SEARCH FOR SINGLE VECTOR-LIKE T QUARK DECAYING INTO ALL HADRONIC FINAL STATES WITH THE ATLAS DETECTOR AND UPDATED FLAVOR TAGGING PARAMETERIZATIONS AND PIXEL SIZE CHARACTERIZATION STUDIES FOR THE HL-LHC UPGRADE

By

Madhuranga Thilakasiri Madugoda Ralalage Don

Bachelor of Science in Physics.
University of Sri Jayewardenepura
Nugegoda, Sri Lanka
2010

Submitted to the Faculty of the
Graduate College of
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2020

SEARCH FOR SINGLE VECTOR-LIKE T QUARK DECAYING INTO ALL HADRONIC FINAL STATES WITH THE ATLAS DETECTOR AND UPDATED FLAVOR TAGGING PARAMETERIZATIONS AND PIXEL SIZE CHARACTERIZATION STUDIES FOR THE HL-LHC UPGRADE

Dissertation Approved:

Dr.Joseph Haley

Dissertation Advisor

Dr.Alexander Khanov

Dr.K.S.Babu

Dr.Subhash Kak

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere gratitude for those great people who were close to me during my graduate studies at the department of physics of the Oklahoma State University, and also for those who enable me to reach this height in my life through their valuable contributions.

My dissertation advisor Dr. Joseph Haley has been a strong fort and a safety net throughout my graduate studies as he has been always extending his helping hand, especially during the time I was getting used for the excruciating codings involved in particle physics. I sincerely appreciate his approach in extending enough freedom in exploring extents in my studies while providing guidance during the time when it should be.

I'm thankful for my committee members Dr.K.S.Babu, Dr.Alexander Khanov, and Dr.S.Kak for their valuable feedbacks and the support they have extended towards the completion of my studies.

I'm thankful for the former and current Oklahoma State ATLAS group members Dr.Flera Rizatdinova, Dr.David Jamin, and Steven Welch for the support they have given to me at various stages in my graduate studies and for creating an enthusiastic environment to do research work. I'm also glad that I had the chance to work with the fellow students Leyland Palmer, Wasikul Islam, Evan Richard Vanderwall, and Egor Antipov in the group.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

I would like to extend my sincere gratitude to Dr.Alexander Khanov for the kind guidance and for teaching the fundamentals of b-tagging studies which enabled me to take my authorship qualification task at CREN with enthusiasm and for Dr.Alessandro Calandri for his immense caring support in implementing the flavor tagging upgrade studies related to my qualification work.

Dr.Nuwan Karunaratne has been an inspiring person during my undergraduate life where he has been continuously encouraging and supporting me to get into the four-year special degree program at my undergraduate university and later in continuing towards graduate studies.

My undergraduate advisor Dr.W.D.A.T.Wijeratne deserves a special appreciation as he has been continuously supporting and guiding me in my undergraduate research work by providing literature and teaching both theoretical and practical experimental methods at the laboratory and the department workshop.

I would like to thank Dr.Pekka Sinervo, Dr.Riccardo Di Sipio, Joel Foo, Sahibjeet Singh, Shannon Egan, Dylan Folsom, and Daniel Levy for their vital contributions towards the analysis described in this dissertation.

I had the greatest opportunity to work with some brilliant people at the ATLAS community where I was able to learn a plethora of vital skills from them. I would like to thank Dr.Daniel Marley, Dr.Alison McCarn for their support in my early days of coding and to Dr.Samuel Meehan, Dr.Giordon Stark, and Danica Marina McDonell for their support in learning continuous integration and analysis preservation work. I'm also thankful to Dr.Petr Jačka and Dr.Oliver Majersky for their support in the production of signal samples towards the analysis described here.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

My sincere gratitude goes to the department of physics at Oklahoma State University for providing an exceptional learning and researching environment through the exceptional faculty and staff and a special thank should go to Susan Cantrell and the crew for providing a vibrant environment with joyous events throughout the year. I really enjoyed the opportunities I gained from being a student at Oklahoma State University which allowed me to work with great people from various backgrounds, enriching my life as a whole.

I would never be able to pursue my graduate studies if not for the support I have received from my family. My mother Shantha and my late father Thilakasiri deserve my heartfelt gratitude for everything they provided me throughout my life and for teaching the value of education by providing me scientific literature since my childhood. I would also like to thank my brother Tharanga and sister-in-law Anusha for taking care of everything back in my country, Sri Lanka which helps me to concentrate on my studies.

I won't manage to come this far in my graduate life if not for the immense support from my beloved wife Erandi, thank you very much for being with me and encouraging me especially during the worst days in my life. Even though we lived apart thousands of miles away, our endless skype hours kept me going and the two instances I had to see you in a year were a bliss to my life. I'm extremely proud of your achievements as a graduate student and am eagerly waiting to join with you in Arizona to begin our lives together.

Finally, I would like to extend my sincere gratitude to the public education system and the people of the Democratic Socialist Republic of Sri Lanka, for providing free education for the children of the nation from elementary to tertiary education as myself is too a result of this great system.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

I dedicate my dissertation to the most precious people in my life, firstly, to my late father Thilakasiri, and my mother Shantha. I would not be the person I am today if not for your unconditional love and dedication throughout my life, you two are the beacons of my life. Secondly to my wife Erandi, for her utmost love and support throughout my graduate life, thank you for your patience and staying with me during the ups and downs.

Name: Madhuranga Thilakasiri, Madogoda Ralalage Don

Date of Degree: July, 2020

Title of Study: SEARCH FOR SINGLE VECTOR-LIKE T QUARK DECAYING INTO ALL HADRONIC FINAL STATES WITH THE ATLAS DETECTOR AND UPDATED FLAVOR TAGGING PARAMETERIZATIONS AND PIXEL SIZE CHARACTERIZATION STUDIES FOR THE HL-LHC UPGRADE

Major Field: Physics

Abstract: This dissertation summarizes two independent research works where the first is related to a Vector-Like Top quark decays to an all hadronic final state with Higgs and a SM top. The second is on flavor tagging parameterizations and pixel size characterization studies for the HL-LHC upgrade. The first analysis searches for a final state with hadronic jets and no leptons, which will be used to reconstruct the single produced vector-like Top quark at different mass and coupling values. The leading and the sub-leading jets in the analysis are tagged either as top, b or Higgs, and based on the tagged status, a 9×9 matrix is defined to recognize control and signal regions for the analysis. A τ_{21} dependent Higgs tagger is used to avoid b jet dependency. Data collected by the ATLAS experiment with a $\sqrt{s}$=13 TeV proton-proton collisions delivered by the Large Hadron Collider from 2015 to 2018 were used. The analysis is currently in the blinded stage and the derived expected upper limits at 95% Confidence Limits for three coupling values of $\kappa = 0.1$, 0.5 and 1.0 are presented here. The second study is related to the upgrades to the ATLAS detector with the introduction of a new Inner Tracker, which will record more statistics at higher pseudorapidity regions. The MV2c10 b tagger was used in defining the best Boosted Decision Tree parameters that results in the highest light-jet rejection as a function of b-tagging efficiency by using a combination of $t\bar{t}$, single-top and Z$'$ samples. A similar analysis was performed on characterizing the impacts of pixel size choices of 25×100 μm^2 vs 50×50 μm^2 using $t\bar{t}$ samples. Finally, the flavor tagging efficiencies and uncertainties for "b", "c", and "light"flavors were derived, which will be used in the High Luminosity Large Hadron Collider era.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

ABBREVIATIONS

BDT	Boosted Decisions Tree
BSM	Beyond Standard Model
CL	Confidence limit
DM	Dark Matter
EM	Electromagnetic interactions
HL−LHC	High Luminosity Large Hadron Collider
LHC	Large Hadron Collider
MC	MonteCarlo
NP	Nuisance Parameter
PL	Profile Likelihood
POI	Parameter of interest
SI	Signal injection factor
SM	Standard Model

CHAPTER I

Introduction

1.1 The Standard Model(SM)

Humans have always been curious to know, what are the fundamental building blocks of nature? From the humble beginnings of the eastern philosophers in modeling that everything is made out of water, fire, air, and earth, we now have a much more elaborate and accurate model to explain the fundamental building blocks of nature. The Standard Model is a theoretical framework to explain how matter constitutes and how it interacts with forces and with itself. The SM defines the following particles to be fundamental particles.

1. Fermions (Quarks and Leptons)
2. Charge carrier particles (fundamental bosons)
3. Higgs boson

In addition, all the above given fundamental particles have an anti-particle that has the same mass but opposite charge and all other quantum numbers. Since all these particles and anti-particles are considered as elementary particles, they are treated as point-like particles that have no internal structure. Over the next sub-sections, the above fundamental particles are discussed more elaborately.

1.1.1 Fermions

Quarks

Proton-proton collision experiments revealed that a proton is composed of elementary particles that are known as **"quarks"**. Further experiments found that quarks have three

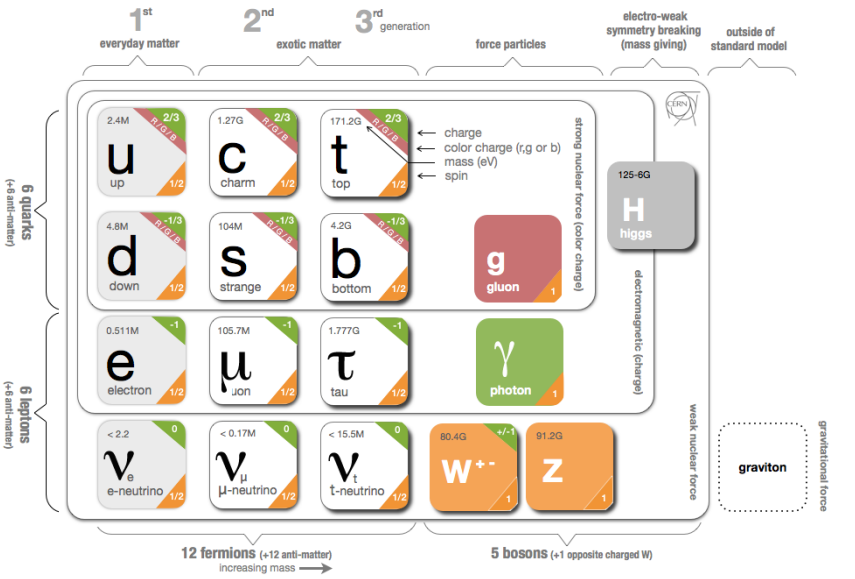


Figure 1.1: Table of elementary particles

identifiable generations that lead to have six flavors for them. The first generation is comprised of two quarks, "**up**" and "**down**" and they are stable in existence compared to the other quarks. As an example, a proton is made out of uniting two up quarks and a single down quark, and a neutron is made out of uniting two down quarks and a single up quark. The second generation has **"charm"** and **"strange"** quarks and the heaviest and the least stable quarks are found in the third generation that are **"top"** and **"bottom"**. Though six flavors can be seen among the quarks, one common feature also can be observed, that is they have fractional electrical charges. Quarks **up**, **charm** and **top** have a fractional charge of $(\frac{+2}{3}e)$[1] and quarks **down**, **strange** and **bottom** are having a fractional charge of $(\frac{-1}{3}e)$. A unique feature of quarks is that they are held together by color charge.

Based on the Quantum Chromo-Dynamics (**QCD**) theory, particles with color charge will interact via the strong force. When it comes to color charge, three colors are define as **"Blue"**, **"Red"** and **"Green"** and quarks unite in such a way that the constituent will become colorless. One of the fascinating properties of the color charge of quarks is that they never exist alone because the strong force will always bid different colors

[1] e: Magnitude of the charge of an electron

together until the resulting object is color neutral. Quarks do have their anti-particles and by uniting with them, they can produce **"Mesons"**' or quarks may also unite as three to form **"Baryons"**. Table 1.2 gives some examples for Mesons and Baryons.

1^{st}	2^{nd}	3^{rd}	$Charge$
$up(u)$	$charm(c)$	$top(t)$	$\frac{+2}{3}$
$down(d)$	$strange(s)$	$bottom(b)$	$\frac{-1}{3}$

Table 1.1: quarks by generation and their fractional electric charge

Baryons (three quarks)	Mesons (quark-anti quark pair)
$Proton = uud$	$\pi^{+} = u\bar{d}$
$Neutron = udd$	$\pi^{-} = d\bar{u}$
$\Sigma = dds$	$J/\Psi = c\bar{c}$

Table 1.2: examples for Bryons and Mesons (Hadrons)

Leptons

Leptons are the other family of fundamental fermions. There are three electrically charged versions which can be ordered from the smallest to largest as follows,

1. Electron
2. Muon
3. Tau

As we have seen in quarks, charged leptons too have their own anti-particles and they also accompany another set of particles (considered massless according to the SM and are chargeless) known as **"neutrinos"**. Charged leptons are capable of interacting with electromagnetic, weak, and gravitational forces. Like quarks, leptons are spin half ($\frac{1}{2}$) fermions as quarks but have no color charge, associate to them. In terms of their lifetime, the electron is stable while muon and tau eventually will decay into electrons and neutrinos/anti-neutrinos.

Lepton	Electric charge
e^- μ^- τ^-	-1
ν_e ν_μ ν_τ	0

Table 1.3: Leptons and their electric charges

1.1.2 Bosons

Charge carrier particles (Fundamental Bosons)

Force carrier particles are another set of elementary particles that can be seen in the SM.

1. Photon - force carrier of electromagnetic interactions
2. Gluons - force carriers of strong interactions
3. W and Z bosons - force carriers of weak interactions

Photons are massless bosons that mediate the electromagnetic force between electrically charged particles. Gluons are also massless and work as the force carriers for the strong interaction between colored particles. There are eight different gluons corresponding to different color/antic-color combinations. Gluons play a major in particle collisions where the majority of the resulting particles are generated from the interactions of the gluons. When the protons collide at the Large Hadron Collider as beams, the gluons in the proton beam are interacting to generate many of the exotic particles, see fig 1.3. There are three force carrier particles for the weak interaction, W^+, W^-, and Z^0. While W^+ and W^- carry a charge of $+1$ and -1, Z^0 is electrically neutral. Unlike gluons and photons, W^+, W^-, and Z^0 are massive particles. Since all these force carrier particles are in the family of bosons, their spin is 1.

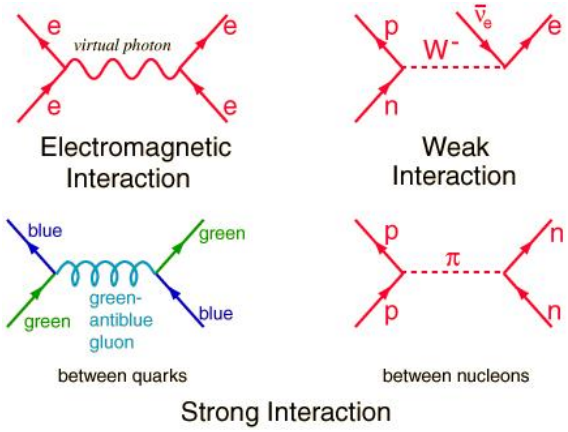


Figure 1.2: Examples for force carrier particles

1.1.2.1 Interactions mediate via (gauge) bosons

Electromagnetic Interactions (EM Interactions)

The EM interactions can be observed when there are charged particles around. Many everyday forces such as magnetism, friction, etc... are results of the EM interactions. The force carrier particle for the EM interactions is the photon. Depending on the energy carried, its nomenclature varies such as radio waves, IR waves, visible light, etc... The photon is a massless particle and travels at the speed of light in a vacuum. At the atomic and molecule level, residual EM interactions can be seen, where electrons in one atom are attracted to the protons in another. Though it is a long-range force, its strength reduces with the inverse of the square of the distance. EM interactions are the result of a U(1) symmetry in the SM[2].

Strong interactions

The strong interaction makes quarks bind tightly into hadrons. This is a short distance interaction and is represented by the gauge group SU $(3)^3$. Strong force employs color charge (on quarks), red, blue, and green (and anti-color on anti-quarks), and the carrier particle is the gluon. Quarks under the influence of strong interactions, always arranged in combination, where the result is colorless. This can be obvious in both baryons and mesons.

Weak interactions

The weak interaction has its influence in a wider spectrum of phenomena where it influences on quarks, leptons, and even on charge less particles. Many decays of particles (such as β decay) into lighter ones can be understood through weak interactions. Weak interaction treats a given generation of quarks as a single particle with two states and therefore, it can operate on a generation of quarks and transform the flavor of the quarks.

[2]Refer to Appendix A.4
[3]Refer to Appendix A.5

This interaction has the shortest range of 10^{-18}m and is represented by the SU (2) group [4]. W^+, W^-, and Z^0 bosons are the force carrier particles.

Higgs Boson

The Higgs boson is a fundamental particle in the SM which is a quantum excitation of the Higgs Field. The Higgs field is an energy field that permeates in the entire universe and is essential to explain how gauge bosons W^+, W^-, and Z^0 get their corresponding masses. Without the Higgs field, all bosons would be massless but, experimental results verify that gauge bosons do have large masses. Higgs boson is behaving slightly different from the other bosons described in the SM family by having "zero" spin (while other bosons have a spin of 1) and it is chargeless and colorless. Being a heavy particle (125-127 GeV/c^2), Higgs boson decays into other SM particles almost immediately.

The existence of the Higgs boson is validating the SM through the mass generation mechanism (where Higgs filed is involved) and opens a broad spectrum for us to explore. Its existence allows us to look for advanced extensions for the SM and eventually will help to explore beyond the horizon of SM (Beyond Standard Model - BSM). One such face of new physics is Vector-Like quarks (**VLQs**), which is the main attention of the search we are involved in.

Force	Strength	Boson	Spin	Mass/GeV
Strong	1	Gluon g	1	0
Electromagnetism	10^{-3}	Photon γ	1	0
Weak	10^{-8}	W boson $W^{\pm}$	1	80.4
		Z boson $Z^{\pm}$	1	91.2
Gravity	10^{-37}	Graviton G	2	0

Table 1.4: The four fundamental forces and their approximately relative strengths. Values are for two fundamental particles kept at a distance of 1 fm $= 10^{-15}$ m [73]

[4] Refer to Appendix A.5

1.1.3 Beyond Standard Model Physics(BSM)

While the SM is successful in supporting many theoretical predictions (in return validating the SM) being made over the past decades, there are several instances where it becomes deficient. A few such examples can be given as,

- CP violation
- Gravitational interaction inclusion
- Dark matter and Dark Energy
- Mass hierarchy problem

The explorable universe mainly consists of matter while their counter-parts (while the universe mostly consists of dark energy and dark matter), the anti-matter is scarce to find. Each particle in the SM has its anti-particle (ex: e^- and e^+) and therefore, laws of physics are to expect a symmetry between matter and anti-matter. This inequality can be addressed via CP violation.

The SM is combining three of the four fundamental interactions electromagnetic, strong and weak yet due to its weak interaction strength, gravity requires higher energies to be included in the SM. This is allowing the physicists to explore new frontiers.

The universe is filled with matter and is expected to slow down in its expansion due to the gravitational pull among the matter. Yet, scientists have observed that the universe is rather expanding at an accelerated rate. This gave rise to the ideology of dark matter and dark energy where they are being held responsible for the observed accelerated expansion of the universe. It is estimated that $\approx$ 68% of the universe is dark energy, $\approx$ 27% of the universe is dark matter and $\approx$ 5% of the universe is regular matter which is explained by the SM. The DM doesn't interact with the interactions described in the SM such as EM and the only interaction so far infer is gravitational interaction. And dark energy too, can not be explained by the SM.

Measurements of the Higgs mass (125GeV) turned out to be far less than one might expect, making its mass in the order of electroweak scale rather than that of the plank scale. The Higgs mass was expected to be heavy with the quadratically diverging quantum-loop corrections thus there has to be either fine-tuning terms for the SM parameters or there is more in the universe to search for (or new physics) that can cancel the divergences.

1.1.4 Vector-Like Quarks(VLQs)

In addressing the shortcomings in the SM, many theoretical models for BSM physics are proposed and one common feature many of them have is the prediction of a new **heavy quark**. The three prominent features of such a heavy quark are,

- Spin is $\frac{1}{2}$
- Under SU(3) transformations, transform as triplets (they have color charge)
- Does not show a chirality preference as SM quarks when interacting with SU(2) isospin weak group

The VLQs are gaining attention among the particle physics community due to several factors. From an experimental point of view, VLQs are the simplest colored fermions that are allowed by experimental data and their addition to the SM could give rise to many new frontiers to study for both theoretical and experimental scientists such as new sources for CP violation.

1.1.4.1 Why the name "Vector-like"?

When the SM quarks undergo $SU(2)_L$ transformation they do so only with left-handed chirality. Consider the charge current lagrangian,

$$\mathcal{L}_W = \frac{g}{\sqrt{2}}(J^{\mu^+}W_\mu^+ + J^{\mu^-}W_\mu^-) \tag{1.1}$$

Since the SM chiral quarks intract with $W^{\pm}$ boson left-handedly,

$$J^{\mu^+} = J_L^{\mu^+} + J_R^{\mu^+} \tag{1.2}$$

where,

$$J_L^{\mu^+} = \bar{u}_L \gamma^\mu d_L = \bar{u}\gamma^\mu \frac{(1-\gamma^5)}{2} d = \frac{\bar{u}\gamma^\mu d - \bar{u}\gamma^\mu\gamma^5 d}{2} = V - A \tag{1.3}$$

and,

$$J_R^{\mu^+} = 0 \tag{1.4}$$

But, on the other hand, VLQs interact with weak interactions in such a way that there is a symmetry in left and right-handedness, which is resulting in a "Vector-Like "charge current see eq. (1.5) . Thus, gets the name Vector-Like quarks.

$$J^{\mu^+} = J_L^{\mu^+} + J_R^{\mu^+} = \bar{u}_L \gamma^\mu d_L + \bar{u}_R \gamma^\mu d_R = \bar{u}\gamma^\mu d = V \tag{1.5}$$

Many BSM theories such as Composite Higgs and Little Higgs are using VLQs in their quantum loop diagrams to cancel the quadratic divergence that arises due to radiative corrections to the Higgs mass during the processes of Higgs production and Higgs decay [27] [71] .

1.1.4.2 VLQ mixing with SM quarks

VLQs could come in seven possible gauge-covariant multiplets including singlets, doublets, and triplets.

$$\begin{array}{llll} T_{L,R}, & B_{L,R}, & & (singlets), \\ \Big(XT\Big)_{L,R}, & \Big(TB\Big)_{L,R}, & \Big(BY\Big)_{L,R}, & (doublets), \\ \Big(XTB\Big)_{L,R}, & \Big(TBY\Big)_{L,R}, & & (triplets). \end{array} \tag{1.6}$$

The new quarks T and B are heavy yet similar to the top and bottom quarks in the SM. They carry a fractional electric charge of $\frac{+2}{3}$ and $\frac{-1}{3}$ respectively. Also, both X and Y too are similar to SM top and bottom, though they have different fractional charges to that of T and B, $\frac{+5}{3}$ and $\frac{-4}{3}$ respectively. The SM top and the bottom quarks are the heaviest among the SM quarks and therefore are preferred by the VLQs to couple to them strongly. Therefore, the VLQs are commonly referred to as "top partners".

This dissertation focuses on the decay of VLQ T and thus, the mixing of the VLQ T with SM top(up type) is described below. The same structure is available for VLQ B when coupling with SM bottom.

$$\begin{pmatrix} t_{L,R} \\ T_{L,R} \end{pmatrix} = \begin{pmatrix} \cos\theta^u_{L,R} & -\sin\theta^u_{L,R} e^{i\phi_u} \\ \sin\theta^u_{L,R} e^{i\phi_u} & \cos\theta^u_{L,R} \end{pmatrix} \begin{pmatrix} t^\alpha_{L,R} \\ T^\alpha_{L,R} \end{pmatrix} \tag{1.7}$$

Angle θ is standing for the mixing angle, ϕ is the phase factor and the α is for the weak eigenstate in the mixing. The introduction of the VLQs as an extension to the SM has ramifications. The changes to the CKM matrix due to VLQs mixing affects the couplings between SM quarks and the SM electroweak bosons Z^0,$W^\pm$, and H^0. Such couplings of VLQs with SM gauge bosons are essential in introducing higher-order quantum loop corrections to the Higgs boson mass studies. The relationship between the VLQ singlets, doublets, and the triplets with the Yukawa couplings(y^q_{ij}; $q = u, d$) and mixing angles can

be expressed as follows.

$$\begin{aligned}\tan 2\theta_L^q &= \frac{\sqrt{2}|y_{34}^q|\upsilon M_0}{(M_0)^2-|y_{33}^q|\upsilon^2/2-|y_{34}^q|\upsilon^2/2} \quad (singlets, triplets),\\ \tan 2\theta_R^q &= \frac{\sqrt{2}|y_{43}^q|\upsilon M_0}{(M_0)^2-|y_{33}^q|\upsilon^2/2-|y_{43}^q|\upsilon^2/2} \quad (doublets)\end{aligned} \tag{1.8}$$

M_0 and υ given in eq. (1.8) are standing for the bare mass of the mixing VLQ mass and the vacuum expectation value.

A general Lagrangian for a given VLQ can be given as follows, which explain it's mixing with the SM quark(s) via SM gague bosons [24],

$$\begin{aligned}\mathcal{L}_H &= -\tfrac{gm_Q}{2m_W}\bar{q}\big(Y_{qQ}^L P_L + Y_{qQ}^R P_R\big)QH + h.c.\\ \mathcal{L}_W &= -\tfrac{g}{\sqrt{2}}\bar{Q}\gamma^\mu\big(V_{Qq}^L P_L + V_{Qq}^R P_R\big)qW_\mu^+ + h.c.\\ &\quad -\tfrac{g}{\sqrt{2}}\bar{q}\gamma^\mu\big(V_{qQ}^L P_L + V_{qQ}^R P_R\big)QW_\mu^+ + h.c.\\ \mathcal{L}_Z &= -\tfrac{g}{\sqrt{2}c_W}\bar{q}\gamma^\mu\big(\pm X_{qQ}^L P_L \pm X_{qQ}^R P_R\big)qZ_\mu + h.c.\end{aligned} \tag{1.9}$$

Here in eq. (1.9), $P_R = \frac{1}{2}(1+\gamma^5)$ and $P_L = \frac{1}{2}(1-\gamma^5)$ are the chiral projection operators, c_w is the cosine of the weak angle the modified CKM matrix (due to the VLQ couplings introduction) explains the couplings $V^{L,R}, X^{L,R}, Y^{L,R}$. The $V_{qQ}^{L,R}$ gives the heavy-light couplings to the W boson, $X_{qQ}^{L,R}$ gives the heavy-light couplings to the Z boson and $Y_{qQ}^{L,R}$ gives the heavy-light couplings to the Higgs boson (if "q "and "Q "are flipped in the aforementioned couplings terms, it corresponds to the light-heavy couplings) for all the signlets, doublet and the triplets combinations. The "Q "and "q "represent heavy and light quarks [24].

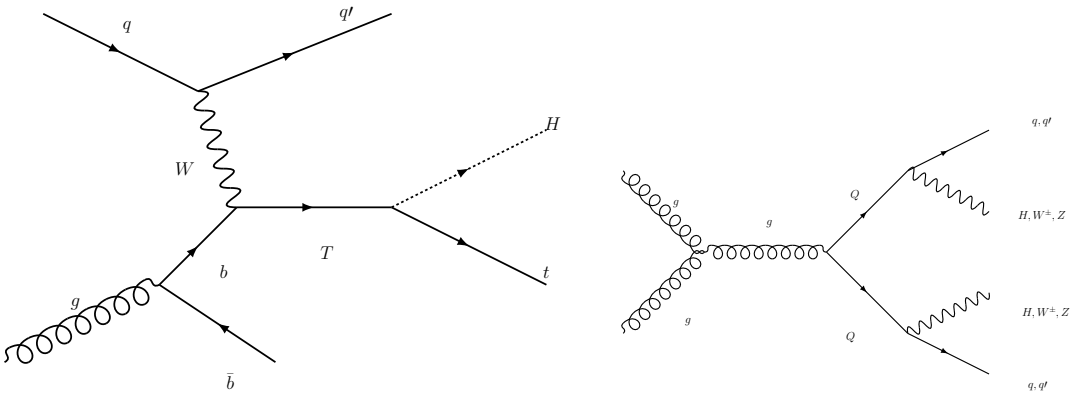


Figure 1.3: VLQ production modes at the LHC single production(left) and pair production(right).

1.1.4.3 VLQs production

There are two distinct ways that the VLQs are produced in the Large Hadron Collider. Single production can become prominent when the mass of the considered VLQ is higher (approx. $>$ 1TeV). Such single produced VLQ interacts with SM quarks and gauge bosons is governed by the electroweak parameters. Even though it has a model dependency, many analyses show a preference towards single produced VLQs as they have a better chance of production at the collision level due to their heavy mass. In comparison, pair production is regarded to be model-independent, allowing researchers to explore many frontiers. Another appealing feature of pair production is, unlike single production, gauge couplings of the VLQs and the gluons are only dependent on the mass of the VLQ in consideration. The expected cross-section for a function of VLQ mass can be observed in fig.1.4 , where pair production cross sections were derived using top++[50] cross-section calculating software and the cross-sections for single produced VLQs were derived using Protos (PROgram for TOp Simulations)[22, 23].

Decaying VLQs can be branched out to various final statuses. For an example VLQ Top can decay as,

$$\begin{aligned} T &\to Ht \\ T &\to Wb \\ T &\to Zt \end{aligned} \tag{1.10}$$

and one can evaluate the branching ratios for each instance based on the existence of the VLQ whether as a SU(2) singlet or a SU(2) doublet. Such calculations of branching ratios as a function of VLQ mass has been carried out using Protos and are shown in fig.1.5

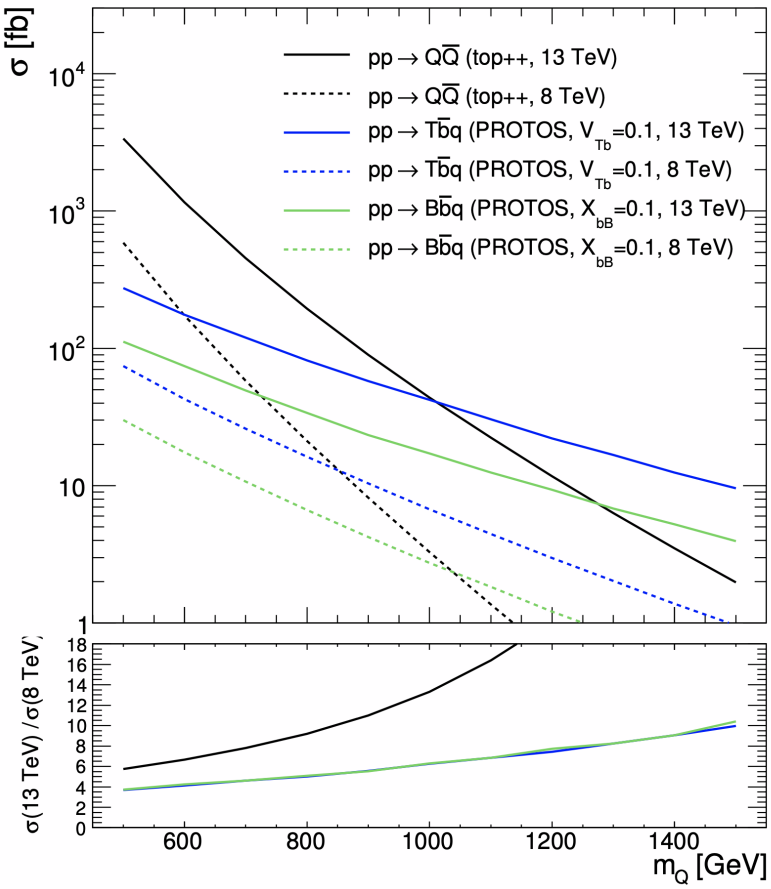


Figure 1.4: VLQ production cross-section as a function of VLQ mass for single and pair productions.

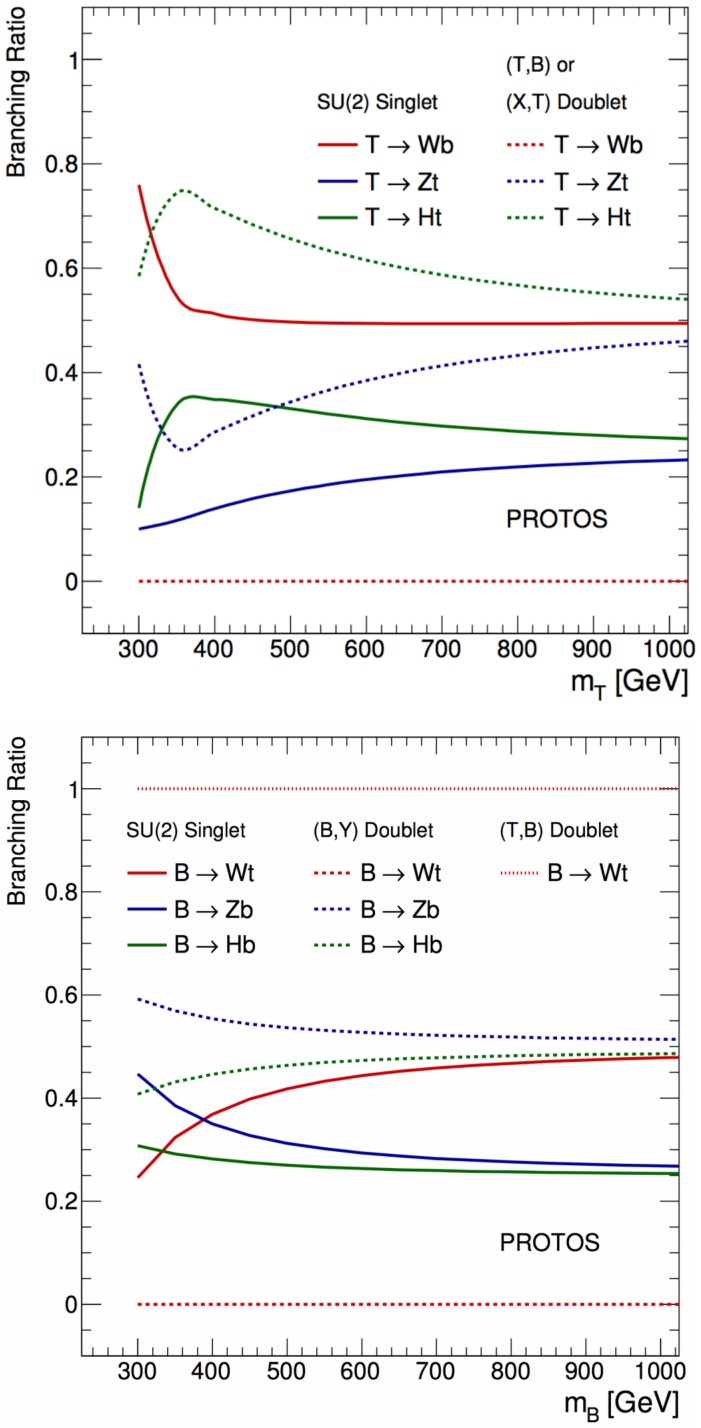


Figure 1.5: Branching ratios as a function of VLQ mass. T branching(top) and B branching(bottom). Singlet decay modes are in solid lines and the doublet decay modes are in dashed lines.

CHAPTER II

The Large Hadron Collider(LHC)

2.1 Introduction

The LHC is the most powerful particle accelerator on the planet. Its construction and development is primarily to address questions unsolved in physics and to improve the understanding of nature. Over 10,000 scientists from over 100 countries and hundreds of universities and research institutes are taking part in the search for new physics. Managed by the European Council for Nuclear Research (Conseil European pour la Reserche Nuclaire, CERN), the complex spreads over a 27 km circular tunnel, that lies 120m below the ground level at the Swiss-French border. It has four main collision points and major detectors are placed at them.

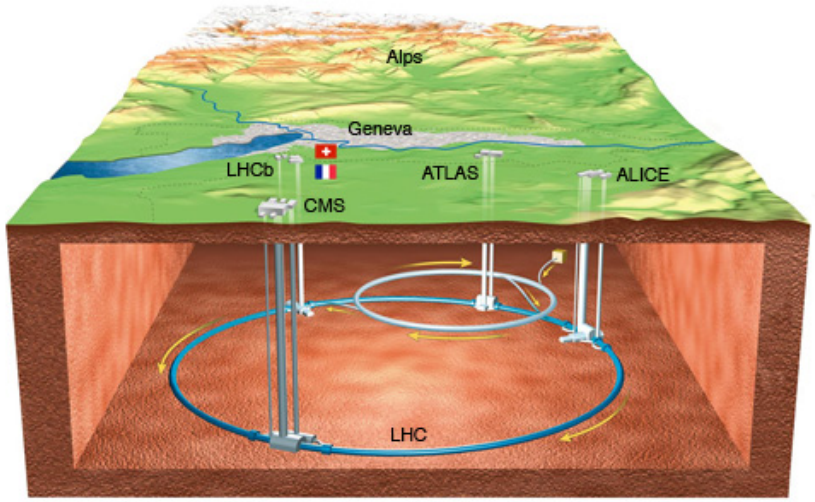


Figure 2.1: The Large Hadron Collider with its four detectors ATLAS, CMS, ALICE and LHCb.

The LHC was developed to broaden our qualitative and quantitative understanding of fundamental particles. Based on the de Broglie relation, the wavelength of a particle is inversely proportional to its momentum, which implies that, in order to probe into smaller wavelengths, momentum should be high. Instead of colliding an accelerated particle on a stationary target, the LHC accelerates two particles in the opposite direction and make them collide. This methodology increases the collision momentum which allows to probe deeper into the particles that are accelerated. For a given circular collider, its maximum obtainable energy depends on the radius of the ring and the strength of the dipole magnetic field used to bend the accelerating particles. Based on the current specifications, the LHC can accelerate a proton beam up to 6.5 TeV energy.

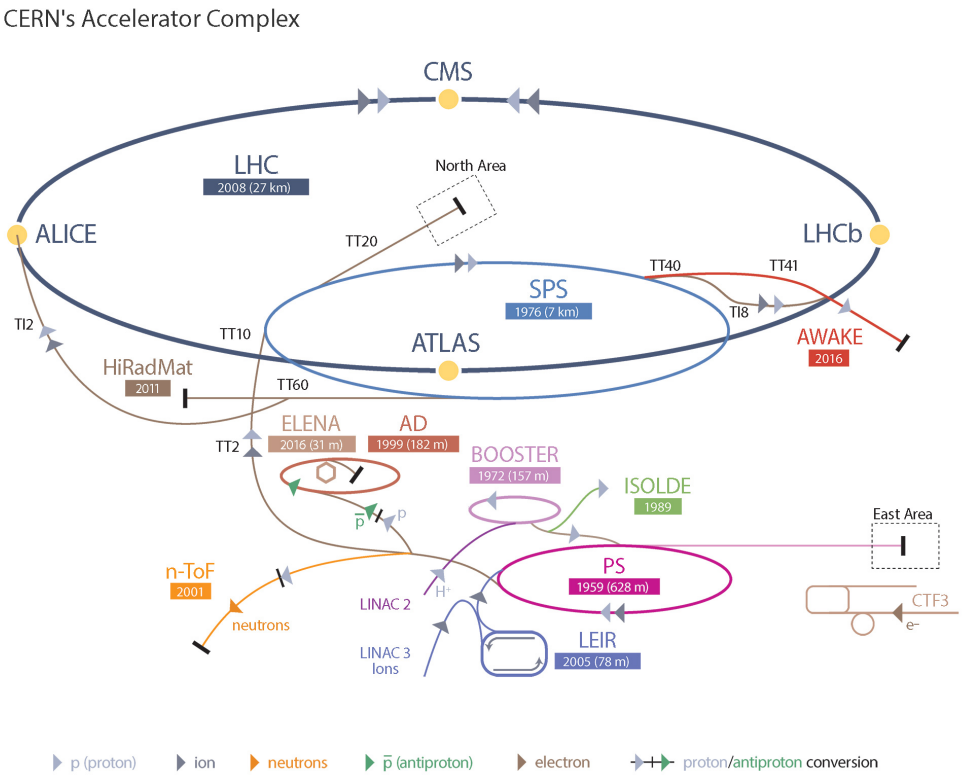


Figure 2.2: CERN's accelerator complex

Currently, the LHC is capable of producing $\sqrt{S} = 13$ TeV energy at the center of mass of a collision of two proton-proton beams. There are four main experimental setups **ATLAS**, **CMS**, **LHCb** and **ALICE**. Out of these four, ATLAS and CMS are general-purpose detectors, LHCb for b-physics studies, and ALICE is focused on heavy-ion collisions.

A series of successive accelerators are used in the CERN accelerator complex before the proton beams are introduced to the final step of the LHC. Protons used to feed the LHC are produced by ripping the electron from the Hydrogen. The protons begin their journey at the linear accelerator (LINAC 2) and get accelerated up to 50 MeV. Then the protons get transferred to the Proton Synchrotron Booster and increase its energy up to 1.4 GeV. Then protons get into the Proton Synchrotron (PS) and increase their energy up to 25 GeV. Then they move to the Super Proton Synchrotron (SPS) and further increase their energy up to 450 GeV. Protons accelerated at the SPS will then be injected into the two beam pipes of the LHC and they will further accelerate inside the LHC (in opposite directions) to 6.5 TeV each. The LHC will make the counter-rotating beams collide at four different locations where the detectors are located as shown in fig 2.2. We have used the data generated from the LHC accelerating the protons to an energy of 6.5 TeV (Run 2) and is expected to increase the energy to 7 TeV by the Run 3.

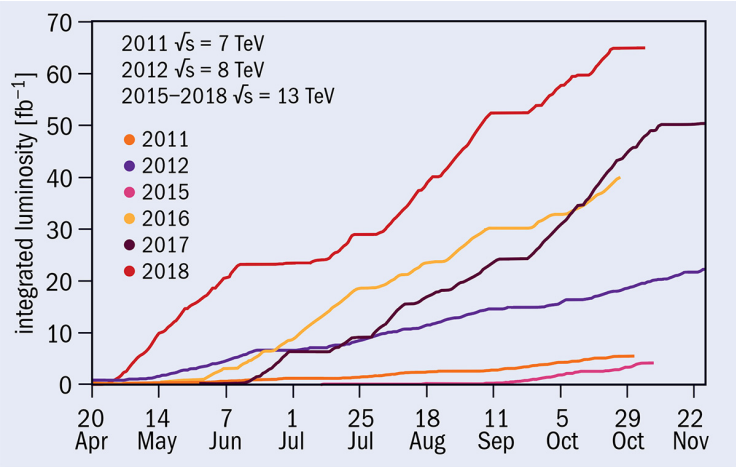


Figure 2.3: Integrated luminosity delivered by the LHC in each operational years - average of ATLAS and CMS

2.1.1 ATLAS Detector

The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector is one of the most advanced and modern particle detectors in the world.

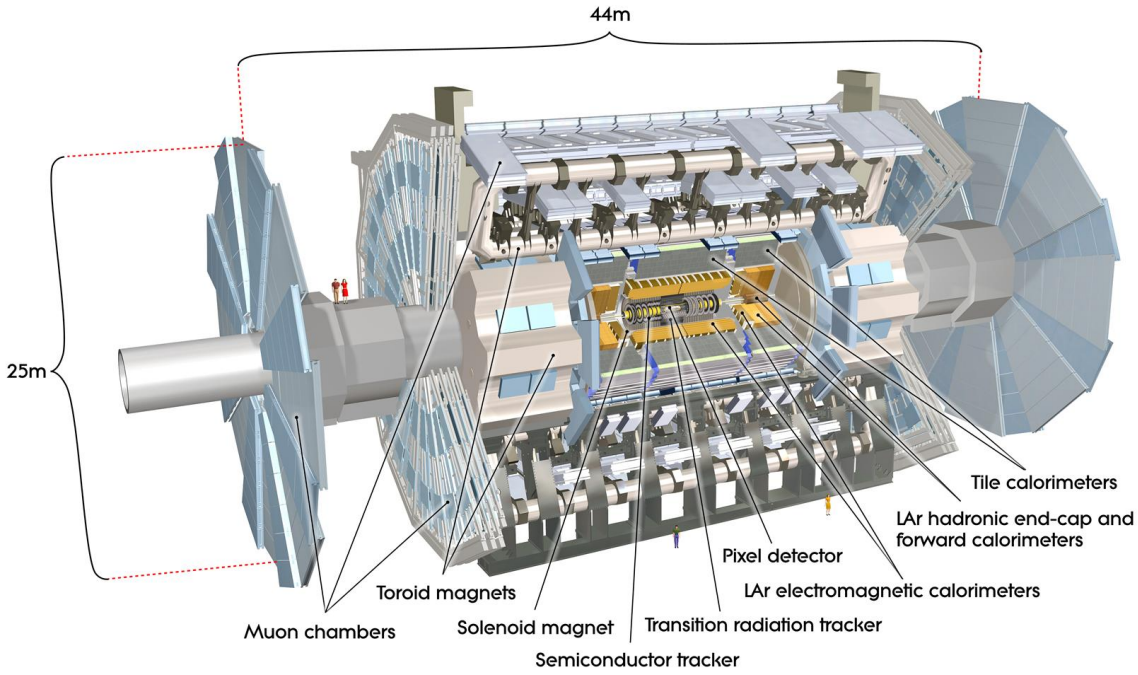


Figure 2.4: The ATLAS Detector

As shown in the figure 2.4, the ATLAS detector is 44 m in length and 25 m in height. The beamline pipe of the LHC is running through the center of the detector and it's placed at a collision point of the LHC. Thus, when particles are colliding, new physics can be studied from the detector. Various detection systems included in the detector will then record energy depositions and traces of particles produced in great numbers.

For all measuring purposes, ATLAS employs a coordinate system where its origin is at the nominal interaction point (IP). The direction of the accelerated particles in the beam pipe is considered as the positive z-axis, the axis point from the IP to the center of the LHC ring considered as the x-axis and the axis pointing toward the surface of the earth is the y-axis. X-Y plane is known as the "Transverse plane". Calculations related to the transverse plane are based on cylindrical coordinates r - ϕ. Azimuthal angle ϕ is measured

from the x-axis around the beam pipe. The polar angle θ described as the angle from the positive z-axis and radial dimension, r, measured as the distance from the beamline. Pseudorapidity (η) is an important term related to the detector and is defined as

$$\eta = -ln\Big[\tan\Big(\frac{\theta}{2}\Big)\Big]. \tag{2.1}$$

In hadron collider physics, η gets is more convenient to use than the polar angle θ because differences in pseudorapidity are Lorentz invariant and production of particles is constant as a function of pseudorapidity. The difference of distance (angular separation of particles) in space is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. When it comes to measuring the physical momenta of a particle produced in the collider, terms transverse momentum p_T, polar angle in the transverse plane ϕ, and the pseudorapidity η are required. Thus, one can obtain the cartesian momenta (p_x,p_y,p_z) as follows. [70]

$p_x = p_T \cos\phi$

$p_y = p_T \sin\phi$

$p_z = p_T \sinh\eta$ (along the beam axis)

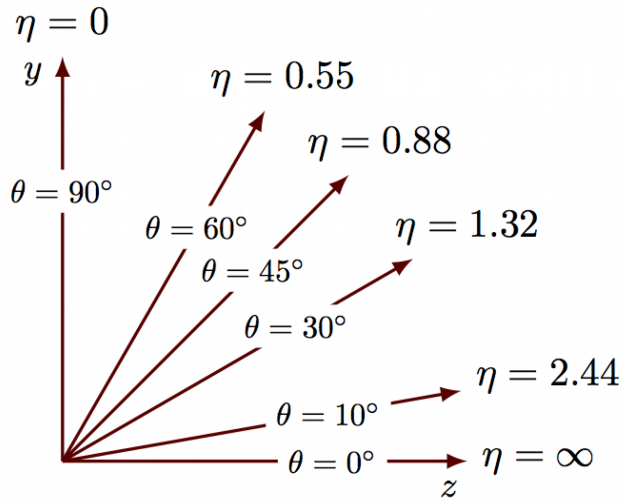


Figure 2.5: Pseudorapidity η and its relation to the polar angle θ. z axis is the beam direction and y axis is directing upwards.

The ATLAS detector is an assembly of various detectors, with each designed to achieve a specific task. Referring fig 2.4 shows us that there three major detecting components available at ATLAS. They are,

1. Inner Detector (ID)
2. Calorimetric System
3. Muon Spectrometer (MS)

The Inner Detector (ID) is the closest to the beam pipe and is designed to detect tracks left by the charged particles. The Calorimetric System then laying around the ID. It's designed to be sensitive for both charged and neutral particles such as hadrons, electrons, and photons. At the outermost layer, we find the Muon Spectrometer (MS) which is designed to detect trajectories of muons that were created by particle collisions.

2.1.1.1 Inner Detector (ID)

The ID (figure 2.6) is designed to be sensitive in such a manner that it provides high precision momentum resolution and is also capable to make fine primary and secondary vertex measurements for the tracks of charged particles that are above a certain level of P_T (usually above 500 MeV) within $\mid \eta \mid < 2.5$. Due to the high particle density near the ID, it can make measurements at high-precision. Thus, its capability can be used to reconstruct trajectories of charged particles. The ID has 3 major components.

