NATIONAL TECHNICAL UNIVERSITY OF ATHENS

UNDERGRADUATE THESIS

Search for a new topophyllic leptophobic Z'_{TC2} boson in the fully hadronic ttbar final state using the CMS detector

---

Author:
Lucy KOTSIOPOULOU

Supervisor:
Dr. Konstantinos KOUSOURIS

Three Member Committee:
Dr. Konstantinos KOUSOURIS
Dr. Georgios TSIPOLITIS
Dr. Georgia KARAPOSTOLI

A thesis submitted in fulfilment of the requirements
for the degree of Integrated Master

in the

Laboratory of Experimental High Energy Physics & Related Technology-Instrumentation

Department of Physics

School of Applied Mathematics and Physical Sciences

Athens, Greece, July 24th, 2023

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΠΡΟΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ

Αναζήτηση νέου top-φιλικού λεπτοφοβικού μποζονίου Z'_{TC2} στην πλήρως αδρονική τελική κατάσταση ζεύγους ttbar με χρήση του ανιχνευτή CMS

Συγγραφεύς:
Λουκία ΚΩΤΣΙΟΠΟΥΛΟΥ

Επιβλέπων:
Δρ. Κωνσταντίνος ΚΟΥΣΟΥΡΗΣ

Τριμελής Επιτροπή:
Δρ. Κωνσταντίνος ΚΟΥΣΟΥΡΗΣ
Δρ. Γεώργιος ΤΣΙΠΟΛΙΤΗΣ
Δρ. Γεωργία ΚΑΡΑΠΟΣΤΟΛΗ

Διπλωματική εργασία που υποβλήθηκε για την εκπλήρωση των απαιτήσεων για το πτυχίο του Integrated Master

στο

Εργαστήριο Πειραματικής Φυσικής Υψηλών Ενεργειών & Σχετικής Τεχνολογίας Οργάνων
Τομέας Φυσικής
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

Αθήνα, Ελλάδα, 24 Ιουλίου, 2023

Search for a new topophyllic leptophobic Z'_{TC2} boson in the fully hadronic ttbar final state using the CMS detector

Lucy Kotsiopoulou

Abstract

This thesis focuses on the experimental search for a new Z' boson decaying to $t\bar{t}$ pair resonances, arising from a Topcolour Assisted Technicolour (TC2) model (Z'_{TC2}) using 2016 data taken by the CMS experiment at the LHC at CERN. TC2 aims to provide an underlying mechanism for ElectroWeak Symmetry Breaking (EWSB) assuming that new interactions couple preferentially to the third generation and especially to the top quark. In the theoretical component, I first review the currently standing Standard Model (SM) of Particle Physics and the CMS experiment, and subsequently present the theoretical elements of TC2 and cross section formulae. The experimental component comprises of an overview of the jet reconstruction techniques used at CMS, such as the anti-k_t algorithm and N-subjettiness, and a small synopsis of the data analysis framework used, ROOT. The final chapter focuses on my own work: analysis of Monte Carlo simulations and data for the fully hadronic $t\bar{t}$ final state. A cut-based Top Tagger is created from zero, to distinguish the reducible and irreducible background signals of SM produced $t\bar{t}$ and QCD respectively. After the insertion of the data, a data driven method is used to fully calculate the QCD background contamination, after which fitting methods are applied to the data, and the log-likelihood function is graphed. Finally, observed Bayesian upper limits for the signal yield are calculated. No significant deviation from the SM background is observed.

Key words: *Z prime, Z' boson, Z'_{TC2} boson, Topcolor Assisted Technicolor, Topcolour Assisted Technicolour, Electroweak symmetry breaking, top quark mass, electroweak scale, ttbar resonances, data-driven method for QCD, cut-based top tagger, subjettiness, anti-k_t algorithm, Bayesian upper limits, CMS Collaboration, CERN.*

Athens, 2023
National Technical University of Athens
School of Applied Mathematics and Physical Sciences

Αναζήτηση νέου top-φιλικού λεπτοφοβικού μποζονίου Z'_{TC2} στην πλήρως αδρονική τελική κατάσταση ζεύγους ttbar με χρήση του ανιχνευτή CMS

Λουκία Κωτσιοπούλου

Περίληψη

Η διπλωματική αυτή εργασία επικεντρώνεται στην αναζήτηση ενός νέου μποζονίου Z′ που διασπάται σε ζεύγος συντονισμών $t\bar{t}$, από το μοντέλο της θεωρίας Topcolour Assisted Technicolour (TC2), χρησιμοποιώντας δεδομένα του 2016 από το πείραμα CMS στο CERN. Η TC2 επιχειρεί να εξηγήσει τον βαθύτερο μηχανισμό της Θραύσης της Ηλεκτρασθενούς Συμμετρίας θεωρώντας πως μία νέα ισχυρή δυναμική έχει προνομιακή ζεύξη στην Τρίτη Γενεά, και ιδιαίτερα στο top quark. Στο θεωρητικό μέρος, πρώτα κάνω μία σύνοψη της τωρινής θεωρίας του Καθιερωμένου Προτύπου των Στοιχειωδών Σωματιδίων (SM) και το πείραμα CMS, και μετέπειτα την αναλυτική θεωρία της TC2 και τους υπολογισμούς ενεργών διατομών. Το πειραματικό μέρος αποτελείται από μία επισκόπηση τεχνικών ανακατασκευής πιδάκων, όπως τον αλγόριθμο anti-k_t και την N-subjettiness, και μία μικρή εισαγωγή στο πρόγραμμα ανάλυσης που χρησιμοποιήθηκε, ROOT. Η τελευταία ενότητα πραγματεύεται την ατομική δουλειά μου: κατασκευάζεται ένας Top Tagger για την ανίχνευση $t\bar{t}$ στην πλήρως αδρονική τελική κατάσταση, για να διαχωριστούν τα δύο υπόβαθρα του $t\bar{t}$ και του QCD SM διαδικασιών. Μετά την προσθήκη και των δεδομένων, μία μέθοδος υποβοηθούμενη-από-δεδομένα υλοποιείται για τον υπολογισμό της μόλυνσης από υπόβαθρο QCD. Εφαρμόζεται στη συνέχεια η διαδικασία προσαρμογής στα δεδομένα, και σχεδιάζεται η συνάρτηση log-likelihood. Τελικά, υπολογίζονται τα άνω όρια Bayes για την παρατηρούμενη ποσότητα σήματος. Δεν παρατηρείται σημαντική απόκλιση του σήματος από το υπόβαθρο του SM.

Λέξεις κλειδιά: *Z τονούμενο, μποζόνιο Z′, μποζόνιο Z'_{TC2}, Topcolor Assisted Technicolor, Topcolour Assisted Technicolour, Θραύση ηλεκτρασθενούς συμμετρίας, μάζα τοπ κουάρκ, ηλεκτρασθενής κλίμακα, συντονισμοί ttbar, data-driven QCD μέθοδος, cut-βασισμένος top tagger, subjettiness, anti-k_t αλγόριθμος, άνω όρια Bayes, Συνεργασία CMS, CERN.*

Αθήνα, 2023

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστημών

There once was a Z′ at CERN
About which we’re going to learn
It’s an interesting particle
The definite article!
For which all the physicists yearn

So, a fermion or boson?
Check its spin to know for sure
An integer or just a half?
But both decay, they don't endure

So you'll have to count them quick
To clear up scientists’ confusion
Assess Z′ and ttbar
Sketch the graphs and draw conclusions

GG Carey, 2023.

Acknowledgments

In this section I would like to give my heartfelt thanks to the people who, over the years and in the last few months, helped me and were part of making this thesis a reality in its final form.

I shall start with my supervisor, Dr. Kousouris, whom I have known for five years now, as he was one of my first physics professors at university, in Physics I (Mechanics). I remember me pressing him to tell me about particle physics, and him replying to me, “Don’t get ahead of yourself”, and, well, look at where we are now. So thank you for our cooperation during this past year during my thesis, for explaining the experimental methods and processes in such a plain and theoretical way, and for all the jokes about my Hallowe’en and goth obsession.

Apart from my supervisor, I would like to thank the rest of the CMS working group at NTUA, for listening to my meeting presentations every Thursday and for their apposite comments, but especially the following colleagues: Anna Zacharopoulou (PhD candidate), for her useful insights on my final result and presentation the day before my thesis defence; Nausika Kalogeropoulou (MSc graduate), for pressing me and convincing me to go ask my questions of my supervisor; Rena Siamarkou (PhD candidate) for finding mistakes in my programming at all hours, coming up to the office just for me to solve my confusion and for so lovingly going through my whole presentation with me before the defence; Thodoris Chatzistavrou (PhD candidate), for using the little time he had to help and explain physics to me when I was confused and depressed, for standing up for me and for being a wonderful companion and friend.

I would like to thank my professors, both at NTUA, and at my pre-university schools for teaching me all that I needed (and some that I didn’t) need to know about sciences, and the arts, and for giving me such wonderful memories of my childhood and university life.

Next, I would like to thank my friends, physicists and not, who at the last minute managed to come to my thesis defence in person and online and support me; it meant the world to me! Especially I would like to mention and thank my online friend at CERN, Nikos, for pressing me to carry on, believing in me, and convincing me that I could achieve what I wanted all on my own. See you in neutrinos!

Finally, my personal thanks will go to my parents D(M) and B(D), for checking and correcting my thesis and presentation, as they have been doing for the last 23 years for all my projects. Specifically my mother, for looking at the pretty pictures, and my father for getting me into physics and mathematics since I was 2 seconds old (maybe 3).

My last and most important thanks though will go to my very best friends: Andreas, Evi and Marc. They were my Formula 1 support crew, and they dropped everything to come and clean and cook and mop for me, while I was busy studying. I couldn’t have done it without you, my Italian badass, Fifi, mi diablito.

Ευχαριστίες

Σε αυτή την ενότητα θα ήθελα να ευχαριστήσω από καρδιάς τους ανθρώπους που όλα αυτά τα χρόνια και τους τελευταίους μήνες με βοήθησαν και συμμετείχαν στην πραγματοποίηση αυτής της διπλωματικής εργασίας στην τελική της μορφή.

Θα ξεκινήσω με τον επιβλέποντα μου, τον κ. Κουσουρή, τον οποίο γνωρίζω εδώ και πέντε χρόνια, καθώς ήταν ένας από τους πρώτους μου καθηγητές φυσικής στο πανεπιστήμιο, στη Φυσική Ι (Μηχανική). Θυμάμαι να τον πιέζω να μου πει για σωματιδιακή φυσική, και εκείνος να μου απαντά, «Μην προτρέχεις», και, τελικά κοιτάξτε πού βρισκόμαστε τώρα. Σας ευχαριστώ λοιπόν για τη συνεργασία μας κατά τη διάρκεια της περασμένης χρονιάς στην διπλωματική μου, για την εξήγηση των πειραματικών μεθόδων και διαδικασιών με τόσο απλό και θεωρητικό τρόπο, και για όλα τα αστεία σχετικά με την λατρεία μου για το Hallowe'en και τα γκοθικά πράγματα.

Εκτός από τον επιβλέποντά μου, θα ήθελα να ευχαριστήσω την υπόλοιπη ομάδα του CMS στο ΕΜΠ, που άκουγαν τις παρουσιάσεις μου κάθε Πέμπτη με διορθωτικά σχόλια, αλλά και ιδιαίτερα τους ακόλουθους συναδέλφους: Άννα Ζαχαροπούλου (υποψήφια διδάκτωρ), για τις διορθώσεις της σχετικά με το τελικό μου αποτέλεσμα και την πρόβα μου την ημέρα πριν την παρουσίασή μου· Ναυσικά Καλογεροπούλου (απόφοιτο MSc), που με πίεζε και με έπεισε να πηγαίνω να ρωτάω στον κ. Κουσουρή· Ρένα Σιαμάρκου (υποψήφια διδάκτωρ) που έβρισκε λάθη στον προγραμματισμό μου σε κάθε στιγμή της ημέρας, που ερχόταν στο γραφείο μόνο και μόνο για να λύσω τις απορίες μου και που με τόση αγάπη κοίταξε όλη την παρουσίασή μαζί μου· Θοδωρή Χατζησταύρου (υποψήφιο διδάκτωρα), που χρησιμοποιούσε τον λίγο χρόνο που είχε για να με βοηθάει και να μου εξηγεί επαναλαμβανόμενα φυσική όταν ήμουν αγχωμένη, που στάθηκε στο πλευρό μου και που ήταν ένας υπέροχος φίλος.

Θα ήθελα επίσης να ευχαριστήσω τους καθηγητές μου, τόσο στο ΕΜΠ όσο και στα προ-πανεπιστημιακά μου σχολεία που μου έμαθαν όλα όσα χρειαζόμουν (και μερικά που δεν χρειαζόμουν) για τις επιστήμες, και τις τέχνες, και που μου χάρισαν τόσο υπέροχες αναμνήσεις από την παιδική μου ηλικία και την πανεπιστημιακή ζωή.

Στη συνέχεια, θα ήθελα να ευχαριστήσω τους φίλους μου, φυσικούς και μη, που την τελευταία στιγμή κατάφεραν να έρθουν στην υπεράσπιση της διπλωματικής μου, αυτοπροσώπως και διαδικτυακώς, και με υποστήριξαν. Σήμαινε τα πάντα για μένα! Ιδιαίτερα θα ήθελα να αναφέρω και να ευχαριστήσω τον διαδικτυακό μου φίλο στο CERN, Νίκο, που με πίεσε να συνεχίσω, που πίστεψε σε μένα και που με έπεισε πως μπορώ να πετύχω ότι ήθελα. Τα λέμε στα νετρίνα!

Τέλος, οι προσωπικές μου ευχαριστίες θα πάνε στους γονείς μου D(M) και B(D), για τον έλεγχο και τη διόρθωση του κειμένου και της παρουσίασής μου, όπως έχουν κάνει τα τελευταία 23 χρόνια για όλες τις εργασίες μου. Συγκεκριμένα, τη μητέρα μου, που κοίταγε απλά τις "όμορφες φωτογραφίες", και τον πατέρα μου που με μύησε στην φυσική και τα μαθηματικά από τότε που ήμουν 2 δευτερολέπτων (ίσως 3).

Ωστόσο, το τελευταίο και σημαντικότερο ευχαριστώ θα πάει στους καλύτερους μου φίλους: τον Ανδρέα, την Εύη και τον Marc. Ήταν το πλήρωμά μου υποστήριξης της Formula 1: άφησαν τα πάντα για να έρθουν να καθαρίσουν, να μαγειρέψουν και να σφουγγαρίσουν το σπίτι, ενώ εγώ εργαζόμουν για την διπλωματική. Δεν θα μπορούσα να τα καταφέρω χωρίς εσάς, my Italian badass, Fifi, mi diablito.

Dedicated to Evi, Andreas
and the kaimaki I promised Marc

Αφιερωμένο στην Εύη, τον Αντρέα,
και το καϊμάκι που υποσχέθηκα στον Marc

Dedicat a Evi, Andreas
i al kaimaki que he promès a Marc

TABLE OF CONTENTS

LIST OF FIGURES

Index of Tables

LIST OF ABBREVIATIONS

ΚΠ	**Κ**αθιερωμένο **Π**ρότυπο
ΦΣΣ	**Φ**υσική **Σ**τοιχειωδών **Σ**ωματιδίων
ΦΥΕ	**Φ**υσική **Υ**ψηλών **Ε**νεργειών
ATLAS	**A** **T**oroidal **L**HC **A**pparatu**s** (LHC Collaboration)
BSM	**B**eyond the **S**tandard **M**odel
BUL	**B**ayesian **U**pper **L**imit
CERN	**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire
CM	**C**entre of **M**ass
CMS	**C**ompact **M**uon **S**olenoid
CR	**C**ontrol **R**egion
ETC	**E**xtended **T**echnicolour
EWSB	**E**lectro**w**eak **S**ymmetry **B**reaking
GSW	**G**lashow, **S**alam and **W**einberg electroweak theory
HEP	**H**igh **E**nergy **P**hysics
LHC	**L**arge **H**adron **C**ollider
MC	**M**onte **C**arlo
NSD	**N**ovel **S**trong **D**ynamics
NLL	**N**egative **L**og-**l**ikelihood
PDF	**P**robability **D**ensity **F**unction
PP	**P**article **P**hysics
QCD	**Q**uantum **C**hromo**d**ynamics
QED	**Q**uantum **E**lectro**d**ynamics
QFT	**Q**uantum **F**ield **T**heory
QM	**Q**uantum **M**echanics
SM	**S**tandard **M**odel
SR	**S**ignal **R**egion
SSB	**S**pontaneous **S**ymmetry **B**reaking
TC	**T**echni**c**olour
TC2	**T**opcolour **assisted** techni**c**olour
VEV	**V**acuum **E**xpectation **V**alue

ΕΚΤΕΤΑΜΕΝΗ ΠΕΡΙΛΗΨΗ ΣΤΗΝ ΕΛΛΗΝΙΚΗ

Η Φυσική Στοιχειωδών Σωματιδίων (ΦΣΣ) ή αλλιώς η Φυσική Υψηλών Ενεργειών (ΦΥΕ) είναι η υποκατηγορία των Φυσικών Επιστημών που επιχειρεί να μελετήσει τα πιο στοιχειώδη αδιαίρετα συστατικά της ύλης, αλλά και να ανακαλύψει νέα. Η πιο πλήρης θεωρία της ΦΣΣ είναι αυτή του Καθιερωμένου Προτύπου (ΚΠ), η οποία ομαδοποιεί τα σωματίδια της ύλης, τα φερμιόνια, και τα σωματίδια-φορείς των δυνάμεων, τα μποζόνια, σε μία ολοκληρωμένη και συνεκτική θεωρία (δείτε εικόνα 2). Το ΚΠ περιγράφει τρεις από τις τέσσερις θεμελιώδεις δυνάμεις που θεωρούνται πως υπάρχουν στην φύση: την Ισχυρή Αλληλεπίδραση, η οποία "κρατά" ενωμένα τα quarks μέσα στα πρωτόνια και τα νετρόνια των ατόμων· την Ηλεκτρομαγνητική, η οποία είναι εμφανής στην καθημερινότητα σε μαγνήτες και καλώδια ρεύματος· την Ασθενή, η οποία ευθύνεται για την πυρηνική διάσπαση-β και τις διεργασίες στους αστέρες. Οι δύο τελευταίες έχουν πια ενοποιηθεί σε μία με βάση την *Ηλεκτρασθενή Θεωρία*, ενώ την τέταρτη αποτελεί η Βαρύτητα, η οποία κυβερνά τον κόσμο μεγάλης κλίμακας.

Τα σωματίδια χωρίζονται στις δύο παραπάνω αναφερόμενες κατηγορίες, τα φερμιόνια και τα μποζόνια, τα σωματίδια ύλης και τα σωματίδια-φορείς. Τα φερμιόνια χωρίζονται περαιτέρω στα *quarks* και στα *λεπτόνια*· τα *up* (**u**) και *down* (**d**) quarks και το ηλεκτρόνιο (**e**) απαρτίζουν τα άτομα, από τα οποία αποτελείται όλη η καθημερινή ύλη. Η κάθε δύναμη "μεταφέρεται" από το αντίστοιχο μποζόνιο: το φωτόνιο (**γ**) για την Ηλεκτρομαγνητική δύναμη, το γκλουόνιο (**g**) για την Ισχυρή και τα μποζόνια $\mathbf{W^+}$, $\mathbf{W^-}$ και **Z** για την Ασθενή. Το πιο πρόσφατο νέο σωματίδιο που ανακαλύφθηκε και που συμπληρώνει την πλήρη εικόνα του ΚΠ είναι το λεγόμενο *μποζόνιο Higgs*, το οποίο μέσω του *μηχανισμού Higgs*, παρέχει τον τρόπο με τον οποίο τα υπόλοιπα σωματίδια αποκτούν μάζα.[i]

Το ΚΠ, εν συντομία, περιγράφεται από την πλήρη (τοπική) ομάδα $\mathbf{SU(3)_C \times SU(2)_L \times U_Y(1)}$, που περιλαμβάνει την ομάδα χρώματος των quarks $\mathbf{SU(3)_C}$ (**C**: colour), και τις ομάδες $\mathbf{SU(2)_L \times U_Y(1)}$ που σχετίζονται με την Ηλεκτρασθενή αλληλεπίδραση. Το σωματίδιο Higgs *σπάει* την Ηλεκτρασθενή συμμετρία στην *ασθενή κλίμακα* ενέργειας, δίνοντας έτσι μάζα στα φερμιόνια, και στα ηλεκτρασθενή μποζόνια, $\mathbf{W^\pm}$ και $\mathbf{Z^0}$:

$$SU(3)_C \times SU(2)_L \times U_Y(1) \rightarrow SU(3)_C \times U_{em}(1)$$

Με αυτόν τον τρόπο *διάσπασης*, τα σωματίδια των εναπομεινάντων ομάδων του ηλεκτρομαγνητισμού και του χρώματος παραμένουν άμαζα, το φωτόνιο και το γκλουόνιο, ενώ τα $\mathbf{W^\pm}$ και **Z** αποκτούν μάζα, όταν η ομάδα *σπάει*, στην *ηλεκτρασθενή κλίμακα*, περίπου O(100) GeV/c^2.[ii]

---

i Thomson 2013; Mee και Manton 2017.

ii Mavromatos 2021.

Για την πειραματική ανίχνευση και μελέτη των σωματιδίων στην ΦΣΣ, απαιτούνται υψηλές ενέργειες για την διερεύνηση των μικρότερων συστατικών της ύλης. Ένα από τα σχετικά πειράματα, που βρίσκεται στο Ευρωπαϊκό Συμβούλιο για την Πυρηνική Έρευνα (Conseil Européen pour la Recherche Nucléaire – CERN) στην Γενεύη της Ελβετίας, είναι ο Μεγάλος Επιταχυντής Αδρονίων (Large Hadron Collider – LHC). Έχει περιφέρεια 27χλμ και επιταχύνει σε τεράστιες ενέργειες και ορμές πρωτόνια σε αντίθετες κατευθύνσεις, ώσπου τελικά συγκρούονται σε τέσσερα διαφορετικά σημεία κατά μήκος του Επιταχυντή (δείτε εικόνα 3). Σε κάθε σημείο, βρίσκεται ένα *πείραμα*, σε συνεργασία δε με ένα από τα αυτά, λεγόμενο CMS (Compact Muon Solenoid – Συμπαγές Σωληνοειδές Μιονίου), πραγματοποιήθηκε αυτή η διπλωματική. Συνοπτικά, ο τρόπος με τον οποίο λειτουργεί, χαρακτηρίζεται από την *κρεμμυδωτή μορφή* του: εάν η σύγκρουση των πρωτονίων θεωρείται πως γίνεται στην μέση του ανιχνευτή, κυλινδρικά γύρω έχουν τοποθετηθεί στρώσεις διαφορετικού είδους ανιχνευτών, που μετρούν την θέση του σωματιδίου, την ορμή, την ενέργεια και άλλες ιδιότητες. Δείτε τις εικόνες 4, 5 και 6 για μερικά σχεδιαγράμματα.[i]

Παρ'όλο που το ΚΠ παρέχει μία εξαιρετικά απλή και πλήρη περιγραφή του μικρόκοσμου, αρκετά φαινόμενα είναι ακόμη ανεξήγητα και υπάρχει ενεργή αναζήτηση για την "λύση" τους στην επιστημονική κοινότητα. Η παρούσα εργασία εξετάζει τρεις δυνατότητες για νέα φυσική και επομένως τρία κίνητρα για την ύπαρξη του αναζητούμενου σωματιδίου Z', το οποίο είναι και το κεντρικό αντικείμενο της. Το πρώτο "μυστήριο", για το οποίο ακόμη δεν υπάρχει ικανοποιητική εξήγηση, αποτελεί η προέλευση της λεγόμενης *ηλεκτρασθενούς* ή *ασθενούς κλίμακας* (*weak* or *electroweak* scale), μερικών O(100) GeV, στην οποία παρατηρείται Αυθόρμητη Θραύση της Συμμετρίας (Spontaneous Symmetry Breaking – SSB) που προκαλείται από το Higgs μεταξύ της Ηλεκτρομαγνητικής και της Ασθενούς δύναμης, επίσης αποκαλούμενη Ηλεκτρασθενής Διάσπαση (ElectroWeak Symmetry Breaking – EWSB). Ενώ είναι γνωστό, ύστερα από την ανακάλυψη του Higgs στο CERN το 2012, πως το σωματίδιο Higgs προκαλεί την EWSB, δεν είναι γνωστό ποιος είναι ούτε ποιός είναι ο βαθύτερος μηχανισμός και η κύρια δυναμική που την διέπει, ούτε ο λόγος που συμβαίνει στην συγκεκριμένη *ηλεκτρασθενή* κλίμακα. Τέλος, ένα από τα κύρια θέματα έρευνας στο CERN, είναι ο λόγος για τον οποίο το *top quark* της Τρίτης Γενεάς, είναι τόσο βαρύτερο από τα υπόλοιπα. Συγκριτικά, το αμέσως επόμενο quark, το bottom (**b**) quark έχει μάζα 4.3 GeV/c^2 (40 φορές μικρότερο), ενώ το ελαφρύτερο στο ΚΠ, το up quark (**u**), έχει μάζα 0.003 GeV/c^2 (10000 φορές μικρότερο) (εικόνα 1).[ii]

---

i Martin και Shaw 2017; Thomson 2013.

ii Hill and Simmons, 2003.

Μία από τις θεωρητικές προτάσεις, για την εξήγηση του φαινομένου της EWSB, αλλά και που προβλέπει το top quark με μεγάλη μάζα, είναι η θεωρία του *Technicolour Υποβοηθούμενη από Topcolour* (Topcolour-Assisted Technicolour – TC2), που αρχικά προτάθηκε από τον θεωρητικό φυσικό στο Fermilab, Christopher T. Hill το 1994 (δείτε εικόνα 7). Η θεωρία βασίζεται στις θεωρίες του Technicolour και Topcolour, αλλά η καινοτομία της έγκειται στο γεγονός ότι θεωρεί ότι το μποζόνιο Higgs αποτελείται από ένα συμπύκνωμα $\langle t\bar{t} \rangle$ ($t\bar{t}$ condensate), και ότι ενσωματώνει Νέα Ισχυρή Δυναμική (Novel Strong Dynamics – NSD) για την εξήγηση της EWSB. Θεωρεί εν ολίγοις ότι το βαρύ top quark είναι το κλειδί για την εξήγηση της EWSB, αφού η μεγάλη του μάζα υποδηλώνει πως έχει μεγάλη σύζευξη με το πεδίο Higgs. Ως αποτέλεσμα των NSD που εισάγονται, προβλέπεται η ύπαρξη ενός πειραματικά ανιχνεύσιμου αποτελέσματος, ενός νέου είδους μποζονίου Z'_{TC2} (στην συνέχεια συμβολιζόμενο με αυτόν τον τρόπο). Ανάμεσα στην πληθώρα των μοντέλων που πρότεινε ο Hill, με άλλους συναδέλφους του, για την εμφάνιση ενός τέτοιου σωματιδίου στην θεωρία είναι το "Μοντέλο IV". Αποτελεί το πιο δημοφιλές σε σύγχρονες πειραματικές αναζητήσεις και διαθέτει και την μεγαλύτερη ενεργό διατομή. Η ίδια αναζήτηση θα διεξαχθεί και στην παρούσα εργασία.[i]

Το Z'_{TC2} θεωρητικά διασπάται μετά την παραγωγή του σε ένα ζεύγος top-antitop quark ($t\bar{t}$) (εικόνα 8), το οποίο, λόγω της αδρανοποίησης, εμφανίζεται ως *πίδακες* (*jets*) στον ανιχνευτή (εικόνα 9). Η αδρανοποίηση (hadronisation) είναι η διαδικασία από την οποία προκύπτουν πίδακες και όχι τροχιές quarks. Ως αποτέλεσμα των αλληλεπιδράσεων της QCD, η ενέργεια του πεδίου του Ισχυρής Αλληλεπίδρασης μεταξύ των δύο quarks (σε απόσταση 10^{-15} m) μετατρέπεται σε περαιτέρω ζεύγη quarks και antiquarks, τα οποία τελικά θα συγκροτήσουν τους πίδακες. Εν ολίγοις: 'συγκροτήματα πολλών quarks από αλληλεπιδράσεις $q\bar{q}$. Το t ($\bar{t}$) διασπάται εν συνεχεία κατά 99.8% σε W^+ (W^-) μποζόνιο και b ($\bar{b}$) quark, και συνεπώς οι αναζητήσεις για ζευγάρια top quark $t\bar{t}$ επικεντρώνονται στην αναζήτηση των προϊόντων αυτών των δύο σωματιδίων. Στην περίπτωση της παραγωγής ζευγαριού $t\bar{t}$, όπως προβλέπεται για το Z'_{TC2}, οι κατηγορίες των τελικών καταστάσεων θα είναι οι εξής: πλήρως αδρονική (fully hadronic) με μόνο quarks, λεπτονική+πίδακες (lepton+jets) με ένα W από το ζευγάρι $t\bar{t}$ να διασπάται σε quarks και το άλλο λεπτονικά, και διλεπτονική (dilepton) με λεπτονικές διασπάσεις και των δύο W (εικόνες 11 και 12). Η παρούσα εργασία θα επικεντρώνεται και θα εξετάζει μόνο την περίπτωση *Πλήρους Αδρονικής* τελικής κατάστασης του συστήματος $t\bar{t}$.[ii]

---

i Harris, Hill καικαι Parke 1999.

ii Lannon, Margaroli και Neu (2012), σελ. 1-3.

Εφόσον η αναζήτηση του Z'_{TC2} συνεπάγεται την αναζήτηση ζευγαριού $t\bar{t}$, η συγγραφεύς ξεκινά την ανάλυσή της με την δημιουργία ενός αλγορίθμου για την εύρεση top quarks (*Top Tagging*) και την διαφοροποίησή τους από το ένα είδους υποβάθρου της ανάλυσης, δηλαδή το *αναγώγιμο* (*reducible*) των διεργασιών QCD. Ξεκινώντας με αρχεία προσομοιώσεων σχετικών γεγονότων Monte Carlo, εφαρμόζονται συνθήκες που αντιστοιχούν σε ιδιότητες του **t** για την ταυτοποίηση του: συνθήκες σχετικά με την μάζα του πίδακα του **t** (που αντικατοπτρίζεται στην μεταβλητή jetMass) (εικόνα 40), την εσωτερική γεωμετρία του πίδακα (συγκεκριμένα πόσοι *υποπίδακες* εμπεριέχονται, που αντικατοπτρίζεται στην μεταβλητή της *Subjettiness* τ_3/τ_1) (εικόνα 39) και τα περιεχόμενα σωματίδια μέσα στον πίδακα (εικόνα 47). Το πείραμα CMS παρέχει την έτοιμη πληροφορία του πόσα **b** quarks εμπεριέχονται μέσα σε έναν πίδακα, η οποία πληροφορία θα χρησιμοποιηθεί για την τρίτη συνθήκη. Επιλέγονται τα ακόλουθα δύο *σημεία εργασίας* (*working points*), τα λεγόμενα *χαλαρά* (*loose*) και *αυστηρά* (*tight*):

$$Loose\ working\ point: \quad Mass\ window\ cut\ A[140,250] + 53\,\%\ Top\ Tagging\ Efficiency$$

$$Tight\ working\ point: \quad Mass\ window\ cut\ A[140,250] + 30\,\%\ Top\ Tagging\ Efficiency + b\ cuts$$

Μετά την κατασκευή των διαγραμμάτων *Αποδοτικότητας* (εικόνες 48 και 49), επιλέγεται η πρώτη συνθήκη ως η καλύτερη. Σε αυτό το τελικό σημείο, γίνεται σύγκριση των παραπάνω συνθηκών και με προσομοιώσεις του αναζητούμενου σωματιδίου Z'_{TC2}, για το οποίο στην συνέχεια θα θεωρηθούν τέσσερις διαφορετικές περιπτώσεις μάζας (εικόνα 54): 1400GeV/c^2, 2000GeV/c^2, 2500GeV/c^2 και 3500GeV/c^2.

Στην ανάλυση προστίθεται και το αρχείο των δεδομένων, και παρατηρείται το ιστόγραμμα της αναλλοίωτης μάζας των δύο πιδάκων για τα έξι συμπεριλαμβανόμενα αρχεία: το υπόβαθρο του $t\bar{t}$, τα τέσσερα σήματα του Z' και τα δεδομένα. Όλα τα ιστογράμματα παρουσιάζονται με νέα κλίμακα με βάση τα χαρακτηριστικά του ανιχνευτή και την ενεργό διατομή της κάθε διεργασίας, ώστε να αναπαρίσταται ο πραγματικά αναμενόμενος αριθμός γεγονότων. Παρατηρώντας τα διαγράμματα (εικόνα 56), οι *χαλαρές* συνθήκες εμφανίζουν σημαντική ποσότητα από το υπόβαθρο της QCD, και γι'αυτό εφαρμόζεται μία μέθοδος *βασισμένη σε δεδομένα* (*data-driven method*) για το υπολογισμό της ποσότητας QCD με χρήση των *αυστηρών* συνθηκών (εικόνα 52).

Έχοντας υπολογίσει το υπόβαθρο και την QCD, εφαρμόζεται μία μέθοδος *προσαρμογής* (*fitting* method) στην οποία οι ελεύθερες παράμετροι είναι: η ποσότητα υποβάθρου QCD **N_QCD**, η ποσότητα υποβάθρου $t\bar{t}$ **N_tt** και η ποσότητα σήματος **N_signal_i**, όπου **i** αναφέρεται στην κάθε περίπτωση μάζας. Υπολογίζεται η συνάρτηση *log-likelihood* που αναζητά τις ελάχιστες τιμές των παραμέτρων αυτών κατά τη διαδικασία προσαρμογής. Αυτή σχεδιάζεται σε συνάρτηση με την τιμή

του **N_signal** για κάθε περίπτωση μάζας (εικόνες 66 και 65). Η ελαχιστοποίηση προκύπτει για αρνητικές τιμές του **N_signal**, με αποτέλεσμα *να μην υπάρχουν ενδείξεις ανακάλυψης του Z' σωματιδίου*, καθώς δεν παρατηρείται κάποια σημαντική απόκλιση του σήματος από το υπόβαθρο στο φάσμα συντονισμού μάζας $t\bar{t}$. Τέλος, υπολογίζονται τα *παρατηρούμενα άνω όρια Bayes*, τα οποία και σχεδιάζονται σε συνάρτηση με την μάζα, ανάλογα της ενεργούς διατομής (εικόνα 67). Η ενεργός διατομή μειώνεται όσο αυξάνεται η μάζα, όπως αναμένεται. Επί του παρόντος, δεν μπορούν να γίνουν συγκρίσεις με πειραματικές τιμές δημοσιεύσεων (όπως τα πιο πρόσφατα από το CMS και ATLAS[i]), καθώς δεν υπολογίστηκαν τα *θεωρητικά άνω όρια* στην παρούσα εργασία.

