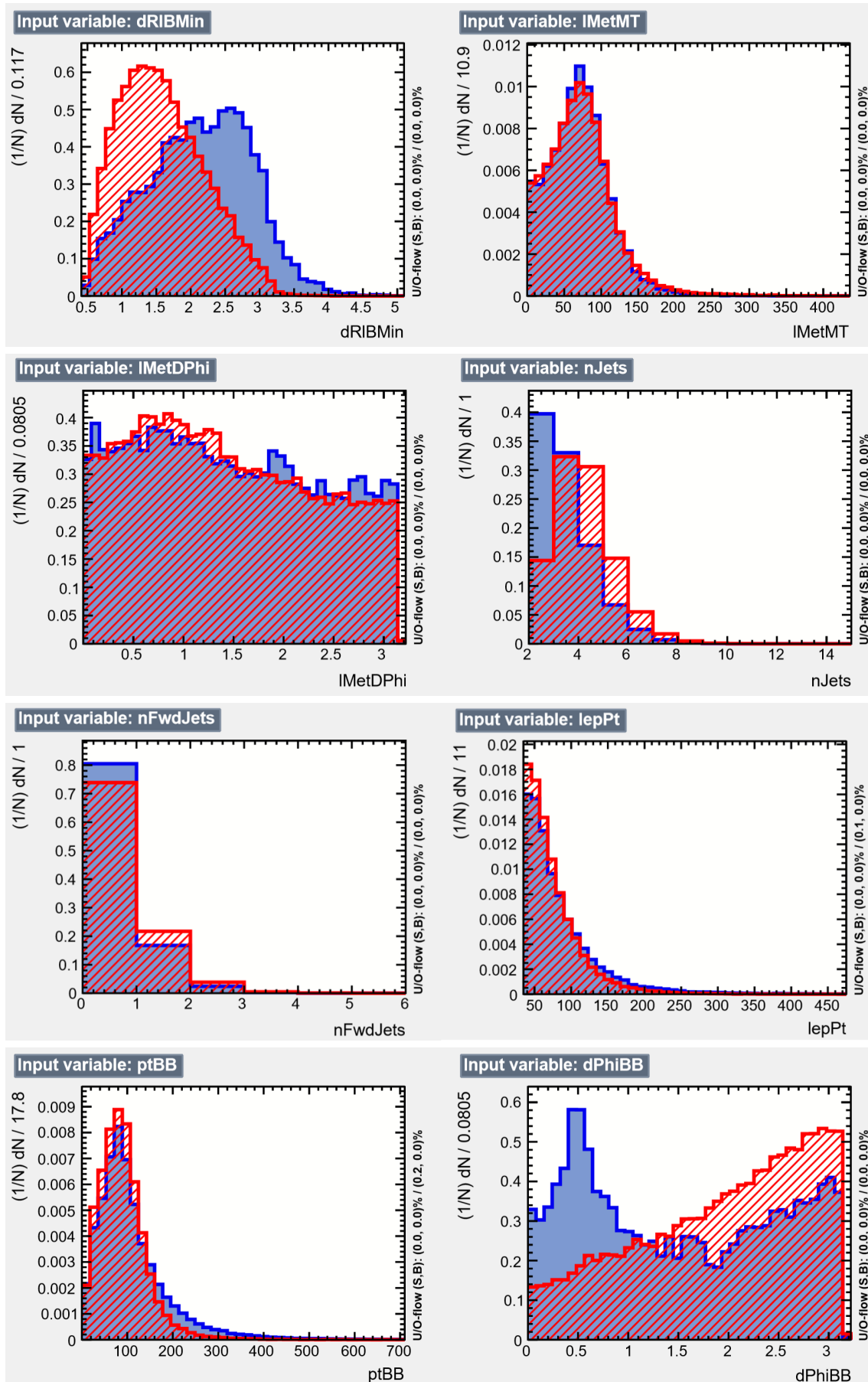
4.12 BDT 2: ορισμένο σε διευρυμένο φασικό χώρο

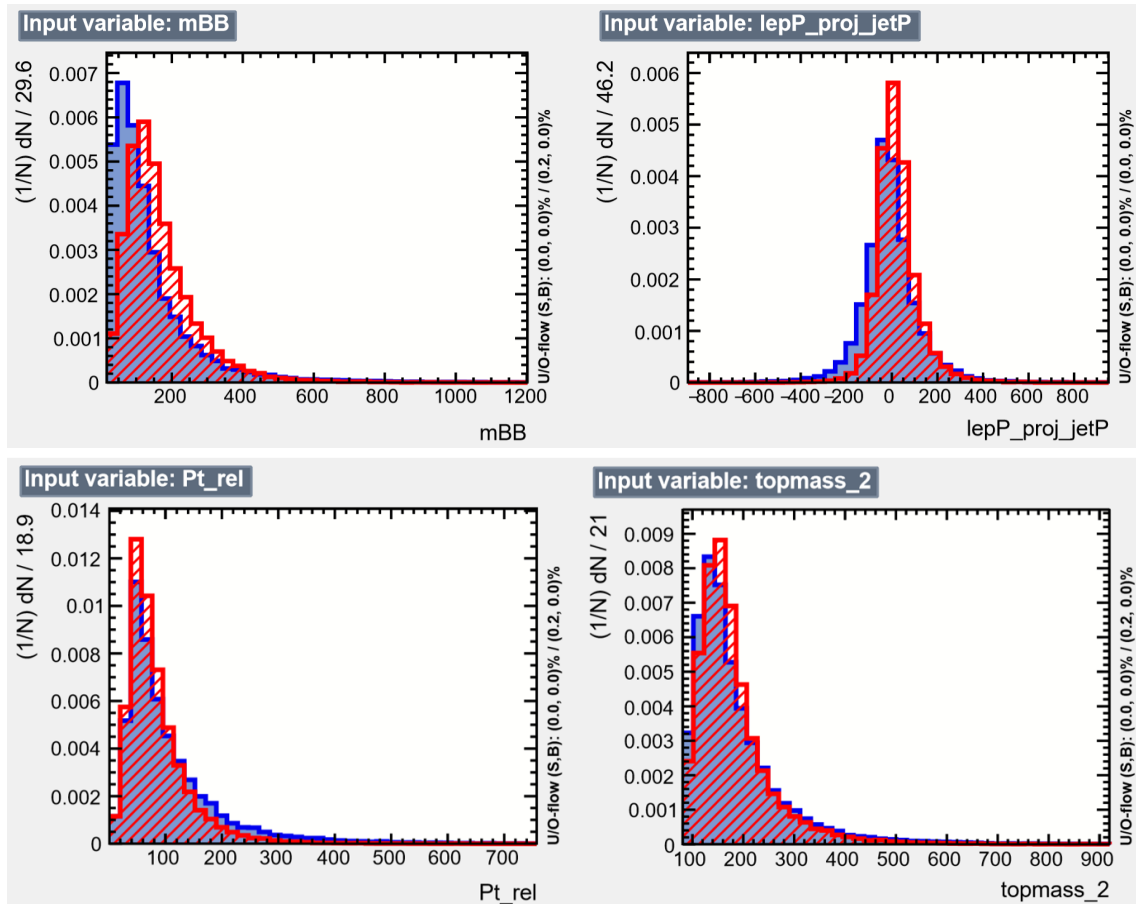
Στην συνέχεια υλοποιήθηκε ένα δεύτερο BDT όπου αφού αυξήθηκε η πολλαπλότητα του αριθμού των πιδάκων από ίσο σε μεγαλύτερο ή ίσο με δύο ($n_{Jets} \geq 2$) και αφαιρέθηκε από τους περιορισμούς επιλογής ο αριθμός των forward πιδάκων προστέθηκαν στην εκπαίδευση ως μεταβλητές. Η επιλογή αυτή να διευρυνθεί ο φασικός χώρος αναμένεται να αυξήσει έντονα τη συνεισφορά των υποβάθρων $t\bar{t}$ και single-top και ειδικά του $t\bar{t}$. Αυτό δεν αποτελεί πρόβλημα καθώς αναμένεται από το BDT που αποτελεί μία «έξυπνη» μεταβλητή, να τα διαχωρίσει.

4.12.1 Μεταβλητές

Οι κατανομές των μεταβλητών αυτού του BDT φαίνονται στα σχήματα που ακολουθούν.