1. Pixel Detector
2. Semi-Conductor (silicon microstrip) Tracker (SCT)
3. Transition Radiation Tracker (TRT)

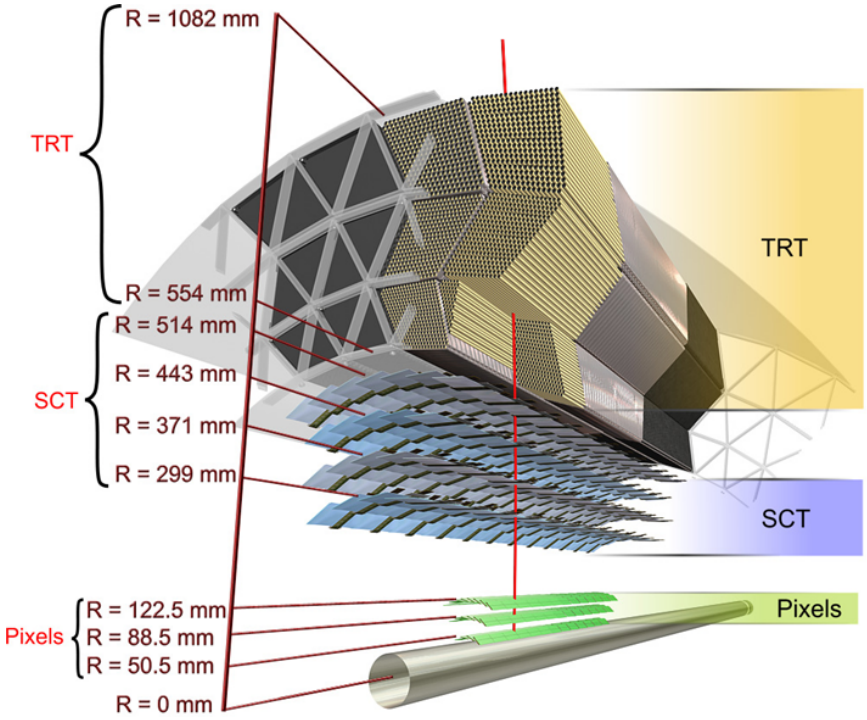


Figure 2.6: Inner Detector

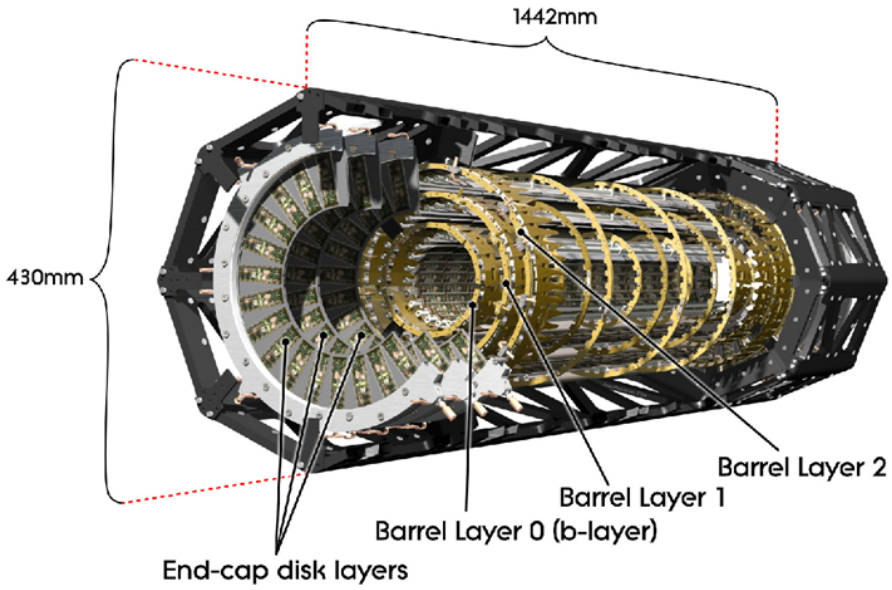


Figure 2.7: Schematic of Pixel Detector

2.1.1.1.1 Pixel Detector (PD) Silicon pixels are used in the PD (figure 2.7) to obtain high-precision measurements for the charged particles that are available near to the point of interaction. It has over 80 million channels and is capable to provide three precise measurements per charge particle on average. The PD is arranged in a central barrel module concentric with the beam pipe and has two end-cap modules perpendicular to the beam axis. As shown in figure 1.6, the pixel detector barrel consists of three layers of barrels, and each barrel has an end-cap contains three wheels. Lying close to the beam pipe, it spreads from 50 to 122.5 mm region in the ATLAS detector. To enhance the fine resolution of charged particle detection, it uses 1744 individual sensors, and single sensor is equipped with 46080 readout channels resulting 80.4 million channels. When a charged particle interacts with a sensor, it creates free electrons which then will be read out by the electrodes in the chip. A hit is registered when enough energy is stored in the sensor.

2.1.1.1.2 Semi-Conductor Tracker (SCT) The SCT (figure 2.8) detects charged particles similar to the PD. It's capable of making eight precision measurements per track and can measure momentum, impact parameter, and vertex position of charged particles in a radial range of 299-514 mm. There are four layers of concentric modules in the central barrel region and it has two sets of end-cap wheels. A single-sided wafer (6 cm in length) glued together on opposite sides at each layer is used as the detection element and it has 768 readout channels. Thus, it provides a total of 6.3 million channels. When a charged particle travels through the detector element it creates free electrons that are read by the electrodes and counted as a hit.

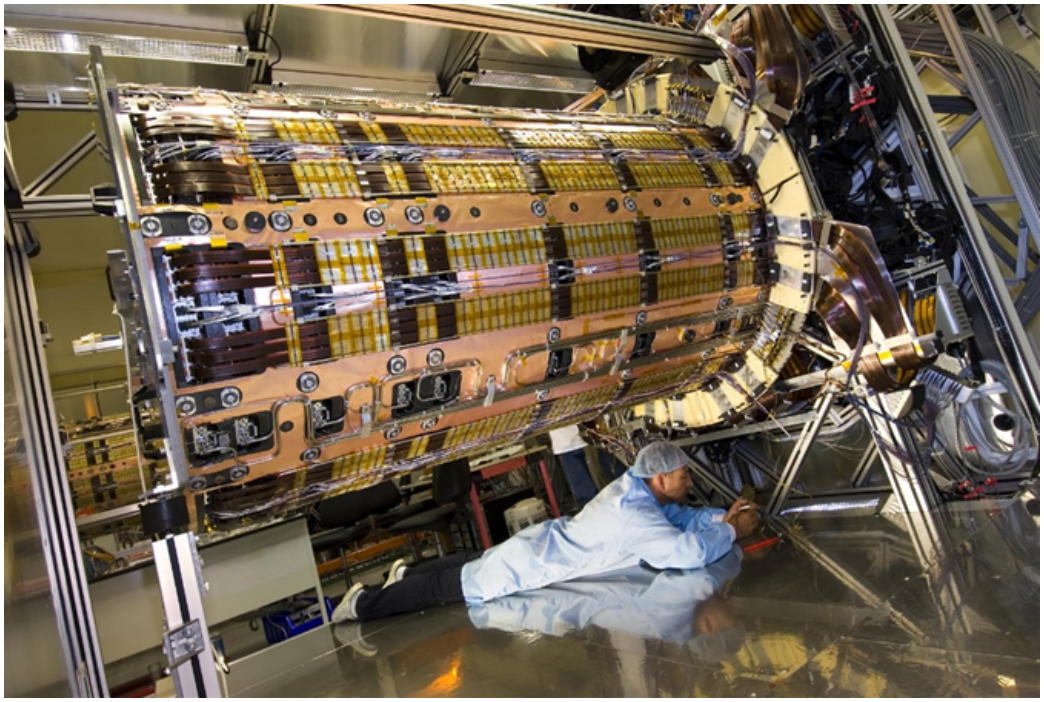


Figure 2.8: Semi-Conductor Tracker

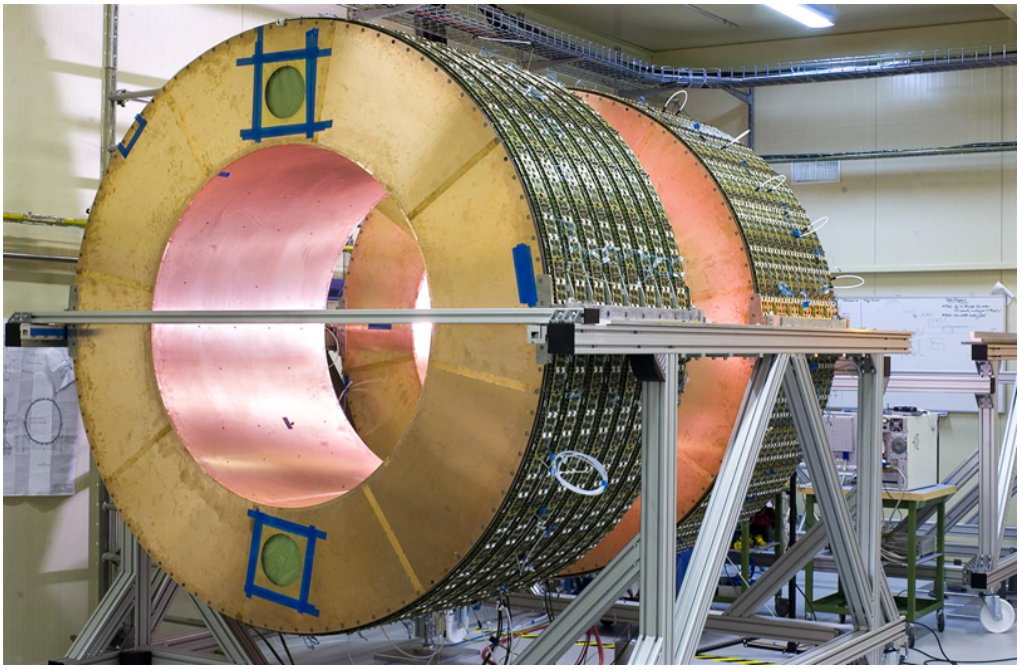


Figure 2.9: Transition Radiation Tracker

2.1.1.1.3 Transition Radiation Tracker (TRT) The last to come in the inner detector system is the TRT as in fig 2.9, which uses straw tube detectors for particle detection. Each cylinder is filled with gas (a mixture of 70% Xe, 27% CO_2 and 3% O_2) with a long anode wire running down the center and the tube acting as a cathode that has a diameter of 4mm. When a charged particle or photons pass through the gas tube, they ionize the gas, resulting in a current. A straw tube is capable to register 36 hits/track. The TRT barrel region has three concentric barrels and 96 modules are arranged in layers. There are 40 end-cap wheels for the TRT and it spreads a radial distance of 554-1082 mm. Thus, the TRT has 351,000 readout channels ready for particle detection.

2.1.1.2 Calorimetric System (CS)

The CS as in fig 2.10 in ATLAS is used to measure the energies of electrons, photons, and hadrons and is also vital to measure the E_T^{miss} of a given event. The CS is capable of providing a coverage of $\mid \eta \mid < 4.9$. The calorimetric system uses two calorimeters, Electromagnetic and Hadronic, which confine most particles before reaching the muon detectors. These calorimeters are "Sampling"calorimeters.

There are alternating layers of absorbers and active materials in the sampling calorimeters and they allow to make measurements of ionizing energies due to the interactions with particles. When particles interact with the absorber, a shower of secondary charged particles is produced that is then measured by the active material. The number of produced secondary particles depends on the energy carried by the incoming particles. Losing all the energy of the incoming particle within the calorimeter is vital for an accurate measurement.

Thus, the calorimeter is designed in such a way that it should be able to prevent the "punch through effect" for the incoming particles. In order to achieve this, two parameters are defined: interaction length (λ), defined as the average distance for a hadron to undergo a hadronic interaction and radiation length (X_0), defined as the average distance traveled

by an electron before electromagnetic interaction in the calorimeter. Referring to figure 2.10, Electromagnetic calorimeter is placed close to the interaction point and is followed by the Hadronic calorimeter.

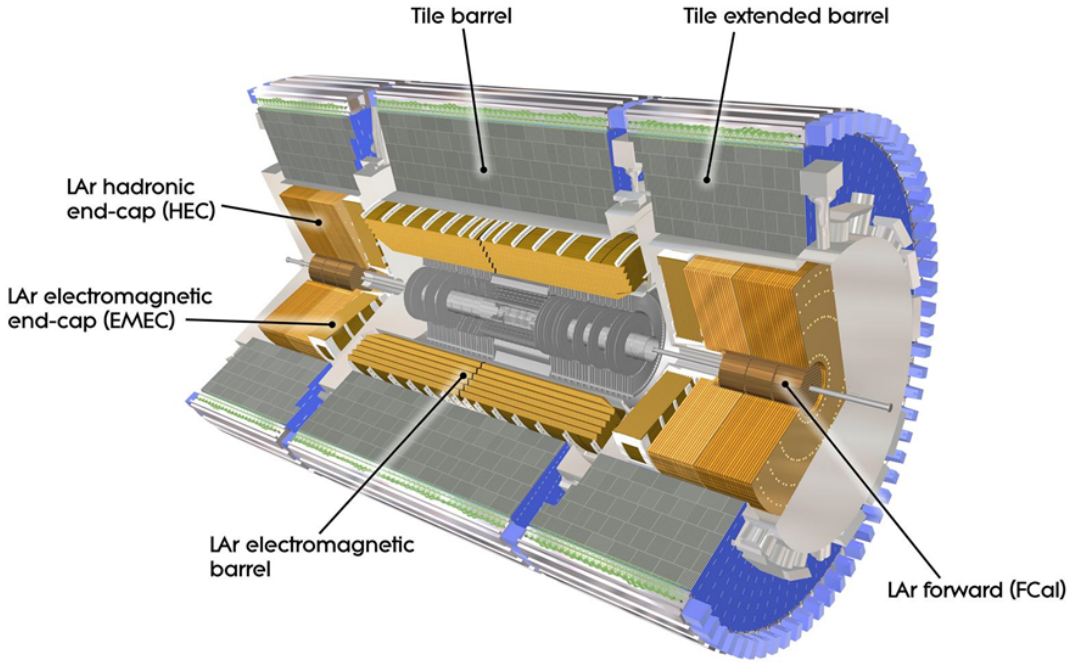


Figure 2.10: Full Calorimetric System

2.1.1.2.1 Electromagnetic Calorimeter (EMC) This is responsible for the energy measurement of the electromagnetically interacting particles up to $\mid \eta \mid < 4.9$. ATLAS has chosen a liquid-Argon/lead detector as its EMC. The EMC measures the EM particle energy (from charged particles and photons) through the energy deposition due to the ionization and absorption. Its structure has two end-cap components that fit into a barrel part. It looks like an accordion in shape and this shape helps the calorimeter to provide complete azimuthal coverage without any cracks. Lead absorbing plates are cladded with stainless steel plates to secure a proper absorption and provides a smooth surface. Three sampling layers in the very center of the calorimeter ($\mid \eta \mid < 2.5$) are providing a platform to make precise measurements. The region ($2.5 < \mid \eta \mid < 3.2$) which covers the overlap pf end-cap and barrel components of the EMC employ two sampling layers. Finally the forward calorimeter cover the region ($3.2 < \mid \eta \mid < 4.9$). The Electromagnetic calorimeter has a capability of 24 radiation lengths in the barrel region, 26 at the end-cap region, and 10 interaction lengths.

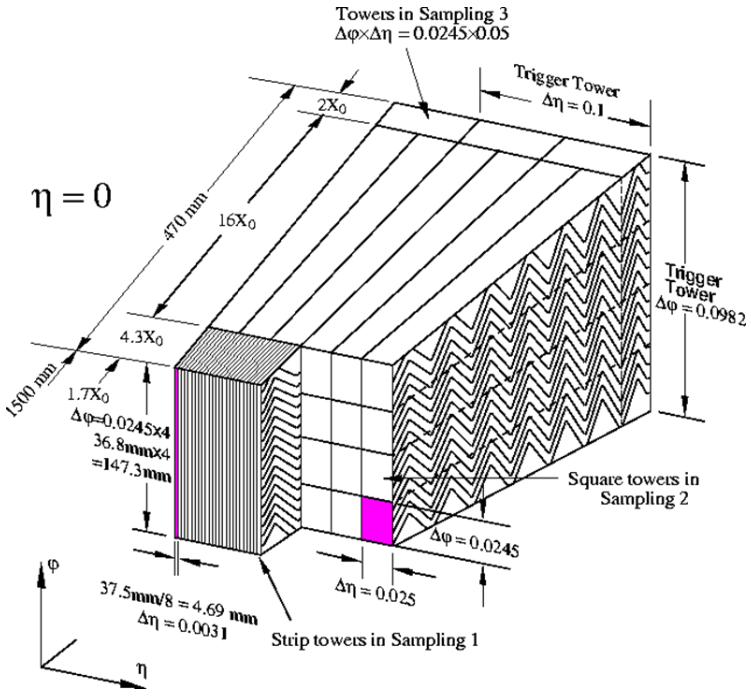


Figure 2.11: Electromagnetic Calorimeter

2.1.1.2.2 Hadronic Calorimeter (HC) The HC is designed to address two main functions.

1. Measure the directions and energies of "jets"[1]

2. Verify the presence of particles (such as neutrinos) that escape ATLAS undetected by studying the imbalance in total transverse momentum

The three major components of the HC are,

1. Scintillating tile calorimeter

2. Hadronic end-cap calorimeter

3. Forward calorimeter

Closer to the interaction point is the scintillating tile calorimeter and it is designed to detect energies of hadronic particles within the range of $\mid \eta \mid < 1.7$. It is then followed by the hadronic end-cap calorimeter which is providing a coverage of $1.5 < \mid \eta \mid < 3.2$. This is made with two wheels per end-cap and it is pointing towards the EM calorimeter. The forward calorimeter covers a range of $3.1 < \mid \eta \mid < 4.9$. Among all the three, the SCT and the hadronic end-cap calorimeters have a higher granularity in $\eta \times \phi$ space when compared with the forward calorimeter. the scintillating tile calorimeter has three longitudinal sampling layers and they have 1.4, 4.0, and 1.8 interaction lengths at η=0. The hadronic end-cap calorimeter has an interaction length of 12 and the forward calorimeter has an interaction length of 9.5. [68].

[1] A jet is a cluster of particles resulting from the hadronization of high-energy quarks/gluons.

Figure 2.12: Scintillating tile calorimeter

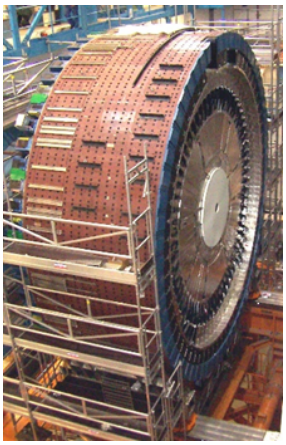


Figure 2.13: Hadronic end-cap calorimeter

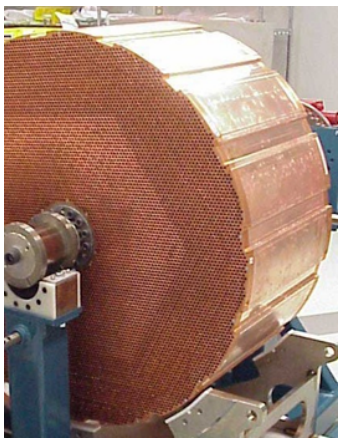


Figure 2.14: Forward calorimeter

2.1.1.3 Muon Spectrometer (MS)

Especially design for the detection of muons, the MS is a robust tracking system that uses four different technologies to produce high-precision tracking. They are,

1. Monitored Drift Tubes (MDT)
2. Cathode Strip Chambers (CSC)
3. Resistive Plate Chambers (RPC)
4. Thin Gap Chambers (TGC)

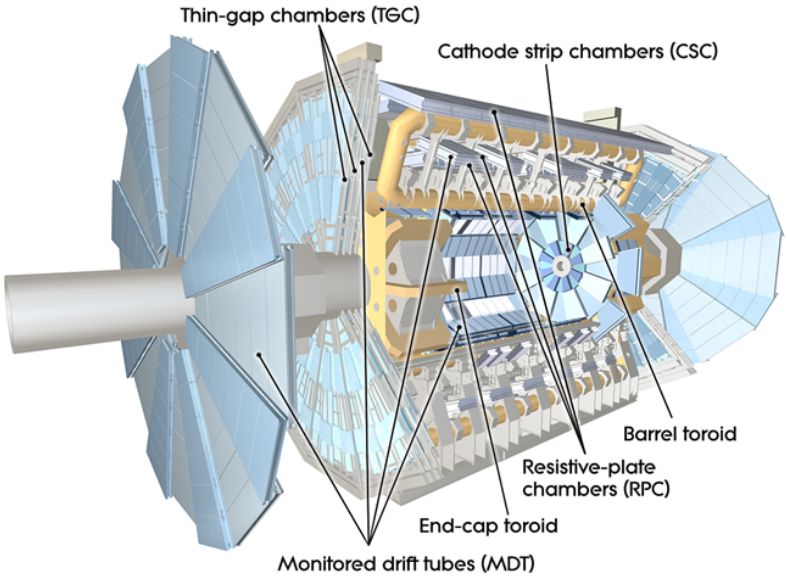


Figure 2.15: Muon System

The MS uses the RPC and TGC to trigger on events with high p_T muons in the range of $|\eta| < 2.4$. The MDT and CSC are providing tracking of muons in the range of $|\eta| < 2.7$, and provide a resolution of 10% for muons with a momentum 1 TeV. Toroidal magnetic fields in a range of $|\eta| < 1.0$ are provided by the Barrel Toroids and enclose the MS. End-cap magnets provide magnetic field for a region of $1.4 < |\eta| < 2.7$. Three cylindrical layers of chambers (MDT and RPC) can be seen in the barrel region and the end-cap region has four wheels on each side (TGC, MDT, and CSC).

2.1.1.4 Magnets

As ATLAS uses trajectories left by particles to study their properties (such as momentum, charge, etc...), all tracking devices are required to be in a magnetic field. The ATLAS detector uses a solenoid magnet to produce a magnetic field for the inner detector and eight magnets at end-cap and toroidal systems to provide a magnetic field to the muon system. Both systems use superconducting magnets which are kept at a temperature close to 4.5 K.

The solenoid magnet is placed in between the inner detector and the calorimetric system. Running parallel to the beamline, it has a length of 5.8m. This is capable of producing a 2T strong magnetic field at the center of the inner detector. The magnet is placed in the same vacuum space where the calorimetric system is to reduce the passive detector material.

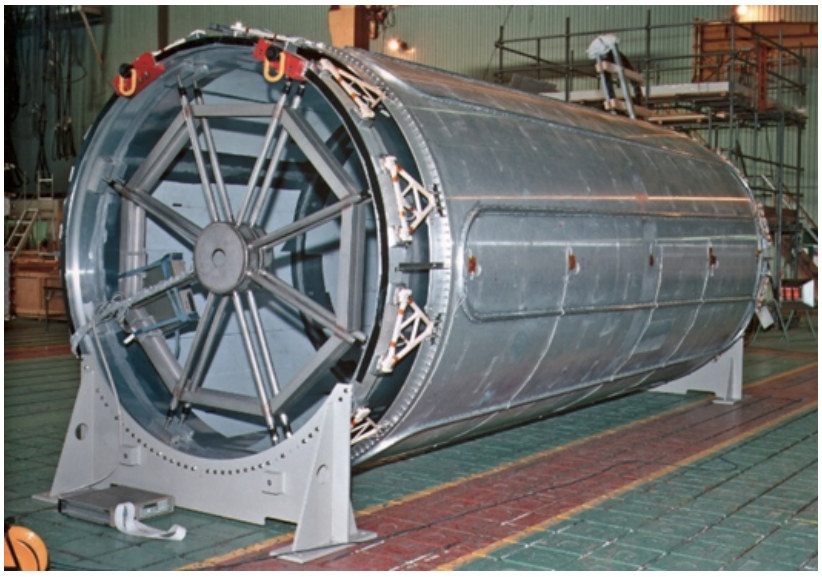


Figure 2.16: Solenoid Magnet

Two end-cap air-core toroids and an air-core barrel toroid belong to the magnet system. These three could have been combined into a single large toroid, but are separate for technical convenience. Each toroid has eight coils that vary from 120 turns for the barrel to 116 turns for end-cap. All the coils in use are superconducting and equipped with cooling circuits and a vacuum system which allows to harness their optimum performance at around 4.5 K. The toroid magnets provide a toroidal magnetic field for the operation of the muon system that is almost perfectly orthogonal to the particle trajectory in any radial direction.

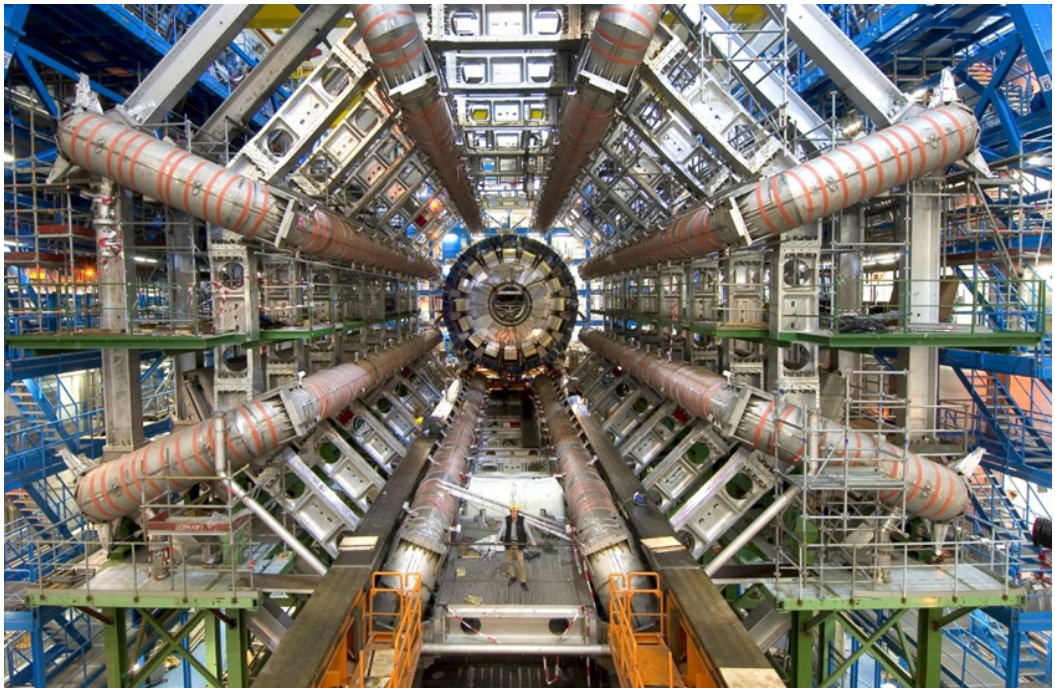


Figure 2.17: Barrel toroid magnet in the underground cavern

2.1.1.5 Trigger and Data Acquisition System (TDAQ)

At operational conditions, the LHC make bunches with 10^{11} of protons to cross each other at 40MHz, resulting in approximately 25 p-p interactions per bunch crossing at the center of ATLAS. Even though there are nearly 2 billion collisions per second, very few events have new physics to study. Therefore, TDAQ has to give priority in saving only the interesting physics events such as,

1. high p_T jets and fermions
2. events with high E_T^{miss}
3. events that have large transverse energy

Thus, TDAQ employ triggers at two different levels level 1 (L1) and the High Level Trigger (HLT) to select and save events for permanent storage. This will ensure that interesting new physics produced at collisions are secured towards further analysis.

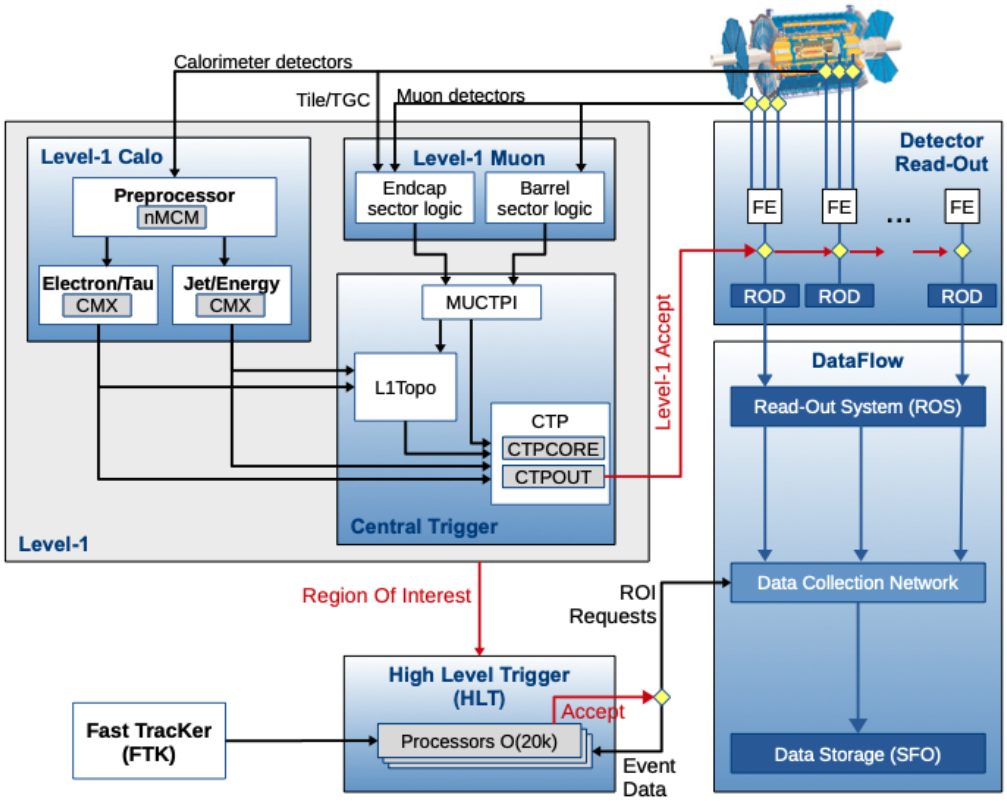


Figure 2.18: ATLAS trigger and data acquisition system for RUN2 [64]

2.1.1.5.1 Level 1 Trigger (L1) The L1 trigger is a hardware-based trigger that rejects the low energy events generated during the collisions. The L1 calorimeter trigger searches for the key features in physics analyses such as leptons (electrons and taus), photons, particle jets, high missing transverse, and total transverse momenta using the information it receives from the electromagnetic and hadronic calorimeters.

The L1 Muon trigger that has two end-cap sections and a barrel section takes the information coming from the muon detectors (high p_T muons) to generate fast trigger signals to the L1 trigger.

The topological trigger (L1 Topo) taking the combined outputs from the L1 calorimeter and L1 muon triggers that are combined using the output merger modules (CMX). The new L1 Topo trigger introduced in RUN2 combines angular separation, invariant mass requirements and total transverse momenta that are coming from the calorimeter and muon triggers to apply topological selections (Regions of Interest-RoI which are the geographical coordinates of the regions in the detector that the selections made by the calorimeter and muon triggers identified to have interesting events. RoI data have information on η, ϕ and p_T measurements that used for events selection). This is useful in suppressing backgrounds expected at various trigger selections used in physics analysis.

Information from the triggers L1 calorimeter, L1 Muon, and L1 Topo are feed into the central trigger processor (CTP) which provides the Level 1 trigger decisions and also feeds the sub-detector readout system with the Level 1 accepted signals and LHC timing signals.

A hardware-based tracking system known as Fast TracKer (FTK) is used to provide a global track reconstruction using the information from the ATLAS silicon tracking detectors and the High-Level Trigger (HLT) will be fed with full-event track information. This is also useful in developing pile-up robust trigger strategies as well.

Upgrades introduced to the L1 trigger during the RUN2 data taking period have reduced the event rate (due to the bunch crossing in the LHC) of 30MHz to 100kHz

whereas during the RUN1 period it was reduced to 75kHz. 25 μs is regarded as the Level 1 decision accept time.

2.1.1.5.2 High Level Trigger (HLT) The ATLAS TDAQ system used to have two separate triggers Level 2 and Event Filter Trigger in further processing stages at the RUN1 and they were merged into a single High-Level Trigger (a software-based trigger) farm during the RUN2. This merge significantly reduced the complexity in the trigger process as well as reduced the data-fetching duplication through code and algorithm duplication. The HLT managed to reduce the CPU resource usage by augmenting the processing inputs within the RoIs. The HLT has an output rate of 1kHz (400 Hz during RUN1) which is reduced from 100 kHz coming from the L1 trigger and has a processing time of 200 ms [64, 28]. Once an event passes through this trigger level are written into different data streams such as,

- Main physics
- Delayed physics
- Express stream
- Trigger level analysis
- Detector calibration

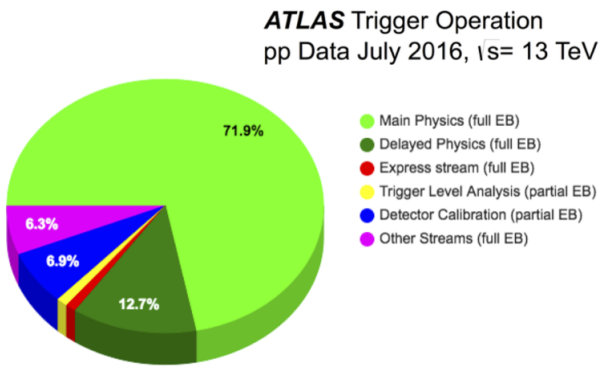


Figure 2.19: Event usage by data streams saved by the HLT in 2016 July [64]

CHAPTER III

Experimental Methods

3.1 Event Reconstruction

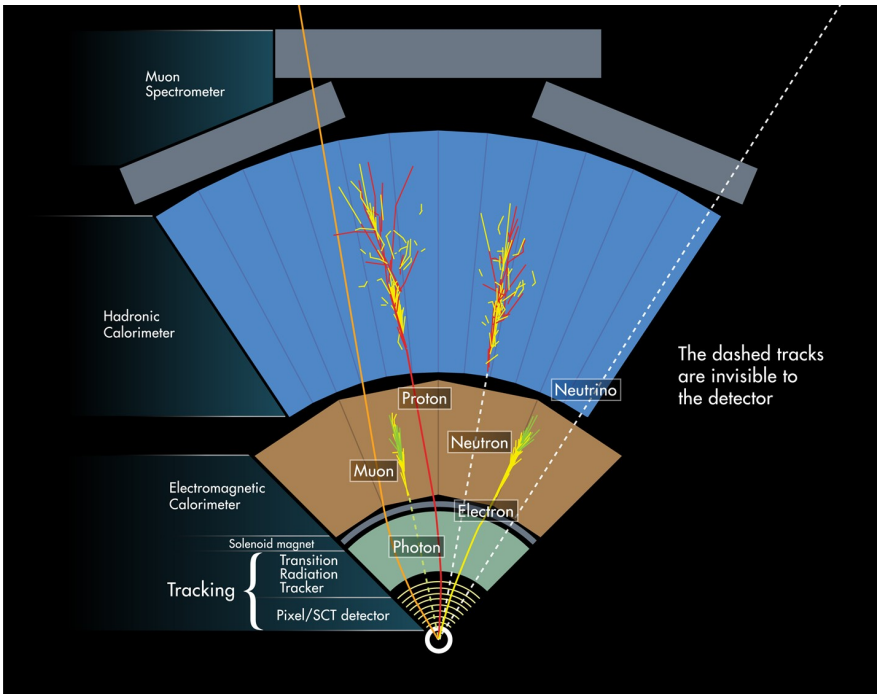


Figure 3.1: Interactions of particles generated from pp collisions with the detector materials which are used for the particle identification [65]

The ATLAS detector as described in the previous chapter is capable of converting collected information from the interactions of highly energized particles with the detector materials into electronic signals or data. In the process, other than particles such as neutrinos, the electrons, hadrons, and photons are interacting with the calorimeters by depositing their energies. The trajectories these particles leave in the detectors can be used along with the applied magnetic field to make observations of the properties of the particles in the study such as their momenta and sign of charge.

As a forensic detective makes deductions based on the shreds of evidence left at a crime site, a group of scientists working on a particular study is required to use the data to make conclusions on their studies. This is done by identifying and reconstructing physics objects such as reconstructed particle trajectory and combining such reconstructed physics objects then could be used for the identification of an analysis requirement(s).

3.1.1 Event simulation

The pp collisions at the LHC produce a plethora of particles and many particle physics analyses are analyzed based on the collected particle data. One concern arises at this stage of an analysis is it's impractical to find the intermediate state of the particular analysis as many of the heavy particles produced at the collision level are decayed almost immediately after their production and the detector is detecting their daughter particles. Thus, it is a common practice to simulate the interesting events and event generators based on "$Monte\ Carlo\ integration$ "method are used in simulating events and they are commonly referred as Monte Carlo event generators.

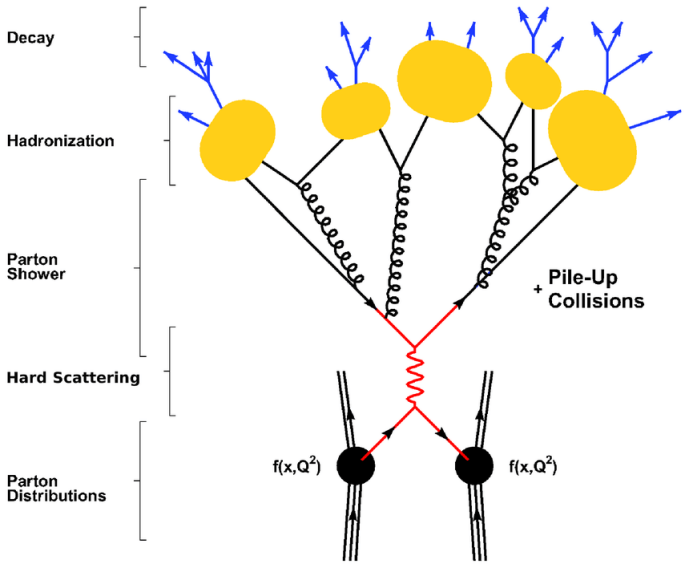


Figure 3.2: Evolution of pp collisions at different stages in the event simulation [29]

Accuracy of the predictions made in physics analyses has a heavy emphasis on how well the events are simulated. Thus, it is essential to simulate events such as hadronization of colored particles, the underlying event, interaction of protons at the collision level, how the final stage particles are interacting with the detector materials, and the simulation of ATLAS detector with greater details. These complex processes are achieved through factorization.

3.1.1.1 Parton factorization

When the two proton beams are colliding at the LHC, protons in the beams are colliding at energies $\sim$ 13 TeV. Such high energies gained by the protons allow them to treat their constituents (the partons) as free particles because they are bound together less strongly than the collision energy due to the effects of asymptotic freedom. Thus, a collision of two protons can be regarded as a collision of two partons in a hard scattering event. When an analysis group is interested in studying a particular physics process, they can derive the cross-section for the production of a final state using the "factorization theorem " [45] as follows,

$$\sigma_{pp\to X} = \sum_{a,b} \int_0^1 dx_a dx_b f_a\left(x_a, \mu_F^2\right) f_b\left(x_b, \mu_F^2\right) \sigma_{ab\to X}\left(x_a p, x_b p, \mu_F^2\right) \tag{3.1}$$

The a and b given in eq. (3.1) are the parton types that can progress towards the final state of X and the $\sigma_{ab\to X}$ is the corss-section for the process of $ab \to X$. The parton distribution function(**PDF**) $f_{i(j)}\left(x_{a(b)}, \mu_F^2\right)$ is describing the probability of a parton $i(j)$ carrying a fractional momentum of $x_{a(b)}$ of the the total momentum of a proton in collision. The hard partonic cross-section, $\sigma_{ab\to X}\left(x_a p, x_b p, \mu_F\right)$ is decribing the $hard\ process$ among the partons interaction. The μ_F^2 is known as the $factorization\ scale$ is used to distinct low momentum, long distance parton density in the proton from the short distance hard process [40]. PDFs do not depend on the physics process in study making them universal and essential ingrediant in event simulations. They are usualy derived

through Deep-inelastic Scatterings observed in hadron colliders.

3.1.1.2 The hard scattering cross section

The cross-section for the process of $\sigma_{ab\to X}$ is providing information on the final state X of the process. If the process of interest is either at Leading Order (LO) in purtabation theory or has fewer of final state particles, only a few matrix elements required to be considered in the determination of $\sigma_{ab\to X}$ and it gets complicated along with the process. In general, the $\sigma_{ab\to X}$ for a n number of orders can be shown as,

$$\sigma_{ab\to X} \sim \underbrace{\sum_{n=0}^{\infty} \int d\Phi_{X+n}}_{\sum_{legs}} \Big| \underbrace{\sum_{i}^{\infty} \mathcal{M}^{i}_{X+n}}_{\sum_{lops}} \Big|^2 \tag{3.2}$$

Summation over all the "real emissions "n are regarded as $legs$ and the summation over all "virtual emissions "i are regarded as $loops$. The dΦ is the phase-space factor and it is regarded to integrate over all additional n legs. The KLN theorem [57] is used to rectify divergences that arise from the loops and legs by doing the cross-sectional calculations at a fixed strong coupling constant α_s.

3.1.1.3 Parton showering

Following the cross-sectional caluculations, parton showering algorihtms are used in studying the effects of higher order QCD contributions which are not covered in the hard process stage. Parton showeing algorithms are used to simulate the radiation of gluons, quarks and bosons from partons. These algorithms assume a collinear splitting of a parton a in to partons b and c can have a differential corss section of [34],

$$d\sigma_{n+1} = d\sigma_{n+1} \frac{\alpha_s}{2\pi} \frac{d\theta^2}{\theta^2} P_{ba}(z,\phi) dz d\phi, \tag{3.3}$$

Angles θ and ϕ are standing for opening and azimuthal for the splitting and P_{ba} is

the probability of parton a splitting into two partons where z is the fraction of parton a's momentum transferred to the parton b in the process. As partons are radiating quarks and gluons, there are three possibilities existing for splitting.

- $q \rightarrow qg$
- $g \rightarrow q\bar{q}$
- $g \rightarrow gg$

MC methods are used to generate values for θ, z, and ϕ by iterating over the possible splittings until two resulting partons individual virtual mass reach the hadronization cut-off energy level of $\sim 1GeV^2$ [75].

3.1.1.4 Hadronization

Once the partons reach to the hadronization cut-off energy level $\mathcal{O} \sim 1GeV^2$, they are beginning to combine to form colorless hadrons. Event generators at this level are inclined to use phenomenological models such as string model and cluster model. Both the models are making the partons to create $q\bar{q}$ pairs as well as the baryons [54, 26, 75].

3.1.1.5 Underlying event and pile-up

As hard scattered partons are providing the data for analyses, it is important to keep them separated from the parton remnants from the proton collisions, which are capable of producing hadrons and interact with the detector. Any hadron generated from pp collisions that have no hard scattering involved in its existence can be regarded as part of the underlying event.

Additional proton-proton collisions in the same bunch crossings are called as pile-up. Pile-up could be treated either as $in- \ time$ when additional proton-proton collisions occurring in the same bunch-crossing or $out- \ of- \ time$: when additional proton-proton collisions occur just before and after the bunch-crossings.

3.1.1.6 Detector simulation

The GEANT 4 particle simulation tool kit [21] is used to simulate the interactions of the final state particles with the detector materials. Electromagnetic and hadronic interactions of the long-lived particles with the detector materials are simulated using a ATLAS detector model [20]. Two simulation formats Fullsim and Fastsim are provided based on analysis requirements. Full sim simulates the interactions of generated particles through detector materials producing energy deposits, interaction period and interaction position in the detector, and as a result, this simulation is consuming a large computation time. As an alternative, ATLFAST II [20] is used to simulate interactions for a simplified detector model.

3.1.2 Physics objcet reconstruction

When the pp collisions are taken place at the detector center, various detecting aspects installed in it are starting to record the events generated. Measurements made by such sub-detectors are then used to reconstruct the stable particles[1] based on various algorithms as needed.

3.1.2.1 Tracks and primary vertex

Stable charged particles generated from pp collisions are depositing their energies while passing through the tracking system as illustrated in fig. 3.1. An $inside-out$ pattern recognition algorithm is used to reconstruct the tracks. Three-dimensional representation of the silicon detector measurements is used to create a $track\ seed$ which is used to build track candidates by using a window search towards the seed direction. Hits recorded at the detectors in the ID that fall into the road window are collected and judged using a simplified Kalman filtering. Once a hit is included in a track candidate, the reconstructed track is fitted in a χ^2 minimization for track fitting. There is a chance of mixing fake

[1] Particles with a minimum mean lifetime(τ) $\approx$ 30ps are regarded as stable particles

tracks and overlapping tracks within the hits considered for the track reconstruction. An $ambiguity\ solving$ algorithm is used in this regard where it employs a reward and a penalty until the best track candidate is determined [47]. The magnetic field generated by the Solenoid Magenet bends the charged particles traveling through the ID to discriminate the particles based on their charge. Identified tracks are then further separated into two categories, "loose"and "tight", based on $p_T > 400$ MeV , $|\eta|$ <2.5 and different number of hits recorded in the ID layers. Track reconstruction efficiency can be calculated for the reconstructed tracks as a function of p_T and $|\eta|$ which can be illustrated as follows [2].

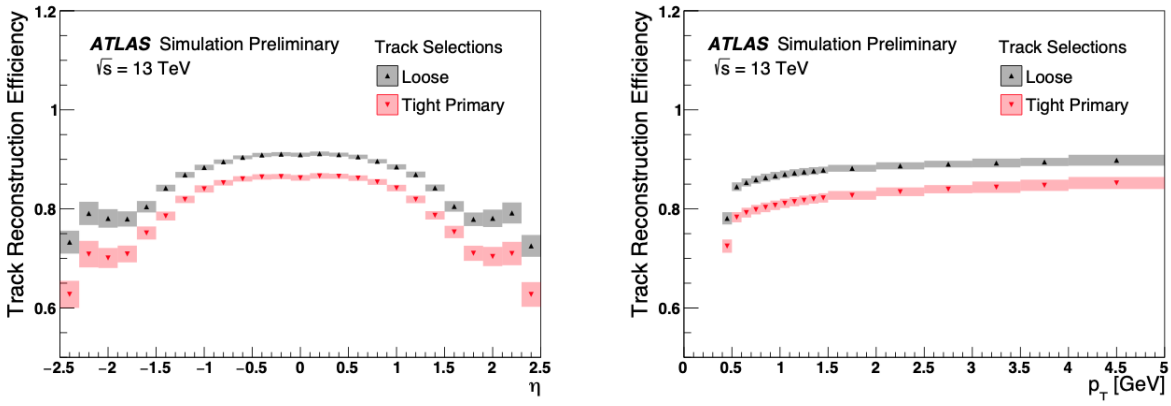


Figure 3.3: Track reconstruction efficiency as a function of η(left) p_T(right) in loose and tight track selections[2].

The primary vertices in the events arising from the pp collisions are reconstructed using a vertex finding algorithm by iterating over the tracks. A seed position for the vertex fitting is selected based on the beam spot in the transverse plane and using the selected seed position as the starting point and considering the parameters of the previously reconstructed tracks as inputs, the algorithm iterates a χ^2 minimization to optimize the position of the vertex. A weight is assigned to each input track to evaluate its compatibility with the vertex. This process is repeated until the best compatibility is derived. Rejected tracks during the iteration are collected and iterated again to find other vertices. At least two tracks with $p_T > 400$ MeV are required to be associated with the vertex in order to become a primary vertex [11, 33].

3.1.2.2 Jets

When a parton undergoes its evolution towards the hadronization stage, sprays of collimated hadrons are generated and start to interact with the detector layers. Such collimated sprays are commonly known as $jets$. As jets arise from the hadronization of different physics objects generated at the pp collisions, they carry essential information for many analyses at ATLAS and therefore require a precise reconstruction.

Jet reconstruction begins with combining the calorimetric cells which recorded the energy deposits of jets, into topologically connected clusters ($topo\ clusters$), based on their signal-to-noise ratio. Such formed clusteres then could be used to reconstruct the hadronic shower generated by a particle inside the calorimeter. The cluster formation is defined based on the cell signal significance which is defined as,

$$\varsigma_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{noise,cell}^{EM}} \tag{3.4}$$

Here, E_{cell}^{EM} is the energy measured in the cell and $\sigma_{noise,cell}^{EM}$ is the average noise in the cell and both of them are measured in the EM scale.

A growing-volume algorithm is used in forming the topo clusters and proceed towards the seed and collect steps. Cells with $\varsigma_{cell}^{EM} > 4$ are regarded as seed cells and form the $proto$ - $clusters$. If an adjacent calorimeter cell (either lying next to or at least have an overlap with the seed cell in (η, ϕ) plane) to the seed cell satisfies $\varsigma_{cell}^{EM} > 2$ (growth cells) is added to the current proto-cluster. This process continues until the the boundary requirement of $\varsigma_{cell}^{EM} > 0$ (boundary cells) it is met but doesn't exceed $\varsigma_{cell}^{EM} > 2$. Finally, if a given topo cluster exceeds a local maximum of $E_{cell}^{EM} > 500$ MeV and have two or more local maxima are spilt in all three dimensions.

The next step in the jet reconstruction is the combination of topo-clusters into jets using tha anti-k_t algorithm [35]. The anti-k_t algorithm is a sequential recombination algorithm and is infra-red and collinear safe. Distance between the particles d_{ij} and distance between the particle and the LHC beam B d_{iB} are defined as,

$$\begin{aligned} d_{ij} &= min(k_{t,i}^{2p}, k_{t,i}^{2p})\Delta_{ij}^2, \\ d_{iB} &= k_{ti}^{2p} \end{aligned} \tag{3.5}$$

Here, $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j^2)$ is the distance between the particles i and j in the (η, ϕ) plane. The transverse momentum, rapidity and the azimuthal angle of particle is given by $k_{t,i}$, η_i and ϕ_i respectively. The anti-k_t algorithm sets the parameter $p = -1$ which makes it distinct from the inclusive k_t algorithm (when $p = 1$). The combination iteration begins with comparing the distances d_{ij} and d_{iB}. If $d_{ij} < d_{iB}$, the topo clusters are combined and continue the iteration until $d_{ij} > d_{iB}$ at which time the clusters are removed and recognized as a jet. This process runs over all topo clusters until they are identified as jets. The anti-k_t can also be used to produce $truth\ jets$ and $track\ jets$ by providing truth particles and the tracks in the ID as inputs. Based on the radius of the jets, they can be seperated as small-radius (small-R) if R $\leq$ 0.6 and large-radius(large-R) if R $>$ 0.6. The analysis described in this dissertation use three such reconstrcuted jets with R=0.2, R=0.4 and R=1.0 with anti-k_t algorithm.