Η συγγραφέας προτείνει την συνέχεια της αναζήτησης και για τα υπόλοιπα τρία μοντέλα, για τα οποία, απ'όσο γνωρίζει μέχρι στιγμής, δεν έχουν διεξαχθεί αναζητήσεις στο LHC.

i The CMS Collaboration 2019; The ATLAS Collaboration 2021.

I. INTRODUCTION

The field of Particle Physics (PP) or High Energy Physics (HEP) aims to study the properties of the smallest constituents of matter, and possibly discover new ones. One could say that the very beginning of the field was the attempt to describe the substance of light in Ancient Greece, by "early scientists" such as Aristotle and Euclid, while the atomists first spoke of the smallest constituent of matter being the atom (from the Greek: atomo/άτομο, meaning indivisible).[i] Following Newton's corpuscular theory of light and Einstein's Nobel Prize for the photoelectric phenomenon, the conclusion finally becomes the following: all massive matter and non-massive matter (such as the photons of light) behave equally as point-like place-determined particles AND as flowing infinitely-extending waves, depending on the circumstances.

It could be considered that PP officially began as a field during one experiment and one theory in 1897 and 1900, namely the discovery of the first subatomic particle, the electron by J.J. Thomson; and the theory by A. Einstein, that light is made up of particles, called *photons*. For the first three decades of the 20th century, PP and its sister field Nuclear Physics, the study of the nucleus of atoms, were to all intents and purposes the same fields. PP then branched off into the study of the interior of the atom, and of the nucleus, until the very elementary particles that cannot be broken up any further.[ii]

Physics as a science generally works in two areas: theory and experiment. An article in the Fermilab Annual Report, “Theorists and Experimentalists: Partners in the Search” explains the phenomenon very accurately:

> One difference between physics and other sciences is that physics has a more or less strict division into experimental and theoretical research. The main reason for this is that the laws of physics depend on mathematics in a more basic way than other sciences. Because the talents and training needed to carry out measurements in the lab are different from those needed to carry out mathematical calculations, physicists tend to take one path or the other at the start of their careers [iii]

This thesis is split into two parts, the **Theoretical Component**, which will lay the foundations of the modern understanding of PP and dive into new physics theories, and the **Experimental Component**, which will focus on experimental techniques used at PP experiments to analyse already known physics and to discover new physics.

---

i Στέφος, Στεργιούλης και Χαριτίδου 2012, σελ. 130.
ii Still 2017, p. 10.
iii “Theorists and Experimentalists: Partners in the Search” 1993, p. 28.

THEORETICAL COMPONENT

The modern picture of PP comprises of a very well tested theory, the Standard Model of Particle Physics (SM), and of the theories aiming to extend the SM and answer mysteries within the field, the theories Beyond the Standard Model (BSM). In the first chapter, I will review the SM, and the fundamental particles and forces that it describes. Following the theoretical foundations of modern PP, I will give an overview of the largest particle physics experimental complex built to date, the Large Hadron Collider (LHC) at the European Organisation for Nuclear Research (CERN - Conseil Européen pour la Recherche Nucléaire) which is located on the border between France and Switzerland. Within CERN there are many collaborations, which each comprise of an *experiment*, located along the circular collider of the LHC. This thesis was done in collaboration with the Compact Muon Solenoid Experiment (CMS Experiment), and so it will be the focus of my review.

One of these elementary particles, known as the top quark, is especially interesting because of its very large mass. As a comparison, the second heaviest particle in the quark category is 40 times lighter (the bottom quark), and the lightest (the up quark) 10000 times lighter than the top quark[i]. This makes it one of the most popular study subjects at particle physics experiments, but it has also become an essential component for developing new BSM theories. Another interesting mystery is describing the fundamental mechanism of *ElectroWeak Symmetry Breaking* (EWSB), and why it happens at the so-called *weak scale*. Since the 1950s the question has arisen of how the fundamental particles acquire mass: in 2012, CERN successfully announced the discovery of the Higgs particle, which completes the SM "particle zoo". The Higgs, in short, gives mass to the rest of the massive particles in SM, by breaking the ElectroWeak symmetry between the quarks and weak bosons at low energies. However, the fundamental mechanism of this EWSB is not known; nor what the Higgs particle actually is. As the top quark, the most massive particle of the SM, has a huge coupling to

Source: Thomson 2013, p. 3.

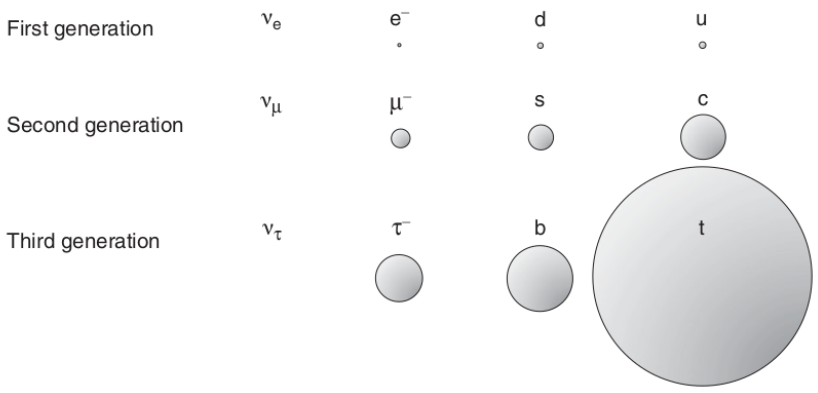


Figure 1: The particles in the three generations of fundamental fermions with the masses indicated by imagined spherical volumes of constant density. In reality, fundamental particles are believed to be point-like.

i Particle Data Group 2000.

the Higgs field, it is hopeful that the further study of it will give fruitful explanations for the origins of EWSB.

Next, I will naturally cover some of the suggested theories to explain EWSB. Specifically, I will focus on the *Topcolour Assisted Technicolour* (TC2) model[i], first suggested by the theoretical physicist Christopher T. Hill at Fermilab in 1994. It is based on the (Extended) Technicolour theory, started in the 1970s, and the Topcolour theory. This specific model provides an explanation of EWSB suggesting that the Higgs boson is actually a condensate of a top and antitop quark, and manages to both explain a possible mechanism for EWSB by also predicting a viable large mass for the top quark. Lastly, to explain these two phenomena, it is needed to insert new types of forces, *Novel Strong Dynamics* (NSD), which in turn provide an experimentally testable consequence: the emergence of a new type of gauge boson, the Z'_{TC2} (**Z** prime, from a TC2 theory), which is theoretically posed to decay to a top-antitop pair.

Following the presentation and analysis of the TC2 model that predicts the Z'_{TC2} boson, I will present the results of calculations of the cross section of the new boson, by researchers at Fermilab and other institutions. The cross section of a particle is basically the property needed to be able to detect it in particle physics experiments, which is what makes it a defining parameter in PP.

Finally I will analyse the various experimental signatures of the Z'_{TC2}; in the case of this thesis, it decays to a top-antitop pair, which can be experimentally detected and isolated as an interesting event, possibly originating from the Z'_{TC2}. The top quark has two ways of decaying, so in the case of a top-antitop pair ($t\bar{t}$), the situation is more complicated, as in the detector one will see three different possible combinations of final particle states: fully hadronic, lepton+jets (or known as semileptonic) and leptonic. This thesis will focus on the first type of decay, which will depend on the detection of two distinct *hadronic jets*, i.e. two energy depositions in the detector, from each top and antitop quark.[ii]

i Hill 1994; Harris, Hill and Parke 1999.
ii Lannon, Margaroli and Neu 2012.

EXPERIMENTAL COMPONENT

In the first chapter on the experimental component, I will first review some special detection methods for hadronic jets within the detector, as is the case for the top-antitop pair. I will analyse the concepts of *Subjettinness* and the a*nti-k_T algorithm* for jet clustering[i], which will be used in the main part of the analysis. Lastly, I will provide an overview of the data analysis framework used, ROOT, and some statistical methods that will be needed later in the text.

The main analysis chapter, and fully my own work, is chapter 5. Here, I start by working with Monte Carlo samples of Z'_{TC2} and $t\bar{t}$ processes, simulations of how the processes should look like when observed in the detector. I will use the methods described above to develop an in-house Top Tagger using simple *cuts* (conditions) on the simulation files, to finally decide on the optimal parameters to reduce the $t\bar{t}$ background process and highlight the Z'_{TC2} signal. Afterwards, I implement a *reversed cut process*, using a *data-driven method*, to quantify the amount of the second background, the QCD background. Finally the fitting process is implemented, and the log-likelihood functions are calculated, to acquire the finishing results, which are the observed Bayesian upper limits of the amount of the signal process.

The last chapter is reserved for my personal conclusions. After shortly mentioning the most recent particle searches of this type, I will comment on the future prospects of discovering such a particle and suggest new possibilities for discovery.[ii]

i Cacciari, Salam and Soyez 2008; Thaler and Tilburg, 2011.

ii The CMS Collaboration 2019; The ATLAS Collaboration 2021.

THEORETICAL COMPONENT

II. REVIEW OF THE STANDARD MODEL AND THE CMS DETECTOR AT CERN

II.i. Review of the Standard Model of Particle Physics

The best and most accurate describing model that PP has to date is the so called Standard Model (SM). It was developed in the 1960s and 70s, and has been verified experimentally numerous times over the years, needing just a few minor changes. However it is incomplete, as there are still unanswered questions concerning the microcosm, such as the origin of dark matter and dark energy[i]. In this thesis, and in the next chapter, the unknown mechanism of ElectroWeak Symmetry Breaking, the origin of the "weak" scale, and the true nature of the Higgs boson will be addressed.

Wikimedia Commons: Miss J, as in Hobbs, 2017

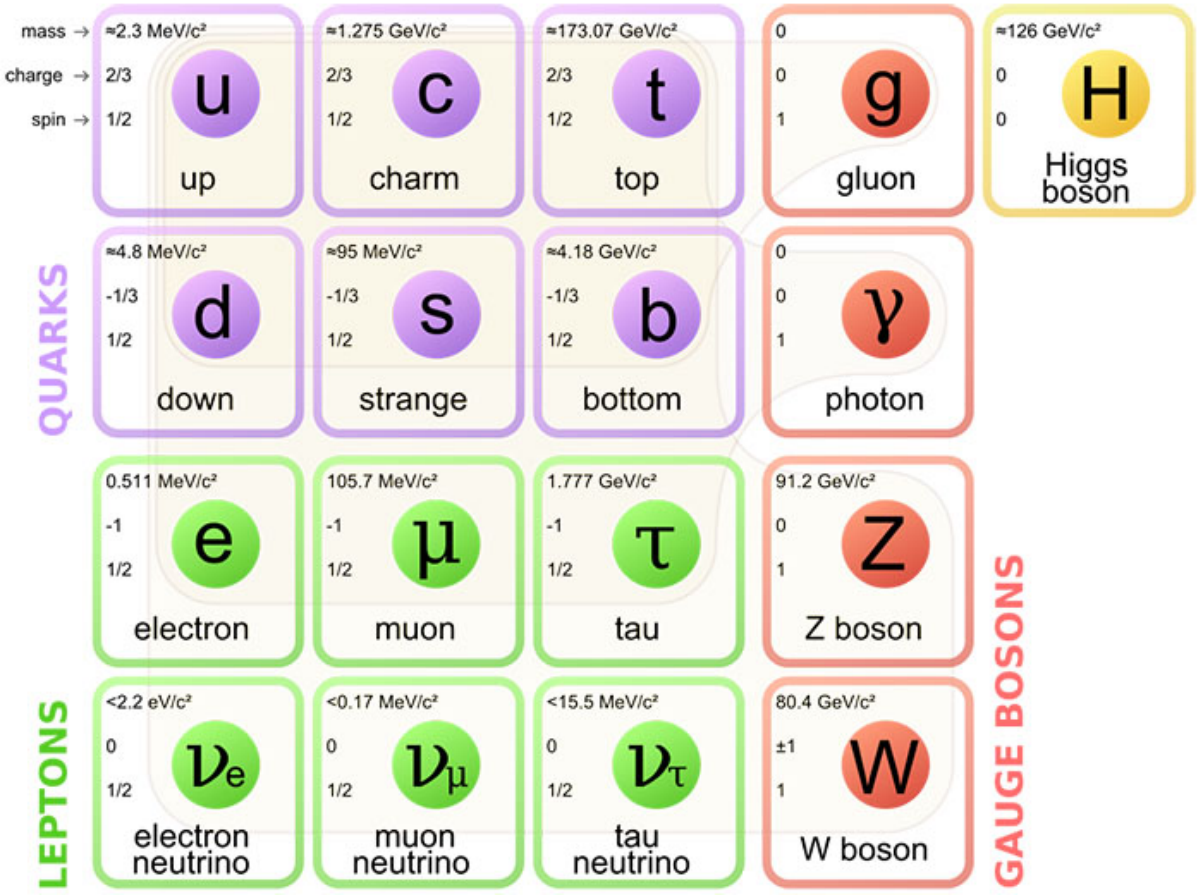


Figure 2: All the particles described in the Standard Model of Particle Physics

i Still 2017, p. 8

The SM uses *fundamental particles*, and *fundamental forces*, to describe the fundamental workings of nature. In this framework, the forces are also portrayed and described as the exchange of other different particles, the *force mediators*.[i] Beginning with the forces, it is considered that there are four fundamental forces in nature[ii]:

- Gravity, which acts on massive objects and mostly determines the large scale universe
- Electromagnetic force, which acts on *charged* objects and is responsible for holding different atoms together
- Strong force, which "glues" together the protons and neutrons inside the atom
- Weak force, which governs beta decay and plays an important role in Star synthesis

If one assumes the Strong force to have intensity 1, then it is followed by the Electromagnetic force at 10^{-2} times weaker, by the Weak force at 10^{-5} times weaker, and finally joined by the weakest force, gravity at 10^{-39} [iii]. The SM succeeds in describing the unified force of three of the above, the united ElectroWeak (Electromagnetism and Weak) force, and the Strong force and its interactions. The only force omitted by the SM is gravity, and incorporating it is one of the most challenging prospects of modern PP.

Providing an overview of all the particles and their workings, the "periodic table" of PP is provided in Figure 2. Each particle in this table is defined by three numbers: its rest mass, spin and charges that govern its interactions. There are of course other properties that they may have, such as magnetic moment and decay rates, but these are not as fundamental. The particles are split into two categories: integer spin *bosons*, so named because they obey Bose-Einstein statistics; and half-integer spin *fermions*, that obey Fermi-Dirac statistics. These *statistics*, are derived from the theory of *Quantum Mechanics* (QM), and in short describe the constraints on particles, and in what *state* they are allowed to be in.[iv] All bosons and fermions also have *antiparticles*, which are complete replicas of the standard particles, with the same mass, but opposite charge, spin or other properties. They constitute what is known as *anti-matter*.

In the SM, there are twelve fundamental fermions, all of half integer spin equal to ½. They also are then divided into two categories, the quarks and leptons, according to whether they feel the strong force; any particle with an electromagnetic charge will feel the Electromagnetic force, and so one is left with the following: the quarks "feel" and can interact via the Strong, Weak and Electromagnetic force, while as the leptons interact via the Electromagnetic and Weak force. It is

i Thomson 2013, p. 1.
ii Mee and Manton 2017, p. 393.
iii Mavromatos 2021, p. 45.
iv Mandl 2013, p. 245; Mee and Manton 2017, p. 394.

noted that the lepton category also contains three types of neutrinos, which are neutrally charged, so they only interact via the Weak force. Apart from this categorisation into quarks and leptons, they are also categorised into three groups, or *generations*, of four particles each.

- The first generation: *up* ($\mathbf{u}$) and *down* ($\mathbf{d}$) quarks, *electron* ($\mathbf{e}$) and *electron neutrino* ($\boldsymbol{\nu_e}$)
- The second: *charm* ($\mathbf{c}$) and *strange* ($\mathbf{s}$) quarks, *muon* ($\boldsymbol{\mu}$) and *muon neutrino* ($\boldsymbol{\nu_\mu}$)
- The third: *top* ($\mathbf{t}$) and *bottom* ($\mathbf{b}$) quarks, *tauon* ($\boldsymbol{\tau}$) and *tau neutrino* ($\boldsymbol{\nu_\tau}$).

As can be seen from figure Figure 2, each generation is a *heavier* "copy" of the first, carrying the same spin and charge values for the respective particles. Particles are symbolised with letters, for example the muon $\boldsymbol{\mu}$, and their antiparticles with the same letter but with a bar on top, $\boldsymbol{\bar{\mu}}$.[i]

Three particles in the first generation are those which make up all the tangible matter in the cosmos. The two particles that make up the nucleus of the atom, the proton and the neutron, consist of solely up and down quarks[ii, iii], while as the electron is that which "orbits" around the nucleus, forming a "probability cloud", which completes the modern picture of the atom. The remainder of the quarks in the other two generations also make up other matter particles, the Baryons, such as the *lambda* ($\boldsymbol{\Lambda}$) made up of $\mathbf{uds}$, and the Mesons, such as the kaon ($\mathbf{K^-}$), made up of $\mathbf{s\bar{u}}$. The leptons do not appear as constituents of any particle, and appear as free particles arising from various physical processes.

In the beginning of the chapter, it was mentioned that the SM comprises of matter particles, discussed above, and of three fundamental forces which are also described by force mediating particles, namely the *gauge bosons*. As opposed to the classical approach of the theory of interactions, in PP *Quantum Field Theory* (QFT) does not treat forces as scalar potentials that act on matter from a distance, but as actual force carrying particles. Each of the three forces in nature is described by a QFT, which describes the exchange of one of the spin-1 gauge bosons. The *photon* ($\boldsymbol{\gamma}$), is the carrier of the Electromagnetic force, and its equivalent QFT is called *Quantum ElectroDynamics* (QED). Similarly, the carrier of the strong force is the *gluon* ($\mathbf{g}$), massless like the photon, while for the weak force there are two massive particles, the charged $\mathbf{W^\pm}$ (particle and anti-particle) and the neutral $\mathbf{Z^0}$. The first governs the weak charged-current interaction, which is responsible for nuclear β-decay and nuclear fusion, and the second the neutral-current interaction for other weak processes. The four fundamental bosons, as well as some of their properties, are depicted in Figure 2.

---

i Mee and Manton 2017, p. 394.

ii The proton is made up of two up quarks and one down quark bound together ($\mathbf{uud}$), while as the neutron consists of two down quarks and one up ($\mathbf{udd}$).

iii Mee and Manton 2017, p. 394.

The fundamental bosons are completed by the Higgs boson, which had been theoretically posited for more than 20 years, yet was only discovered as recently as in 2012 by the particle accelerator at CERN. It is the only spin-0 (*scalar*) particle discovered to date, and constitutes the final piece of the puzzle to complete the SM. It provides the mechanism by which all the other massive particles acquire mass, called ElectroWeak Symmetry Breaking (EWSB). In the following chapters an extensive overview of the Higgs mechanism will be provided.[i]

In the language of *mathematics*, the SM is described in the terms of *group theory* and the *Lagrangian density* (from now on known solely as *Lagrangian*) expressing the interactions between the forces and the particles. Group theory in the particle physics context describes the underlying symmetries within the theory in a neat manner; the Lagrangian contains the terms of these symmetries and describes the relationship between them. As mentioned above, the SM is described by Quantum Field Theories (QFT), and more specifically by their subcategory, *gauge theories*[ii]. In the following paragraphs, various parts of the Lagrangian will be sequentially be presented, along with their group theory related counterparts, so as to finally arrive at the final Lagrangian of the SM. As the purpose of this review is not the fundamentals of group theory nor of the theoretical implications of PP, a general description of the formulae to follow will suffice to complete this small review of basic concepts.

Beginning with the general picture, the complete (local) group that describes the SM is the following[iii]:

$$SU(3)_C \times SU(2)_L \times U_Y(1) \quad \text{(A)}$$

The $\mathbf{SU(3)_C}$ group expresses the *colour* symmetry between the quarks, $\mathbf{SU(2)_L}$ is related to the *isospin* parameter of the electroweak force and lastly $\mathbf{U_Y(1)}$ represents the weak hypercharge, $\mathbf{Y}$.

Starting with the last two groups, the way in which the Electromagnetic and Weak forces can be described by a coherent *electroweak* force is presented. This theory is called the Electroweak Theory, developed by Glashow, Salam and Weinberg (GSW), and it incorporates the leptons and their interactions, including how they acquire mass via the Higgs boson. The part of the Lagrangian that describes the lepton sector is the following:

i Thomson 2013, p. 6.
ii Mavromatos 2021, p. 3
iii The following review is all based on lecture notes by Mavromatos 2021, p. 78, 82, 86-8.

$$L_{lepton} = -\frac{1}{4} F^{a}_{W\,\mu\nu} \frac{1}{4} F^{a\,\mu\nu}_{W} - \frac{1}{4} F^{a}_{B\,\mu\nu} \frac{1}{4} F^{a\,\mu\nu}_{B} + \sum_{f=e,\mu,\tau} (\bar{l}_f i\gamma^{\rho} D_{\rho} l_f + \bar{f}_R i\gamma^{\mu} D^{Y}_{\mu} f_R + g_f \bar{f}_R \varphi^{\dagger} l_f + h.c.) +$$
$$+ (D_{\mu}\varphi)^{\dagger} (D^{\mu}\varphi) + \mu^2 \varphi^{\dagger}\varphi - \lambda(\varphi^{\dagger}\varphi)^2$$

(B)

where g_f is the strength of the weak force (Fermi's constant), γ^{μ} αre the Dirac matrices, λ is a parameter constant and **h.c.** is the hermitian conjugate.

The first terms of the form $F^{a}_{W\,\mu\nu}$ and $F_{B\,\mu\nu}$, respectively, are the field strength tensors describing the weak force bosons W^{μ} and the weak hypercharge[i] bosons B^{μ}. They are given by:

$$F^{a}_{W\,\mu\nu} = \partial_{\mu} W^{a}_{\nu} - \partial_{\nu} W^{a}_{\mu} + i\varepsilon^{abc} W^{b}_{\mu} W^{c}_{\nu}$$
$$F_{B\mu\nu} = \partial_{\mu} B_{\nu} - \partial_{\nu} B_{\mu}$$

(C)

where ϑ_{μ} is the relativistic partial derivative, i is the unitary imaginary number, and ε^{abc} is the Levi-Chivita symbol. The summation in L_{lepton} sums over the three types of leptons (lepton doublets), and contains the interactions between them and the hypercharge. The final term describes the interactions of the Higgs field φ (doublet); this term "breaks" the symmetry of the $SU(2)_L$ group, and as a consequence the $W^{\pm}$ and Z^0 become massive.

After providing a description of the electroweak interactions, now the final group, $SU(3)_C$ will describe the strong interaction. The strong force is mediated by the massless gluons, and is primarily "felt" by the quarks. The quarks, analogous to the charge of the electromagnetic field, have a *colour charge*, red, blue or green. They can also have the "opposite" colour charges, which explains why the theory of *Quantum Chromodynamics* (QCD) describes them in *colour triplets*, symbolised as ψ_f. The QCD Lagrangian is the following:

$$L_{QCD} = -\frac{1}{4} F^{a}_{\mu\nu} F^{a\mu\nu} + \sum_{f=flavour} \bar{\psi}_f (i\gamma^{\mu} D_{\mu} - m_f)\psi_f \quad \text{(D)}$$

where α spans 1,… 8, $D_{\mu} = \partial_{\mu} + i g_s \frac{\lambda^{\alpha}}{2} A^{2}_{\mu}$, g_s is the coupling strength of the strong force, m_f the mass of quark f and f spans the flavour of the quarks.

The first term now is describes the field strengths of the gluons A^{α}_{μ} and is given by:

$$F^{a}_{\mu\nu} = \partial_{\mu} A^{\alpha}_{\nu} - \partial_{\nu} A^{\alpha}_{\mu} + g_s f^{\alpha\beta\gamma} A^{\beta}_{\mu} A^{\gamma}_{\nu} \quad \text{(E)}$$

where $f^{\alpha\beta\gamma}$ are group theory form factors.

The second summation term describes the quarks and their interactions through their symmetry group $SU(3)_C$.

---

i Weak hypercharge is a quantum number that relates the *charge* number (of the electromagnetic forces) with the weak force quantum number, *isospin*.

It should be noted that, although principally it seemed like the electroweak interaction and the strong interaction were two different theories, they indeed are very alike because of *lepton-quark* symmetry. It can be seen as if the SM treats these two different types of particles as symmetrical, with many similar properties.

This completes the full picture of the interactions that the SM describes, including the final part, the Higgs boson. Putting an emphasis on the way this particle provides mass, as it will be a key topic later on, it is explicitly presented that the Higgs terms break the following symmetry:

$$SU(3)_C \times SU(2)_L \times U_Y(1) \rightarrow SU(3)_C \times U_{em}(1) \quad \text{(F)}$$

This way, as the "remaining" groups after the breakage are that of electromagnetism and the colour group, the photon and the gluon will remain massless. However, the three weak force bosons will acquire mass as the $\mathbf{SU(2)_L}$ will be *spontaneously* broken by the EWSB mechanism.

II.ii. The CMS detector at CERN

In order to obtain experimental results in particle physics, high energies are required to "probe" either the interior of known particles, or to "create" new phenomena unseen until now in the laboratory. For this reason, HEP experiments are very often of the form of particle colliders, which accelerate charged particles that collide, producing new particles as products that can then be studied.

Source: Thomson 2017, p. 84 .

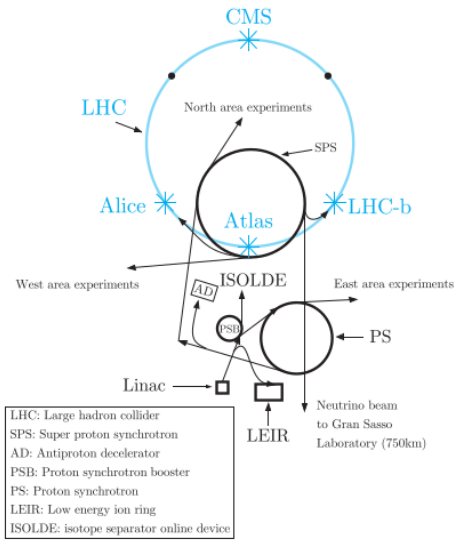


Figure 3: A schematic diagram of the CERN site showing the LHC and some of its other accelerators.

This thesis was done in cooperation with the CMS Experiment (Compact Muon Solenoid Experiment), one of the experimental complexes situated around the Large Hadron Collider (LHC). The LHC is a 27km circumference proton-proton (hadrons) particle collider at the European Organisation for Nuclear Research (CERN - Conseil Européen pour la Recherche Nucléaire). After the protons go through an initial acceleration process, they are directed towards the main circular collider, the LHC, where after executing multiple rounds and acquiring more and more energy and momentum they finally collide at four points, where the experiments are situated (see Figure 3).

Some interesting variables regarding accelerator experiments are the following: the cross section **σ**, the centre of mass energy $\mathbf{E_{CM}}$ and the luminosity **L**. Beginning with the second parameter, in accordance with the theory of Special Relativity, in the Centre of Mass frame (CM) of the colliding constituents, all the energy available from them become available for particle production. Assuming that the crossing angle of these particles is zero, and that the colliding particles are of the same mass (as in the **pp** case of the LHC), the CM energy $\mathbf{E_{CM}}$ will be:

$$E_{CM} = 2E_L \quad \text{(G)}$$

where $\mathbf{E_L}$ is the projectile along the beam energy of each particle. The current CM energies at the LHC have reached up to 13TeV.[i] The CM energy is the parameter that determines the type of particles that can be created and discovered.[ii]

The next quantity, the *cross section*, is related to how probable it is quantum mechanically for a certain interaction. The luminosity is highly related, as it expresses the rate at which events happen within the collider. Their relationship is the following:

$$N = \sigma \int L(t)\,dt \quad \text{(H)}$$

This equation describes that for a given process, the number of interactions **N** that will happen, will be the product of the cross section **σ** (the reaction probability) and the integrated luminosity **L** integrated over the lifetime **t** of the operation of the machine in question.[iii]

Regarding the CMS detector, it is one of the experiments along the beamline at the LHC, and similarly with the other experiments, it uses an "onion-type" formation of detectors (see Figure 4 and Figure 5). The protons, after being accelerated in the LHC tunnel, collide in the central vertex of the detector, causing the particles produced to move in an "outwards" direction, perpendicular to the original beamline axis. For this reason, the "onion" formation includes

Source: The CMS Collaboration et al. 2008, p. 3.

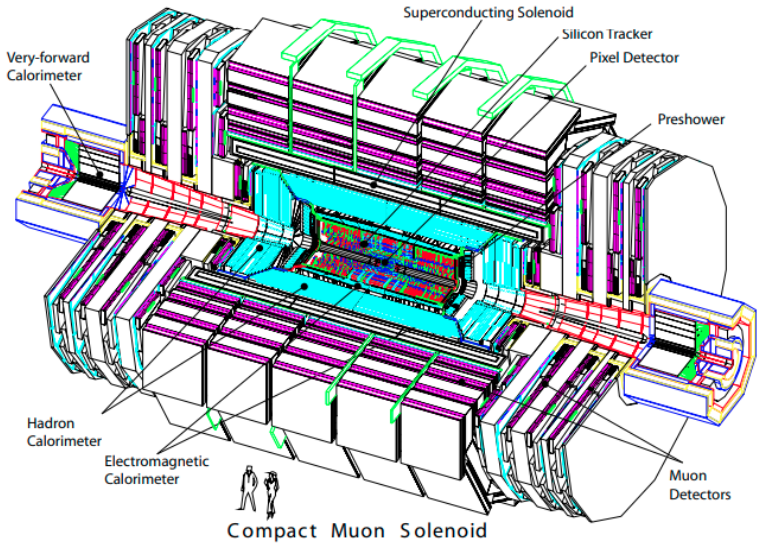


Figure 4: A perspective view of the CMS detector

i Martin and Shaw 2017, p. 78.
ii Thomson 2013, p. 26.
iii Thomson 2013, p. 26

putting different types of detectors around the production vertex; as the particles pass through the different detector layers, different particle properties will be measured.

By looking at Figure 5, one can see some particle tracks of different types, and how they pass through the detector. Briefly, the first tracker at CMS is a silicon tracker, which measures the particles' tracks, while the two calorimeters measure the energy deposition of the particles. One of the main and defining features of the CMS detector is its superconducting solenoid, of 6m internal diameter, providing a magnetic field of 3.8T.[i]

Finally, the coordinates used at particle experiments will be portrayed, as they will be used in the experimental analysis further on in the text. The coordinate system can be seen in Figure 6. The coordinate system at CMS is centred at the nominal collision point of the protons, where the **y** axis points upwards, the **x** axis radially towards the centre of the LHC circle, and the **z** axis along the beam axis tangentially at every point. The polar coordinates used include **φ**, the azimuthial angle measured from the **x** axis in the **xy** plane, and **θ**, the polar angle measured from the **z** axis. Finally, another coordinate system includes the pseudorapidity variable **η**, and **φ**, defined as:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \quad \text{(I)}$$

This variable is the limit of the variable **y** of rapidity, in the limit of negligible mass because of very high energetic particles (mainly *jets*, discussed in the following chapters), given by:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \quad \text{(J)}$$

where **E** is the energy and $\mathbf{p_z}$ the momentum along the **z** axis of the produced particle jets. Rapidity is widely used as a parameter in coordinate systems, as rapidity differences remain *Lorentz invariant* under boosts along the beam axis.[ii]

Finally, a reference will be made to another measured variable within the detector, that is of crucial importance for any particle physics analysis: the transverse (to the beam axis) momentum $\mathbf{p_t}$, given by:[iii]

$$p_t = \sqrt{p_x^2 + p_y^2} \quad \text{(K)}$$

i The CMS Collaboration 2019, p. 2.
ii The CMS Collaboration 2008, p. 2; Thomson 2013, p. 275.
iii Thomson 2013, p. 274.

Source: as in Kuusela and Panaretos 2014 (Barney (2004)). Copyright: CERN, for the benefit of the CMS Collaboration.

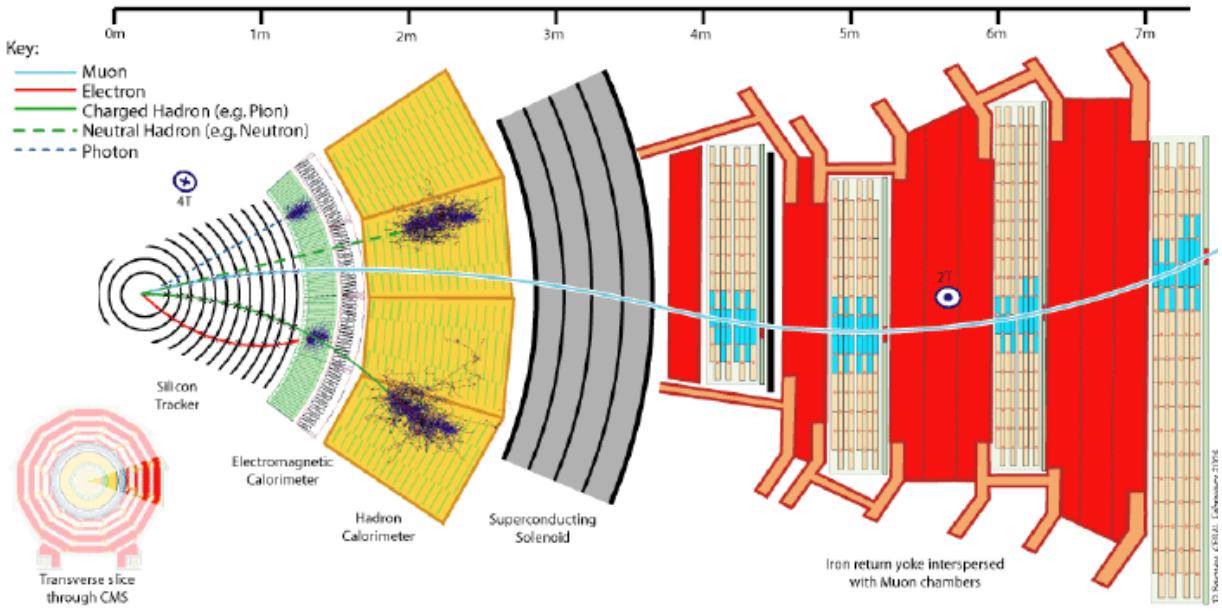


Figure 5: Illustration of the detection of particles at the CMS experiment (Barney, 2004). Each type of a particle leaves its characteristic trace in the various subdetectors of the experiment. This enables identification of different particles as well as the measurement of their energies and trajectories.