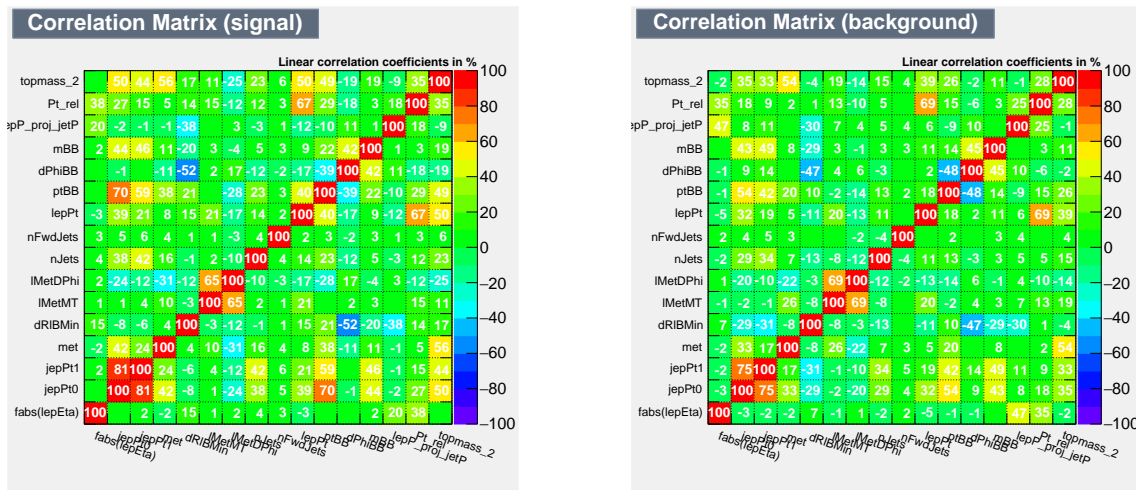






ΣΧΗΜΑ 4.25: Μεταβλητές Boosted Decision Tree.

Ακολουθούν οι πίνακες των συσχετίσεων όλων των μεταβλητών για τις δύο κλάσεις.



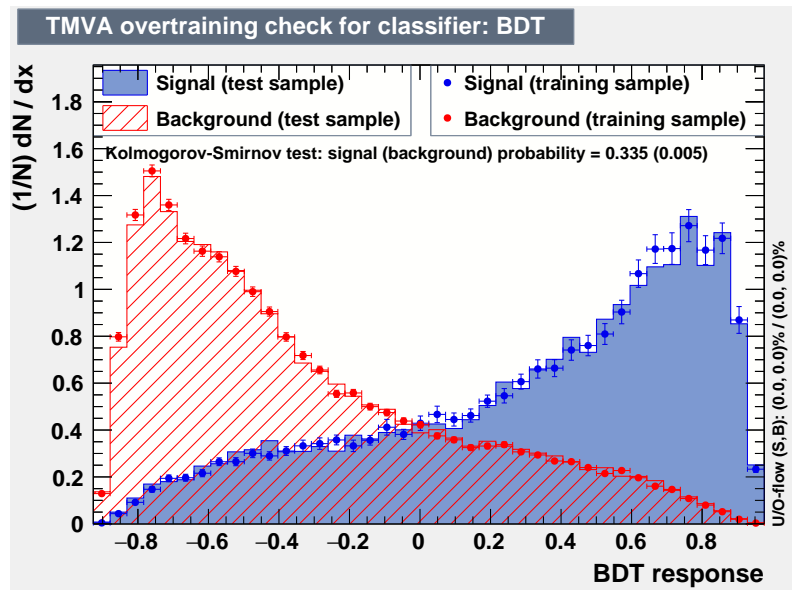
(α)

(β)

ΣΧΗΜΑ 4.26: Πίνακες συσχετίσεων α)για το σήμα και β)για το υπόβαθρο.

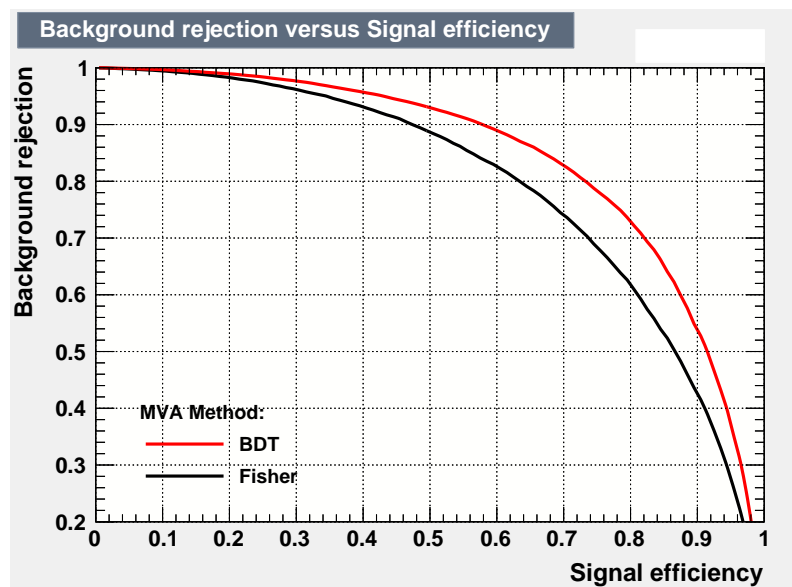
4.12.2 Απόκριση του Boosted Decision Tree - Score

Παρατίθεται η απόκριση του BDT και η καμπύλη ROC.



ΣΧΗΜΑ 4.27: Απόκριση του Boosted Decision Tree.

4.12.3 Καμπύλη ROC



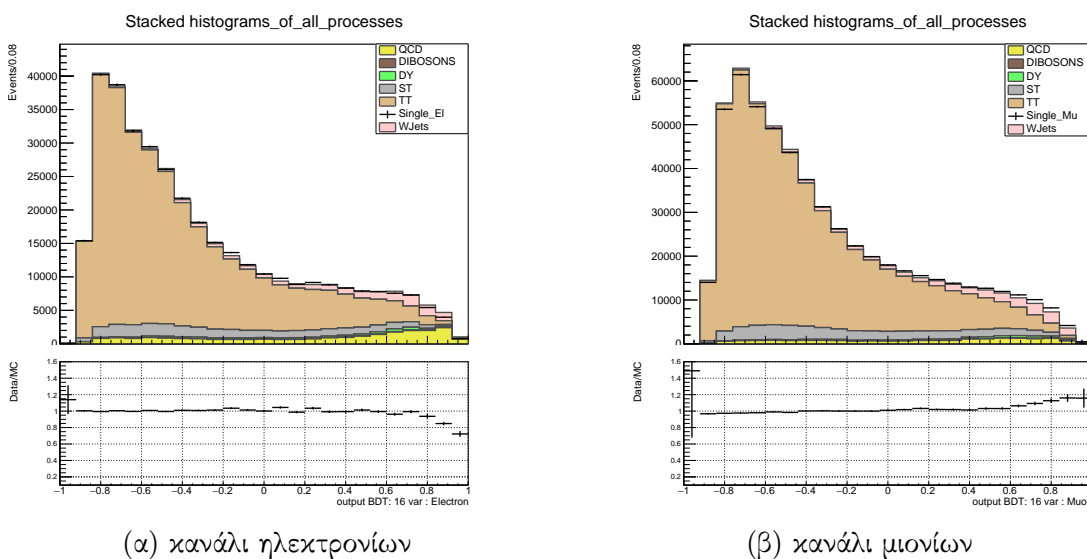
ΣΧΗΜΑ 4.28: Καμπύλη ROC του Boosted Decision Tree.

Η καμπύλη ROC στο Σχήμα 4.28 σε σύγκριση με την αντίστοιχη στο Σχήμα 4.11, εμφανίζει ελαφρώς καλύτερη απόδοση για το δεύτερο BDT που υλοποιήθηκε. Όμως, σκοπός αυτής της μελέτης δεν ήταν η βελτιστοποίηση της διαχωριστικής ικανότητας του σήματος ως προς

το υπόβαθρο, αλλά η δημιουργία μίας μεταβλητής στην οποία οι διάφορες διεργασίες θα παρουσιάζουν διαφορετική κατανομή η καθεμία, ώστε να επιτευχθεί μέσω του fit όσο πιο αξιόπιστη εκτίμηση των γεγονότων του σήματος.

