Search for single production of vector-like quarks decaying into a Wb final state at $\sqrt{s} = 13$ TeV with the ATLAS detector

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Anjishnu Bandyopadhyay
aus
Calcutta, India

Bonn, 26.10.2023

Dieser Forschungsbericht wurde als Dissertation von der
Mathematisch-Naturwissenschaftlichen Fakultät der Universität Bonn angenommen und ist
auf dem Hochschulschriftenserver der ULB Bonn
https://nbn-resolving.org/html/urn:nbn:de:hbz:5-77461 elektronisch
publiziert.

Gutachter und Betreuer:	Prof. Dr. Ian C. Brock
Gutachter:	Prof. Dr. Jochen C. Dingfelder
Tag der Promotion:	14.03.2024
Erscheinungsjahr:	2024

For my grandmother

Acknowledgements

This journey has been a long one. And not a single step of this journey would have been possible without Ian. The support, motivation, inspiration, and academic freedom he provided me is unparalled. To add to that, there were the cricket discussions and banters. I am grateful for the work place atmosphere he created in our group. One thing Ian taught me is how to be optimistic in the face of all adversities.

Beside the obvious, I am thankful to my mother for inculcating my reading habit. Without that nothing would have been possible. I thank my father for being the best first physics teacher one can imagine. I am lucky to have a brother who reminds me every day that humour is the best way of facing everything in life. I am indebted to my grandmother for giving me a second life.

Ankan, thanks for being the perennial safe space. The daily calls made life much better. Sujit, hoping you can keep on challenging me in an unfiltered manner forever. Uddhipan, thanks for all the discussions on what could have been. Tathagata, thanks for teaching the value of a rolling humanities education. Debarati, thanks for reminding me the difference between statistics and probability. Sooraj, thanks for all the philosophical discussions. Arshia, thanks for all our tea sessions. Ripunjay, thanks for the banter and being the perfect hospital partner. Shauri, thanks for the music and everything else.

I want to thank all the colleagues I had during my time at the group. Sebastian, for setting the standards of programming, Jan, for helping me become an IT person, Rui, for all the physics and statistics discussions and being a great Jamboree and DPG room mate, Ozan, for all the discussions on life and politics we had at the office and during our drinks, Irina, for all our tennis chats, and Regina for being the best leader we had, Christian and Rico, for all our dark jokes at the office, Philipp, for the CERN days and the crazy laughs. One person I can't thank enough is Rosi. She provided me the love and warmth that is rare in today's world.

I want to acknowledge the support I received from my colleagues at Humboldt University especially Dustin, Janet, Ferdi, and Heiko. I would like to thank the ATLAS members with whom I worked with. I would like to thank Andrea and Frau Streich for all the support I got during my PI days.

I would like to express my gratitude all the GEDS members that made my CERN days unforgettable. Lastly, a big thanks to all the artists whose works I ingested over these years.

And if I am forgetting someone, thank you for your help and support.

Contents

CHAPTER 1

Prologue

The ΛCDM model of cosmology in combination with the Standard Model of particle physics (SM) provides the best understanding of the universe at both galactic and sub-atomic scales. The gravitational force is explained by Einstein's theory of relativity, which is the basis of the ΛCDM model. The measurements from the Planck Collaboration [1] show that the universe comprises of about 70% dark energy, 25% dark matter, and 5% ordinary matter. The Standard Model of particle physics is a theory that gives a description of the fundamental building blocks of ordinary matter and the interactions among them namely the electroweak and the strong forces. The predictions of this theory have been tested and proven with high precision. A theory that only describes 5% of the universe may seem inconsequential, but the Standard Model can provide an explanation for the mechanism that makes the stars shine, all 1×10^{24} of them. Till date, there exists very limited understanding of both dark matter and dark energy. Dark matter is said to interact using gravitational forces but not electromagnetic forces. Dark energy is the form of energy that is driving the accelerated expansion of the universe.

The discovery of the Higgs boson in 2012 by both the ATLAS and CMS collaborations [2, 3] using high energy proton-proton collisions at the Large Hadron Collider (LHC) ensured that all particles predicted by the SM were experimentally discovered. In the subsequent years, the Higgs boson and other SM particles have been studied in great detail using the data collected during the Run 1 and Run 2 of the LHC. The most precise measurement of the Higgs boson mass is (125.11 ± 0.11) GeV which was performed by the ATLAS collaboration [4]. On the contrary, the quantum corrections for the predicted Higgs mass due to the interactions of the Higgs boson and other SM particles are proportional to the next scale of new physics, which is the Planck scale (1×10^{19} GeV) according to the current theories. Thus, there exists a significant discrepancy among the predicted and measured value of the Higgs boson mass. This problem is called the Higgs mass hierarchy problem. Theorists have proposed numerous theories that solve this problem by cancelling these corrections arising from the interactions of new particles predicted in these theories and the Higgs boson.

The high centre-of-mass energies of the proton-proton collisions at the LHC were the

perfect toolkit to probe the validity of these Beyond Standard Model (BSM) theories. The 13 TeV centre-of-mass energy used in the Run 2 of the LHC from 2015 to 2018 provided an opportunity to study unexplored high-energy domains. The new physics program of the LHC mainly comprised of experimental searches of hypothesised particles using the collision data gathered by the ATLAS and CMS detectors. Predictions from theories like Supersymmetry [5, 6], Composite Higgs models [7, 8], Little Higgs models [9–11] were extensively tested.

Premise and organisation of thesis

One such prediction of various BSM theories is the existence of particles called Vector-like quarks [12–14], whose properties slightly differ from Standard Model quarks. Unlike SM particles, these particles' mass does not arise due to the Higgs mechanism. This ensures that Higgs measurements do not constrain the predictions concerning vector-like quarks. This thesis uses Run 2 LHC collision data collected by the ATLAS detector to search for the existence of vector-like quarks. These particles can be produced both via single production and pair production. In the phenomenological predictions, the single production has a higher production cross-section compared to pair production at higher energies. Thus, this thesis focusses on the single production search. There are several types of vector-like quarks that are hypothesised in theories. Vector-like quarks predominantly couple with third generation SM quarks in association with SM bosons. Among the various decay modes, the decay mode involving a W-boson and a b-quark was studied in this thesis. This choice was motivated by the relative high branching fraction of vector-like quarks to this decay mode. While studying decay modes in LHC collisions, only the the final decay mode can be directly accessed. For example, rather than measuring the W-boson, the decay products of a W-boson are used. For this thesis, both the leptonic and hadronic decay modes were probed. The leptonic decay mode was studied using the 2015-2016 collision dataset whereas the hadronic decay mode was investigated using the full Run 2 collision dataset collected from 2015-2018. The search for vector-like quarks performed in this thesis using the Wb hadronic decay mode is the first of its kind.

This thesis is divided into three parts. The Part I introduces all general concepts, tools, and methodologies that have been used for this thesis. The Part II details the analysis and results of the search using the leptonic decay mode with 2015-2016 data. The work for this part resulted in a publication by the ATLAS collaboration [15]. The Part III describes the analysis and results of the search using the hadronic decay mode with 2015-2018 data. This work is currently in review by the ATLAS collaboration for an upcoming publication.

Part I

General concepts

CHAPTER 2

Theoretical concepts

In this chapter, the key theoretical aspects of particle physics will be introduced. After introducing the basic theoretical terminology that will be used in this thesis, the Standard Model of particle physics (SM) will be discussed in detail. The SM is the best theory to date which can describe the electroweak and the strong forces. Although the SM has been verified to extreme precision, it still has certain drawbacks. These shortcomings of the SM will be described in the following section of this chapter. Then, various Beyond Standard Model theories that mitigate these shortcomings will be presented. Finally, the phenomenology and previous experimental searches of vector-like quarks will be discussed

2.1 Basic terminology in particle physics

Natural units

In the scientific field of particle physics, rather than using the SI units, a system of units called natural units is used. Natural units are units defined in terms of physical constants. In this system of units, the Planck's constant $\hbar$, the gravitational constant G, the speed of light in vacuum c, and the Boltzmann constants k_B are taken as unity i.e. $\hbar = G = c = k_B = 1$. Electronvolt (eV) is the fundamental unit in this system of units. Since most physical processes considered in this thesis occur at high energy, observables like mass and energy will be expressed in units of MeV, GeV, or TeV.

Cross-section and Luminosity

Scattering experiments are mostly used in particle physics to test the various predictions of theoretical models. The cross-section, σ, of a scattering process defined as the effective area where the process occurs is a physical observable that is frequently used to test theoretical models. The cross-section is a measure of the probability of the occurrence of the process

that is being observed. It can be calculated theoretically as

$$\sigma \propto \int |M|^2 \, d\rho, \tag{2.1}$$

where M is the matrix element of the process and $\int d\rho$ is the integral over the whole phase space. The matrix element is a measure of the amplitude of the transition from the initial to the final state of the scattering process that is being observed. The cross-section σ can be denoted in terms of the rate of observed events N_{obs}

$$\frac{dN_{\text{obs}}}{dt} = \sigma \cdot \mathscr{L} \tag{2.2}$$

$$\implies N_{\text{obs}} = \sigma \int \mathscr{L} \, dt, \tag{2.3}$$

where $\mathscr{L}$ is called the instantaneous luminosity and $\int \mathscr{L} dt$ is the integrated luminosity. The unit of cross-section is barn ($1\,\text{b} = 1 \times 10^{-24}\,\text{cm}^2$). Thus, instantaneous luminosity can be expressed in units of $\text{cm}^{-2}\text{s}^{-1}$ and integrated luminosity in units of b^{-1}. In a collider experiment, where two beams with n_1 and n_2 particles each collide in bunches with frequency f, the luminosity can be written as

$$\mathscr{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y}. \tag{2.4}$$

Here, σ_x and σ_y characterise the root mean square transverse beam sizes in the horizontal and vertical directions respectively [6].

Decay width and branching fraction

In scattering experiments, the final state particles generally decay into lighter particles. If such a particle decays with a lifetime τ, then the decay width Γ is defined as

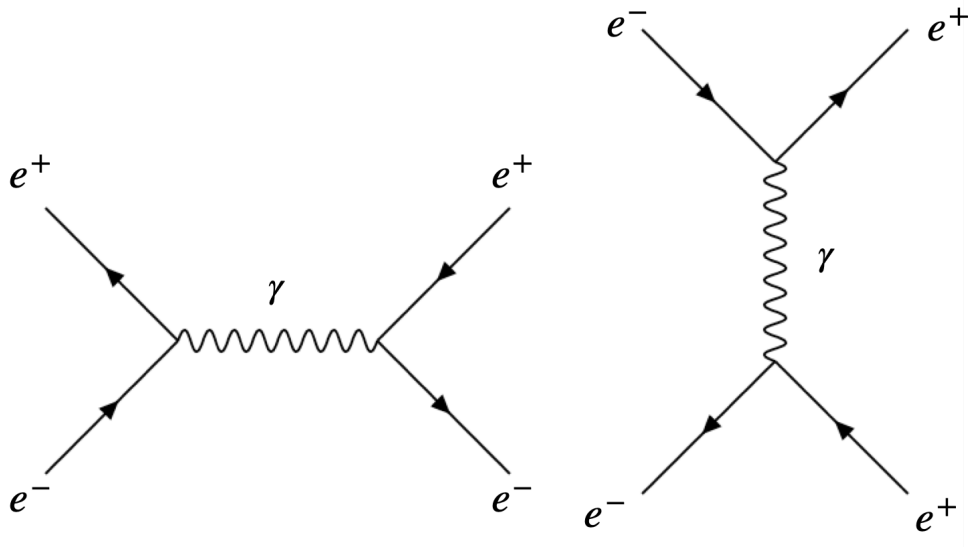
$$\Gamma = \frac{1}{\tau} \tag{2.5}$$

in natural units. Most particles decay via more than one decay mode. Suppose a particle has i decay modes with individual decay widths Γ_i and total decay width Γ_{total} then the branching fraction $\mathscr{B}_i$ for the i-th decay mode is defined as

$$\mathscr{B}_i = \frac{\Gamma_i}{\Gamma_{\text{total}}} \tag{2.6}$$

Feynman diagrams

Richard Feynman formulated an elegant method to represent any scattering process and calculate the corresponding matrix element M using perturbative expansion [16]. It is a way of computing a physical observable by expanding the given observable in a power series of the coupling strength associated with the scattering process. Feynman diagrams are space-time diagrams that illustrate a scattering process and the corresponding matrix element M can be calculated with the help of these diagrams using certain rules. These rules namely Feynman rules, dictate the scheme that needs to be used to compute the analytical expression for M, given a unique Lagrangian density. To illustrate this idea in detail, the Feynman diagrams of Bhabha (electron-positron) scattering will be used.The two Feynman diagrams for Bhabha scattering are shown in Fig. 2.1. In the convention used in this thesis, time goes from left to right in all Feynman diagrams.



(a) s-channel diagram (b) t-channel diagram

Figure 2.1: Feynman diagrams for s-channel and t-channel processes.

Here, incoming particles (anti-particles) are denoted by incoming (outgoing) arrows. At each vertex, all conservation laws are preserved. In Fig. 2.1(a), a "time-like" virtual photon is exchanged during the e^-e^+ scattering, whereas in Fig. 2.1(b) a "space-like" virtual photon is exchanged. The Feynman diagram in Fig. 2.1(a) is called s-channel ($s = (p_{e^-} + p_{e^+})^2$) since $|M| \propto 1/s$, where s is the centre of mass energy of the system. For Fig. 2.1(b), $|M| \propto 1/t$, where $t = (p_{e^-}^{\text{incoming}} - p_{e^-}^{\text{outgoing}})^2$. In Feynman diagrams, each vertex contributes one order of electric charge to M. For Bhabha scattering, that charge is the electric charge e of $e^-$and e^+. Thus, both the diagrams in Fig. 2.1, $|M| \propto \alpha_{\text{EM}}$, where the electromagnetic constant $\alpha_{\text{EM}} = \frac{e^2}{4\pi}$,. In perturbative expansion, such diagrams are called "tree-level" or "leading-order" (LO) diagrams. Higher order terms in the expansion are called next-to-

leading order (NLO), next-to-next-to-leading order (NNLO). and so on. For NLO and NNLO, the number of diagrams increase by a significant amount and $|M|$ is proportional to higher orders of α_{EM}. As $\alpha_{\mathrm{EM}} << 1$, NLO and NNLO contributions are smaller than LO contributions. A NLO diagram depicting loop-corrections to Bhabha scattering is shown in Fig. 2.2.

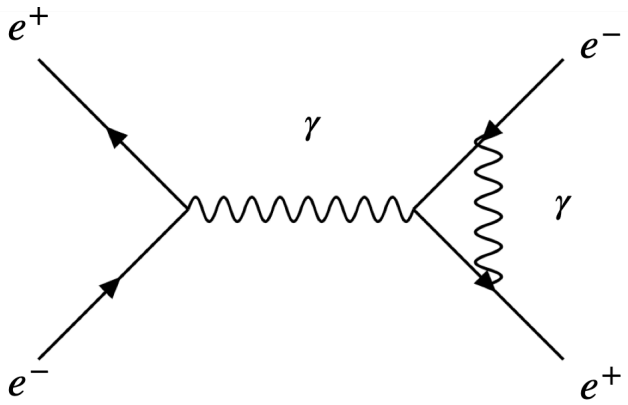


Figure 2.2: A NLO diagrams for the Bhabha Scattering.

2.2 Standard Model of particle physics

The Standard Model (SM) of particle physics is the most accurate and precise theoretical description of the known fundamental particles and their interactions. The content of the SM can be classified into two broad categories, fermions and bosons. Fermions are spin-1/2 particles governed by the Fermi-Dirac statistics and are the fundamental building blocks of ordinary matter. Bosons are integer-spin articles governed by the Bose-Einstein statistics which describe the interactions among the fermions and themselves. Fermions can be further sub-divided into quarks and leptons, whereas the bosons can be classified into gauge bosons and the Higgs boson. The particle content of the SM is shown in Fig. 2.3. Both the leptons and quarks have corresponding anti-particles with opposite quantum numbers. The neutral bosons are their own anti-particles. The W^--boson is the anti-particle of W^+-boson. The lines represent the allowed interactions among the various particles. The details of the SM will be discussed in detail in the following sections.

Lagrangian of the SM

The SM is a combination of quantum field theories (QFT) that provides a complete theoretical description of the electroweak and strong interactions. The type of quantum field theories that constitute the SM are called gauge theories. These theories have certain symmetries, namely gauge symmetries. A simple example of a gauge symmetry is the symmetry of A_μ

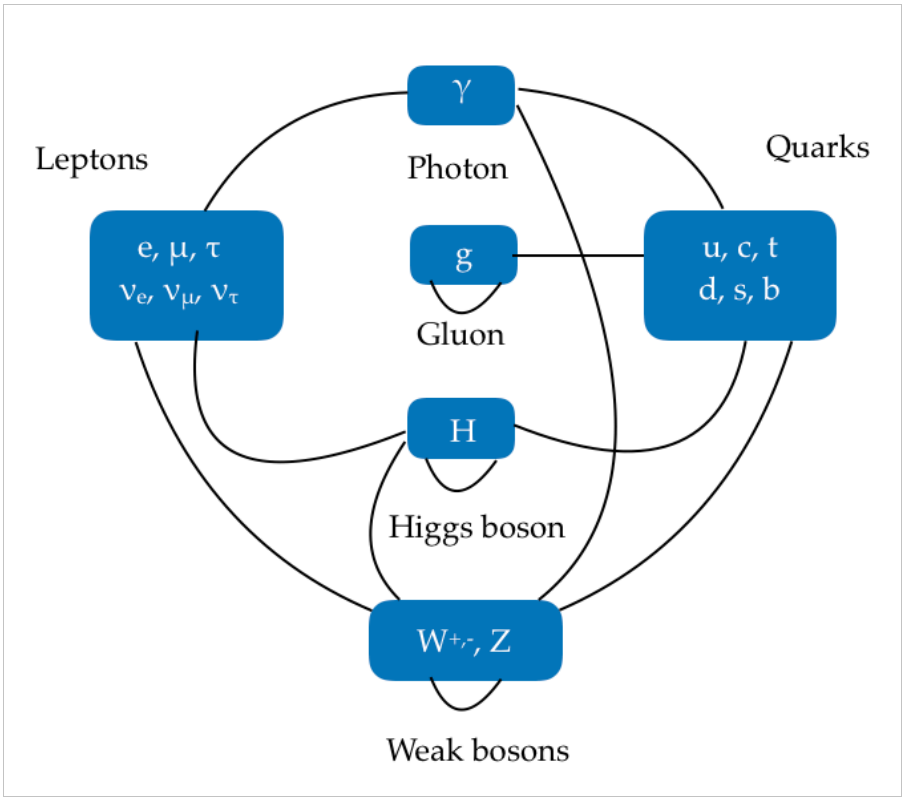


Figure 2.3: A representation of the different particles and allowed interactions of the Standard Model.

in Maxwell's equations. These symmetries imply that a given QFT is invariant under certain mathematical transformations called gauge transformations. Invariance under gauge transformations ensures unitarity is preserved in a QFT. As a result of gauge invariance, the SM can be embedded in a Lie group $SU(3) \times SU(2) \times U(1)$. In mathematical terminology, the fermions are the irreducible representations of this group and the bosons are its generators. Like any other physical theory, a QFT can be described entirely by its Lagrangian, L. Since for a QFT, the fields are the fundamental quantities, and the space-time coordinates are mere labels, the Lagrangian density $\mathcal{L}$ is used ($L = \int dt d^3x \mathcal{L}$). The SM Lagrangian $\mathcal{L}$[1] can be written as [17]

$$\mathcal{L} = \mathcal{L}_{\text{bosons}} + \mathcal{L}_{\text{fermions}} + \mathcal{L}_{\text{Yukawa}} + \mathcal{L}_{\text{Higgs}} \tag{2.7}$$

$$= \left(-\frac{1}{4}F^a_{\mu\nu}F^{a\mu v}\right) + \left(i\bar{\psi}^i\gamma^\mu D_\mu \psi^i\right) + \left(\bar{\psi}^i_L V_{ij}\Phi\psi^j_R + h.c.\right) + \left(\left|D_\mu\Phi\right|^2 - V(\Phi)\right). \tag{2.8}$$

[1] From here, Lagrangian will mean Lagrangian density if not specified otherwise.

Here, $\mathcal{L}_{\text{bosons}}$ and $\mathcal{L}_{\text{fermions}}$ denote the interactions among the fermions and the gauge bosons. $F^a_{\mu\nu}$ is the stress-energy tensor of the gauge bosons and ψ^i is the spinor of the i-th fermion. The Yukawa term, $\mathcal{L}_{\text{Yukawa}}$, is the mass-term for the various particles which arises due to the Higgs mechanism. The Higgs term is responsible for the various SM particles achieving mass. The Higgs term also quantifies the self-interactions of the Higgs field.

2.2.1 Electroweak interactions

The theory of electroweak interactions is a unification of the electromagnetic and weak forces, first proposed by Glashow, Salam, and Weinberg [18–20]. This theory of quantisation of the electromagnetic force known as Quantum Electrodynamics (QED) was first developed in the 1940s be Feynman, Schwinger, and Tomonoga [21–23]. QED was developed as a gauge theory with the generator being the photon. The predictions for Bhabha scattering, as shown in Fig. 2.1, can be computed using QED. Weak interactions describe physical processes like the β-decay. The theory of weak interactions was first developed by Fermi [24] as a 4-point theory to describe the β-decay. In the 1950s, Wu discovered parity violation in weak interactions in a β-decay experiment involving the decay of unstable cobalt-60 to stable nickel-60 [25]. Subsequently, to incorporate parity violation, Feynman, Gell-Mann [26] and Sudarshan, Marshak [27] formulated the V-A theory of weak interactions.

The Glashow-Salam-Weinberg model successfully unified the V-A theory and QED. This new theory can be embedded into a $SU(2)_L \times U(1)_Y$ group. The $SU(2)$ symmetry transformations act on the isospin of the left handed fermions indicated by the subscript L. The $U(1)$ symmetry acts on the hypercharge (Y), a linear combination of electric charge and the third component of the isospin, irrespective of the chirality of the fermions. The left-handed leptons (ℓ_L) are grouped into $SU(2)$ doublets, and the right handed leptons ℓ_R are represented as singlets.

$$\ell_L = \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$$
$$\ell_R = e_R, \mu_R, \tau_R$$

Here, $e, \mu,$ and τ denote the electron, muon, and tau lepton respectively. The SM does not allow the presence of right-handed neutrinos. In electroweak theory, neutrinos are massless. The symmetry group $SU(2)_L \times U(1)_Y$ has four generators, and thus four-vector fields $W^{1,2,3}_\mu, B_\mu$ describe them. The gauge bosons of the elctroweak theory, $W^\pm$, Z, and the

photon (A) can be written as a linear combination of the four fields, $W_\mu^{1,2,3}$ and B_μ.

$$\begin{aligned} W_\mu^\pm &= \sqrt{\frac{1}{2}}\left(W_\mu^1 \mp i W_\mu^2\right) & \tan\theta_W &= \frac{g'}{g} \\ A_\mu &= B_\mu \cos\theta_W + W_\mu^3 \cos\theta_W & e &= g\sin\theta_W = g'\cos\theta_W \\ Z_\mu &= -B_\mu \cos\theta_W + W_\mu^3 \cos\theta_W & \cos\theta_W &= \frac{m_W}{m_Z} \end{aligned} \tag{2.9}$$

Here, θ_W is the Weinberg angle, the tangent of which quantifies the ratio of the $U(1)_Y$ hypercharge coupling g and $SU(2)_L$ isospin coupling g'. The Weinberg angle also relates g, g', and the electric charge e. Since the elctroweak force violates parity conservation, only the left-handed leptons ℓ_L interact with $W^\pm$ via charged interactions, but both ℓ_L and ℓ_R interact with both the Z and the photon via neutral interactions. The first experimental evidence of electroweak theory was the discovery of weak neutral current at CERN in 1973 using the Gargamelle bubble chamber [28]. The W and the Z bosons were discovered later in 1983 at the S$p\bar{p}$S at CERN by the UA1 and the UA2 collaborations [29–32].

For the quark sector, the electroweak interactions work similarly with the quarks being embedded in $SU(2)$ doublets.

$$Q = \begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix}$$

These doublets comprise the flavour eigenstates. Since all quarks are massive,[2] the mass and flavour eigenstates are not identical under the electroweak theory. For the up-type quarks the mass and flavour eigenstates are congruent but for down-type quarks, the mass eigenstates are related to the flavour eigenstates via the Cabibbo–Kobayashi–Maskawa (CKM) matrix [33, 34] in the following manner

$$\begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} \tag{2.10}$$

The values of the CKM matrix have been determined experimentally with the diagonal having values close to unity [6].

2.2.2 Strong interactions

In the 1950s, a plethora of particles like pions, kaons, and baryons that interact via the strong force was discovered. In the early 1960s, Gell-Mann and Ne'eman independently proposed

[2] Neutrinos are massless in the SM and thus we do not have a separate mass eigenstate for leptons.

a $SU(3)$ symmetry to classify baryons and mesons in a "broken eightfold way" with the mesons being embedded in an octet and baryons in a ten-fold representation [35, 36]. This formulation predicted the existence of particles like the Ω baryon which was later discovered in 1964 [37]. Gell-Mann and Zweig independently theorised particles called "quarks" [38, 39].[3] Three quarks with spin $1/2$, namely up , down, and strange were predicted in the $SU(3)$ model with respective electric charges of $-1/3, 2/3,$ and $-1/3$. The apparent violation of Pauli exclusion principle in the discovery of the Δ^{+++} resonance [40] was resolved by the prediction of a new charge called "colour", which could be "red", "green", and "blue" [41, 42].

In the late 1960s, experiments at SLAC [43, 44] using electron-proton collisions indicated a hard-structure inside the proton. Feynman theorised that the proton was made up of charged "partons". Later experiments from both CERN and SLAC confirmed that these partons were indeed the quarks predicted by Gell-Mann. Finally, Quantum Chromodynamics (QCD) was proposed as a complete theory with $SU(3)_C$ (C denotes colour charge) gauge symmetry [45]. The eight massless generators of this symmetry called gluons were predicted to be the carrier of the strong force. The discovery of gluons at PETRA in 1979 [46–49] cemented QCD as the theory of the strong interactions.

Asymptotic freedom and confinement

Unlike photons, gluons self-interact. This property of gluons leads to the fact that the strong force becomes asymptotically weak at increasing energies or decreasing lengths. Asymptotic freedom [50, 51] in QCD happens because in vacuum the "anti-screening" among the gluons exceed the "screening" among the quarks. Thus, the strong coupling constant α_s evolves with energy (Q) as

$$\alpha_s(Q) \propto \frac{1}{\ln\left(\frac{Q}{\Lambda}\right)}. \tag{2.11}$$

This equation is called the Renormalisation Group Equation (RGE). In perturbation theory, any observed physical quantity can be written as a power series of the coupling constant e.g. α_s. In gauge theories like QCD, the power series includes loop divergences. These divergences are done away with by rewriting the theory in terms of "renormalised" physical quantities like the mass or the charge as a function of energy scales. Since a measured quantity like mass or charge cannot change as a function of energy (Q), the energy dependence is absorbed into the coupling α_s via the RGE. The running of α_s has been verified experimentally as shown in Fig. 2.4. From Eq. (2.11), it follows that at length scales less than $1/\Lambda$, where Λ is the QCD scale parameter, quarks do not couple strongly. Experiments have shown that $1/\Lambda \approx 1$ fm, the dimension of light hadrons. At length scales greater than $1/\Lambda$, quarks couple strongly and cannot exist in "bare" form. They form bound states such as mesons (quark anti-quark pair) and baryons (three quarks or anti-quarks). This feature of

[3] Zweig had named the same particles as aces.

QCD is known as confinement. It should be noted that at low energies perturbative methods are not valid as α_s grows exponentially (as shown in Fig. 2.4). Non-perturbative methods like lattice QCD have been successful in predicting experimental results such as masses of light hadrons [52] at high precision.

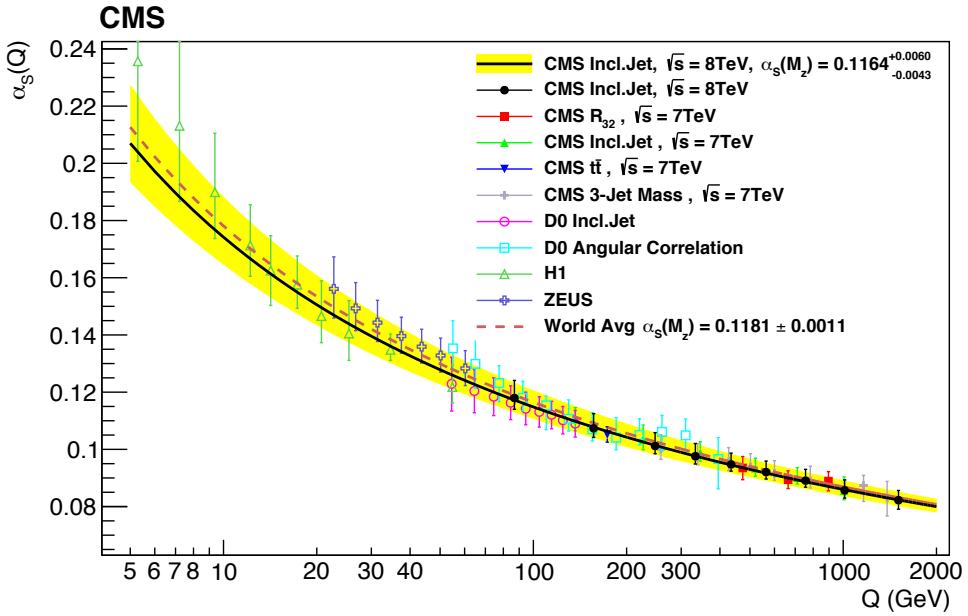


Figure 2.4: The running of α_s as a function of energy (Q) as measured by various experiments [53].

Parton density functions

The LHC collides bunches of protons at very high energies. When two protons collide with each other in a scattering process $pp \to X$, due to asymptotic freedom, this process can be thought as an effective collision of the partons in each of the protons. The partons can be the valence quarks (uud), the gluons, or the "sea" of quark anti-quark pairs of non-valence quarks created from the gluons. The cross-section of the process $\sigma_{pp\to X}$ can be factorised [54] in terms of individual partonic cross-sections $\sigma_{ij\to X}$ as

$$\sigma_{pp\to X} = \sum_{i,j} \int dx_i dx_j f_i\left(x_i, \mu_F^2\right) f_j\left(x_j, \mu_F^2\right) \hat{\sigma}_{ij\to X}\left(x_i, x_j, \mu_R^2, \mu_F^2\right), \tag{2.12}$$

where the two protons have i and j partons with momentum fractions $x_{i/j} = \frac{p_{\text{proton}}}{p_{i/j-\text{th parton}}}$ respectively. Here, $f_i\left(x_i, \mu_F^2\right)$ is the parton density function (PDF) of the i-th parton. It is measure of the probability of finding the i-th parton inside a given proton with momentum fraction x_i of the total momentum of the proton at an energy scale μ_F^2. PDFs are measured experimentally by combining data collected at different colliders. Various sets of measurements

of PDFs like NNPDF2.3 [55] and CT10 NNLO [56] are used by the ATLAS collaboration. The measurements for PDFs are only possible at certain energy scales but they need to be extrapolated to other energies. The dependence of quark and gluon PDFs $f\left(x, \mu_F^2\right)$ on energy scales (μ_F^2) can be parameterised using the DGLAP equations [57–59]

$$\mu_F^2 \frac{\mathrm{d}f_a\left(x, \mu_F^2\right)}{\mathrm{d}\mu_F^2} = \sum_{b \in \{q,g\}} \int_x^1 \frac{\mathrm{d}z}{z} \frac{\alpha_s}{2\pi} \hat{P}_{ba}(z) f_b\left(x/z, \mu_F^2\right) \tag{2.13}$$

where the sum ranges over both quarks (q) and gluons (g), and P_{ab} describes the splitting from parton a to $b (a, b = q, g)$. $P(ab)$ can be itself written as power expansion in terms of the strong coupling constant α_s. The CT10 NNLO PDFs at 8 GeV and 85 GeV for the valence quarks, gluons, and the quark "sea" are shown in Fig. 2.5. These PDFs have been calculated using the 3-loop DGLAP equations [56].

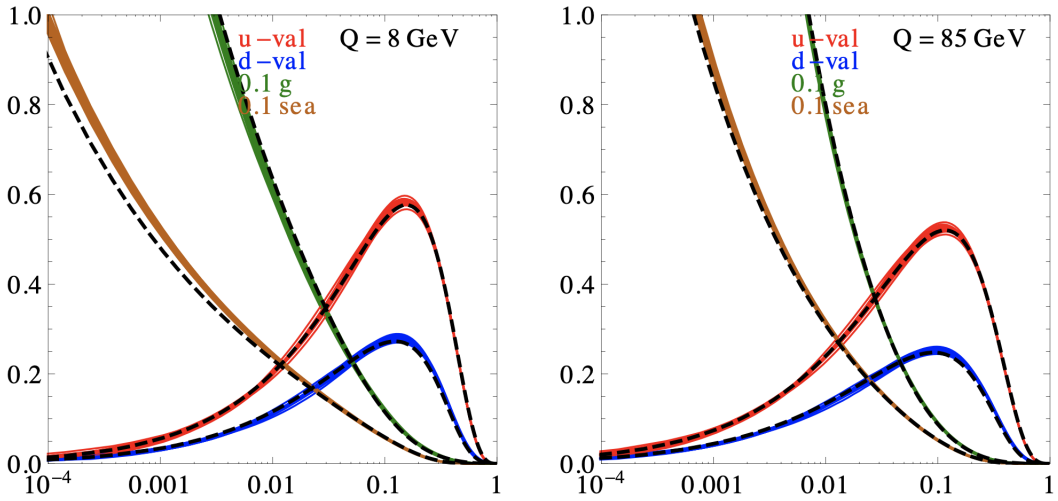


Figure 2.5: The CT10 NNLO PDFs for the valence quarks, gluons, and the quark "sea" at 8 GeV and 85 GeV respectively. The sea and gluon PDFs are scaled by a factor of 0.1. Here the sea is comprised of $2(\bar{u} + \bar{d} + \bar{s})$ [56].

Parton shower and hadronisation

The physical implication of the DGLAP equations is that if a parton a is resolved in the parent hadron at μ_F^2, then it may have been produced via the splitting of parton b at $\mu_F^2 + d\mu_F^2$ [60]. Thus, the hard-scattering of two protons, as performed at the LHC, is accompanied by extra partons which further radiate gluons. These gluons can subsequently decay into quark anti-quark pairs or radiate more gluons. This forms a parton shower. But due to the confinement property of QCD, these quarks hadronise above the QCD scale Λ. This whole process of parton showering and hadronisation is simulated using Monte Carlo methods. A detailed

mathematical description of the showering and hadronisation is shown in Ref. [60], whereas a schematic representation of this process is depicted in Fig. 2.6.

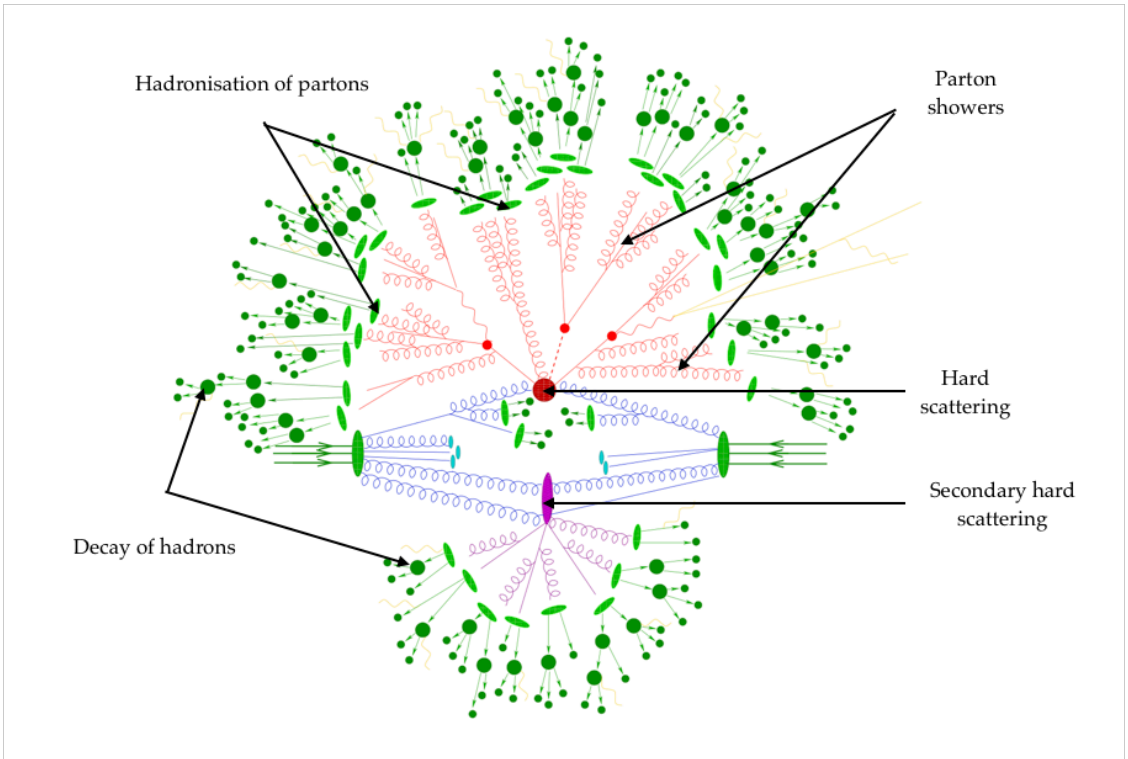


Figure 2.6: A schematic representation of a proton-proton collision as modelled by a Monte Carlo generator [60].

2.2.3 Higgs mechanism

The electroweak theory and QCD Lagrangians do not contain any mass-terms. The weak gauge bosons and the fermions achieve mass in the SM through a mechanism called the Brout-Englert-Higgs mechanism [61, 62]. The Brout-Englert-Higgs mechanism is a way of spontaneously breaking the gauge symmetries by introducing a new field Φ, the Higgs field. To achieve this, an extra term is introduced in the SM Lagrangian $\mathscr{L}_{\text{Higgs}}$ (the fourth term in Section 2.2).

$$\mathscr{L}_{\text{Higgs}} = \left|D_\mu \Phi\right|^2 - V(\Phi) \tag{2.14}$$

$$= \left|D_\mu \Phi\right|^2 - \left(-\mu^2 \Phi^\dagger \Phi + \lambda \left(\Phi^\dagger \Phi\right)^2\right), \tag{2.15}$$

where the Higgs field Φ is a complex doublet. Evidently, the potential $V(\phi)$ has a minima at $\{\Phi\} = 0$ for $\mu^2 < 0$ but for $\mu^2 > 0$ the potential acquires a vacuum expectation value (vev)

$$\{\phi\} = \sqrt{\frac{\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}} \tag{2.16}$$

The ground state Φ_0 can be written as following

$$\Phi_0 = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.17}$$

and this ground state spontaneously breaks the $SU(2)$ symmetry of the potential. Subsequently, according to perturbation theory, the Lagrangian can be expanded around the minima using the substitution $v \to v + h(x)$, where $h(x)$ is a neutral field. This gives rise to mass terms for the electroweak gauge bosons $W^{\pm}$ and Z, and a mass-term for the field h. This neutral field is called the Higgs field and the boson for this field is called the Higgs boson. In addition, this expansion also leads to trilinear (hhh) and quartic Higgs ($hhhh$) self-couplings and couplings of the Higgs boson to the gauge bosons V in the form of $hhVV$ and hVV. The masses can be written in terms of the vev v as

$$m_W = \frac{1}{2} g v, \; m_Z = \frac{1}{2}\sqrt{g^2 + g'^2}\, v, \; m_h = \sqrt{2\lambda} v \tag{2.18}$$

Yukawa term

The Higgs field can also be used to generate mass terms for the fermions. This is achieved by introducing a gauge-invariant term in the SM Lagrangian called the Yukawa term $\mathscr{L}_{\text{Yukawa}}$ (the third term in Section 2.2) by combining the left and right-handed fermions and the Higgs doublet.

$$\mathscr{L}_{\text{Yukawa}} = \left(\bar{\psi}^i_L V_{ij} \Phi \psi^j_R + h.c. \right) \tag{2.19}$$

Replacing Φ by Φ_0 in Eq. (2.19) would yield masses for fermions $m_{l,q}$

$$m_{l,q} = \frac{1}{2} \lambda_{l,q} v \tag{2.20}$$

and yield fermion-Higgs interaction terms like $\bar{f} f h$. The matrix V_{ij} in Eq. (2.19) is synonymous to the CKM quark-mixing matrix introduced in Eq. (2.10). It should be noted that neither v or λ can be predicted theoretically and can only be derived from experimental measurements.

The Higgs boson was discovered by the ATLAS and CMS collaborations in 2012 with a mass of around 125 GeV [2, 3]. The current most precise mass measurements of the Higgs boson from ATLAS and CMS are (124.97 ± 0.24) GeV and (125.11 ± 0.11) GeV respectively [4,

63]. This would suggest $v \approx 246\,\text{GeV}$. Evidence of couplings of the Higgs boson to most of the fermions and the gauge bosons has been observed by both the ATLAS and CMS collaborations.

2.3 Physics of the W-boson

Since the W-boson is an important particle for this thesis, a short review of W-boson physics is presented in this section. The W-boson decays into a quark anti-quark pair ($q\bar{q}$) or a lepton neutrino pair ($l^{\pm}\nu_l(\bar{\nu}_l)$). The branching fraction of the hadronic $q\bar{q}$ decay is about 67% and the branching fraction of the leptonic decay is about 33% with equal branching fractions for the three leptonic flavours ($l = e, \mu, \tau$)) [6]. The discovery of the W-boson by the UA1 and UA2 collaborations [29, 30] was performed using the leptonic decays, especially $W \rightarrow e\nu_e$ and $W \rightarrow \mu\nu_\mu$, since the hadronic decay mode is accompanied by overwhelming hadronic background originating from hard parton scattering processes. Using further data, the UA2 collaboration measured the mass of the W-boson as $M_W = (80.49 \pm 0.43\,(\text{stat.}) \pm 0.24\,(\text{sys.}))\,\text{GeV}$ [64]. A study of the hadronic decays in the UA2 data verified the $W \rightarrow q\bar{q}$ decay mode [65]. The $W \rightarrow q\bar{q}$ decay depends on all the elements of the CKM matrix not involving the top-quark. The Large Electron-Positron collider (LEP) measured $|V_{cs}| = 0.989 \pm 0.14$ [66].

The e^-e^+ LEP collider data was crucial in studying the W-boson. With the centre-of-mass energy of e^-e^+ beams exceeding $2M_W$, the measurement of the pair-production of W^+W^- was performed at LEP. This allowed a window to probe the validity of SM predictions with great precision. The s-channel pair-production processes allowed the measurement of trilinear gauge couplings (TGC) like ZWW and γWW. Precision measurements from LEP proved the existence of TGCs, thus establishing the $SU(2)$ structure of the electroweak theory [66]. Since the longitudinal polarisation of the W-boson is a direct consequence of spontaneous symmetry breaking of the electroweak theory, the W-boson polarisation measurements by LEP were successful in establishing SM as a robust theory of electroweak interactions.

The LHC has also studied the W-boson with extreme precision. The ATLAS collaboration has performed the most precise single measurement of the W-mass with $M_W = (80.370 \pm 0.019)\,\text{GeV}$ [67]. The coupling of the W-boson to the Higgs boson was also observed with the discovery of the $H \rightarrow W^+W^-$ decay mode [68]. In pp collisions, W-bosons are produced mainly via the interaction of the quarks inside the proton. Consequently, the ratio of W^+ and W^- production cross-sections is measure of the difference of u and d valence quark PDFs, and thus a good test of QCD predictions. The ATLAS collaboration has measured this ratio to a precision of 1% [69].

The top-quark, the heaviest SM particle with a mass of around 173 GeV, is the only quark that does not hadronise due to its extremely short lifetime. The top-quark decays into a W-boson and a b-quark with approximately 99% branching fraction. The single production of top-quarks, which has been observed and studied at the LHC with high precision, was

used to measure the value of $|V_{tb}|$ through the tWb vertex [70]. The Wb final state is the focus of this thesis. Here, the Wb final state has been used to search for particles called vector-like quarks, which will be introduced later in this chapter.

2.4 Drawbacks of the Standard Model

The SM has been tested experimentally to a high degree of precision, and almost all experimental results agree with its theoretical predictions. Despite its success, the SM has certain drawbacks and limitations. In this section, some of those drawbacks are discussed in brief.

- **Hierarchy problem:** The quadratic corrections to the Higgs mass at one loop level arise from its interactions with fermions. The largest contribution to these corrections come from the top loop, as shown in Fig. 2.7. The high value of the top-Higgs Yukawa coupling λ_t is the reason for this. The quadratic correction originating from the Feynman diagram in Fig. 2.7 can be written as

$$\delta m_H^2 \propto -|\lambda_t|^2 \Lambda^2 + ... \tag{2.21}$$

  where Λ is the next scale of new physics. For the SM, Λ is the Planck scale ($\approx 1 \times 10^{19}$ GeV). These corrections are more than 30 orders of magnitude larger than the observed mass of the Higgs boson. This discrepancy is known as the Higgs mass hierarchy problem [71, 72]. These corrections need to be cancelled theoretically to solve this problem.

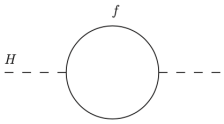


Figure 2.7: Loop corrections to the mass of the Higgs boson arising from top-quark

- **Matter anti-matter asymmetry:** The matter anti-matter asymmetry observed in the universe is $1 : 10^{-9}$. Although the electroweak force is CP violating, the amount of CP violation needed to achieve this asymmetry cannot be accommodated in the SM. Baryon number violation is a necessary condition for this asymmetry which is absent in the SM.

- **Dark matter and dark energy:** The SM is precise description of the particles and interactions present in ordinary matter. But ordinary matter only comprises 5% of the universe. Recent cosmological data [73] has shown that the universe is made of 27% dark matter and 68% dark energy. Relic densities of dark matter indicate that it should interact at the weak scale. But the SM is unable to describe the nature of either dark matter or dark energy.

- **Neutrino masses:** In the SM, neutrinos are massless. Experiments have proven that neutrinos have a finite mass. There is no mechanism in the SM to explain the finite mass of neutrinos.

2.5 Beyond the Standard Model physics

The theories and models that propose solutions to various shortcomings of the SM encompass the sub-field of **Beyond the Standard Model (BSM)** physics. In this section, the major BSM theories that solve those problems will be briefly reviewed.

2.5.1 Supersymmetry

Supersymmetry (SUSY) [5, 6] is an extension of the SM where new generalised symmetries turn a boson into a fermion and vice-versa. In SUSY, for every particle of the SM we have a super-partner hypothesised. If this symmetry was exact, a super-partner boson of the electron with a mass of 0.511 MeV would already been . Since this is not the case, SUSY is broken, and the masses of these super-partners are theorised to be in the TeV scale. SUSY also provides a framework where the electroweak and strong force couplings unite at energies around 10^{15} GeV, commonly known as the Grand Unified Theory (GUT) scale.

One of the important ramifications of SUSY is that it exactly solves the Higgs mass hierarchy problem. As shown above, the one loop corrections to the Higgs mass is proportional to the ultraviolet cutoff of the SM, which is the Planck scale $\approx 10^{19}$ GeV and the Yukawa coupling of the fermions. The top quark being the heaviest quark has the highest Yukawa coupling to the Higgs boson and thus contributes most to these corrections. In SUSY, the super-partner of the top quark, the loop corrections coming from the stop cancel the radiative corrections arising from the top-quark. This happens due to the reversal of sign in bosonic and fermionic loops arising from the spin-statistics theorem. In the minimal supersymmetric Standard Model (MSSM), there exists one super-partner for each SM particle and five Higgs bosons after electroweak symmetry breaking. The lightest supersymetric particle in the spectrum is a good candidate for weakly-interacting massive particle (WIMP) dark-matter. This is one of the most promising solutions to the dark-matter puzzle given the predicted WIMP dark-matter density is consistent with present-day cosmological observations. All the aforementioned features of SUSY has made it the forerunner of BSM theories. But unfortunately, no evidence of SUSY has been discovered till date. The current experimental bounds on SUSY arising from various ATLAS searches are presented in Ref. [74].

2.5.2 Composite Higgs models

Composite Higgs models (CHM) [7, 8] are a class of models which solve the Higgs mass hierarchy problem in a non-supersymmetric manner, first predicted in the 80s. The main idea behind CHM is that the Higgs boson is a composite state arising from new strongly interacting dynamics of a more fundamental theory at a higher energy scale. In such a

framework, the Higgs boson is a pseudo-Nambu Goldstone boson, the mass of which is generated dynamically by breaking a global symmetry of the aforementioned fundamental theory.[4] In CHM, the dynamic scale f defines the compositeness scale and the Higgs is elementary when $\xi = v^2/f^2 \to 0$, v being the Higgs VEV. This fact implies that BSM effects are only visible when ξ deviates from 0. In CHM, the quadratic sensitivity of the Higgs mass, is limited to composite scale f, which from LEP constraints turn out to be 5–10 TeV. Spin-1 resonances and top partners are still needed to cancel the quadratic corrections but here the fine tuning is in the order of 1% [12].

2.5.3 Little Higgs models

Little Higgs models (LHM) [9–11], predicted in the early 2000s, are models where the Higgs mass is shielded from one-loop quadratic divergences using new global symmetries. In LHM, the global symmetry is broken at the TeV scale and so the "litle Higgs" [10], associated with the SM Higgs, is the pseudo-Nambu Goldstone boson of this symmetry. This global symmetry is both spontaneously and explicitly broken "collectively" (A symmetry is broken "collectively" when two or more couplings in the Lagrangian are non-zero) to restore the mass of the "little Higgs". The "littlest Higgs" model achieves this by breaking a SU(5) symmetry to a SO(5) symmetry. This way of breaking the symmetries introduces new sectors which further needs the presence of a vector-like top partner to cancel the Yukawa quadratic divergences coming from the top quark. Here the loop cancellation is achieved by particles of same spin unlike SUSY. Since LHM require the vector-like top partner to be in the TeV scale, the LHC is the perfect test bed to prove or constrain these theories. To add to this elegant solution of the hierarchy problem, LHM phenomenology also include viable WIMP dark matter candidates [11].

2.6 Vector-like quarks

Vector-like quarks (VLQ) are hypothetical spin-1/2 coloured fermions. These coloured SU(3) triplets emerge primarily in non-SUSY BSM theories like CHM and LHM. Although the main SUSY models like MSSM do not predict VLQs, there are some low energy variations that allow the existence of a VLQs [75]. Current experimental bounds from the S parameter alone excludes a chiral SM-like fourth generation quark at 9σ [6]. The Higgs discovery also adds heavy constraints on the existence of a fourth chiral family [76]. Unlike SM chiral quarks, the left and right handed components of VLQs transform in the same manner under SU(2) transformations. This implies that a VLQ interacts via a vector current interaction with the W-boson. Due to this reason, VLQs can have a gauge-invariant mass-term which is disconnected from the Higgs mechanism. Thus, experimental bounds from the Higgs discovery do not put any constraint on the existence of a VLQ. Although VLQs have no Yukawa coupling with the Higgs boson, they can couple to the Higgs via SM chiral quarks.

[4] The pion is the Goldstone boson of QCD. It gets its mass in the same way the Higgs gets its mass in CHM.

A fermionic loop of a VLQ with the top-quark in association with a Higgs boson and a VLQ-Higgs loop can successfully cancel the quadratic divergence of the Higgs mass with slight fine tuning which depends on the mass of the VLQ [12]. Consequently, the search for VLQs is a perfect probe for a plethora of BSM theories in the post-Higgs landscape.

2.6.1 Phenomenology of vector-like quarks

Since VLQs are theorised in many models, a model-independent phenomenology of VLQs is the best tool for their search in collider experiments like the LHC. The simplest such model known as the "Bridge Model" is developed in Ref. [12]. Here, the VLQs (Q) couple with the SM gauge bosons ($V = W^{\pm}/Z$) and the Higgs boson (H) and a SM quark (q) via a dimensionless coupling $c^{qQ}_{\chi=L/R}$ where $\chi = L/R$ denotes the chirality. The Lagrangian for this model is formulated as

$$\mathcal{L} = \sum_{\chi,q,Q} \left[\frac{g_w}{2} \sum_V c^{Qq}_{\chi,V} \bar{Q}_\chi \not{V} q_\chi + c^{Qq}_{\chi,H} H \bar{Q}_{\chi'} q_\chi \right] + \text{h.c.} \ . \tag{2.22}$$

There are four types of VLQs, $Q = T^{+2/3}, B^{-1/3}, Y^{-4/3}, X^{+5/3}$, where the superscript denote the electric charges. In Ref. [13], these four quarks are divided into seven SU(2) multiplets in a completely renormalisable theory.

$$\begin{array}{l} T_{L,R}, \quad B_{L,R} \quad \text{(singlets)}, \\ (X,T)_{L,R}, \quad (T,B)_{L,R}, \quad (B,Y)_{L,R} \quad \text{(doublets)} \\ (X,T,B)_{L,R}, \quad (T,B,Y)_{L,R} \quad \text{(triplets)} \end{array} \tag{2.23}$$

The coupling of the Langrangian in Ref. [13] is defined via their mixing angles $\theta_{\mathrm{L/R}}$. For each multiplet, θ_{L} and θ_{R} are related via the relation

$$\begin{aligned} \tan\theta^q_R &= \frac{m_q}{m_Q} \tan\theta^q_L \quad \text{(singlets, triplets)} \\ \tan\theta^q_L &= \frac{m_q}{m_Q} \tan\theta^q_R \quad \text{(doublets)} \ . \end{aligned} \tag{2.24}$$

$m_q << m_Q$ for the current search program in LHC and thus only the left-handed couplings are relevant for singlets and triplets whereas right-handed couplings matter for the doublets. For both these theories, the VLQs predominantly couple to the gauge and Higgs bosons via the third generation quarks. In this thesis, the searches are performed for Y/T decaying to Wb. The interpretation of the Y quark is done in both the (B,Y) doublet and the (T,B,Y) triplet cases. For the T quark, the benchmark singlet case is analysed. The T quark does not couple to Wb in the (T,B,Y) triplet [13]. Hence, for the Y quark both the right and left-handed couplings are probed whereas for the T quark only the left-handed coupling is studied. The T VLQ can decay to Wb, Zt, and Ht. For $m_T > 600\,\text{GeV}$, the relation among

the branching ratios for these final states can be written as

$$\mathcal{BR}(T \to Wb) : \mathcal{BR}(T \to Zt) : \mathcal{BR}(T \to Ht) \approx 2 : 1 : 1 \tag{2.25}$$

Obviously, the Y VLQ with a charge of $-4/3$ decays exclusively to a Wb final state. The mixing angle of Ref. [13] and the coupling of Ref. [12] can be related as

$$\begin{aligned} c_{L/R}^{Wb} &= \sqrt{2} \cdot \sin\theta_{L/R} \text{ (singlet and doublet model)} \\ c_{L/R}^{Wb} &= 2 \cdot \sin\theta_{L/R} \text{ (triplet model)} \end{aligned} \tag{2.26}$$

where only QWb interactions are allowed. Due to the SU(2) multiplet representation, the renormalisable theory in Ref. [13] is indirectly constrained by LEP measurements on R_b, the partial width of the $Z \to b\bar{b}$ hadronic decay, and the oblique electroweak parameters S and T [77].