Once the jets are reconstructed, they need to be calibrated in order to be used in analyses. An area-based method is used in calibrating small-R jets. The method subtracts the per-event pile-up contribution in the given jet according to its area. The pile-up contribution is determine based on the median p_T density ρ of jets in the (η, ϕ) plane. For a given jet, p_T density is given by $\frac{p_T}{A}$such that area A of that jet is determined using ghost association[2]. The ρ calculations show that there is a residual p_T dependency of in-time pile-up that can be related to the number of primary vertices(N_{PV}) and out-of-

[2]particles with infinitesimal mass and energy which are clustered with the considering jet in the (η, ϕ) plane [35]

time-pie-up with μ. Distribution of median p_T for a given range of μ (mean number of pp collisions per bunch crossing) and selected values of N_{PV} are shown in fig. 3.4

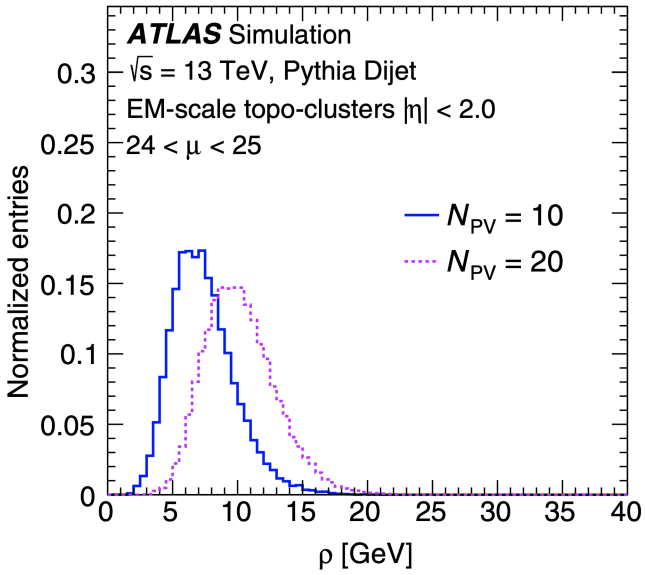


Figure 3.4: Median p_T distribution for N_{PV}values of 10 and 20 at $24 < \mu < 25$ [9]

The calibrated jets are then corrected with an absolute jet energy scale and η calibrations. The jet energy calibration is calculated by matching the reconstructed jets with the truth jets geometrically. The average energy response $\frac{E^{reco}}{E^{truth}}$ is derived as a function of η_{det} (which is pointing from the geometric center of the detector). This average energy response is then used to correct the jet energy. Fig.3.5 shows the relation of energy response to the η_{det} at different truth energy values from full ATLAS simulation.

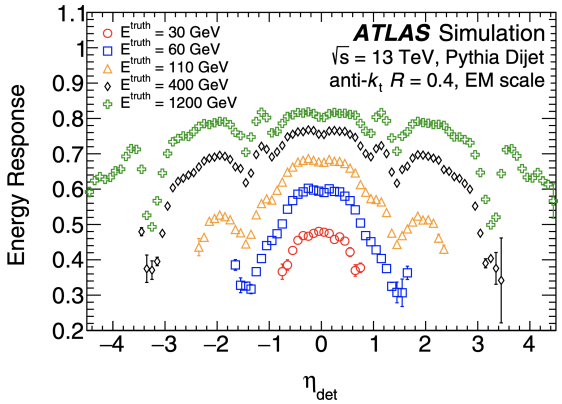


Figure 3.5: Average energy response as a function of η_{det} for selected truth energy values [9]

After the average energy response corrections, residual dependencies can be observed between jet energy and the shape of the jet. These differences are addressed via implementing a global sequential calibration(GSC). Five observables can be found which are capable of improving the jet energy scale independently for each observation, corrections are applied based on p_T^{truth} and η_{det}. Finally, an in-situ calibration applied on the jets in the data from a well-measured objects to correct the possible differences with the simulation[11].

Large-R jets can also be contaminated from pile-up, multi parton interaction etc..., thus requiring them to be calibrated as well. A trimming algorithm [58] using the ratio of the $\frac{p_T^i}{p_T^{jet}}$ (p_T^i is the constituent p_T and p_T^{jet} is the jet p_T)as the selection criterion can be used in this case where it can remove contaminats with a small proportion of hard scatter decay products and the final state radiation. Constituents in the jet use a k_t algorithm to form subjets with size R_{sub} if any selected subjet satisfies the requirement of $\frac{p_T^i}{p_T^{jet}} > f_{cut}$, it will be removed. Those who pass the requirement constitute the trimmed jet corrected from the effects of contaminents. Analysis at Ref. [58] studied the trimming for parameter values of f_{cut} = 0.01, 0.03, 0.05 and R_{sub} = 0.2 and 0.3. The large-R jets in this analysis require $|\eta|$ <2.0 and a p_T >300 GeV. Trimming parameters of R_{sub} = 0.2 and f_{cut} = 0.05 was found to be optimal.

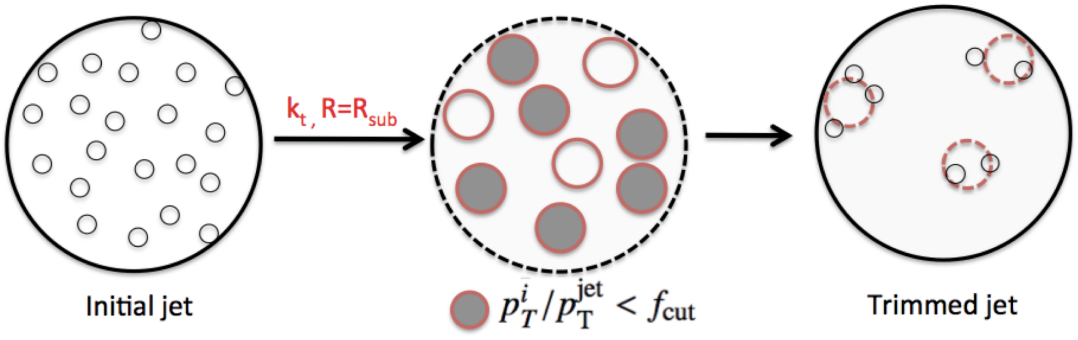


Figure 3.6: Application of trimming algorithm on the large-R jets[58]

Having the impact from soft interactions removed from the trimming procedure, large-R jets then can undergo further corrections for its energy and mass in a similar manner which small-R jets were corrected [12]. An energy response of $\frac{E_{reco}}{E_{truth}}$ and a mass response of $\frac{m_{reco}}{m_{truth}}$ are defined by geometrically matching reconstructed jets with the particle level jets. The average response for both instances are derived using a Gaussian fit applied on the response distribution. The average energy response (as a function of η_{det}) calibrates the simulation and the jet energy scale(JES) correction factor c_{JES} deerived as a function of η_{det} and the jet energy and applied on as a muliplicative scale factor on the jet four moemtum,

$$\begin{aligned} E'_{reco} &= c_{JES}\, E_{reco} \\ m'_{reco} &= c_{JES}\, m_{reco} \end{aligned} \tag{3.6}$$

Jet mass calibration follows the energy calibration and it is sensitivity to the clustering and soft contributions,as it's a key ingredient in physics analyses. An average mass reponse is evaluated in a similar way to the average energy response. The mass calibration application is applied after the jet energy scale calibrations and the correction factor for mass c_{JMS} is then derived as a function of η_{det}, E_{reco} and $log\frac{m_{reco}}{E_{reco}}$,

$$\begin{aligned} E'_{reco} &= c_{JES}\, E_{reco} \\ m'_{reco} &= c_{JES}\, c_{JMS}\, m_{reco} \end{aligned} \tag{3.7}$$

Jet vertex tagging is used to find the compatibility of a given jet with the primary vertex of the supposed interaction. This procedure helps remove any contributions from pile-up events. The Jet vertex tagger uses two inputs R_{p_T} and corrJVF which are combined in a two-dimensional likelihood and fed into a k-nearest neighbor algorithm to discriminate pile-up events from the jets in an analysis [18].

3.1.2.3 b-tagging

The long lifetime experienced by the hadrons containing b quark of $\tau \sim 1.5$ ps [42] ($c\tau \sim$450 μm) has a significance in many BSM related analyses as b quark is a favorable final state particle of them. As some weakly decaying hadrons contain b quarks (b-hadrons), jets arise from such b-hadrons (b-jets) can have secondary vertices (due to the longevity of the b quark). Proper identification of such b-jets is essential in the physics studies such as BSM, because of being a heavy particle, b quark is an ideal candidate to become a decay product of a heavy VLQ. The ATLAS uses a number of algorithms at different levels to identify such b-jets. At lower-levels, parameters such as impact parameter (IP2D, IP3D) and reconstructed secondary vertices are used for b-jet identification, which are later used as inputs for higher-level algorithms for further improvements. Taggers such as SV1 and `JetFitter` are used in the reconstruction of displaced vertices in the b-jet [4].

MV2 (multivariate) is a common high-level discriminant used for b-tagging in ATLAS analyses. MV2 is based on a Boosted Decision Tree (BDT) classifier that intakes the results from IP2D, IP3D, SV1, `JetFitter` and information of p_T and the $|\eta|$ from the jet to discriminate c-jets and $light$-jets from the b-jets. The MV2c10 variant uses simulated background sample with a c-jet background fraction of 7% and a 50%-50% mixture of Z' and $t\bar{t}$ as signal for optimizations [36].

One of the latest high-level discrimination algorithm use in ATLAS is the Deep Learning (DL1) which uses a deep neural network to discriminate c-jets and $light$-jets from the b-jets. The same low-level tagging inputs used in the MV2 classifier are used here as well. To avoid any discrimination on signal and background based on their kinematic (η, p_T) differences, the kinematic distributions per jet flavor are re-weighted to that of the b-jet distribution. These re-weighted weights are then used in the backpropagation updating process during the sample training step. The Adam optimizer used in minimizing categorical cross-entropy and the rectified linear unit(ReLu) activation function used in

the hidden layers. The output layer uses the softmax activation function. At the end of the training, outputs are combined using a log-likelihood into a single discriminant, which is allowing to use the same neural net for b and c tagging [61].

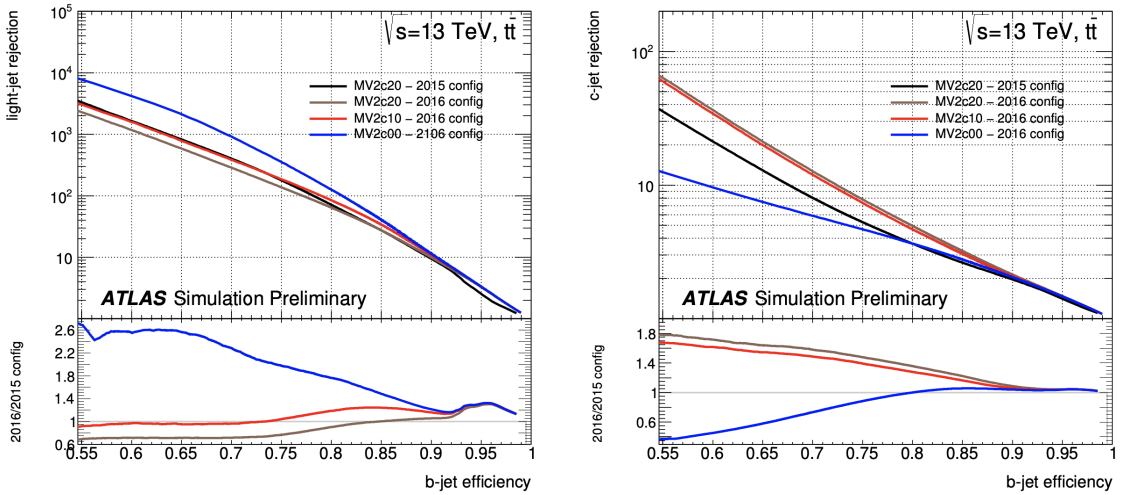


Figure 3.7: $light$-jet(left) and c-jet rejection(right) vs b-jet efficiency for MV2 variants[36, 6]

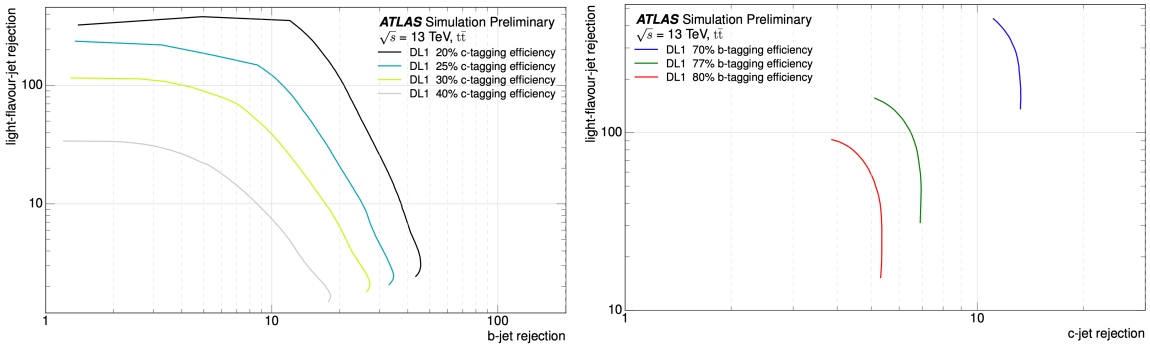


Figure 3.8: DL1 performance for c-tagging(left) and b-tagging(right) [61]

Our analysis used MV2 and DL1 outputs at ntuple levels for the b-taggin optimization studies.

3.1.2.4 Electrons

Electron reconstruction in the central ($|\eta|$ <2.47) region of the ATLAS detector is a stepwise process. A sliding window clustering algorithm with a size of 3×5 in the units of 0.025×0.025 in $\eta \times \phi$ space is defined to be compatible with the granularity of the EM calorimeter for the seed cluster reconstruction. This algorithm searches for windows with energy deposits of E_T >2.5GeV and, if a duplicate is encountered, the seed with the smallest E_T will be removed. If there is more than one track satisfying the requirements, a primary track is defined based on its distance to the seed barycenter, the number of pixels, and SCT hits. Tracks are then associated with all the seed clusters and for the combination of cluster-track pairs, the clusters are reformed as 3×7 unit size in the barrel and 5×5 unit size in the end-cap regions. Pattern recognition algorithms are used to recognizes the energy losses due to interactions of tracks with the detector material in accounting for bremsstrahlung. The track candidates are then fitted either with the pion or electron hypothesis with the ATLAS Global χ^2 track fitter. If a track can not be associate with a cluster, it is regarded as a photon and will be discarded from further analysis. The energy calibration of the clusters to the original electron energy is performed using multivariate methods [13] on simulated MC samples.

The electron identification step is based on a multivariate, log-likelihood method. Information such as the track-based information from the TRT, shape of the electromagnetic shower in the calorimeter, bremsstrahlung effects, track quality and goodness of track-to-cluster matching are brought in together to make a single likelihood value, which will be used as a discriminant in rejecting background(electrons from hadronic jets, photons) in favor of the signal efficiency. Three levels of operating points are studied as $Loose$, $Medium$ and $Tight$ in the Electron identification stage at Ref. [72]

Many ATLAS Analyses demand isolation requirements for the electrons to avoid the further contaminations from the sources such as hadrons. The calorimetric isolation energy $E_T^{cone0.2}$ sum up the transverse energies around the candidate electron cluster using

a cone with ΔR=0.2 and only the clusters with positive energies are taken into account. Then, transverse energy in a rectangular cluster of 0.125×0.175 in (η, ϕ) space centered at the electron barycentre is subtracted. A correction is then applied for possible energy leakages, pile-up, and underlying event. Track isolation $p_T^{varcone0.2}$ is taking the summation of transverse momenta of all the tracks that are reconstructed near the electron yet, are not associated with, within a cone of ΔR=min(0.2, 10GeV/E_T) around the electron track candidate.

Well known events such as $Z \to ee$ and $J/\psi \to ee$ are used to measure the efficiencies of electron reconstruction, identification, and isolation. To avoid possible biases, a $tag\,and\,probe$ method is used where one electron is required to satisfy the strong requirement allowing it to be tagged, and the other electron can be selected only from the Z or J/ψ mass resonance.

3.1.2.5 Muons

The muon reconstruction begins at the ID and MS independently to each other and the information gathered from these subdetectors are then to form muon tracks. Combining hit patterns in the individual layers to segments in the MS, tracks are formed and then fitted to make tracks that are spanning over several layers. Four algorithms are used in the combined muons reconstruction.

- Combined muons (CB): Muons track are independently reconstructed in the MS and ID. Combined tracks are formed using the hits from the MS and the ID applied on a global file. An $outside-in$ algorithm is used in the MS to begin the reconstruction process and then extrapolate towards the ID to match with the ID tracks. The fit quality can be improved by the addition or removal of MS hits.

- Segment-tagged muons (ST): When at least one local track segment in the MDT or CSC is associated with an outward extrapolated track in the ID, it is regarded as a muon. This is useful for identifying muons with low p_T which are capable of

only passing through one layer in the MS or are traveling through reduced accepting regions in the MS.

- Calorimeter-tagged muons (CT): The MS has a lower sensitivity in the regions of the detector that falls into $|\eta|$ <0.1. A track in the ID can be treated as a muon if the calorimeter measurements are compatible with signs of a minimum-ionizing particle.

- Extrapolated muons (ME): The ID has no coverage in the very forward regions at 2.5< $|\eta|$ <2.7. Thus, in muon reconstruction, the candidate is required to traverse at least two layers(three layers in the forwards regions) in the MS. These tracks in the MS, therefore, need to be compatible with the interaction point.

Several factors such as the compatibility of ID and MS measurements, quality in the track fit, and global χ^2 fit are taking into account when a candidate is identified as a muon. Four muon identification selections are described at Ref. [17] and a brief description of the is given bellow.

- Medium: This is the default selection for ATLAS and we have adopted this requirement into our analysis. Only the ME and CB are used in this selection. The CB is required to have at least three hits in the MDT layer and the ME requires at least 3 hits in the MDT/CSC layers.

- Loose: All muon types are allowed and are designed to deliver good quality muons while maximizing the reconstruction efficiency. These are specifically designed for Higgs boson candidate reconstruction in the four-lepton final state [16].

- Tight: These are designed to improve the purity of the muons and only the CB muons with at least two hits in the MS and satisfying the medium selection criteria are considered. Two hits in at least two layers of the MS are required.

- high-p_T: This selection is to maximize the momentum resolution for tracks with p_T > 100GeV and optimized for high-mass Z' and W' searches [14, 15].

The reconstruction efficiency, identification, and isolation requirements for muons are performed similar to the electrons as described in section 3.1.2.4 by referring to well-studied resonances such as $Z \to \mu\mu$ and $J/\psi \to \mu\mu$ events.

3.1.2.6 Missing transverse momentum

Momenta of all physical objects generated in a pp collisions must satisfy the conservation of momentum on the plane transverse to the beam axis. There can be several instances where momentum conservation could break, first from the particles such as neutrinos(which is neutral and weakly interacting) generated from the collisions can escape from the detector with no detection made and secondly if the physics objects are mismodeled. Proper understanding of the missing transverse momentum also helpful in the search for new particles such as "Least Supersymmetric Particle (LSP)"which is regarded as a candidate in the dark matter related studies [60]. The absence in energy is therefore known as the "missing momentum(energy)".

The missing transverse momentum (also referred to as E_T^{miss}) can be evaluated form the negative vector summation of the transverse momenta of all the particles detected in a given event. A soft term (ID track-based or calorimeter-based) is added in E_T^{miss} reconstruction to include any activity that is unassociated from the calibrated objects in the calculation. The track-based soft term reduces the pile-up dependency to the E_T^{miss}. the Calorimeter-based soft term takes all topo clusters that are not associated with jets in the E_T^{miss} calculations.

$$-E^{miss}_{T(x,y)} = E^{miss,e}_{T(x,y)} + E^{miss,\gamma}_{T(x,y)} + E^{miss,\tau}_{T(x,y)} + E^{miss,jets}_{T(x,y)} + E^{miss,\mu}_{T(x,y)} + E^{miss,soft}_{T(x,y)}, \tag{3.8}$$

$$E^{miss}_{T} = \sqrt{(E^{miss}_{T(x)})^2 + (E^{miss}_{T(y)})^2} \tag{3.9}$$

$$\phi^{miss} = \frac{E^{miss}_{T(y)}}{E^{miss}_{T(x)}}. \tag{3.10}$$

where E^{miss}_{T} is the magnitude of missing transverse momentum and ϕ^{miss} is the azimuthal angle [5, 3].

CHAPTER IV

VLQ T→Ht Analysis

4.1 Introduction

This chapter describes an analysis searching for single production of VLQ T quarks decaying into a SM top quark and a Higgs boson in an all-hadronic final state. The daughters(SM top and the Higgs boson) of the decaying T will be "boosted"(have high transverse momentum as they are sharing the energy of the decaying T) making them collimated towards the direction of the parent particle. As discussed in section 3.1.2.2, such a collimated spray of particles can be regarded as a jet. Thus, we require two large-R jets(leading and sub-leading) with p_T > 350 GeV. Furthermore, the leading large-R jet must have p_T > 500 GeV for better trigger performance. The main background expected in this analysis is pair-produced SM top and the multijet events from SM particles such as $gluons$ and/or u, d, s, c and b quarks.

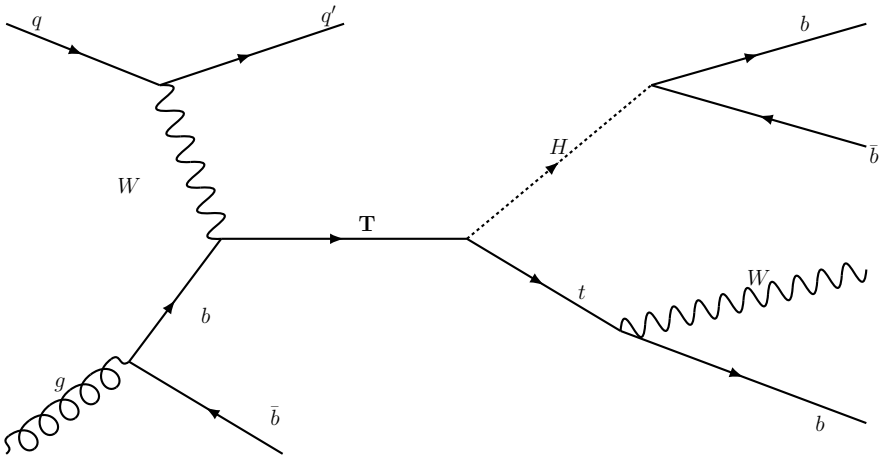


Figure 4.1: A Feynman diagram for the production of VLQ Top and its decay considered for the analysis described.

4.1.1 Data and MC simulated samples

This analysis uses data collected by the ATLAS detector from 2015 through 2018 (Run 2 data). The data set consists of 139 fb^{-1} of integrated luminosity from pp collisions with a center of mass energy of $\sqrt{s}$ = 13 TeV. Only the data collected during the periods when all the subdetectors were operational are considered in this analysis.

4.1.2 Signal production

Simulated samples of the signal process, $pp \rightarrow T \rightarrow Ht$, were generated with the MadGraph5 event generator [25] at leading order in α_s and with the NNPDF23 PDF set. They were then passed to Pythia8[pythia ref] for the parton showering and hadronization. Nominal signal samples for the physics process of $Wb \rightarrow T \rightarrow Ht$ were produced with the VLQ mass values of 1.1 Tev, 1.7 TeV and 2.1 TeV with a coupling (coupling value between the VLQ T to SM particles) value of κ= 1.0. Events in each sample are carrying generator-level weights allowing the samples to re-weight into a different κ with a mass 100 GeV less or to a different coupling value with the same mass. The coupling value κ is varying in ranges defined as, 0.1$\leq$ κ $\leq$0.5 in 0.05 steps and 0.5$\leq$ κ $\leq$1.6 in 0.1 steps. Such variation in VLQ mass and coupling is generating a two-dimensional grid and corresponding leading order cross-sections are tabulated in the table C.4 in Appendix C. A miscalculation in the Lagrangian used in the MadGraph for generating samples for the $T \rightarrow Ht$ was discovered in 2019 (as described in Appendix B) and we have used the bugged samples to test mass and coupling re-weightings and for optimization studies. In collaborating with the effort on regenerating samples with the corrected Lagrangian, we performed a validation study which is described in Appendix B, and the bug-fixed samples are now been used in further optimization studies.

4.1.3 Background production

Two dominant SM backgrounds boosted $t\bar{t}$ and the multijet background are considered in this analysis. The multijet background events are estimated using a data-driven method described in section 4.1.6.3.

The POWHEG-BOX V2 MC generator was used to model the $t\bar{t}$ process using an NLO-calculation. MADGRAPH5_aMC@NLO and SHERPA MC generators were used in calculating alternative matrix-elemet calculations.

The production of $t\bar{t}$ pairs associated with W, Z and Higgs bosons was modeled with the MG5_aMC@NLO event generator along with PYTHIA8 parton shower and hadronization model.

4.1.4 Event selection

An inclusive set of anti-k_t jet triggers with a radius parameter of R =1.0 were used in this analysis for event selections. A trigger efficiency study for 2018 data taking period was carried out in this analysis realized that un-preselected efficiencies were 100% efficient when a requirement of an offline section of at least one large-R jets with p_T >500 GeV and $|\eta|$ <2.0 was imposed. The p_T thresholds for the un-preselected triggers for the rest of the data collection periods from 2015, were varying between 360 GeV and 480 GeV, giving a higher efficiencies for an offline selection during the period.

The event selection requires dijet events where one jet is recognized as a top-quark candidate and the other as a Higgs boson candidate. A primary vertex (which has the largest p_T^2 and the sum is for all tracks with p_T >0.4 GeV)with at least five associated tracks are required for the events.

The $t\bar{t}$ background, requires several steps to reduce its contribution. The $t\bar{t}$ events that decay semi-leptonically are rejected by imposing a requirement of having no well-identified, isolated reconstructed electron or muon candidate. Then the events need to pass through the following requirements to be identified as a candidate for a fully hadronic

decay. There must be at least two large-R jets with p_T > 350 GeV and $|\eta|$ < 2.0 and the leading large-R jet must have p_T > 500 GeV. The candidates then require to have at least three variable radius track jets (VRTrack jets) with p_T > 25 GeV and $|\eta|$ < 2.5. These pre-selection conditions allow to have an event sample of 90 million events towards the analysis.

A multi-step algorithmic approach is then employed on the leading and second-leading large-R jets to classify and sort the events into a 9×9 matrix of regions that will be used throughout this analysis.

- A DNN top-tagging algorithm is used to determine if the jet is tagged as a top-jet candidate
  - Optimized at 80% working point
- A Higgs boson tagger algorithm is used to determine if the jet is tagged as a Higgs boson candidate
  - Optimized at 70% working point
- Number of VRTrackjets jets in a given large-R jet that are also b-tagged and are geometrically matched to the large-R jet(taken into three categories, 0 b-tags, 1 b-tags and ≥ 2 b-tags)[1].
  - MV2c10 b-tagging algorithm optimized at 70% working point

[1]Please refer to Appendix C for taging studies

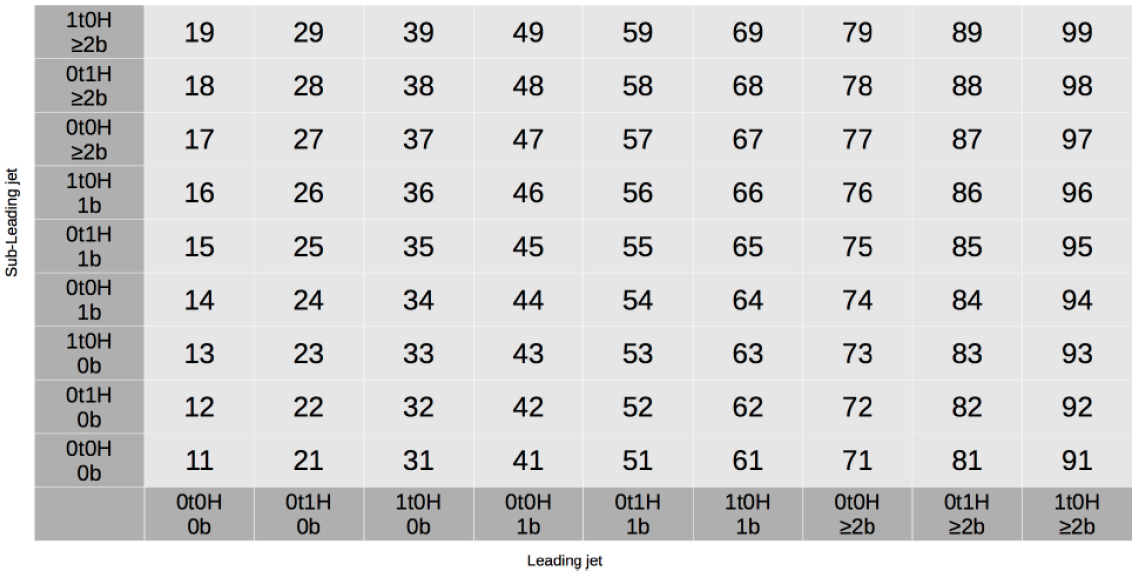


Figure 4.2: Pre-selected events sorted into 9×9 regions based on b, top and Higgs jet tagging. Regions are exlusive and a given event in a region is unique.

As depicted in fig 4.2, the 9×9 regions matrix contains all the events that passed all the aforementioned requirements and any given region is unique. Columns in the matrix define the tagging status of the leading jet and the rows define that of the second leading jet. For instance, region 53 has a total number of events with their leading jet has 1 Higgs-tag and a 1 b-tag and the second-leading jet has only 1 top tag[2].

Three main regions are then defined to recognize the key analysis features such as signal regions, validation regions and the $t\bar{t}$ normalization regions. They can be expressed as follows,

- Regions 68.86,89,98 as signal regions
- Regions 38,49,83,94 as validation regions
- Regions 66,69,96,99 as $t\bar{t}$ normalization regions

[2]Please refer to Appendix C for tagging studies for this analysis

The Signal regions: Events in the signal regions have one top-tagged large-R jet with at least 1 b-tagged subjet and the other is 1 Higgs-tagged with two or more b-tagged jets. As the analysis is focusing on a tH final state, order in which either large-R jets tagged as Higgs or top is irrelevant. Either the leading or second-leading jet will have on top-tag when the events are selected for the signal regions. The sum of four-momenta of these two large-R jets are thus defined as the tH final state candidate.

The Validation regions: These regions closely resemble the signal regions however, they are dominated by background events and therefore are important in carrying studies especially in Higgs tagging. For example, region 38 has its second-leading jet tagged with 1 Higgs and at least two b-tags enabling it to be used in studies such as large-R jet mass distribution and subjet kinematic observables in related to the Higgs.

The $t\bar{t}$ normalization regions: These regions are dominated by $t\bar{t}$ events and are therefore helpful in top-tagging studies in the analysis. The large-R jets with 1 top tag and more than 1 b-tag are regarded as a "real top "candidate because of the misidentification of charm quark (coming from the hadronic decay of W boson) as a top during the b-tagging stage. These regions are similar to the signal regions discussed in the analysis given at Ref [8].

Signal regions: 68, 86, 89, 98
Normalization regions for ttbar: 66, 99, 69, 96
Validation regions: 38, 83, 49, 94

Second Leading jet \ Leading jet	0t0H 0b	0t1H 0b	1t0H 0b	0t0H 1b	0t1H 1b	1t0H 1b	0t0H ≥2b	0t1H ≥2b	1t0H ≥2b
1t0H ≥2b	19	29	39	49	59	69	79	89	99
0t1H ≥2b	18	28	38	48	58	68	78	88	98
0t0H ≥2b	17	27	37	47	57	67	77	87	97
1t0H 1b	16	26	36	46	56	66	76	86	96
0t1H 1b	15	25	35	45	55	65	75	85	95
0t0H 1b	14	24	34	44	54	64	74	84	94
1t0H 0b	13	23	33	43	53	63	73	83	93
0t1H 0b	12	22	32	42	52	62	72	82	92
0t0H 0b	11	21	31	41	51	61	71	81	91

Figure 4.3: Pre-selected events sorted into 9×9 regions based on b, top and Higgs jet tagging. Regions in green are $t\bar{t}$ normalization, pruple are validation regions and in red are signal regions.

4.1.5 Event yields

Once the taggers top, Higgs, and the b were optimized, the event yields for the Run 2 data in each analysis region is illustrated as in fig4.4. The signal regions are left blank as our analysis is still in the "blind "status.

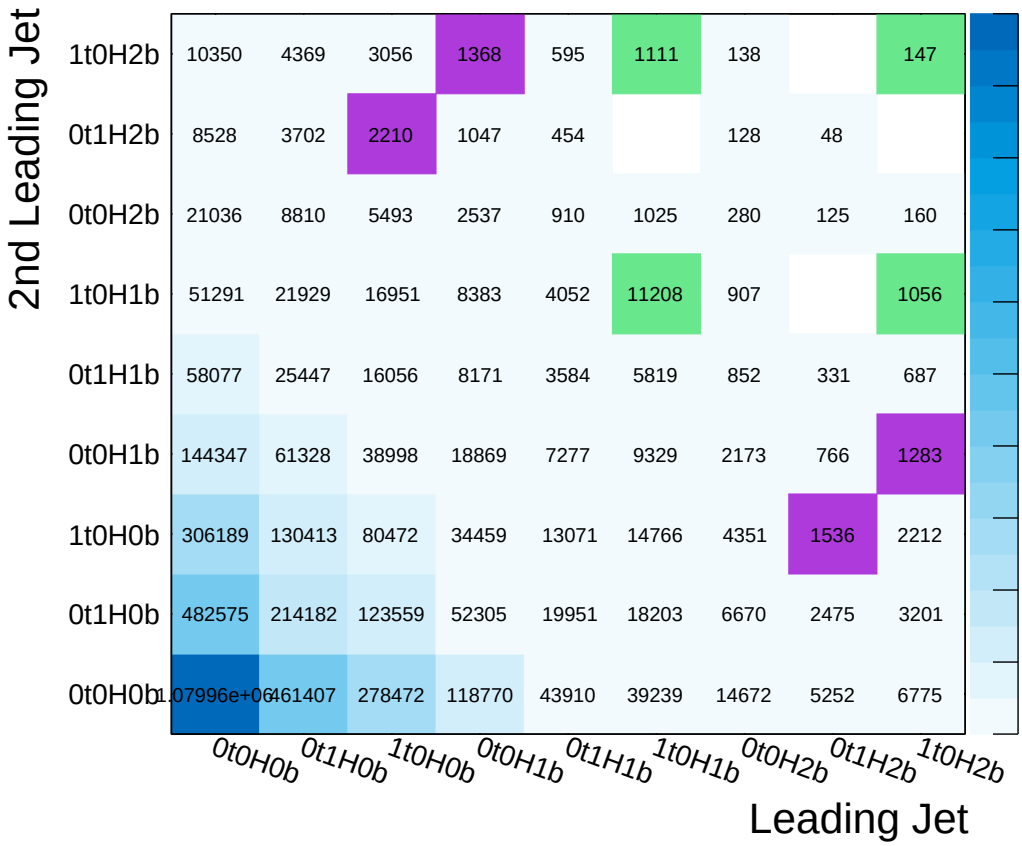


Figure 4.4: Run 2 event yields except for the VLQ signal regions which are blinded. The yields are derived using the current tagging selctions, DNN top tagger at 50% W.P., Higgs tagger with p_T dependent τ_{21} cut and b-tagging of variable radius small-R jets using MV2c10 algorithm at 70% W.P. Regions in green: $t\bar{t}$ control regions, in purple: Validation regions

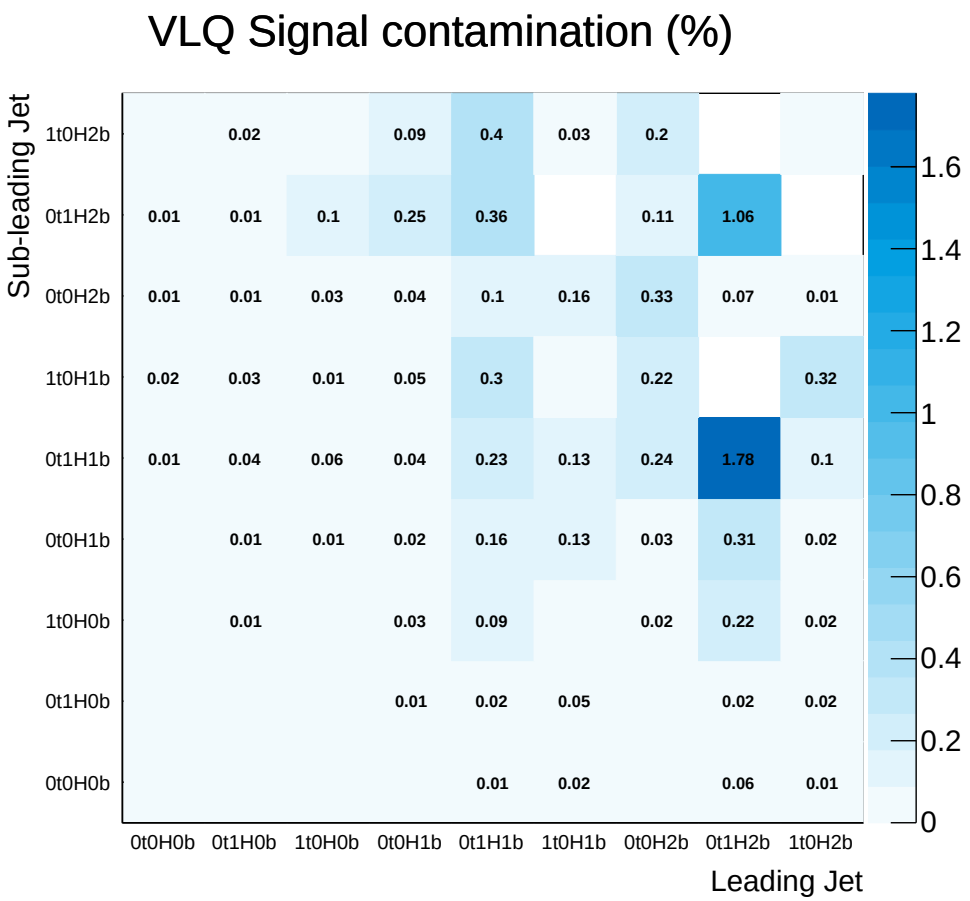


Figure 4.5: The VLQ signal to data in the 9×9 region matrix as a percentage. The signal regions left in white color as the analysis in the blinded stage.

4.1.6 Background estimation

The backgrounds arising from the multijet backgrounds and the $t\bar{t}$ final states with both the top quarks decaying fully hadronic are the two major contributors for background in this analysis. Following them is the background from $t\bar{t}$ final states with one top quark decays semi-leptonically and the daughter lepton being unidentified. MC simulations are used to estimate the background coming from the SM processes that have at least one top quark jet.

4.1.6.1 Background from other SM processes

A small contribution is expected from single-top-quark produced from the Wt-channel into the signal sample. The t-channel process is accounted for in for the multijet background estimation.

4.1.6.2 $t\bar{t}$ background estimation

The $t\bar{t}$ events that arise from the sem-leptonically decaying top quarks are estimated using POWHEG+PHYTHIA8 generators. This analysis does not identify and reject the τ lepton candidates thus, their contributions are added with semi-leptonic decay contributions. The $t\bar{t}$ normalization and the validation regions have contributions from semi-leptonic $t\bar{t}$ events. The other main source for the $t\bar{t}$ background is the $t\bar{t}$ final states that are misreconstructed as the tH final states. These $t\bar{t}$ events are contributing towards the $t\bar{t}$ normalization regions as well as the other validation regions where there are loosen top or b-tagging requirements imposed.

4.1.6.3 Multijet Background estimation

A data-driven approach (ABCD method) was adopted in estimating the multijet background in the tH final states. This approach was inspired by the analysis at Ref [30]. This estimation is based on the analysis regions illustrated as in fig.4.2.

The ABCD method is a widely used likelihood method in physics analyses for estimating multijet background in analysis regions. Adopting the ABCD method into our analysis, we begin by assuming that the ratio among the tagged to untagged events in a tagging state of the leading large-R jet is independent of that of the second-leading jet. For example, Referencing the 9×9 region matrix in fig 4.2, regions 96 and 91 have the same taggings for the leading large-R jet while the second leading jet has one top and one b-tag in region 96 but not in region 91. This is the same for regions 16 and 11. Thus, the ratio between regions 96 to 91 can be compared with the ratios of 16 to 11 and their

ratios can be used to estimate the number of events in the region 96 as follows,

Let A=96, B=91, C=16 and D=11 then,

$$\begin{aligned} \frac{A}{B} = \frac{C}{D} &\Rightarrow \frac{N_{96}}{N_{91}} = \frac{N_{16}}{N_{11}} \\ A = B \cdot \frac{C}{D} &\Rightarrow N_{96} = N_{91} \cdot \frac{N_{16}}{N_{11}} \end{aligned} \tag{4.1}$$

Likewise, this method can be adopted in estimating the number of events in a particular region of any given analysis. The multijet background for the VLQ signal (68, 86, 89, 98) and the validation (66, 69, 96, 99) regions in this analysis were estimated using the ABCD method.

Before the application of the ABCD method into our analysis regions, all the simulated MC backgrounds such as all hadronic, non all hadronic, and single top distributions were subtracted from data at bin by bin level. Regional estimation for the analysis at an elementary level was performed based on equation 4.1. While the event estimation in a selected region based on equation 4.1 serves as an initiation, the assumption of independence between the leading and the second-leading jets in terms of tagging status is not held in a real-life scenario. Thus, a correction factor was introduced to study the regional correlations and make corrections for the multijet background estimation as follows,

$$K_{\Delta A}^{\Delta B} = \frac{G(A_f|B_f) \cdot G(A_i|B_i)}{G(A_i|B_f) \cdot G_(A_f|B_i)} \tag{4.2}$$

Here, $K_{\Delta A}^{\Delta B}$ is the correlation correction factor and ΔA is the change in tagging state in the leading large-R jet and ΔB is that of the second-leading large-R jet. A_f and B_f are standing for the final tagging status of the leading ($lj1$) and second-leading ($lj2$) large-R jets and A_i and B_i are for the initial tagging status of them. To access the number of events in the region of interest, function $G(lj1|lj2)$ is used. For example, G($0b0t0H|2b0t0H$) $= N_{17}$ and G($1b1t0H|2b0t1H$) $= N_{68}$.

The Δ defined in the correlation correction factor $K_{\Delta A}^{\Delta B}$ is referring to the change in the tagging state of the correlation of interest where ΔA is the difference between A_i

and A_f while ΔB is the difference between B_i and B_f. For example, if we are interest to evaluate the correlation correction factor for $\frac{N_{32}}{N_{31}} = \frac{N_{12}}{N_{11}}$ it can be expressed as,

$$K_{1t}^{1h} = \frac{G(1t|1h) \cdot G(0t|0h)}{G(0t|1h) \cdot G(1t|0h)} = \frac{N_{32} \cdot N_{11}}{N_{12} \cdot N_{31}} \tag{4.3}$$

The evaluation of correlation correction factor accross the 9×9 analysis regions provides a symmetry of correlations. For example, if calculation given in eq 4.3 performed for $K_{\Delta A}^{\Delta B}$, by inverting the digits of all the regions in the equation can gives the correlation correction factor $K_{\Delta B}^{\Delta A}$ as,

$$K_{\Delta B}^{\Delta A} = \frac{N_{23} \cdot N_{11}}{N_{21} \cdot N_{13}} = K_{1h}^{1t} \tag{4.4}$$

In the process of deriving correlation correction factors, two methods were tested,

- Calculate factor for all possible variations and average
- Calculate factor for selected variations, take the average and then summed up for certain selections

Correlation correction factors derived by averaging for all possible variations produced nonsensical values as they were affected by the subtraction of non all hadronic MC contributions from data. Thus, the analysis regions that passes through a MC contribution cut ($100 \cdot \frac{N_{MC}}{N_{Data}} < W.P.$) were used with the selected averaging method for four working points(W.P.) 7.5%, 10%, 12.5% and 15% (these working points correspond to the amount of $t\bar{t}$ contamination). Compared to the other W.P., the 7.5% method had lower uncertainties in most bins (except for some bins with higher uncertainties while the other W.P. had higher uncertainties in all the bins). The 7.5% method was further refined by removing certain variations which caused the higher uncertainties.

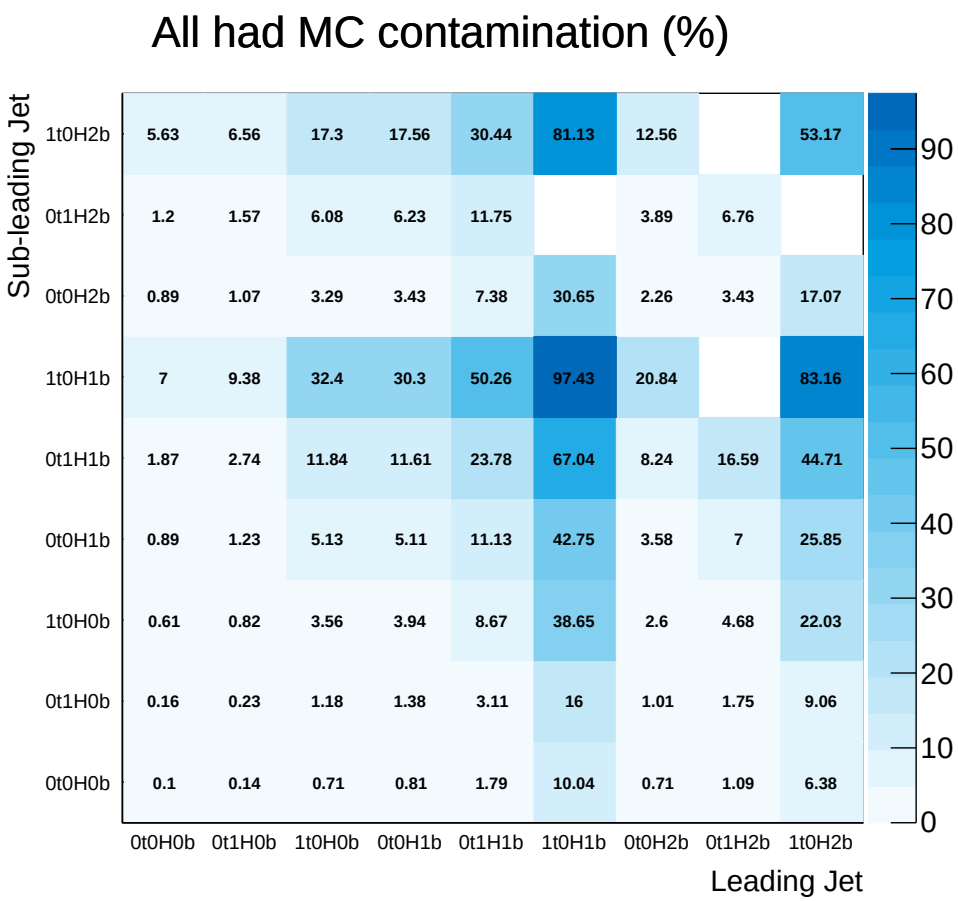


Figure 4.6: Contamination of $t\bar{t}$ all-hadronic Monte Carlo into each region in the 9×9 matrix as a percentage. The signal regions left in white color as the analysis in the blinded stage.

Both the aforementioned methods assumed any correlation correction factors corresponding to $1b$ variations are equivalent to that of $2b$ variations because the $2b$ correction factors calculated from the generalized formula resulted in extremely high uncertainties. This was addressed by summing up certain variations together before the application of a formula similar to that of the generalized formulae used in earlier instances and is known as the $2b$ summed method.

The formulae and their values for correlations that do not require to used $2b$ summed method are given in table 4.1

Correlation Factor	Formula	Value	Error Percentage (%)
K_{1t}^{1b}	$\frac{N_{34}\cdot N_{11}}{N_{14}\cdot N_{31}}$	1.012 ± 0.0059	0.58
K_{1b}^{1t}	$\frac{N_{43}\cdot N_{11}}{N_{13}\cdot N_{41}}$	1.019 ± 0.0064	0.63
K_{1t}^{1t}	$\frac{N_{33}\cdot N_{11}}{N_{13}\cdot N_{31}}$	1.018 ± 0.0051	0.50
K_{1b}^{1b}	$\frac{N_{44}\cdot N_{11}}{N_{14}\cdot N_{41}}$	1.097 ± 0.0079	0.72
K_{1t}^{1h}	$\frac{N_{32}\cdot N_{11}}{N_{12}\cdot N_{31}}$	0.9725 ± 0.0035	0.36
K_{1h}^{1t}	$\frac{N_{23}\cdot N_{11}}{N_{13}\cdot N_{21}}$	0.9859 ± 0.0036	0.37
K_{1b}^{1h}	$\frac{N_{42}\cdot N_{11}}{N_{12}\cdot N_{41}}$	0.9865 ± 0.0045	0.46
K_{1h}^{1b}	$\frac{N_{24}\cdot N_{11}}{N_{14}\cdot N_{21}}$	0.9825 ± 0.0041	0.0042

Table 4.1: All the correlation factors that do not require the $2b$ summed method used on various regions.

The formulae and their values for correlations that used $2b$ summed method are given in table 4.2.

Correlation Factor	Summed Region A	Summed Region B	Summed Region C	Summed Region D	Value	Error Percentage (%)
K_{2b}^{2b}	$77, 78, 87, 88$	$11, 12, 21, 22$	$17, 18, 27, 28$	$71, 72, 81, 82$	1.034 ± 0.0373	3.61
K_{1h}^{2b}	$27, 28, 29, 87, 88$	$11, 12, 13, 41, 42$	$17, 18, 19, 47, 48$	$21, 22, 23, 81, 82$	0.9907 ± 0.0077	0.78
K_{2b}^{1h}	$72, 82, 92, 78, 88$	$11, 21, 31, 14, 24$	$71, 81, 91, 74, 84$	$12, 22, 32, 18, 28$	1.005 ± 0.0093	0.93
K_{1b}^{2b}	$47, 48$	$11, 12$	$17, 18$	$41, 42$	1.049 ± 0.0160	1.53
K_{2b}^{1b}	$74, 84$	$11, 21$	$71, 81$	$14, 24$	1.061 ± 0.0181	1.71
K_{1t}^{2b}	$37, 38$	$11, 12$	$31, 32$	$17, 18$	0.9782 ± 0.0122	1.25
K_{2b}^{1t}	$73, 83$	$11, 21$	$13, 23$	$71, 81$	0.9997 ± 0.0148	1.25

Table 4.2: All the correlation factors that require the $2b$ summed method used on various regions. The computation of $\frac{A \cdot B}{C \cdot D}$ derive the correlation correction factor where A,B,C,D are the sum of the regions decribed in the table.

At the end of the correlation factor studies, the $2b$ summed method was adopted because of its capability in deriving values with lower uncertainties, and this method is, therefore, deriving a second-order estimate for the multijet background.

After the multijet background estimation derived using the correlation correction factor K, another scale factor α was introduced. The scale factor α is used for the $t\bar{t}$ all-hadronic and semi-leptonic backgrounds where a data-driven method is used in deriving its values.

$$\alpha = \frac{Data - Multijet - Unscaled\,MC}{Scaled\,MC} \tag{4.5}$$

The $Scaled\,MC$ term in equation 4.5 represents the summation of all-hadronic and semi-leptonic samples as they have the same production cross-section and $Unscaled\,MC$ is for the other MC processes such as $t\bar{t}$ + $W/Z/H$ and single-top. Summation of all the inclusive yields from four $t\bar{t}$ normalization regions are used in deriving α. The scale factor then scales the MC distributions before they are subtracted from data iteratively where the studies showed that the scale factor value, its uncertainty, and the final multijet background estimation were converging with four iterations. The value was found to be $\alpha = 0.793 \pm 0.011$ when evaluating with employing DNN top tagger and track jets for b-tagging. The Scale factor α was 0.78 ± 0.06 when the TRexFitter was used to fit for the $t\bar{t}$ normalization region which gives a good agreement on the iterative method.