Σύγκριση της απόκρισης του BDT 2 στα Monte Carlo δείγματα και στα δεδομένα.

Ακολουθεί το αποτέλεσμα της εφαρμογής του BDT 2 στα δεδομένα και στα Monte Carlo δείγματα. Χρησιμοποιήθηκαν οι scale factors των control περιοχών για τα δύο κανάλια αντίστοιχα και η διεργασία της qcd εκτιμήθηκε όμοια με πριν από τα δεδομένα.

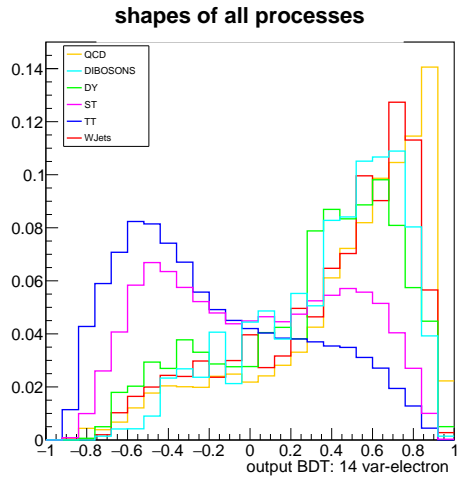


ΣΧΗΜΑ 4.29: Stack ιστόγραμματα της απόκρισης του BDT 2.

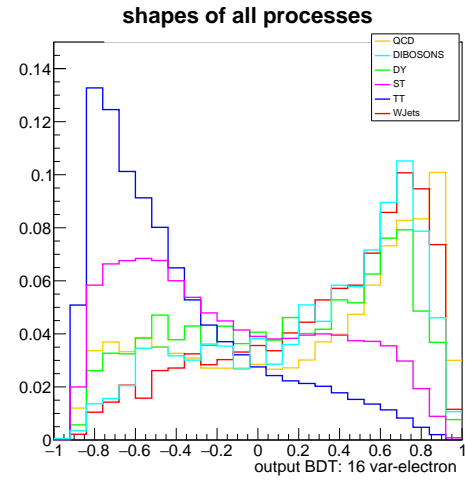
Σύγκριση των κατανομών για όλες τις διεργασίες

Από τα σχήματα των διεργασιών για τα δύο BDT (και στα δύο κανάλια), όπως φαίνονται στις εικόνες (4.30) και (4.31) που ακολουθούν, προκύπτει ότι υπάρχουν διεργασίες και στα δύο με παρόμοιες κατανομές. Αυτές είναι το σήμα $W + b\bar{b}$, με τη Drell-Yan διεργασία, την Diboson και την QCD. Ιδανικά επιθυμείται κάθε διεργασία να διαφοροποιείται σχηματικά όσο το δυνατόν περισσότερο, όμως λόγω κοινών τελικών καταστάσεων δεν είναι πάντα εφικτό.

Ηλεκτρόνια



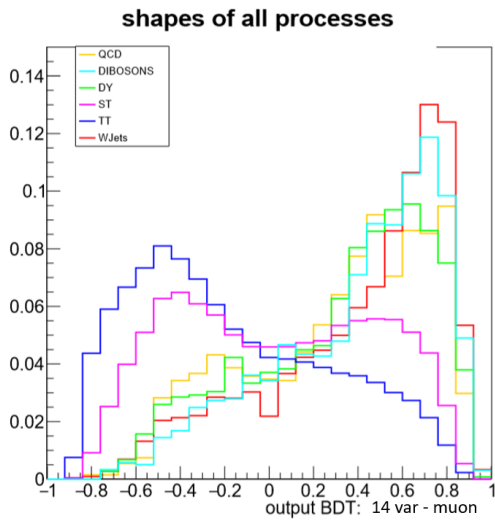
(α) BDT 1



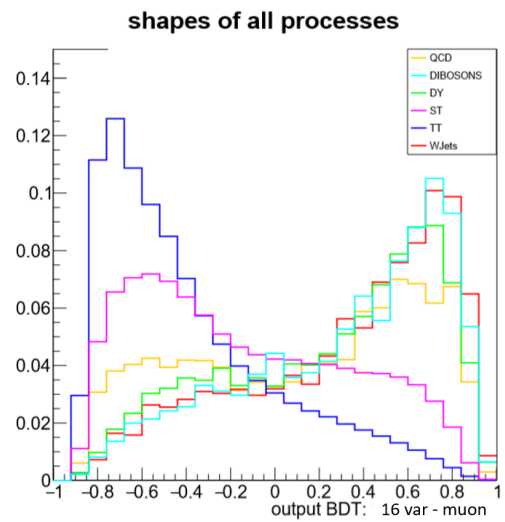
(β) BDT 2

ΣΧΗΜΑ 4.30: Κατανομές στο κανάλι των ηλεκτρονίων .

Μιόνια



(α) BDT 1



(β) BDT 2

ΣΧΗΜΑ 4.31: Κατανομές των BDT στο κανάλι των μιονίων.

4.13 Fits

Για να εκτιμηθεί ο αριθμός των γεγονότων του σήματος, πραγματοποιήθηκε fit του αθροίσματος όλων των διεργασιών, όπως αυτό προέκυψε από τα Monte Carlo, στα δεδομένα. Για το fit χρησιμοποιήθηκε η βιβλιοθήκη RooFit, η οποία χρησιμοποιεί τον αλγόριθμο Minuit για την ελαχιστοποίηση της $-\log(\text{likelihood})$ [21]. Ο αλγόριθμος αυτός είναι ο κύριος που χρησιμοποιείται στη φυσική υψηλών ενεργειών.

Πιθανοφάνεια-likelihood

Ορίζεται ως συνάρτηση πιθανοφάνειας η πυκνότητα πιθανότητας (PDF) που χαρακτηρίζει ένα σύνολο από πειραματικές παρατηρήσεις, προσαρμοσμένες στο σύνολο των δεδομένων[22]. Αν θεωρηθεί αριθμός N ανεξάρτητων παρατηρήσεων του συνόλου $X = [x_1, x_2, \dots, x_n]$ και οι συναρτήσεις πυκνότητας πιθανότητας εξαρτώνται από τις άγνωστες παραμέτρους θ_j , η συνάρτηση πιθανοφάνειας δίνεται:

$$L = \prod_{k=1}^N P(x_i|\theta_j) \quad (4.9)$$

Για μεγαλύτερη ευκολία στον υπολογισμό του μεγιστού χρησιμοποιείται συνήθως ο φυσικός λογάριθμος της συνάρτησης πιθανοφάνειας, όπου λόγω μονοτονίας μεγιστοποιείται στο ίδιο σημείο. Ειδικότερα, προτιμάται η ελαχιστοποίηση της «αρνητικής λογαριθμικής πιθανοφάνειας» (Negative Log Likelihood).

$$-\ln L = -\sum_{k=1}^N \ln P(x_i|\theta_j) \quad (4.10)$$

Στη φυσική συνήθως χρησιμοποιείται η extended likelihood function, όπου ένας επιπλέον όρος (κατανομή Poisson) προστίθεται στη συνάρτηση πιθανοφάνειας.

Η συνάρτηση πυκνότητας πιθανότητας δίνεται:

$$F(x) = f \cdot PDF_{\text{signal}}(x) + (1 - f) \cdot PDF_{\text{background}}(x) \quad (4.11)$$

όπου $f = \frac{N_{\text{signal}}}{N_{\text{signal}} + N_{\text{background}}}$, $1 - f = \frac{N_{\text{background}}}{N_{\text{signal}} + N_{\text{background}}}$

και $N_{\text{expected}} = N_{\text{signal}} + N_{\text{background}}$.

Βήματα για το fit

- Κατασκευάζονται οι κατανομές της απόκρισης του BDT των επιμέρους διεργασιών και των δεδομένων σε ιστογράμματα (binned data).
- Μετατρέπονται σε μη παραμετρικές συναρτήσεις πυκνότητας πιθανότητας (PDF).
- Δημιουργείται ένα αθροιστικό μοντέλο από όλες τις διεργασίες (σήματος και υποβάθρου) με παραμέτρους τον αριθμό των γεγονότων (yields) για κάθε μία από αυτές. Το μοντέλο δίνεται από τη σχέση:

$$F_{model} = N_{Wb\bar{b}}PDF_{Wb\bar{b}} + N_{DY}PDF_{DY} + N_{t\bar{t}}PDF_{t\bar{t}} \\ + N_{St}PDF_{St} + N_{DB}PDF_{DB} + N_{QCD}PDF_{QCD} \quad (4.12)$$

όπου $N_{Wb\bar{b}} + N_{DY} + N_{t\bar{t}} + N_{St} + N_{DB} + N_{QCD} = N_{expected} = N_{data}$.