2.6.2 Experimental searches for vector-like quarks

Both the ATLAS and CMS collaborations have performed extensive searches to look for the existence of VLQs. The two main modes that are searched for are the pair-production and the single-production. Similar to the SM quarks, the pair-production of VLQs is theorised to happen via a strong force induced $gg \to Q\bar{Q}$ process. Therefore, the pair-production searches can be truly model-independent. The pair-production searches performed using the ATLAS detector did not discover any significant excess and hence have provided 95% confidence level mass and coupling limits [78–87]. In Ref. [87], the most comprehensive combination of pair-production searches was performed. A singlet T quark with mass less than 1.31 TeV and a singlet B quark with a mass less than 1.22 TeV are excluded in this combination. This analysis also excluded masses lower than 1.37 TeV for the B and T quarks in the (B, Y) doublet model. For a pair-produced Y quark, all masses less than 1.35 TeV were excluded [82]. The CMS collaboration also has published various limits in different final states [88–98].

Single-production searches are model-dependent because of their electroweak production modes. Although these searches are coupling dependent, they have a significantly higher production cross-section for pp collisions when compared to their pair-production counterparts (Fig. 2.8). The ATLAS and CMS collaborations have published various coupling dependent mass limits for single-production searches [84, 99–108]. Limits between 1 and 2 TeV have been set for various values of the individual couplings. Using the 2015 data collected at the ATLAS detector, coupling limits on the QWb coupling strength for a Y quark of mass 1 TeV in the (B, Y) doublet was quoted as $|c_R^{Wb}| = 0.45$. For a unit coupling, the mass limit for Y quark was 1.44 TeV [109].

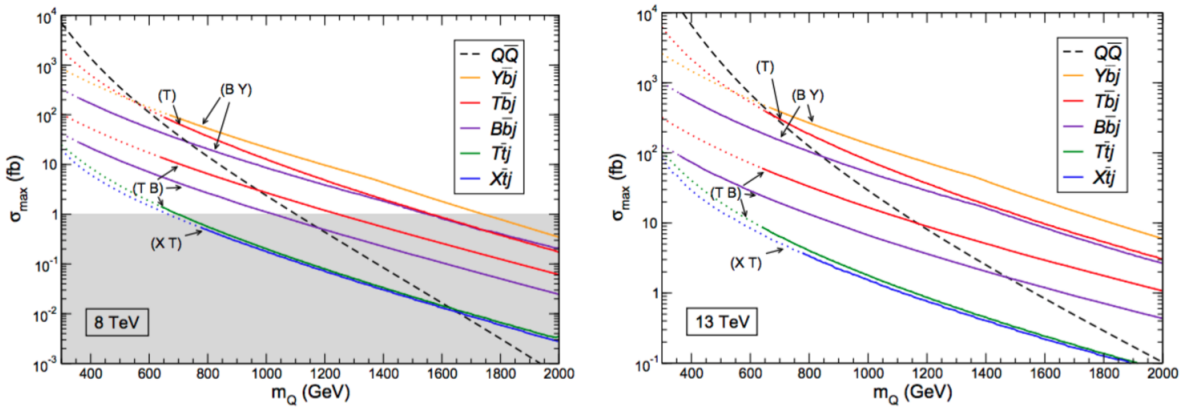
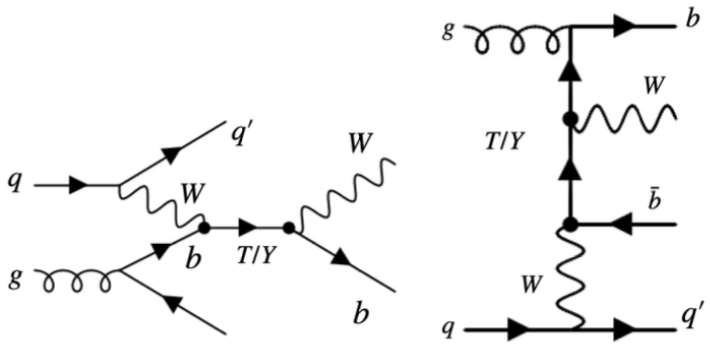


Figure 2.8: A comparison of production cross-sections for pair and single-production final states as a function of VLQ mass for pp collisions at $\sqrt{s} = 8$ TeV and $\sqrt{s} = 13$ TeV respectively [13].

2.6.3 Interference with Standard Model backgrounds

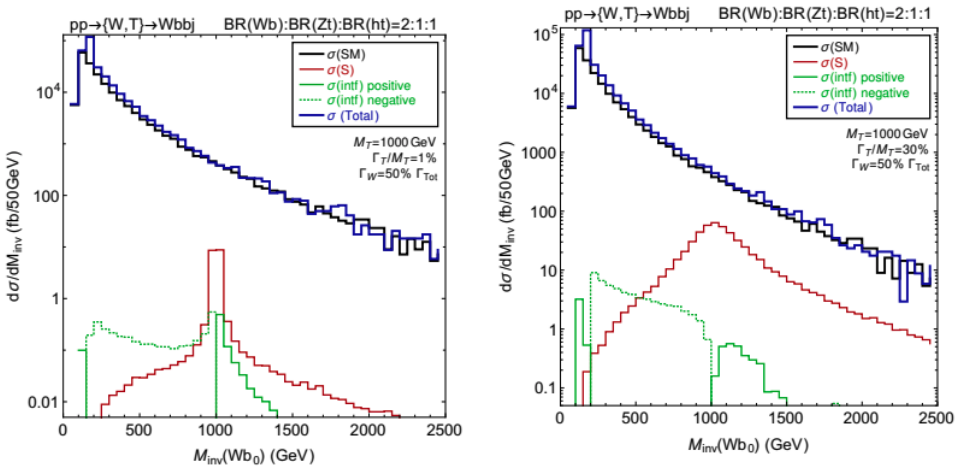
The resonant and non-resonant diagrams of the $T/Y \to Wb$ decay are the s-channel and the t-channel diagrams respectively as depicted in Fig. 2.9. The s-channel diagram contributes mainly to the production cross-section. In pp collisions, the SM backgrounds that have similar final states e.g. single-top processes interfere both constructively and destructively with the signal. When both the s-channel, t-channel, and the SM backgrounds are considered for the $pp \to Wbbj$ final state, a 10% effect of interference on the signal around the mass peak is expected for the case when the VLQ resonance has narrow-width [110]. The effect on the shape of the invariant mass of the VLQ for both the narrow-width and large-width cases is shown in Fig. 2.10



(a) (b)

Figure 2.9: Feynman diagrams for $T/Y \to Wb$ for the (a) s-channel (resonant) and (b) t-channel (non-resonant) processes in pp collisions.

The cross section is proportional to the scattering amplitude of the process squared, hence



(a) (b)

Figure 2.10: Effect of background interference for a VLQ T with a mass of 1 TeV for (a) a narrow-width signal and (b) a large-width signal [110].

the scattering amplitude for the process $pp \to Wbbj$ can be written as:

$$|\mathcal{M}(pp \to Wbbj)|^2 = |\mathcal{M}_{SM} + \mathcal{M}_{VLQ}|^2. \tag{2.27}$$

Therefore,

$$|\mathcal{M}(pp \to Wbbj)|^2 = |\mathcal{M}_{SM}|^2 + |\mathcal{M}_{VLQ}|^2 + 2Re(\mathcal{M}_{SM}\mathcal{M}_{VLQ}). \tag{2.28}$$

For this thesis, this results in a cross-section parametrised as [110]:

$$\sigma(pp \to Wbbj) = \sigma_{\text{SM}} + \kappa^2\sigma_{VLQ} + \kappa\sigma_{\text{int}} \tag{2.29}$$

where κ is the coupling between the probed VLQ and the SM particles it can interact with, and σ_{int} is the contribution of interference part to the cross-section and σ_{int}, is the cross-section arising from the interference.

CHAPTER 3

Experimental apparatus

Particle physics experiments are mostly scattering experiments. In such experiments, particles are accelerated and subsequently collided against each other at high energies. In this thesis, the analysed data has been collected using the ATLAS detector situated at the Large Hadron Collider (LHC). In this chapter, a brief overview of the LHC and the ATLAS detector will be presented. Finally, the details of Monte Carlo simulations and event generation will also be discussed.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [111] is the largest ever hadron collider that was built and is situated at CERN at the Swiss-French border near Geneva. It is a 26.7 km long proton-proton (pp) collider situated about 100 m underground. The LHC was designed to run at a centre-of-mass energy of $\sqrt{s} = 14$ TeV, but was operated at a beam energy of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV in Run 1 and $\sqrt{s} = 13$ TeV in Run 2 [112]. Recently, the Run 3 was started at a beam energy of $\sqrt{s} = 13.6$ TeV. Additionally, the LHC can also perform heavy ion collisions. This thesis is based on the pp collision data taken in Run 2.

The accelerated protons are collided at four different collision points. Two of them are the two multi-purpose detectors ATLAS [113] and CMS [114], and the other two points are the LHCb [115] and ALICE [116] experiments. The ATLAS and the CMS detectors collect data to study both the Standard Model and Beyond Standard Model physics. The LHCb detector is used to study b and c physics whereas the ALICE detector collects heavy ion collision data. Additionally, there are three more smaller experiments situated at the LHC, namely TOTEM [117], MoEDAL [118], and LHCf [119]. TOTEM, situated at the CMS interaction point, studies precise measurements of protons as they emerge from collisions at small angles. MoEDAL, situated at the LHCb interaction point, searches for magnetic monopoles and other exotic particles. This detector has been upgraded to MoEDAL-MAPP [118] to increase the sensitivity of MoEDAL to milli-charged particles and long lived particles. The LHCf, located at the ATLAS interaction point was built to calibrate hadronic interaction

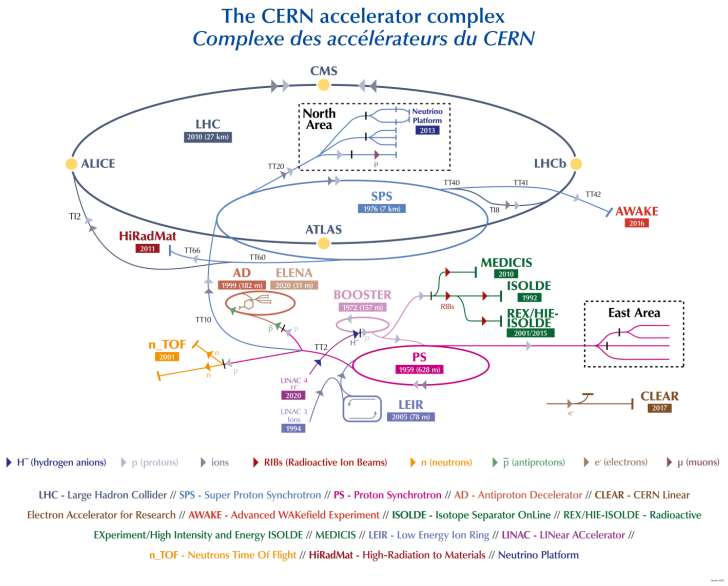


Figure 3.1: A schematic of the CERN accelerator complex [120].

models but it also measures spectra of particles such as photons and neutrons.

As depicted in Fig. 3.1, the protons that are injected into the LHC are accelerated consecutively through a series of smaller accelerators. The initial acceleration is performed at the LINAC (LINAC2 for Run 2 and LINAC4 for Run 3). The hydrogen atoms are stripped of the electron and the resulting protons are accelerated to 50 MeV. This is followed by the booster which further boosts the energy of the protons to 1.4 GeV. Subsequently, the Proton Synchrotron PS accelerates the protons to an energy of 25 GeV. Finally, the last acceleration prior to injection is done by the Super Proton Synchrotron (SPS) where the proton beams are accelerated up to 450 GeV. The LHC performs the final acceleration of each proton beam to $\sqrt{s}/2$ by using eight radio frequency cavities per beam operating at 400 MHz. LHC uses 1232 dipole magnets to curve the proton beams along the circumference of the collider. As protons are positively charged, they repel each other, and thus around 800 quadrupole magnets are used to focus the beams during their journey across the collider. With a revolution frequency of $f = c/27\,\text{km} \approx 11\,\text{kHz}$, and 2808 bunches each containing around 1×10^{11} protons, the instantaneous luminosity can be calculated using Section 2.1 as $\mathcal{O}(1 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1})$. The LHC delivered integrated luminosity of $156\,\text{fb}^{-1}$ during the Run 2. The ATLAS detector recorded $139\,\text{fb}^{-1}$ of data which was suitable for physics analyses. In this thesis, that data has been analysed. The delivered and recorded luminosity as a function of time during stable pp beams is shown in Fig. 3.2. Due to the high instantaneous luminosity, one bunch causes multiple pp collisions. Thus, contributions from these additional pp collisions, (in-time) pileup events, are superimposed over the hard collisions. This integration has a latency period of few nanoseconds which adds to the (out-of-time) pileup effects. The mean number of bunch crossings during the whole Run 2 was 33.7, thus creating significant pileup effects.

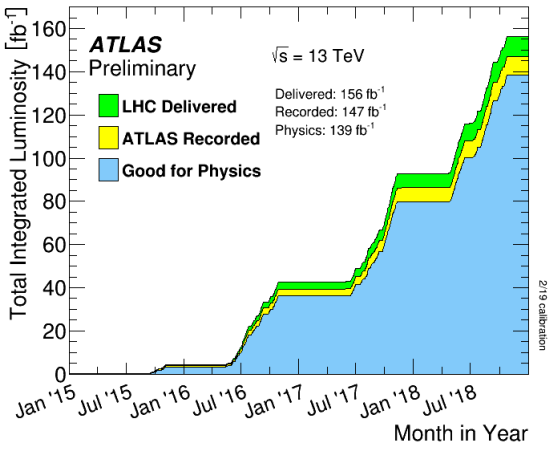


Figure 3.2: Cumulative luminosity as a function of time delivered to ATLAS (green), recorded by ATLAS (yellow), and certified to be good quality data (blue) during stable beams for pp collisions at 13 TeV centre-of-mass energy in Run 2 during 2015-2018 [121].

3.2 The ATLAS detector

The ATLAS (A Toroidal LHC ApparatuS) detector [113] is a general purpose detector built to study the SM and BSM physics in pp collisions. The 44 m long, 25 m high, and 7000 ton heavy detector is the largest detector at the LHC. The detector is managed by the ATLAS collaboration which also performs numerous physics analysis using the data acquired by the detector. An overview of the ATLAS detector is shown in Fig. 3.3.

Co-ordinate system and kinematic observables The measurements of various particles using the ATLAS detector are performed using a right-handed coordinate system. The direction of the accelerating proton beams is taken as the z-axis. The transverse plane is used as the xy plane. The y axis is points upwards towards the surface whereas the x axis is defined as the horizontal direction toward the centre of the LHC ring. Protons being composite particles the centre-of-mass system of the interacting partons moves with respect to the lab system. Thus, the initial momentum of the partons is zero only in the transverse plane. Consequently, accurate momentum measurements can be done using the transverse component of the momentum vector, p_{T}. It is defined as

$$p_{\mathrm{T}} = \sqrt{(p_{\mathrm{x}})^2 + (p_{\mathrm{y}})^2}.$$

Neutrinos are not detected by the ATLAS detector due to the weak interaction, and thus a quantity called $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined in the transverse plane to measure the momentum imbalance. It is defined as the absolute magnitude of the vector imbalance of the momentum in the transverse plane. The spherical co-ordinate system is used to describe the measured physics objects. For a given vector, the polar angle θ is the angle with respect to the z-axis. The

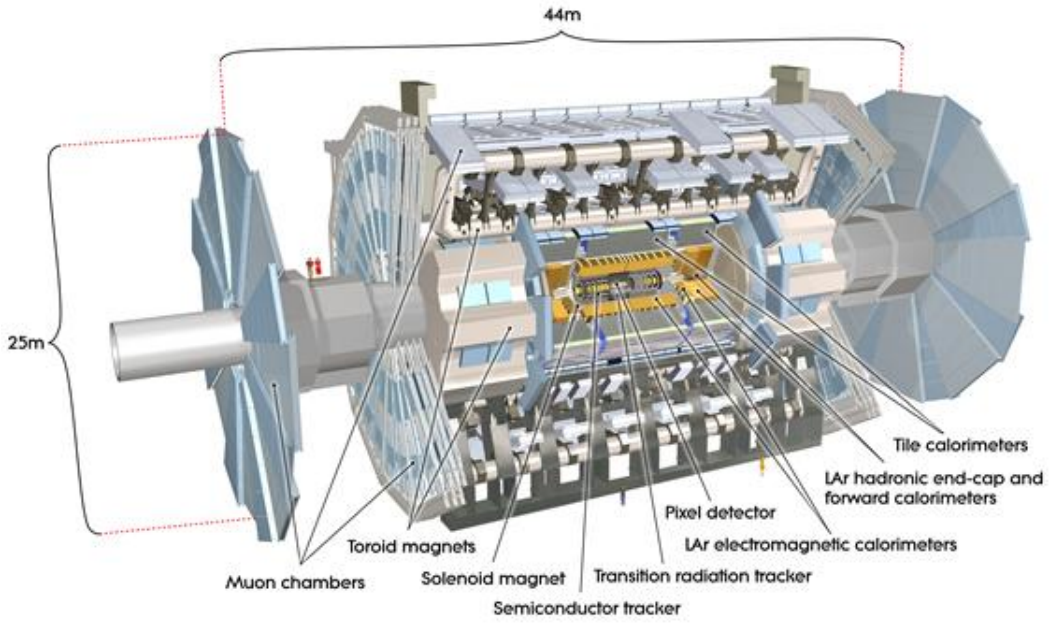


Figure 3.3: An overview of the ATLAS detector [122].

azimuthal angle ϕ is is calculated in the x-y plane with respect to the x-axis. Rather than using θ, the pseudorapidity η, which is the high energy approximation of rapidity y is used in certain cases. ΔR is invariant under a Lorentz transformation. ΔR is defined either using $\Delta\eta$ or Δy. To define certain objects such as jets (See Section 4.3) in Run 2 of the LHC, ΔR uses Δy.

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right) \quad \overset{m\ll E}{\longrightarrow} \quad \eta = -\ln\left(\tan\frac{\theta}{2}\right)$$
$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta/y)^2}$$

Components of the ATLAS detector

Inner detector

The Inner Detector (ID) [123–125] is the detector system situated closest to the beam pipe. It was designed to identify charged particles e.g. electrons and provide momentum and vertex measurements to reconstruct the tracks of particles. The ID is divided into three sub-systems, the Transition Radiation Tracker (TRT), the Semiconductor Tracker (SCT), and the Pixel detector. The ID was designed to achieve a transverse momentum resolution of $\sigma_{p_T}/p_T = 0.05\% p_T \oplus 1\%$. An overview of the ID with its sub-systems is shown in Fig. 3.4.

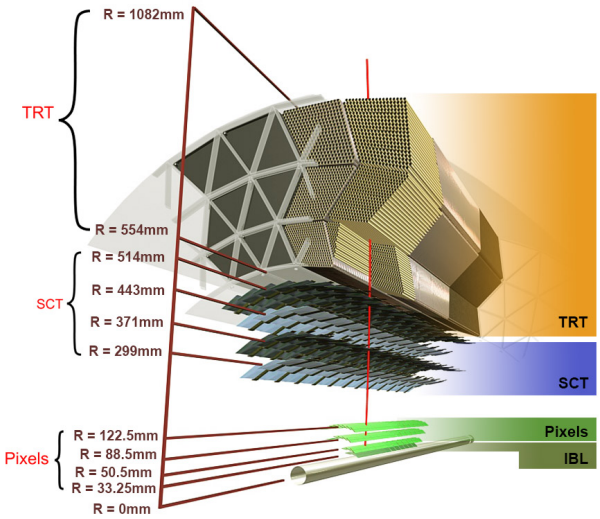


Figure 3.4: An overview of the inner detector of ATLAS. The three detector sub systems, the Pixel detector, SCT, and TRT are also highlighted here [125].

Pixel detector The Pixel detector [126] is the first detector sub-system located near the beam pipe. It is a silicon pixel detector. At first it was designed with 3 cylindrical layers in the barrel region and 3 disks in the two endcap regions. The pixels in the three layers present in the barrel region and the the end-caps have a size in $R - \phi xz$ of 50 µm x 400 µm and a sensor thickness of 250 µm. This provides an intrinsic transverse resolution of 10 µm and a longitudinal resolution of 115 µm. The barrel region gives a coverage of $|\eta| < 1.5$ and the two endcap regions have a coverage of $1.5 < |\eta| < 2.5$.

A fourth cylindrical layer, the Insertable B-layer (IBL) [127] was installed in the long shutdown between Run I and Run 2 to improve the tracking performance and precision. The addition of the IBL to ATLAS improved the resolution of impact parameter (the minimum distance between the track of the particle and the primary collision point) by around 40% for tracks with $p_{\mathrm{T}} < 1\,\mathrm{GeV}$ [128]. It has a a cylindrical structure consisting of a layer of pixel modules placed at about 33 mm from the beam axis. The IBL provides a coverage of $|\eta| < 2.5$.

Semiconductor tracker (SCT) The next detector sub-system located after the Pixel detector, the SCT [129] is a silicon strip detector. It consists of 4 double layers of microstrips in the barrel region and nine planar end-cap discs on each side. In the barrel region, the SCT covers $255\,\mathrm{mm} < R < 549\,\mathrm{mm}$ i.e. $\eta < 1.4$, whereas in the end cap regions it covers $275\,\mathrm{mm} < R < 560\,\mathrm{mm}$ for $839\,\mathrm{mm} < |z| < 2\,735\,\mathrm{mm}$ i.e. $1.4 < |\eta| < 2.5$. The SCT offers a transverse resolution of 17 µm and a longitudinal resolution of 580 µm. Due to a lower expected particle density, silicon strips were preferred over silicon pixels. The usage of strips also allows to optimise the readout channels while maintaining a good spatial resolution. Both the silicon based sub-systems of the ID are kept at low temperature to alleviate the

electronic noise.

Transition Radiation tracker (TRT) Unlike the Pixel detector and SCT, the TRT [130] is a gaseous detector sub-system. It constitutes of 300,000 straw tubes, each having a diameter of 4 mm filled with a gaseous mixture of 70% Xenon, 27% carbon dioxide, and 3% oxygen. During Run 2, a second gas mixture where Xenon was replaced with Argon was used for straw tubes belonging to modules with gas leaks. Each tube centre has a gold-plated tungsten wire with a potential difference of 1.5 kV with the surface. This potential difference amplifies the electrical signal arising from the ionisation of the gas inside the straw tubes.

The straws are kept parallel to the beam pipe in the barrel region where they cover the $560\,\text{mm} < R < 1\,080\,\text{mm}$ and $|z| < 720\,\text{mm}$ region. For both the end-cap regions, the straws are arranged orthogonal to the beam pipe and they cover a region of $827\,\text{mm} < |z| < 2\,774\,\text{mm}$ and $617\,\text{mm} < R < 1\,106\,\text{mm}$. The coverage of the TRT is $|\eta| < 2.0$ and it has a single hit resolution of 120 µm in the barrel and 130 µm in the end-cap regions. Additionally, the TRT also provides the particle ID. However, for the high luminosity upgrade of the LHC, the ID will be replaced by a completely silicon based tracker (ITk) [131].

Calorimeter system

The ATLAS calorimeter system [132] consisting of the electromagnetic calorimeter (ECAL) [133] and the hadronic calorimeter (HCAL) [134] are used to measure the energy of particles traversing the detector. The high energy particles deposit energy in the calorimeters with cascades of secondary particles namely showers. The ECAL measures the energies of particles such as electrons whereas the HCAL takes energy measurements of particles such as pions, kaons, and other neutral particles. The calorimeters provide an expanding coverage of $|\eta| < 4.9$ with ϕ symmetry. Both ECAL and HCAL are sampling calorimeters where active material such as liquid argon and plastic scintillator is placed alternatively with passive absorbing material such as copper, iron, tungsten, and lead. A sliced view of the calorimeter is shown in 3.5

Electromagnetic Calorimeter (ECAL) The ECAL placed around the ID is a sampling calorimeter with lead absorbers kept in an accordion structure where the gaps contain liquid Argon (LAr) as the active material. The LAr ECAL is maintained at a low temperature of 90 K. The barrel region has a coverage of $|\eta| < 1.47$ and the two end-cap regions cover $1.375 < |\eta| < 3.2$. Particles travelling through the ECAL ionise the LAr which helps to measure their respective energies. The absorption power of the ECAL is characterised by radiation lengths. Radiation length, X_0, is the distance an electron traverses before losing $1/e$-th of its energy. For the ECAL, the thickness can be quantified as $22X_0$ for the barrel and $24X_0$ for the end-caps.The energy resolution of the ECAL is $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$ [136].

Hadronic Calorimeter (HCAL) The HCAL too is a sampling calorimeter comprising of three parts, the Tile calorimeter (TileCal), the Hadronic End-cap Calorimeter (HEC) and

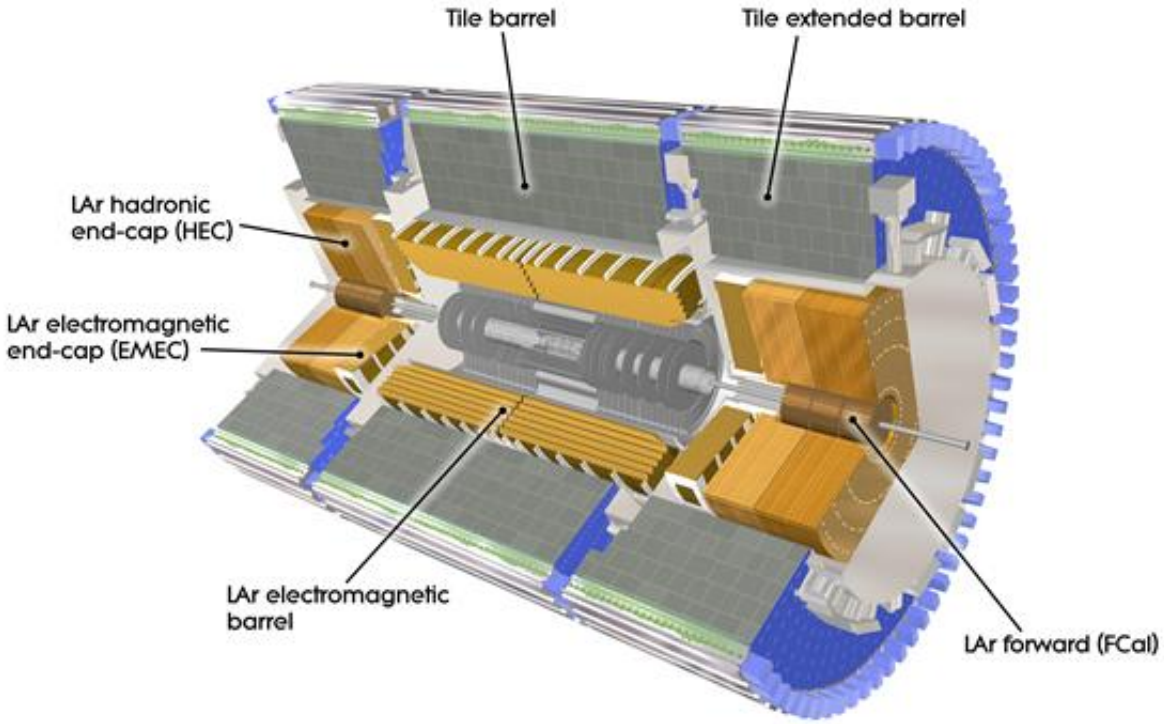


Figure 3.5: A sliced view of the ATLAS calorimeter system [135].

the Forward Calorimeter (FCal). The TileCal, placed around the ECAL, is made up of steel absorber plates and plastic scintillators as the active material which are read through photomultiplier tubes. The central barrel of the TileCal covers $|\eta| < 1$ and the two extended barrels cover $0.8 < |\eta| < 1.7$. The HEC has two wheels per end-cap and is constructed using LAr technology with copper absorber plates. It provides a coverage of $1.5 < |\eta| < 3.2$. Similarly, the FCal also uses LAr technology with copper and tungsten as absorbers. It covers the region $3.1|\eta| < 4.9$. Both the TileCal and HEC have an energy resolution of $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$ whereas the FCal has a resolution of $\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$ [137].

Muon spectrometer (MS)

The final sub-system of the ATLAS detector is the muon spectrometer (MS) [138]. Due to the high mass of muons, they can travel a considerable distance before being detected in the MS. The MS has four detector sub-systems grouped into trigger and precision muon tracking chambers. Additionally, there are two end-cap toroid magnets which assist in muon tracking by providing a magnetic field orthogonal to the muon direction. A layout of the MS is presented in Fig. 3.6.

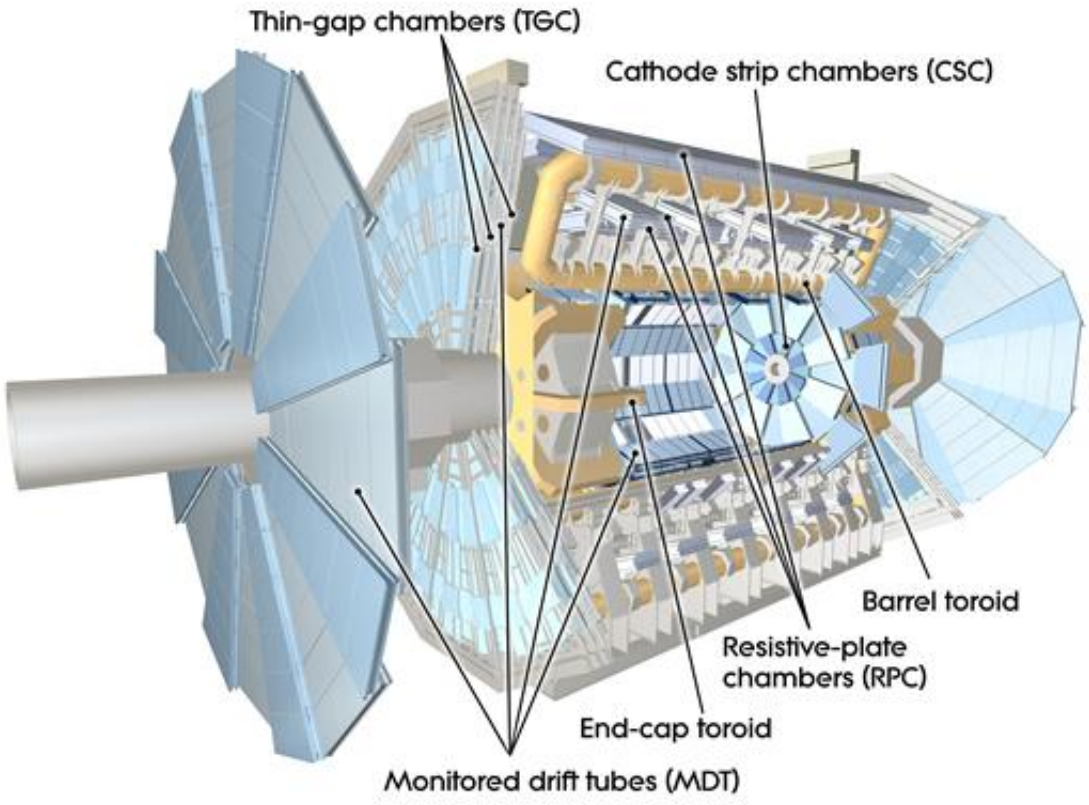


Figure 3.6: A layout of the ATLAS muon spectrometer [139].

The muon trigger chambers provide fast trigger information for the muons. The barrel region, $|\eta| < 1.05$ uses three resistive plate chambers. These plates, having high a potential difference among each other are placed parallel to each other with a gas mixture in the gaps. In addition to the trigger information, these plates also offer a spatial resolution of 10 mm in the $\eta - \phi$ plane. The end-cap regions ($1.05 < |\eta| < 2.4$ of these trigger chambers constitute of multi-wire chambers with a gas mixture in the gaps. The wires are put at a distance of 1.8 mm. These thin gap chambers (TGC) are also capable of measuring ϕ information with a resolution of 5 mm. Each TGC module has a wire plane positioned at high positive voltage, inserted between grounded cathode planes. The TGCs can trigger the high momentum muons.

Track coordinates are measured precisely using two types of chambers: the Monitored Drift Tubes (MDT) at small pseudorapidity values ($|\eta| < 1.4$) and the Cathode Strip Chambers (CSC) at higher pseudorapidity values ($2 < |\eta| < 2.7$). MDTs are aluminium drift tubes filled with an argon-based gas with a central tungsten wire. CSCs are multiwire proportional chambers where the cathodes are divided into strips. The MS was designed to provide a high-p_{T} resolution for muons. A 1 TeV muon can be measured with a 12% p_{T} resolution whereas muons with p_{T} with 10 to 200 GeV can be measured with about 3% resolution.

Magnet system

In addition to the detector sub-systems, the magnet system [140] plays a crucial role in performing tracking, energy, and charge measurements. Due to the Lorentz force, the magnetic field helps to bend the charged particles based on their mass and charge. The central magnet is a superconducting solenoid magnet with a magnetic field of 2 T. This magnet is made of NbTi and is maintained at a low temperature odf 1.8 K using liquid helium. The toroidal magnets are divided in the central barrel region and the two end-caps. These magnets offer inhomogeneous magnetic fields of 0.5 T in the barrel region and 1.5 to 2 T in the end-cap regions.

Trigger and Data Acquisition system

The Trigger and Data Acquisition system (TDAQ) [141] has two components: the data acquisition (DAQ) part which reads out the detector information and the trigger system which receives the detector information and decides which information to retain. The TDAQ receives information at the rate of 40 MHz and filters it down to a rate 1.5 kHz and ensures the saving of the relevant physics information of the measurements. The trigger system is further divided into the Level-1 (L1) and High-Level triggers (HLT). The hardware-based L1 trigger reduces the information rate from 40 MHz to 100 kHz by selecting only reduced granularity information from the calorimeters and the muon spectrometer. The L1, having a fast latency period of 2.5 µs, identifies various regions of interest in the $\eta - \phi$ plane and transmits the information to the HLT. The software-based HLT refines the information rate from 100 kHz to 1.5 kHz based on additional selection criteria. The HLT has a latency period of 550 ms, thus allowing it to apply selection criteria close to offline requirements. A schematic overview of the TDAQ is shown in Fig. 3.7

3.3 Event simulation

To analyse the data collected by ATLAS detector precisely, it is important to simulate the events of the underlying processes. The physical processes are simulated using the statistical methods of Monte Carlo simulations. Subsequently, the evolution of the processes in the detector by passing these events through a simulation of the ATLAS detector. As shown in Eq. (2.12), the cross-section of a process in hadron collisions can be factorised. A set of Monte Carlo generators can simulate the subsequent processes like parton-shower and hadronisation in addition to the hard collisions. The MC generators produces millions of events to model the probability distribution of the underlying physical observables of the high-energy processes. These events are then processed by the ATLAS detector simulation which is coded with the GEANT4 toolkit [143]. Simulating the whole decay chain of particles through the ATLAS detector is computationally expensive. To combat this, ATLAS collaboration has developed faster algorithms with performance similar to GEANT4. One of these algorithms, ATLFAST-II [144] has been used in this thesis.

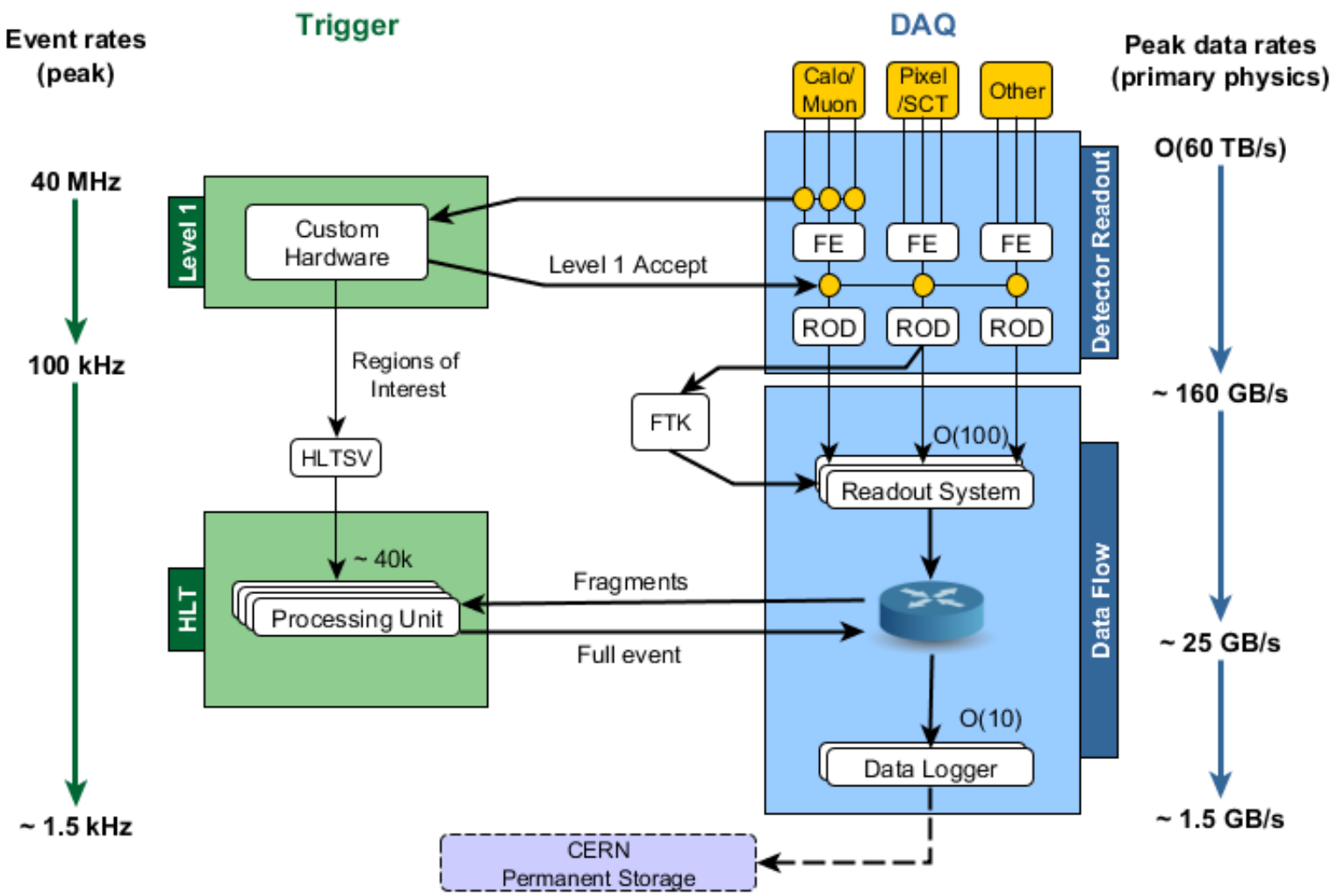


Figure 3.7: A schematic overview of the TDAQ system of ATLAS [142].

CHAPTER 4

Physics object reconstruction

The ATLAS detector detects the various particles that are produced in pp collisions at the LHC by using the complex information from the interactions of these particles with various parts of the detector, which have been discussed in the previous chapter. All particles except neutrinos leave directly measurable traces in the detector. Heavier particles generally decay into their lighter decay products which leave their signatures in the various sub-detectors. We use this information subsequently to reconstruct the various physics objects used in all physics analyses performed by the ATLAS collaboration. In this chapter, the description of more general particle reconstruction methods will preceed the description of more special algorithms used to reconstruct physics objects e.g. leptons. The focus of this chapter will be physics objects used in the analyses performed in this thesis.

4.1 Tracking and Vertexing

Tracking

Charged particles deposit different energy yields as they travel through the tracking detectors according to their physical properties which are called "hits". By combining the hits from pixels and strips of the Pixel and SCT detectors respectively, three-dimensional measurements called space points are created. Three such space points are used to create the tracking seed. An iterative combinatorial algorithm is used to find the tracks from these seeds. To ensure inclusion of seeds that gives good tracks, firstly SCT-only and pixel only seeds are used and then mixture seeds are included. These seeds serve as inputs to a Kalman filter [145]. In addition, space-points from the remaining layers of the pixel and SCT which agree with the preliminary trajectory are also put through the filter. This results in a huge amount of track candidates being created. An ambiguity solver uses a sequential fit after ranking these candidates according to a score to create the track collection. The workflow of this solver as detailed in [145] is briefed in Fig. 4.1 . The solver only accepts candidates with certain properties like $p_{\mathrm{T}} < 0.4\,\mathrm{GeV}$, $|\eta| < 2.5$. The algorithm classifies merged clusters cre-

ated by three charged particles with an efficiency of around 85% and subsequently uses it as an input for the fit. This algorithm is also propagated to TRT hits of the charged particles by reperforming the fit in an "inside-out" manner. A more detailed description of the algorithm can be found in Ref. [146]. Since the tracks that are built in this manner are trajectories of charged particles in a magnetic field, which is almost homogeneous, a helical approximation of the path is quite accurate. Thus, such helical tracks can be completely described by a five dimensional vector, $(d_0, z_0, \phi, \theta, q/|\vec{p}|)$, where d_0 and z_0 are the transverse and longitudinal impact parameters respectively, ϕ and θ are the azimuthal and polar angle respectively, and $q/|\vec{p}|$ is the charge over momentum ratio. Here the standard ATLAS coordinate system is used.

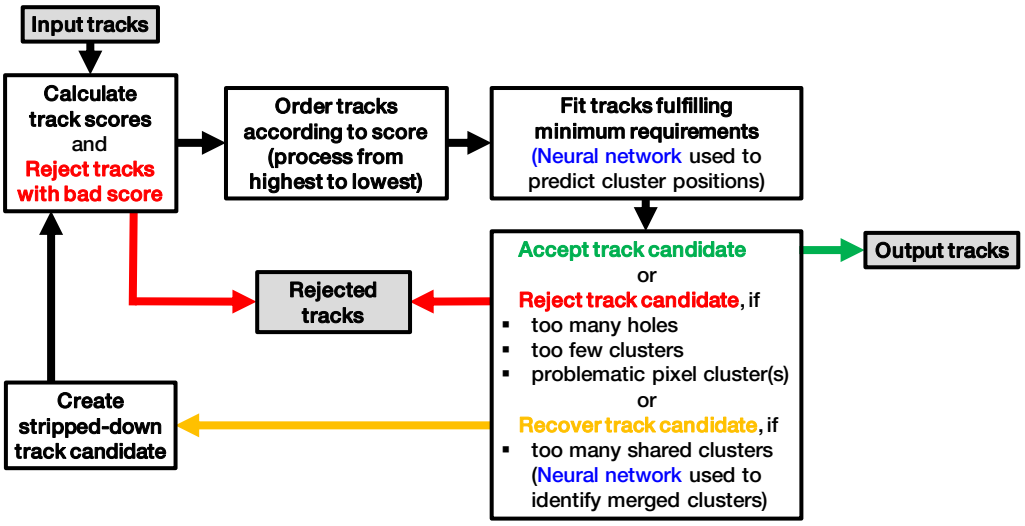


Figure 4.1: A schematic of flow of tracks through the ambiguity solver [145].

Vertexing

Vertexing is the process of attributing *primary vertices* i.e. the point of hard scattering of pp collisions. Suitable track candidates are used to perform vertex finding and vertex fitting, the two main steps of vertexing. The algorithm starts by assigning an initial seed to the primary vertex based on the coordinates of the beam spot. An adaptive vertex-fitting algorithm performs an iterative χ^2 minimisation. Each track used as an input to the fit is assigned a weight which is inversely proportional to its χ^2. A cut-off χ^2_{cutoff} is defined for the iterative fit where the vertex is recalculated at each step by recalculating the track weights. When the last iteration finishes, any track having a final weight which is 7σ away in compatibility with the primary vertex is rejected and considered for reconstructing new primary vertices [147]. If a vertex has at least two associated tracks, it is retained as a primary vertex. The width and position can also be used to gain better precision on the position of the primary vertices. If this constraint is relaxed, reconstruction of the beam spot can be done by using primary

vertices from different events. For each event, only the primary vertex is considered, for which $\sum p_{\mathrm{T}}^2$ is maximum and there are at least five associated tracks. Vertices which fall outside this spot are called *secondary vertices*. Such vertices will be discussed in detail in Section 4.3.4 in the context of reconstruction of B-hadrons.

Topo-clusters

Topological clusters [148] or topo-clusters are built iteratively with calorimeter seed cells that satisfy certain criteria regarding noise thresholds. These seeds are defined at the electromagnetic (EM) scale. Cluster seeds should have significance $\varsigma = |E_{\mathrm{cell}}^{\mathrm{EM}} / \sigma_{\mathrm{cell,\,noise}}^{\mathrm{EM}}| \geq 4$, where $E_{\mathrm{cell}}^{\mathrm{EM}}$ and $\sigma_{\mathrm{cell,\,noise}}^{\mathrm{EM}}$ are the respective expected cell energy and the expected cell noise at the electromagnetic (EM) scale. In the successive iterations to form a topo-cluster, neighbouring cells with $\varsigma \geq 2$ and $\varsigma \geq 0$ are also taken into account. The "4-2-0" reconstruction algorithm is mostly used to define topo-clusters.

4.2 Charged leptons

For this thesis, charged leptons would indicate electrons and muons.[1] The reconstruction of charged leptons, more specifically *prompt leptons*, which are defined as originating from hard scattering, will be described in this section.

4.2.1 Electrons

Electron candidates are reconstructed by using the tracks and energy clusters they leave in the tracker and the EM calorimeter. To reconstruct electrons in ATLAS, fixed-size calorimeter clusters defined by the "sliding window" algorithm [149] was used for the 2015-16 data and the dynamic variable-size "superclusters" is being used for the 2015-2018 dataset. Since in this thesis, electron candidates using these two methods were considered for two different analyses, both will be discussed here.

Reconstruction for 2015-16 data The "sliding window" algorithm starts with the building of towers on the $\eta - \phi$ grid of the calorimeter with $\Delta\eta \times \Delta\phi$ spacing of 0.025×0.025. A fixed-sized rectangular window is slid across the grid to find a local E_{T} maximum above a given E_{T} threshold (2.5 GeV). The cells contained in this window form the cluster used in further reconstruction. Such clusters in the region are then matched with electron tracks coming from the ID with matching requirements of $\Delta\eta < 0.05$ and $-0.10 < \Delta\phi < 0.05$ when compared to the barycentre of the cluster. While travelling in the detector, electrons lose energy due to bremsstrahlung, thus creating further electron-positron pairs from the radiated photon in quite some cases. The negative requirement on $\Delta\phi$ matching is to remove the effects of energy loss due to bremsstrahlung from bending of the track due to the

[1] Both electrons and muons would indicate the respective particle and their anti-particles.

magnetic field in the opposite direction. Track and cluster candidates affected by significant bremsstrahlung are also corrected for the inefficiencies and the non-linearities arising from it. If several tracks fulfil the matching criteria mentioned above, the primary electron is chosen with the track that has no associated photon-conversion vertex and a few more criteria regarding the number of hits in the silicon layers of the detector. Further details of this algorithm can be found in Ref. [150]. The efficiency of this reconstruction algorithm is close to 100% except for $E_{\mathrm{T}} < 30\,\mathrm{GeV}$ as can be seen in Fig. 4.3 when measured in $Z \to e^{-}e^{+}$ events.

Reconstruction for 2015-18 data For the "supercluster" algorithm [151], the variable-size clusters are built dynamically to recover energy losses from bremsstrahlung in a better manner. The inputs to this algorithm are topo-clusters.

The supercluster electron seed is built with a cluster having $E_{\mathrm{T}} > 1\,\mathrm{GeV}$ and four matching hits in the silicon tracking detector. All the clusters around the seed cluster barycentre within a window of $\Delta\phi \times \Delta\eta = 0.075 \times 0.125$ are also taken as satellite clusters to take into account the energy from secondary EM showers. The satellite and the seed cluster both share the same best-matched track. The satellite and the seed clusters together constitute the superclusters. After the superclusters are built for electrons and photons separately, tracks are matched to electron superclusters and conversion vertices are matched to photon superclusters. An iterative ambiguity solver decides whether an electron candidate can come from come from an electron supercluster or a photon supercluster. A flowchart of this process is presented in Fig. 4.2. Further details of this algorithm can be found in Ref. [151].

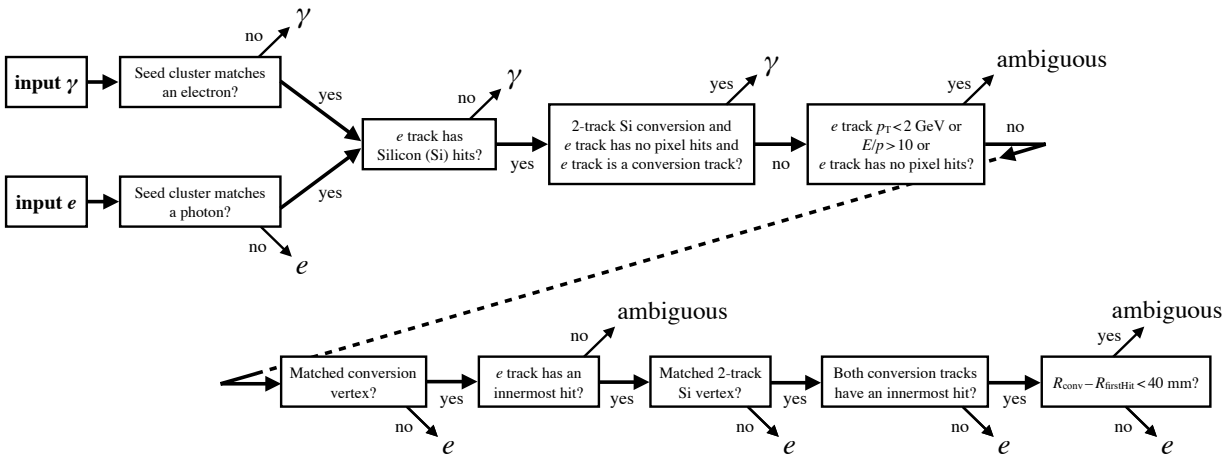


Figure 4.2: A schematic depicting the logic of the ambiguity resolution for particles initially reconstructed both as electrons and photons

Identification, isolation, and calibration

Electron candidates are further classified from non-prompt electrons using a likelihood based multivariate identification algorithm. The variables used are the properties of the primary electron track, the development of the EM shower, and the mutual compatibility of the track and the reconstructed cluster. Based on various thresholds, the electron candidates can be identified as Tight, Loose, or Medium. These three operating points are optimised in 9 $|\eta|$ bins and 12 E_{T} bins in a manner that the electrons satisfy certain requirements on the likelihood discriminant based on a set of predefined efficiencies. For standard electroweak processes, on average the Loose, Medium, and Tight operating points have respective efficiencies of 93%, 88%, and 80%. These efficiencies increase gradually as a function of increasing E_{T}. The Medium and Tight operating points achieve better background rejection for the corresponding trade-off in efficiency reduction. The efficiency of these candidates as a function of η in $Z \to e^+e^-$ events is shown in Fig. 4.3. The dip in efficiency in the region $1.37 < |\eta| < 1.52$ arises due to the gap between the barrel and endcap layers of the calorimeter where a lot of material[2] is placed. Thus, for physics analyses, these electrons are excluded. To further isolate these candidates from various non-prompt electrons certain criteria are placed on corrected sum of E_{T} and p_{T} within a cone of fixed size around the electron. These criteria ensure both calorimeter and track isolation have high efficiencies. The electron energy calibration is done in the detector region of $|\eta| < 2.47$ using $Z \to e^+e^-$ events. This process consists of optimising the energy by minimising the effects of calorimeter material. The relative energy scales of the calorimeter layers are also adjusted for during the calibration. Correcting for local non-uniformities and the energy scale in data are the final steps of this procedure.

4.2.2 Muons

Muons are reconstructed in the ID and the MS separately and then combined to reconstruct the final muon candidate [152]. The reconstruction in the ID is performed using track reconstruction of charged particles as discussed before. A Hough transform is used to search for hits corresponding to segments in various sub-detectors of the MS. The hits from these segments in the middle layer are used as initial seeds in a combinatorial segment-seeded search which is then extended to outer and inner layers of the MS. These segments are stitched together to build the corresponding tracks. Two matching segments are required for a track. In the barrel-endcap region a single high-quality segment is used. All hits associated with a given muon track candidate are fed into a global χ^2 fit. An iterative fit is performed by removing hits which contribute largely to the χ^2 in each step.

The combined ID-MS reconstruction is based on four different algorithms. The "combined" muon method uses a global fit using hits from both ID and MS. For most muons, an outward-in iteration is performed by starting with hits from the MS and then adding hits from the ID in the fit. The "segment-tagged" muon reconstruction uses ID tracks and extrapolates it to

[2] This material has radiation lengths ranging from 5 to 10.

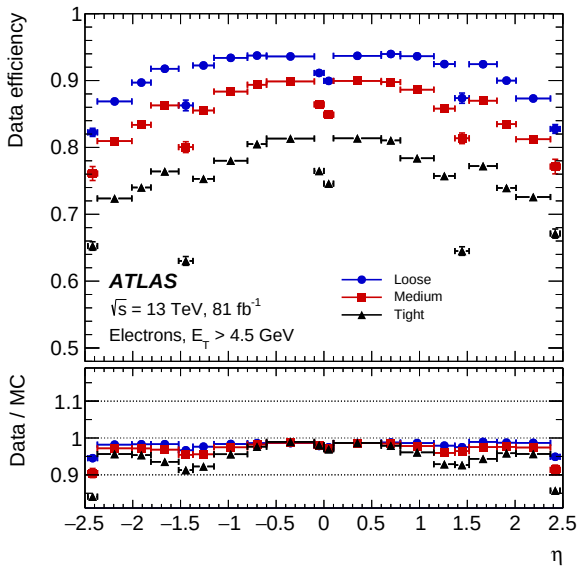


Figure 4.3: The electron identification efficiency in $Z \to e^+e^-$ events as a function of η the Loose, Medium, and Tight operating points [151].

the MS. If at least one track segment in the MDT or CSC chambers matches, it as classified as "segment-tagged". "Calorimeter-tagged" muons have a track in the ID matched to the minimum-ionising particle in the calorimeter. This type of muons recovers muons in the region where the MS is partially instrumented. "Extrapolated" muons are reconstructed by using the MS track loosely compatible with the interaction point. The muon should traverse at least two layers of MS chambers and three forward layers. This method is used in the region $2.5 < |\eta| < 2.7$, since this part is uncovered by the ID.

Identification, isolation, and calibration

Muon identification ensures the suppression of background muons originating from kaon and pion decays. Four muon identification selection points, namely, Tight, Loose, Medium, and High-p_{T} are defined. In ATLAS, the medium selection is used as default. For this selection the combined and extrapolated muons are used. For this selection, the associated systematic uncertainties are the least. High-p_{T} muons are used for muon candidates with $p_{\mathrm{T}} > 100\,\mathrm{GeV}$. The reconstruction efficiency as a function of p_{T} for the medium selection as measured in $Z \to \mu\mu$ and $J/\psi \to \mu\mu$, The efficiencies for 2015-2018 data measured in $J/\psi \to \mu\mu$ events can be seen in Fig. 4.4

Prompt muons are generally well isolated from muons coming from semi-leptonic decays of pions and kaons embedded in jets. To isolate them properly certain track-based and calorimeter-based variables similar to the electron isolation are used. The different isolation selection points all have efficiencies close to 99%. The calibration factors are found to be close to unity [152].

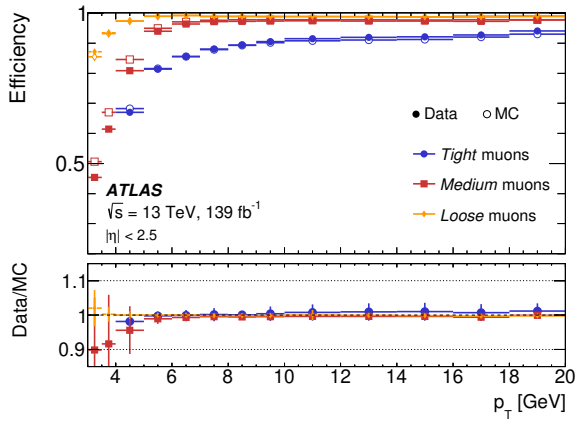


Figure 4.4: The muon identification efficiency in $J/\psi \to \mu^+\mu^-$ events as a function of p_{T} for the Loose, Medium, and Tight operating points [152].