4.1.6.4 Background summary

The table 4.3 is summerizes the observed event yields and the predicted background events for tH signal regions and the $t\bar{t}$ normalization regions.

	$t\bar{t}$ Normalization Regions			tH Signal Regions		
$t\bar{t}$ (all-hadronic)	10 764	±	140	368	±	12
$t\bar{t}$ (non-all-hadronic)	541	±	11	60	±	3
Single top-quark	142	±	8	12	±	2
$t\bar{t}$ +$W/Z/H$	151	±	2	13	±	0.4
Multijet events	2734	±	32	714	±	11
Prediction	14 332	±	144	1166	±	17
Data (139 fb^{-1})	14 332	±	120			

Table 4.3: Predicted or measured numbers of background events in the tH signal regions and $t\bar{t}$ normalization regions

4.1.7 Systematic uncertainties

The sources of systematic uncertainties affecting this analysis are discussed in this section. The main uncertainties considered are detector-related uncertainties, luminosity uncertainty, and modeling uncertainties on background processes.

4.1.7.1 Luminosity uncertainties

The uncertainty for the 2018 data alone is 2.0%, and the uncertainty for the combined Run-2 dataset (2015-2018) is 1.7%. [3] All the backgrounds estimations derived from simulated samples were using this uncertainty and was normalized to the measured integrated luminosity.

4.1.7.2 Detector-related uncertainties

The detector-related uncertainties are included in the estimations of the simulated background and signal samples and they are associated with a plethora of physics objects such as large-R jets, b-tagging, pileup, muons, E_T^{miss} as described in table 4.4. A nuisance parameter was introduced for each uncertainty to evaluate the effects of the detector uncertainties on the measured signal strength using a Profile-likelihood fit. The nominal histograms in the simulated samples contain two variants "Up"and "Down" for each nuisance parameter, which corresponds to a 1σ uncertainty range and they were used in as templates during the fitting stage. The vertical morphing method was used to interpolate between the templates and the nominal histogram for the introduced nuisance parameter.

The systematics were propagated to the $t\bar{t}$ all-hadronic MC samples as they are contributing towards the object tagging and the large-R jet kinematic reconstruction. We used the latest version (by the time of the compilation of this dissertation) AnalysisTop-21.2.98 to produce all the associated samples with the systematic uncertainties described here.

The large-R jet uncertainties are estimated for a variaty of variables such as p_T , τ_{21}^{WTA} (used for Higgs tagging studies), τ_{32}^{WTA}, C_2, D_2, Q_w, $\sqrt{d_{12}}$, $\sqrt{d_{23}}$ (used for top tagging studies) by including a wide selection of systematics at the tree level in sample ntuple production [4]. We have derived uncertainties associated with τ_{21}^{WTA} as the Higgs

[3] This paragraph is adapted from the ATLAS publications committee: https://twiki.cern.ch/twiki/bin/view/Atlas/LuminosityForPhysics

[4] WTA: Winner Takes All

tagging algorithm used in our analysis requires a varying τ_{21} cut to evade the b jet dependancy. A detailed description on the associated uncertainty derivation for τ_{21}^{WTA} is given in Appendix D.

Systematic uncertainties coming from b-tagging scale factors extrapolation for high p_T and extrapolation from charm tagging are included in the b-tagging uncertainties. Additionally, scale factor measurements from *in situ* measurements for large-R top quarks will be included as well.

We use the DNN boosted top-tagger for top-tagging studies as described in appendix C and Ref [10], and its performance associated systematic uncertainties are provided in AnalysisTop 21.2.99. For the time being, systematic uncertainties are only available for contained 80% option for the DNN top tagger and they are still in an early stage of development.

The systematics associated with leptons such as p_T scale, p_T resolution E_T^{miss}, sagitta for muons and energy, and resolution for electrons/photons are included in this analysis. They are mainly used in the studies related to identification, trigger, and kinematics. Systematics for vertex identification and track-based variable are arie due to pileup fake rates and the event weights for hard scattering efficiencies. This analysis includes uncertainties for an overall pileup weight and the multivariate Jet-Vertex-Tagger (JVT). All the detector-related uncertainties considered in this analysis are summarized in table 4.4.

The Higgs tagger in the analysis uses a τ_{21} dependent cut at p$_T$ bin level for optimization purposes and we derived the associated τ_{21} uncertainties using the "RTrack"methodology and they were included in the ntuples produced from AnalaysisTop 21.2.123 version. A detailed description of the τ_{21} uncertainty studies are given in Appendix D.

Source	Description	Plotting Name
b-tagging	Scale extrapolation to high pT Scale extrapolation from charm	weight_ trackjet_ bTagSF_ MV2c10_ 70_ extrapolation weight_ trackjet_ bTagSF_ MV2c10_ 70_ extrapolation_ from_ charm
Pileup	Weight from pileup Weight from jet-vertex-tagger	weight_ pileup weight_ jvt
Muons	p_T Scale p_T resolution MS p_T resolution ID Sagitta MUON_ SAGITTA_ RHO	MUON_ SCALE MUON_ MS MUON_ ID MUON_ SAGITTA_ RES-BIAS
Large-R Jets	Mass Calibration (97 NPs)	Various
Electrons	Energy scale Energy resolution	EG_ SCALE_ ALL EG_ SCALE_ AF2 EG_ RESOLUTION
Missing E_T	Missing E_TScale	MET_ SoftTrk_ Scale
Integrated luminosity	Standard uncertainty of 1.7%	
MC mod-elling	QCD Radiation	weight_ Var3c

Table 4.4: All the detector-related uncertainties considered in the analysis.

4.1.7.3 Modeling uncertainty for $t\bar{t}$ background

Two significant sources can be recognized for modeling uncertainties. First is from the correlation between the signal and $t\bar{t}$ normalization regions. An uncertainty can arise if the relative change in $t\bar{t}$ events in each signal and normalization regions are not proportional as expected. The other uncertainty source is the $t\bar{t} + V$ processes rising from smaller MC derived backgrounds.

4.1.8 Satistical analysis

Particle physics analyses use hypothesis testing at their statistical analysis stage. The null hypothesis H_0 usually takes the background only versus the alternative hypothesis H_1 which includes backgrounds and the signal desired. A frequentist significance test approach with a profile likelihood ratio is commonly practiced. Parameters such as the cross-section of the signal process, signal mass (in our analysis $m(tH)$) are mainly regarded as the parameters of interest (POIs) and the other parameters such as systematic uncertainties are regarded as nuisance parameters (NPs).

In an experiment that measures a certain kinematic variable x, the data that results can be represented as a distribution given by a histogram. Assuming such a constructed histogram has N number of bins, an expectation value for the number of entries(n) in a given bin is,

$$E[n_i] = \mu s_i + b_i \tag{4.6}$$

Here, n_i is the number of entries in the i^{th} bin. The s_i, b_i are the nominal contributions from signal and background to the particular bin. The μ is defining the strength of the signal process and therefore, $\mu = 0$ stands for nominal background-only hypothesis, and $\mu = 1$ stands for the nominal signal hypothesis.

Including the shape and nuisance parameters, the profile likelihood (PL) function can be given as,

$$L(\mu, \theta) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \cdot \prod_{k=1}^{m} \mathcal{G}(\theta_k), \tag{4.7}$$

Here, $\theta = (\theta_s, \theta_b, b_{tot})$ stands for all the nuisance parameters and the $\mathcal{G}(\theta_k)$ is the constriants term for them. The constraint term can take various functional forms depending on the nuisance prameters considered and as we are taking systematic uncertainties for the nuisance parameters, the constraint term becomes a Gaussian constraint with a

probability density function of,

$$\mathcal{G}(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(\theta-\theta_0)^2}{2\sigma^2}\right) \tag{4.8}$$

with a θ_0 of nominal value typically 0 and a standard deviation of σ typically 1. A profile likelihood ratio,

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}, \tag{4.9}$$

can be considered to test a hypothesized value for μ which is ranging between $0 < \lambda < 1$ where larger values of λ suggest a stronger agreement between the data and $\hat{\mu}$. The $\hat{\mu}$ and $\hat{\theta}$ are the estimators that maximize the likelihood function which is known as the unconditional likelihood function. The $\hat{\hat{\theta}}$ is the value for θ (values for nuisance parameters) that maximize the L for the testing μ, which is also known as the conditional maximum-likelihood. The estimator $\hat{\mu}$ is constrained by $0 \leq \hat{\mu} \leq \mu$ and many analyses require the mean number of signal events to be non-negative, making $\hat{\mu}$ to be non-negative as well [63].

A test statistic

$$t_\mu = -2\lambda(\mu), \tag{4.10}$$

can be defined to test the compatibility of data and μ where higher valuef for t_μ suggesting a greater incompatibility. To quantify this incompatibility one can compute the p-value using[5],

$$p_\mu = \int_{t_{\mu,obs}}^{\infty} f(t_\mu|\mu) dt_\mu. \tag{4.11}$$

[5] $t_{\mu,obs}$:t_μ value observed from data
$f(t_\mu|\mu)$:pdf for t_μ

Testing for a discovery of a positive signal, the test statistic can be considered as $\mu = 0$ (under the assumption of $\mu \geq 0$) where rejection of $\mu = 0$ suggest the discovery of the desired signal. For this special case t_μ is taken as $q_0 = t_0$ such that,

$$q_0 = -2\lambda(0), \tag{4.12}$$

where any upward or downward fluctuation is suggesting the presence of the signal sought after. A signal is excluded at 95% confidence level(CL) if $p_1 < 0.05$ where CLs are computed with the asymptotic approximation [48].

The analysis described in this dissertation performs a PL fit for VLQ Top signal strength (μ) regarding the POI. The fitting tool TREXFITTER [67] that interface with `HistFactory` [49] and `RooStats` [74] packages is used to conduct the binned PL studies. As described in the equation 4.7, the statistical and systematic uncertainties for a given bin are treated as Poisson and Gaussian distributions respectively. The notation of γ is used in this analysis to denote the nuisance parameters for a statistical uncertainty associated with a given bin and it is used in the ranking and the correlation plots presented here. In searching for a resonance in the reconstructed T mass m_{tH}, we are using a multijet (as described in section 4.1.6.3), $t\bar{t}$ non all-hadronic MC, $t\bar{t}$ all-hadronic MC and the single-top MC as the background sources for the analysis. A possible signal sample from the table C.4 (that describes the cross-section values for the VLQ mass points at different coupling values) is used in the signal-plus-background fits in the analysis.

4.1.9 Fit to Asimov data results

The analysis is currently practicing a $blind\ analysis$ approach to minimize a possible experimenter's bias towards the analysis [69]. Thus, Asimov data [6] is used in the signal regions for the fits. Three regions were considered in the preliminary level studies where $t\bar{t}$ normalization control region (ttCR) was formed by combining the regions 66,69,96 and 99. The signal regions(SR) 68,86,89 and 98 were combined into two SRs, SR1 (68,86) and SR2 (89,98) for a stable fit as some signal regions were having lower statistics. Later, all the SRs were combined into one SR and its performance during the fits was revealed to be similar to that when there are two SRs. While the fit is performed for the μ in both signal and control regions, it also fits for $t\bar{t}$ scale factor (α as described in eq 4.5) in the ttCR.

The SR is kept blinded from real data by adopting Asimov data which comprised with predicted yields coming from scaled (using the scale factor α) all-hadronic $t\bar{t}$ MC, non all-hadronic $t\bar{t}$ MC, single top MC, and the multijet backgrounds and a VLQ signal sample with a constant cross-section of 10 fb (irrespective of mass and coupling value(κ) of the signal sample) that has a signal injection factor (SI) of 1. The SI factor of 1 is allowing us to perform a signal injection test to realize the expected limits on the cross-section. In the process of deriving an upper limit for the VLQ cross-section as a function of the VLQ mass, we regarded the cross-section for any considering VLQ sample to have an expected value of 10 fb, thus making any limit set on μ represent a limit set on the observed cross-section divided by 10 fb.

The performance of the likelihood fit in an analysis can be improved by removing or symmetrizing the NPs involved. This process is known as pruning. The shape and the normalization components are dropped/pruned if their impact on the nominal distribution of m_{tH} is less than 1% and this method therefore significantly reduces the weight on computing resources. Fig. 4.7 shows the pruning results for this analysis.

[6] Asimov data: A method of estimation where an ensemble of data sets being replaced by a single representative dataset [48].

The one-sided systematic uncertainties observed in this analysis such as CategoryReductionJETMassResWZ and the CategoryReductionJetMassRessHbb were treated with the one-sided symmetrization, in order to obtain a valid up and down systematic variation. The one-sided symmetrization was implemented by regarding the deviation of the number of events in a given bin from its nominal value and take up and down systematic variation to be the same. Elaborating this in layman terms, if the nominal value for the number of events in a given bin is 100 and the one-sided systematic variation observed is 90, once the symmetrization is implemented the up variation will become 110 events and the down variation will be 90.

The two sided symetrization was implemented on the two-sided systematics using an algorithm of,

$$\text{systematic variation in a given bin} = \frac{\sigma_{up} - \sigma_{down}}{2} \tag{4.13}$$

where σ_{up} stand for the systematic up variation and σ_{down} stands for the systematic down variation and the TREXFITTER was used in implementing both symmetrization processes. Symmetrization is, therefore, ensuring that when the PL fit performs on the bins, the up and down systematic variations won't both here more or less events than to that of the nominal value. The fig. 4.9 shows the list of nuisance parameters which are un-symmetrized and the fig. 4.10 shows when they are symmetrized. An Asimov sample of predicted yields form the backgrounds and a VLQ sample with a mass of 2.0 TeV, $\kappa =$ 0.5, and a SI $= 0$ is used here. The one-sided systematics CategoryReductionJETMassResWZ and the CategoryReductionJetMassRessHbb before and after the symmetrization are shown in fig. 4.8.

A comparison of pre-fit and post-fit for the invariant mass distribution of Higgs and top quark jets in the $t\bar{t}$ CR and the SR are given in fig 4.11 and fig. 4.12.

When there are several systematic NPs involved in an analysis, it is important to evaluate which individual NP has a significant impact on the signal strength μ. The TrexFitter is facilitating such studies by ranking the NPs. The impact imposed on the μ by a certain NP($\Delta\mu$) is measured by studying the shift of μ from the nominal fit to another fit that has the considered NP set fixed at its maximum-likelihood estimatior value($\hat{\theta}$) found from the nominal fit ($\hat{\theta} \pm x$) where x is $pre-fit$ impact ($x = \Delta\theta = 1$) or $post-fit$ impact ($x = \Delta\hat{\theta} \leq 1$) with a 1$\sigma$ uncertainty for the NP[7]. The fig. 4.13 describes the NP ranking of the γ NPs and the fig. 4.14 describes the NP ranking for the NP for the systematics considered in this analysis.

The fit results were then used to generate a correlation matrix which enables to study the correlations among the NPs considered. The NPs that have at least 20% of a correlation coefficient with another given NP is only used in deriving the correlation matrix given in fig 4.15.

We initiated the fitting studies by combining the SRs (68,86,89,98) into two SRs and later tested with a single SR which combined all the SRs. The single SR performed within 1% in the expected 95% CL for the VLQ mass of 2.0 TeV, $\kappa = 0.5$, and SI $= 0$ when compared with the two SRs approach during the fitting studies. The aforementioned 1% difference was observed when deriving the upper limit for the signal strength where the single SR had a value of 1.89 while the two SRs were having a value of 1.88. Thus, the single SR choice was selected to continue with the analysis as it is capable of increasing the statistics in the SR while stabilizing the fitting procedures.

[7] $\Delta\hat{\theta}$ is the uncertainty of $\hat{\theta}$

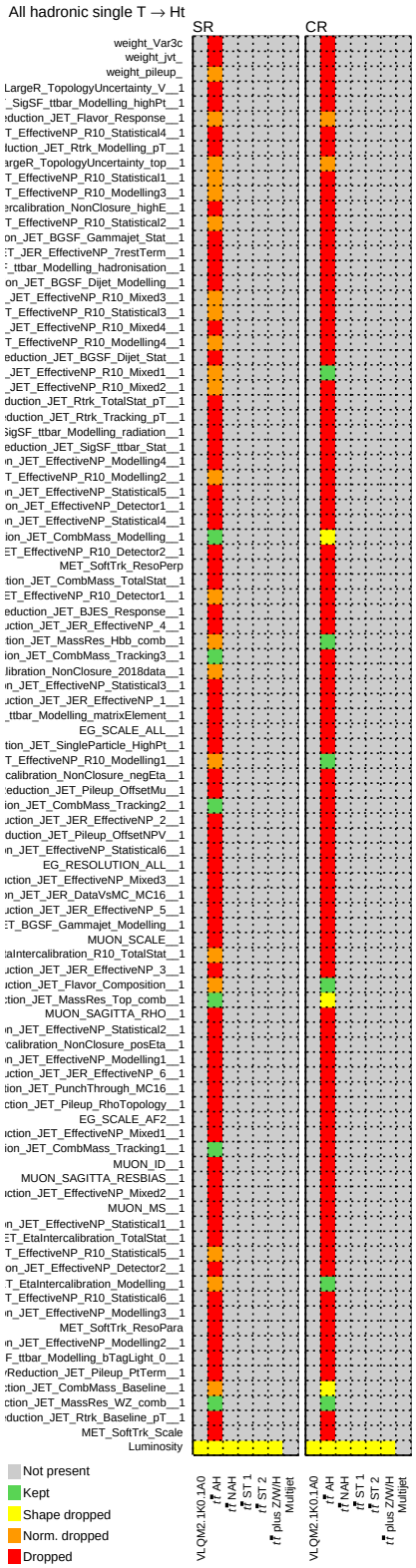


Figure 4.7: Pruned systematic uncertainties are summerized here. Systematics given in red are completely dropped, in yellow have their shape dropped, in orange have their normalization dropped and in green are retaining both the shape and the normalization in the fit.

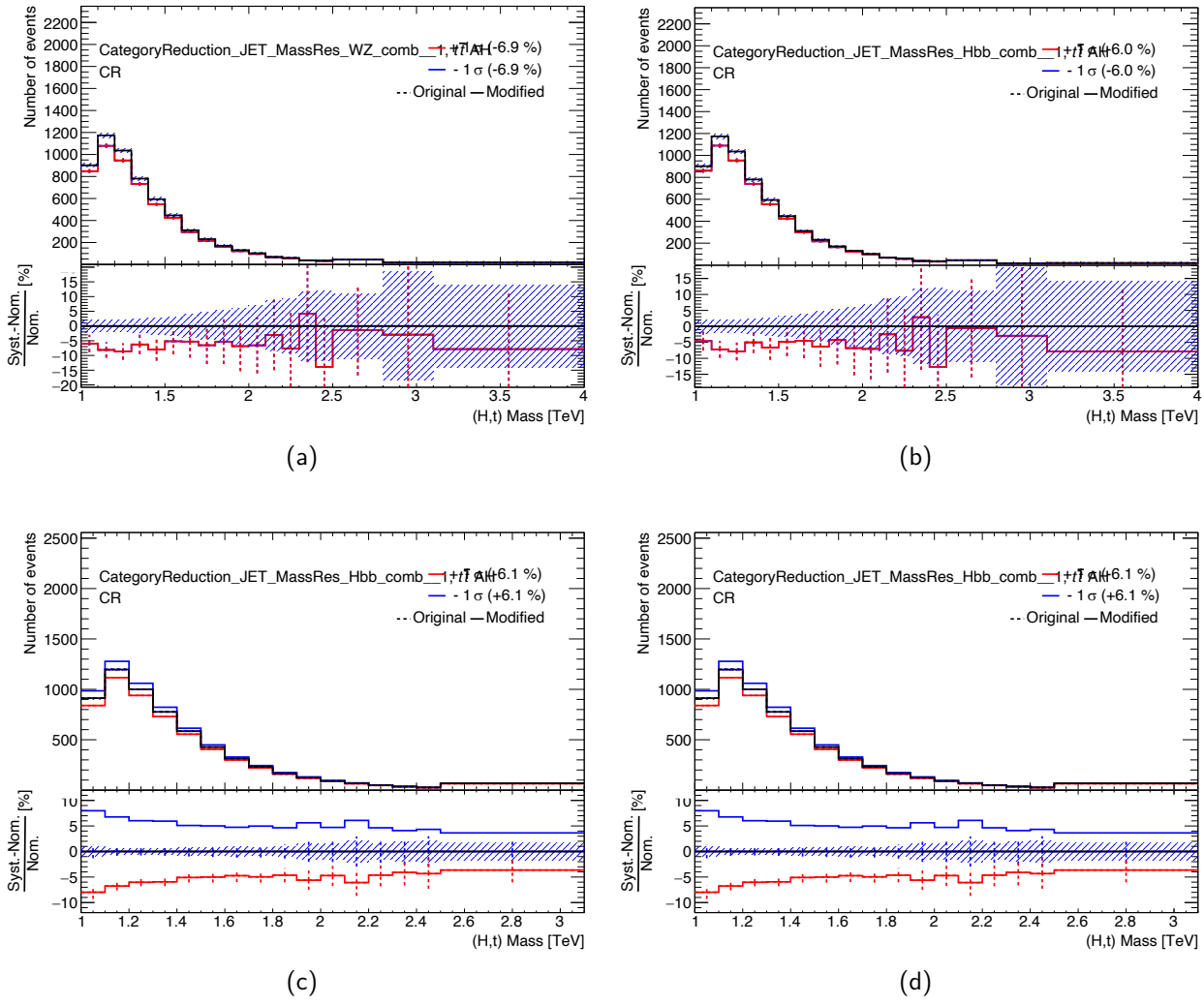


Figure 4.8: The top row shows the unsymmetrized systematics of fig. 4.8a CategoryReductionJETMassResWZ and the fig. 4.8a CategoryReductionJetMassRessHbb. Prior to the symmetrization, no up variation can be seen. The bottom row shows the symmetrized systematics of fig. 4.8c CategoryReductionJETMassResWZ and the fig. 4.8c CategoryReductionJetMassRessHbb and both the up and down variations are now visible. These systematic distributions are from the $t\bar{t}$ normalization region.

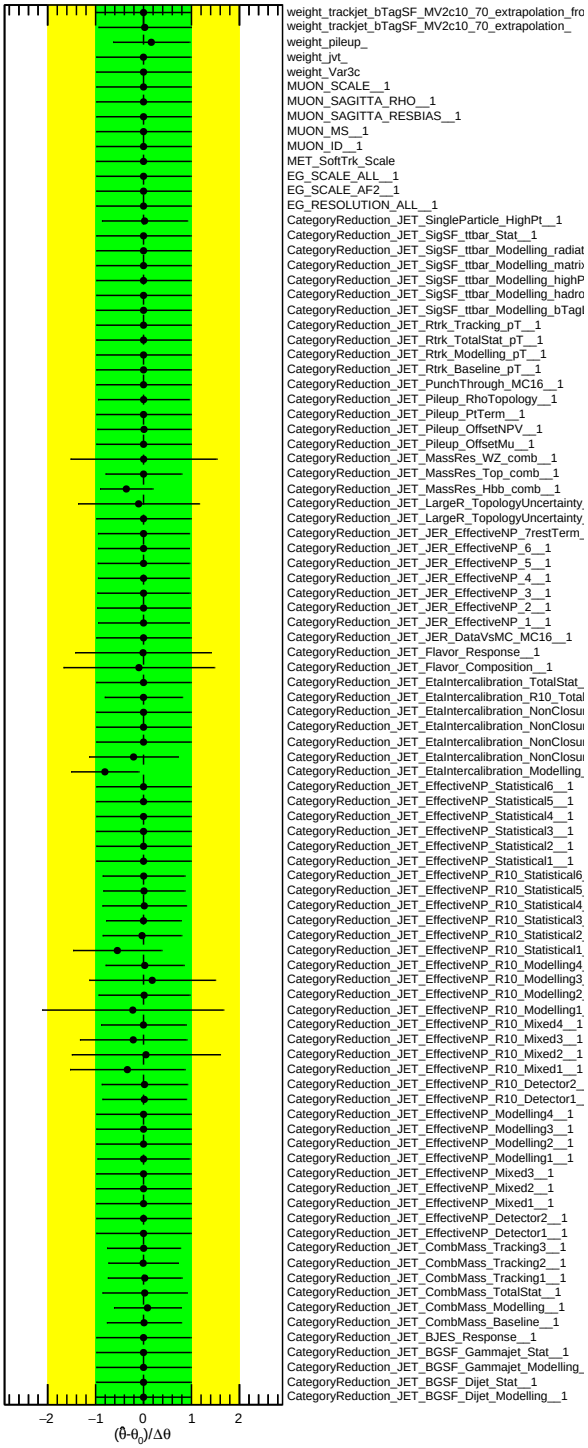


Figure 4.9: Unsymmetrized systematic nuisance parameters with their pulls and constrains for a fit using an Asimov dataset with a VLQ mass of 2 TeV, $\kappa = 0.5$ and SI $= 0$ (background only). If the fit is deviated from 0, the corresponding NP is pulled and is constrained if the error is less than 1.

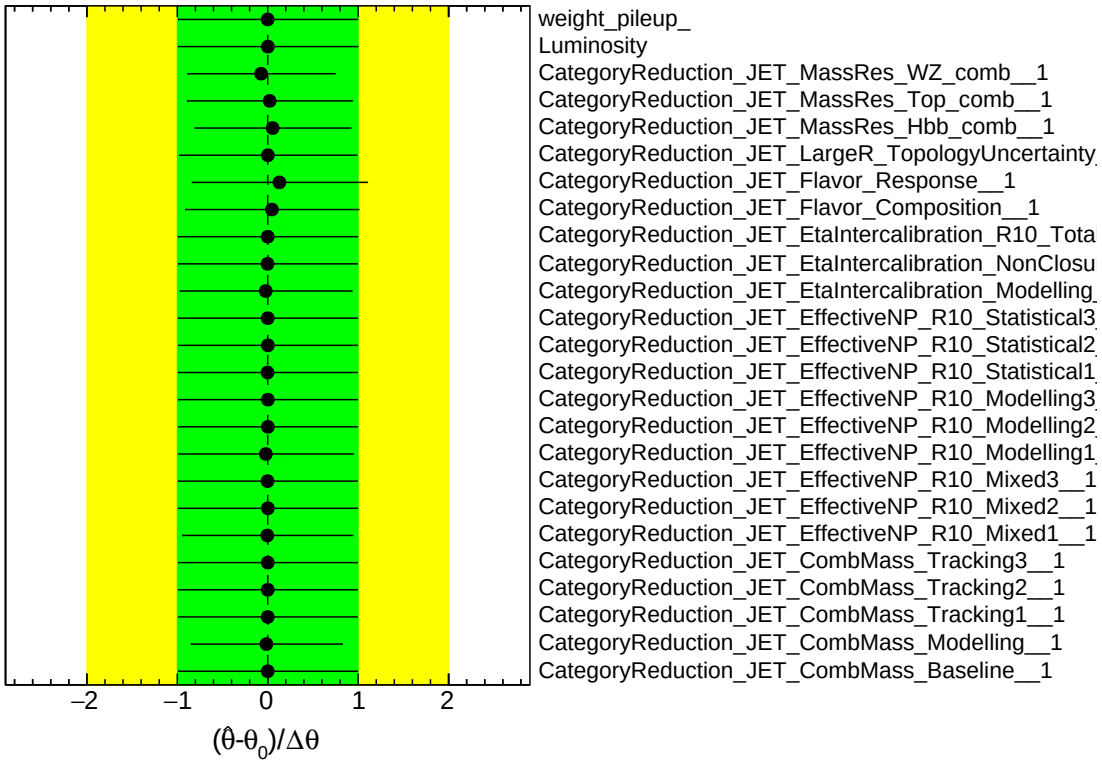
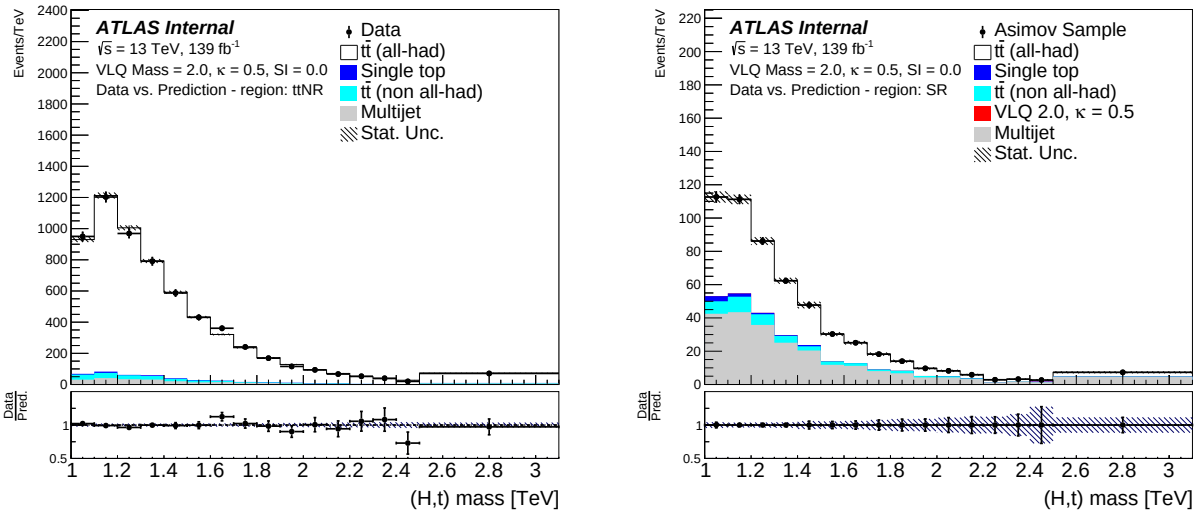
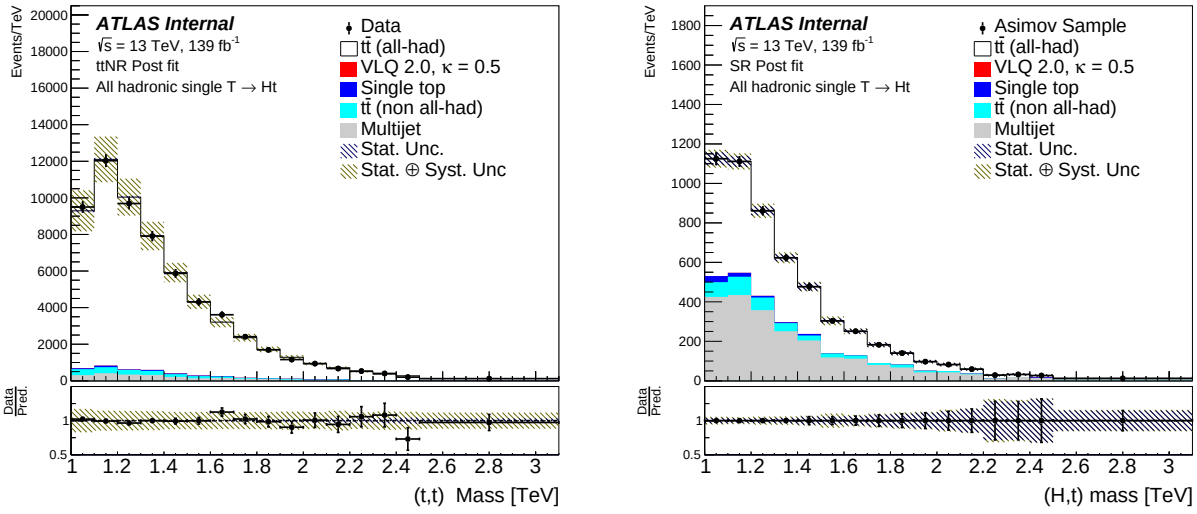


Figure 4.10: Symmetrized systematic nuisance parameters with their pulls and constrains for a fit using an Asimov dataset with a VLQ mass of 2 TeV, $\kappa = 0.5$ and SI $= 0$ (background only). If the fit is deviated from 0, the corresponding NP is pulled and is constrained if the error is less than 1.



(a) (b)

Figure 4.11: Pre-fit invariant mass distributions of Higgs and SM top jets, fig. 4.11a for $t\bar{t}$ normalization region(NR) and fig. 4.11b for the signal region(SR) for an Asimov dataset with a VLQ mass of 2.0 TeV, $\kappa = 0.5$ and SI $= 0$.



(a) (b)

Figure 4.12: Post-fit invariant mass distributions of Higgs and SM top jets, fig. 4.12a for $t\bar{t}$ normalization region(NR) and fig. 4.12b for the signal region(SR) for an Asimov dataset with a VLQ mass of 2.0 TeV, $\kappa = 0.5$ and SI $= 0$.

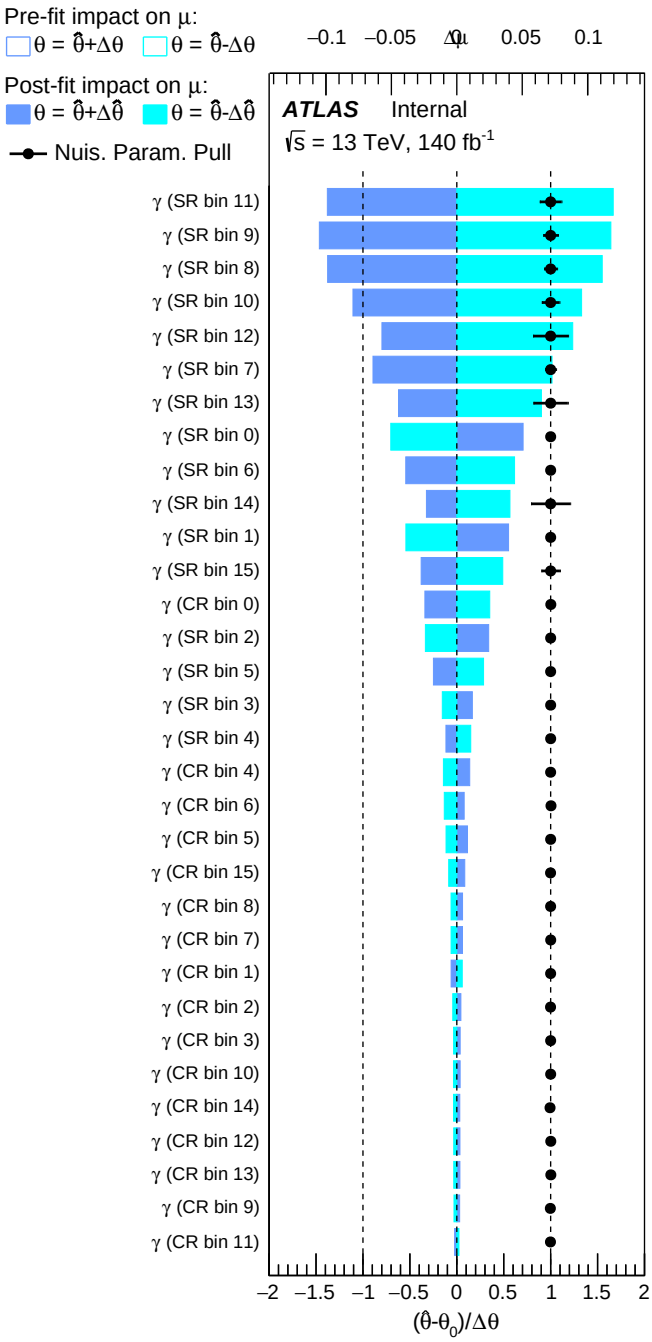


Figure 4.13: The γ nuisance parameters are ranked here based on their impact on the signal strength. An Asimov dataset with a VLQ mass of 2.0 TeV, $\kappa = 0.5$ and SI $= 0$ is used. The opened dark(light) blue boxes represent the impact of each γ NP on the signal strength prior to the fit. The filled dark(light) blue area gives the impact on the signal strength from each γ NP after the fit with $+1$(-1) uncertainty (top axis). The black points in the ranking plot corresponds to the fitted values (bottom axis) of the NPs and their errors are represented by the error bars.

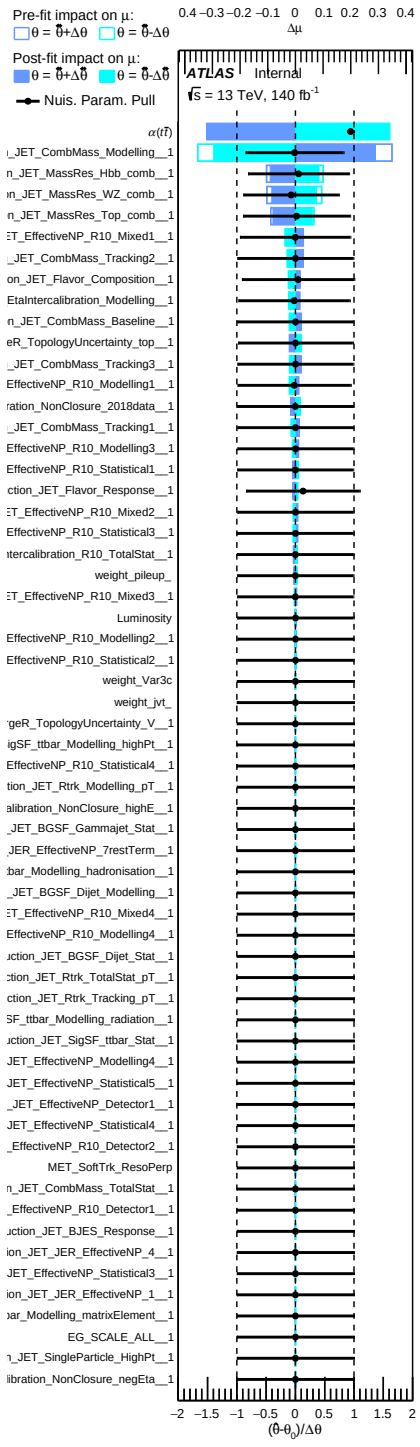


Figure 4.14: The systematic nuisance parameters are ranked here based on their impact on the signal strength. An Asimov dataset with a VLQ mass of 2.0 TeV, $\kappa = 0.5$ and SI $= 0$ is used. The opened dark(light) blue boxes represent the impact of each NP on the signal strength prior to the fit. The filled dark(light) blue area gives the impact on the signal strength from each NP after the fit with $+1$(-1) uncertainty (top axis). The black points in the ranking plot corresponds to the fitted values (bottom axis) of the NPs and their errors are represented by the error bars.

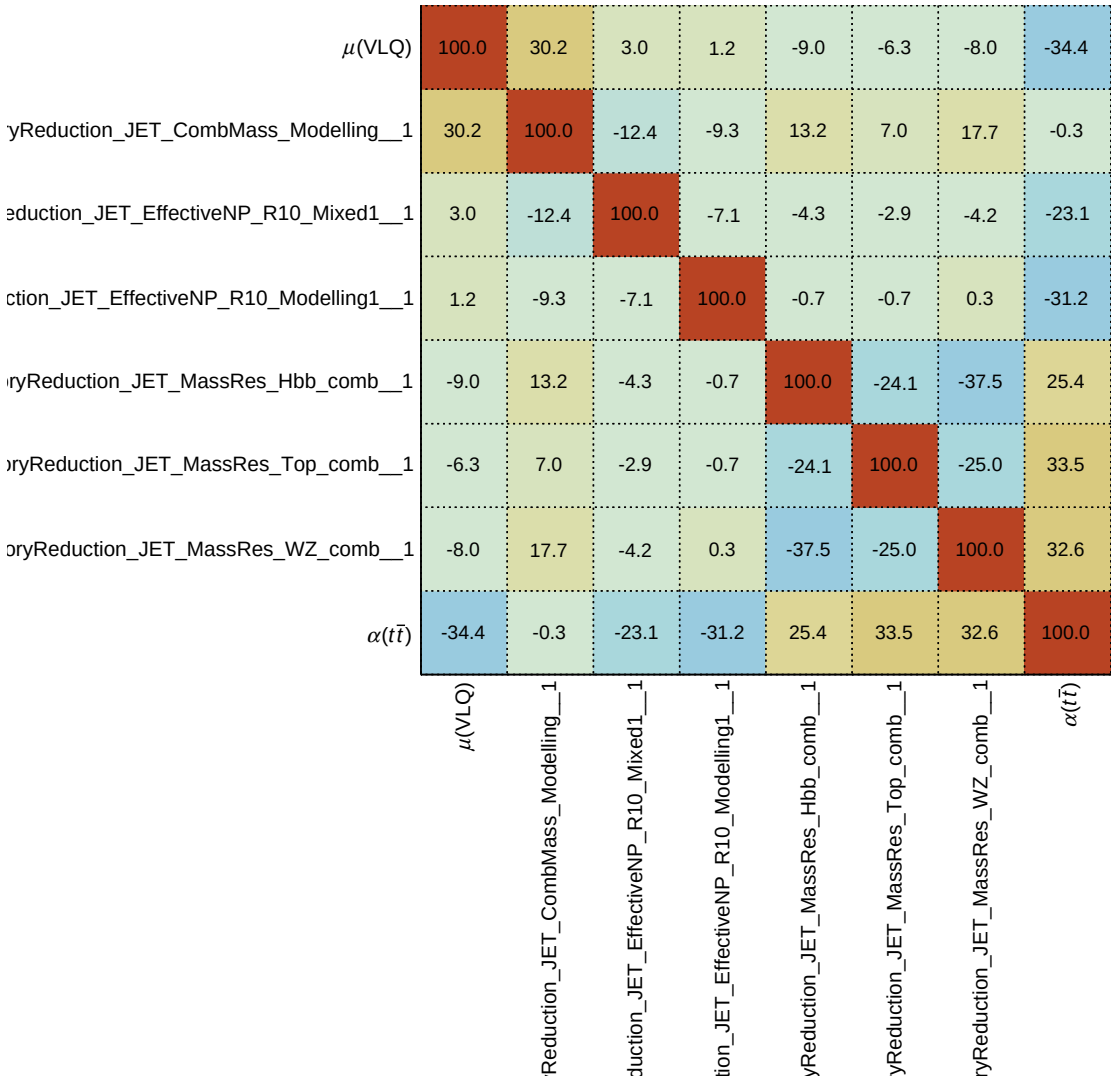


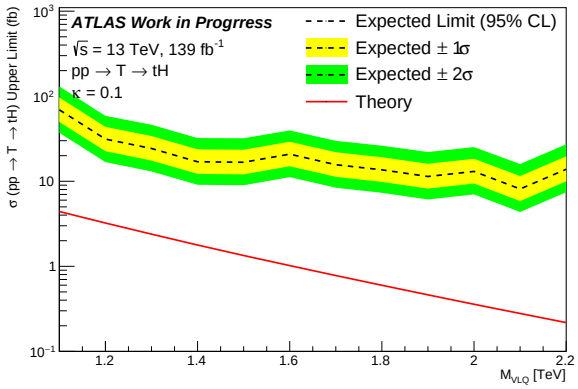
Figure 4.15: The correlation matrix describe the correlations between the NPs in the fit using an Asimov dataset with a VLQ mass of 2.0 TeV, $\kappa = 0.5$ and SI $= 0$. NPs that have at least a 20% correlation among each other are depicted here.

4.1.10 Results

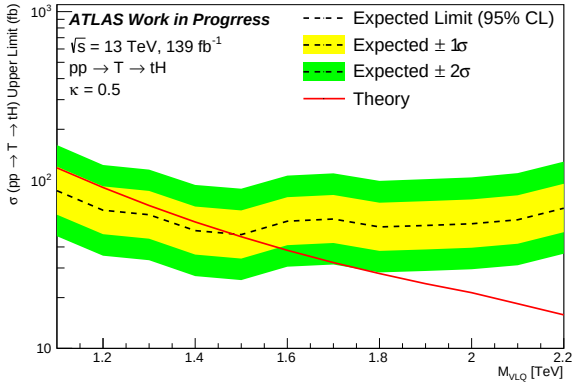
A search for single produced vector-like T quark decays into an all-hadronic final state via $T \rightarrow Ht$ decay mode has been presented with using an integrated luminosity of 139 fb^{-1} collected by the ATLAS detector during the period of 2015 to 2018. The analysis is searching for a final state with boosted Higgs that decays to two b-tagged jets and a boosted SM top-tagged jets. The analysis is currently in the blinded stage and thusly, an Asimov dataset (SI $= 0$) is used in deriving the expected upper limits for the cross-section of the $T \rightarrow Ht$ process as described in the fig. 4.16. The cross-section is tested for three different coupling constant (κ) values of 0.1, 0.5, and 1.0.

The dashed line in all three figures in fig 4.16 is the expected upper limit at the 95% CL on the cross-section for the physics process of $T \rightarrow Ht$ in the absence of m_{tH} signal. The yellow band around the expected upper limit line is corresponding to the ± 1 standard deviation and the green band corresponds to the ± 2 standard deviations from the expected upper limit line. The red line is representing the theoretical upper limits for cross-section for a given coupling constant value. For the coupling constant of κ=0.1, the theoretical upper limit is lower than the expected upper limit at all mass points. As we are currently in the "blinded"stage, the observed upper limit at 95% CL isn't provided here. It will be a solid black line that is showing the upper limit at the presence of the m_{tH} signal. As the solid line is being the upper limit for the cross-section with 95% CL, any cross-section above it would be excluded.

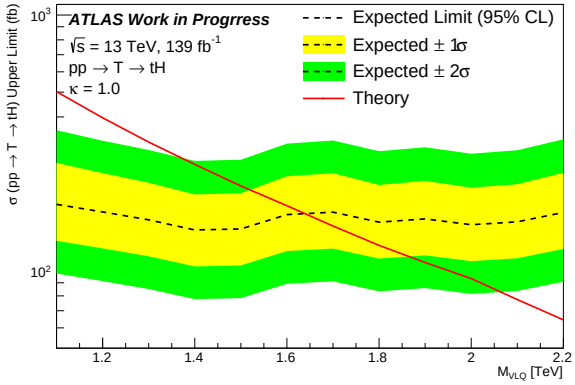
As it was mentioned in the section 1.1.4.3, the single VLQ production is model dependent, and therefore for each mass point, the cross-section for the $T \rightarrow Ht$ process will be different as described in table C.4. It is observed that the expected 95% CL is increasing along with the κ.



(a) $\kappa = 0.1$



(b) $\kappa = 0.5$



(c) $\kappa = 1.0$

Figure 4.16: Upper limits derived for the VLQ cross-section using the CL_s method at κ values of 0.1, 0.5, and 1.0. An Asimov dataset with a SI $= 0$ is used. The expected limit is represented by the dashed line which is deriving from the background only hypothesis. The red line represents the theoretical upper limit

CHAPTER V

Flavor tagging parameterizations and pixel size characterization studies for HL-LHC upgrade

5.1 Introduction

The LHC will undergo an upgrade between 2024 and 2027 to become the High-Luminosity Large Hadron Collider, allowing the LHC to gather an integrated luminosity of 4000 fb^{-1} from the photon-photon collisions as described in the fig. 5.1. The increment in the luminosity up to 7.5×10^{34} $cm^{-2}s^{-1}$ brings a challenge of processing the high rate of data reception and the excessive radiation damage to the ID of the current ATLAS detector. To address these challenges, the ID will be replaced with a new, all-silicon tracking detector (Inner Tracker: ITk). The introduction of the new ITk expects to withstand the radiation conditions under the high luminosity and deliver better tracking performance.

The ITk is comprised with two sub systems,

- The Strip Detector
- The Pixel Detector

Comparing with the ID, which has a pseudo-rapidity (η) coverage [1] of 2.5, the new ITk will have a broader η coverage up to 4.0 with precise tracking efficiency and performance at 200 pile-up events. The Pixel Detector will provide this coverage by using its five-barrel layers and η dependent end-cap rings. The Strip Detector, which will be outside the Pixel detector, is capable of providing a $|\eta| < 2.7$ by using its four-barrel layers and the six end-cap disks as shown in fig. 5.2‘[31].

[1] Refer fig. 2.5

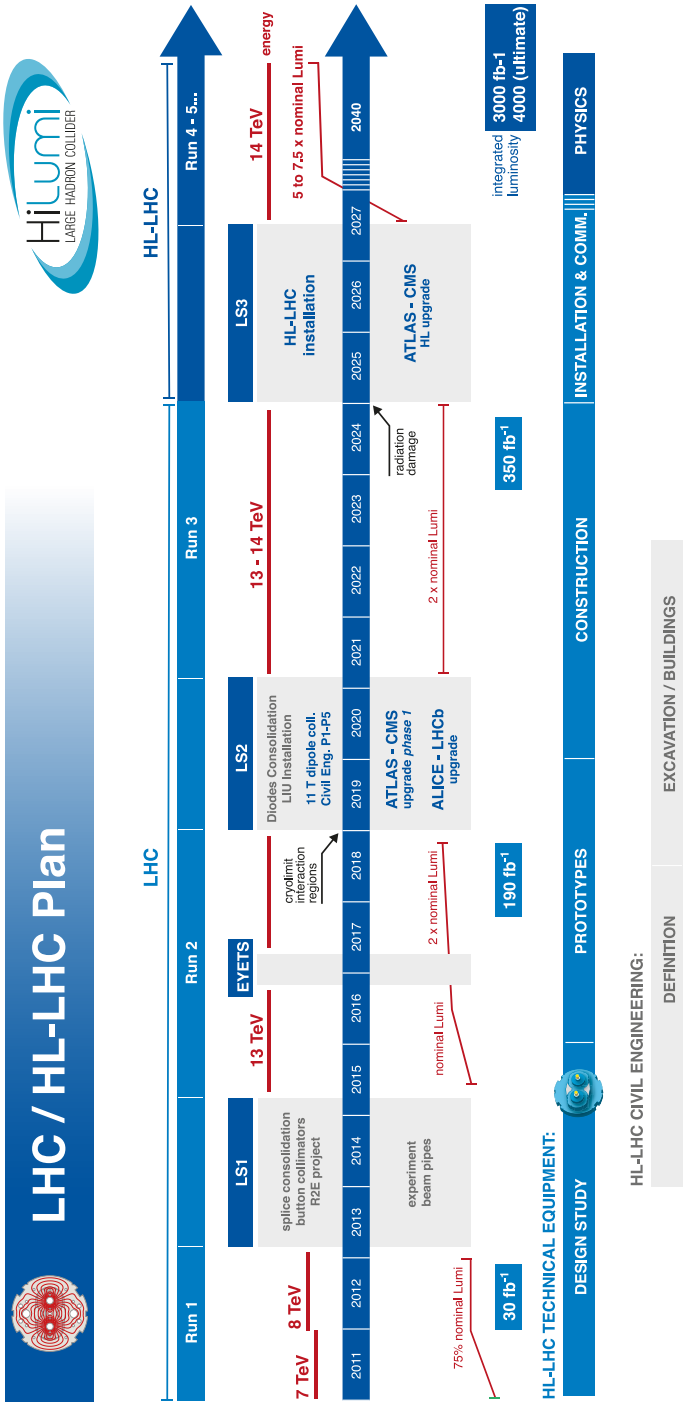


Figure 5.1: Project schedule for the proposed High Luminosity LHC.