- Οι παράμετροι αυτοί δεν είναι σταθεροί αλλά αφήνονται να κυμανθούν μέσα σε κάποια όρια:
 - $N_{W+b\bar{b}}$: ο αριθμός των γεγονότων του σήματος αφήνεται ελεύθερος, χωρίς όρια, καθώς είναι η κύρια παράμετρος που καθορίζεται σε αυτή την ανάλυση και δεν είναι επιθυμητό να υπάρχει bias.
 - $N_{t\bar{t}}$ και N_{DY} : οι αριθμοί των γεγονότων των διεργασιών $t\bar{t}$ και Drell-Yan εκτιμήθηκαν από τις control περιοχές (που αναλύθηκαν παραπάνω) με αβεβαιότητα $\pm 0,01$ και $\pm 0,02$ αντίστοιχα, για ηλεκτρόνια, και για μίονια με σφάλμα $\pm 0,01$ και για τις δύο διεργασίες. Παραμένοντας συντηρητικοί ως προς αυτά τα σφάλματα, τα όρια καθόριστηκαν να κυμαίνονται στο $\pm 5\%$ της αρχικής τους τιμής.
 - N_{St} : τα όρια της single-top διεργασίας ορίστηκαν $\pm 30\%$
 - N_{DB} : η διακύμανση αυτού του yield αφαιρέθηκε να είναι $\pm 50\%$ της αρχικής τιμής.
 - N_{QCD} : Ο αριθμός των γεγονότων της qcd διεργασίας εκτιμήθηκε από τα δεδομένα καθώς και το σχήμα της, και η εκτίμηση αυτή έχει μεγάλη αβεβαιότητα. Για αυτό ο αριθμός αυτός αφήνεται επίσης χωρίς όρια να κυμανθεί ελεύθερα κατά την υλοποίηση του fit.

Μετά τον καθορισμό των παραπάνω βημάτων υλοποιείται το fit.

4.13.1 Toy Monte Carlo

Για την επιλογή του ενός από τα δύο BDT για να υλοποιηθεί το παραπάνω fit, διεξάχθηκαν για κάθε ένα από αυτά ένα toy Monte Carlo. Η ονομασία toy είναι για να διευκρινιστεί η διαφορά από τα Monte Carlo samples παραγωγής γεγονότων προσομοίωσης. Ο λόγος για την υλοποίηση αυτών των toy Monte Carlo είναι ότι η ανάλυση απαιτείται να είναι αμερόληπτη και όχι καθοδηγούμενη από τα συγκεκριμένα δεδομένα. Για την υλοποίηση τους χρησιμοποιήθηκε επίσης η βιβλιοθήκη RooFit.

Βήματα για το toy Monte Carlo

- Δημιουργείται ένα μοντέλο όμοιο με το μοντέλο που προέκυψε από την προσομοίωση (4.12):

$$F_{gen} = N_{injected}PDF_{Wb\bar{b}} + N_{DY}PDF_{DY} + N_{t\bar{t}}PDF_{t\bar{t}} \\ + N_{St}PDF_{St} + N_{DB}PDF_{DB} + N_{QCD}PDF_{QCD}$$

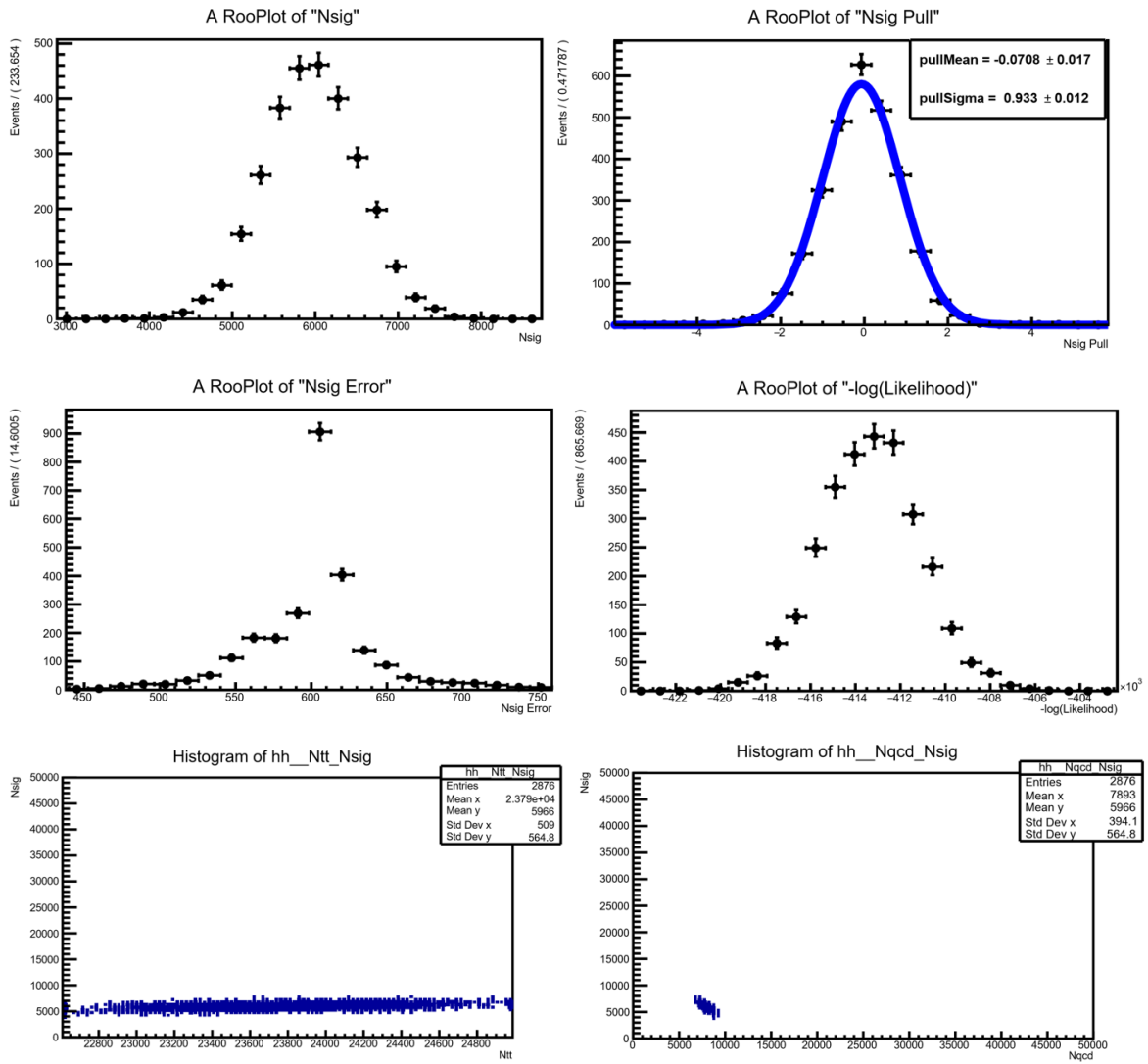
όπου εδώ οι παράμετροι είναι σταθεροί και $N_{injected} = kN_{W+b\bar{b}}^{initial}$ είναι ο αριθμός γεγονότων που απαιτούνται να γεννηθούν από τη διεργασία του σήματος. Το μοντέλο αυτό είναι ο γεννήτορας των ψευδοδεδομένων σε κάθε κύκλο και τα γεγονότα που δημιουργούνται ακολουθούν κατανομή Poisson. Για τον αριθμό των γεγονότων του σήματος, που έγιναν generated, ισχύει η σχέση:

$$N_{gen}^{signal} = N_{gen}^{total} \frac{N_{injected}}{N_{injected} + N_{DY} + N_{t\bar{t}} + N_{St} + N_{DB} + N_{QCD}}$$