4.3 Jets

Due to the confinement property of QCD, strongly interacting particles like quarks and gluons cannot be observed directly. The hadronisation of these particles results in a collimated spray of particles, namely jets. By measuring these jets, an approximate measurement of the kinematic properties of the original parton can be performed. The exact prescription of doing this involves a good jet definition. Jet algorithms that have been developed over the last few decades provide a set of rules according to which a certain group of final state particles can be grouped into a jet depending on distance or energy parameters of those particles.

These algorithms are required to fulfil a set of criteria that were defined by a group of theorists and experimentalists which is known as the Snowmass accord [153]. Those criteria are the following

1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields finite cross sections at any order of perturbation theory;
5. Yields a cross section that is relatively insensitive to hadronisation.

The last three criteria are set to make sure the algorithms are infra-red and collinear (IRC) safe. IRC safety implies the fact that the number of hard jets defined by the given algorithm remains unchanged despite the addition of a further soft emission or a collinear splitting [154].

- Infra-red safety requires real and virtual divergences to cancel if one or more "soft" (low-momenta particles with $p_i \to 0$) partons are added in the process of fragmentation arising from both perturbative and non-perturbative effects.
- Collinear safety requires additional divergences to the amplitude to cancel when if one or more parton splits ($p_i || p_j$) into further collinear partons.

4.3.1 Cone algorithms

Cone algorithms are "top-down" algorithms which assume hadronic activity resulting in jets to be confined inside cones in the $\eta - \phi$ plane [155]. The first jet algorithm was a cone algorithm developed by Sterman and Weinberg for e^+ e^- collisions [154]. Most cone algorithms use stable cones i.e. the momenta of the particles inside a cone of fixed radius R adds up to the direction of the cone's centre. Iterative cone algorithms which have been used in experiments are Iterative Cone-Progressive Removal (IC-PR) and Iterative Cone-Split Merge (IC-SM) algorithms [155]. IC-PR and IC-SM algorithms rely on iteratively building a stable cone around a seed cell having momentum or energy greater than a certain threshold. IC-PR removes the remaining particles inside the cone once a stable cone is formed whereas IC-SM iteratively merges them. Both these algorithms are IRC unsafe. The UA1 algorithm [156], which was the first jet algorithm developed for hadronic collisions, was a simpler version of IC-PR algorithms where an iteration is performed over decreasing energy particles rather than cone direction [155]. The SISCone algorithm [157] is a completely IRC-safe algorithm since it is a seedless algorithm and tries to iteratively find particles using the radius parameter rather than using energy or momenta requirements.

Sequential recombination algorithms

Sequential recombination algorithms can be seen as "bottom-up" [155] algorithms which are based on the assumption that jets are formed due to the successive branching of partons. These algorithms combines these partons iteratively to define jets. All sequential recombination algorithms are IRC safe and are computationally fast. The generalised k_t algorithm, which is mostly used for this class of algorithms can be summarised as following [158]:

1. List the 4-vectors of all the particles present in the event.
2. For each particle i, build inter-particle distances defined as

$$d_{ij} = \min(p_{\mathrm{T}i}^{2k}, p_{\mathrm{T}j}^{2k}) \frac{\Delta R_{ij}^2}{R^2} \tag{4.1}$$

$$\Delta R_{ij}^2 = (y_i - y_j)^2 - (\phi_i - \phi_j)^2 \tag{4.2}$$

and a beam distance for each particle in the momentum space defined as

$$d_{i\mathrm{B}} = p_{\mathrm{T}i}^{2k} \tag{4.3}$$

Here k can take the values 0, 1, or -1 based upon the algorithm. The three cases will be discussed in detail below.

3. Find the minimum in the set $\{d_{ij}, d_{iB}\}$ iteratively.
   - If $d_{ij} < d_{iB}$, both the particles i and j are removed and recombined into a new particle k and added to the list of particles.
   - If $d_{iB} < d_{ij}$ the particle i is called a jet.
4. Go back to step 2 and repeat this procedure till the list of particles is exhausted.

k_t algorithm

In Eq. (4.1), $k = 1$ corresponds to the k_t algorithm [159]. This algorithm implies that a particle with low p_T also has low d_{ij} and thus would be recombined early on the recombination process. This allows the algorithm to be IR-safe in spite of being sensitive to soft parton emissions. Simultaneously, this also results in an undesired sensitivity to pile-up. Some of the Tevatron experiments had used this algorithm.

Cambridge/Aachen (C/A) algorithm

For the C/A algorithm [160], $k = 0$ in Eq. (4.1). The distance here is purely geometrical and independent of momenta of the particles. This has the advantage of declustering and studying jet substructure.

Anti-k_t algorithm

The anti-k_t algorithm [161] is used today to reconstruct jets in all the LHC experiments. This algorithm corresponds to the $k = -1$ case in Eq. (4.1). This leads to the algorithm clustering hard (high-p_T) particles first during recombination. The algorithm results is jets that are insensitive to soft radiation. Anti-k_t jets grow in a circular manner by assimilating soft particles till a given distance R when measured from the jet axis. This property of anti-k_t jets is one of the main motivations for being used across all LHC experiments. A visual representation of the three sequential recombination algorithms and the SISCone algorithm on the $\eta - \phi$ plane with $R = 1$ is shown in Fig. 4.5. In ATLAS, two types of jets are mainly used with two different values of R, small-R jets with $R = 0.4$ and large-R jets with $R = 1.0$,

4.3.2 Small-R jets

In this section, the reconstruction and calibration of small-R jets used in this thesis will be discussed. Here, two jet reconstruction algorithms have been used. The first one, namely EM-Topo jets have been used for the analysis using data collected in 2015-16. For the analysis using the 2015-18 dataset, an algorithm called Particle Flow has been used to reconstruct small-R jets.

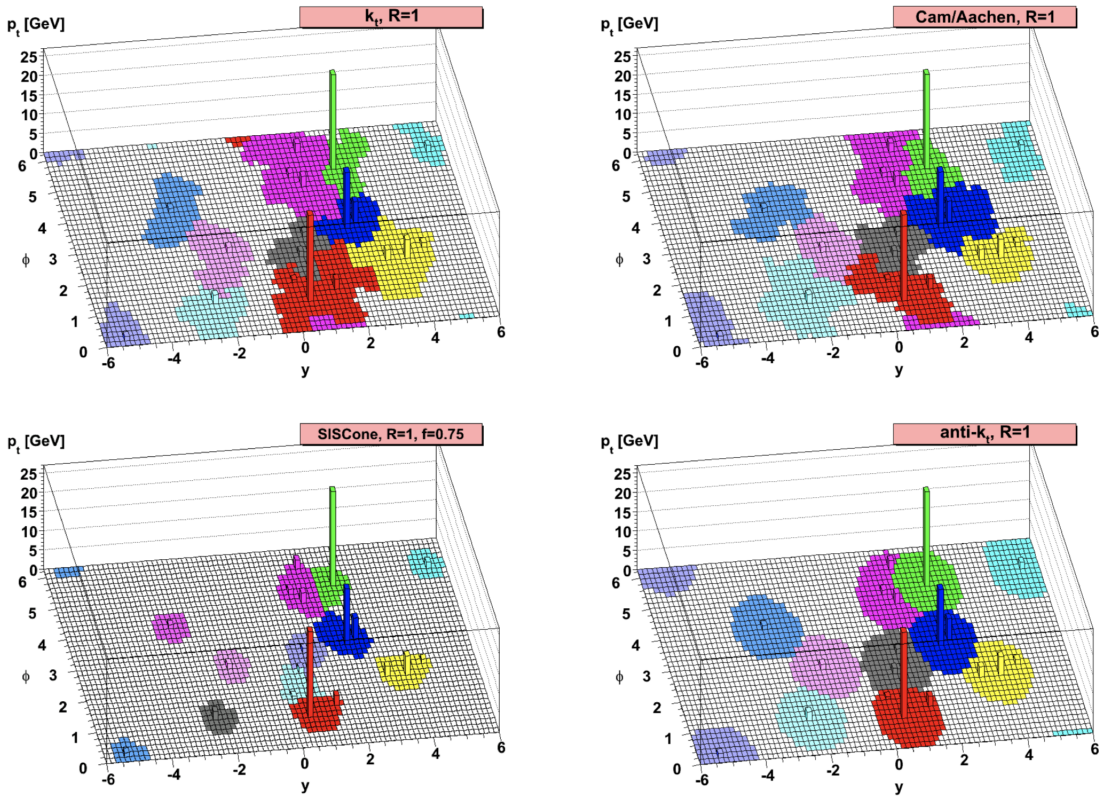


Figure 4.5: A parton-level representation of various jet-algorithms discussed above [161].

EM-Topo jets

EM-Topo jets are calorimeter jets reconstructed using an algorithm that uses topological clusters (defined in Section 4.2.1) as inputs [162]. The topological clusters are calibrated to the electromagnetic (EM) scale prior to feeding them as inputs to the algorithm. Each topo-cluster is assumed to be massless with energy E as the sum of the energies of the cells of the cluster i.e. $E = \sum E_{\text{cell}}$. A four-vector is associated with each topo-cluster with the momentum direction being from the centre of the detector to the energy weighted barycentre of the given topo-cluster. The topo-clusters over a p_{T} threshold of 7 GeV within a radius of $R = 0.4$ are clustered together according to the anti-k_t algorithm. This algorithm truncates when all the topo-cluster four-vectors are replaced by jet four-vectors.

These jets are calibrated to ensure that the jet energy scale is corrected to the truth jet energy scale defined at the particle level. A flowchart showing the stages of the calibration is illustrated in Fig. 4.6.

For the calibration, firstly, origin correction is performed to recalculate the jet's four momenta to the hard-scattering vertex without altering the jet energy. Subsequently, the pileup correction is applied to remove the excess energy inside the jet resulting from in-time

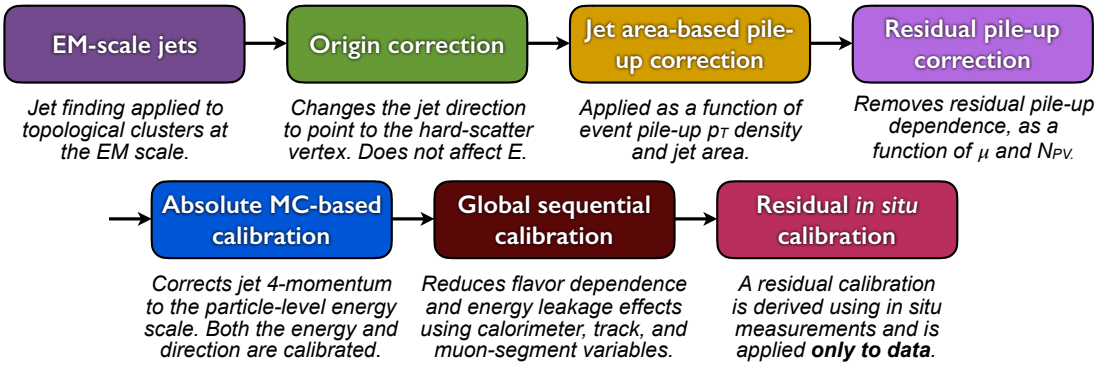


Figure 4.6: A flowchart showing the stages of the calibration of EM-Topo jets [162].

and out-of-time pileup. The corrected $p_{\mathrm{T}}^{\mathrm{corr}}$ and the uncorrected EM-scale reconstructed $p_{\mathrm{T}}^{\mathrm{reco}}$ are related as [162, 163]

$$p_{\mathrm{T}}^{\mathrm{corr}} = p_{\mathrm{T}}^{\mathrm{reco}} - \rho \cdot A - \alpha \cdot (N_{PV} - 1) - \beta \cdot < \mu > \tag{4.4}$$

where ρ is the median energy density of jets in a given event, N_{PV} is the number of primary vertices in that event, $< \mu >$ is the average number of interactions per bunch crossing, and A is the jet area. α and β are the in-time and out-of-time pileup respectively.

In the next step, the jet energy scale (JES) and η-calibration is performed to correct the four-momentum of the jet to the corresponding particle level jet. The biases arising from changes in calorimeter granularity are fixed in this step. This is done in MC-simulation by calibrating the reconstructed jet energy to isolated truth-matched jets within $\Delta R = 0.3$. The average energy response, $E^{\mathrm{reco}}/E^{\mathrm{truth}}$, is parametrised as a function of jet p_{T} and η_{det}, the jet η pointing from the centre of the detector. The energy jet energy response is shown in Fig. 4.7.

The Global Sequential Calibration (GSC) improves the JES resolution by correcting the residual dependencies on both longitudanal and transverse jet properties. For example, the flavour of the initiating particle, a quark or a gluon, affects the calorimeter response of the jet. Five jet observables are used to correct for the four-momenta of the jet iteratively as a function of $p_{\mathrm{T}}^{\mathrm{truth}}$ and η_{det}.

The last stages of the calibration involve accounting for the difference in jet response between data and MC. These differences happen due to imperfect simulation of the detector response in MC. Four *in*-situ calibrations are performed to correct the jet response of both central ($|\eta| < 0.8$) and forward ($0.8 < |\eta| < 4.5$) jets. The η-intercalibration fixes the JES of forward jets by referencing their p_{T} against the p_{T} of well-measured central jets. Subsequently, the central jets are calibrated using Z/γ+jets and dijet events. The Z/γ+jets, where the Z/γ decays into an electron or muon pair, are used to calibrate jets with $p_{\mathrm{T}} <$ 950 GeV. Multijets are then used to calibrate jets with 300 GeV $< p_{\mathrm{T}} <$ 2 000 GeV. The quantity $R_{\textit{in-situ}}$, defined for both data and MC, is the ratio of p_{T} of the jet and the p_{T} of

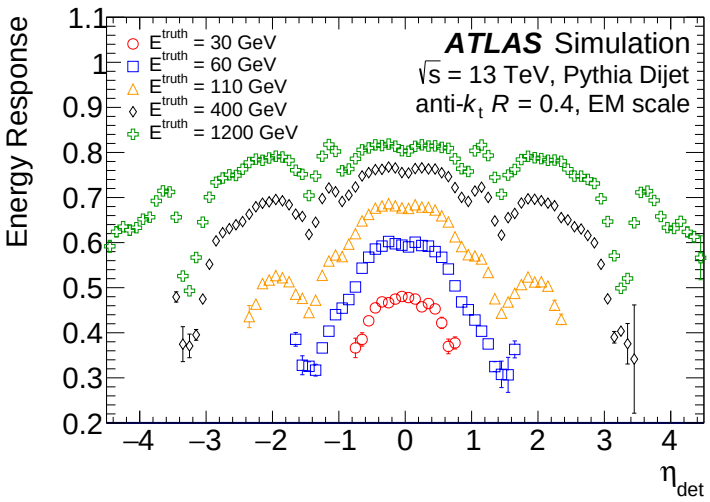


Figure 4.7: The average energy response after origin and pile-up corrections applied, as a function of η_{det} for jets of a truth energy of 30, 60, 110, 400, and 1200 GeV [162].

the reference object. The ratio $R_{\text{data}}/R_{\text{MC}}$ as a function of p_{T} for the three central jet in-situ calibrations and the final combined correction is shown in Fig. 4.8 Systematic uncertainties arising from these methods will be discussed in Chapter 7.

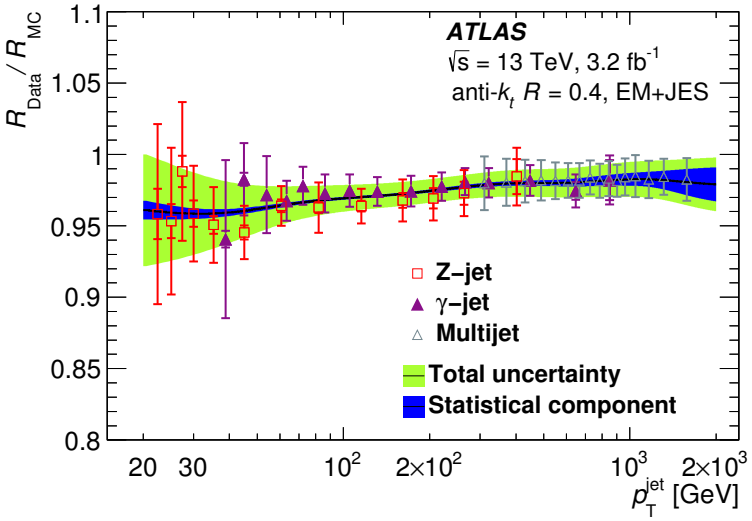


Figure 4.8: $R_{\text{data}}/R_{\text{MC}}$ as a function of p_{T} for Z+jet, γ+jet, and multijet in-situ calibrations. The black line denotes the final derived correction [162].

Particle flow jets

For the reconstruction of EM-topo jets, only calorimeter measurements are used. The particle flow (PFlow) algorithm [164] uses both tracker and calorimeter information to build "particle flow" objects, which are subsequently fed into the jet algorithm as inputs. To avoid double counting of the energy measurement from both the tracker and the calorimeter, the calorimeter measurement is subtracted when the tracker energy is used for the algorithm.

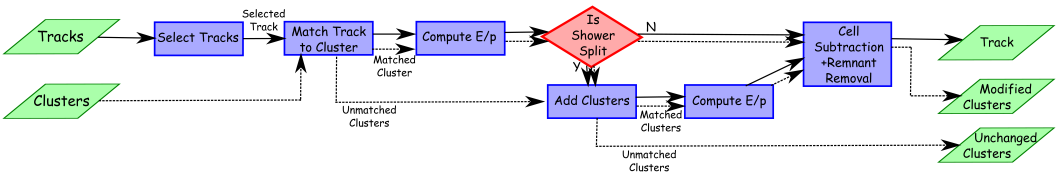


Figure 4.9: A flowchart showing the stages of the Particle flow algorithm [164].

Firstly, the selected tracks are matched to their respective topo-clusters. Since energy can be deposited in multiple clusters due to a split shower, the algorithm adds the relevant extra clusters to a track/cluster system based on their respective probabilities of being associated with the given shower. To perform the subtraction of the calorimeter measurement, the average energy deposited by a particle, $< E_{\text{dep}} >$ is computed. It is calculated as the product of the measured momentum p^{track} and the expected value of $E^{\text{cluster}}/p^{\text{track}}$. $< E^{\text{cluster}}/p^{\text{track}} >$ is the sum of energies of all topo-clusters associated with single particle samples without pileup within a cone of $\Delta R = 0.4$ around the position of the track. In cases where $< E_{\text{dep}} >$ is greater than the energy of the matched clusters, the topo-clusters are removed. For the remaining cases, a cell-by-cell energy subtraction is performed. Subsequently, if the remaining energy is less than $1.5\sigma(E_{\text{dep}})$, the jet constitutes of a single particle track-cluster system. Subsequently, this remnant energy coming from shower fluctuations is removed. All the remaining tracks and clusters are then fed as inputs to the anti-k_t jet reconstruction algorithm. The calibration for these jets is performed in a a similar way as described above for EM-Topo jets. The jet energy response for PFlow jets as a function of η_{det} for jets of various reconstructed energies is shown in Fig. 4.10. A comparison of the combined *in situ* correction i.e. jet response ratio in data and MC for PFlow+JES vs EM+JES is depicted in Fig. 4.11.

4.3.3 Large-R jets

Final states with high p_{T} are associated with massive particles like W-bosons. In most instances, these signatures are hadronic decays of the said particles. The high p_{T} of the decays products results in their collimation or Lorentz boost in the direction of the parent particle. A schematic of the reconstruction and calibration of large-R jets, the details of which will be discussed below, is shown in Fig. 4.12.

Such hadronic decays are reconstructed using the anti-k_t algorithm with a large radius parameter, $R = 1$. The inputs to the algorithm are three dimensional topological clusters.

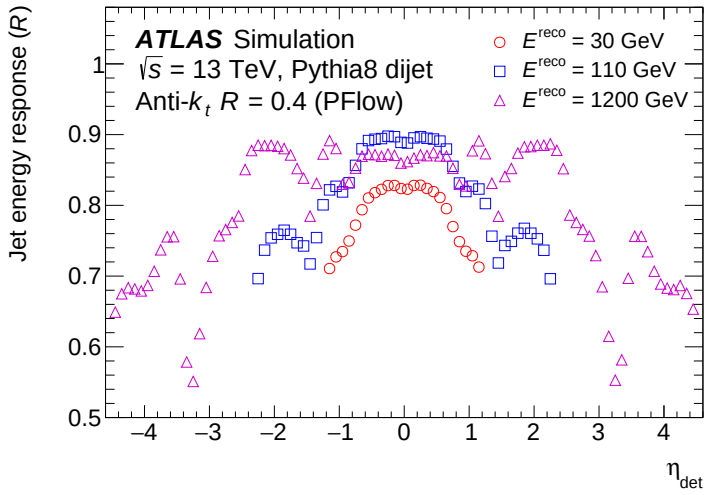


Figure 4.10: Jet energy response as a function of η_{det} for jets with reconstructed energy of 30, 110, and 1200 GeV with PYTHIA 8 sample [165].

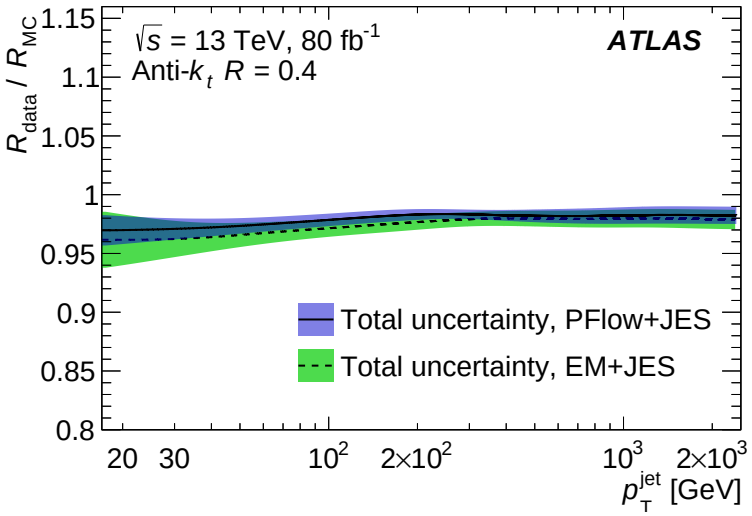


Figure 4.11: A comparison of the combined *in situ* correction i.e. jet response ratio in data and MC for PFlow+JES vs EM+JES [165].

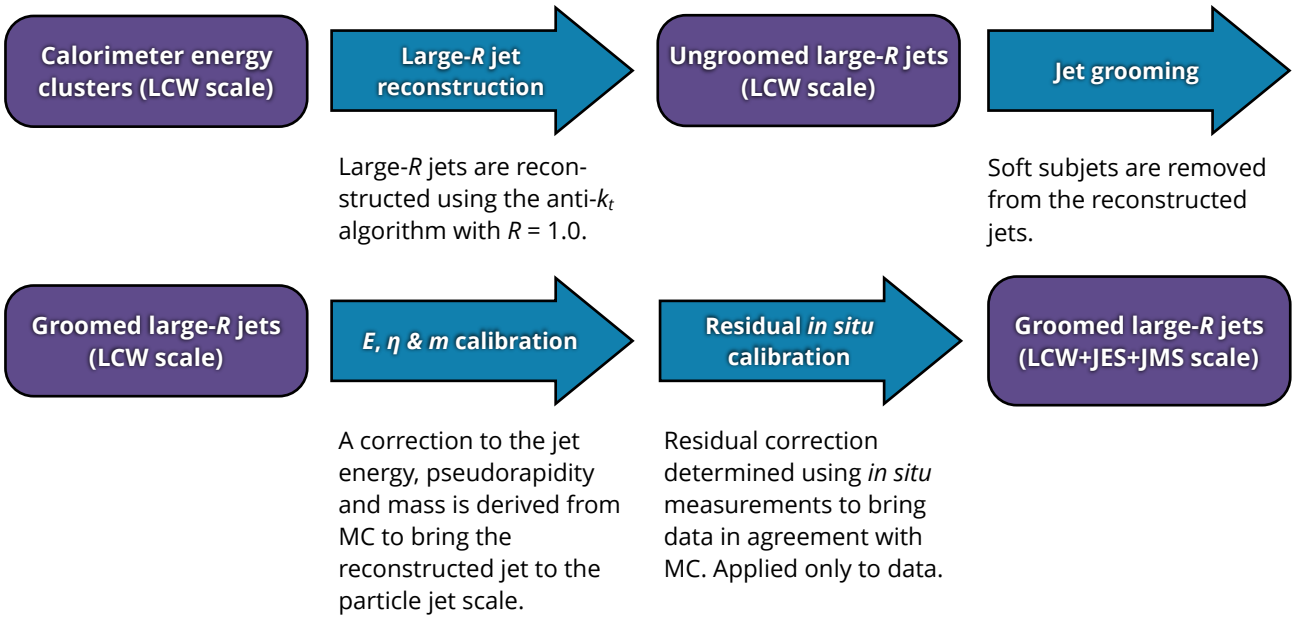


Figure 4.12: A schematic of the reconstruction and calibration of large-R jets used for the analysis in this thesis [166].

These energy clusters are calibrated to the hadronic scale using Local Cluster Reweighting (LCW) [148]. LCW addresses various inefficiencies of the cluster-by-cluster reconstruction of the calorimeter signal. Firstly, the likelihood of a cluster being generated from a electromagnetic (EM) shower, $\mathcal{P}_{\text{clus}}^{\text{EM}}$ is calculated. Thus, $1 - \mathcal{P}_{\text{clus}}^{\text{EM}}$ is the corresponding likelihood for hadronic interactions. The calibration weights are applied dynamically to avoid inefficiencies originating from misclassification. The per-cell weight that is applied for the calibration is

$$w_{\text{cell}}^{\text{cal}} = \mathcal{P}_{\text{clus}}^{\text{EM}} \cdot w_{\text{cell}}^{\text{em-cal}} + \left(1 - \mathcal{P}_{\text{clus}}^{\text{EM}}\right) \cdot w_{\text{cell}}^{\text{had-cal}} \tag{4.5}$$

Here, $w_{\text{cell}}^{\text{em-cal}}$ and $w_{\text{cell}}^{\text{had-cal}}$ are the electromagnetic and hadronic weights applied by the respective calibrations. A schematic of the LCW calibration is shown in Fig. 4.13.

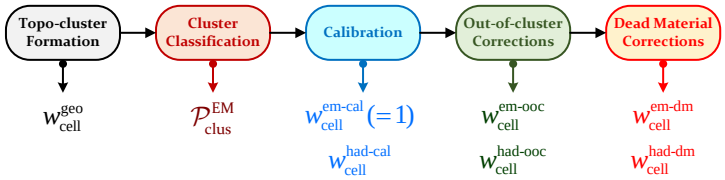
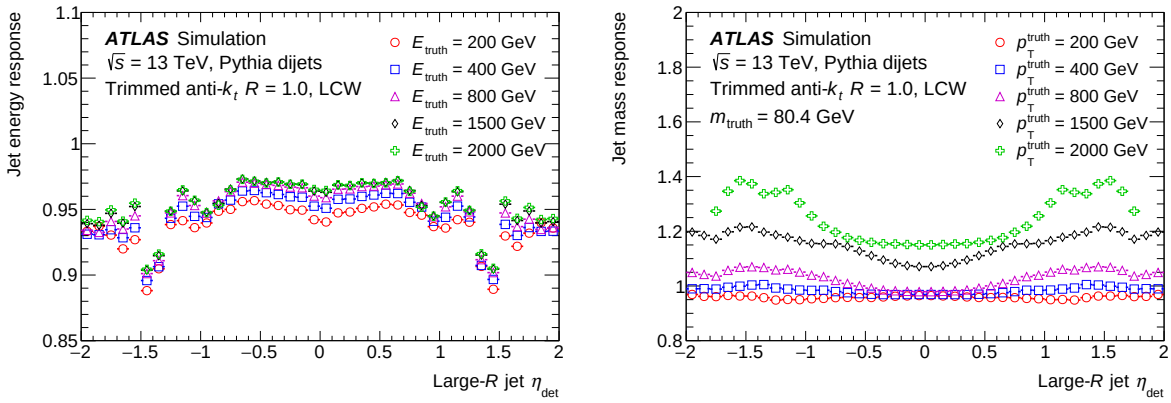


Figure 4.13: A schematic of the LCW calibration method. The corrections include calibrations, out-of-cluster (ooc), and dead material (dm) corrections [148].

After the initial reconstruction, the large-R jets are groomed using a "trimming" algorithm [167]. This procedure mitigates the effects of pile-up and soft emissions. It entails

reclustering the jet constituents using the k_t algorithm into subjets with $R_{\text{sub}=0.2}$. All subjets with $p_{\text{T}}^{\text{subjet}}/p_{\text{T}}^{\text{jet}} < 0.05$ are removed and the jet p_{T} is recalculated from the remaining constituents.

Subsequently, a simulation-based calibration is done against the reference of particle level jets. These jets are simulated by clustering particles in hard scatter interactions. The jet energy scale of the groomed jet is calibrated to the corresponding energy of particle level jets in a sample of simulated dijets. This sample is also used to correct for the η of the jet. Finally, the combined mass [166] of the jet is also calibrated in a similar manner. All these calibrations are applied as a parametrised function of kinematic variables of the large-R jet such as its energy and η_{det}. The jet energy and mass response for jets with varying truth energy and p_{T} are shown in Fig. 4.14[3].



(a) (b)

Figure 4.14: The response for (a) jet energy with truth energies varying from 200 to 2 000 GeV and (b) jet mass for jets with the truth mass of the W-boson and $p_{\text{T}}^{\text{truth}}$ varying from 200 to 2 000 GeV respectively [166].

After applying the simulation-based calibrations, the residual *in situ* calibrations for jet p_{T} and mass are performed *in situ* by using momentum conservation against a reference object that recoils against the jet. To ensure that the jet response is a uniform function of η_{det}, the jet calibration of central reference jets ($|\eta| < 0.8$) determined in dijet events is propagated to the forward detector region ($0.8 < |\eta| < 2.5$). The η-intercalibration method employed for large-R jets is similar to that used for small-R jets. Henceforth, three balance methods, Z+jet, γ+jet, and the multijet balance methods are applied to calibrate the absolute p_{T} scale of the jet. The recoiling systems for these methods are $Z \to e^-e^+/\mu^+\mu^-$, a photon, and calibrated low p_{T} small-R jets respectively.

The Z/γ+jet method balances the p_{T} of the jet against the $p_{\text{T}}^{\text{ref}}$ i.e. $p_{\text{T}}^{Z/\gamma} \cdot |cos\Delta\phi|$, where $\Delta\phi$ is the azimuthal angular separation between the jet and the direction of the Z-boson or the photon. The p_{T} range of the Z-boson for the Z+jet method is up to 500 GeV. Since,

[3] The ATLAS collaboration calibrates large-R jets with $p_{\text{T}} > 200$ GeV and $|\eta| < 2.5$.

the cross-section for γ+jet production is higher than the Z+jets mode, the γ+jet method can probe and thus calibrate large-R jets with higher p_{T}. These two methods only suffice for jets with p_{T} up to 1 TeV. For jets with higher p_{T}, the multijet balance method is employed. It balances the jet p_{T} against a recoil system consisting of small-R jets with a minimum p_{T} of 30 GeV. It is required that the leading small-R jet of the recoil system can have no more than 80% of the p_{T} of the probe jet.

The *in situ* jet mass calibration is done by using two methods, the R_{trk} double ratio method and the forward folding method. R_{trk}, the average calorimeter to track response is defined as $< p_{\mathrm{T}}^{\mathrm{calo}} / p_{\mathrm{T}}^{\mathrm{track}} >$. The double ratio of both $R_{\mathrm{trk}}^{p_{\mathrm{T}}}$ and $R_{\mathrm{trk}}^{\mathrm{m}}$ (the superscript m denotes the initial ratio for mass instead of p_{T}.) in simulations to data is used to correct for the jet mass scale. $R_{\mathrm{trk}}^{\mathrm{m}}$ as a function of the jet p_{T} when compared in data and MC for three distinct generators is depicted in Fig. 4.15. The forward folding method comprises of fitting the W-boson mass and top-quark mass peaks and jet mass response within data and simulations using a χ^2 fit. Finally, all these calibrations are combined using methods described in Ref. [168]. A smoothing is applied to remove any spikes in the combined calibrations due to statistical fluctuations. The combined data to MC jet p_{T} response is shown in Fig. 4.16

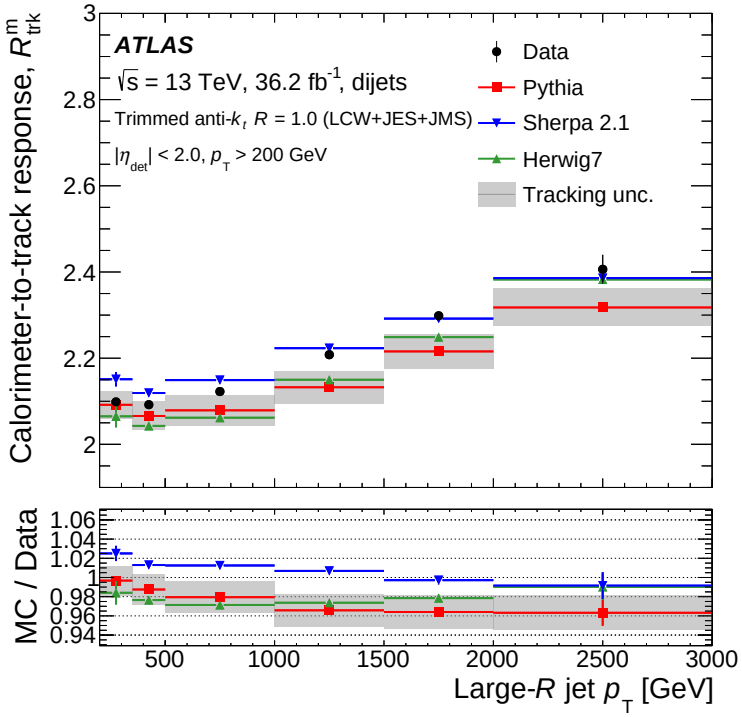


Figure 4.15: Comparison of $R_{\mathrm{trk}}^{\mathrm{m}}$ in data and MC for three different MC generators [166].

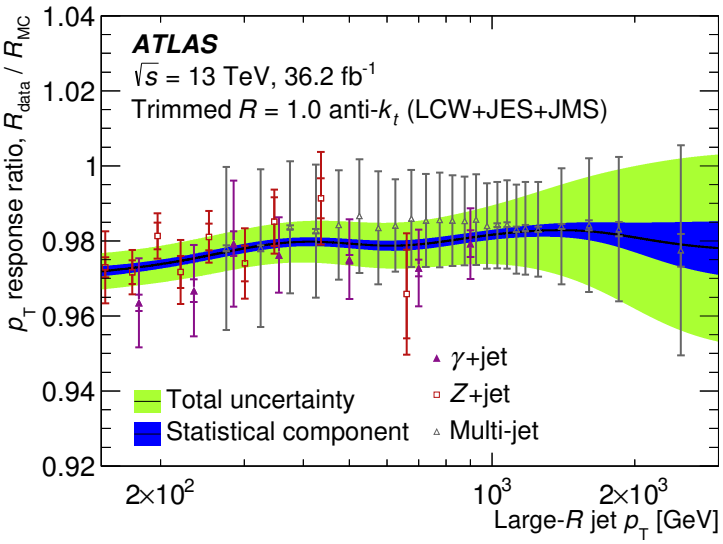


Figure 4.16: The combined data to MC jet p_{T} response plotted as a function of the large-R jet p_{T}. The individual calibrations are also shown here [166].

4.3.4 Jet tagging

Jet vertex tagging

The jet vertex tagging (JVT) [169] algorithm selects jets originating only from the primary vertex of an event. By choosing the jets from hard scattering interactions, the jet vertex tagger mitigates the effects of in-time pileup. The JVT discriminant is based upon a k-nearest neighbour algorithm which builds a two dimensional likelihood with $\mathrm{R}_{p_{\mathrm{T}}}$ and corrJVF. Here, $\mathrm{R}_{p_{\mathrm{T}}}$ quantifies the ratio of scalar sum of p_{T} of tracks originating from the primary vertex and the p_{T} of the calibrated jet and corrJVF is the jet vertex fraction corrected for the dependence on number of primary vertices.

b-tagging

In both the analyses done for this thesis, the final state contains a b-quark which hadronises into a jet. The b-jets are identified by the ATLAS collaboration using classification algorithms. In the leptonic analysis with the 2015-2016 ATLAS data, a boosted decision tree (BDT) based algorithm called MV2 [170] is used. For the hadronic analysis with the 2015-2018 data, a recurrent deep neural network based algorithm called DLIr [171] is deployed to identify b-jets.

The B-hadrons generally have a relatively long lifetime, thus decaying in times in order of 1.5 ps. This implies that such a hadron with a p_{T} of 70 GeV would have a decay length, $< L_{xy} >= 6.4$ mm in the transverse plane before decaying into soft leptons at a secondary vertex [172].

A schematic of such a decay is shown in Fig. 4.17. Three low-level algorithms construct 22

variables which are inputs to the aforementioned classification algorithms in addition to 2 kinematic properties of the jet (p_{T} and η). These are as following

1. IP2D/IP3D tagger: The impact parameter (IP) shown in Fig. 4.17 shows the impact parameter in the transverse plane. The IP2D algorithm takes into account the signed transverse impact parameter significance whereas IP3D uses both signed transverse and longitudinal impact parameters of the B-hadron decays to calculate the ratio of b-jet, c-jet, and light flavour jet probabilities for each track.

2. Secondary vertex (SV) tagging: The SV tagging algorithm finds the singular secondary vertex associated with the decay of a B-hadron.

3. Topological multi-vertex algorithm: The `JetFitter` algorithm reconstructs the complete decay chain by using the weak B-hadron and C-hadron decays.

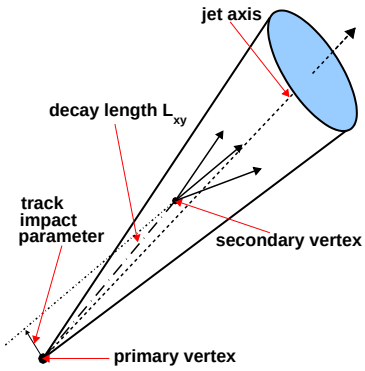


Figure 4.17: A schematic of a B-hadron decay into soft leptons resulting in the creation of a secondary vertex. The track impact parameter is the closest approach distance between the extrapolation of the track and the primary vertex [172].

The MV2 algorithm uses 24 low-level variables to classify the signal (b-jets) from the background (c-jets and light flavoured jets) with a BDT. The MV2c10 algorithm uses 7% of c-jets in the $t\bar{t}$ training sample. The value of the discriminant used for identifying b-jets varies according to the selection efficiency. For 60% and 80% identification efficiencies in 2015-2016 data, the value of the cut is 0.18 and 0.93 respectively [173].

Given the recent advancements in deep learning, the DL1 algorithm replaces the BDT with a deep feed forward neural network. The inputs for this algorithm are the ones used in MV2 and c-tagging `JetFitter` variables. The IP3D algorithm does not take into the correlations among the properties of the tracks inside the decay of a B-hadron. A recurrent neural network (RNN) is capable of learning the sequential dependence among variables.

Thus, the outputs of a new tagger called RNNIP which learns the interdependencies among the tracks are combined with the discriminant of the DL1 tagger [171]. The combined tagger is referred as the DL1r tagger. For this tagger, the training sample was a hybrid of $t\bar{t}$ and Z' events. The inclusion of Z' events ramps up the response of the tagger in high and medium p_{T} range.

The cuts on the final discriminant based on b-jet identification efficiency in inclusive $t\bar{t}$ samples are named operating or working points. For both the taggers, correction factors are applied to the MC samples to compensate for differences between data and simulation in the b-tagging identification efficiencies for b-jets, c-jets, and light flavoured jets [174, 175].

W-tagging

In the hadronic analysis performed with the 2015-2018 ATLAS data, the final state involves a high-p_{T} W-boson. Such a high-p_{T} W-boson is reconstructed using a three variable cut-based W-tagger [176]. The three kinematic variables that are used to build this tagger are

1. m^{comb} : The combined jet mass of the large-R jet, m^{comb} is a weighted average of the calorimeter and track-assisted mass measurements of the jet [166].

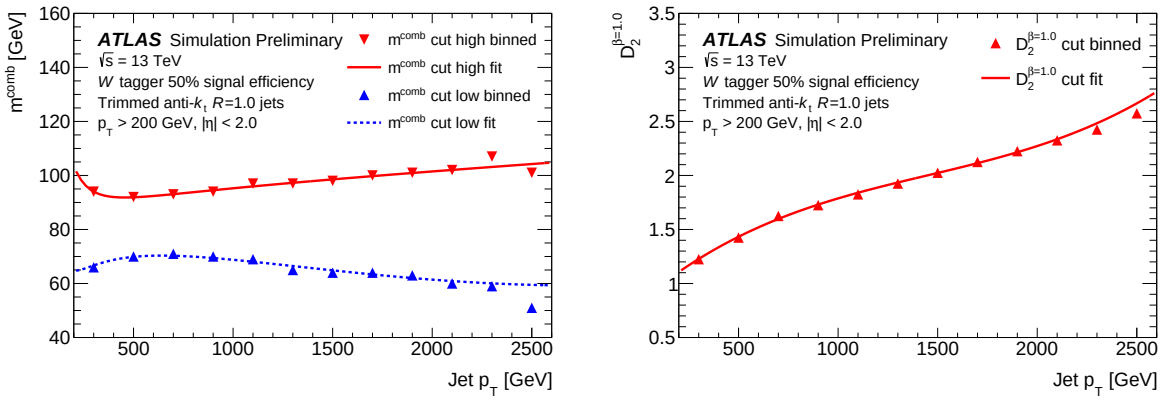
2. $D_2^{\beta=1}$: This variable is a ratio of the three-point energy correlation function, e_3^{β} and the two-point energy correlation, e_2^{β} cubed. Both the energy correlations are defined in Eq. (4.6), where p_{Ti} is the p_{T} of the i-th particle in the jet with n_J particles. Here, p_{TJ} is the transverse momentum of the given jet [177] and ΔR_{ij} represents the Lorentz boost invariant Euclidean distance among the i-th and j-th particle in the $\eta - \phi$ plane. For the W-tagger $D_2^{\beta=1}$ can be quantified by inserting $\beta = 1$ in Eq. (4.6).

$$
\begin{aligned}
e_2^{(\beta)} &= \frac{1}{p_{TJ}^2} \sum_{1 \le i < j \le n_J} p_{Ti} p_{Tj} \Delta R_{ij}^{\beta} \\
e_3^{(\beta)} &= \frac{1}{p_{TJ}^3} \sum_{1 \le i < j < k \le n_J} p_{Ti} p_{Tj} p_{Tk} \Delta R_{ij}^{\beta} \Delta R_{ik}^{\beta} \Delta R_{jk}^{\beta} \\
D_2^{(\beta)} &\equiv \frac{e_3^{(\beta)}}{\left(e_2^{(\beta)}\right)^3}
\end{aligned}
\tag{4.6}
$$

3. N_{trk} : This denotes the number of tracks ghost-associated [178] to the original ungroomed jet. Ghost-association accurately matches the tracks to the calorimeter subjets by making the tracks infinitesimally soft and including them in the jet finding process.

The cuts on these variables are optimised simultaneously as a function of jet p_{T} based on the identification efficiency of the W-boson. Smooth polynomial functions as defined in Ref. [179] are fitted in bins of p_{T} to produce a smooth W-tagger. The tagger is optimised at two working points (WP) , set at 50% and 80% signal efficiencies. The cuts for the 50%

WP for m^{comb} and $D_2^{\beta=1}$ are shown in Fig. 4.18. For m^{comb},a mass window requirement of $60\,\text{GeV} < m_{\text{comb}} < 100\,\text{GeV}$ is used. The highest value p_{T} dependent cut for N_{trk} is $N_{\text{trk}} = 26$.



(a) (b)

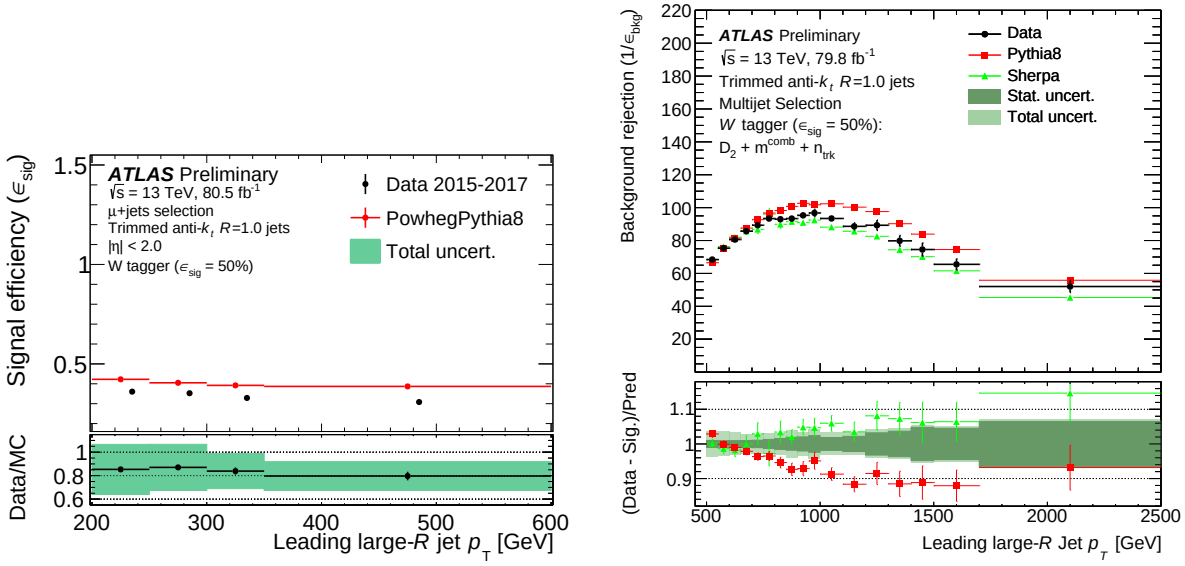
Figure 4.18: The binned and fitted cuts for (a) m^{comb} and (b) $D_2^{\beta=1}$ as a function of jet p_{T} for the three-variable W-tagger at 50% WP [176].

Scale factors quantify the relative difference in tagging efficiencies in data nd MC for both signal (W-jets) and background jets (γ+jets/multijets). The signal scale factors are evaluated using $t\bar{t}$ samples in the lepton-plus-jets channel, whereas the background scale factors are computed with γ+jet and multijet samples. The respective data/MC comparison of signal efficiency and background rejection for the W-tagger at the 50% WP is shown in Fig. 4.19.

4.4 Missing transverse energy

The conservation of momentum implies that the sum of momenta in the transverse plane of the ATLAS detector should be zero. However, weakly interacting particles like neutrinos escape the detector, thus creating a momentum imbalance. The missing transverse energy with magnitude $E_{\text{T}}^{\text{miss}}$ is the negative vector sum of all reconstructed hard objects (leptons, photons, hadronically decaying τ-leptons, and jets). A soft term is added to the calculation of $E_{\text{T}}^{\text{miss}}$ to ensure the inclusion of detector signals, residual energy from both hard scatter, unassigned to any of the previously used reconstructed objects. To mitigate the effect of pile-up, the soft term is calculated by performing track-matching to the primary vertex [180].

In this thesis, for the single-lepton analysis, EM-Topo small-R jets with $p_{\text{T}} > 20\,\text{GeV}$ that also pass the JVT criteria when the jet has $|\eta| < 2.4$ and $p_{\text{T}} < 60\,\text{GeV}$ are used for $E_{\text{T}}^{\text{miss}}$ calculation and for the hadronic analysis PFlow jets fulfilling the same criteria are used.



(a) (b)

Figure 4.19: The (a) signal efficiency and (b) as a function of leading large-R jet p_{T} for the three-variable W-tagger at 50% WP [176].

CHAPTER 5

Statistical methods

In this chapter, an overview of the statistical methods used in both the analyses performed for this thesis will be presented. In the field of experimental high-energy physics, statistical methods are used to compare the predicted parameters of a physical probabilistic model with the observed data. Depending on the definition of probability, the methods can be classified into two categories, frequentist and Bayesian. To quote Ref.[6],"In frequentist statistics, probability is interpreted as the frequency of the outcome of a repeatable experiment" and "In Bayesian statistics, the interpretation of probability is more general and includes degree of belief (called subjective probability). One can then speak of a probability density function (p.d.f.) for a parameter, which expresses one's state of knowledge about where its true value lies.". There is no clear mandate whether one set of methods is better or not. In this thesis, due to certain pragmatic reasons e.g. faster computing time, frequentist methods have been used. A detailed review of Bayesian methods used in particle physics can be found in Ref. [181]. An excellent comparison of frequentist and Bayesian methods is presented in Ref. [182].

5.1 Statistical inference

In any experiment, statistical inference is the process of comparing the observed data to the various parameters of a physical model. Consider an experiment where the data is characterised by a set of outcomes $\vec{x}$. Now, consider a model M which is described by a set of parameters $\vec{\lambda}$. The probability of the data, $\vec{x}$ given the model M, $P(\vec{x}|\vec{\lambda}, M)$, can be defined as a product of the individual probability distribution functions $f(x_i|\vec{\lambda})$. This product is called the *Likelihood*, $L(\vec{\lambda})$.

$$L(\vec{\lambda}) = \prod_{i=1}^{N} f(x_i|\vec{\lambda}) \tag{5.1}$$

Equation 5.1 is valid only if N is fixed and x_i is independent of each other. If N is not fixed and itself is a function of $\vec{\lambda}$, then the likelihood should include that dependency [6]. For example, if N follows a Poisson distribution with mean n, then equation 5.1 would be rewritten as

$$L(\vec{\lambda}) = \frac{n^N}{N!} e^{-N} \prod_{i=1}^{N} f(x_i|\vec{\lambda}). \tag{5.2}$$

Equation 5.2 is often referred to as the extended likelihood [6].

5.2 Parameter estimation

Parameter estimation is the way of finding the "best" value of a parameter in a model based on the available data, including the error on that estimate [183]. In frequentist statistics, parameter estimation is performed using mainly two methods, method of least squares and method of maximum likelihood. Here, only the latter is discussed as it has been solely used in this thesis.

Method of maximum likelihood

In the method of maximum likelihood (ML), the parameters $\vec{\lambda}$ are estimated by maximising the likelihood as defined in Eq. (5.1). Thus, the estimators $\vec{\hat{\lambda}}$, are the solutions to the following equation:

$$\frac{\partial \ln L}{\partial \lambda_i} = 0 \;\; i = 1..n \tag{5.3}$$

The likelihood being a product of functions, can take extremely large values and thus, maximising the log-likelihood is more computationally pragmatic. The ML estimator follows the criteria of being unbiased, consistent, robust, and efficient [6]. If the data are small, both Equation 5.1 or Equation 5.2 can be used. But in high energy physics experiments, the datasets used are generally quite large and thus, binned data is used. In collider experiments, events are binned as a function of a kinematic variable e.g. p_{T}. These events when binned in N bins follow a Poisson distribution in each bin. So the likelihood can be rewritten as

$$L = \prod_{i=i}^{N} {E_i}^{n_i} \cdot \frac{e^{-E_i}}{n_i!}, \tag{5.4}$$

$$E_i = \sum_{j=i}^{N_{\text{proc}}} \mathcal{L} \cdot \sigma_j \cdot \epsilon_{ij}, \tag{5.5}$$

$$= \mathcal{L} \cdot (\mu_{\text{sig}} \cdot \sigma \cdot \epsilon_i^{\text{signal}} + \sum_{j=i}^{N_{\text{bkgs}}} \mu_j^{\text{bkg}} \cdot \sigma_j \cdot \epsilon_{ij}), \tag{5.6}$$

where E_i and n_i are the expected and observed events in bin i respectively. Here, the expected number of events, E_i, is expressed as a sum over the expected events for the signal and all the background processes. In Equation 5.6, $\mathcal{L}$ is the integrated luminosity, σ is the theoretical cross-section, and ϵ represents the per-bin efficiencies in the phase-space of interest. These efficiencies are usually obtained from Monte-Carlo simulations. Here, μ_{sig}, the signal strength, is the main parameter of interest which is estimated. If the ML estimate of $\mu_{\text{sig}} \approx 1$, the existence of signal in data as predicted by the model is implied. A value of $\mu_{\text{sig}} \approx 0$ implies the absence of signal in data. The theoretical predictions for background processes are derived from the Standard Model and thus, in most cases $\mu_j^{\text{bkg}} = 1$ is used. Since the values of ϵ_{ij} are a function of detector acceptance and reconstruction algorithms, their predictions are not always exact. To account for background mismodelling, μ_j^{bkg} is also included as a parameter of interest in the likelihood. As the Standard Model is a highly precise model, the ML estimate of μ_j^{bkg} should be around 1. In reality, rather than maximising the log-likelihood, $-2\ln L$ is minimised. Using Equation 5.4, we get

$$-2 \cdot \ln L(\mu_{\text{sig}}, \mu_j^{\text{bkg}}) = -2 \cdot \sum_{i=1}^{N} \left[n_i \cdot \ln E_i(\mu_{\text{sig}}, \mu_j^{\text{bkg}}) - E_i(\mu_{\text{sig}}, \mu_j^{\text{bkg}}) \right] + \text{const.} \tag{5.7}$$

Since the inputs to our fit are finite Monte Carlo samples, the actual fit to data also takes into account the fact that ϵ_{ij} are generated using Poisson statistics. This leads to the likelihood being further extended. Eventually, the negative binned likelihood minimisation becomes an exercise in numerically solving multidimensional linear differential equations [184].

Nuisance parameters

The uncertainty on the parameter of interest, μ, has both statistical and systematic components. Both these components are incorporated into the likelihood as nuisance parameters. As shown in Eqs. (5.4) and (5.5), the Poisson likelihood is a function of expected events per bin. The experimental uncertainties arising from the modelling of the detector response, reconstruction algorithms of the physics objects affect the efficiencies ϵ_{ij}. Theoretical uncertainties impact the cross-section of both the backgrounds and the signal. Finally, there is also an

uncertainty on the measured integrated luminosity. The statistical uncertainty mainly comes from the limited size of our MC samples. A nuisance parameter is associated with each of these uncertainties and a Gaussian constraint is applied to them. Suppose the integrated luminosity, $\mathcal{L}$ has an uncertainty of 5% ($\sigma_{\mathcal{L}}$). Then, the Poisson likelihood can be written as,

$$L = \prod_{i=i}^{N} \mathcal{P}(n_i|\mu_i)\ \mathcal{G}(\mathcal{L}|\tilde{\mathcal{L}}, \sigma_{\mathcal{L}}) \tag{5.8}$$

for the measured luminosity $\tilde{\mathcal{L}} \pm \sigma_{\mathcal{L}}$.Here, $\mathcal{G}$ is a normalised Gaussian with a mean $\tilde{\mathcal{L}}$ and a standard deviation $\sigma_{\mathcal{L}}$ [185]. Thus, Eq. (5.7) becomes

$$-2 \cdot \ln L(\mu_{\text{sig}}, \mu_j^{\text{bkg}}, \mathcal{L}) = -2 \cdot \sum_{i=1}^{N} \left[n_i \cdot \ln E_i(\mu_{\text{sig}}, \mu_j^{\text{bkg}}) - E_i(\mu_{\text{sig}}, \mu_j^{\text{bkg}}) + \frac{(\mathcal{L} - \tilde{\mathcal{L}})^2}{2\sigma_{\mathcal{L}}^2} \right]. \tag{5.9}$$

For each uncertainty, a nuisance parameter is associated with it, which results in an extra penalty term on the negative log-likelihood. The Gaussian constraint on each of these nuisance parameters is an assumption arising from the central limit theorem, but some other function can also be used to constrain the nuisance parameters.