Source: Delmastro 2014, p. 19.

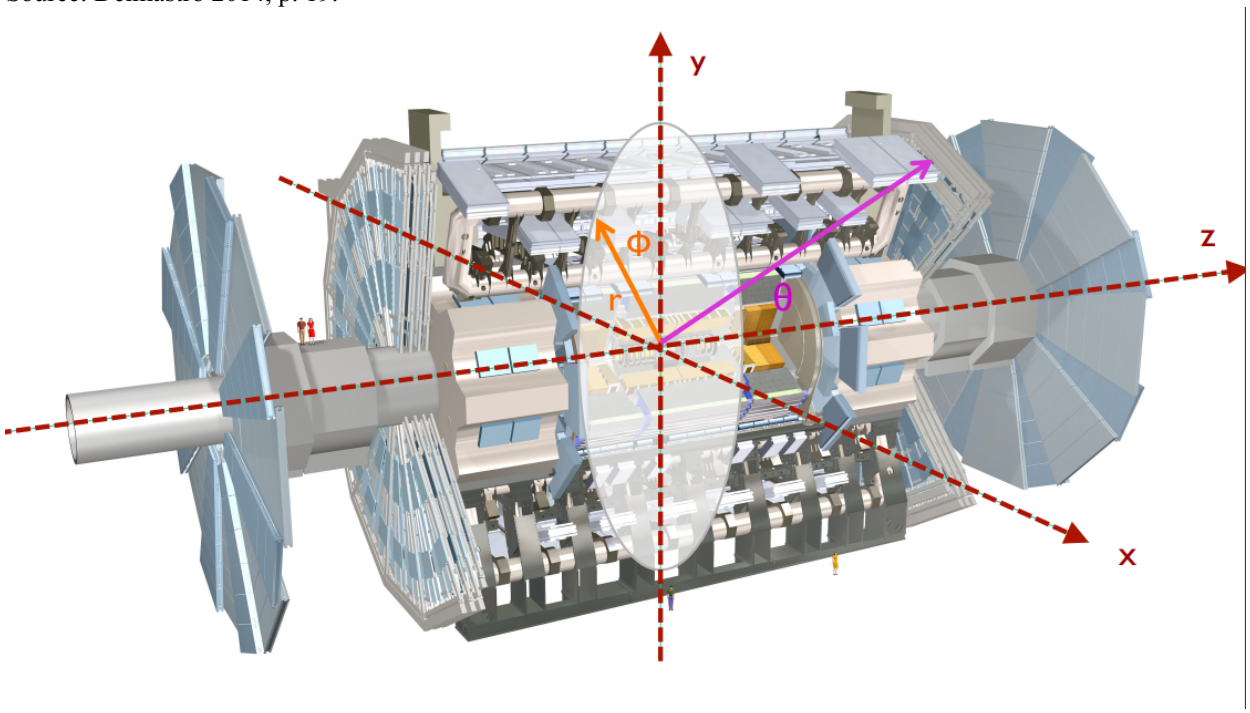


Figure 6: The detector coordinates used at CMS.

III. UNDERLYING THEORIES OF MODELS THAT PREDICT A Z'_{TC2} AND CROSS SECTION PHENOMENOLOGY

In this chapter I will briefly present the motivation, theory and experimentally expected results of the **Z'**, the particle in question. The chapter begins with an insight into some of the unanswered questions by the SM, to be continued in the second subchapter, by one of the solutions to some of the matters the SM has not been able to explain. One of these solutions is the theory of *Topcolour Assisted Technicolour* (TC2), an evolution of the popular Technicolour (TC) and *Extended Technicolour* (ETC) theories first developed during the 1970s by Weinberg and Susskind[i]. The TC2 theories, while solving the "gaps" in the SM, also produce an experimentally observable and testable consequence, the existence of a new type of boson, the **Z′** (or also **Z***, **Z** prime). This testable consequence has the potential of being discovered in particle accelerators, among them being the LHC at CERN, and the final subchapterS will explore calculations of its cross section and its possible detectable decay modes.

III.i. Motivations for the existence of a new Z' boson

As explained in the previous chapter, in the review of the Standard Model of Particle Physics, the last piece of the puzzle was resolved with the discovery of the Higgs boson in 2012 at CERN, and therefore the affirmation of the Higgs mechanism model, how the quarks and leptons acquire mass by interacting with the Higgs field. At the *electroweak scale*, the interactions of the various particles with the Higgs field cause ElectroWeak Symmetry Breaking (EWSB), ultimately "giving" them masses. However, as yet unknown is the fundamental processes and underlying interactions that govern EWSB; the Higgs particle solves the problem superficially, but how it actually acts, or even what it is, remains a mystery. Before the discovery of the Higgs boson, the theory of it had already been proposed, but there were still numerous suggestions on what it actually was. As Kenneth Lane mentions in his review "An Introduction to Technicolor" (1994), on ways that EWSB might manifest itself: "[it may be] a single new particle – the ""Higgs boson"; it may be several such bosons; or a replication of all the known particles; or an infinite tower of new resonances; or something still unimagined".[ii] This concludes the first motivation for necessitating such new physics searches.

i Hill and Simmons, 2003, p.16.
ii Lane, 1994, p. 3.

The second reason is directly linked to EWSB: the *electroweak scale* (or *weak* scale), at which EWSB actually takes place. It is located around the energy of 1 TeV, a scale that arises from the vacuum expectation value (VEV) of the gauge Higgs field, or more specifically, the decay constant of the Nambu-Goldstone bosons created by the Higgs mechanism f_π, that will then be transformed into the longitudinal components of the weak bosons ($W_L^\pm$ and Z_L^0). The weak scale, therefore is given by:

$$f_\pi = v_{weak} = \frac{1}{\sqrt{2\sqrt{2}G_F}} = 246\,GeV \quad \text{(L)}$$

where G_F is the Fermi constant (the strength of the weal interaction).[i] The weak scale opposes the other two fundamental mass scales in nature, the QCD or strong interaction scale, Λ_{QCD}, and the gravitational scale, M_{Planck}, which arise naturally from the underlying theories, such as the SM and Quantum Mechanics. On the contrary, the weak scale does not arise from any underlying theories or symmetries, and usually requires fine-tuning to be produced; this is exactly what TC2 will aim to solve and mimic: "a similar dynamical and natural origin".[ii]

Closing, with the third motivation for the development of TC2, is the mass of the top quark. As its mass is so heavy compared to the rest of the quarks (the next heaviest is the bottom quark, with a mass of 4.3GeV/c^2, 40 times smaller, while as the lightest, the up quark, at 0.003GeV/c^2, is 10000 times smaller[iii]), it naturally has an equally large coupling to the Higgs field. Therefore, "it is natural to wonder whether m_t [the mass of the top] has a different origin than the masses of the other quarks and leptons".[iv] Could the top quark be the key to EWSB?

III.ii. The emergence of a Z'_{TC2} boson in a Topcolour Assisted Technicolour theory

There are many theories which are capable of solving the questions above, some refuted and revolutionised, while some others are still the subject of searches for physics Beyond the Standard Model (BSM). This thesis will focus on the theoretical presentation and following experimental search based on the Topcolour Assisted Technicolour (TC2) model, as numerous, if not most modern experiments search for new physics based on this model, among others, and because many already conducted searches from this theory serve as benchmarks for future ones. The most recent collaborative publication regarding a search for the Z' based on TC2 is by the ATLAS Collaboration in 2021, "Search for $t\bar{t}$ resonances in fully hadronic final states in pp collisions at $\sqrt{s}$ = 13 TeV with

i Lane, 1994, p. 3; Hill 1994, p. 1
ii Hill and Simmons 2003, p. 5
iii Particle Data Group 2000
iv Popovic and Simmons, 1998, p. 2

the ATLAS detector", in which they conduct a search for the TC2 Z'_{TC2}, using previously set benchmarks by other searches based on the same model.

Source: "Theorists and Experimentalists: Partners in the Search" (1993).

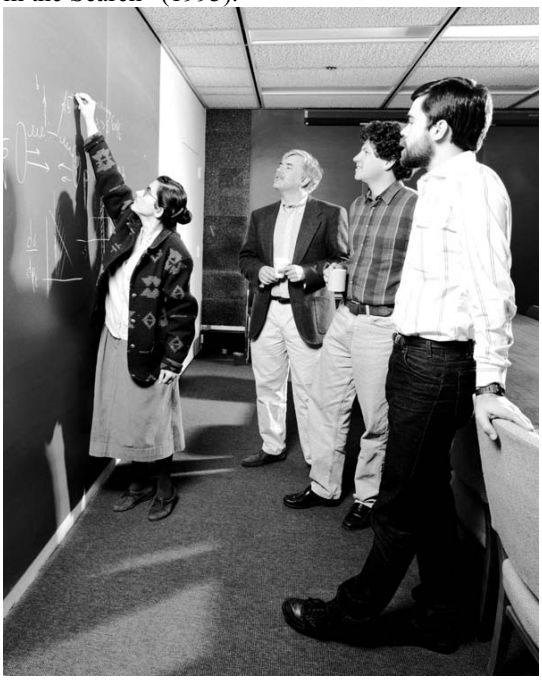


Figure 7: From the left, two experimental physicists, Fritz Dejongh and Vaia Papadimitriou, discuss a question in b physics with theorists Christopher Hill and Andreas Kronfeld

TC2 is the evolution of the TC theory, first developed in the late 1970s, and later on ETC, while also embedding the idea of Topcolour. A main figure in developing much of TC and relevant theories, and the first to suggest TC2, seems to be the American physicist at Fermilab, Christopher T. Hill[i]. He has worked on the development of such models, and also on cross section calculations that may be possible to detect in particle colliders. During his PhD thesis, "Higgs Scalars and the Nonleptonic Weak Interactions" (1977)[ii], Richard Feynman was among the executive committee, while his supervisor was Murray Gell-Mann.

Very briefly, Technicolour theories were developed extensively as they were able to make predictions within the energy scale of EWSB, around 1TeV, while they are one of the most popular realisations of models that can explain EWSB and predict the correct masses for the **W** and **Z** bosons.[iii] The main ideas developed within TC, in trying to explain EWSB, is creating equivalent dynamics, analogous to the strong force and the symmetry that governs QCD. It is therefore assumed that the weak scale interactions are governed by *Novel Strong Dynamics* (NSD), in complete analogy to the Strong Dynamics for QCD. For that reason, the theory develops an added sector, that contains *technifermions*, *techniquarks* and *technigluons*, the details of which will not be delved into here.[iv]

Pure TC manages to describe successfully the breakage of the chiral symmetries of the new fermions of the theory, the *technifermions.* However, it is deemed non-realistic and extended to Extended Technicolour (ETC), which develops a mechanism for transmitting EWSB to the ordinary

i For more information about Topcolour, TC and relevant theories see: Bardeen, Hill and Lindner 1990, Hill 1991 and Hill 1994.

ii Hill 1977.

iii Lane 1994, p.1; Hill and Simmons, 2003, p. 17.

iv Hill and Simmons, 2003, σελ. 6,19, 24.

SM quarks and leptons[i]. As the measurements though of the mass of the top quark got more and more precise, it became clear that the mass of the top quark was of the same order of magnitude as the weak scale where EWSB happens. Theories were then developed that encompassed the top quark as playing a special role in TC theories, possibly as some kind of special techniquark. Some early attempts, but ultimately unacceptable as requiring fine-tuning parameters, were top-quark condensation and Topcolour, which were finally moulded into Topcolour Assisted Technicolour, incorporating the best features from TC and Topcolour.[ii] Using elements from the top-quark condensation models, one of the main themes of TC2 is the postulate that the Higgs boson is not an elementary particle, but a bound state of a top – antitop pair, a $t\bar{t}$ condensate, $\langle t\bar{t} \rangle$. This, in addition to Topcolour dynamics, which can be reviewed in Hill (1991), creates the theory of TC2, which will be analytically presented in the following paragraphs.[iii]

The TC2 model, including solving the problems of EWSB and generating a large top quark mass, also has a testable consequence, which is the existence of a new Z' boson. The outline is the insertion of a new strong gauge force, similar to the NSD of TC, that will preferentially couple to the third generation of quarks. This new force is set to generate the $t\bar{t}$ condensate, that will ultimately be responsible for a large top mass. TC2, however, needs yet another mechanism for all to fall in to place, that of a *tilting mechanism*, which will ensure the enhancement of the production of a $t\bar{t}$ condensate, and block a similar formation of a $b\bar{b}$ condensate.[iv] Without this mechanism, these two condensates could be equal, and the theory would generate masses of very close magnitudes, as opposed to a 40 times difference in mass. One way of incorporating tilting into the theory is to introduce a new boson, a neutral Z'. From here on, the only Z' to be discussed will be that of a TC2 theory, and it shall be denoted as Z'_{TC2}.

This new boson, shall provide an attractive interaction between $t\bar{t}$, and a repulsive one for $b\bar{b}$, so as to create a light bottom quark and a heavy top. The equivalent Z boson in the SM framework works in a similar way, although it is too weak to provide this type of tilting. In the publication by R.M. Harris, C.T. Hill and S.J. Parke, "Cross Section for Topcolor Z'_t decaying to $t\bar{t}$" (1999), there are four different models presented that all produce different Z'_{TC2}, with respect to their resonance widths, their cross sections and the ways in which they actualise the tilting mechanism. The simplest model, referred to as "Model I", and originating in the first paper on TC2 "Topcolor Assisted Technicolor" by C.T. Hill in 1994, only requires embedding an extended

i Hill and Simmons, 2003, σελ. 6,19, 24
ii Hill and Simmons, 2003, p. 18.
iii T. Hill and Simmons (2003), σελ. 18.
iv Harris, Hill and Parke, 1999, p. 2.

electroweak sector, which includes an extra U(1) group that is responsible for the new boson. Model I produces the lowest cross sections out of all four models, while as II and III are similar and produce slightly larger cross section values.[i]

Model IV is a more interesting proposition, not only for its complexity and plausibility, but because it is the most searched for model experimentally. All searches for a Z'_{TC2} boson that arise from a TC2 model are based on the cross section calculations of this model, and as far as the author is aware, there have not been any recent searches for the other models at the LHC (i.e. at the accelerator with the highest centre of mass energy at present). Therefore I shall only present the physics for Model IV, and begin in the next subchapter by providing the relevant Lagrangian and the method to compute the cross sections for the current centre of mass energies at the LHC[ii].

The main idea of the first three models is to extend the SM by adding an extra electroweak sector of the form:

$$SU(2)_W \times U(1)_h \times U(1)_\ell \qquad (M)$$

assuming a stronger coupling g_h of the $U(1)_h$ group, compared to the coupling g_ℓ of the $U(1)_\ell$. These two hypercharge groups will break into a subgroup $U(1)_Y$, effectively giving mass to a new Z'_{TC2} boson.[iii]

The fourth model however, is not characterised by $U(1)_Y$ mixing, and instead continues using the standard coupling constant g_1 of the SM, normalised to fit with the model. It is a non-standard model, meaning that the $U(1)_Y \rightarrow U(1)_h \times U(1)_\ell$ and the generations are grouped differently. In Model IV, the first and third generation of quarks *only* shall be grouped together as having strong couplings to the Z'_{TC2}, as to ensure topcolour tilting this model assumes a leptophobic interaction.[iv]

III.iii. Cross section calculations and predictions for the leptophobic "Model IV"

The tilting mechanism mentioned in the previous section is responsible for the enhancement of the $t\bar{t}$ condensate and not the $b\bar{b}$ one, which is assumed to make up the inner structure of the Higgs boson. This allows the theory to couple preferentially to the third generation, and specifically the top quark, generating its unusual heaviness. In the following paragraphs, I will present the

i Harris, Hill and Parke, 1999, p. 2-3; Hill and Simmons (2003), p. 129.
ii The ATLAS Collaboration, 2021, p. 1; Harris, Hill and Parke, 1999, p. 3.
iii Hill and Simmons 2003, p. 129; Harris, Hill and Parke, 1999, p. 2-3.
iv Harris, Hill and Parke, 1999, p. 5, 8.

equations calculated by C.T. Hill, R.M. Harris and S.J. Parke in 1999 of the decay widths and cross sections based on Model IV discussed above.

Model IV materialises tilting using a leptophobic interaction. Therefore the $\mathbf{Z'_{TC2}}$ is postulated to couple to the first and third of only the quark generations, as follows:

$$\text{quark generations } (1,3) \supset U(1)_2$$

Therefore, during subsequent partial width calculations, decays of the $\mathbf{Z'_{TC2}}$ to all first and third generation quarks will have to be considered. The dominant part of the interaction Lagrangian, as provided in the original publication "Cross Section for Topcolor Z'$_t$ decaying to $t\bar{t}$" (1999), is:

$$L'_{IV} = \left(\tfrac{1}{2} g_1 \cot\theta_H\right) Z'^{\mu} \left(\bar{t}_L \gamma_\mu t_L + \bar{b}_L \gamma_\mu b_L + f_1 \bar{t}_R \gamma_\mu t_R + f_2 \bar{b}_R \gamma_\mu b_R\right) -$$

$$- \bar{u}_L \gamma_\mu u_L - \bar{d}_L \gamma_\mu d_L - f_1 \bar{u}_R \gamma_\mu u_R - f_2 \bar{d}_R \gamma_\mu d_R$$

(N)

As can be seen from the Lagrangian, the $\mathbf{Z'_{TC2}}$ boson in question couples to all first and third generation quarks, for example the terms $\bar{t}_L \gamma_\mu t_L$ and $\bar{d}_L \gamma_\mu d_L$ for the top and down quark respectively, Two additional constants, f_1 and f_2, are added for topcolour tilting, in order to ensure an attractive $t\bar{t}$ channel, and a repulsive $b\bar{b}$ channel, i.e. $f_1 > 0$ and/or $f_2 < 0$. To avoid fine-tuning the theory, the constraint on the mixing angle, $\cot\theta_h >> 1$, is also imposed.

By summing on all colour and spin states, for initial and final states, the cross section for this Model is:

$$\sigma = \frac{9\alpha^2\pi}{16\cos^4\theta_W} \cot^4\theta_H \times z_1 \times \left[\beta\left(1+\frac{1}{3}\beta^2\right) \times z_2 + f_1 \beta\left(1-\beta^2\right)\right] \left[\frac{s}{\left(s - M^2_{Z'_{TC2}}\right)^2 + s\Gamma^2}\right] \theta\left(s - 4m_t^2\right) \quad \text{(O)}$$

where z_1 is a variable I have inputted, which according to the final state, takes the following values:

$$1 + f_1^2 \ , \quad for\ u + \bar{u}\ initial\ state$$

$$1 + f_2^2 \ , \quad for\ d + \bar{d}\ initial\ state$$

while as z_2 takes the following values:

$$1 + f_1^2 \ , \quad for\ t + \bar{t} \quad or \quad u + \bar{u}\ initial\ state$$

$$1 + f_1^2 \ , \quad for\ b + \bar{b} \quad or \quad d + \bar{d}\ initial\ state$$

The last term is added to ensure that the final state will contain a $t\bar{t}$ pair, as the other terms contain coupling to all four quarks in the first and third generations.

As there are couplings to all four quarks, the up, down, bottom and top, the partial widths for each final state are calculated:

$$\Gamma(Z'_{TC2} \to t\bar{t}) = \frac{\alpha \cot^2\theta_H}{8\cos^2\theta_W}\sqrt{M^2_{Z'_{TC2}} - 4m_t^2}\left((1+f_1^2)\left[1-\frac{m_t^2}{M^2_{Z'_{TC2}}}\right]-3f_1\left[\frac{m_t^2}{M^2_{Z'_{TC2}}}\right]\right) \quad \text{(P)}$$

$$\Gamma(Z'_{TC2} \to \theta\bar{\theta}) = \frac{\alpha \cot^2\theta_H M^2_{Z'_{TC2}}}{8\cos^2\theta_W}(1+f_1^2) \quad \text{(Q)}$$

$$\Gamma(Z'_{TC2} \to b\bar{b}) = \frac{\alpha \cot^2\theta_H M_{Z'_{TC2}}}{8\cos^2\theta_W}(1+f_2^2) \quad \text{(R)}$$

$$\Gamma(Z'_{TC2} \to d\bar{d}) = \Gamma(Z'_{TC2} \to b\bar{b}) = \frac{\alpha \cot^2\theta_H M_{Z'_{TC2}}}{8\cos^2\theta_W}(1+f_2^2) \quad \text{(S)}$$

Concluding, the total decay width will be:

$$\Gamma = \frac{\alpha \cot^2\theta_H M_{Z'_{TC2}}}{8\cos^2\theta_W}\left[\sqrt{1-\frac{4m_t^2}{M^2_{Z'_{TC2}}}}\left((1+f_1^2)-(1+f_1^2+3f_1)\frac{m_t^2}{M_{Z'_{TC2}}}\right)+(3+f_1^2+2f_2^2)\right] \quad \text{(T)}$$

The authors of the publication offer a simplification, assuming $f_1 = 1$ and $f_2 = 0$, to ensure a leptophobic, topophyllic and b-phobic (for the right-handed terms) final state, as will also be the focus and search in this thesis. The following formulae are for the cross section and decay width of this special case, and will be those used and referred to as σ_{TC2} and Γ_{TC2} from now on:

$$\sigma_{TC2} = \frac{9\alpha^2\pi}{16\cos^4\theta_W}\cot^4\theta_H \times z_1 \times \left[\beta\left(1+\frac{1}{3}\beta^2\right)\times z_2 + \beta(1-\beta^2)\right]\left[\frac{s}{(s-M^2_{Z'_{TC2}})^2+s\Gamma^2}\right]\theta(s-4m_t^2) \quad \text{(U)}$$

where now z_1 shall be:

$$2\ , \quad for\ u+\bar{u}\ initial\ state$$
$$1\ , \quad for\ d+\bar{d}\ initial\ state$$

and z_2:

$$2\ , \quad for\ t+\bar{t}\ \ or\ \ u+\bar{u}\ initial\ state$$
$$1\ , \quad for\ b+\bar{b}\ \ or\ \ d+\bar{d}\ \ initial\ state$$

The decay width will be:[i]

$$\Gamma = \frac{\alpha \cot^2\theta_H M_{Z'_{TC2}}}{8\cos^2\theta_W}\left[\sqrt{1-\frac{4m_t^2}{M^2_{Z'_{TC2}}}}\left(2-5\frac{m_t^2}{M_{Z'_{TC2}}}\right)+4\right] \quad \text{(V)}$$

The cross section and width above can be used to numerically calculate the cross section values, as a function of the Z'_{TC2} mass over width ratio, for pp collisions at the LHC. As this thesis realises searches for four different possible masses at the width of 1%, only 1% width calculations will be presented further on.

i Harris, Hill and Parke, 1999, p. 8-9

III.iv. Top-antitop pair production and decay modes

Based on Hill's model IV from TC2, the Z'_{TC2} theoretically decays to a top-antitop pair ($t\bar{t}$) (Figure 8). Therefore the attention will be on $t\bar{t}$ pair production processes. During the experimental process, *$t\bar{t}$ resonances* are sought out, massive resonances which pertain to the presence of a massive particle in the $t\bar{t}$ mass spectrum ($m_{t\bar{t}}$). If observed, these resonances will appear as local deviations from the SM $t\bar{t}$ background processes as "lumps" protruding in the mass distribution.[i]

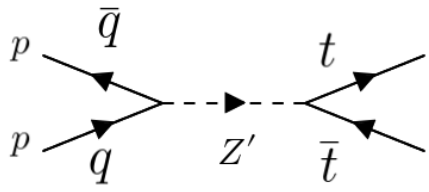


Figure 8: Feynman diagram of the decay of the Z'_{TC2}.

Source: Thomson, 2013, p. 24.

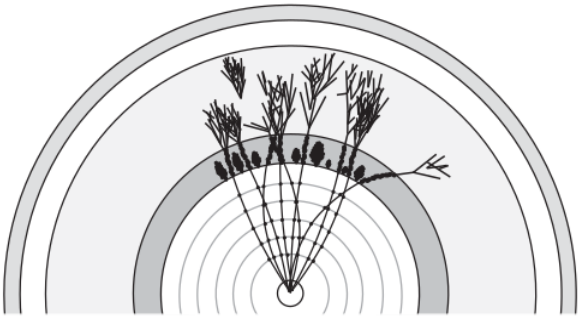


Figure 9: An illustration of the appearance of a jet in a detector. In practice, the individual particles are not resolved.

The top quark is special, for, as opposed to the other five, it has huge mass at 172.69 ± 0.30GeV/c^2 [ii], which gives it interesting properties. One of them is that it can be observed "naked" in the detector, as its *hadronisation* time is larger than its lifetime, meaning that it decays electroweakly into its products, before it has time to hadronise. *Hadronisation* is the process in which *jets* are produced instead of quark tracks; as a result of QCD interactions, the Strong interaction energy between the quarks (at a distance of 10^{-15}m) is transformed into more quark pairs, finally creating a "shower" of quarks, which make up the *jet.* As a result, if one quark q is produced in a given process, pairs of $q\bar{q}$ will be finally be created and show the experimental signature of a jet (see Figure 9). The other quarks hadronise much faster, and there is no time to

Source: Lannon, Margaroli and Neu 2012, p. 3

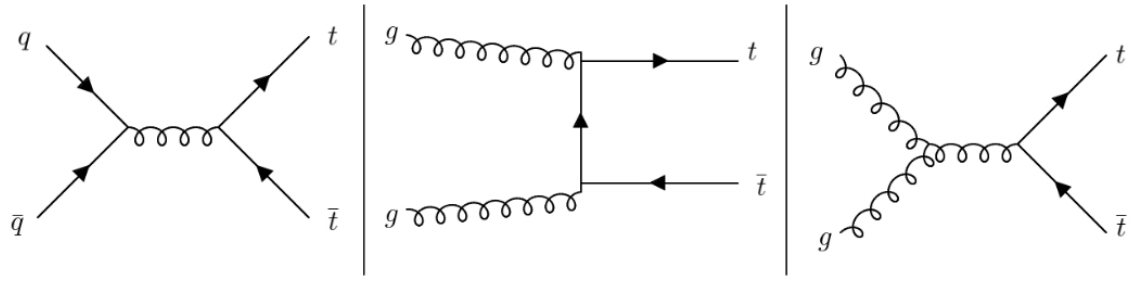


Figure 10: Feynman diagrams for single top quark production. Represented are (a) a LO s-channel diagram, (b) a NLO t -channel diagram, and (c) a NLO W t production diagram

i The ATLAS Collaboration 2021, p. 3
ii Workman et al. 2022

"see" their products in the detector.[i] At particle accelerators, a top pair can be produced in two ways: either through $q\bar{q}$ annihilations or through gluon-gluon fusion (Figure 9) and at the LHC the first method is favoured by 15%, while as Z'_{TC2} production also arises from the first[ii].

Regarding the decay products of the top quark, it is found that at a rate of 99.8% it decays into a **W** boson and a bottom quark **b**, as it is more massive than the **W**. Keeping in mind that the current search will be for $t\bar{t}$ pairs, if the top quark **t** (antitop $\bar{t}$) decays to a W^+ (W^-), then there will three different decay possibilities, because of the two decay possibilities of the **W**: 67.7% to $q\bar{q}$ ($c\bar{s}$ or $u\bar{d}$) and 32.4% to a charged lepton **l** and its corresponding SM doublet neutrino ν_l. Therefore, for the $t\bar{t}$ pair case, there will be three possible final states: *fully hadronic* with only quarks in the final state; *lepton+jets* (or *semileptonic*) where one **W** decays to quarks and the other to leptons; *dileptonic*, with only leptons in the final state. In Figure 11 one can see a table for the various decay possibilities, while as Figure 12 shows a representation of one possible $t\bar{t}$ bar event.[iii]

This search for the Z'_{TC2} particle in $t\bar{t}$ final states will focus on only the *fullly hadronic* final state case. Further on in the text the analysis will be orientated to isolate only jet signals.

Source: Lannon, Margaroli and Neu 2012, p. 3

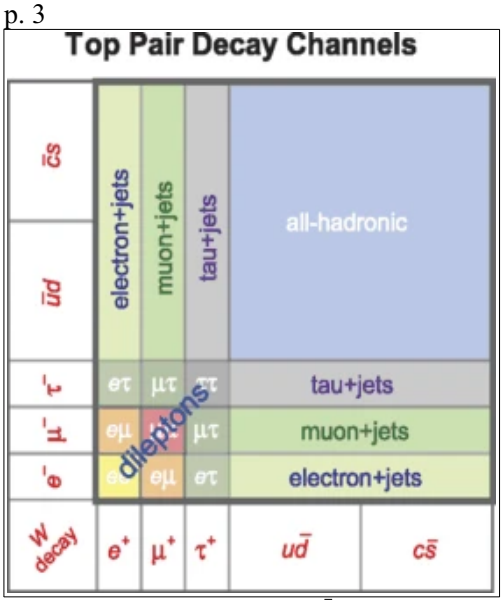


Figure 11: Final states of the $t\bar{t}$ system

Source: Raeky 2009.

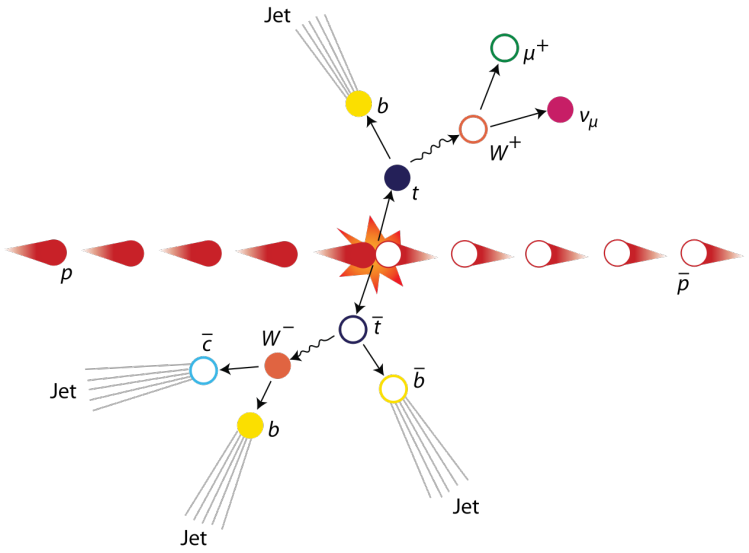


Figure 12: Representation of a possible $t\bar{t}$ final state

i Lannon, Margaroli and Neu 2012, p. 1-3
ii The ATLAS Collaboration 2021, p. 3; Lannon, Margaroli and Neu (2012), σελ. 3.
iii Lannon, Margaroli and Neu 2012, p. 1-3

EXPERIMENTAL COMPONENT

IV. EXPERIMENTAL DETECTION AND ANALYSIS METHODS USED BY CMS

This chapter marks the beginning of the Experimental Component of this text, in which the previous theoretical backgrounds will be seen being used "in action" to acquire real experimental results.

It begins with a small overview on how CMS "reconstructs" jets within the detector, the method in which particles and their properties are identified. The jet reconstruction algorithm used, the anti-k_t method, will be briefly presented, and finally the parameter of N-subjettiness will be introduced, for identifying the number of subjets within a larger main jet.

The next section will briefly refer to the data analysis framework used to actualise the analysis, ROOT, and some basic definitions needed for the final analysis chapter.

The last section will describe the statistical methods used to procure the final result of the analysis: the fitting process, the calculation of the log-likelihood function, and the Bayesian statistical approach to calculating the upper limits for the possible discovery of a Z'_{TC2}.

IV.i. Jet reconstruction techniques

a. The Anti-k_t algorithm

x As in this analysis the main focus will be the production of top (or antitop, from now on both referred to as "top jets") jets, methods are applied during the collection of the data, or the production of the Monte Carlo (MC) simulations, to create *clusters* of the detected experimental signatures. These methods are applied mostly to highly collimated quarks and gluons, as they hadronise very fast after their production; it is therefore necessary to try and reconstruct the jets and identify the original particle that produced them.[i]

The most commonly used algorithm to *cluster* particles into reconstructed jets used at CMS is the *anti-k_t jet clustering algorithm*, and it is one of many methods[ii]. As this is an undergraduate thesis, the author did not apply these methods during any point in her analysis, and instead procured the already reconstructed jet histograms from files from her colleagues. Nevertheless, as this is an

i Bakas *a* 2023, p. 65.
ii Bakas *a* 2023, p. 65.

important method in event collection at CMS, there will be a brief overview in the following paragraphs.

One begins by defining the following distance variables, d_{ij} and d_{iB}:

$$d_{ij} = min\left(k_{ti}^{2p}, k_{tj}^{2p}\right)\frac{\Delta_{ij}^2}{R^2} \quad \text{(W)} \qquad , \qquad d_{iB} = k_{ti}^{2p} \quad \text{(X)}$$

where d_{ij} expresses the distance between the particle i and the pseudojet j, while as d_{iB} is the distance between the particle i and the beamline (B). The first definition includes finding the minimum transverse momentum k_t out of two particles, and multiplying by the distance between them in the η-φ (eta-phi, rapidity-azimuth) plane Δ_{ij}, finally dividing by a jet radius parameter R. The distance Δ_{ij} is defined as:

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\varphi_i - \varphi_j)^2 \quad \text{(Y)}$$

The extra parameter p is what discriminates this algorithm from other clustering methods, as it allows control over the relative power of the energy compared to the geometry of the jets.
The algorithm procedure is as follows: the two types of distances are calculated by looping over all events and particles in the analysis; for each event, the smaller of the two is taken, and accordingly, if the smaller value is the d_{ij} distance, the two particles i and j are grouped together and recombined, while if it is the d_{iB} the i particle is labelled as a "jet", and is removed from the loop over the particles. This loop procedure is continued until all entities have been exhausted and organised into clusters.[i]

This completes the review of the anti-k_t clustering algorithm. Next, the N-subjettiness variable will be introduced, which identifies the number of subjets within an already reconstructed jet.

b. N-subjettiness

In the previous section, an algorithm was discussed in order to cluster and reconstruct jets together, after the production in the detector. After the clustering procedure and the jet reconstruction, in this case of a top and antitop quark, further methods are used to identify subjets within these original reconstructed jets.

i Cacciari, Salam and Soyez 2008, p. 1-2.

In the case of a top jet, there will be three subjets within the originally reconstructed jet structure: the bottom quark, and the products of the **W** boson decaying *hadronically* into two quarks. As mentioned above, this analysis will focus on the fully hadronic final state of the top - antitop pair.