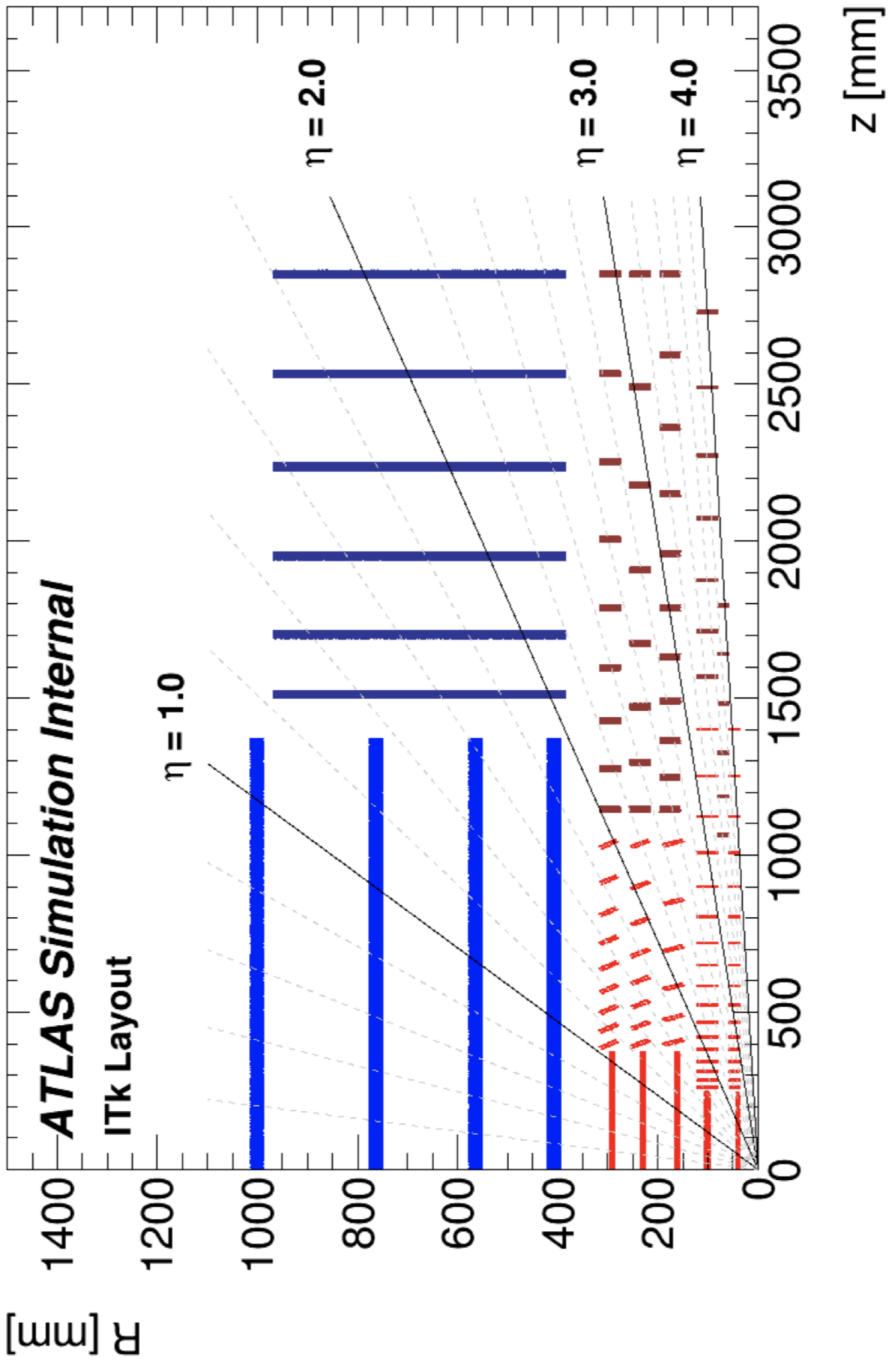


Figure 5.2: Schematic diagram of the Strip and the Pixel detectors in the ITk. The Strip Detector is given in blue and the Pixel Detector is given in red color. The vertical axis indicates the radius of the ITk from the point of interaction, the horizontal axis is the beam axis that origins from the interaction point.

Many physics analyses such as Vector-Like Quarks have an emphasis on the proper identification of b-hadrons with an accuracy as the Vector-Like Quarks have a preference to decay into a heavier SM 3^{rd} generation "b"quark.

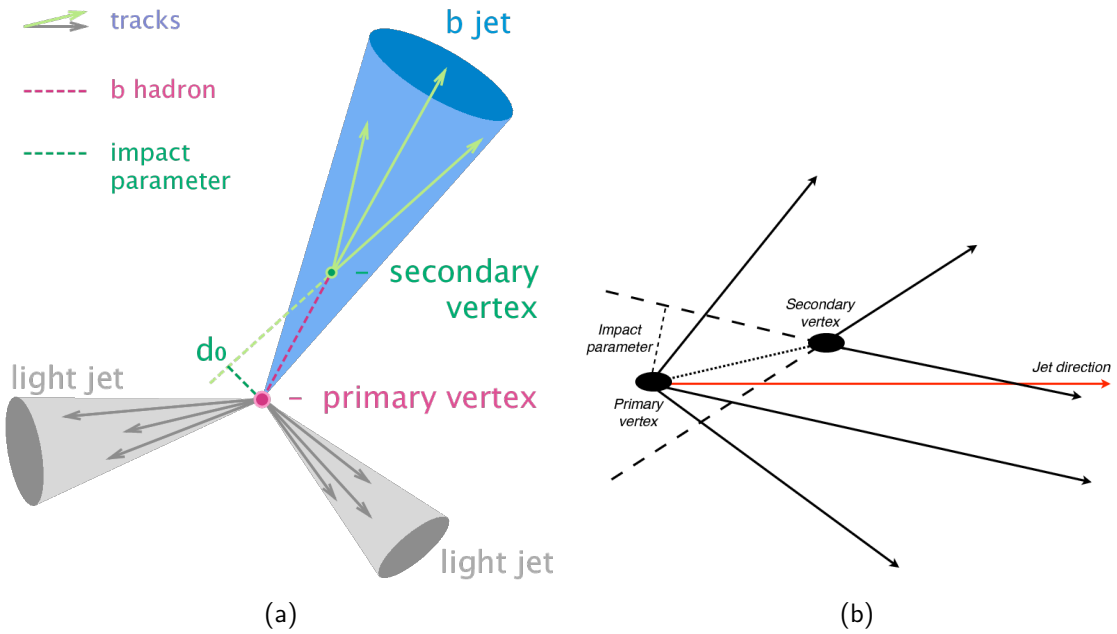


Figure 5.3: Fig. 5.3a shows the origination of a b-jet from the point of collision (primary vertex). The b-hadron travels a further more distance due its longevity, resulting a secondary vertex within the b-jet. The impact parameter (d_0) as given in fig. 5.3b depicts the diviation of the the secondary vertex from the primary vertex [44].

some of the particle jets originating from the primary vertex have the chance of including a b-hadron in them which gives their name of "b-$tagged$"jets or "b-$flavored$"jets. Compared with charm quark and the light hadrons, the b quark has a slightly longer life span ($\sim$1.5 ps or $c\tau$ $\sim$450 μm). This longevity allows the b-hadron to travel farther from the primary vertex to form tracks which can be separated from the tracks originated from the primary vertex through their larger impact parameter (which is illustrated in fig. 5.3) values.

The rate of identifying true b jets is important in developing and optimizing b-tagging (this is true for any flavor tagging efforts) algorithms. This rate is referred to as the "efficiency"and one can define the efficiency for a particular flavor for example b-tagging efficiency [46]. Thus, it is intended by ATLAS to maintain or improve the b-tagging

performance, after the transition from current ID towards the ITk.

Therefore, as part of the ongoing ITk upgrade studies, we were interested to study the following,

1. Define the best performing Boosted Decision Tree (BDT) parameters for the MV2c10 b tagger by using the $t\bar{t}$, single-top and Z' samples. These samples are used mainly in studying the b-tagging performance in the very forward (i.e. $|\eta| >$ 2.7 up to 4.0) region of the ITk.

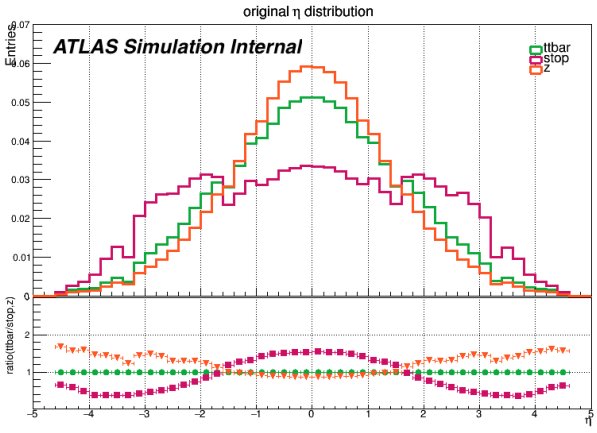
2. Characterize the impact of different pixel size choices in consideration of $25\times100\mu\text{m}^2$ vs $50\times50\mu\text{m}^2$.

3. Provide updated flavor tagging parametrizations for the upgrade physics studies. In particular, we derived both the related efficiencies and uncertainties.

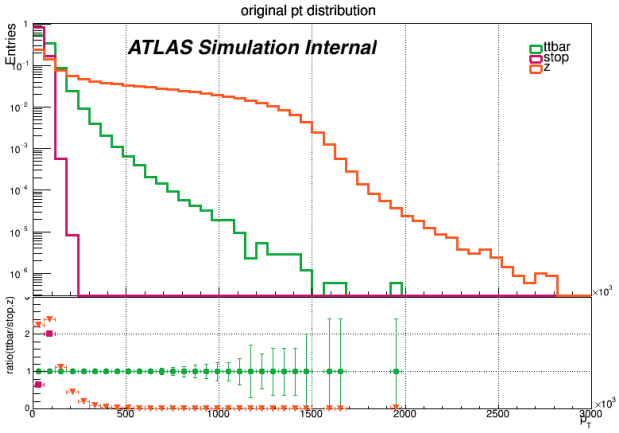
5.2 Definitions of optimized Multivariate algorithm (MV2) parameters for the upgrade studies

As the new ITk will record events in the very forward regions ($|\eta| >$ 2.7 up to 4.0), it is important to evaluate the $b-$tagging efficiencies at different regions in the ITk. For the purpose, we have trained a combination of $t\bar{t}$, Z′, and single-top samples (Hybrid sample) using a multivariate algorithm as the discriminator. Samples studied and presented here correspond to the "Step 3.0 ITk Layout"as defined in Ref. [52]. Each of the three samples provides higher statistics for a particular case such as,

- $t\bar{t}$: Provide statistics with b-jets
- Z′: Provide statistics at higher transverse momenta (p_T)
- single-top: Provide statistics at higher pseudo-rapidity (η)



(a)



(b)

Figure 5.4: Fig. 5.4a shows the η distribution of the $t\bar{t}$, Z′ and single-top samples and the fig. 5.4b refers to the p_T distribution of the same samples.

The hybrid sample is then fed into a Multivariate Algorithm (MV2), which is based on a Boosted Decisions Tree (BDT) algorithm to maximize the light-jet rejection for three $b-$tagging efficiency working points of 70%, 77%, and 85%.

The MV2 algorithm or MV2 tagger is a high-level tagger that takes the lower level flavor taggers such as IP2D, IP3D, SV1, and JetFitter [2] along with the η and p_T of the samples fed in. The MV2 tagger has three variants based on the background fraction of jets that are tagged as "charm (c)" in the training samples [6],

[2]Please refer to Appendix E for the flavor taggers

- MV2c00 : 0% c-jet fraction
- MV2c10 : 7% c-jet fraction
- MV2c20 : 15% c-jet fraction

We have used a MV2c10 tagger which is optimized at c-jet fraction of 7% and the light-jet fraction of 93% in the background while treating b-jets as the signal at the training level. [3] The MV2c10 tagger used in this study employs a Boosted Decisions Tree algorithm for discriminating light-jets in favor of the b-jets using the following parameters,

Parameter	Default value
Number of trees	1000
Depth	30
Minimal node size	0.05%
Cuts	200

Table 5.1: BDT parameters with default values for the MV2 tagger

The default values given in the table 5.1 are the optimized values for MV2c10 used in the RUN 2 as described in Table 2 of Ref. [19].

We used the performance optimization framework of "Run2BtagOptimisationFramework" [4] for both the BDT parameters optimizations for the hybrid sample studies and the pixel size characterization studies. The "Run2BtagOptimisationFramework"allows the BDT parameters to be tweaked and produce new samples on the grid, which later can be feed into the MV2 tagger for classification.

[3] MV2 Tagger for $b-$tagging
[4] Run2BtagOptimisationFramework for $b-$tagging performance studies

We began the training process with the $t\bar{t}$ sample that was generated with 500k events with an average number of interactions per bunch crossing of μ=200 for the ITk layout by splitting the sample into 70% for training and 30% for testing. It was observed during the sample training that the default parameters assigned to the MV2c10 tagger, gives rise to an algorithm overtraining where the BDT algorithm performance for the training sample outperformed the testing sample significantly as shown in fig. 5.5. Overtraining in BDT can arise due to several factors such as attempting to split the nodes until all the leaf nodes are pure in the training sample, could result in the performance in the training sample to be better than when it's on the test sample (one can observe in fig 5.5 that the training performance in solid lines is more closer towards the perfect separation in the upper right corner than to that on the testing sample performance given by the dashed lines) [43].

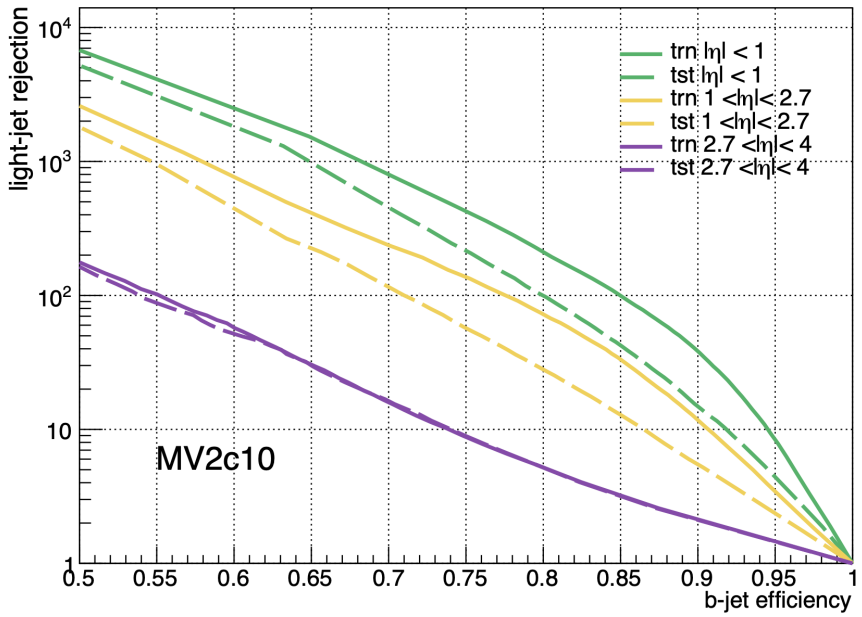
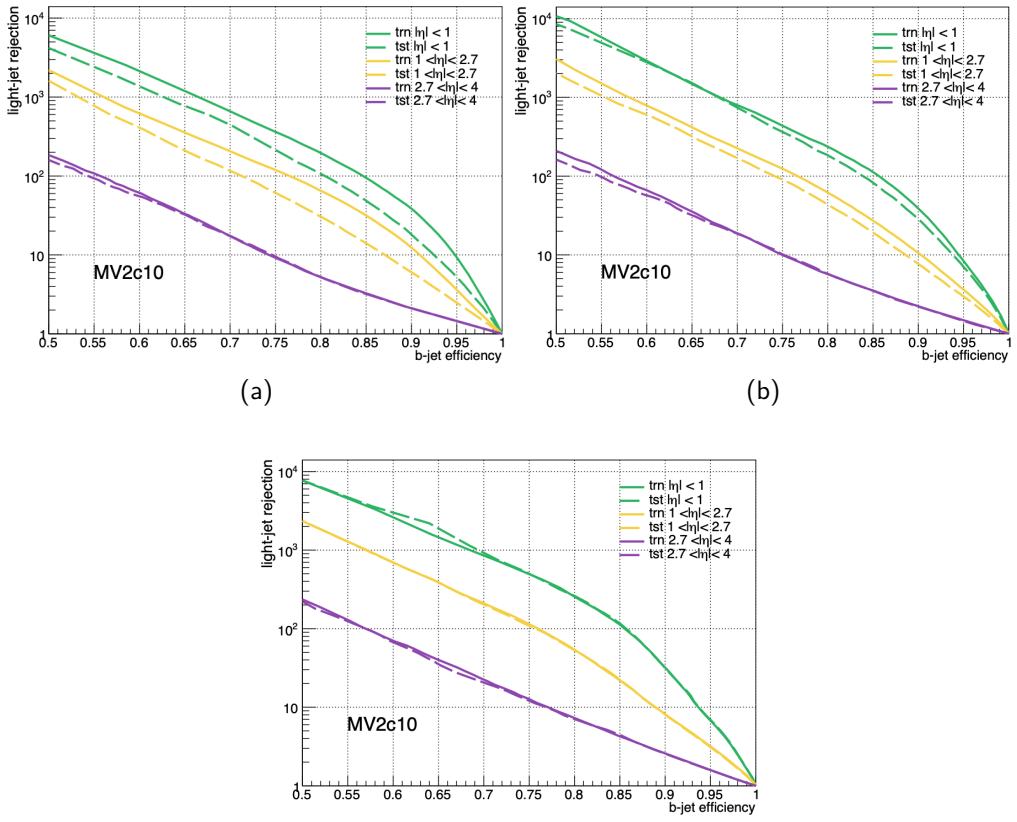


Figure 5.5: Fig. 5.6a shows the light-jet rejection vs $b-$tagging efficiency for the $t\bar{t}$ sample with 500k events and 200 pile-up events trained with the MV2c10 tagger. The dafalut parameters described in the table 5.1 are used here. The solid lines represent the training sample and the dashed lines represent the testing sample. Performance of the MV2c10 tagger was observed through three different pseudo-rapidity regions in the ITk depicted by the colors of green for $|\eta|$ <1.0, yellow for 1.0< $|\eta|$ <2.7 and purple for 2.7< $|\eta|$ <4.0.

This algorithm overtraining is because the size of the simulated $t\bar{t}$ sample (500k events) used in the upgrade studies is a factor of 10 times smaller than that of the samples used in the RUN 2 studies thus, the parameters that were optimal for RUN 2 are not optimal here. In addressing the overtraining, we reduced the depth parameter into lower values (30 being the default depth parameter) of 20, 10 and 3 by reference to page 138 in the user guide for TMVA (Toolkit for Multivariate Data Analysis with ROOT [55]) which implemented the BDT described here [5]. The gradual reduction of the depth parameter reduced the algorithm overtraining, especially in the pseudo-rapidity regions of $|\eta|$ <1.0 and the 1.0< $|\eta|$ <2.7, and it is also observed that the light-jet rejection improved. Results obtained in the depth reduction studies are given in fig. 5.6.



(a) (b)

(c)

Figure 5.6: Light-jet rejection vs $b-$tagging efficiency for the $t\bar{t}$ sample. Fig 5.6a, fig 5.6b and fig 5.6c shows the training and testing performance at depth parameter values of 20,10, and 3, respectively.

[5] Refer to page 138 in the user guide for TMVA

5.2.1 Results

By observing the reduction of algorithm overtraining by reducing the depth parameter, we prepared the hybrid sample by combining all three samples of $t\bar{t}$, single-top and the Z′ into a single sample and split it such that 70% was allocated for training and the rest 30% for testing. After training the hybrid sample under various BDT parameters [6] as shown in table 5.2, highest values for the light-jets rejection at three different $b-$tagging efficiencies of 70%, 77%, and 85% were observed for the BDT parameter values that are highlighted in table 5.2.

BDT parameter test values				Light-jet rejection		
Cuts	Depth	No.T	mns	b-jet eff 70%	b-jet eff 77%	b-jet eff 85%
200	12	1000	0.05%	178.4	78.17	23.51
200	12	1000	0.10%	203.74	85.36	23.44
200	12	500	0.05%	190.55	81.19	23.5
200	12	500	0.10%	198.43	83.47	23.76
100	20	500	0.10%	190.55	80.75	23.67
200	20	500	0.10%	199.15	82.21	23.67
100	20	500	0.10%	196.32	81.53	23.53

Table 5.2: Optimization of light jet rejection at different BDT parameters for different b-tagging efficiency working points. The BDT parameters are Cuts: number of pruning cuts, Depth: tree depth, No.T: number of trees and mns: minimal node size.

[6] Cuts:Number of grid points in variable range used in finding optimal cut in node splitting

Depth:maximum depth of the decision tree allowed

Number of trees:Number of trees in the forest

Minimal node size:Minimum percentage of training events required in a leaf node (default: Classification: 5%, Regression: 0.2%)

Please refer to page 116 in the TMVA Users Guide

5.3 Pixel size choice studies for 25×100μm^2 vs 50×50μm^2

In comparison to the pixel pitch of 50x400μm^2 (or 50×250μm^2 in the Insertable B-Layer, which is the innermost layer in the current pixel dector [59]) available in the current pixel detecter [41], two new pixel pitch configurations are proposed for the new ITk as follows,

- 25×100 μm^2
- 50×50 μm^2

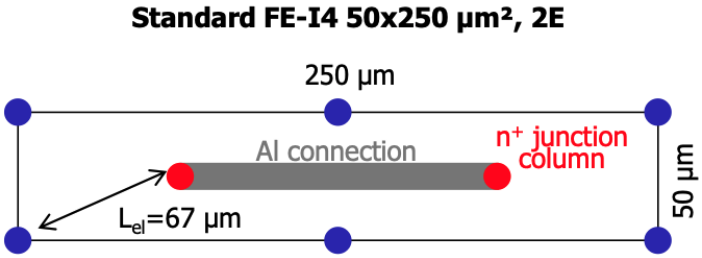


Figure 5.7: Schematic diagram of 50×250μm^2 pixel in the in the Insertable B-Layer. 2E stands for two n$^+$ junction columns depcted in red and L_{el} is the inter electrode distance. [62]

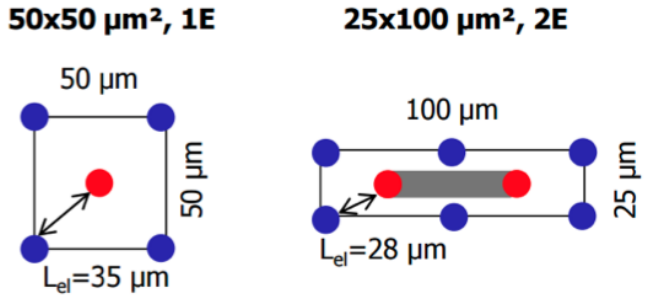


Figure 5.8: Schematic diagrams for proposed pixel size choices of 50×50μm^2 (left) and 25×100μm^2 (right). 1E/2E stands for one/two n$^+$ junction column(s) depcted in red and L_{el} is the inter electrode distance. [62, 53].

We used the following $t\bar{t}$ samples (with ∼500k events) that are simulated according to the pixel configurations mentioned above to train with the MV2 tagger to optimize the rejection of light-jets in favor to the $b-$tagging efficiency.

- 25×1000 μm^2
  - ▶ mc15 14TeV.117050.PowhegPythia P2011C ttbar.recon.AOD.e2176 s3348 s3347 r11004/
- 50×50 μm^2
  - ▶ mc15 14TeV.117050.PowhegPythia P2011C ttbar.recon.AOD.e2176 s3348 s3347 r10899 r11002/

The samples for each pixel configurations were trained using the following Reference Histograms [7],

- 25×1000 μm^2
  - ▶ Ref Histos ConfigV05 ttbar 25x100 Step3 AC Mu200 r39mm StdFormat.root
- 50×50 μm^2
  - ▶ Ref Histos 50x50 Config6 r39mm r11002.root

The light-jet rejection as a function of $b-$tagging efficiency for the pixel pitch choices showed an over-training (training sample over-perform over the test sample performance, if the tagger is trained optimally, the testing sample performance should not either over or underperform). for the BDT parameters described in table 5.1. In the case of 50×50 μm^2 the over-training was most for region 1.0 $<|\eta|<$ 2.7 while for 25×100 μm^2 it was worst for 1.0 $<|\eta|$. Thus, we reduced the depth parameter from 30 to 3 and observed the over-training to be gone. Further studies during the sample training stage showed that depth parameter at a value of 12 yields the least overtraining results for both pixel choices across all the $|\eta|$ regions.

[7] The reference histograms are used for lower-level b-tagging algorithms IP2D and IP3D based on the d0/z0 significances for b/c/light jets hypotheses. Such histograms are used to extract the final log-likelihood ratios and perform the discrimination [38, 39]

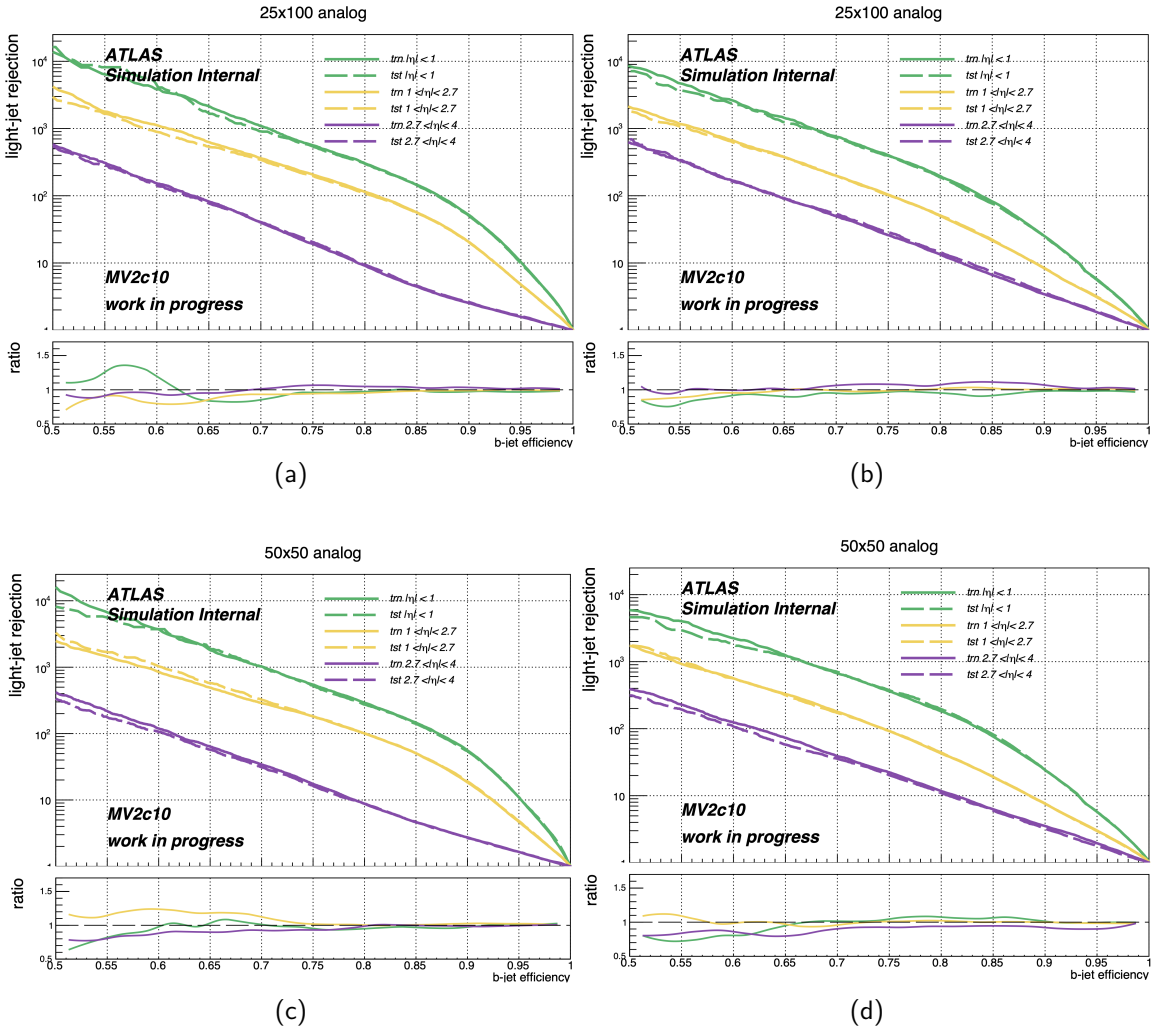


Figure 5.9: Fig (a) shows the light-jet rejection vs $b-$tagging efficiency when the depth parameter $= 30$ for the pixel choice of 25×100 μm^2 and the fig (b) refers to the similar when depth $= 3$ and the overtraining is observed in the region of $1.0 < |\eta|$. Fig (c) shows the light-jet rejection vs $b-$tagging efficiency when the depth parameter $= 30$ for the pixel choice of 50×50 μm^2 and the fig (d) refers to the similar when depth $= 3$ and the overtraining is observed in the region of $1.0 < |\eta| < 2.7$. In both pixel choices, the depth $=3$ addresses overtraining in the afforementioned $|\eta|$ regions. "*trn*"in the legend with a solid line stands for training and "*tst*"with the dashed line is for testing. The ratio in the bottom plot is defined as $\frac{testing}{training}$.

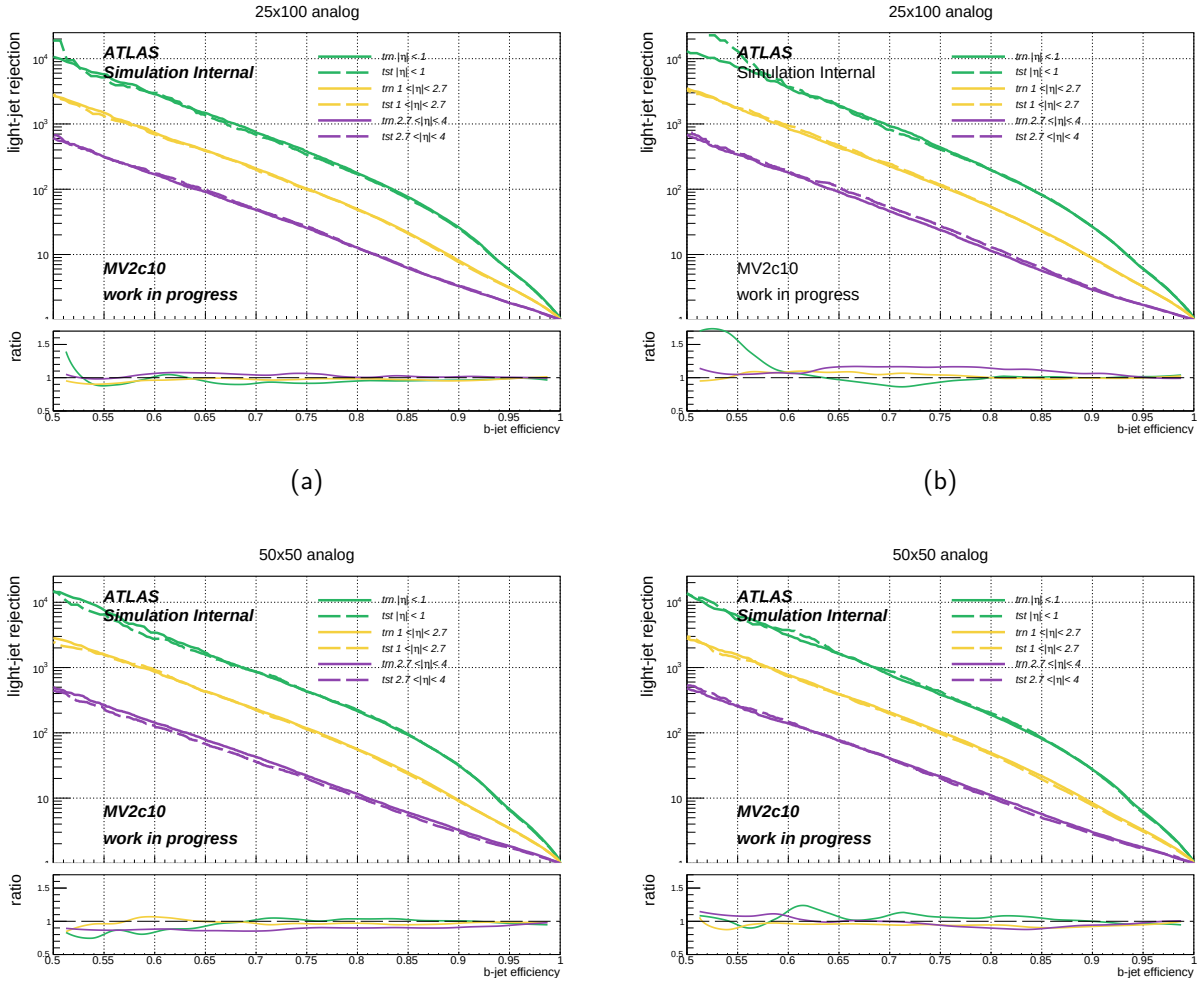


Figure 5.10: Top row corresponds to the pixel choice of 25×100 μm^2 and the fig. 5.10a is when the minimal node size parameter at 0.05% and fig. 5.10b is for minimal node size parameter at 0.10%. Bottom row corresponds to the pixel choice of 50×50 μm^2 and the fig. 5.10c is when the minimal node size parameter at 0.05% and fig. 5.10d is for minimal node size parameter at 0.10%. In all the instances, the depth parameter =12, number of trees parameter = 500 and cuts parameter = 200.

5.3.1 Results

While maintaining the parameters of Depth = 12, Number of trees = 500, cuts = 200 and varying the parameter of minimal node size between 0.05% and 0.10%, it was observed that the overtraining instances are minimal and the light-jet rejection is at maximum for both the pixel choices when minimal node size is at 0.10%. Tables 5.3 and 5.4 are summarizing the light-jet rejection values observed at three different $b-$tagging efficiencies for the pixel size choice studies. Highlighted cells are for the BDT parameter values which gives the best agreement between testing a training sample.

BDT parameter test values				Light-jet rejection		
Cuts	Depth	No.T	mns	b-jet eff 70%	b-jet eff 77%	b-jet eff 85%
200	12	500	0.05%	328.74	125.05	34.26
200	12	500	0.10%	405.79	151.17	38.4
200	30	1000	0.05%	465.24	217.87	72.76
200	03	1000	0.05%	344.45	132.93	36.32

Table 5.3: Optimization of light jet rejection at different BDT parameters for different b-tagging efficiency working points for pixel size pitch 25×100 μm^2.The BDT parameters are Cuts: number of pruning cuts, Depth: tree depth, No.T: number of trees and mns: minimal node size.

BDT parameter test values				Light-jet rejection		
Cuts	Depth	No.T	mns	b-jet eff 70%	b-jet eff 77%	b-jet eff 85%
200	12	500	0.05%	367.79	147.58	39.01
200	12	500	0.10%	348.39	133.12	34.61
200	30	1000	0.05%	454.38	199.85	68
200	03	1000	0.05%	295.63	125.63	32.59

Table 5.4: Optimization of light jet rejection at different BDT parameters for different b-tagging efficiency working points for pixel size pitch 50×50 μm^2.The BDT parameters are Cuts: number of pruning cuts, Depth: tree depth, No.T: number of trees and mns: minimal node size.

As depicted in fig 5.11, we have observed during the sample training studies that the light-jet rejection as a function of $b-$tagging efficiency (at 70%, 75%, and 85%), is greater in the pixel pitch of 25×100 μm^2 when compared with the pixel pitch of 50×50 μm^2 for all the three $|\eta|$ regions we considered. A similar and more detailed analysis given in Ref [32] reports that "The study of the performance obtained with a pixel pitch of 25×100 μm^2 has also been shown to improve the light jet rejection with respect to the 50×50 μm^2 configuration by 10 to 35% in most of the phase space."[8]

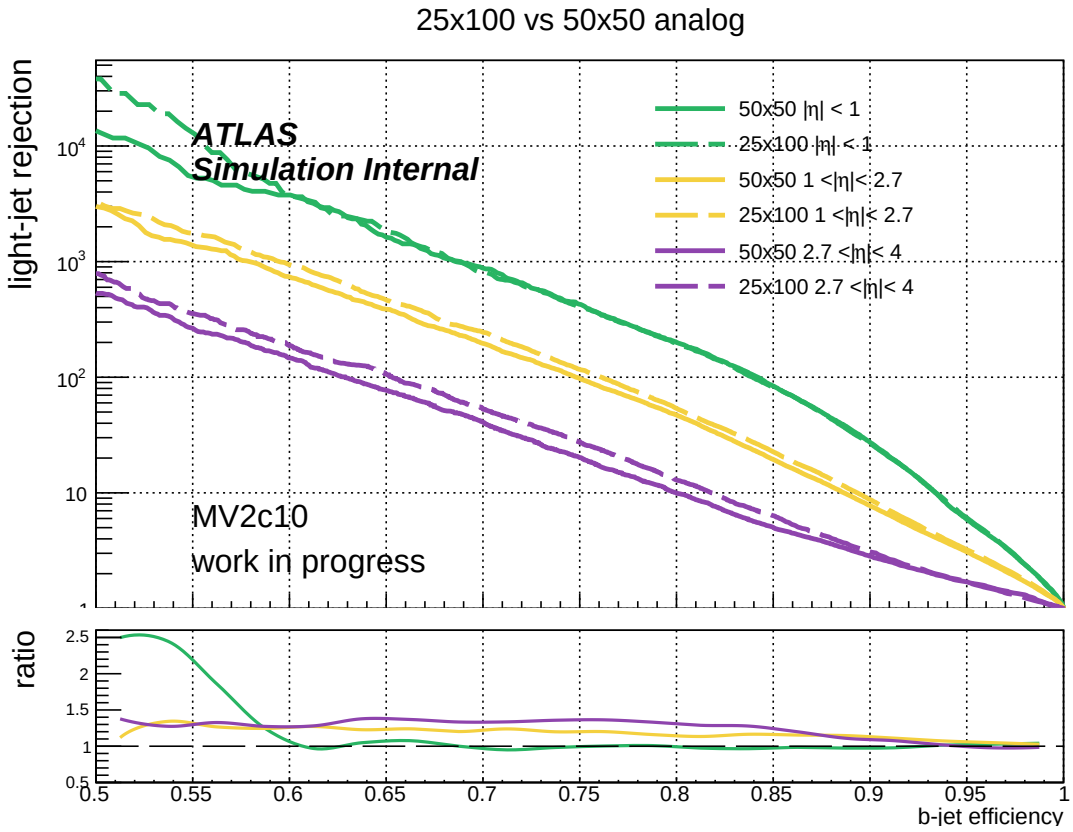


Figure 5.11: Comparison of light-jet rejection as a function of $b-$tagging efficiency for the pixel size choices. Solid line is for 50×50 μm^2 and dashed line is for 25×100 μm^2. The comparison given here are comparing the samples with highlighted BDT parameters in table 5.3 and table 5.4.

[8]The quoted lines are directly adopted from the ATLAS internal note: Expected $b-$tagging Performance with the upgraded ATLAS Inner Tracker Detector at the High-Luminosity LHC

5.4 Flavor tagging efficiency and uncertainty studies

We then studied the tagging efficiencies and uncertainties for the flavors of b-, c-, and light-jets using the following frameworks.

1. Upgrade Performance Functions (UPF): for tagging efficiency calculations

2. Calibration Data Interface (CDI): for uncertainty calculations

The UPF framework is providing parameterized estimations of the performance of the ATLAS detector at the HL-LHC era. The functions used in the UPF framework fully simulated and are applied to the truth level quantities. The flavor tagging functions included in this framework provides the efficiencies for c-jets and light-jets mistagged as b-jets for b-tagging efficiencies of 70% and 85% working points.

Since the UPF framework does not provide the associated flavor tagging uncertainties, the CDI framework was used to derive the uncertainties. It employs the "SFEigen"model, which is based on the eigenvector decomposition of the uncertainty covariance matrix which is corresponding to the kinematics (usually the jet p_T and η) bins provided in the calibration files. Each source of uncertainty (such as different flavors considered here) are varied by $\pm$ 1σ in each bin in the calibration file that is used for the calibration of flavor tagging algorithms. Uncertainties for the aforementioned flavors are derived here at 60%, 70%, 77%, and 85% $b-$tagging efficiency working points.

Having both these frameworks included in the "Run2BtagOptimizationFramework", we have used the layout of RUN2 and the pile-up interaction (μ) option of 200 in our studies. We have referred to the recommendations provided by the Upgrade Physics Activities group. [9]

[9] Please refer to ATLAS twiki: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HighLumiLhcSystematics2018#Experimental_systematics that provides recommendations for treating systematic uncertainties in the HL-LHC projections. Table 5.5 is adapted from this twiki.

Flavor Tagging Nuisance Parameter	Scale Factor for HL-LHC	Comments
FT_EFF_Eigen_B_[1-3]	1/3	inclusive in pt below 300 GeV and for all WP
FT_EFF_Eigen_C_[1-3]	1/3	valid for all WP
FT_EFF_Eigen_L_[1-5]	1/2 (tight and medium WP) 1/3 (loose WP)	quite conservative scaling factors depending on the WP

Table 5.5: Scale Factors adopted in deriving flavor tagging uncertainty derivations.

5.4.1 Results: Efficiencies

The UPF framework has derived efficiencies at following $|\eta|$ and p_T values,

$\|\eta\|$	jet p_T (GeV)
0.2	20
0.6	30
1	40
1.4	60
1.8	110
2.2	270
2.6	450
	750
	1500

Table 5.6: $|\eta|$ and p_T values which the efficiencies are derived at.

Thus, we have derived the efficiencies for flavors "b", "c", and "light"as two sets,

- Efficiency distribution for $|\eta|$ at different jet p_T values
- Efficiency distribution for jet p_T at different $|\eta|$ values

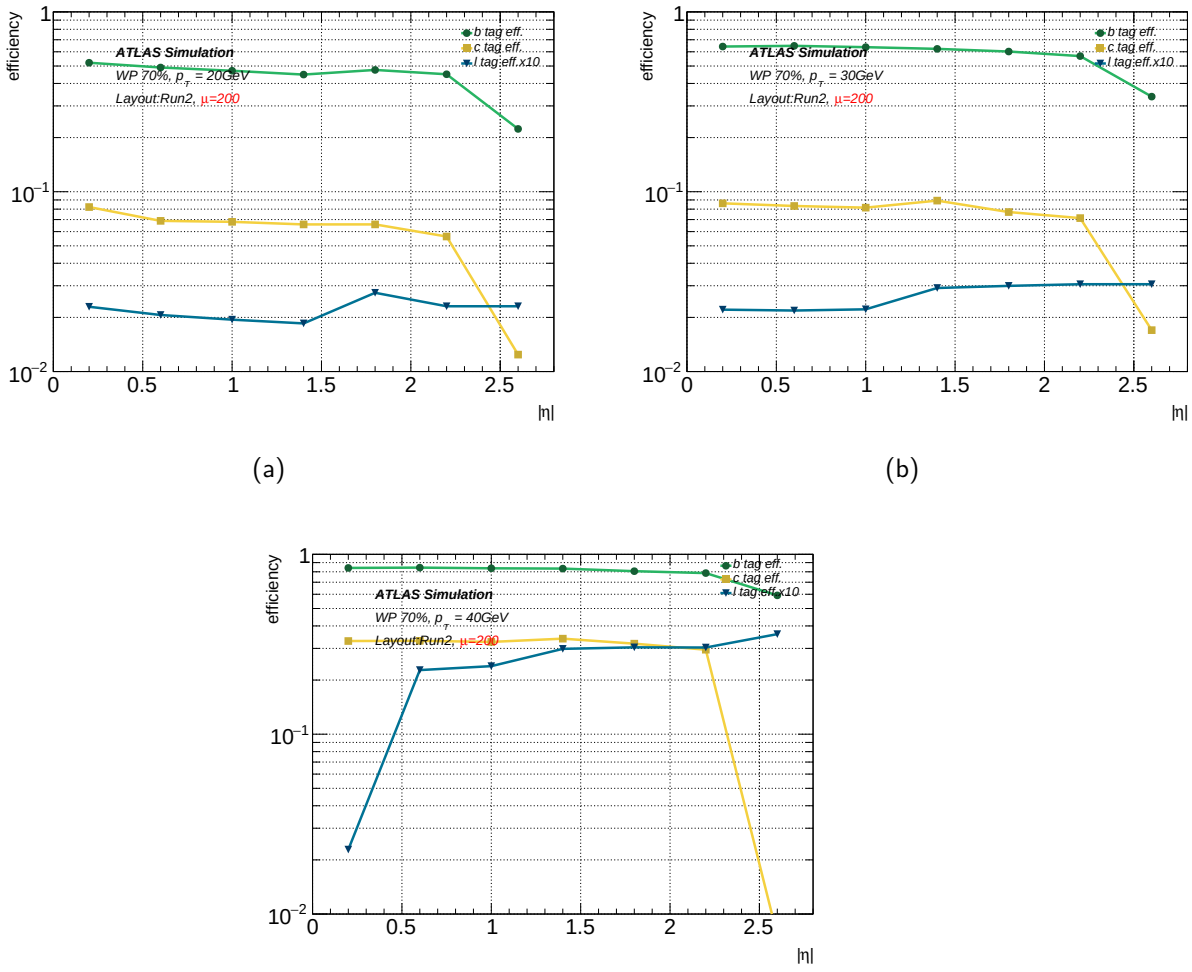


Figure 5.12: Efficiency distribution as a function of $|\eta|$ for jet p_T values of 20,30, and 40 GeV respectively. These efficiencies are derived at 70% $b-$tagging efficiency. $b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue. Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

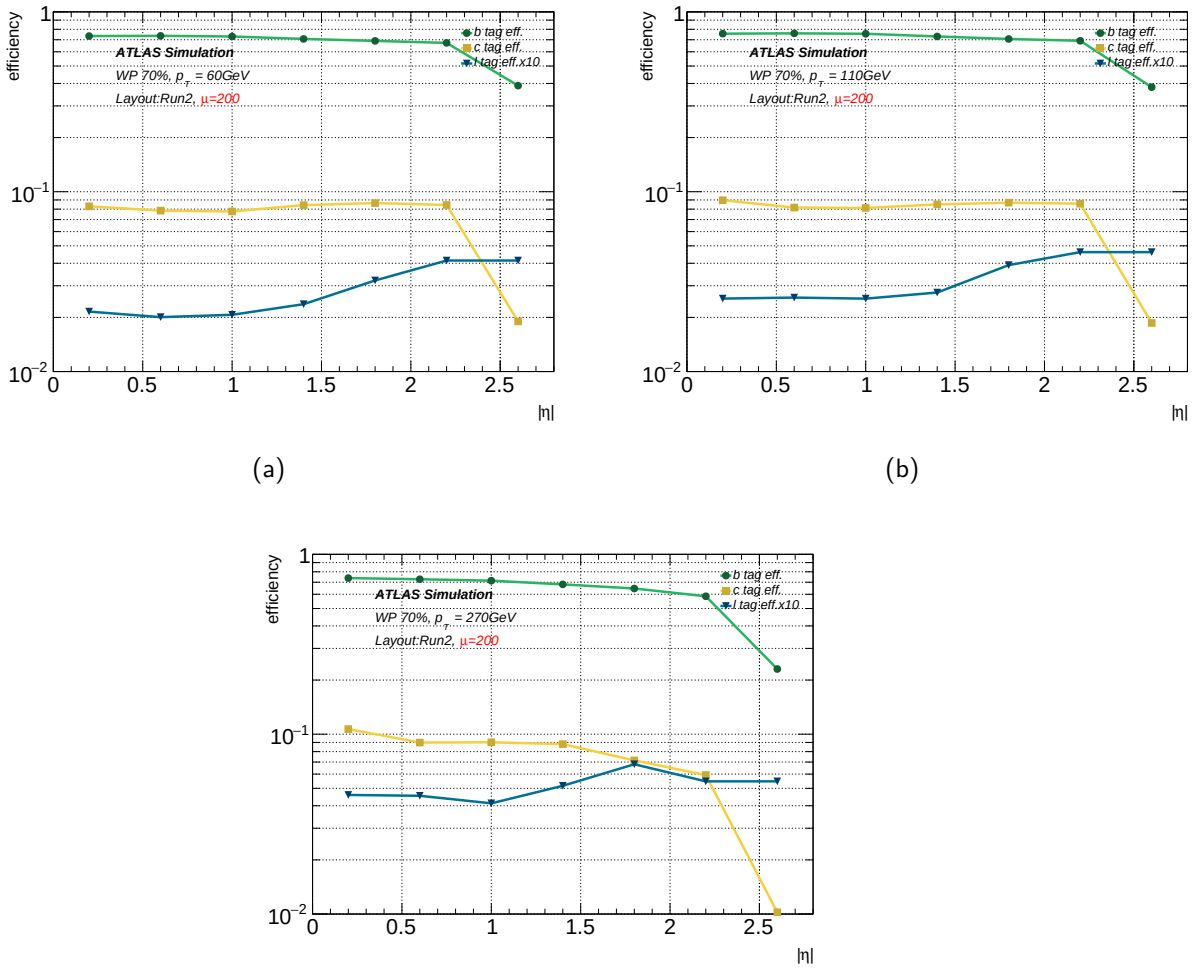


Figure 5.13: Efficiency distribution as a function of $|\eta|$ for jet p_T values of 60,110, and 270 GeV respectively. These efficiencies are derived at 70% $b-$tagging efficiency. $b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