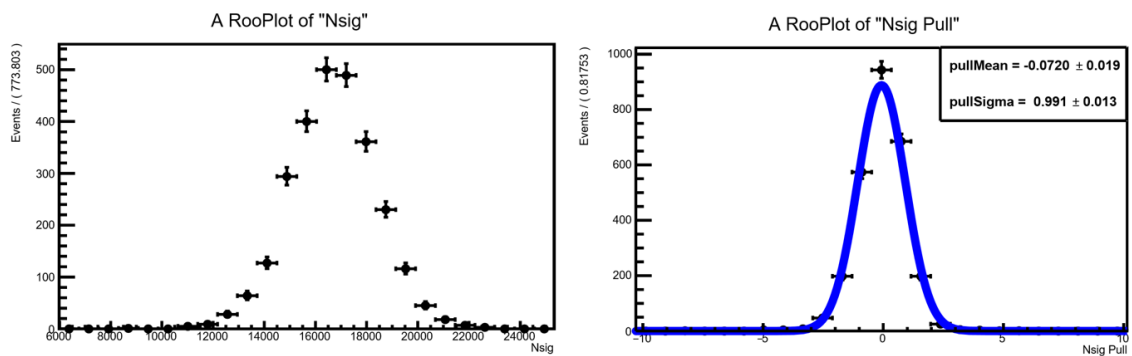
(Και αντίστοιχα ισχύει για κάθε διεργασία.)

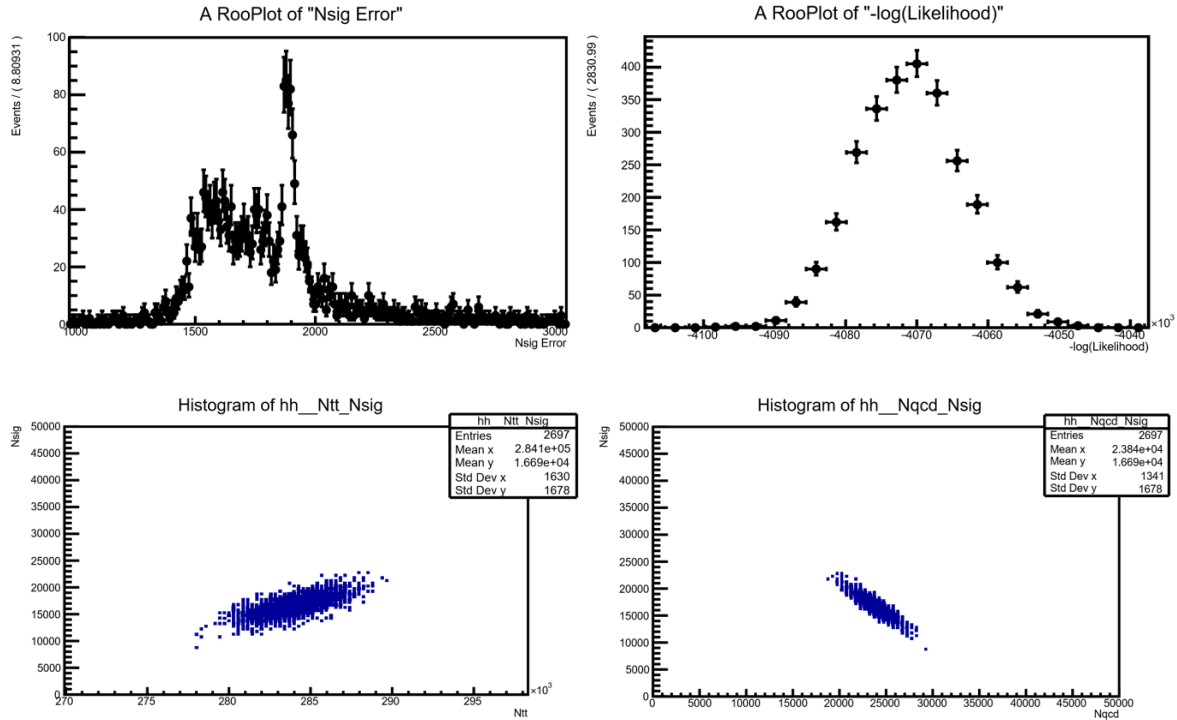
- Πραγματοποιείται ένα extended likelihood fit, του μοντέλου της σχέσης (4.12) στα γεγονότα που παράχθηκαν, με τους περιορισμούς που αναφέρθηκαν παραπάνω.
- Επαναλαμβάνεται η παραγωγή γεγονότων και το fit για ένα κύκλο επαναλήψεων. Στη παρούσα ανάλυση πραγματοποιήθηκαν 4000 επαναλήψεις.

Με αυτή τη διαδικασία συγκρίνοντας τα toys για τα δύο BDTs επιλέχθηκε το ορθότερο αποτέλεσμα. Αυτό είναι που πληρεί τις ιδιότητες της αμεροληψίας (unbiased) και της συνέπειας (consistent). Παρακάτω παρατίθενται τα αποτελέσματα των δύο toys Monte Carlo, για το κανάλι των ηλεκτρονίων, για $N_{injected} = N_{W+b\bar{b}}$, δηλαδή $r_{injected} = 1$.



ΣΧΗΜΑ 4.32: Του Monte Carlo για το BDT-14 μεταβλητές στο κανάλι των ηλεκτρονίων





ΣΧΗΜΑ 4.33: Toy Monte Carlo για το BDT-16 μεταβλητές στο κανάλι των ηλεκτρονίων

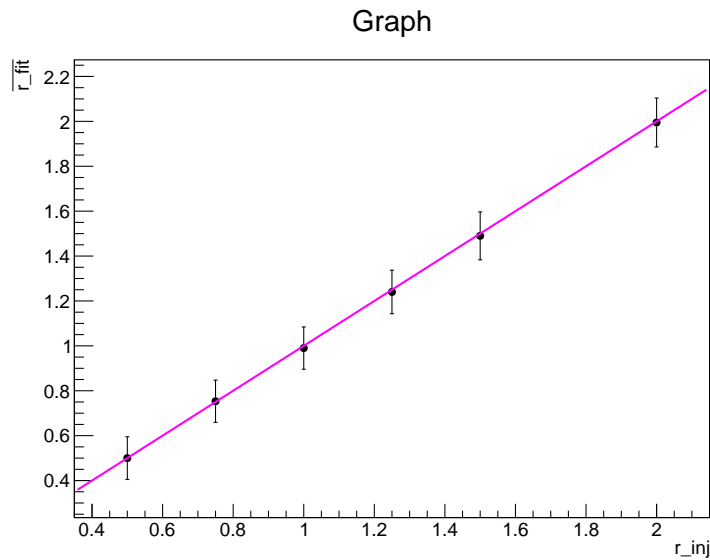
Πρώτα παρατηρείται η pull κατανομή για την παράμετρο N_{sig} (αριθμός των γεγονότων του σήματος). Για να θεωρηθεί η μέθοδος αμερόληπτη και συνεπής, πρέπει η κατανομή του pull να ακολουθεί κανονική κατανομή με μέση τιμή μηδέν και διασπορά ίση με τη μονάδα. Ο υπολογισμός του pull γίνεται μέσω της σχέσης:

$$\frac{N_{sig}^{fit} - N_{sig}^{true}}{error_N_{sig}^{fit}} \sim N(0, 1) \quad (4.13)$$

Όμοια, ο αριθμός γεγονότων του σήματος, όπως προκύπτει από κάθε fit αναμένεται να ακολουθεί κανονική κατανομή με μέση τιμή την πραγματική τιμή, δηλαδή την $N_{injected}$ που ορίζεται στην αρχή. Τέλος, για να είναι η μέθοδος συνεπής θα πρέπει το σφάλμα της εκτίμησης του N_{sig} να συγκλίνει σε μία τιμή και η διασπορά του να μειώνεται όσο αυξάνεται ο αριθμός των επαναλήψεων του toy Monte Carlo, $Var(error_N_{sig}^{fit})_{N \rightarrow \infty} = 0$.