Morphing of shape uncertainties

Systematic uncertainties can be either normalisation or shape uncertainties. Normalisation uncertainties are uncertainties that affect the number of expected events uniformly across all bins e.g. the luminosity uncertainty described in the last section. All uncertainties that affect the per-bin efficiencies are normalisation, shape, or a combination of both uncertainties. These uncertainties are generally computed by altering certain physical parameters in the MC simulations. Only the $\pm 1\sigma$ deviations for these uncertainties are calculated and their effect on the spectrum of physical quantities e.g. p_{T} is calculated.

While minimising the negative log-likelihood, a likelihood scan is performed, and thus, the value of the efficiencies has to be known as a continuous function of a given uncertainty θ to calculate the appropriate penalty term on the log-likelihood. From MC simulations, we achieve the nominal and $\pm 1\sigma$ value of ϵ_{ij}. To define a continuous function, a morphing parameter f is introduced [185]. A simple linear relation can be written as

$$\epsilon_{ij} = \epsilon_{ij}^0 + f \frac{\epsilon_{ij}^+ + \epsilon_{ij}^-}{2}, \tag{5.10}$$

where ϵ_{ij}^0 is the nominal value of ϵ_{ij} and $\epsilon_{ij}^{\pm}$ are the respective values of the efficiencies for $\pm 1\sigma$ of the uncertainty. In `HistFactory` [186], the tool used for likelihood building is this thesis, a more complex function is used for interpolating the values of ϵ_{ij} for any value of

$\sigma(\theta)$.[1] For shape systematics, an exponential parameterisation is used.

$$\epsilon_{ij}(\theta) = \epsilon_{ij}^{0} \cdot \begin{cases} (\epsilon_{ij}^{+}/\epsilon_{ij}^{0})^{\theta} & \theta \geq 0, \\ (\epsilon_{ij}^{-}/\epsilon_{ij}^{0})^{\theta} & \theta < 0. \end{cases} \tag{5.11}$$

For normalisation uncertainties, the exponential parameterisation shown in Eq. (5.11) is used for $|\theta| > 1$, whereas for $|\theta| < 1$, a six degree polynomial of θ is used. The motivation for these choices are discussed in detail in Ref. [186].

Monte Carlo statistical uncertainties

The expected number of events in each bin is calculated from finite Monte Carlo simulations. This leads to a statistical uncertainty on the expected number of events per bin for each process. If E_{ij} events simulated for bin i and process j can be written as $\beta_{ij}\alpha_i$, where α_i is the nominal rate and β_{ij} are the efficiencies, Eq. (5.4) can be rewritten as

$$L = \prod_{i=i}^{N} \left[\left(\sum_{j=\text{proc}} \beta_{ij}\alpha_i \right)^{n_i} \cdot \frac{e^{-\sum_{j=\text{proc}} \beta_{ij}\alpha_i}}{n_i!} \right]. \tag{5.12}$$

Barlow and Beeston [183] proposed to treat each β_{ij} as a nuisance parameter with the nominal value centered at 1 with a Poisson constraint. This introduces a lot of new parameters that can significantly increase the computing time. Thus, only one nuisance parameter β_i is introduced per bin, which represents the overall statistical uncertainty on the total expected number of events. Consequently, in Eq. (5.12), the sum over the processes is dropped. For a Poisson distribution, $\text{Var}[\tilde{n}_i] = \sqrt{\beta_i\alpha_i}$, which leads to the following relation for the observed number of events for bin i n_i[2].

$$n_i = \alpha_i = (1/\sigma_i^{\text{stat unc.}})^2, \tag{5.13}$$

where $\sigma_i^{\text{stat unc.}}$ is the corresponding statistical uncertainty arising due to finite sample size of the Monte Carlo simulations. The Poisson constraint implies that for each β_i, the nominal value, $\beta_i^0 = 1$, and the uncertainties $\beta_i^{+,-} = 1 \pm \sigma_i^{\text{stat unc.}i}$. Each of these uncertainties adds a penalty term in the likelihood similar to Eq. (5.9).

5.3 Hypothesis testing

In particle physics experiments, one of the main goals is to find new particles and processes hypothesised by various new theories. These new theories e.g. the theories predicting

[1] From here onwards θ and $\sigma(\theta)$ will be used synonymously.

[2] $\tilde{n}_i$ is the random variable for the observed events in bin i.

vector-like quarks are tested against already proved theories e.g. the Standard Model to check if these new theories can be accepted or not. This is called hypothesis testing.

While searching for new particles like VLQs, the hypothesis that there are no VLQs is treated as the background-only hypothesis or the null hypothesis. At the LHC, if VLQs are present in data, they will manifest in terms of excess events in addition to the backgrounds predicted by the SM. Consequently, the background-only hypothesis (H_0) is tested against the signal plus background hypothesis (H_1). While testing two hypotheses, either the null hypothesis can be rejected or accepted. If a statement is made regarding the validity of H_0, two types of error can happen.

1. Type I error: H_0 is rejected although it is true. The probability of this happening is denoted by α.

2. Type II error: H_0 is accepted although it is false. The probability of this happening is denoted by β.

The probabilities of these instances happening or not happening is depicted in Table 5.1.

	H_0 is true	H_0 is false
H_0 is rejected	Type I error (Probability α)	Right decision (Probability $(1-\beta)$)
H_0 is accepted	Right decision (Probability $(1-\alpha)$)	Type II error (Probability β)

Table 5.1: Probabilities of occurrence of Type I and Type II errors.

To quantify the difference between H_0 and H_1, an unbiased test-statistic Q is chosen. The choice of Q should fulfill the condition, $\alpha + \beta < 1$ because $\alpha + \beta > 1$ implies that the probability to make either type I or type II error is higher than taking the corresponding right decision. The Neyman-Pearson lemma [187] states that the likelihood ratio Q, is the most powerful test statistic i.e. H_0 (background-only hypothesis) is rejected with the highest probability or equivalently yields the lowest value of β when tested against a hypothesis H_1 (signal plus background hypothesis with $\mu_{\text{sig}} = \mu$) compared to other test statistics.

$$Q = \frac{L(H_1(\mu_{\text{sig}} = \mu))}{L(H_0(\mu_{\text{sig}} = 0))}. \tag{5.14}$$

p-Value and significance

The Q-value for observed data events Q_{obs} maps the outcome of the experiment to a single real-valued number. We can simulate a repitition of the experiment by creating pseudo-experiments. This can be achieved by drawing from a random Poisson distribution with mean N and setting that value as the observed events. A distribution $f(Q|\tilde{N})$ will be generated with $f(Q|\tilde{N} = N, \mu_{\text{sig}} = \mu)$ for a given hypothesis with $\mu_{\text{sig}} = \mu$. The p-value for

a given dataset under a hypothesis, is defined as the the probability for an equally or more 'extreme' outcome than what is observed assuming that hypothesis [188].

$$p = \int_{Q_{\text{obs}}}^{\infty} f(Q|\mu_{\text{sig}} = \mu)\, dQ. \tag{5.15}$$

Since $f(Q|\mu_{\text{sig}})$ is defined to be a monotonically increasing function of μ_{sig}, the p-value would be lower if the data agrees more with the signal plus background hypothesis H_1. Based on the value of Q_{obs}, if the background-only hypothesis H_0 is rejected, then the p-value would be synonymous to α, the probability of making a Type I error. The choice of p-value for which a definitive statement can be made about rejecting H_0 is different in each scientific sub-field. In particle physics, rejecting H_0 generally implies the discovery of a new particle. The threshold used to claim discovery is $p = \alpha = 2.87 \cdot 10^{-7}$. This value corresponds to the integral under a unit Gaussian from 5σ to infinity. Usually, rather than quoting the p-Value, the significance Z defined as

$$Z = \Phi^{-1}(\mathcal{G}(0,1)) \cdot (1 - p), \tag{5.16}$$

where $\Phi^{-1}\mathcal{G}(0,1)$, the inverse of the cumulative distribution of a unit Gaussian, is used for accepting or rejecting the background-only hypothesis. While searching for a new particle, values of 3σ and 5σ are used as thresholds for evidence and discovery of the signal respectively.

5.4 Limit setting

In a search for new particles, when the data agrees with the background-only hypothesis, a high p-value is achieved. Thus, rather than completely rejecting the signal plus background hypothesis, an upper limit is placed on the allowed values for the signal strength, μ, the parameter of interest in the fit. In cases of non-discovery, an upper limit is quoted on physical parameters like the production cross-section or the mass of the particle that is being searched.

In frequentist statistics, the Neyman construction [189] is used to define confidence intervals (CL), which is further used to set upper limits. An upper limit, μ_{up} is set at $(1 - \alpha)\%$ confidence level on the parameter μ if

$$P(Q < \mu_{\text{up}}|\hat{\mu}) = 1 - \alpha, \tag{5.17}$$

where Q is the test statistic being used for Neyman construction and $\hat{\mu}$ is the true value of μ. For 95% CL ($\alpha = 0.05$), this implies that the probability that $\hat{\mu}$ cannot be greater than μ_{up} is 95%.

The CL_s method

Experiments designed to search for new particles and processes do not always yield the desired sensitivity. If the observed number of events undershoots the expected number of events, the maximum likelihood estimator of the signal strength μ is negative. Using strict frequentist confidence levels, for such cases, even zero signal strength is excluded. Although these upper limits are consistent from a statistical perspective, they are "unphysical". Thus, the CL_s method [190] is used for setting limits on the signal strength.

Using Eq. (5.15), $p_\mu^{\text{s+b}}$ is defined as the p-value for the signal plus background hypothesis with signal strength μ and p_0^{b} as the p-value for the background-only (null) hypothesis ($\mu_{\text{sig}} = 0$). The statistic CL_s is defined as

$$CL_s = \frac{p_\mu^{\text{s+b}}}{1 - p_0^{\text{b}}}. \tag{5.18}$$

For a given theoretical model, a point in its parameter space is excluded at $(1-\alpha)\%$ confidence level if $CL_s \leq \alpha$. To find the exact upper limit μ_{up}, a scan of the parameter space is performed to find the point in the phase space where $CL_s = \alpha$. If μ is excluded at 95% or 68% confidence level, $\alpha = 0.05$ and $\alpha = 0.32$ respectively. When the data is less than expected number of events, $p_\mu^{\text{s+b}}$ decreases, but as p_0^{b} increases, $(1 - p_0^{\text{b}})$ decreases. This prevents $CL_s \leq \alpha$. Hence, an "unphysical" limit on the model parameter can be avoided.

The CL_s method being more conservative, has a higher coverage for the model that is being tested. Although the quantity CL_s is not a strict probability, it can be shown that CL_s yields identical values as Bayesian credible intervals do while using constant priors for Poisson counting experiments [190].

5.5 Profile likelihood fit: A toy example

In this section, a toy example will be used to demonstrate the fits similar to the ones presented in this thesis. The toy example demonstrated here can be extended to any of the complex fits that are performed in most ATLAS analyses.

Profile likelihood

The log-likelihood in Eq. (5.9) is written as a function of one nuisance parameter (NP). In reality, there are hundreds of NPs affecting the likelihood. So the likelihood L is a function of μ and the NP vector $\boldsymbol{\theta}$. The maximum likelihood estimators of μ and $\boldsymbol{\theta}$, $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the solutions to Eq. (5.3) for $L(\mu, \boldsymbol{\theta})$. The conditional maximum likelihood estimator $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ is defined as the estimate of $\boldsymbol{\theta}$, where μ is fixed in $L(\mu, \boldsymbol{\theta})$. The profile-likelihood ratio is

constructed as

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \tag{5.19}$$

This definition ensures that $\lambda(\mu)$ does not have an explicit dependence on the NP vector (θ). The process of using a fixed value of $\boldsymbol{\theta}$ for a given λ is known as "profiling" away the NPs. Using the profile-likelihood ratio, a test statistic $q(\mu) = -2\ln\lambda(\mu)$ is defined. It is evident that $q(\mu)$ follows the Neyman-Pearson Lemma. Hence, Eq. (5.15) can be rewritten as

$$p(\mu, \boldsymbol{\theta}) = \int_{q_{\text{obs}}}^{\infty} f(q(\mu)|\mu, \boldsymbol{\theta}) dq. \tag{5.20}$$

Asimov dataset

In order to compute the p-value for a given hypothesis using Eq. (5.15), the function $f(q(\mu))$ needs to be sampled. This sampling is done by generating pseudo-data. For a dataset with expected sensitivity of $\approx 5\sigma$, 10^7 pseudo-data points need to be generated, which is computationally expensive. Instead, a single fit is performed to a dataset called "Asimov dataset" [191]. For a Poisson likelihood with signal strength μ in bin i the Asimov data A_i can be written as

$$A_i = \mu \cdot E_i^{\text{sig}}(\theta = 0) + \sum_{j=\text{bkgs}} E_{ij}^{\text{bkg}}(\theta = 0). \tag{5.21}$$

As proven in Ref. [191], in the asymptotic limit, for this dataset, $\lambda(\mu) = \lambda(\mu_{\text{true}})$ and λ is also independent of $\boldsymbol{\theta}$. From Wilks' theorem [192], it further follows that the $f(q|\mu)$ is a chi-squared distribution at the asymptotic limit. Thus, the p-value for the hypothesis with signal strength μ for an observed value of q_{obs} of the test-statistic is

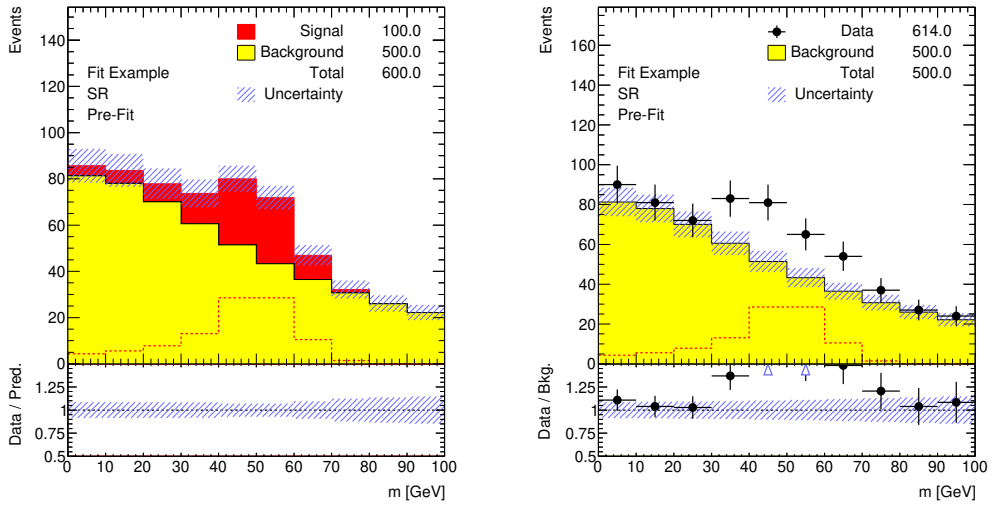
$$p_\mu = 2 \cdot (1 - \Phi\left(\sqrt{q(\mu)}\right)), \tag{5.22}$$

where Φ is the cumulative distribution function of the normalised unit Gaussian. In physics analyses, before performing the fit to data, a fit to the Asimov dataset is performed. The input and output of the Asimov fit for μ is exactly the same by definition. In BSM searches, $\mu = 0$ is used in the Asimov data whereas in SM searches $\mu = 1$ is set to construct the Asimov dataset.

Toy example

In this section, a fit will be performed to simulated data using one signal and one background samples in a single signal region (SR). These fits are done using `TRexFitter` [193], which uses `HistFactory` and `RooStats` [194] for likelihood building and minimisation respectively. This toy example is going to exemplify the ideas explained in the previous sections.

The mass distribution of the final state is used as the kinematic variable for fitting. A Landau function is used to parametrise the falling spectrum of the background, whereas a Crystal Ball function is used to describe the signal with mass 50 GeV. The data is simulated as a Poisson distribution with the total number of expected signal plus background events as its mean. Here, 100 signal and 500 background events are simulated. For the background distribution, three systematic uncertainties are assigned affecting each term in Eq. (5.5). A 2% luminosity uncertainty, a 2% cross-section uncertainty, and a shape dependent uncertainty e.g. jet energy scale are assigned to the background. The luminosity and jet energy scale uncertainty is also applied to the signal. In addition, the per bin Monte Carlo statistical uncertainty is also assigned. The pre-fit distributions of the signal and background, with all the uncertainties are shown in Fig. 5.1. The shape-dependent uncertainty used for the signal and background in shown in Fig. 5.2.



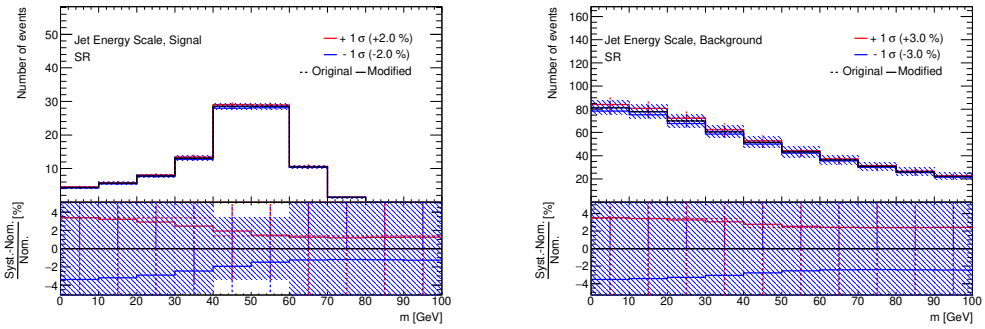
(a) Signal plus background distributions at pre-fit level with all uncertainties.

(b) Data and background distributions at pre-fit level.

Figure 5.1: Mass distribution of simulated signal, background, and data.

For the fit to Asimov data, by definition, the post-fit value is $\hat{\mu}_{\text{sig}} = 1$. The total uncertainty on μ_{sig}, σ_μ, should be close to the uncertainty on μ_{sig} for the corresponding fit to real data. The uncertainty is the square-root of the first diagonal element of the the covariance matrix of the fit parameters, which is the inverse of the matrix of the second derivative of the log-likelihood w.r.t. to the fit parameters. The nuisance parameters (NP) corresponding to the uncertainties, $\boldsymbol{\theta}$, which are profiled should yield a post-fit value $\hat{\boldsymbol{\theta}} = 0$ for the Asimov fit. Since the prior on θ_i is $\mathcal{G}(0,1)$, a post-fit value of the standard deviation significantly less than 1 implies that the uncertainty corresponding to the NP is overestimated. This is known as the fit constraining a NP.

The post-fit behaviour of μ_{sig} and $\boldsymbol{\theta}$ for both the Asimov and data fits are shown in Fig. 5.3 and Fig. 5.4 respectively. The post-fit mass distributions for the signal and background for



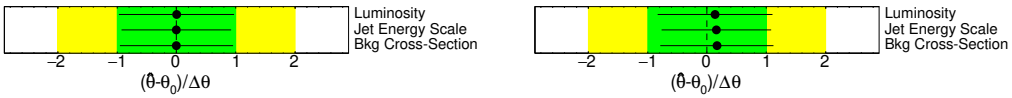
(a) (b)

Figure 5.2: Shape-dependent uncertainty for signal and background.

(a) Post-fit value of μ_{sig} for the fit to Asimov data.

(b) Post-fit value of μ_{sig} for the fit to observed data.

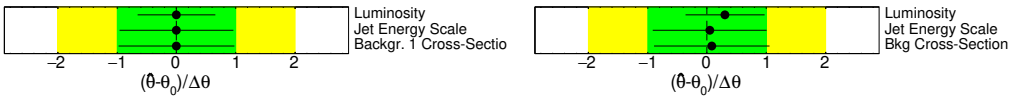
Figure 5.3: Post-fit values of the signal strength μ_{sig} for the Asimov and data fits.



(a) Post-fit value of NPs $\boldsymbol{\theta}$ for the fit to Asimov data.

(b) Post-fit value of NPs $\boldsymbol{\theta}$ for the fit to observed data.

Figure 5.4: Post-fit value of NPs $\boldsymbol{\theta}$ in the toy Asimov and data fits. Here, $\hat{\theta}$ is the profiled value, whereas θ_0 is the prior on each NP. $\Delta\theta$ represents the post-fit constraint on θ.



(a) Post-fit value of NPs $\boldsymbol{\theta}$ for the fit to Asimov data when the luminosity uncertainty is increased to 10%.

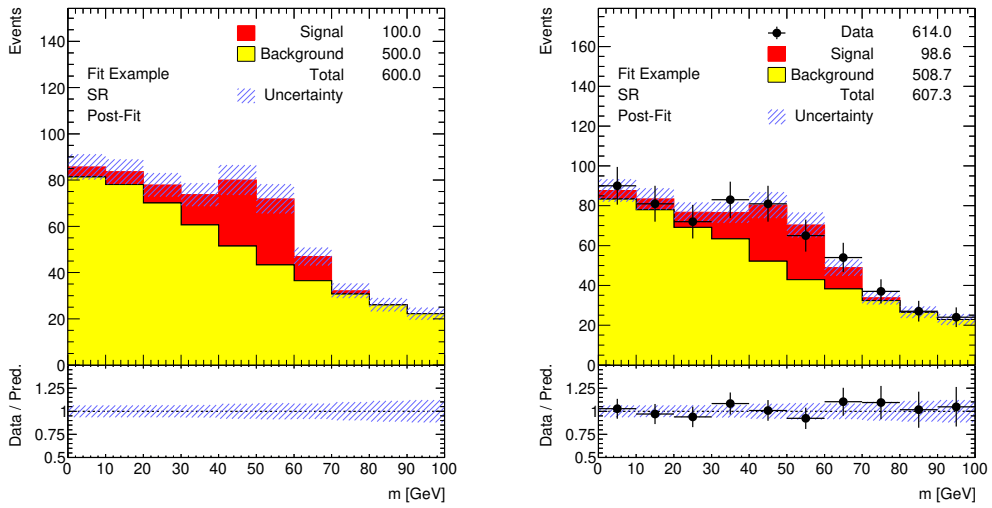
(b) Post-fit value of NPs $\boldsymbol{\theta}$ for the fit to observed data when the luminosity uncertainty is increased to 10%.

Figure 5.5: Post-fit value of NPs $\boldsymbol{\theta}$ for the fit to Asimov data when the luminosity uncertainty is increased to 10%. Here, $\hat{\theta}$ is the profiled value, whereas θ_0 is the prior on each NP. $\Delta\theta$ represents the post-fit constraint on θ. Due to the obvious overestimation, the luminosity NP is constrained.

both the fits are shown in Fig. 5.6. Hence, the observed $\hat{\mu} = 0.98$ is consistent with Figs. 5.1(a) and 5.6(b). The absence of any significant constraint on the NPs in Fig. 5.4 implies that none of the uncertainties are significantly overestimated. If the toy-example is altered to have a 10% luminosity uncertainty, a significant constraint can be noticed. The post-fit NPs for this fit setup are shown in Fig. 5.5. As the signal was successfully extracted from the data with $\hat{\mu} \approx 1$, the expected and observed significance computed from the Asimov and the real data should have similar values. The p-value and the significance for both these fits is shown in Table 5.2. Using these values, it can be inferred that there is strong evidence for a signal with mass 50 GeV in this dataset.

	Asimov data	Observed data
p-value	1.04×10^{-6}	2.16×10^{-6}
Significance	4.74	4.59

Table 5.2: Significance and p-value for accepting the signal plus background hypothesis.



(a) Post-fit mass distribution for the fit to Asimov data.

(b) Mass distribution for the fit to observed data.

Figure 5.6: Post-fit mass-distribution mass distribution for signal and background.

The dataset that was simulated for this toy example contained a signal but in most BSM searches the fitted value of $\hat{\mu}_{\text{sig}} \approx 0$. In such cases, an upper limit is set using the aforementioned CL_s method. The toy fits described above will be repeated but to a dataset, sampled from a Poisson distribution with the expected number of background events as its mean for each bin. Consequently, the Asimov dataset is created by setting $\mu = 0$ in Eq. (5.21). The real and Asimov datasets are shown in Fig. 5.7. Using the dataset shown in Fig. 5.7(a), although a fitted value of $\hat{\mu} = -0.03 \pm 0.18$ is achieved, $\mu = 0$ cannot be excluded. The

main reason for achieving a negative $\hat{\mu}$ is data undershooting the background around the signal peak. This is one of those cases for which the *CLs* method was devised. Using the CL_s method we achieve the following limits.

$$\mu_{\text{up}}^{\text{obs}} = 0.37, \tag{5.23}$$

$$\mu_{\text{up}}^{\text{exp}} = 0.39^{+0.17,+0.39}_{-0.11,-0.17}. \tag{5.24}$$

Here, $\mu_{\text{up}}^{\text{obs}}$ is the observed 95% CL_s limit and $\mu_{\text{up}}^{\text{exp}}$ is the 95% CL_s expected limit. The observed limit is calculated from data whereas the expected limit is derived from Asimov data defined using $\mu = 0$. From Eq. (5.18) and Eq. (5.22), it follows that

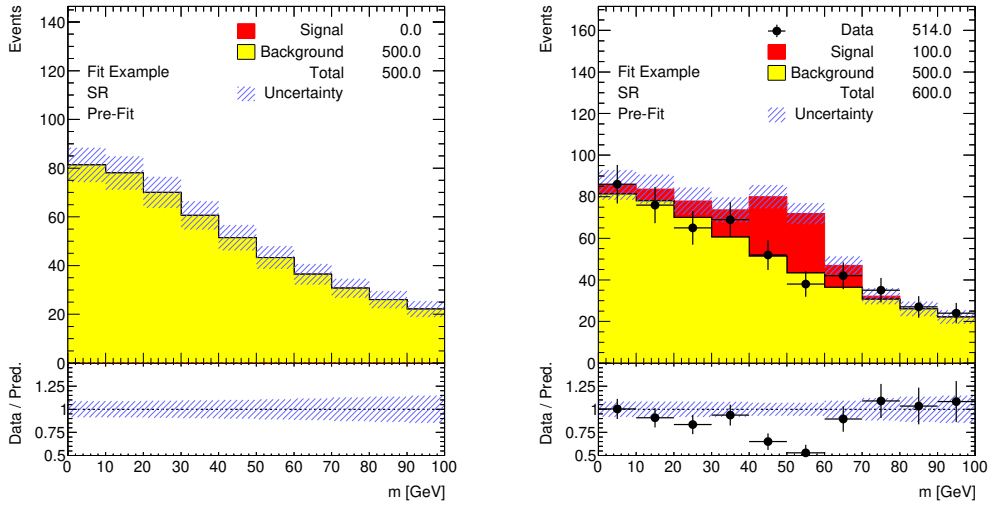
$$CL_s = \frac{1 - \Phi\left(\sqrt{q_\mu}\right)}{\Phi\left(\sqrt{q_{\mu=0}} - \sqrt{q_\mu}\right)}. \tag{5.25}$$

Thus, the 95% CL_s upper limit, μ_{up} is computed numerically by scanning over various values of μ to find the exact μ_{up} where $CL_s = 0.05$. The errors on $\mu_{\text{up}}^{\text{exp}}$ in Eq. (5.24) are the $\pm 1, 2\sigma$ error bands computed using

$$\mu_{\text{up}\pm N} = \sigma\left(\Phi^{-1}(1 - \alpha\Phi(N)) \pm N\right), N = 1, 2 \tag{5.26}$$

$$\sigma^2 = \frac{\mu^2}{q_{\mu,A}}, \tag{5.27}$$

where $q_{\mu,A}$ is the test statistic calculated from Asimov dataset for a given μ. This formula is derived from the asymptotic approximation for the Asimov dataset [191]. Since σ is a function of μ, Eq. (5.27) is solved numerically.



(a) Signal and background distributions at pre-fit level with all uncertainties.

(b) Data and background distributions at pre-fit level.

Figure 5.7: Mass distributions of simulated signal, background, and data with $\mu = 0$ at pre-fit level.

Part II

Single-lepton analysis

CHAPTER 6

Event selection

In this chapter, the analysis strategy and the event selection for the single-lepton analysis using the 2015-2016 ATLAS dataset corresponding to an integrated luminosity of 36.1 fb^{-1}will be discussed. The production cross-sections of the VLQ signals that have been searched for are some orders of magnitude smaller than the main backgrounds that decay into the same decay products. Kinematic regions with higher signal purity need to be defined to increase the chance of discovering the signal. Since, the optimisation of the kinematic cuts for this analysis was not the focus of this thesis, only a brief overview of the optimisation will be presented here.[1]

6.1 Analysis strategy

The analysis discussed here is a search for single production of a Y or T VLQ decaying to a Wb topology, where the W-boson subsequently decays leptonically. Kinematic cuts based on the event topology of the signal production are employed to find a phase-space with high signal to background ratio. The signal process which has been probed, as shown in Fig. 6.1, is characterised by a single lepton (the lepton can be either an electron or a muon), $E_{\mathrm{T}}^{\mathrm{miss}}$ originating from the neutrino of the W-decay, a b-jet originating from the VLQ decay, a second b-jet originating from gluon-splitting, and a light (u/d) flavoured jet. The physics objects used in this analysis are reconstructed using methods described in Chapter 4. If a T/Y VLQ exists, theoretical and experimental constraints (discussed in Section 2.6.2) require it to be relatively massive. Thus, the decay products of such a massive particle would have to be directionally opposite and have high-p_{T} to achieve energy-momentum conservation. The outgoing light-flavoured jet generally falls in the forward region of the detector and thus has high values of $|\eta|$. The second b-jet originating from the gluon splitting tends to fall outside of detector acceptance in most cases. The Feynman diagram shown in Fig. 6.1 justifies the fact that excluding the b-jet coming from the VLQ decay, additional hard jets

[1] Further details about this can be found in the PhD thesis of Dustin Biedermann from Humboldt University, Berlin [195].

(jets with $p_{\mathrm{T}} > 75\,\mathrm{GeV}$ for this analysis) should be very few in any signal event. Since, the light-flavoured jet or the second b-jet can be hard, rather than vetoing an event with a second hard-jet, events having a second hard-jet within a certain ΔR separation are vetoed. The largest background contribution while probing such single-top-like final states come from top-quark pair production ($t\bar{t}$), single top-quark production, and W-boson production in association with jets (W+jets). The modelling of the backgrounds and the signal will be discussed in the next section.

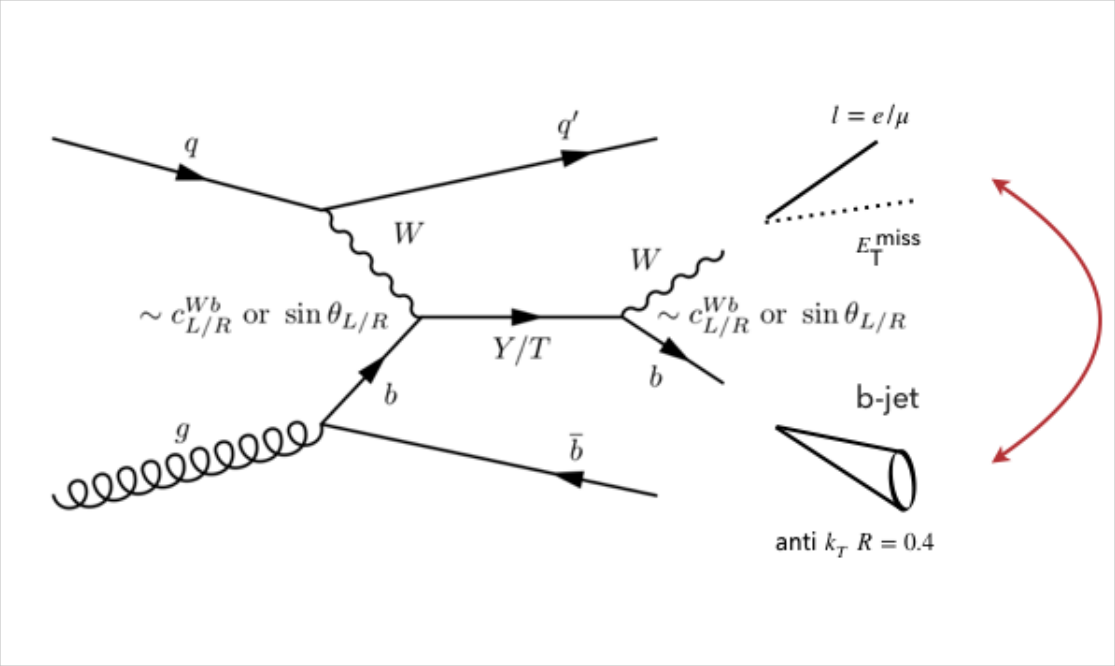


Figure 6.1: Leading order Feynman diagram for the $T/Y \to Wb$ decaying leptonically. The final state objects used for the search is also shown here. The two-sided red arrow signifies the high angular separation between the lepton and the leading b-jet that is necessary to achieve energy-momentum conservation during the decay of a relatively high-mass Y/T.

In addition to the signal enriched region, two background enriched control regions are defined. The control regions are used to constrain a potential mis-modelling of the two main backgrounds namely $t\bar{t}$ and W+jets. A binned maximum likelihood fit using both the signal and control regions is used to test the signal-plus-background (SM) hypothesis versus the background-only hypothesis to find the existence of a possible signal. The reconstructed vector-like quark mass (m_{VLQ}) is used as a discriminant to separate signal against all the SM backgrounds. The m_{VLQ} is calculated by performing a vector sum of the the Lorentz vectors of the electron, leading b-jet, and the neutrino. It is expected that $E_{\mathrm{T}}^{\mathrm{miss}}$ equals the projection of the neutrino four momentum in the transverse plane. The z-component of the neutrino is calculated by solving the quadratic equation resulting from setting the Lorentz-invariant mass of the lepton-neutrino system to the mass of the W-boson. If no real solution exists, the

$E_{\mathrm{T}}^{\mathrm{miss}}$ vector is varied by a certain value such that exactly one real solution exists. In cases where the quadratic equation yields two real solutions, the solution with the lowest value of $|p_{\nu,z}|$ is chosen.

6.2 Data, signal, and background modelling

Data samples

The data used for this analysis was pp collision data collected by the ATLAS detector in 2015 and 2016 at $\sqrt{s} = 13\,\mathrm{TeV}$. The peak recorded luminosity during this period was $1.37 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with a bunch spacing of 25 ns. The average number of simultaneous pp interactions per bunch crossing, or "pile-up" was measured to be approximately 24. To ensure high data quality, data events pertaining to the Good Run Lists (GRL) [1] are chosen. The integrated luminosity of these GRL data events delivered by ATLAS in 2015-16 was $36.1\,\mathrm{fb}^{-1}$.

Background processes

The simulated background processes used in the analysis are top-pair production ($t\bar{t}$), W/Z+jets, single-top, and other SM processes such as diboson production, $t\bar{t}$ V (top-pair production associated with a W/Z boson), and $t\bar{t}$ H (top-pair production associated with a Higgs boson). Monte Carlo (MC) simulation of these processes were performed either through the full GEANT4 [143] simulation of the ATLAS detector. or through a faster simulation process namely AFII which uses a parametrised version of the calorimeter showers [196]. The effects of in-time and-out-of-time pile-up arising from additional pp collisions in same and neighbouring bunch crossings were simulated in PYTHIA 8.186 [197] by overlaying minimum bias interactions. EVTGEN [198] was used to simulate decays of heavy flavour hadrons for all processes except the ones that are generated using the SHERPA [199] MC generator. A list of various parameters used for the simulated background processes are shown in Section 6.2. These parameters include the Parton Density Function (PDF) set, the tuned parameters to model multi-parton interactions, and the order of perturbative QCD. Certain backgrounds such as non-prompt leptons, jets misidentified as leptons which are collectively called fakes, were derived using data-driven methods. For this analysis, the Matrix Method developed by the Top Physics group of the ATLAS collaboration was used to model these data-driven backgrounds.

[1] In the course of standard LHC operations, the ATLAS detector may encounter various challenges that hinder the proper functioning of its sub-detector systems. Although most of these challenges do not significantly affect data collection, there are certain detector conditions that render luminosity blocks unsuitable for analysis. To address this, Good Runs Lists (GRLs) are meticulously maintained using an algorithm to document the luminosity blocks of data suitable for physics analyses.

Simulated signal modelling

The Y/T VLQ signal samples used in this analysis were simulated using MADGRAPH5_aMC@NLO 2.2.3 [200] with the four flavour scheme at leading order. The NNPDF2.3 [55] sets were used as parton distribution functions while PYTHIA 8.210 [201] was used to modelling parton shower and hadronisation. As both T and Y decay to a Wb final state, the kinematic properties of both the signals are similar, and thus only Y signals were produced and the results corresponding to the Tqb final state were derived from them. This was also motivated by the fact that the acceptance of $T \to Zt/Ht$ is negligible for this final state. The Yqb signals were produced with an equal left and right handed coupling parameter $k_T = 0.5$ corresponding to the VLQ model of Ref. [14]. The coupling parameters k_T and $c^{Wb}_{\mathrm{L/R}}$ (as introduced in Section 2.6) can be written in terms of each other as $k_T \approx c^{Wb}_{\mathrm{L/R}}/\sqrt{2}$. The leading order cross-sections used for the simulated signals were normalised to the next to leading order calculation performed in Ref. [12]. This normalisation is performed with the the narrow-width approximation (NWA). Since the production cross-section and decay widths of the VLQ signals are coupling and mass dependent, with higher mass and couplings the NWA is no longer valid. The NWA to non-NWA correction done at leading order in MADGRAPH5_aMC@NLO was propagated to next to leading order. Signal samples were produced in the range of 800 to 2 000 GeV with a step size of 100 GeV.

Modelling of interference effects

Due to similarity in the final topologies, the SM backgrounds and the VLQ signal can have significant interference effects among themselves (See Section 2.6.3 for details)[2]. As the shape of the invariant mass distribution is a function of the coupling and the interference a reweighting procedure was developed specially for this analysis[3] to reweight the signal distribution with interference to any desired point on the mass-coupling grid. The reweighting factors are per-event weights that were applied to each signal event of the simulated signal samples. The factors can be represented as

$$r\left(m_{Wb}; c^{Wb}, c_0^{Wb}\right) = \frac{K_{\mathrm{VLQ}} f_{\mathrm{VLQ}}\left(m_{Wb}; c^{Wb}\right) + \sqrt{K_{\mathrm{SM}} \cdot K_{\mathrm{VLQ}}} f_{\mathrm{I}}\left(m_{Wb}; c^{Wb}\right)}{f_{\mathrm{VLQ}}\left(m_{Wb}; c_0^{Wb}\right)} \tag{6.1}$$

where K_{VLQ} is derived from Ref. [12] and K_{SM} is assumed to be unity. The functions f are m_{Wb} distributions derived from high statistics samples of the target coupling c^{Wb}, and the

[2] The interference effects of the T/Y VLQ signal with the SM backgrounds as used in this analysis were overestimated. The correct magnitude of the interference is calculated in Ref. [110] which was published two years after the publication of this analysis. Certain methods were developed for this thesis and for this analysis assuming large interference effects among the SM backgrounds and the VLQ signal. In hindsight, while these methods may not have been necessary for dealing with the interference effects of this analysis, but they can be useful for other new physics searches with large interference effects among signal and background.

[3] This procedure was developed by Tobias Kupfer, Didier Alexandre, Ferdinand Schenck, and Dustin Biedermann at TU Dortmund and Humboldt University of Berlin.

nominal coupling c_0^{Wb}. A polynomial fit is performed to the histogram of r to achieve a smooth reweighting function.

Table 6.1: MC generators used to simulate the signals and different background processes. Parameters such as the parameter tune for the underlying event, PDF set, and the highest-order perturbative QCD (pQCD) accuracy used for the normalisation of each sample is listed here. All background processes were generated at NLO in QCD. The LO cross-sections calculated for the Yqb signal processes in the simulation were normalised to the NLO theoretical cross-section taken from Ref. [202]

Process	MC Generator + parton showering/hadronisation	Tuned parameters	PDF set	Cross-section order in pQCD
$\boldsymbol{Yqb}$	MADGRAPH5_AMC@NLO 2.2.3 + PYTHIA 8.210	A14	NNPDF2.3	NLO
$\boldsymbol{t\bar{t}}$	POWHEG-BOX 2.0 + PYTHIA 6.428	P2012	CT10	NNLO+NNLL
Single top	POWHEG-BOX 1.0 + PYTHIA 6.428	P2012	CT10	NNLO+NNLL
Dibosons WW, WZ, ZZ	SHERPA 2.1.1	Default	CT10	NLO
$\boldsymbol{W/Z}$ **+ jets**	SHERPA 2.2.0	Default	CT10	NNLO
$t\bar{t}V$	MADGRAPH5_AMC@NLO 2.2.3 + PYTHIA 8.210	A14	NNPDF2.3	NLO
$t\bar{t}H$	MADGRAPH5_AMC@NLO 2.2.3 + HERWIG++ 2.7.1	CTEQ6L1	CT10	NLO

6.3 Pre-selection

Before optimising the signal-region kinematic cuts, a broad event-selection is employed based on the topology of the final state as shown in Figure 6.1. This set of cuts, in addition to certain event-quality cuts, is defined as the preselection region. The preselection region is characterised by events which fulfil the following criteria:

1. Events should contain a primary vertex (see Section 4.1) , the jets in the event should match standard ATLAS quality criteria [203].

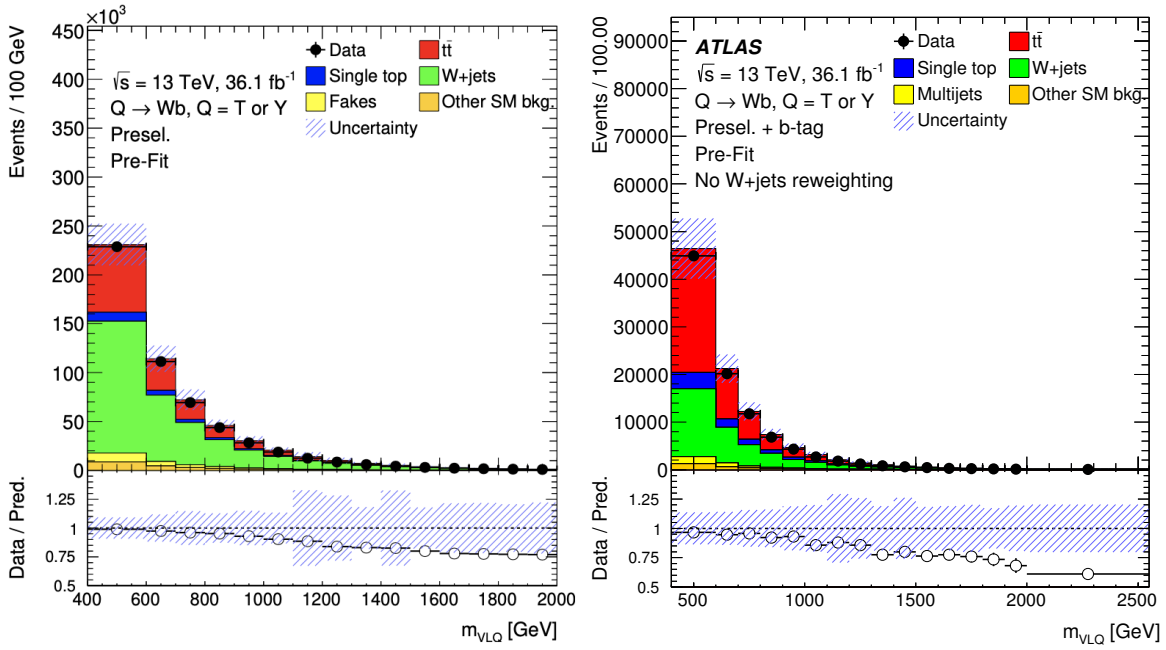
2. Since a single-lepton final-state is probed, events are required to pass a set of single electron or single muon triggers [195, 204, 205].

3. A single electron or muon with $p_T > 28\,\text{GeV}$ trigger matched to the leptons that pass the single-electron or muon triggers should be present in all the events.

4. At least two jets with $p_T > 25\,\text{GeV}$ are required to be present in each event among which one should be central ($|\eta| < 2.5$). In addition to this, forward jets ($2.5 < |\eta| < 4.5$) with $p_T > 40\,\text{GeV}$ are also considered.

5. The E_T^{miss} is required to be greater than 120 GeV to get rid of SM diboson and multijet events.

The kinematic distribution of these events for the reconstructed VLQ mass, m_{VLQ} is shown in Fig. 6.2(a). Pre-selection events with an extra requirement of the leading jet being b-tagged is depicted in Fig. 6.2(b). This requirement results in a significant suppression of the W+jets background. A significant mis-modelling is observed in the high m_{VLQ} region in both the afore-mentioned regions. This mis-modelling has been solved using background reweighting for W+jets as a systematic uncertainty in the final fit. This will be discussed in detail in Chapter 7.



(a) (b)

Figure 6.2: Distribution of the reconstructed VLQ mass, m_{VLQ} for the observed data and SM background predictions in the (a) pre-selection, (b) pre-selection + b-tag regions respectively. The uncertainty band is the statistical and systematic uncertainties in quadrature. [15, 195]

6.4 Signal region

The signal region (SR) is optimised to find a region where the signal-background ratio is high. Certain kinematic cuts, in addition to the pre-selection cuts are utilised to suppress the main SM backgrounds. Since the analysis searches for T/Y signals of varying coupling and masses, it is impractical to optimise the kinematic cuts for all these mass-points. A signal with a mass and coupling close to the existing ATLAS limits [99] is chosen. A Y signal and a mass of 1.2 TeV with a coupling of $(c_L^{Wb})^2 + (c_R^{Wb})^2 = 0.5$ is used to optimise the signal-region cuts.

The following signal region cuts in addition to the preselection cuts were decided upon after the optimisation.

1. Events containing at least one b-jet are chosen and the leading jet is required to be b-tagged.

2. The leading jet has a $p_{\mathrm{T}} > 350\,\mathrm{GeV}$. This helps to suppress the $t\bar{t}$ and W+jets background.

3. To achieve energy-momentum balance for the decay of a high mass Y/T, the b-jet should recoil against the lepton originating from the W-boson as depicted in Figure 6.1. Thus, only events wih high azimuthal separation i.e. $|\Delta\phi(\text{lepton, leading-jet}) > 2.5|$ are chosen.

4. An additional angular cut between the lepton and central jets, $\Delta R_{\mathrm{min}} > 2$, ΔR_{min} being the minimum angular separation between the lepton and the central jets, is also applied for the signal region events.

5. Events containing any hard ($p_{\mathrm{T}} > 75\,\mathrm{GeV}$) central jet excluding the leading jet are rejected if $|\Delta R\ (\text{leading-jet, hard central-jets})| > 2.7$ or $|\Delta R\ (\text{leading-jet, hard central-jets})| < 1.2$.

6. Events are required to have at least one forward ($2.5 < |\eta| < 4.5$) jet with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$. The p_{T} cut helps to suppress the single-top background.

These kinematic cuts result in a SR which has a $S/B = 0.65$ and $S/\sqrt{B} = 13.7$ for the Y signal with a mass of 1.2 TeV at a coupling of $(c_L^{Wb})^2 + (c_R^{Wb})^2 = 0.5$.

6.5 Control regions

A precise understanding of the SM backgrounds is pivotal to this analysis. After the SR cuts are applied, the two main irreducible backgrounds are $t\bar{t}$ and W+jets. Two kinematic regions orthogonal to the SR called control regions (CR) are defined to model the two respective backgrounds.

W+jets CR

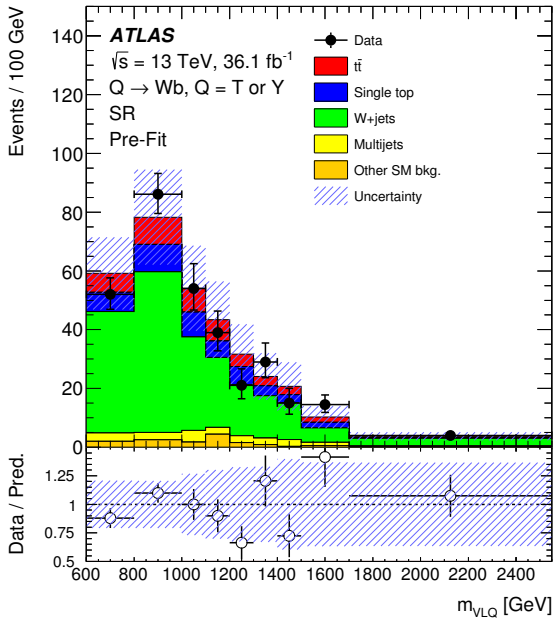
The W+jets CR is defined by changing or reversing the following SR cuts. The cut on leading jet p_{T} is lowered to 250 GeV to increase W+jets statistics. The forward jet requirement in the SR is removed. The cut on azimuthal separation between the lepton and the leading jet is reversed to ensure orthogonality between the W+jets CR and the SR. The veto on hard central jets applied in the SR is also removed to increase statistics.

$t\bar{t}$ CR

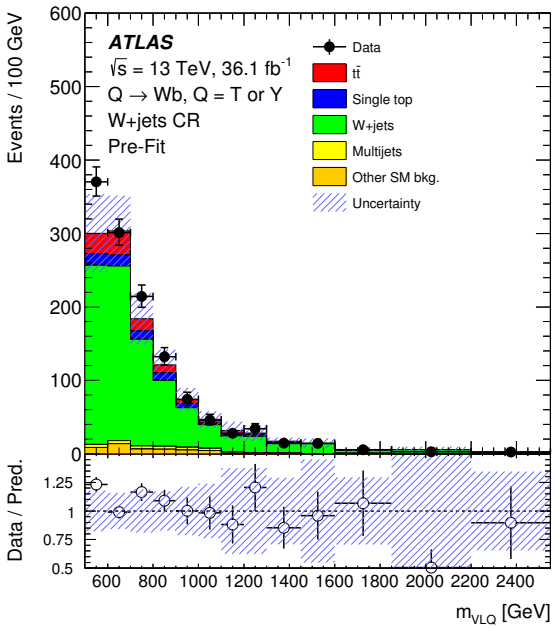
For the $t\bar{t}$ CR, the leading jet p_{T} cut is lowered to 200 GeV as it significantly increases the statistics. The topology of the $t\bar{t}$ final state requires a high central jet multiplicity. Thus, the veto on the hard central jets as used in the SR is reversed. In addition to the leading jet, each event is required to have at least one hard ($p_{\mathrm{T}} > 75$ GeV) central jet with $|\Delta R\ (\text{leading-jet, hard central-jets})| > 2.7$ or $|\Delta R\ (\text{leading-jet, hard central-jets})| < 1.2$.

The distributions of the m_{VLQ} in all these three regions after applying all these cuts is shown in Fig. 6.3. A tabular summary of these three regions is presented in Table 6.2. The event yields in these three regions for all the backgrounds is depicted in Table 6.3.

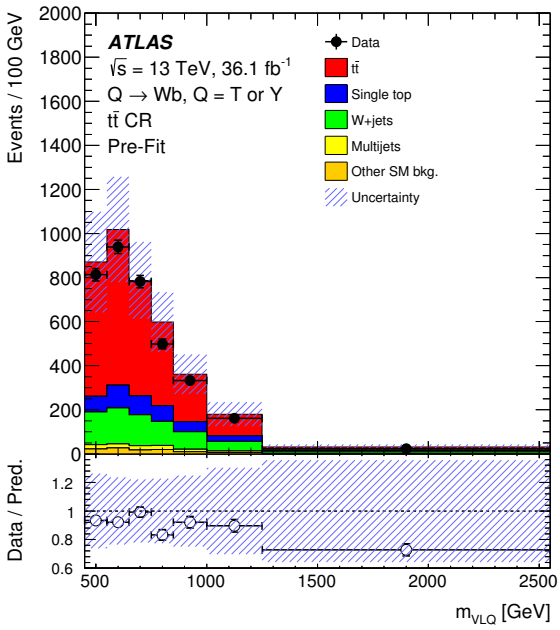
Since the CRs are defined to be orthogonal to the SR, negligible signal contamination is expected. This was explicitly checked for Y signals of mass 0.8 to 1.6 TeV in steps of 200 GeV having a coupling of $(c_L^{Wb})^2 + (c_R^{Wb})^2 = 0.5$. The $S/\sqrt{B}$ ratio varied from $0.08 - 5.7$ in the CRs.



(a)



(b)



(c)

Figure 6.3: Distribution of the reconstructed VLQ mass, m_{VLQ} for the observed data and SM background predictions in the (a) SR, (b) W+jets CR, and (c) $t\bar{t}$ CR respectively. These distributions are at a pre-fit level i.e. before the fit to data. The uncertainty band is calculated by adding all statistical and systematic uncertainties added in quadrature.

Table 6.2: Summary of common pre-selection requirements and selection requirements for the SR compared to those for the $t\bar{t}$ and W+jets CRs. All other selection requirements are the same for all three regions.

Requirement \ Region	SR	$t\bar{t}$ CR	W+jets CR
Pre-selection			
Leptons		1	
$E_{\mathrm{T}}^{\mathrm{miss}}$		$> 120\,\mathrm{GeV}$	
Central jets ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$)		≥ 1	
Selection			
b-tagged jets	≥ 1	≥ 1	1
Leading b-jet p_{T}	$> 350\,\mathrm{GeV}$	$> 200\,\mathrm{GeV}$	$> 250\,\mathrm{GeV}$
$\lvert\Delta\phi$(lepton, b-tagged jet)$\rvert$	> 2.5	> 2.5	≤ 2.5
Jets ($p_{\mathrm{T}} > 75\,\mathrm{GeV}$) with ΔR(jet, b-tagged jet) < 1.2 or ΔR(jet, b-tagged jet) > 2.7	0	≥ 1	–
ΔR(lepton, jets)	> 2.0	–	> 2.0
Forward jets ($p_{\mathrm{T}} > 40\,\mathrm{GeV}$)	≥ 1	≥ 1	–

Table 6.3: Event yields in the SR, $t\bar{t}$ CR, and W+jets CR for the data and SM background predictions before the fit to data. The uncertainties include all statistical and systematic uncertainties.

Source	SR	$t\bar{t}$ CR	W+jets CR
$t\bar{t}$	65 ± 48	2983 ± 907	111 ± 40
Single top	63 ± 56	520 ± 408	68 ± 57
W+jets	324 ± 66	909 ± 168	923 ± 125
Multijet e	22 ± 25	35 ± 40	0 ± 4
Multijet μ	7 ± 7	93 ± 95	26 ± 27
Z+jets, diboson	20 ± 5	105 ± 19	50 ± 7
$t\bar{t}\,V$	0.3 ± 0.1	22 ± 2	1.6 ± 0.2
$t\bar{t}\,H$	0 ± 0	7.4 ± 1.2	0.2 ± 0.1
Total	501 ± 106	4672 ± 1089	1179 ± 155
Data	497	4227	1274

CHAPTER 7

Systematic uncertainties

For every experimental measurement, in addition to the statistical uncertainties, there are associated systematic uncertainties that have to be taken into account. In this chapter, the systematic uncertainties that influence the single-lepton analysis will be described. The experimental uncertainties that arise from reconstruction of the final state objects are derived by the Combined Performance (CP) groups of the ATLAS collaboration. In addition to the experimental uncertainties, modelling and theoretical uncertainties are used. The effects of these uncertainties are evaluated on the relevant signal and background processes by calculating the difference in event yields and/or difference in shape as a function of various kinematic variables in the regions defined in Table 6.2. When an uncertainty affects the event yields uniformly as a function of a kinematic variable, the uncertainty is called a normalisation uncertainty, whereas a shape uncertainty only affects the shape of a kinematic variable. Most uncertainties affect both shape and normalisation of the signal and background processes. The uncertainties are used as nuisance parameters in the statistical analysis. For almost all uncertainties, both $\pm 1\sigma$ variations are calculated and used as templates in the final analysis.

7.1 Experimental uncertainties

The luminosity uncertainty on the 2015-2016 ATLAS dataset was measured as 2.1%. The uncertainty is derived on the baseline luminosity measurement using the LUCID-2 detector [206] with a methodology similar to the one discussed in Ref. [207]. This uncertainty is applied uniformly to all simulated signal and background background processes as a normalisation uncertainty.