A widespread used method in identifying the jet substructure geometry, used in the last ten years, is calculating the so-called variable *N-subjettiness*. This variable can be used as a distinguishing parameter between boosted hadronically decaying objects and QCD backgrounds. As will be seen later in the analysis, a boosted phase space of high transverse momentum $\mathbf{p_t}$ cut will be applied for the analysis purposes; within this phase space decaying top jets collimate and it is possible to reconstruct all the products within one jet, as seen in Figure 13. Additionally, a QCD background will also be present during the current analysis, which, unlike the boosted top jets, does not contain "lobes" of three energy deposits. This is what N-subjettiness exploits, by effectively measuring the number of subjets within an already reconstructed jet. The discriminating variable of N-subjettiness therefore aligns exactly with the purposes of this thesis.[i]

Source: as seen in Bakas *a* (2023), p. 17.

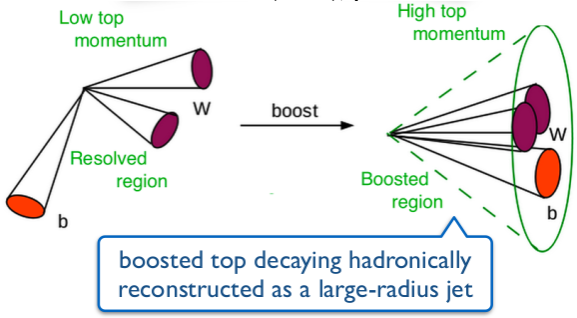


Figure 13: Non boosted top jets as opposed to boosted with high p_t jets. The second case can easily be reconstructed as a main jet with three subjets.

Assuming an already reconstructed main jet, **N** candidate subjets are originally theorised.[ii] N-subjettiness is defined as an *inclusive jet shape*, symbolised as τ_N, as follows:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} min\{\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k}\} \quad (Z)$$

where the **k** index spans the constituent particles within the reconstructed jet, $\mathbf{p_t}$ is their transverse momentum, and **ΔR** is defined as follows:

$$\Delta R_{J,k} = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2} \quad (A.a)$$

This variable represents the distance between a constituent particle **k** and a candidate subjet **J** in the **η-φ** (eta-phi, rapidity-azimuth) plane, which are popularly used coordinates used in the detector (see Chapter II.ii). This variable is one of which will be calculated later on during the analysis part of this text.

i Thaler and Tilburg, 2011, p. 2-3.

ii In this analysis, the correct number of subjets will be theorised, based on the expectations of having three subjets in top jet events. In more complex situations, minimisation procedures can be applied to calculate **N**.

The normalisation constant d_0 is taken as follows:

$$d_0 = \sum_k p_{T,k} R_0 \quad \text{(A/b)}$$

where R_0 is the radius that was used in the jet clustering algorithm to initially reconstruct the original jet.

By calculating τ_N, one manages to quantify *how subjetty* a jet is, or similarly, how accurate the original guesswork on the number of subjets is. Looking at the limits of the variable, if N-subjettiness is calculated to be around 0, it means that all the subjets are very close to the reconstructed jet and are aligned with it, meaning the assumption was correct and that the jet contains N or fewer subjets. On the contrary, if τ_N is large, it signifies that the energy of the subjets are distributed far from the jet centre, meaning that one has "failed" to intuit the approximate number of subjets. In this case, the jet in reality has at least N+1 subjets.

As an example, the case of distinguishing W^+W^- production jets and QCD jets will be used. In Figure 14 event displays are shown of their energy deposit geometry; it is clear that the W jets exhibit a more distinguishable dijet topology, with two subjets from the quarks in the hadronic final state, while as the QCD jet "splits" multiple times, making its geometry appear more disperse. For this reason, it is at first logical to assume that because the W jets contain two principal subjets, and the QCD jet could be categorised as having one subjet, that the τ_2 and τ_1 respectively would be the correct distinguishing parameters. However, observing the histograms of τ_1 and τ_2 for both types of events (Figure 15), they do not seem to provide much distinguishing power. The ratio of these variables though, remarkably, is the optimal discriminating variable, as can be seen in Figure 16 (a). In 16 (b) a multivariate analysis is shown, in which the optimal relation between τ_1 and τ_2 can be calculated.[i]

In conclusion, the ratio of the N-subjettiness pertaining to the boosted hadronically decaying jets signal to that of the QCD background provides significant discriminating power between the two. In the next chapter, it will be shown how these variables are used in practice, as they will form a fundamental part of the preliminary analysis.

i Thaler and Tilburg, 2011, p. 3-5.

Source: Thaler and Tilburg (2011), p. 4.

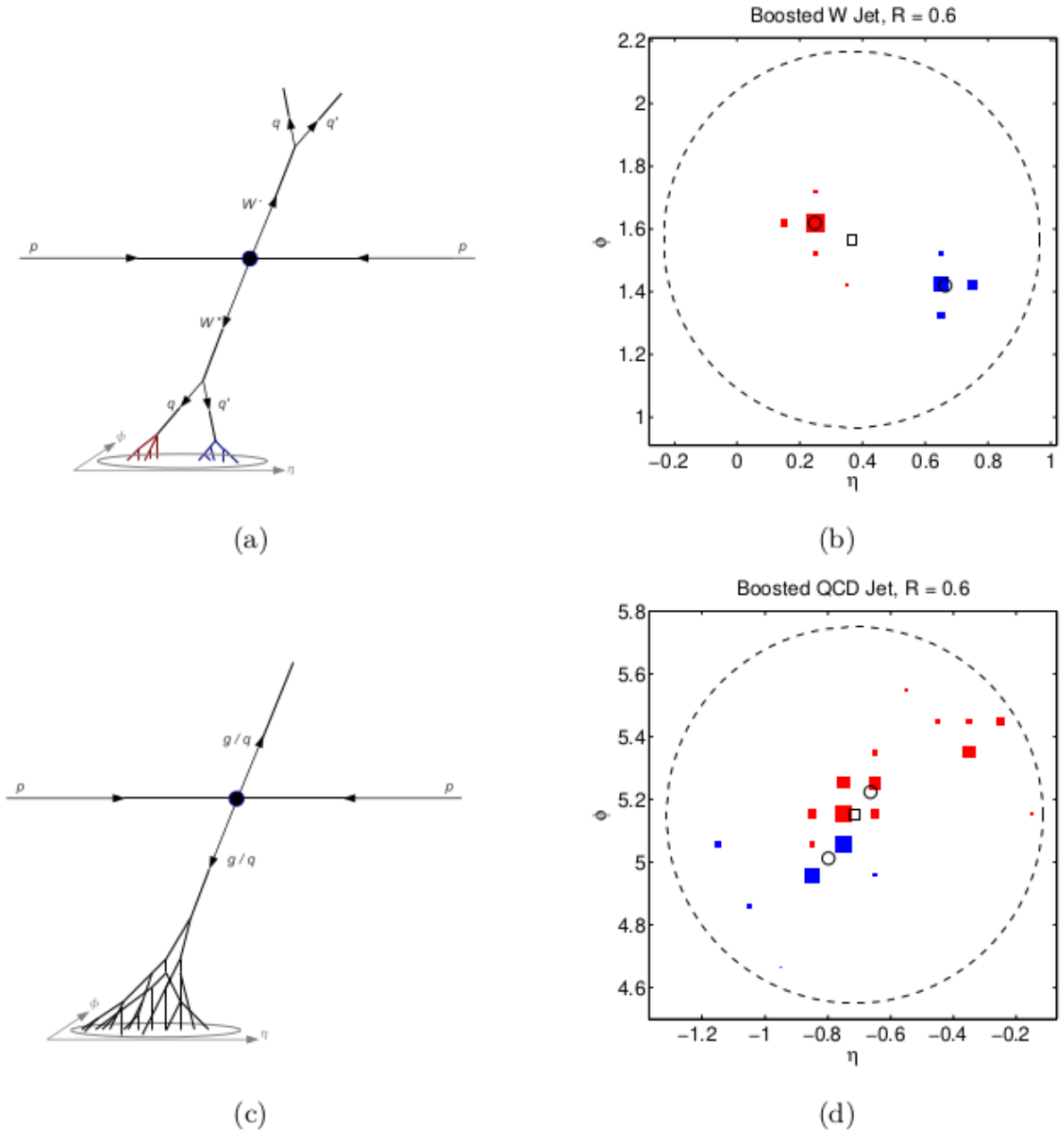
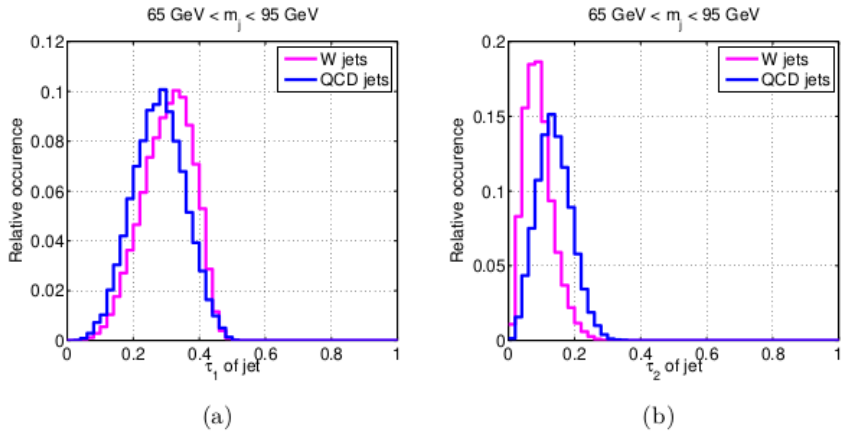


Figure 14: Left: Schematic of the fully hadronic decay sequences in (a) W^+W^- and (c) dijet QCD events. Whereas a W jet is typically composed of two distinct lobes of energy, a QCD jet acquires invariant mass through multiple splittings. Right: Typical event displays for (b) W jets and (d) QCD jets with invariant mass near m W. The jets are clustered with the anti-k_T jet algorithm [31] using R = 0.6, with the dashed line giving the approximate boundary of the jet. The marker size for each calorimeter cell is proportional to the logarithm of the particle energies in the cell. The cells are colored according to how the exclusive k_T algorithm divides the cells into two candidate subjets. The open square indicates the total jet direction and the open circles indicate the two subjet directions. The discriminating variable τ_2/τ_1 measures the relative alignment of the jet energy along the open circles compared to the open square.

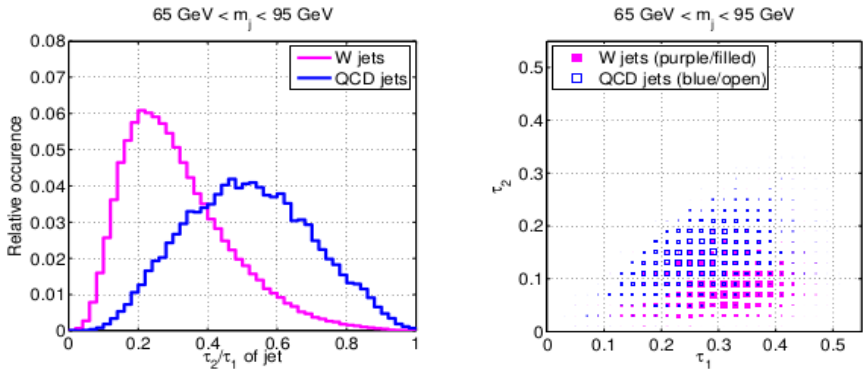
Source: Thaler and Tilburg (2011), p. 5.



(a) (b)

Figure 15: Distributions of (a) τ_1 and (b) τ_2 for boosted W and QCD jets. For these plots, we impose an invariant mass window of 65 GeV < m_{jet} < 95 GeV on jets of R = 0.6, p_t > 300 GeV, and $|\eta|$ < 1.3. By themselves, the τ_N do not offer that much discriminating power for boosted objects beyond the invariant mass cut.

Source: Thaler and Tilburg (2011), p. 5.



(a) (b)

Figure 16: (a): Distribution of τ_2/τ_1 for boosted W and QCD jets. The selection criteria are the same as in Fig. 2. One sees that the τ_2/τ_1 ratio gives considerable separation between W jets and QCD jets beyond the invariant mass cut. (b): Density plot in the τ_1–τ_2 plane. Marker sizes are proportional to the number of jets in a given bin. In principle, a multivariate cut in the τ_1–τ_2 plane would give further distinguishing power.

IV.ii. Processing physical quantities using the ROOT framework by CERN

The ROOT data analysis framework is what will be used to analyse the data and MC simulations during the analysis. This information is stored in the form of *histograms* in files of the type "file.root", while as the programmes written to analyse the histograms are written in files of the type "file.C" (in the C++ programming language). The histograms are organised by variable within the ROOT file within what is called a *TTree*, for example: the scalar variable *mJJ* (the invariant mass of two jets); the jet variable *jetMass* (the mass of one produced jet); the boolean variable *jetIsBtag* (if the jet is a B or not); the integer variable *nJets* (the number of jets during each event);

and others. The jet variables will have as many histograms as the number of jets during each event; the jets are ordered by their transverse momentum $\mathbf{p_t}$, so the jet with the largest $\mathbf{p_t}$ will be labelled as [0], the *leading* jet, while as the next one will be the *subleading* jet, [1] and so on. In the case of this analysis, the interest is turned to two decaying top and antitop jets, so there will only be two to categorise, into leading and subleading.

The MC files used two simulation programmes, namely *pythia* and *madgraph*; *pythia* simulates the hadronisation process, while as *madgraph* uses information from Feynman diagrams and randomly generated numbers to calculate decay widths.

During the analysis and the presentation of the histograms **`hist`** of various variables later on in the text, they will appear with different *scaling factors*, according to what is wished to be observed or calculated. Within the code, the scaling is applied using the command **`hist→Scale(scaling_factor)`**, which simply multiplies each entry in the histogram by the **scaling_factor** number. In Table 1 the different types of scaling are listed.

Type of Scaling	**scaling_factor**
None (pure number of events)	-
Unity	`1.0/hist→Integral()`
Cross Section	$\frac{\sigma_i \cdot L}{N_{generated\ events}}$

Table 1: Types of histogram scaling

The fist type basically includes no modifications of the histogram, and shows exactly the number of events generated in the simulation or measured in the collected data. The second is a scaling to unity, by dividing the histogram by its integral; this is used for *shape comparison* between histograms. The last type is labelled *cross section scaling* and is used for the MC simulations when data is inserted so as to reflect the real number of events that would have been measured in the detector. σ_i is the cross section of the process, and is shown in Table 2 further down, while as the luminosity is taken as $\mathbf{L=36fb^{-1}}$ and $\mathbf{N_{generated\ events}}$ is the number of original events produced for each MC simulation file.

IV.iii. Statistical methods utilised: fitting procedures, log-likelihood calculation and Bayesian upper limits

When reaching the final result during the analysis, the final step will be to apply a fitting procedure to the data, with three free variables: the amount of $t\bar{t}$ background **N_tt**, the amount of QCD background **N_QCD** and the amount of signal **N_signal**. These will be provided by carefully analysed histograms from MC simulations, and the fitting programme will be required to find the optimum solution for the three variables.

This procedure is accompanied by the calculation of the negative log-likelihood distribution (multiplied by a factor of two) (symbolised as **-2ΔlnL**), which minimises the fitting process and finds the optimal values of each free variable. It is plotted as a function of the **N_signal** value, as this is the interesting variable. This value is basically the optimal result of the fitting process, and most likely value for the estimation of the value the analysis is in search of. In the following paragraphs, a few mathematical definitions will be presented.

Assuming that the parameter **θ** is the theoretical value one wishes to calculate, and **X** the data sample of experimentally measured results, it is desired to calculate an *estimator function* for **θ**:

$$\hat{\theta}=\hat{\theta}(X) \quad \text{(A.c)}$$

The **N** experimental measurements of the data sample **X**, **x**, will be regarded as a *random variable*, with a probability density of **p(x|θ)**, with **θ** being a set of **r** parameters (the theoretical values wished to be estimated). Assuming that the sample measurements are independent, the likelihood **L** of set **X** will be:

$$L(X|\vec{\theta})=p(\vec{x}_1|\vec{\theta})\,p(\vec{x}_2|\vec{\theta})\ldots p(\vec{x}_N|\vec{\theta}) \quad \text{(A.d)}$$

The estimation of the values **θ** will be calculated by maximising **L**:

$$\hat{\vec{\theta}}=argmax_{\vec{\theta}}\,L(X|\vec{\theta}) \quad \text{(A.e)}$$

In practice, the *log-likelihood* function is calculated, as:

$$lnL(X|\vec{\theta})=\sum_{i=1}^{N}\ln[\,p(\vec{x}_i|\vec{\theta})] \quad \text{(A.f)}$$

and then it is minimised as:

$$\frac{\partial lnL}{\partial\vec{\theta}}_{\vec{\theta}=\hat{\vec{\theta}}}=0 \quad \text{(A.g)}$$

The calculation of the log-likelihood in this way makes calculations easier, as now additions instead of multiplications will be present, and also the result is unaltered, as the **ln** function is monotonous and has the same minimum and maximum values as its argument.[i]

The fitting method used later on will calculate and graph the *negative log-likelihood* function multiplied by 2, and one will be able to see where it is minimised.

If the calculated and graphed log-likelihood function is taken through a reversal process, so as to calculate the original likelihood function **L**, one can estimate the *Bayesian Upper Limit* (BUL) .

The goal is that of calculating a range in which the theoretical value that is being searched for belongs to, as a range is a "better result" than a single value. The *Bayesian* approach, as opposed to the *frequentist* approach, treats the unknown value **θ** as a random variable.

Source: Kousouris 2023, p. 5.24

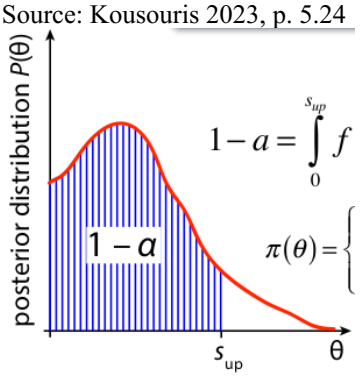


Figure 17: Bayesian approach for BUL

Assuming now that the unknown variable looked for is now labelled **μ**, in the case of this analysis the **N_signal**, the posterior probability **p(μ|x)**, will be given by the following formula based on the Bayesian approach:

$$p(\mu|x)=\frac{p(x|\mu)\,\pi(\mu)}{p(x)} \quad \text{(A.h)}$$

where **π(μ)** is called the *prior probability*, and expresses any knowledge about the theoretical value known beforehand. In the analysis later on, a flat prior with **π(μ)=1** will be assumed. As it wishes to calculate the range in which **μ** will be found with a certain probability level, named *credibility* **α**, the following equation can be solved to calculate the upper limit μ_{up}:

$$\int_0^{\mu_{up}} p(\mu)\,d\mu=(1-\alpha)\int_0^{\infty} p(\mu)\,d\mu \quad \text{(A.i)}$$

The **p(μ)** will be given by the "reversed" **-2ΔlnL** function:

$$L=p(x|\mu)=e^{-2\Delta lnL} \quad \text{(A.j)}$$

In this case, the **x** variable will be **N_signal**, and μ_{up} will be the upper limit sought for.

i Kousouris 2023, p. 2, 4, 6

V. ANALYSIS OF MC SIMULATIONS AND EXPERIMENTAL DATA

As has been discussed above, the elusive Z'_{TC2} particle during the current search theoretically decays to to a top – antitop quark pair ($t\bar{t}$), based on the model used in this thesis. The objective is to search for a "lump" in the data that pertains to the discovery of a new boson. By searching for $t\bar{t}$ resonances in the mass spectrum ($m_{t\bar{t}}$), any signs of new physics will appear as a localised deviation ("lump") from the SM $t\bar{t}$ background, as it is naturally monotonously steeply falling. For that reason, events that contain $t\bar{t}$ pairs will be searched for, with one obvious background signal being the rest of particles produced within the detector, the SM *QCD background*. This QCD background will be named as *reducible*, as with necessary conditions it can be successfully suppressed; however, there is one more background signal that cannot be so easily ignored, namely the *irreducible* background, which will be the plethora of $t\bar{t}$ pairs created, but that do not originate from the Z'_{TC2}. As will be seen further on, a number of conditions, named *cuts,* will be used to distinguish the various backgrounds, and to find the optimal ones for this analysis.[i]

As this is an undergraduate thesis, I will use ready made MC simulation files from my colleagues, which I will analyse using the ROOT framework. The two files in use will be one that simulates ttbar events, and that contains both $t\bar{t}$ events and QCD background, to first distinguish the reducible and irreducible backgrounds; and the second will be files that simulate Z'_{TC2} production, for various different masses and widths. These will be discussed later on in the relevant section.

V.i. Z'_{TC2} and $t\bar{t}$ background variable plots

Before diving into the analysis, I will present some preliminary histogram plots of some useful variable distributions. This will allow for a first insight into the objective of distinguishing the background from the signal and the variables that will be needed later on. The two MC files that will be used for this preliminary analysis are:

- "TT_TuneCUETP8M2T4_13TeV-powheg-pythia8.root", for simulating the $t\bar{t}$ background
- "ZprimeToTT_M3500_W35_TuneCP2_Psweights_13TeV-madgraph-pythiaMLM-pythia8_20UL.root", for simulating the Z' signal of mass 3500GeV/c^2 and width 1%.

i The information and analysis methods described in the following sections have been provided by the author's supervisor Dr. Kousouris, along with additional support and guidelines by her PhD doctoral student colleagues, Eirini Siamarkou and Theodoros Chatzistavrou. The author has tried to implement this information as accurately as possible, but there may be inaccuracies. When the source is other than mentioned here, it will be cited separately.

The exposition of histograms will begin with the scalar quantities, namely: dPhiJJ (the angle between the two jets produced), met (the missing transverse momentum during each event), nJets (the number of jets produced during the event), and nLeptons (the number of leptons produced in the event).

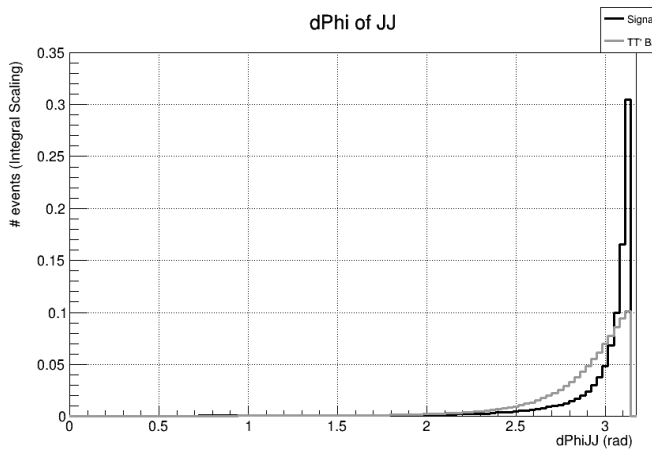


Figure 19: Histogram of dPhiJJ for the Z'_{TC2} signal and the $t\bar{t}$ background

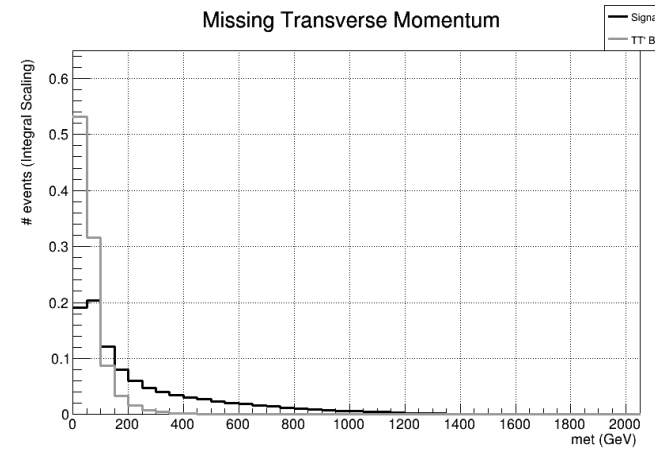


Figure 18: Histogram of met for the Z'_{TC2} signal and the $t\bar{t}$ background

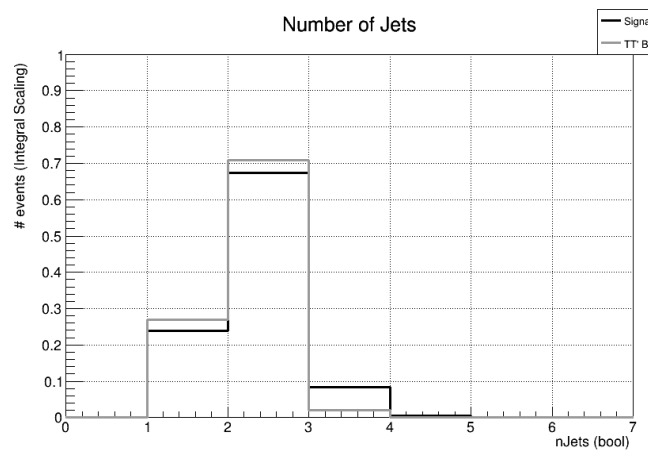


Figure 21: Histogram of nJets for the Z'_{TC2} signal and the $t\bar{t}$ background

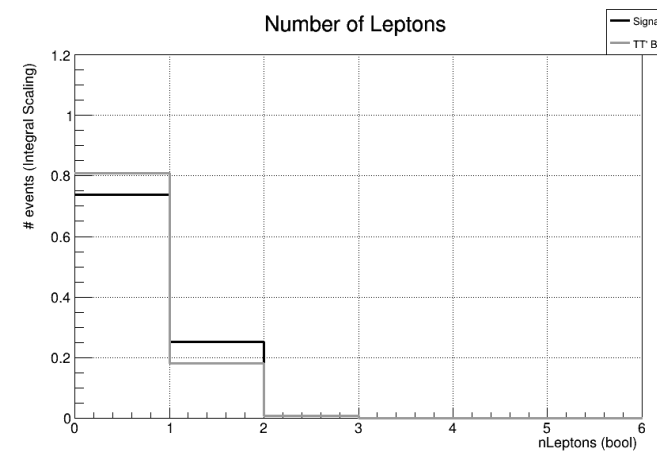


Figure 20: Histogram of nLeptons for the Z'_{TC2} signal and the $t\bar{t}$ background

The dPhi variable peaks mostly at pi, as is logical because in the centre of mass (CM) rest frame the top and antitop jets will be produced "back-to-back" in a *pencil-like* formation. The missing energy transfer variable is the missing energy that the detector had not measured, and usually refers to neutrinos or errors due to the resolution of the detector. In the present analysis there are no neutrinos, so it naturally peaks at zero. The nLeptons variable shows the leptons in the final state which, as has been seen, appears in two possible final states of a $t\bar{t}$ pair. Finally, in the analysis two jets are required, which is the maximum number shown in the histogram.

The jet quantities split into the leading (denoted as [0]) and subleading (denoted as [1]) jets, are: **jetEta** (pseudorapidity of the jet), **jetMass** (the mass of one reconstructed jet), **jetPhi** (the small angle between one jet and the φ axis) and **jetPt** (the transverse momentum of the jet).

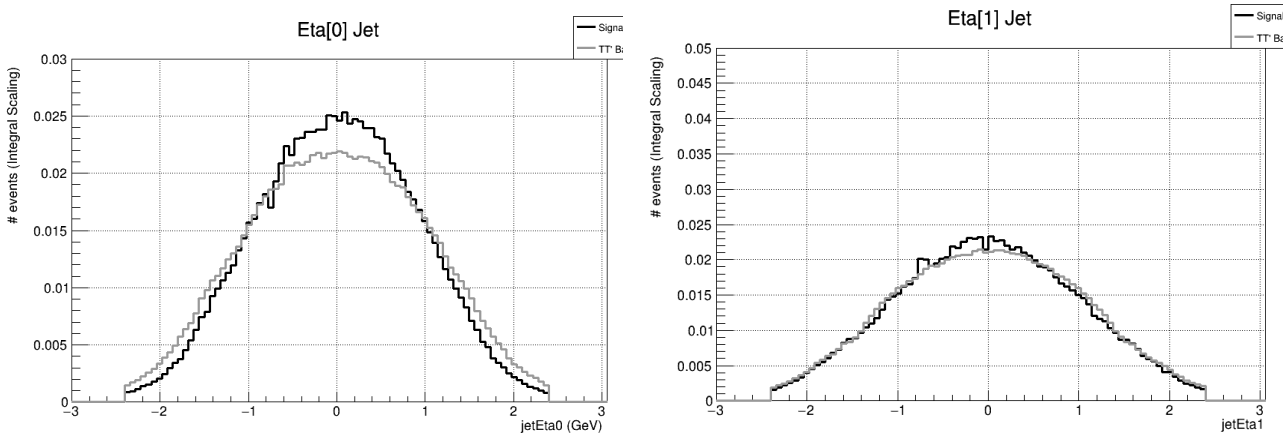


Figure 22: Histogram of jetEta for the Z'_{TC2} signal and the $t\bar{t}$ background separately for the leading and subleading jets.

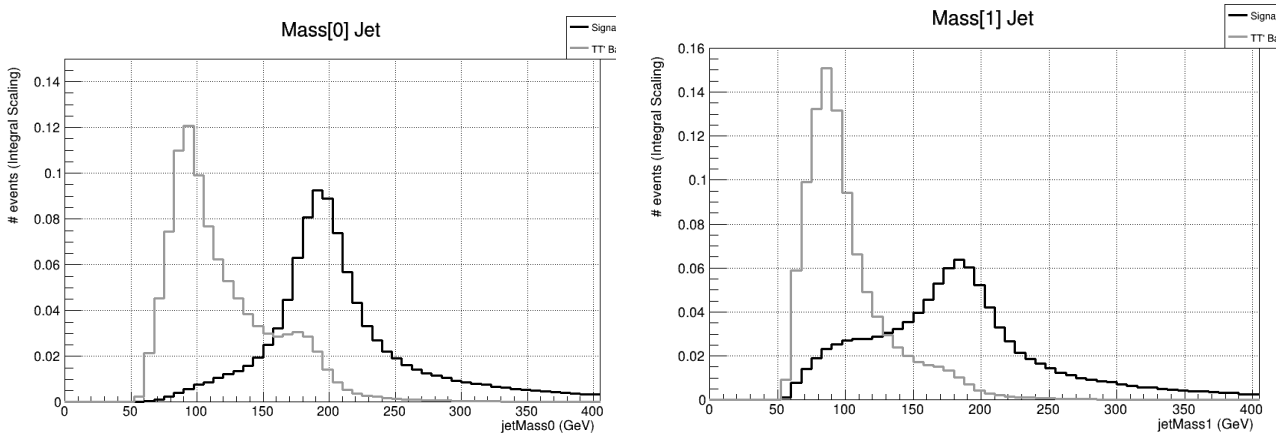


Figure 23: Histogram of jetMass for the Z'_{TC2} signal and the $t\bar{t}$ background separately for the leading and subleading jets.

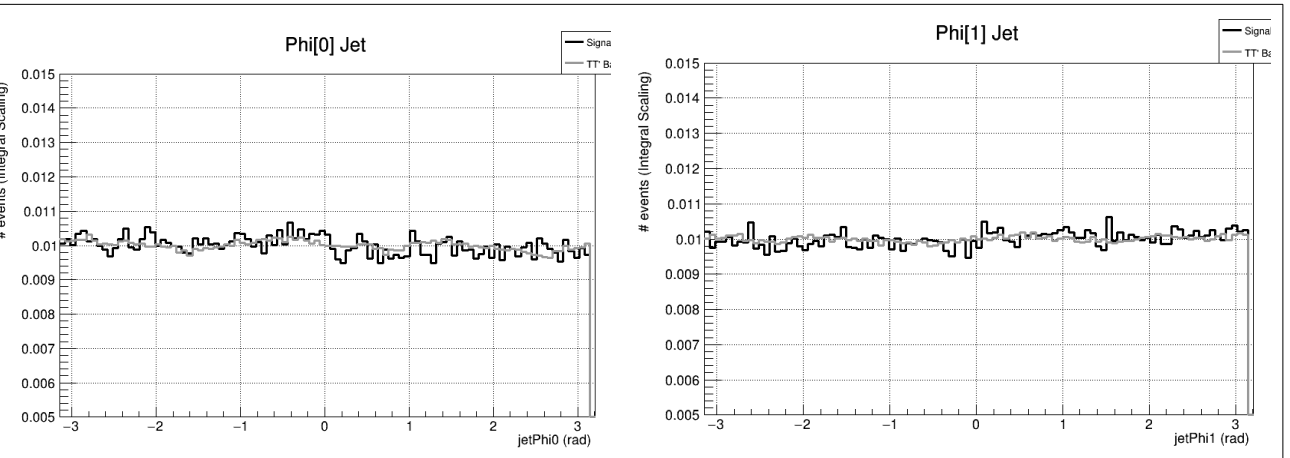


Figure 24: Histogram of jetPhi for the Z'_{TC2} signal and the $t\bar{t}$ background separately for the leading and subleading jets.

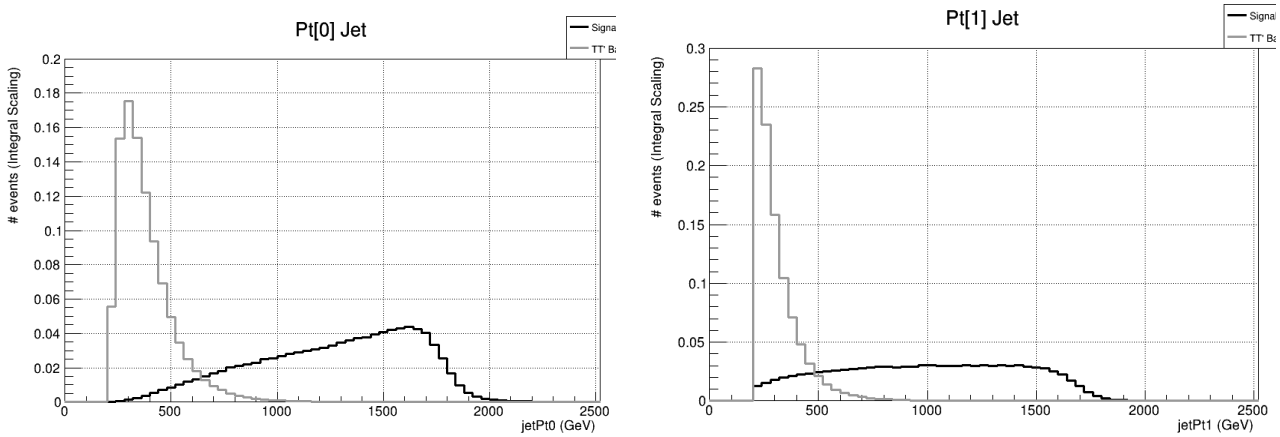


Figure 25: Histogram of jetPt for the Z'_{TC2} signal and the $t\bar{t}$ background separately for the leading and subleading jets.

As can been seen, the pseudorapidity variable spans from approximately -3 to 3 (allowing for some imperfections), and presents a smooth peak at 0, implying that the jets are produced perpendicular to the beam axis. The jetMass variables present peaks around the mass of the top quark, while as the earlier peak of the background around 90 GeV is "iconic", meaning that it is not a real peak, but a pileup of events due to an original cut applied. The same explanation for the "peak" applies to the jetPt histograms, while as the jetPhi variable appears relatively random.