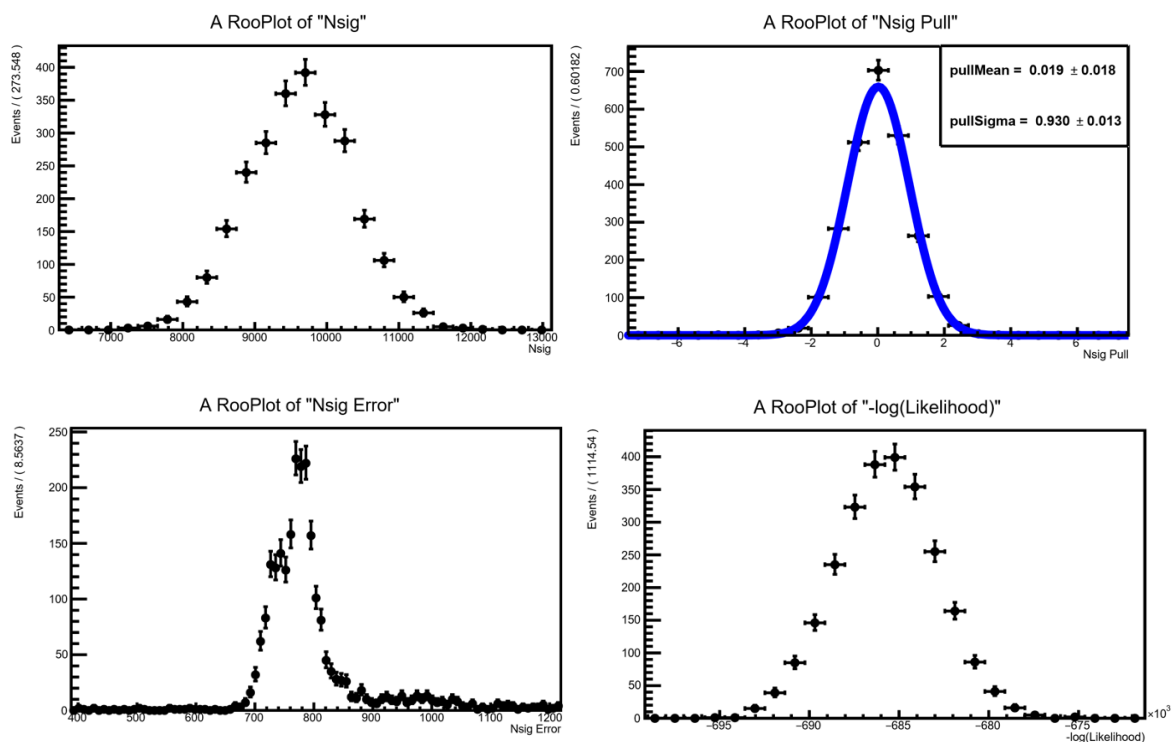
Τα δύο toys, έχουν κατανομή pull που πληρεί τα κριτήρια που αναφέρονται παραπάνω. Όμως, το σφάλμα για την παράμετρο του σήματος στην περίπτωση του BDT με τις 16 μεταβλητές, δεν κατανέμεται γύρω από μία τιμή αλλά εμφανίζει δύο κορυφές. Επομένως, το fit υλοποιήθηκε στο BDT με τις 14 μεταβλητές, του οποίου το toy MC δεν παρουσιάζει κάποια δυσμορφία.

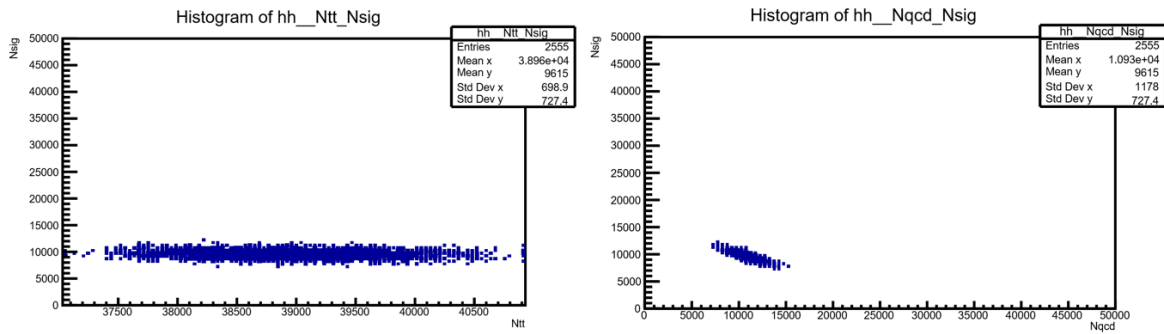
Στη συνέχεια υλοποιήθηκαν toys για το BDT με τις 14 μεταβλητές για διαφορετικές τιμές του $N_{injected}$. Διαπιστώθηκε ότι η μέση τιμή της εκτίμησης του αριθμού γεγονότων του σήματος είναι ο αρχικός αριθμός γεγονότων του σήματος που παράγεται στα ψευδοδοδεμένα ($N_{injected}$) και το σφάλμα παραμένει σχεδόν σταθερό. Ακολουθεί το διάγραμμα r_{fit} ως προς $r_{injected}$.



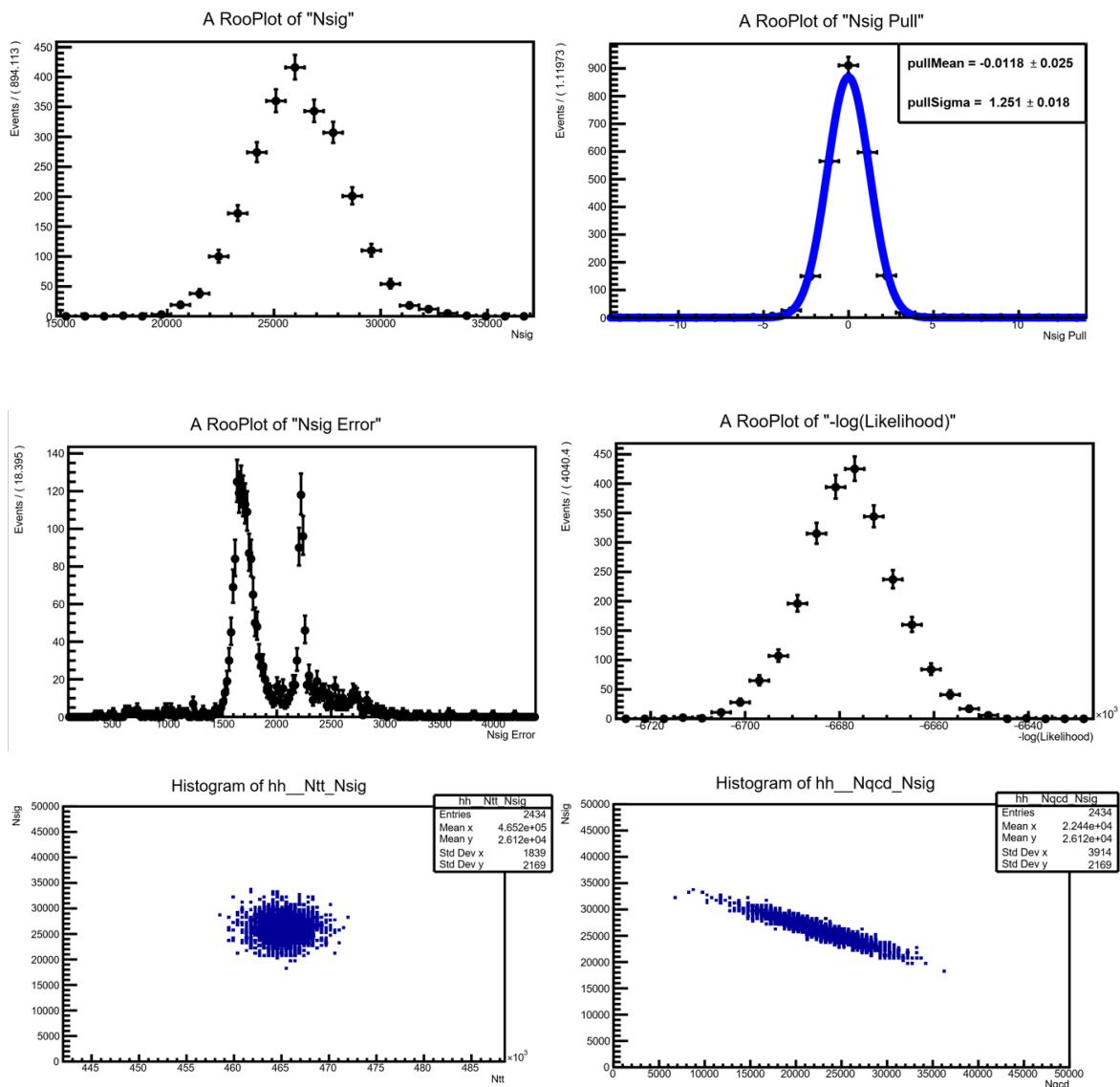
ΣΧΗΜΑ 4.34: Τιμές του r_{fit} ως προς $r_{injected}$ για toys MC του BDT-14 μεταβλητές.