The accuracy of the reconstruction algorithms in both data and MC depend upon the number of primary interactions. To account for the differences in data and MC an event weight is multiplied for each event to derive the pileup uncertainty. The efficiency of JVT as described in Section 4.3.4 does not match in MC and data. The JVT efficiency is measured using $Z(\rightarrow \mu\mu) +$ jets events. The ratio of the efficiencies in data and MC is applied as a

scale factor per event, whereas the uncertainties on these scale factors are propagated to the event yields.

For the leptons (e, μ), the difference in trigger efficiency, energy scale, and resolution differ in MC simulations and data. The calibration is done in $Z \to \ell\ell$ events to evaluate the lepton scale factors for efficiency measurements. The uncertainties in these scale factors are used in this analysis. The same calibration procedure also yields the lepton energy scale and resolution uncertainties. A cross-check of these uncertainty measurements is also performed using $J/\psi \to \ell\ell$ decays [208].

The JES uncertainties are propagated from the calibrations done on the jet collection described in Section 4.3.2. The uncertainties include pile-up, GSC punch-through, and uncertainties propagated from the Z/γ + jets, multijet balance, and η-intercalibration. For the single-lepton analysis, 21 correlated JES uncertainties were used as nuisance parameters. Although the most accurate JES uncertainties' configuration include 80 nuisance parameters, a reduced configuration of 21 nuisance parameters is the most appropriate for a search performed for this thesis as recommended by the Jet $E_{\mathrm{T}}^{\mathrm{miss}}$ CP working group [162].

The jet energy resolution (JER) is the average resolution of the ATLAS detector for the energy measurement of jets. As this is measured on data, the MC simulation jets are smeared using the measured resolution. The JER measurement used for the smearing was measured in Run-1 data but was extrapolated to Run-2 conditions [209]. Here, all JER uncertainties were combined into one nuisance parameter. Since the smearing is an error extrapolation, only the $+1\sigma$ error can be calculated.

The "hard term" of the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) is determined based on the reconstructed hard objects, with uncertainties considered in the calculation, as well as the uncertainties associated with the resolution and scale calculation of the "soft term" which is composed of unassociated inner detector tracks.

The b-tagging MV2 classifier, described in Section 4.3.4, is not perfectly accurate. Based on the used working point, misclassification of b-jets, c-jets, and light flavoured jets happen, which is further used to define tagging uncertainties. Three types of tagging efficiencies for the b-jets, c-jets, and light flavoured jets contribute to the construction of three uncorrelated uncertainties. flavour, p_{T}, and η dependent scale factors are applied to each jet. Uncertainties on these scale factors are eigen decomposed into 27 nuisance parameters. Sixteen of these parameters relate to light flavour mistagging, three of them are associated with b-tagging, six of them arise from c-tagging, and two are extrapolation uncertainties. Since the scale factors are only applied up to 700 GeV, a high-p_{T} extrapolation eigenvector component is used for jets in the high-p_{T} range.

7.2 Modelling uncertainties

Modelling uncertainties that affect the MC simulation are used in this analysis. These uncertainties affect the event generation preceding the detector simulation. These have an impact on both signal and background processes depending on the MC event generator

being used.

$t\bar{t}$ For $t\bar{t}$ samples, the choice of parton shower algorithm is varied. The nominal MC sample of POWHEG-BOX + PYTHIA is compared against the variation sample of POWHEG-BOX + HERWIG++ to construct the parton shower uncertainty. The difference between MC samples generated and showered using MADGRAPH5_AMC@NLO + HERWIG++ and POWHEG-BOX + HERWIG++ is propagated to the nominal POWHEG-BOX + PYTHIA sample as a generator uncertainty. Additionally, another uncertainty is derived by varying the h_{damp} parameter, which alters the amount of initial and final-state radiation (ISR/FSR). No cross-section uncertainty is used for $t\bar{t}$ as the normalisation for this background is a free parameter in the fit. Further details are described in Chapter 8.

Single-top Similar uncertainties for ISR/FSR and parton shower are determined for the single-top samples. As the single-top samples comprise three individual samples (s-channel, t-channel, and Wt), the evaluation of these uncertainties was also done separately. For the generator uncertainty, the variation among the MADGRAPH5_AMC@NLO + HERWIG++ sample and the POWHEG-BOX + HERWIG++ and sample is propagated to the nominal POWHEG-BOX + PYTHIA sample. The ISR/FSR uncertainty was evaluated similarly to the uncertainty on $t\bar{t}$. The fragmentation uncertainty is taken as the difference between POWHEG-BOX + PYTHIA and POWHEG-BOX + HERWIG++ samples. Due to significant interference among the Wt and the $t\bar{t}$ processes, the Wt MC samples are created using two schemes, each of which has a different way of handling the interference. The diagram removal (DR) sample is used as default, whereas the diagram subtraction (DS) sample is used the variation. A 6.8% cross-section uncertainty is applied to this background in all regions.

W+jets W+jets samples are created using MADGRAPH5_AMC@NLO + PYTHIA and compared against the nominal SHERPA samples as the generator uncertainty. To account for the differences in light and heavy flavour W+jets in signal and control regions, the shape differences between the m_{VLQ} template for heavy plus light flavour and heavy flavour is used as an uncertainty for W+jets. Similar to the $t\bar{t}$ background, the W+jets normalisation factor is also a free parameter in the fit.

Other SM backgrounds For the other SM backgrounds like Z+jets and diboson production, the cross-section uncertainty on the NNLO production cross-section is used as the sole modelling uncertainty. A 5% normalisation uncertainty is applied to both these backgrounds. A decorrelated normalisation uncertainty of 100% is applied to both electron and muon channel multijet processes.

Signal samples The modelling uncertainty for the Y/T signal sample was evaluated as a PDF uncertainty using the PDF4LHC15 prescription [210]. In this analysis, certain MC

signal samples were produced using a faster simulation algorithm (AFII) that uses parameterised showers rather than using the full detector simulation (FullSim). The differences in efficiencies for the m_{VLQ} templates were studied for 900 GeV, 1 500 GeV, 1 600 GeV, and 1 800 GeV. To account for the differences, the AFII samples were scaled up by 7.5% $\pm$ 2.5%.

Background reweighting uncertainties

As depicted in Fig. 6.2, a significant mismodelling among the MC prediction and the data is present for the reconstructed vector-like quark mass, m_{VLQ}. It is observed that the major contributor to this mismodelling is the leading jet p_{T} in the W+jets background. A reweighting of the leading jet p_{T} of the W+jets sample is performed, which is subsequently used as a systematic uncertainty. The reweighting is done in the preselection plus b-tag region. Firstly, all backgrounds other than the W+jets background are subtracted from the number of observed data events. Then, bin-by-bin reweighting weights were evaluated by scaling the remaining W+jets events to the data events. In the low mass region these bin-by-bin weights lie between 0.8 to 1.1, whereas in the high mass region they are around 0.4 [195]. The original and reweighted leading jet p_{T} distributions in the preselection plus b-tag region is shown in Fig. 7.1. These reweighting factors are applied to the W+jets events in all regions and the resulting distributions are used as a systematic uncertainty.

For the multijet background originating from fake leptons, a disagreement between MC and data was observed in the multijet validation region. An electron p_{T} reweighting was performed to fix this disagreement. This reweighted template was used as a decorrelated uncertainty on the fake electrons [195].

The experimental and modelling uncertainties in the signal and the control regions are added up in quadrature before the fit. The shaded region in the bottom panels shown of the three figures shown in Fig. 6.3 quantify the shape and normalisation dependent total uncertainty on the background MC predictions.

Signal reweighting uncertainties

The reweighting procedure for the signal samples described in Section 6.2 has statistical uncertainties associated to it. The Gaussian propagation of that uncertainty is used for all the reweighted signal samples. To test the efficiency of the reweighting procedure, signal samples with $\kappa = 0.5$ and $\kappa = 0.1$ are simulated and reweighted to the other coupling. When either of the samples at both of these couplings is the target sample, an average expected yield difference of 2.5% is observed. Thus, a 2.5% reweighting uncertainty is used for all signal samples on the mass coupling grid to account for this disagreement.

Interfering background

Certain parts of the interfering background are not part of the simulated SM background samples. A low statistics sample of this background was produced, and it was observed

that the shape of this background is similar to other major SM backgrounds ($t\bar{t}$, single-top, W+jets). Therefore, an additional shape uncertainty was assigned to the $t\bar{t}$ sample by normalising the missing interfering background to the event yields of the $t\bar{t}$ sample in the signal region[1].

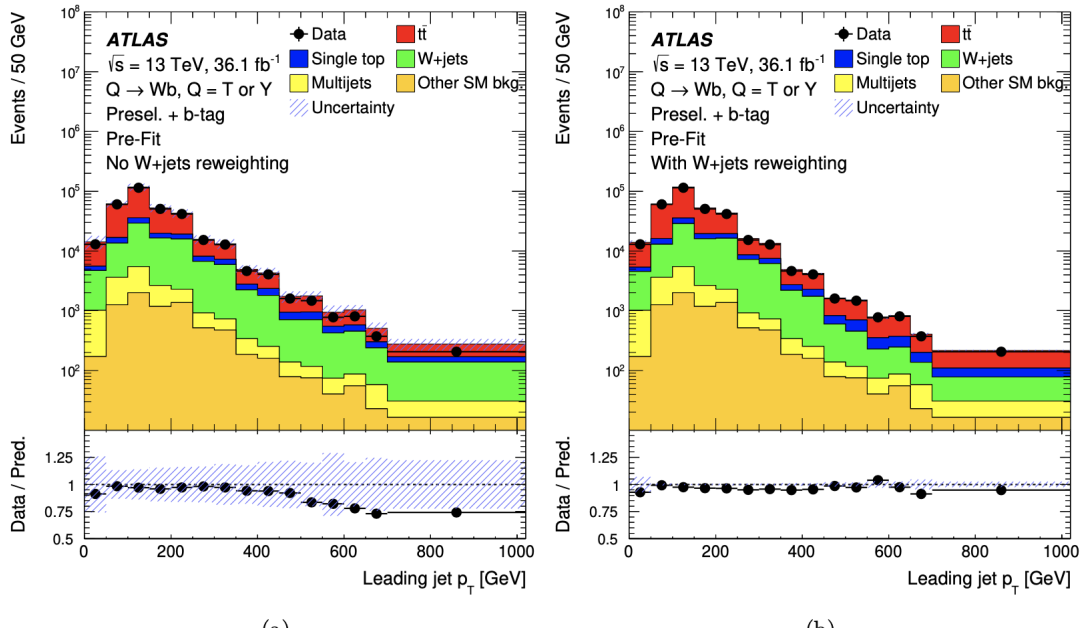


Figure 7.1: Leading jet p_{T} distributions in the Preselection plus b-tag region before and after applying the W+jets leading jet p_{T} reweighting [195].

[1] It should be noted that the effect of the interference was overestimated in this analysis, and thus the resulting interfering background was also overestimated. In reality, the interference effects are negligible. So, this uncertainty is a conservative addition to the full set of uncertainties.

CHAPTER 8

Statistical analysis

In this chapter, the details of the statistical analysis performed on the events selected in Chapter 6 will be discussed. The statistical methods that are used here are described in detail in Chapter 5. Using the m_{VLQ} distribution as a discriminating variable, the search for T/Y is performed.

8.1 Profile likelihood fits

All statistical fits employed for this analysis are profile likelihood fits. The likelihoods are built with the binned m_{VLQ} distributions in the control (CR) and signal regions (SR) as shown in Fig. 6.3. The profile likelihoods include all systematic and statistical uncertainties as nuisance parameters (NP) according to the methods detailed in Section 5.2 and Section 5.5. The systematic NPs originate from the uncertainties discussed in Chapter 7 and the statistical NPs are Monte Carlo uncertainties added into the likelihood using Eq. (5.12). The NPs affect both the shape and normalisation of the signal and background templates of the discriminating variable, m_{VLQ}. All the NPs are incorporated into the likelihood using a Gaussian prior. Each NP affects the profile likelihood via a penalty term according to Eqs. (5.8) and (5.9). If a NP affects the normalisation only, the associated signal or background yields are varied in each m_{VLQ} bin by a fixed $\pm x\%$, where $\pm x\%$ denote the $\pm 1\sigma$ variation. When a NP only affects the shape of the m_{VLQ} distribution, the $\pm 1\sigma$ variation of the m_{VLQ} distribution changes the yields across the mass range in a way that the total yields of the varied distribution remains unchanged. A NP that affects both shape and normalisation varies the yields per bin to change both the shape and normalisation of the associated process.

Pruning, symmetrisation, and smoothing This analysis incorporates hundreds of NPs into the fits. To avoid, over-constraining the fit the variations above a certain threshold are included. If a certain NP does not affect the normalisation of a given processes by more than a percent, the normalisation component is dropped. Subsequently, if a NP does not

change the shape of at least one m_{VLQ} bin by more than a percent the shape component of the NP is dropped for the associated process. This way of dropping certain components of systematic uncertainties is called pruning. The $\pm 1\sigma$ variations for all uncertainties are not the same. Including these unequal variations can lead to certain instabilities in the fit. There also exists uncertainties e.g. jet energy resolution, where only either the $+1\sigma$ variation or the -1σ variation is available. In both these cases, the $\pm 1\sigma$ variations are symmetrised for both the shape and normalisation components. Here, the symmetrisation is performed by either averaging the magnitude of the up and down variations, or by mirroring the variation that is present. The magnitude of an uncertainty can differ significantly from one bin to the other of the m_{VLQ} distribution. This results in ragged shape variations which can lead to subsequent fit instabilities. Such occurrences are avoided by employing a method called smoothing, where the resulting distribution is mostly smooth across the whole mass range. An example of an uncertainty before and after smoothing and symmetrisation is depicted in Fig. 8.1

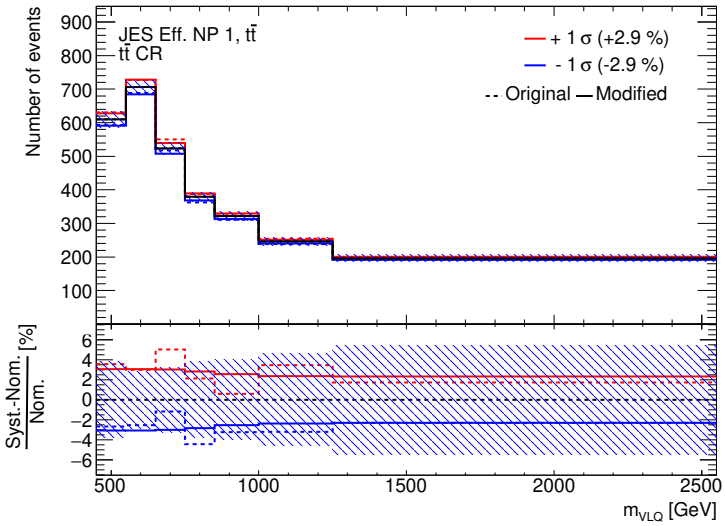


Figure 8.1: The $\pm 1\sigma$ m_{VLQ} distribution of a jet energy scale NP for the $t\bar{t}$ background in the $t\bar{t}$ CR before (dotted line) and after (solid line) applying smoothing and symmetrisation.

Background-only hypothesis testing

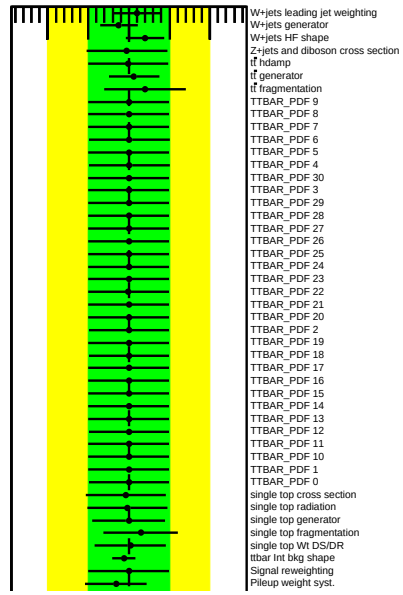
For the null or background-only hypothesis, $\mu_{\mathrm{signal}} = 0$ is used. To test the validity of the SM against the data, the null hypothesis is fitted against the data in the CRs and the SR using a maximum log-likelihood fit. Here, the normalisation factors for the two major backgrounds, $t\bar{t}$ and W+jets are free parameters in the fit. Thus, in this fit, μ_j^{bkg} (See Eq. (5.6)), where $j \in t\bar{t}, W+\mathrm{jets}$ is the parameter of interest (POI) A very good post-fit agreement of the total yields of data and the SM backgrounds is observed within uncertainties. A comparison of data versus the post-fit yields of the SM backgrounds in the SR and both the CRs is shown in Fig. 8.6. The post-fit yields and uncertainties of each of the individual

SM backgrounds in comparison to the data for the SR, $t\bar{t}$ CR, and W+jets CR are tabulated in Table 8.1. The uncertainties for each of these backgrounds are calculated by performing a Gaussian propagation of the systematic uncertainties tabulated in Table 8.2 and Monte Carlo (MC) statistical uncertainties. This propagation includes the correlations between all the nuisance parameters (NPs) included in the fit. The relative effect of the systematic uncertainties on the total inclusive expected SM background yields determined from the VLQ candidate invariant mass distribution after the fit to data using the background-only hypothesis are listed in the last column as percentage in the last column of Table 8.2. The second column in Table 8.2 explains the way the uncertainties are assigned. The pulls of the individual NPs after the fit to data using the background-only hypothesis is shown in Fig. 8.2 , Fig. 8.3, and Fig. 8.4. All the pulls of the NPs are within $\pm 1\sigma$ and most of the NPs are not significantly constrained. The ones that are significantly constrained are due to the constraints coming from data. This implies that these uncertainties were overestimated at the pre-fit level. The methodology for the construction of such a plot is explained in Section 5.5. The observed values for both the free parameters, the normalisation factors of $t\bar{t}$ and W+jets are 0.95 ± 0.26 and 1.18 ± 0.19 respectively. The total errors on the respective normalisation factors are 0.26 and 0.19. The two parameters have a 36% correlation among themselves. The normalisation factors for $t\bar{t}$ and W+jets have respective anti-correlations of 27% and 45% with the cross-section uncertainty of single-top background. The W+jets leading jet reweighting NP has an anti-correlation of 22.8% . The post-fit distributions for the discriminant m_{VLQ} for the SM backgrounds compared to data in the SR, W+jets CR, and $t\bar{t}$ CR is shown in Fig. 8.6. Th post-fit distributions for other kinematic variables are shown in Appendix A. As observed in Fig. 8.6(a), the data distribution of m_{VLQ} agrees to the fitted background yields within uncertainties in all bins except the bin centred at 1.6 TeV. A fit to data using the 1.6 TeV Y signal excluding interference effects with a coupling $\sqrt{(c_{\mathrm{L}}^{Wb})^2 + (c_{\mathrm{R}}^{Wb})^2} \approx 0.5$ with plus background hypothesis yields a significance of 1.1σ. Thus, the absence of any significant excess over observed data implies that the SM background-only hypothesis is accepted.

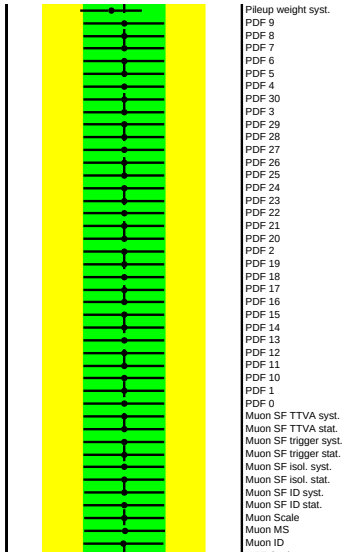
8.2 Limit setting

Since the SM background-only hypothesis is accepted after the fits described above, upper limits on the signal strengths of Y/T VLQ defined by the theoretical model is calculated. A mass-coupling grid is used to set mass and coupling upper limits. The 95% upper limits are set using the procedure described in Section 5.4. For the limit setting procedure, both expected and observed limits are calculated. The expected limits are calculated by fitting to Asimov datasets. Asimov datasets are constructed from expected Monte Carlo (MC) distributions of the background and signal processes in both SR and CRs using Eq. (5.21). Therefore, for a given VLQ signal hypothesis with coupling $c_{L/R}^{Wb}$ and mass m, a unique signal-plus-background Asimov dataset is built by setting $\mu = 1$ in Eq. (5.21), where $\mu = 1$ corresponds to the theoretical cross-section predicted in Ref. [12].

Depending upon the interpretation of the T/Y signal, the calculated interference of the

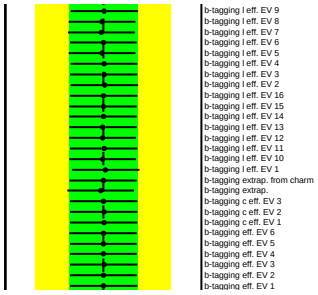


(a)

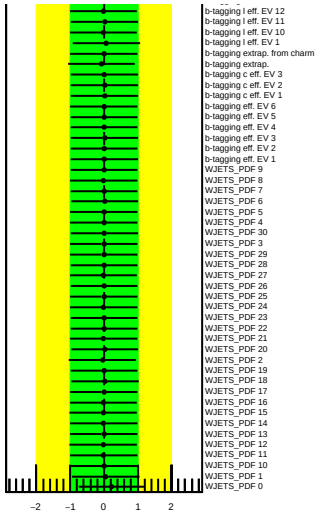


(b)

Figure 8.2: Post-fit pulls on a subset of the NPs after the fit to data in SR and both the CRs using the background-only hypothesis.



(a)



(b)

Figure 8.3: Post-fit pulls on a subset of the NPs after the fit to data in SR and both the CRs using the background-only hypothesis.

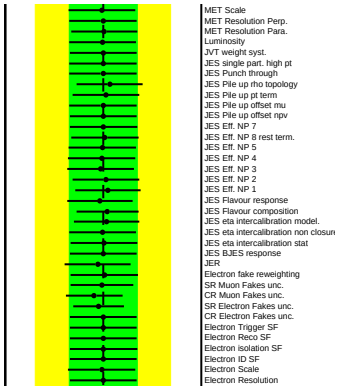


Figure 8.4: Post-fit pulls on a subset of the NPs after the fit to data in SR and both the CRs using the background-only hypothesis.

Table 8.1: Post-fit event yields of individual SM backgrounds in the SR, $t\bar{t}$ CR, and W+jets CRs compared to data yields after the maximum log-likelihood fit using the background-only hypothesis. The uncertainties depicted here are the total post-fit uncertainties of each background process. These uncertainties include the correlated effect of both systematic and statistical nuisance parameters.

Source	SR	$t\bar{t}$ CR	W+jets CR
$t\bar{t}$	58 ± 21	2715 ± 295	100 ± 29
Single top	29 ± 15	271 ± 118	34 ± 18
W+jets	373 ± 45	1052 ± 143	1077 ± 84
Multijet e	22 ± 20	35 ± 40	0 ± 4
Multijet μ	7 ± 7	92 ± 71	26 ± 20
Z+jets, diboson	20 ± 5	102 ± 20	50 ± 8
$t\bar{t}\,V$	0.3 ± 0.1	21 ± 3	1.6 ± 0.3
$t\bar{t}\,H$	0 ± 0	7 ± 1	0.2 ± 0.1
Total	500 ± 30	4300 ± 210	1290 ± 70
Data	497	4227	1274

signal with the SM background and the chirality of the signal changes. In the (T, B, Y) triplet interpretation, limits are set on the Y left-handed (LH) signal, whereas the Y RH signal is probed for the (B, Y) doublet representation. For the T singlet case, limits are calculated for the T LH signal. The interference with the signal and the background is significant for the triplet and singlet representations, whereas for the doublet representation the interference is several orders of magnitude smaller than the previous two cases. The interference pattern changes the shape of the m_{VLQ} distribution based on the coupling. Thus, rather than calculating mass limits, mass-dependent upper limits were produced for the $QWb, Q = T/Y$ coupling. Since the interference is destructive, a dedicated method was developed for this thesis to compute coupling limits for the VLQ signals.[1]

8.2.1 Coupling limits with interference

The interference between the signals and the SM backgrounds is destructive, and thus changes the m_{VLQ} distribution significantly. The most notable effect is observed near the mass peak of the Y LH and the T LH signals. For the Y LH signal, the signal (S) plus interference (I) template has negative entries for adjacent masses greater than the mass of the probed signal, whereas, the T LH S+I distribution has negative entries for adjacent

[1] As discussed in Section 6.2, these interference effects as used in this analysis were overestimated. The correct magnitude of the interference is calculated in Ref. [110] which was published two years after the publication of this analysis. The method to calculate coupling limits for signals with considerable destructive interference was developed assuming large interference effects. In hindsight, although these methods may not have been necessary for dealing with the interference effects of this analysis, they can be useful for other new physics searches with significant destructive interference effects between signal and background.

Table 8.2: Systematic uncertainties considered in all the fits as nuisance parameters. An uncertainty that affects normalisation only for both signal and background processes is denoted by "N", whereas "SN" means that the uncertainty affects both shape and normalisation and "F" depicts a floating normalisation uncertainty on the given background. The relative systematic uncertainties for the inclusive expected SM background event yields calculated from the m_{VLQ} distribution after the fit to the background-only hypothesis are given in the last column in percentage.

Systematic uncertainty	Type	SM background [%]
Luminosity	N	2.1
Pile-up	SN	0.3
Reconstructed objects:		
Electron efficiency, energy scale, resolution	SN	0.9
Muon efficiency, momentum scale, resolution	SN	0.7
Jet vertex tagger	SN	0.1
Jet energy scale	SN	6.4
Jet energy resolution	SN	2.7
Missing transverse momentum	SN	0.3
b-tagging efficiency for b-jets	SN	0.8
b-tagging efficiency for c-jets	SN	1.8
b-tagging efficiency for light-flavour jets	SN	8.4
Background model:		
$t\bar{t}$ modelling: ISR/FSR	SN	0.1
$t\bar{t}$ modelling: generator	SN	2.5
$t\bar{t}$ modelling: parton shower/hadronisation	SN	3.8
$t\bar{t}$ modelling: interfering background shape	S	0.2
Single-top cross-section	N	6.7
Single-top modelling: ISR/FSR	SN	0.05
Single-top modelling: generator	SN	0.4
Single-top modelling: DS/DR	SN	4.7
Single-top modelling: parton shower/hadronisation	SN	1.9
W+jets modelling: generator	SN	0.7
W+jets modelling: reweighting	S	4.4
W+jets heavy flavour	S	0.03
Diboson + Z+jets normalisation	N	0.2
Multijet normalisation	N	3.7
Multijet reweighting	S	2.1
$t\bar{t}$ background normalisation factor	F	26
W+jets background normalisation factor	F	19

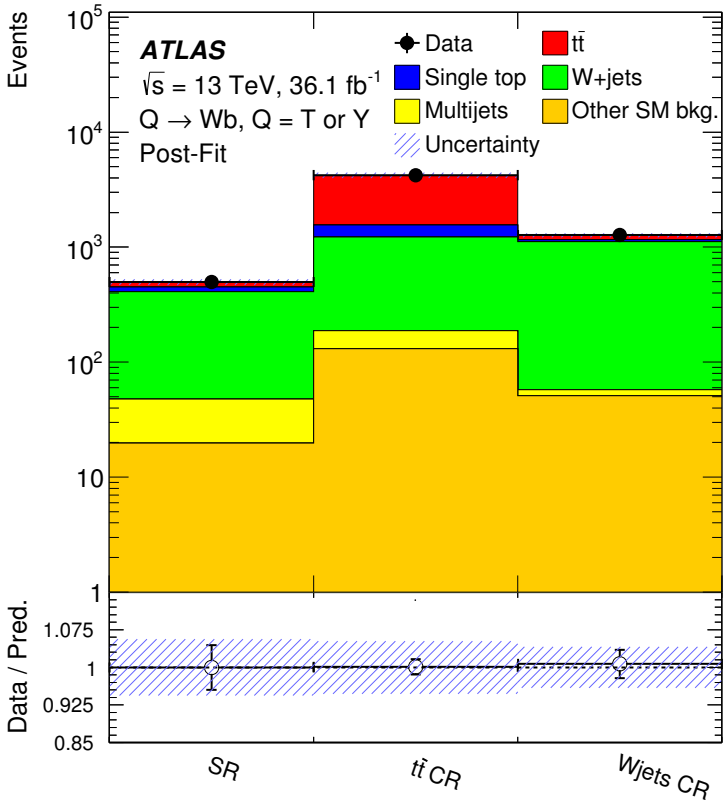
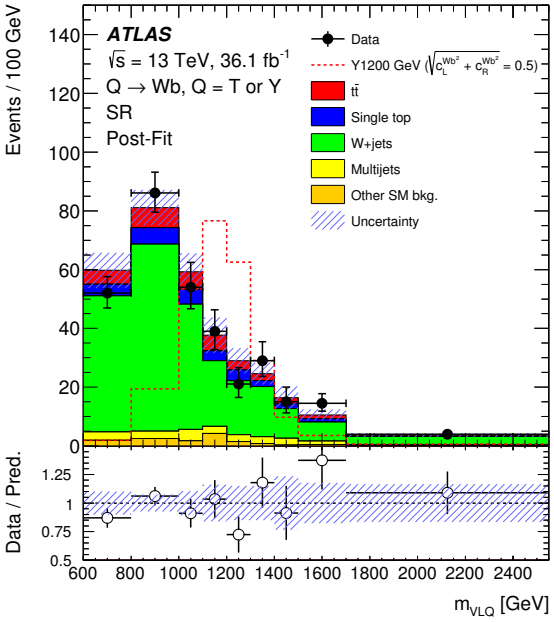


Figure 8.5: Post-fit background yields in the SR and in the two CRs after the fit to the data in the CRs and SR under the null or background-only hypothesis. The bottom panel shows the ratio of data to the fitted SM background yields. The band represents the total (statistical and systematic) uncertainty after the maximum log-likelihood fit.

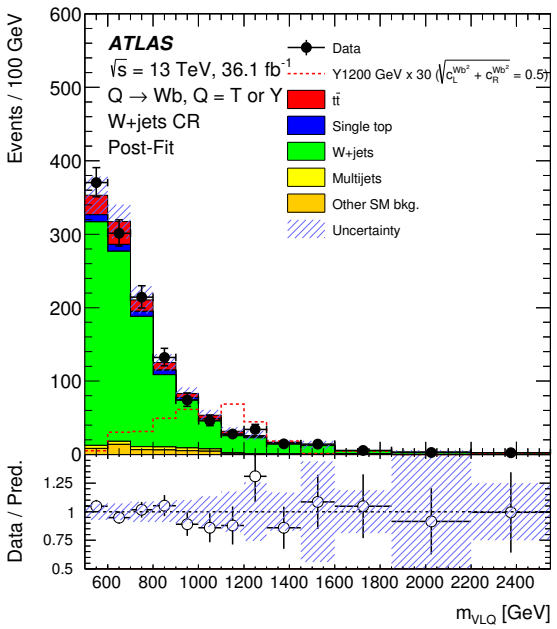
masses less than the mass of the probed signal. The S+I templates for each mass point for a given coupling is computed by reweighting the m_{VLQ} distribution of the initial coupling without interference. The reweighting procedure is performed using the methods described in Section 6.2. The S+I m_{VLQ} template for certain initial and target couplings for two YLH and TLH signals are shown in Fig. 8.7.

Due to the presence of negative S+I event yields, the binning showed in Fig. 8.7 cannot be used to set limits on signal strength. The shape and normalisation of these S+I templates being coupling dependent is another reason for developing an iterative method for limit setting. The steps of this method are as follows:

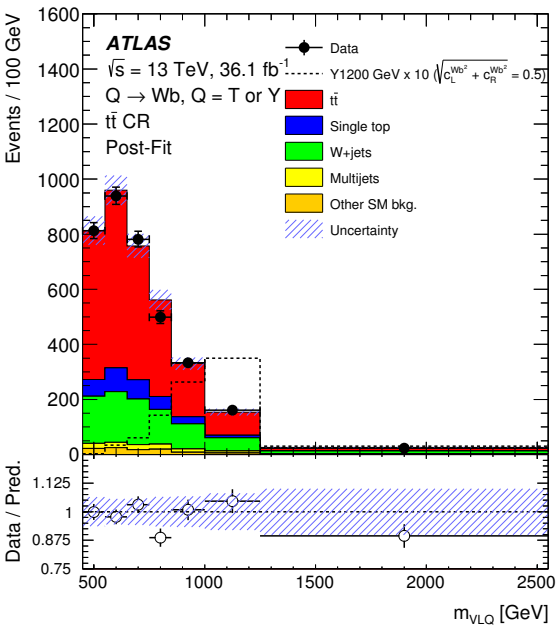
1. For each mass point on the mass-coupling grid, a MC signal sample with initial coupling c_{L}^{Wb} or $c_{\mathrm{R}}^{Wb} = 0.5$ is reweighted to a S+I m_{VLQ} template with the same signal. Since YLH and TLH m_{VLQ} distributions contain negative entries, a new binning is chosen such that the bins with positive event yields "eat" away the negative yields



(a)



(b)



(c)

Figure 8.6: Distribution of the reconstructed VLQ mass, m_{VLQ} for the observed data and SM background fitted yields in the (a) SR, (b) W+jets CR, and (c) $t\bar{t}$ CR respectively. These distributions are at a post-fit level i.e. after the fit to data in SR and CRs using the background-only hypothesis. The uncertainty band is calculated by adding all statistical and systematic uncertainties added in quadrature. A m_{VLQ} distribution for a Y signal with a coupling of $\sqrt{(c_{\mathrm{L}}^{Wb})^2 + (c_{\mathrm{R}}^{Wb})^2} \approx 0.5$ without considering any interference effects is overlaid here; for better representation, it is multiplied by a factor of 30 in the W+jets CR and by a factor of 10 in the $t\bar{t}$ CR.

originating from the destructive interference. The binning used to achieve this for each mass point differs and is tabulated in Table 8.3. It can be observed that for lower masses coarser binnings can achieve the suppression of the negative entries and at higher masses a finer binning can be used to achieve the balancing of negative entries in the S+I m_{VLQ} distribution. No rebinning is needed for the YRH S+I template. Binnings used for YLH and TLH signals for couple of mass points are shown in Fig. 8.9.

2. An upper limit on the cross-section of S+I, $\sigma_S^{\mathrm{limit}} + \sigma_I^{\mathrm{limit}}$ is set using the rebinned template for c_{L}^{Wb} for both YLH and TLH. The upper limit on the cross-section is also calculated for c_{R}^{Wb} but the same binning is used. The calculated upper limits are 95% CL_s limits derived using the profile likelihood fit methodology described in Section 5.4. The interference limit term $\sigma_I^{\mathrm{limit}}$ is subtracted from the cross-section limit $\sigma_S^{\mathrm{limit}} + \sigma_I^{\mathrm{limit}}$. Subsequently, using this cross-section and the theory cross-section the $c_{\mathrm{L,R}}^{Wb,\mathrm{limit}}$ is calculated.

3. After the first iteration, for a given mass point, if $|c_{\mathrm{L,R}}^{Wb,\mathrm{limit}} - 0.5| < 0.1$, the final coupling limit is 0.5. If not, a new S+I m_{VLQ} template is constructed at $c_{\mathrm{L,R}}^{Wb,\mathrm{limit}}$. Step 2 is repeated until the input coupling and the computed coupling limit converge for each mass point. The convergence criterion used here is $|c_{\mathrm{L,R}}^{Wb,\mathrm{limit}} - c_{\mathrm{L,R}}^{Wb,\mathrm{input}}| < 0.1$. The iterative procedure of subsequent reweighting and limit setting is stopped when convergence is reached.

4. When convergence is reached, the "final" expected and observed limits are quoted as the coupling limits for the whole mass range. The number of iterations required for convergence varies for each mass point. The expected limits that are calculated in this iterative procedure use the rebinned S+I template to build the corresponding Asimov dataset for the all points of interest in the mass-coupling grid.

A flowchart of this iterative procedure is shown in Fig. 8.8. The expected and observed coupling limits on c_{L}^{Wb} and c_{R}^{Wb} for the singlet, triplet, and doublet interpretations for each probed mass point is shown in Table 8.4, Table 8.5 and Table 8.6 respectively. The mass ranges probed for each interpretation differ. This is because the total event yields for the S+I template become negative for masses greater than 1 600 GeV for YLH signals and masses greater than 1 200 GeV for TLH. Due to this, the aforementioned iterative procedure cannot be used.[2] As the S+I m_{VLQ} distribution has no negative event yields, the full mass range is probed. The iterative procedure was explicitly tested to be agnostic against the starting coupling values. The converged coupling limits remain invariant for any choice of starting coupling values. As discussed in Section 2.6.1, the mixing angle $\theta_{\mathrm{L/R}}$ of the renormalisable model [13] and the coupling parameter $c_{\mathrm{L/R}}^{Wb}$ of the "bridge" model [12] are related. Thus the limits on $c_{\mathrm{L/R}}^{Wb}$ can be also recast as the limits on $|\sin\theta_{\mathrm{L/R}}|$. The expected and observed limits on the three interpretations of T/Y VLQs as a function of m_{VLQ} are shown in Fig. 8.10.

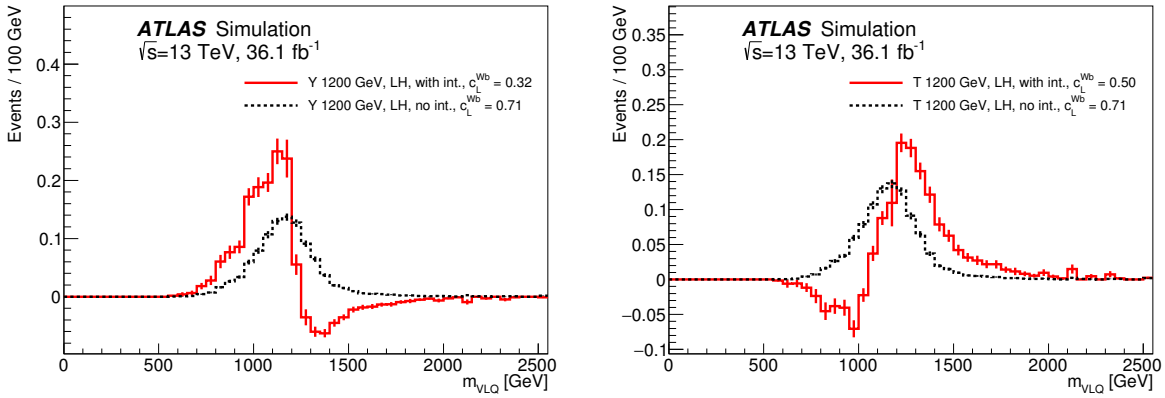
[2] This happens due to the overestimation of the interference effects for YLH and TLH.

Here, for the T singlet interpretation the observed upper exclusion limit at 95% confidence level on $|\sin\theta_L|$ (c_L^{Wb}) is 0.18 (0.25) for a T LH quark with m_{VLQ} = 800 GeV , which increases to 0.35 (0.49) for a T LH quark with a mass of m_{VLQ} = 1 200 GeV. In the (B,Y) doublet interpretation, the upper exclusion limit at 95% confidence level on $|\sin\theta_R|$ (c_R^{Wb}) is 0.17 (0.24) for a signal with m_{VLQ} = 800 GeV, whereas the same upper limit is 0.80 (1.13) for Y RH quarks with a mass of m_{VLQ} = 1 900 GeV. Observed exclusion upper limits for Y LH signals in a (T,B,Y) triplet are $|\sin\theta_L|$ (c_L^{Wb}) vary between 0.16 (0.31) and 0.39 (0.78) for masses between 800–1 600 GeV. Indirect electroweak constraints originating from LEP parameters such as S, T, and R_b on $\sin\theta_{L/R}$ have been derived in Ref. [13]. These constraints as a function of m_{VLQ} have been overlayed on the coupling limit plots in Fig. 8.10. As shown in Fig. 8.10(c), for the (B,Y) doublet interpretation, the exclusion limits set on the mixing parameter $\sin\theta_R$are comparable with the exclusion limits from electroweak precision observables in the mass range between about 900–1 250 GeV. The constraints from S and T parameters exclude masses till around 1 250 GeV. Additionally, the constraints from R_b, the partial width for $Z \to bb$ over the total hadronic Z width in LEP data are comparable to the observed limits in masses up to 1 200 GeV. However, these constraints in the singlet and triplet interpretations are not comparable to the observed exclusion coupling limits.

Since the interference effects in the Y RH case is negligible, mass limits are computed for the Y RH case and the no interference cases. All the mass limits are computed for each mass point at an interval of 100 GeV using m_{VLQ} distributions with couplings of $1/\sqrt{2}$. These limits are shown individually and are also compared against each other in Fig. 8.12. The observed (expected) mass limit below which Y RH quarks are excluded at 95% confidence level is about 1.65 TeV (1.81 TeV) for a right-handed coupling value of $c_R^{Wb} = 1/\sqrt{2}$. Evidently, if the interference effects are negligible, the mass limits are comparable with and without those effects.

Table 8.3: Binning used for different Y/T mass points that is used to determine coupling limits for signals where interference with the SM background is included. Since the interference effects for the right-handed Y signals are negligible, the same binning is used for the full mass range.

Signal	Mass in [GeV]	Binning in [GeV]
Y LH	800	[600, 650, 700, 750, 2550]
Y LH	900	[600, 650, 700, 800, 850, 2550]
Y LH	1000	[600, 650, 700, 750, 800, 850, 900, 950, 2550]
Y LH	1100	[600, 650, 700, 750, 800, 850, 900, 950, 1000, 2550]
Y LH	1200	[600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 2550]
Y LH	1300	[600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 2550]
Y LH	1400	[600, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 2550]
Y LH	1500	[650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150s, 1200, 1250, 1300, 1350, 1400, 2550]
Y LH	1600	[600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 2550]
T	800	[600, 800, 900, 1000, 1500, 2550]
T LH	900	[600, 900, 1000, 1200, 1500, 2550]
T LH	1000	[600, 1000, 1100, 1400, 2550]
T LH	1100	[600, 1150, 1250, 1600, 2550]
T LH	1200	[600, 1250, 1350, 1500, 1650, 2550]
Y RH		[600, 800, 1000, 1100, 1200, 1300, 1400, 1500, 1700, 2550]



(a) (b)

Figure 8.7: Distribution of m_{VLQ} for 8.7(a) Y LH signal with a mass of 1 200 GeV with a target coupling of 0.71 and an initial coupling of 0.32 and 8.7(b) T LH signal with a mass of 1 200 GeV with a target coupling of 0.5 and an initial coupling of 0.71. For both cases, the signal template with initial coupling without interference is reweighted to the signal plus interference template at the final coupling [15].

Table 8.4: Observed and expected 95% CL_s upper limits on $|\sin\theta_{\mathrm{L}}|$ and c_{L}^{Wb} for a left-handed Y quark for the (T, B, Y) triplet interpretation with masses of 800–1 600 GeV. The $\pm1\sigma$ and $\pm2\sigma$ uncertainties in the expected coupling limits are also tabulated.

Y mass [GeV]	Observed limit on $\lvert\sin\theta_{\mathrm{L}}\rvert$	Expected limit on $\lvert\sin\theta_{\mathrm{L}}\rvert^{+1\sigma/+2\sigma}_{-1\sigma/-2\sigma}$	Observed limit on c_{L}^{Wb}	Expected limit on $c_{\mathrm{L}}^{Wb\,+1\sigma/+2\sigma}_{\quad-1\sigma/-2\sigma}$
800	0.16	$0.20\,^{0.04/0.09}_{0.03/0.05}$	0.31	$0.40\,^{0.08/0.19}_{0.06/0.11}$
900	0.14	$0.15\,^{0.03/0.07}_{0.02/0.04}$	0.28	$0.30\,^{0.06/0.13}_{0.05/0.08}$
1000	0.16	$0.15\,^{0.03/0.06}_{0.02/0.04}$	0.32	$0.29\,^{0.05/0.12}_{0.04/0.08}$
1100	0.23	$0.22\,^{0.03/0.08}_{0.03/0.06}$	0.47	$0.43\,^{0.07/0.15}_{0.07/0.12}$
1200	0.20	$0.16\,^{0.03/0.07}_{0.02/0.04}$	0.40	$0.33\,^{0.06/0.13}_{0.05/0.09}$
1300	0.25	$0.21\,^{0.04/0.08}_{0.03/0.06}$	0.49	$0.43\,^{0.08/0.16}_{0.07/0.12}$
1400	0.18	$0.25\,^{0.05/0.10}_{0.04/0.07}$	0.36	$0.51\,^{0.09/0.20}_{0.08/0.14}$
1500	0.32	$0.35\,^{0.08/0.18}_{0.06/0.10}$	0.64	$0.70\,^{0.16/0.37}_{0.12/0.20}$
1600	0.39	$0.40\,^{0.11/0.28}_{0.07/0.12}$	0.78	$0.80\,^{0.21/0.56}_{0.14/0.24}$

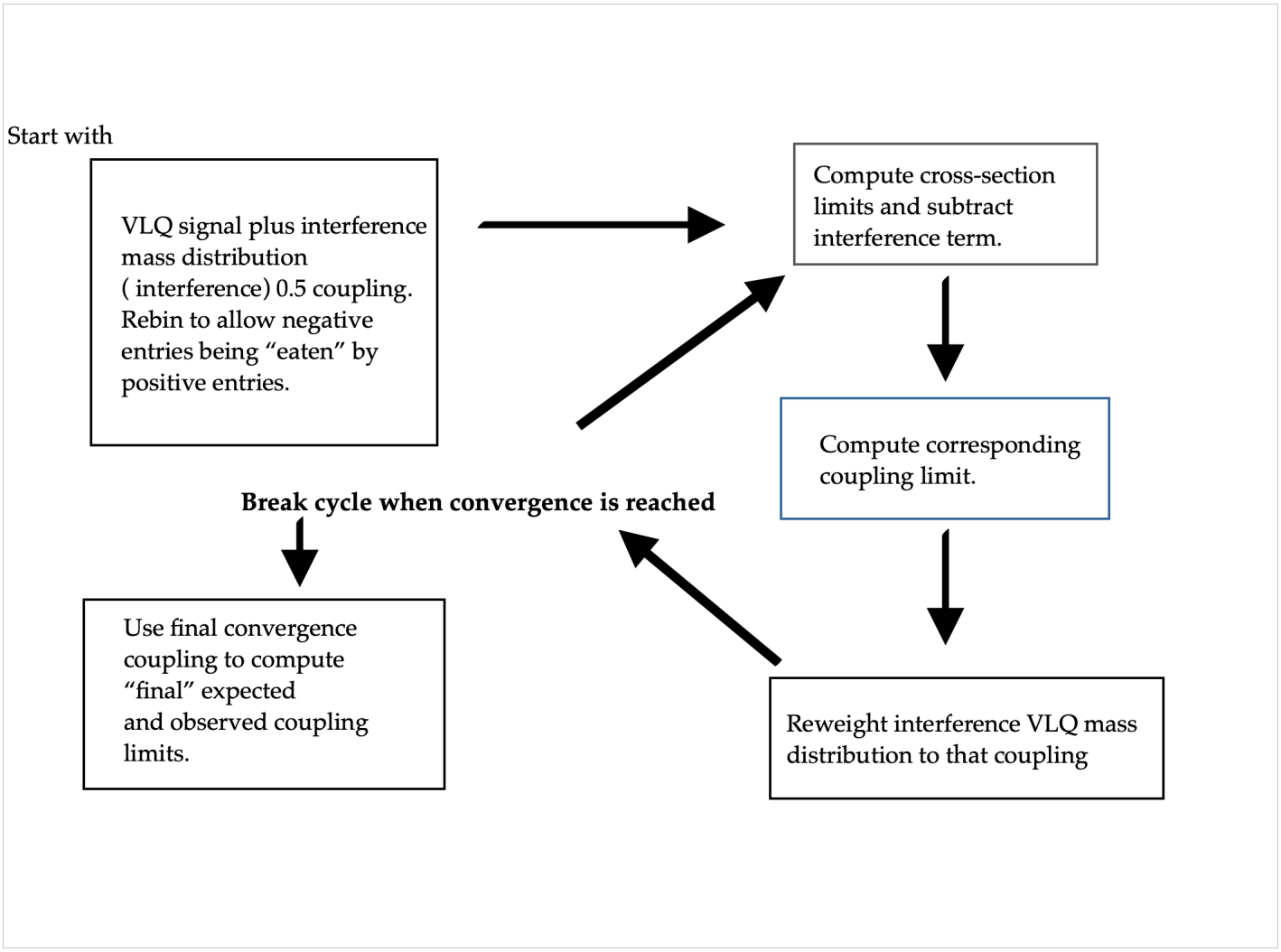


Figure 8.8: A flowchart of the limit setting procedure developed for this analysis for signals with significant interference with the SM background.

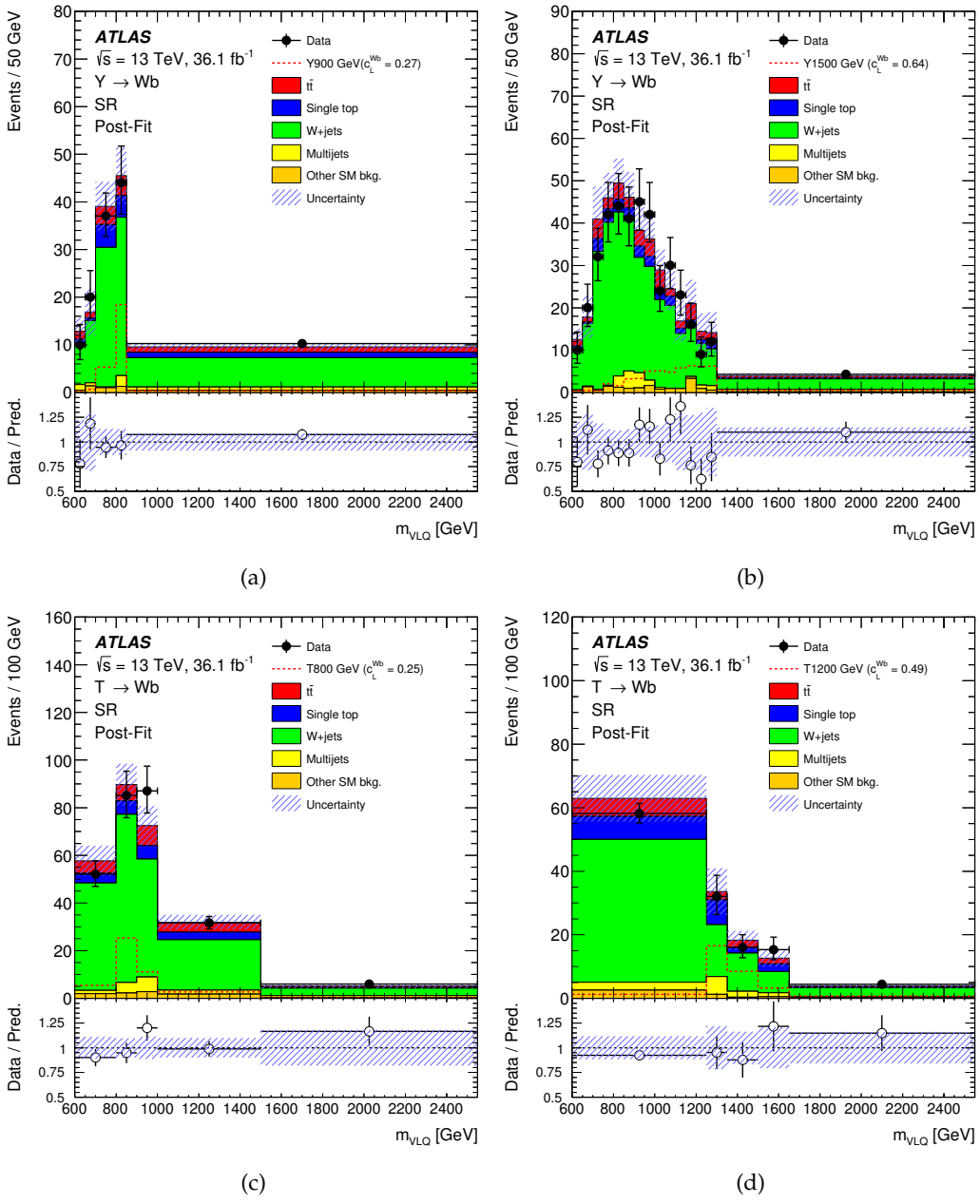
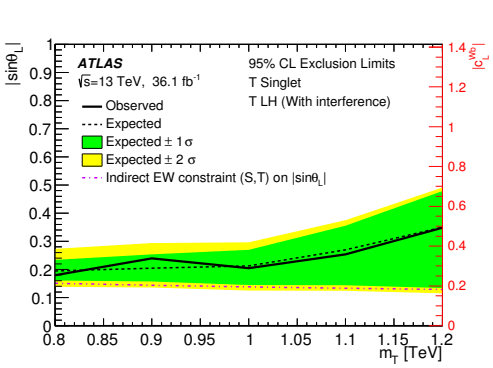
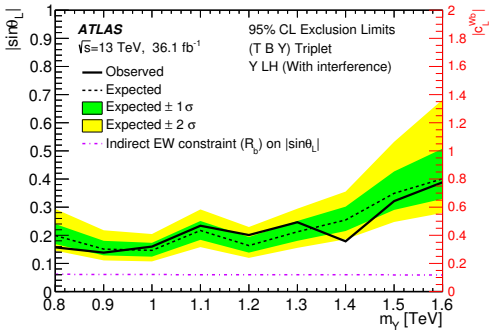


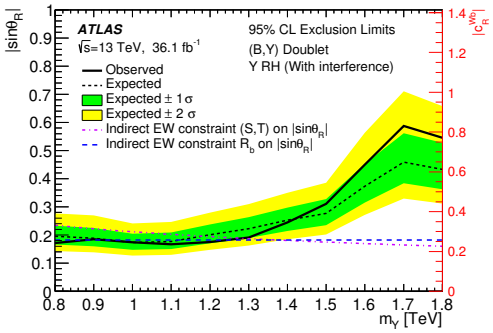
Figure 8.9: Distributions of the VLQ candidate mass m_{VLQ} after the fit to the background-only hypotheses for four different binnings selected for four different signal masses for both YLH and TLH signals. The m_{VLQ} distribution for 8.9(a) a left-handed Y signal with mass 900 GeV and coupling c_{L}^{Wb} = 0.26, 8.9(b) a left-handed Y signal with mass 1 500 GeV and coupling c_{L}^{Wb} = 0.64, 8.9(c) a left-handed T signal with mass of 800 GeV and coupling c_{L}^{Wb} = 0.25 and 8.9(d) a left-handed T signal with mass 1 200 GeV and coupling c_{L}^{Wb} = 0.49 are shown here. Each m_{VLQ} distribution include interference effects which are "eaten" away for the signal in the wide bin. The lower panels show the ratio of data to the fitted background yields. The error bars represent the statistical uncertainty in the data. The band represents the total systematic uncertainty after the maximum-likelihood fit. For comparison purposes, the rebinned signal plus interference m_{VLQ} distributions at the aforementioned couplings have been overlayed (dotted line).



(a)



(b)



(c)

Figure 8.10: Observed (solid line) and expected (short-dashed line) 95% CL_s limits on 8.10(a) the mixing angle $|\sin\theta_L|$ and the coupling value c_L^{Wb} for the singlet T-quark interpretation under the assumption $\mathcal{B}(T \to Wb) \approx 0.5$ 8.10(b) $|\sin\theta_L|$ and c_L^{Wb} for the (T, B, Y) triplet interpretation, and 8.10(c) $|\sin\theta_R|$ and c_R^{Wb} for the (B, Y) doublet interpretation assuming a branching ratio $\mathcal{B}(Y \to Wb) = 1$. All the limits are plotted as a function of the VLQ mass. The surrounding bands correspond to $\pm$1and $\pm$ 2 standard deviations around the expected limit. Indirect constraints from electroweak precision observables on the mixing angles, which originate from LEP measurements, are only valid for the mixing angles. These constraints are either from the S and T parameters (dashed-dotted line) or from the R_b values (long-dashed line) are also overlayed [13] on the coupling limits.