Another jet quantity depicted as a histogram, but calculated by the author within the ROOT programme, is the absolute value of the cosine of the angle θ^* between the jet and the z axis in the CM rest frame. In the rest frame, the two jets produced, and as they are produced as a two-body decay, they should have a pencil-like formation, i.e. be produced exactly "back-to-back". From QFT calculations, it is expected that the $t\bar{t}$ process should make large angles with the z axis (so a small cosine), while as the Z'_{TC2} should exhibit small angles (cosine near 1), as can be seen in Figure 26.

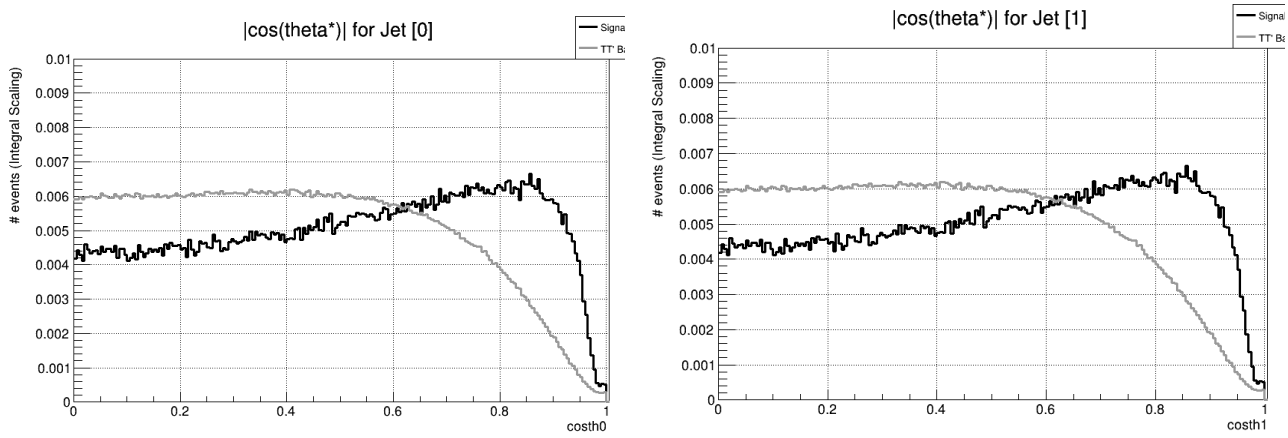


Figure 26: Histogram of jet-abs(cosθ*) for the Z'_{TC2} signal and the $t\bar{t}$ background separately for the leading and subleading jets.

Finally, I present the histograms for the invariant mass of the two jets together, mJJ. This distribution for the Z'_{TC2} will have a peak at the mass of the Z'_{TC2}, as determined by the mathe hypothesis of the MC file used at the time. The $t\bar{t}$ background presents a steeply falling distribution. The variable mTTbarparton is also plotted for the signal Z'_{TC2}, which is a MC constructed variable, aimed to simulate the mass of the partons produced, i.e. the top and anti-top quarks. It is not a variable that is measured in the detector, and theoretically it ought to be equal to the mJJ variable. They are shown in Figure 27.

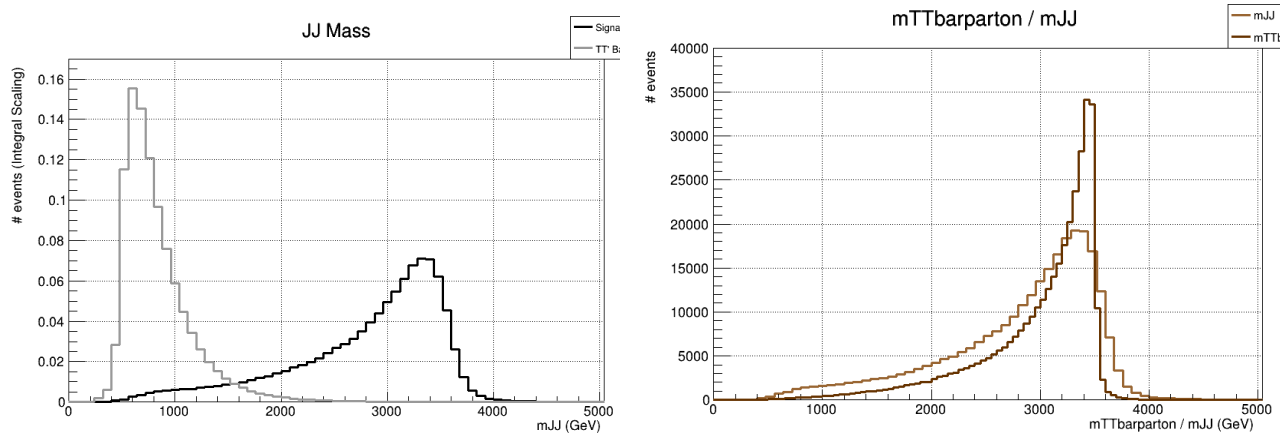


Figure 27: Histogram of mJJ for the Z'_{TC2} signal and the $t\bar{t}$ background, and an overlay of the mJJ and mTTbarparton variables for the Z'_{TC2} signal.

V.ii. Top Tagger Development

The main analysis will begin with searching for $t\bar{t}$ pair events in the detector. As $t\bar{t}$ events are required for the pinning down of a possible Z'_{TC2} boson, the main analysis begins with procuring the correct cuts for purely $t\bar{t}$ events. In this analysis, therefore, the $t\bar{t}$ will count as the "signal", whereas the "rest" of the events will be characterised as QCD processes and basically of no use, as the final objective is to have a $t\bar{t}$ enriched final state. This section will therefore be dedicated to developing a *Top Tagger*, a mechanism to distinguish top jets (so the analysis will be per jet) from their QCD background.

As discussed in Chapter IV.i.b, jet reconstruction when working with $t\bar{t}$ pair states is easier and clearer in the phase space of high transverse momentum, p_t, to obtain high p_t jets. Therefore, the analysis is improved by applying the following p_t cuts for all quantities:

$$p_t \, cuts: \quad p_t > 400 \;\; GeV/c \qquad (5.I)$$

In Figures 28 and 29 the $\mathbf{p_t}$ inclusive histograms (for the leading and subleading jets together) are shown with the pre $\mathbf{p_t}$ and post cuts. As the jets have been reconstructed and are stored as information within the root file, first the following quantity is calculated:

$$DR = \sqrt{D\varphi^2 + D\eta^2} \qquad \text{(A.k)}$$

where **DR** shall be the angular separation (the distance calculated between the centre of the reconstructed jet) (imagined as a cone shape) and the originating top parton, in this case a top or antitop quark in the φ - η plane.

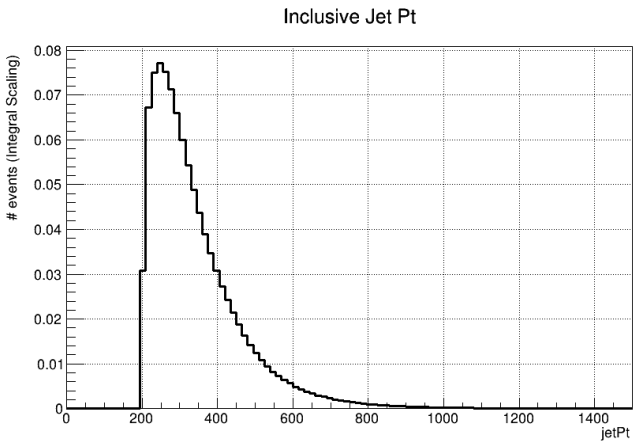


Figure 29: Inclusive histogram of the transverse momentum p_t, including both jets, scaled to unity.

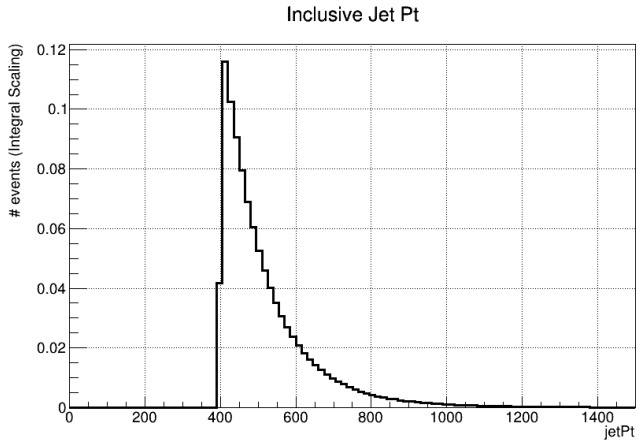


Figure 28: Inclusive histogram of the transverse momentum p_t, including both jets, with the p_t cuts $p > 400$, scaled to unity.

The process is as follows: looping over all reconstructed jets, the quantity **DR** is computed per event, and a histogram is filled with the *minimum* angular distance **DRmin**. By minimising this distance, it will be possible to basically observe which top quarks have been determined by the

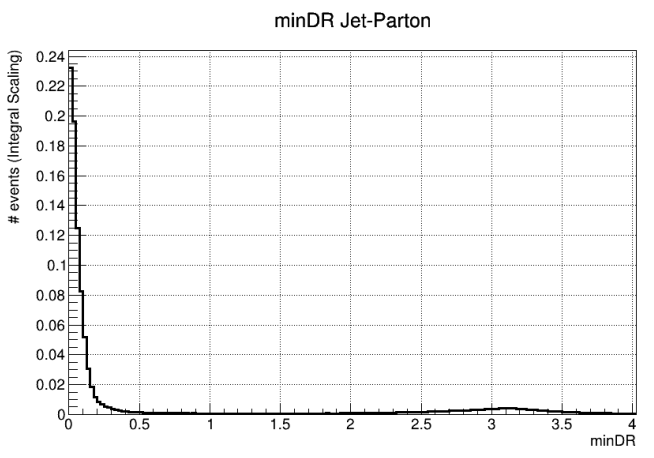


Figure 31: Histogram of the angular separation DRmin between the reconstructed jet and the parton (anti-parton). To the left is the original diagram (A)

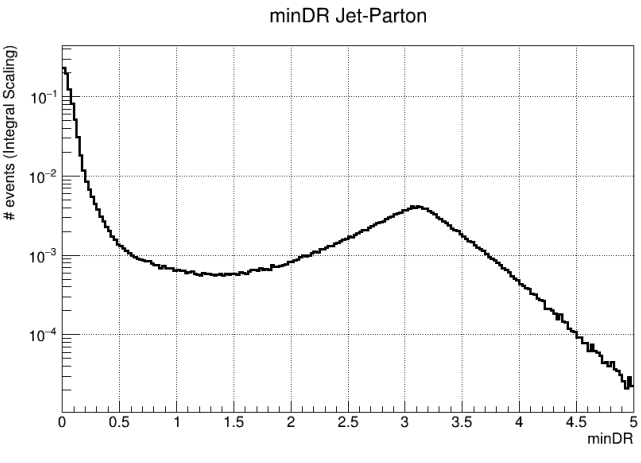


Figure 30: Histogram of the angular separation Drmin between the reconstructed jet and the parton (anti-parton). The right histogram depicts it in logarithmic scale (B), both scaled to unity

trigger level as within the jet or not. The events that will be characterised as within the jets will be referred to as *matched jets* (top jets), while the events well without the jet shall be *non-matched jets* (namely the QCD background).

The histograms of **DRmin** in normal and logarithmic scale are shown in Figures 31 and 30. As can be seen from the regular plot, there is a distinct drop in number of events after about **DRmin=0.5**, which will also determine the first cut condition. As the purpose of this diagram was to find top quarks matched to the reconstructed jets, it follows that the maximum of events will be those. Therefore the tail following the sudden drop in events will be the remaining QCD background. By observing the curve better in the logarithmic scale, the first cuts are determined:

$$Matched\ jets: \quad DRmin < 0.5 \qquad (5.II.i)$$

$$Non-matched\ jets: \quad DRmin > 2 \qquad (5.II.ii)$$

Following the 5.II cuts, the subjettiness variables τ_{32} ($=\tau_3/\tau_2$) and τ_{31} ($=\tau_3/\tau_1$) for matched and non-matched jets are plotted, to discern which of the two is best for distinguishing between $t\bar{t}$ and QCD signals, first without the p_t cuts. They can be seen in Figure 33 and Figure 32, and also in Figures 35 and 34, as p_t binned (clearly without the $p_t > 400$ cuts).

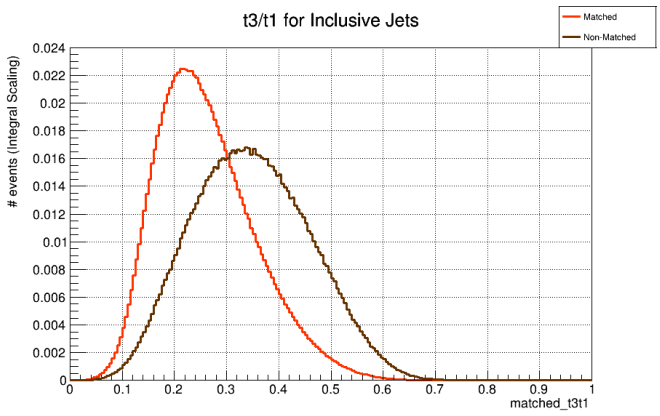


Figure 32: The subjettiness variable τ_{31} plotted for inclusive jets for matched jets overlaid with non-matched jets, with no added cuts.

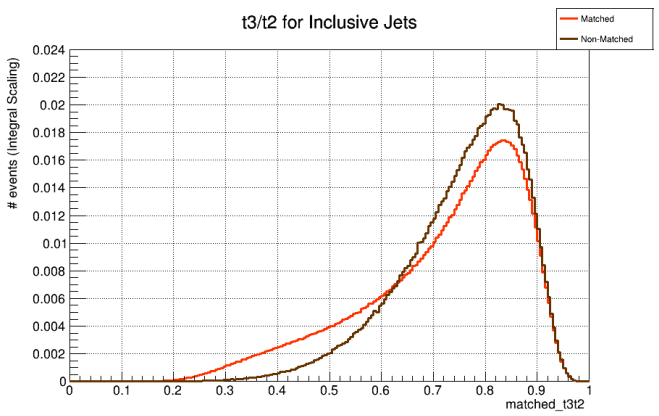


Figure 33: The subjettiness variable τ_{32} plotted for inclusive jets for matched jets overlaid with non-matched jets, with no added cuts.

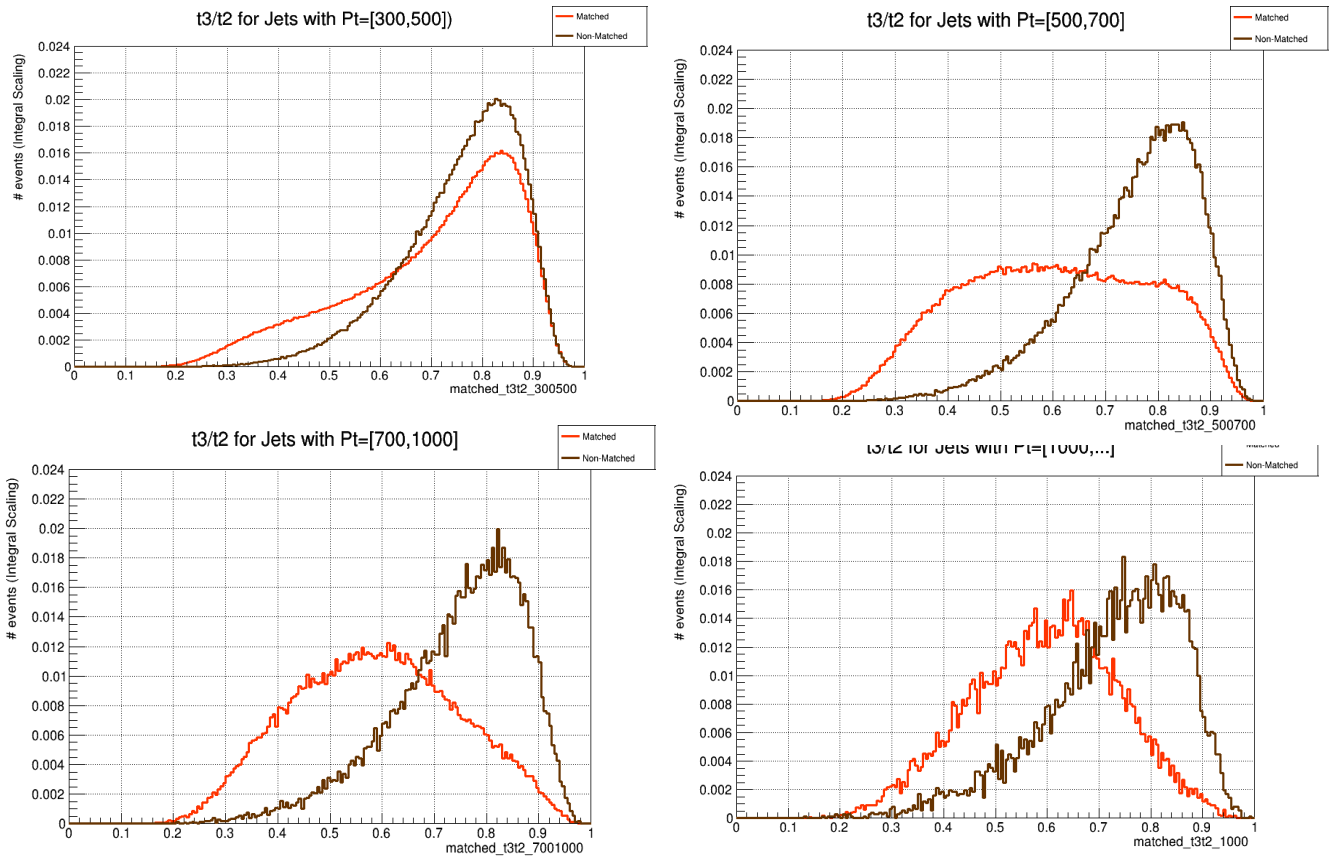


Figure 34: The p_t binned subjettiness variable τ_{32} plotted for inclusive jets for matched jets overlaid with non-matched jets.

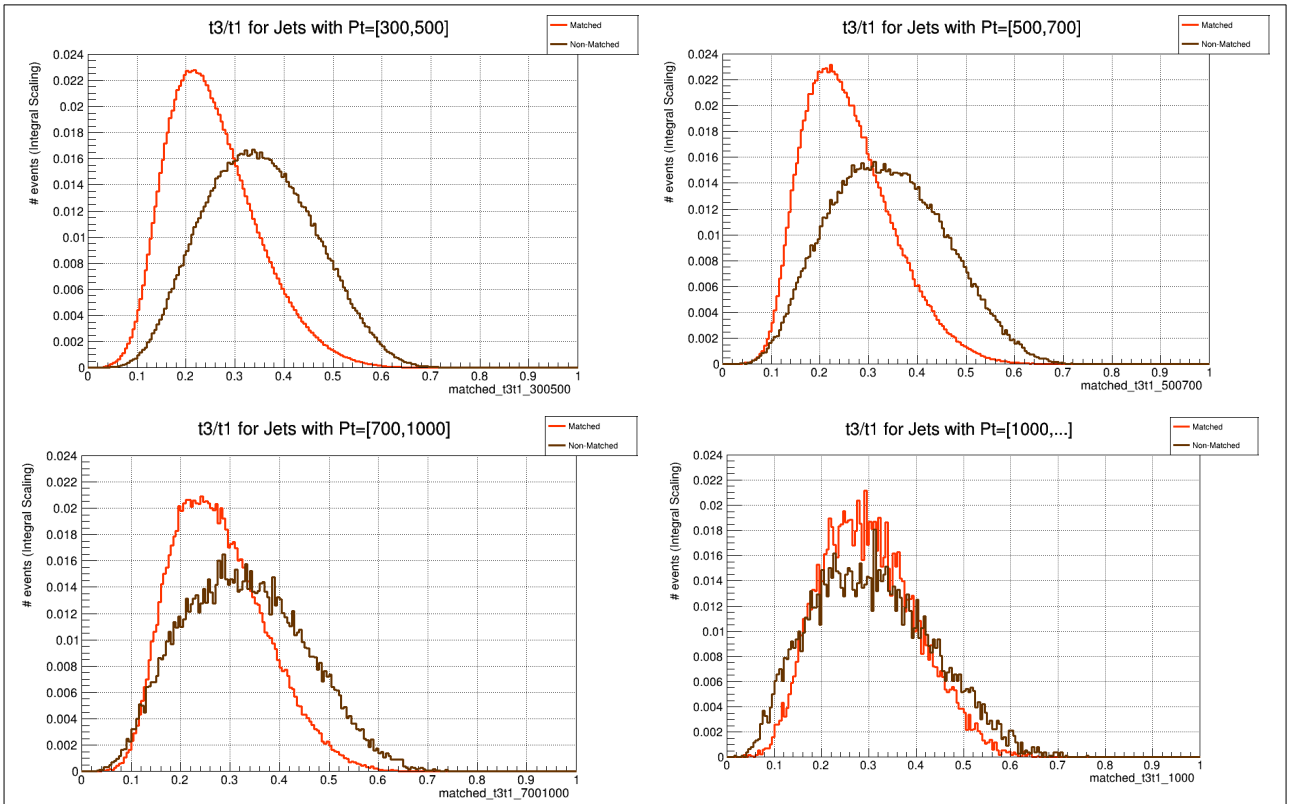


Figure 35: The p_t binned subjettiness variable τ_{31} plotted for inclusive jets for matched jets overlaid with non-matched jets.

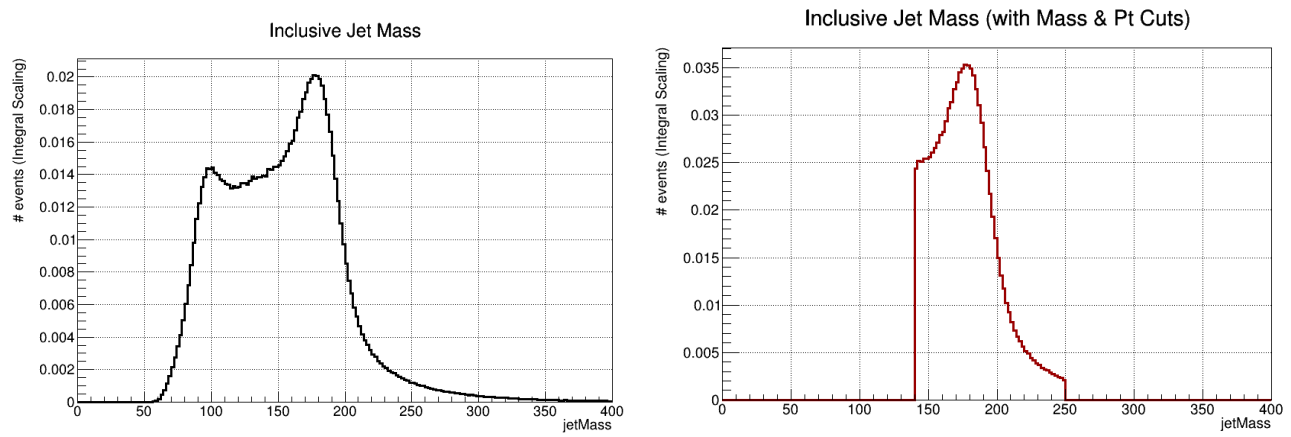


Figure 36: Plotting of the jetMass histogram for inclusive jets, without any initial cuts (left) and then with the p_t>400 cut and mass cuts 140<jetMass<250.

Now, with the $\mathbf{p_t}$ cuts in place, another one is applied, regarding the mass of the top jet. By observing the jetMass variable peaks (Figure 36), the mass cut is originally chosen as:

$$Top\,mass\,cuts: \quad 140 < jetMass < 250 \quad (5.II.iii)$$

By applying the top mass cuts as well as the $\mathbf{p_t}$ cuts, the subjettiness variables are plotted again (Figure 37):

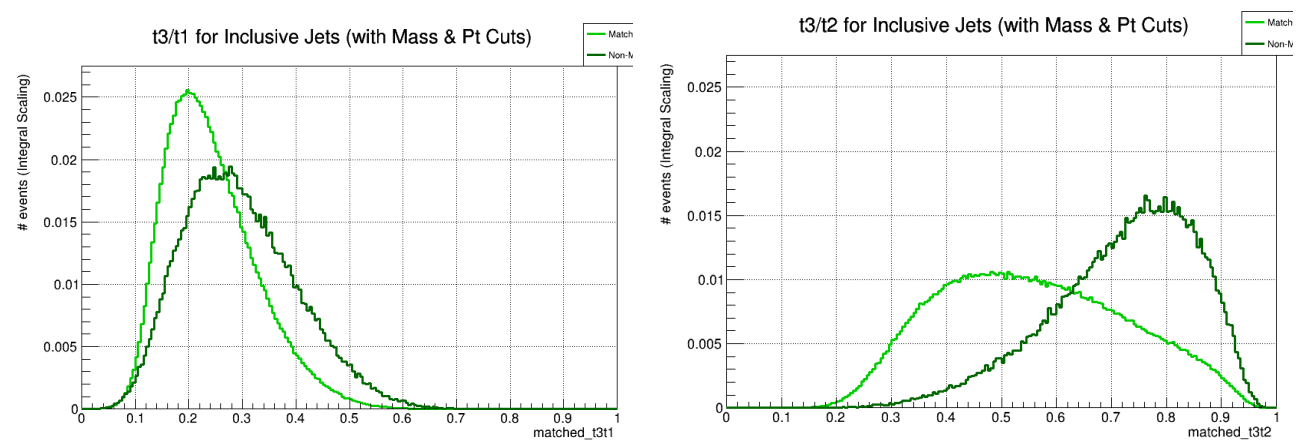


Figure 37: The subjettiness variables τ_{31} and τ_{32} plotted for inclusive jets for matched jets overlaid with non-matched jets, with p_t>400 and 140<jetMass<250 cuts..

Observing the above histograms, it becomes clear that the $\boldsymbol{\tau_{31}}$ subjettiness variable is the best to distinguish the matched from the non-matched jets. From now on, it will be the only subjettiness variable used.

The 2D histograms (Figure 38) of **jetMass vs. τ_{31}** and **jetMass vs. τ_{32}** are also plotted, with and without the mass cuts, separately for matched and non-matched jets.

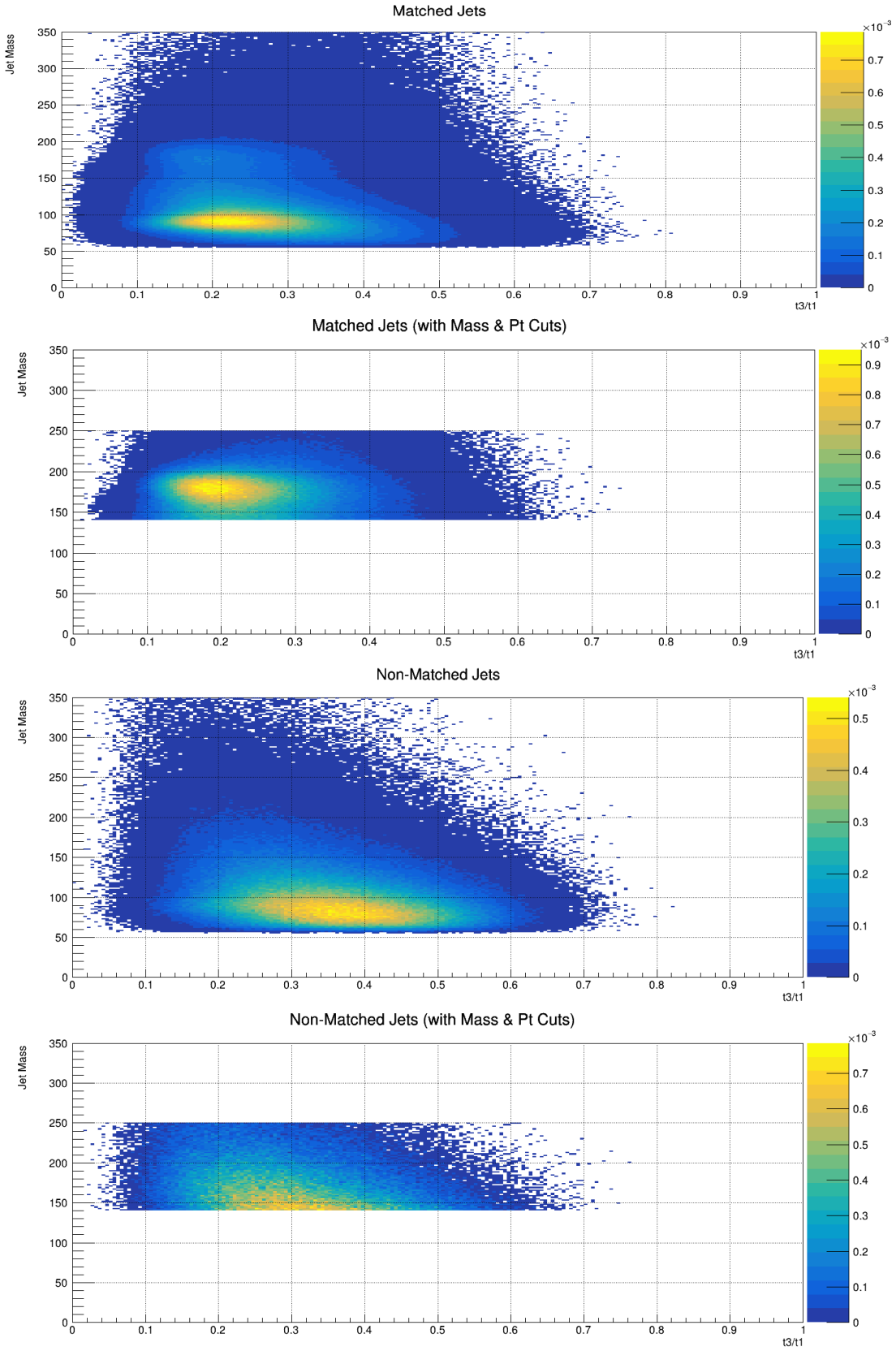


Figure 38: 2D histogram of the jetMass vs. the subjettiness variable τ_{31}, for matched and non-matched jets, with and without the mass cuts

It is observed that the matched jets contain the exact peak for jetMass, as is required, and shows that the analysis before this stage had been done correctly. Specifically, the “yellow peak” is collected and is more prominent when the mass cuts are applied, while on the contrary when the mass cuts are applied to the non-matched jets, it is much fainter as there is nearly no peak.

The mass cuts originally chosen are now joined by two more, so as to explore more possibilities. The subjettiness variable is plotted (Figure 39) for these different mass windows, however it does not change much, as expected, as it depends on the structure of the jet and not its mass.

Overall, these are the three mass windows considered:

$$Mass\ window\ A: 140 < jetMass < 250$$
$$Mass\ window\ B: 150 < jetMass < 230$$
$$Mass\ window\ C: 150 < jetMass < 210$$
$$(5.II.iv)$$

The histograms of the jetMass with the above cuts applied can be seen in Figure 40.

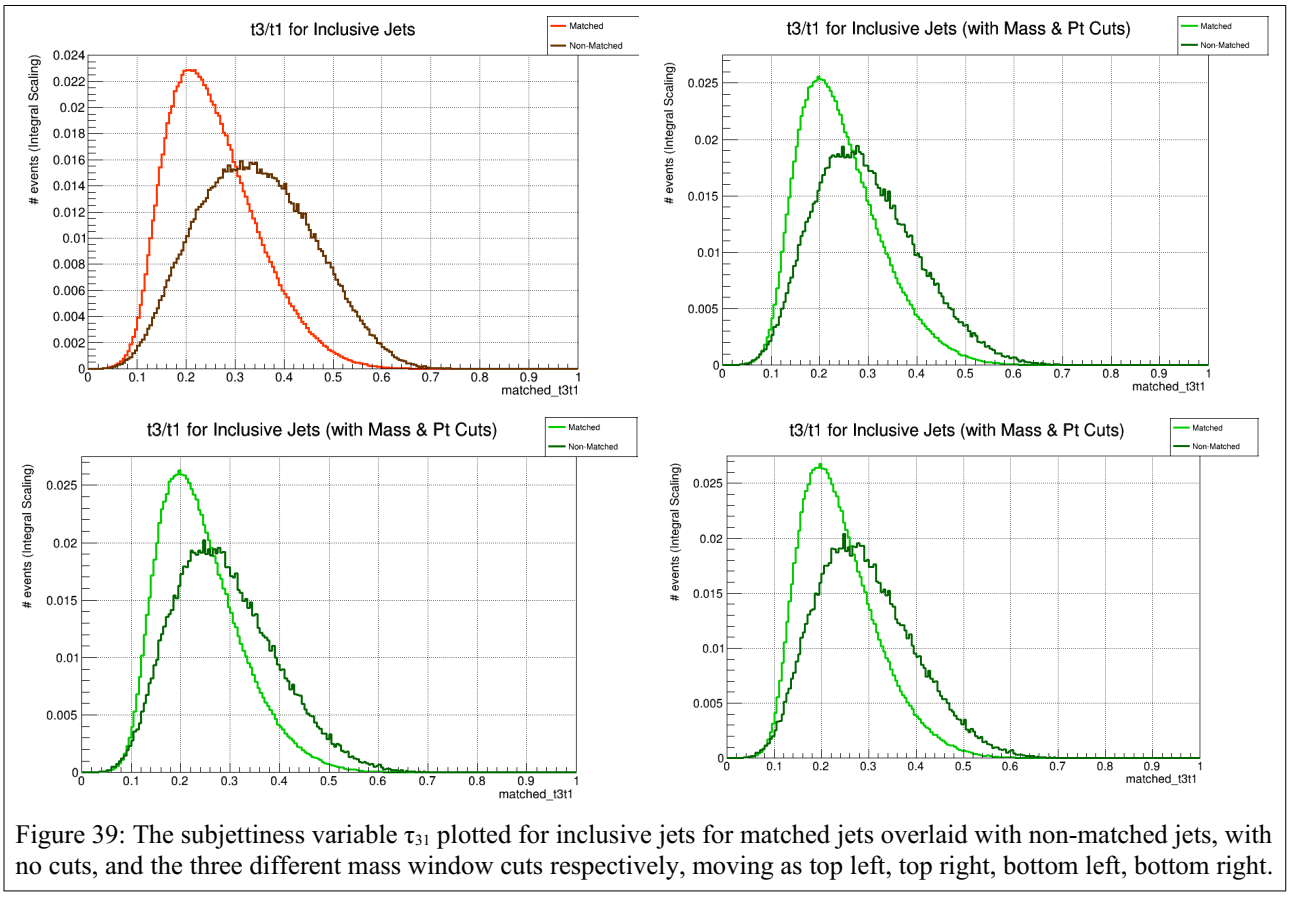


Figure 39: The subjettiness variable τ_{31} plotted for inclusive jets for matched jets overlaid with non-matched jets, with no cuts, and the three different mass window cuts respectively, moving as top left, top right, bottom left, bottom right.