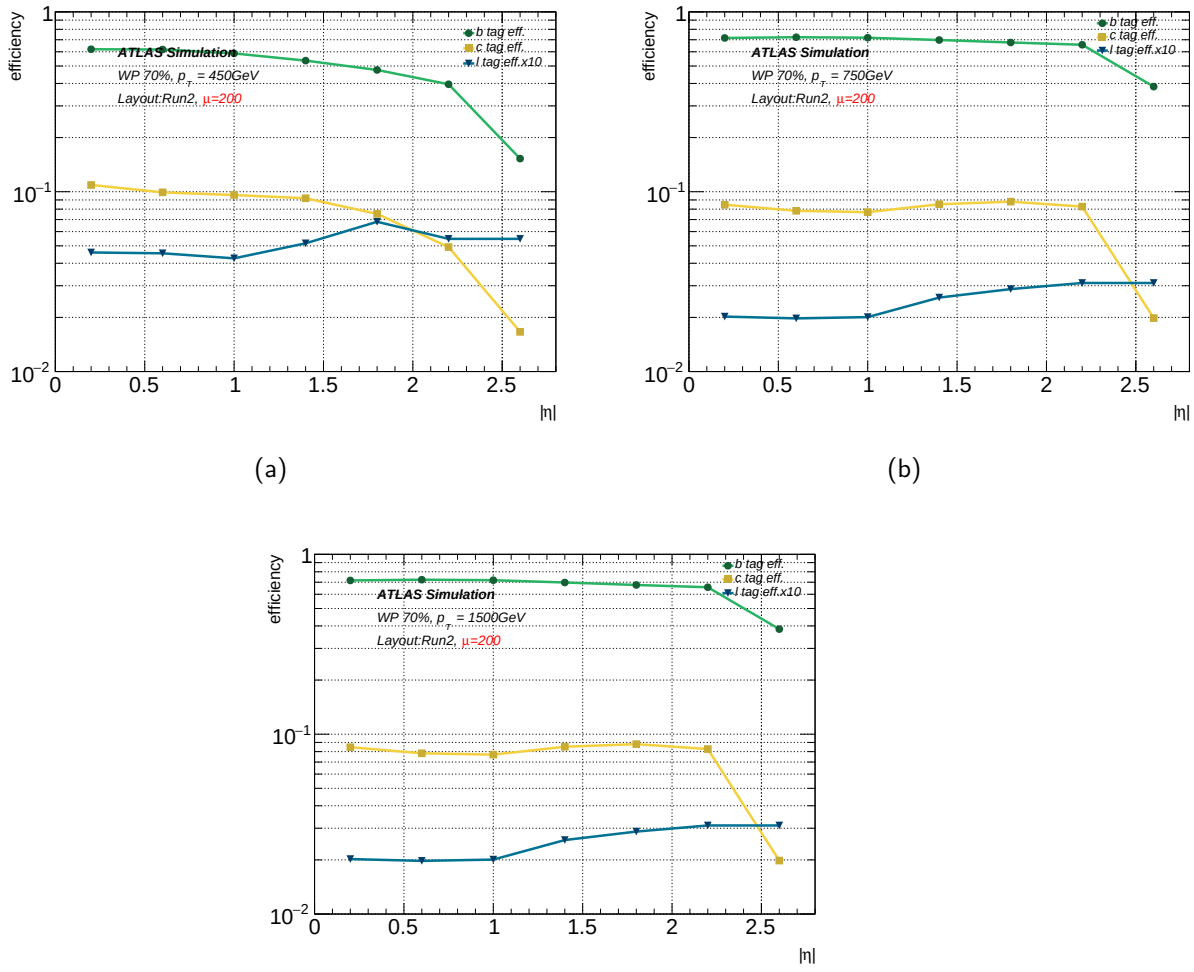


Figure 5.14: Efficiency distribution as a function of $|\eta|$ for jet p_T values of 450,750, and 1500 GeV respectively. These efficiencies are derived at 70% $b-$tagging efficiency. $b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

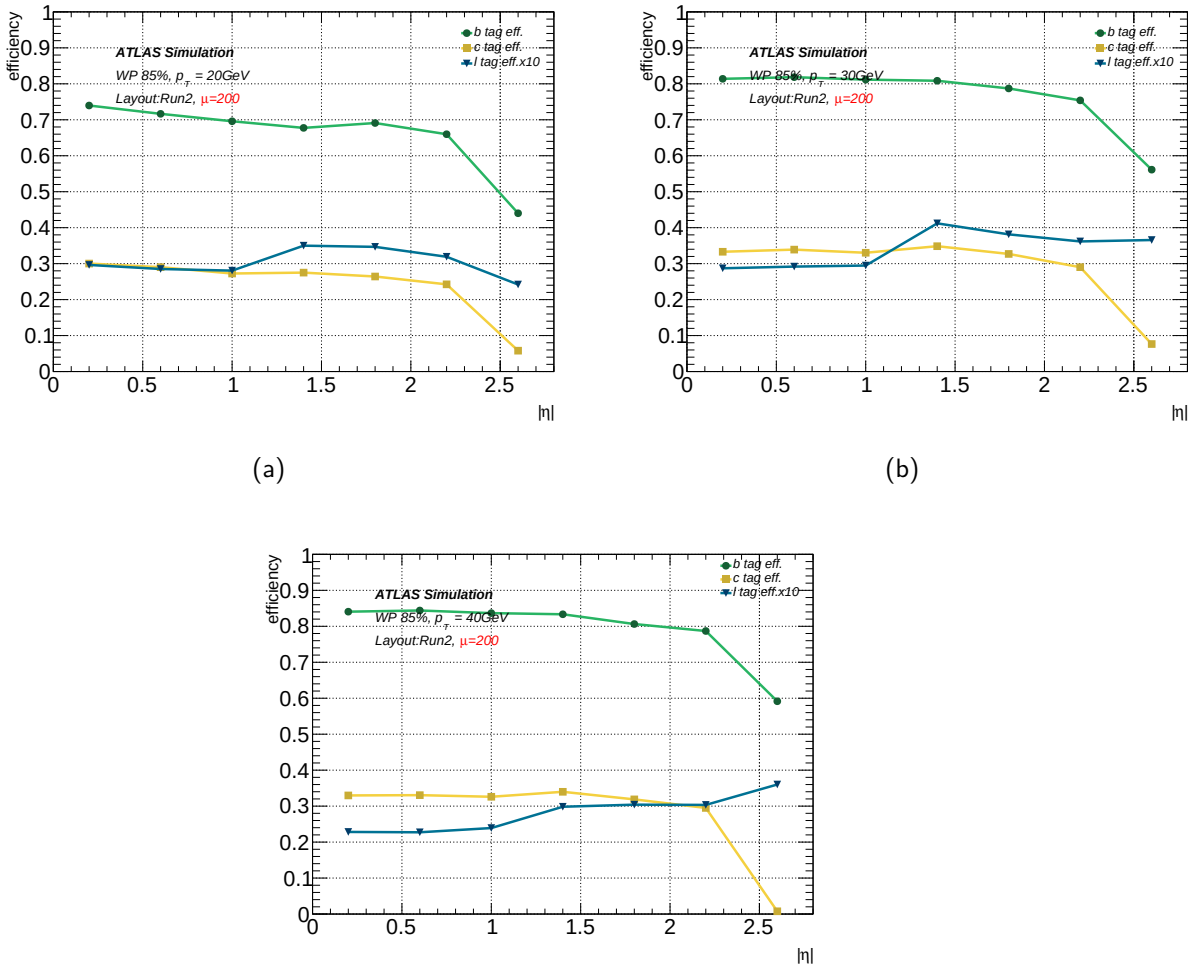


Figure 5.15: Efficiency distribution as a function of $|\eta|$ for jet p_T values of 20, 30, and 40 GeV respectively. These efficiencies are derived at 85% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

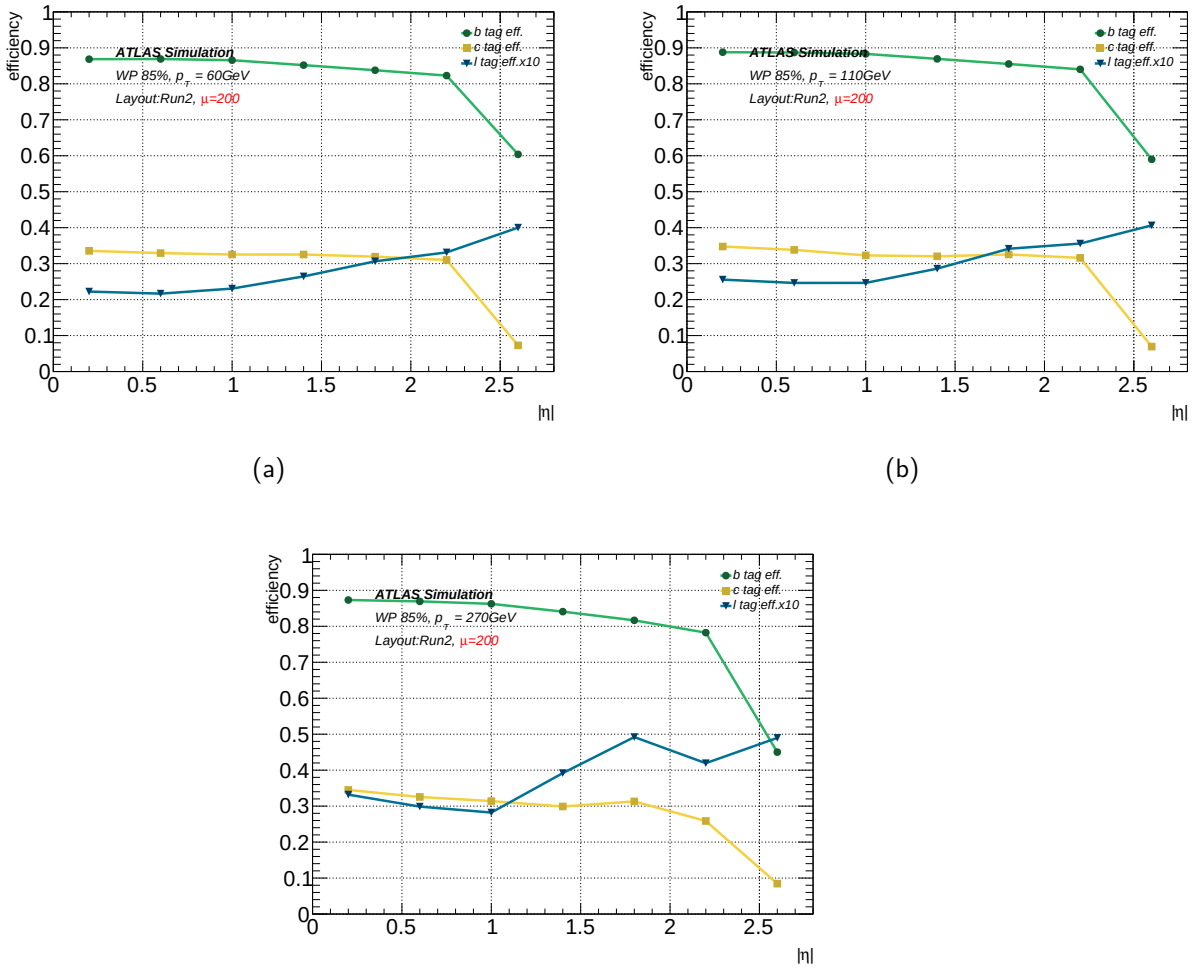


Figure 5.16: Efficiency distribution as a function of $|\eta|$ for jet p_T values of 60, 110, and 270 GeV respectively. These efficiencies are derived at 85% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

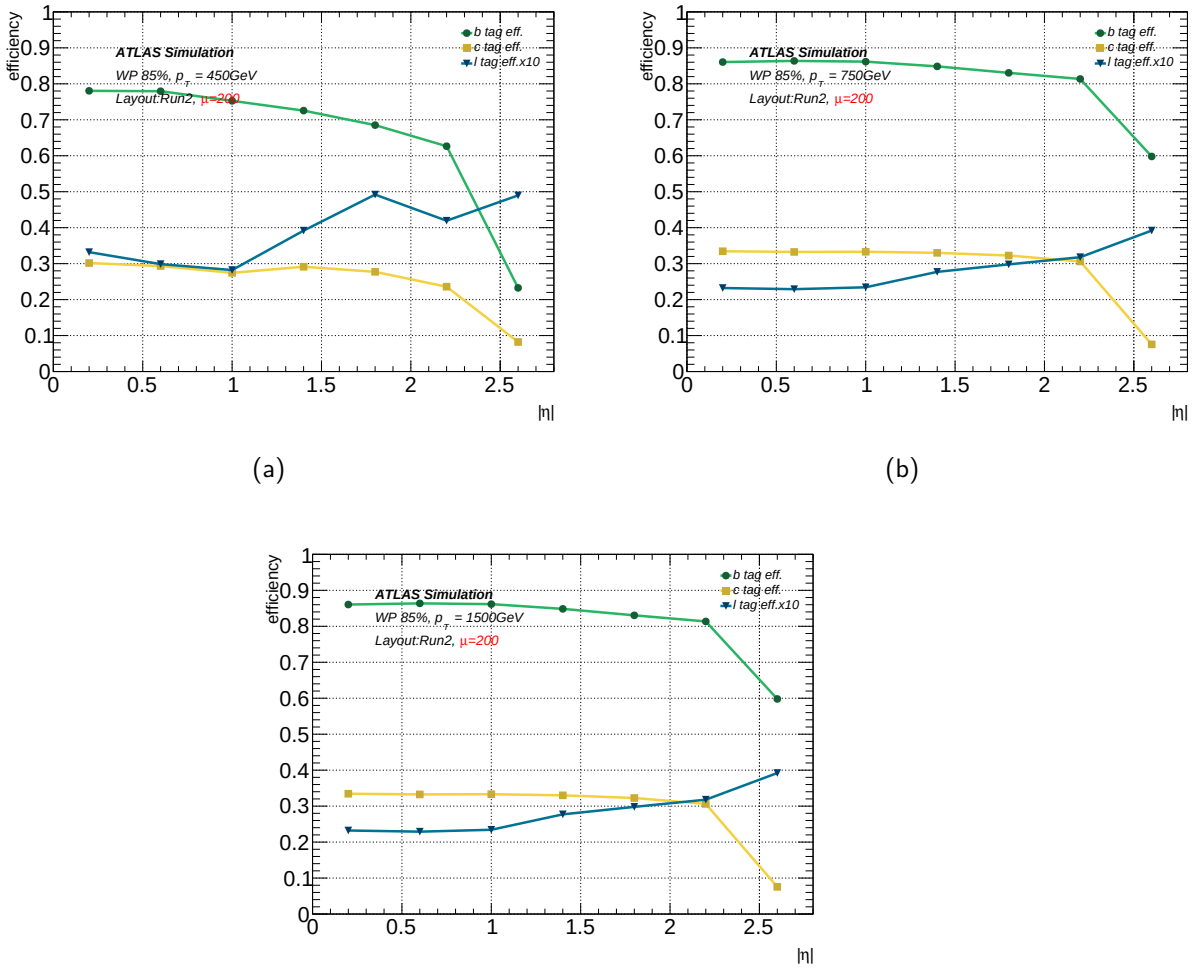


Figure 5.17: Efficiency distribution as a function of $|\eta|$ for jet p_T values of 450,750, and 1500 GeV respectively. These efficiencies are derived at 85% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

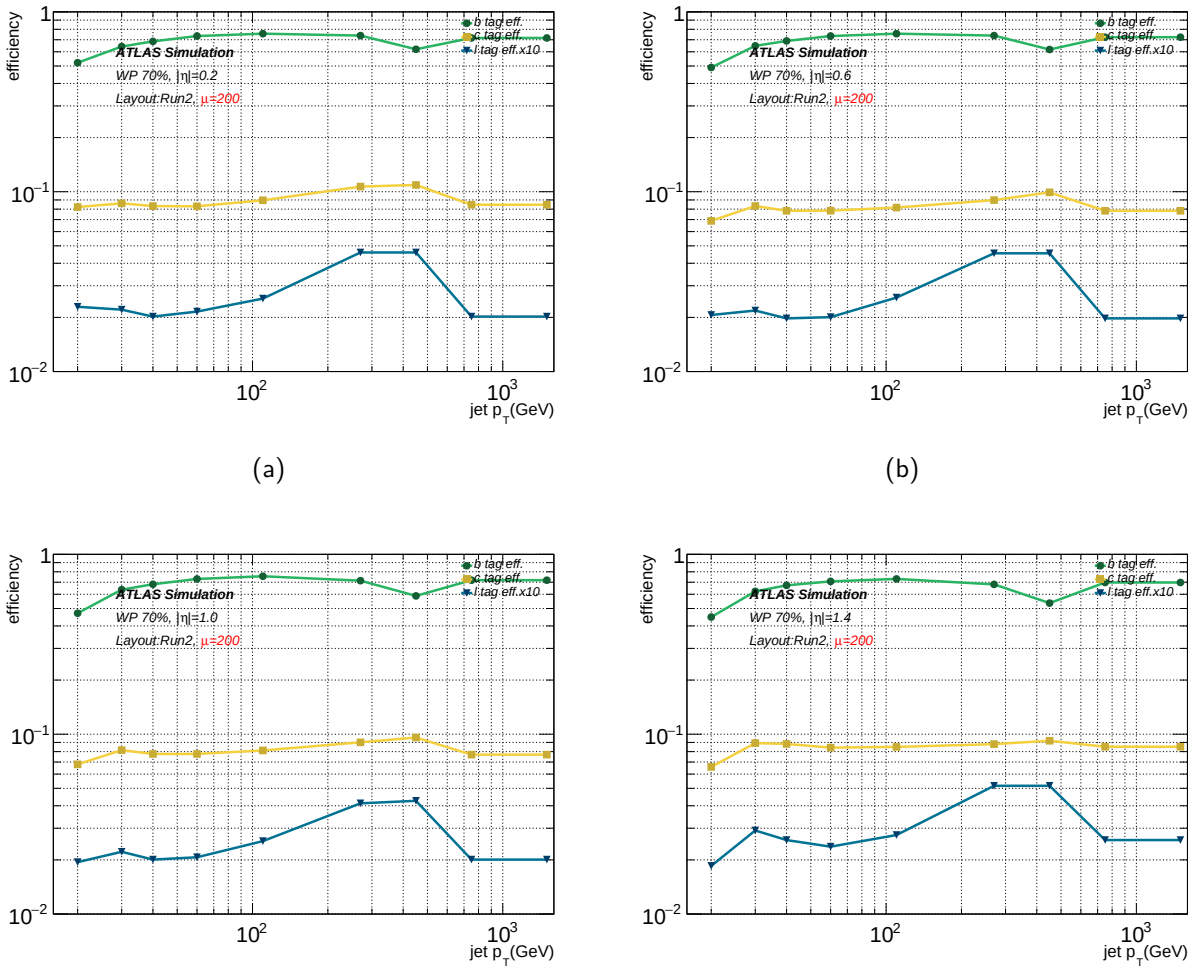


Figure 5.18: Efficiency distribution as a function of jet p_T for $|\eta|$ values of 0.2, 0.6,1.0, and 1.4 respectively. These efficiencies are derived at 70% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

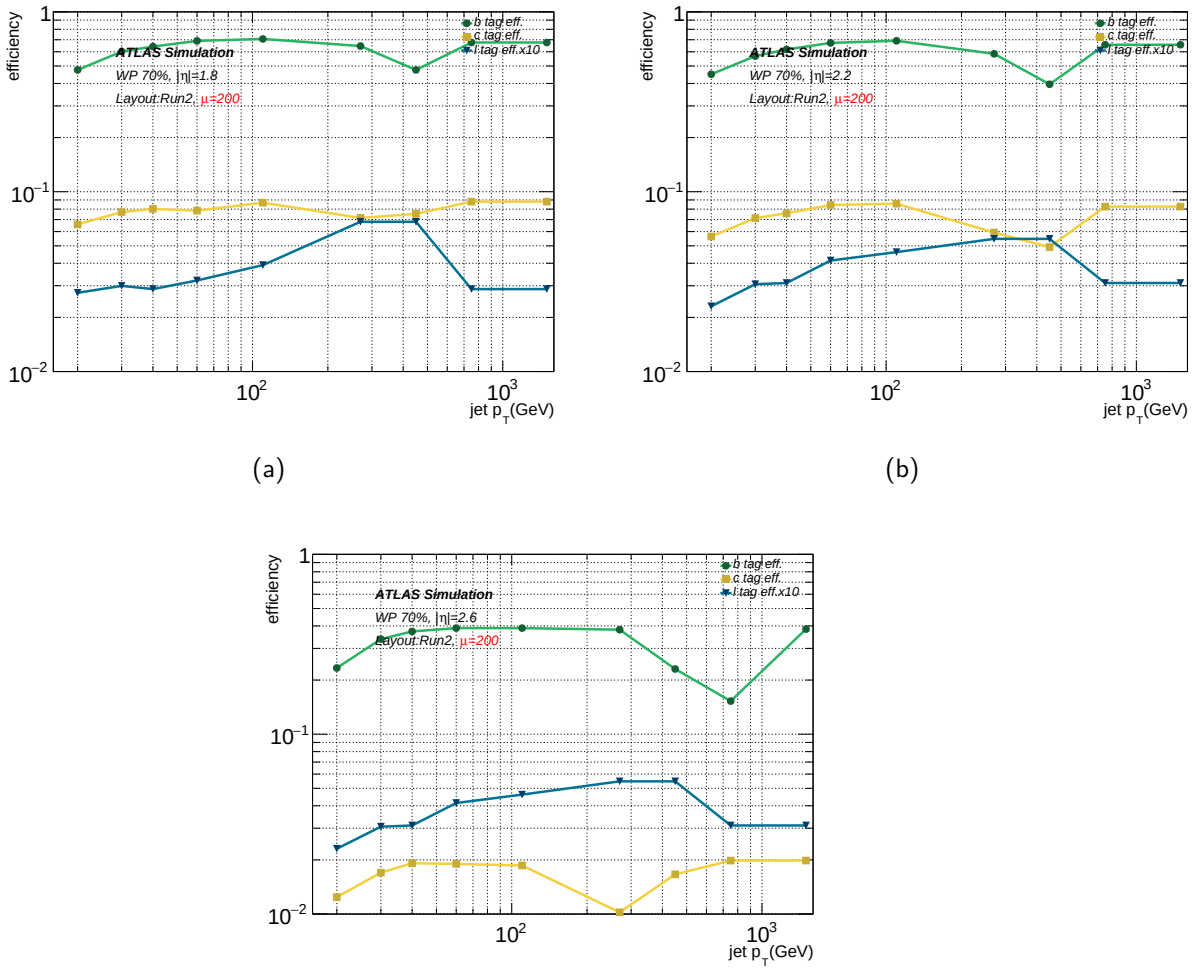


Figure 5.19: Efficiency distribution as a function of jet p_T for $|\eta|$ values of 1.8, 2.2, and 2.6 respectively. These efficiencies are derived at 70% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

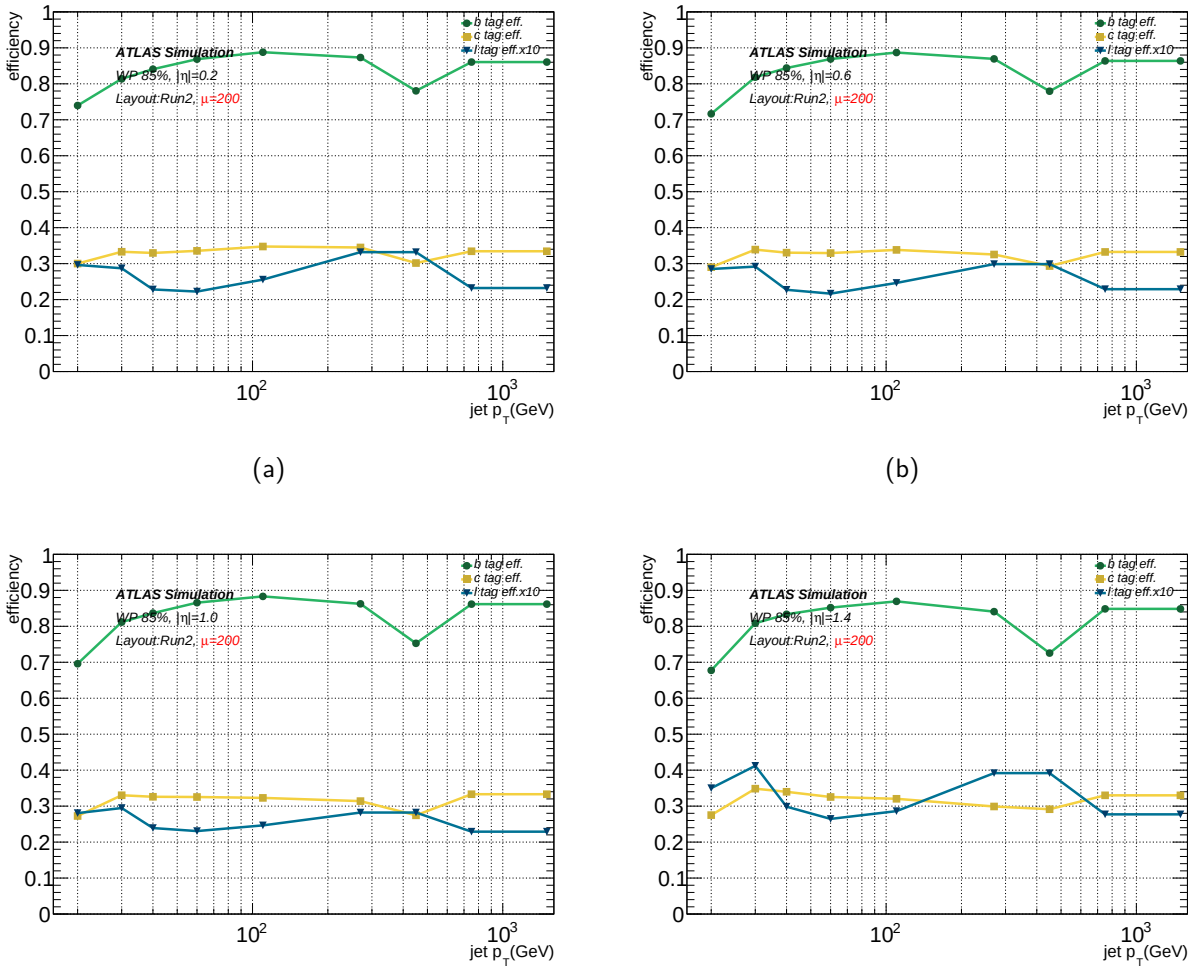
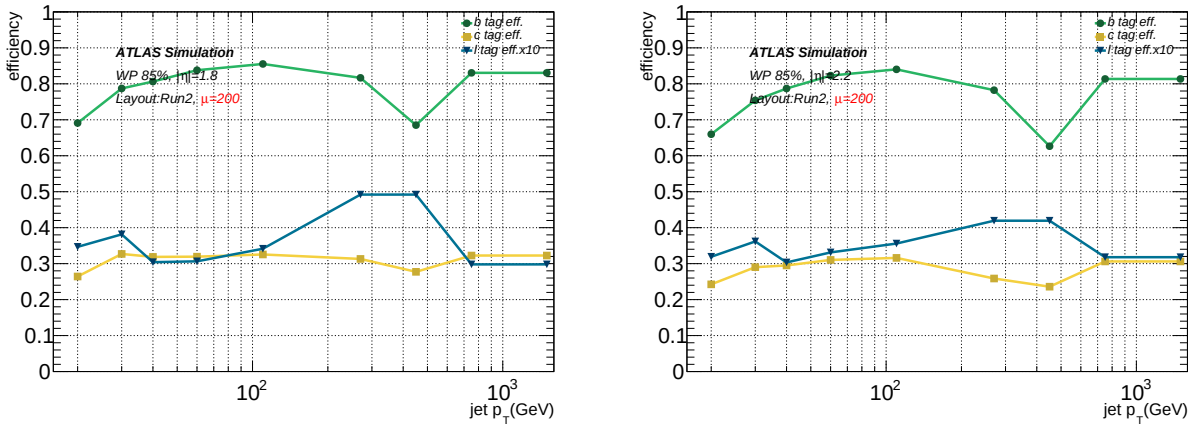
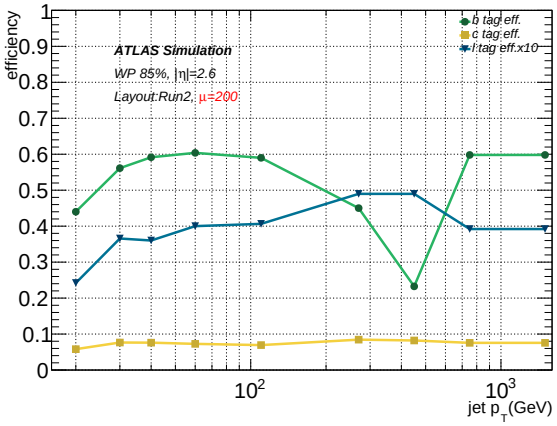


Figure 5.20: Efficiency distribution as a function of jet p_T for $|\eta|$ values values of 0.2, 0.6,1.0, and 1.4 respectively. These efficiencies are derived at 85% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.



(a) (b)



(c)

Figure 5.21: Efficiency distribution as a function of jet p_T for $|\eta|$ values values of 1.8, 2.2, and 2.6 respectively. These efficiencies are derived at 85% $b-$tagging efficiency.$b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.Light-jet tagging efficiency values multiplied by 10 to depict along with the other two jet efficiencies.

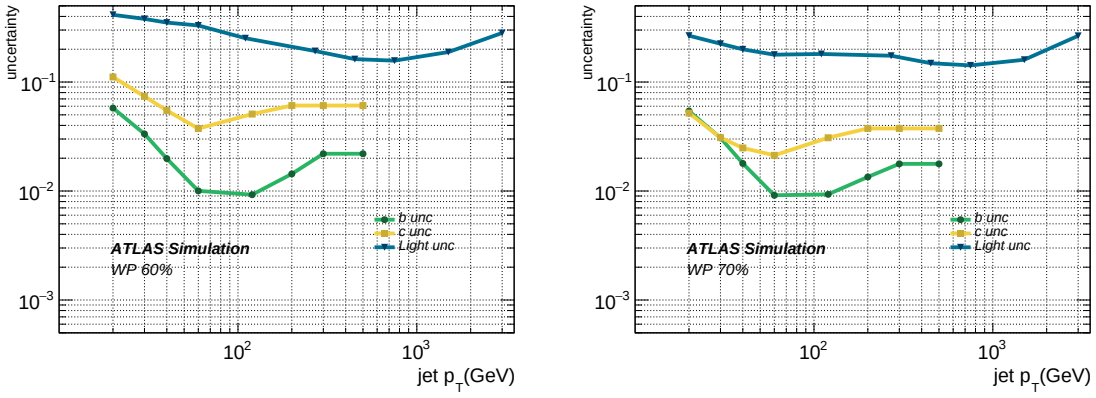
Higher efficiencies were observed for all the flavors at the 85% working point. For both the working points and at all the $|\eta|$ and the jet p_T distributions, we observed that the light jet efficiencies are lower than the other two flavors (Referring to figures from 5.12 to 5.16, the light jet efficiency values are multiplied by a factor of 10 to depict along with the other flavors). At $|\eta|$=2.6, b and c jet efficiencies record their lowest while light-jet efficiency is maximal. For both the working points, when the efficiencies are observed with respect to the jet p_T distribution, a sudden drop can be seen at a jet p_T value of 450 GeV for the b jet efficiency curves. The c jet efficiencies are given as a function of p_T remains almost constant except for $|\eta|$=2.6 at 70% working point.

5.4.2 Results: Uncertainties

The CDI framework has derived the flavor tagging uncertainties with adhering to the recommendations given at the table 5.5. The uncertainties for each flavor were derived at the following p_T values.

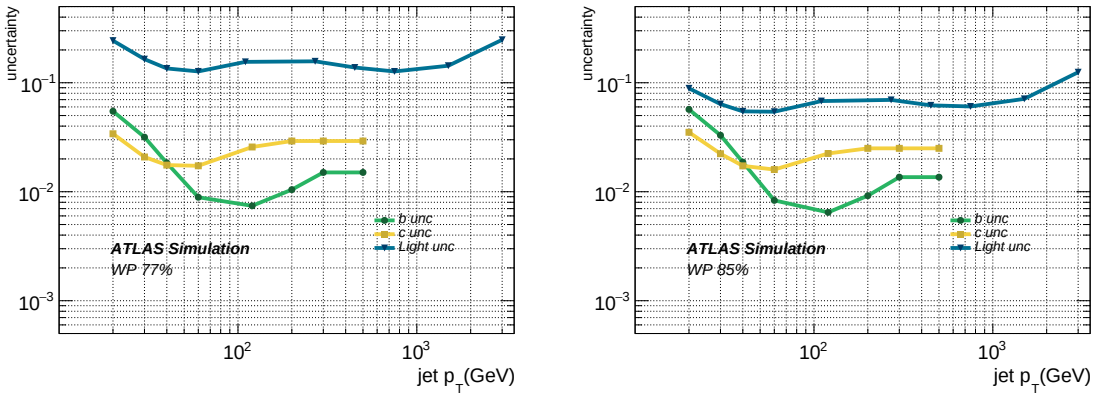
b and c flavors (GeV)	light flavor (GeV)
20	20
30	30
40	40
60	60
120	110
200	270
300	450
500	750
	1500
	3000

Table 5.7: p_T values of each flavor that the uncertainties are derived for.



(a) (b)

Figure 5.22: Uncertainty distribution as a function of jet p_T. Fig (a) is for 60% working point and fig (b) is for 70% working point. $b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.



(a) (b)

Figure 5.23: Uncertainty distribution as a function of jet p_T. Fig (a) is for 77% working point and fig (b) is for 85% working point. The $b-$tagging efficiency is given in green, c-tagging in yellow, and $light$-tagging in blue.

At all the working points and the jet p_T values, we observed that the uncertainties were higher for light jets when compared with the other two flavors. At higher working points, b-jet uncertainties at lower p_T values are higher than the c-jet uncertainties, but when the jet p_T increases, the c-jet uncertainties because higher than the b-jet uncertainties.

CHAPTER VI

Conclusions

This dissertation presents two independent research works. The first analysis searches for the single production of vector-like Top quark that decays into an all-hadronic final state (with no leptons) via $T \rightarrow Ht$ decay mode. Data collected by the ATLAS detector with a $\sqrt{s}$=13 TeV proton-proton collisions delivered by the LHC with an integrated luminosity of 139fb^{-1} during the period from 2015 to 2018 are used. Two back-to-back boosted jets originating from the decaying VLQ T are considered they are tagged as SM top, b or Higgs boson based on the final all-hadronic decay products. A 9×9 matrix is defined for the control and signal regions in the analysis where each region in the matrix contains unique events based on the tagging status of the leading and the sub-leading jets. The dominant multijet background was estimated with the ABCD method which used a correlation correction factor to study the regional correlation among the regions in the 9×9 region matrix. The analysis currently in the blinded stage thus, have derived the expected upper limits at 95% confidence level for three coupling constant values of 0.1, 0.5, and 1.0. It has been observed that the expected upper limit increases as the coupling constant increases.

The LHC plans to undergo an upgrade from 2024 to 2027 to increase the recording luminosity to become the HL-LHC by gathering an integrated luminosity of 4000fb^{-1} from the proton-proton collisions. The ATLAS detector, therefore, upgrades itself by introducing a new Inner Tracker to record events at a broader pseudorapidity and the second analysis given here studies firstly, defining the optimal BDT parameters for a combination of single-top, $t\bar{t}$, and Z$^{'}$ samples (these samples provide statistics with b-jets,

high pseudo-rapidity and high p_T respectively) using MV2c10 b-tagger, which provided the optimal light-jet rejection at three $b-$tagging efficiency working points of 70%, 77%, and 85%. During the optimization studies, observed over-training issues when using the default BDT parameters from the RUN 2 studies were addressed by reducing the tree depth parameter, and later all the BDT parameters were optimized to result in optimal light-jet rejection. Secondly, two pixel size pitch configurations of 25$\times$100 μm^2 and 50$\times$50 μm^2 were studied using configuration based $t\bar{t}$ samples. Both the samples were separately trained by the MV2c10 tagger to get the optimal light-jet rejection at 70%, 77% and 85% $b-$tagging efficiencies and observed that they get optimized under the same BDT parameters, yet 25$\times$100 μm^2 configuration has better light-jet rejection over 50$\times$50 μm^2 configuration. Finally, the tagging efficiencies (at $b-$tagging efficiency working points of 70% and 85%) and uncertainties (at $b-$tagging efficiency working points of 60%,70%,77%, and 85%) for the flavors of b-, c- and light-jets were derived which will be used in the HL-LHC era.

References

[1] Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. Technical Report ATLAS-CONF-2011-102, CERN, Geneva, Jul 2011.

[2] Early Inner Detector Tracking Performance in the 2015 data at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-051, CERN, Geneva, Dec 2015. .

[3] Expected performance of missing transverse momentum reconstruction for the ATLAS detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-023, CERN, Geneva, Jul 2015.

[4] Expected performance of the ATLAS b-tagging algorithms in Run-2. Technical Report ATL-PHYS-PUB-2015-022, CERN, Geneva, Jul 2015.

[5] Performance of missing transverse momentum reconstruction for the ATLAS detector in the first proton-proton collisions at at $\sqrt{s}$= 13 TeV. Technical Report ATL-PHYS-PUB-2015-027, CERN, Geneva, Jul 2015.

[6] Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. Technical Report ATL-PHYS-PUB-2016-012, CERN, Geneva, Jun 2016.

[7] Variable Radius, Exclusive-k$_T$, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS. Technical Report ATL-PHYS-PUB-2017-010, CERN, Geneva, Jun 2017.

[8] Morad et al. Measurements of $t\bar{t}$ differential cross-sections of highly boosted top quarks decaying to all-hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector. *Phys. Rev. D*, 98(1):012003, 2018.

[9] M. Aaboud et al. Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 96(7):072002, 2017.

[10] Morad Aaboud et al. Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC. *Eur. Phys. J. C*, 79(5):375, 2019.

[11] Aaboud, Morad and others. Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC. *Eur. Phys. J. C*, 77(5):332, 2017.

[12] Aaboud, Morad and others. In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector. *Eur. Phys. J. C*, 79(2):135, 2019.

[13] Georges Aad et al. Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data. *Eur. Phys. J. C*, 74(10):3071, 2014.

[14] Georges Aad et al. Search for high-mass dilepton resonances in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev. D*, 90(5):052005, 2014.

[15] Georges Aad et al. Search for new particles in events with one lepton and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 09:037, 2014.

[16] Georges Aad et al. Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector. *Phys. Rev. D*, 91(1):012006, 2015.

[17] Georges Aad et al. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV. *Eur. Phys. J. C*, 76(5):292, 2016.

[18] Georges Aad et al. Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector. *Eur. Phys. J. C*, 76(11):581, 2016.

[19] Georges Aad et al. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019.

[20] Aad, G. and others. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010.

[21] Agostinelli, S. and others. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.

[22] J. Aguilar-Saavedra. Protos: Program for top simulations, tech. rep. http://jaguilar.web.cern.ch/jaguilar/protos. .

[23] J.A. Aguilar-Saavedra. Identifying top partners at LHC. *JHEP*, 11:030, 2009.

[24] Aguilar-Saavedra, J.A. and Benbrik, R. and Heinemeyer, S. and Pérez-Victoria, M. Handbook of vectorlike quarks: Mixing and single production. *Phys. Rev. D*, 88(9):094010, 2013.

[25] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.

[26] Bo Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys. Rept.*, 97:31–145, 1983.

[27] Nima Arkani-Hamed, Andrew G Cohen, Emanuel Katz, and Ann E Nelson. The littlest higgs. *Journal of High Energy Physics*, 2002(07):034–034, jul 2002.

[28] Moritz Backes. The ATLAS Trigger System : Ready for Run-2. Technical Report ATL-DAQ-PROC-2015-041, CERN, Geneva, Oct 2015.

[29] Julia Bauer and Thomas Muller. Prospects for the observation of electroweak top quark production with the cms experiment. 04 2020.

[30] Peter Berta, Kyle James Read Cormier, Ye Chen, Riccardo Di Sipio, Petr Jacka, Marino Romano, Yuji Yamazaki, Pekka Sinervo, Roman Lysak, Shima Shimizu Noda, Jiri Kvita, Jan Palicka, Rupert Leitner, Federica Fabbri, Francesco Spano, Cassandra Lindsey Miller, and Robin Leigh Hayes. Measurements of $t\bar{t}$ differential cross-sections in the all-hadronic channel using highly boosted top quarks in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector: Analysis of the full 2015 and 2016 dataset. Technical Report ATL-COM-PHYS-2016-1696, CERN, Geneva, Nov 2016.

[31] Clement Charles Bordereau, Nihal Brahimi, Alessandro Calandri, Arnaud Duperrin, Laurent Vacavant, Stefan Guindon, Alexander Khanov, Madhuranga Madugoda Ralalage Don, and Thomas Strebler. Optimization of the b-tagging algorithms for the Phase 2 ITk Pixel TDR. Technical Report ATL-COM-UPGRADE-2018-010, CERN, Geneva, May 2018.

[32] Bordereau, Clement Charles and Brahimi, Nihal and Calandri, Alessandro and Duperrin, Arnaud and Guindon, Stefan and Khanov, Alexander and Madugoda Ralalage Don, Madhuranga and Strebler, Thomas and Vacavant, Laurent and Jezequel, Stephane. Expected b-tagging performance with the upgraded ATLAS Inner Tracker detector at the High-Luminosity LHC. Technical Report ATL-COM-PHYS-2019-1163, CERN, Geneva, Sep 2019. Aim is to make this an FTAG CP pub-note.

[33] Boutle, S. and others. Primary vertex reconstruction at the ATLAS experiment. *J. Phys. Conf. Ser.*, 898(4):042056, 2017.

[34] Buckley, Andy and others. General-purpose event generators for LHC physics. *Phys. Rept.*, 504:145–233, 2011.

[35] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[36] Alessandro Calandri. Flavour tagging algorithms and performance at the ATLAS experiment. Technical Report ATL-PHYS-PROC-2016-105, CERN, Geneva, Aug 2016.

[37] Alessandro Calandri. Flavour tagging algorithms and performance at the ATLAS experiment. *PoS*, LHCP2016:180, 2016.

[38] Alessandro Calandri, Marco Battaglia, Francesco Armando Di Bello, Yann Coadou, Andrea Coccaro, Markus Cristinziani, Geoffrey Gilles, Tobias Golling, Daniel Guest, Richard John Hawkings, Zihao Jiang, Michael Kagan, Vadim Kostyukhin, Marie Christine Lanfermann, Emmanuel Le Guirriec, Michela Paganini, Christopher Samuel Pollard, Andrea Sciandra, Jonathan Shlomi, Lauren Tompkins, and Michael Ughetto. Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run. Technical Report ATL-COM-PHYS-2017-505, CERN, Geneva, May 2017.

[39] Alessandro Calandri, Geoffrey Gilles, Remi Zaidan, Andrea Coccaro, Valerio Dao, Marco Battaglia, Antonio Miucci, Francesco Armando Di Bello, Eloi Le Quilleuc, Yulia Rodina, Jonathan Shlomi, Rima El Kosseifi, Richard John Hawkings, Giacinto Piacquadio, Laurent Vacavant, Markus Cristinziani, Vadim Kostyukhin, Mazuza Ghneimat, Yann Coadou, and Kazuya Mochizuki. Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run. Technical Report ATL-COM-PHYS-2016-374, CERN, Geneva, Apr 2016.

[40] J M Campbell, J W Huston, and W J Stirling. Hard interactions of quarks and gluons: a primer for LHC physics. *Reports on Progress in Physics*, 70(1):89–193, dec 2006.

[41] M. Capeans, G. Darbo, K. Einsweiler, M. Elsing, T. Flick, M. Garcia-Sciveres, C. Gemme, H. Pernegger, O. Rohne, and R. Vuillermet. ATLAS Insertable B-Layer Technical Design Report. 9 2010.

[42] Hai-Yang Cheng. Phenomenological Study of Heavy Hadron Lifetimes. *JHEP*, 11:014, 2018.

[43] Yann Coadou. Boosted Decision Trees and Applications. *EPJ Web Conf.*, 55:02004, 2013.

[44] Andrea Coccaro. Track Reconstruction and b-Jet Identification for the ATLAS Trigger System. Technical Report arXiv:1112.0180. ATL-DAQ-PROC-2011-051, CERN, Geneva, Dec 2011. Comments: 7 pages, 10 figures, conference proceedings for ACAT 2011.

[45] Collins, John C. and Soper, Davison E. and Sterman, George F. *Factorization of Hard Processes in QCD*, volume 5, pages 1–91. 1989.

[46] Ian Connelly. Performance and calibration of b-tagging with the ATLAS experiment at LHC Run-2. *EPJ Web Conf.*, 164:07025, 2017.

[47] Cornelissen, T. and Elsing, M. and Gavrilenko, I. and Liebig, W. and Moyse, E. and Salzburger, A. The new ATLAS track reconstruction (NEWT). *J. Phys. Conf. Ser.*, 119:032014, 2008.

[48] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

[49] Kyle Cranmer, George Lewis, Lorenzo Moneta, Akira Shibata, and Wouter Verkerke. HistFactory: A tool for creating statistical models for. Technical report.

[50] Czakon, Michal and Mitov, Alexander. Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders. *Comput. Phys. Commun.*, 185:2930, 2014.

[51] Celine Degrande, Claude Duhr, Benjamin Fuks, David Grellscheid, Olivier Mattelaer, and Thomas Reiter. UFO - The Universal FeynRules Output. *Comput. Phys. Commun.*, 183:1201–1214, 2012.

[52] Markus Elsing, Danilo Giugni, Noemi Calace, Nora Emilia Pettersson, Andreas Salzburger, Izaac Gregory Sanderswood, Ben Smart, Laurent Vacavant, Francesco Costanza, Nathan Peter Readioff, Sharka Todorova-Nova, Alexander Khanov, Thomas Strebler, Helen Hayward, Nigel Hessey, Joel Hengwei Foo, Yee Chinn Yap, Jike Wang, Peter Sutcliffe, Neal Hartman, Steven Welch, Sebastien Michal, Matthias Hamer, Joleen Pater, Alexandre Bitadze, Martin Janda, Vit Zahradnik, Mauro Citterio, Simone Pagan Griso, Aleksandra Dimitrievska, Cristiano Alpigiani, Andrzej Smykiewicz, Atanu Pathak, Igor Gavrilenko, Diptaparna Biswas, Hector De la Torre, Nguyen Phuong Dang, Marcin Wladyslaw Wolter, Guennadi Borissov, Graham Richard Lee, Valentina Maria Cairo, Swagato Banerjee, Nicholas Adam Styles, Matthias Danninger, Marianna Testa, Diego Alvarez Feito, Thanos Manousos, Mika Huhtinen, Kenneth Wraight, Craig Buttar, Tony Affolder, Ian Wilmut, Charles Peter Evans, Arnaud Duperrin, Chen Zhou, Ke Li, Liza Mijovic, and Maximilian Goblirsch-Kolb. Final Report of the second ITk Pixel Layout Task Force. Technical Report ATL-COM-ITK-2018-053, CERN, Geneva, Dec 2018. Hi Claudia, I finally added the 3 plots you wanted to appendix 3, such that this can become an INT note now. M.

[53] Tobias Flick and ATLAS Collaboration. The Phase II ATLAS Pixel Upgrade: The Inner Tracker (ITk). Technical Report ATL-ITK-PROC-2016-006, CERN, Geneva, Nov 2016.

[54] Stefan Gieseke, Patrick Kirchgaeßer, and Simon Plätzer. Baryon production from cluster hadronisation. *Eur. Phys. J. C*, 78(2):99, 2018.

[55] Andreas Hocker et al. TMVA - Toolkit for Multivariate Data Analysis. 3 2007.

[56] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.

[57] Kinoshita, Toichiro and Ukawa, Akira. Mass singularities of feynman amplitudes. In Huzihiro Araki, editor, *International Symposium on Mathematical Problems in Theoretical Physics*, pages 55–58, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[58] David Krohn, Jesse Thaler, and Lian-Tao Wang. Jet Trimming. *JHEP*, 02:084, 2010.

[59] A. La Rosa. ATLAS IBL Pixel Upgrade. *Nucl. Phys. B Proc. Suppl.*, 215:147–150, 2011.

[60] A.B. Lahanas. LSP as a Candidate for Dark Matter. *Lect. Notes Phys.*, 720:35–68, 2007.

[61] Marie Lanfermann. Deep Learning in Flavour Tagging at the ATLAS experiment. *PoS*, EPS-HEP2017:764, 2018.

[62] J. Lange et al. 3D silicon pixel detectors for the High-Luminosity LHC. *JINST*, 11(11):C11024, 2016.

[63] Luca Lista. *Statistical Methods for Data Analysis in Particle Physics*, volume 909. Springer, 2016.

[64] A.Ruiz Martínez. The Run-2 ATLAS Trigger System. *J. Phys. Conf. Ser.*, 762(1):012003, 2016.

[65] Joao Pequenao and Paul Schaffner. How ATLAS detects particles: diagram of particle paths in the detector. Jan 2013.

[66] Giacinto Piacquadio and Christian Weiser. A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS. *J. Phys. Conf. Ser.*, 119:032032, 2008.

[67] Michele Pinamonti, Loic Valery, Alexander Held, and Tomas Dado. Trexfitter.

[68] P. Puzo. ATLAS calorimetry. *Nucl. Instrum. Meth. A*, 494:340–345, 2002.

[69] Aaron Roodman. Blind analysis in particle physics. *eConf*, C030908:TUIT001, 2003.

[70] Raghunath Sahoo. Relativistic Kinematics. 4 2016.

[71] Schmaltz, Martin and Tucker-Smith, David. Little Higgs review. *Ann. Rev. Nucl. Part. Sci.*, 55:229–270, 2005.

[72] The ATLAS collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. 6 2016.

[73] Thomson, Mark. *Modern particle physics*. Cambridge University Press, New York, 2013.

[74] Wouter Verkerke and David Kirkby. The RooFit toolkit for data modeling, 2003.

[75] B.R. Webber. Hadronization. In *Summer School on Hadronic Aspects of Collider Physics*, pages 49–77, 11 1994.

eccntformat[1].

APPENDIX A

Symmetry groups

A.1 Symmetry groups

Symmetry groups are vital in the particle physics theories. If a given theory is invariant under the transformation by any symmetry group, it will result in the corresponding quantum numbers and conservation laws.

A.2 Group

To be considered as a group, any group G needs to have elements a, inverse elements a^{-1}, a unit element 1, and also, the satisfaction of following multiplication rules.

- if a,$b \ni$ G, then $c = a \cdot b \ni$ G
- $a \cdot (\mathrm{b} \cdot \mathrm{c}) = (a \cdot b) \cdot c$
- $a \cdot 1 = 1 \cdot a = a$
- $a \cdot a^{-1} = a^{-1} \cdot a = 1$
- A group is called "Abelian" if $(a \cdot b) = (b \cdot a)$

A.3 The product of groups

The product of any two given groups A and B is attained by considering the multiplication of the elements by pairs.

$$(a_1 \cdot b_1) \cdot (a_2 \cdot b_2) = ((a_1 \cdot a_2) \cdot (b_1 \cdot b_2)) \tag{A.1}$$

Groups can be represented by matrices (usually as n $\times$ n) and the followings are some examples for commonly used matrices in the group theory.