Table 8.5: Observed and expected 95% CL_s upper limits on $|\sin\theta_L|$ and c_L^{Wb} for a left-handed T quark for the T singlet interpretation with masses of 800–1 200 GeV. The $\pm1\sigma$ and $\pm2\sigma$ uncertainties in the expected limits are also tabulated.

T mass [GeV]	Observed limit on $\lvert\sin\theta_L\rvert$	Expected limit on $\lvert\sin\theta_L\rvert$ $^{+1\sigma/+2\sigma}_{-1\sigma/-2\sigma}$	Observed limit on c_L^{Wb}	Expected limit on c_L^{Wb} $^{+1\sigma/+2\sigma}_{-1\sigma/-2\sigma}$
800	0.18	$0.19\,^{0.04/0.08}_{0.03/0.06}$	0.25	$0.27\,^{0.06/0.11}_{0.05/0.08}$
900	0.24	$0.20\,^{0.05/0.09}_{0.05/0.07}$	0.34	$0.29\,^{0.07/0.13}_{0.07/0.10}$
1000	0.20	$0.21\,^{0.06/0.08}_{0.07/0.09}$	0.29	$0.30\,^{0.08/0.12}_{0.10/0.12}$
1100	0.25	$0.27\,^{0.09/0.11}_{0.13/0.15}$	0.36	$0.38\,^{0.12/0.15}_{0.18/0.21}$
1200	0.35	$0.35\,^{0.13/0.14}_{0.22/0.23}$	0.49	$0.49\,^{0.18/0.20}_{0.31/0.33}$

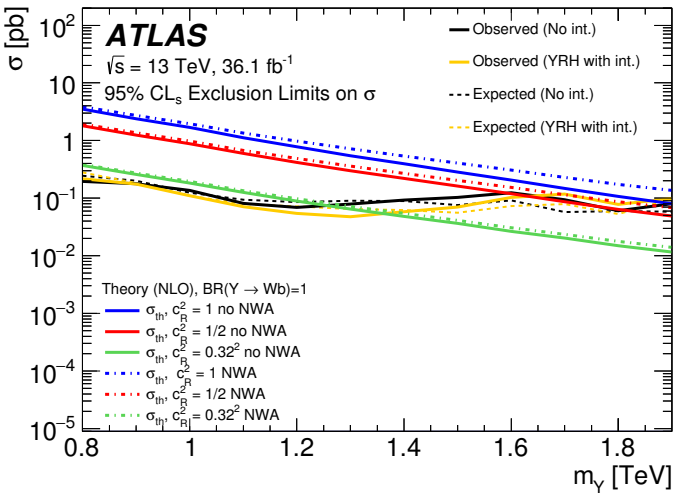


Figure 8.11: Observed (solid line) and expected (short-dashed line) 95% CL_s limits on cross-section times branching ratio for the right-handed Y quark as a function of m_{VLQ}. Here, the branching ratio $\mathcal{B}(Y \to Wb)$ is set to one. The computed theoretical NLO cross-sections for different coupling values are shown for the calculation using the narrow-width approximation (dashed-dotted lines) and without using the narrow-width approximation (solid lines) as described in the text.

Table 8.6: Observed and expected 95% CL_s upper limits on $|\sin\theta_R|$ and c_R^{Wb} for a right-handed Y quark for theoretical (B, Y) doublet interpretation with masses of 800–1 900 GeV. The $\pm1\sigma$ and $\pm2\sigma$ uncertainties in the expected limits are also tabulated.

Y mass [GeV]	Observed limit on $\lvert\sin\theta_R\rvert$	Expected limit on $\lvert\sin\theta_R\rvert^{+1\sigma/+2\sigma}_{-1\sigma/-2\sigma}$	Observed limit on c_R^{Wb}	Expected limit on $c_R^{Wb\ +1\sigma/+2\sigma}_{\ -1\sigma/-2\sigma}$
800	0.17	$0.20^{0.04/0.08}_{0.03/0.05}$	0.24	$0.28^{0.05/0.12}_{0.04/0.07}$
900	0.18	$0.19^{0.04/0.08}_{0.03/0.05}$	0.26	$0.27^{0.05/0.11}_{0.04/0.07}$
1000	0.17	$0.17^{0.03/0.07}_{0.03/0.05}$	0.25	$0.25^{0.04/0.10}_{0.04/0.07}$
1100	0.17	$0.18^{0.03/0.07}_{0.03/0.05}$	0.24	$0.25^{0.05/0.10}_{0.04/0.07}$
1200	0.17	$0.20^{0.04/0.08}_{0.03/0.05}$	0.25	$0.28^{0.05/0.11}_{0.04/0.08}$
1300	0.19	$0.22^{0.04/0.09}_{0.03/0.06}$	0.27	$0.31^{0.06/0.12}_{0.05/0.08}$
1400	0.24	$0.25^{0.05/0.10}_{0.04/0.07}$	0.35	$0.36^{0.06/0.14}_{0.05/0.10}$
1500	0.31	$0.28^{0.05/0.11}_{0.04/0.07}$	0.44	$0.39^{0.07/0.15}_{0.06/0.11}$
1600	0.45	$0.37^{0.08/0.19}_{0.06/0.10}$	0.64	$0.53^{0.11/0.27}_{0.08/0.14}$
1700	0.59	$0.46^{0.10/0.25}_{0.08/0.13}$	0.83	$0.65^{0.15/0.36}_{0.11/0.18}$
1800	0.55	$0.43^{0.09/0.22}_{0.07/0.12}$	0.77	$0.61^{0.13/0.32}_{0.10/0.17}$
1900	0.80	$0.79^{0.24/0.69}_{0.18/0.29}$	1.13	$1.11^{0.34/0.98}_{0.26/0.41}$

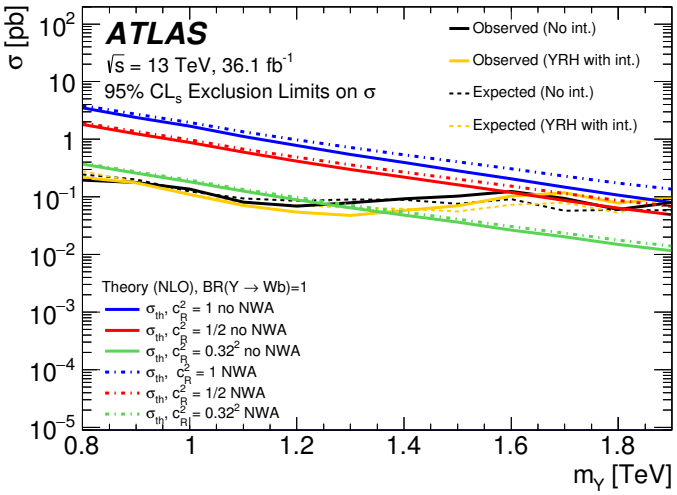


Figure 8.12: Observed (solid line) and expected (short-dashed line) 95% CL_s limits on cross-section times branching ratio for the case of the right-handed Y quark as a function of VLQ mass. Here, the branching ratio $\mathcal{B}(Y \to Wb)$ is set to one. The theoretical NLO cross-sections for different coupling values are shown for both with (dashed-dotted lines) and without (solid lines) the narrow-width approximation as described in Section 2.6.3.

Ranking plot

In an analysis, the statistical results can be either statistically limited or systematically limited. Since, in this analysis, the limits on signal strength are the results of interest, the effects of systematic uncertainties on the signal strengths are explicitly checked. Evidently, the fitted signal strength with and without using the systematic uncertainties as nuisance parameters differ significantly. This implies that this analysis is systematically limited i.e. the sensitivity of the results can improve when the uncertainties are controlled. The fitted signal strength μ for a signal-plus-background fit to data in SR and CRs including both systematic and statistical uncertainties and only statistical uncertainties for a Y RH signal with m_{VLQ} of 1 200 GeV are $\mu = -0.50 \pm 0.70$ and $\mu = -0.10 \pm 0.42$ respectively, which demonstrates that the search is hence limited by systematic uncertainties rather than statistical uncertainties. The top 20 systematic uncertainties with the largest impact is shown in Fig. 8.13. The pre-fit (post-fit) impact of each nuisance parameter on the signal strength μ, $\Delta\mu$, is calculated by comparing the nominal best-fit value of μ with the result of the fit while fixing the considered nuisance parameter to its best-fit value, $\hat{\theta}$, moved by its pre-fit (post-fit) uncertainties, $\pm\Delta\theta(\pm\Delta\hat{\theta})$. The fit used for measuring this impact is a fit to data for the signal-plus-background using the respective signal.

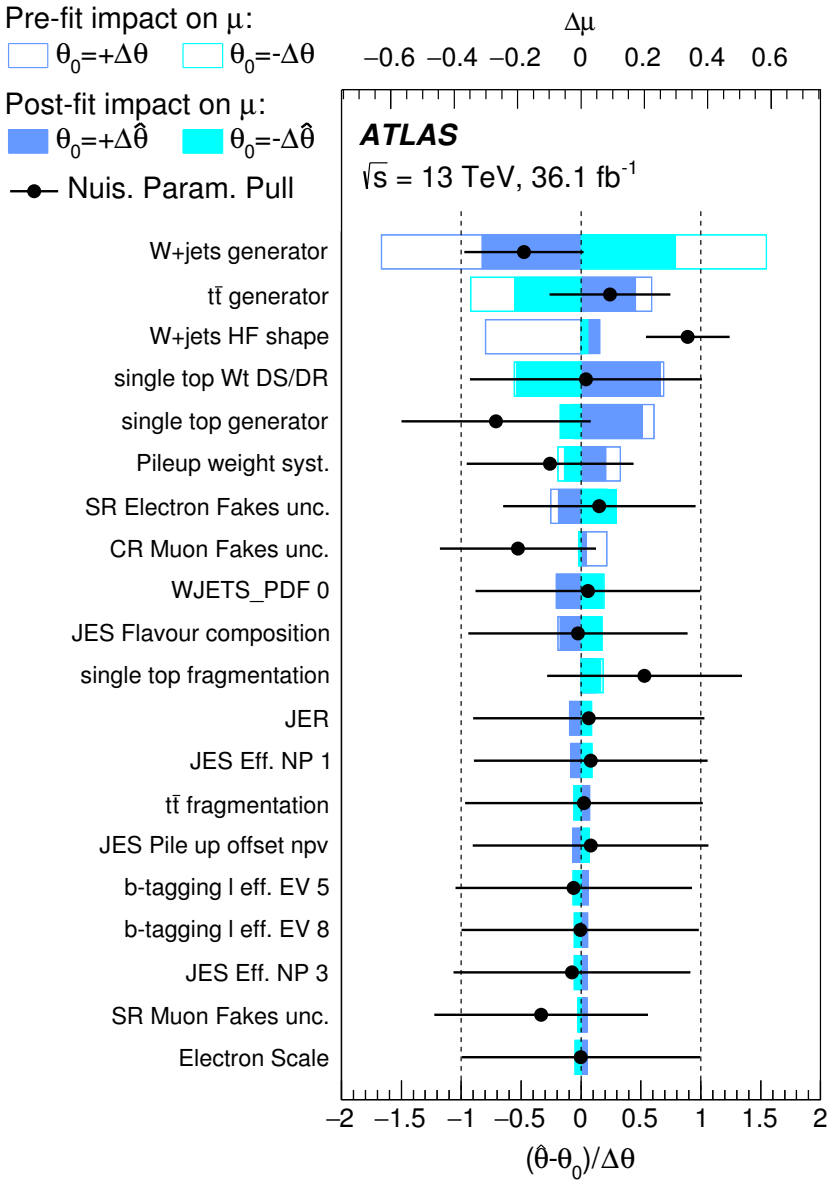


Figure 8.13: Ranking of the nuisance parameters included in the fit according to their impact on the the measured signal strength, μ, for the right-handed Y signal of mass 1 200 GeV with coupling of $c_R^{Wb} \approx 0.28$ and considering interference of the signal with the SM background extracted from a fit on data with this 1200 GeV Y signal injected and fitted. Only the top 20 nuisance parameters are shown without including those related to MC statistical uncertainties.The empty rectangles denote the pre-fit impact on μ and the filled ones to the post-fit impact on μ, both corresponding to the upper scale. The pulls of the nuisance parameters relative to their nominal values, θ_0. These pulls and their relative post-fit errors, correspond to the lower scale.

Part III

Hadronic analysis

CHAPTER 9

Event selection

In this chapter, the analysis strategy, optimisation of the kinematic cuts, and background modelling for the hadronic analysis using the 2015-2018 dataset corresponding to an integrated luminosity of 139 fb^{-1} will be discussed. The search for single production of a T/Y decaying into a Wb topology has up to now been performed only in the single-lepton final state i.e. $pp \to Q(T/Y) + q + X$ with $Q \to b + W$ and $W \to \ell + \nu$. This analysis is the first ever $T/Y \to Wb$ search where the hadronic decay mode od W-boson i.e. $W \to q\bar{q}$ is probed.

9.1 Analysis strategy

The analysis discussed here is a search for single production of a Y or T VLQ decaying to a Wb topology, where the W-boson subsequently decays hadronically. Kinematic cuts based on the event topology of the signal production are used to find a phase-space with high signal to background ratio. Since the limits from the previous single-lepton analysis as discussed in this thesis are in the TeV range, the signals probed in this analysis are in that mass range. The decay products of such a massive particle would have high angular separation and high-p_{T} to achieve energy-momentum conservation. The leading order Feynman diagram for such a signal is shown in Fig. 9.1. The hadronic decay products of the W-boson would have high p_{T} and being Lorentz-boosted can be reconstructed using a large-R anti-k_t (R =1) jet. Such a large-R jet is characterised by the presence of multiple prongs associated with the "subjets" from the hadronic decay of the W-boson. The large-R jet objects used here are described in Section 4.3.3. The identification of the hadronically decaying W-boson is performed using a 3-variable W-tagger as discussed in Section 4.3.4. Similar to the single-lepton final state, the outgoing light-flavoured jet generally falls in the forward region of the detector and thus has high values of $|\eta|$. The second b-jet originating from the gluon splitting tends to fall outside of detector acceptance in most cases.

The dominant background contribution for this search is multijet production, which predominantly constitutes jets from light-flavour quarks and gluons. The sub-dominant background contributions originate from top-quark pair production ($t\bar{t}$), single top-quark

production, and W/Z-boson production in association with jets (W/Z+jets). The modelling of the sub-dominant backgrounds and the signal will be discussed in the next section. A binned maximum likelihood fit using only the signal region is used to test the signal-plus-background (SM) hypothesis versus the background-only hypothesis to find the existence of a possible signal. The reconstructed vector-like quark mass (m_{VLQ})is used as a discriminant to separate signal against all the SM backgrounds. The m_{VLQ} is calculated by taking a vector sum of the Lorentz vectors of the leading large-R jet at a given working point efficiency of W-tagging and the leading small-R jet.

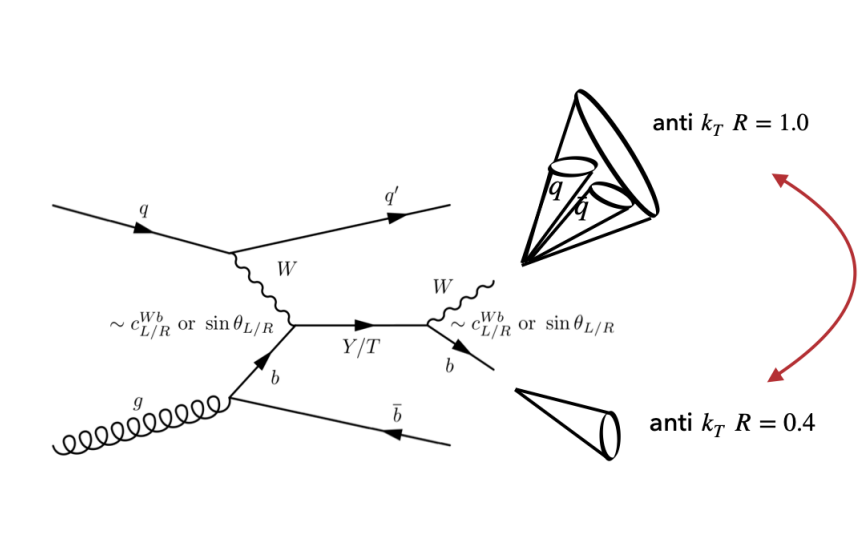


Figure 9.1: Leading order Feynman diagram for the $T/Y \to Wb$ decaying hadronically. The final state objects used for the search is also shown here. The two-sided red arrow signifies the high angular separation between the leading large-R W-jet and the leading small R jet that is necessary to achieve energy-momentum conservation during the decay of a relatively high-mass Y/T.

9.2 Data and simulated signal and background processes

Data samples

The data used for this analysis was pp collision data collected by the ATLAS detector from 2015 to 2018 at $\sqrt{s} = 13\,\mathrm{TeV}$. The peak recorded luminosity during this period was $2.1 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ with a bunch spacing of 25 ns. The average number of simultaneous pp interactions per bunch crossing, , $\langle\mu\rangle$ was measured to be 33.7 with the highest measurement of 36.1 being in 2018 . High data quality events are only events by selecting events pertaining to the Good Run Lists (GRL). The integrated luminosity of these GRL data events delivered by ATLAS in 2015-18 was 139.1 fb^{-1}.

Background processes

Similar to the single-lepton analysis, most of the SM backgrounds were simulated using Monte Carlo distributions. However, in this hadronic analysis, the major irreducible background, the final estimate of the multijets are not simulated. They are derived from the data using the ABCD method (see Section 9.6 for further details). Although the final estimate used in the analysis is data-driven, the simulated dijet estimate of the multijets is used in the ABCD method.

For this analysis, the final hadronic topology is dominated by the strong interaction. The final state probed in this search is a multijet one, but is quite close to the dijet topology given the two pronged nature of the Wb decay. Dijets and multijet topologies have served as a precise test of the strong interaction in the SM [211, 212]. The inclusive cross-section of the dijet process has also been measured double differentially using 2015 ATLAS data [213]. Monte Carlo distributions of the dijet process have been produced by the ATLAS collaboration using the PYTHIA 8.230 [197], SHERPA 2.2.5 [199] , HERWIG 7.1.3 [214], and MADGRAPH5_AMC@NLO 2.3.3.1 [200] + PYTHIA 8 generators. In this thesis, the default dijet (multijet) MC sample is the PYTHIA 8.230 sample. The samples are produced in slices of p_{T} of the leading truth anti-k_t jets. The SHERPA 2.2.5 samples have been used for uncertainty calculations. The HERWIG 7.1.3 MC sample was tested to calculate uncertainties but was finally not used due to its low statistics. For the PYTHIA 8 sample, the NNPDF2.3 [55] PDF set is used alongside the A14 tune for parton shower and multi-parton interactions.

$t\bar{t}$ events are produced using the NLO generator POWHEG-BOX v2 [215–217] interfaced with the NLO NNPDF3.0 [218] PDF set in the matrix element calculations. The samples are normalised to the cross-section calculated at next-to-next-to leading order (NNLO) in QCD including resummation of next-to-next-to-leading logarithmic soft gluon terms with Top++2.0 [219–225]. Parton showering, hadronization and the underlying event are simulated using PYTHIA v8.230 with the LO NNPDF2.3 PDF set and the A14 set of tuned parameters [226]. The top-quark kinematics in $t\bar{t}$ events are corrected to account for NLO electroweak effects [227]. The h_{damp} parameter in POWHEG-BOX which controls matrix element to parton shower matching and effectively regulates the high-p_{T} radiation, was set to $1.5m_t$ [228], where $m_t = 172.5\,\mathrm{GeV}$.

Single top-quark processes were generated with POWHEG-BOX v2 with the NNPDF3.0 NLO PDF set and interfaced with PYTHIA 8.230 with the A14 [226] tune using the NNPDF2.3 LO PDF set. Overlaps between the $t\bar{t}$ and Wt final states were removed using the "diagram removal" scheme (DR) [229, 230]. The "diagram subtraction" scheme (DS) [230] was used as an alternative method, and the full difference between the two methods assigned as an uncertainty [231].

W/Z+jets events were generated with the SHERPA 2.2.1 generator. The matrix-element calculation was performed with up to two partons at NLO and up to four partons at leading order (LO). The matrix-element calculation was merged with the SHERPA parton shower [232] using the ME+PS@NLO prescription [233]. The PDF set used for the matrix-element calculation was the NNPDF3.0 [218] NNLO PDF set.

Signal Modelling

Simulated events for T quark signal processes were generated at LO in the four-flavour scheme with the MADGRAPH5_AMC@NLO 2.6.5 generator using the NNPDF2.3 PDF set, interfaced to PYTHIA 8 for parton showering and hadronisation. The model from [234] is used in the computation of the matrix elements, and all tree-level processes are included. The coupling between the T quark and the gauge boson, κ, is set to 1.0 in the generated samples (1 to 2.7 TeV in 200 GeV steps). The matrix element-based weights are used to reweight these samples to other values of κ and mass.

Details regarding the generation of signal and background samples are tabulated in Table 9.1

Table 9.1: The generators employed to model the signals and various background processes are described, along with the parameter tune for the underlying event, PDF set, and highest-order perturbative QCD (pQCD) accuracy used for the normalisation of each sample. With the exception of the Tqb signals, all processes were generated at NLO in QCD. The LO cross-sections calculated for the Tqb signal processes in the simulation were normalised to the NLO theoretical cross-section taken from Ref. [234].

Process	Generator plus parton showering/hadronisation	Tuned parameters	PDF set: Gen. + show./had.	Inclusive cross-section order in pQCD
Tqb	MADGRAPH5_AMC@NLO 2.2.3 + PYTHIA 8.210	A14	NNPDF2.3 + NNPDF2.3	NLO
$t\bar{t}$	POWHEG-BOX v2 + PYTHIA 8.230	A14	NNPDF3.0 + NNPDF2.3	NNLO
Single-top	POWHEG-BOXv2 + PYTHIA 8.230	A14	NNPDF3.0 + NNPDF2.3	NNLO
W/Z **+ jets**	SHERPA 2.2.1	Default	NNPDF3.0	NNLO
Multijet	PYTHIA 8.186	A14	NNPDF2.3	LO

9.3 Pre-selection

Before optimising the signal-region kinematic cuts, a broad event-selection is employed based on the topology of the final state as shown in Fig. 9.1. This set of cuts, in addition to certain event-quality cuts, is defined as the Pre-selection region. The pre-selection region is characterised by events which fulfil the following criteria:

1. Events should be in the Good Run List (GRL) of the 2015-2018 ATLAS dataset. Events with LAr calorimeter noise burst and data corruption and incomplete events are vetoed.

2. Events should contain a primary vertex (see Section 4.1), the jets in the event should be match standard ATLAS quality criteria [203].

3. All events that have any charged lepton (electron/muon) (see Section 4.2.1 and Section 4.2.2) present are vetoed.

4. Since the W-boson decays into a large-R jet, events are required to pass a set of large-R jet triggers [204, 205, 235, 236].

5. Events are required to have exactly one W-tagged (50% working point 3 variable-tagger) large-R jet with a $p_{\mathrm{T}} > 500\,\mathrm{GeV}$. This cut is motivated by the fact that the trigger efficiencies for all large-R triggers used here are 100% for $p_{\mathrm{T}} > 500\,\mathrm{GeV}$.

6. The leading small-R jet is required to have transverse momentum $p_{\mathrm{T}} > 350\,\mathrm{GeV}$. All the signals searched in this analysis have masses greater than 1 TeV and this cut is used to suppress multijet background contribution beyond the region of interest of the search.

7. At least one forward jet ($p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $2.5 < |\eta| < 4.5$) is required in the event as shown in Fig. 9.1.

The kinematic distributions of the transverse momentum (p_{T}) , mass, ϕ and η distributions of the leading large-R jet, the transverse momentum (p_{T}), the ϕ, and η distribution of the leading small-R jet of the leading small-R jet after applying the pre-selection cuts are shown in Figs. 9.2 and 9.3 respectively. The angular distance ΔR between the leading large-R jet and the leading small-R jet ΔR(leading W-tagged-jet, leading small-R jet) and the azimuthal separation between the leading W-tagged large-R jet and the leading small-R jet are illustrated in Fig. 9.4. The multijet events used in Fig. 9.2, Fig. 9.3, and Fig. 9.4 are the PYTHIA 8 MC samples. The disagreement among the data and backgrounds for all variables arises from the multijet MC sample. Thus, a data-driven estimate is used to search for the VLQs. The reconstructed vector-like mass (m_{VLQ}) is used as a discriminant for this analysis. Here, m_{VLQ} is calculated as a vector sum of the Lorentz vectors of the leading small-R jet and the leading large-R jet.

9.4 Overlap removal and definition of leading small-R jet

As depicted in Fig. 9.4, the distributions of the angular separation variables between the leading large-R jet and the leading small-R jet, $\Delta R(W, \text{lead small-R jet})$ and $\Delta\phi(W, \text{lead small-R jet})$ show a significant overlap among both these objects. Thus, an overlap removal between these two objects is performed to reconstruct these objects better . To achieve this, a re-definition of the small-R jet is performed. The following steps are executed to achieve this.

1. If for an event, $\Delta R(\text{leading large-R jet}, \text{leading small-R jet}) > 1$, the leading small-R jet is used without any overlap removal.
2. If for an event, $\Delta R(\text{leading large-R jet}, \text{leading small-R jet}) < 1$ but the $\Delta R(\text{leading large-R jet}, \text{sub-leading small-R jet}) > 1$ the sub-leading small-R jet is re-defined as the leading small-R jet.
3. If for an event, $\Delta R(\text{leading large-R jet}, \text{leading small-R jet}) < 1$ and $\Delta R(\text{leading large-R jet}, \text{sub-leading small-R jet}) < 1$, the event is vetoed.

9.5 Definition of signal, control, and validation regions

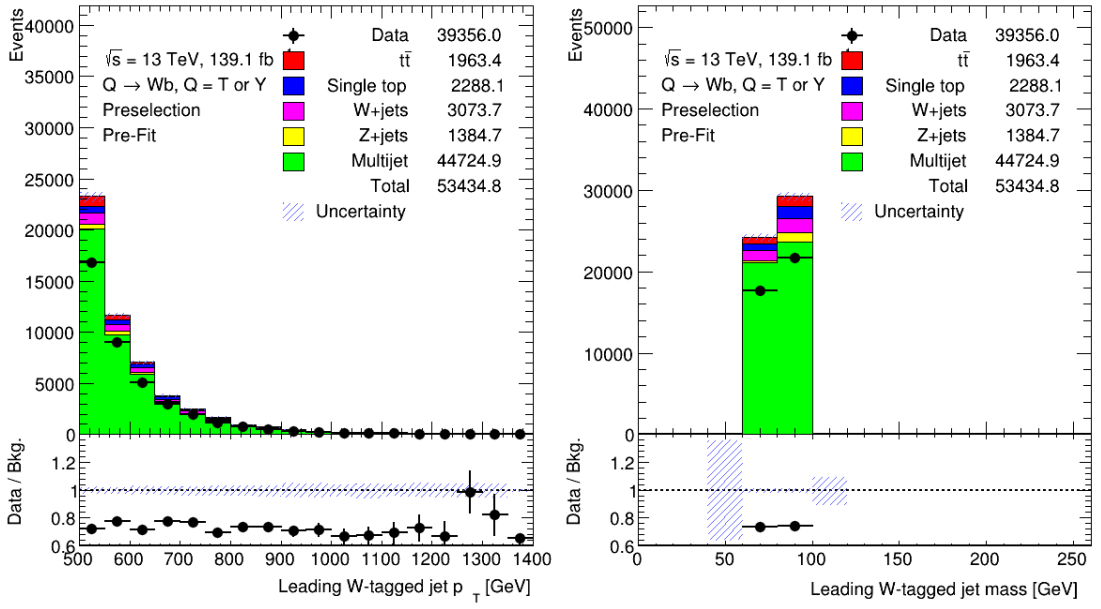
There are six regions which are defined for this analysis by making selection cuts on two variables, the working point efficiency of W-tagging and b-jet multiplicity. These six regions constitute of one signal region (SR), one validation region (VR), and four control regions. These two variables are chosen for the ABCD method to estimate the multijet background (see section Section 9.6)[1]. In Fig. 9.5, the b-jet multiplicity is plotted on the x-axis and the W-tagging WP on the y-axis. In this analysis, all b-tagged jets are defined at 70% selection efficiency of the DL1r b-tagger as detailed in Section 4.3.4. This working point was chosen because it provided an optimal balance between light-flavoured jet suppression (hence suppressing the multijet background) and event statistics needed for this search. The signal over background ratio was optimised over the mass-coupling phase space that is probed in this analysis. In addition to the cuts detailed below, all these regions have the following cuts in common.

1. Pre-selection cuts defined in Section 9.3 but with the leading small-R jet redefined using overlap removal as described in Section 9.4. Additionally, the W-tagging working point is tuned accordingly e.g. for the signal region the working point is unchanged, whereas for the to be validation region the leading large-R jet is W-tagged for the loose selection (80% working) but not tight selection (50% WP)
2. The p_{T} cut for the leading forward jet is increased to 40 GeV. This ensures considerable background suppression.
3. The leading small-R jet should be central i.e. $|\eta(\text{leading small-}R\text{ jet}| < 2.5$.

Signal region

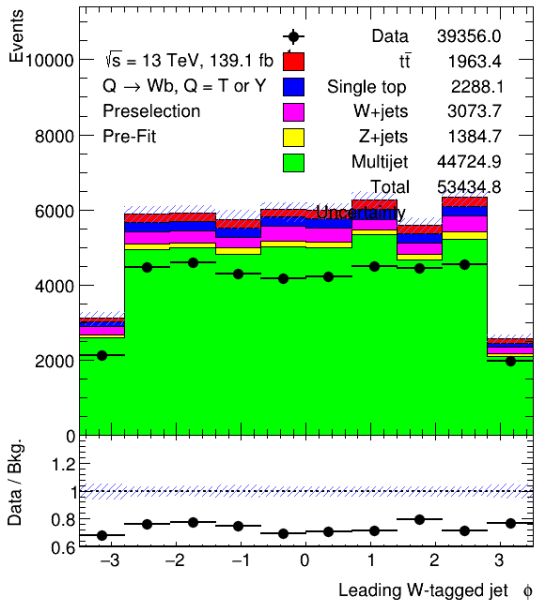
Region A1, represented as SR, contains those events which pass the following criteria: the leading large-R jet should be W-tagged at tight selection (50% WP). The events should contain at least one b-tagged jet, and the leading small-R jet as defined in Section 9.4 (denoted by j_1) should lie within the central region of the detector i.e $|\eta_{j_1}| < 2.5$.

[1] A detailed explanation for the choice of these two variables and the ABCD method is provided in Section 9.6

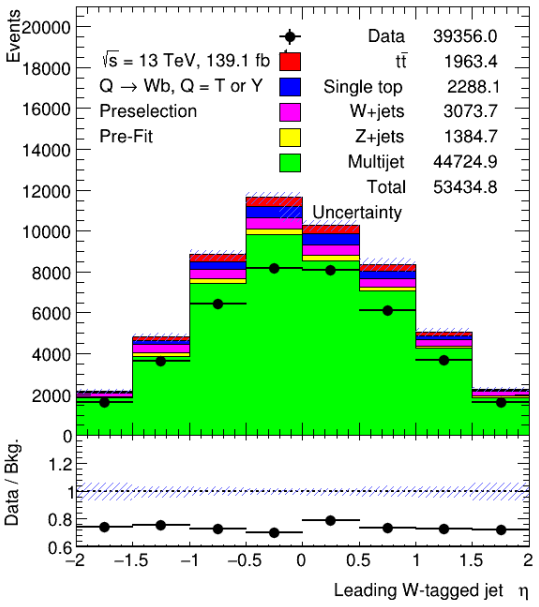


(a)

(b)

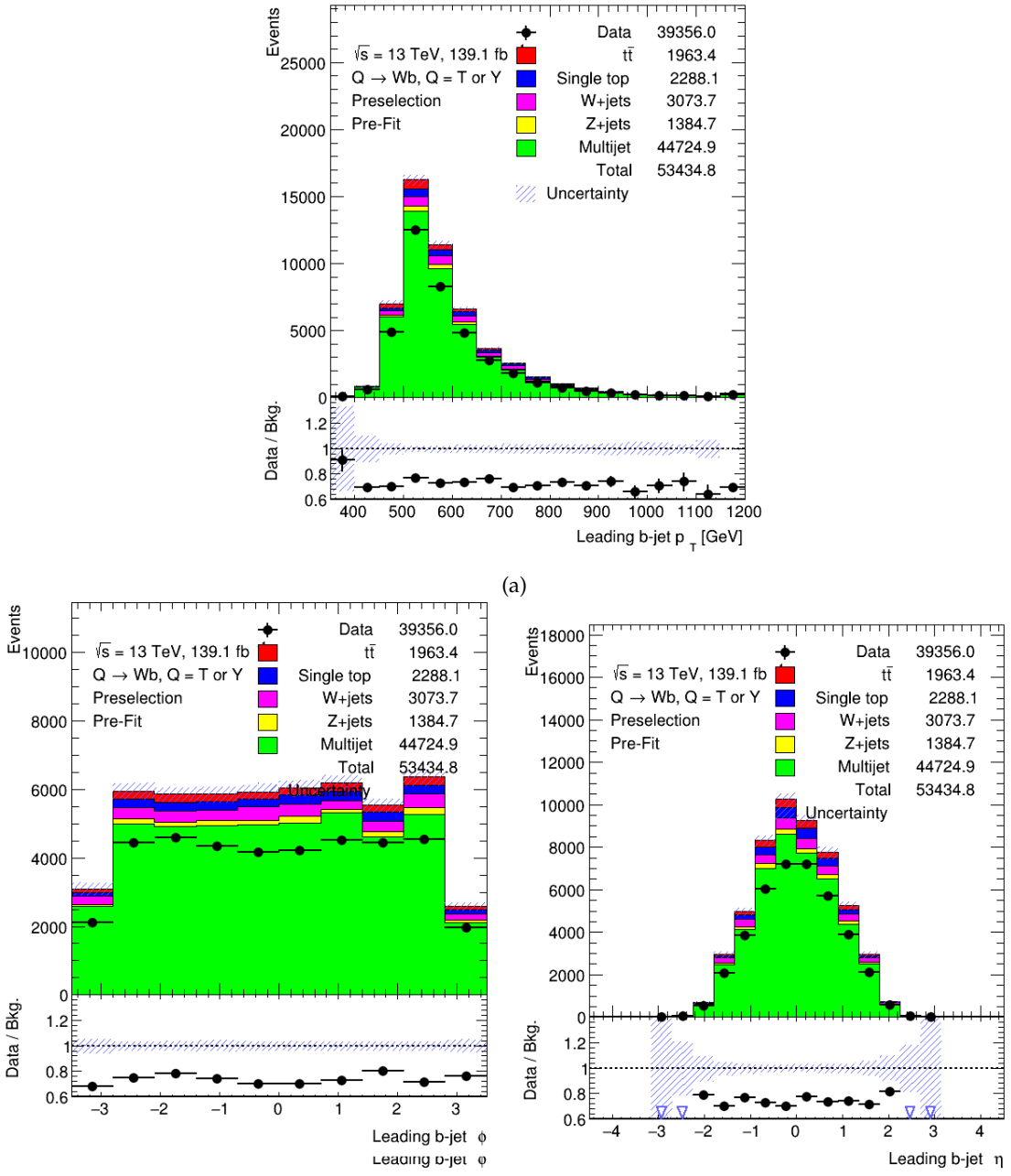


(c)



(d)

Figure 9.2: Data and Standard Model Monte Carlo distributions of the (a) p_{T}, (a) mass, (c) ϕ and (d) η of leading large-R jet after applying the pre-selection cuts. The shaded error band depicts the statistical uncertainty of the SM prediction, the error attached to the data points is the corresponding statistical uncertainty.

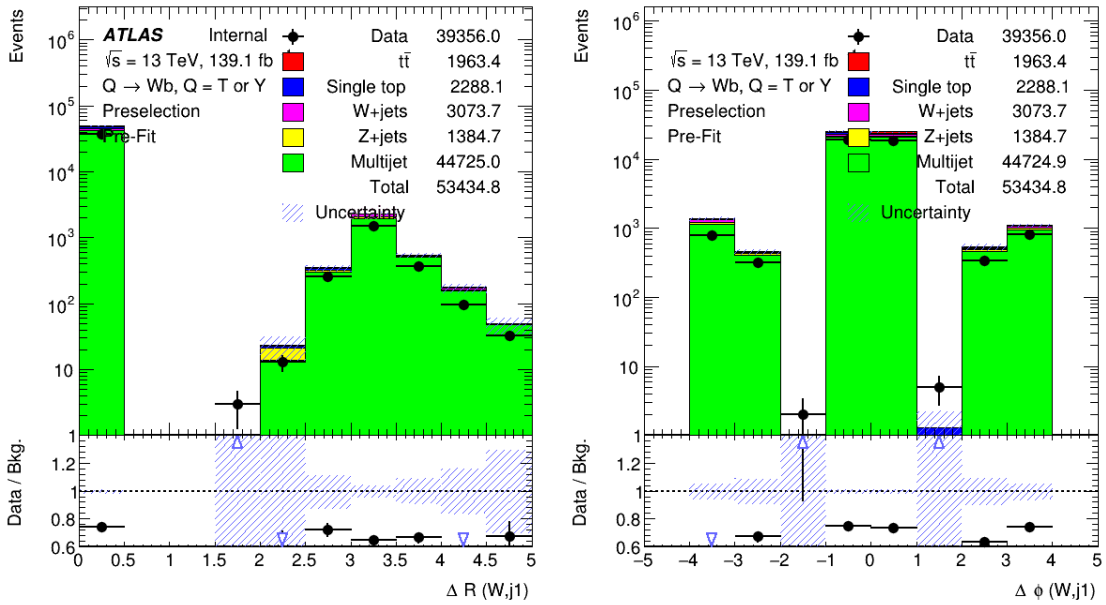


(a)

(b)

(c)

Figure 9.3: Data and Standard model Monte Carlo distributions of the (a) p_{T}, (b) ϕ and (c) η of leading b-jet after applying the pre-selection cuts. The shaded error band depicts the statistical uncertainty of the SM prediction, the error attached to the data points is the corresponding statistical uncertainty.



(a) (b)

Figure 9.4: Data and Standard model Monte Carlo distributions of the (a) ΔR, (b) $\Delta\phi$ between the leading large-R jet and the leading small-R jet after applying the pre-selection cuts. The shaded error band depicts the statistical uncertainty of the SM prediction, the error attached to the data points is the corresponding statistical uncertainty.

Validation region

Validation region (VR) A is designed to be kinematically close to the SR and partly orthogonal to the SR. The selection criteria include the leading large-R jet to be W-tagged at loose (80% WP) not tight (50% WP) with a requirement of at least one b-tagged jet, where the leading small-R jet lies in the central region of the detector. As this region is kinematically close to the SR, similar performance of background modelling and statistical fitting as compared to the SR is expected.

Control region

Regions B, C, D, and D1, defined as the CRs have the following selection cuts.

- **Region B:** This region is defined using events which should at least have a large-R jet, and the leading large-R jet should not be W-tagged at the loose selection. Events are required to have at least one b-tagged jet, where the leading small-R jet lies in a central region of the detector.
- **Regions C, D and D1:** These three regions contain events in which none of the small-R jets is b-tagged and the leading small-R jet should be a central jet. The selection criteria

that differentiates the events in these three regions is the requirement of leading large-R jet not to be W-tagged in the case of region C, W-tagged at tight, and loose-not-tight selection for regions D and D1 respectively.

A pictorial representation of the regions is shown in Fig. 9.5

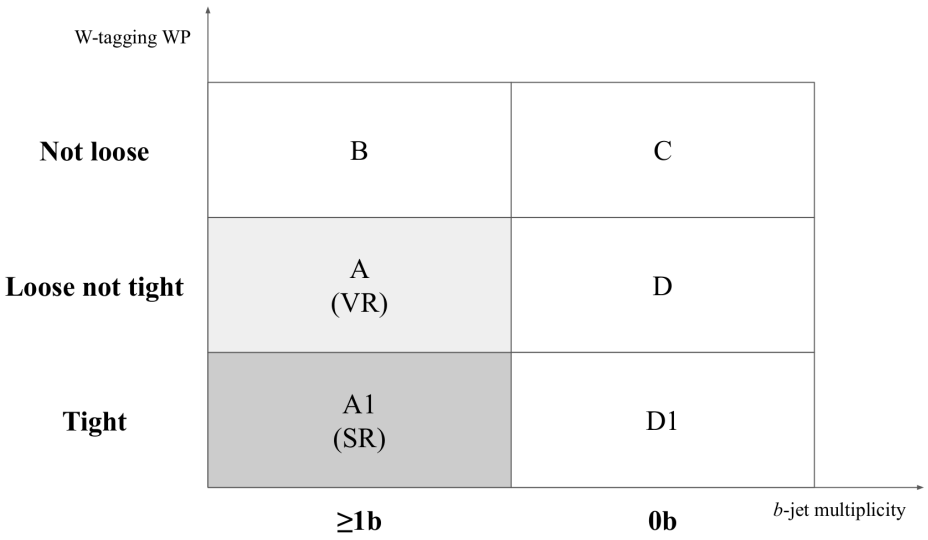


Figure 9.5: A representation of all the ABCD regions defined on two uncorrelated variables for the hadronic channel. Region A1 is the signal region; region A is the validation region and the other four regions B, C, D, D1 are control regions.

9.6 Multijet background estimation

The data and MC events selected in the Pre-selection region reveal a significant mismodelling of the multijet background. The disagreement between data and MC is clearly observed in Figs. 9.2 to 9.4. As the multijet background is the main irreducible background, it has to be modelled using data-driven methods. Here, a data-driven estimation of the multijet background using the ABCD method [237] is performed[2]. Similar ATLAS collaboration analyses where BSM physics is searched using fully hadronic states also employ the ABCD method to model the multijet background [239, 240]. The ABCD method used in this analysis is motivated by the ABCD method used in the BSM search of $W' \to tb$ in the hadronic final state [240]. The ABCD method uses two uncorrelated variables in orthogonal regions. By combining these two variables four kinematic regions, A, B, C, and D are created. The region A is a signal enriched region where the background of interest is modelled. The ABCD method requires the regions B, C, and D to be signal deficient. Consequently, the ABCD method formulates that the number of modelled background events in region A as

[2] The first iteration of the ABCD method for this VLQ search was first developed in Ref. [238].

$$N_{\mathrm{A}} = N_{\mathrm{B}} \times \frac{N_{\mathrm{D}}}{N_{\mathrm{C}}} \tag{9.1}$$

where N_j isolate the number of background events in region $j \in A, B, C, D$.

As discussed in Section 4.3.4, the W-tagging methodology used in this thesis does not use any b-tagging information. Additionally, the VLQ T/Y decay is probed using W and b tagged jets. Thus, using b-tagging and W-tagging information for the ABCD method is justified.

9.6.1 Choice of variables

The working point efficiency of W-tagging is selected as one of the variables for the ABCD method. The working points used in this thesis correspond to the 3-variable W-tagger discussed in Section 4.3.4. The two selected working points (WP) are 80% referred as *loose*, and 50% referred as *tight*. The tight WP is a subset of the loose WP because it has a stricter criteria. The three kinematic regions defined by using the selections based on these two working points are tight, loose-not-tight, and not loose. They are described in detail below:

- Tight: This region is constructed with events containing large-R jets that are W-tagged at 50% WP.
- Loose-not-tight: This region is constructed with events which have large-R jets that are W-tagged at 80% WP but not tagged at 50% WP.
- Not loose: This region contains all the events which have large-R jets that are not W-tagged at 80% WP.

The multiplicity of b-tagged jets is chosen as the second variable. The two kinematic regions which are defined based on this variable are $0b$ and $\geq 1b$. Both these regions are described in detail below:

- $0b$: This region is designed using events which do not possess any small-R jets that are b-tagged.
- $\geq 1b$: This region is comprised of events which have at least one small-R jet that is b-tagged.

In this thesis, in addition to the four regions, an additional validation and a control region have also been defined. Prior to directly checking the prediction of the ABCD method in the SR, the method is validated in the VR. The 6 regions needed for the ABCD method is detailed in Section 9.6.1. To calculate the background estimate of multijets in SR A1, the control regions B, C, and D1 are used whereas to estimate the multijet background events in VR A, the control regions B, C, and D are used.

9.6.2 Closure tests for ABCD method

Before proceeding with the estimation of the multijet background using the ABCD method, two preliminary closure tests are performed. These are described in detail below:

Test I One of the fundamental assumptions of the ABCD method is the control regions being signal deficient. A closure test of this assumption was performed by checking the $S/\sqrt{B}$ and S/B yield ratio in all the regions. The signal samples used for this test are the Tqb samples with m_{VLQ} of 1.1 TeV and 1.5 TeV and respective couplings of $\kappa = 0.25$ and $\kappa = 0.45$. These signals are chosen in the mass-coupling phase space because they are close to the limits calculated in the single lepton analysis performed in his thesis (see Chapter 8). The background signals chosen for this closure test are the Monte Carlo samples mentioned in Table 9.1. In Fig. 9.6 the $S/\sqrt{B}$ and S/B ratio for both the signal samples in all the regions are shown. The S/B in all the CRs is less than 1%, thus validating the tested assumption.

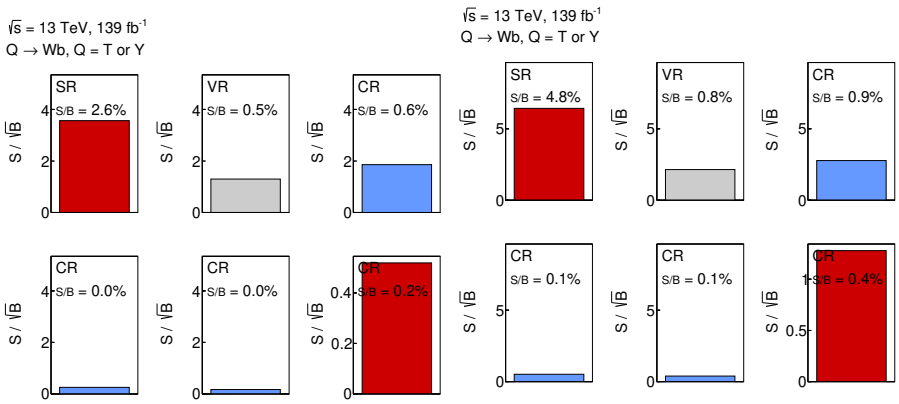


Figure 9.6: Signal-background ratio in all the six ABCD regions for (a) 1.1 TeV signal with $\kappa = 0.25$ and (b) 1.5 TeV signal with $\kappa = 0.45$

Test II Prior to using the data-driven ABCD method, Eq. (9.1) needs to be validated on the multijet MC samples. If the two chosen variables are perfectly uncorrelated, Eq. (9.1) should hold true for the multijet MC samples. The calculated event yields in SR A1 by using regions B, C, D1 is compared against the event yields of multijet MC samples. The same comparison is performed for VR A by using regions B, C, and D. Table 9.2 depicts the yields of multijet MC in the SR and VR versus the calculated yields. A respective difference of 9% and 17% is observed in the SR and VR. This implies a slight correlation among the variables chosen. Thus, to incorporate this correlation in the data-driven ABCD method, a correction factor R_{corr} is introduced.

Table 9.2: Event yields of multijet Pythia8 background from the MC sample in SR A1 and VR A versus the expected yield calculated using Eq. (9.1). The errors presented here are purely statistical.

Multijet	SR A1	VR A
MC yields	$14\,142 \pm 276$	$63\,147 \pm 591$
Expected yields	$15\,489 \pm 183$	$76\,054 \pm 729$
Difference to MC	9%	17%

9.6.3 Implementation of the ABCD method

The ABCD method is implemented with the inclusion of the correction factor R_{corr}. Since the ABCD method is designed to calculate the yields of the multijet background, the other backgrounds are subtracted from the data yields. Using the subtraction step, the expected data-driven yields for multijets are calculated for the CRs B, C, D, and D1. These data-driven multijet yields for B, C, and D are inserted into Eq. (9.1) to calculate the yields for VR A and the yields for B, C, D1 are used to calculate the yields for SR A1. In this analysis, since the search of T/Y VLQ is conducted across the whole range of the discriminant m_{VLQ}, both the shape and the normalisation of the multijet background are important. Thus, the ABCD method is implemented in each bin i with the correction factor R_{corr} calculated per bin. The final formula for the per bin ABCD method to calculate the final multijet yields for SR A and VR A1 can be written as

$$N_{\text{A/A1}}^{\text{Est. multijet}}[i] = R_{\text{corr}}[i] \times (N_{\text{B}}^{\text{Data}}[i] - N_{\text{B}}^{\text{Other bkg}}[i]) \times \frac{(N_{\text{D/D1}}^{\text{Data}}[i] - N_{\text{D/D1}}^{\text{Other bkg}}[i])}{(N_{\text{C}}^{\text{Data}}[i] - N_{\text{C}}^{\text{Other bkg}}[i])}, \tag{9.2}$$

where $R_{\text{corr}}[i]$ is the per-bin correction factor calculated from multijet MC samples. $R_{\text{corr}}[i]$, a measure of the residual correlation among the two variables. The per-bin correction factor is calculated using the multijet MC samples. The MC m_{VLQ} distributions for the CRs are shown in Fig 9.7. Evidently, these MC samples have a significant mismodelling across the full m_{VLQ} range. Calculating the R_{corr} factor from these distributions can lead to erroneous calculations. Thus, the normalisation of the multijet samples is fixed by fitting the multijet MC yields to data with a single bin maximum log likelihood fit under a background-only hypothesis. In these fits, the normalisation of the other SM backgrounds are not allowed to vary. The m_{VLQ} distributions after scaling the normalisation of multijets for CRs B, C, D, and D1 are depicted in Fig. 9.8. Using these normalisation scaled multijet yields, the R_{corr} factor is calculated according to Eq. (9.3).

$$R_{\text{corr}}[i] = \frac{N_{\text{A/A1}}^{\text{normalisation scaled multijet MC}}[i]}{N_{\text{B}}^{\text{normalisation scaled multijet MC}}[i]} \times \frac{N_{\text{C}}^{\text{normalisation scaled multijet MC}}[i]}{N_{\text{D/D1}}^{\text{normalisation scaled multijet MC}}[i]} \tag{9.3}$$

Here, $N_{\mathrm{A}}^{\text{normalisation scaled multijet MC}}[i]$ is used while calculating the ABCD yields for VR A and $N_{\mathrm{A1}}^{\text{normalisation scaled multijet MC}}[i]$ is used to calculate the ABCD yields for SR A1. For the VR A, the expected S/B is less than 1%. Consequently, the one bin fit of multijet MC yields to data under the background-only hypothesis is a fair approximation while calculating the R_{corr} factor. For calculation of R_{corr}, the assumption of using the background-only hypothesis for the SR A1 in the one bin fit is to fix the multijet normalisation not perfect, but it is better than using significantly mismodelled MC distributions to derive the R_{corr} factor. Furthermore, it is observed that the respective R_{corr} distributions for calculating the ABCD yields in VR A and SR A1 have certain statistical fluctuations. In order to smoothen these fluctuations, a third order polynomial fit is to derive the final R_{corr} distributions. The functions describing the shapes and values for $R_{\mathrm{corr}}\,[i]$ as a function of m_{VLQ} in GeV for the VR A and SR A1 are detailed in Eq. (9.4). Finally, these distributions are used in Eq. (9.2) to calculate the yields per-bin for multijets in the SR A1 and VR A.

$$\begin{aligned} f(x) &= -1.01\times10^{-10}x^3 + 5.7\times10^{-7}x^2 - 1.1\times10^{-3}x + 1.43 \;\;, x = m_{\mathrm{VLQ}} \text{ in GeV for VR A} \\ f(x) &= -3.7\times10^{-11}x^3 + 2.46\times10^{-7}x^2 - 5.7\times10^{-4}x + 1.14 \;\;, x = m_{\mathrm{VLQ}} \text{ in GeV for SR A1} \end{aligned} \tag{9.4}$$

The m_{VLQ} distributions for multijets calculated with the aforementioned ABCD method are compared against the data after adding the other SM backgrounds. These distributions are shown in Fig. 9.10. The uncertainties depicted in the plots include the statistical and systematic uncertainties for both multijets and other SM backgrounds. These uncertainties are described in detail in Chapter 10. The calculated yields with all the uncertainties compared to the other SM backgrounds and the observed data in both SR A1 and VR A are tabulated in Table 9.3 For the SR A1, a good shape and normalisation agreement within uncertainties is observed among the data and the sum of estimated multijet background and MC estimates of other SM backgrounds for the discriminating variable m_{VLQ}. For the VR A, although an agreement of shape among the data and and the sum of estimated multijet background and MC estimates of other SM backgrounds is observed, a discrepancy is observed in the normalisation. It should be noted that this discrepancy is within all uncertainties.

Table 9.3: Result showing the estimated multijet background with all the uncertainties in the validation region for all the kinematic distributions for the hadronic channel. The uncertainties depicted here include the statistical and systematic uncertainties for both multijets and other SM backgrounds. These uncertainties are described in detail in Chapter 10.

Distribution	Event yield (SR A1)	Event yield (VR A)
ABCD multijet estimate	$11\,220 \pm 743$	$68\,594 \pm 13\,554$
Other SM bkg.	$1\,995 \pm 863$	$2\,133 \pm 736$
Data	12 923	62 409

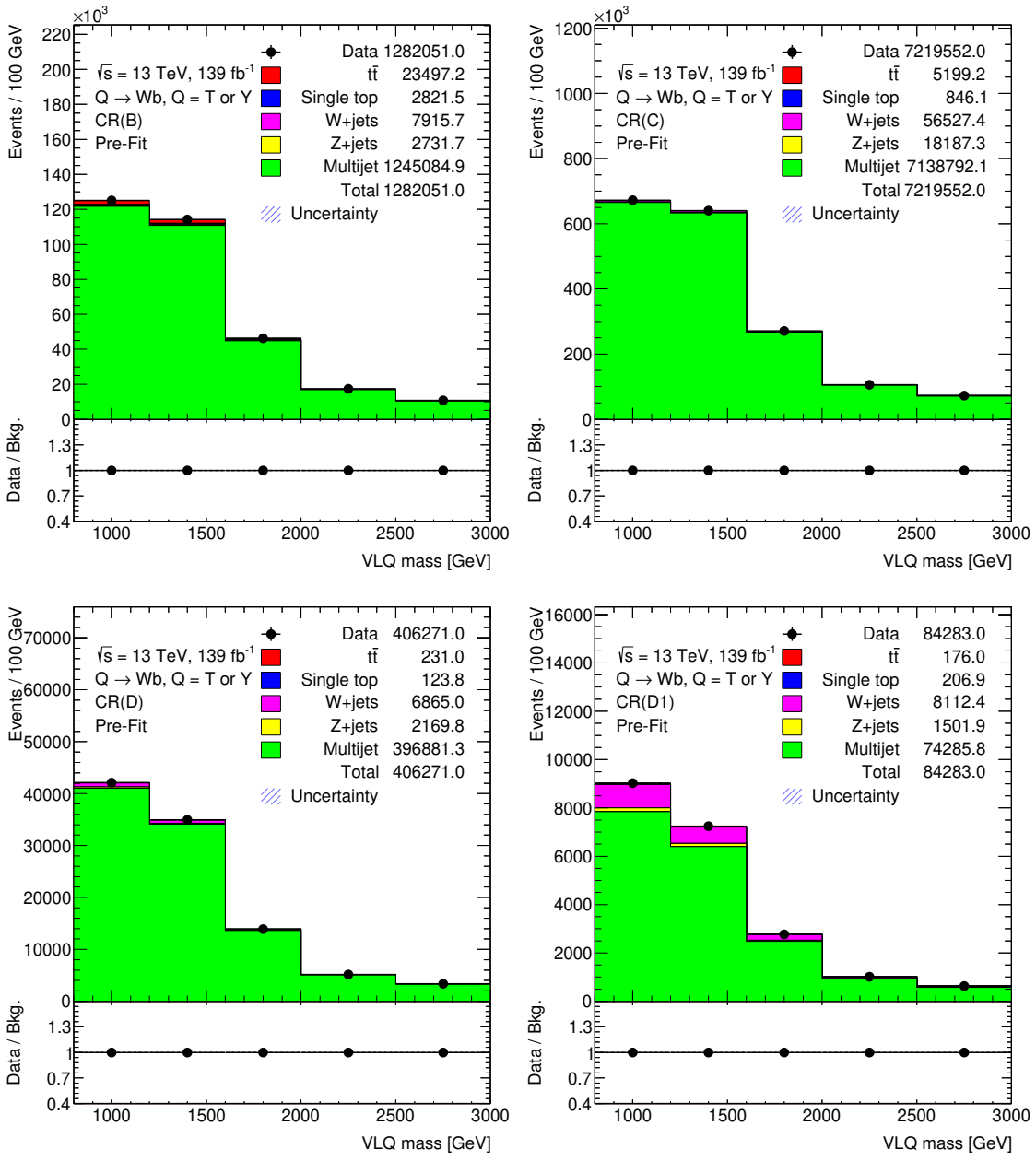


Figure 9.7: MC m_{VLQ} distributions for CRs B, C, D and D1.