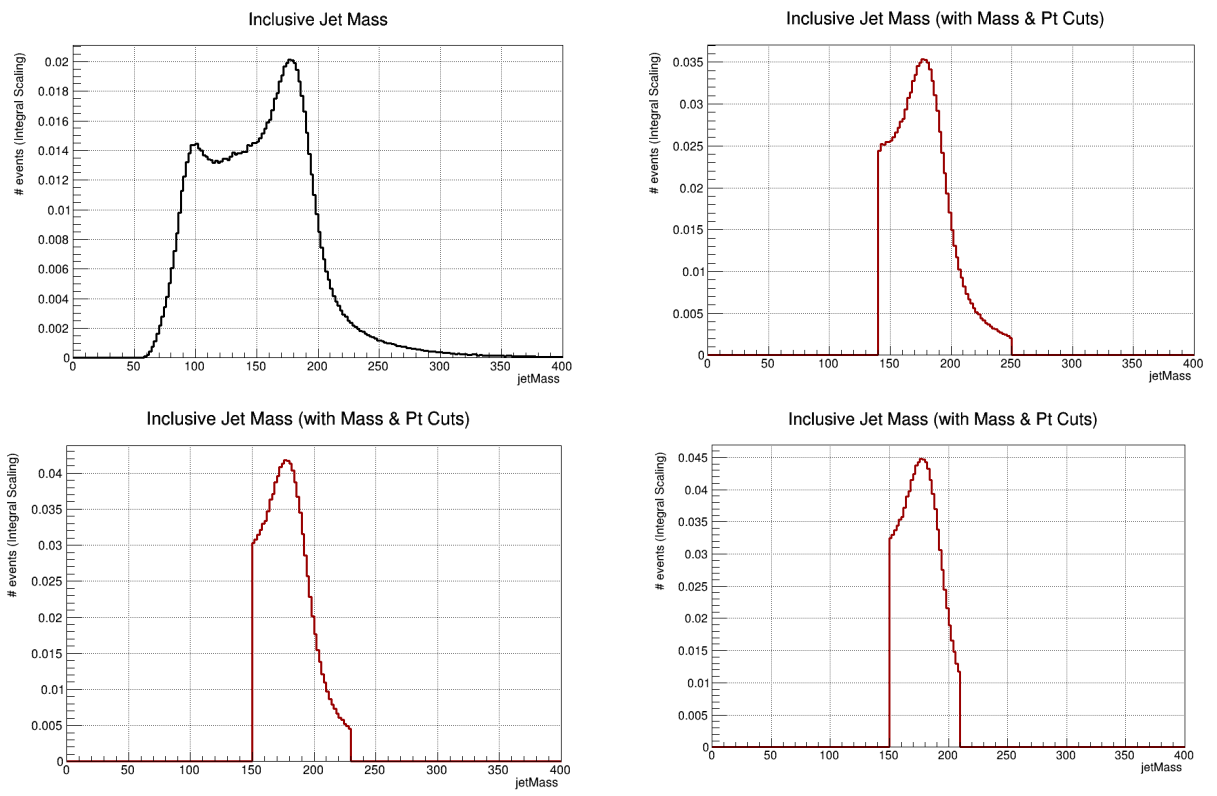


Figure 40: The histograms of jetMass for the different mass windows, with p_t cuts: all mass, window A, B and C

With these diagrams, the following quantity is calculated, the Top Quark Tagging Efficiency:

$$Efficiency = \frac{number\ of\ \ jets\ that\ \ pass\ mass\ cut + \tau_{31}\ cut\ with\ p_t > 400}{number\ of\ \ jets\ with\ p_t > 400} \quad \text{(A.l)}$$

The mass cuts will vary depending on which window has been chosen, A, B or C, while the $\mathbf{p_t}$ cuts are the same for all histograms from now on. The $\boldsymbol{\tau_{31}}$ is characterised as the integral from 0 up to different points **x** every time of the $\boldsymbol{\tau_{31}}$ histogram:

$$\tau_{31}\ cut: \quad \int_0^x \tau_{31} \quad \text{(5.II.v)}$$

The quantity is then calculated by calculating the numerator every time for different **x** points, and dividing every time by the integral of the whole histogram. For each **x**, which will scan all the bins in the histogram in this analysis, this variable is plotted, which results in the Efficiency diagram. The **y'y** axis will be the *fake rate* (percentage of non-matched jets), by integrating $\boldsymbol{\tau_{31}}$ for non-matched jets; the **x'x** axis will be the *top tagging rate* (percentage of successfully tagged matched (top) jets), by integrating the matched jets. The Efficiency can be seen in a normal and logarithmic scale in Figure 41.

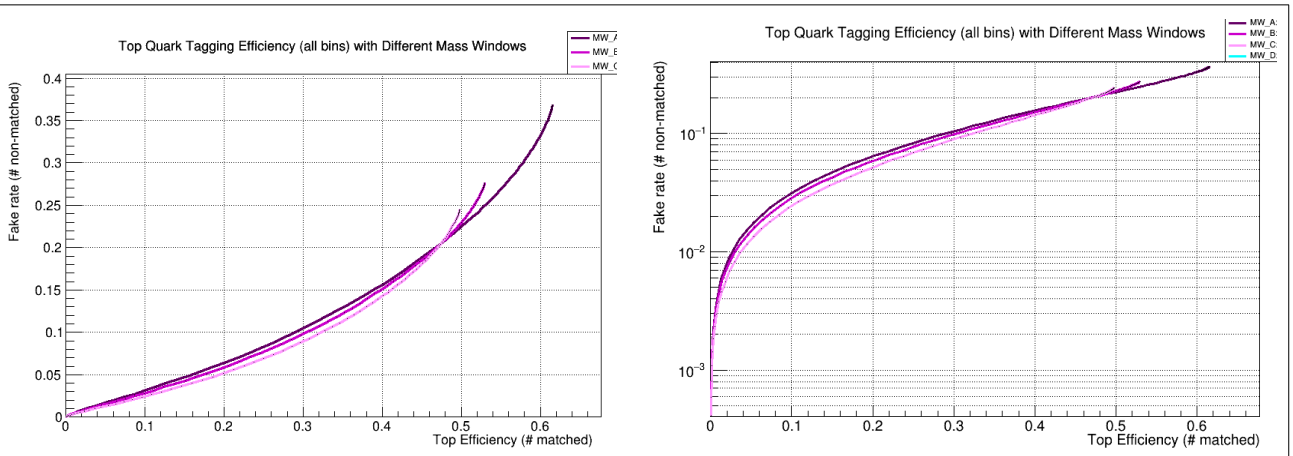


Figure 41: The Top Tagging Efficiency diagram plotted for three different mass windows, A, B and C.

One may observe that not all the curves reach unity, as may be implied by equation (A.l). This is logical, as the mass cut for each curve above has decreased the number of events that will be integrated, more and more as the mass window becomes smaller. So when compared together, they will not all have the same limits.

With the Efficiency diagram in hand, two *working points* are chosen, to continue the analysis:

$$\textit{Loose working point}: \quad \textit{Mass window cut } A[140,250] + 53\,\% \textit{ Top Tagging Efficiency} \left(\int \tau_{31} < 0.53 \right)$$

$$\textit{Tight working point}: \quad \textit{Mass window cut } C[150,210] + 30\,\% \textit{ Top Tagging Efficiency} \left(\int \tau_{31} < 0.30 \right)$$

(5.II.vi)

These points can be easily seen in the diagrams: the loose working point is for Top Efficiency equal to 0.53 and Fake Rate equal to approximately 0.25. Similarly, the tight point will have efficiency 0.30 and fake rate 0.08. It is clear from the very shape of the diagram that one could never each 100% efficiency with 0% fake rate, which would be the ideal case. No matter which points chosen, there will always be a percentage of "good signal" lost, and "bad signal" remaining. For this reason two points are chosen at this stage, to later make more tests to finally choose one final working point.

The mass cuts that relate to the two chosen working points can be seen in the previous Figure 40, while as the final $\boldsymbol{\tau_{31}}$ variable with loose and tight cuts can be seen in Figure 42 and Figure 43 for matched and non-matched jets.

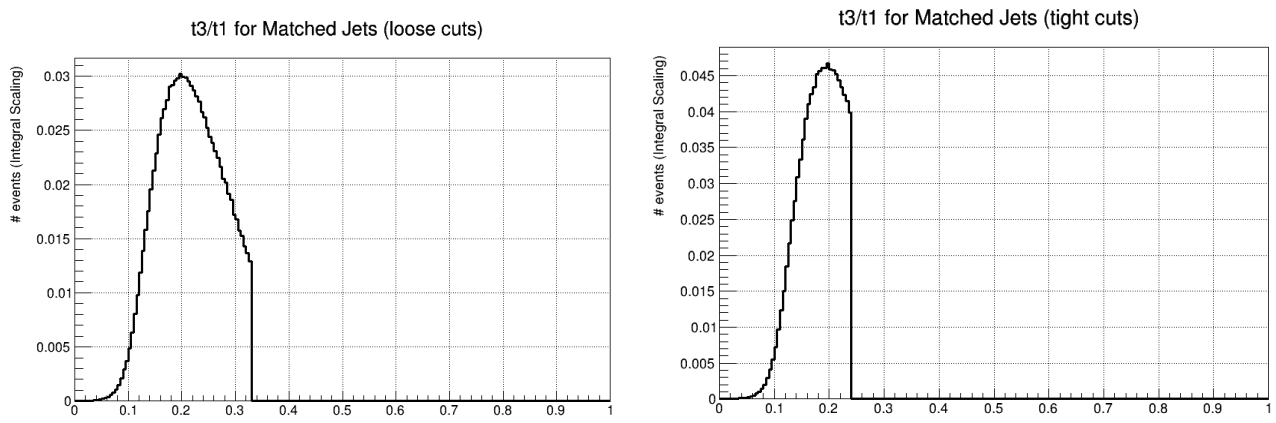


Figure 42: The subjettiness variable τ_{31} with the tight and loose working point cuts applied for matched jets.

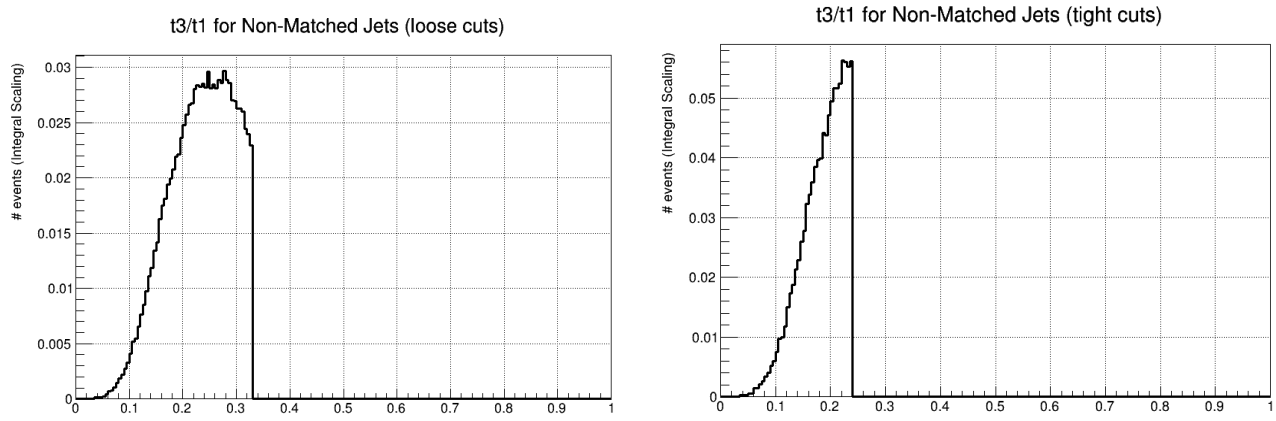


Figure 43: The subjettiness variable τ_{31} with the tight and loose working point cuts applied for non-matched jets.

Continuing the analysis, the p_t *efficiency* diagram is plotted, separately for matched jets, non-matched, for loose and tight working points, giving four graphs in total. This efficiency is created by dividing the p_t histograms with the loose or tight cut conditions applied, divided by the total p_t histogram with only the p_t cuts. It is a way of monitoring the efficiency as a function of the transverse momentum, p_t. The p_t diagrams, with only the p_t cuts, the tight and loose cuts, for

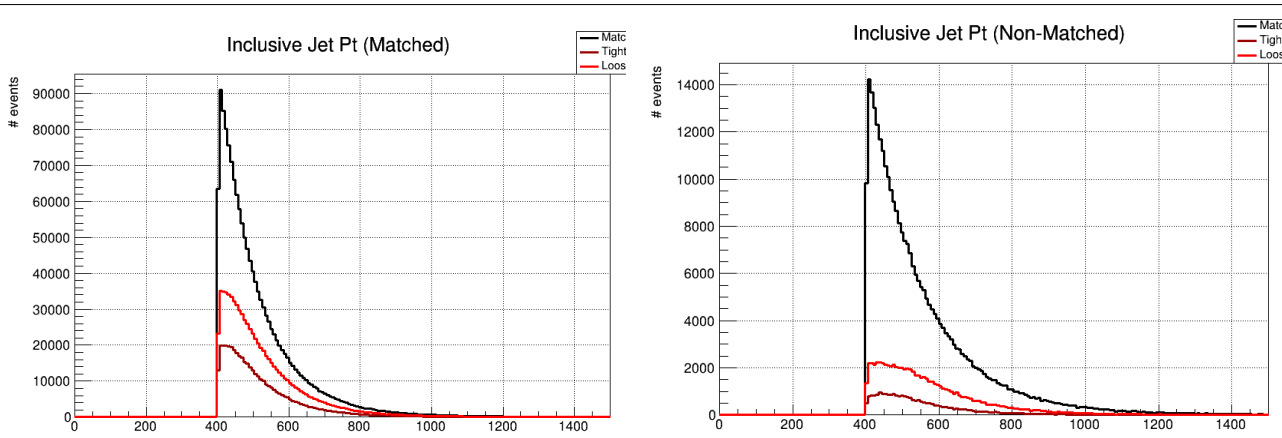


Figure 44: The p_t histograms with the following cuts: p_t only, loose, and tight. They are not normalised.

matched and non-matched jets can be seen in Figure 44, while as the $\mathbf{p_t}$ efficiencies can be seen in Figure 45, where the loose and tight cuts are overlaid together in the same canvas for matched and non-matched jets.

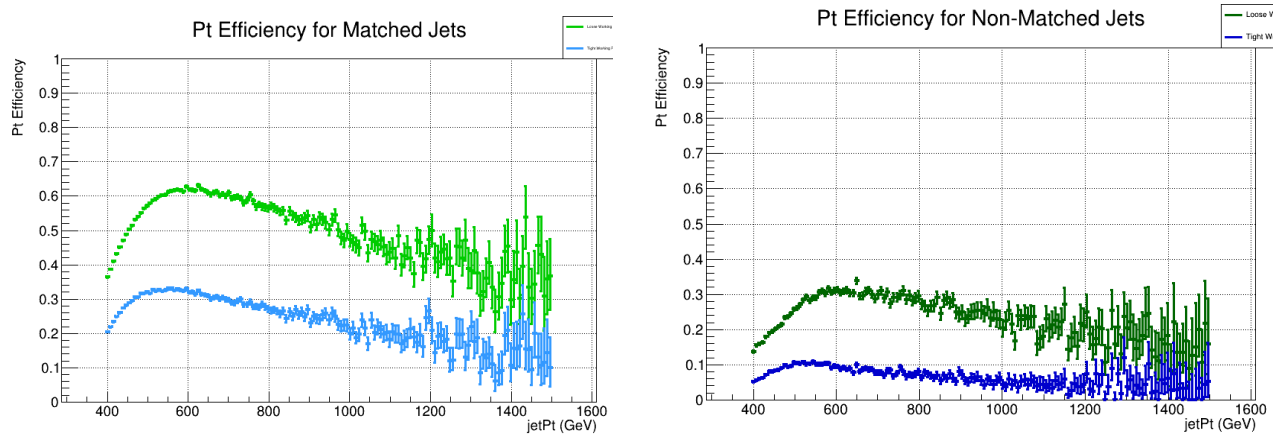


Figure 45: The p_t efficiencies plotted for matched and non-matched jets. The green curves represent the efficiency with the loose working points, while as the blue utilise the tight working point cuts.

The loose cuts clearly have a better efficiency, both for tagging matched jets and for non-matched jets.

The 2D histogram with only $\mathbf{p_t}$ cuts of the jetPt vs. jetMass is also plotted in Figure 46 for matched jets.

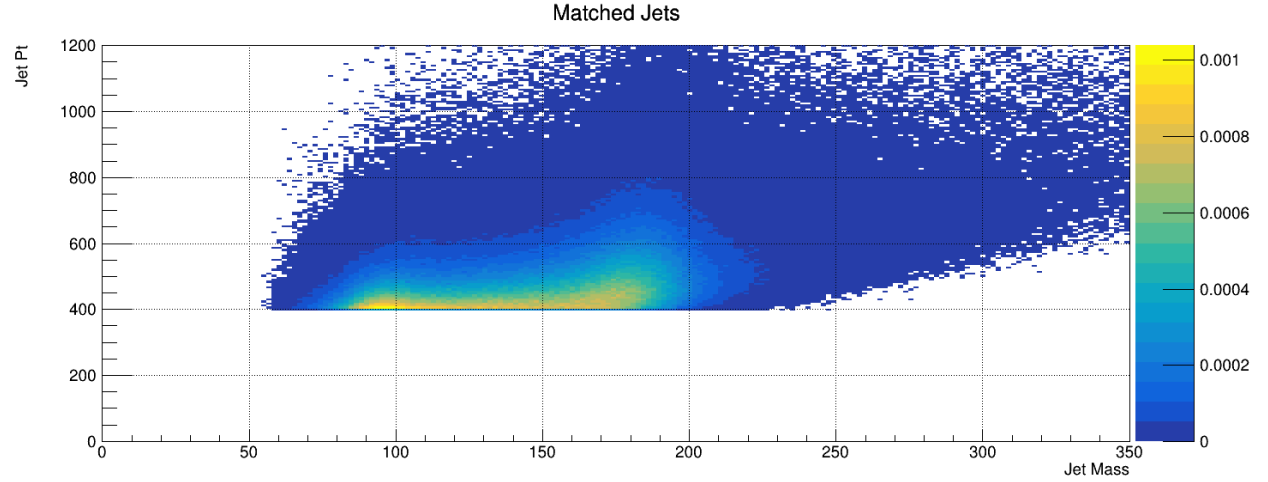


Figure 46: 2D histogram of the jetPt vs. jetMass, for matched jets, with only p_t cuts.

So far, two elements for tagging top quark jets have been utilised: referring to the mass of the jet, and the substructure geometry of the jet. CMS provides one more tagging element automatically within the ROOT file, which recognises the presence of a bottom quark within a jet. As mentioned in chapter III.iv, the top quark decays into a $\mathbf{W}$ boson and a bottom quark. By utilising the ready

variable jetNBSubDCSV and setting it equal to 1, the cuts become even more restrictive, and hopefully more accurately tag a top jet. The following cut is consequently applied:

$$b\,cut: \quad jetNBSubDCSV = 1 \qquad (5.II.vi)$$

This variable, with the 5.II.vi cuts applied, and previous cuts, is shown in Figure 47.

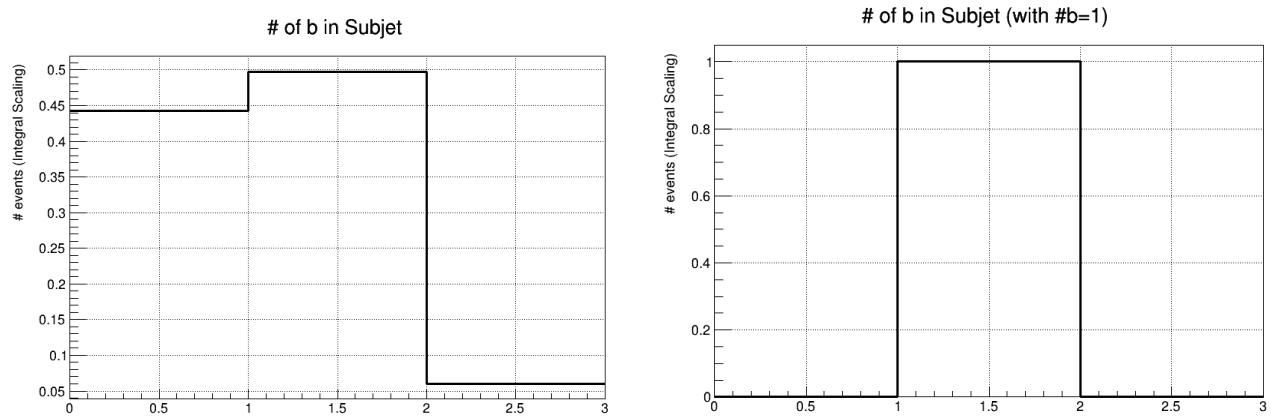


Figure 47: The jetNBSubDCSV histogram with only p_t cuts and with the condition that it is equal to 1 (with p_t cuts).

The Top Tagging Efficiency diagram is then replotted, overlaying the earlier curves with loose and tight cuts, with the same curves, but by applying cut 5.II.vi as well. In Figure 48 it is seen for two different plotting axes (zoomed in and out).The $\mathbf{p_t}$ efficiencies are also replotted, with the extra b cut condition in Figure 49.

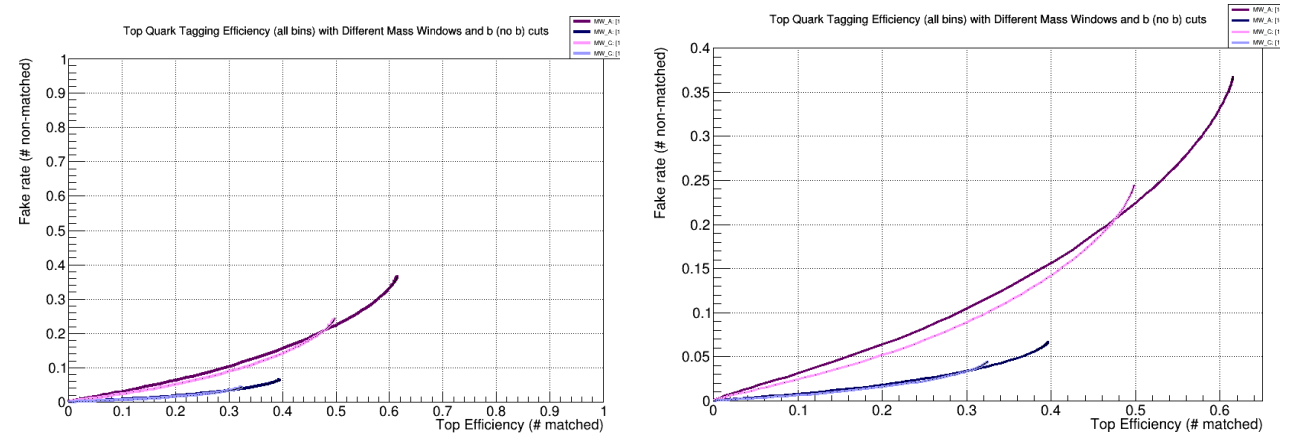


Figure 48: The Top Tagging Efficiency diagram plotted for the loose and tight working points, and for the loose and tight but the added condition of having a bottom quark subjet. To the left the axes are set to unity, while as to the right it is zoomed.

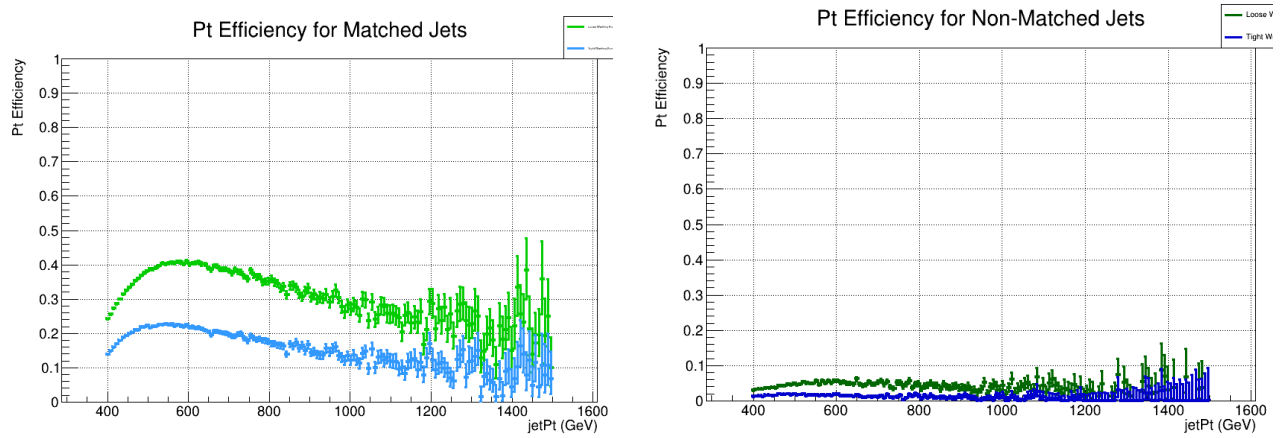


Figure 49: The p_t efficiencies plotted for matched and non-matched jets. The green curves represent the efficiency with the loose working points, while as the blue utilise the tight working point cuts. The extra condition regarding a b subjet is utilised.

It is observed that both efficiencies drop compared to the case without the **b** cuts, but the efficiency for non-matched jets reduces radically. This is a very good sign, showing that the background is significantly reduced when using this new variable.

By inspecting Figure 48, two NEW working points are set:

$$\textit{NEW loose working point}: \quad \textit{Mass window cut } A[140,250] + 53\,\% \textit{ Top Tagging Efficiency}$$

$$\textit{NEW Tight working point}: \quad \textit{Mass window cut } A[140,250] + 30\,\% \textit{ Top Tagging Efficiency} + b \textit{ cuts} \qquad \text{(5.II.vii)}$$

In summary, the NEW loose cuts and NEW tight cuts basic difference is the utilisation of **b** cuts or not. The more subtle difference is the percentage of τ_{31} that is kept, while the mass window remains as that of A for both.

In the next section, the Sensitivity will be calculated for each new working point, and will be plotted for different mass values of the Z'_{TC2} which this thesis is in search of. From now on, the NEW loose and tight working points will be simply known as *loose* and *tight* working points.

a. Sensitivity diagram

With the two types of cuts in hand, the loose and tight working points:

$$Loose\ working\ point: \quad Mass\ window\ cut\ A[140,250] + 53\,\%\ Top\ Tagging\ Efficiency$$

$$Tight\ working\ point: \quad Mass\ window\ cut\ A[140,250] + 30\,\%\ Top\ Tagging\ Efficiency + b\ cuts$$

(5II.vii)

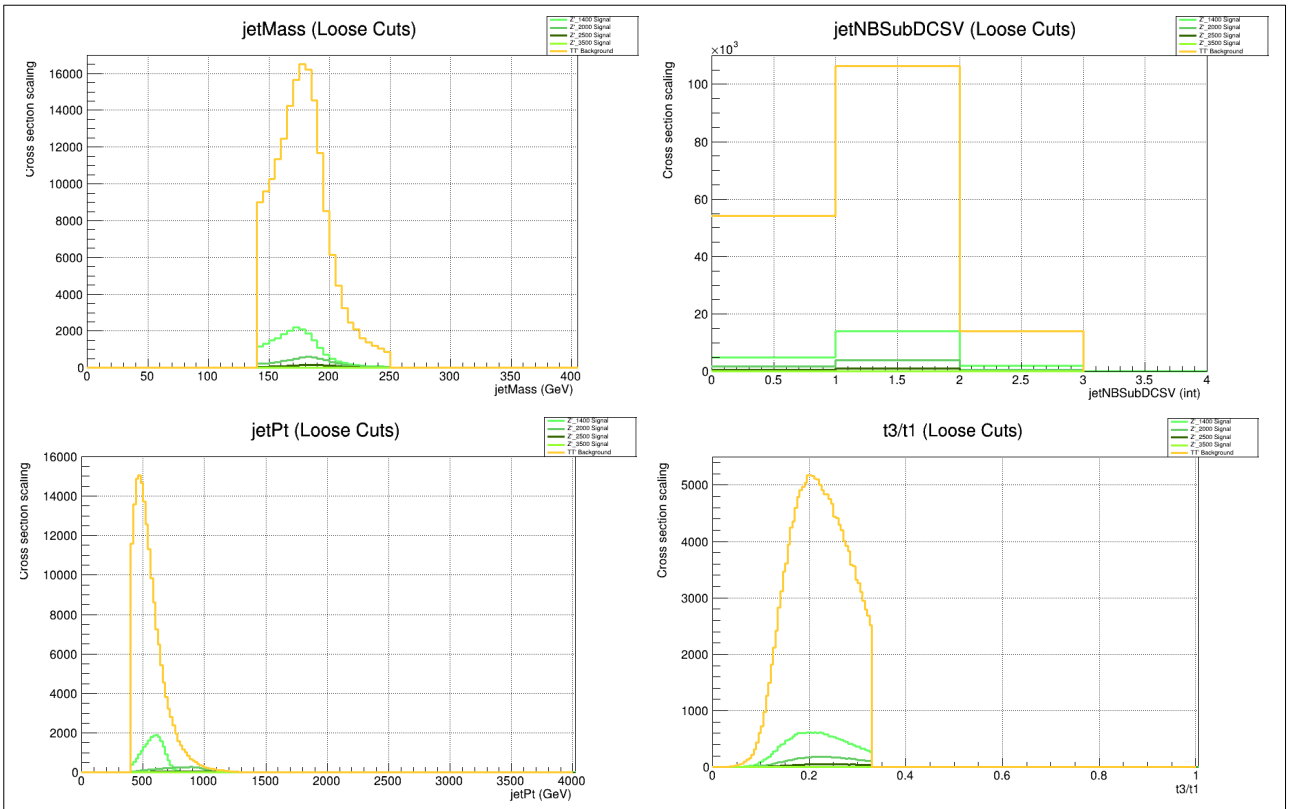


Figure 50: The histograms of jetMass, jetNBSubDCSV, jetPt and τ_{31} with loose cuts for the background and four Z'_{TC2} signals.

of course still retaining the high $\mathbf{p_t}$ phase space:

$$p_t\ cuts: \quad p_t > 400 \ \ GeV/c \qquad (5.I)$$

The final variable in the Top Tagger analysis will be calculated, this time as a function of the mass of the real signal, the $\mathbf{Z'_{TC2}}$. The different masses for the $\mathbf{Z'_{TC2}}$ will be procured from the following ROOT files:

- "ZprimeToTT_M1400_W14_TuneCP2_Psweights_13TeV-madgraph-pythiaMLM-pythia8_20UL.root", for simulating the Z' signal of mass 1400GeV/c^2 and width 1%.
- "ZprimeToTT_M2000_W20_TuneCP2_Psweights_13TeV-madgraph-pythiaMLM-pythia8_20UL.root", for simulating the Z' signal of mass 2000GeV/c^2 and width 1%.

- "ZprimeToTT_M2500_W25_TuneCP2_Psweights_13TeV-madgraph-pythiaMLM-pythia8_20UL.root", for simulating the Z' signal of mass 2500GeV/c^2 and width 1%.
- "ZprimeToTT_M3500_W35_TuneCP2_Psweights_13TeV-madgraph-pythiaMLM-pythia8_20UL.root", for simulating the Z' signal of mass 3500GeV/c^2 and width 1%.

All the variables determining the loose and tight cuts, separately for loose and for tight cuts, are plotted as overlays with the background and the four Z' mass signals in Figure 50 and Figure 51. The loose cuts and tight cuts will be applied respectively to BOTH the background events ($\mathbf{t\bar{t}}$ and QCD) and the signal events ($\mathbf{Z'_{TC2}}$), so as each process has gone through the same cut conditions.

The plots now that the $\mathbf{Z'_{TC2}}$ signal will been overlaid, and to prepare for when the data is also plotted are, so as to depict the true number of events expected to be obtained in the detector, *cross section scaled*. All variables earlier were depicted as scaled to unity, but the histograms will now be multiplied by the following numbers depicted in the following table. The exact formula and more about this type of scaling have been discussed in IV.ii.

Process	Cross section (pb)[i]	Scaling factor
$t\bar{t}$ background	832 (for all final states)	0.760265179
Z'_{TC2} with m=1400 and Γ=1%	0.9095	0.251950026
Z'_{TC2} with m=2000 and Γ=1%	0.1662	0.047517521
Z'_{TC2} with m=2500 and Γ=1%	0.04749	0.013247152
Z'_{TC2} with m=3500 and Γ=1%	0.005105	0.001532231

Table 2: Cross sections and scaling factors used for each process in the analysis

As can be seen in the plots, all the cuts have been applied correctly, and it is clear that the $\mathbf{Z'_{TC2}}$ signals are significantly smaller that the background signal, especially as the mass becomes larger. The last mass of 3500GeV/c^2 is practically negligible, as is logical by inspecting the small scaling value and cross section that it has. In the mJJ mass spectrum (in Figure 52 in normal and logarithmic scale) one can see the overlay of the background and the four signals. The $\mathbf{Z'_{TC2}}$ signals are also nearly negligible.

By observing the mass distributions for the different Z' cases (Figure 53)[ii], the following mass windows are selected by the author:

i These cross section values have been procured from the source (Bakas *b* 2023). They are from previously created MC simulations, which use *generic* cross section values for all relevant Z' models. Still, this thesis has focused only on the theoretical presentation and interest in the TC2 model, as it is the one most been searched for

ii These are selected from distributions only containing the $\mathbf{p_t}$ cuts.