Ακολουθούν τα toys MC στο κανάλι των μιονίων.





ΣΧΗΜΑ 4.35: Toy Monte Carlo για το BDT-14 μεταβλητές στο κανάλι μιονίων

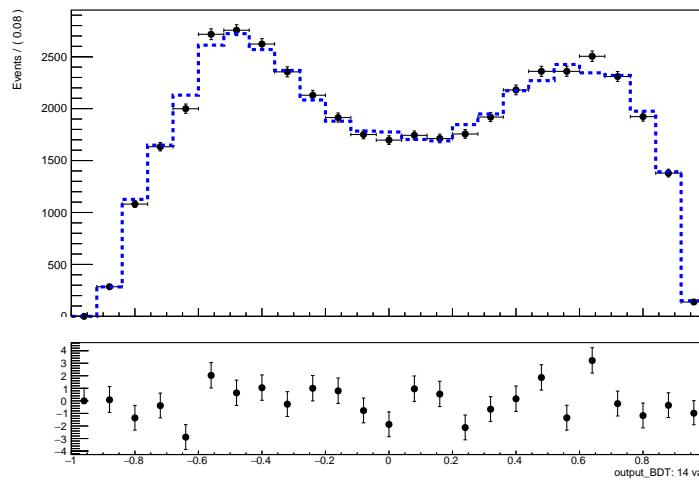


ΣΧΗΜΑ 4.36: Toy Monte Carlo για το BDT-16 μεταβλητές στο κανάλι μιονίων

Για τα $toys$ στο κανάλι των μιονίων παρατηρείται όμοια συμπεριφορά στην κατανομή του σφάλματος. Συνεπώς, επιλέχθηκε και στο μιονικό κανάλι το BDT-14 μεταβλητές ως η μέθοδος από την οποία θα εκτιμηθεί ο αριθμός των γεγονότων του σήματος.

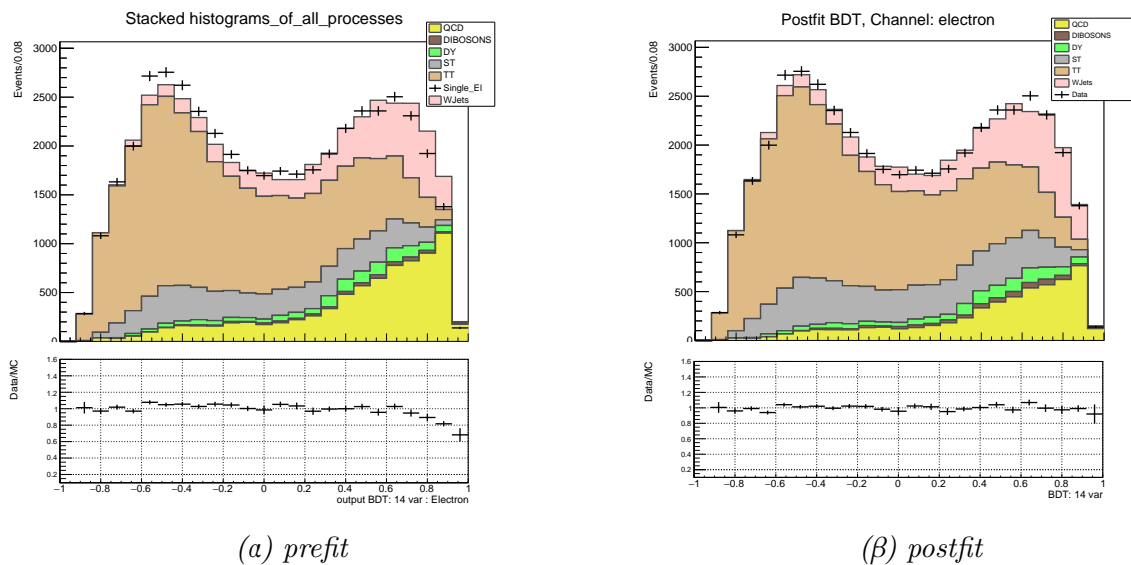
4.13.2 Αποτελέσματα του fit

Στην εικόνα (4.37) φαίνεται το fit για το BDT-14 μεταβλητές, για τα ηλεκτρόνια, και ακολουθούν τα stack διαγράμματα πριν και μετά το fit καθώς και ο πίνακας με τον αριθμό των γεγονότων.



ΣΧΗΜΑ 4.37: Fit του BDT στο κανάλι των ηλεκτρονίων

pre-fit και post-fit του BDT



(a) prefit

(b) postfit

ΣΧΗΜΑ 4.38: BDT-14 μεταβλητές prefit και postfit

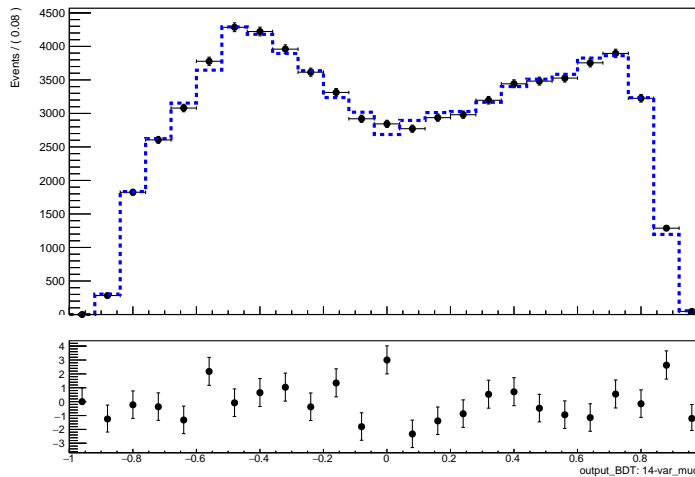
| Yields | | |
|----------------|---------|---------------------|
| | prefit | postfit |
| $W + b\bar{b}$ | 5999,7 | $6304,3 \pm 437,7$ |
| Drell-Yan | 1450,3 | $1522,8 \pm 119,2$ |
| $t\bar{t}$ | 23798,8 | $23935,2 \pm 238,7$ |
| Single top | 5742,5 | $7465,2 \pm 173,4$ |
| Diboson | 345,9 | $518,8 \pm 246,3$ |
| Qcd | 7879,9 | $5441,9 \pm 354,2$ |
| Data | 45217,0 | |

ΠΙΝΑΚΑΣ 4.3: Αριθμός γεγονότων για κάθε διεργασία πριν και μετά το fit για τα ηλεκτρόνια.

Από τον παραπάνω πίνακα προκύπτει:

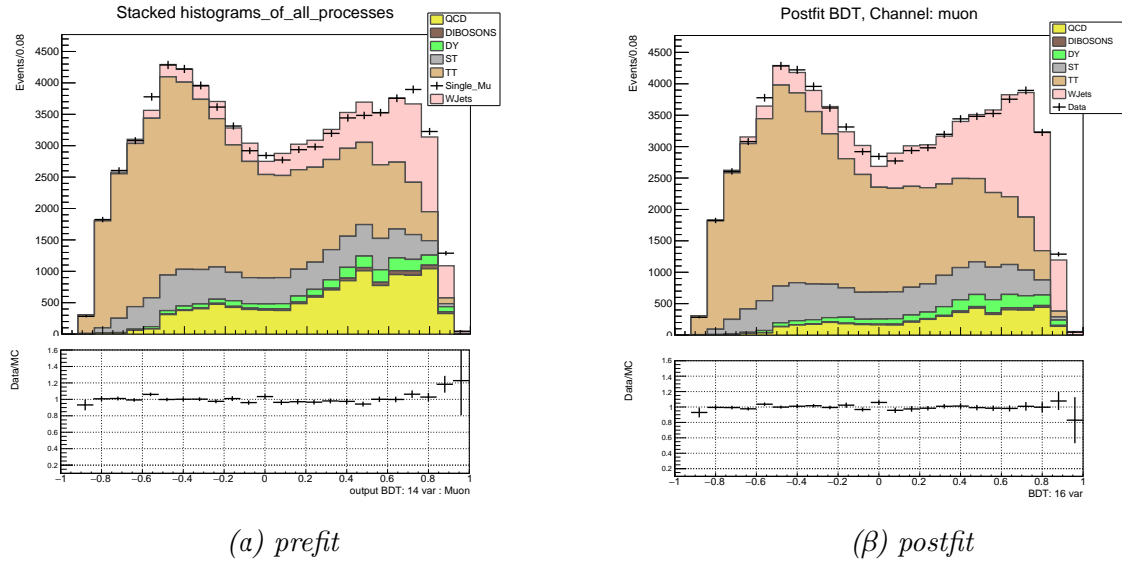
- $r_{W+b\bar{b}}^{electron} = 1,05 \pm 0,07$

Για το κανάλι των μιονίων προέκυψαν τα αποτελέσματα που ακολουθούν.