- Hermitian $\rightarrow M^{\dagger} = M$
- Unitary $\rightarrow U^{\dagger} = U^{-1}$
- Special $\rightarrow det \mid S \mid = 1$

A.4 Unitary group U(1)

The group that represents the unitary transformations acting on complex one dimensional (1-D) vectors is known as U(1). Its generator is the unit operator I and it transforms the phase of a given wave function. The following transformations are acting on the 1-D vector space identified with the electric charge.

$$U = \exp(i\alpha Q) \tag{A.2}$$

where, Q is the charge operator. There are two ways this transformation can be used,

- **Global**: α is a constant (no dependence on space)
- **Local**: α is depending on spatial position $\alpha(x)$

The operation of the charge operator Q on a wave function of a particle yields the charge of the particle,

$$Q\Psi = q\Psi \tag{A.3}$$

Thus, when the wave function Ψ transforms globally under an U(1) transformation,

$$\Psi \rightarrow \Psi^{'} = \exp(i\alpha Q)\Psi \tag{A.4}$$

gives the following for a single particle.

$$\Psi^{'} = \exp(i\alpha q)\Psi \tag{A.5}$$

But, the transformation under the local U(1) transformation is more important in particle physics. The transformation group for this instance is in the form of,

$$U = \exp(i\alpha(x)Q) \tag{A.6}$$

This is an example for a "gauge transformation"[1] and one can consider the application of this transformation on a Lagrangian (L) of a fermion with charge q as follows.

$$L \sim \bar{\Psi}\gamma_{\mu}\partial^{\mu}\Psi \tag{A.7}$$

and under the local transformation, Lagrangian will become,

$$L \rightarrow L^{'} \sim L + \bar{\Psi}(iq\gamma_{\mu}\partial^{\mu}\alpha)\Psi \neq L \tag{A.8}$$

for the invariance of the Lagrangian under the transformation, α has to vanish. Thus, we are forced to replace ∂^{μ} by the covariant derivative,

$$D^{\mu} = \partial^{\mu} + iQA^{\mu} \tag{A.9}$$

[1] Term "gauge" stands for size

where A^μ is a scalar field that ensures the transformation invariance. Thus, A^μ has to undergo the following transformation.

$$A^\mu \rightarrow A^\mu - \frac{1}{Q}\partial^\mu \alpha \tag{A.10}$$

A^μ is now identified as the "photon" which is the mediating gauge boson of the electromagnetic field and the corresponding fermion is the electron. Thus, the local U(1) symmetry associated with the electric charge leads us to realize the existence of the photon and eventually the electromagnetic interactions.

A.5 SU(2) group

The special unitary (SU) group transformation acting on 2-D complex vectors is known as the SU(2) group. The generators for the SU(2) group are the Pauli matrices (σ_1,σ_2, and σ_3). The SU(2) group is used to describe the weak isospin (an exact symmetry of the SM), which gives rise to the mediating gauge bosons of the weak interactions in the SM. SU(2) group provides 3 vector fields (bosons) W_1, W_2, and W_3 ($\vec{W}$). The weak interaction interacts with fermions (quarks and leptons) and is capable of changing their flavor upon interaction. We are now aware that both left-handed, quark and lepton generations are coming as doublets (two) and an example of the interaction of these doublets with the vector field $\vec{W}$ is as follows.

$$\begin{pmatrix}\bar{u} & \bar{d}\end{pmatrix}\left(g\vec{W}\cdot\frac{\vec{\sigma}}{2}\right)\begin{pmatrix}u\\ d\end{pmatrix} = \begin{pmatrix}\bar{u} & \bar{d}\end{pmatrix}\frac{g}{2}\begin{pmatrix}W_3 & W_1 - iW_2\\ W_1 + iW_2 & -W_3\end{pmatrix}\begin{pmatrix}u\\ d\end{pmatrix} \tag{A.11}$$

for the convenience, we can re-write it as follows.

Define, $W^+ = \dfrac{W_1 - iW_2}{\sqrt{2}}$ and $W^- = \dfrac{W_1 + iW_2}{\sqrt{2}}$, we get

$$\begin{pmatrix}\bar{u} & \bar{d}\end{pmatrix}\frac{g}{2}\begin{pmatrix}W_3 & \sqrt{2}W^+\\ \sqrt{2}W^- & -W_3\end{pmatrix}\begin{pmatrix}u\\ d\end{pmatrix} \tag{A.12}$$

The above result is true for all the generations of quarks and leptons. The W^+, W^- are the two charged massive gauge bosons that interact with fermions and change their flavor (ex: W^+ acting on d quark turn it into a u quark and vise versa). W_3 (which is Z^0 boson) does not affect the flavor of the fermion it's acting on.

Pauli matrices:

$$\sigma_1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \sigma_2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}, \sigma_3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix} \tag{A.13}$$

A.6 SU(3) Group

SU(3) is the special unitary group transformation acting on the 3-D complex vectors. The generators for the SU(3) group are derived from the "Gell-Mann" matrices(λ_i ; i=1...8) and there are 8 of them. The SU(3) group is important to explore "Quantum Chromodynamics(QCD)" which explains how quarks and gluons (strong interaction mediators) have colors, how they confine to be colorless, and how they interact. Thus, SU(3) is laying a foundation to understand the strong interaction in the SM. The quarks carry a color charge with them: red(r), blue(b) and green(g) (anti-quarks have anti-colors, anti-red($\bar{r}$), anti-blue($\bar{b}$), anti-green($\bar{g}$)) and gluons carry a pair of color,anti-color (ex: r$\bar{b}$). In QCD, the strong interaction is invariant under SU(3) transformations and therefore it has an exact symmetry, which is known as the SU(3) color symmetry.

The three colors in the QCD can be defined as the SU(3) color states as follows.

$$r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, g = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{A.14}$$

For a given quark wave function Ψ,

$$\Psi = \begin{pmatrix} \Psi_r \\ \Psi_b \\ \Psi_g \end{pmatrix} = \begin{pmatrix} r \\ b \\ g \end{pmatrix} \tag{A.15}$$

The following representation is an example of the interaction of the color triplet (three colors) with the 8 gluons.

$$\text{Lagrangian } L \sim \bar{\Psi}\lambda_1 G_1 \Psi = \begin{pmatrix} \bar{r} & \bar{b} & \bar{g} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} G_1 \begin{pmatrix} r \\ b \\ g \end{pmatrix} = \bar{r} G_1 b + \bar{b} G_1 r \tag{A.16}$$

The above example represents the annihilation of the r quark and creating a b quark by emitting a gluon G_1. Thus, G_1 can be written as,

$$G_1 \sim \frac{1}{\sqrt{2}}(\bar{r}b + \bar{b}r) \tag{A.17}$$

Similarly, we can derive the rest of the gluons using the rest of the Gell-Mann matrices and they are as follows.

$$G_2 \sim \frac{-i}{\sqrt{2}}(r\bar{b} - b\bar{r}),\ \ G_3 \sim \frac{-1}{\sqrt{2}}(r\bar{r} - b\bar{b}) \tag{A.18}$$

$$G_4 \sim \frac{1}{\sqrt{2}}(b\bar{g} + g\bar{b}),\ \ G_5 \sim \frac{-i}{\sqrt{2}}(b\bar{g} - g\bar{b}) \tag{A.19}$$

$$G_6 \sim \frac{1}{\sqrt{2}}(g\bar{r} + r\bar{g}),\ \ G_7 \sim \frac{-i}{\sqrt{2}}(g\bar{r} - r\bar{g}),\ \ G_8 \sim \frac{1}{\sqrt{6}}(r\bar{r} + b\bar{b} - 2g\bar{g}) \tag{A.20}$$

Theoretically, there should be 9 gluons (3 $\times$ 3 colors $=$ 9) but, one of the gluons has the following arrangement which makes it to be colorless (net color charge in the arrangement $=$ 0). Therefore, operating it on any quark or gluon makes no change in them.

$$G_9 \sim \frac{1}{\sqrt{3}}(r\bar{r} + b\bar{b} + g\bar{g}) \tag{A.21}$$

The 8 gluon states we defined above can be arranged in such a way that the arrangement will start to appear as a charge operator,

$$\text{example: } r\bar{b} = \frac{G_1 - iG_2}{\sqrt{2}} \tag{A.22}$$

This result is representing the exchange of gluons among the quarks when they change their color.

APPENDIX B

VLQ signal sample validation studies

The analysis presented in this dissertation uses single produced VLQ Top signal samples. During the 2019 falls, it was found that the Lagrangian defined for the leading order UFO (Universal FeynRules Output) [51] model which is used in generating signal samples using the PYTHIA8 and MADGRAPH5 generators was having a miscalculation for the $T \rightarrow Ht$ decay process of,

$$\mathcal{L} = \bar{t}\gamma^{\mu}(c_L^{tZ}P_L + c_R^{tZ}P_R)TZ_{\mu} + \bar{t}(c_L^{tH}P_L + c_R^{tH}P_R)TH + h.c. \tag{B.1}$$

Where the SM top and the VLQ Top are misplaced in the Lagrangian, such that it needed to be rather given as,

$$\mathcal{L} = \bar{T}\gamma^{\mu}(c_L^{tZ}P_L + c_R^{tZ}P_R)tZ_{\mu} + \bar{T}(c_L^{tH}P_L + c_R^{tH}P_R)tH + h.c. \tag{B.2}$$

The misplacements of the SM and VLQ top quarks in the Lagrangian do not affect the gauge interaction through expanding the Higgs term as given in eq. B.1 for the wrong Lagrangian and in eq. B.2 for the corrected Lagrangian shows that the chirality is flipped, which is causing the mismodeling in the physics process of interest.

Expanding only the Higgs term in eq. B.1,

$$\bar{t}(c_L^{tH}P_L + c_R^{tH}P_R)TH = c_L^{tH}\bar{t}_R T_L H + c_R^{tH}\bar{t}_L T_R H \tag{B.3}$$

and in eq. B.2,

$$\bar{T}(c_L^{tH}P_L + c_R^{tH}P_R)tH = c_L^{tH}\bar{T}_R t_L H + c_R^{tH}\bar{T}_L t_R H \tag{B.4}$$

it is obvious that the chiralities in both the SM top and the VLQ Top are flipped. Due to the VLQ mixing with the SM 3^{rd} generation quarks as described in section 1.7, the flips in chiralities in both the SM top and VLQ Top, therefore, have ramifications in the production level of the signal samples. Once the mistake is identified, the corrected UFO model was prepared and before entering into the large scale signal sample production which will be eventually used by many analyses across the particle physics community, a validation study was required to identify which coupling constant(s) (κ) should be chosen to produce the VLQ signal samples.

We have participated in validating the VLQ samples produced by the $qb(T \rightarrow Ht)$ production mode that are mediated via a W boson and finally producing VLQ Top that undergoes $T \rightarrow Ht$. These samples were required by analyses such as all-hadronic $T \rightarrow Ht$ and 1-lep Ht/Zt+X thus, we have used a framework named "MG_Reweight_testinATLAS2" [1]

[1] given at: https://gitlab.cern.ch/avroy/MG_Reweight_testinATLAS2

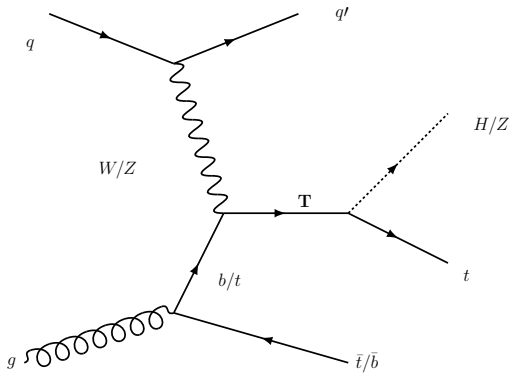
that produces the $T \rightarrow Ht$ events based on the corrected UFO model files using the MadGraph5 and Pythia8 MC generators.

The MG_Reweight_testinATLAS2 framework is producing signal samples with a nominal κ value of 1.0 and has the following embedded re-weighting options.

- Produced nominal mass can be re-weighted into a 100 GeV less mass point. Ex: 1100 GeV can be re-weighted into 1000 GeV
- Produced nominal κ (1.0 by default) can be re-weighted in to an array of κ values. Ex: κ = 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6.

We used the aforementioned features of the MG_Reweight_testinATLAS2 framework with an approach described bellow for the validation studies.

1. 55000 events were generated for the $Wb \rightarrow T \rightarrow Ht$ process with
   - Nominal mass points of 1100, 1700 and 2100 GeV
   - Nominal κ of 1.0 and 0.4
   - Left/Right handed (LH/RH) chirality
2. 11000 events were generated for testing with
   - Nominal mass points of 1100, 1600, 1700, 2000 and 2100 GeV
   - Nominal κ of 0.1, 0.4[2], 0.6 and 1.6
   - Left/Right handed (LH/RH) chirality



(a)

Figure B.1: Feynman diagram for the $Wb \rightarrow T \rightarrow Ht$ process.

The sample generation was based on the Feynman diagram given in fig. B.1a and we then reweighted the nominal samples to the same mass and κ values in the test samples to study how compared are they.

[2]only for the mass points of 1100, 1700 and 2100 GeV

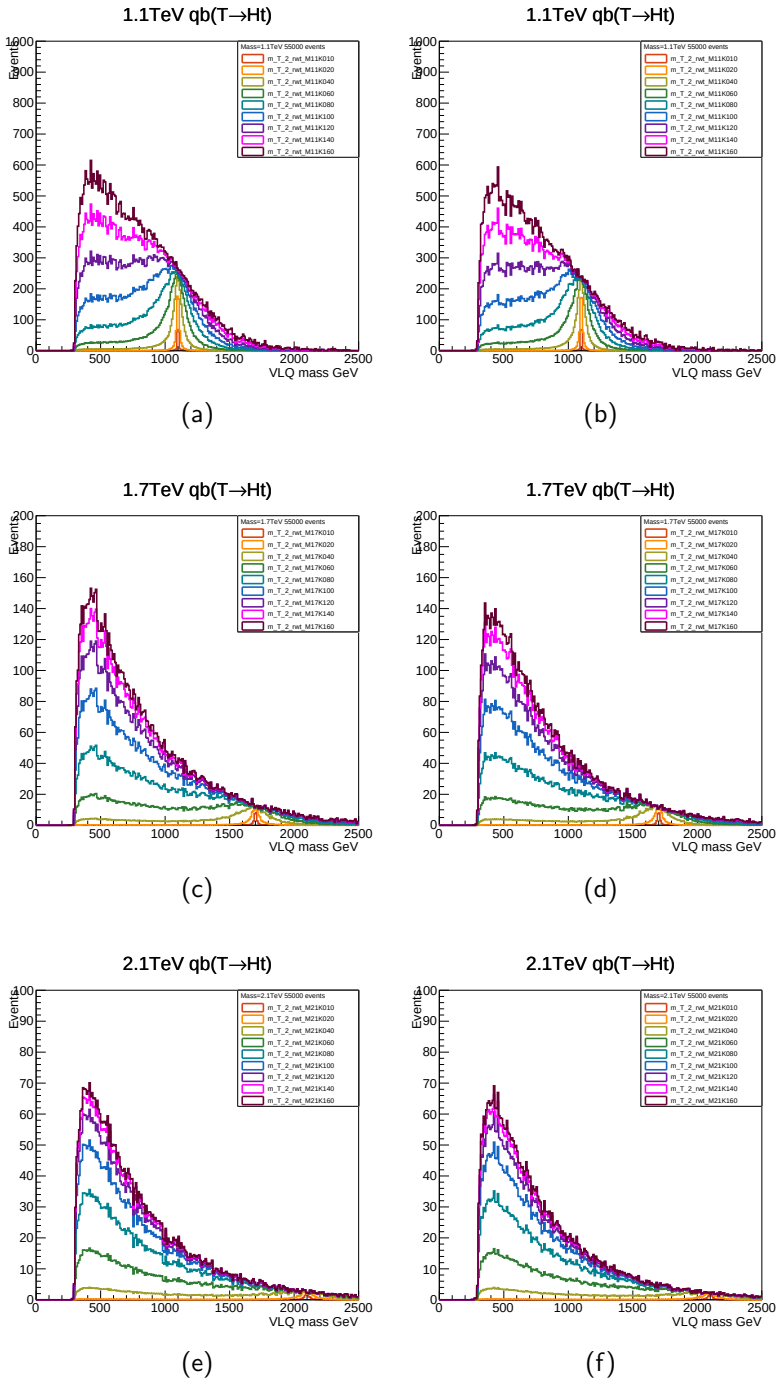
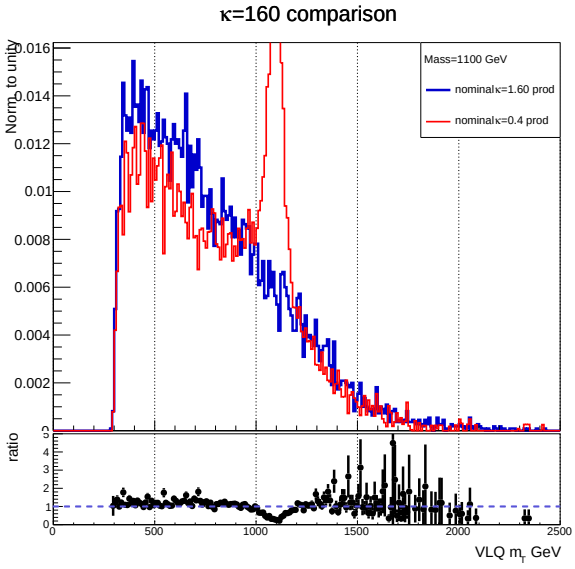


Figure B.2: The number of events distribution for the generated VLQ mass from $Wb \rightarrow T \rightarrow Ht$ with a nominal $\kappa = 1.0$ reweighted into an array of κ values. Fig. B.2a, fig. B.2c ,and fig. B.2e for left handed chirality and Fig. B.2b, fig. B.2d ,and fig. B.2f for right handed chirality.

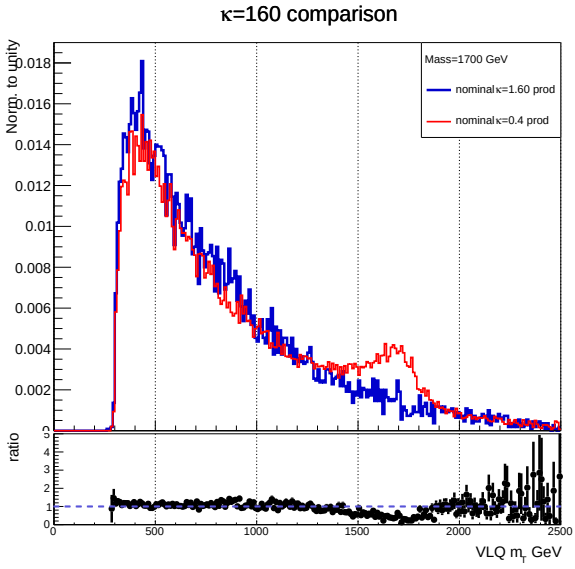
Since the kinematics between LH and RH samples show no significant difference and the left-handedness is more preferred when VLQ Top interact with SM top via weak interactions, it was decided to continue the study with the LH samples under the assumption

of results would be sufficient to interpret for the RH samples as well.

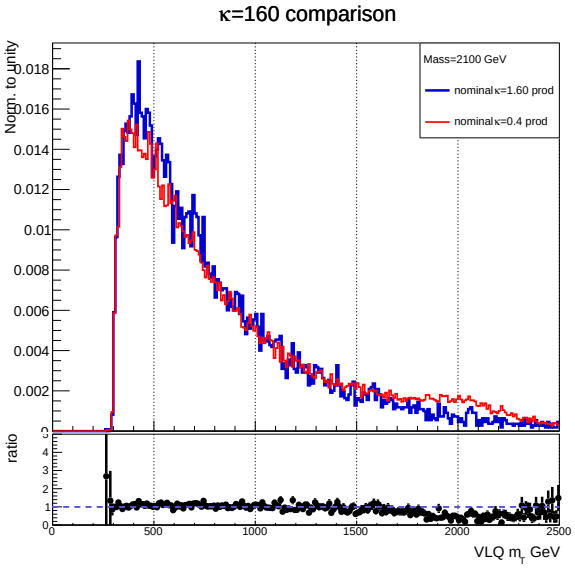
When re-weighting the samples produced with a nominal κ of 0.4, it was observed that its performance is inadequate while re-weighting into higher κ values.



(a)



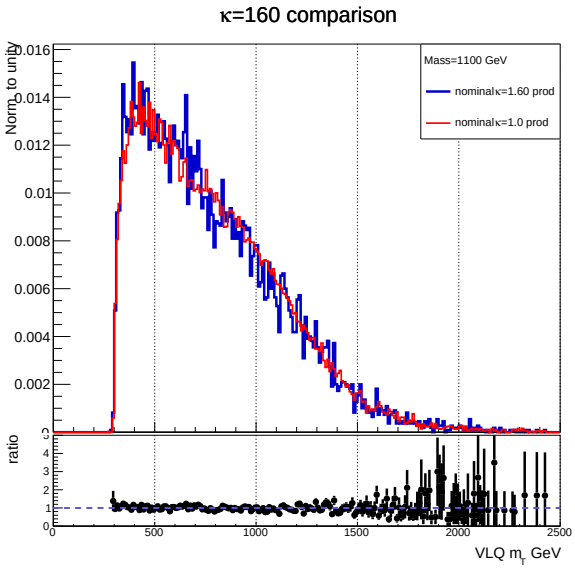
(b)



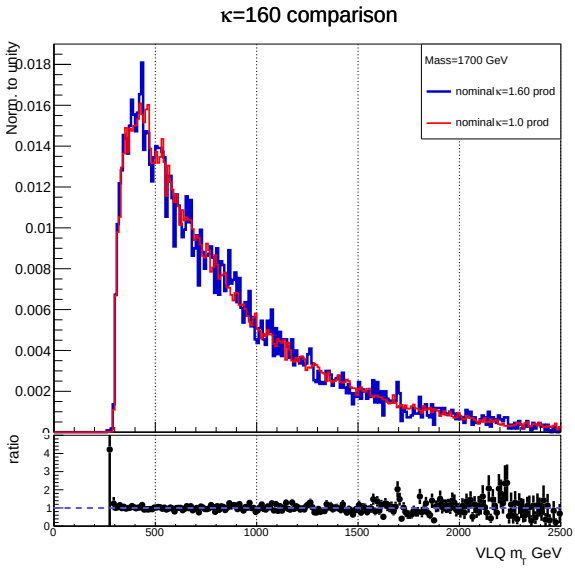
(c)

Figure B.3: Re-weighting of nominal samples produced with a κ=0.4 into κ=1.6. Fig. B.3a is for 1.1 TeV, fig. B.3b is for 1.7 TeV, fig. B.3c is for 2.1 TeV.

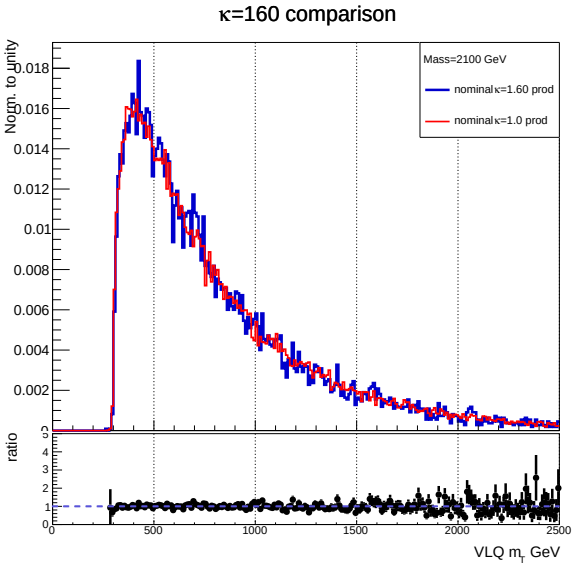
In comparison, samples produced with a nominal $\kappa = 1.0$, had no such inadequacy when reweighting into higher κ values.



(a)



(b)



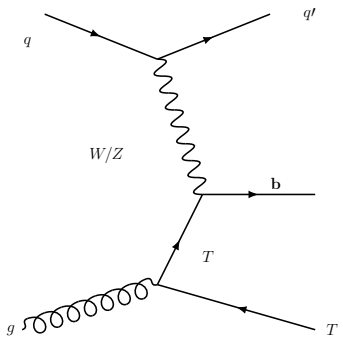
(c)

Figure B.4: Re-weighting of nominal samples produced with a κ=1.0 into κ=1.6. Fig. B.4a is for 1.1 TeV, fig. B.4b is for 1.7 TeV, fig. B.4c is for 2.1 TeV.

Therefore, it was decided to continue with the samples that are produced with a nominal κ of 1.0 for further validation studies.

The evaluation of the selection of $\kappa = 1.0$ was then tested by comparing the nominal samples produced with the mass points of 1100, 1700, 2100 GeV with the test samples made with the mass points of 1100, 1600, 1700, 2000 and 2100 GeV and κ of 0.1, 0.4[3], 0.6 and 1.6. For all the given mass points and the κ values, kinematic variables such as p_T (b-jet, Higgs jet, SM top jet, W jet, VLQ Top) and mass (SM top, W, VLQ Top) were used. Few of such comparison studies are presented in fig[give fig references].

It was brought into attention that there is another production mode as in fig. B.5 which can be considered in the $Wb \rightarrow T \rightarrow Ht$ process and we have tested the kinematic effects of the new process's inclusion into the sample production level at three stages,



(a)

Figure B.5: The second Feynman diagram for the $Wb \rightarrow T \rightarrow Ht$ process.

1. Produce samples with 25000 events including both production modes with mass points of 1700, 2100 GeV and nominal $\kappa = 0.4$ and $\kappa = 1.0$.
2. Compare samples produced with the production mode depicted in the Feynman diagram in fig. B.1 with the samples produced including both the diagrams.
3. Reweight the new samples (with both the production modes included) into different κ values to evaluate the reweighting performance.

When both the sample production modes were compared, a significant difference did not observe as shown in fig.[fig references]. The re-weighting studies for the new samples realized that the samples produced with a nominal $\kappa = 1.0$ show no significant difference when they are re-weighted to samples produced with a nominal $\kappa = 0.4$. Thus, the following conclusions were made towards the large scale sample production.

- Use the corrected Lagrangian based new UFO files in sample production.
- Set nominal κ to 1.0.
- Generate 250,000 events for mass points from 1.1 TeV to 2.3 TeV in 200 GeV steps.
- The generator weights will allow the generated samples to re-weight into 20 different κ values from 0.1 to 1.6 for the nominal mass and masses less than 100 GeV.

[3]only for the mass points of 1100, 1700 and 2100 GeV

APPENDIX C

$V \rightarrow Ht$ tagging and optimization studies

Three distinctive tagging algorithms are used in this analysis to identify top quarks, b-hadron jets (b-jets) and Higgs boson which will be used to recognize potential VLQ candidates that undergo in a $T \rightarrow tH$ where $H \rightarrow b\bar{b}$ and $t \rightarrow Wb$ with the W boson decays hadronically. Then the three algorithms are optimized harmoniously by a binned PL fitting using the `TRexFitter`.

C.1 Tagging optimization

The fit is evaluating a 95% CL on the VLQ signal strength (μ) for three regions defined as,

- SR1: signal regions 68 and 86 combined
- SR2: signal regions 89 and 98 combined
- ttCR: combination of $t\bar{t}$ notmalization regions (66+69+96+99).

Due to the limited statistics in the signal regions, four signal regions were combined to form two regions SR1 and SR2, allowing the fitting to more stable. The scale factor α described in eq. 4.5 is also considered in the fitting process along with the μ for the ttCR. As our analysis is still in the un-blinded phase, we avoided using Run2 data in the SR1 and SR2, and conducted the optimization studies with the Asimov data [1] which is defined in our analysis as the summation of predicted yields in the SR's and a VLQ sample with a cross-section of 10 fb. The summation of scaled non all hadronic MC samples, scaled $t\bar{t}$ all hadronic samples, single top, and multijet contributions are considered as the predicted yields in the signal regions.

The optimization studies used three differnet VLQ samples with nominal masses of 1.1, 1.7 and 2.1 TeV produced with a nominal coupling constant (κ) of 1.0. These samples can be re-weighted in to mass values of 1.0, 1.6 and 2.0 TeV while each of these mass points can be re-weighted to different κ values of κ = 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9,1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6.

An approach of setting a 95% CL on μ was selected to test with different VLQ masses and their κ values to determine the best taggers. Throughout the optimization studies, all the VLQ samples considered were set to have an expected cross-section of 10 fb irrespective of the mass and the κ considered making the limit on the μ is a limit implied on the observed cross-section for the given mass and κ given in table C.4 divided by 10 fb.

[1] Asimov data: A method of estimation where an ensemble of data sets being replaced by a single representative dataset [48].

The taggers that produced the lowest expected limits on masses were considered as the best selections and this analysis is more sensitive towards masses $\geq$ 1.6 TeV.

C.2 b-tagging

Our analysis is interested in high p_T objects associated with b-hadrons such as b-jets. Thus, in optimizing the acceptance of high p_T b-jets, we made a preference in selecting jets coming from the ID information (track jets) over the jets form at the calorimeter (calo jets). Even though the calo jets provide a higher background rejection, along with the other tagging optimizations discussed in sections XX and YY, an overall improvement on signal vs background yield can be seen when the track jets are in use as depicted in fig. 4.4. The variable track jets were preferred over the fixed radius track jets (R=0.4) as they reduce the radius parameter of the reconstructed track jets in the high p_T domain. Also, this selection of variable track jets improves the background rejection in the Higgs tagging studies.

The AnalysisTop 21.2.98 version provides two b-tagging algorithms MV2c10 and DL1 and both the algorithms have four defined working points at 60%,70%,77%, and 85%. During the b-tagging optimization studies, Higgs tagger kept at a fixed $\tau_{21}$and the top tagger was set constant at contained 80% selection. The aforementioned working points were then tested in comparing expected 95% CL on μ with VLQ masses of 1.6 TeV and higher with the coupling constants (κ) values of 0.5, 1.0, and 1.5. It's obvious from fig. that at all the coupling values and VLQ masses that the 70% working point in MV2c10 tagger has better performance while higher the VLQ mass and κ, DL1 tagger at 70% performs closer with MVc10 tagger at 70%.

Feature	Criterion
	EM Topo Jets / Track jets
Jet collection	`AntiKt4EMTopo/AntiKt2PV0/AntiKtVR30Rmax4Rmin02`
Jet selection	$p_T > 25$ GeV $\|\eta\| < 2.5$ JVT cut if applicable
Algorithm	`MV2c10`
Operating point	Fixed Eff = 70
CDI	`2017-21-13TeV-MC16-CDI-2017-12-22_v1`

Table C.1: Event selection criteria for b-tagging.

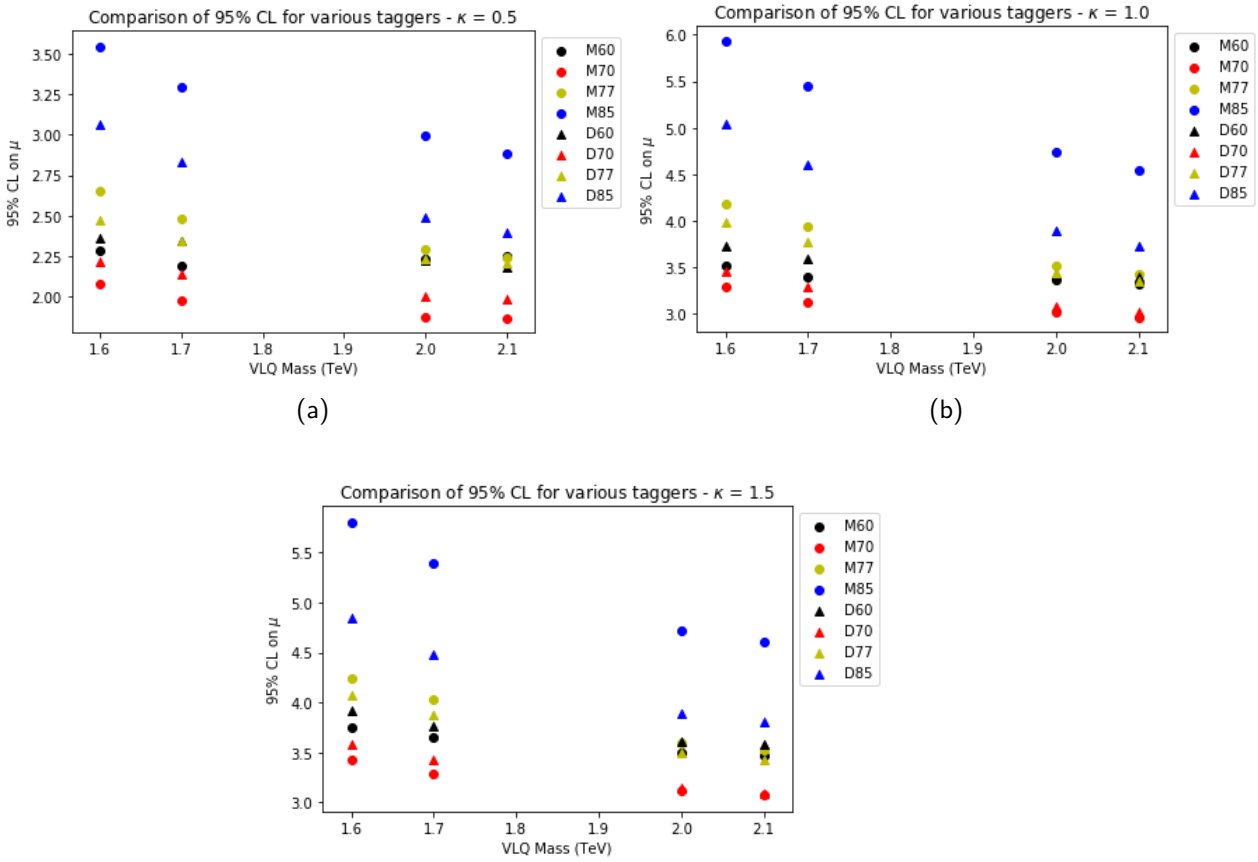


Figure C.1: Performance of working points in the two b-tagging algorithms studied in comparing the 95% CL on μ with differnt VLQ mass points derived from a binned PL by the `TRExFitter`. Figures C.1a, C.1b, C.1c refer to fits for κ values of 0.5, 1.0 and 1.5.

C.3 Top tagging

We were inspired by the all-hadronic top-quark differential cross-section analysis[10] in studying our $t\bar{t}$ normalization regions, and they were using a top-quark tagger that requires an explicit mass and τ_{32} requirements. But the DNN (Deep Neural Network) top-tagger we employed in our analysis is capable of identifying top-quarks without having a dependency on using an explicit mass cut implied on the large-R jet candidate. The DNN top-tagger is discriminating the jets coming from gluons, bosons and lighter quarks (u, d, s) in favor of the top-quark jets by using high-level variables feeding into a deep neural network that processes the inputs with an $Adam(AdaptiveMomentEstimation)$ [56] optimizer.

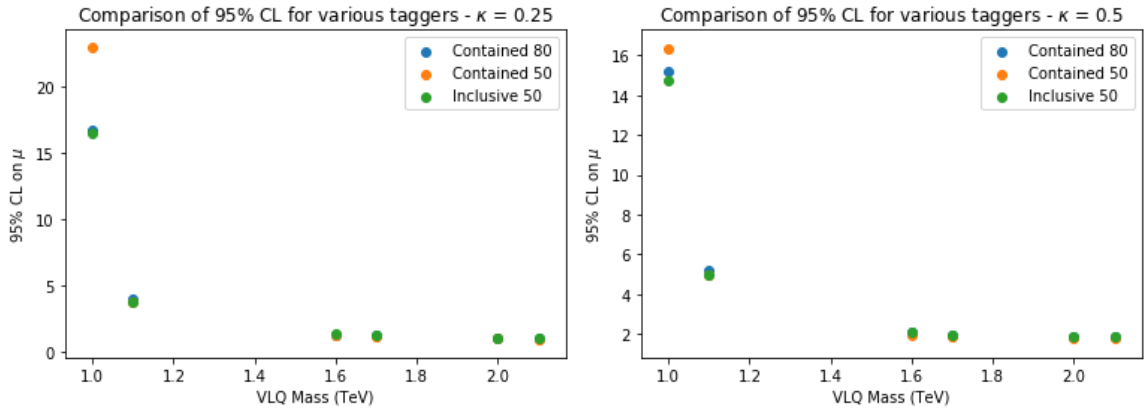
Since we have a Higgs candidate involvement in our analysis, we are required to overcome possible overlapping of Higgs-tagged and top-tagged jets. This is to avoid identification of a given large-R jet as top-tagged and Higgs-tagged, allowing the application of ABCD method for background evaluation across all the signal, $t\bar{t}$ normalization, and the validation regions from the multijet events in the analysis.

This is achieved first, by optimizing the Higgs tagger using a mass window defined for the large-R jet (as described at section C.4 in appendix C). A mass window of 140 - 225 GeV was then defined for the DNN top-tagger making any large-R jet candidate that belongs to this window to be identified as a top-tagged large-R jet. This definition of the mass window for the DNN top-tagger splits the mass window defined for the large-R jet into two and the lower mass region is then used for the Higgs tagging studies.

The DNN top-tagger is providing two tagging options based on either the decay products from the top-quark are fully contained in the large-R jet or not and thus, they are named as $contained$ and $inclusive$ taggers. The tagger is referred to as $inclusive$ when the samples with decay products are not fully contained in the large-R jet and if it is contained, the tagger is then known as the $contained$ tagger. Both the tagging options have two working points 50% and 80% making four sub-options to test for.

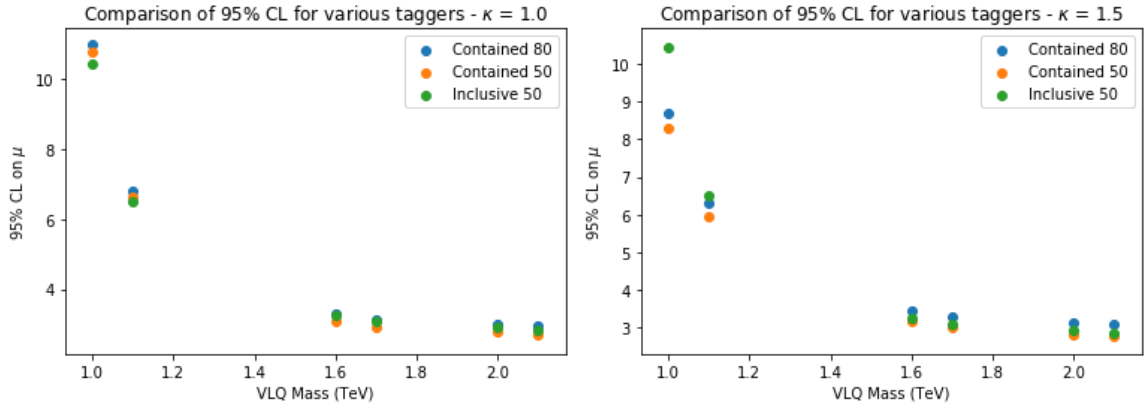
During the top-tagging optimization studies, the b-tagging was kept at 70% working point for the MV2c10 tagger and the Higgs-tagging was kept at the fixed τ_{21}cut at 0.45. Then the aforementioned contained and inclusive DNN top taggers were trained at the two working points of 50% and 80%, where, for the contained tagger, it required the signal jets to be matched with a truth groomed jet with a mass $\geq$ 140 GeV, truth top-quark and at least one ghost b hadron has to be ghost matched with a truth groomed jet. For the inclusive tagger, it required the signal jets to be matched with a truth top-quark. The inclusive top tagger with 80% working point was dropped later as it was having a higher amount of background in the signal regions. The top tagging scale factor systematics were not considered during the optimization as they were not available for all the taggers and the b-tagging scale factor systematics were used with pruning them.

The optimization was performed on VLQ masses of 1.0, 1.1, 1.6, 1.7, 2.0 and 2.1 TeV at κ values of 0.25, 0.5, 1.0 and 1.5 by setting an expected CL of 95% on the μ. From the fig.C.2 one can see that for VLQ masses $\geq$ 1.6 TeV and for higher κ values, the contained tagger at 50% working point has the lowest values for the 95% CL on μ. Along with this observation and due to the fact that this analysis is less sensitive to lower VLQ masses of 1.0 and 1.1 TeV, the contained tagger at 50% working point was determined as the best choice.



(a)

(b)



(c)

(d)

Figure C.2: Performance of working points in the three top-tagging algorithms studied in comparing the 95% CL on μ with differnt VLQ mass points derived from a binned PL by the `TRExFitter`. Figures C.2a, C.2b, C.2c and C.2d refer to fits for κ values of 0.25, 0.5, 1.0 and 1.5.

Feature	Criterion
Algorithm	anti-k_t
R-parameter	1.0
Input constituent	LCTopo
Grooming algorithm	Trimming
f_{cut}	0.05
R_{trim}	0.2
Analysis release number	21.2.98
CalibArea tag	00-04-81
Calibration configuration	JES_MC16recommendation_FatJet_JMS_comb_19Jan2018.config
Calibration sequence (Data)	EtaJES_JMS_Insitu
Calibration sequence (MC)	EtaJES_JMS
Selection requirements	
Observable	Requirement
p_T	> 300 GeV
$\|\eta\|$	< 2.0
Mass	> 100 GeV
Boosted object tagger	
Object	Working point
top	80%
$X \to bb$	TBD

Table C.2: Large-R jet reconstruction criteria for top tagging.

C.4 Higgs boson tagging

A simple Higgs boson taging algorithm is used in this analysis primirily requering following requirements,

- Jet mass window of 100 - 140 GeV
- N subjettiness of τ_{21}

The τ_{21}given by the ratio of $\tau_{21} = \frac{\tau_2}{\tau_1}$ describes how well a jet can be expressed in terms of subjets contained in it and therefore, the τ_{21}can be used as a discriminator in discriminating jets with two-prong structure from those who do not. Thus, these selections were applied and optimized for the capture of two-pronged action of $H \to bb$ decay.

C.4.1 Higgs-jet efficiency and background jet rejection evaluation

The VLQ signal samples with mass points of 1.1, 1.7, and 2.1 TeV (produced with κ =1) that include events with information of particle level and reconstruction level were used in the efficiency evaluation studies. A reconstruction level jet in a given event is considered to be Higgs-matched if it is within a $\Delta R \leq 1.0$ from a particle level jet which is ghost matched with two b-hadrons. The jets involved were screened through the Higgs-tagging algorithm and of those pass through the Higgs-tagger parameters were tagged as Higgs jets. The efficiency of tagging was determined by taking the ratio of Higgs-tagged jets to Higgs-matched jets for all the Higgs-matched jets.

The evaluation of background rejection was assessed on two major background sources in the analysis,

- top jets decaying to all-hadronic final state
- multijet

The Higgs-tagging algorithm is then applied on all the large-R jets considered and the rejection is defined as the ratio of the total number of large-R jet to the number of large-R jets that are mis-tagged as Higgs jets. An MC sample purely comprised of $t\bar{t}$ events with an all-hadronic decay was selected in the rejection of all the top jets and data measured during the period of 2015-2016run with large-R jets that are unassociated with any b-tagged small-R jets were used in the multijet background rejection.

C.4.2 Optimization of the Higgs-taggger

Once the efficiency and the background rejection methods are established, optimization of the Higgs tagger begun with varying the two parameters mentioned earlier. The mass window was explored by setting the upper mass cut at 145 GeV and the τ_{21} at a value of 0.45 to achieve an equilibrium between the tagging efficiency and the background rejection. The lower mass cut was then varied and observed that the increment in the lower mass cut is decreasing the background rejection versus the Higgs-tagging efficiency as shown in fig and it was decided that the lower, mass cut at 100 GeV is optimal as any increment beyond that point significantly decrease the efficiency. This observation was universal through all the VLQ mass points considered as well as the rejection of top and multijet backgrounds.

A similar study was performed by setting a lower mass cut of 95 GeV and have the upper mass cut to be variable while a τ_{21} cut of 0.45 is applied and is depicted in figC.3. Based on the behavior of the Higgs-tagging efficiency as a function of the higher mass cut variation, the upper mass cut at 140 GeV was selected to be the optimal selection.

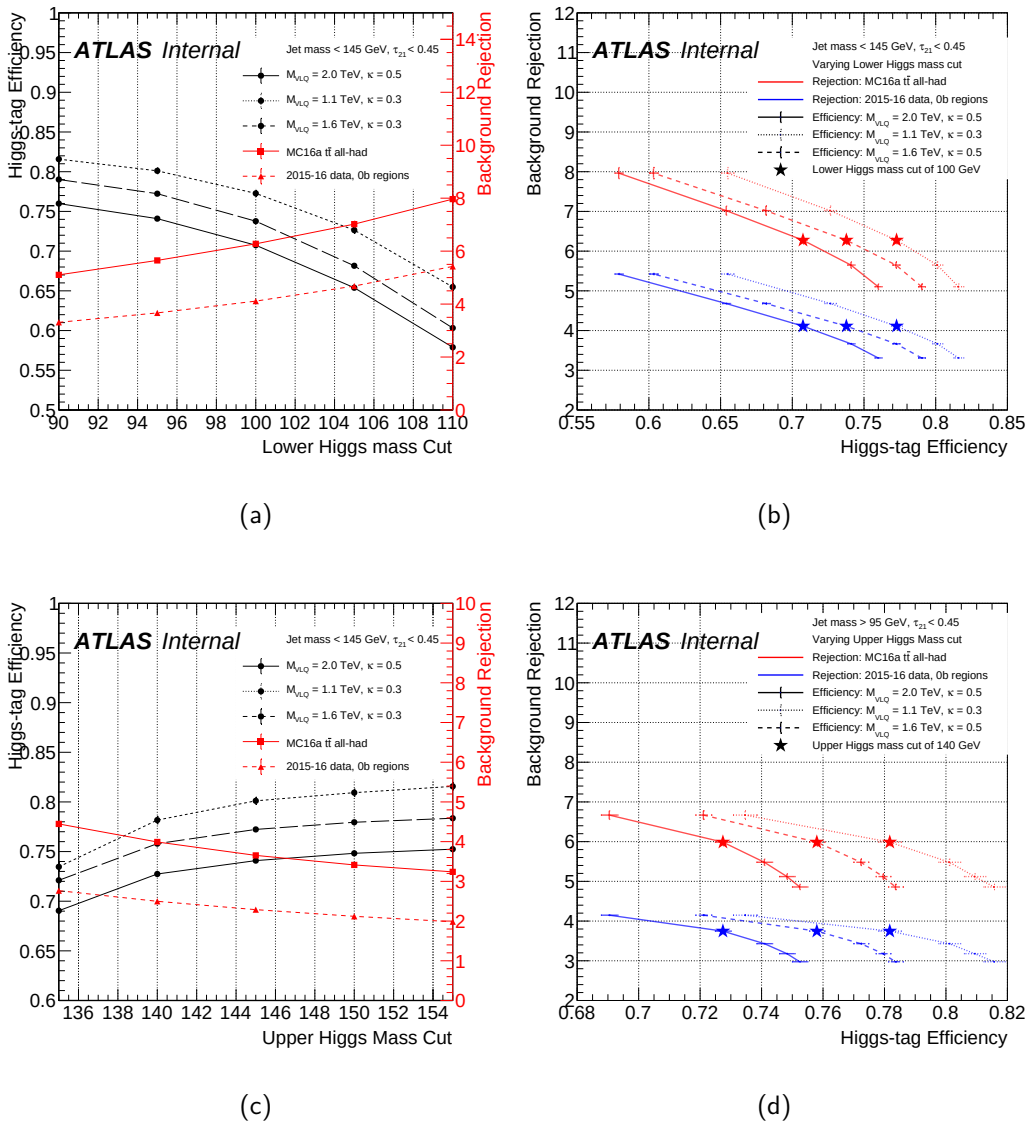
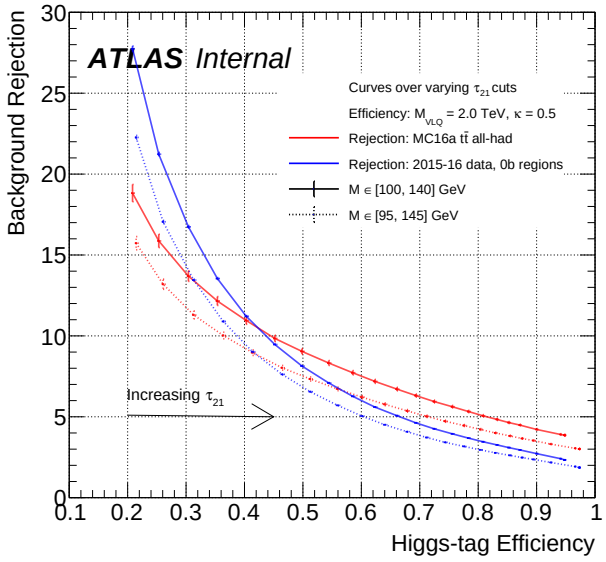


Figure C.3: Figure C.3a in the top left is showing the Higgs-tagging efficiency as a function of lower mass cut distribution (in black) and the background rejection as a function of lower mass cut. Figure C.3b in top right showing the background rejection as a function of Higgs-tagging efficiency and lower mass cut of 100 GeV is highlighted by using a star. Figure C.3c in the bottom left is showing the Higgs-tagging efficiency as a function of upper mass cut distribution (in black) and the background rejection as a function of upper mass cut. Figure C.3d in bottom right showing the background rejection as a function of Higgs-tagging efficiency and upper mass cut of 140 GeV is highlighted by using a star.