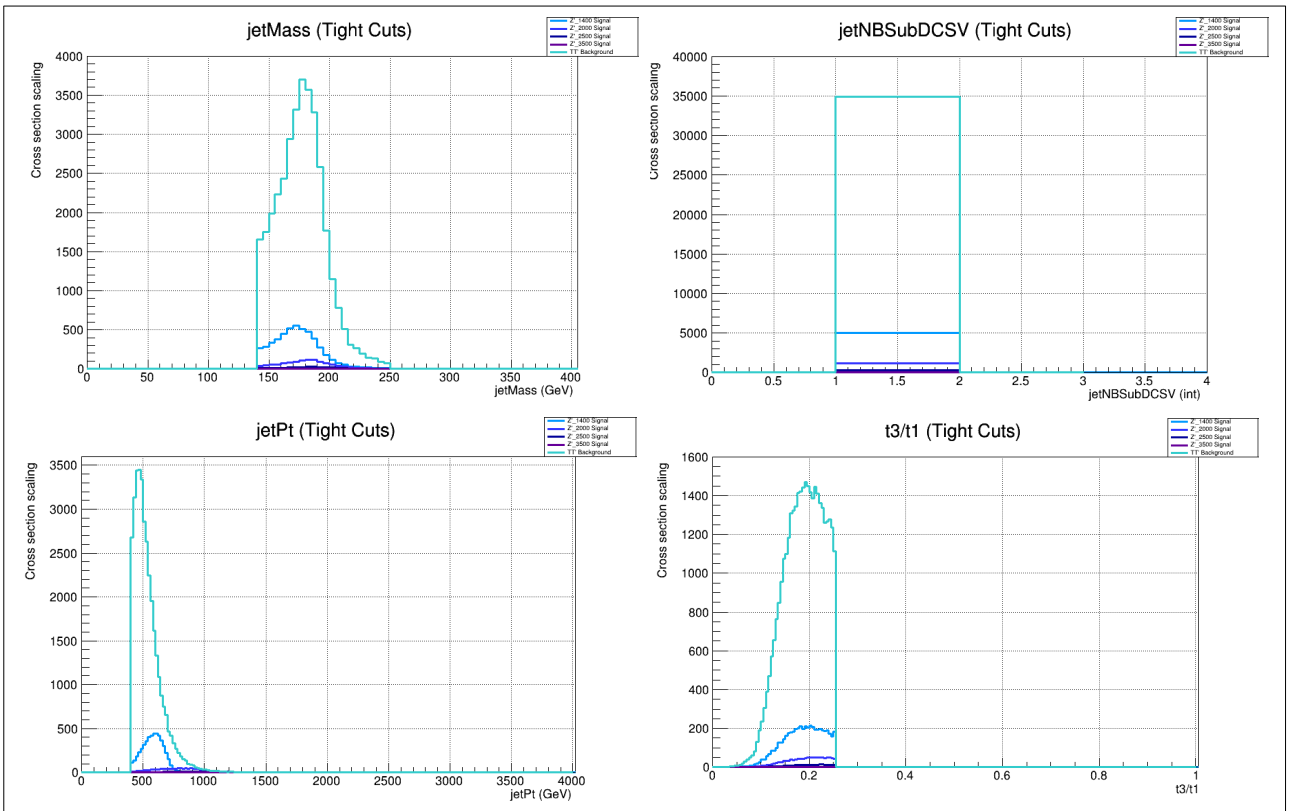


Figure 51: The histograms of jetMass, jetNBSubDCSV, jetPt and τ_{31} with tight cuts for the background and four Z'_{TC2} signals.

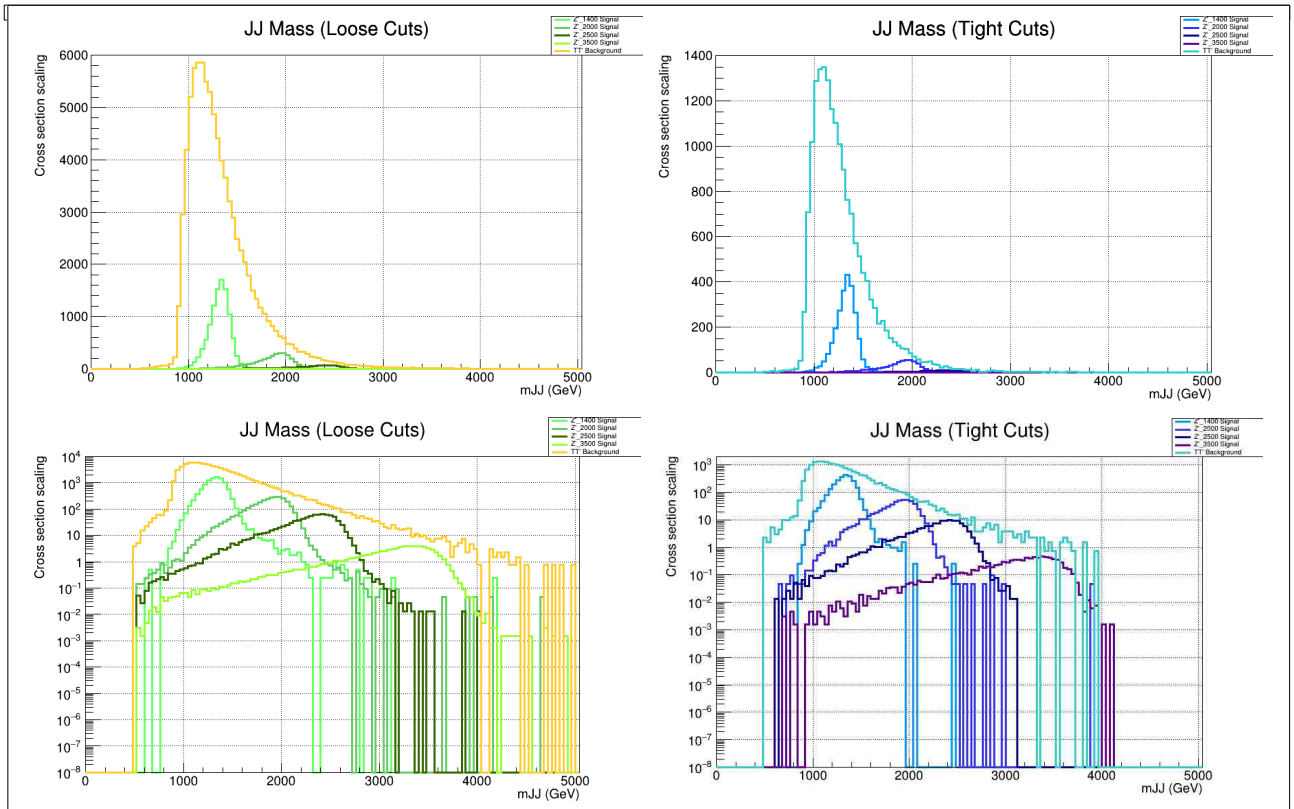


Figure 52: The histograms of mJJ for each of the four Z'_{TC2} signals and the background signals, overlaid together for tight and loose cut cases. The large signal is the tt background, while as the Z'_{TC2} signals decrease in events as the mass increases, and can be seen in the legend. Underneath the same distributions are shown, in logarithmic scale.

Process	Mass window (GeV/c^2)
Z'_{TC2} with m=1400 and Γ=1%	[900, 1600]
Z'_{TC2} with m=2000 and Γ=1%	[1400,2300]
Z'_{TC2} with m=2500 and Γ=1%	[1800,2800]
Z'_{TC2} with m=3500 and Γ=1%	[2500,4000]

Table 3: Mass windows chosen to select the peak for each Z' process

The sensitivity will determine which working point is better based on the search for Z'_{TC2} signal, with the background known from the Top Tagger development. This background will contain $t\bar{t}$

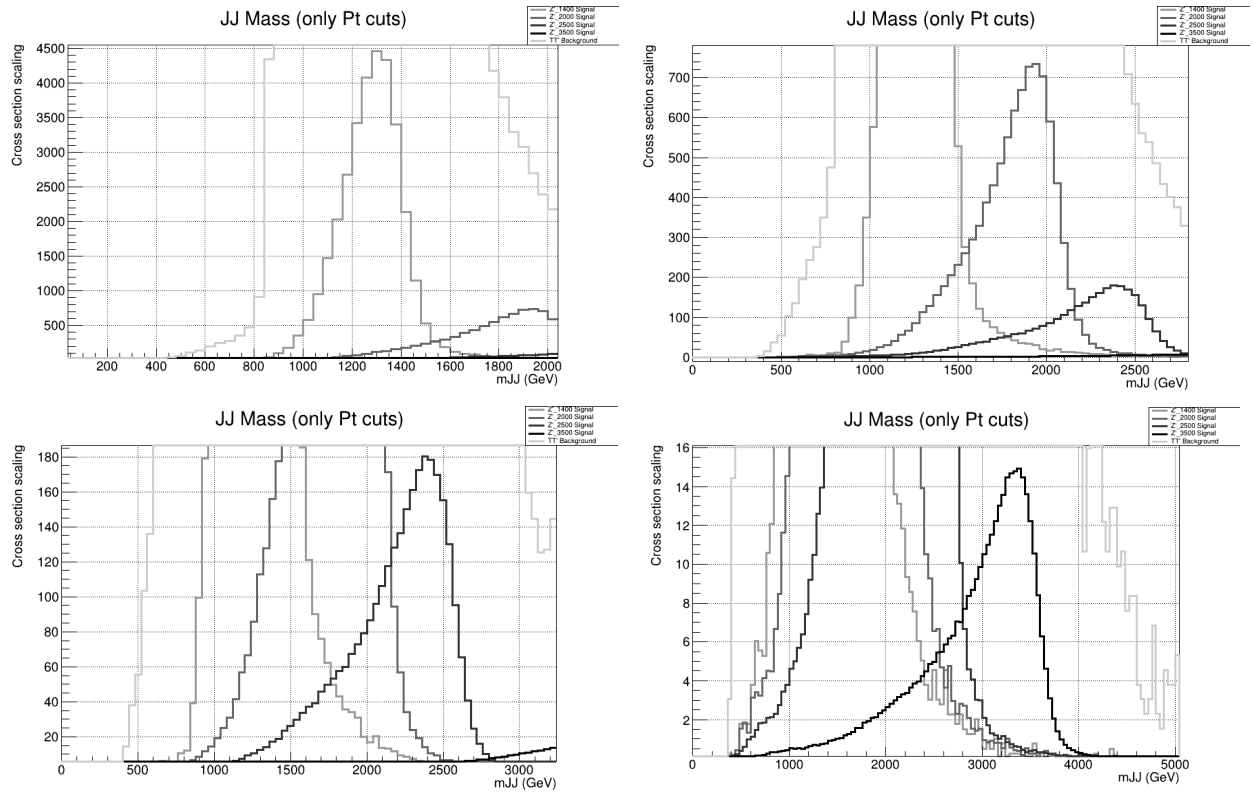


Figure 53: The histograms of mJJ for each of the four Z'_{TC2} signals, taken from the zoomed in overlays of all four together plus the tt background. In each plot, the plot whose peak is perfectly within the canvas boundaries is that of interest: Top left: mass 1400; top right: mass 2000; bottom left: mass 2500; bottom right: mass 3500.

(background for this analysis) events and QCD events. The sensitivity is calculated by the following formula:

$$sensitivity = \frac{signal}{\sqrt{background + signal}}$$

where **sensitivity** and **background** imply taking the integral of the mJJ distribution in the mass windows of Table 3 of the background process and the different Z'_{TC2} process. In total four sensitivities will be calculated: one for each Z'_{TC2} case, by integrating in each mass window. Two plots will arise, the four calculated sensitivities, for each cut case: loose and tight working points.

The sensitivity diagram, plotted as a function of each of the four Z'_{TC2} masses, is shown in Figure 54.

Both plots contain an unnatural drop at mass 2000; after careful inspection, it is the author's opinion

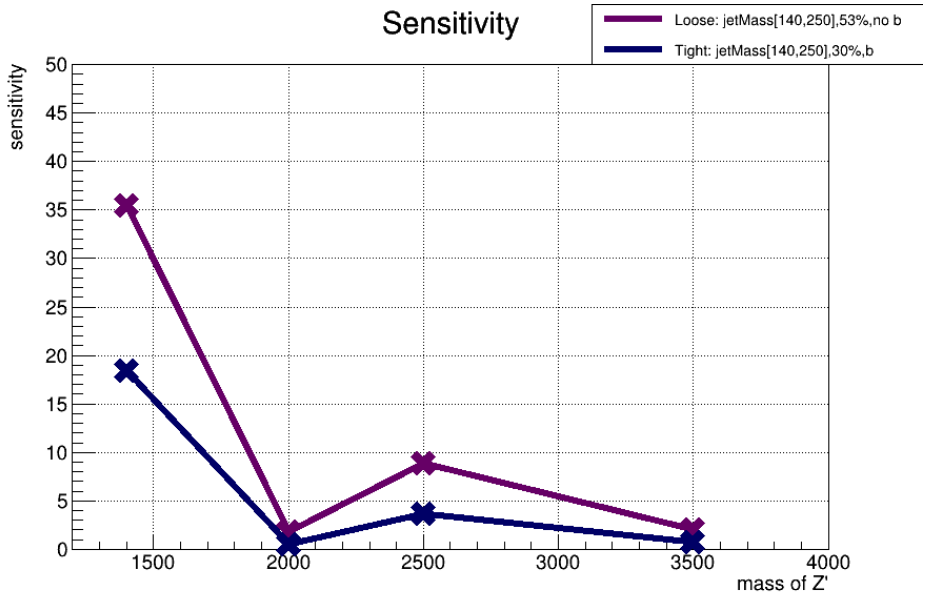


Figure 54: Sensitivity diagram, comparing the loose and tight cut working points, as a function of the Z'_{TC2} mass.

that the ratio of background events and signal events for this mass is just enough to cause the drop, as can be seen in Figure 50. The peak for mass 1400 coincides with the peak of the $t\bar{t}$ background, and for the other two masses the ration of background and signal is relatively stable; however, at mass 2000, the signal appears with a large peak, and at the same time the background is steeply falling with no peak in that mass window, creating the anomalous drop.

The final verdict is that, when comparing the signal to the $t\bar{t}$ background, the loose cuts provide better distinguishing power between the two. Even thought the loose cuts contain more background QCD noise, based on the Top Tagger development earlier, the tight cuts appear to "cut" too much signal with the "b cut" condition.

In the next section, data will be added to the analysis, which contains much more QCD background than any of the previous files. A data-driven method to calculate this amount of QCD will be used, and finally a fitting procedure will be applied to obtain the final constraints of this search.

V.iii. Data analysis and comparison with MC simulations

In the previous analysis the *loose* cuts were deemed as the optimal option, as opposed to the *tight* cuts. Hence, the plots for all five relevant variables (jetMass, jetNBSubDCSV, jetPt, t3/t1 and finally mJJ) are plotted together, including the Data file from **CMS data taken from the LHC in 2016**. The file is the following:

- "JetHT_Run2016-17Jul2018.root", 2016 data file, The CMS Collaboration, CERN

They are first plotted for the loose cuts case in Figure 55, and in Figure 56a for mJJ. Of course the data file has gone through the same cut process as the other processes.

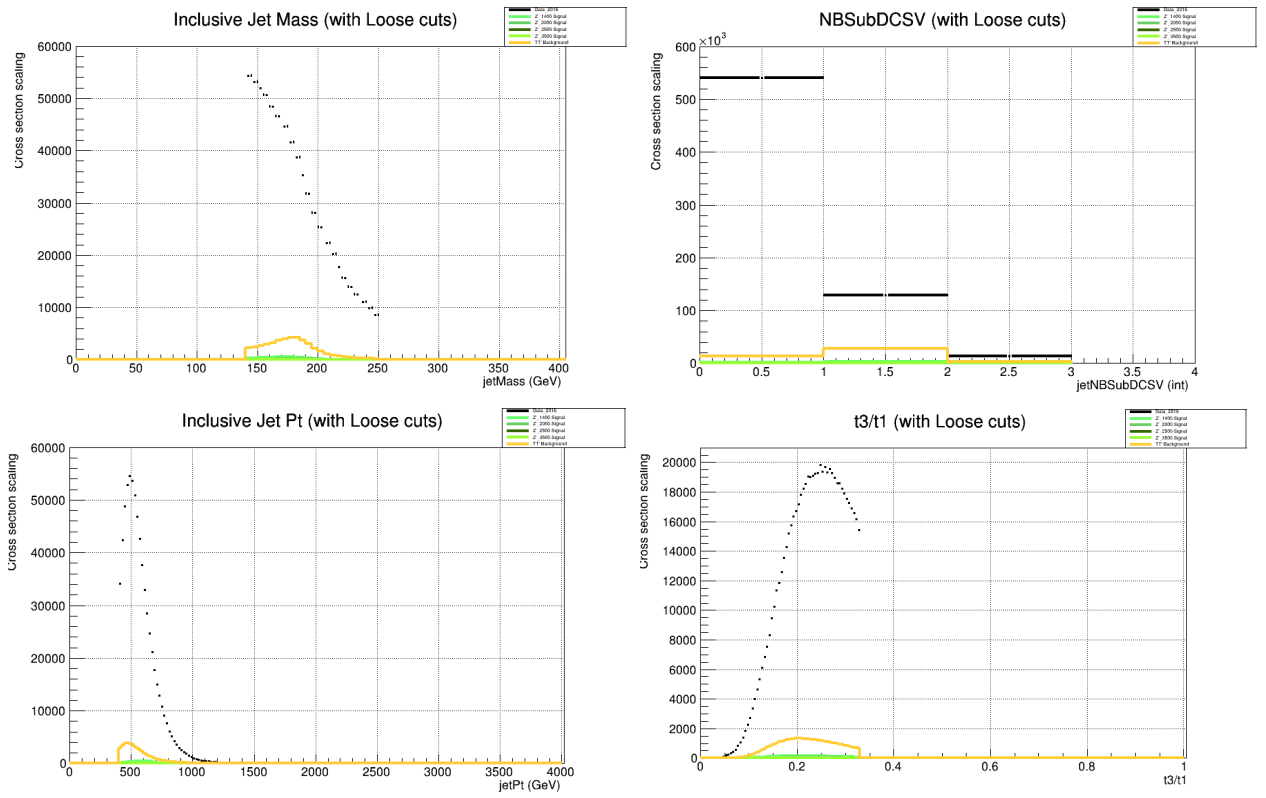


Figure 55: The histograms of the four relevant variables used during the cut process, for loose cuts. Overlaid are: 2016 data, tt background process, four different Z'_{TC2} processes for different mass cases.

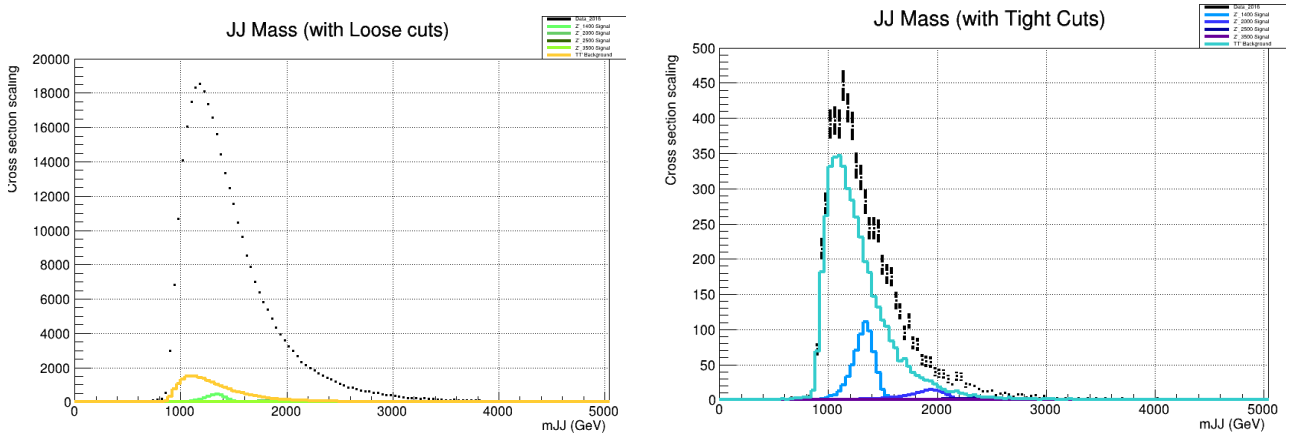


Figure 56: The mJJ histogram with Loose and Tight cuts. Overlaid are: 2016 data, tt background process, four different Z'_{TC2} processes for different mass cases.

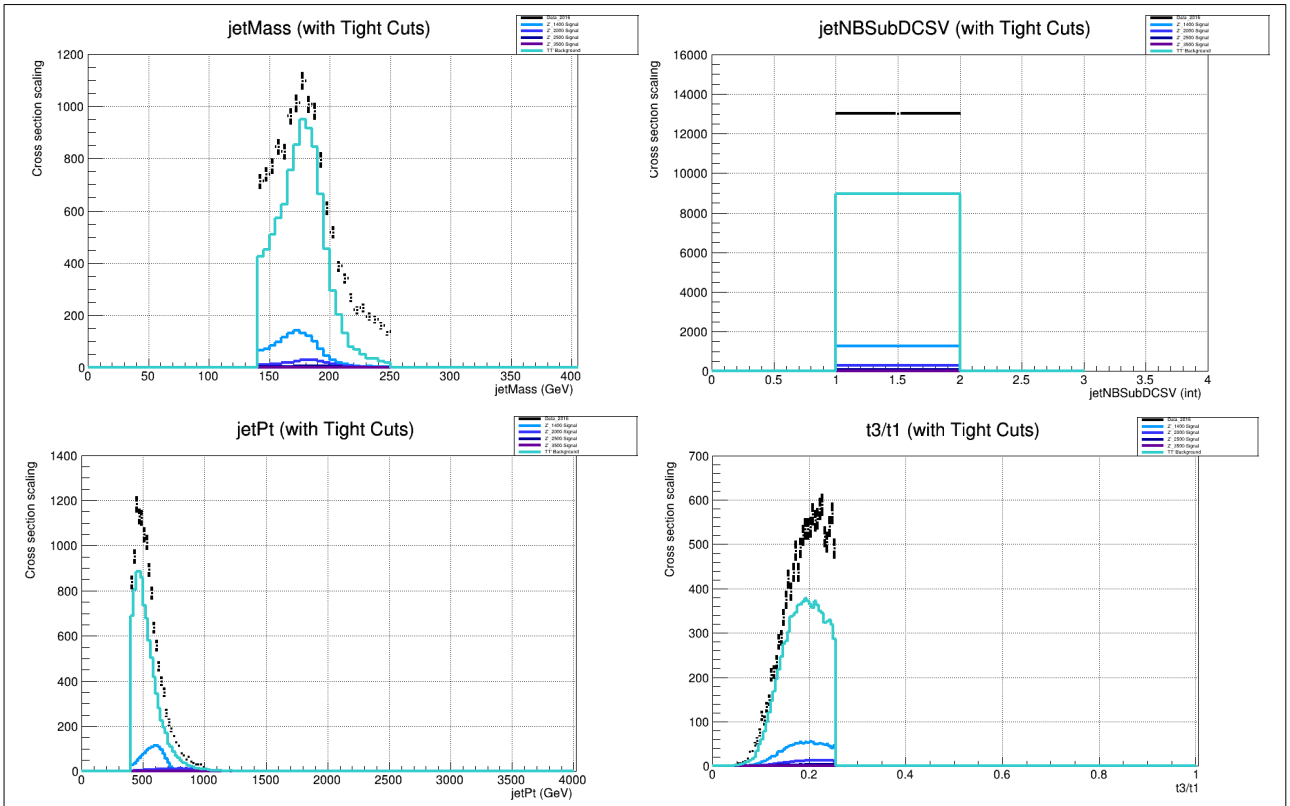


Figure 57: The histograms of the four relevant variables used during the cut process, for tight cuts. Overlaid are: 2016 data, tt background process, four different Z'_{TC2} processes for different mass cases.

As can be seen initially from these plots (Figures 55 and 56a), the data appears to be very "far away" from the $t\bar{t}$ background. This seems to be a case of *QCD flooding*, as the whole analysis up to this point has been to identify $t\bar{t}$ jet events. Applying the chosen cuts, therefore, should have yielded similar distributions for data and $t\bar{t}$ background, which is not the case here. Therefore the same diagrams for the *tight* cuts are also plotted in and Figure 56b for mJJ.

For the tight cuts, there is a much better agreement between the data and the $\mathbf{t\bar{t}}$ process, and so henceforth the chosen cuts will be using the **tight working point**. The difference in agreement now is purely because of QCD background, which was not as present in the Top Development scheme.

Now a *reversal method* will be implemented to calculate the amount of QCD present in the data. This method basically implies reversing the cuts applied, to attain the opposite phase space in order to calculate the "noise". This region will be called the *Control Region* (CR), while as earlier the work was being done in the *Signal Region* (SR). It is recalled that the tight cuts are the following:

$$\textit{Tight working point}: \quad \textit{Mass window cut } A\,[140,250] + 30\,\%\,\textit{Top Tagging Efficiency} + b\,\textit{cuts} \tag{5II.viii}$$

The defining variable, ensuring that the events after applying the above cut are produced by a top jet, is **NBSubDCSV**, which gives the information on whether there is a **b** quark within the jet or not. By "reversing" all the quantities in the cut, the phase space would be irrelevant to the analysis, and the QCD *contamination* would not be able to be calculated. The main variable is the number of bottom quarks in the jets, and that is the one that will be reversed. The following cut therefore, is applied to all variables, keeping every other condition constant and unchanged:

$$\textit{Reversed } b\,\textit{cuts}: \quad NBSubDCSV = 0 \tag{5.III.i}$$

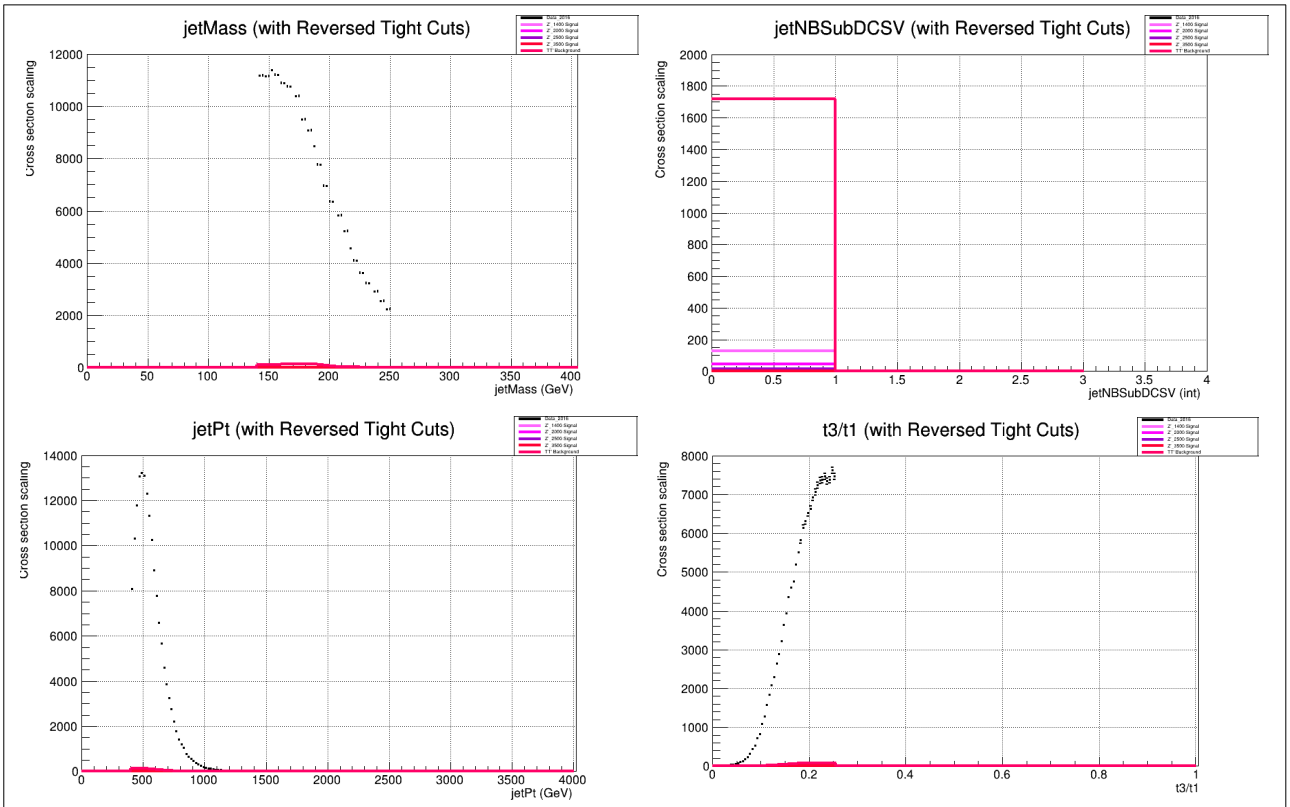


Figure 58: The histograms of the four relevant variables used during the cut process, for reversed b cuts. Overlaid are: 2016 data, tt background process, four different Z'_{TC2} processes for different mass cases.

The plots of all five relevant variables with Reversed Tight Cuts can be seen in Figure 58 and Figure 59.

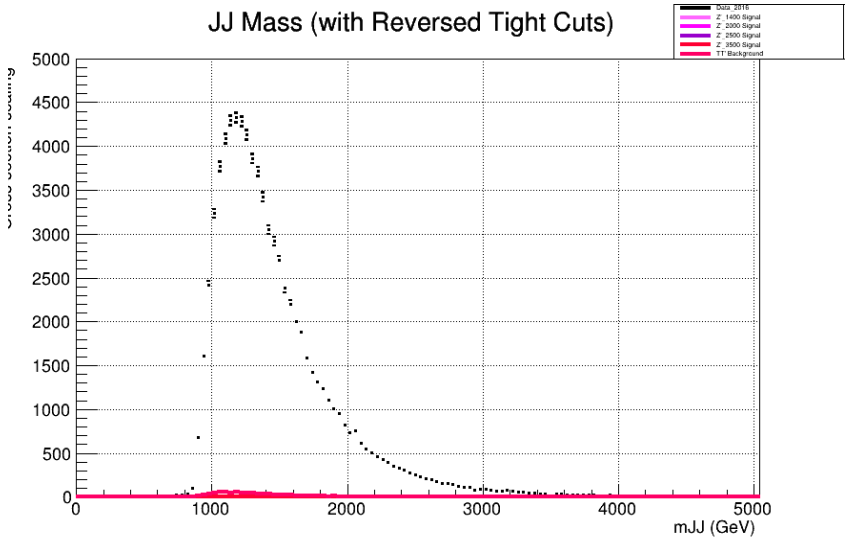


Figure 59: mJJ histogram for reversed b cuts. Overlayed are: 2016 data, tt background process, four different Z'_{TC2} processes for different mass cases.

As can be seen from the NBSubDCSV plot in Figure 58, the *reversed b cuts* have been applied correctly, while all other conditions have remained as they were. Observing the mJJ distribution, it is seen that the QCD contamination from the $\mathbf{t\bar{t}}$ process and the $\mathbf{Z'_{TC2}}$ signal processes are nearly negligible.

The Reversing process is now as follows: The *contamination events* are now subtracted from the data events in the CR for the mJJ variable, so as to procure the "pure" data distribution. This distribution is then normalised to unity to create a PDF, called the *Template*.

The parameter $\mathbf{N_{QCD}}$ is calculated, expressing the amount of QCD present in the analysis:

$$N_{QCD} = N_{data}^{SR} - N_{tt}^{SR} = 4040.13$$

The signal, separately for each $\mathbf{Z'_{TC2}}$ mass case should also be subtracted. However, to avoid complications and because it is so negligible, it will be omitted and only the $\mathbf{t\bar{t}}$ events will be subtracted. This $\mathbf{N_{QCD}}$ is now multiplied by the Template created earlier, to create the final *QCD histogram*. All four created histograms throughout this process are shown in Figure 60.

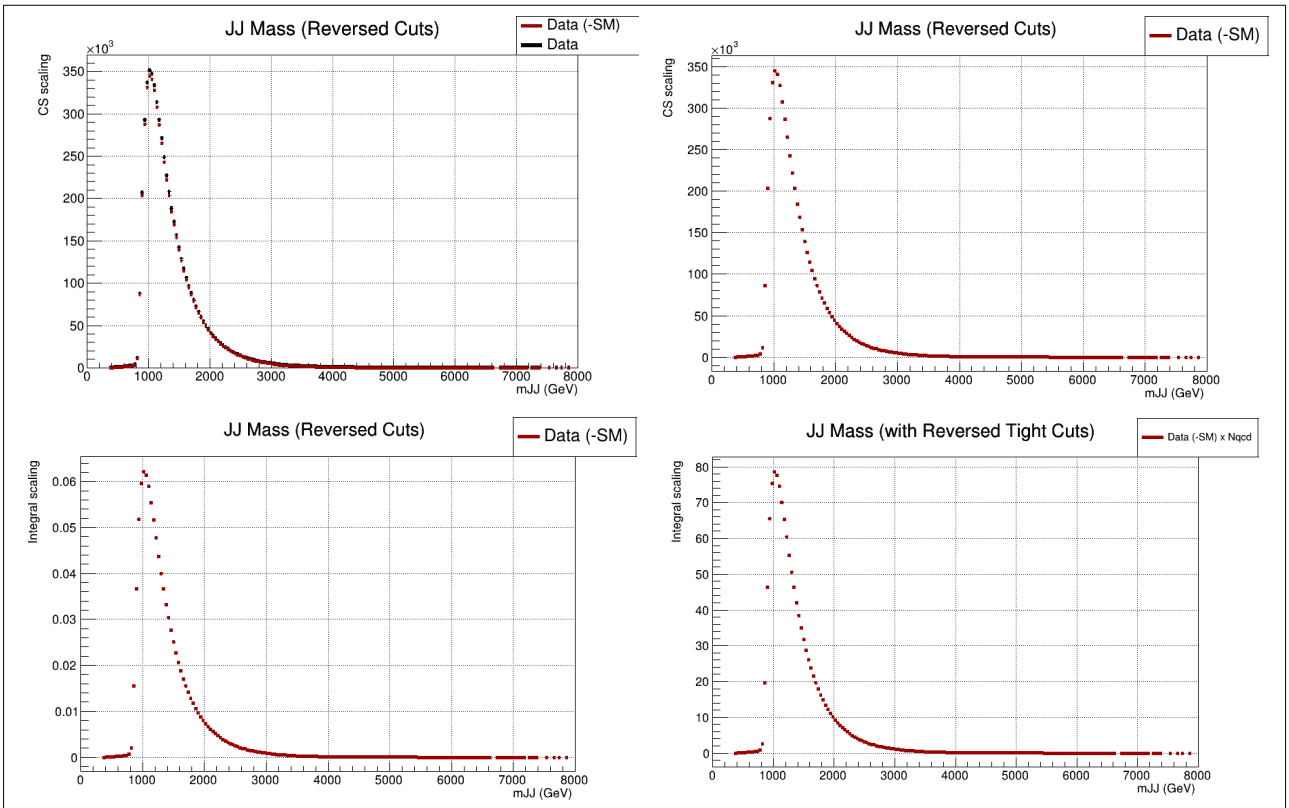


Figure 60: The mJJ histograms during the removal of contamination process. Top left: the data distribution in the CR without the removal of the contamination overlaid with the CR data after the removal of the contamination; top right: the CR data after the removal of the contamination; bottom left: the pdf distribution of the CR data after the contamination; bottom right: the pdf distribution of the CR data after the contamination multiplied by the N_{QCD} parameter.