ΣΧΗΜΑ 4.39: Fit του BDT-14 μεταβλητές στο κανάλι των μιονίων

pre-fit και post-fit του BDT



ΣΧΗΜΑ 4.40: BDT-14 μεταβλητές prefit και postfit, κανάλι μιονίων

| Yields | | |
|--------------|---------|-----------------------|
| | prefit | postfit |
| $W+b\bar{b}$ | 9601,2 | $15256, 2 \pm 827, 1$ |
| Drell-Yan | 2128,8 | $2235, 2 \pm 185, 3$ |
| $t\bar{t}$ | 38981,7 | $38981, 7 \pm 712, 5$ |
| Single top | 9011,5 | $9326, 4 \pm 1180, 4$ |
| Diboson | 580,1 | $290, 0 \pm 553, 4$ |
| Qcd | 10971,8 | $4653, 2 \pm 1289, 4$ |
| Data | 71275 | |

ΠΙΝΑΚΑΣ 4.4: Αριθμός γεγονότων για κάθε διεργασία πριν και μετά το fit για τα μίονια

Από τον παραπάνω πίνακα προέκυψε:

- $r_{W+b\bar{b}}^{\mu on} = 1, 59 \pm 0, 09$

Οι δύο συντελεστές για το κανάλι των ηλεκτρονίων και το μιονίων δεν βρίσκονται σε συμφωνία. Η ενεργός διατομή υπολογίστηκε από το αποτέλεσμα στο κανάλι των ηλεκτρονίων.

4.14 Ενεργός Διατομή - Cross-Section

Η ολική ενεργός διατομή της διεργασίας $W \rightarrow e\nu + bb$ δίνεται από τη σχέση:

$$\sigma = \frac{N_{sig}}{A\epsilon\mathcal{L}} \quad (4.14)$$

όπου N_{sig} είναι η εκτίμηση του αριθμού γεγονότων του σήματος, που προέκυψε από το fit, \mathcal{L} είναι η luminosity του πειράματος και οι συντελεστές A και ϵ είναι η acceptance και η efficiency του ανιχνευτή.

Σε αυτή την ανάλυση μετρήθηκε η fiducial ενεργός διατομή:

$$\sigma = \frac{N_{sig}}{\mathcal{L}} = 0,176 \pm 0,012 pb \quad (4.15)$$

Η θεωρητική τιμή της ενεργού διατομής δίνεται αθροίζοντας για όλα τα δείγματα του σήματος, το γινόμενο του λόγου των γεγονότων που πέρασαν τα selection cuts προς τα γεγονότα που έγιναν generated σε κάθε Monte Carlo δείγμα επί την ενεργό διατομή του αντίστοιχου δείγματος.

$$\sigma_{theor} = \sum_i \frac{N_{passed}}{generated \ events \ in \ MC_i} \sigma_{MC,i} = 0,167 pb \quad (4.16)$$

4.15 Συμπεράσματα

Στην ανάλυση που υλοποιήθηκε, κατασκευάστηκε ένα BDT σε μία προσπάθεια να δημιουργηθεί ένας εκλεπτυσμένος τρόπος ανάδειξης του σήματος ($W + b\bar{b}$ διεργασία). Αφού διερευνήθηκαν κάποιες control περιοχές για το καθορισμό κάποιων διεργασιών του υποβάθρου, υλοποιήθηκε fit για τα δύο κανάλια διάσπασης του W .

Στα fit αυτά για τα δύο κανάλια δεν έχουμε συμφωνία των συντελεστών $r_{W+b\bar{b}}$. Πρέπει να αναφερθεί ότι τα μόνια υποφέρουν ως προς τη μοντελοποίηση της qcd. Η διπλωματική αυτή δεν καταπιάστηκε με την επίλυση αυτού του προβλήματος, αλλά αφήνεται προς μελλοντική μελέτη. Για αυτό η μέτρηση πραγματοποιήθηκε μόνο στο κανάλι των ηλεκτρονίων.

Η πειραματική τιμή που μετρήθηκε είναι σε ικανοποιητική συμφωνία με τη θεωρητική τιμή, μέσα στα όρια των στατιστικών σφαλμάτων.

Βιβλιογραφία

- [1] Ian JR Aitchison and Anthony JG Hey. *Gauge Theories in Particle Physics: A Practical Introduction: From Relativistic Quantum Mechanics to QED*. Vol. 1. CRC Press, 2012.
- [2] David Griffiths. *Introduction to elementary particles*. John Wiley & Sons, 2008.
- [3] ΒΑΓΙΟΝΑΚΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ Ε. *ΣΩΜΑΤΙΔΙΑΚΗ ΦΥΣΙΚΗ Μια Εισαγωγή στη Βασική Δομή της Ύλης*. 2008.
- [4] *The Large Hadron Collider*. <https://home.cern/science/accelerators/large-hadron-collider>. Accessed: 2020-08-02.
- [5] *How an accelerator works*. <https://home.cern/science/accelerators/how-accelerator-works>. Accessed: 2020-08-02.
- [6] CMS Collaboration et al. *The CMS experiment at the CERN LHC*. 2008.
- [7] CMS collaboration et al. “Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s}=7$ TeV”. In: *arXiv preprint arXiv:1107.4789* (2011).
- [8] Γεώργιος Μπάκας. “Μελέτη Διαφορικής Ενεργού Διατομής της παραγωγής ζεύγους top-antitop quark στον ανιχνευτή CMS”. In: (2018).
- [9] CMS collaboration et al. “Measurement of the production cross section for a W boson and two b jets in pp collisions at $\sqrt{s}=8$ TeV”. In: *Available on CMS information server* (2014).
- [10] CMS collaboration et al. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}=8$ TeV”. In: *arXiv preprint arXiv:1502.02701* (2015).
- [11] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti-kt jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (2008), p. 063.
- [12] *Machine Learning*. <https://www.mathworks.com/discovery/machine-learning.html>. Accessed: 2020-05-30.
- [13] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition, Edition*. 2009.
- [14] Byron P Roe et al. “Boosted decision trees as an alternative to artificial neural networks for particle identification”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 543.2-3 (2005), pp. 577–584.
- [15] Andreas Hoecker et al. “TMVA-toolkit for multivariate data analysis”. In: *arXiv preprint physics/0703039* (2007).
- [16] *Introduction to Monte Carlo for Particle Physics Study*. <https://indico.cern.ch/event/92209/contributions/2114409/attachments/1098701/1567290/CST2010-MC.pdf>. Accessed: 2020-08-24.

-
- [17] CMS collaboration et al. “Study of W boson production in PbPb and pp collisions at $\sqrt{s} = 2.76$ TeV”. In: *arXiv preprint arXiv:1205.6334* (2012).
- [18] Kirti Ranjan, Ashutosh Bhardwaj, et al. “Single-Top Quark Production at CMS”. In: *Universe* 5.1 (2019), p. 19.
- [19] CMS collaboration, CMS Collaboration, et al. *Pileup jet identification*. 2013.
- [20] Lidija Radovanovic. “Data-driven QCD background estimation to W production in association with jets using the ATLAS detector”. PhD thesis. University of Cape Town, 2016.
- [21] W Verkerke and D Kirkby. *RooFit users manual v2. 91*. 2011.
- [22] Luca Lista. *Statistical methods for data analysis in particle physics*. Vol. 941. Springer, 2017.