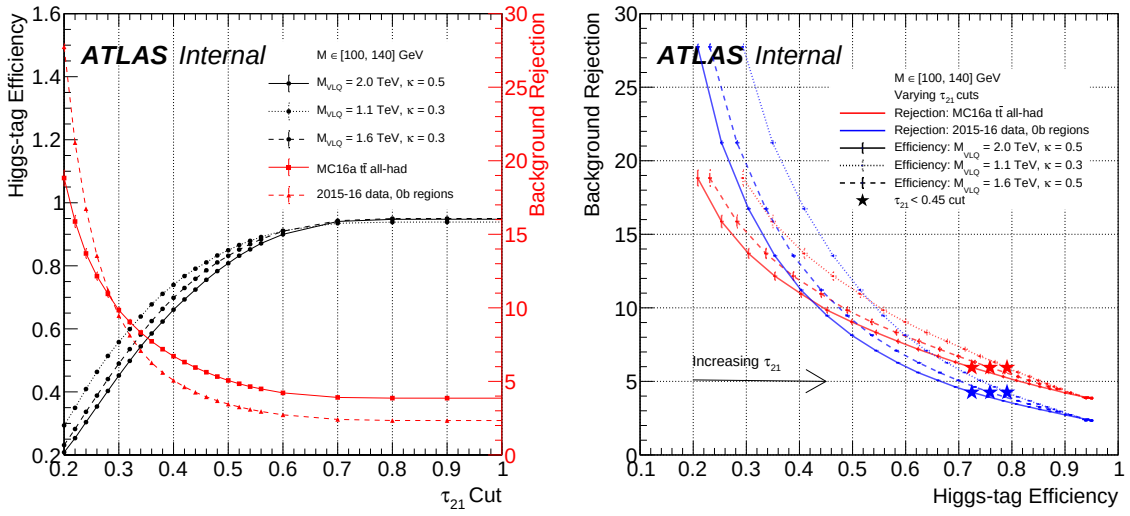
A study on two mass windows defined as loose [95,145] GeV and tight [100,140] GeV was performed as a comparison using a varying τ_{21}distribution. Efficiencies derived with respect to the Higgs jets coming from the VLQ mass sample of 2.0 TeV were used here. Based on fig C.4 the tight mass window performs better at all the efficiency points when compared with the loose mass window.



(a)

Figure C.4: Background rejection as a function of Higgs-tagging efficiency when the τ_{21}vary from 0.2 - 1.0 (τ_{21}= 1.0 is when there is no τ_{21}applied).

As shown from fig. C.4 the mass window of [100, 140] GeV has the best performance, it was then used to find the optimum τ_{21}cut for the analysis. Based on the performance shown in fig. C.5, a τ_{21}cut at 0.45 showed the best performance with an efficiency greater than 70% in rejecting the background in favor of the Higgs-tagging. The Higgs-tagging efficiency was observed to be dependent on the mass of the VLQ signal which produces the Higgs jets it is tagging. It is believed this dependency arises as a result of the correlation between the τ_{21}values with the p_T of the Higgs jet. This analysis is therefore explored the performance of a p_T dependent τ_{21}cut as it can make the efficiency of the tagger to be independent from the VLQ signal mass.



(a) (b)

Figure C.5: Higgs-tagger efficiency and background rejection performance for mass window of [100, 140] GeV with varying τ_{21}cut applied. Figure C.5a in left is the Higgs-tagging efficiency as a function of τ_{21}cut (in black) and background rejection as a fuction of τ_{21}cut (in red). Figure C.5b in right is the background rejection as a function of Higgs-tagging efficiency and the τ_{21}cut at 0.45 is highlighted in a star.

The table C.3 summarizes the Higgs tagging efficiencies and the background rejections for the Higgs-tagger with a mass window of 100 - 140 GeV and a fixed τ_{21}cut at 0.45.

VLQ Sample	Efficiency		
$m_{VLQ} = 2.0$ TeV, $\kappa = 0.5$	0.741	$\pm$	0.003
$m_{VLQ} = 1.1$ TeV, $\kappa = 0.3$	0.801	$\pm$	0.004
$m_{VLQ} = 1.6$ TeV, $\kappa = 0.3$	0.772	$\pm$	0.002

Background Sample	Rejection		
$t\bar{t}$ all-hadronic	5.65	$\pm$	0.08
2015-2016 data, no b-tags	3.738	$\pm$	0.006

Table C.3: Higgs tagging efficiencies for signal VLQ mass points and corresponding κ values and the background rejections for the prominant backgrounds.

The calculated efficiencies and the rejections for the Higgs-tagger were then plotted by binning them with respect to large-R jet p_T , mass and $|\eta|$ as follows.

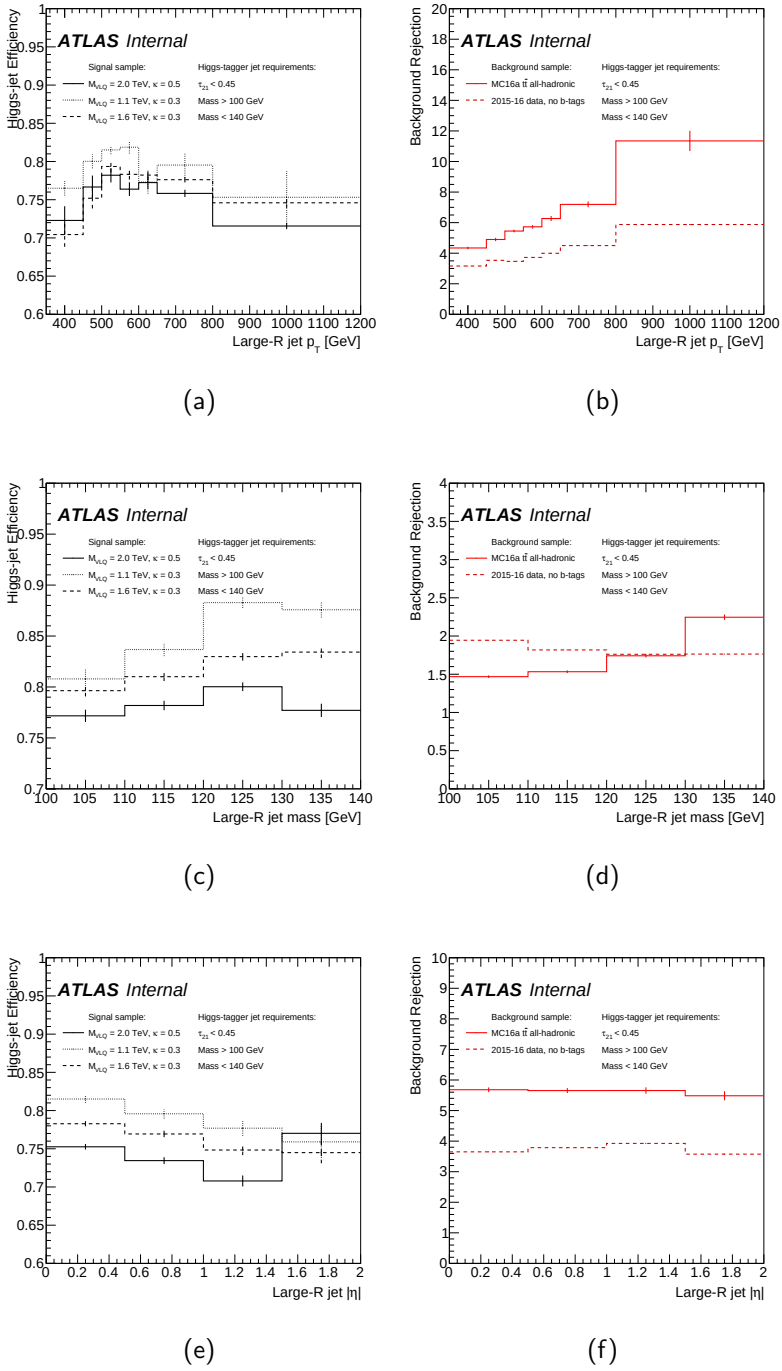


Figure C.6:
Figure C.6a is the tag efficiency and Figure C.6b is the background rejection w.r.t large-R jet p_T .
Figure C.6c is the tag efficiency and Figure C.6d is the background rejection w.r.t large-R jet mass.
Figure C.6e is the tag efficiency and Figure C.6f is the background rejection w.r.t large-R jet $|\eta|$.

C.4.3 Variale radius track jets

The Higgs tagging studies discussed in the previous section were conducted using a fixed radius (R=0.4) track jets for b-tagging and ntuple production using AnalysisTop version of 21.2.45. But the later AnalysisTop versions ($>$ 21.2.72) are replacing the fixed radius track jet with the variable radius track jets [7]. A study was conducted to find the Higgs tagging efficiencies and the background rejection as a function of varying $\tau_{21}$and they are depicted in fig. C.7.

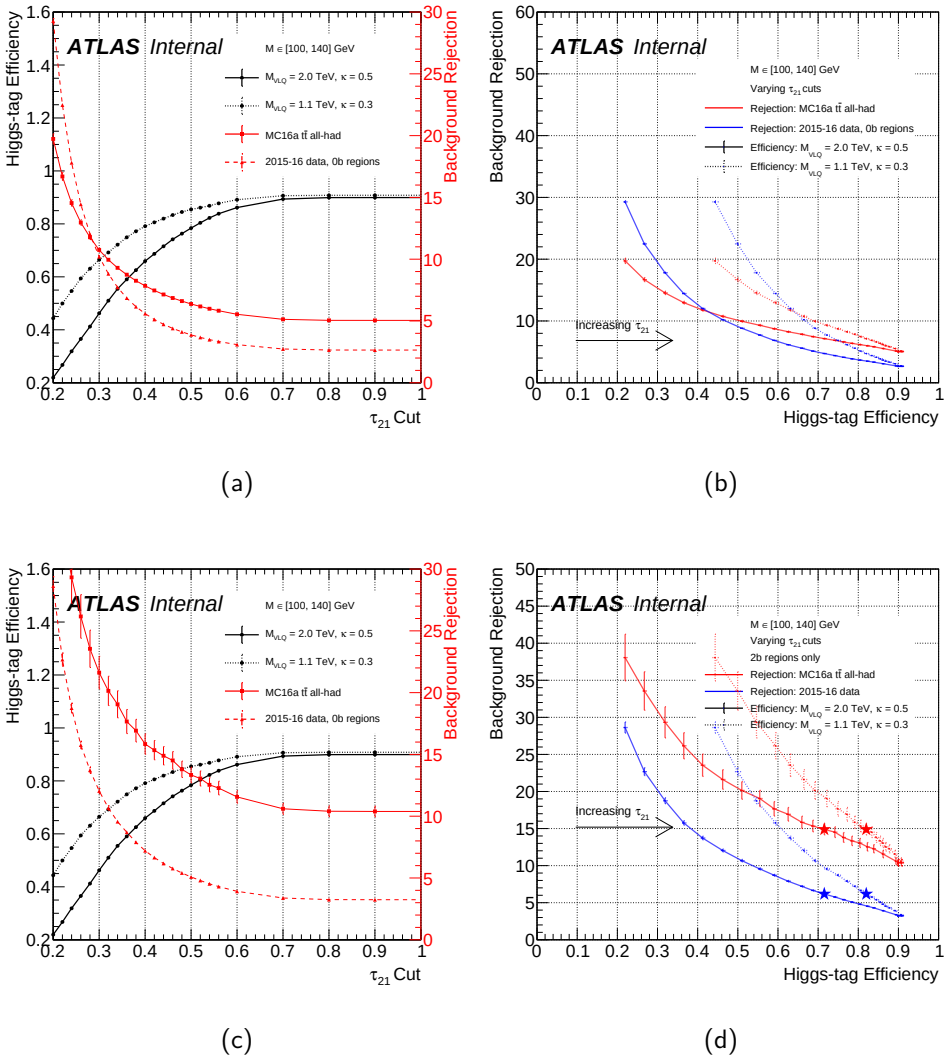
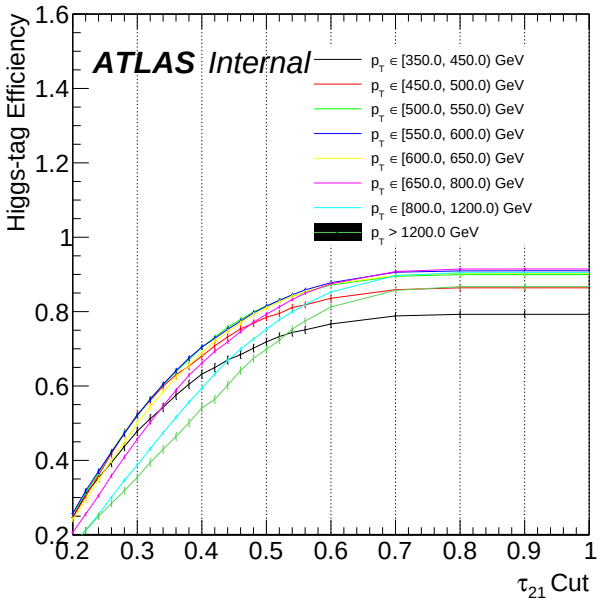


Figure C.7:
Figure C.7a is the tag efficiency and Figure C.7b is the background rejection for the [100 - 140] GeV mass window for varying τ_{21}cut.
Figure C.7c is the tag efficiency and Figure C.7d is the background rejection for the [100 - 140] GeV mass window for varying τ_{21}cut and two b-tagged regions only. the τ_{21}cut at 0.45 is highlighted in a star

C.4.4 The large-R jet p_T dependence of τ_{21}

It was realized from the studies mentioned before that the τ_{21}cut at 0.45 on all large-R jets was providing the best Higgs tagging efficiencies and the background rejections but, it causes to have an uneven performance across the p_T distribution. To address this issue, it was decided to apply a specific τ_{21}cut for each p_T bin so that the efficiency distribution in the p_T bins considered can be flattened into a single working point. Working points 50% and 70% were identified to be satisfying this requirement. The process of identifying these working points was performed as follows.

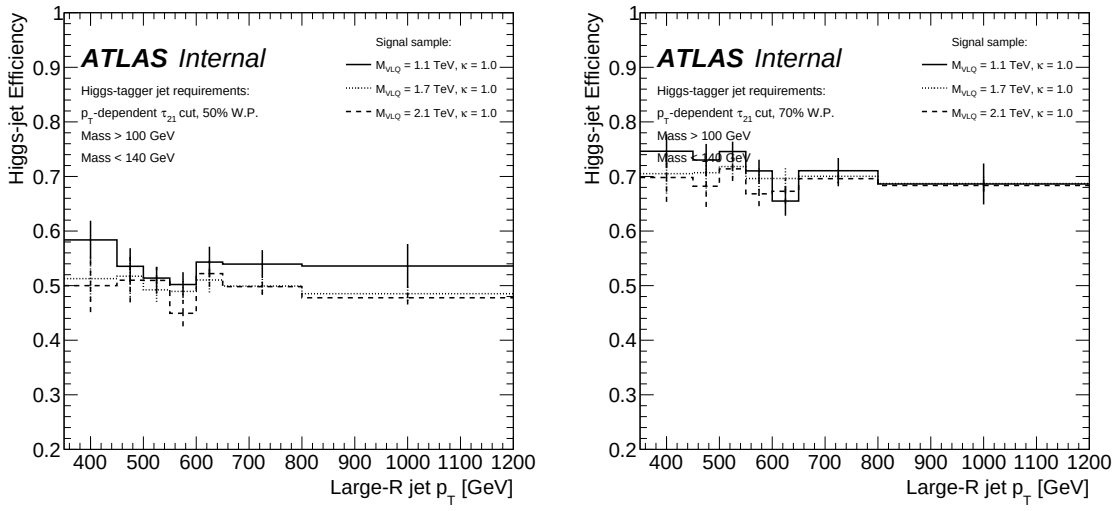
- For a given range of τ_{21}cuts, ntuples were generated with the calculated Higgs jet efficiencies.
- Find the τ_{21}value that gives the closest efficiency to the selected working point for each p_T bin.
- Arrange the τ_{21}cuts into an array corresponding to each p_T bins considered which are to be used in the Higgs tagger.
- Assess the efficiency of the τ_{21}independent Higgs tagger.



(a)

Figure C.8: Higgs-tagging efficiency as a function of varying τ_{21}cut for different p_T bins. The varying τ_{21}cut is to achieve a single working point across all p_T bins though no convergence can be seen. The figure comprises the VLQ signal samples of mass points 1.1, 1.7 and 2.1 TeV with $\kappa = 1.0$.

The figure C.8a is realizing that a single τ_{21}cut can not be defined to provide a consistent efficiency for the p_T spectrum rather, each p_T bin requires to have a specific τ_{21}cut. This understanding was tested on a selected single sample, for example, VLQ signal mass of 1.7 TeV from the MC16a campaign with a κ of 1.0, by defining τ_{21}cuts that provide a flattened efficiency working point throughout the p_T spectrum.

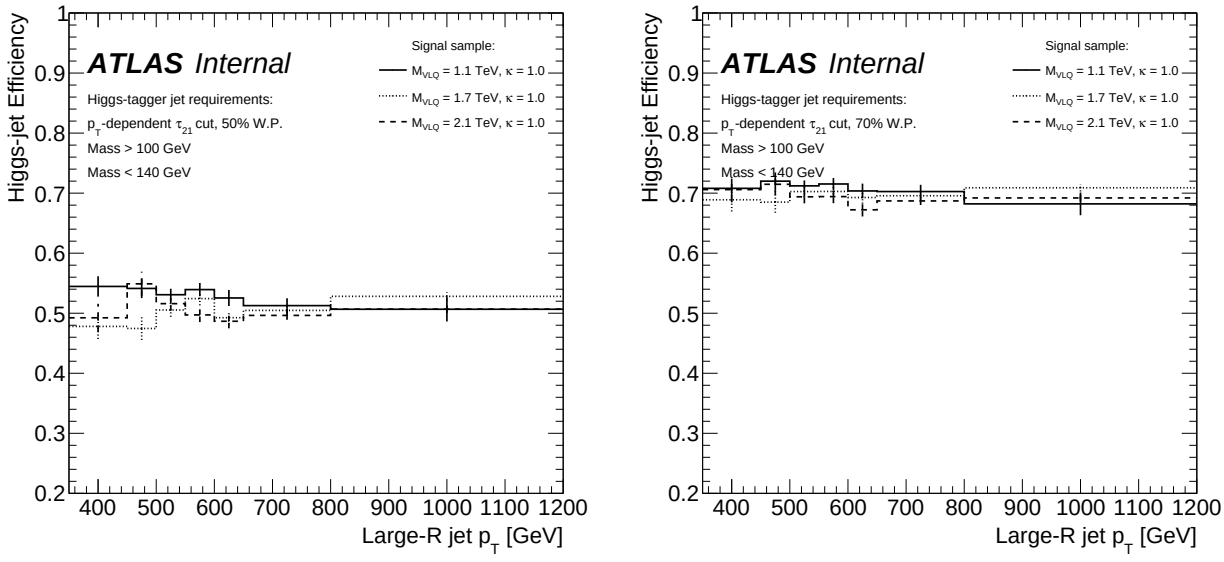


(a) (b)

Figure C.9: Higgs tagging efficiency as a function of large-R jet p_T for all MC16a VLQ signal samples. The τ_{21}values gave fig.C.9a 50% W.P. and C.9b for 70% W.P. from the 1.7 TeV signal sample.

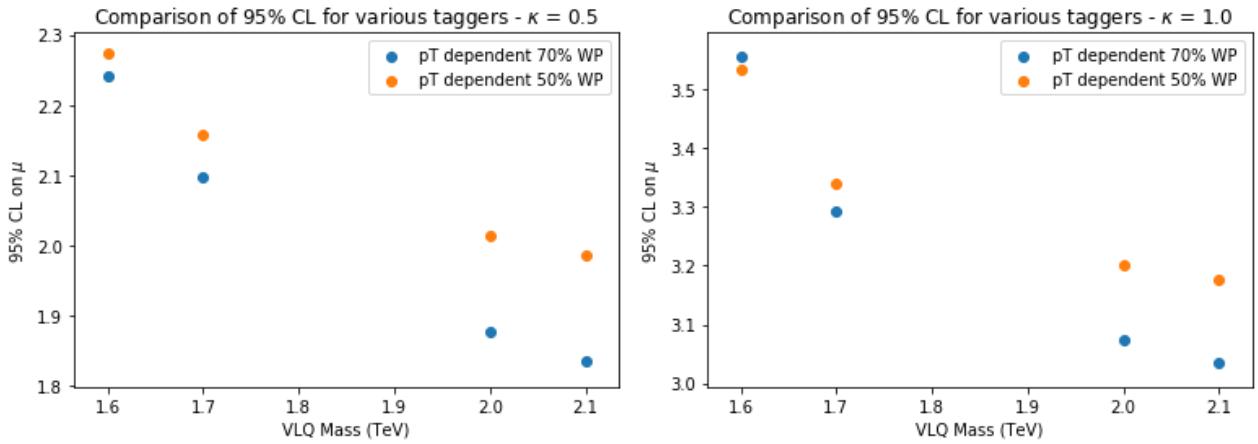
With the confidence came from the before mentioned sanity-check, the τ_{21}values were derived for all the signal mass points for all the MC16 campaigns. The fig C.10 is showing the efficiencies at 50% and 70% working points when all MC campaigns are combined for each mass point. It is also obvious from fig. C.10 that the efficiency is different for the mass points considered such that 1.1 TeV has the highest efficiency while the 2.1 TeV has the lowest.

In the optimization stage for the Higgs tagger, the 80% contained top tagger (no top tagging scale factor systematics were used), and the 70% MV2c10 b-tagger were selected. When tested on VLQ signal mass points $\geq$ 1.6 TeV and on κ values of 0.5, 1.0, and 1.5, it was found that the Higgs tagger at 50% W.P. has the best performance when expected 95% CLs were compared on μ and the results are given in fig C.11.

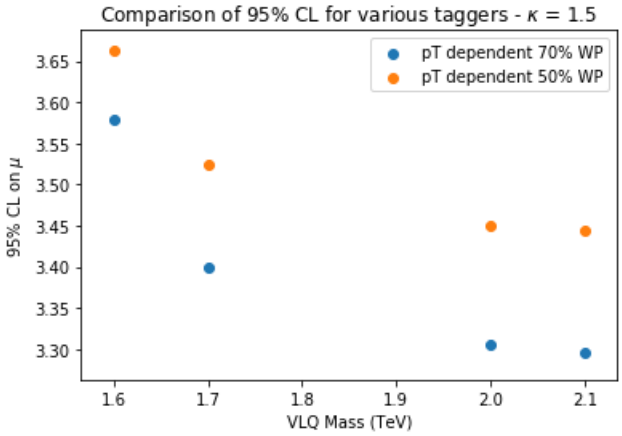


(a) (b)

Figure C.10: Higgs tagging efficiency as a function of large-R jet p_T for all VLQ signal samples (MC16 combined campaigns). The τ_{21}values agave fig.C.10a 50% W.P. and C.10b for 70% W.P. for the combined signal samples.



(a) (b)



(c)

Figure C.11: Performance of working points in the two Higgs-tagging algorithms studied in comparing the 95% CL on μ with differnt VLQ mass points derived from a binned PL by the `TRExFitter`. Figures C.11a, C.11b, C.11c refer to fits for κ values of 0.5, 1.0 and 1.5.

	Cross-section (fb) for W-mediated $pp \to T \to Ht$ with singlet T								
m_T:	κ: 0.1	0.2	0.3	0.4	0.5	0.6	0.8	1.0	1.6
1000 GeV	6.11	24.3	53.4	89.3	158	232	415	648	1660
1100 GeV	4.41	17.5	38.5	72.9	118	174	318	502	1290
1200 GeV	3.23	12.8	29.4	55	90.2	135	251	397	1020
1300 GeV	2.39	9.47	22.4	42.5	70.5	107	201	319	820
1400 GeV	1.78	7.05	17.3	33.4	56.3	86.2	164	261	671
1500 GeV	1.34	5.47	13.5	26.7	46	71.1	136	215	553
1600 GeV	1.02	4.25	10.8	21.8	38.2	59.7	114	180	462
1700 GeV	0.779	3.32	8.72	18.2	32.3	50.5	95.2	150	384
1800 GeV	0.599	2.62	7.16	15.4	27.8	43.3	80.4	126	323
1900 GeV	0.462	2.1	6	13.3	24.2	37.5	68.8	108	276
2000 GeV	0.358	1.71	5.1	11.7	21.4	32.8	59.8	93.5	239
2100 GeV	0.279	1.4	4.46	10.3	18.4	27.6	49.5	77.4	198
2200 GeV	0.218	1.17	3.96	9.19	15.8	23.2	41.3	64.5	165

Table C.4: Leading order cross-section (in fb) for W-mediated $pp \to T \to Ht$ for a left handed singlet VLQ T mass(m_T) and coupling(κ) at a center-of-mass energy if 13 TeV.

APPENDIX D

The τ_{21} uncertainty studies

D.1 Derivation of τ_{21} uncertainties

As described in Section C.4 in appendix C and fig. C.8, a varying τ_{21} cut is used for various p_T binns in the Higgs tagging optimization studies. As we consider data from Run 2, associated uncertainties with τ_{21} need to be determined and so far the consolidated recommendations provide jet uncertainties for the data collection period of 2015 - 2017. Since the pileup profile for 2018 is between those from 2015+2016 and 2017 data it is assumed that the uncertainties derived from the 2015-2017 data period can be generalized for 2018 data as well.

The τ_{21} uncertainties are derived in a process with several steps. RTrack methodology is used in assessing the systematic uncertainties in jet substructure variables and it is carried out by calculating double ratios which will be used as the inputs in the step of final uncertainty estimations.

Under the assumptions of systematic uncertainties of calorimeter measurements are largely uncorrelated with the tracking uncertainties and track jet mass is not measured better than the calorimeter jet mass, two components are defined as follows,

1. Ratio and double-ratio calculations of calorimeter-to-track jet mass for various MC
   - $r^{m}_{track\,jet} = \frac{m^{jet}}{m^{track\,jet}}$
   - $R^{m}_{r\,track\,jet} = \frac{r^{m,\,data}_{track\,jet}}{r^{m,\,MC}_{track\,jet}}$
2. Tracking uncertainty evaluation
   - $R^{m}_{r\,track\,syst} = \frac{r^{m,\,syst}_{track\,jet}}{r^{m,\,data}_{track\,jet}}$

The ratio calculation between the calorimeter and track jets in data and MC becomes important in the systematic uncertainty derivation as it indicates the quality in the detector effects simulations where well modelled simulation should give $R^{m}_{r\,track\,jet}$ as 1. The dijet samples generated from MC generators PYTHIA8, SHERPA, and HERWIG7 are used in the "RTrackUncertaintyGuide"package [1] which we used in deriving τ_{21} systematic uncertainties. The PYTHIA is used as the basline MC generator and the an additional uncertainty referred to as modeling uncertainty is derived using the SHERPA and HERWIG7 generators.

[1] given at https://twiki.cern.ch/ twiki/bin/viewauth/AtlasProtected/RTrackUncertaintyGuide

The systematic uncertainties are derived in bins either as calorimeter jet mass and jet p_T or the ratio of $\frac{m}{p_T}$. We adopted the ratio calculation of $\frac{m}{p_T}$ for binning as it is recommended for the boosted jets. The "RTrackUncertaintyGuide"is so far deriving the uncertainties inclusive to η and expects to explore the pseudo rapidity dependance with increasing datasets. During the uncertainty evaluations, only the mean of $r^m_{track\,jet}$ which is $\left\langle r^m_{track\,jet} \right\rangle$ considered. The Jet/Etmiss group is currently providing four variations towards the final systematic uncertainties as,

1. Baseline: measures how $R^m_{r\,track\,jet}$ is deviated from 1 using PYTHIA8
2. Modeling: measures how $R^m_{r\,track\,jet}$ is deviated from 1 using different MC generators (in our case HERWIG7 and SHERPA)
3. Tracking: take the quadratic sum of the effects from various uncertainty sources
    - tc1: track variation covers track reconstruction efficiency, dense environment efficiency and the impact paramter reconstruction
    - tc2: track variation covering the frake rate
    - tc3: track varioation covering the sagitta bias
4. Statistical: Uncertainty coming from the MC samples and datasets statistical power.

The "RTrackUncertaintyGuide"then takes,

Moment	XAOD jet attribute name
N subjettiness	Tau1, Tau2, Tau3, Tau21, Tau32, Tau1_wta, Tau2_wta, Tau3_wta, Tau21_wta, Tau32_wta
kT splitting scale	Split12, Split23, Split34
Energy correlations	ECF1, ECF2, ECF3, C2, D2
Qw	Qw

Table D.1: The JSS variables considered in RTrackUncertaintyGuide tool

jet substructure (JSS) variables and produces TH3 histograms from the ntuples which are then converted into TH2 histograms that are binned as $\frac{calo}{track}$, p_T , m or $\frac{m}{p_T}$. The TH2 histograms are then process to make plots and RTrack maps which will be used to make smoothed histograms. We selected the $\frac{m}{p_T}$ binning distribution of,

- 0.0 - 0.1 in 0.05 steps
- 0.1 - 0.3 in 0.1 steps
- 0.3 - 0.7 in 0.2 steps
- 0.7 -1.0

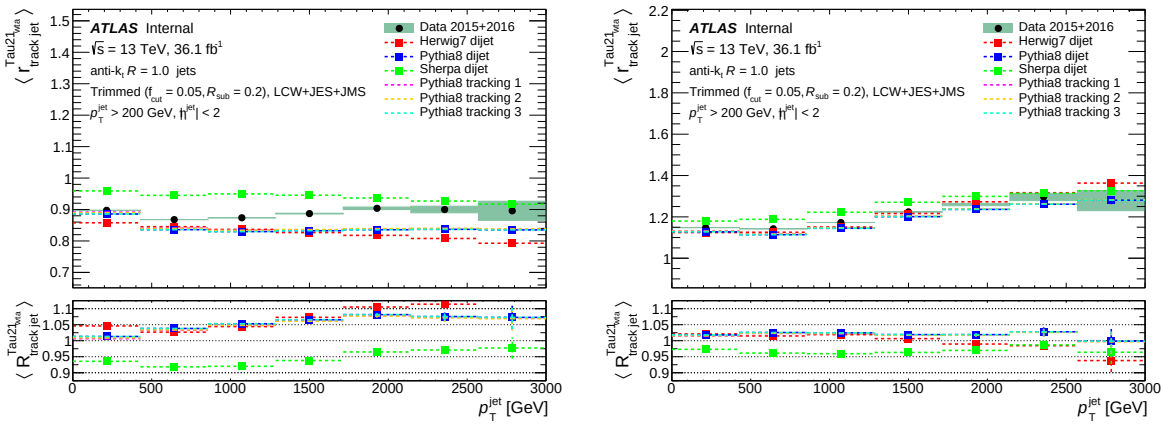
The results from the RTrackUncertaintyGuide are then passed to the "JetUncertaintyProvider" [2] which is combining the results from the RTrack maps generated in $\frac{m}{p_T}$ binnings into a single smoothed 2D histogram for a given JSS variable (in our analysis, τ_{21}) for PYTHIA8 as baseline and HERWIG7 and SHERPA as modeling.

The smoothing is a four-step procedure where,

- Step0: gives the raw histograms
- Step1a: gives smoothed raw histograms for baseline and modeling
- Step1b: smoothed raw histogram of the difference between baseline and modeling
- Step1c: maximum absolute deviation of baseline and modeling
- Step2: smoothed maximum absolute deviation histogram

D.2 Results

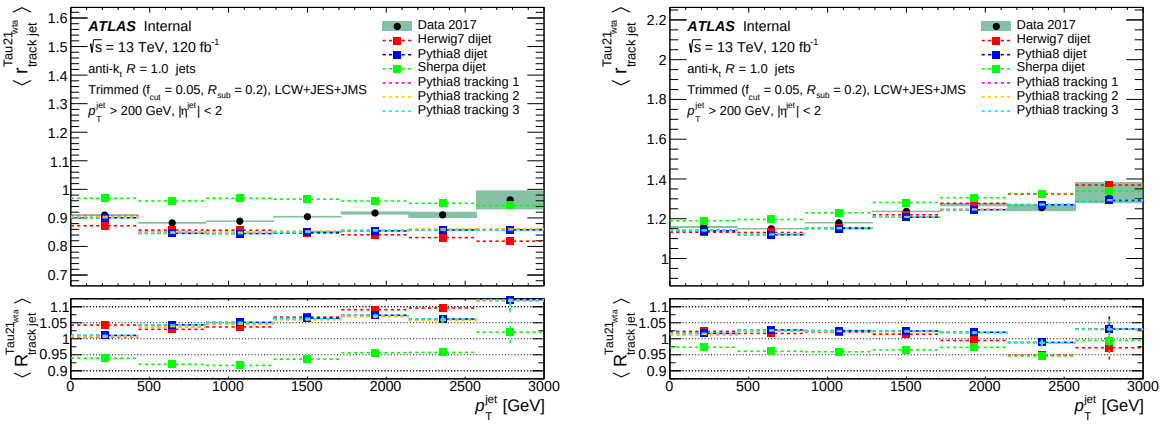
Using the JETM6 derivation files for rel21, we first derived the distribution of MC generators in the aforementioned $\frac{m}{p_T}$ regions. It is recommended in the RTrackUncertaintyGuide tool to compare 2015+2016 data with the MC16a campaign MC and 2017 data with the MC16d.



(a) (b)

Figure D.1: Figure D.2a for $\frac{m}{p_T}$ = 0.0 - 0.05 and D.2b for $\frac{m}{p_T}$ = 0.05 - 0.1 comparing 2015+2016 data with MC16a campaign.

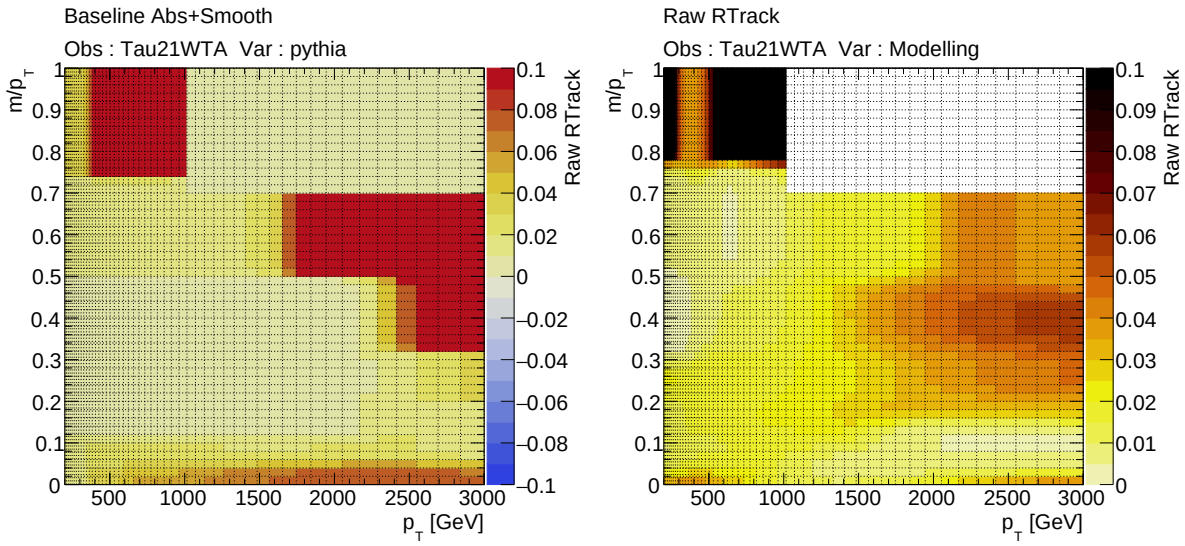
[2] given at: https://gitlab.cern.ch/atlas-jetetmiss-substructure/RTrackUncertaintyProcessing



(a) (b)

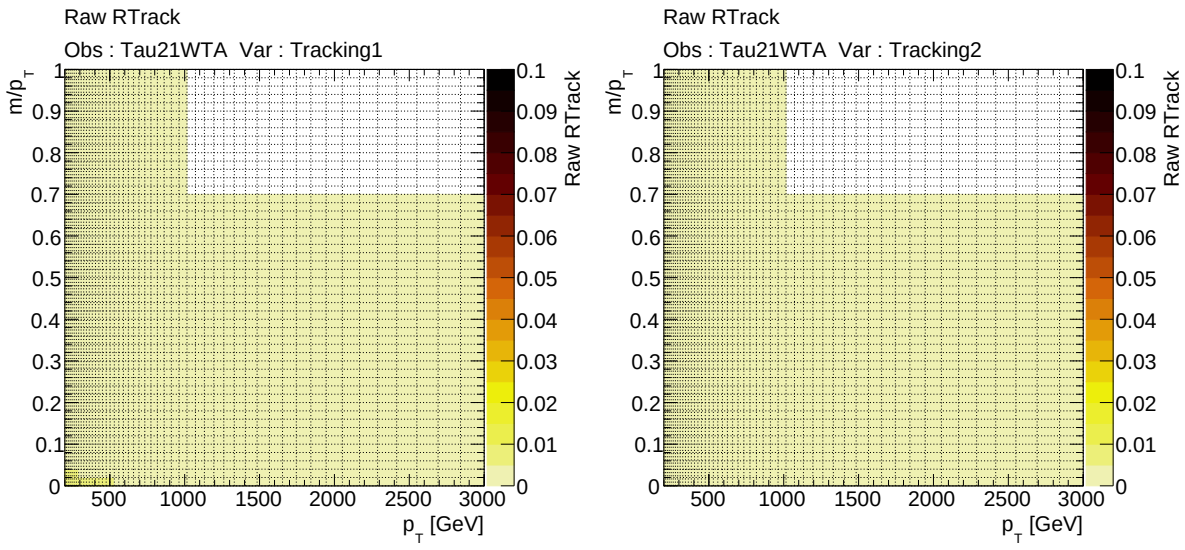
Figure D.2: Figure D.2a for $\frac{m}{p_T}$ = 0.0 - 0.05 and D.2b for $\frac{m}{p_T}$ = 0.05 - 0.1 comparing 2017 data with MC16d campaign.

And the resulting smoothes 2Dhistograms for each MC generators are as follows,

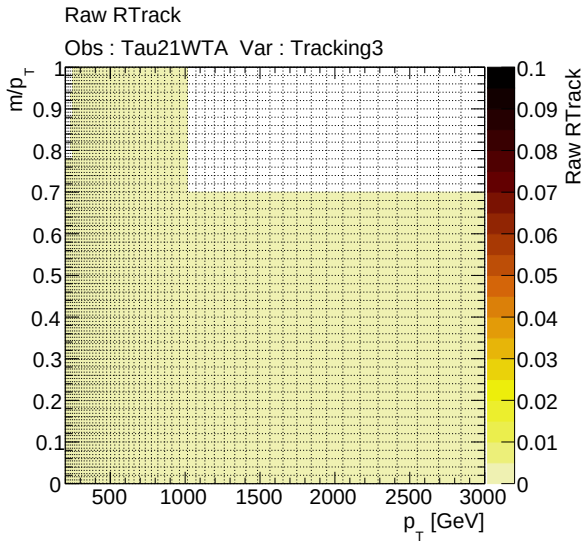


(a) (b)

Figure D.3: Figure D.3a for PYTHIA8 and D.3b for Modeling (HERWIG7 and SHERPA) smoothed histograms for 2015+2016 data and MC16a samples.



(a) (b)



(c)

Figure D.4: Figure D.4a for tc1, D.4b for tc2 and D.4c for tc3 smoothed histograms.

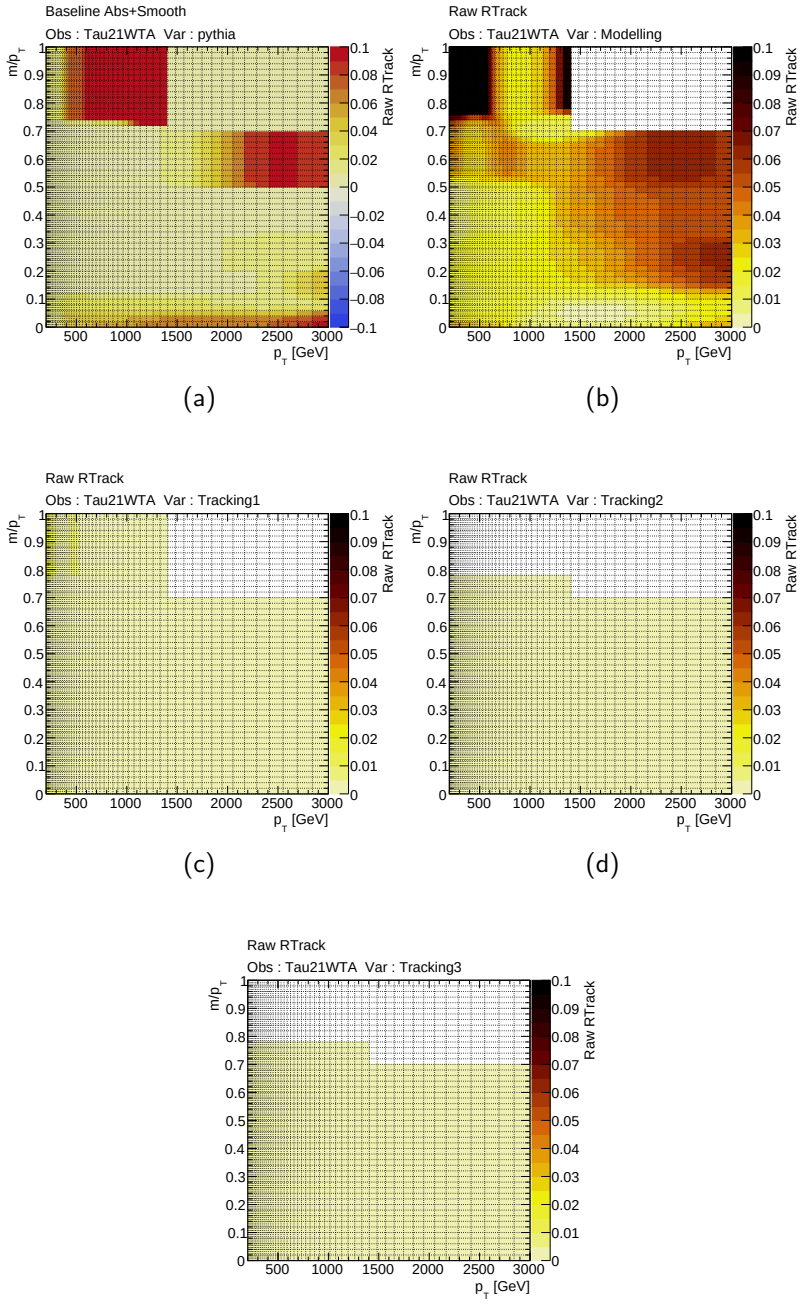


Figure D.5: Figure D.5a for PYTHIA8 and D.5b for Modeling (HERWIG7 and SHERPA), figure D.5c for tc1, D.5d for tc2 and D.5e for tc3 smoothed histograms for 2017 data and MC16d samples.

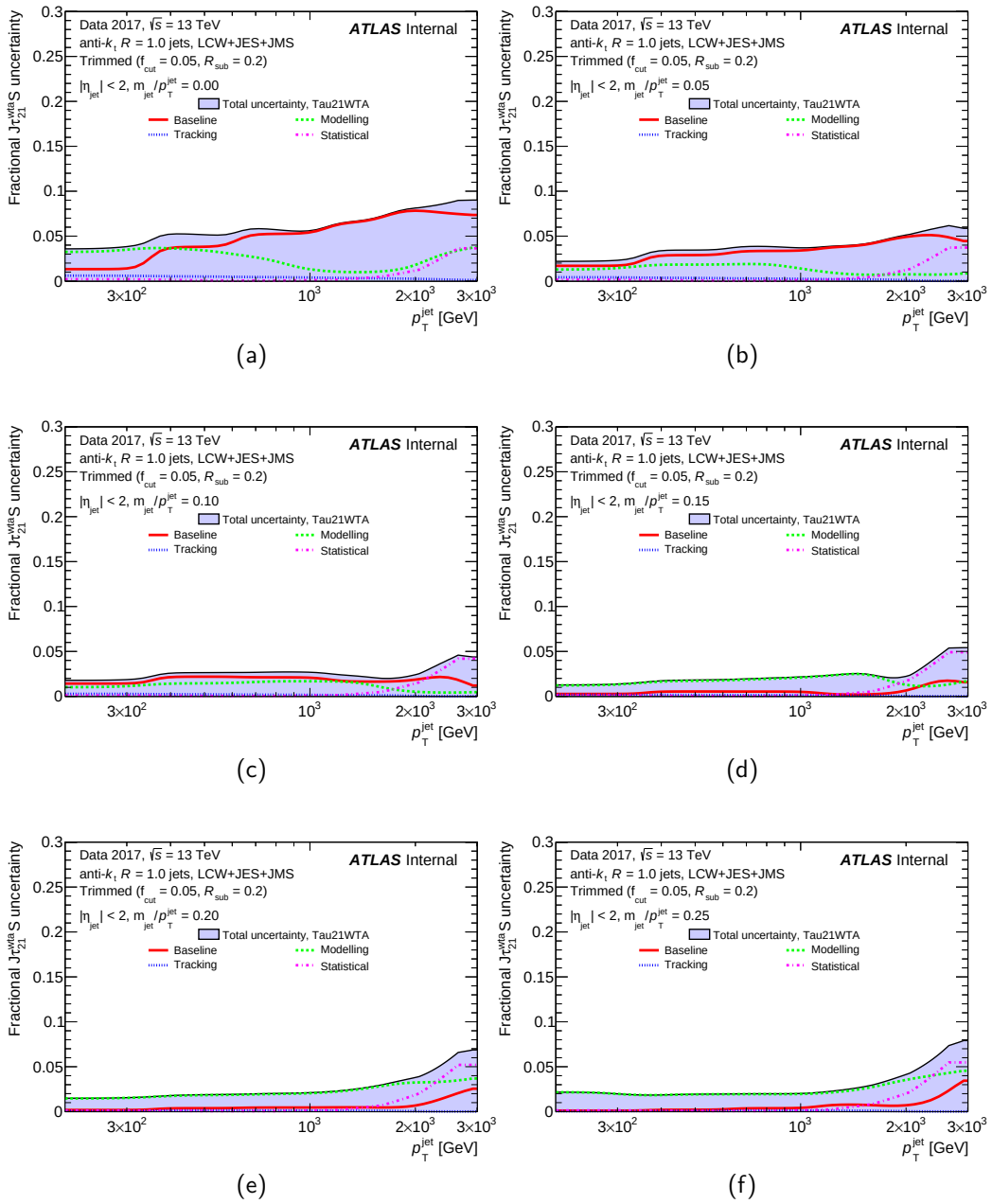


Figure D.6: Figure D.6a is showing the total uncertainty distribution for TAU21_WTA derived at $\frac{m_{jet}}{p_T^{jet}}$=0.0, fig D.6b is for $\frac{m_{jet}}{p_T^{jet}}$=0.05, fig D.6c is for $\frac{m_{jet}}{p_T^{jet}}$=0.10, fig D.6d is for $\frac{m_{jet}}{p_T^{jet}}$=0.15, fig D.6e is for $\frac{m_{jet}}{p_T^{jet}}$=0.2, and fig D.6f is for $\frac{m_{jet}}{p_T^{jet}}$=0.25 using 2017 data and MC16d samples.

APPENDIX E

Low level flavor taggers

E.1 Impact parameter based taggers: IP

Low-level flavor taggers used in ATLAS are based on the reconstructed track jets using the ID in the region of acceptance $|\eta| < 2.5$. The insertable B-Layer which introduced in RUN 2, has improved the vertex reconstruction and track extrapolation which in result, enhanced the b-tagging performance in the low-to-medium jet p_T region. Following track-based taggers are commonly used in b-tagging such as,

- IP2D
- IP3D

The IP2D and IP3D are impact parameter based taggers which use a Log Likelihood Ratio to discriminate whether the tracks associated to the jets [1] are compatible with the primary vertex hypothesis or not. Two important parameters of,

- d_0 : transverse impact paramter
- $z_0 \sin\theta$: longitudinal impact parameter

are defining the IP2D and IP3D taggers, where the d_0 is the closest distance from the primary vertex to the track and $z_0 \sin\theta$ is the closest distance among the track and the primary vertex in the longitudinal plane in the (r, ϕ) space. The IP2D tagger uses the transverse impact parameter significance $(\frac{d_0}{\sigma_{d_0}})$ and IP3D uses both the longitudinal impact paramter signifiance $(\frac{z_0 \sin\theta}{\sigma_{z_0 \sin\theta}})$ and the transverse impact paramter as the discriminating variable.

Tracks originating from the b hadrons have larger impact parameters as the b-hadrons lifetime is longer than c quark and the lighter hadrons. This gives rise to a secondary vertex which is clearly separated from the primary vertex. A sign can, therefore, be assigned based on the location of the secondary vertex if it is ahead or behind to the from the primary vertex, relative to the direction of the jet. [1, 6, 37]

[1] Tracks and jets are associated via spatial matching ΔR in the (η, ϕ) space $\Delta R = \sqrt{(\eta_{jet} - \eta_{track})^2 + (\phi_{jet} - \phi_{track})^2}$. ΔR is 0.25 for jet p_T around 20GeV and 0.25 for jets with jet p_T around 150 GeV [1]

E.2 Secondary vertex finding algorithm: SV

Another low-level tagger used in b-tagging is the secondary vertex finding algorithm. It is reconstructing an inclusive displaced secondary vertex within the jet. The SV algorithm search for all the track pairs for the two-track hypothesis and reject any of them if they were found to be originated from other long-live particles such as K_s or Λ. If any track pair passes, a new vertex is fitted with all the tracks from the accepted two-track vertices and any outlier will be removed iteratively. [1, 6]

E.3 Decay Chain Multi-vertex Algorithm: JetFitter

The Jetfitter algorithm tries to reconstruct a full b-hadron decay chain using the topological structure of weak b- and c-hadron decays in the jet. A common line which the primary vertex and the b and c vertices lie is determined using a Kalman filter which also finds the flight path of the b-hadron and the position of the b- and c-hadrons. [66]

VITA

Madhuranga Thilakasiri Madugoda Ralalage Don

Candidate for the Degree of

Doctor of Philosophy

Dissertation: SEARCH FOR SINGLE VECTOR-LIKE T QUARK DECAYING INTO ALL HADRONIC FINAL STATES WITH THE ATLAS DETECTOR AND UPDATED FLAVOR TAGGING PARAMETERIZATIONS AND PIXEL SIZE CHARACTERIZATION STUDIES FOR THE HL-LHC UPGRADE

Major Field: Experimental High Energy Physics

Biographical:

Education:

Earned Bachelor of Science in Physics with first class honors at Univeristy of Sri Jayewardenepura, Sri Lanka in 2010.

Completed the requirements for the degree of Doctor of Philosophy with a major in Experimental High energy Physics at Oklahoma State University in July 2020.

Experience:

Graduate Research/Teaching Assistant at Oklahoma State University in Stillwater Oklahoma from 2013 to 2020.

Research Scientist at Industrial Technology Institute in Sri Lanka from 2012 to 2013.

Teaching Assistant at University of Sri Jayewardenepura in Nugegoda, Sri Lanka from 2010 to 2011.

Professional Affiliations:

Member of Amercian Physical Society