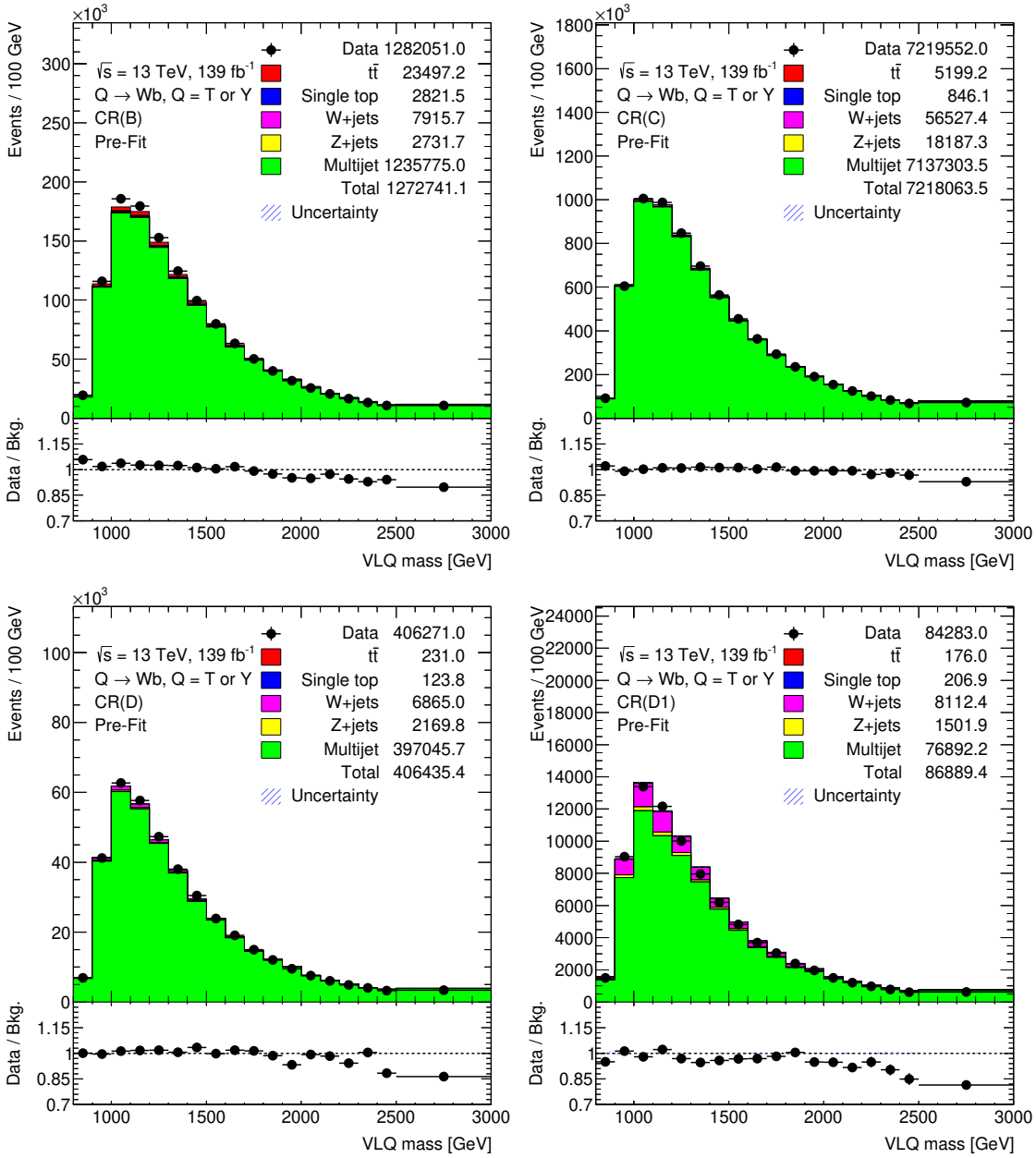
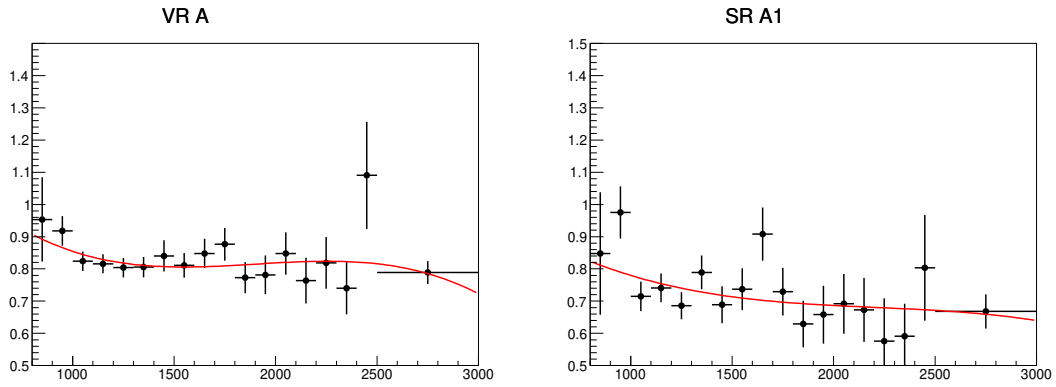


Figure 9.8: m_{VLQ} distributions for CRs B, C, D and D1 after scaling the normalisation of multijets to data using one bin maximum log likelihood fit. The yields of the other SM backgrounds are constants in these fits.



(a) (b)

Figure 9.9: Distribution of R_{corr} obtained using scaled multijet Pythia8 MC for (a) VR A and (b) SR A1. The red line in each of the figures is the fitted third degree polynomial as mentioned in Eqn. 9.4, which is used to smooth out the R_{corr} distribution. The final estimate is computed using the parametrised polynomial.

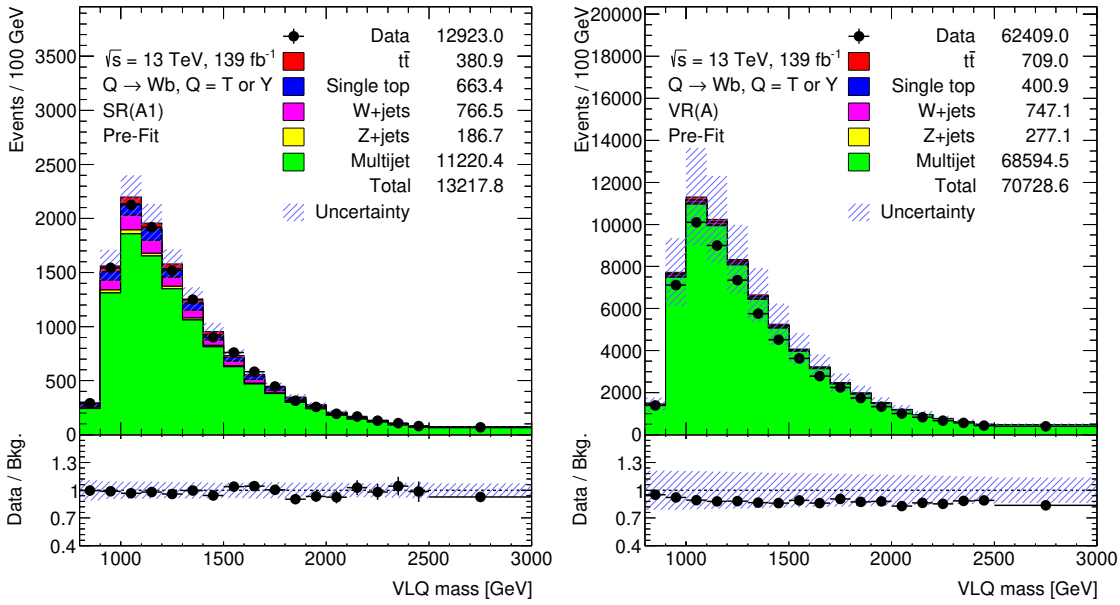


Figure 9.10: Estimated multijet background from the ABCD method including all the uncertainties in the signal region and the validation region respectively. Here, all the uncertainties for all other backgrounds are also included. These uncertainties will be discussed in detail in Chapter 10

CHAPTER 10

Systematic uncertainties

In this chapter, the uncertainties used in this analysis will be described. The uncertainties are applied to both the signal and the background. As discussed in Chapter 9, the main background, the multijets are modelled using the ABCD method. The uncertainties on the estimated multijet are assigned due to various parts of the ABCD method. For the other SM backgrounds and the VLQ T/Y signal, both experimental and statistical uncertainties are applied. These uncertainties will be discussed in this chapter. Similar to the treatment of uncertainties in the 1-lepton analysis discussed in Chapter 7, the uncertainties are used as nuisance parameters where the $\pm 1\sigma$ variations are inputs to the likelihood fit. The MC statistical uncertainties are implemented using the Barlow Beeston method as detailed in Chapter 5. The statistical uncertainties that arise from the ABCD method are also incorporated as separate uncertainties. All the nuisance parameters are used as either shape and/or normalisation uncertainties.

10.1 Experimental and theoretical uncertainties

The luminosity uncertainty on the 2015-2018 ATLAS dataset was measured to beam 1.7%[1]. The uncertainty is derived on the baseline luminosity measurement using the LUCID-2 detector [206] with a methodology discussed in Ref. [241]. This uncertainty is applied uniformly to all simulated signal and background background processes as a normalisation uncertainty. Due to the presence of a lepton veto, no lepton uncertainties have been included in this analysis.

The small-R jets used in this analysis are particle flow jets. The jet vertex tagger efficiencies, as measured in $Z(\rightarrow \mu\mu) +$ jets events, do not match in data and MC and thus the ratio is used as a per-event scale factor. The uncertainties on these scale factors are used as pileup uncertainties for all SM backgrounds and the signal yields.

[1] Although, the recent ATLAS measurements [241] has measured the uncertainty to be 0.83%, the samples used in this thesis were calculated with the previous recommendations of 1.7%

The JES uncertainties are propagated through the calibrations performed on the Particle Flow jet collection described in Section 4.3.2. "Relative in situ" dijet eta-intercalibration, "absolute in situ" Z+jet balance, "absolute in situ" gamma+jet balance, "absolute in situ" multi-jet balance, "absolute in situ" propagation of single particle and test beam uncertainties, pileup uncertainties, flavour-related uncertainties, punch-through uncertainty, and MC non-closure uncertainty are the individual sources of the JES uncertainties. The high-p_{T} region of interest is covered using single-particle uncertainties. In the ATLAS recommendations the most accurate JES uncertainties' configuration include 100 nuisance parameters, a reduced configuration of 30 nuisance parameters is the most appropriate for a search performed for this thesis as recommended by the Jet $E_{\mathrm{T}}^{\mathrm{miss}}$ CP working group [162].

The jet energy resolution (JER) is the measured average resolution of the ATLAS detector for the energy of jets. This being measured on data, the MC simulation jets are smeared using the resolution measured in data. Here, the JER uncertainties were calculated using the noise term evaluated in zero bias data using random cones and dijet asymmetry across p_{T} and η. Similar to the JES uncertainties, the most accurate uncertainty configuration has 34 nuisance parameters, but for this analysis using 8 JER nuisance parameters are the recommended configuration.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ hard term is calculated from the reconstructed hard objects. The corresponding uncertainties in the calculation are taken in addition to the uncertainties on the soft term resolution and scale calculation i.e. for the unassociated inner detector tracks.

The b-tagging DL1r classifier is used to identify b-jets in this analysis. The b-tagging working point is chosen to be 70% to optimise the signal to multijet background ratio. The uncertainties arise due to the misclassification of b-jets, c-jets, and light-flavoured jets. The tagging scale factors are calibrated using di-lepton $t\bar{t}$ events. The uncertainties arising from this are propagated to every event. These uncertainties are further eigen decomposed to 18 tagging nuisance parameters. Since the scale factors are only applied up to 300 GeV, a high-p_{T} extrapolation eigenvector component is used for jets in the high-p_{T} range. Additionally, a $c \to \tau$ b-tag extrapolation uncertainty is used as a nuisance parameter.

For the large-R jets, the "in situ" uncertainties are calculated for the jet energy scale (JES), jet energy resolution (JER), jet mass scale (JMS), and jet mass resolution (JMR). The "in-situ" corrections against reference objects and subsequent dijet η-intercalibration fixes the central value of JES. The uncertainties from these calibrations arise from detector effects and statistical modelling. Additional uncertainty components come from flavour response i.e. jet response differences between quarks and gluons, flavour composition accounting for quark gluon fraction, topology of V-bosons and top quarks. Similarly for the JER, as done for small-R jets, the dijet balance asymmetry is used to evaluate the JER. The random cone method is not used for large-R jets. Since the relative jet response is 1 within uncertainty, the JES smearing is not applied. The JES uncertainties propagated through JER calculation and the data/MC differences are taken as the JER uncertainties while smearing the MC to calculate all the JER uncertainties. The Forward Folding technique as described in Ref. [242] fits the W-boson and top-quark mass peaks in lepton-plus-jets $t\bar{t}$ events in order to calculate both the mass scale and resolution difference of large-R jets between data and MC. This new

approach improves upon the existing R_{trk} method, where the double-ratio of calorimeter versus track quantities in data and MC measures the uncertainty in large-R jets. These two complementary approaches are summed and the intermediary mass bins are interpolated. The JMS uncertainties arise from both these methods, the interpolation, and response differences among QCD jets and from W/Z/top jets. The JMR is also calculated using the Forward Folding method. The uncertainties arise from the method and the subsequent interpolation among mass bins.

The three variable W-tagging scale factors are calibrated using $t\bar{t}$ events in the lepton-plus-jets channel. The underlying systematic uncertainties are individually propagated to the signal (W-jet) scale factors. Subsequently, a quadratic addition of all the components is performed to calculate the final scale factor uncertainty. The uncertainty components used are statistical, theory, $t\bar{t}$ modelling, large-R jet, b-tagging,and other experimental uncertainties. For the high p_{T} range (pT > 600 GeV), the scale factors are extrapolated and hence the corresponding uncertainties are also propagated. For this thesis, 19 uncorrelated nuisance parameters are used to include the W-tagging scale factor uncertainties.

Theoretical cross-section uncertainties are used for the simulated SM backgrounds. For $t\bar{t}$, a $+2.4\%/-3.5\%$ uncertainty is assigned, for the single-top-quark backgrounds an uncertainty of $\pm 2.5\%$ is used whereas for both W/Z+jets a $\pm 6\%$ uncertainty is applied. All these cross-section uncertainties are implemented as normalisation nuisance parameters.

10.2 Modelling uncertainties on simulated backgrounds

The modelling uncertainties are assigned to single-top quark and W+jet events as they are the secondary irreducible backgrounds following the multijet background yields. Uncertainties related to initial-state radiation/final-state radiation (ISR/FSR) and parton shower were assessed for the single-top samples. Since the single-top samples consist of three distinct components (the s-channel, t-channel, and Wt processes), the evaluation of these uncertainties was performed separately for each process. To account for generator uncertainty, the variation between the MADGRAPH5_AMC@NLO +HERWIG++ sample and the POWHEG-BOX +HERWIG++ sample was propagated to the nominal POWHEG-BOX +PYTHIA sample. The ISR/FSR uncertainty was determined similarly to the uncertainty estimation for $t\bar{t}$ events. For the fragmentation uncertainty, the difference between the POWHEG-BOX +PYTHIA sample and the POWHEG-BOX +HERWIG++ sample was considered. Due to significant interference between the Wt and $t\bar{t}$ processes, two different schemas were employed to generate the Wt Monte Carlo (MC) samples. The default sample was the diagram removal (DR) method, while the variation sample was the diagram subtraction (DS) method. For the W+jets events, a variational sample reconstructed using the Lund jet plane is considered as the variational sample [243].

10.3 Multijet modelling uncertainties

The ABCD method used to estimate the multijet distributions entails a set of operations which correspond to certain uncertainties which are propagated to the final m_{VLQ} distribution to calculate various uncertainties. These uncertainties can be categorised as follows:

1. Experimental and theoretical uncertainties of other SM backgrounds described above propagated through the ABCD method. Thus, in place of the SM backgrounds, the corresponding up and down variations are used to estimate the ABCD up and down variations for the multjet estimate. Consequently, for all experimental and theoretical uncertainties, each nuisance parameter is also associated with the calculated multijet estimate.

2. Statistical uncertainties that arise due to the arithmetical operations of the method calculated using standard Gaussian propagation. These uncertainties are assigned per bin as a separate nuisance parameter.

3. Generator uncertainties which are calculated by calculating R_{corr} using the SHERPA dijet MC sample. The methods which are used for PYTHIA 8 are replicated using the SHERPA dijet MC. The difference between the PYTHIA 8 estimation and SHERPA estimation is used as an uncertainty.

4. The statistical uncertainties arising from the polynomial fit performed to model R_{corr} for the ABCD estimate.

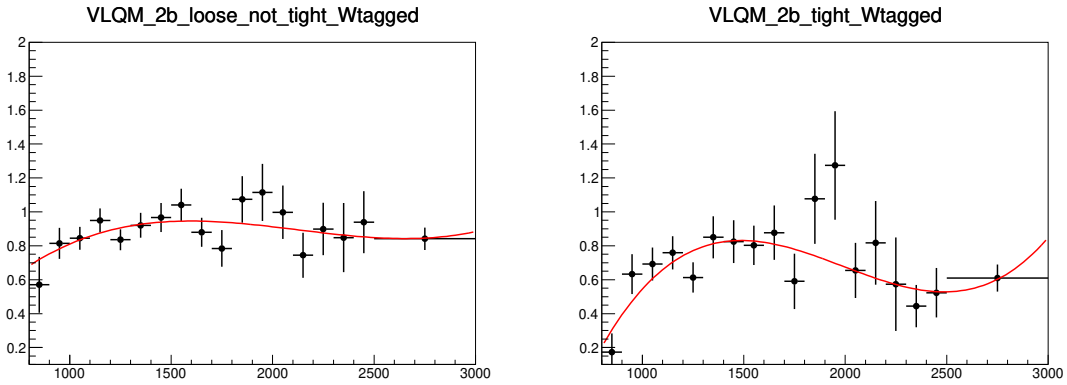
5. A non-closure uncertainty to measure the non-closure of the shape of the estimated multijet background using ABCD method in the VR after fitting the estimate to data solely in the VR.

Generator uncertainties

In this thesis, R_{corr} has been calculated using the PYTHIA 8 MC. The corresponding calculation was also performed using SHERPA dijet MC samples. The scaled dijet method of calculating R_{corr} is used for SHERPA in place of PYTHIA 8. The difference between the distribution using SHERPA for calculating R_{corr} and the ABCD estimate is used as an uncertainty. The smoothing of the R_{corr} distribution is performed similarly while calculating the estimated multijet using SHERPA MC samples. As it will be shown in Chapter 11, this uncertainty has one of the highest impacts on the signal strength. In the SR and VR, the uncertainty for SHERPA is shown in Fig. 10.2.

Non-closure multijet VR shape uncertainty

As shown in Fig. 10.3(a), the ABCD estimate of the multijet background in the VR summed with the SM backgrounds have a disagreement with data within uncertainties. To study this



(a) (b)

Figure 10.1: Distribution of R_{corr} obtained using scaled multijet Sherpa MC for (a) VR A and (b) SR A1. The red line in each of the figure is the fitted third degree polynomial used to smooth out the R_{corr} distribution. The final Sherpa uncertainty estimate is computed using the parametrised polynomial.

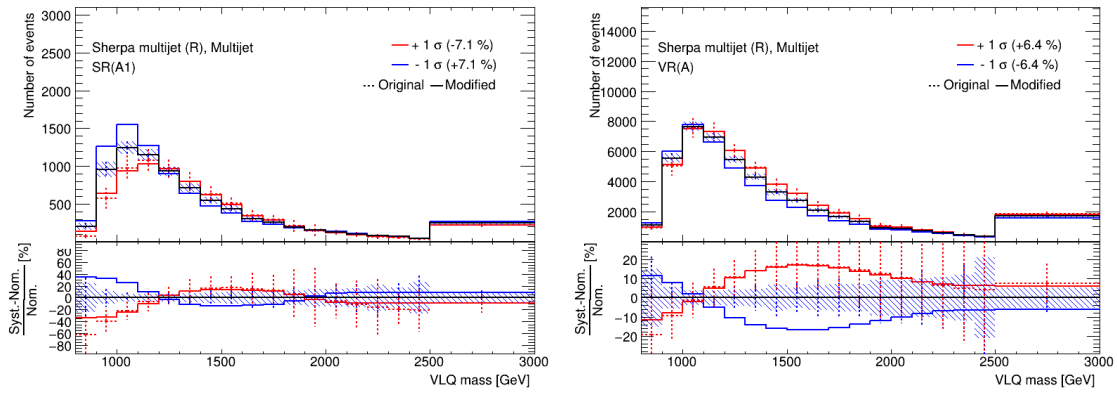


Figure 10.2: Generator uncertainties in the SR and VR SHERPA dijet MC for calculation of R_{corr} using scaled dijet method with third degree polynomial fit.

further, the m_{VLQ} distribution of the multijet ABCD estimate plus the SM backgrounds is fitted against the data in the VR against the background-only hypothesis to observe if this disagreement persists after the fit. As depicted in Fig. 10.3(b), in the post-fit m_{VLQ} distribution, a non-closure exists in the shape agreement between the data and fitted backgrounds. To account for any disagreement in the post-fit m_{VLQ} distributions in the SR, this non-closure is accounted for in the SR a s a shape uncertainty. Thus, the final fit takes into account this uncertainty as a single shape-only nuisance parameter. The $+1\sigma$ variation for each bin i is created as follows.

$$(\text{Non closure multijet variation SR})_i = (\text{Non closure multijet SR})_i \cdot \frac{(\text{Data - Post-fit SM non multijet bkg.})_i}{(\text{Post-fit Multijet estimate in VR})_i}.$$

The shape of the final uncertainty is shown in Fig. 10.3.

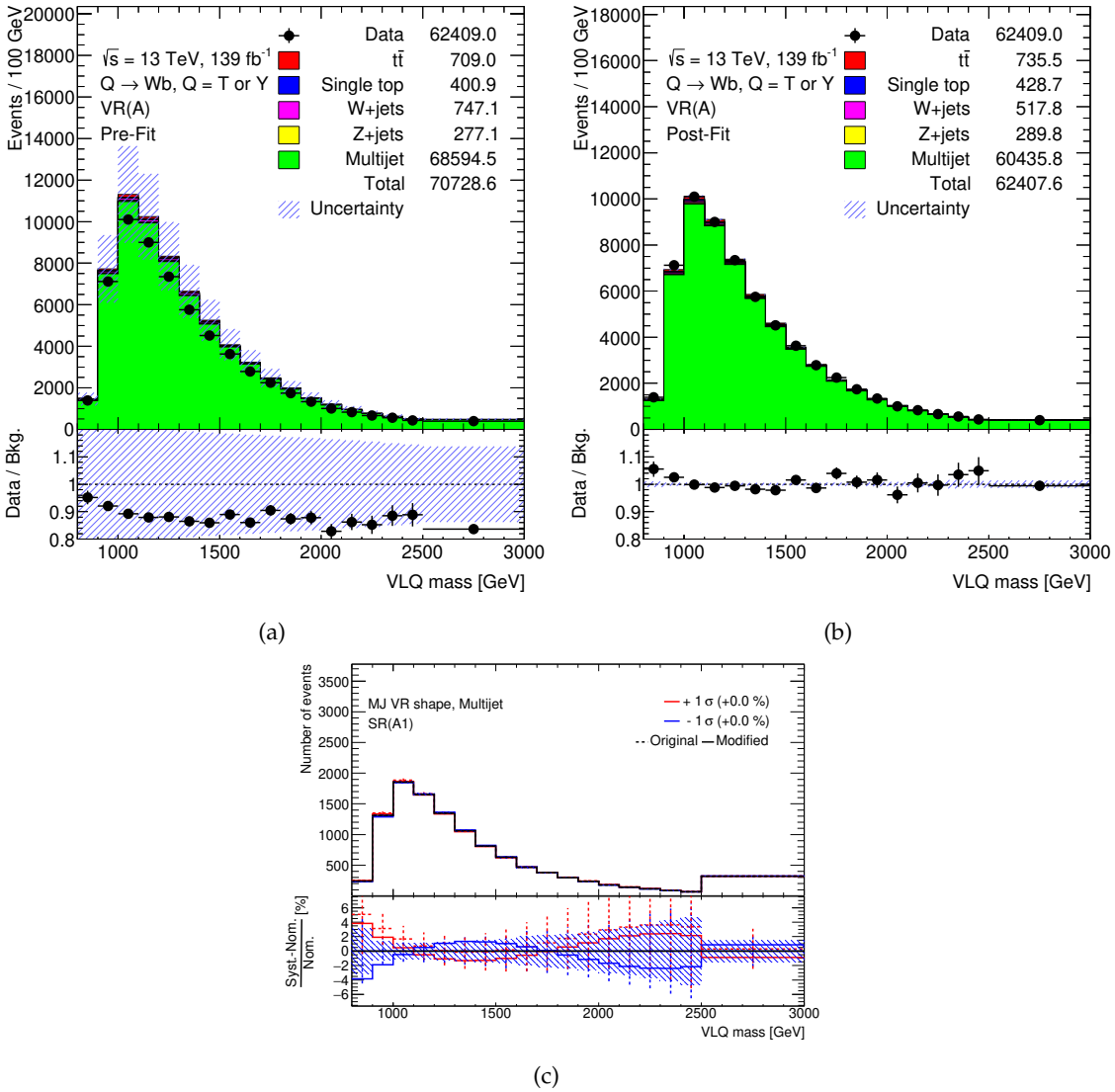


Figure 10.3: Distributions for m_{VLQ} for (a) VR pre-fit (b) VR post-fit after a profile likelihood fit to data against the background-only hypothesis (c) SR shape of the non-closure uncertainty.

CHAPTER 11

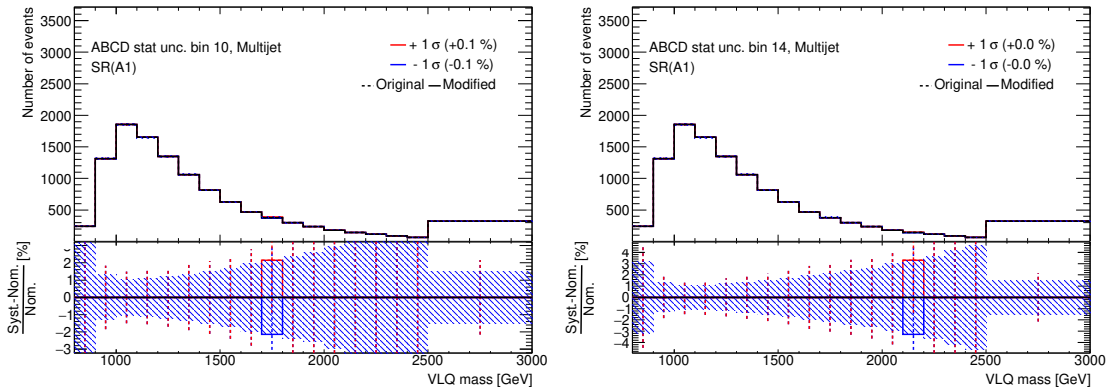
Statistical analysis

In this chapter, the statistical analysis done with the events selected in Chapter 9 to search for T/Y events will be elaborated. Statistical methods such as hypothesis testing, limit setting that are used in this chapter are detailed in Chapter 5. In this analysis, the reconstructed m_{VLQ} is used as a discriminating variable to distinguish between signal and background events. The main irreducible background for this search, the multijet events, are estimated using the control regions (CR). For the statistical analysis, the estimated multijet background is used in as an input. Thus, all the fits in this statistical analysis are profile likelihood fits performed solely in the signal region (SR).

11.1 Profile likelihood fits

The profile likelihoods are built with the binned m_{VLQ} distribution SR as illustrated in Fig. 9.10. All systematic and statistical uncertainties described in Chapter 10 are incorporated as nuisance parameters (NP) in accordance with the methods elaborated in Section 5.2 and Section 5.5.

Akin to the single-lepton analysis, this analysis also employs symmetrisation, pruning, and smoothing. For pruning, a 1% shape and normalisation threshold is applied i.e. any nuisance parameter where the shape or normalisation variation for all bins is less than 1% is not included in the profile likelihood. The smoothing and symmetrisation is done similar to the single-lepton analysis done in this thesis. The Monte Carlo statistical uncertainties for all non-multijet Standard Model backgrounds are added per-bin using the Barlow-Beeston lite method. Since the multijet background is estimated using a data-driven method, rather than using MC statistical uncertainties, the a per-bin statistical uncertainty is assigned. This statistical uncertainty arises from the Gaussian propagation of all the uncertainties coming from the arithmetical operations of the ABCD method. One nuisance parameter is used for each bin of the m_{VLQ} distribution in the SR. The statistical uncertainty assigned for the ABCD estimate in the tenth and fourteenth bin in the SR is illustrated in Fig. 11.1.



(a) (b)

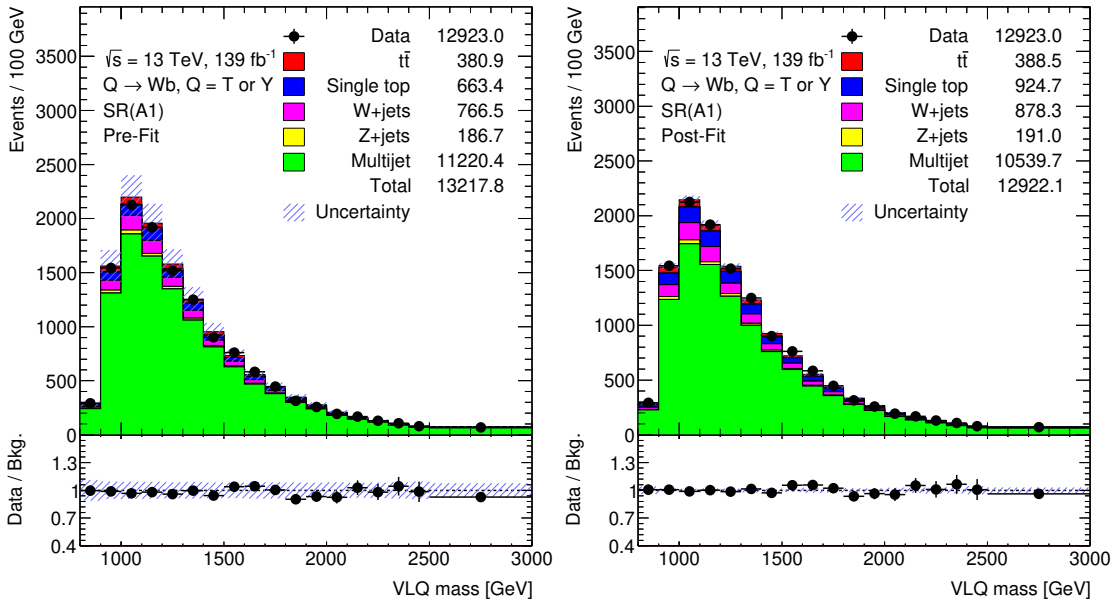
Figure 11.1: The per-bin statistical uncertainty assigned for the ABCD background as per-bin NPs show for the (a) 10th m_{VLQ} bin in the SR and (b) the 14th m_{VLQ} bin in the SR.

Background-only hypothesis testing

The background-only hypothesis is tested to check for any disagreement between the data and the expected SM background yields. The m_{VLQ} distributions of the individual backgrounds and the data in the SR is used i this fit. As $\mu_{\mathrm{signal}} = 0$ is fixed in this fit, the normalisation of the multijets, μ_j^{bkg} (See Eq. (5.2)), where $j \in$ multijets is the parameter of interest (POI) in this profile likelihood fit. A very good post-fit agreement is observed. The $\mu_j^{\mathrm{multijets}}$ is measured as 0.94 ± 0.08 in the fit. The pre and post-fit m_{VLQ} distributions are shown in Fig. 11.2. The postfit distributions for other kinematic variables are depicted in Appendix B. The good agreement between data and expected background yields in Fig. 11.2(b) implies that the data agrees with the SM (background)-only hypothesis within uncertainties. The individual post-fit background yields compared with data are shown in Table 11.1. The pull plots for the nuisance parameters for the individual components is shown in Fig. 11.3, Fig. 11.4, Fig. 11.5, and Fig. 11.6. No significant pulls are observed. The Sherpa R_{corr} factor nuisance parameter is the only NP constrained by data. Thus, no presence of any possible excess is observed in the search for T/Y VLQs in the SR defined for this analysis.

11.2 Limit setting

Since there is no presence of any significant excess in the fitted m_{VLQ} distributions, upper limits have been computed at 95% confidence level on the signal strengths of T and Y VLQ signals using the CLs method as described in Section 5.4. The fits used here are profile likelihood fits. Here, the expected and observed limits have been calculated for various coupling (κ) values for masses in the range 1 to 2.7 TeV in 100 GeV steps. For the expected limits, a unique Asimov dataset is built for each given (κ, m) pair by using $\mu = 1$ in Eq. (5.21).



(a) (b)

Figure 11.2: The pre and post-fit distributions for m_{VLQ} in the SR before and after the fit to data for the background-only hypothesis. All the uncertainties are included in this fit as nuisance parameters. A good agreement between data and expected background yields in (b) implies that the data agrees with the SM (background)-only hypothesis. The lower panel depicts the ratio of data to the fitted SM background yields. The band represents the total (statistical and systematic) uncertainty after the maximum log-likelihood fit.

Table 11.1: Post-fit background yields after the fit to data for the background-only hypothesis in the SR. The uncertainties represent the addition of all individual uncertainty components using Gaussian propagation.

	SR(A1)
$t\bar{t}$	392 ± 49
Single top	992 ± 786
W+jets	949 ± 612
Z+jets	193 ± 33
Multijet	$10\,395 \pm 796$
Total background	$12\,922 \pm 311$
Data	12 923

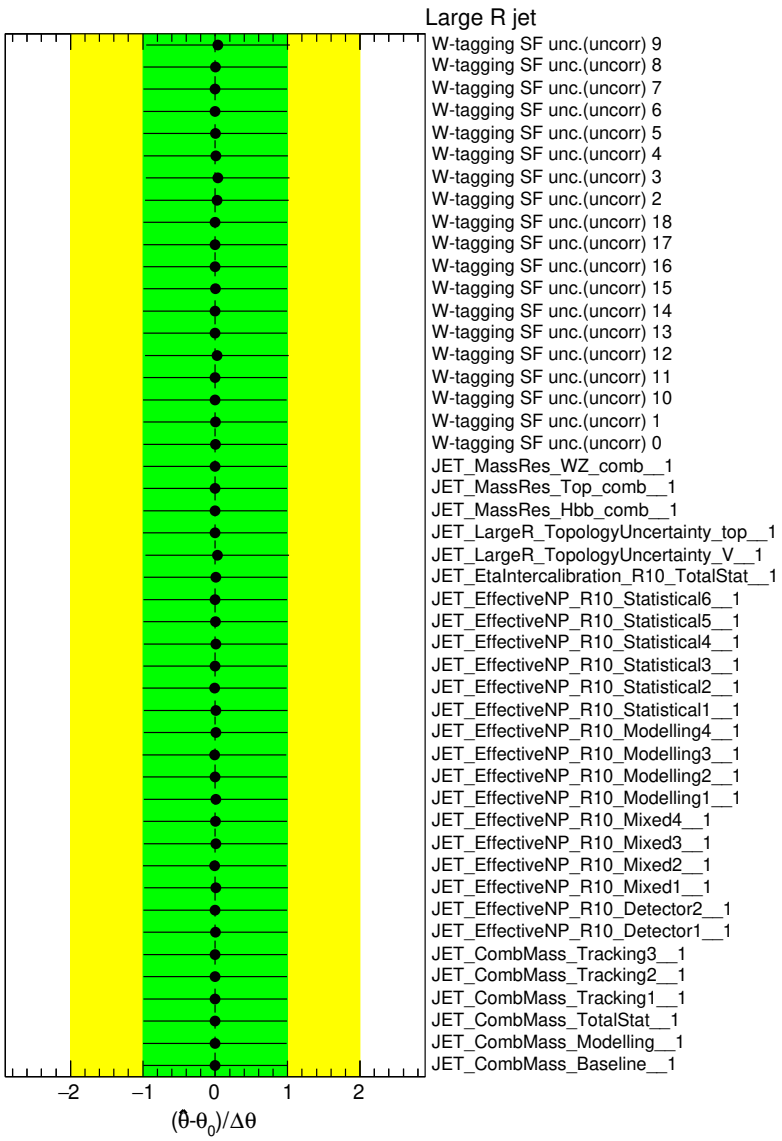


Figure 11.3: Pull plot of the nuisance parameters arising from large-*R* jets that were included in the fit after pruning using the 1% threshold.

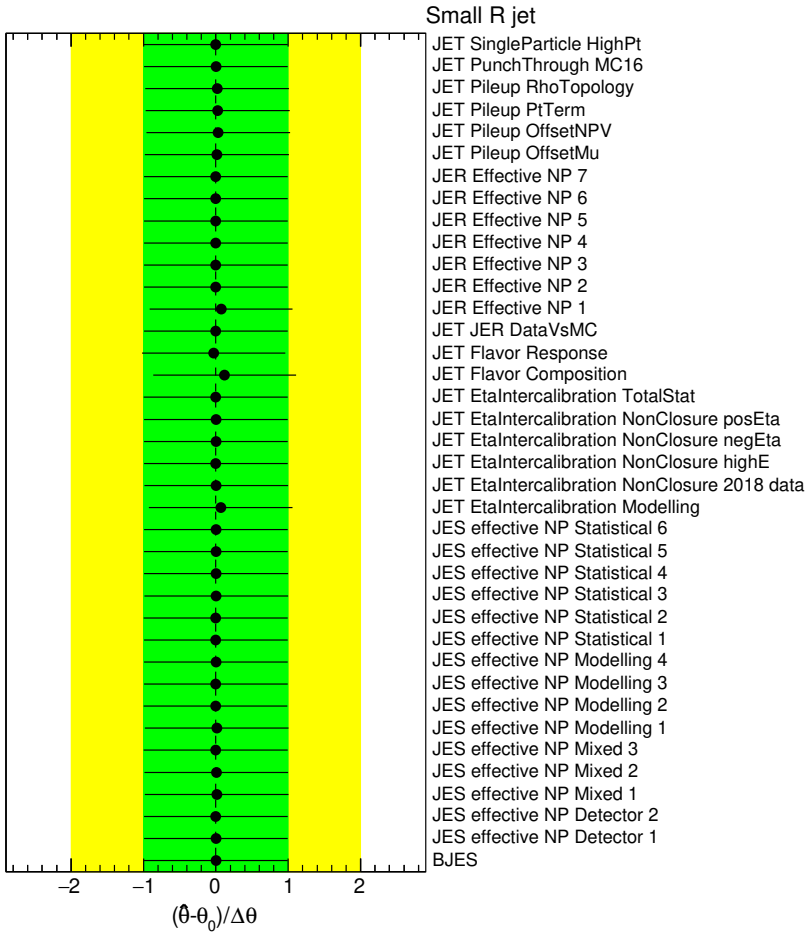


Figure 11.4: Pull plot of the nuisance parameters arising from small-R jets that were included in the fit after pruning using the 1% threshold.

The theoretical cross-sections corresponding to the $\mu = 1 \forall \kappa, m$ are the 4-flavour NLO cross-sections predicted in Ref. [234]. The m_{VLQ} distribution of a 1.5 TeV T-singlet at $\kappa = 0.5$ scaled to the theoretical cross-section in the SR is shown in Fig. 11.7

The signal MC samples created for this search are the T singlet samples at various (κ, m) values. If the interference effects are neglected, the T quark and Y quark differ by the branching fraction. In case of $Y \to Wb$ the branching fraction is 100% and it is approximately 50% for $T \to Wb$. As shown in Fig. 2.10, the interference plays a role only in the low mass range, which is suppressed by the stringent p_{T} cuts on the W-tagged jet and the leading small-R jet of this analysis. Thus, the 95% upper limits were calculated for the T signals and the for the Y case they were calculated by scaling the T limits by the corresponding branching fraction. This is possible because without the inclusion of the negligible interference effects the m_{VLQ} distributions are identical. The mass limits for the T singlet case corresponding to $\kappa = 0.5$ and $\kappa = 0.7$ are shown in Fig. 11.8. Since the limits are coupling dependent, they

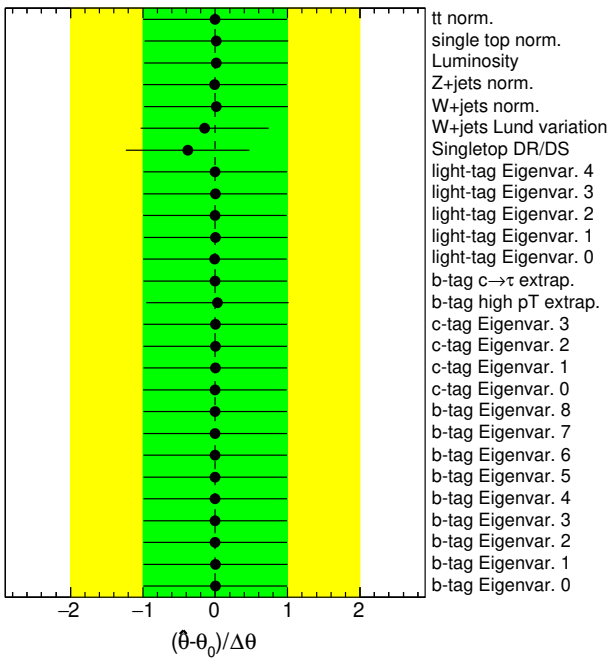


Figure 11.5: Pull plot of the nuisance parameters arising from flavour-tagging and other theoretical uncertainties that were included in the fit after pruning using the 1% threshold.

are computed separately for both the couplings and are shown here. Both the expected and observed mass limits for T and Y quarks at $\kappa = 0.5$ is 1.45 TeV and 1.4 TeVwhereas for $\kappa = 0.7$ they are 1.85 TeV and 1.45 TeV. The 95% CL oberved cross-section times branching ratio limit calculated at the mass limit is 0.1 pb for the T quark whereas it is 0.2 pb for the Y quark.

Since we observe a slight excess in the observed limit compared to expected limit around the 1.6 TeV mass, a fit to data is performed for the signal-plus-background hypothesis. The signal used for the fit is a 1.6 TeV T VLQ with a coupling of 0.5 ($\kappa = 0.5$). The observed significance for this fit is 1.69σ.

Subsequently, for the T-singlet scenario, the 1-dimensional mass limits are reinterpreted to derive the 2-dimensional limits on the mass versus coupling plane, These limits are shown in Fig. 11.9. The exclusion contours are calculated for various cases of the ratio of the width of the resonance (Γ) and the mass of the VLQ resonance. In Fig. 11.9, the coupling (κ) values above the Γ/M countours are excluded for each mass point. Thus, for a 1.4 TeV T-singlet, all $\kappa > 0.68$ are excluded when Γ/M is 0.3. Similarly, for a 1.2 TeV T-singlet, all $\kappa > 0.45$ are excluded when Γ/M is 0.1. The crossing of the observed limits above the expected limits is a consequence of the slight excess of the observed limits in Fig. 11.8. This re-interpretation of 1-dimensional limits to 2-dimensional limits was done following the prescription in Ref. [244]. This prescription works in the narrow-width scenario for the resonance i.e. $\Gamma/M < 0.5$

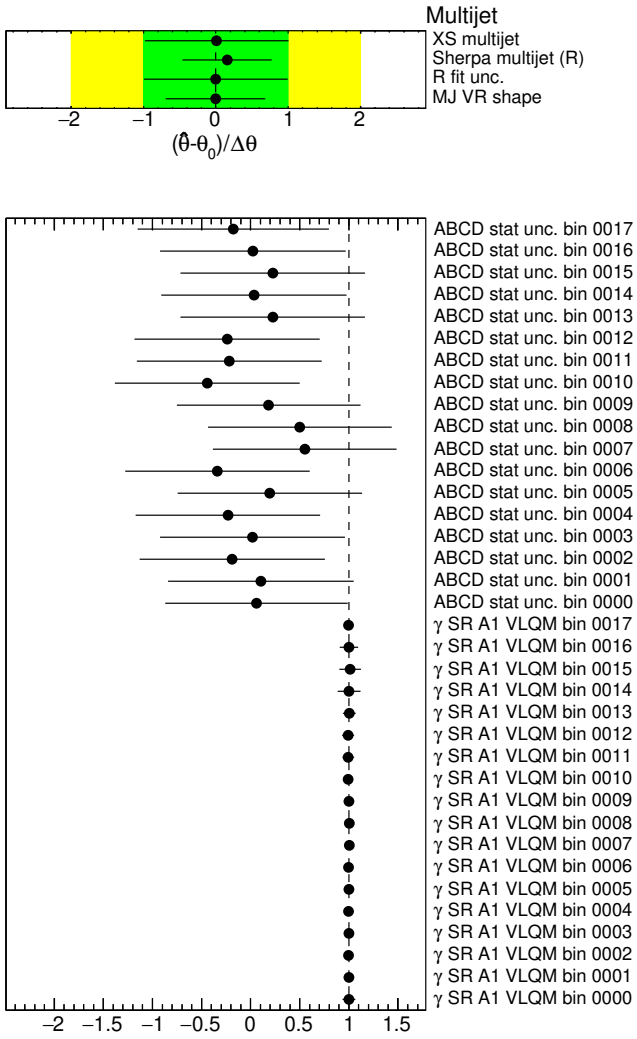


Figure 11.6: Pull plot of the nuisance parameters assigned to the estimated multijet background and other statistical uncertainties that were included in the fit after pruning using the 1% threshold. This also includes the per-bin MC statistical uncertainty.

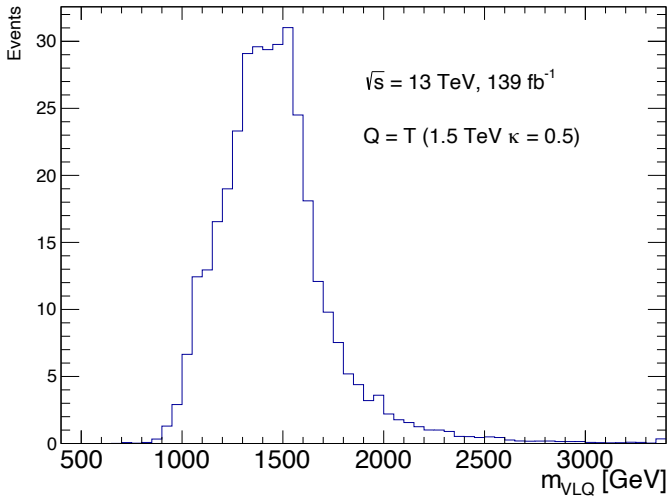
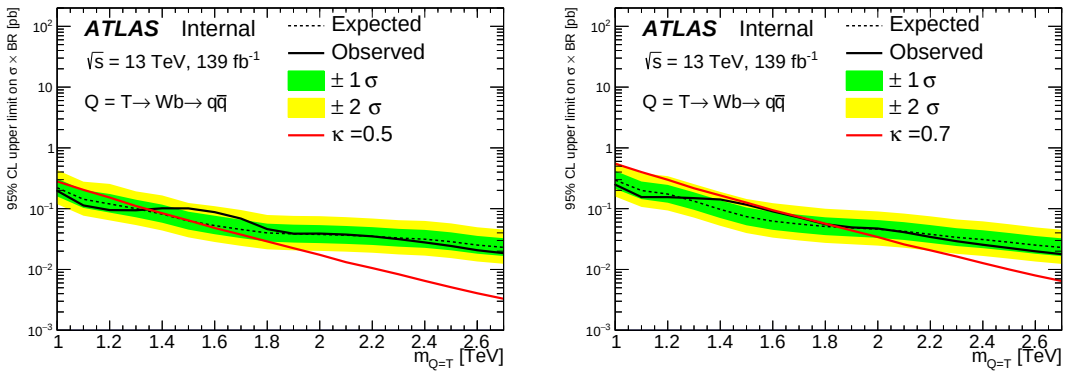


Figure 11.7: The m_{VLQ} distribution of a 1.5 TeV T-singlet at $\kappa = 0.5$ in the SR.



(a) (b)

Figure 11.8: Expected (dotted) and observed (dashed) cross-section limits times branching fraction for T VLQ as a function of m_{VLQ} .Limit are computed for signals with couplings (a) $\kappa = 0.5$ and (b) $\kappa = 0.7$. the theoretical prediction, the branching ratio BR$(T \to Wb)$ is set to $\frac{1}{2}$. The surrounding bands correspond to ± 1 and ± 2 standard deviations around the expected limit.

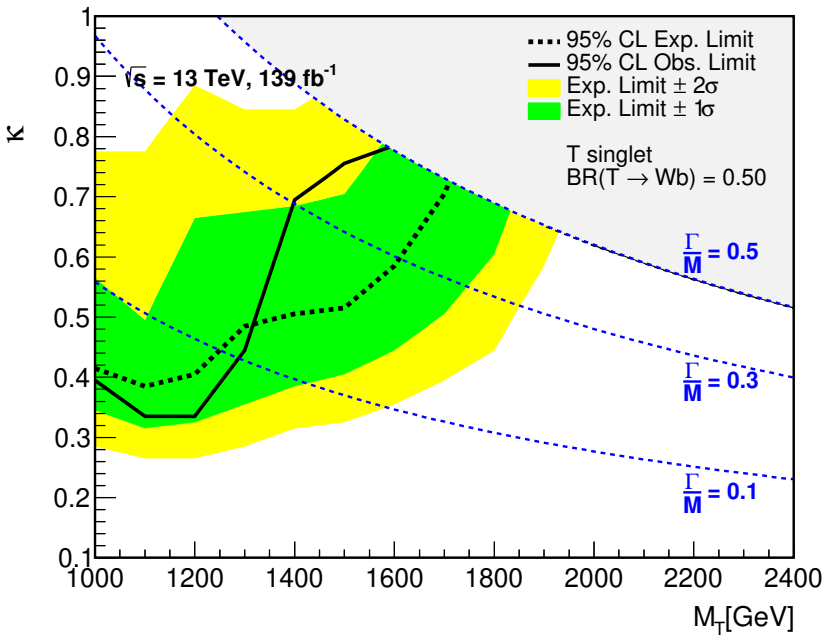


Figure 11.9: Observed (dashed) and expected (dotted) 95% CL exclusion limits on the coupling constant κ as a function of the T VLQ mass in the singlet scenario. All κ values above the black contour lines are excluded at each mass point for the corresponding Γ/M scenario. The shaded bands correspond to ± 1 and ± 2 standard deviations around the expected limit. Limits are only presented in the regime $\Gamma/M \leq 50\%$, where the theory calculations are valid.

Ranking plot

The pre-fit (post-fit) impact of each nuisance parameter on the signal strength μ, $\Delta\mu$, is calculated by comparing the nominal best-fit value of μ with the result of the fit while fixing the considered nuisance parameter to its best-fit value, $\hat{\theta}$, moved by its pre-fit (post-fit) uncertainties, $\pm\Delta\theta(\pm\Delta\hat{\theta})$. The fit used for measuring this impact is a fit to data for the signal-plus-background using the 1.6 TeV T VLQ with a coupling of 0.5 ($\kappa = 0.5$). The ranking of the NPs is shown in Fig. 11.10. It can be observed that the non-closure uncertainty and the generator uncertainty on the multijet by using Sherpa instead of Pythia in the ABCD method have the highest impacts. The data significantly constrains the Sherpa uncertainty. An anti-correlation in the impact is observed which is significantly reduced in the post-fit case. For the fit to data under the signal plus background hypothesis used to create this ranking plot, an anti-correlation of 18% is observed between the signal strength μ_{signal} and μ_{multijet} which is denoted as k_{MJ} in Fig. 11.10.

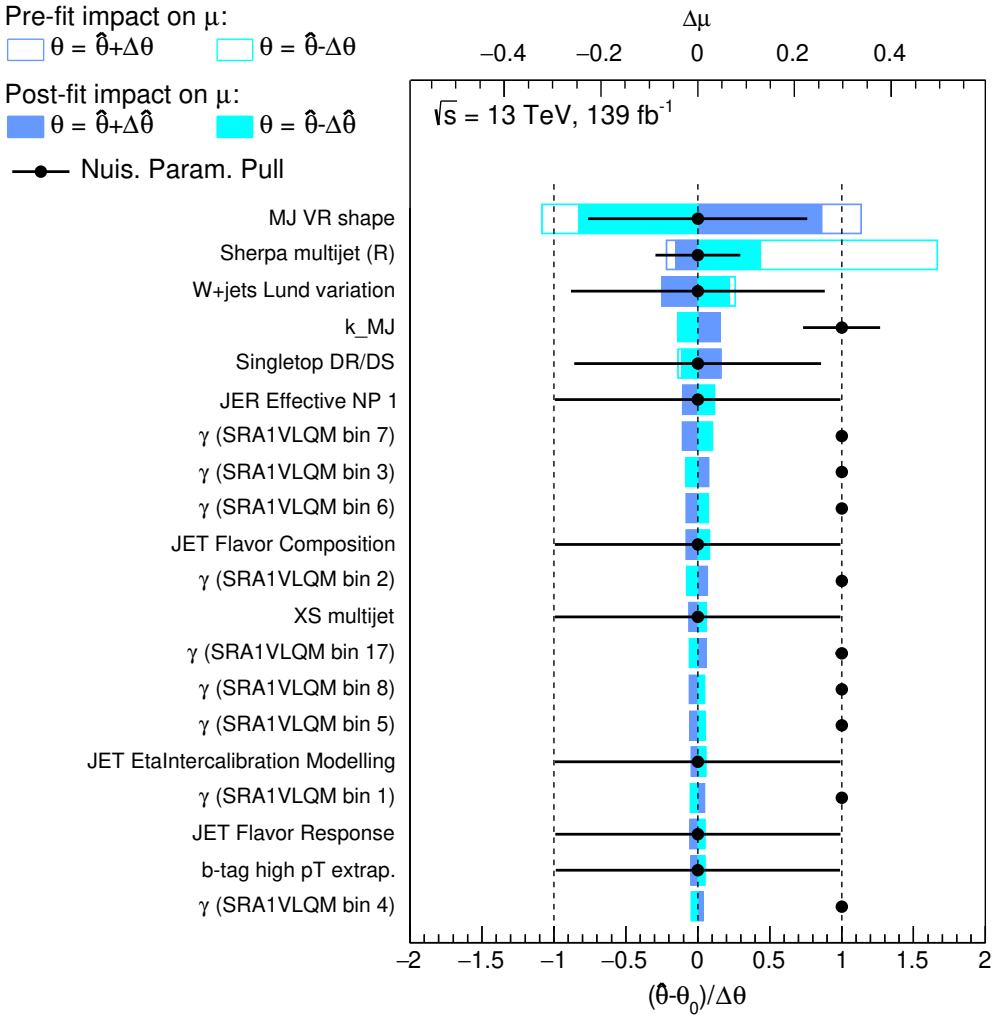


Figure 11.10: Ranking of the nuisance parameters included in the fit according to their impact on the the measured signal strength, μ, for the right-handed Y signal of mass 1.6 TeV with coupling of $\kappa = 0.5$ Only the top 20 nuisance parameters are shown.The empty rectangles denote the pre-fit impact on μ and the filled ones to the post-fit impact on μ, both corresponding to the upper scale. The pulls of the nuisance parameters relative to their nominal values, θ_0. These pulls and their relative post-fit errors, correspond to the lower scale

CHAPTER 12

Epilogue

In this thesis, a search for the single production of T/Y VLQs decaying to a Wb final state was performed using pp collision data collected by the ATLAS detector during Run 2 of the LHC. The thesis covers two analyses: one where the W-boson decays leptonically into a lepton and a neutrino, and the other where the W-boson decays hadronically into a light-flavoured quark-anti-quark pair. The single-lepton analysis uses 36.1 fb^{-1} of pp collision data collected during 2015-16, whereas the hadronic analysis uses 139 fb^{-1} of data collected from 2015 to 2018. In both analyses, no significant excess over the Standard Model background was observed in the data. Therefore, limits were set on the signal strength predicted by the corresponding theoretical models [12, 234]. These models probed in this thesis have model-dependent coupling parameters ($c^{Wb}_{\mathrm{L/R}}, \kappa$, and $\kappa = c^{Wb}_{\mathrm{L/R}}/\sqrt{2}$), and the cross-section limits were re-interpreted to give coupling limits as a function of mass. According to the models, the VLQs can exist in several multiplets, and the coupling strength can vary accordingly. A precise summary of the two analyses performed in this thesis is provided. In the single-lepton analysis, the statistical analysis was the main focus, while for the hadronic analysis, the entire analysis workflow was conducted for this thesis.