All variables now go through this removal of contamination process, and are *stacked* on top of other, as seen in Figures 61 and 62. The stacking process includes stacking the $t\bar{t}$ background events *on top of* the QCD distribution calculated above, by practically adding the events. This addition ought to sum to the data events, as the QCD distribution is basically the "subtraction" of the $t\bar{t}$ background from the data. As can be seen below, this is true except for a small incoherence, in which the $t\bar{t}$ background appears slightly increased. The supervisor of the author assured her that this has been seen in other analyses as well, and is from imperfections due to the MC simulations' creation process.

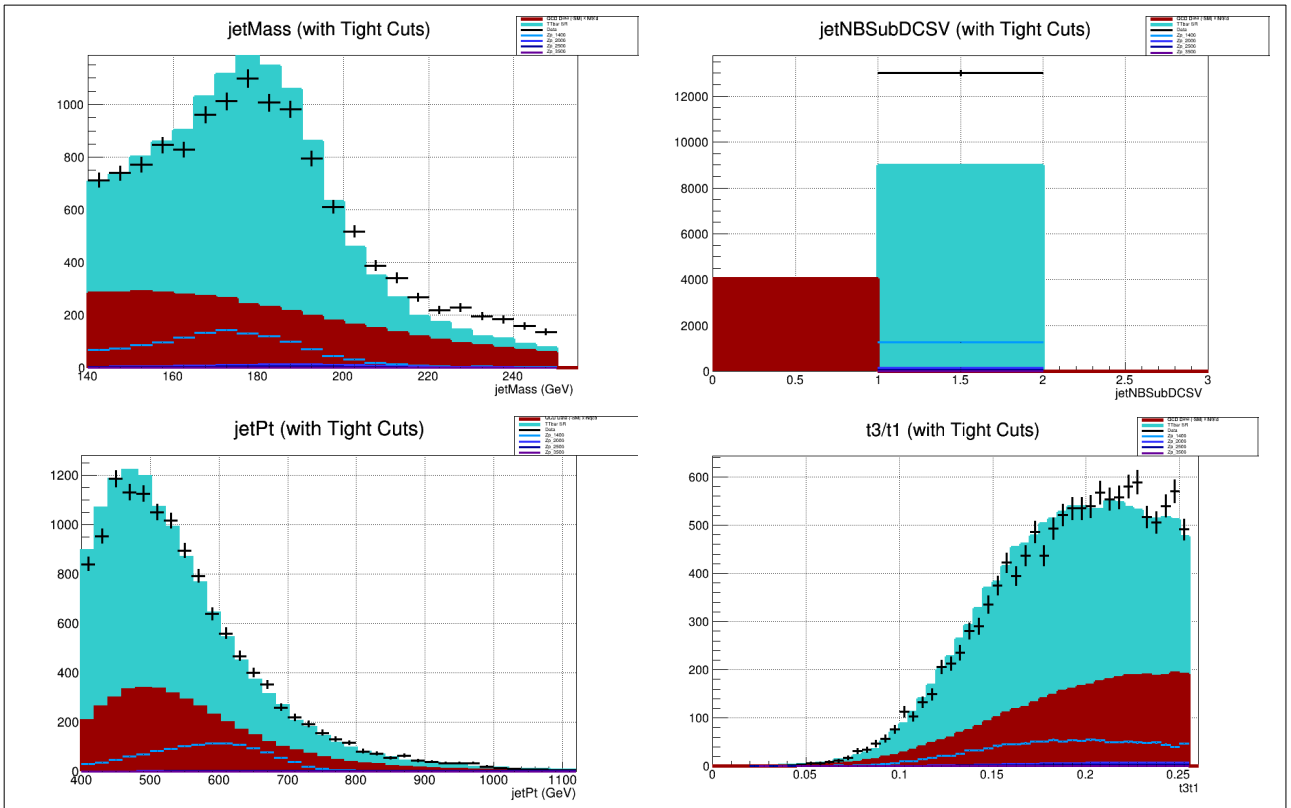


Figure 61: Stacked histograms for all four relevant variables: in red in the QCD distribution calculated above, in teal the tt background and in black above with error-bars is the signal. The lined histograms underneath are the Z'_{TC2} signal distributions, which are nearly negligible.

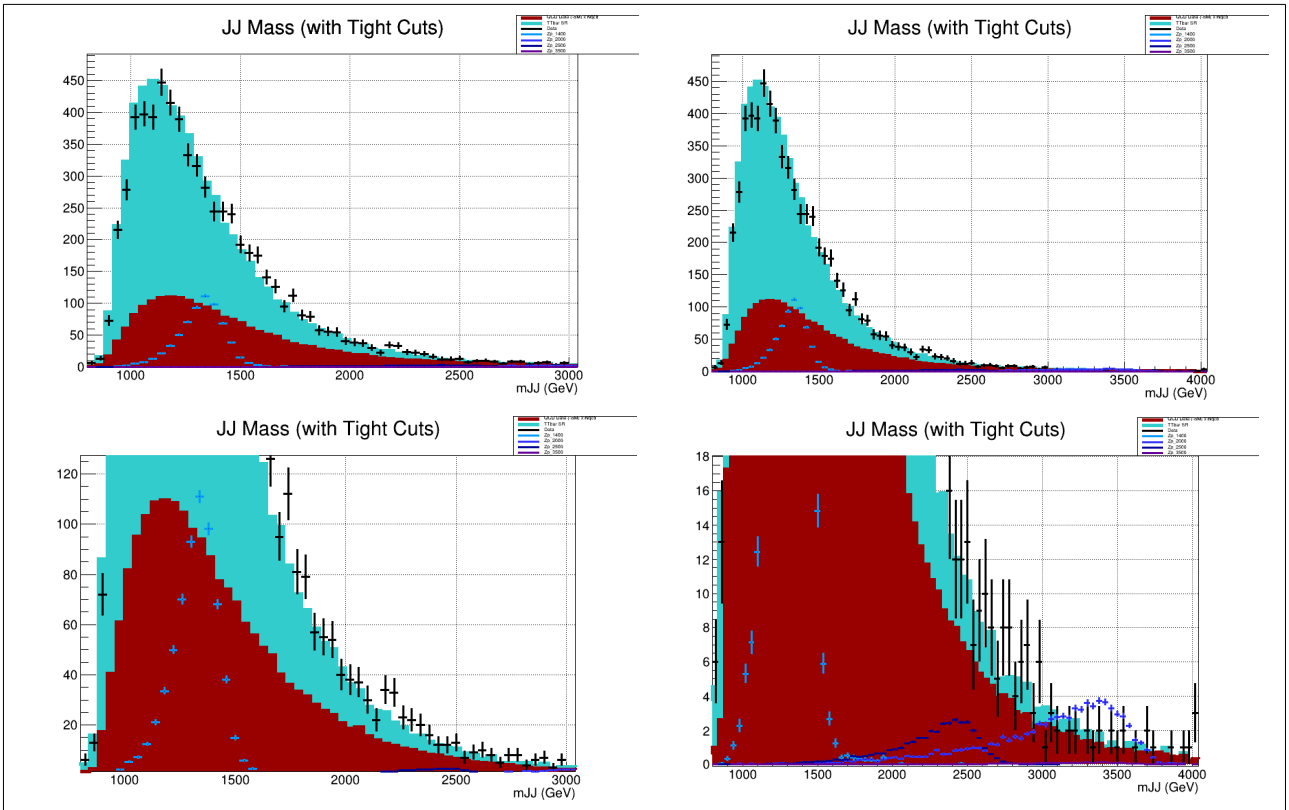


Figure 62: Stacked histograms for the mJJ variable: in red in the QCD distribution calculated above, in teal the tt background and in black above with error-bars is the signal. The lined histograms underneath are the Z'_{TC2} signal distributions, which are nearly negligible. Top figures are the mJJ distribution for different x axes, and underneath zoomed profiles to outline the signal distributions better.

At this point it will be noted, that in order to calculate the QCD background another MC file could have been used, as in the case of the $t\bar{t}$ background, instead of implementing the data-driven method. The reason was that the MC simulations made for QCD are difficult and often inaccurate, and would not yield a good result.

V.iv. Fitting process and Bayesian upper limits calculation

As the author is not familiar with the fitting process implemented using the ROOT framework, to do the fitting to data a ready programme (root code file) was used implemented by her PhD candidate colleagues, Eirini Siamarkou and Theodoros Chatzistavrou.

During the fitting process, first an only background implementation was used, by inserting the histograms from the stacks in the previous section as the "data", "$t\bar{t}$ background" and "QCD background" files. The fitting parameters are the **N_tt** and **N_QCD**, respectively for the "amount" of $t\bar{t}$ and QCD background. The negative log-likelihood (**-2ΔlnL**) function is also plotted. See Figures 63 and 64. The **-2ΔlnL** function is plotted assuming the **N_tt** variable as its parameter, and as can be seen in the plot, its minimum is at the value of approximately **3500 N_tt**.

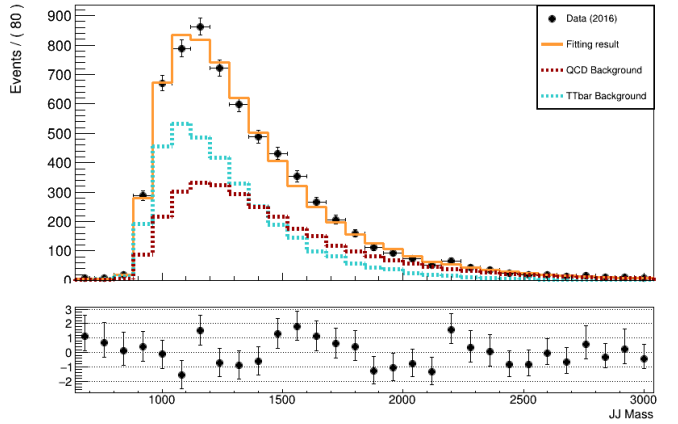


Figure 63: Fitting result for only the background hypothesis with no signal. The parameters are N_tt and N_QCD. Underneath the pull distribution is shown. The binning is divided by two compared to all previous distributions

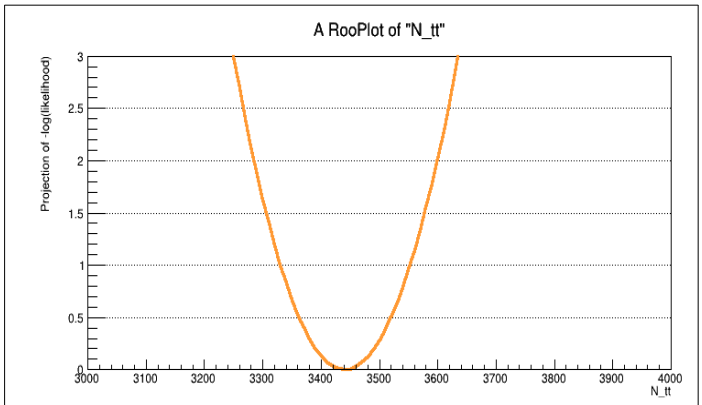


Figure 64: The NLL distribution for the $t\bar{t}$ background.

Adding in the signal, a different fitting procedure will be followed for each Z'_{TC2} mass case, and the **-2ΔlnL** function will be plotted with the parameters of the **N_signal** variable, for each case. As can be seen in the fitting distributions, the signal is extremely weak, and peaks at negative values, as ought to after observing the **-2ΔlnL**. In the 1400 case one can see the the signal takes its negative peak value on the **y** axis. The **-2ΔlnL** function all have their minimums at negative values, which basically means there is no signal detected, and that there is *no discovery*. As the **-2ΔlnL** function spans as nearly a Gaussian distribution, it will contain parts in the negative axis and the positive, if the minimum value is negative.

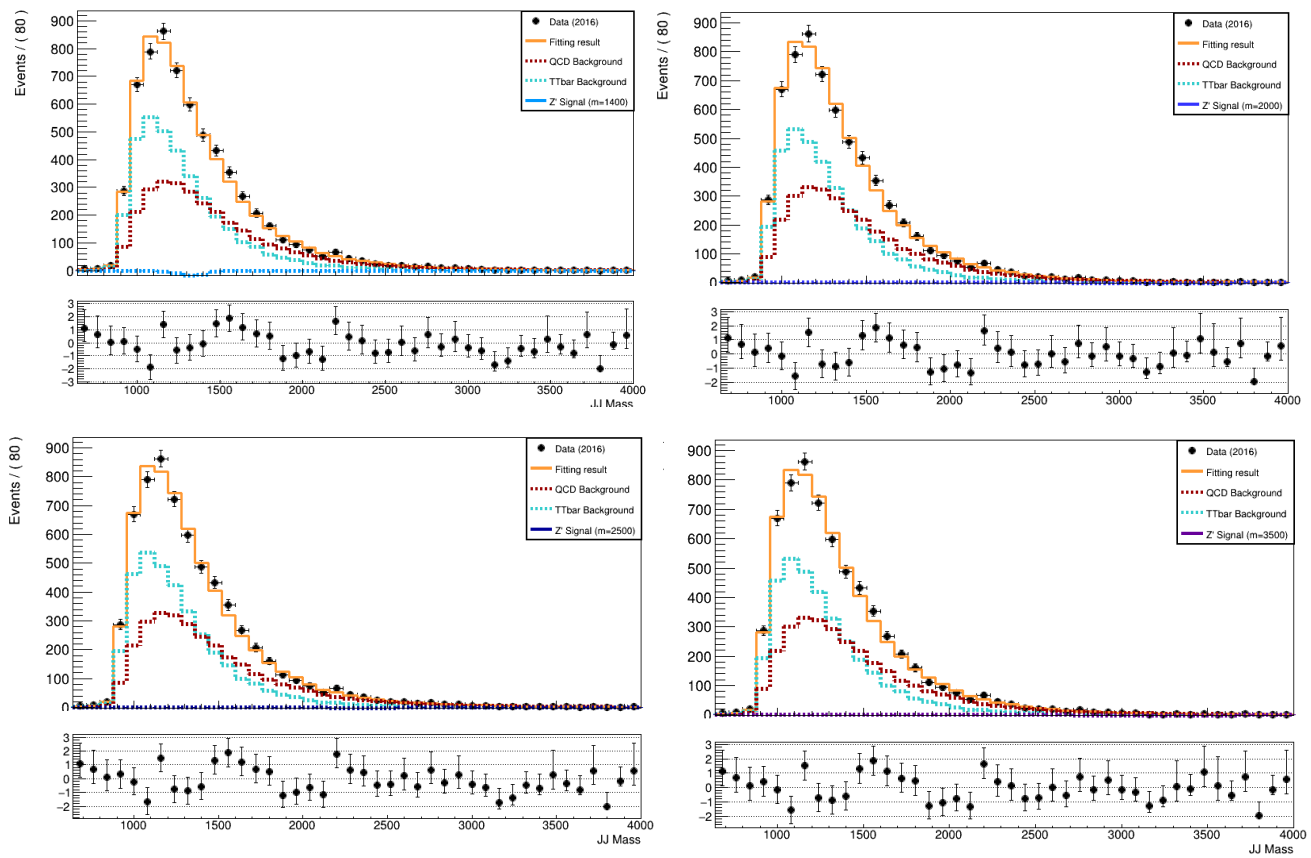


Figure 65: Fitting result containing the parameters of N_QCD, N_tt and N_signal, separately for each mass hypothesis. Underneath the pull distribution is shown. The binning is divided by two compared to all previous distributions

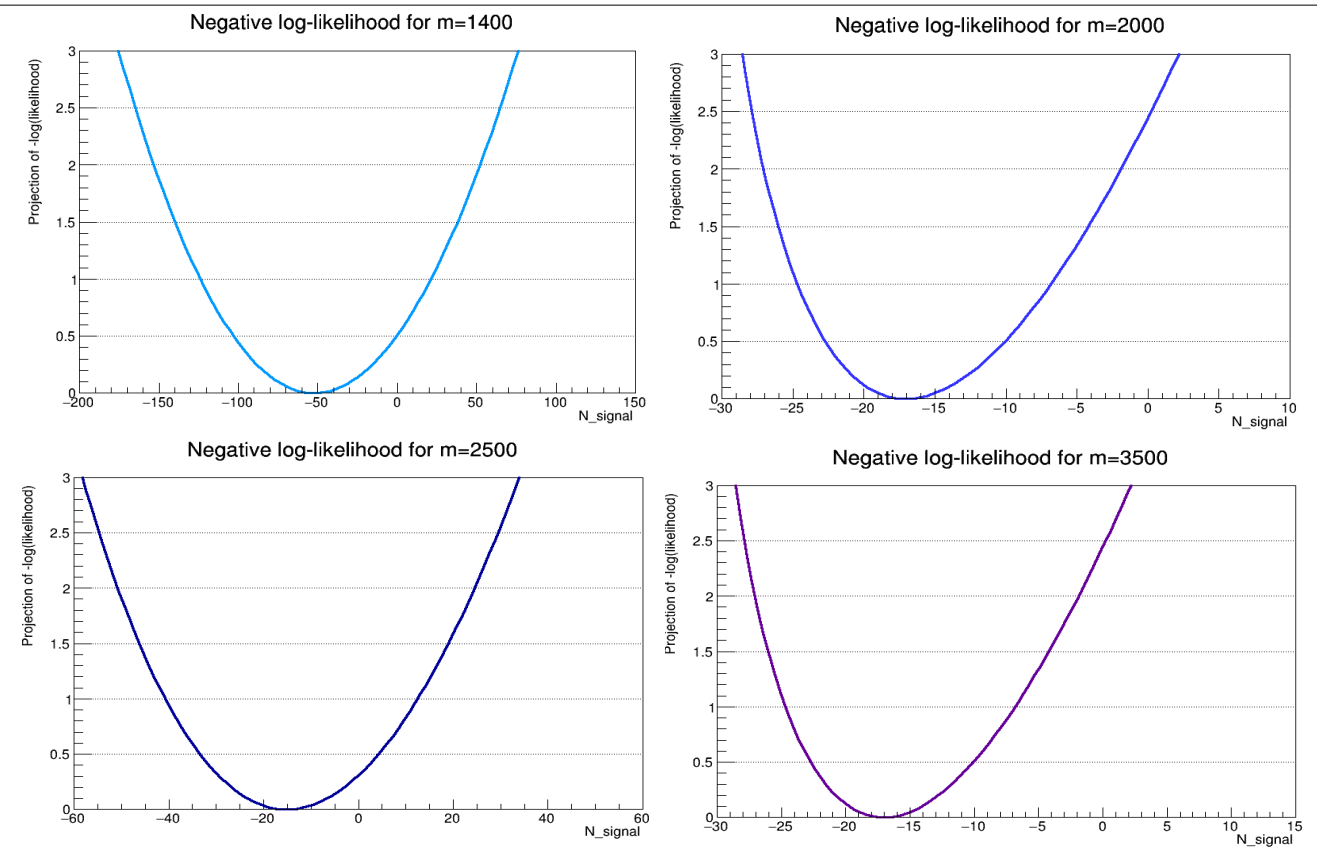


Figure 66: NLL distributions for each mass hypothesis, with the x axis parameter being N_signal. Top left: 1400, top right 2000, bottom left: 2500, bottom right: 3500.

Next the Bayesian upper limits will be calculated, for which the upper limit for the null result will be presented.

The Bayesian upper limits are calculated in the following way (refer to chapter IV.iii for a theoretical overview): the **-2ΔlnL** distributions, separately for each mass case, are "converted" into a Gaussian distribution **L** in the following way:

$$L = p(x|\mu) \rightarrow e^{-2\Delta lnL} \quad \text{(A.m)}$$

where **x** is the unknown parameter that is trying to be calculated, and **μ** is the "known" value. **p(x|μ)** is the conditional probability of **x**, with known **μ**. This distribution is graphed in Figure 67.

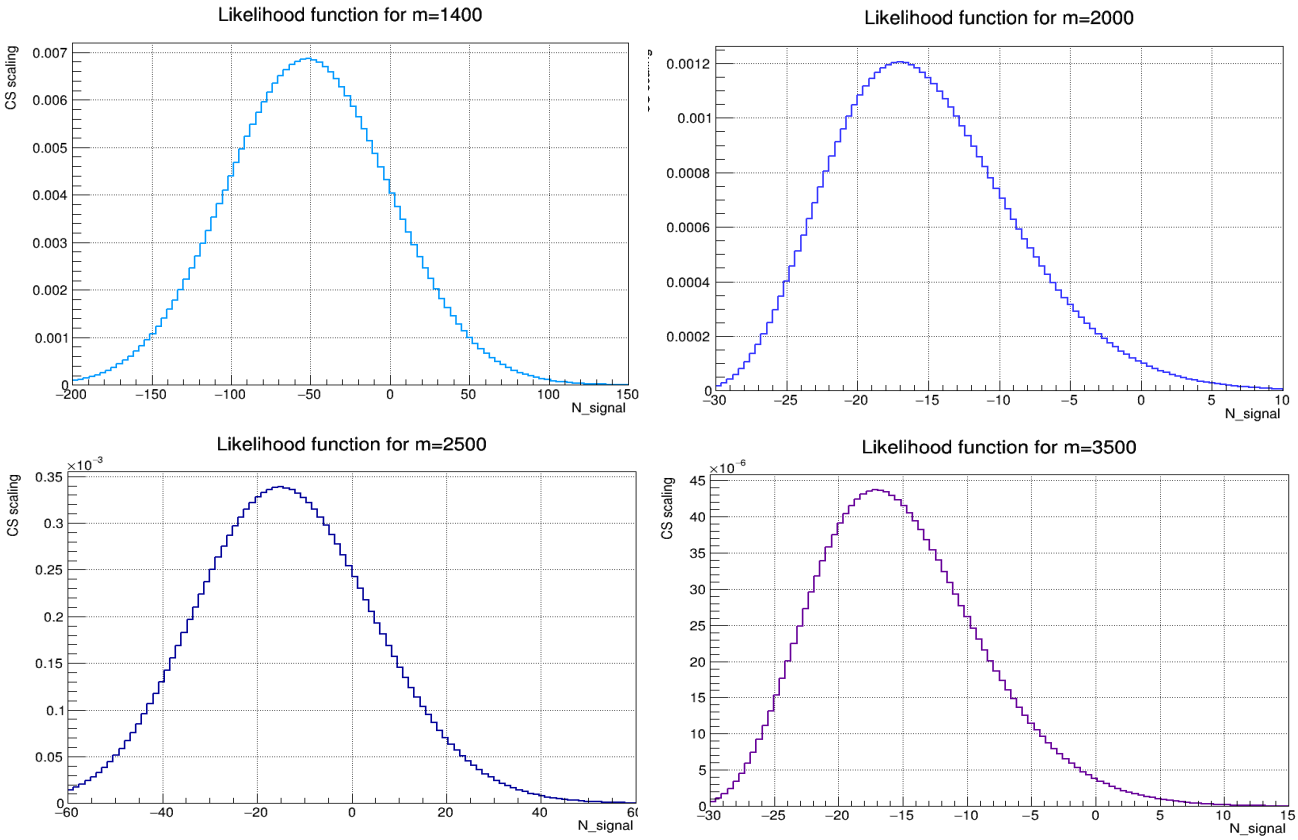


Figure 67: NLL Gaussian distributions for each mass hypothesis, with the x axis parameter being N_signal. Top left: 1400, top right 2000, bottom left: 2500, bottom right: 3500.

Next, the the upper limit μ_{up} is calculated in the following way, where **α** is the *credibility parameter*:

$$\int_0^{\mu_{up}} p(\mu)d\mu = (1-\alpha)\int_n^{\infty} p(\mu)d\mu$$

α shall take two values, 0.1 and 0.05 (or similarly a 90% and 95% Confidence Level). In Figure 68 μ_{up} is plotted as a function of the four masses and in Table 4 these values can be explicitly seen.

Process	μ_{up}	
	$\alpha = 0.1$	$\alpha = 0.05$
Z'_{TC2} with m=1400 and Γ=1%	57.25	74.75
Z'_{TC2} with m=2000 and Γ=1%	6.2	7.8
Z'_{TC2} with m=2500 and Γ=1%	24.6	30.6
Z'_{TC2} with m=3500 and Γ=1%	7.125	8.925

Table 4: Final Bayesian limits for the amount of N_signal for each mass hypothesis

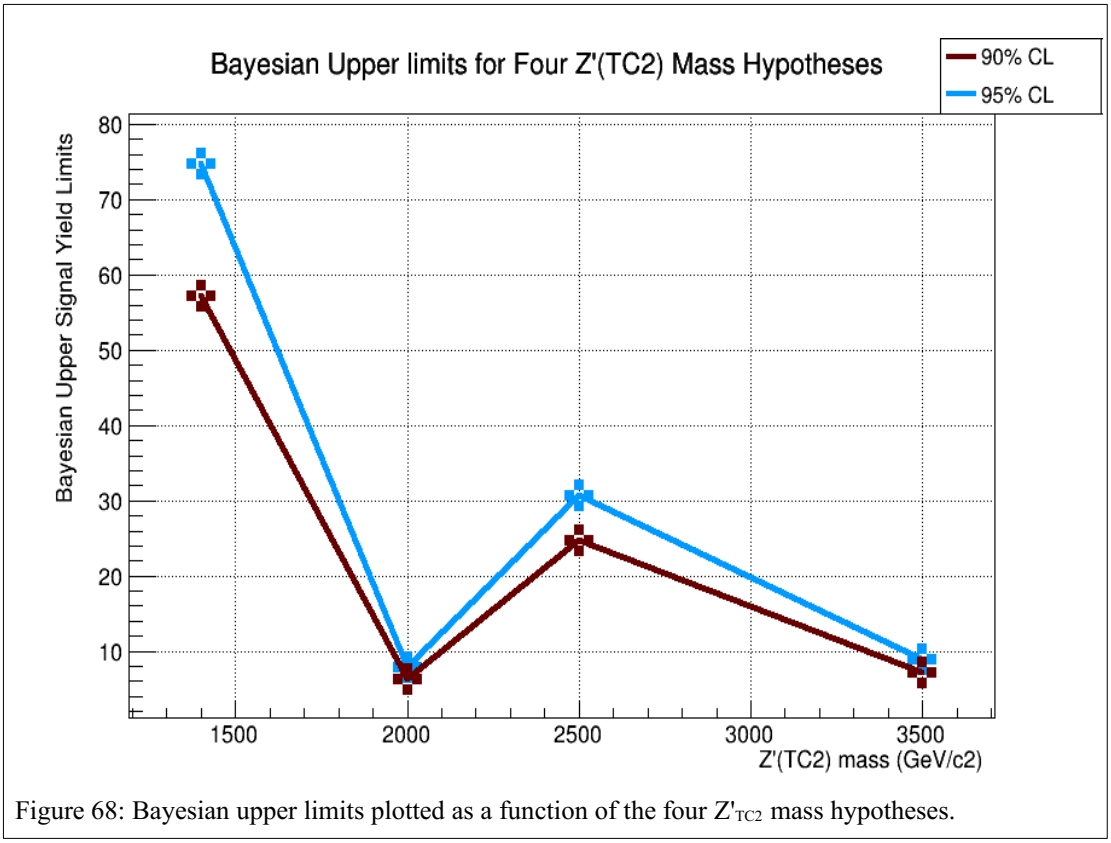


Figure 68: Bayesian upper limits plotted as a function of the four Z'_{TC2} mass hypotheses.

VI. CONCLUSIONS AND FUTURE PROSPECTS

As can be seen from the last diagram of the analysis (Figure 68), the amount of signal observed, **N_signal**, which is analogous to the cross section (it is effectively a cross section multiplicative constant), falls steeply as mass increases. This is an expected result, as the increase of mass in a particle makes is "harder" to detect: it requires more energy to be produced, and it will be a less likely occurring process, as nature will not favour its production. The result of this will be the lack of sufficient data for a proper statistical analysis; as can be seen for the mass at 2000, the statistics here are problematic. There was also an anomaly at this mass value, observed in the Sensitivity diagram, which had originally been attributed to "just as it occurred from the calculations". As this mass seems to exhibit deviations from the others, the author suggests a further inspection of the physics for this case.

The most recent collaborative publications in the search for a Z' are from CMS and ATLAS, namely "Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s}$ = 13 TeV" (2019) and "Search for $t\bar{t}$ resonances in fully hadronic final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector" (2021) respectively. The first search is conducted for all final $t\bar{t}$ states (fully hadronic, lepton+jets, dileptonic), while the second only for the fully hadronic final state. Unfortunately, these results cannot be compared to those in this thesis, as only the *observed Bayesian limits* were calculated and not the *expected* theoretical limits. Therefore, there cannot be any comparison with previous results, nor can a coherent conclusion about this thesis' final result be inferred. The next step would be to calculate these expected upper limits for a proper comparison, or to calculate the *Confidence Level intervals*, using the *Frequentist* method, and not the Bayesian one.

Finally, the author would like to suggest to the scientific community that the Z' boson arising from models I, II and II from source [i][i] should also be searched for. The cross section calculations for these three have already been calculated for the centre of mass energies for the Tevatron, and had presented significantly smaller cross sections than for Model IV. However, with the possibility now of the LHC at 13TeV, detecting them may be have fruitful results. The possible cross sections for the LHC for Model IV had already been calculated by [ii][ii], and to the best knowledge of the author, have not been calculated for models I-III. In the future, it may be a task I am interested in undertaking. I am optimistic that these new searches may provide useful insights into new physics.

i Harris, Hill and Parke, 1999.
ii Harris and Jain 2012.

The link for the presentation of the thesis defence can be found at this URL:

https://docs.google.com/presentation/d/

1I4lpH4eqBn6JfpoZLN8qK054UBToF698Ji0yeV3mYSM/edit?usp=sharing

VII. BIBLIOGRAPHY

Sources in Greek

Στέφος, Α, Στεργιούλης, Ε & Χαριτίδου, Γ 2012, Ιστορία της αρχαίας ελληνικής γραμματείας, Παιδαγωγικό Ινστιτούτο.

G. Bakas *a*, "Μελέτη γωνιακών κατανομών ζεύγους Top Quark με τον ανιχνευτή CMS στο LHC και αναζήτηση νέας φυσικής" (2023), *PhD defence presentation.*

K. Kousouris, *Statistical Methods in Physics (1ο εξάμηνο μεταπτυχιακού)*, course material, προεβλήθη στις 20 Ιουλίου 2023.

F. Mandl, *Στατιστική Φυσική*, trans. Α. Αγγελάκη, Γ. Θεοδώρου και Α. Ακύλας, 2nd ed., (Α.Γ. Πνευματικός, Αθήνα, 2013)

N. Mavromatos, *Θεωρητική Φυσική Θεμελιωδών Αλληλεπιδράσεων (9ο εξάμηνο)*, course material, προεβλήθη στις 18 Ιουνίου 2023.

Particle Data Group, LBNL, *Το Καθιερωμένο Πρότυπο των Σψματιδίων και Αλληλεπιδράσεων*, educational poster (2000).

Sources in English

The ATLAS Collaboration, "Search for $t\bar{t}$ resonances in fully hadronic final states in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector" (2021), https://arxiv.org/abs/2005.05138.

G. Bakas *b*, "Study of top quark pair angular distributions with the CMS detector of the LHC at CERN and search for new physics" (2023), *PhD thesis.*

A. Bardeen, C.T. Hill and M. Lindner, "Minimal Dynamical Symmetry Breaking of the Standard Model" (1990), 10.1103/PhysRevD.41.1647 .

M. Cacciari, G.P. Salam and G. Soyez, “The anti-k_t jet clustering algorithm” (2008), https://arxiv.org/pdf/0802.1189.pdf.

The CMS Collaboration et al., “The CMS experiment at the CERN LHC” (2008), *https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08004/pdf*.

——, “Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s}$ = 13 TeV” (2019), https://arxiv.org/abs/1810.05905.

M. Delmastro, *Experimental Particle Physics: 1. units, quantities, kinematics, measurements*, conference slides, *esipap* (2014), https://indico.cern.ch/event/294651/contributions/671927/attachments/552039/760667/Delmastro_ESIPAP2014_1.pdf

R.M. Harris et al., “Cross Section for Topcolor Z'_t decaying to $t\bar{t}$” (1999), https://arxiv.org/abs/hep-ph/9911288.

—— and S. Jain,, “Cross Sections for Leptophobic Topcolor Z’ decaying to top-antitop” (2012), “https://arxiv.org/abs/1112.4928.

C. T. Hill, PhD thesis, California Institute of Technology, 1977.

——, “Topcolor: top quark condensation in a gauge extension of the standard model” (1991), 10.1016/0370-2693(91)91061-Y .

——, “Topcolor Assisted Technicolor” (1994), https://arxiv.org/abs/hep-ph/9411426.

—— and R,M. Harris and S.J. Parke, “Cross Section for Topcolor Z'_t decaying to $t\bar{t}$” (1999), *https://arxiv.org/pdf/hep-ph/9911288.pdf*.

—— and E. Simmons, “Strong dynamics and electroweak symmetry breaking” (2003), https://arxiv.org/pdf/hep-ph/0203079.pdf.

B. Hobbs, “The Standard Model of particle physics is brilliant and completely flawed” (2017), *ABC Science*, “*https://www.abc.net.au/news/science/20*”.

M. Kuusela and V. Panaretos, “Empirical Bayes unfolding of elementary particle spectra at the Large Hadron Collider” (2014), *https://www.researchgate.net/publication/260003686_Empirical_Bayes_unfolding_of_elementary_particle_spectra_at_the_Large_Hadron_Collider*.

K. Lane, “An Introduction to Technicolor” (1994), https://arxiv.org/abs/hep-ph/9401324.

K. Lannon, F. Margaroli and C. Neu, “Measurements of the production, decay and properties of the top quark: a review” (2012), https://arxiv.org/abs/1201.5873.

N. Manton and N. Mee, *The Physical World: An Inspirational Tour of Fundamental Physics* (Oxford University Press, Oxford, 2017).

B.R. Martin and G. Shaw, “Particle Physics” (2017), 4th ed. UK: John Wiley & Sons.

M.B. Popovic and E.H. Simmons, “A Heavy Top Quark From Flavor-Universal Colorons” (1998), *https://arxiv.org/abs/hep-ph/9806287v3*.

Raeky, άρθρο Wikipedia "Top Quark", <https://en.wikipedia.org/wiki/Top_quark>, seen 4th June 2023, (2009).

B. Still, *Particle Physics Brick by Brick: atomic and subatomic physics explained… in lego®*, (Firefly Books Ltd., New York, 2017).

“Theorists and Experimentalists: Partners in the Search”, *Fermilab Annual Report* (1993), https://lss.fnal.gov/archive/annual/fermilab-annual-1993.pdf.

J. Thale and K.V. Tilburg, “Identifying Boosted Objects with N-subjettiness” (2011), *https://arxiv.org/pdf/1011.2268.pdf*.

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

R.L. Workman et al. (Particle Data Group), Prog. Theor. Exp. Phys. (2022), 083C01 (2022.)