Single-lepton analysis

The masses of the VLQs searched in this analysis are greater than 800 GeV, decaying to a Wb with $W \to \ell + \nu$, where $\ell = e/\mu$. To search for such topologies originating from a heavy VLQ, a high p_{T} ($p_{\mathrm{T}} > 350$ GeV in the signal region) b-jet with high angular separation from the lepton is used. The VLQ production typology also includes a light-flavored forward jet. The main backgrounds in this analysis are W+jets and $t\bar{t}$. Two separate control regions are used to model and constrain each of these backgrounds, in addition to the signal-enriched signal region. Subsequently, profile likelihood fits were performed by combining these three regions and using the reconstructed m_{VLQ} as the discriminant to search for the VLQs. No excess was observed, and mass and coupling limits were set. For the T/Y VLQ, interference between signal and backgrounds can exist. This analysis initially overestimated these

interference effects, which were correctly calculated to be negligible in Ref. [110] two years after the publication of this analysis. To account for the overestimated interference effects, a statistical method for calculating coupling limits was developed in this thesis. In hindsight, although these methods may not have been necessary for dealing with the interference effects of this analysis, they can be useful for other new physics searches with significant destructive interference effects between signal and background. Due to these interference effects, the m_{VLQ} distributions changed significantly based on the multiplet representation. For the singlet and triplet cases, the interference was overestimated. In the case of the doublet representation, the interference effects were negligible, and thus, mass and coupling limits were set for the Y RH representation in the (B, Y) doublet. As shown in Fig. 8.10(c), the limits on the mixing parameter $\sin\theta_{\mathrm{R}}$ for the (B, Y) doublet were comparable to the indirect constraints derived from LEP electroweak measurements in the mass range 900–1 250 GeV. In Fig. 8.12, it can be observed that for $c_{\mathrm{R}}^{Wb} = 1/\sqrt{2}$,masses below 1.6 TeV are excluded, and for $c_{\mathrm{R}}^{Wb} = 0.32$, masses below 1.21 TeV are excluded. The figure also shows that the limits with and without interference are comparable.

Hadronic analysis

The search for $T/Y \to Wb \to q\bar{q}$ using the 2015-18 pp collision data is the first analysis of its kind in this channel. Given the limits set on VLQ masses in analyses using the 2015-16 data are in the TeV range, the signals probed in this analysis have masses greater than 1 TeV. A particle with such a mass, when decaying to $Wb \to q\bar{q}$, has a signature with a Lorentz-boosted W-boson decay. So, the expected decay products of the W-boson are clustered into a large-R jet. The reconstruction of the W-jet is performed using a 3-variable W-tagger. The p_{T} of the leading large-R jet is chosen to be greater than 500 GeV to achieve maximum trigger efficiency. The p_{T} cut on the leading small-R jet, which is redefined using an overlap removal procedure, is 350 GeV to perform background suppression. Similar to the single-lepton analysis, a forward jet is also selected here. The overlap removal procedure ensures high angular separation between the leading large-R and small-R jets. As with other hadronic analyses, the main irreducible background is the multijet background. Due to significant mismodelling in the MC simulation, the background is simulated using data-driven ABCD methods. In this analysis, the W-tagging working point and b-jet multiplicity are used as orthogonal variables to construct signal, control, and validation regions, which are subsequently used to estimate the multijet background in the signal and validation regions. For this thesis, the ABCD method was improved to accurately calculate the multijet background. Finally, a profile likelihood fit is performed in the signal region to search for the existence of T/Y quarks. No significant excess was observed in data (see Fig. 11.2(b)), and thus mass and mass-coupling limits were set. The 95% upper limits for the T-singlet signals were calculated, and for the Y case, they were calculated by scaling the T limits by the corresponding branching fraction. This is possible because, without the inclusion of interference, there is no difference in the m_{VLQ} distributions that are utilised to set these limits. The mass limits for the T singlet case corresponding to $\kappa = 0.5$ and $\kappa = 0.7$ are shown

in Fig. 11.8. Two-dimensional mass-dependent coupling limits are shown in Fig. 11.9 for various Γ/M scenarios.

The focus in this thesis was on two orthogonal decay modes of the Wb final state to search for single-production of T/Y vector-like quarks. Thus, the search scanned the entire available phase space for the $T/Y \to Wb$ decay mode during Run 2 of the LHC using ATLAS data. It is worth noting that the single-lepton analysis for the full Run 2 dataset is also being explored by the ATLAS collaboration independently. The statistical methods developed to handle interference effects can be used in further analyses if the effects are significant. The hadronic analysis was the first of its kind while studying $T/Y \to Wb$. This analysis used advanced methods that use jet substructure. The methods developed here to improve the ABCD method can easily be applied to similar hadronic searches. The complexity of reconstructing jet structure has been one of the reasons behind the difficulties of studying hadronic final states. Evidently, the methods used in this thesis demonstrate the potential of using advanced jet-taggers to explore statistically rich hadronic final states for future searches in the upcoming runs of the LHC.

Bibliography

[1] Planck Collaboration,
Planck 2018 results. I. Overview and the cosmological legacy of Planck,
Astron. Astrophys. **641** (2020) A1, arXiv: 1807.06205 [astro-ph.CO]
(cit. on p. 1).

[2] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys.Lett. **B716** (2012) 1,
arXiv: 1207.7214 [hep-ex] (cit. on pp. 1, 16).

[3] CMS Collaboration,
Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC,
Phys.Lett. **B716** (2012) 30, arXiv: 1207.7235 [hep-ex] (cit. on pp. 1, 16).

[4] ATLAS Collaboration, *Measurement of the Higgs boson mass with $H \to \gamma\gamma$ decays in 140 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, (2023),
arXiv: 2308.07216 [hep-ex] (cit. on pp. 1, 16).

[5] S. P. Martin, *A Supersymmetry primer*,
Adv. Ser. Direct. High Energy Phys. **18** (1998) 1, ed. by G. L. Kane,
arXiv: hep-ph/9709356 (cit. on pp. 2, 19).

[6] Particle Data Group, *Review of Particle Physics*,
Progress of Theoretical and Experimental Physics **2020** (2020), ISSN: 2050-3911
(cit. on pp. 2, 6, 11, 17, 19, 20, 57, 58).

[7] D. B. Kaplan, H. Georgi and S. Dimopoulos, *Composite Higgs Scalars*,
Phys. Lett. B **136** (1984) 187 (cit. on pp. 2, 19).

[8] N. Vignaroli, *Discovering the composite Higgs through the decay of a heavy fermion*,
JHEP **07** (2012) 158, arXiv: 1204.0468 [hep-ph] (cit. on pp. 2, 19).

[9] N. Arkani-Hamed, A. G. Cohen and H. Georgi,
Electroweak symmetry breaking from dimensional deconstruction,
Phys. Lett. B **513** (2001) 232, arXiv: hep-ph/0105239 (cit. on pp. 2, 20).

[10] N. Arkani-Hamed et al., *The Littlest Higgs*, JHEP **07** (2002) 034,
arXiv: hep-ph/0206021 (cit. on pp. 2, 20).

[11] N. Arkani-Hamed et al.,
Phenomenology of electroweak symmetry breaking from theory space, JHEP **08** (2002) 020,
arXiv: hep-ph/0202089 (cit. on pp. 2, 20).

[12] O. Matsedonskyi, G. Panico and A. Wulzer,
On the Interpretation of Top Partners Searches, JHEP **12** (2014) 097,
arXiv: 1409.0100 [hep-ph] (cit. on pp. 2, 20–22, 76, 91, 98, 147).

[13] J. A. Aguilar-Saavedra et al., *Handbook of vectorlike quarks: Mixing and single production*,
Phys. Rev. D **88** (2013) 094010, arXiv: 1306.0572 [hep-ph]
(cit. on pp. 2, 21–23, 98, 99, 103).

[14] M. Buchkremer et al., *Model Independent Framework for Searches of Top Partners*,
Nucl. Phys. B **876** (2013) 376, arXiv: 1305.4172 [hep-ph] (cit. on pp. 2, 76).

[15] ATLAS Collaboration, *Search for single production of vector-like quarks decaying into Wb in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **05** (2019) 164,
arXiv: 1812.07343 [hep-ex] (cit. on pp. 2, 78, 100).

[16] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*,
Phys. Rev. **76** (6 1949) 769 (cit. on p. 7).

[17] F. Halzen and A. D. Martin, *Quarks and Leptons*, Wiley and Sons, 1984 (cit. on p. 9).

[18] S. Glashow, *Partial Symmetries of Weak Interactions*, Nucl.Phys. **22** (1961) 579
(cit. on p. 10).

[19] A. Salam, *Weak and Electromagnetic Interactions*, Conf.Proc. **C680519** (1968) 367
(cit. on p. 10).

[20] S. Weinberg, *A Model of Leptons*, Phys.Rev.Lett. **19** (1967) 1264 (cit. on p. 10).

[21] R. P. Feynman,
Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction,
Phys. Rev. **80** (3 1950) 440 (cit. on p. 10).

[22] J. Schwinger, *On Quantum-Electrodynamics and the Magnetic Moment of the Electron*,
Phys. Rev. **73** (4 1948) 416 (cit. on p. 10).

[23] S. Tomonaga,
On a relativistically invariant formulation of the quantum theory of wave fields,
Prog. Theor. Phys. **1** (1946) 27 (cit. on p. 10).

[24] E. Fermi, *Tentativo di una teoria dell'emissione dei raggi beta*, Ric. Sci. **4** (1933) 491
(cit. on p. 10).

[25] C. S. Wu et al., *Experimental Test of Parity Conservation in Beta Decay*,
Phys. Rev. **105** (4 1957) 1413 (cit. on p. 10).

[26] R. P. Feynman and M. Gell-Mann, *Theory of the Fermi Interaction*,
Phys. Rev. **109** (1 1958) 193 (cit. on p. 10).

[27] E. Sudarshan and R. Marshak, "The nature of the four-fermion interaction",
Conference on Mesons and Newly discovered Particles, (Padua,Venice, 1957)
(cit. on p. 10).

[28] Gargamelle Neutrino Collaboration, *Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment*, Phys. Lett. B **46** (1973) 138 (cit. on p. 11).

[29] UA1 Collaboration, *Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV*, Phys. Lett. B **122** (1983) 103 (cit. on pp. 11, 17).

[30] UA2 Collaboration, *Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider*, Phys. Lett. B **122** (1983) 476 (cit. on pp. 11, 17).

[31] UA2 Collaboration, *Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider*, Phys. Lett. B **129** (1983) 130 (cit. on p. 11).

[32] UA1 Collaboration, *Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider*, Phys. Lett. B **126** (1983) 398 (cit. on p. 11).

[33] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531 (cit. on p. 11).

[34] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652 (cit. on p. 11).

[35] M. Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, (1961) (cit. on p. 12).

[36] Y. Ne'eman, *Derivation of strong interactions from a gauge invariance*, Nucl. Phys. **26** (1961) 222, ed. by R. Ruffini and Y. Verbin (cit. on p. 12).

[37] V. E. Barnes et al., *Observation of a Hyperon with Strangeness Minus Three*, Phys. Rev. Lett. **12** (8 1964) 204 (cit. on p. 12).

[38] M. Gell-Mann, *A Schematic Model of Baryons and Mesons*, Phys. Lett. **8** (1964) 214 (cit. on p. 12).

[39] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*, tech. rep. CERN-TH-401, CERN, 1964, URL: http://cds.cern.ch/record/352337 (cit. on p. 12).

[40] J. Piffaretti et al., *Pi+ and pi- Scattering from C-12 at 150-MeV*, Phys. Lett. B **67** (1977) 289 (cit. on p. 12).

[41] M. Han and Y. Nambu, *Three Triplet Model with Double SU(3) Symmetry*, Phys. Rev. **139** (1965) B1006, ed. by T. Eguchi (cit. on p. 12).

[42] O. Greenberg, *Spin and Unitary Spin Independence in a Paraquark Model of Baryons and Mesons*, Phys. Rev. Lett. **13** (1964) 598 (cit. on p. 12).

[43] E. D. Bloom et al., *High-Energy Inelastic e p Scattering at 6-Degrees and 10-Degrees*, Phys. Rev. Lett. **23** (1969) 930 (cit. on p. 12).

[44] M. Breidenbach et al., *Observed behavior of highly inelastic electron-proton scattering*, Phys. Rev. Lett. **23** (1969) 935 (cit. on p. 12).

[45] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the Color Octet Gluon Picture*, Phys. Lett. B **47** (1973) 365 (cit. on p. 12).

[46] TASSO Collaboration, *Evidence for Planar Events in e+ e- Annihilation at High-Energies*, Phys. Lett. B **86** (1979) 243 (cit. on p. 12).

[47] PLUTO Collaboration, *Evidence for Gluon Bremsstrahlung in e+ e- Annihilations at High-Energies*, Phys. Lett. B **86** (1979) 418 (cit. on p. 12).

[48] D. P. Barber et al., *Discovery of Three Jet Events and a Test of Quantum Chromodynamics at PETRA Energies*, Phys. Rev. Lett. **43** (1979) 830 (cit. on p. 12).

[49] JADE Collaboration, *Observation of Planar Three Jet Events in e+ e- Annihilation and Evidence for Gluon Bremsstrahlung*, Phys. Lett. B **91** (1980) 142 (cit. on p. 12).

[50] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Nonabelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) 1343, ed. by J. Taylor (cit. on p. 12).

[51] D. Gross and F. Wilczek, *Asymptotically Free Gauge Theories. 1*, Phys.Rev. **D8** (1973) 3633 (cit. on p. 12).

[52] S. Durr et al., *Ab Initio Determination of Light Hadron Masses*, Science **322** (2008) (cit. on p. 13).

[53] CMS Collaboration, *Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8$ TeV and cross section ratios to 2.76 and 7 TeV*, JHEP **03** (2017) 156, arXiv: 1609.05331 [hep-ex] (cit. on p. 13).

[54] J. C. Collins, D. E. Soper and G. F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, arXiv: hep-ph/0409313 (cit. on p. 13).

[55] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244, arXiv: 1207.1303 [hep-ph] (cit. on pp. 14, 76, 113).

[56] J. Gao et al., *CT10 next-to-next-to-leading order global analysis of QCD*, Phys. Rev. D **89** (2014) 033009, arXiv: 1302.6246 [hep-ph] (cit. on p. 14).

[57] Y. L. Dokshitzer, *Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics.*, Sov. Phys. JETP **46** (1977) 641 (cit. on p. 14).

[58] V. Gribov and L. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438 (cit. on p. 14).

[59] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. B **126** (1977) 298 (cit. on p. 14).

[60] S. Höche, "Introduction to parton-shower event generators",
Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders, 2015 235, arXiv: 1411.4085 [hep-ph] (cit. on pp. 14, 15).

[61] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
Phys.Rev.Lett. **13** (1964) 508 (cit. on p. 15).

[62] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
Phys.Rev.Lett. **13** (1964) 321 (cit. on p. 15).

[63] CMS Collaboration,
A measurement of the Higgs boson mass in the diphoton decay channel,
Phys. Lett. B **805** (2020) 135425, arXiv: 2002.06398 [hep-ex] (cit. on p. 16).

[64] UA2 collaboration,
A Precise Determination of the W and Z Masses at the {CERN} $\bar{p}p$ Collider,
Phys. Lett. B **241** (1990) 150 (cit. on p. 17).

[65] UA2 Collaboration,
A Measurement of two jet decays of the W and Z bosons at the CERN $\bar{p}p$ collider,
Z. Phys. C **49** (1991) 17 (cit. on p. 17).

[66] S. Mele, *Physics of W bosons at LEP*, Phys. Rept. **403-404** (2004) 255,
arXiv: hep-ex/0407042 (cit. on p. 17).

[67] ATLAS Collaboration,
Measurement of the W-boson mass in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector,
Eur. Phys. J. C **78** (2018) 110, [Erratum: Eur.Phys.J.C 78, 898 (2018)],
arXiv: 1701.07240 [hep-ex] (cit. on p. 17).

[68] ATLAS Collaboration,
Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector,
Phys. Rev. D **92** (2015) 012006, arXiv: 1412.2641 [hep-ex] (cit. on p. 17).

[69] ATLAS Collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **759** (2016) 601,
arXiv: 1603.09222 [hep-ex] (cit. on p. 17).

[70] A. Giammanco and R. Schwienhorst,
Single top-quark production at the Tevatron and the LHC,
Rev. Mod. Phys. **90** (3 2018) 035001 (cit. on p. 18).

[71] G. 't Hooft et al.,
"Naturalness, Chiral Symmetry, and Spontaneous Chiral Symmetry Breaking",
Recent Developments in Gauge Theories. Proceedings, Nato Advanced Study Institute, Cargese, France, August 26 - September 8, 1979, ed. by G. 't Hooft, vol. 59, 1980 p. 135
(cit. on p. 18).

[72] M. Veltman, *The Infrared - Ultraviolet Connection*, Acta Phys.Polon. **B12** (1981) 437
(cit. on p. 18).

[73] Planck Collaboration, *Planck 2015 results. XIII. Cosmological parameters*, Astron. Astrophys. **594** (2016) A13, arXiv: 1502.01589 [astro-ph.CO] (cit. on p. 18).

[74] ATLAS Collaboration, *SUSY Summary Plots March 2022*, tech. rep., CERN, 2022, URL: http://cds.cern.ch/record/2805985 (cit. on p. 19).

[75] S. P. Martin, *Extra vector-like matter and the lightest Higgs scalar boson mass in low-energy supersymmetry*, Phys. Rev. D **81** (2010) 035004, arXiv: 0910.2732 [hep-ph] (cit. on p. 20).

[76] O. Eberhardt et al., *Impact of a Higgs boson at a mass of 126 GeV on the standard model with three and four fermion generations*, Phys. Rev. Lett. **109** (2012) 241802, arXiv: 1209.1101 [hep-ph] (cit. on p. 20).

[77] M. E. Peskin and T. Takeuchi, *A New constraint on a strongly interacting Higgs sector*, Phys. Rev. Lett. **65** (1990) 964 (cit. on p. 22).

[78] ATLAS Collaboration, *Search for Pair Production of a New b' Quark that Decays into a Z Boson and a Bottom Quark with the ATLAS Detector*, Phys. Rev. Lett. **109** (2012) 071801, arXiv: 1204.1265 [hep-ex] (cit. on p. 22).

[79] ATLAS Collaboration, *Search for production of vector-like quark pairs and of four top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, JHEP **08** (2015) 105, arXiv: 1505.04306 [hep-ex] (cit. on p. 22).

[80] ATLAS Collaboration, *Search for pair production of a new heavy quark that decays into a W boson and a light quark in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, Phys. Rev. D **92** (2015) 112007, arXiv: 1509.04261 [hep-ex] (cit. on p. 22).

[81] ATLAS Collaboration, *Search for pair production of up-type vector-like quarks and for four-top-quark events in final states with multiple b-jets with the ATLAS detector*, JHEP **07** (2018) 089, arXiv: 1803.09678 [hep-ex] (cit. on p. 22).

[82] ATLAS Collaboration, *Search for pair production of heavy vector-like quarks decaying to high-p_T W bosons and b quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **10** (2017) 141, arXiv: 1707.03347 [hep-ex] (cit. on p. 22).

[83] ATLAS Collaboration, *Search for pair production of vector-like top quarks in events with one lepton, jets, and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector*, JHEP **08** (2017) 052, arXiv: 1705.10751 [hep-ex] (cit. on p. 22).

[84] ATLAS Collaboration, *Search for pair- and single-production of vector-like quarks in final states with at least one Z boson decaying into a pair of electrons or muons in pp collision data collected with the ATLAS detector*, Phys. Rev. D **98** (2018) 112010, arXiv: 1806.10555 [hep-ex] (cit. on p. 22).

[85] ATLAS Collaboration, *Search for pair production of heavy vector-like quarks decaying into hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. D **98** (2018) 092005, arXiv: 1808.01771 [hep-ex] (cit. on p. 22).

[86] ATLAS Collaboration, *Search for pair production of heavy vector-like quarks decaying into high-p_T W bosons and top quarks in the lepton-plus-jets final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08** (2018) 048, arXiv: 1806.01762 [hep-ex] (cit. on p. 22).

[87] ATLAS Collaboration, *Combination of the searches for pair-produced vector-like partners of the third-generation quarks at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Rev. Lett. **121** (2018) 211801, arXiv: 1808.02343 [hep-ex] (cit. on p. 22).

[88] CMS Collaboration, *Search for a Vectorlike Quark with Charge 2/3 in t + Z Events from pp Collisions at $\sqrt{s} = 7$ TeV*, Phys. Rev. Lett. **107** (2011) 271802, arXiv: 1109.4985 [hep-ex] (cit. on p. 22).

[89] CMS Collaboration, *Search for pair produced fourth-generation up-type quarks in pp collisions at $\sqrt{s} = 7$ TeV with a lepton in the final state*, Phys. Lett. B **718** (2012) 307, arXiv: 1209.0471 [hep-ex] (cit. on p. 22).

[90] CMS Collaboration, *Search for vectorlike charge 2/3 T quarks in proton–proton collisions at $\sqrt{(s)} = 8$ TeV*, Phys. Rev. D **93** (2016) 012003, arXiv: 1509.04177 [hep-ex] (cit. on p. 22).

[91] CMS Collaboration, *Search for pair-produced vectorlike B quarks in proton–proton collisions at $\sqrt{s} = 8$ TeV*, Phys. Rev. D **93** (2016) 112009, arXiv: 1507.07129 [hep-ex] (cit. on p. 22).

[92] CMS Collaboration, *Inclusive search for a vector-like T quark with charge $\frac{2}{3}$ in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **729** (2014) 149, arXiv: 1311.7667 [hep-ex] (cit. on p. 22).

[93] CMS Collaboration, *Search for vector-like T quarks decaying to top quarks and Higgs bosons in the all-hadronic channel using jet substructure*, JHEP **06** (2015) 080, arXiv: 1503.01952 [hep-ex] (cit. on p. 22).

[94] CMS Collaboration, *Search for pair production of vector-like T and B quarks in single-lepton final states using boosted jet substructure techniques at $\sqrt{s} = 13$ TeV*, JHEP **11** (2017) 085, arXiv: 1706.03408 [hep-ex] (cit. on p. 22).

[95] CMS Collaboration, *Search for pair production of vector-like quarks in the $bW\bar{b}W$ channel from proton–proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **779** (2018) 82, arXiv: 1710.01539 [hep-ex] (cit. on p. 22).

[96] CMS Collaboration, *Search for vector-like T and B quark pairs in final states with leptons at $\sqrt{s} = 13$ TeV*, JHEP **08** (2018) 177, arXiv: 1805.04758 [hep-ex] (cit. on p. 22).

[97] CMS Collaboration, *Search for vector-like quarks in events with two oppositely charged leptons and jets in proton–proton collisions at* $\sqrt{s} = 13\,TeV$, Eur. Phys. J. C **79** (2019) 364, arXiv: 1812.09768 [hep-ex] (cit. on p. 22).

[98] CMS Collaboration,
Search for pair production of vector-like quarks in the fully hadronic final state,
Phys. Rev. D **100** (2019) 072001, arXiv: 1906.11903 [hep-ex] (cit. on p. 22).

[99] ATLAS Collaboration, *Search for single production of vector-like quarks decaying into* Wb *in* pp *collisions at* $\sqrt{s} = 8\,TeV$ *with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 442, arXiv: 1602.05606 [hep-ex] (cit. on pp. 22, 79).

[100] ATLAS Collaboration, *Search for the production of single vector-like and excited quarks in the* Wt *final state in* pp *collisions at* $\sqrt{s} = 8\,TeV$ *with the ATLAS detector*,
JHEP **02** (2016) 110, arXiv: 1510.02664 [hep-ex] (cit. on p. 22).

[101] ATLAS Collaboration, *Search for single production of a vector-like* T *quark decaying into a Higgs boson and top quark with fully hadronic final states using the ATLAS detector*,
Phys. Rev. D **105** (2022) 092012, arXiv: 2201.07045 [hep-ex] (cit. on p. 22).

[102] CMS Collaboration, *Search for single production of a heavy vector-like* T *quark decaying to a Higgs boson and a top quark with a lepton and jets in the final state*,
Phys. Lett. B **771** (2017) 80, arXiv: 1612.00999 [hep-ex] (cit. on p. 22).

[103] CMS Collaboration, *Search for single production of vector-like quarks decaying to a* Z *boson and a top or a bottom quark in proton–proton collisions at* $\sqrt{s} = 13\,TeV$,
JHEP **05** (2017) 029, arXiv: 1701.07409 [hep-ex] (cit. on p. 22).

[104] CMS Collaboration, *Search for electroweak production of a vector-like quark decaying to a top quark and a Higgs boson using boosted topologies in fully hadronic final states*,
JHEP **04** (2017) 136, arXiv: 1612.05336 [hep-ex] (cit. on p. 22).

[105] CMS Collaboration, *Search for a heavy resonance decaying to a top quark and a vector-like top quark at* $\sqrt{s} = 13\,TeV$, JHEP **09** (2017) 053, arXiv: 1703.06352 [hep-ex] (cit. on p. 22).

[106] CMS Collaboration, *Search for single production of a vector-like* T *quark decaying to a* Z *boson and a top quark in proton–proton collisions at* $\sqrt{s} = 13\,TeV$,
Phys. Lett. B **781** (2018) 574, arXiv: 1708.01062 [hep-ex] (cit. on p. 22).

[107] CMS Collaboration, *Search for single production of vector-like quarks decaying to a top quark and a* W *boson in proton–proton collisions at* $\sqrt{s} = 13\,TeV$,
Eur. Phys. J. C **79** (2019) 90, arXiv: 1809.08597 [hep-ex] (cit. on p. 22).

[108] CMS Collaboration,
Search for electroweak production of a vector-like T *quark using fully hadronic final states*,
JHEP **01** (2020) 036, arXiv: 1909.04721 [hep-ex] (cit. on p. 22).

[109] ATLAS Collaboration, *Search for single production of vector-like quarks decaying into* Wb *in* pp *collisions at* $\sqrt{s} = 13\,TeV$ *with the ATLAS detector*, ATLAS-CONF-2016-072, 2016, URL: https://cds.cern.ch/record/2206226 (cit. on p. 22).

[110] A. Deandrea et al., *Single production of vector-like quarks: the effects of large width, interference and NLO corrections*, JHEP **08** (2021) 107, arXiv: 2105.08745 [hep-ph] (cit. on pp. 23, 24, 76, 94, 148).

[111] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001 (cit. on p. 25).

[112] E. Todesco and J. Wenninger, *Large Hadron Collider momentum calibration and accuracy*, Phys. Rev. Accel. Beams **20** (2017) 081003 (cit. on p. 25).

[113] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on pp. 25, 27).

[114] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 25).

[115] A. A. Alves Jr. et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 25).

[116] K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 25).

[117] G. Anelli et al., *The TOTEM experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08007 (cit. on p. 25).

[118] J. Pinfold, "MoEDAL-MAPP, an LHC Dedicated Detector Search Facility", *2022 Snowmass Summer Study*, 2022, arXiv: 2209.03988 [hep-ph] (cit. on p. 25).

[119] T. Sako, "LHCf: A new experiment to study very forward particle emission at LHC", *29th International Cosmic Ray Conference*, 2005 (cit. on p. 25).

[120] E. Lopienska, *The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022*, (2022), General Photo, URL: https://cds.cern.ch/record/2800984 (cit. on p. 26).

[121] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-2 of the LHC*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Publications_and_Conference_Results (cit. on p. 27).

[122] J. Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 28).

[123] ATLAS Collaboration, *ATLAS Inner Detector: Technical Design Report, Volume 1*, ATLAS-TDR-4; CERN-LHCC-97-016, 1997, URL: https://cds.cern.ch/record/331063 (cit. on p. 28).

[124] ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, Eur. Phys. J. C **70** (2010) 787, arXiv: 1004.5293 [hep-ex] (cit. on p. 28).

[125] K. Potamianos, *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*, tech. rep., 15 pages, EPS-HEP 2015 Proceedings: CERN, 2016, arXiv: 1608.07850, URL: https://cds.cern.ch/record/2209070 (cit. on pp. 28, 29).

[126] ATLAS Collaboration, *ATLAS Pixel Detector: Technical Design Report*, ATLAS-TDR-11; CERN-LHCC-98-013, 1998, URL: https://cds.cern.ch/record/381263 (cit. on p. 29).

[127] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-TDR-19; CERN-LHCC-2010-013, 2010, URL: https://cds.cern.ch/record/1291633 (cit. on p. 29), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009, 2012, URL: https://cds.cern.ch/record/1451888.

[128] Y. Takubo, *ATLAS IBL operational experience*, tech. rep., CERN, 2017, URL: https://cds.cern.ch/record/2235541 (cit. on p. 29).

[129] ATLAS Collaboration, *Operation and performance of the ATLAS semiconductor tracker in LHC Run 2*, JINST **17** (2021) P01013, arXiv: 2109.02591 [physics.ins-det] (cit. on p. 29).

[130] ATLAS Collaboration, *Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties*, JINST **12** (2017) P05002, arXiv: 1702.06473 [hep-ex] (cit. on p. 30).

[131] ATLAS Collaboration, *ATLAS Inner Tracker Strip Detector: Technical Design Report*, ATLAS-TDR-025; CERN-LHCC-2017-005, 2017, URL: https://cds.cern.ch/record/2257755 (cit. on p. 30).

[132] ATLAS Collaboration, *ATLAS Calorimeter Performance: Technical Design Report*, ATLAS-TDR-1; CERN-LHCC-96-040, 1996, URL: https://cds.cern.ch/record/331059 (cit. on p. 30).

[133] ATLAS Collaboration, *ATLAS Liquid Argon Calorimeter: Technical Design Report*, ATLAS-TDR-2; CERN-LHCC-96-041, 1996, URL: https://cds.cern.ch/record/331061 (cit. on p. 30).

[134] ATLAS Collaboration, *ATLAS Tile Calorimeter: Technical Design Report*, ATLAS-TDR-3; CERN-LHCC-96-042, 1996, URL: https://cds.cern.ch/record/331062 (cit. on p. 30).

[135] J. Pequenao, "Computer Generated image of the ATLAS calorimeter", 2008, URL: https://cds.cern.ch/record/1095927 (cit. on p. 31).

[136] ATLAS Collaboration, *ATLAS LAr Calorimeter Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-027; CERN-LHCC-2017-018, 2017, URL: https://cds.cern.ch/record/2285582 (cit. on p. 30).

[137] ATLAS Collaboration, *ATLAS Tile Calorimeter Phase-II Upgrade: Technical Design Report*, ATLAS-TDR-028; CERN-LHCC-2017-019, 2017, URL: https://cds.cern.ch/record/2285583 (cit. on p. 31).

[138] ATLAS Collaboration, *ATLAS Muon Spectrometer: Technical Design Report*, ATLAS-TDR-10; CERN-LHCC-97-022, CERN, 1997, URL: https://cds.cern.ch/record/331068 (cit. on p. 31).

[139] J. Pequenao, "Computer generated image of the ATLAS Muons subsystem", 2008, URL: https://cds.cern.ch/record/1095929 (cit. on p. 32).

[140] ATLAS Collaboration, *ATLAS Magnet System: Magnet Project Technical Design Report, Volume 1*, ATLAS-TDR-6; CERN-LHCC-97-018, 1997, URL: https://cds.cern.ch/record/338080 (cit. on p. 33).

[141] ATLAS Collaboration, *ATLAS TDAQ System Phase-I Upgrade: Technical Design Report*, ATLAS-TDR-023; CERN-LHCC-2013-018, 2013, URL: https://cds.cern.ch/record/1602235 (cit. on p. 33).

[142] ATLAS Collaboration, *Public TDAQ Results for Run-2 of the LHC*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ApprovedPlotsDAQ#Notes (cit. on p. 34).

[143] GEANT4 Collaboration, S. Agostinelli et al., GEANT4 – *a simulation toolkit*, Nucl. Instrum. Meth. A **506** (2003) 250 (cit. on pp. 33, 75).

[144] ATLAS Collaboration, *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*, ATL-PHYS-PUB-2010-013, 2010, URL: https://cds.cern.ch/record/1300517 (cit. on p. 33).

[145] ATLAS Collaboration, *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, Eur. Phys. J. C **77** (2017) 673, arXiv: 1704.07983 [hep-ex] (cit. on pp. 35, 36).

[146] T. Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, J. Phys. Conf. Ser. **119** (2008) 032014, ed. by R. Sobie, R. Tafirout and J. Thomson (cit. on p. 36).

[147] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: https://cds.cern.ch/record/2037717 (cit. on p. 36).

[148] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: 1603.02934 [hep-ex] (cit. on pp. 37, 49).

[149] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, URL: https://cds.cern.ch/record/1099735 (cit. on p. 37).

[150] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79** (2019) 639, arXiv: 1902.04655 [hep-ex] (cit. on p. 38).

[151] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton–proton collision data*, JINST **14** (2019) P12006, arXiv: 1908.00005 [hep-ex] (cit. on pp. 38, 40).

[152] ATLAS Collaboration, *Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **81** (2021) 578, arXiv: 2012.00578 [hep-ex] (cit. on pp. 39–41).

[153] J. E. Huth et al., "Toward a standardization of jet definitions", *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90)*, 1990 0134 (cit. on p. 41).

[154] G. F. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, Phys. Rev. Lett. **39** (1977) 1436 (cit. on pp. 41, 42).

[155] G. P. Salam, *Towards Jetography*, Eur. Phys. J. C **67** (2010) 637, arXiv: 0906.1833 [hep-ph] (cit. on p. 42).

[156] UA1 Collaboration, *Physics results of the UA1 collaboration at the CERN proton-antiproton collider*, eConf **C840723** (1984) 634, ed. by A. Zichichi (cit. on p. 42).

[157] G. P. Salam and G. Soyez, *A Practical Seedless Infrared-Safe Cone jet algorithm*, JHEP **05** (2007) 086, arXiv: 0704.0292 [hep-ph] (cit. on p. 42).

[158] M. Cacciari, G. P. Salam and G. Soyez, *FastJet user manual*, Eur. Phys. J. C **72** (2012) 1896, arXiv: 1111.6097 [hep-ph] (cit. on p. 42).

[159] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. D **48** (1993) 3160, arXiv: hep-ph/9305266 (cit. on p. 43).

[160] Y. L. Dokshitzer et al., *Better jet clustering algorithms*, JHEP **08** (1997) 001, arXiv: hep-ph/9707323 (cit. on p. 43).

[161] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 [hep-ph] (cit. on pp. 43, 44).

[162] ATLAS Collaboration, *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** (2017) 072002, arXiv: 1703.09665 [hep-ex] (cit. on pp. 44–46, 84, 130).

[163] *Pile-up subtraction and suppression for jets in ATLAS*, tech. rep. ATLAS-CONF-2013-083, CERN, 2013, URL: https://cds.cern.ch/record/1570994 (cit. on p. 45).

[164] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: 1703.10485 [hep-ex] (cit. on p. 47).

[165] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Eur. Phys. J. C **81** (2020) 689, arXiv: 2007.02645 [hep-ex] (cit. on p. 48).

[166] ATLAS Collaboration, *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** (2019) 135, arXiv: 1807.09477 [hep-ex] (cit. on pp. 49–52, 54).

[167] D. Krohn, J. Thaler and L.-T. Wang, *Jet Trimming*, JHEP **02** (2010) 084, arXiv: 0912.1342 [hep-ph] (cit. on p. 49).

[168] ATLAS Collaboration, *Jet energy measurement with the ATLAS detector in proton–proton collisions at $\sqrt{s} = 7$ TeV*, Eur. Phys. J. C **73** (2013) 2304, arXiv: 1112.6426 [hep-ex] (cit. on p. 51).

[169] ATLAS Collaboration, *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** (2016) 581, arXiv: 1510.03823 [hep-ex] (cit. on p. 52).

[170] ATLAS Collaboration, *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** (2019) 970, arXiv: 1907.05120 [hep-ex] (cit. on p. 52).

[171] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATL-PHYS-PUB-2017-013, 2017, URL: https://cds.cern.ch/record/2273281 (cit. on pp. 52, 54).

[172] ATLAS Collaboration, *b-tagging in dense environments*, ATL-PHYS-PUB-2014-014, 2014, URL: https://cds.cern.ch/record/1750682 (cit. on pp. 52, 53).

[173] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, URL: https://cds.cern.ch/record/2160731 (cit. on p. 53).

[174] ATLAS Collaboration, *Measurement of b-tagging efficiency of c-jets in $t\bar{t}$ events using a likelihood approach with the ATLAS detector*, ATLAS-CONF-2018-001, 2018, URL: https://cds.cern.ch/record/2306649 (cit. on p. 54).

[175] ATLAS Collaboration, *Measurement of the c-jet mistagging efficiency in $t\bar{t}$ events using pp collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS detector*, Eur. Phys. J. C **82** (2022) 95, arXiv: 2109.10627 [hep-ex] (cit. on p. 54).

[176] ATLAS Collaboration, *Boosted hadronic vector boson and top quark tagging with ATLAS using Run 2 data*, ATL-PHYS-PUB-2020-017, 2020, URL: https://cds.cern.ch/record/2724149 (cit. on pp. 54–56).

[177] A. J. Larkoski, I. Moult and D. Neill, *Power Counting to Better Jet Observables*, JHEP **12** (2014) 009, arXiv: 1409.6298 [hep-ph] (cit. on p. 54).

[178] M. Cacciari, G. P. Salam and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv: 0802.1188 [hep-ph] (cit. on p. 54).

[179] ATLAS Collaboration,
Performance of top-quark and W-boson tagging with ATLAS in Run 2 of the LHC,
Eur. Phys. J. C **79** (2019) 375, arXiv: 1808.07858 [hep-ex] (cit. on p. 54).

[180] ATLAS Collaboration,
E_T^{miss} performance in the ATLAS detector using 2015–2016 LHC pp collisions,
ATLAS-CONF-2018-023, 2018, URL: https://cds.cern.ch/record/2625233
(cit. on p. 55).

[181] G. D'Agostini,
Bayesian inference in processing experimental data: principles and basic applications,
Reports on Progress in Physics () (cit. on p. 57).

[182] R. D. Cousins, *Why isn't every physicist a Bayesian?*, Am. J. Phys. **63** (1995) 398
(cit. on p. 57).

[183] R. J. Barlow, *Statistics: a guide to the use of statistical methods in the physical sciences*,
Manchester physics series, Wiley, 1989,
URL: https://cds.cern.ch/record/213033 (cit. on pp. 58, 61).

[184] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*,
Comput. Phys. Commun. **77** (1993) 219 (cit. on p. 59).

[185] J. Conway,
"Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra",
PHYSTAT 2011, 2011 115, arXiv: 1103.0354 [physics.data-an] (cit. on p. 60).

[186] K. Cranmer et al.,
HistFactory: A tool for creating statistical models for use with RooFit and RooStats,
tech. rep. CERN-OPEN-2012-016, New York U., 2012,
URL: https://cds.cern.ch/record/1456844 (cit. on pp. 60, 61).

[187] J. Neyman, E. S. Pearson and K. Pearson,
On the problem of the most efficient tests of statistical hypotheses,
Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** (1933) 289 (cit. on p. 62).

[188] K. Cranmer, "Practical Statistics for the LHC",
2011 European School of High-Energy Physics, 2014 267,
arXiv: 1503.07622 [physics.data-an] (cit. on p. 63).

[189] J. Neyman,
Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability,
Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences **236** (1937) 333, ISSN: 00804614 (cit. on p. 63).

[190] A. L. Read, *Modified frequentist analysis of search results (the CL_s method)*, (2000),
URL: https://cds.cern.ch/record/451614 (cit. on p. 64).

[191] G. Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, Eur. Phys. J. C **71** (2011) 1554, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], arXiv: 1007.1727 [physics.data-an] (cit. on pp. 65, 69).

[192] S. S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, The Annals of Mathematical Statistics **9** (1938) 60, ISSN: 00034851 (cit. on p. 65).

[193] *TRExFitter*, https://trexfitter-docs.web.cern.ch/trexfitter-docs/ (cit. on p. 65).

[194] L. Moneta et al., *The RooStats Project*, PoS **ACAT2010** (2010) 057, ed. by T. Speer et al., arXiv: 1009.1003 [physics.data-an] (cit. on p. 65).

[195] D. Biedermann, *Search for Single Production of Vector-like Quarks Decaying into Wb in pp Collisions at sqrt(s)=13 TeV with the ATLAS Detector*, PhD thesis: Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät, 2020 (cit. on pp. 73, 77, 78, 86, 87).

[196] ATLAS Collaboration, *Performance of the Fast ATLAS Tracking Simulation (FATRAS) and the ATLAS Fast Calorimeter Simulation (FastCaloSim) with single particles*, ATL-SOFT-PUB-2014-001, 2014, URL: https://cds.cern.ch/record/1669341 (cit. on p. 75).

[197] T. Sjöstrand, S. Mrenna and P. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on pp. 75, 113).

[198] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on p. 75).

[199] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP **02** (2009) 007, arXiv: 0811.4622 [hep-ph] (cit. on pp. 75, 113).

[200] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on pp. 76, 113).

[201] S. Mrenna and P. Skands, *Automated parton-shower variations in PYTHIA 8*, Phys. Rev. D **94** (2016) 074005, arXiv: 1605.08352 [hep-ph] (cit. on p. 76).

[202] O. Matsedonskyi, G. Panico and A. Wulzer, *On the interpretation of Top Partners searches*, Journal of High Energy Physics **2014** (2014), URL: https://doi.org/10.1007%2Fjhep12%282014%29097 (cit. on p. 77).

[203] ATLAS Collaboration, *Selection of jets produced in* 13 *TeV proton–proton collisions with the ATLAS detector*, ATLAS-CONF-2015-029, 2015, URL: https://cds.cern.ch/record/2037702 (cit. on pp. 77, 115).

[204] ATLAS Collaboration, *2015 start-up trigger menu and initial performance assessment of the ATLAS trigger using Run-2 data*, ATL-DAQ-PUB-2016-001, 2016, URL: https://cds.cern.ch/record/2136007 (cit. on pp. 77, 115).

[205] ATLAS Collaboration, *Trigger Menu in 2016*, ATL-DAQ-PUB-2017-001, 2017, URL: https://cds.cern.ch/record/2242069 (cit. on pp. 77, 115).

[206] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017. 33 p, URL: http://cds.cern.ch/record/2633501 (cit. on pp. 83, 129).

[207] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv: 1608.03953 [hep-ex] (cit. on p. 83).

[208] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton–proton collision data*, JINST **14** (2019) P03017, arXiv: 1812.03848 [hep-ex] (cit. on p. 84).

[209] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-015, 2015, URL: https://cds.cern.ch/record/2037613 (cit. on p. 84).

[210] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43** (2016) 023001, arXiv: 1510.03865 [hep-ph] (cit. on p. 85).

[211] ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs using $139\,fb^{-1}$ of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **03** (2020) 145, arXiv: 1910.08447 [hep-ex] (cit. on p. 113).

[212] ATLAS Collaboration, *Search for Low-Mass Dijet Resonances Using Trigger-Level Jets with the ATLAS Detector in pp Collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **121** (2018) 081801, arXiv: 1804.03496 [hep-ex] (cit. on p. 113).

[213] ATLAS Collaboration, *Measurement of inclusive jet and dijet cross-sections in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **05** (2018) 195, arXiv: 1711.02692 [hep-ex] (cit. on p. 113).

[214] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv: 1512.01178 [hep-ph] (cit. on p. 113).

[215] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 113).

[216] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 113).

[217] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 113).

[218] R. D. Ball et al., *Parton distributions for the LHC run II*, JHEP **04** (2015) 040, arXiv: 1410.8849 [hep-ph] (cit. on p. 113).

[219] M. Cacciari et al., *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. B **710** (2012) 612, arXiv: 1111.5869 [hep-ph] (cit. on p. 113).

[220] M. Beneke et al., *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl. Phys. B **855** (2012) 695, arXiv: 1109.1536 [hep-ph] (cit. on p. 113).

[221] P. Bärnreuther, M. Czakon and A. Mitov, *Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001, arXiv: 1204.5201 [hep-ph] (cit. on p. 113).

[222] M. Czakon and A. Mitov, *NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels*, JHEP **12** (2012) 054, arXiv: 1207.0236 [hep-ph] (cit. on p. 113).

[223] M. Czakon and A. Mitov, *NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction*, JHEP **01** (2013) 080, arXiv: 1210.6832 [hep-ph] (cit. on p. 113).

[224] M. Czakon, P. Fiedler and A. Mitov, *Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$*, Phys. Rev. Lett. **110** (2013) 252004, arXiv: 1303.6254 [hep-ph] (cit. on p. 113).

[225] M. Czakon and A. Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv: 1112.5675 [hep-ph] (cit. on p. 113).

[226] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: https://cds.cern.ch/record/1966419 (cit. on p. 113).

[227] J. Kuehn, A. Scharf and P. Uwer, *Weak interactions in top-quark pair production at hadron colliders: An update*, Phys. Rev. D **91** (2015) 014020, arXiv: 1305.5773 [hep-ph] (cit. on p. 113).

[228] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling with Sherpa and MG5_aMC@NLO*, ATL-PHYS-PUB-2017-007, 2017, URL: https://cds.cern.ch/record/2261938 (cit. on p. 113).

[229] S. Frixione et al., *Single-top production in MC@NLO*, JHEP **03** (2006) 092, arXiv: hep-ph/0512250 (cit. on p. 113).

[230] S. Frixione et al., *Single-top hadroproduction in association with a W boson*, JHEP **07** (2008) 029, arXiv: 0805.3067 [hep-ph] (cit. on p. 113).

[231] C. D. White, S. Frixione, E. Laenen and F. Maltoni, *Isolating Wt production at the LHC*, JHEP **11** (2009) 074, arXiv: 0908.0631 [hep-ph] (cit. on p. 113).

[232] S. Schumann and F. Krauss, *A parton shower algorithm based on Catani–Seymour dipole factorisation*, JHEP **03** (2008) 038, arXiv: 0709.1027 [hep-ph] (cit. on p. 113).

[233] S. Höche et al., *QCD matrix elements + parton showers. The NLO case*, JHEP **04** (2013) 027, arXiv: 1207.5030 [hep-ph] (cit. on p. 113).

[234] G. Cacciapaglia et al., *Next-to-leading-order predictions for single vector-like quark production at the LHC*, Phys. Lett. B **793** (2019) 206, arXiv: 1811.05055 [hep-ph] (cit. on pp. 114, 141, 147).

[235] ATLAS Collaboration, *Trigger Menu in 2017*, ATL-DAQ-PUB-2018-002, 2018, URL: https://cds.cern.ch/record/2625986 (cit. on p. 115).

[236] ATLAS Collaboration, *Trigger Menu in 2018*, ATL-DAQ-PUB-2019-001, 2019, URL: https://cds.cern.ch/record/2693402 (cit. on p. 115).

[237] A. Loginov, *Strategies of data-driven estimations of ttbar backgrounds in ATLAS*, tech. rep., CERN, 2010, URL: https://cds.cern.ch/record/1287126 (cit. on p. 120).

[238] M. R. Rahman, *Background estimation in the search for single production of vector-like quarks decaying into Wb in pp collisions using a data-driven method at at $\sqrt{s} = 13$ TeV with the ATLAS detector*, 2020 (cit. on p. 120).

[239] ATLAS Collaboration, *Search for $W' \to tb$ decays in the hadronic final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Phys. Lett. B **781** (2018) 327, arXiv: 1801.07893 [hep-ex] (cit. on p. 120).

[240] ATLAS Collaboration, *Search for $t\bar{t}$ resonances in fully hadronic final states in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **10** (2020) 061, arXiv: 2005.05138 [hep-ex] (cit. on p. 120).

[241] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*, (2022), arXiv: 2212.09379 [hep-ex] (cit. on p. 129).

[242] ATLAS Collaboration, *Measurement of the ATLAS Detector Jet Mass Response using Forward Folding with 80 fb^{-1} of $\sqrt{s} = 13$ TeV pp data*, ATLAS-CONF-2020-022, 2020, URL: https://cds.cern.ch/record/2724442 (cit. on p. 130).

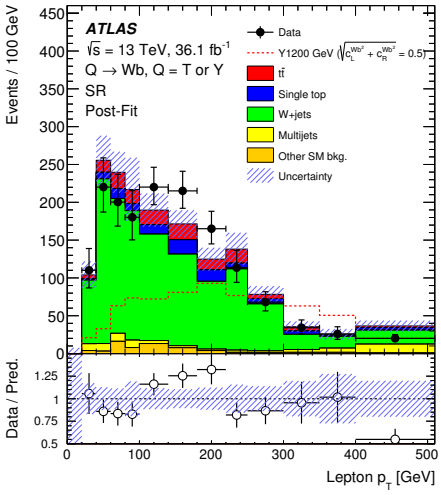
[243] ATLAS Collaboration, *Measurement of the Lund Jet Plane Using Charged Particles in 13 TeV Proton–Proton Collisions with the ATLAS Detector*, Phys. Rev. Lett. **124** (2020) 222002, arXiv: 2004.03540 [hep-ex] (cit. on p. 131).

[244] A. Roy et al.,
Novel interpretation strategy for searches of singly produced vectorlike quarks at the LHC,
Phys. Rev. D **101** (2020) 115027, arXiv: 2003.00640 [hep-ph] (cit. on p. 142).

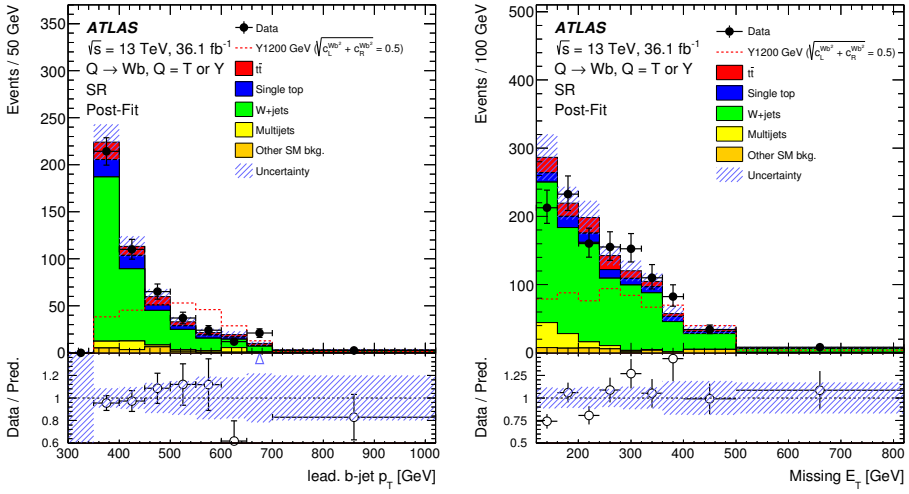
APPENDIX A

Single-lepton analysis post-fit kinematic variables

In this section the m_{VLQ} distributions of various kinematic variables after the fit to data in SR and CRs for the background-only hypothesis is shown.



(a)



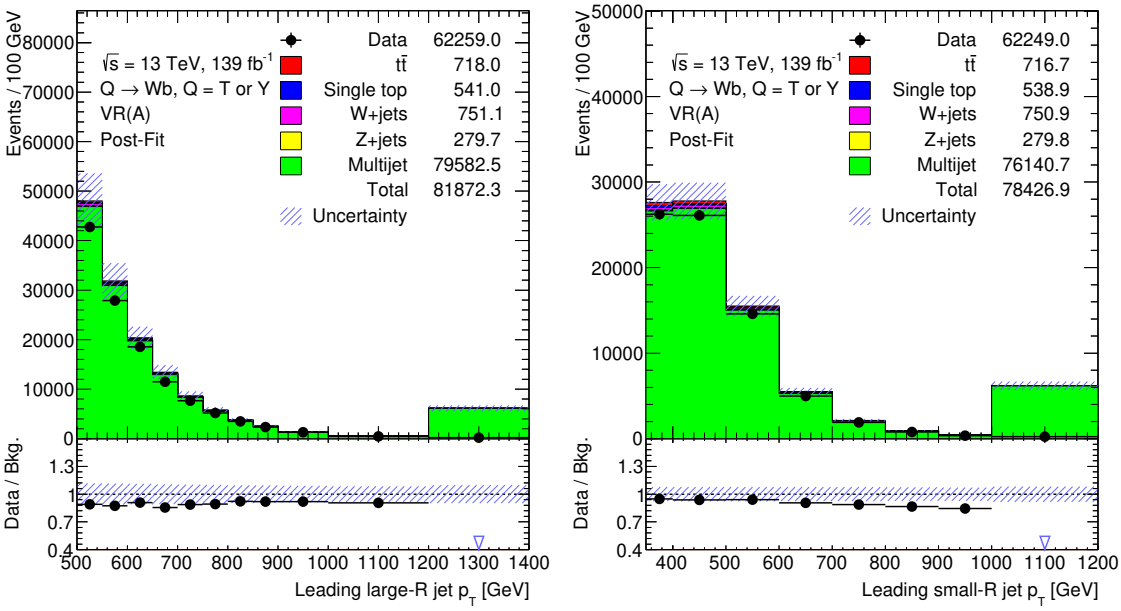
(b) (c)

Figure A.1: Distributions of (a) lepton p_{T}, (b) leading b-tagged jet p_{T}, and (c) $E_{\mathrm{T}}^{\mathrm{miss}}$ in the SR for data and the SM background processes with their post-fit normalisations. The uncertainty band includes statistical and all systematic uncertainties. The red dashed line shows a Y signal with a mass of 1 200 GeV and a coupling of $\sqrt{(c_{\mathrm{L}}^{Wb})^2 + (c_{\mathrm{R}}^{Wb})^2} \approx 0.5$ The first and last bin include the underflow and overflow, respectively.

APPENDIX B

Hadronic analysis post-fit kinematic variables

In this section the m_{VLQ} distributions of various kinematic variables after the fit to data in SR and CRs for the background-only hypothesis is shown. The multijet background estimate used as the input for these fits are calculated separately for the given variables e.g. leading large-R jet p_{T} or leading small-R jet p_{T}. The ABCD method does not work well for these variables at high p_{T} and thus we observe a disagreemnt. This happens due to low statistics in the high p_{T} regions.



(a) (b)

Figure B.1: Distributions of (a) leading large-R jet p_{T} and **??** leading small-R jet p_{T} in the SR for data and the SM background processes with their post-fit normalisations. The uncertainty band includes statistical and all systematic uncertainties.

List of Figures

List of Tables